Київський національний університет імені Тараса Шевченка

Журнал обчислювальної та прикладної

2014 **М А Т Е М А Т И К И** 2 (116)

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 16 червня 2014 р. (протокол № 11)

Свідоцтво про державну реєстрацію КВ 4246 від 26.05.2000

Підписано до друку 16 червня 2014 р.

Журнал обчислювальної та прикладної математики | **2014** **№ 2 (116)** | Journal of Computational & Applied Mathematics

UDC 519.6

MODIFIED NEWTON METHOD FOR ANTENNA POWER SYNTHESIS PROBLEM WITH FIXED NORM OF THE PATTERN

Mykhaylo Andriychuk, Mykola Voytovych

Резюме. Постановка задачі синтезу антен за потужністю доповнена суттєвою фізичною умовою рівності норм заданої і одержаної діаграм. За допомогою методу Лагранжа задача зведена до безумовної мінімізації з невідомим параметром (множником Лагранжа), який відповідає за виконання згаданої умови. Рівнянням Ейлера для цієї задачі є нелінійне інтенальне рівняння типу Гаммерштейна з кубічною залежністю підінтегральної функції від модуля шуканої функції і лінійної залежністю від її аргумента (фазового множника). Рівняння розв'язується модифікованим методом Ньютона. Числові результати продемонстровані на прикладі лінійної антени, яка описується інтегральним перетворенням Фур'є фінітної функції. Виявлено і проаналізовано чисельно процес галуження розв'язків задачі.

Abstract. The problem formulation of the antenna synthesis according to the prescribed power radiation pattern is generalized by taking into account the important physical restriction on the norm of the synthesized pattern. By the Lagrange method, the problem is reduced to an unconditional variational problem with unknown parameter (Lagrange multiplier), which provides the condition of the norm equality. The Lagrange–Euler equation for this problem is a nonlinear integral equation of the Hammerstein type with cubic dependence of the integrand on modulus of the unknown function. The argument (phase) factor of this function is involved in the integrand linearly. The modified Newton method is used to solve this nonlinear equation. Numerical results are demonstrated on the example of the linear antenna. The solution branching is observed numerically and analyzed.

1. Introduction

The antenna power synthesis problem [1] belongs to the phase optimization problems in which the argument (phase) of a desired complex function is not the given function, but it is an additional parameter to be optimized. In contrast to the more usual synthesis problem according to the prescribed amplitude pattern [2], the main term in the functional of the power problem to be minimized, is the mean square difference between not the modulus (amplitudes) of obtained and desired radiation patterns, but between their squared values (powers). Increasing the algebraic degree of the unknown function leads to the higher nonlinearity of the problem and causes new theoretical and computational complications.

Key words. Antenna power synthesis problem, Nonlinear integral equation, Modified Newton method, Branching of solutions

The possibility to provide the local irregularities in the desired pattern more precisely than in the synthesis problem by the given amplitude radiation pattern is one of the advantages of the power synthesis.

In its earlier formulation the power synthesis had some disadvantage connected with absence of any restriction onto the norm of radiation pattern. In this case the radiation pattern is involved in all terms of nonlinear integral equation as an linear multiplier, what admits, in particular, the existence of zero solution. Moreover, it is turned out that this solution is unique at certain combinations of parameters.

To avoid the above disadvantage, and to take into account some physical requirements on the synthesized radiation pattern, we supplement the functional to be optimized by additional condition describing the norm equality of the desired and obtained radiation patterns. Some modifications of this condition were used in [3]–[5] in other formulations of the antenna synthesis problems. Such conditional minimization problem can be reduced to the unconditional one by the Lagrange multipliers method.

The Lagrange-Euler equation for resulting functional is an nonlinear integral equation of the Hammerstein type. It contains the unknown function in the integrand as a cubic algebraic term.

The equation is numerically solved by the modified Newton method. Since the problem has nonunique solutions, they can be separated only by the appropriate choice of the initial approximations. Different types of solutions were found and analyzed. Their branching points were observed numerically as well. Of course, such an approach does not investigate the branching process completely. This question is a subject of special studies. For this purpose the approach based on the complex polynomial presentation of the solutions [6] can be appied.

Some results of this paper were announced in [7].

2. Problem formulation

The current u on the antenna and radiation pattern f generated by it, are connected by the relation

$$f = Au, \tag{1}$$

where A is a linear bounded operator. The antenna synthesis problem according to the prescribed power radiation pattern F^2 consists in minimization of the functional [1]

$$\sigma_\alpha(u) = ||F^2 - |f|^2||_2^2 + \alpha||u||_1^2, \tag{2}$$

where $\|\cdot\|_1$, $\|\cdot\|_2$ are the mean square norms in the spaces of the currents and radiation patterns, respectively, $\alpha > 0$ is a given positive coefficient (weight factor); further we assume $\|F\|_2^2 = 1$. We supplement this functional by the condition

$$||f||_2^2 = 1. \tag{3}$$

Using the Lagrange multipliers method, we reduce the problem (2)-(3) to minimization of the functional

$$\sigma_{\alpha,\mu}(u) = ||F^2 - |f|^2||_2^2 + \alpha||u||_1^2 - \mu||f||_2^2 \tag{4}$$

with undefined coefficient (Lagrange multiplier) μ. Fixing μ, we denote by u_μ and f_μ the current u minimizing $\sigma_{\alpha,\mu}(u)$ and radiation pattern f generated by it, respectively. Then the condition (3) may be considered as the transcendental equation for determining μ. Another way to solve the problem is to find u, f and μ simultaneously.

The Lagrange-Euler equation for the functional (4) can be written in the form

$$\alpha f - 2AA^*[(F^2 - |f|^2)f] - \mu AA^* f = 0. \tag{5}$$

Here A^* is the operator adjoint to A. After f and μ is found from (5), (3), the desired field distribution u is calculated as

$$u = \frac{1}{\alpha}\left(2A^*[(F^2 - |f|^2)f] + \mu A^* f\right). \tag{6}$$

Equation (3) may be supplemented to (5), and they together may be considered as the equation system for determining f and μ. The modified Newton method described in [9] in the context of similar systems of equation, can be applied to system (5), (3). In order to use it, we convert equations (5), (3) to the convenient form

$$\Phi(f,\mu) \equiv \alpha f - 2AA^*[(F^2 - |f|^2)f] - \mu AA^* f = 0, \tag{7}$$

$$\Psi(f) \equiv ||f||^2 - 1 = 0. \tag{8}$$

The next approximation to the unknown f and μ is calculated in the used method as

$$f_{p+1} = f_p + \delta f'_p + i\delta f''_p, \tag{9}$$

$$\mu_{p+1} = \mu_p + \delta\mu_p, \tag{10}$$

where $\delta f'_p$, $\delta f''_p$, $\delta\mu_p$ are found from the linear equation system

$$\begin{bmatrix} [\alpha - \mu_p AA^* - 2AA^*(F^2 \\ -|f_p|^2) + 4AA^*(f'_p)] & 4AA^*(f'_0 f''_0) & -AA^*(f'_0) \\ 4AA^*(f'_p f''_p) & [\alpha - \mu_p AA^* - 2AA^*(F^2- \\ |f_p|^2) + 4AA^*(f''_p)] & -AA^*(f''_0) \\ 2f'_p & -2f''_p & 0 \end{bmatrix} \times$$

$$\times \begin{bmatrix} \delta f'_p \\ \delta f''_p \\ \delta\mu_p \end{bmatrix} = \begin{bmatrix} -\Phi'_p \\ -\Phi''_p \\ -\Psi_p \end{bmatrix}. \tag{11}$$

In the case when the parameter μ is fixed, equation (8) does not participate in the system, then the last row in system (11), as well as the last column in its matrix are omitted.

3. Numerical results

The proposed approach has been tested on the example of the synthesis problem for the linear antenna of limited length, which is described by the Fourier transform operator mapping on the compactly bounded functions. The desired power pattern F^2 is assumed to be given also as a compactly bounded function. In this case the operators A, A^*, and the kernel $K(\xi_1, \xi_2)$ of the operator AA^* have the forms

$$f(\xi) = Au \equiv \int\limits_{-1}^{1} u(x) e^{icx\xi} dx, \tag{12}$$

$$A^* g = \frac{c}{2\pi} \int\limits_{-1}^{1} g(\xi) e^{-icx\xi} d\xi, \tag{13}$$

$$K(\xi, \xi') = \frac{\sin c(\xi - \xi\prime)}{\pi(\xi - \xi\prime)}, \tag{14}$$

where x is the normalized coordinate on the antenna, $\xi = \sin\vartheta / \sin\vartheta_0$ is the generalized angular coordinate in the far zone, $2\vartheta_0$ is the angle where the prescribed power pattern F^2 differs from zero, $c = ka\sin\vartheta_0$ is the characteristic physical parameter, $2a$ is the antenna length.

The Lagrange-Euler equation for the functional (4) for this example has the form

$$\alpha f(\xi') - \frac{2}{\pi} \int_{-1}^{1} \frac{\sin c(\xi - \xi\prime)}{\pi(\xi - \xi\prime)} [(F^2(\xi) - |f(\xi)|^2) f(\xi)] d\xi - \\ - \mu \int_{-1}^{1} \frac{\sin c(\xi - \xi\prime)}{\pi(\xi - \xi\prime)} f(\xi) d\xi = 0. \tag{15}$$

The numerical results are presented for the prescribed power patterns $F^2(\xi) \equiv 1/2$ and $F^2(\xi) \equiv \cos(\pi x/2)$, $|x| \leq 1$; for $|x| > 1$ these functions equal zero.

The main result of the optimization is the mean-square deviation $\sigma_0 = ||F^2 - |f|^2||_2^2$ of the power patterns (the first term in functional (4)), two other terms have the auxiliary sense. Dependencies of σ_0 on the parameter c are shown for these two prescribed patterns in Fig.1, respectively, for different solutions of equation (15). The results depend essentially on the parameter α (weight factor in functional (4)); its values are given in the figures. The solutions branch at some values of c; these values are denoted by c_{n1} (the index n values relate to different α).

The real solutions to (15) exist for two given $F^2(\xi)$ and all values of c (dashed lines in the figures). Different behaviour of σ_0 for different F^2 at small c is

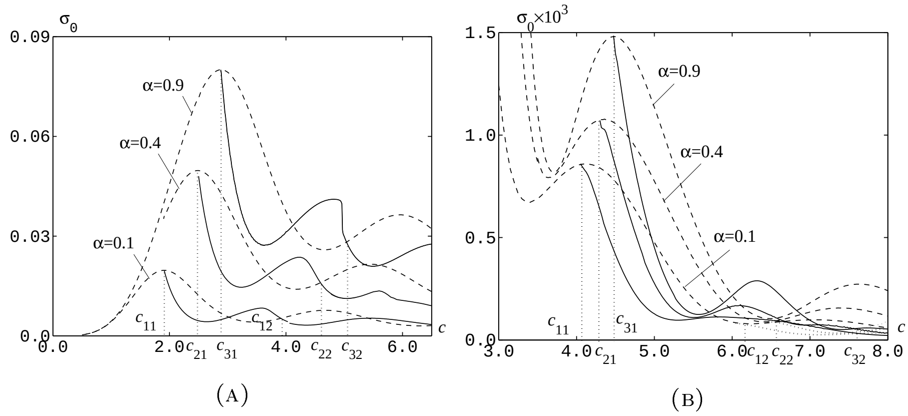


Fig. 1. The mean-square deviation of power patterns for different solutions to (15); (A) $F^2(\xi) \equiv 0.5$; (B) $F^2(\xi) = \cos^2(\pi\xi/2)$

explained by the fact, that the real solution to (15) is asymptotically constant at $c \to 0$. Therefore, the function $F^2(\xi) \equiv \text{const}$ can be easily approximated at small c. This property is inherent only to this given F^2.

Note, that at $\mu = 0$ and fixed α equation (15) has only zero solution at small c. This is explained by the fact that the second term in functional (4) is dominant at small c.

At the points $c = c_{n1}$ the complex solutions with odd phase functions $\arg f(-\xi) = -\arg f(\xi)$ (solid lines) branch off from the real solutions. The branching points approximately coincide with the first maximums of σ_0 as a function of c for the real solution. It is easy to check from (6) that the current distribution on the antenna, which generates the pattern with odd $\arg f(\xi)$, is real, but it has zero points in the interval $x \in [-1, 1]$. This fact is important from the practical point of view, because in this case no phase transformer device is needed for its creation.

The next characteristic points in Fig.1 are the points $c = c_{n2}$ where two new complex solutions simultaneously arise with odd and even phases, respectively. They have the same $|f(\xi|$ and hence the same $\sigma_0(u)$. However, the current $u(x)$ is different for these solutions. One of them, corresponding the the odd $\arg f(\xi)$, is real and has zero points in $x \in [-1, 1]$, whereas the second one, with the even phase ($\arg f(-\xi) = \arg f(\xi)$) is even complex function (in some cases it also can have zeros on the antenna). The solutions with odd phase branch off from one of the same type (that is, with the odd phase), whereas the solution with even phase branch off from the real one; both them arise at the same point c_{n2}. Consequently, at least four solutions exist at $c > c_{n2}$: one real (that is, with zero phase), one with even phase, and two with odd phases. In Fig. 1 the results only for one solution with odd phase are presented.

Note that the evenness of the phase distributions $\arg f(\xi)$ and $\arg u(x)$ is caused by the symmetry of the given function $F^2(\xi)$ and both intervals $x \in [-1, 1]$ and $\xi \in [-1, 1]$.

As it is investigated so far, the solution behavior for the considered problem is qualitatively similar to that for the synthesis problem according to the amplitude radiation pattern (the rigorous solutions to this problem see in [8], [9]). However, this analogy can be not complete: the problem considered here has nonlinearity of the higher degree and can have additional solutions different in the behavior from those of the mentioned problem.

At fixed c the current norm almost does not depend on α for the solutions of all types. This is caused by the fact that this norm is hardly affected by the radiation pattern norm, which is fixed in our formulation.

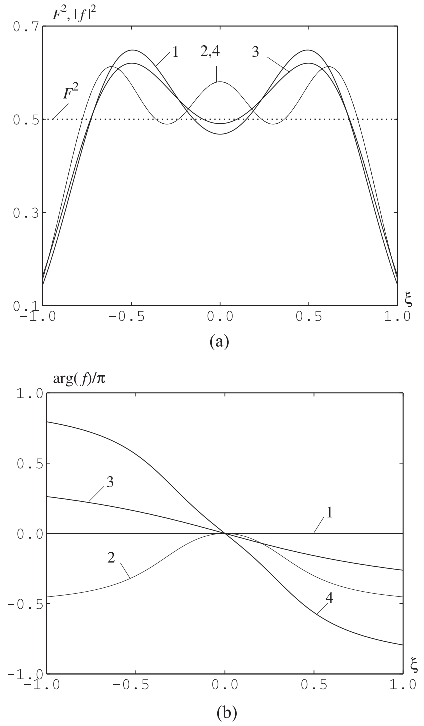


FIG. 2. Power (a) and phase (b) obtained patterns; $F^2(\xi) \equiv 0.5$; $\alpha = 0.4$; $c = 5$

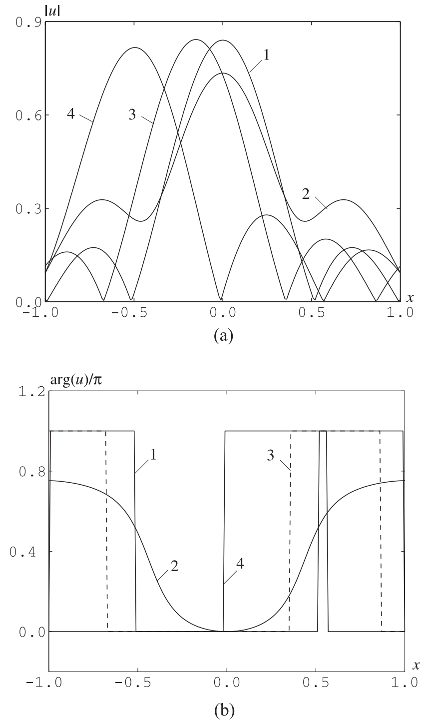


FIG. 3. Amplitude (a) and phase (b) distributions of the currents; $F^2(\xi) \equiv 0.5$; $\alpha = 0.4$; $c = 5$

The optimal directivity patterns $f(\xi)$ and the currents $u(x)$ which create them, corresponding to the solutions of different type for the desired function $F^2(\xi) \equiv$ const, are presented in Figs. 2, 3; the parameters are shown in the captions. The curves are labeled as follows: (1) – real solution; (2) – first solution with odd phase; (3) – second solution with odd phase; (4) – solution with even phase. Analogous results for the case $F^2(\xi) \equiv \cos(\pi x/2)$ are shown in Figs. 4, 5. In this case the amplitude of power pattern in all solutions almost coincides with the desired one.

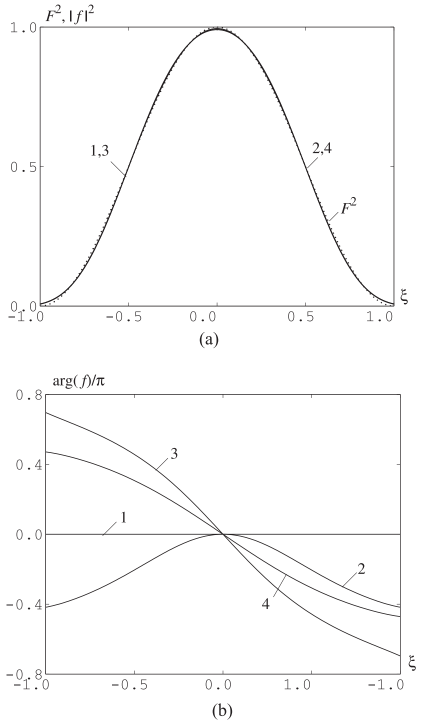


FIG. 4. Power (a) and phase (b) obtained patterns; $F^2(\xi) = \cos^2(\pi\xi/2)$; $\alpha = 0.9$; $c = 7$

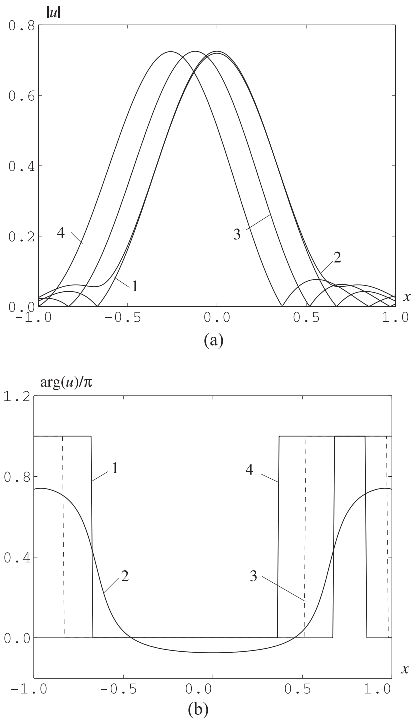


FIG. 5. Amplitude (a) and phase (b) distributions of the currents; $F^2(\xi) = \cos^2(\pi\xi/2)$; $\alpha = 0.9$; $c = 7$

Bibliography

1. Savenko P. O. Nonlinear Problems of Radiating Systems / P. O. Savenko – Lviv: IAPMM NASU, 2002 (In Ukrainian).
2. Andriychuk M. I. Synthesis of Antennas according to Amplitude Radiation Pattern: Numerical Methods and Algorithms / M. I. Andriychuk, N. N. Voitovich, P. A. Savenko, V. P. Tkachuk. – Kiev: Naukova Dumka, 1993 (In Russian).
3. Voitovich N. N. Antenna design for given pattern of amplitude of radiation (Semenov's method) / N. N. Voitovich // Radio Eng. and Electron. Phys. – 1972. – Vol. 17., No 12. – P. 2000-2005.
4. Bakhrakh L. D. Synthesis of Radiating Systems (Theory and Methods of Design) / L. D. Bakhrakh, S. D. Kremenetskiy – Moscow: Sov. Radio, 1974 (In Russian).
5. Katsenelenbaum B. Z. Optimal current synthesis on a circle / B. Z. Katsenelenbaum, M. Y. Shalukhin // J. Comm. Tech. Electron. – 1989. – Vol. 34., No 5. – P. 25-33.
6. Bulatsyk O. O. A complex polynomial representation of solution to the antenna synthesis problem by the power pattern / O. O. Bulatsyk // Direct and Inverse Problems in Electrodynamic and Acoustic Wave Theory (DIPED-2013). – Lviv. – 2013. – P. 166-170.
7. Andriychuk M. I. Antenna synthesis according to power radiation pattern with condition of norm equality / M. I. Andriychuk, N. N. Voitovich // Direct and Inverse Problems in Electrodynamic and Acoustic Wave Theory (DIPED-2013). – Lviv. – 2013. – P. 137-140.

8. Voitovich N. N. Antenna synthesis by amplitude radiation pattern and modified phase problem. Appendix in: B. Z. Katsenelenbaum. Electromagnetic Fields – Restrictions and Approximation / N. N. Voitovich. – Weinheim: WILEY-VCH. – 2003. – P. 191-234.
9. Bulatsyk O. O. Phase Optimization Problems / O. O. Bulatsyk, B. Z. Katsenelenbaum, Yu. P. Topolyuk, N. N. Voitovich. – Weinheim: WILEY-VCH, 2010.

Mykhaylo Andriychuk, Mykola Voytovych,
Pidstrygach Institute for Applide Problems of Mechanics and Mathematics, 3b Naukova Str., Lviv, 79060, Ukraine

Received 20.05.2014

Журнал обчислювальної та прикладної математики | 2014 № 2 (116) | Journal of Computational & Applied Mathematics

UDC 519.6

THE ALTERNATING METHOD APPLIED TO TWO-POINT BOUNDARY VALUE PROBLEMS

George Baravdish, B. Tomas Johansson

Резюме. Альтернуючий ітераційний метод Козлова-Мазьї, що був запропонований для обернених крайових задач для рівнянь в частинних похідних, застосовано до дво-точкової крайової задачі для звичайного диференціального рівняння другого порядку. Досліджено випадок лінійного диференціального оператора другого порядку. Зокрема, подано критерій збіжності як зв'язок між коефіцієнтами диференціального оператора і кінцевим моментом часу інтервалу. Для нелінійного диференціального оператора виведено деякі формули, за допомогою яких можна довести збіжність. Однак, як показали чисельні експерименти, знаходження критерію збіжності в нелінійному випадку є нетривіальною задачею.

Abstract. The alternating iterative method of Kozlov and Maz'ya, originally proposed for inverse boundary value problems for partial differential operators, is applied to a two-point boundary value problem for a second-order ordinary differential operator. The case of a linear second-order operator is investigated in detail. In particular, a criteria for convergence expressing a relationship between the coefficients of this operator and the final time of the interval is given. For nonlinear operators some formulas are derived on which a proof of convergence can be obtained. However, as is highlighted by a numerical example, finding criteria on the problem to guarantee convergence of the alternating method in the nonlinear case is nontrivial.

1. Introduction

The alternating iterative method was proposed in 1989 by V. A. Kozlov and V. G. Maz'ya [33] to solve some inverse ill-posed problems such as the Cauchy problem for a self-adjoint and strongly elliptic operator and data reconstruction for hyperbolic operators. An advantage with the alternating method is that one solves well-posed problems for the same type of governing partial differential operator in the solution domain as in the ill-posed problem and there is no parameter involved in the procedure. These properties have made the alternating method a popular choice in engineering applications and we give a brief survey on some of these results and applications before introducing the problem to be studied.

For general applications and implementation of the alternating method for Cauchy problems for time-independent operators (typically the Laplace operator), see [35, 23, 8, 16, 42, 40, 29, 24]. Relaxation to speed up the convergence has been introduced and examined in [29, 30, 25, 27]. Generalization of the alternating method to the Stokes system was undertaken in [7] and to

Key words. Heat equation, mixed problem, Rothe's method, boundary integral equation method, trigonometrical quadrature method.

the Helmholtz operator in [26], see also [10]. The alternating method for the Laplace equation was extended to unbounded domains in [13]. Convergence for some nonlinear operators was shown in [41, 4]. The various numerical implementations have mainly been performed using the boundary element method or integral equations, which is natural when only boundary data is updated. Implementation using the finite element method and error estimates suitable for adaptive methods were given in [5]. In that work it was also shown that the alternating method for elliptic problems can be interpreted as the minimization of a certain functional.

The aim of the present study is to show that the alternating method can be applied also to some two-point boundary value problems for a second-order operator. Specifically, we study

$$\begin{cases} u''(t) + f(t,u) = 0, \quad \text{in} \quad I, \\ u(0) = \varphi, \\ u(T) = \psi. \end{cases} \tag{1}$$

Here, $I = [0,T]$, where $T > 0$ is a real number, and $f : I \times X \to X$. We do not strive after the most general situation nor to have a method that can be compared with the many advanced numerical methods already presented in the literature for ordinary differential operators of the form (1). Instead, as pointed out above, we are interested solely in the alternating method and to add some more knowledge around this procedure, in particular, to give some classes of functions f for which the iterative method converges and to give some f for which there is no convergence. Thus, for simplicity, we concentrate on (1) when f is a continuous function, and where the space X is $\mathbb{R}^n$ or a Hilbert space; potentially X can be a Banach space. In fact, the main part of this study is devoted to the linear case when $f = Q(t)u$ with $Q(t) = A + B(t)$ being a smooth positive self-adjoint operator on X, and to show convergence of the alternating method in this case thereby generalizing the similar situation in [33] to time-dependent operators. One can of course have a higher order differential operator as well as different type of boundary conditions but we do not investigate that further here.

There are many applications leading to a model of the form (1), for example, deflection of cantilever beams under certain load [11], plate deflection theory [2], confinement of a plasma column using radiation pressure [47], heat transfer in fins [32], the study of tumour growth [1, 52], cell oxygen uptake [36, 39] and in modelling the distribution of heat sources in the human head [19, 44] to only mention a few.

Partly due to its many applications, there is an overwhelming literature on two-point boundary value problems and it is not within the scope of this study to give a general overview; instead below we point towards some references for (1) and within these the reader can find further references.

Existence and uniqueness of a solution to (1) is nontrivial. In the case $X = \mathbb{R}^n$, existence of a solution was settled in [48, 49]. For existence of a solution when X is a Banach space, see [12, 51, 43]. General references for second-order differential equations in Banach spaces are [18, Chapter 2], [21, Chapter 2 Section 7] and [46, Chapter 5 Section 3].

For general ideas on the numerical solution of (1), see [3, Chapter 11] and [31]. An excellent overview of both theoretical and numerical findings for (1), starting with the seminal paper of E. Picard [45], is given in the introduction in [14].

Let us then describe the method that we shall use to obtain a numerical approximation to (1). Following the original paper on the alternating method [33], the algorithm is:

(i) Make an initial guess η_0 of $u'(0)$. Then the first approximation u_0 is obtained by solving

$$\begin{cases} u_0''(t) + f(t, u_0) = 0, & \text{in } I, \\ u_0(0) = \varphi, \\ u_0'(0) = \eta_0. \end{cases} \tag{2}$$

(ii) Having obtained u_{2k}, the approximation u_{2k+1} for $k \geq 0$, is obtained by solving

$$\begin{cases} u_{2k+1}''(t) + f(t, u_{2k+1}) = 0, & \text{in } I, \\ u_{2k+1}(T) = \psi, \\ u_{2k+1}'(T) = u_{2k}'(T). \end{cases} \tag{3}$$

(iii) Then u_{2k+2} is obtained by solving

$$\begin{cases} u_{2k+2}''(t) + f(t, u_{2k+2}) = 0, & \text{in } I, \\ u_{2k+2}(0) = \psi, \\ u_{2k+2}'(0) = u_{2k+1}'(0). \end{cases} \tag{4}$$

The procedure then continues by iterating in the last two steps. Clearly, the initial value problems solved in each step are well-posed.

As mentioned above, we shall mainly concentrate on the linear case and in Section 2, we investigate the situation when $f(t, u) = Q(t)u$, with $Q(t) = A + B(t)$ being a self-adjoint linear smooth operator generating a (vector) sine and cosine function. Convergence of the alternating procedure is shown under a restriction on the final time T, see Theorem 2.2. We remark that the conditions on $Q(t)$ can be relaxed such that Q can be a differential operator on the space X, thus the results obtained can be applied to time-dependent hyperbolic problems as well. The results in Section 2 builds on Chapter 5 in [6], where the setting was $\mathbb{R}^n$.

To gain more insight and to be able to state a condition that is more easy to check for convergence of the alternating method, in Section 3 a linear and scalar equation is examined when $X = \mathbb{R}$ and $f(t, u) = q(t)u$. It is shown that provided that the smallest eigenvalue for some two-point boundary value problems in the interval I is greater than one then the method converges for $0 < T_1 < T$, see Theorem 3.4. In Section 3.1, we describe a class of functions f for which the alternating method diverges. In Section 4, we briefly investigate the nonlinear case. We derive some formulas for the iterates on which a proof of convergence can be based. However, this needs some monotonicity results for the solution and the function f. As is highlighted by a numerical example in Section 4.1, the alternating method can converge in the nonlinear case without the iterates being monotonically increasing (decreasing) towards the analytical solution. Thus, a full proof of the convergence in the nonlinear case seems intricate and beyond the scope of this study. In Section 4.1, we also suggest and briefly investigate a modification in the sense of linearization in the alternating procedure. This modification appears to converge for classes of functions where the original alternating method diverges. This merit further investigations of this linearization but it is not pursued here but deferred to future work.

2. The alternating procedure for second order linear equations

We start by first introducing some notation. The space $C(I;X)$ is the set of all continuous functions $v : I \to X$ and endowed with the usual supremum norm

$$\|v\|_\infty = \max_{0\leq t\leq T} |v(t)|.$$

Similarly, $C^k(I;X)$ is the space of k-times differentiable functions with the k-th derivative being continuous (supremum norm) and $k \geq 1$ an integer. The spectral radius of an operator Q is defined as usual,

$$r(Q) = \sup\{|\lambda|; \lambda \in \sigma(Q)\}.$$

We are interested in solving (1) in the case when $f(t,u) = Q(t)u$. We assume that

$$Q(t) = A + B(t), \tag{5}$$

where A is a linear operator generating a cosine function, i.e. a function $c(t)$ mapping into the space of bounded operators on X and satisfying $c(t+s) + c(t-s) = c(t)c(s)$, where $t, s \geq 0$, and $c(0) = I$, see further [18, Section 2] for criteria on A to guarantee existence of such a function $c(t)$. Moreover, $B(t)$ maps into the space of bounded linear operators on X and is twice strongly continuously differentiable and the domain of $B(t)$ has to contain the domain of A. Furthermore, $Q(t)$ is assumed to be self-adjoint and positive. This latter condition will in particular guarantee that in the case of $\mathbb{R}^n$, the initial value problems used in the alternating method will not be stiff.

We study the linear second-order differential equation with two-point boundary value conditions:

$$\begin{cases} u'' + Q(t)u = 0, & \text{in } I, \\ u(0) = \varphi, \\ u(T) = \psi, \end{cases} \tag{6}$$

where $u \in C^2([0,T];X)$ and Q as above; for the boundary data φ, $\psi \in X$.

It is known, see [38], that for problem (6) there exists functions $S(t)$ and $C(t)$, commonly denoted the (vector) sine and cosine function respectively, that satisfy

$$S(0) = C'(0) = 0, \quad S'(0) = C(0) = I.$$

Provided that the spectral radius $r(C^*(T)S'(T)) < 1$, then $S(T)$ has an inverse and the solution to (6) can be given as

$$u(t) = S(t)S(T)^{-1}(\psi - C(T)\varphi) + C(t)\varphi. \tag{7}$$

This will be verified in the next section.

For simplicity, we shall assume that X is a Hilbert space mainly to simplify the use of adjoint operators; most of the derivations can be justified also in a Banach space.

1. **Properties of the sine and cosine functions.** The solutions $S(t)$ and $C(t)$ need not be self-adjoint although $Q(t)$ is. By C^* and S^* we mean the adjoint of C and S respectively, i.e. the adjoint of $C(t)$ and $S(t)$ for $t \in I$. For the sake of completeness we include a proof of the following.

Lemma 1. *The solutions $S(t)$ and $C(t)$ to problem* (6) *satisfy the identities:*

$$S'^*(t)C(t) - S^*(t)C'(t) = I, \tag{8}$$

and

$$S'(t)C^*(t) - C'(t)S^*(t) = I. \tag{9}$$

The elements $S^*(t)$ and $C^*(t)$ are the adjoint operators of $S(t)$ and $C(t)$, and I is the identity.

Proof. Due to the smoothness assumption on Q, we can differentiate the left-hand side of equality (8) to formally obtain

$$\begin{aligned}
&\frac{d}{dt}(S'^*(t)C(t) - S^*(t)C'(t)) = \\
&\quad = S''^*(t)C(t) + S'^*(t)C'(t) - S'^*(t)C'(t) - S^*(t)C''(t) = \\
&\quad = S''^*(t)C(t) - S^*(t)C''(t) = \\
&\quad = -(Q(t)S(t))^*C(t) + S^*(t)Q(t)C(t) = \\
&\quad = -S^*(t)Q(t)C(t) + S^*(t)Q(t)C(t) = 0.
\end{aligned}$$

The equality (8) then follows by formally integrating this using the initial conditions for the $S(t)$ and $C(t)$ and their derivatives.

To prove (9), we first show that S^*S' and C^*C' are self-adjoint We have

$$\frac{d}{dt}(S^*S' - S'^*S) = S^*S'' - S''^*S = S^*QS - S^*QS = 0.$$

Again, formally integrating using that $S(0) = 0$, it follows that $S^*S' = S'^*S$. Similarly, one can show that $C^*C' = C'^*C$.

Define the following operator matrix

$$B = \begin{pmatrix} -C'^*(t) & C^*(t) \\ S'^*(t) & -S^*(t) \end{pmatrix}. \tag{10}$$

This matrix is a left inverse of

$$A = \begin{pmatrix} S(t) & C(t) \\ S'(t) & C'(t) \end{pmatrix},$$

that is $BA = I$, and this is straightforward to check by formal matrix multiplication using (8) together with $S^*S' = S'^*S$ and $C^*C' = C'^*C$. Thus, B is the inverse of A and using that therefore $BA = I$, i.e.

$$\begin{pmatrix} S(t) & C(t) \\ S'(t) & C'(t) \end{pmatrix} \begin{pmatrix} -C'^*(t) & C^*(t) \\ S'^*(t) & -S^*(t) \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \tag{11}$$

gives (9). □

We note that from (11) follows immediately that also SC^* and $S'C'^*$ are self-adjoint, which we state as a separate result.

Lemma 2. *The operators SC^* and $S'C'^*$ are self-adjoint.*

We then verify that provided $r(C^*(T)S'(T)) < 1$ then (7) is a well-defined solution to (6).

Lemma 3. *Assume that $r(C^*(T)S'(T)) < 1$. Then the inverse of $S(T)$ exists.*

Proof. This is a standard application of the Neumann series in combination with the relation (8). Indeed,

$$(I - C^*(T)S'(T)) \sum_{j=0}^{k-1} (C^*(T)S'(T))^j \quad = \quad (I - (C^*(T)S'(T))^k). \tag{12}$$

Letting k tend to infinity one can conclude, since $r(C^*(T)S'(T)) < 1$, that $(I - C^*(T)S'(T))$ has an inverse. Applying (8) the result follows. □

2. **Convergence of the alternating procedure for (6).** The alternating procedure for problem (6) was given in the introduction. For clarity, we state the steps again. The element u_{2k} satisfies the initial value problem

$$\begin{cases} u''_{2k} + Q(t)u_{2k} = 0, \text{ in } I, \\ u_{2k}(0) = \varphi, \\ u'_{2k}(0) = u'_{2k-1}(0), \end{cases} \tag{13}$$

where $u'_0(0) = \eta$ is arbitrary. The solution to this problem is given by

$$u_{2k}(t) = S(t)u'_{2k-1}(0) + C(t)\varphi. \tag{14}$$

The element u_{2k+1} is constructed as the solution to

$$\begin{cases} u''_{2k+1} + Q(t)u_{2k+1} = 0, \text{ in } I, \\ u_{2k+1}(T) = \psi, \\ u'_{2k+1}(T) = u'_{2k}(T), \end{cases} \tag{15}$$

with solution

$$u_{2k+1}(t) = (S(t)C^*(T) - C(t)S^*(T))u'_{2k}(T) + (C(t)S'^*(T) - S(t)C'^*(T))\psi. \tag{16}$$

To verify that this indeed is a solution one can use that SC^* and $S'C'^*$ are self-adjoint according to Lemma 2 together with (8)-(9).

We shall then establish convergence of the alternating algorithm (convergence was shown in [33] for time-independent operators).

Theorem 1. *Let u be a solution to problem (6) and let $C(t)$ and $S(t)$ be the fundamental cosine and sine solutions to this problem. Let u_k be the k-th approximate solution generated by the alternating procedure. If $r(C^*(T)S'(T)) < 1$, where r is the spectral radius, then*

$$\|u_{2k} - u\|_\infty \leq C_1\delta^k$$

and

$$\|u_{2k+1} - u\|_\infty \leq C_2\delta^k,$$

where C_1 and C_2 are positive constants and $\delta \in (r(C^*(T)S'(T)), 1)$.

Proof. The solution u_{2k+1} to (15) is given by (16) and this gives

$$\begin{aligned} u_{2k+1}(t) &= (S(t)C^*(T) - C(t)S^*(T))u'_{2k}(T) + \\ &\quad +(C(t)S'^*(T) - S(t)C'^*(T))\psi = \\ &= Z_1(t)u'_{2k}(T) + Z_2(t)\psi. \end{aligned} \tag{17}$$

In particular, calculating $u'_{2k-1}(0)$ and using that the solution to (13) is given by (14) tedious but straightforward calculations show that

$$\begin{aligned} u_{2k}(t) &= S(t)\sum_{j=0}^{k-1}(C^*(T)S'(T))^j\, C'^*(T)(C(T)\varphi - \psi) + \\ &\quad +S(t)(C^*(T)S'(T))^k\eta + C(t)\varphi. \end{aligned} \tag{18}$$

Using this expression in (17) one derives

$$\begin{aligned} u_{2k+1}(t) &= Z_1(t)S'^*(T)(C^*(T)S'(T))^k\eta + \\ &\quad +Z_1(t)S'^*(T)\sum_{j=0}^{k-1}(C^*(T)S'(T))^jC'^*(T)(C(T)\varphi - \psi) + \\ &\quad +Z_1(t)C'^*(T)\varphi + Z_2(t)\psi. \end{aligned}$$

Similar to the proof of Lemma 3 it follows from identity (8) that

$$\begin{aligned} \sum_{j=0}^{k-1}(C^*(T)S'(T))^j &= (I - C^*(T)S'(T))^{-1}(I - (C^*(T)S'(T))^k) = \\ &= S^{-1}(T)C'^*(T)^{-1}(I - (C^*(T)S'(T))^k). \end{aligned} \tag{19}$$

Employing this in (18) one obtains

$$\begin{aligned} u_{2k}(t) &= S(t)(I - (C^*(T)S'(T))^k)S^{-1}(T)C'^*(T)^{-1}C'^*(T)(C(T)\varphi - \psi) + \\ &\quad +S(t)(C^*(T)S'(T))^k\eta + C(t)\varphi = \\ &= S(t)(C^*(T)S'(T))^k(\eta - S^{-1}(T)(C(T)\varphi - \psi)) + \\ &\quad +S(t)S^{-1}(T)(C(T)\varphi - \psi) + C(t)\varphi. \end{aligned}$$

Similarly, using (19) in (17)

$$\begin{aligned} u_{2k+1}(t) &= Z_1(t)S_1'^*(T)(C^*(T)S'(T))^k(\eta - S^{-1}(T)(C(T)\varphi - \psi)) + \\ &\quad +S(t)S^{-1}(T)(C(T)\varphi - \psi) + C(t)\varphi. \end{aligned}$$

Next, from Lemma 3 the element $S^{-1}(T)$ exists and thus the solution to problem (6) is given by (7). Using this, we finally have

$$u_{2k}(t) - u(t) = S(t)(C^*(T)S'(T))^k(\eta - u'(0))$$

and similarly

$$u_{2k+1}(t) - u(t) = Z_1(t)S_1'^*(T)(C^*(T)S'(T))^k(\eta - u'(0)).$$

Taking norms and making use of the identity

$$r(Q) = \limsup_{k\to\infty} \|Q^k\|^{1/k}, \tag{20}$$

we get

$$\|u_{2k} - u\|_\infty \leq \|S\|_\infty \|(C^*(T)S'(T))^k\| \|\eta - u'(0)\| \leq C_1\delta^k$$

and

$$\|u_{2k+1} - u\|_\infty \leq \|Z_1\|_\infty \|S_1'^*(T)\| \|(C^*(T)S'(T))^k\| \|\eta - u'(0)\| \leq C_2\delta^k,$$

where $\delta \in (r(C^*(T)S'(T)), 1)$. Thus, the result follows. □

Remark 1. One can relax the conditions on the operator $Q = A + B(t)$. In fact, one can impose conditions such that $B(t)$ can be a differential operator and thus the problem studied can model for example the Dirichlet problem for a hyperbolic equation, see [37]. This then generalizes the results in [33] for the Dirichlet problem for hyperbolic operators to include time-dependent coefficients. Note though that the Dirichlet problem for the hyperbolic problem has only a unique solution when T is irrational, see [20]. Note also that generalizing to include equations with a term $V(t)u'$ is considerable more difficult in the Banach space setting, see [18, Chapter 8].

Remark 2. Consider the partial differential operator

$$\Delta u + f(u) = 0 \quad \text{in } \Omega$$

supplied with Dirichlet boundary conditions, where Ω is an annular smooth domain in $\mathbb{R}^n$. Searching for a radial solution, $u(r)$, leads to the equation

$$u''(r) + \frac{n-1}{r}u'(r) + f(u(r)) = 0.$$

Substituting $s = r^{2-n}$ gives

$$u''(s) + \rho(s)f(u(s)) = 0,$$

with boundary conditions $u(s_1) = \varphi$ and $u(s_2) = \psi$, see further [34]. Thus, with f of the above form, the results also apply to problems for the Laplace equation. Nonlinear functions f will be discussed in Section 4, thus the alternating method could potentially be applied to semi-linear problems for the Laplace operator.

3. A scalar equation

The results in the previous section are in an abstract setting and as remarked at the end of the previous section the operator $Q(t)$ could even in fact be a partial differential operator. Since the present study has as one of its aims to study the alternating method for ordinary differential equations, we simplify in this section and replace $Q(t)$ by $q^2(t)$, where $q(t)$ is a scalar real-valued function and $X = \mathbb{R}$, and study a classical scalar second-order two point boundary value problem,

$$\begin{cases} u'' + q^2(t)u = 0, \text{ in } I, \\ u(0) = \varphi, \\ u(T) = \psi, \end{cases} \tag{21}$$

where $q \in C(0,T)$. The condition for convergence of the alternating method stated in Lemma 1 is in the case of (21) reduced to $|c(T)s'(T)| < 1$, where c and s are the usual fundamental solutions. To give conditions on the function q and final time T for which this condition is satisfied, we shall need the following two lemmas below. These essentially follow from classical comparison theorems for Sturm-Liouville operators; for completeness we give the proofs. For an overview of history and results on Sturm-Liouville comparison theory, see [15, 17, 50].

Lemma 4. *Let $a, b \in C[0,T]$, and let y be a nontrivial solution of*

$$\begin{cases} y'' + a^2(t)y = 0, \\ y'(0) = 0. \end{cases}$$

Suppose that y has its first positive zero at $t = T$, and let z be a nontrivial solution of the equation

$$\begin{cases} z'' + b^2(t)z = 0, \\ z'(0) = 0, \end{cases}$$

with $b^2(t) > a^2(t)$ on $(0,T)$. Then there exists τ with $0 < \tau < T$, such that $z(\tau) = 0$.

Proof. Without loss of generality we can assume that $y(0) = 1$. Therefore, by the assumption that y has no zeros in $0 < t < T$, we find that y is positive on this interval. Using the governing equation, it follows that y' is decreasing on $(0,T)$. Assume then that z has no zeros in $(0,T)$, for instance, that z is positive on $(0,T)$. Let $w = y'z - yz'$; then $w(0) = 0$ and using $y(T) = 0$ gives $w(T) = y'(T)z(T) \leq 0$ since y' is decreasing and z is positive. However,

$$w' = y''z + y'z' - y'z' - yz'' = yz(b^2 - a^2) > 0,$$

which is a contradiction. □

Similarly, one can show a result about zeros of the derivative.

Lemma 5. *Let $a, b \in C[0,T]$, and let y be a nontrivial solution of*

$$\begin{cases} y'' + a^2(t)y = 0, \\ y(0) = 0. \end{cases}$$

Suppose that y' has its first positive zero at $t = T$, and let z be a solution of

$$\begin{cases} z'' + b^2(t)z = 0, \\ z(0) = 0, \end{cases}$$

with $b^2(t) > a^2(t)$ *on* $(0,T)$. *Then there exists* τ *with* $0 < \tau < T$, *such that* $z'(\tau) = 0$.

Proof. We can assume that $y'(0) = 1$, and since by assumption y' does not have any zero on $(0,T)$ one can conclude that y is positive on $(0,T)$. Assume that z' has no zeros in $(0,T)$, for instance, that z' is positive on $(0,T)$. This gives that z is positive on $(0,T)$ since $z(0) = 0$. Let $w = y'z - yz'$, then $w(0) = 0$ and using that $y'(T) = 0$ together with the positiveness of y and z' imply $w(T) = -y(T)z'(T) \leq 0$. However,

$$w' = y''z + y'z' - y'z' - yz'' = yz(b^2 - a^2) > 0,$$

which is a contradiction. □

To derive properties of the fundamental solutions c and s, we shall use the above lemmas together with the following two eigenvalue problems to compare zeros of the solutions. Let λ_{DN} be the first eigenvalue of the following problem*

$$\begin{cases} u'' + \lambda q^2(t)u = 0, \text{ in } I, \\ u(0) = 0, \\ u'(T) = 0, \end{cases} \tag{22}$$

and let λ_{ND} be the first eigenvalue of the problem†

$$\begin{cases} u'' + \lambda q^2(t)u = 0, \text{ in } I, \\ u'(0) = 0, \\ u(T) = 0. \end{cases} \tag{23}$$

Lemma 6. *Let* λ_{DN} *and* λ_{ND} *be defined as above. If* $1 < \min\{\lambda_{DN}, \lambda_{ND}\}$, *then the alternating procedure converges on every interval* $[0,T_1]$, $0 < T_1 < T$.

Proof. The fundamental solution $c(t)$ satisfies $c(0) = 1$ and $c'(0) = 0$. Clearly, from the governing equation for this function, $c'(t)$ is non-positive on the interval $(0,T)$ implying that $c(t)$ is decreasing on this interval. Suppose that $c(T_1) = 0$ for some $0 < T_1 < T$. Then, from Lemma 4 with $T = T_1$ and $a^2 = q^2$ and $b^2 = \lambda_{ND}q^2$, we conclude that the solution to (23) is zero for $t = \tau$ with $0 < \tau < T_1$. However, then the eigenfunction solution to (23) would be identically zero, which is a contradiction. Therefore, we find that $c(t)$ do no change sign on $[0,T_1]$ and we can conclude that $0 < c(t) < 1$ on $[0,T_1]$. A similar conclusion can be made using Lemma 5 for $s'(t)$, and therefore $0 < c(t)s'(t) < 1$ on $[0,T_1]$. Thus, the condition for convergence in Theorem 1 is satisfied. □

It is then possible to state a convergence result for the alternating method involving a condition on the coefficient q and the final time T.

Theorem 2. *If* $T \leq (2\max_{0\leq t\leq T}|q(t)|)^{-1}\pi$, *then the alternating procedure converges as a geometric progression on the interval* $(0,T)$.

*The subscript DN refers to a Dirichlet condition at $t = 0$ and a Neumann condition at $t = T$.

†The subscript ND refers to a Neumann condition at $t = 0$ and a Dirichlet condition at $t = T$.

Proof. Put $a^2(t) = q^2(t)$ and $b^2(t) = \max_{0\le t\le T} q^2(t) = M^2$. Once can check that $1 < \min\{\lambda_{DN}, \lambda_{ND}\}$ in the interval $[0, T_1]$, where $T_1 \le \pi/2M$. Thus, the conclusion follows from Lemma 6. □

3. **Non-convergence for the alternating method.** As mentioned in the introduction, we are interested in a class of equations for which the alternating method do not converge. Guided by the results in the previous section, we can then give such a class of equations.

Consider the following problem:

$$\begin{cases} u'' - q^2(t)u = 0, \text{ in } I, \\ u(0) = \varphi, \\ u(T) = \psi, \end{cases} \tag{24}$$

where $q \in C[0, T]$. Let c and s be the fundamental solutions corresponding to this equation. Examining the proof of Theorem 1 it is clear that the alternating method do not converge if $|s'(T)c(T)| > 1$. We adjust T, if necessary, such that c and s do not have any zeros for $0 < t < T$. We shall then show that $|s'(T)c(T)| > 1$ holds for the fundamental solutions to (24).

Proposition 1. *Let c and t be the fundamental solutions corresponding to the equation* (24)*. Then $|s'(T)c(T)| > 1$.*

Proof. Since T is chosen such that c and s do not have any zeros in $0 < t < T$ and since $c(0) = 1$ we conclude that c is positive on $(0, T)$. Hence, it follows from the equation (24) that c'' is positive, which implies that c' is increasing on $(0, T)$. Thus, $c(T) > c(0) = 1$. In similar way, one can show that $s'(T) > s'(0) = 1$. □

Therefore, since $|s'(T)c(T)| > 1$, we can conclude that the alternating method applied to (24) will not converge.

4. Nonlinear operators

In this section we shall investigate the nonlinear case

$$\begin{cases} u''(t) + f(u(t)) = 0, \quad \text{in} \quad I, \\ u(0) = \varphi, \\ u(T) = \psi. \end{cases} \tag{25}$$

For simplicity, we assume that u takes values in $\mathbb{R}$. We shall further assume that there exists a unique solution to problem (25). The existence and uniqueness of a solution is a nontrivial matter, and there are plenty of results and conditions in the literature. A good place to start is Chapter 1 in [9]. From that chapter it follows that under a Lipschitz condition on f there exists a time-interval where existence and uniqueness of a solution to (25) holds. Note that only assuming that f is continuous and positive will not guarantee uniqueness, for counterexamples, see [22].

We shall write down the solution to each of the first four steps in the alternating method to be able to derive some general expressions for the generated elements η_k and ζ_k.

To generate an initial guess for the alternating method, let

$$\begin{cases} v''(t) = 0, & \text{in } I, \\ v(0) = \varphi, \\ v(T) = \psi, \end{cases} \tag{26}$$

that is

$$v(t) = \frac{T-t}{T}\varphi + \frac{t}{T}\psi.$$

Then define $\eta_0 = v'(0) = \frac{1}{T}(\psi - \varphi)$. With this initial guess, the first approximation u_0 in the alternating procedure is given by

$$\begin{cases} u_0''(t) + f(u_0(t)) = 0, & \text{in } I, \\ u_0(0) = \varphi, \\ u_0'(0) = \eta_0, \end{cases} \tag{27}$$

with formal solution

$$u_0(t) = \varphi + t\eta_0 - \int_0^t (t-\tau) f(u_0(\tau))\, d\tau = \frac{T-t}{T}\varphi + \frac{t}{T}\psi - \int_0^t (t-\tau) f(u_0(\tau))\, d\tau,$$

where in the last equality the expression for the element η_0 was used. The derivative of u_0 at $t = T$ is calculated from this as

$$u_0'(t) = \frac{1}{T}(\psi - \varphi) - \int_0^t f(u_0(\tau))\, d\tau,$$

giving

$$\zeta_1 = u_0'(T) = \frac{1}{T}(\psi - \varphi) - \int_0^T f(u_0(\tau))\, d\tau.$$

The next approximation u_1 is found from

$$\begin{cases} u_1''(t) + f(u_1(t)) = 0, & \text{in } I, \\ u_1(T) = \psi, \\ u_1'(T) = \zeta_1, \end{cases} \tag{28}$$

with solution

$$u_1(t) = \psi + (t-T)\zeta_1 + \int_t^T (t-\tau) f(u_1(\tau))\, d\tau.$$

Inserting the expression for ζ_1,

$$u_1(t) = \psi + \frac{t-T}{T}(\psi - \varphi) - (t-T)\int_0^T f(u_0(\tau))\, d\tau + \int_t^T (t-\tau) f(u_1(\tau))\, d\tau.$$

From this, the derivative of u_1 at zero is

$$\eta_2 = u_1'(0) = \frac{1}{T}(\psi - \varphi) - \int_0^T \left(f(u_0(\tau)) - f(u_1(\tau))\right) d\tau.$$

Then u_2 is constructed as the solution to

$$\begin{cases} u_2''(t) + f(u_2(t)) = 0, & \text{in } I, \\ u_2(0) = \varphi, \\ u_2'(0) = \eta_2, \end{cases} \tag{29}$$

and formally

$$\begin{aligned} u_2(t) &= \varphi + t\eta_2 - \int_0^t (t-\tau) f(u_2(\tau))\, d\tau = \\ &= \varphi + \frac{t}{T}(\psi - \varphi) - t\int_0^T (f(u_0(\tau)) - f(u_1(\tau)))\, d\tau - \\ &- \int_0^t (t-\tau) f(u_2(\tau))\, d\tau. \end{aligned}$$

Calculating the derivative at $t = T$ we obtain

$$\zeta_3 = u_2'(T) = \frac{1}{T}(\psi - \varphi) - \int_0^T (f(u_0(\tau)) - f(u_1(\tau)) + f(u_2(\tau)))\, d\tau. \tag{30}$$

Then u_3 is constructed,

$$\begin{cases} u_3''(t) + f(u_3(t)) = 0, \quad \text{in} \quad I, \\ u_3(T) = \psi, \\ u_3'(T) = \zeta_3, \end{cases} \tag{31}$$

having the solution

$$u_3(t) = \psi + (t-T)\zeta_3 + \int_t^T (t-\tau) f(u_3(\tau))\, d\tau$$

or by using the expression for ζ_3,

$$\begin{aligned} &u_3(t) = \psi + \\ &\quad + (t-T)\left(\frac{1}{T}(\psi - \varphi) - \int_0^T (f(u_0(\tau)) - f(u_1(\tau)) + f(u_2(\tau)))\, d\tau\right) + \\ &\quad + \int_t^T (t-\tau) f(u_3(\tau))\, d\tau. \end{aligned}$$

From this expression, we have the derivative

$$\begin{aligned} \eta_4 = u_3'(0) = \frac{1}{T}(\psi - \varphi) - \\ - \int_0^T (f(u_0(\tau)) - f(u_1(\tau)) + f(u_2(\tau)) - f(u_3(\tau)))\, d\tau\Big). \end{aligned} \tag{32}$$

Note that (30) and (32) justifies the term alternating method, since the sign appear to alternate with each iteration.

We further observe that

$$\zeta_3 - \zeta_1 = \int_0^T (f(u_1(\tau)) - f(u_2(\tau)))\, d\tau$$

and

$$\eta_4 - \eta_2 = \int_0^T (f(u_3(\tau)) - f(u_2(\tau)))\, d\tau.$$

Continuing by iterating in the last two steps, a simple induction step reveals,

Proposition 2. *Let* $\{\eta_{2k}\}_{k=0}^{\infty}$ *and* $\{\zeta_{2k+1}\}_{k=0}^{\infty}$ *be generated from the alternating procedure. Then*

$$\eta_{2k+2} - \eta_{2k} = \int_0^T \left(f(u_{2k+1}(\tau)) - f(u_{2k}(\tau))\right) d\tau$$

and

$$\zeta_{2k+3} - \zeta_{2k+1} = \int_0^T \left(f(u_{2k+1}(\tau)) - f(u_{2k+2}(\tau))\right) d\tau.$$

Now, note that if f was a positive increasing function and if the approximations u_k generated by the alternating method satsfied $u_{k+1} \geq u_k$, then one can conclude that $\{\eta_{2k}\}$ will be an increasing sequence and $\{\zeta_{2k}\}$ a decreasing sequence. Thus, provided these could be bounded from above and below, one could establish a convergence proof. Another possibility is that the odd approximations $\{u_{2k+1}\}$ are all above each of the even approximations $\{u_{2k}\}$.

However, it appears rather difficult to find conditions on the function f and the final time T to have such conditions satisfied. In fact, in the next section, we shall take a rather simple function f and show numerically that the sequences $\{\eta_{2k}\}$ and $\{\zeta_{2k}\}$ do not need to be monotone, and still there appears to be convergence.

4. **A numerical example for a nonlinear problem.**

Let

$$\begin{cases} u''(t) + \frac{1}{2}\sin(2u(t)) = 0, & \text{in } I, \\ u(0) = 0, \\ u(T) = \psi. \end{cases} \tag{33}$$

Here, $f(u) = \frac{1}{2}\sin(2u(t))$ is Lipschitz with constant $L = 1$. Hence, from [9, p. 5] there is a unique solution to (33) for $T < 2\sqrt{2}$. In fact, we assume that *psi* is chosen such that we have the following explicit expression for the solution,

$$u(t) = \arcsin \frac{e^{2t} - 1}{e^{2t} + 1}. \tag{34}$$

The initial guess is constructed as in the previous section. The initial value problems needed to be solved in each iteration step of the alternating procedure are solved with the Matlab function ODE45 (Matlab version R2013b on a computer with Windows 8.2 and an Intel(R) Core(TM) i3-3217U Central Unit Processor (CPU) at 1.8GHz).

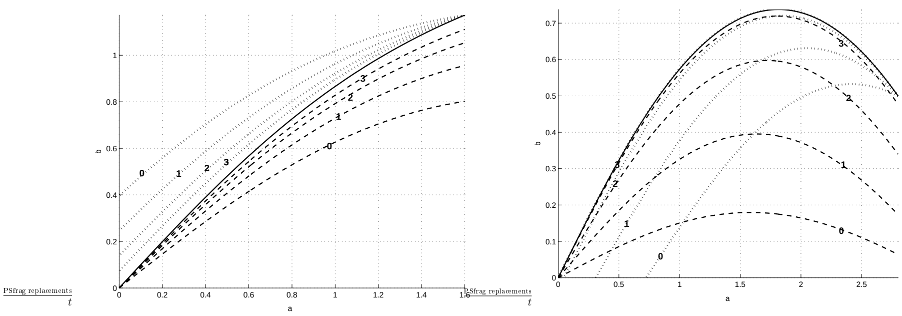
In Fig. 1(a) we present the results obtained after 8 iterations (that is u_7 is the final approximation; the corresponding value for k for the solution u_{2k} and u_{2k+1}, respectively, is marked out on each approximation) obtained with $T = 1.6$ and ψ generated from (34). As can be seen from this figure there is convergence towards the solution to (33). Moreover, a monotone behaviour of the approximations, expected due to Proposition 2, is present. In fact, with these solutions together with the function f and Proposition 2, the sequences η_k and ζ_k should both be positive and decreasing. This has been checked for and is the case in the numerical simulations.

Increasing T there is convergence of the similar kind up to about $T = 1.8$, where the method starts to become slower and eventually does not converge.

Choosing instead $T = 2.8$ and taking $\psi = 0.5$, one can see that monotonicity is no longer present in the sense that some even iterates u_{2k} intersects some odd iterates u_{2m+1}; this is shown in Fig. 1(b). In this case, we used the Matlab function bvpc4 to generate an approximation to the solution to (33) to test convergence against with $\psi = 0.5$ formula (34) does not give the sought solution.

One can also change sign of the function f and run the procedure with $-f$. This causes problems with the alternating method and only for small values of T there seems to be convergence. For example, the method diverges for $T = 1$ and $\psi = 1$ as is highlighted in Fig. 2(a). Note that changing sign was shown in the linear case in Section 3.1 to generate non-convergent sequences in the alternating procedure.

We remark that we have also tried a linearization in the alternating method in the sense that f is instead evaluated on the solution from the previous step. This new linearized procedure does not give any significant improvement for (33). However, changing to $-f$ this linearized procedure appears to converge for $T = 1$ and $\psi = 1$ as shown in Fig. 2(b).



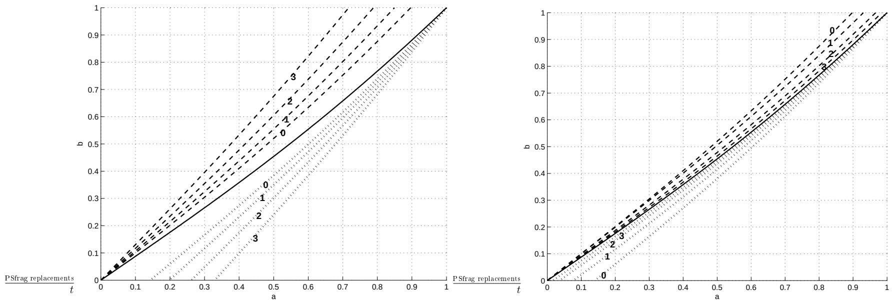
a) $T = 1.6$ and $\psi = u(1.6)$, u from (4.34)

b) $T = 2.8$ and $\psi = 0.5$

FIG. 1. The solutions u_{2k} (- - -) and u_{2k+1} ($\cdots$), and the analytical solution u (—) for various T and ψ.

5. CONCLUSION

The alternating method [33] was investigated for two-point boundary value problems for second order time-dependent differential operators. Convergence was established in the linear case extending [33] to the time-dependent case with the operators taking values in a Hilbert space (potentially the similar analysis can be carried over to the Banach space setting). In the scalar case, a criteria involving the coefficients of the operator and the final time were given to guarantee convergence. It was also shown that changing sign of a term in the differential operator generates equations for which the alternating method does not converge. Moreover, for nonlinear operators, expressions were derived



a) $T = 1$ and $\psi = 1$ with $-f(u)$ b) $T = 1$, $\psi = 1$ with $-f(u)$ and linearization

FIG. 2. The solutions u_{2k} (- - -) and u_{2k+1} ($\cdots$), and the analytical solution u (—) for various T and ψ.

on which a proof of convergence can potentially be obtained. However, as was highlighted by numerical examples, to pin-point precise criteria on the operator and final time to have a proof of convergence also in the nonlinear case seem difficult. A linearization was suggested such that linear differential equations were solved at each iteration step and this linearization turned out to converge in some cases where the orignal alternating method did not converge. This merits further investigations and is deferred to future work.

Bibliography

1. Adam J. A. A simplified mathematical model of tumor growth / J. A. Adam // Math. Biosci. – 1986 – Vol. 81. – P. 224-229.
2. Agarwall R. P. Boundary value problems occurring in plate deflection theory / R. P. Agarwall, G. Akrivis // J. Comput. Appl. Math. – 1982 – Vol. 8. – P. 145-154.
3. Atkinson K. E. Numerical Solution of Ordinary Differential Equations / K. E. Atkinson, W. Han, D. Stewart. – Hoboken, NJ: John Wiley & Sons, Inc., 2009.
4. Avdonin S. Iterative methods for solving a nonlinear boundary inverse problem in glaciology / S. Avdonin, V. Kozlov, D. Maxwell, M. Truffer // J. Inverse Ill-Posed Problems. – 2009 – Vol. 17. – P. 239-258.
5. Baranger T. N. On the alternating method for Cauchy problems and its finite element discretisation, / T. N. Baranger, B. T. Johansson, R. Rischette. – Springer Proceedings in Mathematics & Statistics. – Ed. L. Beilina. – 2013. – P. 183-197.
6. Bastay G. Iterative Methods for Ill-Posed Boundary Value Problems / G. Bastay // Linköping Studies in Science and Technology. – Dissertations. – 1995. – 392 p.
7. Bastay G. An alternating method for the stationary Stokes system / G. Bastay, G. Johansson G., V. Kozlov, D. Lesnic // ZAMM. – 2006 – Vol. 86. – P. 268-280.
8. Baumeister J. On iterative methods for solving ill-posed problems modeled by partial differential equations / J. Baumeister, A. Leitão // J. Inv. Ill-Posed Probl. – 2001 – Vol. 9. – P. 13-29.
9. Bernfeld S. R. An introduction to Nonlinear Boundary Value Problems / S. R. Bernfeld, V. Lakshmikantham. – New-York, London: Academic Press, Inc, 1974.
10. Berntsson F. An alternating iterative procedure for the Cauchy problem for the Helmholtz equation / F. Berntsson, V. Kozlov, L. Mpinganzima, B. O. Turesson // Inverse Probl. Sci. Engn. – 2014 – Vol. 22. – P. 45-62.
11. Bisshopp K. E. Large deflection of cantilever beams, / K. E. Bisshopp, D. C. Drucker // Quart. Appl. Math. – 1945 – Vol. 3. – P. 272-275.

12. Chandra J. Existence of solutions of boundary value problems for nonlinear second-order systems in a Banach space / J. Chandra, V. Lakshmikantham, A. R. Mitchell // Nonlinear Anal. – 1978 – Vol. 2. – P. 157-168.
13. Chapko R. An alternating boundary integral based method for a Cauchy problem for Laplace equation in semi-infinite domains / R. Chapko, B. T. Johansson // Inverse Probl. Imaging. – 2008 – Vol. 3. – P. 317-333.
14. De Coster C. Two-point Boundary Value Problems: lower and upper solutions / C. De Coster, P. Habets. – Amsterdam: Elsevier B. V., 2006.
15. Diaz J. B. Sturm comparison theorems for ordinary and partial differential equations / J. B. Diaz, J. R. McLaughlin // Bull. Amer. Math. Soc. – 1969 – Vol. 75. – P. 335-339.
16. Engl H. W. A Mann iterative regularization method for elliptic Cauchy problems / H. W. Engl, A. Leitão // Numer. Funct. Anal. Optim. – 2001 – Vol. 22. – P. 861-884.
17. Everitt W. N. Charles Sturm and the development of Sturm-Liouville theory in the years 1900 to 1950, in Eds. / W. O. Amrein, A. M. Hinz, D. B. Pearson // Sturm-Liouville Theory. – Birkhäuser. – Basel. – 2005. – P. 45-74,
18. Fattorini H. O. Second Order Linear Differential Equations in Banach Spaces / H. O. Fattorini. – Amsterdam: North-Holland Publishing Co., 1985.
19. Flesch U. The distribution of heat sources in the human head: a theoretical consideration / U. Flesch // J. Theoret. Biol. – 1975 – Vol. 54. – P. 285-287.
20. Fox D. W. The Dirichlet problem for the wave equation / D. W.Fox, C. Pucci // Ann. Mat. Pura Appl. – 1958 – Vol. 4. – P. 155-182.
21. Goldstein J. A. Semigroups of Linear Operators and Applications / J. A. Goldstein. – New York: The Clarendon Press, Oxford University Press, 1985.
22. Guo D. J. Multiple solutions of two-point boundary value problems of ordinary differential equations in Banach spaces / D. J.Guo, V. Lakshmikantham // J. Math. Anal. Appl. – 1988 – Vol. 129. – P. 211-222.
23. Dinh Nho Hào Methods for Inverse Heat Conduction Problems. – / Nho Hào Dinh // Frankfurt/Main: Peter Lang, 1998.
24. Helsing J. Fast reconstruction of harmonic functions from Cauchy data using integral equation techniques / J. Helsing, B. T. Johansson // Inv. Pr. Sci. Engn. – 2010 – Vol. 18. – P. 381-399.
25. Johansson B. T. A relaxation of the alternating method for the reconstruction of a stationary flow from incomplete boundary data Advances in Boundary Integral Methods / B. T.Johansson, D. Lesnic. – Eds. H. Power, A. La Roccaб S. J. Baxter. – // Proceedings of the Seventh UK Conference on Boundary Integral Methods. – University of Nottingham. – UK. – 2009. – P. 161-169.
26. Johansson B. T. An alternating method for Cauchy problems for Helmholtz type operators in non-homogeneous medium, / B. T. Johansson, V. A. Kozlov // IMA J. Appl. Math. – 2009 – Vol. 74. – P. 62-73.
27. Johansson B. T. Relaxation of alternating iterative algorithms for the Cauchy problem associated with the modified Helmholtz equation / B. T. Johansson, L. Marin // CMC Comput. Mater. Continua. – 2010 – Vol. 13. – P. 153-188.
28. Jourhmane M. An alternating method for an inverse Cauchy problem, / M. Jourhmane, A. Nachaoui // Numer. Algorithms. – 1999 – Vol. 21. – P. 247-260.
29. Jourhmane M. Convergence of an alternating method to solve the Cauchy problem for Poission's equation / M. Jourhmane, A. Nachaoui // Appl. Anal. – 2002 – Vol. 81. – P. 1065-1083.
30. Jourhmane M. Relaxation procedures for an iterative algorithm for solving the Cauchy problem for the Laplace equation / M. Jourhmane, D. Lesnic, N. S. Mera // Eng. Anal. Boundary Elements. – 2004 – Vol. 28. – P. 655-665.
31. Keller H. B. Numerical Solution of Two Point Boundary Value Problems / H. B. Keller. – Philadelphia, Pa.: SIAM, 1976.

32. Keller H. H. Radiation heat transfer for annular fins of trapezoidal profile / H. H. Keller, E. S. Holdrege // International Journal of High Performance Computing Applications. – 1970 – Vol. 92. – P. 113-116.
33. Kozlov V. A. On iterative procedures for solving ill-posed boundary value problems that preserve differential equations / V. A. Kozlov, V. G. Maz'ya // Algebra i Analiz. – 1989 – Vol. 1. – P. 144-170. (English transl.: // Leningrad Math. J. – 1990 – Vol. 1. – P. 1207-1228).
34. Lee M. N. On the positive solutions for semi-linear elliptic equations on annular domain with non-homogeneous Dirichlet Boundary conditions / M. N. Lee, S. S. Lin // J. Math. Anal. Appl. – 1994 – Vol. 181. – P. 348-361.
35. Lesnic D. An iterative boundary element method for solving numerically the Cauchy problem for the Laplace equation / D. Lesnic, L. Elliott, D. B. Ingham // Eng. Anal. Bound. Elem. – 1997 – Vol. 20. – P. 123-133.
36. Lin H. S. Oxygen diffusion in a spherical cell with nonlinear oxygen uptake kinetics / H. S. Lin // J. Theoret. Biol. – 1976 – Vol. 60. – P. 449-457.
37. Lin Y. Time-dependent perturbation theory for abstract evolution equations of second order / Y. Lin // Studia Math. – 1998 – Vol. 130. – P. 263-274.
38. Lutz D. On bounded time-dependent perturbations of operator cosine functions / D. Lutz // Aequationes Math. – 1981 – Vol. 23. – P. 197-203.
39. McElwain D. L. S. A re-examination of oxygen diffusion in a spherical cell with Michaelis–Menten oxygen uptake kinetics / D. L. S. McElwain // J. Theoret. Biol. – 1978 – Vol. 71. – P. 255-263.
40. Marin L. Boundary element method for the Cauchy problem in linear elasticity / L. Marin, L. Elliott, D. B. Ingham, D. Lesnic // Eng. Anal. Boundary Elem. – 2001 – Vol. 25. – P. 783-793.
41. Maxwell D. Determining glacier velocities and stresses with inverse methods: an iterative scheme / D. Maxwell, M. Truffer, S. Avdonin, M. Stuefer // Journal Glae. – 2008 – Vol. 54. – P. 888-898.
42. Mera N. S. The boundary element solution of the Cauchy steady heat conduction problem in an anisotropic medium / N. S. Mera, L. Elliott, D. B. Ingham, D. Lesnic // Int. J. Numer. Meth. Engng. – 2000 – Vol. 49. – P. 481-499.
43. Mönch H. Boundary value problems for nonlinear ordinary differential equations of second order in Banach spaces / H. Mönch // Nonlinear Anal. – 1980 – Vol. 4. – P. 985-999.
44. Pandey R. K. On the convergence of a finite difference method for a class of singular boundary value problems arising in physiology / R. K. Pandey, A. K. Singh // J. Comput. Appl. Math. – 2004 – Vol. 166. – P. 553-564.
45. Picard E. Sur l'application des méthodes d'approximations successives á l'étude de certaines équations différentielles ordinaires / E. Picard // J. de Math. – 1893 – Vol. 9. – P. 217-271.
46. Prilepko A. I. Methods for Solving Inverse Problems in Mathematical Physics / A. I. Prilepko, D. G. Orlovsky, I. A. Vasin. – New-York: Marcel Dekker, Inc., 2000.
47. Robert S. Solution of Troesch's two-point boundary value problems by shooting techniques / S. Robert, J. Shipman // J. Comput. Physics. – 1972 – Vol. 10. – P. 232-241.
48. Scorza Dragoni G. II problema dei valori ai limiti studiato in grande per le equazioni differenziali del secondo ordine / Dragoni G. Scorza // Math. Ann. – 1931 – Vol. 105. – P. 133-143.
49. Scorza Dragoni G. Sul problema dei valori ai limiti per i systemi di equazioni differenziali del secondo ordine / Dragoni G. Scorza // Boll. Un. Mat. Ital. – 1935 – Vol. 14. – P. 225-230.
50. Simon B. Sturm oscillation and comparison theorems / B. Simon. – Eds. W. O. Amrein, A. M. Hinz, D. B. Pearson // Sturm-Liouville Theory. – Birkhäuser. – Basel. – 2005. – P. 29-43.
51. Walter W. Existence theorems for a two-point boundary value problem in Banach space / W. Walter // Math. Ann. – 1979 – Vol. 244. – P. 55-64.

52. Ward J. P., King J. R. Mathematical modelling of avascular-tumour growth / J. P. Ward, J. R. King // IMA J. Math. Appl. Medicine Biology. – 1997. – Vol. 14. – P. 39-69.

George Baravdish,
ITN, Campus Norrköping, Linköping University, Sweden;

Tomas Johansson,
ITN, Campus Norrköping, Linköping University, Sweden;
EAS, Aston University, Birmingham, UK

Received 20.03.2014

Журнал обчислювальної та прикладної математики | 2014 № 2 (116) | Journal of Computational & Applied Mathematics

UDC 519.6

AN ALTERNATING BOUNDARY INTEGRAL BASED METHOD FOR A CAUCHY PROBLEM FOR KLEIN-GORDON EQUATION

Roman Chapko, Dmytro Laba

Резюме. Розглядається чисельне розв'язування задачі Коші для рівняння Клейна-Гордона у двозв'язній плоскій області. Зважаючи на некоректність цієї лінійної оберненої задачі, використано альтернуючий метод, який володіє регуляризуючими властивостями. Це приводить до розв'язування двох мішаних крайових задач на кожній ітерації. Ці мішані задачі наближено розв'язуються методом граничних інтегральних рівнянь. Приведено результати чисельних експериментів.

Abstract. We consider the numerical solution of a Cauchy problem for the Klein-Gordon equation in a planar double connected domain. Due to the ill-posedness of this linear inverse problem the alternating method with regularization properties is used. It leads to two mixed well-posed boundary value problems on every iteration. These problems are solved by boundary integral equation method. Numerical examples are presented.

1. Introduction

Let D be a double connected domain in $\mathbb{R}^2$ with inner and outer boundaries Γ_1 and Γ_2, respectively. We suppose that $\Gamma_1, \Gamma_2 \in C^3$ (see Fig. 1). Let ν denote the outward unit normal on boundary.

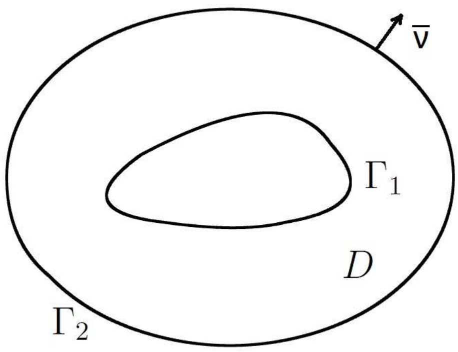


Fig. 1. An example of a double connected domain

Given the sufficiently smooth continuous functions f_1 and f_2, we consider the Cauchy problem of finding a function $u \in C^2(D) \cap C^1(\bar{D})$ which satisfies

Key words. Klein-Gordon equation; Cauchy problem; Double connected domain; Single- and double layer potentials; Integral equations; Alternating method.

the Klein-Gordon equation

$$\Delta u - \varkappa^2 u = 0 \quad \text{in} \quad D \tag{1}$$

and the boundary value conditions

$$u = f \quad \text{and} \quad \frac{\partial u}{\partial \nu} = g \quad \text{on} \quad \Gamma_2. \tag{2}$$

In (1) $\varkappa > 0$ is a given constant. In particular we are interested in finding the Cauchy data on the inner boundary Γ_1.

For the uniqueness of a solution to the Cauchy problem (1), (2) see, for example, [2]. The solution does not in general depend continuously on the data, i.e. the problem is ill-posed in the sense of Hadamard, thus making classical methods inappropriate.

We shall employ the so-called alternating iterative method proposed in [6] and successfully applied in several engineering problems, see for example [5] and [8]. The use of the alternating method with an integral equation approach for the Laplace equation is discussed in [3]. The details of alternating procedure for the case of the Klein-Gordon equation are listed in section 4. In each iteration, mixed direct problems are solved in the solution domain D. There are the Dirichlet-Neumann mixed boundary value problem

$$\Delta w - \varkappa^2 w = 0 \quad \text{in} \quad D, \tag{3}$$

$$w = h \quad \text{on} \quad \Gamma_1, \qquad \frac{\partial w}{\partial \nu} = g \quad \text{on} \quad \Gamma_2 \tag{4}$$

and Neumann-Dirichlet mixed boundary value problem

$$\Delta v - \varkappa^2 v = 0 \quad \text{in} \quad D, \tag{5}$$

$$\frac{\partial v}{\partial \nu} = p \quad \text{on} \quad \Gamma_1, \qquad v = f \quad \text{on} \quad \Gamma_2. \tag{6}$$

For the direct problems in this study, we propose and investigate a numerical method based on the potential theory. Instead, the problems are each reduced to boundary integral equations over Γ_1 and Γ_2. This approach makes the implementation of the alternating method very efficient.

2. Integral equation method for the mixed problems

2.1. Dirichlet-Neumann mixed problem

The problem (3), (4) will be solved by reducing to the system of integral equations of the first kind. We represent the solution $w \in C^2(D) \cap C^1(\bar{D})$ as a combination of a single- and a double-layer potential

$$w(x) = \int\limits_{\Gamma_1} \varphi_1(y)\Phi(x,y)\mathrm{d}s(y) + \int\limits_{\Gamma_2} \varphi_2(y)\frac{\partial \Phi(x,y)}{\partial \nu(y)}\mathrm{d}s(y), \quad x \in D, \tag{7}$$

where φ_1 and φ_2 are unknown continuous densities, $\Phi(x,y) = \frac{1}{2\pi}K_0(\varkappa|x-y|)$ is a fundamental solution of the equation (3) in term of the modified Hankel function K_0 [1].

From the continuity of the single-layer potential and the normal derivative of the double-layer potential we obtain for the problem (3), (4) the following system of integral equations of the first kind

$$\begin{cases} \int\limits_{\Gamma_1} \varphi_1(y)\Phi(x,y)\mathrm{d}s(y) + \int\limits_{\Gamma_2} \varphi_2(y)\dfrac{\partial\Phi(x,y)}{\partial\nu(y)}\mathrm{d}s(y) = h(x), \quad x \in \Gamma_1, \\ \int\limits_{\Gamma_1} \varphi_1(y)\dfrac{\partial\Phi(x,y)}{\partial\nu(x)}\mathrm{d}s(y) + \\ \qquad + \dfrac{\partial}{\partial\nu(x)}\int\limits_{\Gamma_2} \varphi_2(y)\dfrac{\partial\Phi(x,y)}{\partial\nu(y)}\mathrm{d}s(y) = g(x), \quad x \in \Gamma_2. \end{cases} \tag{8}$$

It is known that modified Bessel functions have the following asymptotic properties [1] $K_0(z) \sim \ln\frac{1}{z}, \quad z \to 0$ and $K_1(z) \sim \frac{1}{z}, \quad z \to 0$. Thus, we obtained the system of integral equations of the first kind which contains kernels with logarithmic singularity as well as a hypersingularity.

Using the Maue type expression [7] the second equation from (8) could be rewritten in the following way

$$\int\limits_{\Gamma_1} \varphi_1(y)\Phi(x,y)\mathrm{d}s(y) + \int\limits_{\Gamma_2} \frac{\partial\varphi_2}{\partial\theta}(y)\frac{\partial\Phi(x,y)}{\partial\theta(x)}\mathrm{d}s(y) -$$
$$-\varkappa^2 \int\limits_{\Gamma_2} \varphi_2(y)\Phi(x,y)[\nu(x)\cdot\nu(y)]\mathrm{d}s(y) = g(x), \quad x \in \Gamma_2,$$

where θ denotes the unit tangential vector for Γ_2.

For the future numerical implementation we consider a parametrization of the system (8). We assume that the domain boundaries have the parametric representations

$$\Gamma_i = \{x_i(t) = (x_{i1}(t), x_{i2}(t)), \quad t \in [0, 2\pi]\}, \qquad i = 1, 2,$$

where $x_i : \mathbb{R} \to \mathbb{R}^2$ are C^3 and 2π–periodic with $|x_i'(t)| > 0$ for all $t \in [0, 2\pi]$. As a result of the parametrization of the system (8) we obtain

$$\begin{cases} \dfrac{1}{2\pi}\int\limits_0^{2\pi} [\mu_1(\tau)H_{11}(t,\tau) + \mu_2(\tau)H_{12}(t,\tau)]\,\mathrm{d}\tau = h(t), \\ \dfrac{1}{2\pi}\int\limits_0^{2\pi} \left[\mu_1(\tau)H_{21}(t,\tau) + \mu_2'(\tau)\hat{H}_{22}(t,\tau) + \mu_2(\tau)H_{22}(t,\tau)\right]\mathrm{d}\tau = g(t), \end{cases} \tag{9}$$

where $\mu_i(t) = \varphi_i(x_i(t))$, $i = 1, 2$, $h(t) = h(x_1(t))$, $g(t) = 2g(x_2(t))|x_2'(t)|$. The representation of kernels of the obtained system is listed below

$$H_{11}(t,\tau) = K_0(\varkappa|r_{11}(t,\tau)|)|x_1'(\tau)|,$$
$$H_{12}(t,\tau) = \varkappa K_1(\varkappa|r_{12}(t,\tau)|)\frac{r_{12}(t,\tau)\cdot\nu_2(\tau)}{|r_{12}(t,\tau)|}|x_2'(\tau)|,$$
$$H_{21}(t,\tau) = -2\varkappa K_1(\varkappa|r_{21}(t,\tau)|)\frac{r_{21}(t,\tau)\cdot\nu_2(t)}{|r_{21}(t,\tau)|}|x_1'(\tau)||x_2'(t)|,$$

$$\hat{H}_{22}(t,\tau) = -2\varkappa K_1(\varkappa|r_{22}(t,\tau)|)\frac{[r_{22}(t,\tau)\cdot x_2'(t)]}{|r_{22}(t,\tau)|},$$

$$H_{22}(t,\tau) = 2\varkappa^2 K_0(\varkappa|r_{22}(t,\tau)|)[x_2'(t)\cdot x_2'(\tau)].$$

Here we introduced the notation $r_{ij}(t,\tau) = x_i(t) - x_j(\tau)$.

Next we express the system of integral equations (9) in the specific form to be able to apply the trigonometrical quadrature rules. The system of integral equations in the following form is ready for application of the numerical methods

$$\begin{cases} \frac{1}{2\pi}\int_0^{2\pi}[\mu_1(\tau)(H_{11}^1(t,\tau)\ln\frac{4}{e}\sin^2\frac{t-\tau}{2}+ \\ \qquad + H_{11}^2(t,\tau)) + \mu_2(\tau)H_{12}(t,\tau)]\mathrm{d}\tau = h(t), \\ \frac{1}{2\pi}\int_0^{2\pi}[\mu_1(\tau)H_{21}(t,\tau) + \mu_2'(\tau)\cot\frac{\tau-t}{2}+ \\ \qquad + \mu_2(\tau)(H_{22}^1(t,\tau)\ln\frac{4}{e}\sin^2\frac{t-\tau}{2} + H_{22}^2(t,\tau))]\mathrm{d}\tau = g(t). \end{cases} \tag{10}$$

Here kernels are represented as follows

$$H_{11}^1(t,\tau) = -\frac{1}{2}I_0(\varkappa|x_1(t)-x_1(\tau)|)|x_1'(\tau)|,$$

$$H_{22}^1(t,\tau) =$$

$$= \varkappa^2\Bigg[\frac{I_0(\varkappa|r_{22}(t,\tau)|)+I_2(\varkappa|r_{22}(t,\tau)|)}{2|r_{22}(t,\tau)|^2}r_{22}(t,\tau)\cdot x_2'(t)r_{22}(t,\tau)\cdot x_2'(\tau)$$

$$- \ I_0(\varkappa|r_{22}(t,\tau)|)r_{22}(t,\tau)\cdot\nu_2(t)|x_2'(t)|x_2'(t)\cdot x_2'(\tau)+$$

$$+ \ \frac{I_1(\varkappa|r_{22}(t,\tau)|)}{\varkappa|r_{22}(t,\tau)|^3}r_{22}(t,\tau)\cdot\nu_2(t)|x_2'(t)|r_{22}(t,\tau)\cdot\nu_2(\tau)|x_2'(\tau)|\Bigg],$$

$$H_{ii}^2(t,\tau) = H_{ii}(t,\tau) - H_{ii}^1(t,\tau)\ln\frac{4}{e}\sin^2\frac{t-\tau}{2}, \quad t\neq\tau, \quad i=1,2$$

with diagonal terms

$$H_{22}^1(t,t) = -\frac{1}{2}\varkappa^2|x_2'(t)|^2, \quad H_{11}^2(t,t) = -\frac{1}{2}\ln\frac{e\varkappa^2|x_1'(t)|^2}{4}|x_1'(t)| - \gamma|x_1'(t)|$$

and

$$H_{22}^2(t,t) = \varkappa^2\ln\frac{e\varkappa^2|x_2'(t)|^2}{4}|x_2'(t)|^2-$$

$$-\frac{1}{6}+\frac{1}{3}\frac{x_2'(t)\cdot x_2'''(t)}{|x_2'(t)|^2}+\frac{1}{2}\frac{|x_2''(t)|^2}{|x_2'(t)|^2}-\frac{(x_2'(t)\cdot x_2''(t))^2}{|x_2'(t)|^4}+\varkappa^2(\frac{1}{2}-\gamma)|x_2'(t)|^2,$$

where I_0 and I_1 are the modified Bessel functions and γ is the Euler constant [1].

For $m\in\mathbb{N}\cup\{0\}$ and $0<\alpha<1$, by $C^{m,\alpha}[0,2\pi]$ we denote the space of m-times uniformly Hölder continuously differentiable and 2π-periodic functions furnished with the usual Hölder norm. Using the Riesz theory [7] we can conclude that for given functions $h \in C^{m+1,\alpha}[0,2\pi]$, $g \in C^{m,\alpha}[0,2\pi]$ the system of integral equations (10) provides a unique solution $\mu_1 \in C^{m,\alpha}[0,2\pi]$ and $\mu_2 \in C^{m+1,\alpha}[0,2\pi]$.

Clearly, we have according to (7) the following representation for the normal derivative on the boundary Γ_1

$$\frac{\partial w}{\partial \nu}(x) = -\frac{1}{2}\varphi_1(x)+$$

$$+\int\limits_{\Gamma_1} \varphi_1(y)\frac{\partial \Phi(x,y)}{\partial \nu(x)}\mathrm{d}s(y) + \int\limits_{\Gamma_2} \varphi_2(y)\frac{\partial^2 \Phi(x,y)}{\partial \nu(x)\partial \nu(y)}\mathrm{d}s(y), \quad x \in \Gamma_1,$$

Taking into account the parametric representation of Γ_i, $i = 1,2$ and by some transformation in the kernels we obtain

$$\begin{gathered}\frac{\partial w}{\partial \nu}(x_1(t)) = -\frac{1}{2}\mu_1(t)+ \\ +\frac{1}{2\pi}\int\limits_0^{2\pi}\left[\mu_1(\tau)\left(L_{11}(t,\tau)\ln\frac{4}{e}\sin^2\frac{t-\tau}{2}\mathrm{d}\tau + L_{12}(t,\tau)\right)+ \right. \\ \left. +\mu_2(\tau)L_2(t,\tau)\right]\mathrm{d}\tau, \quad t\in[0;2\pi]\end{gathered} \tag{11}$$

with kernels

$$L_{11}(t,\tau) = \frac{\varkappa}{2}I_1(\varkappa|r_{11}(t,\tau)|)\frac{r_{11}(t,\tau)\cdot\nu_1(t)}{|r_{11}(t,\tau)|}|x_1'(\tau)|,$$

$$L_{12}(t,\tau) = L_1(t,\tau) - L_{11}(t,\tau)\ln\frac{4}{e}\sin^2\frac{t-\tau}{2}, \quad t\neq\tau,$$

$$L_{12}(t,t) = \frac{x_1''(t)\cdot\nu_1(t)}{2|x_1'(t)|}.$$

2.2. Neumann-Dirichlet mixed problem

For solving the mixed boundary value problem (5), (6) we use the similar boundary integral equations approach as described in the previous section.

The solution to the problem (5), (6) inside the domain could be represented as the following sum of potentials

$$v(x) = \int\limits_{\Gamma_1}\varphi_1(y)\frac{\partial\Phi(x,y)}{\partial\nu(y)}\mathrm{d}s(y) + \int\limits_{\Gamma_2}\varphi_2(y)\Phi(x,y)\mathrm{d}s(y), \quad x \in D.$$

As in the previous section, using the boundary conditions, we obtain the system of integral equations which after parametrization and all needed transformations is represented like

$$\begin{cases} \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi}[\mu_1'(\tau)\cot\frac{\tau-t}{2}+\mu_1(\tau)(\tilde{H}_{11}^1(t,\tau)\ln\frac{4}{e}\sin^2\frac{t-\tau}{2}+ \\ \qquad +\tilde{H}_{11}^2(t,\tau))+\mu_2(\tau)\tilde{H}_{12}(t,\tau)]\mathrm{d}\tau=p(t), \\ \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi}[\mu_1(\tau)\tilde{H}_{21}(t,\tau)+ \\ \qquad +\mu_2(\tau)(\tilde{H}_{22}^1(t,\tau)\ln\frac{4}{e}\sin^2\frac{t-\tau}{2}+\tilde{H}_{22}^2(t,\tau))]\mathrm{d}\tau=f(t). \end{cases} \tag{12}$$

Here the kernels are smooth functions and their differential properties are dependent from smoothness of the boundaries Γ_i. Using approach described earlier in this section, one can check the existence and uniqueness of the solution to the system (12).
Again we have the following way to calculate the function values on the inner boundary Γ_1

$$v(x)=\frac{1}{2}\varphi_1(x)+\int_{\Gamma_1}\varphi_1(y)\frac{\partial\Phi(x,y)}{\partial\nu(y)}\mathrm{d}s(y)+\int_{\Gamma_2}\varphi_2(y)\Phi(x,y)\mathrm{d}s(y),\quad x\in\Gamma_1.$$

The corresponding formula for the function values in terms of parametric representation of the boundary curve Γ_1 can be obtained

$$\begin{aligned} v(x_1(t))&=\frac{1}{2}\mu_1(t)+ \\ +\frac{1}{2\pi}\int_0^{2\pi}\Bigg[\mu_1(\tau)\left(\tilde{L}_{11}(t,\tau)\ln\frac{4}{e}\sin^2\frac{t-\tau}{2}+\tilde{L}_{12}(t,\tau)\right)&+ \\ +\mu_2(\tau)\tilde{L}_2(t,\tau)\Bigg]\mathrm{d}\tau.& \end{aligned}$$

3. Numerical solution of integral equations

3.1. Quadrature method

To discretize our integral equations of the first kind we suggest quadrature method. Let $M\in\mathbb{N}$ and $t_j=\dfrac{j\pi}{M}$, $\quad j=0,\ldots,2M-1$. For approximation of corresponding integrals we use the following trigonometrical quadratures [4, 7]

$$
\begin{gathered}
\frac{1}{2\pi}\int\limits_0^{2\pi} f(\tau)\mathrm{d}\tau \approx \frac{1}{2M}\sum_{j=0}^{2M-1} f(t_j),\\
\frac{1}{2\pi}\int\limits_0^{2\pi} f(\tau)\ln\frac{4}{e}\sin^2\frac{t-\tau}{2}\mathrm{d}\tau \approx \sum_{j=0}^{2M-1} R_j(t)f(t_j),\\
\frac{1}{2\pi}\int\limits_0^{2\pi} f'(\tau)\cot\frac{\tau-t}{2}\mathrm{d}\tau \approx \sum_{j=0}^{2M-1} T_j(t)f(t_j).
\end{gathered} \tag{13}
$$

Here the weight functions R_j and T_j are defined as

$$
R_j(t) = -\frac{1}{M}\left[\frac{1}{2} + \sum_{i=1}^{M-1}\frac{1}{i}\cos i(t-t_j) + \frac{\cos M(t-t_j)}{2M}\right]
$$

and

$$
T_j(t) = -\frac{1}{M}\sum_{i=1}^{M-1} i\cos i(t-t_j) - \frac{1}{2}\cos M(t-t_j).
$$

After application quadrature formulas (13) and performing collocation using the nodes of interpolation we obtain the system of linear equations with respect to unknown $\tilde{\mu}_\ell(t_j) \approx \mu_\ell(t_j)$, $\ell = 1, 2$, $j = 0, \dots, 2M-1$

$$
\left\{
\begin{aligned}
&\sum_{j=0}^{2M-1}\tilde{\mu}_1(t_j)H_{11}^1(t_k,t_j)R_j(t_k) + \frac{1}{2M}\sum_{j=0}^{2M-1}\tilde{\mu}_1(t_j)H_{11}^2(t_k,t_j) +\\
&\quad + \frac{1}{2M}\sum_{j=0}^{2M-1}\tilde{\mu}_2(t_j)H_{12}(t_k,t_j) = h(t_k), \quad k = 0,\dots,2M-1,\\
&\frac{1}{2M}\sum_{j=0}^{2M-1}\tilde{\mu}_1(t_j)H_{21}(t_k,t_j) + \sum_{j=0}^{2M-1}\tilde{\mu}_2(t_j)T_j(t_k) -\\
&\quad - \sum_{j=0}^{2M-1}\tilde{\mu}_2(t_j)H_{22}^1(t_k,t_j)R_j(t_k) -\\
&\quad - \frac{1}{2M}\sum_{j=0}^{2M-1}\tilde{\mu}_2(t_j)H_{22}^2(t_k,t_j) = g(t_k), \quad k = 0,\dots,2M-1.
\end{aligned}
\right. \tag{14}
$$

Finally, we have the following representation for approximate solution to Dirichlet-Neumann mixed problem (3), (4) in the domain D

$$
w(x) \approx \frac{1}{2M}\sum_{j=0}^{2M-1}\tilde{\mu}_1(t_j)K_0(\varkappa|x-x_1(t_j)|)|x_1'(t_j)| +
$$

$$
+\frac{1}{2M}\sum_{j=0}^{2M-1}\tilde{\mu}_2(t_j)\varkappa K_1(\varkappa|x-x_2(t_j)|)\frac{[(x-x_2(t_j))\cdot\nu_2(t_j)]}{|x-x_2(t_j)|}|x_2'(t_j)|, \quad x \in D.
$$

Taking into account (11) the numerical approximation for the normal derivative on Γ_1 can be calculated as

$$\frac{\partial w}{\partial \nu}(x_1(t_k)) \approx -\frac{1}{2}\tilde{\mu}_1(t_k) + \sum_{j=0}^{2M-1} \tilde{\mu}_1(t_j) L_{11}(t_k, t_j) R_j(t_k) +$$

$$+\frac{1}{2M}\sum_{j=0}^{2M-1} \tilde{\mu}_1(t_j) L_{12}(t_k, t_j) + \frac{1}{2M}\sum_{j=0}^{2M-1} \tilde{\mu}_2(t_j) L_2(t_k, t_j), \quad k = 0, \dots, 2M-1.$$

Numerical solution of the system (12) is realized in the similar way.

3.2. Numerical experiments for mixed problems

Let's choose the domain with following boundaries (see Fig. 2)

$$\Gamma_1 = \{x(t) = (0.5\cos(t) + 0.5\cos(2t) - 0.25,\ \sin(t)), t \in [0, 2\pi]\}$$

and

$$\Gamma_2 = \{x(t) = (0.3\cos(t) + 0.25,\ 0.2\sin(t)), t \in [0, 2\pi]\}.$$

The boundary conditions for the Dirichlet-Neumann problem are given as

$$h(x) = 0.5x_1, \quad x \in \Gamma_1, \qquad g(x) = 0.05x_2^2, \quad x \in \Gamma_2$$

and for the Neumann-Dirichlet problem we choose

$$p(x) = e^{-x_2}, \quad x \in \Gamma_1, \qquad f(x) = 0.25\sin(x_1 + x_2), \quad x \in \Gamma_2.$$

For both problems we state $\varkappa = 1$.

The maximum norm errors of the obtained numerical solution values on Γ_1 for the Dirichlet-Neumann problem (3), (4) and calculated values of the normal derivative on Γ_1 for the Neumann-Dirichlet problem (5), (6) are listed for various values of the mesh size M in the Table 1. Note, that as the "exact" solutions we use the approximation solutions obtained by our numerical method with $M = 128$.

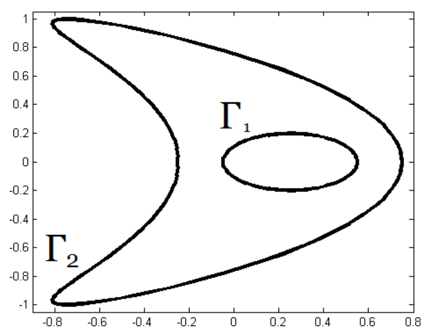


Fig. 2. Solution domain 1

Tabl. 1. Errors of the numerical solutions for the mixed problems

M	$\|\|\frac{\partial w}{\partial \nu} - \frac{\partial w_{ex}}{\partial \nu}\|\|_{C(\Gamma_1)}$	$\|\|v - v_{ex}\|\|_{C(\Gamma_1)}$
4	$1.631718 \cdot 10^{-3}$	$5.145063 \cdot 10^{-3}$
8	$2.131915 \cdot 10^{-5}$	$3.133429 \cdot 10^{-4}$
16	$8.192651 \cdot 10^{-10}$	$4.243675 \cdot 10^{-9}$
32	$3.295214 \cdot 10^{-14}$	$5.041247 \cdot 10^{-13}$

4. An alternating method for the Cauchy problem

4.1. An alternating procedure

To obtain the solution to Cauchy problem (1), (2) we use the alternating iterative procedure.

Each iteration of alternating procedure requires solving one of the mixed boundary value problems and finding Cauchy data on the inner domain boundary. These problems are numerically solved by application of integral equations method described in the above sections.

In problem definitions (3), (4) and (5), (6) functions f and g are the same as in the Cauchy problem (1), (2).

The functions p and h will be substituted with solution approximations during the alternating procedure run.

The alternating procedure of solving Cauchy problem (1), (2) runs as follows

- The first approximation $u^{(0)}$ to the solution is obtained by solving the problem (5), (6), with $p = p_0$, where p_0 is an arbitrary initial guess.
- Having constructed $u^{(2k)}$, we find $u^{(2k+1)}$ by solving (3), (4), with $h = u^{(2k)}\big|_{\Gamma_1}$.
- To obtain $u^{(2k+2)}$ the problem (5), (6) is solved with $p = \left.\dfrac{\partial u^{(2k+1)}}{\partial \nu}\right|_{\Gamma_1}$.

The following result about the convergence of alternating procedure can be obtained using the similar approach as in [3].

Theorem 1. *Suppose that Cauchy problem (1), (2) with appropriate input data f and g has a bounded solution. Let u_k be the k-th approximate solution in the alternating procedure. Then the following is true:*

$$\lim_{k\to\infty} \|u - u_k\|_{L^2(D)} = 0$$

for any sufficiently smooth initial data element p_0 which starts the procedure.

Also we have to note that alternating procedure which is applied to solve Cauchy problem is a regularizing method [3].

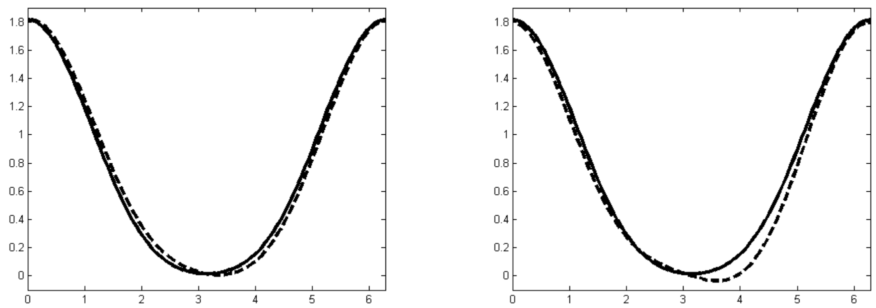
4.2. Numerical experiments for the Cauchy problem

In the numerical experiments we will use the solutions to the mixed problems for generating the input functions for problem (1), (2); i.e. we solve the mixed problem with predefined input functions, calculate the Cauchy data on both boundaries and as a result we got the input data for (1), (2) as well as the solution and it's normal derivative values on the inner boundary (the approximate solution will be compared with this values for checking the results). Please also note that the constant $\varkappa$ is set to one in the following numerical experiments.

Example 1. In the first example we will use the same domain as on Fig. 2. We generate input data for Cauchy problem by solving mixed problem (3), (4) with

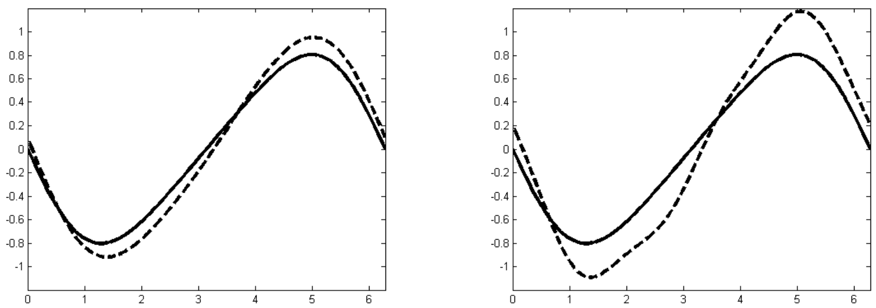
$$h(x) = 6(x_1^2 + x_2^2), \quad x \in \Gamma_1, \qquad g(x) = 3\sin(x_1 + x_2), \quad x \in \Gamma_2.$$

With $M = 128$ and zero initial guess which starts the alternating procedure, we obtain the results reflected in Fig. 3 and Fig. 4 for function and normal derivative reconstructions in case of exact input and input data with noise. The solid line (—) denotes the graph of exact solution and the dashed line (- - -) denotes the numerical solution obtained by alternating procedure.



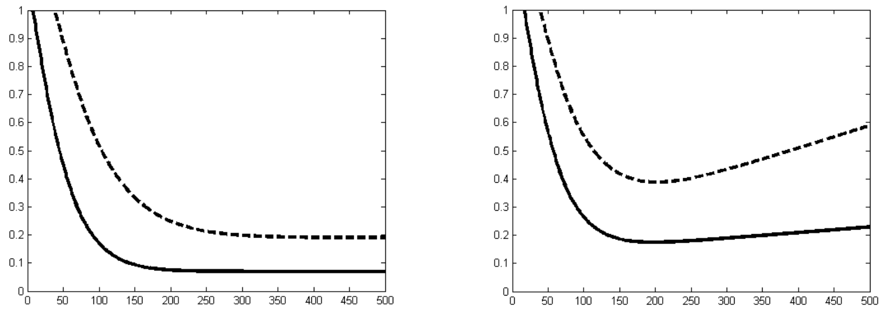
Exact data, $n = 500$ Data with 3% noise, $n = 185$

FIG. 3. Function values on the inner boundary Γ_1 for Ex. 1



Exact data, $n = 500$ Data with 3% noise, $n = 181$

FIG. 4. Normal derivative values on the inner boundary Γ_1 for Ex. 1



Exact data   Data with 3% noise

FIG. 5. C-error of function (—) and normal derivative (- - -) on Γ_1 for Ex. 1

Example 2. Assume that boundaries have the following parametric representations (see Fig. 6)

$$\Gamma_1 = \{x(t) = (0.5\cos(t),\ 0.5\sin(t)), \quad t \in [0, 2\pi]\}$$

and

$$\Gamma_2 = \{x(t) = (2\cos(t),\ \sin(t)), \quad t \in [0, 2\pi]\}.$$

To obtain input functions for this numerical example we solve the mixed boundary value problem (5), (6) with

$$p(x) = x_1 + x_2, \quad x \in \Gamma_1,$$
$$f(x) = 0.5x_1, \quad x \in \Gamma_2.$$

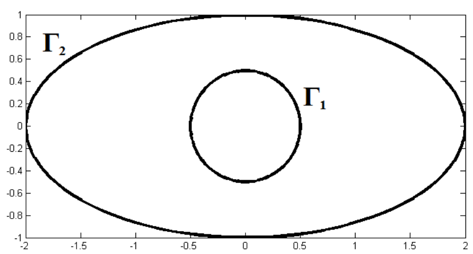
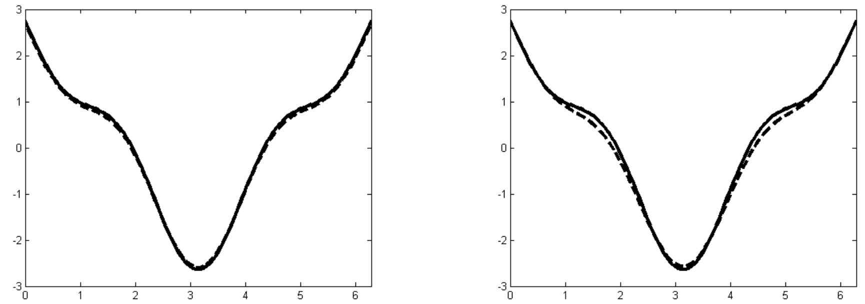


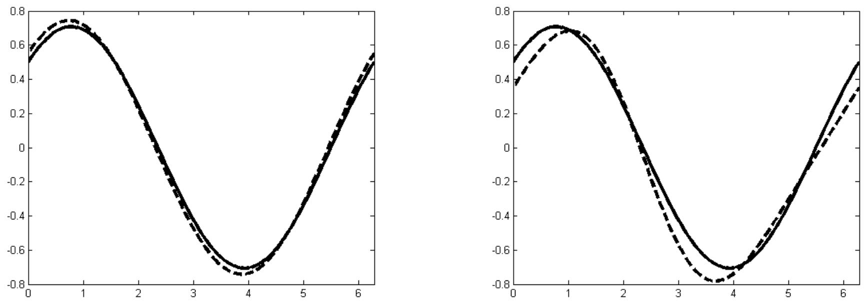
FIG. 6. Solution domain 2

The results of Cauchy data reconstruction on Γ_1 are presented in Fig. 7 and Fig. 8. The corresponding C-errors on every iteration step are reflected in Fig. 9



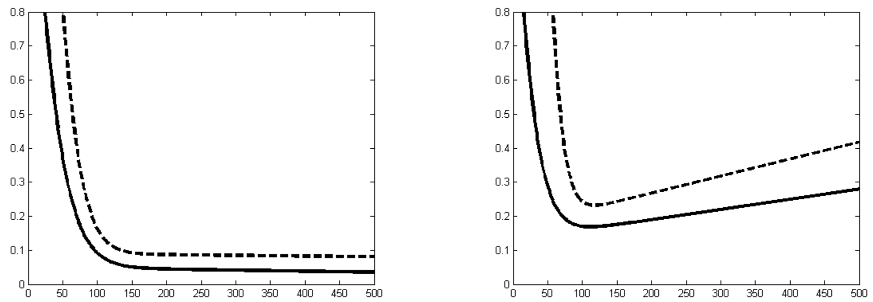
Exact data, $n = 500$ Data with 3% noise, $n = 110$

FIG. 7. Function values on the inner boundary Γ_1 for Ex. 2



Exact data, $n = 500$ Data with 3% noise, $n = 121$

FIG. 8. Normal derivative values on the inner boundary Γ_1 for Ex. 2



Exact data Data with 3% noise

FIG. 9. C-error of solution function (—) and normal derivative (- - -) on Γ_1 for Ex. 2

As one can observe from the above numerical examples, a satisfactory quality for the reconstruction of the boundary function and the normal derivative on the inner boundary is obtained with a reasonable stability against noisy data.

Bibliography

1. Abramowitz M. Handbook of Mathematical Functions With Formulas, Graphs and Mathematical Tables / M. Abramowitz, I. Stegun. – New York: Dover Publications, 1972.
2. Calderón A.-P. Uniqueness in the Cauchy problem for partial differential equations / A.-P. Calderón // Amer. J. Math. – 1958. – Vol. 80. – P. 16-36.
3. Chapko R. An alternating boundary integral based method for a Cauchy problem for the Laplace equation in semi-infinite regions / Roman Chapko, B. Thomas Johanson // Inverse problems and imaging. – 2008. – Vol. 2, No 3. – P. 317-333.
4. Chapko R. On a quadrature method for a logarithmic integral equation of the first kind / R. Chapko, R. Kress // In Agarwal, ed.: World Scientific Series in Applicable Analysis. – 1993. – Vol. 2. – Contributions in Numerical Mathematics. – World Scientific, Singapore – P. 127-140.
5. Jourhmane M. An alternating method for an inverse Cauchy problem / M. Jourhmane, A. Nachaoui // Numer. Algorithms. – 1999. – Vol. 21. – P. 247-260.
6. Kozlov V. A. On iterative procedures for solving ill-posed boundary value problems that preserve differential equations / V. A. Kozlov, V. G. Maz'ya // Algebra i Analiz. – 1989. – Vol. 1. – P. 144-170.
7. Kress R. Linear integral equations / R. Kress. – New-York: Springer-Verlag, 1989. – 367 p.
8. Lesnic D. An iterative boundary element method for solving numerically the Cauchy problem for the Laplace equation / D. Lesnic, L. Elliot, D. B. Ingham // Eng. Anal. Bound. Elem. – 1997. – Vol. 20. – P. 123-133.

Roman Chapko, Dmytro Laba,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine

Received 14.05.2014

Журнал обчислювальної та прикладної математики | 2014 № 2 (116) | Journal of Computational & Applied Mathematics

UDC 519.6

THREE-POINT DIFFERENCE SCHEMES OF HIGH-ORDER ACCURACY FOR SECOND-ORDER NONLINEAR ORDINARY DIFFERENTIAL EQUATIONS WITH BOUNDARY CONDITIONS OF THIRD KIND

Marta Król, Myroslav Kutniv

Резюме. Для нелінійних звичайних диференціальних рівнянь другого порядку з похідною в правій частині та крайовими умовами третього роду побудовано та обґрунтовано триточкові різницеві схеми високого порядку точності на нерівномірній сітці. Побудовано також апроксимацію потоку крайової задачі у вузлах сітки. Для обчислення розв'язку різницевих схем використовуються ітераційні методи. Доведено існування та єдиність розв'язку цих схем, встановлено оцінку точності. Ефективність триточкових різницевих схем шостого порядку точності проілюстрована на прикладах.

Abstract. Three-point difference schemes of hight-order accuracy on a non-equidistant grid for the second-order nonlinear ordinary differential equations with derivative in the right-hand side and boundary conditions of the third kind is constructed and justified. We also construct an approximation of flow for boundary value problem at grid nodes. Iterative methods were used to compute the solution of difference schemes. We prove the existence and uniqueness of the solution of this schemes and determine the order of accuracy. The efficiency of a three-point difference schemes of sixth-order accuracy is illustrated by an examples.

1. Introduction

An approach for construction of exact three-point difference scheme (ETDS) and three-point difference schemes (TDS) of high-order accuracy on a equidistant grid for the nonlinear problems of the form

$$\frac{d}{dx}\left[k(x)\frac{du}{dx}\right] = -f\left(x,u\right), \quad x \in (0,1), \quad u(0) = \mu_1, \quad u(1) = \mu_2$$

was suggested in [8, 7]. These results on a non-equidistant grid were generalized and developed in [6] and for monotone boundary value problems in [5, 1].

In the present paper the effective algorithmic implementation ETDS, proposed in [9], was developed via the truncated TDS for a nonlinear ODEs

$$\frac{d}{dx}\left[k(x)\frac{du}{dx}\right] = -f\left(x,u,\frac{du}{dx}\right), \quad x \in (0,1), \tag{1}$$

Key words. Nonlinear ordinary differential equations, boundary conditions of third kind, exact three-point difference schemes, three-point difference schemes of hight-order accuracy, iterative methods.

with a boundary conditions

$$k(0)\frac{du(0)}{dx} - \beta_1 u(0) = -\mu_1, \quad -k(1)\frac{du(1)}{dx} - \beta_2 u(1) = -\mu_2, \tag{2}$$

where $k(x), f(x,u,\xi)$ are given functions, a $\beta_1, \beta_2, \mu_1, \mu_2$ are given numbers. To find the coefficients and right-hand side of TDS at each node's $x_j, j = 1, 2, ..., N-1$ of the non-equidistance grid we need to solve two auxiliary initial value problems for nonlinear ODEs and two initial value problems for linear ODEs on the intervals $[x_{j-1}, x_j]$ (forward) and $[x_j, x_{j+1}]$ (backward). Moreover, to find right-hand sides difference boundary conditions we need to solve initial value problems for nonlinear and linear ODEs on the intervals $[x_0, x_1]$ (forward) and $[x_{N-1}, x_N]$ (backward). These initial value problems can be solved by executing only one step with an arbitrary one-step method order of accuracy $\bar{m} = 2[(m+1)/2]$ (m is a given the positive integer, $[\cdot]$ denotes the entire part of the number in this brackets). As a result the implementations ETDS which received from truncated TDS of rank $\bar{m}$, for which it is proved that it has an order of accuracy $\bar{m}$. Constructed approaching flow $k(x)du/dx$ at the grid nodes, the order of accuracy of which is the same as the solution, that is of $\bar{m}$.

2. Existence and uniqueness of the solution

Sufficient conditions for the existence and uniqueness of weak solutions of the problem (1), (2) are given by following statement.

Theorem 1. *Let the following assumptions be satisfied*

$$0 < c_1 \le k(x) \le c_2 \quad \forall x \in [0,1], \quad k(x) \in Q^1[0,1], \tag{3}$$

$$\begin{aligned} f_{u\xi}(x) &\equiv f(x,u,\xi) \in Q^0[0,1] \quad \forall u, \xi \in R^1, \\ f_x(u,\xi) &\equiv f(x,u,\xi) \in C(R^2) \quad \forall x \in [0,1], \end{aligned} \tag{4}$$

$$|f(x,u,\xi) - f_0(x)| \le c(|u|)[g(x) + |\xi|] \quad \forall x \in [0,1],\ u, \xi \in R^1, \tag{5}$$

$$[f(x,u,\xi) - f(x,v,\eta)](u-v) \le 0 \quad \forall x \in [0,1],\ u, v, \xi, \eta \in R^1, \tag{6}$$

$$\beta_1 > 0, \quad \beta_2 > 0. \tag{7}$$

Then the problem (1),(2) has a unique solution $u(x) \in W_2^1(0,1)$, *with* $u(x),\ k(x)\dfrac{du}{dx} \in C[0,1]$.

Here $c(t)$ is a continuous function, $f_0(x) \in L_2(0,1)$, $g(x) \in L_1(0,1)$, c_1, c_2, c_3 are some real constants, $Q^p[0,1]$ is the class of functions having p piece-wise continuous derivatives and a finite number of discontinuity points of first kind.

The proof can be found in [9].

3. Algorithmic implementation of the exact three-point difference schemes

On the interval $(0,1)$ we introduce the non-equidistant grid

$$\hat{\omega}_h = \{x_j \in (0,1),\ j = 1, 2, ..., N-1,\ h_j = x_j - x_{j-1} > 0,\ h_1 + h_2 + ... + h_N\}$$

such the discontinuity points of functions $k(x), f(x,u,\xi)$ coincide with the nodes of the grid $\hat{\omega}_h$. Denote by ρ the set of all discontinuity points and assume that N in such that $\rho \subseteq \hat{\omega}_h$. At discontinuity points the solution of problem (1),(2) should satisfy the continuity conditions

$$u(x_i - 0) = u(x_i + 0), \quad k(x)\frac{du}{dx}\bigg|_{x=x_i-0} = k(x)\frac{du}{dx}\bigg|_{x=x_i+0} \qquad \forall x_i \in \rho.$$

For problem (1),(2) in paper [9] is constructed ETDS of the form

$$(au_{\bar{x}})_{\hat{x},j} = -\varphi(x_j, u), \quad j = 1, 2, ..., N-1, \tag{8}$$

$$\begin{aligned} &\frac{1}{\hbar_0}\left(a_1 u_{x,0} - \beta_1 u_0\right) = -\varphi(x_0, u), \\ -&\frac{1}{\hbar_N}\left(a_N u_{\bar{x},N} + \beta_2 u_N\right) = -\varphi(x_N, u), \end{aligned} \tag{9}$$

where

$$u_{\bar{x},j} = \frac{u_j - u_{j-1}}{h_j}, \quad u_{\hat{x},j} = \frac{u_{j+1} - u_j}{\hbar_j}, \quad u_{x,j} = \frac{u_{j+1} - u_j}{h_{j+1}},$$

$$a(x_j) = \left[\frac{1}{h_j} V_1^j(x_j)\right]^{-1}, \quad \hbar_j = \frac{h_j + h_{j+1}}{2}, \quad \hbar_0 = 0,5h_1, \quad \hbar_N = 0,5h_N,$$

$$\varphi(x_j, u) = [\hbar_j V_1^j(x_j)]^{-1} \int\limits_{x_{j-1}}^{x_j} V_1^j(\xi) f\left(\xi, u, \frac{du}{d\xi}\right) d\xi +$$

$$+ [\hbar_j V_2^j(x_j)]^{-1} \int\limits_{x_j}^{x_{j+1}} V_2^j(\xi) f\left(\xi, u, \frac{du}{d\xi}\right) d\xi,$$

$$\varphi(x_0, u) = [\hbar_0 V_1^1(x_1)]^{-1} \int\limits_{x_0}^{x_1} V_2^0(\xi) f\left(\xi, u, \frac{du}{d\xi}\right) d\xi + \hbar_0^{-1}\mu_1,$$

$$\varphi(x_N, u) = [\hbar_N V_1^N(x_N)]^{-1} \int\limits_{x_{N-1}}^{x_N} V_1^N(\xi) f\left(\xi, u, \frac{du}{d\xi}\right) d\xi + \hbar_N^{-1}\mu_2,$$

$$V_1^j(x) = \int\limits_{x_{j-1}}^{x} \frac{dt}{k(t)}, \quad V_2^j(x) = \int\limits_{x}^{x_{j+1}} \frac{dt}{k(t)}.$$

First of all, take into account, that since

$$(-1)^{\alpha+1} \int\limits_{x_{j+(-1)^\alpha}}^{x_j} V_\alpha^j(\xi) f\left(\xi, u(\xi), \frac{du}{d\xi}\right) d\xi =$$

$$= (-1)^\alpha V_\alpha^j(x_j) Z_\alpha^j(x_j, u) + Y_\alpha^j(x_j, u) - u_{j+(-1)^\alpha},$$

where $Y_\alpha^j(x,u), Z_\alpha^j(x,u), j = 2-\alpha,\ 3-\alpha, ..., N+1-\alpha,\ \alpha = 1,2$ are the solutions of the initial value problems

$$\begin{gathered}\frac{dY_\alpha^j(x,u)}{dx} = \frac{Z_\alpha^j(x,u)}{k(x)}, \quad \frac{dZ_\alpha^j(x,u)}{dx} = -f\left(x, Y_\alpha^j(x,u), \frac{Z_\alpha^j(x,u)}{k(x)}\right), \\ x_{j-2+\alpha} < x < x_{j-1+\alpha}, \\ Y_\alpha^j(x_{j+(-1)^\alpha}, u) = u_{j+(-1)^\alpha}, \quad Z_\alpha^j(x_{j+(-1)^\alpha}, u) = k(x)\frac{du}{dx}\bigg|_{x=x_{j+(-1)^\alpha}},\end{gathered} \tag{10}$$

and $\bar{V}_\alpha^j(x) = (-1)^{\alpha+1}V_\alpha^j(x)$ are the solutions of the initial value problems

$$\begin{gathered}\frac{d\bar{V}_\alpha^j(x)}{dx} = \frac{1}{k(x)}, \quad x_{j-2+\alpha} < x < x_{j-1+\alpha}, \\ \bar{V}_\alpha^j(x_{j+(-1)^\alpha}) = 0, \quad j = 2-\alpha, 3-\alpha, ..., N+1-\alpha, \quad \alpha = 1,2.\end{gathered} \tag{11}$$

Obviously, the right-hand side of the ETDS can be written as

$$\varphi(x_j, u) = \frac{1}{\hbar_j}\sum_{\alpha=1}^{2}(-1)^\alpha\left[Z_\alpha^j(x_j,u) + (-1)^\alpha\frac{Y_\alpha^j(x_j,u) - u_{j+(-1)^\alpha}}{V_\alpha^j(x_j)}\right], \tag{12}$$

$$\varphi(x_0, u) = \frac{1}{\hbar_0}\left[Z_2^0(x_0,u) + \frac{Y_2^0(x_0,u) - u_1}{V_2^0(x_0)} + \mu_1\right], \tag{13}$$

$$\varphi(x_N, u) = \frac{1}{\hbar_N}\left[-Z_1^N(x_N,u) + \frac{Y_1^N(x_N,u) - u_{N-1}}{V_1^N(x_N)} + \mu_2\right]. \tag{14}$$

Therefore, to construct the ETDS (8), (9), (12)-(14) for $j = 2-\alpha, 3-\alpha, ..., N+1-\alpha$ it is necessary to solve initial value problems (10), (11) in the forward direction ($\alpha = 1$), and in the backward direction ($\alpha = 2$). We will solve then numerically by using one-step methods:

$$\begin{aligned}Y_\alpha^{(\bar{m})j}(x_j,u) &= u_{j+(-1)^\alpha} + (-1)^{\alpha+1}h_{j-1+\alpha}\times \\ &\times \Phi_1\left(x_{j+(-1)^\alpha}, u_{j+(-1)^\alpha}, \left(k\frac{du}{dx}\right)_{j+(-1)^\alpha}, (-1)^{\alpha+1}h_{j-1+\alpha}\right),\end{aligned} \tag{15}$$

$$\begin{aligned}Z_\alpha^{(m)j}(x_j,u) &= \left(k\frac{du}{dx}\right)_{j+(-1)^\alpha} + (-1)^{\alpha+1}h_{j-1+\alpha}\times \\ &\times \Phi_2\left(x_{j+(-1)^\alpha}, u_{j+(-1)^\alpha}, \left(k\frac{du}{dx}\right)_{j+(-1)^\alpha}, (-1)^{\alpha+1}h_{j-1+\alpha}\right),\end{aligned} \tag{16}$$

$$\bar{V}_\alpha^{(\bar{m})j}(x_j) = (-1)^{\alpha+1}h_{j-1+\alpha}\Phi_3(x_{j+(-1)^\alpha}, 0, (-1)^{\alpha+1}h_{j-1+\alpha}), \tag{17}$$

where $\Phi_1(x,u,v,h), \Phi_2(x,u,v,h), \Phi_3(x,u,h)$ are increment functions,

$$\left(k\frac{du}{dx}\right)_{j+(-1)^\alpha} = k(x)\frac{du}{dx}\bigg|_{x=x_{j+(-1)^\alpha}},$$

$Z_\alpha^{(m)j}(x_j,u)$ approximates the values $Z_\alpha^j(x_j,u)$ with an order of accuracy m, $Y_\alpha^{(\bar{m})j}(x_j,u)$ and $\bar{V}_\alpha^{(\bar{m})j}(x_j)$ approximate $Y_\alpha^j(x_j,u)$ and $\bar{V}_\alpha^j(x_j)$, respectively, with accuracy order $\bar{m}$.

If $k(x)$ and the right-hand side of the differential equation $f(x,u,\xi)$ are differentiated a sufficient number of times, then there exist expansions

$$\begin{aligned} Y_\alpha^j(x_j,u) =& Y_\alpha^{(\bar m)j}(x_j,u)+\\ &+[(-1)^{\alpha+1}h_{j-1+\alpha}]^{\bar m+1}\psi_\alpha^j(x_{j+(-1)^\alpha},u)+O(h_{j-1+\alpha}^{\bar m+2}), \end{aligned} \tag{18}$$

$$\begin{aligned} Z_\alpha^j(x_j,u) =& Z_\alpha^{(m)j}(x_j,u)+\\ &+[(-1)^{\alpha+1}h_{j-1+\alpha}]^{m+1}\tilde\psi_\alpha^j(x_{j+(-1)^\alpha},u)+O(h_{j-1+\alpha}^{m+2}), \end{aligned} \tag{19}$$

$$\begin{aligned} \bar V_\alpha^j(x_j) =& \bar V_\alpha^{(\bar m)j}(x_j)+\\ &+[(-1)^{\alpha+1}h_{j-1+\alpha}]^{\bar m+1}\bar\psi_\alpha^j(x_{j+(-1)^\alpha})+O(h_{j-1+\alpha}^{\bar m+2}). \end{aligned} \tag{20}$$

If in the ETDS (8), (9), (12)-(14) the exact solutions of the corresponding initial value problems (10), (11) are approximated by numerical solutions, the following truncated TDS of rank $\bar m$ is obtained:

$$(a^{(\bar m)}y_{\bar x}^{(\bar m)})_{\hat x,j} = -\varphi^{(\bar m)}(x_j,y^{(\bar m)}),\quad j=1,2,\dots,N-1, \tag{21}$$

$$\begin{aligned} &\frac{1}{\hbar_0}\left(a_1^{(\bar m)}y_{x,0}^{(\bar m)}-\beta_1 y_0^{(\bar m)}\right) = -\varphi^{(\bar m)}(x_0,y^{(\bar m)}),\\ -&\frac{1}{\hbar_N}\left(a_N^{(\bar m)}y_{\bar x,N}^{(\bar m)}+\beta_2 y_N^{(\bar m)}\right) = -\varphi^{(\bar m)}(x_N,y^{(\bar m)}), \end{aligned} \tag{22}$$

where

$$a^{(\bar m)}(x_j) = \left[\frac{1}{h_j}V_1^{(\bar m)}(x_j)\right]^{-1},\quad j=1,2,\dots,N,$$

$$\varphi^{(\bar m)}(x_j,u) = \hbar_j^{-1}\sum_{\alpha=1}^{2}(-1)^\alpha\left[Z_\alpha^{(m)j}(x_j,u)+(-1)^\alpha\frac{Y_\alpha^{(\bar m)j}(x_j,u)-u_{j+(-1)^\alpha}}{V_\alpha^{(\bar m)j}(x_j)}\right],$$

$$\varphi^{(\bar m)}(x_0,u) = \frac{1}{\hbar_0}\left[Z_2^{(m)0}(x_0,u)+\frac{Y_2^{(\bar m)0}(x_0,u)-u_1}{V_2^{(\bar m)0}(x_0)}+\mu_1\right],$$

$$\varphi^{(\bar m)}(x_N,u) = \frac{1}{\hbar_N}\left[-Z_1^{(m)N}(x_N,u)+\frac{Y_1^{(\bar m)N}(x_N,u)-u_{N-1}}{V_1^{(\bar m)N}(x_N)}+\mu_2\right].$$

We need the following assertion to prove the existence and uniqueness of a solution to TDS (21), (22) and to establish its accuracy.

Lemma 1. *Let*

$$0<c_1\le k(x)\le c_2\quad \forall x\in[0,1],\quad k(x)\in Q^{m+1}[0,1],$$

$$f(x,u,\xi)\in\bigcup_{j=1}^{N}C^m\left([x_{j-1},x_j]\times R^2\right).$$

Then one has the following estimates

$$\left|a^{(\bar m)}(x_j)-a(x_j)\right|\le M\,|h|^{\bar m},\quad j=1,2,\dots,N, \tag{23}$$

$$\begin{aligned} &\varphi^{(\bar{m})}(x_j,u)-\varphi(x_j,u)= \\ &=\left\{h_j^{m+1}\left[k(x)\left(\psi_1^j(x,u)-\bar{\psi}_1^j(x)k(x)\frac{du}{dx}\right)-\tilde{\psi}_1^j(x,u)\right]_{x=x_j+0}\right\}_{\hat{x}}+ \\ &+O\left(\frac{h_j^{m+2}+h_{j+1}^{m+2}}{\hbar_j}\right),\quad j=1,2,\ldots,N-1, \end{aligned}\tag{24}$$

$$\begin{aligned} &\varphi^{(\bar{m})}(x_0,u)-\varphi(x_0,u)= \\ &=\frac{h_1^{m+1}}{\hbar_0}\left[k(x)\left(\psi_1^1(x,u)-\bar{\psi}_1^1(x)k(x)\frac{du}{dx}\right)-\tilde{\psi}_1^1(x,u)\right]_{x=x_1}+ \\ &+O\left(\frac{h_1^{m+2}}{\hbar_0}\right), \end{aligned}\tag{25}$$

$$\begin{aligned} &\varphi^{(\bar{m})}(x_N,u)-\varphi(x_N,u)= \\ &=-\frac{h_N^{m+1}}{\hbar_N}\left[k(x)\left(\psi_1^N(x,u)-\bar{\psi}_1^N(x)k(x)\frac{du}{dx}\right)-\tilde{\psi}_1^N(x,u)\right]_{x=x_N}+ \\ &+O\left(\frac{h_N^{m+2}}{\hbar_N}\right), \end{aligned}\tag{26}$$

if m is odd and

$$\begin{aligned} &\varphi^{(\bar{m})}(x_j,u)-\varphi(x_j,u)= \\ &=\left\{h_j^{m}\left[k(x)\left(\psi_1^j(x,u)-\bar{\psi}_1^j(x)k(x)\frac{du}{dx}\right)\right]_{x=x_j+0}\right\}_{\hat{x}}+ \\ &+O\left(\frac{h_j^{m+1}+h_{j+1}^{m+1}}{\hbar_j}\right),\quad j=1,2,\ldots,N-1, \end{aligned}\tag{27}$$

$$\begin{aligned} &\varphi^{(\bar{m})}(x_0,u)-\varphi(x_0,u)= \\ &=\frac{h_1^{m}}{\hbar_0}\left[k(x)\left(\psi_1^1(x,u)-\bar{\psi}_1^1(x)k(x)\frac{du}{dx}\right)\right]_{x=x_1}+O\left(\frac{h_1^{m+1}}{\hbar_0}\right), \end{aligned}\tag{28}$$

$$\begin{aligned} &\varphi^{(\bar{m})}(x_N,u)-\varphi(x_N,u)= \\ &=-\frac{h_N^{m}}{\hbar_N}\left[k(x)\left(\psi_1^N(x,u)-\bar{\psi}_1^N(x)k(x)\frac{du}{dx}\right)\right]_{x=x_N}+O\left(\frac{h_N^{m+1}}{\hbar_N}\right), \end{aligned}\tag{29}$$

if m is even.

Proof. The estimate (23) follows from relation (20). Actually,

$$a^{(\bar{m})}(x_j)-a(x_j)=\frac{h_j[V_1^j(x_j)-V_1^{(\bar{m})j}(x_j)]}{V_1^j(x_j)V_1^{(\bar{m})j}(x_j)}=O(h_j^{\bar{m}}).$$

Let us prove (24)-(29). Note that

$$\begin{aligned}\varphi^{(\bar m)}(x_j,u)-\varphi(x_j,u)=\hbar_j^{-1}\sum_{\alpha=1}^{2}(-1)^{\alpha}\Big\{Z_{\alpha}^{(m)j}(x_j,u)-Z_{\alpha}^{j}(x_j,u)+\\ +(-1)^{\alpha}\left[\frac{Y_{\alpha}^{(\bar m)j}(x_j,u)-u_{j+(-1)^{\alpha}}}{V_{\alpha}^{(\bar m)j}(x_j)}-\frac{Y_{\alpha}^{j}(x_j,u)-u_{j+(-1)^{\alpha}}}{V_{\alpha}^{j}(x_j)}\right]\Big\}.\end{aligned} \tag{30}$$

$$\begin{aligned}\varphi^{(\bar m)}(x_0,u)-\varphi(x_0,u)=\frac{1}{\hbar_0}\Big\{Z_2^{(m)0}(x_0,u)-Z_2^{0}(x_0,u)+\\ +\frac{Y_2^{(\bar m)0}(x_0,u)-u_1}{V_2^{(\bar m)0}(x_0)}-\frac{Y_2^{0}(x_0,u)-u_1}{V_2^{0}(x_0)}\Big\},\end{aligned} \tag{31}$$

$$\begin{aligned}\varphi^{(\bar m)}(x_N,u)-\varphi(x_N,u)=\frac{1}{\hbar_N}\Big\{-Z_1^{(m)N}(x_N,u)+Z_1^{N}(x_N,u)+\\ +\frac{Y_1^{(\bar m)N}(x_N,u)-u_{N-1}}{V_1^{(\bar m)N}(x_N)}-\frac{Y_1^{N}(x_N,u)-u_{N-1}}{V_1^{N}(x_N)}\Big\}.\end{aligned} \tag{32}$$

From Lemma 3.4 (see [4, p.102]) and the equalities

$$V_{\alpha}^{j}(x_j)=\frac{h_{j-1+\alpha}}{k_{j+(-1)^{\alpha}}}+O(h_{j-1+\alpha}^2),$$

$$Y_{\alpha}^{j}(x_j,u)-u_{j+(-1)^{\alpha}}=(-1)^{\alpha+1}h_{j-1+\alpha}\left.\frac{du}{dx}\right|_{x=x_{j+(-1)^{\alpha}}}+O(h_{j-1+\alpha}^2),$$

$$j=2-\alpha,\ 3-\alpha,...,N+1-\alpha,\quad \alpha=1,2$$

we obtain

$$\begin{aligned}&Z_{\alpha}^{(m)j}(x_j,u)-Z_{\alpha}^{j}(x_j,u)=\\ &\quad=-[(-1)^{\alpha+1}h_{j-1+\alpha}]^{m+1}\tilde\psi_1^{j-1+\alpha}(x_{j+(-1)^{\alpha}},u)+O(h_{j-1+\alpha}^{m+2}),\end{aligned} \tag{33}$$

$$\begin{aligned}&\frac{Y_{\alpha}^{(\bar m)j}(x_j,u)-u_{j+(-1)^{\alpha}}}{V_{\alpha}^{(\bar m)j}(x_j)}-\frac{Y_{\alpha}^{j}(x_j,u)-u_{j+(-1)^{\alpha}}}{V_{\alpha}^{j}(x_j)}=-(-1)^{\alpha+1}h_{j-1+\alpha}^{\bar m}\times\\ &\times\left[k(x)\left(\psi_1^{j-1+\alpha}(x,u)-\bar\psi_1^{j-1+\alpha}(x)k(x)\frac{du}{dx}\right)\right]_{x=x_{j+(-1)^{\alpha}}}+\\ &+O(h_{j-1+\alpha}^{\bar m+1}),\quad j=2-\alpha,\ 3-\alpha,...,N+1-\alpha,\quad \alpha=1,2.\end{aligned} \tag{34}$$

Then the equalities (30)-(32) are reduced to estimates (25), (26), and

$$\begin{aligned}
&\varphi^{(\bar m)}(x_j,u)-\varphi(x_j,u)=\frac{1}{\hbar_j}\left\{h_{j+1}^{m+1}\times\right.\\
&\times\left[k(x)\left(\psi_1^{j+1}(x,u)-\bar\psi_1^{j+1}(x)k(x)\frac{du}{dx}\right)-\tilde\psi_1^{j+1}(x,u)\right]_{x=x_{j+1}}-\\
&\left.-h_j^{m+1}\left[k(x)\left(\psi_1^{j}(x,u)-\bar\psi_1^{j}(x)k(x)\frac{du}{dx}\right)-\tilde\psi_1^{j}(x,u)\right]_{x=x_{j-1}}\right\}+\\
&+O\left(\frac{h_j^{m+2}+h_{j+1}^{m+2}}{\hbar_j}\right),
\end{aligned}\tag{35}$$

for odd m, and to (28), (29), and

$$\begin{aligned}
&\varphi^{(\bar m)}(x_j,u)-\varphi(x_j,u)=\\
&=\frac{1}{\hbar_j}\left\{h_{j+1}^{m}\left[k(x)\left(\psi_1^{j+1}(x,u)-\bar\psi_1^{j+1}(x)k(x)\frac{du}{dx}\right)\right]_{x=x_{j+1}}-\right.\\
&\left.-h_j^{m}\left[k(x)\left(\psi_1^{j}(x,u)-\bar\psi_1^{j}(x)k(x)\frac{du}{dx}\right)\right]_{x=x_{j-1}}\right\}+\\
&+O\left(\frac{h_j^{m+1}+h_{j+1}^{m+1}}{\hbar_j}\right),
\end{aligned}\tag{36}$$

for even m.

Since

$$\begin{aligned}
&\left[k(x)\left(\psi_1^{j}(x,u)-\bar\psi_1^{j}(x)k(x)\frac{du}{dx}\right)\right]_{x=x_{j-1}}=\\
&\qquad=\left[k(x)\left(\psi_1^{j}(x,u)-\bar\psi_1^{j}(x)k(x)\frac{du}{dx}\right)\right]_{x=x_{j}}+O(h_j),\\
&\tilde\psi_1^{j}(x_{j-1},u)=\tilde\psi_1^{j}(x_j,u)+O(h_j),
\end{aligned}$$

it follows from (35) and (36) that the estimates (24) and (27) hold.

On the basis of the above-obtained results, one can prove the following assertion.

Theorem 2. *Let the assumptions of Theorem* 1 *and Lemma* 1 *hold. Then there exists an* $h_0>0$ *such that for all* $\{h_j\}_{j=1}^N$ *with* $|h|=\max\limits_{1\le j\le N}h_j\le h_0$ *and TDS (21), (22) has a unique solution, whose accuracy is characterized by the estimate*

$$\left\|y^{(\bar m)}-u\right\|_{1,2,\hat{\hat\omega}_h}^{*}=\left[\left\|y^{(\bar m)}-u\right\|_{0,2,\hat{\hat\omega}_h}^{2}+\left\|k\frac{dy^{(\bar m)}}{dx}-k\frac{du}{dx}\right\|_{0,2,\hat{\hat\omega}_h}^{2}\right]^{1/2}\le M\,|h|^{\bar m},$$

where

$$\|u\|_{0,2,\hat{\omega}_h} = \left\{ \sum_{j=0}^{N} \hbar_j u_j^2 \right\}^{1/2},$$

$$k(x)\frac{dy^{(\bar{m})}}{dx}\bigg|_{x=x_j} = a_{j-1+\alpha} y^{(\bar{m})}_{\bar{x},j-1+\alpha} + Z^{(m)j}_{\alpha}\left(x_j, y^{(\bar{m})}\right) + + (-1)^{\alpha} \frac{Y^{(\bar{m})j}_{\alpha}\left(x_j, y^{(\bar{m})}\right) - y^{(\bar{m})}_{j+(-1)^{\alpha}}}{V^{(\bar{m})j}_{\alpha}(x_j)}, \quad \alpha = 1, 2,$$

$$j = 1, 2, ..., N-1,$$

$$k(x)\frac{dy^{(\bar{m})}}{dx}\bigg|_{x=x_0} = \beta_1 y_0^{(\bar{m})} - \mu_1, \quad k(x)\frac{dy^{(\bar{m})}}{dx}\bigg|_{x=x_N} = -\beta_2 y_N^{(\bar{m})} + \mu_2,$$

and M is a constant independent of $|h|$.

Proof. Let us consider the operators

$$B_h^{(\bar{m})} u_j = \begin{cases} -\dfrac{1}{\hbar_0}\left(a_1^{(\bar{m})} u_{x,0} - \beta_1 u_0\right), & j = 0, \\ -(a^{(\bar{m})} u_{\bar{x}})_{\hat{x},j}, & j = 1, 2, ..., N-1, \\ \dfrac{1}{\hbar_N}\left(a_N^{(\bar{m})} u_{\bar{x},N} + \beta_2 u_N\right), & j = N, \end{cases} \tag{37}$$

$$A_h^{(\bar{m})}(x_j, u) = \begin{cases} -\dfrac{1}{\hbar_0}\left(a_1^{(\bar{m})} u_{x,0} - \beta_1 u_0\right) - \varphi^{(\bar{m})}(x_0, u), & j = 0, \\ -(a^{(\bar{m})} u_{\bar{x}})_{\hat{x},j} - \varphi^{(\bar{m})}(x_j, u), & j = 1, 2, ..., N-1, \\ \dfrac{1}{\hbar_N}\left(a_N^{(\bar{m})} u_{\bar{x},N} + \beta_2 u_N\right) - \varphi^{(\bar{m})}(x_N, u), & j = N, \end{cases} \tag{38}$$

which is defined in the finite-dimensional space of grid functions $H(\hat{\bar{\omega}}_h)$ with the scalar products

$$(u, v)_{\hat{\bar{\omega}}_h} = \sum_{\xi \in \hat{\omega}_h} \hbar(\xi) u(\xi) v(\xi) + \hbar_0 u_0 v_0 + \hbar_N u_N v_N,$$

$$(u, v)_{\hat{\omega}_h^+} = \sum_{\xi \in \hat{\omega}_h^+} h(\xi) u(\xi) v(\xi), \quad \hat{\omega}_h^+ = \hat{\omega}_h \cup x_N,$$

and the norms

$$\|u\|_{0,2,\hat{\bar{\omega}}_h} = (u, u)^{1/2}_{\hat{\bar{\omega}}_h}, \quad \|u\|_{0,2,\hat{\omega}_h^+} = (u, u)^{1/2}_{\hat{\omega}_h^+},$$

$$\|u\|_{1,2,\hat{\bar{\omega}}_h} = \left(\|u\|^2_{0,2,\hat{\bar{\omega}}_h} + \|u_{\bar{x}}\|^2_{0,2,\hat{\omega}_h^+}\right)^{1/2}, \quad \|u\|_{B_h^{(\bar{m})}} = \left(B_h^{(\bar{m})} u, u\right)^{1/2}_{\hat{\bar{\omega}}_h}.$$

Because (see proof of Theorem 2 in [9])

$$(\varphi(x, u) - \varphi(x, v), u - v)_{\hat{\bar{\omega}}_h} \le -c_1 \int_0^1 \left\{ \frac{d}{d\eta}\left[\hat{u}(\eta) - \hat{v}(\eta) - u(\eta) + v(\eta)\right] \right\}^2 d\eta,$$

$$\hat{u}(\eta) = u(x_j)\frac{V_1^j(\eta)}{V_1^j(x_j)} + u(x_{j-1})\frac{V_2^{j-1}(\eta)}{V_1^j(x_j)}, \quad x_{j-1} \le \eta \le x_j,$$

then due to (23)-(29) $\exists h_0 > 0$ such that $\forall \{h_j\}_{j=1}^N$ with $|h| = \max\limits_{1\le j\le N} h_j \le h_0$ the following estimation holds:

$$0 < \tilde{c}_1 \le a^{(\bar{m})}(x) \quad \forall x \in \hat{\omega}_h^+,$$

$$\left(\varphi^{(\bar{m})}(x,u) - \varphi^{(\bar{m})}(x,v), u - v\right)_{\hat{\bar{\omega}}_h} \le 0.$$

Then, from the Green's first difference formula [3, p.26]) and inequality (see [3, p.39])

$$\gamma_1 \|u\|^2_{0,2,\hat{\bar{\omega}}_h} \le \left(u^2_{\bar{x}}, 1\right)_{\hat{\omega}_h^+} + \beta_1 u_0^2 + \beta_2 u_N^2, \ \gamma_1 > 0, \tag{39}$$

it follows that

$$\begin{aligned}
&\left(A_h^{(\bar{m})}(x,u) - A_h^{(\bar{m})}(x,v), u - v\right)_{\hat{\bar{\omega}}_h} = \left(a^{(\bar{m})}\left(u_{\bar{x}} - v_{\bar{x}}\right)^2, 1\right)_{\hat{\omega}_h^+} + \\
&+ \beta_1\left(u_0 - v_0\right)^2 + \beta_2\left(u_N - v_N\right)^2 - \\
&- \left(\varphi^{(\bar{m})}(x,u) - \varphi^{(\bar{m})}(x,v), u - v\right)_{\hat{\bar{\omega}}_h} \ge \|u - v\|^2_{B_h^{(\bar{m})}} = \\
&= \left(a^{(\bar{m})}\left(u_{\bar{x}} - v_{\bar{x}}\right)^2, 1\right)_{\hat{\omega}_h^+} + \beta_1\left(u_0 - v_0\right)^2 + \beta_2\left(u_N - v_N\right)^2 \ge \\
&\ge \min\{\tilde{c}_1, 1\}\left[\left((u_{\bar{x}} - v_{\bar{x}})^2, 1\right)_{\hat{\omega}_h^+} + \beta_1(u_0 - v_0)^2 + \beta_2(u_N - v_N)^2\right] \ge \\
&\ge \min\{\tilde{c}_1, 1\}\,\gamma_1\|u - v\|^2_{0,2,\hat{\bar{\omega}}_h}.
\end{aligned} \tag{40}$$

Therefore, if $|h| \le h_0$, then $A_h^{(\bar{m})}(x,u)$ is strongly monotone operator, and the TDS (21), (22) has a unique solution $y^{(\bar{m})}(x), x \in \hat{\bar{\omega}}_h$ (see [2, p.461]).

For error $z(x) = y^{(\bar{m})}(x) - u(x), x \in \hat{\bar{\omega}}_h$ of difference scheme (21), (22) will have a problem

$$\begin{aligned}
&-\left[a^{(\bar{m})}(x)z_{\bar{x}}(x)\right]_{\hat{x}} - \left(\varphi^{(\bar{m})}(x,y^{(\bar{m})}) - \varphi^{(\bar{m})}(x,u)\right) = \\
&\quad = \varphi^{(\bar{m})}(x,u) - \varphi(x,u) + \left[\left(a^{(\bar{m})}(x) - a(x)\right)u_{\bar{x}}(x)\right]_{\hat{x}}, \quad x \in \hat{\omega}_h,
\end{aligned} \tag{41}$$

$$\begin{aligned}
&-\frac{1}{\hbar_0}\left(a_1^{(\bar{m})}z_{x,0} - \beta_1 z_0\right) - \left(\varphi^{(\bar{m})}(x_0,y^{(\bar{m})}) - \varphi^{(\bar{m})}(x_0,u)\right) = \\
&\quad = \varphi^{(\bar{m})}(x_0,u) - \varphi(x_0,u) + \frac{1}{\hbar_0}\left(a_1^{(\bar{m})} - a_1\right)u_{x,0},
\end{aligned} \tag{42}$$

$$\begin{aligned}
&\frac{1}{\hbar_N}\left(a_N^{(\bar{m})}z_{x,N} - \beta_2 z_N\right) - \left(\varphi^{(\bar{m})}(x_N,y^{(\bar{m})}) - \varphi^{(\bar{m})}(x_N,u)\right) = \\
&\quad = \varphi^{(\bar{m})}(x_N,u) - \varphi(x_N,u) - \frac{1}{\hbar_N}\left(a_N^{(\bar{m})} - a_N\right)u_{\bar{x},N}.
\end{aligned} \tag{43}$$

From (41)-(43) we obtain

$$\begin{aligned}\left(A_h^{(\bar m)}(x,y^{(\bar m)})-A_h^{(\bar m)}(x,u),z\right)_{\hat{\bar\omega}_h} &= \\ &= \left(\left(\left(a^{(\bar m)}-a\right)u_{\bar x}\right)_{\hat x},z\right)_{\hat\omega_h}-\left(a_N^{(\bar m)}-a_N\right)u_{\bar x,N}z_N+ \\ &+\left(a_1^{(\bar m)}-a_1\right)u_{x,0}z_0+\left(\varphi^{(\bar m)}(x,u)-\varphi(x,u),z\right)_{\hat{\bar\omega}_h}.\end{aligned} \tag{44}$$

Using the relations (23)-(29), Cauchy-Bunyakovsky-Schwartz inequality, formula of summation by parts (see [3, p.25]), evaluate the expression on the right-hand side equality (44)

$$\begin{aligned}&\left(\left(\left(a^{(\bar m)}-a\right)u_{\bar x}\right)_{\hat x},z\right)_{\hat\omega_h}-\left(a_N^{(\bar m)}-a_N\right)u_{\bar x,N}z_N+\left(a_1^{(\bar m)}-a_1\right)u_{x,0}z_0= \\ &=\left(\left(a^{(\bar m)}-a\right)\ u_{\bar x},z_{\bar x}\right)_{\hat\omega_h^+}\le\left\|a^{(\bar m)}-a\right\|_{0,2,\hat\omega_h^+}\|u_{\bar x}\|_{0,2,\hat\omega_h^+}\|z_{\bar x}\|_{0,2,\hat\omega_h^+}\le \\ &\le M\,|h|^{\bar m}\,\|z_{\bar x}\|_{0,2,\hat\omega_h^+}\le\frac{M\,|h|^{\bar m}}{\tilde c_1}\,\|z\|_{B_h^{(\bar m)}},\end{aligned} \tag{45}$$

$$\begin{aligned}\left(\varphi^{(\bar m)}(x,u)-\varphi(x,u),z\right)_{\hat{\bar\omega}_h}&\le M\,|h|^{m+1}\,\|z_{\bar x}\|_{0,2,\hat\omega_h^+}\le \\ &\le\frac{M\,|h|^{m+1}}{\tilde c_1}\,\|z\|_{B_h^{(\bar m)}},\end{aligned} \tag{46}$$

if m is odd;

$$\left(\varphi^{(\bar m)}(x,u)-\varphi(x,u),z\right)_{\hat{\bar\omega}_h}\le M\,|h|^{m}\,\|z_{\bar x}\|_{0,2,\hat\omega_h^+}\le\frac{M\,|h|^{m}}{\tilde c_1}\,\|z\|_{B_h^{(\bar m)}}, \tag{47}$$

if m is even.

Taking into account the relations (40), (44)-(47) is the true estimation

$$\|z\|^2_{B_h^{(\bar m)}}\le\left(A_h^{(\bar m)}(x,y^{(\bar m)})-A_h^{(\bar m)}(x,u),z\right)_{\hat{\bar\omega}_h}\le M\,|h|^{\bar m}\,\|z\|_{B_h^{(\bar m)}}.$$

Hence it follows that $\|z\|_{B_h^{(\bar m)}}\le M\,|h|^{\bar m}$. So on the basis of equivalence of norms $\|\cdot\|_{1,2,\hat{\bar\omega}_h}$, $\|\cdot\|_{B_h^{(\bar m)}}$, we obtain

$$\|z\|_{1,2,\hat{\bar\omega}_h}\le M\,|h|^{\bar m}. \tag{48}$$

Due to (23), (48), (33), (34), (15)-(17) we have

$$\left|\left(k\frac{dz}{dx}\right)_0\right|\le\beta_1\,|z_0|\le M\,|h|^{\bar m},\quad\left|\left(k\frac{dz}{dx}\right)_N\right|\le\beta_2\,|z_N|\le M\,|h|^{\bar m}, \tag{49}$$

$$\left|\left(k\frac{dz}{dx}\right)_j\right| \le \left|a_{j-1+\alpha}^{(\bar m)} - a_{j-1+\alpha}\right|\left|y_{\bar x,j-1+\alpha}^{(\bar m)}\right| + |a_{j-1+\alpha}|\,|z_{\bar x,j-1+\alpha}| +$$
$$+\left|Z_\alpha^{(m)j}(x_j,y^{(\bar m)}) + (-1)^\alpha \frac{Y_\alpha^{(\bar m)j}(x_j,y^{(\bar m)}) - y_{j+(-1)^\alpha}^{(\bar m)}}{V_\alpha^{(\bar m)j}(x_j)} -\right.$$
$$\left.-Z_\alpha^{(m)j}(x_j,u) - (-1)^\alpha \frac{Y_\alpha^{(\bar m)j}(x_j,u) - u_{j+(-1)^\alpha}}{V_\alpha^{(\bar m)j}(x_j)}\right| +$$
$$+\left|Z_\alpha^{(m)j}(x_j,u) - Z_\alpha^{j}(x_j,u)\right| +$$
$$+\left|\frac{Y_\alpha^{(\bar m)j}(x_j,u) - u_{j+(-1)^\alpha}}{V_\alpha^{(\bar m)j}(x_j)} - \frac{Y_\alpha^{j}(x_j,u) - u_{j+(-1)^\alpha}}{V_\alpha^{j}(x_j)}\right| \le$$
$$\le M\,|h|^{\bar m} +$$
$$+\left|\Phi\left(x_{j+(-1)^\alpha}, y_{j+(-1)^\alpha}^{(\bar m)}, \left(k\frac{dy^{(\bar m)}}{dx}\right)_{j+(-1)^\alpha}, (-1)^{\alpha+1}h_{j-1+\alpha}\right) -\right.$$
$$\left.-\Phi\left(x_{j+(-1)^\alpha}, u_{j+(-1)^\alpha}, \left(k\frac{du}{dx}\right)_{j+(-1)^\alpha}, (-1)^{\alpha+1}h_{j-1+\alpha}\right)\right|,$$
$$\alpha = 1,2, \quad j = 1,2,...,N-1,$$

where

$$\Phi(x,u,v,h) = v + h\Phi_2(x,u,v,h) - \frac{\Phi_1(x,u,v,h)}{\Phi_3(x,u,h)}.$$

Because

$$\Phi_1(x,u,v,0) = \frac{v}{k(x)}, \quad \Phi_3(x,0,0) = \frac{1}{k(x)},$$

so using the Theorem on finite increments, we obtain

$$\Phi(x,u,v,h) = \Phi(x,u,v,0) + h\frac{\partial\Phi(x,u,v,\bar h)}{\partial h} = h\frac{\partial\Phi(x,u,v,\bar h)}{\partial h}, \quad \bar h \in (0,h).$$

Then

$$\left|\left(k\frac{dz}{dx}\right)_j\right| \le M\,|h|^{\bar m} +$$
$$+ h_{j-1+\alpha}\left|\frac{\partial\Phi\left(x_{j+(-1)^\alpha}, y_{j+(-1)^\alpha}^{(\bar m)}, \left(k\frac{dy^{(\bar m)}}{dx}\right)_{j+(-1)^\alpha}, (-1)^{\alpha+1}\bar h_{j-1+\alpha}\right)}{\partial h} -\right.$$
$$\left.-\frac{\partial\Phi\left(x_{j+(-1)^\alpha}, u_{j+(-1)^\alpha}, \left(k\frac{du}{dx}\right)_{j+(-1)^\alpha}, (-1)^{\alpha+1}\bar h_{j-1+\alpha}\right)}{\partial h}\right| \le \tag{50}$$

$$\le M\,|h|^{\bar m} + h_{j-1+\alpha}\left|\frac{\partial^2\Phi(x_{j+(-1)^\alpha},\bar u_{j+(-1)^\alpha},\bar v_{j+(-1)^\alpha},\bar h_{j-1+\alpha})}{\partial h\partial u}\right|\left|z_{j+(-1)^\alpha}\right| +$$

$$+ h_{j-1+\alpha}\left|\frac{\partial^2\Phi(x_{j+(-1)^\alpha},\bar u_{j+(-1)^\alpha},\bar v_{j+(-1)^\alpha},\bar h_{j-1+\alpha})}{\partial h\partial v}\right|\left|\left(k\frac{dz}{dx}\right)_{j+(-1)^\alpha}\right| \le$$

$$\le M\,|h|^{\bar m} + |h|\,M_1\left|\left(k\frac{dz}{dx}\right)_{j+(-1)^\alpha}\right|, \quad \alpha = 1,2, \quad j = 1,2,...,N-1,$$

where $\bar u_j = u_j + \theta_j z_j,\ 0 < \theta_j < 1, \bar v_j = \left(k\frac{du}{dx}\right)_j + \eta_j\left(k\frac{dz}{dx}\right)_j,\ 0 < \eta_j < 1,\ j = 0,1,2,...,N$.

Consistently applying inequalities (49), (50), we obtain

$$\left|\left(k\frac{dz}{dx}\right)_j\right| \le M\,|h|^{\bar m}, \quad j = 0,1,2,...,N.$$

Hence

$$\left\|k\frac{dz}{dx}\right\|_{0,2,\hat{\bar\omega}_h} \le M\,|h|^{\bar m}.$$

Therefore, taking into account (48), we will have $\|z\|^*_{1,2,\hat{\bar\omega}_h} \le M\,|h|^{\bar m}$.

For solving the nonlinear TDS order of accuracy $\bar m$ (21), (22) apply the iteration method.

Theorem 3. *Let the conditions of Theorem 2 are satisfied. Then*

$$\left|\varphi^{(\bar m)}(x,u) - \varphi^{(\bar m)}(x,v)\right| \le \tilde L\,|u-v|,$$

there exist an $h_0 > 0$ such that for all $\{h_j\}_{j=1}^N$ with $|h| \le h_0$,

$$0 < \tilde c_1 \le a^{(\bar m)}(x),$$

$$\left(A_h^{(\bar m)}(x,u) - A_h^{(\bar m)}(x,v), u-v\right)_{\hat{\bar\omega}_h} \ge \|u-v\|^2_{B_h^{(\bar m)}},$$

the iteration method

$$B_h^{(\bar m)}\frac{y^{(\bar m,n)} - y^{(\bar m,n-1)}}{\tau} + A_h^{(\bar m)}(x, y^{(\bar m,n-1)}) = 0, \quad x \in \hat{\bar\omega}_h, \quad n = 1,2,..., \tag{51}$$

$$y^{(\bar m,0)}(x) = \frac{\mu_1 + \mu_2 + \mu_1\beta_2 V_2^{(\bar m)}(x) + \mu_2\beta_1 V_1^{(\bar m)}(x)}{\beta_1 + \beta_2 + \beta_1\beta_2 V_1^{(\bar m)}(1)}, \quad x \in \hat{\bar\omega}_h,$$

$$V_1^{(\bar m)}(x_j) = \sum_{k=1}^{j} V_1^{(\bar m)k}(x_k), \quad V_2^{(\bar m)}(x_j) = \sum_{k=j+1}^{N} V_1^{(\bar m)k}(x_k)$$

with

$$\tau = \tau_0 = \left(1 + \frac{\tilde L}{\gamma_1 \min\{\tilde c_1, 1\}}\right)^{-2}$$

converges and for the error we have

$$\left\|y^{(m,n)} - u\right\|^*_{1,2,\hat{\bar\omega}_h} \le M(|h|^{\bar m} + q^n), \quad q = \sqrt{1-\tau_0}, \tag{52}$$

where the operators $B_h^{(\bar m)}, A_h^{(\bar m)}(x,u)$ are determined by the formulas (37), (38),

$$k(x)\frac{dy^{(\bar m,n)}}{dx}\bigg|_{x=x_0} = \beta_1 y_0^{(\bar m,n)} - \mu_1, \quad k(x)\frac{dy^{(\bar m,n)}}{dx}\bigg|_{x=x_N} = -\beta_2 y_N^{(\bar m,n)} + \mu_2,$$

$$\begin{aligned} k(x)\frac{dy^{(\bar m,n)}}{dx}\bigg|_{x=x_j} =& a_{j-1+\alpha} y_{\bar x,j-1+\alpha}^{(\bar m,n)} + Z_\alpha^{(m)j}\left(x_j, y^{(\bar m,n)}\right) \\ &+ (-1)^\alpha \frac{Y_\alpha^{(\bar m)j}\left(x_j, y^{(\bar m,n)}\right) - y_{j+(-1)^\alpha}^{(\bar m,n)}}{V_\alpha^{(\bar m)j}(x_j)}, \quad \alpha = 1,2, \\ & j = 1,2,...,N-1, \end{aligned}$$

and M is a constant independent of $|h|, m, n$.

Proof. According to Theorem 2 we have

$$\begin{aligned} \left\|y^{(\bar m,n)} - u\right\|_{1,2,\hat{\bar\omega}_h}^* \le \left\|y^{(\bar m)} - u\right\|_{1,2,\hat{\bar\omega}_h}^* + \left\|y^{(\bar m,n)} - y^{(\bar m)}\right\|_{1,2,\hat{\bar\omega}_h}^* \le \\ \le M|h|^{\bar m} + \left\|y^{(\bar m,n)} - y^{(\bar m)}\right\|_{1,2,\hat{\bar\omega}_h}^* . \end{aligned} \tag{53}$$

Considering that the $f(x,u,\xi) \in \bigcup_{j=1}^{N} C^{\bar m}([x_{j-1},x_i]\times R^2)$, we obtain

$$\left|\varphi^{(\bar m)}(x,u) - \varphi^{(\bar m)}(x,v)\right| \le \tilde L\,|u-v|, \quad x \in \hat{\bar\omega}_h.$$

Using the Cauchy-Bunyakovsky-Schwartz inequality and (39) we get an estimate

$$\begin{aligned} \left(A_h^{(\bar m)}(x,u) - A_h^{(\bar m)}(x,v), w\right)_{\hat{\bar\omega}_h} &\le \|u-v\|_{B_h^{(\bar m)}} \|w\|_{B_h^{(\bar m)}} + \\ + \left\|\varphi^{(\bar m)}(x,u) - \varphi^{(\bar m)}(x,v)\right\|_{0,2,\hat{\bar\omega}_h} \|w\|_{0,2,\hat{\bar\omega}_h} &\le \\ \le \|u-v\|_{B_h^{(\bar m)}} \|w\|_{B_h^{(\bar m)}} + \tilde L\,\|u-v\|_{0,2,\hat{\bar\omega}_h} \|w\|_{0,2,\hat{\bar\omega}_h} &\le \\ \le \left(1 + \frac{\tilde L}{\gamma_1 \min\{\tilde c_1, 1\}}\right) \|u-v\|_{B_h^{(\bar m)}} \|w\|_{B_h^{(\bar m)}} . \end{aligned}$$

We put $w = \left(B_h^{(\bar m)}\right)^{-1}\left(A_h^{(\bar m)}(x,u) - A_h^{(\bar m)}(x,v)\right)$, then

$$\begin{aligned} \left\|\left(B_h^{(\bar m)}\right)^{-1}\left(A_h^{(\bar m)}(x,u) - A_h^{(\bar m)}(x,v)\right)\right\|_{B_h^{(\bar m)}} \le \\ \le \left(1 + \frac{\tilde L}{\gamma_1 \min\{\tilde c_1, 1\}}\right) \|u-v\|_{B_h^{(\bar m)}} . \end{aligned} \tag{54}$$

From (41), (54) it follows

$$\left(A_h^{(\bar{m})}(x,u) - A_h^{(\bar{m})}(x,v), \left(B_h^{(\bar{m})}\right)^{-1}\left(A_h^{(\bar{m})}(x,u) - A_h^{(\bar{m})}(x,v)\right)\right)_{\hat{\bar{\omega}}_h} \le$$

$$\le \left(1 + \frac{\tilde{L}}{\gamma_1 \min\{\tilde{c}_1, 1\}}\right)^2 \|u - v\|^2_{B_h^{(\bar{m})}} \le$$

$$\le \left(1 + \frac{\tilde{L}}{\gamma_1 \min\{\tilde{c}_1, 1\}}\right)^2 \left(A_h^{(\bar{m})}(x,u) - A_h^{(\bar{m})}(x,v), u - v\right)_{\hat{\bar{\omega}}_h}.$$

Therefore [3, p.353], the iteration method (51) converges in the space $H_{B_h^{(\bar{m})}}$. As the norms $\|\cdot\|_{1,2,\hat{\bar{\omega}}_h}$, $\|\cdot\|_{B_h^{(\bar{m})}}$ are equivalent, then the error can be estimated as

$$\left\|y^{(\bar{m},n)} - y^{(\bar{m})}\right\|_{1,2,\hat{\bar{\omega}}_h} \le M_1 q^n.$$

In addition

$$\left|\left(k\frac{dy^{(\bar{m},n)}}{dx}\right)_0 - \left(k\frac{dy^{(\bar{m})}}{dx}\right)_0\right| \le \beta_1 \left|y_0^{(\bar{m},n)} - y_0^{(\bar{m})}\right| \le$$

$$\le M_1 \left\|y^{(\bar{m},n)} - y^{(\bar{m})}\right\|_{1,2,\hat{\bar{\omega}}_h},$$

$$\left|\left(k\frac{dy^{(\bar{m},n)}}{dx}\right)_N - \left(k\frac{dy^{(\bar{m})}}{dx}\right)_N\right| \le \beta_2 \left|y_N^{(\bar{m},n)} - y_N^{(\bar{m})}\right| \le$$

$$\le M_2 \left\|y^{(\bar{m},n)} - y^{(\bar{m})}\right\|_{1,2,\hat{\bar{\omega}}_h},$$

$$\left|\left(k\frac{dy^{(\bar{m},n)}}{dx}\right)_j - \left(k\frac{dy^{(\bar{m})}}{dx}\right)_j\right| \le \left|a_{j-1+\alpha}^{(\bar{m})}\right| \left|y_{\bar{x},j-1+\alpha}^{(\bar{m},n)} - y_{\bar{x},j-1+\alpha}^{(\bar{m})}\right| +$$

$$+ \left|Z_\alpha^{(m)j}(x_j, y^{(\bar{m},n)}) - Z_\alpha^{(m)j}(x_j, y^{(\bar{m})})\right| + \frac{1}{\left|V_\alpha^{(\bar{m})j}(x_j)\right|} \left|y_{j+(-1)^\alpha}^{(\bar{m},n)} - y_{j+(-1)^\alpha}^{(\bar{m})}\right| +$$

$$+ \frac{1}{\left|V_\alpha^{(\bar{m})j}(x_j)\right|} \left|Y_\alpha^{(\bar{m})j}(x_j, y^{(\bar{m},n)}) - Y_\alpha^{(\bar{m})j}(x_j, y^{(\bar{m})})\right| \le$$

$$\le M_3 \left\|y^{(\bar{m},n)} - y^{(\bar{m})}\right\|_{1,2,\hat{\bar{\omega}}_h} +$$

$$+ \left[\left|\frac{\partial}{\partial u} Z_\alpha^{(m)j}(x_j, u)\right|_{u=\tilde{y}} + \frac{1}{\left|V_\alpha^{(\bar{m})j}(x_j)\right|} \left|\frac{\partial}{\partial u} Y_\alpha^{(\bar{m})j}(x_j, u)\right|_{u=\bar{y}}\right] \times$$

$$\times \left\|y^{(\bar{m},n)} - y^{(\bar{m})}\right\|_{0,2,\hat{\bar{\omega}}_h} \le$$

$$\le M_3 \left\|y^{(\bar{m},n)} - y^{(\bar{m})}\right\|_{1,2,\hat{\bar{\omega}}_h}, \quad j = 1, 2, ..., N,$$

$$\left\| k\frac{dy^{(\bar{m},n)}}{dx} - k\frac{dy^{(\bar{m})}}{dx} \right\|_{0,2,\hat{\bar{\omega}}_h} \le M \left\| y^{(\bar{m},n)} - y^{(\bar{m})} \right\|_{1,2,\hat{\bar{\omega}}_h}.$$

Hence we get that

$$\left\| y^{(\bar{m},n)} - y^{(\bar{m})} \right\|^{*}_{1,2,\hat{\bar{\omega}}_h} \le Mq^n. \tag{55}$$

From the inequality (53), (55) implies the following estimate (52).

From a practical point of view to find a solution TDS (21), (22) will eventually need to use an iteration method of Newton. Linearizing (21), (22) taking into account the equality

$$\varphi^{(\bar{m})}(x_j, y^{(\bar{m})}) = \hbar_j^{-1} \sum_{\alpha=1}^{2} \left[\frac{h_{j-1+\alpha}}{2} f\left(x_{j+(-1)^\alpha}, y^{(\bar{m})}_{j+(-1)^\alpha}, \left.\frac{dy^{(\bar{m})}}{dx}\right|_{x=x_{j+(-1)^\alpha}} \right) \right] +$$
$$+ O\left(\frac{h_j^2 + h_{j+1}^2}{\hbar_j} \right), \quad j = 1, 2, ..., N-1,$$
$$\varphi^{(\bar{m})}(x_0, y^{(\bar{m})}) = f\left(x_1, y_1^{(\bar{m})}, \left.\frac{dy^{(\bar{m})}}{dx}\right|_{x=x_1} \right) + \frac{1}{\hbar_0}\mu_1 + O(h_1),$$
$$\varphi^{(\bar{m})}(x_N, y^{(\bar{m})}) = f\left(x_{N-1}, y_{N-1}^{(\bar{m})}, \left.\frac{dy^{(\bar{m})}}{dx}\right|_{x=x_{N-1}} \right) + \frac{1}{\hbar_N}\mu_2 + O(h_N),$$

$$\left.\frac{dy^{(\bar{m})}}{dx}\right|_{x=x_{j+(-1)^\alpha}} = y^{(\bar{m})}_{\bar{x},j-1+\alpha} + O\left(\frac{h^2_{j-1+\alpha}}{\hbar_j} \right),$$
$$j = 2-\alpha,\ 3-\alpha, ..., N+1-\alpha, \quad \alpha = 1, 2,$$

then modified Newton iteration method will be a form

$$\begin{gathered}
\left(a^{(\bar{m})} \nabla y_{\bar{x}}^{(\bar{m},n)} \right)_{\hat{x},j} + \frac{h_j}{2\hbar_j} \frac{\partial f\left(x_{j-1}, y_{j-1}^{(\bar{m},n-1)}, \left.\frac{dy^{(\bar{m},n-1)}}{dx}\right|_{x=x_{j-1}} \right)}{\partial u} \nabla y_{j-1}^{(\bar{m},n)} + \\
+ \frac{h_{j+1}}{2\hbar_j} \frac{\partial f\left(x_{j+1}, y_{j+1}^{(\bar{m},n-1)}, \left.\frac{dy^{(\bar{m},n-1)}}{dx}\right|_{x=x_{j+1}} \right)}{\partial u} \nabla y_{j+1}^{(\bar{m},n)} + \\
+ \frac{h_j}{2\hbar_j} \frac{\partial f\left(x_{j-1}, y_{j-1}^{(\bar{m},n-1)}, \left.\frac{dy^{(\bar{m},n-1)}}{dx}\right|_{x=x_{j-1}} \right)}{\partial \xi} \nabla y_{\bar{x},j}^{(\bar{m},n)} + \\
+ \frac{h_{j+1}}{2\hbar_j} \frac{\partial f\left(x_{j+1}, y_{j+1}^{(\bar{m},n-1)}, \left.\frac{dy^{(\bar{m},n-1)}}{dx}\right|_{x=x_{j+1}} \right)}{\partial \xi} \nabla y_{x,j}^{(\bar{m},n)} = \\
= -\varphi^{(\bar{m})}\left(x_j, y^{(\bar{m},n-1)} \right) - \left(a^{(\bar{m})} y_{\bar{x}}^{(\bar{m},n-1)} \right)_{\hat{x},j}, \quad j = 1, 2, ..., N-1,
\end{gathered} \tag{56}$$

$$\begin{aligned}
&\frac{1}{\hbar_0}\left(a_1^{(\bar m,n)}\nabla y_{x,0}^{(\bar m,n)}-\beta_1\nabla y_0^{(\bar m,n)}\right)+\\
&\quad+\frac{\partial f\left(x_1,y_1^{(\bar m,n-1)},\left.\frac{dy^{(\bar m,n-1)}}{dx}\right|_{x=x_1}\right)}{\partial u}\nabla y_1^{(\bar m,n)}+\\
&\quad+\frac{\partial f\left(x_1,y_1^{(\bar m,n-1)},\left.\frac{dy^{(\bar m,n-1)}}{dx}\right|_{x=x_1}\right)}{\partial \xi}\nabla y_{x,0}^{(\bar m,n)}=\\
&\quad=-\varphi^{(\bar m)}\left(x_0,y^{(\bar m,n-1)}\right)-\frac{1}{\hbar_0}\left(a_1^{(\bar m,n)}y_{x,0}^{(\bar m,n-1)}-\beta_1 y_0^{(\bar m,n-1)}\right),
\end{aligned}\tag{57}$$

$$\begin{aligned}
&-\frac{1}{\hbar_N}\left(a_N^{(\bar m,n)}\nabla y_{\bar x,N}^{(\bar m,n)}+\beta_2\nabla y_N^{(\bar m,n)}\right)+\\
&\quad+\frac{\partial f\left(x_{N-1},y_{N-1}^{(\bar m,n-1)},\left.\frac{dy^{(\bar m,n-1)}}{dx}\right|_{x=x_{N-1}}\right)}{\partial u}\nabla y_{N-1}^{(\bar m,n)}+\\
&\quad+\frac{\partial f\left(x_{N-1},y_{N-1}^{(\bar m,n-1)},\left.\frac{dy^{(\bar m,n-1)}}{dx}\right|_{x=x_{N-1}}\right)}{\partial \xi}\nabla y_{\bar x,N}^{(\bar m,n)}=\\
&\quad=-\varphi^{(\bar m)}\left(x_N,y^{(\bar m,n-1)}\right)+\frac{1}{\hbar_N}\left(a_N^{(\bar m,n)}y_{\bar x,N}^{(\bar m,n-1)}+\beta_2 y_N^{(\bar m,n-1)}\right),
\end{aligned}\tag{58}$$

$$y_j^{(\bar m,n)}=y_j^{(\bar m,n-1)}+\nabla y_j^{(\bar m,n)},\quad j=1,2,...,N-1,\quad n=1,2,... \tag{59}$$

4. Numerical examples

Example 1. Let us consider boundary value problem

$$\begin{gathered}
\frac{d^2u}{dx^2}=\pi^2\exp(u),\quad x\in(0,1),\\
\frac{du(0)}{dx}-u(0)=-\frac{\pi}{\sqrt{3}}+\ln 1,5,\quad -\frac{du(1)}{dx}-u(1)=-\sqrt{3}\pi-\ln 2,
\end{gathered}\tag{60}$$

with the exact solution

$$u(x)=-\ln\left(2\cos^2\left(\frac{\pi}{2}\left(x-\frac{1}{3}\right)\right)\right).$$

Since $f(x,u,\xi)=-\pi^2\exp(u)$ it follows that condition (5) is satisfied if we take $f_0(x)\equiv 0, c(t)=\pi^2\exp(t)$, $g(x)\equiv 1$. Besides we have

$$[f(x,u,\xi)-f(x,v,\eta)]\,(u-v)=-\pi^2\exp(\theta u+(1-\theta)v)(u-v)^2\le 0,\quad 0<\theta<1.$$

Thus, due to Theorem 1 the problem has a unique solution.

For numerical solution of problem (60) on the equidistance grid $\bar{\omega}_h = \{x_j = jh, j = 0, 1, ..., N, h = 1/N\}$ we use TDS of the sixth order of accuracy ($m = 6$)

$$\begin{aligned} & y^{(6)}_{\bar{x}x,j} = -\varphi^{(6)}(x_j, y^{(6)}), \quad j = 1, 2, \ldots, N-1, \\ & \frac{2}{h}\left(y^{(6)}_{x,0} - \beta_1 y^{(6)}_0\right) = -\varphi^{(6)}(x_0, y^{(6)}), \\ & -\frac{2}{h}\left(y^{(6)}_{\bar{x},N} + \beta_2 y^{(6)}_N\right) = -\varphi^{(6)}(x_N, y^{(6)}), \end{aligned} \tag{61}$$

with

$$\varphi^{(6)}(x_j, u) = h^{-1}\sum_{\alpha=1}^{2}(-1)^\alpha\left[Z^{(6)j}_\alpha(x_j, u) + (-1)^\alpha \frac{Y^{(6)j}_\alpha(x_j, u) - u_{j+(-1)^\alpha}}{h}\right],$$

$$\varphi^{(6)}(x_0, u) = \frac{2}{h}\left[Z^{(6)0}_2(x_0, u) + \frac{Y^{(6)0}_2(x_0, u) - u_1}{h} + \mu_1\right],$$

$$\varphi^{(6)}(x_N, u) = \frac{2}{h}\left[-Z^{(6)N}_1(x_N, u) + \frac{Y^{(6)N}_1(x_N, u) - u_{N-1}}{h} + \mu_2\right],$$

$$\beta_1 = 1, \quad \beta_2 = 1, \quad \mu_1 = \frac{\pi}{\sqrt{3}} - \ln 1,5, \quad \mu_2 = \sqrt{3}\pi + \ln 2,$$

and $Y^{(6)j}_\alpha(x, u), Z^{(6)j}_\alpha(x, u)$ are numerical solutions of initial value problems

$$\begin{aligned} & \frac{dY^j_\alpha(x, u)}{dx} = Z^j_\alpha(x, u), \quad \frac{dZ^j_\alpha(x, u)}{dx} = -f\left(x, Y^j_\alpha(x, u), Z^j_\alpha(x, u)\right), \\ & \qquad x_{j-2+\alpha} < x < x_{j-2+\alpha}, \\ & Y^j_\alpha\left(x_{j+(-1)^\alpha}, u\right) = u_{j+(-1)^\alpha}, \quad Z^j_\alpha\left(x_{j+(-1)^\alpha}, u\right) = \left.\frac{du}{dx}\right|_{x=x_{j+(-1)^\alpha}}, \end{aligned} \tag{62}$$

$$j = 2-\alpha, 3-\alpha, ..., N+1-\alpha, \quad \alpha = 1, 2$$

computed by a explicit Runge-Kutta method of the sixth-order of accuracy (see Table 6.1 [10, p.189]).

To determine the solution of the difference scheme (61) the modified Newton method (56)-(59) will be used. System linear algebraic equations (56)-(58) for the unknowns $\nabla y^{(6,n)}(x)$, $x \in \hat{\omega}_h$ we solved by Gaussian elimination for linear system with a tridiagonal matrix.

Numerical results are given in Table 1. To evaluate the convergence rate in practice, we introduced the following quantities

$$er = \left\|z^{(6)}\right\|^*_{1,2,\bar{\omega}_h} = \left\|y^{(6)} - u\right\|^*_{1,2,\bar{\omega}_h}, \quad p = \log_2 \frac{\left\|z^{(6)}\right\|^*_{1,2,\bar{\omega}_h}}{\left\|z^{(6)}\right\|^*_{1,2,\bar{\omega}_{h/2}}}.$$

In the following example the implementation of the TDS uses the $h - h/2$ a posteriori estimation to achieve a given accuracy EPS. The comparison with the true error Er shows that this accuracy is actually achieved.

TABL. 1. Numerical results for problem (60).

N	Er	p
16	$0,2241 \cdot 10^{-5}$	
32	$0,3522 \cdot 10^{-7}$	6
64	$0,5514 \cdot 10^{-9}$	6
128	$0,8642 \cdot 10^{-11}$	6

Example 2. Let us consider the boundary value problem

$$\begin{gathered} \frac{d^2u}{dx^2} = 3u\frac{du}{dx}, \quad x \in (0,1), \\ \frac{du(0)}{dx} = -1,5/\cosh^2(0,75), \\ -\frac{du(1)}{dx} - u(1) = 1,5/\cosh^2(0,75) + \tanh(0,75). \end{gathered} \tag{63}$$

The exact solution is $u(x) = \tanh\left(\frac{3(1-2x)}{4}\right)$.

The numerical results which have been obtained for difference scheme of order of accuracy 6 are given in Table 2

TABL. 2. Numerical results for problem (63).

EPS	N	Er
10^{-4}	2048	$0,1323 \cdot 10^{-5}$
10^{-6}	2048	$0,4816 \cdot 10^{-7}$
10^{-8}	4096	$0,4078 \cdot 10^{-9}$

BIBLIOGRAPHY

1. Gnativ L. B. Generalized three-point difference schemes of high order of accuracy for nonlinear ordinary differential equations of second order / L. B. Gnativ, M. V. Kutniv, A. I. Chukhrai // Journal of Mathematical Sciences.– 2010.– Vol. 167, No 1.– P. 62-75.
2. Trenogin V. A. Functional Analysis / V. A. Trenogin.– Moscow: Nauka, 1980 (in Russian).
3. Samarskii A. A. Numerical Methods for Grid Equations Vol. 2, Iterative Methods / A. A. Samarskii, E. S. Nikolaev.– Basel, Boston, Berlin: Birkhäuser Verlag, 1989.
4. Gavrilyuk I. P. Exact and Truncated Difference Schemes for Boundary Value ODEs. (International Seris of Numerical Mathematics Vol. 159) / I. P. Gavrilyuk, M. Hermann, V. L. Makarov, M. V. Kutniv.– Basel: Springer AG, 2011.
5. Kutniv M. V. Accurate three-point difference schemes for second-order monotone ordinary differential equations and their implementation / M. V. Kutniv // Computational Mathematics and Mathematical Physics.– 2000.– Vol. 40, No 3.– P. 368-382.
6. Kutniv M. V. Modified three-point difference schemes of high-accuracy order for second order nonlinear ordinary differential equations / M. V. Kutniv // Computational Methods in Applied Mathematics.– 2003.– Vol. 3, No 2.– P. 287-312.

7. Kutniv M. V. Accurate three-point difference schemes for second-order nonlinear ordinary differential equations and their implementation / M. V. Kutniv, V. L. Makarov, A. A. Samarskii // Computational Mathematics and Mathematical Physics. – 1999. – Vol. 39, No 1. – P. 45-60.
8. Makarov V. L. Exact three-point difference schemes for second-order nonlinear ordinary differential equations and their implementation / V. L. Makarov, A. A. Samarskii // Soviet Math. Dokl. – 1991. – Vol. 41, No 4. – P. 495-500.
9. Gnativ L. B. Exact three-point difference schemes for second order nonlinear differential equations with boundary conditions of the third kind / L. B. Gnativ, M. Król, M. V. Kutniv // J. Numer. Appl. Math. – 2012. – No 3 (109). – P. 34-52.
10. Hairer E. Solving Ordinary Differential Equations I, Nonstiff Problems / E. Hairer, S. P. Nørsett, G. Wanner. – Berlin, Heidelberg, New York: Springer Verlag, 1987.

Marta Król, Myroslav Kutniv,
Rzeszow University of Technology,
8, Powstancow Warszawy str., 35-959, Rzeszow, Polska

Received 21.05.2014

Журнал обчислювальної та прикладної математики | 2014 № 2 (116) | Journal of Computational & Applied Mathematics

UDC 519.6

INTERPOLATING FUNCTIONAL POLYNOMIAL FOR THE APPROXIMATE SOLUTION OF THE BOUNDARY VALUE PROBLEM

Volodymyr Makarov, Igor Demkiv

Резюме. У роботі, застосовуючи функціональний поліном Ньютона побудований на континуальній множині вузлів, будується інтерполяційний функціональний поліном n-го порядку для наближення до розв'язку крайової задачі другого порядку.

Abstract. Interpolating functional polynomial of order for the approximation to the solution of the boundary value problem of the second order is constructed and justified in this paper. This is done using Newton functional polynomial constructed on a continual set of knots.

1. Introduction

Many authors investigated the generalization of the classical theory of one variable functions interpolation to the case of nonlinear functionals and operators (see for example [1, 2, 3, 4, 5, 6, 7, 8]). In particular, in [9] it is suggested to seek for Newton-type interpolants in the class of functional polynomials of the following form

$$P_n(x(\cdot)) = K_0 + \\ + \sum_{s=1}^{n} \int_0^1 \int_{z_1}^1 ... \int_{z_{s-1}}^1 K_s(\overrightarrow{z}^s) \prod_{i=1}^{s} [x(z_i) - x_{i-1}(z_i)]\, dz_s \dots dz_1, \tag{1}$$

where $x_i\,(z) \in Q[0,1]$, $i = 0, 1, \dots$ are arbitrary, fixed elements from the space $Q[0,1]$. Which is a space of piecewise continuous on the interval $[0,1]$ functions with a finite number of discontinuity points of the first kind. For determination of the kernels K_0, $K_s\,(\overrightarrow{z}^s)$, $s = \overline{1,n}$ a following continual set of knots

$$x^n\left(z, \vec{\xi}^n\right) = x_0\,(z) + \sum_{i=1}^{n} H\,(z - \xi_i)\,[x_i\,(z) - x_{i-1}\,(z)]\,, \quad z \in [0,1]\,, \tag{2}$$

$$\vec{\xi}^n = (\xi_1, \xi_2, \dots, \xi_n) \in \overline{\Omega}_n = \\ = \{\vec{z}^n = (z_1, z_2, \dots, z_n) : 0 \le z_1 \le z_2 \le \dots \le z_n \le 1\}\,,$$

was introduced and continual interpolation conditions of the form

Key words. Newton's functional polynomial, continual set of nodes, boundary value problem, interpolating functional polynomial.

$$P_n^I\left(x^n\left(\cdot,\vec{\xi^n}\right)\right)=F\left(x^n\left(\cdot,\vec{\xi^n}\right)\right)\ ,\quad \forall\vec{\xi^n}\in\overline{\Omega}_n,$$

were set, where $H(z)$ is a Heaviside function.

In the above-mentioned work, it was shown that the necessary conditions for polynomial (1) to be interpolating on the continual knots (2) are the determination of its kernels according to the following formulas

$$K_0=F\left(x_0\left(\cdot\right)\right),$$

$$K_s\left(\vec{z^s}\right)=(-1)^s\prod_{i=1}^{s}\left[x_i\left(z_i\right)-x_{i-1}\left(z_i\right)\right]^{-1}\frac{\partial^s}{\partial z_1\ldots\partial z_s}F\left(x^s\left(\cdot,\vec{z^s}\right)\right)\ ,$$

$$s=\overline{1,n}\,.$$

To ensure sufficient condition for polynomial $P_n\left(x\left(\cdot\right)\right)$ to be interpolating on continual knots (2) the following substitution rules satisfaction

$$\frac{\partial^p}{\partial z_1\partial z_2\ldots\partial z_p}\left[F\left(x^{p+1}\left(\cdot,\vec{z}^{p+1}\right)\right)\Big|_{z_{p+1}=z_p}\right]=$$
$$=\left[\frac{\partial^p}{\partial z_1\partial z_2\ldots\partial z_p}F\left(x^{p+1}\left(\cdot,\vec{z}^{p+1}\right)\right)\right]\Bigg|_{z_{p+1}=z_p}\frac{x_{p+1}\left(z_p\right)-x_{p-1}\left(z_p\right)}{x_p\left(z_p\right)-x_{p-1}\left(z_p\right)}\ ,\quad p=\overline{1,n-1} \tag{3}$$

were required.

The purpose of this paper is to develop and study the interpolating functional polynomial for approximation of the solution of the second order boundary value problem.

2. Statement of the problem

One must apply the Newton type functional polynomial of the form (1), (2) and construct the approximation to the solution of the following boundary value problem.

$$U''\left(x;q\left(\cdot\right)\right)-q\left(x\right)U\left(x;q\left(\cdot\right)\right)=-f\left(x\right),\quad x\in\left(0,1\right), \tag{4}$$

$$U\left(0;q\left(\cdot\right)\right)=0,\quad U\left(1;q\left(\cdot\right)\right)=0. \tag{5}$$

3. Solution of the problem

When the function $f(x)$ is fixed, one can consider solution of the problem (4), (5) as non-linear operator with respect to $q(x)$. We introduce the following continual interpolating knots

$$q^n\left(x,\overrightarrow{\xi}^n\right)=\sum_{i=1}^{n}\frac{1}{n}H\left(x-\xi_i\right), \tag{6}$$

where

$$0\le\xi_1\le\xi_2\le\ldots\le\xi_n\le1 \tag{7}$$

and the frame of these knots are

$$q_i(x) = \frac{i}{n}, \quad i = \overline{0, n}.$$

Let us write the following $n-$ degree interpolating functional polynomial of Newton type

$$U_n(x; q(\cdot)) = \sum_{i=0}^{n} \int_0^1 \int_{z_1}^1 \dots \int_{z_{i-1}}^1 K_i(x; q(\cdot)) \prod_{p=1}^{i} n\left(q(z_p) - \frac{p}{n}\right) d\overrightarrow{z}_p, \tag{8}$$

where

$$K_i(x; q(\cdot)) = (-1)^i \frac{\partial^i}{\partial z_1 \dots \partial z_i} U\left(x; q^i\left(x; \overrightarrow{z}^i\right)\right), \tag{9}$$

$$i = \overline{1, n}, \quad K_0(x; q(\cdot)) = U(x; 0).$$

According to Theorem 2.1 from [9] the necessary and sufficient condition for polynomial (8), (9) to be interpolating for solution of the boundary problem (4), (5) on a continual set of interpolating knots (6), (7), i.e. the following conditions were met

$$U\left(x; q^n\left(\cdot, \overrightarrow{\xi}^n\right)\right) = U_n\left(x; q^n\left(\cdot, \overrightarrow{\xi}^n\right)\right), \quad \forall \overrightarrow{\xi}^n \in \Omega_n, \tag{10}$$

is the following substitution rules to be applicable

$$\left[\frac{\partial}{\partial \xi_{i-1}} U\left(x; q^i\left(\cdot, \overrightarrow{\xi}^i\right)\right)\right]_{\xi_i = \xi_{i-1}} = \frac{1}{2} \frac{\partial}{\partial \xi_{i-1}} U\left(x; q^i\left(\cdot; \overrightarrow{\xi}^i\Big|_{\xi_i = \xi_{i-1}}\right)\right), \quad i = \overline{2, n}. \tag{11}$$

The following statement is fulfilled.

Lemma 1. *Let the solution of boundary value problem (4), (5) be considered as non-linear operator with respect to $q(x)$. Then it satisfies the substitution rule (11).*

Proof. Consider the following boundary problem

$$U''\left(x; q^i\left(\cdot; \overrightarrow{\xi}^i\right)\right) - \sum_{p=1}^{i} \frac{1}{n} H(x - \xi_p) U\left(x; q^i\left(x; \overrightarrow{\xi}^i\right)\right) = -f(x), \quad x \in (0, 1), \tag{12}$$

$$U\left(0; q^i\left(\cdot; \overrightarrow{\xi}^i\right)\right) = 0, \quad U\left(1; q^i\left(\cdot; \overrightarrow{\xi}^i\right)\right) = 0. \tag{13}$$

As consequences from (12), (13) we have following two boundary value problems with the same differential operator

$$\frac{d^2}{dx^2}\left[\frac{\partial}{\partial \xi_{i-1}} U\left(x; q^i\left(\cdot; \overrightarrow{\xi}^i\right)\right)\right]_{\xi_i = \xi_{i-1}} -$$

$$-\sum_{p=1}^{i}\frac{1}{n}\,H\left(x-\xi_p\right)\Big|_{\xi_i=\xi_{i-1}}\left[\frac{\partial}{\partial\xi_{i-1}}U\left(x;q^i\left(\cdot;\overrightarrow{\xi}^{\,i}\right)\right)\right]_{\xi_i=\xi_{i-1}}= \tag{14}$$

$$=\frac{1}{n}\frac{d}{d\xi_{i-1}}\,H\left(x-\xi_{i-1}\right)U\left(x;q^i\left(\cdot,\overrightarrow{\xi}^{\,i}\right)\right)\Big|_{\xi_i=\xi_{i-1}},$$

$$\left.\left[\frac{\partial}{\partial\xi_{i-1}}U\left(x;q^i\left(\cdot;\overrightarrow{\xi}^{\,i}\right)\right)\right]_{\xi_i=\xi_{i-1}}\right|_{x=0,1}=0, \tag{15}$$

$$\frac{d^2}{dx^2}\left[\frac{\partial}{\partial\xi_{i-1}}U\left(x;q^i\left(\cdot;\overrightarrow{\xi}^{\,i}\right)\right)\right]_{\xi_i=\xi_{i-1}}-$$

$$-\sum_{p=1}^{i}\frac{1}{n}\,H\left(x-\xi_p\right)\Big|_{\xi_i=\xi_{i-1}}\frac{\partial}{\partial\xi_{i-1}}U\left(x;q^i\left(\cdot;\overrightarrow{\xi}^{\,i}\Big|_{\xi_i=\xi_{i-1}}\right)\right)= \tag{16}$$

$$=\frac{2}{n}\frac{d}{d\xi_{i-1}}H\left(x-\xi_{i-1}\right)U\left(x;q^i\left(\cdot,\overrightarrow{\xi}^{\,i}\Big|_{\xi_i=\xi_{i-1}}\right)\right),$$

$$\left.\frac{\partial}{\partial\xi_{i-1}}U\left(x;q^i\left(\cdot;\overrightarrow{\xi}^{\,i}\Big|_{\xi_i=\xi_{i-1}}\right)\right)\right|_{x=0,1}=0. \tag{17}$$

Note that right hand sides of their differential equations differ only by numerical multiplier. Comparison of boundary value problems (14), (15) and (16), (17) proves the lemma.

To construct the interpolant (8), (9) one must find the solution of the problems (12), (13) at $i=\overline{0,n}$. Then we have

$$U\left(x;0\right)=K_0\left(x;q\left(\cdot\right)\right)=\int_0^1 G_0\left(x,\xi\right)f\left(\xi\right)d\xi,$$

$$U\left(x;q^i\left(\cdot;\overrightarrow{\xi}^{\,i}\right)\right)=\int_0^1 G_i\left(x,\xi\right)f\left(\xi\right)d\xi,\qquad i=\overline{1,n},$$

where $G_i\left(x,\xi\right),\ i=\overline{0,n}$ are Green's functions of the corresponding boundary value problems

$$G_0\left(x,\xi\right)=\begin{cases}x\left(1-\xi\right), & 0\le x\le\xi,\\ \xi\left(1-x\right), & \xi\le x\le 1,\end{cases}$$

.

$$G_i\left(x,\xi\right)=\frac{1}{V_{1,i}\left(1\right)}\begin{cases}V_{1,i}\left(x\right)V_{2,i}\left(\xi\right), & 0\le x\le\xi,\\ V_{2,i}\left(x\right)V_{1,i}\left(\xi\right), & \xi\le x\le 1.\end{cases}$$

Here $V_{1,i}\left(x\right),\ V_{2,i}\left(x\right)$ are solutions of the following Cauchy problems:

$$\frac{d^2V_{\alpha i}\left(x\right)}{dx^2}-\sum_{p=1}^{i}\frac{1}{n}H\left(x-\xi_p\right)V_{\alpha i}\left(x\right)=0,\qquad x\in\left(0,1\right),\qquad \alpha=1,2;$$

$$V_{1i}\left(0\right)=0;\qquad \frac{dV_{1i}\left(0\right)}{dx}=1;\qquad V_{2i}\left(1\right)=0;\qquad \frac{dV_{2i}\left(1\right)}{dx}=-1.$$

It is quite simple to find functions $V_{1,i}(x)$, $V_{2,i}(x)$ in explicit form because the differential equations which they satisfy have a piecewise constant coefficient. In particular at $i = 1$ we obtain

$$V_{11}(x) = \begin{cases} x, & 0 \le x \le \xi_1, \\ \sqrt{n}\sinh\dfrac{1}{\sqrt{n}}(x-\xi_1) + x\cosh\dfrac{1}{\sqrt{n}}(x-\xi_1), & \xi_1 \le x \le 1, \end{cases}$$

$$V_{21}(x) = \begin{cases} \sqrt{n}\sinh\dfrac{1}{\sqrt{n}}(1-x), & \xi_1 \le x \le 1, \\ -\cosh\dfrac{1}{\sqrt{n}}(1-\xi_1)(x-\xi_1) + \sqrt{n}\sinh\dfrac{(1-\xi_1)}{\sqrt{n}}, & 0 \le x \le \xi_1. \end{cases}$$

4. Conclusions

Thus, Newton type interpolating functional polynomial of $n-$degree of form (8), (9) was obtained. This polynomial will be the approximation to the solution of the boundary value problem (4), (5).

Bibliography

1. Ulm S. On generalized divided differences / S. Ulm // Izv AN ESSR. Ser. Phys. and math. sciences. – 1967. – V. 16, No 1. – P. 13-26 (in Russian).
2. Prenter P. M. Lagrange and Hermite interpolation in Banach spaces / P. M. Prenter // Appr. Theory. – 1971. –Vol. 4, No 4. – P. 419-432.
3. Porter W. A. Data interpolation: causality structure and system identification / W. A. Porter // Inf. and Contr. – 1975. – Vol. 29, No 3. – P. 217-233.
4. Kergin P. A natural interpolation of functions / P. Kergin // J. Approx. Theory. – 1980. – Vol. 19, No 4. – P. 278-293.
5. Micchelli C. A. A constructive approach to Kergin interpolation in Rk / C. A. Micchelli // Rocky Mountain J. Math. – 1980. – No 10. – P. 485-497.
6. Andersson M. Complex Kergin Interpolation / M. Andersson, M. Passare // J. Approx. Theory. – 1991. – No 64. – P. 214-225.
7. Filipsson L. Kergin interpolation in Banach spaces / L. Filipsson // J. Approx. Theory. – 2004. – No 127. – P. 108-123.
8. Yanovich L. A. Approximate calculation of continual integrals on Gauss measures / L. A. Yanovich. – Mn: Sciense and technics, 1976. – 384 p. (in Russian).
9. Makarov V. L. Necessary and sufficient conditions for the existence of interpolating functional polynomial on continual set of knots / V. L. Makarov, I. I. Demkiv, B. R. Mykhalchuk // Reports of NAS of Ukraine. – 2003. – No 7. – P. 7-12.

Volodymyr Makarov,
Institute of Mathematics, National Academy of Sciences,
3, Tereschenkivs'ka Str., Kyiv, 01601, Ukraine;

Igor Demkiv,
National University "Lvivska Politechnika",
12 S. Bandery Str., Lviv, 79013, Ukraine

Received 21.05.2014

Журнал обчислювальної та прикладної математики | 2014 № 2 (116) | Journal of Computational & Applied Mathematics

UDC 519.6

FD-METHOD FOR SOLVING THE STURM–LIOUVILLE PROBLEM WITH POTENTIAL THAT IS THE DERIVATIVE OF A FUNCTION OF BOUNDED VARIATION

Volodymyr Makarov, Nataliia Romaniuk, Igor Lazurchak

Резюме. Розглядається скалярна задача Штурма-Ліувілля з потенціалом, що є похідною від функції обмеженої варіації, та крайовими умовами Діріхле. Викладена основа реалізації FD-методу у випадку, коли функція $\bar{q}(x)$, що наближає потенціал $q(x)$, є тотожнім нулем, а також у загальному випадку. Встановлені достатні умови суперекспоненціальної збіжності FD-методу та оцінки його точності, які є значним посиленням та узагальненням відповідних результатів, отриманих в попередніх роботах.

Abstract. We consider a scalar Sturm-Liouville problem with the Dirichlet boundary conditions where the potential $q(x)$ is assumed to be a derivative of the function with bounded variation. The application of the abstract FD-method scheme to such eigenvalue problem is studied in the scope of this work. In addition to the general case when the function $\bar{q}(x)$ approximating $q(x)$ is assumed to be arbitrary we study the case when $\bar{q}(x)$ is equal to zero everywhere. We obtain new sufficient conditions for the super-exponential convergence of the FD-method and its accuracy estimates which essentially generalize similar results obtained in the earlier works.

MSC 2010: 65L15, 65L20, 34B09, 34B24, 34L16, 34L20.

1. Introduction

Most of the current technological and industrial advancements in electronics rely on the increasingly rigorous quantum-mechanical models. The models where the discontinuities of the potential are essential to represent the modelled phenomena and can not be disregarded. Mathematically such models can be represented as follows (*the one particle, many center Hamiltonian*):

$$\mathcal{H} = -\Delta + \sum_{\alpha \in \aleph} \gamma_\alpha \delta_\alpha(\cdot), \quad (1)$$

where Δ is a Laplace operator in $L^2(R^d)$, d stands for the dimension of the configuration space, $\aleph$ is a discrete, countable at most, subset in R^d, $\delta_\alpha(\cdot)$ is a Dirac delta function at the point α (i.e. a single measure concentrated at α) (see [1]). $\mathcal{H}$ describes the energy of the quantum mechanical particle which moves under the influence of an "interaction potential" created by the "point source" forces γ_α, located at α. We will denote this function as $\delta(x)$ and refer to it as Dirac delta function (**DDF**).

Key words. Sturm-Liouville problem, Dirac delta function potential, distribution potential, functional-discrete method, super-exponential convergence rate.

Dirac delta function (DDF) *potentials* had been used for modelling of atomic and molecular systems including atomic lattices, quantum heterostructures, semiconductors, organic fluorescent materials, solar cells etc. (see [1, 2, 3] and citations of them). Among recent applications of (1) one may mention the novel structure of quantum waveguide [2] based on the modelling with the same potential as in (1) having the finite numbers of delta functions. This type of potentials are called *Dirac comb* by the authors of [2]. History of the studies, mathematical properties and the visualization for some of the models involving such discontinuous potentials as well as various physical applications are summarized in [3].

Linear Sturm-Liouville problem with *distribution potentials* are extensively studied theoretically (for example see [4]). The authors of [5] derive the total regularized trace formula of differential Sturm–Liouville operators on a finite closed interval with singular potentials $q(x)$ that are not locally integrable functions and such that $\int q(x)dx \in BV_c[0,\pi]$ in the sense of distributions (the definition of $BV_c[0,\pi]$ will be given shortly). During the technical revision of [5] author of [6] found a simple proof for the case of potential $q\left(x\right)=\delta\left(x-\frac{\pi}{2}\right)$. Note that if $q\left(x\right)\in L_1$ then Theorem 1 from [5] contains the results of [7]. Independently from [5] the authors of [8] received the spectral asymptotic and the trace formula on the interval $[0,l]$ for the class of potentials, which may contain finite of sum δ-functions.

In the current paper we study an eigenvalue problem for the Hamiltonian having the form (1) with $d=1$, $\aleph=\{\alpha\}$, $\alpha\in(0,1)$, which is stated as follows:

$$\frac{d^2u(x)}{dx^2}+\left(\lambda-q(x)\right)u(x)=0,\ x\in(0,1),\ u(0)=0,\ u(1)=0, \tag{2}$$

where

$$q(x)=\frac{d\sigma(x)}{dx}$$

and $\sigma(x)$ is a function of bounded variation.

We start by summarizing some useful facts from the real analysis. Since $\sigma(x)$ is the function of bounded variation, the following representation is valid:

$$\sigma(x)=h(x)+\psi(x)+\chi(x),$$

with $h(x)$, $\psi(x)$, $\chi(x)$ being the jump function, the absolutely continuous function and the singular function correspondingly (see. [9], p.347). The singular part $\chi(x)$ has at most countable number of discontinuities which coincide with those of the jump function $h(x)$. Let us enumerate these discontinuity points in the ascending order and denote them as $x_p\in(0,1),\ p=1,2,\dots,\ x_1<x_2<\dots$, then $h(x)=\sum_p\gamma_p H\left(x-x_p\right)$, where γ_p are real numbers, $H(z)$ is the Heaviside function. From now on we assume that $\sigma(x)$ belongs to the class $BV_c[0,1]$. That is the class of functions with bounded variation and which are right continuous at any point $x\in(0,1)$ and continuous at the endpoints $x=0$ and $x=1$.

An essential role in the proof of FD-method's convergence rely on the following result:

Theorem 1. *([10], p.481) Let $\sigma(x) \in BV_c[0,1]$ and a function $f(x)$ be continuous on the segment $[0,1]$, then the following inequality holds true:*

$$\left| \int_0^1 f(x) d\sigma(x) \right| \le \max_{x \in [0,1]} |f(x)| \, \|\sigma\|_v \, ,$$

where $\|\sigma\|_v = \operatorname{var}\{\sigma(x); 0, 1\}$.

Due to the importance of the model there exist a large number of software packages for the numerical solution of the singular scalar Sturm-Liouville problems. Most notable FORTRAN packages are SL02F [11] and SLEDGE [12] implementing the Pruess method, SLEIGN [13, 14] and SLEIGN2 [15] – shooting method based on the Prüfer transformation. MATSLISE package [16] implements the Constant Perturbation Methods (CPM) and the Line Perturbation Methods (LPM) in MATLAB.

The code of SLEIGN2 became a considerable improvement of SLEIGN code. It covers more problem cases than other software packages, existent at that moment. Among other things the developers of SLEIGN2 expand the list of singular self-adjoint problems compatible with the package. Such list along with problem's classification, numerical examples and the package documentation can be found in [15]. The mentioned FORTRAN codes is available as a part of SLTSTPAK package (see [17]). Its implementation details as well as 60 test problem application examples are given [18]. Taking in to account the joint interest from different application areas, and the lack of common interface for the mentioned software packages the developers (V. Ledoux and rest of authors) created MATSLISE. It offers an interactive graphical user interface for various Sturm-Liouville problem solvers and the ability to control the parameters of the solver on-the-fly. Aside of that it contains some useful solution visualization tools (see [19]).

In spite of the large amount of implementations none of the mentioned packages can handle DDF potentials directly.

The purpose of the current work is to study, justify and propose algorithm implementation of the FD-method for eigenvalue problem for the Sturm-Liouville operator (2) with the potential being the derivative of the function with bounded variation such as

$$q(x) = \sum_{p=1}^{k} \gamma_p \delta(x - x_p) + \psi'(x) \, , \; x_p \in (0,1), \; p = \overline{1,k}.$$

The results, presented here, extends the results reported in [20] in the linear case ($N(u) \equiv 0$), where the potential $q(x)$ have only one singularity ($k = 1$). Aside of that the current work contains the generalization of section 5 from [21], where the FD-method (with $\bar{q}(x) \equiv 0$) considered in application to (2) with $q(x) = a\delta\left(x - \frac{1}{2}\right), \; a > 0$.

In section 2 we apply the simplest version of the FD-method, when the function $\bar{q}(x)$, approximating the potential $q(x)$, is zero everywhere. The necessary conditions of the applied method's convergence is given. We show that under such conditions the method will converge super-exponentially. The practical implications of the technique proposed here lie in the fact that theoretical estimates on the lowest eigenvalue number for which the method is justified to converge, are more close to the number obtained experimentally. It may be considered as an improvement of the similar conditions from theorem 1 [8]. In the end of the section we present some numerical experiments to justify our theoretical results. The algorithm of general FD-method scheme ($\bar{q}(x) \not\equiv 0$) along with its justification is given in section 3. The results of a numerical calculation presented in the end of the section illustrate the effectiveness of the proposed algorithm.

2. FD-method for $\bar{q}(x) \equiv 0$

To find the approximate solution of the problem (2) we shall apply the FD-method of the m-th rank with the function $\bar{q}(x) \equiv 0$. Detailed justification for the choice of the FD-method scheme used here will be given in section 3 dealing with the general case $\bar{q}(x) \not\equiv 0$. The m-th rank approximate solution will be sought in the form of a finite sum

$$\overset{m}{u}_n(x) = \sum_{j=0}^{m} u_n^{(j)}(x), \ \overset{m}{\lambda}_n = \sum_{j=0}^{m} \lambda_n^{(j)}, \tag{3}$$

where every summand in (3) is obtained from the solution of the recurrent sequence of problems

$$\begin{aligned} &\frac{d^2 u_n^{(j+1)}(x)}{dx^2} + \lambda_n^{(0)} u_n^{(j+1)}(x) = -\sum_{p=0}^{j} \lambda_n^{(j+1-p)} u_n^{(p)}(x) + q(x) u_n^{(j)}(x), \\ &u_n^{(j+1)}(0) = 0, \ u_n^{(j+1)}(1) = 0, \ x \in (0,1), \ j = 0,1,...,m-1, \\ &u_n^{(0)} = \sqrt{2}\sin(n\pi x), \ \lambda_n^{(0)} = (n\pi)^2, \end{aligned} \tag{4}$$

supplied by the solvability condition

$$\lambda_n^{(j+1)} = -\sum_{p=1}^{j} \lambda_n^{(j+1-p)} \int_0^1 u_n^{(p)}(x)\, u_n^{(0)}(x) dx + \int_0^1 q(x) u_n^{(j)}(x)\, u_n^{(0)}(x) dx$$

and the following orthogonality condition

$$\int_0^1 u_n^{(j+1)}(x)\, u_n^{(0)}(x) dx = 0,$$

which guaranties the uniqueness of the solution to (4). Let us represent the solution to (4) using the generalized Green's function approach:

$$u_n^{(j+1)}(x) = \int_0^1 g_n(x,\xi)\left[-\sum_{p=0}^{j}\lambda_n^{(j+1-p)}u_n^{(p)}(\xi)+q(\xi)u_n^{(j)}(\xi)\right]d\xi =$$
$$= -\sum_{p=0}^{j}\lambda_n^{(j+1-p)}\int_0^1 g_n(x,\xi)u_n^{(p)}(\xi)d\xi + \int_0^1 g_n(x,\xi)u_n^{(j)}(\xi)d\sigma(\xi), \quad (5)$$
$$\lambda_n^{(j+1)} = \int_0^1 q(\xi)u_n^{(j)}(\xi)u_n^{(0)}(\xi)d\xi = \int_0^1 u_n^{(j)}(\xi)u_n^{(0)}(\xi)d\sigma(\xi),$$

where

$$g_n(x,\xi) = \left[\frac{(x-H(x-\xi))\cos(n\pi x)}{\pi n} - \frac{\sin(n\pi x)}{2\pi^2 n^2}\right]\sin(n\pi\xi) +$$
$$+\frac{\sin(n\pi x)(\xi - H(\xi - x))\cos(n\pi\xi)}{\pi n} = g_{n,1}(x,\xi)+g_{n,2}(x,\xi),$$
$$g_{n,1}(x,\xi) = \frac{(x-H(x-\xi))\cos(n\pi x)}{\pi n}\sin(n\pi\xi) + \quad (6)$$
$$+\frac{\sin(n\pi x)(\xi-H(\xi-x))}{\pi n}\cos(n\pi\xi),$$
$$g_{n,2}(x,\xi) = -\frac{\sin(n\pi x)}{2\pi^2 n^2}\sin(n\pi\xi).$$

The generalized Green's function $g_n(x,\xi)$ has the following properties:

$$g_n(x,\xi) = g_n(\xi,x), \ g_n(x,\xi) = g_n(1-x,1-\xi),$$
$$\int_0^1 g_n(x,\xi)\sin(n\pi x)dx = 0, \ \int_0^1 g_n(x,\xi)\sin(n\pi\xi)d\xi = 0, \quad (7)$$
$$|g_n(x,\xi)| \le \frac{1}{\pi n}+\frac{1}{2(\pi n)^2} \le \frac{7}{6\pi n}.$$

Representation (5) along with the properties of Green function (7) and the results of theorem 1 allows us to obtain the following recurrent system of inequalities

$$\left\|u_n^{(j+1)}\right\|_\infty \le \|g_n\|_\infty \left(\sum_{p=1}^{j}\left|\lambda_n^{(j+1-p)}\right|\left\|u_n^{(p)}\right\|_\infty + \left\|u_n^{(j)}\right\|_\infty \|\sigma\|_v\right),$$
$$\left|\lambda_n^{(j+1)}\right| \le \sqrt{2}\left\|u_n^{(j)}\right\|_\infty \|\sigma\|_v, \quad (8)$$
$$j = 0,1,...,m-1.$$

One can deduce from (8) that

$$\left\|u_n^{(j+1)}\right\| \le M_n \sum_{p=0}^{j}\left\|u_n^{(j-p)}\right\|\left\|u_n^{(p)}\right\|_\infty,$$

where $M_n = \sqrt{2}\left\|g_n\right\|_\infty \left\|\sigma\right\|_v \leq \sqrt{2}\frac{7}{6\pi n}\left\|\sigma\right\|_v$.

To obtain the solution of (8) we use the generating functions method (see [22]). It gives us the following sequence of estimates for the solution

$$\left\|u_n^{(j)}\right\|_\infty \leq 2\sqrt{2}\frac{(2j-1)!!}{(2j+2)!!}(4M_n)^j \leq \sqrt{2}\frac{(4M_n)^j}{(j+1)\sqrt{\pi j}},$$
$$\left|\lambda_n^{(j+1)}\right| \leq 4\left\|\sigma\right\|_v \frac{(2j-1)!!}{(2j+2)!!}(4M_n)^j \leq 2\left\|\sigma\right\|_v \frac{(4M_n)^j}{(j+1)\sqrt{\pi j}},$$
$$j = 0, 1, ..., m-1,$$

where $(2j)!! = 2\cdot 4\cdot ...\cdot 2j$, $(2j+1)!! = 1\cdot 3\cdot ...\cdot (2j+1)$. These estimates along with the assumptions regarding the form of $\sigma(x)$ yields the next result.

Theorem 2. *Let $\sigma(x) \in BV_c[0,1]$ and the following condition holds true*

$$r_n \overset{def}{=} 4M_n = 4\sqrt{2}\left\|g_n\right\|_\infty \left\|\sigma\right\|_v < 1, \tag{9}$$

then the FD-method for the Sturm–Liouville problem (2) converges super-exponentially. Moreover the error estimates satisfy (10), (11)

$$\left\|u_n - \overset{m}{u}_n\right\|_\infty = \left\|u_n - \sum_{j=0}^{m} u_n^{(j+1)}\right\|_\infty \leq \frac{\sqrt{2}r_n^{m+1}}{(m+2)\sqrt{\pi(m+1)}(1-r_n)}, \tag{10}$$

$$\left|\lambda_n - \overset{m}{\lambda}_n\right| = \left|\lambda_n - \sum_{j=0}^{m}\lambda_n^{(j+1)}\right| \leq \frac{2\left\|\sigma\right\|_v r_n^m}{(m+1)\sqrt{\pi m}(1-r_n)}. \tag{11}$$

This result is a considerable extension and generalization of the similar results of section 5 from [21], as well as the results of theorem 1 from [8]. In order to show that let us recall the similar result from [8]. If $\sigma(x) \in BV_c[0,1]$ and

$$n > \frac{1}{4\pi}\left(\frac{681}{16}\left\|\sigma\right\|_v + 1\right) \overset{def}{=} n_b \tag{12}$$

then the following representation (in the notation of current work) is valid

$$\lambda_n = (\pi n)^2 - \int_0^1 \left[u_n^{(0)}(x)\right]^2 d\sigma(x) - \int_0^1\int_0^1 k_n(\xi_1,\xi_2)d\sigma(\xi_1)d\sigma(\xi_2) + \nu'_{n,2}(\sigma),$$

where

$$k_n(\xi_1,\xi_2) \equiv \frac{1}{4\pi n}\sum_{i=1}^{2}(1-\cos(2\pi n\xi_i))\sin(2\pi n\xi_{3-i})\times$$
$$\times\left[\frac{2}{\pi}\Theta(2\pi\xi_{3-i}) + (-1)^{i-1}sgn(\xi_2-\xi_1)\right],$$
$$\Theta(t) = (\pi - t)/2,$$

$$\left|\nu'_{n,2}(\sigma)\right| \leq \left\|\sigma\right\|_v^2 \frac{4.4 + 467\left\|\sigma\right\|_v + 2\left\|\sigma\right\|_v^2}{\left(\pi n - \frac{1}{4}\right)^2} \overset{def}{=} \gamma_b\left(n, \left\|\sigma\right\|_v\right). \tag{13}$$

At the same time, it follows from theorem 2 that

$$\lambda_n = (\pi n)^2 + \lambda_n^{(1)} + \lambda_n^{(2)} + R_n^{(3)}, \tag{14}$$

where

$$\lambda_n^{(1)} = \int_0^1 \left[u_n^{(0)}(x)\right]^2 d\sigma(x),\ \lambda_n^{(2)} = \int_0^1 u_n^{(0)}(x) u_n^{(1)}(x) d\sigma(x),$$

$$u_n^{(1)}(x) = \int_0^1 g_n(x,\xi) u_n^{(0)}(\xi) d\sigma(\xi),$$

while the residual term $R_n^{(3)}$ satisfies

$$\left|R_n^{(3)}\right| \leq \frac{2\left\|\sigma\right\|_v r_n^2}{3\sqrt{2\pi}(1-r_n)}, \tag{15}$$

as long as (9) holds. To make the comparison of the estimates (13) and (15) more convenient, we employ the estimate for r_n

$$r_n \overset{def}{=} 4\sqrt{2}\left\|g_n\right\|_\infty \left\|\sigma\right\|_v \leq 4\sqrt{2}\left[\frac{1}{\pi n} + \frac{1}{2(\pi n)^2}\right]\left\|\sigma\right\|_v \overset{def}{=} r_{n,1} \leq$$

$$\leq \frac{14\sqrt{2}}{3\pi n}\left\|\sigma\right\|_v \overset{def}{=} r_{n,2}.$$

Then the estimate (15) could be replaced by the estimate

$$\left|R_n^{(3)}\right| \leq \frac{2\left\|\sigma\right\|_v r_{n,1}^2}{3\sqrt{2\pi}(1-r_{n,1})} \overset{def}{=} \gamma_m\left(n, \left\|\sigma\right\|_v\right), \tag{16}$$

valid for all n such that

$$n > \frac{2\sqrt{2}}{\pi}\left(\left\|\sigma\right\|_v + \sqrt{\left\|\sigma\right\|_v^2 + \frac{\sqrt{2}}{4}\left\|\sigma\right\|_v}\right) \overset{def}{=} n_m. \tag{17}$$

By comparing (12) and (17) it is easy to see that

$$n_b > n_m,\ \forall \left\|\sigma\right\|_v \in [0,\infty),\ \lim_{\left\|\sigma\right\|_v \to \infty}(n_b - n_m) = \infty,$$

i.e. the condition (17) is less strict than the condition (12). Let us now compare estimates (16) and (13) for the residual terms for $n > n_b$, when both estimates make sense. For the clarity we remove the second summand from

$$\lambda_n^{(2)} = \int_0^1 u_n^{(0)}(x) \int_0^1 g_{n,1}(x,\xi) u_n^{(0)}(\xi) d\sigma(\xi) d\sigma(x) +$$

$$+ \int_0^1 u_n^{(0)}(x) \int_0^1 g_{n,2}(x,\xi) u_n^{(0)}(\xi) d\sigma(\xi) d\sigma(x) = \lambda_{n,1}^{(2)} + \lambda_{n,2}^{(2)}$$

(see (6)) and combine it with $R_n^{(3)}$. One can observe, afterwards, that

$$\nu_{n,2}'(\sigma) = R_n^{(3)} + \lambda_{n,2}^{(2)},$$

which after taking the norm of both sides lead us to the estimate for

$$\left|\nu'_{n,2}(\sigma)\right| \leq \left|R_n^{(3)}\right| + \left|\lambda_{n,2}^{(2)}\right| \leq \frac{2\left\|\sigma\right\|_v r_{n,1}^2}{3\sqrt{\pi 2}(1-r_{n,1})} + \frac{\left\|\sigma\right\|_v^2}{(n\pi)^2} \stackrel{def}{=} \tilde{\gamma}_m(n, \left\|\sigma\right\|_v).$$

Using the elementary computations we see that

$$\gamma_b(n, \left\|\sigma\right\|_v) > \tilde{\gamma}_m(n, \left\|\sigma\right\|_v), \ \forall \left\|\sigma\right\|_v \geq 0.$$

Consequently, we have shown that the second-rank FD-method could be more efficient than the approach suggested in [21] from the accuracy standpoint.

Example 2.1. Let us consider problem (2) with the potential $q(x) = \delta(x-a)$ for $\bar{q}(x) \equiv 0$, where a is a real number and $a \in (0,1)$. The algorithm of FD method described above is *exactly realizable* (see [23]) in this case.

Let us denote

$$I_0(x) = g_n(t,a), \quad I_j(x) = \int_0^1 g_n(x,t)\, I_{j-1}(t)\, dt, \ j = 1, 2, \ \ldots .$$

By applying and the so-called *sifting* or *sampling property* for function $f \in C^1[0,1]$, which reads as

$$\int_0^1 f(x)\,\delta(x-a)\,dx = f(a), \ \ a \in (0,1)$$

to (5) we obtain the following formulas for approximations of eigenvalues:

$$\begin{aligned}
\lambda_n^{(1)} &= \left[u_n^{(0)}(a)\right]^2, \ \ \lambda_n^{(2)} = \left[u_n^{(0)}(a)\right]^2 I_0(a), \\
\lambda_n^{(3)} &= \left[u_n^{(0)}(a)\right]^2 \left(-\left[u_n^{(0)}(a)\right]^2 I_1(a) + \left[I_0(a)\right]^2\right), \\
\lambda_n^{(4)} &= \left[u_n^{(0)}(a)\right]^2 \left(\left[u_n^{(0)}(a)\right]^4 I_2(a) - 3\left[u_n^{(0)}(a)\right]^2 I_0(a)\, I_1(a) + \left[I_0(a)\right]^3\right), \\
\lambda_n^{(5)} &= \left[u_n^{(0)}(a)\right]^2 \left(-\left[u_n^{(0)}(a)\right]^6 I_3(a) + \right. \\
&\quad + \left[u_n^{(0)}(a)\right]^4 \left(4I_0(a)\, I_2(a) + 2\left[I_1(a)\right]^2\right) - \\
&\quad \left. - 6\left[u_n^{(0)}(a)\right]^2 \left[I_0(a)\right]^2 I_1(a) + \left[I_0(a)\right]^4\right).
\end{aligned}$$

By setting $a = \dfrac{1}{\sqrt{2}}$ we obtain

$$I_0\left(\frac{1}{\sqrt{2}}\right) = \frac{\left(\sqrt{2}-1\right)\sin\left(\pi n\sqrt{2}\right)}{2n\pi} + \frac{\cos\left(\pi n\sqrt{2}\right) - 1}{4n^2\pi^2},$$

$$I_1\left(\frac{1}{\sqrt{2}}\right)=\frac{\left(3\sqrt{2}-4\right)\cos\left(\pi n\sqrt{2}\right)+1}{12n^2\pi^2}+ \\ +\frac{\left(\sqrt{2}-1\right)\sin\left(\pi n\sqrt{2}\right)}{4n^3\pi^3}+\frac{3}{16}\frac{\cos\left(\pi n\sqrt{2}\right)-1}{n^4\pi^4},$$

$$I_2\left(\frac{1}{\sqrt{2}}\right)=-\frac{\left(2\sqrt{2}-3\right)\sin\left(\pi n\sqrt{2}\right)+1}{24n^3\pi^3}+\frac{\left(3\sqrt{2}-4\right)\cos\left(\pi n\sqrt{2}\right)+1}{16n^4\pi^4}+ \\ +\frac{3\left(\sqrt{2}-1\right)\sin\left(\pi n\sqrt{2}\right)}{16n^5\pi^5}+\frac{5}{32}\frac{\cos\left(\pi n\sqrt{2}\right)-1}{n^6\pi^6},$$

$$I_3\left(\frac{1}{\sqrt{2}}\right)=-\frac{\left(30\sqrt{2}-43\right)\cos\left(\pi n\sqrt{2}\right)-2}{1440n^4\pi^4}-\frac{\left(2\sqrt{2}-3\right)\sin\left(\pi n\sqrt{2}\right)}{24n^5\pi^5}+ \\ +\frac{5}{96}\frac{\left(3\sqrt{2}-4\right)\cos\left(\pi n\sqrt{2}\right)+1}{n^6\pi^6}+\frac{5}{32}\frac{\left(\sqrt{2}-1\right)\sin\left(\pi n\sqrt{2}\right)}{n^7\pi^7}+ \\ +\frac{35}{256}\frac{\cos\left(\pi n\sqrt{2}\right)-1}{n^8\pi^8}.$$

From here we derive analytical expressions for the corrections to eigenvalues:

$$\lambda_n^{(1)}=1-\cos\left(\pi n\sqrt{2}\right),$$

$$\lambda_n^{(2)}=\frac{\sqrt{2}-1}{4n\pi}\left[2\sin\left(\pi n\sqrt{2}\right)-\sin\left(2\pi n\sqrt{2}\right)\right]+ \\ +\frac{1}{8n^2\pi^2}\left[4\cos\left(\pi n\sqrt{2}\right)-\cos\left(2\pi n\sqrt{2}\right)-3\right],$$

$$\lambda_n^{(3)}=\frac{1}{48n^2\pi^2}\left[\left(27-15\sqrt{2}\right)\cos\left(\pi n\sqrt{2}\right)-\left(36-24\sqrt{2}\right)\cos\left(2\pi n\sqrt{2}\right)-\right. \\ \left.-\left(-13+9\sqrt{2}\right)\cos\left(3\pi n\sqrt{2}\right)-4\right]+ \\ +\frac{\sqrt{2}-1}{8n^3\pi^3}\left[-5\sin\left(\pi n\sqrt{2}\right)+4\sin\left(2\pi n\sqrt{2}\right)-\sin\left(3\pi n\sqrt{2}\right)\right]- \\ -\frac{1}{16n^4\pi^4}\left[15\cos\left(\pi n\sqrt{2}\right)-6\cos\left(2\pi n\sqrt{2}\right)+\cos\left(3\pi n\sqrt{2}\right)-10\right],$$

$$\lambda_n^{(4)}=\frac{1}{96n^3\pi^3}\left[\left(33-26\sqrt{2}\right)\sin\left(\pi n\sqrt{2}\right)-\left(102-74\sqrt{2}\right)\sin\left(2\pi n\sqrt{2}\right)+\right. \\ \left.+\left(93-66\sqrt{2}\right)\sin\left(3\pi n\sqrt{2}\right)-\left(27-19\sqrt{2}\right)\sin\left(4\pi n\sqrt{2}\right)\right]+ \\ +\frac{1}{128n^4\pi^4}\left[\left(84\sqrt{2}-160\right)\cos\left(\pi n\sqrt{2}\right)+\left(260-168\sqrt{2}\right)\cos\left(2\pi n\sqrt{2}\right)+\right. \\ \left.+\left(-160+108\sqrt{2}\right)\cos\left(3\pi n\sqrt{2}\right)+\left(35-24\sqrt{2}\right)\cos\left(4\pi n\sqrt{2}\right)+25\right]-$$

$$- \frac{3}{32n^5\pi^5}\left[-14\sin\left(\pi n\sqrt{2}\right)+14\sin\left(2\pi n\sqrt{2}\right)-6\sin\left(3\pi n\sqrt{2}\right)+\right.$$
$$\left.+\sin\left(4\pi n\sqrt{2}\right)\right]\left(\sqrt{2}-1\right)+\frac{5}{128n^6\pi^6}\left[56\cos\left(\pi n\sqrt{2}\right)-\right.$$
$$\left.-28\cos\left(2\pi n\sqrt{2}\right)+8\cos\left(3\pi n\sqrt{2}\right)-\cos\left(4\pi n\sqrt{2}\right)-35\right].$$

Symbolic and numerical computations were carried out using the computer algebra system Maple 17.00 (where `Digits=50`). The exact values of first four smallest eigenvalues are:

$$\lambda_1^{ex} \approx 11.02252382511,\ \ \lambda_2^{ex} \approx 41.34074086778,$$
$$\lambda_3^{ex} \approx 89.10712301833,\ \ \lambda_4^{ex} \approx 158.4324892201.$$

Numerical results are given in Table 1, where we show the absolute error of approximation to the eigenvalue $\left|\lambda_n^{ex} - \overset{m}{\lambda}_n\right|$, $n = \overline{1,4}$ calculated by the FD-method with the rank $m = \overline{1,7}$.

TABL. 1. Convergence of FD-method for the eigenvalues λ_n, $n = \overline{1,4}$.

m	$\left\lvert\lambda_1^{ex} - \overset{m}{\lambda}_1\right\rvert$	$\left\lvert\lambda_2^{ex} - \overset{m}{\lambda}_2\right\rvert$	$\left\lvert\lambda_3^{ex} - \overset{m}{\lambda}_3\right\rvert$	$\left\lvert\lambda_4^{ex} - \overset{m}{\lambda}_4\right\rvert$
0	1.1529194	1.8623232	$2.8068340 \cdot 10^{-1}$	$5.1881880 \cdot 10^{-1}$
1	$1.13335918 \cdot 10^{-1}$	$4.1070777 \cdot 10^{-3}$	$3.9480386 \cdot 10^{-3}$	$8.1111546 \cdot 10^{-3}$
2	$7.74223271 \cdot 10^{-3}$	$5.4659978 \cdot 10^{-3}$	$3.5908153 \cdot 10^{-5}$	$2.0458308 \cdot 10^{-5}$
3	$2.41326302 \cdot 10^{-4}$	$2.2361009 \cdot 10^{-4}$	$2.7688079 \cdot 10^{-6}$	$4.7899759 \cdot 10^{-6}$
4	$1.80327662 \cdot 10^{-5}$	$1.7567730 \cdot 10^{-5}$	$2.2495782 \cdot 10^{-8}$	$9.7346306 \cdot 10^{-8}$
5	$2.80813804 \cdot 10^{-6}$	$2.7903081 \cdot 10^{-6}$	$1.7826757 \cdot 10^{-9}$	$1.1955865 \cdot 10^{-9}$
6	$8.40809762 \cdot 10^{-8}$	$8.3989549 \cdot 10^{-8}$	$5.1188146 \cdot 10^{-11}$	$1.0727859 \cdot 10^{-10}$
7	$1.70181022 \cdot 10^{-8}$	$1.7004392 \cdot 10^{-8}$	$7.0536476 \cdot 10^{-13}$	$1.6910131 \cdot 10^{-12}$

One can see that the method converges for all eigenvalues including $n = 1$, even though condition (9) of theorem 2 is satisfied for $n \geq 2$ only.

3. General scheme of FD-method (for $\bar{q}(x) \not\equiv 0$)

If condition (9) is not valid, one has to apply the general FD-method technique. We intend to consider this case in the present section. For this purpose we embed problem (2) into the more general parametrical problem set

$$\frac{\partial^2 u(x,t)}{\partial x^2}+\left\{\lambda(t)-\sum_{p=1}^{k}\gamma_p\delta(x-x_p)-\hat{\psi}'(x)-\right.$$
$$\left.-t\left[\psi'(x)-\hat{\psi}'(x)\right]\right\}u(x,t)=0, \tag{18}$$
$$x\in(0,1),\ u(0,t)=u(1,t)=0,$$

where $\psi(x)$ is the absolutely continuous function while $\hat{\psi}(x)$ stands for its piecewise linear approximation,

$$\hat{\psi}(x) = \psi(x_p)\frac{x_{p+1}-x}{x_{p+1}-x_p} + \psi(x_{p+1})\frac{x-x_p}{x_{p+1}-x_p},$$

$$\hat{\psi}'(x) = \psi_{x,p} = \frac{\psi(x_{p+1})-\psi(x_p)}{x_{p+1}-x_p},$$

$$x \in [x_p, x_{p+1}], \ \ p = \overline{0,k},$$

$$0 = x_0 < x_1 < ... < x_{k+1} = 1.$$

We look for the solution (18) in the form of series

$$u_n(x,t) = \sum_{j=0}^{\infty} u_n^{(j)}(x)\,t^j, \quad \lambda_n(t) = \sum_{j=0}^{\infty} \lambda_n^{(j)} t^j. \tag{19}$$

We substitute expressions (19) into (18) and then compare the coefficients in front of the equal powers of t. It gives us the following recurrence sequence of boundary problems:

$$\begin{cases} L_n^{(0)} u_n^{(j+1)}(x) \equiv \dfrac{d^2 u_n^{(j+1)}(x)}{dx^2} + \\ + \left\{ \lambda_n^{(0)} - \sum\limits_{p=1}^{k} \gamma_p \delta(x-x_p) - \hat{\psi}'(x) \right\} u_n^{(j+1)}(x) = \\ = -\sum\limits_{l=0}^{j} \lambda_n^{(j+1-l)} u_n^{(l)}(x) + \left[\psi'(x) - \hat{\psi}'(x)\right] u_n^{(j)}(x) \equiv \\ \equiv -F_n^{(j+1)}(x), \ x \in (0,1), \\ u_n^{(j+1)}(0) = u_n^{(j+1)}(1) = 0, \end{cases} \tag{20}$$

$$\lambda_n^{(j+1)} = \int_0^1 u_n^{(0)}(x)\left[\psi'(x) - \hat{\psi}'(x)\right] u_n^{(j)}(x)\,dx, \tag{21}$$

$$\int_0^1 u_n^{(0)}(x)\,u_n^{(j+1)}(x)\,dx = 0, \tag{22}$$

$$j = 0, 1, \ldots .$$

Here the pair $\left\{\lambda_n^{(0)}, u_n^{(0)}(x)\right\} = \{\lambda_n(0), u_n(0)\}$ is the solution of the basic problem

$$\begin{gathered} \frac{\partial^2 u_n^{(0)}(x)}{\partial x^2} + \left\{ \lambda_n^{(0)} - \sum_{p=1}^{k} \gamma_p \delta(x-x_p) - \hat{\psi}'(x) \right\} u_n^{(0)}(x) = 0, \ x \in (0,1), \\ u_n^{(0)}(0) = u_n^{(0)}(1) = 0, \end{gathered} \tag{23}$$

The sufficient conditions for the convergence of the series for $u_n(x,t)$ and $\lambda_n(t)$ at $t = 1$, where $u_n(x) = u_n(x,1)$, $\lambda_n = \lambda_n(1), n = 1, 2, ...$, will be presented later. But first we give the algorithmic implementation of the FD-method.

Let us rewrite the problem (23) in the alternative form

$$\begin{gathered}\frac{\partial^2 u_n^{(0)}(x)}{\partial x^2}+\left\{\lambda_n^{(0)}-\hat{\psi}'(x)\right\}u_n^{(0)}(x)=0,\\ x\in(0,x_1)\cup(x_1,x_2)\cup...\cup(x_k,1),\\ u_n^{(0)}(0)=u_n^{(0)}(1)=0,\end{gathered} \tag{24}$$

$$\left.\begin{aligned}\left[u_n^{(0)}(x)\right]_{x=x_p}&=u_n^{(0)}(x_p+0)-u_n^{(0)}(x_p-0)=0,\\ \left[\frac{du_n^{(0)}(x)}{dx}\right]_{x=x_p}&=\frac{du_n^{(0)}(x_p+0)}{dx}-\frac{du_n^{(0)}(x_p-0)}{dx}=\gamma_p u_n^{(0)}(x_p),\\ &p=\overline{1,k}.\qquad\text{(matching conditions)}\end{aligned}\right\} \tag{25}$$

On the intervals $[x_p,x_{p+1}),p=\overline{0,k-1}$ and $[x_k,1]$ the solutions of equation (24) can be written as follows

$$\begin{aligned}u_n^{(0)}(x)=A_{p,n}^{(0)}\sin&\left(\sqrt{\mu_{n,p}^{(0)}}(x-x_p)\right)+\\ &+B_{p,n}^{(0)}\cos\left(\sqrt{\mu_{n,p}^{(0)}}(x-x_p)\right),\quad x\in[x_p,x_{p+1}),\end{aligned}$$

$$p=\overline{0,k-1},\qquad B_{0,n}^{(0)}=0,$$

$$u_n^{(0)}(x)=A_{k,n}^{(0)}\sin\left(\sqrt{\mu_{n,k}^{(0)}}(1-x)\right),x\in[x_k,1],$$

where

$$\mu_{n,p}^{(0)}=\lambda_n^{(0)}-\psi_{x,p}.$$

The calculation of constants $A_{p,n}^{(0)}$, $p=\overline{0,k}$, $B_{p,n}^{(0)}$, $p=\overline{1,k-1}$ is performed using the combination of conditions (25) which when applied to the representation of solutions lead us to the following homogeneous system:

$$\begin{aligned}-A_{p-1,n}^{(0)}&\sin\left(\sqrt{\mu_{n,p-1}^{(0)}}(x_p-x_{p-1})\right)-\\ &-B_{p-1,n}^{(0)}\cos\left(\sqrt{\mu_{n,p-1}^{(0)}}(x_p-x_{p-1})\right)+B_{p,n}^{(0)}=0,\end{aligned}$$

$$\begin{aligned}-A_{p-1,n}^{(0)}&\sqrt{\mu_{n,p-1}^{(0)}}\cos\left(\sqrt{\mu_{n,p-1}^{(0)}}(x_p-x_{p-1})\right)+\\ &+B_{p-1,n}^{(0)}\sqrt{\mu_{n,p-1}^{(0)}}\sin\left(\sqrt{\mu_{n,p-1}^{(0)}}(x_p-x_{p-1})\right)+\\ &+\sqrt{\mu_{n,p}^{(0)}}A_{p,n}^{(0)}-\gamma_p B_{p,n}^{(0)}=0,\quad p=\overline{1,k-1},\quad B_{0,n}^{(0)}=0,\end{aligned} \tag{26}$$

$$-A^{(0)}_{k-1,n}\sin\left(\sqrt{\mu^{(0)}_{n,k-1}}\left(x_k-x_{k-1}\right)\right)-$$
$$-B^{(0)}_{k-1,n}\cos\left(\sqrt{\mu^{(0)}_{n,k-1}}\left(x_k-x_{k-1}\right)\right)+$$
$$+A^{(0)}_{k,n}\sin\left(\sqrt{\mu^{(0)}_{n,k}}\left(1-x_k\right)\right)=0,$$
$$-A^{(0)}_{k-1,n}\sqrt{\mu^{(0)}_{n,k-1}}\cos\left(\sqrt{\mu^{(0)}_{n,k-1}}\left(x_k-x_{k-1}\right)\right)+$$
$$+B^{(0)}_{k-1,n}\sqrt{\mu^{(0)}_{n,k-1}}\sin\left(\sqrt{\mu^{(0)}_{n,k-1}}\left(x_k-x_{k-1}\right)\right)-$$
$$-A^{(0)}_{k,n}\left[\sqrt{\mu^{(0)}_{n,k}}\cos\left(\sqrt{\mu^{(0)}_{n,k}}\left(1-x_k\right)\right)+\right.$$
$$\left.+\gamma_k\sin\left(\sqrt{\mu^{(0)}_{n,k}}\left(1-x_k\right)\right)\right]=0.$$

We look for the roots of determinant $\Delta\left(\lambda^{(0)}_n\right)$ of system (26) which are different from $\psi_{x,p}$, $p=\overline{0,k}$. Every eigenvalue of problems (24)-(25) is the zero of determinant $\Delta\left(\lambda^{(0)}_n\right)$ having the multiplicity 1. The eigenvalues form a monotonically increasing sequence $\lambda^{(0)}_1<\lambda^{(0)}_2<...<\lambda^{(0)}_n<...$ which tends to infinity.

For the given $\lambda^{(0)}_n$ the solution to system (26) can be determined only up to a constant factor which we calculate from the normalization condition

$$\left\|u^{(0)}_n\right\|_0=\left\{\int\limits_0^1\left[u^{(0)}_n(x)\right]^2dx\right\}^{\frac{1}{2}}=1.$$

The sequence of the normalized eigenfunctions $\left\{u^{(0)}_n(x)\right\}_{n=1}^{\infty}$ form a complete orthonormal system in $L_2[0,1]$. The above mentioned facts follow from the results of chapter 12 in [10].

Let us, move on to the solution of the recurrent sequence of problems (20)-(22). First we rewrite these equations in the equivalent form

$$\begin{gathered}\hat{L}^{(0)}_n u^{(j+1)}_n(x)\equiv\frac{d^2u^{(j+1)}_n(x)}{dx^2}+\mu^{(0)}_n u^{(j+1)}_n(x)=-F^{(j+1)}_n(x),\\ x\in(0,x_1)\cup(x_1,x_2)\cup...\cup(x_k,1),\\ \mu^{(0)}_n(x)=\mu^{(0)}_{n,p},\ \ x\in(x_p,x_{p+1}),\ \ p=\overline{0,k},\\ u^{(j+1)}_n(0)=u^{(j+1)}_n(1)=0,\end{gathered}\tag{27}$$

$$\left.\begin{aligned}&\left[u^{(j+1)}_n(x)\right]_{x=x_p}=0,\\ &\left[\frac{du^{(j+1)}_n(x)}{dx}\right]_{x=x_p}=\gamma_p u^{(j+1)}_n(x_p),p=\overline{1,k}.\end{aligned}\right\}\text{(matching conditions)}$$

Whereupon, its solution possess a representation

$$u_n^{(j+1)}(x) = A_{p,n}^{(j+1)} \sin\left(\sqrt{\mu_{n,p}^{(0)}}(x - x_p)\right) + \\ + B_{p,n}^{(j+1)} \cos\left(\sqrt{\mu_{n,p}^{(0)}}(x - x_p)\right) - \\ - \int\limits_{x_p}^{x} \frac{\sin\left(\sqrt{\mu_{n,p}^{(0)}}(x - \xi)\right)}{\sqrt{\mu_{n,p}^{(0)}}} F_n^{(j+1)}(\xi)\, d\xi,\ x \in [x_p, x_{p+1}),$$

$$p = \overline{0, k-1}, \quad B_{0,n}^{(j+1)} = 0,$$

$$u_n^{(j+1)}(x) = A_{k,n}^{(j+1)} \sin\left(\sqrt{\mu_{n,k}^{(0)}}(1 - x)\right) + \\ + \int\limits_{x}^{1} \frac{\sin\left(\sqrt{\mu_{n,k}^{(0)}}(x - \xi)\right)}{\sqrt{\mu_{n,k}^{(0)}}} F_n^{(j+1)}(\xi)\, d\xi, \quad x \in [x_k, 1].$$

By combining (27) and the matching conditions we obtain the following system for coefficients $A_{p,n}^{(j+1)}$, $B_{p,n}^{(j+1)}$:

$$- A_{p-1,n}^{(j+1)} \sin\left(\sqrt{\mu_{n,p-1}^{(0)}}(x_p - x_{p-1})\right) - B_{p-1,n}^{(j+1)} \cos\left(\sqrt{\mu_{n,p-1}^{(0)}}(x_p - x_{p-1})\right) + \\ + B_{p,n}^{(j+1)} = - \int\limits_{x_{p-1}}^{x_p} \frac{\sin\left(\sqrt{\mu_{n,p-1}^{(0)}}(x_p - \xi)\right)}{\sqrt{\mu_{n,p-1}^{(0)}}} F_n^{(j+1)}(\xi)\, d\xi,$$

$$- A_{p-1,n}^{(j+1)} \sqrt{\mu_{n,p-1}^{(0)}} \cos\left(\sqrt{\mu_{n,p-1}^{(0)}}(x_p - x_{p-1})\right) + B_{p-1,n}^{(j+1)} \sqrt{\mu_{n,p-1}^{(0)}} \times \\ \times \sin\left(\sqrt{\mu_{n,p-1}^{(0)}}(x_p - x_{p-1})\right) + \sqrt{\mu_{n,p}^{(0)}} A_{p,n}^{(j+1)} - \gamma_p B_{p,n}^{(j+1)} = \\ = - \int\limits_{x_{p-1}}^{x_p} \cos\left(\sqrt{\mu_{n,p-1}^{(0)}}(x_p - \xi)\right) F_n^{(j+1)}(\xi)\, d\xi, \tag{28}$$

$$p = \overline{1, k-1}, \qquad B_{0,n}^{(j+1)} = 0,$$

$$- A_{k-1,n}^{(j+1)} \sin\left(\sqrt{\mu_{n,k-1}^{(0)}}(x_k - x_{k-1})\right) - \\ - B_{k-1,n}^{(j+1)} \cos\left(\sqrt{\mu_{n,k-1}^{(0)}}(x_k - x_{k-1})\right) + \\ + A_{k,n}^{(j+1)} \sin\left(\sqrt{\mu_{n,k}^{(0)}}(1 - x_k)\right) =$$

$$= -\int\limits_{x_k}^{1} \frac{\sin\left(\sqrt{\mu_{n,k}^{(0)}}\,(x_k-\xi)\right)}{\sqrt{\mu_{n,k}^{(0)}}} F_n^{(j+1)}(\xi)\,d\xi -$$

$$-\int\limits_{x_{k-1}}^{x_k} \frac{\sin\left(\sqrt{\mu_{n,k-1}^{(0)}}\,(x_k-\xi)\right)}{\sqrt{\mu_{n,k-1}^{(0)}}} F_n^{(j+1)}(\xi)\,d\xi,$$

$$- A_{k-1,n}^{(j+1)}\sqrt{\mu_{n,k-1}^{(0)}}\cos\left(\sqrt{\mu_{n,k-1}^{(0)}}\,(x_k-x_{k-1})\right) +$$

$$+ B_{k-1,n}^{(j+1)}\sqrt{\mu_{n,k-1}^{(0)}}\sin\left(\sqrt{\mu_{n,k-1}^{(0)}}\,(x_k-x_{k-1})\right) -$$

$$- A_{k,n}^{(j+1)}\left[\sqrt{\mu_{n,k}^{(0)}}\cos\left(\sqrt{\mu_{n,k}^{(0)}}\,(1-x_k)\right) +\right.$$

$$\left.+\gamma_k \sin\left(\sqrt{\mu_{n,k}^{(0)}}\,(1-x_k)\right)\right] = -\int\limits_{x_k}^{1}\cos\left(\sqrt{\mu_{n,k}^{(0)}}\,(x_k-\xi)\right) F_n^{(j+1)}(\xi)\,d\xi -$$

$$-\int\limits_{x_{k-1}}^{x_k}\cos\left(\sqrt{\mu_{n,k-1}^{(0)}}\,(x_k-\xi)\right) F_n^{(j+1)}(\xi)\,d\xi +$$

$$+\gamma_k\int\limits_{x_k}^{1}\frac{\sin\left(\sqrt{\mu_{n,k}^{(0)}}\,(x_k-\xi)\right)}{\sqrt{\mu_{n,k}^{(0)}}} F_n^{(j+1)}(\xi)d\xi.$$

The left-hand-side matrix of this system of linear algebraic equations is degenerate since it coincides with that of the system (26). For the solution of (28) to exist it is necessary and sufficient that the vector composed from the right-hand-side coefficients is orthogonal to the eigenvector of the conjugate matrix.

Let us introduce the following vectors

$$\vec{Y}_n^{(j+1)} = \left\{ A_{0,n}^{(j+1)}, \underbrace{A_{1,n}^{(j+1)}, B_{1,n}^{(j+1)}}, \ldots, \underbrace{A_{k-1,n}^{(j+1)}, B_{k-1,n}^{(j+1)}}, A_{k,n}^{(j+1)} \right\}^T,$$

$$\vec{H}_n^{(j+1)} = \left\{ \vec{H}_{n,p}^{(j+1)} \right\}_{p=\overline{1,k}}^T,$$

$$\vec{H}_{n,p}^{(j+1)} = \left\{ -\int\limits_{x_{p-1}}^{x_p} \frac{\sin\left(\sqrt{\mu_{n,p-1}^{(0)}}\,(x_p-\xi)\right)}{\sqrt{\mu_{n,p-1}^{(0)}}} F_n^{(j+1)}(\xi)\,d\xi, \right.$$

$$\left. -\int\limits_{x_{p-1}}^{x_p}\cos\left(\sqrt{\mu_{n,p-1}^{(0)}}\,(x_p-\xi)\right) F_n^{(j+1)}(\xi)\,d\xi \right\},$$

$$p = \overline{1, k-1},$$

$$\vec{H}_{n,k}^{(j+1)} = \left\{ -\int\limits_{x_k}^{1} \frac{\sin\left(\sqrt{\mu_{n,k}^{(0)}}\,(x_k - \xi)\right)}{\sqrt{\mu_{n,k}^{(0)}}} F_n^{(j+1)}(\xi)\,d\xi - \right.$$
$$- \int\limits_{x_{k-1}}^{x_k} \frac{\sin\left(\sqrt{\mu_{n,k-1}^{(0)}}\,(x_k - \xi)\right)}{\sqrt{\mu_{n,k-1}^{(0)}}} F_n^{(j+1)}(\xi)\,d\xi,$$
$$- \int\limits_{x_k}^{1} \cos\left(\sqrt{\mu_{n,k}^{(0)}}\,(x_k - \xi)\right) F_n^{(j+1)}(\xi)\,d\xi -$$
$$- \int\limits_{x_{k-1}}^{x_k} \cos\left(\sqrt{\mu_{n,k-1}^{(0)}}\,(x_k - \xi)\right) F_n^{(j+1)}(\xi)\,d\xi +$$
$$\left. + \gamma_k \int\limits_{x_k}^{1} \frac{\sin\left(\sqrt{\mu_{n,k}^{(0)}}\,(x_k - \xi)\right)}{\sqrt{\mu_{n,k}^{(0)}}} F_n^{(j+1)}(\xi)\,d\xi \right\}$$

and denote the matrix of the system (26) as D_n. Then systems (26), (28) could be presented in the matrix-vector form

$$D_n \vec{Y}_n^{(0)} = \vec{0}, \;\; D_n \vec{Y}_n^{(j+1)} = \vec{H}_n^{(j+1)}, \;\; j = 0, 1, \ldots. \tag{29}$$

If $\vec{Z}_n^T$ is the eigenvector (row) that corresponds to the null eigenvalue of the matrix D_n, i.e.

$$\vec{Z}_n^T D_n = \vec{0},$$

then the necessary and sufficient condition of the solvability of system (29) is

$$\vec{Z}_n^T \vec{H}_n^{(j+1)} = 0. \tag{30}$$

It is easy to show that condition (30) is equivalent to the integral condition having the form

$$\int\limits_0^1 F_n^{(j+1)}(x)\, u_n^{(0)}(x)\,dx = 0. \tag{31}$$

Next we wind from (31) or, equivalently, from (30) that

$$\begin{aligned} \lambda_n^{(j+1)} = &-\sum_{p=1}^{j} \lambda_n^{(j-p+1)} \int\limits_0^1 u_n^{(0)}(x)\, u_n^{(p)}(x)\,dx + \\ &+ \int\limits_0^1 u_n^{(0)}(x) \left[\psi'(x) - \hat{\psi}'(x)\right] u_n^{(j)}(x)\,dx. \end{aligned} \tag{32}$$

Since the solution of system of linear algebraic equations (29) is found with the accuracy up to a constant factor, $u_n^{(j+1)}(x)$ is found with the same accuracy.

The constant factor can be calculated from the orthogonality condition (22), and formula (32) is transformed to (21).

The aforementioned results give us all information necessary to apply FD-method to some concrete problem. They however are not so useful to get the sufficient conditions of its convergence and the corresponding accuracy estimates (both a-priory and a-posteriori).

To get those estimates we propose an alternative approach. Relying on the completeness of the orthonormalized system $\left\{u_n^{(0)}(x)\right\}_{n=1}^{\infty}$ in $L_2[0,1]$, we write down the solution to problem (20) in the following form:

$$u_n^{(j+1)}(x) = -\sum_{\substack{p=1\\p\neq n}}^{\infty}\int_0^1 F_n^{(j+1)}(\xi)\,u_p^{(0)}(\xi)\,d\xi\,\frac{u_p^{(0)}(x)}{\lambda_n^{(0)}-\lambda_p^{(0)}}.$$

It lead us to the estimate

$$\begin{aligned}\left\|u_n^{(j+1)}\right\| &\leq M_n\left\|F_n^{(j+1)}\right\| \leq \\ &\leq M_n\left\{\sum_{l=1}^{j}\left|\lambda_n^{(j+1-l)}\right|\left\|u_n^{(l)}\right\| + \left\|\left[\psi'(x) - \hat{\psi}'(x)\right]u_n^{(j)}(x)\right\|\right\},\end{aligned} \tag{33}$$

where

$$M_n = \max\left\{\frac{1}{\lambda_n^{(0)}-\lambda_{n-1}^{(0)}}, \frac{1}{\lambda_{n+1}^{(0)}-\lambda_n^{(0)}}\right\}. \tag{34}$$

Let us introduce a function

$$\omega(\psi') = \max_{0\leq p\leq k}\max_{x\in[x_p,x_{p+1}]}\left|\int_{x_p}^{x_{p+1}}\frac{\psi'(x)-\psi'(t)}{x_{p+1}-x_p}dt\right|.$$

Then by substituting (21) into (33) we receive the sequence of estimates

$$\begin{aligned}&\left\|u_n^{(j+1)}\right\| \leq M_n\left\{\sum_{l=1}^{j}\left|\lambda_n^{(j+1-l)}\right|\left\|u_n^{(l)}\right\| + \omega(\psi')\left\|u_n^{(j)}\right\|\right\},\\ &\left|\lambda_n^{(j+1)}\right| \leq \omega(\psi')\left\|u_n^{(j)}\right\|,\end{aligned} \tag{35}$$

that lead to the following inequality

$$\left\|u_n^{(j+1)}\right\| \leq M_n\omega(\psi')\sum_{l=0}^{j}\left\|u_n^{(j-l)}\right\|\left\|u_n^{(l)}\right\|. \tag{36}$$

The solution of inequality (36) be obtained via the generating functions method. It has a following form (see [24])

$$\begin{aligned}\left\|u_n^{(j+1)}\right\| &\leq \left(4M_n\omega(\psi')\right)^{j+1}2\frac{(2j+1)!!}{(2j+4)!!} \leq \\ &\leq \frac{[4M_n\omega(\psi')]^{j+1}}{(j+2)\sqrt{\pi(j+1)}} = \frac{\hat{r}_n^{j+1}}{(j+2)\sqrt{\pi(j+1)}}.\end{aligned} \tag{37}$$

Inequality (37) permit us to get the corresponding inequality for the eigenvalue from (35)

$$\left|\lambda_n^{(j+1)}\right| \le \omega\left(\psi'\right)\hat{r}_n^j\, 2\frac{(2j-1)!!}{(2j+2)!!} \le \omega\left(\psi'\right)\frac{\hat{r}_n^j}{(j+1)\sqrt{\pi j}}. \tag{38}$$

Using estimates (37), (38) one can easily deduce that the next statement is correct

Theorem 3. *Let*

$$\sigma(x) = \sum_{p=1}^{k}\gamma_p H(x - x_p) + \psi(x) \tag{39}$$

and the following condition holds true

$$\hat{r}_n \overset{def}{=} 4M_n\omega\left(\psi'\right) < 1,$$

then the FD-method for the Sturm–Liouville problem (18), (39) converges super-exponentially. Moreover the following error estimates are valid:

$$\left\|u_n - \overset{m}{u}_n\right\| \le \left\|u_n - \sum_{j=0}^{m}u_n^{(j)}\right\| \le \frac{\hat{r}_n^{m+1}}{(m+2)\sqrt{\pi(m+1)}\,(1-\hat{r}_n)}, \tag{40}$$

$$\left|\lambda_n - \overset{m}{\lambda}_n\right| \le \left|\lambda_n - \sum_{j=0}^{m}\lambda_n^{(j)}\right| \le \frac{\omega\left(\psi'\right)\hat{r}_n^m}{(m+1)\sqrt{\pi m}\,(1-\hat{r}_n)}. \tag{41}$$

Remark 3.1. *In order to to understand the behavior of $\hat{r}_n$ with respect to n one can use (34) and theorem 2. They lead to the estimates on the denominator from* (34)

$$\begin{aligned}
\lambda_n^{(0)} - \lambda_{n-1}^{(0)} &= \pi^2(2n-1) + \\
&+ 2\sum_{p=1}^{k}\gamma_p\left[\sin^2(n\pi x_p) - \sin^2((n-1)\pi x_p)\right] + R_n^{(2)} - R_{n-1}^{(2)} \ge \\
&\ge \pi^2(2n-1) - 4\sum_{p=1}^{k}|\gamma_p| - \frac{2\sum_{p=1}^{k}|\gamma_p|}{2\sqrt{\pi}}\left[\frac{\hat{r}_n}{1-\hat{r}_n} + \frac{\hat{r}_{n-1}}{1-\hat{r}_{n-1}}\right], \\
\lambda_{n+1}^{(0)} - \lambda_n^{(0)} &\ge \pi^2(2n+1) - 4\sum_{p=1}^{k}|\gamma_p| - \frac{\sum_{p=1}^{k}|\gamma_p|}{\sqrt{\pi}}\left[\frac{\hat{r}_{n+1}}{1-\hat{r}_{n+1}} + \frac{\hat{r}_n}{1-\hat{r}_n}\right],
\end{aligned}$$

These estimates are valid under condition (9), i.e. estimates (40), (41) and (10), (11) from the theorem 2 are valid under the same restriction on n. However, $\hat{r}_n$ has a reserve of easing the restrictions on n up to its complete exclusion. This reserve caused by the occurrence of factor $\omega\left(\psi'\right)$ in $\hat{r}_n$ that will relax the restrictions on n provided that function $\psi'(x)$ is, at least, piecewise continuous function from $Q^0[0,1]$, i.e. $\psi(x) \in C[0,1] \cap Q^1[0,1]$.

Remark 3.2. *If the conditions of theorem 3 are met then the series $u_n(x,t) = \sum_{j=0}^{\infty}u_n^{(j)}(x)t^j$, $\lambda_n(t) = \sum_{j=0}^{\infty}\lambda_n^{(j)}t^j$ are absolutely convergent* *for* $|t| \leq 1$. *Moreover they approximate the exact solution of given problem* $u_n(x) = u_n(x,1) = \sum_{j=0}^{\infty} u_n^{(j)}(x)$, $\lambda_n = \lambda_n(1) = \sum_{j=0}^{\infty} \lambda_n^{(j)}$.

Example 3.1. We applied the FD-method to problem (2) with the potential $q(x) = \delta\left(x - \frac{1}{2}\right) + 100x$ in the following cases: a) $\hat{\psi}'(x) \equiv 0$, $k = 1$, $x_1 = \frac{1}{2}$, $\gamma_1 = 1$; b) the interval $(0,1)$ is partitioned into two equal subintervals ($\hat{\psi}'(x) \not\equiv 0$, $k = 1$, $x_1 = \frac{1}{2}$, $\gamma_1 = 1$); c) the interval $(0,1)$ is partitioned into four equal subintervals ($\hat{\psi}'(x) \not\equiv 0$, $k = 3$, $x_1 = \frac{1}{4}, x_2 = \frac{1}{2}, x_3 = \frac{3}{4}$, $\gamma_1 = 0$, $\gamma_2 = 1$, $\gamma_3 = 0$).

TABL. 2. Convergence of FD-method for eigenvalue λ_1

m	a) $\hat{\psi}'(x) \equiv 0$, $k=1$, $\left\lvert\lambda_1^{ex} - \overset{m}{\lambda}_1\right\rvert$	b) $\hat{\psi}'(x) \not\equiv 0$, $k=1$, $\left\lvert\lambda_1^{ex} - \overset{m}{\lambda}_1\right\rvert$	c) $\hat{\psi}'(x) \not\equiv 0$, $k=3$, $\left\lvert\lambda_1^{ex} - \overset{m}{\lambda}_1\right\rvert$
0	39.79669103	2.270616222	$2.168801379 \cdot 10^{-1}$
1	10.20330897	$8.341737964 \cdot 10^{-1}$	$6.083140294 \cdot 10^{-2}$
2	2.135818380	$1.901098870 \cdot 10^{-2}$	$5.300909434 \cdot 10^{-5}$
3	2.135818380	$3.157060409 \cdot 10^{-3}$	$4.333553271 \cdot 10^{-6}$
4	1.226920389	$2.930165507 \cdot 10^{-4}$	$1.367746278 \cdot 10^{-8}$
5	1.226920389	$2.102813177 \cdot 10^{-5}$	$5.850330410 \cdot 10^{-10}$
6	$9.509541771 \cdot 10^{-1}$	$4.743628885 \cdot 10^{-6}$	$3.835005760 \cdot 10^{-12}$
7	$9.509541771 \cdot 10^{-1}$	$5.240882809 \cdot 10^{-8}$	$9.702842701 \cdot 10^{-14}$
8	$8.506978298 \cdot 10^{-1}$	$7.286716281 \cdot 10^{-8}$	$1.229092383 \cdot 10^{-15}$
9	$8.506978298 \cdot 10^{-1}$	$2.930256199 \cdot 10^{-9}$	$1.865391361 \cdot 10^{-17}$
10	$8.276761403 \cdot 10^{-1}$	$1.032042190 \cdot 10^{-9}$	$4.064792983 \cdot 10^{-19}$
11	$8.276761403 \cdot 10^{-1}$	$1.038538699 \cdot 10^{-10}$	$3.423104476 \cdot 10^{-21}$
12	$8.508842593 \cdot 10^{-1}$	$1.221151730 \cdot 10^{-11}$	$1.238050539 \cdot 10^{-22}$
13	$8.508842593 \cdot 10^{-1}$	$2.481662360 \cdot 10^{-12}$	$3.497226425 \cdot 10^{-25}$
14	$9.094304891 \cdot 10^{-1}$	$8.479672332 \cdot 10^{-14}$	$3.323469489 \cdot 10^{-26}$
15	$9.094304891 \cdot 10^{-1}$	$4.980766446 \cdot 10^{-14}$	$1.068874105 \cdot 10^{-28}$
16	1.000506593	$1.155397490 \cdot 10^{-15}$	$7.886548397 \cdot 10^{-30}$
17	1.000506593	$8.676674901 \cdot 10^{-16}$	$9.000481917 \cdot 10^{-32}$
18	1.125540512	$6.964067548 \cdot 10^{-17}$	$1.660871600 \cdot 10^{-33}$
19	1.125540512	$1.270480995 \cdot 10^{-17}$	$4.098653028 \cdot 10^{-35}$
20	1.288866993	$2.045466355 \cdot 10^{-18}$	$2.760733186 \cdot 10^{-37}$

We computed the exact eigenvalue (further denoted by λ_1^{ex}) and its approximation (denoted by λ_1) using the computer algebra system Maple 17.00 (`Digits=100`). The smallest exact eigenvalue of the problem, considered here, is equal to

$$\lambda_1^{ex} \approx 51.568550194800485588919739351190684390 85.$$

The absolute errors of approximations $\left\lvert\lambda_1^{ex} - \overset{m}{\lambda}_1\right\rvert$ to smallest eigenvalue λ_1 obtained using the FD-method of rank $m = \overline{1,20}$ in the cases a)-c) are presented in table 2.

One can see from the table 2 that the simplest form of the FD-method a) (with $\hat{\psi}'(x) \equiv 0$) for the first eigenvalue is divergent while the FD-method converges when the interval is partitioned into two or more subintervals. The convergence rate is doubled with increase in the number of subdivision points (from one to three).

Bibliography

1. Solvable Models in Quantum Mechanics [Second edition. With an appendix by Pavel Exner] / S. Albeverio, F. Gesztesy, R. Hoegh-Krohn, H. Holden. – Providence, State RI.: AMS Chelsea Publishing, 2005.
2. Hassan S. Ashour. Coherent Tunneling through Quantum Wire Tailored by Gaussian Profile / Hassan S. Ashour // Journal of Modern Physics. – 2011. – No 2. – P. 124-130.
3. Belloni M. The infinite well and Dirac delta function potentials as pedagogical, mathematical and physical models in quantum mechanics / M. Belloni, R. W. Robinett // Physics Reports. – 2014. – No 540. – P. 25-122.
4. Sadovnichii V. A. Traces of operators / V. A. Sadovnichii, V. E. Podolskii // Uspekhi Mat. Nauk. – 2006. – Vol. 61, No 5. — P. 89-156 (in Russian).
5. Savchuk A. M. Trace Formula for Sturm–Liouville Operators with Singular Potentials / A. M. Savchuk, A. A. Shkalikov // Mathematical Notes. – 2001. – Vol. 69, No 3. – P. 387-400.
6. Savchuk A. M. First-order regularized trace of the Sturm-Liouville operator with δ – potential / A. M. Savchuk // Uspekhi Mat. Nauk. – 2000. – Vol. 55, No 6. – P. 155-156 (in Russian).
7. Vinokurov V. A. The eigenvalue and the trace of the Sturm–Liouville operator as differentiable functions of an integrable potential / V. A. Vinokurov, V. A. Sadovnichii // Dokl. Ross. Akad. Nauk. – 1999. – Vol. 365, No 3. – P. 295-297 (in Russian).
8. Vinokurov V. A. Asymptotics of eigenvalues and eigenfunctions and the trace formula for the potential containing δ-functions / V. A. Vinokurov, V. A. Sadovnichiy // Diff. Equations. -– 2002. — Vol. 38, No 6. – P. 735-751 (in Russian).
9. Kolmogorov A. N. Elements of the Theory of Functions and Functional Analysis / A. N. Kolmogorov, S. V. Fomin. – M.: Nauka, 1976 (in Russian).
10. Atkinson F. V. Discrete and continuous boundary problems / F. V. Atkinson. – M.: Mir, 1968 (in Russian).
11. Marletta M. Automatic solution of Sturm-Liouville problems using the Pruess method / M. Marletta, J. D. Pryce // Journal of Computational and Applied Mathematics. – 1992. – Vol. 39, No 1. – P. 57-78.
12. Pruess S. Mathematical software for Sturm-Liouville problems / S. Pruess, C. T. Fulton // ACM Trans. Math. Software. — September, 1993. – Vol. 19, No 3. – P. 360-376.
13. Bailey P. B. ALGORITHM 700: A FORTRAN software package for Sturm-Liouville problems / P. B. Bailey, B. S. Garbor, H. G. Kaper, A. Zettl // ACM Transactions on Mathematical Software. – December, 1991. – Vol. 17, No 4. – P. 500-501.
14. Bailey P. B. Eigenvalue and Eigenfunction Computaitions for Sturm–Liouville Problems / P. B. Bailey, B. S. Garbor, H. G. Kaper, A. Zettl // ACM Transactions on Mathematical Software. – December, 1991. – Vol. 17, No 4. – P. 491-499.
15. Bailey P. B. Algorithm 810: The SLEIGN2 Sturm-Liouville Code / P. B. Bailey, W. N. Everitt, A. Zettl // ACM Transactions on Mathematical Software (TOMS). — June, 2001. — Vol. 27, No 2. – P. 143-192.
16. Ledoux V. A MATLAB Package for the Numerical Solution of Sturm-Liouville and Schrődinger Equations / V. Ledoux, M. Van Daele, G. Vanden Berghe // ACM Transactions on Mathematical Software. — December, 2005. — Vol. 31, No 4. – P. 532-554.
17. Pryce J. D. Algorithm 789: SLTSTPAK: a test package for Sturm-Liouville solvers / J. D. PRYCE // ACM Transactions on Mathematical Software. — March, 1999. — Vol. 25, No 1. – P. 58-69.
18. Pryce J. D. A Test Package for Sturm-Liouville Solvers / J. D. PRYCE // ACM Transactions on Mathematical Software. — March, 1999. — Vol. 25, No 1. – P. 21-57.

19. Ledoux V. Study of Special Algorithms for Solving Sturm-Liouville and Schrődinger Equations / V. Ledoux. – // Ph. D. Thesis. – Department of Applied Mathematics and Computer Science. – Ghent University. – 2007.
20. Makarov V. L. An exponentially convergent functional-discrete method for solving Sturm-Liouville problems with a potential including the Dirac δ-function / V. L. Makarov, N. O. Rossokhata, D. V. Dragunov // J. Comput. Appl. Math. – 2013. – No 250. – P. 39-57.
21. Makarov V. L. Experimental and analytical studies of properties of the FD-method components when applying to the Sturm-Liouville problems / V. L. Makarov, N. M. Romaniuk, I. I. Lazurchak // Transactions of the Institute of Mathematics, NAS Ukraine. – 2013. – Vol. 10, No 1. – P. 145-170 (in Ukrainian).
22. Makarov V. L. FD-method – exponential convergence rate / V. L. Makarov // Computational and Applied Mathematics. – 1997. – No 82. – P. 69-74 (in Russian).
23. Makarov V. L. The FD-method for first-order linear hyperbolic differential equations with piecewise smooth coefficients / V. L. Makarov, V. V. Vinokur // J. of Mathematical Sciences. – 1995. – Vol. 77, No 5. – P. 3399-3405.
24. Makarov V. L. FD-method for Sturm-Liouville Problems. Exponential Rate of Convergence / V. L. Makarov, O. L. Ukhanev // Applied Mathematics and Informatics. -– Tbilisi University Press. – 1997. – Vol. 2. – P. 1-19.

Volodymyr Makarov, Nataliia Romaniuk,
Institute of Mathematics, National Academy of Sciences,
3, Tereschenkivs'ka Str., Kyiv, 01601, Ukraine;

Igor Lazurchak,
Drogobych State Pedagogical University,
34, I. Franka Str., Drogobych, 82100, Ukraine

Received 21.05.2014

Журнал обчислювальної та прикладної математики | 2014 № 2 (116) | Journal of Computational & Applied Mathematics

UDC 519.6+517.983.54

IMPLICIT ITERATION METHOD OF SOLVING LINEAR EQUATIONS WITH APPROXIMATING RIGHT-HAND MEMBER AND APPROXIMATELY SPECIFIED OPERATOR

Oleg Matysik

Резюме. У гільбертовому просторі досліджується неявний метод ітерацій розв'язування лінійних рівнянь з ненегативним самоспряженим і несамоспряженим обмеженим оператором. Доведено збіжність методу у випадку апріорного вибору числа ітерацій у вихідній нормі гільбертового простору, в припущенні, що похибки є не тільки в правій частині рівняння, а й в операторі. Отримано оцінки похибки і апріорний момент зупинки.

Abstract. The article deals with the study of the implicit method of solving linear equations with nonnegative self-adjoint and nonself-adjoint limited operator in Hilbert space. It aims at proving the method convergence in case of a priori choice of the number of iterations in the basic norm of Hilbert space on the assumption of existing errors not only in the equation right-hand member but in the operator as well. Error estimation and a priori stop moment are obtained.

1. Problem statement

Let H and F be Hilbert spaces and $A \in \pounds(H, F)$, i. e. A is a linear continuous operator functioning from H to F. It is assumed that zero belongs to operator spectrum A, but it is not its characteristic constant. The following equation is solved

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

The problem of searching for element $x \in H$ by element $y \in F$ is incorrect, for arbitrary small disturbances in the right-hand member y may result in arbitrary disturbances in solution.

Let us suppose that the accurate development $x^* \in H$ of equation (1) exists and is the unique one. We shall search for it with the help of iteration process

$$(E + \alpha^2 A^{2k})x_{n+1} = (E - \alpha A^k)^2 x_n + 2\alpha A^{k-1} y, x_0 = 0, k \in N, \tag{2}$$

where E is an identity operator while α is an iteration parameter.

We consider that operator A and the right-hand member of equation (1) are specified approximately, i.e. approximation $y_\delta, \|y - y_\delta\| \le \delta$ is known instead of y, and operator $A_\eta, \|A - A_\eta\| \le \eta$ is known instead of operator A. Suppose $0 \in Sp(A_\eta), Sp(A_\eta) \subseteq [0, M]$. Then method (2) will look

$$(E + \alpha^2 A_\eta^{2k})x_{n+1} = (E - \alpha A_\eta^k)^2 x_n + 2\alpha A_\eta^{k-1} y_\delta, x_0 = 0, k \in N. \tag{3}$$

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

The case of approximate right-member of equation y_δ and faithful operator A for the method under consideration (3) has been studied in monograph [1]. It deals with a priori and a posteriori choice of a regularization parameter and the case of non-unique solution of problem (1), as well as with proving the method convergence in Hilbert space energy norm.

Let us prove the method convergence (3) in case of a priori choice of a regularization parameter in solving the equation $A_\eta x = y_\delta$ with the approximate operator A_η and the approximate right-hand member y_δ and obtain a priori estimated errors.

2. The case of self-adjoint nonnegative operators

Let H equal F, $A = A^* \geq 0$, $A_\eta = A_\eta^* \geq 0$, $Sp(A_\eta) \subseteq [0, M]$, $0 < \eta \leq \eta_0$. The iteration method (3) will be presented in the following way:

$$x_\eta = g_n(A_\eta) y_\delta, \tag{4}$$

where $g_n(\lambda) = \lambda^{-1}\left[1 - \dfrac{(1-\alpha\lambda^k)^{2n}}{(1+\alpha^2\lambda^{2k})^n}\right]$. There have been obtained in [1-2] the conditions for functions $g_n(\lambda)$ with $\alpha > 0$:

$$\sup_{0\leq\lambda\leq M} |g_n(\lambda)| \leq \gamma n^{1/k}, \gamma = 2k\alpha^{1/k}, n > 0, \tag{5}$$

$$\sup_{0\leq\lambda\leq M} \lambda^s |1 - \lambda g_n(\lambda)| \leq \gamma_s n^{-s/k}, (n > 0), 0 < s < \infty, \gamma_s = \left(\frac{s}{2k\alpha e}\right)^{s/k}, \tag{6}$$

(here s is the degree of source representability of exact solution $x^* = A^s z, s > 0, \|z\| \leq \rho$),

$$\sup_{0\leq\lambda\leq M} |1 - \lambda g_n(\lambda)| \leq \gamma_0, \gamma_0 = 1, n > 0, \tag{7}$$

$$\sup_{0\leq\lambda\leq M} \lambda |1 - \lambda g_n(\lambda)| \to 0, n \to \infty. \tag{8}$$

The following is valid:

Lemma 1. *Let $A = A^* \geq 0$, $A_\eta = A_\eta^* \geq 0$, $\|A_\eta - A\| \leq \eta$, $Sp(A_\eta) \subseteq [0, M]$, $(0 < \eta \leq \eta_0)$, $\alpha > 0$ and conditions (7), (8) be satisfied. Then $\|G_{n\eta} v\| \to 0$ at $n \to \infty$, $\eta \to 0$ $\forall v \in N(A)^\perp = \overline{R(A)}$, where $N(A) = \{x \in H | Ax = 0\}$ and $G_{n\eta} = E - A_\eta g_n(A_\eta)$.*

Proof. We have

$$\begin{aligned}
\|G_{n\eta} v\| &= \|(E - A_\eta g_n(A_\eta))v\| = \\
&= \left\| \int_0^M (1 - \lambda g_n(\lambda)) dE_\lambda v \right\| = \left\| \int_0^M \frac{(1-\alpha\lambda^k)^{2n}}{(1+\alpha^2\lambda^{2k})^n} dE_\lambda v \right\| \leq \\
&\leq \left\| \int_0^\varepsilon \frac{(1-\alpha\lambda^k)^{2n}}{(1+\alpha^2\lambda^{2k})^n} dE_\lambda v \right\| + \left\| \int_\varepsilon^M \frac{(1-\alpha\lambda^k)^{2n}}{(1+\alpha^2\lambda^{2k})^n} dE_\lambda v \right\|.
\end{aligned}$$

$$\left\| \int_{\varepsilon}^{M} \frac{(1-\alpha\lambda^k)^{2n}}{(1+\alpha^2\lambda^{2k})^n} dE_\lambda v \right\| \le q^n(\varepsilon) \left\| \int_{\varepsilon}^{M} dE_\lambda v \right\| \to 0, n \to \infty,$$

as for $\lambda \in [\varepsilon, M]$

$$\frac{(1-\alpha\lambda^k)^2}{(1+\alpha^2\lambda^{2k})^n} \le q(\varepsilon) < 1.$$

$$\left\| \int_{0}^{\varepsilon} \frac{(1-\alpha\lambda^k)^{2n}}{(1+\alpha^2\lambda^{2k})^n} dE_\lambda v \right\| \le \left\| \int_{0}^{\varepsilon} dE_\lambda v \right\| = \|E_\varepsilon v\| \to 0, \quad \varepsilon \to 0$$

owing to integrated spectrum properties [3-4]. Consequently, $\|G_{n\eta}v\| \to 0$ at $n \to \infty$, $\eta \to 0$. Lemma 1 is proved. □

The convergence condition for method (3) is given by

Theorem 1. *Let* $A = A^* \ge 0$, $A_\eta = A_\eta^* \ge 0$, $\|A_\eta - A\| \le \eta$, $Sp(A_\eta) \subseteq [0, M]$, $(0 < \eta \le \eta_0)$, $\alpha > 0$, $y \in R(A)$, $\|y - y_\delta\| \le \delta$ *and conditions* (5), (7), (8) *be satisfied. Let us choose parameter* $n = n(\delta, \eta)$ *in approximation* (3) *so that* $(\delta + \eta)n^{1/k}(\delta, \eta) \to 0$ *at* $n(\delta, \eta) \to \infty$, $\delta \to 0$, $\eta \to 0$. *Then* $x_{n(\delta,\eta)} \to x^*$ *at* $\delta \to 0$, $\eta \to 0$.

Proof. According to (4) we have $x_n = g_n(A_\eta)y_\delta$. Then

$$x_n - x^* = g_n(A_\eta)y_\delta - x^* = -G_{n\eta}x^* + G_{n\eta}x^* + g_n(A_\eta)y_\delta - x^* =$$

$$= -G_{n\eta}x^* + (E - A_\eta g_n(A_\eta))x^* + g_n(A_\eta)y_\delta - x^* = -G_{n\eta}x^* + g_n(A_\eta)(y_\delta - A_\eta x^*).$$

Condition (5) being as follows $\|g_n(A_\eta)\| \le \sup\limits_{0\le\lambda\le M} |g_n(\lambda)| \le \gamma n^{1/k}$, then

$$\|y_\delta - A_\eta x^*\| \le \|y_\delta - y\| + \|y - A_\eta x^*\| =$$

$$= \|y_\delta - y\| + \|Ax^* - A_\eta x^*\| \le \delta + \|A - A_\eta\|\|x^*\| \le \delta + \eta\|x^*\|.$$

Consequently,

$$\|x_{n(\delta,\eta)} - x^*\| \le \|G_{n\eta}x^*\| + \|g_n(A_\eta)(y_\delta - A_\eta x^*)\| \le \|G_{n\eta}x^*\| + \gamma n^{1/k}(\delta + \eta\|x^*\|).$$

As appears from Lemma 1, $\|G_{n\eta}x^*\| \to 0$ at $n \to \infty$, $\eta \to 0$, and according to the condition of Theorem 1, $n^{1/k}(\delta + \eta) \to 0$ at $\delta \to 0$, $\eta \to 0$. Thus, $\|x_{n(\delta,\eta)} - x^*\| \to 0$, $\delta \to 0$, $\eta \to 0$. Theorem 1 is proved. □

Theorem 2. *Let* $A = A^* \ge 0$, $A_\eta = A_\eta^* \ge 0$, $\|A_\eta - A\| \le \eta$, $Sp(A_\eta) \subseteq [0, M]$, $(0 < \eta \le \eta_0)$, $\alpha > 0$, $y \in R(A)$, $\|y_\delta - y\| \le \delta$ *and conditions* (5), (6) *be satisfied. If the exact solution is source representable, i.e.* $x^* = A^s z$, $s > 0$, $\|z\| \le \rho$, *then error estimation is equitable*

$$\|x_{n(\delta,\eta)} - x^*\| \le \gamma_0 c_s \eta^{\min(1,s)}\rho + \gamma_s n^{-s/k}\rho + \gamma n^{1/k}(\delta + \eta\|x^*\|), 0 < s < \infty.$$

Proof. Using the source representability of the exact solution we have

$$\begin{aligned} \|G_{n\eta}x^*\| = \|G_{n\eta}A^s z\| \le \|G_{n\eta}(A^s - A_\eta^s)z\| + \|G_{n\eta}A_\eta^s z\| \le \\ \le \gamma_0 c_s \eta^{\min(1,s)}\rho + \gamma_s n^{-s/k}\rho, \end{aligned} \tag{9}$$

as according to Lemma 1.1 [5, p. 91] $\|A_\eta^s - A^s\| \le c_s\eta^{\min(1,s)}$, $c_s = const$, ($c_s \le 2$ for $0 < s \le 1$). Then

$$\|x_{n(\delta,\eta)} - x^*\| \le \gamma_0 c_s \eta^{\min(1,s)}\rho + \gamma_s n^{-s/k}\rho + \gamma n^{1/k}(\delta + \eta\|x^*\|), 0 < s < \infty. \quad (10)$$

Theorem 2 is proved. □

If the right side of estimation (10) is minimized by n, we get the meaning of a priori stop moment:

$$n_{opt} = \left[\frac{s\gamma_s\rho}{\gamma\left(\delta + \|x^*\|\eta\right)}\right]^{k/(s+1)} = d_s\rho^{k/(s+1)}\left[\delta + \eta\|x^*\|\right]^{-k/(s+1)},$$

where $d_s = \left(\dfrac{s\gamma_s}{\gamma}\right)^{k/(s+1)} = \left(\dfrac{s}{2k}\right)^{(s+k)/(s+1)} \alpha^{-1}e^{-s/(s+1)}$. Consequently,

$$n_{opt} = \left(\frac{s}{2k}\right)^{(s+k)/(s+1)} \alpha^{-1}e^{-s/(s+1)}\rho^{k/(s+1)}\left[\delta + \eta\|x^*\|\right]^{-k/(s+1)}.$$

Let us substitute n_{opt} in estimation (10) to get

$$\|x_{n(\delta,\eta)} - x^*\|_{opt} \le \gamma_0 c_s\eta^{\min(1,s)}\rho + \gamma_s\rho\left(d_s\rho^{k/(s+1)}\right)^{-s/k}\left(\delta + \eta\|x^*\|\right)^{s/(s+1)} +$$

$$+\gamma\left(\delta + \eta\|x^*\|\right)d_s^{1/k}\rho^{1/(s+1)}\left(\delta + \eta\|x^*\|\right)^{-1/(s+1)} =$$

$$= \gamma_0 c_s\eta^{\min(1,s)}\rho + \left(\delta + \eta\|x^*\|\right)^{s/(s+1)}\left(d_s^{-s/k}\gamma_s\rho^{1/(s+1)} + \gamma d_s^{1/k}\rho^{1/(s+1)}\right) =$$

$$= \gamma_0 c_s\eta^{\min(1,s)}\rho + \rho^{1/(s+1)}c_s'\left(\delta + \eta\|x^*\|\right)^{s/(s+1)},$$

where

$$c_s' = d_s^{-s/k}\gamma_s + \gamma d_s^{1/k} = \left(s^{1/(s+1)} + s^{-s/(s+1)}\right)\gamma^{s/(s+1)}\gamma_s^{1/(s+1)} =$$

$$= \left(\frac{s}{2k}\right)^{s(1-k)/(k(s+1))}(1+s)e^{-s/(k(s+1))}.$$

Hence

$$\|x_{n(\delta,\eta)} - x^*\|_{opt} \le c_s\eta^{\min(1,s)}\rho +$$

$$+\left(\frac{s}{2k}\right)^{s(1-k)/(k(s+1))}(1+s)e^{-s/(k(s+1))}\rho^{1/(s+1)}\left(\delta + \eta\|x^*\|\right)^{s/(s+1)}.$$

Note. *Optimal error estimation does not depend on* α, *whereas* n_{opt} *depends on* α. *Since there are no contingencies concerning* α *upwards* ($\alpha > 0$), *it is possible to choose* α *so as to make* $n_{opt} = 1$. *For that it is enough to take*

$$\alpha_{opt} = \left(\frac{s}{2k}\right)^{(s+k)/(s+1)} e^{-s/(s+1)}\rho^{k/(s+1)}\left[\delta + \eta\|x^*\|\right]^{-k/(s+1)}.$$

3. The case of nonself-adjoint operators

In case of nonself-adjoint problem iteration method (3) will be presented as

$$\begin{gathered}\left[E + \alpha^2(A_\eta^* A_\eta)^{2k}\right]x_{n+1} = \left[E - \alpha(A_\eta^* A_\eta)^k\right]^2 x_n + \\ +2\alpha(A_\eta^* A_\eta)^{k-1}A_\eta^* y_\delta, \quad x_0 = 0, \ k \in N.\end{gathered} \quad (11)$$

It can be written as follows:

$$x_n = g_n(A_\eta^* A_\eta) A_\eta^* y_\delta. \tag{12}$$

It follows from Lemma 1 that

Lemma 2. *Let* $A, A_\eta \in \pounds(H, F), \|A_\eta - A\| \le \eta, \|A_\eta\|^2 \le M, \alpha > 0$ *and conditions* (7), (8) *be satisfied. Then*

$$\|K_{n\eta} v\| \to 0 \text{ at } n \to \infty, \eta \to 0, \forall v \in N(A)^\perp = \overline{R(A^*)}, \tag{13}$$

$$\|\tilde{K}_{n\eta} z\| \to 0 \text{ at } n \to \infty, \eta \to 0, \forall z \in N(A^*)^\perp = \overline{R(A)}, \tag{14}$$

where $K_{n\eta} = E - A_\eta^* A_\eta g_n(A_\eta^* A_\eta)$, $\tilde{K}_{n\eta} = E - A_\eta A_\eta^* g_n(A_\eta A_\eta^*)$.

Lemma 2 is used for proving the following theorem.

Theorem 3. *Let* $A, A_\eta \in \pounds(H, F)$, $\|A - A_\eta\| \le \eta$, $\|A_\eta\|^2 \le M$, $(0 < \eta \le \eta_0)$, $\alpha > 0$, $y \in R(A)$, $\|y_\delta - y\| \le \delta$ *and conditions* (5), (7), (8) *be satisfied. Parameter* $n = n(\delta, \eta)$ *is chosen so as to get*

$$(\delta + \eta)^2 n^{1/k}(\delta, \eta) \to 0 \text{ at } n(\delta, \eta) \to \infty, \delta \to 0, \eta \to 0. \tag{15}$$

Then $x_{n(\delta,\eta)} \to x^*$ *at* $\delta \to 0, \eta \to 0$.

Proof. For approximation error $x_{n(\delta,\eta)}$ we have

$$x_{n(\delta,\eta)} - x^* = -K_{n\eta} x^* + g_n(A_\eta^* A_\eta) A_\eta^* (y_\delta - A_\eta x^*). \tag{16}$$

We see $\|g_n(A_\eta^* A_\eta) A_\eta^*\| = \|g_n(A_\eta^* A_\eta)(A_\eta^* A_\eta)^{1/2}\| \le \gamma_* n^{1/(2k)}$, where

$$\gamma_* = \sup_{n>0} \left(n^{-1/(2k)} \sup_{0 \le \lambda \le M} \lambda^{1/2} |g_n(\lambda)| \right) \le 2k^{1/2} \alpha^{1/(2k)} \quad [1, \ p. \ 141].$$

Since $\|y_\delta - A_\eta x^*\| \le \|y_\delta - y\| + \|y - A_\eta x^*\| = \|y_\delta - y\| + \|Ax^* - A_\eta x^*\| \le \delta + \eta\|x^*\|$, it follows that $\|g_n(A_\eta^* A_\eta) A_\eta^* (y_\delta - A_\eta x^*)\| \le 2k^{1/2} \alpha^{1/(2k)} n^{1/(2k)} (\delta + \|x^*\|\eta)$. That is why

$$\|x_{n(\delta,\eta)} - x^*\| \le \|K_{n\eta} x^*\| + \|g_n(A_\eta^* A_\eta) A_\eta^* (y_\delta - A_\eta x^*)\| \le \|K_{n\eta} x^*\| +$$
$$+ 2k^{1/2} \alpha^{1/(2k)} n^{1/(2k)} (\delta + \eta\|x^*\|).$$

Let us show that $\|K_{n\eta} x^*\| \to 0$ at $n \to \infty$, $\eta \to 0$. Actually,

$$\|K_{n\eta} x^*\| = \left\|(E - A_\eta^* A_\eta g_n(A_\eta^* A_\eta)) x^*\right\| =$$

$$= \left\| \int_0^{\|A_\eta^* A_\eta\|} (1 - \lambda g_n(\lambda)) dE_\lambda x^* \right\| = \left\| \int_0^{\|A_\eta^* A_\eta\|} \frac{(1 - \alpha\lambda^k)^{2n}}{(1 + \alpha^2 \lambda^{2k})^n} dE_\lambda x^* \right\| \le$$

$$\le \left\| \int_0^{\varepsilon} \frac{(1 - \alpha\lambda^k)^{2n}}{(1 + \alpha^2 \lambda^{2k})^n} dE_\lambda x^* \right\| + \left\| \int_\varepsilon^{\|A_\eta^* A_\eta\|} \frac{(1 - \alpha\lambda^k)^{2n}}{(1 + \alpha^2 \lambda^{2k})^n} dE_\lambda x^* \right\|.$$

Then

$$\left\| \int\limits_{\varepsilon}^{\|A_\eta^* A_\eta\|} \frac{(1-\alpha\lambda^k)^{2n}}{(1+\alpha^2\lambda^{2k})^n} dE_\lambda x^* \right\| \le q^n(\varepsilon) \left\| \int\limits_{\varepsilon}^{\|A_\eta^* A_\eta\|} dE_\lambda x^* \right\| \to 0, \quad n\to\infty,$$

as for $\lambda \in \left[\varepsilon, \|A_\eta^* A_\eta\|\right]$, $\dfrac{(1-\alpha\lambda^k)^2}{1+\alpha^2\lambda^{2k}} \le q(\varepsilon) < 1$.

$$\left\| \int\limits_0^{\varepsilon} \frac{(1-\alpha\lambda^k)^{2n}}{(1+\alpha^2\lambda^{2k})^n} dE_\lambda x^* \right\| \le \left\| \int\limits_0^{\varepsilon} dE_\lambda x^* \right\| = \|E_\varepsilon x^*\| \to 0, \quad \varepsilon \to 0$$

owing to integrated spectrum properties [3–4].

From statement (15) $n^{1/k}(\delta+\eta)^2 \to 0$ at $n\to\infty$, $\delta\to 0$, $\eta\to 0$. Hence $2k^{1/2}\alpha^{1/(2k)}n^{1/(2k)}\left(\delta+\eta\|x^*\|\right)\to 0$, $n\to\infty$, $\delta\to 0$, $\eta\to 0$. Thus,

$$\left\|x_{n(\delta,\eta)} - x^*\right\| \to 0, \quad n\to\infty, \quad \delta\to 0, \quad \eta\to 0.$$

Theorem 3 is proved. □

The following is valid

Theorem 4. *Let* $A, A_\eta \in \pounds(H,F)$, $\|A-A_\eta\| \le \eta$, $\|A_\eta\|^2 \le M$, $(0<\eta\le\eta_0)$, $\alpha>0$, $y\in R(A)$, $\|y_\delta - y\| \le \delta$. *If the exact solution can be represented as* $x^* = |A|^s z$, $s>0$, $\|z\|\le\rho$, $|A| = (A^*A)^{1/2}$ *and conditions* (5), (6) *are satisfied, then estimation error is real*

$$\left\|x_{n(\delta,\eta)} - x^*\right\| \le \gamma_0 c_s\left(1+|\ln\eta|\right)\eta^{\min(1,s)}\rho +$$
$$+\gamma_{s/2}n^{-s/(2k)}\rho + 2k^{1/2}\alpha^{1/(2k)}n^{1/(2k)}\left(\delta + \|x^*\|\eta\right), \ 0<s<\infty.$$

Proof. In case of sourcewise representable exact solution $x^* = |A|^s z = (A^*A)^{s/2} z$ owing to (6) we get $\sup\limits_{0\le\lambda\le M} \lambda^{s/2}|1-\lambda g_n(\lambda)| \le \gamma_{s/2} n^{-s/(2k)}$, where $\gamma_{s/2} = \left(\dfrac{s}{4k\alpha e}\right)^{s/(2k)}$. Then

$$\|K_{n\eta}|A_\eta|^s z\| = \left\||A_\eta|^s\left[E - A_\eta^*A_\eta g_n\left(A_\eta^*A_\eta\right)\right]z\right\| =$$
$$= \left\|\left(A_\eta^*A_\eta\right)^{s/2}\left[E - A_\eta^*A_\eta g_n\left(A_\eta^*A_\eta\right)\right]z\right\| \le \gamma_{s/2}n^{-s/(2k)}\rho.$$

Hence

$$\|K_{n\eta}x^*\| = \|K_{n\eta}|A|^s z\| = \|K_{n\eta}\left(|A_\eta|^s - |A|^s\right)z\| +$$
$$+\|K_{n\eta}|A_\eta|^s z\| \le \gamma_0 c_s\left(1+|\ln\eta|\right)\eta^{\min(1,s)}\rho + \gamma_{s/2}n^{-s/(2k)}\rho,$$

since according to [5, p. 92] we have $\||A_\eta|^s - |A|^s\| \le c_s\left(1+|\ln\eta|\right)\eta^{\min(1,s)}$, $c_s = const$, ($c_s \le 2$ for $0<s\le 1$). Following (16)

$$\begin{aligned}\left\|x_{n(\delta,\eta)} - x^*\right\| &\le \|K_{n\eta}x^*\| + \gamma_* n^{1/(2k)}\left(\delta+\|x^*\|\eta\right) = \|K_{n\eta}x^*\| + \\ &+ 2k^{1/2}\alpha^{1/(2k)}n^{1/(2k)}\left(\delta+\|x^*\|\eta\right) \le \gamma_0 c_s\left(1+|\ln\eta|\right)\eta^{\min(1,s)}\rho + \\ &+ \gamma_{s/2}n^{-s/(2k)}\rho + 2k^{1/2}\alpha^{1/(2k)}n^{1/(2k)}\left(\delta+\|x^*\|\eta\right), \quad 0<s<\infty.\end{aligned} \tag{17}$$

Theorem 4 is proved. □

By minimizing the right-hand member (17) at n, the meaning of a priori stop moment is obtained:

$$n_{opt} = \left(\frac{s\gamma_{s/2}}{\gamma_*}\right)^{2k/(s+1)} \rho^{2k/(s+1)} \left(\delta + \|x^*\|\,\eta\right)^{-2k/(s+1)} =$$
$$= (4k)^{-(s+k)/(s+1)} s^{(2k+s)/(s+1)} e^{-s/(s+1)} \alpha^{-1} \rho^{2k/(s+1)} \left(\delta + \|x^*\|\,\eta\right)^{-2k/(s+1)}.$$

The substitution of n_{opt} into estimation (17) allows obtaining the optimal error estimation for the method of iterations (11)

$$\left\|x_{n(\delta,\eta)} - x^*\right\|_{opt} \leq \gamma_0 c_s \left(1 + |\ln\eta|\right) \eta^{\min(1,s)} \rho +$$
$$+ c''_s \rho^{1/(s+1)} \left(\delta + \|x^*\|\,\eta\right)^{s/(s+1)}, \quad 0 < s < \infty,$$

where

$$c''_s = \left(s^{1/(s+1)} + s^{-s/(s+1)}\right) \gamma_*^{s/(s+1)} \gamma_{s/2}^{1/(s+1)} =$$
$$= s^{s(1-2k)/(2k(s+1))} (s+1)(4k)^{s(k-1)/(2k(s+1))} e^{-s/(2k(s+1))}.$$

To sum it up,

$$\left\|x_{n(\delta,\eta)} - x^*\right\|_{opt} \leq c_s \left(1 + |\ln\eta|\right) \eta^{\min(1,s)} \rho + s^{s(1-2k)/(2k(s+1))} (s+1) \times$$
$$\times (4k)^{s(k-1)/(2k(s+1))} e^{-s/(2k(s+1))} \rho^{1/(s+1)} \left(\delta + \|x^*\|\,\eta\right)^{s/(s+1)}, 0 < s < \infty.$$

Bibliography

1. Matysik O. V. Explicit and implicit iteration procedures of solving ill-posed problems / O. V. Matysik. — Brest: BrSU named after A. S. Pushkin, 2014. -– 213 p. (in Russian).
2. Matysik O. V. The implicit iterative method of solving operator equations of first kind in Hilbert space / O. V. Matysik // Reports of National Academy of Sciences. -– 2012. -– Vol. 56, No 6. -– P. 28–33. (In Russian).
3. Kantorovich L. V. Functional analysis in normalized spaces / L. V. Kantorovich, G. P. Akilov. -– Moscow: Phismathgiz, 1959. – 680 p. (in Russian).
4. Lyusternik L. A. Elements of functional analysis / L. A. Lyusternik, V. I. Sobolev. -– Moscow: Nauka, 1965. -– 520 p. (in Russian).
5. Vainikko G. M. Iteration procedures in ill-posed problems / G. M. Vainikko, A. Yu. Veretennikov. -– Moscow: Nauka, 1986. -– 178 p. (in Russian).

Oleg Matysik, Brest State University named after A. S. Pushkin, 21, Kosmonavtov boulevard, Brest, 224016, Belarus

Received 21.05.2014

Журнал обчислювальної та прикладної математики | 2014 № 2 (116) | Journal of Computational & Applied Mathematics

UDC 519.6

ON THE BOUNDARY INTEGRAL EQUATIONS METHOD FOR EXTERIOR BOUNDARY VALUE PROBLEMS FOR INFINITE SYSTEMS OF ELLIPTIC EQUATIONS OF SPECIAL KIND

Yuriy Muzychuk

Резюме. В тривимірних обмежених областях з ліпшицевою межею розглянуто зовнішні граничні задачі для нескінченних систем еліптичних рівнянь спеціального трикутного вигляду зі змінними коефіцієнтами. Сформульовано варіаційні постановки задач Діріхле, Неймана та Робіна та встановлено їхню коректність у відповідних просторах Соболєва. За допомогою введеного поняття q-згортки отримано аналоги першої та другої формул Гріна та побудовано інтегральні зображення розв'язків розглянутих задач у випадку сталих коефіцієнтів. Досліджено властивості інтегральних операторів та коректність отриманих систем граничних інтегральних рівнянь.

Abstract. Boundary value problems for infinite triangular systems of elliptic equations with variable coefficients are considered in exterior 3d Lipschitz domains. Variational formulations of Dirichlet, Neumann and Robin problems are received and their well-posedness in corresponding Sobolev spaces is established. Via the introduced q-convolution the analogues of the first and the second Green's formulae are obtained and integral representations of the generalized solutions for formulated problems in the case of constant coefficients are built. We investigate the properties of integral operators and well-posedness of received systems of boundary integral equations.

1. Introduction

The method of boundary integral equations (BIEs) can be applied to a wide class of boundary value problems (BVPs) for elliptic partial differential equations (PDEs). Theoretical aspects of this method have been well investigated in the literature, see, e.g.,[1, 2], and the references therein. The main advantage of the BIEs method is the reduction by one of the dimension of the problem by switching to unknown functions that are defined only on the domain's boundary. It is particularly suited for exterior problems in unbounded domains. Numerous engineering applications confirm the efficiency of this method.

In the case of initial-boundary value problems for evolution equations, the BIEs method can be used both for the BVP investigations and for their effective numerical solution, see, e.g.,[3, 4, 5, 6]. But since the time and space variables are intertwined in the kernel of boundary integral operators it makes the application of this method more complicated. Therefore when solving the

Key words. Boundary value problems; boundary integral equations; elliptic equation; infinite system; variational formulation.

BIEs that depend on the time and space variables, besides the Galerkin or collocation methods, specialized approaches for consideration of the time variable are used. Such composite methods have been studied in the works cited above. They have certain characteristics that define the features of the algorithm implementation. For instance, usage of the so-called Convolution Quadrature method [5] or the Laguerre transform of the time variable [7, 8] leads to solving sequences of BIEs.

In [9] the BIEs method was used for finding solutions of interior BVPs for infinite triangular systems which one could obtain from evolution equations by the Laguerre transform in the time domain. The idea of this method lies in the generalization of the concept of the potential on infinite sequences of functions that depend only on the space variables. Herewith the convolutions of the Cauchy data of the unknown solution with fundamental solution of the infinite system and its normal derivative are used. Application of such convolution of the infinite sequences to the particular problem leads to a sequence of BIEs that has he same operator of the left-hand side and the expression in the right-hand side contains solutions of the previous BIEs. In this paper we extend this approach for exterior problems.

Traditionally the BIEs method is used for BVPs with constant coefficients, since in case of variable coefficients PDE's fundamental solutions are generally not explicitly available. Still on the stage of investigation of the well-posedness of BVPs we will consider a system with variable coefficients. Note that such problems can be treated as some generalization of BVPs that arise as a result of the application of the Laguerre transform to the non-stationary problems.

The paper is organized as follows. In Section 2 we formulate the Dirichlet, Neumann and Robin BVPs for some kind of infinite triangular system consisting of elliptic PDEs with variable coefficients. We consider these problems in appropriate Sobolev spaces and show their well-posedness. Then we introduce the notion of sequences and a new operation on them - the q-convolution of sequences. In this section we also consider variational formulations of the corresponding BVPs and arrive at the analogues of the first and the second Green's formulae. In Section 4 we obtain the integral representation of the solution of the BVPs with constant coefficients and establish a relationship between the Cauchy data of some generalized solution and corresponding BIEs which we study in the following Section 5.

2. Formulation of the BVPs and basic relations

Let $\Omega \subset \mathbb{R}^3$ be a bounded and simply connected domain with Lipschitz boundary Γ and $\Omega^+ := \mathbb{R}^3 \setminus \overline{\Omega}$ be an exterior domain. We consider an infinite system in Ω^+

$$\begin{cases} Pu_0 = f_0\ , \\ c_{1,0}u_0 + Pu_1 = f_1\ , \\ c_{2,0}u_0 + c_{2,1}u_1 + Pu_2 = f_2\ , \\ \quad \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ c_{k,0}u_0 + c_{k,1}u_1 + ... + c_{k,k-1}u_{k-1} + Pu_k = f_k\ , \\ \quad \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \end{cases} \tag{1}$$

where $u_0, u_1, \ ..., \ u_k, \ ...$ are unknown functions, $c_{i,j}$ $(i, j \in \mathbb{N}_0 := \mathbb{N} \cup \{0\})$ are some given measurable and bounded in Ω^+ functions with $c_{i,j} = 0$ when $j \geq i$; f_i $(i \in \mathbb{N}_0)$ are given in Ω^+ functions (functionals). In a formal second order differential operator

$$(Pu)(x) := -\sum_{i,j=1}^{3} \frac{\partial}{\partial x_j}\left[a_{i,j}(x)\frac{\partial u(x)}{\partial x_i}\right] + a_0(x)u(x), \quad x \in \Omega^+, \tag{2}$$

the functions $a_{i,j}$ $(i, j = 1, 2, 3)$ and a_0 are measurable and bounded and satisfy the conditions:

$$a_{i,j}(x) = a_{j,i}(x) \ \ (i, j = 1, 2, 3) \text{ for almost all } x \in \Omega^+,$$

$$\sum_{i,j=1}^{3} a_{i,j}(x)\xi_i\xi_j \geq \alpha \sum_{i=1}^{3} \xi_i^2 \ \text{ for arbitrary } \xi_1, \xi_2, \xi_3 \in \mathbb{R} \text{ and almost all } x \in \Omega^+, \tag{3}$$

with some constant $\alpha > 0$ and

$$a_0(x) > 0 \text{ for almost all } x \in \Omega^+. \tag{4}$$

Let the unit normal vector $\bar{\nu}(x) = (\nu_1(x), \nu_2(x), \nu_3(x))$ to Γ be directed outwards of Ω^+. We investigate BVPs for system (1) that consist in finding its solutions that satisfy one of the following conditions on the boundary Γ

(i) Dirichlet condition:

$$u_k|_\Gamma = \tilde{h}_k, \ k \in \mathbb{N}_0, \tag{5}$$

(ii) Neumann condition:

$$\partial_{\bar{\nu}} u_k|_\Gamma = \tilde{g}_k, \ k \in \mathbb{N}_0, \tag{6}$$

(iii) Robin condition:

$$\left(\partial_{\bar{\nu}} u_k - (b_{k,0}u_0 + b_{k,1}u_1 + ... + b_{k,k-1}u_{k-1} + b_{k,k}u_k)\right)|_\Gamma = \tilde{g}_k, \ k \in \mathbb{N}_0, \tag{7}$$

where $\tilde{h}_i, \ \tilde{g}_i$ $(i \in \mathbb{N}_0)$ are given functions (functionals) on Γ, $b_{i,j} \in L^\infty(\Gamma)$ $(i, j \in \mathbb{N}_0)$ are given functions on Γ with $b_{i,j} = 0$ when $j > i \geq 0$, $b_{i,i} \geq \tilde{b}_i > 0$, $\tilde{b}_i$ – constants. In other words, we will consider the Dirichlet problem (1), (5), the Neumann problem (1), (6) and the Robin problem (1), (7).

Note that the triangular form of system (1) allows us to consequently find the unknown functions $u_k, \ k \in \mathbb{N}_0$. This way when solving the k-th equation ($k \geq 1$) we assume that all solutions $u_i, \ 0 \leq i \leq k-1$, have been found on previous steps and move them to the right hand side of the equation. For instance, we will use this approach for the investigation of the well-posedness of the previously mentioned BVPs. But it isn't suitable for their numerical solution with usage of potentials since it requires additional calculation of volume potentials for combinations of functions $u_i, \ 0 \leq i \leq k-1$, found on previous steps. The method introduced in [9] regarding the interior problems for system (1) allows us to avoid this and build an efficient algorithm for their numerical solution.

We will use the Lebesgue space $L_2(\Omega^+)$ and Sobolev spaces $H^1(\Omega^+)$ and $H_0^1(\Omega^+)$ of real-valued scalar functions and dual to them $\widetilde{H}^{-1}(\Omega^+) :=$

$\left(H^1(\Omega^+)\right)'$ and $H^{-1}(\Omega^+) := \left(H^1_0(\Omega^+)\right)'$, correspondingly. Under $\mathcal{D}(\Omega^+)$ and $\mathcal{D}'(\Omega^+)$ we will understand the spaces of all test functions and distributions on them.

The following bilinear form

$$a_{\Omega^+}(u,v) := \int_{\Omega^+} \left[\sum_{i,j=1}^{3} a_{i,j}(x) \frac{\partial u(x)}{\partial x_i} \frac{\partial v(x)}{\partial x_j} + a_0(x) u(x) v(x) \right] dx \qquad (8)$$

is well defined for any functions $u, v \in H^1(\Omega^+)$. It is known, see, e.g.,[10] and [3, 6] for the case of constant coefficients, one can consider $a_{\Omega^+}(\cdot,\cdot)$ as an inner product and introduce in $H^1(\Omega^+)$ a new norm $|||u||| := (a_{\Omega^+}(u,u))^{1/2}$, which is equivalent to the usual one under the conditions (3) and (4). It is obvious that this form is $H^1(\Omega^+)$-elliptic.

In $H^1(\Omega^+)$ we will consider the following subspace

$$H^1(\Omega^+, P) := \left\{ \ u \in H^1(\Omega^+) \ \middle| \ Pu \in L_2(\Omega^+) \ \right\}, \qquad (9)$$

equipped with the norm

$$||u||_{H^1(\Omega^+,P)} := \left(\ ||u||^2_{H^1(\Omega^+)} + ||Pu||^2_{L_2(\Omega^+)} \ \right)^{1/2}. \qquad (10)$$

Let $\gamma_0^+ : H^1(\Omega^+) \to H^{1/2}(\Gamma)$ be the trace operator and $\gamma_1^+ : H^1(\Omega^+, P) \to H^{-1/2}(\Gamma)$ be the conormal derivative operator, which coincides with the conormal derivative

$$\partial_{\bar{\nu}} u(x) := \sum_{i,j=1}^{3} a_{i,j}(x) \frac{\partial u(x)}{\partial x_i} \nu_j(x), \ x \in \Gamma$$

in case of functions from $H^2(\Omega^+)$, a sufficiently smooth boundary Γ and continuous on $\overline{\Omega^+}$ coefficients $a_{i,j}$ $(i,j = 1,2,3)$. It is known ([1], Theorem 4.4), that for functions $u \in H^1(\Omega^+, P)$ and $v \in H^1(\Omega^+)$ the first Green's formula holds

$$(Pu, v)_{\Omega^+} = a_{\Omega^+}(u,v) + \langle \gamma_1^+ u, \gamma_0^+ v \rangle_\Gamma. \qquad (11)$$

where $(\cdot,\cdot)_{\Omega^+}$ and $\langle\cdot,\cdot\rangle_\Gamma$ denote the $L_2(\Omega^+)$ the inner product and the duality between $H^{-1/2}(\Gamma)$ and $H^{1/2}(\Gamma)$, correspondingly. If $u \in H^1(\Omega^+)$ then the form $a_{\Omega^+}(\cdot,\cdot)$ can also be used for the definition of $Pu \in H_0^{-1}(\Omega^+)$

$$\langle Pu, v \rangle_{\Omega^+,1,0} := a_{\Omega^+}(u,v), \ \forall v \in H^1_0(\Omega^+). \qquad (12)$$

Here $\langle\cdot,\cdot\rangle_{\Omega^+,1,0}$ denotes the duality between $H^{-1}(\Omega^+)$ and $H^1_0(\Omega^+)$.

Let X be an arbitrary linear space over the field of real numbers, $\mathbb{Z}$ – the set of integers. By X^∞ we denote a linear space of mappings $\mathbf{u} : \mathbb{Z} \to X$ satisfying $u(k) = 0$ when $k < 0$. For any element $\mathbf{u} \in X^\infty$ we have $u_k \equiv (\mathbf{u})_k := \mathbf{u}(k), k \in \mathbb{Z}$, and will write it as $\mathbf{u} := (u_0, u_1, ..., u_k, ...)^\top$. Henceforth we will call elements of X^∞ sequences.

We will use triangular matrix operators $\mathbf{C} : (L_2(\Omega^+))^\infty \to (L_2(\Omega^+))^\infty$ and $\mathbf{B} : (L_2(\Gamma))^\infty \to (L_2(\Gamma))^\infty$ that act as $(\mathbf{Cu})_k = \sum_{l=0}^{k} c_{k,l} \cdot (\mathbf{u})_l, \ k \in \mathbb{N}_0$, and $(\mathbf{Bu})_k = \sum_{l=0}^{k} b_{k,l} \cdot (\mathbf{u})_l, \ k \in \mathbb{N}_0$, where $c_{k,l}$ and $b_{k,l}$ are the coefficients of the system (1) and of the Robin boundary condition (7), correspondingly.

The following denotations of sequences are used

$$\mathbf{a}_{\Omega^+}(\mathbf{u},\mathbf{v}) := (a_{\Omega^+}(u_0,v_0),\ a_{\Omega^+}(u_1,v_1),\ ...)^\top,\ \ \mathbf{u},\mathbf{v} \in (H^1(\Omega^+))^\infty,$$

and

$$(\mathbf{u},\mathbf{v})_X := ((u_0,v_0)_X,\ (u_1,v_1)_X,\ ...)^\top,\ \ \mathbf{u},\mathbf{v} \in (X)^\infty,$$

where X is some Hilbert space. In the same manner we will denote sequences for duality pairing. For example, if $\mathbf{u} \in H^{-1/2}(\Gamma)$ and $\mathbf{v} \in H^{1/2}(\Gamma)$ we will use the notation $\langle \mathbf{u},\mathbf{v}\rangle_\Gamma := (\langle u_0,v_0\rangle_\Gamma, \langle u_1,v_1\rangle_\Gamma, ...)^\top$. Analogously, linear functionals on sequences will be treated as component-wise. For the sequence $\mathbf{u} \in (H^1(\Omega^+))^\infty$ we introduce the definition of an exterior trace as a sequence of traces of its components, i.e. $\gamma_0^+\mathbf{u} := (\gamma_0^+u_0, \gamma_0^+u_1, ...)^\top$ will be called an exterior trace of the sequence $\mathbf{u}$ on the surface Γ. If $\mathbf{u} \in (H^1(\Omega^+,P))^\infty$ the sequence $\gamma_1^+\mathbf{u} := (\gamma_1^+u_0, \gamma_1^+u_1, ...)^\top$ will denote an exterior conormal derivative of the sequence $\mathbf{u}$ on the domain's boundary.

Taking into account previous definitions, generalized solutions of the Dirichlet, Neumann and Robin BVPs for system (1) can be defined in the following way.

Definition 1. Let $\mathbf{f} \in (H^{-1}(\Omega))^\infty$ and $\tilde{\mathbf{h}} \in (H^{1/2}(\Gamma))^\infty$. Sequence $\mathbf{u} \in (H^1(\Omega^+))^\infty$ is called a *generalized solution of the Dirichlet problem* (1)*,* (5) if it satisfies the variational equality

$$\mathbf{a}_{\Omega^+}(\mathbf{u},\mathbf{v}) + (\mathbf{C}\mathbf{u},\mathbf{v})_{\Omega^+} = \langle \mathbf{f},\mathbf{v}\rangle_{\Omega^+,1,0},\ \ \forall \mathbf{v} \in (H_0^1(\Omega^+))^\infty, \tag{13}$$

and the boundary condition

$$\gamma_0^+\mathbf{u} = \tilde{\mathbf{h}} \quad \text{on } \Gamma. \tag{14}$$

Definition 2. Let $\mathbf{f} \in (\widetilde{H}^{-1}(\Omega))^\infty$ and $\tilde{\mathbf{g}} \in (H^{-1/2}(\Gamma))^\infty$. Sequence $\mathbf{u} \in (H^1(\Omega^+))^\infty$ is called a *generalized solution of the Neumann problem* (1)*,* (6) if it satisfies the variational equality

$$\mathbf{a}_{\Omega^+}(\mathbf{u},\mathbf{v}) + (\mathbf{C}\mathbf{u},\mathbf{v})_{\Omega^+} = \langle \mathbf{f},\mathbf{v}\rangle_{\Omega^+,1} - \langle \tilde{\mathbf{g}},\gamma_0^+\mathbf{v}\rangle_\Gamma,\ \ \forall \mathbf{v} \in (H^1(\Omega^+))^\infty. \tag{15}$$

Here $\langle\cdot,\cdot\rangle_{\Omega^+,1}$ denotes the duality between $\widetilde{H}^{-1}(\Omega^+)$ and $H^1(\Omega^+)$.

Definition 3. Let $\mathbf{f} \in (\widetilde{H}^{-1}(\Omega))^\infty$ and $\tilde{\mathbf{g}} \in (H^{-1/2}(\Gamma))^\infty$. Sequence $\mathbf{u} \in (H^1(\Omega^+))^\infty$ is called a *generalized solution of the Robin problem* (1)*,* (7) if it satisfies the variational equality

$$\begin{aligned}\mathbf{a}_{\Omega^+}(\mathbf{u},\mathbf{v}) + (\mathbf{C}\mathbf{u},\mathbf{v})_{\Omega^+} + \langle \mathbf{B}\gamma_0^+\mathbf{u},\gamma_0^+\mathbf{v}\rangle_\Gamma &= \\ = \langle \mathbf{f},\mathbf{v}\rangle_{\Omega^+,1} - \langle \tilde{\mathbf{g}},\gamma_0^+\mathbf{v}\rangle_\Gamma,\ \ \forall \mathbf{v} &\in (H^1(\Omega^+))^\infty.\end{aligned} \tag{16}$$

Theorem 1. *The Dirichlet boundary value problem* (1)*,* (5) *has a unique generalized solution.*

Proof. The triangular form of the system (13) gives us opportunity to consider its equations one after another and apply the same standard procedure for investigation of variational equations (see, e.g. [2]) on each step of the proof. Let's start with the first equation:

$$a_{\Omega^+}(u_0,v) = \langle f_0,v\rangle_{\Omega^+,1,0},\ \ \forall v \in H_0^1(\Omega^+).$$

According to the trace theorem for each function $\tilde{h}_k \in H^{1/2}(\Gamma)$ there exists a (non-unique) element $\tilde{u}_k \in H^1(\Omega^+)$ that $\gamma_0^+ \tilde{u}_k = \tilde{h}_k$. Therefore, we can obtain the following variational equation for the difference $u_0 - \tilde{u}_0 =: w \in H_0^1(\Omega^+)$

$$a_{\Omega^+}(w, v) = \langle \tilde{f}_0, v\rangle_{\Omega^+,1,0} := \langle f_0, v\rangle_{\Omega^+,1,0} - a_{\Omega^+}(\tilde{u}_0, v), \quad \forall v \in H_0^1(\Omega^+). \qquad (17)$$

Due to the $H^1(\Omega^+)$-ellipticity of the bilinear form and the boundedness of the functional $\tilde{f}_0$ on $H_0^1(\Omega^+)$ according to the Lax-Milgram theorem this equation has a unique solution $w \in H_0^1(\Omega^+)$. This proves existence of the unique function $u_0 \in H^1(\Omega^+)$ that is a generalized solution of the first problem.

When considering the second variational equation we move the function u_0 into the right hand side of the corresponding equation and for the difference $u_1 - \tilde{u}_1 =: w \in H_0^1(\Omega^+)$ we arrive at the variational equation that differs from (17) only by the right hand side. Therefore, by using the previous considerations we prove the assertion of the theorem for the solution u_1. Obviously, acting this way on each succeeding step we will obtain the variational equation (17) with the following right hand side

$$\langle \tilde{f}_k, v\rangle_{\Omega^+,1,0} := \langle f_k, v\rangle_{\Omega^+,1,0} - \sum_{i=0}^{k-1} (c_{k,i}\tilde{u}_i, v)_{\Omega^+} - a_{\Omega^+}(\tilde{u}_k, v),$$
$$\forall v \in H_0^1(\Omega^+), \quad k \in \mathbb{N}.$$

Here $u_i \, (i = \overline{0, k-1})$ are generalized solutions of the problems considered on the previous steps. As can be seen $\tilde{f}_k \in H^{-1}(\Omega^+)$. Hence, there exists a unique generalized solution of the current BVP. Therefore, for each BVP with an arbitrary index $k \in \mathbb{N}$ the generalized solution $u_k \in H^1(\Omega^+)$ exists and is unique. □

Theorem 2. *The Robin boundary value problem* (1), (7) *has a unique generalized solution.*

Proof. Let's consider the first equation of system (16):

$$a_{\Omega^+}(u_0, v) + b_{\Gamma,0}(u_0, v) = \langle f_0, v\rangle_{\Omega^+,1} - \langle \tilde{g}_0, \gamma_0^+ v\rangle_\Gamma, \quad \forall v \in H^1(\Omega^+). \qquad (18)$$

Here the bilinear form $b_{\Gamma,k}(\cdot,\cdot)$ $(k \in \mathbb{N}_0)$ is expressed through traces of elements of space $H^1(\Omega^+)$ on the boundary Γ:

$$b_{\Gamma,k}(u, v) = \int_\Gamma b_{k,k}(x)\gamma_0^+ u(x)\gamma_0^+ v(x) dS_x, \; u, v \in H^1(\Omega^+).$$

As long as $b_{k,k} \in L^\infty(\Gamma)$ and $\gamma_0 u, \gamma_0 v \in H^{1/2}(\Gamma) \subset L_2(\Gamma)$, such integral exists. Expression

$$\tilde{a}_{\Omega^+}(u, v) := a_{\Omega^+}(u, v) + b_{\Gamma,0}(u, v), \quad u, v \in H^1(\Omega^+), \qquad (19)$$

can be treated as some bilinear form for $u, v \in H^1(\Omega^+)$. Obviously, it is $H^1(\Omega^+)$-elliptic.

On the other hand, taking into account the estimate

$$|\langle \tilde{g}_0, \gamma_0^+ v\rangle_\Gamma| \leq ||\tilde{g}_0||_{H^{-1/2}(\Gamma)} ||\gamma_0^+ v||_{H^{1/2}(\Gamma)} \leq C||\tilde{g}_0||_{H^{-1/2}(\Gamma)} ||v||_{H^1(\Omega^+)}$$

the functional

$$\langle \tilde{f}_0, v\rangle_{\Omega^+,1} := \langle f_0, v\rangle_{\Omega^+,1} - \langle \tilde{g}_0, \gamma_0^+ v\rangle_\Gamma$$

is an element of $\widetilde{H}^{-1}(\Omega^+)$. Then, according to the Lax-Milgram theorem there exists a unique solution $u_0 \in H^1(\Omega^+)$ of the equation (18).

Next we follow the scheme, used in the proof of the previous theorem. Let's consider the equation with an arbitrary index $k \in \mathbb{N}$. After moving all items that contain functions u_i $(i = \overline{0, k-1})$ into the right hand side, this equation takes the form:

$$a_{\Omega^+}(u_k, v) + b_{\Gamma,k}(u_k, v) = \langle \tilde{f}_k, v\rangle_{\Omega^+,1}, \quad \forall v \in H^1(\Omega^+), \quad k \in \mathbb{N}, \tag{20}$$

where

$$\langle \tilde{f}_k, v\rangle_{\Omega^+,1} := \langle f_k, v\rangle_{\Omega^+,1} - \langle \tilde{g}_k, \gamma_0^+ v\rangle_\Gamma - \sum_{i=0}^{k-1} (c_{k,i} u_i, v)_{\Omega^+} - \sum_{i=0}^{k-1} \langle b_{k,i} \gamma_0^+ u_i, \gamma_0^+ v\rangle_\Gamma.$$

Clearly, $\tilde{f}_k \in \widetilde{H}^{-1}(\Omega^+)$. Since the obtained variational equation differs from (18) only in the right hand side, we arrive at the conclusion that there exists its unique solution $u_k \in H^1(\Omega^+)$. Thus we've shown the existence and the uniqueness of each component of the solution of variational system (16). □

As a conclusion of the previous theorem we obtain

Theorem 3. *The Neumann boundary value problem* (1), (6) *has a unique generalized solution.*

Note that condition (4) is a characteristic feature of PDEs obtained from the evolution equations by means of the Laguerre transform. Without such constraint the bilinear form will be just coercive. In this case the existence and the uniqueness of the solutions of BVPs for system (1) can be investigated according to the Fredholm theory, see, e.g., [1, 2], or by considering the variational formulations in corresponding weighted Sobolev spaces [11].

We shall now use the well known procedure (see, e.g. [12], chapter 7) to transform variational problems to the equivalent ones in the operator form. We first consider the variational equation (13) and suppose that the sequence $\mathbf{u} \in (H^1(\Omega^+))^\infty$ is its solution. Bearing in mind (12), we can rewrite it in the following way:

$$\langle \mathbf{Pu}, \mathbf{v}\rangle_{\Omega^+,1,0} + (\mathbf{Cu}, \mathbf{v})_{\Omega^+} = \langle \mathbf{f}, \mathbf{v}\rangle_{\Omega^+,1,0}, \quad \forall \mathbf{v} \in (H_0^1(\Omega^+))^\infty, \tag{21}$$

where the matrix operator $\mathbf{P}$ acts on $\forall \mathbf{u} \in (H^1(\Omega^+))^\infty$ by the rule:

$$(\mathbf{Pu})_k = Pu_k, \quad k \in \mathbb{N}_0.$$

Taking into account the embedding of spaces $H_0^1(\Omega^+) \subset L_2(\Omega^+) \subset H^{-1}(\Omega^+)$, the equality (21) may be presented as

$$\langle \mathbf{Pu}, \mathbf{v}\rangle_{\Omega^+,1,0} + \langle \mathbf{Cu}, \mathbf{v}\rangle_{\Omega^+,1,0} = \langle \mathbf{f}, \mathbf{v}\rangle_{\Omega^+,1,0}, \quad \forall \mathbf{v} \in (H_0^1(\Omega^+))^\infty.$$

After introducing the notation

$$\mathbf{G} := \mathbf{P} + \mathbf{C}, \tag{22}$$

the previous equality can be given in the form of the operator equation

$$\mathbf{Gu} = \mathbf{f} \quad \text{in} \quad (H^{-1}(\Omega^+))^\infty. \tag{23}$$

Thus, the generalized solution of the Dirichlet problem (1), (5) is the solution of the operator equation (23) and satisfies the same boundary condition (5) or its sequence analogue (14). And vice versa, it is easy to see, that the solution of (23), (14) is a generalized solution of the Dirichlet problem (1), (5).

In order to get the operator equation for the Neumann and the Robin problems we will use the Green's formula in the form of (11) instead of (12). We will consider the generalized solutions in space $(H^1(\Omega^+, P))^\infty$ and assume $\mathbf{f} \in (L_2(\Omega^+))^\infty$. Thus, let the sequence $\mathbf{u} \in (H^1(\Omega^+, P))^\infty$ be the generalized solution of the Robin problem (1), (7) i.e. it satisfies the variational equation (16). If we apply the formula (11) to this equation, we get

$$(\mathbf{Pu}, \mathbf{v})_{\Omega^+} - \langle \gamma_1^+ \mathbf{u}, \gamma_0^+ \mathbf{v} \rangle_\Gamma + (\mathbf{Cu}, \mathbf{v})_{\Omega^+} + \\ + \langle \mathbf{B} \gamma_0^+ \mathbf{u}, \gamma_0^+ \mathbf{v} \rangle_\Gamma = (\mathbf{f}, \mathbf{v})_{\Omega^+} - \langle \tilde{\mathbf{g}}, \gamma_0^+ \mathbf{v} \rangle_\Gamma,$$

or

$$(\mathbf{Gu} - \mathbf{f}, \mathbf{v})_{\Omega^+} + \langle \mathbf{B} \gamma_0^+ \mathbf{u} - \gamma_1^+ \mathbf{u} + \tilde{\mathbf{g}}, \gamma_0^+ \mathbf{v} \rangle_\Gamma = 0, \ \ \forall \mathbf{v} \in (H^1(\Omega^+))^\infty. \tag{24}$$

After substitution of an arbitrary element $\mathbf{v} \in (\mathcal{D}(\Omega^+))^\infty$ into (24) we come to the following equality

$$\langle \mathbf{Gu} - \mathbf{f}, \mathbf{v} \rangle_{\Omega^+} = 0,$$

where $\langle \cdot, \cdot \rangle_{\Omega^+}$ is based on the duality between $\mathcal{D}'(\Omega^+)$ and $\mathcal{D}(\Omega^+)$. Thus,

$$\mathbf{Gu} = \mathbf{f} \ \text{ in } (\mathcal{D}'(\Omega^+))^\infty.$$

Since $\mathbf{f} \in (L_2(\Omega^+))^\infty$, the previous equation can be understood as

$$\mathbf{Gu} = \mathbf{f} \ \text{ in } (L_2(\Omega^+))^\infty. \tag{25}$$

Therefore, after substitution of any sequence $\mathbf{v} \in (H^1(\Omega^+))^\infty$ into (24) we arrive at the relation

$$\langle \mathbf{B} \gamma_0^+ \mathbf{u} - \gamma_1^+ \mathbf{u} + \tilde{\mathbf{g}}, \gamma_0^+ \mathbf{v} \rangle_\Gamma = 0 \ \ \forall \mathbf{v} \in (H^1(\Omega^+))^\infty,$$

that, by taking into account that values of the trace operator $\gamma_0^+ : H^1(\Omega^+) \to H^{1/2}(\Gamma)$ fill in the whole space $H^{1/2}(\Gamma)$, is an equivalent form of the Robin boundary condition

$$\gamma_1^+ \mathbf{u} - \mathbf{B} \gamma_0^+ \mathbf{u} = \tilde{\mathbf{g}} \ \text{ in } \ (H^{-1/2}(\Gamma))^\infty. \tag{26}$$

Thus, we have shown that the generalized solution of the Robin problem can be characterized by the operator equation (25) and the boundary condition (26). Analogously it can be shown that the generalized solution of the Neumann problem can be characterized by the same operator equation (25) and the Neumann boundary condition

$$\gamma_1^+ \mathbf{u} = \tilde{\mathbf{g}} \quad \text{in} \ \ (H^{-1/2}(\Gamma))^\infty. \tag{27}$$

Conversely, it is obvious that every solution of the problem (25), (26) (resp. (27)) satisfies the variational problem (16) (resp. (15)).

Note that boundary conditions (26) and (27), as in the theory of elliptic equations, will be referred to as the natural boundary conditions.

3. BVPs in convolution terms

As we have outlined in the introduction, all theoretical and practical aspects of the BIEs method are well known in case of its application to the BVPs for the first equation of the system (1) considered separately as well as for this system as a whole but with a finite number of equations. Henceforth our goal will be to obtain a formula for the solutions of BVPs and appropriate BIEs for the infinite system. Similarly to the previous section, we will use the fact that system (1) is triangular and will develop a recurrent process of the calculation of the components of the solution. To avoid additional volume potentials in the solution representation we will move the components that were found on the previous steps to the right-hand side of the current equation. For this purpose we introduce the following convolution operation on sequences.

Let X, Y and Z be arbitrary linear spaces and $q : X \times Y \to Z$ – some mapping.

Definition 4. By the q-convolution of sequences $\mathbf{u} \in X^\infty$ and $\mathbf{v} \in Y^\infty$ we understand a sequence $\mathbf{w} \in Z^\infty$ that is defined according to the following rule

$$\mathbf{w} = \mathbf{u} \underset{q}{\circ} \mathbf{v}, \tag{28}$$

where $w_n \equiv (\mathbf{u} \underset{q}{\circ} \mathbf{v})_n := \sum_{i=0}^{n} q\left(u_{n-i}, v_i\right)$, when $n \geq 0$, and $w_n = 0$ when $n < 0$.

We will simplify the notation of the q-convolution for some mappings. For instance, in case of $q(u,v) := \langle u, v\rangle_{\Omega^+,1,0}$ we will write $\mathbf{u} \underset{\Omega^+,1,0}{\circ} \mathbf{v} := \mathbf{u} \underset{q}{\circ} \mathbf{v}$.

Consider a sequence $\mathbf{u} \in \left(H^1(\Omega^+)\right)^\infty$ that satisfies the equation (23). Let's substitute it into this equation and, treating the result as equality of elements from $(H^{-1}(\Omega^+))^\infty$ and taking

$$q(w,v) = \langle w, v\rangle_{\Omega^+,1,0},\ v \in H_0^1(\Omega^+),\ w \in H^{-1}(\Omega^+),$$

we apply the q-convolution with an arbitrary sequence $\mathbf{v} \in \left(H_0^1(\Omega^+)\right)^\infty$ to both sides of this equality. After that we arrive at the following variational equation

$$(\mathbf{Gu}) \underset{\Omega^+,1,0}{\circ} \mathbf{v} = \mathbf{f} \underset{\Omega^+,1,0}{\circ} \mathbf{v}, \quad \forall \mathbf{v} \in \left(H_0^1(\Omega^+)\right)^\infty. \tag{29}$$

Thus, the generalized solution of the Dirichlet problem (1), (5) can be characterized by the variational equality (29) and the boundary condition (14).

Now we assume that sequence $\mathbf{u} \in \left(H^1(\Omega^+, P)\right)^\infty$ satisfies the operator equation (25). We apply the q-convolution with some arbitrary sequence $\mathbf{v} \in \left(H^1(\Omega^+)\right)^\infty$ to both of its sides as elements of $(L_2(\Omega^+))^\infty$, taking $q(w,v) = (w,v)_{\Omega^+}$, $v \in H^1(\Omega^+)$, $w \in L_2(\Omega^+)$. As a result we get

$$(\mathbf{Gu}) \underset{\Omega^+}{\circ} \mathbf{v} = \mathbf{f} \underset{\Omega^+}{\circ} \mathbf{v}, \quad \forall \mathbf{v} \in \left(H^1(\Omega^+)\right)^\infty. \tag{30}$$

Thus, the generalized solution of the Robin boundary value problem can be characterized by the variational equality (30) and the boundary condition (26).

Obviously, this property also holds for the generalized solution of the Neumann boundary value problem.

Let's obtain for operator **G** the analogue of the first Green's formula using the q-convolution of sequences. At first note that the component of the q-convolution in the left hand side of (30) with an arbitrary index $k \in \mathbb{N}_0$ after application of the first Green's formula (11) can be written as

$$\begin{aligned}\left((\mathbf{Gu}) \underset{\Omega^+}{\circ} \mathbf{v} \right)_k = \sum_{i=0}^{k} a_{\Omega^+}\left(u_i,\ v_{k-i}\right) + \sum_{i=0}^{k} \left\langle \gamma_1^+ u_i, \gamma_0^+ v_{k-i} \right\rangle_\Gamma + \\ + \sum_{i=1}^{k} \left(\sum_{j=0}^{i-1} c_{i,j} u_j,\ v_{k-i} \right)_{\Omega^+}.\end{aligned} \tag{31}$$

Henceforth we assume that the sum expressions are equal to zero if their last index is less than the first one i.e. in case of $k = 0$ the last item in the previous formula is absent.

Consider a sequence $\left(\Phi_0^+(\mathbf{u},\mathbf{v}),\ \Phi_1^+(\mathbf{u},\mathbf{v}),\ ...,\ \Phi_k^+(\mathbf{u},\mathbf{v}),\ ...\right)^\top$, components of which are such expressions:

$$\begin{aligned}&\Phi_0^+(\mathbf{u},\mathbf{v}) := a_{\Omega^+}\left(u_0,\ v_0\right),\\ &\Phi_k^+(\mathbf{u},\mathbf{v}) := \sum_{i=0}^{k} a_{\Omega^+}\left(u_i,\ v_{k-i}\right) + \sum_{i=1}^{k} \left(\sum_{j=0}^{i-1} c_{i,j} u_j,\ v_{k-i} \right)_{\Omega^+},\ k \in \mathbb{N}_0.\end{aligned} \tag{32}$$

Definition 5. Sequence

$$\mathbf{\Phi}^+(\mathbf{u},\mathbf{v}) = \left(\Phi_0^+(\mathbf{u},\mathbf{v}),\ \Phi_1^+(\mathbf{u},\mathbf{v}),\ ...,\ \Phi_k^+(\mathbf{u},\mathbf{v}),\ ...\right)^\top,\ \mathbf{u},\mathbf{v} \in \left(H^1(\Omega^+)\right)^\infty,$$

defined by the formula (32) is called a bilinear form associated with operator **G**.

Such notation of the bilinear form gives us ability to present the relation (31) in the following way

$$\begin{aligned}(\mathbf{Gu}) \underset{\Omega^+}{\circ} \mathbf{v} = \mathbf{\Phi}^+(\mathbf{u},\mathbf{v}) + \gamma_1^+ \mathbf{u} \underset{\Gamma}{\circ} \gamma_0^+ \mathbf{v},\\ \forall\, \mathbf{u} \in \left(H^1(\Omega^+, P)\right)^\infty,\ \ \mathbf{v} \in \left(H^1(\Omega^+)\right)^\infty,\end{aligned} \tag{33}$$

and treat it as the first Green's formula for the operator **G**. Note that for the left part of the variational equality (29) we can analogously obtain the expression

$$(\mathbf{Gu}) \underset{\Omega^+,1,0}{\circ} \mathbf{v} = \mathbf{\Phi}^+(\mathbf{u},\mathbf{v}), \quad \forall\, \mathbf{u} \in \left(H^1(\Omega^+)\right)^\infty,\ \mathbf{v} \in \left(H_0^1(\Omega^+)\right)^\infty, \tag{34}$$

when using the equality (12).

In general, due to the triangular structure of operator **C**, definition of the second Green's formula may be complicated. In order to apply the classical approach, see, e.g. [2], we need an additional condition on the operator **C**

$$(\mathbf{Cu}) \underset{\Omega^+}{\circ} \mathbf{v} = (\mathbf{Cv}) \underset{\Omega^+}{\circ} \mathbf{u}, \quad \forall\, \mathbf{u},\mathbf{v} \in (L_2(\Omega^+))^\infty, \tag{35}$$

which provides the symmetry of the operator $\mathbf{G}$ with regard to the operation of q-convolution. Then applying (33) twice to the couple of sequences $\mathbf{u}, \mathbf{v} \in \left(H^1(\Omega^+, P)\right)^\infty$ we arrive at the following variational equality.

Theorem 4. *For sequences $\boldsymbol{u}, \boldsymbol{v} \in \left(H^1(\Omega^+, P)\right)^\infty$ the following equality holds:*

$$(\boldsymbol{G}\boldsymbol{u}) \underset{\Omega^+}{\circ} \boldsymbol{v} - (\boldsymbol{G}\boldsymbol{v}) \underset{\Omega^+}{\circ} \boldsymbol{u} = \gamma_1^+ \boldsymbol{u} \underset{\Gamma}{\circ} \gamma_0^+ \boldsymbol{v} - \gamma_1^+ \boldsymbol{v} \underset{\Gamma}{\circ} \gamma_0^+ \boldsymbol{u}. \tag{36}$$

We treat it as the second Green's formula for the operator $\mathbf{G}$. Further in this paper we suppose the operator $\mathbf{C}$ satisfies (35).

4. Integral representation of the solution

Green's formulae and fundamental solutions of the operator $\mathbf{G}$ are the key ingredients of the integral representation of the solutions of the BVPs. As usual we call the sequence $\widetilde{\mathbf{E}}(x,y) = \left(\widetilde{E}_0(x,y),\ \widetilde{E}_1(x,y),\ ...\right)^\top,\ x, y \in \mathbb{R}^3$, a fundamental solution of the operator $\mathbf{G}$, if it satisfies the equation

$$\mathbf{G}\widetilde{\mathbf{E}} = \boldsymbol{\delta}_y \text{ in } \left(\mathcal{D}'(\mathbb{R}^3)\right)^\infty,$$

where $\boldsymbol{\delta}_y(x) = (\delta_y(x),\ \delta_y(x),\ ...)^\top$ and $\delta_y(\cdot) := \delta(\cdot - y)$ is Dirac's delta-function. Henceforth we also assume this operator has constant coefficients and particularly

$$P := -\Delta + \kappa^2. \tag{37}$$

The condition (35) can be rewritten in the form

$$\sum_{k=1}^{n}\sum_{i=0}^{k-1} c_{k,i}\xi_i\eta_{n-k} = \sum_{k=1}^{n}\sum_{i=0}^{k-1} c_{k,i}\eta_i\xi_{n-k}, \quad \forall n \in \mathbb{N},\ \forall \boldsymbol{\xi}, \boldsymbol{\eta} \in \mathbb{R}^\infty. \tag{38}$$

The last feature is natural for system (1) which is obtained as a result of the Laguerre transformation with parameter $\sigma > 0$ of the heat ($\kappa = \sqrt{\sigma}$) or the wave ($\kappa = \sigma$) equation [8]. Note that γ_1^+ now denotes a normal derivative operator. We also recall the well-known fundamental solution of the operator P:

$$\widetilde{E}_0(x,y) := \frac{e^{-\kappa|x-y|}}{4\pi|x-y|},\ x, y \in \mathbb{R}^3. \tag{39}$$

In [13] and references therein the construction of such solutions for the operator $\mathbf{G}$ with constant coefficients has been considered. For instance, if system (1) corresponds to the wave equation, then the fundamental solution's components for the operator $\mathbf{G}$ have the following presentation

$$\widetilde{E}_i(x,y) := \frac{e^{-\kappa|x-y|}}{4\pi|x-y|}\mathcal{L}_i(\kappa|x-y|),\ \ \forall i \in \mathbb{N}_0,\ \ x, y \in \mathbb{R}^3, \tag{40}$$

where $\mathcal{L}_i$ denotes the Laguerre polynomial [16].

By using the q-convolution we build sequences that in analogy to the theory of elliptic equations can be also called potentials. For that we use a sequence $\mathbf{E}(x,y) = (E_0(x,y),\ E_1(x,y),\ ...)^\top$, where

$$E_i(x,y) := \widetilde{E}_i(x,y) - \widetilde{E}_{i-1}(x,y),\ i \in \mathbb{N},\ \ E_0(x,y) = \widetilde{E}_0(x,y),\ x, y \in \mathbb{R}^3, \tag{41}$$

It was shown in [9] that $\mathbf{E}$ is the solution of the equation

$$\mathbf{G}\mathbf{E} = \overline{\boldsymbol{\delta}_y} \text{ in } \left(\mathcal{D}'(\mathbb{R}^3)\right)^{\infty}, \tag{42}$$

where $\overline{\boldsymbol{\delta}_y}(x) = (\,\delta_y(x),\ 0,\ 0,\ \dots)^{\top}$.

Definition 6. Let $\boldsymbol{\lambda} \in \left(H^{1/2}(\Gamma)\right)^{\infty}$ and $\boldsymbol{\mu} \in \left(H^{-1/2}(\Gamma)\right)^{\infty}$. Sequences

$$\mathbf{V}\boldsymbol{\mu}(x) := (\mathbf{V}\boldsymbol{\mu})(x) \equiv \boldsymbol{\mu}(\cdot) \underset{\Gamma}{\circ} \mathbf{E}(x-\cdot),\ x \in \Omega^+, \tag{43}$$

and

$$\mathbf{W}\boldsymbol{\lambda}(x) := (\mathbf{W}\boldsymbol{\lambda})(x) \equiv \partial_{\bar{\nu}(\cdot)}\mathbf{E}(x-\cdot) \underset{\Gamma}{\circ} \boldsymbol{\lambda}(\cdot),\ x \in \Omega^+, \tag{44}$$

are called the single and the double layer potentials of the operator $\mathbf{G}$ on the surface Γ, correspondingly.

Lemma 1. *For arbitrary sequences $\boldsymbol{\lambda} \in \left(H^{1/2}(\Gamma)\right)^{\infty}$ and $\boldsymbol{\mu} \in \left(H^{-1/2}(\Gamma)\right)^{\infty}$ the layer potentials $\boldsymbol{u} = \boldsymbol{V\mu}$ and $\boldsymbol{u} = \boldsymbol{W\lambda}$ are the solutions of the homogeneous equation*

$$\boldsymbol{Gu} = \boldsymbol{0} \ \ (in\ \mathbb{R}^3 \setminus \Gamma). \tag{45}$$

Proof. Proof of the lemma regarding the domain Ω can be found in lemma 5.3 [7] and in case of the domain Ω^+ can be done analogously. □

Similarly to the layer potentials $\mathbf{V}$ and $\mathbf{W}$, by means of the q-convolution we can define the volume potential for the domain Ω^+ and use it to obtain a partial solution of the system (1). Since in this case the difference from the interior problems discussed in [9] is minor we will consider only problems for the homogeneous system (45).

Let $\gamma_0^- : H^1(\Omega) \to H^{1/2}(\Gamma)$ be a trace operator, $\gamma_1^- : H^1(\Omega, P) \to H^{-1/2}(\Gamma)$ be a normal derivative operator and $[\gamma_0 u] := \gamma_0^+ u - \gamma_0^- u$, $[\gamma_1 u] := \gamma_1^+ u - \gamma_1^- u$ are their jumps across the boundary Γ.

Theorem 5. *For the sequence $\boldsymbol{u} \in \left(H^1(\mathbb{R}^3 \setminus \Gamma, P)\right)^{\infty}$ which satisfies the equation (45) in $\mathbb{R}^3 \setminus \Gamma$ the following representation takes place*

$$\boldsymbol{u}(x) = \boldsymbol{W\lambda}(x) - \boldsymbol{V\mu}(x),\ x \in \mathbb{R}^3 \setminus \Gamma, \tag{46}$$

where $\boldsymbol{\lambda} := [\gamma_0 \boldsymbol{u}]$ and $\boldsymbol{\mu} := [\gamma_1 \boldsymbol{u}]$.

Proof. As we can see, the layer potentials consist of the components

$$\begin{aligned} &(V_j\mu)(x) := \langle \mu(\cdot),\ E_j(x-\cdot)\,\rangle_\Gamma,\ \mu \in H^{-1/2}(\Gamma); \\ &(W_j\lambda)(x) := \langle \partial_{\bar{\nu}(\cdot)} E_j(x-\cdot),\ \lambda(\cdot)\,\rangle_\Gamma,\ \lambda \in H^{1/2}.(\Gamma),\ \ j \in \mathbb{N}_0. \end{aligned} \tag{47}$$

Let some function $u_0 \in H^1(\mathbb{R}^3 \setminus \Gamma, P)$ satisfy the equation $Pu = 0$ in $\mathbb{R}^3 \setminus \Gamma$. Then the third Green's formula holds

$$u_0(x) = (W_0\lambda_0)(x) - (V_0\mu_0)(x),\ x \in \mathbb{R}^3 \setminus \Gamma, \tag{48}$$

where $\lambda_0 := [\gamma_0 u_0]$ and $\mu_0 := [\gamma_1 u_0]$. Note that this formula can be derived from the first equality in (36) if we take $v_0(\cdot) = E_0(x,\cdot)$. For the explanation of the corresponding procedure and some aspects of usage of this formula see, e.g. [1, 14] and [3, 4] for the case of operator (37).

We can use this approach for the following components of the sequence $\mathbf{u}$ as well. Let us assume we also have a function $u_1 \in H^1(\mathbb{R}^3 \setminus \Gamma, P)$ provided the pair u_0 and u_1 satisfies the second equation in (45). Then from the second equality in (36) we obtain:

$$\begin{aligned} &-(c_{1,0}v_0 + Pv_1, u_0)_{\Omega^+} - (Pv_0, u_1)_{\Omega^+} = \\ &= \langle \gamma_1^+ u_1,\ \gamma_0^+ v_0 \rangle_\Gamma + \langle \gamma_1^+ u_0,\ \gamma_0^+ v_1 \rangle_\Gamma - \langle \gamma_1^+ v_1,\ \gamma_0^+ u_0 \rangle_\Gamma - \\ &- \langle \gamma_1^+ v_0,\ \gamma_0^+ u_1 \rangle_\Gamma. \end{aligned} \tag{49}$$

If we take $v_0(\cdot) = E_0(x,\cdot)$ and $v_1(\cdot) = E_1(x,\cdot)$ and keep in mind the first two equalities of (42) we obtain for $\forall x \in \Omega^+$:

$$\begin{aligned} -u_1(x) = \langle \gamma_1^+ u_1,\ \gamma_0^+ E_0 \rangle_\Gamma + \langle \gamma_1^+ u_0,\ \gamma_0^+ E_1 \rangle_\Gamma - \\ - \langle \gamma_1^+ E_1,\ \gamma_0^+ u_0 \rangle_\Gamma - \langle \gamma_1^+ E_0,\ \gamma_0^+ u_1 \rangle_\Gamma. \end{aligned}$$

If we use the second Green's formula for the interior domain Ω [9] we will have

$$0 = -\langle \gamma_1^- u_1,\ \gamma_0^+ E_0 \rangle_\Gamma - \langle \gamma_1^- u_0,\ \gamma_0^+ E_1 \rangle_\Gamma + \langle \gamma_1^+ E_1,\ \gamma_0^- u_0 \rangle_\Gamma + \langle \gamma_1^+ E_0,\ \gamma_0^- u_1 \rangle_\Gamma.$$

Therefore, by adding the last two formulae we obtain the representation formula for the component u_1 for $\forall x \in \Omega^+$:

$$u_1(x) = (W_0\lambda_1)(x) + (W_1\lambda_0)(x) - (V_0\mu_1)(x) - (V_1\mu_0)(x). \tag{50}$$

It is straightforward to see that there is the same representation formula for $\forall x \in \Omega$.

Now we consider the equality in (36) with index $k > 1$. After the substitution $v_0(\cdot) = E_0(x,\cdot)$, $v_1(\cdot) = E_1(x,\cdot)$, ..., and $v_k(\cdot) = E_k(x,\cdot)$ all components in it's left hand side will disappear except $(Pv_0, u_k)_{\Omega^+}$. As in previous cases from $(Pv_0, u_k)_{\Omega^+}$ we get $u_k(x)$ for $\forall x \in \Omega^+$ and 0 for $\forall x \in \Omega$. The rest of the proof repeats the same operations as for $k = 1$. □

Main properties of the potentials $\mathbf{V}$ and $\mathbf{W}$ have been studied in the aforementioned work [9]. Here we recall some of them. Let us consider the boundary operators

$$\begin{aligned} &\mathbf{V} : (H^{-1/2}(\Gamma))^\infty \to (H^{1/2}(\Gamma))^\infty, \quad \mathbf{K}' : (H^{-1/2}(\Gamma))^\infty \to (H^{-1/2}(\Gamma))^\infty, \\ &\mathbf{K} : (H^{1/2}(\Gamma))^\infty \to (H^{1/2}(\Gamma))^\infty, \qquad \mathbf{D} : (H^{1/2}(\Gamma))^\infty \to (H^{-1/2}(\Gamma))^\infty, \end{aligned}$$

defined by means of q-convolution in the following way:

$$\begin{aligned} &(\mathbf{V}\boldsymbol{\mu})_i := \sum_{j=0}^{i} V_j \mu_{i-j}, \quad (\mathbf{K}\boldsymbol{\lambda})_i := \sum_{j=0}^{i} K_j \lambda_{i-j}, \\ &(\mathbf{K}'\boldsymbol{\mu})_i := \sum_{j=0}^{i} K'_j \mu_{i-j}, \quad (\mathbf{D}\boldsymbol{\lambda})_i := \sum_{j=0}^{i} D_j \lambda_{i-j},\ i \in \mathbb{N}_0, \end{aligned}$$

for arbitrary sequences $\boldsymbol{\lambda} \in \left(H^{1/2}(\Gamma)\right)^\infty$ and $\boldsymbol{\mu} \in \left(H^{-1/2}(\Gamma)\right)^\infty$. Components of these operators are defined as follows:

$$V_j\mu := \gamma_0^+ V_j\mu, \; D_j\lambda := -\gamma_1^+ W_j\lambda, \;\; j \in \mathbb{N}_0,$$

$$K_0'\mu := \gamma_1^+ V_0\mu - \frac{1}{2}\mu, \;\; K_j'\mu := \gamma_1^+ V_j\mu, \;\; j \in \mathbb{N},$$

$$K_0\lambda := \gamma_0^+ W_0\lambda + \frac{1}{2}\lambda, \;\; K_j\lambda := \gamma_0^+ W_j\lambda, \;\; j \in \mathbb{N}.$$

Hence, according to the theorem 5 the generalized solution of the homogeneous system (45) can be given by its trace and the normal derivative on the boundary – the Cauchy data. As it can be seen from the boundary conditions (5) and (7), in each of the boundary problems these data are incomplete. To get the complete Cauchy data we need to consider corresponding BIEs that can be obtained by means of the presentation (46). Note that this is the so-called direct approach [2] to replacement of BVPs by BIEs and in our case it could be implemented taking into account the results obtained in [14, 8]. As a result, the following theorem defines the relation between the Cauchy data of some generalized solution of the homogeneous system and BIEs.

Theorem 6. ***(i)*** *If a pair of sequences* $(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \left(H^{1/2}(\Gamma)\right)^\infty \times \left(H^{-1/2}(\Gamma)\right)^\infty$ *are the Cauchy data of some generalized solution of the equation* (45)*, then they satisfy both equations*

$$\left(\frac{1}{2}\boldsymbol{I} - \boldsymbol{K}\right)\boldsymbol{\lambda} + \boldsymbol{V}\boldsymbol{\mu} = 0 \quad in \;\; (H^{1/2}(\Gamma))^\infty \tag{51}$$

and

$$\boldsymbol{D}\boldsymbol{\lambda} + \left(\frac{1}{2}\boldsymbol{I} + \boldsymbol{K}'\right)\boldsymbol{\mu} = 0 \quad in \;\; (H^{-1/2}(\Gamma))^\infty. \tag{52}$$

(ii) *If a pair of sequences* $(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \left(H^{1/2}(\Gamma)\right)^\infty \times \left(H^{-1/2}(\Gamma)\right)^\infty$ *satisfy one of the equations* (51) *or* (52)*, then they satisfy the second one and are the Cauchy data of some generalized solution of the equation* (45).

Note that for the integral representation of the solution of the PDEs with variable coefficients it is possible to use a parametrix (Levi function) associated with a fundamental solution of corresponding operator with frozen coefficients [11].

5. Boundary integral equations

Theorem 6 gives us reason for the replacement of boundary value problems with corresponding boundary integral equations in regards to the Cauchy datum that is not given explicitly in the formulation of the problem. Due to the similarity of the boundary integral equations that are obtained for interior and exterior problems we will demonstrate this procedure for the Dirichlet problem (1), (5) only. In this case the boundary condition contains the given sequence $\boldsymbol{\lambda} = \tilde{\mathbf{h}} \in (H^{1/2}(\Gamma))^\infty$. Then, taking into account the equation (51), after substitution of the given trace into it we will obtain the following boundary integral equation of the first kind in regards to the sequence $\boldsymbol{\mu}$:

$$\mathbf{V}\boldsymbol{\mu} = \left(-\frac{1}{2}\mathbf{I} + \mathbf{K}\right)\tilde{\mathbf{h}} \quad \text{in} \quad (H^{1/2}(\Gamma))^{\infty}. \tag{53}$$

If we substitute the known trace into the equation (52), we will come to the following boundary integral equation of the second kind

$$\left(\frac{1}{2}\mathbf{I} + \mathbf{K}'\right)\boldsymbol{\mu} = -\mathbf{D}\tilde{\mathbf{h}} \quad \text{in} \quad (H^{-1/2}(\Gamma))^{\infty}. \tag{54}$$

Theorem 7. *The normal derivative of the generalized solution* $\boldsymbol{u} \in (H^1(\Omega, P))^{\infty}$ *of the Dirichlet problem* (1), (5) *satisfies both boundary integral equations* (53) *and* (54). *Conversely, if a sequence* $\boldsymbol{\mu} \in (H^{-1/2}(\Gamma))^{\infty}$ *is a solution of one of the boundary integral equations* (53) *or* (54) *then it will satisfy the other one and the function built by formula* (46) *with* $\boldsymbol{\lambda} = \tilde{\boldsymbol{h}}$ *will be the generalized solution of the Dirichlet problem* (1), (5).

Proof. Since boundary integral equations (53) and (54) are only modifications of the relations (51) and (52), then the validity of the direct and the inverse statements of this theorem is granted by the theorem 6. □

Obtained sequences of BIEs have some important recurrent properties. Consider the BIEs (53). It can be reduced to a sequence of equations

$$V_0\mu_k = -\frac{1}{2}\tilde{h}_k + \sum_{i=0}^{k} K_{k-i}\tilde{h}_i - \sum_{i=0}^{k-1} V_{k-i}\mu_i \quad \text{in} \quad H^{1/2}(\Gamma), \quad k \in \mathbb{N}_0.$$

Applying the same approach for equations (54) we get the following sequences of BIEs of the second kind

$$\frac{1}{2}\mu_k + K_0'\mu_k = -\sum_{i=0}^{k} D_{k-i}\tilde{h}_i - \sum_{i=0}^{k-1} K_{k-i}'\mu_i \quad \text{in} \quad H^{-1/2}(\Gamma), \; k \in \mathbb{N}_0,$$

As we see, after the application of q-convolution to the BVPs in the operator form, all of the obtained sequences of BIEs will have the same important property. It consists in the fact that their boundary operators in the left hand sides remain the same for each $k \in \mathbb{N}_0$. Solvability of such integral equations and numerical methods for their solution are well studied in the literature. At the other point of view, the structure of the obtained BIEs allows us to build efficient algorithms for their numerical solution. The same applies for BIEs that correspond to other BVPs. Such equations are discussed in details in [13].

Thus, variational problems for infinite triangular systems, which consist of elliptic equations with variable coefficients, have been formulated and their well-posedness has been shown. By using the q-convolution of sequences, in the case of constant coefficients the representation of generalized solutions in the form of potentials has been obtained, with which variational problems have been reduced to triangular systems of BIEs. Components of the solution of the system of BIEs can consistently be found from the relevant equations which differ only in the right hand side. In this case the right hand side consists of the components of the solutions, found on previous steps, besides of the given Cauchy data. A numerical method for the solution of such systems, developed on the basis of the boundary elements method in [15], gives us ability to efficiently solve the considered boundary problems.

Acknowledgements. The author expresses his sincere gratitude to Prof. W. Wendland for the attention to this research and valuable comments, and to Prof. R. Chapko for ongoing support.

Bibliography

1. McLean W. Strongly Elliptic Systems and Boundary Integral Equations / William McLean. – Cambridge: Cambridge University Press, 2000. – 372 p.
2. Hsiao G. C. Boundary Integral Equations / George Hsiao, Wolfgang Wendland. – Berlin: Springer-Verlag, 2008. – 640 p.
3. Bamberger A. Formulation variationnelle espace-temps pour le calcul par potentiel retarde de la diffraction d'une onde acoustique (I) / A. Bamberger, T. Ha-Duong // Math. Methods Appl. Sci. – 1986. – No 8 (3). – P. 405-435.
4. Bamberger A. Formulation variationnelle pour le calcul de la diffraction d'une onde acoustique par une surface rigide / A. Bamberger, T. Ha-Duong // Math. Methods Appl. Sci. – 1986. – No 8 (4). – P. 598-608.
5. Lubich C. On the multistep time discretization of linear initial-boundary value problems and their boundary integral equations / Christian Lubich // Numer. Math. – 1994. – No 67. – P. 365-389.
6. Laliena A. R. A distributional version of Kirchhoff's formula / A. R. Laliena, F. J. Sayas // J. Math. Anal. Appl. – 2009. – No 359. – P. 197-208.
7. Litynskyy S. On week solutions of boundary value problems for an infinite system of elliptic equations / Svyatoslav Litynskyy, Yuriy Muzychuk, Anatoliy Muzychuk // Visn. Lviv. Univ. Ser. Prykl. Matem. Inform. – 2009. – No 15. – P. 52-70 (in Ukrainian).
8. Chapko R. On the numerical solution of initial boundary value problems by the Laguerre transformation and boundary integral equations / Roman Chapko, Rainer Kress // Integral and Integrodifferential Equations: Theory, Methods and Applications. Series in Mathematical Analysis and Applications. – 2000. – No 2. – P. 55-69.
9. Muzychuk Yu. A. On variational formulations of inner boundary value problems for infinite systems of elliptic equations of special kind / Yuriy Muzychuk, Roman Chapko // Matematychni Studii. – 2012. – No 38 (1). – P. 12-34.
10. Sybil Yu. Three-dimensional elliptic boundary value problems for an open Lipschitz surface / Yuriy Sybil // Matematychni Studii. – 1997. – No 8. – P. 79-96.
11. Chkadua O. Analysis of segregated boundary-domain integral equations for mixed variable-coefficient BVPs in exterior domains / O. Chkadua, S. Mikhailov, D. Natroshvili. – Eds. C. Constanda, P. Harris // Integral Methods in Science and Engineering: Computational and Analytic Aspects. – Springer (Birkhauser). – Boston. – ISBN 978-0-8176-8237-8. – 2011. – P. 109-128.
12. Dautray R. Mathematical analysis and numerical methods for science and technology. Vol. 2. Functional and Variational Methods / Robert Dautray, Jacques-Louis Lions. – Berlin: Springer, 2000. – 592 p.
13. Muzychuk Yu. A. On the boundary integral equation method for boundary-value problems for a system of elliptic equations of the special type in partially semi-infinite domains / Yuriy Muzychuk, Roman Chapko // Reports of the National Academy of Sciences of Ukraine. – 2012. – No 11. – P. 20-27 (in Ukrainian).
14. Costabel M. Boundary integral operators on Lipschitz domains: elementary results / Martin Costabel // SIAM J.Math.Anal. – 1988. – No 19. – P. 613-626.
15. Litynskyy S. Boundary elements method for some triangular system of boundary integral equations / Svyatoslav Litynskyy, Yuriy Muzychuk // Proceedings of XIV International

Seminar/Workshop on Direct and Inverse problem of Electromagnetic and Acoustic Wave Theory (DIPED-2009). – 2009. – P. 208-211.
16. Funaro D. Polynomial Approximations of Differential Equations / Daniele Funaro. – Springer-Verlag, 1992. – 305 p.

Yuriy Muzychuk,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine

Received 12.03.2014

Журнал обчислювальної та прикладної математики | 2014 № 2 (116) | Journal of Computational & Applied Mathematics

UDC 519.6

MATRIX CONTINUED FRACTIONS FOR SOLVING THE POLYNOMIAL MATRIX EQUATIONS

Anastasiya Nedashkovska

Резюме. Розглянуто алгоритм розв'язування поліноміальних матричних рівнянь. Запропонована рекурентна формула розвинення розв'язку в ланцюговий матричний дріб. Доведено збіжність методу. Наведено результати чисельних експериментів, що підтверджують справедливість теоретичних викладок..

Abstract. The article deals with the algorithm for solving the polynomial matrix equations. Recurrent formula for decomposition solution by the matrix continued fractions is proposed. The convergence of the method is proved and results of the numerical experiments that confirm the validity of the calculations are provided.

1. Introduction

The most simple matrix equations were being solved in the second half of the nineteenth century [1]. In default of a common approach polynomial matrix equations were resolved for a specific partial case.

A new approach for solving equations of the form

$$A_n X^n + A_{n-1} X^{n-1} + \ldots + A_1 X + A_0 = 0, \tag{1}$$

is proposed in this paper. Here the coefficients $A_i \in \mathbb{R}^{p\times p}$ $\left(i = \overline{1,m}\right)$ and unknowns $X \in \mathbb{R}^{p\times p}$ are set on the ring of no commutative matrices.

For example we can consider quadratic equation

$$XAX + X + B = 0, \tag{2}$$

where A and B are nonzero square matrices of order n with constant coefficients and X is unknown square matrix of order n.

The equation can be written in the form

$$(XA + E)\,X = -B.$$

Or, assuming the existence of the inverse matrix, in form $(XA + E)^{-1}$,

$$X = -\,(XA + E)^{-1}\,B.$$

For convenience here this notation will be used:

$$-\,(XA + E)^{-1}\,B = -\frac{B}{E + XA}.$$

Key words. Polynomial matrix equations; The matrix continued fractions; The convergence of the method.

Then, using the insertion method to solve equation (2), the following expansion of X into a continued fraction is written:

$$X = -\cfrac{B}{E - \cfrac{BA}{E - \cfrac{BA}{E - \cfrac{BA}{\ddots}}}} \tag{3}$$

Using the similar transformations to solve the matrix equation

$$AX + XB + XFX + C = 0 \tag{4}$$

we obtain formal expansion of X into the following continued fraction

$$X = -F^{-1}B + \cfrac{AF^{-1}B - C}{A - F^{-1}BF + \cfrac{AF^{-1}BF - CF}{-F^{-1}BF + \cfrac{AF^{-1}B - C}{\ddots \atop A - F^{-1}BF + \cfrac{AF^{-1}BF - CF}{\ddots}}}} \tag{5}$$

Or using the Prinhcheym's notation for continued fractions

$$X = -F^{-1}B + \frac{AF^{-1}B - C|}{|A - F^{-1}BF} + \frac{AF^{-1}BF - CF|}{|-F^{-1}BF} + \ldots + \frac{AF^{-1}B - C|}{|A - F^{-1}BF} + \ldots$$

It is known [1] that the problem of optimal control for discrete stationary control system is reduced to a discrete Riccati equation

$$A^T XA - X - A^T XB \left(R + B^T XB\right)^{-1} B^T XA + Q = 0. \tag{6}$$

Here matrices A with dimension $n \times n$ and B with dimension $n \times m$ describes the state of the system

$$x(k+1) = Ax(k) + Bu(k).$$

And symmetric matrices Q and R defines quality criteria

$$J = \sum_{k=0}^{\infty} \left[x^T(k) Qx(k) + u^T(k) Ru(k)\right].$$

Herewith R is positive defined and Q is positive semi defined.

It turns out that the matrix continued fractions can be used for solving the discrete Riccati equation (6). After regrouping its members obtain

$$A^T X \left(A - E - B \left(R + B^T XB\right)^{-1} B^T XA\right) + Q = 0,$$

or

$$A^T X \left(A - E - B \left(R + B^T XB\right)^{-1} B^T XBB^{-1}A\right) + Q = 0.$$

From this we obtain

$$A^T X \left[A - E - B \left(R + B^T XB\right)^{-1} \left(R + B^T XB - R\right) B^{-1}A\right] + Q = 0$$

and

$$A^T X \left[A - E - BB^{-1}A + B\left(R + B^T X B\right)^{-1} R B^{-1} A\right] + Q = 0.$$

So,

$$X = -\left(A^{-1}\right)^T Q \left[A - E - BB^{-1}A + B\left(R + B^T X B\right)^{-1} R B^{-1} A\right]^{-1}.$$

Thus, the following recurrent formula can be written for the Riccati equation:

$$X = -\frac{\left(A^{-1}\right)^T Q\Big|}{\Big|E + BB^{-1}A - A - B\dfrac{RB^{-1}A}{R + B^T X B}}. \tag{7}$$

Using composition (7) for equation (6) with numerical or symbolic elements, the following expansion of X into a continued fraction can be written:

$$\begin{aligned} X = -\frac{\left(A^{-1}\right)^T Q\Big|}{|E + BB^{-1}A - A} - B\frac{RB^{-1}A|}{|R} - B^T\frac{\left(A^{-1}\right)^T QB\Big|}{|E + BB^{-1}A - A} - \\ - \ldots - B\frac{RB^{-1}A|}{|R} - B^T\frac{\left(A^{-1}\right)^T QB\Big|}{|E + BB^{-1}A - A} - \ldots \end{aligned} \tag{8}$$

It is easy to see, comparing the expansions in continued fractions for equations (2), (4) and (6), that all of them are derived from a certain kind of schemes that does not fit into the framework of a single method. Moreover, algorithms for expansions of solutions in continued fractions are not known for algebraic numeric equations with two higher orders too.

2. The computational scheme of the method

The algorithm of expansions into the periodic branched continued fraction

$$x = p_0 + \sum_{i=1}^{n-1} \frac{p_i|}{|-q_i} + \sum_{i=1}^{n-1} \frac{p_i|}{|-q_i} + \ldots + \sum_{i=1}^{n-1} \frac{p_i|}{|-q_i} + \ldots \tag{9}$$

for polynomial numerical equations

$$x^n + a_1 x^{n-1} + a_2 x^{n-2} + \ldots + a_{n-1} x + a_n = 0 \tag{10}$$

was proposed in [2]. Unknown coefficients p_i and q_i of the fraction (10) are defined as solutions of systems of linear algebraic equations. However, this scheme cannot be trivially moved in case of solving matrix polynomial equations because non commutative multiplication of matrices. But a similar algorithm can be constructed.

Theorem 1. *A solution to equation (1) of the n th order can be represented in the form of an infinite periodic continued fraction with $(n-1)$ branches.*

Proof. Suppose that matrices $(X - Q_k)^{-1}$ $\left(k = \overline{1, n-1}\right)$ are invertibles and consider the equality

$$X = P_0 + \sum_{k=1}^{n-1} \left(X - Q_k\right)^{-1} P_k, \tag{11}$$

were $P_k \in \mathbb{R}^{p\times p}\,(k=0,1,\ldots,n-1)$ and $Q_k \in \mathbb{R}^{p\times p}\,(k=1,2,\ldots,n-1)$ are square matrices with unknown elements. To define them, the method of undetermined coefficients can be used. We will look for such items $p_{k,i,j}$ $\big(i = 1,2,\ldots,p;\ j=1,2,\ldots,p\big)$ and $q_{k,i,j}$ $\big(i=1,2,\ldots,p;\ j=1,2,\ldots,p\big)$ of matrices P_k and Q_k accordingly, that equations (1) and (11) will be equivalent.

Put additional,

$$Q_k = q_k \cdot E, \tag{12}$$

where E identity matrices and their dimensions are equal p. Easy to see that in this case

$$(X-Q_k)\,(X-Q_1)\times\ldots\times(X-Q_{k-1})\,(X-Q_{k+1})\times\ldots\times(X-Q_{n-1}) = \prod_{k=1}^{n-1}(X-Q_k)\,.$$

We reduce fractions in (11) to a common denominator and get

$$\begin{gathered} X = \left[\prod_{k=1}^{n-1}(X-Q_k)\right]^{-1}\cdot\left[\prod_{k=1}^{n-1}(X-Q_k)\,P_0 + \prod_{k=2}^{n-1}(X-Q_k)\,P_1 + \right.\\ + (X-Q_1)\prod_{k=3}^{n-1}(X-Q_k)\,P_2 + \ldots + \prod_{k=1}^{l-1}(X-Q_k)\prod_{k=l+1}^{n-1}(X-Q_k)\,P_l + \ldots + \\ \left. + \prod_{k=1}^{n-2}(X-Q_k)\,P_{n-1}\right]. \end{gathered} \tag{13}$$

Whence we obtain the following equation:

$$\begin{gathered} \left[\prod_{k=1}^{n-1}(X-Q_k)\right]X - \left[\prod_{k=1}^{n-1}(X-Q_k)\,P_0 + \prod_{k=2}^{n-1}(X-Q_k)\,P_1 + \right.\\ + (X-Q_1)\prod_{k=3}^{n-1}(X-Q_k)\,P_2 + \ldots + \prod_{k=1}^{l-1}(X-Q_k)\prod_{k=l+1}^{n-1}(X-Q_k)\,P_l + \ldots + \\ \left. + \prod_{k=1}^{n-2}(X-Q_k)\,P_{n-1}\right]. \end{gathered}$$

For each of the products we can write:

$$\begin{aligned} &-\prod_{k=1}^{n-1}(X-Q_k) = -\Big[X^n + X^{n-1}\,(-1)^{n-1}\,Q_1Q_2\ldots Q_{n-1} + \\ &+ X^{n-2}(-1)^{n-2}(Q_1Q_2\ldots Q_{n-2} + Q_1Q_2\ldots Q_{n-3}Q_{n-1} + \ldots + Q_2Q_3\ldots Q_{n-1}) + \\ &+ \ldots + X^2\,(Q_1Q_2 + Q_1Q_3 + \ldots + Q_{n-2}Q_{n-1}) - X\,(Q_1 + Q_2 + \ldots + Q_{n-1})\Big]\,; \end{aligned}$$

$$\begin{aligned} &\prod_{k=1}^{n-1}(X-Q_k)\,P_0 = X^{n-1}P_0 + X^{n-2}\,(-1)^{n-1}\,Q_1Q_2\ldots Q_{n-1}P_0 + \\ &+ X^{n-3}(-1)^{n-3}(Q_1Q_2\ldots Q_{n-2} + Q_1Q_2\ldots Q_{n-3}Q_{n-1} + \ldots + Q_2Q_3\ldots Q_{n-1})P_0 \\ &+ \ldots + X(Q_1Q_2 + Q_1Q_3 + \ldots + Q_{n-2}Q_{n-1})P_0 - (Q_1 + Q_2 + \ldots + Q_{n-1})P_0; \end{aligned}$$

$$\prod_{k=2}^{n-1}(X-Q_k)P_1=X^{n-2}P_1+X^{n-3}(-1)^{n-2}Q_2Q_3\dots Q_{n-2}P_1+X^{n-4}(-1)^{n-3}\cdot$$
$$\cdot(Q_2Q_3\dots Q_{n-2}+Q_2Q_3\dots Q_{n-3}Q_{n-1}+\dots+Q_3Q_4\dots Q_{n-1})P_1+\dots+$$
$$+(Q_2Q_3+Q_2Q_4+\dots Q_{n-2}Q_{n-1})P_1-(Q_2+Q_3+\dots+Q_{n-1})P_1;$$

$$\prod_{k=2}^{l-1}(X-Q_k)\prod_{k=l+1}^{n-1}(X-Q_k)P_l=X^{n-2}P_l+X^{n-3}(-1)^{n-2}\cdot$$
$$\cdot Q_1Q_2\dots Q_{l-1}Q_{l+1}\dots Q_{n-1}P_l+X^{n-4}(-1)^{n-3}(Q_1Q_2\dots Q_{l-1}Q_{l+1}\dots Q_{n-2}+$$
$$+Q_1Q_2\dots Q_{l-1}Q_{l+1}\dots Q_{n-3}Q_{n-1}+\dots+Q_2Q_3\dots Q_{l-1}Q_{l+1}\dots Q_{n-2}Q_{n-1})P_l$$
$$+\dots+X(Q_2Q_3+Q_2Q_4+\dots+Q_{l-1}Q_{l+1}+Q_{l-1}Q_{l+2}+\dots+Q_{n-2}Q_{n-1})P_l-$$
$$-(Q_1+Q_2+\dots+Q_{l-1}+Q_{l+1}+\dots+Q_{n-1})P_l;$$

$$\prod_{k=1}^{n-2}(X-Q_k)P_{n-1}=X^{n-2}P_{n-1}+X^{n-3}(-1)^{n-2}Q_1Q_2\dots Q_{n-2}Q_{n-1}P_{n-1}+$$
$$+X^{n-4}(-1)^{n-3}(Q_1Q_2\dots Q_{n-3}+Q_1Q_2\dots Q_{n-4}Q_{n-2}+\dots+Q_2Q_3\dots Q_{n-2})P_{n-1}$$
$$+\dots+X(Q_1Q_2+Q_2Q_3+\dots+Q_{n-3}Q_{n-2})P_{n-1}-(Q_1+Q_2+\dots+Q_{n-2})P_{n-1}.$$

We now sum up the right sides of the equalities above, with simultaneously grouping the coefficients of identical powers of X. Equating coefficients of identical powers of X, we obtain the following system of equations for the determination of $P_k\,(k=0,1,2,\dots,n-1)$ and $Q_k\,(k=1,2,\dots,n-1)$:

$$\begin{aligned}
&(-1)^{n-1}Q_1Q_2\dots Q_{n-1}+P_0=A_1;\\
&(-1)^{n-2}\prod_{k=1}^{n-1}\prod_{l=1}^{k-1}Q_l\prod_{l=k+1}^{n-1}Q_l-\sum_{k=1}^{n-1}P_k+\\
&\qquad+(-1)^{n-1}Q_1Q_2\dots Q_{n-1}P_0=A_2;\\
&(-1)^{n-3}\sum_{k=1}^{n-2}\sum_{l=k+1}^{n-2}(1-\delta_{kl})\prod_{r=1}^{k-1}Q_r\prod_{r=k+1}^{l-1}Q_r\prod_{r=l+1}^{n-2}Q_r+\\
&\qquad+\sum_{k=1}^{n-1}\prod_{r=1}^{k-1}Q_r\prod_{r=k+1}^{n-1}Q_rP_k+(-1)^{n-1}\sum_{k=1}^{n-1}\prod_{r=1}^{k-1}Q_rP_0=A_3;\\
&\dots\\
&\sum_{k=1}^{n-1}Q_k+\sum_{k=2}^{n-1}\sum_{l=k+1}^{n-1}Q_kQ_lP_1+\dots+\sum_{k=1}^{n-1}\sum_{l=k+1}^{n-1}(1-\delta_{kr})Q_kQ_lP_r+\\
&+\dots+\sum_{k=1}^{n-2}\sum_{l=k+1}^{n-2}Q_kQ_lP_{n-1}=A_{n-1};\\
&\dots\\
&\sum_{k=1}^{n-1}Q_kP_1+\dots+\sum_{k=1}^{n-1}(1-\delta_{kr})Q_rP_r+\dots+\sum_{k=1}^{n-2}Q_kP_{n-1}+\\
&\qquad+\sum_{k=1}^{n-1}Q_kP_0=A_n,
\end{aligned}\tag{14}$$

where $\delta_{kl}=\begin{cases}1 \text{ if } k=l,\\ 0 \text{ if } k\neq l.\end{cases}$

If all the chosen Q_k are pairwise different, then the latter system of n equations in n unknowns (14) will become linear relatively unknown P_k $(k = 0, 1, 2, \ \ldots, n-1)$ and will have a unique solution. Using composition law (11) for X, we obtain the following expansion in terms of matrix branched continued fractions. If the left matrix multiplication $(X - Q_k)^{-1} P_r$ is denoted as $\dfrac{P_r}{X - Q_k}$, the recurrent formula for X will look as such:

$$X = P_0 + \sum_{k=1}^{n-1} \frac{P_r}{X - Q_k}. \tag{15}$$

Applying now the composition (14), we obtain the expanse of matrix branched continued fraction

$$X = P_0 + \sum_{k_1=1}^{n-1} \cfrac{P_{k_1}}{P_0 - Q_{k_1} + \sum\limits_{k_2=1}^{n-1} \cfrac{P_{k_2}}{P_0 - Q_{k_2} + \sum\limits_{k_3=1}^{n-1} \cfrac{}{\ddots + \cfrac{}{\sum\limits_{k_m=1}^{n-1} \cfrac{P_{k_m}}{P_0 - Q_{k_m} + \cfrac{}{\ddots}}}}}} \tag{16}$$

which is what had to be proved. □

To calculate the solution on the computer systems the recurrent formula (15) is sufficient. But for analytical writing solution and research of its existence and convergence approaching fractions shall use the theory of branched continued fraction for expanse (16). But solving equations (1) and (2), (4) and (6) requires a detailed study of convergence and computational stability of the matrix branched continued fraction.

Some sufficient signs of convergence for matrix branched continued fractions have been proposed in [3].

But the convergence of the branched fraction does not necessarily mean the convergence to the solution of the corresponding equation (1), (3) or (5). So we will focus on this aspect in more detail and consider the branched continued fraction

$$\sum_{k_l=1}^{N} \frac{a_{k_l}|}{|b_{k_l}} + \sum_{k_2=l}^{N} \frac{a_{k_1k_2}|}{|b_{k_1k_2}} + \sum_{k_3=1}^{N} \frac{a_{k_1k_2k_3}|}{|b_{k_1k_2k_3}} + \ldots + \sum_{k_i=1}^{N} \frac{a_{k_1k_2k_3\ldots k_i}|}{|b_{k_1k_2k_3\ldots k_i}} + \ldots \tag{17}$$

Here $a_{k_1k_2k_3\ldots k_i}$ and $b_{k_1k_2k_3\ldots k_i}$ are square matrices of dimension $p \times p$. In [2] and [3] the following sufficient signs have been obtained.

Theorem 2. *If the solution of polynomial matrix equation exists and belongs to the interval* $[-N, N]$*, then the expansion by some iterative procedure into the matrix branched continued fraction (17) with elements that satisfy the conditions*

$$\left\| b_{k(s)}^{-1} \right\| \leq \frac{1}{\left\| a_{k(s)+N} \right\|} \ \ (k(s) \in [1, N]; s = 1, 2, 3, \ldots)$$

converges to this solution.

Theorem 3. *If the solution of polynomial matrix equation exists and belongs to the interval* $\left[-\sum\limits_{k(s)=1}^{N}\left\|a_{k(s)}\right\|, \sum\limits_{k(s)=1}^{N}\left\|a_{k(s)}\right\|\right]$, *then the expansion by some iterative procedure into the matrix branched continued fraction (17) with elements that satisfy the conditions*

$$\left\|b_{k(s)}^{-1}\right\| \leq \frac{1}{1+\sum\limits_{k(s+1)=1}^{N}\left\|a_{k(s+1)}\right\|} \quad (s=1,2,3,\ldots)$$

converges to this solution.

These signs can be used to analyze the convergence of matrix continued fractions (3), (5), (8) and (16). Also, they are simple and easy to use. The theorems 2 and 3 can be used in practice, particularly in computer algebra systems, and serve as a basis for other sufficient signs for matrix branched continued fraction.

Note also, that if signs of convergence are valid, the iterative process (16) can finish if the inequality

$$\|X_{k+1}-X_k\| \leq \epsilon$$

is valid. Here ϵ – given calculation accuracy. This follows from the fact that in conditions of the theorem 2 and the theorem 3 the absolutely convergent numerical majorizing branched fractions build for matrix branched continued fractions (16). And its approach fractions form a monotone sequence.

Estimate the complexity of the algorithm. To obtain $P_k\ \left(k=\overline{0,n-1}\right)$ and $Q_k\ \left(k=\overline{1,n}\right)$ for the system of equations (14) we need to specify the pairwise different values for all matrix elements of Q_k. Then, doing generally up to the principal term n^5p^3 operations of multiplication and n^5p^3 operations of addition, we obtain the block system of linear algebraic equations with order n to determine P_k. For its solution need to complete an additional n^3p^3 operations of multiplication and n^3p^3 operations of addition. One iteration using the recurrent formula (11) requires the implementation of $2np^3$ operations of multiplication and np^3 operations of addition.

3. Numerical experiments

To verify the practical effectiveness of this approach, a series of numerical experiments were done in Mat Lab environment. In particular matrix equation

$$X^3+A_2X^2+A_1X+A_0=0,$$

was being solved. Here matrix coefficients were equal

$$A_2=\begin{pmatrix} 2.0000 & -3.0000 & -5.0000\\ 0.2200 & 0.2510 & 0.2500\\ 0.2200 & -0.2340 & -0.1300 \end{pmatrix};\ A_1=\begin{pmatrix} 1.0000 & 6.0000 & -5.0000\\ 0.2500 & 0.2200 & 0.2510\\ 0.2340 & -0.1300 & 0.2200 \end{pmatrix};$$

$$A_0=\begin{pmatrix} 136.0000 & 139.0000 & 134.0000\\ -272.0240 & -269.0270 & -282.0490\\ -350.2980 & -358.7900 & -336.5740 \end{pmatrix}.$$

The recurrent formula $X = P_0 + (Q_1 + X)^{-1} P_1 + (Q_2 + X)^{-1} P_2$ was being used to calculate X.

The matrix coefficients were set as

$$Q_1 = \begin{pmatrix} 0.96 & 0 & 0 \\ 0 & 0.96 & 0 \\ 0 & 0 & 0.96 \end{pmatrix}; \ Q_1 = \begin{pmatrix} 1.92 & 0 & 0 \\ 0 & 1.92 & 0 \\ 0 & 0 & 1.92 \end{pmatrix}.$$

Then from equations (15) the following values were calculated

$$P_2 = \begin{pmatrix} 139.9739 & 121.2717 & 130.3833 \\ -285.0969 & -287.0854 & -293.3430 \\ -364.5170 & -374.3781 & -358.9099 \end{pmatrix};$$

$$P_1 = \begin{pmatrix} -141.6651 & -135.9117 & -139.7833 \\ 285.4805 & 281.1371 & 293.8120 \\ 364.9166 & 373.8342 & 351.8643 \end{pmatrix}; P_0 = \begin{pmatrix} 0.8800 & 3.0000 & 5.0000 \\ -0.2200 & 2.6290 & -0.2500 \\ -0.2200 & 0.2340 & 3.0100 \end{pmatrix};$$

For the initial approximation X_0 was chosen zero matrix and the following approximate value of the unknown matrix was received

$$X = \begin{pmatrix} 12.3600 & 147.9411 & -107.2121 \\ -28.9221 & -290.3746 & 224.4685 \\ -36.9221 & -363.6585 & 282.0369 \end{pmatrix}$$

with the following results

Number of iteration	30	40	50	60	70
Norm of difference	0.3015	$6.1725E-04$	$9.0636E-06$	$4.9470E-08$	$6.9768E-09$

Thus, this approach can be applied to solve scientific and technical problems in generalized models of V. Leontyev and so on. However, the task of building a more subtle signs of convergence for periodic matrix branched continued fractions with broader areas of convergence is still open.

Bibliography

1. Ikramov H. D. Numeric solution of matrix equations / H. D. Ikramov. – Oxford: Pergamon Press, 1984. – 192 p.
2. Nedashkovskyy M. O. Solving of non-linear polynomial equation by branching chain fractions / M. O. Nedashkovskyy // "Computing" International Scientific Journal. – ISSN 1727-6209. – 2003. – Vol. 2. – Issue 1.
3. Nedashkovskyy M. O. The convergence of branched continued fractions to solutions of polynomial matrix equations / M. O. Nedashkovskyy // VII th International Conference INTERPOR 2008. – Porous Materials. – Theory and Experiment. – Bydgoszcz. Lubostroń. – 20-22, October 2008. – Kazimierz Wielki Bydgoszcz University. – 2008. – P. 63-64.

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

Received 21.05.2014

Журнал обчислювальної та прикладної математики | **2014 № 2 (116)** | Journal of Computational & Applied Mathematics

UDC 519.6

A GRADIENT METHOD OF SOLVING INVERSE EIGENVALUE PROBLEM

Bohdan Podlevskyi, Oksana Yaroshko

Резюме. Розглядається обернена задача на власні значення, яка включає класичні адитивні та мультиплікативні спектральні задачі. Показано спосіб зведення оберненої задачі до багатопараметричної задачі на власні значення. Запропоновано чисельний метод відшукання наближеного розв'язку спектральної задачі шляхом розв'язання еквівалентної їй варіаційної задачі. Проведено числові експерименти для ілюстрації роботи методу.

Abstract. It is investigated the inverse eigenvalue problem that includes classic additive and multiplicative spectral problems. It is presented the method of transformation of the inverse eigenvalue problem to the direct multi-parameter one. It is proposed the numerical method of calculating the approximate solution of the spectral problem by solving the equivalent variation problem. There are several numerical experiments presented in order to illustrate the behavior of the method.

1. Introduction

The problem of reconstruction of the matrix of some given structure based on the given spectral data is well known as the inverse eigenvalue problem, or in other words, the inverse spectral problem.

Such problems arise in a wide area of analysis investigations and mathematical physics, namely in the systems of control and identifications, the structural analysis, the modeling of mechanical systems and so on.

The major common point of all these applications is the fact that the physical parameters of some system should be restored based on the given dynamical parameters of the same system. If we describe the physical parameters mathematically and present them in a form of a matrix, we get an inverse eigenvalue problem.

As it was mentioned above, the needed matrix should have some given structure. Such structural constraints are not unsubstantial – they add sense to the spectral problem. Beside that, they define the different types of inverse spectral problem: additive, multiplicative, multi-parameter, structural etc.

There are two main questions regarding the eigenvalue problem: theoretical one, concerning the existence of the solution, and practical one, about the numerical method of finding this solution. There is provided a lot of literature concerning the conditions of solubility and uniqueness of the solution for different types of inverse spectral problem. A variety of methods of calculating the approximate solution of the mentioned problem is also listed in different

Key words. Eigenvalue problem, inverse problem, variation problem, functional, iterative procedure.

sources. See, for example, [1] - [11] and the cited literature). In this article we will discuss another numerical algorithm of solving the inverse eigenvalue problem, assuming that the solution exists.

2. Inverse Eigenvalue Problem

Let's consider the following inverse spectral problem.

Problem GIEP (General Inverse Eignevalue Problem).

Provided it is given the complex matrices of dimensions $n \times n$*:* $A_0, A_1, \ldots, A_m \in C^{n\times n}$ *and the collection of numbers* $\lambda = \{\lambda_1, ..., \lambda_m\} \in C^m$*.*

Find such parameters $p = \{p_1, ..., p_m\} \in C^m$ *that the eigenvalues of the matrix*

$$A(p) = A_0 + p_1 A_1 + ... + p_m A_m \tag{1}$$

coincide with the given set of numbers $\lambda = \{\lambda_1, ..., \lambda_m\} \in C^m$*.*

This problem involves classic partial cases of *additive* and *multiplicative* inverse spectral problems:

Problem AIEP (Additive Inverse Eigenvalue Problem).

Let A *be a given matrix and* $\lambda = \{\lambda_1, ..., \lambda_m\} \in C^m$ *be a given set of numbers.*

Find the diagonal matrix $D = \operatorname{diag}(p_1, \ldots, p_m)$*,* $p_1, p_2, \ldots, p_m \in C^m$*, such that the matrix* $A + D$ *has the eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_m$*.*

Problem MIEP (Multiplicative Inverse Eigenvalue Problem).

Let A *be a given matrix and* $\lambda = \{\lambda_1, ..., \lambda_m\} \in C^m$ *be a given set of numbers.*

Find the diagonal matrix $D = \operatorname{diag}(p_1, \ldots, p_m)$*,* $p_1, p_2, \ldots, p_m \in C^m$*, such that the matrix* AD *has the eigenvalues* $\lambda_1, \lambda_2, \ldots, \lambda_m$*.*

The question of solvability of such kind of problems, namely AIEP, is widely explored in the literature (see, for example, [2], [4], [11]). Beside the theoretical results there is a lot of numerical methods constructed for solving the additive inverse eigenvalue problem (see, for example, [1], [3], [5], [7] - [9]).

In this survey we propose another method of finding the approximate solution of the problem (2.1) in the real Euclidian space. This method is based on a gradient procedure.

3. Preliminary

Consider the multi-parameter spectral problem in the Euclidian space E^n:

$$T(\lambda)\, x \equiv Ax - \lambda_1 B_1 x - ... - \lambda_m B_m x = 0 \tag{2}$$

where $\lambda = \{\lambda_1, ..., \lambda_m\} \in E^m$ are spectral parameters, $x = (x_1, ..., x_n) \in E^n$, $A, B_1, ..., B_m$ are some linear operators that act in the real Euclidian space E^n.

The multi-parameter eigenvalue problem, linear towards the spectral parameters, consists in finding a vector of spectral parameters $\lambda = \{\lambda_1, \ ... \ , \lambda_m\} \in E^m$ such that there exists a non-trivial solution $x \in E^n \backslash \{0\}$ of the equation (3.1).

Let's put the variation problem of minimization of the following functional in correspondence to the spectral problem (3.1):

$$F(u) = \frac{1}{2} \left\| T(\lambda)\, x \right\|_H^2, \ \forall u = \{x, \lambda\} \in H = (E^n \backslash \{0\}) \oplus E^m \tag{3}$$

The problem of minimization of the functional (3.2) consists in finding such set of parameters $\lambda = \{\lambda_1, ..., \lambda_m\} \in E^m$ and corresponding vector $x \in E^n \backslash \{0\}$ that the functional $F(u)$ reaches its minimum value:

$$F(u) \to \min_u, \quad u \in U \subset H, \tag{4}$$

where U is the set with points $u = \{x, \lambda\}$ that satisfy the equation (3.1), H is an Euclidian space with the scalar product and the norm defined in a standard way:

$$(u, v)_H = (u_1, u_2)_{E^n} + (v_1, v_2)_{E^m}, \|u\|_H = \sqrt{\|u_1\|^2_{E^n} + \|v_1\|^2_{E^m}},$$

$$u = \{u_1, v_1\}, v = \{u_2, v_2\}, u_1, u_2 \in E^n, v_1, v_2 \in E^m.$$

In the article [8] it is shown that the spectral problem (3.1) and the variation problem (3.3) are equivalent. This means that each eigen pair $\{x, \lambda\}$ of the problem (3.1) is the point of minimum $u = \{x, \lambda\}$ of the functional (3.2), and vice-versa.

This result lets us construct the gradient procedure of numerical solving of the problem (3.3) and thus, of the problem (3.1), in the following form:

$$u_{k+1} = u_k - \gamma(u_k) \nabla F(u_k), \quad k = 0,\ 1,\ 2,\ ... \tag{5}$$

The relation (3.4) describes the whole class of methods that differ only by the choice of the step value $\gamma(u_k)$.

In this article we will calculate the value $\gamma_k = \gamma(u_k)$ at each step of the process by using the formula:

$$\gamma_k = \frac{F(u_k)}{||\nabla F(u_k)||^2_H} \tag{6}$$

From this point here, in order to make the formulas more easy to read, we will omit the index H in the denotation of the scalar product and the norm.

So, the iteration process can be written as following:

$$u_{k+1} = u_k - \frac{F(u_k)}{||\nabla F(u_k)||^2} \nabla F(u_k), \tag{7}$$

where the gradient of the functional has the structure

$$\nabla F(u) = \left\{ (T^*Tx, e_1), ..., (T^*Tx, e_n), \left(Tx, \frac{\partial T}{\partial \lambda_1} x \right), ..., \left(Tx, \frac{\partial T}{\partial \lambda_m} x \right) \right\} \tag{8}$$

Here $T \equiv T(\lambda)$, and $e_i \in E^n$ is the vector, the i-th co-ordinate of which is equal to 1 and all the others co-ordinates are 0.

If the starting approximation is chosen in some sense close enough to the eigenvector and the vector of eigenvalues, then the iteration process (3.6) converges to the stationary point of the functional (3.2) $u^* = \{x^*, \lambda^*\}$. In this point the minimum of the functional is reached. Note, this means that the process converges to the eigenvector x^* and the vector of eigenvalues λ^* of the problem (3.1).

Thus, for the iteration process, described above, the following theorem is true:

Theorem 1. [8] *Let the gradient of the functional (3.6) satisfies the Lipchitz condition*

$$\|\nabla F(u) - \nabla F(z)\| \leq L\,\|u - z\|\,, \quad \forall u, z \in U, \;\; L > 0 \tag{9}$$

where U is a closed convex set that contains the solution u^*. If for some starting approximation $u_0 = (x_0, \lambda^{(0)}) \in U$ the following condition is true

$$0 < \gamma_0 \equiv \gamma(u_0) \leq 1/2L, \tag{10}$$

then the iteration process (3.6) converges to the point of minimum of the functional (3.2) $u^* = \{x^*, \lambda^*\}$ and, thus, to the eigenvector x^* and the vector of eigenvalues λ^* of the problem (3.1). Which means that the relations below are true:

$$\lim_{k\to\infty} \rho(u_k, U_*) = \lim_{k\to\infty} \rho(u_k, u^*) = 0 \tag{11}$$

$$\lim_{k\to\infty} F(u_k) = F(u^*) = 0 \tag{12}$$

4. Algorithm of solving an inverse spectral problem

Consider an inverse eigenvalue problem of type (2.1) with the real matrices $A_0, A_1, ..., A_m \in E^{n\times n}$, and where the pairs $\left\{\lambda_k, x^k\right\}_{k=1}^m$ are the eigen pairs of the matrix $A(p)$. Here $\lambda = \{\lambda_1, ..., \lambda_m\} \in E^m$, $x^k \in H = E^n\backslash\{0\}$, $k = 1, 2, ..., m$, E is the real Euclidian space.

By using the definition of an eigen value and a corresponding eigen vector, we can write as following:

$$A\,(c)\,x^k = \lambda_k x^k, \quad x^k \in H, \quad k = 1, ..., m$$

Thus, we get the system of m equations to find the parameters $p_1, ..., p_m$:

$$\left\{ \begin{array}{c} ((A_0 - \lambda_1 I) + p_1 A_1 + ... + p_m A_m)\,x^1 = 0, \\ ... \\ ((A_0 - \lambda_m I) + p_1 A_1 + ... + p_m A_m)\,x^m = 0, \end{array} \right. \tag{13}$$

Let's transform this system so that it has the structure (3.1). For this reason consider the matrix operators $\mathbf{A}$, $\mathbf{B}_i\ :\ \mathbf{H} \to \mathbf{H}$, $\mathbf{H} = \bigoplus_{k=1}^{m} E^{n\times n}$, $i = 1, ..., m$:

$$\mathbf{A} = \begin{pmatrix} (A_0 - \lambda_1 I) & & 0 \\ & \ddots & \\ 0 & & (A_0 - \lambda_m I) \end{pmatrix} \tag{14}$$

$$\mathbf{B}_i = \begin{pmatrix} -A_i & & 0 \\ & \ddots & \\ 0 & & -A_i \end{pmatrix} \tag{15}$$

In case $\mathbf{x} = (x^1,\ x^2,\ \dots\ ,\ x^m)^T \in \mathbf{H}$, we get

$$\mathbf{A}\mathbf{x} = ((A_0 - \lambda_1 I)x^1,\ (A_0 - \lambda_2 I)x^2,\ \dots\ ,\ (A_0 - \lambda_m I)x^m),$$

$$\mathbf{B}_i\mathbf{x} = (-A_i x^1,\ -A_i x^2,\ \dots\ ,\ -A_i x^m).$$

Now we can proceed from the problem (4.1) to the problem of type (3.1) in the space $\mathbf{H}$.

$$T(p) \equiv \mathbf{A}\mathbf{x} - p_1\mathbf{B}_1\mathbf{x} - \dots - p_m\mathbf{B}_m\mathbf{x} = 0 \qquad (16)$$

So, we configured the problem of finding such set of parameters $p_1, \dots, p_m$, that the equation (4.4) has a non-trivial solution $\mathbf{x} \in \mathbf{H}\backslash\{\mathbf{0}\}$.

Now let's put a variation problem in correspondence to the problem (4.4):

$$F(\mathbf{u}) \to \min_{\mathbf{u}}, \quad \mathbf{u} \in \mathbf{U} \subset \mathbf{H},$$

where

$$F(\mathbf{u}) = \frac{1}{2}\left\|T(p)\,\mathbf{x}\right\|^2_{\tilde{\mathbf{H}}},\ \forall \mathbf{u} = \{\mathbf{x}, p\} \in \tilde{\mathbf{H}} = \mathbf{H} \oplus E^m \qquad (17)$$

The task is to find the set of parameters $p = \{p_1, \dots, p_m\} \in E^m$ and the corresponding vector $\mathbf{x} \in \mathbf{H}\backslash\{\mathbf{0}\}$, for which the functional $F(\mathbf{u})$ reaches its minimum value. For this variation problem we will apply the iteration process (3.6).

So, the algorithm consists of the following steps:

Step 1. Select the starting approximation.

Step 2. Build the matrices A, B_i, $i = 1, \dots, m$ by using the formulas (4.2), (4.3).

Step 3. **for** $k = 0, 1, 2, \dots$ until the exactness is reached **do**:

Step 4. Calculate $T(p_k)$ by using the formula (4.4).

Step 5. Calculate $F(u_k)$ by using the formula (4.5).

Step 6. Calculate $\nabla F(u_k)$ by using the formula (3.7).

Step 7. Calculate the next approximation u_{k+1} of the solution by using the formula (3.6).

end *for k*

Step 8. Extract the p_{k+1} components of the vector $u_{k+1} = \{x_{k+1}, p_{k+1}\}$ which is the needed approximate solution.

Step 9. End.

5. Numerical experiments

Let's demonstrate how the algorithm works on two examples below.

Example 1. [7]. Consider the following inverse eigenvalue problem:

$$A_0 = \begin{pmatrix} 1.5 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, A_1 = \begin{pmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, A_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$A_3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, A_4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix},$$

$$A(p) = A_0 + p_1 A_1 + p_2 A_2 + p_3 A_3 + p_4 A_4.$$

Where $n = m = 4$, the eigen values are given: $\lambda = \{0, 2, 2, 4\}$. The exact solution of the problem is also known $p^* = \{1, 1, 1, 1\}$.

Let's choose the starting approximation of the parametrs $p_i, i = 1, 2, 3, 4$.

a) $p^{(0)} = \{1.1, 0.9, 1.1, 0.9\}$, as it is proposed in [7], which is quite close to the exact solution;

b) $p^{(0)} = \{0.1, 0.2, 0.3, 0.4\}$, which strongly differs from the exact solution.

Note, that in [7] the starting approximation is given only for the parameters. However, to apply the method proposed in this article we also need to select the approximation of the eigen vectors $x^{(0)}$. In order to choose the correct values $x^{(0)}$, we did the following calculations: for the given parameters $p^{(0)} = \left\{p_1^{(0)}, ..., p_m^{(0)}\right\}$ we built the matrix $A\left(p^{(0)}\right) = A_0 + p_1^{(0)} A_1 + ... + p_m^{(0)} A_m$ and found the eigen values and the corresponding eigen vectors of this matrix by using the software application Matlab. Then we accepted these eigen vectors as the starting approximation $x^{(0)}$ for the method (3.6).

There were used two different stop conditions for the iteration process:

1. The value of the functional becomes zero, which means that $F\left(u^{(k+1)}\right) < \varepsilon$, $\varepsilon = 10^{-9}$, where $u^{(k+1)} = \left\{x^{(k+1)}, p^{(k+1)}\right\}$ is the k-th approximation of the solution of the problem, $k = 0, 1, ...$

2. The norm of deviation between the values of the parameters on two iterations $p^{(k)}$ and $p^{(k+1)}$ becomes sufficiently small: $\left\|p^{(k)} - p^{(k+1)}\right\| < \varepsilon$, $\varepsilon = 10^{-9}$, $k = 0, 1, ...$

The results received in the cases a) and b) of starting approximations are presented in the Table 1 and the Table 2 respectively. Note, that each table contains two approximate solutions that correspond to two stop conditions of the iteration process.

TABL. 1. Approximate solutions of Example 1, case a

	p^*	$p^{(0)}$	$p^{(m+1)}$, Stop cond 1	$p^{(m+1)}$, Stop cond 2
	1	1.1	1.0000403787	1.0000000123
	1	0.9	1.0000199351	1.0000000061
	1	1.1	1.0000266736	1.0000000081
	1	0.9	0.9999747545	0.9999999923
F	9.37e-31	0.02	8.4065983153e-11	9.2005057095e-18
$\|p - p^*\|$	0	0.2	5.8109138139e-5	1.7748612738e-8

In the tables it is also given the value of the functional in the point of starting approximation, $F^0 = F\left(u^{(0)}\right)$, the point of approximate solution, $F = F(u)$, TABL. 2. Approximate solutions of Example 1, case b

	p^*	$p^{(0)}$	$p^{(m+1)}$, Stop cond 1	$p^{(m+1)}$, Stop cond 2
	1	0.1	0.9999582549	0.9999999786
	1	0.2	0.9999782404	0.9999999888
	1	0.3	0.9999714938	0.9999999854
	1	0.4	1.0000265629	1.0000000135
F	9.37e-31	1.89	7.1596842832e-11	1.9824859739e-17
$\|p-p^*\|$	0	1.52	6.1109094783e-5	3.1269595091e-8

and the point of exact solution, $F^* = F(u^*)$. In this way it can be seen that the value of the functional decreases, as it was expected.

Example 2. Consider the given inverse spectral problem, where the matrices A_i are the Toeplitz matrices. Note, that similarly to the previous example here $n = m$.

$$A(p) = A_0 + p_1 A_1 + ... + p_n A_n$$

$$A_0 = O, A_1 = I,$$

$$A_2 = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \ddots & \vdots \\ 0 & 1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 & 1 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}, ..., A_n = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & \ddots & \ddots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \ddots & \ddots & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

Let's solve this problem for $n = 5$. In this case the exact values of the parameters are $p^* = \{-1.8,\ 1.9,\ 2.5, 0.08,\ 1.2\}$.

The chosen starting approximations of the parameters $p_i, i = 1, ..., 5$ are the following: a) $p^{(0)} = \{-1,\ 1,\ 1,\ -1,\ 1\}$; b) $p^{(0)} = \{0,\ 1,\ 1,\ 0,\ 1\}$.

In order to select the starting approximations of the eigen vectors $x^{(0)}$ we did the same calculations as it was explained in the Example 1. The received results are presented in the Table 3 and the Table 4 for two cases of starting approximations.

Similarly to the Example 1, there were used two conditions to stop the iteration process. By analyzing the received values of the functional in the points of starting approximation, approximate solution and exact solution it can be seen that they go down to the minimum (zero) value, as expected.

Let us also note, that the analyzed examples had been solved by using two variants of the Newton method presented in the articles [1] and [3]. This experiment showed that the iteration processes of the Newton methods [1] and [3] do not converge to the exact solution if the selected starting approximation strongly differs from the exact values. The method presented in this survey, on the contrary, does converge to the exact solution in case of the same starting approximations.

Tabl. 3. Approximate solutions of Example 2, case a

	p^*	$p^{(0)}$	$p^{(m+1)}$, Stop cond 1	$p^{(m+1)}$, Stop cond 2
	-1.8	-1	-1.8000000001	-1.8000000001
	1.9	1	1.9000000048	1.9000000184
	2.5	1	2.4999999976	2.4999999865
	0.08	-1	0.0800000037	0.0799999959
	1.2	1	1.1999999837	1.1999999674
F	1.94e-29	12.62	3.7755871274e-10	1.1228316897e-9
$\|p-p^*\|$	0	2.22	1.7159808703e-8	3.9948275161e-8

Tabl. 4. Approximate solutions of Example 2, case b

	p^*	$p^{(0)}$	$p^{(m+1)}$, Stop cond 1	$p^{(m+1)}$, Stop cond 2
	-1.8	0	-1.8000048843	-1.8000048843
	1.9	1	1.9000646818	1.9000646818
	2.5	1	2.4999533405	2.4999533404
	0.08	0	0.0799945432	0.0799945431
	1.2	1	1.1998789745	1.1998789744
F	1.94e-29	17.56	3.5156195443e-6	3.5156195443e-6
$\|p-p^*\|$	0	2.52	1.4512640324e-4	1.4512640324e-4

Bibliography

1. Biegler-Konig F. W. A Newton iteration process for inverse eigenvalue problems / F. W. Biegler-Konig // Numer. Math. – 1981. – Vol. 37, No 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. Podlevskyi B. M. Newton's method for the solution of inverse spectral problems / B. M. Podlevskyi, O. S. Yaroshko // J. Mathematical Sciences. – 2013. – Vol. 174, No 3. – P. 156-165.
4. Friedland S. Inverse eigenvalue problems / S. Friedland // Linear Algebra Appl. – 1977. – Vol. 17, No 1. – P. 15–51.
5. Friedland S. The formulation and analysis of numerical methods for inverse eigenvalue problems / S. Friedland, J. Nocedal, M. L. Overton // SIAM J. Numer. Anal. – 1987. – Vol. 24, No 3. – P. 634–667.
6. Graham M. L. Inverse Problems in Vibration / M. L. Graham, M. T. Gladwell. – New York: Kluwer Academic Publishers. – 2005.
7. Dai H. A Numerical Method for Solving Inverse Eigenvalue Problems / H. Dai // Mathematical Modelling and Numerical Analysis M2AN. – 1999. – Vol. 33, No 5. – P. 1003-1017.
8. Khlobystov V. V. A gradient descent method for solving of one class of nonlinear multiparameter eigenvalue problem / V. V. Khlobystov, B. M. Podlevskyi, O. S. Yaroshko // Trudy Instituta Matematiki Akad. Nayk Belarusi. – 2014 (in Russian).
9. Kublanovskaya V. N. On one approach to solving an inverse eigenvalue problem / V. N. Kublanovskaya // Zapiski Nauch. Semin. Leningr. Otdel. Mat. Inst. Akad. Nauk SSSR. – 1970. – Vol. 18. – P. 138–149 (in Russian).

10. Li L. Sufficient conditions for the solvability of algebraic inverse eigenvalue problems / L. Li // Linear Algebra Appl. – 1995. – Vol. 221. – P. 117–129.
11. Chu M. T. Inverse Eigenvalue Problems / Moody T. Chu // SIAM Rev. – 1998. – Vol. 40. – P. 1-39.

Bohdan Podlevskyi,
Pidstrygach Institute for Applide Problems
of Mechanics and Mathematics,
3b Naukova Str., Lviv, 79060, Ukraine;

Bohdan Podlevskyi, Oksana Yaroshko,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine

Received 18.05.2014

Журнал обчислювальної та прикладної математики | 2014 № 2 (116) | Journal of Computational & Applied Mathematics

UDC 519.6

TWO-STEP COMBINED METHOD FOR SOLVING NONLINEAR OPERATOR EQUATIONS

Stepan Shakhno, Halyna Yarmola

Резюме. У статті вивчено напівлокальну збіжність двокрокового комбінованого методу для розв'язування нелінійних операторних рівнянь, побудованого на базі двох методів з порядками збіжності $1+\sqrt{2}$. Аналіз збіжності проведено за узагальнених умов Ліпшиця для перших і других похідних та поділених різниць першого порядку.

Abstract. In this paper we study a semilocal convergence of the two-step combined method for solving nonlinear operator equations. It method is based on two methods of convergence orders $1+\sqrt{2}$. Convergence analysis is provided for generalized Lipschits condition for Frechet derivates of the first and second orders and for divided differences of the first order.

1. Introduction

Consider the equation

$$H(x) \equiv F(x) + G(x) = 0, \tag{1}$$

where F and G are nonlinear operators, defined on a convex subset D of a Banach space X with values in a Banach space Y. F is a Fréchet-differentiable operator, G is a continuous operator, differentiability of which is not required.

The well-known Newton's method cannot be applied, as differentiability of operator H is required. For solving nonlinear equation (1) very often use the two-point iterative process [1]

$$x_{n+1} = x_n - A_n^{-1}(F(x_n) + G(x_n)), \quad n = 0, 1, \ldots, \tag{2}$$

where $A_n = A(x_{n-1}, x_n) \in L(X, Y)$. The convergence analysis of the method (2) in general and for $A_n = F'(x_n)$, $A_n = F'(x_n) + G(x_{n-1}; x_n)$, $A_n = H(x_{n-1}; x_n)$ and its modifications was provided by authors [1, 2, 3, 4, 5, 6, 18]. Here $G(x; y)$ ($H(x; y)$) is a first order divided difference of the operator $G(H)$ at the points x and y [13, 14, 15]. In papers [7, 11] we researched a semilocal convergence of the method (2) for $A_n = F'(x_n) + G(x_{n-1}; x_n)$ and $A_n = F'(x_n) + G(2x_n - x_{n-1}; x_{n-1})$.

In works [10, 12] we proposed a two-step method that is based on the methods with the convergence orders $1+\sqrt{2}$ [9, 17]. Its iterative formula is:

$$\begin{aligned} x_{n+1} &= x_n - \Big[F'\Big(\frac{x_n + y_n}{2}\Big) + G(x_n; y_n)\Big]^{-1} H(x_n), \\ y_{n+1} &= x_{n+1} - \Big[F'\Big(\frac{x_n + y_n}{2}\Big) + G(x_n; y_n)\Big]^{-1} H(x_{n+1}), \quad n = 0, 1, \ldots. \end{aligned} \tag{3}$$

Key words. Generalized Lipschitz condition, nondifferentiable operator, semilocal convergence.

We provided a local and a semilocal convergence analysis for method (3) under classical Lipschitz conditions for the first and second order derivatives and divided differences of the first order and established the convergence order. Also we showed results of the numerical solving of the nonlinear equations and systems of nonlinear equations by this iterative process. In paper [8] we proved the local convergence theorem of the (3) under generalized Lipschitz conditions.

In this paper, we study the semilocal convergence of the method (3) under generalized Lipschitz conditions for the first and second order derivatives and divided differences of the first order. These conditions are more general and include classical Lipschitz conditions. Therefore our results have the theoretical interest.

2. Preliminaries

We will need the following definition and lemmas [8, 16].

Definition 7. Let G 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 $G(x;y)$, which satisfies the condition

$$G(x;y)(x-y) = G(x) - G(y)$$

is called a divided difference of the first order of G at the points x and y.

In the study of iterative methods very often use the Lipschitz conditions with constant L. Parameter L under Lipschitz conditions does not necessarily has to be a constant, but may also be a positive integrable function. In work [16] Wang suggested generalized Lipschitz conditions for the derivative operator in which instead of constant there was used a certain positive integrable function. In the work [9] we introduce analogous generalized Lipschitz conditions for the divided difference of the first order operator.

Let us denote as $U_0 = \{x : \ \|x - x_0\| \le r_0\}$ a closed ball of radius r_0 with center at the point x_0. If L in Lipschitz conditions is a positive integrable function, we consider the conditions

$$\|F'(x) - F'(y)\| \le \int_0^{\|x-y\|} L(u)du, \ x, y \in U_0 \tag{4}$$

and

$$\|G(x;y) - G(u;v)\| \le \int_0^{\|x-u\|+\|y-v\|} M(z)dz, \ x, y, u, v \in U_0, \tag{5}$$

where L and M are positive integrable functions. Lipschitz conditions (4) and (5) we will call generalized Lipschitz conditions or Lipschitz conditions with the L (or M) average. Note that in the case of constants L and M we obtain from (4) and (5) the classical Lipschitz conditions.

Lemma 1. [16]. *Let* $h(t) = \dfrac{1}{t}\displaystyle\int_0^t L(u)du, \ 0 \le t \le r,$ *where* $L(u)$ *is a positive integrable function that is nondecreasing monotonically in* $[0, r]$*. Then* $h(t)$ *is nondecreasing monotonically with respect to* t*.*

Lemma 2. [8]. *Let* $g(t) = \dfrac{1}{t^3}\displaystyle\int_0^t N(u)(t-u)^2 du,\ 0 \le t \le r$, *where* $N(u)$ *is a positive integrable function that is nondecreasing monotonically in* $[0, r]$. *Then* $g(t)$ *is a nondecreasing monotonically with respect to* t.

3. Semilocal convergence analysis of the two-step iterative process (3)

We can show the following semilocal convergence theorem for the method (3). Imposed terms guarantee the convergence of the iterative process (3) to the solution x^* and its uniqueness.

Theorem 1. *Let* F *and* G *be nonlinear operators, defined on an open convex subset* D *of a Banach space* X *with values in a Banach space* Y. F *is a Fréchet-differentiable operator,* G *is a continuous operator, differentiability of which is not required. Assume that the linear operator* $A_0 = F'\Big(\dfrac{x_0+y_0}{2}\Big) + G(x_0; y_0)$, *where* $x_0,\ y_0 \in D$, *is invertible and in* $U_0 = \{x:\ \|x - x_0\| \le r_0\} \subset D$ *the Lipschitz conditions are fulfilled*

$$\|A_0^{-1}(F'(x) - F'(y))\| \le \int_0^{\|x-y\|} L(z)dz, \qquad (6)$$

$$\|A_0^{-1}(F''(x)h - F''(y)h)\| \le \|h\| \int_0^{\|x-y\|} N(z)dz,\ h \in X, \qquad (7)$$

$$\|A_0^{-1}(G(x;y) - G(u;v))\| \le \int_0^{\|x-u\|+\|y-v\|} M(z)dz, \qquad (8)$$

where L, M, *and* N *are positive integrable and nondecreasing monotonically functions.*

Let a, c $(c > a)$, r_0 *be nonnegative numbers such that*

$$\|x_0 - y_0\| \le a, \quad \|A_0^{-1}(F(x_0) + G(x_0))\| \le c, \qquad (9)$$

$$r_0 \ge \frac{c}{1-\gamma}, \quad \int_0^{(2r_0-a)/2} L(z)dz + \int_0^{2r_0-a} M(z)dz < 1, \qquad (10)$$

$$\gamma = \frac{\frac{1}{8}c\int_0^c N(z)\Big(1-\frac{z}{c}\Big)^2 dz + \int_0^{(c-a)/2} L(z)dz + \int_0^{c-a} M(z)dz}{1 - \int_0^{(2r_0-a)/2} L(z)dz - \int_0^{2r_0-a} M(z)dz}, \quad 0 \le \gamma < 1.$$

Then the iterative process (3) is well-defined and sequences $\{x_n\}_{n\ge0}$, $\{y_n\}_{n\ge0}$ *generated by it remain in* U_0 *and converge to the solution* x^* *of equation (1) and, for all* $n \ge 0$, *the following inequalities are satisfied*

$$\|x_n - x_{n+1}\| \le t_n - t_{n+1}, \quad \|y_n - x_{n+1}\| \le s_n - t_{n+1}, \qquad (11)$$

$$\|x_n - x^*\| \le t_n - t^*, \quad \|y_n - x^*\| \le s_n - t^*, \qquad (12)$$

where sequences $\{t_n\}_{n\ge0}$ *and* $\{s_n\}_{n\ge0}$ *defined by the formulas*

$$t_0 = r_0, \quad s_0 = r_0 - a, \quad t_1 = r_0 - c,$$

$$t_{n+1}-t_{n+2}=$$
$$=\frac{1}{8c^3}\frac{\int_0^c N(z)(c-z)^2dz(t_n-t_{n+1})^3}{1-\int_0^{(t_0-t_{n+1}+s_0-s_{n+1})/2}L(z)dz-\int_0^{t_0-t_{n+1}+s_0-s_{n+1}}M(z)dz}+$$
$$+\frac{1}{c-a}\frac{\Big[\int_0^{(c-a)/2}L(z)dz+\int_0^{c-a}M(z)dz\Big](t_n-t_{n+1})(s_n-t_{n+1})}{1-\int_0^{(t_0-t_{n+1}+s_0-s_{n+1})/2}L(z)dz-\int_0^{t_0-t_{n+1}+s_0-s_{n+1}}M(z)dz},\quad n\geq 0, \tag{13}$$

$$t_{n+1}-s_{n+1}=$$
$$=\frac{1}{8c^3}\frac{\int_0^c N(z)(c-z)^2dz(t_n-t_{n+1})^3}{1-\int_0^{(t_0-t_n+s_0-s_n)/2}L(z)dz-\int_0^{t_0-t_n+s_0-s_n}M(z)dz}+$$
$$+\frac{1}{c-a}\frac{\Big[\int_0^{(c-a)/2}L(z)dz+\int_0^{c-a}M(z)dz\Big](t_n-t_{n+1})(s_n-t_{n+1})}{1-\int_0^{(t_0-t_n+s_0-s_n)/2}L(z)dz-\int_0^{t_0-t_n+s_0-s_n}M(z)dz},\quad n\geq 0 \tag{14}$$

are nonincreasing nonnegative and converge to certain t^* such that

$$r_0-\frac{c}{1-\gamma}\leq t^*<t_0.$$

Proof. Let us show by the mathematical induction method that, for all $k\geq 0$

$$t_{k+1}\geq s_{k+1}\geq t_{k+2}\geq r_0-\frac{c}{1-\gamma}\geq 0, \tag{15}$$

$$t_{k+1}-t_{k+2}\leq\gamma(t_k-t_{k+1}),\quad t_{k+1}-s_{k+1}\leq\gamma(t_k-t_{k+1}) \tag{16}$$

are satisfied. For $k=0$, from (13) and (14), we get

$$t_1-t_2=\frac{1}{c^3}\frac{\int_0^c N(z)(c-z)^2dz(t_0-t_1)^3}{1-\int_0^{(t_0-t_1+s_0-s_1)/2}L(z)dz-\int_0^{t_0-t_1+s_0-s_1}M(z)dz}+$$
$$+\frac{1}{c-a}\frac{\Big[\int_0^{(c-a)/2}L(z)dz+\int_0^{c-a}M(z)dz\Big](t_0-t_1)(s_0-t_1)}{1-\int_0^{(t_0-t_1+s_0-s_1)/2}L(z)dz-\int_0^{t_0-t_1+s_0-s_1}M(z)dz}$$

and

$$t_2=r_0-c-\Big[\frac{\frac{1}{8}c\int_0^c N(z)(1-\frac{z}{c})^2dz+\int_0^{(c-a)/2}L(z)dz+\int_0^{c-a}M(z)dz}{1-\int_0^{(2r_0-a)/2}L(z)dz-\int_0^{2r_0-a}M(z)dz}\Big]c\geq$$
$$\geq r_0-(1+\gamma)c=r_0-\frac{(1-\gamma^2)c}{1-\gamma}\geq r_0-\frac{c}{1-\gamma}\geq 0.$$

Similarly, we have

$$t_1-s_1=\frac{1}{8c^3}\int_0^c N(z)(c-z)^2dz(t_0-t_1)^3+$$
$$+\frac{1}{c-a}\Big[\int_0^{(c-a)/2}L(z)dz+\int_0^{c-a}M(z)dz\Big](t_0-t_1)(s_0-t_1)$$

and

$$s_1 = r_0 - c - \Big[\frac{1}{8}c\int_0^c N(z)\Big(1-\frac{z}{c}\Big)^2 dz + \int_0^{(c-a)/2} L(z)dz + \int_0^{c-a} M(z)dz\Big]c.$$

From the last equalities it follows that

$$t_1 \geq t_2,\ s_1 \geq t_2,\ t_1 \geq s_1 \geq t_2 \geq r_0 - \frac{c}{1-\gamma} \geq 0.$$

Assume that that inequalities (15) and (16) are satisfied for $k = \overline{0, n-1}$. Then, for $k = n$, we obtain

$$\begin{gathered} t_{n+1} - t_{n+2} = \\ \frac{1}{8c^3}\frac{\int_0^c N(z)(c-z)^2 dz (t_n - t_{n+1})^3}{1 - \int_0^{(t_0 - t_{n+1} + s_0 - s_{n+1})/2} L(z)dz - \int_0^{t_0 - t_{n+1} + s_0 - s_{n+1}} M(z)dz} + \\ + \frac{1}{c-a}\frac{\Big[\int_0^{(c-a)/2} L(z)dz + \int_0^{c-a} M(z)dz\Big](t_n - t_{n+1})(s_n - t_{n+1})}{1 - \int_0^{(t_0 - t_{n+1} + s_0 - s_{n+1})/2} L(z)dz - \int_0^{t_0 - t_{n+1} + s_0 - s_{n+1}} M(z)dz} \leq \\ \leq \frac{\frac{1}{8}c\int_0^c N(z)(1-\frac{z}{c})^2 dz + \int_0^{(c-a)/2} L(z)dz + \int_0^{c-a} M(z)dz}{1 - \int_0^{(2r_0 - a)/2} L(z)dz - \int_0^{2r_0 - a} M(z)dz}(t_n - t_{n+1}) = \\ = \gamma(t_n - t_{n+1}), \\ t_{n+1} - s_{n+1} = \frac{1}{8c^3}\frac{\int_0^c N(z)(c-z)^2 dz (t_n - t_{n+1})^3}{1 - \int_0^{(t_0 - t_n + s_0 - s_n)/2} L(z)dz - \int_0^{t_0 - t_n + s_0 - s_n} M(z)dz} + \\ + \frac{1}{c-a}\frac{\Big[\int_0^{(c-a)/2} L(z)dz + \int_0^{c-a} M(z)dz\Big](t_n - t_{n+1})(s_n - t_{n+1})}{1 - \int_0^{(t_0 - t_n + s_0 - s_n)/2} L(z)dz - \int_0^{t_0 - t_n + s_0 - s_n} M(z)dz} \leq \\ \leq \frac{\frac{1}{8}c\int_0^c N(z)(1-\frac{z}{c})^2 dz + \int_0^{(c-a)/2} L(z)dz + \int_0^{c-a} M(z)dz}{1 - \int_0^{(2r_0 - a)/2} L(z)dz - \int_0^{2r_0 - a} M(z)dz}(t_n - t_{n+1}) = \\ = \gamma(t_n - t_{n+1}) \end{gathered}$$

and

$$\begin{gathered} t_{n+1} \geq s_{n+1} \geq t_{n+2} \geq t_{n+1} - \gamma(t_n - t_{n+1}) \geq \\ \geq r_0 - \frac{1-\gamma^{n+2}}{1-\gamma}c \geq r_0 - \frac{c}{1-\gamma} \geq 0. \end{gathered}$$

So, we prove, that $\{t_n\}_{n\geq 0}$ and $\{s_n\}_{n\geq 0}$ are nonincreasing, nonnegative sequences and converge to $t^* \geq 0$.

Let us prove, by mathematical induction, that the iterative process (3) is well-defined and inequalities (11) are satisfied for all $n \geq 0$.

Taking into account (9) and that $t_0 - t_1 = c$, we establish that $x_1 \in U_0$ and (11) are satisfied for $n = 0$.

Denote $A_n = F'\Big(\frac{x_n + y_n}{2}\Big) + G(x_n; y_n)$. Using the Lipschitz conditions (6) and (8), we have

$$\|I - A_0^{-1} A_{n+1}\| = \|A_0^{-1}[A_0 - A_{n+1}]\| \leq$$
$$\leq \Big\|A_0^{-1}[F'\Big(\frac{x_0 + y_0}{2}\Big) - F'\Big(\frac{x_{n+1} + y_{n+1}}{2}\Big) + G(x_0; y_0) - G(x_{n+1}; y_{n+1})]\Big\| \leq$$
$$\leq \int_0^{(\|x_0 - x_{n+1}\| + \|y_0 - y_{n+1}\|)/2} L(z)dz + \int_0^{\|x_0 - x_{n+1}\| + \|y_0 - y_{n+1}\|} M(z)dz \leq$$
$$\leq \int_0^{(t_0 - t_{n+1} + s_0 - s_{n+1})/2} L(z)dz + \int_0^{t_0 - t_{n+1} + s_0 - s_{n+1}} M(z)dz \leq$$
$$\leq \int_0^{(t_0 + s_0)/2} L(z)dz + \int_0^{t_0 + s_0} M(z)dz < 1.$$

According to the Banach lemma on the invertible operator, A_{n+1} is invertible and

$$\|A_{n+1}^{-1} A_0\| \leq$$
$$\leq \Big(1 - \int_0^{(\|x_0 - x_{n+1}\| + \|y_0 - y_{n+1}\|)/2} L(z)dz - \int_0^{\|x_0 - x_{n+1}\| + \|y_0 - y_{n+1}\|} M(z)dz\Big)^{-1}.$$

Let us prove that iterative process (3) is well-defined for $k = n + 1$. Taking into account the definition of the first order divided difference, conditions (6), (8) and identity [17]

$$F(x) - F(y) - F'\Big(\frac{x + y}{2}\Big)(x - y) = \frac{1}{4} \int_0^1 (1 - t)\Big[F''\Big(\frac{x + y}{2} + \frac{t}{2}(x - y)\Big) -$$
$$- F''\Big(\frac{x + y}{2} + \frac{t}{2}(y - x)\Big)\Big] dt (x - y)(x - y),$$

we obtain

$$\|A_0^{-1} H(x_{n+1})\| =$$
$$= \|A_0^{-1}\Big[F(x_{n+1}) - F(x_n) - F'\Big(\frac{x_n + x_{n+1}}{2}\Big)(x_{n+1} - x_n) +$$
$$+ F'\Big(\frac{x_n + x_{n+1}}{2}\Big)(x_{n+1} - x_n) - F'\Big(\frac{x_n + y_n}{2}\Big)(x_{n+1} - x_n) +$$
$$+ G(x_{n+1}) - G(x_n) - G(x_n; y_n)(x_{n+1} - x_n)]\,\| \leq$$
$$\leq \frac{1}{8} \int_0^{\|x_n - x_{n+1}\|} N(z)(\|x_n - x_{n+1}\| - z)^2 dz +$$
$$+ \int_0^{\|y_n - x_{n+1}\|/2} L(z)dz \|x_n - x_{n+1}\| +$$
$$+ \int_0^{\|y_n - x_{n+1}\|} M(z)dz \|x_n - x_{n+1}\|.$$

Denote

$$A_n = \frac{1}{8}\int_0^{\|x_n - x_{n+1}\|} N(z)(\|x_n - x_{n+1}\| - z)^2 dz,$$

$$B_n = \int_0^{\|y_n - x_{n+1}\|/2} L(z)dz,\ C_n = \int_0^{\|y_n - x_{n+1}\|} M(z)dz,$$

$$Q_{n+1} = 1 - \int_0^{(\|x_0 - x_{n+1}\| + \|y_0 - y_{n+1}\|)/2} L(z)dz - \int_0^{\|x_0 - x_{n+1}\| + \|y_0 - y_{n+1}\|} M(z)dz.$$

Hence, taking into account lemmas 1, 2 and inequalities (11), we have

$$\|x_{n+1} - x_{n+2}\| = \|A_{n+1}^{-1} H(x_{n+1})\| \le \|A_{n+1}^{-1} A_0\| \|A_0^{-1} H(x_{n+1})\| \le$$
$$\le \frac{A_n + [B_n + C_n]\|x_n - x_{n+1}\|}{Q_{n+1}} =$$
$$= \frac{A_n \|x_n - x_{n+1}\|^3}{Q_{n+1}\|x_n - x_{n+1}\|^3} + \frac{[B_n + C_n]\|x_n - x_{n+1}\|\|y_n - x_{n+1}\|}{Q_{n+1}\|y_n - x_{n+1}\|} \le$$

$$\le \frac{A_0 \|x_n - x_{n+1}\|^3}{Q_{n+1}\|x_0 - x_1\|^3} + \frac{[B_0 + C_0]\|x_n - x_{n+1}\|\|y_n - x_{n+1}\|}{Q_{n+1}\|y_0 - x_1\|} \le$$
$$\le \frac{1}{8(t_0 - t_1)^3} \frac{\int_0^{t_0 - t_1} N(z)(t_0 - t_1 - z)^2 dz (t_n - t_{n+1})^3}{1 - \int_0^{(t_0 - t_{n+1} + s_0 - s_{n+1})/2} L(z)dz - \int_0^{t_0 - t_{n+1} + s_0 - s_{n+1}} M(z)dz} +$$
$$+ \frac{1}{s_0 - t_1} \frac{\left[\int_0^{(s_0 - t_1)/2} L(z)dz + \int_0^{s_0 - t_1} L(z)dz\right](t_n - t_{n+1})(s_n - t_{n+1})}{1 - \int_0^{(t_0 - t_{n+1} + s_0 - s_{n+1})/2} L(z)dz - \int_0^{t_0 - t_{n+1} + s_0 - s_{n+1}} M(z)dz} =$$
$$= \frac{1}{8c^3} \frac{\int_0^{c} N(z)(c - z)^2 dz (t_n - t_{n+1})^3}{1 - \int_0^{(t_0 - t_{n+1} + s_0 - s_{n+1})/2} L(z)dz - \int_0^{t_0 - t_{n+1} + s_0 - s_{n+1}} M(z)dz} +$$
$$+ \frac{1}{c - a} \frac{\left[\int_0^{(c - a)/2} L(z)dz + \int_0^{c - a} M(z)dz\right](t_n - t_{n+1})(s_n - t_{n+1})}{1 - \int_0^{(t_0 - t_{n+1} + s_0 - s_{n+1})/2} L(z)dz - \int_0^{t_0 - t_{n+1} + s_0 - s_{n+1}} M(z)dz} =$$
$$= t_{n+1} - t_{n+2}$$

and

$$\|x_{n+2} - y_{n+2}\| = \|A_{n+1}^{-1} H(x_{n+2})\| \le \|A_{n+1}^{-1} A_0\| \|A_0^{-1} H(x_{n+2})\| \le$$
$$\le \frac{A_{n+1} + [B_{n+1} + C_{n+1}]\|x_n - x_{n+1}\|}{Q_{n+1}} =$$
$$= \frac{A_{n+1}\|x_{n+1} - x_{n+2}\|^3}{Q_{n+1}\|x_n - x_{n+1}\|^3} + \frac{[B_{n+1} + C_{n+1}]\|x_{n+1} - x_{n+2}\|\|y_{n+1} - x_{n+2}\|}{Q_{n+1}\|y_{n+1} - x_{n+2}\|} \le$$
$$\le \frac{A_0\|x_{n+1} - x_{n+2}\|^3}{Q_{n+1}\|x_0 - x_1\|^3} + \frac{[B_0 + C_0]\|x_{n+1} - x_{n+2}\|\|y_{n+1} - x_{n+2}\|}{Q_{n+1}\|y_0 - x_1\|} \le$$
$$\le \frac{1}{8(t_0 - t_1)^3} \frac{\int_0^{t_0 - t_1} N(z)((t_0 - t_1) - z)^2 dz (t_{n+1} - t_{n+2})^3}{1 - \int_0^{(t_0 - t_{n+1} + s_0 - s_{n+1})/2} L(z)dz - \int_0^{t_0 - t_{n+1} + s_0 - s_{n+1}} M(z)dz} +$$

$$+\frac{1}{s_0-t_1}\frac{\left[\int_0^{(s_0-t_1)/2}L(z)dz+\int_0^{s_0-t_1}M(z)dz\right](t_{n+1}-t_{n+2})(t_{n+2}-s_{n+1})}{1-\int_0^{(t_0-t_{n+1}+s_0-s_{n+1})/2}L(z)dz-\int_0^{t_0-t_{n+1}+s_0-s_{n+1}}M(z)dz}=$$

$$=\frac{1}{8c^3}\frac{\int_0^c N(z)(c-z)^2dz(t_{n+1}-t_{n+2})^3}{1-\int_0^{(t_0-t_{n+1}+s_0-s_{n+1})/2}L(z)dz-\int_0^{t_0-t_{n+1}+s_0-s_{n+1}}M(z)dz}+$$

$$+\frac{1}{c-a}\frac{\left[\int_0^{(c-a)/2}L(z)dz+\int_0^{c-a}M(z)dz\right](t_{n+1}-t_{n+2})(s_{n+1}-t_{n+2})}{1-\int_0^{(t_0-t_{n+1}+s_0-s_{n+1})/2}L(z)dz-\int_0^{t_0-t_{n+1}+s_0-s_{n+1}}M(z)dz}=$$

$$=t_{n+2}-s_{n+2}.$$

Thus, the iterative process (3) is well-defined for all $n\geq 0$. Hence it follows that

$$\|x_n-x_k\|\leq t_n-t_k,\ \|y_n-x_k\|\leq s_n-t_k, \|y_n-y_k\|\leq s_n-s_k,\quad 0\leq n\leq k, \tag{17}$$

i.e., the sequence $\{x_n\}_{n\geq 0}$ and $\{y_n\}_{n\geq 0}$ are fundamental in a Banach space X and convergence to x^*. From (17) for $k\to\infty$ it follows inequalities (12). Let us show that x^* is the solution of the equation (1). Indeed,

$$\|A_0^{-1}H(x_{n+1})\|\leq\frac{1}{8}\int_0^{\|x_n-x_{n+1}\|}N(z)(\|x_n-x_{n+1}\|-z)^2dz+$$

$$+\int_0^{\|y_n-x_{n+1}\|/2}L(z)dz\|x_n-x_{n+1}\|+\int_0^{\|y_n-x_{n+1}\|}M(z)dz\|x_n-x_{n+1}\|\leq$$

$$\leq\frac{1}{24}N(\|x_n-x_{n+1}\|)\|x_n-x_{n+1}\|^3+\int_0^{\|y_n-x_{n+1}\|/2}L(z)dz\|x_n-x_{n+1}\|+$$

$$+\int_0^{\|y_n-x_{n+1}\|}M(z)dz\|x_n-x_{n+1}\|\to 0,\quad \text{when } n\to\infty.$$

Thus, $H(x^*)=0$. The theorem is proven. □

Theorem 2. *Let F and G be nonlinear operators, defined on an open convex subset D of a Banach space X with values in a Banach space Y. F is a Fréchet-differentiable operator, G is a continuous operator, differentiability of which is not required. Assume that:*
1) conditions of Theorem 1 are satisfied;
2) r_0 from Theorem 1 additionally satisfies condition

$$\gamma_1=\frac{\frac{1}{8}r_0\int_0^{r_0}N(z)\left(1-\frac{z}{r_0}\right)^2dz+\int_0^{(r_0-a)/2}L(z)dz+\int_0^{r_0-a}M(z)dz}{1-\int_0^{(2r_0-a)/2}L(z)dz-\int_0^{2r_0-a}M(z)dz}<1. \tag{18}$$

Then the iterative process (3) is well-defined and generated by it $\{x_n\}_{n\geq 0}$ belongs to U_0 and converges to the unique solution x^* of the equation $F(x)=0$ in U_0.

Proof. To show the uniqueness, we assume that there exists a second solution x^{**}.

Using the approximation

$$x_{n+1}-x^{**}=x_n-x^{**}-A_n^{-1}[H(x_n)-H(x^{**})]=$$
$$=A_n^{-1}\big[F'\Big(\frac{x_n+y_n}{2}\Big)(x_n-x^{**})-F(x_n)+F(x^{**})\big]+$$
$$+A_n^{-1}\big[G(x_n;y_n)-G(x_n;x^{**})\big](x_n-x^{**}),$$

we obtain

$$\|x_{n+1}-x^{**}\|\le\Big\|A_n^{-1}\Big[F'\Big(\frac{x_n+y_n}{2}\Big)(x_n-x^{**})-F(x_n)+F(x^{**})\Big]\Big\|+$$
$$+\|A_n^{-1}\big[G(x_n;y_n)-G(x_n;x^{**})\big](x_n-x^{**})\|\le$$
$$\le\Big\|A_n^{-1}\Big[F'\Big(\frac{x_n+x^{**}}{2}\Big)(x_n-x^{**})-F(x_n)+F(x^{**})\Big]\Big\|+$$
$$+\Big\|A_n^{-1}\Big[F'\Big(\frac{x_n+y_n}{2}\Big)-F'\Big(\frac{x_n+x^{**}}{2}\Big)\Big](x_n-x^{**})\Big\|+$$
$$+\|A_n^{-1}\big[G(x_n;y_n)-G(x_n;x^{**})\big](x_n-x^{**})\|\le$$
$$\le\|A_n^{-1}A_0\|\Big\|A_0^{-1}\Big[F(x_n)-F(x^{**})-F'\Big(\frac{x_n+x^{**}}{2}\Big)(x_n-x^{**})\Big]\Big\|+$$
$$+\|A_n^{-1}A_0\|\Big\|A_0^{-1}\Big[F'\Big(\frac{x_n+y_n}{2}\Big)-F'\Big(\frac{x_n+x^{**}}{2}\Big)\Big]\Big\|\|x_n-x^{**}\|+$$
$$+\|A_n^{-1}A_0\|\|A_0^{-1}\big[G(x_n;y_n)-G(x_n;x^{**})\big]\|\|x_n-x^{**}\|\le$$
$$\le\frac{1}{4}\frac{\int_0^1(1-t)\int_0^{t\|x_n-x^{**}\|}N(z)dzdt}{Q_n}\|x_n-x^{**}\|^2+$$
$$+\frac{\int_0^{\|y_n-x^{**}\|/2}L(z)dz}{Q_n}\|x_n-x^{**}\|+\frac{\int_0^{\|y_n-x^{**}\|}M(z)dz}{Q_n}\|x_n-x^{**}\|=$$
$$=\frac{\frac{1}{4}\int_0^{\|x_n-x^{**}\|}N(z)\int_{z/\|x_n-x^{**}\|}^1(1-t)dzdt\|x_n-x^{**}\|^2}{Q_n}+$$
$$+\frac{\int_0^{\|y_n-x^{**}\|/2}L(z)dz+\int_0^{\|y_n-x^{**}\|}M(z)dz}{Q_n}\|x_n-x^{**}\|\le$$
$$=\frac{\frac{1}{8}\int_0^{\|x_n-x^{**}\|}N(z)\big(1-\frac{z}{\|x_n-x^{**}\|}\big)^2dz\|x_n-x^{**}\|^2}{Q_n}+$$
$$+\frac{\int_0^{\|y_n-x^{**}\|/2}L(z)dz+\int_0^{\|y_n-x^{**}\|}M(z)dz}{Q_n}\|x_n-x^{**}\|\le$$
$$\le\gamma_1\|x_n-x^{**}\|\le\ldots\le\gamma_1^{n+1}\|x_0-x^{**}\|,$$

which implies $x^{**}=\lim_{n\to\infty}x_n=x^*$. The theorem is proven. □

Let $L(z)=L=\text{const}$, $N(z)=N=\text{const}$ and $M(z)=M=\text{const}$. Then we get the following result.

Theorem 3. *Let F and G be nonlinear operators, defined on an open convex subset D of a Banach space X with values in a Banach space Y. F is a Fréchet-differentiable operator, G is a continuous operator, differentiability of which is* *not required. Assume that the linear operator* $A_0 = F'\Big(\dfrac{x_0+y_0}{2}\Big) + G(x_0;y_0)$, *where* $x_0,\ y_0 \in D$, *is invertible and in* $U_0 = \{x:\ \|x-x_0\| \le r_0\} \subset D$ *the Lipschitz conditions are fulfilled*

$$\|A_0^{-1}(F'(x)-F'(y))\| \le L\|x-y\|,$$
$$\|A_0^{-1}(F''(x)h-F''(y)h)\| \le N\|x-y\|\|h\|,\ h\in X,$$
$$\|A_0^{-1}(G(x;y)-G(u;v))\| \le M(\|x-u\|+\|y-v\|),$$

where L, M *and* N *are positive numbers.*

Let a, c $(c>a)$, r_0 *be nonnegative numbers such that*

$$\|x_0-y_0\| \le a, \quad \|A_0^{-1}(F(x_0)+G(x_0))\| \le c,$$
$$r_0 \ge \frac{c}{1-\gamma}, \quad (L/2+M)(2r_0-a)<1,$$
$$\gamma = \frac{c^2N/24+(L/2+M)(c-a)}{1-(L/2+M)(2r_0-a)}, \qquad 0\le\gamma<1.$$

Then the iterative process (3) is well-defined and sequences $\{x_n\}_{n\ge0}$, $\{y_n\}_{n\ge0}$ *generated by it remain in* U_0 *and converge to the solution* x^* *of equation (1) and, for all* $n\ge0$, *the following inequalities are satisfied*

$$\|x_n-x_{n+1}\| \le t_n-t_{n+1}, \quad \|y_n-x_{n+1}\| \le s_n-t_{n+1},$$
$$\|x_n-x^*\| \le t_n-t^*, \quad \|y_n-x^*\| \le s_n-t^*,$$

where sequences $\{t_n\}_{n\ge0}$ *and* $\{s_n\}_{n\ge0}$ *defined by the formulas*

$$t_0=r_0, \quad s_0=r_0-a, \quad t_1=r_0-c,$$
$$t_{n+1}-t_{n+2}=$$
$$\frac{N(t_n-t_{n+1})^3/24+(L/2+M)(t_n-t_{n+1})(s_n-t_{n+1})}{1-(L/2+M)(t_0-t_{n+1}+s_0-s_{n+1})}, \quad n\ge0, \tag{19}$$
$$t_{n+1}-s_{n+1}=$$
$$\frac{N(t_n-t_{n+1})^3/24+(L/2+M)(t_n-t_{n+1})(s_n-t_{n+1})}{1-(L/2+M)(t_0-t_n+s_0-s_n)}, \quad n\ge0$$

are nonincreasing nonnegative and converge to certain t^* *such that* $r_0-\dfrac{c}{1-\gamma} \le t^* < t_0$.

Remark 1. *If* $F(x)=0$, $L=0$ *and* $N=0$ *then the sequences* $\{t_n\}_{n\ge0}$ *and* $\{s_n\}_{n\ge0}$, *defined by the formulas (19), reduce to similar ones in* [9] *for the case* $\alpha=1$.

Bibliography

1. Argyros I. K. A unifying local-semilocal convergence analysis and applications for two-point Newton-like methods in Banach space / I. K. Argyros // J. Math. Anal. Appl.– 2004.– Vol. 298.– P. 374-397.
2. 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. 6.– P. 21-28.

3. Argyros I. K. On the convergence of modified Newton methods for solving equations containing a non-differentiable term / I. K. Argyros, H. Ren // J. Comp. App. Math.– 2009.– Vol. 231.– P. 897-906.
4. Hernandez M. A. A uniparametric family of iterative processes for solving nondifferentiable operators / M. A. Hernandez, M. J. Rubio // J. Math. Anal. Appl.– 2002.– Vol. 275.– P. 821-834.
5. Hernandez M. A. The Secant method for nondifferentiable operators / M. A. Hernandez, M. J. Rubio // Appl. Math. Lett.– 2002.– Vol. 15.– P. 395-399.
6. Ren H. A new semilocal convergence theorem for a fast iterative method with nondifferentiable operators / H. Ren, I. K. Argyros // J. Appl. Math. Comp.– 2010.– Vol. 34, No 1-2.– P. 39-46.
7. Shakhno S. M. Convergence analysis of combined method for solving nonlinear equations / S. M. Shakhno, I. V. Melnyk, H. P. Yarmola // Math. methods and physicomechanical fields.– 2013.– Vol. 56, No. 1.– P.31-39 (in Ukrainian).
8. Shakhno S. M. Convergence of the two-step combined method and uniqueness of the solution of nonlinear operator equations / S. M. Shakhno // J. Comp. App. Math.– 2014.– Vol. 261.– P. 378-386.
9. Shakhno S. M. On an iterative algorithm with superquadratic convergence for solving nonlinear operator equations / S. M. Shakhno // J. Comp. App. Math.– 2009.– Vol. 231.– P. 222-235.
10. Shakhno S. M. On the two-step method for solving nonlinear equations with nondifferentiable operator / S. M. Shakhno, H. P. Yarmola // PAMM – Proc. Appl. Math. Mech.– 2012.– Vol. 12.– P. 617-618.– doi 10.1002/pamm.201210297.
11. Shakhno S.M. Two-point method for solving nonlinear equations with nondifferentiable operator / S. M. Shakhno, H. P. Yarmola // Matematychni Studii.– 2010.– Vol. 36, No 2.– P. 213-220 (in Ukrainian).
12. Shakhno S. M. Two-step method for solving nonlinear equation with nondifferentiable operator / S. M. Shakhno, H. P. Yarmola // J. Comput. Appl. Math.– 2012.– Vol. 109, No 3.– P. 105-115.
13. Ulm S. Extension of Steffensen's method for solving nonlinear operator equations / S. Ulm // Comput. Math. and Math. Phys. – 1964.– Vol. 4, No 6.– P. 1093-1097 (in Russian).
14. Ulm S. On construction of generalized divided differences / S. Yu. Ulm, V. V. Pol // Izv. Akad. Nauk Est. SSR. Ser. Fiz. Mat.– 1969.– Vol. 18, No 1.– P. 100-102 (in Russian).
15. Ulm S. On generalized divided differences / S. Ulm // Izv. Akad. Nauk Est. SSR. Ser. Fiz. Mat.– 1967.–Vol. 16.– P. 13-26, P. 146-156 (in Russian).
16. Wang X. Convergence of Newton's method and uniqueness of the solution of equations in Banach space / X. Wang // IMA J. Numer. Anal.– 2000.– Vol. 20.– P. 123-134.
17. Werner W. Über ein Verfarhren der Ordnung $1 + \sqrt{2}$ zur Nullstellenbestimmung / W. Werner // Numer. Math.– 1979.– Vol. 32.– P. 333-342.
18. Zabrejko P.P. The majorant method in the theory of Newton-Kantorovich approximations and the Pta'k error estimates / P.P. Zabrejko, D.F. Nguen // Numerical Functional Analysis and Optimization.– 1987.– Vol. 9, No 5-6.– P. 671-684.

Stepan Shakhno, Halyna Yarmola,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine

Received 23.04.2014

Журнал обчислювальної та прикладної математики | **2014** **№ 2 (116)** | Journal of Computational & Applied Mathematics

UDC 519.6

NUMERICAL ANALYSIS OF THE GIRKMANN PROBLEM WITH FEM/BEM COUPLING USING DOMAIN DECOMPOSITION

Andriy Styahar

Резюме. Ми розглядаємо поєднану модель для задачі Гіркмана. Ця задача полягає в обчисленні плоского деформованого стану для тіла, що складається з основної частини та тонкої частини, що прикріплена до основної частини. Для побудови наближеного розв'язку цієї задачі ми використовуємо метод граничних елементів (МГЕ) та метод скінченних елементів (МСЕ), поєднані за допомогою алгоритму декомпозиції областей. Наведено результати числових експериментів. Порівняно напружено-деформований стан конструкцій для різних форм оболонок.

Abstract. We consider a coupled model for the Girkmann problem. The problem involves computation of the plane strain state for the body which consists of a massive part and a thin part, which is attached to the massive part. For the numerical solution of this problem we use boundary element method (BEM) and finite element method (FEM) for different parts of the body, which are coupled using domain decomposition. We provide the results of some numerical simulations. The stress-strain state for the structures having shells of different shapes are compared.

1. Introduction

A lot of structures, that occur in engineering, are inhomogeneous and contain thin parts and massive parts. Therefore, it is important to develop both analytical methods and numerical algorithms for the analysis of the stress-strain state of such structures. Different aspects of such problems were discussed in [3, 6, 8, 2] (in [8] the case of the bodies with thin inclusions is considered; in [2] the bodies with thin covers are considered). Papers [3] and [6] are devoted to the numerical solution of the Girkmann problem.

In this article, we solve numerically the Girkmann problem which involves computation of a plane strain state for the body consisting of a massive part and a thin part, which is attached to the massive part. The thin part is modeled using Timoshenko shell theory equations and its stress-strain state is numerically computed using FEM with bubble shape functions. The massive part is modeled using the theory of linear elasticity and the numerical solution is obtained using boundary element method (BEM). The approximate solutions in both parts are connected using domain decomposition algorithm.

The application of domain decomposition method allows us to decouple problems in both parts and solve the problems independently in each part. As a

Key words. Girkmann problem, elasticity theory, Timoshenko shell theory, finite element method, boundary element method, domain decomposition.

result, it is possible to compute the stress-strain state accurately even for small shell thicknesses without having problems with stability issues of the coupled problem.

We compare the stress-strain state for different shapes of the middle line of the shells: circular, parabolic and of the form of chain curve. Although the curves lie close to each other, the stress-strain states in these cases are very different from each other.

2. Problem statement

Let us consider a problem of plane strain of an elastic body which consists of a massive part Ω_1 with the thin part in Ω_2 attached to Ω_1 by its end face (Fig. 1). Let us denote by Γ_i the outer boundary of the bodies in Ω_i, $i = 1, 2$ and by Γ_I the common boundary between bodies in Ω_1 and Ω_2.

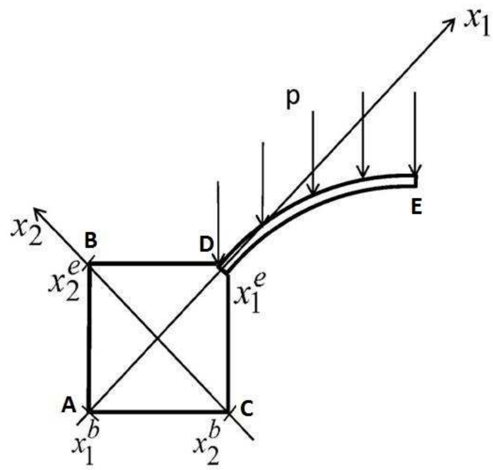


Fig. 1. Elastic Body

The plane strain stress of the body in Ω_1 can be described by

$$\begin{aligned} \frac{\partial \sigma_{11}}{\partial x_1} + \frac{\partial \sigma_{12}}{\partial x_2} &= f_1 \\ \frac{\partial \sigma_{21}}{\partial x_1} + \frac{\partial \sigma_{22}}{\partial x_2} &= f_2 \end{aligned} \qquad (1)$$

that holds for $x \in \Omega_1$, $x = (x_1, x_2)$.

Here $f = (f_1, f_2)$ denotes the volume forces that act on the body in Ω_1.

From the Hook's law it follows that the components of the stress tensor can be written as

$$\sigma_{ij} = \frac{1}{2} E_1 \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right), \qquad i, j = 1, 2,$$

where $u(x) = (u_1(x), u_2(x))$ is the displacement vector with u_i being the displacements in the directions x_i for $i = 1, 2$; E_1 is the Young's modulus of the body in Ω_1. In the following we assume that no volume forces act on the body in Ω_1.

Let us denote by n the outer normal vector to Ω_1, and by τ – the tangent vector.

Equations (1) are considered together with the boundary conditions

$$u_n = 0, u_\tau = 0, x \in \Gamma_D$$

and

$$\sigma_{nn} = 0, \sigma_{n\tau} = 0, x \in \Gamma_N,$$

where $\Gamma_1 = \Gamma_D \cup \Gamma_N$, $\Gamma_D \cap \Gamma_N = \emptyset$; u_n and u_τ are the components of the displacement vector in the coordinate system n, τ. Similarly, σ_{nn} and $\sigma_{n\tau}$ are the components of the stress tensor in the n, τ coordinate system.

For the description of the thin part in Ω_2 we use the equations of Timoshenko shell theory of the form [4]

$$-\frac{1}{A_1}\frac{dT_{11}}{d\xi_1} - k_1 T_{13} = p_1,$$

$$-\frac{1}{A_1}\frac{dT_{13}}{d\xi_1} + k_1 T_{11} = p_3, \tag{2}$$

$$-\frac{1}{A_1}\frac{dM_{11}}{d\xi_1} + T_{13} = m_1, 0 \le \xi_1 \le 1,$$

where v_1, w, γ_1 are the displacements and angle of revolution in the shell; T_{11}, T_{13}, M_{11} are the forces and momentum in the shell; $A_1 = A_1(\xi_1)$, $k_1 = k_1(\xi_1)$ correspond to Lame parameter and middle line curvature parameter; p_1, p_3, m_1 are given functions; it holds

$$T_{11} = \frac{E_2 h}{1 - v_2^2}\varepsilon_{11},\ T_{13} = k' G' h \varepsilon_{13},\ M_{11} = \frac{E_2 h^3}{12\left(1 - v_2^2\right)}\chi_{11},$$

$$\varepsilon_{11} = \frac{1}{A_1}\frac{d\mathrm{v}_1}{d\xi_1} + k_1 w,\ \varepsilon_{13} = \frac{1}{A_1}\frac{dw}{d\xi_1} + \gamma_1 - k_1 \mathrm{v}_1,\ \chi_{11} = \frac{1}{A_1}\frac{d\gamma_1}{d\xi_1},$$

$$p_1 = \left(1 + k_1 \tfrac{h}{2}\right)\sigma_{13}^{+} - \left(1 - k_1 \tfrac{h}{2}\right)\sigma_{13}^{-},$$

$$p_3 = \left(1 + k_1 \tfrac{h}{2}\right)\sigma_{33}^{+} - \left(1 - k_1 \tfrac{h}{2}\right)\sigma_{33}^{-},$$

$$m_1 = \tfrac{h}{2}\left(\left(1 + k_1 \tfrac{h}{2}\right)\sigma_{13}^{+} - \left(1 - k_1 \tfrac{h}{2}\right)\sigma_{13}^{-}\right).$$

Here E_2 is the Young's modulus for the shell, v_2 is the Poisson's ratio; $\sigma_{ij}^{+}, \sigma_{ij}^{-}$, $i, j = 1, 3$ are the components of the stress tensor on the outer ($\xi_3 = \frac{h}{2}$) and inner ($\xi_3 = -\frac{h}{2}$) boundaries of the shell. It is known, that in the case of isotropic bodies we have $k' = \frac{5}{6}$, $G' = \frac{E_2}{2(1+v_2)}$.

At the free end of the thin part we impose boundary conditions either on the displacements v_1, w and γ_1 or on the forces T_{11}, T_{13} and momentum M_{11} in the shell (if the end is subjected to load or free). At the top and bottom outer boundaries of the shell we prescribe to σ_{13}^{+} and σ_{33}^{+} some given stresses.

Remark. The choice of 2D curvilinear coordinate system for the shell as ξ_1, ξ_3 (instead of ξ_1, ξ_2) is based on the fact, that 2D problem is obtained from the 3D case by assuming the body being infinite in the direction of ξ_2.

On the boundary Γ_I, common to both Ω_1 and Ω_2 we prescribe the following coupling conditions:

$$u_n = \mathrm{v}_1 + \xi_3\gamma_1, u_\tau = w,$$

$$\int_{-\frac{h}{2}}^{\frac{h}{2}} \sigma_{nn} d\xi_3 = T_{11}, \int_{-\frac{h}{2}}^{\frac{h}{2}} \sigma_{n\tau} d\xi_3 = T_{13}, \int_{-\frac{h}{2}}^{\frac{h}{2}} \sigma_{nn}\xi_3 d\xi_3 = M_{11}. \tag{3}$$

3. Numerical Approximation of the Model

For the numerical solution of the model domain decomposition algorithm is used. Inside the main part we construct the approximate solution using boundary element method (BEM) applied to the integral equations based on the Green's representation formula for the solution of the following form [1]

$$\frac{1}{2}u_j(x_0)\int_\Gamma (t_i(x)G_{ij}(x,x_0) - F_{ij}(x,x_0)u_i(x))d\Gamma(x), \tag{4}$$

where $\Gamma = \Gamma_1 \cup \Gamma_I$, $x_0 \in \Gamma$;

$G_{ij}(x,\zeta) = C_1(C_2\delta_{ij}\log r - \frac{y_i y_j}{r^2})$ is the matrix Green's function;

$F_{ij}(x,\zeta) = \frac{C_3}{r^2}(C_4(\delta_{ik}y_j + \delta_{jk}y_i - \delta_{ij}y_k) + 2\frac{y_i y_j y_k}{r^2})$ is a co-normal derivative of the matrix Green's function;

$r^2 = y_i y_i$;

$y_i = x_i - \zeta_i$;

$\mu_1 = \frac{E_1}{2(1+\nu_1)}$ is a shear modulus of the body in Ω_1;

$C_1 = -\frac{1}{8\pi\mu(1-\nu_1)}$,

$C_2 = 3 - 4\nu_1$,

$C_3 = -\frac{1}{4\pi(1-\nu_1)}$,

$C_4 = 1 - 2\nu_1$,

In order to apply BEM we divide the boundary $\Gamma_1 \cup \Gamma_I$ of Ω_1 into the elements and then choose the appropriate shape functions $\phi_j(\xi)$, $j = 1, 2, ..., m$, to construct the approximation.

The approximate solution can be written in the form

$$u_i(\xi) = \sum_{j=1}^m u_{ij}\phi_j(\xi), \quad i = 1, 2,$$

$$t_i(\xi) = \sum_{j=1}^m t_{ij}\phi_j(\xi), \quad i = 1, 2, \xi \in \Gamma_1 \cup \Gamma_I,$$

where u_{ij} and t_{ij} are the unknown coefficients that are found by applying Galerkin method to the integral equation (4) (see [1]).

The approximate solution of the boundary value problem inside Ω_2 is found using finite element method with bubble shape functions. On each element the shape functions are given by

$$\Phi_0(\xi) = \frac{1-\xi}{2}, \quad \Phi_1(\xi) = \frac{1+\xi}{2},$$

$$\Phi_j(\xi) = \sqrt{\frac{2j-1}{2}} \int_{-1}^{\xi} P_{j-1}(t)dt, \quad j = 2, 3, ...,$$

where $\xi \in [-1, 1]$ is the local coordinate, obtained by mapping each element onto the inverval [-1,1]; $P_j(t)$ are the Legendre polynomials.

In order to find the approximate solution of the boundary-value problem (2), we apply to the system (2) Galerkin approach.

The approximate solutions in both domains are connected using domain decomposition algorithm (Dirichlet-Neumann scheme) [5]. The domain decomposition algorithm has the following form:

1) set an initial guess λ^0 for the unknown displacements on the interface Γ_I, set $\varepsilon > 0$;

2) for k=0,1,... solve the boundary value problem in Ω_2 with the displacements equal to λ^k to obtain the apporimation for the loads in Ω_1 using (3);

3) solve the corresponding integral equations in Ω_1 to find the displacements u_n^1 and u_τ^1 on Γ_I;

4) update the displacements λ^k on Γ_I:

$$\lambda_1^{k+1} = \lambda_1^k + \theta u_n^1,$$

$$\lambda_2^{k+1} = \lambda_2^k + \theta u_\tau^1,$$

where $\theta > 0$ is a relaxation parameter;

5) if $\|\lambda^{k+1} - \lambda^k\| \geq \varepsilon$ then go to step 2, otherwise the algorithm ends.

It is known, that the Steklov-Poincare equation that corresponds to our problem, possesses a unique solution [7]. Moreover, domain decomposition algorithm converges for appropriately chosen (empirically) relaxation parameter θ $(0 \leq \theta \leq \theta_{max})$ [7].

4. Numerical experiments

Let Ω_1 be a polygon with $x_1^b = -1$, $x_2^b = -1$, $x_1^e = 1$, $x_2^e = 1$. To the main part in Ω_1 a thin body in Ω_2 is attached on its edge. The thickness of the body in Ω_2 is $h = 0.01$ (Fig. 1).

On the boundaries AC and AB the structure is fixed (the displacements are equal to zero); we prescribe a load of $p = 1Pa/m$ on the outer boundary of the body in Ω_2 (Fig. 1); on the edge with the point E the symmetry conditions are set; all the other parts of the outer boundary are traction-free.

We consider the following physical parameters of the bodies: Young's modulus of the main part in Ω_1 is equal to $E_1 = 25000$ MPa, which corresponds to concrete; the Young's modulus of the thin part in Ω_2 is equal to $E_2 = 20580$

MPa, which corresponds to cork. Poisson's ratio of the body in Ω_1 is equal to $\nu_1 = 0.33$, in Ω_2 – $\nu_2 = 0$.

For the numerical solution we use FEM in the shell with bubble shape functions. For the main part we use boundary element method with quadratic shape functions. Problems in both parts are connected using domain decomposition algorithm (Dirichlet-Neumann scheme) [5].

In all the cases under consideration the convergence is obtained in around 5 iterations. The results correspond to a case of 202 boundary elements, 32 finite elements of the fourth order. We find, that the mesh refinement or the change of the order of the shape functions don't change the solution significantly.

Let us consider different cases of the curve shapes, that describe middle line of the body in Ω_2: circle arc, parabola and chain curve. The unknown coefficients of the parametric representation of the curves are chosen in such a way, that all the curves have the same endpoints D and E. Moreover, all the curves are symmetric with respect to the axis, which passes through the point E and is colinear to AB.

In the case of the circle arc the parametric representation has the form

$$x_1(\alpha) = R\sin\alpha,$$

$$x_2(\alpha) = R\cos\alpha, \quad \tfrac{\pi}{4} \le \alpha \le \tfrac{\pi}{2}.$$

Let us choose $R = 5.005$.

In the case of parabola parametric representation has the form

$$x_1(\alpha) = -\tfrac{2-\sqrt{2}}{R}x_2^2 + R,$$

$$x_2(\alpha) = R\cos\alpha, \quad \tfrac{\pi}{4} \le \alpha \le \tfrac{\pi}{2}.$$

In the case of chain curve parametric representation has the form

$$x_1(\alpha) = -\tfrac{4.497}{2}(e^{\frac{x_2}{4.497}} + e^{-\frac{x_2}{4.497}}) + 9.502,$$

$$x_2(\alpha) = R\cos\alpha, \quad \tfrac{\pi}{4} \le \alpha \le \tfrac{\pi}{2}.$$

The graphs of three curves are shown on Fig. 2

We can conclude from Fig. 2, that the graphs of the curves lie close to each other.

Formulae for the calculation of Lame parameter A_1 and curvatures k_1 of the middle line of the shells have the form

$$A_1 = \sqrt{x_1'^2 + x_2'^2},$$

$$k_1 = \tfrac{x_1'' x_2' - x_1' x_2''}{A_1^3}.$$

Let us calculate the stress-strain state for the body depicted on the Fig. 1.

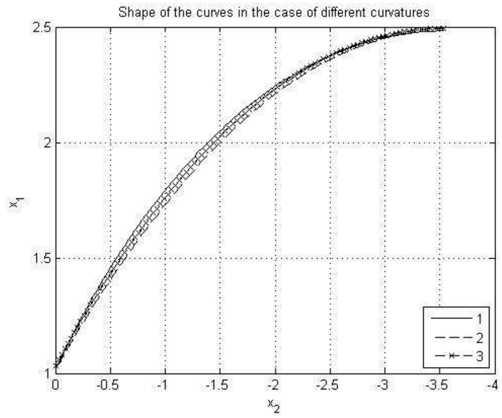


FIG. 2. Middle Line of Different Curves

Fig. 3, 4 show the displacements in the case of different shapes of middle lines, Fig. 5-7 show the momenta that arise on the middle line of Ω_2 in the case of different shapes of middle lines.

Curve 1 on Fig. 3 corresponds to the case of the middle line having the shape of part of the parabola, curve 2 – middle line being the chain curve.

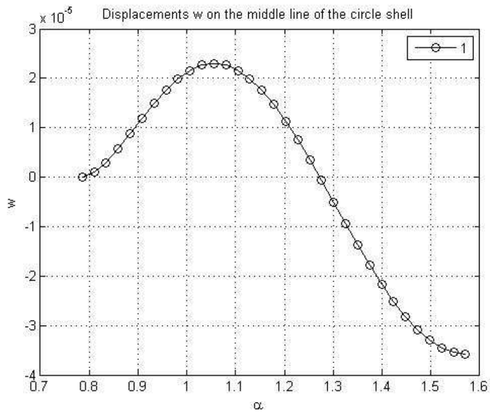


FIG. 3. Displacements w on the middle line of the shell in the case of the circle-shaped shell

On the interface $0 \leq x_2 \leq h$, $x_1 = x_1^e$ we have to set the Neumann condition for the problem in main part, and Dirichlet condition for the problem in the shell. The displacements on the interface for the shell are found using the conditions

$$u_n = v_1 + \xi_3 \gamma_1,$$

$$u_\tau = w.$$

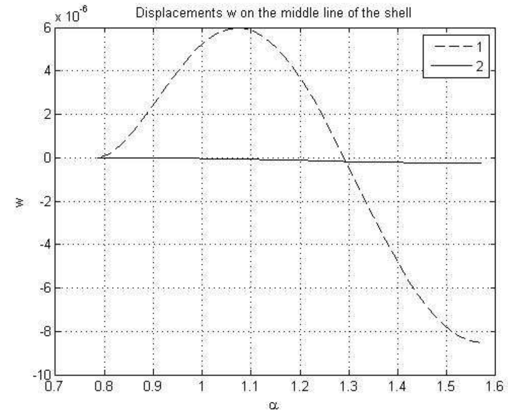


FIG. 4. Displacements w on the middle line of the shell in the case of parabola and chain curve

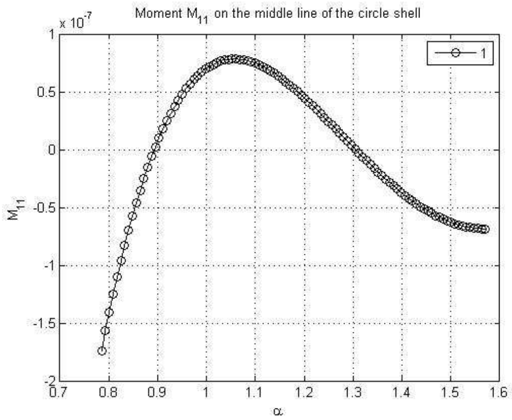


FIG. 5. Momentum m_{11} on the middle line of the shell in the case of the circle-shaped shell

Applying the first condition at the points $\xi_3 = 0$ and $\xi_3 = h/2$, we find that

$$v_1|_{\xi_1=0} = -u_n|_{\xi_3=0},$$

$$\gamma_1|_{\xi_1=0} = \frac{2}{h}(u_n|_{\xi_3=\frac{h}{2}} - u_n|_{\xi_3=0}).$$

Applying the second condition at the point $\xi_3 = 0$, we find that

$$w|_{\xi_1=0} = u_\tau|_{\xi_3=0}.$$

Let us consider the conditions on the loads, that need to be imposed on the interface for the problem in the main part. In order to express $\sigma_{n\tau}$ we use conditions

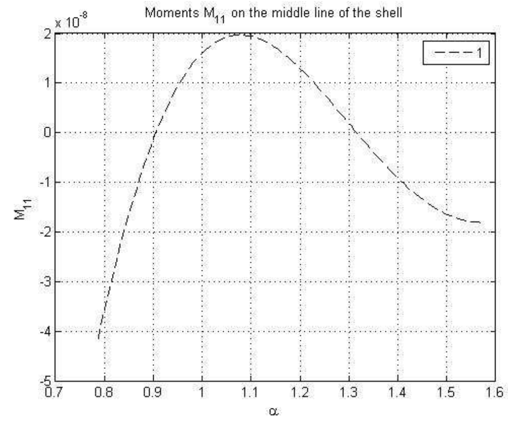


FIG. 6. Momentum m_{11} on the middle line of the shell in the case of parabola

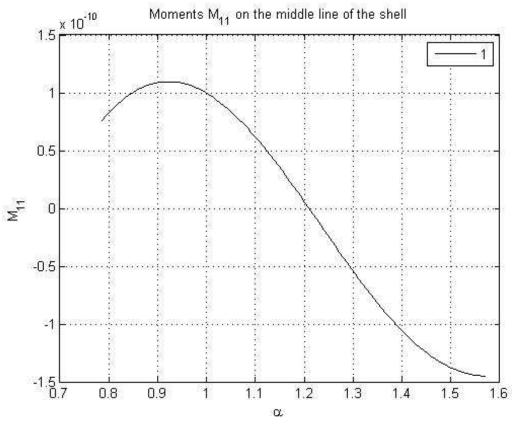


FIG. 7. Momentum m_{11} on the middle line of the shell in the case of chain curve

$$\int_{-\frac{h}{2}}^{\frac{h}{2}} \sigma_{n\tau} d\xi_3 = T_{13}, \quad \sigma_{n\tau}(\xi_3) = \sigma_{13}^{-}|_{\xi_1=0}, \quad \sigma_{n\tau}(\xi_3) = -\sigma_{13}^{+}|_{\xi_1=0}.$$

In order to express σ_{nn} we use conditions

$$\int_{-\frac{h}{2}}^{\frac{h}{2}} \sigma_{nn} d\xi_3 = T_{11}, \quad \int_{-\frac{h}{2}}^{\frac{h}{2}} \sigma_{nn} \xi_3 d\xi_3 = M_{11}.$$

Let us assume that on the interface $\sigma_{n\tau} = a\xi_3^2 + b\xi_3 + c$, $\sigma_{nn} = e\xi_3 + f$, where a, b, c, e, f are the unknown coefficients. These assumptions are based on the fact, that we have three conditions for $\sigma_{n\tau}$ and two conditions on σ_{nn}.

The computations yield

$$\sigma_{nn}(\xi_3) = M_{11}\frac{12}{h^3}\xi_3 + \frac{T_{11}}{h},$$

$$\sigma_{n\tau} = (\frac{3}{h^2}(\sigma_{13}^-|_{\xi_1=0} - \sigma_{13}^+|_{\xi_1=0}) - \frac{6}{h^3}T_{13})\xi_3^2 -$$

$$-\frac{1}{h}(\sigma_{13}^-|_{\xi_1=0} + \sigma_{13}^+|_{\xi_1=0})\xi_3 + \frac{1}{h}(T_{13} - \frac{1}{4}(h(\sigma_{13}^-|_{\xi_1=0} - \sigma_{13}^+|_{\xi_1=0}) - 2T_{13}).$$

From Fig. 3-4 we can conclude, that the smallest displacement in the normal direction is achieved when the middle line of the thin part of the body is a chain curve. The largest displacement in the normal direction arises when the middle line of the thin part is a circle segment.

Fig. 5-7 show, that the smallest momentum is achieved when the middle line of the thin part of the body is a chain curve. The largest momentum arises when the middle line of the thin part is a circle segment.

Therefore, the stress-strain state of the bodies inside the thin part in the case of the Girkmann problem heavily depends on the geometrical parameters of the middle line of the shell (shape, curvature).

5. Conclusions

We conclude, that the stress-strain state of the bodies inside the shell in the case of the Girkmann problem heavily depends on the geometrical parameters of the middle line of the shell (shape, curvature). The elastic body where the shell has the shape of the chain curve, is the best since almost no momentum arises in this case.

The convergence of our algorithm is obtained in around 5 iterations. Therefore, the proposed algorithm can be efficiently applied for the numerical solution of the Girkmann problem.

Bibliography

1. Banerjee P. K. Boundary element methods in engineering science / P. K. Banerjee, R. Butterfield. – McGraw Hill, 1981.
2. Dyyak I. Numerical investigation of a plain strain state for a body with thin cover using domain decomposition / I. Dyyak, Ya. Savula, A. Styahar // Journal of Numerical and Applied Mathematics. – 2012. – Vol. 3(109). – P. 23-33.
3. Niemi A. H. Finite element analysis of the Girkmann problem using the modern hp-version and the classical h-version / A. H. Niemi, I. Babuska, J. Pitkaranta, L. Demkowicz // ICES Report. – 2010.
4. Pelekh B. Generalized shell theory / B. Pelekh. – Lviv, 1978 (in Ukrainian).
5. Quarteroni A. Domain decomposition methods for partial differential equations / A. Quarteroni, A. Valli. – Oxford, 1999.
6. Savula Ya. Coupled boundary and finite element analysis of a special class of two-dimensional problems of the theory of elasticity / Ya. Savula, H. Mang, I. Dyyak, N. Pauk // Computers and Structures. – 2000. – Vol. 75(2). – P. 157-165.
7. Styahar A. Numerical analysis of the strain-stress state for the body with thin inclusion using domain decomposition / A. Styahar, Ya. Savula, I. Dyyak // Mathematical Methods and Ph.-Mech. Fields. – in print (in Ukrainian).

8. Vynnytska L. Numerical investigation of a plain strain state for a body with thin cover using domain decomposition / L. Vynnytska, Ya. Savula // Ph-math. Modelling and Information Technologies. – 2008. – Vol. 7. – P. 21-29 (in Ukrainian).

Andriy Styahar,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine

Received 21.05.2014

Журнал обчислювальної та прикладної математики | 2014 № 2 (116) | Journal of Computational & Applied Mathematics

UDC 517.958:534+519.6

NUMERICAL MODELLING OF SHALLOW-WATER FLOW IN HYDRODYNAMIC APPROXIMATIONS

Petro Venherskyi, Valerii Trushevskyi

Резюме. Сформульовано двовимірну початково-крайову задачу руху води на території водозбору. Для виводу рівнянь руху проводилися усереднення доданків за глибиною потоку та враховувалися умови мілкості потоків. Побудовано відповідну варіаційну задачу, для якої при дискретизації за просторовими змінними використано метод скінченних елементів і за часом – однокрокову рекурентну схему. Для великих чисел Рейнольдса побудовано стабілізаційну схему, що базується на функціях-бульбашках із використанням методу найменших квадратів.Числові результати апробовано на тестових прикладах для різних початкових та крайових умов, у різні моменти часу і при виборі великих значень чисел Рейнольдса.

Abstract. Formulated a two-dimensional initial-boundary value problem of movement of water in the watershed. To derive the equations of motion were carried averaging summands in the depth flow and conditions of shallow flows were taken into account. The variational problem was built for it in discretization for spatial variables used finite element method and time - one-step recurrent scheme. For large Reynolds numbers built stabilization scheme based on functional bubbles by the method of least squares. Numerical results tested on test examples for different initial and boundary conditions, at different times and in selecting high values of Reynolds numbers.

1. Introduction

One of the most important processes of a hydrological cycle concerns to a shallow water flows to which belong rain and channels flows, water flow from a watershed surface, motion of water in ocean, etc. Processes which underlie of this model have wave nature, with wave length is much greater then the vertical dimensions. To describe these processes is possible outgoing from general equations of Navier-Stokes or from equations of Reynolds. From supposition, that the horizontal scales of fluid motion are much more vertical, the average on vertical component of a flow is realized. The detailed derivation of average equations of shallow water from equations of the Reynolds can be found in works [3],[6]. Equations looks like following:

†*Key words.* Variational problem, initial-boundary value problem, Galerkin approximations, shallow-water flow, Navier-Stokes equations, hydrodynamic approximations.

$$\begin{cases} \frac{\partial q_i}{\partial t} + \sum\limits_{j=1}^{2} \frac{\partial}{\partial x_j}\left(\frac{q_i q_j}{h}\right) = \sum\limits_{j=1}^{2} \frac{\partial N_{ij}}{\partial x_j} - \frac{\partial N_p}{\partial x_i} + B_i, \\ N_{ij} \approx \varepsilon_{ij}\left(\frac{\partial q_i}{\partial x_j} + \frac{\partial q_j}{\partial x_i}\right), \\ \frac{\partial(\rho h)}{\partial t} + \sum\limits_{j=1}^{2} \frac{\partial q_j}{\partial x_j} = 0, \quad i,j = 1,2, \end{cases} \tag{1}$$

where ε_{ij} - vortex viscosity coefficient, $q_i = hu_i$ – unknowns of value flows, $B_i = \tau_i|_\xi - \tau_i|_\eta + p_a \frac{\partial h}{\partial x_i} + \rho g h \frac{\partial \eta}{\partial x_i}$ (i = 1,2), $N_p = \rho g \frac{h^2}{2} + hp_a, p_a$ - atmospheric pressure, ξ - free surface of flow, - bottom contour, h - flow depth, $\tau_i|_\xi$ та $\tau_i|_\eta$ - stresses on free surface and bottom contour accordingly.

Average equations of shallow water deduced from general Navier-Stokes equations in the works [1,2],[4],[7]. It looks like

$$\begin{cases} \frac{\partial u_i}{\partial t} + \sum\limits_{j=1}^{2} u_j \frac{\partial u_i}{\partial x_j} + g\frac{\partial h}{\partial x_i} + \frac{\left(u_i - u_i^0\right)R - u_i I}{h} = -g\frac{\partial \eta}{\partial x} - \frac{F_i}{h} - \frac{\partial (R\Lambda)}{\partial x_i}, \\ \frac{\partial h}{\partial t} + \frac{\partial (hu_j)}{\partial x_j} = R - I, \quad i = 1,2, \end{cases} \tag{2}$$

where u_i - unknowns of speed value, h - unknown flow depth, u_{i_0} - velocity on a free surface, g - acceleration of gravity, I - speed of fluid infiltration into the ground, R - rain inflow velocity, η - bottom contour, Λ - speed of falling of rain drops, F_i - items which allow for tangential stresses on the bottom and on the free surface of a flow.

In motion equations from viscous terms there are only tangential stresses on a free surface and at the bottom, others are rejected in conditions of shallow water. In a result of averaging system of equations set by depth of a flow and allowing conditions of shallow water, the third equation of motion will be converted to the hydrostatic law of pressure, which is characteristic for shallow water equations

$$p(z) = p(\xi) - \rho f_3 (\xi - z).$$

For completion of problem formulation equation of shallow-water supplement by an initial and boundary conditions. The boundary conditions in the literature partition on two kinds: those which are set on hard boundary of flow and on opened boundary. On each of boundaries it is necessary to set two conditions: normal and tangent components of stresses or speeds. For model (2) are set only normal components [2]:

on hard boundary

$$q_n = 0 \; or \; q_n = \overline{q}_n;$$

on opened boundary

$$N_{nn} = \overline{N}_{nn}.$$

It is explained to those that in model (1) the terms that take into account votex viscosity are discarded, therefore tangent components of stresses or flows are not set.

Let's consider one more version of assigning of a boundary conditions. Let Ω - projection of a fluid flow on a two-dimension plane. The boundary of area Ω is partitioned on following parts: Γ_B - fixed boundary of a watershed, Γ_R-boundary of a channel (the fluid inflows), Γ_S - opened sea border (the fluid can both inflow and outflow, see Fig. 1).

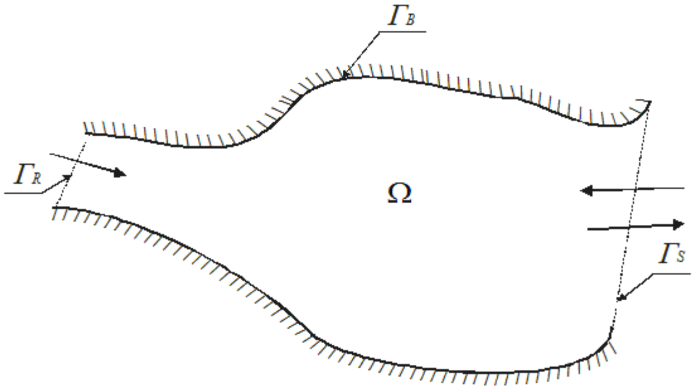


FIG. 1. Projection of a fluid flow on a two-dimension plane

More often boundary conditions for two-dimension problem of shallow water write down [2,4,9-13]:

– on fixed boundary Γ_B of a flow set

$$U \cdot \nu = 0, \quad \nabla U_\tau \cdot \nu = 0,$$

where ν and τ - units normal and tangent to bound of domain, U_τ - tangential components of velocity;

– on boundary of fluid inflow:

$$U \cdot \nu = \hat{U} \cdot \nu, \quad \mu \frac{\partial U}{\partial \nu} \cdot \tau = 0,$$

where μ - coefficient of viscosity;

– on opened sea border the boundary conditions it is possible to set as

$$\frac{\partial U}{\partial \nu} = 0.$$

In considered above shallow water models all items which contain component of stresses are skipped. Component of stresses are saved only on a free surface and on the bottom of flow. Scientific approach, which is submitted in this work saves all components of stresses in motion equations. For solving of shallow water problem the finite element method was selected.

2. Formulation of initial-boundary problem

Suppose that flow of viscous incompressible fluid in each point of time $t \in [0,T], 0 < T < +\infty$, forms on an immovable surface $x_3 = \eta(x_1, x_2)$ of watershed some fluid layer $D = D(t)$ (Fig.2).

Let's designate through $\xi(x,t)$ a free surface of this flow, which contacts to atmosphere, where $x = (x_1, x_2, x_3) \in R^3, \nu$ - unit outward normal of domain

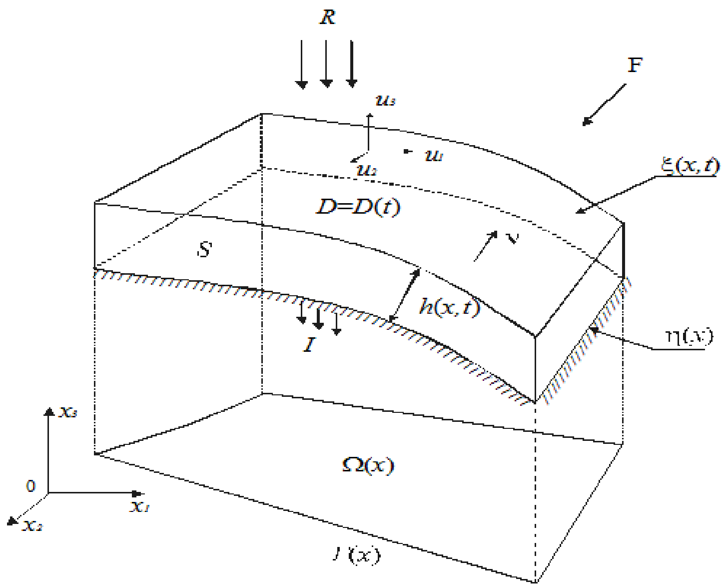


FIG. 2. Model of shallow water flow

$D = D(t)$. Lateral (vertical) surface of this flow, if such exists we shall designate through S. Let's mark, that the part of a surface S can be degenerated in boundary Γ of watershed river. So $\partial D(t) = \eta \cup \xi(t) \cup S$.

Projection of a fluid layer $D(t)$ on a horizontal plane we will denote as Ω. Assume, that boundary γ of domain continuous by Lipschitz.

Let's guess, that a fluid state under the influence of mass forces $F = \{f_i(x)_{i=1}^3$ in each point of time $t \in [0,T], 0 < t < +\infty$ is described by of the Navier-Stokes equations

$$\begin{cases} \rho\left(\frac{\partial u_i}{\partial t} + \sum\limits_{k=1}^{3} \frac{\partial}{\partial x_k}(u_i u_k) - f_i\right) - \sum\limits_{k=1}^{3} \frac{\partial \sigma_{ik}}{\partial x_k} = 0, \\ \sigma_{ij} = -p\delta_{ij} + \tau_{ij}, \\ \tau_{ij} = 2\mu e_{ij}, \\ e_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right), \\ div\ \vec{u} = 0, \quad i,j = 1,2,3, \end{cases} \tag{3}$$

where $div\ \vec{u} = \sum\limits_{k=1}^{3} \frac{\partial u_k}{\partial x_k}, \vec{u} = \{u_i(x,t)\}_{i=1}^3$ and $p = p(x,t)$ - velocity vector and hydrostatic pressure accordingly, $F = fi(x,t)_{i=1}^3$ - vector of mass forces, $\rho = const > 0$ and $\mu = const > 0$- density and viscosity, $\{e_{ij}\}_{i,j=1}^3$, $\{\sigma_{ij}\}_{i,j=1}^3$ - velocity and stresses tensors, δ_{ij} - Kroneker symbol.

Let in an initial time water flow described by conditions

$$u_i|_{t=0} = u_i^0 \ in\ D(0), i = 1,2,3. \tag{4}$$

Except of initial conditions, the equations are necessary supplement by the applicable boundary conditions, which determine interaction of flow water with atmosphere, surface of ground, groundwater etc.The main factors which influence on a fluid state:

– intensive rain precipitations, evaporation of water;
– replenishing of water from channel;
– infiltration of water in soil(groundwater replenishment);
– atmospheric wind, etc.

Attempts to describe characteristic modes of shallow-water flows result in simplification of equations (3) and respective to them boundary conditions and will be reviewed later. At the given stage we will limit by a typical boundary conditions for this equations [2,5,7,9-13]:

$$u = \widehat{u} \; on \; B_u \times (0,T], mes(B_u) > 0, B_u \subset \partial D(t), \tag{5}$$

$$\tau_{ij}\nu_j = \widehat{\tau}_i \; on \; B_\tau \times (0,T], B_\tau \subset \partial D(t) \backslash B_u, \; i,j = 1,2,3, \tag{6}$$

where $u = \{u_i\}_{j=1}^3, \nu = \{\nu_i\}_{j=1}^3$ - unit outward normal of bound $\partial D(t), \nu_i = cos(\nu, x_i)$.

Generally free surface of a flow $\xi(x,t)$ is unknown, therefore it is necessary to set conditions for definition of its position in space in each point time. For finding of a free surface $x_3 = \xi(x_1, x_2, t)$ we shall use a kinematic condition [16]:

$$u_3 + R = \frac{\partial \xi}{\partial t} + u_1^0 \frac{\partial \xi}{\partial x_1} + u_2^0 \frac{\partial \xi}{\partial x_2}, \tag{7}$$

where R – rain velocity, u_1^0, u_2^0 - horizontal components of velocity on a free surface and initial condition

$$\xi|_{t=0} = \xi^0 \; in \, \Omega. \tag{8}$$

On the bottom of flow the fluid can flow in a soil in a direction of an axis x_3

$$u_3 = -I \; on \; [0,T], \tag{9}$$

where I - velocity of seepage water in soil. If I $= 0$ does it mean that surface is impermeable ; I > 0 - fluid particles seepage in a soil with a preset speed; I $<$ 0 - the groundwaters rise on a back surface of ground.

On a base surface for velocity we shall allow for a condition of adhesion

$$u_1 = u_2 = 0. \tag{10}$$

The initial-boundary problem (3)-(10) is difficult to applying for a nature watersheds and requires simplifications. At the first stage (3) we will reduce equation to a undimensional kind. Such form will give a chance to receive numbers, which characterize motion of water (Reynold's number), and also the parameters of equations are such normalized that their values will change in definite limits. At the second stage, allowing conditions of shallow water, neglect terms order of smallness $\varepsilon = \delta/L$ (the maximum thickness of a flow does not exceed the size δ, and characteristic horizontal dimensions value $L, and(\delta/L << 1)$).

All components of stresses in two first equations of motion remain saved after simplification. The following step of simplifications is reduction of a problem dimension at the expense of a depth averaging of equations. After an average is received a two-dimension problem of a water flow in hydrodynamic approximation concerning three unknowns - two components of flow and depth:

$$\begin{cases} \frac{\partial q_i}{\partial t} + \sum\limits_{j=1}^{2} \frac{\partial}{\partial x_j}\left(q_i \frac{q_j}{h}\right) + Gh\left(\frac{\partial h}{\partial x_i} + \frac{\partial \eta}{\partial x_i}\right) - \frac{1}{\rho Re}\sum\limits_{j=1}^{2} \frac{\partial(\tau_{ij} h)}{\partial x_j} - \frac{g|q|q_i}{ReC^2h^2} = 0, \\ \frac{\partial h}{\partial t} + \sum\limits_{j=1}^{2} \frac{\partial q_j}{\partial x_j} = R - I, \\ \tau_{ij} = \mu\left(\frac{\partial(q_i/h)}{\partial x_j} + \frac{\partial(q_j/h)}{\partial x_i}\right), \quad i, j = 1,\ 2, \end{cases} \tag{11}$$

where h - unknown depth, $q = (q_1, q_2)$ - unknown vector of flow, η - bottom contour,ρ - density of water, Re - Reynolds number, τ_{ij} - stresses tensor, μ - viscosity of water, C - Shezi factor,g - gravitational acceleration, $G = \frac{gL}{V_\infty^2}$, L - typical spatial size, V_∞ - typical velocity, R - rain inflow, I - water seepage in a soil.

The first two equations of system are averaged equations of motion, which are parabolic type. Their novelty consists in preservation of addend with internal stresses of a flow, which are essential on surfaces with considerably change gradients. In the literature the hyperbolic equations of a shallow water flow are considered where the stresses only on the bottom and on a free surface of a flow are taking into consideration. In this case it is supposed that the wind stresses are negligible. The third equation of a system is an averaged equation of continuity, which describes a free surface of a flow.

Let's consider a water flow from a surface watershed in a projection on a horizontal plane. Here Ω - two-dimension domain which restricted by curve Γ_B (watershed line) and Γ_P (outflow line), n, ζ - normal and tangent to boundary of area accordingly.

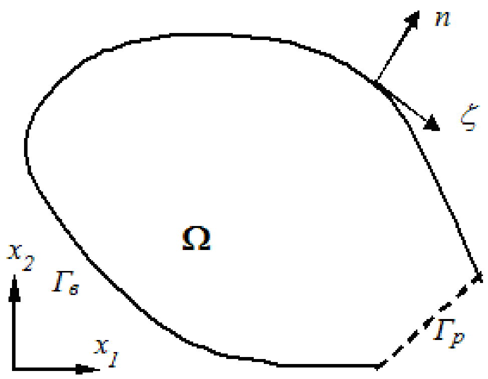


FIG. 3. Water flow projection on a horizontal plane

Equations of system (11) are added by boundary conditions

$$\tau_\zeta|_{\Gamma_B} = 0, q \cdot n|_{\Gamma_B} = 0, \quad q \cdot \zeta|_{\Gamma_P} = 0, q \cdot n|_{\Gamma_P} = \hat{q} \tag{12}$$

and initial conditions

$$h|_{t=0} = h_0, \quad q|_{t=0} = q_{0\text{B}}\ in\,\Omega, \tag{13}$$

where $\hat{q}$ – known water outflow.

In outcome we received a system of three equations (11) for searching components of vector of a flow and depth with boundary (12) and initial (13) conditions. We will decide a problem (11)-(13) by a finite element method (FEM)[8,12-13].

3. Applying of a FEM to the problem solution

According to a procedure of a FEM it is necessary to make a variational formulation. For formulation of variational initial-boundary target setting (11) – (13) we will enter set of allowed functions for flows

$$Q\left(\hat{q}\right) := \left\{ q = \{q_i\}_{i=1}^{2} \in H^1\left(\Omega\right)^2 \,\middle|\, q\cdot n|_{\Gamma_P} = \hat{q},\, q\cdot n|_{\Gamma_B} = 0,\, q\cdot\zeta|_{\Gamma_P} = 0 \right\}$$

and space $Q_0 = Q(0)$. Space of allowed(permissible) functions for depth - $\Phi := L^2(\Omega)$. Let's search a flow as $q = q_* + \bar{q}$ with unknown $q_* \in Q_0, \bar{q}.n = \hat{q}\; on\, \Gamma_p$. Further, for simplicity of identifications we will use instead of q_* identification q.

Let's enter the following forms

$$\begin{cases} a(q,p) = \int\limits_{\Omega} q\cdot p dx, \quad b(w;q,p) = \int\limits_{\Omega} \sum\limits_{i,j=1}^{2} p_i \frac{\partial}{\partial x_j}\left(q_i w_j\right) dx, \\ c\left(h;w,p\right) = \frac{1}{\rho}\int\limits_{\Omega} h \sum\limits_{i,j=1}^{2} \tau_{ij}\left(w\right) \frac{\partial p_i}{\partial x_j} dx, \quad d\left(z;h,p\right) = \frac{1}{2}\int\limits_{\Omega} Gzh\left(\nabla\cdot p\right) dx, \\ l\left(\eta;h,p\right) = \int\limits_{\Omega} G\eta\nabla\cdot\left(hp\right) dx, \quad \bar{R}\left(h,q,p\right) = \int\limits_{\Omega} \frac{g|q|(q\cdot p)}{C^2h^2} dx, \\ \qquad \forall p,\ q,\ w \in Q_0, \\ m\left(q,\theta\right) = \int\limits_{\Omega} \left(\nabla\cdot q\right)\theta dx, \quad \langle s,\theta\rangle = \int\limits_{\Omega} \left(R - I\right)\theta dx, \quad \forall\theta,\ z\in\Phi. \end{cases} \tag{14}$$

Then, take into 3, the variational initial-boundary target setting to become

$$\begin{cases} Given\, q_0 \in Q_0, h_0 \in \Phi; \\ Find\, q \in Q_0, h \in \Phi\; such\; that \\ a\left(q'\left(t\right),p\right) + b\left(q\left(t\right)/h(t); q\left(t\right),p\right) - d\left(h\left(t\right); h\left(t\right),p\right) - \\ -l\left(\eta; h\left(t\right),p\right) + \frac{1}{Re}\left[c\left(h\left(t\right); q/h\left(t\right),p\right) - \bar{R}\left(h\left(t\right); q\left(t\right),p\right)\right] + \\ \qquad +a\left(F(\hat{q}),p\right) = 0, \\ a\left(h'\left(t\right),\theta\right) + m\left(q\left(t\right),\theta\right) + a(V(\hat{q}),\theta) = \langle s\left(t\right),\theta\rangle \forall t \in [0,T]\,, \\ a(q\left(0\right) - q_0, p) = 0,\ a\left(h\left(0\right) - h_0,\theta\right) = 0 \quad \forall p\in Q_0, \forall\theta\in\Phi, \end{cases} \tag{15}$$

where $F(\hat{q})\; and\; V(\hat{q})$ – items accordingly of first and second equations of a system, which are formed by a flow components $\hat{q}$.

We will decide the variational problem with usage of a projective-net scheme of FEM. Let's conduct a discretization of a problem in time. Interval of time [0,T] we will divide into $N_T + 1$ identical parts $[t_k, t_{k+1}]$ by length Δt and we will select approximations for depth and flows as

$$h(x,t) \approx h_{\Delta t}(x,t) = h^k\left(x\right) + H^{k+\frac{1}{2}}\left(x\right)\Delta t\omega\left(t\right), \tag{16}$$

$$q(x,t) \approx q_{\Delta t}(x,t) = q^k\left(x\right) + U^{k+\frac{1}{2}}\left(x\right)\Delta t\omega\left(t\right), \tag{17}$$

where

$$H^{k+\frac{1}{2}} = \frac{h^{k+1} - h^k}{\Delta\, t}, H^{k+\frac{1}{2}} \in \Phi, \quad \forall x \in \Omega,\ \forall t \in [t_k, t_{k+1}], \quad k = 0, \quad \ldots, N_T.$$

It is known, that if we approximate a function by an interpolation polynomial of the first order, the precision greater than Δt^2 cannot be obtained. Therefore at the given stage (phase) we can conduct a linearization of a problem by throwing off terms of the order. By substituting (16) – (17) in a variational problem (15) and ignore terms of the order Δt^2, we receive a linearized problem as the one-step recurrent scheme of integrating in time

$$\left\{\begin{array}{l} Given \quad q^0 \in Q_0,\ h^0 \in such\ that\ \lambda \in (0,\ 1]\ ; \\ Find \quad U^{k+\frac{1}{2}} \in Q_0,\ H^{k+\frac{1}{2}} \in Phi\ ,\ such\ that \\ a(U^{k+\frac{1}{2}}, p) + \\ +\lambda\Delta t\Big[b(q^k/h^k; U^{k+\frac{1}{2}}, p) + b(U^{k+\frac{1}{2}}; q^k/h^k, p) - 2d(H^{k+\frac{1}{2}}; h^k, p) - \\ -l(\eta; H^{k+\frac{1}{2}}, p) + \frac{1}{Re}\Big(c(H^{k+\frac{1}{2}}; q^k/h^k, p) + c(h^k; U^{k+\frac{1}{2}}/h^k, p)\Big)\Big] = \\ = d\left(h^k; h^k, p\right) + l\left(\eta; h^k, p\right) - b\left(q^k/h^k; q^k, p\right) - \\ -\frac{1}{Re}\left[c\left(h^k; q^k/h^k, p\right) - \bar{R}\left(h^k; q^k, p\right)\right] - a(F_{k+1/2}, p), \\ a(H^{k+\frac{1}{2}}, \theta) + \lambda\Delta t m(U^{k+\frac{1}{2}}, \theta) = \\ =< s_{k+1/2}, \theta > -m(q^k, \theta) - a(V_{k+1/2}, \theta), \\ q^{k+1} = q^k + \Delta t U^{k+\frac{1}{2}},\ h^{k+1} = h^k + \Delta t H^{k+\frac{1}{2}},\ k = 0, ..., N_T, \end{array}\right. \tag{18}$$

where $F_{k+1/2} = F(t_k + \Delta t/2), V_{k+1/2} = V(t_k + \Delta t/2), s_{k+1/2} = s(t_k + \Delta t/2)$.

At a discretization of a problem (18) according to space variables are utilised piecewise linear approximatings on triangular elements for flows and piecewise constant approximatings of depthes. Such selection of approximatings allows to eliminate depth of a flow and to receive a system of simple equations only concerning vector of a flow.

For a discretization of a problem according space variables the domain Ω is divided into triangular finite elements. Let's enter the spaces for flows $Q_0^h \subset Q_0, \dim\ Q_0^h = N_p < \infty$ and for depthes $\Phi^h \subset \Phi, dim\ \Phi_0^h = N_e\ <\ \infty$. Let's select piecewise linear approximatings for flows

$$\varphi_i\left(x_1, x_2\right) = \left\{\begin{array}{l} L_i\left(x_1, x_2\right), P_i \in \Omega_e, \\ 0, P_i \notin \Omega_e \end{array}\right.$$

and piecewise constant for depthes

$$\psi_e\left(x_1, x_2\right) = \left\{\begin{array}{l} 1, P \in \Omega_e, \\ 0, P \notin \Omega_e. \end{array}\right.$$

Further using a procedure of a Galorkin method, we will obtain a system of simple equations concerning unknowns of vector of a flow W in nodal values of a grid and vector of depths S in center of gravity of triangles:

$$\begin{cases} Given\ q^0 \in Q_0, \quad h^0 \in \Phi \quad and \quad \lambda \in (0,1] \ ; \\ Find\ U_h^{k+\frac{1}{2}} = \sum\limits_{i=1}^{N_p} W_i^{k+\frac{1}{2}} \varphi_i \in Q_h^0, \quad H_h^{k+\frac{1}{2}} = \sum\limits_{e=1}^{N_e} S_e^{k+\frac{1}{2}} \psi_e \in \Phi^h \\ such,\ \ that\ \ a(U_h^{k+\frac{1}{2}}, p) + \\ +\lambda \Delta t \Big[b(q^k/h^k; U_h^{k+\frac{1}{2}}, p) + b(U_h^{k+\frac{1}{2}}; q^k/h^k, p) - 2d(H_h^{k+\frac{1}{2}}, h^k, p) - \\ -l(\eta, H_h^{k+\frac{1}{2}}, p) + \frac{1}{Re} \left(c(H_h^{k+\frac{1}{2}}; q^k/h^k, p) + c(h^k; U_h^{k+\frac{1}{2}}/h^k, p) \right) \Big] = \\ = d\left(h^k, h^k, p\right) + l\left(\eta, h^k, p\right) - b\left(q^k/h^k; q^k, p\right) - \\ -\frac{1}{Re}\left[c\left(h^k; q^k/h^k, p\right) - \bar{R}\left(h^k; q^k, p\right) \right] - a(F_{k+\frac{1}{2}}, p) \quad \forall p \in Q_0, \\ a(H_h^{k+\frac{1}{2}}, \theta) + \lambda \Delta t m(U_h^{k+\frac{1}{2}}, \theta) = \\ =< s_{k+\frac{1}{2}}, \theta > -m(q^k, \theta) - a(V_{k+\frac{1}{2}}, \theta) \quad \forall \theta \in \Phi, \\ q^{k+1} = q^k + \Delta t U_h^{k+\frac{1}{2}}, \qquad h^{k+1} = h^k + \Delta t H_h^{k+\frac{1}{2}}, \qquad k = 0, ..., N_T. \end{cases} \tag{19}$$

On a Fig. 4 completely sampled equations are sketched on one finite element

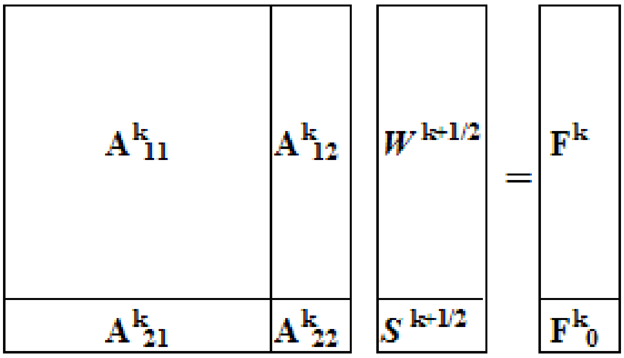


FIG. 4. Diagrammatic representation of a system of simple equations

A_{22} – diagonal matrix. At the expense of condensation of internal parameters we can eliminate depth on one finite element by using a ratio

$$S_e^{k+\frac{1}{2}} = A_{22}^{k^{-1}}(F_0^k - A_{21}^k W_e^{k+\frac{1}{2}}). \tag{20}$$

In outcome we will obtain a system of simple equations concerning two unknowns – flow components

$$(A_{11}^k - A_{12}^k A_{22}^{k^{-1}} A_{21}^k) W_e^{k+\frac{1}{2}} = F^k - A_{12}^k A_{22}^{k^{-1}} F_0^k.$$

4. Stabilization scheme FEM

At large values of Reynold's numbers (Re>100) flows and their gradients change sharply. As outcome the obtained solution of a shallow water problem loses the stability and appears oscillations. On this case, stabilization scheme is obtained, which is based on bubble functions with usage of a least-squares method. As the depth of a fluid is considered as a constant on one finite element, it does not influence behaviour of the solution. In a system (18) the stabilization addend is added to equations of flows in the next view

$$
\begin{aligned}
S(U^{k+\frac{1}{2}}, H^{k+\frac{1}{2}}, p) = M_e \Bigg[\int\limits_{\Omega_e} U^{k+\frac{1}{2}} \cdot p dx + \\
+ \Delta t \lambda \Bigg[\int\limits_{\Omega_e} \sum_{j=1}^{2} \left((\frac{\partial}{\partial x_j} ((q_i^k U_j^{k+\frac{1}{2}}) + (q_j^k U_i^{k+\frac{1}{2}}))/h^k \right) p dx + \\
+ \int\limits_{\Omega_e} \left(U_i^{k+\frac{1}{2}} + \sum_{j=1}^{2} \frac{\partial}{\partial x_j} (q_i^k q_j^k)/h^k + G h^k \frac{\partial \eta}{\partial x_i} - \frac{g|q^k| q_i^k}{ReC^2 \left(h^k\right)^2} \right) \times \\
\times \left(\sum_{j=1}^{2} \frac{\partial}{\partial x_j} ((q_i^k p_j) + (q_j^k p_i))/h^k \right) dx + \int\limits_{\Omega_e} G H^{k+\frac{1}{2}} \frac{\partial \eta}{\partial x_i} p dx \Bigg] - \\
- \int\limits_{\Omega_e} \frac{g|q^k| q_i^k p_i}{ReC^2 (h^k)^2} dx + \int\limits_{\Omega_e} \sum_{j=1}^{2} \frac{\partial}{\partial x_j} ((q_i^k q_j^k)/h^k) p dx + \int\limits_{\Omega_e} G h^k \frac{\partial \eta}{\partial x_i} p dx \Bigg],
\end{aligned}
\tag{21}
$$

where M_e -– stabilization factor on each finite element.

For stabilization factor M_e using the upper-bound estimate μ_0 obtained in the work [6] for approximating scheme of Navier-Stokes equations

$$
\mu_0 = \frac{7}{5} \left(\frac{1}{7kd^2/\Delta^2 - e} \right), \tag{22}
$$

where Δ - square of finite triangle element, $d^2 = l_1^2 + l_2^2 + l_3^2, l_i$ - length of triangle side (i=1, 2, 3), $e = div\, w, w$ - know velocity from previous step, k - kinematic viscosity of a fluid.

5. Test examples

Example 1. Let's consider a problem of shallow water flow from a surface some watershed. All parameters of a problem are set in a dimensionless view. Let's select a test surface watershed $\eta(x, y)$ as Fig. 5, where x, y change from 0 to 2. In an initial time we will enable that h_0 =0.01, q_i=0 (i=1,2). Concerning boundary conditions, we enable, that the water does not outflow and normal component of flow velocities on boundary of domain is equal zero $q\dot{n}$=0. We enable, that constant rain influx R=1, infiltration of a fluid in a ground I=0, coefficient factor Shezi C=60, Reynold's number Re=0.1. Quantity of splitting points of domain 60x60. For the solution of a problem we apply the numeric scheme (19), in which parameters $\lambda = 0.5$, $\Delta t = 0.005$. Let's consider result in a point of time t = 0.195 (quantity of steps in time tt=40).

In a Fig. 5 the depth of a flow H (quantity of water is figured, which collects at the bottom surface with constant rain influx). As the water does not outflow, cavities are filled by the water. From results apparently, that the maximum value of depth is reached in the middle of a bottom surface, where there is a greatest cavity. In the highest points of surface watershed values of depth are approaches to zero, as the water flows down.

The law of conservation of mass for the given example is tested. The conducted calculations have shown, that the volume of the fall out precipitations approximately coincides with water volume on a given surface in the given point of time 0.78105.

In Fig.7 and Fig.8 are figured values components of a fluid flow accordingly on axes x and y. In a Fig.9 the module of a flow is figured. From results it is possible to see, that the flow has zero values in those points of a bottom surface, where the fluid collects and whence the water flows off, in these extreme points water is not gone. The maximum values of a flow are reached in currents, where there is a maximum slope of a bottom surface to horizont.

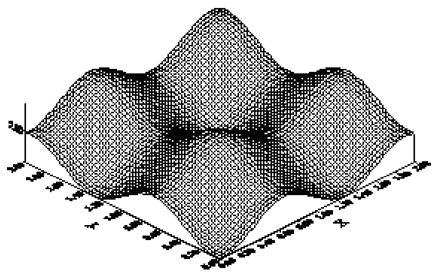


FIG. 5. Bottom surface $\eta(x, y)$

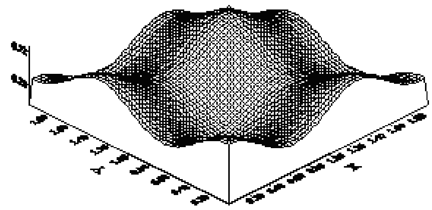


FIG. 6. Flow depth H

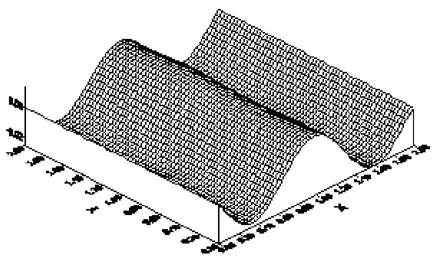


FIG. 7. Flow component Qx

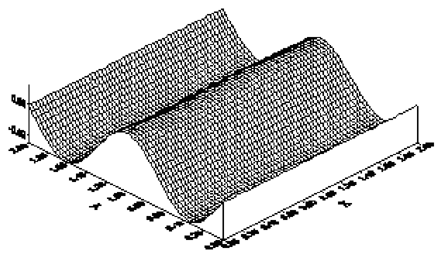


FIG. 8. Flow component Qy

Example 2. By important point at problem solving of shallow water is selection of a Reynold's number values. When parameter receives large values (Re>100), solution obtained with the help of the numeric scheme (19), loses the stability, values of flows and their gradients are very large, as a result of it there are oscillations. In the Fig. 10 the values of depthes of a problem with parameters by given in an example 1 and Reynold's number Re = 150 are figured. On Fig. 11 the values of component flows accordingly on axis x are figured. The results are displayed in a point of time t = 0.073 (quantity of steps in time tt = 15, $\Delta t = 0.005$).

For the solution of this problem the stabilization scheme of a finite element method with stabilization factor (21) was obtained. We apply the stabilization

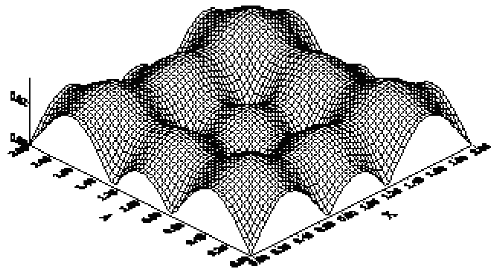


FIG. 9. Module of flow

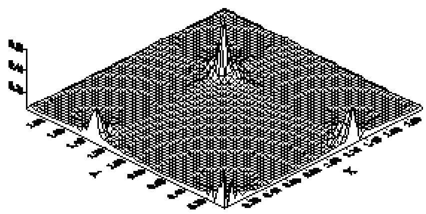


FIG. 10. Flow depth H

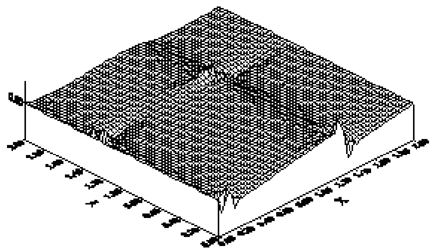


FIG. 11. Flow component Qx

scheme to the solution of our problem with a Reynold's number Re=150 and stabilization factor $M_e = -0.5$. Let's consider computing results in a point of time t = 0.586 (quantity of steps in time tt=60), quantity of splitting points of domain 30x30. In a Fig. 12 the values of depth are figured, the Fig. 13, Fig. 14, Fig. 15 - represent values components and module of a flow accordingly.

From results it is possible to see, that the problem, which has arisen, at applying the numeric scheme (19) to the solution of a problem, is decided positively

The results are smoothed at the expense of the introducing of a stabilization factor. The computing results have shown, that the problems of a shallow water flow can be decided with any values of Reynold's numbers, applying the stabilization scheme of a finite element method.

The law of conservation of mass for the obtained outcomes is executed. The volume of the fall out precipitations coincides with a volume of a fluid on a surface watershed 2.34314.

Example 3. Let's consider a water flow from a surface watershed Fig. 16 (part of Perespil countryside in the Lvov area). Boundary and initial conditions we will select similarly to the previous example, quantity of splitting points of domain 60x60, stabilization a factor M_e = -0.5. Let's consider the results in a point of time t = 0.146 (quantity of steps in time tt=30) with a Reynold's number Re=150. In a Fig. 17 the depth H of a water flow is displayed. For the greater visualization we compare isolines of a watershed surface (Fig. 18)

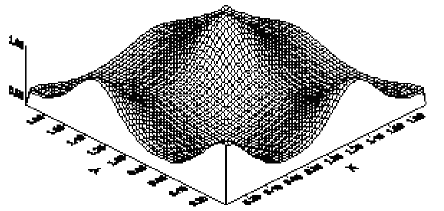


FIG. 12. Flow depth H

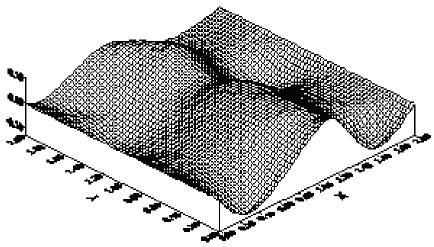


FIG. 13. Flow component Qx

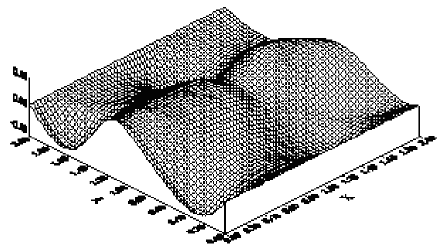


FIG. 14. Flow component Qy

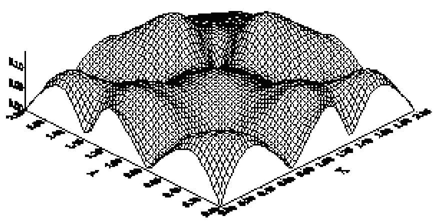


FIG. 15. Module of flow

and depth (Fig. 19). As the water does not outflow, cavities are filled by water. From results we can see, that the filling of a watershed surface by water implements according to isolines.In Fig. 20 is displayed module of flow.

7. CONCLUSIONS

For a selected example with stabilization factor is the laws of conservation of mass and flow of fluids are fair. The obtained model enables to conduct calculations of values of depth and speeds of fluid flows on columbines with rain and lateral influxes for different initial and boundary conditions in different point of time with large values of a Reynold's number.

The above examples indicate that significant influence on the solution of the problem of shallow water on the surface of a watershed has a choice of Reynolds number. For small values of this number of problem can be solved by using numerical scheme (19). Choosing Re> 100, the solution loses its stability (Fig. 10 – Fig. 13). This is because for large values of the Reynolds number solutions of problems may have internal and boundary layers - a very narrow area where most solutions and their gradients change sharply. As a result, numerical solutions, built on the Galerkin scheme, where the parameter discretization is too large to consider all these layers can ostcillate throughout the domain.

Considering it was built stabilization scheme FEM. Applying this scheme to solving problems of shallow water on the surface of a watershed above mentioned problem disappears (Fig.14 - Fig.20).

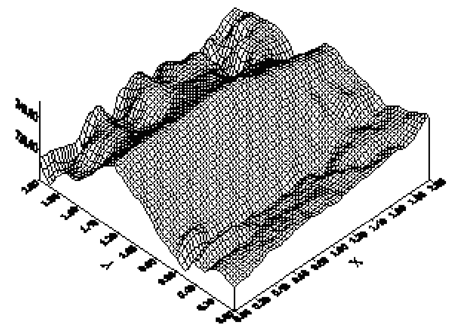


FIG. 16. Bottom surface (x, y)

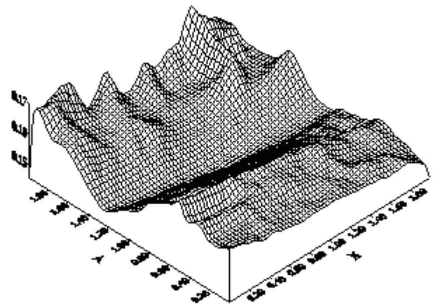


FIG. 17. Flow depth H

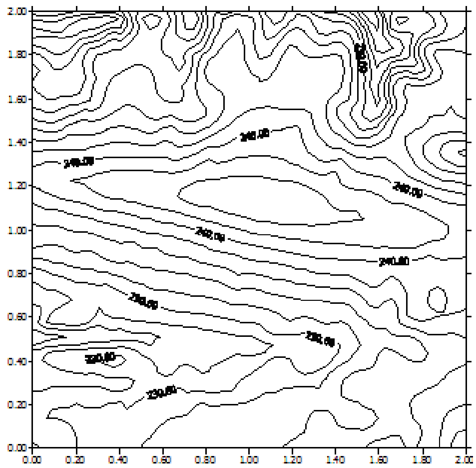


FIG. 18. Isolines of bottom surface η(x,y)

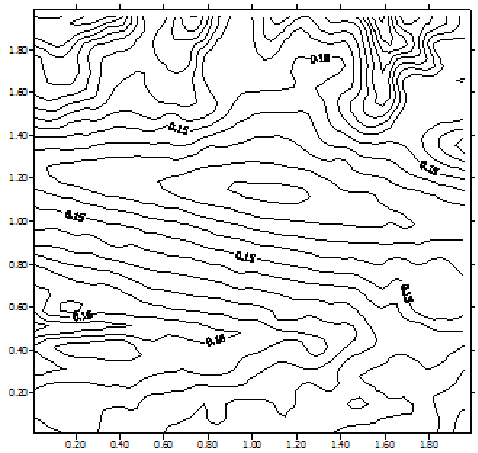


FIG. 19. Isolines of depth surface H(x,y)

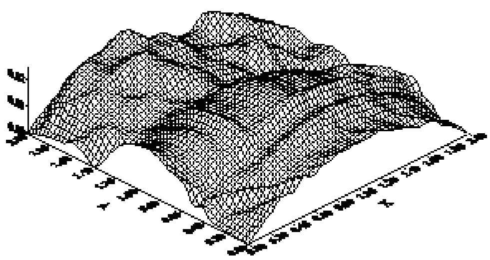


FIG. 20. Module of flow

Thus, based stabilization scheme FEM can be effective in solving the problem of shallow water from any surface water catchment for large Reynolds numbers.

Bibliography

1. Venherskyi P. Mathematical modeling of rain water runoff in the relief area / P. Venherskyi, N. Smushak // Intern. scientific-practical. conf. "GIS technology today": Abstracts ext. – Lviv, 1999. — P. 35.
2. Venherskyi P. Construction of mathematical models of rain water flow on surface watershed / P. Venherskyi, N. Smushak, G. Shinkarenko // Visnyk of Lviv University. A series of mech-mat. – Lviv, 1998. – No 50. – P. 41-45.
3. Venherskyi P. Numerical study of regularized problems of shallow water in kinematic approximation / P. Venherskyi, V. Trushevskyi // Visnyk of Lviv University. A series of App. Math. Inform. – 2003. – No 6. – P. 116-125.
4. Venherskyi P. Numerical modeling of shallow overland flow in the kinematic approximation / P. Venherskyi, V. Trushevskyi // Visnyk of Lviv University. A series of App. Math. Inform. – 1999. – No 1. – P. 44-49.
5. Venherskyi P. Estimation of the depth of recursion of a class of bilateral methods / P. Venherskyi, V. Trushevskyi, P. Senyo // Visnyk of Lviv. University. A series of mechanics and Math. – 1998. – No 50. – P. 45-48.
6. Venherskyi P. Stabilization of the numerical solution of a variational problem of shallow water / P.Venherskyi,V. Trushevskyi, G.Shinkarenko // Visnyk of Lviv University. A series of App. Math. Inform. – 2002. – No 4. – P. 102-109.
7. Venherskyi P. Numerical solution of variational problems of surface flow / P. Venherskyi, V.Trushevskyi, G.Shinkarenko // Visnyk of Kyiv. University. A series of kibernet. – 2002. – No 3. – P. 26-30.
8. Venherskyi P. Applying object-oriented approach for describing algorithms for solving boundary value problems by finite element method / P. Venherskyi, O.Yefremov, B. Stryhaliuk // Visnyk of nat. Univ "Lviv politehnyka." Electronics and telecommunications. — 2002. — No 443. — P. 190-192.
9. Weiyan T. Shallow Water Hydrodynamics: Mathem. Theory Numer. Solution for a Two-dimens. System Shallow Water Equations / T. Weiyan. — Amsterdam: Elsevier, 1992. -– 434 p.
10. Westerink J. J. Tide and Storm Surge Predictions Using a Finite Element Model / J. J. Westerink, R. A. Luettich, A. M. Baptista, N. W. Schefiner, P. Farrar // J. Hydraulic Eng. — 1992. — No 118. — P. 1373-1390.
11. Wunsch O. Numerical simulation of 3d viscous fluid flow and convective mixing in a static mixer / O. Wunsch, G. Bohme // Archive of Applied Mechanics. — 2000. — No 70. — P. 91-102.
12. Yang C.-T. An analysis of flow over a backward-facing step by an assumed stress mixed finite element method. / C.-T. Yang, S. N. Alturi // Numer. Meth. Laminar and Turbulent Flow Proc. 3 Int. Conf. Seattle. — Swansea. — 1983. — P. 302-316.
13. Yang C.-T. An assumed deviatoric stress-pressure-velocity mixed finite element method for unsteady, convective, incompressible viscose flow: Part 2: Computational studies / C.-T. Yang, S. N. Alturi // Int. J. Num. Meth. Fluids. — 1984. -– No 4. — P. 43-69.

Petro Venherskyi, Valerii Trushevskyi,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine

Received 21.05.2014

Журнал обчислювальної та прикладної математики | 2014 № 2 (116) | Journal of Computational & Applied Mathematics

UDC 519.6

EXTENSION OF A CLASS OF NONLINEAR HAMMERSTAIN INTEGRAL EQUATIONS WITH SOLUTIONS REPRESENTED BY COMPLEX POLYNOMIALS

Olena Bulatsyk

Резюме. В роботі розглядається нелінійне інтегральне рівняння типу Гаммерштейна з довільною залежністю від модуля невідомої функції. Розв'язки рівнянь такого типу подаються через поліноми скінчених степенів, параметри яких визначаються із системи, що складається із одного інтегрального і скінченного числа трансцендентних рівнянь. Встановлено існування еквівалентних груп розв'язків нелінійних інтегральних рівнянь, що розглядаються. Одержано необхідні умови для точок галуження і системи рівнянь для їх обчислення. Наведено числові результати для конкретної задачі.

Abstract. An approach, developed before for nonlinear integral Hammerstein equations with the linear dependence on the modulus of unknown function, is generalized to the case of arbitrary differentiable dependency. The approach is based on presentation of the solutions via a complex polynomials of finite degrees. The problem is reduced to a system of integro-transcendental equations. The systems of linear homogeneous equations for the branching points and integro-transcendental equations for the parameters of the solution branches are obtained. Numerical results for a concrete problem are presented.

1. Introduction

Let us consider the nonlinear integral equation of the Hammerstein type

$$\alpha f(\xi) = B[W(|f|)e^{i\arg f}] \equiv \int_a^b K(\xi,\xi')W(|f(\xi')|)\exp(i\arg f(\xi'))d\xi' \qquad (1)$$

with the kernel

$$K(\xi,\xi',c) = \frac{s(\xi)q(\xi') - s(\xi')q(\xi)}{\tau(\xi) - \tau(\xi')} \qquad (2)$$

generated by the linear positive defined integral operator $B : L_2(a,b) \to L_2(a,b)$,

$$(Bg, g) > 0 \qquad (3)$$

for any $g \in L_2(a,b)$;

Key words. Nonlinear integral equation of Hammerstein type, finite-parametric solutions, branching of solutions, phase optimization problem.

$s(\xi)$, $q(\xi)$, $\tau(\xi)$ are real continuous functions such that the function sets $\{\tau^n(\xi)s(\xi)\}$, $\{\tau^n(\xi)q(\xi)\}$ $(n = 0, 1, \ldots)$ are linearly independent;

$W\,(|f(\xi)\,| \in L_2(a, b)$ is a given real piecewise differentiated function.

The general theory of nonlinear integral equations and numerical methods for their solving was intensively developed in recent years (see e.g. [1], [7], [9], [10] and the literature cited there). In previous papers we have considered the nonlinear integral Hammerstein equations without any dependency of the integrand on the modulus of unknown function [11] or with a linear dependency on the modulus [6]. Such types of equations arise in different applications, in particular, in the phase optimization problems of antennas or quasioptical transmitting lines with different restrictions on the solution phase. It was established that the solutions to such equations depend on the finite number of complex parameters which are inverse zeros of polynomials of appropriate degrees (generating polynomials). These parameters are calculated from a system of transcendental equations.

In this paper the approach is generalized to equations with a nonlinear dependence of the integrand on the modulus of unknown function. The results presented here were particularly annonced in [5] and [4].

2. Finite-parametric representation of the solutions

We confine ourselves to the case when the solutions to (1) have no zeros at $\xi \in (a, b)$, and assume that they can be represented in the form

$$f(\xi) = \beta \frac{|f(\xi)| P_N(\tau)}{|P_N(\tau)|}, \tag{4}$$

where β is any complex constant with $|\beta| = 1$ (without loss of generality, we further put $\beta = 1$);

$\tau = \tau(\xi)$, $\tau' = \tau(\xi')$;

$$P_N(\tau) = \prod_{k=1}^{N} (1 - \eta_{Nk}\tau) \tag{5}$$

is a polynomial of a finite degree N with complex pairwise non-conjugated zeros η_{Nk}^{-1}:

$$\eta_{Nk} - \bar{\eta}_{Nm} \neq 0, \quad k, m = 1, 2, ..., N. \tag{6}$$

We call $P_N(\tau)$ as the *generating polynomial.*

It follows from (4) that

$$\exp(i \arg f(\xi)) = \beta \frac{P_N(\tau)}{|P_N(\tau)|}. \tag{7}$$

Introduce the symmetrical polynomial of two real variables

$$R_{N-1}(\tau, \tau') = \frac{2i[P_N(\tau')\bar{P}_N(\tau)]}{(\tau - \tau')} = \sum_{n,m=1}^{N} d_{nm} \tau^{n-1} (\tau')^{m-1} \tag{8}$$

and denote the matrix of its coefficients by $D = \{d_{nm}\}$. The determinant of D equals

$$\det D = (-1)^{[N/2]} \prod_{k,m=1}^{N} (\bar{\eta}_{Nm} - \eta_{Nk}), \tag{9}$$

where the square brackets mean the integer part of the value. This fact follows from the condition 4^0 of the Bezudiant from [8]. Its immediate proof is given in [11]. Due to condition (6), $\det D \neq 0$.

The conditions for the function $f(\xi)$ of the form (4) to be a solution to equation (1) are stated by the following theorem.

Theorem 1. *Let a function $f(\xi)$ of the form (4) have no zeros at $\xi \in [a, b]$. In order that it is a solution to equation (1), it is necessary and sufficient that the parameters η_{Nk} satisfy the following system of the transcendental equations:*

$$|f(\xi)| = \int_a^b K(\xi, \xi')W(|f(\xi')|)\frac{\mathrm{Re}[P_N(\tau')\bar{P}_N(\tau)]}{|P_N(\tau')||P_N(\tau)|}d\xi', \tag{10a}$$

$$\Phi_{Nn}(|f(\xi)|, \eta_{N1}, \eta_{N2}, ... \eta_{NN}) = 0, \quad n = 1, 2, ..., N, \tag{10b}$$

$$\Psi_{Nn}(|f(\xi)|, \eta_{N1}, \eta_{N2}, ... \eta_{NN}) = 0, \quad n = 1, 2, ..., N, \tag{10c}$$

where

$$\Phi_{Nn} = \int_a^b \tau^{n-1} s(\xi)\frac{W(|f(\xi)|)}{|P_N(\tau)|}d\xi, \tag{11a}$$

$$\Psi_{Nn} = \int_a^b \tau^{n-1} q(\xi)\frac{W(|f(\xi)|)}{|P_N(\tau)|}d\xi. \tag{11b}$$

Proof. Necessity. Let function (4) be a solution to equation (1). Substituting (4) into (1) and multiplying the both sides of this equality by $\bar{P}_N(\tau)$, we have

$$\alpha\frac{|f(\xi)|\,|P_N(\tau)|^2}{|P_N(\tau)|} = \bar{P}_N(\tau)\int_a^b K(\xi, \xi')W(|f(\xi)|)\frac{P_N(\tau')}{|P_N(\tau')|}d\xi'. \tag{12}$$

After dividing both its sides by $|P_N(\tau)|$ this equation becomes of the form (10a). On the other hand, after taking the imaginary part from the same result, we have

$$\int_a^b \frac{[s(\xi)q(\xi') - s(\xi')q(\xi)]R_{N-1}(\tau, \tau')}{|P_N(\tau')|}W(|f(\xi)|) \equiv 0. \tag{13}$$

Then, substituting (8) into (13) with interchanging the variables ξ and ξ', we have

$$\sum_{n,m=1}^{N} d_{nm}\left[q(\xi')\int_a^b \frac{\tau^{n-1}s(\xi)W(|f(\xi)|)}{|P_N(\tau)|}d\xi - \right. \\ \left. - s(\xi')\int_a^b \frac{\tau^{n-1}q(\xi)W(|f(\xi)|)}{|P_N(\tau)|}d\xi\right](\tau')^{m-1} \equiv 0. \quad (14)$$

Since the functions $\{\tau^n s\}$, $\{\tau^n q\}$, $n = 0, \ldots, N-1$, are linearly independent, (14) gives

$$\sum_{n=1}^{N} d_{nm}\int_a^b \frac{\tau^{n-1}s\,(\xi)\,W(|f(\xi)|)}{|P_N(\tau)|}d\xi \;=\; 0, \quad n = 1, 2, ..., N, \quad (15a)$$

$$\sum_{n=1}^{N} d_{nm}\int_a^b \frac{\tau^{n-1}q\,(\xi)\,W(|f(\xi)|)}{|P_N(\tau)|}d\xi \;=\; 0, \quad n = 1, 2, ..., N. \quad (15b)$$

Equalities (15) can be considered as two independent systems of linear algebraic equations with respect to the unknown integrals. The determinant of their common matrix D does not equal zero owing to conditions (6), so that the systems have only zero solutions, that is, the transcendental equations (10) are satisfied.

Sufficiency. Let (10) hold at a certain integer N and complex η_{Nk}, $k = 1, 2, ..., N$, satisfying conditions (6). Then, of course, equalities (15) are satisfied, too, and, hence, the identities (14) and (13) hold as well. With the aid of (8), we obtain from (13)

$$\mathrm{Im}\left[\bar{P}_N(\tau)\int_a^b K(\xi,\xi')\frac{W(|f(\xi')|)}{|P_N(\tau')|}P_N(\tau')d\xi'\right] = 0 \quad (16)$$

or, after adding the real function $\alpha|f(\xi)|\,|P_N(\tau)|$ under the imaginary sign,

$$\mathrm{Im}\left[\alpha|f(\xi)|\,|P_N(\tau)| + \bar{P}_N(\tau)\int_a^b K(\xi,\xi')\frac{W(|f(\xi')|)}{|P_N(\tau')|}P_N(\tau')d\xi'\right] = 0. \quad (17)$$

Dividing the both sides of (17) by the real positive function $|P_N(\tau)|$, we obtain

$$\mathrm{Im}\left[\alpha|f(\xi)| + \frac{\bar{P}_N(\tau)}{|P_N(\tau)|}\int_a^b K(\xi,\xi')\frac{W(|f(\xi')|)}{|P_N(\tau')|}P_N(\tau')d\xi'\right] = 0. \quad (18)$$

On the other hand, integral equation (10a) can be written in the form

$$\mathrm{Re}\left[\alpha\,|f(\xi)| + \frac{\bar{P}_N(\tau)}{|P_N(\tau)|}\int_a^b K(\xi,\xi')\frac{W(|f(\xi')|)}{|P_N(\tau')|}P_N(\tau')d\xi'\right] = 0. \quad (19)$$

Equalities (18) and (19) together imply that the expression in their square brackets equals zero, that is, function (4) solves integral equation (1).

End of proof.

Theorem 2. *If the function $f(\xi)$ of the form (4) with $\beta = 1$ solves equation (1), then the functions*

$$f_n(\xi) = \frac{|f(\xi)|P_N(\tau)}{|P_N(\tau)|}\frac{1-\bar{\eta}_{Nn}\tau}{1-\eta_{Nn}\tau}, \quad n = 1, 2, ..., N,$$

solve this equation, too.

Proof. The proof of this theorem is analogous to the proof of the Theorem 2.2 in [6] with substitution $W(|f(\xi)|) = F(\xi) - |f(\xi)|$.

In the simplest case, the theorem is complitely adjusted with the obvious property that if the function $f(\xi)$ solves equation (1), then $\bar{f}(\xi)$ solves this equation, too.

Corollary 1. *The solutions to integral equation (10a) and the system of transcendental equations (10b,10c) make up the equivalent groups inside which the function $|f(\xi)|$ remains the same and the polynomials $P_N(\tau)$ differ only by substitution of any number $s < N$ of the parameters η_k by the complex conjugated ones:*

$$P_N^{(s)}(\tau) = \prod_{m=1}^{s}(1-\eta_{n_m}\tau)\prod_{m=s+1}^{N}(1-\bar{\eta}_{n_m}\tau),$$

where $n_{m_1} \neq n_{m_2}$ if $m_1 \neq m_2$. Such polynomials generate the solutions to (1) with the same $|f(\xi)|$.

Corollary 2. *If there is a solution to equation (1) with two parameters $\eta_1 = -\eta_2$ in the polynomial P_N, which give an even polynomial argument addend, then a solution exists in the same equivalent group, which has an odd argument. In particular, if all parameters of the polynomial P_N can be devided into such symmetrical pairs, what means that the polynomial argument is an even function, then another solution exists in the same equivalent group, which have an odd argument.*

This corollary is justyfied by the following logical considerations. The argument of the factor $p_1(\tau) = (1-\eta_1\tau)(1-\eta_2\tau) = 1-\eta_1^2\tau^2$ is obviously the even function of τ. Substituting η_2 with $\overline{\eta}_2$ according to above theorem gives the factor $p_2(\tau) = (1-\eta_1\tau)(1+\overline{\eta}_1\tau) = 1-|\eta_1|^2\tau^2-(\eta_1-\overline{\eta}_1)\tau$. Its argument is :

$$\arg p_2 = \arctan\frac{2\mathrm{Im}\eta_1\tau}{1-|\eta_1|^2\tau^2}.$$

If N is even integer and $P_N(\tau) = \prod\limits_{n=1}^{N/2}(1-\eta_n^2\tau^2)$, then

$$\widetilde{P}_N(\tau) = \prod_{n=1}^{N/2}(1-|\eta_n|^2\tau^2-(\eta_n-\overline{\eta}_n)\tau)$$

and its argument is

$$\arg \widetilde{P}_N(\tau) = \sum_{n=1}^{N/2} \arctan \frac{2\mathrm{Im}\eta_n \tau}{1 - |\eta_n|^2 \tau^2},$$

which is the odd function of τ.

3. Branching of solutions

For $N = 0$ (real positive solutions) the transcendental equations (10b, 10c) disappear and the only integral equation (10a) remains, which coincides with (1), in which $f(\xi)$ must be substituted by $|f(\xi)|$. This equation has the non-trivial solution but not for all values c and N.

The number of solutions to (1) may change at some values $c = c_j$. Such values are called the branching points. The branching points of solutions to equation (1) are found from the condition that the system of the homogeneous integral equations

$$\begin{aligned} \lambda_n w_n |f| &= B\left[W(|f|) \frac{\mathrm{Im}\left(\bar{P}_N(\tau')P_N(\tau)\right)}{|P_N(\tau')|\,|P_N(\tau)|} v_n + \right. \qquad (20a) \\ &\left. + W'(|f|)|f| \frac{\mathrm{Re}\left(\bar{P}_N(\tau')P_N(\tau)\right)}{|P_N(\tau')|\,|P_N(\tau)|} w_n \right], \\ \lambda_n v_n |f| &= B\left[W(|f|) \frac{\mathrm{Re}\left(\bar{P}_N(\tau)P_N(\tau')\right)}{|P_N(\tau')|\,|P_N(\tau)|} v_n + \right. \qquad (20b) \\ &\left. + W'(|f|)|f| \frac{\mathrm{Im}\left(\bar{P}_N(\tau)P_N(\tau')\right)}{|P_N(\tau')|\,|P_N(\tau)|} w_n \right] \end{aligned}$$

has multiple eigenvalues $\lambda_n = 1$. Here $\{w_n, v_n\}$ are vector-functions; $W' = dW/d(|f|)$. It is easy to check that $\lambda_1 = 1$, $\{v_1 \equiv 1,\ w_1 \equiv 0\}$ is always the eigenpair of (20). These equations are obtained by application of usual pertrubations technique to equation (1) (see e.g. [6]).

There is an obvious way to obtain the transcendental equation system for calculation of the branching points and the polynomial parameters in them. As a rule, the branching of solutions to equation (1) is caused by changing the degree N of the polynomial P_N by one. At the branching points the parameters η_{Nk} of the initial polynomial P_N and parameters $\eta_{N+1,k}$ of the branched polynomial P_{N+1} are connected by the equalities

$$\begin{aligned} \frac{P_N(\tau)}{|P_N(\tau)|} = \frac{P_{N+1}(\tau)}{|P_{N+1}(\tau)|}, \quad \eta_{Nk} = \eta_{N+1,k}, \qquad k = 1, 2, \ldots, N, \\ \mathrm{Im}\eta_{N+1,N+1} = 0. \end{aligned} \qquad (21)$$

At the branching points two new unknown c_0 and $\mathrm{Re}\eta_{N+1,N+1}$ occur. Besides (10), system

$$|f(\xi)| = \int_a^b K(\xi,\xi')W(|f(\xi')|)\frac{\mathrm{Re}[P_{N+1}(\tau')\bar{P}_{N+1}(\tau)]}{|P_{N+1}(\tau')||P_{N+1}(\tau)|}d\xi', \tag{22a}$$

$$\int_a^b \tau^{n-1}s(\xi)\frac{W(|f(\xi)|)}{|P_{N+1}(\tau)|}d\xi = 0, \quad n = 1,2,...,N+1, \tag{22b}$$

$$\int_a^b \tau^{n-1}s(\xi)\frac{W(|f(\xi)|)}{|P_{N+1}(\tau)|}d\xi = 0, \quad n = 1,2,...,N+1 \tag{22c}$$

should hold. Since the new parameter $\eta_{N+1,N+1}$ is real, the integral equation (22a) coincides with (10a) and the kth equation of system (22b) (22c) $k = 1,2,\ldots,N$, is a linear combination of the corresponding equation of system (10b), (10c) and $(k+1)$th equation of (22b) (22c). Hence, at the branching point, besides system (10) only two additional equations

$$\int_a^b \tau^N s(\xi)\frac{F(\xi)-\beta|f(\xi)|}{|P_N(\tau)|(1-\eta_{N+1,N+1}\tau)}d\xi = 0, \tag{23a}$$

$$\int_a^b \tau^N q(\xi)\frac{F(\xi)-\beta|f(\xi)|}{|P_N(\tau)|(1-\eta_{N+1,N+1}\tau)}d\xi = 0 \tag{23b}$$

should hold. On the whole, we have one real integral equation and $2N+2$ transcendental ones for determining the real function $|f(\xi)|$, N complex parameters η_{Nk}, $k = 1,2,\ldots,N$ and real $\eta_{N+1,N+1}$ and c_j.

At the branching points where the polynomial degree changes by two, the equalities

$$\eta_{Nk} = \eta_{N+2,k}, k = 1,\ldots,N \tag{24}$$

are valid. Besides (10), the four additional equations

$$\begin{aligned}&\int_a^b \frac{\tau^{k-1}s(\xi)W(|f(\xi)|)}{|P_N(\tau)|(1-\eta_{N+2,N+1}\tau)(1-\eta_{N+2,N+2}\tau)}d\xi=0,\ n = N+1, N+2;\\&\int_a^b \frac{\tau^{k-1}q(\xi)W(|f(\xi)|)}{|P_N(\tau)|(1-\eta_{N+2,N+1}\tau)(1-\eta_{N+2,N+2}\tau)}d\xi=0,\ n = N+1, N+2,\end{aligned} \tag{25}$$

should be fulfilled with $\eta_{N+2,N+1}$, $\eta_{N+2,N+2}$ satisfying the conditions

$$\eta_{N+2,N+1} = \bar{\eta}_{N+2,N+2} \tag{26}$$

or

$$\mathrm{Im}\eta_{N+2,N+1} = \mathrm{Im}\eta_{N+2,N+2} = 0. \tag{27}$$

Hence, we have $2N+5$ equation for $2N+4$ real unknown: N complex η_{Nk}, $n = 1,2,\ldots,N$, one real c_j, one complex $\eta_{N+2,N+1}$ or two real $\eta_{N+2,N+1}$,

$\eta_{N+2,N+2}$, and $|f(\xi)|$. As it was mentioned in the preceding subsection, the existence of solutions to such a system is low-probable in general case. However, they may exist in the case when

$$W(|f(\xi)|) = W(|f(-\xi)|). \tag{28}$$

Then the solutions are possible, which generate the polynomials with the even modulus

$$|P_N(\tau)| = |P_N(-\tau)|, \qquad |P_{N+2}(\tau)| = |P_{N+2}(-\tau)|. \tag{29}$$

This equality decreases the number of unknowns twice: the parameters $\eta_{N+2,k}$ become imaginary or appear by couples with opposite signs and $\eta_{N+2,n}$, $n = N+1, N+2$ are always imaginary with opposite signs:

$$\mathrm{Re}\eta_{N+2,N+1} = \mathrm{Re}\eta_{N+2,N+2} = 0, \tag{30a}$$

$$\eta_{N+2,N+1} = \bar{\eta}_{N+2,N+2}. \tag{30b}$$

On the other hand, conditions (29) decrease the number of equations twice, as well: N equations of system (10b), (10c) and two additional equations (25) become identities, because they have odd integrands in the left-hand side.

Finally, at fulfilling (28), (29) the solution branching is possible with decreasing the polynomial degree by two if the following transcendental equation system holds:

$$\int_a^b \tau^{2n-1} s(\xi) \frac{W(|f(\xi)|)}{|P_N(\tau)|} d\xi = 0, \quad n = 1, 2, \ldots [N/2], \tag{31a}$$

$$\int_a^b \tau^{2n-2} q(\xi) \frac{W(|f(\xi)|)}{|P_N(\tau)|} d\xi = 0, \quad n = 1, 2, \ldots [(N+1)/2], \tag{31b}$$

$$\int_a^b \frac{\tau^{2[(N+2)/2]-1} s(\xi) W(|f(\xi)|)}{|P_N(\tau)|(1-\eta_{N+2,N+1}\tau)(1-\eta_{N+2,N+2}\tau)} d\xi = 0, \tag{31c}$$

$$\int_a^b \frac{\tau^{2[(N+1)/2]} q(\xi) W(|f(\xi)|)}{|P_N(\tau)|(1-\eta_{N+2,N+1}\tau)(1-\eta_{N+2,N+2}\tau)} d\xi = 0, \tag{31d}$$

where η_{Nk}, $k = 1, \ldots, N$, are either imaginary or appear by couples with alternative signs, and $\eta_{N+2,k}$, $k = N+1, N+2$ are subject to conditions (30). As a result, we have $N+3$ real equations with respect to $N+3$ real unknowns.

4. Numerical results

As an example, we show the numerical results obtained for $W(|f(\xi)|) = 1/2 - |f(\xi)|^2$ and $\alpha = 0.5$. This problem arises in the case when the linear antenna should create the uniform power pattern $F^2 \equiv 1/2$. The calculations were carried out by the Newton method.

The real and imaginary parts of η_{Nk} are shown in Fig. 1. The real parts of solutions are drawn by the dashed lines, the imaginary ones – by the solid lines. The curve numbering corresponds to the indexes Nk at these parameters.

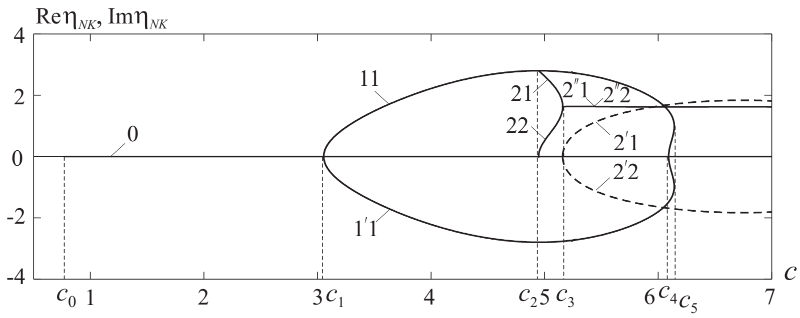


FIG. 1. Real and imag parts of parametrs η_{Nk};
$W(|f(\xi)|) = 1/2 - |f(\xi)|^2$, $\alpha = 0.5$

For $c < c_0 = 0.84$ there are no nontrivial solutions to equation (1) at this α. At $c = c_0$ the solution $f_0(\xi)$ with $N = 0$ arises (curve 0). It starts from $f_0(\xi) \equiv 0$.

At the point $c_1 = 3.05$ two complex conjugate solutions $f_1(\xi)$, $f_{1'}(\xi)$ with $N = 1$ and imaginary η_{11}, $\eta_{1'1}$ respectively, branch off from $f_0(\xi)$ (curves 11, $1'1$). At the point $c_2 = 4.95$, two solutions with $N = 2$ branch off from each solution with $N = 1$. All they make up an equivalent group; we analyze only one of them denoted by $f_2(\xi)$. The solutions $f_1(\xi)$, $f_{1'}(\xi)$ continue to exist. Two more characteristic points, related to them, are c_4 and c_5.

The solution $f_2(\xi)$, arising at $c = c_2$ has two imaginary parameters η_{21}, η_{22} (curves $21, 22$). At $c_3 = 5.16$ the solution $f_2(\xi)$ transforms into $f_{2'}(\xi)$, which has two complex parameters η'_{21}, η'_{22} with $\mathrm{Re}\eta'_{22} = -\mathrm{Re}\eta'_{21}$, $\mathrm{Im}\eta'_{21} = \mathrm{Im}\eta'_{22}$. Curves $2'1$, $2''1$, correspond to $\mathrm{Re}\eta'_{21}$, $\mathrm{Im}\eta'_{21}$ and curves $2'2$, $2''2$, – to $\mathrm{Re}\eta'_{22}$, $\mathrm{Im}\eta'_{22}$, respectively.

When c increases, the solutions with larger N appear, similarly as in the problem of antenna synthesis according to the amplitude pattern [3].

BIBLIOGRAPHY

1. Bauschke H. H. Phase retrieval, error reduction algorithm, and Fienup variants: A view from convex optimization / H. H. Bauschke, P. L. Combettes, D. R. Luke // J. of Optical Society of America. – 2002. – Vol. 19, No 7. – P. 1334-13345.
2. Bulatsyk O. O. Phase Optimization Problems / O. O. Bulatsyk, B. Z. Katsenelenbaum, Yu. P. Topolyuk, N. N. Voitovich. – Weinheim: WILEY-VCH, 2010. – 301 p.
3. Bulatsyk O. O. Properties of nonlinear Hammerstein integral equations connected with modified phase problem / O. O. Bulatsyk, N. N. Voitovich // Direct and Inverse Problems of Electromagnetic and Acoustic Wave Theory (DIPED-2003). – Proc. of Int. Seminar/Workshop, Lviv. – 2003. – P. 135-138.
4. Bulatsyk O. O. Complex polynomials representation of solutions to the antenna synthesis problem by the power pattern / O. O. Bulatsyk // Direct and Inverse Problems of Electromagnetic and Acoustic Wave Theory (DIPED-2013). – Proc. of Int. Seminar/Workshop, Lviv-Tbilisi. – 2013. – P. 166-170.

5. Bulatsyk O. Extension of the class of nonlinear integral equations which permit to present solutions by the complex polynomials / O. Bulatsyk // Modern Problems of Mechanics and Mathematics. Lviv, IAPMM NASU. – 2013. – Vol. 3. – P. 16-18.
6. Bulatsyk O. O. Finite-parametric solutions of a class of Hammerstein nonlinear integral equations / O. O. Bulatsyk, Yu. P. Topolyuk, N. N. Voitovich // J. Numer Appl. Math. – 2010. – Vol. 1 (100). – P. 10-28.
7. Fienup J. R. Phase retrieval algorithm: a comparison 2. / J. R. Fienup // Appl. Opt. – 1982. – Vol. 21. – P. 2758-2769.
8. Lander F. J. The Bezoitiant and inversion of Hankel and Toeplitz matrices / F. J. Lander // Matematicheskie Issledovaniya. – 1974. – Vol. 2. – P. 69-87 (in Russian).
9. Masujima M. Applied Mathematical Methods in Theoretical Physics / M. Masujima. – Weinheim: WILEY- VCH, 2005. – 377 p.
10. Precup R. Methods in Nonlinear Integral Equation / R. Precup. – Dordrecht: Kluver Acad. Publ, 2002. – 232 p.
11. Voitovich N. N. Approximation of compactly supported functions with free phase by functions with bounded spectrum / N. N. Voitovich, Y. P. Topolyuk, O. O. Reshnyak // Fields Institute Communications. – 2000. – Vol. 25. – P. 531-541.

Olena Bulatsyk,
Pidstrygach Institute for Applide Problems
of Mechanics and Mathematics,
3-b Naukova Str., Lviv, 79060, Ukraine

Received 20.05.2014

Taras Shevchenko National University of Kyiv

Journal of Computational and Applied

2014 MATHEMATICS 2 (116)

Founded in 1965

C O N T E N T S