Київський національний університет імені Тараса Шевченка

Журнал обчислювальної та прикладної

2017 **МАТЕМАТИКИ** 3 (126)

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 20 листопада 2017 р. (протокол № 4)

Свідоцтво про державну реєстрацію КВ 4246 від 26.05.2000

Підписано до друку 20 листопада 2017 р.

Журнал обчислювальної та прикладної математики | 2017 № 3 (126) | Journal of Numerical & Applied Mathematics

UDC 519.6

AN ITERATIVE METHOD FOR THE CAUCHY PROBLEM FOR THE LAPLACE EQUATION IN THREE-DIMENSIONAL DOMAINS

I. V. Borachok

Резюме. Ми розглядаємо ітераційний узагальнений метод Ландвебера для задачі Коші для рівняння Лапласа у двозв'язних тривимірних областях. Цей метод є регуляризуючою процедурою для отримання стабільного розв'язку. На кожному кроці ітераційного методу потрібно розв'язати дві коректні прямі задачі для рівняння Лапласа. Кожна пряма задача вирішується методом граничних інтегральних рівнянь із застосуванням проекційного методу Гальоркіна для дискретизації. Наприкінці наведені деякі чисельні результати.

Abstract. We consider an iterative generalized Landweber method for the Cauchy problem for the Laplace equation in doubly connected 3-dimensional domains. This method is a regularizing procedure for obtaining a stable solution to the Cauchy problem, and consists of solving two well-posed direct problems for the Laplace equation at each iteration step. Each direct problem is solved by a boundary integral equations method with a projection Galerkin method for the discretisation. Some numerical results are given and discussed as well at the end.

1. Introduction

The Cauchy problem for the Laplace equation has important applications. For example, it occurs in electrostatics, non-destructive testing, cardiology, leak identification, etc. This problem belongs to the class of ill-posed linear inverse problems, since it is unstable with respect to input data [7] (a small remark here, the input Cauchy data should be compatible [6]). We focus on the numerical solution of this Cauchy problem in three-dimensional doubly connected domains.

The Cauchy problem can be solved numerically in a stable way by combining direct methods, such as for example the boundary integral equations method [4,10–12,15] or the method of fundamental solutions [14] etc, with some regularization strategy, for example, Tikhonov regularization with an appropriate way of selecting the regularization parameter like the Morozov discrepancy principle or the L-curve method [4, 12, 15]. Another approach for numerically solving the Cauchy problem is to use iterative methods, where the choice of the termination of the iterations is part of the regularization. Numerical examples show that iterative methods give good results in the case of noisy data, namely,

Key words. Laplace equation, Cauchy problem, Landweber method, Robin boundary problem, boundary integral equations, projection Galerkin method, Wienert's method, R^3 domains.

we can calculate an approximation with an error being equal to the noise level or even smaller, by selecting a good strategy for the numerical implementation of the iterative approach. Commonly used methods are the alternating method [5, 8, 12] and the Landweber procedure [12] in combination with the boundary integral equations method for solving the direct problems needed in both these iterative algorithms.

In this paper, we apply one recent approach being a generalized Landweber method proposed in [2], for 3-dimensional doubly connected domains. The main difference from the standard Landweber method is that we do not need to use any adjoint operator, that is we do not need to involve any adjoint differential equation.

We then describe more on the problem formulation. Let $D_1 \subset \mathbb{R}^3$, $D_2 \subset \mathbb{R}^3$ be simply connected smooth bounded domains with boundary surfaces Γ_1 and Γ_2, respectively, that satisfy: $\overline{D}_1 \subset D_2$. Let $D = D_2 \setminus \overline{D}_1$ be the solution domain and $\nu = (\nu_1, \nu_2, \nu_3)^t$ the outward unit normal to the boundary of D; this boundary is denoted by $\partial D = \Gamma_1 \cup \Gamma_2$.

The Cauchy problem is then as follows. We need to find a classical solution $u \in C^2(D) \cap C^1(\overline{D})$ of the Laplace equation:

$$\Delta u = 0 \quad \text{in } D \tag{1}$$

that satisfies the boundary conditions:

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

It is not the full solution in D that is of prime interest, it is instead to find (reconstruct) the corresponding Cauchy data $\left\{ u, \dfrac{\partial u}{\partial \nu} \right\}$ on the interior boundary surface Γ_1.

As mentioned, for the numerical solution of the above problem, we apply one adjoint-free Landweber method [2] being a regularizing procedure for obtaining a stable numerical solution [2]. At each step of the iterative procedure, we need to solve the Dirichlet respectively the Robin direct problems for the Laplace equation. We use the boundary integral equations method for solving the required direct problems in the iterative method, and this choice is based on good numerical results for domains in $\mathbb{R}^2$, see [11, 12] as well as for domains in $\mathbb{R}^3$ [4, 5], together with advantages such as reduction of the dimension of the problem and the flexibility in terms of the form of the boundary surfaces. As a stopping rule for the iterations, the Morozov discrepancy principle is used.

The solution of each direct problem is represented as a combination of potentials [4, 9, 12]. Based on this representation, we obtain a system of linear integral equations for finding the unknown densities by requiring that the given Cauchy data should be satisfied. For discretization Wienert's method is applied; it is a Galerkin discrete projection method, where the unknown densities are represented as a linear combination of spherical harmonics [1] and the boundary integrals are rewritten over the unit sphere, and to those obtained integrals certain cubature rules are then applied [13, 16].

An outline of this work is: in Section 2, we consider the iterative algorithm, the boundary integral equations method for one of the direct problems in the procedure (having boundary conditions of Robin type) is given in Section 3 and in Section 4 some numerical results are shown and discussed.

2. The iterative algorithm

We consider one of the iterative methods proposed in [2], in three-dimensional doubly connected domains. At each iteration step, we need to solve one Dirichlet and one Robin boundary value problem for the Laplace equation. The algorithm is as follows:

- The first approximation u_0 of the solution u is calculated by solving the Dirichlet boundary value problem:

$$\Delta u_0 = 0 \quad \text{in } D, \tag{3}$$

$$u_0 = \eta_0 \text{ on } \Gamma_1 \quad \text{and} \quad u_0 = f \quad \text{on } \Gamma_2, \tag{4}$$

  where η_0 is an arbitrary initial starting approximation on the boundary Γ_1.
- Then the element v_0 is obtained by solving the Robin boundary value problem:

$$\Delta v_0 = 0 \quad \text{in } D, \tag{5}$$

$$\frac{\partial v_0}{\partial \nu} + \kappa v_0 = 0 \text{ on } \Gamma_1 \quad \text{and} \quad \frac{\partial v_0}{\partial \nu} + \kappa v_0 = g - \frac{\partial u_0}{\partial \nu} \quad \text{on } \Gamma_2. \tag{6}$$

- Having obtained u_{k-1} and v_{k-1}, the approximation u_k is obtained from the Dirichlet boundary value problem:

$$\Delta u_k = 0 \quad \text{in } D, \tag{7}$$

$$u_k = \eta_k \text{ on } \Gamma_1 \quad \text{and} \quad u_k = f \quad \text{on } \Gamma_2, \tag{8}$$

  where

$$\eta_k = \eta_{k-1} + \gamma v_{k-1}|_{\Gamma_1}, \quad \gamma > 0. \tag{9}$$

- Then the solution v_k is obtained by solving the following Robin boundary value problem:

$$\Delta v_k = 0 \quad \text{in } D, \tag{10}$$

$$\frac{\partial v_k}{\partial \nu} + \kappa v_k = 0 \text{ on } \Gamma_1 \quad \text{and} \quad \frac{\partial v_k}{\partial \nu} + \kappa v_k = g - \frac{\partial u_k}{\partial \nu} \quad \text{on } \Gamma_2. \tag{11}$$

The iterative procedure then continues by iterating in the last two steps. The stopping rule is the Morozov discrepancy principle. The initial approximation is arbitrary for linear problems, and we select it as the zero-function.

The parameter κ in the Robin boundary condition is positive: $\kappa > 0$. The parameter $\gamma > 0$ in the iterative procedure is a relaxation parameter, which is needed for convergence of the algorithm [2].

The Dirichlet and Robin boundary value problems are well-posed in $L^2(D)$ for boundary data from $L^2(\Gamma_1)$ and $L^2(\Gamma_2)$. Moreover, given $f, g \in L^2(\Gamma_2)$ one can show that $\lim\limits_{k\to\infty} \|u - u_k\|_{L^2(D)} = 0$, where u_k is the k-th approximation generated from the above algorithm and u is the solution of the Cauchy problem (1)–(2). Furthermore, for noisy data $\{f^\delta, g^\delta\}$, with $\delta > 0$, we have

$\left\| f^\delta - u_k^\delta \right\|_{L^2(\Gamma_2)} \leq \tau\delta$, for $\tau > 1$, where u_k^δ is the k-th approximation obtained from the iterative algorithm using the noisy data. For further information and details on these estimates, see [2].

3. Numerical solution of the boundary value problems

To solve each of the boundary value problems used in the iterative procedure, we use the boundary integral equations method. In the introduction, we mentioned some advantages of this approach such as reducing the dimension of the problem compared with the dimension of the solution domain, the flexibility of applying it for domains of different shapes or even to unbounded domains, its super-algebraic convergence for analytical data etc.

In [3], it is demonstrated how to solve the Dirichlet boundary value problem using a single-layer representation of the solution. The similar ideas can be applied to the Dirichlet boundary value problem by instead using a combination of single- and double-layer potentials to represent the solution, thereby obtaining an integral equation of the second kind to solve [9].

We then turn to the Robin boundary value problem:

$$\Delta u = 0 \quad \text{in } D, \tag{12}$$

$$\frac{\partial u}{\partial \nu} + \kappa u = h \text{ on } \Gamma_1 \quad \text{and} \quad \frac{\partial u}{\partial \nu} + \kappa u = w \quad \text{on } \Gamma_2, \tag{13}$$

where $\kappa > 0$, $h \in L_2(\Gamma_1)$, $w \in L_2(\Gamma_2)$ are given.

To obtain an integral equation of the second kind, we represent the solution of (12)–(13) as a sum of two single-layer potentials:

$$u(x) = \sum_{l=1}^{2} \int_{\Gamma_l} \varphi_l(y)\Phi(x,y)\, ds(y), \quad x \in D, \tag{14}$$

where $\Phi(x,y) = \dfrac{1}{4\pi|x-y|}$ is a fundamental solution of the Laplace equation in $\mathbb{R}^3$ and $\varphi_l \in C(\Gamma_l)$, $l = 1, 2$, are unknown densities.

From the representation of the solution (14) requiring the boundary conditions (13) to be satisfied, invoking properties of single-layer potentials [9], we obtain a system of linear integral equations for finding the unknown densities:

$$\begin{cases} -\dfrac{1}{2}\varphi_1 + K_{11}\varphi_1 + K_{12}\varphi_2 + \kappa\left(S_{11}\varphi_1 + S_{12}\varphi_2\right) = h, & \text{on } \Gamma_1, \\ \dfrac{1}{2}\varphi_2 + K_{21}\varphi_1 + K_{22}\varphi_2 + \kappa\left(S_{21}\varphi_1 + S_{22}\varphi_2\right) = w, & \text{on } \Gamma_2, \end{cases} \tag{15}$$

where we used the following boundary integral operators for $l, r = 1, 2$:

$$(S_{lr}\psi)(x) = \int_{\Gamma_r} \psi(y)\Phi(x,y)\, ds(y),\ x \in \Gamma_l,\ \psi \in C(\Gamma_r), \tag{16}$$

$$(K_{lr}\psi)(x) = \int_{\Gamma_r} \psi(y)\frac{\partial\Phi(x,y)}{\partial\nu(x)}\, ds(y),\ x \in \Gamma_l,\ \psi \in C(\Gamma_r). \tag{17}$$

Notice here that for the Robin boundary problem the approximation of the solution on the internal boundary surface Γ_1 needed in the above generalized Landweber algorithm, can be obtained as

$$u(x) = (S_{11}\varphi_1)(x) + (S_{12}\varphi_2)(x), \quad x \in \Gamma_1. \tag{18}$$

We assume that the two boundary surfaces can be smoothly mapped one-to-one to the unit sphere $\mathbb{S}^2 = \left\{\widehat{x} \in \mathbb{R}^3 : |\widehat{x}| = 1\right\}$. In that case there exist one-to-one mappings $q_l : \mathbb{S}^2 \to \Gamma_l$, $l = 1, 2$, having smoothly varying Jacobian J_{q_l}, $l = 1, 2$. Therefore, based on (16) and (17), we can rewrite the system of integral equations (15) over the unit sphere:

$$\begin{cases} -\dfrac{1}{2}\phi_1 + \tilde{K}_{11}\phi_1 + \tilde{K}_{12}\phi_2 + \kappa\left(\tilde{S}_{11}\phi_1 + \tilde{S}_{12}\phi_2\right) = \tilde{h}, & \text{on } \mathbb{S}^2, \\ \dfrac{1}{2}\phi_2 + \tilde{K}_{21}\phi_1 + \tilde{K}_{22}\phi_2 + \kappa\left(\tilde{S}_{21}\phi_1 + \tilde{S}_{22}\phi_2\right) = \tilde{w}, & \text{on } \mathbb{S}^2, \end{cases} \tag{19}$$

where $\phi_l(\widehat{x}) = \varphi_l(q_l(\widehat{x}))$, $l = 1, 2$, $\tilde{h}(\widehat{x}) = h(q_1(\widehat{x}))$, $\tilde{w}(\widehat{x}) = w(q_2(\widehat{x}))$ for $\widehat{x} \in \mathbb{S}^2$ and the parametrised integral operators are for $l, r = 1, 2$:

$$(\tilde{S}_{lr}\psi)(\widehat{x}) = \int_{\mathbb{S}^2} \psi(\widehat{y}) L_{lr}(\widehat{x}, \widehat{y})\, ds(y), \quad \psi(\widehat{x}) \in C(\mathbb{S}^2),\ \widehat{x} \in \mathbb{S}^2, \tag{20}$$

and

$$(\tilde{K}_{lr}\psi)(\widehat{x}) = \int_{\mathbb{S}^2} \psi(\widehat{y}) M_{lr}(\widehat{x}, \widehat{y})\, ds(y), \quad \psi(\widehat{x}) \in C(\mathbb{S}^2),\ \widehat{x} \in \mathbb{S}^2, \tag{21}$$

with

$$L_{lr}(\widehat{x}, \widehat{y}) = \begin{cases} J_{q_r}(\widehat{y})\Phi(q_l(\widehat{x}), q_r(\widehat{y})), & l \neq r, \\ \dfrac{R_l(\widehat{x}, \widehat{y})}{|\widehat{x} - \widehat{y}|}, & l = r, \end{cases}$$

$$M_{lr}(\widehat{x}, \widehat{y}) = \begin{cases} -J_{q_r}(\widehat{y})\dfrac{(q_l(\widehat{x}) - q_r(\widehat{y}))^T \nu(q_l(\widehat{x})))}{4\pi|q_l(\widehat{x}) - q_r(\widehat{y})|^3}, & l \neq r, \\ \dfrac{\tilde{R}_l(\widehat{x}, \widehat{y})}{|\widehat{x} - \widehat{y}|}, & l = r, \end{cases}$$

where

$$R_l(\widehat{x}, \widehat{y}) = \frac{J_{q_l}(\widehat{y})}{4\pi} \begin{cases} \dfrac{|\widehat{x} - \widehat{y}|}{|q_l(\widehat{x}) - q_l(\widehat{y})|}, & \widehat{x} \neq \widehat{y} \\ \dfrac{1}{J_{q_l}(\widehat{x})}, & \widehat{x} = \widehat{y} \end{cases}$$

and

$$\tilde{R}_l(\widehat{x}, \widehat{y}) = -R_l(\widehat{x}, \widehat{y}) \begin{cases} \dfrac{(q_l(\widehat{x}) - q_r(\widehat{y}))^T \nu(q_l(\widehat{x})))}{4\pi|q_l(\widehat{x}) - q_r(\widehat{y})|^2}, & \widehat{x} \neq \widehat{y}\,, \\ -\dfrac{2\sum\limits_{j=1}^{3} q_{jl}^{'}(\widehat{x})\nu_j(\widehat{x}) - \sum\limits_{j=1}^{3} q_{jl}^{''}(\widehat{x})\nu_j(\widehat{x})}{2J_{q_l}^2(\widehat{x})}, & \widehat{x} = \widehat{y}\,. \end{cases}$$

From this representation, it can be seen that the integral operators S_{ll} and K_{ll}, $l = 1, 2$, each have a weak singularity.

For the numerical approximation of the integrals in (20) and (21), we next use the following cubature rules for $n^{'} > 0$, see [13, 16]:

– cubature for integrals with a continuous integrand:

$$\int\limits_{\mathbb{S}^2} f(\widehat{y})\, ds(\widehat{y}) \approx \sum_{p^{'}=0}^{2n^{'}+1} \sum_{s^{'}=1}^{n^{'}+1} \widetilde{\mu}_{p^{'}} \widetilde{a}_{s^{'}} f(\widehat{y}_{s^{'}p^{'}}); \tag{22}$$

– cubature for integrals with a weak singularity in the integrand:

$$\int\limits_{\mathbb{S}^2} \frac{f(\widehat{y})}{|\widehat{x}-\widehat{y}|}\, ds(\widehat{y}) \approx \sum_{p^{'}=0}^{2n^{'}+1} \sum_{s^{'}=1}^{n^{'}+1} \widetilde{\mu}_{p^{'}} \widetilde{b}_{s^{'}} f(T_{\widehat{x}}^{-1} \widehat{y}_{s^{'}p^{'}}). \tag{23}$$

In the cubature rules (22)–(23), we use the following cubature points:

$$\widehat{y}_{s^{'}p^{'}} = \Big(\sin\theta_{s^{'}} \cos\varphi_{p^{'}}, \sin\theta_{s^{'}} \sin\varphi_{p^{'}}, \cos\theta_{s^{'}}\Big),$$

with $\varphi_{p^{'}} = \dfrac{p^{'}\pi}{n^{'}+1}$, $\theta_{s^{'}} = \arccos z_{s^{'}}$, where $z_{s^{'}}$ are the zeros of the Legendre polynomials $P_{n^{'}+1}$ [1]. The weights of the cubature rules are: $\widetilde{\mu}_{p^{'}} = \dfrac{\pi}{n^{'}+1}$, $\widetilde{a}_{s^{'}} = \dfrac{2(1-z_{s^{'}}^2)}{((n^{'}+1)P_{n^{'}}(z_{s^{'}}))^2}$, $\widetilde{b}_{s^{'}} = \widetilde{a}_{s^{'}} \displaystyle\sum_{l=0}^{n^{'}} P_l(z_{s^{'}})$. Following [13], we use an orthogonal transformation $T_{\widehat{x}}$ to move the weak singularity in the integrands to appear at the north pole of the sphere; it is present in (23). The transformation $T_{\widehat{x}}$ is defined as follows:

$$T_{\widehat{x}} = D_F(\varphi) D_T(\theta) D_F(-\varphi), \quad x \in \mathbb{S}^2$$

with

$$D_F(\psi) \begin{pmatrix} \cos(\psi) & -\sin(\psi) & 0 \\ \sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad D_T(\psi) \begin{pmatrix} \cos(\psi) & 0 & -\sin(\psi) \\ 0 & 1 & 0 \\ \sin(\psi) & 0 & \cos(\psi) \end{pmatrix}.$$

The cubatures (22)–(23) have exponential convergence for a sufficiently smooth integrand f, see [16].

For discretisation of the system (19), we use a Galerkin projection method. The unknown densities ϕ_l, $l = 1, 2$, are first approximated by a linear combination of real-valued spherical harmonics:

$$\phi_l \approx \widetilde{\phi}_l = \sum_{k=0}^{n} \sum_{m=-k}^{k} \phi_{k,m}^l Y_{k,m}^R, \quad l = 1, 2, \tag{24}$$

where $\phi_{k,m}^l$ are the unknown coefficients, and the real-valued spherical harmonics are:

$$Y_{k,m}^R = \begin{cases} \operatorname{Im} Y_{k,|m|}, & 0 < m \le k, \\ \operatorname{Re} Y_{k,|m|}, & -k \le m \le 0 \end{cases}$$

with $Y_{k,m}$ the spherical harmonics [1].

We consider the following discrete inner product, defined from the cubature rule (22):

$$(v, d) = \sum_{p=0}^{2n+1} \sum_{s=1}^{n+1} \mu_p a_s v(\widehat{y}_{sp}) d(\widehat{y}_{sp}), \quad v, d \in C(\mathbb{S}^2), \tag{25}$$

where the weights and points are generated from (22) for the parameter $n > 0$.

After approximating the unknown densities in (19) by (24), and by applying $(n+1)^2$ times the inner product (25) to (19) with $Y^R_{k,m}$, $k = 0, \dots, n$, $m = -k, \dots, k$, taking into account the representation of the integral operators (20) and (21), we obtain a linear system of equations for finding the unknown coefficients in the representation (24):

$$\begin{cases} \sum\limits_{k=0}^{n} \sum\limits_{m=-k}^{k} \left(\phi^1_{k,m} A^{11}_{kk'mm'} + \phi^2_{k,m} A^{12}_{kk'mm'} \right) = \\ \qquad = \sum\limits_{p=0}^{2n+1} \sum\limits_{s=1}^{n+1} \mu_p a_s \tilde{h}(\widehat{x}_{sp}) Y^R_{k,m}(\widehat{x}_{sp}), \\ \sum\limits_{k=0}^{n} \sum\limits_{m=-k}^{k} \left(\phi^1_{k,m} A^{21}_{kk'mm'} + \phi^2_{k,m} A^{22}_{kk'mm'} \right) = \\ \qquad = \sum\limits_{p=0}^{2n+1} \sum\limits_{s=1}^{n+1} \mu_p a_s \tilde{w}(\widehat{x}_{sp}) Y^R_{k,m}(\widehat{x}_{sp}), \end{cases} \tag{26}$$

for $k' = 0, \dots, n$, $m = -k, \dots, k$, $n = 0, 1, \dots$, with coefficients for $l, r = 1, 2$ given by:

$$\begin{aligned} A^{lr}_{kk'mm'} &= \sum_{p=0}^{2n+1} \sum_{s=1}^{n+1} \sum_{p'=0}^{2n'+1} \sum_{s'=1}^{n'+1} \mu_{p'} \mu_p a_s Y^R_{k',m'}(\widehat{x}_{sp}) \times \\ &\times \left(\begin{cases} \widetilde{a}'_s Y^R_{k,m}(\widehat{y}_{s'p'}) \left(M_{lr}(\widehat{x}_{sp}, \widehat{y}_{s'p'}) + \kappa L_{lr}(\widehat{x}_{sp}, \widehat{y}_{s'p'}) \right), & l \neq r \\ \widetilde{b}'_s Y^R_{k,m}(\widehat{y}^{s'p'}_{sp}) \left(\tilde{R}_l(\widehat{x}_{sp}, \widehat{y}^{s'p'}_{sp}) + \kappa R_l(\widehat{x}_{sp}, \widehat{y}^{s'p'}_{sp}) \right), & l = r \end{cases} \right. \\ &+ \left. \begin{cases} 0, & l \neq r \\ (-1)^l \dfrac{1}{2} Y^R_{k,m}(\widehat{x}_{sp}), & l = r \end{cases} \right), \end{aligned}$$

where $\widehat{y}^{s'p'}_{sp} = T^{-1}_{\widehat{x}_{sp}} \widehat{y}_{s'p'}$.

Calculation of the coefficients $A^{lr}_{kk'mm'}$ requires many operations. We can reduce the number of operations by using sequential calculation of smaller additional matrices [4, 5]. Employing this strategy, we can reduce the number of operations from $O(n^8)$ to $O(n^5)$. The coefficients $A^{lr}_{kk'mm'}$ of the system (26) need only to be calculated once, and can then be used at each step of the generalized Landweber iterative algorithm. In fact, we only need to calculate the right-hand side of the system (26) at each step for different functions $\tilde{h}$ and $\tilde{w}$.

After finding the unknown coefficients $\phi^l_{k,m}$, $l = 1, 2$, from (26), we can find an approximation of the unknown densities $\tilde{\phi}_l$, $l = 1, 2$, from (24).

The solution of the Robin boundary value problem (12)–(13) on the interior surface Γ_1 is given by (18); using the approximation of the densities (24), the cubature rule (22) and the representation of the integral operator (16), an approximation of the solution on Γ_1 is then given by:

$$u(\widehat{x}) \approx \sum_{s'=1}^{n'+1} \sum_{\rho'=0}^{2n'+1} \Big(\widetilde{b}_{s'} \widetilde{\mu}_{\rho'} \tilde{\phi}_1(T_{\widehat{x}}^{-1} \widehat{y}_{s'\rho'}) R_1(\widehat{x}, T_{\widehat{x}}^{-1} \widehat{y}_{s'\rho'}) + \\ + \widetilde{a}_{s'} \widetilde{\mu}_{\rho'} \tilde{\phi}_2(\widehat{y}_{s'\rho'}) L_{12}(\widehat{x}, \widehat{y}_{s'\rho'}) \Big), \quad \widehat{x} \in \Gamma_1.$$

4. Numerical experiments

In this section, we give some numerical examples. The main example is the numerical solution of the Cauchy problem (1)–(2) by using the iterative generalized Landweber algorithm with exact and noisy data. However, we first start by giving results for the Robin boundary value problem (12)–(13) needed in the iterative algorithm, to see how our proposed boundary integral equations method and discretisation perform for this direct problem.

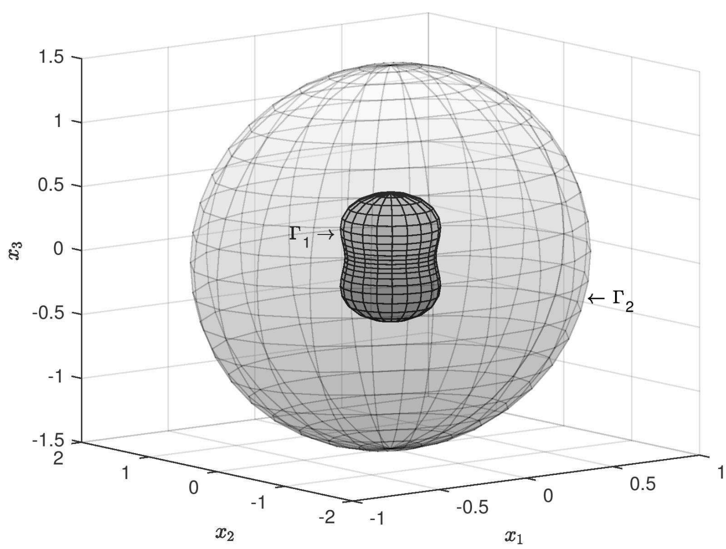


Fig. 1. The solution domain D in Ex. 1

Example 1 (Robin problem (12)–(13)). Let the doubly connected domain D (see Fig. 1) be bounded by the two surfaces:

$\Gamma_l = \{x(\theta, \varphi) = r_1(\theta, \varphi)\,(\sin\theta\cos\varphi, 2\sin\theta\sin\varphi, \cos\theta)\,,\ \theta \in [0, \pi],\ \varphi \in [0, 2\pi]\}\,,$

where radial function r_1 is:

$$r_1(\theta, \varphi) = \frac{1}{2\sqrt{1+\sqrt{2}}} \sqrt{\cos(2\theta) + \sqrt{2 - \sin^2(2\theta)}},$$

and

$$\Gamma_2 = \big\{x(\theta, \varphi) = (\sin\theta\cos\varphi, 1.5\sin\theta\sin\varphi, 1.5\cos\theta)\,,\ \theta \in [0, \pi],\ \varphi \in [0, 2\pi]\big\}.$$

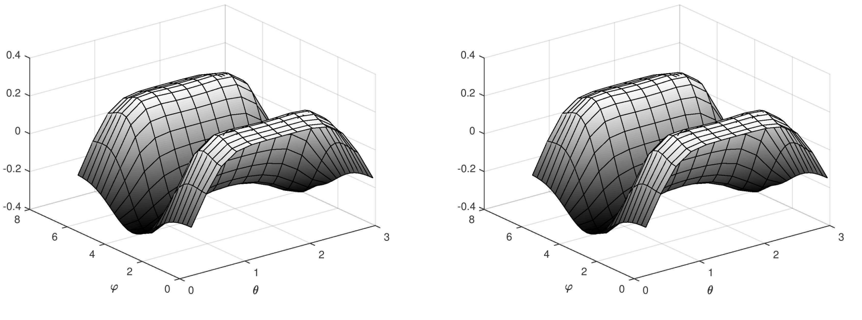
TABL. 1. L_2-errors for the Robin boundary value problem in Ex. 1

$n = n'$	$\dfrac{\|u_{ex} - u_n\|_{L_2(\Gamma_1)}}{\|u_{ex}\|_{L_2(\Gamma_1)}}$
2	5.13-E01
4	1.27-E02
6	6.56-E04
8	2.81-E05
10	1.91-E06
12	1.50-E07

The boundary data needed in the Robin boundary problem are generated from the exact solution: $u_{ex}(x) = x_2^2 - x_3^2 + x_1$, $x = (x_1, x_2, x_3)$, thus we get:

$$\frac{\partial u}{\partial \nu}(x) + \kappa u(x) = \nu_1(x) + 2x_2\nu_2(x) - 2x_3\nu_3(x) + \\ + \kappa(x_2^2 - x_3^2 + x_1), \quad x \in \Gamma_l, \quad l = 1, 2.$$

Values of the relative L_2-errors for the Robin boundary value problem (12)–(13) are presented in Table 1. As we can see from this table, super-algebraic convergence is present. In Fig. 2 are the exact and the numerical approximation for the function values on the internal boundary surface Γ_1, obtained with the discretisation parameters being $n = n' = 12$.



a). exact solution b). approximate solution

FIG. 2. Exact and numerical approximation for the function values on the internal boundary Γ_1 for the solution of the Robin boundary problem in Ex. 1

Example 2 (Cauchy problem (1)–(2)). Let the domain D (see Fig. 3) be bounded by the two surfaces:

$$\Gamma_l = \big\{x(\theta, \varphi) = r_l(\theta, \varphi)\,(\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)\,, \\ \theta \in [0, \pi],\ \varphi \in [0, 2\pi]\big\}, \quad l = 1, 2,$$

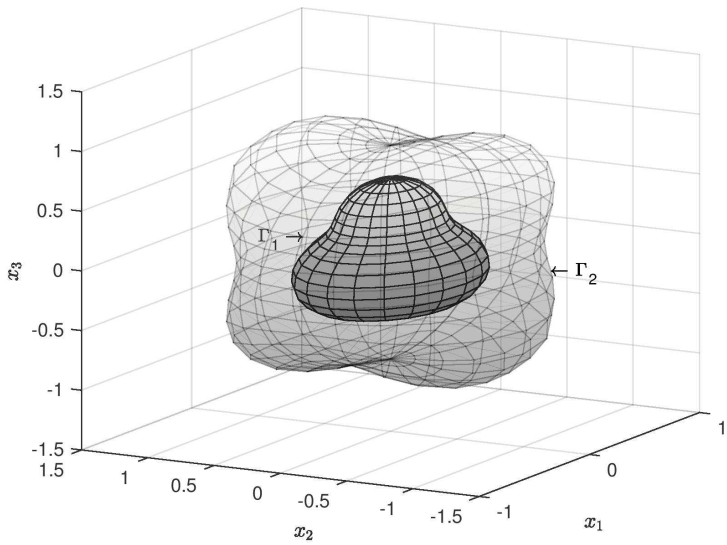
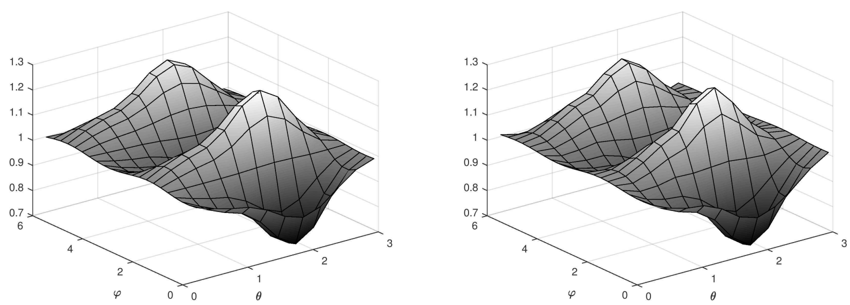


FIG. 3. The solution domain D in Ex. 2



a). exact solution

b). approximate solution

FIG. 4. Reconstruction of the solution on the boundary Γ_1 in Ex. 2 (exact data)

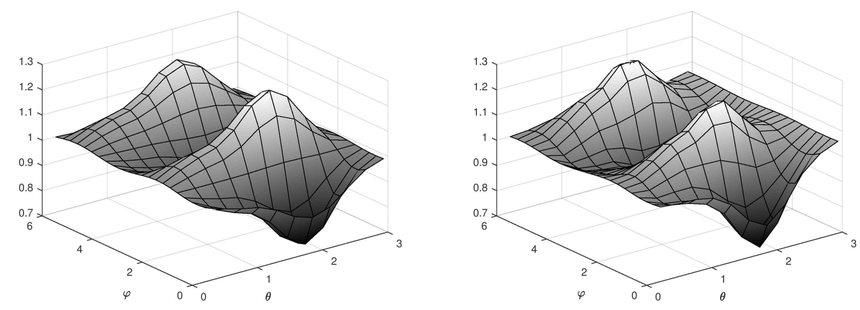
where the radial functions are as follows:

$$r_1(\theta, \varphi) = 0.2 \left(0.6 + \sqrt{4.25 + 2\cos(3\theta)}\right)$$

and

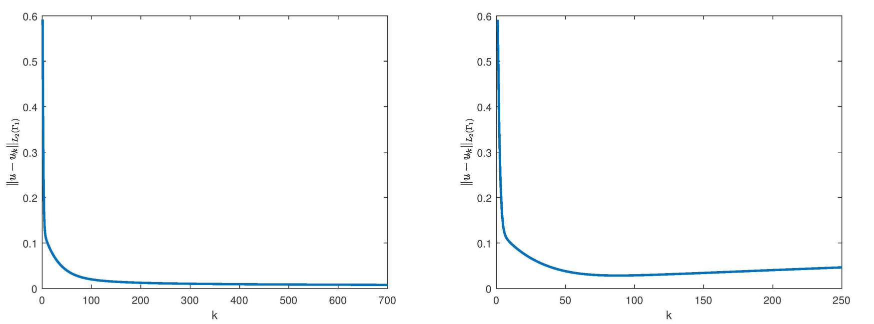
$$r_2(\theta, \varphi) = \sqrt{0.8 + 0.2\left(\cos(2\varphi) - 1\right)\left(\cos(4\theta) - 1\right)}\,.$$

We take a harmonic function $u_{ex}(x) = e^{x_2}\cos x_1 - e^{x_1}\sin x_2$ as an exact solution of the Cauchy problem (1)–(2). The necessary data for the Cauchy problem are generated from the exact solution u_{ex} on the external boundary Γ_2, as in Example 1.



a). exact solution   b). approximate solution

FIG. 5. Reconstruction of the solution on the boundary Γ_1 in Ex. 2 (3% noise)



a). exact data   b). 3% noisy data

FIG. 6. L_2-errors in Ex. 2

The results of the numerical reconstruction of the function u_{ex} by the generalized Landweber algorithm on the boundary Γ_1, for the cases of exact and noisy data, are shown in Figs. 4–5. Values of the relative L_2-errors at each iteration are presented in Fig. 6. In the case of exact data, after 700 iterations, we get

$$\frac{\|u_{ex} - u_{700}\|_{L_2(\Gamma_1)}}{\|u_{ex}\|_{L_2(\Gamma_1)}} = 0.0078$$

and for noisy data after 88 iterations (noise is 3%) we obtain

$$\frac{\|u_{ex} - u_{88}\|_{L_2(\Gamma_1)}}{\|u_{ex}\|_{L_2(\Gamma_1)}} = 0.0283,$$

in both cases the discretisation parameters for the direct boundary value problems are $n' = n = 10$. The relaxation parameter γ for the generalized Landweber method is selected as 0.5 (both for exact and noisy data).

5. Conclusion

We employed a generalized iterative Landweber algorithm, which can be applied to obtain a stable solution to the Cauchy problem, in particular it was used to find a stable approximation of the function values of the solution on the interior boundary surface of doubly connected three-dimensional domains. At each iteration step of the algorithm, we need to solve one Dirichlet and one Robin boundary value problem. Each of these direct boundary problems is solved by an indirect integral equations method in conjunction with a Galerkin method for the discretisation. Applicability of proposed algorithm and discretisation are highlighted by some numerical examples both for direct problems as well as for the Cauchy problem.

Bibliography

1. Abramowitz M. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables Dover Publications / M. Abramowitz, I. A. Stegun. – New-York, 1972.
2. Baravdish G. Adjoint-free iterative methods for the Cauchy problem for second-order elliptic equations / G. Baravdish, I. V. Borachok, R. S. Chapko, B. T. Johansson, M. Slodička // to appear.
3. Borachok I. V. On the numerical solution of a Dirichlet boundary value problem for Laplace Equation in three-dimensional doubly connected domains / I. V. Borachok, R. S. Chapko // Visnyk of the Lviv University. – Series Appl. Math and Informatics. – 2016. – Vol. 18.
4. Borachok I. V. Numerical solution of a Cauchy problem for Laplace equation in 3-dimensional domains by integral equations / I. V. Borachok, R. S. Chapko, B. T. Johansson // Inverse Problems in Science and Engineering. – 2016. – Vol. 9.
5. Borachok I. V. Numerical solution of an elliptic 3-dimensional Cauchy problem by the alternating method and boundary integral equations / I. V. Borachok, R. S. Chapko, B. T. Johansson // J. Inverse Ill-Posed Probl. – 2016. – Vol. 24. – P. 711-725.
6. Calderón A. P. Uniqueness in the Cauchy problem for partial differential equations / A. P. Calderón // Amer. J. Math. – 1958. – Vol. 80. – P. 16-36.
7. Hadamard J. Lectures on Cauchy's Problem in Linear Partial Differential Equations / J. Hadamard. – New Haven: Yale University Press, 1923.
8. 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.
9. Kress R. Linear Integral Equations (2nd ed.) / R. Kress. – Heidelberg: Springer-Verlag, 1999.
10. Chapko R. S. A direct integral equation method for a Cauchy problem for the Laplace equation in 3-dimensional semi-infinite domains / R. S. Chapko, B. T. Johansson // CMES Comput. Model. Eng. Sci. – 2012. – Vol. 85. – P. 105-128.
11. Chapko R. S. On the numerical solution of a Cauchy problem for the Laplace equation via a direct integral equation approach / R. S. Chapko, B. T. Johansson // Inverse Probl. Imaging. – 2012. – Vol. 6. – P. 25-36.
12. Chapko R. S. On the use of an integral equation approach for the numerical solution of a Cauchy problem for Laplace equation in a doubly connected planar domain / R. S. Chapko, B. T. Johansson, Y. Savka // Inverse Probl. Sci. Eng. – 2014. – Vol. 22. – P. 130-149.
13. Ganesh M. A high-order algorithm for obstacle scattering in three dimensions / M. Ganesh, I. G. Graham // J. Comp. Phys. – 2004. – Vol.198. – P. 211-242.
14. Shigeta T. Method of fundamental solutions with optimal regularization techniques for the Cauchy problem of the Laplace equation with singular points / T. Shigeta, D. L. Young // Journal of Computational Physics. – 2009. – Vol. 228. – P. 1903-1915.

15. Sun Yao Indirect boundary integral equation method for the Cauchy problem of the Laplace equation / Yao Sun // Journal of Scientific Computing. – 2017. – P. 469-498.
16. Wienert L. Die Numerische Approximation von Randintegraloperatoren f ür die Helmholtzgleichung im R^3 / L. Wienert // Ph. D. Thesis, University of Göttingen. – 1990.

Ihor Borachok,
Faculty of Applied Mathematics and Informatics,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine.

Received 01.11.2017; revised 17.11.2017

Журнал обчислювальної та прикладної математики | 2017 № 3 (126) | Journal of Numerical & Applied Mathematics

UDC 519.6

ON A BOUNDARY INTEGRAL EQUATION METHOD FOR ELASTOSTATIC CAUCHY PROBLEMS IN ANNULAR PLANAR DOMAINS

R. S. Chapko, B. T. Johansson, M. V. Shkolyk

Резюме. Розглянуто задачу Коші реконструкції поля зсуву (переміщення) планарного кільцеподібного лінійного пружного тіла, коли відомо вектор переміщення та напружень на зовнішній границі. Шукане значення представлене у вигляді еластостатичного потенціалу простого шару по двох границях тіла, що містить дві невідомі густини. Використовуючи задані граничні умови, отримано систему інтегральних рівнянь для знаходження цих густин. Досліджено властивості системи, здійснено дискретизацію за схемою Нистьома та регуляризацію Тіхонова. Наведені чисельні результати показують, що переміщення та відповідне поле напружень на границі, де не задано початкових значень, можна достатньо точно реконструювати як для точних вхдіних даних, так і для даних з похибкою.

Abstract. The Cauchy problem of reconstructing the displacement field of a planar annular linear elastic body from knowledge of the displacement vector and normal stress (traction) on the outer boundary is considered. The sought field is represented in terms of a single-layer elastic potential over the two boundary curves of the body involving two unknown densities. These densities are found by imposing the given boundary conditions, rendering a system of two boundary integrals to be solved for the densities. Properties of this system is investigated, and discretisation is done via a Nyström scheme together with Tikhonov regularization. Numerical results are included showing that the displacement can be accurately reconstructed in a stably way both for exact and noisy data together with the corresponding stress field on the boundary part where no information is initially given.

1. Introduction

Let $D \subset \mathbb{R}^2$ be an annular planar domain with sufficiently smooth boundaries Γ_1 and Γ_2. Each boundary part is a simple closed curve, and Γ_1 is contained in the bounded interior of Γ_2. The domain D is then the bounded region in-between Γ_1 and Γ_2 as illustrated in Fig. 1. We consider D to be a representative for a planar linear isotropic elastic body.

In some applications it is not possible to take measurements throughout the boundary of D. There can be a hostile environment or the body can be partly buried making only a part of the boundary accessible for measurements.

We assume that the external boundary Γ_2 is accessible for measurements but not Γ_1. Our aim is to reconstruct the missing data on Γ_1. We work in the setting of elastostatics (static elastic deformation), and, as mentioned, D is

Key words. Elastostatics, Cauchy problem, boundary integral equation method, trigonometrical quadrature method.

considered as a planar linear isotropic material. The displacement vector $u = (u_1, u_2) \in C^2(D) \cap C^1(\overline{D})$ describes the deformation of D. Under the standard assumptions of elastostatics (in particular small deformations of an isotropic and homogeneous linear elastic material) the displacement field satisfies the Navier equation

$$\mu\Delta u + (\lambda + \mu)\text{grad div} u = 0 \quad \text{in} \quad D, \tag{1}$$

with the constants μ and λ ($\mu > 0, \lambda > -\mu$) being the Lamé coefficients characterizing physical properties of the body.

We assume that the displacement and normal stress (the traction field) can be measured on Γ_2, giving respectively the Dirichlet boundary condition

$$u = f \quad \text{on} \quad \Gamma_2 \tag{2}$$

and Neumann boundary condition

$$Tu = g \quad \text{on} \quad \Gamma_2. \tag{3}$$

The vector functions f and g are given, and are commonly termed as Cauchy data. The element Tu is the stress tensor (due to molecular interactions from the deformation) in the outward unit normal direction to the boundary and is denoted as the traction. The traction can be expressed as

$$Tu = \lambda \text{div}\, u\, \nu + 2\mu(\nu \cdot \text{grad})u + \mu \text{div}(Qu)Q\nu,$$

where ν is the outward unit normal vector to the boundary, and the matrix Q is given by $Q = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. The introduction of the matrix Q makes for an easy way to express the last term in the right-hand side in the definition of Tu in the planar case, which otherwise has to be written in terms of a projection of a rotational field.

The Cauchy problem in elastostatics is then to solve (1)–(3), and in particular to find the displacement and traction on the boundary part Γ_1. Uniqueness is clear from standard results of elliptic equations such as the Holmgren theorem. However, the solution will not in general depend continuously on the data, that is the Cauchy problem is ill-posed. We tactically assume that the data are compatibly such that there exists a displacement field u.

In [3], an overview is given of a regularizing method based on a single-layer approach for the stable numerical solution to the corresponding classical Cauchy problem for the Laplace equation (for both two and three dimensional regions). The method surveyed builds on ideas given in [6] and [1]. We continue the work of [3], by extending the single-layer approach to the above Cauchy problem in elastostatics.

The Cauchy problem for elliptic equations is classical, and it is not possible in this work to give adequate overview and references. To at least guide the reader to some works, see the introduction in [2]. It is stationary heat transfer problems that make up the majority of the works on numerical methods for Cauchy problems, the corresponding results for elasticity is more limited. However, the first and third author of the work [8] have been active on inverse problems in elasticity, see for example [8, 9] and references therein (there are plenty more from these authors). However, the numerics is via the boundary element method or the method of fundamental solutions for simply connected domains. In [4] an iterative regularizing method is developed for the Cauchy problem of elastostatics in a half-plane containing a bounded inclusion.

For the outline of the work, in Section 2, we recall the fundamental solution to the Navier equation and discuss some classical integral formulations. In Section 3, the Cauchy problem is reduced to a system of boundary integral equations by representing the solution in terms of a single-layer solution over the boundary curves giving two unknown densities to determine. Furthermore, by parameterising the boundary curves, a parameterised system of integral equations is obtained. Properties of system is stated, see Theorem 1. Then, in Section 4, the parameterised system is discretised using a Nyström scheme. The discrete linear system obtained is ill-conditioned due to the ill-posedness of the Cauchy problem, hence Tikhonov regularization is invoked for its solution. In Section 5, numerical examples are presented for two different planar regions, showing that accurate and stable numerical results can be obtained both for the displacement and traction on the boundary part Γ_1. Some conclusions are given in the final section, Section 6.

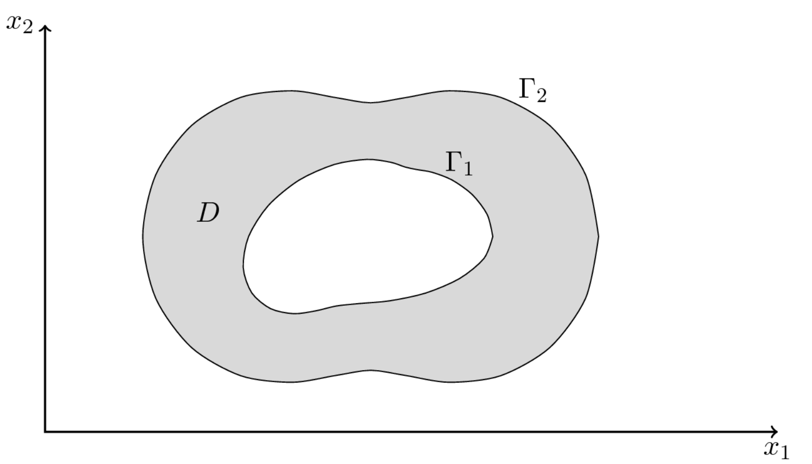


FIG. 1. Example of an annular planar domain D with boundary parts Γ_1 and Γ_2

2. Reduction to integral equations by Betti's formula

Reduction of the Cauchy problem (1)–(3) to a system of integral equations involves the use of the fundamental solution to the equation (1). In this section, we recall that fundamental solution, and for the sake of completeness, we state some direct representation formulas for the solution of (1)–(3). However, these representation formulas will not be further used, instead, in the next section, we introduce an alternative single-layer approach.

It is known [7] that the fundamental solution of the Navier equation (1) is given by

$$\Phi(x,y) = \frac{C_1}{2\pi}\Psi(x,y)I + \frac{C_2}{2\pi}J(x-y), \tag{4}$$

where

$$C_1 = \frac{\lambda + 3\mu}{\mu(\lambda+2\mu)}, \quad C_2 = \frac{\lambda+\mu}{\mu(\lambda+2\mu)},$$

and

$$\Psi(x,y) = \ln\frac{1}{|x-y|}, \quad x, y \in \mathbb{R}^2, \quad x \neq y.$$

Here, I is identity matrix (of size 2×2), J is defined by the formula

$$J(\omega) = \frac{\omega\omega^\top}{|\omega|^2}, \quad \omega \in \mathbb{R}^2 \setminus \{\bar{0}\}.$$

An analogue of the Green's formula for the Laplace equation is the so-called Betti's formula for the Navier equation; details and derivation of this formula can be found in for example [7]. Using Betti's formula, we seek the solution of (1)–(3) in the form

$$u(x) = \int\limits_{\Gamma_1} \left[T_y\Phi(x,y)\right]^\top \psi_1(y) - \Phi(x,y)\psi_2(y)\, ds(y) + B(x), \quad x \in D, \tag{5}$$

where

$$B(x) = \int\limits_{\Gamma_2} \Phi(x,y)g(y) - \left[T_y\Phi(x,y)\right]^\top f(y)\, ds(y).$$

The unknown vector-densities ψ_1 and ψ_2 represent the sought values (Cauchy data) on the inner inaccessible boundary Γ_1, that is

$$\psi_1(x) = u(x) \quad \text{and} \quad \psi_2(x) = Tu(x), \quad x \in \Gamma_1.$$

The representation (5) is then matched against the Cauchy data, that is against the displacement $u(x)$ respectively traction $Tu(x)$ on Γ_2. Using classical jump relations for the potentials in (5), we obtain the following system of integral equations of the second kind,

$$\begin{aligned}
&\frac{1}{2}\psi_1(x) - \int\limits_{\Gamma_1} \left[T_y\Phi(x,y)\right]^\top \psi_1(y)\, ds(y) + \int\limits_{\Gamma_2} \Phi(x,y)\psi_2(y)\, ds(y) = B(x),\\
&\frac{1}{2}\psi_2(x) - T_x\int\limits_{\Gamma_1} \left[T_y\Phi(x,y)\right]^\top \psi_1(y)\, ds(y) +\\
&\qquad + \int\limits_{\Gamma_2} T_x\Phi(x,y)\psi_2(y)\, ds(y) = TB(x),
\end{aligned} \tag{6}$$

where $x \in \Gamma_1$.

The described method of reducing the problem (1)–(3) to the above system of integral equations (IE) is naturally denoted the *direct* integral equation approach. We do not employ this but consider a related alternative strategy based on single-layer potentials.

3. Reduction to integral equations by potential theory

To reduce the Cauchy problem (1)–(3) to a the system of integral equations, we apply what is termed as an *indirect* integral equations approach based on potential theory.

We seek the solution of (1)–(3) as a single–layer elastic potential

$$u(x) = \int_{\Gamma_1} \Phi(x,y)\varphi_1(y)\,ds(y) + \int_{\Gamma_2} \Phi(x,y)\varphi_2(y)\,ds(y), \quad x \in D \tag{7}$$

with unknown vector-densities φ_1 and φ_2. We have the following result.

Proposition 1. *The single-layer potential* (7) *is the solution of the Cauchy problem* (1)–(3) *provided that the densities* φ_1 *and* φ_2 *are solutions of the following system of integral equations*

$$\begin{gathered} \int_{\Gamma_1} \Phi(x,y)\varphi_1(y)\,ds(y) + \int_{\Gamma_2} \Phi(x,y)\varphi_2(y)\,ds(y) = f(x), \quad x \in \Gamma_2, \\ \int_{\Gamma_1} T_x\Phi(x,y)\varphi_1(y)\,ds(y) + \frac{1}{2}\varphi_2(x) + \\ + \int_{\Gamma_2} T_x\Phi(x,y)\varphi_2(y)\,ds(y) = g(x), \quad x \in \Gamma_2. \end{gathered} \tag{8}$$

A proof of the proposition is obtained by matching the representation against the given Cauchy data involving classical jump relations for elastic single-layer potentials (for formulas, see [5, 7]).

There are singularities present in kernels in the above system. It is advantageous, both for theoretical and numerical investigations, to parameterise the system and make the singularities explicit. For the parameterisation, assume that the boundary curves Γ_1 and Γ_2 each have a parametric representation

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

where x_{i1} and x_{i2} are both 2π–periodic and twice continuously differentiable.

Using the representation of the boundary curves, we obtain from (8) the parameterised system of integral equations,

$$\begin{cases} \frac{1}{2\pi}\int\limits_0^{2\pi} K_{21}(t,\tau)\mu_1(\tau)\,d\tau + \frac{1}{2\pi}\int\limits_0^{2\pi} K_{22}(t,\tau)\mu_2(\tau)\,d\tau = f(t), \\ \frac{1}{2\pi}\int\limits_0^{2\pi} N_{21}(t,\tau)\mu_1(\tau)\,d\tau + \frac{1}{2}\frac{\mu_2(t)}{|x_2'(t)|} + \frac{1}{2\pi}\int\limits_0^{2\pi} N_{22}(t,\tau)\mu_2(\tau)\,d\tau = g(t), \end{cases} \tag{9}$$

where

$$K_{ij}(t,\tau) = 2\pi\Phi(x_i(t), x_j(\tau)), \quad i,j = 1,2,$$

$$N_{ij}(t,\tau) = \frac{1}{|x_i'(t)|}\left\{M_{ij}^1(t,\tau) + M_{ij}^2(t,\tau)\right\}, \quad i,j = 1,2,$$

$$M_{ij}^1(t,\tau) = C_3\frac{(x_i(t) - x_j(\tau)) \cdot x_i'(t)}{|x_i(t) - x_j(\tau)|^2}Q, \quad i \neq j,$$

$$M_{ij}^2(t,\tau) = -\frac{(x_i(t) - x_j(\tau)) \cdot Qx_i'(t)}{|x_i(t) - x_j(\tau)|^2}\left[C_3 I + C_4 \widetilde{J}(x_i(t), x_j(\tau))\right],$$
$$t \neq \tau \text{ when } \quad i = j,$$

$$\widetilde{J}(x_i(t), x_j(\tau)) = J(x_i(t) - x_j(\tau)), \quad t \neq \tau \text{ when } \quad i = j,$$

and

$$\widetilde{J}(x_i(t), x_i(t)) = \frac{x_i'(t)\,[x_i'(t)]^\top}{|x_i'(t)|^2}.$$

We have used the notation

$$f(t) = f(x_2(t)), \quad g(t) = g(x_2(t)), \quad \mu_i(\tau) = \varphi_i(x_i(\tau))|x_i'(\tau)|, \quad i = 1, 2,$$

and defined

$$C_3 = -\frac{2\mu}{\lambda + 2\mu} \quad \text{and} \quad C_4 = \frac{4(\lambda+\mu)}{\lambda + 2\mu}.$$

The kernels K_{22} and N_{22} (to be precise the component M_{22}^1) have singularities that can be written in an additive way using special weight functions. Put

$$K_{ii}(t,\tau) = \widetilde{K}_i(t,\tau) - \frac{C_1}{2}\ln\left\{\frac{4}{e}\sin^2\frac{t-\tau}{2}\right\} I, \quad i = 1, 2, \qquad (10)$$

where

$$\widetilde{K}_i(t,\tau) = \begin{cases} K_{ii}(t,\tau) + \frac{C_1}{2}\ln\left\{\frac{4}{e}\sin^2\frac{t-\tau}{2}\right\} I, & t \neq \tau, \\ \frac{C_1}{2}\ln\frac{1}{e|x_i'(t)|^2} I + C_2 \widetilde{J}(x_i(t), x_i(t)), & t = \tau. \end{cases}$$

Similar manipulations can be done for the kernels N_{11} and N_{22}. Denote by

$$M_{ii}^1(t,\tau) = M_i^3(t,\tau) + \frac{C_3}{2}\cot\frac{t-\tau}{2} Q, \quad i = 1, 2.$$

Then,

$$M_i^3(t,\tau) = \begin{cases} M_{ii}^1(t,\tau) - \frac{C_3}{2}\cot\frac{t-\tau}{2} Q, & t \neq \tau, \\ -\frac{C_3}{2}\frac{x_i'(t)\cdot x_i''(t)}{|x_i'(t)|^2} Q, & t = \tau. \end{cases}$$

As a result of these expressions, we obtain

$$N_{ii}(t,\tau) = \widetilde{N}_i(t,\tau) + \frac{C_3}{2|x_i'(t)|}\cot\frac{t-\tau}{2} Q, \quad i = 1, 2, \qquad (11)$$

where

$$\widetilde{N}_i(t,\tau) = \begin{cases} N_{ii}(t,\tau) - \frac{C_3}{2|x_i'(t)|}\cot\frac{t-\tau}{2} Q, & t \neq \tau, \\ \frac{1}{|x_i'(t)|}\left\{M_i^3(t,t) + M_{ii}^2(t,t)\right\}, & t = \tau. \end{cases}$$

Using for example L'Hopital's rule, it is straightforward to verify that the components M_{ii}^2 are at least continuous across $t = \tau$:

$$M_{ii}^2(t,t) = -\frac{x_i'(t)\cdot Qx_i''(t)}{2|x_i'(t)|^2}\left[C_3 I + C_4 \widetilde{J}(x_i(t), x_i(t))\right], \quad i = 1, 2.$$

Introduce the integral operators:

$$(S_{ii}\mu_i)(t) = \tfrac{1}{2\pi}\int\limits_0^{2\pi}\left[\widetilde{K}_i(t,\tau) - \tfrac{C_1}{2}\ln\left\{\tfrac{4}{e}\sin^2\tfrac{t-\tau}{2}\right\}I\right]\mu_i(\tau)\,d\tau,\quad i=1,2,$$
$$(S_{ij}\mu_j)(t) = \tfrac{1}{2\pi}\int\limits_0^{2\pi} K_{ij}(t,\tau)\mu_j(\tau)\,d\tau,\quad i,j=1,2,\quad i\neq j,$$
$$(L_{ii}\mu_i)(t) = \tfrac{1}{2\pi}\int\limits_0^{2\pi}\left[\widetilde{N}_i(t,\tau) + \tfrac{C_3}{2|x_i'(t)|}\cot\tfrac{t-\tau}{2}Q\right]\mu_i(\tau)\,d\tau,\quad i=1,2,$$
$$(L_{ij}\mu_j)(t) = \tfrac{1}{2\pi}\int\limits_0^{2\pi} N_{ij}(t,\tau)\mu_j(\tau)\,d\tau,\quad i,j=1,2,\quad i\neq j.$$

Taking into account the above expressions for the singularities in the kernels, the system of integral equations (9) can be written in operator form:

$$\begin{cases} (S_{21}\mu_1)(t) + (S_{22}\mu_2)(t) = f(t), \\ (L_{21}\mu_1)(t) + \left(\left(\tfrac{1}{2}I + L_{22}\right)\mu_2\right)(t) = g(t). \end{cases} \tag{12}$$

It can then be shown that for the operator corresponding to this system, the following holds.

Theorem 1. *The operator* $\mathbf{M} : L_2[0,2\pi]\times L_2[0,2\pi] \to L_2[0,2\pi]\times L_2[0,2\pi]$ *defined as*

$$\mathbf{M} = \begin{pmatrix} S_{21} & S_{22} \\ L_{21} & \tfrac{1}{2}I + L_{22}\end{pmatrix}$$

is injective and has a dense range.

This follows in the same way as for the corresponding theorem for the Laplace operator; for details in the case of the Laplace operator, see [3].

4. Full discretization and Tikhonov regularization

For the discretization of the system (12) of integral equations, we use quadratures rules that are based on trigonometric interpolation. The quadrature rules presume introducing an equidistant mesh of nodal points,

$$t_j = \frac{\pi}{n}j,\quad j = \overline{0, 2n-1},\quad n\in\mathbb{N}. \tag{13}$$

The operator S_{22} in (12) contains a logarithmic singularity. We therefore use the quadrature

$$\frac{1}{2\pi}\int\limits_0^{2\pi}\ln\left\{\frac{4}{e}\sin^2\frac{t-\tau}{2}\right\}f(\tau)\,d\tau \approx \sum_{j=0}^{2n-1}R_j(t)f(t_j), \tag{14}$$

where $R_j(t)$ is a weight function given by

$$R_j(t) := -\frac{1}{2n}\left\{1 + 2\sum_{k=1}^{n-1}\frac{\cos k(t-t_j)}{k} + \frac{\cos n(t-t_j)}{n}\right\}.$$

For a singularity of the kind contained in the operator L_{22} in (12), we apply instead the quadrature formula

$$\frac{1}{2\pi}\int_0^{2\pi} \cot\frac{\tau - t}{2} f(\tau), d\tau \approx \sum_{j=0}^{2n-1} \widetilde{T}_j(t) f(t_j), \tag{15}$$

with a weight function

$$\widetilde{T}_j(t) := -\frac{1}{n}\sum_{k=1}^{n-1} \sin k(t - t_j) - \frac{1}{2n}\sin n(t - t_j).$$

Since we work with 2π-periodic functions, it natural to use the trapezoidal rule

$$\frac{1}{2\pi}\int_0^{2\pi} f(\tau)\, d\tau \approx \frac{1}{2n}\sum_{j=0}^{2n-1} f(t_j). \tag{16}$$

Derivation of the quadrature formulas (14)–(16), and proof of their order of convergence can be found in [5].

For a partial discretization of the system of integral equations (12), we apply the quadrature formulas (14)–(16) on the equidistant nodal points (13). After then also collocating at these points, we obtain a system of linear equations

$$\begin{cases} \frac{1}{2n}\sum\limits_{j=0}^{2n-1} K_{21}(t_i, t_j)\mu_{1j} + \sum\limits_{j=0}^{2n-1}\left[\frac{1}{2n}\widetilde{K}_2(t_i, t_j) - \frac{C_1}{2} R_j(t_i) I\right]\mu_{2j} = f(t_i), \\ \frac{1}{2n}\sum\limits_{j=0}^{2n-1} N_{21}(t_i, t_j)\mu_{1j} + \\ \left\{\frac{1}{2|x_2'(t_i)|} I + \sum\limits_{j=0}^{2n-1}\left[\frac{1}{2n}\widetilde{N}_2(t_i, t_j) - \frac{C_3}{2|x_2'(t_i)|}\widetilde{T}_j(t_i) Q\right]\right\}\mu_{2j} = g(t_i), \end{cases} \tag{17}$$

where $i = \overline{0, 2n-1}$, and

$$\mu_{kj} \approx \mu_k(t_j), \quad k = 1, 2, \quad j = \overline{0, 2n-1}.$$

In a matrix-vector form, the system (17) can be written as

$$\mathbf{A}\bar{\mu} = \mathbf{F}. \tag{18}$$

As noted earlier, the problem (1)–(3) is ill-posed (there is no continuous dependence with respect to the input data). Hence, the system (12) is also ill-posed. A consequence of this is that the discrete linear system (17) is ill-conditioned, since it is obtained from (12). In order to obtain a stable numerical solution to (12), a regularizing method is needed. One such method is, for example, the classical Tikhonov regularization.

Tikhonov regularization for a linear system $Ax = b$ is based on minimizing the functional

$$\min_x \|Ax - b\|_2^2 + \alpha\|x\|_2^2,$$

where the number $\alpha > 0$ is the regularization parameter to be appropriately chosen.

The minimization problem is reduced to the approximation of x_α from the equality

$$(\alpha I + A^* A)x_\alpha = A^* b,$$

where A^* is adjoint operator of A.

In the case of a discrete system as (17), the usual transposed matrix $A^\top$ acts as an adjoint operator to the matrix A. Therefore, the regularization for (17) consists in finding $\bar{\mu}_\alpha$ from the system

$$(\alpha I + \mathbf{A}^\top \mathbf{A})\bar{\mu}_\alpha = \mathbf{A}^\top \mathbf{F}, \tag{19}$$

where the matrix $\mathbf{A}$ and vector $\mathbf{F}$ are determined in accordance with (17).

Taking into account the representation (7) of the solution to the Cauchy problem (1)–(3) and classical properties of the single-layer potential, the displacement vector u and traction Tu can be constructed on the inner boundary Γ_1 by the formulas

$$u(x) = (S_{11}\varphi_1)(x) + (S_{12}\varphi_2)(x), \quad x \in \Gamma_1$$

and

$$Tu(x) = \left(\left(-\frac{1}{2} I + L_{11} \right) \varphi_1 \right)(x) + (L_{12}\varphi_2)(x), \quad x \in \Gamma_1.$$

We generate an approximation to the quantities in discrete form by the formulas

$$u(x_1(t_i)) \approx \sum_{j=0}^{2n-1} \left[\frac{1}{2n} \widetilde{K}_1(t_i, t_j) - \frac{C_1}{2} R_j(t_i) I \right] \mu_{1j} + \frac{1}{2n} \sum_{j=0}^{2n-1} K_{12}(t_i, t_j)\mu_{2j}, \quad i = \overline{0, 2n-1} \tag{20}$$

and

$$Tu(x_1(t_i)) \approx -\frac{1}{2} \frac{\mu_{1i}}{|x_1'(t_i)|} + \sum_{j=0}^{2n-1} \left[\frac{1}{2n} \widetilde{N}_1(t_i, t_j) - \frac{C_3}{2|x_1'(t_i)|} \widetilde{T}_j(t_i) Q \right] \mu_{1j} + \frac{1}{2n} \sum_{j=0}^{2n-1} N_{12}(t_i, t_j)\mu_{2j}, \quad i = \overline{0, 2n-1}, \tag{21}$$

where μ_{kj} is the solution of the regularized system (19).

5. Numerical experiments

We shall present numerical results for two different configurations.

Example 1. Consider the annular domain of Fig. 2 having boundary curves

$$\Gamma_1 = \left\{ x_1(t) = (1.2\cos t, 1.6\sqrt{0.4\sin^2 t + \cos^2 t}\sin t) : \ t \in [0, 2\pi] \right\},$$
$$\Gamma_2 = \left\{ x_2(t) = (3\cos t, 4\sqrt{0.4\sin^2 t + \cos^2 t}\sin t) : \ t \in [0, 2\pi] \right\}.$$

As the exact solution to compare our numerical reconstructions with, we take

$$u_{ex}(x) = \Phi_1(x, y^*), \quad x \in D,$$

where Φ_1 is the first column of the matrix constituting the fundamental solution Φ in (4), and y^* is an arbitrary point which does not belong to the domain D.

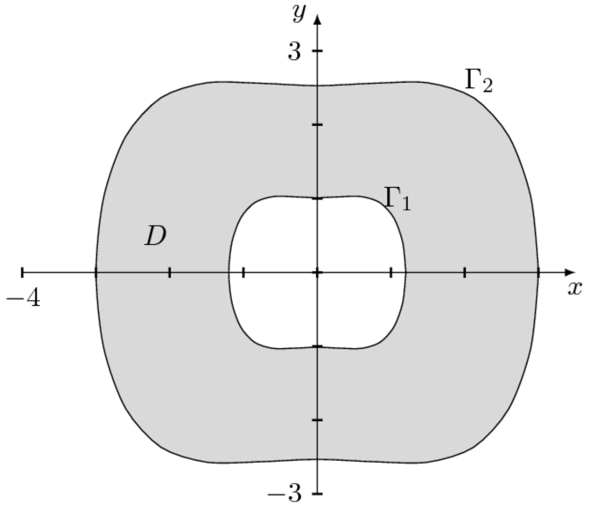


FIG. 2. Domain in Example 1

Then boundary values of the solution u_{ex} can be calculated exactly by the formulas

$$f_{ex\_i}(x) = \Phi_1(x, y^*) \quad \text{and} \quad g_{ex\_i}(x) = T\Phi_1(x, y^*), \quad x \in \Gamma_i,\ i = 1, 2.$$

(A)

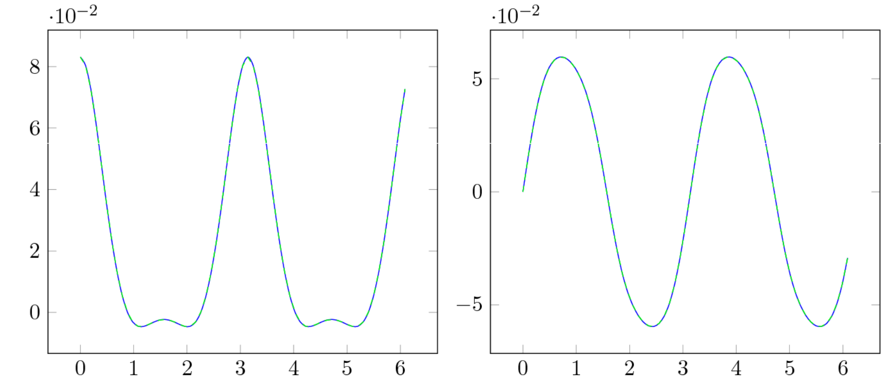
α	$\delta = 0$
E-10	3.94E-4
E-11	9.37E-5
E-12	2.92E-5
E-13	**2.59E − 5**
E-14	1.49E-4
E-15	1.33E-3

(B)

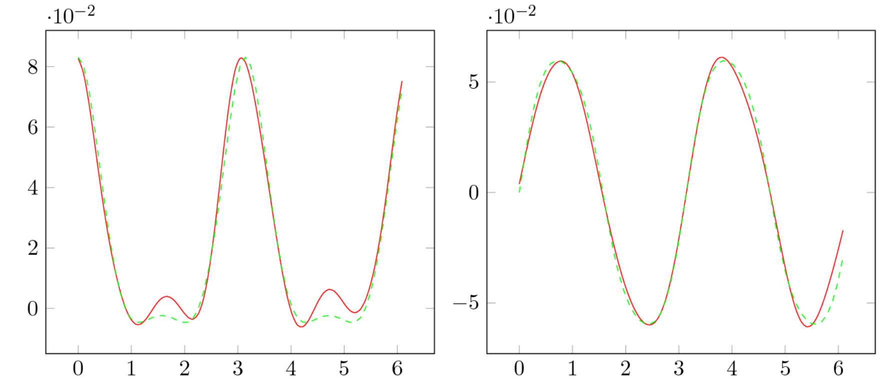
α	$\delta = 0.03$	$\delta = 0.05$
E-2	3.97E-2	4.18E-2
E-3	2.81E-2	4.92E-2
E-4	**3.65E − 3**	**5.36E − 3**
E-5	7.39E-3	8.56E-3
E-6	1.02E-2	1.76E-2
E-7	3.32E-2	5.33E-2

TABL. 1. Error in the reconstructed element f_{11} compared with the exact solution, for different parameters α in the case of (A) exact and (B) noisy data with noise level δ

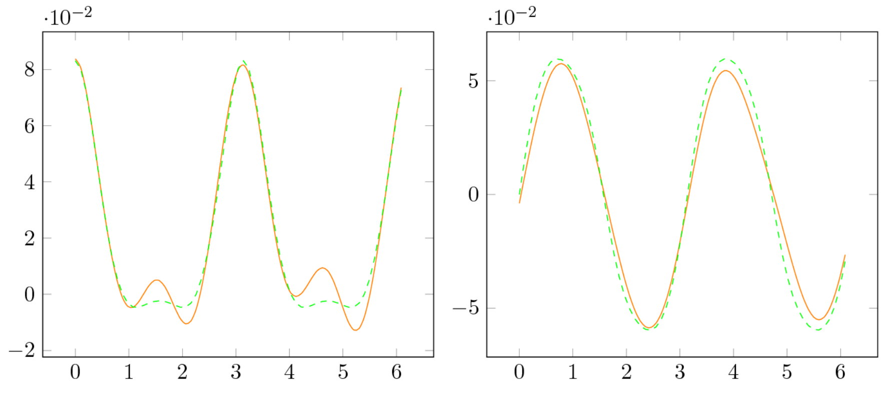
Let the Cauchy data (2) and (3) be generated as $f = f_{ex\_2}$ and $g = g_{ex\_2}$, respectively. Concerning parameters, we use $y^* = (0, 0)$, the Lamé coefficients are $\lambda = 2$, $\mu = 1$, and the discretization parameter $n = 32$ in (13).



(A) $\delta = 0$



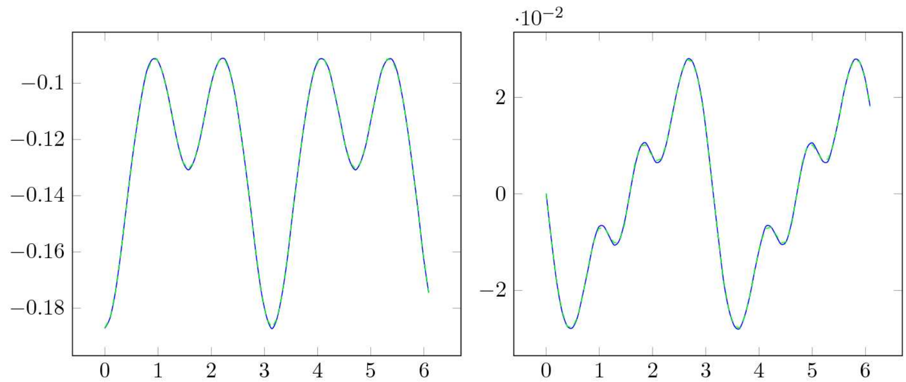
(B) $\delta = 0.03$



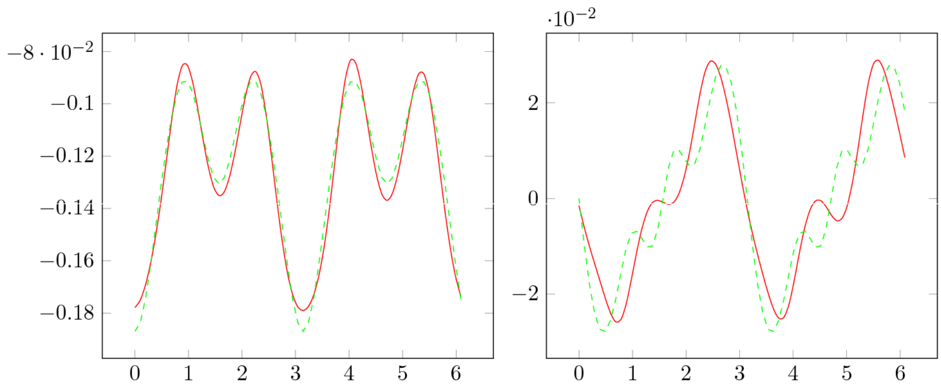
(C) $\delta = 0.05$

FIG. 3. Approximated (———) and exact (- - -) solutions of f_{11} (left) and f_{12} (right) for noise level δ

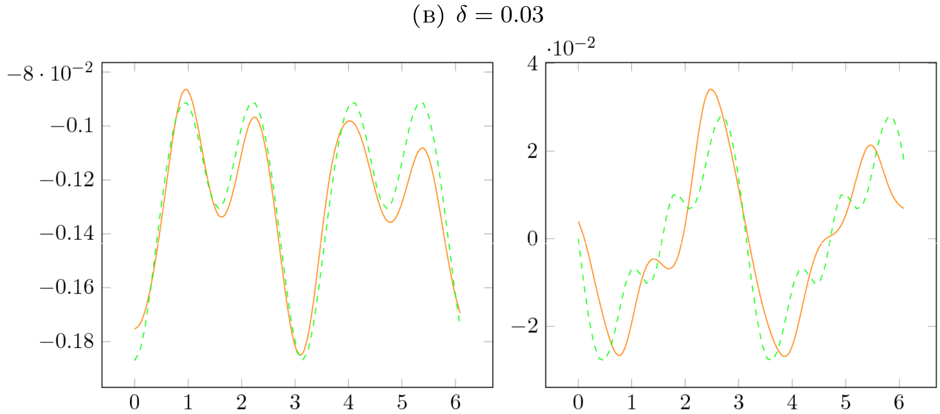
Due to the ill-posedness of the Cauchy problem, we apply Tikhonov regularization as mentioned in the previous section. The regularizing parameter α is chosen by trial and error. The optimal regularization parameter used is as given in Table 1 for exact data and for noisy data having 3% and 5% random pointwise error added into the data, respectively.



(A) $\delta = 0$



(B) $\delta = 0.03$



(C) $\delta = 0.05$

FIG. 4. Approximated (———) and exact (- - -) solutions of g_{11} (left) and g_{12} (right) for noise level δ

The number in bold is the value chosen for α.

To be more precise about noisy data, we point out that noisy data g_δ is generated from the exact value g as follows

$$g_\delta = g + \delta(2\eta - 1)\|g\|_{L_2},$$

with noise level δ and a random value $\eta \in (0,1)$.

The approximation of the displacement $f_1 = (f_{11}, f_{12})$ and traction $g_1 = (g_{11}, g_{12})$ on the inner boundary Γ_1, are calculated according to the formulas (20) and (21). The obtained results are shown in the Fig. 3 and Fig. 4.

As expected, the displacement vector is more accurately reconstructed than the traction. However, it is pleasing to see that also with noisy data, the reconstructions of the traction components follow the exact values. When more noise is added, the accuracy decreases but in a stable manner meaning that the results still resembles the exact values.

To convince the reader that the results presented are not optimised but are of the form to be expected for other configurations and data, we present results for a different domain and set of Cauchy data.

Example 2. In this example, we consider the doubly connected planar domain shown in Fig. 5. The boundary curves have parametric representation given by:

$$\Gamma_1 = \left\{x_1(t) = (0.7\cos t, 0.72\sin t + 0.6\cos^2 t) : \ t \in [0, 2\pi]\right\},$$
$$\Gamma_2 = \left\{x_2(t) = (1.8\cos t, 1.68\sin t + 1.4\cos^2 t) : \ t \in [0, 2\pi]\right\}.$$

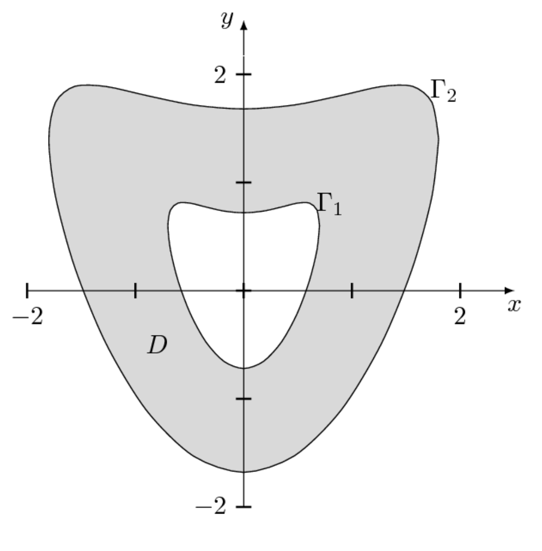


FIG. 5. Domain in Example 2

To have some data to compare against, we generate the Cauchy data artificially. This means that we first solve a Dirichlet boundary value problem, with values on the boundary curves as

$$f_i(x) = \begin{pmatrix} x_1 + x_2 \\ 5x_1 - x_2 \end{pmatrix}, \quad x = (x_1, x_2) \in \Gamma_i, \quad i = 1, 2.$$

Let the Lamé parameters be $\lambda = 2$, $\mu = 2$, and the discretization parameter is set to $n = 32$ in (13).

(A)

α	$\delta = 0$
E-7	6.47E-5
E-8	1.42E-5
E-9	4.27E-6
E-10	$\mathbf{1.22E-6}$
E-11	2.61E-6
E-12	2.59E-5

(B)

α	$\delta = 0.03$	$\delta = 0.05$
E-1	1.52E-1	2.11E-1
E-2	2.45E-1	2.97E-1
E-3	3.78E-2	5.66E-2
E-4	$\mathbf{2.55E-2}$	$\mathbf{3.13E-2}$
E-5	8.57E-2	8.01E-2
E-6	1.71E-1	2.08E-1

TABL. 2. Error in the reconstructed element f_{12} compared with the exact solution, for different parameters α in the case of (a) exact and (b) noisy data with noise level δ

Let the solution of the above Dirichlet problem be given as a single-layer elastic potential (7). After performing the similar manipulations that have been described for the Cauchy problem (that is parameterisation of the obtained system, making singularities explicit and then discretize), we obtain a system of linear equations

$$\sum_{j=0}^{2n-1}\left[\frac{1}{2n}\widetilde{K}_m(t_i,t_j)-\frac{C_1}{2}R_j(t_i)I\right]\mu_{mj}+\frac{1}{2n}\sum_{j=0}^{2n-1}K_{ml}(t_i,t_j)\mu_{lj}=f_m(x_m(t_i)),$$
$$i=\overline{0,2n-1},\quad m=1,2,\quad l=3-m.$$

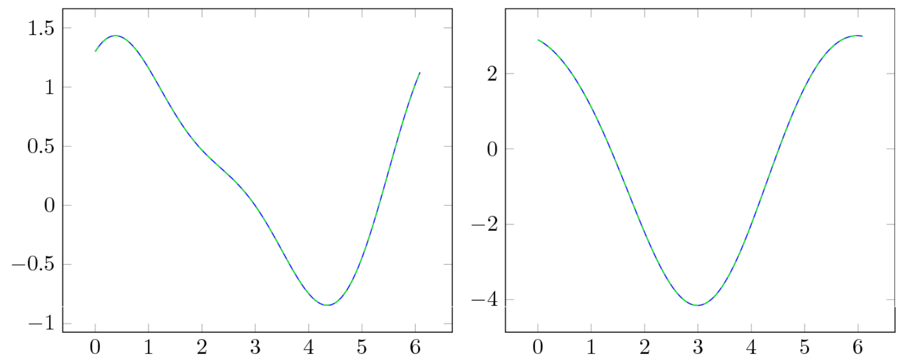
Solving for μ_{mj}, we can then calculate the Neumann boundary values by the formula

$$\begin{aligned} g_m(x_m(t_i)) &\approx \\ &\approx (-1)^m\frac{1}{2}\frac{\mu_{mi}}{|x_m'(t_i)|}+\sum_{j=0}^{2n-1}\left[\frac{1}{2n}\widetilde{N}_m(t_i,t_j)-\frac{C_3}{2|x_m'(t_i)|}\widetilde{T}_j(t_i)Q\right]\mu_{mj}+ \\ &+\frac{1}{2n}\sum_{j=0}^{2n-1}N_{ml}(t_i,t_j)\mu_{lj},\quad i=\overline{0,2n-1},\quad m=1,2,\quad l=3-m. \end{aligned}\tag{22}$$

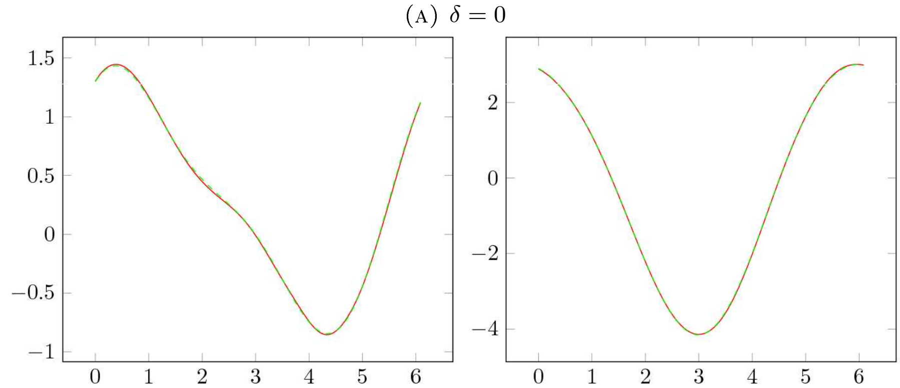
The Cauchy data in (2) and (3) is then generated as $f=f_2$ and $g=g_2$.

As in the previous example, we have to choose a regularization parameter α. The values used are given in bold in Table 2.

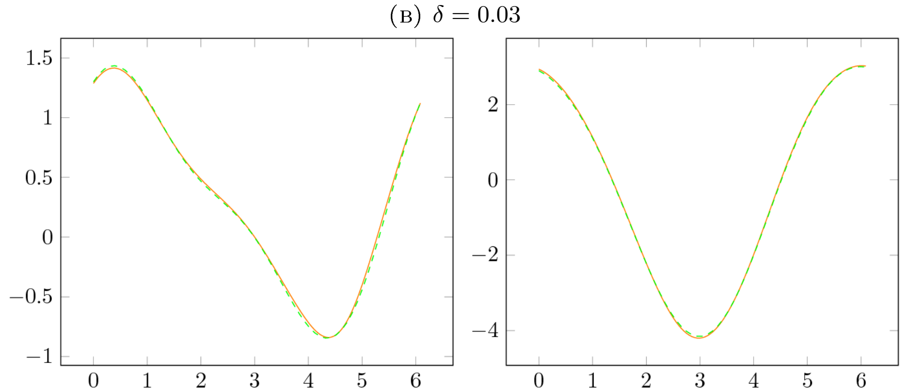
The numerical approximation of the Cauchy data on the inner boundary Γ_1 is found via the formulas (20) and (21). The results obtained are shown in



(A) $\delta = 0$



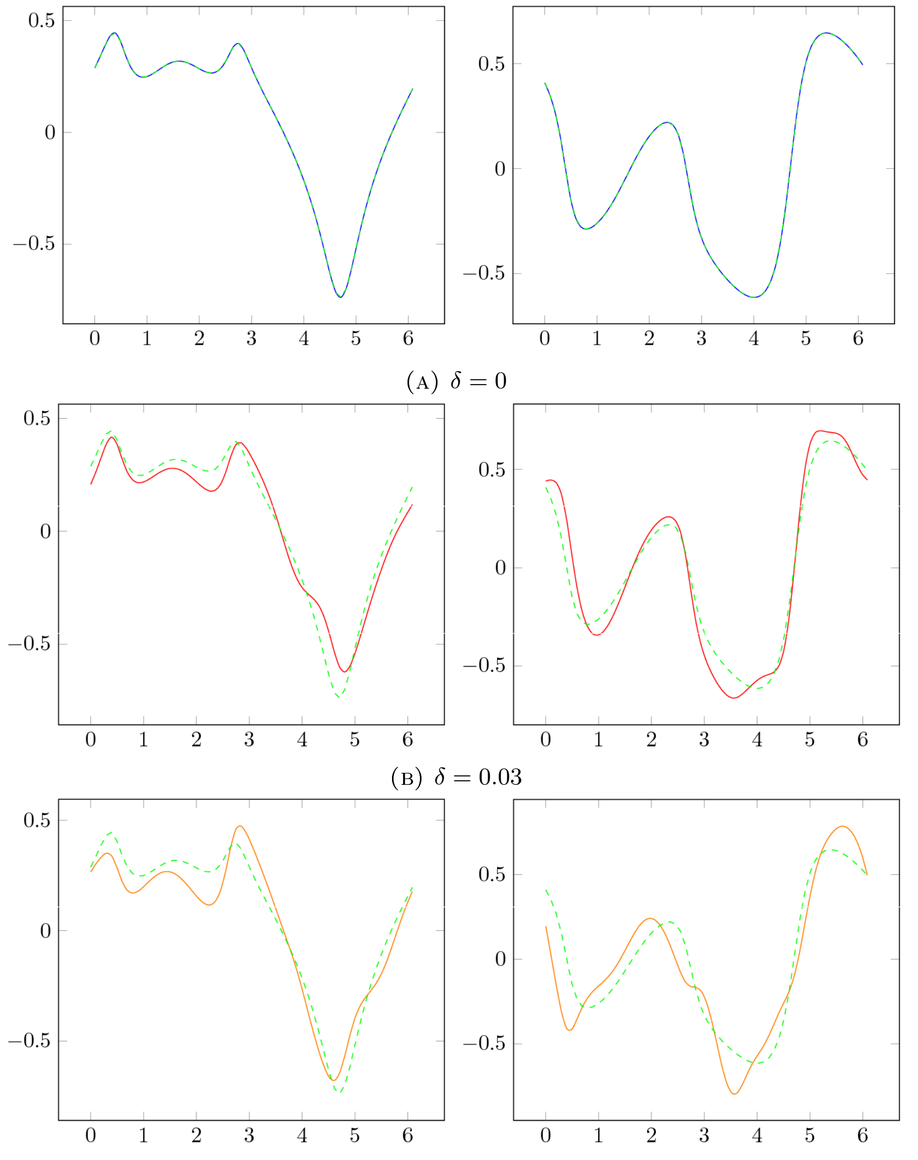
(B) $\delta = 0.03$



(C) $\delta = 0.05$

FIG. 6. Approximated (———) and exact (- - -) solutions of f_{11} (left) and f_{12} (right) for noise level δ

Fig. 6 and Fig. 7. It should be noted that in this example what is denoted as the exact Neumann data in the Cauchy problem is in fact an approximation since it is generated via solving the Dirichlet problem as explained above. But since the direct Dirichlet problem is well-posed and the discretization parameter is sufficiently large ($n = 32$), a high-order accuracy of the data generated by (22) is expected.



(A) $\delta = 0$

(B) $\delta = 0.03$

(C) $\delta = 0.05$

FIG. 7. Generated (- - -) and approximated (———) solutions of g_{11} (left) and g_{12} (right) for noise level δ

The obtained results are similar to those found in the previous example.

The traction vector is also here reconstructed with less accuracy than the placement as expected but follows the exact solution.

6. Conclusion

A regularizing method based on the elastic single-layer potential was derived for the Cauchy problem in elastostatics. The Cauchy data in the form of the displacement and traction is given on the outer boundary curve of a planar annular and linear isotropic body. From the single-layer representation, a system of boundary integrals to be solve for two unknown densities were obtained by matching against the data. It was shown that the system has at most one solution, and that there exists a solution for a dense set of square integrable data. Discretisation was done via a Nyström scheme in conjunction with Tikhonov regularization. Special care was taken to handle the various singularities in the kernels. The suggested approach performs well as verified by two numerical examples. The reconstructions corroborated well both for the displacement vector and traction with the sought solutions, also in the case of noisy data. The traction vector is naturally found with less accuracy. Overall, the outlined approach is a lightweight and flexible method for elastostatic Cauchy problems, and generalizes naturally earlier work [3] on a single-layer approach for the Cauchy problem for the Laplace equation.

Bibliography

1. Cakoni F. Integral equations for inverse problems in corrosion detection from partial Cauchy data / F. Cakoni, R. Kress // Inverse Problems Imaging – 2007. – Vol. 1. – P. 229-245.
2. Chapko R. A direct integral equation method for a Cauchy problem for the Laplace equation in 3-dimensional semi-infinite domains / R. Chapko, B. T. Johansson // CMES Comput. Model. Eng. Sci. – 2012. – Vol. 85. – P. 105-128.
3. Chapko R. A boundary integral approach for numerical solution of the Cauchy problem for the Laplace equation / R. Chapko, B. T. Johansson // Ukr. Math. J. – 2016. – Vol. 68. – P. 1665-1682.
4. Chapko R. On the numerical solution of a Cauchy problem in an elastostatic half-plane with a bounded inclusion / R. Chapko, B. T. Johansson, O. Sobeyko // CMES Comput. Model. Eng. Sci. – 2010. – Vol. 62. – P. 57-75.
5. Kress R. Linear integral equations (3rd ed.) / R. Kress. – New-York: Springer-Verlag, 2014.
6. Kress R. Nonlinear integral equations and the iterative solution for an inverse boundary value problem / R. Kress, W. Rundell // Inverse Problems – 2005. – Vol. 21. – P. 1207-223.
7. Kupradze V. D. Potential methods in the theory of elasticity / V. D. Kupradze. – Jerusalem: Israel Program for Scientific Translations, 1965.
8. Marin L. Conjugate gradient-boundary element method for the Cauchy problem in elasticity / L. Marin, D. N. Hào, D. Lesnic // Quart. J. Mech. Appl. Math. – 2002. – Vol. 55. – P. 227-247.
9. Marin L. Regularized MFS solution of inverse boundary value problems in three-dimensional steady-state linear thermoelasticity / L. Marin, A. Karageorghis, D. Lesnic // Int. J. Solids Struct. – 2016. – Vol. 91. – P. 127-142.

Roman Chapko, Mariia Shkolyk,
Faculty of Applied Mathematics and Informatics,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine;
Tomas Johansson,
School of Mathematics,
Aston University, Birmingham, B4 7ET, UK.

Received 13.07.2017; revised 28.08.2017

Журнал обчислювальної та прикладної математики | 2017 № 3 (126) | Journal of Numerical & Applied Mathematics

UDC 519.6

HETEROGENEOUS MODEL OF THE PROCESS OF THERMAL CONDUCTIVITY IN A MULTILAYERED MEDIUM WITH THIN LAYERS

L. M. Diakoniuk, Ya. H. Savula

Резюме. Ми розглядаємо початково крайову задачу теплопровідності в багатошаровому середовищі з малими товщинами шарів. Побудовано комп'ютерну модель, що дозволяє враховувати малі товщини шарів та уникати труднощів, які пов'язані з чисельною реалізацією задачі. Доведено теорему про неперервність та еліптичність білінійних форм варіаційних рівнянь. Для чисельного дослідження розв'язку використано напіваналітичний метод скінченних елементів.

Abstract. We consider initially the boundary value problem of thermal conductivity in a multilayered medium with small layer thicknesses. A computer model has been constructed, which allows to take into account the small thicknesses of the layers and avoid the difficulties associated with the numerical implementation of the problem. The theorem on the continuity and ellipticity of bilinear forms of variational equations is proved. The semi-analytic finite elements method used for numerical investigation of the solution.

1. Introduction

Modern materials and constructions, that are used in an instrument making, often have a difficult, heterogeneous structure. Natural environments physical processes areinvestigated in that, too in swingeing majority is heterogenous. It is known that at the mathematical design of problems in such environments there are two going neartaking into account of them difficult structure. The first approach envisages the use of process of homogenization, and second – in development of multiscale strategy. At development of the second approach, that allows more exactly to take into account the features of structure of environment, often there are the difficulties, constrained with the use of numeral methods(in particular, at application of Finite Elements Method) in areas that contain the thin including. In such cases build various-scale mathematical models to development of that the devoted works of many authors, in particular [1], [3], [4], [5] [7].

In this work we numerically construct a heterogeneous mathematical model of the process of heat and mass transfer in multilayer environments, where the thicknesses of layers are much smaller than other characteristic sizes.

Key words. Heat equation, heterogeneous model, finite elements method.

2. Formulation of problem

Let's consider the problem of heat conductivity for a multilayered medium of complex shape, which occupies an area

$$V = \bigcup V_k, k = 1, n; V_i \bigcap V_j = \oslash, i \neq j$$

with different thermal characteristics of the material of each layer. Boundary V_k of each regions consists of the lateral surface S_k and front surfaces S_k^- and S_k^+ and is considered Lipschitz (Fig.1).

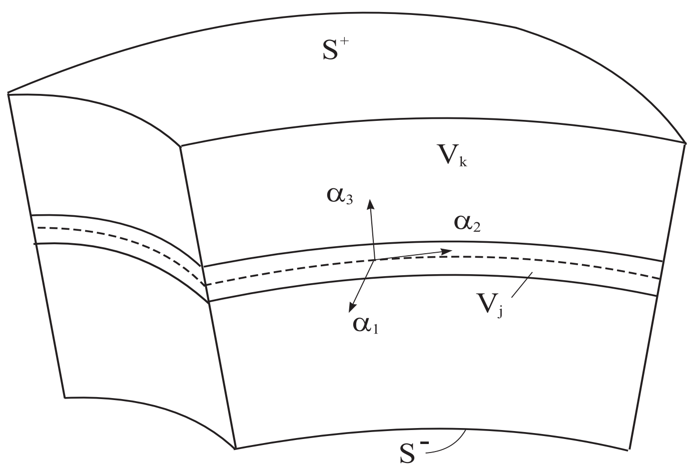


Fig. 1. The domain with thin layer

We denote J_2 the set of indexes of the regions V_k corresponding to "thin" layers whose thickness is small in comparison with other characteristic sizes. We will denote J_3 the set of indices of other areas. We associate each of the regions with some curvilinear coordinate systems related to the median surface of the area. The Lame coefficients of these regions are given by the relations

$$H_{1j} = A_{1j}(1 + k_{1j}\alpha_3^j), H_{2j} = A_{2j}(1 + k_{2j}\alpha_3^j), H_{3j} = 1.$$

Here A_1^j, A_2^j are the Lame coefficients of median surface, k_1^j, k_2^j are coefficient of curvature of the median surface. Let's consider the process of heat conduction in the described region, assuming that on the outer boundary there is a heat exchange according to Newton's law, and on the interfaces there is an ideal contact [2].

3. Transformation the three-dimensional heat transfer problem to two-dimensional in a thin layer

Consider a thin layer, where thickness is small compared with other characteristic of its size, occupying the area V_j. Let us itroduce the curvilinear coordinate system $(\alpha_1^j, \alpha_2^j, \alpha_3^j)$ associated with the median surface Ω_j of the region with the boundary Γ_j. The coordinate lines of this surface are the lines of major curvature. This

$$V_j = \{\alpha_1^j, \alpha_2^j, \alpha_3^j : \ (\alpha_1^j, \alpha_2^j) \in \Omega_j, -\frac{h_j}{2} \leq \alpha_3^j \leq \frac{h_j}{2}\},$$

where Ω_j is a two-dimensional region with a Lipschitz boundary on the median surface of the layer, h_j is the thickness of the layer. We will assume that on the facial surfaces $\alpha_3^j = \frac{h_j}{2}$ and $\alpha_3^j = -\frac{h_j}{2}$ the heat fluxes q_n^+ and q_n^- are given respectively, and on the lateral surface there is a heat exchange according to Newton's law

$$-\lambda_j \frac{\partial T_j}{\partial n}|_S = (T_j - T_c), \tag{1}$$

where λ_j is the coefficient of thermal conductivity, n is the external normal to the surface, T_j is the temperature function of the layer, T_c is the ambient temperature. At the initial moment of time, the temperature distribution is given by the ratio

$$T_j(\alpha_1^j, \alpha_2^j, \alpha_3^j, 0) = T_0^j(\alpha_1^j, \alpha_2^j, \alpha_3^j). \tag{2}$$

The process of thermal conductivity in the orthogonal coordinate system associated with the median surface of the layer can be described with the following equation:

$$c_j \rho_j \frac{\partial T_j}{\partial \tau} = \sum_{l=1}^{2} \frac{1}{H_{1j} H_{2j}} (\frac{\partial}{\partial \alpha_l^j} \lambda_j \frac{H_{1j} H_{2j}}{H_{lj}} \frac{\partial T_j}{\partial \alpha_l}) + q_{vj}, \tag{3}$$

where c_j is a coefficient of specific heat capacity, ρ_j is a coefficient of density, q_{vj} is the density of internal heat sources, τ is the time parameter. Considering that the thickness of the layer is small, we assume that the distribution of the desired function of temperature over the thickness of the layer is according to the linear law. In accordance with this assumption, we will supply the temperature in the region in the form

$$T_j(\alpha_1^j, \alpha_2^j, \alpha_3^j, \tau) = t_1(\alpha_1^j, \alpha_2^j, \tau) + \frac{2\alpha_3^j}{h_j} t_2(\alpha_1^j, \alpha_2^j, \tau). \tag{4}$$

We substitute (4) into (3) and orthogonalize the non-relation of the Bubnov-Galerkin equation to functions $v_1(\alpha_1^j, \alpha_2^j)$ and $\alpha_3^j v_2(\alpha_1^j, \alpha_2^j)$, wheare $v_1(\alpha_1^j, \alpha_2^j)$, $v_2(\alpha_1^j, \alpha_2^j) \in W_2^1(\Omega_j)$.

We select and calculate the integral over the variable α_3^j in the interval $[-\frac{h_j}{2}, -\frac{h_j}{2}]$. At the same time, let's take into account that the element of the volume and we use development in the Macrolena series of quantities $1/A_1^j(1 + k_1^j)\alpha_3^j$, $1/A_2^j(1 + k_2^j)\alpha_3^j$. Having neglected the magnitude $O((h_j k_i^j)^2)$, $i = 1, 2$ and taking into account the fact that $v_1(\alpha_1^j, \alpha_2^j), v_2(\alpha_1^j, \alpha_2^j)$ are arbitrary functions, we obtain the following key equations with the respect to the unknown functions t_1^j, t_2^j :

$$c_j\rho_j h_j \frac{\partial t_1^j}{\partial \tau} + c_j\rho_j \frac{h_j^2}{6}(k_1^j + k_2^j)\frac{\partial t_2^j}{\partial \tau} = \sum_{i=1}^{2}\Big(\frac{h_j}{A_1^j A_2^j}\frac{\partial}{\partial \alpha_i^j}\Big(\lambda_j \frac{A_{3-i}^j}{A_i^j}\frac{\partial t_1^j}{\partial \alpha_i^j}\Big)+$$
$$+\frac{h_j^2}{6A_1^j A_2^j}\frac{\partial}{\partial \alpha_i^j}\Big(\lambda_j \frac{A_{3-i}^j}{A_i^j}(k_{3-i}^j - k_i^j)\frac{\partial t_2^j}{\partial \alpha_i^j}\Big)\Big) + (1 + k_1^j\frac{h_j}{2})(1 + k_2^j\frac{h_j}{2})q_n^+ + \tag{5}$$
$$+(1 - k_1^j\frac{h_j}{2})(1 - k_2^j\frac{h_j}{2})q_n^- - q_1 = 0,$$

$$c_j\rho_j \frac{h_j^2}{6}(k_1^j + k_2^j)\frac{\partial t_1^j}{\partial \tau} + c_j\rho_j \frac{h_j}{3}\frac{\partial t_2^j}{\partial \tau} =$$
$$= \sum_{i=}^{2}\Big(\frac{h_j^2}{6A_1^j A_2^j}\frac{\partial}{\partial \alpha_i^j}\Big(\lambda_j \frac{A_{3-i}^j}{A_i^j}(k_{3-i}^j - k_i^j)\frac{\partial t_1^j}{\partial \alpha_i^j}\Big)+$$
$$+\frac{h_j}{3A_1^j A_2^j}\frac{\partial}{\partial \alpha_i^j}\Big(\lambda_j \frac{A_{3-i}^j}{A_i^j}\frac{\partial t_2^j}{\partial \alpha_i^j}\Big)\Big) + (1 + k_1^j\frac{h_j}{2})(1 + k_2^j\frac{h_j}{2})q_n^+ + \tag{6}$$
$$+(1 - k_1^j\frac{h_j}{2})(1 - k_2^j\frac{h_j}{2})q_n^- + \frac{4\lambda_j}{h_j} - q_2 = 0.$$

We use the following notation

$$q_1 = \int_{-\frac{h_j}{2}}^{\frac{h_j}{2}} q_v(1 + k_1^j\alpha_3^j)(1 + k_2^j\alpha_3^j)d\alpha_3^j,$$

$$q_2 = \frac{2}{h_j}\int_{-\frac{h_j}{2}}^{\frac{h_j}{2}} q_v(1 + k_1^j\alpha_3^j)(1 + k_2^j\alpha_3^j)\alpha_3^j d\alpha_3^j,$$

$$-\lambda_j \frac{\partial T_j}{\partial \alpha_3^j} = q_n^+, \text{ for } \alpha_3^j = \frac{h_j}{2}.$$

By performing similar transformations to the boundary condition on the lateral cylindrical surface, we obtain boundary conditions for functions t_1^j, t_2^j in the form

$$-\Big(\frac{\lambda_j h_j}{A_i^j}\frac{\partial t_1^j}{\partial \alpha_i^j} + \frac{1}{6}\frac{\lambda_j h_j^2}{A_i^j}(k_{3-i}^j - k_i^j)\frac{\partial t_2^j}{\partial \alpha_i^j}\Big)n_i = \alpha\Big(h_j t_1^j + \frac{h_j^2}{6}k_\Gamma^j t_2^j - t_1^c\Big),$$
$$-\Big(\frac{1}{6}\frac{\lambda h_j^2}{A_i^j}(k_{3-i}^j - k_i^j)\frac{\partial t_1^j}{\partial \alpha_i^j} + \frac{h_j}{3}\frac{\lambda_j}{A_i^j}\frac{\partial t_2^j}{\partial \alpha_i^j}\Big)n_i = \alpha\Big(\frac{h_j^2}{6}k_\Gamma^j t_1^j + \frac{h_j}{3}t_2^j - t_2^c\Big), \tag{7}$$

$$t_1^c = \int_{-\frac{h_j}{2}}^{\frac{h_j}{2}} T_c(1 + k_\Gamma^j\alpha_3^j)d\alpha_3^j,$$

$$t_2^c = \frac{2}{h_j}\int_{-\frac{h_j}{2}}^{\frac{h_j}{2}} T_c(1 + k_\Gamma^j\alpha_3^j)\alpha_3^j d\alpha_3^j,$$

$k_\Gamma^j = k_1^j n_1^2 + k_2^j n_2^2$, (n_1, n_2) are the coordinates of the unit normal vector to Γ.

Here the element of the surface area

$$dS = A^j_\Gamma(1 + k^j_\Gamma\alpha^j_3)d\Gamma,$$

where $A^j_\Gamma = A^j_1 n^2_1 + A^j_2 n^2_2$.

In the same way from equation (6)–(7) we also obtain the initial conditions

$$\begin{aligned} &h_j t^j_1(\alpha^j_1, \alpha^j_2, 0) + \frac{h^2_j}{6}(k^j_1 + k^j_2)t^j_2(\alpha^j_1, \alpha^j_2, 0) = t^0_1, \\ &\frac{h^2_j}{6}(k^j_1 + k^j_2)t^j_1(\alpha^j_1, \alpha^j_2, 0) + \frac{h_j}{3}t_2(\alpha^j_1, \alpha^j_2, 0) = t^0_2, \end{aligned} \tag{8}$$

where

$$t^0_1 = \int_{-\frac{h_j}{2}}^{\frac{h_j}{2}} T^j_0(1 + k^j_1\alpha^j_3)(1 + k^j_2\alpha^j_3)d\alpha^j_3,$$

$$t^0_2 = \frac{2}{h_j}\int_{-\frac{h_j}{2}}^{\frac{h_j}{2}} T^j_0(1 + k^j_1\alpha^j_3)(1 + k^j_2\alpha^j_3)\alpha^j_3 d\alpha^j_3.$$

Thus, for a thin layer it is possible to reduce the dimensionality of the problem to two-dimensional relative curvilinear coordinates on the median surface. As a result, we obtain a mathematical model of the process of thermal conductivity in a thin layer, consisting of equations (5), (6), boundary conditions (7) and initial conditions (8).

4. Description of the process of thermal conductivity in a multilayer area

Considering the results obtained in the previous section, the heterogeneous mathematical model of the heat conduction process in a multilayered medium, can be presented as the following system of differential equations of different measurements in spatial coordinate

$$c_j\rho_j\frac{\partial T_j}{\partial\tau} = \sum_{l=1}^{3}\frac{1}{H_{1j}H_{2j}}\left(\frac{\partial}{\partial\alpha^j_l}\lambda_j\frac{H_{1j}H_{2j}}{H_{lj}}\frac{\partial T_j}{\partial\alpha_l}\right) + q_{vj},\ j \in V_j, \tag{9}$$

$$\begin{aligned} &c_j\rho_j h_j\frac{\partial t^{(j)}_1}{\partial\tau} + c_j\rho_j\frac{h^2_j}{6}(k^{(j)}_1 + k^{(j)}_2)\frac{\partial t^{(j)}_2}{\partial\tau} = \\ &+\sum_{i=1}^{2}\Big(\frac{h_j}{A^{(j)}_1A^{(j)}_2}\frac{\partial}{\partial\alpha^{(j)}_i}(\lambda_j\frac{A^{(j)}_{3-i}}{A^{(j)}_i}\frac{\partial t^{(j)}_1}{\partial\alpha^{(j)}_i})+ \\ &+\frac{h^2_j}{6A^{(j)}_1A^{(j)}_2}\frac{\partial}{\partial\alpha^{(j)}_i}(\lambda_j\frac{A^{(j)}_{3-i}}{A^{(j)}_i}(k^{(j)}_{3-i} - k^{(j)}_i)\frac{\partial t^{(j)}_2}{\partial\alpha^{(j)}_i})+ \\ &+(1 + k^{(j)}_1\frac{h_j}{2})(1 + k^{(j)}_2\frac{h_j}{2})q^{(j)+}_n+ \\ &+(1 - k^{(j)}_1\frac{h_j}{2})(1 - k^{(j)}_2\frac{h_j}{2})q^{(j)-}_n - q_1 = 0, \end{aligned} \tag{10}$$

$$c_j\rho_j\frac{h_j^2}{6}(k_1^{(j)}+k_2^{(j)})\frac{\partial t_1^{(j)}}{\partial\tau}+c_j\rho_j\frac{h_j}{3}\frac{\partial t_2^{(j)}}{\partial\tau}=$$
$$=\sum_{i=1}^{2}(\frac{h_j^2}{6A_1^{(j)}A_2^{(j)}}\frac{\partial}{\partial\alpha_i^{(j)}}(\lambda_j\frac{A_{3-i}^{(j)}}{A_i^{(j)}}(k_{3-i}^{(j)}-k_i^{(j)})\frac{\partial t_1^{(j)}}{\partial\alpha_i^{(j)}})+$$
$$+\frac{h_j}{3A_1^{(j)}A_2^{(j)}}\frac{\partial}{\partial\alpha_i^{(j)}}(\lambda_j\frac{A_{3-i}^{(j)}}{A_i^{(j)}}\frac{\partial t_2^{(j)}}{\partial\alpha_i^{(j)}})+ \tag{11}$$
$$+(1+k_1^{(j)}\frac{h_j}{2})(1+k_2^{(j)}\frac{h_j}{2})q_n^{(j)+}+$$
$$+(1-k_1^{(j)}\frac{h_j}{2})(1-k_2^{(j)}\frac{h_j}{2})q_n^{(j)-}+\frac{4\lambda}{h_j}-q_2=0.$$

On the boundary with the external environment, the desired functions must satisfy the relation

$$(-\lambda_j\frac{\partial T_j}{\partial n}-a(T_j-T_{c_k}))|_{S_j}=0, \tag{12}$$

$$-\sum_{i=1}^{2}(\frac{\lambda_j h_j}{A_i^{(j)}}\frac{\partial t_1^(j)}{\partial\alpha_i^{(j)}}+\frac{1}{6}\frac{\lambda_j h_j^2}{A_i^{(j)}}(k_{3-i}^{(j)}-k_i^{(j)})\frac{\partial t_2^{(j)}}{\partial\alpha_i^(j)})n_i=$$
$$=a(h_jt_1^{(j)}+\frac{h_j^{(2)}}{6}k_\Gamma^{(j)}t_2^{(j)}-t_1^c), \tag{13}$$

$$-\sum_{i=1}^{2}(\frac{1}{6}\frac{\lambda_j h_j^2}{A_i^{(j)}}(k_{3-i}^{(j)}-k_i^{(j)})\frac{\partial t_1^(j)}{\partial\alpha_i^{(j)}}+\frac{1}{3}\frac{\lambda_j h_j}{A_i^{(j)}}\frac{\partial t_2^{(j)}}{\partial\alpha_i^(j)})n_i=$$
$$=a(\frac{1}{6}h_j^2k_\Gamma^{(j)}t_1^{(j)}+\frac{h_j^{(2)}}{3}t_2^{(j)}-t_2^c) \tag{14}$$

and initial conditions

$$T_j(\alpha_1^j,\alpha_2^j,\alpha_3^j,0)=T_0^j(\alpha_1^j,\alpha_2^j,\alpha_3^j),\ \text{for } j\in J_3, \tag{15}$$

$$h_jt_1^{(j)}(\alpha_1,\alpha2,0)+\frac{h_j^2}{6}(k_1^{(j)}+k_2^{(j)})t_2^{(j)}(\alpha_1,\alpha2,0)=t_1^0,\ \text{for } j\in J_2, \tag{16}$$

$$\frac{h_j^2}{6}(k_1^{(j)}+k_2^{(j)})t_1^{(j)}(\alpha_1,\alpha_2,0)+\frac{h_j}{3}(t_2^{(j)}(\alpha_1,\alpha_2,0)=t_2^0,\ \text{for } j\in J_2. \tag{17}$$

For a complete description of the mathematical model, it is necessary to introduce the conjugation conditions, which describe the equality of the functions of temperature distribution and heat fluxes on the boundary of the collisions of the regions:

$$T_{1_{j_1}}|_{S_j}=T_{1_{j_2}}, \tag{18}$$

$$\lambda_{j_1}\frac{\partial T_{j_1}}{\partial n}|_{S_j}=\lambda_{j_2}\frac{\partial T_{j_2}}{\partial n}. \tag{19}$$

If the contact layer is thin, it should be taken into account that the temperature on the upper face surface is given by the ratio

$$T_{j_1}=t_1^{(j_2)}+t_2^{(j_2)}$$

and on the bottom, respectively

$$T_{j_1} = t_1^{(j_2)} - t_2^{(j_2)}.$$

Taking into account all the described relations, we obtained a closed system of differential equations of the second order for the determination of unknown unknown temperature functions.

5. Variational formulation

According to the Bubnov-Galerkin method, we construct the variational equations of the problem (9)–(19). For the transformation of the integrals we use the following formula, derived from the Green's formula:

$$-\int_V div(\lambda\nabla u)vdV = \int_V \lambda\nabla u\nabla vdV - \int_S \lambda\frac{\partial u}{\partial n}vdS. \tag{20}$$

Consider the variational problem of thermal conductivity in a multilayered medium with subtle inclusions, that is to find functions u, which satisfy the equation

$$\begin{gathered}\sum_{k\in J_3} A_k(T_k,u^k) + \sum_{j\in J_2} a_j(t_j,u^j) + \sum_{j\in J_2} m^j(t_j',u^j) = \\ = \sum_{k\in J_3}(f_k,u^k) + \sum_{j\in J_2}(f^j,u^j),\end{gathered} \tag{21}$$

$$M^k(T_k,u^k) = \int_{V_k} c^k\rho^k T_k u^k dv,$$

$$\begin{gathered}A^k(T_k,u^k) = \int_{V_k}\lambda^k \operatorname{grad} u \operatorname{grad} T_k dv - \int_{S_k}\lambda^k\frac{\partial T_k}{\partial v}u^k dS + \\ + g_{11}\int_{S^+} aT_{ck}u^{ck}ds + g_{12}\int_{S^-} aT_{ck}u^{ck}ds,\end{gathered}$$

$$\begin{gathered}a^j(t^j,u^j) = \int_{\Omega_j} t_p^{jT} Au^j d\Omega + \int_{\Gamma_j} t^{jT}Gu^j d\Gamma + \\ + g_{21}\int_{\Omega_{j+}}(1+k_1\frac{h}{2})(1+k_2\frac{h}{2})(t_1^{j+}+t_2^{j+})(u_1^{j+}+u_2^{j+})d\Omega + \\ + g_{22}\int_{\Omega_{j-}}(1-k_1\frac{h}{2})(1-k_2\frac{h}{2})(t_1^{j-}-t_2^{j-})(u_1^{j-}-u_2^{j-})d\Omega,\end{gathered}$$

$$m^j(t^j,u^j) = \int_{\Omega_j} t^{jT\prime}\overline{M}u_j A_1^j A_2^j d\alpha_1 d\alpha_2, \tag{22}$$

$$t^T = (t_1,t_2),\ u^T = (u_1,u_2),\ t^{'T} = (\frac{\partial t_1}{\partial\tau},\frac{\partial t_2}{\partial\tau}),$$

$$t_p^T = (\frac{\partial t_1}{\partial\alpha_1},\frac{\partial t_1}{\partial\alpha_2},\frac{\partial t_2}{\partial\alpha_1},\frac{\partial t_2}{\partial\alpha_2})^T,\ u_p^T = (\frac{\partial u_1}{\partial\alpha_1},\frac{\partial u_1}{\partial\alpha_2},\frac{\partial u_2}{\partial\alpha_1},\frac{\partial u_2}{\partial\alpha_2})^T,$$

$$\overline{A} = \begin{pmatrix} \lambda^j h^j & 0 & \frac{\lambda^j (h^j)^2 (k_2^j - k_1^j)}{6} & 0 \\ 0 & \lambda^j h^j & 0 & \frac{\lambda^j (h^j)^2 (k_2^j - k_1^j)}{6} \\ \frac{\lambda^j (h^j)^2 (k_2^j - k_1^j)}{6} & 0 & \lambda^j \frac{h^j}{3} & 0 \\ 0 & \frac{\lambda^j (h^j)^2 (k_2^j - k_1^j)}{6} & 0 & \lambda^j \frac{h^j}{3} \end{pmatrix},$$

$$\overline{M} = \begin{pmatrix} c^j \rho^j h^j & \frac{1}{6} c^j \rho^j (h^j)^2 (k_1^j + k_2^j) \\ \frac{1}{6} c^j \rho^j (h^j)^2 (k_1^j + k_2^j) & \frac{1}{3} c^j \rho^j h^j \end{pmatrix},$$

$$G = \begin{pmatrix} a h^j & a \frac{(h^j)^2}{6} k_\Gamma \\ a \frac{(h^j)^2}{6} k_\Gamma & a \frac{h^j}{3} \end{pmatrix},$$

$$g_{1_i} = 0, \quad g_{2_i} = 1,$$

$$g_{1_i} = 0, \quad g_{2_i} = 1,$$

if the layer containing the outer surface,

$$g_{1_i} = 1, g_{2_i} = 0$$

otherwise, conjugation condition

$$T_{k_1} = T_{k_2}, \tag{23}$$

the initial condition,

$$\sum_{k \in J_3} M^{(k)}(T_k - T_k^0, u^{(k)}) + \sum_{j \in J_2} m^{(j)}(t_j - t_0^j, u^{(j)}) = 0, \ \tau = 0 \tag{24}$$

for arbitrary functions $u^k \in U_k(\Omega_k), u_1^j, u_2^j \in U_j(\Omega_j)$, where, those that implement the main junction conditions. Let us prove the following lemma.

Lemma 1. *The bilinear forms associated with the operator of the problem* (20)–(23) *are symmetric under the homogeneous boundary condition of the third kind.*

Proof. Let us prove that for bilinear forms $a^j(u,v), m^j(u,v)$ the following statements hold true:

1) the domain of the operator of the problem is a dense set.

2) $a^{(j)}(u,v) = a^{(j)}(v,u); m^{(j)}(u,v) = m^{(j)}(v,u)$.

3)The first statement is executed because $C_0^\infty(V) \subset D$.

Obviously, the implementation of the second equality for bilinear forms $a(u,v), m(u,v)$ provided by the symmetry of the matrices $\overline{A}$ and $\overline{M}$ The lemma is proved. The following theorem holds true.

Theorem 1. *Let the condition holds true:*

$$h_j |k_i^j| \le \sqrt{3}, \ j \in J_2, \tag{25}$$

Then the bilinear forms of the problem (20)-(23) *are continuous and elliptic, assuming uniform third-order boundary condition.*

Proof. First we prove, that the theorem holds true in the case $j = 1$, $j \in J_2$; $k = 1$, $k \in J_3$. Then the indices can be neglected.

The proof of the ellipticity and continuity of bilinear forms $A(u, v), M(u, v)$ is is described in [6]. In order to show that the bilinear forms $a(u, v), m(u, v)$ have the property of ellipticity, one must prove that the inequalities are true

$$m(u, u) \geq c_1\|u\|, a(u, u) \geq c_2\|u\|, \text{ where } c_1 > 0, c_2 > 0.$$

We show that the matrices $\overline{A}$ and$\overline{M}$ are positively defined. Let's find eigen values of matrices $\overline{M}$, which are the roots of the algebraic equation. They are

$$\eta_{1,2} = \frac{c_j\rho_j h_j}{3}(2 \pm \sqrt{1 + \frac{h_j^2(k_1^j + k_2^j)^2}{4}})$$

From the fact that $c(\alpha_1^j, \alpha_2^j, 0), c(\alpha_1^j, \alpha_2^j, 0)$-positive functions

$$h_j|k_i^j| \leq \sqrt{3}, i = 1, 2, \tag{26}$$

that the eigen values η_1, η_2 are positive. Then for a quadratic form $m(u, u)$ the following estimation is valid

$$\begin{gathered}\int_\Omega u^T\overline{A}ud\Omega \geq \\ \geq (\frac{\overline{c\rho}h}{3}(2 - \sqrt{1 + \frac{h^2\min_\Omega(k_1 + k_2)^2}{4}}))^2\int_\Omega(u_1^2 + u_2^2)d\Omega = \gamma_1^2\|u\|_{L_2(\Omega)},\end{gathered} \tag{27}$$

where

$$\bar{c} = \min_{(\alpha_1,\alpha_2)\in\Omega}(c), \bar{\rho} = \min_{(\alpha_1,\alpha_2)\in\Omega}(\rho),$$

$$\gamma_1^2 = \frac{\overline{c\rho}h}{3}(2 - \sqrt{1 + \frac{h^2\max_\Omega(k_1 + k_2)^2}{4}}))^2.$$

Thus, $m(u, v)$ is elliptic in space $L_2(\Omega)$ To prove continuity we use the Cauchy-Bunyakovskii inequality.

$$\begin{gathered}|m(t, u)| = |\int_\Omega t^T\overline{M}ud\Omega| = |\int_\Omega c\rho ht_1u_1d\Omega + \int_\Omega\frac{1}{6}c\rho(k_1 + k_2)h^2t_2u_1d\Omega + \\ + \int_\Omega\frac{1}{6}c\rho(k_1 + k_2)h^2t_1u_2d\Omega + \int_\Omega\frac{1}{3}c\rho ht_2u_2d\Omega| \leq \\ \leq \overline{c\rho}h\|t_1\|_V\|u_1\|_V + \frac{1}{6}\overline{c\rho}h^2|\overline{k_1} + \overline{k_2}|\|t_2\|_V\|u_1\|_V + \\ + \frac{1}{6}\overline{c\rho}h^2|\overline{k_1} + \overline{k_2}|\|t_1\|_V\|u_2\|_V + \frac{1}{3}\overline{c\rho}h\|t_2\|_V\|u_2\|_V \leq \\ \leq C_1^2(\|t_1\|_V\|u_1\|_V + \|t_1\|_V\|u_2\|_V + \|t_2\|_V\|u_1\|_V + \|t_2\|_V\|u_2\|_V) = \\ = C_1^2(\|t_1\|_V + \|t_2\|_V)(\|u_1\|_V + \|u_2\|_V) = C_1^2\|t\|\|u\|, \\ \bar{c} = \max_\Omega|c|, \bar{\rho} = \max_\Omega|\rho|, C_1^2 = \overline{c\rho}\max\{h, \frac{h(\overline{k_1} + \overline{k_2})}{6}\}, \\ \overline{k_1} = \max_\Omega|k_1|, \overline{k_2} = \max_\Omega|k_2|,\end{gathered}$$

$$\|u\| = \|u\|_{W_2^1(\Omega)}.$$

To prove the ellipticity of the bilinear form $a(u,v)$, we use the inequality (25). We first find the eigenvalues of a matrix of bilinear form

$$Q(\xi,\eta) = \lambda h\xi^2 + \frac{1}{3}h^2(k_{3-i} - k_i)\xi\eta + \frac{\lambda h}{3}\eta^2, \tag{28}$$

solving the equation for this $u^{(k)} \in U_k, u_1^{(j)}, u_2^{(j)} \in \overline{U}_j$ We obtain the following eigen values

$$y_{1,2} = \frac{2}{3}\lambda h \pm \frac{1}{3}\lambda h\sqrt{(1 + \frac{h^2(k_{3-i} - k_i)^2}{4})}. \tag{29}$$

Referencing 25,the eigen values are positive. So,

$$Q(\xi,\eta) \geq \lambda h(\frac{2}{3} - \frac{1}{3}\sqrt{(1 + \frac{h^2(k_{3-i} - k_i)^2}{4})})(\xi^2 + \eta^2). \tag{30}$$

Similarly, after finding the eigenvalues of a matrix of bilinear form

$$N(\xi,\eta) = ah\xi^2 + ah^2\frac{k_\Gamma}{3}\xi\eta + a\frac{h}{3}\eta^2 \tag{31}$$

we obtain the inequality

$$N(\xi,\eta) \geq a(\frac{2}{3} - \frac{1}{3}\sqrt{1 + \frac{h^2k_\Gamma^2}{4}})(\xi^2 + \eta^2). \tag{32}$$

Taking into account (25), (30), (32) and the Friedrichs inequality, we obtain

$$a(u,u) \geq \mu_1^2\int_\Omega((\frac{\partial u_1}{\partial \alpha_i})^2 +$$

$$+(\frac{\partial u_2}{\partial \alpha_i})^2)d\Omega + \mu_2^2\int_\Omega(u_1^2 + u_2^2)d\Omega + \mu_3^2\int_\Gamma(u_1^2 + u_2^2)d\Gamma \geq$$

$$\geq \gamma_2^2\Big(\int_\Omega((\frac{\partial u_1}{\partial \alpha_i})^2 + (\frac{\partial u_2}{\partial \alpha_i})^2)d\Omega + \int_\Omega(u_1^2 + u_2^2)d\Omega\Big) = \gamma_2^2\|u\|^2,$$

$$\mu_1^2 = \overline{\lambda}h\Big(\frac{2}{3} - \frac{1}{3}\sqrt{1 + \frac{h^2\max_\Omega(k_{3-i} - k_i)^2}{4}}\Big),$$

$$\mu_2^2 = \max\{g_{21}, g_{22}\}\,\frac{(2-\sqrt{3})^2}{4},$$

$$\mu_3^2 = \overline{a}h\Big(\frac{2}{3} - \frac{1}{3}\sqrt{1 + \frac{h^2\max_\Omega(k_\Gamma^2)}{4}}\Big),$$

$$\overline{\lambda} = \min_{(\alpha_1,\alpha_2)\in\Omega}(\lambda),\ \overline{a} = \min_{(\alpha_1,\alpha_2)\in\Gamma}(a)$$

$$\gamma_2^2 = \begin{cases}\min\{\mu_1^2, \mu_2^2\},\ \mu_2 \neq 0,\\ \min\{\frac{1}{2}\mu_1^2, \frac{1}{2}\mu_1^2\mu_4^2\},\ \mu_2 = 0\end{cases}$$

and μ_4^2 is a constant obtained from Friedrichs's inequality.

The continuity of the bilinear form follows from the following inequality

$$|a(t,u)| \leq |\int_\Omega^T \overline{A}ud\Omega| + |\int_\Gamma t^TGud\Gamma| \leq \sigma_1|\int_\Omega\frac{\partial t_1}{\partial \alpha_i}\frac{\partial u_1}{\partial \alpha_i}d\Omega| +$$

$$+\sigma_{2i}|\int_\Omega \frac{\partial t_2}{\partial\alpha_i}\frac{\partial u_1}{\partial\alpha_i}d\Omega| + \sigma_{2i}|\int_\Omega \frac{\partial t_1}{\partial\alpha_i}\frac{\partial u_2}{\partial\alpha_i}d\Omega|+$$
$$+\sigma_3|\int_\Omega \frac{\partial t_2}{\partial\alpha_i}\frac{\partial u_2}{\partial\alpha_i}d\Omega| + \sigma_4|\int_\Omega t_2u_2d\Omega|$$
$$+\sigma_5|\int_\Omega t_1u_1d\Omega| + \sigma_6|\int_\Omega t_2u_1d\Omega| + \sigma_6|\int_\Omega t_1u_2d\Omega| + \sigma_5|\int_\Omega t_2u_2d\Omega|+$$
$$+\sigma_7|\int_\Gamma t_1u_1d\Gamma| + \sigma_8|\int_\Gamma t_2u_1d\Gamma| + \sigma_8|\int_\Gamma t_1u_2d\Gamma| + \sigma_7|\int_\Gamma t_2u_2d\Gamma| \le$$
$$\le C_2(\|t_1\|_V\|u_1\|_V + \|t_1\|_V\|u_2\|_V + \|t_2\|_V\|u_1\|_V + \|t_2\|_V\|u_2\|_V) \le C_2\|t\|\|u\|,$$

where

$$\sigma_1 = \max_\Omega|\frac{\lambda h}{A_i^2}|, \quad \sigma_{2i} = \frac{1}{6}\max_\Omega|\frac{\lambda h^2}{A_i^2}(k_{3-i}-k_i)|,$$
$$\sigma_3 = \frac{1}{3}\max_\Omega|\frac{\lambda h}{A_i^2}|, \quad \sigma_4 = \frac{1}{3}\max_\Omega|\frac{4\lambda}{h}|,$$
$$\sigma_5 = \max_\Omega(1+k_1k_2\frac{h^2}{4}), \sigma_6 = \frac{h}{2}\max_\Omega(k_1+k_2), \sigma_7 = \max_\Gamma ah, \sigma_8 = \max_\Gamma|\frac{ah^2k_\Gamma^2\lambda}{6}|.$$

We have proved the properties of the bilinear forms for any thin and ordinary layer. However, these propositions hold true for the operator of a problem for a multilayered medium, since they are executed for a single term, which derives that they will be executed for the sum in the formula. Note that the inequality (25) can be considered as a criterion for the thin layer. The theorem is proved.

6. Approximation of the solution

To solve the beforementioned variational problem, we discretize the solution in spatial variables. In this case, for sampling functions $T_j(\alpha_1^j, \alpha_2^j, \alpha_3^j, \tau)$ we apply the approximations of the semi-analytic finite elements method and for functions $t_1^{(j)}(\alpha_1^j, \alpha_2^j, \tau), t_2^{(j)}(\alpha_1^j, \alpha_2^j, \tau)$ are the approximations of the finite elements method. According to these methods, we choose the approximation spaces $\{V_h\}$ from space V so that

$$dimV_h \longrightarrow \infty, h \longrightarrow 0,$$
$$\bigcup V_h\text{-tightly enclosed in V.}$$

We will present the unknown functions in the

$$T_j(\alpha_1^j, \alpha_2^j, \alpha_3^j, \tau) = \sum_{k=1}^{M}\sum_{i=1}^{N} T_{ki}^j(\tau)\widetilde{\psi_k}(\alpha_3^j)\widetilde{\phi}_i(\alpha_1^j, \alpha_2^j), \tag{33}$$

$$t_1^{jh}(\alpha_1^j, \alpha_2^j, \tau) = \sum_{i=1}^{N} t_{1i}^j(\tau)\widetilde{\phi}_i(\alpha_1^j, \alpha_2^j), \tag{34}$$

$$t_2^{jh}(\alpha_1^j, \alpha_2^j, \tau) = \sum_{i=1}^{N} t_{2i}^j(\tau)\widetilde{\phi}_i(\alpha_1^j, \alpha_2^j), \tag{35}$$

where $\widetilde{\psi_k}, \widetilde{\phi}_i(\alpha_1^j, \alpha_2^j)$ are basic functions, $T_{ki}^j, t_{1i}^j, t_{2i}^j$−− are unknown coefficients. To approximate the desired solution for the third spatial coordinate, we use the expansion of the desired function in a series of functions-"bubbles". These functions on the interval [-1,1] are given by the relations

$$\widetilde{\psi_1}(\xi) = \frac{1+\xi}{2},\ \widetilde{\psi_2}(\xi) = \frac{1-\xi}{2},\ \widetilde{\psi_i}(\xi) = \Phi_{i-1}(\xi), i = 3, 4...;$$

$$\Phi_{i-1}(\xi) = \sqrt{\frac{2i-1}{2}} \int_{-1}^{\xi} P_{i-1}(t)dt. \tag{36}$$

Here $P_i(t)$ are known Legendre polynomials. It is convenient to use the recurrence formula for calculations

$$\Phi_j(\xi) = \frac{1}{\sqrt{2}(2j-1)}(P_j(\xi) - P_{j-2}(\xi)). \tag{37}$$

The property of the orthogonality of the Legendre polynomial follows an important property of internal forms

$$\psi_i(-1) = \psi_i(-1) = 0,\ i = 3, 4, ... \tag{38}$$

External forms allow to calculate solutions at the borders. It is essentially used for convenient and easy implementation of junction conditions with other areas. In addition, this system of functions has favorable properties in terms of numerical stability. In order to approximate the time-domain solution, we propose to use the well-known Crank-Nicholson difference scheme [6].

7. Numerical example

Based on the constructed heterogeneous mathematical model and the proposed numerical approximations, a program complex was created in the language C# that implements this approach. A series of computational experiments was conducted using it.

Let us consider the problem of stationary heat conductivity in an axisymmetric infinite hollow cylinder with a thin outer coating. The problem is to find the distribution of the function of temperature, if it is known that on the outer and inner parts of the cylinder surface there is a heat exchange according to Newton's law with different values of the temperature of the medium. Coefficients of thermal conductivity of the coating are $\lambda_1 = const$, massive part – $\lambda_2 = const$. For the analysis of stationary heat conductivity in a cylinder, a stationary analogue of the proposed mathematical model is used. Since boundary conditions and geometry do not depend on spatial coordinates the solution of the problem will depend only on one coordinate. This allows us to get rid of the dependence on the relationship between the parameters of the finite-element grid along two coordinate axes, to carry out only the P-adaptive refinement in the radial direction and to investigate its influence on the resulting solution. The mathematical model of the described problem is a system of ordinary differential equations. To analyze its numerical solution, we first find the analytic solution of the classical mathematical model without taking into account the small thickness of the layer.

$$\varepsilon = \frac{\max_V |T_i - T_a n|}{\max_V |T_a n|} 100. \tag{39}$$

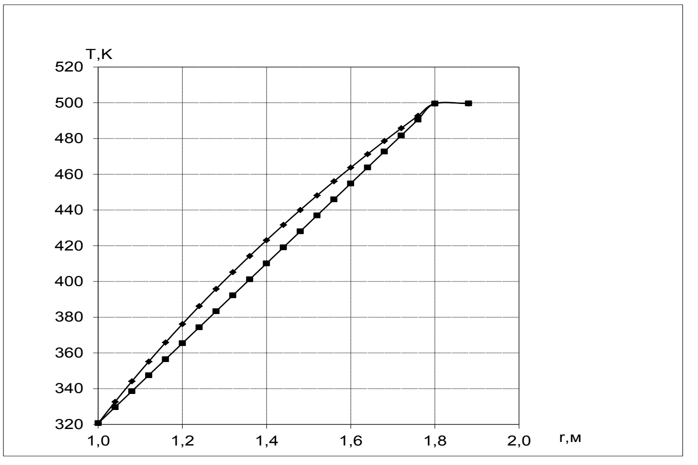


FIG. 2. Graphs of the function of temperature distribution using different numbers of members in expansion in thickness. (The curve with diamonds – analitical solution T_{an}, the curve with squares – numerical solution T_i for $p = 2$)

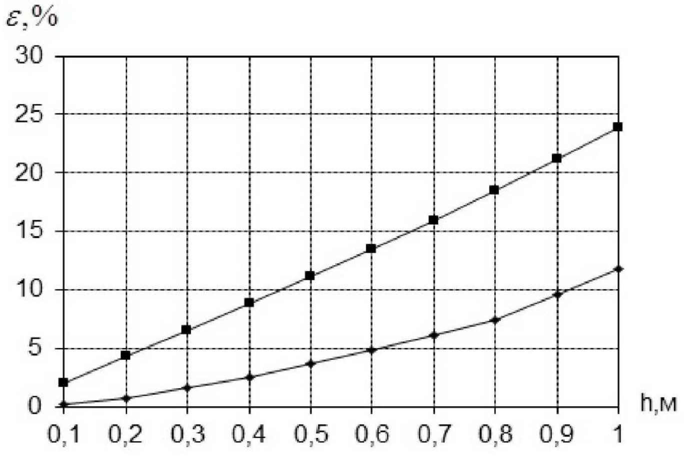


FIG. 3. Charts of the absolute error dependence on the thickness of the thin cylindrical layer (1. $\lambda = 385\frac{Dg}{Kms}$, 2. $\lambda = 3.85\frac{Dg}{Kms}$)

This solution is used to compare the results calculated using the algorithm proposed in the work. In the computational experiment, the effect of the content of a different number of members in the sum (33) was investigated to approximate the solution in thickness. Experiment results are shown in Fig.2.

The "Analytical solution" curve of this figure corresponds to the analytical solution, and the curve "Numerical solution" shows results, obtained with the

TABL. 1. The dependence of the relative error on the content of a different number of basic functions over the thickness of the layer

Number of polynomials	Relative error
2	3,1931
3	0,3321
4	0,0333
5	0,0034
6	0,0040
7	0,0001

algorithm using 2 members of the expansion for the thickness of the lower layer of the cylinder. In a numerical experiment, the solution of the model was also studied in the case of preserving 3, ..., 7 members of the decomposition. The graphs of the obtained solutions in the current scale almost coincide. As it should be expected, with increasing order of approximation, the graph of the numerical solution goes to the analytic solution, which confirms the theoretical conclusion about the convergence of the proposed algorithm. To confirm this, as a criterion for the analysis of approximate solutions (Fig. 3), the relative error rate is used

Here T_{an} is the analytical solution of the problem, T_i is the numerical solution. Table 1 it shows its decline, depending on the increase in the members of the schedule.

8. Conclusion

The sugested heterogeneous model allows to effectively analyze the process of thermal conductivity in multilayer environments, since it avoids the difficulties associated with the application of numerical methods.

Bibliography

1. Dejneka V. S. Models and methods for solving problems with conjugation conditions / I. V. Sergienko, V. S. Dejneka, V. S. Skopetsij. – V. M. Glushkov Institute of Cybernetics of NAS of Ukraine. – Kyiv: Naukova Dumka. – 1998. – 614 p.
2. Savula Ya. The investigation of the variational problem of heat-transfer in multilayered environment with thin inclusions / Ya. Savula, L. Diakoniuk // Visnyk of the Lviv University. — Series Applied Mathematics and Computer Science. – Lviv. -–2001. – Iss. 3. – P. 125-130. (in Ukrainian).
3. Savula Ya. Numerical analysis of advection-diffusion in the continuum with thin canal. / Ya. Savula, V. Koukharskyi, Ye. Chaplia // Numerical Heat Transfer. – Part A: Applications: An International Journal of Computation and Methodology. – 1998. – Vol. 33, Iss. 3. – P. 341-351.
4. Gavrysh V. Nonlineary boundary-value problem of heat conduction for a layered plate with inclusion / V. Gavrysh // Material science. – 2015. – Vol. 51, № 3. – P. 331-339.
5. Pidstrygach Ya. Selected Works // Ya. Pidstrygach. – Kiev: Naukova Dumka, 1995. – 450 p.

6. Savula Ya. Numerical analysis of the problems of mathematical physics by variational methods / Ya. Savula. – Lviv: Publishing Center of I.Franko National University of Lviv, 2004. – 221 p.
7. Quarteroni R. Multifields modelling in numerical simulation of partial differential equations / R. Quarteroni // Multifields modelling in numerical simulation of partial differential equations. – GAMM-Mitteilungen. – 1996. – Vol. 1. – P. 45-63.

Liliia Diakoniuk, Yarema Savula,
Faculty of Applied Mathematics and Informatics,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine.

Received 17.10.2017; revised 03.11.2017

Журнал обчислювальної та прикладної математики | 2017 № 3 (126) | Journal of Numerical & Applied Mathematics

UDC 519.632.4

ON THE APPLICATION OF THE ONE *HP*-ADAPTIVE FINITE ELEMENT STRATEGY FOR NONSYMMETRIC CONVECTION-DIFFUSION-REACTION PROBLEMS

R. G. Drebotiy, H. A. Shynkarenko

Резюме. Ми розглядаємо застосування однієї hp-адаптивної стратегії методу скінченних елементів до розв'язування несиметричних крайових задач конвекції-дифузії-реакції. В основі розглядуваної стратегії лежить процедура вибору на кожному скінчнному елементі між збільшенням його порядку чи поділом, що базується на порівнянні норм наближень до похибки для розглядуваних способів перебудови скінченного елемента. Ми розглядаємо алгоритм адаптування та наводимо обґрунтування ідеї алгоритму у випадку симетричної крайової задачі. Застосовність алгоритму до несиметричних задач ми аналізуємо шляхом розгляду результатів числових експериментів, а також доповнюємо наведені результати теоретичним аналізом можливості зведення вихідної варіаційної задачі до симетричної форми. Ми наводимо дві процедури, що дають змогу перейти від несиметричної задачі до еквівалентної симетричної, або до послідовності симетричних задач, послідовність розв'язків яких збігається до розв'язку вихідної несиметричної задачі. Отриманий результат врешті може бути використаний для побудови комбінованих алгоритмів на основі однієї із схем симетризації та алгоритму hp-адаптування.

Abstract. We consider application of certain hp-adaptive strategy for finite element method for solving nonsymmetric convection-diffusion-reaction boundary value problems. In the base of described strategy lies refinement selection procedure which is used to choose on each finite element between degree increase or bisection. It uses special comparative criteria for norms of approximation to local errors on different refinement patterns. We present the adaptation algorithm itself and proof of idea behind it for symmetric problems. For the case when problem is nonsymmetric we provide corresponding analysis of numerical experiments and also we add pure theoretical analysis of the possibility of bringing given variational problem to symmetric form, taking into account that the algorithm is naturally applicable in the latter case. We describe two approaches that can provide transition from nonsymmetric variational problem to directly equivalent symmetric problem in the first approach or to sequence of symmetric problems, solutions of which forms sequence of functions that is convergent to the solution of initial nonsymmetric problem in the second approach. Obtained result can be used to build algorithms, based on a combination of one of the described symmetrization methods with hp-adaptive scheme.

Key words. Convection-diffusion-reaction problem, finite element method, a posteriori error estimator, adaptive strategy, hp-adaptivity, nonsymmetric problem.

1. Introduction

Space mesh adaptivity today is the major technique which is used to optimize the process of finding the approximate solution by finite element method in various free and commercial engineering simulation tools. Using it also is crucial, since in most cases the nature of considered boundary problem is characterized by highly nonuniform distribution of local errors in the case of uniform mesh. In the context of modeling of convection-diffusion-reaction phenomena, the reason of such error distribution lies in relatively large values of Péclet and Strouhal numbers for the given problem.

Special and natural attention is on so-called hp-adaptive methods [2, 4, 5, 8–10], since they provide most wide approximation capabilities by using both space mesh adaptivity (h-) and element polynomial degree adaptivity (p-). Despite that there are reasonable facts to believe that such algorithms (hp-) can be considered "exotic" in some sense, investigation in that field is still important, since it is proved [8] that there is possibility to obtain exponentially convergent sequence of approximations by using hp-refined meshes.

In this paper we study the possibility of application of hp-adaptive strategy, introduced in [5], to nonsymmetric variational problems. The fact is that the nature of introduced algorithm can be explained only for problems with self-adjoint operators. Despite this, in practice, it can be seen, that algorithm still can be used for nonsymmetric problems which is shown in provided numerical example. The goal is of this example is to demonstrate that algorithm can provide solid results, regardless of the used a posteriori error estimators or adaptation criteria. The second part of this work is the pure theoretical investigation of the possible methods of symmetrization of nonsymmetric variational problems.

The paper structure is the following: in section 2 we define model problem; in section 3 we construct variational formulation; in section 4 we present hp-adaptation algorithm and discuss the main idea behind it; in section 5 we extend algorithm with some specific error estimator; in section 6 we review adaptation criteria which we will use in numerical experiment; in section 7 we provide numerical results for direct application of described algorithm and in section 8 we study two methods of symmetrization of variational problem. Final conclusions are given in section 9.

2. Model boundary value problem

Let us consider the following boundary value problem:

Find function $u = u(x)$ such that

$$\begin{cases} -\left(\mu u'\right)' + \beta u' + \sigma u = f \text{ in } \Omega = (0, L) \\ (\mu u')\big|_{x=0} = \alpha[u(0) - \bar{u}_0], \;\; -(\mu u')\big|_{x=L} = \gamma[u(L) - \bar{u}_L], \end{cases} \tag{1}$$

where

$$\alpha, \gamma \geq 0, \;\; \mu = \mu(x) \geq \mu_0 > 0, \;\; \beta(0) \leq 0, \;\; \beta(L) \geq 0, \;\; \sigma = \sigma(x) \geq 0,$$
$$\sigma(x) - \beta'(x)/2 \geq \sigma_0 > 0 \text{ almost everywhere in } (0, L), \tag{2}$$

$$\mu, \beta, \sigma \in L^{\infty}(0, L), \ f \in L^2(0, L).$$

Considered problem is used in analysis of ecologic phenomena, semiconductors, biology etc. Many real problems of such kind are singularly perturbed [3]. In the terms of differential equation parameters it means that coefficients near highest order derivatives are relatively small in comparison to others. So in this case a second order equation is almost degenerated to first order one. In combination with standard boundary conditions it causes existence of layers near domain's boundary with high solution gradient. Those boundary layers are making the solving of problem by using well-known uniform-mesh-based FEM quite difficult. Such conditions leads to large Péclet and Strouhal criteria and to nonuniform local error distribution.

3. Variational formulation

Using standard approach [1], we can simply define variational problem corresponding to (1): find solution $u \in V$, such that

$$a(u, v) = \langle l, v\rangle \quad \forall v \in V, \tag{3}$$

where

$$\begin{aligned} a(u, v) &:= \int_0^L [\mu u'v' + \beta u'v + \sigma uv]\, dx + \alpha u(0)v(0) + \gamma u(L)v(L), \\ \langle l, v\rangle &:= \int_0^L fv\, dx + \alpha \bar{u}_0 v(0) + \gamma \bar{u}_L v(L), \ \forall u, v \in V := H^1(0, L). \end{aligned} \tag{4}$$

Under conditions (2) problem data satisfies (for details see [6]) conditions of Lax-Milgram theorem [1] and therefore this variational problem is well-posed.

For further needs, let us define energy norm $\|v\|_E = \sqrt{a(v, v)}$.

To discretize obtained variational problem we use general finite element method with high-order polynomial basis functions. In other words, we define some space $V_h \subset V, \dim V_h < +\infty$, of piecewise-polynomial functions and find finite element approximation $u_h \in V_h$ as a solution of variational equation:

$$a(u_h, v_h) = \langle l, v_h\rangle \quad \forall v_h \in V_h. \tag{5}$$

Now if we construct finite basis $\{\varphi_i\}_{i=1}^n$ of space V_h then by expanding $u_h = \sum_{i=1}^n q_i\varphi_i$, where $q_i \in \mathbb{R}, i = \overline{1, n}$ we can clearly see, that (5) is equal to the following system of algebraic linear equations for $q_i, i = \overline{1, n}$:

$$\sum_{i=1}^n q_i a(\varphi_i, \varphi_j) = \langle l, \varphi_j\rangle \quad j = \overline{1, n}. \tag{6}$$

For general reference see [2, 9, 10].

4. *hp*-ADAPTATION ALGORITHM

In this section we briefly present discussion and review of algorithm from [5].

Let us consider finite element mesh $\tau_h = \{K = (x_{k-1}, x_k)\}_{k=1}^n$ where $0 = x_0 < x_1 < \cdots < x_n = L$. Let us define global error approximation space in the form:

$$E_h = \bigoplus_{K \in \tau_h} E_h^K, \tag{7}$$

where space of functions $E_h^K = \{v \in V \,|\, \operatorname{supp} v \subset K\}$ and $\dim E_h^K < +\infty$. Let us define the following variational problem for error approximation:

$$\begin{cases} \text{find } e_h \in E_h \text{ such that} \\ a(e_h, v_h) = \displaystyle\int_\Omega R[u_h] v_h dx \quad \forall v_h \in E_h, \end{cases} \tag{8}$$

where R is the residual:

$$R[u_h] := f - \left(\mu {u_h}'\right)' - \beta {u_h}' - \sigma u_h. \tag{9}$$

It is not hard to see that problem (8) can be decomposed per elements. For each element we have to solve a problem:

$$\begin{cases} \text{find } e_h^K \in E_h^K \text{ such that} \\ a(e_h^K, v_h^K) = \displaystyle\int_K R[u_h] v_h^K dx \quad \forall v_h^K \in E_h^K \end{cases} \tag{10}$$

and then $e_h = \sum_{K \in \tau_h} e_h^K$.

Consider now the case $\beta \equiv 0$, i.e. the problem has symmetric bilinear form. Then the following well-known equality holds:

$$\|u - u_h\|_E^2 = \|u\|_E^2 - \|u_h\|_E^2. \tag{11}$$

Since error estimation problem has the same bilinear form as the original, then for finite element error approximation e_h the equality above also holds:

$$\|e - e_h\|_E^2 = \|e\|_E^2 - \|e_h\|_E^2. \tag{12}$$

From this equality we see that if energy norm of error approximation increases than also increases accuracy of this approximation. Denote the finite element solution on the current mesh as $u_h \in V_h$ and corresponding error $e = u - u_h$. Then (12) we can rewrite as

$$\|u - (u_h + e_h)\|_E^2 = \|u - u_h\|_E^2 - \|e_h\|_E^2. \tag{13}$$

Let us find finite element solution $\tilde{u}_h$ in space $\tilde{V}_h = V_h + E_h \subset V$, where E_h is the error approximation space, defined in (7). For symmetric case we have well-known optimality inequality:

$$\|u - \tilde{u}_h\|_E \leq \|u - \tilde{v}_h\|_E, \ \forall \tilde{v}_h \in \tilde{V}_h. \tag{14}$$

Using now (13), and the fact that $u_h + e_h \in \tilde{V}_h$ we have:

$$\|u - \tilde{u}_h\|_E^2 \leq \|u - u_h\|_E^2 - \|e_h\|_E^2. \tag{15}$$

Decomposing the second term in the right part we obtain inequality:

$$\|u-\tilde{u}_h\|_E^2 \le \|u-u_h\|_E^2 - \sum_{K\in\tau_h} \left\|e_h^K\right\|_E^2. \tag{16}$$

Consider now decomposition of approximation space V_h into local approximation spaces V_h^K, $K\in\tau_h$. Spaces $V_h^K+E_h^K$ are considered as refined local finite element spaces according to transition from current mesh to mesh defined by space $\tilde{V}_h$. In the case when E_h^K consists of piecewise-polynomial functions it directly defines some refinement pattern on element K. For each element K we can consider now several different choices of space E_h^K: $E_1,\dots,E_S$ and taking into account (16) we see, that it is optimal to use refinement pattern defined by the space $E_h^K := E_{s_K}$, $s_K\in\{1,\dots,S\}$ which gives a maximum to a value of $\left\|e_h^K\right\|_E$ in the right part of (16).

So, now we can review the entire algorithm, which consists of two phases:

Initialization:

Compute:

$$\begin{aligned} \mu_0 &= \min_{x\in[0,L]} \mu(x), \\ \sigma_0 &= \min_{x\in[0,L]} \left\{\sigma(x) - \frac{\beta'(x)}{2}\right\}, \\ C &= 2\cdot\left[\min\{\mu_0,\,\sigma_0\}\right]^{-1/2}. \end{aligned} \tag{17}$$

Set τ_h to some initial finite element mesh.

For each finite element $K=(x_{k-1},x_k)\in\tau_h$ we define quadratic bubble function

$$\omega_K(x) := (x_k-x)(x-x_{k-1}). \tag{18}$$

TOL is acceptable relative error level in percent.

p_{max} is the maximum supported degree of polynomial basis function on finite element.

$\theta\in(0,1)$ is fixed value.

Iteration:

Step 1: Find FEM solution u_h on the current mesh τ_h. Define u_h^K as restriction of u_h to the element K and $p_K := \deg(u_h^K)$.

Step 2: For all elements $K\in\tau_h$ compute

$$\eta_K = \frac{C}{\sqrt{p_K(p_K+1)}} \left\|\sqrt{\omega_K} R[u_h]\right\|_{L^2(K)}. \tag{19}$$

Define $\eta := \sqrt{\sum_K \eta_K^2}$.

Then if $\frac{\eta}{\|u_h\|_E}\times 100\% < TOL$ we stop the algorithm, else:

Step 3: Choose elements for refinement.

Compute $\eta_{\max} = \max\limits_K \eta_K$.

We will change those elements K, for which $\eta_K > (1-\theta)\eta_{\max}$. The set of all selected elements we name as A_θ.

Step 4: Mesh modification. For all selected elements $K=(x_{k-1},x_k)\in A_\theta$ choose between bisection and increasing of polynomial degree on it by 1.

Step 4a: If $p_K = p_{\max}$ then we divide element into two with orders (p_K, p_K), otherwise:
Step 4b: Define $X^p(a,b)$ as a space of all polynomials of order p on closed interval $[a,b]$.

Define spaces:

$$\begin{aligned} V^1_{hp}(K) = \{v \in C(K) | v \in X^{p_K}(x_{k-1}, (x_{k-1}+x_k)/2), \\ v \in X^{p_K}((x_{k-1}+x_k)/2, x_k),\, v|_{\partial K} = 0\} \\ V^2_{hp}(K) = \{v \in X^{p_K+1}(K)|\; v|_{\partial K} = 0\}. \end{aligned} \tag{20}$$

Now we solve problem (10) for $E_h^K := V^1_{hp}(K)$ and $E_h^K := V^2_{hp}(K)$. Let us denote obtained solutions as e_h^1 and e_h^2 respectively.

Compute $r_m = \|e_h^m\|_E, \;\; m = 1,2$

Step 5: Consider the difference $\Delta = r_2 - r_1$.
If $\Delta > \delta$ where δ is predefined value, then we increase element degree by 1, otherwise we bisect it into two elements with approximation polynomial degrees (p_K, p_K).
Step 6: Go to **Step 1**.

Idea of described algorithm is clear for symmetric problems. Some numerical experiments are available in [5,6]. Technically we can run algorithm on nonsymmetric problems too, without having any theoretical background in that case. We will try to perform some numerical experiments to show how described algorithm will work in practice for nonsymmetric problem. We describe additional error estimator in next section 5 and additional adaptation criteria in section 6. Using those we will provide corresponding comparative numerical results in section 7 to show that algorithm can provide solid results despite of which combination of estimator and adaptation criteria we use.

5. Error estimator based on fundamental solution

For error indicator η_K, introduced by (19) in section 4, instead of using explicit formula we can use implicit indicator in the form of problem (10) but with special approximation space $E_h^K = \mathrm{span}\{\varphi_K\}$, where:

$$\varphi_K(x) = \begin{cases} c_{11}\varphi_{11}(x) + c_{12}\varphi_{12}(x) & \text{on } x \in [x_{k-1}, x_{k-1/2}], \\ & \varphi_1(x_{k-1}) = 0, \varphi_1(x_{k-1/2}) = 1, \\ c_{21}\varphi_{21}(x) + c_{22}\varphi_{22}(x) & \text{on } x \in [x_{k-1/2}, x_k], \\ & \varphi_2(x_{k-1/2}) = 1, \varphi_2(x_k) = 0, \end{cases} \tag{21}$$

and $\{\varphi_{1i}(x)\}, \{\varphi_{2i}(x)\}$ are the sets of fundamental solutions for equations

$$-\left(\tilde{\mu}_i w'\right)' + \tilde{\beta}_i w' + \tilde{\sigma}_i w = 0, \; i = \overline{1,2} \tag{22}$$

with constant coefficients (selected as mean values of corresponding functions) on corresponding intervals $[x_{k-1}, x_{k-1/2}]$ and $[x_{k-1/2}, x_k]$. Then we solve (10) and use the energy norm of obtained approximation as an error indicator η_K. To find fundamental solutions we solve corresponding quadratic equations

$$-\tilde{\mu}_i \lambda_i^2 + \tilde{\beta}_i \lambda_i + \tilde{\sigma}_i = 0, \; i = \overline{1,2}. \tag{23}$$

Here for each of two equations we have three cases possible:

i. if $\lambda_i^{(1)}, \lambda_i^{(2)} \in \mathbb{R}, \lambda_i^{(1)} \neq \lambda_i^{(2)}$ then
$\varphi_{i1}(x) = \exp(\lambda_i^{(1)}x)$, $\varphi_{i2}(x) = \exp(\lambda_i^{(2)}x)$;

ii. if $\lambda_i^{(1)}, \lambda_i^{(2)} \in \mathbb{R}, \lambda_i^{(1)} = \lambda_i^{(2)}$ then
$\varphi_{i1}(x) = \exp(\lambda_i^{(1)}x)$, $\varphi_{i2}(x) = x\exp(\lambda_i^{(1)}x)$;

iii. if $\lambda_i^{(1)}, \lambda_i^{(2)} \in \mathbb{C}\backslash\mathbb{R}, \lambda_i^{(1)} = \alpha + \beta i, \lambda_2 = \alpha - \beta i$ then
$\varphi_{i1}(x) = \exp(\alpha x)\sin(\beta x)$, $\varphi_{i2}(x) = \exp(\alpha x)\cos(\beta x)$.

6. Element selection criteria

In addition to adding new estimator in previous section, we also will try to run algorithm with different adaptation criteria, used in step 3 to choose elements for refinement procedure. So we will have two criteria:

i. (*"maximum"* criteria) element K is refined if

$$\eta_K > (1-\theta)\eta_{max}, \tag{24}$$

where $\eta_{max} = \max\limits_K \eta_K$ and $\theta \in (0,1)$ is fixed value;

ii. (*"average"* criteria) element K is refined if

$$\frac{\sqrt{N}\eta_K}{\sqrt{\|u_h\|_E^2 + \sum_{K'} \eta_{K'}^2}} 100\% > \varepsilon, \tag{25}$$

where ε is is acceptable tolerance in % for average error level over finite element, N is element count.

7. Numerical example

We consider boundary value problem (1) with the following data

$$\begin{aligned} &\mu = 0.01, \beta = 100.896(x-1)^3, \sigma = 84(2-(x-1)^2), f = 200, \\ &\alpha = \gamma = 10^{14}, \bar{u}_0 = \bar{u}_L = 0, L = 2. \end{aligned} \tag{26}$$

Algorithm parameters are: $TOL = 5\%, p_{max} = 3, \delta = -150, \theta = 0.6, \varepsilon = 20$.

Fig. 1 demonstrates approximation obtained by introduced algorithm using fundamental solution error indicator "maximum" adaptation criteria. Taking into account boundary conditions we can clearly see that we have two boundary layers in the both ends of interval (which we don't see directly in the plot according to very large gradient of approximation near those two points). In tables 1 and 2 we present convergence history for different combinations of introduced error estimators from sections 5 and 4 in combination with "maximum" criteria (24) and "average" criteria (25). Average convergence rate is found using least squares method.

In general we can see from provided numerical examples that:

i. the better choice in according to count of elements, iterations and d.o.f. reached is a combination of the explicit indicator and "maximum" criteria;

ii. there is no large difference between "maximum" and "average" selection criteria;

iii. if we need to have almost monotonic relative error decreasing we need to choose explicit indicator from 4.

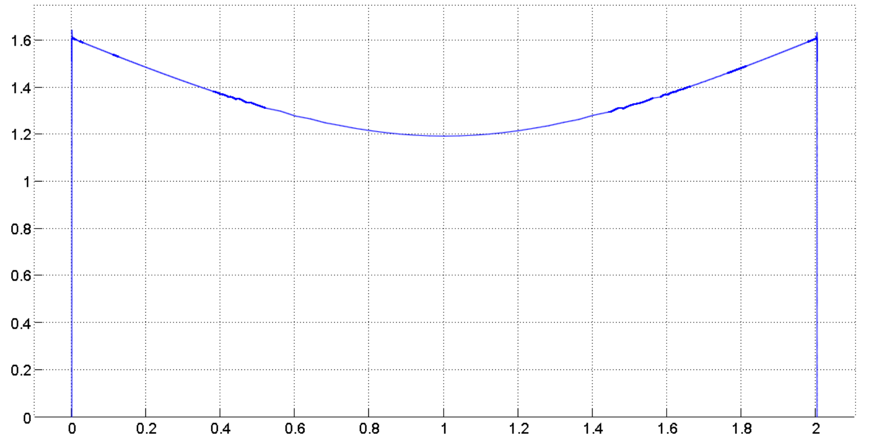


FIG. 1. Approximation to solution of problem with data (26) using implicit error indicator based on fundamental solution basis which was introduced in section 5 combined with the "maximum" criteria (24)

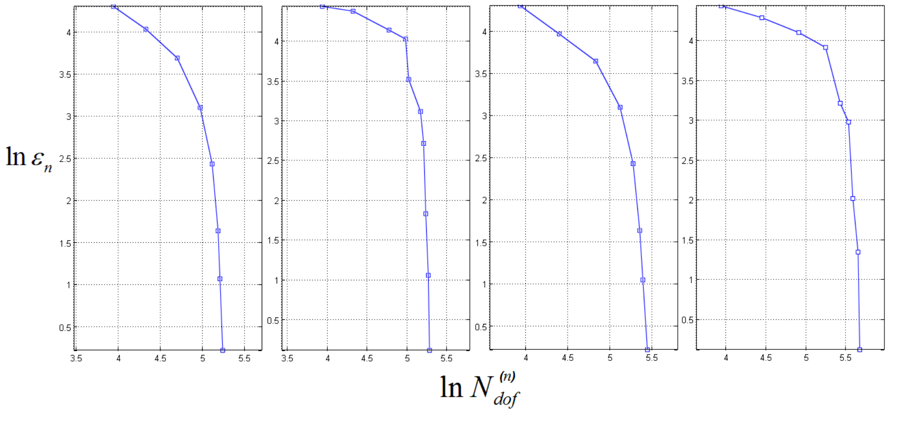


FIG. 2. Dependency between absolute error indicator ϵ_n and number of degrees of freedom $N_{dof}^{(n)}$ in log-log scale for previous results: a) for algorithm with explicit error indicator from section 4 and "maximum" criteria (24); b) for algorithm with indicator based on fundamental solution described in section 5 and "maximum" criteria (24); c) for algorithm with explicit error indicator from section 4 and "average" criteria (25); d) for algorithm with indicator based on fundamental solution described in section 5 and "average" criteria (25)

TABL. 1. Convergence history for problem with data (26) for the "maximum" criteria (24): n is an iteration number, N element count, $N_{dof}^{(n)}$ count of degrees of freedom, $\epsilon_n = \eta$ absolute error indicator, $r_n = \eta \|u_h\|_E^{-1} \times 100\%$ relative error, $p_n = -(\ln \epsilon_n - \ln \epsilon_{n-1}) \times \left(\ln N_{dof}^{(n)} - \ln N_{dof}^{(n-1)}\right)^{-1}$ rate of convergence

Explicit indicator						Fundamental solution indicator					
n	N	$N_{dof}^{(n)}$	ϵ_n	r_n	p_n	n	N	$N_{dof}^{(n)}$	ϵ_n	r_n	p_n
0	50	51	74.00	73.85		0	50	51	84.41	84.25	
1	72	75	56.51	50.94	0.69	1	69	75	79.52	67.86	0.15
2	106	109	40.08	32.25	0.91	2	102	118	62.69	52.61	0.52
3	136	143	22.26	21.22	2.16	3	124	145	56.07	49.96	0.54
4	144	165	11.39	15.61	4.68	4	130	151	33.69	33.22	12.56
5	144	177	5.14	17.90	11.33	5	142	175	22.55	37.16	2.72
6	144	181	2.90	8.39	25.48	6	142	182	15.10	43.72	10.21
7	146	187	1.24	4.57	26.08	7	143	187	6.24	19.58	32.59
						8	145	193	2.88	11.29	24.49
						9	146	196	1.12	4.72	61.11
average rate of convergence 2.66						average rate of convergence 2.38					

TABL. 2. Convergence history for problem with data (26) for the "average" criteria (25).

Explicit indicator						Fundamental solution indicator					
n	N	$N_{dof}^{(n)}$	ϵ_n	r_n	p_n	n	N	$N_{dof}^{(n)}$	ϵ_n	r_n	p_n
0	50	51	74.00	73.85		0	50	51	84.41	84.25	
1	72	81	52.93	52.15	0.72	1	72	85	72.67	71.90	0.29
2	106	125	38.34	32.55	0.74	2	106	135	60.36	50.78	0.40
3	134	167	22.12	21.43	1.89	3	136	189	50.07	48.36	0.55
4	142	195	11.33	15.90	4.31	4	144	227	24.87	35.20	3.81
5	142	211	5.11	18.80	10.08	5	144	252	19.63	70.70	2.26
6	142	219	2.84	8.02	15.76	6	144	266	7.52	21.09	17.73
7	146	231	1.24	4.57	15.56	7	150	284	3.84	14.18	10.26
						8	152	290	1.13	4.75	58.56
average rate of convergence 2.32						average rate of convergence 1.86					

Also, taking into account, that during preparation of this paper the algorithm was tested on several other problems, we can conclude from solid numerical results that the algorithm is applicable in practice in the case of nonsymmetric problems too, despite of which indicators or element selection criteria we use (without any theoretical background). In the next section we provide some pure theoretical analysis in that case.

8. Symmetrization methods

Instead of trying to generalize somehow (11) to nonsymmetric problems to bring similar argument as in remark in section 4, it is natural to try to construct equivalent (in some sense) to (3) but symmetric variational problem.

Here we present two pure theoretical results which can not be used in practice directly but can be considered as a starting point in further investigation in described direction.

8.1. **Equivalent symmetric problem approach.** Let us recall variational equation (1) in expanded form:

$$\begin{aligned}&\int_0^L [\mu u'v' + \beta u'v + \sigma uv]\,dx + \alpha u(0)v(0) + \gamma u(L)v(L) = \\ &= \int_0^L fv\,dx + \alpha\bar{u}_0 v(0) + \gamma\bar{u}_L v(L), \qquad \forall v \in V.\end{aligned} \tag{27}$$

We are free to choose arbitrary function v in (27) in the form: $v = zw$, where both functions z and w are arbitrary, but z is fixed. After substitution into (27) and small algebra we obtain equivalend equation:

$$\begin{aligned}&\int_0^L [\mu z u'w' + \underbrace{(\mu z' + \beta z)}\, u'w + \sigma zuw]\,dx + \\ &+ \alpha z(0)u(0)w(0) + \gamma z(L)u(L)w(L) = \\ &= \int_0^L fzw\,dx + \alpha\bar{u}_0 z(0)w(0) + \gamma\bar{u}_L z(L)w(L), \qquad \forall w \in V.\end{aligned} \tag{28}$$

Lets choose z as a solution of the ordinary differential equation $\mu z' + \beta z = 0$. It is not hard to find partial solution:

$$z(x) = \exp\left\{-\int_0^x \frac{\beta(\xi)}{\mu(\xi)}d\xi\right\}. \tag{29}$$

Substituting (29) into (28) lead us to:

$$\begin{aligned}&\int_0^L [\mu z u'w' + \sigma zuw]\,dx + \alpha u(0)w(0) + \gamma z(L)u(L)w(L) = \\ &= \int_0^L fzw\,dx + \alpha\bar{u}_0 w(0) + \gamma\bar{u}_L z(L)w(L), \qquad \forall w \in V.\end{aligned} \tag{30}$$

It is not hard to see that (3) and (30) are equivalent and furthermore the bilinear form

$$b(u,v) := \int_0^L [\mu z u' w' + \sigma z u w]\, dx + \alpha u(0)w(0) + \gamma z(L)u(L)w(L), \tag{31}$$

in the left part of (30), is symmetric. Corresponding to (30) boundary value problem is:

$$\begin{cases} \textit{find function } u = u(x), \textit{ such that} \\ -\left(\mu z u'\right)' + \sigma z u = fz \text{ on } \Omega = (0,L) \\ (\mu z u')\big|_{x=0} = \alpha[u(0) - \bar{u}_0],\ -(\mu z u')\big|_{x=L} = \gamma z(L)[u(L) - \bar{u}_L]. \end{cases} \tag{32}$$

Visual simplicity of obtained symmetrization procedure and the problem (32), in practice lead us to problem which is technically hard to solve. The reason is in function z (29). Fraction $\frac{\beta(\xi)}{\mu(\xi)}$ is almost proportional to Péclet number for the given problem and in the latter is singular perturbed multiplier z will be the exponent with large negative power. In such conditions it is very problematically to calculate integrals from (30) when we use standard Galerkin discretization according to very large quadrature round-off errors. We investigated numerically the following approaches:

i. trapezoidal rule;
ii. interpolation-type quadrature based on L-splines;
iii. asymptotic formula at $Pe \to +\infty$;
iv. $\tanh - \sinh$ quadratures;
v. adaptive quadratures using previous methods;
vi. implementation of adaptation algorithm using Wolfram Mathematica.

Those approaches even with combination with element-wise scaling of function z does not provide successful practical result.

8.2. **Iterative approach.** The second approach does not provide directly equivalent symmetric problem. Let us suppose that the bilinear form a and linear functional l from (1) satisfy conditions of Lax-Milgram theorem, i.e. a and l are bounded and moreover bilinear form a is V-elliptical. So, there are two positive constants $M > 0$ and $\alpha > 0$ such that:

$$\begin{aligned} a(u,v) &\le M\|u\|_V\|v\|_V, \qquad &\forall u,v \in V, \\ a(u,u) &\ge \alpha\|u\|_V^2, &\forall u \in V. \end{aligned} \tag{33}$$

By the way, where the conditions from (1) guarantees existence of such constants M and α.

Let us construct sequence $\{u^k\}_{k=0}^{\infty} \in V$. We select arbitrary $u^0 \in V$, u^k, $k > 0$ we find from the following *symmetric* variational problem:

$$\begin{cases} \textit{find function } u^k \in V, \textit{ such that} \\ a(u^k, v) + a(v, u^k) = \langle l, v\rangle + a(v, u^{k-1}), \quad \forall v \in V. \end{cases} \tag{34}$$

Under previous conditions for a and l it is not hard to conclude that the sequence is well-defined, i.e. the solution of (34) exists on each step.

Theorem 1. *If $M < 2\alpha$, than $u^k \xrightarrow[k\to\infty]{} u$ in V, where u is the solution of* (3), *moreover*

$$\|u - u^k\|_V \le \left(\frac{M}{2\alpha}\right)^k \|u - u^0\|_V. \tag{35}$$

Proof. Let us define $e^k = u^k - u$. Then substitute $u^k = u + e^k$ into equation from (34). We get:

$$a(u + e^k, v) + a(v, u + e^k) = \langle l, v\rangle + a(v, u + e^{k-1}), \tag{36}$$

or after simplification:

$$a(e^k, v) + a(v, e^k) = a(v, e^{k-1}). \tag{37}$$

Taking $v = e^k$ and using (33) we obtain the following inequality chain:

$$2\alpha\|e^k\|_V^2 \le 2a(e^k, e^k) = a(e^k, e^{k-1}) \le M\|e^k\|_V\|e^{k-1}\|_V. \tag{38}$$

If there exist $k_0 : e^{k_0} = 0_V$ than it is obvious that $u^k = u, \forall k \ge k_0$, i.e. we have convergent sequence and the inequality from theorem statement holds. In other case $\forall k \in \mathbb{N}$ we can divide (38) by $\|e^k\|_V \ne 0$ and we obtain:

$$\|e^k\|_V \le \frac{M}{2\alpha}\|e^{k-1}\|_V. \tag{39}$$

By combining the last recurrent formula we simply get the final estimate (35):

$$\|e^k\|_V \le \left(\frac{M}{2\alpha}\right)^k \|e^0\|_V, \tag{40}$$

and convergence if $M < 2\alpha$.

9. Conclusion

In this paper we studied application of certain *hp*-adaptive algorithm to nonsymmetric problems. We combined this algorithm with different a posteriori error estimators and adaptation criteria to show by numerical experiment that algorithm can be directly applied to nonsymmetric problems. Also we construct several methods of symmetrization of given variational problem and provide corresponding theoretical analysis of those procedures. Two approaches are described. First can be used to build equivalent symmetric problem. In the second approach we built iterative procedure, where by solving symmetric variational problem on each step we can obtain sequence of elements that is convergent in the space of test functions to the solution of the original nonsymmetric problem. We still are working on the problem of theorem applicability to singular perturbed problems and schemes of combining this theorem with adaptive finite element algorithms. Also we are working on practical implementation of both symmetrization schemes which in practice involve building some ad hoc numerical quadratures.

Bibliography

1. Brenner S. The Mathematical Theory of Finite Element Methods / S. Brenner, L. Scott // Springer. – 2008. – 3 ed.
2. Demkowicz L. Computing with *hp*-adaptive finite elements. I. One– and Two Dimensional Elliptic and Maxwell Problems / L. Demkowicz. – Austin: Chapman & Hall. – 2005.
3. Doolan E. P. Uniform numerical methods for problems with initial and boundary layers / E. P. Doolan, J. J. H. Miller, W. H. A. Schilders. – Dublin: BOOLE PRESS. – 1980.
4. Dorfler W. Convergence of an adaptive hp finite element strategy in one space dimension / W. Dorfler, V. Heuveline // Appl. Num. Math. – 2007. – № 57. – P. 1108-1124.
5. Drebotiy R. *hp*-adaptive finite element method for 1d diffusion-convection-reaction boundary value problems / R. Drebotiy, H. Shynkarenko // Politechnika Opolska, Opole, Manufacturing Processes. Actual Problems-2014. – Volume 1: Basic science applications in manufacturing processes. – P. 11-26.
6. Drebotiy R. Symmetrization of diffusion-advection-reaction problem and *hp*-adaptive finite element approximations / R. Drebotiy, H. Shynkarenko // Visnyk of Lviv National University. – Series Applied Mathematics and Informatics. – 2015. – Vol. 23. – P. 55-72.
7. Drebotiy R. Comparison of error indicators and refinement criteria for *hp*-adaptation algorithm for finite element method / R. Drebotiy, H. Shynkarenko // Visnyk Odeskoho Natsionalnoho Universytetu Matematyka i Mekhanika. – 2014. – Vol. 19, Is. 4 (24). – P. 45–57.
8. Schwab Ch. *p* and *hp*-Finite Element Methods / Ch. Schwab. – Oxford: Clarendon Press. – 1998.
9. Solin P. Higher-Order Finite Element Methods / P. Solin, K. Segeth, I. Dolezel. – London: Chapman and Hall. – 2003.
10. Solin P. Static Condensation, Partial Orthogonalization of Basis Functions, and ILU Preconditioning in the hp-FEM / P. Solin, T. Vejchodsky // The University of Texas at El Paso, Department of Mathematical Sciences Research Reports Series, Research Report. – 2007. – № 4.

Roman Drebotiy, Heorgiy Shynkarenko,
Faculty of Applied Mathematics and Informatics,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine;
Heorgiy Shynkarenko,
Polytechnica Opolska,
Prószkowska Str., 76, Opole, 45758, Poland.

Received 01.07.2017; revised 13.09.2017

Журнал обчислювальної та прикладної математики | **2017** **№ 3 (126)** | Journal of Numerical & Applied Mathematics

UDC 519.6

CONVERGENCE ANALYSIS OF A TWO-STEP MODIFICATION OF THE GAUSS-NEWTON METHOD AND ITS APPLICATIONS

R. P. Iakymchuk, S. M. Shakhno, H. P. Yarmola

Резюме. У роботі досліджено збіжність двокрокової модифікації методу Гаусса-Ньютона за узагальнених умов Ліпшиця для похідних першого і другого порядків. Встановлено порядок і радіус збіжності методу, а також область єдиності розв'язку нелінійної задачі про найменші квадрати. Проведено чисельні експерименти на відомих тестових задачах.

Abstract. We investigate the convergence of a two-step modification of the Gauss-Newton method applying the generalized Lipschitz condition for the first- and second-order derivatives. The convergence order as well as the convergence radius of the method are studied and the uniqueness ball of the solution of the nonlinear least squares problem is examined. Finally, we carry out numerical experiments on a set of well-known test problems.

1. Introduction

Let us consider the nonlinear least squares problem [6]:

$$\min f(x) := \frac{1}{2}F(x)^T F(x), \tag{1}$$

where F is a Fréchet differentiable operator defined on $\mathbb{R}^n$ with its values on $\mathbb{R}^m$, $m \geq n$. The best known method for finding an approximate solution of the problem (1) is the Gauss-Newton method, which is defined as

$$x_{k+1} = x_k - [F^{'}(x_k)^T F^{'}(x_k)]^{-1} F^{'}(x_k)^T F(x_k), \ k = 0, 1, 2, ... \tag{2}$$

The convergence analysis of the method (2) under various conditions was conducted in [4, 5]. In paper [11], three free-derivative iterative methods were investigated under the classical Lipschitz conditions. The radius of the convergence ball and the convergence order of these methods were determined. The study of these methods was conducted in the case of both zero and nonzero residuals.

For solving the problem (1), we consider a two-step modification of the Gauss-Newton method [1, 3]

$$\left\{ \begin{array}{l} x_{k+1} = x_k - [F^{'}(z_k)^T F^{'}(z_k)]^{-1} F^{'}(z_k)^T F(x_k), \\ y_{k+1} = x_{k+1} - [F^{'}(z_k)^T F^{'}(z_k)]^{-1} F^{'}(z_k)^T F(x_{k+1}), \ k = 0, 1, 2, ..., \end{array} \right. \tag{3}$$

Key words. Least squares problem, Gauss-Newton method, Lipschitz conditions with L average, radius of convergence, uniqueness ball.

where $z_k = (x_k+y_k)/2$; x_0 and y_0 are given. In case when $m = n$, this method is equivalent to the methods proposed by Bartish [2] and Werner [17]. On each iteration, the method (3) computes the inversion of the matrix $[F^{'}(z_k)^T F^{'}(z_k)]^{-1}$ only once. Because of that, the computation cost of each iteration of the method (3) is roughly the same as of the Gauss-Newton method (2): for calculating y_{k+1}, it is only necessary to perform one backward substitution, which requires $O(n^2)$ floating-point operations (Flops), since the LL^T decomposition of the matrix $F^{'}(z_k)^T F^{'}(z_k)$, which costs $O(n^3)$ ($O(n^3/3)$ to be precise) Flops [6], is computed for x_{k+1}.

The main goal of this paper is to analyze the local convergence of the method (3). Bartish et al. [1] examined the local convergence of this method using the classical Lipschitz condition for derivatives of the second-order, but only for the problem (1) with zero residuals. Instead, we study the convergence of the above-mentioned method using the generalized Lipschitz conditions [15] for derivatives of the first- and second-orders; such conditions employ an integrable function $L(u)$ instead of the Lipschitz constant L. The Lipschitz condition with L average in the inscribe sphere makes us unify the convergence criteria containing the Kantorovich theorem and the Smale α-theory [5, 8, 12, 14, 15]. We prove the convergence of the method (3) for the problem (1) with zero as well as non-zero residuals. Furthermore, we find both the order and the radius of the convergence of the method (3) as well as the uniqueness ball of the solution of the problem (1). We have published some of the results without proofs as an extended abstract [7].

2. Preliminaries

For our study, we present different definitions of the Lipschitz conditions. Let us denote $B(x_*, r) = \{x \in D \subseteq \mathbb{R}^n : \|\mathrm{x} - \mathrm{x}_*\| \leq \mathrm{r}\}$ as an closed ball with the radius r ($r > 0$) at x_*.

Definition 1. The function $F : \mathbb{R}^n \to \mathbb{R}^m$ satisfies the classical Lipschitz condition on $B(x_*, r)$ if

$$\|F(x) - F(y)\| \leq L\|x - y\|,$$

where $x, y \in B(x_*, r)$ and L is the Lipschitz constant.

In Definition 1 L may not necessary be a constant, but it also can be an integrable function $L(u)$.

Definition 2 ([15]). The function $F : \mathbb{R}^n \to \mathbb{R}^m$ satisfies the Lipschitz condition with L average on $B(x_*, r)$ if

$$\|F(x) - F(y)\| \leq \int_0^{\|x-y\|} L(u)du, \quad \forall x \in B(x_*, r),$$

where $L(u)$ is a positive non-decreasing function.

Let $\mathbb{R}^{m\times n}$, $m \geq n$, denote a set of all $m \times n$ matrices. Then, for a full rank matrix $A \in \mathbb{R}^{m\times n}$, its Moore-Penrose pseudo-inverse [6] is defined as $A^{\dagger} = (A^T A)^{-1} A^T$.

Lemma 1 ([13,16]). *Let $A, E \in \mathbb{R}^{m\times n}$. Assume that $C = A+E$, $\|A^\dagger\|\|E\| < 1$, and $rank(A) = rank(C)$. Then,*

$$\|C^\dagger\| \leq \frac{\|A^\dagger\|}{1-\|A^\dagger\|\|E\|}.$$

If $rank(A) = rank(C) = \min(m,n)$, we can obtain

$$\|C^\dagger - A^\dagger\| \leq \frac{\sqrt{2}\|A^\dagger\|^2\|E\|}{1-\|A^\dagger\|\|E\|}.$$

Lemma 2 ([4]). *Let $A, E \in \mathbb{R}^{m\times n}$. Assume that $C = A+E$, $\|EA^\dagger\| < 1$, and $rank(A) = n$, then $rank(C) = n$.*

Lemma 3 ([15]). *Let $h(t) = \dfrac{1}{t}\displaystyle\int_0^t L(u)du$, $0 \leq t \leq r$, where $L(u)$ is a positive integrable function and monotonically non-decreasing on $[0, r]$. Then, $h(t)$ is monotonically non-decreasing with respect to t.*

Lemma 4 ([10]). *Let $g(t) = \dfrac{1}{t^3}\displaystyle\int_0^t N(u)(t-u)^2du$, $0 \leq t \leq r$, where $N(u)$ is a positive integrable function and monotonically non-decreasing on $[0, r]$. Then, $g(t)$ is monotonically non-decreasing with respect to t.*

3. Local Convergence Analysis of Method (3)

In this section, we investigate the convergence and the radius of the convergence ball of the method (3).

Theorem 1. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$, $m \geq n$, be a twice Fréchet differentiable operator on a subset $D \subseteq \mathbb{R}^n$. Assume that the problem (1) has a solution $x_* \in D$ and a Fréchet derivative $F'(x_*)$ has full rank. Suppose that Fréchet derivatives $F'(x)$ and $F''(x)$ on $B(x_*, R) = \{x \in D : \|x - x_*\| \leq R\}$ satisfy the Lipschitz conditions with L and N average:*

$$\|F'(x) - F'(y)\| \leq \int_0^{\|x-y\|} L(u)du, \tag{4}$$

$$\|F''(x) - F''(y)\| \leq \int_0^{\|x-y\|} N(u)du, \tag{5}$$

where L and N are positive non-decreasing functions on $[0, 3R/2]$.

Furthermore, assume function

$$\begin{aligned} h_0(p) \;=\; & (\beta/8)\int_0^p N(u)(p-u)^2du + \beta p\Big(\int_0^{(3/2)p} L(u)du + \int_0^p L(u)du\Big) + \\ & + \sqrt{2}\alpha\beta^2\int_0^p L(u)du - p \end{aligned} \tag{6}$$

has a minimal zero r on $[0, R]$, which also satisfies

$$\beta\int_0^r L(u)du < 1. \tag{7}$$

Then, for all $x_0, y_0 \in B(x_*, r)$ the sequences $\{x_k\}$ and $\{y_k\}$, which are generated by the method (3), are well defined, remain in $B(x_*, r)$ for all $k \geq 0$, and converge to x_* such that

$$\begin{aligned} \rho(x_{k+1}) &\leq \gamma\rho(x_k)^3 + \eta\rho(x_k)\rho(y_k) + \theta\rho(z_k), && (8)\\ \rho(y_{k+1}) &\leq \gamma\rho(x_{k+1})^3 + (\eta/3)(\rho(x_k) + \rho(y_k) + \rho(x_{k+1}))\rho(x_{k+1}) + \\ &\quad +\theta\rho(z_k), && (9)\\ r_{k+1} &= \max\{\rho(x_{k+1}), \rho(y_{k+1})\} \leq qr_k \leq \cdots \leq q^{k+1}r_0, && (10) \end{aligned}$$

where $\rho(x) = \|x - x_*\|$, $r_0 = \max\{\rho(x_0), \rho(y_0)\}$,

$$\begin{aligned} q &= \gamma\rho(x_0)^2 + \theta + \eta, && (11)\\ \gamma &= \frac{\beta\int_0^{\rho(x_0)} N(u)(\rho(x_0) - u)^2 du}{8\rho(x_0)^3\left(1 - \beta\int_0^{\rho(z_0)} L(u)du\right)},\ \theta = \frac{\sqrt{2}\alpha\beta^2\int_0^{\rho(z_0)} L(u)du}{\rho(z_0)\left(1 - \beta\int_0^{\rho(z_0)} L(u)du\right)}, && (12)\\ \eta &= \frac{\beta\int_0^{\rho(x_0)+\rho(y_0)/2} L(u)du}{(2\rho(x_0) + \rho(y_0))/3\left(1 - \beta\int_0^{\rho(z_0)} L(u)du\right)}, && (13)\\ \alpha &= \|F(x_*)\|, \quad \beta = \|(F'(x_*)^T F'(x_*))^{-1}F'(x_*)^T\|. && (14) \end{aligned}$$

Proof. Let choose arbitrary $x_0, y_0 \in B(x_*, r)$. For x_1, y_1 that are generated by (3), we have

$$\begin{gathered} x_1 - x_* = x_0 - x_* - \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T F(x_0) = \\ = \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T \left[F'(z_0)(x_0 - x_*) - F(x_0) + F(x_*)\right] + \\ + \left[F'(x_*)^T F'(x_*)\right]^{-1} F'(x_*)^T F(x_*) - \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T F(x_*) = \\ = \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T \times \\ \times \left[\left(F'\left(\frac{x_0 + x_*}{2}\right)(x_0 - x_*) - F(x_0) + F(x_*)\right) + \right. \\ \left. + \left(F'(z_0) - F'\left(\frac{x_0 + x_*}{2}\right)\right)(x_0 - x_*)\right] + \\ + \left[F'(x_*)^T F'(x_*)\right]^{-1} F'(x_*)^T F(x_*) - \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T F(x_*); \end{gathered}$$

$$\begin{gathered} y_1 - x_* = x_1 - x_* - \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T F(x_1) = \\ = \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T \left[F'(z_0)(x_1 - x_*) - F(x_1) + F(x_*)\right] + \\ + \left[F'(x_*)^T F'(x_*)\right]^{-1} F'(x_*)^T F(x_*) - \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T F(x_*) = \\ = \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T \times \\ \times \left[\left(F'\left(\frac{x_1 + x_*}{2}\right)(x_1 - x_*) - F(x_1) + F(x_*)\right) + \right. \\ \left. + \left(F'(z_0) - F'\left(\frac{x_1 + x_*}{2}\right)\right)(x_1 - x_*)\right] + \end{gathered}$$

$$+\left[F^{'}(x_*)^TF^{'}(x_*)\right]^{-1}F^{'}(x_*)^TF(x_*)-\left[F^{'}(z_0)^TF^{'}(z_0)\right]^{-1}F^{'}(z_0)^TF(x_*).$$

According to Lemma 1 from [17] with the value $\omega = 1/2$ we can write

$$\begin{gathered}F(x)-F(y)-F'\left(\frac{x+y}{2}\right)(x-y)=\\ =\frac{1}{4}\int_0^1(1-t)\left[F''\left(\frac{x+y}{2}+\frac{t}{2}(x-y)\right)-\right.\\ \left.-F''\left(\frac{x+y}{2}+\frac{t}{2}(y-x)\right)\right](x-y)^2dt.\end{gathered}\tag{15}$$

By setting $x = x_*$ and $y = x_0$ in the equation above, we receive

$$\begin{gathered}\left\|F(x_*)-F(x_0)-F'\left(\frac{x_0+x_*}{2}\right)(x_*-x_0)\right\|=\\ =\frac{1}{4}\left\|\int_0^1(1-t)\left[F''\left(\frac{x_0+x_*}{2}+\frac{t}{2}(x_*-x_0)\right)-\right.\right.\\ \left.\left.-\;F''\left(\frac{x_0+x_*}{2}+\frac{t}{2}(x_0-x_*)\right)\right](x_*-x_0)^2dt\right\|\le\\ \le\frac{1}{4}\int_0^1(1-t)\int_0^{t\|x_0-x_*\|}N(u)du\|x_0-x_*\|^2dt=\\ =\frac{1}{8}\int_0^{\rho(x_0)}N(u)\left(1-\frac{u}{\rho(x_0)}\right)^2du\rho(x_0)^2=\frac{1}{8}\int_0^{\rho(x_0)}N(u)(\rho(x_0)-u)^2du,\end{gathered}$$

and also

$$\left\|F^{'}\left(\frac{x_0+y_0}{2}\right)-F^{'}\left(\frac{x_0+x_*}{2}\right)\right\|\le\int_0^{\rho(y_0)/2}L(u)du.$$

Using (4) and (14), we obtain that

$$\|(F^{'}(x_*)^TF^{'}(x_*))^{-1}F^{'}(x_*)^T\|\|F^{'}(x)-F^{'}(x_*)\|\le\beta\int_0^{\rho(x)}L(u)du.$$

According to Lemmas 1 and 2 and that $F^{'}(x)$ has full rank, for all $x\in B(x_*,r)$, the following inequalities hold

$$\|(F^{'}(x)^TF^{'}(x))^{-1}F^{'}(x)^T\|\le\frac{\beta}{1-\beta\int_0^{\rho(x)}L(u)du};\tag{16}$$

$$\begin{gathered}\|(F^{'}(x)^TF^{'}(x))^{-1}F^{'}(x)^T-(F^{'}(x_*)^TF^{'}(x_*))^{-1}F^{'}(x_*)^T\|\le\\ \le\frac{\sqrt{2}\beta^2\int_0^{\rho(x)}L(u)du}{1-\beta\int_0^{\rho(x)}L(u)du}.\end{gathered}\tag{17}$$

By the monotonicity of $L(u)$ and $N(u)$ with Lemmas 3 and 4, functions $\frac{1}{t}\int_0^t L(u)du$ and $\frac{1}{t^3}\int_0^t N(u)(t-u)^2du$ are non-decreasing by t. Hence, from

(6) and (7) it follows that

$$q \le \frac{1}{r_0}\left[\frac{\beta\int\limits_0^{r_0} N(u)(r_0-u)^2du}{8\left(1-\beta\int\limits_0^{r_0} L(u)du\right)} + \frac{\beta r_0 \int\limits_0^{(3/2)r_0} L(u)du + \sqrt{2}\alpha\beta^2\int\limits_0^{r_0} L(u)du}{1-\beta\int\limits_0^{r_0} L(u)du}\right] <$$

$$< \frac{1}{r}\left[\frac{\beta\int\limits_0^{r} N(u)(r-u)^2du}{8\left(1-\beta\int\limits_0^{r} L(u)du\right)} + \frac{\beta r \int\limits_0^{(3/2)r} L(u)du}{1-\beta\int\limits_0^{r} L(u)du} + \frac{\sqrt{2}\alpha\beta^2\int\limits_0^{r} L(u)du}{1-\beta\int\limits_0^{r} L(u)du}\right] \le 1.$$

Thus, by Lemmas 1-4, conditions (4) and (5), and the afore-derived estimates, we obtain

$$\begin{aligned}\|x_1-x_*\| \le \left\|\left[F'(z_0)^TF'(z_0)\right]^{-1}F'(z_0)^T\right\| \times \\ \times\left\|\left(F'\left(\frac{x_0+x_*}{2}\right)(x_0-x_*)-F(x_0)+F(x_*)\right)+\right. \\ \left.+\left(F'(z_0)-F'\left(\frac{x_0+x_*}{2}\right)\right)(x_0-x_*)\right\|+ \\ +\left\|\left[F'(x_*)^TF'(x_*)\right]^{-1}F'(x_*)^TF(x_*)-\left[F'(z_0)^TF'(z_0)\right]^{-1}F'(z_0)^TF(x_*)\right\| \le \\ \le \frac{\beta\rho(x_0)^3\int_0^{\rho(x_0)}N(u)(\rho(x_0)-u)^2du}{8\rho(x_0)^3\left(1-\beta\int_0^{\rho(z_0)}L(u)du\right)}+ \\ +\frac{\beta\rho(x_0)\rho(y_0)\int_0^{\rho(y_0)/2}L(u)du}{\rho(y_0)\left(1-\beta\int_0^{\rho(z_0)}L(u)du\right)}+\frac{\sqrt{2}\alpha\beta^2\rho(z_0)\int_0^{\rho(z_0)}L(u)du}{\rho(z_0)\left(1-\beta\int_0^{\rho(z_0)}L(u)du\right)} < \\ < \gamma\rho(x_0)^3+\eta\rho(x_0)\rho(y_0)+\theta\rho(z_0) < qr_0 < r.\end{aligned}$$

Similarly,

$$\begin{aligned}\|y_1-x_*\| = \left\|\left[F'(z_0)^TF'(z_0)\right]^{-1}F'(z_0)^T\right\| \times \\ \times\left\|\left(F'\left(\frac{x_1+x_*}{2}\right)(x_1-x_*)-F(x_1)+F(x_*)\right)+\right. \\ \left.+\left(F'(z_0)-F'\left(\frac{x_1+x_*}{2}\right)\right)(x_1-x_*)\right\|+ \\ +\left\|\left[F'(x_*)^TF'(x_*)\right]^{-1}F'(x_*)^TF(x_*)-\left[F'(z_0)^TF'(z_0)\right]^{-1}F'(z_0)^TF(x_*)\right\| \le \\ \le \frac{\beta\rho(x_1)^3\int_0^{\rho(x_1)}N(u)(\rho(x_1)-u)^2du}{8\rho(x_1)^3\left(1-\beta\int_0^{\rho(z_0)}L(u)du\right)}+ \\ +\frac{\beta\rho(x_1)\rho(z_0')\int_0^{\rho(z_0')}L(u)du}{\rho(z_0')\left(1-\beta\int_0^{\rho(z_0)}L(u)du\right)}+\frac{\sqrt{2}\alpha\beta^2\rho(z_0)\int_0^{\rho(z_0)}L(u)du}{\rho(z_0)\left(1-\beta\int_0^{\rho(z_0)}L(u)du\right)} \le\end{aligned}$$

$$\leq \gamma\rho(x_1)^3 + (\eta/3)\rho(x_1)(\rho(x_0) + \rho(y_0) + \rho(x_1)) + \theta\rho(z_0) <$$
$$< \gamma\rho(x_0)^3 + (\eta/3)\rho(x_0)(2\rho(x_0) + \rho(y_0)) + \theta\rho(z_0) < qr_0 < r,$$

where $\rho(z_0^{'}) = (\rho(x_0) + \rho(y_0) + \rho(x_1))/2$. Therefore, $x_1, y_1 \in B(x_*, r)$ and both (8) and (9) follow for $k = 0$. Also, (10) is satisfied

$$r_1 = \max\{\|x_1 - x_*\|, \|y_1 - x_*\|\} \leq qr_0.$$

Using mathematical induction, assume that $x_k, y_k \in B(x_*, r)$ and (8)–(10) hold for $k > 0$. Then, from (3) for $k+1$ we obtain that

$$\begin{gathered}
\|x_{k+1} - x_*\| \leq \frac{\beta\rho(x_k)^3 \int_0^{\rho(x_k)} N(u)(\rho(x_k) - u)^2 du}{8\rho(x_k)^3 \left(1 - \beta\int_0^{\rho(z_k)} L(u)du\right)} + \\
+ \frac{\beta\rho(x_k)\rho(y_k) \int_0^{\rho(y_k)/2} L(u)du}{\rho(y_k)\left(1 - \beta\int_0^{\rho(z_k)} L(u)du\right)} + \frac{\sqrt{2}\alpha\beta^2\rho(z_k)\int_0^{\rho(z_k)} L(u)du}{\rho(z_k)\left(1 - \beta\int_0^{\rho(z_k)} L(u)du\right)} \leq \\
\leq \frac{\beta\rho(x_k)^3 \int_0^{\rho(x_0)} N(u)(\rho(x_0) - u)^2 du}{8\rho(x_0)^3 \left(1 - \beta\int_0^{\rho(z_0)} L(u)du\right)} + \\
+ \frac{\beta\rho(x_k)\rho(y_k) \int_0^{\rho(y_0)/2} L(u)du}{\rho(y_0)\left(1 - \beta\int_0^{\rho(z_0)} L(u)du\right)} + \frac{\sqrt{2}\alpha\beta^2\rho(z_k)\int_0^{\rho(z_0)} L(u)du}{\rho(z_0)\left(1 - \beta\int_0^{\rho(z_0)} L(u)du\right)} \leq \\
\leq \gamma\rho(x_k)^3 + \eta\rho(x_k)\rho(y_k) + \theta\rho(z_k) \leq qr_k < r.
\end{gathered} \tag{18}$$

and

$$\begin{gathered}
\|y_{k+1} - x_*\| \leq \frac{\beta\rho(x_{k+1})^3 \int_0^{\rho(x_{k+1})} N(u)(\rho(x_{k+1}) - u)^2 du}{8\rho(x_{k+1})^3 \left(1 - \beta\int_0^{\rho(z_k)} L(u)du\right)} + \\
+ \frac{\beta\rho(x_{k+1})\rho(z_k^{'}) \int_0^{\rho(z_k^{'})} L(u)du}{\rho(z_k^{'})\left(1 - \beta\int_0^{\rho(z_k)} L(u)du\right)} + \frac{\sqrt{2}\alpha\beta^2\rho(z_k)\int_0^{\rho(z_k)} L(u)du}{\rho(z_k)\left(1 - \beta\int_0^{\rho(z_k)} L(u)du\right)} \leq \\
\leq \frac{\beta\rho(x_{k+1})^3 \int_0^{\rho(x_0)} N(u)(\rho(x_0) - u)^2 du}{8\rho(x_0)^3 \left(1 - \beta\int_0^{\rho(z_0)} L(u)du\right)} + \\
+ \frac{\beta\rho(x_{k+1})\rho(z_k^{'}) \int_0^{\rho(z_0^{'})} L(u)du}{\rho(z_0^{'})\left(1 - \beta\int_0^{\rho(z_0)} L(u)du\right)} + \frac{\sqrt{2}\alpha\beta^2\rho(z_k)\int_0^{\rho(z_0)} L(u)du}{\rho(z_0)\left(1 - \beta\int_0^{\rho(z_0)} L(u)du\right)} \leq \\
\leq \gamma\rho(x_{k+1})^3 + (\eta/3)(\rho(x_k) + \rho(y_k) + \rho(x_{k+1}))\rho(x_{k+1}) + \theta\rho(z_k) \leq \\
\leq qr_k < r.
\end{gathered} \tag{19}$$

where $\rho(z_k^{'}) = (\rho(x_k) + \rho(y_k) + \rho(x_{k+1}))/2$. According to (11) and both inequalities (8) and (9), we receive

$$r_{k+1} \quad = \quad \max\{\|x_{k+1} - x_*\|, \|y_{k+1} - x_*\|\} \leq qr_k \leq q^2 r_{k-1} \leq \cdots \leq q^{k+1} r_0.$$

Thus, $x_{k+1}, y_{k+1} \in B(x_*, r)$ and (8)–(10) hold; and also $\lim\limits_{k\to\infty} x_k = x_*$ and $\lim\limits_{k\to\infty} y_k = x_*$. This completes the induction and the proof of Theorem 1. $\square$

In case of zero residual ($\alpha = \|F(x_*)\| = 0$) the results of Theorem 1 are

Corollary 1. *Suppose that x_* satisfies (1), $F(x_*) = 0$, $F(x)$ is a twice Fréchet differentiable operator in $B(x_*, R)$, $F^{'}(x_*)$ has full rank, and both $F^{'}(x)$ and $F^{''}(x)$ satisfy the Lipschitz conditions with L and N average as in (4) and (5), respectively, where L and N are positive non-decreasing functions on $[0, 3R/2]$. Furthrmore, assume function H_0 has a minimal zero r on $[0, R]$, which also satisfies:*

$$\beta \int_0^r L(u)du < 1,$$

where

$$H_0(p) = (\beta/8)\int_0^p N(u)(p-u)^2 du + \beta p\Big(\int_0^{(3/2)p} L(u)du + \int_0^p L(u)du\Big) - p.$$

Then, the Gauss-Newton type method (3) is convergent for all x_0, $y_0 \in B(x_*, r)$ such that

$$\begin{aligned} \rho(x_{k+1}) &\leq \gamma\rho(x_k)^3 + \eta\rho(x_k)\rho(y_k), \\ \rho(y_{k+1}) &\leq \gamma\rho(x_{k+1})^3 + (\eta/3)(\rho(x_k) + \rho(y_k) + \rho(x_{k+1}))\rho(x_{k+1}), \\ r_{k+1} &= \max\{\rho(x_{k+1}), \rho(y_{k+1})\} \leq qr_k \leq \cdots \leq q^{k+1}r_0, \end{aligned}$$

where $\rho(x) = \|x - x_*\|$, $r_0 = \max\{\rho(x_0), \rho(y_0)\}$,

$$q = \frac{\beta\int_0^{\rho(x_0)} N(u)(\rho(x_0)-u)^2du}{8\rho(x_0)\left(1-\beta\int_0^{\rho(z_0)} L(u)du\right)} +$$

$$+ \frac{\beta\rho(x_0)\int_0^{\rho(x_0)+\rho(y_0)/2} L(u)du}{(2\rho(x_0)+\rho(y_0))/3\left(1-\beta\int_0^{\rho(z_0)} L(u)du\right)} < 1,$$

γ, η, β hold in (12)–(14).

Corollary 2. *Convergence order of the iterative method (3) in case of zero residual is equal to $1+\sqrt{2}$.*

Proof. Assume that $a_k = \rho(x_k), b_k = \rho(y_k), k = 0, 1, 2, ...$ Since the residual is equal to zero, i.e. $\alpha = \|F(x_*)\| = 0$, so $\theta = 0$. From the inequalities (18) and (19), we have

$$a_{k+1} \leq a_k(\gamma a_k^2 + \eta b_k), \tag{20}$$

$$\begin{aligned} b_{k+1} &\leq a_{k+1}\left[\gamma a_{k+1}^2 + \eta/3(a_k + a_{k+1} + b_k)\right] \leq \\ &\leq a_{k+1}\left[(\gamma a_k + 2\eta/3)a_k + \eta b_k/3\right] \leq \\ &\leq a_{k+1}a_k\left[\gamma r + \eta\right] = a_{k+1}a_k\phi_1. \end{aligned} \tag{21}$$

From (20) and (21) for large enough k, it follows

$$a_{k+1} \leq a_k(\gamma a_k^2 + \eta b_k) \leq a_k(\gamma a_k^2 + \eta\phi_1 a_k a_{k-1}) \leq a_k^2 a_{k-1}(\gamma + \eta\phi_1) = a_k^2 a_{k-1}\phi_2.$$

From this inequality, we obtain an equation [17]

$$\rho^2 - 2\rho - 1 = 0.$$

The positive root of the latter, which is $\rho_* = 1+\sqrt{2}$, is the order of convergence of the iterative method (3). □

Theorem 2. ***(The uniqueness of solution)*** *Suppose x_* satisfies (1) and $F(x)$ has a continuous derivative $F^{'}(x)$ in the ball $B(x_*, r)$. Moreover, $F^{'}(x_*)$ has full rank and $F^{'}(x)$ satisfies the Lipschitz condition with L average (4). Let $r > 0$ satisfy*

$$\frac{\beta}{r}\int_0^r L(u)(r-u)du + \frac{\alpha\beta_0}{r}\int_0^r L(u)du \le 1, \tag{22}$$

where α and β are defined in (14) and $\beta_0 = \|[F^{'}(x_*)^T F^{'}(x_*)]^{-1}\|$. Then, x_* is a unique solution of the problem (1) in $B(x_*, r)$.

The proof of this theorem is analogous to the one in [4].

4. Applications

In this section, we apply the obtained results to special cases, when, for instance, L is a Lipschitz constant. Then, we immediately receive results of the convergence analysis of the method (3).

Theorem 3. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$, $m \ge n$, be a twice Fréchet differentiable operator in $D \subseteq \mathbb{R}^n$. Assume that (1) has a solution $x_* \in D$ and a Fréchet derivative $F^{'}(x_*)$ has full rank. Suppose that Fréchet derivatives $F^{'}(x)$ and $F^{''}(x)$ on $B(x_*, r) = \{x \in D : \|x - x_*\| \le r\}$ satisfy the Lipschitz conditions:*

$$\begin{aligned} \|F^{'}(x) - F^{'}(y)\| &\le L\|x-y\|, && (23)\\ \|F^{''}(x) - F^{''}(y)\| &\le N\|x-y\|, && (24) \end{aligned}$$

where $x, y \in B(x_*, r)$ and both L and N are positive numbers. Also, the radius $r > 0$ is a root of the equation

$$\beta N r^2 + 60\beta L r + 24\sqrt{2}\alpha\beta^2 L - 24 = 0. \tag{25}$$

Then, for all $x_0, y_0 \in B(x_*, r)$ the sequences $\{x_k\}$ and $\{y_k\}$, which are generated by the method (3), are well defined, remain in $B(x_*, r)$ for all $k \ge 0$, and converge to x_* such that

$$\begin{aligned} \rho(x_{k+1}) &\le \frac{(\beta/24)N\rho(x_k)^3 + \beta L\rho(x_k)\rho(y_k)/2 + \sqrt{2}\alpha\beta^2 L\rho(z_k)}{1 - \beta L\rho(z_k)}, && (26)\\ \rho(y_{k+1}) &\le \frac{(\beta/24)N\rho(x_{k+1})^3 + \beta L\rho(x_{k+1})(\rho(x_{k+1}) + \rho(x_k) + \rho(y_k))/2}{1 - \beta L\rho(z_k)} + \\ &+ \frac{\sqrt{2}\alpha\beta^2 L\rho(z_k)}{1 - \beta L\rho(z_k)}, && (27)\\ r_{k+1} &= \max\{\rho(x_{k+1}), \rho(y_{k+1})\} \le q r_k \le \cdots \le q^{k+1} r_0, && (28) \end{aligned}$$

where $\rho(x) = \|x - x_*\|$, $r_0 = \max\{\rho(x_0), \rho(y_0)\}$,

$$0 < q = \frac{(\beta/24)N\rho(x_0)^2 + \beta L(\rho(x_0) + \rho(y_0)/2) + \sqrt{2}\alpha\beta^2 L}{1 - \beta L\rho(z_0)} < 1, \tag{29}$$

$z_k = (x_k + y_k)/2$ and both α and β are defined in (14).

Proof. Let choose arbitrary $x_0, y_0 \in B(x_*, r)$. According to Lemma 1 from [17] and the proof of Theorem 1, by setting $x = x_*$ and $y = x_0$ in (15), we receive

$$\left\| F(x_*) - F(x_0) - F'\left(\frac{x_0 + x_*}{2}\right)(x_* - x_0)\right\| =$$
$$= \frac{1}{4}\left\| \int_0^1 (1-t)\left[F''\left(\frac{x_0+x_*}{2} + \frac{t}{2}(x_* - x_0)\right) - \right.\right.$$
$$\left.\left. - F''\left(\frac{x_0+x_*}{2} + \frac{t}{2}(x_0 - x_*)\right)\right](x_* - x_0)^2 dt\right\| \leq$$
$$\leq \frac{1}{4}\int_0^1 t(1-t)N\|x_0 - x_*\|^3 dt = \frac{1}{24}N\rho(x_0)^3,$$

and also

$$\left\| F'\left(\frac{x_0+y_0}{2}\right) - F'\left(\frac{x_0+x_*}{2}\right)\right\| \leq L\rho(y_0)/2.$$

Using (23) and (14), we obtain that

$$\|(F'(x_*)^T F'(x_*))^{-1} F'(x_*)^T\| \|F'(x) - F'(x_*)\| \leq \beta L \rho(x).$$

According to that $F'(x)$ has full rank, for all $x \in B(x_*, r)$, the following inequalities hold

$$\|(F'(x)^T F'(x))^{-1} F'(x)^T\| \leq \frac{\beta}{1 - \beta L\rho(x)},$$

$$\|(F'(x)^T F'(x))^{-1} F'(x)^T - (F'(x_*)^T F'(x_*))^{-1} F'(x_*)^T\| \leq \frac{\sqrt{2}\beta^2 L\rho(x)}{1 - \beta L\rho(x)}.$$

Hence, from (25) it follows that

$$0 < q = \frac{(\beta/24)N\rho(x_0)^2 + 3\beta L(\rho(x_0) + \rho(y_0)/2) + \sqrt{2}\alpha\beta^2 L}{1 - \beta L\rho(z_0)} <$$
$$< \frac{(\beta/24)Nr^2 + 3\beta Lr/2 + \sqrt{2}\alpha\beta^2 L}{1 - \beta L r} \leq 1.$$

Thus, by Lemmas 1-4, conditions (23) and (24), and the derived estimates in the proof of Theorem 1, we obtain

$$\|x_1 - x_*\| \leq \frac{(\beta/24)N\rho(x_0)^3 + \beta L\rho(x_0)\rho(y_0)/2 + \sqrt{2}\alpha\beta^2 L\rho(z_0)}{1 - \beta L\rho(z_0)} < qr_0 < r.$$

Similarly,

$$\|y_1 - x_*\| \leq \frac{(\beta/24)N\rho(x_1)^3}{1 - \beta L\rho(z_0)} +$$
$$+ \frac{\beta L\rho(x_1)(\rho(x_1) + \rho(x_0) + \rho(y_0))/2 + \sqrt{2}\alpha\beta^2 L\rho(z_0)}{1 - \beta L\rho(z_0)} \leq$$
$$\leq \frac{(\beta/24)N\rho(x_0)^3 + \beta L\rho(x_0)(2\rho(x_0) + \rho(y_0))/2 + \sqrt{2}\alpha\beta^2 L\rho(z_0)}{1 - \beta L\rho(z_0)} < qr_0 < r.$$

Therefore, $x_1, y_1 \in B(x_*, r)$ and both (26) and (27) follow for $k = 0$. Also, (28) is satisfied

$$r_1 = \max\{\|x_1 - x_*\|, \|y_1 - x_*\|\} \leq qr_0.$$

Using mathematical induction, assume that $x_k, y_k \in B(x_*, r)$ and (28) holds for $k > 0$. Then, for $k+1$ from (3) we obtain that

$$\|x_{k+1} - x_*\| \le \frac{(\beta/24)N\rho(x_k)^3 + \beta L\rho(x_k)\rho(y_k)/2 + \sqrt{2}\alpha\beta^2 L\rho(z_k)}{1 - \beta L\rho(z_k)} \le$$
$$\le \frac{((\beta/24)N\rho(x_0)^2 + \beta L\rho(y_0)/2 + \sqrt{2}\alpha\beta^2 L)r_k}{1 - \beta L\rho(z_0)} = qr_k < r$$

and

$$\|y_{k+1} - x_*\| \le \frac{(\beta/24)N\rho(x_{k+1})^3 + \beta L\rho(x_{k+1})(\rho(x_{k+1}) + \rho(x_k) + \rho(y_k))/2}{1 - \beta L\rho(z_k)} +$$
$$+ \frac{\sqrt{2}\alpha\beta^2 L\rho(z_k)}{1 - \beta L\rho(z_k)} < qr_k < r.$$

According to (29) and both inequalities (26) and (27), we receive

$$r_{k+1} = \max\{\|x_{k+1} - x_*\|, \|y_{k+1} - x_*\|\} \le qr_k \le q^2 r_{k-1} \le \cdots \le q^{k+1} r_0.$$

Thus, $x_{k+1}, y_{k+1} \in B(x_*, r)$ as well as (26), (27), and (28) hold. □

From (25) it follows that the convergence radius of the method (3) is

$$r = \frac{4(1 - \sqrt{2}\alpha\beta^2 L)}{5\beta L + \frac{1}{12}\sqrt{(60\beta L)^2 + 96\beta N(1 - \sqrt{2}\alpha\beta^2 L)}}.$$

For zero residual, Theorem 3 can be formulated as

Corollary 3. *Suppose that x_* satisfies (1), $F(x_*) = 0$, $F(x)$ is a twice Fréchet differentiable operator in $B(x_*, r)$, $F'(x_*)$ has full rank, and both $F'(x)$ and $F''(x)$ satisfy the classic Lipschitz conditions as in (23) and (24), respectively. Moreover, the radius $r > 0$ is a unique positive root of the following equation*

$$\beta N r^2 + 60\beta L r - 24 = 0.$$

Then, the Gauss-Newton type method (3) is convergent for all x_0, $y_0 \in B(x_*, r)$ such that

$$\begin{aligned}
\rho(x_{k+1}) &\le \frac{(\beta/24)N\rho(x_k)^3 + \beta L\rho(x_k)\rho(y_k)/2}{1 - \beta L\rho(z_k)}, \\
\rho(y_{k+1}) &\le \frac{(\beta/24)N\rho(x_{k+1})^3 + \beta L\rho(x_{k+1})(\rho(x_{k+1}) + \rho(x_k) + \rho(y_k))/2}{1 - \beta L\rho(z_k)}, \\
r_{k+1} &= \max\{\rho(x_{k+1}), \rho(y_{k+1})\} \le qr_k \le \cdots \le q^{k+1} r_0,
\end{aligned}$$

where $\rho(x) = \|x - x_*\|$, $r_0 = \max\{\rho(x_0), \rho(y_0)\}$,

$$0 < q = \frac{(\beta/24)N\rho(x_0)^2 + \beta L(\rho(x_0) + \rho(y_0)/2}{1 - \beta L\rho(z_0)} < 1,$$

$z_k = (x_k + y_k)/2$ and β is defined in (14).

From Corollary 3, the convergence radius is

$$r = \frac{4}{5\beta L + \frac{1}{12}\sqrt{(60\beta L)^2 + 96\beta N}} < \frac{2}{5\beta L}$$

that corresponds to the previously received results in [10] for nonlinear equations ($m = n$).

Under the classic Lipschitz condition Theorem 2 for the uniqueness of the solution can be written as follow

Theorem 4. *Suppose x_* satisfies (1) and $F(x)$ has a continuous derivative $F^{'}(x)$ in $B(x_*, r)$. Moreover, $F^{'}(x_*)$ has full rank and $F^{'}(x)$ satisfies the classic Lipschitz condition as in (23). Let $r > 0$ satisfy*

$$\frac{\beta L r}{2} + \alpha\beta_0 L \leq 1.$$

Then, x_* is a unique solution of the problem (1) in $B(x_*, r)$.

5. Numerical experiments

We carried out a set of experiments on widely used test problems and compared the number of iterations under which the Gauss-Newton method (2), the Secant method [11], and the method (3) converge to the solution. We used the same initial points for all methods and the following stopping criteria:

$$\|x_{k+1} - x_k\| \leq \varepsilon \quad \text{and} \quad \|A_{k+1}^T F(x_{k+1})\| \leq \varepsilon,$$

where

- $A_{k+1} = F^{'}(x_{k+1})$ for the Gauss-Newton method (2);
- $A_{k+1} = F^{'}(z_{k+1})$ for the method (3);
- $A_{k+1} = F(x_{k+1}, x_k)$ for the Secant method, $F(x_{k+1}, x_k)$ is the divided difference of the first order of F [11].

Tabl. 1. The number of iterations to the solution with the accuracy $\varepsilon = 10^{-12}$

Example	Gauss-Newton	Secant	M-d (3)
Rosenbrock func. ($n = m = 4$) $x_0 = (-1.2, 1, -1.2, 1)$	5	4	4
Box-3D func. ($n = 3$, $m = 10$) $x_0 = (0, 10, 20)$	7	9	6
Gnedenko-Veibull dist. ($n = 2$, $m = 8$) $x_0 = (1, 1)$	7	–	6
Freidenstein-Ross func. ($n = m = 2$) $x_0 = (0.5, -2)$	43	18	10
Wood func. ($n = 4$, $m = 6$) $x_0 = (-3, -1, -3, -1)$	52	75	50
Bard func. ($n = 3$, $m = 15$) $x_0 = (1, 1, 1)$	10	–	9

In Table 1 we present the amount of iterations spent by each methods to compute an approximation to the solution of the examples from [9, 11] with the accuracy $\varepsilon = 10^{-12}$. The additional initial point y_0 we calculated in the following way: $y_0 = x_0 + 0.01$. The symbol '–' indicates that the Secant method does not converge to the solution with the desired accuracy, however the method converges for the lower accuracy ($\varepsilon = 10^{-8}$).

6. Conclusions

We studied the local convergence of the Gauss-Newton type method (3) under the generalized and classic Lipschitz conditions for the first- and second-order derivatives. We determined the convergence order and the radius of the method (3) as well as proved the uniqueness ball of the solution of the nonlinear least squares problem (1). The method (3) is not only more efficient than the Gauss-Newton and Secant methods in terms of the convergence order, but also in terms of the amount of iterations to the solution on a variety of test problems. Furthermore, the method (3) has promising characteristics for parallelization, which we plan to utilize for constructing and developing new parallel methods for solving the problem (1).

Acknowledgements

The simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at PDC Centre for High Performance Computing (PDC-HPC).

We are in debt to the unknown referee, who made useful remarks which helped a lot to improve the paper.

Bibliography

1. Bartish M. Ia. About applications of a modification of the Gauss-Newton method / M. Ia. Bartish, A. I. Chypurko // Visnyk of Lviv Univ. – Ser. Appl. Math. and Infor. – 1999. – Vol. 1. – P. 3-7. (in Ukrainian).
2. Bartish M. Ia. About one iterative method for solving functional equations / M. Ia. Bartish // Dopov. AN URSR. Ser. A. – 1968. – Vol. 30, № 5. – P. 387-391. (in Ukrainian).
3. Bartish M. Ia. About one modification of the Gauss-Newton method / M. Ia. Bartish, A. I. Chypurko, S. M. Shakhno // Visnyk of Lviv Univ. – Ser. Mech. Math. – 1995. – Vol. 42. – P. 35-38. (in Ukrainian).
4. Chen J. Convergence of Gauss-Newton method's and uniqueness of the solution / J. Chen, W. Li // Applied Mathematics and Computation. – 2005. – Vol. 170. – P. 686-705.
5. Chong C. Convergence behavior of Gauss-Newton's method and extensions of the Smale point estimate theory / C. Chong, N. Hu, J. Wang // Journal of Complexity. – 2010. – Vol. 26. – P. 268-295.
6. Dennis J. M. Numerical Methods for Unconstrained Optimization and Nonlinear Equations / J. M. Dennis, R. B. Schnabel. – New-York: Prentice-Hall, 1983.
7. Iakymchuk R. On the Convergence Analysis of a Two-Step Modification of the Gauss-Newton Method / R. Iakymchuk, S. M. Shakhno // PAMM. – 2014. – Vol. 14. – P. 813-814.
8. Kantorovich L. V. Functional Analysis / L. V. Kantorovich, G. P. Akilov. – Oxford: Pergamon Press, 1982.
9. More J. J. Testing Unconstrained Optimization Software / J. J. More, B. S. Garbow, K. E. Hillstrom // ACM Transactions on Mathematical Software. – 1981. – Vol. 7, № 1. – P. 17-41.

10. Shakhno S. M. Local convergence of a two-step Newton-like method for solving nonlinear equations under the generalized Lipschitz conditions / S. M. Shakhno // Physical and Mathematical Modeling and Information Technology. – 2012. – Vol. 16. – P. 163-172. (in Ukrainian).
11. Shakhno S. M. On an iterative algorithm of order 1.839... for solving the nonlinear least squares problems / S. M. Shakhno, O. P. Gnatyshyn // Applied Mathematics and Computation. – 2005. – Vol. 161. – P. 253-264.
12. Smale S. Newtons method estimates from data at one point / S. Smale // The Merging of Disciplines: New Directions in Pure / editors: R. Ewing, K. Gross, and C. Martin. – New York: Springer-Verlag, 1986. – P. 185-196.
13. Steward G. W. On the continuity of the generalized inverse / G. W. Steward // SIAM J. Appl. Math. – 1960. – Vol. 17, № 1. – P. 33-45.
14. Wang X. H. Convergence of Newton's method and inverse function theorem in Banach space / X. H. Wang // Math. Comp. – 1999. – Vol. 68. – P. 169-186.
15. Wang X. H. Convergence of Newton's method and uniqueness of the solution of equations in Banach space / X. H. Wang // IMA Journal of Numerical Analysis. – 2000. – Vol. 20. – P. 123-134.
16. Wedin P.-Å. Perturbation theory for pseudo-inverses / P.-Å. Wedin // BIT Numerical Mathematics. – 1973. – Vol. 13, № 2. – P. 217-232.
17. Werner W. Über ein Verfahren der Ordnung $1+\sqrt{2}$ zur Nullstellenbestimmung / W. Werner // Numer. Math. – 1979. – Vol. 32. – P. 333-342.

Roman Iakymchuk,
Department of Computational Science and Technology,
School of Computer Science and Communication,
KTH Royal Institute of Technology,
Lindstedtsvaĝen 5, Stockholm, 100 44, Sweden;
Stepan Shakhno, Halyna Yarmola,
Faculty of Applied Mathematics and Informatics,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine.

Received 05.08.2017; revised 07.10.2017

Журнал обчислювальної та прикладної математики | 2017 № 3 (126) | Journal of Numerical & Applied Mathematics

UDC 519.6

APPLICATION OF FINITE ELEMENTS METHOD FOR SOLVING VARIATIONAL PROBLEMS OF CHANNEL FLOWS

Y. V. Kokovska, M. M. Prytula, P. S. Venherskyi

Резюме. Виведено рівняння руху руслового потоку в псевдопризматичному руслі. Побудовано початково-крайову задачу руслового потоку в гідродинамічному наближенні. Сформульовано варіаційну постановку задачі, для якої при дискретизації за просторовою змінною використано метод скінченних елементів з базисними лінійними і квадратичними функціями та при дискретизації за часом – однокрокову рекурентну схему. В умовах рівноваги сил опору і сили земного тяжіння побудовано рівняння кінематичної хвилі, з врахуванням доданку із числом Рейнольдса та другою похідною за просторовою змінною. На тестовому прикладі показано порівняння цих двох підходів з врахуванням зміни градієнтів лінії середнього дна русла.

Abstract. The equation of motion of the channel flow in the pseudo prismatic channel is derived. The initial-boundary value problem of the channel flow in the hydrodynamic approximation is constructed. The variational problem was formulated and solved by method of finite elements with basic linear and quadratic functions for the spatial variable, and at time discretization one step recurrent scheme was constructed. In the conditions of the balance of the forces of resistance and the forces of gravity, the equation of the kinematic wave was derived, taking into account the addition with the number of Reynolds and the second derivative of the spatial variable. The test example shows a comparison of these two approaches, taking into account the change of the gradients of the line of the middle bottom of the channel.

1. Introduction

The transformation of the natural environment and global climate change are causing changes in hydrological systems. The estimation of such changes can be made on the basis of experimental data by comparing the hydrological characteristics before and after anthropogenic impact. However, the possibilities for such estimations are very limited, as the hydro meteorological conditions vary greatly. The main perspectives for the development of research methods and predictions of the behavior of natural hydrological systems are currently solved with the help of their mathematical modeling [1, 13].

In the general study of such an entire system, taking into account all the factors of influence, is a complex and not always appropriate task for study, therefore, only a some part of the region is investigated. The object of research can be the territory of the watershed of the river, which is characterized by

Key words. Variational problem, initial-boundary value problem, Galerkin approximations, channel flow, kinematic and hydrodynamic approximations.

similar climatic conditions and is under the influence of similar factors affecting the movement of fluid. For the description of water streams [2, 5, 10, 11], two approaches are most often used.

One of them is so-called hydrodynamic approach [2, 4, 9], in which the general laws of conservation of momentum, energy, mass are used to describe the processes. In this case, a complicated system of equations is used, usually non-linear, and in many cases, this task is cumbersome to estimate the amount of water.

The second approach is based on the equation of the kinematic wave [3], which are formed in the direction of the flow and occur under conditions of equilibrium of the forces of resistance and forces of gravity. These waves, which mainly affect the formation of the channel flow, which, unlike other types of waves, are formed in different directions and therefore quickly disappear.

In this paper, the flow of water is considered on one of the main elements of the watershed, namely in the inflows and in the main rivers, and these channels will be called pseudoprizmatic. Such channels are formed by moving a curve along a middle bottom line, while it is assumed that the depth of flow is very small compared with the radius of the curvature of the bottom line and the middle line of the free surface is horizontal in any normal section of the flow.

This mathematical model depends on many factors that can change fast enough, so this model must be stable to external and internal influences that significantly modify the solution of the problem. For approximation of the solution linear basic functions were used.

Since the problem is nonlinear, the solution acquires (gets) large positive and negative values, especially in the case of sharp changes of relief of the bottom of the flow. Therefore, the order of approximations of the solution and was shown the feasibility of this approach on different test examples [6].

2. Equation of water flow in pseudoprismatic channel

Choose a coordinate system such that the axis x is directed on the tangent straight to the middle bottom line, and the coordinate lines y and z are straight lines lying in the normal to the bottom of the plane so that y is directed horizontally (Fig. 1).

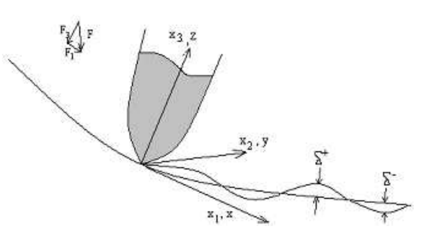


Fig. 1. Form of the channel flow.

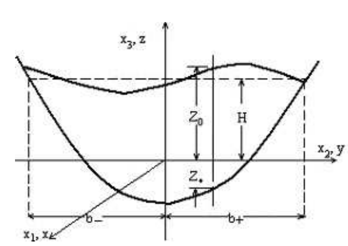


Fig. 2. Cross section of the flow.

The system of equations that characterize the motion of fluid:

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0; \tag{1}$$

$$\frac{\partial u}{\partial t}+\frac{\partial uu}{\partial x}+\frac{\partial uv}{\partial y}+\frac{\partial uw}{\partial z}=X-\frac{1}{\rho}\frac{\partial p}{\partial x}+\frac{1}{\rho}\left(\frac{\partial \tau_{xx}}{\partial x}+\frac{\partial \tau_{xy}}{\partial y}+\frac{\partial \tau_{xz}}{\partial z}\right); \tag{2}$$

$$\frac{\partial v}{\partial t}+\frac{\partial vu}{\partial x}+\frac{\partial vv}{\partial y}+\frac{\partial vw}{\partial z}=Y-\frac{1}{\rho}\frac{\partial p}{\partial y}+\frac{1}{\rho}\left(\frac{\partial \tau_{yx}}{\partial x}+\frac{\partial \tau_{yy}}{\partial y}+\frac{\partial \tau_{yz}}{\partial z}\right); \tag{3}$$

$$\frac{\partial w}{\partial t}+\frac{\partial wu}{\partial x}+\frac{\partial wv}{\partial y}+\frac{\partial ww}{\partial z}=Z-\frac{1}{\rho}\frac{\partial p}{\partial z}+\frac{1}{\rho}\left(\frac{\partial \tau_{zx}}{\partial x}+\frac{\partial \tau_{zy}}{\partial y}+\frac{\partial \tau_{zz}}{\partial z}\right). \tag{4}$$

Equation (1) is the equation of continuity for incompressible fluid, and (2) – (4) the Navier-Stokes equations in which u, v, w and X, Y, Z are projections of the velocity vector v and the vector of acceleration Capacitive forces F on the axis x, y, z.

We integrate the equation (1) with the area of the cross-section of the flow (Fig. 2):

$$\frac{1}{F}\int\limits_{b_-}^{b_+}dy\int\limits_{z_0}^{H}\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}\right)dz=0. \tag{5}$$

We use the differentiation formula under the integral sign, and taking into account the symmetry of the channel as to the XOZ plane, when all integrals of F containing $\frac{\partial}{\partial y}$ are equal zero, we obtain

$$\frac{1}{F}\int\limits_{b_-}^{b_+}dy\int\limits_{z_0}^{H}\left(\frac{\partial u}{\partial x}+\frac{\partial w}{\partial z}\right)dz=0. \tag{6}$$

Since on the surface of the bottom of the flow $z=z_0$ the vector of velocity is zero, then $u_{z=z_0}=0$.

We set the kinematic condition on a free surface:

$$w_{z=H}=\frac{\partial H}{\partial t}+u_{z=H}\frac{\partial H}{\partial x} \tag{7}$$

and the fact that the value $\int\limits_{b_-}^{b_+}\int\limits_{z_0}^{H}udzdy=Q$ is the rate of flow, then equation (6) is written as follows:

$$\frac{\partial Q}{\partial x}+\frac{\partial H}{\partial t}B=0. \tag{8}$$

Let us turn to the equations of motion. It is obvious that for flow in the gravity field $X=g\sin\delta, Y=0, Z=-g\cos\delta=-g_*$, where g-acceleration of gravity, δ – the sharp angle between the horizontal plane and the tangent to the line of the middle bottom. We integrate equation (4) for z and express the value of pressure:

$$\begin{gathered}\frac{p}{\rho}=\frac{p_H}{\rho}+(H-z_0)g_*-w^2+\frac{\partial}{\partial t}\int\limits_{z_0}^{H}wdz+\frac{\partial}{\partial x}\int\limits_{z_0}^{H}wudz+\\+\frac{\partial}{\partial y}\int\limits_{z_0}^{H}wvd\xi+\frac{1}{\rho}\left(\frac{\partial}{\partial x}\int\limits_{z_0}^{H}\tau_{zx}dz+\frac{\partial}{\partial y}\int\limits_{z_0}^{H}\tau_{zy}dz-\tau_{zz}\right),\end{gathered}\tag{9}$$

We substitute this value of p into equation (2), and we integrate the result by the area of the cross section F, we obtain:

$$\begin{gathered}\frac{\partial}{\partial t}\int\limits_{b_-}^{b_+}dy\int\limits_{z_*}^{z_0}udz+\frac{\partial}{\partial x}\int\limits_{b_-}^{b_+}dy\int\limits_{z_*}^{z_0}u^2dz=\\=g\left(\sin\delta-\frac{\partial z_0}{\partial x}\right)\int\limits_{b_-}^{b_+}\int\limits_{z_*}^{z_0}dzdy-\frac{1}{\rho}\int\limits_{b_-}^{b_+}\left(\tau_{zx}\right)_{z=z_*}dy+\varepsilon.\end{gathered}\tag{10}$$

where ε – additions that do not significantly affect the solution of the problem.

We use the expression defined in the hydraulics of turbulent flows

$$\frac{1}{\rho gF}\int\limits_{b_-}^{b_+}\left(\tau_{zx}\right)_{z=z_*}dy=\frac{Q^2}{K^2}=\frac{U^2}{C^2R},\tag{11}$$

where $K=CF\sqrt{R}$ – channel capacity; R – hydraulic radius; C – coefficient of Chezy. Then equation (10) will be written as:

$$\frac{1}{g}\left(\frac{\partial U}{\partial t}+U\frac{\partial\alpha U}{\partial x}-\frac{\alpha-1}{F}U\frac{\partial F}{\partial t}\right)=i-\frac{\partial H}{\partial x}-\frac{U^2}{C^2R}+\varepsilon.\tag{12}$$

If in (12) we neglect a addition ε, we obtain an hydrodynamic equation of one-dimensional unstable, slowly changing motion.

3. Initial-boundary problem of the channel flow in hydrodynamic approximation

If $Q=UF$, then from (8) follows that equation will be written as:

$$\frac{\partial(UF)}{\partial x}+B\frac{\partial H}{\partial t}=0$$

From where:

$$\frac{\partial(UF)}{\partial x}+\frac{\partial F}{\partial t}=0;\tag{13}$$

In equation (12) we neglect by addition ε, then equation will be written:

$$\frac{1}{g}\frac{\partial U}{\partial t}+\frac{\alpha}{g}U\frac{\partial U}{\partial x}-\frac{\alpha-1}{g}\frac{U}{F}\frac{\partial F}{\partial t}+\frac{1}{B}\frac{\partial F}{\partial x}+\frac{U^2}{C^2R}=i,\tag{14}$$

where U - flow velocity and F – cross-sectional area; $g = 9{,}8\ m/s^2$ – acceleration of gravity; C=const– coefficient of Chezy; $i = \sin\delta$, where δ – the angle of the midline of the channel bottom to the x-axis; $B = b_+ - b_-$ – width of the channel; R=const – hydraulic radius; α– parameter adjustments of movement.

Complement these equations by initial

$$U\,|_{t=0} = u_0(x), F\,|_{t=0} = f_0(x) \ \ on \ \ [0, L] \tag{15}$$

and boundary conditions

$$U(t, 0) = 0, F(t, 0) = 0. \tag{16}$$

obtain initial-boundary problem of the unknown – the flow velocity U and cross-sectional area F.

So, system of equations (13)–(16) describe initial-boundary problem of fluid flow in open pseudoprizmatic channel.

3.1. **Variational problem.** Choose spaces of allowable functions $H := L^2(\Omega)$, $V := H^1(\Omega)$, where $\Omega = [0,\ \ L]$.

To construct the variational problem multiply equation (13) an arbitrary function $\varphi \in V$, and the (14) – $\psi \in V$ and integrate the results by region Ω.

Input such bilinear form:

$$a(u, f, \varphi) = \int\limits_{\Omega} u\frac{\partial f}{\partial x}\varphi dx; \ \ b(u, \varphi) = \int\limits_{\Omega} u\varphi dx; \ \ c(u, \varphi) = \int\limits_{\Omega} \frac{\partial u}{\partial x}\varphi dx;$$

$$d(u, f, \varphi) = \int\limits_{\Omega} uf\varphi dx;$$

and linear functional

$$l(\varphi) = \int\limits_{\Omega} i\varphi dx.$$

Then variational formulation of initial-boundary problem (13)–(16) can be written as:

$$\begin{cases} Given: \quad u_0, f_0 \in H; \\ Find\ a\ pair: \quad (u, f) \in L^2(0, T; V \times V) \ \ such \ \ that \\ a(u(t), f(t), \varphi) + a(f(t), u(t), \varphi) + b(f'(t), \varphi) = 0; \\ \frac{1}{g}b(u'(t), \psi) + \frac{\alpha}{g}a(u(t), u(t), \psi) + \\ +\frac{1}{B}c(f(t), \psi) + \frac{1}{C^2R}d(u(t), u(t), \psi) - \\ -\frac{\alpha-1}{g}d(w(t), f'(t), \psi) = \langle l, \psi \rangle\,, \ \forall t \in (0,\ T], \\ b(u(0) - u_0, \varphi) = 0, b(f(0) - f_0, \psi) = 0, \ \forall\varphi, \psi \in V. \end{cases} \tag{17}$$

The solution to this problem will be search using the finite elements method.

4. Discretization variation problem in time variable

Divide the length of time [0,T] in $N_T + 1$ equal parts $[t_j, t_{j+1}]$ with length $\Delta t = t_{j+1} - t_j$, $j = 0, ..., N_T$. On each interval $[t_j, t_{j+1}]$ looking solutions of (5). Solutions$u(x,t), f(x,t) \in L^2(0,T;V)$ to this problem approximate by polynomials form

$$\begin{cases} u_{\Delta t}(x,t) = \{1-\omega(t)\}\, u^j(x) + \omega(t) u^{j+1}(x); \\ f_{\Delta t}(x,t) = \{1-\omega(t)\}\, f^j(x) + \omega(t) f^{j+1}(x); \\ t \in [t_{j,} t_{i+1}], j = 0, 1, ..., N_T - 1, \omega(t_j, t) = \frac{t-t_j}{\Delta t} \end{cases} \tag{18}$$

with unknown functions $u^j(x), f^j(x) \in V_h$.

For functional $l(x,t) \in V_h^1$ in problem (17) will use the following approximation

$$l_{\Delta t}(x,t) = l_{j+1/2} = l(t_{j+1/2}, x). \tag{19}$$

Then recurrent scheme [12, 14] will be written as:

$$\begin{cases} Given: \Delta t, \omega(t) = const > 0,\ u^j, f^j \in V \times V. \\ Find: u^{j+1}, f^{j+1} \in V \times V, such\quad that: \\ b\left(f^{j+1/2}, \varphi\right) + \Delta t \gamma a\left(u^j, f^{j+1/2}, \varphi\right) + \\ + \Delta t \gamma a\left(u^{j+1/2}, f^j, \varphi\right) + \Delta t \gamma a\left(f^{j+1/2}, u^j, \varphi\right) + \\ + \Delta t \gamma a\left(f^j, u^{j+1/2}, \varphi\right) = - a\left(u^j, f^j, \varphi\right) - a\left(f^j, u^j, \varphi\right); \\ \frac{1}{g} b\left(u^{j+1/2}, \psi\right) + \frac{\alpha}{g}\Delta t \beta\left[a\left(u^j, u^{j+1/2}, \psi\right) + a\left(u^{j+1/2}, u^j, \psi\right)\right] + \\ \frac{1}{B}\Delta t \beta c\left(f^{j+1/2}, \psi\right) + \frac{2}{C^2 R}\Delta t \beta d\left(u^j, u^{j+1/2}, \psi\right) - \\ - \frac{\alpha - 1}{g} d(w^j, f^{j+1/2}, \psi) = \\ = \left\langle l_{j+1/2}, \psi\right\rangle - \frac{\alpha}{g} a\left(u^j, u^j, \psi\right) - \frac{1}{B} c\left(f^j, \psi\right) - \frac{1}{C^2 R} d\left(u^j, u^j, \psi\right); \\ u^{j+1} = u^j + \Delta t u^{j+1/2}, f^{j+1} = f^j + \Delta t f^{j+1/2}. \end{cases} \tag{20}$$

The scheme provides that the initial solution (u^0, f^0) defined by initial conditions (16).

5. Discretization of variation problem for spatial variables

Choose a sequence of finite spaces approximations V_h of the space V with properties $\dim V_h \xrightarrow[h \to 0]{} \infty$. Then (u_h, v_h) – semi discrete approximation of solution *(u,f)*.

The interval [0, L] divide using sequence equally spaced units: $x_i = i \cdot h, i = 0, ..., N, h = \frac{L}{N}$ on N finite segments $[x_i, x_{i+1}], i = 0, 1, ..., N-1$.

Choose a base $\{\varphi_j\}_{j=1}^N$, $\{\psi_i\}_{i=1}^M$ in space approximations V_h.

Define functions $u_h^j(x) = \sum_{i=1}^N U_i^j \varphi_i(x), f_h^j(x) = \sum_{i=1}^N F_i^j \varphi_i(x)$ a schedule of for the basis functions $\{\varphi_i\}_{i=1}^N$, $\{\psi_i\}_{i=1}^M$ and unknown coefficients $U = \{U_i\}_{i=1}^M, F = \{F_i\}_{i=1}^N$.

Continuous piecewise defined basis functions $\{\varphi_i(x)\}_{i=1}^N$ of the space V_h chosen as linear polynomials, and $\{\psi_i(x)\}_{i=1}^M$- in the form of quadratic functions.
Functions $\{\varphi_i(x)\}_{i=1}^N$ and $\{\psi_i(x)\}_{i=1}^M$ denote as:

$$\varphi_i(x) = \begin{cases} 0, & 0 \le x \le x_{i-1}, \\ \frac{x-x_{i-1}}{h}, & x_{i-1} \le x \le x_i, \\ \frac{x_{i+1}-x}{h}, & x_i \le x \le x_{i+1}, \\ 0, & x_i \le x \le L. \end{cases}$$

$$\psi_i(x) = \begin{cases} 0, & 0 \le x \le x_{i-2}, \\ \frac{2(x-x_{i-2})(x-x_{i-1})}{h^2}, & x_{i-2} \le x \le x_{i-1}, \\ \frac{4(x-x_{i-1})(x-x_i)}{-h^2}, & x_{i-1} \le x \le x_i, \\ \frac{2(x-x_i)(x-x_{i+1})}{h^2}, & x_i \le x \le x_{i+1}, \\ 0, & x_{i+1} \le x \le L. \end{cases}$$

Overlaid matrices we obtain recurrent scheme as follows[7, 8]:

$$\begin{cases} Given: \ \ \Delta t, \gamma, \beta = const > 0; \ u^j, f^j \in R^n. \\ Find: \ u^{j+1}, f^{j+1} \in R^n, \\ such \ that: \\ \left[B1 + \Delta t\gamma A1\left(u^j\right) + \Delta t\gamma A2\left(u^j\right)\right] f^{j+1/2} + \\ + \left[\Delta t\gamma A3\left(f^j\right) + \Delta t\gamma A4\left(f^j\right)\right] u^{j+1/2} = \\ = -AP1\left(u^j, f^j\right) - AP2\left(f^j, u^j\right)\left[\frac{1}{B}\Delta t\beta C + \frac{\alpha-1}{g} D2(w^j)\right] f^{j+1/2} + \\ + \frac{1}{g}B2 + \frac{\alpha}{g}\Delta t\beta\left(A5\left(u^j\right) + A6\left(u^j\right)\right) + \\ + \frac{1}{C^2R} 2\Delta t\beta D1(u^j) u^{j+1/2} = \\ = L_{j+1/2} - \frac{\alpha}{g} AP3(u^j, u^j) - \frac{1}{B} CP(f^j) - \frac{1}{C^2R} DP(u^j, u^j) \\ u^{j+1} = u^j + \Delta t u^{j+1/2}, f^{j+1} = f^j + \Delta t f^{j+1/2}. \end{cases} \tag{21}$$

In this system, the values of the parameters of recurrent equations γ and β we choose from the conditions of their stability and provide the desired accuracy.

6. Equation of motion of water in the channel in the approach of the kinematic wave

So, the simplified equations of water in the form of equations of the kinematic wave [3]

$$\frac{\partial F}{\partial t} + \frac{3}{2} C\sqrt{UF}\frac{\partial F}{\partial x} - \frac{1}{Re}\frac{\partial^2 F}{\partial x^2} = Bw, \tag{22}$$

where $F = F(x,t)$ – cross-sectional area; $B = b_+(x,y) - b_-(x,y) = const$ – width of the channel; w – side inflow; Re – Reynolds number; i – slope of the bottom.

Initial and boundary conditions:

$$\begin{aligned} &F_{t=0} = F_0, \\ &\left(-\beta \tfrac{\partial F}{\partial x} + (1-\beta) F\right)\Big|_{x=0} = 0, \\ &\left(\gamma \tfrac{\partial F}{\partial x} + (1-\gamma) F\right)\Big|_{x=L} = 0, \quad \gamma, \beta > 0. \end{aligned} \tag{23}$$

Enter the denotation

$$\begin{aligned} &(h, \varphi) := \int\limits_{\Omega} h\varphi dx, \quad c\,(h, \varphi) := \int\limits_{\Omega} \nabla h \cdot \nabla \varphi dx, \\ &b\,(\xi; h, \varphi) := \int\limits_{\Omega} \xi^{m-1} h\alpha \cdot \nabla \varphi dx, \end{aligned} \tag{24}$$

$$\langle l, \varphi \rangle := \int\limits_{\Omega} R\varphi dx - \int\limits_{p} \hat{q}\varphi dx, \quad \forall \xi, h, \varphi \in V \tag{25}$$

Taking into account the designation (24), (25), variational formulation of the problem will look like:

$$\begin{cases} Given: \ h^0 \in V \ and \ \lambda \in (0, 1]; \\ Find: \ H^{k+\frac{1}{2}} \in V, \\ (H^{k+\frac{1}{2}}, \varphi) + \Delta t \lambda (mb(h^k; H^{k+\frac{1}{2}}, \varphi) + \frac{1}{Re} c(H^{k+\frac{1}{2}}, \varphi)) = \\ = \left\langle l_{k+1/2}, \varphi \right\rangle - b(h^k; h^k, \varphi) - \frac{1}{Re} c(h^k, \varphi) \quad \forall \varphi \in V, \\ h^{k+1} = h^k + \Delta t H^{k+\frac{1}{2}}, \ \ k = 0, ..., N_T. \end{cases} \tag{26}$$

The constructed variational problem of channel flow in kinematic approximation (26) makes it possible to find the depth of the flow in any point of time.

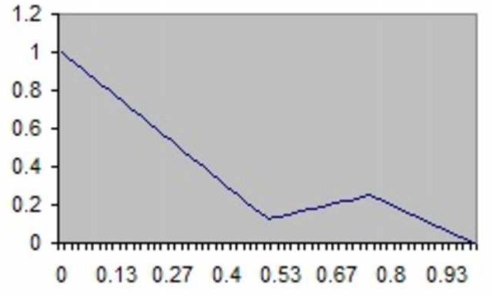


Fig. 3. Form of the bottom of the channel

7. Analysis of numerical experiments

We will test the obtained models on the test examples. The first example shows an effective use of quadratic approximations to eliminate the oscillation of solutions of hydrodynamic problem. The second example shows finding a solution of the problem of kinematic approximation, taking into account the

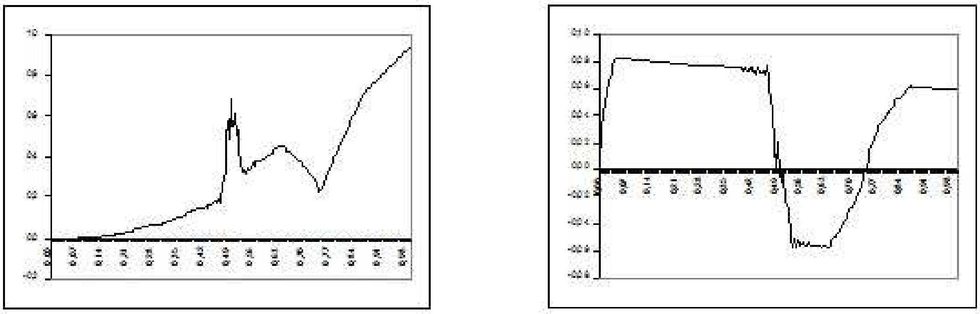


FIG. 4. Cross-sectional area and velocity (linear approximation 1000FE)

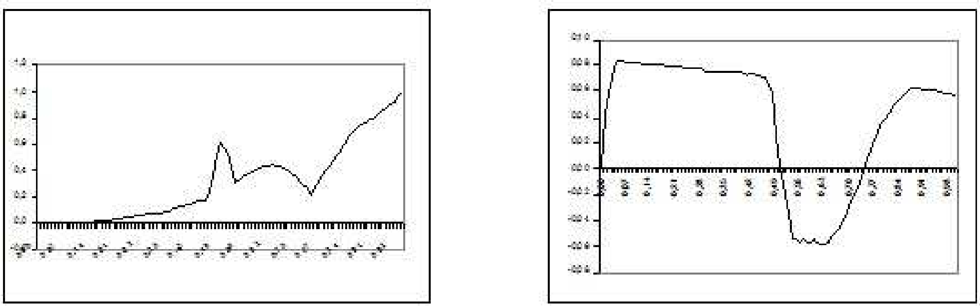


FIG. 5. Cross-sectional area and velocity (quadratic approximation 500FE)

addition with the second derivative. But the line of the middle bottom in examples 1 and 2 is the same.

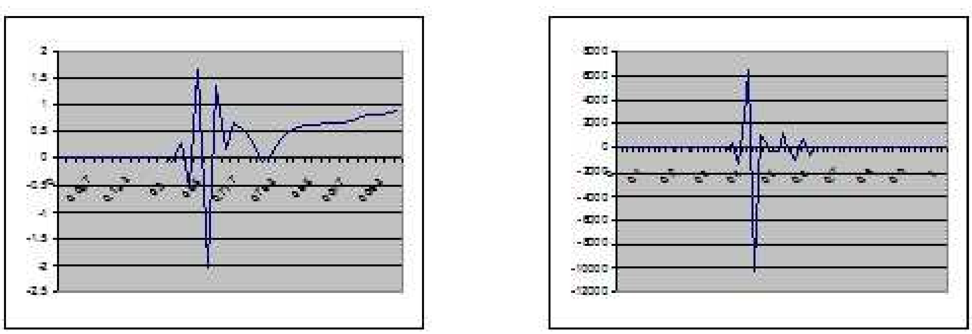
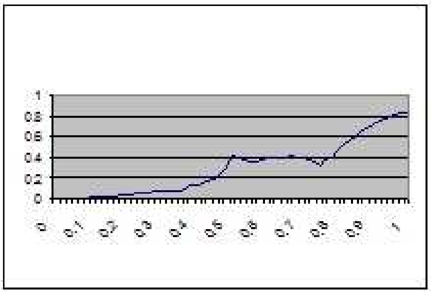


FIG. 6. Cross-sectional area and velocity (kinematic wave $\frac{1}{Re} = 0$)

Example 1. Input data: α=1, $0 \le x \le 1$, $0 \le t \le 1$, $\Delta t = 0.0001$, B=8, g=9.8, C=60, R=1, $F_0 = x^2$.



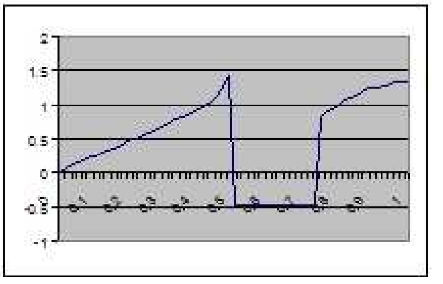


FIG. 7. Cross-sectional area and velocity (kinematic wave $Re = 20$)

Example 2. Input data: α=1, $0 \le x \le 1$, $0 \le t \le 1$, $\Delta t = 0.0001$, B=20, g=9.8, C=60, R=0, 1/Re=0 and Re = 20, $F(0,t) = 0$, $\frac{\partial F}{\partial x}\big|_{x=1} = 0$, $F_0 = x^2$, $U = C\sqrt{VF}$.

8. Conclusions

In this paper, a model of fluid motion in open pseudo prismatic channel in the hydrodynamic approximation, which is described by a system of equations with unknown variables of velocity and area cross-section of the flow, was constructed. In conditions of balance of the forces of resistance and gravity for this model the equation of the kinematic wave was written. In it the addition with the number of Reynolds and with the second derivative for spatial variable was taking into account. The initial-boundary problem was set for both approaches and its variational formulation was written. The variational problem was solving using the finite elements method. The choice of linear and quadratic basis functions was investigated in discretization a problem for a spatial variable and in application one-time recurrent integration scheme in time.

The obtained solutions of the problem are tested on examples with a complex relief of the bottom of the channel. In the model of hydrodynamic approximation, the expediency of increasing the order of approximation schemes for a spatial variable in approximations of velocity of flow and in the kinematic wave model the use of a regularized multiplier are shown. The test example shows a comparison of the two approaches, taking into account the change of the gradients of the line of the middle bottom of the channel.

Bibliography

1. Burakov D. A. Two-dimensional nonstationary mathematical model of water catchment / D. A. Burakov, E. D. Karepova, V. V. Shaidurov // Notes on Numerical Fluid Mechanics and Multidisciplinary Design. – 2011. – № 115. – P. 253-266.
2. Venherskyi P. S. One of approaches of modeling of processes of river-bed flow of fluid / P. S. Venherskyi, Y. V. Kokovska // Visnyk of Lviv University. – Ser. of App. Math. and Inform. – 2009. – № 15. – P. 178-195. (in Ukrainian).
3. Venherskyi P. S. The equation of the surface flow in the kinematic waves approximation for channels with uneven bottom / P. S. Venherskyi, V. M. Kyrylych, Y. V. Kokovska // Applied problems of mechanics and mathematics. – 2016. – № 14. – P. 31-36.
4. Hladkyi A. V. Numerical-analytical methods for wave processes studies. / A. V. Hladkyi, I. V. Sergijenko, V. V. Skopeckyi. – Science opinion. – Kyiv. – 2001. – 452 p.

5. Kartvelishvili N. A. Flows in undeformed channels / N. A. Kartvelishvili. – L.: Gidrometeoizdat. – 1973. – 279 p.
6. Kokovska Y. V. On comparison of two approaches of numerical modelling of channel flow of fluid / Y. V. Kokovska, P. S. Venherskyi // XXI All-Ukrainian Scientific Conference "Modern Problems of Applied Mathematics and Informatics". – Collection of scientific works. – Lviv. – 2015. – P. 183-184.
7. Kokovska Y. Numerical solution of variational problems of surface and channel flows by finite elements method / Y. Kokovska, M. Prytula, P. Venherskyi // Ukrainian International Conference on Applied Mathematics. – Lviv. – 2017. – P. 86-87.
8. Kokovska Y. Use linear and quadratic approximations for solution of problem of fluid flow in chanel / Y. Kokovska // Visnyk of Lviv. University. – Ser. of App. Math. and Inform. – 2012. – No 18. – P. 165-175. (in Ukrainian).
9. Kokovska Y. Mathematical modeling of fluid flows in pseudoprismatic channels. / Y. Kokovska // Modern Problems of Mechanics and Mathematics: Ya.S.Pidstrygach Institute of Applied Problems of Mechanics and Mathematics of the National Academy of Sciences of Ukraine. – Lviv. – 2011. – P. 189-190.
10. Kuchment L. S. Dynamic-stochastic models of river flow formation / L. S. Kuchment, A. N. Gelfan. – RAN, Institute of Water Problems. – M.: Science. – 1993. – 112 p.
11. Ladikov-Royev Yu. P. Mathematical modeling of the process of formation of stream flow and stream erosion of soils / Yu. P. Ladikov-Royev, V. I. Panchuk, G. A. Chechko // Management and Informatics Problems. – 2000. – № 3. – P. 76-85.
12. Savula Y. G. Numerical analysis of the problems of mathematical physics by variational methods / Y. G. Savula // I. Franko National univ. of Lviv. – Lviv. – 2004. – P. 222.
13. Susidko M. M. Mathematical modeling of formation processes as a basis for forecasting systems / M. M. Susidko // Science Collection of the KNU "Hydrology, Hydrochemistry and Hydroecology". – 2000. – Vol. 1. – P. 32-40.
14. Shynkarenko G. A. Projection-net methods for solving initial-boundary problems / G. A. Shynkarenko. – Kiev: NMC VO. – 1991. – 87 p.

Yaryna Kokovska, Mykola Prytula, Petro Venherskyi,
Faculty of Applied Mathematics and Informatics,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine.

Received 14.10.2017; revised 27.10.2017

Журнал обчислювальної та прикладної математики 2017 № 3 (126) Journal of Numerical & Applied Mathematics

UDC 519.6

APPLICATION OF (G'/G) – EXPANSION METHOD TO TWO KORTEWEG – DE VRIES TYPE DYNAMIC SYSTEMS

I. S. Mykhailiuk, M. M. Prytula

Резюме. Метод G'/G розвинення [12] застосовано до двох нелінійних динамічних ситем типу Кортевега – де Фріза [20]. Для обох систем побудовано розв'язки типу біжучих хвиль у формі гіперболічних, раціональних і тригонометричних функцій. Отримані результати порівняно з результатами, отриманими $tanh$- методом [4] і графічно проаналізовано.

Abstract. The (G'/G) – expansion method [15] is applied to two Korteweg – de Vries type nonlinear dynamic systems [1]. For both systems the traveling wave solutions in the form of hyperbolic, rational and trigonometric functions are constructed. The obtained results are compared to ones derived by means of the tanh – method [6] and graphically analyzed.

1. Introduction

Solutions to nonlinear evolution equations (NEE) play a crucial role in mathematical physics, therefore more and more scientists from all over the world dedicate their studies to investigate such equations. Nonlinear wave phenomena appear in various scientific and engineering fields, such as fluid mechanics, plasma physics, optical fibers, biology, solidstate physics, chemical kinematics, chemical physics and geochemistry.

With the advent of computers many effective numeric methods for finding approximate solutions to partial differential equations (PDEs) appeared. On the other hand, the creation of modern powerful computer algebra systems, such as MATLAB, MATHEMATICA and MAPLE, simplified the analytical investigation of NEEs, assisting mathematicians in their tiny computations. Hence during the past five decades a wide variety of analytical methods for finding exact solutions to NEEs was developed.

Recently, the (G'/G) – expansion method, firstly introduced by Wang et al. [15], has become widely used for many PDEs. It turned out that the method just mentioned provides solutions in a more general form compared to other analytical methods (e.g. the tanh – method [6]). What is more, with a certain choice of arbitrary parameters in the (G'/G) – expansion method some well-known solutions to PDEs can be rediscovered.

In paper [14], the authors constructed soliton solutions for two Korteweg–de Vries (KdV) type nonlinear dynamic systems [1, 3] by means of the tanh –

Key words. (G'/G) – expansion method, Korteweg – de Vries type dynamic system, soliton solution.

method [6]. In this work, we investigate these systems using the (G'/G) – expansion method and construct solutions in more general form. The rest of the paper is organized as follows. In Section 2, we describe the (G'/G) – expansion method [15] for finding traveling wave solutions to nonlinear evolution equations. In Section 3, we provide a brief overview of the main generalizations of the method being discussed. In Sections 4 and 5, we apply the method to two nonlinear KdV type dynamic systems [1,3], and analyze the obtained solutions. Finally, in Section 6, we summarize our results.

2. Description of the $\left(\frac{G'}{G}\right)$ – expansion method

Suppose that a nonlinear equation, say in two independent variables x and t, is given by

$$P\left(u, u_t, u_x, u_{tt}, u_{xx}, u_{xt}, ...\right) = 0, \tag{1}$$

where $u = u\left(x, t\right)$ is an unknown function, P is a polynomial in $u = u\left(x, t\right)$ and its various partial derivatives, in which the highest order derivatives and nonlinear terms are involved. In the following we give the main steps of the (G'/G) – expansion method [15].

Step 1. Combining independent variables x and t into one variable

$$\xi = x - Vt, \tag{2}$$

we suppose that $u\left(x, t\right) = u\left(\xi\right)$. Traveling wave variable (2) permits us to reduce Eq. (1) to an ordinary differential equation (ODE) for $u\left(x, t\right) = u\left(\xi\right)$

$$P\left(u, -Vu', u', V^2u'', -Vu'', u'', ...\right) = 0. \tag{3}$$

Step 2. Suppose that the solution to ODE (3) can be expressed by a polynomial in (G'/G) as follows:

$$u(\xi) = \sum_{i=0}^{m} \alpha_i \left(\frac{G'}{G}\right)^i, \tag{4}$$

where $G = G\left(\xi\right)$ satisfies the second order linear ODE in the form of

$$G'' + \lambda G' + \mu G = 0, \tag{5}$$

$\alpha_i \left(i = \overline{0, m}\right), \lambda, \mu$ are constants to be determined later, $\alpha_m \neq 0$. The positive integer m can be determined by considering the homogeneous balance between the highest order derivatives and nonlinear terms appearing in ODE (3).

Step 3. By substituting (4) into Eq. (3) and using the second order LODE (5), collecting all terms with the same order of (G'/G) together, the left-hand side of Eq. (3) is converted into another polynomial in (G'/G). Equating each coefficient of this polynomial to zero yields a set of algebraic equations for $\alpha_i \left(i = \overline{0, m}\right), \lambda$ and μ.

Step 4. Assuming that the constants $\alpha_i \left(i = \overline{0, m}\right), \lambda, \mu$ and V can be obtained by solving the algebraic equations in Step 3, since the general solutions to the second order linear ODE (5) have been well known for us, then substituting $\alpha_i \left(i = \overline{0, m}\right), \lambda, \mu, V$ and the general solutions to Eq. (5) into (4) we obtain traveling wave solutions to the original nonlinear evolution equation (1).

As it was already mentioned, the solution to Eq. (5) is well-known for us and can be easily derived by the Euler method:

$$G(\xi)=\begin{cases}\left(A_1\sinh\frac{\xi\sqrt{\lambda^2-4\mu}}{2}+A_2\cosh\frac{\xi\sqrt{\lambda^2-4\mu}}{2}\right)e^{-\frac{1}{2}\lambda\xi}, \\ \qquad\qquad\qquad if\ \lambda^2-4\mu>0, \\ (A_1+A_2\xi)\,e^{-\frac{1}{2}\lambda\xi}, \quad if\ \lambda^2-4\mu=0, \\ \left(A_1\sin\frac{\xi\sqrt{4\mu-\lambda^2}}{2}+A_2\cos\frac{\xi\sqrt{4\mu-\lambda^2}}{2}\right)e^{-\frac{1}{2}\lambda\xi}, \\ \qquad\qquad\qquad if\ \lambda^2-4\mu<0.\end{cases} \tag{6}$$

3. Main generalizations of the (G'/G) – expansion method

Since 2008, when the (G'/G) – expansion method was introduced by Wang et al. [15], many modifications and generalizations of the algorithm have been developed, each of which concerned different aspect of the method. Therefore, it is worth classifying them by that aspect.

3.1. **Homogeneous balance value.** The classical method [15] assumed that the homogeneous balance value, which determines a degree of polynomial (4), is a positive integer. In paper [4] the authors used a transform to handle the equations with negative or fractional homogeneous balance value. Let m be a value of balance for a certain equation. If $m=\frac{p}{q}$ is a fraction in the lowest terms, then we set the solution

$$u(\xi)=v^{\frac{p}{q}}(\xi),$$

and when m is a negative integer, then we set

$$u(\xi)=v^{m}(\xi),$$

then substitute the new expression for $u(\xi)$ into (3) and recompute the balance value for a new equation, which is now guaranteed to be a positive integer [4].

3.2. **Representation of the solution to NEE.** Another way to modify the original method is to replace the polynomial in $\left(\frac{G'}{G}\right)$ with a more general structure.

In works [2] and [16] the solution was suggested to be found in the following form:

$$u(\xi)=a_0+\sum_{i=1}^{n}\left[a_i\left(\frac{G'}{G}\right)^{i}+b_i\left(\frac{G'}{G}\right)^{i-1}\sqrt{\sigma\left(1+\frac{1}{\mu}\left(\frac{G'}{G}\right)^{2}\right)}\right],$$

and, moreover, the function $G=G(\xi)$ was found as a solution to simplified equation

$$G''+\mu G=0,$$

where μ is a constant to be determined.

Yet another form of the solution representation was introduced in papers [21], [17] and [13], namely the solution was supposed to have the following form:

$$u(\xi) = \sum_{i=-n}^{n} \alpha_i \left(\frac{G'}{G}\right)^i,$$

i. e. the expansion included the terms with negative degrees.

As it is shown in the corresponding works, both mentioned representations of function $u = u(\xi)$ yield more general solutions to certain NEEs [2,13,16,17,21].

3.3. **Auxiliary equation for function $G = G(\xi)$.** Other modifications of the method affected the form of the auxiliary equation, which in the classical $\left(\frac{G'}{G}\right)$ – expansion method is of the form (5). One of the most frequently used equations was the nonlinear one of the following form:

$$GG'' = AG^2 + BGG' + C\left(G'\right)^2,$$

where the prime denotes the derivative with respect to ξ; A, B, C are all real parameters.

This improvement of the method was firstly introduced by Liu et al. in [5] to obtain more general solutions to NEEs in comparison with the classical method. It was successfully applied to some well-known equations of mathematical physics, among other, in works [5, 7–12].

3.4. **Coefficient of the polynomial in $\left(\frac{G'}{G}\right)$.** One more generalization of the original method was the idea to find a solution to NEEs as a polynomial in $\left(\frac{G'}{G}\right)$ with variable coefficients [20], namely

$$u(\xi) = \sum_{i=1}^{n} \alpha_i(X) \left(\frac{G'}{G}\right)^i + \alpha_0(X),$$

where $\alpha_i = \alpha_i(X)\left(i = \overline{0, n}\right), \xi = \xi(X)$ are functions to be determined. As in the classical method, function $G = G(\xi)$ satisfies Eq. (5). The rest of the algorithm remains the same, except that at the third step one need to solve a system of ordinary differential equations rather than algebraic ones.

The described idea was successfully used to solve some NEEs in papers [18–20].

4. Application: Example 1

Consider the following Korteweg – de Vries (KdV) type nonlinear dynamic system [1, 3]

$$\begin{cases} u_t = u_{xxx} - v_x, \\ v_t = -2v_{xxx} - uv_x. \end{cases} \tag{7}$$

Let us solve system (7) by use of the (G'/G) – expansion method.

Step 1. Introducing traveling wave variable $\xi = x - Vt$, we reduce system (7) to a system of ODE for $u = u(\xi)$ and $v = v(\xi)$

$$\begin{cases} -Vu' = u''' - v', \\ -Vv' = -2v''' - uv'. \end{cases} \tag{8}$$

Suppose that the solution to system (8) can be expressed by polynomials in (G'/G) as follows:

$$u(\xi) = \sum_{i=0}^{m} \alpha_i \left(\frac{G'}{G}\right)^i, \quad v(\xi) = \sum_{i=0}^{n} \beta_i \left(\frac{G'}{G}\right)^i. \tag{9}$$

Considering the homogeneous balance between u''' and v', v''' and uv' in the first and the second equations of system (8) correspondingly, we obtain a simple system of algebraic equations

$$\begin{cases} m + 3 = n + 1, \\ n + 3 = m + n + 1, \end{cases} \tag{10}$$

from which it can be easily found that $m = 2$ and $n = 4$.

Step 2. Considering (9) and (10), we find the solution to system (8) in the following form:

$$\begin{cases} u(\xi) = \alpha_2 \left(\frac{G'}{G}\right)^2 + \alpha_1 \left(\frac{G'}{G}\right) + \alpha_0, \\ v(\xi) = \beta_4 \left(\frac{G'}{G}\right)^4 + \beta_3 \left(\frac{G'}{G}\right)^3 + \beta_2 \left(\frac{G'}{G}\right)^2 + \beta_1 \left(\frac{G'}{G}\right) + \beta_0, \end{cases} \tag{11}$$

where function $G = G(\xi)$ satisfies the second order linear ODE (5), λ, μ, V, α_i $\left(i = \overline{0,2}\right)$, β_j $\left(j = \overline{0,4}\right)$ are all constants to be determined later, $\alpha_2 \neq 0$, $\beta_4 \neq 0$.

Step 3. Substituting (11) into system (8) and collecting all terms with the same power of $\left(\frac{G'}{G}\right)$ together, the left-hand sides of equations (8) are converted into another polynomials in $\left(\frac{G'}{G}\right)$. Equating each coefficient of these polynomials to zero yields a set of simultaneous algebraic equations for $\lambda, \mu, V, \alpha_i \left(i = \overline{0,2}\right), \beta_j \left(j = \overline{0,4}\right)$ as follows:

– from *the first equation* in (8):

$$\begin{aligned}
0:&\quad \alpha_1\lambda^2\mu + 6\alpha_2\lambda\mu^2 + 2\alpha_1\mu^2 - \beta_1\mu + \alpha_1\mu V = 0 \\
1:&\quad \alpha_1\lambda^3 + 6\alpha_2\lambda^2\mu + 8\alpha_2\mu\left(\lambda^2 + 2\mu\right) + 8\alpha_1\lambda\mu - \beta_1\lambda - 2\beta_2\mu + \\
&\quad +V\left(\alpha_1\lambda + 2\alpha_2\mu\right) = 0 \\
2:&\quad 8\alpha_2\lambda\left(\lambda^2 + 2\mu\right) + 7\alpha_1\lambda^2 + 36\alpha_2\lambda\mu + 8\alpha_1\mu - 2\beta_2\lambda - 3\beta_3\mu - \\
&\quad -\beta_1 + V\left(2\alpha_2\lambda + \alpha_1\right) = 0 \\
3:&\quad 8\alpha_2\left(\lambda^2 + 2\mu\right) + 30\alpha_2\lambda^2 + 12\alpha_1\lambda + 24\alpha_2\mu - 3\beta_3\lambda - 4\beta_4\mu - \\
&\quad -2\beta_2 + 2\alpha_2 V = 0 \\
4:&\quad 54\alpha_2\lambda + 6\alpha_1 - 4\beta_4\lambda - 3\beta_3 = 0 \\
5:&\quad 24\alpha_2 - 4\beta_4 = 0;
\end{aligned}$$

– from *the second equation* in (8):

$$
\begin{aligned}
0: \quad & -\alpha_0\beta_1\mu - 2\beta_1\lambda^2\mu - 12\beta_2\lambda\mu^2 - 12\beta_3\mu^3 - 4\beta_1\mu^2 + \beta_1\mu V = 0 \\
1: \quad & -\alpha_0\beta_1\lambda - \alpha_1\beta_1\mu - 2\alpha_0\beta_2\mu - 2\beta_1\lambda^3 - 28\beta_2\lambda^2\mu - 72\beta_3\lambda\mu^2 - \\
& -16\beta_1\lambda\mu - 48\beta_4\mu^3 - 32\beta_2\mu^2 + \beta_1\lambda V + 2\beta_2\mu V = 0 \\
2: \quad & -\alpha_1\beta_1\lambda - 2\alpha_0\beta_2\lambda - \alpha_2\beta_1\mu - 2\alpha_1\beta_2\mu - 3\alpha_0\beta_3\mu - \alpha_0\beta_1 - 16\beta_2\lambda^3 - \\
& -114\beta_3\lambda^2\mu - 14\beta_1\lambda^2 - 216\beta_4\lambda\mu^2 - 104\beta_2\lambda\mu - 120\beta_3\mu^2 - 16\beta_1\mu + \\
& +2\beta_2\lambda V + 3\beta_3\mu V + \beta_1 V = 0 \\
3: \quad & -\alpha_2\beta_1\lambda - 2\alpha_1\beta_2\lambda - 3\alpha_0\beta_3\lambda - 2\alpha_2\beta_2\mu - 3\alpha_1\beta_3\mu - 4\alpha_0\beta_4\mu - \\
& -\alpha_1\beta_1 - 2\alpha_0\beta_2 - 54\beta_3\lambda^3 - 296\beta_4\lambda^2\mu - 76\beta_2\lambda^2 - 336\beta_3\lambda\mu - \\
& -24\beta_1\lambda - 304\beta_4\mu^2 - 80\beta_2\mu + 3\beta_3\lambda V + 4\beta_4\mu V + 2\beta_2 V = 0 \\
4: \quad & -2\alpha_2\beta_2\lambda - 3\alpha_1\beta_3\lambda - 4\alpha_0\beta_4\lambda - 3\alpha_2\beta_3\mu - 4\alpha_1\beta_4\mu - \alpha_2\beta_1 - \\
& -2\alpha_1\beta_2 - 3\alpha_0\beta_3 - 128\beta_4\lambda^3 - 222\beta_3\lambda^2 - 784\beta_4\lambda\mu - 108\beta_2\lambda - \\
& -228\beta_3\mu - 12\beta_1 + 4\beta_4\lambda V + 3\beta_3 V = 0 \\
5: \quad & -3\alpha_2\beta_3\lambda - 4\alpha_1\beta_4\lambda - 4\alpha_2\beta_4\mu - 2\alpha_2\beta_2 - 3\alpha_1\beta_3 - 4\alpha_0\beta_4 - 488\beta_4\lambda^2 - \\
& -288\beta_3\lambda - 496\beta_4\mu - 48\beta_2 + 4\beta_4 V = 0 \\
6: \quad & -\alpha_2\left(4\beta_4\lambda + 3\beta_3\right) - 4\alpha_1\beta_4 + 2\left(-120\beta_4\lambda - 60\left(3\beta_4\lambda + \beta_3\right)\right) = 0 \\
7: \quad & -4\alpha_2\beta_4 - 240\beta_4 = 0.
\end{aligned}
$$

In addition to this, the highest order coefficients in (11) are supposed to be nonzero:

$$\alpha_2 \neq 0, \ \beta_4 \neq 0. \tag{12}$$

Step 4. Solving the system of algebraic equations from the previous step with conditions (12) with the aid of MATHEMATICA yields *four sets of solutions*:

– Set 1.

$$
\begin{gathered}
V = \lambda^2 - 4\mu, \quad \alpha_0 = -\lambda^2 - 56\mu, \quad \alpha_1 = -60\lambda, \quad \alpha_2 = -60, \\
\beta_1 = -120\left(\lambda^3 + 2\lambda\mu\right), \quad \beta_2 = -240\left(2\lambda^2 + \mu\right), \\
\beta_3 = -720\lambda, \quad \beta_4 = -360,
\end{gathered} \tag{13}
$$

where λ, μ and β_0 are arbitrary constants.

– Set 2.

$$
\begin{gathered}
V = 4\mu - \lambda^2, \quad \alpha_0 = -3\left(3\lambda^2 + 8\mu\right), \quad \alpha_1 = -60\lambda, \quad \alpha_2 = -60, \\
\beta_1 = -720\lambda\mu, \quad \beta_2 = -360\left(\lambda^2 + 2\mu\right), \\
\beta_3 = -720\lambda, \quad \beta_4 = -360,
\end{gathered} \tag{14}
$$

where λ, μ and β_0 are arbitrary constants.

– Set 3.

$$
\begin{gathered}
V = \lambda^2, \quad \mu = 0, \quad \alpha_0 = -\lambda^2, \quad \alpha_1 = -60\lambda, \quad \alpha_2 = -60, \\
\beta_1 = -120\lambda^3, \quad \beta_2 = -480\lambda^2, \quad \beta_3 = -720\lambda, \quad \beta_4 = -360,
\end{gathered} \tag{15}
$$

where λ and β_0 are arbitrary constants.

– Set 4.

$$
\begin{gathered}
V = -\lambda^2, \quad \mu = 0, \quad \alpha_0 = -9\lambda^2, \quad \alpha_1 = -60\lambda, \quad \alpha_2 = -60, \\
\beta_1 = 0, \quad \beta_2 = -360\lambda^2, \quad \beta_3 = -720\lambda, \quad \beta_4 = -360,
\end{gathered} \tag{16}
$$

where λ and β_0 are arbitrary constants.

Finally, substituting solutions (13)–(16) with the general solution to linear ODE (5) into representation (11) we obtain *four separate sets* of traveling wave solutions to the KdV type dynamic system (7) as follows.

Solutions set 1. Constants set (13) yields three families of solutions:

– when $\lambda^2 - 4\mu > 0$, we get the family of hyperbolic functions solutions

$$\begin{cases} u(\xi) = & -\dfrac{15\left(A_1^2-A_2^2\right)\sigma}{\left(A_1 \sinh\frac{\xi\sqrt{\sigma}}{2}+A_2\cosh\frac{\xi\sqrt{\sigma}}{2}\right)^2} - \sigma, \\ v(\xi) = & -\dfrac{15\left(A_1^2-A_2^2\right)\sigma^2\left(4A_2A_1\sinh\xi\sqrt{\sigma}+2\left(A_1^2+A_2^2\right)\cosh\xi\sqrt{\sigma}+A_1^2-A_2^2\right)}{2\left(A_1\sinh\frac{\xi\sqrt{\sigma}}{2}+A_2\cosh\frac{\xi\sqrt{\sigma}}{2}\right)^4} + \\ & +\beta_0 + 120\mu\left(\lambda^2-\mu\right), \end{cases} \tag{17}$$

where $\xi = x - \left(\lambda^2 - 4\mu\right)t$, $\sigma = \lambda^2 - 4\mu$, A_1, A_2, β_0 are arbitrary constants; in particular, setting $\lambda = \pm\sqrt{\frac{8}{3}}|k_1|$, $\mu = -\frac{1}{3}k_1^2$, $A_1 = 0$, $\beta_0 = a_{20}$, we obtain exactly the soliton solution, found by means of the tanh – method in [14];

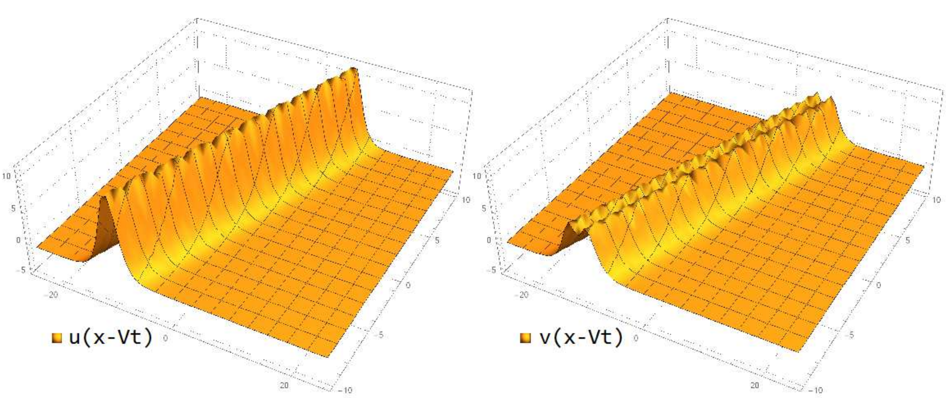


FIG. 1. Hyperbolic functions solution (17) when $A_1 = 1$, $A_2 = 1.2$, $\lambda = 2.2$, $\mu = 1$, $\beta_0 = -460.8$

– when $\lambda^2 - 4\mu = 0$, we get the family of rational functions solutions

$$\begin{cases} u(\xi) = & -\dfrac{60A_2^2}{(A_2\xi+A_1)^2}, \\ v(\xi) = & \dfrac{360\left(\mu^2(A_2\xi+A_1)^4-A_2^4\right)}{(A_2\xi+A_1)^4} + \beta_0, \end{cases} \tag{18}$$

where $\xi = x$, A_1, A_2, β_0 are arbitrary constants;

– when $\lambda^2 - 4\mu < 0$, we get the family of trigonometric functions solutions

$$\begin{cases} u(\xi) = & -\dfrac{15\left(A_1^2+A_2^2\right)\sigma}{\left(A_1 \sin\frac{\xi\sqrt{\sigma}}{2}+A_2\cos\frac{\xi\sqrt{\sigma}}{2}\right)^2} + \sigma, \\ v(\xi) = & -\dfrac{15\left(A_1^2+A_2^2\right)\sigma^2\left(-4A_2A_1\sin\xi\sqrt{\sigma}+2\left(A_1^2-A_2^2\right)\cos\xi\sqrt{\sigma}+A_1^2+A_2^2\right)}{2\left(A_1\sin\frac{\xi\sqrt{\sigma}}{2}+A_2\cos\frac{\xi\sqrt{\sigma}}{2}\right)^4} + \\ & +\beta_0 + 120\mu\left(\lambda^2-\mu\right), \end{cases} \tag{19}$$

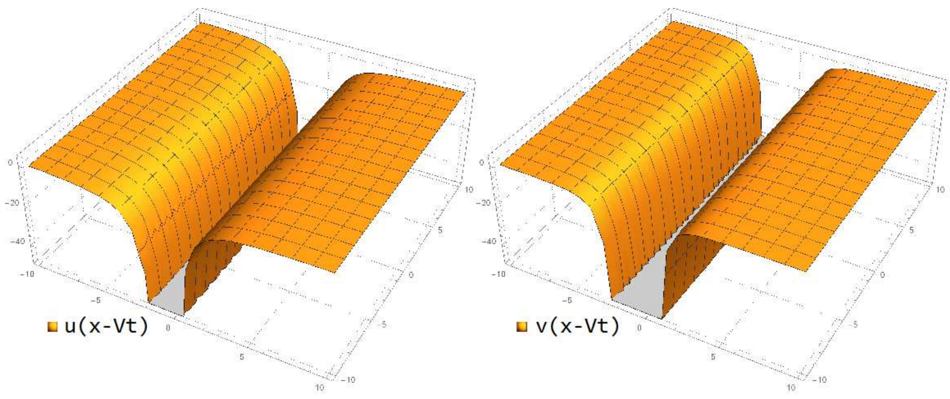


FIG. 2. Rational functions solution (18) when $A_1 = 1,\ A_2 = 1.2,\ \lambda = 1,\ \mu = 0.25,\ \beta_0 = -22.5$

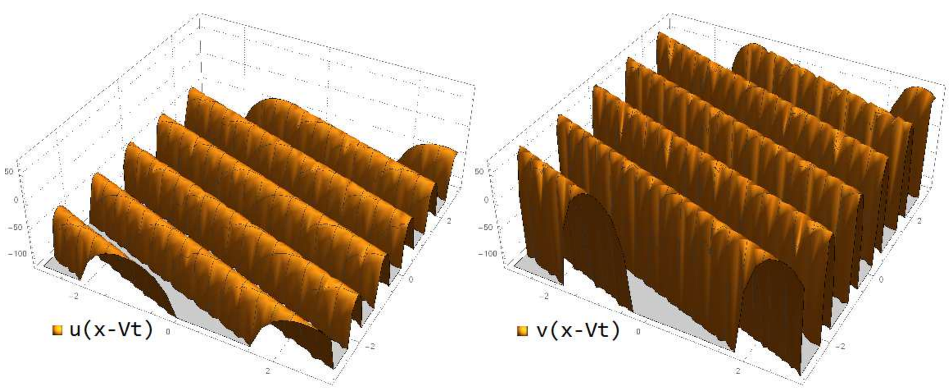


FIG. 3. Trigonometric functions solution (19) when $A_1 = 1,\ A_2 = 1.2,\ \lambda = 1,\ \mu = 1,\ \beta_0 = 0$

where $\xi = x - \left(\lambda^2 - 4\mu\right) t,\ \sigma = 4\mu - \lambda^2,\quad A_1, A_2, \beta_0$ are arbitrary constants.

Solutions set 2. Constants set (14) yields three families of solutions:

– when $\lambda^2 - 4\mu > 0$, we get the family of hyperbolic functions solutions

$$\begin{cases} u\left(\xi\right) = & -\dfrac{15\sigma\left(A_1^2 - A_2^2\right)}{\left(A_1 \sinh\frac{\xi\sqrt{\sigma}}{2} + A_2 \cosh\frac{\xi\sqrt{\sigma}}{2}\right)^2} - 9\sigma \\ v\left(\xi\right) = & -\dfrac{45\left(A_1^2 - A_2^2\right)^2\sigma^2}{2\left(A_1 \sinh\frac{\xi\sqrt{\sigma}}{2} + A_2 \cosh\frac{\xi\sqrt{\sigma}}{2}\right)^4} + \beta_0 + 360\mu^2, \end{cases} \tag{20}$$

where $\xi = x + \left(\lambda^2 - 4\mu\right) t,\ \sigma = \lambda^2 - 4\mu,\quad A_1, A_2, \beta_0$ are arbitrary constants;

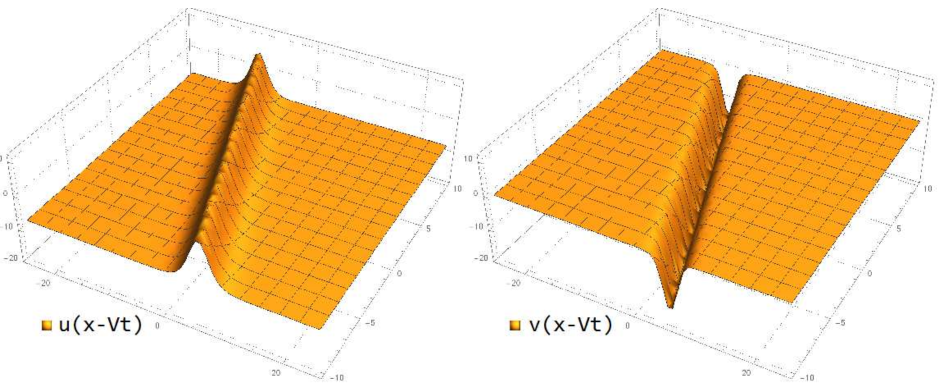


FIG. 4. Hyperbolic functions solution (20) when $A_1 = 1$, $A_2 = 1.2$, $\lambda = 2.2$, $\mu = 1$, $\beta_0 = -360$

– when $\lambda^2 - 4\mu = 0$, we get the family of rational functions solutions

$$\begin{cases} u(\xi) = & -\frac{60A_2^2}{(A_2\xi + A_1)^2}, \\ v(\xi) = & \frac{360\left(\mu^2(A_2\xi+A_1)^4 - A_2^4\right)}{(A_2\xi+A_1)^4} + \beta_0, \end{cases} \tag{21}$$

where $\xi = x$, A_1, A_2, β_0 are arbitrary constants; note that solutions (21) coincide with corresponding family (18) from the first set.

– when $\lambda^2 - 4\mu < 0$, we get the family of trigonometric functions solutions

$$\begin{cases} u(\xi) = & \frac{-15\sigma\left(A_1^2+A_2^2\right)}{\left(A_1 \sin\frac{\xi\sqrt{\sigma}}{2} + A_2 \cos\frac{\xi\sqrt{\sigma}}{2}\right)^2} + 9\sigma, \\ v(\xi) = & -\frac{45\left(A_1^2+A_2^2\right)^2\sigma^2}{2\left(A_1 \sin\frac{\xi\sqrt{\sigma}}{2} + A_2 \cos\frac{\xi\sqrt{\sigma}}{2}\right)^4} + \beta_0 + 360\mu^2, \end{cases} \tag{22}$$

where $\xi = x + \left(\lambda^2 - 4\mu\right)t$, $\sigma = 4\mu - \lambda^2$, A_1, A_2, β_0 are arbitrary constants.

Solutions set 3. Constants set (15) yields two families of solutions:

– when $\lambda > 0$, we get the family of hyperbolic functions solutions

$$\begin{cases} u(\xi) = & \frac{15\lambda^2\left(A_2^2 - A_1^2\right)}{\left(A_1 \sinh\frac{\xi|\lambda|}{2} + A_2 \cosh\frac{\lambda\xi}{2}\right)^2} - \lambda^2, \\ v(\xi) = & \frac{2A_1A_2\left(A_2^2-A_1^2\right)\left(\beta_0+30\lambda^4\right)\sinh\xi|\lambda| + A_1A_2\left(A_1^2+A_2^2\right)\beta_0 \sinh 2\xi|\lambda|}{2\left(A_1 \sinh\frac{\xi|\lambda|}{2} + A_2 \cosh\frac{\lambda\xi}{2}\right)^4} + \\ & + \frac{-\left(A_1^4-A_2^4\right)\left(\beta_0+30\lambda^4\right)\cosh\lambda\xi - \frac{3}{4}\left(A_1^2-A_2^2\right)^2\left(20\lambda^4-\beta_0\right)}{2\left(A_1 \sinh\frac{\xi|\lambda|}{2} + A_2 \cosh\frac{\lambda\xi}{2}\right)^4} + \\ & + \frac{\frac{1}{4}\left(A_1^4+6A_2^2A_1^2+A_2^4\right)\beta_0 \cosh 2\lambda\xi}{2\left(A_1 \sinh\frac{\xi|\lambda|}{2} + A_2 \cosh\frac{\lambda\xi}{2}\right)^4}, \end{cases} \tag{23}$$

where $\xi = x - \lambda^2 t$, A_1, A_2, β_0 are arbitrary constants;

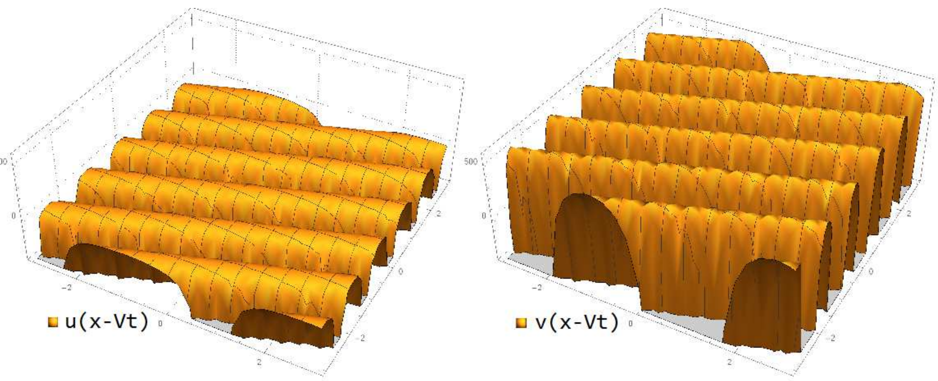


FIG. 5. Trigonometric functions solution (22) when $A_1 = 1$, $A_2 = 1.2$, $\lambda = 1$, $\mu = 1$, $\beta_0 = 360$

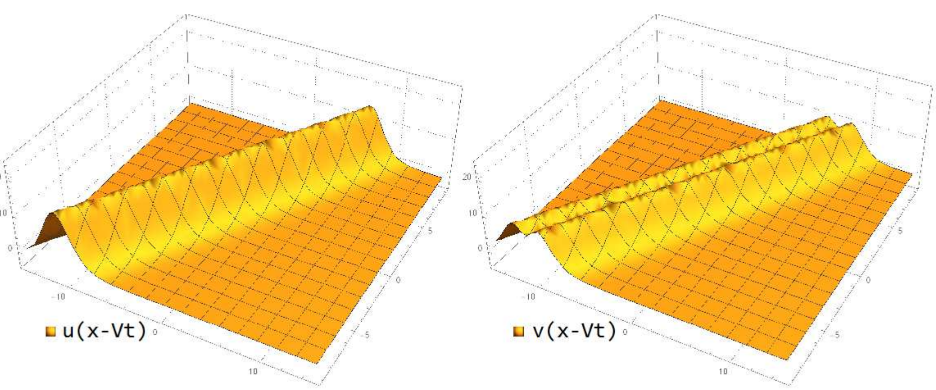


FIG. 6. Hyperbolic functions solution (23) when $A_1 = 1$, $A_2 = 1.2$, $\lambda = 1$, $\beta_0 = 0$

– when $\lambda = 0$, we get the family of rational functions solutions

$$\begin{cases} u(\xi) = & -\frac{60A_2^2}{(A_2\xi + A_1)^2}, \\ v(\xi) = & \beta_0 - \frac{360A_2^4}{(A_2\xi + A_1)^4}, \end{cases} \% \tag{24}$$

where $\xi = x$, A_1, A_2, β_0 are arbitrary constants; note that solutions (24) coincide with corresponding family (18) from the first set.

Solutions set 4. Constants set (16) yields two families of solutions:

– when $\lambda > 0$, we get the family of hyperbolic functions solutions

$$\begin{cases} u(\xi) = & -\frac{3\lambda^2\left(6A_2A_1\sinh\xi|\lambda| + A_1^2(3\cosh\lambda\xi + 7) + A_2^2(3\cosh\lambda\xi - 7)\right)}{2\left(A_1\sinh\frac{\xi|\lambda|}{2} + A_2\cosh\frac{\lambda\xi}{2}\right)^2}, \\ v(\xi) = & \beta_0 - \frac{45\left(A_1^2 - A_2^2\right)^2\lambda^4}{2\left(A_1\sinh\frac{\xi|\lambda|}{2} + A_2\cosh\frac{\lambda\xi}{2}\right)^4}, \end{cases} \tag{25}$$

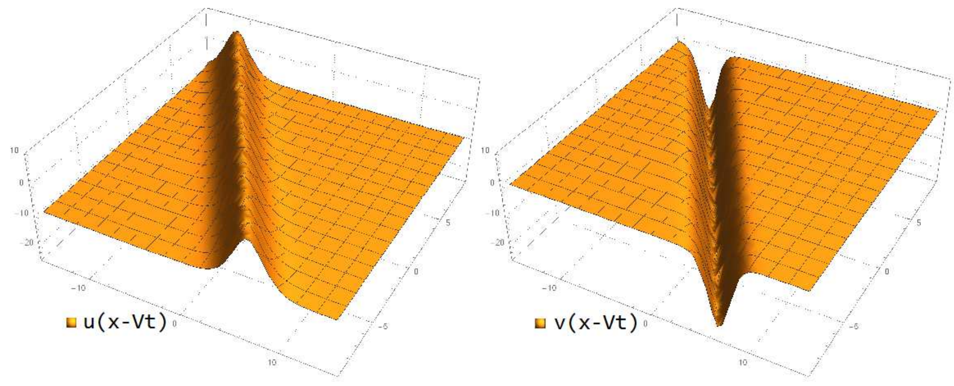


FIG. 7. Hyperbolic functions solution (25) when $A_1 = 1$, $A_2 = 1.2$, $\lambda = 1$, $\beta_0 = 0$

where $\xi = x + \lambda^2 t$, A_1, A_2, β_0 are arbitrary constants;

– when $\lambda = 0$, we get the family of rational functions solutions

$$\begin{cases} u(\xi) = & -\frac{60A_2^2}{(A_2\xi + A_1)^2}, \\ v(\xi) = & \beta_0 - \frac{360A_2^4}{(A_2\xi + A_1)^4}, \end{cases} \tag{26}$$

where $\xi = x$, A_1, A_2, β_0 are arbitrary constants; note that solutions (26) coincide with corresponding family (18) from the first set.

5. Application: Example 2

Consider the following Korteweg – de Vries (KdV) type nonlinear dynamic system [1]

$$\begin{cases} u_t = u_{xxx} + uu_x - vv_x, \\ v_t = -2v_{xxx} - uv_x. \end{cases} \tag{27}$$

Let us solve system (27) by use of the (G'/G) – expansion method.

Step 1. Introducing traveling wave variable $\xi = x - Vt$, we reduce system (27) to a system of ODE for $u = u(\xi)$ and $v = v(\xi)$

$$\begin{cases} -Vu' = u''' + uu' - vv', \\ -Vv' = -2v''' - uv'. \end{cases} \tag{28}$$

Suppose that the solution to system (28) can be expressed by polynomials in (G'/G) as follows:

$$u(\xi) = \sum_{i=0}^{m} \alpha_i \left(\frac{G'}{G}\right)^i, \quad v(\xi) = \sum_{i=0}^{n} \beta_i \left(\frac{G'}{G}\right)^i. \tag{29}$$

Considering the homogeneous balance between u''' and vv', v''' and uv' in the first and the second equations of system (28) correspondingly, we obtain a simple system of algebraic equations

$$\begin{cases} m+3=2n+1. \\ n+3=m+n+1, \end{cases} \tag{30}$$

from which it can be easily found that $m=2$ and $n=2$.

Step 2. Considering (29) and (30), we find the solution to system (28) in the following form:

$$\begin{cases} u(\xi)=\alpha_2\left(\frac{G'}{G}\right)^2+\alpha_1\left(\frac{G'}{G}\right)+\alpha_0, \\ v(\xi)=\beta_2\left(\frac{G'}{G}\right)^2+\beta_1\left(\frac{G'}{G}\right)+\beta_0, \end{cases} \tag{31}$$

where function $G=G(\xi)$ satisfies the second order linear ODE (5), λ, μ, V, α_i $\left(i=\overline{0,2}\right)$, β_j $\left(j=\overline{0,2}\right)$ are all constants to be determined later, $\alpha_2\neq 0$, $\beta_2\neq 0$.

Step 3. Substituting (31) into system (28) and collecting all terms with the same power of $\left(\frac{G'}{G}\right)$ together, the left-hand sides of equations (28) are converted into another polynomials in $\left(\frac{G'}{G}\right)$. Equating each coefficient of these polynomials to zero yields a set of simultaneous algebraic equations for $\lambda,\mu,V,\alpha_i\left(i=\overline{0,2}\right),\beta_j\left(j=\overline{0,2}\right)$ as follows:

– from *the first equation* in (28):

$$\begin{aligned}
0:\quad & \alpha_1\lambda^2\mu+6\alpha_2\lambda\mu^2+2\alpha_1\mu^2+\alpha_0\alpha_1\mu-\beta_0\beta_1\mu+\alpha_1\mu V=0 \\
1:\quad & \alpha_1\lambda^3+6\alpha_2\lambda^2\mu+8\alpha_2\mu\left(\lambda^2+2\mu\right)+8\alpha_1\lambda\mu+\alpha_0\left(\alpha_1\lambda+2\alpha_2\mu\right)+ \\
& +\alpha_1^2\mu-\beta_0\left(\beta_1\lambda+2\beta_2\mu\right)-\beta_1^2\mu+V\left(\alpha_1\lambda+2\alpha_2\mu\right)=0 \\
2:\quad & 8\alpha_2\lambda\left(\lambda^2+2\mu\right)+7\alpha_1\lambda^2+36\alpha_2\lambda\mu+\alpha_1\left(\alpha_1\lambda+2\alpha_2\mu\right)+ \\
& +\alpha_0\left(2\alpha_2\lambda+\alpha_1\right)+8\alpha_1\mu+\alpha_1\alpha_2\mu-\beta_1\left(\beta_1\lambda+2\beta_2\mu\right)- \\
& -\beta_0\left(2\beta_2\lambda+\beta_1\right)-\beta_1\beta_2\mu+V\left(2\alpha_2\lambda+\alpha_1\right)=0 \\
3:\quad & 8\alpha_2\left(\lambda^2+2\mu\right)+30\alpha_2\lambda^2+\alpha_2\left(\alpha_1\lambda+2\alpha_2\mu\right)+12\alpha_1\lambda+ \\
& +\alpha_1\left(2\alpha_2\lambda+\alpha_1\right)+24\alpha_2\mu+2\alpha_0\alpha_2-\beta_2\left(\beta_1\lambda+2\beta_2\mu\right)- \\
& -\beta_1\left(2\beta_2\lambda+\beta_1\right)-2\beta_0\beta_2+2\alpha_2V=0 \\
4:\quad & 54\alpha_2\lambda+\alpha_2\left(2\alpha_2\lambda+\alpha_1\right)+2\alpha_2\alpha_1+6\alpha_1-\beta_2\left(2\beta_2\lambda+\beta_1\right)- \\
& -2\beta_1\beta_2=0 \\
5:\quad & 2\alpha_2^2+24\alpha_2-2\beta_2^2=0;
\end{aligned}$$

– from *the second equation* in (28):

$$\begin{aligned}
0:\quad & -\alpha_0\beta_1\mu-2\beta_1\lambda^2\mu-12\beta_2\lambda\mu^2-4\beta_1\mu^2+\beta_1\mu V=0 \\
1:\quad & -2\left(\beta_1\lambda^3+6\beta_2\lambda^2\mu+8\beta_2\mu\left(\lambda^2+2\mu\right)+8\beta_1\lambda\mu\right)+ \\
& +V\left(\beta_1\lambda+2\beta_2\mu\right)-\alpha_0\left(\beta_1\lambda+2\beta_2\mu\right)+\alpha_1\beta_1(-\mu)=0 \\
2:\quad & -\alpha_1\left(\beta_1\lambda+2\beta_2\mu\right)-\alpha_0\left(2\beta_2\lambda+\beta_1\right)+\alpha_2\beta_1(-\mu)- \\
& -2\left(8\beta_2\lambda\left(\lambda^2+2\mu\right)+7\beta_1\lambda^2+36\beta_2\lambda\mu+8\beta_1\mu\right)+ \\
& +V\left(2\beta_2\lambda+\beta_1\right)=0 \\
3:\quad & -\alpha_2\left(\beta_1\lambda+2\beta_2\mu\right)-\alpha_1\left(2\beta_2\lambda+\beta_1\right)-2\alpha_0\beta_2- \\
& -2\left(8\beta_2\left(\lambda^2+2\mu\right)+30\beta_2\lambda^2+12\beta_1\lambda+24\beta_2\mu\right)+2\beta_2V=0 \\
4:\quad & -\alpha_2\left(2\beta_2\lambda+\beta_1\right)-2\alpha_1\beta_2-2\left(54\beta_2\lambda+6\beta_1\right)=0 \\
5:\quad & -2\alpha_2\beta_2-48\beta_2=0.
\end{aligned}$$

In addition to this, the highest order coefficients in (31) are supposed to be nonzero:

$$\alpha_2 \neq 0,\ \beta_2 \neq 0. \tag{32}$$

Step 4. Solving the system of algebraic equations from the previous step with conditions (32) with the aid of MATHEMATICA yields *four sets of solutions*:

- Set 1.

$$\begin{gathered}\alpha_0 = -2\lambda^2 - 16\mu + V, \quad \alpha_1 = -24\lambda, \quad \alpha_2 = -24,\\ \beta_0 = \sqrt{2}\left(-\lambda^2 - 8\mu + 2V\right), \quad \beta_1 = -12\sqrt{2}\lambda, \quad \beta_2 = -12\sqrt{2},\end{gathered} \tag{33}$$

  where λ, μ and V are arbitrary constants.
- Set 2.

$$\begin{gathered}\alpha_0 = -2\lambda^2 - 16\mu + V, \quad \alpha_1 = -24\lambda, \quad \alpha_2 = -24,\\ \beta_0 = \sqrt{2}\left(\lambda^2 + 8\mu - 2V\right), \quad \beta_1 = 12\sqrt{2}\lambda, \quad \beta_2 = 12\sqrt{2},\end{gathered} \tag{34}$$

  where λ, μ and V are arbitrary constants.
- Set 3.

$$\begin{gathered}\mu = 0, \quad \alpha_0 = V - 2\lambda^2, \quad \alpha_1 = -24\lambda, \quad \alpha_2 = -24,\\ \beta_0 = 2\sqrt{2}V - \sqrt{2}\lambda^2, \quad \beta_1 = -12\sqrt{2}\lambda, \quad \beta_2 = -12\sqrt{2},\end{gathered} \tag{35}$$

  where λ and V are arbitrary constants.
- Set 4.

$$\begin{gathered}\mu = 0, \quad \alpha_0 = V - 2\lambda^2, \quad \alpha_1 = -24\lambda, \quad \alpha_2 = -24,\\ \beta_0 = \sqrt{2}\lambda^2 - 2\sqrt{2}V, \quad \beta_1 = 12\sqrt{2}\lambda, \quad \beta_2 = 12\sqrt{2}\end{gathered} \tag{36}$$

  where λ and V are arbitrary constants.

Finally, substituting solutions (33)–(36) with the general solution to linear ODE (5) into representation (31) we obtain *four separate sets* of traveling wave solutions to the KdV type dynamic system (27) as follows.

Solutions set 1. Constants set (33) yields three families of solutions:

- when $\lambda^2 - 4\mu > 0$, we get the family of hyperbolic functions solutions

$$\begin{cases} u(\xi) = & -\frac{6\left(A_1^2 - A_2^2\right)\sigma}{\left(A_1 \sinh\frac{\xi\sqrt{\sigma}}{2} + A_2 \cosh\frac{\xi\sqrt{\sigma}}{2}\right)^2} - 2\lambda^2 + 8\mu + V,\\ v(\xi) = & -\frac{3\sqrt{2}\left(A_1^2 - A_2^2\right)\sigma}{\left(A_1 \sinh\frac{\xi\sqrt{\sigma}}{2} + A_2 \cosh\frac{\xi\sqrt{\sigma}}{2}\right)^2} - \sqrt{2}\sigma + 2\sqrt{2}V,\end{cases} \tag{37}$$

  where $\xi = x - Vt$, $\sigma = \lambda^2 - 4\mu$, A_1, A_2, V are arbitrary constants;
- when $\lambda^2 - 4\mu = 0$, we get the family of rational functions solutions

$$\begin{cases} u(\xi) = & \frac{A_2^2\left(\xi^2 V - 24\right) + 2A_2 A_1 \xi V + A_1^2 V}{(A_2\xi + A_1)^2},\\ v(\xi) = & 2\sqrt{2}\left(V - \frac{6A_2^2}{(A_2\xi + A_1)^2}\right),\end{cases} \tag{38}$$

  where $\xi = x - Vt$, A_1, A_2, V are arbitrary constants;

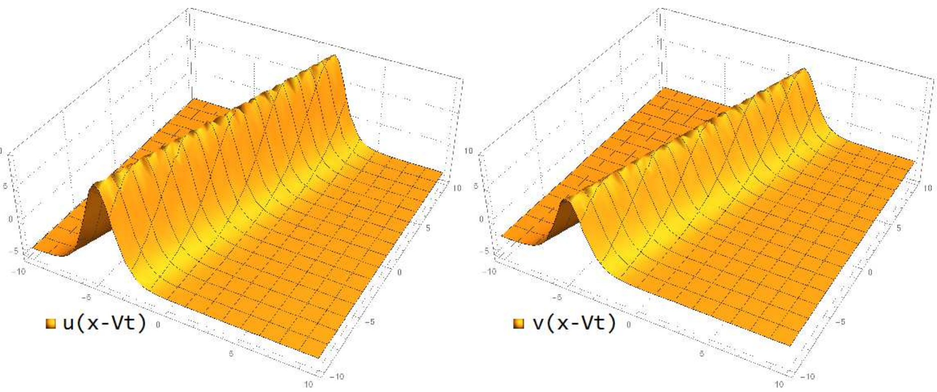


FIG. 8. Hyperbolic functions solution (37) when $A_1 = 1$, $A_2 = 1.2,\ \lambda = 2.5,\ \mu = 1,\ V = 0.3$

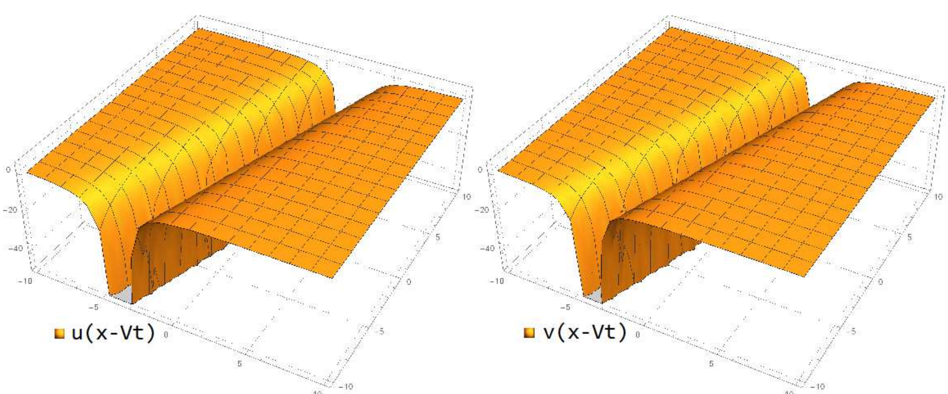


FIG. 9. Rational functions solution (38) when $A_1 = 1,\ A_2 = 1.2,\ \lambda = 2,\ \mu = 1,\ V = 0.3$

– when $\lambda^2 - 4\mu < 0$, we get the family of trigonometric functions solutions

$$\begin{cases} u(\xi) = & -\frac{6\sigma\left(A_1^2+A_2^2\right)}{\left(A_1 \sin\frac{\xi\sqrt{\sigma}}{2}+A_2\cos\frac{\xi\sqrt{\sigma}}{2}\right)^2} - 2\lambda^2 + 8\mu + V, \\ v(\xi) = & -\frac{3\sqrt{2}\left(A_1^2+A_2^2\right)\sigma}{\left(A_1 \sin\frac{\xi\sqrt{\sigma}}{2}+A_2\cos\frac{\xi\sqrt{\sigma}}{2}\right)^2} + \sqrt{2}\sigma + 2\sqrt{2}V, \end{cases} \tag{39}$$

where $\xi = x - Vt,\ \sigma = 4\mu - \lambda^2,\ \ A_1, A_2, V$ are arbitrary constants.

Solutions set 2. Constants set (34) yields three families of solutions:

– when $\lambda^2 - 4\mu > 0$, we get the family of hyperbolic functions solutions

$$\begin{cases} u(\xi) = & -\frac{6\left(A_1^2-A_2^2\right)\sigma}{\left(A_1 \sinh\frac{\xi\sqrt{\sigma}}{2}+A_2\cosh\frac{\xi\sqrt{\sigma}}{2}\right)^2} - 2\lambda^2 + 8\mu + V, \\ v(\xi) = & \frac{3\sqrt{2}\left(A_1^2-A_2^2\right)\sigma}{\left(A_1 \sinh\frac{\xi\sqrt{\sigma}}{2}+A_2\cosh\frac{\xi\sqrt{\sigma}}{2}\right)^2} + \sqrt{2}\sigma - 2\sqrt{2}V, \end{cases} \tag{40}$$

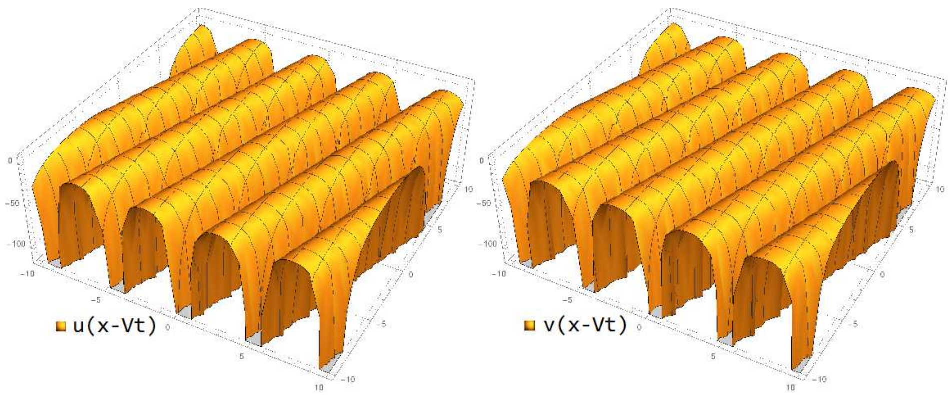


FIG. 10. Trigonometric functions solution (39) when $A_1 = 1$, $A_2 = 1.2$, $\lambda = 1.5$, $\mu = 1$, $V = 0.3$

where $\xi = x - Vt$, $\sigma = \lambda^2 - 4\mu$, A_1, A_2, V are arbitrary constants; in particular, setting $A_1 = 0$, $\sigma = 4k_1^2$, $V = a_{10} - 16k_1^2$, we obtain exactly the soliton solution, found by means of the tanh – method in [14].

– when $\lambda^2 - 4\mu = 0$, we get the family of rational functions solutions

$$\begin{cases} u(\xi) = & \frac{A_2^2(\xi^2 V - 24) + 2A_2A_1\xi V + A_1^2 V}{(A_2\xi + A_1)^2}, \\ v(\xi) = & 2\sqrt{2}\left(\frac{6A_2^2}{(A_2\xi + A_1)^2} - V\right), \end{cases} \tag{41}$$

where $\xi = x - Vt$, A_1, A_2, V are arbitrary constants;

– when $\lambda^2 - 4\mu < 0$, we get the family of trigonometric functions solutions

$$\begin{cases} u(\xi) = & -\frac{6(A_1^2 + A_2^2)\sigma}{\left(A_1 \sin\frac{\xi\sqrt{\sigma}}{2} + A_2 \cos\frac{\xi\sqrt{\sigma}}{2}\right)^2} - 2\lambda^2 + 8\mu + V, \\ v(\xi) = & \frac{3\sqrt{2}(A_1^2 + A_2^2)\sigma}{\left(A_1 \sin\frac{\xi\sqrt{\sigma}}{2} + A_2 \cos\frac{\xi\sqrt{\sigma}}{2}\right)^2} - \sqrt{2}\sigma - 2\sqrt{2}V, \end{cases} \tag{42}$$

where $\xi = x - Vt$, $\sigma = 4\mu - \lambda^2$, A_1, A_2, V are arbitrary constants.

Solutions set 3. Constants set (35) yields two families of solutions:

– when $\lambda > 0$, we get the family of hyperbolic functions solutions

$$\begin{cases} u(\xi) = & \frac{(V - 2\lambda^2)(2A_1A_2 \sinh \xi|\lambda| + (A_1^2 + A_2^2)\cosh \lambda\xi) - (A_1^2 - A_2^2)(10\lambda^2 + V)}{2\left(A_1 \sinh\frac{\xi|\lambda|}{2} + A_2 \cosh\frac{\lambda\xi}{2}\right)^2}, \\ v(\xi) = & \frac{(2V - \lambda^2)(2A_1A_2 \sinh \xi|\lambda| + (A_1^2 + A_2^2)\cosh \lambda\xi) - (A_1^2 - A_2^2)(5\lambda^2 + 2V)}{\sqrt{2}\left(A_1 \sinh\frac{\xi|\lambda|}{2} + A_2 \cosh\frac{\lambda\xi}{2}\right)^2}, \end{cases} \tag{43}$$

where $\xi = x - Vt$, A_1, A_2, V are arbitrary constants;

– when $\lambda = 0$, we get the family of rational functions solutions

$$\begin{cases} u(\xi) = & \frac{A_2^2(\xi^2 V - 24) + 2A_2A_1\xi V + A_1^2 V}{(A_2\xi + A_1)^2}, \\ v(\xi) = & 2\sqrt{2}\left(V - \frac{6A_2^2}{(A_2\xi + A_1)^2}\right), \end{cases} \tag{44}$$

where $\xi = x - Vt$, A_1, A_2, V are arbitrary constants.

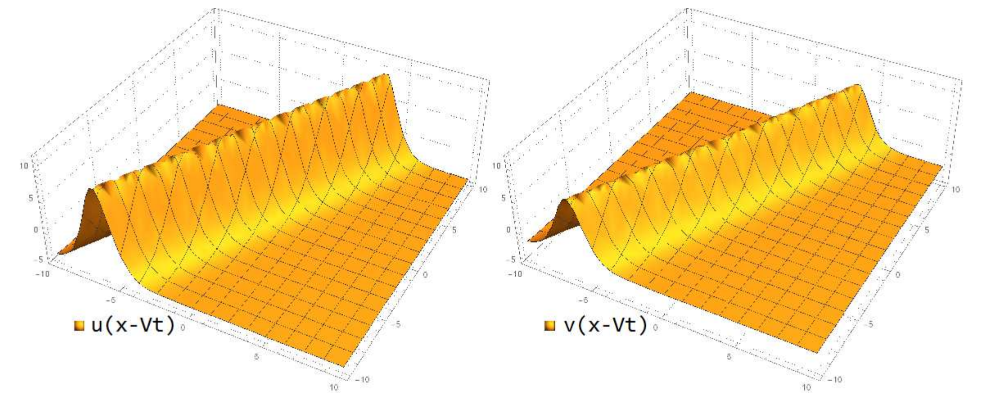


FIG. 11. Hyperbolic functions solution (43) when $A_1 = 1$, $A_2 = 1.2$, $\lambda = 1.5$, $V = 0.5$

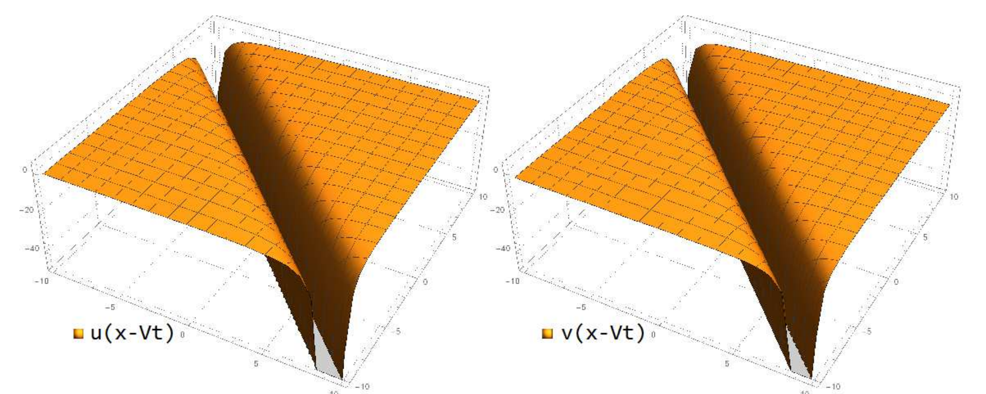


FIG. 12. Rational functions solution (44) when $A_1 = 1$, $A_2 = 1.2$, $\lambda = 0$, $V = -1$

Solutions set 4. Constants set (36) yields two families of solutions:

– when $\lambda > 0$, we get the family of hyperbolic functions solutions

$$\begin{cases} u(\xi) = & \dfrac{\left(V-2\lambda^2\right)\left(2A_1A_2\sinh\xi|\lambda|+\left(A_1^2+A_2^2\right)\cosh\lambda\xi\right)-\left(A_1^2-A_2^2\right)\left(10\lambda^2+V\right)}{2\left(A_1\sinh\frac{\xi|\lambda|}{2}+A_2\cosh\frac{\lambda\xi}{2}\right)^2}, \\ v(\xi) = & \dfrac{\left(\lambda^2-2V\right)\left(2A_1A_2\sinh\xi|\lambda|+\left(A_1^2+A_2^2\right)\cosh\lambda\xi\right)+\left(A_1^2-A_2^2\right)\left(5\lambda^2+2V\right)}{\sqrt{2}\left(A_1\sinh\frac{\xi|\lambda|}{2}+A_2\cosh\frac{\lambda\xi}{2}\right)^2}, \end{cases} \tag{45}$$

where $\xi = x - Vt$, A_1, A_2, V are arbitrary constants;

– when $\lambda = 0$, we get the family of rational functions solutions

$$\begin{cases} u(\xi) = & \dfrac{A_2^2\left(\xi^2V-24\right)+2A_2A_1\xi V+A_1^2V}{(A_2\xi+A_1)^2}, \\ v(\xi) = & 2\sqrt{2}\left(\dfrac{6A_2^2}{(A_2\xi+A_1)^2}-V\right), \end{cases} \tag{46}$$

where $\xi = x - Vt$, A_1, A_2, V are arbitrary constants.

6. Conclusion

The (G'/G) – expansion method was successfully used to derive exact traveling wave solutions to two KdV type nonlinear dynamic systems [1, 3].

The method was implemented in computer system MATHEMATICA, with the aid of which we obtained the solutions in the form of hyperbolic, rational and trigonometric functions for both systems. Moreover, it is shown that with a certain choice of arbitrary parameters in both systems it is possible to rediscover the soliton solutions, found by means of the tanh – method in [14], and hence the solutions obtained in the present paper are of more general forms.

The correctness of the obtained results was assured by putting them back into the original systems with the aid of MATHEMATICA. Most of the obtained solutions were graphically analyzed.

The main advantage of the method is that it provides solutions with relatively many arbitrary parameters, and thus these solutions are often more general compared to other analytical methods. As it was shown in Section 3, there exist certain modifications of the method to provide solutions in more general form in comparison with the classical $\left(\frac{G'}{G}\right)$ – expansion method [15], therefore the authors plan to use them for further investigations.

Finally, the method is confirmed to be suitable for implementation in modern computer algebra systems.

Bibliography

1. Drinfeld V. G. Lie algebras and equations of Korteweg–de Vries type / V. G. Drinfeld, V. V. Sokolov // Itogi Nauki i Tekhniki. – Ser. Sovrem. Probl. Mat. Nov. Dostizh. – M.: VINITI, 1984. – № 15. – P. 81-180. (in Russin).
2. Guo S. The extended (G'/G) – expansion method and its applications to the Whitham–Broer–Kaup–Like equations and coupled Hirota–Satsuma KdV equations / S. Guo, Y. Zhou // Applied Mathematics and Computation. – 2010. – № 215. – P. 3214-3221.
3. Hentosh O. Ye. Differential-Geometric and Lie-Algebraic Foundations of Investigating Nonlinear Dynamical Systems on Functional Manifolds / O. Ye. Hentosh, M. M. Prytula, A. K. Prykarpatskyi. – Lviv: Publishing center of the Ivan Franko National University of Lviv, 2006. – 408 p. (in Ukrainian).
4. Jabbari A. A generalization of (G'/G) – expansion method and its application to nonlinear reaction–diffusion equations arising in mathematical biology / A. Jabbari, J. Manafian Heris, H. Kheiri // International Journal of Biomathematics. – 2014. – Vol. 7. – № 3. – 10 p.
5. Liu X. Application of Improved (G'/G) – Expansion Method to Traveling Wave Solutions of Two Nonlinear Evolution Equations / X. Liu, W. Zhang, Z. Li // Advances in Applied Mathematics and Mechanics. – 2012. – Vol. 4. – № 1. – P. 122-130.
6. Malfliet W. Solitary wave solutions of nonlinear wave equation / W. Malfliet // Am. J. Phys. – 1992. – Vol. 60. – P. 650-654.
7. Naher H. Algorithms and Some New Traveling Wave Solutions via New Extension of (G'/G) – expansion Method of Coupled Nonlinear Equations / H. Naher // British Journal of Mathematics and Computer Science. – 2015. – № 11. – P. 1-13.
8. Naher H. New approach of (G'/G) – expansion method and new approach of generalized (G'/G) – expansion method for nonlinear evolution equation / H. Naher, F. Abdullah // AIP ADVANCES. – 2013. – № 3. – 15 p.
9. Naher H. New approach of (G'/G) – expansion method and new approach of generalized (G'/G) – expansion method for ZKBBM equation / H. Naher // Journal of Egyptian Mathematical Society. – 2015 – Vol. 23, № 1. – C. 42-48.

10. Naher H. New Approach of (G'/G) – expansion Method for RLW Equation / H. Naher, F. Abdullah // Research Journal of Applied Sciences, Engineering and Technology. – 2014. – № 7. – P. 4864-4871.
11. Nur Alam M. Application of New Generalized (G'/G) –Expansion Method to the (3+1)-Dimensional Kadomtsev-Petviashvili Equation / M. Nur Alam, M. Ali Akbar, M. Hafez, F. Belgacem // Italian journal of pure and applied mathematics. – 2016. – № 36. – P. 415-428.
12. Nur Alam M. Application of the Novel (G'/G) – Expansion Method to the Regularized Long Wave Equation / M. Nur Alam, F. Belgacem // Waves Wavelets Fractals Adv. Anal. – 2015. – № 1. – P. 51-67.
13. Shehata A. The traveling wave solutions of the perturbed nonlinear Schrodinger equation and the cubic–quintic Ginzburg Landau equation using the modified (G'/G) – expansion method / A. R. Shehata // Applied Mathem. and Computation. – 2010. – № 217. – P. 1-10.
14. Starchak M. Soliton solutions of nonlinear dynamical systems of KdV–type equations / M. Starchak, M. Prytula // Visnyk of the Lviv University. – Series Applied Mathematics and Computer Science. – 2009. – Vol. 15. – P. 71-80. (in Ukrainian).
15. Wang M. The (G'/G) – expansion method and travelling wave solutions of nonlinear evolution equations in mathematical physics / M. Wang, X. Li, J. Zhang // Physics Letters. – 2008. – № 372. – P. 417-423.
16. Zayed E. Applications of an Extended (G'/G) – Expansion Method to Find Exact Solutions of Nonlinear PDEs in Mathematical Physics / E. Zayed, S. Al-Joudi // Mathematical Problems in Engineering. – 2010. – Vol. 2010. – 19 p.
17. Zayed E. Applications of an Improved (G'/G) – expansion Method to Nonlinear PDEs in Mathematical Physics / E. Zayed, S. Al-Joudi // Numerical Analysis and Applied Mathematics, International Conference. – 2009. – Vol. 1. – P. 371-376.
18. Zhang J. A generalized (G'/G) – expansion method and its applications / J. Zhang, X. Wei, Y. Lu // Physics Letters A. – 2008. – № 372. – P. 3653-3658.
19. Zhang S. A generalized (G'/G) – expansion method and its application to the (2+1)-dimensional Broer–Kaup equations / S. Zhang, W. Wang, J. Tong // Applied Mathematics and Computation. – 2009. – № 209. – P. 399-404.
20. Zhang S. A generalized (G'/G) – expansion method for the mKdV equation with variable coefficients / S. Zhang, J. Tong, W. Wang // Physics Letters A. – 2008. – № 372. – P. 2254-2257.
21. Zhou Y. Application of Modified (G'/G) – Expansion Method to Traveling Wave Solutions for Whitham Broer Kaup-Like Equations / Y. Zhou, C. Li // Commun. Theor. Phys. Beijing, China. – 2009. – Vol. 51. – C. 664-670.

Ivan Mykhailiuk, Mykola Prytula,
Faculty of Applied Mathematics and Informatics,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine.

Received 05.09.2017; revised 27.10.2017

Журнал обчислювальної та прикладної математики | 2017 № 3 (126) | Journal of Numerical & Applied Mathematics

UDC 519.6

DETERMINATION THE QUANTITY OF EIGENVALUE FOR TWO-PARAMETER EIGENVALUE PROBLEMS IN THE PRESCRIBED REGION

B. M. Podlevskyi

Резюме. Запропоновано алгоритм знаходження кількості власних значень двопараметричних спектральних задач у деякій заданій області. В основі алгоритму лежить принцип аргумента аналітичної функції однієї змінної. Наведено чисельні результати для нелінійної двопараметричної задачі на власні значення.

Abstract. An algorithm for finding the number of eigenvalues of two-parameter spectral problems in a given region is proposed. At the heart of the algorithm lies the principle of the argument of the analytic function of one variable. Numerical results for a nonlinear two-parameter eigenvalue problems are given.

1. Introduction

The multiparameter eigenvalue problems $T(\lambda)x = 0$ with operator-valued functions $T(\lambda): R^m \to L(H)$ ($L(H)$ – the set of linear bounded operators operating in a finite-dimensional Hilbert space$H$), which depends on several spectral parameters λ, have a classical analysis of their source. In particular, they arise in solving boundary value problems for differential equations with partial derivatives by separating the variables.

In abstract formulation, they are written in the form of a system of equations

$$T(\lambda)u \equiv \left(A_k - \sum_{i=1}^{m} \lambda_i B_{ki}\right) u_k = 0, \quad k = 1, 2, ..., m, \tag{1}$$

if the operator-function $T(\lambda)$ linearly depends on the spectral parameters $\lambda_i \in R$, $i = 1, 2, \ldots, m$, $A, B_i, A_k, B_{ki} \in L(H)$, $k, i = 1, 2, \ldots, m$.

An algebraic two-parameter eigenvalue problem as a partial case of a spectral problem (1) is written in the form of a system of two homogeneous linear equations

$$\begin{aligned} T_1(\lambda,\mu) &\equiv (A_1 + \lambda B_1 + \mu C_1)x = 0, \\ T_2(\lambda,\mu) &\equiv (A_2 + \lambda B_2 + \mu C_2)y = 0, \end{aligned} \tag{2}$$

where $A_i,\ B_i,\ C_i$ are the square matrices of the nth order. We will define our eigenvalue sets (in our case that are eigen pairs $(\lambda,\ \mu)$) such that the system (2) has non-trivial solutions $x \neq 0$ and $y \neq 0$.

Key words. Two-parameter eigenvalue problem, number of eigenvalues, principle of argument.

It is obvious that own pairs are solutions of the system of two nonlinear algebraic equations

$$\begin{aligned} f(\lambda, \mu) &\equiv \det(A_1 + \lambda B_1 + \mu C_1) = 0. \\ g(\lambda, \mu) &\equiv \det(A_2 + \lambda B_2 + \mu C_2) = 0. \end{aligned} \tag{3}$$

In this work the problem of finding the number of real roots of the system (3), which are in a certain region of the change of spectral parameters $(\lambda,\ \mu)$, is considered.

2. Preliminaries

An algorithm for finding the number of zeros of an analytic function in a given region, as well as some approximations to each of them, which can then be specified using iterative methods, in particular by the Newton method or its two-way analogues (see, for example, [5, 9]), is based on the ratio, which implies, in particular, the principle of the argument of the analytic function (see, for example, [2]):

Integral $\frac{1}{2\pi i}\int\limits_{\Gamma} \varphi(\lambda)\frac{f'(\lambda)}{f(\lambda)}\,d\lambda$ *is equal to the difference between the sum of values that takes the function $\varphi(\lambda)$ in the zeros of the function $f(\lambda)$ lying in inside the domain G, bounded by the curve Γ and the sum of the values that takes the same function $\varphi(\lambda)$ in the poles of the function $f(\lambda)$ that lying in inside of Γ, that is,*

$$\frac{1}{2\pi i}\int\limits_{\Gamma} \varphi(\lambda)\frac{f'(\lambda)}{f(\lambda)}\,d\lambda = \sum_{j=1}^{m} \nu_j \varphi(\alpha_j) - \sum_{j=1}^{n} \mu_j \varphi(\beta_j). \tag{4}$$

Here $\varphi(\lambda)$ is an analytic function in the domain G; $f(\lambda)$ is analytic in G everywhere, except for the finite number of poles $\beta_j \in G$, $j = 1,\ 2,\ \dots,\ n$, and $f(\lambda) \neq 0$ in G everywhere except for the finite number of zeros $\alpha_j \in G$, $j = 1,\ 2,\ \dots,\ m$; ν_j and μ_j is the multiplicity of zero and the order of the pole, respectively.

In particular, if we take $\varphi(\lambda) \equiv 1$, then we get that

$$\frac{1}{2\pi i}\int\limits_{\Gamma} \frac{f'(\lambda)}{f(\lambda)}\,d\lambda = \sum_{j=1}^{m} \nu_j - \sum_{j=1}^{n} \mu_j, \tag{5}$$

that is, the integral is equal to the difference between the number of zeros and the poles of function $f(\lambda)$ lying inside of Γ, taking into account their multiplicities (the so-called principle of the argument).

If the analytic function $f(\lambda)$ does not have poles in G, then the principle of argument (5) allows us to determine the number of all its zeros that lie in the domain G. However, this does not allow you to localize each of them.

To locate the zeros we use again the relation (4). Taking now $\varphi(\lambda) = \lambda^k$, $k = 1, 2,\ \dots$, we get the following statement.

Suppose that the analytic function $f(\lambda)$ does not have poles in G, but has in G, taking into account the multiplicity, the m zeros $\lambda_1, \lambda_2, \dots, \lambda_m$ and has no zeros on the boundary Γ of the domain G, then the number m is determined in *accordance with the principle of the argument*

$$m = s_0 = \frac{1}{2\pi\, i}\int\limits_{\Gamma} \frac{f'(\lambda)}{f(\lambda)}\, d\lambda \tag{6}$$

and the relationship is true

$$\sum_{j=1}^{m} (\lambda_j)^k = s_k\,, \quad k = 1,\ \dots\ , m, \tag{7}$$

where

$$s_k = \frac{1}{2\pi\, i}\int\limits_{\Gamma} \lambda^k \frac{f'(\lambda)}{f(\lambda)}\, d\lambda, \qquad k = 1,\ 2,\ \dots \quad . \tag{8}$$

The right-hand side of (7) is nothing but symmetrical functions of the roots $\lambda_1, \lambda_2, \dots, \lambda_m$ inside of Γ, from which, in principle, roots can be found, for example:

If $m = 1$ than

$$\lambda_1 = \frac{1}{2\pi i}\int\limits_{\Gamma} \frac{\lambda f'(\lambda)}{f(\lambda)} d\lambda.$$

If $m = 2$

$$s_1 \equiv \lambda_1 + \lambda_2 = \frac{1}{2\pi i}\int\limits_{\Gamma} \frac{\lambda f'(\lambda)}{f(\lambda)} d\lambda$$

and

$$s_2 \equiv \lambda_1^2 + \lambda_2^2 = \frac{1}{2\pi i}\int\limits_{\Gamma} \frac{\lambda^2 f'(\lambda)}{f(\lambda)} d\lambda.$$

This will give us $\lambda_1\lambda_2 = \frac{1}{2}(\lambda_1 + \lambda_2)^2 - \frac{1}{2}(\lambda_1^2 + \lambda_2^2)$ and, consequently, we find λ_1 and λ_2 by solving a square equation. This procedure can be continued in an obvious way for $m = k$. Another approach, when the system (7) is solved directly, it was considered in the work [6, 7].

For the functions of one variable or one-parameter spectral problems, the principle of the argument (6) and the formulae of the principle of argument (7) and (8) have been repeatedly used for solving various problems (see, for example, [1, 3, 4, 6-8, 10]).

In this paper, based on the principle of the argument of the function of one variable, the algorithm for finding the number of eigenvalues of a two-parameter spectral problem in a given region of changing of the spectral parameters is proposed.

3. Number roots of a system of two real equations with two real variables

Let us consider a two-parameter spectral problem (2), whose eigenvalues λ, μ we will seek as the roots of the system of nonlinear equations (3), where the functions $f(\lambda, \mu)$ and $g(\lambda, \mu)$ are real functions of real variables.

For this purpose we will construct the function $u = f + ig$ and we will require that it be analytic and have no poles inside a certain region G. Then, as is known, the number m of roots $\nu = \lambda + i\mu$ of a function u in the region G, which is bounded by a curve Γ, that is, common solutions $(\lambda,\ \mu)$ of equations $f(\lambda,\mu) = 0$, $g(\lambda,\mu) = 0$, follows from the principele of argument of the analytic function (6), that is,

$$m = \frac{1}{2\pi i}\int\limits_{\Gamma}\frac{u'(\nu)}{u(\nu)}d\nu.$$

Taking into account that

$$\frac{1}{2\pi i}\int\limits_{\Gamma}\frac{u'(\nu)}{u(\nu)}d\nu = \frac{1}{2\pi i}\int\limits_{\Gamma} d\log u(\nu) = \frac{1}{2\pi}\int\limits_{\Gamma} d\phi,$$

where

$$\phi = \arg\log u(\nu) = \arctan\frac{g}{f} + \pi n, \tag{9}$$

we obtaine

$$m = \frac{1}{2\pi}\int\limits_{\Gamma} d(\arctan\frac{g}{f} + n\pi).$$

Consider the curve Γ with its parametric representation $\lambda = \lambda(t)$; $\mu = \mu(t)$; $0 \leqslant t \leqslant 1$. From (9) we have

$$d\phi = \frac{gdf - fdg}{f^2 + g^2}$$

If ϕ we replace the differentiation by t, we obtain

$$\frac{1}{2\pi}\int\limits_{\Gamma} d\phi = \frac{1}{2\pi}\int\limits_{\Gamma}\frac{d\phi}{dt}dt$$

Moreover, if we consider our expression $d\phi$ along the curve, we will have:

$$\frac{1}{2\pi}\int\limits_{\Gamma}\frac{d\phi}{dt}dt = \int\limits_{0}^{1}\frac{g\left(\frac{df}{d\lambda}\frac{d\lambda}{dt} + \frac{df}{d\mu}\frac{d\mu}{dt}\right) - f\left(\frac{dg}{d\lambda}\frac{d\lambda}{dt} + \frac{dg}{d\mu}\frac{d\mu}{dt}\right)}{f^2 + g^2}dt \tag{10}$$

Consequently, the number of eigenvalues m of the system of equations (3) is calculated by formula (10), in which the integral is replaced by some quadrature formula, for example, rectangles.

4. Numerical example

Let us consider a nonlinear two-parameter spectral problem

$$\begin{aligned} T_1(\lambda,\mu)\mathbf{x} &\equiv \begin{pmatrix} \lambda^2 - \mu^2 & 1 \\ 1 & 1 \end{pmatrix}\mathbf{x} = 0,\ \ \mathbf{x} \in R^2, \\ T_2(\lambda,\mu)\mathbf{y} &\equiv \begin{pmatrix} 2\lambda & 2 \\ 1 & 1 \end{pmatrix}\mathbf{y} = 0,\ \ \mathbf{y} \in R^2, \end{aligned} \tag{11}$$

and calculate the number of eigenvalues lying in different areas.

As was noted above, the eigenvalues of the problem (11) are solutions of the system of two nonlinear algebraic equations

$$\begin{aligned} f(\lambda,\mu) &= \det T_1(\lambda,\mu) = \lambda^2-\mu^2-1=0, \\ g(\lambda,\mu) &= \det T_2(\lambda,\mu) = 2\lambda\mu-2=0. \end{aligned} \tag{12}$$

It is easy to verify that the system (12) has two solutions:

$$(\lambda,\mu)_{1,2} = (\pm 1,272; 0,786).$$

The number of solutions m of the system (12) was calculated by the formula (10), in which the integral was replaced by the quadrature formula of rectangles, and the circle with center (λ^*,μ^*) and radius ρ^* was chosen as the boundary of Γ. The value of the functions (determinant) f and g and their derivatives on the boundary of the region (circle) were calculated on the basis of the LU-decomposition of the matrices $T_1(\lambda,\mu)$ and $T_2(\lambda,\mu)$ [6, 7].

Numerical calculations are carried out for different choices of the radius of circle and its center. The results are presented in Table 1. The first column of the table shows the coordinates of the center (λ^*,μ^*) of the circle, in the second column is its radius ρ^*, and in the third the number m of eigenvalues lying in that circle.

TABL. 1. Number eigenvalues of the problem (4.1)

(λ^*,μ^*)	ρ^*	m	(λ^*,μ^*)	ρ^*	m
(0.0, 0.0)	1.0	0	(0.0, 1.0)	2.0	2
(0.0, 1.0)	1.0	0	(0.0, 1.0)	2.0	1
(1.0, 1.0)	1.0	1	(-1.0, 0.0)	2.0	1

5. Conclusion

In this paper, based on the principle of the argument of the analytic function of one variable, an algorithm for finding the number of real eigenvalues of the system of two determinantal equations, that is, the real eigenvalues of a two-parameter spectral problem in a given region of changing of the spectral parameters, is proposed.

The numerical experiments performed for various problems have shown the effectiveness of the algorithm in the sense that for calculating the number of eigenvalues in a given region, there is no need for great accuracy in the calculation of the integral, and this does not require, in turn, a large partition of the integration boundary. This significantly reduces the calculation time, but, at the same time, it is sensitive to the choice of the boundary of the area. The algorithm ceases to work when the eigenvalues (though one) falls on the boundary that we preset. In this case, it is necessary to correct the boundary.

Bibliography

1. Abramov A. A. The Argument Principle in a Spectral Problem for Systems of Ordinary Differential Equations with Singularities / A. A. Abramov, V. I. Ul'yanova, L. F. Yukhno // Comput. Math. Math. Phys. – 1998. – Vol. 38. – P. 57-63.
2. Bitsadze A. V. Fundamentals of the theory of analytic functions of a complex variable. 3rd ed. / A. V. Bitsadze. – Moscow: Nauka, 1984. (in Rusian).
3. Delves L. M. A numerical method for locating the zeros of analytic function / L. M. Delves, J. N. Lyness // Math. Comput. – 1967. – Vol. 21. – P. 543-561.
4. Kartyschov S. V. Numerical method of solution of the eigenvalue problem with nonlinear entrance of the spectral parameter for sparse matrixes / S. V. Kartyschov // Zh. Vychisl. Mat. Mat. Fiz. – 1989. – Vol. 29. – P. 1898-1903. (in Russian).
5. Podlevs'kyi B. M. Bilateral analog of the Newton method for determination of eigenvalues of nonlinear spectral problems / B. M. Podlevs'kyi // J. Mathematical Sciences. – 2009. – Vol. 160, № 3. – P. 357-367.
6. Podlevskyi B. M. Bilateral methods for solving of nonlinear spectral problems / B. M. Podlevskyi. – Kyiv: Nauk dumka, 2014. (in Ukrainian).
7. Podlevskyi B. M. Calculating the exact derivatives of matrix determinant / B. M. Podlevskyi // Visnyk Lviv Univer. – Ser. Appl. Math. Inform. – 2013. – № 20. – P. 42-48. (in Ukrainian).
8. Podlevskyi B. M. Some computational aspects of algorithms for solving nonlinear two-parameter eigenvalue problems / B. M. Podlevskyi // J. Numer. Appl. Math. – 2010. – № 1 (100). – P. 100-110.
9. Podlevskii B. M. On Certain Two-Sided Analogues of Newton's Method for Solving Nonlinear Eigenvalue Problems / B. M. Podlevskii // Comput. Math. Math. Phys. – 2007. – Vol. 47, № 11. – P. 1745-1755.
10. Sakurai T. On Certain Two-Sided Analogues of Newton's Method for Solving Nonlinear Eigenvalue Problems / T. Sakurai, P. Kravanja, H. Sugiura, M. Van Barel // J. Comput. Appl. Math. – 2003. – Vol. 152. – P. 467-480.

Bohdan Podlevskyi,
Pidstrygach Institute for Applied Problems of Mechanics
and Mathematics, National Academy of Sciences of Ukraine,
3-b, Naukova Str., Lviv, 79060, Ukraine.

Received 11.10.2017; revised 02.11.2017

Журнал обчислювальної та прикладної математики | 2017 № 3 (126) | Journal of Numerical & Applied Mathematics

UDC 517.988:519.632

CONSTRUCTION OF TWO-SIDED APPROXIMATIONS TO POSITIVE SOLUTIONS OF BOUNDARY VALUE PROBLEMS FOR SEMILINEAR ELLIPTIC SYSTEMS

M. V. Sidorov

Резюме. Розглядається однорідна задача Діріхле для системи напівлінійних еліптичних рівнянь. Для побудови двобічних наближень до додатного розв'язку цієї системи використовуються методи теорії напівупорядкованих просторів, зокрема, результати В. І. Опойцева про розв'язність операторних рівнянь з гетеротонним оператором. Можливості і ефективність розробленого метода продемонстрована обчислювальним експериментом для системи Лане-Емдена.

Abstract. A homogeneous Dirichlet problem for a system of semilinear elliptic equations is considered. To construct two-sided approximations to a positive solution of this system, methods of the theory of semiordered spaces, in particular, the results of V.I. Opoĭcev on the solvability of operator equations with a heterotone operator are used. The possibilities and effectiveness of the developed method is demonstrated by a numerical experiment for the Lane-Emden system.

1. Introduction

Let us consider a homogeneous Dirichlet problem for a system of semilinear elliptic equations:

$$-\Delta u_i = f_i(\mathbf{x}, u_1, \ldots, u_n) \quad \text{in} \quad \Omega \subset \mathbb{R}^m, \tag{1}$$

$$u_i|_{\partial\Omega} = 0, \quad i = 1, 2, \ldots, n, \tag{2}$$

or in a vector form

$$-\Delta \mathbf{u} = \mathbf{f}(\mathbf{x}, \mathbf{u}) \quad \text{in} \quad \Omega \subset \mathbb{R}^m,$$

$$\mathbf{u}|_{\partial\Omega} = \boldsymbol{\theta},$$

where $\mathbf{x} = (x_1, \ldots, x_m)$, $\mathbf{u} = (u_1, \ldots, u_n)$, $-\Delta\mathbf{u} = (-\Delta u_1, \ldots, -\Delta u_n)$, $\mathbf{f} = (f_1, \ldots, f_n)$, $\boldsymbol{\theta} = (0, \ldots, 0)$, Δ is the Laplace operator,

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \ldots + \frac{\partial^2}{\partial x_m^2}.$$

Let us assume that $\Omega \subset \mathbb{R}^m$ is bounded domain with a piecewise smooth boundary $\partial\Omega$, functions $f_i(\mathbf{x}, \mathbf{u})$, $i = 1, 2, \ldots, n$, are non negative and continuous on the set of variables $\mathbf{x}$, $\mathbf{u}$, if $\mathbf{x} \in \Omega$, $u_i > 0$, $i = 1, 2, \ldots, n$.

The problem (1), (2) is a mathematical model of many stationary processes that are considered in chemical kinetics, biology, combustion theory etc. [12]. Many works [1,2,6,9,10,12,16, etc.] are devoted to the investigation of problem

Key words. Positive solution; semilinear elliptic systems; heterotone operator; two-sided approach; Lane-Emden system.

(1), (2). But the focus in these works was mainly on clarifying the conditions of existence and uniqueness of the positive solution of the problem or on the conditions of having a solution with radial symmetry for a case where Ω is a unit sphere, and an effective algorithm for numerical finding the solution was not proposed.

The purpose of this work is to develop the iterative methods for solving the boundary value problem (1), (2), which have a two-sided nature of convergence to the desired solution. Two-sided approximate methods of solving the nonlinear operator equations based on the theory of nonlinear operators in semiordered spaces were developed in [4, 5, 7, 8, 13, 14]. This work continues the research begun in [5] and distributes it to systems of nonlinear equations.

2. Some information from the theory of nonlinear operators in spaces with cones

Let us consider some concepts and facts from the theory of nonlinear operators in semiordered spaces that will be used further [7, 13, 14].

Let E be a real Banach space, and θ is a zero element of space E. A closed convex set $\mathcal{K} \subset E$ is called a cone, if from the fact that $x \in \mathcal{K}$, $x \neq \theta$, follows $\alpha x \in \mathcal{K}$ with $\alpha \geq 0$ and $-x \notin \mathcal{K}$.

Any cone $\mathcal{K} \subset E$ allows to enter in space E a semiordering by rule: $x \leqslant y$, if $y - x \in \mathcal{K}$. Elements $x \geqslant \theta$ (i.e. $x \in \mathcal{K}$) are called positive. The set of elements $\langle y, z \rangle$ of a semiordered space, which consists of those $x \in E$ for which $y \leqslant x \leqslant z$, is called a cone segment.

Normal cones are important class of cones for application of the theory of semiordered spaces in computational mathematics. A cone $\mathcal{K}$ is called normal if there exists a number $N(\mathcal{K}) > 0$, that from $\theta \leqslant x \leqslant y$ follows $\|x\| \leq N(\mathcal{K})\,\|y\|$. In this case, it is said that the norm is semimonotonic. If $N(\mathcal{K}) = 1$, then the cone is called acute and it is said that the norm is monotonous.

Let us consider the definitions of some classes of operators in spaces with cone.

The operator $T : E \to E$ is called positive if it leaves invariant the cone $\mathcal{K}$, i.e. $T(x) \in \mathcal{K}$ for anyone $x \in \mathcal{K}$.

The operator $T : E \to E$ is called heterotone (or mixed monotone [3, 13, etc.]), if it allows a diagonal representation $T(x) \equiv \hat{T}(x, x)$, where the companion operator $\hat{T} : E \times E \to E$ monotonically increases with respect to the first argument and decreases with respect to the second one, i.e.

a) if $y_1 \leqslant y_2$, then $\hat{T}(y_1, z) \leqslant \hat{T}(y_2, z)$ for all $z \in E$;

b) if $z_1 \leqslant z_2$, then $\hat{T}(y, z_1) \geqslant \hat{T}(y, z_2)$ for all $y \in E$.

A cone segment $\langle y_0, z_0 \rangle$ is called strongly invariant for a heterotone operator T, if

$$\hat{T}(y_0, z_0) \geq y_0, \quad \hat{T}(z_0, y_0) \leq z_0.$$

Let us fixate some nonzero element $u_0 \in \mathcal{K}$ and denote by $K(u_0)$ a set of such elements $x \in \mathcal{K}$, for which we can specify such $\alpha, \beta > 0$, that

$$\alpha u_0 \leqslant x \leqslant \beta u_0.$$

A positive heterotone operator T is called pseudoconcave, if $\hat{T}(y,z) \in K(u_0)$ for any $y, z \in \mathcal{K}$, $y \neq \theta$, $z \neq \theta$, and for any $v, w \in K(u_0)$ i $\tau \in (0;1)$

$$\hat{T}\left(\tau v, \frac{1}{\tau} w\right) \geqslant \tau \hat{T}(v,w),$$

and the sign of equality is impossible here.

A pseudoconcave operator T is called u_0-pseudoconcave, if for any $v, w \in K(u_0)$ and $\tau \in (0;1)$ you can find such $\eta(v,w,\tau) > 0$, that

$$\hat{T}\left(\tau v, \frac{1}{\tau} w\right) \geqslant \tau[1 + \eta(v,w,\tau)]\hat{T}(v,w).$$

Properties and the problem of constructing approximate solutions of operator equations with a heterotone operator have been considered in [3, 4, 11, 13, 14]. In particular, the following assertion holds [13, 14]: if the cone $\mathcal{K}$ is normal, the operator $\hat{T}$ is completely continuous, for T there is a strongly invariant cone segment $\langle y_0, z_0 \rangle$, and the system $\hat{T}(y,z) = y$, $\hat{T}(z,y) = z$ on $\langle y_0, z_0 \rangle$ has no solutions such that $y \neq z$, then the iterative process, which is formed by the rule $y_{n+1} = \hat{T}(y_n, z_n)$, $z_{n+1} = \hat{T}(z_n, y_n)$, $n = 0, 1, 2, \ldots$, starting from the point (y_0, z_0), two-sided converges to the unique on $\langle y_0, z_0 \rangle$ fixed point x^* of the operator T:

$$y_0 \leqslant y_1 \leqslant \ldots \leqslant y_n \leqslant \ldots \leqslant x^* \leqslant \ldots \leqslant z_n \leqslant \ldots \leqslant z_1 \leqslant z_0.$$

It is known [13, 14], that the system $\hat{T}(y,z) = y$, $\hat{T}(z,y) = z$ on $\langle y_0, z_0 \rangle$ has no solutions such that $y \neq z$, if T – u_0-pseudoconcave operator.

3. Construction of two-sided approximations

To analyze the problem (1), (2) and construct two-sided approximations to its positive solution, we will use the methods of the theory of nonlinear operators in semiordered spaces [7, 13, 14].

Let $\mathbf{C}^n(\bar{\Omega}) = \{\mathbf{u} = (u_1, \ldots, u_n) : u_i \in C(\bar{\Omega}),\ i = 1, \ldots, n\}$ be a Banach space of continuous in $\bar{\Omega} = \Omega \cup \partial\Omega$ vector-valued functions with a norm $\|\mathbf{u}\|_n = \max\{\|u_1\|, \ldots, \|u_n\|\}$, where $\|u_i\| = \max\limits_{\mathbf{x}\in\bar{\Omega}} |u_i(\mathbf{x})|$. Let us define in $\mathbf{C}^n(\bar{\Omega})$ a cone

$$\boldsymbol{\mathcal{K}}_+ = \{\mathbf{u} = (u_1, \ldots, u_n) \in \mathbf{C}^n(\bar{\Omega}) : u_i(\mathbf{x}) \geq 0,\ \mathbf{x} \in \bar{\Omega},\ i = 1, \ldots, n\}$$

of vector-valued functions with non negative coordinates. Notice that the cone $\boldsymbol{\mathcal{K}}_+$ in $\mathbf{C}^n(\bar{\Omega})$ is normal (and even acute) [7, 13, 14].

Using cone $\boldsymbol{\mathcal{K}}_+$ in space $\mathbf{C}^n(\bar{\Omega})$ we introduce a semiordering by the rule: for $\mathbf{u}, \mathbf{v} \in \mathbf{C}^n(\bar{\Omega})$ $\mathbf{u} \leqslant \mathbf{v}$, if $\mathbf{v} - \mathbf{u} \in \boldsymbol{\mathcal{K}}_+$, i.e.

$$\mathbf{u} \leqslant \mathbf{v}, \text{ if } u_i(\mathbf{x}) \leq v_i(\mathbf{x}) \text{ for all } \mathbf{x} \in \bar{\Omega} \text{ and for } i = 1, \ldots, n.$$

From the problem (1), (2) we go over to the system of integral equations of Hammerstein

$$u_i(\mathbf{x}) = \int\limits_{\Omega} G(\mathbf{x}, \boldsymbol{\xi}) f_i(\boldsymbol{\xi}, u_1(\boldsymbol{\xi}), \ldots, u_n(\boldsymbol{\xi})) d\boldsymbol{\xi}, \quad i = 1, \ldots, n, \tag{3}$$

or in a vector form

$$\mathbf{u}(\mathbf{x}) = \int\limits_{\Omega} G(\mathbf{x}, \boldsymbol{\xi})\mathbf{f}(\boldsymbol{\xi}, \mathbf{u}(\boldsymbol{\xi}))d\boldsymbol{\xi},$$

where $G(\mathbf{x}, \boldsymbol{\xi})$ is Green's function of the first boundary value problem for the operator $-\Delta$ in the domain Ω, $\mathbf{x} = (x_1, \ldots, x_m)$, $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_m)$.

The solution (generalized) of the problem (1), (2) will be called the vector-valued function $\mathbf{u}^* \in \mathbf{C}^n(\bar{\Omega})$, which is the solution of the system (3).

Let us introduce a nonlinear integral operator $\mathbf{T}$ acting in $\mathbf{C}^n(\bar{\Omega})$ by the rule defined by the right-hand side of the system of equations (3):

$$\begin{gathered}\mathbf{T}(\mathbf{u}) = \int\limits_{\Omega} G(\mathbf{x}, \boldsymbol{\xi})\mathbf{f}(\boldsymbol{\xi}, \mathbf{u}(\boldsymbol{\xi}))d\boldsymbol{\xi} = \\ = \left(\int\limits_{\Omega} G(\mathbf{x}, \boldsymbol{\xi}) f_1(\boldsymbol{\xi}, u_1(\boldsymbol{\xi}), \ldots, u_n(\boldsymbol{\xi}))d\boldsymbol{\xi}, \ldots, \right. \\ \left. \int\limits_{\Omega} G(\mathbf{x}, \boldsymbol{\xi}) f_n(\boldsymbol{\xi}, u_1(\boldsymbol{\xi}), \ldots, u_n(\boldsymbol{\xi}))d\boldsymbol{\xi}\right).\end{gathered} \tag{4}$$

Since $f_i(\mathbf{x}, u_1, \ldots, u_n) \geq 0$, if $\mathbf{x} \in \Omega$, $i = 1, \ldots, n$, and $G(\mathbf{x}, \boldsymbol{\xi}) \geq 0$, $\mathbf{x}, \boldsymbol{\xi} \in \Omega$, $\mathbf{x} \neq \boldsymbol{\xi}$, then the operator $\mathbf{T}$ is positive, that is, it leaves invariant a cone $\boldsymbol{\mathcal{K}}_+$: $\mathbf{T}(\boldsymbol{\mathcal{K}}_+) \subset \boldsymbol{\mathcal{K}}_+$.

Let us assume that the vector-valued function $\mathbf{f}(\mathbf{x}, \mathbf{u})$ allows a diagonal representation $\mathbf{f}(\mathbf{x}, \mathbf{u}) = \hat{\mathbf{f}}(\mathbf{x}, \mathbf{u}, \mathbf{u})$, where continuous on the set of variables $\mathbf{x}$, $\mathbf{v}$, $\mathbf{w}$ the functions $\hat{f}_i(\mathbf{x}, \mathbf{v}, \mathbf{w}) = \hat{f}_i(\mathbf{x}, v_1, \ldots, v_n, w_1, \ldots, w_n)$ monotonically increases with respect to all v_i and monotonically decreases with respect to all w_i, $i = 1, \ldots, n$, for all $\mathbf{x} \in \Omega$. Then the operator $\mathbf{T}$ of the form (4) will be heterotone with the companion operator

$$\begin{gathered}\hat{\mathbf{T}}(\mathbf{v}, \mathbf{w}) = \int\limits_{\Omega} G(\mathbf{x}, \boldsymbol{\xi})\hat{\mathbf{f}}(\boldsymbol{\xi}, \mathbf{v}(\boldsymbol{\xi}), \mathbf{w}(\boldsymbol{\xi}))d\boldsymbol{\xi} = \\ = \left(\int\limits_{\Omega} G(\mathbf{x}, \boldsymbol{\xi})\hat{f}_1(\boldsymbol{\xi}, v_1(\boldsymbol{\xi}), \ldots, v_n(\boldsymbol{\xi}), w_1(\boldsymbol{\xi}), \ldots, w_n(\boldsymbol{\xi}))d\boldsymbol{\xi}, \ldots, \right. \\ \left. \int\limits_{\Omega} G(\mathbf{x}, \boldsymbol{\xi})\hat{f}_n(\boldsymbol{\xi}, v_1(\boldsymbol{\xi}), \ldots, v_n(\boldsymbol{\xi}), w_1(\boldsymbol{\xi}), \ldots, w_n(\boldsymbol{\xi}))d\boldsymbol{\xi}\right).\end{gathered} \tag{5}$$

Operators $\mathbf{T}$ and $\hat{\mathbf{T}}$ are completely continuous [7,13,14].

In a cone $\boldsymbol{\mathcal{K}}_+$ we will define a strongly invariant cone segment $\langle \mathbf{v}^0, \mathbf{w}^0 \rangle$ by conditions

$$\hat{\mathbf{T}}(\mathbf{v}^0, \mathbf{w}^0) \geqslant \mathbf{v}^0, \quad \hat{\mathbf{T}}(\mathbf{w}^0, \mathbf{v}^0) \leqslant \mathbf{w}^0,$$

i.e.

$$\int\limits_{\Omega} G(\mathbf{x},\boldsymbol{\xi})\hat{f}_i(\boldsymbol{\xi}, v_1^0(\boldsymbol{\xi}),\ldots,v_n^0(\boldsymbol{\xi}), w_1^0(\boldsymbol{\xi}),\ldots,w_n^0(\boldsymbol{\xi}))d\boldsymbol{\xi} \geq v_i^0(\mathbf{x}) \text{ for all } \mathbf{x}\in\bar{\Omega},$$

$$\int\limits_{\Omega} G(\mathbf{x},\boldsymbol{\xi})\hat{f}_i(\boldsymbol{\xi}, w_1^0(\boldsymbol{\xi}),\ldots,w_n^0(\boldsymbol{\xi}), v_1^0(\boldsymbol{\xi}),\ldots,v_n^0(\boldsymbol{\xi}))d\boldsymbol{\xi} \leq w_i^0(\mathbf{x}) \text{ for all } \mathbf{x}\in\bar{\Omega},$$

$i = 1,\ldots,n.$

If the boundary $\partial\Omega$ of the domain Ω consists of a finite number of pieces of lines $\sigma_i(\mathbf{x}) = 0$, $i = 1, 2, \ldots, s$, where each $\sigma_i(\mathbf{x})$ is an elementary function, then using the R-functions method [15] one can construct in the form of a single analytic expression an elementary function $\omega(\mathbf{x})$ such that:

a) $\omega(\mathbf{x}) > 0$ in Ω;
b) $\omega(\mathbf{x}) = 0$ on $\partial\Omega$;
c) $|\nabla\omega(\mathbf{x})| \neq 0$ on $\partial\Omega$.

Then a strongly invariant cone segment can be searched in the form

$$\left\langle \mathbf{v}^0, \mathbf{w}^0 \right\rangle == \langle \alpha\omega(\mathbf{x}), \beta\omega(\mathbf{x})\rangle,$$

where $\alpha = (\alpha_1,\ldots,\alpha_n)$, $\beta = (\beta_1,\ldots,\beta_n)$, $0 \leq \alpha_i < \beta_i$, satisfy the system of inequalities

$$\int\limits_{\Omega} G(\mathbf{x},\boldsymbol{\xi})\hat{f}_i(\boldsymbol{\xi}, \alpha_1\omega(\boldsymbol{\xi}),\ldots,\alpha_n\omega(\boldsymbol{\xi}), \beta_1\omega(\boldsymbol{\xi}),\ldots,\beta_n\omega(\boldsymbol{\xi}))d\boldsymbol{\xi} \geq \alpha_i\omega(\mathbf{x})$$

for all $\mathbf{x}\in\bar{\Omega}$,

$$\int\limits_{\Omega} G(\mathbf{x},\boldsymbol{\xi})\hat{f}_i(\boldsymbol{\xi}, \beta_1\omega(\boldsymbol{\xi}),\ldots,\beta_n\omega(\boldsymbol{\xi}), \alpha_1\omega(\boldsymbol{\xi}),\ldots,\alpha_n\omega(\boldsymbol{\xi}))d\boldsymbol{\xi} \leq \beta_i\omega(\mathbf{x})$$

for all $\mathbf{x}\in\bar{\Omega}$, $\quad i = 1,\ldots,n.$

Let us create an iterative process according to the scheme

$$\mathbf{v}^{(k+1)} = \hat{\mathbf{T}}(\mathbf{v}^{(k)}, \mathbf{w}^{(k)}), \quad \mathbf{w}^{(k+1)} = \hat{\mathbf{T}}(\mathbf{w}^{(k)}, \mathbf{v}^{(k)}), \quad k = 0, 1, 2, \ldots,$$
$$\mathbf{v}^{(0)} = \mathbf{v}^0, \quad \mathbf{w}^{(0)} = \mathbf{w}^0,$$

i.e.

$$v_i^{(k+1)}(\mathbf{x}) = \int\limits_{\Omega} G(\mathbf{x},\boldsymbol{\xi})\hat{f}_i(\boldsymbol{\xi}, v_1^{(k)}(\boldsymbol{\xi}),\ldots,v_n^{(k)}(\boldsymbol{\xi}), w_1^{(k)}(\boldsymbol{\xi}),\ldots,w_n^{(k)}(\boldsymbol{\xi}))d\boldsymbol{\xi}, \tag{6}$$

$$w_i^{(k+1)}(\mathbf{x}) = \int\limits_{\Omega} G(\mathbf{x},\boldsymbol{\xi})\hat{f}_i(\boldsymbol{\xi}, w_1^{(k)}(\boldsymbol{\xi}),\ldots,w_n^{(k)}(\boldsymbol{\xi}), v_1^{(k)}(\boldsymbol{\xi}),\ldots,v_n^{(k)}(\boldsymbol{\xi}))d\boldsymbol{\xi}, \tag{7}$$

$k = 0, 1, 2, \ldots,$

$$v_i^{(0)}(\mathbf{x}) = v_i^0(\mathbf{x}), \quad w_i^{(0)}(\mathbf{x}) = w_i^0(\mathbf{x}), \quad i = 1,\ldots,n. \tag{8}$$

Given that the strong invariance of the constructed cone segment and heterotony of the operator $\mathbf{T}$, for which operator $\hat{\mathbf{T}}$ is an companion one, we can conclude that the sequence $\{\mathbf{v}^{(k)}(\mathbf{x})\}$ does not decrease behind the cone $\boldsymbol{\mathcal{K}}_+$, and the sequence $\{\mathbf{w}^{(k)}(\mathbf{x})\}$ does not increase behind the cone $\mathcal{K}_+$. In addition, from the normality of the cone $\mathcal{K}_+$ and completely continuity of the operator $\hat{\mathbf{T}}$ implies the existence of limits $\mathbf{v}^*(\mathbf{x})$ and $\mathbf{w}^*(\mathbf{x})$ of these sequences. Thus, the following inequalities hold:

$$\mathbf{v}^0 = \mathbf{v}^{(0)} \leqslant \mathbf{v}^{(1)} \leqslant \ldots \leqslant \mathbf{v}^{(k)} \leqslant \ldots \leqslant \mathbf{v}^* \leqslant \\ \leqslant \mathbf{w}^* \leqslant \ldots \leqslant \mathbf{w}^{(k)} \leqslant \ldots \leqslant \mathbf{w}^{(1)} \leqslant \mathbf{w}^{(0)} = \mathbf{w}^0.$$

The vector-valued functions $\mathbf{v}^* = (v_1^*, \ldots, v_n^*)$ and $\mathbf{w}^* = (w_1^*, \ldots, w_n^*)$ are a solution of the system of equations

$$\mathbf{v}^* = \hat{\mathbf{T}}(\mathbf{v}^*, \mathbf{w}^*), \quad \mathbf{w}^* = \hat{\mathbf{T}}(\mathbf{w}^*, \mathbf{v}^*),$$

i.e. the systems

$$v_i^*(\mathbf{x}) = \int\limits_\Omega G(\mathbf{x}, \boldsymbol{\xi}) \hat{f}_i(\boldsymbol{\xi}, v_1^*(\boldsymbol{\xi}), \ldots, v_n^*(\boldsymbol{\xi}), w_1^*(\boldsymbol{\xi}), \ldots, w_n^*(\boldsymbol{\xi})) d\boldsymbol{\xi},$$

$$w_i^*(\mathbf{x}) = \int\limits_\Omega G(\mathbf{x}, \boldsymbol{\xi}) \hat{f}_i(\boldsymbol{\xi}, w_1^*(\boldsymbol{\xi}), \ldots, w_n^*(\boldsymbol{\xi}), v_1^*(\boldsymbol{\xi}), \ldots, v_n^*(\boldsymbol{\xi})) d\boldsymbol{\xi}, \ i = 1, \ldots, n.$$

If we have received that $\mathbf{v}^* = \mathbf{w}^* = \mathbf{u}^*$, then $\mathbf{u}^*$ is the unique on the cone segment $\langle \mathbf{v}^0, \mathbf{w}^0 \rangle$ fixed point of the operator $\mathbf{T}$, and hence, $\mathbf{u}^*$ is the unique on $\langle \mathbf{v}^0, \mathbf{w}^0 \rangle$ solution of the boundary value problem (1), (2).

Sufficient condition for the implementation of equality $\mathbf{v}^* = \mathbf{w}^*$ is the condition [3] of the existence of such $\alpha \in (0; 1)$, that

$$\left\| \hat{\mathbf{T}}(\mathbf{v}, \mathbf{w}) - \hat{\mathbf{T}}(\mathbf{w}, \mathbf{v}) \right\|_n \leq \alpha \|\mathbf{v} - \mathbf{w}\|_n \text{ for all } \mathbf{v}, \mathbf{w} \in \langle \mathbf{v}^0, \mathbf{w}^0 \rangle .$$

Let the functions $\hat{f}_i(\mathbf{x}, v_1(\mathbf{x}), \ldots, v_n(\mathbf{x}), w_1(\mathbf{x}), \ldots, w_n(\mathbf{x}))$, $i = 1, \ldots, n$, for all positive numbers $v_1, \ldots, v_n$, $w_1, \ldots, w_n$ and for all $\mathbf{x} \in \Omega$ satisfy the inequality

$$\left| \hat{f}_i(\mathbf{x}, \mathbf{v}, \mathbf{w}) - \hat{f}_i(\mathbf{x}, \mathbf{w}, \mathbf{v}) \right| \leq L_i \max\{|v_1 - w_1|, \ldots, |v_n - w_n|\}, \\ i = 1, \ldots, n, \tag{9}$$

where $L_i > 0$, $i = 1, \ldots, n$.

Then there will be an estimate

$$\left\| \hat{\mathbf{T}}(\mathbf{v}, \mathbf{w}) - \hat{\mathbf{T}}(\mathbf{w}, \mathbf{v}) \right\|_n \leq LM \|\mathbf{v} - \mathbf{w}\|_n, \tag{10}$$

where $L = \max\{L_1, \ldots, L_n\}$, $M = \max\limits_{x \in \bar{\Omega}} \int\limits_\Omega G(\mathbf{x}, \boldsymbol{\xi}) d\boldsymbol{\xi}$.

In addition, on the basis of estimate (10) we obtain that

$$\left\| \mathbf{w}^{(k)} - \mathbf{v}^{(k)} \right\|_n = \left\| \hat{\mathbf{T}}(\mathbf{w}^{(k-1)}, \mathbf{v}^{(k-1)}) - \hat{\mathbf{T}}(\mathbf{v}^{(k-1)}, \mathbf{w}^{(k-1)}) \right\|_n \leq \\ \leq LM \left\| \mathbf{w}^{(k-1)} - \mathbf{v}^{(k-1)} \right\|_n \leq \cdots \leq (LM)^k \left\| \mathbf{w}^{(0)} - \mathbf{v}^{(0)} \right\|_n.$$

Hence, the following theorem holds.

Theorem 1. *Let a heterotone operator* $\mathbf{T}$ *of the form (4) for which operator* $\hat{\mathbf{T}}$ *of the form (5) is an companion one, has a strongly invariant cone segment* $\left\langle \mathbf{v}^0, \mathbf{w}^0 \right\rangle$ *and the inequalities (9) are executed, moreover* $LM < 1$*. Then the iteration process (6–(8) converges to the unique on* $\left\langle \mathbf{v}^0, \mathbf{w}^0 \right\rangle$ *solution* $\mathbf{u}^*$ *of the boundary value problem (1), (2), and the following inequalities*

$$\mathbf{v}^0 = \mathbf{v}^{(0)} \leqslant \mathbf{v}^{(1)} \leqslant \ldots \leqslant \mathbf{v}^{(k)} \leqslant \ldots \leqslant \mathbf{u}^* \leqslant$$
$$\leqslant \ldots \leqslant \mathbf{w}^{(k)} \leqslant \ldots \leqslant \mathbf{w}^{(1)} \leqslant \mathbf{w}^{(0)} = \mathbf{w}^0$$

are satisfied and

$$\left\| \mathbf{w}^{(k)} - \mathbf{v}^{(k)} \right\|_n \le (LM)^k \left\| \mathbf{w}^{(0)} - \mathbf{v}^{(0)} \right\|_n. \tag{11}$$

Another condition that ensures the uniqueness of the positive solution of the boundary value problem (1), (2) is u_0-pseudoconcavity of the operator $\mathbf{T}$ of the form (4) [13, 14].

Suppose that for all positive numbers $v_1, \ldots, v_n$, $w_1, \ldots, w_n$ and any $\tau \in (0, 1)$ the inequalities

$$\hat{f}_i \left(\mathbf{x}, \tau \mathbf{v}, \frac{1}{\tau} \mathbf{w} \right) > \tau \hat{f}_i(\mathbf{x}, \mathbf{v}, \mathbf{w}), \quad \mathbf{x} \in \Omega, \quad i = 1, \ldots, n, \tag{12}$$

are performed.

Let us denote $u_0(\mathbf{x}) = \int\limits_\Omega G(\mathbf{x}, \boldsymbol{\xi}) d\boldsymbol{\xi}$. Then [7, 13, 14] for any $\mathbf{v}, \mathbf{w} \in \mathcal{K}_+$ there are such $\alpha_i(\mathbf{v}, \mathbf{w}) > 0$, $\beta_i(\mathbf{v}, \mathbf{w}) > 0$, $\tilde{\alpha}_i(\mathbf{v}, \mathbf{w}) > 0$, $\tilde{\beta}_i(\mathbf{v}, \mathbf{w}) > 0$, $i = 1, \ldots, n$, that

$$\alpha_i(\mathbf{v}, \mathbf{w}) u_0(\mathbf{x}) \le \int\limits_\Omega G(\mathbf{x}, \boldsymbol{\xi}) \hat{f}_i \boldsymbol{\xi}, \mathbf{v}(\boldsymbol{\xi}), \mathbf{w}(\boldsymbol{\xi})) d\boldsymbol{\xi} \le \beta_i(\mathbf{v}, \mathbf{w}) u_0(\mathbf{x}), \ i = 1, \ldots, n,$$

$$\tilde{\alpha}_i(\mathbf{v}, \mathbf{w}) u_0(\mathbf{x}) \le \int\limits_\Omega G(\mathbf{x}, \boldsymbol{\xi}) \left[\hat{f}_i \left(\mathbf{x}, \tau \mathbf{v}, \frac{1}{\tau} \mathbf{w} \right) - \tau \hat{f}_i(\mathbf{x}, \mathbf{v}, \mathbf{w}) \right] d\boldsymbol{\xi} \le$$
$$\le \tilde{\beta}_i(\mathbf{v}, \mathbf{w}) u_0(\mathbf{x}), \ i = 1, \ldots, n.$$

Hence we will have that

$$\int\limits_\Omega G(\mathbf{x}, \boldsymbol{\xi}) \hat{f}_i \left(\mathbf{x}, \tau \mathbf{v}, \frac{1}{\tau} \mathbf{w} \right) d\boldsymbol{\xi} \ge \tilde{\alpha}_i(\mathbf{v}, \mathbf{w}) u_0(\mathbf{x}) + \tau \int\limits_\Omega G(\mathbf{x}, \boldsymbol{\xi}) \hat{f}_i(\mathbf{x}, \mathbf{v}, \mathbf{w}) d\boldsymbol{\xi} \ge$$
$$\ge \tau \left(1 + \frac{\tilde{\alpha}_i(\mathbf{v}, \mathbf{w})}{\tau \beta_i(\mathbf{v}, \mathbf{w})} \right) \int\limits_\Omega G(\mathbf{x}, \boldsymbol{\xi}) \hat{f}_i(\mathbf{x}, \mathbf{v}, \mathbf{w}) d\boldsymbol{\xi}, \ i = 1, \ldots, n,$$

i.e.

$$\hat{\mathbf{T}} \left(\tau \mathbf{v}, \frac{1}{\tau} \mathbf{w} \right) \ge \tau [1 + \eta(\mathbf{v}, \mathbf{w}, \tau)] \hat{\mathbf{T}}(\mathbf{v}, \mathbf{w}), \tag{13}$$

where $\eta(\mathbf{v}, \mathbf{w}, \tau) = \min \left\{ \dfrac{\tilde{\alpha}_1(\mathbf{v}, \mathbf{w})}{\tau \beta_1(\mathbf{v}, \mathbf{w})}, \ldots, \dfrac{\tilde{\alpha}_n(\mathbf{v}, \mathbf{w})}{\tau \beta_n(\mathbf{v}, \mathbf{w})} \right\}$.

Inequality (13) means u_0-pseudoconcavity of the operator $\mathbf{T}$.

Hence, the following theorem holds.

Theorem 2. *Let a heterotone operator* $\mathbf{T}$ *of the form (4), for which the operator* $\hat{\mathbf{T}}$ *of the form (5) is an companion one, has a strongly invariant cone segment* $\langle \mathbf{v}^0, \mathbf{w}^0 \rangle$ *and the inequalities (12) are performed. Then the iteration process (6) – (8) converges to the unique positive solution* $\mathbf{u}^* \in \langle \mathbf{v}^0, \mathbf{w}^0 \rangle$ *of the boundary value problem (1), (2), and the following inequalities*

$$\mathbf{v}^0 = \mathbf{v}^{(0)} \leqslant \mathbf{v}^{(1)} \leqslant \ldots \leqslant \mathbf{v}^{(k)} \leqslant \ldots \leqslant \mathbf{u}^* \leqslant$$
$$\leqslant \ldots \leqslant \mathbf{w}^{(k)} \leqslant \ldots \leqslant \mathbf{w}^{(1)} \leqslant \mathbf{w}^{(0)} = \mathbf{w}^0$$

are satisfied.

Note that the advantage of constructed two-sided iterative processes is that at each k iteration we have a convenient a posteriori estimation of the error for an approximate solution $\mathbf{u}^{(k)}(\mathbf{x}) = \dfrac{1}{2}(\mathbf{w}^{(k)}(\mathbf{x}) + \mathbf{v}^{(k)}(\mathbf{x}))$:

$$\left\|\mathbf{u}^* - \mathbf{u}^{(k)}\right\|_n \le \frac{1}{2}\left\|\mathbf{w}^{(k)} - \mathbf{v}^{(k)}\right\|_n.$$

Then, if accuracy $\varepsilon > 0$ is given, then the iterative process should be carried out before the inequality $\max\{\max\limits_{\mathbf{x}\in\bar{\Omega}}(w_1^{(k)}(\mathbf{x}) - v_1^{(k)}(\mathbf{x})), \ldots, \max\limits_{\mathbf{x}\in\bar{\Omega}}(w_n^{(k)}(\mathbf{x}) - v_n^{(k)}(\mathbf{x}))\} < 2\varepsilon$ will be performed and with accuracy ε it can be assumed that $\mathbf{u}^*(\mathbf{x}) \approx \mathbf{u}^{(k)}(\mathbf{x})$.

Also, based on the inequality (11) we can obtain an estimate for the number of iterations required to achieve the given accuracy. Indeed, from the inequalities

$$\left\|\mathbf{u}^* - \mathbf{u}^{(k)}\right\|_n \le \frac{1}{2}\left\|\mathbf{w}^{(k)} - \mathbf{v}^{(k)}\right\|_n \le \frac{(LM)^k}{2}\left\|\mathbf{w}^{(0)} - \mathbf{v}^{(0)}\right\|_n < \varepsilon$$

we find that to achieve accuracy ε

$$k_0(\varepsilon) = \left[\frac{\ln \dfrac{\left\|\mathbf{w}^{(0)} - \mathbf{v}^{(0)}\right\|_n}{2\varepsilon}}{\ln \dfrac{1}{LM}}\right] + 1$$

iterations must be done, where the square brackets denote an integer part of the number.

4. Numerical experiment

The construction of the two-sided approximations to the positive solution of the boundary value problem (1), (2) will be demonstrated on the system of two Lane-Emden equations with a homogeneous Dirichlet condition:

$$-\Delta u_1 = u_2^{p_1}, \quad -\Delta u_2 = u_1^{-p_2} \quad \text{in} \quad \Omega, \tag{14}$$

$$u_1|_{\partial\Omega} = u_2|_{\partial\Omega} = 0, \tag{15}$$

where $p_1 > 0$, $p_2 > 0$.

The construction of two-sided approximations to the positive solution of the Lane-Emden equation $-\Delta u = u^p$ for $p = \dfrac{1}{2}$ was made in [5].

The questions of the existence and uniqueness of the solution of problem (14), (15) in the case when Ω is a sphere of radius R, $p_1 > 0$, $p_2 < 0$ were investigated in [2].

The functions $f_1(\mathbf{x}, u_1, u_2) = u_2^{p_1}$, $f_2(\mathbf{x}, u_1, u_2) = u_1^{-p_2}$ are positive and continuous on a set of variables, if $u_1, u_2 > 0$, and allow a diagonal representation by using the functions

$$\hat{f}_1(\mathbf{x}, v_1, v_2, w_1, w_2) = v_2^{p_1}, \quad \hat{f}_2(\mathbf{x}, v_1, v_2, w_1, w_2) = w_1^{-p_2}. \tag{16}$$

The problem (14), (15) is replaced by the equivalent system of Hammerstein integral equations

$$u_1(\mathbf{x}) = \int\limits_\Omega G(\mathbf{x}, \boldsymbol{\xi}) u_2^{p_1}(\boldsymbol{\xi}) d\boldsymbol{\xi}, \quad u_2(\mathbf{x}) = \int\limits_\Omega G(\mathbf{x}, \boldsymbol{\xi}) u_1^{-p_2}(\boldsymbol{\xi}) d\boldsymbol{\xi}. \tag{17}$$

With the system (17) we will associate a heterotone operator

$$\mathbf{T}(u_1, u_2) = \left(\int\limits_\Omega G(\mathbf{x}, \boldsymbol{\xi}) u_2^{p_1}(\boldsymbol{\xi}) d\boldsymbol{\xi}, \int\limits_\Omega G(\mathbf{x}, \boldsymbol{\xi}) u_1^{-p_2}(\boldsymbol{\xi}) d\boldsymbol{\xi} \right), \tag{18}$$

for which the companion operator has the form

$$\hat{\mathbf{T}}(v_1, v_2, w_1, w_2) = \left(\int\limits_\Omega G(\mathbf{x}, \boldsymbol{\xi}) v_2^{p_1}(\boldsymbol{\xi}) d\boldsymbol{\xi}, \int\limits_\Omega G(\mathbf{x}, \boldsymbol{\xi}) w_1^{-p_2}(\boldsymbol{\xi}) d\boldsymbol{\xi} \right).$$

Condition (12) for functions (16) leads to inequalities

$$\hat{f}_1\left(\mathbf{x}, \tau v_1, \tau v_2, \frac{1}{\tau} w_1, \frac{1}{\tau} w_2\right) = (\tau v_2)^{p_1} > \tau \hat{f}_1(\mathbf{x}, v_1, v_2, w_1, w_2) = \tau v_2^{p_1},$$
$$\hat{f}_2\left(\mathbf{x}, \tau v_1, \tau v_2, \frac{1}{\tau} w_1, \frac{1}{\tau} w_2\right) = \left(\frac{1}{\tau} w_1\right)^{-p_2} > \tau \hat{f}_2(\mathbf{x}, v_1, v_2, w_1, w_2) = \tau w_1^{-p_2},$$

whereof $\tau^{p_1 - 1} > 1$, $\tau^{p_2 - 1} > 1$, i.e. $0 < p_1 < 1$, $0 < p_2 < 1$.

For the operator (18) a strongly invariant cone segment will be search in the form $\langle \mathbf{v}^0, \mathbf{w}^0 \rangle$, where

$$\begin{aligned}
&\mathbf{v}^0(\mathbf{x}) = (v_1^0(\mathbf{x}), v_2^0(\mathbf{x})) = (\alpha_1 \omega(\mathbf{x}), \alpha_2 \omega(\mathbf{x})), \\
&\mathbf{w}^0(\mathbf{x}) = (w_1^0(\mathbf{x}), w_2^0(\mathbf{x})) = (\beta_1 \omega(\mathbf{x}), \beta_2 \omega(\mathbf{x})), \\
&0 \le \alpha_1 < \beta_1, \quad 0 \le \alpha_2 < \beta_2,
\end{aligned}$$

and the function $\omega(\mathbf{x})$ satisfies the conditions a) - c) of section 3.

The system of inequalities for the determination α_1, α_2, β_1, β_2 has the form:

$$\begin{aligned}
&\alpha_2^{p_1}\int\limits_{\Omega} G(\mathbf{x},\boldsymbol{\xi})\omega^{p_1}(\boldsymbol{\xi})d\boldsymbol{\xi} \ge \alpha_1\omega(\mathbf{x}),\\
&\beta_1^{-p_2}\int\limits_{\Omega} G(\mathbf{x},\boldsymbol{\xi})\omega^{-p_2}(\boldsymbol{\xi})d\boldsymbol{\xi} \ge \alpha_2\omega(\mathbf{x}),\\
&\beta_2^{p_1}\int\limits_{\Omega} G(\mathbf{x},\boldsymbol{\xi})\omega^{p_1}(\boldsymbol{\xi})d\boldsymbol{\xi} \le \beta_1\omega(\mathbf{x}),\\
&\alpha_1^{-p_2}\int\limits_{\Omega} G(\mathbf{x},\boldsymbol{\xi})\omega^{-p_2}(\boldsymbol{\xi})d\boldsymbol{\xi} \le \beta_2\omega(\mathbf{x}) \text{ for all } \mathbf{x}\in\Omega.
\end{aligned}\tag{19}$$

Hence, the following theorem holds.

Theorem 3. *Let $0 < p_1 < 1$, $0 < p_2 < 1$ and the system (19) has a solution $(\alpha_1, \alpha_2, \beta_1, \beta_2)$ such that $0 \le \alpha_1 < \beta_1$, $0 \le \alpha_2 < \beta_2$. Then the iterative process*

$$v_1^{(k+1)}(\mathbf{x}) = \int\limits_{\Omega} G(\mathbf{x},\boldsymbol{\xi})(v_2^{(k)}(\boldsymbol{\xi}))^{p_1}d\boldsymbol{\xi},\ v_2^{(k+1)}(\mathbf{x}) = \int\limits_{\Omega} G(\mathbf{x},\boldsymbol{\xi})(w_1^{(k)}(\boldsymbol{\xi}))^{-p_2}d\boldsymbol{\xi},$$

$$w_1^{(k+1)}(\mathbf{x}) = \int\limits_{\Omega} G(\mathbf{x},\boldsymbol{\xi})(w_2^{(k)}(\boldsymbol{\xi}))^{p_1}d\boldsymbol{\xi},\ w_2^{(k+1)}(\mathbf{x}) = \int\limits_{\Omega} G(\mathbf{x},\boldsymbol{\xi})(v_1^{(k)}(\boldsymbol{\xi}))^{-p_2}d\boldsymbol{\xi},$$

$k = 0, 1, 2, \ldots,$

where $v_1^{(0)}(\mathbf{x}) = \alpha_1\omega(\mathbf{x})$, $v_2^{(0)}(\mathbf{x}) = \alpha_2\omega(\mathbf{x})$, $w_1^{(0)}(\mathbf{x}) = \beta_1\omega(\mathbf{x})$, $w_2^{(0)}(\mathbf{x}) = \beta_2\omega(\mathbf{x})$, converges to the unique positive solution $(u_1^*(\mathbf{x}), u_2^*(\mathbf{x}))$ of system (14), (15), and besides, for all $\mathbf{x}\in\bar\Omega$ the following inequalities

$$\alpha_1\omega(\mathbf{x}) = v_1^{(0)}(\mathbf{x}) \le v_1^{(1)}(\mathbf{x}) \le \ldots \le u_1^*(\mathbf{x}) \le \ldots \le w_1^{(1)}(\mathbf{x}) \le w_1^{(0)}(\mathbf{x}) = \beta_1\omega(\mathbf{x}),$$

$$\alpha_2\omega(\mathbf{x}) = v_2^{(0)}(\mathbf{x}) \le v_2^{(1)}(\mathbf{x}) \le \ldots \le u_2^*(\mathbf{x}) \le \ldots \le w_2^{(1)}(\mathbf{x}) \le w_2^{(0)}(\mathbf{x}) = \beta_2\omega(\mathbf{x})$$

are satisfied.

A computational experiment was carried out for the values $p_1 = \dfrac{1}{4}$, $p_2 = \dfrac{1}{3}$, if $m = 2$ and $\Omega = \{\mathbf{x} = (x_1, x_2) : |\mathbf{x}| < 1\}$ is unit circle. For this domain we have $\omega(\mathbf{x}) == \dfrac{1}{2}(1 - x_1^2 - x_2^2)$, $G(\mathbf{x},\boldsymbol{\xi}) = \dfrac{1}{2\pi}\ln\dfrac{\rho r_{\mathbf{x}\boldsymbol{\xi}^1}}{r_{\mathbf{x}\boldsymbol{\xi}}}$, where $\rho = \sqrt{\xi_1^2 + \xi_2^2}$, points $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^1$ are symmetric with respect to the circle of the unit radius, $r_{\mathbf{x}\boldsymbol{\xi}}$, $r_{\mathbf{x}\boldsymbol{\xi}^1}$ are distances between points $\mathbf{x}$, $\boldsymbol{\xi}$ and $\mathbf{x}$, $\boldsymbol{\xi}^1$ accordingly. The solution of the system of inequalities (19) is, for example, numbers $\alpha_1 = 0.332$, $\alpha_2 = 0.959$, $\beta_1 = 0.418$, $\beta_2 = 1.364$. Accuracy $\varepsilon = 10^{-4}$ was reached on the sixth iteration.

An obtained approximate solution

$$u_1^{(6)}(\mathbf{x}) = \frac{v_1^{(6)}(\mathbf{x}) + w_1^{(6)}(\mathbf{x})}{2}, \quad u_2^{(6)}(\mathbf{x}) = \frac{v_2^{(6)}(\mathbf{x}) + w_2^{(6)}(\mathbf{x})}{2}$$

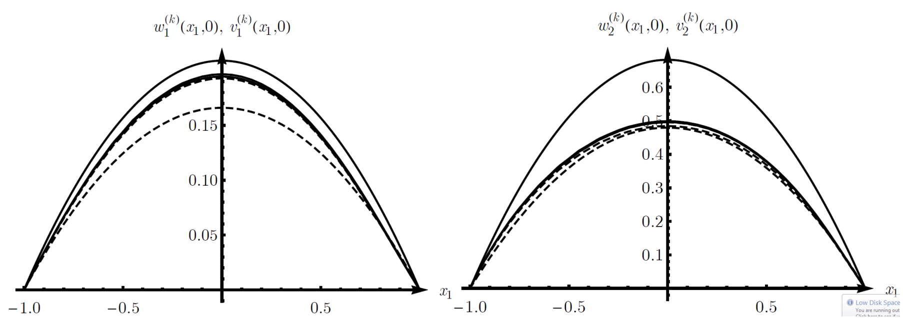
has a radial symmetry.

TABL. 1. The values of the error estimation of the approximate solution

Number of iteration k	$\varepsilon_1^{(k)}$	$\varepsilon_2^{(k)}$
0	$0.22 \cdot 10^{-1}$	$0.10 \cdot 10^{0}$
1	$0.88 \cdot 10^{-2}$	$0.19 \cdot 10^{-1}$
2	$0.19 \cdot 10^{-2}$	$0.72 \cdot 10^{-2}$
3	$0.71 \cdot 10^{-3}$	$0.16 \cdot 10^{-2}$
4	$0.16 \cdot 10^{-3}$	$0.61 \cdot 10^{-3}$
5	$0.60 \cdot 10^{-4}$	$0.13 \cdot 10^{-3}$
6	$0.13 \cdot 10^{-4}$	$0.51 \cdot 10^{-4}$

TABL. 2. The values of the approximate solution at the points $\mathbf{x}_i = (0.25i, 0)$, $i = 0, 1, 2, 3$

$\mathbf{x}_i = (0.25i, 0)$	(0,0)	(0.25,0)	(0.5,0)	(0.75,0)
$u_1^{(6)}(\mathbf{x}_i)$	0.1946	0.1806	0.1397	0.0752
$u_2^{(6)}(\mathbf{x}_i)$	0.4960	0.4674	0.3781	0.2192



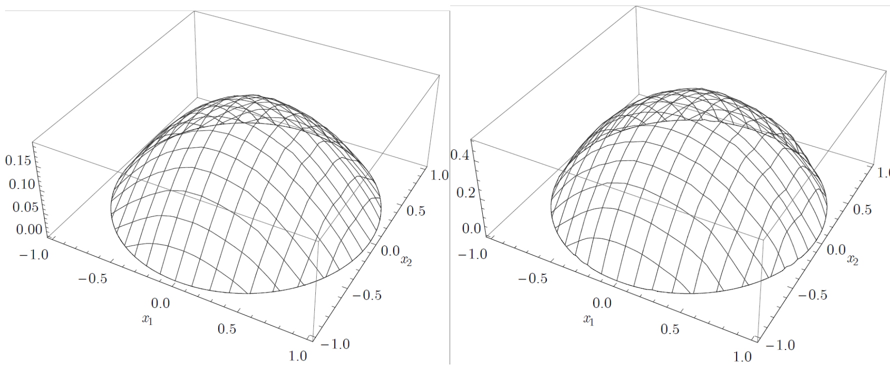
(A) (B)

FIG. 1. Graphs of cross-sections of upper and lower approximations $w_1^{(k)}(x_1, 0)$, $v_1^{(k)}(x_1, 0)$ (a) and $w_2^{(k)}(x_1, 0)$, $v_2^{(k)}(x_1, 0)$ (b), $k = 0, 2, 4, 6$

Table 1 gives the data on how the estimate

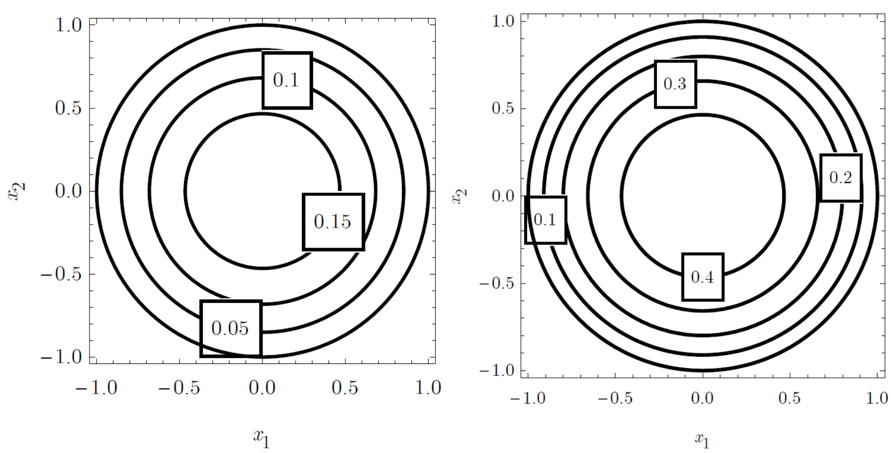
$$\varepsilon_i^{(k)} = \max_{\mathbf{x} \in \bar{\Omega}} \frac{1}{2} \left| w_i^{(k)}(\mathbf{x}) - v_i^{(k)}(\mathbf{x}) \right|$$

the norm of error $\left\| u_i^* - u_i^{(k)} \right\|$ of an approximate solution $u_i^{(k)}(\mathbf{x})$, $i = 1, 2$, is changed, depending on the iteration number k, $k = 0, 1, \ldots, 6$. Table 2 shows the values, found with accuracy $\varepsilon = 10^{-4}$ of the approximate solution $u_1^{(6)}(\mathbf{x})$, $u_2^{(6)}(\mathbf{x})$ at points located on the ray $\varphi = 0$. It was found that $\left\| u_1^{(6)} \right\| = 0.1946$, $\left\| u_2^{(6)} \right\| = 0.4960$.



(A) (B)

FIG. 2. Surfaces of approximate solutions $u_1^{(6)}(\mathbf{x})$ (a) and $u_2^{(6)}(\mathbf{x})$ (b)



(A) (B)

FIG. 3. Contour lines of approximate solutions $u_1^{(6)}(\mathbf{x})$ (a) and $u_2^{(6)}(\mathbf{x})$ (b)

Fig. 1 shows the graphs of the cross-sections of the upper $w_1^{(k)}(\mathbf{x})$, $w_2^{(k)}(\mathbf{x})$ and the lower $v_1^{(k)}(\mathbf{x})$, $v_2^{(k)}(\mathbf{x})$ approximations at $x_2 = 0$ for $k = 0, 2, 4, 6$. Fig. 2, 3 show the surfaces of the approximate solutions $u_1^{(6)}(\mathbf{x})$, $u_2^{(6)}(\mathbf{x})$ and their contour lines respectively.

5. Conclusions

The paper proposed a method of constructing the two-sided approximations to a positive solution of the homogeneous Dirichlet problem for a system of semilinear elliptic equations. The numerical experiment, conducted for the Lane-Emden system, demonstrated the possibilities and effectiveness of the method. The proposed approach to numerical solution of semilinear systems can be used in solving various applications, the mathematical models of which is the problem (1), (2).

The limitation of using the proposed method may be due to the fact that the Green's function of the first boundary value problem for an operator $-\Delta$ is known only for a certain number of classical domains. When considering the problem (1), (2) in the domains of non classical geometry or in domains for which the Green's function is known, but has a complex analytic expression, to construct the corresponding (1), (2) system of integral equations, can be used an approach based on the corresponding Green's quasi-function [15].

Bibliography

1. Cui R. Uniqueness of the positive solution for a class of semilinear elliptic systems / R. Cui, Y. Wang, J. Shi // Nonlinear Analysis: Theory, Methods & Applications. – 2007. – Vol. 67, № 6. – P. 1710-1714.
2. Dalmasso R. Existence and uniqueness of positive radial solutions for the Lane–Emden system / R. Dalmasso // Nonlinear Analysis: Theory, Methods & Applications. – 2004. – Vol. 57, № 3. – P. 341-348.
3. Guo D. Coupled fixed points of nonlinear operators with applications / D. Guo, V. Lakshmikantham // Nonlinear Analysis: Theory, Methods & Applications. – 1987. – Vol. 11, № 5. – P. 623-632.
4. Kolosov A. I. A class of boundary value problems reducible to an equation with a heterotonic operator / A. I. Kolosov // Differentsial'nye Uravneniya. – 1985. – Vol. 21, № 11. – P. 1884-1891.
5. Kolosova S. V. On the construction of two-sided approximations to the positive solution of the Lane-Emden equation / S. V. Kolosova, V. S. Lukhanin, M. V. Sidorov // Visnyk of Zaporizhzhya National University. – Physical and mathematical Sciences. – 2015. – № 3. – P. 107-120. (in Russian).
6. Korman P. On Lane-Emden type systems / P. Korman, J. Shi // Discrete Contin. Dyn. Syst. – 2005. – P. 510-517.
7. Krasnosel'skij M. A. Positive Solutions of Operator Equations / M. A. Krasnosel'skij. – Moscow: Fizmatgiz, 1962. (in Russian).
8. Kurpel' N. S. Two-sided Operator Inequalities and their Application / N. S. Kurpel', B. A. Shuvar. – Kiev: Naukova Dumka, 1980. (in Russian).
9. Li C. A degree theory framework for semilinear elliptic systems / C. Li, J. Villavert // Proceedings of the American Mathematical Society. – 2016. – Vol. 144, № 9. – P. 3731-3740.
10. Maniwa M. Uniqueness and existence of positive solutions for some semilinear elliptic systems / M. Maniwa // Nonlinear Analysis: Theory, Methods & Applications. – 2004. – Vol. 59, № 6. – P. 993-999.
11. Moore J. Existence of multiple quasifixed points of mixed monotone operators by iterative techniques / J. Moore // Applied Mathematics and Computation. – 1981. – Vol. 9, № 2. – P. 135-141.
12. Pao C. V. Nonlinear Parabolic And Elliptic Equations / C. V. Pao. – New-York: Plenum Press, 1992.
13. Opoĭcev V. I. Generalization of the theory of monotone and concave operators / V. I. Opoĭcev // Trudy Moskov. Mat. Obshch. – 1978. – Vol. 36. – P. 237-273. (in Russian).
14. Opoĭcev V. I. Nonlinear Operators in Spaces with a Cone / V. I. Opoĭcev, T. A. Khurodze. – Tbilisi: Izdatel'stvo Tbilisskogo Universiteta, 1984. (in Russian).
15. Rvacev V. L. Theory of R-functions and its Some Applications / V. L. Rvacev. – Kiev: Naukova Dumka, 1982. (in Russian).

16. Troy W. C. Symmetry properties in systems of semilinear elliptic equations / W. C. Troy // Journal of Differential Equations. – 1981. – Vol. 42., № 3. – P. 400-413.

Maksim Sidorov,
Kharkiv National University of Radio Electronics,
14, Nauki Ave., Kharkiv, 61166, Ukraine.

Received 24.08.2017; revised 24.09.2017

Журнал обчислювальної та прикладної математики | 2017 № 3 (126) | Journal of Numerical & Applied Mathematics

UDC 517.983.27

SINC APPROXIMATION OF ALGEBRAICALLY DECAYING FUNCTIONS

D. O. Sytnyk

Резюме. В роботі запропоновано узагальнення Sinc інтерполяційного методу, яке дозволяє наближати на $\mathbb{R}$ функції спадаючі алгебраїчно. Подібно до класичної Sinc інтерполяції ми формулюємо два типи оцінок похибки. Перший стосується загального класу функцій, що мають алгебраїчний порядок спадання на $\mathbb{R}$. Оцінки похибки другого типу є справедливими для випадку коли порядок спадання функції відомий у смузі комплексної площини навколо дійсної осі. Теоретичні викладки підкріплені чисельними експериментами.

Abstract. An extension of sinc interpolation on $\mathbb{R}$ to the class of algebraically decaying functions is developed in the paper. Similar to the classical sinc interpolation we establish two types of error estimates. First covers a wider class of functions with the algebraic order of decay on $\mathbb{R}$. The second type of error estimates governs the case when the order of function's decay can be estimated everywhere in the horizontal strip of complex plane around $\mathbb{R}$. The numerical examples are provided.

1. Introduction

We begin by introducing some necessary notation. Let

$$\operatorname{sinc}(x) = \frac{\sin \pi x}{\pi x},$$

$$S\{k,h\}(x) = \operatorname{sinc}\left(\frac{x}{h} - k\right), \quad h > 0,\ k \in \mathbb{Z}. \tag{1}$$

By $H^1(D_d)$ in the paper we denote the class of functions $f(x)$ analytic in the horizontal strip D_d

$$D_d = \{z = x + iy \quad x \in (-\infty, \infty),\ |y| \le d\}, \tag{2}$$

and such, that the quantity

$$N_1(f, D_d) \equiv \int_{\partial D_d} |f(z)| dz,$$

is bounded. Next, for some given $h > 0$ and integer $N > 0$ we define a sinc interpolation polynomial as

$$C_N\{f,h\}(x) = \sum_{k=-N}^{N} f(kh) S\{k,h\}(x). \tag{3}$$

Key words. Sinc methods, sinc interpolation, algebraically decaying functions, Lambert-W function, polynomial order of convergence, approximation on real-line.

The following classical result characterize the accuracy of interpolation of $f \in H^1(D_d)$ by $C_N\{f,h\}(x)$ for the case, when $f(s)$ is exponentially decaying.

Theorem (Stenger [6, p. 137]) *Assume that the function $f \in H^1(D_d)$ is bounded by*

$$|f(x)| \le L\mathrm{e}^{-\alpha|x|}, \quad \forall x \in \mathbb{R}, \tag{4}$$

with some $\alpha, L > 0$. Then the error of $2N+1$ term sinc interpolation of $f(x)$ by $C_N\{f,h\}(x)$, satisfies the following estimate

$$\begin{gathered}\sup_{x\in\mathbb{R}} |f(x) - C_N\{f,h\}(x)| \le c\mathcal{E}_N,\\ \mathcal{E}_N = N^{1/2}\mathrm{e}^{-\sqrt{\pi d\alpha N}},\end{gathered} \tag{5}$$

provided that

$$h = \sqrt{\frac{\pi d}{\alpha N}}. \tag{6}$$

Here $c > 0$ is some constant dependent on f, d, α and independent on N. In this paper we extend the results of the above theorem to a class of algebraically decaying functions on $\mathbb{R}$. All theoretical considerations are given in sections 1,2. Section 3 is devoted to numerical examples and discussion.

2. Interpolation of functions with algebraic decay on real line

In this section we study the convergence of sinc interpolation for the class of algebraically decaying functions. Specifically, we consider the situation when a function $f(x)$ satisfies

$$|f(x)| \le \frac{L}{1+|x|^\alpha}, \quad \forall x \in \mathbb{R} \tag{7}$$

instead of inequality (4), convenient for the classical sinc methods [6].

Theorem 1. *Assume that the function $f \in H^1(D_d)$ has an algebraic decay defined by* (7) *with some $\alpha > 1$, $L > 0$. Then the error of $2N+1$-term sinc interpolation* (3) *satisfies the following estimate*

$$\begin{gathered}\sup_{x\in\mathbb{R}} |f(x) - C_N\{f,h\}(x)| \le c\mathcal{E}_N, \quad \forall x \in \mathbb{R},\\ \mathcal{E}_N = \frac{\alpha^\alpha (N+1)^{1-\alpha}}{(\alpha-1)(\pi d)^\alpha}\left(\mathbf{W}\left(\frac{\pi d}{\alpha}\left(\frac{\alpha-1}{\pi d}\right)^{\frac{1}{\alpha}}(N+1)^{\frac{\alpha-1}{\alpha}}\right)\right)^\alpha,\end{gathered} \tag{8}$$

provided that h in (3) *is chosen as*

$$h = \frac{\pi d}{\alpha}\left(\mathbf{W}\left(\frac{\pi d}{\alpha}\left(\frac{\alpha-1}{\pi d}\right)^{\frac{1}{\alpha}}(N+1)^{\frac{\alpha-1}{\alpha}}\right)\right)^{-1}. \tag{9}$$

Here $\mathbf{W}[\cdot]$ denotes a positive branch of the Lambert-W function, $c = c_1 N_1(f, D_d) + 2L$ and $c_1 > 1$ is the constant independent of N:

$$c_1 = \frac{(\pi d)^{2(\alpha-1)}(\alpha-1)^2}{(\pi d)^{2(\alpha-1)}(\alpha-1)^2 - \alpha^{2\alpha}\mathbf{W}^{2\alpha}\left(\frac{\pi d}{\alpha}\sqrt[\alpha]{\frac{\alpha-1}{\pi d}}\right)}. \tag{10}$$

Proof. For any fixed h the error of sinc interpolation can be represented as follows [6, equation (3.1.29)]

$$|f(x) - C_N\{f,h\}(x)| \le |f(x) - C_\infty\{f,h\}(x)| + \sum_{|k|>N} |f(kh)|\,.$$

Bound of the first term on the right-hand side of this formula was obtained in Theorem 3.1.3 from [6]. For $x \in \mathbb{R}$ this term satisfies

$$|f(x) - C_\infty\{f,h\}(x)| \le \frac{N_1(f,D_d)}{2\pi d \sinh \frac{\pi d}{h}} \le \frac{c_1 N_1(f,D_d)}{\pi d} \mathrm{e}^{-\frac{\pi d}{h}}, \tag{11}$$

where $c_1 > 1$ is some constant to be determined later. For the second term we get

$$\begin{aligned} \sum_{|k|>N} |f(kh)| \le \le 2L \sum_{k=N+1}^{\infty} (kh)^{-\alpha} \le 2L \int\limits_{N+1}^{\infty} (th)^{-\alpha} dt \\ \le \frac{2L(N+1)^{1-\alpha}}{(\alpha-1)h^\alpha}. \end{aligned} \tag{12}$$

The above sequence of inequalities is justified as long as $f(x)$ satisfy (7) with some $\alpha > 1$. For such $f(x)$, truncation error (12) decays algebraically as $N \to \infty$. In order to balance it with exponentially decaying discretization error (11) one needs to solve for h the equation

$$\frac{\mathrm{e}^{-\frac{\pi d}{h}}}{c_2} = \frac{(N+1)^{1-\alpha}}{(\alpha-1)h^\alpha}. \tag{13}$$

Let $s = \frac{\pi d}{\alpha} h^{-1}$ and assume that $c_2 > 0$ is some fixed parameter. Then, equation (13) takes the form

$$\frac{\pi d}{\alpha} \left(\frac{\alpha-1}{c_2} (N+1)^{\alpha-1} \right)^{\frac{1}{\alpha}} = s\mathrm{e}^s,$$

which has a unique solution

$$s = \mathbf{W}\left(\frac{\pi d}{\alpha} \left(\frac{\alpha-1}{c_2} (N+1)^{\alpha-1} \right)^{\frac{1}{\alpha}} \right).$$

Next, we set $c_2 = \pi d$ and substitute back the expression for s in terms of h to obtain (9). The proof of (8) is straightforward

$$\begin{aligned} |f(x) - C_N\{f,h\}(x)| \le (c_1 N_1(f,D_d) + 2L) \frac{(N+1)^{1-\alpha}}{(\alpha-1)h^\alpha} \le \\ \le c \frac{\alpha^\alpha (N+1)^{1-\alpha}}{(\alpha-1)(\pi d)^\alpha} \left(\mathbf{W}\left(\frac{\pi d}{\alpha} \left(\frac{\alpha-1}{\pi d} \right)^{\frac{1}{\alpha}} (N+1)^{\frac{\alpha-1}{\alpha}} \right) \right)^\alpha. \end{aligned}$$

Now, let us come back to the determination of c_1. The smallest c_1 suitable for (11) can be defined as follows

$$c_1 = \sup_{N \in \mathbb{Z}_+} \left\{ \frac{e^{\frac{\pi d}{h}}}{2 \sinh \frac{\pi d}{h}} \right\} = \max_{N \in \mathbb{Z}_+} \left(1 - e^{-\frac{2\pi d}{h}} \right)^{-1}.$$

Its not hard to see that the maximum is attained at $N = 0$. Therefore, the value of c_1:

$$c_1 = \left(1 - \exp\left(-2\alpha \mathbf{W}\left(\frac{\pi d}{\alpha}\sqrt[\alpha]{\frac{\alpha-1}{\pi d}}\right)\right)\right)^{-1}$$

is clearly greater than one, for any $\alpha > 1$, $d > 0$. To get (10) we apply the identity $\exp(-\mathbf{W}(x)) = \mathbf{W}(x)/x$ to the above formula for c_1 and rearrange the result accordingly

$$c_1 = \left(1 - \frac{\alpha^{2\alpha}}{(\pi d)^{2(\alpha-1)}(\alpha-1)^2}\left(\mathbf{W}\left(\frac{\pi d}{\alpha}\sqrt[\alpha]{\frac{\alpha-1}{\pi d}}\right)\right)^{2\alpha}\right)^{-1}$$
$$= \frac{(\pi d)^{2(\alpha-1)}(\alpha-1)^2}{(\pi d)^{2(\alpha-1)}(\alpha-1)^2 - \alpha^{2\alpha}\mathbf{W}^{2\alpha}\left(\frac{\pi d}{\alpha}\sqrt[\alpha]{\frac{\alpha-1}{\pi d}}\right)}.$$

The presence of $\mathbf{W}(x)$ in estimate (8) makes it harder to perceive the asymptotic behavior of the interpolation error intuitively. To fix that we recall a well-established result [5] on the asymptotic properties of $\mathbf{W}(x)$, valid for any $x > e$:

$$\ln x - \ln(\ln x) + \frac{\ln(\ln x)}{2\ln x} \le \mathbf{W}(x) \le \ln x - \ln(\ln x) + \frac{e\ln(\ln x)}{(e-1)\ln x}.$$

By using the above inequality along with the definition of $\mathbf{W}(x)$ and (13) we transform (8) in the following way

$$|f(x) - C_N\{f,h\}(x)| \le \frac{c}{\mathrm{e}^{\alpha s}} \le c\left(\frac{\ln\left(\frac{\pi d}{\alpha}\left(\frac{\alpha-1}{\pi d}\right)^{\frac{1}{\alpha}}(N+1)^{\frac{\alpha-1}{\alpha}}\right)}{\frac{\pi d}{\alpha}\left(\frac{\alpha-1}{\pi d}\right)^{\frac{1}{\alpha}}(N+1)^{\frac{\alpha-1}{\alpha}}}\right)^{\alpha} \le$$
$$\le \frac{c}{(\pi d)^{\alpha-1}}\left(\frac{N+1}{\alpha-1}\right)^{1-\alpha}\ln^{\alpha}\left(\pi d\left(\frac{\alpha-1}{\alpha^{\alpha}}\right)^{\frac{1}{\alpha-1}}(N+1)\right);$$

whence it is clear that the error of sinc interpolation provided by Theorem 1 is asymptotically equal to $(N+1)^{1-\alpha}\ln^{\alpha}(N+1)$ as $N \to \infty$. To analyze the error for small N we note that, in the view of (13), $\mathcal{E}_N$ is bounded by the exponent with a strictly decreasing negative argument. Consequently, for any $\alpha > 1$, $x \in \mathbb{R}$, the error $\sup_{x\in\mathbb{R}} |f(x) - C_N\{f,h\}(x)|$ lies within the interval $[0, c]$ and decreases as $N \to \infty$.

One might conclude from the foregoing analysis that a simple asymptotic formula $W(x) \approx \ln(x)$ can be used to redefine h (9) in terms of logarithms, which are computationally more favorable than the Lambert-W function. To explore this possibility we set

$$h = \frac{\pi d}{\alpha}\left(\ln\left(\frac{\pi d}{\alpha}\left(\frac{\alpha-1}{c_2}\right)^{\frac{1}{\alpha}}(N+1)^{\frac{\alpha-1}{\alpha}}\right)\right)^{-1},$$

and study the corresponding error terms of the approximation. Discretization error (11) is positive and monotonically decreasing in N for any $c_2 > 0$, since

h is monotonic. The principal part $\frac{(N+1)^{1-\alpha}}{(\alpha-1)h^{\alpha}}$ of truncation error (12) has one global maximum at $N = N_0$:

$$N_0 = \left(\frac{\alpha}{\pi\, d}\right)^{\frac{\alpha}{\alpha-1}} \exp\left(\frac{\alpha}{\alpha-1}\right)\left(\frac{\alpha-1}{c_2}\right)^{-\frac{1}{(\alpha-1)}} - 1.$$

To guarantee a monotonous decrease of the truncation error for all $N \geq 0$ we must require $N_0 = 0$, which yields $c_2 = (\alpha-1)\left(\frac{\pi d}{\alpha \mathrm{e}}\right)^{\alpha}$. The aforementioned formula for h is thereby reduced to

$$h = \frac{\pi d}{\alpha + (\alpha-1)\ln(N+1)}. \tag{14}$$

For such h, the error of sinc interpolation will be bounded by (8) with

$$\mathcal{E}_N = \frac{(N+1)^{1-\alpha}}{(\alpha-1)(\pi d)^{\alpha}} \left(\alpha + (\alpha-1)\ln(N+1)\right)^{\alpha}, \tag{15}$$

and $c = (\alpha-1)\left(\frac{\pi d}{\alpha e}\right)^{\alpha} N_1(f, D_d) + 2L$. The main concern with (15), is the presence of additional summand α when compared to (8).

Remark 1. *The definition of h from Theorem 1 can not be simplified by adopting $W(x) \approx \ln(x)$, since such simplification, as described by* (14), (15), *would make the approximation method ineffective for large α.*

With an additional a-priory knowledge about $f(x)$ we should be able to improve the convergence properties of $C_N\{f,h\}(x)$ described by Theorem 1. The following improvement of (8) offers a more realistic balance of discretization and truncation errors, presuming that both $N_1(f, D_d)$ and L are known.

Corollary 4. *Assume that the function $f(x)$ satisfies the conditions of Theorem 1. If*

$$h = \frac{\pi d}{\alpha}\left(\mathbf{W}\left(\frac{\pi d}{\alpha}\left(\frac{N_1(f,D_d)(\alpha-1)}{\pi d L}\right)^{\frac{1}{\alpha}}(N+1)^{\frac{\alpha-1}{\alpha}}\right)\right)^{-1}, \tag{16}$$

then the error of sinc interpolation fulfills estimate (8), *with $c = (c_1+2)L$ and $\mathcal{E}_N$ given by*

$$\mathcal{E}_N = \frac{(N+1)^{1-\alpha}}{(\alpha-1)} h^{-\alpha}.$$

Formula (16) was obtained in the same way as (9), except this time we set

$$c_2 = \frac{\pi d L}{N_1(f, D_d)}.$$

3. Interpolation of functions with algebraic decay in the strip

Corollary 4 is difficult to apply as it is, because the evaluation of $N_1(f, D_d)$ requires computation of the contour integral over ∂D_d. In order to make this result more applicable we note, that if $f \in H^1(D_r)$, for some $r > 0$, then $\lim\limits_{x\to\pm\infty} f(x+iy) = 0$ uniformly with respect to $y \in [d, d]$, for all $d \in (0, r)$ [2, Proposition 6]. Hence, for any $r > 0$ there exist a nonempty subspace of $H^1(D_r)$, such that its elements f satisfy

$$|f(z)| \leq \frac{L}{1+|z|^{\alpha}}, \quad \forall z \in D_d, \tag{17}$$

with some $d \in (0, r)$.

Theorem 2. *Assume that the function $f(z)$ is analytic in the horizontal strip D_d, $d > 0$. If $f(z)$ is bounded by* (17) *with some $\alpha > 1$, $L > 0$, then the error of sinc interpolation* (3) *satisfies the following estimate*

$$\begin{aligned} &\sup_{x\in\mathbb{R}} |f(x) - C_N\{f,h\}(x)| \leq c\mathcal{E}_N, \\ &\mathcal{E}_N = \frac{\alpha^{\alpha}(N+1)^{1-\alpha}}{(\alpha-1)(\pi d)^{\alpha}} h^{\alpha}, \end{aligned} \tag{18}$$

provided that

$$h = \frac{\pi d}{\alpha}\left(\mathbf{W}\left(\frac{\pi d}{\alpha}\left(\frac{4\beta(\alpha-1)}{\pi d}\right)^{\frac{1}{\alpha}}(N+1)^{\frac{\alpha-1}{\alpha}}\right)\right)^{-1}, \tag{19}$$

with $\beta = \min\left\{\frac{1}{\operatorname{sinc}(\alpha^{-1})}, \left(\frac{2}{d}\right)^{\alpha-1} B\left(\frac{\alpha}{2}-\frac{1}{2}, \frac{\alpha}{2}+\frac{1}{2}\right)\right\}$. *Here $B(\cdot,\cdot)$ is the beta function, $c = 2(c_1\beta+1)L$ and c_1 is the constant dependent on α, d.*

Proof.

$$\int\limits_{-\infty}^{+\infty} |f(x+id)|\,dx \leq \int\limits_{-\infty}^{+\infty} \frac{L dx}{1+|x+id|^{\alpha}} = 2L\int\limits_{0}^{+\infty} \frac{dx}{1+(x^2+d^2)^{\frac{\alpha}{2}}}, \tag{20}$$

$$\int\limits_{0}^{+\infty} \frac{dx}{1+(x^2+d^2)^{\alpha/2}} \leq \int\limits_{0}^{+\infty} \frac{dx}{1+x^{\alpha}} = \lim_{x\to\infty} \frac{x\ \Phi\left(-x^{\alpha},1,\alpha^{-1}\right)}{\alpha} =$$
$$= \lim_{x\to+\infty}\left|\frac{x\ \Phi\left(-x^{\alpha},1,\alpha^{-1}\right)}{\alpha}\right| = \lim_{\substack{\Re z\to+\infty \\ \Im z\to 0}}\left|\frac{z\Phi\left(-z^{\alpha},1,\alpha^{-1}\right)}{\alpha}\right|.$$

Here $\Re z$ and $\Im z$ is real and imaginary part of z correspondingly. To evaluate the last limit we employ Corollary 1 from [3]. It offers a convergent expansion of Hurwitz-Lerch zeta function $\Phi(z,s,a)$ when its second parameter s has integer value

$$z\Phi\left(z^{\alpha},1,\frac{1}{\alpha}\right) = \pi\left(sgn\left\{\operatorname{Arg}(\alpha\ln(z))\right\}i + \cot\frac{\pi}{\alpha}\right) - \sum_{k=1}^{\infty}\frac{z^{1-\alpha k}}{1/\alpha-k}. \tag{21}$$

The expression on the right of (21) is bounded and uniformly convergent to the left-hand side for any $\alpha > 1$, $|z| > 1$, such that $z^{\alpha} \notin (-\infty,-1)\cup(1,\infty)$. Therefore

$$\lim_{\substack{\Re z\to+\infty \\ \Im z\to 0}}\left|\frac{z\Phi\left(-z^{\alpha},1,\alpha^{-1}\right)}{\alpha}\right| = \frac{\pi}{\alpha}\sqrt{1+\cot^2\frac{\pi}{\alpha}} - \frac{1}{\alpha}\sum_{k=1}^{\infty}\lim_{\substack{\Re z\to+\infty \\ \Im z\to 0}}\frac{z^{1-\alpha k}}{1/\alpha-k},$$

which leads us to the bound

$$\int_{-\infty}^{+\infty} |f(x+id)|\, dx \leq \frac{2\pi L}{\alpha}\sqrt{1+\cot^2\frac{\pi}{\alpha}} = 2L\,\mathrm{sinc}^{-1}\left(\frac{1}{\alpha}\right). \tag{22}$$

For large d, the integral from (20) can be estimated as follows

$$\begin{aligned}\int_{0}^{+\infty} \frac{1}{1+(x^2+d^2)^{\alpha/2}}dx \leq \int_{0}^{+\infty} \frac{1}{(x^2+d^2)^{\alpha/2}}dx &= \\ &= \frac{\sqrt{\pi} d^{1-\alpha}\Gamma\left((\alpha-1)/2\right)}{2\Gamma\left(\alpha/2\right)} = \\ &= \frac{d^{1-\alpha}\Gamma\left((\alpha-1)/2\right)\Gamma\left((\alpha+1)/2\right)}{2^{2-\alpha}\Gamma(\alpha)} \leq \\ &\leq \frac{1}{2}B\left(\frac{\alpha}{2}-\frac{1}{2},\frac{\alpha}{2}+\frac{1}{2}\right)\left(\frac{2}{d}\right)^{\alpha-1}.\end{aligned}$$

To obtain the above estimate we used a well-known multiplication theorem [1, p. 4] for Gamma function $\Gamma(\cdot)$. The next bound is a direct consequence of the above formula and (20)

$$\int_{-\infty}^{+\infty} |f(x+id)|\, dx \leq 2LB\left(\frac{\alpha}{2}-\frac{1}{2},\frac{\alpha}{2}+\frac{1}{2}\right)\left(\frac{2}{d}\right)^{\alpha-1}. \tag{23}$$

By combining bounds (22), (23) and taking in to account the fact that the expression on the right of (17) is invariant with respect to $z \to \bar{z}$ we arrive at the following estimate

$$N_1(f,D_d) \leq 4L\min\left\{\frac{1}{\mathrm{sinc}\left(\frac{1}{\alpha}\right)}, \left(\frac{2}{d}\right)^{\alpha-1} B\left(\frac{\alpha}{2}-\frac{1}{2},\frac{\alpha}{2}+\frac{1}{2}\right)\right\}.$$

To finalize the proof, we evaluate (16) assuming that the value of $N_1(f,D_d)$ is equal to its estimate provided by the previous formula. This will get us (19).

4. Examples and discussion

In this section we consider several examples of the developed approximation method. As measure of experimental error we use a discrete norm

$$err = \max_{\forall x\in X} \left|f(x) - C_N\{f,h\}(x)\right|,$$

defined on a uniform grid $X = \left\{jh/2 \mid j = \overline{-2N,2N}\right\}$. With such choice of X the specified discrete norm ought to capture the contribution from both the descretization and truncation parts of the error. To experimentally check the convergence of $C_N\{f,h\}(x)$ we repeat the approximation procedure on a sequence of grids determined by

$$N_i \in \{1,2,4,8,16,32,64,128,256,512,1024\},$$

and the corresponding h_i evaluated by one of the formulas (9), (16) or (19).

Example 1. Let
$$f(x) = \frac{4}{2 + x^{2a}},$$
where $a \geq 2$ is integer. Then, the largest possible value of d such that $f(x)$ remains analytic in D_d, is equal to $\sqrt[2\alpha]{2} \sin \frac{\pi}{a}$. To simplify the computation of $N_1(f, D_d)$ we set $d = \frac{\sqrt[2a]{2}}{2} \sin \frac{\pi}{a}$, $a = 2$, then $N_1(f, D_d) \approx 4.550125680$, $L \approx 4.5$, $\alpha = 4$. The behaviour of an error $err(x) = f(x) - C_{32}\{f, h\}(x)$ for three different values of h is depicted in Fig. 1.

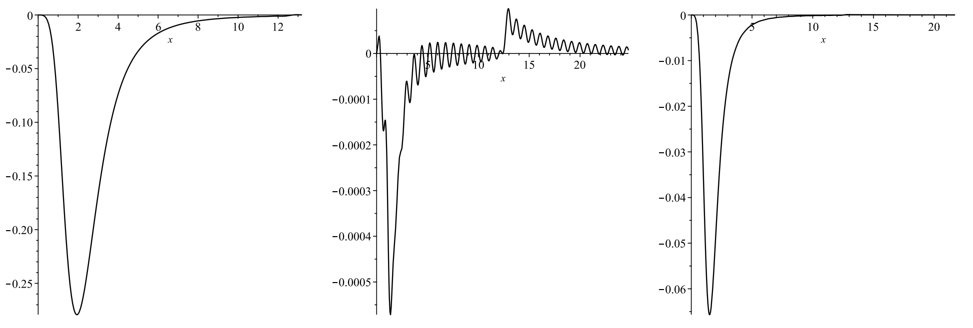


FIG. 1. Graphs of $err(x) = f(x) - C_{32}\{f, h\}(x)$ from Example 1 for h calculated by (9) – left, (16) – center, (19) – right

Predictably the value of h calculated by (16) is superior to those calculated by (19) and (9). One can see a discernible bump in the error function at $x_0 = N_6 h \approx 12.3792$. The values of $err(x)$ on the left of x_0 corresponds to the discretization error, whilst the values on the right of x_0 corresponds to the truncation error. The magnitude of those errors are almost match. This highlight the fact that the chosen h is really close to theoretically optimal value (16).

Example 2. In this example we set $f(x) \in H^1(D_d)$ as
$$f(x) = \frac{6 \cos 2x}{(5 + \cos^2 x)(1 + x^4)},$$
and choose formula (9) for the evaluation of h. The function $f(x)$ is meromorphic and bounded in D_d for any d smaller than the imaginary part of zeros of $\left(5 + \cos^2 x\right)\left(1 + x^4\right)$. The zeros of the polynomial part of this expression lie closer to the real line than any zero of $5 + \cos^2 x$, so $d \leq \Im \sqrt[4]{-1} = \frac{\sqrt{2}}{2} \approx .707106781186550$. Therefore it is safe to set $d = 0.7$. For given $f(x)$ we can also explicitly find the parameters of algebraic decay bound (7): $L = f(0) = 1$, $\alpha = 4$.

Note, that for a more general function $f(x)$ the corresponding L, α can be calculated numerically from a sequence of its values. For explicitly given $f(x)$ the possible values of d can be calculated numerically as well, for example using `Analytic` routine from Maple [4].

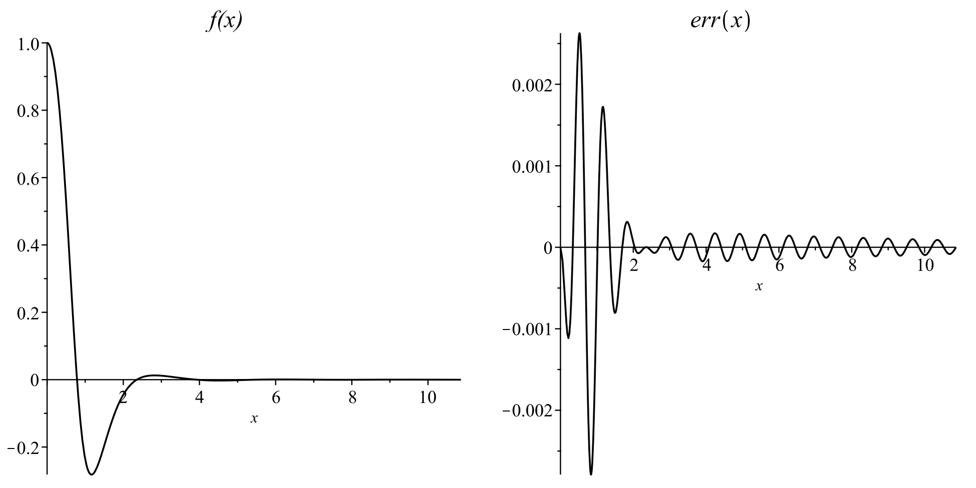


FIG. 2. Graphs of $f(x)$ and $err(x) = f(x) - C_{32}\{f,h\}(x)$ from Example 2

The graphs of the approximated function $f(x)$ and the error of its interpolation by $C_{32}\{f,h\}(x)$ are given in Fig. 2.

The precise values of err_i for $i = 1,\ldots,11$ are presented in Table 1. Here we additionally supply the theoretical estimate $\mathcal{E}_{N_i}$ defined in Theorem 1 and the value of $c_i = err_i/\mathcal{E}_{N_i}$.

TABL. 1. Result of the numerical experiments for $f(x)$ from Example 2. The step size h is calculated by (9), the quantities $\mathcal{E}_N$ and c are evaluated with help of (8)

i	N_i	err_i	$\mathcal{E}_{N_i}$	c_i
1	1	0.164468448	0.04709645766	3.49216175
2	2	0.06868780928	0.02952007611	2.326816808
3	4	0.05758701686	0.01520376206	3.787682064
4	8	0.03584624921	0.006430513883	5.574398852
5	16	0.0096295153	0.002280722496	4.222133695
6	32	0.00277964663	0.0006985817398	3.978985524
7	64	0.001039781276	0.0001901179719	5.469137218
8	128	0.0001265620194	4.706647235E-05	2.689005848
9	256	6.005526369E-05	1.079496434E-05	5.563266519
10	512	5.048493593E-06	2.325942889E-06	2.170514855
11	1024	2.594213457E-06	4.758456168E-07	5.451796476

The data from in Table 1 demonstrates that the approximation method presented by Theorem 1 converges to $f(x)$. The of observed approximation error is consistent with the estimate provided by (8). Moreover the estimated value of c from (8) remains bounded by 5.6 for all $i = \overline{1,6}$. All this prove the effectiveness of the developed method.

Bibliography

1. Bateman H. Higher transcendental functions. 1983 edition / H. Bateman. – Bombay: Tata-mcgrawhill Book Company Ltd., 1953. – Vol. 1.
2. Butzer P. L. Basic Relations Valid for the Bernstein Space B^p_σ and Their Extensions to Functions from Larger Spaces with Error Estimates in Terms of Their Distances from B^p_σ / P. L. Butzer, G. Schmeisser, R. L. Stens // Journal of Fourier Analysis and Applications. – 2013. – Apr. – Vol. 19, № 2. – P. 333-375.
3. Ferreira E. New properties of the lerch's transcendent / E. Ferreira, A. Kohara, J. Sesma // Journal of Number Theory. – 2017. – Vol. 172. – P. 21-31. Access mode: http://www.sciencedirect.com/science/article/pii/S0022314X16302219
4. Maplesoft maple, Maplesoft, a division of Waterloo Maple Inc. – Waterloo, Ontario, 2016. Access mode: https://www.maplesoft.com/products/Maple/
5. Corless R. M. On the LambertW function / R. M. Corless, G. H. Gonnet, D. E. G. Hare et al. // Advances in Computational Mathematics. – 1996. – Vol. 5, № 1. – P. 329-359. Access mode: http://dx.doi.org/10.1007/BF02124750
6. Stenger F. Numerical methods based on Sinc and analytic functions / F. Stenger. – Berlin, Heidelberg, London, etc.: Springer-Verlag, 1993. – Vol. 20 of Springer Series in Computational Mathematics. P. xv + 565. ISBN: 0-387-94008-1 (New-York), 3-540-94008-1 (Berlin) 0-387-94008-1 (New York), 3-540-94008-1 (Berlin).

Dmytro Sytnyk,
Institute of Mathematics, NAS of Ukraine,
3, Tereschenkivska Str., Kyiv, 01004, Ukraine.

Received 26.10.2017; revised 16.11.2017

Taras Shevchenko National University of Kyiv
Journal of Numerical and Applied
2017 MATHEMATICS 3 (126)
Founded in 1965

CONTENTS