Applications of Mathematics

Josef Dalík; Helena Růžičková

An explicit modified method of characteristics for the one-dimensional nonstationary convection-diffusion problem with dominating convection

Applications of Mathematics, Vol. 40 (1995), No. 5, 367–380

Persistent URL: http://dml.cz/dmlcz/134300

Terms of use:

© Institute of Mathematics AS CR, 1995

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.

This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* http://dml.cz

AN EXPLICIT MODIFIED METHOD OF CHARACTERISTICS FOR THE ONE-DIMENSIONAL NONSTATIONARY CONVECTION-DIFFUSION PROBLEM WITH DOMINATING CONVECTION

Josef Dalík, Brno, and Helena Růžičková, Brno

(Received December 8, 1993)

Summary. We describe a numerical method for the equation $u_t + pu_x - \varepsilon u_{xx} = f$ in $(0,1) \times (0,T)$ with Dirichlet boundary and initial conditions which is a combination of the method of characteristics and the finite-difference method. We prove both an a priori local error-estimate of a high order and stability. Example 3.3 indicates that our approximate solutions are disturbed only by a minimal amount of the artificial diffusion.

Keywords: method of characteristics, finite differences, convection-diffusion problem, local error-estimate, stability

AMS classification: 65M06, 65M25, 65M12

Introduction

We will consider the problem to find $u = u(x,t)$ such that

$$\begin{aligned} &u_t + pu_x - \varepsilon u_{xx} = f \quad \text{in} \quad \mathrm{Q} = (0,1) \times (0,T), \\ &u(x,0) = u_0(x) \quad \text{in} \quad (0,1) \quad \text{and} \\ &u(0,t) = \varphi(t), u(1,t) = \psi(t) \quad \text{in} \quad \langle 0,T). \end{aligned} \tag{1}$$

Here the function p is continuous and its partial derivative p_x is bounded in $\mathbb{R} \times \langle 0,T\rangle$, $f \in \mathbb{C}(\mathrm{Q})$, $u_0 \in \mathbb{C}^2\,\langle 0,1\rangle$, φ, $\psi \in \mathbb{C}^1\,\langle 0,T\rangle$, the functions p, f, u_0'', φ', ψ' are hölderian and compatibility conditions of order one in the sense of Ladyzhenskaya, Solonnikov, Uraltseva [7] are satisfied at the points $[0,0]$, $[1,0]$. We remark that the closure $\bar{\mathrm{Q}}$ is the union of Q and its boundary Γ.

Problem (1) is a model of various diffusion processes in flowing media. According to [7], Chap. IV, Th. 5.2, it has a unique classical solution u and its partial derivatives up to any given order are continuous under correspondingly stronger assumptions. If convection dominates (i.e. $\varepsilon \ll 1 + |p|$) then it is typical that there exist narrow strips in Q, called *boundary* or *internal layers*, in which the gradient of u is extremely large.

In the paper Douglas, Russell [6], combinations of the method of characteristics and the finite-difference method appeared which approximate values of u at a time t_j by values of u at a suitable time $t_i < t_j$. Such methods are called modified methods of characteristics (MMOC) in Noye [9], Allen & Khosravani [1], Bugai [4] and in other papers.

In this work we apply the ideas from Dalík [5] to the approximate solution of (1). As a result, we obtain an explicit MMOC in which the time t_i is chosen optimally so that the approximate solutions neither oscillate nor contain any visible amount of the so-called artificial diffusion. Especially, the breadths of approximations of the parabolic layers are in a good agreement with the breadths of their exact patterns. Our method is applicable under the condition $3\varepsilon h_t \leqslant h_x^2$, valid for the discretization steps h_x and h_t. For some special subregions $\mathsf{A} \subseteq \mathsf{Q}$, we prove that the error $u(a) - u_a$ at nodes $a \in \mathsf{A}$ is proportional to $h_x^2 + h_t^4$. A comparable estimate has been presented in the paper Tourigny, Süli [11].

We express by $|u(a) - u_a| = O(h_x^2 + h_t^4)$ the fact that $|u(a) - u_a| \leqslant C(h_x^2 + h_t^4)$. Constants like C do never depend on ε, h_x, h_t, but, in general, they do depend on partial derivatives of u up to the order four and also on the time T. We denote by $\emptyset$ the empty set, by $\overline{ab}$ the segment connecting points $a, b \in \bar{\mathsf{Q}}$ and use the symbol $\neg(c)$ for the negation of a condition (c).

2. Discretization, part one

Let n, k be positive integers. We put

$$\begin{aligned} h_x &= \frac{1}{n+1},\ x_m = m h_x \quad \text{for} \quad m = 0, 1, \dots, n+1, \\ h_t &= \frac{T}{k}, t_j = j h_t \quad for \quad j \in \langle 0, k \rangle\,, \quad \text{and} \\ Q_h &= \{[x_m, t_j];\, m = 1, \dots, n, j = 1, \dots, k\}. \end{aligned}$$

To each node $a \in Q_h$, we relate one *equation for* an approximation u_a of the exact value $u(a)$.

Let $a = [x_m, t_j] \in Q_h$ be fixed. We choose a real $i \in \langle 0, j)$ (the algorithm of this choice will be described in Sect. 3) and define a function $\tilde{x}(t)$ as a unique solution of the initial-value problem

$$\frac{\mathrm{d}\tilde{x}}{\mathrm{d}t} = p(\tilde{x}, t) \quad \text{for} \quad t \in \langle t_i, t_j \rangle, \ \tilde{x}(t_j) = x_m. \tag{2}$$

We put

$$\mathsf{C}_{a,i} = \{[\tilde{x}(t), t];\ t_i \leqslant t \leqslant t_j\}$$

and suppose that $\mathsf{C}_{a,i} \subset \bar{\mathsf{Q}}$. Then $u_t + pu_x = \frac{\mathrm{d}u}{\mathrm{d}t}$ and the equation (1) acquires the form

$$\frac{\mathrm{d}u}{\mathrm{d}t} - \varepsilon u_{xx} = f \tag{3}$$

at points from $\mathsf{C}_{a,i}$. If we put $z = [\tilde{x}(t_i), t_i]$ and integrate (3) along $\langle t_i, t_j \rangle$ then we obtain

$$u(a) - u(z) - \varepsilon \int_{t_i}^{t_j} u_{xx}(\tilde{x}, t)\,\mathrm{d}t = \int_{t_i}^{t_j} f(\tilde{x}, t)\,\mathrm{d}t. \tag{4}$$

We approximate the integrals

$$\int_{t_i}^{t_j} f\,\mathrm{d}t \tag{5}$$

by the Simpson rule with step $h_t/2$ and denote by I_f the resulting value,

$$\int_{t_i}^{t_j} u_{xx}\,\mathrm{d}t \tag{6}$$

by the value $(t_j - t_i)u_{xx}(z)$.

Remark. The function $\tilde{x}(t)$ has to be approximated by a numerical solution of (2). Because of (5), it is necessary to compute approximations $\tilde{x}_\iota$ of $\tilde{x}(t_\iota)$ for $\iota = j, j-0.5, \ldots, i$. If the requirement

$$\tilde{x}(t_\iota) = \tilde{x}_\iota + (t_j - t_\iota)O(h_t^4) \tag{7}$$

is satisfied then the error of this approximation does not decrease the order of the resulting errors at nodes. For example the classical Runge-Kutta method with step $-h_t/2$ gives an approximation satisfying (7) for a sufficiently large class of functions p—see Lambert [8]. For the sake of formal simplicity, we neglect the errors of the approximations $\tilde{x}_\iota$.

If we insert the approximations (5), (6) into (4) then we obtain

$$u(a) - u(z) - \varepsilon(t_j - t_i)u_{xx}(z) = I_f + e_t \tag{8}$$

and well-known error-estimates, see for example Beresin, Shidkov [2], give us the following estimate of the error e_t.

Lemma 2.1. *If $f \in \mathbb{C}^4(\mathsf{C}_{a,i})$*[1] *and $u_{xx} \in \mathbb{C}^1(\mathsf{C}_{a,i})$ then*

$$e_t = (t_j - t_i)O[(t_j - t_i)\varepsilon + h_t^4].$$

3. Discretization, part two

If we determine the value of i and approximate $u(z)$, $u_{xx}(z)$ in (8) then we obtain the equation for u_a, whose general form is

$$u_a - \sum_{b\in R(a)} \gamma(ab)u_b = \varrho_a. \tag{9}$$

A number e_a is called an *error of the equation for u_a* whenever

$$u(a) - \sum_{b\in R(a)} \gamma(ab)u(b) = \varrho_a + e_a. \tag{10}$$

The set $R(a)$ and the coefficients $\gamma(ab)$, ϱ_a will be defined for each of the following schemes I, II, III separately.

Let us consider the conditions

$$\varepsilon h_t/h_x^2 \leqslant \frac{1}{3}, \tag{c1}$$

$$\frac{1}{6} \leqslant \varepsilon j h_t/h_x^2. \tag{c2}$$

If (c1) and (c2) are true then there exist integers ι such that

$$1 \leqslant \iota \leqslant j \quad \text{and} \quad \frac{1}{6} \leqslant \varepsilon\iota h_t/h_x^2 \leqslant \frac{1}{3}.$$

[1] $g \in \mathbb{C}^\iota(\mathsf{C}_{a,i})$ means $g(\tilde{x}(t), t) \in \mathbb{C}^\iota\left\langle t_i, t_j\right\rangle$ for any $g = g(x,t)$ and $\iota = 0, 1, \ldots$

We denote by d the largest of these ι, put

$$v = \varepsilon d h_t / h_x^2$$

and formulate the condition

$$\text{(c3)} \qquad \mathsf{C}_{a,j-d} \subset \bar{\mathsf{Q}}.$$

In case of (c1), (c2) and (c3) we put $i = j - d$, denote by l the index from $\{0, 1, \ldots, n\}$ which satisfies

$$x_l \leqslant \tilde{x}(t_i) < x_{l+1}$$

(if $l = n$ then we admit $\bar{x}(t_i) = x_{l+1}$), define

$$\alpha = (x_{l+1} - \tilde{x}(t_i))/h_x, \quad \beta = (\tilde{x}(t_i) - x_l)/h_x$$

and formulate the condition

$$\text{(c4)} \qquad 0 < l < n.$$

Scheme I. If (c1), (c2), (c3) and (c4) then we put

$$\begin{aligned}
b_\iota &= [x_{\iota+l-1}, t_i] \quad \text{for} \quad \iota = 0, 1, 2, 3, \\
R(a) &= \{b_0, b_1, b_2, b_3\} - \Gamma, \quad S(a) = \{b_0, b_1, b_2, b_3\} \cap \Gamma, \\
\gamma(ab_0) &= \alpha[v - (1 - \alpha^2)/6], \\
\gamma(ab_1) &= (1 + \alpha)\alpha(1 + \beta)/2 - (2\alpha - \beta)v, \\
\gamma(ab_2) &= (1 + \beta)\beta(1 + \alpha)/2 - (2\beta - \alpha)v, \\
\gamma(ab_3) &= \beta[v - (1 - \beta^2)/6], \\
\varrho_a &= I_f + \sum_{b \in S(a)} \gamma(ab)u(b) \quad \text{and} \\
\mathsf{C}_a &= \mathsf{C}_{a,i}, \quad \mathsf{L}_a = \overline{b_0 b_3}.
\end{aligned}$$

Lemma 3.1. *If* (c1), (c2), (c3), (c4), $u \in \mathbb{C}^4(\mathsf{L}_a)$, $u_{xx} \in \mathbb{C}^1(\mathsf{C}_a)$ *and* $f \in \mathbb{C}^4(\mathsf{C}_a)$ *then*

$$e_a = \mathrm{d} h_t O(h_x^2 + h_t^4).$$

Proof. Let P be the 3rd-degree polynomial satisfying $P(x_\iota) = u(x_\iota, t_i)$ for $\iota = l - 1$, l, $l + 1$, $l + 2$. If we approximate the value $u(z)$ and $u_{xx}(z)$ in (8) by $P(\tilde{x}(t_i))$ and $P''(\tilde{x}(t_i))$, respectively, then we obtain the identity

$$u(a) - \sum_{b \in R(a)} \gamma(ab)u(b) = \varrho_a + e_t + e_x.$$

Using the well-known form of the remainder term $u(x,t_i) - P(x)$, one can see that

$$e_x = O(\varepsilon\,\mathrm{d}h_t h_x^2 + h_x^4).$$

This result, Lemma 2.1 and the inequalities $\frac{1}{6} \leqslant v \leqslant \frac{1}{3}$ yield the above estimate of $e_a = e_t + e_x$. □

We omit the simple proof of the next statement.

Lemma 3.2. *If* (c1), (c2), (c3) *and* (c4) *then the following assertions* (a), (b) *are true.*

(a) $\gamma(ab) \geqslant 0$ *for all* $b \in R(a) \cup S(a)$.

(b) $\sum\limits_{b\in R(a)} \gamma(ab) \leqslant 1$.

Motivation. We have chosen the index i in such a way that the following requirements (a), (b) are satisfied.

(a) *The value of v and equivalently of d is as large as possible*: Then the distance between b_ι and C_a is small in comparison to the difference $t_j - t_i = \mathrm{d}h_t$ for $\iota = 0,1,2,3$. Consequently, the value u_a depends on the approximate values of u at the nodes from a narrow strip along $\mathsf{C}_{a,0}$ only. In Fig. 1, the strip is schematically indicated by dotted curves.

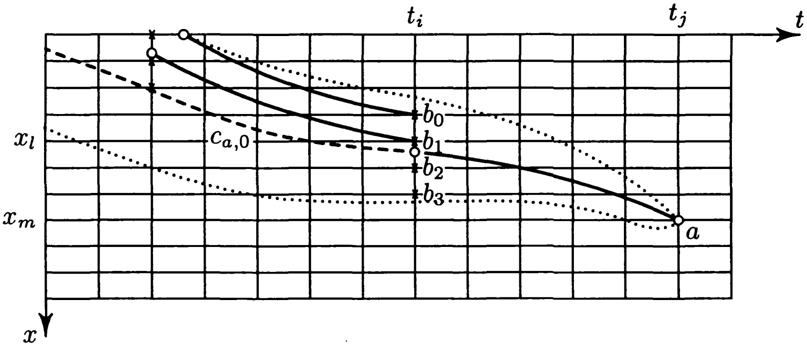

Fig. 1

This property, shared also by Schemes II, III, enables us to obtain rather narrow approximations of parabolic layers for problems with dominating convection. Some more exact information can be found in [5].

(b) $\gamma(ab_\iota) \geqslant 0$ for $\iota = 0,1,2,3$ *and the approximations of the layers are smooth*: Nonnegativity of γ is a precondition for the stability of approximate solutions.

Nonetheless, these essential inequalities are valid under weaker conditions. Namely, we have

$$\frac{1}{6} \leqslant v \leqslant \frac{1}{2} \Rightarrow \gamma(ab_\iota) \geqslant 0 \quad \text{for} \quad \iota = 0,1,2,3.$$

Hence instead of $v \leqslant \frac{1}{3}$ we could require $v \leqslant \frac{1}{2}$ which would better satisfy (a). We do not recommend this change, because approximations of layers computed with values of v near to $\frac{1}{2}$ are not smooth enough. This fact can be observed in the following example.

Example 3.3. The problem

$$\begin{aligned} u_t - 0.01u_{xx} &= 0 \quad \text{in} \quad (0,1)\times(0,2),\\ u(x,0) &= \mathrm{e}^{-(10x-5)^2} \quad \text{in} \quad (0,1),\\ u(0,t) &= \frac{1}{\sqrt{4t+1}}\mathrm{e}^{-\frac{25}{4t+1}} = u(1,t) \quad \text{in} \quad \langle 0,2) \end{aligned}$$

has an exact solution $u(x,t) = \frac{1}{\sqrt{4t+1}}\mathrm{e}^{-\frac{(10x-5)^2}{4t+1}}$. It has been solved by Scheme I with steps

(a) $h_x = 0.1$ and $h_t = 0.25$. Then $v = \frac{1}{4}$ and $u_a - \frac{1}{4}(u_{b_0} + 2u_{b_1} + u_{b_2}) = 0$.

(b) $h_x = 0.1$ and $h_t = 0.5$ by admitting $v \leqslant \frac{1}{2}$. Then $v = \frac{1}{2}$ and $u_a - \frac{1}{2}(u_{b_0} + u_{b_2}) = 0$.

(If $a = [x_m, t_j]$ then $b_\iota = [x_{\iota+m-1}, t_{j-1}]$ in (a) and (b).) In Fig. 2, we compare linear splines with nodes $x = 0, 0.1, \ldots, 1$, related to

the exact values of u (full curves),

the approximate values of u obtained by (a) (dotted curves),

the approximate values of u obtained by (b) (dashed curves)

at the time-levels $t = 0.5, 1, 1.5$.

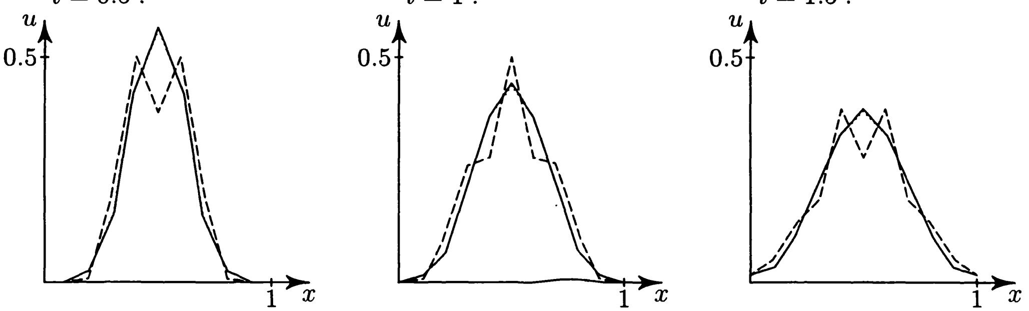

Fig. 2

Scheme II. If (c1), (c2), (c3) and $\neg$(c4) then either $l = 0$ or $l = n$. In case of $l = 0$ we put

$$b_\iota = [x_{\iota-1}, t_i] \quad \text{for} \quad \iota = 1, 2, 3,$$
$$R(a) = \{b_1, b_2, b_3\} - \Gamma, \quad S(a) = \{b_1, b_2, b_3\} \cap \Gamma,$$
$$\gamma(ab_1) = \alpha + \beta v, \quad \gamma(ab_2) = \beta(1 - 2v), \quad \gamma(ab_3) = \beta v,$$
$$\mathsf{C}_a = \mathsf{C}_{a,i}, \quad \mathsf{L}_a = \overline{b_1 b_3}$$

and in case of $l = n$ we put

$$b_\iota = [x_{\iota+n-1}, t_i] \quad \text{for} \quad \iota = 0, 1, 2,$$
$$R(a) = \{b_o, b_1, b_2\} - \Gamma, \quad S(a) = \{b_0, b_1, b_2\} \cap \Gamma,$$
$$\gamma(ab_0) = \alpha v, \ \gamma(ab_1) = \alpha(1 - 2v), \ \gamma(ab_2) =' \beta + \alpha v,$$
$$\mathsf{C}_a = \mathsf{C}_{a,i}, \quad \mathsf{L}_a = \overline{b_0 b_2}.$$

Lemma 3.4. *If* (c1), (c2), (c3), $\neg$(c4), $u \in \mathbb{C}^4(\mathsf{L}_a)$, $u_{xx} \in \mathbb{C}^1(\mathsf{C}_a)$ *and* $f \in \mathbb{C}^4(\mathsf{C}_a)$ *then*

$$e_a = O(h_x^2) + \mathrm{d}h_t 0(h_x^2 + h_t^4).$$

Sketch of proof. In case of $l = 0$ we obtain Scheme II by inserting $\alpha u(x_0, t_i) + \beta u(x_1, t_i)$ instead of $u(z)$ and $\beta v[u(x_0, t_i) - 2u(x_1, t_i) + u(x_2, t_i)]$ instead of $\varepsilon(t_j - t_i) u_{xx}(z)$ into (8). □

Lemma 3.5. *If* (c1), (c2), (c3) *and* $\neg$(c4) *then the following assertions* (a), (b) *are true.*

(a) $\gamma(ab) \geqslant 0$ *for all* $b \in R(a) \cup S(a)$.

(b) $\sum\limits_{b \in R(a)} \gamma(ab) \leqslant \frac{5}{6}$.

Scheme III. If (c1) and $\neg$[(c2) and (c3)] then we put

$$i = \inf\{s \in \mathbb{R};\ \mathsf{C}_{a,s} \subset \bar{\mathsf{Q}}\},$$
$$R(a) = \emptyset,$$
$$\varrho_a = I_f + u(\tilde{x}(t_i), t_i),$$
$$\mathsf{C}_a = \mathsf{C}_{a,i} \ \text{ and } \ \mathsf{L}_a = \emptyset.$$

Lemma 3.6. *If* (c1), $\neg$[(c2) and (c3)], u_{xx} *is bounded on* C_a *and* $f \in \mathbb{C}^4(\mathsf{C}_a)$ *then*

$$e_a = O(h_x^2) + \mathrm{d}h_t O(h_t^4).$$

Remark. Of course, the value $u(\tilde{x}(t_i), t_i)$ cannot be computed exactly. Lemma 3.6 prescribes the minimal order of the admissible error.

4. A priori local error-estimate and stability

We formulate the main results (Theorems 4.1 and 4.3) in this section and prove them in the next one.

Let $\mathsf{A} \subset \bar{\mathsf{Q}}$ be arbitrary. We put

$$A_h = \mathsf{A} \cap Q_h.$$

and call A *hereditary* if

$$a \in A_h \Rightarrow \mathsf{C}_a \cup \mathsf{L}_a \subseteq \mathsf{A}.$$

Theorem 4.1. *Let $\mathsf{A} \subseteq \bar{\mathsf{Q}}$ be hereditary and $u \in \mathbb{C}^4(\mathsf{L}_a), u_{xx} \in \mathbb{C}^1(\mathsf{C}_a), f \in \mathbb{C}^4(\mathsf{C}_a)$ for all $a \in A_h$. Then*

$$\max_{a \in A_h} |u(a) - u_a| = O(h_x^2 + h_t^4).$$

We illustrate this order of convergence by the following example.

Example 4.2. It is easy to see that the problem

$$\begin{gathered} u_t + (1 - 0.5x)u_x - 0.01u_{xx} = -1.5x^3 + 3x^2 - 0.06x \quad \text{in} \quad (0,1) \times (0,\infty), \\ u_0(x) = 2x^2 + (\mathrm{e}^{0.02(1-x)} - 1)/(\mathrm{e}^{0.02} - 1) \quad \text{in} \quad (0,1), \\ u(0,t) = 1,\, u(1,t) = 2 \quad \text{in} \quad \langle 0,\infty) \end{gathered}$$

has an exact stationary solution $u_{st}(x) = 1 + x^3$. For time-levels near to 4, the approximate solution does not depend on t any more. Therefore the values $u_{[x_m,4]}$ are approximations of $u_{st}(x_m)$ for $m = 1, \ldots, n$. In Tab. 1, we compare maximal relative errors of these approximations obtained by various discretization steps.

h_x	h_t	maximal relative error at $t = 4$
0.2	0.2	1.759 %
0.1	0.138	0.44 %
0.05	0.1	0.08 %

Tab. 1

Remark. In the estimate from Theorem 4.1, the coefficient at $h_x^2 + h_t^4$ depends on the values of partial derivatives of u up to the order four at various points from A. If layers appear in A then these values and, consequently, the coefficient may be extremely large. Hence Theorem 4.1 is valuable only if no layers intersect A.

According to Raviart [10], an approximate solution of the problem (1) does not oscillate provided the matrix M of the resulting system of equations is *monotone* (i.e. M^{-1} does exist and is non-negative).

Theorem 4.3. *The matrix of the system of equations for u_a, $a \in Q_h$, is monotone.*

The authors have devoted much effort to an application of this method to a two-dimensional analogue of the problem (1).

5. Proof of theorems 4.1, 4.3

We first express the errors $u(a) - u_a$ as linear combinations of errors of equations.

We denote by N^I and N^{II}, N^{III} the set of nodes a such that the equation for u_a corresponds to Scheme I and II, III, respectively. Then we have

$$\sum_{b \in R(a)} \gamma(ab) \left\{ \begin{array}{c} \leqslant 1 \\ \leqslant \frac{5}{6} \\ = 0 \end{array} \right\} \quad \text{whenever} \quad \left\{ \begin{array}{c} a \in N^I \\ a \in N^{II} \\ a \in N^{III} \end{array} \right\} \tag{11}$$

and all coefficients $\gamma(ab)$ are non-negative according to Lemmas 3.2, 3.5.

We say that a sequence $r = a_1 a_2 \ldots a_\mu$ is a *string (of nodes)* whenever $\mu \geqslant 1$, $a_1 \in Q_h$ and $a_\iota \in R(a_{\iota-1})$ for $\iota = 2, \ldots, \mu$. Then we put $\bar{r} = a_1$, $\underline{r} = a_\mu$, $|r| = \mu$ and call $|r|$ the *length* of r. If $s = b_1 b_2 \ldots b_\nu$ is another string and $b_1 \in R(a_\mu)$ then we denote by rs the string $a_1 a_2 \ldots a_\mu b_1 \ldots b_\nu$. We call r the *initial substring* in rs and write $r \prec rs$.

We relate the integer $\mu_a = 1 + [j/d]$ to each node $a = [x_m, t_j]$.

Lemma 5.1. *Let r be a string such that $\bar{r} = a$. Then $|r| \leqslant \mu_a$.*

Proof. If we denote $a = [x_m, t_j]$, $r = a_1 a_2 \ldots a_\mu$ and $a_\iota = [x_{m(\iota)}, t_{j(\iota)}]$ for $\iota = 1, \ldots, \mu$ then $t_{j(\iota)} = t_{j(\iota+1)} + \mathrm{d}h_t$ for $\iota = 1, \ldots, \mu - 1$ and we obtain

$$jh_t = t_j = t_{j(1)} = t_{j(\mu)} + (\mu - 1)\,\mathrm{d}h_t \geqslant (\mu - 1)\,\mathrm{d}h_t.$$

□

Let $a \in Q_h$. We put

$$R_\iota(a) = \{r\,;\, \bar{r} = a \quad \text{and} \quad |r| = \iota\} \quad \text{for} \quad \iota = 1, \ldots, \mu_a$$

and

$$R^+(a) = R_1(a) \cup \ldots \cup R_{\mu_a}(a).$$

We say that a set S of strings is an *antichain (with a root a)* whenever $a \in Q_h$, $S \subseteq R^+(a)$ and $r \not\prec s$ for any $r, s \in S$.

Obviously we have $R_1(a) = \{a\}$, $R_2(a) = \{ab;\, b \in R(a)\}$ and $R^+(a) = \{r\,;\, \bar{r} = a\}$ by Lemma 5.1.

To each string $r = a_1 a_2 \dots a_\mu$, we relate the value

$$\gamma^*(r) = \begin{cases} 1 & \text{in the case} \quad |r| = 1 \quad \text{and} \\ \gamma(a_1a_2)\gamma(a_2a_3)\dots\gamma(a_{\mu-1}a_\mu) & \text{otherwise.} \end{cases}$$

We will take advantage of the fact that

$$\gamma^*(r) = \gamma^*(s)\gamma^*(\underline{s}t) = \gamma^*(s\underline{t})\gamma^*(t)$$

for all strings r, s, t satisfying $r = st$.

Lemma 5.2. *Let S be an antichain with a root a. Then*

$$\sum_{r\in S} \gamma^*(r) \begin{cases} = 1 & \textit{if } S = \{a\}, \\ \leqslant \sum\limits_{b\in R(a)} \gamma(ab) & \textit{otherwise.} \end{cases}$$

Proof. Since this statement is obvious in the case $S \subseteq \{a\}$, we suppose that $S \not\subseteq \{a\}$. Then $|S| = \max\limits_{r\in S} |r| \geqslant 2$.

Step 1. Let $|S| = 2$. This, together with the fact that S is an antichain, gives $S \subseteq \{ab;\, b \in R(a)\}$ and we have

$$\sum_{r\in S} \gamma^*(r) \leqslant \sum_{b\in R(a)} \gamma(ab).$$

Step 2. Assume that $|S| > 2$ and the statement holds for all antichains T such that $|T| < |S|$. We put

$$S_b = \{s \in R^+(b)\,;\, as \in S\} \quad \text{for each} \quad b \in R(a).$$

Then S_b is an antichain with the root b and $|S_b| < |S|$. We conclude

$$\sum_{s\in S_b} \gamma^*(s) \leqslant \sum_{c\in R(b)} \gamma(bc) \leqslant 1$$

by our assumption and by (11). Hence

$$\begin{aligned} \sum_{r\in S} \gamma^*(r) &= \sum_{b\in R(a)} \sum_{s\in S_b} \gamma^*(as) \\ &= \sum_{b\in R(a)} \gamma(ab) \sum_{s\in S_b} \gamma^*(s) \leqslant \sum_{b\in R(a)} \gamma(ab). \end{aligned}$$

□

If S is an antichain with a root a such that $S \neq \{a\}$ then

$$(12)\qquad \sum_{r\in S} \gamma^*(r) \begin{Bmatrix} \leqslant 1 \\ \leqslant \frac{5}{6} \\ = 0 \end{Bmatrix} \quad \text{whenever} \quad \begin{Bmatrix} a \in N^I \\ a \in N^{II} \\ a \in N^{III} \end{Bmatrix}$$

by Lemma 5.2 and by (11).

The following basic relation between the nodal errors of approximation and the errors of equations has been proved in [5].

Lemma 5.3. *If $a \in Q_h$ then*

$$u(a) - u_a = \sum_{r\in R^+(a)} \gamma^*(r) e_{\underline{r}}.$$

Since the order of error $e_{\underline{r}}$ for $\underline{r} \in N^I$ is different from that for $\underline{r} \in N^{II} \cup N^{III}$ (see Lemmas 3.1, 3.4, 3.6), we put

$$R^J(a) = \{r \in R^+(a);\ \underline{r} \in N^J\}$$

and estimate each of the sums $\sum_{r\in R^J(a)} \gamma^*(r)$ for $J = I, II, III$ separately.

Lemma 5.4. *Let $a \in Q_h$ be arbitrary. Then the following assertions* (a)–(c) *are true.*

(a) $\sum\limits_{r\in R^I(a)} \gamma^*(r) \leqslant \mu_a.$

(b) $\sum\limits_{r\in R^{II}(a)} \gamma^*(r) \leqslant 6.$

(c) $\sum\limits_{r\in R^{III}(a)} \gamma^*(r) \leqslant 1.$

Proof of (a). As $R_\iota(a)$ is an antichain with root a, we have

$$\sum_{r\in R_\iota(a)} \gamma^*(r) \leqslant 1 \quad \text{for} \quad \iota = 1, \dots, \mu_a$$

by (12). Hence

$$\sum_{r\in R^I(a)} \gamma^*(r) \leqslant \sum_{r\in R^+(a)} \gamma^*(r) \leqslant \mu_a.$$

Proof of (b). For $\iota = 1, 2, \dots, \mu_a$, let $S_\iota(a)$ denote the set of those strings from $R^{II}(a)$ in which the nodes from N^{II} appear in exactly ι positions. Then, obviously,

$$(13)\qquad R^{II}(a) = S_1(a) \cup \ldots \cup S_{\mu_a}(a)$$

and $S_\iota(a)$ is an antichain for $\iota = 1, \ldots, \mu_a$. We prove the assertion

$$\sum_{r \in S_\iota(a)} \gamma^*(r) \leqslant \left(\frac{5}{6}\right)^{\iota-1} \tag{14}$$

for $\iota = 1, \ldots, \mu_a$ by induction:

Step 1. If $\iota = 1$ then (14) follows by (12).

Step 2. Assume that (14) is true for some $\iota < \mu_a$. Then

$$S_{\iota+1}(a) = \{st;\ s \in S_\iota(a) \quad \text{and} \quad \underline{s}t \in S_2(\underline{s})\}$$

and, at the same time,

$$\sum_{v \in S_2(\underline{s})} \gamma^*(v) \leqslant \sum_{b \in R(\underline{s})} \gamma(\underline{s}b) \leqslant \frac{5}{6}$$

for each $\underline{s} \in N^{II}$ according to (12). Hence

$$\sum_{r \in S_{\iota+1}(a)} \gamma^*(r) = \sum_{s \in S_{\iota}(a)} \gamma^*(s) \sum_{v \in S_2(\underline{s})} \gamma^*(v) \leqslant \frac{5}{6} \sum_{s \in S_\iota(a)} \gamma^*(s) \leqslant \left(\frac{5}{6}\right)^{\iota}.$$

The assertion (b) is an immediate consequence of (13) and (14).

Proof of (c). This statement follows by the fact that $R^{III}(a)$ is an antichain and by (12). □

Proof of Theorem 4.1. A consecutive application of Lemmas 5.3, 5.4, 3.1, 3.4 and 3.6 yields

$$|u(a) - u_a| \leqslant \sum_{J=I,II,III} \sum_{r \in R^J(a)} \gamma^*(r)|e_{\underline{r}}| = O(h_x^2 + h_t^4)$$

for each $a \in A_h$. □

Proof of Theorem 4.3. Let M be the matrix of the system of equations for u_a, $a \in Q_h$. Then the elements of M are

$$m_{ab} = \begin{cases} 1 & \text{in the case} \quad b = a, \\ -\gamma(ab) & \text{in the case} \quad b \in R(a), \\ 0 & \text{otherwise} \end{cases}$$

for all $a, b \in Q_h$. According to Bramble, Hubbard [3], M is monotone whenever the following conditions (a)–(c) are satisfied:

(a) $a \neq b \Rightarrow m_{ab} \leqslant 0$: This is true by Lemmas 3.2, 3.5.

(b) There is a nonempty set $I \subseteq Q_h$ such that

$$\sum_{b \in Q_h} m_{ab} > 0 \Leftrightarrow a \in I \quad \text{and} \quad \sum_{b \in Q_h} m_{ab} = 0 \Leftrightarrow a \notin I:$$

According to (11), this assertion holds for

$$I = \Big\{a \in Q_h;\ \sum_{b \in R(a)} \gamma(ab) < 1\Big\}$$

and $N^{II} \cup N^{III} \subseteq I$.

(c) For every $a \in Q_h$ there exist nodes $b \in I$ and $c_1, \ldots, c_q$ such that each of the numbers $m_{ac_1}, m_{c_1c_2}, \ldots, m_{c_qb}$ is non-zero: Let $r \in R^+(a)$ satisfy $r \not\prec s$ for all $s \in R^+(a)$. If $|r| = 1$ then $r = a \in I$ because $R(a) = \emptyset$. If $|r| > 1$ then (c) is obviously satisfied by nodes $a, c_1, \ldots, c_q, b$ such that $r = ac_1c_2 \ldots c_qb$. □

References

[1] *M.B. Allen, A. Khosravani*: Solute transport via alternating-direction collocation using the modified method of characteristics. Advances in Water Recources *15* (1992), 125–132.

[2] *I.S. Beresin, N.P. Shidkov*: Numerical methods I. Nauka, Moscow, 1966. (In Russian.)

[3] *J.H. Bramble, B.E. Hubbard*: New monotone type approximations for elliptic problems. Math. Comp. (1964), no. 18, 349–367.

[4] *D.A. Bugai*: Accuracy analysis of the eulerian-lagrangian numerical schemes for the convection-diffusion equation. Preprint.

[5] *J. Dalík*: A finite difference method for a two-dimensional convection-diffusion problem with dominating convection. Submitted to publication.

[6] *J. Dougals Jr., T.F. Russell*: Numerical methods for convection dominated diffusion problems based on combining the method of characteristics with finite elements or finite difference procedures. SIAM J. Numer. Anal. (1982), no. 19, 871–885.

[7] *O.A. Ladyzhenskaya, V.A. Solonnikov, N.N. Uraltseva*: Linear and quasilinear equations of parabolic type. Nauka, Moscow, 1967. (In Russian.)

[8] *J.D. Lambert*: Computational Methods in Ordinary Differential Equations. John Wiley & Sons, London, 1973.

[9] *J.B. Noye*: Finite-difference methods for solving the one-dimensional transport equation. Numerical modeling: Application to Marine Systems (J. Noye, ed.). Elsevier, North Holland, 1987, pp. 231–256.

[10] *P.A. Raviart*: Les méthodes d'élements finis en mécanique des fluides II. 3. Edditions Eyrolles, Paris, 1981.

[11] *Y. Tourigny, E. Süli*: The finite element method with nodes moving along the characteristics for convection-diffusion equations. Numer. Math. (1991), no. 59, 399–412.

Author's address: *Josef Dalík*, katedra matematiky a deskriptivní geometrie stavební fakulty VUT, Žižkova 17, 662 37 Brno, Czech Republic; *Helena Růžičková*, katedra matematiky a deskriptivní geometrie strojní fakulty VUT, Technická 2, 616 69 Brno, Czech Republic.