AUTHOR QUERY FORM

ELSEVIER	**Journal:** Journal of Computational and Applied Mathematics **Article Number:** 12330	**Please e-mail your responses and any corrections to:** **E-mail: corrections.esch@elsevier.river-valley.com**

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof. **It is crucial that you NOT make direct edits to the PDF using the editing tools as doing so could lead us to overlook your desired changes.** Rather, please request corrections by using the tools in the Comment pane to annotate the PDF and call out the changes you would like to see. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof.

Location in article	**Query / Remark: Click on the Q link to find the query's location in text** **Please insert your reply or correction at the corresponding line in the proof**
Q1	Your article is registered as belonging to the Special Issue/Collection entitled "Advanced NumericalMethod". If this is NOT correct and your article is a regular item or belongs to a different Special Issue please contact s.sekar@elsevier.com immediately prior to returning yo This is correct.
Q2	Please confirm that given names and surnames have been identified correctly and are presented in the desired order and please carefully verify the spelling of a The name is also correct.
Q3	Have we correctly interpreted the following funding source(s) and country names you cited in your article: Hungarian Scientific Research Fund? Funding source is OK.
	Please check this box or indicate your approval if you have no corrections to make to the PDF file

Thank you for your assistance.

The other modifications are also correct. Thank you very much for your effort.

ARTICLE IN PRESS

Journal of Computational and Applied Mathematics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

On some discrete qualitative properties of implicit finite difference solutions of nonlinear parabolic problems

Róbert Horváth

Department of Analysis, Budapest University of Technology and Economics and MTA-ELTE Numerical Analysis and Large Networks Research Group, Egry J. u. 1, H-1111 Budapest, Hungary

ARTICLE INFO

Article history:
Received 16 January 2019
Received in revised form 7 June 2019

Keywords:
Nonlinear parabolic problems
Finite difference method
Qualitative properties

ABSTRACT

In this paper we investigate two special qualitative properties of the finite difference solutions of one-dimensional nonlinear parabolic initial boundary value problems. The first property says that the number of the so-called L-level points, or specially the number of the zeros, of the solution must be non-increasing in time. The second property requires a similar property for the number of the local maximizers and minimizers. First we recall a theorem that guarantees the above properties for the solution of a special second order nonlinear parabolic problem. Then we generate the numerical solution with the implicit Euler finite difference method and show that the obtained numerical solution satisfies the discrete versions of the above properties without any requirements on the mesh parameters. We close the paper with some numerical tests.

1. Introduction

The qualitative properties of partial differential equations and their numerical solutions are under intensive research nowadays. Partial differential equations generally model some real-life phenomena of some applied sciences such as physics, chemistry, biology, etc. These phenomena possess some characteristic properties, so it is natural that similar properties are required also for their mathematical and numerical models. From the point of view of the applications, it is useful to show that the mathematical models satisfy the required properties, moreover, in the numerical models the qualitative properties are essentially guaranteed by restricting somehow the spatial mesh and the time step.

For initial boundary value problems of parabolic equations, which can be considered as equations that model for example heat conduction phenomena, the linear case has been investigated thoroughly in the literature. The most frequently investigated properties are the maximum–minimum principles, the non-negativity and non-positivity preservation properties and the maximum norm contractivity properties, e.g. [1–6]. Monotonicity and sign-stability properties were also discussed [7,8]. In the last decades the interest of the researchers turned to nonlinear problems, e.g. [9–15], where the qualitative properties can be guaranteed by much more complicated assumptions than in the linear case.

In this paper we concentrate on two special properties of one-dimensional nonlinear parabolic problems. The first property is the monotone change of the number of the level points of the solution function. To the author's knowledge, this property has not been investigated for the numerical solution of nonlinear problems yet. The special version of this property says that the number of the zeros of the solution function must be non-increasing in time. This property is similar to the sign-stability property that says that the number of the sign-changes of the solution function must be non-increasing in time. Let us notice, however, that these properties are not the same. For example, the function $x \mapsto |\sin x|$ has one sign-change on the interval $[0, 2\pi]$ but it has three zeros, namely at $x = 0$, $x = \pi$ and $x = 2\pi$. The second investigated

E-mail address: rhorvath@math.bme.hu.

https://doi.org/10.1016/j.cam.2019.06.046
0377-0427/© 2019 Published by Elsevier B.V.

property will be the monotone change of the number of the local maximizers and minimizers of the solution function. This property was already investigated in [15] but only for semi-linear problems and for the explicit Euler method. In this paper, we formulate new results and generalize the previous ones in the sense that we consider nonlinear equations and general time-dependent Dirichlet boundary conditions. Nevertheless, our result is restrictive in the sense that it is formulated only for the implicit Euler finite difference solution of the continuous problem.

The paper is structured as follows. In Section 2, we formulate the initial boundary value problem to be investigated and review some results from the literature that show the validity of the two investigated qualitative properties. In Section 3, the discrete equivalents of the continuous properties are defined and proven for the implicit Euler finite difference solution of the problem. In this Section 4, we verify our theoretic results on some numerical test problems.

2. The continuous problem and its qualitative properties

Let us introduce the notations $Q_T = (0,T) \times (0,1)$, $\bar{Q}_T = [0,T] \times [0,1]$ and $Q_{\bar{T}} = (0,T] \times (0,1)$, where T is a fixed positive number. Let us consider the nonlinear initial boundary value problem

$$\begin{aligned} u'_t &= r(t,x,u,u''_{xx}),\ (t,x) \in Q_{\bar{T}}, \\ u(0,x) &= u_0(x),\ x \in [0,1], \\ u(t,0) &= \nu_0(t),\ u(t,1) = \nu_1(t),\ t \in [0,T] \end{aligned} \tag{1}$$

for the unknown function $u : \bar{Q}_T \to \mathbb{R}$, $(t,x) \mapsto u(t,x)$, where r, ν_0, ν_1 and u_0 are suitable, sufficiently smooth functions that guarantee the unique solvability of the problem such that $u \in C(\bar{Q}_T) \cup C^{1,2}(Q_{\bar{T}})$. We always assume the compatibility conditions $u_0(0) = \nu_0(0)$ and $u_0(1) = \nu_1(0)$. In problems that are described by Eq. (1) (e.g. heat conduction, diffusion) the variables t and x denote the time and space variables, respectively. We will adopt this terminology in this paper.

Definition 1. Let $\phi : [a,b] \to \mathbb{R}$ be a continuous function defined on the real interval $[a,b]$. Let L be a fixed real number. If an interval $[\alpha,\beta] \subset [a,b]$ $(\beta \geq \alpha)$ satisfies the properties
(a) $\phi(y) = L$ for all $y \in [\alpha,\beta]$,
(b) there does not exist an interval $[\alpha',\beta'] \subset [a,b]$ such that $\beta' - \alpha' > \beta - \alpha$, $[\alpha,\beta] \subset [\alpha',\beta']$ and $\phi(y) = L$ for all $y \in [\alpha',\beta']$,
then the interval $[\alpha,\beta]$ is called the (generalized) L-level point of the function ϕ.

The number of the L-level points of a function ϕ can be any natural number or infinity and will be denoted by $\zeta^L_{\phi|_{[a,b]}}$. For the case $L = 0$ and $\alpha = \beta$, the L-level points give the usual zeros of the function. For example,

$$\zeta^0_{x\mapsto 0|_{[0,1]}} = 1,\ \zeta^0_{x\mapsto \sin x|_{[0,2\pi]}} = 3,\ \zeta^0_{x\mapsto -\sin^- x|_{[0,2\pi]}} = 2$$

($\sin^-$ denotes the usual negative part of the sin function).

Definition 2. Let $\phi : [a,b] \to \mathbb{R}$ be a continuous function defined on the real interval $[a,b]$. If an interval $[\alpha,\beta] \subset [a,b]$ $(\beta \geq \alpha)$ satisfies the properties
(a) there exists a real constant C such that $\phi(y) = C$ for all $y \in [\alpha,\beta]$,
(b) there does not exist an interval $[\alpha',\beta'] \subset [a,b]$ such that $\beta' - \alpha' > \beta - \alpha$, $[\alpha,\beta] \subset [\alpha',\beta']$ and $\phi(y) = C$ for all $y \in [\alpha',\beta']$,
(c) there exists an interval $[\alpha'',\beta''] \subset [a,b]$ such that $\alpha'' < \alpha$ $(\alpha'' = \alpha$ if $\alpha = a)$, $\beta'' > \beta$ $(\beta'' = \beta$ if $\beta = b)$ and $\phi(y) \leq (\geq) C$ for all $y \in [\alpha'',\beta'']$,
then the interval $[\alpha,\beta]$ is called the (generalized) local maximizer (minimizer) of the function ϕ.

The number of the local maximizers (minimizers) of a function ϕ can be any natural number or infinity and will be denoted by $\mu_{\phi|_{[a,b]}}$. For example, for the number of the local maximizers we have

$$\mu_{x\mapsto const|_{[0,1]}} = 1,\ \mu_{x\mapsto \sin x|_{[0,2\pi]}} = 2,\ \mu_{x\mapsto -\sin^- x|_{[0,2\pi]}} = 2.$$

It is easy to see that (in the above sense) all continuous functions possess at least a local maximizer and a minimizer. Between any two maximizers (minimizers) there is at least one minimizer (maximizer).

Let us introduce the following subsets of the set $\bar{Q}_T$: the parabolic boundary $\bar{Q}_T \setminus Q_{\bar{T}}$ will be denoted by Γ and the final time level set $\{T\} \times [0,1]$ by $\mathcal{T}$. Let v be an arbitrary continuous function defined on $\bar{Q}_T$. The number of the L-level points of v on the set $\mathcal{T}$ is defined as the number of the L-level points of the one-variable function $x \mapsto v(T,x)$ $(x \in [0,1])$ and will be denoted by $\zeta^L_{v|_{\mathcal{T}}}$. The number of the L-level points of v on the parabolic boundary is defined as the number of the L-level points of the one-variable function

$$y \mapsto \begin{cases} v(-y,0) & y \in [-T,0], \\ v(0,y) & y \in (0,1), \\ v(y-1,1) & y \in [1,1+T], \end{cases}$$

Please cite this article as: R. Horváth, On some discrete qualitative properties of implicit finite difference solutions of nonlinear parabolic problems, Journal of Computational and Applied Mathematics (2019), https://doi.org/10.1016/j.cam.2019.06.046.

where we think of the parabolic boundary Γ as an unbent one-dimensional interval, and it is denoted by $\zeta^L_{v|\Gamma}$. The numbers of the local maximizers and minimizers of v on $\mathcal{T}$ and Γ are defined similarly and are denoted by $\mu_{v|\mathcal{T}}$ and $\mu_{v|\Gamma}$, respectively.

Let us turn back to problem (1). The relation between the above defined quantities for the solution u of problem (1) was revealed in paper [16].

Theorem 1 ([16]). *Let us assume that the function $r : (0,T] \times (0,1) \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is continuous on its domain of definition. Furthermore, let us assume that*

(a) $r(t,x,z,0) = 0$ for all $t \in (0,T]$, $x \in (0,1)$ and $z \in \mathbb{R}$, and

(b) $r(t,x,z,w_1) \le r(t,x,z,w_2)$ for all $t \in (0,T]$, $x \in (0,1)$, $z \in \mathbb{R}$ and real numbers $w_1 \le w_2$ (r is non-decreasing in its fourth argument).

Then for any fixed real number L we have

$$\zeta^L_{u|\mathcal{T}} \le \zeta^L_{u|\Gamma}$$

for the L-level points, moreover,

$$\mu_{u|\mathcal{T}} \le \mu_{u|\Gamma}$$

is valid for the local maximizers (minimizers) of the solution of problem (1).

Remark 1. Conditions (a) and (b) restrict the choice for the function r essentially. For example, the linear right-hand side $u''_{xx} + u$ is not allowed because r does not satisfy requirement (a). However, functions in the form $r(t,x,z,w) = r_1(t,x,z)r_2(w)$, where r_2 is non-decreasing (non-increasing), $r_2(0) = 0$ and $r_1(t,x,z) \ge (\le)0$ satisfy the assumptions of the theorem. Thus, the theorem is valid, for example, for the right-hand sides

$$u''_{xx}, \quad t(u''_{xx})^3, \quad u^2\sinh(u''_{xx}).$$

Remark 2. In Nickel's original paper, the function r is allowed to depend also on the first derivative of the solution. For the discrete case, we have to leave out this dependence because the finite difference approximation of the first derivative of a differentiable function at a local extremizer is not necessarily zero. We generally do not know even the sign of the approximation.

In the next section we show that the numerical solution obtained by the implicit Euler method possesses the discrete versions of the above properties.

3. Qualitative properties of the implicit Euler finite difference solution

We construct the implicit Euler finite difference solution of problem (1). For the full discretization we define a space–time mesh as follows. We choose the discretization points $x_i = i\Delta x$ ($i = 0,\dots,n+1$), where n is a fixed natural number and $\Delta x = 1/(n+1)$. The discrete time levels are defined as $t_k = k\Delta t$ ($k = 0,\dots,n_T$) with $\Delta t = T/n_T$, where n_T is another fixed natural number. The approximation of the solution u (the so-called numerical solution) at the point $(k\Delta t, i\Delta x)$ will be denoted by u_i^k, and these values can be computed by solving the system of nonlinear algebraic equations

$$\begin{aligned}
&\frac{u_i^{k+1}-u_i^k}{\Delta t} = r\left(t_{k+1}, x_i, u_i^{k+1}, \frac{u_{i-1}^{k+1}-2u_i^{k+1}+u_{i+1}^{k+1}}{\Delta x^2}\right), \quad \begin{matrix} i = 1,\dots,n \\ k = 0,\dots,n_T-1 \end{matrix}, \\
&\qquad u_i^0 = u_0(x_i),\ i = 0,\dots,n+1, \\
&\qquad u_0^k = \nu_0(t_k),\ u_{n+1}^k = \nu_1(t_k),\ k = 0,\dots,n_T.
\end{aligned} \tag{2}$$

In order to formulate the discrete version of Theorem 1, we define the number of L-level points and the number of local maximizers (minimizers) of the numerical solutions as follows. We construct the bilinear two-dimensional interpolation function of the mesh function u_i^k on the set $\bar{Q}_T$. Let us denote this interpolation function by $p(t,x)$. Let us notice that albeit the obtained interpolation function is piecewise quadratic, it gives a piecewise linear interpolation function in both t and x variables. Thus the discrete equivalents of the values $\zeta^L_{u|\mathcal{T}}$, $\zeta^L_{u|\Gamma}$, $\mu_{u|\mathcal{T}}$ and $\mu_{u|\Gamma}$ can be defined as $\zeta^L_{p|\mathcal{T}}$, $\zeta^L_{p|\Gamma}$, $\mu_{p|\mathcal{T}}$ and $\mu_{p|\Gamma}$, respectively. To prove the discrete equivalent of Theorem 1 we have to show the relations $\zeta^L_{p|\mathcal{T}} \le \zeta^L_{p|\Gamma}$ and $\mu_{p|\mathcal{T}} \le \mu_{p|\Gamma}$. This is the goal of the remainder of this section of the paper.

We introduce the following terminology. For the sake of simplicity, we say that the value u_i^k (or the mesh point (t_k, x_i)) is L-positive (L-negative) if $u_i^k > (<)L$. We say that the sequence of the mesh points

$$(t_{n_T}, x_i), (t_{n_T}, x_{i+1}), (t_{n_T}, x_{i+2}), \dots, (t_{n_T}, x_j)$$

form an L-positive (L-negative) group of mesh points of $\mathcal{T}$ if all $u_i^{n_T}, u_{i+1}^{n_T}, \dots, u_j^{n_T}$ values are L-positive (L-negative), and if $i \ne 0$, then $u_{i-1}^{n_T} \le (\ge)L$, and if $j \ne n+1$, then $u_{j+1}^{n_T} \le (\ge)L$. These types of groups can be defined for all other time levels and also for the parabolic boundary Γ. In the last case, we think of Γ as an unbent one-dimensional interval.

Let us choose an L-positive (L-negative) group of mesh points $\mathcal{G}$ of $\mathcal{T}$. Let us denote the set of all L-positive (L-negative) mesh points (t_k, x_i) that can be connected with the points of $\mathcal{G}$ by a path (in graph theoretic sense) through adjacent L-positive (L-negative) mesh points by $B_\mathcal{G}$. The properties of these sets is discussed in the next lemma.

Lemma 1. *Let us suppose that the assumptions of Theorem 1 for the function r are fulfilled. Let L be a fixed real value. If the implicit Euler finite difference numerical solution of problem (1) produced by (2) exists, then the above constructed $B_\mathcal{G}$ sets have the properties:*

(i) $B_\mathcal{G} \cap \Gamma \neq \emptyset$ (that is $B_\mathcal{G}$ always reaches the parabolic boundary), and

(ii) for any two different (L-positive or L-negative) groups $\mathcal{G}_1$ and $\mathcal{G}_2$ we have $B_{\mathcal{G}_1} \cap B_{\mathcal{G}_2} = \emptyset$.

Proof. To show property (i) for L-positive groups (the proof for L-negative groups is similar), we assume the opposite of the statement, that is we assume that $B_\mathcal{G} \cap \Gamma = \emptyset$. We choose a point $(t_{k^\star}, x_{i^\star}) \in B_\mathcal{G}$ (due to the indirect assumption we have $k^\star > 0$ and $0 < i^\star < n+1$) that satisfies the property

$$u_{i^\star}^{k^\star} = \max_{(t_k,x_i)\in B_\mathcal{G}} \{u_i^k\}.$$

Because $u_{i^\star-1}^{k^\star} \leq u_{i^\star}^{k^\star}$ and $u_{i^\star+1}^{k^\star} \leq u_{i^\star}^{k^\star}$, it is valid that

$$\frac{u_{i^\star-1}^{k^\star} - 2u_{i^\star}^{k^\star} + u_{i^\star+1}^{k^\star}}{\Delta x^2} \leq 0$$

and because of the assumptions (a) and (b) in Theorem 1 we have

$$\begin{aligned} \frac{u_{i^\star}^{k^\star} - u_{i^\star}^{k^\star-1}}{\Delta t} &= r\left(t_{k^\star}, x_{i^\star}, u_{i^\star}^{k^\star}, \frac{u_{i^\star-1}^{k^\star} - 2u_{i^\star}^{k^\star} + u_{i^\star+1}^{k^\star}}{\Delta x^2}\right) \\ &\leq r\left(t_{k^\star}, x_{i^\star}, u_{i^\star}^{k^\star}, 0\right) = 0. \end{aligned}$$

This implies that $u_{i^\star}^{k^\star-1} \geq u_{i^\star}^{k^\star}$, which means that $(t_{k^\star-1}, x_{i^\star}) \in B_\mathcal{G}$ and $u_{i^\star}^{k^\star-1} = u_{i^\star}^{k^\star}$. Repeating the previous reasoning, we get that the points $(t_k, x_{i^\star})$ $(k = k^\star - 1, k^\star - 2, \ldots, 1, 0)$ are all in the set $B_\mathcal{G}$, that is $(t_0, x_{i^\star}) = (0, x_{i^\star})$ is in both $B_\mathcal{G}$ and Γ, which contradicts the indirect assumption.

The statement (ii) is trivial if one of the groups is L-positive and the other one is L-negative. So it is enough to show the property for two L-positive groups (for L-negative groups the proof is similar). We apply indirect proof again. Let us suppose that $\mathcal{G}_1$ and $\mathcal{G}_2$ are two different L-positive groups of $\mathcal{T}$ such that $B_{\mathcal{G}_1} \cap B_{\mathcal{G}_2} \neq \emptyset$. Then there is a shortest path γ (the length of the path is measured in temporal and spatial step sizes) through adjacent mesh points (t_k, x_i) from one of the points (t_{n_T}, x_{i_1}) of $\mathcal{G}_1$ to one of the points (t_{n_T}, x_{i_2}) of $\mathcal{G}_2$ such that u_i^k is L-positive at all mesh points in the path. This path separates all the mesh points into three mutually disjoint sets: the first set (denoted by γ_a) contains the inner mesh points of the domain bordered by γ and $\mathcal{T}$, and the points $(t_{n_T}, x_{i_1+1}), (t_{n_T}, x_{i_1+2}), \ldots, (t_{n_T}, x_{i_2-1})$. The second set is the set of the mesh points of γ, and the third set (denoted by γ_b) is the complement of the union of the previous two sets. It can be seen from the construction that a point from γ_a cannot be adjacent to a point from γ_b.

The set γ_a is not empty because there must be a mesh point between $\mathcal{G}_1$ and $\mathcal{G}_2$ in $\mathcal{T}$ with solution value not greater than L. We choose a point $(t_{k^{\star\star}}, x_{i^{\star\star}}) \in \gamma_a$ that satisfies the property

$$u_{i^{\star\star}}^{k^{\star\star}} = \min_{(t_k,x_i)\in\gamma_a} \{u_i^k\}.$$

We trivially have $u_{i^{\star\star}}^{k^{\star\star}} \leq L$. Because $u_{i^{\star\star}-1}^{k^{\star\star}} \geq u_{i^{\star\star}}^{k^{\star\star}}$ and $u_{i^{\star\star}+1}^{k^{\star\star}} \geq u_{i^{\star\star}}^{k^{\star\star}}$, it is valid that

$$\frac{u_{i^{\star\star}-1}^{k^{\star\star}} - 2u_{i^{\star\star}}^{k^{\star\star}} + u_{i^{\star\star}+1}^{k^{\star\star}}}{\Delta x^2} \geq 0$$

and because of the properties (a) and (b) of r in Theorem 1 we have

$$\begin{aligned} \frac{u_{i^{\star\star}}^{k^{\star\star}} - u_{i^{\star\star}}^{k^{\star\star}-1}}{\Delta t} &= r\left(t_{k^{\star\star}}, x_{i^{\star\star}}, u_{i^{\star\star}}^{k^{\star\star}}, \frac{u_{i^{\star\star}-1}^{k^{\star\star}} - 2u_{i^{\star\star}}^{k^{\star\star}} + u_{i^{\star\star}+1}^{k^{\star\star}}}{\Delta x^2}\right) \\ &\geq r\left(t_{k^{\star\star}}, x_{i^{\star\star}}, u_{i^{\star\star}}^{k^{\star\star}}, 0\right) = 0. \end{aligned}$$

This implies that $u_{i^{\star\star}}^{k^{\star\star}-1} \leq u_{i^{\star\star}}^{k^{\star\star}}$, which means that $(t_{k^{\star\star}-1}, x_{i^{\star\star}}) \in \gamma_a$ and $u_{i^{\star\star}}^{k^{\star\star}-1} = u_{i^{\star\star}}^{k^{\star\star}}$. Repeating the previous reasoning, we get that the points $(t_k, x_{i^{\star\star}})$ $(k = k^{\star\star} - 1, k^{\star\star} - 2, \ldots, 1, 0)$ are all in the set γ_a, which contradicts the fact that one of these points must belong to γ. ■

Remark 3. It can be seen from the proof of the lemma that the assumptions of the lemma are enough to guarantee the maximum (minimum) principle for the implicit Euler finite difference solution. That is, the maximum (minimum) value of the numerical solution u_i^k is taken also at some mesh point of the parabolic boundary Γ. Specially, if the maximum (minimum) is taken at an inner point $(t_{k'}, x_{i'}) \in Q_{\bar{T}}$ of the mesh, then we have $u_{i'}^{k'} = u_{i'}^{k'-1} = \cdots = u_{i'}^{0}$, that is the maximum (minimum) is taken also at the zeroth time level.

Based on this lemma we are ready to formulate the discrete version of Theorem 1.

Theorem 2. *Under the assumptions of Theorem 1 for the function r, if the implicit Euler finite difference numerical solution of problem (1) produced by (2) exists, then it satisfies the relations*

$$\zeta^L_{p|\mathcal{T}} \leq \zeta^L_{p|\Gamma}, \quad \mu_{p|\mathcal{T}} \leq \mu_{p|\Gamma}$$

for the number of L-level points (L is a fixed real number) and for the number of the local maximizers (minimizers).

Proof. First we prove the relation $\zeta^L_{p|\mathcal{T}} \leq \zeta^L_{p|\Gamma}$. Let L be a fixed real number.

If $u_i^{n_T} = L$ $(i = 0, \ldots, n+1)$, then $\zeta^L_{p|\mathcal{T}} = 1$ and $\zeta^L_{p|\Gamma} \geq 1$, since the values $u_0^{n_T}$ and $u_{n+1}^{n_T}$ belong also to Γ. Thus the required relation is satisfied. In the sequel we assume that not all values $u_i^{n_T}$ are equal to L. We consider the L-positive and L-negative groups of $\mathcal{T}$, and we number them with natural numbers in positive x direction. These groups are always disjoint, moreover, there is exactly one L-level point of the piecewise linear function $p(t,x)$ between any two adjacent groups.

Based on the properties (i)–(ii), we can assign an (not necessarily unique) L-positive or L-negative mesh point on the parabolic boundary Γ to each L-positive and L-negative group of $\mathcal{T}$, respectively. The ordering of these points of Γ is the same as that of the groups of $\mathcal{T}$, moreover, the function p has at least one L-level point between any two adjacent points. The possible L-level points in the corners $(T,0)$ and $(T,1)$ belong to both $\mathcal{T}$ and Γ. This shows the validity of the first statement.

Now we prove the second statement $\mu_{p|\mathcal{T}} \leq \mu_{p|\Gamma}$ for local maximizers (for local minimizers the proof is similar). We exclude the case when all values $u_i^{n_T}$ $(i = 0, \ldots, n+1)$ are the same. In this case the statement is trivially true. Let the local maximizers of the function $x \mapsto p(T,x)$ be numbered in positive x direction. There is exactly one local minimizer between each two adjacent local maximizers. If the first maximizer does not contain the point $(T,0)$ then this point is in a local minimizer, and similarly, if the last maximizer does not contain the point $(T,1)$ then this point is in a local minimizer.

Let the maximum value at the ith local maximizer be denoted by C_i, respectively. We assign positive values ε_i to each local maximizer such that $L_i = C_i - \varepsilon_i$ is greater than the maximum values at the neighboring local minimizers. Let us construct the L_i-positive groups that contain the mesh points of the ith local maximizers. Let these groups be denoted by $\mathcal{G}_i$. Then we construct the set of points $B^{L_i}_{\mathcal{G}_i}$. According to Lemma 1, these sets reach the parabolic boundary Γ at some points Q_i (which is not uniquely defined). We show that if $i \neq j$, then $B^{L_i}_{\mathcal{G}_i} \cap B^{L_j}_{\mathcal{G}_j} = \emptyset$. If $L_i = L_j$, then this follows from Lemma 1 directly. If we have $L_i < L_j$, then instead of the L_j-positive group $\mathcal{G}_j$ we consider the L_i-positive group that contains the mesh points of the jth maximizer. Let us denote this group by $\mathcal{G}_{j'}$. The choice of the ε values guarantees that $\mathcal{G}_i$ and $\mathcal{G}_{j'}$ are disjoint L_i-positive groups. Lemma 1 gives that $B^{L_i}_{\mathcal{G}_i} \cap B^{L_i}_{\mathcal{G}_{j'}} = \emptyset$. The inclusion $B^{L_j}_{\mathcal{G}_j} \subset B^{L_i}_{\mathcal{G}_{j'}}$ shows that the sets $B^{L_i}_{\mathcal{G}_i}$ and $B^{L_j}_{\mathcal{G}_j}$ must be disjoint.

Let us consider the L_i-positive groups of the mesh points on Γ that contains Q_i, respectively. Based on the above considerations these groups must be disjoint. Each group contains at least one local maximizer of p on Γ. Let us choose one of these. In this way we have assigned a local maximizer on Γ to each local maximizer on $\mathcal{T}$, which shows the relation $\mu_{p|\mathcal{T}} \leq \mu_{p|\Gamma}$ and completes the proof. ■

4. Some numerical examples

4.1. A simple verification of the result

In the first example we verify our theoretical result using the simple test problem

$$\begin{aligned} u'_t &= (1+u^2)(u''_{xx})^3, \ (t,x) \in Q_{\bar{T}}, \\ u(0,x) &= x^2 \sin(3\pi x), \ x \in [0,1], \\ u(t,0) &= \sin(2\pi t), \ u(t,1) = -\sin(2\pi t), \ t \in [0,T], \end{aligned} \tag{3}$$

which problem trivially satisfies the conditions posed in Theorem 1. We construct the implicit Euler finite difference solution of the problem using the parameters $T = 0.51$, $n = 20$ $(\Delta x = 1/21)$, $\Delta t = 1/100$. The approximation of the initial function u_0 can be seen on the left panel of Fig. 1. The right panel shows the approximation at the final time level $T = 0.51$.

We consider first the 0-level points of the numerical solution. The sign of the numerical solution values can be seen in Fig. 2. $+$, $*$ and o signs denote positive, negative and zero values, respectively. It can be seen easily that $\zeta^0_{p|\mathcal{T}} = 3$ and

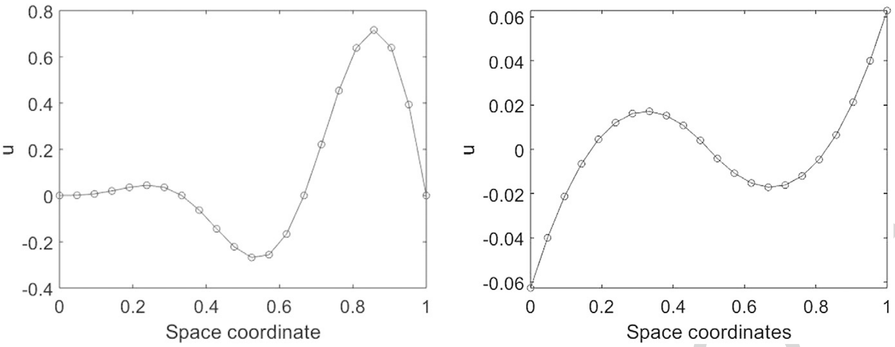

Fig. 1. Approximation of the initial function and the solution at $T = 0.51$ for the problem (3).

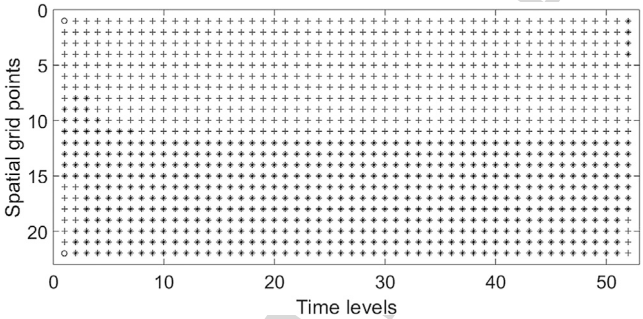

Fig. 2. The sign of the numerical solution values of problem (3). $+$, $*$ and o symbols denote positive, negative and zero values, respectively. In the figure, the coordinates are shifted by 1, which means that the sign of the value u_i^k can be found at the coordinates $(k+1, i+1)$.

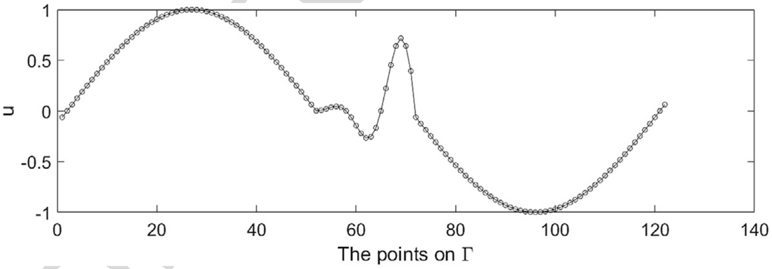

Fig. 3. The numerical solution values on the unbent parabolic boundary Γ. The figure shows the elements of the vector $[u_0^{n_T}, u_0^{n_T-1}, \ldots, u_0^0, u_1^0, \ldots, u_{n+1}^0, u_{n+1}^1, \ldots, u_{n+1}^{n_T}]$.

$\zeta^0_{p|\Gamma} = 6$. The number of the L-level points for other nonzero values L can be counted similarly. For example, for $L = 0.1$ we have $\zeta^{0.1}_{p|\mathcal{T}} = 0$ and $\zeta^{0.1}_{p|\Gamma} = 4$. These results verify the validity of our first statement in Theorem 2.

It can be seen from the right panel of Fig. 1 that the number of the local maximizers and minimizers of the numerical solution at the n_Tth time level is 2. The number of the local maximizers and minimizers of the numerical solution on Γ can be calculated using the plot (Fig. 3) of the numerical solution on the unbent set Γ. For the number of the local maximizers (minimizers) we have $\mu_{p|\mathcal{T}} = 2$ ($\mu_{p|\Gamma} = 4$), which verifies the second part of Theorem 2.

4.2. A counterexample for the explicit Euler method

In the second example we show that the investigated properties are not valid in general. While the implicit Euler method behaves well from the point of view of the investigated qualitative properties, the explicit Euler method does not possess these properties without some restrictions on the discretization parameters.

Please cite this article as: R. Horváth, On some discrete qualitative properties of implicit finite difference solutions of nonlinear parabolic problems, Journal of Computational and Applied Mathematics (2019), https://doi.org/10.1016/j.cam.2019.06.046.

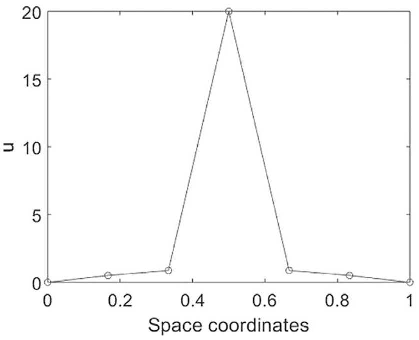

Fig. 4. The approximation of the initial function of problem (4) on the mesh with $n = 5$.

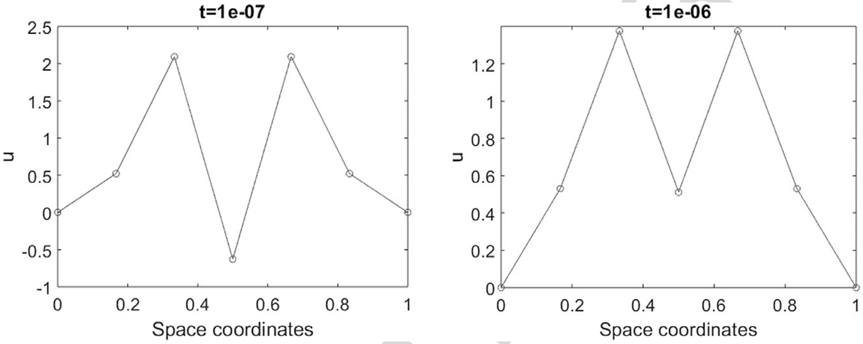

Fig. 5. Approximation of the solution of problem (4) at $t = 10^{-7}$ and $t = 10^{-6}$.

We solve the problem

$$\begin{aligned} &u'_t = (u''_{xx})^3, \ (t,x) \in Q_{\bar{T}}, \\ &u(0,x) = u_0(x) = \sin(\pi x) \cdot \left(1 + \left(19 e^{-\frac{(x-1/2)^2}{2 \cdot 0.01^2}}\right)\right), \ x \in [0,1], \\ &u(t,0) = u(t,1) = 0, \ t \in [0,T], \end{aligned} \tag{4}$$

with the explicit Euler method

$$\begin{aligned} &\frac{u_i^{k+1} - u_i^k}{\Delta t} = r\left(t_k, x_i, u_i^k, \frac{u_{i-1}^k - 2u_i^k + u_{i+1}^k}{\Delta x^2}\right), \quad \begin{matrix} i = 1, \dots, n \\ k = 0, \dots, n_T - 1 \end{matrix}, \\ &u_i^0 = u_0(x_i), \ i = 0, \dots, n+1, \\ &u_0^k = u_{n+1}^k = 0, \ k = 0, \dots, n_T. \end{aligned} \tag{5}$$

We set the discretization parameters to $n = 5$ ($\Delta x = 1/6$) and $\Delta t = 10^{-8}$. The approximation of the initial function on this mesh can be seen in Fig. 4. Taking into the account also the boundary conditions, we obtain that $\zeta^0_{p|_\Gamma} = 2$ and for the local maximizers (minimizers) we have $\mu_{p|_\Gamma} = 1(2)$, independently of the choice of T. At the 10th time level ($T = 10^{-7}$) we get the approximation of the left panel of Fig. 5. For this approximation we have $\zeta^0_{p|_\mathcal{T}} = 4$ and for the local maximizers (minimizers) we have $\mu_{p|_\mathcal{T}} = 2(3)$. We see that the explicit Euler finite difference method with the above parameters does not fulfill any of the required qualitative properties. We can see (right panel of Fig. 5) that the condition for the local maximizers (minimizers) is not fulfilled even after 100 time steps.

We can obtain qualitatively more correct numerical solution if we refine the mesh. If we set $n = 25$ and $\Delta t = 10^{-12}$ (such a small time step is needed to maintain the stability of the scheme) and perform 10^5 iterations, then we obtain the qualitatively adequate approximation of the solution at the same time level $t = 10^{-7}$ as before (Fig. 6).

5. Summary and conclusions

We have shown that the implicit Euler finite difference numerical solution of the nonlinear problem (1) satisfies two special properties. It does not allow the number of the L-levels and the number of the local maximizers and minimizers

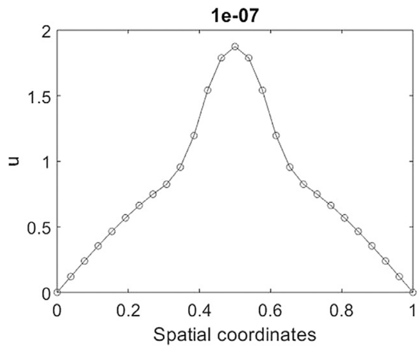

Fig. 6. Approximation of the solution of problem (3) at $t = 10^{-7}$ using the mesh parameters $n = 25$ and $\Delta t = 10^{-12}$.

of the solution to grow. We verified our results on some numerical test problems. We can conclude from the results that the implicit Euler finite difference solution can be used to solve the problem (1) without any restriction from the point of view of the above qualitative properties. We guess that, as in the linear case, the above properties are valid for the explicit Euler and the Crank–Nicolson methods only under strict conditions for the mesh. To find sufficient conditions to these methods would be an interesting task for future research.

Acknowledgments

This research was supported by the Hungarian Scientific Research Fund OTKA, Nos. K112157 and SNN125119. Q3

References

[1] I. Faragó, R. Horváth, Discrete maximum principle and adequate discretizations of linear parabolic problems, SIAM Sci. Comput. 28 (2006) 2313–2336.
[2] I. Faragó, R. Horváth, S. Korotov, Discrete maximum principles for FE solutions of nonstationary diffusion-reaction problems with mixed boundary conditions, Numer. Methods Partial Differential Equations 27 (3) (2011) 702–720.
[3] H. Fujii, Some remarks on finite element analysis of time-dependent field problems, in: Y. Yamada, R.H. Gallagher (Eds.), Theory and Practice in Finite Element Structural Analysis, University of Tokyo Press, Tokyo, 1973, pp. 91–106.
[4] V.A. Kondratev, E.M. Landis, Qualitative theory of second order linear partial differential equations, in: Yu. V. Egorov, M.A. Shubin (Eds.), Encyclopaedia of Mathematical Sciences, Vol. 32, Springer Verlag, 1991.
[5] T. Vejchodský, S. Korotov, A. Hannukainen, Discrete maximum principle for parabolic problems solved by prismatic finite elements, Math. Comput. Simulation 80 (8) (2010) 1758–1770.
[6] T.I. Zelenyak, M.P. Vishnevskii, M.M. Lavrentiev, Qualitative Theory of Parabolic Equations, Part 1, De Gruyter, 1997.
[7] R. Horváth, On the monotonicity conservation in numerical solutions of the heat equation, Appl. Numer. Math. 42 (2008) 189–199.
[8] R. Horváth, On the sign-stability of numerical solutions of one-dimensional parabolic problems, Appl. Math. Model. 32 (8) (2008) 1570–1578.
[9] J. Below, C. De Coster, Qualitative theory for parabolic problems under dynamical boundary conditions, J. Inequal. Appl. 5 (2000) 467–486.
[10] I. Faragó, R. Horváth, J. Karátson, S. Korotov, Qualitative properties of nonlinear parabolic operators, J. Math. Anal. Appl. 448 (1) (2017) 473–497.
[11] I. Faragó, J. Karátson, S. Korotov, Discrete maximum principles for the FEM solution of some nonlinear parabolic problems, Electron. Trans. Numer. Anal. 36 (2009) 149–167.
[12] T.B. Gyulov, M.N. Koleva, L.G. Vulkov, Efficient finite difference method for optimal portfolio in a power utility regime-switching model, Int. J. Comput. Math. (2018) http://dx.doi.org/10.1080/00207160.2018.1474207.
[13] R. Horváth, On the sign-stability of finite difference solutions of semilinear parabolic problems, Lect. Notes Comput. Sci. 5434 (2009) 305–313.
[14] A. Pauthier, P. Polacik, Large-time behavior of solutions of parabolic equations on the real line with convergent initial data, Nonlinearity 31 (2018) 4423–4441.
[15] M. Tabata, A finite difference approach to the number of peaks of solutions for semilinear parabolic problems, J. Math. Soc. Japan 32 (1) (1980) 171–192.
[16] K. Nickel, Gestaltaussagen über Lösungen parabolischer Differentialgleichungen, J. Reine Angew. Math. 211 (1962) 78–94.

Please cite this article as: R. Horváth, On some discrete qualitative properties of implicit finite difference solutions of nonlinear parabolic problems, Journal of Computational and Applied Mathematics (2019), https://doi.org/10.1016/j.cam.2019.06.046.