Second international conference *Supercomputer technologies of mathematical modelling*, Yakutsk, Russia, July 8-11, 2013.

# RELIABLE NUMERICAL MODELS FOR PARABOLIC PROBLEMS

**István Faragó**

***Eötvös Loránd University and MTA-ELTE Numerical Analysis and Large Networks Research Group, Budapest, Hungary***
faragois@cs.elte.hu

**Abstract**

By constructing mathematical and numerical models in order to describe some real-life problem, we require that these models have different qualitative properties, which typically arise from some basic principles of the modelled phenomena. In this paper we investigate this question for continuous and discrete models. We give the conditions for the discretization parameters under which the qualitative properties are preserved.

**Keywords:** numerical modelling, monotone schemes, finite difference, finite element.

**Acknowledgements** The work is supported by Hungarian National Research Fund OTKA No. K81403.

## 1 Introduction

When we construct mathematical and/or numerical models in order to model or solve a real-life problem, these models should have different qualitative properties, which typically arise from some basic principles of the modelled phenomena. E.g., many processes, varying in time, have such properties as monotonicity, non-negativity preservation and maximum principles. The discretization can qualitatively deform the mathematical models: certain qualitative properties which are inherent in the original real-life process are not preserved. Therefore, our goal is to guarantee quality preservation. It is almost obvious that the complexity of a model defines its tractableness: for structurally simple models, usually, it is easier to give qualitative characterization and/or define its solution. (For complex problems, in general, it is even impossible.) We note that the operator splitting method is a powerful tool to decompose a complex time-dependent problem into a sequence of simpler sub-problems, for which the required qualitative properties could be checked easier.

For some details and proofs we refer to the references [2] - [5].

## 2 Qualitative properties of the linear operators for the continuous models

Let $\Omega$ denote a bounded, simply connected domain in $\mathbb{R}^d$ $(d \in \mathbb{N}^+)$ with a Lipschitz-continuous boundary $\partial\Omega$. We introduce the sets

$$Q_\tau = \Omega\times(0,\tau), \qquad \bar{Q}_\tau = \bar{\Omega}\times[0,\tau], \qquad Q_{\bar{\tau}} = \Omega\times(0,\tau], \qquad \Gamma_\tau = (\partial\Omega\times[0,\tau])\cup(\Omega\times\{0\})$$

for any arbitrary positive number $\tau$. The set $\Gamma_\tau$ is usually called *parabolic boundary.* For some fixed number $T > 0$, we consider the linear partial differential operator

$$L \equiv \frac{\partial}{\partial t} - \sum_{0 \le |\varsigma| \le \delta} a_\varsigma \frac{\partial^{|\varsigma|}}{\partial^{\varsigma_1} x_1 \dots \partial^{\varsigma_d} x_d} \equiv \frac{\partial}{\partial t} - \sum_{0 \le |\varsigma| \le \delta} a_\varsigma D^\varsigma, \tag{1}$$

where $\delta$ is the order of the operator, $\varsigma_1, \dots, \varsigma_d$ denote non-negative integers, $|\varsigma|$ is defined as $|\varsigma| = \varsigma_1 + \dots + \varsigma_d$ for the multi-index $\varsigma = (\varsigma_1, \dots, \varsigma_d)$, and the coefficient functions $a_\varsigma : Q_T \to \mathbb{R}$ are bounded and sufficiently smooth in the set $Q_T$.

Typically the function $v \in \operatorname{dom} L$ describes the values of a physical quantity in the domain $\bar{Q}_T$, that is, the dependence of the quantity on place and time. The above mentioned physical property can be connected by the following definition.

**Definition.** Operator (1) is said to be *monotone* if for all $t^\star \in (0, T)$ and $v_1, v_2 \in \operatorname{dom} L$ such that $v_1|_{\Gamma_{t^\star}} \ge v_2|_{\Gamma_{t^\star}}$ and $(Lv_1)|_{Q_{\bar{t}^\star}} \ge (Lv_2)|_{Q_{\bar{t}^\star}}$, the relation $v_1|_{Q_{\bar{t}^\star}} \ge v_2|_{Q_{\bar{t}^\star}}$ holds.

**Definition.** The operator $L$ is called *non-negativity preserving* (NP) when for any $v \in \operatorname{dom} L$ and $t^\star \in (0, T)$ such that $v|_{\Gamma_{t^\star}} \ge 0$ and $(Lv)|_{Q_{\bar{t}^\star}} \ge 0$, the relation $v|_{Q_{\bar{t}^\star}} \ge 0$ holds.

Clearly, these properties of the linear operator (1) are equivalent.

Often we may need only certain characterization of $v$, which does not require the knowledge of $v$ in the whole domain. From a practical point of view, only such estimates are suitable which include only the known input data. This kind of estimations is called *maximum-minimum principles.*

For the operators various maximum-minimum principles are defined and used in the literature, because they well characterize the operator $L$ itself (cf. [1] and references therein). Now we list four possible variants of them.

We say that the operator $L$ satisfies the *weak maximum-minimum principle* (WMP) if for any function $v \in \operatorname{dom} L$ and any $t^\star \in (0, T)$ the inequalities

$$\min\{0, \min_{\Gamma_{t^\star}} v\} + t^\star \min\{0, \inf_{Q_{\bar{t}^\star}} Lv\} \le v(x,t) \le \max\{0, \max_{\Gamma_{t^\star}} v\} + t^\star \max\{0, \sup_{Q_{\bar{t}^\star}} Lv\}$$

are valid for all $(x,t) \in \bar{Q}_{t^\star}$. The operator $L$ satisfies the *strong maximum-minimum principle* (SMP) if the inequalities

$$\min_{\Gamma_{t^\star}} v + t^\star \cdot \min\{0, \inf_{Q_{\bar{t}^\star}} Lv\} \le \min_{\bar{Q}_{t^\star}} v \le \max_{\bar{Q}_{t^\star}} v \le \max_{\Gamma_{t^\star}} v + t^\star \cdot \max\{0, \sup_{Q_{\bar{t}^\star}} Lv\}$$

are satisfied.

When the sign of $Lv$ is known, then it is possible that the estimates involve only the known values of $v$ on the parabolic boundary. These types of maximum-minimum principles are called *boundary maximum-minimum principles*, which are frequently used in proofs of the uniqueness theorems. We say that the operator $L$ satisfies the *weak boundary maximum-minimum principle* (WBMP) if the inequalities $\min\{0, \min_{\Gamma_{t^\star}} v\} \le \min_{\bar{Q}_{t^\star}} v \le \max_{\bar{Q}_{t^\star}} v \le \max\{0, \max_{\Gamma_{t^\star}} v\}$ hold. We say that the *strong boundary maximum-minimum principle* (SBMP) is true if for any function $v \in \operatorname{dom} L$ and any $t^\star \in (0, T)$ such that $Lv|_{Q_{\bar{t}^\star}} \ge 0$, the relations $\min_{\Gamma_{t^\star}} v = \min_{\bar{Q}_{t^\star}} v \le \max_{\bar{Q}_{t^\star}} v = \max_{\Gamma_{t^\star}} v$ hold.

The implications between the different qualitative properties are shown in Figure 1. (Here *cal*1 denotes the constant 1 function, and the solid arrows mean unconditional implications while the dashed ones are true only under the indicated assumptions.) As one can easily observe, for $a_0 = 0$ the non-negativity property (NP) implies all the other properties.

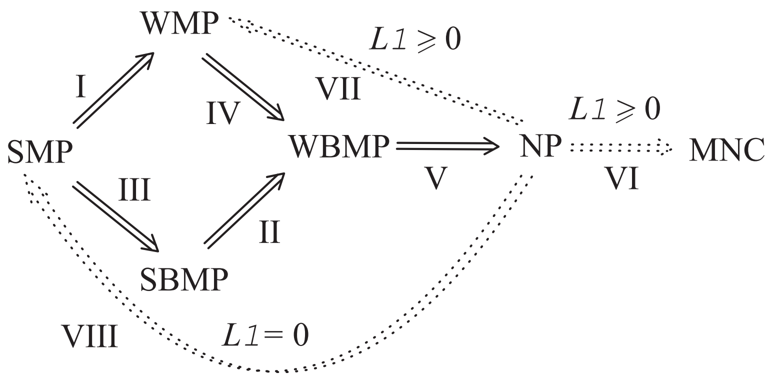


Рис. 1: Implications between the qualitative properties.

## 3 Discrete analogues of the qualitative properties - reliable discrete models

In this part we present the natural discrete analogs of the qualitative properties formulated in the the previous section for the continuous models.

First, we introduce some notations.

Let us assume that the sets $\mathcal{P} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ and $\mathcal{P}_\partial = \{\mathbf{x}_{N+1}, \ldots, \mathbf{x}_{N+N_\partial}\}$ consist of different vertices in $\Omega$ and on $\partial\Omega$, respectively. We set $\bar{N} = N + N_\partial$ and $\bar{\mathcal{P}} = \mathcal{P} \cup \mathcal{P}_\partial$. Let $T$ and $\Delta t < T$ be two arbitrary positive numbers. Moreover, let us suppose that the integer $M$ satisfies the condition $M\Delta t \le T < (M+1)\Delta t$ and introduce the set $\mathcal{R} = \{t_n = n\Delta t \,|\, n = 0, 1, \ldots, M\}$. For any values $\tau$ from the set $\mathcal{R}$ we introduce the notations

$$\mathcal{R}_\tau = \{t \in \mathcal{R} \,|\, 0 < t < \tau\}, \quad \mathcal{R}_{\bar{\tau}} = \{t \in \mathcal{R} \,|\, 0 < t \le \tau\}, \quad \mathcal{R}^0_{\bar{\tau}} = \{t \in \mathcal{R} \,|\, 0 \le t \le \tau\},$$

and the sets

$$\mathcal{Q}_\tau = \mathcal{P} \times \mathcal{R}_\tau, \qquad \bar{\mathcal{Q}}_\tau = \bar{\mathcal{P}} \times \mathcal{R}^0_{\bar{\tau}}, \qquad \mathcal{Q}_{\bar{\tau}} = \mathcal{P} \times \mathcal{R}_{\bar{\tau}}, \qquad \mathcal{G}_\tau = (\mathcal{P}_\partial \times \mathcal{R}^0_{\bar{\tau}}) \cup (\mathcal{P} \times \{0\}).$$

Linear mappings that map from the space of real-valued functions defined on $\bar{\mathcal{Q}}_{t_M}$ to the space of real-valued functions defined on $\mathcal{Q}_{t_M}$ are called *discrete (linear) mesh operators*. We define the qualitative properties of the discrete mesh operators in an analogous way to those in the linear partial differential operator case. For simplicity, we will formulate only one of them, which plays central role in our further investigations. (The ordering relations for vectors and matrices in the sequel are always meant elementwise.)

**Definition.** We say that the discrete mesh operator $\mathcal{L}$ is *discrete non-negativity preserving* (DNP) if for any $\nu \in \operatorname{dom} \mathcal{L}$ and any $t^\star \in \mathcal{R}_{t_M}$ such that $\min_{\mathcal{G}_{t^\star}} \nu \ge 0$ and $\mathcal{L}\nu|_{\mathcal{Q}_{\bar{t}^\star}} \ge 0$, the relation $\nu|_{\mathcal{Q}_{\bar{t}^\star}} \ge 0$ holds.

Let us introduce two special mesh functions, $\mathbb{1}$ and $\mathbb{t}$, defined on $\bar{\mathcal{Q}}_{t_M}$ with the following equalities: $\mathbb{1}(\mathbf{x}_i, t_n) = 1$, $\mathbb{t}(\mathbf{x}_i, t_n) = n\Delta t$ for all $(\mathbf{x}_i, t_n) \in \bar{\mathcal{Q}}_{t_M}$. (These mesh functions are the discrete analogue of the continuous functions $v(\mathbf{x}, t) = 1$ and

$v(\mathbf{x}, t) = t$, respectively, associated with the mesh $\bar{\mathcal{Q}}_{t_M}$.) We will also use the notation $\mathbb{m}$ for the mesh function $(\mathbb{m})_i^n = n$. The implications between the different discrete qualitative properties are shown in Figure 2.

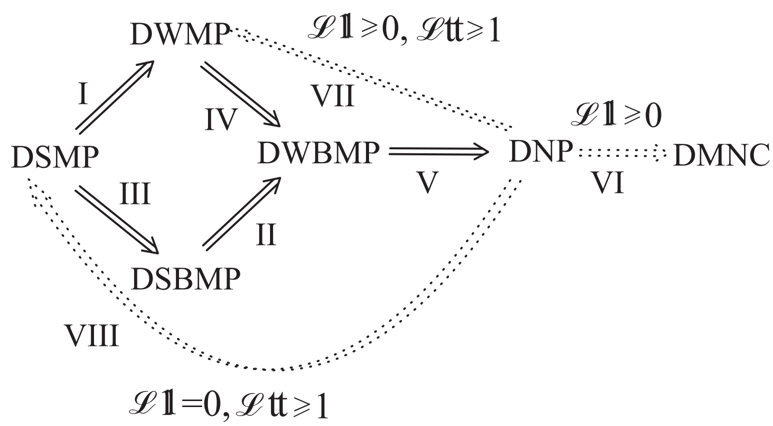


Рис. 2: Implications between the discrete qualitative properties.

From the illustrated implications we can directly observe the validity of the following statement.

**Theorem 1.** Under the conditions $\mathcal{L}\mathbb{1} = 0$ and $\mathcal{L}\mathbb{t} = 1$ the discrete non-negativity property (DNP) implies all the other properties.

# 4 Two-level discrete mesh operators

In the sequel, the values $\nu(\mathbf{x}_i, n\Delta t)$ of the function $\nu$ defined in $\bar{\mathcal{Q}}_{t_M}$ will be denoted by $\nu_i^n$. Similar notation is applied to the function $\mathcal{L}\nu$. We introduce the vectors

$$\boldsymbol{\nu}^n = [\nu_1^n, \ldots, \nu_{\bar{N}}^n], \quad \boldsymbol{\nu}_0^n = [\nu_1^n, \ldots, \nu_N^n], \quad \boldsymbol{\nu}_\partial^n = [\nu_{N+1}^n, \ldots, \nu_{\bar{N}}^n].$$

In many numerical methods, the discrete mesh operators have a special form, namely, they are defined as

$$(\mathcal{L}\nu)_i^n = (\mathbf{X}_1^{(n)}\boldsymbol{\nu}^n - \mathbf{X}_2^{(n)}\boldsymbol{\nu}^{n-1})_i, \quad i = 1, \ldots, N, \quad n = 1, \ldots, M, \tag{2}$$

where $\mathbf{X}_1^{(n)}, \mathbf{X}_2^{(n)} \in \mathbb{R}^{N\times\bar{N}}$ are some given matrices. The term "two-level method"refers to the fact that two discrete time levels are involved into the definition of the mesh operator. Sometimes such a method is also called "one-step method". In order to give the connections between the qualitative properties of such a type of mesh operators, we reformulate the conditions in Theorem 1, see also Figure 2. We have already introduced the notation $\mathbf{e} = [1, \ldots, 1] \in \mathbb{R}^{\bar{N}}$. The $N$-element and the $(\bar{N} - N)$-element version of this vector will be denoted by $\mathbf{e}_0$ and $\mathbf{e}_\partial$, respectively, i.e., $\mathbf{e} = [\mathbf{e}_0 | \, \mathbf{e}_\partial]$. Then the condition $\mathcal{L}\mathbb{1} = 0$ reads as

$$(\mathbf{X}_1^{(n)} - \mathbf{X}_2^{(n)})\mathbf{e} = \mathbf{0}$$

while the condition $\mathcal{L}\mathbb{t} \geq 1$ means that

$$\mathbf{X}_1^{(n)}(\Delta t n\mathbf{e}) - \mathbf{X}_2^{(n)}(\Delta t(n-1)\mathbf{e}) = \Delta t(n(\mathbf{X}_1^{(n)} - \mathbf{X}_2^{(n)})\mathbf{e} + \mathbf{X}_2^{(n)}\mathbf{e}) \geq \mathbf{e}_0.$$

If $(\mathbf{X}_1^{(n)} - \mathbf{X}_2^{(n)})\mathbf{e} = \mathbf{0}$ $(n = 1, \dots, M)$, then the above condition reduces to $\Delta t \mathbf{X}_2^{(n)}\mathbf{e} \geq \mathbf{e}_0$. Since $\mathbf{X}_2^{(n)}\mathbf{e} = \mathbf{X}_1^{(n)}\mathbf{e}$, we have

**Theorem 2.** If a non-negativity preserving discrete mesh operator of type (2) has the properties

$$(\mathbf{X}_1^{(n)} - \mathbf{X}_2^{(n)})\,\mathbf{e} = \mathbf{0}, \quad \Delta t \mathbf{X}_1^{(n)}\,\mathbf{e} \geq \mathbf{e}_0 \;\text{ or }\; \Delta t \mathbf{X}_2^{(n)}\,\mathbf{e} \geq \mathbf{e}_0, \tag{3}$$

then the operator possesses all the discrete qualitative properties.

How to guarantee the DNP property? To this aim, we introduce the following convenient partition of the matrices $\mathbf{X}_1^{(n)}$ and $\mathbf{X}_2^{(n)}$:

$$\mathbf{X}_1^{(n)} = [\mathbf{X}_{10}^{(n)} | \mathbf{X}_{1\partial}^{(n)}], \quad \mathbf{X}_2^{(n)} = [\mathbf{X}_{20}^{(n)} | \mathbf{X}_{2\partial}^{(n)}], \tag{4}$$

where $\mathbf{X}_{10}^{(n)}$ and $\mathbf{X}_{20}^{(n)}$ are square matrices from $\mathbb{R}^{N\times N}$, and $\mathbf{X}_{1\partial}^{(n)}, \mathbf{X}_{2\partial}^{(n)} \in \mathbb{R}^{N\times N_\partial}$.

Then the following statement is valid.

**Theorem 3.** Let us suppose that the matrices $\mathbf{X}_{10}^{(n)}$ $(n = 1, \dots, M)$ of the discrete mesh operator $\mathcal{L}$ defined in (2) are regular. Then $\mathcal{L}$ possesses the discrete non-negativity preservation property if and only if the following relations hold for all $n = 1, \dots, M$,

(P1) $(\mathbf{X}_{10}^{(n)})^{-1} \geq \mathbf{0}$,
(P2) $-(\mathbf{X}_{10}^{(n)})^{-1}\mathbf{X}_{1\partial}^{(n)} \geq \mathbf{0}$,
(P3) $(\mathbf{X}_{10}^{(n)})^{-1}\mathbf{X}_2^{(n)} \geq \mathbf{0}$.

Hence, summarizing the results, we have

**Theorem 4.** Under the conditions (3) and (P1)-(P3) the mesh operator of type (2) has all discrete qualitative properties.

# 5 Qualitative properties of the discrete heat conduction mesh operator in the 1D

In this section we consider the one-dimensional heat conduction operator with a constant coefficient, which is assumed, for simplicity, to be equal to one, i.e.,

$$L \equiv \frac{\partial}{\partial t} - \frac{\partial^2}{\partial x^2}. \tag{5}$$

On a fixed uniform mesh we consider the one-step discrete mesh operator $\mathcal{L}$ , obtained by finite difference method and having the form (2)-(4) with $N_\partial = 2$, $\bar{N} = N + 2$ and

$$\mathbf{X}_{10}^{(n)} = \text{tridiag}\left[-\frac{\theta}{h^2}, \frac{1}{\Delta t} + 2\frac{\theta}{h^2}, -\frac{\theta}{h^2}\right] \in \mathbb{R}^{N\times N},$$

$$\mathbf{X}_{20}^{(n)} = \text{tridiag}\left[\frac{1-\theta}{h^2}, \frac{1}{\Delta t} - 2\frac{1-\theta}{h^2}, \frac{1-\theta}{h^2}\right] \in \mathbb{R}^{N\times N},$$

$$\mathbf{X}_{1\partial}^{(n)} = -\frac{\theta}{h^2}\mathbf{E}; \;\; \mathbf{X}_{2\partial}^{(n)} = \frac{1-\theta}{h^2}\mathbf{E}, \;\; \text{where } \mathbf{E} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}^T \in \mathbb{R}^{N\times 2}.$$

Let us apply Theorem 4. By direct calculation we get that for this discrete mesh operator the conditions in (3) are satisfied. (In the second inequality we verify the condition $\mathbf{X}_1^{(n)}\,\mathbf{e} \geq \mathbf{e}_0$.) Since $\mathbf{X}_{10}^{(n)}$ is an M-matrix, therefore its inverse is nonnegative. Therefore, the validity of (P1) and (P2) is straightforward. Since (P3) can be written as

$$(\mathbf{X}_{10}^{(n)})^{-1}\mathbf{X}_2^{(n)} = (\mathbf{X}_{10}^{(n)})^{-1}[\mathbf{X}_{2,0}^{(n)}, \mathbf{X}_{2\partial}^{(n)}] = (\mathbf{X}_{10}^{(n)})^{-1}\mathbf{X}_{2,0}^{(n)} + (\mathbf{X}_{10}^{(n)})^{-1}\mathbf{X}_{2\partial}^{(n)} \geq \mathbf{0},$$

due to the obvious relation $(\mathbf{X}_{10}^{(n)})^{-1}\mathbf{X}_{2\partial}^{(n)} \geq \mathbf{0}$, we get the following statement.

**Theorem 5.** The finite difference discrete mesh operator has all the discrete qualitative properties if the condition

$$(\mathbf{X}_{10}^{(n)})^{-1}\mathbf{X}_{2,0}^{(n)} \geq \mathbf{0} \tag{6}$$

is satisfied.

Clearly, to satisfy (6) the condition $\mathbf{X}_{2,0}^{(n)} \geq \mathbf{0}$ is sufficient. Hence, we get

**Theorem 6.** Under the condition $\Delta t/h^2 \leq 0.5$ the finite difference discrete mesh operator has all the discrete qualitative properties.

However, the necessity of this condition is not clear. Due to the special structure of these matrices (they are uniformly continuant, symmetrical tridiagonal matrices), they have some special qualitative properties, which will be considered in the sequel.

Hereafter we investigate the real, uniformly continuant, symmetrical tridiagonal matrices

$$\mathbf{X}_{10} = \text{tridiag}[-z, 2\tilde{w}, -z];\;\; \mathbf{X}_{20} = \text{tridiag}[s, \tilde{p}, s]. \tag{7}$$

We assume that $z > 0$ and $s > 0$. Then, we can consider the equivalent form of the matrices,

$$\mathbf{X}_{10} = z \cdot \text{tridiag}[-1, 2w, -1];\;\; \mathbf{X}_{20} = s \cdot \text{tridiag}[1, p, 1], \tag{8}$$

where $w = \tilde{w}/z$ and $p = \tilde{p}/s$. The inverse of such a matrix can be defined directly, with the aid of the one-pair matrix $\mathbf{G} = (G_{ij})$, depending on the parameter $w$ as

$$G_{i,j} = \begin{cases} \gamma_{i,j}, & \text{if } i \leq j \\ \gamma_{j,i}, & \text{if } j \leq i \end{cases} \tag{9}$$

and for $w > 1$ we have $\gamma_{i,j} = \dfrac{\text{sh}(i\vartheta)\text{sh}(N+1-j)\vartheta}{\text{sh}\vartheta\text{sh}(N+1)\vartheta}$ with $\vartheta = \text{arch}(w)$. Hence we have the relation $\mathbf{X}_{10}^{-1} = (1/z)\mathbf{G}$, thus, for the matrix $\mathbf{X}_{pr} = \mathbf{X}_{10}^{-1}\mathbf{X}_{20}$ a direct computation verifies the validity of the relation

$$\mathbf{X}_{pr} = \frac{s}{z}\left[(2w+p)\mathbf{G} - \mathbf{I}_0\right], \tag{10}$$

where $\mathbf{I}_0 \in \mathbb{R}^{N\times N}$ denotes the unit matrix. Using (10), we obtain the following statement.

**Theorem 7.** Let us suppose that $w > 1$. Then $\mathbf{X}_{pr} \in \mathbb{R}^{N\times N}$ is non-negative for an arbitrary fixed $N$ if and only if the conditions

$$2w + p > 0, \quad \gamma_{i,i} \geq \frac{1}{2w+p}, \;\; i = 1, 2, \ldots, N \tag{11}$$

are fulfilled.

For the diagonal elements of the matrix $\mathbf{X}_{pr}$ the relation

$$\min\{\gamma_{i,i}, \;\; i = 1, 2, \ldots, N\} = \gamma_{1,1} = \gamma_{N,N}$$

holds. Hence we get that $\mathbf{X}_{pr} \in \mathbb{R}^{N\times N}$ is non-negative for an arbitrary fixed $N$ if and only if the conditions (11) and

$$\frac{\mathrm{sh}(N\vartheta)}{\mathrm{sh}((N+1)\vartheta)} \geq \frac{1}{2w+p} \tag{12}$$

are satisfied. Obvously the following relations are true:

$$\sup\left\{\frac{\mathrm{sh}(N\vartheta)}{\mathrm{sh}((N+1)\vartheta)};\ N\in\mathbb{N}\right\} = \mathrm{ch}(\vartheta)-\mathrm{sh}(\vartheta) = \exp(-\vartheta),$$

$$\exp(-\vartheta) = \exp(-\mathrm{arch}(w)) = \exp\left(\ln\left[w+\sqrt{w^2-1}\right]^{-1}\right) = \left[w+\sqrt{w^2-1}\right]^{-1}.$$

Therefore, from some sufficiently large $N_0 \in \mathbb{N}$ the relation $\mathbf{X}_{pr} \geq \mathbf{0}$ may be true only if the condition $\left[w+\sqrt{w^2-1}\right]^{-1} > \dfrac{1}{2w+p}$, i.e., the condition

$$p > -w + \sqrt{w^2-1} \tag{13}$$

is fulfilled. This proves the following

**Theorem 8.** Assume that $z>0$, $s>0$ and $w>1$. If, for some number $N_0 \in \mathbb{N}$, the conditions (11) and (12) are satisfied, then all matrices $\mathbf{X}_{pr} \in \mathbb{R}^{N\times N}$ with $N \geq N_0$ are non-negative. Moreover, there exists such a number $N_0$ if and only if the condition (11) (13) holds.

Let us analyze the conditions for different values of $N_0$. When $N_0 = 1$, then due to the relation $\dfrac{\mathrm{sh}\vartheta}{\mathrm{sh}(2\vartheta)} = \dfrac{1}{2\mathrm{ch}\vartheta} = \dfrac{1}{2w}$, (12) results in the condition

$$p \geq 0. \tag{14}$$

Since $\dfrac{\mathrm{sh}(2\vartheta)}{\mathrm{sh}(3\vartheta)} = \dfrac{2\mathrm{ch}(\vartheta)}{4\mathrm{ch}^2(\vartheta)-1} = \dfrac{2w}{4w^2-1}$, for $N_0 = 2$ the condition (12) results in the assumption

$$p \geq -\frac{1}{2w}. \tag{15}$$

Let us apply the above results to the finite difference discrete mesh operator, by using the notation $q = \Delta t/h^2$. In case $\theta = 0$ we have $\mathbf{X}_{pr} = \mathrm{tridiag}[q, 1-2q, q]$, hence the condition is $q \leq 0.5$. In case $\theta = 1$ we have $\mathbf{X}_{pr} = (\mathrm{tridiag}[-q, 1+2q, -q])^{-1}$. Since $\mathrm{tridiag}[-q, 1+2q, -q]$ is an M-matrix, hence for this case we do not have any condition for the choice of $q$. Let us assume that $\theta \in (0,1)$.

Then we can use the form (8) with the choice

$$z = \frac{\theta q}{\Delta t}, \quad s = \frac{(1-\theta)q}{\Delta t}, \quad w = \frac{1+2\theta q}{2\theta q}, \quad p = \frac{1-2(1-\theta)q}{(1-\theta)q}. \tag{16}$$

Since for the considered $\theta$ we have $z > 0$, $s >$ and $w > 1$, Theorem 7 and its consequences are applicable.

Using (14), we directly get that the condition of the non-negativity preservation for all $N = 1, 2, \ldots$ is the condition

$$q \leq \frac{1}{2(1-\theta)}. \tag{17}$$

| $\theta$ | $N=1$ | $N=2$ | $N=\infty$ |
|---|---|---|---|
| 0 | 0.5 | 0.5 | 0.5 |
| $0.5-(12q)^{-1}$ | 0.8333 | 0.9574 | 0.9661 |
| 0.5 | 1 | $2\sqrt{3}/3$ | $2(2-\sqrt{2})$ |
| 1 | $\infty$ | $\infty$ | $\infty$ |

Таблица 1: Non-negativity providing upper bounds for $q$ in the different finite difference mesh operators.

However, the non-negativity preservation for all $N=2,3,\ldots$ should be guaranteed by the weaker condition (15), which, in our case, yields the upper bound

$$q \le \frac{-1+2\theta+\sqrt{1-\theta(1-\theta)}}{3\theta(1-\theta)}. \tag{18}$$

Our aim is to get the largest value for $q$ under which the non-negativity preservation for sufficiently large values $N$ still holds. The necessary condition (13) results in the bound

$$q \le \frac{1-\sqrt{1-\theta}}{\theta(1-\theta)}. \tag{19}$$

We can summarize our results as follows.

**Theorem 9.** The finite difference discrete mesh operator $\mathcal{L}$ is non-negativity preserving (and hence, it has all discrete qualitative properties) for each $N\ge 1$ if and only if the condition (17) holds. It is non-negativity preserving for each $N\ge 2$ only under the condition (18). There exists a number $N_0 \in \mathbb{N}$ such that $\mathcal{L}$ is non-negativity preserving for each $N \ge N_0$ if and only if the weaker condition (19) is satisfied.

We demonstrate our results on some special choices of $\theta$. The results are shown in Table 1.

# References


[1] Ladyzhenskaya, O. A., Solonnikov, V. A., Ural'tseva, N. N. Linear and quasilinear equations of parabolic type. Nauka, Moscow,1967.

[2] Faragó, I. Tarvainen, P. Qualitative analysis of one-step algebraic models with tridiagonal Toeplitz matrices, Period. Math. Hungar., 1997.

[3] Faragó, I. Horváth, R. Discrete maximum principle and adequate discretizations of linear parabolic problems. SIAM Scientific Computing, 2006.

[4] Faragó, I. Horváth, R. Continuous and discrete parabolic operators and their qualitative properties, IMA Numerical Analysis, 2008.

[5] Faragó, I., Korotov,S., Szabó,T. Non-negativity preservation of the discrete nonstationary heat equation in 1D and 2D, Aplimat-Journal of Applied Mathematics, 2010.