Mathematical Institute
Academy of Sciences, Czech Republic
Žitná 25, 115 67 Praha 1

# Method of lines and conservation of nonnegativity

## Semidiscrete solution of the linear parabolic problem does not conserve nonnegativity

Tomáš Vejchodský

e-mail: vejchod@math.cas.cz

Colloquium 09/03/2004

University of Texas at El Paso
Department of Mathematical Sciences

# Method of lines and conservation of nonnegativity

## Semidiscrete solution of the linear parabolic problem does not conserve nonnegativity

Tomáš Vejchodský

e-mail: vejchod@math.utep.edu

Colloquium 09/03/2004

## Introduction

- Absolute temperature, density, concentration – nonnegative
- Mathematical models – maximum (comparison) principle
- Discrete models – discrete maximum principle

**Applications:**
physics (heat conduction, nuclear), engineering, economy

# The heat conduction problem

**Classical formulation:**

$$\begin{aligned} \partial_t u(x,t) - \Delta u(x,t) &= 0 \qquad &&\text{in } \Omega \times (0,T), \\ u(x,t) &= 0 &&\text{on } \partial\Omega \times [0,T], \\ u(x,0) &= u_0(x) &&\text{in } \Omega, \end{aligned}$$

where $T > 0$,
$\Omega \subset \mathbb{R}^d$ polyhedral domain, $d \in \{1, 2, 3, \dots\}$ arbitrary,
$u$ temperature, $u_0$ initial condition – sufficiently smooth.

**Comparison principle:**

$$u_{01} \le u_{02} \text{ in } \Omega \quad \Longrightarrow \quad u_1 \le u_2 \text{ in } \Omega \times (0,T).$$

**Definition:**
The problem conserves nonnegativity $\overset{\text{def}}{\Longleftrightarrow}$ $(\forall u_0 \ge 0 \;\Rightarrow\; u \ge 0)$.

Comparison principle $\Longleftrightarrow$ nonnegativity conservation.

## Numerical approaches

- Method of lines:
  $x$ discretized, $t$ continuous $\Rightarrow$ system of ODE
  (Solver of ODE's $\Rightarrow$ full discretization.)

- Rothe's method:
  $x$ continuous, $t$ discretized $\Rightarrow$ series of elliptic problems
  (Elliptic problems solver $\Rightarrow$ full discretization.)

# Weak formulation

**Classical formulation:**

$$\begin{aligned} \partial_t u(x,t) - \Delta u(x,t) &= 0 && \text{in } \Omega \times (0,T), \\ u(x,t) &= 0 && \text{on } \partial\Omega \times [0,T], \\ u(x,0) &= u_0(x) && \text{in } \Omega. \end{aligned}$$

**Weak formulation:**

find $u \in H_0^1(\Omega)$ such that $\partial_t u \in L^2(\Omega)$ for a.e. $t \in (0,T)$ and

$$\int_\Omega \partial_t u v \, \mathrm{d}x + \int_\Omega \nabla u \cdot \nabla v \, \mathrm{d}x = 0 \quad \forall v \in H_0^1(\Omega), \text{ a.e. } t \in [0,T],$$

$$u(x,0) = u_0(x) \quad \text{in } \Omega.$$

Initial condition: $u_0 \in H_0^1(\Omega)$.

$$H^1(\Omega) = \{v \in L^2(\Omega) : \partial_{x_i} v \in L^2(\Omega)\}$$

$$H_0^1(\Omega) = \{v \in H^1(\Omega) : v|_{\partial\Omega} = 0\}$$

## Finite elements

$T_h$ ............. simplicial partition of $\Omega$.
$V_{h0} \subset H_0^1(\Omega)$ ... finite element space
(continuous and piecewise linear functions based on $T_h$).
$V_{h0} = \text{span}\{\varphi_1, \varphi_2, \dots, \varphi_N\}$.

Acute type condition:
1D ... empty
2D ... all angles in triangulation $\leq \pi/2$
3D ... all dihedral angles between faces of all tetrahedra $\leq \pi/2$

($\Rightarrow$ off-diagonal entries of the sitffness matrix $A$ are $\leq 0$
$\Rightarrow$ $A^{-1} \geq 0$ $\Rightarrow$ discrete maximum principle for elliptic problems.)

# Semidiscretization

Semidiscrete Galerkin problem: find $\bar{u}_h \in C^1([0,T], V_{h0})$ such that

$$\int_\Omega \partial_t \bar{u}_h v_h \,\mathrm{d}x + \int_\Omega \nabla \bar{u}_h \cdot \nabla v_h \,\mathrm{d}x = 0 \quad \forall v_h \in V_{h0},$$

$$\bar{u}_h(x,0) = \bar{u}_{h0}(x) \qquad \text{in } \Omega.$$

$\bar{u}_{h0}$ . . . projection of $u_0$ into $V_{h0}$.

$$\bar{u}_h(x,t) = \sum_{j=1}^{N} y_j(t)\varphi_j(x) \quad \Updownarrow \quad v_h = \varphi_i$$

$$M\dot{y}(t) + Ay(t) = 0$$

$$y(0) = y_0$$

Mass matrix: $M_{ij} = \int_\Omega \varphi_i \varphi_j \,\mathrm{d}x$.

Stiffness matrix: $A_{ij} = \int_\Omega \nabla\varphi_i \cdot \nabla\varphi_j \,\mathrm{d}x$.

Vector of coefficients: $y(t) = (y_1(t), y_2(t), \dots, y_N(t))^\top$.

Initial condition: $y_0 = (y_{01}, y_{02}, \dots, y_{0N})^\top$.

Exact solution: $y(t) = \exp(-M^{-1}At)y_0, \quad t \geq 0$.

## Properties of $M$ and $A$

**Definiton:** Matrix $Q \geq 0 \overset{\text{def}}{\Longleftrightarrow} \forall i,j \; Q_{ij} \geq 0$.

**Definiton:** $\mathcal{Z} = \left\{K \in \mathbb{R}^{N \times N} : \forall i \neq j \; K_{ij} \leq 0, N \in \mathbb{N}\right\}$

- $M \geq 0$ (nonnegativity of FE basis functions)
- $M$, $A$ – Gramm matrices
  (nonsingular, symmetric, positive definite)
- $M$, $A$ – irreducible and sparse (if the mesh is sufficiently fine)
- $A \in \mathcal{Z}$ (acute type condition)
- $A^{-1} > 0$ ($A$ irreducible M-matrix)
- $M^{-1} \notin \mathcal{Z}$ (both positive and negative off-diagonal entries in $M^{-1}$)

## Semidiscrete nonnegative conservation

Recall semidiscrete solution: $y(t) = \exp(-M^{-1}At)y_0, \quad t \geq 0.$

semidiscrete problem conserves nonnegativity

$$\Updownarrow$$

$$y_0 \geq 0 \quad \Rightarrow \quad y(t) \geq 0 \text{ for all } t \geq 0$$

$$\Updownarrow$$

$$\exp(-M^{-1}At) \geq 0 \text{ for all } t \geq 0$$

## Preliminaries

**Theorem:** $Q \in \mathbb{R}^{N \times N}$ irreducible.

$$\exp(-Qt) \geq 0 \text{ for all } t \geq 0 \quad \Leftrightarrow \quad Q \in \mathcal{Z}$$

*Proof.* See [Varga, 1963], page 257, Theorem 8.1. □

| |
|---|
| semidiscrete problem conserves nonnegativity |
| $\Updownarrow$ |
| $M^{-1}A \in \mathcal{Z}$ (if $M^{-1}A$ irreducible) |

Recall: $\mathcal{Z} = \left\{ K \in \mathbb{R}^{N \times N} : \forall i \neq j \;\; K_{ij} \leq 0, N \in \mathbb{N} \right\}$

## Irreducibility of $M^{-1}A$

**Theorem:** $Q \in \mathbb{R}^{N\times N}$ nonsingular. $Q$ irreducible $\Leftrightarrow$ $Q^{-1}$ irreducible.

*Proof.* $Q$ reducible:

$$PQP^\top = \begin{pmatrix} A_1 & B \\ 0 & A_2 \end{pmatrix},$$

$$\left(PQP^\top\right)^{-1} = P^{-\top}Q^{-1}P^{-1} = \begin{pmatrix} A_1^{-1} & -A_1^{-1}BA_2^{-1} \\ 0 & A_2^{-1} \end{pmatrix},$$

$Q^{-1}$ reducible. $\square$

## Irreducibility of $M^{-1}A$

**Theorem:** $P, Q \in \mathbb{R}^{N\times N}$, $P \geq 0$, $Q \geq 0$. $P$ irreducible and $\text{diag}\, Q \neq 0$ $\Rightarrow$ $PQ$ and $QP$ irreducible.

*Proof.*

$$Q = \underbrace{D}_{\text{diag}\, Q} + \underbrace{O}_{\text{off}-\text{diag}\, Q}$$

$$\text{digraph}(P) = \text{digraph}(DP) = \text{digraph}(PD)$$

$$PQ = \underbrace{PD}_{\text{digraph}(P)} + \underbrace{PO}_{\text{addition edges}}$$

$\square$

$A$ irreducible $\overset{\text{Th.}}{\Longrightarrow} A^{-1}$ irreducible $\overset{\text{Th.}}{\Longrightarrow} A^{-1}M$ irreducible

$\overset{\text{Th.}}{\Longrightarrow} M^{-1}A$ irreducible

$$M^{-1}A \notin \mathcal{Z}$$

**Definiton:**

The set of matrices with zeros, where $M \in \mathbb{R}^{N\times N}$ has zeros:

$$\mathcal{M}_M = \left\{K \in \mathbb{R}^{N\times N} : \forall i,j \quad M_{ij} = 0 \;\Rightarrow\; K_{ij} = 0\right\}.$$

**Theorem:** $M \in \mathbb{R}^{N\times N}$ nonnegative, nonsingular, irreducible,
$\exists i \neq j \;\; M_{ij} = 0$, and $\forall k \;\; M_{kk} \neq 0$.
$A \in \mathcal{M}_M$ nonsingular, irreducible, $A^{-1} \geq 0$.
$\Rightarrow \quad M^{-1}A \notin \mathcal{Z}$.

Recall: $\mathcal{Z} = \left\{K \in \mathbb{R}^{N\times N} : \forall i \neq j \;\; K_{ij} \leq 0, N \in \mathbb{N}\right\}$

*Proof.* Assume that $M^{-1}A \in \mathcal{Z}$.

$$M^{-1}A = \underbrace{D}_{\text{diagonal}} - \underbrace{Q}_{\geq 0 \text{ with diag} \neq 0} \quad \Leftrightarrow \quad \underbrace{MD}_{\in \mathcal{M}_M} - \underbrace{A}_{\in \mathcal{M}_M} = \underbrace{MQ}_{\in \mathcal{M}_M}$$

$i \neq j$, $M_{ij} = 0$. $M$ irreducible $\Rightarrow \exists j_m \neq i \; M_{i,j_m} \neq 0$.
$M^{-1}A$ irreducible $\Rightarrow$ $Q$ irreducible $\Rightarrow$
$\Rightarrow \exists j_1, j_2, \ldots, j_m : Q_{j_1,j} \neq 0,\; Q_{j_2,j_1} \neq 0,\; \ldots,\; Q_{j_m,j_{m-1}} \neq 0$.

$$(MQ)_{ij} = M_{i,j_1} \underbrace{Q_{j_1,j}}_{\neq 0} + \sum_{k \neq j_1} \underbrace{M_{ik}}_{\geq 0} \underbrace{Q_{kj}}_{\geq 0} \quad \begin{cases} \neq 0 & \text{if} \quad M_{i,j_1} \neq 0 \Rightarrow * \\ = ? & \text{and} \quad M_{i,j_1} = 0 \end{cases}$$

$$(MQ)_{i,j_1} = M_{i,j_2} \underbrace{Q_{j_2,j_1}}_{\neq 0} + \underbrace{\ldots}_{\geq 0} \quad \begin{cases} \neq 0 & \text{if} \quad M_{i,j_2} \neq 0 \Rightarrow * \\ = ? & \text{and} \quad M_{i,j_2} = 0 \end{cases}$$

$$\vdots \qquad\qquad \vdots$$

$$\vdots \qquad\qquad M_{j,j_{m-1}} = 0$$

$$(MQ)_{i,j_{m-1}} = \underbrace{M_{i,j_m}}_{\neq 0} \underbrace{Q_{j_m,j_{m-1}}}_{\neq 0} + \underbrace{\ldots}_{\geq 0} \quad \neq 0 \Rightarrow *$$

$* \; MQ \notin \mathcal{M}_M$

$\square$

## Conclusion

$M^{-1}A \notin \mathcal{Z}$ (and $M^{-1}A$ irreducible)

$\Downarrow$

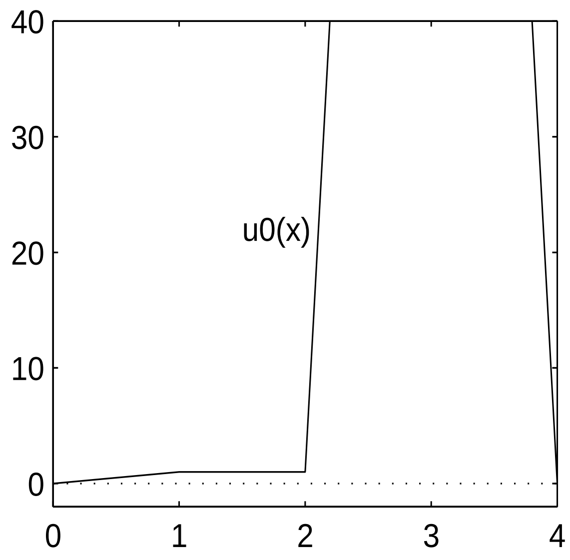
semidiscrete problem does not conserve nonnegativity

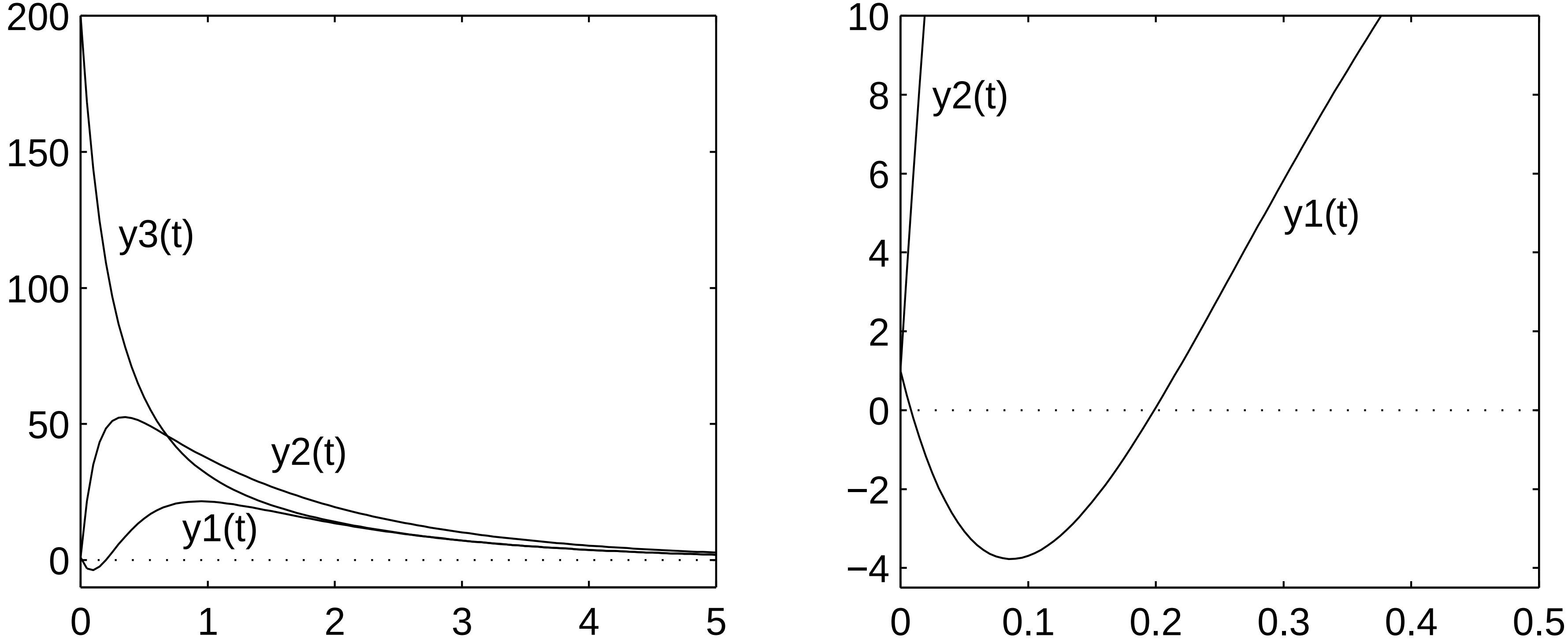
**Corollary:** If the simplicial partition of $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$, satisfies the acute type condition and if it is fine enough
then the semidiscrete problem does not conserve nonnegativity,
i.e., $\exists \bar{u}_{h0} \geq 0$ and $(x,t) \in \Omega \times (0,\infty)$ such that $\bar{u}_h(x,t) < 0$.

# Example in 1D

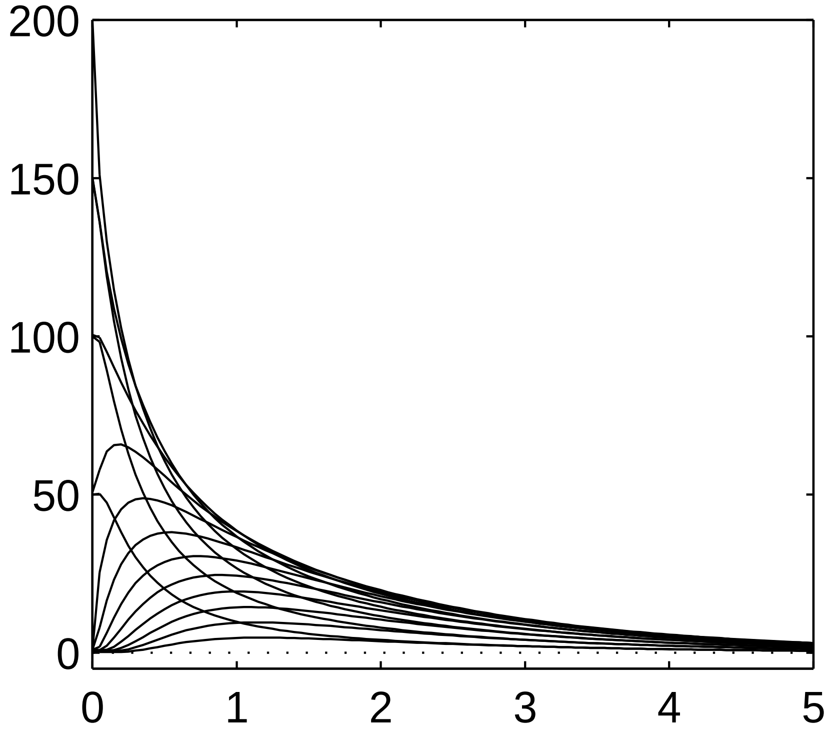
$$\dot{u}(x,t) - u'(x,t) = 0 \qquad \text{in } (0,4)\times(0,T),$$
$$u(0,t) = 0 \quad u(4,t) = 0 \qquad \text{for } t \geq 0,$$
$$u(x,0) = u_0(x) \qquad \text{in } (0,4),$$

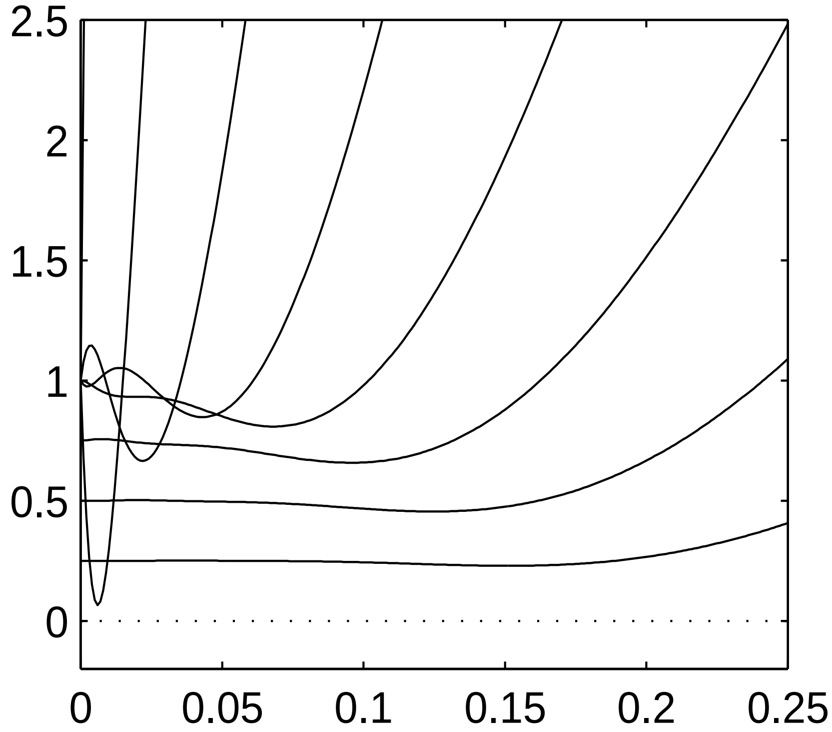
$$u_0(x) = \varphi_1(x) + \varphi_2(x) + 200\varphi_3(x)$$





200
150
100
50
0
y3(t)
y2(t)
y1(t)
0
1
2
3
4
5
10
8
6
4
2
0
−2
−4
y2(t)
y1(t)
0
0.1
0.2
0.3
0.4
0.5


200
150
100
50
0
0
1
2
3
4
5


2.5
2
1.5
1
0.5
0
0
0.05
0.1
0.15
0.2
0.25

## Nonnegativity for $t \geq t_0$

**Theorem:** Consider sufficiently fine mesh, acute type condition.
Then $\exists t_0 \in \mathbb{R}\ \forall t \geq t_0\ \forall \bar{u}_{h0} \geq 0 \quad \bar{u}_h(x,t) \geq 0$ in $\Omega$.

*Proof.*

$$\left[\exp(-M^{-1}At)\right]_{ij} = \underbrace{v_{i1}\bar{v}_{1j}}_{>0} \mathrm{e}^{-\lambda_1 t} + \sum_{k=2}^{N} v_{ik}\bar{v}_{kj}\mathrm{e}^{-\lambda_k t}, \quad i,j = 1,2,\ldots,N.$$

$V = (v_{ij})$ ... columns – eigenvectors of $M^{-1}A$
$V^{-1} = (\bar{v}_{ij})$
$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ □

## Concluding remarks

- Nonnegativity occurs only close to $t = 0$.

- Given initial condition
  $\Rightarrow$ sufficiently fine mesh $\Rightarrow$ nonnegative solution.

- Mass lumping.

**Thank you for your attention.**