Electronic Transactions on Numerical Analysis.
Volume 36, pp. 149-167, 2010.
Copyright © 2010, Kent State University.
ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

DISCRETE MAXIMUM PRINCIPLES FOR THE FEM SOLUTION OF SOME NONLINEAR PARABOLIC PROBLEMS*

ISTVÁN FARAGÓ†, JÁNOS KARÁTSON†, AND SERGEY KOROTOV‡

Dedicated to Richard S. Varga on the occasion of his 80th birthday

Abstract. Discrete maximum principles are established for finite element approximations of nonlinear parabolic problems. The conditions on the space and time discretizations are similar to the usual conditions for linear problems.

Key words. nonlinear parabolic problems, discrete maximum principle, finite element method

AMS subject classifications. 65M60, 65M50, 35B50

1. Introduction. The numerical approximations of solutions of models described by partial differential equations are naturally required to mirror some basic qualitative properties of the exact solutions. For parabolic equations, such a basic qualitative property is the (continuous) maximum principle (CMP). Several variants of CMPs exist; see, e.g., [16, 26]. Its discrete analogues, the so-called discrete maximum principles (DMPs) were first presented and analysed for the case of parabolic problems in the papers [17, 22]. If the finite element method (FEM) is employed for the spatial discretization, then the corresponding DMPs are normally ensured by imposing certain geometrical restrictions on the spatial meshes used; see, e.g., [7, 8, 10, 12, 17, 18, 19, 20, 24, 25, 29, 30] and the references therein. For elliptic problems, an extra requirement for the mesh is to provide irreducibility of the stiffness matrix [6]. This property is not always easy to ensure [8, p.5]. In contrast, there is no such requirement in the parabolic case, cf. [17, 12] and Theorem 4.1 below. For parabolic problems, the other most important condition to satisfy the DMP is that the time-steps often have to be chosen between certain lower and upper bounds with respect to the space mesh. In general, $\Delta t = O(h^2)$ must hold [10, 13]; this well-known requirement also appears for finite difference discretizations [11] and in the context of convergence in maximum norm [27]. A related important discrete qualitative property of the numerical solutions is the so-called nonnegativity preservation, analysed in the context of DMPs, e.g., in [10, 12].

In this paper, we prove discrete maximum principles for nonlinear parabolic problems, which has never been considered so far according to the authors' knowledge.

The paper is organized as follows. In Section 2, we formulate the nonlinear parabolic problem. The discretization scheme is given in detail in Section 3. Some preliminaries on linear problems and the maximum principle are given in Section 4. The DMP and related nonnegativity preservation, and the conditions for their validity are presented in Section 5: we first consider two types of growth conditions for the reaction terms, then we discuss sufficient geometric conditions on the FE meshes used, and finally we give two relevant real-life examples.

*Received March 16, 2009. Accepted for publication December 9, 2009. Published online on March 25, 2010. Recommended by J. Li. This work was supported by the Hungarian Research Grant OTKA no. K 67819, by HAS under the Bolyai János Scholarship, by the Academy Research Fellowship no. 208628 and project no. 124619 from the Academy of Finland.

†Department of Applied Analysis and Computational Mathematics, Eötvös Loránd University, H–1518, Budapest, Pf. 120, Hungary (farago, karatson@cs.elte.hu).

‡Department of Mathematics, Tampere University of Technology, P.O. Box 553, FI–33101 Tampere, Finland (sergey.korotov@tut.fi).

ETNA
Kent State University
http://etna.math.kent.edu

2. The problem. In the sequel, we consider the following mixed nonlinear parabolic problem. Find a function $u = u(x,t)$, such that

$$\frac{\partial u}{\partial t} - \operatorname{div}\Big(k(x,t,u,\nabla u)\nabla u\Big) + q(x,t,u) = f(x,t) \quad \text{in } Q_T := \Omega \times (0,T), \tag{2.1}$$

where Ω is a bounded domain in $\mathbb{R}^d$ and $T > 0$. The boundary and initial conditions are given by

$$u(x,t) = g(x,t) \quad \text{for} \quad (x,t) \in \Gamma_D \times [0,T], \tag{2.2}$$

$$k(x,t,u,\nabla u)\tfrac{\partial u}{\partial \nu} + s(x,t,u) = \gamma(x,t) \quad \text{for} \quad (x,t) \in \Gamma_N \times [0,T], \tag{2.3}$$

$$u(x,0) = u_0(x) \quad \text{for} \quad x \in \Omega, \tag{2.4}$$

respectively. We impose the following

ASSUMPTION 2.1.

(A1) Ω is a bounded polytopic domain in $\mathbb{R}^d$ with a Lipschitz continuous boundary $\partial\Omega$; $\Gamma_N, \Gamma_D \subset \partial\Omega$ are open sets, such that $\Gamma_N \cap \Gamma_D = \emptyset$ and $\overline{\Gamma}_N \cup \overline{\Gamma}_D = \partial\Omega$.

(A2) The scalar functions $k : \overline{Q}_T \times \mathbb{R}^{d+1} \to \mathbb{R}$, $q : \overline{Q}_T \times \mathbb{R} \to \mathbb{R}$ and $s : \overline{\Gamma}_N \times [0,T] \times \mathbb{R} \to \mathbb{R}$ are measurable and bounded. Moreover, q and s are continuously differentiable with respect to their last variable ξ, on their domains of definition. Further, $f \in L^\infty(Q_T)$, $\gamma \in L^2(\Gamma_N \times [0,T])$, $g \in L^\infty(\Gamma_D \times [0,T])$ and $u_0 \in L^\infty(\Omega)$.

(A3) There exist positive constants μ_0 and μ_1, such that

$$0 < \mu_0 \le k(x,t,\xi,\eta) \le \mu_1 \tag{2.5}$$

for all $(x,t,\xi,\eta) \in \Omega \times (0,T) \times \mathbb{R} \times \mathbb{R}^d$.

(A4) Let $2 \le p_1$ if $d = 2$, or $2 \le p_1 < \frac{2d}{d-2}$ if $d > 2$. Further, let $2 \le p_2 < 2.5$ if $d = 2$ or 3 and $p_2 = 2$ if $d > 3$. There exist constants $\alpha_1, \alpha_2, \beta_1, \beta_2 \ge 0$, such that for any $x \in \Omega$ (or $x \in \Gamma_N$, respectively), $t \in (0,T)$ and $\xi \in \mathbb{R}$,

$$0 \le \frac{\partial q(x,t,\xi)}{\partial \xi} \le \alpha_1 + \beta_1|\xi|^{p_1-2}, 0 \le \frac{\partial s(x,t,\xi)}{\partial \xi} \le \alpha_2 + \beta_2|\xi|^{p_2-2}. \tag{2.6}$$

We define weak solutions in the usual way as follows. Let $H^1_D(\Omega) := \{u \in H^1(\Omega) : u_{|\Gamma_D} = 0\}$. A function $u : Q_T \to \mathbb{R}$ is called the weak solution of the problem (2.1)–(2.4) if u is continuously differentiable with respect to t and $u(.,t) \in H^1_D(\Omega)$ for all $t \in (0,T)$ and satisfies the relation

$$\int_\Omega \frac{\partial u}{\partial t} v\,dx + \int_\Omega \Big(k(x,t,u,\nabla u)\nabla u \cdot \nabla v + q(x,t,u)v\Big)\,dx + \int_{\Gamma_N} s(x,t,u)v\,d\sigma \tag{2.7}$$
$$= \int_\Omega f v\,dx + \int_{\Gamma_N} \gamma v\,d\sigma, \qquad \forall v \in H^1_D(\Omega), \quad t \in (0,T).$$

Further,

$$u = g \quad \text{on} \quad [0,T] \times \Gamma_D, \qquad u|_{t=0} = u_0 \quad \text{in} \quad \Omega. \tag{2.8}$$

Here and in the sequel, equality of functions in Lebesgue or Sobolev spaces is understood almost everywhere.

ETNA
Kent State University
http://etna.math.kent.edu

3. Discretization scheme. The discretization of problem (2.1)–(2.4) is built up in a standard way. The presentation below is an adoption of the discretization in [12] to the nonlinear case.

3.1. Semidiscretization in space. Let $\mathcal{T}_h$ be a finite element mesh over the solution domain $\Omega \subset \mathbb{R}^d$, where h stands for the discretization parameter. We choose basis functions $\phi_1, ..., \phi_{\bar{m}}$, assumed to be continuous and to satisfy

$$\phi_i \geq 0 \quad (i = 1, \ldots, \bar{m}), \qquad \sum_{i=1}^{\bar{m}} \phi_i \equiv 1. \tag{3.1}$$

Let there exist node points $P_i \in \overline{\Omega}$, $i = 1, \ldots, \bar{m}$, such that

$$\phi_i(P_j) = \delta_{ij}, \tag{3.2}$$

where δ_{ij} is the Kronecker symbol. (These conditions hold, e.g., for standard linear, bilinear or prismatic FEM.) Let V_h denote the finite element subspace spanned by the above basis functions:

$$V_h = \text{span}\{\phi_1, ..., \phi_{\bar{m}}\} \subset H^1(\Omega).$$

Now, let $m < \bar{m}$ be such that

$$P_1, ..., P_m \tag{3.3}$$

are the vertices that lie in Ω or on Γ_N, and let

$$P_{m+1}, ..., P_{\bar{m}} \tag{3.4}$$

be the vertices that lie on $\overline{\Gamma}_D$. Then the basis functions $\phi_1, ..., \phi_m$ satisfy the homogeneous Dirichlet boundary condition on Γ_D, i.e., $\phi_i \in H^1_D(\Omega)$. We define

$$V_h^0 = \text{span}\{\phi_1, ..., \phi_m\} \subset H^1_D(\Omega).$$

Then the semidiscrete problem for (2.7) with initial-boundary conditions (2.8) reads as follows: find a function $u_h = u_h(x,t)$, such that

$$u_h(x,0) = u_0^h(x), \quad x \in \Omega,$$

$$u_h(.,t) - g_h(.,t) \in V_0^h, \quad t \in (0,T),$$

and

$$\int_\Omega \frac{\partial u_h}{\partial t} v_h \, dx + \int_\Omega \Big(k(x,t,u_h,\nabla u_h)\nabla u_h \cdot \nabla v_h + q(x,t,u_h) v_h \Big) \, dx + \int_{\Gamma_N} s(x,t,u_h) v_h \, d\sigma$$

$$= \int_\Omega f v_h \, dx + \int_{\Gamma_N} \gamma v \, d\sigma, \qquad \forall v_h \in V_0^h, \quad t \in (0,T). \tag{3.5}$$

ETNA
Kent State University
http://etna.math.kent.edu

In the above formulae, the functions u_0^h and $g_h(.,t)$ (for any fixed t) are suitable approximations of the given functions u_0 and $g(.,t)$, respectively. In particular, we will use the following form to describe g_h:

$$g_h(x,t) = \sum_{i=1}^{m_\partial} g_i^h(t)\phi_{m+i}(x), \tag{3.6}$$

where

$$m_\partial := \bar{m} - m.$$

We note that, based on the consistency of the initial and boundary conditions ($g(s,0) = u_0(s), s \in \partial\Omega$), we obtain

$$g(P_{m+i},0) = u_0(P_{m+i}), \quad i = 1,\dots,m_\partial.$$

We seek the numerical solution of the form

$$u_h(x,t) = \sum_{i=1}^{m} u_i^h(t)\phi_i(x) + g_h(x,t) \tag{3.7}$$

and notice that it is sufficient that u_h satisfies (3.5) for $v_h = \phi_i$, $i = 1,2,\dots,m$, only. Then, introducing the notation

$$\mathbf{u}^h(t) = [u_1^h(t),\dots,u_m^h(t), g_1^h(t),\dots,g_{m\partial}^h(t)]^T, \tag{3.8}$$

we are led to the following Cauchy problem for the system of ordinary differential equations:

$$\mathbf{M}\frac{\mathrm{d}\mathbf{u}^h}{\mathrm{d}t} + \mathbf{G}(\mathbf{u}^h(t)) = \mathbf{f}(t), \tag{3.9}$$

$$\mathbf{u}^h(0) = \mathbf{u}_0^h = [u_0(P_1),\dots,u_0(P_m), g_1^h(0),\dots,g_{m\partial}^h(0)]^T, \tag{3.10}$$

where

$$\mathbf{M} = [M_{ij}]_{m\times\bar{m}}, \quad M_{ij} = \int_\Omega \phi_j(x)\phi_i(x)\,dx, \tag{3.11}$$

$$\mathbf{G}(\mathbf{u}^h(t)) = [G(\mathbf{u}^h(t))_i]_{i=1,\dots,m},$$

$$G(\mathbf{u}^h(t))_i = \int_\Omega \Big(k(x,t,u_h,\nabla u_h)\nabla u_h\cdot\nabla\phi_i + q(x,t,u_h)\phi_i\Big)\,dx + \int_{\Gamma_N} s(x,t,u_h)\phi_i\,d\sigma(x),$$

$$\mathbf{f}(t) = [f_i(t)]_{i=1,\dots,m}, \quad f_i(t) = \int_\Omega f(x,t)\phi_i(x)\,dx + \int_{\Gamma_N}\gamma(x,t)\phi_i(x)\,d\sigma(x).$$

The solution $\mathbf{u}^h = \mathbf{u}^h(t)$ of problem (3.9)–(3.10) is called the semidiscrete solution. Its existence and uniqueness is ensured by Assumption 2.1, since then $\mathbf{G}$ is locally Lipschitz continuous.

ETNA
Kent State University
http://etna.math.kent.edu

3.2. Full discretization. In order to get a fully discrete numerical scheme, we choose a time-step Δt and denote the approximation to $\mathbf{u}^h(n\Delta t)$ and $\mathbf{f}(n\Delta t)$ by $\mathbf{u}^n$ and $\mathbf{f}^n$ (for $n = 0, 1, 2, \ldots, n_T$, where $n_T \Delta t = T$), respectively. To discretize (3.9) in time, we apply the so-called θ-method with some given parameter $\theta \in (0, 1]$.

We note that the case $\theta = 0$, which is otherwise also acceptable, will be excluded later by condition (5.16). This gives no strong difference, since the presence of $\mathbf{M}$ makes the scheme not explicit even for $\theta = 0$.

We then obtain a system of nonlinear algebraic equations of the form

$$\mathbf{M}\frac{\mathbf{u}^{n+1} - \mathbf{u}^n}{\Delta t} + \theta \mathbf{G}(\mathbf{u}^{n+1}) + (1-\theta)\mathbf{G}(\mathbf{u}^n) = \mathbf{f}^{(n,\theta)} := \theta \mathbf{f}^{n+1} + (1-\theta)\mathbf{f}^n, \tag{3.12}$$

$n = 0, 1, \ldots, n_T - 1$, which can be rewritten as a recursion

$$\mathbf{M}\mathbf{u}^{n+1} + \theta \Delta t \mathbf{G}(\mathbf{u}^{n+1}) = \mathbf{M}\mathbf{u}^n - (1-\theta)\Delta t \mathbf{G}(\mathbf{u}^n) + \Delta t\, \mathbf{f}^{(n,\theta)}, \tag{3.13}$$

with $\mathbf{u}^0 = \mathbf{u}^h(0)$. Furthermore, we will use the notation

$$\mathbf{P}(\mathbf{u}^{n+1}) := \mathbf{M}\mathbf{u}^{n+1} + \theta \Delta t \mathbf{G}(\mathbf{u}^{n+1}), \;\; \mathbf{Q}(\mathbf{u}^n) := \mathbf{M}\mathbf{u}^n - (1-\theta)\Delta t \mathbf{G}(\mathbf{u}^n). \tag{3.14}$$

Then, the iteration procedure (3.13) can be also written as

$$\mathbf{P}(\mathbf{u}^{n+1}) = \mathbf{Q}(\mathbf{u}^n) + \Delta t\, \mathbf{f}^{(n,\theta)}. \tag{3.15}$$

We note that finding $\mathbf{u}^{n+1}$ in (3.15) requires the solution of a nonlinear algebraic system. The mass matrix $\mathbf{M}$ is positive definite, and it follows from Assumption 2.1 that $\mathbf{u} \mapsto \mathbf{G}(\mathbf{u})$ has positive semidefinite derivatives. Therefore, by the definition in (3.14), the function $\mathbf{u} \mapsto \mathbf{P}(\mathbf{u})$ has regular derivatives. This ensures the unique solvability of (3.15) and, under standard local Lipschitz conditions on the coefficients, also the convergence of the damped Newton iteration; see, e.g., [14].

4. Preliminaries: linear problems and the maximum principle. An important and widely studied special case of (2.1)–(2.4) is the linear problem with Dirichlet boundary conditions

$$\frac{\partial u}{\partial t} - k\Delta u + c(x,t)u = f(x,t), \tag{4.1}$$

$$u = g \quad \text{on} \quad [0,T] \times \partial\Omega, \qquad u|_{t=0} = u_0 \quad \text{in} \quad \Omega, \tag{4.2}$$

where $k > 0$ is constant and $c \geq 0$. If the data and solution are assumed to be sufficiently smooth, then problem (4.1)–(4.2) is known to satisfy the *continuous maximum principle*, see [12], which is a starting point for our study:

$$\begin{aligned} &\min\{0; \min_{\Gamma_{t_1}} u\} + t_1 \min\{0; \min_{Q_{t_1}} f\} \leq u(x,t_1) \leq \\ &\leq \max\{0; \max_{\Gamma_{t_1}} u\} + t_1 \max\{0; \max_{Q_{t_1}} f\} \end{aligned} \tag{4.3}$$

for all $x \in \Omega$ and any fixed $t_1 \in (0,T)$, where $Q_{t_1} := \Omega \times [0,t_1]$, and Γ_{t_1} denotes the parabolic boundary, i.e., $\Gamma_{t_1} := (\partial\Omega \times [0,t_1]) \cup (\Omega \times \{0\})$. A related property, which follows

ETNA
Kent State University
http://etna.math.kent.edu

from the above [11], is the *continuous nonnegativity preservation principle*: relations $f \geq 0$, $g \geq 0$ and $u_0 \geq 0$ imply that

$$u(x,t) \geq 0 \tag{4.4}$$

for all $(x,t) \in Q_T$.

In the discrete case, the ODE system (3.9) now becomes a linear system

$$\mathbf{M}\frac{\mathrm{d}\mathbf{u}^h}{\mathrm{d}t} + \mathbf{K}\mathbf{u}^h(t) = \mathbf{f}, \tag{4.5}$$

where $\mathbf{K}_{ij} = \int\limits_{\Omega} \Big(k\,\nabla\phi_i \cdot \nabla\phi_j + c\phi_i\phi_j \Big)$. The full discretization is

$$\mathbf{M}\frac{\mathbf{u}^{n+1} - \mathbf{u}^n}{\Delta t} + \theta\mathbf{K}\mathbf{u}^{n+1} + (1-\theta)\mathbf{K}\mathbf{u}^n = \mathbf{f}^{(n,\theta)} := \theta\mathbf{f}^{n+1} + (1-\theta)\mathbf{f}^n. \tag{4.6}$$

Then (3.14)–(3.15) can be simplified: introducing the matrices

$$\mathbf{A} := \mathbf{M} + \theta\Delta t\mathbf{K}, \qquad \mathbf{B} := \mathbf{M} - (1-\theta)\Delta t\mathbf{K}, \tag{4.7}$$

equation (4.6) now can be rewritten as

$$\mathbf{A}\mathbf{u}^{n+1} = \mathbf{B}\mathbf{u}^n + \Delta t\,\mathbf{f}^{(n,\theta)}. \tag{4.8}$$

To formulate the discrete maximum principle, let us define the following values:

$$g^n_{min} = \min\{g^n_1, \dots, g^n_{m_\partial}\}, \quad g^n_{max} = \max\{g^n_1, \dots, g^n_{m_\partial}\}, \tag{4.9}$$

$$u^n_{min} = \min\{g^n_{min}, u^n_1, \dots, u^n_m\}, \quad u^n_{max} = \max\{g^n_{max}, u^n_1, \dots, u^n_m\}, \tag{4.10}$$

for $n = 0, 1, \dots, n_T$, and

$$f^{(n,n+1)}_{min} := \inf_{\substack{x\in\Omega,\\ \tau\in(n\Delta t,(n+1)\Delta t)}} f(x,\tau), \qquad f^{(n,n+1)}_{max} := \sup_{\substack{x\in\Omega,\\ \tau\in(n\Delta t,(n+1)\Delta t)}} f(x,\tau), \tag{4.11}$$

for $n = 0, 1, \dots, n_T - 1$. If f is only in $L^\infty(\Omega)$, then the above infima and suprema will mean essential infima and suprema, respectively. Then the discrete analogue of the continuous maximum principle (4.3) can be formulated as follows:

$$\begin{aligned} &\min\{0, g^{(n+1)}_{min}, u^{(n)}_{min}\} + \Delta t \min\{0, f^{(n,n+1)}_{min}\} \leq \\ &u^{n+1}_i \leq \max\{0, g^{(n+1)}_{max}, u^{(n)}_{max}\} + \Delta t \max\{0, f^{(n,n+1)}_{max}\}. \end{aligned} \tag{4.12}$$

This will be denoted by DMP and it corresponds to the continuous maximum principle for one time-level, i.e., when $t_1 \in [n\Delta t, (n+1)\Delta t]$.

It was proved that the full discretization of the linear problem satisfies the DMP (4.12) in the following case:

THEOREM 4.1. [17, 12] *Let the basis functions satisfy (3.1)–(3.2), and let the following conditions hold for the matrices (4.7):*

(i) $A_{ij} \leq 0 \quad (i \neq j,\ i = 1, \dots, m,\ j = 1, \dots, \bar{m})$;

(ii) $B_{ii} \geq 0 \quad (i = 1, \dots, m)$.

ETNA
Kent State University
http://etna.math.kent.edu

Then the Galerkin solution of the problem (4.1)–(4.2), combined with the θ-method in the time discretization, satisfies the discrete maximum principle (4.12).

We note that in the original form (see, e.g., [12, Theorem 6]) it is also assumed that $K_{ij} \leq 0$ $(i \neq j,\ i = 1, ..., m,\ j = 1, ..., \bar{m})$. However, now by our assumption $\theta > 0$, using (3.1) and (3.11) we have $M_{ij} \geq 0$. Hence it follows from assumption (i) and (4.7) that $K_{ij} = (1/\theta\Delta t)(A_{ij} - M_{ij}) \leq 0$.

The above result has been extended recently to mixed boundary value problems [13]. Let the boundary conditions in (4.2) be replaced by

$$u = g \quad \text{on} \quad [0,T] \times \Gamma_D, \qquad k\,\nabla u \cdot \nu = q \quad \text{on} \quad [0,T] \times \Gamma_N^0, \tag{4.13}$$

$$k\,\nabla u \cdot \nu + \sigma u = \varrho \quad \text{on} \quad [0,T] \times \Gamma_N^1,$$

where $\sigma > 0$ is constant. If the conditions of Theorem 4.1 hold and $q \leq 0$, then

$$u_i^{n+1} \leq \max\{0, g_{max}^{(n+1)}, u_{max}^{(n)}\} + \Delta t \max\{0, f_{max}^{(n,n+1)}\} + \frac{1}{\theta}\max\{0, \Big(\frac{\varrho}{\sigma}\Big)_{max}^{(n,n+1)}\}. \tag{4.14}$$

In [13] a constant σ is considered for simplicity, in which case σ is simply a constant factor above and $\varrho_{max}^{(n,n+1)}$ is defined analogously to (4.11). However, their proof can be rewritten exactly in the same way for a variable coefficient $\sigma = \sigma(x,\tau)$, simply by estimating ϱ/σ by its suprema, in which case we have the DMP (4.14) with

$$\Big(\frac{\varrho}{\sigma}\Big)_{max}^{(n,n+1)} := \sup_{\substack{x \in \Gamma_N^1, \\ \tau \in (n\Delta t, (n+1)\Delta t)}} \frac{\varrho(x,\tau)}{\sigma(x,\tau)}. \tag{4.15}$$

REMARK 4.2. The indices $1, ..., m$ that arise in (4.10) now correspond to node points in the interior of Ω or on Γ_N, as in (3.3), and accordingly, the other m_∂ indices involved in $g_{max}^{(n+1)}$ in (4.14) correspond to the values on Γ_D. That is, whereas the DMP (4.12) involves the values of g on $\partial\Omega$, the DMP (4.14) involves the values of g on Γ_D only.

REMARK 4.3. Before we turn from linear problems to the nonlinear case, we note that the above-mentioned CMP results directly imply the same form of the CMP for some nonlinear problems as well. Namely, since the solution u is fixed, first we can replace $c(x,t)$ by $c(x,t,u(x))$ which just means that we have another fixed variable coefficient of u. Therefore, if $c(x,t,u(x)) \geq 0$ then the same CMP (4.3) holds. More generally, monotone nonlinear lower order terms can be rewritten as a nonlinear coefficient multiplied by the solution u, and we can then use the above result. Details of this technique are given below, since it will be used as well in our study of the DMP in the sequel.

5. The discrete maximum principle for the nonlinear problem.

5.1. Reformulation of the problem.

We can rewrite problem (2.7) as follows. Let

$$r(x,t,\xi) := \int_0^1 \frac{\partial q}{\partial \xi}(x,t,\alpha\xi)\,d\alpha, \quad z(x,t,\xi) := \int_0^1 \frac{\partial s}{\partial \xi}(x,t,\alpha\xi)\,d\alpha, \tag{5.1}$$

for any $x \in \Omega$, $t > 0$, $\xi \in \mathbb{R}$, and

$$\hat{f}(x,t) := f(x,t) - q(x,t,0), \quad \hat{\gamma}(x,t) := \gamma(x,t) - s(x,t,0),$$

ETNA
Kent State University
http://etna.math.kent.edu

for any $x \in \Omega$, or respectively, $x \in \Gamma_N$, $t > 0$. Then the Newton-Leibniz formula yields for all x, t, ξ that

$$q(x,t,\xi) - q(x,t,0) = r(x,t,\xi)\,\xi, \qquad s(x,t,\xi) - s(x,t,0) = z(x,t,\xi)\,\xi.$$

Subtracting $q(x,t,0)$ and $s(x,t,0)$ from (2.1) and (2.3), respectively, we thus obtain that problem (2.7) is equivalent to

$$\int_\Omega \frac{\partial u}{\partial t} v\,dx + B(u;u,v) = \int_\Omega \hat{f} v\,dx + \int_{\Gamma_N} \hat{\gamma} v\,d\sigma, \qquad \forall v \in H^1_D(\Omega), \quad t \in (0,T), \tag{5.2}$$

where

$$B(w;u,v) := \int_\Omega \Big(k(x,t,w,\nabla w)\nabla u \cdot \nabla v + r(x,t,w)uv \Big)\,dx + \int_{\Gamma_N} z(x,t,w)uv\,d\sigma, \tag{5.3}$$

for any $w, u, v \in H^1_D(\Omega)$. The semidiscretization of the problem reads as follows: find a function $u_h = u_h(x,t)$, such that

$$u_h(x,0) = u_0^h(x), \quad x \in \Omega,$$

$$u_h(.,t) - g_h(.,t) \in V_0^h, \quad t \in (0,T),$$

and

$$\int_\Omega \frac{\partial u_h}{\partial t} v_h\,dx + B(u_h;u_h,v_h) = \int_\Omega \hat{f} v_h\,dx + \int_{\Gamma_N} \hat{\gamma} v_h\,d\sigma, \qquad \forall v_h \in V_0^h, \quad t \in (0,T).$$

Proceeding as in (3.7)–(3.9), the Cauchy problem for the system of ordinary differential equations (3.9) takes the following form:

$$\mathbf{M}\frac{\mathrm{d}\mathbf{u}^h}{\mathrm{d}t} + \mathbf{K}(\mathbf{u}^h)\mathbf{u}^h = \hat{\mathbf{f}}, \tag{5.4}$$

$$\mathbf{u}^h(0) = \mathbf{u}_0^h = [u_0(P_1), \ldots, u_0(P_m), g_1^h(0), \ldots, g_{m_\partial}^h(0)]^T, \tag{5.5}$$

where $\mathbf{M}$ is as in (3.9),

$$\mathbf{K}(\mathbf{u}^h) = [K(\mathbf{u}^h)_{ij}]_{m\times\bar{m}}, \quad K(\mathbf{u}^h)_{ij} = B(u_h;\phi_j,\phi_i),$$

$$\hat{\mathbf{f}}(t) = [\hat{f}_i(t)]_{i=1,\ldots,m}, \quad \hat{f}_i(t) = \int_\Omega \hat{f}(x,t)\phi_i(x)\,dx + \int_{\Gamma_N} \hat{\gamma}(x,t)\phi_i(x)\,d\sigma(x). \tag{5.6}$$

The full discretization reads as

$$\mathbf{M}\mathbf{u}^{n+1} + \theta\Delta t\mathbf{K}(\mathbf{u}^{n+1})\mathbf{u}^{n+1} = \mathbf{M}\mathbf{u}^n - (1-\theta)\Delta t\mathbf{K}(\mathbf{u}^n)\mathbf{u}^n + \Delta t\,\hat{\mathbf{f}}^{(n,\theta)}. \tag{5.7}$$

Since we have set $\mathbf{G}(\mathbf{u}^h) = \mathbf{K}(\mathbf{u}^h)\mathbf{u}^h$ in (3.9), the expressions (3.14)–(3.15) become

$$\mathbf{P}(\mathbf{u}^{n+1}) = \big(\mathbf{M} + \theta\Delta t\mathbf{K}(\mathbf{u}^{n+1})\big)\,\mathbf{u}^{n+1}, \qquad \mathbf{Q}(\mathbf{u}^n) = \big(\mathbf{M} - (1-\theta)\Delta t\,\mathbf{K}(\mathbf{u}^n)\big)\,\mathbf{u}^n,$$

respectively. Letting

$$\mathbf{A}(\mathbf{u}^h) := \mathbf{M} + \theta\Delta t\mathbf{K}(\mathbf{u}^h),\ \mathbf{B}(\mathbf{u}^h) := \mathbf{M} - (1-\theta)\Delta t\mathbf{K}(\mathbf{u}^h), \qquad \forall \mathbf{u}^h \in \mathbb{R}^{\bar{m}}, \tag{5.8}$$

the iteration procedure (5.7) takes the form

$$\mathbf{A}(\mathbf{u}^{n+1})\mathbf{u}^{n+1} = \mathbf{B}(\mathbf{u}^n)\mathbf{u}^n + \Delta t\,\hat{\mathbf{f}}^{(n,\theta)}, \tag{5.9}$$

which is similar to (4.8), but now the coefficient matrices depend on $\mathbf{u}^{n+1}$ or on $\mathbf{u}^n$.

ETNA
Kent State University
http://etna.math.kent.edu

5.2. The DMP: problems with sublinear growth. Let us consider Assumption 2.1, where we let $p_1 = p_2 = 2$ in (A4), i.e., we have instead

(A4′) There exist constants $\alpha_1, \alpha_2 \geq 0$ such that for any $x \in \Omega$ (or $x \in \Gamma_N$), $t \in (0, T)$ and $\xi \in \mathbb{R}$,

$$0 \leq \frac{\partial q(x,t,\xi)}{\partial \xi} \leq \alpha_1, \qquad 0 \leq \frac{\partial s(x,t,\xi)}{\partial \xi} \leq \alpha_2. \tag{5.10}$$

In what follows, we will need the standard notion of (patch-)regularity of the considered meshes, cf. [3].

DEFINITION 5.1. Let $\Omega \subset \mathbb{R}^d$ and let us consider a family of FEM subspaces $\mathcal{V} = \{V_h\}_{h\to 0}$. The corresponding family of FE meshes will be called *regular* if there exist constants $c_0, c_1 > 0$, such that for any $h > 0$ and basis function ϕ_p,

$$c_1 h^d \leq \text{meas}(\text{supp}\,\phi_p), \quad \text{diam}(\text{supp}\,\phi_p) \leq c_0 h, \tag{5.11}$$

where $meas$ denotes d-dimensional measure and $supp$ denotes the support, i.e., the closure of the set where the function does not vanish, and

$$\text{meas}(\partial(\text{supp}\,\phi_p)) \leq c_2 h^{d-1}, \tag{5.12}$$

where $meas$ denotes $(d-1)$-dimensional measure of the boundary of $\text{supp}\,\phi_p$.

Note that (5.11) also implies

$$\text{meas}(\text{supp}\,\phi_p) \leq c_3 h^d. \tag{5.13}$$

In fact, (5.12) also follows from (5.11) under certain natural but additional assumptions, e.g., if $\text{supp}\,\phi_p$ is convex, as is the case for linear, bilinear or prismatic elements.

THEOREM 5.2. *Let problem (2.1)–(2.4) satisfy Assumption 2.1, such that we let $p_1 = p_2 = 2$ in (2.6), i.e., (A4) reduces to (A4′) above. Let us consider a family of finite element subspaces $\mathcal{V} = \{V_h\}_{h\to 0}$, where the basis functions satisfy (3.1)–(3.2) and the family of associated FE meshes is regular as in Definition 5.1. Let the following assumptions hold:*

(i) for any $i = 1, ..., m,\ j = 1, ..., \bar{m}$ $(i \neq j)$, if $\text{meas}(\text{supp}\,\phi_i \cap \text{supp}\,\phi_j) > 0$, *then*

$$\nabla\phi_i \cdot \nabla\phi_j \leq 0 \ \text{ on } \Omega \quad \text{and} \quad \int_\Omega \nabla\phi_i \cdot \nabla\phi_j \leq -K_0\, h^{d-2}, \tag{5.14}$$

with some constant $K_0 > 0$ independent of i, j and h;

(ii) the mesh parameter h satisfies

$$h < h_0 := \frac{2\mu_0 K_0}{c_2\alpha_2 + \sqrt{c_2^2\alpha_2^2 + 4\mu_0 K_0 c_2 \alpha_1}}; \tag{5.15}$$

(iii) we have

$$\Delta t \geq \frac{c_2 h^2}{\theta\big(\mu_0 K_0 - \alpha_1 c_2 h^2 - \alpha_2 c_2 h\big)}; \tag{5.16}$$

(iv) if $\theta < 1$, then

$$\Delta t \leq \frac{1}{(1-\theta)\, R(h)}, \tag{5.17}$$

ETNA
Kent State University
http://etna.math.kent.edu

where

$$R(h) := \mu_1 N(h) + \alpha_2\, G(h) + \alpha_1, \tag{5.18}$$

using the notation

$$N(h) := \max_{i=1,\dots,m} \frac{\int_\Omega |\nabla \phi_i|^2}{\int_\Omega \phi_i^2}, \qquad G(h) := \max_{i=1,\dots,m} \frac{\int_{\Gamma_N} \phi_i^2}{\int_\Omega \phi_i^2}. \tag{5.19}$$

Then for all $\mathbf{u}^h \in \mathbb{R}^{\bar m}$*, the matrices* $\mathbf{A}(\mathbf{u}^h)$ *and* $\mathbf{B}(\mathbf{u}^h)$*, defined in (5.8), have the following properties:*

(1) $A(\mathbf{u}^h)_{ij} \le 0 \quad (i \ne j,\ i = 1, \dots, m,\ j = 1, \dots, \bar m)$*;*

(2) $B(\mathbf{u}^h)_{ii} \ge 0 \quad (i = 1, \dots, m)$*.*

Proof. (1) We have

$$A(\mathbf{u}^h)_{ij} := M_{ij} + \theta \Delta t K(\mathbf{u}^h)_{ij} = \int_\Omega \phi_j \phi_i \; + \; \theta \Delta t B(u_h; \phi_j, \phi_i) \tag{5.20}$$

$$= \int_\Omega \phi_j \phi_i + \theta \Delta t \left[\int_\Omega \Big(k(x,t,u_h,\nabla u_h) \nabla \phi_j \cdot \nabla \phi_i + r(x,t,u_h) \phi_j \phi_i \Big) + \int_{\Gamma_N} z(x,t,u_h) \phi_j \phi_i \right].$$

Let $\Omega_{ij} := \operatorname{supp} \phi_i \cap \operatorname{supp} \phi_j$ and $\Gamma_{ij} := \partial \Omega_{ij}$. Here, by (3.1) and (5.12)-(5.13),

$$\int_\Omega \phi_j \phi_i \le \operatorname{meas}(\Omega_{ij}) \le c_2 h^d \quad \text{and} \quad \int_{\Gamma_N} \phi_j \phi_i \le \operatorname{meas}(\Gamma_{ij}) \le c_2 h^{d-1} \tag{5.21}$$

and, similarly,

$$\int_\Omega r(x,t,u_h) \phi_j \phi_i \le \alpha_1 c_2 h^d \quad \text{and} \quad \int_{\Gamma_N} z(x,t,u_h) \phi_j \phi_i \le \alpha_2 c_2 h^{d-1}, \tag{5.22}$$

since by (5.1), r and z inherit (5.10). By (2.5) and (5.14),

$$\int_\Omega k(x,t,u_h,\nabla u_h) \nabla \phi_j \cdot \nabla \phi_i \le -\mu_0 K_0\, h^{d-2}. \tag{5.23}$$

Altogether, we obtain

$$A(\mathbf{u}^h)_{ij} \le c_2 h^d \left[1 + \; \theta \Delta t \left(-\frac{\mu_0 K_0}{c_2} \frac{1}{h^2} \; + \alpha_1 + \frac{\alpha_2}{h} \right) \right].$$

Since $h < h_0$ for h_0 defined in (5.15), it readily follows that we have a negative coefficient of $\theta \Delta t$ above, and from (5.16) we obtain that the expression in the large brackets is nonpositive. Hence $A(\mathbf{u}^h)_{ij} \le 0$.

(2) Analogously to (5.20), we have

$$B(\mathbf{u}^h)_{ii} := M_{ii} - (1-\theta) \Delta t K(\mathbf{u}^h)_{ii} \ge 0,$$

if and only if

$$\int_\Omega \phi_i^2 \ge (1-\theta) \Delta t \left[\int_\Omega \Big(k(x,t,u_h,\nabla u_h) |\nabla \phi_i|^2 + r(x,t,u_h) \phi_i^2 \Big) + \int_{\Gamma_N} z(x,t,u_h)\, \phi_i^2 \right]. \tag{5.24}$$

ETNA
Kent State University
http://etna.math.kent.edu

The latter holds for all Δt if $\theta = 1$ (i.e. the scheme is implicit). If $\theta < 1$, we estimate the expression in brackets in (5.24) by

$$\int_\Omega \left(\mu_1 |\nabla \phi_i|^2 + \alpha_1 \phi_i^2\right) + \int_{\Gamma_N} \alpha_2 \, \phi_i^2 \leq \; R(h) \cdot \int_\Omega \phi_i^2.$$

It readily follows that (5.24) holds for all Δt that satisfies (5.18). ∎

Now we can derive the corresponding *discrete maximum principle:*

COROLLARY 5.3. *Let the conditions of Theorem 5.2 hold, and let* $\hat{\gamma}(x,t) := \gamma(x,t) - s(x,t,0) \leq 0$. *Then*

$$u_i^{n+1} \leq \; \max\{0, g_{max}^{(n+1)}, u_{max}^{(n)}\} + \Delta t \max\{0, \hat{f}_{max}^{(n,n+1)}\}. \tag{5.25}$$

Proof. Our reformulated problem has the right-hand side $\hat{f}(x,t) := f(x,t) - q(x,t,0)$, which is in $L^\infty(Q_T)$ by Assumption 2.1 (A2). Further, by (5.2)–(5.3), we have the Neumann boundary condition

$$k(x,t,u,\nabla u)\, \nabla u \cdot \nu + z(x,t,u) u = \hat{\gamma}(x,t) \qquad \text{on } \Gamma_N,$$

where $z \geq 0$ and $\hat{\gamma} \leq 0$. We can rewrite our boundary conditions to match (4.13): let Γ_N^0 and Γ_N^1 be the portions where $z \equiv 0$ and $z > 0$, respectively. Then, by assumption, $q := \hat{\gamma}_{|\Gamma_N^0} \leq 0$ and $\varrho := \hat{\gamma}_{|\Gamma_N^1} \leq 0$. Therefore (4.14) can be applied (with $\hat{f}$) and its last term can be dropped, whence we obtain (5.25). ∎

REMARK 5.4. Note that the DMP (5.25) involves the values of g on Γ_D; see also Remark 4.2. Besides that, (5.25) is formally identical to the upper part of (4.12), and could in fact be derived from it directly as an alternate proof. Namely, one can apply Theorem 4.1 as an algebraic result for the ODE system (5.4). Here $\mathbf{f}$ is replaced by $\hat{\mathbf{f}}$, which also involves the values of $\hat{\gamma}$; see (5.6). However, by our assumption $\hat{\gamma} \leq 0$. Therefore, we obtain a further upper bound by dropping the integrals with $\hat{\gamma}$, and we are thus led to (5.25).

REMARK 5.5. For various popular finite elements one has $N(h) = O(h^{-2})$ and $G(h) = O(h^{-1})$ in (5.19). Therefore,

$$R(h) = O(h^{-2})$$

in (5.18). Let us first consider $N(h)$. Exact formulae for $\int_T |\nabla \phi_i|^2$ and $\int_T \phi_i^2$ on elements T have been derived for simplicial elements in any dimension ([2], [5, p. 201]) bilinear elements in 2D [12] and prismatic elements in 3D [28] showing that

$$\int_T |\nabla \phi_i|^2 = O(h^{d-2}) \quad \text{and} \quad \int_T \phi_i^2 = O(h^d). \tag{5.26}$$

This immediately yields $N(h) = O(h^{-2})$, since the integral over Ω equals the integral over the support which consists of a bounded number of elements. For $G(h)$, we have $\int_{\Gamma_N} \phi_i^2 \leq \int_{\partial(\operatorname{supp} \varphi_i)} \phi_i^2$. The latter is an integral over the finite union of $(d-1)$-dimensional elements of the above types. Hence we can apply the $(d-1)$-dimensional formula to obtain $\int_{\Gamma_N} \phi_i^2 \leq O(h^{d-1})$. This implies $G(h) = O(h^{-1})$.

REMARK 5.6. *(Discussion of the assumptions in Theorem 5.2.)*

(i) Assumption (i) can be ensured by suitable geometric properties of the space mesh; see subsection 5.4 below.

(ii) The value of h_0 contains given or computable constants from the assumptions on the coefficients, the mesh regularity and geometry.

ETNA
Kent State University
http://etna.math.kent.edu

(iii) The lower bound in (5.16) is asymptotically

$$\Delta t \geq O(h^2), \tag{5.27}$$

as $h \to 0$, and the constants are similarly computable.

(iv) If $\theta = 1$, i.e., the scheme is implicit, then there is no upper restriction on Δt. If $\theta < 1$, then Remark 5.5 shows that for many popular elements one has $R(h) = O(h^{-2})$ in (5.18). This has been proved so far for simplicial elements in any dimension, bilinear elements in 2D and prismatic elements in 3D. Hence, $\Delta t \leq O(h^2)$ as $h \to 0$, which yields with (5.27) the usual condition

$$\Delta t = O(h^2) \tag{5.28}$$

as $h \to 0$, for the space and time discretizations. In addition, the lower bound in (5.16) must be smaller than the upper bound in (5.17): in view of the factor $1 - \theta$ in the latter, this gives a restriction on θ to be close enough to 1.

REMARK 5.7. Let us consider problem (2.1)–(2.4) with principal parts only, i.e., when $q \equiv s \equiv 0$:

$$\frac{\partial u}{\partial t} - \operatorname{div}\Big(k(x,t,u,\nabla u)\nabla u\Big) = f(x,t) \quad \text{in } Q_T := \Omega \times (0,T),$$

$$u(x,t) = g(x,t) \quad \text{for} \quad (x,t) \in \Gamma_D \times [0,T],$$

$$k(x,t,u,\nabla u)\tfrac{\partial u}{\partial \nu} = \gamma(x,t) \quad \text{for} \quad (x,t) \in \Gamma_N \times [0,T],$$

and

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

Then Assumptions (ii)-(iv) of Theorem 5.2 become much simplified, since $\alpha_1 = \alpha_2 = 0$. Namely, assumption (ii) is dropped since formally $h_0 = \infty$, i.e., there is no upper bound on h. Assumptions (iii)-(iv) read as follows:

$$\Delta t \geq \frac{c_2}{\theta \mu_0 K_0} h^2; \qquad \text{if } \theta < 1, \text{ then } \quad \Delta t \leq \frac{1}{\mu_1 (1-\theta)} \min_{i=1,\dots,m} \frac{\int_\Omega \phi_i^2}{\int\limits_\Omega |\nabla \phi_i|^2}. \tag{5.29}$$

Let us now return to the statement (5.25). By reversing signs in Corollary 5.3, we obtain the corresponding discrete minimum principle:

COROLLARY 5.8. *Let the conditions of Theorem 5.2 hold, and let* $\hat{\gamma}(x,t) := \gamma(x,t) - s(x,t,0) \geq 0$. *Then*

$$u_i^{n+1} \geq \min\{0, g_{min}^{(n+1)}, u_{min}^{(n)}\} + \Delta t \min\{0, \hat{f}_{min}^{(n,n+1)}\}. \tag{5.30}$$

An important special case is the discrete *nonnegativity preservation principle*, the discrete analogue of (4.4):

THEOREM 5.9. *Let the conditions of Theorem 5.2 hold, and let* $\hat{f} \geq 0$, $g \geq 0$, $\hat{\gamma} \geq 0$ *and* $u_0 \geq 0$. *Then the discrete solution satisfies*

$$u_i^n \geq 0, \qquad \forall n = 0, 1, \dots, n_T, \ i = 1, \dots, m.$$

ETNA
Kent State University
http://etna.math.kent.edu

Proof. Assumptions $\hat{f} \geq 0$, $g \geq 0$ and $\hat{\gamma} \geq 0$ imply $g_{min}^{(n+1)} \geq 0$ and $\hat{f}_{min}^{(n,n+1)}$ for all n and i. Hence (5.30) becomes

$$u_i^{n+1} \geq \min\{0, u_{min}^{(n)}\}.$$

Here assumption $u_0 \geq 0$ implies $u_{min}^{(0)} \geq 0$. Hence we obtain by induction that $u_{min}^{(n)} \geq 0$ for all n. ∎

By Theorem 5.9, u^h is nonnegative at each node point. Properties (3.1)–(3.2) of the basis functions imply that the FEM solution $u^h(., n\Delta t)$ is also nonnegative for all time levels $n\Delta t$. If, in addition, we extend the solutions to Q_T with values between those on the neighbouring time levels, e.g., with the method of lines, then we obtain that the discrete solution satisfies

$$u^h \geq 0 \quad \text{on} \quad Q_T.$$

5.3. The DMP: problems with superlinear growth. In this subsection we allow stronger growth of the nonlinearities q and s than above, i.e., we return to Assumption 2.1 (A4). For this we need some extra technical assumptions and results. Let us first summarize the additional conditions.

ASSUMPTION 5.10.

(B1) We restrict ourselves to the case of an implicit scheme:

$$\theta = 1.$$

(B2) V_h is made up by linear, bilinear or prismatic elements.

(B3) The coefficient on Γ_N satisfies $\hat{\gamma}(x,t) := \gamma(x,t) - s(x,t,0) \equiv 0$. Further, $\Gamma_D \neq \emptyset$.

(B4) The exact solution satisfies $u(.,t) \in W^{1,q}(\Omega)$ for some $q > 2$ (if $d = 2$) or some $q \geq 2d/(d-(d-2)(p_1-2))$ (if $d \geq 3$) for all $t \in [0,T]$.

(B5) The discretization satisfies $M_{p_1} := \sup_{t\in[0,T]} \|u(.,t) - u_h(.,t)\|_{L^{p_1}(\Omega)} < \infty$.

Now, by [1], under Assumption 2.1 (A4), we recall the Sobolev embedding estimates

$$\|v\|_{L^{p_1}(\Omega)} \leq C_{\Omega,p_1}\|v\|_{H_D^1}, \quad \|v\|_{L^{p_2}(\Gamma_N)} \leq C_{\Gamma_N,p_2}\|v\|_{H_D^1}, \quad \forall v \in H_D^1(\Omega), \tag{5.31}$$

with some constants $C_{\Omega,p_1}, C_{\Gamma_N,p_2} > 0$ independent of v.

LEMMA 5.11. *Let V_h be made up by linear, bilinear or prismatic elements. Then there exists a constant $c_{p_2} > 0$, such that*

$$\|v\|_{L^{p_2}(\Gamma_N)} \leq c_{p_2} h^{-1}\|v\|_{L^2(\Omega)}, \qquad \forall v \in V_h. \tag{5.32}$$

Proof. We have

$$\|v\|_{H_D^1}^2 := \int_\Omega |\nabla v|^2 \leq \int_\Omega v^2 \max_{v\in V_h} \frac{\int_\Omega |\nabla v|^2}{\int_\Omega v^2} \leq c_{p_2} \cdot R(h) \int_\Omega v^2,$$

where $R(h)$ comes from (5.18) and, as seen before, satisfies $R(h) = O(h^{-2})$. This, combined with (5.31), yields the required estimate. ∎

Now we consider the full discretization (5.7) for $\theta = 1$:

$$\mathbf{M}\mathbf{u}^{n+1} + \Delta t \mathbf{K}(\mathbf{u}^{n+1})\mathbf{u}^{n+1} = \mathbf{M}\mathbf{u}^n + \Delta t\, \hat{\mathbf{f}}^{(n)}. \tag{5.33}$$

ETNA
Kent State University
http://etna.math.kent.edu

Let $u^{n+1} \in V_h$ denote the function with coefficient vector $\mathbf{u}^{n+1}$, and let $f^n(x) := f(x, n\Delta t)$. Then, by the definition of the mass and stiffness matrices, (5.33) implies

$$\int_\Omega u^{n+1} v + \Delta t\, B(u^{n+1}; u^{n+1}, v) = \int_\Omega u^n v + \Delta t \Big(\int_\Omega \hat{f}^n v + \int_{\Gamma_N} \hat{\gamma}^n v \Big), \quad \forall v \in V_h. \tag{5.34}$$

Here, from Assumption 5.10 (B3), the integral on Γ_N vanishes. Furthermore, recall that $\hat{f} \in L^\infty(Q_T)$ by Assumption 2.1 (A2).

LEMMA 5.12. *If Assumption 5.10 holds, then for all $t \in [0, T]$,*

$$\|u(.,t)\|_{L^{p_1}(\Omega)} \le \|u^0\|_{L^{p_1}(\Omega)} + T(\text{meas}(\Omega))^{\frac{1}{p_1}} \|\hat{f}\|_{L^\infty(Q_T)}.$$

Proof. Let $v = |u|^{p_1-2}u$, which satisfies $\nabla v = (p_1 - 1)|u|^{p_1-2}\nabla u$. By assumption (B4), $|\nabla u| \in L^q(\Omega)$, and it is easy to see from the condition on q that $|u|^{p_1-2} \in L^{q'}(\Omega)$, where $(1/q) + (1/q') = 1/2$. This implies by Hölder's inequality that $|\nabla v| \in L^2(\Omega)$. That is, for all fixed t we have $v(.,t) \in H^1_D(\Omega)$. Hence, we can set it in (5.2):

$$\int_\Omega \frac{\partial u}{\partial t} |u|^{p_1-2} u \, dx + B(u; u, |u|^{p_1-2}u) = \int_\Omega \hat{f} |u|^{p_1-2} u \, dx, \qquad \forall v \in H^1_D(\Omega), \quad t \in (0, T), \tag{5.35}$$

where we have used that $\hat{\gamma} \equiv 0$. Let

$$N(t) := \|u(.,t)\|^{p_1}_{L^{p_1}(\Omega)} = \int_\Omega |u(x,t)|^{p_1} \, dx.$$

Then $N'(t) = \int_\Omega p_1 \, |u|^{p_1-2} u \, \frac{\partial u}{\partial t} \, dx$. Further, using (5.3) and that $\nabla v = (p_1 - 1)|u|^{p_1-2}\nabla u$, we obtain

$$B(u; u, |u|^{p_1-2}u) = \int_\Omega \Big(k(x,t,u,\nabla u)(p_1 - 1)|u|^{p_1-2}|\nabla u|^2$$

$$+ r(x,t,u)|u|^{p_1} \Big) \, dx + \int_{\Gamma_N} z(x,t,u)|u|^{p_1} \, d\sigma \ge 0.$$

Hence, the left-hand side of (5.35) is estimated below by $N'(t)/p_1$. Using Hölder's inequality for the right-hand side of (5.35), we obtain

$$\frac{1}{p_1} N'(t) \le \|\hat{f}(.,t)\|_{L^{p_1}(\Omega)} \|u(.,t)\|^{p_1-1}_{L^{p_1}(\Omega)} \le (\text{meas}(\Omega))^{\frac{1}{p_1}} \|\hat{f}\|_{L^\infty(Q_T)} \, N(t)^{\frac{p_1-1}{p_1}}.$$

Excluding the trivial case $u \equiv 0$, we can divide by $N(t)^{\frac{p_1-1}{p_1}}$ and integrate from 0 to t to obtain

$$N(t)^{\frac{1}{p_1}} - N(0)^{\frac{1}{p_1}} \le T(\text{meas}(\Omega))^{\frac{1}{p_1}} \|\hat{f}\|_{L^\infty(Q_T)},$$

which is the desired estimate. ∎

LEMMA 5.13. *(1) If Assumptions 5.10 (B1) and (B3) hold, then the norms $\|u^n\|_{L^2(\Omega)}$ are bounded independently of n and V_h:*

$$\|u^n\|_{L^2(\Omega)} \le \|u^0\|_{L^2(\Omega)} + T(\text{meas}(\Omega))^{\frac{1}{2}} \|\hat{f}\|_{L^\infty(Q_T)} =: K_{L_2}.$$

ETNA
Kent State University
http://etna.math.kent.edu

(2) If all assumptions in Assumption 5.10 hold, then the norms $\|u^n\|_{L^{p_1}(\Omega)}$ *are bounded independently of* n *and* V_h*:*

$$\|u^n\|_{L^{p_1}(\Omega)} \le M_{p_1} + \|u^0\|_{L^{p_1}(\Omega)} + T(\text{meas}(\Omega))^{\frac{1}{p_1}}\|\hat{f}\|_{L^\infty(Q_T)} =: K_{p_1,\Omega}\,.$$

Proof. (1) Setting $v = u^{n+1}$ in (5.34), we obtain

$$\int_\Omega (u^{n+1})^2 + \Delta t\, B(u^{n+1}; u^{n+1}, u^{n+1}) = \int_\Omega u^n u^{n+1} + \Delta t \int_\Omega \hat{f}^n u^{n+1}. \tag{5.36}$$

To estimate below, the bilinear form can be dropped from the left-hand side since it is coercive. Using Cauchy-Schwarz inequalities, we have

$$\|u^{n+1}\|^2_{L^2(\Omega)} \le \|u^n\|_{L^2(\Omega)}\|u^{n+1}\|_{L^2(\Omega)} + \Delta t\|\hat{f}^n\|_{L^2(\Omega)}\|u^{n+1}\|_{L^2(\Omega)}.$$

Dividing by $\|u^{n+1}\|_{L^2(\Omega)}$ and repeating the argument n times, we obtain

$$\|u^{n+1}\|_{L^2(\Omega)} \le \|u^0\|_{L^2(\Omega)} + (n+1)\Delta t\|\hat{f}^n\|_{L^2(\Omega)},$$

where the r.h.s. is bounded since $(n+1)\Delta t \le T$ and $\|\hat{f}^n\|_{L^2(\Omega)} \le (\text{meas}(\Omega))^{\frac{1}{2}}\|\hat{f}\|_{L^\infty(Q_T)}$.

(2) It follows directly from Lemma 5.12 and assumption (B5). ∎

Lemmas 5.11 and 5.13 imply

COROLLARY 5.14. *We have*

$$\|u^n\|_{L^{p_2}(\Gamma_N)} \le K_{p_2,\Gamma_N} h^{-1},$$

where the constant $K_{p_2,\Gamma_N} > 0$ *is bounded independently of* n *and* V_h*.*

THEOREM 5.15. *Let problem (2.1)–(2.4) satisfy Assumption 2.1 and Assumption 5.10. Let us consider a family of finite element subspaces* $\mathcal{V} = \{V_h\}_{h\to 0}$*, where the family of associated FE meshes is regular as in Definition 5.1. Let the following assumptions hold:*

(i) for any $i = 1, \dots, m,\ j = 1, \dots, \bar{m}$ $(i \ne j)$*, if* $\text{meas}(\text{supp}\,\phi_i \cap \text{supp}\,\phi_j) > 0$*, then*

$$\nabla\phi_i \cdot \nabla\phi_j \le 0 \ \text{ on } \Omega \quad \text{and} \quad \int_\Omega \nabla\phi_i \cdot \nabla\phi_j \le -K_0\, h^{d-2} \tag{5.37}$$

with some constant $K_0 > 0$ *independent of* i, j *and* h*;*

(ii) the mesh parameter h *satisfies* $h < h_0$*, where* $h_0 > 0$ *is the first positive root of the equation*

$$-\frac{\mu_0 K_0}{c_2}\frac{1}{h^2} + \alpha_1 + \frac{\alpha_2}{h} + \frac{\beta_1 K_{p_1,\Omega}^{p_1-2}}{h^{\gamma_1}} + \frac{\beta_2 K_{p_2,\Gamma_N}^{p_2-2}}{h^{\gamma_2}} = 0, \tag{5.38}$$

where the numbers $0 < \gamma_1, \gamma_2 < 2$ *are defined below in (5.40), (5.41), respectively;*

(iii) we have

(5.39)

$$\Delta t \ge \frac{c_2 h^2}{\theta\big(\mu_0 K_0 - c_2\alpha_1 h^2 - c_2\alpha_2 h - c_2\beta_1 K_{p_1,\Omega}^{p_1-2} h^{2-\gamma_1} - c_2\beta_2 K_{p_2,\Gamma_N}^{p_2-2} h^{2-\gamma_2}\big)}.$$

Then the matrices $\mathbf{A}(\mathbf{u}^{n+1})$ *and* $\mathbf{B}(\mathbf{u}^n)$*, defined in (5.8)–(5.9), have the following properties:*

(1) $A(\mathbf{u}^{n+1})_{ij} \le 0, \quad i \ne j,\ i = 1, \dots, m,\ j = 1, \dots, \bar{m}$*;*

(2) $B(\mathbf{u}^n)_{ii} \ge 0, \quad i = 1, \dots, m.$

ETNA
Kent State University
http://etna.math.kent.edu

Proof. We follow the proof of Theorem 5.2. As a first difference, instead of u_h in the arguments, we must consider the functions u^{n+1} (for $\mathbf{A}$) and u^n (for $\mathbf{B}$) that have the coefficient vectors $\mathbf{u}^{n+1}$ and $\mathbf{u}^n$, respectively.

(1) Since we now have (2.6) instead of (5.10), the first estimate in (5.22) is replaced by

$$\int_\Omega r(x,t,u^{n+1})\phi_j\phi_i \le \int_\Omega \big(\alpha_1+\beta_1|u^{n+1}|^{p_1-2}\big)\phi_j\phi_i \le \alpha_1 \text{meas}(\Omega_{ij})+\beta_1\int_{\Omega_{ij}}|u^{n+1}|^{p_1-2}\,.$$

Here the first term is bounded by $\alpha_1 c_2 h^d$ as before. To estimate the second term, we use Hölder's inequality:

$$\int_{\Omega_{ij}}|u^{n+1}|^{p_1-2} \le \|u^{n+1}\|^{p_1-2}_{L^{p_1}(\Omega_{ij})}\|1\|^2_{L^{p_1}(\Omega_{ij})}.$$

For the first factor, we use Lemma 5.13 (2) to find that

$$\|u^{n+1}\|^{p_1-2}_{L^{p_1}(\Omega_{ij})} \le \|u^{n+1}\|^{p_1-2}_{L^{p_1}(\Omega)} \le K^{p_1-2}_{p_1,\Omega}\,.$$

The second factor satisfies, by (5.21),

$$\|1\|^2_{L^{p_1}(\Omega_{ij})} = \big(\text{meas}(\Omega_{ij})\big)^{2/p_1} \le c_2 h^{\frac{2d}{p_1}} \equiv c_2 h^{d-\gamma_1},$$

with

$$\gamma_1 := d - \frac{2d}{p_1} < 2, \tag{5.40}$$

since from Assumption 2.1 (A4), we have $\frac{2d}{p_1} > d-2$. Hence,

$$\int_{\Omega_{ij}}|u^{n+1}|^{p_1-2} \le K^{p_1-2}_{p_1,\Omega} c_2 h^{d-\gamma_1},$$

and, altogether,

$$\int_\Omega r(x,t,u^{n+1})\phi_j\phi_i \le \alpha_1 c_2 h^d + \beta_1 K^{p_1-2}_{p_1,\Omega} c_2 h^{d-\gamma_1}\,.$$

Similarly,

$$\int_{\Gamma_N} z(x,t,u^{n+1})\phi_j\phi_i \le \alpha_2 c_2 h^{d-1} + \beta_2\int_{\Gamma_{ij}}|u^{n+1}|^{p_2-2},$$

and here, for $d=2,3$ we use Corollary 5.14 and (5.13) to obtain

$$\int_{\Gamma_{ij}}|u^{n+1}|^{p_2-2} \le \|u^{n+1}\|^{p_2-2}_{L^{p_2}(\Gamma_{ij})}\|1\|^2_{L^{p_2}(\Gamma_{ij})} \le \|u^{n+1}\|^{p_2-2}_{L^{p_2}(\Gamma_N)}\big(\text{meas}(\Gamma_{ij})\big)^{2/p_2}$$

$$\le K^{p_2-2}_{p_2,\Gamma_N} c_2 h^{2-p_2+\frac{2(d-1)}{p_2}} \equiv K^{p_2-2}_{p_2,\Gamma_N} c_2 h^{d-\gamma_2},$$

where

$$\gamma_2 := d-2+p_2-\frac{2(d-1)}{p_2} < 2, \tag{5.41}$$

from assumption $p_2 \leq 2.5$. Summing up, using the above and (5.23), we obtain

$$A(\mathbf{u}^h)_{ij} \leq c_2 h^d \left[1 + \theta \Delta t \left(-\frac{\mu_0 K_0}{c_2} \frac{1}{h^2} + \alpha_1 + \frac{\alpha_2}{h} + \frac{\beta_1 K_{p_1,\Omega}^{p_1-2}}{h^{\gamma_1}} + \frac{\beta_2 K_{p_2,\Gamma_N}^{p_2-2}}{h^{\gamma_2}} \right) \right] .$$

Since $h < h_0$ for h_0 defined in (5.38), it follows that we have a negative coefficient of $\theta \Delta t$ above, and from (5.39) we obtain that the expression in the large brackets is nonpositive. Hence $A(\mathbf{u}^h)_{ij} \leq 0$.

(2) For the implicit scheme, $\mathbf{B}(\mathbf{u}^n)$ coincides with the mass matrix $\mathbf{M}$, whose diagonal entries are positive. $\square$

Similarly to the sublinear case, we can derive the corresponding discrete maximum, minimum and nonnegativity preservation principles. We only formulate here the latter:

COROLLARY 5.16. *Let the conditions of Theorem 5.15 hold, and let $\hat{f} \geq 0$, $g \geq 0$, $\hat{\gamma} \geq 0$ and $u_0 \geq 0$. Then the discrete solution satisfies $u_i^n \geq 0$, for $n = 0, 1, ..., n_T$, $i = 1, ..., m$.*

5.4. Geometric properties of the space mesh. In order to satisfy condition (5.37), the most direct way is to require

$$\nabla \phi_i \cdot \nabla \phi_j \leq -K_0 \, h^{-2} \tag{5.42}$$

pointwise on the common support of these basis functions. In view of well-known formulae (see, e.g., [2, 7, 25, 30]), the above condition has a nice geometric interpretation: in the case of simplicial meshes, it is sufficient if the employed mesh is uniformly acute [4, 25]. In the case of bilinear elements, condition (5.42) is equivalent to the so-called strict non-narrowness of the meshes; see [12, 19]. The case of prismatic finite elements is treated in [18].

These conditions are sufficient but not necessary. For instance, for linear elements, some obtuse interior angles may occur in the simplices of the meshes, just as for linear problems (see, e.g., [24]), or one can require (5.42) only on a proper subpart of each intersection of supports with asymptotically nonvanishing measure, see more details in [21]. These weaker conditions may allow in general easier mesh adaptive procedures that preserve the DMP.

5.5. Examples. We give two real-life examples where discrete nonnegativity can be derived for suitable discretizations.

(a) Nonlinear heat conduction.

Heat conduction in a body $\Omega \subset \mathbb{R}^3$ with nonlinear diffusion coefficient is often described by the model

$$\frac{\partial u}{\partial t} - \operatorname{div} \Big(k(x,t,u) \nabla u \Big) = f(x,t) \tag{5.43}$$

in $Q_T := \Omega \times (0,T)$, where $T > 0$ is the time interval considered; see, e.g., [15]. The usual boundary and initial conditions are

$$u(x,t) = g(x,t) \quad \text{for} \quad (x,t) \in \Gamma_D \times [0,T], \tag{5.44}$$

$$k(x,t,u) \tfrac{\partial u}{\partial \nu} = \gamma(x,t) \quad \text{for} \quad (x,t) \in \Gamma_N \times [0,T], \tag{5.45}$$

and

$$u(x,0) = u_0(x) \quad \text{for} \quad x \in \Omega, \tag{5.46}$$

ETNA
Kent State University
http://etna.math.kent.edu

where all coefficients are bounded nonnegative measurable functions and k has a positive lower bound. The function u describes the temperature. Hence $u \geq 0$.

(b) Reaction-diffusion problems.

A reaction-diffusion process in a body $\Omega \subset \mathbb{R}^d$, $d = 2$ or 3, is often described by the model

$$\frac{\partial u}{\partial t} - \operatorname{div}\Big(k(x,t)\nabla u\Big) + q(x,u) = f(x,t) \tag{5.47}$$

in $Q_T := \Omega \times (0,T)$. The boundary and initial conditions are

$$u(x,t) = g(x,t) \quad \text{for} \quad (x,t) \in \Gamma_D \times [0,T], \tag{5.48}$$

$$k(x,t)\tfrac{\partial u}{\partial \nu} + s(x,u) = \gamma(x,t) \quad \text{for} \quad (x,t) \in \Gamma_N \times [0,T], \tag{5.49}$$

and

$$u(x,0) = u_0(x) \quad \text{for} \quad x \in \Omega, \tag{5.50}$$

The function u describes the concentration. Hence $u \geq 0$. Here the coefficients k, f, g, γ and u_0 are bounded nonnegative measurable functions and k has a positive lower bound. Further, q and s desribe the rate of reaction in the body and on the transmission boundary, respectively. Hence $q(x,0) = s(x,0) = 0$ for all x. In various examples the reaction process is such that q and s grow along with u. Further, the rate is at most polynomial, i.e., we may assume that the growth conditions (2.6) are satisfied. For instance, $q(x,u) = u^\sigma$ for some $\sigma > 1$ in some autocatalytic chemical reactions, or $q(x,u) = \frac{1}{\varepsilon}\frac{u}{u+\kappa}$ describes the Michaelis-Menten reaction in enzyme kynetics [9, 23].

In both examples, we have $\hat{f} = f \geq 0$, $g \geq 0$, $\hat{\gamma} = \gamma \geq 0$ and $u_0 \geq 0$. Therefore we can use Theorem 5.9 and Corollary 5.16, respectively, to derive the discrete nonnegativity principle:

THEOREM 5.17. *Let the full discretization satisfy the conditions of Theorem 5.2 for problem (5.43)–(5.46), or the conditions of Theorem 5.15 for problem (5.47)–(5.50). Then the discrete solution satisfies $u_i^n \geq 0$, for $n = 0, 1, \ldots, n_T$, $i = 1, \ldots, m$.*

In particular, for problem (5.43)–(5.46) we can use the simplified assumptions (5.29) for Theorem 5.2, as given in Remark 5.7.

Consequently, as pointed out after Theorem 5.9, if we extend the solutions to Q_T with values between those on the neighbouring time levels, e.g., by the method of lines, then the discrete solution satisfies $u^h \geq 0$, on Q_T.

REFERENCES

[1] R. A. ADAMS, *Sobolev Spaces*, Academic Press, New York-London, 1975.
[2] J. BRANDTS, S. KOROTOV, AND M. KŘÍŽEK, *Dissection of the path-simplex in $\mathbb{R}^n$ into n path-subsimplices*, Linear Algebra Appl., 421 (2007), pp. 382–393.
[3] J. BRANDTS, S. KOROTOV, AND M. KŘÍŽEK, *On the equivalence of regularity criteria for triangular and tetrahedral finite element partitions*, Comput. Math. Appl., 55 (2008), pp. 2227–2233.
[4] J. BRANDTS, S. KOROTOV, M. KŘÍŽEK, AND J. ŠOLC, *On nonobtuse simplicial partitions*, SIAM Rev., 51 (2009), pp. 317–335.
[5] P. G. CIARLET, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1978.
[6] P. G. CIARLET, *Discrete maximum principle for finite difference operators*, Aequationes Math., 4 (1970), pp. 338–352.

ETNA
Kent State University
http://etna.math.kent.edu

[7] P. G. CIARLET AND P. A. RAVIART, *Maximum principle and uniform convergence for the finite element method*, Comput. Methods Appl. Mech. Engrg., 2 (1973), pp. 17–31.

[8] A. DRAGANESCU, T. F. DUPONT, AND L. R. SCOTT, *Failure of the discrete maximum principle for an elliptic finite element problem*, Math. Comp., 74 (2005), pp. 1–23.

[9] J. I. DÍAZ, *Applications of symmetric rearrangement to certain nonlinear elliptic equations with a free boundary*, in Nonlinear differential equations (Granada, 1984), Research Notes in Mathematics, 132, Pitman, Boston, MA, 1985, pp. 155–181.

[10] I. FARAGÓ AND R. HORVÁTH, *Discrete maximum principle and adequate discretizations of linear parabolic problems*, SIAM J. Sci. Comput., 28 (2006), pp. 2313–2336.

[11] I. FARAGÓ AND R. HORVÁTH, *Continuous and discrete parabolic operators and their qualitative properties*, IMA J. Numer. Anal., 29 (2009), pp. 606-631.

[12] I. FARAGÓ, R. HORVÁTH, AND S. KOROTOV, *Discrete maximum principle for linear parabolic problems solved on hybrid meshes*, Appl. Numer. Math., 53 (2005), pp. 249–264.

[13] I. FARAGÓ, R. HORVÁTH, AND S. KOROTOV, *Discrete maximum principles for FE solutions of nonstationary diffusion-reaction problems with mixed boundary conditions*, Numer. Methods Partial Differential Equations, to appear.

[14] I. FARAGÓ AND J. KARÁTSON, *Numerical Solution of Nonlinear Elliptic Problems via Preconditioning Operators. Theory and Applications.* Advances in Computation, Vol. 11, NOVA Science Publishers, New York, 2002.

[15] J. FRANCU, *Monotone operators. A survey directed to applications to differential equations*, Appl. Math., 35 (1990), pp. 257–301.

[16] A. FRIEDMANN, *Partial Differential Equations of Parabolic Type*, Prentice-Hall, Englewood Cliffs, NJ, 1964.

[17] H. FUJII, *Some remarks on finite element analysis of time-dependent field problems*, in Theory and Practice in Finite Element Structural Analysis, Y. Yamada and R. H. Gallagher, eds., Tokyo University Press, Tokyo, 1973, pp. 91–106.

[18] A. HANNUKAINEN, S. KOROTOV, AND T. VEJCHODSKÝ, *Discrete maximum principles for FE solutions of the diffusion-reaction problem on prismatic meshes*, J. Comput. Appl. Math., 226 (2009), pp. 275–287.

[19] R. HORVÁTH, *Sufficient conditions of the discrete maximum-minimum principle for parabolic problems on rectangular meshes*, Comput. Math. Appl., 55 (2008), pp. 2306–2317.

[20] J. KARÁTSON AND S. KOROTOV, *Discrete maximum principles for finite element solutions of nonlinear elliptic problems with mixed boundary conditions*, Numer. Math., 99 (2005), pp. 669–698.

[21] J. KARÁTSON AND S. KOROTOV, *A discrete maximum principle in Hilbert space with applications to nonlinear cooperative elliptic systems*, SIAM J. Numer. Anal., 47 (2009), pp. 2518-2549.

[22] H. B. KELLER, *The numerical solution of parabolic partial differential equations*, in Mathematical Methods for Digital Computers, A. Ralston and H.S. Wilf, eds., New York, 1960, pp. 135–143.

[23] H. B. KELLER, *Elliptic boundary value problems suggested by nonlinear diffusion processes*, Arch. Rational Mech. Anal., 35 (1969), pp. 363–381.

[24] S. KOROTOV, M. KŘÍŽEK, AND P. NEITTAANMÄKI, *Weakened acute type condition for tetrahedral triangulations and the discrete maximum principle*, Math. Comp., 70 (2001), pp. 107–119.

[25] M. KŘÍŽEK AND QUN LIN, *On diagonal dominance of stiffness matrices in 3D*, East-West J. Numer. Math., 3 (1995), pp. 59–69.

[26] O. A. LADYZENSKAJA, V. A. SOLONNIKOV, AND N. N. URAL'CEVA, *Linear and Quasilinear Equations of Parabolic Type*, Translations of Mathematical Monographs, Vol. 23,American Mathematical Society, Providence, RI, 1968.

[27] V. THOMÉE, *Galerkin Finite Element Methods for Parabolic Problems*, Springer, Berlin, 1997.

[28] T. VEJCHODSKÝ, S. KOROTOV, AND A. HANNUKAINEN, *Discrete maximum principle for parabolic problems solved by prismatic finite elements*, Math. Comput. Simulation, to appear.

[29] T. VEJCHODSKÝ AND P. SOLÍN, *Discrete maximum principle for higher-order finite elements in 1D*, Math. Comp., 76 (2007), pp. 1833–1846.

[30] J. XU AND L. ZIKATANOV, *A monotone finite element scheme for convection-diffusion equations*, Math. Comp., 68 (1999), pp. 1429–1446.