MATHEMATICS OF COMPUTATION
Volume 77, Number 261, January 2008, Pages 11–19
S 0025-5718(07)02021-2
Article electronically published on May 14, 2007

ON THE EXISTENCE OF MAXIMUM PRINCIPLES IN PARABOLIC FINITE ELEMENT EQUATIONS

VIDAR THOMÉE AND LARS B. WAHLBIN

ABSTRACT. In 1973, H. Fujii investigated discrete versions of the maximum principle for the model heat equation using piecewise linear finite elements in space. In particular, he showed that the lumped mass method allows a maximum principle when the simplices of the triangulation are acute, and this is known to generalize in two space dimensions to triangulations of Delauney type. In this note we consider more general parabolic equations and first show that a maximum principle cannot hold for the standard spatially semidiscrete problem. We then show that for the lumped mass method the above conditions on the triangulation are essentially sharp. This is in contrast to the elliptic case in which the requirements are weaker. We also study conditions for the solution operator acting on the discrete initial data, with homogeneous lateral boundary conditions, to be a contraction or a positive operator.

1. INTRODUCTION

Let Ω be a bounded polyhedral domain in $\mathbb{R}^d$ and consider the problem

$$u_t + Au = 0 \quad \text{in } \Omega, \quad \text{for } t > 0, \qquad \text{with } u = g \quad \text{on } \partial\Omega, \quad \text{for } t \geq 0, \qquad u(0) = v \quad \text{in } \Omega. \tag{1.1}$$

Here

$$Au = -\sum_{k,l=1}^{d} \frac{\partial}{\partial x_k}\big(a_{kl}\frac{\partial u}{\partial x_l}\big) + \sum_{k=1}^{d} b_k \frac{\partial u}{\partial x_k},$$

where the coefficients $a_{kl},\ b_k \in \mathbb{C}^1(\bar{\Omega})$, and $(a_{kl}(x))$ is a symmetric and uniformly positive definite matrix on $\bar{\Omega}$.

The maximum principle for (1.1) asserts that, if $Q_T = \Omega \times (0,T)$, with $T > 0$, the maximum and the minumum of a solution $u \in \mathbb{C}^2(Q_T) \cap \mathbb{C}(\bar{Q}_T)$ over $\bar{Q}_T$ occur on the parabolic boundary, $\big(\partial\Omega \times [0,T]\big) \cup \big(\Omega \times \{t=0\}\big)$. As a consequence of this we find at once a bound for $|u|$ in Q_T, namely

$$\|u\|_{Q_T} = \max\big(\|g\|_{\partial\Omega\times[0,T]}, \|v\|_{\Omega}\big), \quad \text{where } \|u\|_V = \max_{\bar{V}} |u|. \tag{1.2}$$

Here V denotes a set in $\mathbb{R}^d$ or $\mathbb{R}^{d+1}$; when $V = \Omega$ we normally omit this subscript.

Received by the editor October 16, 2006 and, in revised from, November 10, 2006.
2000 *Mathematics Subject Classification.* Primary 65M12, 65M60.
Key words and phrases. Maximum principle, parabolic equations, finite elements, lumped mass.
The authors were partly supported by the U.S. National Science Foundation under Grant DMS 0310539.

Since constant functions satisfy the differential equation in (1.1), it is well known and easy to prove that (1.2) is equivalent to the positivity condition

$$u \geq 0 \quad \text{in } Q_T, \quad \text{if } v \geq 0 \quad \text{in } \Omega \quad \text{and} \quad g \geq 0 \quad \text{on } \partial\Omega \times [0,T]. \tag{1.3}$$

In the case of homogeneous Dirichlet boundary conditions, i.e., when $g = 0$, it follows from (1.2) that the solution operator $E(t)$, the semigroup defined by $u(t) = E(t)v$ on the continuous functions which vanish on $\partial\Omega$, is a contraction, or

$$\|E(t)v\| \leq \|v\|, \quad \text{for } t \geq 0. \tag{1.4}$$

We shall be concerned with spatially semidiscrete approximations of (1.1) based on continuous, piecewise linear finite elements, defined on a family of triangulations $\mathcal{T}_h = \{\tau\}$ of $\bar{\Omega}$ into closed simplices τ, such that any face of any τ is either a subset of the boundary $\partial\Omega$ or a face of another $\tau \in \mathcal{T}_h$. We set $h = \max_{\tau \in \mathcal{T}_h} \operatorname{diam}(\tau)$. We associate with $\mathcal{T}_h$ the finite dimensional spaces

$$S_h = \{\chi \in \mathcal{C}(\bar{\Omega}) : \ \chi|_\tau \text{ linear for } \tau \in \mathcal{T}_h\} \quad \text{and} \quad S_h^0 = \{\chi \in S_h : \chi = 0 \quad \text{on } \partial\Omega\}.$$

For each t, let $g_h(t)$ be the restriction to $\partial\Omega$ of a function in S_h and let $v_h \in S_h$ with $v_h = g_h(0)$ on $\partial\Omega$. The semidiscrete standard Galerkin finite element problem associated with (1.1) is then to find $u_h(t) \in S_h$ for $t \geq 0$ such that

$$\begin{aligned} &(u_{h,t}, \chi) + A(u_h, \chi) = 0, \quad \forall \chi \in S_h^0, \ t > 0, \\ &\qquad \text{with } u_h(t) = g_h(t) \quad \text{on } \partial\Omega, \quad \text{for } t \geq 0, \quad \text{and } u_h(0) = v_h, \end{aligned} \tag{1.5}$$

where $(f,g) = \int_\Omega f(x)\, g(x)\, dx$ and, with $\underline{b} = (b_1, \dots, b_d)^T$,

$$A(f,g) = \int_\Omega \Big(\sum_{k,l=1}^d a_{kl} \frac{\partial f}{\partial x_l} \frac{\partial g}{\partial x_k} + \sum_{k=1}^d b_k \frac{\partial f}{\partial x_k} g \Big)\, dx = A_0(f,g) + (\underline{b} \cdot \nabla f, g).$$

It is natural to ask whether an analogue of the maximum principle (1.2) holds for the discrete problem (1.5), or whether

$$\|u_h\|_{Q_T} = \max\big(\|g_h\|_{\partial\Omega \times [0,T]}, \|v_h\|_\Omega\big). \tag{1.6}$$

We shall demonstrate below that, in general, this is not the case. In Fujii [2] it was shown, for special families of triangulations $\mathcal{T}_h$ with all angles acute, that the backward Euler method may satisfy a maximum principle. His result requires a certain lower bound for the time step and therefore does not imply the same for the semidiscrete method by letting the time step tend to 0.

As a preparation for our analysis we express the semidiscrete problem (1.5) in matrix form: Let $\{P_i\}_{i=1}^n$ denote the nodes of $\mathcal{T}_h$ in the interior of Ω, and $\{P_{n+i}\}_{i=1}^m$ those on $\partial\Omega$, and let $\{\Phi_i\}_{i=1}^{n+m} \subset S_h$ be the standard basis of pyramid functions defined by $\Phi_i(P_j) = \delta_{ij}$. The mass and stiffness matrices are then $\mathcal{M} = (m_{ij})$ and $\mathcal{S} = (s_{ij})$, where $m_{ij} = (\Phi_i, \Phi_j)$ and $s_{ij} = A(\Phi_i, \Phi_j), \ i,j = 1:n$. To include the boundary terms, we also set $\mathcal{B} = (b_{ij})$ and $\mathcal{Z} = (z_{ij})$ with $b_{ij} = A(\Phi_i, \Phi_{n+j}), \ z_{ij} = (\Phi_i, \Phi_{n+j}), \ i = 1:n, \ j = 1:m$. We now also introduce the vector $\alpha(t) = (\alpha_1(t), \dots, \alpha_n(t))^T$ of nodal values of $u_h(t)$ and correspondingly $\widetilde{g}(t) = (\widetilde{g}_{n+1}(t), \dots, \widetilde{g}_{n+m}(t))^T$, where $\widetilde{g}_j(t) = g_h(P_j, t)$ and $\widetilde{v} = (v_h(P_1), \dots, v_h(P_n))^T$. Thus $u_h(t) = \sum_{i=1}^n \alpha_i(t)\Phi_i + \sum_{j=1}^m \widetilde{g}_{n+j}(t)\, \Phi_{n+j}$, and we may hence write (1.5) as

$$\mathcal{M}\frac{d\alpha}{dt} + \mathcal{S}\alpha = -\mathcal{B}\widetilde{g} - \mathcal{Z}\frac{d\widetilde{g}}{dt}, \quad \text{for } t \geq 0, \quad \text{with } \alpha(0) = \widetilde{v}. \tag{1.7}$$

Since the last term in (1.7) cannot be bounded by $\|g_h\|_{\partial\Omega\times[0,T]}$, it is already now clear that the full discrete maximum principle (1.6) cannot hold.

We now introduce the discrete semigroup $E_h(t)$ on S_h^0 by setting $E_h(t)v_h = u_h(t)$, where $u_h(t)$ is the solution of (1.5) with boundary data $g_h(t) = 0$. It has been shown by specific counterexamples (see, e.g., [6], Chapter 6) that $E_h(t)$ does not generally satisfy the analogue of (1.4),

$$\|E_h(t)v_h\| \le \|v_h\|, \quad \text{for } t \ge 0. \tag{1.8}$$

Our first goal in this paper is to show that, in fact, (1.8) cannot hold for any triangulation $\mathcal{T}_h$ which is "fine" enough. We continue to show that $E_h(t)$ cannot be positive in the sense that $E_h(t)v_h \ge 0$ if $v_h \ge 0$. Thus in neither case (1.6) can be valid. These results will be shown in Section 2 below. For weaker maximum-norm stability estimates than (1.8), see [6] and references therein.

We now turn to the lumped mass method, which results from replacing the mass matrix $\mathcal{M}$ in (1.7) by a diagonal matrix $\mathcal{D}$ with diagonal elements $d_{ii} = \sum_{j=1}^n m_{ij}$ and also setting $\mathcal{Z} = 0$, or

$$\mathcal{D}\frac{d\alpha}{dt} + \mathcal{S}\alpha = -\mathcal{B}\widetilde{g}, \quad \text{for } t \ge 0, \quad \text{with } \alpha(0) = \widetilde{v}. \tag{1.9}$$

This may also be written in variational form, replacing the inner product in (1.5) by a quadrature approximation, or

$$\begin{aligned}(u_{h,t}, \chi)_h + A(u_h, \chi) = 0, \quad \forall \chi \in S_h^0,\ t > 0,&\\ \text{with } u_h(t) = g_h(t) \quad \text{on } \partial\Omega, \quad \text{for } t \ge 0, \quad \text{and } u_h(0) = v_h,&\end{aligned} \tag{1.10}$$

where

$$(\psi, \chi)_h = \sum_{\tau\in\mathcal{T}_h} Q_{\tau,h}(\psi\,\chi), \quad Q_{\tau,h}(f) = \frac{\text{meas}(\tau)}{d+1}\sum_{P_j\in\bar{\tau}} f(P_j) \approx \int_\tau f\,dx.$$

In this lumped mass case, to be discussed in Section 3 below, we shall show that the discrete maximum principle (1.6) holds if and only if the off-diagonal elements of the stiffness matrix $\mathcal{S}$ are nonpositive and if $\mathcal{B} \le 0$, elementwise. In one space dimension, this is always the case. In two space dimensions, with $A = -\Delta$, this is equivalent to the condition that each edge of $\mathcal{T}_h$, not entirely on $\partial\Omega$, is of Delauney type, in the sense that the sums of the opposing angles in the two triangles containing it are at most π. In fact, if P_i and P_j are neighbors, i.e., if P_iP_j is an edge of $\mathcal{T}_h$, and if α and β are the angles opposite P_iP_j, then $(\nabla\Phi_i, \nabla\Phi_j) = -\sin(\alpha+\beta)/(4\sin\alpha\,\sin\beta) < 0, = 0$, or > 0 when $\alpha+\beta < \pi, = \pi$, or $> \pi$, respectively; see, e.g., Strang and Fix [5, p. 78], Xu and Zikatanov [7], Drăgănescu, Dupont and Scott [1], or [6]. In [2] the condition used was that each of these angles is $\le \pi/2$, and, for higher space dimensions, similar conditions of "acute" type are used. In Xu and Zikatanov [7] sharp conditions of Delauney type were given for $\mathcal{S}^{-1}$ to be nonpositive in any number of space dimensions.

The stationary discrete elliptic problem corresponding to (1.5) and (1.10) is

$$A(w_h, \chi) = 0, \quad \forall \chi \in S_h^0, \quad \text{with } w_h = g_h \quad \text{on } \partial\Omega, \tag{1.11}$$

If $A(\chi, \chi)$ is positive definite on S_h^0, (1.11) has a unique solution and may be written in matrix form, with α and $\widetilde{g}$ the vectors of nodal values of w_h and g_h,

$$\mathcal{S}\alpha = -\mathcal{B}\widetilde{g}, \quad \text{or} \quad \alpha = -\mathcal{S}^{-1}\mathcal{B}\widetilde{g}. \tag{1.12}$$

The maximum principle for (1.11), or

$$\|w_h\|_\Omega \le \|g_h\|_{\partial\Omega}, \tag{1.13}$$

is now equivalent to the positivity condition $\alpha \ge 0$ for $\widetilde{g} \ge 0$, and hence, by (1.12), $\mathcal{S}^{-1}\mathcal{B} \le 0$ is necessary and sufficient for (1.13).

Consider now the discrete parabolic problem (1.10) with $v_h = 0$ at interior nodes and $g_h(t) = g_h =$ constant in time, and assume that $A(\chi, \chi)$ is positive definite on S_h^0 and the maximum principle holds for (1.10). The solution of (1.10) then converges to the solution of (1.11) as $t \to \infty$, and (1.13) thus also holds for the discrete stationary problem. Hence $\mathcal{S}^{-1}\mathcal{B} \le 0$ is a necessary condition for (1.6) in this case.

In two dimensions and with $A = -\Delta$, by the results already mentioned, under Delauney conditions on $\mathcal{T}_h$, we have $\mathcal{S}^{-1} \ge 0$ and $\mathcal{B} \le 0$ and hence (1.13) holds. In Ruas Santos [4] examples were given of triangulations of non-Delauney type, for which the elliptic maximum principle still holds. In $\mathbb{R}^3$, Korotov, Křížek, and Neittaanmäki [3] showed an elliptic maximum principle for tetrahedral decompositions of nonacute type. (For certain convection dominated elliptic cases, cf. [7] and references therein.) Thus the parabolic maximum principle demands more stringent conditions than (1.13). In Section 3 we also discuss conditions for contractivity and positivity of the solution operator $\bar{E}_h(t)$ on S_h^0.

2. The standard Galerkin method

In this section we shall show that, in general, a maximum principle cannot hold for the semidiscrete standard Galerkin method (1.5). We shall show that, in fact, the discrete solution operator $E_h(t)$ on S_h^0, thus with $g_h(t) = 0$, is, in general, neither contractive nor positive. For our analysis, we define a node P_i of $\mathcal{T}_h$ to be strictly interior if all its neighbors are interior, and a near-boundary node if it is interior but not strictly interior. We set $\sigma_i = \operatorname{supp}(\Phi_i)$.

Theorem 2.1. *Assume that* $\operatorname{div} \underline{b} = 0$ *and that* $\mathcal{T}_h$ *is such that each near-boundary node has a strictly interior neighbor. Then* $E_h(t)$ *cannot be a contraction.*

The condition $\operatorname{div} \underline{b} = 0$ *is superfluous if for each near-boundary node* P_j *and an associated strictly interior neighbor* P_i*, we have* $\operatorname{meas}(\sigma_j \cap \sigma_i) \ge c \operatorname{meas}(\sigma_i)$ *with* $c > 0$ *and if* h *is sufficiently small.*

Proof. Setting $t = 0$ in (1.7), with $\widetilde{g} = 0$ and $\alpha(0) = \underline{1} = (1, 1, \dots, 1)^T$, we have

$$\mathcal{M}\beta = -\gamma = -\mathcal{S}\underline{1}, \quad \text{where } \beta = \alpha'(0). \tag{2.1}$$

If $E_h(t)$ were a contraction, then we would have $\beta \le 0$ elementwise, and hence $\gamma \ge 0$, and we shall show that this is not possible under the assumptions of the theorem. Let $w_h = \sum_{j=1}^n \Phi_j \in S_h$ correspond to the vector $\underline{1}$ above.

We assume first that $\operatorname{div} \underline{b} = 0$. In this case we have for the element γ_i of γ, corresponding to a strictly interior node P_i, using integration by parts,

$$\gamma_i = \sum_{j=1}^n s_{ij} = A(\Phi_i, w_h) = (\underline{b} \cdot \nabla\Phi_i, w_h) = -(\operatorname{div} \underline{b}\ \Phi_i, w_h) = 0, \tag{2.2}$$

since $w_h = 1$ on σ_i. Hence $\sum_{j=1}^n m_{ij}\beta_j = 0$ for β as in (2.1). Here $m_{ij} \ge 0$, with $m_{ij} > 0$ if and only if $j = i$ or if P_j is a neighbor of P_i, and if $\beta \le 0$, then $\beta_j = 0$

for the corresponding j. Thus by the assumption on $\mathcal{T}_h$ we have $\beta = 0$ and hence $\gamma = 0$. But γ cannot be zero since

$$\gamma \cdot \underline{1} = \mathcal{S}\,\underline{1} \cdot \underline{1} = A(w_h, w_h) = A_0(w_h, w_h) - \tfrac{1}{2}\big(\operatorname{div} \underline{b}\, w_h, w_h\big) > 0. \tag{2.3}$$

We now turn to the case of a general $\underline{b}$ and assume again that $\beta \le 0$. For P_i a strictly interior node we now have, by (2.1) and (2.2), that $m_{ii}|\beta_i| \le \gamma_i \le C\text{meas}(\sigma_i)$. Since $m_{ii} \ge c\,\text{meas}(\sigma_i)$, with $c > 0$, it follows that $|\beta_i| \le C$ for a positive constant C. If P_j is a near-boundary node, let P_i be a strictly interior neighboring node. Then $m_{ij}|\beta_j| \le \sum_{l=1}^n m_{il}|\beta_l| = \gamma_i \le C\text{meas}(\sigma_i)$, and, since, by assumption, $m_{ij} = (\Phi_i, \Phi_j) \ge c\,\text{meas}(\sigma_i \cap \sigma_j) \ge c\,\text{meas}(\sigma_i)$, we conclude also now that $|\beta_j| \le C$. Thus, for all interior nodes P_i, $\gamma_i \le C \sum_{j=1}^n m_{ij} \le C\text{meas}(\sigma_i)$, and hence $\gamma \cdot \underline{1} \le C$. But, by the first part of (2.3),

$$\gamma \cdot \underline{1} = \mathcal{S}\underline{1} \cdot \underline{1} \ge c\|\nabla w_h\|_{L_2}^2 - C.$$

Here $\nabla w_h = 0$ on all interior simplices $\tau \in \mathcal{T}_h$, so that only boundary simplices contribute to the first term on the right. For a boundary simplex τ, which has a full $(d-1)$-dimensional face F_τ on $\partial\Omega$, we have, with d_τ the distance from the interior vertex of τ to the hyperplane containing F_τ, and with $|F_\tau|$ the $(d-1)$-dimensional measure of F_τ,

$$\|\nabla w_h\|_{L_2(\tau)}^2 \ge d_\tau^{-2}\text{meas}(\tau) \ge c d_\tau^{-1}\,|F_\tau| \ge ch^{-1}\,|F_\tau|, \quad \text{with } c > 0.$$

Hence after summation over these τ, since $\bigcup_\tau F_\tau = \partial\Omega$, we conclude that $\gamma \cdot \underline{1} \ge ch^{-1} - C$. For small h this contradicts the boundedness of $\gamma \cdot \underline{1}$. □

We note that, in one space dimension, the first assumption about $\mathcal{T}_h$ holds if there are at least three interior nodes, and when $d = 2$, the second assumption is satisfied when the triangulation is fine enough, and the angles in the triangles of $\bigcup_i \sigma_i$ are bounded below where the union is taken over all i such that P_i is a strictly interior neighbor of a near-boundary node. To see that some condition on $\mathcal{T}_h$ is needed, we consider the case when there is only one interior node P_1 and $\operatorname{div} \underline{b} = 0$. The system (1.7) then reduces to the scalar equation

$$\|\Phi_1\|_{L_2}^2 \alpha_1' + A_0(\Phi_1, \Phi_1)\alpha_1 = 0, \quad \text{for } t \ge 0, \quad \alpha_1(0) = \widetilde{v} = (v_h, \Phi_1).$$

The solution is then the exponentially decreasing function $u_h(t) = \exp(-t\lambda)\widetilde{v}$, with $\lambda = A_0(\Phi_1, \Phi_1)/\|\Phi_1\|_{L_2}^2 > 0$, and the solution operator is a contraction.

Theorem 2.2. *Assume that* $\operatorname{div} \underline{b} \le 0$ *and that* $\mathcal{T}_h$ *is such that there exists a strictly interior node,* P_1 *say, such that any neighbor of* P_1 *has an interior neighbor which is not a neighbor of* P_1. *Then* $E_h(t)$ *cannot be a positive operator.*

The condition $\operatorname{div} \underline{b} \le 0$ *is not needed if* h *is sufficiently small.*

Proof. If $E_h(t)$ is a positive operator, then, by (1.7) with $\widetilde{g} = 0$, we have that $\mathcal{E}(t) = e^{-\mathcal{K}t} \ge 0$, elementwise, where $\mathcal{K} = (k_{ij}) = \mathcal{M}^{-1}\mathcal{S}$. Since $\mathcal{E}(t) = I - \mathcal{K}t + O(t^2)$ as $t \to 0$, we see that then all off-diagonal elements of $\mathcal{K}$ are nonpositive. We shall show that this is impossible.

Let P_i be any interior node $\ne P_1$ which is not a neighbor of P_1. Since $\mathcal{M}\mathcal{K} = \mathcal{S}$ and since $m_{i1} = s_{i1} = 0$, we have $\sum_{j \ne 1} m_{ij}k_{j1} = 0$. Hence $k_{j1} = 0$ when $m_{ij} > 0$, i.e., when $j = i$ and when j is such that P_j is a neighbor of P_i. By our assumption about $\mathcal{T}_h$ this shows that actually $k_{j1} = 0$ for all $j \ne 1$; this is also true when P_j and P_1 are neighbors. Thus the first column of $\mathcal{K}$ only contains one possible nonzero element, namely k_{11}.

For the stiffness matrix $\mathcal{S}$ we have $\sum_{j=1}^{n} s_{j1} = A(1,\Phi_1) = 0$ while, if $\operatorname{div} \underline{b} \le 0$,

$$s_{11} = A_0(\Phi_1,\Phi_1) - \tfrac{1}{2}\big(\operatorname{div} \underline{b}\, \Phi_1, \Phi_1\big) > 0.$$

In the case of a general $\underline{b}$, we have $s_{11} \ge (ch^{-2} - C)\operatorname{meas}(\sigma_1)$, with $c > 0$, and hence $s_{11} > 0$ for h small. Thus, in either case, the first column of $\mathcal{S}$ has elements of different signs whereas this is not the case for $\mathcal{M}$, in contradiction to $\mathcal{M}\mathcal{K} = \mathcal{S}$. $\square$

This time we remark that our assumption about $\mathcal{T}_h$ is satisfied in one dimension if there are five or more interior nodes. For the example following Theorem 2.1 above, with only one interior node, $E_h(t)$ is also a positive operator, which shows that some condition on $\mathcal{T}_h$ is needed in Theorem 2.2.

3. The lumped mass method

In this section we consider the lumped mass method and give necessary and sufficient conditions for the maximum principle to hold and also for the contractivity and positivity of the operator $\bar{E}_h(t)$ on S_h^0.

Theorem 3.1. *The maximum principle* (1.6) *holds for the semidiscrete parabolic lumped mass problem* (1.10) *if and only if the off-diagonal elements of $\mathcal{S}$ and all elements of $\mathcal{B}$ are nonpositive.*

Proof. With the notation in the introduction we have, from (1.9),

$$\alpha(t) = \bar{\mathcal{E}}(t)\alpha(0) - \int_0^t \bar{\mathcal{E}}(t-s)\mathcal{D}^{-1}\mathcal{B}\,\widetilde{g}(t)\,ds, \text{ with } \bar{\mathcal{E}}(t) = e^{-\mathcal{H}t},\ \mathcal{H} = \mathcal{D}^{-1}\mathcal{S}. \tag{3.1}$$

As in (1.3) for the continuous case, the maximum principle (1.6) is equivalent to

$$\alpha(t) \ge 0 \quad \text{for } t \ge 0 \quad \text{if } \alpha(0) \ge 0 \quad \text{and} \quad \widetilde{g}(t) \ge 0 \quad \text{for } t \ge 0.$$

Thus, if (1.6) holds, it follows from (3.1), with $\widetilde{g}(t) = 0$, that $\bar{\mathcal{E}}(t) \ge 0$. Since

$$\bar{\mathcal{E}}(t) = I - t\mathcal{H} + O(t^2), \quad \text{as } t \to 0, \tag{3.2}$$

all off-diagonal elements of $\mathcal{H}$ must then be nonpositive, and since $\mathcal{D}$ is diagonal with positive elements, the off-diagonal elements of $\mathcal{S} = \mathcal{D}\mathcal{H}$ are also nonpositive. Setting $\alpha(0) = 0$ and $\widetilde{g}(t) \ge 0$ in (3.1), we now find that it is also necessary that $\mathcal{B} \le 0$.

Conversely, if the off-diagonal elements of $\mathcal{S}$ are nonpositive, this holds also for $\mathcal{H}$. Writing $\mathcal{H} = \mathcal{P} - \mathcal{Q}$ where $\mathcal{P}$ is diagonal and $\mathcal{Q} \ge 0$, and setting $\mathcal{J} = I + k\mathcal{P}$, we have, for k small,

$$(I + k\mathcal{H})^{-1} = (\mathcal{J} - kQ)^{-1} = (I - k\mathcal{J}^{-1}\mathcal{Q})^{-1}\mathcal{J}^{-1} = \sum_{l=0}^{\infty} k^l(\mathcal{J}^{-1}Q)^l\mathcal{J}^{-1} \ge 0.$$

Since

$$\bar{\mathcal{E}}(t) = \lim_{n\to\infty}\Big(I + \frac{t}{n}\mathcal{H}\Big)^{-n}, \tag{3.3}$$

this shows $\bar{\mathcal{E}}(t) \ge 0$. If also $\mathcal{B} \le 0$, it follows that $\bar{\mathcal{E}}(t-s)\mathcal{D}^{-1}\mathcal{B} \le 0$. Hence, if $\alpha(0) \ge 0$ and $\widetilde{g}(t) \ge 0$ for $t \ge 0$, we obtain by (3.1) that $\alpha(t) \ge 0$ for $t \ge 0$, which shows our claim. $\square$

We note that the condition of the theorem may also be expressed as

$$A(\Phi_i,\Phi_j)\le 0 \quad \text{for } i=1:n,\ j=1:n+m,\ i\ne j. \tag{3.4}$$

As already remarked, in two dimensions and with $A=-\Delta$, this condition is equivalent to the triangulation being of Delauney type. When $A=-\Delta+\underline{b}\cdot\nabla$, (3.4) holds for h sufficiently small if the triangulation is strictly Delauney in the sense that $\alpha+\beta\le\gamma_0<\pi$, for all angles α, β associated with the edges of $\mathcal{T}_h$, not on $\partial\Omega$.

When $\operatorname{div}\underline{b}=0$, the condition (3.4) implies that $\mathcal{S}$ is diagonally dominant, or $\sum_{j\ne i}|s_{ij}|\le s_{ii}$ for $i=1:n$. In fact, since $\sum_{j=1}^{n+m}\Phi_j=1$ in Ω, we have $\sum_{j=1}^{n+m}A(\Phi_i,\Phi_j)=A(\Phi_i,1)=0$, so that $\sum_{j=1}^{n}s_{ij}=-\sum_{j=1}^{m}A(\Phi_i,\Phi_{n+j})\ge 0$, with equality when P_i is strictly interior (and thus not a neighbor of the P_{n+j}). Hence $\sum_{j\ne i}|s_{ij}|=-\sum_{j\ne i}s_{ij}\le s_{ii}$, with equality if P_i is strictly interior. We also note that at a strictly interior node P_i, if $\operatorname{div}\underline{b}=0$, the diagonal dominance in the i^{th} row is equivalent to $s_{ij}\le 0$ for $j\ne i$. This follows from $\sum_{j\ne i}|s_{ij}|\le s_{ii}=-\sum_{j\ne i}s_{ij}$. Note that the row sums of $\mathcal{S}$ corresponding to stictly interior nodes P_i are always zero and that for all these rows to be diagonally dominant, it is thus necessary and sufficient that $s_{ij}\le 0$ for $j\ne i$.

We shall now give necessary and sufficient conditions for the solution operator $\bar E_h(t)$ with homogeneous boundary conditions to be contractive. In view of the above discussion, in the case $\operatorname{div}\underline{b}=0$, these conditions are essentially concerned with the properties of the rows corresponding to near-boundary nodes.

Theorem 3.2. *The semigroup $\bar E_h(t)$ on S_h^0 is a contraction if and only if $\mathcal{S}$ is diagonally dominant.*

Proof. If $\bar E_h(t)$ is a contraction, so is the matrix $\bar{\mathcal{E}}(t)$ in (3.1) with respect to the vector maximum-norm $|\cdot|_\infty$. Since $\bar{\mathcal{E}}(t)=I-\mathcal{H}t+O(t^2)$ as $t\to 0$, we have $|\bar{\mathcal{E}}(t)|_\infty=\max(1-th_{ii}+t\sum_{j\ne i}|h_{ij}|)+O(t^2)$. For this norm to be bounded by 1, we find at once by taking t small that it is necessary that $\sum_{j\ne i}|h_{ij}|\le h_{ii}$ for $i=1:n$, so that $\mathcal{H}$ is diagonally dominant. Since $\mathcal{D}$ is a positive diagonal matrix, $\mathcal{S}=\mathcal{D}\mathcal{H}$ is then also diagonally dominant.

Conversely, if we know that $\mathcal{S}$, and hence also $\mathcal{H}=\mathcal{D}^{-1}\mathcal{S}$, is diagonally dominant, it is easy to see that

$$|(I+k\mathcal{H})^{-1}|_\infty\le 1, \quad \text{for } k>0. \tag{3.5}$$

In fact, set $w=(I+k\mathcal{H})^{-1}v$ and let $|w_j|=|w|_\infty$. Then

$$(1+kh_{jj})|w_j|=|v_j-k\sum_{l\ne j}h_{jl}w_l|\le|v|_\infty+kh_{jj}|w|_\infty,$$

from which $|w|_\infty\le|v|_\infty$, which shows (3.5). By (3.3) this implies $|\bar{\mathcal{E}}(t)|_\infty\le 1$, for $t\ge 0$, so that $\bar E_h(t)$ is a contraction. □

We now express the corresponding result for the positivity of $\bar E_h(t)$.

Theorem 3.3. *The semigroup $\bar E_h(t)$ is positive if and only if $s_{ij}\le 0$ for $j\ne i$.*

Proof. This time the positivity of $\bar{\mathcal{E}}(t)$, together with (3.2), shows at once that the off-diagonal elements $\mathcal{H}$ are nonpositive. Since $\mathcal{S}=\mathcal{D}\mathcal{H}$, this shows the only if part of the theorem.

Conversely, if the off-diagonal elements of $\mathcal{H}$ are nonpositive, one finds as in the last part of the proof of Theorem 3.1 that $\bar{\mathcal{E}}(t)\ge 0$. □

We next give two examples of two-dimensional triangulations for $A = -\Delta$ to show that neither of the conditions in Theorems 3.2 and 3.3 implies the other.

Example 3.1. The first example is a triangulation which gives a diagonally dominant stiffness matrix $\mathcal{S}$ but which has some $s_{ij} > 0,\ j \neq i$.

Let $\Omega = (0,1) \times (0,1)$ and start with a uniform partition into axes parallel squares, divided into triangles by their diagonals, from the lower left to the upper right hand corners. Let B_0, B_1, B_2 be three neighboring nodes on the interior of the horizontal lower boundary, N_1, N_2 the neighbors above B_1, B_2, and, in turn, Q_1, Q_2 their neighbors above. In the rectangle $N_1N_2Q_2Q_1$, change the original triangulation by erasing the edge N_1Q_2 and inserting an interior node P_0 connected to N_1, N_2, Q_1 and Q_2 such that the edge N_1N_2 becomes non-Delauney, while the rest of the edges are all Delauney. Then all rows of the modified $\mathcal{S}$ except those corresponding to N_1 and N_2 are diagonally dominant. For these we have (with a slight abuse of notation), noting that $A(\Phi_{N_1}, \Phi_{B_0}) = A(\Phi_{N_2}, \Phi_{B_1}) = 0$,

$$s_{N_lN_l} + s_{N_1N_2} + \sum_j s_{N_lP_j} + A(\Phi_{N_l}, \Phi_{B_l}) \equiv a_l + b - c_l - d = 0, \quad l = 1,\ 2,$$

where the summations are taken over the interior nodes P_j different from N_l, including the new node P_0. Here $a_l > 0$ and $c_l \geq 0$ since N_lP_j are Delauney. Furthermore, $b = s_{N_1N_2} = s_{N_2N_1} > 0$ since N_1N_2 is non-Delauney, and $d = -A(\Phi_{N_1}, \Phi_{B_1}) = -A(\Phi_{N_2}, \Phi_{B_2}) > 0$. The condition of diagonal dominance for $\mathcal{S}$ is that $b + c_l \leq a_l$ for $l = 1, 2$, and this holds if $d \geq 2b$, which is satisfied for a suitable choice of P_0.

Example 3.2. The second example is a triangulation which gives a stiffness matrix $\mathcal{S}$ with $s_{ij} \leq 0$ for all $j \neq i$ but which is not diagonally dominant.

We choose two points $B_1,\ B_2$ on a straight portion of $\partial\Omega$, and we let N be an interior point of Ω, so that NB_1B_2 is an equilateral triangle. By erecting suitable obtuse triangles Q_1NB_1 and Q_2NB_2 outside the triangle NB_1B_2, we can arrange that the edges NB_1 and NB_2 are non-Delauney. We complete the triangulation so that the rest of the edges are Delauney. Since all interior edges are Delauney, we have $s_{ij} \leq 0$ for $i \neq j$. We consider now the row corresponding to N in $\mathcal{S}$. We have

$$s_{NN} + \sum_j s_{NP_j} + \big(A(\Phi_N, \Phi_{B_1}) + A(\Phi_N, \Phi_{B_2})\big) \equiv a - c + d = 0,$$

where the summation is over interior nodes $P_j \neq N$. This row is not diagonally dominant since, by construction, $d > 0$, which implies $a < c$, or $s_{NN} < \sum_j |s_{NP_j}|$.

We finally remark that, for any given fixed triangulation $\mathcal{T}_h$, and given a_{ij}, it is possible to have a $\underline{b}$ such that the semigroup $\bar{E}_h(t)$ on S_h^0 cannot be contractive in maximum-norm. For, with $u_h(t) = \bar{E}_h(t)v_h$, we have

$$\tfrac{1}{2}\frac{d}{dt}\|u_h\|_h^2 = -A_0(u_h, u_h) + \tfrac{1}{2}(\operatorname{div}\underline{b}\ u_h, u_h), \quad \text{where } \|\cdot\|_h = (\cdot,\cdot)_h^{1/2}. \tag{3.6}$$

Since S_h^0 is finite dimensional, and hence all norms on it are equivalent, we can choose $\operatorname{div}\underline{b}$ large enough for the right hand side in (3.6) to be bounded below by $c\,\|u_h\|_h^2$, and hence such that $\|u_h(t)\|_h \to \infty$, and thus also $\|u_h(t)\| \to \infty$ as $t \to \infty$.

References

[1] A. Drăgănescu, T.F. Dupont, and L.R. Scott, *Failure of the discrete maximum principle for an elliptic finite element problem*, Math. Comp. 74 (2004), 1–23. MR2085400 (2005f:65148)

[2] H. Fujii, *Some remarks on finite element analysis of time-dependent field problems*, in Theory and Practice in Finite Element Structural Analysis, University of Tokyo Press, Tokyo, 1973, pp. 91–106.

[3] S. Korotov, M. Křížek, and P. Neittaanmäki, *Weakened acute type condition for tetrahedral triangulations and the discrete maximum principle,* Math. Comp. 70 (2001), 107–119. MR1803125 (2001i:65126)

[4] V. Ruas Santos, *On the strong maximum principle for some piecewise linear finite element approximate problems of nonpositive type*, J. Fac. Sci. Univ. Tokyo Sect. IA Math. 29 (1982), 473–491. MR672072 (84b:65118)

[5] G. Strang and G.J. Fix, *An Analysis of the Finite Element Method*, Prentice Hall, Englewood Cliffs, N.J., 1973. MR0443377 (56:1747)

[6] V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, Second edition, Springer-Verlag, Berlin, Heidelberg, 2006. MR2249024 (2007b:65003)

[7] J. Xu and L. Zikatanov, *A monotone finite element scheme for convection-diffusion equations*, Math. Comp. 68 (1999), 1429–1446. MR1654022 (99m:65225)

Department of Mathematics, Chalmers University of Technology, S-41296 Göteborg, Sweden

E-mail address: thomee@math.chalmers.se

Department of Mathematics, Cornell University, Ithaca, New York 14853

E-mail address: wahlbin@math.cornell.edu