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

\begin{equation}
\begin{aligned}
\frac{\partial u}{\partial t} + Au &= 0 \quad \text{in } \Omega, \quad \text{for } t > 0, \\
\text{with } u &= g \quad \text{on } \partial \Omega, \quad \text{for } t \geq 0, \\
\text{and } u(0) &= v \quad \text{in } \Omega.
\end{aligned}
\end{equation}

Here

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

where the coefficients $a_{kl}$, $b_k \in C^1(\bar{\Omega})$, 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 minimum of a solution $u \in C^2(Q_T) \cap C(\bar{Q}_T)$ over $Q_T$ occur on the parabolic boundary, $(\partial \Omega \times [0, T]) \cup (\Omega \times \{t = 0\})$. As a consequence of this we find at once a bound for $|u|$ in $Q_T$, namely

\begin{equation}
\|u\|_{Q_T} = \max (\|g\|_{\partial \Omega \times [0, T]}, \|v\|_{\Omega}), \quad \text{where } \|u\|_V = \max _V |u|.
\end{equation}

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.

©2007 American Mathematical Society
Reverts to public domain 28 years from publication

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use
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

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

In the case of homogeneous Dirichlet boundary conditions, i.e., when \( \partial v = 0 \) on \( \partial \Omega \), 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

\[ \|E(t)v\| \leq \|v\|, \quad \text{for} \quad t \geq 0. \]

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

\[ S_h = \{ \chi \in C(\Omega) : \chi|_\tau \text{ linear for } \tau \in T_h \} \quad \text{and} \quad S_h^0 = \{ \chi \in S_h : \chi = 0 \quad \text{on} \quad \partial \Omega \}. \]

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

\[
\begin{align*}
(\alpha_h(t), \chi) + A(u_h(t), \chi) &= 0, \quad \forall \chi \in S_h^0, \quad t > 0, \\
&\quad \text{with } u_h(t) = g_h(t) \quad \text{on } \partial \Omega, \quad \text{for } t \geq 0, \quad \text{and } u_h(0) = v_h,
\end{align*}
\]

where \( (f, g) = \int_\Omega f(x)g(x)\,dx \) and, with \( h = (b_1, \ldots, b_d)^T \),

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

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

\[ \|u_h\|_{Q_T} = \max \left( \|g_h\|_{\partial \Omega \times [0, T]}, \|v_h\|_{\Omega} \right). \]

We shall demonstrate below that, in general, this is not the case. In Fujii [2] it was shown, for special families of triangulations \( T_h \) 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 \( T_h \) in the interior of \( \Omega \), 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

\[ M = (m_{ij}) \quad \text{and} \quad S = (s_{ij}), \quad \text{where} \quad m_{ij} = (\Phi_i, \Phi_j) \quad \text{and} \quad s_{ij} = A(\Phi_i, \Phi_j), \quad i, j = 1 : n. \]

To include the boundary terms, we also set \( B = (b_{ij}) \) and \( Z = (z_{ij}) \) with \( b_{ij} = A(\Phi_i, \Phi_{n+j}), \quad z_{ij} = (\Phi_i, \Phi_{n+j}), \quad i = 1 : n, \quad j = 1 : m. \) We now also introduce the vector \( \alpha(t) = (\alpha_1(t), \ldots, \alpha_n(t))^T \) of nodal values of \( u_h(t) \) and correspondingly \( \bar{\alpha}(t) = (\bar{\alpha}_{n+1}(t), \ldots, \bar{\alpha}_{n+m}(t))^T \), where \( \bar{\alpha}(t) = g_h(P_j, t) \) and \( \bar{v} = (v_h(P_1), \ldots, v_h(P_n))^T. \)

Thus \( u_h(t) = \sum_{i=1}^n \alpha_i(t) \Phi_i + \sum_{j=1}^m v_{n+j}(t) \Phi_{n+j} \), and we may hence write (1.5) as

\[
M \frac{d\alpha}{dt} + S\alpha = -B\bar{\alpha} - 2\frac{d\bar{\alpha}}{dt}, \quad \text{for} \quad t \geq 0, \quad \text{with} \quad \alpha(0) = \bar{v}. \]
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 (1.4) 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).

\[
(1.8) \quad \| E_h(t)v_h \| \leq \| v_h \|, \quad \text{for } t \geq 0.
\]

Our first goal in this paper is to show that, in fact, (1.8) cannot hold for any triangulation \( T_h \) which is “fine” enough. We continue to show that \( E_h(t) \) cannot be positive in the sense that \( E_h(t)v_h \geq 0 \) if \( v_h \geq 0 \). Thus in neither case (1.8) 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 \( M \) in (1.7) by a diagonal matrix \( D \) with diagonal elements \( d_{ii} = \sum_{j=1}^n m_{ij} \) and also setting \( Z = 0 \), or

\[
(1.9) \quad D \frac{d\alpha}{dt} + S\alpha = -B\bar{g}, \quad \text{for } t \geq 0, \quad \text{with } \alpha(0) = \bar{v}.
\]

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

\[
(1.10) \quad (u_{h,t},\chi)_h + A(u_h,\chi) = 0, \quad \forall \chi \in S_h^0, \quad t > 0,
\]

\[
\text{with } u_h(t) = g_h(t) \quad \text{on } \partial\Omega, \quad \text{for } t \geq 0, \quad \text{and } u_h(0) = v_h,
\]

where

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

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 \( S \) are nonpositive and if \( B \leq 0 \), 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 \( T_h \), 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 \( \pi \). In fact, if \( P_i \) and \( P_j \) are neighbors, i.e., if \( P_iP_j \) is an edge of \( T_h \), and if \( \alpha \) and \( \beta \) are the angles opposite \( P_iP_j \), then \((\nabla \Phi_i,\nabla \Phi_j) = -\sin(\alpha + \beta)/(4 \sin \alpha \sin \beta) < 0\), 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 \( \leq \pi/2 \), 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 \( S^{-1} \) to be nonpositive in any number of space dimensions.

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

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

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 \( \alpha \) and \( \bar{g} \) the vectors of nodal values of \( w_h \) and \( g_h \),

\[
(1.12) \quad S\alpha = -B\bar{g}, \quad \text{or } \alpha = -S^{-1}B\bar{g}.
\]
The maximum principle for (1.11), or
\[ \|w_h\|_\Omega \leq \|g_h\|_{\partial\Omega}, \]
is now equivalent to the positivity condition \( \alpha \geq 0 \) for \( \tilde{g} \geq 0 \), and hence, by (1.12), \( S^{-1}B \leq 0 \) 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 \( S^{-1}B \leq 0 \) is a necessary condition for (1.16) in this case.

In two dimensions and with \( A = -\Delta \), by the results already mentioned, under Delauney conditions on \( T_h \), we have \( S^{-1} \geq 0 \) and \( B \leq 0 \) 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, Krížek, 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 \( \tilde{E}_h(t) \) 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 (1.3). 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 \( T_h \) 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 \( \sigma_i = \text{supp} (\Phi_i) \).

**Theorem 2.1.** Assume that \( \text{div } h = 0 \) and that \( T_h \) is such that each near-boundary node has a strictly interior neighbor. Then \( E_h(t) \) cannot be a contraction.

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

**Proof.** Setting \( t = 0 \) in (1.17), with \( \tilde{g} = 0 \) and \( \alpha(0) = 1 = (1, 1, \ldots, 1)^T \), we have
\[ M\beta = -\gamma = -S1, \]
where \( \beta = \alpha'(0) \).

If \( E_h(t) \) were a contraction, then we would have \( \beta \leq 0 \) elementwise, and hence \( \gamma \geq 0 \), 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 \( 1 \) above.

We assume first that \( \text{div } h = 0 \). In this case we have for the element \( \gamma_i \) of \( \gamma \), corresponding to a strictly interior node \( P_i \), using integration by parts,
\[ \gamma_i = \sum_{j=1}^{n} s_{ij} = A(\Phi_i, w_h) = (h \cdot \nabla \Phi_i, w_h) = -\langle \text{div } h, \Phi_i, w_h \rangle = 0, \]
since \( w_h = 1 \) on \( \sigma_i \). Hence \( \sum_{j=1}^{n} m_{ij} \beta_j = 0 \) for \( \beta \) as in (2.2). Here \( m_{ij} \geq 0 \), with \( m_{ij} > 0 \) if and only if \( j = i \) or if \( P_j \) is a neighbor of \( P_i \), and if \( \beta \leq 0 \), then \( \beta_j = 0 \)
for the corresponding $j$. Thus by the assumption on $T_h$ we have $\beta = 0$ and hence $\gamma = 0$. But $\gamma$ cannot be zero since
\begin{equation}
(2.3) \quad \gamma \cdot 1 = S1 \cdot 1 = A(w_h, w_h) = A_0(w_h, w_h) - \frac{1}{2}(\text{div } b w_h, w_h) > 0.\end{equation}

We now turn to the case of a general $h$ and assume again that $\beta \leq 0$. For $P_i$ a strictly interior node we now have, by (2.1) and (2.2), that $m_{ii} |\beta_i| \leq \gamma_i \leq C \text{meas}(\sigma_i)$. Since $m_{ii} \geq c \text{meas}(\sigma_i)$, with $c > 0$, it follows that $|\beta_i| \leq C$ for a positive constant $C$. If $P_i$ is a near-boundary node, let $P_i$ be a strictly interior neighboring node. Then $m_{ij} |\beta_j| \leq \sum_{l=1}^n m_{il} |\beta_l| = \gamma_i \leq C \text{meas}(\sigma_i)$, and, since, by assumption, $m_{ij} = (\Phi_i, \Phi_j) \geq c \text{meas}(\sigma_i \cap \sigma_j) \geq c \text{meas}(\sigma_j)$, we conclude also now that $|\beta_j| \leq C$. Thus, for all interior nodes $P_i$, $\gamma_i \leq C \sum_{j=1}^n m_{ij} \leq C \text{meas}(\sigma_i)$, and hence $\gamma \cdot 1 = C$. But, by the first part of (2.3),
\begin{equation}
\gamma \cdot 1 = S1 \cdot 1 \geq c ||\nabla w_h||_{L^2}^2 - C.
\end{equation}

Here $\nabla w_h = 0$ on all interior simplices $\tau \in T_h$, so that only boundary simplices contribute to the first term on the right. For a boundary simplex $\tau$, which has a full $(d-1)$-dimensional face $F_\tau$ on $\partial \Omega$, we have, with $d_\tau$ the distance from the interior vertex of $\tau$ to the hyperplane containing $F_\tau$, and with $|F_\tau|$ the $(d-1)$-dimensional measure of $F_\tau$,
\begin{equation}
||\nabla w_h||_{L^2(\tau)}^2 \geq d_\tau^2 \text{meas}(\tau) \geq c d_\tau^{-1} |F_\tau| \geq c h^{-1} |F_\tau|, \quad \text{with } c > 0.
\end{equation}

Hence after summation over these $\tau$, since $\bigcup_k F_\tau = \partial \Omega$, we conclude that $\gamma \cdot 1 \geq c h^{-1} - C$. For small $h$ this contradicts the boundedness of $\gamma \cdot 1$. 

We note that, in one space dimension, the first assumption about $T_h$ 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 $T_h$ is needed, consider the case when there is only one interior node $P_1$ and $\text{div } b = 0$.

The system (1.7) then reduces to the scalar equation
\begin{equation}
||\Phi_i||_{L^2}^2 \alpha_1' + A_0(\Phi_i, \Phi_i) \alpha_1 = 0, \quad \text{for } t \geq 0, \quad \alpha_1(0) = \bar{v} = (v_h, \Phi_1).
\end{equation}

The solution is then the exponentially decreasing function $u_h(t) = \exp(-\lambda t)\bar{v}$, 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 $\text{div } b \leq 0$ and that $T_h$ 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 $\text{div } b \leq 0$ is not needed if $h$ is sufficiently small.

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

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

For the stiffness matrix $S$ we have $\sum_{j=1}^n s_{j,1} = A(1, \Phi_1) = 0$ while, if $\text{div} b \leq 0$,
\[ s_{1,1} = A_0(\Phi_1, \Phi_1) - \frac{1}{2} (\text{div} b \Phi_1, \Phi_1) > 0. \]
In the case of a general $b$, we have $s_{1,1} \geq (c h^{-2} - C) \text{meas} (\sigma_1)$, with $c > 0$, and hence $s_{1,1} > 0$ for $h$ small. Thus, in either case, the first column of $S$ has elements of different signs whereas this is not the case for $M$, in contradiction to $MK = S$. □

This time we remark that our assumption about $T_h$ 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 $T_h$ 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 $E_h(t)$ on $S_0^h$.

**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 $S$ and all elements of $B$ are nonpositive.

**Proof.** With the notation in the introduction we have, from (1.9),
\[(3.1) \quad \alpha(t) = \tilde{E}(t) \alpha(0) - \int_0^t \tilde{E}(t - s) D^{-1} B \tilde{g}(t) \, ds, \quad \text{with} \quad \tilde{E}(t) = e^{-Ht}, \quad H = D^{-1} S.\]
As in (1.3) for the continuous case, the maximum principle (1.6) is equivalent to
\[\alpha(t) \geq 0 \quad \text{for} \quad t \geq 0 \quad \text{if} \quad \alpha(0) \geq 0 \quad \text{and} \quad \tilde{g}(t) \geq 0 \quad \text{for} \quad t \geq 0.\]
Thus, if (1.6) holds, it follows from (3.1), with $\tilde{g}(t) = 0$, that $\tilde{E}(t) \geq 0$. Since
\[(3.2) \quad \tilde{E}(t) = I - t H + O(t^2), \quad \text{as} \quad t \to 0,\]
all off-diagonal elements of $H$ must then be nonpositive, and since $D$ is diagonal with positive elements, the off-diagonal elements of $S = D H$ are also nonpositive. Setting $\alpha(0) = 0$ and $\tilde{g}(t) \geq 0$ in (3.1), we now find that it is also necessary that $B \leq 0$.
Conversely, if the off-diagonal elements of $S$ are nonpositive, this holds also for $H$. Writing $H = P - Q$ where $P$ is diagonal and $Q \geq 0$, and setting $J = I + k P$, we have, for $k$ small,
\[ (I + k H)^{-1} = (J - k Q)^{-1} = (I - k J^{-1} Q)^{-1} = \sum_{l=0}^{\infty} k^l (J^{-1} Q)^l J^{-1} \geq 0. \]
Since
\[(3.3) \quad \tilde{E}(t) = \lim_{n \to -\infty} \left( I + \frac{t}{n} H \right)^{-n}, \]
this shows $\tilde{E}(t) \geq 0$. If also $B \leq 0$, it follows that $\tilde{E}(t - s) D^{-1} B \leq 0$. Hence, if $\alpha(0) \geq 0$ and $\tilde{g}(t) \geq 0$ for $t \geq 0$, we obtain by (3.1) that $\alpha(t) \geq 0$ for $t \geq 0$, which shows our claim. □
We note that the condition of the theorem may also be expressed as
\[(3.4)\]
\[A(\Phi_i, \Phi_j) \leq 0 \quad \text{for } i = 1: n, \ j = 1: n + m, \ i \neq j.\]

As already remarked, in two dimensions and with \(A = -\Delta\), this condition is equivalent to the triangulation being of Delauney type. When \(A = -\Delta + \bar{b} \cdot \nabla\), \[(3.4)\]
holds for \(h\) sufficiently small if the triangulation is strictly Delauney in the sense that \(x + \beta \leq \gamma_0 < \pi\), for all angles \(x, \beta\) associated with the edges of \(T_h\), not on \(\partial\Omega\).

When \(\text{div} \bar{b} = 0\), the condition \[(3.3)\]
implies that \(S\) is diagonally dominant, or \(\sum_{j \neq i} |s_{ij}| \leq s_{ii}\) for \(i = 1: n\). In fact, since \(\sum_{j=1}^{n+m} A(\Phi_i, \Phi_j) = A(\Phi_i, 1) = 0\), so that \(\sum_{j=1}^{n} s_{ij} = -\sum_{j=1}^{m} A(\Phi_i, \Phi_{n+j}) \geq 0\), with equality when \(P_i\) is strictly interior (and thus not a neighbor of the \(P_{n+j}\)). Hence \(\sum_{j \neq i} |s_{ij}| = -\sum_{j \neq i} s_{ij} \leq s_{ii}\), with equality if \(P_i\) is strictly interior. We also note that at a strictly interior node \(P_i\), if \(\text{div} \bar{b} = 0\), the diagonal dominance in the \(i^{th}\) row is equivalent to \(s_{ij} \leq 0\) for \(j \neq i\). This follows from \(\sum_{j \neq i} |s_{ij}| \leq s_{ii} = -\sum_{j \neq i} s_{ij}\). Note that the row sums of \(S\) corresponding to strictly interior nodes \(P_i\) are always zero and that for all these rows to be diagonally dominant, it is thus necessary and sufficient that \(s_{ij} \leq 0\) for \(j \neq i\).

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 \(\text{div} \bar{b} = 0\), 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 \(S\) is diagonally dominant.

**Proof.** If \(\bar{E}_h(t)\) is a contraction, so is the matrix \(\bar{E}(t)\) in \[(3.1)\] with respect to the vector maximum-norm \(|\cdot|_\infty\). Since \(\bar{E}(t) = I - \mathcal{H}t + O(t^2)\) as \(t \to 0\), we have \(|\bar{E}(t)|_\infty = \max(1 - th_{ii} + t\sum_{j \neq i} |h_{ij}|) + O(t^2)\). For this norm to be bounded by 1, we find at once by taking \(t\) small that it is necessary that \(\sum_{j \neq i} |h_{ij}| \leq h_{ii}\) for \(i = 1: n\), so that \(\mathcal{H}\) is diagonally dominant. Since \(\mathcal{D}\) is a positive diagonal matrix, \(S = \mathcal{D}\mathcal{H}\) 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
\[(3.5)\]
\[|(I + k\mathcal{H})^{-1}|_\infty \leq 1, \quad \text{for } k > 0.\]

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 \neq j} h_{jl}w_l| \leq |v|_\infty + kh_{jj}|w|_\infty,
from which \(|w|_\infty \leq |v|_\infty\), which shows \[(3.5)\]. By \[(3.5)\] this implies \(|\bar{E}(t)|_\infty \leq 1\), for \(t \geq 0\), 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 \(s_{ij} \leq 0\) for \(j \neq i\).

**Proof.** This time the positivity of \(\bar{E}(t)\), 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 \(\bar{E}(t) \geq 0\).


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 \( S \) 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 \( N_1 N_2 Q_2 Q_1 \), change the original triangulation by erasing the edge \( N_1 Q_2 \) and inserting an interior node \( P_0 \) connected to \( N_1, N_2, Q_1 \) and \( Q_2 \) such that the edge \( N_1 Q_2 \) becomes non-Delauney, while the rest of the edges are all Delauney. Then all rows of the modified \( S \) except those corresponding to \( N_1 \) and \( N_2 \) are diagonally dominant. For these we have (with a slight abuse of notation), noting that \( A(\Phi_{N_1}, \Phi_{B_0}) = A(\Phi_{N_2}, \Phi_{B_1}) = 0 \),

\[
s_{N_1 N_2} + s_{N_1 P_j} + A(\Phi_{N_1}, \Phi_{B_1}) = c_l = b - c_l - d = 0, \quad l = 1, 2,
\]

where the summations are taken over the interior nodes \( P_j \) different from \( N_i \), including the new node \( P_0 \). Here \( a_l > 0 \) and \( c_l \geq 0 \) since \( N_i P_j \) are Delauney. Furthermore, \( b = s_{N_1 N_2} = s_{N_2 N_1} > 0 \) since \( N_1 N_2 \) 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 \( S \) 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 \( S \) 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 \( \Omega \), so that \( NB_1 B_2 \) is an equilateral triangle. By erecting suitable obtuse triangles \( Q_1 N B_1 \) and \( Q_2 N B_2 \) outside the triangle \( NB_1 B_2 \), we can arrange that the edges \( N B_1 \) and \( N B_2 \) 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 \( S \). We have

\[
s_{N N} + \sum_j s_{N P_j} + (A(\Phi_N, \Phi_{B_1}) + A(\Phi_N, \Phi_{B_2})) = a - c + d = 0,
\]

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 \( T_h \), and given \( a_{ij} \), it is possible to have a \( b \) such that the semigroup \( E_h(t) \) on \( S_h^0 \) cannot be contractive in maximum-norm. For, with \( u_h(t) = E_h(t) u_h \), we have

\[
\frac{1}{2} \frac{d}{dt} \|u_h\|^2 = -A_0(u_h, u_h) + \frac{1}{2} (\text{div}_h u_h, u_h), \quad \text{where} \quad \| \cdot \| = (\cdot, \cdot)_{h}^{1/2}.
\]

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


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

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use