

# ON POSITIVITY AND MAXIMUM-NORM CONTRACTIVITY IN TIME STEPPING METHODS FOR PARABOLIC EQUATIONS

A.H. SCHATZ<sup>1</sup>, V. THOMÉE<sup>2</sup>, AND L.B. WAHLBIN<sup>3</sup>

**Abstract** — In an earlier paper the last two authors studied spatially semidiscrete piecewise linear finite element approximations of the heat equation and showed that, in the case of the standard Galerkin method, the solution operator of the initial-value problem is neither positive nor contractive in the maximum-norm for small time, but that for the lumped mass method these properties hold, if the triangulations are essentially of Delaunay type. In this paper we continue the study by considering fully discrete analogues obtained by discretization also in time. The above properties then carry over to the backward Euler time stepping method, but for other methods the results are more restrictive. We discuss in particular the  $\theta$ -method and the  $(0, 2)$  Padé approximation in one space dimension.

**Keywords:** parabolic equations, time-stepping, finite elements, lumped mass, positivity, maximum-norm contractivity.

**2000 Mathematics Subject Classification:** 65M12, 65M60.

## 1. Introduction

We consider the initial boundary value problem

$$u_t - \Delta u = 0 \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega, \quad \text{for } t \geq 0, \quad \text{with } u(0) = v, \quad (1.1)$$

in a convex polygonal domain  $\Omega \subset \mathbb{R}^2$ . By the maximum principle the maximum and minimum of a smooth solution are attained for  $t = 0$ . This implies that the solution operator defined by  $E(t)v = u(t)$  is a positive operator in the sense that

$$\text{if } v \geq 0 \quad \text{in } \Omega, \quad \text{then } E(t)v \geq 0 \quad \text{in } \Omega, \quad \text{for } t \geq 0,$$

and also that  $E(t)$  is a (nonstrict) contraction in the maximum-norm, i.e.,

$$\|E(t)v\|_\infty = \|u(t)\|_\infty \leq \|v\|_\infty, \quad \text{for } t \geq 0, \quad \text{where } \|v\|_\infty = \sup_{x \in \Omega} |v(x)|.$$

---

<sup>1</sup>*Department of Mathematics, Cornell University, Ithaca N.Y. 14853, USA. E-mail: schatz@math.cornell.edu*

<sup>2</sup>*Department of Mathematical Sciences, Chalmers University of Technology, S-41296 Göteborg, Sweden. E-mail: thomee@chalmers.se*

<sup>3</sup>*Department of Mathematics, Cornell University, Ithaca N.Y. 14853, USA. E-mail: wahlbin@math.cornell.edu*

The paper [8] considered the spatially semidiscrete finite element analogue of (1.1) in  $S_h$ , the set of continuous piecewise linear functions on a triangulation  $\mathcal{T}_h$ , to find  $u_h(t) \in S_h$  such that, with  $(v, w) = \int_{\Omega} vw \, dx$ ,  $A(v, w) = (\nabla v, \nabla w)$ ,

$$(u_{h,t}, \chi) + A(u_h, \chi) = 0, \quad \text{for } \chi \in S_h, \, t \geq 0, \quad \text{with } u_h(0) = v_h \approx v. \tag{1.2}$$

Defining the semidiscrete solution operator by  $E_h(t)v_h = u_h(t)$ , it was shown that  $E_h(t)$  can essentially neither be a positive operator nor a contraction in  $\|\cdot\|_{\infty}$  for all  $t > 0$ .

The lumped mass variant of (1.2) was also investigated, i.e., to find  $\bar{u}_h(t) \in S_h$  such that

$$(\bar{u}_{h,t}, \chi)_h + A(\bar{u}_h, \chi) = 0, \quad \text{for } \chi \in S_h, \, t \geq 0, \quad \text{with } \bar{u}_h(0) = v_h, \tag{1.3}$$

where  $(\cdot, \cdot)_h$  is a specific quadrature version of  $(\cdot, \cdot)$ , namely

$$(v, w)_h = \sum_{\tau \in \mathcal{T}_h} Q_{\tau,h}(v\bar{w}), \quad \text{with } Q_{\tau,h}(f) = \frac{1}{3} \text{area}(\tau) \sum_{j=1}^3 f(P_{\tau,j}) \approx \int_{\tau} f \, dx.$$

In this case it was shown that, for the solution operator defined by  $\bar{E}_h(t)v_h = \bar{u}_h(t)$ , if the triangulation  $\mathcal{T}_h$  is essentially of Delaunay type, then  $\bar{E}_h(t)$  is a positive operator and a contraction in  $\|\cdot\|_{\infty}$  for  $t \geq 0$ .

Our purpose in this paper is to study the corresponding problems of positivity and maximum-norm contractivity for time discrete versions of the above problems. We begin by studying time discretization in a more general Banach space framework, following Bolley and Crouzeix [1] and Kovács [5].

Let  $\mathcal{B}$  be a Banach space with norm  $\|\cdot\|$ , and assume that  $-A$  generates a contraction  $C_0$  semigroup  $E(t) = e^{-At}$  in  $\mathcal{B}$ . The semigroup  $E(t)$  is then the solution operator of the abstract initial value problem

$$u' + Au = 0, \quad \text{for } t \geq 0, \quad \text{with } u(0) = v, \tag{1.4}$$

and the solution of this problem thus satisfies

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

If  $\mathcal{B}$  is an ordered Banach space we say that  $E(t)$  is positive, or  $E(t) \geq 0$ , if  $u(t) = E(t)v \geq 0$  for  $v \geq 0$ , where 0 is the zero element of  $\mathcal{B}$ .

Let  $r(z)$  be an  $A$ -stable rational function, i.e., with  $|r(z)| \leq 1$  for  $\text{Re } z \geq 0$ , and assume in addition consistency, i.e., that  $r(z)$  approximates the exponential  $e^{-z}$  in the sense that  $r(z) = 1 - z + O(z^2)$  as  $z \rightarrow 0$ ; for brevity we shall call such a rational function  $A$ -correct below. For the time discretization of (1.4), let  $k$  be a time step and  $t_n = nk$ . An approximate solution at  $t = t_n$  is then

$$U^n = E_k^n v, \quad \text{where } E_k = r(kA), \tag{1.5}$$

and we may ask if the time-stepping operator  $E_k^n$  is contractive in  $\mathcal{B}$ , and/or positive if  $\mathcal{B}$  is ordered.

We note that the time stepping operator  $E_k^n$  is contractive for all  $n \geq 1$  if and only if  $E_k$  is contractive, and positive if and only if  $E_k$  is. The most basic example is the backward Euler method, with  $r(z) = 1/(1+z)$ . In this case the contraction property of  $E_k$  is immediate. In fact we have

$$E_k v = (I + kA)^{-1} v = \int_0^{\infty} e^{-t} E(kt) v \, dt, \quad \text{for } k > 0, \, v \in \mathcal{B}, \tag{1.6}$$

and hence, if  $E(t)$  is a contraction,

$$\|E_k\| \leq \int_0^\infty e^{-t} \|E(kt)\| dt \leq \int_0^\infty e^{-t} dt = 1.$$

If  $\mathcal{B}$  is ordered, with a closed positive cone  $\mathcal{B}^+ = \{v \in \mathcal{B}, v \geq 0\}$ , then  $E_k$  is also positive as follows at once from (1.6).

We want to ask if these properties carry over to more general rational functions  $r(z)$  in the above framework, and in the concrete cases referred to above.

Consider first the case that  $\mathcal{B}$  is a Hilbert space  $\mathcal{H}$  and  $A$  is a selfadjoint positive definite, not necessarily bounded, linear operator with a compact inverse. Then, if  $\{\lambda_j\}_{j=1}^\infty$  are the eigenvalues of  $A$  and  $\{\phi_j\}_{j=1}^\infty$  a corresponding basis of orthonormal eigenfunctions, we have for the solution operator of (1.4),

$$E(t)v = \sum_{j=1}^\infty e^{-\lambda_j t} (v, \phi_j) \phi_j \quad \text{and} \quad \|E(t)v\| \leq \sup_j e^{-\lambda_j t} \|v\| = e^{-\lambda_1 t} \|v\| \leq \|v\|,$$

so that  $E(t)$  is a contraction in  $\mathcal{H}$ . For the corresponding time stepping operator  $E_k$  in (1.5) we have, if  $|r(\lambda)| \leq 1$  for  $\lambda \geq 0$ , that

$$E_k v = \sum_{j=1}^\infty r(k\lambda_j) (v, \phi_j) \phi_j \quad \text{and} \quad \|E_k v\| = \sup_j |r(k\lambda_j)| \|v\| \leq \|v\|, \quad \text{for } k > 0,$$

so that  $E_k$  is a contraction for any  $k > 0$ . In particular this holds in the  $L_2$  norm for the concrete initial boundary value problem (1.1) and its spatial discretizations by the standard Galerkin method (1.2) and the lumped mass method (1.3).

However, we shall see that when we consider contractivity in maximum-norm and the related property of positivity, things are more complicated, and that, in general, these properties of  $E_k = r(kA)$  hold only in exceptional cases.

In Section 2 below we discuss some general results within the framework of the Banach space  $\mathcal{B}$ . We first show that if  $r(z)$  is bounded for  $\operatorname{Re} z \geq 0$ , then we may write

$$r(z) = \int_0^\infty g(t) e^{-zt} dt + r(\infty), \quad \text{with } g \in L_1(\mathbb{R}_+),$$

and we say that  $r(z)$  is of *positive type* if  $g(t) \geq 0$  and  $r(\infty) \geq 0$ . In this case  $E_k = r(kA)$  is a contraction if  $r(0) = 1$ , and positive if  $\mathcal{B}$  is ordered with  $\mathcal{B}^+$  closed.

Unfortunately, for  $r(z)$  to be of positive type is exceptional and can only happen for first order approximations to  $e^{-z}$ . If  $\|E(t)\| \rightarrow 0$  as  $t \rightarrow \infty$  and  $|r(\infty)| < 1$ , however,  $E_k$  is a contraction for large  $k$ , even when  $r(z)$  is not of positive type.

In Section 3 we give some examples relating to positivity and contractivity in the Banach space  $\mathcal{C}_0(\mathbb{R})$  of continuous functions on  $\mathbb{R}$  which vanish at  $\pm\infty$ , with norm  $\|\cdot\|_\infty$ . For  $A = D := d/dx$  we show that positivity and contractivity for all  $k > 0$  requires  $r(z)$  to be of positive type. If  $A = -\varepsilon D^2 + D$ , with  $\varepsilon > 0$ , then  $E_k$  cannot be positive or contractive for any  $k > 0$ , for arbitrarily small  $\varepsilon$ , unless  $r(z)$  is of positive type. For  $A = -D^2$  we give a simple criterion on  $r(z)$  which excludes positivity and contractivity of  $E_k$ , but also an example of an  $r(z)$  which is not of positive type but for which  $E_k$  is both positive and contractive, for all  $k > 0$ .

In Sections 4, 5, and 6 we then study the positivity and maximum-norm contractivity of such time discretization operators, applied to the spatially continuous problem (1.1) and to its spatially semidiscrete standard Galerkin and lumped mass analogues (1.2) and (1.3), respectively. In each of these sections, as also earlier, we discuss in some detail the  $\theta$ -method defined by

$$r_\theta(z) = \frac{1 - (1 - \theta)z}{1 + \theta z} = \frac{1}{\theta} \frac{1}{1 + \theta z} - \frac{1 - \theta}{\theta}, \quad 0 \leq \theta \leq 1. \tag{1.7}$$

For the spatially continuous problem we show that, except for the backward Euler method ( $\theta = 1$ ),  $E_k = r_\theta(kA)$  cannot be positive for any  $k > 0$  and not maximum-norm contractive for small  $k$ . The finite element versions were studied in Fujii [3], where sufficient conditions for the maximum-principle were given in terms of properties of the triangulation  $\mathcal{T}_h$ . For the standard Galerkin method it was shown that the maximum-principle holds when  $\mathcal{T}_h$  is of Delaunay type if  $k$  is bounded above and below in a specific way. Since neither positivity nor contraction holds for the spatially semidiscrete problem, we show here that these properties also cannot be satisfied for the corresponding time stepping methods, for  $k$  small. The lumped mass method is more advantageous in that a maximum-principle holds for  $k$  small when  $\mathcal{T}_h$  is of Delaunay type. We include versions of these results below.

As an example of an  $A$ -correct rational function, which is of higher order than first, we also consider the  $(0, 2)$  Padé approximation of  $e^{-z}$  defined by

$$r(z) = r_{02}(z) = \frac{1}{1 + z + \frac{1}{2}z^2}. \tag{1.8}$$

We study the corresponding time discrete operator  $r_{02}(kA)$ , restricting ourselves to one space dimension, and using uniform meshes in the spatially discrete cases. We shall see that, even under these restrictive assumptions, the corresponding operator  $E_k = r_{02}(kA)$  is neither positive nor maximum-norm contractive for small  $k$ , in any of the three concrete cases considered, but that these properties hold for larger  $k$ .

## 2. Some results on time stepping in a Banach space framework

We first discuss a representation of a rational function, bounded in the right half-plane of  $\mathbb{C}$ , of the infinitesimal generator of a  $C_0$  semigroup in a Banach space.

**Lemma 2.1.** *Let  $r(z)$  be a rational function, bounded for  $\operatorname{Re} z \geq 0$ , and let  $-A$  be the infinitesimal generator of a bounded  $C_0$  semigroup  $E(t) = e^{-At}$  on the Banach space  $\mathcal{B}$ . Then  $E_k := r(kA)$  is well defined for  $k > 0$  and we may write*

$$E_k = r(kA) = \int_0^\infty g(t) E(kt) dt + r(\infty) I, \quad \text{for } k > 0. \tag{2.1}$$

If  $\{\zeta_j\}_{j=1}^J$  with multiplicities  $\{m_j\}_{j=1}^J$  are the poles of  $r(z)$ , then  $g(t)$  has the form

$$g(t) = \sum_{j=1}^J P_j(t) e^{\zeta_j t}, \tag{2.2}$$

where  $P_j$  is a polynomial of degree  $m_j - 1$ .

*Proof.* Noting that

$$\frac{1}{(z - \zeta)^{n+1}} = \int_0^\infty \frac{1}{n!} t^n e^{-zt} e^{\zeta t} dt, \quad \text{for } \operatorname{Re}(z - \zeta) > 0,$$

we find by partial fraction decomposition of  $r(z)$  that, with  $g(t)$  as in (2.2),

$$r(z) = \int_0^\infty e^{-zt} g(t) dt + r(\infty), \quad \text{for } \operatorname{Re} z \geq 0. \quad (2.3)$$

The representation (2.1) then follows by replacing  $z$  by  $kA$ , and thus  $e^{-zt}$  by  $E(kt)$ , see, e.g., [6], p. 20.  $\square$

For the backward Euler method we have  $r(z) = 1/(1+z)$ , and (2.3) holds with  $g(t) = e^{-t}$ ,  $r(\infty) = 0$ , and for  $r_\theta(z)$  defined in (1.7),  $g(t) = \theta^{-2}e^{-\theta t}$  and  $r(\infty) = -\theta^{-1}(1-\theta)$ . For  $r_{02}(z)$  as in (1.8), we have  $g(t) = 2e^{-t} \sin t$ ,  $r(\infty) = 0$ , as follows from

$$\begin{aligned} r_{02}(z) &= \frac{2}{(z+1+i)(z+1-i)} = i \left( \frac{1}{z+1+i} - \frac{1}{z+1-i} \right) \\ &= i \int_0^\infty (e^{-(z+1+i)t} - e^{-(z+1-i)t}) dt = 2 \int_0^\infty e^{-zt} e^{-t} \sin t dt. \end{aligned} \quad (2.4)$$

We recall that  $r(z)$  is said to be of positive type if  $g(t) \geq 0$ ,  $r(\infty) \geq 0$ . We remark, cf. [1], that by a well-known theorem of Bernstein this holds if and only if  $r(x)$  is completely monotone on  $\mathbb{R}_+$ , or if  $(-1)^n r^{(n)}(x) \geq 0$  for all  $n \geq 0$ ,  $x \in \mathbb{R}_+$ .

We now give an upper bound for  $\|E_k\|$ , which shows that if  $E(t)$  is a contraction, then  $E_k$  is a contraction if  $r(z)$  is of positive type and  $r(0) = 1$ , cf. [1], [5].

**Theorem 2.1.** *Let the rational function  $r(z)$  satisfy (2.3), and let  $E(t)$  be a contraction. Then we have, for  $E_k := r(kA)$ ,*

$$\|E_k\| \leq \int_0^\infty |g(t)| dt + |r(\infty)|, \quad \text{for } t > 0. \quad (2.5)$$

*If  $r(z)$  is of positive type and  $r(0) = 1$ , then  $E_k$  is a contraction.*

*Proof.* The inequality (2.5) follows at once from (2.1). If  $r(z)$  is of positive type and  $r(0) = 1$ , then (2.3) with  $z = 0$  shows that the right hand side of (2.5) equals 1, and thus  $E_k$  is a contraction.  $\square$

Following [1] and [5], we also have the following immediate consequence of Lemma 2.1.

**Theorem 2.2.** *Let  $\mathcal{B}$  be ordered, with its positive cone  $\mathcal{B}^+$  closed. If  $E(t)$  is positive and  $r(z)$  is of positive type, then  $E_k = r(kA)$  is also positive.*

Unfortunately, for  $r(z)$  to be of positive type is exceptional. For example, for the  $\theta$ -method,  $E_k = r_\theta(kA)$ , we have  $r(\infty) = -\theta^{-1}(1-\theta) < 0$  for  $\theta < 1$ , so that  $r_\theta(z)$  is not of positive type, and (2.5) only shows

$$\|E_k\| \leq \int_0^\infty \theta^{-2} e^{-t/\theta} dt + \theta^{-1}(1-\theta) = 2\theta^{-1} - 1.$$

For the Crank-Nicolson method we have  $\theta = \frac{1}{2}$ ,  $r(\infty) = -1$ , and (2.5) only shows  $\|E_k\| \leq 3$ . Neither is  $r_{02}(z)$  of positive type since the sign of  $g(t) = 2e^{-t} \sin t$  varies.

It was shown in [1] that  $r(z)$  cannot be of positive type for approximations of order higher than first. For the convenience of the reader we include a proof.

**Theorem 2.3.** *Assume that  $r(z)$  is  $A$ -correct and of positive type. Then  $r(z)$  cannot approximate  $e^{-z}$  to second order.*

*Proof.* If  $r(z)$  approximates  $e^{-z}$  to second order, we have  $r(z) = e^{-z} + O(z^3) = 1 - z + \frac{1}{2}z^2 + O(z^3)$  as  $z \rightarrow 0$ , so that  $r^{(j)}(0) = (-1)^j$ ,  $j = 0, 1, 2$ , where, by (2.3),

$$r^{(j)}(0) = \int_0^\infty (-t)^j g(t) dt + r(\infty)\delta_{0j}.$$

Since  $g(t) \geq 0$ , the Cauchy-Schwarz inequality yields

$$1 = r'(0)^2 = \left( \int_0^\infty t g(t) dt \right)^2 \leq \int_0^\infty g(t) dt \int_0^\infty t^2 g(t) dt \leq r(0) r''(0) = 1.$$

Here equality requires  $g(t)$  and  $t^2 g(t)$  to be proportional, which is impossible. □

It follows that the only  $A$ -stable Padé approximation of  $e^{-z}$  of positive type is the backward Euler method. This does not exclude that a time discretization operator  $E_k = r(kA)$  of higher order than first could be a contraction in special cases, such as the Hilbert space situation discussed in Section 1, in which any  $A$ -stable rational function corresponds to a contraction.

We note that the bound in (2.5) can be relatively small, even when  $E_k$  is not a contraction. For instance, for  $E_k = r_{02}(kA)$  we have, in any Banach space,

$$\|E_k\| = \|r_{02}(kA)\| \leq 2 \int_0^\infty e^{-t} |\sin t| dt = \frac{1 + e^{-\pi}}{1 - e^{-\pi}} = 1.0903, \quad \text{for } k > 0. \quad (2.6)$$

This is a special case of the following lemma which will also be used later.

**Lemma 2.2.** *We have*

$$2 \int_0^\infty e^{-\gamma t} |\sin t| dt = \frac{1 + e^{-\gamma\pi}}{1 - e^{-\gamma\pi}} r_{02}(\gamma - 1), \quad \text{for } \gamma > 1.$$

*Proof.* By (2.4) with  $z = \gamma - 1 \in \mathbb{R}_+$  we find

$$2 \int_0^\infty e^{-\gamma t} \sin t dt = r_{02}(\gamma - 1).$$

Hence, setting  $s = t - \pi$  in the second integral below,

$$\begin{aligned} X &:= 2 \int_0^\pi e^{-\gamma t} \sin t dt = r_{02}(\gamma - 1) - 2 \int_\pi^\infty e^{-\gamma t} \sin t dt \\ &= r_{02}(\gamma - 1) + e^{-\gamma\pi} 2 \int_0^\infty e^{-\gamma s} \sin s ds = (1 + e^{-\gamma\pi}) r_{02}(\gamma - 1). \end{aligned}$$

By the same change of variable as above,

$$Y := 2 \int_0^\infty e^{-\gamma t} |\sin t| dt = X + 2 \int_\pi^\infty e^{-\gamma t} |\sin t| dt = X + e^{-\gamma\pi} Y,$$

from which the result follows. □

We next show that  $E_k = r(kA)$  could be a contraction for some positive  $k$ , even when this does not hold for all such  $k$ .

**Theorem 2.4.** *Assume  $\|E(t)\| \rightarrow 0$  as  $t \rightarrow \infty$  and that  $r(z)$  is an  $A$ -correct rational function with  $|r(\infty)| < 1$ . Then there is a  $k_0 > 0$  such that  $E_k = r(kA)$  is a contraction for  $k \geq k_0$ .*

*Proof.* Since  $g \in \mathcal{C}(\mathbb{R}) \cup L_1(\mathbb{R}_+)$ , we have, by (2.1) and Lebesgue’s theorem,

$$\|E_k\| \leq \int_0^\infty |g(t)| \|E(kt)\| dt + |r(\infty)| \rightarrow |r(\infty)| < 1, \quad \text{as } k \rightarrow \infty,$$

which shows the assertion. □

Since  $r(\infty) \geq 0$  is not necessary for contractivity, in general, the following theorem now shows that positivity and contractivity are not equivalent properties.

**Theorem 2.5.** *Let  $\mathcal{B}$  be ordered, with  $\mathcal{B}^+$  closed, and assume  $\|E(t)\| \rightarrow 0$  as  $t \rightarrow \infty$  and that  $r(z)$  is an  $A$ -correct rational function. Then  $r(\infty) \geq 0$  is a necessary condition for  $E_k = r(kA)$  to be positive for large  $k$ .*

*Proof.* Again, since  $g \in \mathcal{C}(\mathbb{R}) \cup L_1(\mathbb{R}_+)$ , we have, by Lebesgue’s theorem, for  $x \in \mathbb{R}$ ,

$$E_k v(x) = \int_0^\infty g(t) E(kt)v(x) dt + r(\infty)v(x) \rightarrow r(\infty)v(x), \quad \text{as } k \rightarrow \infty,$$

which shows the claim. □

### 3. Some examples in $\mathcal{C}_0(\mathbb{R})$ using Fourier transformation

In this section we shall illustrate the above concepts and results concerning positivity and contractivity with some examples in the Banach space  $\mathcal{C}_0(\mathbb{R})$  of continuous functions on  $\mathbb{R}$  which vanish at  $\pm\infty$ , normed with the maximum-norm  $\|\cdot\|_\infty$ . We note that this Banach space is ordered, with a closed positive cone  $\mathcal{B}^+$ .

We first give an example where positivity of  $E_k = r(kA)$  requires  $r(z)$  to be of positive type. The example is provided by a convection diffusion equation with small or no diffusion. In particular, the operator  $E_k$  cannot be positive for any  $k > 0$ , for all  $E(t)$  corresponding to parabolic equations, unless  $r(z)$  is of positive type.

**Theorem 3.1.** *Let  $\mathcal{B} = \mathcal{C}_0(\mathbb{R})$  and let  $E_\varepsilon(t)$  be the semigroup generated in  $\mathcal{B}$  by  $-A_\varepsilon$ , where  $A_\varepsilon = -\varepsilon d^2/dx^2 + d/dx$ , with  $\varepsilon \geq 0$ . Let further  $r(z)$  be an  $A$ -correct rational function. Then, in order for  $E_{0,k} = r(kA_0)$  to be positive for some  $k > 0$ , it is necessary that  $r(z)$  be of positive type. Also, unless  $r(z)$  is of positive type,  $E_{\varepsilon,k}$  has to be non-positive for  $\varepsilon$  sufficiently small.*

*Proof.* We express  $E_\varepsilon(t)$  in terms of Fourier transforms as

$$E_\varepsilon(t)v(x) = \mathcal{F}^{-1}(e^{-t(\varepsilon\xi^2+i\xi)}\widehat{v})(x), \quad \text{where } \widehat{v}(\xi) = \mathcal{F}v(\xi) = \int_{\mathbb{R}} e^{-ix\xi}v(x) dx. \quad (3.1)$$

We begin by considering the hyperbolic case,  $\varepsilon = 0$ . Then the semigroup is the translation semigroup  $E_0(t)v(x) = v(x + t)$  on  $\mathbb{R}$ , and, by Lemma 2.1,

$$E_{0,k}v(x) = \int_0^\infty g(t)v(x + kt) dt + r(\infty)v(x), \quad \text{for } k > 0. \tag{3.2}$$

Let  $k > 0$  be fixed, and assume  $E_{0,k} \geq 0$ . If  $g(t) < 0$  on some interval  $(a, b) \subset \mathbb{R}_+$ , then, with  $v \geq 0$  such that  $v(kt)$  has support in  $(a, b)$ , we have  $E_{0,k}v(0) < 0$ , in contradiction to our assumption. Similarly,  $r(\infty) < 0$  is impossible.

If  $\varepsilon > 0$  we note that  $E_\varepsilon(t) \geq 0$  by the maximum-principle for parabolic equations. Further, for sufficiently smooth  $v \in \mathcal{B}$ , and  $x \in \mathbb{R}$ ,  $t \geq 0$ , we have  $E_\varepsilon(t)v(x) \rightarrow E_0(t)v(x)$  as  $\varepsilon \rightarrow 0$ . In fact, since  $|e^{-t\varepsilon\xi^2} - 1| \leq Ct\varepsilon\xi^2$ , (3.1) shows

$$\|(E_\varepsilon(t) - E_0(t))v\|_\infty \leq Ct\varepsilon \int_{-\infty}^\infty \xi^2 |\widehat{v}(\xi)| d\xi = Ct\varepsilon \|\widehat{v''}\|_{L_1}. \tag{3.3}$$

Hence, if  $g(t) < 0$  in  $(a, b)$  and  $v$  is as above, we have, by Lebesgue's theorem,

$$E_{\varepsilon,k}v(0) = \int_0^\infty g(t)E_\varepsilon(kt)v dt + r(\infty)v(0) \rightarrow E_{0,k}v(0), \quad \text{as } \varepsilon \rightarrow 0,$$

and thus  $E_{\varepsilon,k}v(0) < 0$  for small  $\varepsilon$ , so that  $E_{\varepsilon,k}$  is then not positive. □

Staying in the situation of Theorem 3.1 we now turn to contractivity. For this we show that the bound in (2.5) is attained in the case  $\varepsilon = 0$ , and is approached in the limit as  $\varepsilon \rightarrow 0$ , for  $\varepsilon > 0$ . In particular, this shows that  $E_k$  cannot be a contraction in  $\mathcal{C}_0(\mathbb{R})$  for all  $E(t)$  corresponding to parabolic equations unless  $r(z)$  is of positive type.

**Theorem 3.2.** *Under the assumptions of Theorem 3.1 we have, for any  $k > 0$ ,*

$$\|E_{\varepsilon,k}\|_\infty \rightarrow \|E_{0,k}\|_\infty = \int_0^\infty |g(t)| dt + |r(\infty)|, \quad \text{as } \varepsilon \rightarrow 0. \tag{3.4}$$

*If  $r(z)$  is  $A$ -correct, and if  $E_{0,k}$  is a contraction for some  $k > 0$ , then  $r(z)$  is of positive type. Also, if  $E_{\varepsilon,k}$  is a contraction for some  $k > 0$  and arbitrarily small  $\varepsilon > 0$ , then  $r(z)$  is of positive type.*

*Proof.* We first show the equality in (3.4). For  $k > 0$  fixed, taking  $v = v_n \in \mathcal{B}$  in (3.2) such that  $\|v_n(k\cdot) - \text{sgn}(g)\|_{L_1(\mathbb{R}_+)} \rightarrow 0$  as  $n \rightarrow \infty$  and  $\|v_n\|_\infty = 1$ ,  $v_n(0) = \text{sgn}(r(\infty))$ , we find, using (3.2) with  $x = 0$ , as  $n \rightarrow \infty$ ,

$$\|E_{0,k}\|_\infty \geq \|E_{0,k}v_n\|_\infty / \|v_n\|_\infty \geq |E_{0,k}v_n(0)| \rightarrow \int_0^\infty |g(t)| dt + |r(\infty)|.$$

Thus the opposite inequality to (2.5) holds, which shows equality. In particular,  $\|E_{0k}\|_\infty = 1$  if and only if  $r(z)$  is of positive type, since otherwise the right side of (2.5) is greater than  $r(0) = 1$ .

For the last statement of the theorem, assume  $r(z)$  is not of positive type. Then we may fix  $v \in \mathcal{B}$ , sufficiently smooth, with  $\|v\|_\infty = 1$  such that  $\|E_{0k}v\|_\infty > 1$ , and therefore  $\|E_{\varepsilon,k}v\|_\infty > 1$  for  $\varepsilon$  sufficiently small, since, by (3.3),

$$\|(E_{\varepsilon,k} - E_{0,k})v\|_\infty \leq \int_0^\infty |g(t)| \|(E_\varepsilon(t) - E_0(t))v\|_\infty dt \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0.$$

Hence  $E_{\varepsilon,k}$  cannot be a contraction, □

We now consider time stepping methods  $E_k = r(kA)$  for the standard heat equation, i.e., with  $A = -d^2/dx^2$ , using  $A$ -correct rational functions  $r(z)$ . If  $r(\infty) = 0$ , then since  $\mathcal{F}(r(kA)v) = r(k\xi^2)\widehat{v}$  and  $r(k\xi^2) \in L_1(\mathbb{R})$  we have, for  $v \in L_1(\mathbb{R}) \cap \mathcal{B}$ ,

$$E_k v = \mathcal{F}^{-1}(r(k\xi^2)\widehat{v}) = f_k * v, \quad \text{where } f_k = \mathcal{F}^{-1}(r(k\xi^2)). \tag{3.5}$$

Setting  $f(x) = f_1(x) = \mathcal{F}^{-1}(r(\xi^2))$ , we have  $f_k(x) = k^{-1/2}f(k^{-1/2}x)$ .

In the case of the backward Euler method we then have  $f(x) = \mathcal{F}^{-1}((1 + \xi^2)^{-1}) = \frac{1}{2}e^{-|x|}$ , and since  $f \geq 0$ , the positivity of  $E_k$  follows at once from (3.5). Further,  $\|E_k\|_\infty = \|f_k\|_{L_1} = \|f\|_{L_1} = 1$ , showing again that  $E_k$  is a contraction.

For the  $\theta$ -method we correspondingly find

$$E_k v = f_k * v - \theta^{-1}(1 - \theta)v, \quad \text{with } f(x) = f_1(x) = \frac{1}{2}\theta^{-1}e^{-|x|/\theta}.$$

We next show that  $E_k$  cannot be positive or contractive for  $\theta < 1$ .

**Theorem 3.3.** *Let  $\mathcal{B} = C_0(\mathbb{R})$  and  $A = -d^2/dx^2$ . Then  $E_k = r_\theta(kA)$ , with  $0 < \theta < 1$ , can neither be positive nor contractive, for any  $k > 0$ .*

*Proof.* Let  $v \in \mathcal{B}$  with  $v \geq 0$ ,  $v(0) = 1$ , and with compact support, and set  $v_\varepsilon(x) = v(x/\varepsilon)$ . Then  $\|f_k * v_\varepsilon\|_\infty \rightarrow 0$  as  $\varepsilon \rightarrow 0$  so that  $E_k v_\varepsilon(0) \rightarrow -\theta^{-1}(1 - \theta) < 0$ , which shows that  $E_k$  cannot be positive.

Choosing instead  $v_\varepsilon \in \mathcal{B}$  with  $\|v_\varepsilon\|_\infty = 1$ ,  $v_\varepsilon(0) = -1$ , and  $v_\varepsilon(x) = 1$  for  $\varepsilon \leq |x| \leq \varepsilon^{-1}$ , we obtain  $E_k v_\varepsilon(0) \rightarrow \|f_k\|_{L_1} + \theta^{-1}(1 - \theta) = \theta^{-1} > 1$ , which shows that  $E_k$  cannot be contractive. □

We note that the last statement is not in conflict with Theorem 2.4, since  $\|E(t)\|_\infty = 1$  for  $t \geq 0$ .

We now show that for the standard heat equation, for a certain class of rational functions,  $E_k$  is neither positive nor contractive. By numerical calculation, we found that this class includes the  $A$ -correct subdiagonal Padé approximations  $r_{m-j,m}(z)$ ,  $m = 2, \dots, 5$ ,  $j = 1, 2$ . We shall assume that  $r(z)$  has poles at  $z_j = \rho_j e^{i\theta_j}$ ,  $j = 0, 1, \dots, J$ , with  $\frac{1}{2}\pi < \theta_j \leq \pi$ . In case of nonreal poles,  $\theta_j < \pi$ ,  $\bar{z}_j$  is also a pole. We further assume that the pole  $z_0$  is distinguished in the sense that

$$z_0 \text{ is nonreal and simple, and } \sqrt{\rho_0} \sin(\frac{1}{2}\theta_0) < \sqrt{\rho_j} \sin(\frac{1}{2}\theta_j), \quad j = 1, \dots, J. \tag{3.6}$$

**Theorem 3.4.** *Let  $\mathcal{B} = C_0(\mathbb{R})$  and  $A = -d^2/dx^2$ . Let  $r(z)$  be an  $A$ -correct rational function with  $r(\infty) = 0$ , and assume that the poles of  $r(z)$  satisfy (3.6). Then  $E_k = r(kA)$  can neither be positive nor contractive, for any  $k > 0$ .*

*Proof.* In this case we may write

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} r(\xi^2) e^{ix\xi} d\xi = i \sum_j \text{Res}(r(\xi_j^2) e^{ix\xi_j}), \quad \text{for } x > 0, \tag{3.7}$$

where the summation is over the poles of  $r(\xi^2)$  with  $\text{Im } \xi_j > 0$ . We shall show that  $f(x)$  changes sign, so that, in particular, by (3.5),  $E_k$  cannot be positive. Further, since  $\|E_k\|_\infty = \|f_k\|_{L_1} = \|f\|_{L_1}$ , independently of  $k$ , and since  $\int_{-\infty}^{\infty} f(x) dx = \mathcal{F}f(0) = r(0) = 1$ , we may then also infer that  $\|f\|_{L_1} > 1$  so that  $E_k$  cannot be a contraction.

By partial fraction decomposition,

$$r(z) = \frac{a}{z - z_0} + \frac{\bar{a}}{z - \bar{z}_0} + \tilde{r}(z),$$

where  $z_0$  or  $\bar{z}_0$  are not poles of  $\tilde{r}(z)$ . The corresponding poles of  $r(\xi^2)$  with  $\text{Im } \xi_j > 0$  are

$$\xi_1 = \sqrt{z_0} = \sqrt{\rho_0} e^{i\frac{1}{2}\theta_0} = \alpha + i\beta, \quad \xi_2 = \sqrt{\bar{z}_0} = \sqrt{\rho_0} e^{-i(\frac{1}{2}\theta_0 + \pi)} = -\alpha + i\beta.$$

Note that  $\beta = \sqrt{\rho_0} \sin(\frac{1}{2}\theta_0)$ . Since  $e^{ix\xi_l} = e^{-x\beta}(\cos \alpha x \pm i \sin \alpha x)$  for  $l = 1, 2$ , we see that, for  $x > 0$ ,  $f(x)$  takes the form

$$f(x) = -\frac{e^{-\beta x}}{\alpha^2 + \beta^2}(\alpha \text{Im}(ae^{i\alpha x}) + \beta \text{Re}(ae^{i\alpha x}) + g_1(x)) = e^{-\beta x}(b \sin(\alpha x + \phi) + g_2(x)),$$

where  $b$  is real, and, by (3.6),  $g_2(x) = o(1)$  as  $x \rightarrow +\infty$ . Obviously,  $f(x)$  thus has to change sign. □

We remark that the above argument also shows that  $E_k = r(kA)$  cannot be positive in the ordered Banach space  $\mathcal{B} = L_2(\mathbb{R})$  unless  $r(z)$  is of positive type. However, if  $r(z)$  is  $A$ -correct, then  $E_k$  is a contraction in  $L_2(\mathbb{R}_+)$ , which provides another example that positivity and contractivity are not equivalent properties.

Together with (2.4), Theorem 3.4 shows, in particular, that  $E_k = r_{02}(kA)$  is not a contraction in  $\mathcal{C}_0(\mathbb{R})$ . However, we shall now see that in this case, a considerably sharper bound than (2.6) holds.

**Theorem 3.5.** *With  $\mathcal{B}$  and  $A$  as in Theorem 3.4, and  $E_k = r_{02}(kA)$ , we have  $\|E_k\|_\infty \approx 1.0014$ .*

*Proof.* In this case we find from (3.7)

$$f(x) = \frac{2}{\pi} \int_0^\infty \frac{\cos x\xi}{\xi^4 + 2\xi^2 + 2} d\xi, \quad \text{for } x > 0,$$

and we shall use the formula ([4], p. 411, 3.733 #1)

$$\int_0^\infty \frac{\cos x\xi d\xi}{\xi^4 + 2b^2\xi^2 \cos 2\tau + b^4} = \frac{\pi}{2b^3} e^{-xb \cos \tau} \frac{\sin(\tau + xb \sin \tau)}{\sin 2\tau}$$

to obtain

$$f(x) = b^{-1} e^{-xb \cos \tau} \sin(\tau + xb \sin \tau), \quad \text{with } b = 2^{1/4}, \tau = \pi/8.$$

We now use the transformation of variables  $y = \tau + x b \sin \tau$ , to find

$$\|f\|_{L_1} = 2 \int_0^\infty |f(x)| dx = \frac{\sqrt{2}}{\sin \tau} e^{\tau \cot \tau} \int_\tau^\infty e^{-y \cot \tau} |\sin y| dy.$$

By Lemma 2.2 we have, using  $r_{02}(\cot \tau - 1) = 2 \sin^2 \tau$ ,

$$\int_0^\infty e^{-y \cot \tau} |\sin y| dy = \frac{1}{2} \frac{1 + e^{-\pi \cot \tau}}{1 - e^{-\pi \cot \tau}} r_{02}(\cot \tau - 1) = \sin^2 \tau \left( 1 + \frac{2}{e^{\pi \cot \tau} - 1} \right),$$

and, by a simple calculation, since  $\sin y > 0$  in  $(0, \tau)$ ,

$$\int_0^\tau e^{-y \cot \tau} |\sin y| dy = \int_0^\tau e^{-y \cot \tau} \sin y dy = \sin^2 \tau (1 - 2 \cos \tau e^{-\tau \cot \tau}).$$

Hence, using also the fact that  $2 \sin \tau \cos \tau = \sin(\pi/4) = 1/\sqrt{2}$ ,

$$\|E_k\|_\infty = \|f\|_{L_1} = \frac{\sqrt{2}}{\sin \tau} e^{\tau \cot \tau} \left( \int_0^\infty - \int_0^\tau \right) = 1 + \frac{2\sqrt{2} \sin \tau e^{\tau \cot \tau}}{e^{\pi \cot \tau} - 1} \approx 1.0014.$$

□

We close this section by giving an example of an  $A$ -correct rational function  $r(z)$  which is not of positive type but for which  $E_k = r(kA)$ , with  $\mathcal{B}$  and  $A$  as in Theorem 3.4, is both positive and contractive. This shows that a rational function does not have to be of positive type for these properties to hold.

**Theorem 3.6.** *Let  $\mathcal{B}$  and  $A$  be as in Theorem 3.4. Then there exists an  $A$ -correct rational function  $r(z)$  with  $r(\infty) = 0$ , which is not of positive type, such that  $E_k = r(kA)$  is positive and contractive.*

*Proof.* We consider a convex combination of the backward Euler rational function and  $r_{02}(kA)$ , corresponding to

$$r(z) = \nu/(1+z) + (1-\nu)r_{02}(z), \quad \text{with } \nu \in (0, 1).$$

Here,  $g(t) = \nu e^{-t} + (1-\nu) 2 \sin t e^{-t}$  and  $r(\infty) = 0$ , and, in the above notation,

$$f(x) = \mathcal{F}^{-1}r(\xi^2)(x) = \nu \frac{1}{2} e^{-x} + (1-\nu) b^{-1} e^{-x b \cos \tau} \sin(\tau + x b \sin \tau), \quad \text{for } x > 0.$$

We want to choose  $\nu$  so that  $f(x) \geq 0$ , in which case  $E_k$  is positive and contractive as above, but such that  $g(t)$  does not have constant sign, and thus  $r(z)$  is not of positive type. But the latter requirement is satisfied if  $2(1-\nu) > \nu$ , i.e., if  $\nu < 2/3$ , and, since  $b \cos \tau \approx 1.0987 > 1$ , the first requirement holds if  $(1-\nu)/b < \nu/2$ , or  $\nu > 2/(2+b) \approx 0.6271$ . Thus, for  $\nu \in (0.63, 0.66)$ ,  $r(z)$  has the desired properties. □

### 4. The spatially continuous problem

We now consider the initial-boundary value problem (1.1), or (1.4) with  $A = -\Delta$ , in the Banach space  $\mathcal{B} = \mathcal{C}_0(\Omega)$  of continuous functions in  $\bar{\Omega}$ , vanishing on  $\partial\Omega$ , with norm  $\|\cdot\|_\infty$ , where  $\Omega$  is a convex polygonal domain in  $\mathbb{R}^2$ , and recall that the solution operator  $E(t)$  is both positive and a contraction for  $t \geq 0$ . As we have seen in Section 1 this implies that the backward Euler method  $E_k = (I + kA)^{-1}$  inherits these properties, for all  $k > 0$ . Also, we note that

$$\|E(t)v\|_\infty \leq C \|AE(t)v\|_{L_2}^{\frac{1}{2}} \|E(t)v\|_{L_2}^{\frac{1}{2}} \leq Ct^{-\frac{1}{2}} \|v\|_{L_2} \leq Ct^{-\frac{1}{2}} \|v\|_\infty, \tag{4.1}$$

so that, in particular,  $\|E(t)\|_\infty \rightarrow 0$  as  $t \rightarrow \infty$ . (This convergence is, in fact, exponential, but we shall not have reason to go into the details.) Thus, by Theorem 2.4,  $E_k = r(kA)$  is a contraction for large  $k$  when  $r(z)$  is an  $A$ -correct rational function with  $|r(\infty)| < 1$ .

We next turn to the  $\theta$ -method defined by  $E_k = r_\theta(kA)$  with  $r_\theta(z)$  as in (1.7). We shall see that even though the backward Euler operator ( $\theta = 1$ ) is positive and contractive for any  $k > 0$ , this is not the case for  $\theta < 1$ . Note that  $r_\theta(z)$  is  $A$ -stable for  $1/2 \leq \theta \leq 1$ .

**Theorem 4.1.** *When  $1/2 \leq \theta < 1$  the time stepping operator  $E_k = r_\theta(kA)$  for (1.1) cannot be positive for any  $k > 0$ .*

*Proof.* Let  $k > 0$  and  $\theta$  with  $1/2 \leq \theta < 1$  be fixed. By (1.7) we may write

$$E_k v = r_\theta(kA)v = \theta^{-1}(I + k\theta A)^{-1}v - \theta^{-1}(1 - \theta)v. \tag{4.2}$$

By (1.6) and (4.1) we find

$$\begin{aligned} \|(I + \theta kA)^{-1}v\|_\infty &\leq \int_0^\infty e^{-t} \|E(\theta kt)v\|_\infty dt \\ &\leq C(\theta k)^{-\frac{1}{2}} \int_0^\infty t^{-\frac{1}{2}} e^{-t} dt \|v\|_{L_2} = C(\theta k)^{-\frac{1}{2}} \|v\|_{L_2}. \end{aligned} \tag{4.3}$$

Now choose  $x_0 \in \Omega$  and  $v = v_\varepsilon$  such that  $0 \leq v_\varepsilon \leq 1$ ,  $v_\varepsilon(x_0) = 1$ , and  $\text{supp}(v_\varepsilon) \subset \{x; |x - x_0| \leq \varepsilon\}$ . Then  $\|v_\varepsilon\|_{L_2} \leq C\varepsilon$ , and hence, with  $v = v_\varepsilon$ , the first term in (4.2) then tends to zero at  $x_0$  as  $\varepsilon \rightarrow 0$ . Hence  $E_k v_\varepsilon(x_0) \rightarrow -(1 - \theta)/\theta < 0$ , so that  $E_k v_\varepsilon(x_0) < 0$  for  $\varepsilon$  small and thus  $E_k$  nonpositive.  $\square$

**Theorem 4.2.** *Let  $1/2 \leq \theta < 1$ , let  $\lambda_1$  be the smallest eigenvalue of  $A = -\Delta$ , and let  $k_0$  be so small that*

$$\frac{1}{\theta} \frac{1}{1 + k_0\theta\lambda_1} + \frac{1}{\theta} > 2. \tag{4.4}$$

*Then  $E_k = r_\theta(kA)$  cannot be a contraction for  $k \leq k_0$ .*

*Proof.* Let  $\Psi_1$  be the first eigenfunction of  $A$ , normalized so that, with  $x_0 \in \Omega$ ,  $\|\Psi_1\|_\infty = \Psi_1(x_0) = 1$ . Let  $\Psi_{1,\varepsilon}$  be a modification of  $\Psi_1$  in  $\{x; |x - x_0| \leq \varepsilon\}$  with  $\|\Psi_{1,\varepsilon}\|_\infty = 1$ ,  $\Psi_{1,\varepsilon}(x_0) = -1$ . Then we have, by (4.3),

$$\|(I + k\theta A)^{-1}(\Psi_{1,\varepsilon} - \Psi_1)\|_\infty \leq C(k\theta)^{-1/2} \|\Psi_{1,\varepsilon} - \Psi_1\|_{L_2} \rightarrow 0, \quad \text{as } \varepsilon \rightarrow 0.$$

Since  $(I + k\theta A)^{-1}\Psi_1(x_0) = (1 + k\theta\lambda_1)^{-1}$  we conclude, by (4.2) and (4.4),

$$\begin{aligned} E_k \Psi_{1,\varepsilon}(x_0) &= \frac{1}{\theta} (I + k\theta A)^{-1}\Psi_{1,\varepsilon}(x_0) - \frac{1 - \theta}{\theta} \Psi_{1,\varepsilon}(x_0) \\ &\rightarrow \frac{1}{\theta} \frac{1}{1 + k\theta\lambda_1} + \frac{1 - \theta}{\theta} > 1, \quad \text{as } \varepsilon \rightarrow 0, \quad \text{for } k \leq k_0. \end{aligned}$$

Thus  $E_k$  cannot be a contraction for  $k \leq k_0$ .  $\square$

Note that for the Crank-Nicolson method ( $\theta = 1/2$ ), (4.4) holds for any  $k_0 > 0$ . Thus  $E_k$  then cannot be a contraction for any  $k > 0$ . However, we have the following consequence of Theorem 2.4.

**Theorem 4.3.** *Let  $1/2 < \theta < 1$ . Then there is a  $k_1 > 0$  such that  $E_k = r_\theta(kA)$  is a contraction for  $k \geq k_1$ .*

We shall now discuss the (0,2) Padé method, in one space dimension. Here  $\mathcal{B} = \mathcal{C}_0(\Omega)$ ,  $\Omega = (0, 1)$ ,  $A = -d^2/dx^2$ , and (1.4) reduces to

$$u_t = u_{xx} \quad \text{in } \Omega, \quad u(x, t) = 0 \quad \text{for } x = 0, 1, \quad \text{for } t \geq 0, \quad u(\cdot, 0) = v. \tag{4.5}$$

Clearly the solution operator  $E(t)$  is still both positive and a contraction for  $t \geq 0$ , and analogously for the backward Euler operator, for  $k > 0$ . Further we have

$$\|E(t)v\|_\infty \leq \|E(t)v\|_{L_2}^{\frac{1}{2}} \|\nabla E(t)v\|_{L_2}^{\frac{1}{2}} \leq Ct^{-\frac{1}{4}}\|v\|_{L_2}, \quad \text{for } t > 0. \tag{4.6}$$

We first show that  $E_k = r_{02}(kA)$  is neither positive nor a contraction in  $\mathcal{C}_0(\Omega)$  for all  $k > 0$ , by reducing this to the case  $\Omega = \mathbb{R}$ , already covered by Theorem 3.4.

**Theorem 4.4.** *The operator  $E_k = r_{02}(kA)$  associated with (4.5) is neither positive nor contractive in  $\mathcal{C}_0(\Omega)$  for all  $k > 0$ .*

*Proof.* Assume  $E_k$  is positive for all  $k > 0$ . Then by a transformation of variables this holds also when the interval  $\Omega = (0, 1)$  is replaced by  $\Omega_\omega = (-\omega, \omega)$ , for any  $\omega > 0$ . We denote the solution operator on  $\Omega_\omega$  by  $E^{(\omega)}(t)$  and the corresponding time stepping operator by  $E_k^{(\omega)}$ . But with  $\tilde{E}(t)$  and  $\tilde{E}_k$  the corresponding operators for the problem on all of  $\mathbb{R}$ ,

$$\tilde{E}_k v(x) - E_k^{(\omega)} v(x) = 2 \int_0^\infty e^{-t} \sin t \left( \tilde{E}(t) - E^{(\omega)}(t) \right) v(x) dt. \tag{4.7}$$

We claim that  $E^{(\omega)}(t)v(x) \rightarrow \tilde{E}(t)v(x)$  as  $\omega \rightarrow \infty$  for each  $x \in \mathbb{R}$ ,  $t > 0$ . In fact, the function  $w_\omega(x, t) = \tilde{E}(t)v(x) - E^{(\omega)}(t)v(x)$  satisfies the heat equation on  $\Omega_\omega \times \mathbb{R}_+$ , has initial values 0, and  $|w_\omega(x, t)| = |\tilde{E}(t)v(x)| \leq \|v\|_\infty$  for  $x = \pm\omega$ ,  $t > 0$ . Hence, by the maximum-principle,  $|w_\omega(x, t)|$  is bounded by the solution of the heat equation in the finite domain, with initial values 0, and boundary values  $\|v\|_\infty$  for  $x = \pm\omega$ . It is then easily seen that for fixed  $(x, t) \in \mathbb{R} \times \mathbb{R}_+$ ,  $w_\omega(x, t) \rightarrow 0$  as  $\omega \rightarrow \infty$ . Hence, using Lebesgue's theorem in (4.7), we conclude that  $E_k^{(\omega)} v(x) \rightarrow \tilde{E}_k v(x)$  as  $\omega \rightarrow \infty$  for each  $x \in \mathbb{R}$ ,  $k > 0$ . Since  $E_k^{(\omega)} v \geq 0$  for  $v \geq 0$  and each  $\omega > 0$ , we must have  $\tilde{E}_k v(x) \geq 0$  for  $v \geq 0$ , which is in contradiction to Theorem 3.4.

In the same way, if  $E_k$  is a contraction for all  $k > 0$ , this also holds for  $E_k^{(\omega)}$  and hence for  $\tilde{E}_k$ , again in contradiction to Theorem 3.4. □

By the same argument, the corresponding result holds for any rational function satisfying the conditions of Theorem 3.4.

We now show some positive results, for larger  $k$ . We first note that contractivity follows at once from Theorem 2.4 and (4.6):

**Theorem 4.5.** *For (4.5) there is a  $k_1 > 0$  such that  $E_k = r_{02}(kA)$  is a contraction for  $k \geq k_1$ .*

We next exhibit positivity of  $E_k = r_{02}(kA)$  for large  $k$ .

**Theorem 4.6.** *For (4.5) there is a  $k_1 > 0$  such that  $E_k = r_{02}(kA) \geq 0$ , for  $k \geq k_1$ .*

*Proof.* We first show that if  $v \geq 0$ , then the solution  $u(x, t) = E(t)v(x)$  of (4.5) is decreasing for  $t \geq 0.1$ , for all  $x \in \Omega$ . Recalling that the eigenfunctions and -values of  $A$  are  $\phi_l(x) = \sqrt{2} \sin(\pi l x)$  and  $\lambda_l = \pi^2 l^2$  for  $l = 1, 2, \dots$ , we have

$$E(t)v(x) = \int_0^1 G(x, y, t)v(y) dy, \quad \text{where } G(x, y, t) = 2 \sum_{l=1}^\infty e^{-\pi^2 l^2 t} \sin(\pi l x) \sin(\pi l y).$$

By termwise differentiation we find, for  $x, y \in \Omega$ ,

$$\frac{\partial}{\partial t} G(x, y, t) = -2\pi^2 \sin(\pi x) \sin(\pi y) e^{-\pi^2 t} \left( 1 + \sum_{l=2}^{\infty} l^2 e^{-\pi^2(l^2-1)t} \frac{\sin(\pi l x)}{\sin(\pi x)} \frac{\sin(\pi l y)}{\sin(\pi y)} \right).$$

Denoting the sum by  $\tilde{S}(x, y, t)$ , we have, since  $|\sin(\pi l x)/\sin(\pi x)| \leq l$  for  $x \in \Omega$ ,

$$|\tilde{S}(x, y, t)| \leq \tilde{S}(t) := \sum_{l=2}^{\infty} l^4 e^{-\pi^2(l^2-1)t}, \quad \text{for } x, y \in \Omega.$$

Here  $\tilde{S}(t)$  is decreasing in  $t$ , and we find that  $\tilde{S}(0.096) = 0.974$ . Hence  $|\tilde{S}(x, y, t)| \leq 1$  for  $t \geq 0.096$ , and thus  $G(x, y, t)$  is decreasing for  $t \geq 0.096$ , for all  $x, y \in \Omega$ . Therefore  $E(t)v(x)$  is decreasing for  $t \geq 0.096$ ,  $x \in \Omega$ . Now fix  $\tau \in (0, \pi)$  and let  $k\tau \geq 0.096$ . Then

$$E_k v(x) = r_{02}(kA)v(x) = \sum_{l=0}^{\infty} I_l, \quad \text{where } I_l := 2 \int_{2\pi l}^{2\pi(l+1)} e^{-t} \sin t E(kt)v(x) dt.$$

Since  $E(kt)v(x)$  is decreasing for  $t \geq \tau \in (0, \pi)$ , and thus in each  $I_l$  with  $l \geq 1$ , we find easily that  $I_l \geq 0$  for  $l \geq 1$ , and, if we choose  $\tau = 2.34$ ,

$$I_0 \geq 2 \left( \int_{\tau}^{\pi} - \int_{\pi}^{2\pi} \right) e^{-t} |\sin t| dt E(k\pi)v(x) = (e^{-\tau}(\cos \tau + \sin \tau) - e^{-2\pi}) E(k\pi)v \geq 0.$$

Hence  $E_k v(x) \geq 0$  if  $k \geq k_1 = 0.096/\tau = 0.0416$ . □

### 5. The standard Galerkin finite element method

We now consider the two-dimensional spatially semidiscrete standard Galerkin problem (1.2). Defining the discrete Laplacian  $\Delta_h : S_h \rightarrow S_h$  by

$$-(\Delta_h \psi, \chi) = (\nabla \psi, \nabla \chi), \quad \forall \psi, \chi \in S_h, \tag{5.1}$$

this may also be expressed as

$$u_{h,t} - \Delta_h u_h = 0, \quad \text{for } t > 0, \quad \text{with } u_h(0) = v_h, \tag{5.2}$$

and we write  $E_h(t) = e^{\Delta_h t}$  for the solution operator of this problem.

In this and the following section we shall use as our Banach space  $\mathcal{B} = S_h$ , with norm  $\|\cdot\|_{\infty}$ . We know from [8] that  $E_h(t)$  is neither a positive operator nor a contraction in  $\mathcal{B}$  for small  $t$ . More precisely, we have the following two theorems, shown under technical assumptions satisfied by normal triangulations.

**Theorem 5.1.** *Assume 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 for small  $t$ .*

**Theorem 5.2.** *Assume that  $\mathcal{T}_h$  is such that each near-boundary node has a strictly interior neighbor. Then  $E_h(t)$  cannot be a contraction in  $\|\cdot\|_{\infty}$  for small  $t$ .*

We now show that, for a fixed triangulation  $\mathcal{T}_h$ ,  $E_h(t)$  is both positive and contractive for large  $t$ . Here and below, we denote by  $\{P_i\}_{i=1}^N$  the nodes of  $\mathcal{T}_h$  in the interior of  $\Omega$ , and let  $\{\Phi_i\}_{i=1}^N \subset S_h$  be the standard basis of pyramid functions, defined by  $\Phi_i(P_j) = \delta_{ij}$ .

**Theorem 5.3.** *For  $\mathcal{T}_h$  given,  $E_h(t) \geq 0$  for  $t$  sufficiently large. Also,  $\|E_h(t)\|_\infty \rightarrow 0$  as  $t \rightarrow \infty$ . In particular,  $E_h(t)$  is a contraction for  $t$  large.*

*Proof.* With  $\{\phi_j^h\}_{j=1}^N$  and  $\{\lambda_j^h\}_{j=1}^N$  the eigenfunctions and -values of the positive definite operator  $A = -\Delta_h$  in  $S_h$ , we have, for  $v_h \in S_h$  such that  $v_h \geq 0$ ,  $v_h \not\equiv 0$ ,

$$E_h(t)v_h = \sum_{j=1}^N e^{-\lambda_j^h t} (v_h, \phi_j^h) \phi_j^h = e^{-\lambda_1^h t} ((v_h, \phi_1^h) \phi_1^h + o(1)), \quad \text{as } t \rightarrow \infty,$$

where we have used that  $\lambda_2^h > \lambda_1^h$ . It is known by the Perron-Frobenius theorem that  $\phi_1^h(P_i) > 0$  for all mesh-points  $P_i$ , and it follows that  $(v_h, \phi_1^h) \phi_1^h(P_i) > 0$  and hence that  $E_h(t)v_h(P_i) > 0$ , for  $t$  large,  $i = 1, \dots, N$ . In particular, this holds for each of the basis functions  $\Phi_l$ , and, since these are finitely many, there is a  $t_0 > 0$  such that  $E_h(t)\Phi_l \geq 0$  for  $l = 1, \dots, N$  and  $t \geq t_0$ . This shows the first part of the theorem.

Since  $\|E_h(t)\|_{L_2} \leq C e^{-\lambda_1^h t} \leq C e^{-\lambda_1 t}$ , the second part of the theorem follows at once from the equivalence of different norms on a finite-dimensional space.  $\square$

We next show that under a mild assumption on the family  $\{\mathcal{T}_h\}$ , we have, independently of  $h$ , the maximal diameter of the triangles  $\tau \in \mathcal{T}_h$ , that  $\|E_h(t)\|_\infty \rightarrow 0$  as  $t \rightarrow \infty$ . Thus, in particular,  $E_h(t)$  is a contraction for large  $t$ , uniformly in  $h$ . Our assumption is the following:

$$\{\mathcal{T}_h\} \text{ is shape regular and } h_{\min} \geq ch^\gamma, \text{ for some } c > 0, \gamma \geq 1. \tag{5.3}$$

Such conditions occur, e.g., in systematic mesh refinements.

**Theorem 5.4.** *Assume that  $\{\mathcal{T}_h\}$  satisfies (5.3). Then we have, with  $C$  independent of  $h$ ,*

$$\|E_h(t)v_h\|_\infty \leq Ct^{-1}\|v_h\|_\infty, \quad \text{for } t > 0, v_h \in S_h.$$

*Proof.* We write, with  $I_h : \mathcal{C}_0(\Omega) \rightarrow S_h$  the standard interpolation operator,

$$\|E_h(t)v_h\|_\infty \leq \|(E_h(t) - I_h E(t))v_h\|_\infty + \|I_h E(t)v_h\|_\infty = I + II.$$

By Sobolev’s inequality and a standard smoothing estimate for  $E(t)$ ,

$$II \leq \|E(t)v_h\|_\infty \leq C\|E(t)v_h\|_{H^2} \leq Ct^{-1}\|v_h\|_{L_2} \leq Ct^{-1}\|v_h\|_\infty, \quad \text{for } t > 0.$$

Further, under our assumptions on  $\{\mathcal{T}_h\}$ , we note that by [7], Lemma 6.4,

$$\|\chi\|_\infty \leq C\ell_h^{1/2}\|\nabla\chi\|_{L_2}, \quad \forall \chi \in S_h, \quad \text{where } \ell_h = \max(1, \log(1/h)).$$

and, by [2], Theorem 2.1, we have the nonsmooth data error estimate

$$\|\nabla(E_h(t)P_h - E(t))v\|_{L_2} \leq Cht^{-1}\|v\|_{L_2}, \quad \text{for } t > 0, \quad \text{if } v \in L_2, v_h = P_h v, \tag{5.4}$$

where  $P_h$  is the  $L_2$ -projection onto  $S_h$ . Thus, using also a standard error estimate for the interpolant, we find

$$\begin{aligned} I &\leq C\ell_h^{1/2}(\|\nabla(E_h(t) - E(t))v_h\|_{L_2} + \|\nabla(I_h - I)E(t)v_h\|_{L_2}) \\ &\leq C\ell_h^{1/2}(ht^{-1}\|v_h\|_{L_2} + h\|E(t)v_h\|_{H^2}) \leq Ct^{-1}\|v_h\|_{L_2} \leq Ct^{-1}\|v_h\|_\infty, \end{aligned}$$

which completes the proof.  $\square$

In order to study time-discretizations of the semidiscrete problem (1.2), or (5.2), it will be convenient to express it in matrix form. For this purpose, we identify a function  $\chi = \sum_{j=1}^N \chi_j \Phi_j \in S_h$  with the vector  $\tilde{\chi} = (\chi_1, \dots, \chi_N)^T$ . Letting  $\mathcal{M} = (m_{ij})$  and  $\mathcal{S} = (s_{ij})$ , with  $m_{ij} = (\Phi_i, \Phi_j)$  and  $s_{ij} = A(\Phi_i, \Phi_j)$ , be the mass and stiffness matrices, respectively, the application of the operator  $-\Delta_h$  corresponds to multiplication by the matrix  $\mathcal{K} = \mathcal{M}^{-1}\mathcal{S}$ . In fact, if  $\psi = \sum_i \alpha_i \Phi_i$ ,  $-\Delta_h \psi = \sum_i \beta_i \Phi_i$ , then (5.1) may be written  $\mathcal{M}\beta \cdot \tilde{\chi} = \mathcal{S}\alpha \cdot \tilde{\chi}$  for all  $\tilde{\chi} \in \mathbb{R}^N$ , or  $\beta = \mathcal{K}\alpha$ . Note that  $-\Delta_h$  is positive definite with respect to  $(v, w)_{L_2}$  and  $\mathcal{K}$  with respect to  $\mathcal{M}\alpha \cdot \beta$ .

The initial value problem (1.2) then takes the form, with  $\alpha_i(t) = u_h(P_i, t)$ ,

$$\mathcal{M} \frac{d\alpha}{dt} + \mathcal{S}\alpha = 0, \quad \text{for } t \geq 0, \quad \text{with } \alpha(0) = \tilde{v}_h = (v_h(P_1), \dots, v_h(P_N))^T,$$

and we introduce the solution operator matrix  $\mathcal{E}_h(t) = e^{-t\mathcal{K}}$ , where  $\mathcal{K} = \mathcal{M}^{-1}\mathcal{S}$ . The solution operator  $E_h(t)$  of (5.2) is then positive if and only if  $\mathcal{E}_h(t)$  is positive (or  $\mathcal{E}_h(t) \geq 0$ , elementwise), and a contraction in  $\|\cdot\|_\infty$  if and only if  $\mathcal{E}_h(t)$  is a contraction in  $|\cdot|_\infty$ , where  $|\tilde{\chi}|_\infty = \max_i |\chi_i| = \|\chi\|_\infty$  is the vector maximum-norm, since  $\|E_h(t)\|_\infty = |\mathcal{E}_h(t)|_\infty$ .

We now consider, with  $r(z)$  an  $A$ -correct rational function, a single step time discretization  $E_{hk}^n v_h$  of (5.2), where  $E_{hk} = r(-k\Delta_h)$ , or, in matrix formulation,

$$\alpha^n = \mathcal{E}_{hk}^n \alpha^0, \quad \text{for } n \geq 0, \quad \text{with } \alpha^0 = \tilde{v}_h, \quad \text{where } \mathcal{E}_{hk} = r(k\mathcal{K}). \tag{5.5}$$

We observe that the negative results of Theorems 5.1 and 5.2 immediately imply that the time stepping operator  $E_{hk}$  cannot be positive or contractive for small  $k$ .

**Theorem 5.5.** *Let  $r(z)$  be  $A$ -correct, and let  $E_{hk} = r(-k\Delta_h)$ . If  $\mathcal{T}_h$  satisfies the assumption of Theorem 5.1, then  $E_{hk}$  cannot be positive for small  $k$ . If  $\mathcal{T}_h$  satisfies the assumption of Theorem 5.2, then  $E_{hk}$  cannot be a contraction for small  $k$ .*

*Proof.* Using the matrix representation (5.5), both results follow from

$$\lim_{n \rightarrow \infty} \mathcal{E}_{h,t/n}^n = \lim_{n \rightarrow \infty} \left( \mathcal{I} + \frac{t}{n} \mathcal{K} + O\left(\frac{t^2}{n^2}\right) \right)^{-n} = e^{-t\mathcal{K}} = \mathcal{E}_h(t),$$

since positivity or contractivity of  $\mathcal{E}_{hk}$ , for small  $k$ , would imply the corresponding property for  $\mathcal{E}_h(t)$ , and thus of  $E_h(t)$ , for  $t > 0$ , in contradiction to Theorems 5.1 and 5.2.  $\square$

This does not exclude that these properties could hold for larger  $k$ . For instance, the following is an immediate consequence of Theorems 2.4, 2.5, and 5.4.

**Theorem 5.6.** *Assume that (5.3) holds, and let  $r(z)$  be  $A$ -correct with  $|r(\infty)| < 1$ . Then the operator  $E_{hk} = r(-k\Delta_h)$  is contractive for large  $k$ , uniformly in  $h$ . Further,  $r(\infty) \geq 0$  is a necessary condition for  $E_{hk}$  to be positive for  $k$  large.*

We now turn to a discussion of the  $\theta$ -method, corresponding to the rational function  $r_\theta(z)$  in (1.7). In this case the time stepping operator  $E_{hk} = r_\theta(-k\Delta_h)$  is defined, in matrix form, by

$$\mathcal{M}(\alpha^{n+1} - \alpha^n)/k + \mathcal{S}(\theta\alpha^{n+1} + (1 - \theta)\alpha^n) = 0, \quad \text{for } n \geq 0, \quad \text{with } \alpha^0 = \tilde{v}, \tag{5.6}$$

or by

$$\mathcal{E}_{hk} = r_\theta(k\mathcal{M}^{-1}\mathcal{S}) = (\mathcal{M} + \theta k\mathcal{S})^{-1}(\mathcal{M} - (1 - \theta)k\mathcal{S}).$$

We first demonstrate that  $\mathcal{E}_{hk}$ , and thus  $E_{hk}$ , is positive and contractive when  $\mathcal{T}_h$  is essentially of Delaunay type, and the time step  $k$  is bounded above and below in a specific way, cf. [3]. Recall that  $\mathcal{T}_h$  is of Delaunay type if the sum of the angles opposite each edge of  $\mathcal{T}_h$  is at most  $\pi$ , and that for an interior edge  $P_iP_j$  this is equivalent to  $s_{ij} \leq 0$ .

**Theorem 5.7.** *Let the matrices  $\mathcal{M}$  and  $\mathcal{S}$  be such that  $s_{ij} \leq 0$  for  $j \neq i$ , and*

$$(1 - \theta)ks_{ii} \leq m_{ii}, \quad \forall i, \quad \text{and} \quad \theta k|s_{ij}| \geq m_{ij}, \quad \text{for } j \neq i. \quad (5.7)$$

*Then  $E_{hk} = r_\theta(-k\Delta_h)$  is positive. If, in addition,  $\sum_{j \neq i} |s_{ij}| \leq s_{ii}$ , for all  $i$ , then  $E_{hk}$  is also a contraction.*

*Proof.* The inequality on the right in (5.7) means that  $m_{ij} + \theta ks_{ij} \leq 0$  for all  $j \neq i$ , so that  $\mathcal{M} + \theta k\mathcal{S}$  is a Stieltjes matrix, i.e., a positive definite symmetric matrix with nonpositive off-diagonal elements, and thus has a nonnegative inverse. Further, the inequality on the left in (5.7) shows that  $\mathcal{M} - (1 - \theta)k\mathcal{S}$  is nonnegative. Together these facts show that  $\mathcal{E}_{hk} \geq 0$ , and thus also  $E_{hk} \geq 0$ .

With  $\beta = \alpha^{n+1}$ ,  $\alpha = \alpha^n$ , (5.6) can be written

$$(\mathcal{M} + \theta k\mathcal{S})\beta = (\mathcal{M} - (1 - \theta)k\mathcal{S})\alpha. \quad (5.8)$$

To show that  $\mathcal{E}_{hk}$  is a contraction, or  $|\mathcal{E}_{hk}|_\infty \leq 1$ , it suffices to show that this implies

$$|\beta|_\infty \leq |\alpha|_\infty. \quad (5.9)$$

Let  $P_i$  be an arbitrary interior node. Then the  $i$ th equation in (5.8) is

$$(m_{ii} + \theta ks_{ii})\beta_i = - \sum_{j \neq i} (m_{ij} + \theta ks_{ij})\beta_j + \sum_j (m_{ij} - (1 - \theta)ks_{ij})\alpha_j.$$

From this it follows that

$$(m_{ii} + \theta ks_{ii})|\beta_i| \leq \sum_{j \neq i} |m_{ij} + \theta ks_{ij}| |\beta|_\infty + \sum_j |m_{ij} - (1 - \theta)ks_{ij}| |\alpha|_\infty.$$

Then, if  $P_i$  is chosen so that  $|\beta_i| = |\beta|_\infty$ , we have, using (5.7), that

$$\begin{aligned} (m_{ii} + \theta ks_{ii})|\beta|_\infty &\leq \left( \theta k \sum_{j \neq i} |s_{ij}| - \sum_{j \neq i} m_{ij} \right) |\beta|_\infty \\ &\quad + \left( (m_{ii} - (1 - \theta)ks_{ii}) + \sum_{j \neq i} (m_{ij} + (1 - \theta)k|s_{ij}|) \right) |\alpha|_\infty. \end{aligned}$$

Using the diagonal dominance of  $\mathcal{S}$  it follows that  $\sum_j m_{ij} |\beta|_\infty \leq \sum_j m_{ij} |\alpha|_\infty$ , and hence that (5.9) holds. Thus  $E_{hk}$  is a contraction.  $\square$

Recall that for  $P_i$  strictly interior, i.e., if  $P_i$  has no neighbor on  $\partial\Omega$ , and if  $s_{ij} \leq 0$  for  $j \neq i$ , then, since  $\sum_j s_{ij} = \sum_j A(\Phi_i, \Phi_j) = A(\Phi_i, 1) = 0$ , we have  $\sum_{j \neq i} |s_{ij}| = s_{ii}$ , so that diagonal dominance is automatic at such nodes. Thus the condition  $\sum_{j \neq i} |s_{ij}| \leq s_{ii}$  is associated with

the behavior of the triangulation near  $\partial\Omega$ , cf., e.g., [8], Examples 3.1 and 3.2, which show that  $s_{ij} \leq 0$ , for  $j \neq i$ , and diagonal dominance of  $\mathcal{S}$  are independent properties.

For the backward Euler method, (5.7) reduces to  $k \geq \max_{j \neq i} (m_{ij}/|s_{ij}|)$ . We note that this condition requires  $s_{ij} = A(\Phi_i, \Phi_j) < 0$  for  $P_i$  and  $P_j$  neighbors, i.e., that  $\mathcal{T}_h$  is strictly Delaunay, in the sense that the sum of the angles opposite  $P_i P_j$  is strictly less than  $\pi$ . In particular, this cannot hold for the standard triangulation  $\mathcal{T}_h$  of a square  $\Omega$  obtained by first subdividing  $\Omega$  into smaller square and then dividing these by their diagonals, because then there will be neighbors for which  $s_{ij} = 0$ . As a positive example, if  $\Omega$  is an equilateral triangle, and  $\mathcal{T}_h$  consists of equilateral triangles, thus with all angles  $= \pi/3$ , then one may show that, if  $P_i$  and  $P_j$  are neighbors, then  $s_{ii} = 2\sqrt{3}$ ,  $s_{ij} = -\sqrt{3}/3$  and  $m_{ii} = h^2\sqrt{3}/4$ ,  $m_{ij} = h^2\sqrt{3}/24$ . The above condition (5.7) is then

$$\frac{1}{\theta} \frac{h^2}{8} \leq k \leq \frac{1}{1-\theta} \frac{h^2}{8},$$

and this can only be satisfied if  $\theta \geq 1/2$ . For the backward Euler case, the condition reduces to the one-sided condition  $k \geq h^2/8$ , and for Crank-Nicolson, we must have  $k = h^2/4$ . We remark that  $r_\theta(\infty) = -(1-\theta)/\theta < 0$  for  $\theta < 1$  so that by Theorem 5.6 an upper bound for  $k$  is required for positivity of  $\mathcal{E}_{hk}$ .

We now turn to the (0,2) Padé method, and we restrict ourselves again to the one-dimensional case, with  $S_h$  based on uniform partitions, using  $x_j = jh$ ,  $j = 1, \dots, N$ ,  $h = 1/(N+1)$ . The solution operator for the semidiscrete version of (4.5) is then  $E_h(t) = e^{\Delta_h t}$  where  $\Delta_h$  is defined by the one-dimensional analogue of (5.1). For the corresponding solution operator matrix we have  $\mathcal{E}_h(t) = e^{-\mathcal{K}t}$  with  $\mathcal{K} = \mathcal{M}^{-1}\mathcal{S}$ , where now, with  $\mathcal{J}$  the tridiagonal matrix with elements 1 on the two bidiagonals and other elements 0,

$$\mathcal{M} = (m_{ij}) = \frac{1}{6} h(4\mathcal{I} + \mathcal{J}) \quad \text{and} \quad \mathcal{S} = (s_{ij}) = h^{-1}(2\mathcal{I} - \mathcal{J}). \tag{5.10}$$

We note that in this case, cf. (4.6), since  $\|\nabla E_h(t)v_h\|_{L_2} \leq Ct^{-1/2}\|v_h\|_{L_2}$ ,

$$\|E_h(t)v_h\|_\infty \leq \|E_h(t)v_h\|_{L_2}^{1/2} \|\nabla E_h(t)v_h\|_{L_2}^{1/2} \leq Ct^{-1/4}\|v_h\|_{L_2} \leq Ct^{-1/4}\|v_h\|_\infty. \tag{5.11}$$

In particular,  $E_h(t)$  is a contraction for large  $t$ , uniformly in  $h$ .

As in the two-dimensional case treated above,  $E_h(t)$  is neither positive nor a contraction for  $k$  small, and as in Theorem 5.5 this also holds for  $E_{hk} = r_{02}(-k\Delta_h)$ . However, this operator is both positive and contractive for larger  $k$ . The contractivity follows at once by Theorem 2.4 and (5.11):

**Theorem 5.8.** *For our spatially one-dimensional problem there is  $k_0 > 0$ , independent of  $h$ , such that  $E_{hk} = r_{02}(-k\Delta_h)$  is a contraction for  $k \geq k_0$ .*

Direct calculation of the matrix norm  $|\mathcal{E}_{hk}|_\infty$  indicates that for  $h = 1/10, 1/20, 1/30$ , and  $1/40$ , we may choose  $k_0 = 0.011, 0.010, 0.010$ , and  $0.010$ , respectively.

We now show positivity for large  $k$ .

**Theorem 5.9.** *For our spatially one-dimensional problem, for  $h \leq 0.1$ ,  $E_{hk} = r_{02}(-k\Delta_h)$  is positive for  $k \geq k_1 = 0.5$ .*

*Proof.* In this one-dimensional situation the eigenvectors and -values of  $\mathcal{K}$  are

$$\{\phi_l^h(x_j)\}_{j=1}^N = \{\sqrt{2h} \sin(l\pi x_j)\}_{j=1}^N \quad \text{and} \quad \lambda_l^h = \frac{6}{h^2} \frac{1 - \cos(\pi lh)}{2 + \cos(\pi lh)}, \quad l = 1, \dots, N,$$

By eigenvector expansion we may write

$$\mathcal{E}_{kh} v_h(x_j) = \sum_{n=1}^N H_{jn}^{hk} v_h(x_n), \quad \text{where} \quad H_{jn}^{hk} = 2h \sum_{l=1}^N r_{02}(k\lambda_l^h) \sin(\pi l x_j) \sin(\pi l x_n),$$

and we want to show  $H_{jn}^{hk} \geq 0$  for  $1 \leq j, n \leq N$ . We write

$$H_{jn}^{hk} = 2h \sin(\pi x_j) \sin(\pi x_n) r_{02}(k\lambda_1^h) (1 + \tilde{S}_{jn}^{hk})$$

where

$$\tilde{S}_{jn}^{hk} = \sum_{l=2}^N \frac{r_{02}(k\lambda_l^h)}{r_{02}(k\lambda_1^h)} \frac{\sin(\pi l x_j)}{\sin(\pi x_j)} \frac{\sin(\pi l x_n)}{\sin(\pi x_n)},$$

and it now suffices to show  $|\tilde{S}_{jn}^{hk}| \leq 1$  for  $1 \leq j, n \leq N$ . As in the proof of Theorem 4.6, using  $\lambda_l^h \geq \lambda_l = \pi^2 l^2$  and  $|\sin(lx)/\sin x| \leq l$ , we find, with  $\kappa_h = \lambda_1^h/\pi^2$ ,

$$|\tilde{S}_{jn}^{hk}| \leq \tilde{S}^{hk} := \sum_{l=2}^N l^2 \frac{1 + k\lambda_1^h + \frac{1}{2}(k\lambda_1^h)^2}{1 + k\lambda_l + \frac{1}{2}(k\lambda_l)^2} \rightarrow \kappa_h^2 \sum_{l=2}^{\infty} \frac{1}{l^2} \approx 0.635 \kappa_h^2, \quad \text{as } k \rightarrow \infty,$$

Here  $\kappa_h \leq 1/(1 - \pi^2 h^2/6)$ , and hence, for the limit to be less than 1, we need to have at least 2 interior mesh-points. Clearly  $\tilde{S}^{hk}$  is increasing in  $h$  and decreasing in  $k$  and we find  $\tilde{S}^{0.1,0.5} = 0.9351 < 1$ . Hence for  $k \geq 0.5$  and  $h \leq 0.1$  we have  $H_{jn}^{hk} \geq 0$  for  $1 \leq j, n \leq N$ .  $\square$

The above value of  $k_1$  appears pessimistic. Direct calculation to determine  $k_1$  such that  $\mathcal{E}_{hk}$  is positive for  $k \geq k_1$ , for  $h = 1/10, 1/20, 1/30$ , and  $1/40$ , give  $k_1 = 0.018, 0.017, 0.017$ , and  $0.015$ , respectively.

### 6. The lumped mass finite element method

Consider now the lumped mass spatially semidiscrete parabolic problem (1.3), and let  $\bar{E}_h(t) = e^{\bar{\Delta}_h t}$  be its solution operator. where  $\bar{\Delta}_h : S_h \rightarrow S_h$  is the discrete Laplacian defined by

$$-(\bar{\Delta}_h \psi, \chi)_h = (\nabla \psi, \nabla \chi), \quad \forall \psi, \chi \in S_h. \tag{6.1}$$

Thus the problem (1.3) is of the form (1.4) with  $A = -\bar{\Delta}_h$ , in  $\mathcal{B} = S_h$  with norm  $\|\cdot\|_\infty$ .

In matrix form, the problem (1.3) may be written

$$\mathcal{D} \frac{d\alpha}{dt} + \mathcal{S} \alpha = 0, \quad \text{for } t \geq 0, \quad \text{with } \alpha(0) = v,$$

where  $\mathcal{D} = (d_{ij})$  is the diagonal matrix with elements  $d_{ij} = (\Phi_i, \Phi_j)_h$  and  $\mathcal{S}$  is the stiffness matrix. The solution operator matrix is then  $\mathcal{E}_h(t) = e^{-\mathcal{H}t}$  where  $\mathcal{H} = \mathcal{D}^{-1} \mathcal{S}$ , and, as for  $E_h(t)$  in Section 5,  $\bar{E}_h(t)$  is positive or contractive if and only if this holds for  $\bar{\mathcal{E}}_h(t)$ . We know from [8] that  $\bar{E}_h(t)$  is both a positive operator and contraction in maximum-norm for  $t \geq 0$ , provided the triangulation is essentially of Delaunay type, or more precisely, we have the following theorem.

**Theorem 6.1.** *The solution operator matrix  $\bar{E}_h(t)$  of (1.3) is positive for all  $t \geq 0$  if and only if  $s_{ij} \leq 0$  for  $j \neq i$ . It is a contraction for all  $t \geq 0$  if and only if  $\mathcal{S}$  is diagonally dominant.*

By the argument of the proof of Theorem 5.4 we find, even without diagonal dominance, and under a mild assumption on  $\{\mathcal{T}_h\}$ , that  $\|\bar{E}_h(t)\|_\infty \rightarrow 0$  as  $t \rightarrow \infty$ , and hence, in particular, that  $\bar{E}_h(t)$  is a contraction for large  $t$ , independently of  $h$ . In this case, we use the analogue of (5.4) for  $\bar{E}(t)$  from [2], Theorem 4.4.

**Theorem 6.2.** *Assume that the family  $\{\mathcal{T}_h\}$  satisfies (5.3). Then we have*

$$\|\bar{E}_h(t)v_h\|_\infty \leq Ct^{-1}\|v_h\|_\infty, \quad \text{for } t > 0, \quad v_h \in S_h.$$

Theorem 6.1 easily carries over to the backward Euler method:

**Theorem 6.3.** *The backward Euler operator  $\bar{E}_{hk} = (I - k\bar{\Delta}_h)^{-1}$  is positive for all  $k > 0$  if and only if  $s_{ij} \leq 0$  for  $j \neq i$ . It is a contraction for all  $k > 0$  if and only if  $\mathcal{S}$  is diagonally dominant.*

*Proof.* For positivity the sufficiency of the condition follows from Theorem 6.1 and the representation formula (1.6), or, in matrix form,

$$\bar{E}_{hk} = (\mathcal{I} + k\mathcal{H})^{-1} = \int_0^\infty e^{-t} \mathcal{E}_h(kt) dt. \tag{6.2}$$

On the other hand, if  $\bar{E}_{hk}$  is positive for small  $k$ , then  $\bar{\mathcal{E}}_h(t)$  is positive since

$$\bar{\mathcal{E}}_h(t) = e^{-\mathcal{H}t} = \lim_{n \rightarrow \infty} (\mathcal{I} + \frac{t}{n}\mathcal{H})^{-n} = \lim_{n \rightarrow \infty} \bar{\mathcal{E}}_{h,t/n}, \quad \text{for } t \geq 0. \tag{6.3}$$

Hence  $\bar{E}_h(t) \geq 0$ , and thus  $s_{ij} \leq 0$  for  $j \neq i$  by Theorem 6.1.

By Theorem 6.1,  $\bar{E}_h(t)$ , and thus  $\bar{\mathcal{E}}_h(t)$ , is a contraction for  $t \geq 0$  if  $\mathcal{S}$  is diagonally dominant, and by (6.2) this implies that  $\bar{E}_{hk}$  is also a contraction, for  $k > 0$ . Also, if  $\bar{E}_{hk}$  is a contraction, so is  $\bar{\mathcal{E}}_h(t)$  by (6.3), and hence  $\mathcal{S}$  is diagonally dominant by Theorem 6.1.  $\square$

We next consider more general  $A$ -correct rational functions, and single step time discretization operators  $\bar{E}_{hk} = r(-k\bar{\Delta}_h)$ , or in matrix form, with  $\mathcal{D}$  and  $\mathcal{S}$  as above,

$$\bar{E}_{hk} = r(k\mathcal{H}), \quad \text{where } \mathcal{H} = \mathcal{D}^{-1}\mathcal{S}.$$

As in Theorems 5.5 and 6.3 we have the following necessary conditions.

**Theorem 6.4.** *If the time stepping operator  $\bar{E}_{hk} = r(-k\Delta_h)$  is positive for  $k$  small, then  $s_{ij} \leq 0$  for  $j \neq i$ . If  $\bar{E}_{hk}$  is a contraction for  $k$  small, then  $\mathcal{S}$  is diagonally dominant.*

We have the following immediate consequence of Theorems 2.4 and 6.2.

**Theorem 6.5.** *Assume that (5.3) holds. Then the operator  $\bar{E}_{hk} = r_{02}(-k\bar{\Delta}_h)$  is contractive for large  $k$ , uniformly in  $h$ . Also, if  $r(z)$  is  $A$ -correct, then  $r(\infty) \geq 0$  is a necessary condition for  $\bar{E}_{hk}$  to be positive for  $k$  large.*

We now turn to the  $\theta$ -method defined by

$$\bar{\mathcal{E}}_{hk} = (\mathcal{D} + \theta k \mathcal{S})^{-1}(\mathcal{D} - (1 - \theta)k \mathcal{S}) = r_\theta(k\mathcal{H}), \quad \mathcal{H} = \mathcal{D}^{-1}\mathcal{S}, \quad 0 \leq \theta \leq 1, \quad (6.4)$$

The following sufficient conditions are essentially contained in [3]. Note that, except for the backward Euler method ( $\theta = 1$ ),  $k$  is required to be bounded above.

**Theorem 6.6.** *Assume that  $s_{ij} \leq 0$  for  $j \neq i$ , and that*

$$(1 - \theta)k s_{ii} \leq d_{ii}, \quad \text{for all } i, \quad \text{with } 0 \leq \theta \leq 1. \quad (6.5)$$

Then  $\bar{E}_{hk} = r_\theta(-k\bar{\Delta}_h) \geq 0$ . If, in addition,  $\mathcal{S}$  is diagonally dominant, then  $\bar{E}_{hk}$  is also a contraction.

*Proof.* By (6.5),  $\mathcal{D} - (1 - \theta)k\mathcal{S} \geq 0$ . Further,  $\mathcal{D} + \theta k\mathcal{S}$  is a Stieltjes matrix, so that  $(\mathcal{D} + \theta k\mathcal{S})^{-1} \geq 0$ . Thus  $\bar{\mathcal{E}}_{hk} \geq 0$  by (6.4).

For the proof of contractivity, we write  $\beta = \bar{\mathcal{E}}_{hk}\alpha$  as

$$(\mathcal{D} + \theta k \mathcal{S})\beta = (\mathcal{D} - (1 - \theta)k \mathcal{S})\alpha.$$

The equation in this system corresponding to the vertex  $P_i$  is

$$(d_{ii} + \theta k s_{ii})\beta_i + \theta k \sum_{j \neq i} s_{ij}\beta_j = (d_{ii} - (1 - \theta)k s_{ii})\alpha_i - (1 - \theta)k \sum_{j \neq i} s_{ij}\alpha_j.$$

With  $i$  such that  $|\beta_i| = |\beta|_\infty$ , (6.5) and the diagonal dominance of  $\mathcal{S}$  imply that

$$\begin{aligned} (d_{ii} + \theta k s_{ii})|\beta|_\infty &\leq \theta k \sum_{j \neq i} |s_{ij}||\beta|_\infty + (d_{ii} - (1 - \theta)k s_{ii})|\alpha|_\infty \\ &\quad + (1 - \theta)k \sum_{j \neq i} |s_{ij}||\alpha|_\infty \leq \theta k s_{ii}|\beta|_\infty + d_{ii}|\alpha|_\infty. \end{aligned}$$

and hence  $|\beta|_\infty \leq |\alpha|_\infty$ . Thus  $\bar{\mathcal{E}}_{hk}$  is a contraction in  $|\cdot|_\infty$ , and  $\bar{E}_{hk}$  in  $\|\cdot\|_\infty$ .  $\square$

We close with some results for the (0, 2) Padé approximation, again restricted to one space dimension and uniform mesh. The spatially discrete solution operator is now  $\bar{E}_h(t) = e^{\bar{\Delta}_h t}$  where  $\bar{\Delta}_h$  is defined by the one-dimensional version of (6.1), with  $(v, w)_h = h \sum_{j=1}^N v(x_j) w(x_j)$ . The corresponding solution operator matrix is then  $\bar{\mathcal{E}}_h(t) = e^{-\mathcal{H}t}$ , where  $\mathcal{H} = \mathcal{D}^{-1}\mathcal{S}$ , with  $\mathcal{D} = h\mathcal{I}$  and  $\mathcal{S} = h^{-1}(2\mathcal{I} - \mathcal{J})$ , cf. (5.10). We now know that  $\bar{E}_h(t) \geq 0$  and  $\|\bar{E}_h(t)\|_\infty \leq 1$  for  $t \geq 0$ , and also that, as in (5.11), that  $\|\bar{E}_h(t)\|_\infty \leq Ct^{-1/4}$ . To study  $\bar{E}_{hk} = r_{02}(-k\bar{\Delta}_h)$ , we write

$$\bar{\mathcal{E}}_{hk} = r_{02}(k\mathcal{H}) = r_{02}(\lambda\bar{\mathcal{H}}), \quad \text{where } \lambda = k/h^2, \quad \bar{\mathcal{H}} = h^2\mathcal{H} = 2\mathcal{I} - \mathcal{J}.$$

We show the following.

**Theorem 6.7.** *The operator  $\bar{E}_{hk} = r_{02}(-k\bar{\Delta}_h)$  cannot be positive for small  $\lambda$  if  $N \geq 4$ , nor can it be a contraction for small  $\lambda$  if  $N \geq 9$ .*

*Proof.* Since  $|\bar{\mathcal{H}}|_\infty = 4$  we have, by expansion of  $\mathcal{E}_{hk} = r_{02}(\lambda\bar{\mathcal{H}})$ , for small  $\lambda$ ,

$$\bar{\mathcal{E}}_{hk} = \bar{\mathcal{E}}^0 + O(\lambda^5), \quad \text{where } \bar{\mathcal{E}}^0 = \mathcal{I} - \lambda\bar{\mathcal{H}} + \frac{1}{2}\lambda^2\bar{\mathcal{H}}^2 - \frac{1}{4}\lambda^4\bar{\mathcal{H}}^4.$$

By the form of  $\bar{\mathcal{H}}$ , the elements of the fourth bidiagonal of  $\bar{\mathcal{H}}^4$  equal 1 (provided  $N \geq 4$ ), and thus the corresponding elements of  $\bar{\mathcal{E}}_{hk}$  equal  $-\frac{1}{4}\lambda^4 < 0$ . Hence  $\bar{\mathcal{E}}_{hk}$  cannot be a positive operator for  $\lambda$  small.

Turning now to contractivity, we note that if  $5 \leq i \leq N - 4$  (requiring  $N \geq 9$ ), then the  $i$ th row sums of  $\mathcal{H}$ ,  $\mathcal{H}^2$  and  $\mathcal{H}^4$  equal zero, so that  $\sum_j (\bar{\mathcal{E}}_{hk}^0)_{ij} = 1$ . Further,  $(\bar{\mathcal{E}}_{hk}^0)_{i,i\pm 4} = -\frac{1}{4}\lambda^4$ , and hence  $|\bar{\mathcal{E}}_{hk}^0|_\infty \geq \sum_j |(\bar{\mathcal{E}}_{hk}^0)_{ij}| \geq 1 + \lambda^4$ , for such an  $i$ . Hence  $|\bar{\mathcal{E}}_{hk}|_\infty \geq 1 + \frac{1}{2}\lambda^4$  for  $\lambda$  small. This completes the proof.  $\square$

Since  $|\bar{\mathcal{E}}_h(t)|_\infty \rightarrow 0$  as  $t \rightarrow \infty$ , it follows by Theorem 2.4 that  $\bar{\mathcal{E}}_{hk}$  is also contractive for sufficiently large  $k$ .

**Theorem 6.8.** *There is a  $k_0 > 0$ , independent of  $h$ , such that  $\bar{E}_{hk} = r_{02}(-k\bar{\Delta}_h)$  is a contraction for  $k \geq k_0$ .*

Direct calculation of the matrix norm  $|\bar{\mathcal{E}}_{hk}|_\infty$ , for  $h = 1/10, 1/20, 1/30$ , and  $1/40$ , shows that we may choose  $k_0 = 0.007, 0.009, 0.010$ , and  $0.010$ , respectively.

We now show that  $\bar{E}_{hk}$  is positive for  $k$  large.

**Theorem 6.9.** *There is a  $k_1 > 0$ , independent of  $h$ , such that  $\bar{E}_{hk} = r_{02}(-k\bar{\Delta}_h)$  is positive if  $k \geq k_1$ .*

*Proof.* The proof is modelled after that of Theorem 4.6. We first show that, if  $v_h \geq 0$  in  $\Omega$ , then, for any  $j$ ,  $E_h(t)v_h(x_j)$  is decreasing for  $t \geq 0.31$ . We may write

$$\bar{\mathcal{E}}_h(t)v_h(x_j) = \sum_{n=1}^N \bar{G}_{jn}^h(t)v_h(x_n), \quad j = 1, \dots, N,$$

where, since the eigenvectors of  $-\bar{\Delta}_h$  are  $\{\phi_l^h(x_j)\}_{j=1}^N = \{\sqrt{2h} \sin(l\pi x_j)\}_{j=1}^N$ , we have

$$\bar{G}_{jn}^h(t) = 2h \sum_{l=1}^N e^{-\bar{\lambda}_l^h t} \sin(l\pi x_j) \sin(l\pi x_n), \quad \text{with } \bar{\lambda}_l^h = \frac{2(1 - \cos(\pi lh))}{h^2}.$$

By differentiation we obtain

$$\begin{aligned} \frac{\partial}{\partial t} \bar{G}_{jn}^h(t) &= -2h \sum_{l=1}^N \bar{\lambda}_l^h e^{-\bar{\lambda}_l^h t} \sin(l\pi x_j) \sin(l\pi x_n) \\ &= -2h \bar{\lambda}_1^h e^{-\bar{\lambda}_1^h t} \sin(\pi x_j) \sin(\pi x_n) (1 + \tilde{S}_{jn}^h(t)), \end{aligned}$$

and, again using  $|\sin(lx)/\sin x| \leq l$ , we find

$$|\tilde{S}_{jn}^h(t)| = \left| \sum_{l=2}^N \frac{\bar{\lambda}_l^h}{\bar{\lambda}_1^h} e^{-(\bar{\lambda}_l^h - \bar{\lambda}_1^h)t} \frac{\sin(l\pi x_j)}{\sin(\pi x_j)} \frac{\sin(l\pi x_n)}{\sin(\pi x_n)} \right| \leq \sum_{l=2}^N \frac{\bar{\lambda}_l^h}{\bar{\lambda}_1^h} l^2 e^{-(\bar{\lambda}_l^h - \bar{\lambda}_1^h)t}.$$

We note that since  $(2/\pi)^2 \xi^2 \leq 2(1 - \cos \xi) \leq \xi^2$  for  $\xi \in (0, \pi)$ , we have

$$4l^2 \leq \bar{\lambda}_l^h \leq \pi^2 l^2, \quad \text{for } l = 1, \dots, N.$$

Hence, since  $\mu e^{-\mu t}$  is decreasing and  $\mu^{-1}e^{\mu t}$  increasing in  $\mu$ , for  $t \geq 1/\mu$ ,

$$|\tilde{S}_{jn}^h(t)| \leq \frac{4}{\pi^2} \sum_{j=2}^{\infty} l^4 e^{-(4l^2 - \pi^2)t} < 1, \quad \text{for } t \geq 0.31.$$

The remaining part of the proof follows that of Theorem 4.6, again with  $\tau = 2.34$ , giving the result stated with  $k_1 = 0.31/2.34 \approx 0.133$ .  $\square$

As in an earlier case, the  $k_1$  above is pessimistic. Direct calculation to determine  $k_1$  such that  $\bar{\mathcal{E}}_{hk}$  is positive for  $k \geq k_1$ , for  $h = 1/10, 1/20, 1/30$ , and  $1/40$ , gives  $k_1 = 0.014, 0.016, 0.017$ , and  $0.017$ , respectively.

**Acknowledgments.** The work of A.H. Schatz and V. Thomée was partly supported by the U.S. National Science Foundation under Grant DMS 0612599.

## References

1. C. Bolley and M. Crouzeix, *Conservation de la positivité lors de la discrétisation des problèmes d'évolution paraboliques*, RAIRO Anal. numér. **12** (1978), pp. 237–245.
2. P. Chatzipantelidis, R. D. Lazarov, and V. Thomée, *Some error estimates for the lumped mass finite element method for a parabolic problem*, Math. Comp., to appear.
3. 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.
4. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products (corrected and enlarged edition)*, Academic Press, 1980.
5. M. Kovács, *On positivity, shape, and norm-bound preservation of time-stepping methods for semigroups*, J. Math. Anal. Appl. **304** (2005), pp. 115–136.
6. A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Springer-Verlag, New York, 1983.
7. V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, Second edition, Springer-Verlag, Berlin Heidelberg, 2006.
8. V. Thomée and L.B. Wahlbin, *On the existence of maximum principles in parabolic finite element equations*, Math. Comp. **77** (2008), pp. 11–19.