Parabolic evolution problems

1. Formalism. The Gelf'and triple

Let V and H be two Hilbert spaces on $\mathbb{C}$, with $V \subset H$, V dense in H, the canonic injection of V into H being continuous. We note $(\,.\,,.)$ the inner product in H, $|\,.\,|$ its associated norm and $\|\,.\,\|$ the norm in V. By Riesz's Theorem, to each bounded antilinear form on H we can associate a unique element $u \in H$ such that this form is the map

$$v \mapsto (u, v) \qquad \text{from } H \text{ to } \mathbb{C}\,;$$

reciprocally, an element $u \in H$ defines in this way a bounded antilinear map on H. We can thus identify H to its antidual H' (that is, we decide to represent every bounded antilinear map on H by the corresponding element u). The space $H \equiv H'$ is thus identified to a subspace of V', the antidual space to V. We thus have

$$V \subset H \equiv H' \subset V'.$$

Moreover, H is dense and continuously embedded into V' and we can use the same notation for the inner product in H and for the duality between V' and V.

Proof. Let C_1 denote the boundedness constant in the injection of V into H; we thus have $\forall\, v \in V$, $|v| \leq C_1\|v\|$. Given $u \in H$, the map $v \mapsto (u, v)$ is bounded on H and hence on V. Therefore, the antilinear map defined by $\ell_u(v) = (u, v)$ satisfies $\ell_u \in V'$. Also we have

$$\|u\|_{H'} := \sup_{0 \neq v \in H} \frac{|(u,v)|}{|v|} = |u| \quad \text{and} \quad \|\ell_u\|_{V'} := \sup_{0 \neq v \in V} \frac{|(u,v)|}{\|v\|} = C_1|u|.$$

This proves that the map $u \mapsto \ell_u$ (which is clearly linear) is bounded from H to V'. Besides, it is injective. In fact, if we have $\ell_u = 0$, then we have $(u, v) = 0$ for all $v \in V$ and by density of V in H, $(u, v) = 0$ for all $v \in V$. We thus derive that $u = 0$, which proves the injectivity.

We can thus identify u and ℓ_u, that is, identify H with a subspace of V' and keep the same notation $(\,.\,,.)$ for the inner product and the antiduality V', V. More precisely, this means that:

> when $\ell \in V'$ and $v \in V$, the quantity (ℓ, v) represents the antilinear map ℓ applied to the element v,
> when $\ell \in H$ and $v \in H$, the quantity (ℓ, v) represents the inner product in H of the elements ℓ and v.

Let us remark that for $\ell \in H$ and $v \in V$ the quantity (ℓ, v) is defined in two different ways, but that both definitions coincide.

It remains to prove that H' is dense in V'. Indeed, otherwise, there would exist a linear map L, bounded on V', vanishing on H' and not identically zero on V'. Since V' is a Hilbert space, we have that $(V')^* = V$, that is, there exists $v \in V$ such that $\forall\, w \in V'$, $L(w) = (w, v)$. We then have $(w, v) = L(w) = 0$ for all $w \in H$. Taking $w = v$, we obtain that $v = 0$. Hence $L = 0$, which contradicts the starting point. □

Let now $a(.\,,.) : V \times V \to \mathbb{C}$ be a bounded sesquilinear form. We can associate to it the operator $A \in \mathcal{L}(V, V')$ defined by

$$\forall\, v, w \in V, \qquad (Av, w) = a(v, w).$$

The continuity means $\forall\, v, w \in V$, $|(Av, w)| \leq M\,\|v\|\,\|w\|$, or equivalently $\|Av\|_{V'} \leq M\|v\|$, or also $\|A\|_{\mathcal{L}(V,V')} \leq M$. Let us recall that the adjoint operator $A^* \in \mathcal{L}(V, V')$ and the adjoint form a^* are defined by

$$(A^* v, w) = a^*(v, w) = \overline{a(w, v)}.$$

We now introduce the space

$$D(A) = \{v \in V\,; Av \in H\}.$$

It is easy to see that, endowed with the norm $\|v\|_{D(A)} = (\|v\|^2 + |Av|^2)^{1/2}$, this is a Hilbert space. Besides, if A is an isomorphism from V to V' (which happens for instance when the sesquilinear form $a(.\,,.)$ is V-elliptic, by Lax-Milgram's Theorem), then A is an isomorphism from $D(A)$ to H. From the density of H in V', we derive that of $D(A) = A^{-1}H$ in $V = A^{-1}V'$. Since V is dense in H, we have also that $D(A)$ is dense in H.

We can recurrently define

$$D(A^{k+1}) = \{v \in V\,; Av \in D(A^k)\}.$$

Example 1. Let Ω be a nonempty bounded open set in $\mathbb{R}^2$,

$$V = H_0^1(\Omega),\ H = L^2(\Omega), \quad (u, v) = \int_\Omega u\,\overline{v}\,dx,\ a(u, v) = \int_\Omega \nabla u\,\nabla\overline{v}\,dx, \quad \|v\| = (a(v, v))^{1/2}.$$

We are in the situation of V dense in H, since $\mathcal{D}(\Omega)$, which is contained in V, is dense in H. The space V', defined by identifying H to its antidual space, is denoted $H^{-1}(\Omega)$. We have here $D(A) = \{v \in H_0^1(\Omega); \Delta v \in L^2(\Omega)\}$ and $Av = -\Delta v$. If Ω is convex (or if the boundary of Ω is of class C^2), by a regularity theorem we deduce that $D(A) = H^2(\Omega) \cap H_0^1(\Omega)$. When the boundary of Ω is C^∞, we can also prove that $D(A^k) = \{v \in H^{2k}(\Omega); v, \Delta v, \ldots \text{and } \Delta^{k-1}v \in H_0^1(\Omega)\}$.

Example 2. We now take

$$V = H^1(\Omega),\ H = L^2(\Omega), \quad (u, v) = \int_\Omega u\,\overline{v}\,dx,\ a(u, v) = \int_\Omega (\nabla u\,\nabla\overline{v} + u\,\overline{v})\,dx, \quad \|v\| = (a(v, v))^{1/2}.$$

We are thus in the situation of V being dense in H. If Ω is convex (or if the boundary of Ω is of class C^2), by a regularity theorem we can prove that $D(A) = \{v \in H^2(\Omega)\,; \frac{\partial v}{\partial n} = 0 \text{ sur } \partial\Omega\}$ and then $Av = -\Delta v + v$. When the boundary of Ω is C^∞, we can also prove that $D(A^k) = \{v \in H^{2k}(\Omega); \frac{\partial v}{\partial n}, \frac{\partial \Delta v}{\partial n}, \ldots \text{and } \frac{\partial \Delta^{k-1} v}{\partial n} = 0 \text{ on } \partial\Omega\}$.

Remark: In the case of Example 1, we have another identification

$$\mathcal{D}(\Omega) \subset H_0^1(\Omega) \subset L^2(\Omega) \subset H^{-1}(\Omega) \subset \mathcal{D}'(\Omega),$$

since the density of $\mathcal{D}(\Omega)$ in $H_0^1(\Omega)$ allows to identify $H^{-1}(\Omega)$ to a subspace of the space of distributions on Ω. This identification is not possible in Example 2. Indeed, for instance the antilinear map $\ell(v) = \int_{\partial\Omega} \overline{v}(x)\, d\sigma(x)$ belongs to $(H^1(\Omega))'$, and does not vanish identically. However, ℓ is clearly null on $\mathcal{D}(\Omega)$, and cannot therefore be identified to an element of $\mathcal{D}'(\Omega)$.

Sectoriality. Let us now assume that the sesquilinear form a is V-elliptic, that is, there exists $a_0 > 0$ such that $\forall\, v \in V$, we have $\operatorname{Re} a(v,v) \geq a_0 \|v\|^2$. By continuity we also have $|a(v,v)| \leq M \|v\|^2$. Writing $a(v,v) = \rho(v) e^{i\gamma(v)} \|v\|^2$, we thus have $a_0 \leq \rho \cos\gamma$ and $\rho \leq M$. With $\alpha = \operatorname{Arccos}(a_0/M)$, we have $0 \leq \alpha < \pi/2$ and the sectoriality property

$$\forall\, v \in V, \quad a(v,v) \in S_\alpha \quad \text{where} \quad S_\alpha := \{z \in \mathbb{C}\,;\, z = 0 \text{ or } |\operatorname{Arg} z| \leq \alpha\}. \tag{1}$$

Let us also remark that for all $z \notin S_\alpha$, the operator $zI - A$ is an isomorphism from V to V'. To see that, let us write $z = |z| e^{i\beta}$ and take

$B = A - zI$ if $\operatorname{Re} z \leq 0$,
$B = ie^{-i\beta}(A - zI)$ if $\beta \in (\alpha, \frac{\pi}{2})$,
$B = -ie^{-i\beta}(A - zI)$ if $\beta \in (-\frac{\pi}{2}, -\alpha)$.

It is sufficient to prove that $(B.\,,.)$ is V-elliptic, which implies by Lax-Milgram's Theorem that B (and consequently $zI - A$) is an isomorphism from V to V'.

The result follows readily when $\operatorname{Re} z \leq 0$, since we then have $\operatorname{Re}(Bv, v) \geq \operatorname{Re} a(v,v) \geq a_0 \|v\|^2$.

In case $\beta \in (-\frac{\pi}{2}, -\alpha)$, we have $\operatorname{Re}(Bv, v) = \operatorname{Im}(e^{-i\beta} a(v,v)) = \rho(v) \sin(\gamma(v) - \beta) \|v\|^2$ (with the notations above) and hence $\operatorname{Re}(Bv, v) \geq a_0 \sin(|\beta| - \alpha) \|v\|^2$. We can likewise prove this bound when $\beta \in (\alpha, \frac{\pi}{2})$.

From what precedes it follows that for all $z \notin S_\alpha$, the operator $zI - A$ is also an isomorphism from $D(A)$ onto H.

2. Resolvent estimates

We now consider a linear operator A, defined on a subspace $D(A) \subset H$, with values on H, and we suppose that $D(A)$ is dense in H. Such an operator is said to be unbounded. Given $\alpha \in [0, \frac{\pi}{2}]$, we associate to it the sector $S_\alpha = \{z \in \mathbb{C}\,;\, z = 0 \text{ or } |\operatorname{Arg} z| \leq \alpha\}$.

Definition. We say that the operator A is mα-accretive if

$$\begin{gathered} \forall\, v \in D(A), \qquad (Av, v) \in S_\alpha. \\ \text{and if } \forall z \notin S_\alpha, \qquad zI - A \text{ is an isomorphism from } D(A) \text{ to } H. \end{gathered} \tag{2}$$

When $\alpha = \frac{\pi}{2}$ we say that the operator is m-accretive. The operators A studied in the previous section are examples of mα-accretive operators. The operator $R(z) = (zI - A)^{-1}$ is called the resolvent of A at the point z.

Theorem 1. *Let A be an mα-accretive operator. Then for all $z \notin S_\alpha$ we have the estimates*

$$\|(zI - A)^{-1}\|_{H \to H} \leq \frac{1}{d(z, S_\alpha)}, \qquad \|A(zI - A)^{-1}\|_{H \to H} \leq \frac{|z|}{d(z, S_\alpha)}.$$

Moreover, the maps $z \mapsto (zI - A)^{-1}$ from S_α^c (complement of S_α) to $\mathcal{L}(H, H)$ and $\mathcal{L}(H, D(A))$ are continuous and infinitely differentiable (in the sense of $\mathbb{C}$).

Conversely, if we assume that $\forall\, z \notin S_\alpha$ $zI - A$ is an isomorphism from $D(A)$ to H and that, for all $z \neq 0$ with $|\arg z| = \alpha + \frac{\pi}{2}$, $\|(zI - A)^{-1}\|_{H \to H} \leq \frac{1}{|z|}$, then A is mα-accretive.

Proof. Let v and f be related by $(zI - A)v = f$.

a) We have $(Av, v) \in S_\alpha$. Thus $d(z, S_\alpha)\, |v|^2 \leq |\, z\, |v|^2 - (Av, v)| = |(f, v)| \leq |f|\, |v|$. Therefore, we have $|v| \leq |f|/d(z, S_\alpha)$, from where the first estimate follows.

b) We also have

$$d(z, S_\alpha)\, |Av|^2 \leq |\, z\, |Av|^2 - |z|^2 (Av, v)| = |z|\, |(Av, Av - z\, v)| = |z|\, |(Av, f)| \leq |z|\, |Av|\, |f|.$$

From this, we deduce that $|Av| \leq |z|\, |f|/d(z, S_\alpha)$, and therefore the second estimate holds.

c) We notice that

$$(z_1 I - A)^{-1} - (z_2 I - A)^{-1} = (z_2 I - A)^{-1}(z_2 - z_1)(z_1 I - A)^{-1},$$

and hence

$$\|R(z_1) - R(z_2)\|_{H \to H} \leq \frac{|z_2 - z_1|}{d(z_1, S_\alpha) d(z_2, S_\alpha)}, \text{ and } \|A(R(z_1) - R(z_2))\|_{H \to H} \leq \frac{|z_2||z_2 - z_1|}{d(z_1, S_\alpha) d(z_2, S_\alpha)},$$

which proves the continuity results.

d) We start from the expression $\frac{1}{z_1 - z_2}(R(z_1) - R(z_2)) = -(z_2 I - A)^{-1}(z_1 I - A)^{-1}$. Taking the limit as $z_2 - z_1 \to 0$, we deduce that R is differentiable and $R'(z_1) = -R(z_1)^2$.

e) Converse part. We now have for all $u \in D(A)$ and $z = |z|\, i\, e^{i\varepsilon\alpha}$, $\varepsilon = \pm 1$,

$$|(zI - A)\, u|^2 \geq |z|^2\, |u|^2,$$

$$\text{i.e.} \qquad |z|^2\, |u|^2 - 2\,\mathrm{Re}\, \bar{z}(Au, u) + |Au|^2 \geq |z|^2\, |u|^2.$$

Dividing by $|z|$

$$2\ \mathrm{Re}\left(e^{i(\frac{\pi}{2} - \varepsilon\alpha)}(Au, u)\right) + \frac{1}{|z|}\, |Au|^2 \geq 0,$$

from where, making $|z|$ tend to infinity, $\mathrm{Re}\left(e^{i(\frac{\pi}{2} - \varepsilon\alpha)}(Au, u)\right) \geq 0$, that is, $(Au, u) \in S_\alpha$. □

3. Functions of operators. Analytic semigroups

Let A be an mα-accretive operator on H, and let us consider a rational fraction r bounded on S_α. We can therefore write r as

$$r(z) = r(\infty) + \sum_j \frac{r_j}{(\alpha_j - z)^{m_j}},$$

with $\alpha_j \notin S_\alpha$, $m_j \in \mathbb{N}^*$. It is then natural to define the operator $r(A)$ by

$$r(A) = r(\infty) I + \sum_j\ r_j((\alpha_j I - A)^{-1})^{m_j}.$$

We have then

- $r(A) \in \mathcal{L}(H,H)$, $\qquad (r(A) \in \mathcal{L}(H,D(A))$, if $r(\infty)=0)$.
- $r(A)+s(A) = (r+s)(A)$, and $r(A)s(A) = (rs)(A)$, if r and s are two rational functions bounded on S_α.

We admit the following theorem:

Theorem 2. *Let $\alpha \in [0, \frac{\pi}{2}]$. There exists a constant $1 \le C_\alpha \le 2+\frac{2}{\sqrt{3}}$ such that for all $m\alpha$-accretive operator A and for all rational fraction r bounded on the sector S_α, we have*

$$\|r(A)\|_{H\to H} \le C_\alpha \sup_{z\in S_\alpha} |r(z)|.$$

Moreover, when $\alpha = \frac{\pi}{2}$, we have $C_{\pi/2} = 1$.

Corollary 3. *When the function f is a uniform limit of rational fractions r_n on S_α, the relation $f(A) = \lim_{n\to\infty} r_n(A)$ defines an operator $f(A) \in \mathcal{L}(H,H)$ and it follows that*

$$\|f(A)\|_{H\to H} \le C_\alpha \sup_{z\in S_\alpha} |f(z)|.$$

Proof. From the previous theorem we have

$$\|r_n(A) - r_p(A)\|_{H\to H} \le C_\alpha \|r_n - r_p\|_{L^\infty(S_\alpha)} \to 0 \quad \text{as } \ n \text{ and } p \to \infty.$$

The sequence $r_n(A)$ is a Cauchy sequence and therefore convergent in the complete space $\mathcal{L}(H,H)$. The corollary follows readily from this (it is clear, in particular, that $f(A)$ does not depend on the choice of the sequence of r_n). ☐

The following properties can be readily verified

$$f(A)+g(A) = (f+g)(A), \qquad f(A)g(A) = (fg)(A) = g(A)f(A). \tag{3}$$

Lemma 4. *For all $\alpha \in [0, \frac{\pi}{2})$, we have*

$$\forall\, n \ge 1, \qquad \sup_{z\in S_\alpha} |e^{-z} - \big(\frac{1}{1+z/n}\big)^n| \le \frac{6}{n\cos^2\alpha}.$$

Proof. Take $\varphi_n(z) = e^{-z} - \big(\frac{1}{1+z/n}\big)^n$.

Clearly $|\varphi_n(z)| \le 2$, which proves the lemma for $n = 1,2,3$.

Remark 1. It holds $|e^{-u} - \frac{1}{1+u}| \le \frac{3}{2}|u|^2$, for all u with $\operatorname{Re} u \ge 0$.

In fact we have (Taylor) $e^{-u} = 1 - u + \int_0^1 (1-s)e^{-su}\,ds\; u^2$ and $\frac{1}{1+u} = 1 - u + \frac{u^2}{1+u}$, from where

$$|e^{-u} - \tfrac{1}{1+u}| = |\int_0^1 (1-s)e^{-su}\,ds - \tfrac{1}{1+u}|\,|u|^2 \le \tfrac{3}{2}\,|u|^2.$$

Remark 2. It holds $\forall\, x \in \mathbb{R}$, $1 + x \le e^x$.

We write

$$\varphi_n(z) = (e^{-z/n} - \frac{1}{1+z/n}) \sum_{k=0}^{n-1} e^{-kz/n}(\frac{1}{1+z/n})^{n-k-1}.$$

From the two remarks, writing $\rho = \operatorname{Re} z$, (therefore $\rho \geq |z| \cos \alpha$), it follows that

$$|\varphi_n(z)| \leq \frac{3}{2}\, \frac{|z|^2}{n^2}\, n \left(\frac{1}{1+\operatorname{Re} z/n}\right)^{n-1} \leq \frac{3}{2\,n}\, \frac{1}{\cos^2\alpha}\, \frac{\rho^2}{(1+\rho/n)^{n-1}}.$$

The right hand side reaches its maximum with $\rho = \frac{2n}{n-3}$, and thus

$$|\varphi_n(z)| \leq \frac{6}{n\cos^2\alpha}\, \frac{n^2\,(n-3)^{n-3}}{(n-1)^{n-1}} \leq \frac{6}{n\cos^2\alpha}.$$

□

Corollary 5. *Let α and β satisfy $0 \leq \alpha < \alpha+\beta < \frac{\pi}{2}$. Then for all $t \in S_\beta$ the function $E(t) = \exp(-tA)$ is well defined. Moreover $E(t) \in \mathcal{L}(H,H)$ and $\|E(t)\|_{H\to H} \leq 1$.*

Proof. Let us take $f(z) = e^{-tz}$ and $r_n(z) = (1+\frac{tz}{n})^{-n}$. By the previous lemma, we have $\|f - r_n\|_{L^\infty(S_\alpha)} \leq 6/(n\cos^2(\alpha+\beta))$, which permits to apply corollary **??** and to obtain the bound $\|E(t)\|_{H\to H} \leq C_\alpha$. A better bound is obtained by noticing that $\frac{t}{n}A$ is m$(\alpha+\beta)$-accretive, and thus by theorem **??**, $\|(I+\frac{t}{n}A)^{-1}\|_{H\to H} \leq 1/d(-1, S_{\alpha+\beta}) = 1$. Therefore $\|r_n(A)\|_{H\to H} \leq 1$ and by taking the limit $\|E(t)\|_{H\to H} \leq 1$. □

Remark. This corollary is valid in particular with $t = 0$ and we have $E(0) = I$.

Theorem 6. *Let $\alpha \in [0, \frac{\pi}{2})$. The family of operators $E(t)$, $t \geq 0$, satisfies the following properties*

- $\forall\, t, s \geq 0, \qquad E(t+s) = E(t)\, E(s),$
- $\forall\, t \geq 0, \qquad \|E(t)\|_{H\to H} \leq 1,$
- $\forall\, u_0 \in H,$ *the map $t \mapsto E(t)\, u_0$ is continuous from $\mathbb{R}^+$ to H.*

We say that this family is a semigroup of contractions strongly continuous on H and that the operator A is the infinitesimal generator of this semigroup.

Proof. a) The first property is called the semigroup property (we would say group if it were valid $\forall t, s \in \mathbb{R}$). This property follows from the relation $e^{-(t+s)z} = e^{-tz}e^{-sz}$, see (**??**). The second property, seen in Corollary **??**, justifies the term contraction.

b) Let us now prove the third property, which is referred to the strong continuity of E in H. Let $u_0 \in H$ be given. For $0 \leq s < t$, we have $|E(t)\, u_0 - E(s)\, u_0| = |E(s)(E(t-s) - E(0))\, u_0| \leq |(E(t-s) - E(0))\, u_0|$. Therefore, it suffices to prove the continuity in 0.

1st case. If $u_0 \in D(A)$. We then take $f = (I+A)u_0$, and thus $u_0 = (I+A)^{-1} f$. Consequently, using corollary **??**

$$|(E(t) - E(0))u_0| = |(E(t) - E(0))(I+A)^{-1} f| \leq C_\alpha \sup_{z\in S_\alpha} \Big|\frac{e^{-tz}-1}{1+z}\Big|\, |f|.$$

$$\text{For } |z| \geq 1/\sqrt{t}, \text{ we have } \Big|\frac{e^{-tz}-1}{1+z}\Big| \leq \frac{2}{|z|} \leq 2\sqrt{t},$$

$$\text{for } |z| \leq 1/\sqrt{t},\ \Big|\frac{e^{-tz}-1}{1+z}\Big| \leq |e^{-tz}-1| \leq e^{\sqrt{t}} - 1.$$

This yields $\quad \lim_{t\to 0} |(E(t)-E(0))u_0| = 0$.

2nd case. If $u_0 \in H$. Using the contraction property, we obtain for all $v_0 \in D(A)$

$$|(E(t)-E(0))u_0| \le |(E(t)-E(0))v_0| + 2\,|u_0 - v_0|.$$

Continuity in 0 follows from the previous case together with the density of $D(A)$ in H. □

Remark. The theorem is still valid for $\alpha = \frac{\pi}{2}$ (some small changes are needed in order to define $E(t)$ in this case). It is also valid for all t and $s \in S_{\frac{\pi}{2}-\alpha}$.

Theorem 7. *For all $t \in S_\beta$ with $t \neq 0$ and $0 \le \alpha < \alpha + \beta < \frac{\pi}{2}$, and for all integer $k \ge 0$, the operator $E(t) \in \mathcal{L}(H, D(A^k))$. Furthermore, the map $t \mapsto E(t)$ from S_β to $\mathcal{L}(H,H)$ is differentiable (holomorphic), and we have the estimate*

$$\forall\, t > 0, \qquad \|E^{(k)}(t)\|_{H\to H} = \|A^k E(t)\|_{H\to H} \le k!\, \frac{1}{t^k \cos^k \alpha}.$$

Proof. a) Let $f_k(z) = (1+z)^k\, e^{-tz}$. Proceeding as in lemma **??**, f_k is the uniform limit in S_α of $(1+z)^k (1+tz/n)^{-n}$. We can thus define $f_k(A) \in \mathcal{L}(H,H)$. From the relation $e^{-tz} = (1+z)^{-k} f_k(z))$ and from $(I+A)^{-k} \in \mathcal{L}(H, D(A^k))$, we deduce that

$$E(t) = (I+A)^{-k}\, f_k(A) \in \mathcal{L}(H, D(A^k)).$$

b) Derivability. Let $K_2 = \sup\{|\zeta^2 e^{-\zeta}|\,;\zeta \in S_{\alpha+\beta}\}$. For non vanishing t and $s \in S_\beta$ we have by Taylor's formula

$$\forall\, z \in S_\alpha, \qquad |e^{-tz} - e^{-sz} + (t-s)\, z\, e^{-sz}| \le \frac{K_2}{2} \max_{0\le x\le 1} \frac{|t-s|^2}{|xt + (1-x)s|^2}.$$

Using corollary **??** we obtain that

$$\|E(t) - E(s) + (t-s) A\, E(s)\|_{H\to H} \le C_\alpha \frac{K_2}{2} \max_{0\le x\le 1} \frac{|t-s|^2}{|xt + (1-x)s|^2},$$

from where derivability of $E(.)$ follows, as well as the formula $E'(t) = -A\, E(t)$. By induction we also obtain $E^{(k)}(t) = (-A)^k\, E(t)$.

c) Let $t > 0$; we take $s(\theta) = t + r\, e^{i\theta}$ with $r = t \sin\beta$, $\theta \in [0, 2\pi)$. We thus have $s(\theta) \in S_\beta$. Let now u and v be arbitrary elements of H, and let $\varphi(z) = (E(z)u, v)$. From the previous paragraph it follows that the function φ is holomorphic in S_β and also $|\varphi(z)| \le \|E(z)\|_{H\to H} |u|\,|v| \le |u|\,|v|$. From the residue's theorem it follows that

$$\frac{1}{k!}\, \varphi^{(k)}(t) = \frac{1}{2\pi i} \int_0^{2\pi} \frac{\varphi(s(\theta))}{(s(\theta))^{k+1}}\, ds(\theta).$$

Therefore

$$|\varphi^{(k)}(t)| \le \frac{k!}{2\pi} \int_0^{2\pi} \frac{|u|\,|v|}{r^k}\, d\theta = \frac{k!\,|u|\,|v|}{(t \sin\beta)^k}.$$

Since on the other hand $\varphi^{(k)}(t) = (E^{(k)}(t)u, v)$ and the bound holds for all u and $v \in H$, we obtain

$$\|E^{(k)}(t)\|_{H\to H} \le \frac{k!}{(t \sin\beta)^k},$$

from where the lemma follows by taking the limit of β to $\frac{\pi}{2} - \alpha$. □

Glossary. The differentiability property in S_β is stated: the semigroup $E(.)$ is analytic (also holomorphic).

Corollary 8. *For $u_0 \in H$, the following problem*

$$(P) \qquad \begin{cases} \text{find } u \in C^1((0,\infty); D(A)) \cap C^0([0,\infty); H) \text{ such that} \\ \forall\, t > 0, \qquad u'(t) + A\,u(t) = 0, \\ \qquad\qquad\qquad u(0) = u_0, \end{cases}$$

has a unique solution, given by $u(t) = E(t)\,u_0$. Moreover, for all integer $k \geq 0$, $u \in C^\infty((0,\infty); D(A^k))$. If, in addition, $u_0 \in D(A^\ell)$, then $u \in C^0([0,\infty); D(A^\ell)) \cap C^\ell([0,\infty); H)$ and

$$\forall\, k, \ell \geq 0,\ \forall\, t > 0, \qquad |u^{(k+\ell)}(t)| = |A^k u^{(\ell)}(t)| \leq \frac{k!}{(t\cos\alpha)^k}\ |A^\ell\, u_0|. \tag{4}$$

Proof. It is clear that $u(t)$ is a solution to (P) since $E'(t) + A\,E(t) = 0$. The C^∞ regularity on $(0,\infty)$ follows from the previous theorem. The continuity at 0 in H is a consequence of the strong continuity of the semigroup. The problem is linear, so to prove uniqueness of solution to (P) it suffices to show that $u_0 = 0$ implies $u(t) = 0$. In fact, if u is a solution of (P) and $u_0 = 0$, we have

$$\frac{d}{dt}\,|u(t)|^2 = (u'(t), u(t)) + (u(t), u'(t)) = 2\,\mathrm{Re}(u'(t), u(t)) = -2\,\mathrm{Re}(Au(t), u(t)) \leq 0.$$

From here we deduce that $|u(t)|^2 \leq |u(0)|^2 = 0$, and therefore we get the uniqueness.

If now $u_0 \in D(A)$, it is easily verified that $(I + \frac{t}{n}A)^{-1}Au_0 = A(I + \frac{t}{n}A)^{-1}u_0$ and by induction $(I+\frac{t}{n}A)^{-n}Au_0 = A(I+\frac{t}{n}A)^{-n}u_0$. By taking the limit (let us recall that $E(t) = \lim_{n\to\infty}(I+\frac{t}{n}A)^{-n}$) it follows that $AE(t)u_0 = E(t)A\,u_0$. Likewise, if $u_0 \in D(A^\ell)$ we have $A^\ell E(t)u_0 = E(t)A^\ell\,u_0$. We remark that $u^{(\ell)}$ is the solution of (P) with u_0 substituted by $(-A)^\ell\,u_0$, from where $u^{(\ell)} \in C^0([0,\infty); H)$. The bound (**??**) is thus a consequence of theorem **??**. □

Remark. Smoothing effects and irreversibility. We have just seen that the solution of (P) has the following smoothing effect. Admitting the initial value to belong only to H, then for all $t > 0$ and all integer k, we have $u(t) \in D(A^k)$ (spaces more regular than H; see examples 1 and 2 in the first paragraph). It follows then that for the backwards problem

$$(P^r) \qquad \begin{cases} \forall\, 0 \leq t < T, \qquad u'(t) + A\,u(t) = 0, \\ \qquad\qquad\qquad u(T) = v_T, \end{cases}$$

to have a solution, we should have $v_T \in D(A^\infty) := \cap_k D(A^k)$. In fact, even in this case, we would be able to construct examples where there is no solution on the interval $(0, T)$. We say that equation (P) is irreversible.

When the operator A proceeds from a sesquilinear elliptic form a such as the one described in paragraph 1, the problem (P) can be written in the following equivalent form

$$(P) \qquad \begin{cases} \text{find } \ u \in C^1((0,\infty); V) \cap C^0([0,\infty); H) \text{ such that} \\ \forall\, t > 0, \forall\, v \in V, \quad (u'(t), v) + a(u(t), v) = 0, \\ \qquad\qquad\qquad u(0) = u_0. \end{cases}$$

4. Example : the heat equation. Convection-diffusion

Let Ω be a bounded open set in $\mathbb{R}^2$. We consider the spaces $H = L^2(\Omega)$ and $V = H^1_0(\Omega)$ or also $V = H^1(\Omega)$. We take

$$(u, v) = \int_\Omega \rho(x)\, u(x)\, \bar{v}(x)\, dx,$$

$$a(u, v) = \int_\Omega \big(a(x)\nabla u \nabla \bar{v} + \vec{b(x)}.\vec{\nabla} u\, \bar{v} + c(x)\, u\, \bar{v}\big)\, dx$$

We make the following hypotheses: ρ and $c \in L^\infty(\Omega\,;\mathbb{R})$, $a \in C^1(\overline{\Omega}\,;\mathbb{R})$, $\vec{b} \in C^1(\overline{\Omega}\,;\mathbb{R}^2)$, $\operatorname{div}\vec{b} = 0$, in Ω, and

$$\forall\, x \in \overline{\Omega}, \quad 0 < \rho_0 \le \rho(x) \le \rho_M, \quad 0 < a_0 \le a(x) \le a_M, \quad 0 \le c_0 \le c(x) \le c_M.$$

In the case $V = H^1(\Omega)$, we will also assume that $c_0 > 0$ and $\vec{b}.\vec{n} = 0$ on the boundary $\partial\Omega$. It is clear that $(.\,,.)$ defines on $L^2(\Omega)$ an inner product equivalent to the usual one and that the sesquilinear form $a(.\,,.)$ is bounded on $V \times V$. Besides, integrating by parts, we have (if $v \in V$)

$$\int_\Omega \vec{b}.\vec{\nabla} v\, \bar{v}\, dx = \int_\Omega \operatorname{div}(v\vec{b})\, \bar{v}\, dx = -\int_\Omega \vec{b}.\vec{\nabla}\bar{v}\, v\, dx,$$

and hence each term of these equations is purely imaginary. Therefore

$$\forall\, v \in V, \qquad \operatorname{Re} a(v, v) \ge a_0\, |v|^2_{1,\Omega} + c_0\, |v|^2_{0,\Omega},$$

which proves the ellipticity of a.

According to the theory developed in the previous paragraphs, we have a mα-accretive operator A, and the problem

$$(P) \qquad \begin{cases} \text{find } u \in C^1((0,\infty); V) \cap C^0([0,\infty); H) \text{ such that} \\ \forall\, t > 0, \forall\, v \in V, \qquad (u'(t), v) + a(u(t), v) = 0, \\ \qquad\qquad\qquad\qquad u(0) = u_0, \end{cases}$$

has a unique solution. Formally, (P) can be written in terms of a (so called parabolic) partial differential equation problem: if $V = H^1_0(\Omega)$ (resp. if $V = H^1(\Omega)$),

$$(P) \qquad \begin{cases} u \in C^1((0,\infty); V) \cap C^0([0,\infty); H) \\ \rho(x)\, \frac{\partial u}{\partial t} - \operatorname{div}(a\vec{\nabla} u) + \vec{b}.\vec{\nabla} u + c\, u = 0, & \text{for } t > 0,\ x \in \Omega, \\ u(t, x) = 0, \quad (\text{resp. } \frac{\partial u}{\partial n} = 0) & \text{for } t > 0,\ x \in \partial\Omega, \\ u(0, x) = u_0(x), & \text{for } x \in \Omega. \end{cases}$$

Remark. If the initial value u_0 is real valued, it is clear that the solution u is also real valued. Indeed u and $\bar{u}$ are both solutions and solution is unique.

On the other hand, corollary **??** provides smoothness results in time $u \in C^\infty((0,\infty); H) \cap C^\ell([0,\infty); H)$ whenever $u_0 \in D(A^\ell)$, $\ell \ge 0$. Under this same hypothesis, we also have smoothness in space: $u \in C^0([0,\infty); D(A^\ell)) \cap C^{\ell-k}([0,\infty); D(A^k))$. In order to explicit these results, we can use the regularity theorems for elliptic problems. They give

- Under the regularity assumptions above on the coefficients a, ρ, $\vec{b}$, c and if, either Ω is convex, or its boundary is of class C^2, then for $V = H_0^1(\Omega)$ (resp. $V = H^1(\Omega)$)

$$D(A) = \{v \in H^2(\Omega)\,;\, v = 0 \text{ on } \partial\Omega\}, \qquad (\text{resp. } D(A) = \{v \in H^2(\Omega)\,;\, \frac{\partial v}{\partial n} = 0 \text{ on } \partial\Omega\}),$$

and if $v \in D(A)$, $\qquad Av = \dfrac{1}{\rho}\,(-\text{div}\,(a\vec{\nabla} v) + \vec{b}.\vec{\nabla} v + c\,v).$

- Furthermore if we assume that the boundary of Ω, as well as the coefficients a, ρ, $\vec{b}$, c, are of class C^∞, then, for all integer $k \geq 1$, we have

$$\begin{aligned} D(A^k) &= \{v \in H^{2k}(\Omega) \cap H_0^1(\Omega)\,;\, v = Av = \cdots = A^{k-1}v = 0 \text{ on } \partial\Omega\}, && \text{if } V = H_0^1(\Omega), \\ D(A^k) &= \{v \in H^{2k}(\Omega)\,;\, \tfrac{\partial v}{\partial n} = \tfrac{\partial Av}{\partial n} = \cdots = \tfrac{\partial A^{k-1}v}{\partial n} = 0 \text{ on } \partial\Omega\}), && \text{if } V = H^1(\Omega). \end{aligned}$$

5. Time approximation: semi-discretization in time

We now turn our attention to the approximation of the problem. In order to get a finite dimensional problem, it is necessary to simultaneously approximate in the space variables (which will be done here by a finite element method) and in the time variable (for which techniques of differential equations are preferred). Nevertheless, each of these discretizations carries out very different difficulties and before examining the full discretization (the only realistic one) it is pedagogically useful to examine time and space semi-discretizations separately. We begin by the time discretization of our problem (P)

$$(P) \qquad \begin{cases} u'(t) + A\,u(t) = 0, & \forall\, t > 0, \\ u(0) = u_0, \end{cases}$$

which is nothing but a differential equation in infinite dimension. We choose a time step $\Delta t > 0$ (which will tend to zero) and consider the time steps $t_n = n\Delta t$. (For the sake of simplicity we have chosen a fixed time step, but our approach could be extended also to variable time steps, with some additional technical difficulties). We denote $u(t_n)$ the exact solution of problem (P) taken at time t_n, and u^n the approximated solution obtained by one of the schemes below.

Although not necessary, we initially choose $u^0 = u_0$. We then obtain u^{n+1} departing from u^n by using the formulas

- Scheme 1. $\dfrac{u^{n+1} - u^n}{\Delta t} + A\,u^n = 0,$ (explicit) Euler method,
- Scheme 2. $\dfrac{u^{n+1} - u^n}{\Delta t} + A\,u^{n+1} = 0,$ implicit Euler method,
- Scheme 3. $\dfrac{u^{n+1} - u^n}{\Delta t} + A\,\dfrac{u^n + u^{n+1}}{2} = 0,$ Crank-Nicolson method,
- Scheme 4. $u^{n+1} = r(\Delta t\,A)\,u^n,$ Runge-Kutta schemes,

(for this last scheme, r is a rational fraction). Let us remark that scheme 4 includes the first three ones: it is enough to take $r(z) = 1 - z$ to obtain Euler's scheme, $r(z) = (1+z)^{-1}$ for the implicit Euler; the Crank-Nicolson scheme corresponds to $r(z) = \frac{1-z/2}{1+z/2}$. Scheme 2 is said to be implicit because to obtain u^{n+1} from u^n we need to solve a linear system. Scheme 3 is also implicit.

Study of the forward Euler scheme. In order to compute u^n, it is necessary that $u_0 \in D(A^n)$, and then $u^n = (I - \Delta t\, A)^n\, u_0$. In general, this hypothesis is not satisfied for all n (but we need large values of n if we want to reach a given time point $\tau > 0$ with a time step Δt, necessarily small in order to obtain good accuracy). In general, this scheme (unlike what happens for ordinary differential equations) cannot therefore be used here. Moreover, for stability reasons that will be evoked in the sequel, even when we would have $u_0 \in D(A^\infty)$, the method would not be actually efficient.

Study of the backward Euler scheme. Let us first remark that this scheme is equivalently written in the form $(I + \Delta t\, A)u^{n+1} = u^n$. The operator A being mα-accretive, the same property applies to $\Delta t\, A$. By theorem **??** it follows that $(I + \Delta t\, A)^{-1}$ is well defined, belongs to $\mathcal{L}(H, H) \cap \mathcal{L}(H, D(A))$ and also $\|(I + \Delta t\, A)^{-1}\|_{H\to H} \leq 1/d(-1, S_\alpha) = 1$. The solution of scheme 2 is then given by

$$u^n = (I + \Delta t\, A)^{-n} u_0.$$

Stability property. In addition to this, if v^n denotes the solution of this scheme for initial value $v^0 = v_0$ we have the stability property

$$\forall\, n \geq 0, \qquad |u^n - v^n| \leq |u_0 - v_0|,$$

that is, the initial error is not amplified by the scheme.

Smoothing property. Assuming only that $u_0 \in H$, we have $u^1 \in D(A)$, $u^2 \in D(A^2)$,..., $u^n \in D(A^n)$.

Convergence properties.

a) The case of non smooth data: we only assume $u_0 \in H$, and obtain that

$$|u(t_n) - u^n| \leq C_\alpha \frac{6\Delta t}{t_n\, \cos^2 \alpha}\; |u_0| \tag{5}$$

Proof. We remark that

$$u(t_n) - u^n = (E(t_n) - (I + \Delta t\, A)^{-n})\, u_0 = (\exp(-t_n\, A) - (I + \frac{t_n}{n}\, A)^{-n})\, u_0.$$

We then apply corollary **??** to the mα-accretive operator $t_n\, A$ and then lemma **??**, thus obtaining

$$\|\exp(-t_n\, A) - (I + \frac{t_n}{n}\, A)^{-n}\|_{H\to H} \leq C_\alpha \sup_{z\in S_\alpha} |e^{-z} - \big(\frac{1}{1+z/n}\big)^n| \leq C_\alpha \frac{6}{n\, \cos^2 \alpha}.$$

To obtain (**??**) we just have to notice that $n = t_n/\Delta t$.

□

Remark. When we look at an error estimate like that given by (**??**), we have to consider the observation time t_n as fixed; then n goes to infinity when Δt tends to zero. We see that the error estimate is worse for small t_n and improves later on.

b) Case of smooth data : $u_0 \in D(A)$. In this case we have

$$\forall\, n \geq 2, \qquad |u(t_n) - u^n| \leq C_\alpha \frac{3\Delta t}{\cos\alpha}\ |Au_0|. \tag{6}$$

Proof. We take

$$\varphi_n(z) = \frac{e^{-nz} - (1+z)^{-n}}{z},$$

and hence have $\Delta t\, A\varphi_n(\Delta t\, A) = E(t_n) - (I + \Delta t\, A)^{-n}$ and therefore $u(t_n) - u^n = \Delta t\, \varphi_n(\Delta t\, A) Au_0$. By corollary **??**, it is enough to prove the bound $|\varphi_n(z)| \leq 3/\cos\alpha$, for all $z \in S_\alpha$. To do that we remark that

$$\varphi_n(z) = \frac{e^{-z} - \frac{1}{1+z}}{z} \sum_{k=0}^{n-1} e^{-kz}\ \frac{1}{(1+z)^{n-1-k}}.$$

We then use the two remarks to lemma **??**, take $\rho = \operatorname{Re} z$, (thus $\rho \geq |z|\cos\alpha$) and obtain

$$|\varphi_n(z)| \leq \frac{3}{2}\ \frac{\rho}{\cos\alpha}\ n\ \frac{1}{(1+\rho)^{n-1}} \leq \frac{3}{2\cos\alpha}\ \frac{n(n-2)^{n-2}}{(n-1)^{n-1}} \leq \frac{3}{\cos\alpha},$$

since the second term reaches its maximum when $\rho = 1/(n-2)$. □

Remark. We see that, when the initial value is smooth, the error estimate is uniform in $0(\Delta t)$. We have not treated the case $n = 1$, which is of little interest, but we would be able to show directly that $\varphi_1(z) = \int_0^1 e^{-sz}\, ds - \frac{1}{1+z}$, and therefore $|u(t_1) - u^1| \leq 2\,\Delta t\,|Au_0|$.

Study of the Crank-Nicolson scheme. This one is written as

$$u^{n+1} = (I - \tfrac{\Delta t}{2}\, A)(I + \tfrac{\Delta t}{2}\, A)^{-1} u^n, \qquad \text{i.e. } u^{n+1} = r(\Delta t\, A)\, u^n, \quad \text{taking } r(z) = \frac{1 - \frac{z}{2}}{1 + \frac{z}{2}}.$$

Stability property. We have still (with the same notations as for the previous scheme) the stability property

$$\|r(\Delta t\, A)^n\|_{H \to H} \leq 1, \qquad \text{and therefore} \quad \forall\, n \geq 0, \quad |u^n - v^n| \leq |u_0 - v_0|.$$

Proof. It is enough to see that $\|r(\Delta t\, A)\|_{H\to H} \leq 1$. Indeed, let $w = r(\Delta t\, A)v$ and let us take $\varphi = (I + \frac{\Delta t}{2}\, A)^{-1} v$. We have

$$|v|^2 - |w|^2 = |(I + \tfrac{\Delta t}{2}\, A)\varphi|^2 - |(I - \tfrac{\Delta t}{2}\, A)\varphi|^2 = 2\,\Delta t\ \operatorname{Re}(A\varphi, \varphi) \geq 0,$$

and therefore $|w| \leq |v|$. □

Unlike the Euler implicit scheme, there is now no smoothing phenomenon, nor loss of smoothness as that of the explicit Euler method. It is easy to verify that if $u_0 \in D(A^\ell)$, then $u^n \in D(A^\ell)$, but nothing better occurs.

Convergence properties.

a) Case of smooth data: $u_0 \in D(A^2)$. In this case we have

$$\forall\, n \geq 2, \qquad |u(t_n) - u^n| \leq C_\alpha\ \max\big(2, \frac{5}{6\, e\ \cos\alpha}\big)\ \Delta t^2\ |A^2 u_0|. \tag{7}$$

Proof. We consider

$$\varphi_n(z) = \frac{e^{-nz} - r(z)^n}{z^2}, \qquad \text{and we hence have} \quad u(t_n) - u^n = \Delta t^2\, \varphi_n(\Delta t\, A)\, A^2 u_0.$$

Because of corollary **??**, it suffices to show that $|\varphi_n(z)| \le \max\big(2, \frac{5}{6\,e\,\cos\alpha}\big)$ for $z \in S_\alpha$.

If $z \in S_\alpha$ and $|z| \ge 1$. We then straightforwardly have $|\varphi_n(z)| \le 2/|z|^2 \le 2$.

If $z \in S_\alpha$ and $|z| \le 1$. We write

$$\varphi_n(z) = \frac{e^{-z} - r(z)}{z^3}\; z \sum_{k=0}^{n-1} e^{-kz}\; r(z)^{n-1-k}. \tag{8}$$

We first remark (Taylor's formula)

$$e^{-z} - \frac{1-\frac{z}{2}}{1+\frac{z}{2}} = \big(1 - z + \frac{z^2}{2} - \frac{z^3}{2}\int_0^1 (1-s)^2\, e^{-sz}\, ds\big) - \big(1 - z + \frac{z^2}{2} - \frac{z^3}{4}\frac{1}{1+\frac{z}{2}}\big),$$

and hence

$$\Big|\; e^{-z} - \frac{1-\frac{z}{2}}{1+\frac{z}{2}}\;\Big| \le \big(\frac{1}{2}\int_0^1 (1-s)^2\, ds + \frac{1}{4}\big)\,|z|^3 = \frac{5}{12}\,|z|^3.$$

Taking $z = \rho e^{i\theta}$, we have $|\theta| \le \alpha$ and

$$|r(z)|^2 = \Big|\frac{1-\frac{z}{2}}{1+\frac{z}{2}}\Big|^2 = \frac{1+\frac{\rho^2}{4} - \rho\cos\theta}{1+\frac{\rho^2}{4} + \rho\cos\theta} \le \frac{1+\frac{\rho^2}{4}\cos^2\theta - \rho\cos\theta}{1+\frac{\rho^2}{4}\cos^2\theta + \rho\cos\theta} = \Big|\frac{1-\frac{\rho}{2}\cos\theta}{1+\frac{\rho}{2}\cos\theta}\Big|^2.$$

On the other hand we easily verify that, for $x \in [0,1]$, we have $\frac{1-x/2}{1+x/2} \le e^{-x}$. From this we derive the bound

$$|r(z)| \le |e^{-z}| = e^{-\rho\cos\theta} \le e^{-\rho\cos\alpha}.$$

Going back to (**??**) and using that for $x > 0$, $e^{-x} \le \frac{1}{e\,x}$, we obtain

$$|\varphi_n(z)| \le \frac{5}{12}\rho\, n\, e^{-(n-1)\rho\cos\alpha} \le \frac{5}{12}\,\frac{n}{(n-1)\,e\,\cos\alpha} \le \frac{5}{6\,e\,\cos\alpha}.$$

□

b) Case of non smooth data, $u_0 \in H$. The method used for the implicit Euler method does not allow to obtain an interesting bound here: we have

$$\sup_{z\in S_\alpha} \big|e^{-z} - r(z/n)^n\big| \ge 1,$$

since $|r(\infty)| = 1$. However, using the density of $D(A^2)$ in H and $\|r(\Delta t\, A)^n\|_{H\to H} \le 1$, it follows readily from a) that

$$\sup_{n\ge 0} |u(t_n) - u^n| \to 0, \qquad \text{as } \Delta t \to 0.$$

Exercise. Show that if $u_0 \in D(A)$, we have $\forall\, n \ge 2, \quad |u(t_n) - u^n| \le K C_\alpha\, \Delta t\, |A u_0|$.

General scheme. The general one step methods (as for instance Runge-Kutta methods) applied to a linear homogeneous differential problem (P) can be reduced (in general) in this simple case to schemes of the form

$$u^0 = u_0, \quad \text{and for } n \geq 0, \qquad u^{n+1} = r(\Delta t\, A)\, u^n, \tag{9}$$

where r is a rational fraction. We assume that the operator A is mα-accretive. The scheme is then well defined if the rational fraction is bounded in S_α, which we therefore assume. To analyse these methods we use the following definitions

Definitions. We say that the rational function $r(.)$

- is an approximation of order p if $r(z) = e^{-z} + 0(|z|^{p+1})$ in a neighbourhood of $z = 0$,
- is $A(\alpha)$-acceptable if $|r(z)| \leq 1$ for all $z \in S_\alpha$,
- is strongly $A(\alpha)$-acceptable if $|r(z)| < 1$ for all $z \in S_\alpha$, $z \neq 0$ and if $|r(\infty)| < 1$.

When $\alpha = \frac{\pi}{2}$, we say A-acceptable instead of $A(\frac{\pi}{2})$-acceptable.

Stability property. When the rational fraction r is $A(\alpha)$-acceptable we have the stability property

$$\forall\, n \geq 0, \quad \|r(\Delta t\, A)^n\|_{H\to H} \leq C_\alpha, \qquad \text{and thus} \qquad |u^n - v^n| \leq C_\alpha |u_0 - v_0|.$$

Indeed this follows readily from corollary **??**. This is actually a stability result: the initial error is amplified at most by a factor C_α.

Remark. When the rational fraction is A-acceptable, which is the case for both previous schemes, we can replace C_α by $C_{\frac{\pi}{2}} = 1$. Indeed if the operator A is mα-accretive, it is a fortiori m-accretive.

Smoothing property. When $r(\infty) = 0$, assuming $u_0 \in D(A^\ell)$, we have that $u^1 \in D(A^{\ell+1})$, $u^2 \in D(A^{\ell+2})$,..., $u^n \in D(A^{\ell+n})$. If $r(\infty) \neq 0$ there is no smoothing effect, but, if the rational fraction is $A(\alpha)$-acceptable, we have that $u_0 \in D(A^\ell)$ implies that $u^n \in D(A^\ell)$.

Convergence properties.

Theorem 9. *We make the following hypotheses:*

- *the operator A is mα-accretive,* $0 \leq \alpha < \frac{\pi}{2}$,
- *r is an approximation of order p, and r is $A(\alpha)$-acceptable.*

*Then there exists a constant K such that $\forall\, u_0 \in D(A^p)$, the solution of scheme (**??**) satisfies*

$$\forall\, n \geq 2, \qquad |u(t_n) - u^n| \leq \frac{K}{\cos\alpha}\, \Delta t^p\ |A^p u_0|. \tag{10}$$

If in addition the rational fraction r is strongly $A(\alpha)$-acceptable, we have for all $u_0 \in H$,

$$\forall\, n \geq 2, \qquad |u(t_n) - u^n| \leq \frac{K'}{t_n^p\, (\cos\alpha)^{p+1}}\, \Delta t^p\ |u_0|. \tag{11}$$

Proof. a) Case of smooth data.

We take

$$\varphi_n(z) = \frac{e^{-nz} - r(z)^n}{z^p}, \qquad \text{and therefore have} \quad u(t_n) - u^n = \Delta t^p\, \varphi_n(\Delta t\, A)\, A^p u_0.$$

Because of corollary **??**, it suffices to show that $C_\alpha\, |\varphi_n(z)| \le \frac{K_r}{\cos\alpha}$ for $z \in S_\alpha$. To do that we first remark that there exists a real number $\gamma > 0$ such that

$$\forall\, z \in S_\alpha \;\text{ with } |z| \le \gamma\, (\cos\alpha)^{1/p} \qquad |r(z)| \le e^{-\operatorname{Re} z/2}.$$

Indeed $e^{\operatorname{Re} z/2}|r(z)| = e^{-\operatorname{Re} z/2} + 0(|z|^{p+1})$, and hence $\quad e^{\operatorname{Re} z/2}|r(z)| \le 1 - |z|\cos\alpha/2 + C|z|^{p+1}$.

If $z \in S_\alpha$ and $|z| \ge \gamma\, (\cos\alpha)^{1/p}$. We then have $|\varphi_n(z)| \le 2/|z|^p \le \frac{2}{\gamma^p \cos\alpha}$.

If $z \in S_\alpha$ and $|z| \le \gamma\, (\cos\alpha)^{1/p}$. We write

$$\varphi_n(z) = \frac{e^{-z} - r(z)}{z^{p+1}}\; z \sum_{k=0}^{n-1} e^{-kz}\; r(z)^{n-1-k}.$$

Since the method is of order p we have

$$|e^{-z} - r(z)| \le K_1 |z|^{p+1}, \qquad \text{for } |z| \le \gamma.$$

It follows from this (recall that for $x \ge 0$, $e^{-x} \le \frac{1}{e\,x}$) that

$$|\varphi_n(z)| \le K_1\, n\, |z| e^{-(n-1)\operatorname{Re} z/2} \le \frac{2\, K_1\, n\, |z|}{(n-1)\, e\, \operatorname{Re} z} \le \frac{4\, K_1}{e\, \cos\alpha}.$$

This proves the result with $K = C_\alpha\, \max(\frac{2}{\gamma^p}, \frac{4\,K_1}{e})$.

b) Case of non smooth data. We take

$$\varphi_n(z) = e^{-nz} - r(z)^n, \qquad \text{so that} \quad u(t_n) - u^n = \varphi_n(\Delta t\, A) u_0.$$

It is enough to prove that $C_\alpha\, |\varphi_n(z)| \le \frac{K'}{n^p\, (\cos\alpha)^{p+1}}$ for $z \in S_\alpha$.

If $z \in S_\alpha$ and $|z| \le \gamma\, (\cos\alpha)^{1/p}$. We write

$$\varphi_n(z) = (e^{-z} - r(z)) \sum_{k=0}^{n-1} e^{-kz}\; r(z)^{n-1-k}.$$

and have thus, noticing that for $x \ge 0$, $e^{-x} \le \left(\frac{p+1}{e\,x}\right)^{p+1}$,

$$|\varphi_n(z)| \le K_1\, |z|^{p+1} n e^{-(n-1)\operatorname{Re} z/2} \le K_1\, \frac{n\, (2\, (p+1)\, |z|)^{p+1}}{\big(e\, (n-1)\, \operatorname{Re} z\big)^{p+1}} \le \frac{K_1}{n^p}\, \frac{\big(4(p+1)\big)^{p+1}}{\big(e\, \cos\alpha\big)^{p+1}}.$$

If $z \in S_\alpha$ and $|z| \ge \gamma\, (\cos\alpha)^{1/p}$. We first remark that there exists $\delta > 0$ such that $|r(z)| \le e^{-\delta(\cos\alpha)^{1/p}}$ for all these values of z. We then have

$$\begin{aligned} |\varphi_n(z)| &\le e^{-n \operatorname{Re} z} + e^{-n\,\delta(\cos\alpha)^{1/p}} \le e^{-n\gamma(\cos\alpha)^{1+1/p}} + e^{-n\,\delta(\cos\alpha)^{1/p}} \\ &\le \left(\frac{p}{e\, n\, \gamma(\cos\alpha)^{1+1/p}}\right)^p + \left(\frac{p}{e\, n\, \delta(\cos\alpha)^{1/p}}\right)^p \le \frac{p^p}{e^p\, n^p}\, \frac{1}{(\cos\alpha)^{p+1}} \left(\frac{1}{\gamma^p} + \frac{1}{\delta^p}\right). \end{aligned}$$

We then take $K' = C_\alpha \max\!\left(K_1 \left(\frac{4(p+1)}{e}\right)^{p+1}, (\frac{p}{e\,\gamma})^p + (\frac{p}{e\,\delta})^p\right)$.

□

Remarks. a) If we only assume that $u_0 \in H$, the rational fraction is $A(\alpha)$-acceptable and of order ≥ 1, we can prove, as for the Crank-Nicolson method,

$$\sup_{n\geq 0} |u(t_n) - u^n| \to 0, \qquad \text{when } \Delta t \to 0.$$

b) If we only assume $u_0 \in D(A^{p-k})$, $0 \leq k \leq p$, the rational fraction is strongly $A(\alpha)$-acceptable and of order p, it is not very difficult to modify the proof of b) to obtain the intermediate result

$$\forall\, n \geq 2, \qquad |u(t_n) - u^n| \leq \frac{K}{t_n^k\,(\cos\alpha)^{k+1}}\, \Delta t^p\ |A^{p-k}u_0|. \tag{12}$$

6. Hilbertian integrals. Operator integrals

Let Γ be an oriented simple curve in the complex plane. We consider the maps

$$\begin{array}{ll} z \mapsto f(z) & \qquad z \mapsto B(z) \\ \Gamma \to H & \qquad \Gamma \to \mathcal{L}(H,H), \end{array}$$

- We will say that the integral $\int_\Gamma f(z)\,dz$ is well defined if for all $v \in H$ the integral $\int_\Gamma (f(z), v)\,dz$ is well defined and there exists and element $I_1 \in H$ such that

$$\forall\, v \in H, \qquad (I_1, v) = \int_\Gamma (f(z), v)\,dz$$

  We will then write $I_1 = \int_\Gamma f(z)\,dz$.

- We will say that the integral $\int_\Gamma B(z)\,dz$ is well defined if, for all u and $v \in H$, the integral $\int_\Gamma (B(z)u, v)\,dz$ is well defined and there exists an element $I_2 \in \mathcal{L}(H,H)$ such that

$$\forall\, u,\ v \in H, \qquad (I_2\, u, v) = \int_\Gamma (B(z)u, v)\,dz$$

  We will then write $I_2 = \int_\Gamma B(z)\,dz$.

A sufficient condition for the integral I_1 to be well defined is that the function f be continuous on Γ with values in H and that $\int_\Gamma |f(z)|\,|dz| < +\infty$. Likewise, a sufficient condition for I_2 to be well defined is that the function B be continuous on Γ with values in $\mathcal{L}(H,H)$ and that $\int_\Gamma \|B(z)\|_{H\to H}\,|dz| < +\infty$.

Remark. Assume that in a neighbourhood of the curve Γ, for all u, $v \in H$, the maps $z \mapsto (f(z), v)$ and $z \mapsto (B(z)u, v)$ are holomorphic. Then we have

$$\int_\Gamma f(z)\,dz = \int_{\Gamma'} f(z)\,dz \qquad \text{and} \qquad \int_\Gamma B(z)\,dz = \int_{\Gamma'} B(z)\,dz,$$

for every homotopic deformation Γ' of Γ in this neighbourhood and with the same end-points.

Lemma 10. *Assume that A is an isomorphism from $D(A)$ onto H and that A is $m\alpha$-accretive, $\alpha \in [0, \frac{\pi}{2}]$; let $\beta \in (0, \pi-\alpha)$. Then*

$$\forall\, \lambda \notin S_{\alpha+\beta}, \qquad (\lambda I - A)^{-1} = \frac{1}{2\pi i}\int_{\Gamma_{\alpha+\beta}} \frac{1}{\lambda - z}(z\,I - A)^{-1}\,dz,$$

where $\Gamma_{\alpha+\beta}$ denotes the (counterclockwise) oriented boundary of the sector $S_{\alpha+\beta}$.

Proof. Let us take

$$J = \int_{\Gamma_{\alpha+\beta}} \frac{1}{\lambda - z}(z\,I - A)^{-1}\; dz.$$

We first prove that this integral is well defined. To do that we derive from theorem **??** the bound $\|(z\,I - A)^{-1}\|_{H\to H} \le \frac{1}{d(z,S_\alpha)} = \frac{1}{|z|\,\sin\beta}$, which proves convergence of the integral for large $|z|$. On the other hand it is easy to prove that

$$\forall\, z \;\text{ with } |z| \le \frac{1}{2\,\|A^{-1}\|_{H\to H}}, \qquad \|(z\,I - A)^{-1}\|_{H\to H} \le 2\|A^{-1}\|_{H\to H},$$

which ensures convergence of the integral in a neighbourhood of 0.

We now denote Γ_M the part consisting of points $z \in \Gamma_{\alpha+\beta}$ with $|z| \le M$, and take $C_M = \{z = M\,e^{i\theta}\,; \alpha+\beta \le \theta \le 2\pi - \alpha - \beta\}$, oriented in the sense of decreasing values of θ. Let us remark that, for $M > |\lambda|$, by homotopy, the integral $\int_{\Gamma_M \cup C_M} \frac{1}{\lambda - z}(z\,I - A)^{-1}\; dz$ does not depend on M, and besides

$$\| \int_{C_M} \frac{1}{\lambda - z}(z\,I - A)^{-1}\; dz\|_{H\to H} \le 2\,\pi\, \frac{1}{M - |\lambda|}\; \frac{M}{M\,\sin\beta} \to 0 \quad \text{ as } M \to \infty,$$

and hence we obtain

$$J = \int_{\Gamma_M \cup C_M} \frac{1}{\lambda - z}(z\,I - A)^{-1}\; dz = \int_{\gamma(z,\varepsilon)} \frac{1}{\lambda - z}(z\,I - A)^{-1}\; dz,$$

where $\gamma(z,\varepsilon)$ is the circle of center z and radius ε (small enough). Taking $\varepsilon \to 0$, we get $J = 2\pi i(\lambda\,I - A)^{-1}$, which proves the lemma. □

Theorem 11. *Assume that A is an isomorphism from $D(A)$ onto H and that A is mα-accretive, $\alpha \in [0, \frac{\pi}{2}]$. Let $\beta \in (0, \pi - \alpha)$ and let f be a holomorphic function in the interior of $S_{\alpha+\beta}$, continuous on $S_{\alpha+\beta}$, and such that $f(z) = 0(z^{-\varepsilon})$ when $|z| \to \infty$ for some $\varepsilon > 0$. Then the operator $f(A) \in \mathcal{L}(H,H)$ is well defined and*

$$f(A) = \frac{1}{2\pi i} \int_{\Gamma_{\alpha+\beta}} f(z)\,(z\,I - A)^{-1}\; dz.$$

Proof. The previous lemma proves the result when f is a rational fraction with simple poles in the complement of $S_{\alpha+\beta}$. Then using corollary **??** it holds for all limit, uniform in S_α, of such fractions. Let now f be a function satisfying the hypotheses of the theorem. It then satisfies the Cauchy formula in the interior of $S_{\alpha+\beta}$

$$f(z) = \frac{1}{2\pi i} \int_{\Gamma_{\alpha+\beta}} f(\sigma)\, \frac{1}{\sigma - z}\; d\sigma. \qquad \text{We take} \qquad f_M(z) = \frac{1}{2\pi i} \int_{\Gamma_M} f(\sigma)\, \frac{1}{\sigma - z}\; d\sigma,$$

and thus have $f_M \to f$ uniformly in S_α as $M \to \infty$. Approximating in f_M the integral by the rectangle rule

$$f_M(z) \simeq \frac{1}{2\pi i} \sum_j f(\sigma_j) \frac{\sigma_{j+1} - \sigma_j}{\sigma_j - z},$$

we obtain that f is a uniform limit in S_α of rational fractions r_n with simple poles on $\Gamma_{\alpha+\beta}$. Noticing finally that $f(.)$ is the uniform limit of $f(. + \varepsilon)$ as $\varepsilon > 0 \to 0$, we deduce that f is a uniform limit in S_α of rational fractions $r_n(. - \varepsilon_n)$ with simple poles on $\Gamma_{\alpha+\beta} - \varepsilon_n$, that is to say, in the exterior of $S_{\alpha+\beta}$. □

Remark 1. In particular this theorem proves that for $0 \leq \alpha < \alpha + \beta < \frac{\pi}{2}$,

$$E(t) = \frac{1}{2\pi i} \int_{\Gamma_{\alpha+\beta}} e^{-t\sigma} (\sigma I - A)^{-1} \, d\sigma.$$

Remark 2. If we add the hypothesis of f being holomorphic in a neighbourhood of 0, the theorem holds true without the isomorphism hypothesis for A, with the condition of deforming $\Gamma_{\alpha+\beta}$ (in $S^c_{\alpha+\beta}$) in a small circular arc surrounding 0.

7. Space approximation. Finite element semidiscretization

We now go back to problem (P) written in the form

$$(P) \qquad \begin{cases} \text{find } u \in C^1((0,\infty); V) \cap C^0([0,\infty); H) \text{ such that} \\ \forall\, t > 0, \forall\, w \in V, \quad (u'(t), w) + a(u(t), w) = 0, \\ \qquad\qquad\qquad\qquad u(0) = u_0. \end{cases}$$

In order to approximate this problem we employ an internal method defined by a finite dimensional subspace $V_h \subset V$ (we can for instance choose a conforming finite element space) and we consider the problem

$$(P_h) \qquad \begin{cases} \text{find} \quad u_h \in C^1([0,\infty); V_h) \text{ such that} \\ \forall\, t > 0, \forall\, w_h \in V_h, \quad (u_h'(t), w_h) + a(u_h(t), w_h) = 0, \\ \qquad\qquad\qquad u_h(0) = u_{0h}, \quad \text{given in } V_h. \end{cases}$$

The introduction of the operator $A_h \in \mathcal{L}(V_h, V_h)$ defined by

$$\forall\, v_h, w_h \in V_h, \qquad A_h\, v_h \in V_h \;\text{ and } \qquad (A_h\, v_h, w_h) = a(v_h, w_h),$$

allows us to write (P_h) in the equivalent form

$$(P_h) \qquad \begin{cases} \forall\, t > 0, \qquad u_h'(t) + A_h\, u_h(t) = 0, \\ \qquad\qquad\qquad u_h(0) = u_{0h}. \end{cases}$$

This problem has a unique solution (it is a linear homogenous differential system in finite dimension). We write it in the form

$$u_h(t) = E_h(t)\, u_{0h},$$

by taking

$$E_h(t) := \exp(-t\, A_h) = I - t\, A_h + \cdots + \frac{(-t)^k}{k!}\, A_h^k + \cdots = \frac{1}{2\pi i} \int_{\Gamma_{\alpha+\beta}} e^{-tz}\, (zI - A_h)^{-1}\, dz.$$

Let us remark that, for exactly the same reasons as for A, and for the same value $\alpha = \text{Arccos}\,(a_0/M)$, the operator A_h is mα-accretive on the space V_h endowed with the inner product of H. Since we are in finite dimension, the operator A_h is necessarily bounded $D(A_h) = V_h$.

Matrix form. The operator A_h is an abstract operator, whose use is convenient for the sake of numerical analysis, but is not directly adapted to effective computations. For that we need to introduce a basis $\{\psi_j\}_{j=1}^N$ of V_h. We then write

$$m_{ij} = (\psi_j, \psi_i), \quad k_{ij} = a(\psi_j, \psi_i), \quad M = (m_{ij}), \quad K = (k_{ij}).$$

The matrix M is called the mass matrix, and K the stiffness matrix. If we denote $U(t) = (u_1(t), u_2(t), \ldots, u_N(t))^T$ the column vector of the components of $u_h(t)$ in basis ψ_j, the problem (P_h) is equivalent to the differential system

$$\begin{cases} \forall\, t > 0, & M\, U'(t) + K\, U(t) = 0, \\ & U(0) = U_{0h}. \end{cases}$$

This system is the one that is directly used in computations. The matrix $M^{-1}K$ is the matrix that represents the operator A_h in basis $\{\psi_j\}_{j=1}^N$.

Error estimates. In order to study the error we will use the following operators

- P_h orthogonal projection of H onto V_h, (it is in fact defined from V' onto V_h) by $\quad P_h\varphi \in V_h$ and $\forall\, w_h \in V_h$, $(P_h\varphi, w_h) = (\varphi, w_h)$.
- $T = A^{-1} \in \mathcal{L}(V', V), \qquad T_h = A_h^{-1} P_h \in \mathcal{L}(V', V_h), \qquad Q_h = A_h^{-1} P_h A \in \mathcal{L}(V, V_h).$

Let $f \in V'$ be given, take $u = T\, f$ and $u_h = T_h\, f = Q_h\, u$. We then have

$$\begin{cases} u \in V, \quad \text{and } \forall\, w \in V, \\ \quad a(u, w) = (f, w), \end{cases} \qquad \begin{cases} u_h \in V_h, \quad \text{and } \forall\, w_h \in V_h, \\ \quad a(u_h, w_h) = (f, w_h). \end{cases}$$

The error bounds between u and u_h are the main object of the approximation theory of elliptic problems by the finite element method. We have in particular

$$\forall\, v_h \in V_h, \qquad a(u - u_h, v_h) = 0,$$

from where by Céa's lemma and the Aubin-Nitsche duality argument,

$$\begin{array}{l} \|u - u_h\| \le \frac{M}{a_0}\ \inf_{v_h \in V_h} \|u - v_h\|, \\ [10pt] |u - u_h| \le M\, \|u - u_h\| \sup\{\inf_{w_h \in V_h} \|T^* g - w_h\|\ ; g \in H, |g| = 1\}. \end{array} \tag{13}$$

We will make in the sequel the hypothesis

$$(\mathcal{V}_1) \qquad \forall\, f \in H, \qquad \inf_{v_h \in V_h} \|Tf - v_h\| + \inf_{v_h \in V_h} \|T^* f - v_h\| \le C\, h\, |f|.$$

Because of (**??**) we deduce from this hypothesis the following property

$$(\mathcal{H}_1) \qquad \forall\, f \in H, \qquad |Tf - T_h f| \le C\, h^2\, |f|.$$

Remark. If we are in the frame of the example described in paragraph 4, assuming Ω to be a convex polygon, $(\mathcal{V}_1)$ follows from the estimates

$$\inf_{v_h \in V_h} \|u - v_h\|_{H^1(\Omega)} \le C\, h\ |u|_{H^2(\Omega)} \qquad \text{and} \quad |Tf|_{H^2(\Omega)} \le C\, |f|_{L^2(\Omega)},$$

which are typical results for elliptic problems.

The operator Q_h is called “the elliptic projection operator onto V_h”.

Without any smoothness requirement on the initial value, we have the following result

Theorem 12. *Assume* $(\mathcal{H}_1)$ *and* $u_0 \in H$*. Then the following error estimate holds*

$$\forall t > 0, \qquad |u(t) - u_h(t)| \leq |P_h u_0 - u_{0h}| + \frac{C\, h^2}{t\, \cos^3 \alpha}\, |u_0|.$$

Moreover

$$\forall t > 0, \qquad t^k\, |u^{(k)}(t) - u_h^{(k)}(t)| \leq \Big(\frac{k!}{\cos^k \alpha} |P_h u_0 - u_{0h}| + \frac{C\, h^2}{t\, \cos^{k+3} \alpha}\, |u_0| \Big).$$

Let us remark that, with the choice $u_{0h} = P_h u_0$, this provides an error $0(h^2/t)$.

Proof. a) If $u_{0h} = P_h u_0$. We write (with $0 \leq \alpha < \alpha + \beta < \frac{\pi}{2}$)

$$\begin{aligned} u(t) - u_h(t) &= E(t)\, u_0 - E_h(t)\, P_h u_0 \\ &= \frac{1}{2\pi i} \int_{\Gamma_{\alpha+\beta}} e^{-tz} \big((zI - A)^{-1} u_0 - (zI - A_h)^{-1} P_h u_0 \big)\, dz. \end{aligned} \tag{14}$$

On the other hand

$$\begin{aligned} (zI - A)^{-1} u_0 - (zI - A_h)^{-1} P_h u_0 &= T(zT - I)^{-1} u_0 - (zT_h - I)^{-1} T_h u_0 \\ &= (zT_h - I)^{-1} (zT_h - I) T (zT - I)^{-1} u_0 - (zT_h - I)^{-1} T_h (zT - I)(zT - I)^{-1} u_0 \\ &= (zT_h - I)^{-1} (T_h - T)(zT - I)^{-1} u_0. \end{aligned}$$

By theorem 1 (for $z \in \Gamma_{\alpha+\beta}$)

$$\|(zT - I)^{-1}\|_{H \to H} = \|A(zI - A)^{-1}\|_{H \to H} \leq \frac{|z]}{d(z, S_\alpha)} = \frac{1}{\sin \beta}.$$

We have the same bound for $\|(zT_h - I)^{-1}\|_{H \to H}$, and thus from hypothesis $(\mathcal{H}_1)$, substituting in equation (**??**),

$$|u(t) - u_h(t)| \leq \frac{1}{2\pi} \int_{\Gamma_{\alpha+\beta}} |e^{-tz}|\, \frac{C\, h^2}{\sin^2 \beta}\, |dz|\; |u_0| = \frac{C\, h^2}{2\, \pi\, \sin^2 \beta}\; \frac{2}{t\, \cos(\alpha + \beta)}\; |u_0|.$$

Taking $\beta = \frac{\pi}{4} - \frac{\alpha}{2}$, and noticing that $2 \sin \beta \geq \sin 2\beta = \cos \alpha$, we deduce that

$$|u(t) - u_h(t)| \leq \frac{C\, h^2}{\pi\, t\, \sin^3 \beta}\; |u_0| \leq \frac{8\, C\, h^2}{\pi\, t\, \cos^3 \alpha}\; |u_0|.$$

b) If $u_{0h} \neq P_h u_0$, we use the relation $u_h(t) = E_h(t)\, P_h u_0 + E_h(t)(u_{0h} - P_h u_0)$, and the bound $|E_h(t)(u_{0h} - P_h u_0)| \leq |u_{0h} - P_h u_0|$.

c) For the bound concerning k-th order derivatives, we write

$$\begin{aligned} t^k\, (u^{(k)}(t) - u_h^{(k)}(t)) = \; & \tfrac{1}{2\pi i} \int_{\Gamma_{\alpha+\beta}} (-tz)^k\, e^{-tz} \big((zI - A)^{-1} u_0 - (zI - A_h)^{-1} P_h u_0 \big)\, dz \\ & + t^k\, E^{(k)}(t)\, (u_{0h} - P_h u_0). \end{aligned}$$

The first term on the right hand side is bounded as in a). For the second one we use the bound (**??**) of corollary **??**. □

Let us now consider the case of a smooth initial data: $u_0 \in D(A)$.

Theorem 13. *Assume* $u_0 \in D(A)$ *and* $(\mathcal{H}_1)$. *Then*

$$\forall\, t > 0, \qquad |u(t) - u_h(t)| \le |u_0 - u_{0h}| + \frac{C\,h^2}{\cos\alpha}\,|A\,u_0|.$$

Proof. a) Hypothesis $(\mathcal{H}_1)$ implies

$$|u(t) - P_h u(t)| \le |u(t) - Q_h u(t)| = |(T - T_h)\,A\,u(t)| \le C\,h^2\,|A\,u(t)| \le C\,h^2\,|A\,u_0|.$$

Likewise, using the bound (**??**) of corollary **??**

$$|u'(t) - P_h u'(t)| \le |u'(t) - Q_h u'(t)| = |(T - T_h)\,A\,u'(t)| \le C\,h^2\,|A\,u'(t)| \le \frac{C\,h^2}{t\,\cos\alpha}\,|A\,u_0|.$$

In particular, we then deduce the bounds

$$|P_h u(t) - Q_h u(t)| \le 2C\,h^2\,|A\,u_0|, \quad \text{and} \quad |P_h u'(t) - Q_h u'(t)| \le \frac{2\,C\,h^2}{t\,\cos\alpha}\,|A\,u_0|.$$

Let us now take

$$e_h(t) = Q_h u(t) - u_h(t), \quad \text{thus} \quad u(t) - u_h(t) = e_h(t) + u(t) - Q_h u(t).$$

Because of the previous bounds, to prove the theorem, it is sufficient to show that

$$|e_h(t)| \le |e_h(0)| + \frac{K\,h^2}{\cos\alpha}\,|A\,u_0|.$$

b) Since u and u_h are solutions to (P), we have

$$e_h'(t) + A_h e_h(t) = Q_h\,u'(t) + P_h\,A\,u(t) = (Q_h - P_h)\,u'(t).$$

The solution of this linear differential equation with right hand side is given by Duhamel's principle (also called the method of variations of constants)

$$\begin{aligned} e_h(t) &= E_h(t)\,e_h(0) + \textstyle\int_0^t E_h(t-\sigma)\,(Q_h - P_h)\,u'(\sigma)\,d\sigma, \\ &= E_1 + E_2 + E_3, \end{aligned}$$

with $E_1 = E_h(t)\,e_h(0)$, $E_2 = \int_{t/2}^t E_h(t-\sigma)\,(Q_h - P_h)\,u'(\sigma)\,d\sigma$, $E_3 = \int_0^{t/2} E_h(t-\sigma)\,(Q_h - P_h)\,u'(\sigma)\,d\sigma$.

We clearly have $|E_1| \le |e_h(0)|$ and

$$|E_2| \le \int_{t/2}^t \frac{2\,C\,h^2}{\sigma\cos\alpha}\,|A\,u_0|\,d\sigma = \frac{2\,\log 2\,C\,h^2}{\cos\alpha}\,|A\,u_0|.$$

For the third term we first integrate by parts

$$E_3 = E_h(t/2)(Q_h - P_h)\,u(t/2) - E_h(t)(Q_h - P_h)\,u(0) - \int_0^{t/2} A_h\,E_h(t-\sigma)\,(Q_h - P_h)\,u(\sigma)\,d\sigma,$$

and therefore

$$|E_3| \leq 2\,C\,h^2\,|A\,u_0| + 2\,C\,h^2\,|A\,u_0| + 2\,C\,h^2\,|A\,u_0| \int_0^{t/2} \|A_h\,E_h(t-\sigma)\|_{H\to H}\,d\sigma.$$

From corollary **??** we have the bound

$$\|A_h\,E_h(t-\sigma)\|_{H\to H} \leq \frac{1}{(t-\sigma)\cos\alpha} \qquad \text{from where} \qquad |E_3| \leq C\,h^2\,(4 + \frac{2\log 2}{\cos\alpha})\,|A\,u_0|,$$

which completes the proof of the theorem.

□

8. Space semi-discretizations: higher order methods

In this paragraph we will need to use the non integer powers of mα-accretive operators, $0 \leq \alpha \leq \frac{\pi}{2}$. To do that we introduce the following functions

$$\varphi_s(z) = \frac{z^s}{(1+z)^s}, \quad \psi_s(z) = \frac{1}{(1+z)^s},$$

where $z \in S_\alpha$, $0 \leq s < 1$. We will choose the principal determination for the powers (that is, the one that is continuous and real valued for real z).

Definitions. For $0 \leq s < 1$ and $k \in \mathbb{N}$

$$\begin{aligned}
&D(A^s) := \{u \in H\,;\exists\, v \in H \ \text{ with } \ u = \psi_s(A)\,v\},\\
&\text{if } u \in D(A^s), \qquad A^s\,u = \varphi_s(A)\,v,\\
&D(A^{k+s}) := \{u \in D(A^k)\,; A^k\,u \in D(A^s)\},\\
&\text{if } u \in D(A^{k+s}), \qquad A^{k+s}\,u = A^s(A^k\,u).
\end{aligned}$$

We have the following properties

Theorem 14. *a) If $0 \leq s \leq s+s'$, $\quad D(A^{s+s'}) \subset D(A^s) \cap D(A^{s'})$.*
Besides, if $u \in D(A^{s+s'})$, then $A^{s'}\,u \in D(A^s)$ and $A^{s+s'}\,u = A^s\,(A^{s'}\,u)$.

b) If $u_0 \in D(A^s)$, then, $\forall\, t \geq 0$, $E(t)\,u_0 \in D(A^s)$ and we have $A^s\,E(t)\,u_0 = E(t)\,A^s\,u_0$.

c) For $s \in [0,1]$, the operator A^s is m-$s\alpha$-accretive.

d) When A proceeds from a bounded V-elliptic hermitic form, that is, when

$$A = A^* \quad \textit{and} \quad \forall\, v \in V, \quad a_0\,\|v\|^2 \leq (Av,v) \leq M\,\|v\|^2,$$

$$\textit{then } D(A^{1/2}) = V \quad \textit{and} \quad \forall\, v \in V, \quad \sqrt{a_0}\,\|v\| \leq |A^{1/2}v| \leq \sqrt{M}\,\|v\|.$$

Proof. a) It is sufficient to look at the case $s+s' \leq 1$, which we suppose henceforth. We remark that $\psi_{s+s'}(z) = \psi_s(z)\,\psi_{s'}(z)$, and $\varphi_{s+s'}(z) = \varphi_s(z)\,\varphi_{s'}(z)$ and that these functions satisfy the hypotheses of corollary **??**. The condition $u \in D(A^{s+s'})$ implies (with $v \in H$) that

$$u = \psi_{s+s'}(A)\,v = \psi_{s'}(A)\,(\psi_s(A)\,v) = \psi_s(A)\,(\psi_{s'}(A)\,v),$$

and we therefore deduce that $D(A^{s+s'}) \subset D(A^s) \cap D(A^{s'})$. In addition to this

$$A^{s'} u = \varphi_{s'}(A)\,(\psi_s(A)\,v) = \psi_s(A)\,(\varphi_{s'}(A)\,v),$$

which proves that $A^{s'} u \in D(A^s)$ and

$$A^s\,(A^{s'} u) = \varphi_s(A)\,(\varphi_{s'}(A)\,v) = \varphi_{s+s'}(A)\,v = A^{s+s'} u.$$

b) This follows directly from the relation $E(t)\,\psi_s(A) = \psi_s(A)\,E(t)$.

c) Let $\lambda \notin S_{s\alpha}$. We take $f(z) = \frac{1}{\lambda - z^s}$. The function f satisfies the hypotheses of corollary **??** ; besides, $f(\lambda\psi_s - \varphi_s) = \psi_s$. Let now $u \in D(A)^s$, and hence $u = \psi_s(A)v$, with $v \in H$. We have

$$f(A)\,(\lambda I - A^s)u = f(A)\,(\lambda\psi_s(A) - \varphi_s(A))\,v = \psi_s(A)\,v = u.$$

This proves that $\lambda I - A^s$ is an isomorphism from $D(A^s)$ to H and that $f(A) = (\lambda I - A^s)^{-1}$. From theorem **??**, to show the $s\alpha$-accretiveness, it suffices to prove that $\|f(A)\|_{H\to H} \le |\lambda|^{-1}$ when $\lambda = \rho\, e^{\pm i(s\alpha+\pi/2)}$. To do that we examine the two possible cases

- $\lambda = \rho\, e^{i(s\alpha+\frac{\pi}{2})}$. We take $B = e^{i(\frac{\pi}{2}-\alpha)}\,A$ and $g(z) = f(e^{-i(\frac{\pi}{2}-\alpha)}z)$. It is clear that $f(A) = g(B)$ and that the operator B is m-accretive. We thus have

$$\begin{aligned}\|f(A)\|_{H\to H} \le \sup_{\operatorname{Re} z\ge 0} |g(z)| &= \sup_{\operatorname{Re} z\ge 0} \frac{1}{|\rho\, e^{i(s\alpha+\frac{\pi}{2})} - e^{-is(\frac{\pi}{2}-\alpha)}\,z^s|} \\ &= \sup_{\operatorname{Re} z\ge 0} \frac{1}{|\rho + e^{i(1-s)\frac{\pi}{2}}\,z^s|} \le \frac{1}{\rho}.\end{aligned}$$

- $\lambda = \rho\, e^{-i(s\alpha+\frac{\pi}{2})}$. We argue as in the previous case, taking now $B = e^{-i(\frac{\pi}{2}-\alpha)}\,A$ and $g(z) = f(e^{i(\frac{\pi}{2}-\alpha)}z)$.

d) This results directly from the relation $(Av, v) = |A^{1/2}v|^2$.

□

We now introduce the hypothesis, for strictly positive integer k,

$$(\mathcal{H}_k) \qquad \forall\, f \in D(A^{(k-1)/2}), \qquad |Tf - T_h f| \le C\,h^{k+1}\,|A^{(k-1)/2} f|.$$

Remark. If we set ourselves in the frame of the example described in paragraph 4, it is rather easy to prove that $D(A^{1/2}) = V$ (see lemma **??** below); besides, the regularity theorems for elliptic problems prove that, whenever the boundary of Ω is smooth enough as well as the coefficients of the sesquilinear form a, then $D(A^{k/2}) \subset H^k(\Omega)$, even for odd values of k. Hypothesis $(\mathcal{H}_k)$ seems then natural in first sight (for finite elements based on polynomials of degree k). This is indeed the case when

we are in dimension 1

we are in dimension 2 on a rectangle with periodic conditions,

this may be the case for an open polygon Ω. For $k > 1$ the regularity theorems are not directly valid (except in very particular situations via reflection techniques) but only on weighted spaces.

Grid refinement techniques allow $(\mathcal{H}_k)$ to hold. Another possibility would be the introduction of functions of singularities in the finite element space.

when Ω has a smooth boundary, in order to satisfy the hypothesis we are led to use curved finite elements. It is possible to slightly modify the formalism of isoparametric elements to keep $\Omega_h = \Omega$, $V_h \subset V$, and satisfy $(\mathcal{H}_k)$. We will not enter into these details.

Theorem 15. *Let $k \geq 1$ be an integer. Under the assumptions $u_0 \in D(A^{(k+1)/2})$ and $(\mathcal{H}_k)$ we have,*

$$\forall\, t > 0, \qquad |u(t) - u_h(t)| \leq |u_0 - u_{0h}| + \frac{C\, h^{k+1}}{\cos\alpha}\, |A^{(k+1)/2}\, u_0|.$$

Proof. We just have to take again the proof of theorem **??** using the estimates

$$|(T - T_h)\, A\, u(t)| \leq C\, h^{k+1}\, |A^{(k+1)/2}\, u(t)| \leq C\, h^{k+1}\, |A^{(k+1)/2}\, u_0|,$$

and

$$|(T - T_h)\, A\, u'(t)| \leq C\, h^{k+1}\, |A^{(k+1)/2}\, u'(t)| \leq C\, \frac{h^{k+1}}{t\, \cos\alpha}\, |A^{(k+1)/2}\, u_0|.$$

□

In order to study the case of non smooth data we will need more precise hypotheses on the approximation method. To do that we introduce a new set of hypotheses depending on the parity of the index.

$$(\mathcal{V}_{2\ell+1}) \qquad \begin{cases} \forall\, f \in D(A^{\ell}), & \inf_{v_h \in V_h}\ \|T f - v_h\| \leq C\, h^{2\ell+1}\, |A^{\ell} f|, \\ \forall\, g \in D(A^{*\ell}), & \inf_{v_h \in V_h}\ \|T^* g - v_h\| \leq C\, h^{2\ell+1}\, |A^{*\ell} g|, \end{cases}$$

$$(\mathcal{V}_{2\ell}) \qquad \begin{cases} \forall\, f \in D(A^{\ell-1/2}), & \inf_{v_h \in V_h}\ \|T f - v_h\| \leq C\, h^{2\ell}\, |A^{\ell-1/2} f|, \\ \forall\, g \in D(A^{*(\ell-1/2)}), & \inf_{v_h \in V_h}\ \|T^* g - v_h\| \leq C\, h^{2\ell}\, \|A^{*\ell-1} g\|. \end{cases}$$

Lemma 16. *Let $k \geq 1$ be an integer.*

Conditions $(\mathcal{V}_k)$ and $(\mathcal{V}_1)$ imply $(\mathcal{H}_k)$,

Conditions $(\mathcal{V}_1)$ and $(\mathcal{V}_2)$ imply $\forall\, f \in H$, $\quad \|T_h(T - T_h) f\| \leq C h^3 |f|$,

Conditions $(\mathcal{V}_1)$ and $(\mathcal{V}_3)$ imply $\forall\, f \in H$, $\quad |T_h(T - T_h) f| \leq C h^4 |f|$.

Proof. Let $u = T\, f$ and $u_h = T_h\, f$. We then have

$$\begin{cases} u \in V, \quad \text{and } \forall\, w \in V, \\ \qquad a(u, w) = (f, w), \end{cases} \qquad\qquad \begin{cases} u_h \in V_h, \quad \text{and } \forall\, w_h \in V_h, \\ \qquad a(u_h, w_h) = (f, w_h). \end{cases}$$

Since $V_h \subset V$ we have, $\forall\, w_h \in V_h$, $a(u - u_h, w_h) = 0$. The estimates (**??**) are still valid

$\|u - u_h\| \leq \frac{M}{a_0}\ \inf_{v_h \in V_h} \|u - v_h\|,$
$[10pt] |u - u_h| \leq M\, \|u - u_h\| \sup\{\inf_{w_h \in V_h} \|T^* g - w_h\|\ ; g \in H, |g| = 1\}.$

a) For $Tf - T_h f = u - u_h$, we get from $(\mathcal{V}_1)$ and $(\mathcal{V}_k)$

$$|Tf - T_h f| \leq M\, \|u - u_h\|\, C\, h \leq \frac{M^2}{a_0}\, C\, h^k\, |A^{(k-1)/2} f|\, C\, h.$$

b) We first remark that $T_h(T-T_h)f = (T_h-T)(T-T_h)f + T(T-T_h)f$. Firstly we bound the first term by $(\mathcal{V}_1)$ and $(\mathcal{H}_1)$

$$\|(T_h-T)(T-T_h)f\| \le C\,h\,|(T-T_h)f| \le C\,h\,C\,h^2\,|f|.$$

For the second one $\varphi = T(T-T_h)f = T(u-u_h)$ we write

$$a_0\,\|\varphi\|^2 \le \mathrm{Re}(a(T(u-u_h),\varphi) = \mathrm{Re}(a(u-u_h, T^*\varphi - w_h) \le M\|u-u_h\|\;\|T^*\varphi - w_h\|.$$

We then use that $\|u-u_h\| \le C\,h\,|f|$ and that $\inf_{w_h\in V_h}\|T^*\varphi - w_h\| \le C\,h^2\|\varphi\|$.

c) We proceed likewise, using the bounds

$$|(T_h-T)(T-T_h)f| \le C\,h^2\,|(T-T_h)f| \le C\,h^2\,C\,h^2\,|f|,$$

$$|\varphi|^2 = (T(u-u_h),\varphi) = a(u-u_h,(T^*)^2\varphi - w_h) \le M\|u-u_h\|\;\|(T^*)^2\varphi - w_h\|,$$

therefore $\|u-u_h\| \le C\,h\,|f|$, and finally $\inf_{w_h\in V_h}\|T^*T^*\varphi - w_h\| \le C\,h^3|A^*T^*\varphi| = C\,h^3|\varphi|$.

□

Theorem 17. *We assume $u_0 \in H$, $(\mathcal{V}_1)$ and $(\mathcal{V}_k)$, $k=1,2$ or 3. In the case $k=2$ we also suppose that*

$$\forall\, v_h \in V_h, \quad |A_h^{1/2}v_h| \le K\|v_h\|. \tag{15}$$

Then we have the bound

$$\forall\, t>0, \qquad |u(t)-u_h(t)| \le |P_hu_0 - u_{0h}| + \frac{C\,h^{k+1}}{t^{(k+1)/2}\cos^{(k+5)/2}\alpha}\,|u_0|.$$

Proof. We only consider the values $k=2$ or 3, since the result for $k=1$ is given by theorem **??**.

a) Let us first look at the case where $u_{0h} = A_hP_h(A^{-1}u_0)$. For $0 \le \alpha \le \alpha+\beta < \frac{\pi}{2}$ we have

$$\int_{\Gamma_{\alpha+\beta}} e^{-tz}\,dz = 0, \quad \text{and} \quad u_0 = zA^{-1}u_0 - (zI-A)A^{-1}u_0,$$

from where

$$u(t) = E(t)\,u_0 = \frac{1}{2\pi i}\int_{\Gamma_{\alpha+\beta}} e^{-tz}(zI-A)^{-1}u_0\,dz = \frac{1}{2\pi i}\int_{\Gamma_{\alpha+\beta}} z\,e^{-tz}(zI-A)^{-1}A^{-1}u_0\,dz.$$

Likewise

$$u_h(t) = \frac{1}{2\pi i}\int_{\Gamma_{\alpha+\beta}} z\,e^{-tz}(zI-A_h)^{-1}A_h^{-1}A_hP_h(A^{-1}u_0)\,dz.$$

We then have (see the proof of theorem **??**)

$$u(t)-u_h(t) = \frac{1}{2\pi i}\int_{\Gamma_{\alpha+\beta}} z\,e^{-tz}(zT_h-I)^{-1}(T_h-T)(zT-I)^{-1}A^{-1}u_0\,dz,$$

with $\|(zT-I)^{-1}\|_{H\to H} \le \frac{1}{\sin\beta}$ and $\|(zT_h-I)^{-1}\|_{H\to H} \le \frac{1}{\sin\beta}$.

1st case: $k=2$. We bound by $(\mathcal{H}_2)$

$$|(T_h-T)(zT-I)^{-1}A^{-1}u_0| \le Ch^3\,|A^{1/2}(zT-I)^{-1}A^{-1}u_0| = Ch^3\,|A^{1/2}(zI-A)^{-1}u_0|.$$

By corollary **??** we have

$$\|A^{1/2}(zI-A)^{-1}\|_{H\to H} \le C_\alpha \sup_{\zeta\in S_\alpha} \Big|\frac{\zeta^{1/2}}{z-\zeta}\Big| \le \frac{C_\alpha}{|z|^{1/2}\sin\beta},$$

and thus $|(T_h-T)(zT-I)^{-1}A^{-1}u_0| \le \frac{C\,h^3}{|z|^{1/2}\sin\beta}\,|u_0|$. Consequently

$$|u(t)-u_h(t)| \le \frac{1}{2\pi}\int_{\Gamma_{\alpha+\beta}} |z\,e^{-tz}|\,\frac{C\,h^3}{|z|^{1/2}\,\sin^2\beta}\,|u_0|\,|dz| = \frac{C\,h^3\,|u_0|}{\pi\,t^{1.5}\,\sin^2\beta\,\cos^{1.5}(\alpha+\beta)}\int_0^\infty x^{1/2}\,e^{-x}\,dx.$$

We finish by taking $\beta = \frac{\pi}{4}-\frac{\alpha}{2}$ and by noticing that $2\sin\beta \ge \cos\alpha$.

2nd case: $k=3$. We bound by $(\mathcal{H}_3)$

$$|(T_h-T)(zT-I)^{-1}A^{-1}u_0| \le Ch^4\,|A(zT-I)^{-1}A^{-1}u_0| = Ch^4\,|(zT-I)^{-1}u_0|,$$

from where $|(zT_h-I)^{-1}(T_h-T)(zT-I)^{-1}A^{-1}u_0| \le \frac{C\,h^4}{\sin^2\beta}\,|u_0|$, and

$$\begin{aligned}|u(t)-u_h(t)| &\le \frac{1}{2\pi}\int_{\Gamma_{\alpha+\beta}} |z\,e^{-tz}|\,\frac{C\,h^4}{\sin^2\beta}\,|u_0|\,|dz| = \frac{C\,h^4\,|u_0|}{\pi\,t^2\,\sin^2\beta\,\cos^2(\alpha+\beta)}\int_0^\infty x\,e^{-x}\,dx,\\ &\le \frac{16\,C\,h^4\,|u_0|}{\pi\,\cos^4\alpha},\end{aligned}$$

with the choice $\beta = \frac{\pi}{4}-\frac{\alpha}{2}$.

b) When $u_{0h} \ne A_hP_h(A^{-1}u_0)$, we have to add the term

$$b = |E_h(t)(u_{0h} - A_hP_h(A^{-1}u_0))| \le |u_{0h} - P_hu_0| + |e_2|,$$

with

$$e_2 = E_h(t)(P_hu_0 - A_hP_hA^{-1}u_0) = E_h(t)A_h^2T_h(T_h-T)u_0,$$

to the previous bound.

If $k=2$ we derive from (**??**) and from lemma **??**

$$|A_h^{1/2}T_h(T-T_h)u_0| \le K\,\|T_h(T-T_h)u_0\| \le K\,C\,h^3|u_0|,$$

from where

$$|e_2| \le \|E_h(t)\,A_h^{3/2}\|_{H\to H}C\,h^3|u_0| \le \frac{K}{t^{1.5}\,\cos^{1.5}\beta}\,h^3|u_0|.$$

Likewise if $k=3$

$$|T_h(T-T_h)u_0| \le Ch^4|u_0|,$$

and therefore

$$|e_2| \le \|E_h(t)\,A_h^2\|_{H\to H}C\,h^4|u_0| \le \frac{2\,C}{t^2\,\cos^2\beta}\,h^4|u_0|.$$

□

We are left to justifying the adequacy of hypothesis (**??**), which is a consequence of the following lemma.

Lemma 18. *We suppose that the operator A_h proceeds from a bounded V-elliptic sesquilinear form and also that there exists a constant C such that*

$$\forall\, v_h \in V_h, \qquad |(A_h - A_h^*)v_h| \le C\,\|v_h\|.$$

Then, there exists a constant K such that condition (**??**) *holds.*

Proof. Let us take

$$B_h = \frac{1}{2}\,(A_h + A_h^*), \qquad C_h = \frac{1}{2}\,(A_h - A_h^*).$$

From theorem **??**, we have $\sqrt{a_0}\,\|v_h\| \le |B_h^{1/2} v_h| \le \sqrt{M}\,\|v_h\|$. Besides we have , $\forall\,\varepsilon > 0$,

$$A_h^{1/2} v_h = A_h^{1/2}(\varepsilon I + A_h)^{-1}(\varepsilon I + A_h)\,v_h = \frac{1}{2\pi i}\int_{\Gamma_{\alpha+\beta}} \frac{z^{1/2}}{\varepsilon + z}\,(zI - A_h)^{-1}\,(\varepsilon I + A_h)\,v_h\;dz.$$

The formula also holds for $\varepsilon = 0$, since the integral is still convergent. We then have

$$(A_h^{1/2} - B_h^{1/2})\,v_h = \frac{1}{2\pi i}\int_{\Gamma_{\alpha+\beta}} z^{-1/2}\,\big((zI - A_h)^{-1}A_h - (zI - B_h)^{-1}B_h\big)\,v_h\;dz. \tag{16}$$

We write

$$\begin{aligned}\big((zI-A_h)^{-1}A_h - (zI-B_h)^{-1}B_h\big)\,v_h &= z\big((zI-A_h)^{-1} - (zI-B_h)^{-1}\big)\,v_h \\ &= z\,(zI-A_h)^{-1}C_h(zI-B_h)^{-1}\,v_h.\end{aligned}$$

We now use the bounds

$$\|(zI-A_h)^{-1}\|_{H\to H} \le \frac{1}{|z|\,\sin\beta} \quad\text{and}\quad \|(zI-A_h)^{-1}\|_{H\to H} \le 2\,\|A_h^{-1}\|_{H\to H} \quad\text{if}\quad 2|z| \le 1/\|A_h^{-1}\|_{H\to H},$$

$$\|(zI-B_h)^{-1}\|_{H\to H} \le \frac{1}{|z|\,\sin(\alpha+\beta)} \quad\text{and}\quad \|(zI-B_h)^{-1}\|_{H\to H} \le 2\,\|A_h^{-1}\|_{H\to H} \quad\text{if}\quad 2|z| \le 1/\|B_h^{-1}\|_{H\to H},$$

$$\begin{aligned}|C_h(zI-B_h)^{-1}\,v_h| &\le \tfrac{C}{2}\,\|(zI-B_h)^{-1}\,v_h\| \le \tfrac{C}{2\sqrt{a_0}}\,|(zI-B_h)^{-1}\,B_h^{1/2}v_h| \\ &\le \tfrac{C}{2\sqrt{a_0}}\,\|(zI-B_h)^{-1}\|_{H\to H}\,|B_h^{1/2}v_h| \le \tfrac{C\sqrt{M}}{2\sqrt{a_0}}\,\|(zI-B_h)^{-1}\|_{H\to H}\,\|v_h\|.\end{aligned}$$

We apply these bounds to (**??**), which yields

$$|(A_h^{1/2} - B_h^{1/2})\,v_h| \le C'\int_0^\infty z^{1/2}\|(zI - A_h)^{-1}\|_{H\to H}\|(zI - B_h)^{-1}\|_{H\to H}\;dz\,\|v_h\| \le C\text{"}\,\|v_h\|.$$

Finally we obtain (**??**) with $K = C\text{"} + \sqrt{M}$.

Remark. The hypotheses of this lemma are easily verified in the frame of paragraph 4. An analogous proof allows to show that, in this context, $D(A^{1/2}) = V$. This last result has been proven for several elliptic operators, but it is not general. This is the famous Kato's problem. □

9. Full discretization

Up to now we have considered the problem

$$(P) \qquad \begin{cases} u'(t) + A\,u(t) = 0, & t > 0,\\ u(0) = u_0, & \end{cases}$$

and its time and space semi-discretizations

$$(P_{\Delta t}) \quad \begin{cases} u^{n+1} &= r(\Delta t\, A)u^n, \\ \ \ u^0 &= u_0, \end{cases} \qquad\qquad (P_h) \quad \begin{cases} u_h'(t) + A_h\, u_h(t) = 0, \quad t > 0, \\ \qquad u_h(0) = u_{0h}. \end{cases}$$

We now turn our attention to the fully discretized problem, defined by

$$(P_{ap}) \quad \begin{cases} u_h^0 = u_{0h}, \\ u_h^{n+1} = r(\Delta t\, A_h)u_h^n, \qquad n \geq 0. \end{cases}$$

Remark. This compact writing is well suited to numerical analysis of scheme (P_{ap}), but not to effective computations. In order to implement this scheme, we have to write it in matrix form. Let $U_h^n = (u_1^n, u_2^n, \dots, u_N^n)^T$ be the column vector of coefficients of u_h^n in basis $\{\psi_j\}_{j=1}^N$ of V_h, M and K be the mass and stiffness matrices defined in paragraph 7. The operator A_h is then represented by the matrix $M^{-1}K$ in this basis. The recurrence relation (P_{ap}) reads then

$$U_h^0 = U_{0h} \text{ given and, for } n \geq 0, \qquad U_h^{n+1} = r(\Delta t\, M^{-1}K)\, U_h^n.$$

This can be rewritten in a way closer to implementation as

- $M\, U_h^{n+1} = (M - \Delta t\, K)\, U_h^n$, in the explicit Euler case $(r(z) = 1 - z)$,
- $(M + \Delta t\, K)\, U_h^{n+1} = M\, U_h^n$, in the implicit Euler case $(r(z) = \frac{1}{1+z})$,
- $(M + \frac{\Delta t}{2}\, K)\, U_h^{n+1} = (M - \frac{\Delta t}{2}\, K)\, U_h^n$, in the Crank-Nicolson case $(r(z) = \frac{1-z/2}{1+z/2})$,
- writing in the form : $r(z) = \prod_{j=1}^q \frac{\alpha_j - \beta_j\, z}{1+\gamma_j\, z}$ the rational fraction, the general scheme is

$$U_h^{n+1} = \prod_{j=1}^q (M + \gamma_j \Delta t\, K)^{-1} (\alpha_j\, M - \beta_j \Delta t\, K)\, U_h^n.$$

Remarks. 1) Every learned numericist knows that one never computes the matrices M^{-1}, or $M^{-1}K$ (these matrices would in general be full whereas M and K are sparse), and even less $r(\Delta t\, M^{-1}K)$. We just solve linear systems and carry out matrix-vector multiplication of type $(M + \gamma_j \Delta t\, K)U = (\alpha_j\, M - \beta_j \Delta t\, K)\, V$.

2) The so called *explicit Euler* scheme is not really explicit. When we use finite elements without numerical integration, this scheme requires the solution of a linear system of the same type $M\, U = \dots$ as for the implicit Euler scheme $(M + \Delta t\, K)\, U = \dots$. The use of some numerical integration formulae (the so called *lumped mass* methods) would allow to have a diagonal matrix M and recover the advantages of the explicit character, but they do not fit in the frame of our present study, where we assume the forms $a(.\,,.)$ and $(.\,,.)$ to be exactly computed.

The von Neumann criterion.

$$(vN) \qquad \forall\, \lambda \in Sp(A_h), \qquad |r(\lambda\, \Delta t)| \leq 1.$$

The condition above, called von Neumann's criterion is a necessary condition for stability. Indeed, let λ be an eigenvalue of A_h (that is, of $M^{-1}K$), v_h being a corresponding eigenvector. If in(P_{ap}) we substitute the initial value u_{0h} by the perturbed data $u_{0h} + \varepsilon\, v_h$, the solution to the scheme u_h^n is replaced by $u_h^n + \varepsilon\, r(\lambda\, \Delta t)^n v_h$. Perturbations of u_{0h} see their components on the eigenvector v_h amplified by a factor $r(\lambda\, \Delta t)^n$. It follows then that, if $|r(\lambda\, \Delta t)| > 1$, the scheme is violently unstable.

When the form a is hermitic (i.e., when the operators A and A_h are selfadjoint) this condition (vN) is also a sufficient condition for stability. In fact, by taking an orthonormal basis of eigenvectors of A_h (such a basis exists and diagonalises A_h) we obtain

$$\|r(\Delta t\, A_h)^n\|_{H\to H} = \sup_{\lambda\in Sp(A_h)} |r(\lambda\, \Delta t)|^n \leq 1.$$

Theorem **??** gives another sufficient condition: r is A(α)-acceptable. Indeed, since the operator A_h is mα-accretive, we have

$$r \text{ is A}(\alpha)\text{-acceptable} \quad \Longrightarrow \quad \|r(\Delta t\, A_h)^n\|_{H\to H} \leq C_\alpha.$$

Remark. Unlike in the situation of paragraph 5, the explicit Euler scheme can now be employed. Let us now assume, for simplicity's sake, the operator A_h to be selfadjoint; then in the present case, where $r(z) = 1 - z$

$$(vN) \quad \text{is equivalent to} \quad \Delta t \|A_h\|_{H\to H} \leq 2.$$

Typically in a finite element method, the quantity $\|A_h\|_{H\to H}$ behaves like $\frac{c}{\underline{h}^2}$ where $\underline{h}$ is the smallest diameter of the triangles. The Euler method works under the assumption that $\Delta t \leq \frac{2\,\underline{h}^2}{c}$, but this condition requires very small time steps ($\underline{h}$ is very small when desired precision is high, it presents a squared exponent, the situation worsens with grid refinements,...) which makes this explicit method of poor efficiency. The same occurs to any other explicit scheme.

We now turn to the study of the error, first without any smoothness assumption on the initial data.

Theorem 19. *Assume that the hypotheses "$u_0 \in H$, $(\mathcal{V}_1)$, $(\mathcal{V}_k)$, $k = 1, 2$ or 3, r strongly A(α)-acceptable and r of order p" hold. Then we have*

$$\forall\, n \geq 1, \qquad |u(t_n) - u_h^n| \leq |P_h u_0 - u_{0h}| + \frac{C\, h^{k+1}}{t_n^{(k+1)/2} \cos^{(k+5)/2} \alpha}\, |u_0| + \frac{C\, \Delta t^p}{t_n^p \cos^{p+1} \alpha}\, |u_{0h}|.$$

Proof. We write

$$|u(t_n) - u_h^n| \leq |u(t_n) - u_h(t_n)| + |(E_h(t_n) - r(\Delta t\, A_h)^n)\, u_{0h}|.$$

The first term in the right hand side is bounded by theorem **??**, the second one by the estimate (**??**) which still holds with A and u_0 replaced by A_h and u_{0h}. □

Let us now look at the case where we have some additional regularity on the data.

Theorem 20. *Assume that the hypotheses "$u_0 \in D(A)$, $(\mathcal{H}_1)$, r is A(α)-acceptable and of order 1" hold. Then*

$$\forall\, n \geq 1, \qquad |u(t_n) - u_h^n| \leq C\big(|u_0 - u_{0h}| + \frac{h^2 + \Delta t}{\cos \alpha}\, |Au_0|\big).$$

Proof. We write

$$|u(t_n)-u_h^n| \le |u(t_n)-E_h(t_n)\,Q_h u_0| + |(E_h(t_n)-r(\Delta t\,A_h)^n)\,Q_h u_0| + |r(\Delta t\,A_h)^n)(Q_h u_0-u_{0h})|.$$

By theorem **??**

$$|u(t_n)-E_h(t_n)\,Q_h u_0| \le |u_0-Q_h u_0| + \frac{C\,h^2}{\cos\alpha}\,|Au_0| \le \frac{C'\,h^2}{\cos\alpha}\,|Au_0|,$$

and we then use formula (**??**) with $p=1$ and A substituted by A_h

$$|(E_h(t_n)-r(\Delta t\,A_h)^n)\,Q_h u_0| \le C\,\frac{\Delta t\,|A_h\,Q_h u_0|}{\cos\alpha} \le C\,\frac{\Delta t\,|Au_0|}{\cos\alpha}.$$

We finally have

$$|r(\Delta t\,A_h)^n)(Q_h u_0-u_{0h})| \le C_\alpha |Q_h u_0-u_{0h}| \le C_\alpha |u_0-u_{0h}| + C\,h^2\,|Au_0|.$$

□

Theorem 21. *Assume that the hypotheses "$u_0 \in D(A^2)$, $(\mathcal{V}_1)$, $(\mathcal{V}_k)$, $k=1,2$ or 3, r is $A(\alpha)$-acceptable and of order 2" hold. Then*

$$\forall\, n\ge 1, \qquad |u(t_n)-u_h^n| \le C\big(|u_0-u_{0h}| + \frac{h^{k+1}+\Delta t^2}{\cos\alpha}\,|A^2u_0|.$$

If in addition the rational fraction r is strongly $A(\alpha)$-acceptable and of order $p\ge 2$, we have

$$\forall\, n\ge 1, \qquad |u(t_n)-u_h^n| \le C\big(|u_0-u_{0h}| + \frac{h^{k+1}}{\cos\alpha}\,|A^2u_0| + \frac{\Delta t^p}{t_n^{p-2}\,\cos^{p-1}\alpha}\,|A^2u_0|\big).$$

Proof. a) Being $R_h u_0 = A_h^{-2}\,P_h\,A^2u_0 = T_h^2\,A^2u_0$ we write

$$|u(t_n)-u_h^n| \le |u(t_n)-E_h(t_n)\,R_h u_0| + |(E_h(t_n)-r(\Delta t\,A_h)^n)\,R_h u_0| + |r(\Delta t\,A_h)^n)(R_h u_0-u_{0h})|.$$

By theorem **??**

$$|u(t_n)-E_h(t_n)\,R_h u_0| \le |u_0-R_h u_0| + \frac{C\,h^{k+1}}{\cos\alpha}\,|A^{(k+1)/2}u_0|,$$

we then use equation (**??**) with $p=2$ and A substituted by A_h

$$|(E_h(t_n)-r(\Delta t\,A_h)^n)\,R_h u_0| \le C\,\frac{\Delta t^2}{\cos\alpha}\,|P_h\,A^2u_0| \le C\,\frac{\Delta t^2}{\cos\alpha}\,|A^2u_0|,$$

$$|r(\Delta t\,A_h)^n)(R_h u_0-u_{0h})| \le C_\alpha |R_h u_0-u_{0h}| \le C_\alpha(|u_0-u_{0h}| + |u_0-R_h u_0|.$$

In order to finish it suffices to remark that $|A^{(k+1)/2}u_0| \le C\,|A^2u_0|$ and that

$$\begin{aligned} |u_0-R_h u_0| &\le |u_0-Q_h u_0| + |Q_h u_0-R_h u_0| = |u_0-Q_h u_0| + |T_h\,A\,u_0-T_h^2\,A^2u_0| \\ &\le |T_h\,(T-T_h)\,A^2u_0| \le C\,h^{k+1}\,|A^{(k+1)/2}u_0|, \end{aligned}$$

by lemma **??**.

b) It is enough to do the following modification to the previous bounds. From formula (**??**) with $k=p-2$ and A substituted by A_h

$$|(E_h(t_n)-r(\Delta t\,A_h)^n)\,R_h u_0| \le C\,\frac{\Delta t^p}{t_n^{p-2}\,\cos^{p-1}\alpha}\,|P_h\,A^2u_0| \le C\,\frac{\Delta t^2}{t_n^{p-2}\,\cos^{p-1}\alpha}\,|A^2u_0|.$$

□

References

[1] Vidar Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, Springer Verlag, (1997).