Chapitre 4. Les problèmes d'évolution paraboliques

1. Formalisme. Le triplet de Gelf'and

Soit V et H deux espaces de Hilbert sur $\mathbb{C}$, avec $V \subset H$, V dense dans H, l'injection canonique de V dans H étant continue. On note $(\,.\,,.)$ le produit scalaire dans H, $|\,.\,|$ la norme associée, $\|\,.\,\|$ la norme dans V. D'après le théorème de Riesz, à toute forme antilinéaire continue sur H est associé de manière unique un élément $u \in H$ telle que cette forme soit l'application

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

réciproquement un élément $u \in H$ définit ainsi une forme antilinéaire continue sur H. On peut donc identifier H à son antidual H' (c'est à dire décider de représenter toute forme antilinéaire continue sur H par l'élément u correspondant). L'espace $H \equiv H'$ s'identifie alors à un sous-espace de l'antidual V' de V. On a ainsi

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

De plus H est dense dans V', s'y injecte continûment, et on peut utiliser la même notation pour le produit scalaire dans H et la dualité entre V' et V.

Preuve. Notons C_1 la constante de continuité de l'injection de V dans H ; ainsi on a $\forall\, v \in V$, $|v| \le C_1\|v\|$. Soit $u \in H$, l'application $v \mapsto (u, v)$ est continue sur H donc a fortiori sur V. Donc la forme antilinéaire définie par $\ell_u(v) = (u, v)$ vérifie $\ell_u \in V'$. De plus on a

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

Cela montre que l'application $u \mapsto \ell_u$ (qui est clairement linéaire) est continue de H dans V'. De plus elle est injective. En effet, si on a $\ell_u = 0$, alors on a $(u, v) = 0$ pour tout $v \in V$, et par densité de V dans H, $(u, v) = 0$ pour tout $v \in H$; on en déduit $u = 0$, ce qui montre bien l'injectivité.

On peut donc identifier u et ℓ_u, c'est à dire identifier H à un sous-espace de V', et garder la même notation $(\,.\,,.)$ pour le produit scalaire dans H et l'antidualité V', V. Plus précisément cela signifie que :

lorsque $\ell \in V'$ et $v \in V$ la quantité (ℓ, v) représente la forme antilinéaire ℓ appliquée à l'élément v,
lorsque $\ell \in H$ et $v \in H$ la quantité (ℓ, v) représente le produit scalaire dans H des éléments ℓ et v.
Notons que, pour $\ell \in H$ et $v \in V$ la quantité (ℓ, v) est définie de deux manières différentes, mais que ces deux définitions coïncident.

Il reste à montrer que H' est dense dans V'. En effet sinon il existerait une forme L linéaire continue sur V', nulle sur H', et non identiquement nulle sur V'. Puisque V' est un espace de Hilbert on a $(V')^* = V$, c'est à dire, il existe $v \in V$ tel que $\forall\, w \in V'$, $L(w) = (w, v)$. On a $(w, v) = L(w) = 0$ pour tout $w \in H$, en particulier en prenant $w = v$ on obtient $v = 0$. Il en résulte que $L = 0$, d'où contradiction. □

Soit maintenant $a(.,.) : V \times V \to \mathbb{C}$ une forme sesquilinéaire continue ; on peut lui associer l'opérateur $A \in \mathcal{L}(V, V')$ défini par

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

La continuité s'écrit $\forall\, v, w \in V$, $|(Av, w)| \leq C\, \|v\|\, \|w\|$, ou de manière équivalente $\|Av\|_{V'} \leq C\, \|v\|$, ou encore $\|A\|_{V \to V'} \leq C$. Rappelons que l'opérateur adjoint $A^* \in \mathcal{L}(V, V')$ et la forme adjointe a^* sont définis par

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

On introduit maintenant l'espace

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

Il est facile de voir que, muni de la norme $\|v\|_{D(A)} = (\|v\|^2 + |Av|^2)^{1/2}$, cet espace est un espace de Hilbert. De plus si A est un isomorphisme de V sur V' (ce qui est en particulier le cas lorsque la forme bilinéaire $a(.,.)$ est V-elliptique, d'après le théorème de Lax-Milgram) A est alors un isomorphisme de $D(A)$ sur H ; de la densité de H dans V' on déduit celle de $D(A) = A^{-1}H$ dans $V = A^{-1}V'$; V étant dense dans H on a donc aussi $D(A)$ dense dans H.

Par récurrence on définit

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

Exemple 1. Soit Ω un ouvert borné non vide de $\mathbb{R}^2$,

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

On est bien dans la situation V dense dans H puisque $\mathcal{D}(\Omega)$ qui est contenu dans V est dense dans H. Par définition l'espace V', défini par l'identification de H à son antidual, est noté $H^{-1}(\Omega)$. On a ici, $D(A) = \{v \in H_0^1(\Omega); \Delta v \in L^2(\Omega)\}$ et $Av = -\Delta v$. Si de plus Ω est convexe (ou si la frontière de Ω est de classe C^2) par un théorème de régularité pour les problèmes elliptiques, on en déduit $D(A) = H^2(\Omega) \cap H_0^1(\Omega)$. Lorsque la frontière Ω est de classe C^∞ on peut montrer de même $D(A^k) = \{v \in H^{2k}(\Omega); v, \Delta v, \ldots \text{et } \Delta^{k-1} v \in H_0^1(\Omega)\}$.

Exemple 2. On prend maintenant

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

On est a fortiori encore dans la situation V dense dans H. Si Ω est convexe (ou si la frontière de Ω est de classe C^2) par un théorème de régularité on peut montrer que $D(A) = \{v \in H^2(\Omega)\, ; \frac{\partial v}{\partial n} = 0 \text{ sur } \partial\Omega\}$ et alors $Av = -\Delta v + v$. Lorsque la frontière Ω est de classe C^∞ on peut montrer aussi $D(A^k) = \{v \in H^{2k}(\Omega); \frac{\partial v}{\partial n}, \frac{\partial \Delta v}{\partial n}, \ldots \text{et } \frac{\partial \Delta^{k-1} v}{\partial n} = 0 \text{ sur } \partial\Omega\}$.

Remarque : Dans le cas de l'exemple 1 on a une identification supplémentaire

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

en effet la densité de $\mathcal{D}(\Omega)$ dans $H_0^1(\Omega)$ permet d'identifier $H^{-1}(\Omega)$ à un sous-espace de l'espace des distributions sur Ω. Cette identification n'est pas loisible dans l'exemple 2. En effet par exemple, la forme antilinéaire $l(v) = \int_{\partial\Omega} \overline{v}(x)\, d\sigma(x)$ appartient à $(H^1(\Omega))'$, elle est non nulle. Or elle est clairement nulle sur $\mathcal{D}(\Omega)$, elle ne peut donc pas s'identifier à un élément de $\mathcal{D}'(\Omega)$.

Sectorialité. On suppose maintenant que la forme sesquilinéaire a est V-elliptique et continue. Il existe donc deux constantes $a_0 > 0$ et M telles que

$$\forall\, v \in V, \qquad \operatorname{Re} a(v,v) \geq a_0 \, \|v\|^2 \quad \text{et} \quad |a(v,v)| \leq M \, \|v\|^2. \tag{1}$$

Posons $a(v,v) = \rho(v) e^{i\gamma(v)} \, \|v\|^2$, on a donc $a_0 \leq \rho \cos\gamma$ et $\rho \leq M$. Soit $\alpha = \operatorname{Arccos}(a_0/M)$, on a ainsi $0 \leq \alpha < \pi/2$ et la propriété de sectorialité

$$\forall\, v \in V, \quad a(v,v) \in S_\alpha \quad \text{où} \quad S_\alpha := \{z \in \mathbb{C} \,;\, z = 0 \text{ ou } |\operatorname{Arg} z| \leq \alpha\}. \tag{2}$$

Remarquons aussi que, pour tout $z \notin S_\alpha$, l'opérateur $zI - A$ est un isomorphisme de V sur V'. En effet écrivons $z = |z| e^{i\beta}$ et posons

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

Il suffit de montrer que $(B.\,,.)$ est V-elliptique, ce qui entraînera par le théorème de Lax-Milgram que B (et par suite $zI - A$) est un isomorphisme de V sur V'.

Le résultat est immédiat lorsque $\operatorname{Re} z \leq 0$, on a alors $\operatorname{Re}(Bv, v) \geq \operatorname{Re} a(v,v) \geq a_0 \, \|v\|^2$.

Dans le cas $\beta \in]-\frac{\pi}{2}, -\alpha[$, on a $\operatorname{Re}(Bv,v) = \operatorname{Im}(e^{-i\beta} a(v,v)) = \rho(v) \sin(\gamma(v) - \beta) \, \|v\|^2$ (avec les notations utilisées précédemment), on a donc $\operatorname{Re}(Bv, v) \geq a_0 \sin(|\beta| - \alpha) \, \|v\|^2$. On démontre similairement cette même majoration dans le cas $\beta \in]\alpha, \frac{\pi}{2}[$.

Il résulte de ce qui précède que, pour tout $z \notin S_\alpha$ l'opérateur $zI - A$ est aussi un isomorphisme de $D(A)$ sur H.

2. Estimations sur la résolvante

On considère maintenant un opérateur linéaire A, défini sur un sous-espace $D(A) \subset H$, à valeurs dans H, et on suppose $D(A)$ dense dans H. Un tel opérateur est dit non borné sur H. Soit $\alpha \in [0, \frac{\pi}{2}]$, on lui associe le secteur $S_\alpha = \{z \in \mathbb{C} \,;\, z = 0 \text{ ou } |\operatorname{Arg} z| \leq \alpha\}$.

Définition. On dit que l'opérateur A est mα-accrétif si

$$\begin{aligned} &\forall\, v \in D(A), \qquad (Av, v) \in S_\alpha. \\ &\text{et si } \forall z \notin S_\alpha, \qquad zI - A \text{ est un isomorphisme de } D(A) \text{ sur } H. \end{aligned} \tag{3}$$

Lorsque $\alpha = \frac{\pi}{2}$ on parle d'opérateur m-accrétif. Les opérateurs A étudiés au paragraphe précédent fournissent des exemples d'opérateurs mα-accrétifs. L'opérateur $R(z) = (zI - A)^{-1}$ s'appelle la résolvante de A au point z.

Théorème 1. *Soit A un opérateur mα-accrétif. Alors pour tout $z \notin S_\alpha$ on a les estimations suivantes*

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

De plus les applications $z \mapsto (zI - A)^{-1}$ de S_α^c (complémentaire de S_α) dans $\mathcal{L}(H,H)$ et $\mathcal{L}(H, D(A))$ sont continues et indéfiniment dérivables (au sens de $\mathbb{C}$).

Réciproquement si, $\forall\, z \notin S_\alpha$, $zI - A$ est un isomorphisme de $D(A)$ sur H et si de plus, pour tout $z \neq 0$ avec $|\arg z| = \alpha + \frac{\pi}{2}$, $\|(zI - A)^{-1}\|_{H \to H} \leq \frac{1}{|z|}$, alors A est un opérateur mα-accrétif.

Démonstration. Soit v et f liés par la relation $(zI - A)v = f$.

a) On a $(Av, v) \in S_\alpha$. Donc $d(z, S_\alpha)\,|v|^2 \leq |\,z\,|v|^2 - (Av, v)| = |(f, v)| \leq |f|\,|v|$. Il en résulte que $|v| \leq |f|/d(z, S_\alpha)$, d'où la première estimation.

b) On a aussi
$d(z, S_\alpha)\,|Av|^2 \leq |\,z\,|Av|^2 - |z|^2(Av, v)| = |z|\,|(Av, Av - z\,v)| = |z|\,|(Av, f)| \leq |z|\,|Av|\,|f|$.
On en déduit donc $|Av| \leq |z|\,|f|/d(z, S_\alpha)$, d'où la seconde estimation.

c) On remarque que

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

d'où l'on déduit

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

ce qui montre les continuités.

d) On part de $\frac{1}{z_1 - z_2}(R(z_1) - R(z_2)) = -(z_2 I - A)^{-1}(z_1 I - A)^{-1}$. Par passage à la limite lorsque $z_2 - z_1 \to 0$ on en déduit que R est dérivable et que $R'(z_1) = -R(z_1)^2$.

e) Réciproque. On a alors pour tout $u \in D(A)$ et tout $z = |z|\,i\,e^{i\varepsilon\alpha}$, $\varepsilon = \pm 1$,

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

En divisant par $|z|$ on obtient

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

d'où, en faisant tendre $|z|$ vers l'infini, Re $\left(e^{i(\frac{\pi}{2} - \varepsilon\alpha)}(Au, u)\right) \geq 0$, c'est à dire $(Au, u) \in S_\alpha$. □

3. Fonctions d'opérateurs. Semi-groupes analytiques

Soit A un opérateur mα-accrétif sur H, et considérons une fraction rationnelle r bornée sur S_α. On peut donc l'écrire sous la forme

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

avec $\alpha_j \notin S_\alpha$, $m_j \in \mathbb{N}^*$. Il est alors naturel de définir l'opérateur $r(A)$ par

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

On a alors

- $r(A) \in \mathcal{L}(H, H)$, $\qquad (r(A) \in \mathcal{L}(H, D(A))$, si $r(\infty) = 0)$.
- $r(A) + s(A) = (r + s)(A)$, et $r(A)s(A) = (rs)(A)$, si r et s sont deux fractions rationnelles bornées dans S_α.

Nous admettrons le théorème suivant :

Théorème 2. *Soit* $\alpha \in [0, \frac{\pi}{2}]$. *Il existe une constante* $1 \leq C_\alpha \leq 2 + \frac{2}{\sqrt{3}}$ *telle que, pour tout opérateur* A *m*α*-accrétif et toute fraction rationnelle* r *bornée dans le secteur* S_α,

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

De plus, lorsque $\alpha = \frac{\pi}{2}$, *on a* $C_{\pi/2} = 1$.

Corollaire 3. *Lorsque la fonction* f *est une limite uniforme de fractions rationnelles* r_n *dans* S_α, *la relation* $f(A) = \lim_{n\to\infty} r_n(A)$ *définit un opérateur* $f(A) \in \mathcal{L}(H, H)$ *et on a*

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

Démonstration. D'après le théorème précédent on a

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

La suite $r_n(A)$ est de Cauchy et donc convergente dans l'espace complet $\mathcal{L}(H, H)$. Le corollaire s'en déduit aisément (il est en particulier clair que $f(A)$ ne dépend pas du choix effectué pour la suite des r_n). □

Les propriétés suivantes sont de vérification immédiate

$$f(A) + g(A) = (f + g)(A), \qquad f(A)g(A) = (fg)(A) = g(A)f(A). \tag{4}$$

Lemme 4. *Pour tout* $\alpha \in [0, \frac{\pi}{2}[$, *on a*

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

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

On a clairement $|\varphi_n(z)| \leq 2$, ce qui montre le lemme pour $n = 1, 2, 3$.

Remarque 1. On a $|e^{-u} - \frac{1}{1+u}| \leq \frac{3}{2}|u|^2$, pour tout u avec $\operatorname{Re} u \geq 0$.

En effet on a (Taylor) $e^{-u} = 1 - u + \int_0^1 (1-s)e^{-su}\, ds\; u^2$ et $\frac{1}{1+u} = 1 - u + \frac{u^2}{1+u}$, d'où

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

Remarque 2. On a $\forall\, x \in \mathbb{R}$, $1 + x \leq e^x$.

On écrit

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

On déduit des deux remarques, en posant $\rho = \operatorname{Re} z$, (on a alors $\rho \geq |z| \cos\alpha$)

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

Le membre de gauche est maximum lorsque $\rho = \frac{2n}{n-3}$, on a donc

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

$\square$

Corollaire 5. *Soit α et β vérifiant $0 \leq \alpha < \alpha+\beta < \frac{\pi}{2}$. Alors pour tout $t \in S_\beta$ la fonction $E(t) = \exp(-tA)$ est bien définie. De plus $E(t) \in \mathcal{L}(H, H)$ et $\|E(t)\|_{H\to H} \leq 1$.*

Démonstration. Posons $f(z) = e^{-tz}$ et $r_n(z) = (1+\frac{tz}{n})^{-n}$. D'après le lemme précédent on a $\|f - r_n\|_{L^\infty(S_\alpha)} \leq 6/(n\cos^2(\alpha+\beta))$, ce qui permet d'appliquer le corollaire 3 et d'obtenir la majoration $\|E(t)\|_{H\to H} \leq C_\alpha$. Mais on obtient une meilleure majoration en remarquant que $\frac{t}{n}A$ est m(α+β)-accrétif, donc par le théorème 1, $\|(I+\frac{t}{n}A)^{-1}\|_{H\to H} \leq 1/d(-1, S_{\alpha+\beta}) = 1$; par suite $\|r_n(A)\|_{H\to H} \leq 1$ et par passage à la limite $\|E(t)\|_{H\to H} \leq 1$.

$\square$

Remarque. Ce corollaire s'applique en particulier à $t = 0$, et on a $E(0) = I$.

Théorème 6. *Soit $\alpha \in [0, \frac{\pi}{2}[$. La famille d'opérateurs $E(t)$, $t \geq 0$, vérifie les propriétés suivantes*

- $\forall\, t, s \geq 0, \qquad E(t+s) = E(t)\,E(s),$
- $\forall\, t \geq 0, \qquad \|E(t)\|_{H\to H} \leq 1,$
- $\forall\, u_0 \in H$, *l'application $t \mapsto E(t)\,u_0$ est continue de $\mathbb{R}^+$ dans H.*

On dit que cette famille constitue un semi-groupe de contraction fortement continu sur H et que l'opérateur A est le générateur infinitésimal de ce semi-groupe.

Démonstration. a) La première propriété est la propriété dite de semi-groupe (on parlerait de groupe si elle était vraie $\forall\, t, s \in \mathbb{R}$). Elle se déduit de la relation $e^{-(t+s)z} = e^{-tz}e^{-sz}$, voir (4). La deuxième propriété, vue au corollaire précédent, justifie le terme contraction.

b) Montrons la troisième propriété, qui se dit E est fortement continu sur H. Soit $u_0 \in H$ donné. Pour $0 \leq s < t$, on a $|E(t)\,u_0 - E(s)\,u_0| = |E(s)(E(t-s) - E(0))\,u_0| \leq |(E(t-s) - E(0))\,u_0|$. Il suffit donc de montrer la continuité en 0.

1^{er} cas. Si $u_0 \in D(A)$. On pose alors $f = (I+A)u_0$, d'où $u_0 = (I+A)^{-1}f$. Par suite, en utilisant le corollaire 3

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

$$\text{Pour } |z| \geq 1/\sqrt{t}, \text{ on a } |\frac{e^{-tz}-1}{1+z}| \leq \frac{2}{|z|} \leq 2\sqrt{t},$$

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

on en déduit $\qquad \lim_{t\to 0} |(E(t) - E(0))u_0| = 0.$

2^{nd} cas. Si $u_0 \in H$. En utilisant la propriété de contraction, on obtient pour tout $v_0 \in D(A)$

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

La continuité en 0 se déduit du cas précédent joint à la densité de $D(A)$ dans H.

$\square$

Remarque. Le théorème reste valable en fait pour $\alpha = \frac{\pi}{2}$ (il manque une petite étape pour définir $E(t)$ que nous n'aborderons pas ici). Il est aussi valable pour tout t et $s \in S_{\frac{\pi}{2}-\alpha}$.

Théorème 7. *Pour tout $t \in S_\beta$ avec $t \neq 0$ et $0 \leq \alpha < \alpha + \beta < \frac{\pi}{2}$, et pour tout entier $k \geq 0$, l'opérateur $E(t) \in \mathcal{L}(H, D(A^k))$. De plus l'application $t \mapsto E(t)$ de S_β dans $\mathcal{L}(H,H)$ est dérivable (holomorphe), et on a l'estimation*

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

Démonstration. a) Posons $f_k(z) = (1+z)^k\, e^{-tz}$. En procédant comme au lemme 4, f_k est la limite uniforme dans S_α de $(1+z)^k(1+tz/n)^{-n}$. On peut donc définir $f_k(A) \in \mathcal{L}(H,H)$. On déduit de la relation $e^{-tz} = (1+z)^{-k} f_k(z)$ et de $(I+A)^{-k} \in \mathcal{L}(H, D(A^k))$

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

b) Dérivabilité. Posons $K_2 = \sup\{|\zeta^2 e^{-\zeta}|\,; \zeta \in S_{\alpha+\beta}\}$. Pour t et s non nuls $\in S_\beta$ on a par la formule de Taylor

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

En utilisant le corollaire 3 on en déduit

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

d'où l'on déduit la dérivabilité de $E(.)$ et la formule $E'(t) = -A\, E(t)$. Par récurrence on obtient aussi $E^{(k)}(t) = (-A)^k\, E(t)$.

c) Soit $t > 0$; on pose $s(\theta) = t + r\, e^{i\theta}$ avec $r = t \sin\beta$, $\theta \in [0, 2\pi[$. On a donc $s(\theta) \in S_\beta$. Soit maintenant u et v arbitraire dans H, on pose $\varphi(z) = (E(z)u, v)$. Il résulte du paragraphe précédent que la fonction φ est holomorphe dans S_β, de plus on $|\varphi(z)| \leq \|E(z)\|_{H\to H} |u|\, |v| \leq |u|\, |v|$. Il résulte du théorème des résidus

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

Par suite

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

Or $\varphi^{(k)}(t) = (E^{(k)}(t)u, v)$, la majoration étant valable pour tout u et $v \in H$, on en déduit

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

d'où le lemme en faisant tendre β vers $\frac{\pi}{2} - \alpha$. $\square$

Glossaire. La propriété de dérivabilité dans S_β s'énonce : le semi-groupe $E(.)$ est analytique (ou encore holomorphe). Par ailleurs on dit que l'opérateur A engendre le semi-groupe E, ou encore que A est le générateur infinitésimal du semi-groupe E.

Corollaire 8. *Étant donné $u_0 \in H$, le problème suivant*

$$(P) \qquad \begin{cases} \textit{trouver } u \in C^1(]0,\infty[; D(A)) \cap C^0([0,\infty[; H) \textit{ tel que} \\ \forall\, t > 0, \qquad u'(t) + A\,u(t) = 0, \\ \qquad\qquad\qquad u(0) = u_0, \end{cases}$$

admet une solution et une seule, donnée par $u(t) = E(t)\,u_0$, de plus on a, pour tout entier $k \geq 0$, $u \in C^\infty(]0,\infty[; D(A^k))$. Si de plus $u_0 \in D(A^\ell)$, alors on a $u \in C^0([0,\infty[; D(A^\ell)) \cap C^\ell([0,\infty[; H)$ et

$$\forall\, k, \ell \geq 0, \; \forall\, t > 0, \qquad |u^{(k+\ell)}(t)| = |A^k u^{(\ell)}(t)| \leq \frac{k!}{(t\cos\alpha)^k}\,|A^\ell\,u_0|. \tag{5}$$

Démonstration. Il est clair que $u(t)$ est solution de (P) puisque $E'(t) + A\,E(t) = 0$. La régularité C^∞ sur $]0,\infty[$ provient du théorème précédent, la continuité en 0 dans H se déduit de la forte continuité du semi-groupe. Le problème étant linéaire, pour montrer l'unicité de la solution de (P) il suffit de montrer que $u_0 = 0$ entraîne $u(t) = 0$. En effet, si u est solution de (P) et $u_0 = 0$, on a

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

On en déduit $|u(t)|^2 \leq |u(0)|^2 = 0$, d'où l'unicité.

Si maintenant $u_0 \in D(A)$, on vérifie facilement que $(I + \frac{t}{n}A)^{-1}Au_0 = A(I + \frac{t}{n}A)^{-1}u_0$ et par récurrence $(I + \frac{t}{n}A)^{-n}Au_0 = A(I + \frac{t}{n}A)^{-n}u_0$. Par passage à la limite (rappelons que $E(t) = \lim_{n\to\infty}(I + \frac{t}{n}A)^{-n}$) on en déduit $AE(t)u_0 = E(t)A\,u_0$. De même, si $u_0 \in D(A^\ell)$ on a $A^\ell E(t)u_0 = E(t)A^\ell\,u_0$. On remarque alors que $u^{(\ell)}$ est solution du problème (P) avec u_0 remplacé par $(-A)^\ell\,u_0$, d'où $u^{(\ell)} \in C^0([0,\infty[; H)$. La majoration (5) est alors une conséquence du théorème 7. $\square$

Remarque. Effets régularisants et irréversibilité. Nous venons de voir que la solution du problème (P) présente le phénomène de régularisation suivant. En supposant seulement la donnée initiale dans H, alors pour tout $t > 0$, et tout entier k, on a $u(t) \in D(A^k)$ (espaces plus réguliers que H, voir les exemples 1 et 2 du premier paragraphe). Il en résulte que, pour que le problème rétrograde

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

ait une solution, il faudrait avoir $v_T \in D(A^\infty) := \cap_k D(A^k)$. En fait même dans ce cas, on pourrait construire des exemples où il n'y aurait pas de solution sur l'intervalle $(0, T)$. On dit que l'équation (P) est irréversible.

Lorsque l'opérateur A provient d'une forme sesquilinéaire elliptique a, comme celle décrite au paragraphe 1, le problème (P) s'écrit sous la forme équivalente suivante

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

4. Exemple : l'équation de la chaleur. Convection-diffusion

Soit Ω un ouvert borné de $\mathbb{R}^2$. On considère les espaces $H = L^2(\Omega)$ et $V = H_0^1(\Omega)$ ou encore $V = H^1(\Omega)$. On prend

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

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

On fait les hypothèses : ρ et $c \in L^\infty(\Omega\,;\mathbb{R})$, $a \in C^1(\overline{\Omega}\,;\mathbb{R})$, $\vec{b} \in C^1(\overline{\Omega}\,;\mathbb{R}^2)$, $\operatorname{div}\vec{b} = 0$, dans Ω, et

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

Dans le cas du choix $V = H^1(\Omega)$, on supposera de plus $c_0 > 0$ et $\vec{b}.\vec{n} = 0$ sur le bord $\partial\Omega$. Il est clair que $(.\,,.)$ définit bien sur $L^2(\Omega)$ un produit scalaire équivalent au produit scalaire usuel, que la forme sesquilinéaire $a(.\,,.)$ est bien continue sur $V \times V$. Par ailleurs, intégrant par parties, on a (si $v \in V$)

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

chacun des membres de ces équations est donc imaginaire pur. Par suite

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

ce qui montre l'ellipticité de a.

D'après la théorie développée dans les paragraphes précédents, on est bien en présence d'un opérateur A mα- accrétif, et le problème

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

admet une solution unique. Formellement il se traduit dans le problème d'équations aux dérivées partielles (dite paraboliques) suivant, si $V = H_0^1(\Omega)$ (resp. si $V = H^1(\Omega)$),

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

Remarque. Si la donnée initiale u_0 est à valeurs réelles, il est clair que la solution u est aussi à valeurs réelles. En effet u et $\bar{u}$ sont simultanément solutions et la solution est unique.

Le corollaire 8 fournit d'autre part des résultats de régularité, en temps $u \in C^\infty(]0,\infty[;H) \cap C^\ell([0,\infty[;H)$ dès que $u_0 \in D(A^\ell)$, $\ell \ge 0$. On a aussi, sous cette même hypothèse, de la régularité en espace : $u \in C^0([0,\infty[;D(A^\ell)) \cap C^{\ell-k}([0,\infty[;D(A^k))$. Pour mieux traduire ces résultats, on peut utiliser les théorèmes de régularité pour les problèmes elliptiques. Ils montrent que

- Sous les hypothèses de régularité faites précédemment sur les coefficients a, ρ, $\vec{b}$, c et si, soit Ω est convexe, soit sa frontière est de classe C^2, alors pour $V = H_0^1(\Omega)$ (resp. $V = H^1(\Omega)$)

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

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

- Lorsque l'on suppose en outre la frontière de Ω, ainsi que les coefficients a, ρ, $\vec{b}$, c, de classe C^∞, alors pour tout entier $k \geq 1$, on a

$$\begin{array}{rcll} D(A^k) & = & \{v \in H^{2k}(\Omega) \cap H_0^1(\Omega)\,; v = Av = \cdots = A^{k-1}v = 0 \text{ sur } \partial\Omega\}, & \text{si } V = H_0^1(\Omega), \\ D(A^k) & = & \{v \in H^{2k}(\Omega)\,; \frac{\partial v}{\partial n} = \frac{\partial Av}{\partial n} = \cdots = \frac{\partial A^{k-1}v}{\partial n} = 0 \text{ sur } \partial\Omega\}), & \text{si } V = H^1(\Omega). \end{array}$$

5. Approximation en temps : semi-discrétisation en temps

Nous nous intéressons maintenant à l'approximaton du problème. Pour nous ramener à un problème en dimension finie, il est nécessaire de discrétiser simultanément par rapport aux variables d'espace (ce qui sera fait ici par une méthode d'éléments finis) et par rapport à la variable de temps (ou on utilisera plutôt des techniques d'équations différentielles). Cependant chacune de ces discrétisations amène des difficultés bien distinctes, et avant d'examiner la discrétisation totale (la seule réaliste) il est utile pédagogiquement d'examiner séparément les semi-discrétisations en temps et en espace. Nous commençons par la discrétisation temporelle de notre problème (P)

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

qui n'est autre qu'une équation différentielle en dimension infinie. Pour cela nous choisissons un pas de temps $\Delta t > 0$ (destiné à tendre vers zéro), et considérons les instants $t_n = n\Delta t$. (Pour simplifier nous avons choisi un pas de temps constant mais notre approche s'étendrait aussi bien à des pas variables, au prix de quelques difficultés techniques). Nous noterons $u(t_n)$ la solution exacte du problème (P) considéré à l'instant t_n, et u^n la solution approchée obtenu par l'un des schémas ci-après

Bien que ce ne soit pas nécessaire, on choisit tout d'abord $u^0 = u_0$. On obtient ensuite u^{n+1} à partir de u^n en utilisant les formules

- Schéma 1. $\quad \dfrac{u^{n+1} - u^n}{\Delta t} + A\,u^n = 0, \qquad$ méthode d'Euler (explicite),
- Schéma 2. $\quad \dfrac{u^{n+1} - u^n}{\Delta t} + A\,u^{n+1} = 0, \qquad$ méthode d'Euler implicite,
- Schéma 3. $\quad \dfrac{u^{n+1} - u^n}{\Delta t} + A\,\dfrac{u^n + u^{n+1}}{2} = 0, \qquad$ schéma de Crank-Nicolson,
- Schéma 4. $\quad u^{n+1} = r(\Delta t\,A)\,u^n, \qquad$ schémas de Runge-Kutta,

(pour le dernier schéma, r est une fraction rationnelle). Notons que le schéma 4 englobe les trois précédents : il suffit de prendre $r(z) = 1 - z$ pour retrouver le schéma d'Euler, $r(z) = (1+z)^{-1}$ pour obtenir Euler implicite, le schéma de Crank-Nicolson correspond à $r(z) = \frac{1-z/2}{1+z/2}$. Le schéma 2 est dit implicite car le calcul de u^{n+1} à partir de u^n nécessite la résolution d'un système linéaire, le schéma 3 est aussi implicite.

Étude du schéma d'Euler-explicite. Pour pouvoir calculer u^n il faut que $u_0 \in D(A^n)$, on a alors $u^n = (I - \Delta t\, A)^n\, u_0$. Cette hypothèse n'est pas en général vérifiée pour tout n (or on a besoin de grandes valeurs de n si on veut pouvoir dépasser un instant $\tau > 0$ donné, avec un pas de temps Δt nécessairement petit pour obtenir une bonne précision). Ce schéma (contrairement à ce qui se passe pour les équations différentielles ordinaires) est donc ici généralement inutilisable. Bien plus, pour des raisons de stabilité qui seront évoquées par la suite, même dans le cas où l'on aurait $u_0 \in D(A^\infty)$ il ne serait toujours pas réellement opérant.

Étude du schéma d'Euler-implicite. Notons tout d'abord que ce schéma s'écrit de manière équivalente sous la forme $(I + \Delta t\, A)u^{n+1} = u^n$. L'opérateur A et mα-accrétif, il en est donc de même de l'opérateur $\Delta t\, A$. Il résulte du théorème 1 que $(I + \Delta t\, A)^{-1}$ est bien défini, appartient à $\mathcal{L}(H, H) \cap \mathcal{L}(H, D(A))$ et de plus $\|(I + \Delta t\, A)^{-1}\|_{H \to H} \leq 1/d(-1, S_\alpha) = 1$. La solution du schéma 2 est donc donnée par

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

Propriété de stabilité. De plus, si v^n désigne la solution de ce schéma pour la donnée initiale $v^0 = v_0$ on a la propriété de stabilité

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

l'erreur initiale n'est pas amplifiée par le schéma.

Propriété régularisante. En supposant seulement $u_0 \in H$, on a ensuite $u^1 \in D(A)$, $u^2 \in D(A^2)$,..., $u^n \in D(A^n)$.

Propriétés de convergence.

a) Cas d'une donnée peu régulière, on suppose seulement $u_0 \in H$, alors on a

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

Démonstration. On remarque que

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

On applique le corollaire 3 à l'opérateur mα-accrétif $t_n\, A$ et ensuite le lemme 4 d'où l'on déduit

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

pour obtenir (6) il suffit de remarquer que $n = t_n/\Delta t$.

□

Remarque. Quand on regarde une estimation d'erreur comme celle donnée par (6), on doit prendre l'habitude de considérer l'instant d'observation t_n comme fixé ; alors n tend vers l'infini quand Δt tend vers zéro. Nous voyons que l'estimation d'erreur est plus mauvaise pour t_n petit et s'améliore ensuite.

b) Cas d'une donnée régulière : $u_0 \in D(A)$. Dans ce cas on a

$$\forall\, n \geq 2, \qquad |u(t_n) - u^n| \leq C_\alpha \frac{3\Delta t}{\cos\alpha}\ |Au_0|. \tag{7}$$

Démonstration. On pose

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

on a donc $\Delta t\, A\varphi_n(\Delta t\, A) = E(t_n) - (I + \Delta t\, A)^{-n}$, d'où $u(t_n) - u^n = \Delta t\, \varphi_n(\Delta t\, A) A u_0$. D'après le corollaire 3, il suffit de montrer la majoration $|\varphi_n(z)| \leq 3/\cos\alpha$, pour tout $z \in S_\alpha$. Pour cela on remarque que

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

On utilise alors les deux remarques du lemme 4, on pose $\rho = \operatorname{Re} z$, (donc $\rho \geq |z|\cos\alpha$) et on obtient

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

en effet le deuxième membre est maximum lorsque $\rho = 1/(n-2)$. □

Remarque. Nous voyons que, lorsque la donnée initiale est régulière, l'estimation d'erreur est uniformément en $0(\Delta t)$. Nous n'avons pas traité le cas $n = 1$ qui est de peu d'intérêt ; mais on pourrait montrer directement que $\varphi_1(z) = \int_0^1 e^{-sz}\, ds - \frac{1}{1+z}$, d'où $|u(t_1) - u^1| \leq 2\,\Delta t\, |Au_0|$.

Étude du schéma de Crank-Nicolson. Il s'écrit

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

Propriété de stabilité. On a encore (avec les mêmes notations que pour le schéma précédent) la propriété de stabilité

$$\|r(\Delta t\, A)^n\|_{H\to H} \leq 1, \qquad \text{d'où} \quad \forall\, n \geq 0, \quad |u^n - v^n| \leq |u_0 - v_0|.$$

Démonstration. Il suffit de voir que $\|r(\Delta t\, A)\|_{H\to H} \leq 1$. En effet, soit $w = r(\Delta t\, A)v$ et posons $\varphi = (I + \frac{\Delta t}{2}\, A)^{-1} v$. On a

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

d'où $|w| \leq |v|$. □

Contrairement au schéma précédent, il n'y a pas cette fois de phénomène de régularisation, ni d'ailleurs celui de perte de régularité comme lors de la méthode d'Euler explicite. On vérifie aisément que si $u_0 \in D(A^\ell)$ alors $u^n \in D(A^\ell)$, mais pas mieux.

Propriétés de convergence.

a) Cas d'une donnée régulière : $u_0 \in D(A^2)$. Dans ce cas on a

$$\forall\, n \geq 2, \qquad |u(t_n) - u^n| \leq C_\alpha \, \max\big(2, \frac{5}{6\, e\, \cos\alpha}\big)\, \Delta t^2 \; |A^2 u_0|. \tag{8}$$

Démonstration. On pose

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

Compte tenu du corollaire 3, il suffit de montrer que $|\varphi_n(z)| \leq \max\big(2, \frac{5}{6\,e\,\cos\alpha}\big)$ pour $z \in S_\alpha$.

Si $z \in S_\alpha$ et $|z| \geq 1$. On a alors de manière évidente $|\varphi_n(z)| \leq 2/|z|^2 \leq 2$.

Si $z \in S_\alpha$ et $|z| \leq 1$. On écrit

$$\varphi_n(z) = \frac{e^{-z} - r(z)}{z^3} \; z \; \sum_{k=0}^{n-1} \; e^{-kz} \; r(z)^{n-1-k}. \tag{9}$$

On remarque d'abord que (formule de Taylor)

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

d'où

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

En posant $z = \rho e^{i\theta}$, on a $|\theta| \leq \alpha$ et

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

Or on vérifie facilement que, pour $x \in [0,1]$, on a $\frac{1-x/2}{1+x/2} \leq e^{-x}$. On en déduit la majoration

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

En revenant à (9) et en utilisant que pour $x > 0$, $e^{-x} \leq \frac{1}{e\,x}$, on obtient

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

□

b) Cas d'une donnée peu régulière, $u_0 \in H$. La méthode utilisée pour Euler implicite ne permet pas ici d'obtenir une majoration intéressante : on a

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

car $|r(\infty)| = 1$. Cependant en utilisant la densité de $D(A^2)$ dans H et $\|r(\Delta t\, A)^n\|_{H\to H} \leq 1$, on déduit facilement du a) que

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

Exercice. Montrer que si $u_0 \in D(A)$, on a $\forall\, n \geq 2, \quad |u(t_n) - u^n| \leq K C_\alpha\, \Delta t\, |A u_0|$.

Schéma général. Les méthodes à un pas générales (comme par exemple les méthodes de Runge-Kutta) appliquées au problème différentiel linéaire homogène (P) se réduisent (généralement) dans ce cas simple à un schéma de la forme

$$u^0 = u_0, \quad \text{et, pour } n \geq 0, \qquad u^{n+1} = r(\Delta t\, A)\, u^n, \tag{10}$$

où r est une fraction rationnelle. Nous supposons que l'opérateur A est mα-accrétif, le schéma est alors bien défini dès que la fraction rationnelle est bornée dans S_α, ce que nous supposerons donc. Pour analyser ces méthodes nous utiliserons les définitions suivantes

Définitions. On dit que la fraction rationnelle $r(.)$

- est une approximation d'ordre p si $\quad r(z) = e^{-z} + 0(|z|^{p+1})$ au voisinage de $z = 0$,
- est $A(\alpha)$-acceptable si $\quad |r(z)| \leq 1$ pour tout $z \in S_\alpha$,
- est fortement $A(\alpha)$-acceptable si $\quad |r(z)| < 1$ pour tout $z \in S_\alpha$, $z \neq 0$ et si $|r(\infty)| < 1$.

Lorsque $\alpha = \frac{\pi}{2}$, on dit A-acceptable au lieu de $A(\frac{\pi}{2})$-acceptable.

Propriété de stabilité. Lorsque la fraction rationnelle r est $A(\alpha)$-acceptable on a la propriété de stabilité

$$\forall\, n \geq 0, \quad \|r(\Delta t\, A)^n\|_{H \to H} \leq C_\alpha, \qquad \text{d'où} \quad |u^n - v^n| \leq C_\alpha |u_0 - v_0|.$$

C'est en effet une conséquence immédiate du corollaire 3. C'est bien un résultat de stabilité : l'erreur initiale est amplifiée au plus par le facteur C_α.

Remarque. Lorsque la fraction rationnelle est A-acceptable, ce qui était le cas dans les deux schémas précédents, on peut remplacer de plus C_α par $C_{\frac{\pi}{2}} = 1$, en effet si l'opérateur A est mα- accrétif, il est a fortiori m-accrétif.

Propriété régularisante. Lorsque $r(\infty) = 0$, en supposant $u_0 \in D(A^\ell)$, on a ensuite $u^1 \in D(A^{\ell+1})$, $u^2 \in D(A^{\ell+2})$,..., $u^n \in D(A^{\ell+n})$. Si $r(\infty) \neq 0$ il n'y a pas de phénomène de régularisation, mais, si la fraction rationnelle est $A(\alpha)$-acceptable, on a $u_0 \in D(A^\ell)$ entraîne $u^n \in D(A^\ell)$.

Propriétés de convergence.

Théorème 9. *On fait les hypothèses suivantes :*

- *l'opérateur A est mα-accrétif,* $\quad 0 \leq \alpha < \frac{\pi}{2}$,
- *r est une approximation d'ordre p, et r est $A(\alpha)$-acceptable.*

Alors il existe une constante K telle que, $\forall\, u_0 \in D(A^p)$, la solution du schéma (10) vérifie

$$\forall\, n \geq 2, \qquad |u(t_n) - u^n| \leq \frac{K}{\cos\alpha}\, \Delta t^p\; |A^p u_0|. \tag{11}$$

Si de plus la fraction rationnelle r est fortement $A(\alpha)$-acceptable, alors on a, pour tout $u_0 \in H$,

$$\forall\, n \geq 2, \qquad |u(t_n) - u^n| \leq \frac{K'}{t_n^p\, (\cos\alpha)^{p+1}}\, \Delta t^p\; |u_0|. \tag{12}$$

Démonstration. a) Cas d'une donnée régulière.

On pose

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

Compte tenu du corollaire 3, il suffit de montrer que $C_\alpha\,|\varphi_n(z)| \le \frac{K_r}{\cos\alpha}$ pour $z \in S_\alpha$. Pour cela on remarque d'abord qu'il existe un réel $\gamma > 0$ tel que

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

en effet $e^{\operatorname{Re} z/2}|r(z)| = e^{-\operatorname{Re} z/2} + 0(|z|^{p+1})$, d'où $\quad e^{\operatorname{Re} z/2}|r(z)| \le 1 - |z|\cos\alpha/2 + C|z|^{p+1}$.

Si $z \in S_\alpha$ et $|z| \ge \gamma\,(\cos\alpha)^{1/p}$. On a alors $|\varphi_n(z)| \le 2/|z|^p \le \frac{2}{\gamma^p \cos\alpha}$.

Si $z \in S_\alpha$ et $|z| \le \gamma\,(\cos\alpha)^{1/p}$. On écrit

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

Puisque la méthode est d'ordre p on a

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

Il en résulte (rappelons que pour $x \ge 0$, $e^{-x} \le \frac{1}{e\,x}$)

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

Cela montre le résultat avec $K = C_\alpha\ \max(\frac{2}{\gamma^p}, \frac{4\,K_1}{e})$.

b) Cas d'une donnée peu régulière. On pose

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

Il suffit de montrer que $C_\alpha\,|\varphi_n(z)| \le \frac{K'}{n^p\,(\cos\alpha)^{p+1}}$ pour $z \in S_\alpha$.

Si $z \in S_\alpha$ et $|z| \le \gamma\,(\cos\alpha)^{1/p}$. On écrit

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

on a donc, en remarquant que pour $x \ge 0$, $e^{-x} \le \big(\frac{p+1}{e\,x}\big)^{p+1}$,

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

Si $z \in S_\alpha$ et $|z| \ge \gamma\,(\cos\alpha)^{1/p}$. On remarque d'abord qu'il existe $\delta > 0$ tel que $|r(z)| \le e^{-\delta(\cos\alpha)^{1/p}}$ pour toutes ces valeurs de z. On a alors

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

On prend alors $K' = C_\alpha \max(K_1\,\big(\frac{4(p+1)}{e}\big)^{p+1}, (\frac{p}{e\,\gamma})^p + (\frac{p}{e\,\delta})^p\,\big)$.

$\square$

Remarques. a) Si on suppose seulement $u_0 \in H$, la fraction rationnelle $A(\alpha)$-acceptable et d'ordre ≥ 1, on peut montrer, comme pour Crank-Nicolson,

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

b) Si on suppose seulement $u_0 \in D(A^{p-k})$, $0 \leq k \leq p$, la fraction rationnelle fortement $A(\alpha)$-acceptable et d'ordre p, il n'est pas très difficile de modifier la démonstration du b) pour obtenir le résultat intermédiaire

$$\forall\, n \geq 2, \qquad |u(t_n) - u^n| \leq \frac{K}{t_n^k\,(\cos\alpha)^{k+1}}\,\Delta t^p\; |A^{p-k} u_0|. \tag{13}$$

6. Intégrales hilbertiennes. Intégrales d'opérateurs

Soit Γ une courbe simple orientée du plan complexe. On considère les applications

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

- On dira que l'intégrale $\int_\Gamma f(z)\,dz$ est bien définie si pour tout $v \in H$ l'intégrale $\int_\Gamma (f(z), v)\,dz$ est bien définie et s'il existe un élément $I_1 \in H$ tel que

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

On écrira alors $I_1 = \int_\Gamma f(z)\,dz$.

- On dira que l'intégrale $\int_\Gamma B(z)\,dz$ est bien définie si pour tout u et $v \in H$ l'intégrale $\int_\Gamma (B(z)u, v)\,dz$ est bien définie et s'il existe un élément $I_2 \in \mathcal{L}(H,H)$ tel que

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

On écrira alors $I_2 = \int_\Gamma B(z)\,dz$.

Une condition suffisante pour que l'intégrale I_1 soit bien définie est que la fonction f soit continue sur Γ à valeurs dans H et que $\int_\Gamma |f(z)|\,|dz| < +\infty$. De même une condition suffisante pour que l'intégrale I_2 soit bien définie est que la fonction B soit continue sur Γ à valeurs dans $\mathcal{L}(H,H)$ et que $\int_\Gamma \|B(z)\|_{H \to H}\,|dz| < +\infty$.

Remarque. Lorsque dans un voisinage de la courbe Γ, pour tout $u,\, v \in H$, les applications $z \mapsto (f(z), v)$ et $z \mapsto (B(z)u, v)$ sont holomorphes, alors on a

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

pour toute déformation homotope Γ' de Γ restant dans ce voisinage et conservant les mêmes extrémités.

Lemme 10. *Soit $A \in \mathcal{L}(D(A), H)$ un opérateur $m\alpha$-accrétif, $\alpha \in [0, \frac{\pi}{2}]$; on suppose de plus que A est un isomorphisme de $D(A)$ sur H ; soit $\beta \in]0, \pi - \alpha[$. Alors*

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

où $\Gamma_{\alpha+\beta}$ désigne la frontière orientée (dans le sens contraire des aiguilles d'une montre) du secteur $S_{\alpha+\beta}$.

Démonstration. Posons

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

Montrons d'abord que cette intégrale est bien définie. Pour cela on déduit du théorème 1 la majoration $\|(z\, I - A)^{-1}\|_{H \to H} \le \frac{1}{d(z, S_\alpha)} = \frac{1}{|z| \sin\beta}$, ce qui montre la convergence de l'intégrale pour $|z|$ grand. D'autre part il est facile de montrer que

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

ce qui assure la convergence de l'intégrale au voisinage de 0.

On note maintenant Γ_M la portion formée des $z \in \Gamma_{\alpha+\beta}$ avec $|z| \le M$, et on pose $C_M = \{z = M\, e^{i\theta}\, ; \alpha + \beta \le \theta \le 2\pi - \alpha - \beta\}$, C_M sera orienté dans le sens des θ décroissants. Remarquons que, pour $M > |\lambda|$, par homotopie, l'intégrale $\int_{\Gamma_M \cup C_M} \frac{1}{\lambda - z} (z\, I - A)^{-1}\ dz$ ne dépend pas de M, de plus

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

on en déduit

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

où $\gamma(z, \varepsilon)$ est le cercle de centre z et de rayon $\varepsilon > 0$ (assez petit). On obtient, par passage à la limite quand $\varepsilon \to 0$, $J = 2\pi i (\lambda \sigma \tau\, I - A)^{-1}$, ce qui montre le lemme. $\square$

Théorème 11. *Soit $A \in \mathcal{L}(D(A), H)$ un opérateur $m\alpha$-accrétif, $\alpha \in [0, \frac{\pi}{2}]$; on suppose que A est un isomorphisme de $D(A)$ sur H ; soit $\beta \in]0, \pi - \alpha[$. Soit f une fonction holomorphe à l'intérieur de $S_{\alpha+\beta}$, continue dans $S_{\alpha+\beta}$, et telle que $f(z) = 0(z^{-\varepsilon})$ quand $|z| \to \infty$ pour un $\varepsilon > 0$. Alors l'opérateur $f(A) \in \mathcal{L}(H, H)$ est bien défini, de plus*

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

Démonstration. Le lemme précédent montre le résultat lorsque f est une fraction rationnelle ayant ses pôles simples et appartenant au complémentaire de $S_{\alpha+\beta}$. Donc en utilisant le corollaire 3 il est vrai pour toute limite, uniforme dans S_α, de telles fractions. Soit maintenant f une fonction satisfaisant aux hypothèses du théorème, elle vérifie la formule de Cauchy à l'intérieur de $S_{\alpha+\beta}$

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

on a ainsi $f_M \to f$ uniformément dans S_α quand $M \to \infty$. En approchant dans f_M l'intégrale par la formule des rectangles

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

on obtient que f est limite uniforme dans S_α de fractions rationnelles r_n à pôles simples sur $\Gamma_{\alpha+\beta}$. En remarquant enfin que $f(.)$ est la limite uniforme de $f(.+\varepsilon)$ quand $\varepsilon > 0 \to 0$, on en déduit que f est limite uniforme dans S_α des fractions rationnelles $r_n(.-\varepsilon_n)$ à pôles simples sur $\Gamma_{\alpha+\beta} - \varepsilon_n$, c'est à dire à l'extérieur de $S_{\alpha+\beta}$. □

Remarque 1. En particulier ce théorème montre que, pour $0 \leq \alpha < \alpha + \beta < \frac{\pi}{2}$,

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

Remarque 2. Si on rajoute l'hypothèse que f est holomorphe au voisinage de 0, le théorème reste vrai sans l'hypothèse d'isomorphisme pour A, à condition de déformer $\Gamma_{\alpha+\beta}$ (dans $S^c_{\alpha+\beta}$) en un petit arc de cercle contournant 0.

7. Approximation en espace. Semi-discrétisation par éléments finis

On revient maintenant au problème (P) écrit sous sa forme

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

Pour l'approcher on utilise une méthode interne définie par un sous-espace de dimension finie $V_h \subset V$ (on pourra par exemple choisir un espace d'éléments finis conformes) et on considère le problème

$$(P_h) \qquad \begin{cases} \text{trouver} \quad u_h \in C^1([0,\infty[; V_h) \text{ tel que} \\ \forall\, t > 0, \forall\, w_h \in V_h, \quad (u_h'(t), w_h) + a(u_h(t), w_h) = 0, \\ \qquad u_h(0) = u_{0h}, \quad \text{donné dans } V_h. \end{cases}$$

L'introduction de l'opérateur $A_h \in \mathcal{L}(V_h, V_h)$ défini par

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

permet d'écrire (P_h) sous la forme équivalente

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

Ce problème admet une solution unique (c'est un système différentiel linéaire homogène en dimension finie). Nous l'écrirons sous la forme

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

en posant

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

Notons que, pour exactement les mêmes raisons que pour A, et pour la même valeur $\alpha =$Arccos (a_0/M), l'opérateur A_h est mα-accrétif sur l'espace V_h muni du produit scalaire de H. Puisqu'on est en dimension finie, l'opérateur A_h est nécessairement borné $D(A_h) = V_h$.

Écriture matricielle. L'opérateur A_h est un opérateur abstrait, dont l'utilisation est agréable pour l'analyse numérique, mais il n'est pas directement adapté aux calculs effectifs. Pour ceux-là on a besoin d'introduire une base $\{\psi_j\}_{j=1}^N$ de V_h. On pose alors

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

La matrice M s'appelle la matrice de masse, et K la matrice de rigidité. Si on note $U(t) = (u_1(t), u_2(t), \ldots, u_N(t))^T$ le vecteur colonne des composantes de $u_h(t)$ dans la base des ψ_j, le problème (P_h) équivaut au système différentiel

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

C'est ce système qui est directement utilisé pour les calculs. La matrice $M^{-1}K$ est la matrice qui représente l'opérateur A_h dans la base $\{\psi_j\}_{j=1}^N$.

Estimations d'erreur. Pour faire l'étude de l'erreur nous utiliserons les opérateurs suivants

- P_h opérateur de projection orthogonale de H sur V_h, (il est défini en fait de V' sur V_h) par $\quad P_h\varphi \in V_h$ et $\forall\, w_h \in V_h$, $(P_h\varphi, w_h) = (\varphi, w_h)$.
- $T = A^{-1} \in \mathcal{L}(V', V)$, $\quad T_h = A_h^{-1} P_h \in \mathcal{L}(V', V_h)$, $\quad Q_h = A_h^{-1} P_h A \in \mathcal{L}(V, V_h)$.

Soit $f \in V'$ donné et posons $u = T\, f$ et $u_h = T_h\, f = Q_h\, u$, alors on a

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

La majoration d'erreur entre u et u_h est l'objet de la théorie d'approximation des problèmes elliptiques par la méthode des éléments finis. On a en particulier

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

d'où, en utilisant le lemme de Céa et la technique de dualité de Aubin-Nitsche,

$$\begin{aligned} &\|u - u_h\| \le \tfrac{M}{a_0}\ \inf_{v_h \in V_h} \|u - v_h\|, \\ &|u - u_h| \le M\, \|u - u_h\|\ \sup\{\inf_{w_h \in V_h} \|T^* g - w_h\|\ ; g \in H, |g| = 1\}. \end{aligned} \tag{14}$$

Nous ferons dans ce qui suit l'hypothèse

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

Compte tenu de (14) on déduit de cette hypothèse la propriété suivante

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

Remarque. Si on se place dans le cadre de l'exemple décrit paragraphe 4, en supposant Ω polygonal convexe, $(\mathcal{V}_1)$ se déduit des estimations

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

qui sont des résultats typiques pour les problèmes elliptiques.

L'opérateur Q_h est appelé : opérateur de projection elliptique sur V_h.

Sans hypothèse de régularité sur la donnée initiale, on a le résultat suivant

Théorème 12. *Sous les hypothèses $(\mathcal{H}_1)$ et $u_0 \in H$, on a l'estimation d'erreur*

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

De plus

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

Notons que, avec le choix $u_{0h} = P_h u_0$, cela donne une erreur en $0(h^2/t)$.

Démonstration. a) Si $u_{0h} = P_h u_0$. On écrit (avec $0 \leq \alpha < \alpha + \beta < \frac{\pi}{2}$)

$$\begin{aligned} u(t) - u_h(t) &= E(t)\,u_0 - E_h(t)\,P_h u_0 \\ &= \frac{1}{2\pi i} \int_{\Gamma_{\alpha+\beta}} e^{-tz} \big((zI - A)^{-1} u_0 - (zI - A_h)^{-1} P_h u_0\big)\,dz. \end{aligned} \tag{15}$$

Or

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

D'après le théorème 1 (pour $z \in \Gamma_{\alpha+\beta}$)

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

On a la même majoration pour $\|(zT_h - I)^{-1}\|_{H \to H}$, donc par l'hypothèse $(\mathcal{H}_1)$, en reportant dans la formule (15),

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

En prenant $\beta = \frac{\pi}{4} - \frac{\alpha}{2}$, et en remarquant que $2\sin\beta \geq \sin 2\beta = \cos\alpha$, on en déduit

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

b) Si $u_{0h} \neq P_h u_0$, on utilise la relation $u_h(t) = E_h(t)\,P_h u_0 + E_h(t)(u_{0h} - P_h u_0)$, et la majoration $|E_h(t)(u_{0h} - P_h u_0)| \leq |u_{0h} - P_h u_0|$.

c) Pour la majoration concernant les dérivées $k^{\text{ièmes}}$, on écrit

$$t^k\,(u^{(k)}(t) - u_h^{(k)}(t)) = \frac{1}{2\pi i}\int_{\Gamma_{\alpha+\beta}} (-tz)^k\, e^{-tz}\left((zI-A)^{-1}u_0 - (zI-A_h)^{-1}P_h u_0\right) dz \\ + t^k\, E^{(k)}(t)\,(u_{0h} - P_h u_0).$$

Le premier terme du membre de droite se majore comme au a). Pour le second on utilise la majoration (5) du corollaire 8. □

Regardons maintenant le cas d'une donnée initiale régulière : $u_0 \in D(A)$.

Théorème 13. *Sous les hypothèses $u_0 \in D(A)$ et $(\mathcal{H}_1)$ on a*

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

Démonstration. a) L'hypothèse $(\mathcal{H}_1)$ entraîne

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

de même en utilisant la majoration (5) du corollaire 8

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

On en déduit en particulier les majorations

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

Posons maintenant

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

Compte tenu des majorations précédentes, pour obtenir le théorème, il suffit de montrer que

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

b) Puisque u et u_h sont solutions de (P), on a

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

La solution de cette équation différentielle linéaire avec second membre est donnée par le principe de Duhamel (encore appelé méthode de variations de la constante)

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

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

On a clairement $|E_1| \leq |e_h(0)|$ et

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

Pour le troisième terme on intègre d'abord par parties

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

d'où

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

D'après le corollaire 8 on a la majoration

$$\|A_h\,E_h(t-\sigma)\|_{H\to H} \leq \frac{1}{(t-\sigma)\cos\alpha} \qquad \text{d'où} \qquad |E_3| \leq C\,h^2\,(4 + \frac{2\log 2}{\cos\alpha})\,|A\,u_0|,$$

ce qui achève la démonstration du théorème.

□

8. Semi-discrétisation en espace : méthodes d'ordres plus élevés

Dans ce paragraphe nous aurons besoin d'utiliser les puissances non entières de l'opérateur, supposé mα-accrétif, A, $0 \leq \alpha \leq \frac{\pi}{2}$. Pour cela on introduit les fonctions suivantes

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

où $z \in S_\alpha$, $0 \leq s < 1$. On choisira la détermination principale pour les puissances (c'est à dire détermination continue et réelle pour z réel).

Définitions. Pour $0 \leq s < 1$ et $k \in \mathbb{N}$

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

On a les propriétés suivantes

Théorème 14. *a) Si $0 \leq s \leq s+s'$, $\quad D(A^{s+s'}) \subset D(A^s) \cap D(A^{s'})$.*

De plus, si $u \in D(A^{s+s'})$, alors $A^{s'}\,u \in D(A^s)$ et $A^{s+s'}\,u = A^s\,(A^{s'}\,u)$.

b) Si $u_0 \in D(A^s)$, alors, $\forall\, t \geq 0$, $E(t)\,u_0 \in D(A^s)$ et on a $A^s\,E(t)\,u_0 = E(t)\,A^s\,u_0$.

c) Pour $s \in [0,1]$, l'opérateur A^s est m-sα-accrétif.

d) Lorsque A provient d'une forme hermitienne, continue et V-elliptique, c'est à dire lorsque

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

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

Démonstration. a) Il suffit de regarder le cas $s+s' \leq 1$, ce que nous supposons donc. On remarque que l'on a $\psi_{s+s'}(z) = \psi_s(z)\,\psi_{s'}(z)$, et $\varphi_{s+s'}(z) = \varphi_s(z)\,\varphi_{s'}(z)$ et que ces fonctions vérifient les hypothèses du corollaire 3. La condition $u \in D(A^{s+s'})$ entraîne (avec $v \in H$)

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

on en déduit $D(A^{s+s'}) \subset D(A^s) \cap D(A^{s'})$. De plus

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

ce qui montre que $A^{s'}u \in D(A^s)$ et

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

b) Cela résulte directement de la relation $E(t)\,\psi_s(A) = \psi_s(A)\,E(t)$.

c) Soit $\lambda \notin S_{s\alpha}$. Posons $f(z) = \frac{1}{\lambda - z^s}$, la fonction f vérifie les hypothèses du corollaire 3 ; de plus $f(\lambda\psi_s - \varphi_s) = \psi_s$. Soit maintenant $u \in D(A)^s$, donc $u = \psi_s(A)v$, avec $v \in H$. On a

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

Cela montre que $\lambda I - A^s$ est un isomorphisme de $D(A^s)$ sur H et que $f(A) = (\lambda I - A^s)^{-1}$. D'après la réciproque du théorème 1, pour montrer le caractère $s\alpha$-accrétif, il suffit de montrer que $\|f(A)\|_{H\to H} \leq |\lambda|^{-1}$ lorsque $\lambda = \rho\, e^{\pm i(s\alpha+\pi/2)}$. Pour cela on regarde les deux cas possibles

- $\lambda = \rho\, e^{i(s\alpha+\pi/2)}$. On pose $B = e^{i(\frac{\pi}{2}-\alpha)}\,A$ et $g(z) = f(e^{-i(\frac{\pi}{2}-\alpha)}z)$. Il est clair que $f(A) = g(B)$ et que l'opérateur B est m-accrétif. On a donc

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

- $\lambda = \rho\, e^{-i(s\alpha+\pi/2)}$. On raisonne comme au cas précédent en prenant maintenant $B = e^{-i(\frac{\pi}{2}-\alpha)}\,A$ et $g(z) = f(e^{i(\frac{\pi}{2}-\alpha)}z)$.

d) Cela résulte directement de la relation $(Av, v) = |A^{1/2}v|^2$.

$\square$

Nous introduisons maintenant l'hypothèse, pour k entier strictement positif,

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

Remarque. Si on se place dans le cadre de l'exemple décrit au paragraphe 4, il est assez facile de montrer que $D(A^{1/2}) = V$ (voir le lemme 18 plus loin) ; de plus les théorèmes de régularité pour les problèmes elliptiques montrent que, lorsque la frontière de Ω est assez régulière ainsi que les différents coefficients de la forme sesquilinéaire a, alors $D(A^{k/2}) \subset H^k(\Omega)$, même pour les valeurs impaires de k. L'hypothèse $(\mathcal{H}_k)$ paraît alors à première vue assez naturelle (pour des éléments finis basés sur des polynômes de degré k). C'est effectivement le cas lorsque

l'on est en dimension 1,

ou encore en dimension 2 sur un rectangle avec des conditions périodiques,

cela peut être le cas pour un ouvert Ω polygonal. Pour $k > 1$ les théorèmes de régularité ne sont plus vrais directement (hormis dans quelques cas très particuliers via des techniques de réflexion) mais seulement dans des espaces avec poids. Des techniques de raffinement de maillage permettent de conserver $(\mathcal{H}_k)$. Une autre possibilité serait l'introduction de fonctions de singularités dans l'espace d'éléments finis.

lorsque Ω a une frontière régulière, pour vérifier l'hypothèse on est conduit à utiliser des éléments finis courbes. Il est possible de modifier légérement le formalisme des éléments isoparamétriques pour garder $\Omega_h = \Omega$, $V_h \subset V$, et vérifier $(\mathcal{H}_k)$. Nous ne rentrerons pas dans ces détails.

Théorème 15. *Soit $k \geq 1$ un entier. Sous les hypothèses $u_0 \in D(A^{(k+1)/2})$ et $(\mathcal{H}_k)$ on a,*

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

Démonstration. Il suffit de reprendre la démonstration du théorème 13 en utilisant les estimations

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

et

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

□

Pour étudier le cas d'une donnée peu régulière nous aurons besoin d'hypothèses plus précises sur la méthode d'approximation. On introduit pour cela un nouveau jeu d'hypothèses dépendant de la parité de l'indice

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

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

Lemme 16. *Soit $k \geq 1$ un entier.*

Les conditions $(\mathcal{V}_k)$ et $(\mathcal{V}_1)$ entraînent $(\mathcal{H}_k)$,

Les conditions $(\mathcal{V}_1)$ et $(\mathcal{V}_2)$ entraînent $\forall\, f \in H, \quad \|T_h(T - T_h) f\| \leq C h^3 |f|,$

Les conditions $(\mathcal{V}_1)$ et $(\mathcal{V}_3)$ entraînent $\forall\, f \in H, \quad |T_h(T - T_h) f| \leq C h^4 |f|.$

Démonstration. Posons $u = T\, f$ et $u_h = T_h\, f$, alors on a

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

Puisque $V_h \subset V$ on a, $\forall\, w_h \in V_h$, $a(u-u_h, w_h) = 0$. Les estimations (14) sont encore valables

$$\|u-u_h\| \leq \tfrac{M}{a_0}\ \inf_{v_h\in V_h} \|u-v_h\|,$$

$$|u-u_h| \leq M\,\|u-u_h\|\ \sup\{\inf_{w_h\in V_h} \|T^*g-w_h\|\ ; g\in H, |g|=1\}.$$

a) Pour $Tf - T_hf = u-u_h$, on déduit de $(\mathcal{V}_1)$ et de $(\mathcal{V}_k)$

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

b) On remarque d'abord que $T_h(T-T_h)f = (T_h-T)(T-T_h)f + T(T-T_h)f$. On majore le premier terme par $(\mathcal{V}_1)$ et $(\mathcal{H}_1)$

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

Pour le deuxième terme $\varphi = T(T-T_h)f = T(u-u_h)$ on écrit

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

On utilise ensuite que $\|u-u_h\| \leq C\,h\,|f|$ et $\inf_{w_h\in V_h}\|T^*\varphi-w_h\| \leq C\,h^2\|\varphi\|$.

c) On procède de même, en utilisant les majorations

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

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

ensuite $\|u-u_h\| \leq C\,h\,|f|$, et enfin $\inf_{w_h\in V_h}\|T^*T^*\varphi-w_h\| \leq C\,h^3|A^*T^*\varphi| = C\,h^3|\varphi|$.

□

Théorème 17. *On fait les hypothèses $u_0\in H$, $(\mathcal{V}_1)$, $(\mathcal{V}_k)$, $k=1,2$ ou 3. Lorsque $k=2$ on suppose en outre*

$$\forall\, v_h\in V_h,\quad |A_h^{1/2}v_h| \leq K\|v_h\|. \tag{16}$$

Alors, on a la majoration

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

Démonstration. Nous considérons seulement les valeurs $k=2$ ou 3, le résultat pour $k=1$ étant fourni par le théorème 12.

a) Regardons tout d'abord le cas où $u_{0h} = A_hP_h(A^{-1}u_0)$. Pour $0\leq\alpha\leq\alpha+\beta<\frac{\pi}{2}$ on a

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

d'où

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

De même

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

On a donc (voir la démonstration du théorème 12)

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

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

1^{er} cas : $k=2$. On majore par $(\mathcal{H}_2)$

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

Par le corollaire 3 on a

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

d'où $|(T_h-T)(zT-I)^{-1}A^{-1}u_0| \le \frac{C\,h^3}{|z|^{1/2}\sin\beta}\,|u_0|$. Par suite

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

On conclut en prenant $\beta = \frac{\pi}{4}-\frac{\alpha}{2}$ et en remarquant que $2\sin\beta \ge \cos\alpha$.

2^{nd} cas : $k=3$. On majore par $(\mathcal{H}_3)$

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

d'où $\quad |(zT_h-I)^{-1}(T_h-T)(zT-I)^{-1}A^{-1}u_0| \le \frac{C\,h^4}{\sin^2\beta}\,|u_0|$, et

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

avec le choix $\beta = \frac{\pi}{4}-\frac{\alpha}{2}$.

b) Lorsque $u_{0h} \ne A_hP_h(A^{-1}u_0)$, il faut rajouter à la majoration précédente le terme supplémentaire

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

avec

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

Si $k=2$ on déduit de (16) et du lemme 16

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

d'où

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

De même si $k=3$

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

d'où

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

□

Il reste à justifier l'adéquation de l'hypothèse (16), cela est une conséquence du lemme suivant

Lemme 18. *On suppose que l'opérateur A_h provient d'une forme sesquilinéaire continue et V-elliptique et que de plus il existe une constante C telle que*

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

Alors il existe une constante K telle que la condition (16) *soit satisfaite.*

Démonstration. Posons

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

D'après le théorème 14, on a $\sqrt{a_0}\,\|v_h\| \leq |B_h^{1/2} v_h| \leq \sqrt{M}\,\|v_h\|$. Par ailleurs on a , $\forall\,\varepsilon > 0$,

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

l'intégrale restant convergente, la formule est aussi valide pour $\varepsilon = 0$. On a donc

$$(A_h^{1/2} - B_h^{1/2})\,v_h = \frac{1}{2\pi i}\int_{\Gamma_{\alpha+\beta}} z^{-1/2}\left((zI - A_h)^{-1}A_h - (zI - B_h)^{-1}B_h\right) v_h\,dz. \tag{17}$$

On écrit

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

On utilise maintenant les majorations

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

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

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

On reporte enfin ces majorations dans (17), ce qui donne

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

Finalement on obtient (16) avec $K = C" + \sqrt{M}$.

Remarque. Les hypothèses de ce lemme sont facilement vérifiées dans le cadre du paragraphe 4. Une démonstration analogue permet de montrer que, dans ce contexte, $D(A^{1/2}) = V$. Ce dernier résultat a été montré pour de nombreux opérateurs elliptiques, mais il n'est pas général. C'est le fameux problème de Kato. $\square$

9. Discrétisation totale

Jusqu'à présent nous avons considéré le problème

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

et ses semi-discrétisations temporelle et spatiale

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

Nous nous intéressons maintenant au schéma totalement discrétisé, défini par

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

Remarque. Cette écriture condensée est bien adaptée à l'analyse numérique du schéma (P_{ap}), mais pas directement aux calculs effectifs. Pour mettre en œuvre ce schéma, il faut l'écrire sous forme matricielle. Soit $U_h^n = (u_1^n, u_2^n, \ldots, u_N^n)^T$ le vecteur colonne des composantes de u_h^n dans la base $\{\psi_j\}_{j=1}^N$ de V_h, M et K les matrices de masse et de rigidité définies au paragraphe 7, l'opérateur A_h est alors représenté par la matrice $M^{-1}K$ dans cette base. La relation de récurrence (P_{ap}) se réécrit alors

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

De manière plus directement utilisable cela se réécrit

- $M\, U_h^{n+1} = (M - \Delta t\, K)\, U_h^n$, dans le cas Euler-explicite $(r(z) = 1 - z)$,
- $(M + \Delta t\, K)\, U_h^{n+1} = M\, U_h^n$, dans le cas Euler-implicite $(r(z) = \frac{1}{1+z})$,
- $(M + \frac{\Delta t}{2}\, K)\, U_h^{n+1} = (M - \frac{\Delta t}{2}\, K)\, U_h^n$, dans le cas Crank-Nicolson $(r(z) = \frac{1-z/2}{1+z/2})$,
- en écrivant sous la forme : $r(z) = \prod_{j=1}^q \frac{\alpha_j - \beta_j\, z}{1+\gamma_j\, z}$ la fraction rationnelle, le schéma général devient
$$U_h^{n+1} = \prod_{j=1}^q (M + \gamma_j \Delta t\, K)^{-1}(\alpha_j\, M - \beta_j \Delta t\, K)\, U_h^n.$$

Remarques. 1) Tout numéricien averti sait que l'on ne calcule jamais les matrices M^{-1}, ni $M^{-1}K$ (ces matrices seraient généralement pleines alors que M et K sont creuses), encore moins $r(\Delta t\, M^{-1}K)$. On effectue seulement des résolutions de systèmes linéaires et des produits matrices-vecteurs du type $(M + \gamma_j \Delta t\, K)U = (\alpha_j\, M - \beta_j \Delta t\, K)\, V$.

2) Le schéma dit *Euler-explicite* n'est plus réellement explicite. Lorsque l'on utilise une méthode d'éléments finis sans intégration numérique, ce schéma requiert la résolution d'un système linéaire de même type $M\, U = \dots$ que pour Euler-implicite $(M + \Delta t\, K)\, U = \dots$. L'utilisation de certaines formules d'intégration numérique (méthodes dites *lumped mass*) permettrait d'avoir une matrice M diagonale et de retrouver les avantages du caractère explicite, mais elles ne rentrent pas dans le cadre de l'étude faite ici, où nous supposons les formes $a(.\,,.)$ et $(.\,,.)$ calculées exactement.

Le critère de von Neumann.

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

La condition écrite ci-dessus, appelée critère de von Neumann, est une condition nécessaire de stabilité. En effet, soit λ une valeur propre de A_h (c'est-à-dire de $M^{-1}K$), v_h un vecteur propre correspondant. Si dans (P_{ap}) on remplace la donnée initiale u_{0h} par une donnée perturbée $u_{0h} + \varepsilon\, v_h$, la solution du schéma u_h^n est remplacée par $u_h^n + \varepsilon\, r(\lambda\, \Delta t)^n v_h$. Les perturbations de u_{0h} voient leurs composantes sur le vecteur propre v_h amplifiées par le facteur $r(\lambda\, \Delta t)^n$. Il en résulte que, si $|r(\lambda\, \Delta t)| > 1$, le schéma est violemment instable.

Lorsque la forme a est hermitienne (i.e. lorsque les opérateur A et A_h sont auto-adjoints) cette condition (vN) est aussi une condition suffisante de stabilité. En effet, en se mettant dans une base orthonormée de vecteurs propres de A_h (une telle base existe et diagonalise A_h) on obtient

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

Le théorème 2 fournit un autre condition suffisante : r est A(α)-acceptable. En effet, puisque l'opérateur A_h est mα-accrétif, on a

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

Remarque. Contrairement à la situation du paragraphe 5, le schéma d'Euler-explicite peut maintenant être utilisé. Supposons pour simplifier l'opérateur A_h auto-adjoint, alors dans le cas présent où $r(z) = 1 - z$

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

Typiquement dans une méthode d'éléments finis, la quantité $\|A_h\|_{H \to H}$ se comporte comme $\frac{c}{\underline{h}^2}$ où $\underline{h}$ est le plus petit diamètre des triangles. La méthode d'Euler est utilisable sous la condition $\Delta t \leq \frac{2\, \underline{h}^2}{c}$, mais cette condition exige des très petits pas de temps ($\underline{h}$ est d'autant plus petit que la précision souhaitée est grande, il arrive avec un facteur carré, la situation empire avec des raffinements de maillage,...) ce qui rend cette méthode explicite peu efficace. Il en serait de même pour tout autre schéma explicite.

Nous en venons maintenant à l'étude de l'erreur, tout d'abord sans hypothèse de régularité sur la donnée initiale

Théorème 19. *Sous les hypothèses $u_0 \in H$, $(\mathcal{V}_1)$ et $(\mathcal{V}_k)$, $k = 1, 2$ ou 3, r est fortement $A(\alpha)$-acceptable et d'ordre p, on a*

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

Démonstration. On écrit

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

Le premier terme du membre de droite est majoré par le théorème 17, le deuxième par l'estimation (12) qui reste valable avec A et u_0 remplacé par A_h et u_{0h}. □

Regardons maintenant le cas où l'on a un peu plus de régularité sur la donnée.

Théorème 20. *Sous les hypothèses $u_0 \in D(A)$, $(\mathcal{H}_1)$, r est $A(\alpha)$-acceptable et d'ordre 1, on a*

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

Démonstration. On écrit

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

D'après le théorème 13

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

on utilise ensuite la formule (11) avec $p = 1$ et A remplacé par A_h

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

enfin on a

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

□

Théorème 21. *Sous les hypothèses $u_0 \in D(A^2)$, $(\mathcal{V}_1)$, $(\mathcal{V}_k)$, k =1,2 ou 3, r est $A(\alpha)$-acceptable et d'ordre 2, on a*

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

Si de plus la fraction rationnelle r est fortement $A(\alpha)$-acceptable et d'ordre $p \geq 2$, on a

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

Démonstration. a) On écrit avec $R_h u_0 = A_h^{-2}\, P_h\, A^2 u_0 = T_h^2\, A^2 u_0$

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

D'après le théorème 15

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

on utilise ensuite la formule (11) avec $p = 2$ et A remplacé par A_h

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

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

Pour terminer il suffit de remarquer que $|A^{(k+1)/2} u_0| \le C\, |A^2 u_0|$ et que

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

d'après le lemme 16.

b) Il suffit d'apporter la modification suivante aux majorations précédentes. D'après la formule (13) avec $k = p - 2$ et A remplacé par A_h

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

□

Référence

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