15$^{\text{th}}$ Internet Seminar 2011/12

Operator Semigroups for Numerical Analysis

The 15$^{\text{th}}$ Internet Seminar on Evolution Equations is devoted to operator semigroup methods for numerical analysis. Based on the Lax Equivalence Theorem we give an operator theoretic and functional analytic approach to the numerical treatment of evolution equations.

The lectures are at a beginning graduate level and only assume basic familiarity with functional analysis, ordinary and partial differential equations, and numerical analysis.

Organised by the European consortium "International School on Evolution Equations", the annual Internet Seminars introduce master-, Ph.D. students and postdocs to varying subjects related to evolution equations. The course consists of three phases.

- In Phase 1 (October-February), a weekly lecture will be provided via the ISEM website. Our aim is to give a thorough introduction to the field, at a speed suitable for master's or Ph.D. students. The weekly lecture will be accompanied by exercises, and the participants are supposed to solve these problems.
- In Phase 2 (March-May), the participants will form small international groups to work on diverse projects which complement the theory of Phase 1 and provide some applications of it.
- Finally, Phase 3 (3-9 June 2012) consists of the final one-week workshop at the Heinrich–Fabri Institut in Blaubeuren (Germany). There the teams will present their projects and additional lectures will be delivered by leading experts.

ISEM team 2011/12:

Virtual lecturers:	András Bátkai (Budapest) Bálint Farkas (Budapest) Petra Csomós (Innsbruck) Alexander Ostermann (Innsbruck)
Website:	`https://isem-mathematik.uibk.ac.at`
Further Information:	`isem@uibk.ac.at`

Description of the course

The course concentrates on the numerical solution of initial value problems of the type

$$\begin{aligned} u'(t) &= Au(t) + f(t), \quad t \geq 0, \\ u(0) &= u_0 \in D(A), \end{aligned}$$

where A is a linear operator with dense domain of definition $D(A)$ in a Banach space X, and u_0 is the initial value. A model example is the Laplace operator $A = \Delta$ with appropriate domain in the Hilbert space $L^2(\Omega)$. In this case the above partial differential equation describes heat conduction inside Ω. One way of finding a solution to this initial value problem is to imitate the way in which one solves linear ordinary differential equations with constant coefficients: First define the exponential e^{tA} in suitable way. Then the solution of the homogeneous problem is given by this fundamental operator applied to the initial value u_0, i.e., $u(t) = \mathrm{e}^{tA}u_0$. This is where operator semigroup theory enters the game: the fundamental operators $T(t) := \mathrm{e}^{tA}$ form a so-called strongly continuous semigroup of bounded linear operators on the Banach space X. That is to say the functional equation $T(t+s) = T(t)T(s)$ and $T(0) = I$ holds together with the continuity of the orbits $t \mapsto T(t)u_0$. If such a semigroup exists, we say that the initial value problem is well-posed. Once existence and uniqueness of solutions are guaranteed, the following numerical aspects appear.

- In most cases the operator A is complicated and numerically impossible to work with, so one approximates it via a sequence of (simple) operators A_m hoping that the corresponding solutions e^{tA_m} (expected to be easily computable) converge to the solution of the original problem e^{tA} in some sense. This procedure is called *space discretisation*. This discretisation may indeed come from a spatial mesh (e.g., for a finite difference method) or from some not so space-related discretisations, e.g., from Fourier-Galerkin methods.

- Equally hard is the computation of the exponential of an operator A. One idea is to approximate the exponential function $z \mapsto \mathrm{e}^z$ by functions r that are easier to handle. A typical example, known also from basic calculus courses, is the backward Euler scheme $r(z) = (1-z)^{-1}$. In this case the approximation means $r(0) = r'(0) = \mathrm{e}^0$, i.e., the first two Taylor coefficients of r and of the exponential function coincide. This leads to the following idea. If $r(tA)$ is approximately the same as e^{tA} for small values of t (up to an error of magnitude t^2), we may take the n^{th} power of it. To compensate for the growing error, we take decreasing time steps as n grows and obtain

  $$\big[r(\tfrac{t}{n}A)\big]^n \approx \big[\mathrm{e}^{\frac{t}{n}A}\big]^n = \mathrm{e}^{tA}$$

  by the semigroup property. This procedure is called *temporal discretisation.*

- Due to numerical reasons, one is usually forced to combine the above two methods and add further spice to the stew: operator splitting. This is usually done when the operator A has a complicated structure, but decomposes into a finite number of parts that are easier to handle.

In semigroup theory the above methods culminate in the famous Lax Equivalence Theorem and Chernoff's Theorem, describing precisely the situation when these methods work. In this course we shall develop the basic tools from operator semigroup theory needed for such an abstract treatment of discretisation procedures.

Topics to be covered include:

- initial value problems and operator semigroups,
- spatial discretisations, Trotter–Kato theorems, finite element and finite difference approximations,
- fractional powers, interpolation spaces, analytic semigroups,
- the Lax Equivalence Theorem and Chernoff's Theorem, error estimates, order of convergence, stability issues,
- temporal discretisations, rational approximations, Runge–Kutta methods, operator splitting procedures,
- applications to various differential equations, like inhomogeneous problems, non-autonomous equations, semilinear equations, Schrödinger equations, delay differential equations, Volterra equations,
- exponential integrators.

Some of these topics will be elaborated on in Phase 2, where the students will have the possibility to work on projects which are related to active research.

Lecture 1

What is the Topic of this Course?

The ultimate aim of these notes is quickly formulated: We would like to develop those functional analytic tools that allows us to adopt methods for ordinary differential equations (ODEs) to solve some classes of time-dependent partial differential equations (PDEs) numerically.

Let us illustrate this idea by recalling first the most trivial one of all ODEs. For a matrix $A \in \mathbb{R}^{d\times d}$ consider the initial value problem

$$\begin{cases} \dot{u}(t) = Au(t), \\ u(0) = u_0. \end{cases}$$

We know that the solution to such an ordinary differential equation is given by

$$u(t) = \mathrm{e}^{tA}u_0,$$

where e^{tA} is the exponential function of the matrix tA defined by the power series

$$\mathrm{e}^{tA} = \sum_{n=0}^{\infty} \frac{t^n A^n}{n!},$$

which converges absolutely and uniformly on every compact interval of $\mathbb{R}$. Here the numerical challenge is, especially for large matrices, to calculate this exponential function in an effective and accurate way.

The exponential function of a matrix plays an important role not only because it solves the linear problem above, but it also occurs in more complicated problems where a nonlinearity is present, like in the equation

$$\dot{v}(t) = Av(t) + F(t, v(t)).$$

To solve such an equation by iterative methods the variation of constants formula plays an essential role, stating that the solution $v(t)$ of this nonlinear equation satisfies

$$v(t) = \mathrm{e}^{tA}v(0) + \int_0^t \mathrm{e}^{(t-s)A}F(s, v(s))\,\mathrm{d}s.$$

Here again the exponential function of a matrix appears. Of course, here further numerical issues arise, such as the calculation of integrals.

There is a multitude of theoretical methods for the calculation of such exponentials, each of them leading to some possible numerical treatment of the problem. We mention those that will be important for us in this course:

1. by means of the Jordan normal form,

2. by means of the Cauchy's integral formula, more precisely, by using the identity
$$\frac{1}{2\pi \mathrm{i}}\oint \frac{\mathrm{e}^{\lambda}}{\lambda - z}\,\mathrm{d}\lambda = \mathrm{e}^{z},$$
3. by using other formulae for the exponential function, say
$$\mathrm{e}^{x} = \lim_{n\to\infty}\Big(1+\frac{x}{n}\Big)^{n} = \lim_{n\to\infty}\Big(1-\frac{x}{n}\Big)^{-n}.$$

Let us start by looking at the first of the suggestions on the above list. Theory tells us that we "only" have to bring A to Jordan normal form, and then the exponential function can be simply read off. The situation is even better if we can find a basis of orthogonal eigenvectors. Then we can bring the matrix A to diagonal form by a similarity transformation $S^{-1}AS = D = \operatorname{diag}(\lambda_1, \ldots, \lambda_d)$, and hence the exponential becomes
$$\mathrm{e}^{tA} = S\mathrm{e}^{tD}S^{-1} = S\operatorname{diag}(\mathrm{e}^{t\lambda_1}, \ldots, \mathrm{e}^{t\lambda_d})S^{-1}.$$
Of course, other numerical difficulties are hidden in calculating the Jordan normal form or the similarity transformation S. Still this very idea proves itself to be useful for partial differential equations. Let us illustrate this idea on the next example.

1.1 The heat equation

Consider the one-dimensional heat equation, say, on the interval $(0,\pi)$
$$\begin{aligned}\partial_t w(t,x) &= \partial_{xx} w(t,x), \quad t>0\\ w(0,x) &= w_0(x),\end{aligned}$$
with homogeneous Dirichlet boundary conditions
$$w(t,0) = w(t,\pi) = 0.$$
We can rewrite this equation (without the initial condition) as a linear ordinary differential equation
$$\dot{u}(t) = Au(t), \quad t>0 \tag{1.1}$$
in the infinite dimensional Hilbert space $\mathrm{L}^2(0,\pi)$. To do this define the operator
$$(Ag)(x) := g''(x) = \frac{\mathrm{d}^2}{\mathrm{d}x^2}g(x)$$
with domain
$$\begin{aligned}D(A) := \Big\{g \in \mathrm{L}^2(0,\pi) : g \text{ cont. differentiable on } [0,\pi],&\\ g'' \text{ exists a.e., } g'' \in \mathrm{L}^2,\ g'(t) - g'(0) = \textstyle\int_0^t g''(s)\,\mathrm{d}s \text{ for } t \in [0,\pi]&\\ \text{and } g(0) = g(\pi) = 0\Big\}.&\end{aligned}$$
Note that the definition of the domain has two ingredients: a condition that the differential operator on the right-hand side of the equation has values in the underlying space (in this case L^2), and

boundary conditions. The initial value is a function $f \in \mathrm{L}^2(0,\pi)$, $f = w_0$, and we look for a continuous function $u : [0,\infty) \to \mathrm{L}^2(0,\pi)$ that is differentiable on $(0,\infty)$ and satisfies equation (1.1) with $u(0) = f$. Formally the solution of this problem is given by the exponential function "e^{tA}" applied to the initial value f. Our aim is now to give a mathematical meaning to the expression "$u(t) = \mathrm{e}^{tA} f$".

First of all, we calculate the eigenvalues of this operator. These are $-n^2$ with corresponding eigenvectors

$$f_n(x) = \sqrt{\frac{2}{\pi}} \sin(nx) \quad \text{for } n \in \mathbb{N},$$

that is,

$$A f_n = -n^2 f_n. \tag{1.2}$$

Note that we have normalised the eigenvectors so that $\|f_n\|_2 = 1$. It is also easy to see that these eigenfunctions are mutually orthogonal with respect to the L^2 scalar product, i.e.,

$$\langle f_n, f_m \rangle := \int_0^\pi f_n(x)\overline{f_m(x)}\,\mathrm{d}x = \begin{cases} 1, & \text{for } n = m \\ 0, & \text{otherwise.} \end{cases}$$

The linear span $\operatorname{lin}\{f_n : n \in \mathbb{N}\}$ of these functions is dense in $\mathrm{L}^2(0,\pi)$, so altogether we obtain a orthonormal basis of eigenvectors of A. As a consequence, every function $f \in \mathrm{L}^2(0,\pi)$ can be written as a series

$$f = \sum_{n=1}^{\infty} \langle f, f_n \rangle f_n, \tag{1.3}$$

where the convergence has to be understood in the L^2 norm. We call $\langle f, f_n \rangle$ the (generalised) **Fourier coefficients** of f.

For $f \in \operatorname{lin}\{f_n : n \in \mathbb{N}\}$, $f = \sum_{n=1}^{N} a_n f_n$, the action of A is simple:

$$Af = \sum_{n=1}^{N} a_n A f_n = \sum_{n=1}^{N} -n^2 a_n f_n.$$

One expects that such a formula should hold true for functions for which the series on the right-hand side converges in $\mathrm{L}^2(0,\pi)$.

Proposition 1.1. *Consider the linear operator M on $\mathrm{L}^2(0,\pi)$ with domain*

$$D(M) := \Big\{ f \in \mathrm{L}^2(0,\pi) : \sum_{n=1}^{\infty} n^4 |\langle f, f_n \rangle|^2 < \infty \Big\}$$

defined by

$$Mf := \sum_{n=1}^{\infty} -n^2 \langle f, f_n \rangle f_n.$$

Then $A = M$, i.e., $D(A) = D(M)$, and for $f \in D(A)$ we have $Af = Mf$. In particular we have

$$Af = \sum_{n=1}^{\infty} -n^2 \langle f, f_n \rangle f_n \quad \text{for all } f \in D(A) = D(M).$$

Proof. Suppose $f \in D(A)$. Then we integrate by parts twice(!) and obtain

$$\begin{aligned}\sqrt{\frac{\pi}{2}}\langle Af, f_n\rangle &= \int\limits_0^\pi f''(x)\sin(nx)\,\mathrm{d}x = f'(x)\sin(nx)\Big|_{x=0}^{x=\pi} - n\int\limits_0^\pi f'(x)\cos(nx)\,\mathrm{d}x\\ &= -n\int\limits_0^\pi f'(x)\cos(nx)\,\mathrm{d}x\\ &= -nf(x)\cos(nx)\Big|_{x=0}^{x=\pi} - n^2\int\limits_0^\pi f(x)\sin(nx)\,\mathrm{d}x = -n^2\sqrt{\frac{\pi}{2}}\langle f, f_n\rangle,\end{aligned}$$

where in the last step we used the boundary conditions $f(0) = f(\pi) = 0$. Since $Af \in \mathrm{L}^2$, its Fourier coefficients are square summable. Whence, $f \in D(M)$ follows. This shows $D(A) \subseteq D(M)$. We also see that

$$Af = \sum_{n=1}^\infty -n^2\langle f, f_n\rangle f_n \quad \text{holds for all } f \in D(A).$$

It only remains to show the other inclusion $D(M) \subseteq D(A)$. To see that, it suffices to note that A is surjective (this is “classical”) and M is injective, so $A = M$ because M extends A (see Exercises 3 and 4.) □

Intuitively, the result above states that A *has diagonal form* with respect to the basis of eigenvectors, and is given by

$$A = \mathrm{diag}(-1, -2^2, \dots, -n^2, \dots).$$

Thus, the exponential of this operator can be immediately defined as

$$\mathrm{e}^{tA} := \mathrm{diag}(\mathrm{e}^{-t}, \mathrm{e}^{-t4}, \dots, \mathrm{e}^{-tn^2}, \dots),$$

meaning that

$$\mathrm{e}^{tA}f = \sum_{n=1}^\infty \mathrm{e}^{-tn^2}\langle f, f_n\rangle f_n.$$

We have to show that this is a meaningful definition. As a first step, let us show that the formula above gives rise to a continuous function.

Proposition 1.2. *Let $f \in \mathrm{L}^2(0,\pi)$. Then for every $t \geq 0$ the series*

$$\mathrm{e}^{tA}f := \sum_{n=1}^\infty \mathrm{e}^{-tn^2}\langle f, f_n\rangle f_n$$

is convergent and defines a function $u(t) = \mathrm{e}^{tA}f$ which is continuous on $[0,\infty)$ with values in $\mathrm{L}^2(0,\pi)$.

Proof. Since for every $n \in \mathbb{N}$ and $t \geq 0$ the inequality $|\mathrm{e}^{-tn^2}| \leq 1$ holds, the sequence $(\mathrm{e}^{-tn^2}\langle f, f_n\rangle)$ is square summable, and the series

$$\sum_{n=1}^{\infty} \mathrm{e}^{-tn^2}\langle f, f_n\rangle f_n$$

that defines $u(t) = \mathrm{e}^{tA}f$ converges in $\mathrm{L}^2(0,\pi)$.

We now prove the continuity at a given $t \geq 0$. Let $\varepsilon > 0$ be given, and choose $n_0 \in \mathbb{N}$ so that

$$\sum_{n=n_0+1}^{\infty} |\langle f, f_n\rangle|^2 \leq \varepsilon.$$

If $t = 0$, then in the following we consider only $h \geq 0$, and if $t > 0$ we additionally suppose $|h| \leq t$. This way we can write

$$\begin{aligned}\big\|\mathrm{e}^{tA}f - \mathrm{e}^{(t+h)A}f\big\|_2^2 &= \big\langle \mathrm{e}^{tA}f - \mathrm{e}^{(t+h)A}f, \mathrm{e}^{tA}f - \mathrm{e}^{(t+h)A}f\big\rangle \\ &= \sum_{n=1}^{\infty}\big|\mathrm{e}^{-(t+h)n^2} - \mathrm{e}^{-tn^2}\big|^2\big|\langle f, f_n\rangle\big|^2 \leq \sum_{n=1}^{n_0}\big|\mathrm{e}^{-(t+h)n^2} - \mathrm{e}^{-tn^2}\big|^2\big|\langle f, f_n\rangle\big|^2 + 2\varepsilon \\ &\leq \sum_{n=1}^{n_0}\big|\mathrm{e}^{-hn^2} - 1\big|^2\big|\langle f, f_n\rangle\big|^2 + 2\varepsilon.\end{aligned}$$

We can finish the proof by choosing $|h|$ so small that the first finitely many terms contribute at most ε. □

Hence, this exponential function provides a candidate to be the solution of (1.1). Let us prove that it is indeed the solution.

Proposition 1.3. *For $f \in \mathrm{L}^2(0,\pi)$ we define $u(t) := \mathrm{e}^{tA}f$. Then $u(t) \in D(A)$ holds for all $t > 0$, and u is differentiable on $(0,\infty)$ with derivative $Au(t)$. That is, u solves the initial value problem*

$$\begin{aligned}\dot{u}(t) &= Au(t), \quad t > 0 \\ u(0) &= f.\end{aligned}$$

Proof. The initial condition is fulfilled by (1.3). Note that for all $t > 0$ and $n \in \mathbb{N}$ we have

$$|\mathrm{e}^{-tn^2}n^2| \leq \mathrm{e}^{-\frac{t}{2}n^2}\frac{2\mathrm{e}^{-1}}{t} \quad \text{for all } n \in \mathbb{N}. \tag{1.4}$$

From this estimate, using the characterisation in Proposition 1.1, we obtain that $u(t) \in D(M) = D(A)$ for each $t > 0$. Define

$$\begin{aligned}v(t) &:= Au(t), \\ u_n(s) &:= \mathrm{e}^{-sn^2}\langle f, f_n\rangle f_n, \\ \text{and} \qquad v_n(s) &:= -n^2\mathrm{e}^{-sn^2}\langle f, f_n\rangle f_n.\end{aligned}$$

Then $\dot{u}_n = v_n$, and both functions are continuous on $[t/2, 3t/2]$ with values in L^2. From inequality (1.4) we obtain that the following two series

$$u(s) = \sum_{n=1}^{\infty} u_n(s) \qquad \text{and} \qquad v(s) = \sum_{n=1}^{\infty} \dot{u}_n(s)$$

have summable numerical majorants for $s \in [t/2, 3/2t]$. This implies that u is differentiable and that we can interchange summation and differentiation, whence $\dot{u}(t) = v(t) = Au(t)$ follows. □

Let us put the above in an abstract, operator theoretic perspective.

Proposition 1.4. *For $t \geq 0$ define $T(t)f := \mathrm{e}^{tA}f$. Then $T(t)$ is a bounded linear operator on $\mathrm{L}^2(0,\pi)$ for each $t \geq 0$. The mapping T satisfies*

$$T(t+s) = T(t)T(s) \quad \textit{and} \quad T(0) = I, \textit{ the identity operator on } \mathrm{L}^2.$$

For each $f \in \mathrm{L}^2(0,\pi)$ the function $t \mapsto T(t)f$ is continuous on $[0,\infty)$.

Proof. As we saw in Proposition 1.2, the inequality

$$\|\mathrm{e}^{tA}f\|_2^2 = \langle \mathrm{e}^{tA}f, \mathrm{e}^{tA}f\rangle \leq \sum_{n=1}^{\infty} \mathrm{e}^{-2tn^2}|\langle f, f_n\rangle|^2 \leq \sum_{n=1}^{\infty} |\langle f, f_n\rangle|^2 = \|f\|_2^2$$

holds. It is moreover clear that the mapping $f \mapsto \mathrm{e}^{tA}f$ is linear, and from the previous inequality we obtain that it is bounded with operator norm

$$\|\mathrm{e}^{tA}\| \leq 1.$$

The identity $T(t+s) = T(t)T(s)$ follows from the properties of the exponential function and the definition of e^{tA}. The relation $T(0) = I$ was discussed in Proposition 1.3, the continuity of the mapping $t \mapsto T(t)f$ follows from Proposition 1.2. □

From the properties above we can coin a new definition.

Definition 1.5. Let X be a Banach space, and let the mapping $T : [0,\infty) \to \mathscr{L}(X)$ have[1] the properties:

a) For all $t, s \in [0,\infty)$

$$\begin{cases} T(t+s) = T(t)T(s) \\ T(0) = I, \text{ the identity operator on } X. \end{cases}$$

b) For all $x \in X$ the mapping

$$t \mapsto T(t)x \in X$$

is continuous.

Then T is called a **strongly continuous** one-parameter **semigroup**[2] of bounded linear operators on the Banach space X. We abbreviate this long expression sometimes to *strongly continuous semigroup*, or simply to *semigroup*.

The semigroup constructed in Proposition 1.4 is called the (Dirichlet) **heat semigroup** on $[0,\pi]$. To sum up, we can state the following.

Conclusion 1.6. Initial value problems lead to semigroups.

[1] Here and later on, $\mathscr{L}(X)$ denotes the set of bounded linear operators on X.

[2] By an alternative terminology one may call such an object a C_0*-semigroup*.

1.2 The shift semigroup

Now that we have the new mathematical notion of *one-parameter semigroups* we want to study them in detail. This, as a matter of fact, is one of the aims of this course. Before doing so let us consider another example.

Take

$$X = \mathrm{BUC}(\mathbb{R}) := \bigl\{f : \mathbb{R} \to \mathbb{R} : f \text{ is uniformly continuous and bounded}\bigr\},$$

which is a Banach space with the supremum norm

$$\|f\|_\infty := \sup_{s\in\mathbb{R}} |f(s)|.$$

The additive (semi)group structure of $\mathbb{R}$ naturally induces a semigroup on this Banach space by setting

$$(S(t)f)(s) = f(t+s), \quad \text{for } f \in X,\ s \in \mathbb{R},\ t \geq 0.$$

One readily sees that $S(t)$ is a bounded linear operator on X, in fact a linear isometry. The semigroup property follows immediately from the definition. From the uniform continuity of $f \in X$ we conclude that

$$t \mapsto S(t)f$$

is continuous, i.e., that S is a strongly continuous semigroup on $X = \mathrm{BUC}(\mathbb{R})$, called the **left shift semigroup**.

Let us investigate whether this semigroup S solves some initial value problem such as (1.1). Again the heuristics of exponential functions helps: Given e^{tA} for a matrix $A \in \mathbb{R}^{d\times d}$, we can "calculate" the exponent by differentiating this exponential function at 0:

$$A = \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{tA}\Big|_{t=0}.$$

What happens in the case of the shift semigroup S? The semigroup S is not even continuous for the operator norm (why?). So let us look at differentiability of the **orbit map** $t \mapsto S(t)f$ for some given $f \in X$, called **strong differentiability**. The limit

$$\lim_{h\to 0} \frac{1}{h}(S(h)f - f) = \lim_{h\to 0} \frac{f(h+\cdot) - f(\cdot)}{h}$$

must exist in the sup-norm of X. We immediately find a suitable candidate for the limit: Since the limit must exist pointwise on $\mathbb{R}$, it cannot be anything else than f'. Hence, the function f must be at least differentiable so that the limit can exist. For f differentiable with f' being uniformly continuous we have

$$\sup_{s\in\mathbb{R}} \left|\frac{f(h+s)-f(s)}{h} - f'(s)\right| = \sup_{s\in\mathbb{R}} \left|\frac{1}{h}\int_s^{s+h} \bigl(f'(r) - f'(s)\bigr)\,\mathrm{d}r\right| \leq \varepsilon,$$

for all h with $|h| \leq \delta$, where $\delta > 0$ is sufficiently small, chosen for the arbitrarily given $\varepsilon > 0$ from the uniform continuity of f'. This shows that if $f, f' \in X$, then we have

$$\lim_{h\to 0} \left\|\frac{f(h+\cdot)-f(\cdot)}{h} - f'(\cdot)\right\|_\infty = \lim_{h\to 0}\sup_{s\in\mathbb{R}} \left|\frac{f(h+s)-f(s)}{h} - f'(s)\right| = 0.$$

Note that for the derivative of $S(t)f$ at arbitrary $t \in \mathbb{R}$ we obtain by the same argument

$$\frac{\mathrm{d}}{\mathrm{d}t}(S(t)f) = S(t)f'.$$

This means that for $f, f' \in X$ the orbit function $u(t) = S(t)f$ solves the differential equation

$$\begin{cases} \dot{u}(t) = Au(t) \\ u(0) = f, \end{cases}$$

where $(Af)(s) = f'(s)$ with domain

$$D(A) := \big\{ f : f, f' \in \mathrm{BUC}(\mathbb{R}) \big\}.$$

We can therefore formulate the parallel of Conclusion 1.6:

Conclusion 1.7. To a semigroup there exists a corresponding initial value problem.

1.3 What is the topic of this course?

At this point we hope to have motivated the study of strongly continuous semigroups from the analytic or PDE point of view. To solve an initial value problem $\dot{u}(t) = Au(t)$, one has to define a semigroup e^{tA}.

The numerical analysis aspects are now the following:

- The operator A is complicated, and numerically impossible to treat, so one approximates it via a sequence of operators A_m and hopes that the corresponding solutions (expected to be easily calculated) e^{tA_m} converge to the solution of the original problem e^{tA} (in a sense yet to be made precise). This procedure is called *space discretisation*, and may indeed come from a spatial mesh (e.g., for a finite element method) or from some not so space-related discretisation, like for Fourier-Galerkin methods, an instance of which we have seen in Section 1.1.

- Equally hard is to determine the exponential function of a matrix (or operator) A (see the list of suggestions on page 1). So a different idea is to approximate the exponential function $z \mapsto \mathrm{e}^z$ by functions r that are easier to handle. A typical example, known also from basic calculus courses, is that of the implicit Euler scheme $r(z) = (1-z)^{-1}$. In this case the approximation means $r(0) = 1$ and $r'(0) = 1$, i.e., the first two Taylor coefficients of the two functions coincide. Heuristically we obtain that $r(tA)$ for a small t is approximately the same as e^{tA} (up to an error of magnitude t^2), we may take the n^{th} power and to compensate the growing error we would obtain, we take the time step smaller and smaller as n grows. We obtain
$$\big(r(\tfrac{t}{n}A)\big)^n \approx \big(\mathrm{e}^{\frac{t}{n}A}\big)^n = \mathrm{e}^{tA},$$
where the semigroup property was used. This procedure is called *time discretisation.*

- Due to numerical reasons one is usually forced to combine the two methods above, and sometimes even by adding a further spice to the stew: operator splitting. This is usually done when operator A has a complicated structure, but decomposes into a finite number of parts that are easier to handle.

- The theory presented above is the basis in extending known ODE methods to time dependent partial differential equations and will allow us to use the variation of constants formula for inhomogeneous or semilinear equations. Hence the convergence analysis of various iteration methods will depend on this theory.

In semigroup theory the above methods culminate in the famous Lax-Chernoff Equivalence Theorem that describes precisely the situation when these methods work. In this course we shall develop the basic tools from operator semigroup theory needed for such an abstract treatment of discretisation procedures.

Exercises

1. Prove that $\sin(nx)$, $n \in \mathbb{N}$, form a complete orthogonal system in $\mathrm{L}^2(0,\pi)$, compute the L^2 norms.

2. Analogously to what is presented in Section 1.1, study the heat equation with **Neumann boundary conditions**:

$$\begin{aligned}\partial_t u(t,x) &= \partial_{xx} u(t,x), \quad t>0\\ u(0,x) &= f(x),\\ \partial_x u(t,0) &= \partial_x u(t,\pi) = 0.\end{aligned}$$

3. Let X be a Banach space and $A_1 : X \to X$ and $A_2 : X \to X$ linear maps such that

- $D(A_1) \subset D(A_2)$ and A_1 is a restriction of A_2
- A_1 is surjective and A_2 is injective.

Show that $A_1 = A_2$.

4. Consider the Hilbert space ℓ^2 of square summable complex sequences.

a) Prove that
$$\mathrm{c}_{00} = \big\{(x_n) \in \ell^2 : x_n = 0 \text{ except for finitely many } n\big\}$$
is a dense linear subspace of ℓ^2.

b) For $m = (m_n)$ an arbitrarily fixed sequence of complex numbers, and $x = (x_n) \in \mathrm{c}_{00}$ define
$$(M_m x)_n = (m_n x_n), \quad \text{i.e., componentwise multiplication.}$$
Give such a necessary and sufficient condition on m that $M_m : \mathrm{c}_{00} \to \mathrm{c}_{00}$ becomes a continuous linear operator with respect to the ℓ^2 norm.

c) Under this condition, prove that M_m extends continuously and linearly to ℓ^2, give a formula for this linear operator, and compute its norm.

d) Give a necessary and sufficient condition on m so that M_m has a continuous inverse.

e) Give a necessary and sufficient condition on m so that e^{tM_m} is defined analogously to Section 1.1.

5. Let $p \in [1, \infty)$ and consider the Banach space $\mathrm{L}^p(\mathbb{R})$. Prove that the formula

$$(S(t)f)(s) := f(t+s) \quad \text{for } f \in \mathrm{L}^p,\ s \in \mathbb{R},\ t \geq 0$$

defines a strongly continuous semigroup on L^p. What happens for $p = \infty$?

6. Let $\mathrm{F_b}(\mathbb{R})$ denote the linear space of all bounded $\mathbb{R} \to \mathbb{R}$ functions. Define

$$(S(t)f)(s) := f(t+s) \quad \text{for } f \in \mathrm{F_b}(\mathbb{R}),\ s \in \mathbb{R},\ t \geq 0.$$

Prove that S is a semigroup, i.e., satisfies Definition 1.5.a). Prove that each of the following spaces is a Banach space with the supremum norm $\|\cdot\|_\infty$ and invariant under $S(t)$ for all $t \geq 0$. Is S a strongly continuous semigroup on these spaces?

a) $\mathrm{F_b}(\mathbb{R})$.

b) $\mathrm{C_b}(\mathbb{R})$ = the space bounded and continuous functions.

c) $\mathrm{C_0}(\mathbb{R})$ = the space bounded and continuous functions vanishing at infinity.

7. Determine the set of those $f \in \mathrm{BUC}(\mathbb{R})$ for which $t \mapsto S(t)f$ is differentiable (S denotes the left shift semigroup).

8. Let S be the left shift semigroup on $\mathrm{BUC}(\mathbb{R})$, and T be the heat semigroup from Section 1.1. Prove the following assertions:

a) $t \mapsto S(t)$ is nowhere continuous for the operator norm.

b) $t \mapsto T(t)$ is not continuous for the operator norm at 0.

c) $t \mapsto T(t)$ is continuous for the operator norm on $(0, \infty)$.

9. Explain why it is not possible to define the heat semigroup for negative time values.

Lecture 2

Fundamentals of One-Parameter Semigroups

Last week we motivated the study of strongly continuous semigroups by standard PDE examples. In this lecture we begin with the thorough investigation of these mathematical objects, and recall first a definition from Lecture 1. Here and later on, X denotes a Banach space, and $\mathscr{L}(X)$ stands for the Banach space of bounded linear operators acting on X.

Definition 2.1. Let $T : [0,\infty) \to \mathscr{L}(X)$ be a mapping.

a) We say that T has the **semigroup property** if for all $t, s \in [0,\infty)$ the identities

$$\begin{aligned} & T(t+s) = T(t)T(s) \\ \text{and} \qquad & T(0) = I, \text{ the identity operator on } X, \end{aligned}$$

hold.

b) Suppose $Y \subseteq X$ is a linear subspace and for all $f \in Y$ the mapping

$$t \mapsto T(t)f \in X$$

is continuous. Then T is called **strongly continuous on** Y. If $Y = X$ we just say **strongly continuous**.

c) A strongly continuous mapping T possessing the semigroup property is called a **strongly continuous one-parameter semigroup** of bounded linear operators on the Banach space X. Often we shall abbreviate this terminology to **semigroup**.

2.1 Basic properties

Let us observe some elementary consequences of the semigroup property and the strong continuity, respectively. The first result reflects again the exponential function: Semigroups can grow at most exponentially.

Proposition 2.2. *a) Let $T : [0,\infty) \to \mathscr{L}(X)$ be a strongly continuous function. Then for all $t \geq 0$ we have*

$$\sup_{s\in[0,t]} \|T(s)\| < \infty,$$

*that is to say, T is **locally bounded**.*

b) Let $T : [0,\infty) \to \mathscr{L}(X)$ be a strongly continuous semigroup. Then there are $M \geq 1$ and $\omega \in \mathbb{R}$ such that

$$\|T(t)\| \leq M\mathrm{e}^{\omega t} \quad \textit{holds for all } t \geq 0.$$

We call the semigroup T of **type** (M, ω) if it satisfies the exponential estimate above with the particular constants M and ω. Note already here that the type of a semigroup may change if we pass to an equivalent norm on X.

Proof. a) For $f \in X$ fixed, the mapping $T(\cdot)f$ is continuous on $[0, \infty)$, hence bounded on compact intervals $[0, t]$, i.e.,

$$\sup_{s \in [0,t]} \|T(s)f\| < \infty.$$

The uniform boundedness principle, see Supplement, Theorem 2.28, implies the assertion.

b) By part a) we have

$$M := \sup_{s \in [0,1]} \|T(s)\| < \infty.$$

Take $t \geq 0$ arbitrary and write $t = n + r$ with $n \in \mathbb{N}$ and $r \in [0, 1)$. From this representation we obtain by using the semigroup property that

$$\begin{aligned} \|T(t)\| = \|T(r)T(1)^n\| \leq M\|T(1)^n\| &\leq M\|T(1)\|^n \\ \leq M(\|T(1)\| + 1)^n \leq M(\|T(1)\| + 1)^t &= M\mathrm{e}^{\omega t} \end{aligned}$$

with $\omega = \log(\|T(1)\| + 1)$. □

Hence, orbits of strongly continuous semigroups are **exponentially bounded**. The *greatest lower bound* of these exponential bounds plays a special role in the theory, hence, we give it a name.

Definition 2.3. For a strongly continuous semigroup T its **growth bound**[1] is defined by

$$\omega_0(T) := \inf\bigl\{\omega \in \mathbb{R} : \text{ there is } M = M_\omega \geq 1 \text{ with } \|T(t)\| \leq M\mathrm{e}^{\omega t} \text{ for all } t \geq 0\bigr\}.$$

Remark 2.4. 1. A strongly continuous semigroup T is of type (M, ω) for all $\omega > \omega_0(T)$ and for some $M = M_\omega$. In general, however, it is *not* of type $(M, \omega_0(T))$ for any M. A simple example is the following. Let $X = \mathbb{C}^2$ and let the matrix semigroup given by

$$T(t) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

Here $\omega_0 = 0$, but clearly T is not bounded, i.e., not of type $(M, 0)$ for any M.

2. For a matrix $A \in \mathbb{R}^{d \times d}$ we define $T(t) = \mathrm{e}^{tA}$. This semigroup T is of type $(1, \|A\|)$ as the trivial norm estimate

$$\|\mathrm{e}^{tA}\| \leq \mathrm{e}^{t\|A\|}$$

shows. In contrast to this, in infinite dimensional situation it can happen that a semigroup is not of type $(1, \omega)$ for any ω, even though $\omega_0(T) = -\infty$. This is an extremely important fact, which causes major difficulties in stability questions of approximation schemes (see Exercise 4).

The definition of an operator semigroup above comprises of the algebraic semigroup property, and the analytic property of strong continuity. We shall see next that these two properties combine well, and we provide some means for verifying strong continuity.

Proposition 2.5. *a) Let $T : [0, \infty) \to \mathscr{L}(X)$ be a locally bounded mapping with the semigroup property, and let $f \in X$. If the mapping $T(\cdot)f$ is* right continuous *at* 0, *i.e., $T(h)f \to f$ for $h \searrow 0$, then it is continuous everywhere.*

[1] Also called the first Lyapunov exponent.

b) A mapping T with the semigroup property is strongly continuous on X if and only if it is locally bounded and there is a dense subset $D \subseteq X$ on which T is strongly continuous.

Proof. a) Fix $f \in X$ and $t > 0$, and set $M := \sup_{[0,2t]} \|T(s)\|$. Then

$$\begin{aligned} T(t+h)f - T(t)f &= T(t)(T(h)f - f), \quad \text{ if } 0 < h, \\ T(t+h)f - T(t)f &= T(t+h)(f - T(-h)f), \quad \text{ if } -t < h < 0. \end{aligned}$$

Summarizing, for $|h| \leq t$ we obtain

$$\|T(t+h)f - T(t)f\| \leq M\|f - T(|h|)f\|,$$

which converges to 0 for $|h| \to 0$ by the assumption.

b) In view of Proposition 2.2 one implication is straightforward. So we turn to the other one, and suppose that T is locally bounded and strongly continuous on a dense subspace D. Take an arbitrary $f \in X$ and some $\varepsilon > 0$. Set $M := \sup_{s\in[0,1]} \|T(s)\|$ and note that $M \geq 1$. By denseness there is $g \in D$ with $\|f - g\| \leq \frac{\varepsilon}{3M}$, whence

$$\|T(h)f - f\| \leq \|T(h)f - T(h)g\| + \|T(h)g - g\| + \|g - f\| \leq \frac{\varepsilon}{3} + M\frac{\varepsilon}{3M} + \frac{\varepsilon}{3M} \leq \varepsilon$$

follows if h is sufficiently small, chosen to $\varepsilon/3$ by the right continuity of $T(\cdot)g$. This shows that the orbit map $T(\cdot)f$ is right continuous at 0, and from part a) even continuity everywhere can be concluded. □

2.2 The infinitesimal generator

One main message in Lecture 1 was that if we have a *semigroup*, then there is a *differential equation* so that the semigroup provides the solutions. Looking for the equation, we now consider the differentiability of orbit maps as in Section 1.2.

Lemma 2.6. *Take a semigroup T and an element $f \in X$. For the orbit map $u : t \mapsto T(t)f$, the following properties are equivalent:*

(i) u is differentiable on $[0, \infty)$,

(ii) u is right differentiable at 0.

If u is differentiable, then

$$\dot{u}(t) = T(t)\dot{u}(0).$$

Proof. We only have to show that (ii) implies (i). Analogously to the proof of Proposition 2.5, one has

$$\begin{aligned} \lim_{h\searrow 0} \tfrac{1}{h}\left(u(t+h) - u(t)\right) &= \lim_{h\searrow 0} \tfrac{1}{h}\left(T(t+h)f - T(t)f\right) = T(t) \lim_{h\searrow 0} \tfrac{1}{h}\left(T(h)f - f\right) \\ &= T(t) \lim_{h\searrow 0} \tfrac{1}{h}\left(u(h) - u(0)\right) = T(t)\,\dot{u}(0), \end{aligned}$$

by the continuity of $T(t)$. Hence u is right differentiable on $[0, \infty)$.

On the other hand, for $-t \leq h < 0$, we write

$$\begin{aligned}&\tfrac{1}{h}\left(T(t+h)f - T(t)f\right) - T(t)\dot{u}(0)\\ &\qquad = T(t+h)\left(\tfrac{1}{h}\left(f - T(-h)f\right) - \dot{u}(0)\right) + T(t+h)\dot{u}(0) - T(t)\dot{u}(0).\end{aligned}$$

As $h \nearrow 0$, the first term on the right-hand side converges to zero by the first part and by the boundedness of $\|T(t+h)\|$ for $h \in [-t, t]$. The other term converges to zero by the strong continuity of T. Hence u is also left differentiable, and its derivative is

$$\dot{u}(t) = T(t)\,\dot{u}(0)$$

for all $t \geq 0$. □

Hence, the derivative $\dot{u}(0)$ of the orbit map $u(t) = T(t)f$ at $t = 0$ determines the derivative at each point $t \in [0, \infty)$. We therefore give a name to the operator $f \mapsto \dot{u}(0)$.

Definition 2.7. The **infinitesimal generator**, or simply **generator** A of a semigroup T is defined as follows. Its **domain of definition** is given by

$$D(A) := \{f \in X : T(\cdot)f \text{ is differentiable in } [0, \infty)\},$$

and for $f \in D(A)$ we set

$$Af := \tfrac{\mathrm{d}}{\mathrm{d}t} T(t)f|_{t=0} = \lim_{h \searrow 0} \tfrac{1}{h}\left(T(h)f - f\right).$$

As we hoped for, a semigroup yields solutions to some linear ODE in the Banach space X.

Proposition 2.8. *The generator A of a strongly continuous semigroup T has the following properties.*

a) $A : D(A) \subseteq X \to X$ is a linear operator.

b) If $f \in D(A)$, then $T(t)f \in D(A)$ and

$$\tfrac{\mathrm{d}}{\mathrm{d}t} T(t)f = T(t)Af = AT(t)f \quad \text{for all } t \geq 0.$$

c) For a given $f \in D(A)$, the semigroup T provides the solutions to the initial value problem

$$\begin{aligned}\dot{u}(t) &= Au(t), \quad t \geq 0\\ u(0) &= f\end{aligned}$$

via $u(t) := T(t)f$.

Proof. a) Linearity follows immediately from the definition because we take the limit of linear objects as $h \searrow 0$.

b) Take $f \in D(A)$ and $t \geq 0$. We have to show that $T(\cdot)T(t)f$ is right differentiable at 0 with derivative $T(t)Af$. From the continuity of $T(t)$ we obtain

$$T(t)Af = T(t) \lim_{h \searrow 0} \frac{T(h)f - f}{h} = \lim_{h \searrow 0} \frac{T(h)T(t)f - T(t)f}{h}.$$

By the definition of A this further equals $AT(t)f$.

Part c) is just a reformulation of b). □

We now investigate infinitesimal generators further.

Proposition 2.9. *The generator A of a strongly continuous semigroup T has the following properties.*

a) For all $t \geq 0$ and $f \in X$, one has

$$\int_0^t T(s)f \,\mathrm{d}s \in D(A),$$

where the integral has to be understood as the Riemann integral of the continuous function $s \mapsto T(s)f$, see Supplement.

b) For all $t \geq 0$, one has

$$\begin{aligned} T(t)f - f &= A \int_0^t T(s)f \,\mathrm{d}s \quad \textit{if } f \in X, \\ &= \int_0^t T(s)Af \,\mathrm{d}s \quad \textit{if } f \in D(A). \end{aligned}$$

Proof. a) For $g := \int_0^t T(s)f \,\mathrm{d}s$ we calculate the difference quotients

$$\begin{aligned} \frac{T(h)g - g}{h} &= \frac{1}{h}\Big(T(h)\int_0^t T(s)f \,\mathrm{d}s - \int_0^t T(s)f \,\mathrm{d}s\Big) = \frac{1}{h}\Big(\int_0^t T(h+s)f \,\mathrm{d}s - \int_0^t T(s)f \,\mathrm{d}s\Big) \\ &= \frac{1}{h}\Big(\int_h^{t+h} T(s)f \,\mathrm{d}s - \int_0^t T(s)f \,\mathrm{d}s\Big) = \frac{1}{h}\Big(\int_t^{t+h} T(s)f \,\mathrm{d}s - \int_0^h T(s)f \,\mathrm{d}s\Big). \end{aligned}$$

Since the integrands here are continuous, we can take limits as $h \searrow 0$ and obtain

$$\lim_{h \searrow 0} \frac{T(h)g - g}{h} = T(t)f - f.$$

This yields $g \in D(A)$ and $Ag = T(t)f - f$.

b) Take $f \in D(A)$, then by Proposition 2.8.b) the identity $AT(t)f = T(t)Af$ holds, hence $v(t) := AT(t)f$ defines a continuous function. For $h > 0$ define the continuous functions $v_h(t) := \frac{1}{h}(T(t+h)f - T(t)f)$. Then we have

$$\|v_h(t) - v(t)\| \leq \|T(t)\| \Big\| \frac{1}{h}(T(h)f - f) - Af \Big\|.$$

From this and the definition of A we conclude (by using the local boundedness of T) that v_h converges to v uniformly on every compact interval. This yields

$$\int_0^t v_h(s) \,\mathrm{d}s \to \int_0^t v(s) \,\mathrm{d}s \quad \text{as } h \searrow 0.$$

We have calculated the limit of the left-hand side in part b): It equals

$$T(t)f - f = A \int_0^t T(s)f \, \mathrm{d}s,$$

which completes the proof. □

Before turning our attention to the main result of this section, let us recall what a *closed operator* is. For a linear operator A defined on a linear subspace $D(A)$ of a Banach space X, we define the **graph norm** of A by

$$\|f\|_A := \|f\| + \|Af\| \quad \text{for } f \in D(A).$$

Then, indeed, $\|\cdot\|_A$ is a norm on $D(A)$. The operator A is called **closed** if $D(A)$ is complete with respect to this graph norm, i.e., if $D(A)$ is a Banach space with this graph norm $\|\cdot\|_A$. The following proposition yields simple yet useful reformulations of the closedness of a linear operator, we leave out the proof.

Proposition 2.10. *Let A be a linear operator with domain $D(A)$ in X. The following assertions are equivalent.*

(i) A is a closed operator.

(ii) For every sequence $(x_n) \subseteq D(A)$ with $x_n \to x$ and $Ax_n \to y$ in X for some $x, y \in X$ one has $x \in D(A)$ and $Ax = y$.

If A is injective the properties above are further equivalent to the following:

(iii) The inverse A^{-1} of A is a closed operator.

The main result of this section summarises the basic properties of the generator.

Theorem 2.11. *The generator of a semigroup is a closed and densely defined linear operator that determines the semigroup uniquely.*

Proof. Let $(f_n) \subseteq D(A)$ be a Cauchy sequence in $D(A)$ with respect to the graph norm. Since for all $f \in D(A)$ the inequalities

$$\|f\| \leq \|f\|_A \quad \text{and} \quad \|Af\| \leq \|f\|_A$$

hold, we conclude that (f_n) and (Af_n) are Cauchy sequences in X with respect to the norm $\|\cdot\|$. Hence, they converge to some $f \in X$ and $g \in X$, respectively. For $t > 0$ we have

$$T(t)f_n - f_n = \int_0^t T(s)Af_n \, \mathrm{d}s$$

by Proposition 2.9. If we set $u_n(s) := T(s)Af_n$ and $u(s) := T(s)g$, then $u_n \to u$ uniformly on $[0, t]$, since T is locally bounded. So we can pass to the limit in the identity above, and obtain

$$T(t)f - f = \int_0^t T(s)g \, \mathrm{d}s.$$

From this we deduce that $t \mapsto T(t)f$ is differentiable at 0 with derivative $u(0) = g$. This means precisely $f \in D(A)$ and $Af = g$. To conclude, we note

$$\|f - f_n\|_A = \|f - f_n\| + \|Af - Af_n\| \to 0 \quad \text{as } n \to \infty,$$

i.e., $f_n \to f$ in graph norm. Therefore, A is a closed operator.

We now show that $D(A)$ is dense in X. Let $f \in X$ be arbitrary and define

$$v(t) := \frac{1}{t} \int\limits_0^t T(s)f \, \mathrm{d}s \quad (t > 0).$$

By Proposition 2.9 we obtain $v(t) \in D(A)$. Since $s \mapsto T(s)f$ is continuous, we have $v(t) \to T(0)f = f$ for $t \searrow 0$.

Suppose S is a semigroup with the same generator A as T. Let $f \in D(A)$ and $t > 0$ be fixed, and consider the function $u : [0, t] \to X$ given by $u(s) := T(t-s)S(s)f$. Then u is differentiable and its derivative is given by the product rule, see Supplement, Theorem 2.31,

$$\tfrac{\mathrm{d}}{\mathrm{d}s}u(s) = (\tfrac{\mathrm{d}}{\mathrm{d}s}T(t-s))S(s)f + T(t-s)\tfrac{\mathrm{d}}{\mathrm{d}s}(S(s)f) = -AT(t-s)S(s)f + T(t-s)AS(s)f.$$

By using that the semigroup and its generator commute on $D(A)$, see Proposition 2.8.b), we obtain that the right-hand term is 0, so u must be constant. This implies

$$S(t)f = u(t) = u(0) = T(t)f,$$

i.e., the bounded linear operators $S(t)$ and $T(t)$ coincide on the dense subspace $D(A)$, hence they must be equal everywhere. □

2.3 Two basic examples

Shift semigroups

Recall from Exercise 1.5 the shift semigroups on the spaces $\mathrm{L}^p(\mathbb{R})$ with $p \in [1, \infty)$. For $f \in \mathrm{L}^p(\mathbb{R})$ we define

$$(S(t)f)(s) := f(t + s) \quad \text{for } s \in \mathbb{R},\ t \geq 0.$$

Then $S(t)$ is a linear isometry on $\mathrm{L}^p(\mathbb{R})$, moreover, S has the semigroup property. We call S the **left shift semigroup** on $\mathrm{L}^p(\mathbb{R})$.

Proposition 2.12. *For $p \in [1, \infty)$ the left shift semigroup S is strongly continuous on $\mathrm{L}^p(\mathbb{R})$.*

To identify the generator of S we first define

$$\begin{aligned}\mathrm{W}^{1,p}(\mathbb{R}) := \big\{ f \in \mathrm{L}^p(\mathbb{R}) : f \text{ is continuous},\\ \text{there exists } g \in \mathrm{L}^p(\mathbb{R}) \text{ with } f(t) - f(0) = \textstyle\int_0^t g(s)\,\mathrm{d}s \text{ for } t \in \mathbb{R} \big\}.\end{aligned}$$

Note that $\mathrm{W}^{1,p}(\mathbb{R})$ is a linear subspace of $\mathrm{L}^p(\mathbb{R})$ and for $f \in \mathrm{W}^{1,p}(\mathbb{R})$ the L^p function g as in the definition exists uniquely. We call it the **derivative** of f, and use the notation $f' := g$. In fact, the function f is almost everywhere differentiable and it derivative equals g almost everywhere. We define a norm on $\mathrm{W}^{1,p}(\mathbb{R})$ by

$$\|f\|^p_{\mathrm{W}^{1,p}} := \|f\|^p_p + \|f'\|^p_p.$$

It is not hard to see that this turns $\mathrm{W}^{1,p}(\mathbb{R})$ into a Banach space.

Proposition 2.13. *The generator A of the left shift semigroup S on* $\mathrm{L}^p(\mathbb{R})$ *is given by*

$$D(A) = \mathrm{W}^{1,p}(\mathbb{R}), \quad Af = f'.$$

The proof is left as Exercise 5.

We now turn to more complicated shifts with boundary conditions. Consider the Banach space $\mathrm{L}^p(0,1)$. For $t \geq 0$ and $f \in \mathrm{L}^p(\mathbb{R})$ define

$$S_0(t)f(s) := \begin{cases} f(t+s) & \text{if } s \in [0,1],\ t+s \leq 1, \\ 0 & \text{if } s \in [0,1],\ t+s > 1. \end{cases}$$

It is easy to see that $S_0(t)$ is a bounded linear operator and that S_0 has the semigroup property. For $t \geq 1$ we have $S_0(t) = 0$, hence $S_0(t)^n = 0$ for $t > 0$ and $n \in \mathbb{N}$ wit $n > \frac{1}{t}$, i.e., $S_0(t)$ is a nilpotent operator. That is why S_0 is called the **nilpotent left shift** on L^p(0,1).

Proposition 2.14. *The nilpotent left shift S_0 is a strongly continuous semigroup on* $\mathrm{L}^p(0,1)$.

We want to identify the generator of S_0. For this purpose we define

$$\begin{aligned} \mathrm{W}^{1,p}_{(0)}(0,1) := \big\{ f \in \mathrm{L}^p(0,1) : {} & f \text{ is continuous on } [0,1], \\ & \text{there exists } g \in \mathrm{L}^p(0,1) \text{ with } f(t) - f(0) = \textstyle\int_0^t g(s)\,\mathrm{d}s \text{ for } t \in [0,1], \\ & \text{and } f(1) = 0 \big\}. \end{aligned}$$

Similarly to the above, every $f \in \mathrm{W}^{1,p}_{(0)}(0,1)$ has a derivative $f' \in \mathrm{L}^p(0,1)$, and we can define a norm on $\mathrm{W}^{1,p}_{(0)}(0,1)$ by

$$\|f\|^p_{\mathrm{W}^{1,p}_{(0)}} := \|f\|^p_p + \|f'\|^p_p,$$

making it a Banach space.

Proposition 2.15. *The generator A of the nilpotent left shift S_0 on* $\mathrm{L}^p(0,1)$ *is given by*

$$D(A) = \mathrm{W}^{1,p}_{(0)}(0,1), \quad Af = f'.$$

The proof of these results is left as Exercise 6.

The Gaussian semigroup

Consider again the heat equation, but now on the entire $\mathbb{R}$:

$$\begin{aligned} \partial_t w(t,x) &= \partial_{xx} w(t,x), \quad t \geq 0,\ x \in \mathbb{R} \\ w(0,x) &= w_0(x), \quad x \in \mathbb{R}. \end{aligned} \tag{2.1}$$

Here w_0 is a function on $\mathbb{R}$ providing the initial heat profile. We follow the rule of thumb of Lecture 1 and seek the solution to this problem as an orbit map of some *semigroup*. To find a candidate for this semigroup we first make some formal computations by using the Fourier transform, which is given for $f \in \mathrm{L}^1(\mathbb{R})$ by the Fourier integral

$$\widehat{f}(\xi) := \mathscr{F}(f)(\xi) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\xi x} f(x)\,\mathrm{d}x.$$

(We remark that with some extra work the next arguments can be made precise.) Recall that $\mathscr{F}$ maps differentiation to multiplication by the Fourier variable $\mathrm{i}\xi$, i.e., $\mathscr{F}(\partial_x f(x))(\xi) = \mathrm{i}\xi\mathscr{F}(f)(\xi)$. If we take the Fourier transform of equation (2.1) with respect to x and interchange the actions of $\mathscr{F}$ and ∂_t, we obtain

$$\begin{aligned} \partial_t \widehat{w}(t,\xi) &= -\xi^2 \widehat{w}(t,\xi) \quad t \geq 0,\ \xi \in \mathbb{R} \\ \widehat{w}(0,\xi) &= \widehat{w}_0(\xi), \quad \xi \in \mathbb{R}. \end{aligned}$$

This is an ODE for $\widehat{w}$, which is easy to solve:

$$\widehat{w}(t,\xi) = \mathrm{e}^{-t|\xi|^2}\widehat{w}_0(\xi).$$

To get w back we take the inverse Fourier transform of this solution:

$$w(t,\cdot) = \mathscr{F}^{-1}(\widehat{w}(t,\cdot)) = \frac{1}{\sqrt{2\pi}}\mathscr{F}^{-1}(\mathrm{e}^{-t|\cdot|^2}) * \mathscr{F}^{-1}(\widehat{w}_0),$$

where we used that $\mathscr{F}^{-1}$ maps products to convolutions. At this point we only have to remember that

$$\mathscr{F}^{-1}(\mathrm{e}^{-t|\cdot|^2})(x) = \frac{1}{\sqrt{2t}}\mathrm{e}^{-\frac{|x|^2}{4t}}.$$

So if we set

$$g_t(x) := \frac{1}{\sqrt{4\pi t}}\mathrm{e}^{-\frac{|x|^2}{4t}} \quad (t > 0),$$

then the candidate for the solution to (2.1) takes the form

$$w(t) = g_t * w_0 \quad \text{for } t > 0.$$

Let us collect some properties of the function g_t.

Remark 2.16. 1. Consider the **standard Gaussian function**

$$g(x) := \frac{1}{\sqrt{4\pi}}\mathrm{e}^{-\frac{x^2}{4}}.$$

Then $g \geq 0$, $\|g\|_1 = 1$ and g belongs to $\mathrm{L}^p(\mathbb{R})$ for all $p \in [1,\infty]$.

2. We have $g_t(x) = \frac{1}{\sqrt{t}}g\big(\frac{x}{\sqrt{t}}\big)$, hence $g_t \geq 0$, $\|g_t\|_1 = 1$ and

$$\lim_{t\searrow 0} \int\limits_{|x|>r} g_t(s)\,\mathrm{d}s = 0 \quad \text{for all } r > 0 \text{ fixed.}$$

The function

$$G(t,x,y) := g_t(x-y) \quad (t > 0,\ x \in \mathbb{R},\ y \in \mathbb{R})$$

is called the **heat** or **Gaussian kernel** on $\mathbb{R}$ and gives rise to a semigroup, called the **heat** or **Gaussian semigroup**.

Proposition 2.17. *Let $p \in [1,\infty)$. For $f \in \mathrm{L}^p(\mathbb{R})$ and $t > 0$ define*

$$(T(t)f)(x) := (g_t * f)(x) = \frac{1}{\sqrt{4\pi t}}\int\limits_{\mathbb{R}} f(y)\mathrm{e}^{-\frac{(x-y)^2}{4t}}\,\mathrm{d}y = \int\limits_{\mathbb{R}} f(y)G(t,x,y)\,\mathrm{d}y,$$

and set $\qquad T(0)f := f.$

Then $T(t)$ is a linear contraction on $\mathrm{L}^p(\mathbb{R})$, and T is a strongly continuous semigroup.

Proof. Let $f \in \mathrm{L}^p(\mathbb{R})$. By Young's inequality and since $g_t \in \mathrm{L}^1(\mathbb{R})$, we obtain that the convolution $g_t * f$ exists and

$$\|g_t * f\|_p \leq \|g_t\|_1 \cdot \|f\|_p = \|f\|_p.$$

In particular, $g_t * f$ belongs to $\mathrm{L}^p(\mathbb{R})$. Since linearity of $f \mapsto g_t * f$ is obvious, we obtain that $T(t)$ is a linear contraction.

To prove the semigroup property, we employ the Fourier transform. To this end fix $f \in \mathrm{L}^1(\mathbb{R}) \cap \mathrm{L}^p(\mathbb{R})$. Then we can take the Fourier transform of $g_t * (g_s * f)$, and we obtain

$$\mathscr{F}(g_t * (g_s * f)) = \sqrt{2\pi}\mathscr{F}(g_t) \cdot \mathscr{F}(g_s * f) = (2\pi)\mathscr{F}(g_t) \cdot \mathscr{F}(g_s) \cdot \mathscr{F}(f)$$

(we use here that $\mathscr{F}$ maps convolution to product). Recall from the above that

$$\mathscr{F}(g_t)(\xi) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-t\xi^2}, \quad \text{therefore,} \quad \mathscr{F}(g_t)(\xi) \cdot \mathscr{F}(g_s)(\xi) = \frac{1}{2\pi}\mathrm{e}^{-(t+s)\xi^2} = \frac{1}{\sqrt{2\pi}}\mathscr{F}(g_{t+s})(\xi).$$

This yields

$$\mathscr{F}(g_t * (g_s * f)) = \sqrt{2\pi}\mathscr{F}(g_{t+s}) \cdot \mathscr{F}(f) = \mathscr{F}(g_{t+s} * f),$$

hence $g_t * (g_s * f) = g_{t+s} * f$. Therefore, $T(t)T(s)f = T(t+s)f$ holds for $f \in \mathrm{L}^1(\mathbb{R}) \cap \mathrm{L}^p(\mathbb{R})$. By the continuity of the semigroup operators and by the denseness of this subspace in L^p, we obtain the equality everywhere.

From the properties of g_t listed in Remark 2.16.2, it follows that $g_t * f \to f$ in $\mathrm{L}^p(\mathbb{R})$ if $t \searrow 0$. Hence the semigroup T is strongly continuous. □

2.4 Powers of generators

It is a crucial ingredient in the definition of the infinitesimal generator A of a strongly continuous semigroup T that $D(A)$ consists precisely of those elements f for which the orbit map $u(t) = T(t)f$ is differentiable. One expects that if even Af belongs to $D(A)$, then u is *twice* continuously differentiable. This, indeed follows from Proposition 2.8.b):

$$\dot{u}(t) = Au(t) = AT(t)f = T(t)Af,$$

hence $\dot{u}$ is a differentiable function if $Af \in D(A)$. This motivates the next construction.

We set $D(A^0) = X$ and $A^0 = I$, and for $n \in \mathbb{N}$ we define

$$\begin{aligned} D(A^n) &:= \bigl\{f \in D(A^{n-1}) : A^{n-1}f \in D(A)\bigr\}, \\ A^n f &= AA^{n-1}f \quad \text{for } f \in D(A^n) \end{aligned}$$

by recursion. Then $D(A^1)$ is just an alternative notation for $D(A)$. These are all linear subspaces of X, and by intersecting them we introduce

$$D(A^\infty) := \bigcap_{n \in \mathbb{N}} D(A^n).$$

These spaces line up in a hierarchy

$$X = D(A^0) \supseteq D(A) \supseteq D(A^2) \supseteq \cdots \supseteq D(A^n) \supseteq \cdots \supseteq D(A^\infty).$$

The space $D(A^n)$ consists of those elements for which the orbit map is n-times continuously differentiable. Are there such (nonzero) vectors at all? Yes, there are, and actually quite many:

Proposition 2.18. *Let A be a generator of a semigroup. Then for $n \in \mathbb{N}$ the spaces $D(A^n)$ and $D(A^\infty)$ are dense in X.*

Proof. Since $D(A^\infty)$ is contained in $D(A^n)$, it suffices to prove the assertions for the former only. To do that we need some preparations. Let $f \in X$ be fixed. For a smooth function φ with compact support, $\operatorname{supp}(\varphi) \subseteq (0,\infty)$, define

$$f_\varphi := \int_0^\infty \varphi(s)T(s)f \,\mathrm{d}s.$$

We first show that $f_\varphi \in D(A)$. For $h > 0$ we can write

$$\begin{aligned}\frac{T(h)f_\varphi - f_\varphi}{h} &= \frac{1}{h}\int_0^\infty \varphi(s)T(h+s)f\,\mathrm{d}s - \frac{1}{h}\int_0^\infty \varphi(s)T(s)f\,\mathrm{d}s\\ &= \frac{1}{h}\int_h^\infty \varphi(s-h)T(s)f\,\mathrm{d}s - \frac{1}{h}\int_0^\infty \varphi(s)T(s)f\,\mathrm{d}s\\ &= \int_h^\infty \frac{\varphi(s-h)-\varphi(s)}{h}T(s)f\,\mathrm{d}s - \frac{1}{h}\int_0^h \varphi(s)T(s)f\,\mathrm{d}s.\end{aligned}$$

If we let $h \searrow 0$, then the second term converges to $\varphi(0)T(0)f = 0$, while the first term has the limit

$$-\int_0^\infty \varphi'(s)T(s)f\,\mathrm{d}s.$$

This yields $f_\varphi \in D(A)$ and $Af_\varphi = f_{-\varphi'}$. We conclude $f_\varphi \in D(A^\infty)$ by induction.

We turn to the actual proof and suppose in addition to the above that $\varphi \geq 0$ and that $\int_0^\infty \varphi(s)\mathrm{d}s = 1$. We set $\varphi_n(s) = n\varphi(ns)$ and $f_n := f_{\varphi_n}$. For given $\varepsilon > 0$ we choose a $\delta > 0$ such that $\|T(s)f - f\| \leq \varepsilon$ holds for all $s \in [0,\delta]$. If $n \in \mathbb{N}$ is sufficiently large, then $\operatorname{supp}\varphi_n \subseteq (0,\delta)$, hence we obtain

$$\begin{aligned}\|f_n - f\| &= \Big\|\int_0^\infty \varphi_n(s)T(s)f\,\mathrm{d}s - f\Big\| = \Big\|\int_0^\infty \varphi_n(s)T(s)f\,\mathrm{d}s - f\int_0^\infty \varphi_n(s)\,\mathrm{d}s\Big\|\\ &= \Big\|\int_0^\infty \varphi_n(s)\big(T(s)f - f\big)\,\mathrm{d}s\Big\| \leq \int_0^\infty \varphi_n(s)\big\|T(s)f - f\big\|\,\mathrm{d}s\\ &\leq \sup_{s\in[0,\delta]}\|T(s)f - f\| \cdot \int_0^\delta \varphi_n(s)\,\mathrm{d}s \leq \varepsilon.\end{aligned}$$

This shows that $f_n \to f$ in X. □

Since the generator A is closed its domain $D(A)$ is a Banach space with the graph norm $\|\cdot\|_A$. Is any of the spaces $D(A^n)$ dense in this Banach space? To answer this question, we first introduce the following general notion.

Definition 2.19. A subspace D of the domain $D(A)$ of a linear operator $A : D(A) \subseteq X \to X$ is called a **core** for A if D is dense in $D(A)$ for the *graph norm*, defined by

$$\|f\|_A = \|f\| + \|Af\| .$$

We shall often use the following result stating that dense invariant subspaces are dense also in the graph norm.

Proposition 2.20. *Let A be the generator of a semigroup T, and let D be a linear subspace of $D(A)$ that is $\|\cdot\|$-dense in X and invariant under the semigroup operators $T(t)$. Then D is a core for A.*

Proof. For $f \in D(A)$ we prove that f belongs to the $\|\cdot\|_A$-closure of D. First, we take a sequence $(f_n) \subset D$ such that $f_n \to f$ in X. Since for each n the maps

$$s \mapsto T(s)f_n \in D \quad \text{and} \quad s \mapsto AT(s)f_n = T(s)Af_n \in X$$

are continuous, the map $s \mapsto T(s)f_n \in D$ is even continuous for the graph norm $\|\cdot\|_A$. From this it follows that the Riemann integral

$$\int_0^t T(s)f_n \,\mathrm{d}s$$

belongs to the $\|\cdot\|_A$-closure of D (use approximating Riemann sums!). Similarly, the $\|\cdot\|_A$-continuity of $s \mapsto T(s)f$ for $f \in D(A)$ implies that

$$\Big\|\frac{1}{t}\int_0^t T(s)f \,\mathrm{d}s - f\Big\|_A \to 0 \qquad \text{as } t \searrow 0 \text{ and}$$

and $$\Big\|\frac{1}{t}\int_0^t T(s)f_n \,\mathrm{d}s - \frac{1}{t}\int_0^t T(s)f \,\mathrm{d}s\Big\|_A \to 0 \qquad \text{as } n \to \infty \text{ and for each } t > 0.$$

This proves that for every $\varepsilon > 0$ we can find $t > 0$ and $n \in \mathbb{N}$ such that

$$\Big\|\frac{1}{t}\int_0^t T(s)f_n \,\mathrm{d}s - f\Big\|_A < \varepsilon. \qquad \square$$

We now can easily answer the question from the above.

Proposition 2.21. *Let A be a generator of a semigroup. Then each of the spaces $D(A^n)$ for $n \in \mathbb{N}$ and $D(A^\infty)$ is a core for A.*

Proof. All the spaces occurring in the assertion are invariant under $T(t)$, and by Proposition 2.18 they are dense in X. Hence the assertion follows from Proposition 2.20. $\square$

2.5 Resolvent of generators

We saw in Lecture 1 that spectral analysis, more precisely, the determination of eigenvalues and eigenfunctions of the Dirichlet Laplacian led to a construction of the semigroup generated by this operator. We conclude this lecture by some basic spectral properties of semigroup generators. Let us begin with the following fundamental spectral theoretic notions.

Definition 2.22. Let A be a closed linear operator defined on a linear subspace $D(A)$ of a Banach space X.

a) The **spectrum** of A is the set

$$\sigma(A) := \big\{\lambda \in \mathbb{C} : \lambda - A : D(A) \to X \text{ is not bijective}\big\}.$$

b) The **resolvent set** of A is $\rho(A) := \mathbb{C} \setminus \sigma(A)$, i.e.,

$$\rho(A) := \big\{\lambda \in \mathbb{C} : \lambda - A : D(A) \to X \text{ is bijective}\big\}.$$

c) If $\lambda \in \rho(A)$ then $\lambda - A$ is injective, hence has an algebraic inverse $(\lambda - A)^{-1}$. We call this operator the resolvent of A at point λ and denote it by

$$R(\lambda, A) := (\lambda - A)^{-1}.$$

Note that if $\lambda \in \rho(A)$, the operator $\lambda - A$ is both injective and surjective, i.e., its algebraic inverse

$$(\lambda - A)^{-1} : X \to D(A)$$

is defined on the entire X. Since A is closed so are $\lambda - A$ and its inverse. As consequence of the closed graph theorem, see Supplement, Theorem 2.32, we immediately obtain that $(\lambda - A)^{-1}$ is bounded.

Proposition 2.23. *For a closed linear operator A and for $\lambda \in \rho(A)$ we have*

$$(\lambda - A)^{-1} = R(\lambda, A) \in \mathscr{L}(X).$$

Let us recall also the next fundamental properties of spectrum and the resolvent.

Proposition 2.24. *Let X be a Banach space and let A be a closed linear operator with domain $D(A) \subseteq X$. Then the following assertions are true:*

a) The resolvent set $\rho(A)$ is open, hence its complement, the spectrum $\sigma(A)$ is closed.

b) The mapping

$$\rho(A) \ni \lambda \mapsto R(\lambda, A) \in \mathscr{L}(X)$$

is complex differentiable. Moreover, for $n \in \mathbb{N}$ we have

$$\frac{\mathrm{d}^n}{\mathrm{d}\lambda^n} R(\lambda, A) = (-1)^n n! R(\lambda, A)^{n+1}.$$

Proof. Statement a) follows from the following Neumann series representation of the resolvent: For $\mu \in \rho(A)$ with $|\lambda - \mu| < \frac{1}{\|R(\mu,A)\|}$, we have

$$R(\lambda, A) = \sum_{k=0}^{\infty} (\lambda - \mu)^k R(\mu, A)^{k+1}.$$

Assertion b) follows from the power series representation in a) and from the fact that a power series is always a Taylor series. □

To prove that the resolvent set of a generator A is non-empty, and to relate the resolvent of A to the semigroup T, the first step is provided by the next lemma.

Lemma 2.25. *Let T be a strongly continuous semigroup with generator A. Then for all $\lambda \in \mathbb{C}$ and $t > 0$ the following identities hold:*

$$\begin{aligned} \mathrm{e}^{-\lambda t}T(t)f - f &= (A-\lambda)\int_0^t \mathrm{e}^{-\lambda s}T(s)f\,\mathrm{d}s \qquad \text{if } f \in X,\\ &= \int_0^t \mathrm{e}^{-\lambda s}T(s)(A-\lambda)f\,\mathrm{d}s \qquad \text{if } f \in D(A). \end{aligned}$$

Proof. Observe that $S(t) = \mathrm{e}^{-\lambda t}T(t)$ is also a strongly continuous semigroup with generator $B = A - \lambda$, see Exercise 3 b). Hence, we can apply Proposition 2.9.b). □

With the help of this lemma we obtain the next, important relations between the resolvent of the generator and the semigroup.

Proposition 2.26. *Let T be a strongly continuous semigroup of type (M, ω) with generator A. Then the following assertions are true:*

a) For all $f \in X$ and $\lambda \in \mathbb{C}$ with $\operatorname{Re}\lambda > \omega$ we have

$$R(\lambda, A)f = \int_0^\infty \mathrm{e}^{-\lambda s}T(s)f\,\mathrm{d}s = \lim_{N\to\infty}\int_0^N \mathrm{e}^{-\lambda s}T(s)f\,\mathrm{d}s.$$

b) For all $f \in X$, $\lambda \in \mathbb{C}$ with $\operatorname{Re}\lambda > \omega$ and $n \in \mathbb{N}$ we have

$$R(\lambda, A)^n f = \frac{1}{(n-1)!}\int_0^\infty s^{n-1}\mathrm{e}^{-\lambda s}T(s)f\,\mathrm{d}s.$$

c) For all $\lambda \in \mathbb{C}$ with $\operatorname{Re}\lambda > \omega$ we have

$$\|R(\lambda, A)^n\| \le \frac{M}{(\operatorname{Re}\lambda - \omega)^n}. \tag{2.2}$$

Proof. From Lemma 2.25 and by taking limit as $t \to \infty$ we conclude that for $\operatorname{Re}\lambda > \omega$ we have

$$\begin{aligned} -f &= (A-\lambda)\int_0^\infty \mathrm{e}^{-\lambda s}T(s)f\,\mathrm{d}s \qquad \text{if } f \in X,\\ &= \int_0^\infty \mathrm{e}^{-\lambda s}T(s)(A-\lambda)f\,\mathrm{d}s \qquad \text{if } f \in D(A). \end{aligned}$$

Since this expression gives a bounded operator, a) is proved. To show b), notice that

$$R(\lambda, A)^n f = \frac{(-1)^{n-1}}{(n-1)!}\frac{\mathrm{d}^{n-1}}{\mathrm{d}\lambda^{n-1}}R(\lambda, A)f = \frac{1}{(n-1)!}\int_0^\infty s^{n-1}\mathrm{e}^{-\lambda s}T(s)f\,\mathrm{d}s.$$

Finally, to see c) we make a norm estimate and obtain

$$\begin{aligned}\|R(\lambda,A)^n f\| &\le \frac{1}{(n-1)!}\int_0^\infty s^{n-1}\mathrm{e}^{-\operatorname{Re}\lambda s} M\mathrm{e}^{\omega s}\|f\|\,\mathrm{d}s \le \frac{M\|f\|}{(n-1)!}\int_0^\infty s^{n-1}\mathrm{e}^{(\omega-\operatorname{Re}\lambda)s}\,\mathrm{d}s\\ &= \frac{M}{(\operatorname{Re}\lambda-\omega)^n}\|f\|. \end{aligned}$$ □

Let us summarise the above as follows.

Conclusion 2.27. If A is the generator of an operator semigroup T, then it is closed, densely defined, and a suitable right half plane belongs to its resolvent set, where the estimate (2.2) holds. The resolvent operators are given by the **Laplace transform** of the semigroup.

2.6 Supplement

We collect here some standard results used in this lecture.

The strong operator topology

At this point, we do not want to give the definition of the strong operator topology, but just point out what *convergence* and *boundedness* mean in this setting.

Let X, Y be Banach spaces and let $(T_n) \subseteq \mathscr{L}(X,Y)$ be a sequence of bounded linear operators between X and Y. We say that the sequence (T_n) **converges strongly** to $T \in \mathscr{L}(X,Y)$, if

$$T_n x \to Tx \quad \text{holds in } Y \text{ as } n\to\infty \text{ for all } x \in X.$$

For the purposes of this course, this is the correct notion of convergence, being, as a matter of fact, nothing else than pointwise convergence.

A subset $\mathscr{K} \subseteq \mathscr{L}(X,Y)$ is called **strongly bounded** (or bounded poinwise) if for all $x \in X$ we have

$$\sup\bigl\{\|Tx\| : T \in \mathscr{K}\bigr\} < \infty.$$

Next, we list some classical functional analysis results concerning these two notions.

Theorem 2.28 (Uniform Boundedness Principle)**.** *Let X, Y be Banach spaces and suppose $\mathscr{K} \subseteq \mathscr{L}(X,Y)$ is strongly bounded, i.e., for all $x \in X$ we have*

$$\sup\bigl\{\|Tx\| : T \in \mathscr{K}\bigr\} < \infty.$$

*Then $\mathscr{K}$ is **uniformly bounded** that is*

$$\sup\bigl\{\|T\| : T \in \mathscr{K}\bigr\} < \infty.$$

This theorem has the following important consequence:

Theorem 2.29. *Let X, Y be Banach spaces, and let $(T_n) \subseteq \mathscr{L}(X,Y)$ be a sequence such that $(T_n x) \subseteq Y$ converges for all $x \in X$. Then*

$$Tx := \lim_{n\to\infty} T_n x$$

defines a bounded linear operator on X.

Theorem 2.30. *Let X, Y be Banach spaces, let $T \in \mathscr{L}(X,Y)$ and let $(T_n) \subseteq \mathscr{L}(X,Y)$ be a norm bounded sequence. Then the following assertions are equivalent:*

(i) For every $x \in X$ we have $T_n x \to Tx$ in X.

(ii) There is a dense subspace $D \subseteq X$ such that for all $x \in X$ we have $T_n x \to Tx$ in X.

(iii) For every compact set $K \subseteq X$ we have $T_n x \to Tx$ in X uniformly for $x \in K$.

By adapting the classical proof of the product rule of differentiation and by making use of the theorem above one can easily prove next result.

Theorem 2.31 (Product rule)**.** *Let $u : [a,b] \to X$ be differentiable, and let $F : [a,b] \to \mathscr{L}(X,Y)$ be strongly continuous such that for every $t \in [a,b]$ the mapping*

$$Fu : s \mapsto F(s)u(t) \in Y$$

is differentiable. Then $s \mapsto F(s)u(s) \in Y$ is differentiable, and we have

$$\tfrac{\mathrm{d}}{\mathrm{d}t}(Fu)(t) = \tfrac{\mathrm{d}}{\mathrm{d}t}F(t) \cdot u(t) + F(t) \cdot \tfrac{\mathrm{d}}{\mathrm{d}t}u(t),$$

where $\tfrac{\mathrm{d}}{\mathrm{d}t}F(t) \cdot u(t)$ denotes the derivative of $s \mapsto F(s)u(t)$ at $s = t$.

The last result we wish to recall from functional analysis is the closed graph theorem.

Theorem 2.32 (Closed Graph Theorem)**.** *Let X be a Banach space, and let $A : X \to Y$ be a closed and linear operator with dense domain $D(A)$ in X. Then A is bounded if and only if $D(A) = X$.*

The Riemann integral

Denote by $\mathrm{C}([a,b];X)$ the space of continuous X-valued functions on $[a,b]$, which becomes a Banach space with the supremum norm. For a continuous function $u \in \mathrm{C}([a,b];X)$ we define its **Riemann integral** by approximation through Riemann sums. Let us briefly sketch the idea how to do this. For $P = \{a = t_1 < t_2 < \cdots < t_n = b\} \subseteq [a,b]$ we set

$$\delta(P) = \max\bigl\{t_{j+1} - t_j : j = 0, \dots, n-1\bigr\},$$

and call P a **partition** of $[a,b]$ and $\delta(P)$ the **mesh** of P. We define the **Riemann sum** of u corresponding to the partition P by

$$S(P,u) := \sum_{j=0}^{n-1} u(t_j)(t_{j+1} - t_j),$$

where n is the number of elements in P. From the uniform continuity of u on the compact interval $[a,b]$ it follows that there exists $x_0 \in X$ such that $S(P,u)$ converges to x_0 if $\delta(P) \to 0$. More precisely, for all $\varepsilon > 0$ there is $\delta > 0$ such that

$$\|S(P,u) - x_0\| < \varepsilon$$

whenever $\delta(P) < \delta$. We call this $x_0 \in X$ the Riemann integral of f an denote it by

$$\int\limits_a^b u(s)\,\mathrm{d}s.$$

The Riemann integral enjoys all the usual properties known for scalar valued functions. Some of them are collected in the next proposition.

Proposition 2.33. *a) There is a sequence of Riemann sums $S(P_n, u)$ with $\delta(P_n) \to 0$ converging to the Riemann integral of u.*

b) The Riemann integral is a bounded linear operator on the space $\mathrm{C}([a, b]; X)$ with values in X.

c) If $T \in \mathscr{L}(X, Y)$, then

$$T \int_a^b u(s)\,\mathrm{d}s = \int_a^b Tu(s)\,\mathrm{d}s.$$

d) If $u : [a, b] \to X$ is continuous, then

$$v(t) := \int_0^t u(s)\,\mathrm{d}s$$

is differentiable with derivative u.

e) If $u : [a, b] \to X$ is continuously differentiable, then

$$u(b) - u(a) = \int_a^b u'(s)\,\mathrm{d}s$$

holds.

For the proof of these assertions one can take the standard route valid for scalar-valued functions.

Exercises

1. For $A \in \mathscr{L}(X)$ and $t \geq 0$ define

$$T(t) = \mathrm{e}^{tA} := \sum_{n=0}^{\infty} \frac{t^n A^n}{n!}.$$

Prove that T is strongly continuous semigroup, which is even continuous for the operator norm on $[0, \infty)$ and consists of continuously invertible operators. Determine its generator.

2. Give an example of a Hilbert space and a bounded (i.e., of type $(M, 0)$) strongly continuous semigroup thereon which is not a contraction.

3. a) For a strongly continuous semigroup T and an invertible transformation R define $S(t) := R^{-1}T(t)R$. Prove that S is a strongly continuous semigroup as well. Determine its growth bound and its generator.

b) For a strongly continuous semigroup T and $z \in \mathbb{C}$ define $S(t) := \mathrm{e}^{tz}T(t)$. Prove that S is a strongly continuous semigroup, determine its growth bound and its generator.

c) For a strongly continuous semigroup T and $\alpha \geq 0$ define $S(t) := T(\alpha t)$. Prove that S is a strongly continuous semigroup, determine its growth bound and its generator.

4. Give an example of strongly continuous semigroup T with $\omega_0(T) = -\infty$ but with $M_\omega \geq 2$ for all exponents $\omega \in \mathbb{R}$.

5. Prove Proposition 2.13.

6. Prove Propositions 2.14 and 2.15.

7. Consider the closed subspace

$$\mathrm{C}_{(0)}([0,1]) := \bigl\{f \in \mathrm{C}([0,1]) : f(1) = 0\bigr\}$$

of the Banach space $\mathrm{C}([0,1])$ of continuous functions on $[0,1]$. Define the nilpotent left shift semigroup thereon and determine its generator.

8. Determine the generator of the Gaussian semigroup on $\mathrm{L}^2(\mathbb{R})$ from Section 2.3.

9. Let $p \in [1,\infty)$ and consider the Gaussian semigroup T on $\mathrm{L}^p(\mathbb{R})$. Prove that for all $t > 0$ and $r \in [p,\infty]$ the operator $T(t)$ is bounded from L^p to $\mathrm{L}^r(\mathbb{R})$.

Lecture 3

Approximation of Semigroups – Part 1

The main topic of this lecture will be to establish approximation theorems for operator semigroups. Consider the abstract initial value problem (Cauchy problem)

$$\begin{cases} \dot{u}(t) = Au(t), \quad t \geq 0 \\ u(0) = u_0 \in X, \end{cases} \tag{ACP}$$

where we suppose that A generates the strongly continuous semigroup T on X. In many applications we are able to construct a sequence of approximating operators A_n which generate strongly continuous semigroups T_n and converge to A in some sense. The question is if this implies the convergence of the semigroups T_n to T?

These approximations usually involve some numerical methods: Either some approximation of the operator A (for example finite differences, see Example 3.7 below), or an approximation of the solution u of a stationary problem $Au = f$ (for example finite element or spectral method, see Example 3.6).

After working through the examples and the exercises, we will see that operator norm convergence would be simply too much to expect, and to weaken this type of convergence, the pointwise one is our next bet. Therefore, in this lecture we shall investigate the *strong convergence* of semigroups, i.e., the property

$$T_n(t)f \to T(t)f \quad \text{as } n \to \infty \quad \text{for all } f \in X \tag{3.1}$$

uniformly for t in compact intervals of $[0, \infty)$.

Remark 3.1. If the convergence stated above holds for the semigroups T_n, T, then the uniform boundedness principle, Theorem 2.28, immediately implies that $\|T_n(t)\|$ has to remain bounded as $n \to \infty$ for all $t \geq 0$. More is true: There exist constants $M \geq 1$, $\omega \in \mathbb{R}$ such that

$$\|T_n(t)\| \leq M\mathrm{e}^{\omega t} \qquad \text{holds for all } n \in \mathbb{N},\ t \geq 0. \tag{3.2}$$

We leave the proof as Exercise 1. This exponential inequality, called *stability condition*, provides a necessary condition to have convergence of the semigroups as in (3.1).

3.1 Generator approximations

Usually, after discretising a differential operator, we end up with a matrix, hence not an operator on the original space, but rather on $\mathbb{C}^n$. So the next important point is that we not only have approximating operators, but also approximating spaces. This motivates our general setup.

Assumption 3.2. Let X_n, X be Banach spaces and assume that there are bounded linear operators $P_n : X \to X_n$, $J_n : X_n \to X$ with the following properties:

- There is a constant $K > 0$ with $\|P_n\|$, $\|J_n\| \leq K$ for all $n \in \mathbb{N}$,

- $P_n J_n = I_n$, the identity operator on X_n, and
- $J_n P_n f \to f$ as $n \to \infty$ for all $f \in X$.

An important remark on our notation. The symbol $\|\cdot\|$ refers here to the operator norm in $\mathscr{L}(X, X_n)$ and $\mathscr{L}(X_n, X)$, respectively. We use the convention that if it is clear from the context, we often do not distinguish in the notation between the norms on different spaces.

Example 3.3 (spectral method). Consider the spaces $X = \ell^2$ and $X_n = \mathbb{C}^n$ with the Euclidian norm and define for $f = (f_k) \in \ell^2$ the operator

$$P_n : \ell^2 \to \mathbb{C}^n, \quad P_n f := (f_1, \ldots, f_n),$$

and for $y = (y_1, \ldots, y_n) \in \mathbb{C}^n$ the operator

$$J_n : \mathbb{C}^n \to \ell^2, \quad J_n(y_1, y_2, \ldots, y_n) := (y_1, \ldots, y_n, 0, \ldots).$$

Clearly, $J_n P_n$ equals the projection onto the first n coordinates. For this example all the above mentioned properties are satisfied, in particular, the last one because

$$\|J_n P_n f - f\|^2 = \sum_{k=n+1}^{\infty} |f_k|^2 \to 0 \text{ as } n \to \infty.$$

This is essentially the same example as if we take $X = L^2(0,1)$, $X_n = \mathbb{C}^n$, $P_n f :=$ the first n Fourier coefficients of f, and $J_n(y_1, \ldots, y_n) :=$ the finite trigonometric sum built from the coefficients $y_1, \ldots, y_n$ (*spectral method*, see Appendix A.2).

Example 3.4 (finite difference). In this example we try to capture the standard discretisation of continuous functions through grid points in an abstract way (*finite difference method*, see Appendix A.1). Let

$$X := \{f \in \mathrm{C}([0,1]) : f(1) = 0\} = \mathrm{C}_{(0)}([0,1]) \quad \text{and} \quad X_n = \mathbb{C}^n,$$

both with the respective maximum norm, see also Exercise 7 in Lecture 1. We define

$$(P_n f)_k := f(\tfrac{k}{n}), \quad k = 0, \ldots, n-1,$$

and

$$J_n(y_0, \ldots, y_{n-1}) := \sum_{k=0}^{n-1} y_k B_{n,k}(x),$$

where for $k \in \{0, \ldots n-1\}$ and $x \in [0,1]$ we have

$$B_{n,k}(x) = \begin{cases} n\big(x - \frac{k}{n}\big) & \text{if } x \in \big[\frac{k-1}{n}, \frac{k}{n}\big), \\ n\big(\frac{k+1}{n} - x\big) & \text{if } x \in \big[\frac{k}{n}, \frac{k+1}{n}\big), \\ 0 & \text{otherwise,} \end{cases}$$

the hat functions. Then $P_n J_n = I_{\mathbb{C}^n}$ and for $n \to \infty$ the convergence $J_n P_n f \to f$ hold true, see Exercise 2.

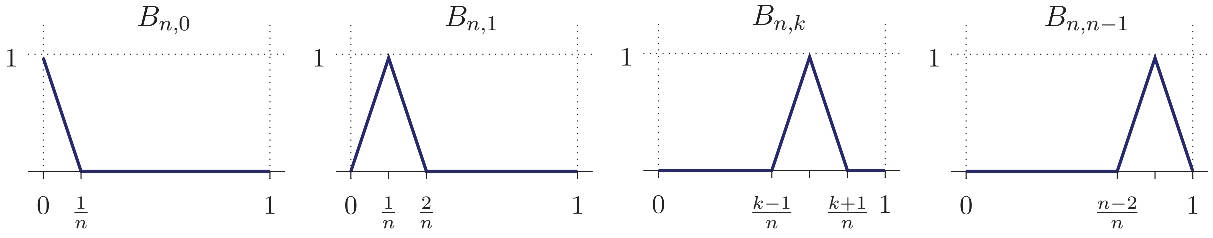

Figure 3.1: The hat functions

After setting the stage for the problems, let us turn our attention back to approximation problems and introduce a general assumption on the convergence of the generators. When examining the examples, we learn that the following setup is quite natural.

Assumption 3.5. Suppose that the operators A_n, A generate strongly continuous semigroups on X_n and X, respectively, and that there are constants $M \geq 0$, $\omega \in \mathbb{R}$ such that the stability condition (3.2) holds. Further suppose that there is a dense subset $Y \subset D(A)$ such that for all $g \in Y$ there is a sequence $y_n \in D(A_n)$ satisfying

$$\|y_n - P_n g\|_{X_n} \to 0 \quad \text{and} \quad \|A_n y_n - P_n A g\|_{X_n} \to 0 \quad \text{as } n \to \infty. \tag{3.3}$$

Clearly, by Assumption 3.2, the convergence stated above is equivalent to

$$\|J_n A_n y_n - Ag\| \to 0 \quad \text{as } n \to \infty$$

for all $g \in Y$. We will freely make use of this equivalence later on, depending on which formulation is more convenient in the given situation. In applications we typically (but not necessarily) have $P_n Y \subset D(A_n)$ and $y_n = P_n g$.

Example 3.6 (spectral method). Taking the setup of Example 3.3 and motivated by the heat equation introduced in Section 1.1, we define the operator on $f = (f_k) \in X$ as

$$Af := (-k^2 f_k)_{k \in \mathbb{N}} \quad \text{with} \quad D(A) = \big\{(f_k) \in X \,:\, (k^2 f_k) \in X\big\}.$$

Further, for $f \in D(A)$, we define

$$A_n(P_n f) := P_n A f, \quad \text{i.e.,} \quad A_n(y_1, y_2, \ldots, y_n) := (-y_1, -4y_2, \ldots, -n^2 y_n).$$

See also Appendix A for some details about the spectral method.

Analogously, motivated by the Schrödinger equation, we define the operator B by

$$Bf := (\mathrm{i}k^2 f_k) \quad \text{for } f = (f_k) \in D(B) := D(A).$$

The approximating operators are then $B_n(y_1, \ldots, y_n) := (\mathrm{i}y_1, \mathrm{i}4y_2, \ldots, \mathrm{i}n^2 y_n)$.

Example 3.7 (finite difference). Continuing Example 3.4, define the generator

$$Af := f' \quad \text{with} \quad D(A) := \big\{f \in \mathrm{C}^1([0,1]) : f(1) = 0\big\}.$$

For $y = (y_0, \ldots y_{n-1}) \in X_n$, we define

$$(A_n y)_k := n(y_{k+1} - y_k) \quad \text{for } k := 0, \ldots n-2 \quad \text{and} \quad (A_n y)_{n-1} := -n y_{n-1},$$

being the standard first-order finite difference scheme. By using that $y = P_n f$, we can write it in a slightly different form:

$$(A_n P_n f)_k := n\big(f(\tfrac{k+1}{n}) - f(\tfrac{k}{n})\big) \quad \text{for } k = 0, \dots, n-1.$$

Then, by the mean value theorem, we obtain

$$\begin{aligned}\|J_n A_n P_n f - Af\|_\infty &= \left\| \sum_{k=0}^{n-1} \frac{f(\frac{k+1}{n}) - f(\frac{k}{n})}{\frac{1}{n}} B_{n,k} - f' \right\|_\infty = \left\| \sum_{k=0}^{n-1} f'(\xi_k) B_{n,k} - f' \right\|_\infty \\ &\leq \left\| f' - \sum_{k=0}^{n-1} f'(\tfrac{k}{n}) B_{n,k} \right\|_\infty + \max_{k=0,\dots,n-1} |f'(\tfrac{k}{n}) - f'(\xi_k)| \\ &\leq \left\| f' - \sum_{k=0}^{n-1} f'(\tfrac{k}{n}) B_{n,k} \right\|_\infty + \omega(f', \tfrac{1}{n}) \to 0 \quad \text{as } n \to \infty.\end{aligned}$$

Here, $\omega(f', s)$ is the modulus of (uniform) continuity of the function f' defined as usual by

$$\omega(f', s) := \sup\big\{|f'(x) - f'(y)| \sup |x - y| \leq s\big\}.$$

An important observation concerning this example is the following. If we assume a bit more regularity and take $f \in \mathrm{C}^2([0,1])$, then we obtain

$$(A_n P_n f - P_n A f)_k = \frac{f(\frac{k+1}{n}) - f(\frac{k}{n})}{\frac{1}{n}} + f'(\tfrac{k}{n}) = f''(\xi_k)\frac{1}{2n},$$

by Taylor's formula, and hence,

$$\|A_n P_n f - P_n A f\| \leq \frac{\|f''\|_\infty}{2n}.$$

This means that, in this example, we not only have convergence, but even *first-order convergence* for twice differentiable functions.

The following Proposition 3.8 will guarantee that for all $f \in \mathrm{C}^2([0,1])$ which remain in C^2 under the semigroup, this convergence carries over to the convergence of the semigroups. More precisely, we have that

for all $\qquad f \in D(A^2) = \big\{f \in \mathrm{C}^2([0,1]) : f(0) = f'(0) = f''(0) = 0\big\}$

and for all $t > 0$ there is $C > 0$ such that

$$\|T_n(t) P_n f - P_n T(t) f\| \leq C \frac{\|f\|_{\mathrm{C}^2}}{n}.$$

Motivated by the previous example, we can formulate our first approximation result on semigroups.

Proposition 3.8. *Suppose that Assumptions 3.2 and 3.5 hold, that $P_n Y \subset D(A_n)$, and that Y is a Banach space invariant under the semigroup T satisfying*

$$\|T(t)\|_Y \leq M e^{\omega t}.$$

If there are constants $C > 0$ and $p \in \mathbb{N}$ with the property that for all $f \in Y$

$$\|A_n P_n f - P_n A f\|_{X_n} \leq C \frac{\|f\|_Y}{n^p},$$

then for all $t > 0$ there is $C' > 0$ such that

$$\|T_n(t) P_n f - P_n T(t) f\|_{X_n} \leq C' \frac{\|f\|_Y}{n^p}.$$

Moreover, this convergence is uniform in t on compact intervals.

In this case we say that we have **convergence of order** p. Also notice that, as discussed in Proposition 2.20, the subspace Y will be a core for the generator A (being dense and invariant under the semigroup), and hence $(\lambda - A)Y \subset X$ will be also a dense set for some/all $\lambda \in \rho(A)$. In many examples we will take $Y := D(A^l)$ for some $l \in \mathbb{N}$.

Proof. For simplicity, we first carry out the proof in the special case $X_n = X$ and $J_n = P_n = I$. It is clear that for $f \in Y$, we have $Af = A_n f + (A - A_n) f$. Application of the fundamental theorem of calculus to the continuously differentiable function $[0, t] \ni s \mapsto T_n(t - s) T(s) f$ implies that the variation of constants formula

$$T(t) f = T_n(t) f + \int_0^t T_n(t - s)(A - A_n) T(s) f \, \mathrm{d}s$$

holds. Therefore, we have

$$\begin{aligned} \|T(t) f - T_n(t) f\| &\leq \int_0^t M \mathrm{e}^{\omega(t-s)} \|(A - A_n) T(s) f\| \, \mathrm{d}s \\ &\leq \int_0^t M \mathrm{e}^{\omega(t-s)} \frac{C}{n^p} \|T(s) f\|_Y \, \mathrm{d}s \leq M^2 \mathrm{e}^{\omega t} \cdot t \cdot \frac{C}{n^p} \|f\|_Y. \end{aligned}$$

From this the assertion follows. The general case can be considered by applying the fundamental theorem of calculus to the modified function $[0, t] \ni s \mapsto T_n(t - s) P_n T(s) f$ to obtain the variation of constants formula

$$P_n T(t) f = T_n(t) P_n f + \int_0^t T_n(t - s) \big(P_n A - A_n P_n\big) T(s) f \, \mathrm{d}s.$$

From here, the argument is the same as above. □

3.2 Resolvent approximations

We will see that in many applications the situation is slightly more complicated than we had before.

Example 3.9 (spectral method). Going back to Example 3.6, we can see quickly that it is difficult and unnatural to obtain similar estimates on the convergence of the generators. But we can immediately infer the following.

If $g = (g_n) \in X$, then there is $f = (f_n) \in D(A)$ such that $g = Af$, $f = A^{-1}g$. Hence,

$$\begin{aligned}\|J_nP_nf - f\| &= \|J_nP_nA^{-1}g - A^{-1}g\| = \|(0,\dots,0,(n+1)^{-2}g_{n+1},\dots)\| \\ &\leq \frac{1}{(n+1)^2}\|J_nP_ng - g\| \leq \frac{1}{(n+1)^2}\|g\|.\end{aligned}$$

Since, by definition, $A_nP_n = P_nA$, this implies that

$$\|J_nA_n^{-1}P_n - A^{-1}\| \leq \frac{1}{(n+1)^2},$$

which means that it is natural to expect the convergence of the resolvent operators.[1]

Hence, in order to infer the convergence of the semigroups, we have to prove the analogue of Proposition 3.8 but now with resolvent convergence.

Proposition 3.10. *Assume that Assumptions 3.2 and 3.5 hold. If there are $C > 0$ and $p \in \mathbb{N}$ such that*

$$\|A_n^{-1}P_n - P_nA^{-1}\| \leq \frac{C}{n^p},$$

then for all $t > 0$ there is $C' > 0$ such that

$$\|T_n(t)P_ng - P_nT(t)g\|_{X_n} \leq C'\frac{\|g\|_{A^2}}{n^p}$$

for all $g \in D(A^2)$. Furthermore, this convergence is uniform in t on compact intervals[2].

A surprising and important observation here is that although we assume operator norm convergence of the resolvents, we do not get back norm convergence of the approximating semigroups in general. This observation will be illustrated in an example before the proof.

Remark 3.11. 1. It is clear that instead of the convergence of A_n^{-1} to A^{-1} we can assume $J_n(\lambda - A_n)^{-1}P_n \to (\lambda - A)^{-1}$ for some $\lambda \in \rho(A) \cap \rho(A_n)$.

2. In concrete cases, as we shall also see in the examples below, these results are far from being sharp. To obtain better results, we also need more structural properties of the approximation.

Example 3.12 (spectral method). We refer to Example 3.6 again, and define the semigroups generated by A and B as

$$T(t)f := \mathrm{e}^{tA}f = (\mathrm{e}^{-t}f_1, \mathrm{e}^{-4t}f_2, \dots, \mathrm{e}^{-k^2t}f_k, \dots)$$

and

$$S(t)f := \mathrm{e}^{tB}f = (\mathrm{e}^{\mathrm{i}t}f_1, \mathrm{e}^{\mathrm{i}4t}f_2, \dots, \mathrm{e}^{\mathrm{i}k^2t}f_k, \dots).$$

The approximating semigroups are $T_n(t) = \operatorname{diag}(\mathrm{e}^{-t}, \mathrm{e}^{-4t}, \dots, \mathrm{e}^{-n^2t})$, that is,

$$J_nT_n(t)P_n = J_nP_nT(t).$$

Similarly, we have $S_n(t) = \operatorname{diag}(\mathrm{e}^{\mathrm{i}t}, \mathrm{e}^{\mathrm{i}4t}, \dots, \mathrm{e}^{\mathrm{i}n^2t})$. We conclude that

$$\|J_nP_nT(t)f - T(t)f\|^2 = \sum_{k=n+1}^{\infty} \mathrm{e}^{-2k^2t}|f_k|^2 \leq \mathrm{e}^{-2(n+1)^2t}\sum_{k=n+1}^{\infty}|f_k|^2 \leq \mathrm{e}^{-2(n+1)^2t}\|f\|^2,$$

[1] Recall that $A^{-1} = -R(0,A)$, the resolvent defined in Lecture 2.

[2] We use the notation $\|g\|_{A^2} := \|g\| + \|A^2f\|$ for the graph norm of A^2.

which shows that for $t > 0$ we get a convergence in operator norm being quicker than any polynomial.

For the other example, however, we observe that

$$\|J_n P_n S(t)f - S(t)f\|^2 = \sum_{k=n+1}^{\infty} |\mathrm{e}^{\mathrm{i}k^2t} f_k|^2 = \sum_{k=n+1}^{\infty} |f_k|^2.$$

Let us introduce again $f = (f_n) \in X$ for $g = (g_n) \in D(B)$ such that $f = Bg$, $g = B^{-1}f$. Then we can repeat the argument:

$$\begin{aligned}\|J_n P_n S(t)g - S(t)g\|^2 &= \sum_{k=n+1}^{\infty} |\mathrm{e}^{\mathrm{i}k^2t} g_k|^2 = \sum_{k=n+1}^{\infty} |\tfrac{1}{k^2} f_k|^2 \\ &\leq \frac{1}{(n+1)^4} \sum_{k=n+1}^{\infty} |f_k|^2 \leq \frac{1}{(n+1)^4}\|f\|^2 = \frac{1}{(n+1)^4}\|Bg\|^2.\end{aligned}$$

This shows that in this case we can only recover the convergence order $p = 2$ for $g \in D(B)$. Thus, we have to be careful even with this simple example.

Proof of Proposition 3.10. In order to simplify matters, we again concentrate first on the calculations in the case $X_n = X$, $J_n = P_n = I$. Let us start by fixing some $t_0 > 0$. Then for all $t \in [0, t_0]$ we obtain that

$$\begin{aligned}&\big(T_n(t) - T(t)\big)A^{-1}f \\ &\quad= \underbrace{T_n(t)(A^{-1} - A_n^{-1})f} + A_n^{-1}(T_n(t) - T(t))f + \underbrace{(A_n^{-1} - A^{-1})T(t)f}.\end{aligned} \tag{3.4}$$

It is clear from the stability assumption that the first and the last term of this sum converge to 0 in the operator norm and at the desired rate, i.e.,

$$\|T_n(t)(A^{-1} - A_n^{-1})f\| \leq M\mathrm{e}^{\omega t_0}\|A^{-1} - A_n^{-1}\| \cdot \|f\|,$$

and

$$\|(A^{-1} - A_n^{-1})T(t)f\| \leq M\mathrm{e}^{\omega t_0}\|A^{-1} - A_n^{-1}\| \cdot \|f\|.$$

Hence, we have to concentrate on the middle term. Instead of this term, we consider first a more symmetric one where the fundamental theorem of calculus comes to help. Indeed, let us first show that

$$A_n^{-1}\big(T(t) - T_n(t)\big)A^{-1}h = \int_0^t T_n(t-s)\left(A^{-1} - A_n^{-1}\right)T(s)h\,\mathrm{d}s \tag{3.5}$$

holds for all $h \in X$ and $t > 0$. To this end, observe that the function

$$[0,t] \ni s \mapsto T_n(t-s)A_n^{-1}T(s)A^{-1}h \in X$$

is differentiable by Theorem 2.31, and its derivative is

$$\begin{aligned}\tfrac{\mathrm{d}}{\mathrm{d}s}\left(T_n(t-s)A_n^{-1}T(s)A^{-1}h\right) &= T_n(t-s)\left(-A_nA_n^{-1}T(s) + A_n^{-1}T(s)A\right)A^{-1}h \\ &= T_n(t-s)\left(A^{-1} - A_n^{-1}\right)T(s)h.\end{aligned}$$

Hence, the fundamental theorem of calculus yields formula (3.5).

Now we can see that for $h \in X$ the inequality

$$\begin{aligned}\|A_n^{-1}(T(t) - T_n(t))A^{-1}h\| &\leq \int_0^t M\mathrm{e}^{\omega(t-s)} \cdot \|A^{-1} - A_n^{-1}\| \cdot \|T(s)h\| \,\mathrm{d}s \\ &\leq t_0 M^2 \mathrm{e}^{\omega t_0} \|A^{-1} - A_n^{-1}\| \cdot \|h\|\end{aligned}$$

holds. Summarising the estimates from above, we conclude that for $g \in D(A^2)$ we can introduce $f = Ag$ and $h = Af$ to obtain that

$$\|T_n(t)g - T(t)g\| \leq \|A^{-1} - A_n^{-1}\| M\mathrm{e}^{\omega t_0}(t_0 M\|A^2 g\| + 2\|Ag\|),$$

which yields the desired estimate. □

3.3 The First Trotter–Kato Theorem

We turn our attention now to the general approximation theorems with the weakest possible assumptions. We start by investigating convergence of generators. As we have seen before, convergence of operators and convergence of the corresponding resolvents are connected.

Lemma 3.13. *Suppose that Assumption 3.2 is satisfied and that A_n, A are closed operators on X_n and X, respectively, such that there is some $\lambda \in \rho(A_n) \cap \rho(A)$ for all $n \in \mathbb{N}$ and there is a constant $M \geq 0$ with the property*

$$\|R(\lambda, A_n)\| \leq M \quad \textit{for all } n \in \mathbb{N}.$$

Then the following assertions are equivalent.

(i) There is a dense subset $Y \subset D(A)$ such that $(\lambda - A)Y$ is dense in X, and for all $f \in Y$ there is a sequence $f_n \in D(A_n)$ satisfying $\|f_n - P_n f\|_{X_n} \to 0$ for which

$$\|A_n f_n - P_n Af\|_{X_n} \to 0 \quad \textit{as } n \to \infty. \tag{3.6}$$

(ii) $\|R(\lambda, A_n)P_n f - P_n R(\lambda, A)f\|_{X_n} \to 0$ as $n \to \infty$ for all $f \in X$.

Proof. (i) ⇒ (ii): It is sufficient to show the convergence for the dense subspace $(\lambda - A)Y$. So let us take $f \in Y$ and $g := (\lambda - A)f$. By assumption, we can choose a sequence $f_n \in D(A_n)$ so that $(f_n - P_n f) \to 0$ and $(A_n f_n - P_n Af) \to 0$, hence

$$g_n := (\lambda - A)g_n$$

satisfies $(g_n - P_n g) \to 0$. Therefore, we obtain

$$\begin{aligned}\|R(\lambda, A_n)P_n g - P_n R(\lambda, A)g\| &\leq \|R(\lambda, A_n)P_n g - R(\lambda, A_n)g_n\| + \|R(\lambda, A_n)g_n - P_n R(\lambda, A)g\| \\ &\leq \|R(\lambda, A_n)\| \cdot \|P_n g - g_n\| + \|f_n - f\| \to 0 \qquad \text{as } n \to \infty.\end{aligned}$$

(ii) ⇒ (i): We set $Y := D(A)$. For given $f \in D(A)$ let $g := (\lambda - A)f$ and $f_n := R(\lambda, A_n)P_n g$. We can see that

$$\begin{aligned} A_n f_n &= A_n R(\lambda, A_n)P_n g = \lambda R(\lambda, A_n)P_n g - P_n g, \\ \text{and} \qquad P_n Af &= P_n AR(\lambda, A)g = \lambda P_n R(\lambda, A)g - P_n g.\end{aligned}$$

Hence, from the assumption it follows

$$(A_n f_n - P_n Af) \to 0.$$

□

Now we can show that, assuming stability and using ideas presented in the previous section, this convergence of the generators is equivalent to the (strong) convergence of the approximating semigroups. Though unnecessary, for the sake of completeness we formulate as appears in textbooks.

Theorem 3.14 (First Trotter–Kato Approximation Theorem)**.** *Suppose that Assumption 3.2 is satisfied and that A_n, A generate strongly continuous semigroups in X_n and X, respectively, and that there are $M \geq 0$, $\omega \in \mathbb{R}$ such that the stability condition* (3.2) *holds. Then the following are equivalent.*

(i) There is a dense subspace $Y \subset D(A)$ such that there is $\lambda > 0$ with $(\lambda - A)Y$ being dense in X. Furthermore, for all $f \in Y$ there is a sequence $f_n \in D(A_n)$ satisfying

$$\|f_n - P_n f\| \to 0 \textit{ and } \quad \|A_n f_n - P_n A f\| \to 0 \quad \textit{as } n \to \infty. \tag{3.7}$$

(ii) $\|R(\lambda, A_n)P_n f - P_n R(\lambda, A)f\| \to 0$ as $n \to \infty$ for all $f \in X$ and some/all $\lambda > \omega$.

(iii) $\|T_n(t)P_n f - P_n T(t)f\| \to 0$ as $n \to \infty$ for all $f \in X$ uniformly for t in compact intervals.

Proof. In view of Lemma 3.13, only the equivalence of (ii) and (iii) has to be shown.

(iii) ⇒ (ii): From the integral representation of the resolvent we see that

$$\|R(\lambda, A_n)P_n f - P_n R(\lambda, A)f\| \leq \int_0^\infty \mathrm{e}^{-\lambda t}\|T(t)f - T_n(t)f\| \,\mathrm{d}t.$$

The desired convergence follows then from the Lebesgue dominated convergence theorem.

(ii) ⇒ (iii): We repeat here the arguments from the proof of Proposition 3.10 in a careful way. Then we can see how we can refine our arguments. From the uniform boundedness principle, Theorem 2.28 and from the stability condition it follows that it suffices to show strong convergence on a dense subset. Fixing $t_0 > 0$ we obtain for all $t \in [0, t_0]$ that

$$\begin{aligned}&(T_n(t) - T(t))R(\lambda, A)f = \\ &\quad = \underbrace{T_n(t)(R(\lambda, A) - R(\lambda, A_n))f} + R(\lambda, A_n)(T_n(t) - T(t))f + \underbrace{(R(\lambda, A_n) - R(\lambda, A))T(t)f}.\end{aligned}$$

It is clear from the stability assumption that the first and the last term of this sum converge to 0, that is,

$$\|T_n(t)(R(\lambda, A) - R(\lambda, A_n))f\| \leq M\mathrm{e}^{\omega t_0}\|R(\lambda, A) - R(\lambda, A_n)f\| \to 0,$$

and

$$\|(R(\lambda, A) - R(\lambda, A_n))T(t)f\| \leq \|(R(\lambda, A) - R(\lambda, A_n))T(t)f\| \to 0$$

as $n \to \infty$. Note that, since the set $\{T(t)f : t \in [0, t_0]\}$ is compact, the second term converges uniformly (i.e., independently of t) to zero, see Proposition 2.30.

Hence, we have to concentrate on the middle term. Here, repeating previous arguments from the proof of Proposition 3.10, we obtain that for all $h \in X$

$$\|R(\lambda, A_n)(T(t) - T_n(t))R(\lambda, A)h\| \leq \int_0^t M\mathrm{e}^{\omega(t-s)} \cdot \|(R(\lambda, A) - R(\lambda, A_n))T(s)h\| \,\mathrm{d}s.$$

Observe again that the set $\{T(t)h : t \in [0, t_0]\}$ is compact and hence the integrand converges uniformly in $s \in [0, t_0]$. Thus, we obtain that the middle term also converges to zero in the desired way. □

3.4 Exercises

1. Prove the exponential estimate, the stability condition, from Remark 3.1.

2. Consider the operators J_n and P_n in Example 3.4 and show that for each $f \in X$, $J_n P_n f \to f$, i.e., for each $f \in \mathrm{C}_{(0)}([0,1])$ we have

$$\sum_{k=0}^{n-1} f(\tfrac{k}{n}) B_{n,k}(x) \to f(x)$$

as $n \to \infty$, uniformly in $x \in [0,1]$.

3. Let $X := L^1(0,1)$, $X_n = \mathbb{C}^n$, and define the operators

$$J_n(y_1, \dots, y_n) := \sum_{k=1}^{n} y_k \cdot \chi_{[(k-1)/n, k/n]},$$

$$(P_n f)_k := n \cdot \int\limits_{\frac{k-1}{n}}^{\frac{k}{n}} f(x) dx,$$

and the norm

$$\|(y_k)\|_n := \frac{1}{n} \sum_{k=1}^{n} |y_k|$$

for $(y_k) \in X_n$. Here χ stands for the characteristic function of a set. Prove that this scheme satisfies the conditions of Assumptions 3.2. Perform analogous calculations to Example 3.7.

4. Finish the proof of Proposition 3.10.

5. Solve the exercises in Appendix A.

Lecture 4

The Lax Equivalence Theorem

We continue here the study of approximation theorems for semigroups by changing the field of space discretisations to time discretisations. Consider the initial value problem (abstract Cauchy problem)

$$\begin{cases} \dot{u}(t) = Au(t), \quad t \geq 0 \\ u(0) = u_0 \in X, \end{cases} \tag{ACP}$$

where we suppose that A generates the strongly continuous semigroup T on the Banach space X. We saw in the previous Lecture 3 how to put spatial discretisations in an abstract setting. Generally, however, it is not usual to approximate the solution of an abstract Cauchy problem by exponential functions, but by some time discretisations, for example by using a finite difference scheme. They are in the focus of our interest this week.

Definition 4.1. Let T be semigroup with generator A, and consider the abstract Cauchy problem (ACP) on X. Consider further a strongly continuous function $F : [0, \infty) \to \mathscr{L}(X)$ with $F(0) = I$.

a) Suppose that there is $D \subseteq D(A)$ a dense subspace in X such that

$$\lim_{h \searrow 0} \frac{F(h)T(t)f - T(t+h)f}{h} = 0$$

holds for all $f \in D$ locally uniformly in t. Then we call F a **consistent finite difference scheme** (or finite difference method). To be more precise, we say that F is consistent with (ACP) on the subspace D.

b) A consistent finite difference scheme is called **stable**, if for all $t_0 > 0$ there is a constant $M \geq 1$ such that

$$\|F(h)^n\| \leq M$$

holds for all $h \geq 0$ and $n \in \mathbb{N}$ with $hn \leq t_0$.

c) A consistent finite difference scheme is called **convergent**, if for all $t > 0$, $h_k \to 0$, $n_k \to \infty$ with $h_k n_k \in [0, t]$ and $h_k n_k \to t$ we have

$$T(t)f = \lim_{k \to \infty} F(h_k)^{n_k} f$$

for all $f \in X$.

We mention the following examples.

Example 4.2. The semigroup T, i.e., the function $F(h) = T(h)$ is the best possible approximation. Certainly, this example has all the properties from the definition above, but is irrelevant from the numerical point of view.

The next example is an extremely important one, motivated by Euler's formula for the exponential function.

Example 4.3 (**Implicit Euler scheme**). By Proposition 2.26.a) we know that there is $\omega \in \mathbb{R}$ such that every $\lambda \geq \omega$ belongs to the resolvent set $\rho(A)$ of A. For $h \in (0, \frac{1}{\omega}]$ we can define

$$F(h) = \tfrac{1}{h} R(\tfrac{1}{h}, A) = (I - hA)^{-1}, \quad \text{and} \quad F(0) = I.$$

(For $h > \frac{1}{\omega}$ we may set $F(h) = F(\frac{1}{\omega})$, but this is not important, because we shall be interested in small h values.) This numerical scheme is called **implicit Euler scheme**. We shall investigate its properties later.

Example 4.4 (**Crank–Nicolson scheme**). We define the **Crank–Nicolson scheme** by

$$F(h) = (I + \tfrac{h}{2}A)(I - \tfrac{h}{2}A)^{-1} \quad \text{for } h \in (0, \tfrac{1}{\omega}] \quad \text{and} \quad F(0) = I.$$

Note that in our terminology the abstract Cauchy problem (ACP), and, in particular, information about the operator A is already incorporated in the finite difference scheme F. Hence, if we speak for example of the implicit Euler scheme, we mean the implicit Euler scheme for that particular problem, and not the implicit Euler scheme in general.

Consistency means in a way that the finite difference scheme is locally (i.e., for h small) a good approximation, in other words that the local error $\|F(h)f - T(h)f\|$ is small. The condition above is in applications hard to verify, since T is a priori unknown. Motivated by the finite dimensional ODE case, we relate the consistency condition to the derivative of F at $t = 0$.

Proposition 4.5. *Let $Y \subseteq D(A)$ be a Banach space that is dense in X and is continuously embedded in the Banach space $D(A)$. Suppose that it is invariant under the semigroup T so that the restriction*[1] *is a strongly continuous semigroup again. Then a finite difference scheme F is consistent with* (ACP) *on Y if and only if*

$$Af = \lim_{h \searrow 0} \frac{F(h)f - f}{h} = F'(0)f \tag{4.1}$$

holds for all $f \in Y$.

Proof. Suppose first that F is consistent. Since $Y \subseteq D(A)$, we have by definition

$$\lim_{h \searrow 0} \frac{T(h)f - f}{h} = Af \quad \text{for } f \in Y.$$

By specialising $t = 0$ in the definition of consistency we obtain

$$0 = \lim_{h \searrow 0} \frac{F(h)f - T(h)f}{h} = \lim_{h \searrow 0} \frac{F(h)f - f + f - T(h)}{h}.$$

This yields the convergence in (4.1).

For the other direction, suppose that (4.1) holds. Then the function $G : [0,1] \to \mathscr{L}(Y, X)$ defined by

$$G(h) := \begin{cases} F'(0) & \text{for } h = 0, \\ \dfrac{F(h) - I}{h} & \text{for } h \in (0,1] \end{cases}$$

[1] We restrict, of course, the semigroup operators: $T|_Y(t) = T(t)|_Y$.

is strongly continuous. First of all note that for $h > 0$ indeed $G(h) \in \mathscr{L}(Y, X)$, since Y is continuously embedded in $D(A)$ hence in X. On the other hand we have $G(0) = A \in \mathscr{L}(Y, X)$, again by the continuous embedding $Y \subseteq D(A)$. The strong continuity on $(0, 1]$ is obvious, whereas at 0 it follows from the assumption (4.1). In particular, we obtain from Proposition 2.2 that

$$\Big\|\frac{F(h) - I}{h}\Big\|_{\mathscr{L}(Y,X)} \le M \quad \text{for all } h \in [0, 1] \text{ and for some } M \ge 0.$$

The same arguments yield

$$\Big\|\frac{T(h) - I}{h}\Big\|_{\mathscr{L}(Y,X)} \le M \quad \text{for all } h \in [0, 1] \text{ and for some } M \ge 0.$$

We now prove consistency on Y. Take $t_0 > 0$. By assumption, T is strongly continuous on the Banach space Y, whence for $f \in Y$ fixed the set

$$C := \big\{T(s)f \,:\, t \in [0, t_0]\big\} \subseteq Y$$

is compact (being the continuous image of a closed interval). By using Proposition 2.30, we conclude

$$\frac{F(h) - I}{h} g \to Ag \quad \text{and} \quad \frac{T(h) - I}{h} g \to Ag$$

uniformly for $g \in C$ as $h \to 0$. This implies the uniform convergence

$$\frac{F(h) - T(h)}{h} T(t)f \to 0$$

for $t \in [0, t_0]$ as $h \to 0$, i.e., consistency. □

4.1 The Lax Equivalence Theorem

It turned out quite early that for partial differential equations finite difference schemes do not always converge. Famous examples are due to Richardson[2] or Courant, Friedrichs and Lewy[3].

Using the notation and the notions above, we can formulate the following fundamental result.

Theorem 4.6 (Lax Equivalence Theorem[4]). *For a consistent finite difference scheme, stability is equivalent to convergence.*

Proof. Suppose first that the consistent finite different scheme F is convergent but not stable. Fix $t_0 > 0$ such that

$$\sup\big\{\|F(h)^n\| : h \ge 0,\ n \in \mathbb{N},\ nh \in [0, t_0]\big\} = \infty,$$

and take a sequence $h_k \to 0$ with $n_k h_k \in [0, t_0]$ and $\|F(h_k)^{n_k}\| \to \infty$. By passing to the subsequence we may suppose that $n_k h_k \to t$ for some $t \in [0, t_0]$. Convergence and the uniform boundedness principle, Theorem 2.28, now imply boundedness of $\|F(h_k)^{n_k}\|$, hence a contradiction. So F is stable.

[2] L. F. Richardson, Weather Prediction by Numerical Process. Cambridge University Press, London 1922.

[3] R. Courant, K. Friedrichs, H. Lewy, "Über die partiellen Differenzengleichungen der mathematischen Physik," Math. Annalen **100** (1928), 32–74.

[4] P. D. Lax and R. D. Richtmyer, "Survey of the stability of linear finite difference equations", Comm. Pure Appl. Math. **9** (1956), 267–293.

Fix $t > 0$ and take sequences $h_k \to 0$, $n_k \to \infty$ with $h_k n_k \in [0, t]$ and $h_k n_k \to t$. Notice first of all that, by the strong continuity of T, it suffices to prove

$$F(h_k)^{n_k} f - T(n_k h_k) f \to 0.$$

Now, one can use the well-known algebraic identity on the difference of two n^{th} powers to obtain the "telescopic sum"

$$F(h_k)^{n_k} f - T(n_k h_k) f = F(h_k)^{n_k} f - T(h_k)^{n_k} f = \sum_{j=0}^{n_k - 1} F(h_k)^{n_k - 1 - j} \big(F(h_k) - T(h_k) \big) T(h_k)^j f. \quad (4.2)$$

From this point on, the proof is a standard epsilon-argument. Taking $f \in D$ and fixing $\varepsilon > 0$, it follows by the consistency assumption in Definition 4.1 a) that there is $N \in \mathbb{N}$ so that

$$\| F(h_k) T(s) f - T(h_k) T(s) f \| \le \varepsilon h_k$$

holds $s \in [0, t]$ for all $k \ge N$ and $t \in [0, t_0]$. For $k \ge N$ we can now estimate (4.2) as follows

$$\| F(h_k)^{n_k} f - T(n_k h_k) f \| \le \sum_{j=0}^{n_k - 1} \| F(h_k)^{n_k - 1 - j} \| \cdot \varepsilon h_k \le \sum_{j=0}^{n_k - 1} M \cdot \varepsilon h_k \ \le M t \varepsilon.$$

This proves the convergence on the space D. The claim follows then from the stability condition and the denseness of the set D in X. □

Applications of the Lax equivalence theorem are numerous. We list here some of them.

Corollary 4.7 (Implicit Euler Scheme)**.** *Assume that the operator A generates the strongly continuous semigroup T of type $(M, 0)$ where $M \ge 1$. Then for all $f \in X$ we have*

$$T(t) f = \lim_{n \to \infty} \left(\tfrac{n}{t} R(\tfrac{n}{t}, A) \right)^n f = \lim_{n \to \infty} \left(I - \tfrac{t}{n} A \right)^{-n} f.$$

Proof. Stability follows from the properties of the generator discussed in Lecture 2, especially from equation (2.2). Let us introduce the function

$$F(t) := \begin{cases} I & \text{for } t = 0, \\ \tfrac{1}{t} R(\tfrac{1}{t}, A) & \text{for } t > 0. \end{cases}$$

Then, from the identity $\lambda R(\lambda, A) - I = A R(\lambda, A)$ we see that for $f \in D(A)$

$$\frac{F(h) f - f}{h} = \tfrac{1}{h} R(\tfrac{1}{h}, A) A f.$$

In order to apply the Lax equivalence theorem 4.6, we need to check consistency, i.e., the convergence of this expression to Af as $h \to 0$ (use Proposition 4.5). The proof can be finished by applying the following result, which we state separately because of its importance. □

Proposition 4.8. *Let A be a closed, densely defined operator. Assume that there are $M \ge 1$ and $\omega \in \mathbb{R}$ such that for all $\lambda > \omega$, we have $\lambda \in \rho(A)$ and $\| \lambda R(\lambda, A) \| \le M$. Then*

a) $\lambda R(\lambda, A) f \to f$ for all $f \in X$ as $\lambda \to \infty$, and

b) $\lambda R(\lambda, A) A f \to A f$ for all $f \in D(A)$ as $\lambda \to \infty$.

Proof. Taking $g \in D(A)$, we see that $\lambda R(\lambda, A)g = R(\lambda, A)Ag + g$. By assumption,

$$\|R(\lambda, A)Ag\| \leq \frac{M}{\lambda}\|Ag\|,$$

and hence $\lambda R(\lambda, A)g \to g$ as $\lambda \to \infty$. By the denseness of $D(A)$ and the boundedness, the convergence follows for all $g \in X$. The second statement is an immediate consequence of the first one. □

The operators $\lambda R(\lambda, A)$, $\lambda > 0$, are called **Yosida approximants**.

Remark 4.9. It can be proved that although the Crank–Nicolson scheme is consistent it is not stable for every generator. The left shift semigroup on $\mathrm{C}_0(\mathbb{R})$ or on $\mathrm{L}^1(\mathbb{R})$ provides a notable counterexample. We will come back to this problem in later lectures.

The following approximation formula is usually called **Lie-Trotter product formula**[5] in functional analysis and has deep applications. We will come back to it in later lectures in more detail.

Corollary 4.10. *Suppose that the operators A, B, and C are generators of strongly continuous semigroups T, S, and U, respectively. Suppose further that*

$$D(A) \cap D(B) = D(C) \text{ and for } f \in D(A) \cap D(B) \text{ we have } Cf = Af + Bf,$$

and that there is $M \geq 1$, $\omega \in \mathbb{R}$ such that

$$\left\|\left(S(\tfrac{t}{n})T(\tfrac{t}{n})\right)^n\right\| \leq M\mathrm{e}^{\omega t}.$$

Then

$$U(t)f = \lim_{n\to\infty} \left(S(\tfrac{t}{n})T(\tfrac{t}{n})\right)^n f$$

for all $f \in X$, locally uniformly in $t \geq 0$.

Proof. We introduce $F(t) = S(t)T(t)$ and check the consistency. For $f \in D(A) \cap D(B)$ we conclude that

$$\frac{F(t)f - f}{t} = S(t)\frac{T(t)f - f}{t} + \frac{S(t)f - f}{t}.$$

Clearly, we have for $f \in D(A) \cap D(B)$ by definition that

$$\begin{aligned} \frac{S(t)f - f}{t} &\to Bf, \\ \frac{T(t)f - f}{t} &\to Af, \\ S(t)g &\to g \quad \text{for all } g \in X \end{aligned}$$

as $t \to 0$. Since the set $\left\{\frac{T(t)f-f}{t} : t \in (0, 1]\right\} \cup \{0\}$ is compact, we can apply Proposition 2.30 to infer that

$$\lim_{t\searrow 0} \frac{F(t)f - f}{t} = Bf + Af = Cf,$$

which proves the assertion. □

[5] H. F. Trotter, "On the product of semi-groups of operators," Proc. Amer. Math. Soc. **10** (1959), 545–551.

4.2 Order of convergence

The quantitative version of the Lax equivalence theorem is quite immediate. Before stating it, we need the following definitions.

Definition 4.11. Let A generate the semigroup T on X and let F be a finite difference scheme. Suppose that there is a dense subspace $Y \subset X$ invariant under the semigroup which is a Banach space and let $p > 0$.

a) The finite difference scheme F is called **consistent of order** $p > 0$ **on** Y, if there is a subspace $Y \subset D(A)$ dense in X and invariant under the semigroup operators T, so that there is a $C > 0$ such that for all $f \in Y$ we have

$$\|F(h)f - T(h)f\| \leq Ch^{p+1}\|f\|_Y. \tag{4.3}$$

b) The finite difference scheme F is called **convergent of order** p **on** Y, if for all $t_0 > 0$ there is $K > 0$ such that for all $g \in Y$ we have

$$\|F(h)^n g - T(nh)g\| \leq Kt_0h^p\|g\|_Y$$

for all $n \in \mathbb{N}$, $h \geq 0$ with $nh \in [0, t_0]$.

We will occasionally say that the finite difference scheme has consistency/convergence order $p > 0$ on the subspace Y. Let us stress that the order p may depend on the subspace Y. This will be extremely important in applications later on.

Proposition 4.12. *Suppose that there is a subspace $Y \subset D(A)$ dense and invariant under the semigroup operators $T(t)$, which is a Banach space with its norm, satisfying*

$$\|T(t)\|_Y \leq M\mathrm{e}^{\omega t}.$$

Let F be a stable finite difference scheme which is consistent of order $p > 0$ on Y. Then the finite difference scheme is convergent of order p on Y.

Proof. The proof goes along the same lines as the one for the Lax equivalence theorem 4.6, the only difference is that we have some bound for the local error $\|F(h) - T(h)\|$. For simplicity, we may take $\omega \geq 0$. Let $t_0 > 0$ be fixed. For $g \in Y$, $n \in \mathbb{N}$ and $h \geq 0$ with $nh \in [0, t_0]$ we can write

$$\begin{aligned}
\|F(h)^n g - T(nh)g\| &\leq \sum_{j=0}^{n-1} \left\|F(h)^{n-1-j}\right\| \left\|(F(h) - T(h))\, T\big(jh\big)g\right\| \\
&\leq \sum_{j=0}^{n-1} MCh^{p+1} \left\|T\big(jh\big)g\right\|_Y \leq \sum_{j=0}^{n-1} MCh^{p+1}M\mathrm{e}^{\omega jh}\|g\|_Y \\
&\leq M^2C\mathrm{e}^{\omega t_0}t_0h^p\|g\|_Y = Kt_0h^p\|g\|_Y.
\end{aligned}$$

□

Remark 4.13. We will elaborate later on the choice of the subspace Y. For example, if $X = \mathrm{L}^2(0, \pi)$, you may think of some Sobolev space $Y \subset X$. It is a dense subset, but a Hilbert space with its own norm. Actually, we will spend quite a lot of time later on with the following two topics: Having a large scale of possible invariant subspaces Y at hand, and developing technical tools to verify the consistency estimate (4.3) for various methods.

It is possible to make the convergence of the implicit Euler scheme quantitative along these lines. Going in this direction, we note first that domains of powers of the generator are good candidates to be such subspaces Y.

Proposition 4.14. *Let A generate the semigroup T of type (M,ω), where $M \geq 1$ and $\omega \in \mathbb{R}$. For $n \in \mathbb{N}$, define*

$$X_n := D(A^n), \quad \|f\|_n := \|f\| + \|A^n f\|,$$

the domain of A^n with its graph norm. Then X_n is a Banach space, the restriction $T_n(t) := T(t)|_{X_n}$ defines a strongly continuous semigroup T_n of the same type (M,ω) in X_n.

Proof. We leave the proof of the fact that this space is a Banach space as Exercise 1. By Proposition 2.18, X_n is invariant and dense in the space X. Notice that for $f \in X_n$,

$$\|T(h)f - f\|_n = \|T(h)f - f\| + \|A^n(T(h)f - f)\| = \|T(h)f - f\| + \|T(h)A^n f - A^n f\| \to 0$$

as $h \to 0$ by the strong continuity of T in X. Finally,

$$\|T(t)f\|_n = \|T(t)f\| + \|A^n T(t)f\| \leq M\mathrm{e}^{\omega t}\|f\| + \|T(t)A^n f\| \leq M\mathrm{e}^{\omega t}\left(\|f\| + \|A^n f\|\right) = M\mathrm{e}^{\omega t}\|f\|_n$$

shows that T_n is of type (M,ω). □

The following can be considered as a very simple special case of a celebrated result by Brenner and Thomée[6] on the convergence of rational approximation schemes.

Corollary 4.15. *Let A generate the semigroup T of type $(M,0)$ where $M \geq 1$. Consider the implicit Euler scheme of Corollary 4.7. Then there is $C > 0$ such that for all $f \in D(A^2)$*

$$\left\|(I - hA)^{-n}f - T(nh)f\right\| \leq Kt_0 h\|f\|_2,$$

holds for all $n \in \mathbb{N}$, $h \geq 0$ such that $nh \in [0,t_0]$.

Proof. Stability follows from (2.2), because A is a generator. We have to deal with consistency. We prove consistency on $D(A^2)$. So take $f \in D(A^2)$, and note that

$$F(h) = (I - hA)^{-1} = \tfrac{1}{h}R(\tfrac{1}{h},A) = AR(\tfrac{1}{h},A) + I \tag{4.4}$$

holds. Hence, by Proposition 2.9, and since $f \in D(A)$ we have

$$\begin{aligned} F(h)f - T(h)f &= AR(\tfrac{1}{h},A)f - A\int_0^h T(s)f\,\mathrm{d}s. \\ &= R(\tfrac{1}{h},A)Af - \int_0^h T(s)Af\,\mathrm{d}s = \int_0^h \left(\tfrac{1}{h}R(\tfrac{1}{h},A) - T(s)\right)Af\,\mathrm{d}s. \end{aligned}$$

In a similar manner as before, we analyse the integrand. Since $g = Af \in D(A)$ we obtain

$$\left(\tfrac{1}{h}R(\tfrac{1}{h},A) - T(s)\right)g = AR(\tfrac{1}{h},A)g - A\int_0^h T(s)g = R(\tfrac{1}{h},A)Ag - \int_0^h T(s)Ag\,\mathrm{d}s.$$

[6] P. Brenner and V. Thomée, “On rational approximations of semigroups,” SIAM J. Numer. Anal. **16** (1979), 683-694.

By (2.2), the inequality

$$\left\|R(\tfrac{1}{h},A)Ag\right\| \le hM\|Ag\|$$

follows. Since we are integrating a bounded function, we have

$$\left\|\int\limits_0^h T(s)Ag\,\mathrm{d}s\right\| \le sM\|Ag\| \le hM\|Ag\|.$$

Summarizing, for $f \in D(A^2)$, the estimate

$$\|F(h)f - T(h)f\| \le h^2(2M\|A^2 f\|)$$

holds. From Proposition 4.12 the assertion follows. □

4.3 Exercises

1. Let A be the generator of a semigroup, and consider the space $X_n = D(A^n)$ with the graph norm.

a) For $n \in \mathbb{N}$ and $x \in D(A^n)$ define $|||x||| := \|x\| + \|Ax\| + \cdots + \|A^n x\|$. Prove that $|||\cdot|||$ and $\|\cdot\|_n$ are equivalent norms.

b) Prove that X_n is a Banach space.

2. Let $X = \ell^2$ and $m = (m_n)$ be a sequence with $\operatorname{Re} m_n \le 0$. Consider the semigroup T generated by the multiplication operator $A = M_m$ and define the Crank–Nicolson method as

$$F(h) = (I + \tfrac{h}{2}A)(I - \tfrac{h}{2}A)^{-1}.$$

a) Show that it is stable.

b) Show that it is consistent.

3. Consider the heat equation of Section 1.1 and show that the implicit Euler scheme converges in the operator norm, and has first order convergence.

4. Solve the exercises in the appendix.

Lecture C

Exercises

1. Let $F : \mathbb{R} \to \mathbb{R}$ be continuously differentiable with $\sup_{x\in\mathbb{R}} |F'(x)| < \infty$. Define the flow $\Phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ as the solution of the nonlinear ODE

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} y(t) = F(y(t)) \\ \quad y(0) = s, \end{cases}$$

i.e., $\Phi(t,s) := y(t)$. Take $X := C_0(\mathbb{R})$ and define

$$\big(T(t)f\big)(s) := f\big(\Phi(t,s)\big)$$

for $t \geq 0$, $s \in \mathbb{R}$.

a) Show that T is a contraction semigroup (i.e., of type $(1,0)$) and identify its generator.

b) What is the corresponding abstract Cauchy problem? Which partial differential equation can we associate with it? Relate the semigroup T to the method of characteristics.

2. Let T be a semigroup on the Banach space X with generator A. Prove that for all $f \in D(A^2)$ we have the Taylor formula

$$T(t)f = f + tAf + \int\limits_0^t (t-s)T(s)A^2 f \mathrm{d}s.$$

Find a general Taylor formula for $f \in D(A^n)$.

3. Let T be a contraction semigroup on the Banach space X with generator A. Prove that

$$\|Af\|^2 \leq 4\|A^2 f\| \cdot \|f\|$$

holds for all $f \in D(A^2)$.

4. Let T be a semigroup of type $(M,0)$ on a Banach space X. For $f \in X$ define

$$|||f||| := \sup\big\{\|T(t)f\| : t \geq 0\big\}.$$

Prove that the norms $\|\cdot\|$ and $|||\cdot|||$ are equivalent, and that T is contraction semigroup for the new norm.

5. Let T be a semigroup on the Banach space X and let $B \in \mathscr{L}(X)$. Define

$$S(t) := \mathrm{e}^{tB}.$$

Prove that the stability condition in the Lie–Trotter product formula, in Corollary 4.10, holds, i.e.

$$\Big\|\big(T(\tfrac{t}{n})S(\tfrac{t}{n})\big)^n\Big\| \leq M\mathrm{e}^{\omega t} \quad \text{for all } t \geq 0$$

with appropriate constants M and ω.

6. Let T be a semigroup with generator A. Prove that the Crank–Nicolson scheme is consistent with the corresponding Cauchy problem on $D(A)$.

7. Prove that the stability of a general finite difference scheme F (from Definition 4.1) is equivalent to each of the following conditions:

(i) There is $t_0 > 0$ and $M \geq 0$ such that

$$\|F(h)^n\| \leq M \quad \text{for all } n \in \mathbb{N},\ h \geq 0 \text{ with } nh \in [0, t_0].$$

(ii) For all $t_0 \geq 0$ there is $M \geq 0$ such that

$$\|F(\tfrac{t}{n})^k\| \leq M \quad \text{for all } t \in [0, t_0],\ n \in \mathbb{N} \text{ and } k = 1, \ldots, n.$$

8. Consider the Runge–Kutta methods based on

a) the Gaussian quadrature with one node

$$\begin{array}{c|c} 1/2 & 1/2 \\ \hline & 1 \end{array}$$

b) the Gaussian quadrature with two nodes

$$\begin{array}{c|cc} 1/2 - \sqrt{3}/6 & 1/4 & 1/4 - \sqrt{3}/6 \\ 1/2 + \sqrt{3}/6 & 1/4 + \sqrt{3}/6 & 1/4 \\ \hline & 1/2 & 1/2 \end{array}$$

c) the Radau IIA quadrature with two nodes

$$\begin{array}{c|cc} 1/3 & 5/12 & -1/12 \\ 1 & 3/4 & 1/4 \\ \hline & 3/4 & 1/4 \end{array}$$

Show the stablity of these methods when applied to

a) the heat equation presented in Section 1.1,

b) any abstract Cauchy problem

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} u(t) = Au(t) \\ u(0) = u_0 \end{cases}$$

with diagonalisable operator A which has a discrete spectrum $\{\lambda_i : i \in \mathbb{N}\}$ with $\lambda_i \leq l \in \mathbb{R}$. Are there any restrictions on the time step?

Lecture 5

Approximation of semigroups – Part 2

In Lectures 3 and 4 we learned about methods for approximating the solutions to the Cauchy problem

$$\begin{cases} \dot{u}(t) = Au(t), \quad t \geq 0 \\ u(0) = u_0 \in D(A), \end{cases} \tag{ACP}$$

provided that A is the *generator* of a semigroup T. In this lecture we try to see the issue from a different perspective: We do not suppose in advance that A is a generator, but approximate it via operators A_n which generate semigroups T_n on X. Then we hope that T_n will converge to some object, and it happens to be a semigroup, whose generator is A or at least coincides with A on a large subspace.

In other words, our aim is to use approximation theorems to prove that a certain operator is the generator of a strongly continuous semigroup of linear operators. This approximation idea for showing well-posedness of (ACP) (i.e., existence of a semigroup generated by A) was already used by Courant, Friedrichs and Lewy[1] (1928), and in innumerous publications ever since.

Recall that in the study of approximation methods for semigroups the convergence of resolvent operators played an essential role. In Lecture 3 such issues were easily handled, again by the fact that A was assumed to be a generator, in particular it had a resolvent. Now, this matter will more complex and somewhat harder, so we begin with the investigation of convergence of resolvents, more precisely, with the connection between the convergence of operators and convergence of resolvents. Since the results are rather technical in nature, we suggest that on the first reading you should skip the proofs and jump to Section 5.2.

5.1 Resolvent convergence

The next example shows that the limit of a convergent sequence of resolvents need not be the resolvent of any operator.

Example 5.1. For a given Banach space X, consider the operators $A_n := -nI$. Then for $\lambda > 0$ we have that $R(\lambda, A_n) = \frac{1}{\lambda+n}I$ converges to 0 in the operator norm as $n \to \infty$. The limit is certainly not a resolvent of any operator (if the Banach space is at least one-dimensional).

Under additional assumptions one can nevertheless ensure that the limit of resolvent operators is again a resolvent of some operator. Before turning to such results, we need some further preparation.

Recall from Lecture 2, particularly from Proposition 2.10, that an operator A is closed if for each sequence $f_n \in D(A)$ such that $f_n \to f$ and $Af_n \to g$, we have $g \in D(A)$ and $Af = g$. Now, we call

[1] R. Courant, K. Friedrichs, H. Lewy, "Über die partiellen Differenzengleichungen der mathematischen Physik," Math. Annalen **100** (1928), 32–74.

an operator B **closable** if it has an extension[2] which is a closed operator. The next proposition shows if B is closable, then it has a smallest closed extension, called the **closure** of B and denoted by$\overline{B}$.

Proposition 5.2. *For an operator B with domain $D(B)$ the following statements hold.*

a) The assertions below are equivalent:

(i) Operator B is closable.

(ii) The closure of the graph of B

$$\overline{\operatorname{graph} B} := \overline{\{(f, Bf) : f \in D(B)\}} \subseteq X \times X$$

(which is a closed subspace of $X \times X$) is the graph of an operator A, i.e., (f, g), $(f, h) \in \overline{\operatorname{graph} B}$ implies $g = h$.

(iii) If $f_n \in D(B)$ with $f_n \to 0$ and $Bf_n \to g$, then $g = 0$.

b) If B is closable, let A be the operator from a). Then A is the smallest closed extension of B.

c) Operator B is closable if and only if $\lambda - B$ is closable for $\lambda \in \mathbb{R}$. We have $\overline{\lambda - B} = \lambda - \overline{B}$.

d) If B has a continuous and injective left inverse C, then B is closable. Moreover, if B has dense range, then $C = \overline{B}^{-1}$.

Proof. We prove d) only, the rest of the assertions is left to the reader as Exercise 1. Suppose $f_n \in D(B)$, $f_n \to 0$ and $Bf_n \to g$. Then $f_n = CBf_n$, so we obtain by continuity that $Cg = 0$. Since C is injective, $g = 0$, and by part a) B is closable.
Assume now that the range of B is dense and let us prove $C\overline{B}f = f$ for all $f \in D(\overline{B})$. Taking $f \in D(\overline{B})$, there is a sequence $f_n \in D(B)$ with $f_n \to f$ and $Bf_n \to \overline{B}f$. From this we conclude

$$f_n = CBf_n \to C\overline{B}f,$$

i.e., $f = C\overline{B}f$. This implies in particular that $\overline{B}$ is injective.
Next we show that $\operatorname{ran}(\overline{B})$ is closed. Suppose $f_n \in D(\overline{B})$ with $\overline{B}f_n \to g$. Then $f_n = C\overline{B}f_n \to Cg$, and we obtain $Cg \in D(\overline{B})$ and $\overline{B}Cg = g$ since $\overline{B}$ is closed. It follows that $g \in \operatorname{ran} \overline{B}$. To conclude the proof we collect the properties of $\overline{B}$: It has dense range by assumption, but as we have proved its range is closed, so $\operatorname{ran} \overline{B} = X$. But $\overline{B}$ is also injective, hence $\overline{B} : D(\overline{B}) \to X$ is bijective, so by Proposition 2.10, the operator $\overline{B}$ is continuously invertible, and of course we have $C = \overline{B}^{-1}$. □

The next proposition connects the convergence of operators to the convergence of their resolvents, i.e., it is in the spirit of Lemma 3.13 from Lecture 3. Note, however, that in contrast to that lemma, generally no equivalence between the two properties can be stated here.

Proposition 5.3. *For every $n \in \mathbb{N}$ let A_n be a densely defined closed operator. Suppose that there is $\lambda \in \bigcap_{n\in\mathbb{N}} \rho(A_n)$ such that*

$$\|R(\lambda, A_n)\| \leq M \quad \textit{for all } n \in \mathbb{N} \textit{ and for some } M \geq 0.$$

Consider the following assertions:

[2] The operator A is an extension of B if $D(B) \subseteq D(A)$ and $A|_{D(B)} = B$.

(i) There is a dense subspace $D \subset X$ and a linear operator $A : D \to X$ such that $(\lambda - A)D$ is dense in X. Furthermore, for all $f \in D$ there are $f_n \in D(A_n)$ with

$$f_n \to f \quad \text{and} \quad A_n f_n \to Af \quad \text{for } n \to \infty.$$

(ii) The limit

$$R(\lambda)f := \lim_{n\to\infty} R(\lambda, A_n)f$$

exists for all $f \in X$ and defines a bounded linear operator with dense range.

Then (i) implies (ii). Under the additional assumption that $R(\lambda)$ has a trivial kernel (ii) implies (i). If (i) holds and $\ker R(\lambda) = \{0\}$, then A is closable, and $R(\lambda, \overline{A}) = R(\lambda)$.

Proof. Take $f \in D$ and set $g = (\lambda - A)f$. Then, by the assumptions, we find $f_n \in D(A_n)$ with $f_n \to f$ and $A_n f_n \to Af$ for $n \to \infty$. Let

$$g_n := (\lambda - A_n)f_n,$$

then $g_n \to g = (\lambda - A)f$ as $n \to \infty$. We now show that the elements $R(\lambda, A_n)g$ form a Cauchy sequence in X. For $n, m \in \mathbb{N}$ we have

$$R(\lambda, A_n)g - R(\lambda, A_m)g = R(\lambda, A_n)(g - g_n) + \Big(R(\lambda, A_n)g_n - R(\lambda, A_m)g_m\Big) + R(\lambda, A_m)(g_m - g).$$

Since $\|R(\lambda, A_n)\| \le M$ for all $n \in \mathbb{N}$, the first and the last term converge to zero as $n, m \to \infty$. For the middle term we have

$$R(\lambda, A_n)g_n - R(\lambda, A_m)g_m = f_n - f_m \to 0$$

as $n, m \to \infty$. Therefore, the limit

$$R(\lambda)g := \lim_{n\to\infty} R(\lambda, A_n)g$$

exists for all $g \in (\lambda - A)D$. By Theorem 2.30 the limit exists for all $g \in \overline{(\lambda - A)D} = X$, and we have $R(\lambda) \in \mathscr{L}(X)$, and (ii) is proved.

To prove (ii) ⇒ (i) suppose that $R(\lambda)$ is injective and let $D := \operatorname{ran} R(\lambda)$, which is dense by assumption. For $f \in D$ set $g = R(\lambda)^{-1}f$, and define $f_n = R(\lambda, A_n)g$. Then we can write

$$\begin{aligned}
A_n f_n - A_m f_m &= A_n R(\lambda, A_n)g - A_m R(\lambda, A_m)g \\
&= (A_n - \lambda)R(\lambda, A_n)g + \lambda R(\lambda, A_n)g - \lambda R(\lambda, A_m)g - (A_m - \lambda)R(\lambda, A_m)g \\
&= \lambda\Big(R(\lambda, A_n)g - R(\lambda, A_m)g\Big),
\end{aligned}$$

which shows that $(A_n f_n)$ is a Cauchy in X. Therefore we can define the linear operator

$$Af := \lim_{n\to\infty} A_n f_n = \lim_{n\to\infty} A_n R(\lambda, A_n)R(\lambda)^{-1}f.$$

For $f \in D$ we have $f_n \to f$ with f_n defined above, so (i) is proved.

Suppose (i) is true, and that $\ker R(\lambda) = \{0\}$. For $f \in D$ let $f_n \to f$ with $f_n \in D(A_n)$ and $A_n f_n \to Af$. Then clearly we have

$$R(\lambda)(\lambda - A)f = \lim_{n\to\infty} R(\lambda, A_n)(\lambda f_n - A_n f_n) = f.$$

Since by assumption $R(\lambda)$ is injective, Proposition 5.2.c) implies that A is closable and $R(\lambda, \overline{A}) = R(\lambda)$. □

We need to find extra conditions to be able to conclude that the operator obtained in (ii) of the previous proposition is injective. The next statement is a first step in this direction.

Proposition 5.4. *For each $n \in \mathbb{N}$ let A_n generate a semigroup of the same type (M, ω) with $\omega \geq 0$. Then the set*

$$\Lambda := \Big\{\mu : \mu > \omega,\ \lim_{n\to\infty} R(\mu, A_n) \text{ exists}\Big\}$$

is either empty or $\Lambda = (\omega, \infty)$.

Proof. We prove that set Λ is both open and relatively closed in (ω, ∞). Using that it is non-empty, by connectedness of (ω, ∞) we obtain the assertion.

First of all recall from Proposition 2.26 that

$$\|R(\mu, A_n)^k\| \leq \frac{M}{(\mu-\omega)^k} \quad \text{holds for all } k \in \mathbb{N} \text{ and } \mu > \omega.$$

If $\mu > \omega$, then we have

$$R(\mu', A_n) = \sum_{k=0}^{\infty} (\mu - \mu')^k R(\mu, A_n)^{k+1}$$

with uniform and absolute convergence in operator norm for all $\mu' > 0$ with $|\mu' - \mu| \leq \delta(\mu - \omega)$. The convergence of this series is even uniform in $n \in \mathbb{N}$. For $\mu \in \Lambda$ and μ' as above, we prove that $\mu' \in \Lambda$. Let $\varepsilon > 0$ and $f \in X$, then there is $N \in \mathbb{N}$ such that

$$\Big\| \sum_{k=N+1}^{\infty} (\mu - \mu')^k R(\mu, A_n)^{k+1} \Big\| \leq \varepsilon \quad \text{for all } n \in \mathbb{N}.$$

Since by assumption $R(\mu, A_n)f$ converges, so does $R(\mu, A_n)^k f$. Hence there is $n_0 \in \mathbb{N}$ such that

$$\Big\| \sum_{k=0}^{N} (\mu - \mu')^k R(\mu, A_n)^{k+1} - \sum_{k=0}^{N} (\mu - \mu')^k R(\mu, A_m)^{k+1} \Big\| \leq \varepsilon$$

whenever $n, m \geq n_0$. Altogether we obtain

$$\|R(\mu', A_n)f - R(\mu', A_m)f\| \leq 3\varepsilon \quad \text{for all } n, m \geq n_0.$$

This proves that $R(\mu', A_n)f$ converges, showing that Λ is open. Let μ be an accumulation point of Λ in (ω, ∞). Then there is $\mu' \in \Lambda$ with $|\mu - \mu'| \leq \frac{\mu' - \omega}{2}$. Thus, by the arguments above, μ belongs to Λ, showing that Λ is closed. □

Now we can state the next fundamental result on convergence of resolvents of generators.

Proposition 5.5. *For $n \in \mathbb{N}$ let A_n generate semigroups of type (M, ω) with $\omega \geq 0$, such that for some $\lambda > \omega$ the limit*

$$R(\lambda)f := \lim_{n\to\infty} R(\lambda, A_n)f$$

exists[3] *for all $f \in X$. If $R(\lambda)$ has dense range, then it is injective and equals the resolvent $R(\lambda, B)$ of a densely defined operator B.*

[3] In other words, $R(\lambda, A) = s - \lim R(\lambda, A_n)$, meaning that the strong limit exists.

Proof. By Proposition 5.4 we can define

$$R(\mu)f := \lim_{n\to\infty} R(\mu, A_n)f$$

for all $\mu > \omega$ and $f \in X$. Clearly $R(\mu)$ is then a bounded linear operator. Since for $\mu, \mu' > \omega$ the resolvent identity

$$R(\mu', A_n) - R(\mu, A_n) = (\mu - \mu')R(\mu, A_n)R(\mu', A_n)$$

holds, by passing to the limit we obtain the equality

$$R(\mu') - R(\mu) = (\mu - \mu')R(\mu)R(\mu') = (\mu - \mu')R(\mu')R(\mu), \tag{5.1}$$

or after rewriting

$$R(\mu') = R(\mu)\Big((\mu - \mu')R(\mu') + I\Big) = \Big((\mu - \mu')R(\mu') + I\Big)R(\mu).$$

From this we can conclude the equalities $\ker R(\mu) = \ker R(\mu')$ and $\operatorname{ran} R(\mu) = \operatorname{ran} R(\mu')$ for all $\mu, \mu' > \omega$.

By the definition of $R(\mu)$ and since $\|(\mu - \omega)R(\mu, A_n)\| \le M$ we have

$$\|R(\mu)\| \le \frac{M}{\mu - \omega} \quad \text{for all } \mu > \omega.$$

In particular $R(\mu) \to 0$ in operator norm for $\mu \to \infty$ and $\mu R(\mu)$ is uniformly bounded for $\mu > \omega$. From (5.1) it follows

$$\mu R(\mu)R(\lambda)f = R(\lambda)f - R(\mu)f + \lambda R(\lambda)R(\mu)f.$$

Hence

$$\lim_{\mu\to\infty} \mu R(\mu)R(\lambda)f = R(\lambda)f.$$

By assumption $\operatorname{ran} R(\lambda)$ is dense, so we conclude by Theorem 2.30 the convergence

$$\lim_{\mu\to\infty} \mu R(\mu)g = g,$$

for all $g \in X$. This also yields that $R(\mu)$ is injective for all $\mu > \omega$.

To conclude the proof, we define $B := \lambda - R(\lambda)^{-1}$ with $D(B) = \operatorname{ran} R(\lambda)$. Then B is a closed and densely defined operator, with $R(\lambda, B) = R(\lambda)$. □

We are now prepared for general approximation theorems, and begin with the *commutative case.*

5.2 Commuting approximations: Generation theorems

In many applications one encounters approximating operators that are bounded and commute. The first result is a simple but rather important special case of a general approximation theorem, the second Trotter–Kato theorem below.

Recall from Exercise 1 in Lecture 2 that the exponential function of a bounded linear operator $B \in \mathscr{L}(X)$ defines a semigroup of type $(1, \|B\|)$ via

$$S(t) = \mathrm{e}^{tB} = \sum_{n=0}^{\infty} \frac{t^n B^n}{n!}.$$

Proposition 5.6. *For $n \in \mathbb{N}$ let $A_n \in \mathscr{L}(X)$ be bounded operators commuting with each other. Suppose the following:*

(i) There exist $M \geq 1$ and $\omega \in \mathbb{R}$ such that

$$\left\|\mathrm{e}^{tA_n}\right\| \leq M\mathrm{e}^{\omega t} \quad \textit{for all } t \geq 0,\ n \in \mathbb{N}.$$

(ii) There is a dense subset $D \subset X$ such that

$$\lim_{n\to\infty} A_n f =: Af \quad \textit{exists for all } f \in D.$$

(iii) The set $(\lambda - A)D$ is dense for some $\lambda > \omega$.

Then operator A is closable and $\overline{A}$ generates a strongly continuous semigroup T given by

$$T(t)f := \lim_{n\to\infty} \mathrm{e}^{tA_n} f$$

for all $f \in X$.

Proof. We first prove that the sequence $(\mathrm{e}^{tA_n} f)$ is convergent for all $f \in X$. To this end, note that for $n, m \in \mathbb{N}$ we have $A_m = A_n + (A_m - A_n)$. Using ideas already presented before, note that the function

$$[0,t] \ni s \mapsto \mathrm{e}^{(t-s)A_m}\mathrm{e}^{sA_n} f$$

is continuously differentiable for all $f \in X$, and its derivative is given by

$$[0,t] \ni s \mapsto \mathrm{e}^{(t-s)A_m}(A_m - A_n)\mathrm{e}^{sA_n} f.$$

Using the fundamental theorem of calculus we can conclude that

$$\mathrm{e}^{tA_m} f - \mathrm{e}^{tA_n} f = \int_0^t \mathrm{e}^{(t-s)A_m}(A_m - A_n)\mathrm{e}^{sA_n} f \,\mathrm{d}s = \int_0^t \mathrm{e}^{(t-s)A_m}\mathrm{e}^{sA_n}(A_m - A_n) f \,\mathrm{d}s,$$

where in the last step we used the commutativity assumption. As a consequence we obtain

$$\|\mathrm{e}^{tA_m} f - \mathrm{e}^{tA_n} f\| \leq tM^2\mathrm{e}^{\omega t}\|A_m f - A_n f\|.$$

This shows that for all $f \in D$, the functions $u_n : [0,\infty) \to \mathscr{L}(X)$ defined by $u_n(t) = \mathrm{e}^{tA_n} f$ form a Cauchy sequence in each of the Banach spaces $\mathrm{C}([0,t_0]; X)$ for $t_0 \geq 0$. Therefore we can define

$$T(t)f := \lim_{n\to\infty} \mathrm{e}^{tA_n} f,$$

and the convergence is uniform on every interval $[0, t_0]$ with $t_0 \geq 0$. From this convergence it follows that the operator is linear with $\|T(t)f\| \leq M\mathrm{e}^{\omega t}\|f\|$ for all $t \geq 0$ and $f \in D$. Hence $T(t)$ extends to a bounded linear operator on X. The next properties are also consequences of the convergence above:

1. We have $T(0) = I$ and $T(t+s) = T(t)T(s)$ for all $t, s \geq 0$.

2. The function $t \mapsto T(t)f$ is continuous for all $f \in D$.

From Proposition 2.5.b) it follows that T is a strongly continuous semigroup on X.

Let us denote by B the generator of T. Our aim i to show that $B = \overline{A}$. Since $\mathrm{e}^{tA_n}f$ converges locally uniformly to $T(t)f$, we conclude by the first Trotter–Kato theorem, Theorem 3.14 that

$$R(\lambda, B)f = \lim_{n\to\infty} R(\lambda, A_n)f.$$

But Proposition 5.3 yields

$$R(\lambda)f = \lim_{n\to\infty} R(\lambda, A_n)f,$$

where the range of $R(\lambda)$ is dense. By Proposition 5.5 the operator $R(\lambda)$ is injective and $\overline{R(\lambda)} = R(\lambda, B)$, so again Proposition 5.3 implies $\overline{R(\lambda)} = R(\lambda, \overline{A})$. These yield $R(\lambda, B) = R(\lambda, \overline{A})$, hence $\overline{A} = B$. □

The previous proposition provides some means for proving that a given operator A, or more precisely its closure $\overline{A}$, is a generator of some semigroup. Now suppose A is an operator for which the implicit Euler scheme is defined (see Example 4.3). If A was a generator, then of course the Euler scheme would be convergent. Let us look at if we can obtain the convergence *without assuming* the generator property of A. We sketch one strategy how to do this. Fix $t > 0$ and take

$$A_h := \tfrac{1}{h}AR(\tfrac{1}{h}, A) = \tfrac{1}{h}(\tfrac{1}{h}R(\tfrac{1}{h}, A) - I) \in \mathscr{L}(X),$$

the **Yosida approximants**, where $h = \frac{t}{n}$. Then $A_h f \to Af$ for all $f \in D(A)$ as $h \searrow 0$ (see Proposition 4.8), and we immediately obtain the convergence of $\mathrm{e}^{\cdot A_h}$ to a semigroup T. To show the convergence of the implicit Euler method, it remains to estimate

$$\Big\|\mathrm{e}^{tA_h}f - \Big(\tfrac{1}{h}R(\tfrac{1}{h}, A)\Big)^n f\Big\| = \Big\|\mathrm{e}^{n\big(\frac{1}{h}R(\frac{1}{h},A)-I\big)}f - \Big(\tfrac{1}{h}R(\tfrac{1}{h}, A)\Big)^n f\Big\|.$$

To be able to do that we need the following general result, which is a straightforward generalisation of a corresponding scalar statement.

Lemma 5.7. *Let $S \in \mathscr{L}(X)$ be a power bounded operator, i.e., suppose $\|S^m\| \leq M$ for all $m \in \mathbb{N}$ and some $M \geq 0$. Then*

$$\Big\|\mathrm{e}^{n(S-I)}f - S^n f\Big\| \leq \sqrt{n}M\|Sf - f\| \tag{5.2}$$

for every $n \in \mathbb{N}$ and $f \in X$.

Proof. To prove that we only need some elementary calculus. Fix $n \in \mathbb{N}$ and, by using the power series representation of the exponential function, note that

$$\mathrm{e}^{n(S-I)} - S^n = \mathrm{e}^{-n}\left(\mathrm{e}^{nS} - \mathrm{e}^n S^n\right) = \mathrm{e}^{-n}\sum_{k=0}^{\infty}\frac{n^k}{k!}\left(S^k - S^n\right). \tag{5.3}$$

For $k, n \in \mathbb{N}_0$ we have

$$S^k - S^n = \begin{cases} \displaystyle\sum_{i=n}^{k-1}(S^{i+1} - S^i) & \text{if } k \geq n, \\ \displaystyle\sum_{i=k}^{n-1}(S^i - S^{i+1}) & \text{if } k < n. \end{cases}$$

By using $\|S^m\| \leq M$, we obtain

$$\|S^k f - S^n f\| \leq |n - k| \cdot M \cdot \|Sf - f\|.$$

Substituteing this in (5.3) we obtain

$$\|e^{n(S-I)}f - S^n f\| \le e^{-n} M \|Sf - f\| \sum_{k=0}^{\infty} \frac{n^k}{k!} |n-k|.$$

By the Cauchy–Schwartz inequality we can estimate this further as

$$\begin{aligned}\|e^{n(S-I)}f - S^n f\| &\le e^{-n} M \|Sf - f\| \left(\sum_{k=0}^{\infty} \frac{n^k}{k!}\right)^{\frac{1}{2}} \left(\sum_{k=0}^{\infty} \frac{n^k}{k!} |n-k|^2\right)^{\frac{1}{2}} \\ &= e^{-n} M \|Sf - f\| (e^n)^{\frac{1}{2}} (ne^n)^{\frac{1}{2}} = \sqrt{n} M \|Sf - f\|.\end{aligned}$$

In the last line we used the identity

$$\sum_{k=0}^{\infty} \frac{n^k}{k!} (n-k)^2 = ne^n.$$ □

Before completing the *"proof"* of the convergence of the Euler scheme, let us formulate a more general product formula. The following important result is shown in the commuting case first, since its proof relies on the second Trotter–Kato approximation theorem (at this point only available for commuting approximations).

Proposition 5.8 (Commuting Chernoff Product Formula)**.** *Consider a function*

$$F : [0, \infty) \to \mathscr{L}(X)$$

with $F(t)F(s) = F(s)F(t)$ for all $t, s > 0$ and $F(0) = I$. Suppose that for some $\omega \in \mathbb{R}$ and $M \ge 0$

$$\big\|F(t)^n\big\| \le M e^{\omega t n} \quad \textit{for all } n \in \mathbb{N}, \tag{5.4}$$

and that there exists $D \subset X$ such that $(\lambda - A)D$ is dense for some $\lambda > 0$ and

$$Af := \lim_{h \searrow 0} \frac{F(h)f - f}{h}$$

exists for all $f \in D$. Then the closure $\overline{A}$ of A generates a bounded strongly continuous semigroup T which is given by

$$T(t)f := \lim_{n \to \infty} \big(F(\tfrac{t}{n})\big)^n f$$

for all $f \in X$. The convergence here is locally uniform in t.

Proof. By replacing $F(t)$ with $e^{-\omega t}F(t)$ and A with $A - \omega$ we may suppose $\omega = 0$. For $h > 0$ define

$$A_h := \frac{F(h) - I}{h} \in \mathscr{L}(X).$$

By assumption we have $A_h f \to Af$ for all $f \in D$ as $h \searrow 0$. Furthermore, using that

$$e^{tA_h} = e^{\frac{t}{h}(F(h)-I)} = e^{-\frac{t}{h}} e^{\frac{t}{h}F(h)},$$

we can estimate

$$\|e^{tA_h}\| \le e^{-\frac{t}{h}} \sum_{n=0}^{\infty} \frac{t^n \|F(h)^n\|}{h^n n!} \le M e^{-\frac{t}{h}} \sum_{n=0}^{\infty} \frac{t^n}{h^n n!} = M.$$

This shows that the conditions of Proposition 5.6 are satisfied, meaning that $\overline{A}$ generates a strongly continuous semigroup T given by

$$T(t)f = \lim_{n\to\infty} \mathrm{e}^{tA_{\frac{t}{n}}} f = \lim_{n\to\infty} \mathrm{e}^{n(F(\frac{t}{n})-I)} f,$$

where the convergence is uniform for $t \in (0, t_0]$, for every $t_0 > 0$. By the assumption $F(0) = I$, we obtain

$$T(t)f = \lim_{n\to\infty} \mathrm{e}^{n(F(\frac{t}{n})-I)} f$$

uniformly on $[0, t_0]$.

On the other hand, by Lemma 5.7 we have for $f \in D$

$$\|F(\tfrac{t}{n})f - f\| = \frac{t}{n}\|A_{\frac{t}{n}} f\|,$$

and hence

$$\left\|\mathrm{e}^{n(F(\frac{t}{n})-I)} f - F(\tfrac{t}{n})^n f\right\| \le \sqrt{n} M \|F(\tfrac{t}{n})f - f\| = \frac{tM}{\sqrt{n}}\|A_{\frac{t}{n}} f\| \to 0$$

as $n \to \infty$ with locally uniform convergence in t. By assumption (5.4) we can apply Theorem 2.30 and conclude the proof (this last mentioned theorem does not explicitly yield the local uniform convergence in t, to obtain that one needs a small twist, see Exercise 6). □

Returning to the Euler scheme note that up to now we did not say a word about the stability, which is certainly needed if we long for convergence.

Proposition 5.9. *For an operator A the following assertions are equivalent:*

(i) There exist constants $M \ge 1$ and $\omega \in \mathbb{R}$ so that $(\omega, \infty) \subseteq \rho(A)$ and

$$\|R(\lambda, A)^n\| \le \frac{M}{(\lambda - \omega)^n} \quad \textit{for all } \lambda > \omega \textit{ and } n \in \mathbb{N}. \tag{5.5}$$

(ii) There exist constants $K \ge 1$ and $\omega' \ge 0$ so that $(\omega', \infty) \subset \rho(A)$ and

$$\left\|\left(\tfrac{1}{h}R(\tfrac{1}{h}, A)\right)^k\right\| \le K\mathrm{e}^{kh\omega'} \quad \textit{for all } k \in \mathbb{N} \textit{ and all } h \in (0, \tfrac{1}{\omega'}) \tag{5.6}$$

(in case $\omega' = 0$ the interval extends to ∞).

Proof. In both implications we use the substitution $h = \frac{1}{\lambda}$. If (ii) is true, then we can write

$$\left\|\left(\lambda R(\lambda, A)\right)^n\right\| \le K\mathrm{e}^{n\frac{\omega'}{\lambda}}.$$

and hence

$$\left\|\left((\lambda - \omega')R(\lambda, A)\right)^n\right\| \le K\mathrm{e}^{n\frac{\omega'}{\lambda}}\left(1 - \tfrac{\omega'}{\lambda}\right)^n \le K,$$

for all $n \in \mathbb{N}$ and $\lambda > \omega'$ meaning that (i) holds with $\omega = \omega'$ and $M = K$.

Suppose now that (i) holds. Then for all $k \in \mathbb{N}$ and $\lambda > \max\{0, \omega\}$ we have

$$\left\|\left(\lambda R(\lambda, A)\right)^k\right\| \le M\tfrac{\lambda^k}{(\lambda-\omega)^k} \le M\mathrm{e}^{k\frac{\omega}{\lambda-\omega}}.$$

So, in case $\omega \le 0$, we can set $\omega' := \omega$, $K := M$ and obtain (ii). Otherwise take $\omega' > \omega > 0$ arbitrary. Then for $\lambda > \omega'$ we have

$$\left\|\left(\lambda R(\lambda, A)\right)^k\right\| \le M\mathrm{e}^{k\frac{\omega}{\lambda}\cdot\frac{1}{1-\frac{\omega}{\lambda}}} \le M\mathrm{e}^{k\frac{\omega'}{\lambda}\cdot\frac{1}{1-\frac{\omega}{\omega'}}}.$$

Hence, (ii) holds with the choice $K = M\mathrm{e}^{\frac{\omega'}{\omega'-\omega}}$. □

Operators satisfying (5.5) are called **Hille–Yosida operators**[4]. We see therefore that the stability of the Euler scheme for A is equivalent to the fact that A is a Hille–Yosida operator.

Theorem 5.10 (Hille–Yosida)**.** *Suppose that A is densely defined Hille-Yosida operator. Then A is the generator of a strongly continuous semigroup T given by the implicit Euler method. More precisely, for every $t_0 > 0$ and $f \in X$ we have*

$$T(t)f = \lim_{n\to\infty} \left(\tfrac{n}{t}R(\tfrac{n}{t}, A)\right)^n f = \lim_{n\to\infty} \left(I - \tfrac{t}{n}A\right)^{-n} f$$

with uniform convergence for $t \in [0, t_0]$.

Proof. Let $\omega' > \max\{\omega, 0\}$. As sketched above we apply Proposition 5.6 to the function

$$F(h) := \begin{cases} I & \text{for } h = 0, \\ \frac{1}{h}R(\frac{1}{h}, A) & \text{for } h \in (0, \frac{1}{\omega'}), \\ \omega' R(\omega', A) & \text{for } h \geq \frac{1}{\omega'}. \end{cases}$$

Stability follows from Proposition 5.9. Further, from the identity $\lambda R(\lambda, A) - I = AR(\lambda, A)$ we conclude that

$$\frac{F(h)f - f}{h} = \tfrac{1}{h}R(\tfrac{1}{h}, A)Af \to Af \quad \text{for } f \in D(A).$$

Since for $\lambda > \omega$ we have $(\lambda - A)D(A) = X$, all the conditions of Proposition 5.6 are satisfied, and the proof is complete. □

5.3 General approximation theorems

We turn our attention to the general form of the approximation theorems presented in the previous section and try to get rid of the commutation assumption.

Theorem 5.11 (Second Trotter–Kato Approximation Theorem)**.** *For $n \in \mathbb{N}$ let A_n generate the semigroup T_n, and suppose that all T_n have the same type (M, ω). Then the following assertions are equivalent:*

(i) There is a densely defined linear operator $A : D \to X$ such that $(\lambda - A)D$ is dense for some $\lambda > \omega$. Moreover, for all $f \in D$ there is $f_n \in D(A_n)$ with

$$f_n \to f \quad \textit{and} \quad A_n f_n \to Af \quad \textit{for } n \to \infty.$$

(ii) The limit

$$R(\lambda)f = \lim_{n\to\infty} R(\lambda, A_n)f$$

exists for all $f \in X$ and for some (and then for all) $\lambda > \omega$. The operator $R(\lambda)$ has dense range.

(iii) There is a semigroup T with generator B such that

$$T_n(t)f \to T(t)f \quad \textit{as } n \to \infty$$

for all $f \in X$ locally uniformly in t.

[4] E. Hille, Functional Analysis and Semigroups, Amer. Math.Soc. Coll. Publ., vol. **31**, Amer. Math. Soc., 1948. and K. Yosida, On the differentiability and the representation of one-parameter semigroups of linear operators, J. Math. Soc. Japan **1** (1948), 1521.

Moreover, under these equivalent conditions, we have $B = \overline{A}$, and $R(\lambda) = R(\lambda, B)$ for all $\lambda > \omega$.

Proof. By rescaling, i.e., by replacing $T_n(t)$ by $\mathrm{e}^{-\omega t}T_n(t)$, we may suppose $\omega = 0$ (cf. Exercise C.3). Proposition 5.3 yields the implication (i) $\Rightarrow$ (ii), and since $R(\lambda)$ is injective by Proposition 5.5, we obtain $R(\lambda) = R(\lambda, \overline{A})$.

Suppose that in (ii) $R(\lambda)$ exists for some $\lambda > 0$, then by Proposition 5.4 $R(\mu)$ exists for all $\mu > 0$, and by Proposition 5.5 $R(\mu)$ are all injective. So Proposition 5.3 yields the implication (ii) $\Rightarrow$ (i). By Proposition 5.5 we have that $R(\lambda) = R(\lambda, B)$ for a closed operator B, and we even obtain $R(\mu) = R(\mu, B)$ (why?). From this we infer

$$\|\lambda^n R(\lambda)^n\| = \|\lambda^n R(\lambda, B)^n\| \leq M \quad \text{for all } \lambda > 0.$$

Hence by the Hille–Yosida theorem, Theorem 5.10 operator B generates a semigroup T, and by the first Trotter–Kato theorem, Theorem 3.14 we see $R(\lambda) = R(\lambda, B)$. Hence (iii) is proved.

The implication (iii) $\Rightarrow$ (i) follows from the first Trotter–Kato Theorem 3.14. □

Now one can easily prove Chernoff's theorem in the following general form. The proof is exactly the same as for the commutative version, one only needs to apply the second Trotter–Kato theorem from above.

Theorem 5.12 (Chernoff Product Formula)**.** *Let*

$$F : [0, \infty) \to \mathscr{L}(X)$$

be a function with $F(0) = I$ such that for some $\omega \geq 0$ and $M \geq 0$ we have

$$\|F(t)^n\| \leq M\mathrm{e}^{\omega n t} \qquad \textit{for all } n \in \mathbb{N},\ t \geq 0.$$

Suppose furthermore that there is $D \subset X$ such that the limit

$$Af := \lim_{h \searrow 0} \frac{F(h)f - f}{h}$$

exists for all $f \in D$, and that $(\lambda - A)D$ is dense for some $\lambda > \omega$. Then the closure $\overline{A}$ of A generates a strongly continuous semigroup T which is given by

$$T(t)f := \lim_{n \to \infty} \left(F(\tfrac{t}{n})\right)^n f$$

for all $f \in X$, and the convergence is uniform for $t \in [0, t_0]$ for each $t_0 > 0$.

5.4 Exercises

1. Prove Proposition 5.2.

2. Prove the identity:

$$\sum_{k=0}^{\infty} \frac{n^k}{k!}(n-k)^2 = n\mathrm{e}^n$$

needed in Lemma 5.7.

3. Prove that in Proposition 5.5 for $\lambda, \mu > \omega$ one has $R(\lambda) = R(\lambda, B)$ and $R(\mu) = R(\mu, B)$ for the same operator B.

4. Consider the Banach space $X := \ell^2$ and recall that for a sequence $m \subseteq \mathbb{C}$ the multiplication operator corresponding to m is denoted by M_m. Now for $n \in \mathbb{N}$ denote by $\mathbf{1}_{\{1,2,\ldots,n}$ the characteristic sequence of the set $\{1, 2, \ldots, n\}$. For a given sequence $m \subseteq \mathbb{C}$ define $m_n := m \cdot \mathbf{1}_{\{1,\ldots,n\}}$ and $A_n := M_{m_n}$ the corresponding multiplication operators. Check the various conditions of the second Trotter–Kato theorem for this sequence of operators.

5. Let A be a generator of a semigroup T on the Banach space X, and let $B \in \mathscr{L}(X)$ be a bounded linear operator. Prove by means of suitable approximations (and not using the Hille–Yosida Theorem) that $A + B$ with $D(A + B) = D(A)$ is a generator of a semigroup.

6. Do the twist in the proof of Proposition 5.8. More precisely, prove that if $F, F_n : [0, t_0] \to \mathscr{L}(X)$ are strongly continuous functions that are uniformly bounded, then the following assertions are equivalent.

(i) $F_n(t)x \to F(t)x$ uniformly on $[0, t_0]$ as $n \to \infty$ for each $x \in X$.

(ii) $F_n(t)x \to F(t)x$ uniformly on $[0, t_0]$ as $n \to \infty$ for each $x \in D$ from a dense subspace D.

(iii) $F_n(t)x \to F(t)x$ uniformly on $[0, t_0] \times K$ as $n \to \infty$ for each compact set $K \subseteq X$.

Lecture 6

The Lumer–Phillips Theorem

In the previous lecture we saw the characterisation of generators of strongly continuous semigroups, called Hille–Yosida theorem. Unfortunately, even in the case of relatively simple problems, it is practically impossible to check all the properties listed: It is already difficult to estimate the operator norm of the resolvent, let alone all powers of it. We also have to make sure that our operator is closed, which also might be a painful task in particular situations.

In this lecture we study a class of operators, for which the above two difficulties may be remedied in a satisfactory way.

6.1 Dissipative operators

Due to their importance, we now return to the study of **contraction semigroups**, i.e., semigroups T where the semigroup operators are contractive, and look for a characterisation of their generator that does not require explicit knowledge of the resolvent. The following is a key notion towards this goal.

Definition 6.1. A linear operator A on a Banach space X is called **dissipative** if

$$\|(\lambda - A)f\| \geq \lambda \|f\| \tag{6.1}$$

for all $\lambda > 0$ and $f \in D(A)$.

Note that it suffices to establish the validity of the inequality above only for unit vectors $f \in X$, $\|f\| = 1$. For $f = 0$ the inequality is trivial, for $f \neq 0$ one can normalise. Note also that we did not require here the density of the domain or any other analytic properties of the operator. To familiarise ourselves with dissipative operators we state some of their basic properties.

Proposition 6.2. *For a dissipative operator A the following properties hold.*

a) $\lambda - A$ is injective for all $\lambda > 0$ and

$$\left\|(\lambda - A)^{-1}g\right\| \leq \frac{1}{\lambda}\|g\|$$

for all g in the range $\operatorname{ran}(\lambda - A) := (\lambda - A)D(A)$.

b) $\lambda - A$ is surjective for some $\lambda > 0$ if and only if it is surjective for each $\lambda > 0$. In that case, one has $(0, \infty) \subset \rho(A)$.

c) A is closed if and only if the range $\operatorname{ran}(\lambda - A)$ *is closed for some (hence all) $\lambda > 0$.*

d) If $\operatorname{ran}(A) \subset \overline{D(A)}$, *e.g., if A is densely defined, then A is closable. Its closure $\overline{A}$ is again dissipative and satisfies* $\operatorname{ran}(\lambda - \overline{A}) = \overline{\operatorname{ran}(\lambda - A)}$ *for all $\lambda > 0$.*

Proof. a) is just a reformulation of estimate (6.1).

To show b) we assume that $(\lambda_0 - A)$ is surjective for some $\lambda_0 > 0$. In combination with a), this yields $\lambda_0 \in \rho(A)$ and $\|R(\lambda_0, A)\| \leq \frac{1}{\lambda_0}$. The series expansion for the resolvent

$$R(\lambda, A_n) = \sum_{k=0}^{\infty} (\lambda_0 - \lambda)^k R(\lambda_0, A_n)^{k+1}$$

yields $(0, 2\lambda_0) \subset \rho(A)$. The dissipativity of A implies that

$$\|R(\lambda, A)\| \leq \frac{1}{\lambda}$$

for $0 < \lambda < 2\lambda_0$. Proceeding in this way, we see that $\lambda - A$ is surjective for all $\lambda > 0$, and therefore $(0, \infty) \subset \rho(A)$.

c) The operator A is closed if and only if $\lambda - A$ is closed for some (hence all) $\lambda > 0$. This is again equivalent to

$$(\lambda - A)^{-1} : \operatorname{ran}(\lambda - A) \to D(A)$$

being closed. By a) this operator is bounded. Hence, by the closed graph theorem, see Theorem 2.32, it is closed if and only if its domain, i.e., $\operatorname{ran}(\lambda - A)$, is closed.

d) Take a sequence $f_n \in D(A)$ satisfying $f_n \to 0$ and $Af_n \to g$. By Proposition 5.2.a) we have to show that $g = 0$. The inequality (6.1) implies that

$$\|\lambda(\lambda - A)f_n + (\lambda - A)w\| \geq \lambda \, \|\lambda f_n + w\|$$

for every $w \in D(A)$ and all $\lambda > 0$. Passing to the limit as $n \to \infty$ yields

$$\|{-\lambda g} + (\lambda - A)w\| \geq \lambda \|w\| \quad \text{and hence} \quad \left\|-g + w - \tfrac{1}{\lambda}Aw\right\| \geq \|w\|.$$

For $\lambda \to \infty$ we obtain that

$$\| - g + w\| \geq \|w\|$$

and by choosing w from the domain $D(A)$ arbitrarily close to $g \in \overline{\operatorname{ran}(A)}$, we see that

$$0 \geq \|g\|.$$

Hence $g = 0$.

In order to verify that $\overline{A}$ is dissipative, take $f \in D(\overline{A})$. By definition of the closure of a linear operator, there exists a sequence $f_n \in D(A)$ satisfying $f_n \to f$ and $Af_n \to \overline{A}f$ when $n \to \infty$. Since A is dissipative and the norm is continuous, this implies that $\|(\lambda - \overline{A})f\| \geq \lambda\|f\|$ for all $\lambda > 0$. Hence $\overline{A}$ is dissipative. Finally, observe that the range $\operatorname{ran}(\lambda - A)$ is dense in $\operatorname{ran}(\lambda - \overline{A})$. Since by assertion c) $\operatorname{ran}(\lambda - \overline{A})$ is closed in X, we obtain the final assertion in d). □

From the resolvent estimate in the Hille–Yosida theorem, Theorem 5.10, it is evident that the generator of a contraction semigroup satisfies the estimate (6.1), and hence is dissipative. On the other hand, as we shall see in a moment, many operators can be shown directly to be dissipative and densely defined. Therefore we reformulate Theorem 5.10 in such a way as to single out the property that ensures that a densely defined, dissipative operator is a generator.

Theorem 6.3 (Lumer–Phillips)**.** *For a densely defined, dissipative operator A on a Banach space X the following statements are equivalent:*

(i) The closure $\overline{A}$ of A generates a contraction semigroup.

(ii) The range $\operatorname{ran}(\lambda - A)$ *is dense in X for some (hence all) $\lambda > 0$.*

Proof. (i) ⇒ (ii) The Hille–Yosida theorem, Theorem 5.10, implies that $\operatorname{ran}(\lambda - \overline{A}) = X$ for all $\lambda > 0$. Since by Proposition 6.2.d) $\operatorname{ran}(\lambda - \overline{A}) = \overline{\operatorname{ran}(\lambda - A)}$, we obtain (ii).

(ii) ⇒ (i) By the same argument, the denseness of the range $\operatorname{ran}(\lambda - A)$ implies that $(\lambda - \overline{A})$ is surjective. Proposition 6.2.b) shows that $(0, \infty) \subset \rho(\overline{A})$, and dissipativity of A implies the estimate

$$\|R(\lambda, \overline{A})\| \leq \frac{1}{\lambda} \quad \text{for } \lambda > 0.$$

This was required in Theorem 5.10 to assure that $\overline{A}$ generated a contraction semigroup. □

The above theorem gains its significance when viewed in the context of the abstract Cauchy problem associated to an operator A.

Remark 6.4. Assume that the operator A is known to be closed, densely defined, and dissipative. Then the Lumer–Phillips theorem, Theorem 6.3 yields the following fact:

In order to ensure that the (time dependent) initial value problem

$$\dot{u}(t) = Au(t),\, u(0) = u_0 \tag{ACP}$$

can be solved for all $u_0 \in D(A)$, it is sufficient to prove that the (stationary) resolvent equation

$$f - Af = g \tag{RE}$$

has solutions for all g in some dense subset in the Banach space X. As an example recall the treatment of the heat equation presented in Section 1.1. In many examples (RE) can be solved explicitly while (ACP) cannot.

Let us investigate the question further how to decide whether an operator is dissipative. When introducing dissipative operators, we had aimed for an easy (or at least more direct) way to characterising generators. Up to now, however, the only way to arrive at the norm inequality (6.1) was by explicit computation of the resolvent and then deducing the norm estimate

$$\|R(\lambda, A)\| \leq \frac{1}{\lambda} \qquad \text{for } \lambda > 0.$$

Fortunately, there is a simpler method that works particularly well in concrete function spaces such as $\mathrm{C}_0(\Omega)$ or $\mathrm{L}^p(\Omega, \mu)$. Due to its importance and since this is the simplest case, we start with the Hilbert space case.

Proposition 6.5. *Let X be a Hilbert space. An operator A is dissipative if and only if for every $f \in D(A)$ we have*

$$\operatorname{Re}\langle Af, f\rangle \leq 0. \tag{6.2}$$

Note that in this theorem the important direction is that (6.2) implies dissipativity. Fortunately, this is also easy to prove.

Proof. Assume (6.2) is satisfied for $f \in D(A)$, $\|f\| = 1$. Then we have

$$\begin{aligned}\|\lambda f - Af\| &\geq |\langle \lambda f - Af, f\rangle| \\ &\geq \operatorname{Re}\langle \lambda f - Af, f\rangle \geq \lambda\end{aligned}$$

for all $\lambda > 0$. This proves one of the implications.

To show the converse, we take $f \in D(A)$, $\|f\| = 1$, and assume that $\|\lambda f - Af\| \geq \lambda$ for all $\lambda > 0$. Consider the normalised elements

$$g_\lambda := \frac{\lambda f - Af}{\|\lambda f - Af\|}.$$

Then for all $\lambda > 0$ we have

$$\lambda \leq \|\lambda f - Af\| = \langle \lambda f - Af, g_\lambda\rangle = \lambda \operatorname{Re}\langle f, g_\lambda\rangle - \operatorname{Re}\langle Af, g_\lambda\rangle.$$

By estimating one of the terms on right-hand side trivially we can conclude the following two inequalities:

$$\lambda \leq \lambda - \operatorname{Re}\langle Af, g_\lambda\rangle \quad \text{and} \quad \lambda \leq \lambda \operatorname{Re}\langle f, g_\lambda\rangle + \|Af\|$$

are valid for each $\lambda > 0$. These yield for $\lambda = n$

$$\operatorname{Re}\langle Af, g_n\rangle \leq 0 \qquad \text{and} \qquad 1 - \frac{1}{n}\|Af\| \leq \operatorname{Re}\langle f, g_n\rangle.$$

Since the unit ball of a Hilbert space is weakly (sequentially) compact, we can take a weakly convergent subsequence (g_{n_k}) with weak limit $g \in H$. Then we obtain

$$\|g\| \leq 1, \qquad \operatorname{Re}\langle Af, g\rangle \leq 0, \qquad \text{and} \qquad \operatorname{Re}\langle f, g\rangle \geq 1.$$

Combining these facts, it follows that $g = f$ and that it satisfies (6.2). □

To introduce the general case we start with a Banach space X and its dual space X'. By the Hahn–Banach theorem, see Theorem 6.16, for every $f \in X$ there exists $\phi \in X'$ such that

$$\phi(f) = \langle f, \phi\rangle = \|f\|^2 = \|\phi\|^2$$

holds. Hence, for every $f \in X$ the following set, called its **duality set**,

$$J(f) := \big\{\phi \in X' : \langle f, \phi\rangle = \|f\|^2 = \|\phi\|^2\big\}, \tag{6.3}$$

is nonempty. Such sets allow a new characterisation of dissipativity.

Proposition 6.6. *An operator A is dissipative if and only if for every $f \in D(A)$ there exists $j(f) \in J(f)$ such that*

$$\operatorname{Re}\langle Af, j(f)\rangle \leq 0. \tag{6.4}$$

If A is the generator of a strongly continuous contraction semigroup, then (6.4) *holds for all $f \in D(A)$ and arbitrary $\phi \in J(f)$.*

Proof. Assume (6.4) is satisfied for $f \in D(A)$, $\|f\| = 1$, and some $j(f) \in J(f)$. Then $\langle f, j(f)\rangle = \|j(f)\|^2 = 1$ and

$$\|\lambda f - Af\| \geq |\langle \lambda f - Af, j(f)\rangle| \geq \operatorname{Re}\langle \lambda f - Af, j(f)\rangle \geq \lambda$$

for all $\lambda > 0$. This proves the important implication. The other implication is only included for the sake of completeness, you may skip this part on the first reading.

To show the converse, we take $f \in D(A)$, $\|f\| = 1$, and assume that $\|\lambda f - Af\| \geq \lambda$ for all $\lambda > 0$. Choose $\phi_\lambda \in J(\lambda f - Af)$ and consider the normalised elements

$$\psi_\lambda := \frac{\phi_\lambda}{\|\phi_\lambda\|}.$$

Then, similarly to the proof of Proposition 6.5, the inequalities

$$\begin{aligned}\lambda \leq \|\lambda f - Af\| &= \langle \lambda f - Af, \psi_\lambda\rangle = \lambda \operatorname{Re}\langle f, \psi_\lambda\rangle - \operatorname{Re}\langle Af, \psi_\lambda\rangle \\ &\leq \min\{\lambda - \operatorname{Re}\langle Af, \psi_\lambda\rangle,\ \lambda \operatorname{Re}\langle f, \psi_\lambda\rangle + \|Af\|\}\end{aligned}$$

are valid for each $\lambda > 0$. This yields for $\lambda = n$

$$\operatorname{Re}\langle Af, \psi_n\rangle \leq 0 \qquad \text{and} \qquad 1 - \frac{1}{n}\|Af\| \leq \operatorname{Re}\langle f, \psi_n\rangle.$$

Let ψ be a weak* accumulation point of (ψ_n), which exists by the Banach–Alaoglu theorem, see Theorem 6.17. Then

$$\|\psi\| \leq 1, \qquad \operatorname{Re}\langle Af, \psi\rangle \leq 0, \qquad \text{and} \qquad \operatorname{Re}\langle f, \psi\rangle \geq 1.$$

Combining these facts, it follows that ψ belongs to $J(f)$ and satisfies (6.4).

Finally, suppose that A generates a contraction semigroup T on X. Then, for every $f \in D(A)$ and arbitrary $\phi \in J(f)$, we have

$$\operatorname{Re}\langle Af, \phi\rangle = \lim_{h \searrow 0}\Big(\frac{\operatorname{Re}\langle T(h)f, \phi\rangle}{h} - \frac{\operatorname{Re}\langle f, \phi\rangle}{h}\Big) \leq \limsup_{h \searrow 0}\Big(\frac{\|T(h)f\| \cdot \|\phi\|}{h} - \frac{\|f\|^2}{h}\Big) \leq 0.$$

This completes the proof. □

Remark 6.7. Note that the requirement in (6.4) can be relaxed in many applications to

$$\operatorname{Re}\langle Af, j(f)\rangle \leq \omega \tag{6.5}$$

for some given $\omega \geq 0$. Operators with this property are called **quasi-dissipative**. Clearly, if A is quasi-dissipative, then $A - \omega$ is dissipative.

6.2 Examples

We continue here with a discussion of these new notions and results in concrete examples. We begin with identifying the duality sets $J(f)$ for some classical function spaces.

Example 6.8. 1. Let Ω be a locally compact Hausdorff space (for example an open or a closed subset of $\mathbb{R}^d$). Consider

$$X := \mathrm{C}_0(\Omega) := \bigl\{f : f \text{ is continuous and vansihes at infinity}\bigr\}.$$

This is a Banach space with the supremum norm $\|\cdot\|_\infty$. For $0 \neq f \in X$, the set $J(f) \subset X'$ contains (multiples of) all point measures supported by those points $s_0 \in \Omega$ where $|f|$ reaches its maximum. More precisely,

$$\Bigl\{\overline{f(s_0)} \cdot \delta_{s_0} : \; s_0 \in \Omega \text{ and } |f(s_0)| = \|f\|_\infty\Bigr\} \subset J(f). \tag{6.6}$$

2. Let $(\Omega, \mathscr{A}, \mu)$ be a σ-finite measure space, let $p \in [1, \infty)$ and $X := \mathrm{L}^p(\Omega, \mathscr{A}, \mu)$. Then $X' = \mathrm{L}^q(\Omega, \mathscr{A}, \mu)$, where $\frac{1}{p} + \frac{1}{q} = 1$. For $0 \neq f \in X$ define

$$\phi(s) := \begin{cases} \overline{f(s)} \cdot |f(s)|^{p-2} \cdot \|f\|^{2-p} & \text{if } f(s) \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{6.7}$$

Then

$$\phi \in J(f) \subset \mathrm{L}^q(\Omega, \mathscr{A}, \mu).$$

We note here without proof that for the reflexive L^p spaces (i.e., for $1 < p < \infty$), as for every Banach space with a strictly convex dual, the sets $J(f)$ are singletons. Hence, for $p \in (1, \infty)$ one has $J(f) = \{\phi\}$, while for $p = 1$ every function $\phi \in \mathrm{L}^\infty(\Omega, \mathscr{A}, \mu)$ satisfying

$$\|\phi\|_\infty \leq \|f\|_1 \qquad \text{and} \qquad \phi(s)\,|f(s)| = \overline{f(s)}\,\|f\|_1 \quad \text{if } f(s) \neq 0$$

belongs to $J(f)$, i.e., on the set $\{s \in \Omega : f(s) = 0\}$ we can give arbitrary values to ϕ as long as they are smaller than $\|f\|_1$.

3. It is easy, but important, to determine $J(f)$ in case of $f \in H$, H a Hilbert space. After the canonical identification of H with its dual H', the duality set of $f \in H$ is

$$J(f) = \{f\}.$$

Hence, a linear operator on H is dissipative if and only if

$$\operatorname{Re}\langle Af,\, f\rangle \leq 0$$

for all $f \in D(A)$ in accordance with Proposition 6.5.

Let us list now some important operators where dissipativity can be tested. For simplicity, we concentrate here only on the point how to test dissipativity.

Example 6.9. Consider the Laplace operator with Dirichlet boundary conditions from Section 1.1, i.e., we take $X = \mathrm{L}^2(0, \pi)$ and consider the operator

$$(Af)(x) := f''(x) = \frac{\mathrm{d}^2}{\mathrm{d}x^2} f(x)$$

with domain

$$\begin{aligned} D(A) := \Bigl\{ f \in \mathrm{L}^2(0,\pi) : {}& f \text{ cont. differentiable on } [0,\pi], \\ & f'' \text{ exists a.e., } f'' \in \mathrm{L}^2,\ f'(x) - f'(0) = \textstyle\int_0^x f''(s)\,\mathrm{d}s \text{ for } x \in [0,\pi] \\ & \text{and } f(0) = f(\pi) = 0 \Bigr\}. \end{aligned}$$

Clearly,

$$\langle Af, f\rangle = \int_0^\pi f''(s)\overline{f(s)}\,\mathrm{d}s = -\int_0^\pi f'(s)\overline{f'(s)}\,\mathrm{d}s = -\|f'\|^2 \le 0,$$

showing the dissipativity of A.

The previous example immediately gives rise to certain generalisations.

Example 6.10. Let $A = M_m$ be a multiplication operator on ℓ^2 with the sequence $m = (m_n)$. Then A is dissipative if and only if $\operatorname{Re} m_n \le 0$ for all $n \in \mathbb{N}$.

Let us analyse now the second derivative in the space of continuous functions. We consider however Neumann boundary conditions.

Example 6.11. Let us consider in $X := \mathrm{C}([0,1])$ the Laplace operator with Neumann boundary conditions given by

$$Af := f'', \quad D(A) := \left\{f \in \mathrm{C}^2([0,1]) \,:\, f'(0) = f'(1) = 0\right\}.$$

To show dissipativity, we use the description of $J(f)$ from Example 6.8.1. Take $f \in D(A)$ and $s_0 \in [0,1]$ such that $|f(s_0)| = \|f\|$. Then by (6.6) we have $\overline{f(s_0)}\delta_{s_0} \in J(f)$. Clearly, the real-valued function

$$g(x) = \operatorname{Re}\left(\overline{f(s_0)}f(x)\right)$$

takes its maximum at $x = s_0$, meaning that if $s_0 \in (0,1)$, then

$$\operatorname{Re}\langle f'', \overline{f(s_0)}\delta_{s_0}\rangle = \left(\operatorname{Re}\overline{f(s_0)}f\right)''(s_0) = g''(s_0) \le 0.$$

If $s_0 = 0$ or $s_0 = 1$, then the boundary condition $f'(s_0) = 0$ implies $g'(s_0) = 0$, and hence $g''(s_0) \le 0$ also in these cases. Hence A is dissipative.

Example 6.12. Consider now the first derivative in various function spaces.

1. Let $X = \mathrm{L}^2(\mathbb{R})$ and $Af = f'$ with

$$D(A) = \mathrm{C}^1_\mathrm{c}(\mathbb{R}) := \left\{f \in \mathrm{C}^1(\mathbb{R}) :\ \text{the support of } f \text{ is compact}\right\}.$$

Then

$$\langle Af, f\rangle = \int_\mathbb{R} f' \cdot \overline{f} = -\int_\mathbb{R} f \cdot \overline{f'} = -\langle f, Af\rangle = -\overline{\langle Af, f\rangle}$$

for $f \in D(A)$, showing that

$$\langle Af, f\rangle + \overline{\langle Af, f\rangle} = 0, \qquad \text{i.e.,} \quad \langle Af, f\rangle \in \mathrm{i}\mathbb{R}.$$

This means that both A and $-A$ are dissipative.

2. Turning our attention to the space of continuous functions, consider

$$X = \mathrm{C}_{(0)}([0,1]) = \{f \in \mathrm{C}([0,1]) \,:\, f(1) = 0\}$$

and $Af = f'$ with $D(A) = \{f \in \mathrm{C}^1([0,1]) \cap X : f' \in X\}$. Suppose f takes its maximum at $s_0 \in [0,1]$. Similarly to Example 6.11 define again the real-valued function

$$g(x) = \mathrm{Re}\left(\overline{f(s_0)}f(x)\right).$$

Then in case $s_0 \in (0,1)$ it follows that

$$\mathrm{Re}\langle f', \overline{f(s_0)}\delta_{s_0}\rangle = \left(\mathrm{Re}\,\overline{f(s_0)}f\right)'(s_0) = g'(s_0) = 0.$$

Since by definition $g'(1) = 0$, we only have to check the case when $s_0 = 0$. But then clearly $g'(s_0) \leq 0$. Hence A is dissipative.

6.3 Perturbations

As an application, let us mention some basic perturbation results. The idea behind perturbation theorems is always the same: We start with a generator A and assume that the operator B is "nice enough". Then $A + B$ generates a semigroup. Let us clarify what "nice enough" could mean here.

As a warm-up, let us recall the results from Exercise 5.5.

Theorem 6.13. *If A generates a semigroup T of type (M, ω) and $B \in \mathscr{L}(X)$, then $A + B$ with $D(A + B) = D(A)$ generates a semigroup S of type $(M, \omega + \|B\|)$.*

Proof. First we change to the operator to $A - \omega$ and then use the renorming procedure presented in Exercise C.4. Then we can assume without the loss of generality that A generates a semigroup of type $(1, 0)$, i.e., a contraction semigroup.

As a next step, we show that the operator $A + B$ has non-empty resolvent set. More precisely, if $\lambda > 0$, we can use the identity

$$\lambda - A - B = (I - BR(\lambda, A))\,(\lambda - A), \tag{6.8}$$

showing that if $\|BR(\lambda, A)\| < 1$, then $\lambda \in \rho(A + B)$ and

$$R(\lambda, A + B) = R(\lambda, A)\sum_{n=0}^{\infty}(BR(\lambda, A))^n. \tag{6.9}$$

By assumption, A is a generator of a contraction semigroup, and hence $\lambda\|R(\lambda, A)\| \leq 1$. Hence, if $\lambda > \|B\|$, then $\lambda \in \rho(A + B)$ and (6.9) holds.

We present here two strategies to continue.

a) Clearly, $A + B - \|B\|$ is dissipative, i.e.,

$$\mathrm{Re}\langle (A + B)f, j(f)\rangle = \mathrm{Re}\langle Af, j(f)\rangle + \mathrm{Re}\langle Bf, j(f)\rangle \leq 0 + \|B\| \cdot \|f\| \cdot \|j(f)\|$$

by the dissipativity of A and the boundedness of B. Since, $\lambda - (A+B)$ is surjective for $\lambda > \|B\|$, we have by the Lumer–Phillips theorem, Theorem 6.3 that $A + B$ generates a semigroup of type $(1, \|B\|)$.

b) We may also use the results of Exercise C.5 and see that for A and B the conditions of Chernoff's theorem, Theorem 5.12 are satisfied. Hence, $A + B$ generates a semigroup. See also Exercise 5.5. □

Clearly, we can immediately extend the previous proof to some unbounded perturbations.

Theorem 6.14. *Let A generate a contraction semigroup and let B be dissipative. Suppose $D(A) \subset D(B)$ and that there is a $\lambda > 0$ with the property that $BR(\lambda, A) \in \mathscr{L}(X)$ and*

$$\|BR(\lambda, A)\| < 1.$$

Then $A + B$ with domain $D(A + B) = D(A)$ generates a contraction semigroup.

We close this lecture by the following example: Recall from Lecture 2 the Gaussian semigroup T on $\mathrm{L}^p(\mathbb{R})$, where $p \in [1, \infty)$. For $f \in \mathrm{L}^p(\mathbb{R})$ we have

$$(T(t)f)(x) := (g_t * f)(x) = \frac{1}{\sqrt{4\pi t}} \int_{\mathbb{R}} f(y) \mathrm{e}^{-\frac{(x-y)^2}{4t}} \,\mathrm{d}y \quad \text{if } t > 0,$$

and $\qquad T(0)f := f.$

The generator of T is the Laplace operator

$$\Delta f = f'', \quad D(\Delta) = \mathrm{W}^{2,p}(\mathbb{R}).$$

The semigroup T consists of contractions, or equivalently, Δ is dissipative (cf. Example 6.9). If $v \in \mathrm{L}^\infty(\mathbb{R})$, then the multiplication operator $B = M_v$ is bounded on $\mathrm{L}^p(\mathbb{R})$. So by Theorem 6.13 the operator $\Delta + B$ with domain $\mathrm{W}^{2,p}(\mathbb{R})$ generates a semigroup. However, we want to consider not necessarily bounded multiplications operators, say we suppose $v \in \mathrm{L}^q(\mathbb{R})$ for some $q \geq 1$. To establish the estimate $\|BR(\lambda, \Delta)\| < 1$ we first need to make sure that the for $f \in \mathrm{L}^p(\mathbb{R})$ the function $v \cdot R(\lambda, \Delta)f$ belongs to $\mathrm{L}^p(\mathbb{R})$. To show that one may use Hölder's inequality:

$$\|v \cdot R(\lambda, \Delta)f\|_p \leq \|v\|_q \cdot \|R(\lambda, \Delta)f\|_r \tag{6.10}$$

where $\frac{1}{p} = \frac{1}{r} + \frac{1}{q}$, and here we have to suppose $q \geq p$. This shows that we need to estimate the operator norm of

$$R(\lambda, \Delta) : \mathrm{L}^p(\mathbb{R}) \to \mathrm{L}^r(\mathbb{R}).$$

To this end, recall from (the solution of) Exercise 2.9 that

$$\|T(t)f\|_r \leq ct^{-\frac{1}{2}\left(\frac{1}{p}-\frac{1}{r}\right)} \|f\|_p = ct^{-\frac{1}{2q}} \|f\|_p \quad \text{for all } t > 0.$$

By using this estimation and by taking the Laplace transform of $T(t)f$ (see Proposition 2.26.a)) we obtain the following estimate for the resolvent:

$$\|R(\lambda, \Delta)f\|_r \leq c\|f\|_p \int_0^\infty t^{-\frac{1}{2q}} \mathrm{e}^{-\lambda t} \,\mathrm{d}t = c\|f\|_p \Gamma\big(1 - \tfrac{1}{2q}\big) \lambda^{\frac{1}{2q}-1} \tag{6.11}$$

if $\frac{1}{2q} < 1$, i.e., if $q > \frac{1}{2}$. Now we are prepared for the following result:

Proposition 6.15. *Consider the Laplace operator Δ with $D(\Delta) = \mathrm{W}^{2,p}(\mathbb{R})$. Let $q \geq p$ and let $v \in \mathrm{L}^q(\mathbb{R})$ be a function with $\mathrm{Re}\, v \leq 0$. Define $B := M_v$ the multiplication operator by v with domain $D(B) = \mathrm{L}^p(\mathbb{R}) \cap \mathrm{L}^r(\mathbb{R})$ (where $\frac{1}{p} = \frac{1}{r} + \frac{1}{q}$). Then $\Delta + B$ with domain $\mathrm{W}^{2,p}(\mathbb{R})$ generates a contraction semigroup.*

Proof. We check the conditions of Theorem 6.14. The dissipativity of B follows from the assumption on the range of v. The condition $\|BR(\lambda, \Delta)\| < 1$ for λ large follows from inequalities (6.10) and (6.11) and from the assumption that $q \geq p > \frac{1}{2}$. □

6.4 Supplement

The Hahn–Banach Theorem

Let X be a Banach space. A linear functional $\phi : X \to \mathbb{C}$ is called **bounded** if there is a constant such that

$$\|\phi(f)\| \leq M\|f\| \quad \text{for all } f \in X.$$

The set

$$X' := \big\{\phi : \phi \text{ is a bounded linear functional on } X\big\}$$

of all bounded linear functionals is a linear space, and becomes a Banach space with the **functional norm**

$$\|\phi\| := \sup_{\substack{f\in X\\ \|f\|\leq 1}} |\phi(x)| = \sup_{\substack{f\in X\\ \|f\|\leq 1}} |\langle f, \phi\rangle|.$$

Here we used the convenient notation $\phi(f) = \langle f, \phi\rangle$. If $\phi \in X'$ then

$$|\langle f, \phi\rangle| \leq \|\phi\| \cdot \|f\|$$

holds for all $f \in X$. The space X' is called the **dual space** of X. That X' is large enough for every Banach space is highly non-trivial, and is actually the statement of the Hahn–Banach[1] theorem. Note however that in specific examples the dual space can be determined.

Theorem 6.16 (Hahn–Banach)**.** *Let X be a Banach space, and let X' be its dual space. Then the following assertions are true:*

a) For $f \in X$, $f \neq 0$ there is $\phi \in X'$ with $\phi(f) = \|f\|$ and $\|\phi\| = 1$. Or, which is the same, for every $0 \neq f \in X$ there is $\phi \in X'$ with $\phi(f) = \|f\|^2 = \|\phi\|^2$.

b) For $f, g \in X$ one has $f = g$ if and only if $\langle f, \phi\rangle = \langle g, \phi\rangle$ for all $\phi \in X'$.

c) A subspace Y is dense in X if and only the zero functional is the only bounded linear functional that vanishes on Y.

The Banach–Alaoglu Theorem

Let $\phi_n, \phi \in X'$. We call ϕ_n **weak*-convergent** to ϕ if for all $f \in X$

$$\langle f, \phi_n - \phi\rangle \to 0 \quad \text{holds as } n \to \infty.$$

The functional ϕ is called the **weak*-limit** of the sequence, and if exists, then it is obviously unique. We call ϕ a **weak*-accumulation** point of the sequence (ϕ_n) if for all $f \in X$ and $\varepsilon > 0$ there is a subsequence (ϕ_{n_k}) with

$$|\langle f, \phi_{n_k} - \phi\rangle| \leq \varepsilon \qquad \text{for all } k \in \mathbb{N}.$$

Obviously, if (ϕ_n) has a weak*-convergent subsequence ϕ, then ϕ is an accumulation point of the sequence. The converse implication is in general not true. The next rather weak formulation of a central result from functional analysis suffices for our purposes.

[1] H. Hahn: Über lineare Gleichungssysteme in linearen Räumen. Journal für die reine und angewandte Mathematik **157** (1927), 214-229. and S. Banach: Sur les fonctionelles linéaires. In: Studia Mathematica **1** (1929), 211-216.

Theorem 6.17 (Banach–Alaoglu[2])**.** *Let X be a Banach space and consider its dual space. Let*

$$\mathrm{B}' := \big\{\phi \in X' : \|\phi\| \le 1\big\} \subseteq X'$$

be the unit ball in X'. Then every sequence $(\phi_n) \subseteq \mathrm{B}'$ has a weak-accumulation point in B'. If X is reflexive or separable, then every sequence $(\phi_n) \subseteq \mathrm{B}'$ has a weak*-convergent subsequence with limit in B'.*

6.5 Exercises

1. Let $\Omega = (0,\pi) \times (0,\pi)$ and define on $\mathrm{L}^2(\Omega)$ the operator A as

$$Af = \Delta f, \quad D(A) := \big\{f \in \mathrm{C}^2(\Omega) \,:\, \text{the support of } f \text{ is compact}\big\}.$$

Show that A is dissipative and its closure generates a contraction semigroup.

2. Let $X = \mathrm{C}[-1,0]$ and $0 < \tau_1 < \tau_2 < \ldots < \tau_n = 1$. Consider the operator $Af := f'$ with

$$D(A) := \left\{ f \in \mathrm{C}^1[-1,0] \,:\, f'(0) = \sum_{i=1}^{n} c_i f(-\tau_i) \right\},$$

where $c_i \in \mathbb{C}$, $i = 1, \ldots, n$. This operator plays an important role in the theory of delay differential equations. Show that A is quasi-dissipative.

3. Give a necessary and sufficient condition on $m : \Omega \to \mathbb{C}$ such that the multiplication operator M_m is dissipative (with maximal domain) in $L^p(\Omega)$.

4. Suppose that A generates a contraction semigroup and $B : D(B) \to X$ satisfies $D(A) \subseteq D(B)$ and has the following property: There is $a \in [0, \frac{1}{2})$ and $b > 0$ such that

$$\|Bx\| \le a\|Ax\| + b\|x\| \quad \text{for all } x \in D(A).$$

Prove that for large $\lambda > 0$ one has $\|BR(\lambda, A)\| < 1$.

5. Let $X = \mathrm{C}_0(\mathbb{R})$ and $Af = f'' + f'$ with $D(A) = \big\{f \in \mathrm{C}^2(\mathbb{R}) \cap X \,:\, f'' + f' \in X\big\}$. Show that it generates a contraction semigroup.

[2] L. Alaoglu, Weak topologies of normed linear spaces. Ann. Math. **41** (1940), 252–267.

Lecture 7

Complex Powers of Closed Operators

Having finished with the theoretical questions on well-posedness of evolution equations, we now turn our attention on technical matters which will be extremely important in proving convergence rates for various discretisation procedures. Note that in Lectures 3 and 4 the existence of Banach space Y invariant under some given semigroup T was of enormous importance. Our aim in this lecture is to arm you with important examples for such spaces. As a warm up, let us first summarise the results of Exercise 4.1.

Proposition 7.1. *Let A be the generator of a semigroup of type (M,ω) in the Banach space X, and consider the space $X_n = D(A^n)$ with the graph norm which we denote by $\|\cdot\|_{A^n}$.*

a) For $n \in \mathbb{N}$ and $f \in D(A^n)$ define $|||f|||_n := \|f\| + \|Af\| + \cdots + \|A^n f\|$. Then $|||\cdot|||_n$ and $\|\cdot\|_{A^n}$ are equivalent norms.

b) The spaces X_n are Banach spaces and are invariant under the semigroup T. If we set $T_n(t) := T(t)|_{X_n}$, then T_n is a semigroup of type (M,ω) on X_n.

For applications these spaces are quite often too small and some intermediate spaces are needed. The purpose of this lecture is to find some possible candidates for such invariant subspaces that fit well in the scale of $D(A^n)$, $n = 1, 2, \ldots$.

To motivate this a bit further, let us consider the next example:

Example 7.2. Recall from Lecture 1 the multiplication operator M on ℓ^2 by the sequence $-n^2$, which corresponds to the Dirichlet Laplacian on $[0,\pi]$ after *diagonalisation* (more precisely after applying the spectral theorem for selfadjoint operators)

$$D(M) = \{(x_n) \in \ell^2 : (n^2 x_n) \in \ell^2\} \quad \text{and} \quad M(x_n) = (-n^2 x_n).$$

For $\alpha \geq 0$ define

$$D((-M)^\alpha) = \{(x_n) \in \ell^2 : (n^{2\alpha} x_n) \in \ell^2\} \quad \text{and} \quad (-M)^\alpha(x_n) = (n^{2\alpha} x_n).$$

(The minus sign here is only a matter of convention.) It is not hard to see that $(-M)^\alpha$ is a closed operator, hence $D((-M)^\alpha)$ is a Banach space with the graph norm. Equally easy is to see that $(-M)^k$ is indeed the k^{th} power of $(-M)$ for $k \in \mathbb{N}$, and that the semigroup T defined by

$$T(t)(x_n) = (\mathrm{e}^{-n^2 t} x_n) \in \ell^2$$

leaves this space invariant (much more(!) is true). Hence the spaces $D((-M)^\alpha)$ fulfill the requirements formulated above.

Thus we set out for the quest for fractional powers of closed operators. For the purposes of this lecture we shall leave semigroups (almost completely) behind, and develop some beautiful operator theoretic notions.

7.1 Complex powers with negative real part

We want to define complex powers of operators A, i.e., we want to plug in A into the function $F(x) = x^z$ where $z \in \mathbb{C}$ is fixed. This means that we want to develop a *functional calculus* for this particular function F and for some reasonable class of operators. To be able to do that we shall need the complex power functions defined on the complex plane. Let $\log : \mathbb{C} \setminus (-\infty, 0] \to \mathbb{C}$ be the principal branch of the logarithm, i.e., $\log(\lambda) = \log|\lambda| + \mathrm{i}\arg(\lambda)$, where we have fixed the function arg with values in $(-\pi, \pi)$. Since log is holomorphic, we can define the holomorphic function $\lambda \mapsto \lambda^z = \mathrm{e}^{z\log(\lambda)}$ on $\mathbb{C} \setminus (-\infty, 0]$ for any given $z \in \mathbb{C}$. Now the basic idea comes from Cauchy's integral theorem for this particular situation:

$$a^z = \oint \frac{\lambda^z}{\lambda - a}\,\mathrm{d}\lambda$$

where we integrate along a curve that passes around $a \notin (-\infty, 0]$ in the positive direction and avoids the negative real axis. Therefore, by analogy, or motivated by multiplication operators (cf. Exercise 2) we have to give meaning to expressions like

$$\oint \lambda^z R(\lambda, A)\,\mathrm{d}\lambda.$$

Of course the curve that we are integrating over has to lie in the resolvent set of A and pass around the spectrum of $\sigma(A)$ in the positive direction. Two difficulties arise here immediately: the spectrum may be unbounded, hence the integration curve has to be unbounded (and anyway the term "passing around" does not make sense any more), and convergence issues for the integral have to be taken care of. This section includes a fair amount of technicalities, but the single idea has been explained above. The next assumption tackles both mentioned difficulties as we shall shortly see.

Assumption 7.3. Suppose for $A : D(A) \to X$ one has $(-\infty, 0] \subseteq \rho(A)$ and

$$\|R(\lambda, A)\| \le \frac{M}{1 + |\lambda|} \quad \text{for all } \lambda \le 0 \text{ and some } M \ge 0.$$

All operators[1] A occurring in this section will be assumed to have the property above. The next is an important example for such operators, leading back for a moment to semigroups.

Example 7.4. If A generates a strongly continuous semigroup of type (M', ω) with $\omega < 0$, then, as consequence of (2.2) in Proposition 2.26, we see that for $\lambda > 0$

$$\|R(\lambda, A)\| \le \frac{M}{\lambda - \omega} \le \frac{M'}{\lambda + 1}.$$

Hence $-A$ satisfies the above estimate in Assumption 7.3 for some M'.

The next fundamental result shows that although only $(-\infty, 0] \subseteq \rho(A)$ was assumed, one gains a sector around the negative real axis, where the resolvent can be estimated satisfactorily well.

[1] Some authors use the names *sectorial operator* or *positive operator* for objects having this property. We decided not to give them a name.

Proposition 7.5. *Suppose A is as in Assumption 7.3. Then there is $\theta_0 \in (\frac{\pi}{2}, \pi)$ and $r_0 > 0$ such that the set*

$$\Lambda := \big\{z \in \mathbb{C} : |\arg(z)| \in (\theta_0, \pi]\big\} \cup \big\{z \in \mathbb{C} : |z| \leq r_0\big\} \subseteq \rho(A)$$

belongs to the resolvent set of A. Moreover, there is $M_0 \geq 0$ so that for every $\lambda \in \Lambda$ one has

$$\|R(\lambda, A)\| \leq \frac{M_0}{1 + |\lambda|}. \tag{7.1}$$

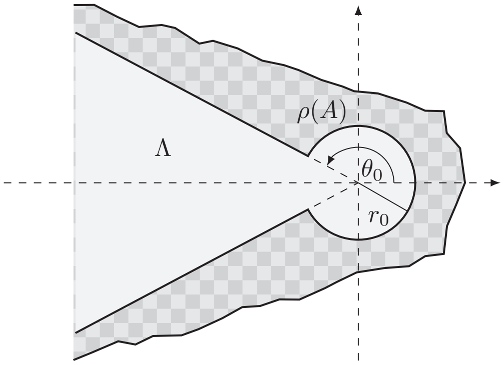

Figure 7.1: The resolvent set of A and the set Λ

Proof. First of all note that for some $r_0 > 0$ the closed ball $\overline{\mathrm{B}}(0, r_0)$ is contained in $\rho(A)$, since $\rho(A)$ is open. So on this ball the resolvent is bounded. On the other hand, we have $\mu \in \rho(A)$ and

$$R(\mu, A) = \sum_{k=0}^{\infty} (\lambda - \mu)^k R(\lambda, A)^{k+1}$$

whenever $|\mu - \lambda| < \|R(\lambda, A)\|^{-1}$, i.e., the open ball $\mathrm{B}(\lambda, -\frac{\lambda}{M})$ is contained in $\rho(A)$. From this the first assertion follows for $\theta_0 = \pi - \arctan(\frac{1}{M})$. If $|\arg \lambda| \in (\theta_0, \pi]$, then

$$\begin{aligned}\|R(\mu, A)\| &\leq \sum_{k=0}^{\infty} |\operatorname{Re}\mu - \mu|^k \frac{M^k}{(1 + |\operatorname{Re}\mu|)^{k+1}} = \sum_{k=0}^{\infty} |\operatorname{Im}\mu|^k \frac{M^k}{(1 + |\operatorname{Re}\mu|)^{k+1}} \\ &\leq \frac{M_1}{1 + |\operatorname{Re}\mu|} \leq \frac{M_0}{1 + |\mu|}.\end{aligned}$$ □

Remark 7.6. The next two estimates will be crucial for proving convergence of some integrals and for estimating them:

1. By the proposition above we have

$$\|R(\lambda, A)\| \leq \frac{M_0}{|\lambda|} \qquad \text{for all } \lambda \in \Lambda,\ |\lambda| > r_0 > 0,$$

and

$$\|R(\lambda, A)\| \leq M_1 \qquad \text{for all } \lambda \in \Lambda,\ |\lambda| \leq r_0.$$

2. For $\lambda \in \mathbb{C} \setminus (-\infty, 0]$ we have
$$\left|\lambda^z\right| = |\lambda|^{\operatorname{Re} z} \mathrm{e}^{-\operatorname{Im} z \cdot \arg(\lambda)} \leq |\lambda|^{\operatorname{Re} z} \mathrm{e}^{\pi |\operatorname{Im} z|} = M_2 |\lambda|^{\operatorname{Re} z}$$
for every fixed z. In particular for $\operatorname{Re} z < 0$, we have a decay as $|\lambda| \to \infty$.

We shall often use these estimates without further mentioning. Next we turn our attention to integration paths. To abbreviate a little we shall call a piecewise continuously differentiable path **admissible** if it belongs to Λ and goes from $\infty \mathrm{e}^{\mathrm{i}\theta}$ to $\infty \mathrm{e}^{-\mathrm{i}\theta}$ for some $\theta \in (\theta_0, \pi)$. Important examples for admissible curves are given by the following parametrisations:

Example 7.7. let $\theta \in (\theta_0, \pi)$ and let $\gamma_1(s) = s\mathrm{e}^{\mathrm{i}\theta} + a$ and let $\gamma_2(s) = s\mathrm{e}^{-\mathrm{i}\theta} + a$, $s \in [0, \infty)$. For $a > 0$ sufficiently small the curve $\gamma = -\gamma_1 + \gamma_2$ is admissible.

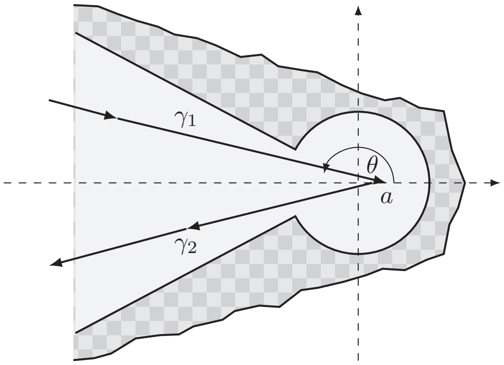

Figure 7.2: An admissible curve γ

Here is the first result giving meaning to the expression we sought for.

Lemma 7.8. *For γ an admissible curve and $z \in \mathbb{C}$ with $\operatorname{Re} z < 0$ the complex path integral*
$$\frac{1}{2\pi\mathrm{i}} \int_\gamma \lambda^z R(\lambda, A) \, \mathrm{d}\lambda \in \mathscr{L}(X)$$
converges in operator norm locally uniformly in $\{z : \operatorname{Re} z < 0\}$, and is independent of γ.

Proof. The integrand is holomorphic, and since
$$\|\lambda^z R(\lambda, A)\| \leq \frac{M|\lambda|^{\operatorname{Re} z} \mathrm{e}^{\pi |\operatorname{Im} z|}}{1 + |\lambda|} \tag{7.2}$$
holds, it follows that the integral is absolutely and locally uniformly convergent.

The independence of the integral from γ follows from Cauchy's integral theorem and from the estimate above. □

The next result shows that our new definition for the power would be consistent with the usual one.

Proposition 7.9. *For $n \in \mathbb{N}$ and $z = -n$ we have*
$$A^z = A^{-n} = \frac{1}{2\pi\mathrm{i}} \int_\gamma \lambda^{-n} R(\lambda, A) \, \mathrm{d}\lambda.$$

Proof. We may assume that γ is an admissible curve of the form given in Example 7.7. Let us consider the part of γ inside of $\mathrm{B}(0,r)$ that we close on the left by a circle arc around 0 of radius r. Hence we obtain the closed curve γ_r. The residue theorem applied to

$$\lambda^{-n}R(\lambda,A)=\sum_{k=0}^{\infty}(-1)^k\lambda^{k-n}(-A)^{-k+1}=-\sum_{k=0}^{\infty}\lambda^{k-n}A^{-k+1}$$

yields

$$\frac{1}{2\pi\mathrm{i}}\int\limits_{\gamma_r}\lambda^{-n}R(\lambda,A)\,\mathrm{d}\lambda=A^{-n},$$

since γ_r is negatively oriented. If we let $r\to\infty$ we obtain the assertion by the estimate in (7.2). □

Now we can create a definition out of what we have seen.

Definition 7.10. For $z\in\mathbb{C}$ with $\operatorname{Re}z<0$ define the operator

$$A^z:=\frac{1}{2\pi\mathrm{i}}\int\limits_{\gamma}\lambda^zR(\lambda,A)\,\mathrm{d}\lambda. \tag{7.3}$$

We call A^z the **power** of A.

As one might have expected we have the following algebraic property.

Proposition 7.11. *For $z,w\in\mathbb{C}$ with $\operatorname{Re}z,\operatorname{Re}w<0$ we have*

$$A^zA^w=A^{z+w}.$$

Proof. Take two admissible curves γ and $\tilde\gamma$ such that γ lies to the left of $\tilde\gamma$. Then we have

$$A^z=\frac{1}{2\pi\mathrm{i}}\int\limits_{\gamma}\lambda^zR(\lambda,A)\,\mathrm{d}\lambda\quad\text{and}\quad A^w=\frac{1}{2\pi\mathrm{i}}\int\limits_{\tilde\gamma}\mu^wR(\mu,A)\,\mathrm{d}\mu.$$

We calculate the product of the two powers

$$A^zA^w=\frac{1}{(2\pi\mathrm{i})^2}\int\limits_{\tilde\gamma}\int\limits_{\gamma}\lambda^z\mu^wR(\mu,A)R(\lambda,A)\,\mathrm{d}\lambda\,\mathrm{d}\mu=\frac{1}{(2\pi\mathrm{i})^2}\int\limits_{\tilde\gamma}\int\limits_{\gamma}\frac{\lambda^z\mu^w}{\lambda-\mu}\big(R(\mu,A)-R(\lambda,A)\big)\,\mathrm{d}\lambda\,\mathrm{d}\mu$$

by the resolvent identity. We can continue by Fubini's theorem

$$\begin{aligned}&=\frac{1}{2\pi\mathrm{i}}\int\limits_{\tilde\gamma}R(\mu,A)\frac{1}{2\pi\mathrm{i}}\int\limits_{\gamma}\frac{\lambda^z\mu^w}{\lambda-\mu}\,\mathrm{d}\lambda\,\mathrm{d}\mu-\frac{1}{2\pi\mathrm{i}}\int\limits_{\gamma}R(\lambda,A)\frac{1}{2\pi\mathrm{i}}\int\limits_{\tilde\gamma}\frac{\lambda^z\mu^w}{\lambda-\mu}\,\mathrm{d}\mu\,\mathrm{d}\lambda\\&=\frac{1}{2\pi\mathrm{i}}\int\limits_{\tilde\gamma}\mu^z\mu^wR(\mu,A)\,\mathrm{d}\mu-0=A^{w+z},\end{aligned}$$

where we also used Cauchy's integral theorem. □

Before turning to the full definition including positive fractional powers we study the properties of A^z as a function of $z\in\{w:\operatorname{Re}w<0\}$.

Proposition 7.12. *The mapping*

$$\{w : \operatorname{Re} w < 0\} \ni z \mapsto A^z \in \mathscr{L}(X)$$

is holomorphic.

Proof. Since the integrand in (7.3) is holomorphic and since the integral is locally uniformly convergent in $\{w : \operatorname{Re} w < 0\}$ (see Lemma 7.8), the assertion follows immediately. □

The next result provides important formulas in which the path integral is replaced by integration on the real line. To ensure convergence at 0 we need to have a condition on the exponent.

Proposition 7.13. *For $z \in \mathbb{C}$ with $-1 < \operatorname{Re} z < 0$ we have*

$$A^z = \frac{\sin(\pi z)}{\pi} \int_0^\infty s^z R(-s, A)\, \mathrm{d}s = -\frac{\sin(\pi z)}{\pi} \int_0^\infty s^z (s + A)^{-1}\, \mathrm{d}s. \tag{7.4}$$

Proof. Choose the admissible curve γ as in Example 7.7. Then

$$\begin{aligned}
A^z &= \frac{1}{2\pi \mathrm{i}} \int_\gamma \lambda^z R(\lambda, A)\, \mathrm{d}\lambda \\
&= -\frac{1}{2\pi \mathrm{i}} \int_0^\infty (s\mathrm{e}^{\mathrm{i}\theta} + a)^z R(s\mathrm{e}^{\mathrm{i}\theta} + a, A) \mathrm{e}^{\mathrm{i}\theta}\, \mathrm{d}s + \frac{1}{2\pi \mathrm{i}} \int_0^\infty (s\mathrm{e}^{-\mathrm{i}\theta} + a)^z R(s\mathrm{e}^{-\mathrm{i}\theta} + a, A) \mathrm{e}^{-\mathrm{i}\theta}\, \mathrm{d}s \\
&= -\frac{1}{2\pi \mathrm{i}} \int_0^\infty \mathrm{e}^{\mathrm{i}\theta(z+1)} (s + a\mathrm{e}^{-\mathrm{i}\theta})^z R(s\mathrm{e}^{\mathrm{i}\theta} + a, A)\, \mathrm{d}s + \frac{1}{2\pi \mathrm{i}} \int_0^\infty \mathrm{e}^{-\mathrm{i}\theta(z+1)} (s + \mathrm{e}^{\mathrm{i}\theta} a)^z R(s\mathrm{e}^{-\mathrm{i}\theta} + a, A)\, \mathrm{d}s.
\end{aligned}$$

If we let $a \to 0$ and $\theta \nearrow \pi$, then we obtain

$$= -\frac{1}{2\pi \mathrm{i}} \int_0^\infty \mathrm{e}^{\mathrm{i}\pi(z+1)} s^z R(-s, A)\, \mathrm{d}s + \frac{1}{2\pi \mathrm{i}} \int_0^\infty \mathrm{e}^{-\mathrm{i}\pi(z+1)} s^z R(-s, A)\, \mathrm{d}s = \frac{\sin(\pi z)}{\pi} \int_0^\infty s^z R(-s, A)\, \mathrm{d}s.$$

This passage to the limit is allowed, since we can estimate the integrand

$$\|\mathrm{e}^{\mathrm{i}\theta(z+1)} (s + a\mathrm{e}^{-\mathrm{i}\theta})^z R(s\mathrm{e}^{\mathrm{i}\theta} + a, A)\| \leq K \frac{s^{\operatorname{Re} z}}{1 + s},$$

which is integrable near $s = 0$ since $\operatorname{Re} z > -1$ and is integrable near ∞ since $\operatorname{Re} z < 0$. Hence we can apply Lebesgue's dominated convergence theorem. □

A trivial consequence of this theorem is the identity

$$a^\alpha = -\frac{\sin(\pi a)}{\pi} \int_0^\infty \frac{s^\alpha}{s + a}\, \mathrm{d}s \tag{7.5}$$

for $a \in (-1, 0)$. This is one of the scalar identities motivating the formulas behind fractional powers of operators.

Proposition 7.14. *For $\alpha \in \mathbb{C}$ with $\operatorname{Re}\alpha \in (-1,0)$ we have*

$$\|A^\alpha\| \le M\frac{|\sin(\pi\alpha)|}{\sin(\pi \operatorname{Re}\alpha)}.$$

In particular, the mapping

$$(-1,0) \ni \alpha \mapsto A^\alpha \in \mathscr{L}(X)$$

is uniformly bounded.

Proof. We use the representation (7.4) from Proposition 7.13. For $\alpha \in (-1,0)$ we have

$$\|A^\alpha\| = \Big\|\frac{\sin(\pi\alpha)}{\pi}\int_0^\infty s^\alpha R(-s,A)\,\mathrm{d}s\Big\| \le \frac{|\sin(\pi\alpha)|}{\pi}\int_0^\infty s^{\operatorname{Re}\alpha}\frac{M}{1+s}\,\mathrm{d}s = M\frac{|\sin(\pi\alpha)|}{\sin(\pi\operatorname{Re}\alpha)},$$

by identity (7.5). For real α the assertion follows from this trivially. □

Corollary 7.15. *If A is densely defined, then $T(t) := A^{-t}$, $t > 0$ and $T(0) = I$ defines a strongly continuous semigroup.*

Proof. The mapping T has the semigroup property by Proposition 7.11. Since $T : [0,1] \to \mathscr{L}(X)$ is bounded by Proposition 7.14, it suffices to check the strong continuity at 0 on the dense subspace $D(A)$ (see Proposition 2.5). For $t \in (0,1)$ and $f \in D(A)$ we have by Proposition 7.13 and by identity (7.5) for $a = 1$ that

$$A^{-t}f - f = \frac{\sin(\pi t)}{\pi}\int_0^\infty s^{-t}\big(R(-s,A) - R(-s,I)\big)\,\mathrm{d}s = \frac{\sin(\pi t)}{\pi}\int_0^\infty \frac{s^{-t}}{1+s}R(-s,A)(I-A)f\,\mathrm{d}s.$$

From this it follows

$$\|A^{-t}f - f\| \le M\frac{\sin(\pi t)}{\pi}\int_0^\infty \frac{s^{-t}}{1+s}\|R(-s,A)\|\cdot\|(I-A)f\|\,\mathrm{d}s \le \frac{\sin(\pi t)}{\pi}\int_0^\infty \frac{s^{-t}}{(1+s)^2}\,\mathrm{d}s\|(I-A)f\|,$$

which converges to 0 as $t \searrow 0$. □

7.2 Complex powers

We expect that $A^z = (A^{-z})^{-1}$ should hold, so A^z should be injective. The first result tells that this intuition—unlike many others concerning complex powers—is true.

Proposition 7.16. *For $z \in \mathbb{C}$ with $\operatorname{Re} z < 0$ the operator A^z is injective.*

Proof. Let $n \in \mathbb{N}$ be such that $-n < \operatorname{Re} z$ and take $w := -n - z$. Then we have

$$A^z A^w = A^w A^z = A^{z+w} = A^{-n}.$$

By Proposition 7.9, the operator A^{-n} is the n^{th} power of the inverse A^{-1} of A so it is injective, hence so are A^z and A^w. □

The result above allows us to formulate the next definition.

Definition 7.17. Let $z \in \mathbb{C}$. If $\operatorname{Re} z < 0$, then the operator A^z is defined in (7.3). If $\operatorname{Re} z > 0$, then we set

$$D(A^z) := \operatorname{ran}(A^{-z}) \quad \text{and} \quad A^z := (A^{-z})^{-1},$$

which exists by Proposition 7.16. If $\operatorname{Re} z = 0$, then we define

$$D(A^z) := \bigl\{f \in X : A^{z-1} f \in D(A)\bigr\} \quad \text{and} \quad A^z f := AA^{z-1} f.$$

In particular, we set $A^0 = I$. The operator A^z is called the **complex power** of A.

First, we study algebraic properties of the complex powers A^z.

Proposition 7.18. *a) For $z \in \mathbb{C}$ with $\operatorname{Re} z < -n$, $n \in \mathbb{N}$ we have that*

$$\operatorname{ran}(A^z) \subseteq D(A^n) \quad \textit{and} \quad A^n A^z f = A^{n+z} f \quad \textit{for all } f \in X.$$

b) For $z \in \mathbb{C}$ with $\operatorname{Re} z < 0$, $f \in D(A^n)$, $n \in \mathbb{N}$ we have

$$A^z f \in D(A^n) \quad \textit{and} \quad A^z A^n f = A^n A^z f.$$

c) For $z \in \mathbb{C}$ with $0 \leq \operatorname{Re} z < n$ we have

$$D(A^z) = \bigl\{f \in X : A^{z-n} f \in D(A^n)\bigr\} \quad \textit{and} \quad A^z f = A^n A^{z-n} f.$$

Proof. a) We first prove the assertion for $n = 1$ and assume $\operatorname{Re}(z) < -1$. Let γ be an admissible curve. Since

$$\|\lambda^z AR(\lambda, A)\| = \|\lambda^z(\lambda R(\lambda, A) - I)\| \leq (M_0 + 1)|\lambda|^{\operatorname{Re} z} \mathrm{e}^{\pi |\operatorname{Im} z|},$$

and since $\lambda^z AR(\lambda, A)$ is bounded on compact parts of γ, we see that the integral A^z converges in the norm of $\mathscr{L}(X, D(A))$. Since A is closed we obtain

$$AA^z = \frac{1}{2\pi \mathrm{i}} \int\limits_\gamma \lambda^{z+1} R(\lambda, A) \, \mathrm{d}\lambda - \frac{1}{2\pi \mathrm{i}} \int\limits_\gamma \lambda^z \, \mathrm{d}\lambda.$$

By closing γ on the right by large circle arc of radius $r > 0$ and by letting $r \to \infty$, we see that integral on the right hand side is 0 by Cauchy's integral theorem. Hence the assertion follows.

For general $n \in \mathbb{N}$ we can argue inductively. Indeed, let $n \in \mathbb{N}$, $n \geq 2$, and let $z \in \mathbb{C}$ be with $\operatorname{Re} z < -n$. Then $\operatorname{Re}(z + 1) < -(n - 1)$, hence $\operatorname{ran}(A^{z+1}) \subseteq D(A^{n-1})$ and $A^{n-1} A^{z+1} f = A^{n+z} f$ follows for $f \in X$ by the induction hypothesis. We already proved $A^{z+1} = AA^z$. From these the assertion follows.

b) Since $R(\lambda, A)$ and A^n commute on $D(A^n)$, it follows that A^z and A^n commute on $D(A^n)$. This implies the assertion.

c) If $\operatorname{Re} z > 0$, then $D(A^z) = \operatorname{ran}(A^{-z})$. By Proposition 7.11 we have $A^{-n} = A^{-z} A^{z-n}$. Hence $f \in \operatorname{ran}(A^{-z})$ if and only if $A^{z-n} f \in \operatorname{ran}(A^{-n}) = D(A^n)$ and the asserted equality follows. Suppose $\operatorname{Re} z = 0$, and notice that the assertion is true for $n = 1$ by the definition of A^z. For $n \in \mathbb{N}$, $n \geq 2$ we have

$$A^{z-n} = A^{1-n} A^{z-1}.$$

This implies that $A^{z-n} f \in D(A^n)$ if and only if $A^{z-1} \in D(A)$. □

The next result is the extension of the "semigroup property" from Proposition 7.11.

Proposition 7.19. *For $z, w \in \mathbb{C}$ with $\operatorname{Re} z < \operatorname{Re} w$ the following assertions are true:*

a) One has $D(A^w) \subseteq D(A^z)$ and $A^z f = A^{z-w} A^w f$ for all $f \in D(A^w)$

b) For every $f \in D(A^w)$ we have $A^z f \in D(A^{w-z})$ and $A^w f = A^{w-z} A^z f$.

c) If $f \in D(A^z)$ and $A^z f \in D(A^{w-z})$, then $f \in D(A^w)$.

Proof. a) Let $n \in \mathbb{N}$ satisfy $n > \operatorname{Re} w$, and let $f \in D(A^w)$, then by Proposition 7.18.c) we have $A^{w-n} f \in D(A^n)$. Proposition 7.11 yields $A^{-n+z} f = A^{z-w} A^{-n+w} f \in D(A^n)$, so actually again by Proposition 7.18.c) we conclude $x \in D(A^z)$.

b) Let $f \in D(A^w)$ and let $n \in \mathbb{N}$ satisfy $n > \operatorname{Re} w$ and $n > \operatorname{Re} w - \operatorname{Re} z$. Then we can write

$$A^{-n+w-z} A^w f = A^{-n+w-z} A^{z-w} A^w f = A^{-n} A^w f,$$

hence by Proposition 7.18.c) we obtain $A^z \in D(A^{w-z})$ and $A^{w-z} A^z f = A^w f$.

c) Take $f \in D(A^z)$ such that $A^z f \in D(A^{w-z})$. Let $n \in \mathbb{N}$ satisfy $n > \operatorname{Re} w$ and $n > \operatorname{Re} w - \operatorname{Re} z$. Proposition 7.11 yields

$$A^{w-2n} f = A^{w-n-z} A^{z-n} f = A^{w-n-z} A^{-n} A^z f = A^{-n} A^{w-z-n} A^z f.$$

By Proposition 7.18.c) the right-hand side belongs to $D(A^{2n})$, so again this proposition gives $f \in D(A^w)$. The equality

$$A^w f = A^{2n} A^{w-2n} f = A^{2n} A^{-n} A^{w-z-n} A^z f = A^{w-z} A^z f$$

also follows. □

7.3 Domain embeddings

As mentioned in the introduction, our main interest in powers of operators lies in the excellent properties of their domains. Hence, we turn to study various norms on $D(A^z)$ for $\operatorname{Re} z > 0$.

Proposition 7.20. *a) For $z \in \mathbb{C}$ the operator A^z is closed.*

b) For $\operatorname{Re} z > 0$ the graph norm of A^z is equivalent to

$$\|f\|_{A^z} := \|A^z f\| \quad \textit{for all } f \in D(A^z).$$

c) For $z, w \in \mathbb{C}$ with $0 \leq \operatorname{Re} z < \operatorname{Re} w$ the embedding

$$D(A^w) \hookrightarrow D(A^z)$$

is continuous.

Proof. a) If $\operatorname{Re}(z) \neq 0$, either A^z or A^{-z} is bounded, hence both of them are closed. If $\operatorname{Re}(z) = 0$, then $A^z = A A^{z-1}$, where A^{z-1} is bounded. By Exercise 1 the product is closed.

b) Since A^z has bounded inverse A^{-z}, we have $\|f\| \leq \|A^{-z}\| \cdot \|A^z f\|$. From this it follows that the graph norm is equivalent to $\| \cdot \|_{A^z}$.

c) By Proposition 7.19.a) we have $D(A^w) \subseteq D(A^z)$ and

$$A^{z-w} A^w f = A^z f \quad \text{for all } f \in D(A^w),$$

hence $\|A^z\| \leq \|A^{z-w}\| \cdot \|A^w\|$. □

To be able to relate the various norms $\|\cdot\|_{A^\alpha}$ more precisely, we need the next alternative formula for complex powers.

Proposition 7.21 (Balakrishnan's formula)**.** *For $\alpha \in \mathbb{C}$ with $0 < \operatorname{Re}\alpha < 1$ we have*

$$A^\alpha f = \frac{\sin(\pi\alpha)}{\pi} A \int_0^\infty s^{\alpha-1}(s+A)^{-1} f \, \mathrm{d}s \quad = \frac{\sin(\pi\alpha)}{\pi} A \int_0^\infty s^{\alpha-1}(s+A)^{-1} f \, \mathrm{d}s \quad \textit{for all } f \in D(A).$$

Proof. Since $-1 < \operatorname{Re}\alpha - 1 < 0$ we obtain from (7.4) in Proposition 7.13

$$A^{\alpha-1} f = \frac{\sin(\pi\alpha)}{\pi} \int_0^\infty s^{\alpha-1}(s+A)^{-1} f \, \mathrm{d}s. \tag{7.6}$$

Since $s^{\alpha-1}(s+A)^{-1} f \in D(A)$ for every $s > 0$ and since

$$\int_0^\infty s^{\alpha-1}(s+A)^{-1} A f \, \mathrm{d}s$$

is a convergent improper integral, the closedness of A implies that the right-hand side in (7.6) belongs to $D(A)$ and that

$$A A^{\alpha-1} f = \frac{\sin(\pi\alpha)}{\pi} A \int_0^\infty s^{\alpha-1}(s+A)^{-1} f \, \mathrm{d}s = \frac{\sin(\pi\alpha)}{\pi} \int_0^\infty s^{\alpha-1}(s+A)^{-1} A f \, \mathrm{d}s.$$

By Proposition 7.19.a) we have $A^\alpha f = A A^{\alpha-1} f$, hence the statement is proved. □

Remark 7.22. The above proof can be modified to yield the following more general statement: For $\alpha, \beta \in \mathbb{C}$ with $0 < \operatorname{Re}\alpha < \operatorname{Re}\beta \leq 1$ we have

$$A^\alpha f = \tfrac{\sin(\pi(\beta-\alpha))}{\pi} \int_0^\infty s^{\alpha-\beta}(s+A)^{-1} A^\beta f \, \mathrm{d}s = \tfrac{\sin(\pi(\beta-\alpha))}{\pi} \int_0^\infty s^{\alpha-\beta}(s+A)^{-1} A^\beta f \tag{7.7}$$

for all $f \in D(A^\beta)$

We can make use of this representation to obtain finer relations between the $\|\cdot\|_{A^\alpha}$ norms.

Proposition 7.23. *For $\alpha, \beta \in \mathbb{C}$ with $0 < \operatorname{Re}\alpha < \operatorname{Re}\beta < 1$ there is $K_0 \geq 0$ such that the following assertions holds:*

a) For all $f \in D(A^\beta)$

$$\|A^\alpha f\| \leq K_0 \big(t^{\operatorname{Re}\alpha - \operatorname{Re}\beta + 1} \|f\| + t^{\operatorname{Re}\alpha - \operatorname{Re}\beta} \|A^\beta f\| \big) \quad \textit{for all } t > 0. \tag{7.8}$$

b) For all $f \in D(A^\beta)$

$$\|A^\alpha f\| \leq 2K_0 \|f\|^{\operatorname{Re}\beta - \operatorname{Re}\alpha} \cdot \|A^\beta f\|^{1-(\operatorname{Re}\beta - \operatorname{Re}\alpha)}. \tag{7.9}$$

Proof. For $f \in D(A^\beta)$ we have by Remark 7.22 that

$$\begin{aligned}
\|A^\alpha f\| &\le \tfrac{|\sin(\pi(\beta-\alpha))|}{\pi}\Big(\int\limits_0^t \|s^{\alpha-\beta}A^\beta(s+A)^{-1}\|\cdot\|f\|\,\mathrm{d}s + \int\limits_t^\infty \|s^{\alpha-\beta}(s+A)^{-1}\|\cdot\|A^\beta f\|\,\mathrm{d}s\Big)\\
&\le \tfrac{|\sin(\pi(\beta-\alpha))|}{\pi}\Big(\int\limits_0^t \|s^{\alpha-\beta}A^{\beta-1}A(s+A)^{-1}\|\cdot\|f\|\,\mathrm{d}s + \int\limits_t^\infty \|s^{\alpha-\beta}(s+A)^{-1}\|\cdot\|A^\beta f\|\,\mathrm{d}s\Big)\\
&\le \tfrac{|\sin(\pi(\beta-\alpha))|}{\pi}\Big(\|A^{\beta-1}\|\int\limits_0^t s^{\operatorname{Re}\alpha-\operatorname{Re}\beta}\Big(1+\frac{Ms}{1+s}\Big)\,\mathrm{d}s\|f\| + \int\limits_t^\infty s^{\operatorname{Re}\alpha-\operatorname{Re}\beta}\frac{M}{s+1}\,\mathrm{d}s\|A^\beta f\|\Big)\\
&\le \tfrac{|\sin(\pi(\beta-\alpha))|}{\pi}\Big(t^{\operatorname{Re}\alpha-\operatorname{Re}\beta+1}(1+M)\|A^{\beta-1}\|\cdot\|f\| + Mt^{\operatorname{Re}\alpha-\operatorname{Re}\beta}\|A^\beta f\|\Big)\\
&\le K_0\big(t^{\operatorname{Re}\alpha-\operatorname{Re}\beta+1}\|f\| + t^{\operatorname{Re}\alpha-\operatorname{Re}\beta}\|A^\beta f\|\big).
\end{aligned}$$

This proves assertion a).

For $f=0$ the desired inequality (7.9) is trivial. For $f \neq 0$ set $t = \frac{\|A^\beta f\|}{\|f\|}$ in the inequality above to conclude

$$\|A^\alpha f\| \le 2K_0\|f\|^{\operatorname{Re}\beta-\operatorname{Re}\alpha}\|A^\beta f\|^{1-(\operatorname{Re}\beta-\operatorname{Re}\alpha)}. \qquad \square$$

Remark 7.24. The proof above works whenever $A^{\beta-1}$ is bounded, for example also for $\beta = 1$. In particular, we obtain for $\alpha \in [0,1]$

$$\|A^\alpha f\| \le K\|f\|^{1-\alpha}\|Af\|^\alpha \quad \text{for all } f \in D(A), \tag{7.10}$$

the limiting cases $\alpha = 0$ and $\alpha = 1$ being trivial.

With some more work one can prove the following general version of interpolation type inequalities, which we mention here without proof.

Theorem 7.25 (Moment inequality)**.** *For $\alpha < \beta < \gamma$ there is $K \ge 0$ such that*

$$\|A^\beta f\| \le K\|A^\alpha f\|^{\frac{\gamma-\beta}{\gamma-\alpha}}\cdot\|A^\gamma f\|^{\frac{\beta-\alpha}{\gamma-\alpha}} \quad \textit{holds for all } f \in D(A^\gamma).$$

Corollary 7.26. *Let $\alpha \in (0,1]$ and let B be a closed operator such that $D(B) \supseteq D(A^\alpha)$. Then the following assertions are true:*

a) There is a $K \ge 0$ such that

$$\|Bf\| \le K\|A^\alpha f\| \quad \textit{for all } f \in D(A^\alpha).$$

b) There is $K_0 \ge 0$ such that

$$\|Bf\| \le K_0\big(s^\alpha\|f\| + s^{\alpha-1}\|Af\|\big) \quad \textit{holds for all } s > 0 \textit{ and } f \in D(A^\alpha).$$

Proof. a) Since by Exercise 1 the operator $BA^{-\alpha}$ is closed, and since it is by assumption everywhere defined, it is bounded by the closed graph theorem, see Theorem 2.32. Boundedness of $BA^{-\alpha}$ means precisely the assertion.

b) The assertion follows from part a) and Proposition 7.23.a). $\square$

After having seen the fine structure of the embeddings of domains of complex powers, let us close this lecture by returning to the motivating question. To which the last result gives one possible answer.

Proposition 7.27. *Let A generate a semigroup T of type (M,ω) with $\omega < 0$ and consider the powers $(-A)^z$ for $\operatorname{Re} z > 0$. The domain $D((-A)^z)$ is invariant under the semigroup T. The restriction of T to this subspace is a strongly continuous semigroup of bounded linear operators for the norm $\|\cdot\|_{(-A)^z}$. The type of this semigroup is (M,ω).*

Proof. Since the bounded operator $(-A)^{-z}$ commutes with $-R(-\lambda,-A) = R(\lambda,A)$, as a consequence of the convergence of the implicit Euler scheme(see Theorem 5.10), we obtain that $(-A)^{-z}$ commutes with the semigroup operators $T(t)$. This implies that $\operatorname{ran}((-A)^{-z}) = D((-A)^z)$ is invariant under the semigroup. Moreover, we have

$$\|(-A)^z T(t) f\| \le \|T(t)\| \cdot \|(-A)^z f\|,$$

so $T(t) \in \mathscr{L}(D((-A)^z))$. The strong continuity follows from

$$\|(T(t)-I)f\|_{(-A)^z} = \|(-A)^z(T(t)-I)f\| = \|T(t)(-A)^z f - (-A)^z f\|. \qquad \square$$

Exercises

1. Suppose $A : D(A) \to X$ is closed and $B \in \mathscr{L}(X)$ is bounded.

a) Prove that the product AB with

$$D(AB) = \big\{ f \in X : Bx \in D(A) \big\}.$$

is closed.

b) Give an example for A and B such that BA with $D(BA) = D(A)$ is not closed.

2. Let $m = (m_n) \subseteq \mathbb{C}$ be a sequence. Give a sufficient and necessary condition on m so that the multiplication operator M_m fulfills Assumption 7.3. Determine in that case the powers of M_m.

3. Prove that for $t \in \mathbb{R}$ and $f \in D(A^{\mathrm{i}t})$ we have $A^{\mathrm{i}t} f \in D(A^{-\mathrm{i}t})$ and $A^{-\mathrm{i}t} A^{\mathrm{i}t} f = f$.

4. Prove the identity (7.7) in Remark 7.22.

5. Suppose A is densely defined, and take $z \in \mathbb{C}$ with $\operatorname{Re} z < 0$. Prove that $T(t) := A^{zt}$, $t > 0$ and $T(0) = I$ defines a strongly continuous semigroup.

6. Assume we have proved assertion b) in Proposition 7.23. Deduce part a) from that.

7. Let $\alpha \in (0,1)$. Prove that for all $\lambda > 0$ sufficiently large we have

$$\|A^\alpha R(-\lambda, A)\| < 1.$$

Compare this to Exercise 6.5.

8. Prove what has been remaining from Proposition 7.1.

Lecture 8

Intermediate Spaces

In the present lecture we give some further examples of spaces lying in the original Banach space X and being invariant under a given semigroup T semigroup. As a motivation let us first reconsider the convergence of the finite difference scheme from Example 3.7.

Example 8.1. Let

$$X := \{f \in \mathrm{C}([0,1]) : f(1) = 0\} = \mathrm{C}_{(0)}([0,1]) \quad \text{and} \quad X_n = \mathbb{C}^n,$$

both with the respective maximum norm, and introduce the operators

$$(P_n f)_k := f(\tfrac{k}{n}), \quad k = 0, \dots, n-1,$$

and

$$J_n(y_0, \dots, y_{n-1}) := \sum_{k=0}^{n-1} y_k B_{n,k}(x).$$

Recall that the operator

$$Af := f' \quad \text{with} \quad D(A) := \{f \in \mathrm{C}^1([0,1]) : f(1) = 0,\ f'(1) = 0\}$$

is the generator of the nilpotent left shift semigroup on X, and is to be approximated by an appropriate sequence A_n. For $y = (y_0, \dots, y_{n-1}) \in X_n$, we define

$$(A_n y)_k := n(y_{k+1} - y_k) \quad \text{for } k = 0, \dots, n-2 \quad \text{and} \quad (A_n y)_{n-1} := -n y_{n-1},$$

being the standard first-order finite difference scheme. Suppose now that $f \in \mathrm{C}^{1,\alpha}([0,1])$ for $0 < \alpha \leq 1$, i.e, we assume that f' is α-Hölder continuous, see also Example 8.16 below. By handling the real and imaginary parts of f separately, we may assume that f is real and we see that

$$\left|(A_n P_n f - P_n A f)_k\right| = \left|\frac{f(\frac{k+1}{n}) - f(\frac{k}{n})}{\frac{1}{n}} - f'(\tfrac{k}{n})\right| = \left|f'(\xi_k) - f'(\tfrac{k}{n})\right|.$$

This means that there is a constant $C = C(f) > 0$ so that

$$\|A_n P_n f - P_n A f\| \leq \frac{C}{n^\alpha}$$

holds for all $n \in \mathbb{N}$. So the approximation has “order” $\alpha \in (0,1]$. Summarising, though we relaxed the smoothness condition on f, we still get a convergence estimate.

As this example shows, there is another possibility to find an appropriate intermediate space between X and X_1 than taking the fractional powers: Instead of $D(A)$ let us consider the spaces where the semigroup is not necessarily differentiable but only Hölder continuous. The aim of this lecture is to explore this possibility.

8.1 Favard spaces

We start with the basic definitions.

Definition 8.2. Let T be a semigroup of type (M,ω) with $\omega < 0$ and let $\alpha \in (0,1]$. The space

$$F_\alpha := \Big\{ f \in X : \sup_{t>0} \big\| \tfrac{1}{t^\alpha}(T(t)f - f) \big\| < \infty \Big\}$$

with the norm $$\|f\|_{F_\alpha} := \sup_{t>0} \big\| \tfrac{1}{t^\alpha}(T(t)f - f) \big\|$$

is called the **Favard space of order** α (of the semigroup T).

Although we usually do not record it in notation, the Favard space does depend on the semigroup. In most cases it should be always clear from the context Which semigroup is actually meant. However, we shall occasionally write $F_\alpha(T)$ to avoid ambiguity.
Note also that the supremum condition in the definition of F_α depends only on the behaviour of the function $t \mapsto T(t)f$ near zero.

It is easy to see that F_α is a normed linear space. Moreover, from the definition it follows immediately that for $0 < \alpha < \beta \leq 1$ we have

$$D(A) \subseteq F_\beta \subseteq F_\alpha \subseteq X.$$

Moreover, the corresponding norms can be compared:

Lemma 8.3. *a) For $\alpha, \beta \in (0,1)$ with $\alpha < \beta$ the embeddings*

$$D(A) \hookrightarrow F_\beta \hookrightarrow F_\alpha \hookrightarrow X$$

are continuous, where $D(A)$ is equipped with the graph norm (in this case equivalent to $f \mapsto \|Af\|$).

b) The Favard norm $\|\cdot\|_{F_1}$ and the graph norm $\|\cdot\|_1$ are equivalent on $D(A)$.

The proof of this lemma is left as Exercise 3.

Proposition 8.4. *Let T be a semigroup of type (M,ω) with $\omega < 0$, and let $\alpha \in (0,1]$. Then F_α is a Banach space and it is invariant under the semigroup T. For all $t \geq 0$ we have $T(t) \in \mathscr{L}(F_\alpha)$.*

Proof. First we show that F_α is complete. To this end, take a Cauchy sequence $(x_m) \subset F_\alpha$, which is a Cauchy sequence by Lemma 8.3 in X, hence it has a limit $f \in X$. For $t > 0$ fixed we have that

$$\big\| \tfrac{1}{t^\alpha}(T(t)f - f) \big\| = \lim_{m\to\infty} \big\| \tfrac{1}{t^\alpha}(T(t)f_m - f_m) \big\| \leq \limsup_{m\to\infty} \|f_m\|_{F_\alpha},$$

which implies that

$$f \in F_\alpha \quad \text{and} \quad \|f\|_{F_\alpha} \leq \limsup_{m\to\infty} \|f_m\|_{F_\alpha}.$$

Since $f_m - f_n \to f - f_n$ in X as $m \to \infty$, the argumentation above yields

$$\|f - f_n\|_{F_\alpha} \leq \limsup_{m\to\infty} \|f_m - f_n\|_{F_\alpha}.$$

From this we can conclude that for every $\varepsilon > 0$

$$\|f - f_n\|_{F_\alpha} \leq \varepsilon$$

holds if $n \in \mathbb{N}$ sufficiently large. This shows that $f_n \to f$ in F_α, meaning that F_α is complete.

The space F_α is clearly invariant under the semigroup, and we have the estimate

$$\|T(s)f\|_{F_\alpha} = \sup_{t>0} \left\| \tfrac{1}{t^\alpha}(T(t)T(s)f - T(s)f) \right\| = \sup_{t>0} \left\| T(s)\tfrac{1}{t^\alpha}(T(t)f - f) \right\| \le \|T(s)\| \cdot \|f\|_{F_\alpha}$$

showing that $T(s) \in \mathscr{L}(F_\alpha)$ and that $\|T(s)\|_{F_\alpha} \le \|T(s)\|$. □

Remark 8.5. The restriction of the semigroup T to F_α need not be strongly continuous on F_α. This small observation will play an important role later.

The next result is a characterisation of the Favard spaces of T in terms of the generator A. More precisely, it states that resolvent decay in in infinity is connected to Hölder continuity of the semigroup near zero.

Proposition 8.6. *Suppose that A generates a semigroup T of type (M, ω) with $\omega < 0$ and let $\alpha \in (0, 1]$. Then we have for the Favard space of T that*

$$F_\alpha = \Big\{ f \in X : \sup_{\lambda > 0} \|\lambda^\alpha AR(\lambda, A)f\| < \infty \Big\},$$

and the Favard norm $\|\cdot\|_{F_\alpha}$ is equivalent to

$$|||f|||_{F_\alpha} := \sup_{\lambda>0} \|\lambda^\alpha AR(\lambda, A)f\|\,.$$

Proof. That $|||\cdot|||_{F_\alpha}$ is a norm we leave as Exercise 5. Take $f \in F_\alpha$. For $\lambda > 0$ we have by Proposition 2.26 that

$$\lambda^\alpha AR(\lambda, A)f = \lambda^{\alpha+1}R(\lambda, A)f - \lambda^\alpha f = \lambda^{\alpha+1} \int_0^\infty \mathrm{e}^{-\lambda s}(T(s)f - f)\mathrm{d}s,$$

hence $$\|\lambda^\alpha AR(\lambda, A)f\| \le \lambda^{\alpha+1} \int_0^\infty \mathrm{e}^{-\lambda s} s^\alpha \|f\|_{F_\alpha} \mathrm{d}s = \|f\|_{F_\alpha} \int_0^\infty \mathrm{e}^{-r} r^\alpha \mathrm{d}r = \Gamma(\alpha+1)\|f\|_{F_\alpha}.$$

This yields

$$\sup_{\lambda>0} \|\lambda^\alpha AR(\lambda, A)f\| \le \Gamma(\alpha+1)\|f\|_{F_\alpha}. \tag{8.1}$$

Conversely, suppose that $|||f|||_{F_\alpha} := \sup_{\lambda>0} \|\lambda^\alpha AR(\lambda, A)f\| < \infty$. Fix $t > 0$ and recall from Proposition 2.9 that $T(t)f - f = A\int_0^t T(s)f\mathrm{d}s$ holds. Let us decompose f as

$$f = \lambda R(\lambda, A)f - AR(\lambda, A)f =: f_\lambda - g_\lambda,$$

where $\lambda > 0$ is to be specified later. Since $f_\lambda \in D(A)$, we have

$$T(t)f_\lambda - f_\lambda = \int_0^t T(s)Af_\lambda \mathrm{d}s = \lambda \int_0^t T(s)AR(\lambda, A)f\mathrm{d}s,$$

implying that

$$\|T(t)f_\lambda - f_\lambda\| \le \lambda t M \|AR(\lambda, A)f\| = \lambda^{1-\alpha} t M \|\lambda^\alpha AR(\lambda, A)f\|. \tag{8.2}$$

For g_λ we use the trivial estimate

$$\|T(t)g_\lambda - g_\lambda\| \le 2M\|g_\lambda\| = 2M\|AR(\lambda, A)f\| = 2M\lambda^{-\alpha}\|\lambda^\alpha AR(\lambda, A)f\|. \tag{8.3}$$

The estimates in (8.2) and (8.3) imply

$$\begin{aligned}\left\|\tfrac{1}{t^\alpha}(T(t)f - f)\right\| &\le \left\|\tfrac{1}{t^\alpha}(T(t)f_\lambda - f_\lambda)\right\| + \left\|\tfrac{1}{t^\alpha}(T(t)g_\lambda - g_\lambda)\right\| \\ &\le (t\lambda)^{1-\alpha}M\|\lambda^\alpha AR(\lambda, A)f\| + 2M(t\lambda)^{-\alpha}\|\lambda^\alpha AR(\lambda, A)f\|\end{aligned} \tag{8.4}$$

and, by (8.1) we obtain

$$\left\|\tfrac{1}{t^\alpha}(T(t)f - f)\right\| \le M|||f|||_{F_\alpha}(t\lambda)^{1-\alpha} + 2M|||f|||_{F_\alpha}(t\lambda)^{-\alpha}.$$

By choosing $\lambda = \frac{1}{t}$ we see that

$$\left\|\tfrac{1}{t^\alpha}(T(t)f - f)\right\| \le 3M|||f|||_{F_\alpha}, \quad \text{i.e.,} \quad f \in F_\alpha. \tag{8.5}$$

The asserted equality is proved. The equivalence of the norms follows from the estimates in (8.1) and (8.5). □

8.2 Hölder spaces

The next class of intermediate spaces for a semigroup T is connected to the strong continuity of T on subspaces of F_α.

Definition 8.7. Let T be a semigroup of type (M, ω) with $\omega < 0$, and let $\alpha \in (0, 1)$. We define

$$X_\alpha := \Big\{f \in X : \lim_{t\searrow 0}\left\|\tfrac{1}{t^\alpha}(T(t)f - f)\right\| = 0\Big\}$$

and the norm

$$\|f\|_{X_\alpha} = \|f\|_{F_\alpha} = \sup_{t>0}\left\|\tfrac{1}{t^\alpha}(T(t)f - f)\right\|,$$

which makes X_α a normed linear space, called the abstract **Hölder space of order** α (of the semigroup T). To stick to the more precise notation we shall sometime write $X_\alpha(T)$.

It is clear from the definition that the Hölder space X_α is a subspace of the Favard space F_α.

Proposition 8.8. *Let T be a semigroup of type (M, ω) with $\omega < 0$ and let $\alpha \in (0, 1)$. Then X_α is a Banach space and it is invariant under the semigroup T. For all $t \ge 0$ we have $T(t) \in \mathscr{L}(X_\alpha)$.*

The proof of this result is left as Exercise 2. The next result compares the Favard and Hölder spaces of different order.

Proposition 8.9. *For $0 < \alpha < \beta \le 1$ we have*

$$D(A) \hookrightarrow X_\beta \hookrightarrow F_\beta \hookrightarrow X_\alpha \hookrightarrow X$$

with continuous embeddings.

Proof. For $\beta = 1$ we have $D(A) = X_\beta$, so the assertion follows by Lemma 8.3. We may assume $\beta < 1$. The continuity of the embedding $X_\beta \hookrightarrow F_\beta$ is then trivial. Therefore, in view of Lemma 8.3 again, it suffices to prove that $F_\beta \hookrightarrow X_\alpha$ is continuous. For $f \in F_\beta$ we have

$$\left\| \tfrac{1}{t^\alpha}(T(t)f - f)\right\| = t^{\beta-\alpha} \left\| \tfrac{1}{t^\beta}(T(t)f - f)\right\| \leq t^{\beta-\alpha}\|f\|_{F_\beta}.$$

This estimate yields $f \in X_\alpha$, and the by Lemma 8.3 the continuity of the embedding follows, too. □

Although T may not be strongly continuous on F_α on the space X_α, generally smaller than F_α, strong continuity holds.

Proposition 8.10. *Let T be a semigroup of type (M, ω) with $\omega < 0$ and with generator A, and let $\alpha \in (0, 1]$. Then we have for the Hölder space of T that*

$$X_\alpha = \Big\{ f \in F_\alpha : \lim_{t\searrow 0} \|T(t)f - f\|_{F_\alpha} = 0 \Big\},$$

i.e., the space X_α is the space of strong continuity for the function $t \mapsto T(t)$ in F_α. Moreover, we have

$$X_\alpha = \overline{D(A)}^{F_\alpha}.$$

Proof. Let first $\alpha \in (0, 1]$ and take $f \in X_\alpha$. For a given $\varepsilon > 0$ choose $\delta > 0$ so that

$$\left\| \tfrac{1}{s^\alpha}(T(s)f - f)\right\| \leq \frac{\varepsilon}{2M} \qquad \text{holds for } 0 < s < \delta.$$

Then, for $0 < s < \delta$ and $t > 0$, we have

$$\left\| \tfrac{1}{s^\alpha}(T(s) - I)(T(t) - I)f\right\| \leq 2M \left\| \tfrac{1}{s^\alpha}(T(s) - I)f\right\| \leq 2M\frac{\varepsilon}{2M} = \varepsilon.$$

Furthermore, there is $\delta' > 0$ so that if $0 < t < \delta'$, then

$$\|T(t)f - f\| \leq \frac{\delta^\alpha \varepsilon}{2M}.$$

Hence, for $s > \delta$ and $0 < t < \delta'$, we have

$$\left\| \tfrac{1}{s^\alpha}(T(s) - I)(T(t) - I)f\right\| \leq 2M \tfrac{1}{s^\alpha}\|T(t)f - f\| \leq 2M \tfrac{1}{\delta^\alpha}\frac{\delta^\alpha\varepsilon}{2M} = \varepsilon.$$

To sum up, for $t < \delta'$ we have

$$\|T(t)f - f\|_{F_\alpha} = \sup_{s>0} \left\| \tfrac{1}{s^\alpha}(T(s) - I)(T(t) - I)f\right\| \leq \varepsilon.$$

Hence $\lim_{t\searrow 0} \|T(t)f - f\|_{F_\alpha} = 0$, meaning that $t \mapsto T(t)|_{X_\alpha}$ is strongly continuous, and hence a semigroup.

Suppose now that $f \in F_\alpha$ is such that $t \mapsto T(t)f$ is strongly continuous. Using again the argument that the convergence of the Riemann sums in the F_α-norm implies the convergence in the X norm, it follows that

$$\lim_{r\searrow 0} \left\| f - \frac{1}{r}\int_0^r T(s)f\,\mathrm{d}s \right\|_{F_\alpha} = 0,$$

which implies that $f \in \overline{D(A)}^{F_\alpha}$.

Let now $f \in \overline{D(A)}^{F_\alpha}$. Suppose first $\alpha \neq 1$, and take $\varepsilon > 0$. Then there is $g \in D(A)$ so that $\|f - g\|_{F_\alpha} \leq \frac{\varepsilon}{2}$ holds. Furthermore, we have

$$\|T(t)g - g\| \leq tM\|Ag\|.$$

We conclude

$$\left\|\tfrac{1}{t^\alpha}(T(t)f - f)\right\| \leq \left\|\tfrac{1}{t^\alpha}(T(t)g - g)\right\| + \left\|\tfrac{1}{t^\alpha}(T(t)(f-g) - (f-g))\right\| \leq t^{1-\alpha}M\|Ag\| + \|f-g\|_{F_\alpha} \leq \varepsilon$$

for $t > 0$ sufficiently small. This shows that $f \in X_\alpha$ in the case $\alpha \neq 1$. Now, suppose $\alpha = 1$. Then, by Lemma 8.3.b), $X_1 = D(A)$ is a Banach space if equipped with the Favard norm $\|\cdot\|_{F_1}$. So actually we have $\overline{D(A)}^{F_\alpha} = X_1$.

We have therefore proved

$$X_\alpha \subseteq \Big\{f \in X : \lim_{t\searrow 0} \|T(t)f - f\|_{F_\alpha} = 0\Big\} \subseteq \overline{D(A)}^{F_\alpha} \subseteq X_\alpha. \qquad \square$$

Here is a description, analogous to Proposition 8.6, of the Hölder spaces of T in terms the generator of T.

Proposition 8.11. *Suppose that A generates a semigroup of type (M, ω) with $\omega < 0$ and let $\alpha \in (0,1)$. Then we have*

$$X_\alpha = \Big\{f \in X : \lim_{\lambda\to\infty} \|\lambda^\alpha AR(\lambda, A)f\| = 0\Big\}.$$

Proof. Since $\|\lambda R(\lambda, A)\| \leq M$ for all $\lambda > 0$, for $f \in D(A)$ we clearly have $\lim_{\lambda\to\infty} \|\lambda^\alpha AR(\lambda, A)f\| = 0$. By Proposition 8.10 we obtain

$$X_\alpha = \overline{D(A)}^{F_\alpha} \subseteq \Big\{f \in X : \lim_{\lambda\to\infty} \|\lambda^\alpha AR(\lambda, A)f\| = 0\Big\},$$

because the set on the right hand side is clearly closed in the Favard norm.

Suppose now $f \in X$ is such that $\lim_{\lambda\to\infty} \|\lambda^\alpha AR(\lambda, A)f\| = 0$. By (8.4) we have

$$\left\|\tfrac{1}{t^\alpha}(T(t)f - f)\right\| \leq (t\lambda)^{1-\alpha}M\|\lambda^\alpha AR(\lambda, A)f\| + 2M(t\lambda)^{-\alpha}\|\lambda^\alpha AR(\lambda, A)f\|.$$

If we take $\lambda = \frac{1}{t}$ and let $t \to 0$, we obtain the assertion. $\square$

8.3 Higher order intermediate spaces

In order that we can define higher order intermediate spaces, say, between $D(A)$ and $D(A^k)$f for $k \in \mathbb{N}$, we first recall the next result from Proposition 7.1.

Proposition 8.12. *Let A be the generator of a semigroup of type (M, ω) in the Banach space X, and consider the space $X_n = D(A^n)$ with the graph norm which we denote by $\|\cdot\|_{A^n}$.*

a) For $n \in \mathbb{N}$ and $f \in D(A^n)$ define $|||f|||_n := \|f\| + \|Af\| + \cdots + \|A^n f\|$. Then $|||\cdot|||_n$ and $\|\cdot\|_{A^n}$ are equivalent norms.

b) *The spaces X_n are Banach spaces and are invariant under the semigroup T. If we set $T_n(t) := T(t)|_{X_n}$, then T_n is a semigroup of type (M,ω) on X_n.*

By part a) we have that X_n is continuously embedded in X_k for all $k,n \in \mathbb{N}_0$ with $n > k$. With the help of this proposition we can extend the scale of the spaces F_α and X_α for all $\alpha > 0$.

Definition 8.13. Suppose that A generates a semigroup T of type (M,ω) with $\omega < 0$ and let $\alpha > 0$. Set $X_0 = X$, $T_0 = T$ and write $\alpha = k + \alpha'$ with $k \in \mathbb{N}_0$ and $\alpha' \in (0,1]$. We define

$$F_\alpha = F_{\alpha'}(T_k) = \text{the Favard space of } T_k \text{ of order } \alpha'$$
$$X_\alpha = F_{\alpha'}(T_k) = \text{the Hölder space of } T_k \text{ of order } \alpha'.$$

Remark 8.14. 1. For $\alpha \in \mathbb{N}$ this new definition is consistent with the one in Proposition 8.12.

2. For $\alpha \in (0,1)$ this definition gives back what we already had in the previous two sections.

3. If T is of type (M,ω), then these spaces are defined via the rescaled semigroup given by $\mathrm{e}^{-(\omega+1)t}T(t)$.

The next result gives the relation between theses intermediate spaces.

Proposition 8.15. *Let $0 < \alpha < \beta$ and let T be a semigroup. Then the following assertion are true:*

a) *X_α is closed subspace of F_α.*

b) *F_β is contained in X_α and the embedding*

$$F_\beta \hookrightarrow X_\alpha$$

is continuous.

c) *The space X_β is dense in X_α for all $0 \le \alpha < \beta$.*

d) *The spaces F_α and X_α are invariant under the semigroup T. Let T_α be the semigroup T restricted to the abstract Hölder space X_α. Then T_α is a strongly semigroup on X_α.*

Proof. a) Write $\alpha = k + \alpha'$ with $k \in \mathbb{N}_0$ and $\alpha' \in (0,1]$. Then, if $\alpha' \neq 1$ the assertion follows from Proposition 8.8 applied to T_k and α'.

b) Write $\alpha = k + \alpha'$, $\beta = n + \beta'$ with $k,n \in \mathbb{N}_0$ and $\alpha', \beta' \in (0,1]$. If $k = n$, then $\alpha' < \beta'$ and the assertion follows by Proposition 8.9. For $f \in F_\alpha = F_{\alpha'}(T_k)$ we have $f \in X_{\beta'}(T_k) = X_\beta$

$$\tfrac{1}{t^\alpha}(T(t)f - f)$$

In the case $n > k$ we have by Lemma 8.3 the continuous embeddings

$$X_n \hookrightarrow X_{k+1} = X_1(T_k) \hookrightarrow X_{\alpha'}(T_k).$$

Also by Lemma 8.3 we obtain that $F_\alpha = F_{\alpha'}(T_k)$ is continuously embedded in X_k. This finishes the proof of b).

c) Write $\alpha = k + \alpha'$ with $k \in \mathbb{N}_0$ and $\alpha' \in (0,1]$. It suffices to prove that $X_n = D(A^n)$ is dense in X_α if $n \ge k+1$. But for such n the space $X_n = D(A^n)$ is dense in $X_{k+1} = D(A^{k+1})$ by Proposition 2.18. Whereas X_{k+1} is densely and continuously embedded in $X_\alpha = X_{\alpha'}(T_k)$ by Propositions 8.10 and 8.9. This complete the proof.

Assertion d) follows immediately from the definition. □

8.4 Basic examples

As we have seen, the spaces F_α and X_α are related to Hölder continuity of semigroup orbits. The next example underlines this fact and, at least partly, the chosen terminology.

Example 8.16. Consider the first derivative, the generator of the left shift semigroup S in $\mathrm{BUC}(\mathbb{R})$. We determine the Favard and the Hölder spaces for this semigroup. More precisely, we rescale $T(t) = \mathrm{e}^{-t}S(t)$ and then determine the corresponding intermediate spaces. First we suppose $\alpha \in (0,1)$. Recall that a function $f : \mathbb{R} \to \mathbb{C}$ is called **α-Hölder continuous** if

$$\sup_{\substack{s,t\in\mathbb{R}\\ s\neq t}} \frac{|f(t)-f(s)|}{|t-s|^\alpha} < \infty,$$

and it is called **little-α-Hölder continuous** if

$$\lim_{h\searrow 0} \sup_{\substack{s,t\in\mathbb{R}\\ 0<|t-s|<h}} \frac{|f(t)-f(s)|}{|t-s|^\alpha} = 0.$$

For the abstract Favard and Hölder spaces of the left shift we have then

$$F_\alpha = \mathrm{C}_\mathrm{b}^\alpha(\mathbb{R}) := \Big\{ f \in \mathrm{C}_\mathrm{b}(\mathbb{R}) : f \text{ is } \alpha\text{-Hölder continuous} \Big\}$$

and

$$X_\alpha = \mathrm{h}^\alpha(\mathbb{R}) := \Big\{ f \in \mathrm{C}_\mathrm{b}(\mathbb{R}) : f \text{ is little-}\alpha\text{-Hölder continuous} \Big\}.$$

The Favard norm is equivalent to

$$\|f\|_{\mathrm{C}_\mathrm{b}^\alpha(\mathbb{R})} = \|f\|_\infty + \sup_{\substack{s,t\in\mathbb{R}\\ s\neq t}} \frac{|f(t)-f(s)|}{|t-s|^\alpha}.$$

For $\alpha = 1$ we have

$$F_1 = \mathrm{Lip}_\mathrm{b}(\mathbb{R}) = \Big\{ f \in \mathrm{C}_\mathrm{b}(\mathbb{R}) : f \text{ is Lipschitz continuous} \Big\}$$

with equivalent norm

$$\|f\|_{\mathrm{Lip}_\mathrm{b}(\mathbb{R})} = \|f\|_\infty + \sup_{\substack{s,t\in\mathbb{R}\\ s\neq t}} \frac{|f(t)-f(s)|}{|t-s|}.$$

For general $\alpha > 0$, we write $\alpha = k + \alpha'$ with $k \in \mathbb{N}_0$ and $\alpha' \in (0,1]$. Then we have

$$F_\alpha = \mathrm{C}_\mathrm{b}^{k,\alpha'}(\mathbb{R}) := \Big\{ f \in \mathrm{C}_\mathrm{b}(\mathbb{R}) : f \text{ is } k\text{-times differentiable with } f^{(k)} \in \mathrm{C}_\mathrm{b}^{\alpha'}(\mathbb{R}) \Big\}$$

and, if $\alpha \notin \mathbb{N}$

$$X_\alpha = \mathrm{h}^{k,\alpha'}(\mathbb{R}) := \Big\{ f \in \mathrm{C}_\mathrm{b}(\mathbb{R}) : f \text{ is } k\text{-times differentiable with } f^{(k)} \in \mathrm{h}^{\alpha'}(\mathbb{R}) \Big\}.$$

We leave the proof of these assertions as Exercise 4.

To determine the Favard and the Hölder spaces of the left shift on $\mathrm{L}^1(\mathbb{R})$ is more complicated. So we state the result for F_1 only, and that without proof.

Example 8.17. Consider the left shift semigroup in $\mathrm{L}^1(\mathbb{R})$. Then for its Favard space we have

$$F_1 = \mathrm{UBV}_1(\mathbb{R}) := \big\{f \in \mathrm{L}^1(\mathbb{R}) : f \text{ is a.e. equal to } g : \mathbb{R} \to \mathbb{C} \text{ of uniformly bounded variation}\big\}.$$

Recall that a function g is of uniformly bounded variation if

$$\bigvee_{-\infty}^{\infty} g = \sup_{R>0} \bigvee_{-R}^{R} g < \infty,$$

where $\bigvee_{-R}^{R} g$ is the variation of g on $[-R, R]$ given by

$$\bigvee_{-R}^{R} g = \sup\Big\{\sum_{j=1}^{n} |g(t_j) - g(t_{j-1})| : n \in \mathbb{N},\ -R = t_0 < t_1 < \cdots < t_n = R\Big\}.$$

Now, such a function g has at most countable discontinuity points, at which the left and right limits exist. If we modify the function g at these points so that it becomes left continuous, the variation will still remain the same. Now for $f \in \mathrm{UBV}_1(\mathbb{R})$ there is a unique left continuous function g which is of uniformly bounded variation and which coincides with f almost everywhere (more precisely, f is the L^1-equivalence class of g). We define $\bigvee_{-\infty}^{\infty} f := \bigvee_{-\infty}^{\infty} g$. The Favard norm F_1 is then equivalent to

$$f \mapsto \bigvee_{-\infty}^{\infty} f + \|f\|_1.$$

8.5 Relation to fractional powers

Recall from Lecture 7 that if A generates a semigroup of type (M, ω) with $\omega < 0$, then it is possible to define the fraction (or complex powers) of $-A$. In this section we want to compare the domains $D((-A)^\alpha)$ to the abstract intermediate spaces introduce in the above. In the first to auxiliary results we only suppose that $-A$ satisfies the Assumption 7.3, i.e., that the complex powers can be defined as described in Lecture 7.

Lemma 8.18. *Suppose $-A$ is as in Assumption 7.3 and let $\alpha \in (0, 1)$. Then we have*

$$\sup_{s\in[0,\infty)} \|s^\alpha AR(s, A)f\| \le K\|(-A)^\alpha f\|$$

for all $f \in D((-A)^\alpha)$.

Proof. By Proposition 7.13 we have

$$(-A)^{-\alpha} = -\frac{\sin(\pi\alpha)}{\pi} \int_0^\infty s^{-\alpha} R(s, A)\mathrm{d}s = -\frac{\sin(\pi\alpha)}{\pi} \int_0^\infty t(ts)^{-\alpha} R(ts, A)\mathrm{d}s$$

for all $t > 0$. We can then write

$$\begin{aligned}(-A)^{-\alpha} t^\alpha AR(t, A) = &-\frac{\sin(\pi\alpha)}{\pi} \int_0^1 s^{-\alpha} AR(ts, A) \cdot tR(t, A)\mathrm{d}s \\ &-\frac{\sin(\pi\alpha)}{\pi} \int_1^\infty s^{-\alpha-1}(ts)R(ts, A) \cdot AR(t, A)\mathrm{d}s.\end{aligned}$$

Since for $t \geq 0$ we have $\|tR(t,A)\| \leq M$ and $\|AR(t,A)\| \leq M+1$, we obtain

$$\|A^{-\alpha} t^{\alpha} AR(t,A)g\| \leq K\|g\|.$$

If we plug in $g = (-A)^{\alpha} f$, we obtain the assertion. □

Lemma 8.19. *Suppose that $-A$ is as in Assumption 7.3. For $\alpha, \beta \in (0,1)$ with $\alpha < \beta$ we have*

$$\|(-A)^{\alpha} f\| \leq K \sup_{s\in[0,\infty)} \|s^{\beta} AR(s,A)f\|$$

for all $f \in D(A)$. (By the way the right-hand side is finite for all $f \in D(A)$.)

Proof. Let $f \in D(A)$. Then we have $\|s^{\beta} AR(s,A)f\| \leq \frac{Ms^{\beta}}{1+s}\|Af\|$, so the expression on the right-hand side is finite. Moreover, by Proposition 7.21 we have that

$$(-A)^{\alpha} f = \frac{\sin(\pi\alpha)}{\pi} \int_0^{\infty} s^{\alpha-1} R(s,A)Af \mathrm{d}s$$

and we can split the integration into two parts

$$(-A)^{\alpha} f = \frac{\sin(\pi\alpha)}{\pi} \int_0^{1} s^{\alpha-1} AR(s,A)f \mathrm{d}s + \frac{\sin(\pi\alpha)}{\pi} \int_1^{\infty} s^{\alpha-\beta-1} s^{\beta} AR(s,A)^{-1} f \mathrm{d}s.$$

Now the first term can be estimated as

$$\begin{aligned} \frac{\sin(\pi\alpha)}{\pi} \Big\| \int_0^1 s^{\alpha-1} AR(s,A)f \mathrm{d}s \Big\| &\leq (1+M_0)\|f\| \frac{\sin(\pi\alpha)}{\pi} \int_0^1 s^{\alpha-1} \mathrm{d}s \\ &\leq K_0\|f\| \leq K_0(\|AR(1,A)f\| + \|A^{-1}AR(1,A)f\|) \leq K_1 \sup_{s\in[0,\infty)} \|s^{\beta} AR(s,A)f\|. \end{aligned}$$

For the second term we have

$$\begin{aligned} \frac{\sin(\pi\alpha)}{\pi} \Big\| \int_1^{\infty} s^{\alpha-\beta-1} s^{\beta} AR(s,A)f \mathrm{d}s \Big\| &\leq \frac{\sin(\pi\alpha)}{\pi} \int_1^{\infty} s^{\alpha-\beta-1} \mathrm{d}s \cdot \sup_{s\in[0,\infty)} \|s^{\beta} AR(s,A)f\| \\ &\leq K_2 \sup_{s\in[0,\infty)} \|s^{\beta} AR(s,A)f\|. \end{aligned}$$

By putting the two estimates together we conclude the proof. □

Theorem 8.20. *Let A be the generator of a semigroup T of type (M,ω) with $\omega < 0$ and let $0 < \alpha < \beta < 1$. Then*

$$X_{\beta} \subseteq F_{\beta} \hookrightarrow D\big((-A)^{\alpha}\big) \hookrightarrow X_{\alpha} \subseteq F_{\alpha}$$

with continuous embeddings.

Proof. Let $\gamma \in (\alpha, \beta)$. By Proposition 8.6 the Favard norm on F_γ is equivalent to

$$|||f|||_{F_\gamma} := \sup_{\lambda > 0} \|\lambda^\gamma AR(\lambda, A)f\| .$$

Now by Lemma 8.19 we conclude that the normed space $D(A) \subseteq X_\gamma$ with the norm $|||\cdot|||_{F_\gamma}$ is continuously embedded in $D\big((-A)^\alpha\big)$, hence so is its closure $X_\gamma = \overline{D(A)}^{F_\gamma}$ (see also Proposition 8.10). Now, by Proposition 8.15 we have $F_\beta \hookrightarrow X_\gamma$ with continuous embedding, so altogether we obtain the continuous embedding

$$F_\beta \hookrightarrow D\big((-A)^\alpha\big)$$

Lemma 8.18 yields the continuous embedding

$$D\big((-A)^\alpha\big) \hookrightarrow \overline{D(A)}^{F_\alpha} = X_\alpha. \qquad \square$$

8.6 Outlook

We mention here a few facts without proofs. Notice first of all that given a semigroup T of type (M, ω) with $\omega < 0$ we have by definition

$$F_\alpha = \Big\{ f \in X : t \mapsto \psi(t) := \tfrac{1}{t^\alpha}\|T(t)f - f\| \in \mathrm{L}^\infty\big((0,\infty)\big) \Big\}.$$

Since $\omega < 0$, the fact that f belongs to F_α does not say too much about the behaviour of $T(t)f$ for large t, rather it tells quite a lot about small t values. We may try to dig out more information by giving more weight to small t values. Therefore, we consider the measure $\frac{\mathrm{d}t}{t}$ which is absolutely continuous with respect to the Lebesgue measure on $(0, \infty)$ with Radon–Nikodým derivative $\frac{1}{t}$. This measure is even equivalent to the Lebesgue measure so the corresponding L^∞ spaces and the L^∞ norms will be the same. Hence, as a matter of fact, no new information has been obtained. However, we may go over the L^p-scale, where we get a more detailed picture. For $p \in [1, \infty]$ let us denote by $\mathrm{L}^p_*(0, \infty)$ the L^p space with respect to the measure $\frac{\mathrm{d}t}{t}$ from above. For $\alpha \in (0,1)$ and $p \geq 1$ we define

$$\big(X, D(A)\big)_{\alpha,p} := \Big\{ f \in X : t \mapsto \psi(t) := \tfrac{1}{t^\alpha}\|T(t)f - f\| \in \mathrm{L}^p_*(0,\infty) \Big\},$$

with the norm

$$\|f\|_{(X,D(A))_{\alpha,p}} := \|f\| + \|\psi\|_{\mathrm{L}^p_*(0,\infty)}.$$

These spaces are Banach spaces and they are invariant under the semigroup T.

Moreover, it is possible to show—in analogy to Proposition 8.6—that

$$\big(X, D(A)\big)_{\alpha,p} = \Big\{ f \in X : \lambda \mapsto \phi(\lambda) := \lambda^\alpha \|AR(\lambda, A)f\| \in \mathrm{L}^p_*(0,\infty) \Big\},$$

and that

$$|||f|||_{(X,D(A))_{\alpha,p}} := \|\phi\|_{\mathrm{L}^p_*(0,\infty)}$$

defines an equivalent norm. Furthermore, one has the continuous embeddings

$$D(A) \hookrightarrow \big(X, D(A)\big)_{\alpha,p} \hookrightarrow \big(X, D(A)\big)_{\alpha,r} \hookrightarrow \big(X, D(A)\big)_{\alpha,\infty} = F_\alpha$$

Lecture 9

Analytic Semigroups

Recall the Gaussian semigroup from Lecture 2 defined for $f \in \mathrm{L}^2(\mathbb{R})$ by

$$(T(t)f)(x) := (g_t * f)(x) = \frac{1}{\sqrt{4\pi t}} \int\limits_{\mathbb{R}} f(y) \mathrm{e}^{-\frac{(x-y)^2}{4t}} \mathrm{d}y = \int\limits_{\mathbb{R}} f(y) G(t,x,y) \mathrm{d}y,$$

and $\qquad T(0)f := f.$

Then T is a bounded strongly continuous semigroup on $\mathrm{L}^2(\mathbb{R})$. In Exercise 2.8 you were asked to determine its generator, which is

$$Af(s) = f''(s) \quad \text{with} \quad D(A) = \mathrm{H}^2(\mathbb{R}).$$

First of all we make the following observation: One can show that for all $f \in \mathrm{L}^2(\mathbb{R})$ and $t > 0$

$$T(t)f \in \bigl\{g \in \mathrm{C}^\infty(\mathbb{R}) : \text{all derivatives of } g \text{ belong to } \mathrm{L}^2(\mathbb{R})\bigr\} \subseteq D(A),$$

hence the mapping $t \mapsto T(t)f$ is differentiable on $(0,\infty)$ for all $f \in \mathrm{L}^2(\mathbb{R})$. It is also not hard to see that the mapping

$$(0,\infty) \ni t \mapsto tAT(t)$$

is bounded.

The second important fact to observe is the following. For $f \in \mathrm{L}^2(\mathbb{R})$ and $z \in \mathbb{C}$ with $\operatorname{Re} z > 0$ one can define

$$(T(z)f)(x) := \frac{1}{\sqrt{4\pi z}} \int\limits_{\mathbb{R}} f(y) \mathrm{e}^{-\frac{(x-y)^2}{4z}} \mathrm{d}y.$$

Then

$$T : \{z : \operatorname{Re} z > 0\} \to \mathscr{L}(\mathrm{L}^2(\mathbb{R})) \quad \text{is a holomorphic function}$$

and, following the arguments in Lecture 2, one easily proves that $T(z)T(w) = T(z+w)$ holds for all $z, w \in \mathbb{C}$ with $\operatorname{Re} z, \operatorname{Re} w > 0$.

This lecture is devoted to the study of the two properties above from an operator theoretic point of view.

9.1 Analytic semigroups

Let us first define the main objects of our study.

Definition 9.1. For $\theta \in (0, \frac{\pi}{2}]$ consider the sector

$$\Sigma_\theta := \bigl\{z \in \mathbb{C} \setminus \{0\} : |\arg(z)| < \theta\bigr\}.$$

Suppose $T : \Sigma_\theta \cup \{0\} \to \mathscr{L}(X)$ is a function with the following properties:

(i) $T:\Sigma_\theta \to \mathscr{L}(X)$ is holomorphic.

(ii) For all $z, w \in \Sigma_\theta$ we have

$$T(z)T(w) = T(z+w), \quad \text{and} \quad T(0) = I.$$

(iii) For every $\theta' \in (0,\theta)$ the equality

$$\lim_{\substack{z\to 0\\ z\in\Sigma_{\theta'}}} T(z)f = f \quad \text{holds for all } f \in X.$$

Then T is called an **analytic semigroup of angle** θ. Suppose moreover the following.

(iv) For all $\theta' \in (0,\theta)$ we have

$$\sup_{z\in\Sigma_{\theta'}} \|T(z)\| < \infty.$$

Then T is called a **bounded analytic semigroup** of angle θ. The generator of the restriction $T : [0,\infty) \to \mathscr{L}(X)$ is called the **generator** of the analytic semigroup T.

Remark 9.2. Clearly, for an analytic semigroup T the mapping

$$T : (0,\infty) \to \mathscr{L}(X), \quad t \mapsto T(t) \in \mathscr{L}(X)$$

is continuous in the *operator norm*, it is even differentiable. Among others, this continuity has the following consequence: For λ sufficiently large the resolvent of the generator is given by the improper integral

$$R(\lambda, A) = \int_0^\infty \mathrm{e}^{-\lambda t} T(t)\mathrm{d}t$$

convergent now in the *operator norm*, cf. Proposition 2.26.

Proposition 9.3. *Let T be an analytic semigroup of angle $\theta \in (0, \frac{\pi}{2}]$ with generator A. Then the following assertions are true:*

a) For every $r > 0$ and $\theta' \in (0,\theta)$ we have

$$\sup\big\{\|T(z)\| : z \in \Sigma_{\theta'},\ |z| \le r\big\} < \infty.$$

b) For all $\theta' \in (0,\theta)$ there is $\omega = \omega_{\theta'} > 0$ and $M = M_{\theta'} \ge 1$ such that

$$\|T(z)\| \le M\mathrm{e}^{\omega \operatorname{Re} z} \quad \textit{for all } z \in \Sigma_{\theta'}.$$

c) For $\alpha \in (-\theta,\theta)$ and $t \ge 0$ define $T_\alpha(t) := T(\mathrm{e}^{\mathrm{i}\alpha}t)$. Then T_α is a strongly continuous semigroup with generator $\mathrm{e}^{\mathrm{i}\alpha}A$.

Proof. a) and b) The proof is similar to that of Proposition 2.2 and exploits the uniform boundedness principle.

c) That T_α is a strongly continuous semigroup is trivial from the definition. Let γ be the half-line emanating from the origin with angle α to the positive semi-axis. By Proposition 2.26 we have for μ sufficiently large that

$$R(\mu, A_\alpha) = \int_0^\infty \exp(-\mu t)T(\mathrm{e}^{\mathrm{i}\alpha}t)\mathrm{d}t = \mathrm{e}^{-\mathrm{i}\alpha}\int_\gamma \exp(-\mu\mathrm{e}^{-\mathrm{i}\alpha}z)T(z)\mathrm{d}z.$$

By using Cauchy's theorem we can transform the path of integration to the positive semi-axis $\tilde{\gamma}$ (work out the details), and we conclude

$$\begin{aligned}R(\mu, A_\alpha) &= \mathrm{e}^{-\mathrm{i}\alpha} \int_{\tilde{\gamma}} \exp(-\mu \mathrm{e}^{-\mathrm{i}\alpha} z) T(z) \mathrm{d}z = \mathrm{e}^{-\mathrm{i}\alpha} \int_0^\infty \exp(-\mu \mathrm{e}^{-\mathrm{i}\alpha} t) T(t) \mathrm{d}t \\ &= \mathrm{e}^{-\mathrm{i}\alpha} R(\mu \mathrm{e}^{-\mathrm{i}\alpha}, A) = R(\mu, \mathrm{e}^{\mathrm{i}\alpha} A),\end{aligned}$$

if $\mathrm{Re}(\mu \mathrm{e}^{-\mathrm{i}\alpha})$ is sufficiently large. This proves $A_\alpha = \mathrm{e}^{\mathrm{i}\alpha} A$.

□

Next we present some fundamental examples.

Example 9.4. For $A \in \mathscr{L}(X)$ and $z \in \mathbb{C}$ define

$$T(z) = \mathrm{e}^{zA} := \sum_{n=0}^{\infty} \frac{z^n A^n}{n!}.$$

Then T is an analytic semigroup. In Exercise 1 you are asked to prove this.

Example 9.5. The Dirichlet heat semigroup on $\mathrm{L}^2(0,1)$, see Lecture 1, has a bounded analytic semigroup extension[1] of angle $\frac{\pi}{2}$.

More generally, we have the following.

Example 9.6. Let H be a Hilbert space and A a self-adjoint operator on H, i.e., $A = A^*$. Then the spectral theorem tells us that there is an L^2-space and a unitary operator $S : H \to \mathrm{L}^2$ such that

$$SAS^{-1} : \mathrm{L}^2 \to \mathrm{L}^2, \quad SAS^{-1} = M_m,$$

where M_m is a multiplication operator on L^2 by a real-valued function m (with maximal domain). If A is negative, i.e.,

$$\langle Af, f \rangle \leq 0 \quad \text{for all } f \in D(A),$$

then $\sigma(A) \subseteq (-\infty, 0]$ and m takes values in $(-\infty, 0]$. It is easy to prove that

$$T(z) := S^{-1} M_{\mathrm{e}^{zm}} S$$

defines a bounded analytic semigroup $T : \Sigma_{\frac{\pi}{2}} \cup \{0\} \to \mathscr{L}(H)$ generated by A, see Exercise 2, cf. also Exercise 7.2. Of course, we may only assume that A is bounded above by ωI, then by replacing A by $A - \omega$ the same arguments work, and we obtain that $A - \omega$ generates an analytic semigroup.

Example 9.7. The shift semigroup on $\mathrm{L}^p(\mathbb{R})$ is not analytic. Or, more generally, if T is a strongly continuous group which is not continuous for the operator norm at $t = 0$, then T is not analytic. Prove this statement in Exercise 3.

Proposition 9.8. *The generator of a bounded analytic semigroup of angle θ has the following properties. The sector $\Sigma_{\frac{\pi}{2}} + \theta$ belongs to the resolvent set $\rho(A)$ of A, and for all $\theta' \in (0, \theta)$ there is $M_{\theta'} \geq 1$ such that*

$$\|R(\lambda, A)\| \leq \frac{M_{\theta'}}{|\lambda|} \quad \textit{for all } \lambda \in \Sigma_{\frac{\pi}{2}} + \theta'.$$

[1] We also say that a semigroup is analytic if it has an analytic semigroup extension to a sector.

Proof. Let $\theta' \in (0, \theta)$ and $\theta'' \in (\theta', \frac{1}{2}(\frac{\pi}{2} + \theta'))$ be fixed, and set

$$M_{\theta''} := \sup_{z \in \overline{\Sigma}_{\theta''}} \|T(z)\|.$$

For $\alpha \in [-\theta'', \theta'']$ and $t \geq 0$ define $T_\alpha(t) := T(\mathrm{e}^{\mathrm{i}\alpha} t)$. By assumption T_α is a bounded semigroup, and by Proposition 9.3 its generator is $A_\alpha := \mathrm{e}^{\mathrm{i}\alpha} A$. By Proposition 2.26 we have for all $\operatorname{Re}\mu > 0$ that

$$\|R(\mu, A_\alpha)\| \leq \frac{M_{\theta''}}{\operatorname{Re}\mu}.$$

For $\lambda \in \Sigma_{\frac{\pi}{2}+\theta'}$ with $\arg\lambda \geq 0$ we have ($\arg\lambda \leq 0$ goes similarly)

$$\|R(\lambda, A)\| = \|R(\mathrm{e}^{-\mathrm{i}\theta''}\lambda, \mathrm{e}^{-\mathrm{i}\theta''} A)\| \leq \frac{M_{\theta''}}{\operatorname{Re}(\mathrm{e}^{-\mathrm{i}\theta''}\lambda)} \leq \frac{M_{\theta''}}{|\lambda| \sin(\theta'' - \theta')}. \qquad \square$$

In the next section we show the converse of this statement.

9.2 Sectorial operators

We make the following definition out of the properties listed in Proposition 9.8.

Definition 9.9. Let A be a linear operator on the Banach space X, and let $\delta \in (0, \frac{\pi}{2})$. Suppose that the sector

$$\Sigma_{\frac{\pi}{2}+\delta} := \left\{\lambda \in \mathbb{C} \setminus \{0\} : |\arg\lambda| < \tfrac{\pi}{2} + \delta\right\}$$

is contained in the resolvent set $\rho(A)$, and that

$$\sup_{\lambda \in \Sigma_{\frac{\pi}{2}+\delta'}} \|\lambda R(\lambda, A)\| < \infty \quad \text{for every } \delta' \in (0, \delta).$$

Then the operator A is called **sectorial of angle** δ.

Example 9.10. Let H be a Hilbert space and let A be a negative self-adjoint operator on H. By the spectral theorem we have an L^2-space and a unitary operator $S : H \to \mathrm{L}^2$ such that $SAS^{-1} = M_m$ where m takes values in $(-\infty, 0]$. For $\lambda \in \mathbb{C}$ with $|\arg\lambda| < \theta$, $\theta \in (\frac{\pi}{2}, \pi)$ we have

$$\|\lambda R(\lambda, M_m)\| = \Big\|\frac{|\lambda|}{|\lambda - m|}\Big\|_\infty = \frac{|\lambda|}{|\lambda| \sin(\theta)} = \frac{1}{\sin(\theta)},$$

i.e., M_m (hence A) is sectorial of angle δ for all $\delta \in (0, \frac{\pi}{2})$, see also Exercise 4.

The aim of this section is to show that the densely defined sectorial operators are precisely the generators of bounded analytic semigroups. One implication is shown in Proposition 9.8, while the other one will be proved by developing a simple functional calculus[2] for such operators, which—similarly to the fractional powers in Lecture 7—is based on Cauchy's integral formula.

$$\mathrm{e}^{az} = \frac{1}{2\pi\mathrm{i}} \oint \frac{\mathrm{e}^{\lambda z}}{\lambda - a} \mathrm{d}\lambda$$

[2] For a thorough treatment see: M. Haase: The Functional Calculus for Sectorial Operators, vol. 169 of Operator Theory: Advances and Applications, Birkhäuser Basel, 2006., but note first the difference between the definitions of sectorial operators here and in the mentioned monograph.

where we integrate along a curve that passes around a in the positive direction. Therefore, we want to give meaning to expressions like

$$\int_{\gamma} \mathrm{e}^{\lambda z} R(\lambda, A) \mathrm{d}\lambda,$$

where γ is a suitable curve. First, we specify these curves. For given $\eta \in (0, \delta)$ and $r > 0$ consider the curves given by the following parametrisations:

$$\begin{aligned} \gamma_{\eta,r,1}(s) &:= s\mathrm{e}^{-\mathrm{i}(\frac{\pi}{2}+\eta)},\ s \in [r, \infty) \\ \gamma_{\eta,r,2}(s) &:= r\mathrm{e}^{\mathrm{i}s},\ |s| \le \tfrac{\pi}{2} + \eta \\ \gamma_{\eta,r,3}(s) &:= s\mathrm{e}^{\mathrm{i}(\frac{\pi}{2}+\eta)},\ s \in (-\infty, -r]. \end{aligned} \tag{9.1}$$

We call then

$$\gamma_{\eta,r} := -\gamma_{\eta,r,1} + \gamma_{\eta,r,2} + \gamma_{\eta,r,3}$$

an **admissible curve**. The next remark concerns estimates that show the convergence of the path integral in the operator norm.

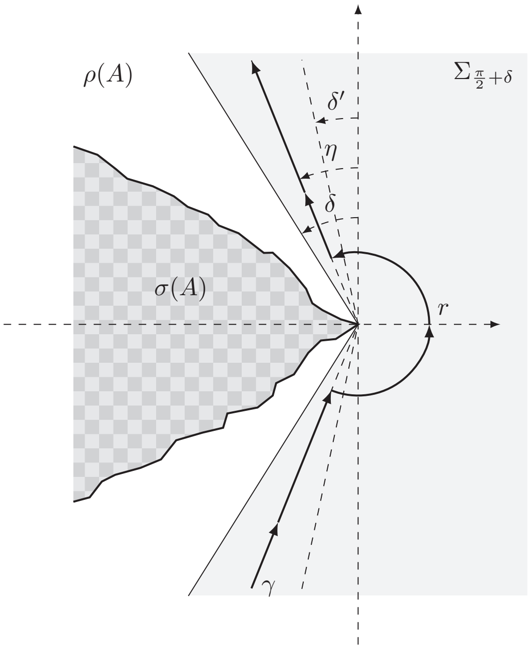

Figure 9.1: An admissible curve $\gamma_{\eta,r}$.

Remark 9.11. 1. If A is a sectorial operator of angle δ, then for every $\delta' \in (0, \delta)$ we have

$$\|R(\lambda, A)\| \le \frac{M_{\delta'}}{|\lambda|}$$

for all $\lambda \in \overline{\Sigma}_{\frac{\pi}{2}+\delta'} \setminus \{0\}$ and some appropriate $M_{\delta'} \ge 1$.

2. For every $z \in \mathbb{C}$ with $|\arg z| < \delta \leq \frac{\pi}{2}$ and $\lambda \in \mathbb{C}$ with $|\arg \lambda| = \frac{\pi}{2} + \eta$ and $\eta \in (\frac{|\arg z| + \delta}{2}, \delta)$ we have

$$\begin{aligned} & |\arg(\lambda) + \arg(z)| \leq \tfrac{\pi}{2} + \eta + \delta \leq \pi - (\delta - \eta) + \delta \leq \tfrac{3\pi}{2} - (\delta - \eta) \\ \text{and} \quad & |\arg(\lambda) + \arg(z)| \geq \tfrac{\pi}{2} + \tfrac{|\arg z| + \delta}{2} - |\arg z| = \tfrac{\pi}{2} + \tfrac{\delta - |\arg z|}{2} \geq \tfrac{\pi}{2} + (\delta - \eta). \end{aligned}$$

Hence for such λ and z we have

$$|\mathrm{e}^{\lambda z}| = \mathrm{e}^{\mathrm{Re}(\lambda z)} = \mathrm{e}^{|\lambda z| \cos(\arg(\lambda) + \arg(z))} \leq \mathrm{e}^{\cos(\frac{\pi}{2} + (\delta - \eta))|\lambda z|} = \mathrm{e}^{-\sin(\delta - \eta)|\lambda z|}.$$

Lemma 9.12. *For a sectorial operator A of angle δ and for an admissible curve $\gamma_{\eta,r}$ with $\eta \in (\frac{|\arg z| + \delta}{2}, \delta)$ and $r > 0$ the integral*

$$\int_{\gamma_{\eta,r}} \mathrm{e}^{\lambda z} R(\lambda, A) \mathrm{d}\lambda$$

converges in operator norm, and its value is independent of $r > 0$ and η.

Proof. For the convergence of the integral only $\gamma_{\eta,r,1}$ and $\gamma_{\eta,r,3}$ need to be considered. By Remark 9.11 for $\lambda \in \mathbb{C}$ with $|\arg \lambda| = \frac{\pi}{2} + \eta$ we can estimate the integrand

$$\|\mathrm{e}^{\lambda z} R(\lambda, A)\| \leq \mathrm{e}^{-\sin(\delta - \eta)|\lambda z|} \frac{M_\eta}{|\lambda|}, \tag{9.2}$$

where the right-hand side converges exponentially fast to 0 for $|\lambda| \to \infty$. This suffices for the convergence of the integral.

The independence of the integral from r and η follows from Cauchy's theorem if we close the angle between two admissible curves by circle arcs around 0 of radius R and let $R \to \infty$. The path integrals on these circle arcs converge to 0 by (9.2). □

The arguments above allow us to make the following definition.

Definition 9.13. Let A be a sectorial operator of angle δ. For $z \in \Sigma_\delta$ and some admissible curve $\gamma = \gamma_{\eta,r}$ with $\eta \in (\frac{|\arg z| + \delta}{2}, \delta)$ we define

$$T(z) = \mathrm{e}^{zA} := \frac{1}{2\pi \mathrm{i}} \int_\gamma \mathrm{e}^{\lambda z} R(\lambda, A) \mathrm{d}\lambda. \tag{9.3}$$

Clearly, this has to be the right definition. Let us check it.

Proposition 9.14. *Let A be a sectorial operator of angle δ. For $T(z)$ from Definition 9.13 the following are true:*

a) $\|T(z)\|$ is uniformly bounded for $z \in \Sigma_{\delta'}$ if $0 < \delta' < \delta$.

b) The map $z \mapsto T(z)$ is holomorphic in Σ_δ.

c) $T(z_1 + z_2) = T(z_1)T(z_2)$ for all $z_1, z_2 \in \Sigma_\delta$.

Proof. a) Let $\delta' \in (0,\delta)$. Given $z \in \Sigma_{\delta'}$ we may choose the path of integration $\gamma_{\eta,r}$ with $\eta \in (\frac{\delta'+\delta}{2},\delta)$ and $r \in (0,\frac{1}{|z|}]$. As in Lemma 9.12 we estimate the integrand: For $\lambda \in \mathbb{C}$ with $|\arg\lambda| = \frac{\pi}{2}+\eta$ and $|\lambda| \geq r$ we have

$$\|\mathrm{e}^{\lambda z}R(\lambda,A)\| \leq \mathrm{e}^{-|\lambda z|\sin(\delta-\eta)}\frac{M_\eta}{|\lambda|}, \tag{9.4}$$

and for $|\lambda| = r$ and $|\arg\lambda| \leq \frac{\pi}{2}+\eta$ we have

$$\|\mathrm{e}^{\lambda z}R(\lambda,A)\| \leq \mathrm{e}^{|\lambda z|}\frac{M_\eta}{|\lambda|} \leq \mathrm{e}^{|\lambda|\frac{1}{r}}\frac{M_\eta}{r} = \mathrm{e}\frac{M_\eta}{r}. \tag{9.5}$$

For the integral in (9.3) these yield (considering the three pieces of the integration path separately)

$$\begin{aligned}\|T(z)\| = \Big\|\frac{1}{2\pi\mathrm{i}}\int_{\gamma_{r,\eta}}\mathrm{e}^{\lambda z}R(\lambda,A)\mathrm{d}\lambda\Big\| &\leq \frac{M_\eta}{\pi}\int_r^\infty \frac{1}{s}\mathrm{e}^{-s|z|\sin(\delta-\eta)}\mathrm{d}s + r\mathrm{e}\frac{M_\eta}{r}\\ &= \frac{M_\eta}{\pi}\int_{|z|r}^\infty \frac{1}{t}\mathrm{e}^{-t\sin(\delta-\eta)}\mathrm{d}t + \mathrm{e}M_\eta.\end{aligned}$$

If specialise $r = \frac{1}{|z|}$, then we obtain

$$\|T(z)\| \leq \frac{M_\eta}{\pi}\int_1^\infty \frac{1}{t}\mathrm{e}^{-t\sin(\delta-\eta)}\mathrm{d}t + \mathrm{e}M_\eta \quad \text{for all } z \in \Sigma_{\delta'}.$$

b) Suppose $K \subseteq \Sigma_\delta$ is a compact set, and let $\delta' \in (0,\delta)$ such that $K \subseteq \Sigma_{\delta'}$ and let $0 < r \leq \inf_{z\in K}\frac{1}{|z|}$. The estimates in the proof of part a) show that the integral defining $T(z)$ converges uniformly on K. Since the integrand $z \mapsto \mathrm{e}^{\lambda z}R(\lambda,A) \in \mathscr{L}(X)$ is holomorphic, so is $T(z)$.

c) Let $z,w \in \Sigma_\delta$ and let γ and $\tilde\gamma$ be two admissible curves as in part a) so that $\tilde\gamma$ lies to the right of γ. We calculate the product

$$\begin{aligned}T(z)T(w) &= \frac{1}{(2\pi\mathrm{i})^2}\int_\gamma\int_{\tilde\gamma}\mathrm{e}^{\mu w}\mathrm{e}^{\lambda z}R(\mu,A)R(\lambda,A)\mathrm{d}\mu\mathrm{d}\lambda\\ &= \frac{1}{(2\pi\mathrm{i})^2}\int_\gamma\int_{\tilde\gamma}\frac{\mathrm{e}^{\mu w}\mathrm{e}^{\lambda z}}{\lambda-\mu}\left(R(\mu,A)-R(\lambda,A)\right)\mathrm{d}\mu\mathrm{d}\lambda\end{aligned}$$

by the resolvent identity. Fubini's theorem yields

$$T(z)T(w) = \frac{1}{2\pi\mathrm{i}}\int_{\tilde\gamma}\mathrm{e}^{\mu w}R(\mu,A)\Big(\frac{1}{2\pi\mathrm{i}}\int_\gamma\frac{\mathrm{e}^{\lambda z}}{\lambda-\mu}\mathrm{d}\lambda\Big)\mathrm{d}\mu - \frac{1}{2\pi\mathrm{i}}\int_\gamma\mathrm{e}^{\lambda z}R(\lambda,A)\Big(\frac{1}{2\pi\mathrm{i}}\int_{\tilde\gamma}\frac{\mathrm{e}^{\mu w}}{\lambda-\mu}\mathrm{d}\mu\Big)\mathrm{d}\lambda.$$

Since $\tilde\gamma$ lies to the right of γ, we have

$$\frac{1}{2\pi\mathrm{i}}\int_{\tilde\gamma}\frac{\mathrm{e}^{\mu w}}{\lambda-\mu}\mathrm{d}\lambda = 0 \quad \text{and} \quad \frac{1}{2\pi\mathrm{i}}\int_\gamma\frac{\mathrm{e}^{\lambda z}}{\lambda-\mu}\mathrm{d}\lambda = \mathrm{e}^{\mu z}.$$

Altogether we conclude

$$T(z)T(w) = \frac{1}{2\pi \mathrm{i}} \int\limits_{\tilde{\gamma}} \mathrm{e}^{\mu z} \mathrm{e}^{\mu w} R(\mu, A) \mathrm{d}\mu = T(z+w). \qquad \square$$

Summarising, given a sectorial operator A, have seen how to construct an analytic semigroup. It will be no surprise to identify its generator.

Proposition 9.15. *Let A be a densely defined sectorial operator of angle δ. Then T given by*

$$T(z) := \mathrm{e}^{zA} := \frac{1}{2\pi \mathrm{i}} \int\limits_{\gamma} \mathrm{e}^{\lambda z} R(\lambda, A) \mathrm{d}\lambda$$

as in Definition 9.13 is a bounded analytic semigroup of angle δ, whose generator is A.

Proof. We only have to prove property (iii) from Definition 9.1. Let us fix $\delta' \in (0, \delta)$, and notice that

$$\frac{1}{2\pi \mathrm{i}} \int\limits_{\gamma} \frac{\mathrm{e}^{\lambda z}}{\lambda} \mathrm{d}\lambda = 1$$

holds for all $z \in \Sigma_{\delta'}$ and for an admissible curve $\gamma = \gamma_{\eta,r}$. For $f \in D(A)$ we have $R(\lambda, A)Af = \lambda R(\lambda, A)f - f$, and hence

$$T(z)f - f = \frac{1}{2\pi \mathrm{i}} \int\limits_{\gamma} \mathrm{e}^{\lambda z} \big(R(\lambda, A) - \frac{1}{\lambda} f\big) f \mathrm{d}\lambda = \frac{1}{2\pi \mathrm{i}} \int\limits_{\gamma} \frac{\mathrm{e}^{\lambda z}}{\lambda} R(\lambda, A) A f \mathrm{d}\lambda$$

for all $z \in \Sigma_{\delta'}$. For $z \to 0$ ($z \in \Sigma_{\delta'}$) we conclude

$$\lim_{\substack{z \to 0 \\ z \in \Sigma_{\delta'}}} \big(T(z)f - f\big) = \frac{1}{2\pi \mathrm{i}} \int\limits_{\gamma} \frac{1}{\lambda} R(\lambda, A) A f \mathrm{d}\lambda,$$

where the passage to the limit is allowed by Lebesgue's dominated convergence theorem. Indeed, we can estimate the integrand by means of inequalities in (9.4) and (9.5):

$$\Big\| \frac{\mathrm{e}^{\lambda z}}{\lambda} R(\lambda, A) A f \Big\| \leq \frac{M_\eta}{|\lambda|^2} \big(1 + \mathrm{e}^{|z|}\big) \|Af\|$$

for all λ that lies on the curve γ.

By Cauchy's theorem we obtain

$$\frac{1}{2\pi \mathrm{i}} \int\limits_{\gamma} \frac{1}{\lambda} R(\lambda, A) A f \mathrm{d}\lambda = 0,$$

which can be seen if we close γ on the right by circle arcs around 0 of radius R and let $R \to \infty$. The integrals on these arcs converge to 0 since A is sectorial. Since we already know that $T(z)$ is uniformly bounded on $\Sigma_{\delta'}$, see Proposition 9.14, we conclude by Theorem 2.30 that

$$\lim_{\substack{z \to 0 \\ z \in \Sigma_{\delta'}}} T(z)f = f \quad \text{for all } f \in X.$$

Therefore T is a bounded analytic semigroup.

Denote for the moment the generator of T by B. Since T is bounded, by Proposition 2.26 we have

$$R(2,B)f = \int\limits_0^\infty \mathrm{e}^{-2t} T(t) f \mathrm{d}t \qquad \text{for all } f \in X.$$

For a fixed $s > 0$ and an admissible curve $\gamma = \gamma_{\eta,1}$ we can write by Fubini's theorem

$$\begin{aligned}\int\limits_0^s \mathrm{e}^{-2t} T(t) f \mathrm{d}t &= \int\limits_0^s \mathrm{e}^{-2t} \frac{1}{2\pi\mathrm{i}} \int\limits_\gamma \mathrm{e}^{\lambda t} R(\lambda, A) f \mathrm{d}\lambda \mathrm{d}t = \frac{1}{2\pi\mathrm{i}} \int\limits_\gamma \int\limits_0^s \mathrm{e}^{(\lambda-2)t} \mathrm{d}t R(\lambda, A) f \mathrm{d}\lambda \\ &= \frac{1}{2\pi\mathrm{i}} \int\limits_\gamma \frac{\mathrm{e}^{s(\lambda-2)} - 1}{\lambda - 2} R(\lambda, A) f \mathrm{d}\lambda \\ &= \frac{1}{2\pi\mathrm{i}} \int\limits_\gamma \frac{\mathrm{e}^{s(\lambda-2)}}{\lambda - 2} R(\lambda, A) f \mathrm{d}\lambda - \frac{1}{2\pi\mathrm{i}} \int\limits_\gamma \frac{R(\lambda, A) f}{\lambda - 2} \mathrm{d}\lambda.\end{aligned}$$

For $s \to \infty$ the first expression converges to 0 since $\mathrm{Re}(\lambda - 2) \leq -1$ for all λ on the curve γ. For the second term we have

$$-\frac{1}{2\pi\mathrm{i}} \int\limits_\gamma \frac{R(\lambda, A) f}{\lambda - 2} \mathrm{d}\lambda = R(2, A) f \quad \text{for all } f \in X.$$

These yield $A = B$. □

Let us summarise what we have proved so far:

Corollary 9.16. *For a densely defined linear operator A on a Banach space X the following are equivalent:*

(i) A is sectorial of angle δ.

(ii) A generates a bounded holomorphic semigroup of angle δ.

9.3 Further characterisations

Analytic semigroups have some fundamental properties needed in calculations and estimates. In this section we investigate the most important properties and develop some other characterisations of generators of analytic semigroups.

Proposition 9.17. *A generator A generates a bounded analytic semigroup if and only if* $\operatorname{ran} T(t) \subseteq D(A)$ *for all $t > 0$ and*

$$\sup_{t>0} \bigl\| tAT(t) \bigr\| < \infty.$$

Proof. Suppose first that A generates a bounded analytic semigroup of angle θ and let $\theta' \in (0, \theta)$. Then by Cauchy's integral formula we have

$$T'(t) = \frac{1}{2\pi \mathrm{i}} \int\limits_{\gamma} \frac{T(z)}{(z-t)^2} \mathrm{d}z,$$

where γ is a circle of radius $r = t\sin(\theta')$ around $t > 0$. From this we conclude

$$\|AT(t)\| = \|T'(t)\| \le \frac{2\pi r}{r^2} \sup_{z \in \theta'} \|T(z)\| \le \frac{2\pi}{t\sin(\theta')} \sup_{z \in \Sigma'_{\theta}} \|T(z)\| \quad \text{for all } t > 0,$$

and that was to be proved.

Conversely suppose that A is the generator of a semigroup T, $\operatorname{ran} T(t) \subseteq D(A)$ for $t > 0$ and $M := \sup_{t>0}\{\|T(t)\|,\ \|tAT(t)\|\} < \infty$. The basic idea is to define the analytic extension by the Taylor series as

$$\sum_{n=0}^{\infty} \frac{(z-t)^n}{n!} \frac{\mathrm{d}^n}{\mathrm{d}t^n} T(t)f.$$

The next step of the proof is now to verify that this definition does indeed make sense and yields an analytic semigroup. By assumption $AT(t) \in \mathscr{L}(X)$ for $t > 0$, hence $A^nT(t) = A^nT^n(t/n) = (AT(t/n))^n \in \mathscr{L}(X)$ and we can write

$$\Big\|\frac{A^nT(t)}{n!}\Big\| = \Big\|\frac{(AT(t/n))^n}{n!}\Big\| \le \frac{n^nM^n}{t^n n!} \le \Big(\frac{M\mathrm{e}}{t}\Big)^n. \tag{9.6}$$

Writing up Taylor's formula with remainder R_n in the integral form we have

$$T(s)f = \sum_{k=0}^{n} \frac{(s-t)^k}{k!} \frac{\mathrm{d}^k}{\mathrm{d}t^k} T(t)f + R_n(t,s)f$$

and

$$R_n(t,s)f = \frac{1}{n!} \int\limits_t^s (s-r)^n \frac{\mathrm{d}^{n+1}}{\mathrm{d}t^{n+1}} T(r)f \mathrm{d}r.$$

By (9.6) we obtain that the series

$$\widetilde{T}(z)f := \sum_{n=0}^{\infty} \frac{(z-t)^n}{n!} \frac{\mathrm{d}^n}{\mathrm{d}t^n} T(t)f$$

is absolutely and uniformly convergent in $\mathscr{L}(X)$ for all $z \in \mathbb{C}$ with $|z - t| \le q \cdot \frac{t}{\mathrm{e}M}$ where $q \in (0,1)$, and that $R_n(t,s) \to 0$ for all $s > 0$ with $|s - t| \le q \cdot \frac{t}{\mathrm{e}M}$. These yield that $\widetilde{T}(t) = T(t)$ for all $t > 0$ and that $\widetilde{T}$ is analytic on the sector Σ_θ with $\theta = \arcsin \frac{1}{\mathrm{e}M}$ and is uniformly bounded on the sectors $\Sigma_{\theta'}$ with $\theta' \in (0, \theta)$. □

Proposition 9.18. *A densely defined linear operator A generates a bounded analytic semigroup if and only if*

$$\Sigma_{\frac{\pi}{2}} = \big\{\lambda \in \mathbb{C} : \operatorname{Re}\lambda > 0\big\} \subseteq \rho(A)$$

and

$$\sup_{\operatorname{Re}\lambda > 0} \|\lambda R(\lambda, A)\| < \infty.$$

Proof. That a generator of a bounded analytic semigroup has the asserted properties follows from Proposition 9.8. For the converse implication notice that the assumptions are almost as in the definition of sectoriality, except we have here the sector $\Sigma_{\frac{\pi}{2}}$. To gain a larger sector one can argue similarly to the proof of Proposition 7.5. □

Proposition 9.19. *A linear operator A generates a bounded analytic semigroup if and only if for some $\alpha \in (0, \frac{\pi}{2})$ both of the operators $\mathrm{e}^{-\mathrm{i}\alpha}A$ and $\mathrm{e}^{\mathrm{i}\alpha}A$ generate bounded strongly continuous semigroups.*

Proof. One implication is already proved in Proposition 9.3.c). For the converse suppose $\mathrm{e}^{-\mathrm{i}\alpha}A$ and $\mathrm{e}^{\mathrm{i}\alpha}A$ generate bounded strongly continuous semigroups. By Proposition 2.26 we have for $\lambda \in \mathbb{C}$ with $\operatorname{Re}\lambda > 0$

$$\|R(\lambda, A)\| = \|R(\mathrm{e}^{\mathrm{i}\alpha}\lambda, \mathrm{e}^{\mathrm{i}\alpha}A)\| \leq \frac{M}{\operatorname{Re}(\mathrm{e}^{\mathrm{i}\alpha}\lambda)} = \frac{M}{\operatorname{Re}\lambda \cdot \cos(\alpha) - \operatorname{Im}\lambda \cdot \sin(\alpha)} \qquad \text{if } \operatorname{Im}\lambda \leq 0,$$

$$\|R(\lambda, A)\| = \|R(\mathrm{e}^{-\mathrm{i}\alpha}\lambda, \mathrm{e}^{-\mathrm{i}\alpha}A)\| \leq \frac{M}{\operatorname{Re}(\mathrm{e}^{-\mathrm{i}\alpha}\lambda)} = \frac{M}{\operatorname{Re}\lambda \cdot \cos(\alpha) + \operatorname{Im}\lambda \cdot \sin(\alpha)} \qquad \text{if } \operatorname{Im}\lambda > 0.$$

So altogether we obtain

$$\|R(\lambda, A)\| \leq \frac{M}{\operatorname{Re}\lambda \cdot \cos(\alpha)} = \frac{M}{\cos^2(\alpha) \cdot |\lambda|},$$

so by Proposition 9.18 the proof is complete. □

About generators of not necessarily bounded analytic semigroups we can say the following.

Proposition 9.20. *For a densely defined linear operator A on the Banach space X the following assertions are equivalent:*

(i) *The operator A generates an analytic semigroup (of some angle).*

(ii) *For some $\omega > 0$ the operator $A - \omega$ generates a bounded analytic semigroup (of some angle).*

(iii) *There is $r > 0$ such that*

$$\big\{\lambda \in \mathbb{C} : \operatorname{Re}\lambda > 0,\ |\lambda| > r\big\} \subseteq \rho(A)$$

and

$$\sup_{\substack{\operatorname{Re}\lambda > 0\\ |\lambda| > r}} \|\lambda R(\lambda, A)\| < \infty.$$

The proof of this assertion is left as Exercise 5.

9.4 Intermediate spaces

In this section we study the intermediate spaces—introduced in Lecture 7 and 8—for analytic semigroups. The first result is yet another characterisation of the Favard and Hölder spaces.

Proposition 9.21. *Let T be an analytic semigroup of type (M, ω) with $\omega < 0$ and with generator A. For $\alpha \in (0, 1]$ one has*

$$F_\alpha = \big\{f \in X : \sup_{t>0} \big\|t^{1-\alpha}AT(t)f\big\| < \infty\big\}$$

with equivalent norm

$$[\![f]\!]_{F_\alpha} := \sup_{t>0} \|t^{1-\alpha} AT(t)f\|.$$

For $\alpha \in (0,1)$ we have

$$X_\alpha = \big\{ f \in X : \lim_{t\to 0} \|t^{1-\alpha} AT(t)f\big\| < \infty \big\}.$$

Proof. It is easy to see that $[\![\cdot]\!]_{F_\alpha}$ is a norm. For every $f \in X$ and $t > 0$ we have

$$t^{1-\alpha}AT(t)f = t^{-\alpha}AT(t)\int_0^t f\mathrm{d}s \quad \text{and} \quad t^{-\alpha}T(t)(T(t)f - f) = t^{-\alpha}T(t)A\int_0^t T(s)f\mathrm{d}s,$$

hence $t^{1-\alpha}AT(t)f = t^{-\alpha}T(t)\big(T(t)f - f\big) - t^{-\alpha}AT(t)\displaystyle\int_0^t (T(s)f - f)\mathrm{d}s.$

If $f \in F_\alpha$, then we obtain for $t > 0$ that

$$\begin{aligned} \|t^{1-\alpha}AT(t)f\| &\le Mt^{-\alpha}\|T(t)f - f\| + \|t^{-\alpha}AT(t)\| \int_0^t s^\alpha \frac{1}{s^\alpha}\|T(s)f - f\|\mathrm{d}s \qquad (9.7)\\ &\le M\|f\|_{F_\alpha} + \|t^{-\alpha}AT(t)\| \int_0^t s^\alpha \frac{1}{s^\alpha}\|T(s)f - f\|\mathrm{d}s \\ &\le M\|f\|_{F_\alpha} + \frac{t}{\alpha+1}\|AT(t)\| \cdot \|f\|_{F_\alpha} \le M_1\|f\|_{F_\alpha}. \end{aligned}$$

Therefore, one inclusion in the first assertion is proved together with the estimate $[\![f]\!]_{F_\alpha} \le M_1\|f\|_{F_\alpha}$. Suppose now that $f \in X$ is such that $[\![f]\!]_{F_\alpha} < \infty$. Then we have for $t > 0$ that the integral on the left-hand side below converges, and we also obtain the equality

$$\int_0^t AT(s)f\mathrm{d}s = A\int_0^t T(s)f\mathrm{d}s.$$

From this can conclude

$$\begin{aligned} \frac{1}{t^\alpha}\|T(t)f - f\| &= \frac{1}{t^\alpha}\Big\|\int_0^t AT(s)f\mathrm{d}s\Big\| = \frac{1}{t^\alpha}\int_0^t s^{\alpha-1}s^{1-\alpha}\|AT(s)f\|\mathrm{d}s \qquad (9.8)\\ &\le \frac{1}{\alpha}[\![f]\!]_{F_\alpha}. \end{aligned}$$

This completes the proof of the statement concerning F_α.

For $f \in X_\alpha$ we obtain by using (9.7) that $t^{1-\alpha}AT(t)f \to 0$ as $t \searrow 0$. Whereas (9.8) implies the converse implication. □

In Lecture 8 we related the domain of fractional powers to the abstract Hölder and Fravard spaces. As an immediate consequence we obtain the next fundamental result.

Corollary 9.22. *Let A generate a bounded analytic semigroup of type (M, ω) with $\omega < 0$, and let $\alpha \geq 0$. Then the following assertions are true:*

a) For each $t \geq 0$ the operator $T(t)$ maps X into $D\big((-A)^\alpha\big)$.

b) For each $t > 0$ the operator $(-A)^\alpha T(t)$ is bounded, and

$$\|(-A)^\alpha T(t)\| \leq M_\alpha t^{-\alpha} \quad \textit{holds for all } t > 0.$$

c) If $\alpha \in (0,1]$ and $f \in D\big((-A)^\alpha\big)$, then

$$\|t^{1-\alpha} AT(t)f\| \leq M_\alpha \|(-A)^\alpha f\| \quad \textit{for all } t > 0.$$

d) If $\alpha \in (0,1]$ and $f \in D\big((-A)^\alpha\big)$, then

$$\|T(t)f - f\| \leq K_\alpha t^\alpha \|(-A^\alpha) f\| \quad \textit{for all } t > 0.$$

Proof. a) For each $t > 0$ the operator $T(t)$ even maps into $D(A^n)$ for all $n \in \mathbb{N}$, see proof of Proposition 9.17.

b) The statement is trivially true for $\alpha = 0$, while for $\alpha = 1$ it follows from Proposition 9.17. By Remark 7.10 we have

$$\|(-A)^\alpha f\| \leq K \|f\|^{1-\alpha} \|Af\|^\alpha \quad \text{for all } f \in D(A),$$

whence we can conclude by Proposition 9.17

$$\|(-A)^\alpha T(t) f\| \leq K \|T(t)f\|^{1-\alpha} \|AT(t)f\|^\alpha \leq \frac{M_\alpha}{t^\alpha} \|f\|.$$

Suppose now $\alpha > 1$. For $\alpha \in \mathbb{N}$ the assertion follows again from Proposition 9.17: For $t > 0$ we have

$$\big\|A^n T(t)\big\| = \big\|(AT(\tfrac{t}{n}))^n\big\| \leq \big\|AT(\tfrac{t}{n})\big\|^n \leq \frac{n^n M^n}{t^n}.$$

Suppose $\alpha \geq 1$. Then we can write $\alpha = n + \alpha'$ with $n \in \mathbb{N}$ and $\alpha' \in (0,1]$. From the above we can conclude

$$\|(-A)^\alpha T(t)\| = \|(-A)^{\alpha'} T(\tfrac{t}{2})(-A)^n T(\tfrac{t}{2})\| \leq \frac{2^{\alpha'} M_{\alpha'}}{t^{\alpha'}} \|(-A)^n T(\tfrac{t}{2})\| \leq \frac{2^{\alpha'} M_{\alpha'} 2^n M_n}{t^{n+\alpha'}} = \frac{M_\alpha}{t^\alpha}$$

for all $t > 0$.

c) By Theorem 8.20 we have the continuous embedding $D\big((-A)^\alpha\big) \hookrightarrow F_\alpha$. In view of Proposition 9.21 the asserted inequality is just a reformulation of this.

d) For $\alpha \in (0,1)$ the statement is just the reformulation of the continuous embedding $D\big((-A)^\alpha\big) \hookrightarrow X_\alpha$, which we proved in Theorem 8.20. Suppose $\alpha = 1$, and let $f \in D(A)$. Then we have

$$\|T(t)f - f\| = \Big\|A \int_0^t T(s) f \,\mathrm{d}s\Big\| \leq Kt \|Af\|. \qquad \square$$

Before stating to Proposition 9.21 analogous characterisation of $(X, D(A))_{\alpha,p}$ spaces, we need to recall[3] the following result.

Proposition 9.23 (Hardy's inequality)**.** *Let $f : (0,\infty) \to \mathbb{R}$ be a positive Lebesgue measurable function, let $\alpha > 0$, and let $p \in [1,\infty)$. Then*

$$\int_0^\infty t^{-\alpha p}\Big(\int_0^t f(s)^p \frac{\mathrm{d}s}{s}\Big)^p \frac{\mathrm{d}t}{t} \leq \frac{1}{\alpha^p}\int_0^\infty t^{-\alpha p} f(t)^p \frac{\mathrm{d}t}{t}.$$

Proposition 9.24. *Let T be an analytic semigroup of type (M,ω) with $\omega < 0$ and with generator A. For $\alpha \in (0,1]$ one has*

$$\big(X, D(A)\big)_{\alpha,p} = \big\{f \in X : t \mapsto \eta(t) = \|t^{1-\alpha} AT(t)f\big\| \in \mathrm{L}^p_*(0,\infty)\big\}$$

with equivalent norm

$$\llbracket f \rrbracket_{(X,D(A))_{\alpha,p}} := \|f\| + \|\eta\|_{\mathrm{L}^p_*(0,\infty)}.$$

Proof. Suppose $f \in \big(X, D(A)\big)_{\alpha,p}$ holds. By (9.7) it suffices to estimate

$$\int_0^\infty \|t^{-\alpha}AT(t)\|^p\Big(\int_0^t \|T(s)f - f\|\mathrm{d}s\Big)^p \frac{\mathrm{d}t}{t}.$$

By Hardy's inequality (see Proposition 9.23) and by Proposition 9.17 we have that

$$\begin{aligned}\int_0^\infty \|t^{-\alpha}AT(t)\|^p\Big(\int_0^t \|T(s)f - f\|\mathrm{d}s\Big)^p \frac{\mathrm{d}t}{t} &\leq M\int_0^\infty t^{-(\alpha+1)p}\Big(\int_0^t s\|T(s)f - f\|\frac{\mathrm{d}s}{s}\Big)^p\frac{\mathrm{d}t}{t}\\ \leq \frac{1}{(\alpha+1)^p}\int_0^\infty s^{-(\alpha+1)p}s^p\|T(s)f-f\|^p\frac{\mathrm{d}s}{s} &= \frac{1}{(\alpha+1)^p}\int_0^\infty s^{-\alpha p}\|T(s)f - f\|^p\frac{\mathrm{d}s}{s}.\end{aligned}$$

Therefore $\llbracket f \rrbracket_{(X,D(A))_{\alpha,p}} \leq M_1\|f\|_{(X,D(A))_{\alpha,p}}$.

Conversely, suppose that $f \in X$ is such that $\eta \in \mathrm{L}^p_*(0,\infty)$. Then again by Hardy's inequality we obtain

$$\begin{aligned}\|f\|^p_{(X,D(A))_{\alpha,p}} &= \int_0^\infty \frac{1}{t^{p\alpha}}\|T(t)f - f\|^p\frac{\mathrm{d}t}{t} \leq \int_0^\infty \frac{1}{t^{p\alpha}}\Big\|\int_0^t AT(s)f\mathrm{d}s\Big\|^p\frac{\mathrm{d}t}{t} = \int_0^\infty \frac{1}{t^{p\alpha}}\Big\|\int_0^t sAT(s)f\frac{\mathrm{d}s}{s}\Big\|^p\frac{\mathrm{d}t}{t}\\ &\leq \frac{1}{\alpha^p}\int_0^\infty s^{-\alpha p}s^p\|AT(s)f\|^p\frac{\mathrm{d}s}{s} = \frac{1}{\alpha^p}\|\eta\|^p_{\mathrm{L}^p_*(0,\infty)}.\end{aligned}$$

□

[3] See, e.g., page 158 of D. J. H. Garling: Inequalities. A Journey into Linear Analysis, Cambridge University Press, 2007.

Exercises

1. Work out the details of Example 9.4.

2. Show that T defined in Example 9.6 is a bounded analytic semigroup.

3. Prove the assertions in Example 9.7.

4. Let X, Y be Banach spaces. Show that if A is a sectorial operator on X and $S : X \to Y$ is continuously invertible then $STS^{-1}T$ is a sectorial operator on Y.

5. Prove Proposition 9.20.

6. Suppose that A generates an analytic semigroup and that $B \in \mathscr{L}(X)$. Prove that $A+B$ generates an analytic semigroup.

Lecture 10

Operator Splitting

In many applications the combined effect of several physical (or chemical, etc.) phenomena is modelled. In these cases one has to solve a partial differential equation where the local time derivative of the modelled physical quantity equals the sum of several operators, describing how this quantity behaves in space. The idea behind operator splitting procedures is that, instead of the sum, we treat each spatial operator separately, i.e., we solve all the corresponding sub-problems. The solution of the original problem is then obtained from the solutions of the sub-problems. Since the sum usually contains operators of different nature, each corresponding to one physical phenomenon, the sub-problems may be easier to solve separately.

Consider the abstract Cauchy problem on the Banach space X of the form

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}u(t) = (A+B)u(t), & t > 0 \\ \quad u(0) = u_0 \end{cases} \tag{10.1}$$

with densely defined, closed, and linear operators A and B. Throughout the lecture we suppose that $D := D(A) \cap D(B)$ is dense in X and $u_0 \in D$.

As an example, we explain how the simplest operator splitting procedure works. The main idea is to choose sub-problems which are easier to handle as the original problem. This can happen if there are particularly well-suited (fast, accurate, reliable, etc.) numerical methods to solve the sub-problems, or if the exact (analytical) solution of at least one of the sub-problems is known. First, one solves the sub-problem corresponding to the operator A on the time interval $[0, h]$ using the original initial value u_0. Then the second sub-problem, corresponding to the operator B, is solved on the same time interval but using the solution of the previous step as initial value. In the next step the sub-problems with A and B are solved on the next time interval $[h, 2h]$, always taking the previous solution as initial value. We repeat this procedure recursively. The corresponding sub-problems can be formulated for $t \in ((k-1)h, kh]$ with $k \in \mathbb{N}$ and $u_{\mathrm{spl},h}(0) = u_0$ as follows:

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}u_A(t) = Au_A(t) \\ u_A((k-1)h) = u_{\mathrm{spl},h}((k-1)h) \end{cases} \quad \text{and} \quad \begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}u_B(t) = Bu_B(t) \\ u_B((k-1)h) = u_A(kh) \end{cases}$$

and we set $u_{\mathrm{spl},h}(kh) = u_B(kh)$.

Operator splittings can be mathematically handled in the same way as finite difference schemes, which were introduced in Definition 4.1, with the help of a strongly continuous function $F : [0, \infty) \to \mathscr{L}(X)$. By applying operator splitting, one computes the numerical solution at time $t > 0$ (more precisely, a sequence of numerical solutions) to problem (10.1) of the form

$$u_{\mathrm{spl},h}(t) = \big(F(\tfrac{t}{n})\big)^n u_0$$

with $h = \frac{t}{n}$ and $n \in \mathbb{N}$. Our aim is to establish splitting procedures being convergent in the sense

$$u(t) = \lim_{n\to\infty} u_{\mathrm{spl},h}(t) = \lim_{n\to\infty} \big(F(\tfrac{t}{n})\big)^n u_0$$

for all (or at least for many) $u_0 \in D$, for all $t \geq 0$. As usual, we set $nh = t$.

There are several splitting procedures in the literature. We collect here the most important ones which are often used in applications as well.

Definition 10.1. The **split solution** to problem (10.1) at time $t > 0$ is defined by

$$u_{\mathrm{spl},h}(t) = \left(F(h)\right)^n u_0$$

for all $u_0 \in D$ and $n \in \mathbb{N}$ with $nh = t$. For the different operator splitting procedures, the strongly continuous function $F : [0, \infty) \to \mathscr{L}(X)$ is formally defined for all $h > 0$ as

Sequential splitting:	$F_{\mathrm{seq}}(h) = \mathrm{e}^{hB}\mathrm{e}^{hA}$,
Marchuk–Strang splitting:	$F_{\mathrm{MS}}(h) = \mathrm{e}^{\frac{h}{2}A}\mathrm{e}^{hB}\mathrm{e}^{\frac{h}{2}A}$,
Lie splitting:	$F_{\mathrm{Lie}}(h) = \bigl(I - hB\bigr)^{-1}\bigl(I - hA\bigr)^{-1}$,
Peaceman–Rachford splitting:	$F_{\mathrm{PR}}(h) = \bigl(I - \frac{h}{2}A\bigr)^{-1}\bigl(I + \frac{h}{2}B\bigr)\bigl(I - \frac{h}{2}B\bigr)^{-1}\bigl(I + \frac{h}{2}A\bigr)$.

Of course, due to the application of operator splitting, there appears a certain **splitting error** in the split solution.

In this chapter we show the convergence of the operator splitting procedures defined above, and we also derive some error bounds for their **local error**

$$\mathscr{E}_{u_0}\bigl(h, u(t)\bigr) = \mathscr{E}\bigl(h, u(t)\bigr) := \|F(h)u(t) - u(t+h)\|,$$

where the exact solution of problem (10.1) is $u(t) = \mathrm{e}^{t(A+B)}u_0$. Hence, the local error can be further rewritten as

$$\mathscr{E}\bigl(h, u(t)\bigr) = \bigl\|F(h)u(t) - \mathrm{e}^{(t+h)(A+B)}u_0\bigr\| = \bigl\|F(h)u(t) - \mathrm{e}^{h(A+B)}u(t)\bigr\|.$$

In order to obtain convergence and some error bounds, we have to investigate the difference $F(h) - \mathrm{e}^{h(A+B)}$. For the sequential splitting[1] this means $\mathrm{e}^{hB}\mathrm{e}^{hA} - \mathrm{e}^{h(A+B)}$ which in general differs from zero (unless $A, B \in \mathbb{C}$).

For simplicity we start with A and B being matrices.

10.1 Matrix case

Consider the problem (10.1) with operators $A, B \in \mathscr{L}(\mathbb{C}^d)$, which corresponds to a linear system of ordinary differential equations.
The exponential function of the matrices A, B can be formulated as power series, and the local error is then defined as

$$\mathscr{E}\bigl(h, u(t)\bigr) = \bigl\|\bigl(F(h) - \mathrm{e}^{h(A+B)}\bigr)u(t)\bigr\|.$$

We start with the investigation of the sequential splitting.

Proposition 10.2. *Consider the operators $A, B \in \mathscr{L}(\mathbb{C}^d)$. If they commute, i.e. $[A, B] := AB - BA = 0$, then the local error of the sequential splitting vanishes.*

[1] Sometimes is called as Lie–Trotter product formula, see Theorem 10.6.

The proof is left as Exercise 2.

In the next step we will prove the convergence of the sequential splitting for general, non-commuting operators.

Theorem 10.3. *The sequential splitting is first-order convergent for $A, B \in \mathscr{L}(\mathbb{C}^d)$.*

Proof. By the Lax equivalence theorem, Theorem 4.6, it is sufficient to show consistency and stability from Definition 4.1, for $t < t_0$. To prove the consistency, we first consider the local error

$$\begin{aligned}\mathscr{E}_{\mathrm{seq}}(t,h) &= \|F(h)u(h) - u(t+h)\| = \big\|\mathrm{e}^{hB}\mathrm{e}^{hA}u(t) - \mathrm{e}^{h(A+B)}u(t)\big\| \\ &\le \big\|\mathrm{e}^{hB}\mathrm{e}^{hA} - \mathrm{e}^{h(A+B)}\big\| \cdot \|u(t)\|.\end{aligned}$$

The local error can be expressed by the power series of the corresponding exponential functions (see also Exercise 1):

$$\begin{aligned}\mathscr{E}_{\mathrm{seq}}(t,h) &\le \|(I + hB + \tfrac{h^2}{2}B^2 + \dots)(I + hA + \tfrac{h^2}{2}A^2 + \dots) \\ &\quad - (I + h(A+B) + \tfrac{h^2}{2}(A+B)^2 + \dots)\| \cdot \|u(t)\| \\ &= \tfrac{h^2}{2}\|(AB - BA) + \dots\| \cdot \|u(t)\| \le \tfrac{h^2}{2}\|[A,B]\| \cdot \|u(t)\| + \mathcal{O}(h^3). \qquad (10.2)\end{aligned}$$

From this estimate consistency follows, since $\frac{1}{h}\mathscr{E}\big(h, u(t)\big) \to 0$ as $h \searrow 0$. We note that from the considerations above, we also conclude that the sequential splitting is of first order.

To show stability, we have to ensure the existence of a constant $M > 0$ such that $\big\|F_{\mathrm{seq}}(\frac{t}{n})^n\big\| \le M$ for all fixed $t < t_0$ and for all $n \in \mathbb{N}$. Since $t < t_0$, the boundedness of A and B implies that

$$\begin{aligned}\big\|F_{\mathrm{seq}}(\tfrac{t}{n})^n\big\| &\le \big\|F_{\mathrm{seq}}(\tfrac{t}{n})\big\|^n = \big\|\mathrm{e}^{\frac{t}{n}B}\mathrm{e}^{\frac{t}{n}A}\big\|^n \le \big\|\mathrm{e}^{\frac{t}{n}B}\big\|^n \cdot \big\|\mathrm{e}^{\frac{t}{n}A}\big\|^n \\ &\le \big(\mathrm{e}^{\frac{t}{n}\|B\|}\big)^n \cdot \big(\mathrm{e}^{\frac{t}{n}\|A\|}\big)^n = \mathrm{e}^{t\|B\|}\mathrm{e}^{t\|A\|} \le \mathrm{e}^{t_0\|B\|}\mathrm{e}^{t_0\|A\|} = \mathrm{e}^{t_0(\|A\|+\|B\|)} \le M. \qquad \square\end{aligned}$$

One can investigate the convergence of the Marchuk–Strang splitting analogously.

Proposition 10.4. *The Marchuk–Strang splitting is of second order for $A, B \in \mathscr{L}(\mathbb{C}^d)$.*

The proof is left as Exercise 3.

We see that in general the behaviour of the commutator of the operators A and B is of enormous importance for these results and also for the investigation of higher order splitting formulae. Here the Baker–Campbell–Hausdorff formula comes to help.[2]

Theorem 10.5 (Baker–Campbell–Hausdorff Formula). *For $A, B \in \mathscr{L}(\mathbb{C}^d)$ and $h \in \mathbb{R}$ we have*

$$\mathrm{e}^{hB}\mathrm{e}^{hA} = \mathrm{e}^{h(A+B)+\Phi(A,B)}$$

with

$$\Phi(A,B) = \tfrac{h^2}{2}[A,B] + \tfrac{h^3}{12}[A-B,[A,B]] - \tfrac{h^4}{24}[B,[A,[A,B]]] + \dots$$

where $[A,B] = AB - BA$ denotes the commutator of A and B (appearing in all terms of the infinite sum above).

For a thorough investigation of operator splitting procedures in the context of matrices, we refer to the monograph by Faragó and Havasi,[3] or to the above cited monograph by Hairer, Lubich and Wanner.

Note that all the above considerations only make sense if $h\|A\|$ and $h\|B\|$ are small, otherwise the large constants will make the convergence very slow. This means that we need other approaches for unbounded operators, or even for matrices coming from discretisation of unbounded operators. One possible approach is presented in the following section.

[2] E. Hairer, Ch. Lubich, and G. Wanner, Geometric Numerical Integration, Springer-Verlag, 2008, Chapter III. 4.

[3] I. Faragó, Á. Havasi, Operator Splittings and their Applications, Mathematics Research Developments, Nova Science Publishers, New York, 2009.

10.2 Exponential splittings

In this section we suppose that the operators A and B are the generators of strongly continuous semigroups on the Banach space X. For the sake of better understanding (since there will be more operators and the corresponding semigroups), we will use the notation $(\mathrm{e}^{tA})_{t\geq 0}$ for the semigroup generated by the operator A, and similarly for the other operators. Therefore, the splittings schemes have the same forms as in Definition 10.1. First we investigate the sequential and Marchuk–Strang splittings.

Under the condition that the operator $A+B$ with the domain $D := D(A)\cap D(B)$ is the generator of a strongly continuous semigroup, the convergence of the sequential splitting was already shown in Corollary 4.10. Since the proof is essentially the same if $A+B$ is not a generator, but its closure, we only state the theorem here. Note that the assertion follows from Chernoff's Theorem, Theorem **??**, applied to the operator F_{seq} defined above (see Exercise 4 as well).

Theorem 10.6 (Lie–Trotter Product Formula[4])**.** *Suppose that the operators A and B are the generators of strongly continuous semigroups. Suppose further that there exist constants $M \geq 1$ and $\omega \in \mathbb{R}$ such that*

$$\left\| \left(\mathrm{e}^{\frac{t}{n}B}\mathrm{e}^{\frac{t}{n}A} \right)^n \right\| \leq M\mathrm{e}^{\omega t} \tag{10.3}$$

holds for all $t \geq 0$ and $n \in \mathbb{N}$. Consider the sum $A+B$ on $D = D(A)\cap D(B)$, and assume that D and $\big(\lambda_0 - (A+B)\big)D$ are dense in X for some $\lambda_0 > \omega$. Then $C = \overline{A+B}$ generates a strongly continuous semigroup given by the sequential splitting, i.e.,

$$\mathrm{e}^{tC}u_0 = \lim_{n\to\infty} \left(\mathrm{e}^{\frac{t}{n}B}\mathrm{e}^{\frac{t}{n}A} \right)^n u_0$$

holds for all $u_0 \in X$ uniformly for t in compact intervals.

Although the theorem above ensures the convergence of the sequential splitting under rather weak conditions, it does not tell us anything about the convergence rate. To obtain certain error bounds, the sum $A+B$ with domain $D = D(A)\cap D(B)$ needs to be a generator as well. This leads to stronger conditions on the operators A and B.

We show the first-order convergence of the sequential splitting following the idea which was presented by Jahnke and Lubich for the Marchuk–Strang splitting.[5] To this end, we need some assumptions.

Assumption 10.7. Suppose that A generates a strongly continuous semigroup on the Banach space X, and let $B \in \mathscr{L}(X)$. Suppose further that there exist a subspace Y such that

$$D(A) \hookrightarrow Y \hookrightarrow X$$

with dense and continuous embeddings. We also assume that $D(A)$ is invariant under the operator B, and that Y is invariant under the semigroup $(\mathrm{e}^{tA})_{t\geq 0}$.

Note that this means in particular that there are constants K_1 and K_2 such that

$$\begin{aligned} \|f\|_Y &\leq K_1\|f\|_A && \text{holds for all } f \in D(A), \text{ and} \\ \|f\| &\leq K_2\|f\|_Y && \text{holds for all } f \in Y. \end{aligned}$$

[4] H. F. Trotter,"On the product of semi-groups of operators," Proc. Amer. Math. Soc. **10** (1959), 545–551.

[5] T. Jahnke and Ch. Lubich, "Error bounds for exponential operator splittings," BIT **40** (2000), 735–744.

We have already seen in the matrix case that the commutator of A and B plays an important role, when seeking an error bound. It motivates us to define it properly also for A, B being generators, and to bound it on an appropriate subspace. For $B \in \mathscr{L}(X)$ the commutator

$$[A, B]f = ABf - BAf$$

is only defined on $D(A)$. However, in some cases we may be able to extend this estimate for all $f \in Y$ for the subspace Y satisfying Assumption 10.7.

Assumption 10.8. Suppose that A generates a strongly continuous semigroup on the Banach space X, and that $B \in \mathscr{L}(X)$. Further suppose that there exists a constant $c_1 \geq 0$ such that

$$\|[A, B]f\| \leq c_1 \|f\|_Y \tag{10.4}$$

holds for all $f \in D(A)$ with Y being the appropriate subspace such as in Assumption 10.7.

Note that in this case the operator $[A, B]$ extends continuously to the entire space Y.

Example 10.9. Suppose that A generates a strongly continuous semigroup on the Banach space X, and let $B \in \mathscr{L}(X)$. Suppose further that there exist constants $\alpha \in (0, 1)$ and $c_1 \geq 0$ such that

$$\|[A, B]f\| \leq c_1 \|(-A)^\alpha f\|$$

holds for all $f \in D(A)$. Then the subspace $Y = D\big((-A)^\alpha\big)$ possesses all the properties listed in Assumption 10.7, see Lecture 7.

Now we are able to state the first-order convergence of the sequential splitting.

Theorem 10.10. *Consider operators A, B and a subspace Y satisfying Assumption 10.7. Under Assumption 10.8 the sequential splitting is first-order convergent.*

Proof. Let $h \in (0, t_0]$. By applying Proposition 4.12, it is sufficient to show the stability and the first-order consistency. Since A is a generator and B is bounded, by Exercise 5.5 the sum $A + B$ with domain $D(A)$ generates a semigroup. After applying the renorming procedure (see Exercise C.4) and shifting the generators in the appropriate way, we may assume that A, B, and $A + B$ generate contraction semigroups, that is, for all $t \geq 0$ we have

$$\|\mathrm{e}^{tA}\| \leq 1, \quad \|\mathrm{e}^{tB}\| \leq 1, \quad \text{and} \quad \|\mathrm{e}^{t(A+B)}\| \leq 1. \tag{10.5}$$

In particular, for $t < t_0$ this proves the stability of F_{seq}. Note that although the rescaling does not effect the order of convergence, it may modify the constants appearing in the estimates and it may have effect how small the time step h has to be chosen.

To show the first-order consistency (see Definition 4.11), we have to ensure the existence of a constant M, depending on c_1 and $\|B\|$, such that for all $f \in D(A) \subset Y$ we have

$$\mathscr{E}_{\mathrm{seq}}(h, f) = \big\|F_{\mathrm{seq}}(h)f - \mathrm{e}^{h(A+B)}f\big\| \leq Mh^2 \|f\|_Y. \tag{10.6}$$

Taylor's Formula from Exercise C.2 implies

$$\mathrm{e}^{hB}g = g + hBg + \int\limits_0^h (h - s)B^2 \mathrm{e}^{sB} g \,\mathrm{d}s$$

for all $g \in X$. In particular, for $g = \mathrm{e}^{hA} f$ we have

$$\mathrm{e}^{hB}\mathrm{e}^{hA} f = \mathrm{e}^{hA} f + hB\mathrm{e}^{hA} f + \int_0^h (h-s)B^2 \mathrm{e}^{sB}\mathrm{e}^{hA} f \mathrm{d}s. \tag{10.7}$$

On the other hand, note that the solution of the initial value problem $\frac{\mathrm{d}}{\mathrm{d}t}v(t) = (A+B)v(t)$, $v(0) = v_0$, can be expressed by the variation-of-constants formula as

$$v(h) = \mathrm{e}^{hA} v_0 + \int_0^h \mathrm{e}^{(h-s)A} B v(s) \mathrm{d}s,$$

which implies

$$\mathrm{e}^{h(A+B)} f = \mathrm{e}^{hA} f + \int_0^h \mathrm{e}^{(h-s)A} B \mathrm{e}^{s(A+B)} f \mathrm{d}s.$$

Using the variation-of-constants formula for the term $v(s) = B\mathrm{e}^{s(A+B)} f$ once again, we obtain

$$\mathrm{e}^{h(A+B)} f = \mathrm{e}^{hA} f + \int_0^h \mathrm{e}^{(h-s)A} B \mathrm{e}^{sA} f \mathrm{d}s + \int_0^h \mathrm{e}^{(h-s)A} B \Big(\int_0^s \mathrm{e}^{(s-r)A} B \mathrm{e}^{r(A+B)} f \mathrm{d}r \Big) \mathrm{d}s. \tag{10.8}$$

Subtracting (10.7) from (10.8), the local error can be written as

$$\begin{aligned} \mathscr{E}_{\mathrm{seq}}(h, f) &= \big\| \mathrm{e}^{hB}\mathrm{e}^{hA} f - \mathrm{e}^{h(A+B)} f \big\| \\ &\le \Big\| hB\mathrm{e}^{hA} f - \int_0^h \mathrm{e}^{(h-s)A} B \mathrm{e}^{sA} f \mathrm{d}s \Big\| + \|(R_1 - R_2) f\| \end{aligned} \tag{10.9}$$

with
$$R_1 f = \int_0^h (h-s) B^2 \mathrm{e}^{sB} \mathrm{e}^{hA} f \mathrm{d}s$$

and
$$R_2 f = \int_0^h \mathrm{e}^{(h-s)A} B \Big(\int_0^s \mathrm{e}^{(s-r)A} B \mathrm{e}^{r(A+B)} f \mathrm{d}r \Big) \mathrm{d}s.$$

For a continuously differentiable function $\eta : \mathbb{R} \to X$, the fundamental theorem of calculus implies that

$$\eta(s) = \eta(h) + \int_h^s \eta'(r) \mathrm{d}r,$$

and therefore

$$h\eta(h) - \int_0^h \eta(s) \mathrm{d}s = h\eta(h) - \int_0^h \eta(h) \mathrm{d}s - \int_0^h \Big(\int_h^s \eta'(r) \mathrm{d}r \Big) \mathrm{d}s = \int_0^h \Big(\int_s^h \eta'(r) \mathrm{d}r \Big) \mathrm{d}s. \tag{10.10}$$

We note that this corresponds to the error of the a first-order quadrature rule (the right rectangular rule). In particular, for

$$\eta(s) = \mathrm{e}^{(h-s)A} B \mathrm{e}^{sA} f,$$

being continuously differentiable because $D(A)$ is invariant under B, we have

$$h\eta(h) - \int_0^h \eta(s)\mathrm{d}s = hB\mathrm{e}^{hA}f - \int_0^h \mathrm{e}^{(h-s)A}B\mathrm{e}^{sA}f\mathrm{d}s$$

being exactly the first term in (10.9) to be estimated. For this special choice of η we obtain

$$\begin{aligned}\eta'(s) &= \mathrm{e}^{(h-s)A}(-A)B\mathrm{e}^{sA}f + \mathrm{e}^{(h-s)A}BA\mathrm{e}^{sA}f\\ &= -\mathrm{e}^{(h-s)A}(AB-BA)\mathrm{e}^{sA}f = -\mathrm{e}^{(h-s)A}[A,B]\mathrm{e}^{sA}f,\end{aligned}$$

and $$\|\eta'(s)\| \le \big\|\mathrm{e}^{(h-s)A}\big\| \cdot \big\|[A,B]\mathrm{e}^{sA}f\big\| \le c_1\|\mathrm{e}^{sA}f\|_Y = c\big\|\mathrm{e}^{sA}\big\|_Y\|f\|_Y,$$

where we used condition (10.4) for $\mathrm{e}^{sA}f \in Y$. Formula (10.10) implies the estimate for the first term in (10.9):

$$\begin{aligned}\Big\|hB\mathrm{e}^{hA}f - \int_0^h \mathrm{e}^{(h-s)A}B\mathrm{e}^{sA}f\mathrm{d}s\Big\| &\le \int_0^h \Big(\int_s^h \|\eta'(r)\|\,\mathrm{d}r\Big)\mathrm{d}s \qquad (10.11)\\ &\le \int_0^h \Big(\int_s^h c\big\|\mathrm{e}^{sA}\big\|_Y\|f\|_Y\mathrm{d}r\Big)\mathrm{d}s \le K\cdot h^2\mathrm{e}^{\omega h}\|f\|_Y \le K'\cdot h^2\|f\|_Y,\end{aligned}$$

for some constant $K' \ge 0$, where we used $h \le t_0$ in the last step. For the second term in (10.9) we have the rough estimate $\|(R_1-R_2)f\| \le \|R_1f\| + \|R_2f\|$ with

$$\begin{aligned}\|R_1f\| &= \Big\|\int_0^h (h-s)B^2\mathrm{e}^{sB}\mathrm{e}^{hA}f\mathrm{d}s\Big\| \le \int_0^h (h-s)\|B^2\|\cdot\big\|\mathrm{e}^{sB}\big\|\cdot\big\|\mathrm{e}^{hA}\big\|\cdot\|f\|\mathrm{d}s\\ &\le \frac{h^2}{2}\|B\|^2\cdot\|f\| \qquad (10.12)\end{aligned}$$

and

$$\begin{aligned}\|R_2f\| &= \Big\|\int_0^h \mathrm{e}^{(h-s)A}B\Big(\int_0^s \mathrm{e}^{(s-r)A}B\mathrm{e}^{r(A+B)}f\mathrm{d}r\Big)\mathrm{d}s\Big\|\\ &\le \int_0^h \big\|\mathrm{e}^{(h-s)A}\big\|\cdot\|B\|\Big(\int_0^s \big\|\mathrm{e}^{(s-r)A}\big\|\cdot\|B\|\cdot\big\|\mathrm{e}^{r(A+B)}\big\|\cdot\|f\|\mathrm{d}r\Big)\mathrm{d}s\\ &\le \frac{h^2}{2}\|B\|^2\cdot\|f\|. \qquad (10.13)\end{aligned}$$

Estimates (10.11), (10.12), and (10.13) imply the desired error bound for (10.9) with an appropriate constant M depending on c_1 and $\|B\|$. □

If operator A generates an analytic semigroup of type $(0,\omega)$ with $\omega < 0$, even stronger estimates hold, requiring bounds only on the norm of the initial value u_0. Recall that in this case there is a constant $M > 0$ so that the estimates

$$\big\|A\mathrm{e}^{tA}\big\| \le \frac{M}{t}, \quad \big\|(A+B)\mathrm{e}^{t(A+B)}\big\| \le \frac{M}{t}, \quad \text{and hence } \big\|A\mathrm{e}^{t(A+B)}\big\| \le \frac{M}{t} \qquad (10.14)$$

hold for all $t > 0$, see Corollary 9.22.

Theorem 10.11. *Suppose that A generates an analytic semigroup of type $(1,\omega)$ with $\omega < 0$ on the Banach space X, and $B \in \mathscr{L}(X)$. Suppose further that Assumptions 10.7 and 10.8 hold with $Y = D\big((-A)^\alpha\big)$ for some $\alpha \in (0,1)$, i.e.,*

$$\|[A,B]f\| \le c\|(-A)^\alpha f\| \quad \text{holds for all } f \in D(A).$$

Then the global error of the sequential splitting is bounded by

$$\|u_{\mathrm{seq},h}(nh) - u(nh)\| \le hM_0 \log(n)\|u_0\|,$$

for all $u_0 \in X$ and $n > 1$, $n \in \mathbb{N}$, $h \ge 0$, $nh \in [0,t_0]$. In particular, the sequential splitting converges in the operator norm like $\frac{\log n}{n}$.

Proof. As before, we assume that all occurring semigroups are contraction semigroups. Applying the telescopic identity, the global error can be written with the help of the local error as

$$\begin{aligned}
\|u_{\mathrm{seq},h}(nh) - u(nh)\| &= \big\|F_{\mathrm{seq}}(h)^n u_0 - \mathrm{e}^{nh(A+B)}u_0\big\| \\
&= \Big\|\sum_{j=0}^{n-1} F_{\mathrm{seq}}(h)^{n-j-1}\big(F_{\mathrm{seq}}(h) - \mathrm{e}^{h(A+B)}\big)\mathrm{e}^{jh(A+B)}u_0\Big\| \\
&\le \sum_{j=0}^{n-1} \|F_{\mathrm{seq}}(h)\|^{n-j-1} \cdot \big\|\big(F_{\mathrm{seq}}(h) - \mathrm{e}^{h(A+B)}\big)\mathrm{e}^{jh(A+B)}u_0\big\| \\
&\le \quad \|F_{\mathrm{seq}}(h)\|^{n-1}\big\|\big(F_{\mathrm{seq}}(h) - \mathrm{e}^{h(A+B)}\big)u_0\big\| \\
&\quad + \sum_{j=1}^{n-1} \|F_{\mathrm{seq}}(h)\|^{n-j-1}\big\|\big(F_{\mathrm{seq}}(h) - \mathrm{e}^{h(A+B)}\big)A^{-1}\big\| \cdot \big\|A\mathrm{e}^{jh(A+B)}u_0\big\|
\end{aligned}$$

for all $u_0 \in D(A)$. Since $\|F_{\mathrm{seq}}(h)\| \le 1$, we obtain

$$\begin{aligned}
\|u_{\mathrm{seq},h}(nh) - u(nh)\| \le{}& \big\|\big(F_{\mathrm{seq}}(h) - \mathrm{e}^{h(A+B)}\big)u_0\big\| \\
&+ \sum_{j=1}^{n-1} \big\|\big(F_{\mathrm{seq}}(h) - \mathrm{e}^{h(A+B)}\big)A^{-1}\big\| \cdot \big\|A\mathrm{e}^{jh(A+B)}u_0\big\|. \qquad (10.15)
\end{aligned}$$

Since $A + B$ generates an analytic semigroup, we have

$$\big\|A\mathrm{e}^{jh(A+B)}\big\| \le \frac{M}{jh}$$

for all $j = 1, \dots, n$. For $g \in X$ we have $A^{-1}g \in D(A) \hookrightarrow D((-A)^\alpha)$, and hence

$$\big\|\big(F_{\mathrm{seq}}(h) - \mathrm{e}^{h(A+B)}\big)A^{-1}g\big\| \le h^2C\|A^{-1}g\|_{(-A)^\alpha} \le h^2C'\|g\|.$$

Hence, for the second term in (10.15) we obtain

$$\sum_{j=1}^{n-1} \big\|\big(F_{\mathrm{seq}}(h) - \mathrm{e}^{h(A+B)}\big)A^{-1}\big\| \cdot \big\|A\mathrm{e}^{jh(A+B)}u_0\big\| \le \sum_{j=1}^{n-1} h^2C'\frac{M}{jh} \le hC'' \log(n).$$

Consider now the case $j=0$, that is, the first term in (10.15). To this end, we have to estimate the operator norm of the local error. By (10.9), we have

$$\mathscr{E}_{\mathrm{seq}}(h,f) \leq \Big\| hB\mathrm{e}^{hA}u_0 - \int\limits_0^h \mathrm{e}^{(h-s)A}B\mathrm{e}^{sA}u_0\mathrm{d}s \Big\| + \|(R_1-R_2)u_0\|\,.$$

Note that by (10.12) and (10.13) the inequality

$$\|(R_1-R_2)\| \leq \|R_1\| + \|R_2\| \leq h^2\|B\|^2.$$

holds. We conclude by means of (10.11) that

$$\begin{aligned}\Big\| hB\mathrm{e}^{hA}u_0 - \int\limits_0^h \mathrm{e}^{(h-s)A}B\mathrm{e}^{sA}u_0\mathrm{d}s \Big\| &\leq \int\limits_0^h \Big(\int\limits_s^h \Big\| \mathrm{e}^{(h-r)A}[A,B]\mathrm{e}^{rA}u_0 \Big\|\,\mathrm{d}r \Big)\mathrm{d}s \\ \leq \int\limits_0^h \Big(\int\limits_s^h \Big\| \mathrm{e}^{(h-r)A}[A,B](-A)^{-\alpha}(-A)^{\alpha}\mathrm{e}^{rA}u_0 \Big\|\,\mathrm{d}r \Big)\mathrm{d}s \\ \leq \int\limits_0^h \Big(\int\limits_s^h \Big\| \mathrm{e}^{(h-r)A}[A,B](-A)^{-\alpha} \Big\| \cdot \|(-A)^{\alpha}\mathrm{e}^{rA}u_0\|\,\mathrm{d}r \Big)\mathrm{d}s,\end{aligned}$$

where by Assumption 10.8 we have $[A,B](-A)^{-\alpha} \in \mathscr{L}(X)$. By using Corollary 9.22, we obtain that

$$\begin{aligned}\Big\| hB\mathrm{e}^{hA}u_0 - \int\limits_0^h \mathrm{e}^{(h-s)A}B\mathrm{e}^{sA}u_0\mathrm{d}s \Big\| &\leq \|[A,B](-A)^{-\alpha}\| \int\limits_0^h \int\limits_s^h \big\|(-A)^{\alpha}\mathrm{e}^{rA}u_0\big\|\,\mathrm{d}r\mathrm{d}s \\ &\leq K \int\limits_0^h \int\limits_s^h \frac{1}{r^{\alpha}}\mathrm{d}r\mathrm{d}s\|u_0\| \leq K'h^{2-\alpha}\|u_0\|.\end{aligned}$$

Putting the pieces together we arrive at

$$\|u_{\mathrm{seq},h}(nh) - u(nh)\| \leq K'h^{2-\alpha}\|u_0\| + hC''\log(n)\|u_0\| \leq M_0 h\log(n)\|u_0\|. \qquad \square$$

The following two results of Jahnke and Lubich about the Marchuk–Strang splitting can be proved by a slightly more detailed analysis, we postpone the proofs to the project phase.

Theorem 10.12. *Suppose that A generates a semigroup of type $(1,\omega)$ with $\omega<0$ on the Banach space X, and $B \in \mathscr{L}(X)$. Suppose further that Assumptions 10.7 and 10.8 hold with $Y = D((-A)^{\alpha})$ for some $\alpha \in (0,1)$, i.e.,*

$$\|[A,B]f\| \leq c\|(-A)^{\alpha}f\| \quad \text{holds for all } f \in D(A). \tag{10.16}$$

Then the Marchuk–Strang splitting is first-order convergent on $D((-A)^{\alpha})$. If in addition there exist constants $c_2 \geq 0$ and $1 \leq \beta \leq 2$ such that

$$\|[A,[A,B]]g\| \leq c_2\big\|(-A)^{\beta}g\big\| \tag{10.17}$$

holds for all $g \in D(A^2)$, then the Marchuk–Strang splitting is convergent of second order on $D((-A)^{\beta})$.

Theorem 10.13. *Suppose that A generates an analytic semigroup on the Banach space X, and $B \in \mathscr{L}(X)$. Suppose further that B leaves $D(A)$ invariant and that conditions* (10.17) *and* (10.16) *hold for all $f, g \in D(A^2)$ with $c_1, c_2 \geq 0$, $\alpha \leq 1$, and $\beta = 1$. Then the global error of the Marchuk–Strang splitting is bounded by*

$$\|u_{\mathrm{MS},h}(nh) - u(nh)\| \leq h^2 M_0 \log(n) \|u_0\|$$

for all $u_0 \in X$.

10.3 Example

Consider the m-dimensional Schrödinger equation on $\Omega = (-\pi, \pi)^m$:

$$\begin{cases} \partial_t w(t,x) = \mathrm{i}\Delta w(t,x) - \mathrm{i}V(x)w(t,x), \quad x \in \Omega,\ t > 0 \\ \ w(0,x) = w_0(x), \quad x \in \Omega \end{cases} \tag{10.18}$$

with periodic boundary conditions, some given initial function w_0, and a C^4 potential $V : \mathbb{R}^m \to \mathbb{C}$ being 2π-periodic in every coordinate direction, and transform the equation to an abstract Cauchy problem in $\mathrm{L}^2(\Omega)$.

Let

$$\mathrm{C}^\infty_{\mathrm{per}}(\Omega) := \big\{ f \in \mathrm{C}^\infty(\mathbb{R}^m) : f \text{ is } 2\pi\text{-periodic in each coordinate direction} \big\},$$

and for $f \in \mathrm{C}^\infty_{\mathrm{per}}(\Omega)$ we define

$$\begin{aligned} \|f\|^2_{\mathrm{H}^1(\Omega)} &:= \|f\|^2_{\mathrm{L}^2(\Omega)} + \|\partial_1 f\|^2_{\mathrm{L}^2(\Omega)} + \cdots + \|\partial_m f\|^2_{\mathrm{L}^2(\Omega)} \\ \|f\|^2_{\mathrm{H}^2(\Omega)} &:= \|f\|^2_{\mathrm{L}^2(\Omega)} + \|\partial_1 f\|^2_{\mathrm{L}^2(\Omega)} + \cdots + \|\partial_m f\|^2_{\mathrm{L}^2(\Omega)} + \sum_{i,j=1}^m \|\partial_i \partial_j f\|^2_{\mathrm{L}^2(\Omega)}. \end{aligned}$$

The completion of

$$\big\{ f|_\Omega : f \in \mathrm{C}^\infty_{\mathrm{per}}(\Omega) \big\}$$

with respect to the norms $\|\cdot\|_{\mathrm{H}^1(\Omega)}$ and $\|\cdot\|_{\mathrm{H}^2(\Omega)}$ is denoted by $\mathrm{H}^1_{\mathrm{per}}(\Omega)$ and $\mathrm{H}^2_{\mathrm{per}}(\Omega)$, respectively. Both are Banach spaces for the respective norms.

Now we split the operator C according to the different physical phenomena:

$$A = \mathrm{i}\Delta \qquad \text{and} \quad B = -M_{\mathrm{i}V}$$

with the corresponding domains

$$D(A) = \mathrm{H}^2_{\mathrm{per}}(\Omega)$$

and

$$D(B) = \mathrm{L}^2(\Omega).$$

Of course, since V is a bounded function, the multiplication operator $M_{\mathrm{i}V}$ is bounded on L^2. By using the Lumer–Phillips theorem, Theorem 6.3, one can show that $\mathrm{i}\Delta$ generates a contraction semigroup T on $\mathrm{L}^2(\Omega)$ (and, since $-\mathrm{i}\Delta$ does this, too, the semigroup operators are invertible, in fact they are unitary). By assumption B leaves $D(A)$ invariant.

In order to prove the first-order convergence of sequential splitting and the second-order convergence of Marchuk–Strang splitting, we have to verify the assumptions of Theorems 10.10 and 10.12, respectively. It remains to bound the norm of the commutators $[A, B]$ and $[A, [A, B]]$ on appropriate subspaces. To get an idea how to choose the subspaces, we *formally* compute the commutators:

$$[A, B]f = ABf - BAf = \Delta(Vf) - V(\Delta f) = (\Delta V)f + 2(\nabla V)^\top \cdot (\nabla f)$$

and

$$[A, [A, B]]g = A\big([A, B]g\big) - [A, B]\big(Ag\big) = \mathrm{i}\big(\Delta(\Delta V)\big)g + 4\mathrm{i}\big(\nabla(\Delta V)\big)^\top \cdot (\nabla g) + 4\mathrm{i}(\Delta V)(\Delta g).$$

They contain first- and second-order derivatives of the functions f and g, and their norms can be estimated as follows:

$$\begin{aligned} &\|[A, B]f\|_2 \le \|\Delta V\|_\infty \cdot \|f\|_2 + \|2(\nabla V)\|_\infty \cdot \|\nabla f\|_2 \le K(\|f\|_2 + \|\nabla f\|_2) \le K\|f\|_{\mathrm{H}^1} \\ \text{and}\quad &\|[A, [A, B]]g\|_2 \le \|\mathrm{i}\big(\Delta(\Delta V)\big)\|_\infty \cdot \|g\|_2 + \|4\mathrm{i}\big(\nabla(\Delta V)\big)\|_\infty \cdot \|\nabla g\|_2 + \|4\mathrm{i}(\Delta V)\|_\infty \cdot \|\Delta g\|_2 \\ &\qquad \le K'(\|g\|_2 + \|\nabla g\|_2 + \|\Delta g\|_2) \le K'\|g\|_{\mathrm{H}^2} \end{aligned}$$

for some constants $K, K' \ge 0$. By using again multiplicators as in Section 1.1 one can compute the fractional powers of $-\mathrm{i}\Delta$ (see also Exercise 7.2), and obtain that with the choice $Y = \mathrm{H}^1_{\mathrm{per}}(\Omega)$ and $\alpha = \frac{1}{2}$ and $\beta = 1$, the assumptions in Theorems 10.10 and 10.12 are fulfilled. Thus we conclude that the sequential splitting is of first order, and the Marchuk–Strang splitting is of second order for this problem.

One can show even more (see Exercise 7): Theorems 10.10 and 10.12 apply also to the semi-discretisation of the problem. By applying a certain operator splitting procedure, the numerical solution of problem (10.18) needs the approximation of the semigroups generated by the operators A and B. The multiplication is a pointwise calculation at every grid point. The semigroup generated by operator $A = \mathrm{i}\Delta$ can be approximated by applying some spectral method, see Appendix A. Using the approximation results presented in Lecture 3, we can imagine how such a combined method works. For more details, we refer to the project phase.

Now we illustrate the results above by presenting some figures showing the behaviour of the global error of the Marchuk–Strang splitting as a function of the time step h. We suppose $m = 1$. In the first case the smooth potential $V(x) = 1 - \cos x$ was used with random initial data in $\mathrm{H}^1_{\mathrm{per}}(-\pi, \pi)$, indicated by red circles, and with random initial data in $\mathrm{H}^2_{\mathrm{per}}(-\pi, \pi)$, indicated by blue stars. In the second case we used the non-smooth potential $V(x) = x + \pi$. One can see that in the case of the smooth potential the convergence is of higher order for the initial data being in $\mathrm{H}^2_{\mathrm{per}}(-\pi, \pi)$ than lying only in $\mathrm{H}^1_{\mathrm{per}}(-\pi, \pi)$. Thus, the numerical experiments are in line with the theoretical results. We note that the numerical experiments suggest that the non-smoothness of the potential does not allow a higher order convergence in general.

Remark 10.14. The commutator conditions stay the same for the corresponding parabolic problem

$$\partial_t w(t, x) = \Delta w(t, x) - V(x)w(t, x),$$

such as the heat equation with special source term, the linearised reaction-diffucion equation, or the imaginary-time Schrödinger equation. Thus, the considerations above apply to them as well.

Exercises

1. Compute the constant in the $\mathcal{O}(h^3)$ term in the formula (10.2).

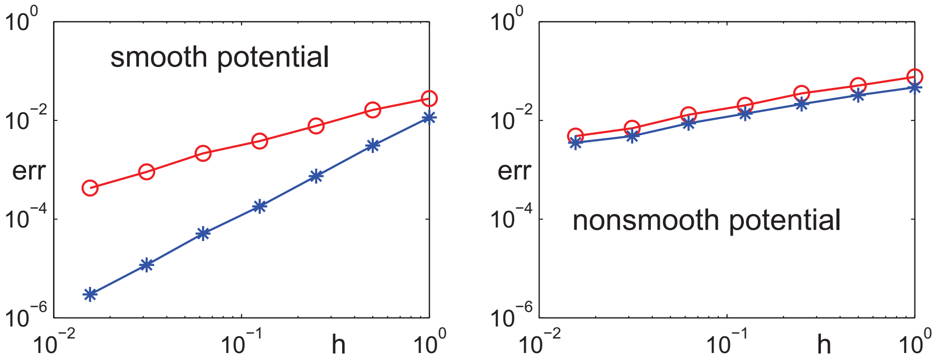

Figure 10.1: Global error of Marchuk–Strang splitting as a function of the time step h.

2. Let X be a Banach space and let $A, B \in \mathscr{L}(X)$. Prove that the following assertions are equivalent:

(i) $[A,B]=0$.

(ii) For all $t \geq 0$ we have $[\mathrm{e}^{tA}, \mathrm{e}^{tB}] = 0$.

Show that under these equivalent conditions one has $\mathrm{e}^{tA}\mathrm{e}^{tB} = \mathrm{e}^{t(A+B)}$.

3. Prove Proposition 10.4.

4. Prove Theorem 10.6.

5. Let A and B be the generators of strongly continuous semigroups. Show that if there exist constants $M \geq 1$ and $\omega \in \mathbb{R}$ such that

$$\left\| \left(\mathrm{e}^{\frac{t}{n}B} \mathrm{e}^{\frac{t}{n}A} \right)^n \right\| \leq M\mathrm{e}^{\omega t} \tag{10.19}$$

holds for all $t \geq 0$ and $n \in \mathbb{N}$, then there exist constants $M_1, M_2 \geq 1$ and $\omega_1, \omega_2 \in \mathbb{R}$ such that

$$\left\| \left(\mathrm{e}^{\frac{t}{n}A} \mathrm{e}^{\frac{t}{n}B} \right)^n \right\| \leq M_1\mathrm{e}^{\omega_1 t}$$

and

$$\left\| \left(\mathrm{e}^{\frac{t}{2n}A} \mathrm{e}^{\frac{t}{n}B} \mathrm{e}^{\frac{t}{2n}A} \right)^n \right\| \leq M_2\mathrm{e}^{\omega_2 t}$$

holds as well for all $t \geq 0$ and $n \in \mathbb{N}$.

6. Work out the details of the conditions appearing in (10.5).

7. Study the space discretisation of the Schrödinger equation (10.18) which you can find as an example in the paper of Jahnke and Lubich. Implement the method together with the sequential and Marchuk–Strang splittings, and solve the equation numerically.

Lecture 11

Operator Splitting with Discretisations

We continue the study of operator splitting procedures and present two more topics concerning these.

In Lecture 10 we investigated the convergence of the splitting procedures in the case when the sub-problems are solved *exactly*. In concrete problems, however, the exact solutions are usually not known. Therefore the use of certain approximation schemes is needed to solve the sub-problems. When a partial differential equation is to be solved by applying operator splitting together with approximation schemes, one usually follows the next steps:

1. The spatial differential operator is split into sub-operators of simpler form.
2. Each sub-operator is approximated by an appropriate space discretisation method (called semi-discretisation). Then we obtain systems of ordinary differential equations corresponding to the sub-operators.
3. Each solution of the semi-discretised system is obtained by using a time discretisation method.

In this lecture we investigate the cases where the solutions of the sub-problems are approximated by using only a time discretisation method or only a space discretisation method. More general cases, where all the steps 1, 2, and 3 are considered, will be investigated in the project phase of this seminar.

As in Lecture 10, we consider the abstract Cauchy problem on the Banach space X of the form

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}u(t) = (A+B)u(t), & t > 0 \\ u(0) = u_0 \end{cases} \tag{11.1}$$

with densely defined, closed linear operators A and B. Throughout the lecture we suppose that the closure C of $A+B$ with domain $D(A) \cap D(B)$ is a generator, and also that the initial value u_0 is taken from $D(A) \cap D(B)$.

Although we started our study of splitting procedures by giving the operators A and B explicitly in the abstract Cauchy problem (11.1), in real-life applications the sum operator C is given. The natural question arises how to split the operator C into the sub-operators A and B (cf. step 1 above). In practice, there are several ways to do this:

a) First we discretise the operator C in space, and split the matrix appearing in the semi-discretised problem (according to some of its structural properties).

b) The operator C describes the combined effect of several different phenomena and is already written as a sum of the corresponding sub-operators.

c) We split the operator according to the space directions (**dimension splitting**).[1]

[1] Sometimes also called as coordinate splitting.

Procedure a) leads to the matrix case which was already presented in Section 10.1, and b) was investigated in Section 10.2. First, we develop some abstract results that are applicable for the dimension splitting. In contrast to the results presented in Section 10.2 (when operator B was bounded), in this case we will have two unbounded operators.

11.1 Resolvent splittings

We turn our attention to the resolvent splittings which have already been introduced in Definition 10.1:

$$\begin{aligned} &\text{Lie splitting:} && F_{\mathrm{Lie}}(h) = \bigl(I - hB\bigr)^{-1}\bigl(I - hA\bigr)^{-1}, \\ &\text{Peaceman–Rachford splitting:} && F_{\mathrm{PR}}(h) = \bigl(I - \tfrac{h}{2}B\bigr)^{-1}\bigl(I + \tfrac{h}{2}A\bigr)\bigl(I - \tfrac{h}{2}A\bigr)^{-1}\bigl(I + \tfrac{h}{2}B\bigr). \end{aligned}$$

We note that for closed and linear operators A, B , there appear the corresponding resolvents in $F_{\mathrm{Lie}}(h)$ and $F_{\mathrm{PR}}(h)$. We only have to assume that $\frac{1}{h}$ and $\frac{2}{h}$ belong to the resolvent sets of both A and B for the Lie and the Peaceman–Rachford splittings, respectively. We are interested in, however, the convergence of the splitting procedures. Since h has to be small then, we may suppose without loss of generality that $\frac{1}{h}$ (and therefore $\frac{2}{h}$) is large enough.

Observe that the terms in Lie and Peaceman–Rachford splittings correspond to the explicit and implicit Euler methods with time steps h or $\frac{h}{2}$. Thus, the resolvent splittings can be considered as the application of operator splitting together with special time discretisation methods.

We prove next the first-order convergence of the Lie splitting following the idea presented by Hansen and Ostermann.[2]

Theorem 11.1. *Let A and B be linear operators, and suppose that there is a $\lambda_0 > 0$ such that $\lambda \in \rho(A) \cap \rho(B)$ for all $\lambda \geq \lambda_0$ (i.e., the Lie splitting is well-defined), and that there exist constants $M \geq 1$ and $\omega \in \mathbb{R}$ such that*

$$\bigl\|\bigl(F_{\mathrm{Lie}}(h)\bigr)^k\bigr\| \leq M\mathrm{e}^{k\omega h} \tag{11.2}$$

holds for all $k \in \mathbb{N}$ and $h \in [0, \frac{1}{\lambda_0}]$ (i.e., the Lie splitting is stable). Suppose further that the closure C of $A + B$ (with $A + B$ having the natural domain $D(A) \cap D(B)$) generates a strongly continuous semigroup on the Banach space X of type (M, ω), and that $D(C^2) \subseteq D(AB) \cap D(A)$ and $AB(\lambda_0 - C)^{-2} \in \mathscr{L}(X)$ hold. Then the Lie splitting is first-order convergent on $D(C^2)$. That is to say, for all $t_0 \geq 0$ we have

$$\|u_{\mathrm{Lie},h}(nh) - u(nh)\| \leq hK\bigl(\|u_0\| + \|Cu_0\| + \|C^2u_0\|\bigr)$$

for all $u_0 \in D(C^2)$ and $nh \in [0, t_0]$, $n \in \mathbb{N}$, $h \in [0, h_0]$, where the constant K may depend on t_0, but not on n and h.

Proof. For better readability we denote the semigroup operators by e^{tC}. By using the definition of the split solution, $u_{\mathrm{Lie},h}(nh) = F_{\mathrm{Lie}}(h)^n u_0$, and the telescopic identity, the error term can be rewritten as

$$u_{\mathrm{Lie},h}(nh) - u(nh) = F_{\mathrm{Lie}}(h)^n u_0 - \mathrm{e}^{nhC}u_0 = \sum_{j=0}^{n-1} F_{\mathrm{Lie}}(h)^{n-j-1}\bigl(F_{\mathrm{Lie}}(h) - \mathrm{e}^{hC}\bigr)\mathrm{e}^{jhC}u_0. \tag{11.3}$$

[2] E. Hansen and A. Ostermann, “Dimension splitting for evolution equations,” Numer. Math. **108** (2008), 557–570.

Next we estimate the local error, i.e., the term $F_{\mathrm{Lie}}(h) - \mathrm{e}^{hC}$ in the expression above. To do that we need to introduce some auxiliary functions, which will also come handy in later lectures.

For all $h > 0$ and $j \in \mathbb{N}$ we define the bounded linear operators $\varphi_j(hC)$ by

$$\varphi_j(hC)f := \frac{1}{h^j}\int\limits_0^h \frac{\tau^{j-1}}{(j-1)!}\mathrm{e}^{(h-\tau)C}f\,\mathrm{d}\tau, \tag{11.4}$$

for all $f \in X$ and $\varphi_0(hC) = \mathrm{e}^{hC}$. These operators are uniformly bounded for $h \in (0, t_0]$. Indeed, this is trivial for $\varphi_0(jC)$, whereas for $j \geq 1$ we have

$$\begin{aligned}\|\varphi_j(hC)f\| &\leq \frac{1}{h^j}\int\limits_0^h \frac{\tau^{j-1}}{(j-1)!}\big\|\mathrm{e}^{(h-\tau)C}f\big\|\mathrm{d}\tau \leq \frac{1}{h^j(j-1)!}\|f\|\int\limits_0^h r^{j-1}M\mathrm{e}^{\omega(h-r)}\mathrm{d}r\\ &\leq \frac{1}{h^j(j-1)!}M_C\mathrm{e}^{\max\{0,\omega\}h}\|f\|\int\limits_0^h r^{j-1}\mathrm{d}r \leq \frac{1}{j!}M\mathrm{e}^{\max\{0,\omega\}t_0}\|f\| = \text{const.}\cdot\|f\|.\end{aligned}$$

Moreover, the operators $\varphi_j(hC)$ satisfy the recurrence relation

$$\varphi_j(hC)f = \tfrac{1}{j!}f + hC\varphi_{j+1}(hC)f \tag{11.5}$$

for all $j = 0, 1, 2, \ldots$ and $f \in X$, see Exercise 1. They also leave $D(C)$ and $D(C^2)$ invariant. In the rest of the proof, we shall use only the following two consequences of (11.5):

$$\big(I - \varphi_0(hC)\big)f = -hC\varphi_1(hC)f \quad \text{and} \quad \big(\varphi_1(hC) - \varphi_0(hC)\big)f = hC\big(\varphi_2(hC) - \varphi_1(hC)\big)f.$$

For $f \in D(C^2)$ we shall derive a form for the local error $F_{\mathrm{Lie}}(h)f - \mathrm{e}^{hC}f$ that allows for the appropriate estimates. So we take $f \in D(C^2) \subset D(A) \cap D(B)$. For the sake of brevity we introduce the following abbreviations for the resolvents:

$$R_A = R(\tfrac{1}{h}, A) \qquad \text{and} \qquad R_B = R(\tfrac{1}{h}, B),$$

then we have

$$F_{\mathrm{Lie}}(h) = \tfrac{1}{h^2}R_BR_A.$$

By using the identity $I = \lambda R(\lambda, A) - AR(\lambda, A)$ for all $\lambda \in \rho(A)$, i.e. $I = \frac{1}{h}R_A - AR_A$ in our case, we express now the local error, i.e., the middle term in the telescopic sum (11.3) in the following form:

$$\begin{aligned}\big(F_{\mathrm{Lie}}(h) - \mathrm{e}^{hC}\big)f &= \big(F_{\mathrm{Lie}}(h) - \varphi_0(hC)\big)f\\ &= F_{\mathrm{Lie}}(h)f - \big(\tfrac{1}{h}R_B - BR_B\big)\big(\tfrac{1}{h}R_A - AR_A\big)\varphi_0(hC)f\\ &= F_{\mathrm{Lie}}(h)f - \big(\tfrac{1}{h^2}R_BR_A - \tfrac{1}{h}BR_BR_A - \tfrac{1}{h}R_BAR_A + BR_BAR_A\big)\varphi_0(hC)f\\ &= F_{\mathrm{Lie}}(h)\big(I - \varphi_0(hC)\big)f + \big(\tfrac{1}{h}BR_BR_A + \tfrac{1}{h}R_BAR_A - BR_BAR_A\big)\varphi_0(hC)f.\end{aligned}$$

Since $f \in D(C^2) \subseteq D(A)$, we can write

$$\begin{aligned}&\big(F_{\mathrm{Lie}}(h) - \mathrm{e}^{hC}\big)f\\ &\quad = F_{\mathrm{Lie}}(h)\big(I - \varphi_0(hC)\big)f + \big(hBF_{\mathrm{Lie}}(h) + hF_{\mathrm{Lie}}(h)A - h^2BF_{\mathrm{Lie}}(h)A\big)\varphi_0(hC)f.\end{aligned} \tag{11.6}$$

Observe that for every $f \in D(C^2) \subseteq D(A)$ we have the following relation:

$$\begin{aligned}\big(hBF_{\mathrm{Lie}}(h) - h^2BF_{\mathrm{Lie}}(h)A\big)f &= hBF_{\mathrm{Lie}}(h)(I - hA)f\\ &= hB(I-hB)^{-1}(I-hA)^{-1}(I-hA)f = hB(I-hB)^{-1}f\\ &= BR_Bf.\end{aligned}$$

For $f \in D(C^2) \subseteq D(B)$ we can rewrite this as:

$$\begin{aligned}\big(hBF_{\mathrm{Lie}}(h) - h^2BF_{\mathrm{Lie}}(h)A\big)f &= BR_Bf\\ &= R_BBf = R_B(\tfrac{1}{h}R_A - AR_A)Bf = \tfrac{1}{h}R_BR_ABf - R_BAR_ABf\\ &= \tfrac{1}{h}R_BR_ABf - R_BR_AABf = hF_{\mathrm{Lie}}(h)Bf - h^2F_{\mathrm{Lie}}(h)ABf.\end{aligned}$$

By inserting this last expression into (11.6), we obtain

$$\big(F_{\mathrm{Lie}}(h) - \mathrm{e}^{hC}\big)f = F_{\mathrm{Lie}}(h)\big(I - \varphi_0(hC)\big)f + hF_{\mathrm{Lie}}(h)(A+B)\varphi_0(hC)f - h^2F_{\mathrm{Lie}}(h)AB\varphi_0(hC)f.$$

By using the identity $\big(I - \varphi_0(hC)\big)f = -hC\varphi_1(hC)f$ we obtain

$$\begin{aligned}\big(F_{\mathrm{Lie}}(h) - \mathrm{e}^{hC}\big)f &= -F_{\mathrm{Lie}}(h)hC\varphi_1(hC)f + hF_{\mathrm{Lie}}(h)C\varphi_0(hC)f - h^2F_{\mathrm{Lie}}(h)AB\varphi_0(hC)f\\ &= F_{\mathrm{Lie}}(h)hC\big(\varphi_0(hC) - \varphi_1(hC)\big)f - h^2F_{\mathrm{Lie}}(h)AB\varphi_0(hC)f.\end{aligned}$$

On the other hand, we have $\big(\varphi_0(hC) - \varphi_1(hC)\big)f = hC\big(\varphi_1(hC) - \varphi_2(hC)\big)f$, and therefore

$$\big(F_{\mathrm{Lie}}(h) - \mathrm{e}^{hC}\big)f = F_{\mathrm{Lie}}(h)h^2C^2\big(\varphi_1(hC) - \varphi_2(hC)\big)f - h^2F_{\mathrm{Lie}}(h)AB\varphi_0(hC)f$$

for all $f \in D(C^2)$.

We now return to the error term. By using the equality above for $f = \mathrm{e}^{jhC}u_0$, and that

$$\varphi_0(hC)f = \mathrm{e}^{hC}f = (\lambda_0 - C)^{-2}\mathrm{e}^{hC}(\lambda_0 - C)^2f$$

holds for all $f \in D(C^2)$, we obtain the following expression for the error term in (11.3):

$$\begin{aligned}&F_{\mathrm{Lie}}(h)^nu_0 - \mathrm{e}^{nhC}u_0\\ &\quad= h^2\sum_{j=0}^{n-1}F_{\mathrm{Lie}}(h)^{n-j}\Big(\big(\varphi_1(hC) - \varphi_2(hC)\big)C^2 - AB(\lambda_0 - C)^{-2}\mathrm{e}^{hC}(\lambda_0 - C)^2\Big)\,\mathrm{e}^{jhC}u_0\\ &\quad= h^2\sum_{j=0}^{n-1}F_{\mathrm{Lie}}(h)^{n-j}\Big(\big(\varphi_1(hC) - \varphi_2(hC)\big)\mathrm{e}^{jhC}C^2 - AB(\lambda_0 - C)^{-2}\mathrm{e}^{(j+1)hC}(\lambda_0 - C)^2\Big)\,u_0\end{aligned}$$

for all $u_0 \in D(C^2)$. Since the terms $\varphi_1(hC)$, $\varphi_2(hC)$, and e^{hC} are uniformly bounded for $h \in (0, t_0]$ and since by assumption the operator $AB(\lambda_0 - C)^{-2}$ is bounded, we obtain the desired estimate:

$$\begin{aligned}\big\|F_{\mathrm{Lie}}(h)^nu_0 - \mathrm{e}^{nhC}u_0\big\| &\le h^2\sum_{j=0}^{n-1}\|F_{\mathrm{Lie}}(h)^{n-j}\|\,\Big((\|\varphi_1(hC)\| + \|\varphi_2(hC)\|)\cdot\big\|\mathrm{e}^{hC}\big\|^j\cdot\|C^2u_0\|\\ &\qquad+\|AB(\lambda_0 - C)^{-2}\|\cdot\big\|\mathrm{e}^{hC}\big\|^{j+1}\cdot\|(\lambda_0^2 - 2\lambda_0C + C^2)u_0\|\Big)\\ &\le h^2nM\mathrm{e}^{\max\{0,\omega\}t}\Big(\text{const.}\cdot M\mathrm{e}^{\max\{0,\omega\}t}\cdot\|C^2u_0\|\\ &\qquad+\text{const.}\cdot M\mathrm{e}^{\max\{0,\omega\}t}\big(\|u_0\| + \|Cu_0\| + \|C^2u_0\|\big)\Big)\\ &\le h^2n\widetilde{K}\left(\|u_0\| + \|Cu_0\| + \|C^2u_0\|\right) = hK\left(\|u_0\| + \|Cu_0\| + \|C^2u_0\|\right)\end{aligned}$$

with a positive constant K depending on t_0. This completes the proof. □

Remark 11.2. The stability condition (11.2) in Theorem 11.1 is satisfied if

$$\big\|(I-hA)^{-1}\big\| \leq 1 \quad \text{and} \quad \big\|(I-hB)^{-1}\big\| \leq 1$$

hold for all $h>0$. In this case, by the Lumer–Phillips theorem, Theorem 6.3 both A and B generate contraction semigroups on X. As a consequence of the convergence result above the operator C generates a contraction semigroup, too.

Following the idea of the proof above, the second-order convergence of the Peaceman–Rachford splitting can also be shown (see the already mentioned paper of Hansen and Ostermann).

Theorem 11.3. *Let A and B be linear operators, and suppose that there is a $\lambda_0>0$ such that $\lambda \in \rho(A)\cap\rho(B)$ for all $\lambda \geq \lambda_0$ (i.e., the Peaceman–Rachford splitting is well-defined on $D(B)$), and that there exist constants $M\geq 1$ and $\omega\in\mathbb{R}$ such that*

$$\big\|\big(F_{\mathrm{PR}}(h)\big)^k (I-\tfrac{h}{2}B)^{-1}\big\| \leq M\mathrm{e}^{k\omega h} \tag{11.7}$$

holds for all $k\in\mathbb{N}$ and $h\in[0,\frac{1}{\lambda_0}]$ (i.e., the Peaceman–Rachford splitting is stable). Suppose further that the closure C of $A+B$ (with $A+B$ having the natural domain $D(A)\cap D(B)$) generates a strongly continuous semigroup on the Banach space X, and that $D(C^2)\subseteq D(AB)\cap D(A)$ and $AB(\lambda_0-C)^{-2}\in\mathscr{L}(X)$ hold. Then the Peaceman–Rachford splitting is second-order convergent on $D(C^3)$. That is, for all $t_0\geq 0$ we have

$$\|u_{\mathrm{PR},h}(t)-u(t)\| \leq h^2 K \sum_{j=0}^{3} \|C^j u_0\|$$

for all $u_0\in D(C^3)$ and $nh\in[0,t_0]$, $n\in\mathbb{N}$, $h\in[0,\frac{1}{\lambda_0}]$, where the constant K depends on t_0, but not on n and t.

Remark 11.4. The stability condition in Theorem 11.3 is satisfied for example if A and B generate contraction semigroups, and $X=H$ is a Hilbert space. Indeed, in this case we have

$$\big\|(I-hA)^{-1}\big\| \leq 1 \quad \text{and} \quad \big\|(I-hB)^{-1}\big\| \leq 1 \quad \text{for all } h>0,$$

and one can prove that also

$$\big\|(I+\tfrac{h}{2}A)(I-\tfrac{h}{2}A)^{-1}\big\| \leq 1 \quad \text{and} \quad \big\|(I+\tfrac{h}{2}B)(I-\tfrac{h}{2}B)^{-1}\big\| \leq 1$$

hold for all $h>0$, see Exercise 4.

As an illustration we give an example of operators A,B,C that satisfy the conditions of Theorem 11.1, hence, for which the Lie splitting is first-order convergent. More details and examples are left to the project phase.

Example 11.5 (Dimension splitting). Consider the heat equation in two dimensions on $\Omega=(0,1)\times(0,1)$ with homogeneous Dirichlet boundary condition:

$$\begin{cases} \partial_t w(t,x,y) = \partial_x\big(a(x,y)\partial_x w(t,x,y)\big) + \partial_y\big(b(x,y)\partial_y w(t,x,y)\big), \quad (x,y)\in\Omega,\ t>0 \\ w(0,x,y)=w_0(x,y), \quad (x,y)\in\Omega \\ w(t,x,y)=0, \quad (x,y)\in\partial\Omega,\ t>0 \end{cases} \tag{11.8}$$

with some given initial function w_0 and functions $a, b \in \mathrm{C}^2(\overline{\Omega})$ being positive on $\overline{\Omega}$. Problem (11.8) can be formulated as an abstract Cauchy problem on the Banach space $X = \mathrm{L}^2(\Omega)$:

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}u(t) = Cu(t), & t > 0 \\ u(0) = u_0 \end{cases}$$

where the operator equals $Cf = \partial_x(a\partial_x f) + \partial_y(b\partial_y f)$ for all $f \in D(C) = \mathrm{H}^2(\Omega) \cap \mathrm{H}^1_0(\Omega)$ and generates an analytic semigroup on X. First of all, we note that by the Sobolev embedding theorem[3] $\mathrm{H}^2(\Omega)$ is continuously embedded in $\mathrm{C}(\overline{\Omega})$ and $\mathrm{H}^4(\Omega)$ is continuously embedded in $\mathrm{C}^2(\overline{\Omega})$. These imply that the boundary conditions can be verified pointwise.

Now we split the operator according to the space directions, i.e., $C = A + B$ with

$$Af = \partial_x(a\partial_x f) \qquad \text{and} \qquad Bf = \partial_y(b\partial_y f)$$

for all $f \in D(C)$. In this special case the domains can be chosen as:[4]

$$\begin{aligned} & D(A) = \big\{f \in X : \partial_{xx}f, \partial_x f \in X, \text{ and } f(0,y) = f(1,y) = 0 \text{ for almost every } y \in (0,1)\big\} \\ \text{and } \; & D(B) = \big\{f \in X : \partial_{yy}f, \partial_y f \in X, \text{ and } f(x,0) = f(x,1) = 0 \text{ for almost every } x \in (0,1)\big\}. \end{aligned}$$

Then operators A, B with the corresponding domains generate analytic semigroups as well.

We have to verify now the assumptions of Theorem 11.1. The operators A, B and $A+B$ generate analytic contraction semigroups on X. To prove the domain condition, we have to show that for all $f \in D(C^2)$ we have $Bf \in D(A)$. For $f \in D(C^2) \subseteq \mathrm{H}^4(\Omega)$ we have $Bf \in \mathrm{H}^2(\Omega)$ and $f = Cf = 0$ on the boundary $\partial\Omega$ of Ω. Then $\partial_x(Bf), \partial_{xx}(Bf) \in X$. Since $f = 0$ on the boundary, its first and second weak tangential derivatives equal zero on $\partial\Omega$. Then from the continuity of Bf it follows that on the horizontal lines of $\partial\Omega$ we have $Bf = \partial_y(b\partial_y f) = Cf - \partial_x(a\partial_x f) = (\partial_x a)(\partial_x f) + a\partial_{xx}f = 0 - 0 = 0$, and on the vertical lines $Bf = \partial_y(b\partial_y f) = (\partial_y b)(\partial_y f) + b\partial_{yy}f = 0$. This yields $X \ni Bf = 0$ on $\partial\Omega$, therefore, $Bf \in D(A)$.

The space $Y := D(C^2)$ equipped with the H^4-norm becomes a Banach space, see Exercise 2. The boundedness of operator $AB(I - C)^{-2}$ follows now from the estimate

$$\big\|AB(I - C)^{-2}\big\| \le \|AB\|_{\mathscr{L}(Y,X)} \cdot \big\|(I - C)^{-2}\big\|_{\mathscr{L}(X,Y)}. \tag{11.9}$$

Indeed, for $f \in Y = D(C^2)$ we have

$$\|ABf\| = \big\|\partial_x\big(a\partial_x\big(\partial_y(b\partial_y f)\big)\big)f\big\| \le \text{const.} \cdot \|f\|_Y.$$

On the other hand for $f \in Y = D(C^2)$ we have

$$\begin{aligned} \|(I - C)^2 f\| \le \|(I - 2C + C^2)f\| \le \|f\| + 2\|Cf\| + \|C^2 f\| &\le \text{const.} \cdot (\|f\| + \|Cf\| + \|C^2 f\|) \\ &\le \text{const.} \cdot \|f\|_Y. \end{aligned}$$

Thus, we have $(I - C)^2 \in \mathscr{L}(Y, X)$ and $(I - C)^2$ is closed. By Proposition 2.10 $(I - C)^{-2}$ is closed, too, and by the closed graph theorem it is bounded, i.e., $(I - C)^{-2} \in \mathscr{L}(X, Y)$. Then estimate (11.9) yields the desired boundedness.

[3] See, e.g., Theorem 4.12 in R. A. Adams, J. J. F. Fournier, Sobolev Spaces, Elsevier, 2003.

[4] A. Ostermann, K. Schratz, "Stability of exponential operator splitting methods for non-contractive semigroups," preprint.

11.2 Operator splitting with space discretisation

In this section we consider operator splittings applied together with a space discretisation method (steps 1 and 2). That is, we assume that the semigroups generated by the sub-operators A, B are approximated by some approximate semigroups.

Consider the abstract Cauchy problem (11.1) on the Banach space X for the sum of the generators A and B. As we did in Assumption 3.2, we define approximate spaces and projection-like operators between the approximate spaces and the original space X.

Assumption 11.6. Let X_m, X be Banach spaces and assume that there are bounded linear operators $P_m : X \to X_m$, $J_m : X_m \to X$ with the following properties:

a) There is a constant $K > 0$ with $\|P_m\|$, $\|J_m\| \leq K$ for all $m \in \mathbb{N}$,

b) $P_m J_m = I_m$, the identity operator on X_m, and

c) $J_m P_m f \to f$ as $m \to \infty$ for all $f \in X$.

As already illustrated in Examples 3.3 and 3.4, the operators P_m together with the spaces X_m usually refer to a kind of space discretisation (cf. Appendix A), the spaces X_m are usually finite dimensional spaces, and the operators J_m refer to the interpolation method how we associate specific elements of the function space to the elements of the approximating spaces.

First we split the operator $C = A + B$ appearing in the original problem (11.1) into the sub-operators A and B. In order to obtain the semi-discretised systems, the sub-operators A and B have to be approximated by operators A_m and B_m for $m \in \mathbb{N}$ fixed. Suppose that the operators A_m and B_m generate the strongly continuous semigroups T_m and S_m on the space X_m, respectively. For the analysis of the convergence, we need to recall the following from Lecture 3.

Assumption 11.7. Suppose that for $m \in \mathbb{N}$ the semigroups T_m and S_m and their generators A_m, B_m satisfy the following conditions:

a) there exist constants $M \geq 1$ and $\omega \in \mathbb{R}$ such that T_m and S_m are all of type (M, ω), and for all $h > 0$, $k, m \in \mathbb{N}$ we have

$$\big\| \big(S_m(h) T_m(h)\big)^k \big\| \leq M \mathrm{e}^{k\omega h}. \tag{11.10}$$

b) We have

$$\lim_{m\to\infty} J_m A_m P_m f = Af \qquad \text{for all } f \in D(A)$$

and

$$\lim_{m\to\infty} J_m B_m P_m f = Bf \qquad \text{for all } f \in D(B).$$

The semigroups T_m, S_m, $m \in \mathbb{N}$, are called **approximate semigroups**, and their generators A_m, B_m, $m \in \mathbb{N}$, are called **approximate generators** if they possess the above properties.

Remark 11.8. From the assumption above and from the first Trotter–Kato approximation theorem, Theorem 3.14, it follows that

$$\lim_{m\to\infty} J_m T_m(h) P_m f = T(h)f$$

and

$$\lim_{m\to\infty} J_m S_m(h) P_m f = S(h)f$$

for all $f \in X$ and locally uniformly in h, where T and S are the semigroups generated by A and B, respectively.

From now on we consider the exponential splittings, that is, the sequential and Marchuk–Strang splittings. We remark that, analogously to the case of exact solutions (i.e., splitting without approximation), the stability condition (11.10) implies the stability of the reversed order and the Marchuk–Strang splittings (cf. Exercise 10.5). This means that the sequential and the Marchuk–Strang splittings fulfill their stability condition (with space discretisation) if the stability condition (11.10) holds. Therefore, it suffices to control only this condition in both cases.

Definition 11.9. For the case of spatial approximation, we define the split solutions of (11.1) as

$$u_{\mathrm{spl},n,m}(t) = J_m\big(F_{\mathrm{spl},m}(h)\big)^n P_m u_0, \quad (t = nh)$$

for $m, n \in \mathbb{N}$ fixed and for $u_0 \in X$. The operators $F_{\mathrm{seq},m}$, describing the splitting procedures together with the space discretisation method, have the form (cf. Definition 10.1):

Sequential splitting: $F_{\mathrm{seq},m}(h) = S_m(h)T_m(h)$,

Marchuk–Strang splitting: $F_{\mathrm{MS},m}(h) = T_m(h/2)S_m(h)T_m(h/2)$.

Definition 11.10. The numerical method for solving problem (11.1) described above is **convergent at a fixed time level** $t > 0$ if for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n, m \geq N$ we have

$$\|u_{\mathrm{spl},n,m}(t) - u(t)\| \leq \varepsilon.$$

This is the usual well-known notion of the convergence of a sequence with two indices and we will use the notation

$$\lim_{n,m\to\infty} u_{\mathrm{spl},n,m}(t) = u(t)$$

to express this.

In order to prove the convergence of operator splitting in this case, we state a modified version of Chernoff's theorem, Theorem 5.12, which is applicable for approximate semigroups as well.

Theorem 11.11 (Modified Chernoff Theorem)**.** *Consider a sequence of strongly continuous functions $F_m : [0,\infty) \to \mathscr{L}(X_m)$, $m \in \mathbb{N}$, satisfying*

$$F_m(0) = I_m \tag{11.11}$$

for all $m \in \mathbb{N}$, and suppose that there exists constants $M \geq 1$, $\omega \in \mathbb{R}$ such that

$$\big\|\big(F_m(t)\big)^k\big\| \leq M\mathrm{e}^{k\omega t} \tag{11.12}$$

holds for all $t \geq 0$ and $m, k \in \mathbb{N}$. Suppose further that the limit

$$\lim_{m\to\infty} \frac{J_m F_m(h) P_m f - J_m P_m f}{h}$$

exists uniformly in $h \in (0, t_0]$, and that

$$Gf := \lim_{h\searrow 0} \lim_{m\to\infty} \frac{J_m F_m(h) P_m f - J_m P_m f}{h} \tag{11.13}$$

exists for all $f \in Y \subset X$, where Y and $(\lambda_0 - G)Y$ are dense subspaces in X for $\lambda_0 > 0$. Then the closure $C = \overline{G}$ of G generates a bounded strongly continuous semigroup U, which is given by

$$U(t)f = \lim_{n,m\to\infty} J_m\big(F_m(h)\big)^n P_m f \tag{11.14}$$

for all $f \in X$ uniformly for t in compact intervals ($t = nh$).

Proof. For $h > 0$ we define

$$G_{h,m} := \frac{F_m(h) - I_m}{h} \in \mathscr{L}(X_m)$$

for all fixed $h \in (0, t_0]$ and $m \in \mathbb{N}$. Observe that for all $f \in Y$ we have

$$\lim_{h \searrow 0} \lim_{m\to\infty} J_m G_{h,m} P_m f = Gf.$$

Then every semigroup $(\mathrm{e}^{tG_{h,m}})_{t\geq 0}$ satisfies

$$\big\|\mathrm{e}^{tG_{h,m}}\big\| \leq \mathrm{e}^{-\frac{t}{h}} \big\|\mathrm{e}^{\frac{t}{h}F_m(h)}\big\| \leq \mathrm{e}^{-\frac{t}{h}} \sum_{k=0}^{\infty} \frac{t^k}{h^k k!} \big\|\big(F_m(h)\big)^k\big\| \leq M\mathrm{e}^{\omega' t} \tag{11.15}$$

for some $\omega' \in \mathbb{R}$ and for every fixed h and m. This shows that the assumptions of the first Trotter–Kato approximation theorem, Theorem 3.14, are fulfilled. Hence we can take the limit in $m \to \infty$ (which is uniform in $h \in (0, t_0]$), and then take limit as $h \to 0$ obtaining that the closure $\overline{G}$ of G generates a strongly continuous semigroup U given by

$$\lim_{h \searrow 0} \lim_{m\to\infty} \big\|U(t)f - J_m \mathrm{e}^{tG_{h,m}} P_m f\big\| = 0 \tag{11.16}$$

for all $f \in X$ and uniformly for t in compact intervals. On the other hand, we have for $t = nh$ by Lemma 5.7 that

$$\begin{aligned}\big\|J_m \mathrm{e}^{tG_{h,m}} P_m f - J_m\big(F_m(h)\big)^n P_m f\big\| &= \Big\|J_m \mathrm{e}^{n(F_m(h)-I_m)} P_m f - J_m\big(F_m(h)\big)^n P_m f\Big\| \\ \leq \sqrt{n}M\|J_m F_m(h) P_m f - J_m P_m f\| &= \frac{\sqrt{n}}{h} M \left\|\frac{J_m F_m(h) P_m f - J_m P_m f}{h}\right\|.\end{aligned} \tag{11.17}$$

Using that $h = \frac{t}{n}$ for some $t \in [0, t_0]$ and taking the limit, we obtain

$$\lim_{h \searrow 0} \lim_{m\to\infty} \big\|J_m \mathrm{e}^{tG_{h,m}} P_m f - J_m\big(F_m(h)\big)^n P_m f\big\| = \lim_{h \searrow 0} \lim_{m\to\infty} \frac{tM}{\sqrt{n}} \left\|\frac{J_m F_m(h) P_m f - J_m P_m f}{h}\right\| = 0 \tag{11.18}$$

for all $f \in Y$, and uniformly for t in compact intervals. The combination of (11.16) and (11.18) yields

$$\begin{aligned}&\lim_{h \searrow 0} \lim_{m\to\infty} \big\|U(t)f - J_m\big(F_m(h)\big)^n P_m f\big\| \\ &\quad\leq \lim_{h \searrow 0} \lim_{m\to\infty} \big\|U(t)f - J_m \mathrm{e}^{tG_{h,m}} P_m f\big\| + \lim_{h \searrow 0} \lim_{m\to\infty} \big\|J_m \mathrm{e}^{tG_{h,m}} P_m f - J_m\big(F_m(h)\big)^n P_m f\big\| = 0\end{aligned}$$

$(t = nh)$ for all $f \in Y$, and uniformly for t in compact intervals. By Theorem 2.30, the statement follows for all $f \in X$. □

In the rest of this lecture, we consider the convergence of the sequential splitting applied together with a space discretisation method.

Lemma 11.12. *Let J_m, P_m, T_m be operators introduced in Assumptions 11.6 and 11.7. Then*

$$\lim_{m\to\infty} \frac{J_m T_m(h) P_m f - J_m P_m f}{h} = \tfrac{1}{h}\big(T(h)f - f\big)$$

holds for all $f \in D(A)$ uniformly in $h \in (0, t_0]$, and

$$\lim_{h \searrow 0} \tfrac{1}{h}\big(T(h)f - f\big) = Af.$$

Proof. Let us investigate the following difference for all $f \in D(A)$:

$$\begin{aligned}
&\left\| \frac{J_m T_m(h) P_m f - J_m P_m f}{h} - \frac{T(h)f - f}{h} \right\| = \frac{1}{h} \left\| \int\limits_0^h J_m A_m T_m(s) P_m f \,\mathrm{d}s - \int\limits_0^h AT(s) f \,\mathrm{d}s \right\| \\
&\quad \le \sup_{s \in [0,t_0]} \|J_m A_m T_m(s) P_m f - AT(s) f\| = \sup_{s \in [0,t_0]} \|J_m T_m(s) A_m P_m f - T(s) Af\| \\
&\quad \le \sup_{s \in [0,t_0]} \big\| J_m T_m(s) P_m (J_m A_m P_m f - Af) + \big(J_m T_m(s) P_m - T(s)\big) Af \big\| \\
&\quad \le \sup_{s \in [0,t_0]} \|J_m\| \cdot \|T_m(s)\| \cdot \|P_m\| \cdot \|J_m A_m P_m f - Af\| + \sup_{s \in [0,t_0]} \big\| \big(J_m T_m(s) P_m - T(s)\big) Af \big\|.
\end{aligned}$$

By Assumption 11.7, the term $\|J_m A_m P_m f - Af\|$ tends to 0 as m tends to infinity. Since $g := Af$ is a fixed element in the Banach space X, $\|J_m T_m(s) P_m g - T(s) g\|$ tends to 0 uniformly in h as $m \to \infty$ because of Remark 11.8. Operators J_m and P_m were assumed to be bounded. The semigroups T_m are of type (M, ω), independently of m. Therefore,

$$\sup_{s \in [0,t_0]} \|T_m(s)\| \le \sup_{s \in [0,t_0]} M \mathrm{e}^{\omega s} \le M \mathrm{e}^{\max\{0,\omega\} t_0} = \text{const.} < \infty.$$

Hence, the difference above tends to 0 uniformly in h. The second limit as $h \searrow 0$ can be obtained by using the definition of the generator:

$$\lim_{h \searrow 0} \frac{T(h)f - f}{h} = Af \qquad \text{for all } f \in D(A).$$

Thus, the statement is proved. □

The same result is true for the semigroup S generated by the operator B, that is,

$$\lim_{h \searrow 0} \lim_{m \to \infty} \frac{J_m S_m(h) P_m f - J_m P_m f}{h} = Bf \tag{11.19}$$

holds for all $f \in D(B)$, where the limit as $m \to \infty$ is locally uniform in h.

Theorem 11.13. *The sequential splitting is convergent at time level $t > 0$ if the stability condition* (11.10) *holds for the approximate semigroups, and the approximate generators satisfy Assumption 11.7.*

Proof. According to the modified Chernoff theorem, Theorem 11.11, the sequential splitting is convergent if the stability (11.12) and the consistency (11.13) hold for the operator

$$F_m(h) = S_m(h) T_m(h). \tag{11.20}$$

The stability condition (11.12) is fulfilled, since we assumed that (11.10) holds. In order to prove the consistency criterion (11.13), we investigate the following limit:

$$\begin{aligned}
&\lim_{h \searrow 0} \lim_{m \to \infty} \frac{J_m S_m(h) T_m(h) P_m f - J_m P_m f}{h} \\
&\quad = \lim_{h \searrow 0} \lim_{m \to \infty} J_m S_m(h) P_m \frac{J_m T_m(h) P_m f - J_m P_m f}{h} \\
&\quad\quad + \lim_{h \searrow 0} \lim_{m \to \infty} \frac{J_m S_m(h) P_m f - J_m P_m f}{h}.
\end{aligned}$$

Remark 11.8 implies

$$\lim_{m\to\infty} J_m S_m(h) P_m f = S(h)f \quad \text{for all } f \in X \text{ and uniformly for } h \in [0, t_0]$$

and $$\lim_{h\searrow 0} S(h)f = f \quad \text{for all } f \in X.$$

Notice further that the set $\left\{\frac{1}{h}(J_m T_m(h)P_m f - J_m P_m f) : h \in (0, t_0]\right\}$ is relatively compact for all $f \in D(A)$, and that on compact sets the strong and the uniform convergence is equivalent due to Theorem 2.30. Then Lemma 11.12 and (11.19) imply that

$$\lim_{h\searrow 0} \lim_{m\to\infty} \frac{J_m F_m(h) P_m f - J_m P_m f}{h} = (A+B)f$$

holds for all $f \in D(A) \cap D(B)$ (see also the proof of Corollary 4.10). This completes the proof. □

We state now the convergence of the Marchuk–Strang splitting, and leave the proof as Exercise 5.

Theorem 11.14. *The Marchuk–Strang splitting is convergent at time level $t > 0$ if the stability condition* (11.10) *holds for the approximate semigroups, and the approximate generators satisfy Assumption 11.7.*

Exercises

1. Prove the recurrence relation (11.5).

2. Let C be the operator from Example 11.5. Prove that the H^4-norm makes $D(C^2)$ a Banach space.

3. Consider the operators A, B and C from Example 11.5 with $a = b = 1$. Show that they generate analytic contraction semigroups.

4. Suppose A generates a contraction semigroup on the Hilbert space H. Prove that the **Cayley transform**

$$G = (I + A)(I - A)^{-1}$$

of A is a contraction.

5. Prove Theorem 11.14.

Lecture 12

Rational Approximations

In the previous lectures we have seen some examples for time discretisation methods, e.g., the explicit and implicit Euler methods, the Crank–Nicolson scheme, and the Radau II A method. We now turn to study numerical approximation schemes $F : [0, \infty) \to \mathscr{L}(X)$ (see Lecture 4) that are defined by means of a rational function r:

$$F(h) = r(hA).$$

Such were the previously mentioned time discretisation methods. In general, we first need to give meaning to the expression $r(hA)$, and—in view of the Lax equivalence theorem, Theorem 4.6—to study consistency and stability of these schemes. We start with the scalar case and make the following definition. Let $r : \mathbb{C} \to \mathbb{C}$ be a rational function, i.e., $r = \frac{P}{Q}$ with P, Q polynomials and $Q \neq 0$, and even if it is not stated we shall usually suppose that P and Q have no common zeros.

Definition 12.1. We call a rational function r a rational approximation of the exponential function of order p, **rational approximation of order** p for short, if there are constants $C, \delta > 0$ such that

$$|r(z) - \mathrm{e}^z| \leq C|z|^{p+1} \quad \text{for all } z \in \mathbb{C} \text{ with } |z| \leq \delta.$$

12.1 The scalar case

Recall the test equation, already seen in Section 1.1 and Appendix B, with the unknown function $u : [0, \infty) \to \mathbb{C}$:

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} u(t) = \lambda u(t), & t > 0 \\ \quad u(0) = u_0, \end{cases} \tag{12.1}$$

where the parameter $\lambda \in \mathbb{C}$ and the initial value $u_0 \in \mathbb{C}$ are given. Of course, the exact solution of problem (12.1) equals $u(t) = \mathrm{e}^{t\lambda} u_0$.

1. Consistency

That r is a rational approximation of order p means by definition that $F(h) := r(h\lambda)$ is a *finite difference scheme (method)* consistent of order p on $X = \mathbb{R}$ with the Cauchy problem (12.1), cf. Definition 4.11. By considering the power series expansion around $z = 0$ one sees immediately that p-order consistency in this case is equivalent to the conditions

$$r(0) = 1,\ r'(0) = 1,\ \ldots,\ r^{(p)}(0) = \exp^{(p)}(0).$$

This yields the next example.

Example 12.2. Consider

$$r(z) := 1 + z + \frac{z^2}{2!} + \dots \dots + \frac{z^s}{s!}.$$

Trivially, r is a rational approximation of order $p = s$. All **explicit s-stage Runge–Kutta methods** of order $p = s$ possess this stability function, see also Example 12.3 below.

If we apply the time discretisation method $F(h) = r(h\lambda)$ to obtain the numerical solution $u_{h,n}$ after n steps with time step h we are led to the recursion

$$u_{h,n} = r(h\lambda)u_{h,n-1}, \quad n \in \mathbb{N}.$$

Next we show that Runge–Kutta methods are of this form.

Example 12.3 (Runge–Kutta methods). Recall the s-stage Runge–Kutta method from Appendix B given by the recursion (B.13), (B.14). For the special right-hand side of problem (12.1) we then have:

$$u_{h,n} = u_{h,n-1} + h\lambda \sum_{i=1}^{s} b_i k_i \tag{12.2}$$

with

$$k_i = u_{h,n-1} + h\lambda \sum_{j=1}^{s} a_{ij} k_j \tag{12.3}$$

with certain coefficients a_{ij}, b_i for $i, j = 1, ..., s$. We introduce the following vectors in $\mathbb{R}^s$:

$$\mathbf{k} = (k_1, ..., k_s)^\top, \quad \mathbf{1} = (1, ..., 1)^\top, \quad \mathbf{b} = (b_1, ..., b_s)^\top,$$

and the matrix $\mathbf{A} = (a_{ij})_{i,j=1,...,s} \in \mathbb{R}^{s\times s}$. Then formulae (12.2) and (12.3) can be written as

$$u_{h,n} = u_{h,n-1} + z\mathbf{b}^\top k$$

and

$$\mathbf{k} = (1 - z\mathbf{A})^{-1}\mathbf{1}u_{h,n-1} \tag{12.4}$$

with $z = h\lambda \in \mathbb{C}$. This implies for all $n \in \mathbb{N}$ that

$$u_{h,n} = u_{h,n-1} + z\mathbf{b}^\top (I - z\mathbf{A})^{-1}\mathbf{1}u_{h,n-1} = \big(1 + z\mathbf{b}^\top (I - z\mathbf{A})^{-1}\mathbf{1}\big)u_{h,n-1}, \tag{12.5}$$

that is, we obtain $u_{h,n} = r(z)u_{h,n-1}$ with $r(z) = 1 + z\mathbf{b}^\top (I - z\mathbf{A})^{-1}\mathbf{1}$ which is a rational function of $z \in \mathbb{C}$. To work out the details of the computations above is left as Exercise 1.

Example 12.4. All the time discretisation methods introduced previously are Runge–Kutta methods. Therefore, the corresponding rational function can be obtained by the derivation in Example 12.3.

explicit Euler method:	$r(z) = 1 + z$
implicit Euler method:	$r(z) = \dfrac{1}{1 - z}$
Crank–Nicolson scheme:	$r(z) = \dfrac{1 + \frac{z}{2}}{1 - \frac{z}{2}}$
Radau II A method:	$r(z) = \dfrac{1 + \frac{2}{5}z + \frac{1}{10}\frac{z^2}{2}}{1 - \frac{3}{5}z + \frac{3}{10}\frac{z^2}{2} - \frac{1}{10}\frac{z^3}{6}}.$

Now let us return to general rational functions

$$r(z) = \frac{P(z)}{Q(z)},$$

with $k = \deg(P)$ and $l = \deg(Q)$, where we suppose that P and Q have no common zeros. If $k, l \in \mathbb{N}_0$ are fixed, the maximal order of approximation to the exponential function is $p = k + l$, see Exercise 5. Such rational approximations r are called **rational Padé approximations**. One can collect the corresponding functions r in the **Padé tableau**, see Table 12.1 for examples.

$l \backslash k$	0	1	2
0	$\dfrac{1}{1}$	$\dfrac{1+z}{1}$	$\dfrac{1+z+\frac{z^2}{2!}}{1}$
1	$\dfrac{1}{1-z}$	$\dfrac{1+\frac{1}{2}z}{1-\frac{1}{2}z}$	$\dfrac{1+\frac{2}{3}z+\frac{2}{3}\frac{z^2}{2!}}{1-\frac{1}{3}z}$
2	$\dfrac{1}{1-z+\frac{z^2}{2!}}$	$\dfrac{1+\frac{1}{3}z}{1-\frac{1}{3}z+\frac{1}{3}\frac{z^2}{2!}}$	$\dfrac{1+\frac{1}{2}z+\frac{1}{6}\frac{z^2}{2!}}{1-\frac{1}{2}z+\frac{1}{6}\frac{z^2}{2!}}$
3	$\dfrac{1}{1-z+\frac{z^2}{2!}-\frac{z^3}{3!}}$	$\dfrac{1+\frac{1}{4}z}{1-\frac{1}{4}z+\frac{1}{2}\frac{z^2}{2!}-\frac{1}{4}\frac{z^3}{3!}}$	$\dfrac{1+\frac{2}{5}z+\frac{1}{10}\frac{z^2}{2!}}{1-\frac{3}{5}z+\frac{3}{10}\frac{z^2}{2!}-\frac{1}{10}\frac{z^3}{3!}}$

Table 12.1: Padé tableau.

2. Stability issues

As discussed in Lecture 4, stability is fundamental if one longs for convergence of the method for all initial values. More precisely, since $u_{h,n} = \big(r(h\lambda)\big)^n u_0$ is expected to be the approximation of the exact solution $u(t) = \mathrm{e}^{t\lambda}u_0$ at time $t = nh$, the recursion $u_{h,n} = r(h\lambda)u_{h,n-1}$ needs to be stable. This motivates the next definition. The set

$$S = S(r) = \big\{z \in \mathbb{C} : |r(z)| \leq 1\big\}$$

is called the **stability region** of the corresponding rational approximation. Also note that, if one starts, say with some Runge–Kutta method as in Example 12.3, and derives a formula for the recursion, the appearing rational function r determines the stability of the method. Hence the rational function is also called **stability function**.

Example 12.5. Consider the following time discretisation methods, their stability functions, and stability regions.

1. For the explicit Euler method we have $r_1(z) = 1 + z$, which implies

   $$S(r_1) = \{z \in \mathbb{C} : |1 + z| \leq 1\}$$

   the closed disc of radius 1, centred at the point -1.

2. The implicit Euler method has stability function $r_2(z) = \frac{1}{1-z}$, hence,

   $$S(r_2) = \{z \in \mathbb{C} : |1 - z| \geq 1\},$$

   which is the exterior of the circle with radius 1 and centre 1.

3. The stability function of the Crank–Nicolson scheme is $r_3(z) = \frac{1+\frac{z}{2}}{1-\frac{z}{2}}$, therefore,

$$S(r_3) = \{z \in \mathbb{C} : \operatorname{Re} z \leq 0\},$$

   i.e., the left half-plane.

The respective stability regions are shown in Figure 12.1.

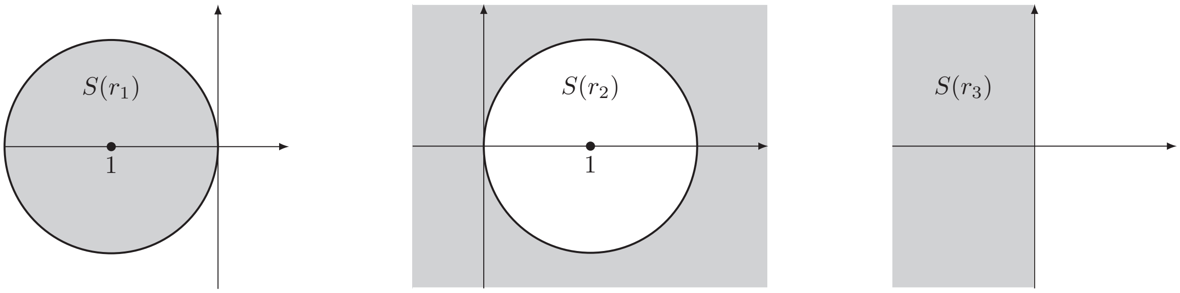

Figure 12.1: Stability regions: explicit Euler, implicit Euler, Crank–Nicolson methods.

From the examples above one can see that the stability of the recursion is not obvious. There is a restrictive condition on $z = h\lambda$. To achieve a stable recursion $u_{h,n} = r(z)u_{h,n-1}$, $n \in \mathbb{N}$, the complex number $z = h\lambda$ has to lie in the stability region S of the method. Since $\lambda \in \mathbb{C}$ is a given parameter in problem (12.6), this yields a condition on the step size h. Thus, if $\operatorname{Re} \lambda \leq 0$, the explicit Euler method is not unconditionally stable in contrast to the the implicit Euler or Crank–Nicolson schemes (cf. Section B.1).

The rational approximations with stability region containing the entire left half-plane are called ***A*-stable**[1]. If the stability region contains a sector

$$\overline{\reflectbox{Σ}}_\alpha = \{z \in \mathbb{C} : |\arg(-z)| \leq \alpha\} = -\overline{\Sigma}_\alpha,$$

for some $\alpha \in [0, \frac{\pi}{2}]$, we speak about **$A(\alpha)$-stability**. (For $\alpha = 0$ we set $\reflectbox{$\Sigma$}_0 = (-\infty, 0)$ and $\Sigma_0 = (0, \infty)$.) Notice that A-stability is the same as $A(\frac{\pi}{2})$-stability. For example, the implicit Euler or Crank–Nicolson schemes are A-stable.

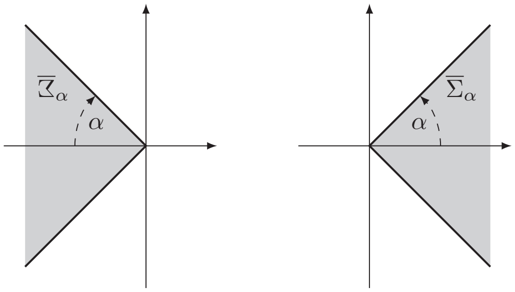
$\overline{\reflectbox{$\Sigma$}}_\alpha$
α
$\overline{\Sigma}_\alpha$
α

Figure 12.2: The sector Σ_α and its reflection.

[1] The terminology is due to G. Dahlquist. According to Hairer, Nørsett and Wanner he said: ‘I didn’t like all these “strong”, “perfect”, “absolute”, “generalized”, “super”, “hyper”, “complete” and so on in mathematical definitions, I wanted something neutral; and having been impressed by David Young’s “property A”, I chose the term “A-stable”.’

Without proof we state an important stability property of Padé approximations first conjectured by Ehle[2], and then proved by Wanner, Hairer, and Nørsett.[3]

Theorem 12.6. *A rational Padé approximation $r = \frac{P}{Q}$ is A-stable if and only if*

$$\deg(Q) - 2 \leq \deg(P) \leq \deg(Q).$$

That is, only those Padé approximations are A-stable whose stability functions appear in the main diagonal or in the first or the second lower sub-diagonal of the Padé tableau.

Remark 12.7. Let $r = \frac{P}{Q}$ be an A-stable rational Padé approximation. If r is diagonal, i.e., $\deg(P) = \deg(Q)$, then $r(\infty) = 1$, otherwise $r(\infty) = 0$.

3. Convergence

The first motivating result shows that $A(\alpha)$-stable rational approximations converge in the scalar case. However, this result will be fundamental later, when we pass to rational approximation of analytic semigroups.

Proposition 12.8. *Let $\alpha \in (0, \frac{\pi}{2}]$, and let r be an $A(\alpha)$-stable rational approximation of order $p \in \mathbb{N}$. Then for all $\theta \in (0, \alpha)$ and $\varepsilon_0 > 0$ there exist constants $C, c \geq 0$ such that*

$$|r(z)^n - \mathrm{e}^{nz}| \leq Cn|z|^{p+1}\mathrm{e}^{-nc|z|} \quad \textit{holds for all } z \in \overline{\Sigma}_\theta \textit{ with } |z| \leq h_0, \textit{ and for all } n \in \mathbb{N}.$$

In particular, we have for all $\lambda \in \Sigma_\alpha$ a constant $K > 0$ such that

$$|r(h\lambda)^n - \mathrm{e}^{t\lambda}| \leq Kn|h|^{p+1}\mathrm{e}^{-nhc} = K\mathrm{e}^{-tc}\frac{t^p}{n^p} \quad (t = nh)$$

for all $h \geq 0$ and $n \in \mathbb{N}$, i.e., the method is convergent of order p.

Proof. First of all, let us fix $C' \geq 0$ so that

$$|r(z) - \mathrm{e}^z| \leq C'|z|^{p+1} \quad \text{for all } z \in \overline{\Sigma}_\alpha \text{ with } |z| \leq h_0.$$

This is possible by the assumption about the approximation order. Note that

$$|\mathrm{e}^z| = \mathrm{e}^{\operatorname{Re} z} \leq \mathrm{e}^{-|z|\cos(\theta)} \quad \text{holds for all } z \in \overline{\Sigma}_\theta.$$

We next claim that for some $c' > 0$ the inequality

$$|r(z)| \leq \mathrm{e}^{-c'|z|} \quad \text{holds for all } z \in \overline{\Sigma}_\theta \text{ with } |z| \leq h_0.$$

We argue by contradiction and assume the contrary, i.e., that for all $n \in \mathbb{N}$ there is $z_n \in \overline{\Sigma}_\theta$ with $|z_n| \leq h_0$ such that

$$|r(z_n)| > \mathrm{e}^{-\frac{|z_n|}{n}}.$$

By passing to a subsequence we may assume that $(z_n) \subseteq \overline{\Sigma}_\theta \cap \overline{\mathrm{B}}(0, h_0)$ is convergent to a limit $z \in \overline{\Sigma}_\theta \cap \overline{\mathrm{B}}(0, h_0)$. Then we obtain $|r(z)| \geq 1$ and the $A(\alpha)$-stability yields $|r(z)| = 1$. By the

[2] B. L. Ehle, "*A*-stable methods and Padé approximations to the exponential," SIAM J. Math. Anal. **4** (1973), 671–680.

[3] G. Wanner, E. Hairer and S. P. Nørsett, "Order stars and stability theorems," BIT Num. Math. **18** (1978), 475–489.

maximum principle for the modulus of holomorphic functions (applied to r on $-\Sigma_\alpha$) we obtain that $z = 0$. Thus we conclude

$$\mathrm{e}^{-\frac{|z_n|}{n}} \leq |r(z_n)| \leq |r(z_n) - \mathrm{e}^{z_n}| + |\mathrm{e}^{z_n}| \leq C'|z_n|^{p+1} + \mathrm{e}^{-|z_n|\cos(\theta)}.$$

Therefore

$$\tfrac{1}{|z_n|}\Big(\mathrm{e}^{|z_n|(\cos(\theta)-\frac{1}{n})} - 1\Big) \leq C'|z_n|^p \mathrm{e}^{|z_n|\cos(\theta)} \to 0$$

as $n \to \infty$. This yields, however, a contradiction.

We obtain therefore the existence of a $c' > 0$ asserted above, and we set $c := \min\{c', \cos(\theta)\}$. By the standard telescopic identity we conclude

$$\begin{aligned}\big|r(z)^n - \mathrm{e}^{nz}\big| &\leq \big|r(z) - \mathrm{e}^z\big| \sum_{j=0}^{n-1} \big|r(z)\big|^j \big|\mathrm{e}^{(n-j-1)z}\big| \leq \big|r(z) - \mathrm{e}^z\big| \sum_{j=0}^{n-1} \mathrm{e}^{-cj|z|}\mathrm{e}^{-c(n-j-1)|z|} \\ &\leq C'|z|^{p+1}\mathrm{e}^{-c(n-1)|z|} \leq C'\mathrm{e}^c|z|^{p+1}\mathrm{e}^{-nc|z|} = C|z|^{p+1}\mathrm{e}^{-nc|z|}\end{aligned}$$

for all $z \in \overline{\Sigma}_\theta$ with $|z| \leq h_0$. □

12.2 Rational functions of operators

Let A be a linear operator on a Banach space X with nonempty resolvent set. Given a rational function $r = \frac{P}{Q}$ we would like to define $r(A)$. First of all, we recall the case when $r = P$ is a polynomial. Suppose $P(z) = z^k$. In this case, as we have already seen, we set $D(A^0) = X$ and $A^0 = I$, and for $k \in \mathbb{N}$ we define

$$\begin{aligned} D(A^k) &:= \big\{f \in D(A^{k-1}) : A^{k-1}f \in D(A)\big\}, \\ A^k f &= AA^{k-1}f \quad \text{for } f \in D(A^k)\end{aligned}$$

by recursion. Then A^k is a closed operator for every $k \in \mathbb{N}_0$, cf. Exercise 4.1. For a general polynomial $P \neq 0$

$$P(z) = a_0 + a_1 z + a_z^2 + \ldots + a_k z^k$$

with $a_k \neq 0$ we set $D(P(A)) := D(A^k)$ and

$$P(A) := a_0 I + a_1 A + a_2 A^2 + \ldots + a_k A^k,$$

which is again a closed operator, see Exercise 3.

Of course, we can write

$$P(z) = a_k(z - z_1)^{m_1}(z - z_2)^{m_2} \cdots (z - z_n)^{m_n}$$

where $z_j \in \mathbb{C}$ are pairwise different. Thus we have the identity

$$P(A) = a_k(A - z_1)^{m_1}(A - z_2)^{m_2} \cdots (A - z_n)^{m_n}.$$

If $P \neq 0$ and the zeros of P all lie in $\rho(A)$ then $(A - z_j)$ are in particular all injective, and we obtain

$$P(A)^{-1} = \frac{1}{a_k}(A - z_1)^{-m_1} \cdots (A - z_n)^{-m_n} = \frac{(-1)^k}{a_k} R(z_1, A)^{m_1} \cdots R(z_n, A)^{m_n} \in \mathscr{L}(X).$$

It is easy to see that $\operatorname{ran}(P(A)^{-1}) = D(A^{\deg(P)})$.

Next we define

$$\mathcal{R}_A = \big\{ r = \tfrac{P}{Q} : P, Q \text{ are polynomials with } Q \text{ having all zeros in } \rho(A) \big\}$$

and for $r \in \mathcal{R}_A$, $r \neq 0$ we set

$$r(A) := P(A)Q(A)^{-1},$$

with

$$D(r(A)) = \big\{ f \in X : Q(A)^{-1} f \in D(P(A)) \big\}.$$

Then $r(A)$ is a closed operator by Exercise 7.1, and we have

$$D(r(A)) = \begin{cases} D\big(A^{\deg(P)-\deg(Q)}\big) & \text{if } \deg(P) \geq \deg(Q) \\ X & \text{otherwise.} \end{cases}$$

Note that $r(A)$ is well-defined, i.e., if $r = \frac{P_1}{Q_1} = \frac{P_2}{Q_2}$ with Q_1, Q_2 having zeros in $\rho(A)$ then

$$P_1(A)Q_1(A)^{-1} = P_2(A)Q_2(A)^{-1}.$$

Finally, let us recall the **partial fraction decomposition** of a rational function. If r is a rational function with poles z_i of order $\nu_i \in \mathbb{N}$, then there is a unique polynomial P_0 and coefficients $c_{ij} \in \mathbb{C}$ such that

$$r(z) = P_0(z) + \sum_{i=1}^{\nu} \sum_{j=1}^{\nu_i} \frac{c_{ij}}{(z - z_i)^j}.$$

This provides yet another evaluation of $r(A)$ for $r \in \mathcal{R}_A$:

$$r(A) = P_0(A) + \sum_{i=1}^{\nu} \sum_{j=1}^{\nu_i} (-1)^j c_{ij} R(z_i, A)^j.$$

We can now at last define $r(hA)$.

Definition 12.9. A finite difference scheme F such that $F(h) = r(hA)$ holds for $h \in [0, h_0]$ with some $h_0 > 0$ is called a **rational approximation scheme**.

We now have a simple **functional calculus** for $r \in \mathcal{R}_A$. For the study of its various algebraic and analytic properties we refer to the Appendix A.6 of the monograph[4] by M. Haase. However to obtain the convergence of the rational approximation method obtained from $r \in \mathcal{R}_A$ further structural properties of r and A are needed. This will be the subject of forthcoming lectures. In the present one we shall give some illustration of results that can be expected, in a situation that is quite near to the scalar case.

[4] M. Haase: The Functional Calculus for Sectorial Operators, vol. 169 of Operator Theory: Advances and Applications, Birkhäuser Basel, 2006.

12.3 Multiplication operators

In this section we consider multiplication operators and start by briefly recalling some results from Exercises 1.4, 7.2 and Examples 7.2, 9.6. Let $(m_n) \subseteq \mathbb{C}$ be a sequence, and consider the multiplication operator $A = M_m$ with maximal domain

$$D(M_m) := \big\{(x_k) \in \ell^2 : (m_k x_k) \in \ell^2\big\}.$$

The norm of M_m is

$$\|M_m\| = \sup_{k\in\mathbb{N}} |m_k|,$$

provided the latter expression is finite. Our standing assumption will be that

$$\big\{m_n : n \in \mathbb{N}\big\} \subseteq \overline{\Sigma}_\delta$$

holds for some $\delta \in [0, \frac{\pi}{2})$. Or in other words:

Assumption 12.10. Let $A = M_m$ be a multiplication operator with spectrum

$$\sigma(M_m) = \overline{\big\{m_n : n \in \mathbb{N}\big\}} \subseteq \overline{\Sigma}_\delta$$

for some $\delta \in [0, \frac{\pi}{2})$.

Under this condition $A = M_m$ generates an analytic contraction semigroup T given by

$$T(t) = \mathrm{e}^{tA} = M_{\mathrm{e}^{tm}}.$$

For $\beta \geq 0$ the fractional power $(-A)^\beta$ of $-A = -M_m$ is given by

$$(-A)^\beta = M_{(-m)^\beta} \quad \text{with maximal domain} \quad D(M_{(-m)^\beta}) = \big\{(x_n) \in \ell^2 : ((-m_k)^\beta x_k) \in \ell^2\big\}.$$

Consider now the abstract Cauchy problem

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} u(t) = Au(t), & t > 0 \\ u(0) = u_0 \end{cases} \tag{12.6}$$

on the Banach space $X = \ell^2$ with $u_0 \in D(A)$. We use some rational approximation schemes to obtain the numerical solution $u_{h,n}$ at time t, i.e., after n steps with time step $h = \frac{t}{n}$.

The first illustrating result states the stability of some suitable schemes and explains why A-stability may be relevant in the general situation of rational approximation schemes.

Proposition 12.11 (**Stability theorem**). *Suppose $A = M_m$ is as above, and let r be an $A(\delta)$-stable rational approximation. Then the estimate*

$$\big\|\big(r(hA)\big)^n\big\| \leq 1$$

holds for all $h > 0$ and $n \in \mathbb{N}$.

Proof. Note that for $z \in \overline{\Sigma}_\delta$ and $h \geq 0$ we have $zh \in \overline{\Sigma}_\delta$. Then by the preparatory remarks we have

$$\big\|\big(r(hA)\big)^n\big\| \leq \sup_{k\in\mathbb{N}} \big\|\big(r(hm_k)\big)^n f\big\| \leq \sup_{z\in\overline{\Sigma}_\delta} |r(z)| \leq 1. \qquad \square$$

□

We can extend the convergence result from the scalar case as follows. Our inspiration is the paper by M.-N. Le Roux[5].

Theorem 12.12 (**Convergence theorem I.**). *Suppose $A = M_m$ is as above. Let r be the stability function of an $A(\alpha)$-stable rational approximation of order p with $|r(\infty)| < 1$ and $\alpha \in (\delta, \frac{\pi}{2}]$. Then there is a constant $K > 0$ such that*

$$\big\|r(hA)^n - \mathrm{e}^{tA}\big\| \leq K\frac{h^p}{t^p} = \frac{K}{n^p} \qquad (t = nh)$$

holds for all $n \in \mathbb{N}$, $t \geq 0$, i.e., one has the convergence of the rational approximation method in the operator norm.

Proof. We have to estimate

$$\big\|r(hA)^n - \mathrm{e}^{tA}\big\| = \sup_{k\in\mathbb{N}}\big|r(hm_k)^n - \mathrm{e}^{tm_k}\big| = \sup_{k\in\mathbb{N}}\big|r(hm_k)^n - \mathrm{e}^{nhm_k}\big|.$$

Since $|r(\infty)| < 1$ we can choose $h_0 > 0$ so large that

$$\sup\big\{z \in \overline{\Sigma}_\delta : |z| \geq h_0\big\} =: r_0 < 1.$$

Suppose first $|hm_k| \leq h_0$. Then by Proposition 12.8 we obtain that

$$\big|r(hm_k)^n - \mathrm{e}^{tm_k}\big| \leq Cn|hm_k|^{p+1}\mathrm{e}^{-nhc|m_k|} = \frac{C}{n^p}|tm_k|^{p+1}\mathrm{e}^{-tc|m_k|} \leq \frac{C'}{n^p}$$

for some constants $C', c > 0$. On the other hand, if $|hm_k| > h_0$, then

$$|r(hm_k)| \leq r_0 < 1.$$

Therefore with some appropriate constant $C'' > 0$ we have

$$|r(hm_k)|^n \leq r_0^n \leq \frac{C''}{n^p}.$$

We also have

$$|\mathrm{e}^{nhm_k}| = \mathrm{e}^{nh\,\mathrm{Re}\,m_k} \leq \mathrm{e}^{-nh\cos(\alpha)|m_k|} = \mathrm{e}^{-nh_0\cos(\alpha)} \leq \frac{C'''}{n^p}.$$

Hence in case $|hm_k| \geq h_0$ we obtain

$$|r(hm_k)^n - \mathrm{e}^{tm_k}\big| \leq \frac{C''' + C''}{n^p}.$$

This and the estimate in the first case finish the proof. □

The drawback of this result is that it tells nothing about the diagonal Padé approximations, e.g., about the Crank–Nicolson scheme. For "smooth" initial data u_0, however, we can recover convergence without the assumption $|r(\infty)| < 1$, hence the next result applies also to the missing case of diagonal Padé approximations.

[5] M.-N. Le Roux, "Semidiscretization in time for parabolic problems," Math. Comp. **147** (1979), 919–931.

Theorem 12.13 (Convergence theorem II.). *Suppose $A = M_m$ is as above. Let r be the stability function of an $A(\alpha)$-stable rational approximation of order p with $\alpha \in (\delta, \frac{\pi}{2}]$. Then for all $\beta \in (0, p]$ there is a constant $K > 0$ such that*

$$\|u_{h,n} - u(t)\| = \|r(hA)^n u_0 - \mathrm{e}^{tA} u_0\| \le K h^\beta \|(-A)^\beta u_0\| \qquad (t = nh)$$

holds for all $n \in \mathbb{N}$, $t \ge 0$ and $u_0 \in D\big((-A)^\beta\big)$.

Proof. We first estimate the term

$$\sup_{\substack{k \in \mathbb{N} \\ m_k \neq 0}} \big| r(hm_k)^n (-m_k)^{-\beta} - \mathrm{e}^{tm_k}(-m_k)^{-\beta} \big|.$$

As before we choose $h_0 > 0$ with

$$\sup\big\{ z \in \overline{\Sigma}_\delta : |z| \ge h_0 \big\} =: r_0 < 1.$$

If $0 < |hm_k| \le h_0$, we obtain by Proposition 12.8 that

$$\begin{aligned} \big| r(hm_k)^n (-m_k)^{-\beta} - \mathrm{e}^{tm_k}(-m_k)^{-\beta} \big| &\le C n h^\beta |hm_k|^{p+1-\beta} \mathrm{e}^{-nhc|m_k|} \\ &= \frac{C h^\beta}{n^{p-\beta}} |tm_k|^{p+1} \mathrm{e}^{-tc|m_k|} \le \frac{C' h^\beta}{n^{p-\beta}} = \frac{C' h^p}{t^{p-\beta}}. \end{aligned}$$

On the other hand, suppose $|hm_k| > h_0$. Then by the $A(\alpha)$-stability we obtain

$$\big| r(hm_k)(-m_k)^{-\beta} - \mathrm{e}^{tm_k}(-m_k)^{-\beta} \big| \le \frac{2}{|m_k|^\beta} \le \frac{2h^\beta}{h_0^\beta}.$$

Therefore, for all $k \in \mathbb{N}$ with $m_k \neq 0$ we obtain

$$\sup_{\substack{k \in \mathbb{N} \\ m_k \neq 0}} \big| r(hm_k)^n (-m_k)^{-\beta} - \mathrm{e}^{tm_k}(-m_k)^{-\beta} \big| \le C'' h^\beta.$$

We now can write

$$\begin{aligned} \|r(hA)^n u_0 - \mathrm{e}^{tA} u_0\|_2^2 &= \sum_{\substack{k \in \mathbb{N} \\ m_k \neq 0}} \big| r(hm_k)^n u_0(k) - \mathrm{e}^{tm_k} u_0(k) \big|^2 \\ &\le \sup_{\substack{k \in \mathbb{N} \\ m_k \neq 0}} \big| r(hm_k)^n - \mathrm{e}^{tm_k} u_0(k) \big|^2 \cdot \|(-A)^\beta u_0\|_2^2 \le C''^2 h^{2\beta} \cdot \|(-A)^\beta u_0\|_2^2. \end{aligned}$$

The assertion is proved. □

Remark 12.14. 1. Of course, it was only for the sake of convenience that we stated the result above on $X = \ell^2$ for suitable multiplication operators $A = M_m$. Essentially the same proofs work for the general setting: Let $(\Omega, \mathscr{A}, \mu)$ be a σ-finite measure space (Ω a nonempty set, $\mathscr{A}$ a σ-algebra, μ a measure), and consider the Banach space $X = \mathrm{L}^2(\Omega, \mathscr{A}, \mu) = \mathrm{L}^2(\Omega)$. Suppose $m : \Omega \to \mathbb{C}$ is a measurable function such that the essential range of m is contained in the sector $\overline{\Sigma}_\alpha$ for some $\alpha \in [0, \frac{\pi}{2})$. Here the **essential range** is defined by

$$\operatorname{essran}(m) := \big\{ z \in \mathbb{C} : m^{-1}(\mathrm{B}(z, \varepsilon)) \text{ has positive } \mu\text{-measure for all } \varepsilon > 0 \big\}.$$

The spectrum of M_m is precisely essran(m), hence $\Sigma_{\pi-\alpha} \subseteq \rho(A)$. The multiplication operator $A = M_m$ with maximal domain

$$D(M_m) := \big\{f \in \mathrm{L}^2(\Omega) : mf \in \mathrm{L}^2(\Omega)\big\}$$

generates an analytic semigroup T given by

$$T(t) = M_{\mathrm{e}^{tm}}.$$

For $\beta \geq 0$ the fractional power $(-A)^\beta$ of $-A = M_{-m}$ is given by

$$(-A)^\beta = M_{(-m)^\beta} \quad \text{with maximal domain} \quad D(M_{(-m)^\beta}) = \big\{f \in \mathrm{L}^2(\Omega) : (-m)^\beta f \in \mathrm{L}^2(\Omega)\big\}.$$

Now the analogues of the results from above can be stated and proved with just a bit more work.

2. By part 1. and the spectral theorem for self-adjoint operators, we see that the results in this section remain valid for non-positive self-adjoint operators A on arbitrary Hilbert spaces.

Exercises

1. Consider an s-stage Runge–Kutta method applied for the problem (12.6) and defined by formulea (12.2) and (12.3)

a) Derive the formulae (12.2), (12.3).

b) Derive the recursion (12.5).

c) Show that the stability function

$$r(z) = \big(1 + z\mathbf{b}^\top(I - z\mathbf{A})^{-1}\mathbf{1}\big)$$

from formula (12.5) is a rational function. That is, $r(z) = \frac{P(z)}{Q(z)}$ with

$$P(z) = \det(I - z\mathbf{A} + z\mathbf{1}\mathbf{b}^\top) \qquad \text{and} \qquad Q(z) = \det(I - z\mathbf{A}).$$

d) Show that if the Runge–Kutta method is of order p, its stability function has the form

$$r(z) = 1 + z + \frac{z^2}{2!} + ... + \frac{z^p}{p!} + \mathcal{O}(z^{p+1}).$$

2. Prove directly that the Crank–Nicolson approximation is A-stable, cf. Example 12.5.

3. Work out the details of Section 12.2.

4. Prove the existence of a partial fraction decomposition for a rational function. *Hint: use complex analysis.*

5. Let $r(z) = \frac{P(z)}{Q(z)}$ with $k = \deg(P)$ and $l = \deg(Q)$, where we suppose that P and Q have no common zeros. Show that if $k, l \in \mathbb{N}_0$ are fixed, the maximal order of approximation to the exponential function is $p = k + l$.

6. Convince yourself about the details of Remark 12.14.

Lecture 13

Rational Approximation and Analytic Semigroups

As we have seen in the previous lecture, rational approximations behave in a nice way for selfadjoint generators. This is due to the fact that

1. we have a well-established stability and convergence theory in the scalar case, and
2. since selfadjoint operators can be considered multiplication operators, we were able to extend the scalar estimates in a uniform way depending only on geometric conditions on the spectrum.

Notice that though we could define rational functions of operators using various formulae in Section 12.2, we could not make direct use of these formulae but needed a more refined and intimate relation between the function of an operator and the original scalar function itself. Such a relation is usually called a functional calculus.

Recall from Lecture 9 the notion of sectorial operators: Let A be a linear operator on the Banach space X, and let $\delta \in (0, \frac{\pi}{2})$. Suppose that the sector

$$\Sigma_{\frac{\pi}{2}+\delta} := \bigl\{\lambda \in \mathbb{C} \setminus \{0\} : |\arg \lambda| < \tfrac{\pi}{2} + \delta\bigr\}$$

is contained in the resolvent set $\rho(A)$, and that

$$\sup_{\lambda \in \Sigma_{\frac{\pi}{2}+\delta'}} \|\lambda R(\lambda, A)\| < \infty \quad \text{for every } \delta' \in (0, \delta).$$

Then the operator A is called **sectorial of angle** δ. For a sectorial operator A we defined

$$T(z) = \mathrm{e}^{zA} := \frac{1}{2\pi \mathrm{i}} \int\limits_{\gamma} \mathrm{e}^{\lambda z} R(\lambda, A) \mathrm{d}\lambda, \quad (z \in \Sigma_\delta) \tag{13.1}$$

with a suitable curve γ. This definition yields a strongly continuous, analytic semigroup in case A is densely defined. We also saw that densely defined sectorial operators are precisely the generators of analytic semigroups.

This lecture is devoted to the study of rational approximation schemes for this class of semigroups. To prove convergence of such schemes (in the spirit of Lecture 12, Section 12.3) we first need to develop a functional calculus, which is a bit more general than the one above for the exponential function.

13.1 The basic functional calculus

Let A be a sectorial operator of angle $\delta \in (0, \frac{\pi}{2})$ and let $\theta \in (\frac{\pi}{2} - \delta, \frac{\pi}{2})$. We consider the sector

$$\reflectbox{Σ}_\theta := \bigl\{z \in \mathbb{C} \setminus \{0\} : |\arg(-z)| < \theta\bigr\} = -\Sigma_\theta = \mathbb{C} \setminus \overline{\Sigma}_{\pi-\theta},$$

and define

$$\mathcal{H}_0^\infty(\reflectbox{Σ}_\theta) := \Big\{F : \reflectbox{Σ}_\theta \to \mathbb{C} : F \text{ is holomorphic} \\ \text{and there are } \varepsilon > 0 \text{ and } C \geq 0 \text{ with } |F(z)| \leq \tfrac{C|z|^\varepsilon}{(1+|z|)^{2\varepsilon}} \text{ for all } z \in \reflectbox{Σ}_\theta\Big\}.$$

We would like to plug the operator A into functions $F \in \mathcal{H}_0^\infty(\reflectbox{Σ}_\theta)$, and as we saw a couple of times before, the operator $F(A)$ will be defined by means of line integrals. First we specify the integration paths. For $\delta' \in (\frac{\pi}{2} - \theta, \delta)$ consider the curves given by the following parametrisations

$$\gamma_{\delta'}^1(s) := s\mathrm{e}^{\mathrm{i}(\frac{\pi}{2}+\delta')} \quad \text{and} \quad \gamma_{\delta'}^2(s) := s\mathrm{e}^{-\mathrm{i}(\frac{\pi}{2}+\delta')} \quad \text{for } s \in [0, \infty).$$

Then we consider the curve $\gamma_{\delta'} := -\gamma_{\delta'}^2 + \gamma_{\delta'}^1$. By an **admissible curve** we shall mean a curve of this type, see Figure 13.1. These ingredients are fixed for remaining of this lecture.

Definition 13.1. Let A be a sectorial operator of angle $\delta > 0$, and let $\theta \in (\frac{\pi}{2} - delta, \frac{\pi}{2}]$. For $F \in \mathcal{H}_0^\infty(\reflectbox{Σ}_\theta)$ we set

$$F(A) := \Phi_A(F) := \frac{1}{2\pi\mathrm{i}} \int_\gamma F(\lambda)R(\lambda, A)\mathrm{d}\lambda$$

where $\gamma = \gamma_{\delta'}$ with $\delta' \in (\frac{\pi}{2} - \theta, \delta)$ is an admissible curve.

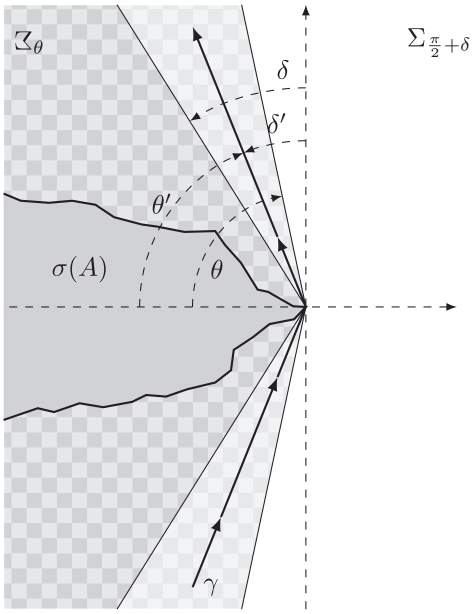
$\reflectbox{$\Sigma$}_\theta$
$\Sigma_{\frac{\pi}{2}+\delta}$
δ
δ'
θ'
$\sigma(A)$
θ
γ

Figure 13.1: An admissible curve $\gamma_{\delta'}$.

Some remarks are in order.

Remarks 13.2. 1. The integral is absolutely convergent because of the assumed decay of $F \in \mathcal{H}_0^\infty(\reflectbox{Σ}_\theta)$ near 0 and ∞ and because of the sectoriality of A. Hence $F(A) \in \mathscr{L}(X)$.

2. It is easy to see that the value of the integral defining $F(A)$ is independent of the particular choice of δ' (use Cauchy's theorem, cf. e.g. Lemma 9.12).

3. The set $\mathcal{H}_0^\infty(\reflectbox{Σ}_\theta)$ is an algebra with the pointwise operations.

Proposition 13.3. *The following assertions are true:*

a) The mapping

$$\Phi_A : \mathcal{H}_0^\infty(\Sigma_\theta) \to \mathscr{L}(X)$$

is linear and multiplicative (i.e., an algebra homomorphism).

b) For $F \in \mathcal{H}_0^\infty(\Sigma_\theta)$ if a closed operator B commutes with the resolvent of A, then it commutes with $F(A)$.

c) For all $F \in \mathcal{H}_0^\infty(\Sigma_\theta)$, $\mu \in \mathbb{C}\backslash\overline{\Sigma}_\theta = \Sigma_{\pi-\theta}$ and for $G(z) := (\mu - z)^{-1}F(z)$ we have that $G \in \mathcal{H}_0^\infty(\Sigma_\theta)$ and

$$G(A) = R(\mu, A)F(A).$$

Proof. a) Linearity follows immediately from the definition. Multiplicativity can be proved based on the resolvent identity, and similarly as in Lecture 7 for the power law of fractional powers, or in Lecture 9 for the semigroup property.

b) The proof is left as an exercise.

c) Let $\mu \in \mathbb{C} \setminus \overline{\Sigma}_\theta$ and let γ be an admissible curve. Then

$$\begin{aligned} R(\mu, A)F(A) &= \frac{1}{2\pi \mathrm{i}} \int_\gamma F(\lambda)R(\mu, A)R(\lambda, A)\mathrm{d}\lambda = \frac{1}{2\pi \mathrm{i}} \int_\gamma F(\lambda)(\mu - \lambda)^{-1}(R(\lambda, A) - R(\mu, A))\mathrm{d}\lambda \\ &= \frac{1}{2\pi \mathrm{i}} \int_\gamma F(\lambda)(\mu - \lambda)^{-1}R(\lambda, A)\mathrm{d}\lambda - \frac{1}{2\pi \mathrm{i}} \int_\gamma F(\lambda)(\mu - \lambda)^{-1}R(\mu, A)\mathrm{d}\lambda = G(A) + 0, \end{aligned}$$

where the second term is 0 by Cauchy's theorem. □

The missing details of the proof above are left as Exercise 1.

The above functional calculus does not include the function $F(z) = \frac{1}{1-z}$ corresponding to the implicit Euler scheme or the exponential function exp. To be able to cover these functions we set

$$\mathcal{E}(\Sigma_\theta) := \mathcal{H}_0^\infty(\Sigma_\theta) + \mathrm{lin}\{\mathbf{1}\} + \mathrm{lin}\{(1 - z)^{-1}\}.$$

Lemma 13.4. *a) The sum defining the linear space $\mathcal{E}(\Sigma_\theta)$ is a direct sum.*

b) The linear space $\mathcal{E}(\Sigma_\theta)$ is an algebra.

c) If $F \in \mathcal{E}(\Sigma_\theta)$ then the function G, defined by $G(z) := F(\frac{1}{z})$, is an element of $\mathcal{E}(\Sigma_\theta)$, too.

Proof. a) Let $F \in \mathcal{E}(\Sigma_\theta)$. Then the limits

$$c := \lim_{\substack{z \to 0 \\ z \in \Sigma_\theta}} F(z) \quad \text{and} \quad d := \lim_{\substack{z \to \infty \\ z \in \Sigma_\theta}} F(z)$$

exist, and we have

$$F(z) = \Big(F(z) - d\mathbf{1} + (d - c)\frac{1}{1 - z}\Big) + d\mathbf{1} - (d - c)\frac{1}{1 - z} = G(z) + d\mathbf{1} - (d - c)\frac{1}{1 - z},$$

where $G \in \mathcal{H}_0^\infty(\Sigma_\theta)$. This yields the assertion.

b) We only have to prove that for $F \in \mathcal{H}_0^\infty(\Sigma_\theta)$ and $G(z) = \frac{1}{1-z}$ one has FG, $G^2 \in \mathcal{E}(\Sigma_\theta)$. This statement about FG is trivial, since even $FG \in \mathcal{H}_0^\infty(\Sigma_\theta)$ is true by definition. As for G^2 we have

$$G^2(z) = \frac{1}{(1-z)^2} = \frac{1}{1-z} + \frac{z}{(1-z)^2},$$

where the second function belongs to $\mathcal{H}_0^\infty(\Sigma_\theta)$. So $G^2 \in \mathcal{E}(\Sigma_\theta)$.

c) We leave the proof as exercise. □

Part a) of the lemma above implies that we can extend the functional calculus to $\mathcal{E}(\Sigma_\theta)$ as follows: For $F \in \mathcal{E}(\Sigma_\theta)$ we introduce the abbreviations

$$F(0) := \lim_{\substack{z \to 0 \\ z \in \Sigma_\theta}} F(z) \quad \text{and} \quad F(\infty) := \lim_{\substack{z \to \infty \\ z \in \Sigma_\theta}} F(z).$$

Then

$$F(z) = G(z) + \frac{F(0) - F(\infty)}{1-z} + F(\infty)\mathbf{1}$$

with $G \in \mathcal{H}_0^\infty(\Sigma_\theta)$. Then we set

$$F(A) := \Phi_A(F) := \Phi_A(G) + (F(0) - F(\infty))R(1, A) + F(\infty)I.$$

Before proving algebraic properties of this extended mapping, i.e., that it is a functional calculus, we show that this new definition is at least consistent with the one developed in Lecture 9 for the exponential function.

Proposition 13.5. *a) Let $F : \Sigma_\theta \to \mathbb{C}$ be a holomorphic function that extends holomorphically to 0 and that, for some $C \geq 0$ and $\varepsilon > 0$, satisfies*

$$|F(z)| \leq \frac{C}{1 + |z|^\varepsilon} \quad \text{for all } z \in \Sigma_\theta.$$

Then $F \in \mathcal{E}(\Sigma_\theta)$ and we have

$$\Phi_A(F) = \frac{1}{2\pi \mathrm{i}} \int_\gamma F(\lambda) R(\lambda, A) \mathrm{d}\lambda,$$

where $\gamma = \gamma_{\delta', a}$ is a suitable curve as depicted in Figure 13.2 (see Eq. 9.1) lying in the domain where F is holomorphic.

b) For $\mu \in \mathbb{C} \setminus \overline{\Sigma}_\theta$ and $F(z) = (\mu - z)^k$, $k \in \mathbb{N}$ we have

$$\Phi_A(F) = R(\mu, A)^k.$$

Proof. a) We can write

$$F(z) = F(z) - \frac{F(0)}{1-z} + \frac{F(0)}{1-z} = G(z) + \frac{F(0)}{1-z}$$

with $G \in \mathcal{H}_0^\infty(\Sigma_\theta)$. Indeed, since $G(0) = 0$ and G is holomorphic at 0 we have $|G(z)| \leq C|z|$ near 0. Besides that the estimate near ∞ remains valid, so we see $F \in \mathcal{E}(\Sigma_\theta)$.

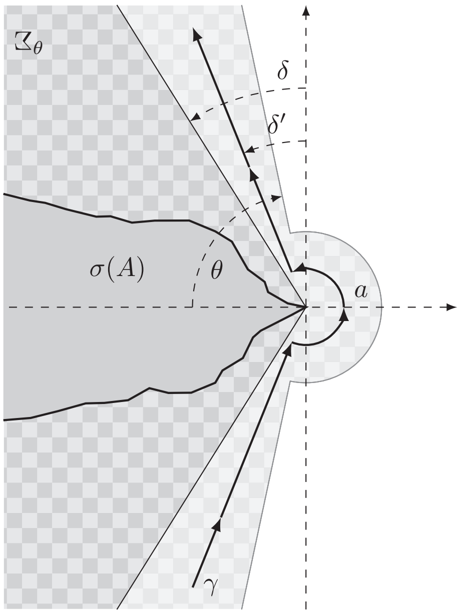

Figure 13.2: The curve $\gamma_{\delta',a}$.

Next notice that the convergence of the integral above is only an issue at ∞ and is assured by the decay of F. First consider the term

$$\frac{1}{2\pi \mathrm{i}} \int\limits_{\delta',a} G(\lambda)R(\lambda,A)\mathrm{d}\lambda.$$

Since $G \in \mathcal{H}_0^\infty(\Sigma_\theta)$ we can let $a \to 0$ and the value of the integral remains unchanged. So we can conclude

$$\frac{1}{2\pi \mathrm{i}} \int\limits_{\gamma} G(\lambda)R(\lambda,A)\mathrm{d}\lambda = \lim_{b\to 0} \frac{1}{2\pi \mathrm{i}} \int\limits_{\gamma_{\delta',b}} G(\lambda)R(\lambda,A)\mathrm{d}\lambda = \Phi_A(G).$$

We now prove

$$R(\mu,A)^k = \frac{1}{2\pi \mathrm{i}} \int\limits_{\gamma} \frac{1}{(\mu-\lambda)^k} R(\lambda,A)\mathrm{d}\lambda.$$

This identity will finish the proofs of both part a) and part b). Consider the curve $\tilde{\gamma} = -\gamma_{\eta,a+|\mu|}$. By Cauchy's theorem

$$\int\limits_{\tilde{\gamma}} \frac{R(\lambda,A)}{(\mu-\lambda)^k}\mathrm{d}\lambda = 0,$$

as can be seen by the usual trick of closing the curve $\tilde{\gamma}$ by increasing circle arcs on the right. Therefore we obtain

$$\begin{aligned}\frac{1}{2\pi \mathrm{i}} \int\limits_{\gamma} \frac{R(\lambda,A)}{(\mu-\lambda)^k}\mathrm{d}\lambda &= \frac{1}{2\pi \mathrm{i}} \int\limits_{\gamma} \frac{R(\lambda,A)}{(\mu-\lambda)^k}\mathrm{d}\lambda + \frac{1}{2\pi \mathrm{i}} \int\limits_{\tilde{\gamma}} \frac{R(\lambda,A)}{(1-\lambda)^k}\mathrm{d}\lambda \\ &= \frac{1}{2\pi \mathrm{i}} \int\limits_{\gamma+\tilde{\gamma}} \frac{R(\lambda,A)}{(\mu-\lambda)^k}\mathrm{d}\lambda = -(-1)^k \frac{\mathrm{d}^{k-1}}{\mathrm{d}z^{k-1}} R(\mu,A) = -(-1)^k(-1)^{k-1}R(\mu,A)^k \\ &= R(\mu,A)^k\end{aligned}$$

by Cauchy's formula for the derivative. □

Proposition 13.6. *The following assertions are true:*

a) The mapping

$$\Phi_A : \mathcal{E}(\Sigma_\theta) \to \mathscr{L}(X)$$

is a unital algebra homomorphism.

b) If a closed operator B commutes with the resolvent of A, then it also commutes with $F(A)$.

c) For $F(z) = z(1-z)^{-1}$ we have

$$F(A) = AR(1,A) = R(1,A) - I.$$

Proof. a) Linearity is immediate just as well the fact that Φ_A preserves the unit. We only have to prove multiplicativity on products $F \cdot G$, $F(1-z)^{-1}$, $(1-z)^{-1}(1-z)^{-1}$. The first case is contained in Proposition 13.3.a), while the second one is in Proposition 13.3.c). It remains to show that for $G(z) = (1-z)^{-2}$ one has

$$G(A) = R(1,A)^2.$$

This is proved in Proposition 13.5.

b) The statement follows directly from the definition and from Proposition 13.3.b).

c) The proof is left as exercise. □

We close this section by the following useful formula, whose proof we nevertheless leave as Exercise 2.

Proposition 13.7. *Let $F : \Sigma_\theta \to \mathbb{C}$ be a holomorphic function that is holomorphic at 0 and at ∞. Then $F \in \mathcal{E}(\Sigma_\theta)$ and we have*

$$\Phi_A(F) = F(\infty) + \frac{1}{2\pi\mathrm{i}} \int\limits_\gamma F(\lambda)R(\lambda,A)\mathrm{d}\lambda,$$

where γ the positively oriented boundary of $\mathrm{B}(0,b) \setminus (\Sigma_{\frac{\pi}{2}-\delta'} \cup \mathrm{B}(0,a))$ for $b > 0$ sufficiently large and $a > 0$ sufficiently small.

13.2 Examples

Exponential function

If A is a sectorial operator of angle $\delta > 0$, then so is hA for every $h \geq 0$: Indeed, we have that

$$\|R(\lambda, hA)\| = \tfrac{1}{h}\big\|R\big(\tfrac{\lambda}{h}, A\big)\big\| \leq \frac{M}{|\lambda|}.$$

For every $\theta \in (\frac{\pi}{2} - \delta, \frac{\pi}{2})$ we have $\exp \in \mathcal{E}(\Sigma_\theta)$, hence we can evaluate

$$\Phi_{hA}(\exp) = \mathrm{e}^{hA}.$$

By Proposition 13.5 this is just the same as the exponential function of hA from Lecture 9, i.e., we have

$$\mathrm{e}^{hA} = \frac{1}{2\pi\mathrm{i}} \int\limits_\gamma \mathrm{e}^{h\lambda} R(\lambda, A)\mathrm{d}\lambda,$$

where γ is a curve as in Proposition 13.5, see Figure 13.2.

Rational functions

Suppose r is an $A(\alpha)$-stable rational function. Then r belongs to $\mathcal{E}(\Sigma_\theta)$ for every $\theta \in (0, \alpha]$. If A is sectorial operator of angle $\delta \in (\frac{\pi}{2} - \alpha, \frac{\pi}{2})$, then so is hA for $h \geq 0$, and we can evaluate $\Phi_A(r)$ by the functional calculus and ask whether this would be the same as $r(hA)$ defined in Lecture 12, Section 12.2. That these indeed coincide can be proved based on the partial fraction decomposition and Proposition 13.5. In particular we have

$$r(A) = r(\infty) + \frac{1}{2\pi \mathrm{i}} \int\limits_\gamma (r(\lambda) - r(\infty)) R(\lambda, A) \mathrm{d}\lambda, \tag{13.2}$$

where γ is a curve as in Proposition 13.5, see Figure 13.2.

Fractional powers

For $\beta > 0$ and $k \in \mathbb{N}$ with $k \geq \beta$ consider the function

$$F_{\beta,k}(z) := \frac{(-z)^\beta}{(1-z)^k}.$$

Then $F_{\beta,k} \in \mathcal{H}_0^\infty(\Sigma_\theta)$, so we can define

$$(-A)^\beta := (I - A)^k F_{\beta,k}(A)$$

with the natural domain

$$D((-A)^\beta) = \big\{ f \in X : F_{\beta,k}(A) f \in D(A^k) \big\}.$$

The next result shows, among others, that the preceding definition is meaningful.

Proposition 13.8. *a) The definition of $(-A)^\beta$ does not depend on the choice of $k \in \mathbb{N}$.*

b) For $h \geq 0$ we have $(-hA)^\beta = h^\beta (-A)^\beta$.

c) For $\eta \in \mathbb{N}$ we have that $(-A)^\beta$ is the usual β^{th} power of $-A$.

d) If $0 \in \rho(A)$, then this new definition coincides with the one in Lecture 7, i.e., for $\beta > 0$ we have

$$(I - A)^k F_{\beta,k}(A) = \Big(\frac{1}{2\pi \mathrm{i}} \int\limits_\gamma (-\lambda)^{-\beta} R(\lambda, A) \mathrm{d}\lambda \Big)^{-1},$$

where γ is admissible curve as in Lecture 7.

The proof of this proposition is left as Exercise 5.

13.3 Convergence of rational approximation schemes

Based on the functional calculus Φ_A developed we study rational approximation schemes for analytic semigroups. We first investigate convergence results, analogous to the ones in Lecture 12.

Theorem 13.9 (**Convergence theorem I.**). *Let A be a sectorial operator of angle $\delta > 0$ and let r be an $A(\alpha)$-stable rational approximation of the exponential function of order p with $|r(\infty)| < 1$ and $\alpha \in (\frac{\pi}{2} - \delta, \frac{\pi}{2}]$. Then there is a constant $K > 0$ such that*

$$\big\|r(hA)^n - \mathrm{e}^{tA}\big\| \leq K\frac{h^p}{t^p} = \frac{K}{n^p} \qquad (t = nh)$$

holds for all $n \in \mathbb{N}$, $t \geq 0$, i.e., one has the convergence of the rational approximation scheme in the operator norm.

Proof. Fix $\delta' \in (\frac{\pi}{2} - \alpha, \delta)$, the admissible curve $\gamma = \gamma_{\delta'}$, and let $\theta' = \frac{\pi}{2} - \delta'$. We set

$$F_n(z) := r(z)^n - \mathrm{e}^{nz},$$

which is of course a function belonging to $\mathcal{E}(\Sigma_\alpha)$. We have to estimate $\|F_n(hA)\|$. Since $F_n(0) = 0$, we have

$$F_n(z) = F_n(z) + F_n(\infty)\frac{z}{1-z} - F_n(\infty)\frac{z}{1-z} = G_n(z) - F_n(\infty)\frac{z}{1-z},$$

with $G_n(z) \in \mathcal{H}_0^\infty(\Sigma_\alpha)$. Thus

$$F_n(hA) = G_n(hA) - F_n(\infty)AR(1, A).$$

Since $F_n(\infty) = r(\infty)^n$ and since $|r(\infty)| < 1$ we obtain

$$\|F_n(\infty)AR(1, A)\| \leq \frac{K'}{n^p} \quad \text{for all } n \in \mathbb{N}. \tag{13.3}$$

We turn to estimating $G_n(hA)$. Since $G_n \in \mathcal{H}_0^\infty(\Sigma_\alpha)$, we have

$$G_n(hA) = \frac{1}{2\pi \mathrm{i}}\int\limits_\gamma G_n(\lambda)R(\lambda, hA)\mathrm{d}\lambda.$$

We shall split the path of integration into two parts: γ_1 is the part of γ that lies outside of $\mathrm{B}(0, h_0)$, while γ_1 is the part inside of this ball. Since $|r(\infty)| < 1$ we can choose $h_0 > 1$ so large that

$$\sup\big\{|r(z)| : z \in \overline{\Sigma}_{\theta'} \text{ and } |z| \geq h_0\big\} =: r_0 < 1.$$

For some constant $C > 0$ we have

$$|r(z) - r(\infty)| \leq \frac{C}{|z|} \quad \text{for all } z \in \overline{\Sigma}_{\theta'} \text{ with } |z| \geq h_0.$$

This and the telescopic formula yield the estimate

$$|r(z)^n - r(\infty)^n| \leq |r(z) - r(\infty)|\sum_{j=0}^{n-1}|r(z)|^{n-1-j}|r(\infty)|^j \leq \frac{Cnr_0^{n-1}}{|z|},$$

from which we obtain

$$|F_n(z) - F_n(\infty)| = |\mathrm{e}^{nz}| + |r(z)^n - r(\infty)^n| \leq \mathrm{e}^{-n\cos(\alpha)|z|} + \frac{Cnr_0^{n-1}}{|z|}.$$

On the other hand we can write

$$\begin{aligned}\Big\|\frac{1}{2\pi\mathrm{i}}\int\limits_{\gamma_2} G_n(\lambda)R(\lambda,A)\mathrm{d}\lambda\Big\| &= \Big\|\frac{1}{2\pi\mathrm{i}}\int\limits_{\gamma_2}\big(F_n(\lambda)-F_n(\infty)\big)R(\lambda,A)+\frac{F_n(\infty)}{1-\lambda}R(\lambda,A)\mathrm{d}\lambda\Big\|\\ &\leq \frac{M}{\pi}\int\limits_{h_0}^{\infty}\Big(\mathrm{e}^{-n\cos(\alpha)s}+\frac{Cnr_0^{n-1}}{s}+\frac{|F_n(\infty)|}{s}\Big)s^{-1}\mathrm{d}s\\ &\leq \frac{M}{\pi}C'\frac{1}{n^p}+\frac{M}{\pi}C'nr_0^{n-1}+\frac{M}{\pi}C'r_0^n\leq\frac{K''}{n^p}\quad\text{for all } n\in\mathbb{N}. \qquad (13.4)\end{aligned}$$

We next estimate the integral on γ_1. Recall from Proposition 12.8 that there are constants $C, c > 0$ so that

$$|r(z)^n-\mathrm{e}^{nz}|\leq Cn|z|^{p+1}\mathrm{e}^{-nc|z|}$$

holds for all $z\in\overline{\Sigma}_{\theta'}$ with $|z|\leq h_0$, and for all $n\in\mathbb{N}$. Whence we conclude

$$|G_n(z)|\leq Cn|z|^{p+1}\mathrm{e}^{-nc|z|}+C'|z|\cdot|F_n(\infty)|.$$

This in turn yields

$$\begin{aligned}\Big\|\frac{1}{2\pi\mathrm{i}}\int\limits_{\gamma_1} G_n(\lambda)R(\lambda,A)\mathrm{d}\lambda\Big\| &\leq 2CMn\int\limits_0^{h_0}s^p\mathrm{e}^{-nsc}\mathrm{d}s+2C'h_0|F_n(\infty)|\\ &\leq\frac{2CMn}{c^{p+1}n^{p+1}}\int\limits_0^{\infty}t^p\mathrm{e}^{-t}\mathrm{d}s+2C'h_0|F_n(\infty)|\\ &=\frac{p!2CM}{c^{p+1}n^p}+2C'h_0|r(\infty)|^n\leq\frac{K'''}{n^p}\quad\text{for all } n\in\mathbb{N}. \qquad (13.5)\end{aligned}$$

By putting everything, i.e., the estimates in (13.3), (13.4) and (13.5), together we conclude the proof. □

An analogue of the next result we already saw in Lecture 12: Convergence for smooth initial data.

Theorem 13.10 (Convergence theorem II.). *Let A be a sectorial operator of angle $\delta > 0$ and let r be an $A(\alpha)$-stable rational approximation of the exponential function of order p with $|r(\infty)| < 1$ and $\alpha\in(\frac{\pi}{2}-\delta,\frac{\pi}{2}]$. For all $\beta\in(0,p]$ there is a constant $K\geq 0$ such that*

$$\|r(hA)^nf-\mathrm{e}^{nhA}f\|\leq Kh^\beta\|(-A)^\beta f\|$$

holds for all $f\in D((-A)^\beta)$, $h>0$ and $n\in\mathbb{N}$.

Proof. We set

$$F_n(z):=(-z)^{-\beta}\big(r(z)^n-\mathrm{e}^{nz}\big).$$

Since $F_n\in\mathcal{H}_0^\infty(\Sigma_\alpha)$, we have

$$F_n(hA)=\frac{1}{2\pi\mathrm{i}}\int\limits_{\gamma}F_n(\lambda)R(\lambda,hA)\mathrm{d}\lambda$$

for some admissible curve γ. Recall from Proposition 12.8 that we have

$$|r(z)^n - \mathrm{e}^{nz}| \leq Cn|z|^{p+1}\mathrm{e}^{-nc|z|}$$

for all $z \in \overline{\Sigma}_\alpha$ with $|z| \leq 1$, and for all $n \in \mathbb{N}$. By this and by splitting γ into two parts γ_1 and γ_2, inside and outside of the ball $\mathrm{B}(0,1)$, we can estimate $\|F_n(hA)\|$ as follows:

$$\begin{aligned}
\|F_n(hA)\| &\leq \frac{1}{2\pi}\int\limits_{\gamma} |F_n(\lambda)| \cdot \|R(\lambda, hA)\| \cdot |\mathrm{d}\lambda| \\
&= \frac{1}{2\pi}\int\limits_{\gamma_1} |F_n(\lambda)| \cdot \tfrac{1}{h}\|R(\tfrac{\lambda}{h}, A)\| \cdot |\mathrm{d}\lambda| + \frac{1}{2\pi}\int\limits_{\gamma_2} |F_n(\lambda)| \cdot \tfrac{1}{h}\|R(\tfrac{\lambda}{h}, A)\| \cdot |\mathrm{d}\lambda| \\
&\leq \frac{CM}{2\pi}\int\limits_{\gamma_1} |\lambda|^{p+1-\beta} n\mathrm{e}^{-cn|\lambda|}\frac{|\mathrm{d}\lambda|}{|\lambda|} + \frac{2M}{2\pi}\int\limits_{\gamma_2} |\lambda|^{-\beta}\frac{|\mathrm{d}\lambda|}{|\lambda|} \\
&\leq \frac{2CM}{2\pi}\int\limits_0^1 s^{p-\beta} n\mathrm{e}^{-cns}\mathrm{d}s + \frac{4M}{2\pi}\int\limits_1^\infty s^{-(\beta+1)}\mathrm{d}s \\
&\leq \frac{2CM}{2\pi}\int\limits_0^\infty s^{p-\beta} n\mathrm{e}^{-cns}\mathrm{d}s + \frac{4M}{2\pi}\int\limits_1^\infty s^{-(\beta+1)}\mathrm{d}s \\
&\leq \frac{2CM}{2\pi}\int\limits_0^\infty s^{p-\beta}\mathrm{e}^{-cs}\mathrm{d}s + \frac{4M}{2\pi}\int\limits_1^\infty s^{-(\beta+1)}\mathrm{d}s = K.
\end{aligned}$$

Let $k \in \mathbb{N}$ be fixed with $k > \beta$, and consider the function

$$G_n(z) := \big(r(z)^n - \mathrm{e}^{nz}\big)(1-z)^{-k} = F_n(z)(-z)^\beta(1-z)^{-k}.$$

Then we have

$$G_n(hA) = (r(hA)^n - \mathrm{e}^{nhA})R(1,hA)^k = F_n(hA)(-hA)^\beta R(1,hA)^k.$$

by Proposition 13.3.c). So for $f \in D((-A)^\beta)$ we can conclude

$$\|(r(hA)^n f - \mathrm{e}^{nhA})f\| \leq \|F_n(hA)(-hA)^\beta f\| \leq Kh^\beta\|(-A)^\beta f\|. \qquad \square$$

13.4 Stability of rational approximation schemes

Finally, let us investigate the stability of of rational approximations. The question is delicate, and we restrict our treatment here to a special case[1] first.

Theorem 13.11. *Suppose that A generates an analytic semigroup, i.e., it satisfies the resolvent condition*

$$\|R(\lambda, A)\| \leq \frac{M}{|\lambda|} \quad \text{over the sector } |\arg\lambda| \leq \tfrac{\pi}{2} + \delta.$$

[1] Ch. Lubich, O. Nevanlinna, "On resolvent conditions and stability estimates," BIT **31** (1991), 293-313.

Let $\pi - \delta < \alpha \leq \frac{\pi}{2}$, and let the rational approximation r be $A(\alpha)$-stable, i.e., suppose that

$$|r(z)| \leq 1 \quad \textit{holds for } |\arg z - \pi| \leq \alpha,$$

and satisfies $|r(\infty)| < 1$. Then there is $K \geq 1$ so that for $h > 0$ and $n \geq 1$ we have

$$\left\|r(hA)^n\right\| \leq K.$$

Proof. The proof is a delicate analysis of the curve integral representation of $r(hA)^n$. To do that, for $n \in \mathbb{N}$ consider the curve which is the union of $\gamma_1 : |\arg z| = \frac{\pi}{2}+\delta$, $|z| \geq 1/n$, and $\gamma_0 : |\arg z| \leq \frac{\pi}{2}+\delta$, $|z| = 1/n$ see Figure 13.3. By the identity in (13.2) we have the representation

$$r(hA)^n - r(\infty)^n = \frac{1}{2\pi \mathrm{i}} \int\limits_{\gamma} \left(r(\lambda)^n - r(\infty)^n\right) R(\lambda, hA) \mathrm{d}\lambda. \tag{13.6}$$

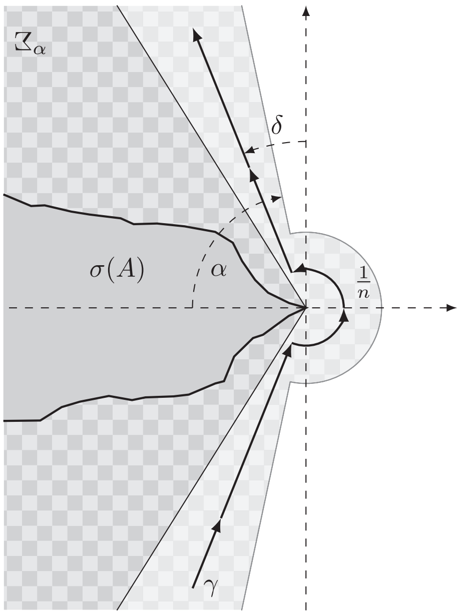

Figure 13.3: The curve γ.

To estimate the right-hand side, we split the integration path into four parts and use the following inequalities. Note first that since r is an approximation of the exponential function, we always have $\rho > 0$ such that there is $C, c > 0$ with

$$|r(z)| \leq |1 + Cz| \leq \mathrm{e}^{-c|z|}$$

for $|z| < \rho$ and $\operatorname{Re} z < 0$. From now on we suppose $n \geq \frac{1}{\rho}$. Further, by the condition $|r(\infty)| < 1$, there is $c_2 > 0$ and $0 < r_0 < 1$ such that for all $\operatorname{Re} z < -c_2$ with $|\arg z - \pi| \leq \alpha$ we have $|r(z)| \leq r_0$. Note the following:

a) If $|z| = 1/n$, then
$$|r(z)^n| \leq (1 + C|z|)^n \leq \mathrm{e}^C.$$

b) If $-\rho \sin(\delta) \leq \operatorname{Re} z \leq -\frac{1}{n} \sin(\delta)$, then
$$|r(z)| \leq \mathrm{e}^{-c|z|}.$$

c) If $-c_2 \leq \operatorname{Re} z \leq -\rho \sin(\delta)$, then

$$|r(z)| \leq 1.$$

d) If $\operatorname{Re} z < -c_2$, then

$$|r(z)| \leq r_0,$$

and $$|r(z)^n - r(\infty)^n| = |r(z) - r(\infty)| \cdot |r(z)^{n-1} + r(z)^{n-2} r(\infty) + \ldots + r(\infty)^{n-1}| \leq \frac{Cnr_0^{n-1}}{|z|}.$$

The contribution of this last term to the path integral (13.6) is

$$\int\limits_{\substack{\gamma \\ \operatorname{Re}\lambda \leq -c_2}} |r(\lambda)^n - r(\infty)^n| \cdot \|R(\lambda, A)\| \cdot |\mathrm{d}\lambda| \leq 2 \int\limits_{c_2/\sin(\delta)}^{\infty} \frac{Cnr_0^{n-1}}{s} \cdot \frac{M}{s} \mathrm{d}s = \frac{2CMnr_0^{n-1}}{c_2} \sin(\delta),$$

which is uniformly bounded in $n \in \mathbb{N}$.

The integrals over the parts defined in a) and c) are clearly bounded. We only have to check the boundedness on part b). But this follows from

$$\int\limits_{\substack{\gamma \\ -\rho\sin(\delta) \leq \operatorname{Re} z \leq -\frac{1}{n}\sin(\delta)}} M \frac{\mathrm{e}^{-Cn|z|}}{|z|} |\mathrm{d}z| \leq 2M \int\limits_{1}^{\infty} \frac{\mathrm{e}^{-Cs}}{s} \mathrm{d}s \leq K'$$

and

$$\int\limits_{\substack{\gamma \\ -\rho\sin(\delta) \leq \operatorname{Re} z \leq -\frac{1}{n}\sin(\delta)}} |r(\infty)|^n \frac{M}{|z|} |\mathrm{d}z| \leq 2M |r(\infty)|^n \int\limits_{1}^{n\rho} \frac{1}{s} \mathrm{d}s = |r(\infty)|^n 2M \log(n\rho) \leq K''.$$

This completes the proof. □

One can extend[2] the previous result and get rid of the condition $|r(\infty)| < 1$. For the sake of completeness we state the result but omit the proof.

Theorem 13.12. *Suppose that A generates an analytic semigroup, i.e., it satisfies the resolvent condition*

$$\|R(\lambda, A)\| \leq \frac{K}{|\lambda|} \quad \textit{over the sector } |\arg \lambda| \leq \tfrac{\pi}{2} + \delta.$$

Let $\pi - \delta < \alpha \leq \frac{\pi}{2}$, and let the rational approximation r be $A(\alpha)$-stable, i.e., suppose that

$$|r(z)| \leq 1 \quad \textit{holds for } |\arg z - \pi| \leq \alpha.$$

Then there is $M \geq 1$ so that for $h > 0$ and $n \geq 1$ we have

$$\big\|r(hA)^n\big\| \leq M.$$

[2] M. Crouzeix , S. Larsson , S. Piskarev , V. Thomée, “The stability of rational approximations of analytic semigroups,” BIT **33** (1993), 74–84.

Exercises

1. Work out the details of the proof of Proposition 13.3.

2. Prove Proposition 13.7.

3. Prove Proposition 13.6.b) and c).

4. Prove Lemma 13.4.c).

5. Prove Proposition 13.8.

Lecture 14

Outlook

In this last lecture we present some further important topics which could have been part of our lectures. Unfortunately, we did not have the time to include them but it will be possible to study these topics among other important subjects in the project phase of the seminar.

14.1 Rational approximations revisited

We saw in Lectures 12 and 13 how functional calculi can help investigate stability and convergence of rational approximation schemes. The Dunford–Riesz calculus from Lecture 13 works for sectorial operators, i.e., for analytic semigroups, and suitable holomorphic functions. We briefly indicate now how to obtain convergence and stability results for general strongly continuous semigroups by means of a functional calculus.

Let A generate a semigroup on the Banach space X. Of course, by rescaling we may suppose that A generates a bounded semigroup T. In this case all operators hA for $h \geq 0$ do so. We would like to define $F(hA)$ for a suitably large class of holomorphic functions that contain at least A-stable rational approximations.

The basic idea is to represent a holomorphic function F as the Laplace transform of a bounded Borel measure μ on $[0, \infty)$, i.e.,

$$F(z) = \int_0^\infty \mathrm{e}^{zs} \mathrm{d}\mu(s) \qquad (\operatorname{Re} z \leq 0).$$

Then we can define

$$F(A) = \int_0^\infty T(s) \mathrm{d}\mu(s),$$

where the integral has to be understood pointwise and in the Bochner sense. Let us consider two simple examples. First, let $\mu = \delta_t$ the point-mass at some $t \geq 0$. Then, of course, we have for the Laplace transform that $F(z) = \mathrm{e}^{tz}$, and hence $F(A) = T(t)$. Second, let μ be the measure which is absolutely continuous with respect to the Lebesgue measure on $[0, \infty)$ with Radon–Nikodym derivative $s \mapsto \mathrm{e}^{-\lambda s}$ ($\lambda \in \mathbb{C}$ with $\operatorname{Re} \lambda > 0$). Then we have

$$\int_0^\infty \mathrm{e}^{zs} \mathrm{d}\mu(s) = \int_0^\infty \mathrm{e}^{zs} \mathrm{e}^{-\lambda s} \mathrm{d}s = (\lambda - z)^{-1}.$$

For $F(A)$ we obtain

$$F(A) = \int_0^\infty T(s) \mathrm{d}\mu(s) = \int_0^\infty T(s) \mathrm{e}^{-\lambda s} \mathrm{d}s = R(\lambda, A) = (\lambda - A)^{-1},$$

by the familiar formula from Proposition 2.26. These examples at least indicate that we may be on the right track. One can prove that in general $F(A)$ is a bounded linear operator and the mapping $F \mapsto F(A) \in \mathscr{L}(X)$ is an algebra homomorphism (the Laplace transforms of bounded Borel measures form an algebra with pointwise operations). This functional calculus is called the **Hille–Phillips calculus**.

The idea to use the Hille–Phillips calculus for the study of rational approximations is due to Hersh and Kato.[1] They also formulated a conjecture that was subsequently(!) answered in the affirmative by Brenner and Thomée:[2]

Theorem 14.1 (**Brenner–Thomée, Stability theorem**). *Let r be an A-stable rational approximation of the exponential function (at least of order 1). Then there exist constants $K \geq 0$, $\omega' \geq 0$ such that for every strongly continuous semigroup T of type (M, ω) with $\omega \geq 0$ and with generator A one has*

$$\|r(hA)^n\| \leq KM\sqrt{n}\mathrm{e}^{\omega\omega' t} \qquad (t = nh).$$

With some extra technicalities one may further refine this result. We remark, however, that in its generality the stability result of Brenner and Thomée, i.e., the $\mathcal{O}(n^{1/2})$ term, is sharp. As for convergence the next one is a basic result; see also Corollary 4.15.

Theorem 14.2 (**Brenner–Thomée, Convergence theorem**). *Let r be an A-stable rational approximation of the exponential function of order p. Then there exist constants $K \geq 0$, $\omega' \geq 0$ such that for every strongly continuous semigroup T of type (M, ω) with $\omega \geq 0$ and with generator A one has for each $f \in D(A^{p+1})$ that*

$$\|r(hA)^n f - \mathrm{e}^{tA} f\| \leq KMth^p \mathrm{e}^{\omega\omega' t}\|A^{p+1} f\| \qquad (t = nh).$$

The proof of these beautiful results, as have been said, rely on the Hille–Phillips calculus or on some variants of it, and on delicate estimates from harmonic analysis. Expect more in the project phase of this seminar!

14.2 Non-autonomous problems

In many applications it is quite natural to consider differential equations with time dependent coefficients, i.e., a non-autonomous evolution equation of the form

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} u(t) = A(t)u(t), & t \geq s \in \mathbb{R} \\ u(s) = f \in X, \end{cases} \tag{14.1}$$

where X is a Banach space and $A(t)$ is a family of (usually unbounded) linear operators on X. As in the autonomous case, the operator family solving a non-autonomous Cauchy problem enjoys certain algebraic properties.

A family $U = (U(t, s))_{t \geq s}$ of linear, bounded operators on a Banach space X is called an (exponentially bounded) **evolution family** if

(i) $U(t, r)U(r, s) = U(t, s), \quad U(t, t) = I$ hold for all $s \leq r \leq t \in \mathbb{R}$,

[1] R. Hersh, T. Kato, "High-accuracy stable difference schemes for well-posed initial-value problems," SIAM Journal on Numerical Analysis **16** (1979), no. 4, 670–682.

[2] P. Brenner, V. Thomée, "On rational approximation of semigroups," SIAM Journal on Numerical Analysis **16** (1979), no. 4, 683–694.

(ii) the mapping $(t,s) \mapsto U(t,s)$ is strongly continuous,

(iii) $\|U(t,s)\| \leq M\mathrm{e}^{\omega(t-s)}$ for some $M \geq 1, \omega \in \mathbb{R}$ and all $s \leq t \in \mathbb{R}$.

In general, and in contrast to the behaviour of semigroups, the algebraic properties of an evolution family do not imply any differentiability on a dense subspace. So we have to add some differentiability assumptions in order to solve a non-autonomous Cauchy problem by an evolution family.

Definition 14.3. An evolution family $U = (U(t,s))_{t \geq s}$ is called the **evolution family solving** (14.1) if for every $s \in \mathbb{R}$ the regularity subspace

$$Y_s := \big\{ g \in X : [s,\infty) \ni t \mapsto U(t,s)g \ \text{ solves (14.1)} \big\}$$

is dense in X.

The well-posedness of (14.1) can be characterised by the existence of a solving evolution family.[3] Hence, application of a suitable numerical method means the approximation of the evolution family U.

To motivate the following, let us consider the scalar case $X = \mathbb{R}$. Then assuming that the function $t \mapsto A(t) \in \mathbb{R}$ is smooth enough, the evolution family can be written explicitly as

$$U(t,s) = \mathrm{e}^{\int_s^t A(r) \mathrm{d}r}. \tag{14.2}$$

It is well-known that this formula holds in general only if the operators $A(t)$ commute, hence, we cannot use it in this form. We may make, however, the following heuristic arguments. Suppose that $A(r)$ generates a semigroup for all $r \geq s$ and that the function $r \mapsto A(r)$ is smooth in some sense. Then choosing a small stepsize $h > 0$, the function $[s, s+h] \ni r \mapsto A(r)$ may be assumed to be constant, for example $A(r) \approx A(s)$ or $A(r) \approx A(s + \frac{h}{2})$. Hence, we arrive at the approximations

$$U(s+h,s) \approx \mathrm{e}^{hA(s)} \qquad \text{or} \qquad U(s+h,s) \approx \mathrm{e}^{hA(s+\frac{h}{2})},$$

respectively. These correspond to the left Riemann sum or the midpoint rule approximation of the integral in (14.2), leading to the simplest possible approximation schemes of a series of methods. Their basic idea is to express the solution $u(t)$ in the form

$$u(t) = \exp(\Omega(t))f,$$

where $\Omega(t)$ is an infinite sum yielded by the formal iteration[4]

$$\begin{aligned} \Omega(t) = {} & \int_0^t A(\tau)\,\mathrm{d}\tau - \frac{1}{2}\int_0^t \Big[\int_0^\tau A(\sigma)\,\mathrm{d}\sigma, A(\tau)\Big]\,\mathrm{d}\tau \\ & + \frac{1}{4}\int_0^t \Big[\int_0^\tau \Big[\int_0^\sigma A(\mu)\,\mathrm{d}\mu, A(\sigma)\Big]\,\mathrm{d}\sigma, A(\tau)\Big]\,\mathrm{d}\tau \\ & + \frac{1}{12}\int_0^t \Big[\int_0^\tau A(\sigma)\,\mathrm{d}\sigma, \Big[\int_0^\tau A(\mu)\,\mathrm{d}\mu, A(\tau)\Big]\Big]\,\mathrm{d}\tau + \cdots \end{aligned} \tag{14.3}$$

[3] See the survey by R. Schnaubelt, “Semigroups for nonautonomous Cauchy problems”, in K. -J. Engel, R. Nagel, *One-Parameter Semigroups for Linear Evolution Equations*, Springer-Verlag, 2000.

[4] W. Magnus, “On the exponential solution of differential equations for a linear operator”, Comm. Pure and Appl. Math. **7** (1954), 639-673.

(with $s = 0$), where $[U, V] = UV - VU$ denotes the commutator of the operators U and V. Methods based on this formal expression are called **Magnus methods** and can be described in the following way: We cut off the infinite series somewhere and approximate the remaining integrals by suitable quadrature rules. Although this method is extensively studied in the finite dimensional case, in Banach spaces only a few papers can be found treating special cases.[5]

We give an idea on a possible approach to prove these formulae. It is possible to transform a non-autonomous Cauchy problem to an autonomous one in a bigger, more complicated space (i.e., introducing an additional differential equation on the time evolution) in the following way. To every evolution family we can associate semigroups on X-valued function spaces. These semigroups, which determine the behaviour of the evolution family completely, are called **evolution semigroups**. Consider the Banach space

$$\mathrm{BUC}(\mathbb{R}, X) = \big\{F : \mathbb{R} \to X : F \text{ is bounded and uniformly continuous}\big\},$$

normed by

$$\|F\| := \sup_{t\in\mathbb{R}} \|F(t)\|, \quad F \in \mathrm{BUC}(\mathbb{R}, X),$$

or any closed subspace of it that is invariant under the right translation semigroup $\mathscr{R}$ defined by

$$(\mathscr{R}(t)F)(s) := F(s - t) \quad \text{for } F \in \mathrm{BUC}(\mathbb{R}, X) \text{ and } s \in \mathbb{R},\ t \geq 0.$$

In the following $\mathscr{X}$ will denote such a closed subspace. We shall typically take $\mathscr{X} = \mathrm{C}_0(\mathbb{R}, X)$, the space of continuous functions vanishing at infinity.

It is easy to check that the following definition yields a strongly continuous semigroup.

Definition 14.4. For an evolution family $U = (U(t, s))_{t\geq s}$ we define the corresponding evolution semigroup $\mathscr{T}$ on the space $\mathscr{X}$ by

$$(\mathscr{T}(t)F)(s) := U(s, s - t)F(s - t)$$

for $F \in \mathscr{X}$, $s \in \mathbb{R}$ and $t \geq 0$. We denote its infinitesimal generator by $\mathscr{G}$.

With the above notation, the evolution semigroup operators can be written as

$$\mathscr{T}(t)F = U(\cdot, \cdot - t)\mathscr{R}(t)F.$$

We can recover the evolution family from the evolution semigroup by choosing a function $F \in \mathscr{X}$ with $F(s) = f$. Then we obtain

$$U(t, s)x = (\mathscr{R}(s - t)\mathscr{T}(t - s)F)(s) \tag{14.4}$$

for every $s \in \mathbb{R}$ and $t \geq s$.

The generator of the right translation semigroup is essentially the differentiation $-\frac{\mathrm{d}}{\mathrm{d}s}$ with domain

$$D(-\tfrac{\mathrm{d}}{\mathrm{d}s}) := \mathscr{X}_1 := \big\{F \in \mathrm{C}^1(\mathbb{R}, X) : F, F' \in \mathscr{X}\big\}.$$

For a family $A(t)$ of unbounded operators on X we consider the corresponding multiplication operator $A(\cdot)$ on the space $\mathscr{X}$ with domain

$$D(A(\cdot)) := \big\{F \in \mathscr{X} : F(s) \in D(A(s))\ \forall\, s \in \mathbb{R},\ \text{and } [s \mapsto A(s)F(s)] \in \mathscr{X}\big\},$$

[5] M. Hochbruck, Ch. Lubich, “On Magnus integrators for time-dependent Schrödinger equations”, SIAM Journal on Numerical Analysis **41** (2003), 945–963.

and defined by

$$(A(\cdot)F)(s) := A(s)F(s) \text{ for all } s \in \mathbb{R}.$$

The main power of evolution semigroups is that one can transform existing semigroup theoretic results to the non-autonomous case and prove the convergence of various approximation schemes. As an illustration, we present a simple convergence result.[6] You will see more on this in the project phase.

Theorem 14.5. *Suppose that problem* (14.1) *is well-posed,* $D(A(t)) := D \subset X$ *and that* $A(t)$ *generates a strongly continuous semigroup for each* $t \in \mathbb{R}$. *If the function* $t \mapsto A(t)f$ *is uniformly continuous for all* $F \in D$, *then*

$$U(t,s)f = \lim_{n\to\infty} \prod_{p=0}^{n-1} \mathrm{e}^{\frac{t-s}{n}A(s+\frac{p(t-s)}{n})}f.$$

Here and later on, for bounded linear operators $L_k \in \mathscr{L}(X)$,

$$\prod_{k=0}^{n-1} L_k := L_{n-1}L_{n-2}\cdots L_0$$

denotes the “time-ordered product”.

Proof. We only give the idea here. In the space $\mathrm{BUC}(\mathbb{R}, X)$ consider the function

$$\mathscr{F}(h) = \mathscr{R}(h)\mathrm{e}^{hA(\cdot)}.$$

Then, for suitable $F \in \mathrm{BUC}(\mathbb{R}, X)$, we get

$$\lim_{h\searrow 0} \frac{\mathscr{F}(h)F - F}{h} = \lim_{h\searrow 0}\left(\mathscr{R}(h)\frac{\mathrm{e}^{hA(\cdot)}F - F}{h} + \frac{\mathscr{R}(h)F - F}{h}\right) = A(\cdot)F - F'.$$

Further, it is easy to check by induction that

$$\big(\mathscr{R}\big(\tfrac{t}{n}\big)\mathrm{e}^{\frac{t}{n}A(\cdot)}\big)^n F(\cdot) = \prod_{p=n}^{1}\big(\mathrm{e}^{\frac{t}{n}A(\cdot-\frac{pt}{n})}\big)\mathscr{R}(t)F(\cdot).$$

Hence, applying Theorem 4.6, we get that the evolution semigroup can be represented as

$$\mathscr{T}(t)F = \lim_{n\to\infty}\prod_{p=n}^{1}\big(\mathrm{e}^{\frac{t}{n}A(\cdot-\frac{pt}{n})}\big)F(\cdot - t).$$

The proof can be finished now by applying (14.4). □

[6] A. Bátkai, P. Csomós, B. Farkas, G. Nickel, “Operator splitting for nonautonomous evolution equations”, J. Funct. Anal. **260** (2011), 2163–2190.

14.3 Geometric properties of semilinear problems

The geometric theory of evolution equations is concerned with qualitative properties of particular solutions (e.g., equilibria, periodic orbits, bifurcations) and their stability properties. In this section we investigate whether numerical methods are able to capture the geometric properties of the exact flow. For simplicity, we restrict our attention to the implicit Euler method. Similar results, however, also hold for a large class of implicit Runge–Kutta methods (among others).

As an example, we consider the semilinear problem

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}u(t) = Au(t) + F(u(t)) \\ \quad u(0) = u_0 \end{cases} \tag{14.5}$$

on a Hilbert space X, together with its implicit Euler discretisation

$$u_{n+1} = u_n + hAu_{n+1} + hF(u_{n+1}) \tag{14.6}$$

for $n \in \mathbb{N}$, $nh = t$. We make the following standard assumptions.

Assumption 14.6. a) The operator A is the generator of a bounded analytic semigroup on X, i.e., a densely defined closed linear operator on X whose resolvent is bounded by

$$\|(\lambda - A)^{-1}\| \le M\,|\lambda|^{-1} \qquad \text{for } \ \lambda \in \Sigma_{\frac{\pi}{2}+\delta} \text{ with some } \delta < \pi/2.$$

In addition, we assume that A itself is invertible with $\|A^{-1}\| \le M$. For a suitable $\alpha \in [0,1)$ let $V = D((-A)^\alpha)$ be the domain of $(-A)^\alpha$ equipped with the norm

$$\|v\|_\alpha = \|(-A)^\alpha v\|.$$

b) The nonlinearity $F : V \to X$ is assumed to be locally Lipschitz bounded, i.e., for every $R > 0$, there exists $L = L(R) < \infty$ such that

$$\|F(v_2) - F(v_1)\| \le L\|v_2 - v_1\|_\alpha \quad \text{for } \|v_i\|_\alpha \le R \quad (i = 1, 2).$$

Reaction-diffusion equations and the incompressible Navier–Stokes equations can be cast in this abstract framework.

Equilibria

The simplest geometric object of (14.5) is an equilibrium point, i.e., a point u_* satisfying

$$Au_* + F(u_*) = 0.$$

It is obvious that u_* is also an equilibrium point of the numerical scheme (14.6). We first linearise F at u_* to obtain

$$F(u) = Bu + G(u)$$

with

$$B = \frac{\mathrm{d}}{\mathrm{d}u}F(u_*) \qquad \text{and} \qquad \frac{\mathrm{d}}{\mathrm{d}u}G(u_*) = 0.$$

From a perturbation argument analogous to Theorem 6.14 we know that the operator $\widetilde{A} = A + B$ generates an analytic semigroup of type $(M, -\nu)$, that is,

$$\|\mathrm{e}^{t\widetilde{A}}\| \le M\mathrm{e}^{-\nu t}. \tag{14.7}$$

We assume that the equilibrium point u_* is asymptotically stable, i.e. $\nu > 0$, and that there exist positive constants R and M such that for any solution of (14.5) with initial value $\|u(0) - u_*\| \le R$ it holds that

$$\|u(t) - u_*\| \le M\mathrm{e}^{-\nu t}, \qquad t \ge 0.$$

Our aim is to show that u_* is an asymptotically stable fixed point of the map (14.6). For this, we subtract

$$u_* = u_* + h\widetilde{A}u_* + hG(u_*)$$

from the numerical method

$$u_{n+1} = u_n + h\widetilde{A}u_{n+1} + hG(u_{n+1})$$

to get the recursion

$$u_{n+1} - u_* = (I - h\widetilde{A})^{-1}(u_n - u_*) + h(I - h\widetilde{A})^{-1}(G(u_{n+1}) - G(u_*)).$$

Solving this recursion, we get

$$u_n - u_* = (I - h\widetilde{A})^{-n}(u_0 - u_*) + h\sum_{j=1}^{n}(I - h\widetilde{A})^{-n-1+j}(G(u_j) - G(u_*)). \tag{14.8}$$

Our estimate below requires the following stability bounds for the sectorial operator A. For any $\mu < \nu$ there exists a maximal step size h_0 such that the following bounds hold for $h \in (0, h_0]$ and all $j \in \mathbb{N}$

$$\big\|(I - h\widetilde{A})^{-j}\big\| \le M\mathrm{e}^{-\mu jh}$$
$$\big\|A^{-\alpha}(I - h\widetilde{A})^{-j}\big\| \le M\frac{\mathrm{e}^{-\mu jh}}{(jh)^{\alpha}}.$$

Let $\varepsilon_n = \|u_n - u_*\|$. Taking norms in (14.8) and inserting these bounds, we obtain

$$\varepsilon_n \le M\mathrm{e}^{-\nu nh}\varepsilon_0 + h\sum_{j=1}^{n}\frac{M\mathrm{e}^{-\nu(n+1-j)h}}{((n+1-j)h)^{\alpha}}\varepsilon_j^2.$$

Solving the Gronwall type inequality, we get the following theorem.

Theorem 14.7. *Let u_* be an asymptotically stable equilibrium point of* (14.5) *with ν given by* (14.7)*, and let $\mu < \nu$. Under the above assumptions there exist positive constants h_0, R, and M such that the following holds: For $h \in (0, h_0]$ and $\|u_0 - u_*\| \le R$, the implicit Euler discretisation* (14.6) *fulfills the bound*

$$\|u_n - u_*\| \le M\mathrm{e}^{-\mu nh} \qquad \text{for all } n \in \mathbb{N}_0.$$

Periodic orbits

Our next aim is to study the approximation of an asymptotically stable periodic orbit of (14.5) by an invariant closed curve of (14.6). Here, we follow closely an article by Lubich and Ostermann who studied this question for implicit Runge–Kutta methods.[7] Let

$$S\big(t, u(\tau)\big) = u(t + \tau) \qquad \text{for all } t \ge 0$$

[7] Ch. Lubich, A. Ostermann, “Runge-Kutta time discretization of reaction-diffusion and Navier–Stokes equations: Nonsmooth-data error estimates and applications to long-time behaviour,” Applied Numerical Math. **22** (1996), 279–292.

denote the nonlinear semigroup on V of (14.5) and let

$$S_h(u_n) = u_{n+1} \qquad \text{for all } n \in \mathbb{N}$$

its numerical discretisation, defined by (14.6). We assume that problem (14.5) has an asymptotically stable periodic orbit $\Gamma = \{u_*(t) : 0 \leq t \leq t_p\}$ with period t_p, i.e.,

$$S\big(t_p, u_*(t)\big) = u_*(t) \qquad \text{for all } t \geq 0$$

and the Fréchet derivative of S along the periodic orbit has $\lambda = 1$ as a simple eigenvalue while the remaining part of the spectrum is bounded in modulus by a number less than 1. Then the following theorem holds.

Theorem 14.8. *Consider problem* (14.5) *under the above assumptions. Then, for sufficiently small time step h, the implicit Euler discretisation* (14.6) *has an invariant closed curve Γ_h in V, i.e., $S_h(\Gamma_h) = \Gamma_h$, which is uniformly asymptotically stable.*

If the periodic solution lies in $\mathrm{L}^1([0, t_p], V)$ over a period, then the Hausdorff distance with respect to the norm of V between Γ_h and Γ is bounded by

$$\mathrm{dist}_{\mathrm{H}}(\Gamma_h, \Gamma) \leq Ch. \tag{14.9}$$

For more details and for the proof we refer to the project phase of the seminar.

14.4 Exponential integrators

Exponential integrators are numerical methods applied to solve semilinear evolution equations of the form

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} u(t) = Au(t) + G(t, u(t)) \\ \quad u(0) = u_0. \end{cases} \tag{14.10}$$

Instead of discretising the differential equation directly, exponential integrators approximate the mild solution given by the variation-of-constants formula

$$u(t) = \mathrm{e}^{tA} u_0 + \int_0^t \mathrm{e}^{(t-\tau)A} G(\tau, u(\tau))\, \mathrm{d}\tau.$$

The simplest numerical method is obtained by replacing the nonlinearity G under the integral by its known value at $t = 0$. For the approximation u_1 of $u(h)$ this yields

$$u_1 = \mathrm{e}^{hA} u_0 + \int_0^h \mathrm{e}^{(h-\tau)A} G(0, u_0)\, \mathrm{d}\tau = \mathrm{e}^{hA} u_0 + h\varphi_1(hA) G(0, u_0)$$

with function φ_1 already defined in (11.4). This numerical scheme is called **exponential Euler method**. When integrating the equation from t_n to $t_{n+1} = t_n + h_n$ the method has the form

$$u_{n+1} = \mathrm{e}^{h_n A} u_n + h_n \varphi_1(h_n A) G(t_n, u_n) = u_n + h_n \varphi_1(h_n A)(Au_n + G(t_n, u_n)).$$

Here, $h_n > 0$ is the time step and u_{n+1} is the numerical approximation to the exact solution at time t_{n+1}. The second representation is preferable from a numerical point of view since its implementation requires only one evaluation of a matrix function.

In a similar way, by introducing stages and/or using more sophisticated interpolation methods, exponential Runge–Kutta and exponential multistep methods can be derived. We will come back to these numerical schemes in the project part of the seminar. Here, we will only illustrate how this approach is used for linear problems. We closely follow the presentation given in the recent survey article by Hochbruck and Ostermann.[8]

Consider the linear inhomogeneous evolution equation

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}u(t) = Au(t) + F(t) \\ u(0) = u_0 \end{cases} \tag{14.11}$$

where A generates a strongly continuous semigroup and the inhomogeneity F is sufficiently smooth. The solution of (14.11) at time

$$t_{n+1} = t_n + h_n, \qquad t_0 = 0, \qquad n \in \mathbb{N}_0$$

is given by the variation-of-constants formula

$$u(t_{n+1}) = \mathrm{e}^{h_n A} u(t_n) + \int_0^{h_n} \mathrm{e}^{(h_n - \tau)A} F(t_n + \tau)\,\mathrm{d}\tau. \tag{14.12}$$

In order to obtain a numerical scheme, we approximate the function F by an interpolation polynomial with prescribed nodes $c_1, \ldots, c_s$. The resulting integrals can be computed analytically. This yields the **exponential quadrature rule**

$$u_{n+1} = \mathrm{e}^{h_n A} u_n + h_n \sum_{i=1}^{s} b_i(h_n A) F(t_n + c_i h_n), \tag{14.13}$$

where the weights b_i are linear combinations of the entire functions φ_j, $j \in \mathbb{N}$ defined in (11.4).

Example 14.9. For $s = 1$, the interpolation polynomial is the constant polynomial $F(t_n + c_1 h_n)$ and we get the numerical scheme

$$u_{n+1} = u_n + h_n \varphi_1(h_n A)(A u_n + F(t_n + c_1 h_n)).$$

The choice $c_1 = 0$ yields the exponential Euler quadrature rule, while $c_1 = \frac{1}{2}$ corresponds to the exponential midpoint rule.

Example 14.10. For $s = 2$, the interpolation polynomial has the form

$$p(t_n + \tau) = F(t_n + c_1 h_n) + \frac{F(t_n + c_2 h_n) - F(t_n + c_1 h_n)}{(c_2 - c_1) h_n} (\tau - c_1 h_n),$$

and we obtain the weights

$$\begin{aligned} b_1(z) &= \frac{c_2}{c_2 - c_1}\varphi_1(z) - \frac{1}{c_2 - c_1}\varphi_2(z) \\ b_2(z) &= -\frac{c_1}{c_2 - c_1}\varphi_1(z) + \frac{1}{c_2 - c_1}\varphi_2(z). \end{aligned}$$

The choice $c_1 = 0$ and $c_2 = 1$ yields the exponential trapezoidal rule.

[8] M. Hochbruck, A. Ostermann, “Exponential integrators,” Acta Numerica **19** (2010), 209–286.

A more general class of quadrature methods is obtained by only requiring that the weights $b_i(h_nA)$ are uniformly bounded in $h_n \geq 0$. In order to analyse these schemes, we expand the right-hand side of equation (14.12) into a Taylor series with remainder in integral form

$$\begin{aligned} u(t_{n+1}) &= \mathrm{e}^{h_nA}u(t_n) + \int_0^{h_n} \mathrm{e}^{(h_n-\tau)A}F(t_n+\tau)\mathrm{d}\tau \\ &= \mathrm{e}^{h_nA}u(t_n) + h_n\sum_{k=1}^{p}\varphi_k(h_nA)h_n^{k-1}F^{(k-1)}(t_n) \\ &\quad + \int_0^{h_n}\mathrm{e}^{(h_n-\tau)A}\int_0^{\tau}\frac{(\tau-\xi)^{p-1}}{(p-1)!}F^{(p)}(t_n+\xi)\mathrm{d}\xi\mathrm{d}\tau. \end{aligned}$$

This is compared to the Taylor series of the numerical solution (14.13):

$$\begin{aligned} u_{n+1} &= \mathrm{e}^{h_nA}u_n + h_n\sum_{i=1}^{s}b_i(h_nA)F(t_n+c_ih_n) \\ &= \mathrm{e}^{h_nA}u_n + h_n\sum_{i=1}^{s}b_i(h_nA)\sum_{k=0}^{p-1}\frac{h_n^kc_i^k}{k!}F^{(k)}(t_n) \\ &\quad + h_n\sum_{i=1}^{s}b_i(h_nA)\int_0^{c_ih_n}\frac{(c_ih_n-\tau)^{p-1}}{(p-1)!}F^{(p)}(t_n+\tau)\mathrm{d}\tau. \end{aligned}$$

Obviously the error $e_n = u_n - u(t_n)$ satisfies

$$e_{n+1} = \mathrm{e}^{h_nA}\,e_n - \delta_{n+1} \tag{14.14}$$

with
$$\delta_{n+1} = \sum_{j=1}^{p}h_n^j\psi_j(h_nA)f^{(j-1)}(t_n) + \delta_{n+1}^{(p)},$$

where
$$\psi_j(h_nA) = \varphi_j(h_nA) - \sum_{i=1}^{s}b_i(h_nA)\frac{c_i^{j-1}}{(j-1)!}$$

and
$$\begin{aligned} \delta_{n+1}^{(p)} &= \int_0^{h_n}\mathrm{e}^{(h_n-\tau)A}\int_0^{\tau}\frac{(\tau-\xi)^{p-1}}{(p-1)!}F^{(p)}(t_n+\xi)\mathrm{d}\xi\mathrm{d}\tau \\ &\quad - h_n\sum_{i=1}^{s}b_i(h_nA)\int_0^{c_ih_n}\frac{(c_ih_n-\tau)^{p-1}}{(p-1)!}F^{(p)}(t_n+\tau)\mathrm{d}\tau. \end{aligned}$$

We are now ready to state our convergence result.

Theorem 14.11. *Let A generate a strongly continuous semigroup and let $F^{(p)} \in \mathrm{L}^1(0,t_0)$. For the numerical solution of problem* (14.11) *consider the exponential quadrature rule* (14.13) *with uniformly bounded weights $b_i(h_nA)$ for $h_n \geq 0$. If the method satisfies the order conditions*

$$\psi_j(h_nA) = 0 \qquad \textit{for all } j = 1,\dots,p, \tag{14.15}$$

then it is convergent of order p. More precisely, the error bound

$$\|u_n - u(t_n)\| \le C \sum_{j=0}^{n-1} h_j^p \int_{t_j}^{t_{j+1}} \|F^{(p)}(\tau)\| \, \mathrm{d}\tau$$

holds uniformly on $t_n \in [0, t_0]$, with a constant C that depends on t_0 but is independent of the chosen time step sequence h_j.

Proof. Solution of the error recursion (14.14) yields the estimate

$$\|e_n\| \le \sum_{j=0}^{n-1} \|\mathrm{e}^{(t_n - t_j)A}\| \cdot \|\delta_j^{(p)}\|.$$

The desired bound follows from the stability bound and the assumption on the weights. □

The End ... of Phase 1

Appendix A

Basic Space Discretisation Methods

Partial differential equations (PDEs) appear in all fields in which mathematical models are applied to describe the time evolution of certain space-dependent quantities. Since these PDEs might be of complicated form, it is often difficult or even impossible to solve them analytically. Therefore, one has to apply numerical schemes to obtain an approximation to the exact solution. In the present Appendix we deal with *space discretisation* methods, i.e., with numerical procedures approximating the spatial differential operator appearing in the PDE.

Let us consider the following linear PDE in a general form with the differential operator L, and the unknown function $w : [0, \infty) \times \Omega \to \mathbb{R}$ for $t \in [0, \infty)$ and $x \in \Omega \subseteq \mathbb{R}^d$:

$$\begin{aligned} \partial_t^\alpha w(t,x) &= Lw(t,x), \quad t > 0, x \in \Omega \\ w(0,x) &= w_0(x), \quad x \in \Omega, \end{aligned} \tag{A.1}$$

subject to appropriate boundary conditions. Here Ω denotes an open set, and $\alpha \in \mathbb{N}$ is the order of the time derivative. We are merely interested in the cases $\alpha = 1$, when $\partial_t w$ appears on the left-hand side of equation (A.1), or $\alpha = 2$ with $\partial_{tt} w$. We note that for $\alpha = 0$, the case $f(x) = Lw(x)$ with a given function $f : \Omega \to \mathbb{R}$ is also possible.

Example A.1. a) *Heat equation*

$$\partial_t w(t,x) = \partial_{xx} w(t,x)$$

in one dimension with $\alpha = 1$, $x \in (0, \pi)$, and $Lw(t,x) := \partial_{xx} w(t,x)$. Together with the boundary condition $w(t,0) = w(t,\pi) = 0$, this problem was already investigated in Section 1.1.

b) *Transport equation*

$$\partial_t w(t,x) = \partial_x w(t,x)$$

in one dimension with $\alpha = 1$, $x \in (0,1)$, and $Lw(t,x) := \partial_x w(t,x)$. Together with the boundary condition $w(t,1) = 0$, this problem was already investigated in Section 1.2.

c) *Wave equation*

$$\partial_{tt} w(t,x) = \partial_{xx} w(t,x)$$

in one dimension with $\alpha = 2$, $x \in (0,1)$, and $Lw(t,x) := \partial_{xx} w(t,x)$ with $w(t,0) = \partial_t w(t,0) = 0$. We will come back to this example later on during the lectures.

The main idea behind the simplest discretisation type[1] of PDEs is the following. First we discretise the operator L on the right-hand side with respect to the space variable x. By this we obtain an ordinary differential equation which is then solved by using time discretisation methods.

In what follows we discuss two ways of approximating the operator L and briefly introduce the two main classes of space discretisation methods: finite differences and Galerkin methods.

[1] It is called the *method of lines.*

A.1 Finite difference methods

Here we discuss finite difference methods in the one-dimensional case for the heat and transport equation, that is, for $d = 1$ and $\Omega = (a, b)$. In order to discretise problems like (A.1) in 1D, we divide the interval (a, b) into N pieces of sub-intervals with length $\Delta x = \frac{b-a}{N}$. Then the points $x_j = a + j\Delta x$, $j = 0, ..., N$, are called **grid points**.

Now, we would like to approximate the exact solution at time level $t \geq 0$ and at the points x_j, i.e., $w(t, x_j) = w(t, a+j\Delta x)$ by the values $w_j(t)$ for $j = 0, ..., N$. To this end we use Taylor's formula with respect to the second variable:

$$\begin{aligned} w(t, x_{j+1}) &= w(t, x_j + \Delta x) \\ &= w(t, x_j) + \Delta x \cdot \partial_x w(t, x_j) + \tfrac{1}{2}(\Delta x)^2 \cdot \partial_{xx} w(t, x_j) + \cdots , \end{aligned}$$

or

$$\begin{aligned} w(t, x_{j-1}) &= w(t, x_j - \Delta x) \\ &= w(t, x_j) - \Delta x \cdot \partial_x w(t, x_j) + \tfrac{1}{2}(\Delta x)^2 \cdot \partial_{xx} w(t, x_j) + \cdots . \end{aligned}$$

The first- and second-order partial derivatives of w with respect to the space variable x, appearing in the expression of Lw, can now be written as:

$$\begin{aligned} \partial_x w(t, x_j) &= \frac{w(t, x_{j+1}) - w(t, x_j)}{\Delta x} + \mathcal{O}\big((\Delta x)^2\big), \quad \text{or} \\ \partial_x w(t, x_j) &= \frac{w(t, x_j) - w(t, x_{j-1})}{\Delta x} + \mathcal{O}\big((\Delta x)^2\big), \quad \text{and} \\ \partial_{xx} w(t, x_j) &= \frac{w(t, x_{j+1}) - 2w(t, x_j) + w(t, x_{j-1})}{(\Delta x)^2} + \mathcal{O}\big((\Delta x)^3\big), \end{aligned}$$

where $\left|\frac{\mathcal{O}((\Delta x)^p)}{(\Delta x)^p}\right| \leq$ const. for small values of Δx. Neglecting the higher-order terms motivates us to define the following approximation formulae to the spatial derivatives:

$$\partial_x w(t, x_j) \approx \frac{w_{j+1}(t) - w_j(t)}{\Delta x}, \tag{A.2}$$

or

$$\partial_x w(t, x_j) \approx \frac{w_j(t) - w_{j-1}(t)}{\Delta x},$$

and

$$\partial_{xx} w(t, x_j) \approx \frac{w_{j+1}(t) - 2w_j(t) + w_{j-1}(t)}{(\Delta x)^2}$$

for $j = 1, ..., N-1$. In general, one can derive the approximating formula for any derivative by using the Taylor series expansion of the function $w(t, x_j)$, and taking into account that $x_j = x_{j\pm k} \pm k\Delta x$ for any k with $x_{j\pm k} \in [a, b]$.

Example A.2. For Example A.1.a), the spatially discretised[2] problem takes the form:

$$\tfrac{\mathrm{d}}{\mathrm{d}t} w_j(t) = \frac{w_{j+1}(t) - 2w_j(t) + w_{j-1}(t)}{(\Delta x)^2}, \quad \text{for } j = 1, ..., N-1.$$

The cases $j = 0$ and $j = N$ are given by the boundary condition $w(0, t) = 0$ as $w_0(t) = w_N(t) = 0$ for all $t \geq 0$. The ordinary differential equations above can be formulated as a system of ordinary

[2] It is sometimes called *semi-discretisation.*

differential equations $\frac{\mathrm{d}}{\mathrm{d}t}W(t) = MW(t)$ with the matrix

$$M = \frac{1}{(\Delta x)^2} \cdot \begin{pmatrix} -2 & 1 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -2 & 1 \\ 0 & 0 & \cdots & 0 & 1 & -2 \end{pmatrix} \in \mathbb{R}^{(N-1)\times(N-1)} \tag{A.3}$$

and the vector $W(t) = \big(w_1(t), ..., w_{N-1}(t)\big) \in \mathbb{R}^{N-1}$. We note that matrix M is of special form, it is a *tridiagonal matrix*, i.e., $M = \frac{1}{(\Delta x)^2}\mathrm{tridiag}(1, -2, 1)$ that has non-zero elements only in its main diagonal and sub-diagonals. Linear systems of this type are much easier to treat numerically.

Example A.3. Example A.1.b) can be spatially discretised as follows:

$$\tfrac{\mathrm{d}}{\mathrm{d}t} w_j(t) = \frac{w_{j+1}(t) - w_j(t)}{\Delta x}, \quad \text{for } j = 1, ..., N$$

with $w_0(t) = 0$ for all $t \geq 0$. Introducing the vector $W(t) = \big(w_1(t), ..., w_N(t)\big) \in \mathbb{R}^N$ and the matrix $M = \frac{1}{\Delta x}\mathrm{tridiag}(-1, 1, 0) \in \mathbb{R}^{N\times N}$, we have the system of ordinary differential equations $\frac{\mathrm{d}}{\mathrm{d}t}W(t) = MW(t)$.

A.2 Galerkin methods

In contrast to finite difference methods, which approximate the exact solution at certain grid points, Galerkin methods use a linear combination of some basis functions. This time let us formulate the problem on the abstract Hilbert space H which is equipped with the inner product $\langle\cdot,\cdot\rangle$. Then for some operator $A : D(A) \subseteq H \to H$ and a given $f \in H$ we consider the following problem:

$$f = Au \quad \text{in } H. \tag{A.4}$$

For instance, let $Au = u''$ on $H = \mathrm{L}^2(0, \pi)$, see Section 1.1. Taking finite dimensional subspaces $H_m \subset H$, $\dim H_m = m$, with a corresponding basis $\{\varphi_1^m, ..., \varphi_m^m\} \subset H_m$, each element $u_m \in H_m$ can be written as the linear combination of the basis functions, i.e.,

$$u_m = \sum_{k=1}^m c_k \varphi_k^m$$

with coefficients $c_k \in \mathbb{C}$, $k = 1, ..., m$. The main idea behind the Galerkin methods is that the exact solution $u \in H$ is approximated by a sequence $(u_m) \subset H_m$ for $m \to \infty$. For notational simplicity we shall drop the superscript from φ_j^m and write φ_j only.

Take the inner product of both sides of problem (A.4) with φ_j for $j = 1, ..., m$:

$$\langle f, \varphi_j\rangle = \langle Au, \varphi_j\rangle, \quad j = 1, ..., m.$$

We define the Galerkin methods by replacing u by $u_m = \sum_{k=1}^m c_k\varphi_k$ in the equation above. Using the linearity of A, we obtain

$$\langle f, \varphi_j\rangle = \left\langle A\sum_{k=1}^m c_k\varphi_k, \varphi_j\right\rangle = \sum_{k=1}^m c_k\big\langle A\varphi_k, \varphi_j\big\rangle, \quad j = 1, ..., m.$$

The definitions of the vectors

$$\Phi := \begin{pmatrix} \langle f, \varphi_1 \rangle \\ \vdots \\ \langle f, \varphi_m \rangle \end{pmatrix} \quad \text{and} \quad C := \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix}$$

and the matrix

$$A_m = \begin{pmatrix} \big\langle A\varphi_1, \varphi_1 \big\rangle & \dots & \big\langle A\varphi_m, \varphi_1 \big\rangle \\ \big\langle A\varphi_1, \varphi_2 \big\rangle & \dots & \big\langle A\varphi_m, \varphi_2 \big\rangle \\ \vdots & & \vdots \\ \big\langle A\varphi_1, \varphi_m \big\rangle & \dots & \big\langle A\varphi_m, \varphi_m \big\rangle \end{pmatrix}$$

enable us to formulate the problem as a system of linear equations

$$A_m C = \Phi.$$

The idea and “procedure” of the Galerkin methods is the following.

1. Choose a subspace $H_m \subset D(A) \subset H$ with $\dim H_m = m$, and a basis $\{\varphi_1, ..., \varphi_m\} \subset H_m$.
2. Solve the system $A_m C = \Phi$. Its solution is $C = \big(c_1, ..., c_m\big)$.
3. The approximation u_m to u is then constructed as $u_m = \sum_{k=1}^m c_k \varphi_k$.

Under appropriate assumptions one can prove that $u_m \to u$ as $m \to \infty$ in the H-norm.

We remark that in some special cases the Galerkin method can be formulated as a variational problem called **Ritz method**, see Exercise 2.

The question arises how to choose the basis functions φ_j, $j = 1, ..., m$. There are two basic possibilities: (i) eigenfunctions of A, (ii) functions being zero outside a small interval where they piecewise coincide with low order polynomials. The corresponding methods are called **spectral method** and **finite element method**, respectively.

Example A.4. Consider the one-dimensional *Poisson equation* on $\mathrm{L}^2(0,\pi)$ with a given function f and homogeneous boundary condition:

$$\begin{aligned} \tfrac{\mathrm{d}^2}{\mathrm{d}x^2} w(x) &= f(x), \quad x \in (0,\pi) \\ w(0) &= w(\pi) = 0. \end{aligned} \tag{A.5}$$

Here we have $A = \frac{\mathrm{d}^2}{\mathrm{d}x^2}$ and $D(A) = \mathrm{H}^2(0,\pi) \cap \mathrm{H}_0^1(0,\pi)$. In order to solve this equation by using the Galerkin method, we choose a finite dimensional subspace $H_m \subset D(A)$ and the appropriate basis functions φ_j, $j = 1, ..., m$. As before, we define the approximation $w_m \in H_m$ as:

$$w_m(x) := \sum_{k=1}^m c_k \varphi_k(x)$$

with coefficients c_k, $k = 1, ..., m$, to be determined from the system

$$\sum_{k=1}^m c_k \langle A\varphi_k, \varphi_j \rangle = \langle f, \varphi_j \rangle$$

$$\sum_{k=1}^m c_k \int_0^\pi (A\varphi_k)(x)\varphi_j(x)\mathrm{d}x = \int_0^\pi f(x)\varphi_j(x)\mathrm{d}x$$

$$\sum_{k=1}^m c_k \int_0^\pi \tfrac{\mathrm{d}^2}{\mathrm{d}x^2}\varphi_k(x)\varphi_j(x)\mathrm{d}x = \int_0^\pi f(x)\varphi_j(x)\mathrm{d}x.$$

Integrating the left-hand side by parts, we obtain

$$\sum_{k=1}^{m} c_k \left(\tfrac{\mathrm{d}}{\mathrm{d}x}\varphi_k(x)\varphi_j(x)\Big|_{x=0}^{x=\pi} - \int_0^\pi \tfrac{\mathrm{d}}{\mathrm{d}x}\varphi_k(x)\tfrac{\mathrm{d}}{\mathrm{d}x}\varphi_j(x) \right),$$

and hence,

$$\sum_{k=1}^{m} -c_k \int_0^\pi \tfrac{\mathrm{d}}{\mathrm{d}x}\varphi_k(x)\tfrac{\mathrm{d}}{\mathrm{d}x}\varphi_j(x)\mathrm{d}x = \int_0^\pi f(x)\varphi_j(x)\mathrm{d}x \quad \text{for } j = 1, ...m, \tag{A.6}$$

where we used the "boundary condition" of the basis functions (they vanish at 0 and at π).

In what follows we determine the approximate solution $w_m(x)$ of problem (A.5) by applying two different sets of basis functions which correspond to spectral and finite element methods, respectively.

a) *Spectral method* (cf. Example 3.3): Since $A = \frac{\mathrm{d}^2}{\mathrm{d}x^2}$ with $D(A) = \mathrm{H}^2(0,\pi) \cap \mathrm{H}_0^1(0,1)$, we can choose the finite dimensional subspace as $\mathrm{lin}\{\sin(jx) : j = 1, ..., m\}$ (cf. Exercise 1.1). Then equation (A.6) results in

$$\begin{aligned}\sum_{k=1}^{m} -c_k \int_0^\pi \tfrac{\mathrm{d}}{\mathrm{d}x}\sin(kx)\tfrac{\mathrm{d}}{\mathrm{d}x}\sin(jx)\mathrm{d}x &= \sum_{k=1}^{m} -c_k \int_0^\pi k\cos(kx) j\cos(jx)\mathrm{d}x \\ &= \sum_{k=1}^{m} -c_k jk \int_0^\pi \cos(kx)\cos(jx)\mathrm{d}x = \int_0^\pi f(x)\sin(jx)\mathrm{d}x.\end{aligned}$$

Due to the orthogonality of basis functions, the Kronecker delta δ_{jk} appears on the left-hand side:

$$\sum_{k=1}^{m} -c_k jk\delta_{jk}\frac{\pi}{2} = \int_0^\pi f(x)\sin(jx)\mathrm{d}x,$$

which leads to the values

$$c_j = -\frac{2}{j^2\pi}\int_0^\pi f(x)\sin(jx)\mathrm{d}x \quad \text{for all } j = 1, ..., m.$$

The approximation $w_m(x)$ to $w(x)$ is then

$$w_m(x) = -\frac{2}{\pi}\sum_{k=1}^{m}\frac{1}{k^2}\sin(kx)\int_0^\pi f(s)\sin(ks)\mathrm{d}s.$$

In this case, the matrix A_m has entries $(A_m)_{jk} = -jk\frac{\pi}{2}\delta_{jk}$ only in its main diagonal which contains then the square numbers $1, ..., m^2$ multiplied by $-\frac{\pi}{2}$.

b) *Finite element method*: Another possible choice for basis functions are the functions

$$\varphi_j(x) := \begin{cases} 0, & \text{for } x < (j-1)\Delta x \\ \dfrac{x}{\Delta x} - (j-1), & \text{for } (j-1)\Delta x \le x < j\Delta x \\ (j+1) - \dfrac{x}{\Delta x}, & \text{for } j\Delta x \le x < (j+1)\Delta x \\ 0, & \text{for } (j+1)\Delta x \le x, \end{cases}$$

which are sometimes called "hat functions", see Example 3.4. Here $\Delta x = \frac{\pi}{m}$ for some $m \in \mathbb{N}$. Their first derivative exists piecewise and can be calculated easily:

$$\tfrac{\mathrm{d}}{\mathrm{d}x}\varphi_j(x) := \begin{cases} 0, & \text{for } x < (j-1)\Delta x \\ \dfrac{1}{\Delta x}, & \text{for } (j-1)\Delta x < x < j\Delta x \\ -\dfrac{1}{\Delta x}, & \text{for } j\Delta x < x < (j+1)\Delta x \\ 0, & \text{for } (j+1)\Delta x < x. \end{cases}$$

Equation (A.6) yields

$$\begin{aligned} \sum_{k=1}^{m} -c_k \int_0^{\pi} \tfrac{\mathrm{d}}{\mathrm{d}x}\varphi_k(x)\tfrac{\mathrm{d}}{\mathrm{d}x}\varphi_j(x)\mathrm{d}x &= \Delta x \left(-c_{j-1}\frac{-1}{(\Delta x)^2} - c_j\frac{2}{(\Delta x)^2} - c_{j+1}\frac{-1}{(\Delta x)^2} \right) \\ &= \frac{1}{\Delta x}(c_{j-1} - 2c_j + c_{j+1}) = \int_0^{\pi} f(x)\varphi_j(x)\mathrm{d}x \quad \text{for } j = 1, ..., m, \end{aligned}$$

with the basis functions φ_j defined above. The matrix A_m has now the tridiagonal form $A_m = \frac{1}{\Delta x}\mathrm{tridiag}(1, -2, 1)$.

Note that the basis functions φ_j, $j = 1, ..., m$, do not belong to $\mathrm{H}^2(0, \pi)$, hence the matrix A_m has to be defined by using (A.6), i.e., the weak formulation is necessary here.

Exercises

1. Consider the heat equation in two dimensions for $(x, y) \in \Omega = (0, \pi) \times (0, \pi)$ and $t \geq 0$:

$$\partial_t w(t, x, y) = \partial_{xx} w(t, x, y) + \partial_{yy} w(t, x, y)$$

with the boundary condition

$$w(t, x, y) = 0 \qquad \text{on } \partial\Omega,$$

where $\partial\Omega$ denotes the boundary of Ω. Derive the form of the corresponding matrix obtained when applying finite differences to discretise the operator $L = \partial_{xx} + \partial_{yy}$ (cf. matrix M in (A.3)).

2. Let H be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$, and $A : D(A) \subset H \to H$ be a linear densely defined operator possessing the following properties:

a) A is *symmetric* on $D(A)$, that is, $\langle Au, v \rangle = \langle u, Av \rangle$ for all $u, v \in D(A)$, and

b) A is *strongly elliptic*, that is, there exists a constant $c > 0$ such that $\langle Au, u \rangle \geq c\|u\|^2$ for all $u \in D(A)$.

For all $v \in D(A)$ and a given element $f \in H$ define the functional $F : D(A) \to \mathbb{R}$ by

$$F(v) := \langle Av, v \rangle - 2\langle f, v \rangle.$$

Show that if $Au = f$ for $u \in D(A)$ then the functional F is minimal, i.e. $F(u) < F(v)$ for all $v \in D(A)$, $v \neq u$.

3. Derive the form of matrix A_m in Examples A.4.a) and b).

Appendix B

Basic Time Discretisation Methods

A standard numerical approach for solving partial differential equations is the method of lines. There the problem is first discretised with respect to the space variable(s), leading to a system of ordinary differential equations. These ODEs are then solved using time discretisation methods. Basic space discretisations have already been introduced in Appendix A. Here we collect some facts about time discretisations. Our basic reference is the textbook *Solving ordinary differential equations I: Nonstiff problems*, by E. Hairer, S.P. Nørsett, and G. Wanner.[1]

Throughout the lectures t will denote the time variable. Let $f : [t_0, t_{\max}] \times \mathbb{R}^m \to \mathbb{R}^m$ be a continuous function, $y_0 \in \mathbb{R}^m$. We consider the following initial value problem on the time interval $[t_0, t_{\max}]$:

$$\begin{cases} y'(t) = f\big(t, y(t)\big) \\ y(t_0) = y_0, \end{cases} \tag{B.1}$$

where $y : [t_0, t_{\max}] \to \mathbb{R}^m$ is the unknown function to be determined.

In order to solve problem (B.1) numerically, we divide the time interval $[t_0, t_{\max}]$ into N pieces:

$$t_0 < t_1 < t_2 < \cdots < t_N = t_{\max}$$

with **variable time steps** $h_n = t_{n+1} - t_n$, $n = 0, ..., N-1$. Such a sub-divison of the time interval is often called a **time grid**. In the case of an equidistant grid with $h_0 = h_1 = \cdots = h_{N-1} =: h$, the lenght $h = \frac{t_{\max} - t_0}{N}$ is usually referred to as **constant time step**. The numerical scheme consists in approximating the exact solution $y(t_n)$ at time levels t_n for $n = 0, ..., N$. This approximation is called a **numerical solution** to problem (B.1) and is denoted by y_n. We note that for an equidistant time grid, the time levels are computed as $t_n = t_0 + nh$ (especially, $t_n = nh$ with $t_0 = 0$).

B.1 Euler's method

The simplest time discretisation method was introduced by L. Euler[2] in 1768. His idea was to replace the derivative y' by an approximation to the tangent, i.e., quotient by the difference quotient. Since we have

$$f(t, y) = y' = \lim_{h \to 0} \frac{y(t+h) - y(t)}{h}$$

by the initial value problem (B.1), we obtain **Euler's method** as

$$f(t_n, y_n) = \frac{y_{n+1} - y_n}{t_{n+1} - t_n},$$

[1] E. Hairer, S.P. Nørsett, G. Wanner, Solving ordinary differential equations I: Nonstiff problems. Springer Verlag, 2008.

[2] L. Euler, Institutionum Calculi Integralis. Volumen Primum, Opera Omnia **XI**, 1768.

usually written as

$$y_{n+1} = y_n + h_n f(t_n, y_n). \tag{B.2}$$

In order to demonstrate a drawback of Euler's method for problems we have in mind, and to motivate a modification, let us consider the problem (1.1)(1.2) appearing in Lecture 1:

$$\begin{cases} y'(t) = -k^2 y(t), \\ y(t_0) = y_0, \end{cases} \tag{B.3}$$

for $n \in \mathbb{N}$ with $t_0 = 0$ and $y_0 = 1$. Note that the exact solution $y(t) = \mathrm{e}^{-k^2 t}$ is positive for all t and bounded by 1. After the first step with equidistant step size h, application of Euler's method (B.2) yields

$$y_1 = y_0 + h(-k^2 y_0) = (1 - hk^2)y_0 = 1 - hk^2.$$

Note that the choice $h > \frac{1}{k^2}$ leads to $y_1 < 0$. The second step yields

$$y_2 = y_1 + h(-k^2 y_1) = (1 - hk^2) + h(-k^2)(1 - hk^2) = (1 - hk^2)^2 > 0.$$

It is easy to see that the numerical solution has the following form after n steps:

$$y_n = (1 - hk^2)^n.$$

For $h > \frac{1}{k^2}$ the value of y_n changes sign at each step, i.e., (y_n) is an alternating sequence for $n \in \mathbb{N}$. Moreover, in contrast to the boundedness of the exact solution $y(t)$, the sequence $\big(|y_n|\big)$ is unbounded for this particular choice of h. These phenomena suggest us (i) to choose the time step carefully, or (ii) to modify Euler's method.

Note that we have the freedom to choose the time level appearing in the argument of the function f. Taking t_{n+1} instead of t_n (i.e., we use the tangent at t_{n+1}), we obtain the **implicit Euler method**:

$$y_{n+1} = y_n + h_n f(t_{n+1}, y_{n+1}). \tag{B.4}$$

For the implicit Euler method, the numerical solution of problem (B.3) after one step reads

$$y_1 = y_0 + h(-k^2 y_1) \quad \Longrightarrow \quad (1 + hk^2)y_1 = y_0 \quad \Longrightarrow \quad y_1 = \frac{1}{1 + hk^2} \cdot y_0 = \frac{1}{1 + hk^2}.$$

The second step yields

$$y_2 = y_1 + h(-k^2 y_2) \quad \Longrightarrow \quad (1 + hk^2)y_2 = y_1 \quad \Longrightarrow \quad y_2 = \frac{1}{1 + hk^2} \cdot y_1 = \frac{1}{(1 + hk^2)^2}.$$

Hence, the numerical solution after n steps,

$$y_n = \frac{1}{(1 + hk^2)^n}$$

results in positive values for all choices of the step size h. Moreoverm it is bounded by 1.

The method is called implicit, because we have to solve a nonlinear system of equations to compute y_{n+1}. The question arises whether one can determine y_{n+1} from (B.4). It is, however, garanteed by the implicit function theorem. We note that explicit and implicit Euler methods are often called forward and backward Euler methods, respectively.

B.2 Runge's method

In order to derive the method introduced by C. Runge[3] in 1905, let us consider the problem (B.1) again, and integrate it between t_n and $t_{n+1} = t_n + h_n$:

$$\int_{t_n}^{t_n+h_n} y'(t)\,\mathrm{d}t = \int_{t_n}^{t_n+h_n} f\big(t, y(t)\big)\,\mathrm{d}t$$

$$y(t_n + h_n) = y(t_n) + \int_{t_n}^{t_n+h_n} f\big(t, y(t)\big)\,\mathrm{d}t. \tag{B.5}$$

The question is now how to approximate the above integral. The first two possibilities

$$\int_{t_n}^{t_n+h_n} f(t, y(t))\,\mathrm{d}t = h_n f(t_n, y_n) + \mathcal{O}(h_n^2)$$

or

$$= h_n f(t_{n+1}, y_{n+1}) + \mathcal{O}(h_n^2)$$

yield the explicit and implicit Euler methods, respectively. Instead of using only the left or right grid point, the function f can also be evaluated at the midpoint $t_n + \frac{h_n}{2}$:

$$\int_{t_n}^{t_n+h_n} f\big(t, y(t)\big)\,\mathrm{d}t = h_n f\big(t_n + \tfrac{h_n}{2}, y(t_n + \tfrac{h_n}{2})\big) + \mathcal{O}(h_n^3),$$

which is already of order 2. To approximate the unknown term $y(t_n + \frac{h_n}{2})$, we use Taylor's formula:

$$y\left(t_n + \tfrac{h_n}{2}\right) = y(t_n) + \frac{h_n}{2} f(t_n, y_n) + \mathcal{O}(h_n^2). \tag{B.6}$$

We note that this corresponds to an explicit Euler step. **Runge's method** for computing y_{n+1} is then defined by

$$y_{n+1} = y_n + h_n f\big(t_n + \tfrac{h_n}{2}, y_n + \tfrac{h_n}{2} f(t_n, y_n)\big). \tag{B.7}$$

B.3 Runge–Kutta methods

Following the idea of the derivation of Runge's method, one can generalise it by applying certain quadrature rules, being common in numerical integration, to approximate the integral appearing in formula (B.5). As already seen, application of left or right Riemann sums lead to explicit or implict Euler method, respectively. The midpoint rule combined with (B.6) yields Runge's method. For the trapezoidal rule we obtain the **Crank–Nicolson method**

$$y_{n+1} = y_n + \tfrac{h_n}{2}\big(f(t_n, y_n) + f(t_{n+1}, y_{n+1})\big) \tag{B.8}$$

being of great importance in practice.

[3] C. Runge, "Über die numerische Auflösung totaler Differentialgleichungen", Göttinger Nachr. (1905), 252–257.

In case of general *quadrature rules*, for $s \in \mathbb{N}$ we define $c_1, ..., c_s$ distinct real numbers between 0 and 1. Then we define the corresponding collocation polynomial u of degree s whose derivative coincides with the function f at the collocation points $t_n + c_i h_n$ for $i = 1, ..., s$, that is,

$$u'(t_n + c_j h_n) = f\big(t_n + c_j h_n, u(t_n + c_j h_n)\big), \quad \text{for } j = 1, ..., s \tag{B.9}$$

with $$u(t_n) = y_n.$$

The numerical solution y_{n+1} of problem (B.1) at time level $t_{n+1} = t_n + h_n$ is then defined as

$$y_{n+1} := u(t_n + h_n). \tag{B.10}$$

In order to compute $u(t_n + h_n)$, let us denote $k_j := u'(t_n + c_j h_n)$ for $j = 1, ..., s$. Note that the Lagrange interpolation polynomials

$$l_i(\tau) = \prod_{\substack{i=1 \\ i \neq m}}^{s} \frac{\tau - c_m}{c_i - c_m}$$

satisfy $l_i(c_j) = \delta_{ij}$, where δ_{ij} denotes the Kronecker delta. Therefore, $\sum_{i=1}^{s} k_i l_i(c_j) = k_j$, which further equals to $u'(t_n + c_j h_n)$. By the Lagrange interpolation form, we obtain

$$u'(t_n + \tau h_n) = \sum_{i=1}^{s} k_i l_i(\tau). \tag{B.11}$$

Due to relation (B.10), the numerical solution y_{n+1} can be computed by integrating between 0 and 1 both sides of equation (B.11) above:

$$\begin{aligned}
\int_0^1 u'(t_n + \tau h_n)\, \mathrm{d}\tau &= \int_0^1 \sum_{i=1}^{s} k_i l_i(\tau)\, \mathrm{d}\tau \\
\tfrac{1}{h_n}\big(u(t_n + \tau h_n)\big)\Big|_{\tau=0}^{\tau=1} &= \sum_{i=1}^{s} k_i \underbrace{\int_0^1 l_i(\tau)\, \mathrm{d}\tau}_{=:b_i} \\
\tfrac{1}{h_n}\big(u(t_n + h_n) - u(t_n)\big) &= \sum_{i=1}^{s} k_i b_i.
\end{aligned}$$

Then we obtain

$$u(t_n + h_n) = u(t_n) + h_n \sum_{i=1}^{s} b_i k_i.$$

Since $u(t_n) = y_n$ by construction, the numerical solution defined by (B.10) has the form

$$y_{n+1} = y_n + h_n \sum_{i=1}^{s} b_i k_i. \tag{B.12}$$

In order to determine the values of $k_i = f\big(u(t_n+c_ih_n, u(t_n+c_ih_n)\big)$, $i = 1, ..., s$, we have to compute $u(t_n + c_ih_n)$. To this end, let us integrate (B.11) again, but this time between 0 and c_i:

$$\begin{aligned}
\int_0^{c_i} u'(t_n + \tau h_n)\, \mathrm{d}\tau &= \int_0^{c_i} \sum_{j=1}^{s} k_j l_j(\tau)\, \mathrm{d}\tau \\
\tfrac{1}{h_n}\big(u(t_n + \tau h_n)\big)\Big|_{\tau=0}^{\tau=c_i} &= \sum_{j=1}^{s} k_j \underbrace{\int_0^{c_i} l_j(\tau)\, \mathrm{d}\tau}_{=:a_{ij}} \\
\tfrac{1}{h_n}\big(u(t_n + c_ih_n) - u(t_n)\big) &= \sum_{i=j}^{s} k_j a_{ij}.
\end{aligned}$$

Hence, we obtain

$$u(t_n + c_ih_n) = u(t_n) + h_n \sum_{j=1}^{s} a_{ij}k_j, \quad \text{for } i = 1, ..., s.$$

This means, application of quadrature rules to approximate the integral in formula (B.5) leads to **collocation methods** defined by

$$y_{n+1} = y_n + h_n \sum_{i=1}^{s} b_i k_i \tag{B.13}$$

with
$$k_i = f\big(t_n + c_ih_n, y_n + h_n \sum_{j=1}^{s} a_{ij}k_j\big), \quad \text{for } i = 1, ..., s \tag{B.14}$$

where
$$b_i = \int_0^1 l_i(\tau)\, \mathrm{d}\tau \quad \text{and} \quad a_{ij} = \int_0^{c_i} l_j(\tau)\, \mathrm{d}\tau, \quad \text{for } i, j = 1, ..., s$$

and
$$l_i(\tau) = \prod_{\substack{i=1\\ i\neq m}}^{s} \frac{\tau - c_m}{c_i - c_m} \quad \text{are the Lagrange interpolation polynomials.}$$

We note here that for special choice of coefficients c_i, $i = 1, ..., s$, the order of collocation methods can even equal $2s$.

One obtain a possible generalisation to collocation methods by choosing the coefficients a_{ij}, b_i, c_i arbitrary, instead of assigning them the special values above. For a fixed $s \in \mathbb{N}$ and some coefficients a_{ij}, b_i, c_i for $i, j = 1, ..., s$, time discretisation methods of the form (B.13), (B.14) are called **s-stage Runge–Kutta methods**. We note that if $a_{ij} = 0$ for $i \leq j$ we have explicit, otherwise we have implicit Runge–Kutta methods. Notice that all collocation methods are implicit Runge–Kutta methods.

It is common to collect the coefficients a_{ij}, b_i, c_i of a Runge–Kutta method in the Butcher tableau proposed by J. C. Butcher[4] in 1964:

[4] J. C. Butcher, "On Runge–Kutta processes of high order", J. Austral. Math. Soc. **IV** (1964), 179–194.

$$\begin{array}{c|cccc} c_1 & a_{11} & a_{12} & \cdots & a_{1s} \\ c_2 & a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_s & a_{s1} & a_{s2} & \cdots & a_{ss} \\ \hline & b_1 & b_2 & \cdots & b_s \end{array}$$

Example B.1. Examples for s-stage Runge–Kutta methods.

1. Explicit Euler method ($s = 1$):

$$\begin{array}{c|c} 0 & \\ \hline & 1 \end{array}$$

2. Implicit Euler method ($s = 1$):

$$\begin{array}{c|c} 1 & 1 \\ \hline & 1 \end{array}$$

3. Runge–Kutta method based on Gaussian quadrature with one node ($s = 1$):

$$\begin{array}{c|c} 1/2 & 1/2 \\ \hline & 1 \end{array}$$

4. Runge's method ($s = 2$):

$$\begin{array}{c|cc} 0 & & \\ 1/2 & 1/2 & \\ \hline & 0 & 1 \end{array}$$

5. Crank–Nicolson method ($s = 2$):

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1/2 & 1/2 \\ \hline & 1/2 & 1/2 \end{array}$$

6. Runge–Kutta method based on Radau II A quadrature with two nodes ($s = 2$):

$$\begin{array}{c|cc} 1/3 & 5/12 & \text{-}1/12 \\ 1 & 3/4 & 1/4 \\ \hline & 3/4 & 1/4 \end{array}$$

B.4 Exercises

1. Prove first- and second-order consistency of the explicit Euler method and the explicit method of Runge, respectively.

2. Construct Euler's number e with the help of Euler's method.
Hint: Consider the differential equation $y' = y$ with the initial condition $y(0) = 1$.

3. Derive the Runge–Kutta methods based on Gaussian and Radau II A rules. *Hint: For Gauß, the nodes are $c_1 = \frac{1}{2}$ for $s = 1$ and $c_{1,2} = \frac{1}{2} \pm \frac{\sqrt{3}}{6}$ for $s = 2$. For Radau II A, the nodes are $c_1 = 1$ for $s = 1$ and $c_1 = \frac{1}{3}$, $c_2 = 1$ for $s = 2$.*

4. Analyse the Runge–Kutta methods based on Gaussian and Radau II A rules for the problem $y' = -k^2 y$. Are there any conditions for the step size?

5. Derive the Butcher Tableau for the Crank–Nicolson scheme (B.8).