Stability Analysis of Finite Difference Schemes for the 1D Heat Equation on an Unbounded Spatial Domain

Bachelor's Thesis

Written by: Réka Dávid

Analytical Mathematics

Advisor: Dr. Ágnes Havasi

Réka Dávid
Department of Applied Analysis and Computational Mathematics
Eötvös Loránd University, Faculty of Science, Institute of Mathematics

Eötvös Loránd University
Faculty of Science
2015

Acknowledgement

Firstly, I would like to express my sincere gratitude to my advisor Dr. Ágnes Havasi for the continuous support, her patience, motivation, and immense knowledge. Her guidance helped me in all the time of writing this thesis.

Besides my advisor, I would like to thank my family for supporting me throughout my education.

Contents

Chapter 1

Introduction

Ever since the early days of computing, the use of computers to solve scientific problems has been a driving force. Most real physical processes are governed by partial differential equations (PDEs). A PDE is an equation stating a relationship between a function of two or more independent variables and the partial derivatives of this function with respect to these independent variables. In many cases, simplifying approximations are made to reduce the governing PDEs to ordinary differential equations (ODEs) or even to algebraic equations. However, because of the ever increasing requirement for more accurate modeling, engineers and scientists are more and more involved to solve the actual PDEs that govern the physical problem being investigated.

Most studies are devoted to numerically approximate the solution of partial differential equations by finite difference methods. Our expectation for the numerical solution is that it converges to the exact solution by refinement of the grid. We will see that, if the difference scheme is consistent with the PDE, the stability is necessary and sufficient for providing convergence. The aim of the present work is to investigate conditions for stability of the one-dimensional *heat conduction problem* when discretized with finite difference schemes on an unbounded spatial domain.

Wolfgang Hackbusch's The Concept of Stability in Numerical Mathematics [1] was followed for the main structure of the thesis. First, we define the one-dimensional heat conduction problem on two Banach-spaces $L^2(\mathbb{R})$ and $C(\mathbb{R})$ of continuous functions in spatial variable x with the respective norms. Next, we examine the exact solution of the problem and some of its qualitative properties. In the third chapter, we introduce the operator semigroup background for the further investigations. Then, we move on to the discretization of the heat conduction problem by finite difference methods. We introduce the discrete spaces l^∞ and l^2 and the transfer operators to provide mapping from the continuous spaces onto the discrete ones. We also define the difference scheme in terms of that shift operator for the numerical approximation and apply our considerations on a programming task. In Chapter 5, we present the classical theory of Lax and Richtmyer stating that for a consistent scheme stability is sufficient and necessary for convergence. Then we derive some simple sufficient and necessary conditions for stability. Next, we introduce the theory of Fourier analysis, define the Fourier synthesis and verify further conditions for stability. The schemes examined so far in the thesis are, at best, conditionally stable. Hence, at last, to obtain unconditionally stable schemes, we admit the implicit difference schemes and some more conditions for stability.

Chapter 2

The heat conduction problem

The *heat conduction problem* describes the distribution of temperature in a given region over time. We restrict our considerations to one spatial variable x. The changing in time is proportional to the second derivative of the temperature in space, yields the parabolic equation

$$\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2} \quad \text{for } \ t > 0, x \in \mathbb{R} \tag{2.1}$$

where α, the *heat conduction coefficient*, is assumed to be constant. We restrict α to be 1 (otherwise transform $t \to \alpha t$ or $x \to \sqrt{\alpha}x$). Let B be a space of functions in spatial variable x. The desired solution of the equation is denoted by $u(t, x)$ of the independent variables x and t. By $u(t) = u(t, \cdot) \in B$ we denote the function partially evaluated at t. Hence, $u(t)(x)$ and $u(t, x)$ are equivalent notations. Let $I = [0, T]$ be the interval in which t varies. Then, the domain of $u(\cdot, \cdot)$ is the set $\Sigma := I \times \mathbb{R}$ with $I = [0, T]$ while $x \in \mathbb{R}$. The spatial domain $\mathbb{R}$ is chosen – also for the periodic case – as unbounded to avoid boundary conditions. If the solutions are assumed to be 2π-periodic in x, we use the domain $\Sigma := I \times [0, 2\pi]$ with $u(t, 0) = u(t, 2\pi)$. It is possible to rewrite the problem to an abstract Cauchy problem with differential operator A. We assume that the domain of A is defined as $D_A := \{v \in B : Av \in B \text{ is defined}\} \subset B$, hence the partial differential equation takes the following form: the task is to find a continuous function $u : I \to D_A$ which solves the Cauchy-problem

$$\begin{cases} \frac{\mathrm{d}u(t)}{\mathrm{d}t} = Au(t) & \text{for } \forall t \in I \\ u(0) := u_0 & \text{for some } u_0 \in D_A \subset B \end{cases} \tag{2.2}$$

Focusing on the differential operator A, we discuss the model

$$A := \frac{\partial^2}{\partial x^2}$$

As it is well known, the solution of problem (2.1) with a continuous initial function is given by

$$u(t, x) = \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} u_0(\varepsilon) \exp\left(\frac{-\left(x - \varepsilon\right)^2}{4t} \right) d\varepsilon \quad \text{for } \ t > 0, x \in \mathbb{R}. \tag{2.3}$$

The solution satisfies (2.1) and provides point-wise convergence $\lim_{t\to 0} u(t,x) = u_0(x)$ for all $x \in \mathbb{R}$. The representation (2.3) shows that the solution is infinitely often differentiable at $t > 0$ although the initial value $u_0(x)$ is only continuous. However, the solution exists only for $t > 0$.

Our considerations are built on two Banach spaces, generally denoted by B:

- $B := C(\mathbb{R})$ is the space of complex-valued and uniformly continuous functions with finite supremum $\|u\|_B = \|u\|_\infty := \sup\{|u(x)| : x \in \mathbb{R}\}$
- $B := L^2(\mathbb{R})$ is the space of complex-valued, measurable and square-integrable functions with a finite L^2-norm $\|u\|_B = \|u\|_2 := \sqrt{\int_{\mathbb{R}} |v(x)|^2 dx}$. This Banach space is also a Hilbert space with the dot product

$$\langle u, v\rangle := \int_{(R)} u(x)\overline{v(x)}dx.$$

For the 2π-periodic case, we can redefine the spaces as

- $C_{per}(\mathbb{R}) := \{v \in C(\mathbb{R}) : v(x) = v(x + 2\pi)\ \forall x \in \mathbb{R}\}$
- $L^2_{per}(\mathbb{R}) := \{v \in L^2(\mathbb{R}) : v(x) = v(x + 2\pi) \text{ for almost all } x \in \mathbb{R}\}$.

Further on, we will verify a qualitative property of the exact solution of the *heat equation*, namely, that its norm is not increasing in time. The next proposition is necessary to prove Lemma 2.3.

Proposition 2.1. *The equality*

$$\sqrt{4\pi t} = \int_{-\infty}^{\infty} \exp\left(\frac{-\delta^2}{4t}\right) d\delta$$

holds for $t > 0$.

Proof. *Suppose that*

(1) means applying polar-coordinate substitution

(2) means applying $u = r^2$ substitution.

Denote by I the integral on the right-hand side of (1.3). Then

$$\begin{aligned}
I^2 &:= \int_{-\infty}^{\infty} \exp\left(\frac{-\delta^2}{4t}\right) d\delta \int_{-\infty}^{\infty} \exp\left(\frac{-y^2}{4t}\right) dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(\frac{-(\delta^2+y^2)}{4t}\right) d\delta dy \overset{(1)}{=} \\
&= \int_0^{\infty}\int_0^{2\pi} r\cdot \exp(-r^2\frac{1}{4t})d\varphi dr = 2\pi\int_0^{\infty} r\cdot\exp(-r^2\frac{1}{4t})dr \overset{(2)}{=} \pi\int_0^{\infty}\exp(-u\frac{1}{4t})du = \\
&= \pi\left| -4t\cdot\exp(-u\frac{1}{4t})\right|_0^{\infty} = 4t\pi
\end{aligned}$$

From $I = \sqrt{4t\pi}$ the equation follows. ■

Lemma 2.2 (∞-norm case)**.** *Let $u(t) \in B := C(\mathbb{R})$ be a solution of problem (2.1). The inequality $\|u(t)\|_\infty \leqslant \|u_0\|_\infty$ holds for $\forall t \geqslant 0$.*

Proof. *We suppose* $u : I \times \mathbb{R} \to D_A \subset C(\mathbb{R})$ *and* $\frac{\partial^2 u}{\partial x^2} \in C(\mathbb{R})$. *Let us define* $c := \|u_0\|_\infty$ *and take the magnitude of the solution (2.3). Then*

$$|u(t)| = \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} |u_0(\varepsilon)| \exp\left(\frac{-(x-\varepsilon)^2}{4t}\right) d\varepsilon \overset{def}{\leqslant} \frac{c}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} \exp\left(\frac{-(x-\varepsilon)^2}{4t}\right) d\varepsilon \overset{2.1}{=} \|u_0\|_\infty .$$

We take the supremum of both sides, thus $\|u(t)\|_\infty \leqslant \|u_0\|_\infty$ *follows.* ■

Lemma 2.3 (2-norm case)**.** *Let* $u(t) \in B := L^2(\mathbb{R})$ *be a solution of problem (2.1). The inequality* $\|u(t)\|_2 \leqslant \|u(0)\|_2$ *holds for* $\forall t \geqslant 0$.

Proof. *We suppose that* $u : I \times \mathbb{R} \to D_A \subset L^2(\mathbb{R})$ *and* $\frac{\partial^2 u}{\partial x^2} \in L^2(\mathbb{R})$*, furthermore* $t'' > t' > 0$. *Hence,*

$$2\int_{t'}^{t''} \int_{\mathbb{R}} u(t) \frac{\partial^2}{\partial x^2} u(t) dx dt = 2\int_{t'}^{t''} \left[u(t)\frac{\partial}{\partial x}u(t)\right]_{-\infty}^{+\infty} dt - 2\int_{t'}^{t''} \int_{\mathbb{R}} \frac{\partial u(t)}{\partial x} \cdot \frac{\partial u(t)}{\partial x} dx dt$$

where the rule of integration by parts was used for the inner integral. Since the solution decays towards plus and minus infinity, the first term on the right-hand side is zero. On the other hand

$$\|u(t'')\|_2^2 - \|u(t')\|_2^2 \overset{def}{=} \int_{\mathbb{R}} u(t'')^2 dx - \int_{\mathbb{R}} u(t')^2 dx = \int_{\mathbb{R}} \int_{t'}^{t''} \frac{\partial}{\partial t} u(t)^2 dt dx$$
$$= 2\int_{t'}^{t''} \int_{\mathbb{R}} u(t) \frac{\partial}{\partial t} u(t) dx dt = -2\int_{t'}^{t''} \int_{\mathbb{R}} \frac{\partial u(t)}{\partial x} \cdot \frac{\partial u(t)}{\partial x} dx dt \leqslant 0$$

where we used the Newton-Leibnitz rule and the fact that the solution u is continuous, thus the integrals can be interchanged. Our conclusion is that $\|u(t)\|_2$ *is weakly decreasing, i.e.* $\|u(t)\|_2 \leqslant \|u_0\|_2$ *for* $\forall t \geqslant 0$. ■

Chapter 3

Semigroup of solution operators

To analyze the consistency, convergence and stability of finite difference methods applied for the *heat equation* later, the notion of operator semigroups will be helpful. In this chapter, besides Hackbusch's book [1], we rely on [11], [12] and chapter 9 of [5]. We know that solution $u(t) \in B$ is *strong* or *classical*, if it satisfies problem (2.1) in $\mathbb{R}$ with initial function u_0 from B_0, where $B_0 := C^\infty(\mathbb{R}) \cap B$, a dense subset of B. For the abstract formulation of the problem, we can define the *solution operator* $S(t)$ as

$$S(t) : u_0 \mapsto u(t)$$

at fixed $t \geqslant 0$. So, the solution operator $S(t)$ maps the initial function $u_0 \in B_0$ to the solution at time t. It is easy to show that $S(t)$ is a linear mapping. Moreover, since $\|u(t)\|_B \leqslant \|u_0\|_B$, therefore $S(t)$ is bounded and continuous too. Hence, $S(t) \in \mathcal{L}(B_0, B)$ for all $t \geqslant 0$. The definition allows the derivation of some properties of the solution from the properties of $S(t)$.

The translation of the function u_0 satisfies the property $S(t+s) = S(t)S(s)$. So that, by using $S(t)$, it is possible to redefine problem (2.1) to a general starting point $s \geqslant 0$.

Proposition 3.1. *The property $S(t)S(s) = S(t+s)$ holds for all $t, s \geqslant 0$:*

Proof. *Suppose the case of strong solution, thus $S(t) \in \mathcal{L}(B_0, B)$ and $S(\tau)u_0 = u(\tau)$. Set $u(s)$ equal to $S(s)u_0 := \hat{u}_0$ for a fixed $s \geqslant 0$. Stepping further, $\hat{u}(t) := u(t+s) = S(t)\hat{u}_0$. Then,*

$$\frac{\partial \hat{u}(t)}{\partial t} = \frac{\partial u(t+s)}{\partial t} \stackrel{def}{=} Au(t+s) = A\hat{u}(t)$$

is problem (2.1) with a possible starting point $\hat{u}_0$. It yields to the property

$$S(t+s)u_0 = u(t+s) = \hat{u}(t) = S(t)\hat{u}_0 = S(t)S(s)u_0 \implies S(t+s) = S(t)S(s). \blacksquare$$

Next, we show the property that $S(t)$ and A commute, which results in the fact that on D_A $S(t)$ maps into D_A. We can call $\{S(t), 0 \geqslant t\}$ a one-parameter semigroup with neutral element $S(0) = I$ and A as the infinitesimal generator of the semigroup.

Proposition 3.2. *A and $S(t)$ commute, i.e. $AS(t) = S(t)A$ on B_0.*

Proof. *Consider a strong solution $u(t) = S(t)u_0$ for $u_0 \in B_0$. Then,*

$$A[S(t)u_0] = Au(t) \overset{(2.2)}{=} \frac{\mathrm{d}u(t)}{\mathrm{d}t} = \lim_{h\to 0} \frac{u(t+h)-u(h)}{h} \overset{def}{=} \lim_{h\to 0} \frac{1}{h}[S(t+h)-S(t)]u_0 \overset{3.1}{=}$$
$$= S(t) \lim_{h\to 0} \frac{1}{h}[S(h)-S(0)]u_0 = S(t) \lim_{h\to 0} \frac{u(h)-u_0}{h} = S(t)\frac{\partial u_0}{\partial t} = S(t)[Au_0]$$

where we used that $S(0) = I$ and the continuity of $S(t)$ on B_0. $AS(t) = S(t)A$ follows. ■

Since $S(t)[Au_0] \in B$ is defined for all $u_0 \in B_0$, also $A[S(t)u_0]$ has this property. This proves that $S(t) : D_A \to D_A$ for $t \geqslant 0$. So far, $S(t)$ was defined on B_0 and it was satisfying:

- $S(t+s) = S(t)S(s) \quad t, s > 0$,
- $S(0) = I$,
- $Av = \lim_{t\to 0} \frac{S(t)v-v}{t}$ for $\forall v \in D_A$.

In the case of $B = C(\mathbb{R})$, the elements of which are uniformly continuous, $\lim_{t\to 0} \|S(t)v - v\| = 0$ also follows for all $v \in B_0$. Another notation of $S(t)$ is e^{At}. Now, we recall the density of B_0 and that $\|u(t)\|_B \leqslant \|u_0\|_B$ has been proved for the heat conduction problem. We infer that operator $S(t)$ can be extended uniquely and continuously onto B.

Theorem 3.3. *Suppose $S(t) \in \mathcal{L}(B_0, B)$ and B_0 is dense in B. Then, $S(t)$ can be extended uniquely and continuously onto B. Furthermore, the extended $S_e(t) \in \mathcal{L}(B, B)$ and the original $S(t) \in \mathcal{L}(B_0, B)$ have equal norms, i.e.* $\sup\{\|S(t)v\|_B / \|v\|_B : v \in B_0\backslash\{0\}\} = \sup\{\|S_e(t)v\|_B / \|v\|_B : v \in B\backslash\{0\}\}$.

The point of the extension to B is that, due to the density of $B_0 \subset B$, for all $v \in B$ there exists a sequence $(v_{0,n}) \subset B_0$ for which $\lim v_{0,n} = v$. Since $v_{0,n} \in B_0$, $\forall n \in \mathbb{N}$, $S(t)$ can be applied to each element $v_{0,n}$ of this sequence, therefore the extension of $S(t)$ onto B can be defined as: $S(t)v := \lim S(t)v_{0,n}$, where $v \in B$. To continue the thread, the existence of a strong solution (2.3) cannot be always guaranteed for $u_0 \in B\backslash B_0$, but there is a resulting function for any $u_0 \in B$ such that $u(t) := S(t)u_0 \in B$ is a *weak* or *generalized solution*. The weak solution is a function for which the derivatives may not all exist but which nonetheless satisfies the differential equation in some precisely defined sense, namely the *variational equation*; we multiply the equation (2.1) by a smooth test function $v \in C^\infty(\mathbb{R})$ and integrate both sides over $\mathbb{R}$.

In the generalized case, the solutions are $u(t) := S(t)u_0 \in B$ with $u_0 \in B$. We would like to provide that $u(t) = S(t)u_0$ tends to $u_0 \in B$ while $t \to 0$. It can be shown that if u_0 is uniformly continuous, then for the solution $u(t)$ of the heat conduction problem $u \to u_0$ uniformly on $\mathbb{R}$. In other words, $\|u(t) - u_0\|_B \to 0$, so $\lim_{t\to 0} \|[S(t) - S(0)]u_0\|_B = 0$. Therefore, in this case the semigroup has the property $\lim_{t\to 0} \|[S(t) - S(0)]u_0\|_B = 0$ too, i.e. the semigroup is continuous. Proposition 3.1 of the case where $u_0 \in B_0$ is also true for the presented one, since B_0 is dense in B. The range $t \geqslant 0$ may be reduced to an interval $[0, \tau]$ with $\tau > 0$, which leads to the following

Proposition 3.4. *Suppose that $K_\tau := \sup_{0\leqslant t\leqslant \tau} \|S(t)\|_{B\to B} < \infty$ for $\tau > 0$. Then,*

$$\|S(t)\|_{B\leftarrow B} \leqslant K_\tau^{\lceil \frac{t}{\tau} \rceil} \quad \textit{for any } t \geqslant 0$$

where $\lceil x \rceil := \min\{n \in \mathbb{Z} : x \leqslant n\}$.

Proof. *We will divide the proof into three cases:*

(1). Suppose that $t = 0$, *then*

$$\left\lceil \frac{t}{\tau} \right\rceil = 0 \text{ thus } K_\tau^{\lceil \frac{t}{\tau} \rceil} = 1 \text{ and } S(t) = I \text{ yield to } \|S(t)\|_{B \leftarrow B} = 1.$$

(2). Suppose that $t \in (0, \tau]$. *Then*

$$\left\lceil \frac{t}{\tau} \right\rceil = 1 \text{ thus } \|S(t)\|_{B \to B} \leqslant \sup_{0 \leqslant t \leqslant \tau} \|S(t)\|_{B \to B} \stackrel{def}{=} K_\tau^1.$$

(3). Suppose that $t > 0$ *is out of the interval* $[0, \tau]$. *There exists* $n \in \mathbb{Z}$ *and* δ *with* $0 < \delta < \tau$ *so that* $t = n\tau + \delta$. *Then, by the semigroup property (3.1) we have*

$$\|S(t)\|_{B \to B} = \|S(\delta) S(\tau)^n\|_{B \to B} .$$

Now, $t = n\tau + \delta$ *implies*

$$n = \frac{t - \delta}{\tau} \leqslant \frac{t}{\tau}$$

and since $\|S(0)\|_{B \to B} = 1$ *yields* $K_\tau \geqslant 1$, *we have*

$$\|S(t)\|_{B \to B} \leqslant K_\tau^{\lceil \frac{\delta}{\tau} \rceil} (K_\tau^{\lceil \frac{\tau}{\tau} \rceil})^n = K_\tau^{n + \lceil \frac{\delta}{\tau} \rceil} = K_\tau^{\lceil \frac{n\tau + \delta}{\tau} \rceil} = K_\tau^{\lceil \frac{t}{\tau} \rceil}.$$

For each case, the inequality with $\tau > 0$ *follows.* ■

Further on in the next chapters, we shall refer to the inequality with $\tau = T$ and $t \in I = [0, T]$, hence $\lceil \frac{t}{T} \rceil = 1$. Moreover, $S(t)$ is supposed to be uniformly bounded on $I = [0, T]$:

$$\|S(t)\|_{B \leftarrow B} \leqslant K_T \quad \text{ for } \forall t \in I = [0, T]. \tag{3.1}$$

For the heat equation, the condition (3.1) is satisfied with $K_T = 1$, since we showed that the norm of the solution was not increasing in time and $\|S(t)\|_{B \leftarrow B} = \sup_{u_0 \neq 0} \{\|u(t)\|_B \,/\, \|u_0\|_B\} \leqslant \|u_0\|_B \,/\, \|u_0\|_B = 1$.

Chapter 4

Discretization of the heat conduction problem

We move on to the discretization of the *heat conduction problem* for numerical approaches. In this chapter, our considerations build on chapter 3 of [9]. The properties of the generalized case resulted in a uniformly continuous mapping from $u_0 \in B$ into $u(t) \in B$ on the bounded interval $t \in [0, T]$. Since the spatial variable x is defined on $\mathbb{R}$, it is possible to create an infinite grid of step size $\Delta x > 0$ as

$$G_{\Delta x} := \{x = v\Delta x : v \in \mathbb{Z}\}.$$

The time variable ranges in the interval $[0, T]$, so we define the finite grid of step size $\Delta t > 0$ as

$$I_{\Delta t} := \{t = \mu\Delta t \leqslant T : \mu \in \mathbb{N}_0\}.$$

Correspondingly, from the Cartesian product of both grids, it is also possible to define the rectangular grid

$$\Sigma_{\Delta x}^{\Delta t} := \{(t, x) \in \Sigma : \frac{x}{\Delta x} \in \mathbb{Z} \text{ and } \frac{t}{\Delta t} \in \mathbb{N}_0\}. \tag{4.1}$$

where we used the notation $\Sigma := I \times \mathbb{R}$. The step sizes Δt and Δx are chosen independently, but are connected by $\lambda = \frac{\Delta t}{\Delta x^2}$. Using the rectangular grid as the domain, we can define the grid function $U : \Sigma_{\Delta x}^{\Delta t} \to \mathbb{C}$ with the notation

$$U_v^\mu := u(\mu\Delta t, v\Delta x) \quad \text{for } (\mu\Delta t, v\Delta x) \in \Sigma_{\Delta x}^{\Delta t} \tag{4.2}$$

For a fixed parameter $t = \mu\Delta t$, the infinite sequences on $\Sigma_{\Delta x}^{\Delta t}$ are

$$\begin{aligned} U^\mu &:= \{u(\mu\Delta t, v\Delta x) : v \in \mathbb{Z}\} = \{U_v^\mu\}_{v \in \mathbb{Z}} \\ U_v &:= \{u(\mu\Delta t, v\Delta x) : \mu \in \mathbb{N}_0\} = \{U_v^\mu\}_{\mu \in \mathbb{N}_0} \end{aligned} \tag{4.3}$$

The set of two-sided infinite grid functions together with the operations of addition and scalar multiplication form a linear space, on which we will use two different norms, leading to the normed spaces l^2 and l^∞:

$$\begin{aligned} \|U\|_{l^2} &:= \sqrt{\Delta x \sum_{v\in\mathbb{Z}} |U_v|^2} \quad \text{for the space } l^2\text{, a Hilbert-space} \\ \|U\|_{l^\infty} &:= \sup\{|U_v| : v \in \mathbb{Z}\} \quad \text{for the space } l^\infty\text{, a Banach-space} \end{aligned} \tag{4.4}$$

4.1 Transfer operators r and p

The continuous Banach space can be mapped onto the linear space of grid functions via a transfer operator. The transfer operator r is defined as a *restriction*

$$r := r_{\Delta x} : B \to l^p,$$

where Δx is the step size of the infinite grid $G_{\Delta x}$. We need to distinguish the mapping in the sense of $B = C(\mathbb{R}) \vee L^2(\mathbb{R})$. An obvious choice of $ru \in l^\infty$ is

$$u \in C(\mathbb{R}) \Longrightarrow ru \in l^\infty \text{ with } (ru)_j := u(j\Delta x) \quad \text{for } j \in \mathbb{Z}$$

By substituting it into (4.4), the ∞-norm can be calculated as the supremum of the absolute values of the discrete $(ru)_j$ points. A choice $ru \in l^2$ is not that self-evident, since $L^2(\mathbb{R})$ functions have no well-defined point evaluations. In this case, the point evaluation is replaced by the integral mean value of the function on the sub-interval $[x_j - \Delta x/2, x_j + \Delta x/2]$ around the grid point x_j:

$$u \in L^2(\mathbb{R}) \Longrightarrow ru \in l^2 \text{ with } (ru)_j := \frac{1}{\Delta x}\int_{(j-\frac{1}{2})\Delta x}^{(j+\frac{1}{2})\Delta x} u(x)dx \quad \text{for } j \in \mathbb{Z}$$

We get the 2-norm as the square root of the sum of discrete squared integrals multiplied by Δx. We can suppose that the norm of r is bounded with respect to the norms $\|\cdot\|_{l^\infty \leftarrow C(\mathbb{R})}$ and $\|\cdot\|_{l^2 \leftarrow L^2(\mathbb{R})}$

$$\|r_{\Delta x}\|_{l^p \leftarrow B} \leqslant C_r \quad \text{for } \forall \Delta x > 0 \tag{4.5}$$

We show that the restrictions considered above satisfy the condition (4.5) with $C_r := 1$. In the case of the 2-norm we use the *Cauchy-Schwarz inequality*:

Theorem 4.1 (The Cauchy-Schwarz inequality)**.** *For all vectors x and y of a dot product space it is true that*

$$|\langle x, y\rangle|^2 \leqslant \langle x, x\rangle\langle y, y\rangle,$$

Equivalently, by taking the square root of both sides, and referring to the norms of vectors generated by the dot product, the inequality is written as

$$|\langle x, y\rangle| \leqslant \|x\| \, \|y\| \, .$$

Moreover, the two sides are equal iff x and y are linearly dependent.

Lemma 4.2. *The restrictions r satisfy the condition (4.5) with $C_r = 1$ with respect to both considered norms.*

Proof. *The basic idea is that from $B \to l^p$ we need to prove the inequality $\|ru\|_{l^p} \leqslant \|u\|_B$, thus $\|r_{\Delta x}\|_{l^p \leftarrow B} \leqslant 1$ is also satisfied.*
(1). For $p = \infty$, it is possible to use the definition of supremum of function u, then:

$$\|ru\|_{l^\infty} \overset{def}{=} \sup\{|u(j\Delta x)| : j \in \mathbb{Z}\} \leqslant \sup\{|u(x)|\}, x \in \mathbb{R}\} = \|u\|_{B=C(\mathbb{R})}$$

where $u(x)$ is the continuous function and $u(j\Delta x)$, $j \in \mathbb{Z}$ represents a subset of function values.
(2). For $p = 2$,

$$\|ru\|_{l^2}^2 = \Delta x \sum\nolimits_j |(ru)_j|^2 = \frac{1}{\Delta x} \sum\nolimits_j \Big| \int_{(j-\frac{1}{2})\Delta x}^{(j+\frac{1}{2})\Delta x} u(x)dx \Big|^2,$$

we apply the Cauchy-Schwarz inequality for the interior of the sum, then:

$$\|ru\|_{l^2}^2 \leqslant \sum\nolimits_j \int_{(j-\frac{1}{2})\Delta x}^{(j+\frac{1}{2})\Delta x} |u(x)|^2 dx = \int_{\mathbb{R}} |u(x)|^2 dx = \|u(x)\|_{B=L^2}^2. \ \blacksquare$$

So far, we have been dealing with *restriction operator* r. The *prolongation* p acts in the reverse direction, such that the operator maps the discrete space onto the continuous space, so

$$p := p_{\Delta x} : l^p \to B.$$

We can also suppose that p is bounded with respect to the norms $\|\cdot\|_{B:=C(\mathbb{R})\leftarrow l^\infty}$ and $\|\cdot\|_{B:=L^2(\mathbb{R})\leftarrow l^2}$ as

$$\|p_{\Delta x}\|_{B\leftarrow l^p} \leqslant C_p \quad \text{for } \forall \Delta x > 0. \tag{4.6}$$

Furthermore, suppose $rp = I$, which indicates that p is a right-inverse of r. We also need to distinguish the mapping in the sense of l^∞ and l^2.

Lemma 4.3. *The prolongations with the following choices of p satisfy condition (4.6) with $C_p = 1$:*
(1) p is a piecewise linear interpolation:

$$v \in l^\infty \mapsto pv \in C(\mathbb{R}) \text{ with } (pv)(x) = \vartheta v_j + (1-\vartheta)v_{j+1}$$
$$\text{where } x = (j+\vartheta)\Delta x,\ j \in \mathbb{Z},\ \vartheta \in [0,1)$$

(2) p is a piecewise constant interpolation:

$$v \in l^2 \mapsto pv \in L^2(\mathbb{R}) \text{ with } (pv)(x) = v_j$$
$$\text{where } x \in \Big[\Big(j - \frac{1}{2}\Big)\Delta x, \Big(j + \frac{1}{2}\Big)\Delta x\Big),\ j \in \mathbb{Z}.$$

Proof. *The basic idea is that from $l^p \to B$ we need to prove inequality $\|pv\|_B \leqslant \|v\|_{l^p}$, thus $\|r_{\Delta x}\|_{B\leftarrow l^p} \leqslant 1$ is also satisfied. During the poof, we use the triangle inequality and that function v is continuous.*

In case (1):

$$\|(pv)(x)\|_{B=C(\mathbb{R})} = \sup_{j\in\mathbb{Z}} \sup\{|\vartheta v_j + (1-\vartheta)v_{j+1}| : \vartheta \in [0,1)\} \leqslant$$
$$\leqslant \sup_j(\max\{|v_j|, |v_{j+1}|\}) = \sup_j\{|v_j|\} = \|v\|_{l^\infty}$$

where, we use the fact that pv is linear and inhomogeneous on interval $[x_j, x_{j+1}]$*, so its supremum on the subinterval equals* $\sup\{|v_j|, |v_{j+1}|\}$.
In case (2):

$$\|(pv)(x)\|^2_{B=L^2(\mathbb{R})} = \int_{\mathbb{R}} |(pv)(x)|^2 dx = \sum_{j\in\mathbb{Z}} \int_{(j-\frac{1}{2})\Delta x}^{(j+\frac{1}{2})\Delta x} 1|(pv)(x)|^2 dx$$
$$\leqslant \sum_{j\in\mathbb{Z}} \left(\int_{(j-\frac{1}{2})\Delta x}^{(j+\frac{1}{2})\Delta x} 1dx\right)\left(\int_{(j-\frac{1}{2})\Delta x}^{(j+\frac{1}{2})\Delta x} |v_j|^2 dx\right) = \Delta x \sum_{j\in\mathbb{Z}} \int_{(j-\frac{1}{2})\Delta x}^{(j+\frac{1}{2})\Delta x} |v_j|^2 dx = \|v\|^2_{l^2}$$

then taking the square root of both sides gives the desired form. ■

We say l^p with $p \in \{2, \infty\}$ is suited to B if condition (4.5) and (4.6) hold for $\forall \Delta x > 0$.

4.2 Difference Schemes

The next task is to define the time stepping operator that maps $U^{\mu+1}$ to U^μ. In this section, we rely on lecture 3 of [10], chapter 6 of [8] and chapter 2 of [7]. We prescribe U^0 via $U^0_v = ru_0$. The *explicit difference scheme* is defined as

$$U^{\mu+1}_v = \sum_{j\in\mathbb{Z}} a_j U^\mu_{v+j}, \tag{4.7}$$

where $a_j \in \mathbb{R}$ are fixed constants. In practice $\sum_{j\in\mathbb{Z}}$ is a finite sum, all a_j vanish. The difference scheme describes a linear mapping $U^\mu \to U^{\mu+1}$ and motivates to define the linear *difference operator*

$$C : l^p \to l^p \text{ where } (CU)_v := \sum_{j\in\mathbb{Z}} a_j U_{v+j} \quad \text{for } U \in l^p \tag{4.8}$$

where $p \in \{2, \infty\}$ and $C^j : U^\mu \mapsto U^{\mu+j}$. As a result, a shorter form is $U^{\mu+1} = CU^\mu$, and a μ-fold application leads to $U^\mu = C^\mu U^0$. The coefficients a_j are assumed to be scalars. If the scalar equation $\frac{\mathrm{d}}{\mathrm{d}t}u = Au$ for $u : I \times \mathbb{R} \mapsto \mathbb{R}$ is replaced by a vector-valued equation, the coefficients a_j become $N \times N$ matrices. The difference operator C (and in this case a_j) may depend on the parameter λ and step size Δt, thereby also on Δx via $\lambda = \frac{\Delta t}{\Delta x^2}$. Thus, C will be denoted as $C = C(\lambda, \Delta t)$ for the rest of the chapter. On the other hand, we can define the *shift operator*

$$E_j : l^p \mapsto l^p, \text{ where } (E_j U)_v := U_{j+v}.$$

Since the shift operator does not change the l^p-norm of U, therefore $\|E_j\|_{l^p \leftarrow l^p}$ is equal to 1. The operator C can be replaced by

$$C := \sum_{j \in \mathbb{Z}} a_j E_j = \sum_{j \in \mathbb{Z}} C_j. \tag{4.9}$$

A simple representative example of a Δt-dependent difference operator can be given. Let operator $C(\lambda)$ be a suitable *difference operator* for $A = \frac{\partial^2}{\partial x^2}$ of the equation 2.1. Then the *differential operator* $A = \frac{\partial^2}{\partial x^2} + b$ can be discretized by $C'(\lambda, \Delta t) := C(\lambda) + \Delta t \cdot b$.

In the case of $A = \frac{\partial^2}{\partial x^2}$, the difference quotient

$$\frac{u(t + \Delta t, x) - u(t, x)}{\Delta t} \quad \text{for } \frac{\partial u}{\partial t}$$

and the second difference quotient

$$\frac{u(t, x - \Delta x) - 2u(t, x) + u(t, x + \Delta x)}{\Delta x^2} \quad \text{for } \frac{\partial^2 u}{\partial x^2}$$

are obvious choices and lead together with $\lambda = \frac{\Delta t}{\Delta x^2}$ to

$$\frac{u(t + \Delta t, x) - u(t, x)}{\Delta t} = \frac{2u(t, x) - u(t, x - \Delta x) - u(t, x + \Delta x)}{\Delta x^2}$$

hence

$$U_v^{\mu+1} = \lambda U_{v-1}^{\mu} + (1 - 2\lambda) U_v^{\mu} + \lambda U_{v+1}^{\mu},$$

and the presented *difference scheme* is as follows:

$$C : l^p \mapsto l^p, \qquad (CU)_v = \lambda U_{v-1} + (1 - 2\lambda) U_v + \lambda U_{v+1} \qquad \text{for } U \in l^p. \tag{4.10}$$

Next, we will apply this explicit scheme in a programming task with different values of the parameter λ, and investigate its convergence.

The programming task. Consider the heat conduction problem with initial function

$$u(x, 0) := 1.$$

Since the second derivative with respect to x is equal to zero and so the time derivative of the temperature is zero by the heat equation, the exact solution of the problem is $u(x, t) \equiv 1$. We perturb the initial function at $x = 0$ by a small value, namely, $1e - 5$, and solve the problem by using the above explicit method with $\Delta t = 0.01$ for $\lambda = 0.45$ and $\lambda = 0.7$ (and $\Delta x = \sqrt{\Delta t / \lambda}$). We expect that the numerical solution will not differ very much from the constant 1 solution of the unperturbed problem. It is important to mention that we set the task of the problem on an infinite grid for spatial variable x, however to calculate the numerical solution, we have to restrict the computation to the grid points of a finite interval. Here we use the interval $[-5, 5]$ and the time interval was chosen as $[0, 1]$.

The source code of the task can be found in Appendix A and Appendix B. Since the applied explicit formula relies on one more neighboring point at both sides of the chosen spatial interval $[xmin, xmax]$, we have to start the computation from the wider interval $[xmin - \Delta x \cdot tmax/\Delta t, xmax + \Delta x \cdot tmax/\Delta t]$. Figure 4.1 shows that the solution is spoiled for $\lambda = 0.7$, even if we refine the grid, however for $\lambda = 0.45$ the numerical solution is very close to the constant 1 solution that we would obtain by using the unperturbed constant 1 initial condition. The different colors belong to different time layers, as it can be seen below.

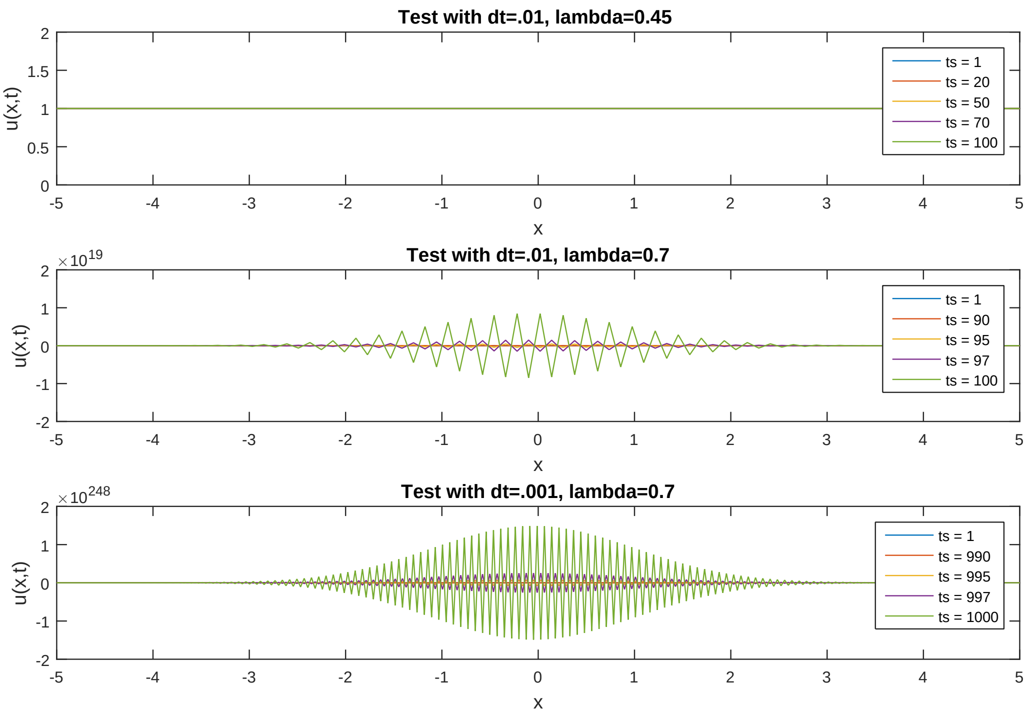

Figure 4.1: The results of the programming task for different values of parameter λ.

In the legends, ts should be understood as the number of time-steps of length Δt that were made during the calculation. In the following we would like to get an explanation to the bad behavior of the numerical solution in this example. We will see that for certain choices of the parameter λ instability arises, which prevents the numerical solution from being convergent.

Chapter 5

Consistency, convergence, stability

Let $B_0 \subset D_A$ be a dense subset of B and l^p with $p \in \{2, \infty\}$ be suited to B. Furthermore, we denote the solution by $u(t) = S(t)u_0$ where $u_0 \in B_0$. The local discretization error E is defined by

$$\mathrm{E(t)} := \frac{1}{\Delta t}[ru(t + \Delta t) - C(\lambda, \Delta t)ru(t)].$$

The difference scheme $C(\lambda, \Delta t)$ is called consistent with respect to the linear spaces l^p with $p \in \{2, \infty\}$ if for all $u_0 \in B_0$

$$\sup\{\|\mathrm{E}(t)\|_{l^p} : 0 \leqslant t \leqslant T - \Delta t\} \to 0 \quad \text{ as } \Delta t \to 0.$$

Using error E, the latter condition is equivalent to

$$\sup_{0 \leqslant t \leqslant T - \Delta t} \|[rS(\Delta t) - C(\lambda, \Delta t)r]S(t)u_0\|_{l^p} = \mathcal{O}(\Delta t) \quad \text{ for all } u_0 \in B_0.$$

Next, the definition of convergence and stability refer to the whole Banach space B, not only to a dense subset B_0. Now, we can suppose the generalized solution denoted by $u(t) = S(t)u_0$ for all $u_0 \in B$. The difference scheme $C(\lambda, \Delta t)$ is called convergent with respect to l^p if

$$\|ru(t) - C(\lambda, \Delta t)^{\mu} ru_0\|_{l^p} \to 0 \quad \text{ for } \Delta t \to 0 \text{ and } \mu\Delta t \to t \in I = [0, T].$$

The difference scheme is called stable with respect to l^p if

$$\sup\{\|C(\lambda, \Delta t)^{\mu}\|_{l^p \leftarrow l^p} : \Delta t \geqslant 0, \mu \in \mathbb{N}_0, 0 \leqslant \Delta t \leqslant T\} < \infty.$$

In this case, the stability constant is defined by

$$K = K(\lambda) := \sup\{\|C(\lambda, \Delta t)^{\mu}\|_{l^p \leftarrow l^p} : \Delta t \geqslant 0, \mu \in \mathbb{N}_0, 0 \leqslant \Delta t \leqslant T\}. \tag{5.1}$$

If the stability condition holds only for certain values of λ, the scheme is called *conditionally stable*, otherwise *unconditionally stable*. We will see that the schemes examined so far are, at best, conditionally stable. Before, we show that consistency and stability with some technical assumptions imply convergence, so stability is sufficient for convergence.

Theorem 5.1 (convergence theorem)**.** *Suppose*

- *r is bounded with respect to l^p,*
- *$S(t)$ satisfies assumption (3.1),*
- *l^p stability of the difference scheme $C(\lambda, \Delta t)$,*
- *l^p consistency,*

then the difference scheme is convergent with respect to l^p.

Proof. *We divide the proof into two cases, based on $u_0 \in B_0$ or $u_0 \in B$.*
Case (1): we suppose $u_0 \in B_0$ and define $u(t) = S(t)u_0$. We split the discretization error E *as*

$$ru(t) - C(\lambda, \Delta t)^\mu ru_0 = r[u(t) - u(\mu\Delta t)] + [rS(\Delta t)^\mu - C(\lambda, \Delta t)^\mu r]u_0,$$

and use the telescopic sum

$$rA^\mu - B^\mu r = \sum_{v=0}^{\mu-1} B^v [rA - Br] A^{\mu-v-1}$$

with $A := S(\Delta t)$, $B := C(\lambda, \Delta t)$ and (4.5) and (5.1). Then

$$\begin{aligned}
&\|ru(t) - C(\lambda, \Delta t)^\mu ru_0\|_{l^p} \leqslant \|r\|_{l^p \leftarrow B} \|u(t) - u(\mu\Delta t)\|_B \\
&+ \sum_{v=0}^{\mu-1} \|\{rS(\Delta t) - C(\lambda, \Delta t)r\} u((\mu - v - 1)\Delta t)\|_{l^p} \leqslant \\
&\leqslant K_r \|u(t) - u(\mu\Delta t)\|_B + \sum_{v=0}^{\mu-1} K(\lambda)\Delta t \|\mathrm{E}((\mu - v - 1)\Delta t)\|_{l^p}.
\end{aligned}$$

Since $u_0 \in B_0 \subset D_A$, the solution is strong, therefore continuous. The first term is a zero sequence, which tends to zero while $\mu\Delta t \to t$. Because of the consistency assumption, the local discretization error tends to zero uniformly, i.e. $\forall \varepsilon > 0 \; \exists \Delta t > 0$ for which $\|\mathrm{E}((\mu - v - 1)\Delta t)\|_{l^p} \leqslant \varepsilon$. We infer that the telescopic sum is bounded with $K(\lambda)T\varepsilon$, which yields that the whole sum tends to zero and proves convergence in the case of an initial function $u_0 \in B_0$.
Case (2): we suppose a general initial function $u_0 \in B$ and a given $\varepsilon > 0$. One finds $u_0^* \in B_0$ that is sufficiently close to $u_0 \in B$, so that $\|u_0 - u_0^*\|_B \leqslant \varepsilon/[3K_r \max \|S(t)\|_{B \leftarrow B}, K(\lambda)]$. The associated solution $u^*(t) = S(t)u_0^*$ satisfies

$$\|r[u(t) - u^*(t)]\|_{l^p} \leqslant K_r \|S(t)[u_0 - u_0^*]\|_B \leqslant K_r \|S(t)\|_{B \leftarrow B} \|u_0 - u_0^*\|_B \leqslant \frac{\varepsilon}{3}$$

and

$$\|C(\lambda, \Delta t)^\mu r[u_0 - u_0^*]\|_{l^p} \leqslant K(\lambda)K_r \|u_0 - u_0^*\|_B \leqslant \frac{\varepsilon}{3}.$$

Together with $\|ru^*(t) - C(\lambda, \Delta t)^\mu ru_0^*\|_{l^p} \leqslant \frac{\varepsilon}{3}$ from case (1) for sufficiently small Δt and $t - \mu\Delta t$, it follows that $\|ru(t) - C(\lambda, \Delta t)^\mu ru_0\|_{l^p} \leqslant \varepsilon$. Therefore, convergence is shown for a general initial function. ■

Theorem 5.2 (stability theorem). *Choose l^p to be suited to B, so that conditions (3.1), (4.5) and (4.6) hold. Then, l^p convergence implies l^p stability.*

The latter theorem implies that stability is also necessary for convergence. The point of necessity of stability for providing convergence is that there are not any sequences $\Delta t_v > 0$, $\mu_v \in \mathbb{N}_0$ with $0 \leqslant \mu_v \Delta t_v \leqslant T$ for which the $\|\cdot\|_{l^p \leftarrow l^p}$ norm of the difference scheme is not bounded. Due to the compactness of $I = [0, T]$, there is a subsequence with $(\mu_v \Delta t_v) \to t \in I$, which yields that $\|ru(t) - C(\lambda, \Delta t_v)^{\mu_v} ru_0\|_{l^p} \to 0$ for all $u_0 \in B$. Therefore the l^p-norm of $C(\lambda, \Delta t_v)^{\mu_v} ru_0$ is bounded for sufficiently large values of ν. One concludes that $C_v := C(\lambda, \Delta t_v)^{\mu_v} r$ is a point-wise bounded sequence of operators, therefore uniformly bounded. Since p is a bounded inverse of r, the boundedness of $\|C(\lambda, \Delta t_v)^{\mu_v}\|_{l^p \leftarrow l^p}$ also follows.

The conclusion of theorems 5.1 and 5.2 is that if we suppose (3.1), (4.5), (4.6) and l^p consistency, then l^p convergence and l^p stability are equivalent. In other words, if the difference scheme is consistent with the PDE, then stability is necessary and sufficient for providing convergence.

So far, we restricted the analysis to the l^p space with $p \in \{2, \infty\}$. A more involved analysis is necessary to describe the properties for l^p with $2 < p < \infty$. On the one hand, for $p > 2$ the Hilbert structure is lost and certain properties also change between the case $p < \infty$ and $p = \infty$. On the other hand, if stability estimates hold for both $p = 2$ and $p = \infty$, these bounds imply corresponding estimates for the l^p and L^p setting for $2 < p < \infty$. In the following lemma, let $\|\cdot\|_{p \leftarrow p}$ be the operator norm of $\mathcal{L}(l^p, l^p)$ or $\mathcal{L}(L^p, L^p)$.

Lemma 5.3 (Riesz-Thorin theorem). *Assume $1 \leqslant p_1 \leqslant q \leqslant p_2 \leqslant \infty$. Then*

$$\|\cdot\|_{q \leftarrow q} \leqslant \|\cdot\|_{p_1 \leftarrow p_1}^{\alpha} \|\cdot\|_{p_2 \leftarrow p_2}^{\beta}$$

with $\alpha = \frac{q - p_1}{p_2 - p_1}$ and $\beta = \frac{p_2 - q}{p_2 - p_1}$.

It is clear to see that $\alpha + \beta = 1$. As a conclusion, p_1 stability $\|\cdot\|_{p_1 \leftarrow p_1} \leqslant M_1$ and p_2 stability $\|\cdot\|_{p_2 \leftarrow p_2} \leqslant M_2$ imply q stability in the form

$$\|\cdot\|_{q \leftarrow q} \leqslant M := M_1^{\alpha} M_2^{\beta}.$$

Hence, if a criterion yields both l^2 and l^∞ stability, then l^p stability holds for all $2 \leqslant p \leqslant \infty$.

5.1 Sufficient and Necessary Conditions for Stability

The following results belong to the classical stability theory of Lax-Richtmeyer [13].

We assume that $\|C(\lambda, \Delta t)\|_{l^p \leftarrow l^p} \leqslant 1 + K_\lambda \Delta t$. Then

$$\|C(\lambda, \Delta t)^{\mu}\|_{l^p \leftarrow l^p} \leqslant \|C(\lambda, \Delta t)\|_{l^p \leftarrow l^p}^{\mu} \leqslant (1 + K_\lambda \Delta t)^{\mu} \leqslant (e^{K_\lambda \Delta t})^{\mu} \leqslant e^{K_\lambda T},$$

therefore the difference scheme is l^p stable with stability constant $K(\lambda) := e^{T K_\lambda}$. The coefficients a_j of $C(\lambda, \Delta t)$ can be used to estimate $\|C(\lambda, \Delta t)\|_{l^p \leftarrow l^p}$ as

$$\|C(\lambda, \Delta t)\|_{l^p \leftarrow l^p} = \left\| \sum\nolimits_j a_j E_j \right\|_{l^p \leftarrow l^p} \leqslant \sum\nolimits_j |a_j| \, \|E_j\|_{l^p \leftarrow l^p} = \sum\nolimits_j |a_j|.$$

Combining the previous statements, we can state the following criterion for stability.

Proposition 5.4. *We suppose that* $\sum_j |a_j| \leqslant 1 + K_\lambda \Delta t$ *for all* $\Delta t > 0$. *Then, the difference scheme* $C(\lambda, \Delta t)$ *is stable with stability constant* $K(\lambda) := e^{K_\lambda T}$.

If $a_j \geqslant 0$ holds for the coefficients, a difference scheme is called positive and it maps non-negative initial functions into non-negative solutions. The positive difference schemes with $\sum_j a_j = 1 + \mathcal{O}(\Delta t)$ are stable with respect to the l^2 and l^∞ norms. To mention an example, for $\lambda \in (0, 1\backslash 2]$ the scheme (4.10) is positive, and the latter criterion ensures l^2 and l^∞ stability. We can also formulate a criterion for instability.

Proposition 5.5. *We suppose that* $\sum a_j \geqslant 1 + \Delta t c(\Delta t)$ *with* $\lim_{\Delta t \to 0} c(\Delta t) = \infty$. *Then the difference scheme is unstable with respect to the* l^2 *and* l^∞ *norms.*

In the present chapter, we only assume a necessary statement for proving 5.5 in chapter 6. We suppose that $\lim_{\Delta t \to 0} c(\Delta t) = \infty$. Then

$$\sup\{[1 + \Delta t c(\Delta t)]^\mu : \mu \in \mathbb{N}_0, \Delta t > 0, \mu \Delta t \leqslant T\} = \infty. \tag{5.2}$$

Later on, we will use that (5.2) also implies that $C_p := \sup_{\Delta t > 0} c(\Delta t) < \infty$ is possible with setting $c(\Delta t) := \frac{\sqrt[\mu]{K} - 1}{\Delta t}$ for any $K > 0$ and all positive integer $\mu \leqslant T/\Delta t$. Therefore, the μth-root of K satisfies the inequality

$$\sqrt{K}^\mu \leqslant 1 + C_p \Delta t. \tag{5.3}$$

An interesting question is whether a stable scheme remains stable after a perturbation. We show that the stability of $C(\lambda, \Delta)$ implies the stability of the scheme after perturbation in both cases that $C(\lambda, \Delta t)$ and perturbation $D(\lambda, \Delta t)$ are commutative operators or not.

Lemma 5.6 (perturbation lemma)**.** *Let* $C(\lambda, \Delta t)$ *be* l^p *stable with stability constant* $K(\lambda)$. *Suppose that a perturbation operator* $D(\lambda, \Delta t)$ *is bounded by*

$$\|D(\lambda, \Delta t)\|_{l^p \to l^p} \leqslant C_D \Delta t.$$

Then

$$C'(\lambda, \Delta t) := C(\lambda, \Delta t) + D(\lambda, \Delta t)$$

is again l^p *stable with stability constant*

$$K'(\lambda) \leqslant K(\lambda) e^{K(\lambda) C_D T},$$

where $I = [0, T]$ *is the given interval. This result holds also in the case where* $C(\lambda, \Delta t)$ *and* $D(\lambda, \Delta t)$ *are non-commutative operators, which is interesting in the case of matrix-valued coefficients* a_j.

A simple application of the *perturbation lemma* is the following one. Let us consider the *differential operator* A in $du/dt = Au$ be $A = A_1 + A_0$, where A_1 contains derivatives of at least first order, while $A_0 u = a_0 u$ is the term of order zero. The discretization also yields $C(\lambda, \Delta t) = C_1(\lambda, \Delta t) + C_0(\lambda, \Delta t)$, where a consistent discretization of C_0 can be estimated as $\|C_0(\lambda, \Delta t)\|_{l^p \leftarrow l^p} = \mathcal{O}(\Delta t)$. So that, by the *perturbation lemma* the stability of C_1 implies the stability of C. Therefore, it suffices to investigate A without terms of order zero. In this case, $A1 = 0$, where $1 \in l^\infty$ the constant function with value 1, shows that $u := 1$ is a solution of the *heat conduction problem*. This implies the special consistency condition $\sum_j a_j = 1$ for the coefficients of $C(\lambda, \Delta t)$ in (4.10).

Next, we introduce the spectral radius

$$\rho(A) := \sup\{|\lambda| : \lambda \text{ singular value of } A\}.$$

Using section 6.5 of [5], we define λ as regular value of A, if $\lambda I - A$ is bijective, and the inverse $(\lambda I - A)^{-1} \in \mathcal{L}(B, B)$ exists. Otherwise, λ is a singular value of A, i.e. the inverse $(\lambda I - A)^{-1} \in \mathcal{L}(B, B)$ does not exist. Furthermore, $p(A^\mu)$ equals $p(A)^\mu$ and $p(A) \leqslant \|A\|$ is valid for any associated norm. These properties with (5.3) yield a necessary condition for stability with respect to the l^2 and l^∞ norms as

$$\rho(C(\lambda, \Delta t)) \leqslant 1 + \mathcal{O}(\Delta t). \tag{5.4}$$

In the case that the operator C of the difference scheme is normal, it is possible to show that the spectral radius is equal to the l^2 norm of $C(\lambda, \Delta t)$ and the l^2 stability is equivalent to the latter necessary condition (5.4). Let $C(\lambda, \Delta t)$ be denoted by C.

Proposition 5.7. *We suppose that $C \in \mathcal{L}(l^2, l^2)$ is normal, i.e. C commutes with the adjoint operator C^*. Then $p(C) = \|C\|_{l^2 \leftarrow l^2}$ holds and l^2 stability is equivalent to $\rho(C) \leqslant 1 + \mathcal{O}(\Delta t)$.*

Proof. *First, we prove that $\left\|C^2\right\|_{l^2 \leftarrow l^2} = \|C^* C\|_{l^2 \leftarrow l^2}$. From*

$$\langle CCu, CCu \rangle_{l^2} = \langle C^* CCu, Cu \rangle_{l^2} = \langle CC^* Cu, Cu \rangle_{l^2} = \langle C^* Cu, C^* Cu \rangle_{l^2}$$

it follows that

$$\left\|C^2\right\|^2_{l^2 \leftarrow l^2} = \sup_{\|u\|_{l^2}=1} \langle CCu, CCu \rangle_{l^2} = \sup_{\|u\|_{l^2}=1} \langle C^* Cu, C^* Cu \rangle_{l^2} = \|C^* C\|^2_{l^2 \leftarrow l^2}.$$

Since also

$$\|C^* C\|_{l^2 \leftarrow l^2} = \sup_{\|u\|_{l^2} = \|v\|_{l^2} = 1} \langle u, C^* C v \rangle = \sup_{\|u\|_{l^2} = \|v\|_{l^2} = 1} \langle Cu, Cv \rangle_{l^2} \geqslant \sup_{\|u\|_{l^2}=1} \langle Cu, Cu \rangle_{l^2} = \|C\|^2_{l^2 \leftarrow l^2},$$

$\|C\|^2_{l^2 \leftarrow l^2} \geqslant \left\|C^2\right\|_{l^2 \leftarrow l^2}$ is shown. Using the property of the operator norm $\left\|C^2\right\|_{l^2 \leftarrow l^2} \leqslant \|C\|^2_{l^2 \leftarrow l^2}$, the equality $\left\|C^2\right\|_{l^2 \leftarrow l^2} = \|C\|^2_{l^2 \leftarrow l^2}$ is proved. Therefore, $\|C^n\|_{l^2 \leftarrow l^2} = \|C\|^n_{l^2 \leftarrow l^2}$ also follows for all $n = 2^k$, where $k \in \mathbb{N}$.
Since C is a bounded linear operator, we can use Gelfand's formula:

$$\rho(C(\lambda, \Delta t)) = \lim_{n \to \infty} \sqrt[n]{\|C^n\|_{l^2 \leftarrow l^2}}.$$

Applying the results of (1), the equality $\rho(C) = \|C\|_{l^2 \leftarrow l^2}$ follows. We can use that (5.4) is necessary, while $\|C\|_{l^2 \leftarrow l^2} \leqslant 1 + \mathcal{O}(\Delta t)$ is sufficient. Since we showed $\rho(C) = \|C\|_{l^2 \leftarrow l^2}$, both inequalities are identical. ■

One may ask whether a similar condition holds for more general operators. For this purpose, we introduce the *almost normal operators*: $C(\lambda, \Delta t)$ is almost normal if

$$\|C(\lambda, \Delta t)C^*(\lambda, \Delta t) - C^*(\lambda, \Delta t)C(\lambda, \Delta t)\|_{l^2 \leftarrow l^2} \leqslant M(\Delta t)^2 \|C(\lambda, \Delta t)\|^2_{l^2 \leftarrow l^2}$$

for some constant $M < \infty$.

Proposition 5.8. *If $C(\lambda, \Delta t)$ is almost normal, then l^2 stability is equivalent to the estimate*

$$\|C(\lambda, \Delta t)\|_{l^2 \leftarrow l^2} \leqslant 1 + \mathcal{O}(\Delta t).$$

Proof. *Since we stated that $\|C(\lambda, \Delta t)\|_{l^2 \leftarrow l^2} \leqslant 1 + \mathcal{O}(\Delta t)$ is sufficient, it is only needed to prove necessity. Again, $C(\lambda, \Delta t)$ is denoted by C. First we prove that the $(C^*)^\mu C^\mu$ can be reordered into $(C^*C)^\mu$ in at most $\mu^2/2$, where one step consists of $C^*C \mapsto CC^*$. In the case of $\mu = 1$, no permutation is required and $0 \leqslant 1^2/2$ proves the start of the induction approach. Suppose the induction hypothesis for μ. Then*

$$(C^*C)^{\mu+1} \hookrightarrow_{\mu^2 \backslash 2} (C^*)^\mu C^\mu C^* C \hookrightarrow_\mu (C^*)^{\mu+1} C^{\mu+1},$$

and $\frac{\mu^2}{2} + 1 \leqslant \frac{(\mu+1)^2}{2}$ prove the hypothesis. Next, we show that the interchange perturbs the operator norm at most by $M(\Delta) \|C\|^{2\mu}_{l^2 \leftarrow l^2}$. In the following formula C_j, $0 \leqslant j \leqslant 2\mu$ are either C or C^* and we focus on one interchange. Furthermore, we use that $\|C^*\|_{l^2 \leftarrow l^2} = \|C\|_{l^2 \leftarrow l^2}$. Then

$$\begin{aligned}
&\|C_1..C_v \underline{CC^*} C_{v+3}..C_{2\mu} - C_1..C_v \underline{C^*C} C_{v+3}..C_{2\mu}\|_{l^2 \leftarrow l^2} \leqslant \\
&\leqslant \|C_1..C_v\|_{l^2 \leftarrow l^2} \|CC^* - C^*C\|_{l^2 \leftarrow l^2} \|C_{v+3}..C_{2\mu}\|_{l^2 \leftarrow l^2} \leqslant \\
&\overset{def}{\leqslant} \|C\|^v_{l^2 \leftarrow l^2} \left[M(\Delta t)^2 \|C\|^2_{l^2 \leftarrow l^2} \right] \|C\|^{2\mu - v - 2}_{l^2 \leftarrow l^2} = M(\Delta t)^2 C^{2\mu}_{l^2 \leftarrow l^2}.
\end{aligned}$$

for all $0 \leqslant v \leqslant 2\mu - 2$. Using the aforementioned,

$$\begin{aligned}
\|C^\mu\|^2_{l^2 \leftarrow l^2} &= \sup_{\|u\|_{l^2} = 1} \langle u, (C^*)^\mu C^\mu u \rangle_{l^2} \geqslant \\
&\geqslant \sup_{\|u\|_{l^2} = 1} \langle u, (C^*C)^\mu u \rangle_{l^2} - \frac{M}{2} (\mu \Delta t)^2 \|C\|^{2\mu}_{l^2 \leftarrow l^2} \|u\|^2_{l^2} = \\
&= \left[1 - \frac{M}{2} (\mu \Delta t)^2 \right] \|C\|^{2\mu}_{l^2 \leftarrow l^2}.
\end{aligned}$$

Restricting μ and Δt by $\mu \Delta t \leqslant \min\{1 \backslash \sqrt{M}, T\}$, we obtain $1 - \frac{M}{2}(\mu \Delta t)^2 \geqslant \frac{1}{2}$. Then

$$\frac{1}{2} \|C\|^{2\mu}_{l^2 \leftarrow l^2} \leqslant \|C^\mu\|^2_{l^2 \leftarrow l^2} \overset{def}{\leqslant} K(\lambda)^2,$$

so that $\|C\| \leqslant \sqrt[2\mu]{2K(\lambda)^2}$ and from (5.3), inequality $\|C\|_{l^2 \leftarrow l^2} \leqslant 1 + \mathcal{O}(\Delta t)$ follows. ■

These inequalities are the most convenient conditions proving stability.

However, even if $\|C(\lambda, \Delta t)\|_{l^2 \leftarrow l^2} \geqslant c > 1$, it may happen that the powers still stay bounded for a constant K, $\|C(\lambda, \Delta t)^\mu\|_{l^2 \leftarrow l^2} \leqslant K$. In this case one may try to find an equivalent norm which behaves easier. Since the square $\|U\|_{l^2}^2 = \sum_{i \in \mathbb{Z}} |U_i|^2$ is a quadratic form, one way is to introduce another quadratic form $Q(U)$ such that

$$\frac{1}{K_1} \|U\|_{l^2}^2 \leqslant Q(U) \leqslant K_1 \|U\|_{l^2}^2$$

which describes the equivalence of the norms $\sqrt{\|\cdot\|_{l^2}}$ and $\sqrt{Q(\cdot)}$. We may put the question of whether the stability property holds for $Q(U)$. Let us suppose that the growth of $Q(\cdot)$ for one time step $U^{\mu+1} = C(\lambda, \Delta t) U^\mu$ is limited by

$$0 \leqslant Q(U^{\mu+1}) - Q(U^\mu) \leqslant K_2 \Delta t \left(\|U^\mu\|_{l^2}^2 + \left\|U^{\mu+1}\right\|_{l^2}^2 \right) + K_3 \Delta t,$$

then for $\Delta t \leqslant 1/(2K_1 K_2)$, the norms $Q(U^\mu)$ and $\|U^\mu\|_{l^2}$ stay bounded. More precisely

$$\|U^\mu\|_{l^2}^2 \leqslant \left\{ (2K_1^2 - 1) \left\|U^0\right\|_{l^2}^2 + \frac{K_3}{2K_2} \right\} \left(\frac{1 + K\Delta t}{1 - K\Delta t} \right)^\mu - \frac{K_3}{2K_2} \quad \text{for } \mu \in \mathbb{N}.$$

Chapter 6

Fourier analysis

To derive further conditions for stability, we consider f as a 2π-periodical function in $L(\mathbb{R})^2_{2\pi}$, meaning the 2π-periodic extension of $L^2(0, 2\pi)$ to $\mathbb{R}$. In general, we call a function *α-periodical* if

$$f(x + \alpha) = f(x) \quad \text{for any } x \in \mathbb{R},$$

where in our case $\alpha = 2\pi$. We call a series trigonometric if it has the form

$$\sum_{k=1}^{\infty} [a_k \cos(kx) + b_k \sin(kx)],$$

where the partial sum of the $2n + 1$ terms is called trigonometric polynomial of order n

$$T_n(x) := \frac{a_0}{2} + \sum_{k=1}^{n} [a_k \cos(kx) + b_k \sin(kx)].$$

Using chapter 5 of [2] and the notes of [4] to understand the periodical sign, we start by assuming that there exists a series of sine waves with wave numbers $k \in \mathbb{N}$, which converges to function f as its sum on $\mathbb{R}$. So that, the task is to find coefficients $b_1, \ldots b_n, \ldots$ for the following problem

$$f(x) = b_1 \sin(x) + \ldots + b_n \sin(nx) + \ldots \quad \text{for any } x \in \mathbb{R}.$$

Such a series, consisting only of sine waves and summing up to f, does not alway exist. So one may apply the linear combination of $\{\sin(kx),\ k = 1, 2, 3, \ldots\}$ and $\{\cos(kx),\ k = 0, 1, 2, \ldots\}$ to provide solvability for all α-periodic functions, at 'almost each' x. In the theory of the Fourier series, our aim is to approximate function f with *trigonometric polynomials*. It is possible to show that the trigonometric series defined above may converge uniformly to f only if the coefficients are defined as

$$a_k := \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(kx) dx \quad \text{for } k = 0, 1, 2, ..$$
$$b_k := \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(kx) dx \quad \text{for } k = 1, 2, 3, ..$$

and this is called the Fourier series of f. If a_k, b_k are chosen to be the above Fourier coefficients, then

$$\int_0^{2\pi} (f(x) - T_n(x))^2 dx$$

reaches its minimum. The next question is that the setting of such a series to f under which conditions provides the uniform convergence and generates f. In the case of a sufficiently smooth 2π-periodic function, the Fourier series converge uniformly with respect to $\|\cdot\|_\infty$ to f. For general L^2 functions, the convergence holds only in the sense of L^2-norm. There are various conditions for the uniform convergence. More details can be found in chapter 1 of [3] and chapter 6 of [6]. Furthermore, using the Euler formula and ordering $T_n(x)$ via $\exp(ix)$ with $i = \sqrt{-1}$, it can be rewritten as

$$T_n(x) = \sum_{k=-n}^{n} c_k \exp(-ikx),$$

with

$$c_k := \frac{1}{2\pi} \int_0^{2\pi} f(x) \exp(ikx) dx,$$

where $c_{-k} = \bar{c}_k$ for $k \in \{1, 2 \ldots\}$. We conclude that it is possible to find connections with the n-dimensional Euclidean space. For example, the Fourier coefficients can be specified as the components of a vector, based on an orthonormal basis. We suppose that

$$\begin{aligned} T_n(x) &:= \sum_{k=-n}^{n} c_k \exp(-ikx) \\ T'_n(x) &:= \sum_{k=-n}^{n} c'_k \exp(-ikx), \end{aligned}$$

then

$$\frac{1}{2\pi} \int_0^{2\pi} T_n(x) T'_n(x) dx = \sum_{k=-n}^{n} c_k c'_k.$$

Therefore, the multiple integration of two trigonometric polynomials is equal to the dot product of the representative coefficients. More generally, it is well-known that for a function $f \in L^2(0, \pi)$ Parseval's equality

$$\frac{1}{2\pi} \int_0^{2\pi} |f(x)|^2 dx = \sum_{k=-\infty}^{\infty} |c_k|^2$$

holds. In the following, the associated Fourier series is

$$\frac{1}{\sqrt{2\pi}} \sum_{\alpha \in \mathbb{Z}} \varphi_\alpha \exp(i\alpha x) \quad \text{with} \quad \varphi_\alpha := \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(x) \exp(-i\alpha x) dx.$$

6.1 Transform operator $\mathcal{F}$

According to the Parseval equality, with the notation introduced before:

$$\|f\|_{L^2(0,2\pi)} = \|\varphi\|_{l^2}\,. \tag{6.1}$$

On the one hand, the transfer from $f \in L^2(0,2\pi)$ to its Fourier coefficients $\varphi \in l^2$ is the *Fourier analysis*, and the operator of which will be denoted by F:

$$\mathcal{F} : f \mapsto \varphi.$$

On the other hand, it is also possible to define the inverse mapping, which is called the *Fourier synthesis*, denoted by $\mathcal{F}^{-1}$. In the following, we rely on chapter 5 of [6].

For $p = 2$, we showed that the solutions U^μ of the difference scheme were l^2 sequences, so that we can apply the Fourier synthesis and denote the associated 2π-periodic function by $\hat{U}^\mu \in L^2(0,2\pi)$ as

$$\hat{U}^\mu := \mathcal{F}^{-1}U^\mu \quad \text{where} \quad \hat{U}^\mu(\xi) = \frac{1}{\sqrt{2\pi}}\sum_{\alpha\in\mathbb{Z}} U^\mu_\alpha \exp(i\alpha\xi) \quad \text{for } \xi \in \mathbb{R}$$

with the coefficients $(U^\mu_\alpha)_{\alpha\in\mathbb{Z}}$. Therefore, it is possible to derive further stability conditions in terms of the obtained for $L^2(0,2\pi) \to L^2(0,2\pi)$ operators. First, one may put the question of how the difference operator $C(\lambda, \Delta t) = C$ should be transformed for $\hat{U}^\mu$. We use the difference scheme $U^{\mu+1} = CU^\mu$, then

$$\hat{U}^{\mu+1} = \mathcal{F}^{-1}U^{\mu+1} = \mathcal{F}^{-1}CU^\mu = \mathcal{F}^{-1}C\mathcal{F}\mathcal{F}^{-1}U^\mu = \hat{C}\hat{U}^\mu$$

with $\hat{C} := \mathcal{F}^{-1}C\mathcal{F}$.

Using isometry (6.1) and the definitions of operators $\mathcal{F}$ and $\mathcal{F}^{-1}$ one can prove three important assertions:

1. $\|\mathcal{F}\|_{l^2\leftarrow L^2(0,2\pi)} = \left\|\mathcal{F}^{-1}\right\|_{L^2(0,2\pi)\leftarrow l^2} = 1$, i.e. $\mathcal{F}$ and $\mathcal{F}^{-1}$ are unitary,
2. for all functions $M \in \mathcal{L}(l^2, L^2(0,2\pi))$, the $\|M\|_{L^2(0,2\pi)\leftarrow l^2} = \|M\mathcal{F}\|_{L^2(0,2\pi)\leftarrow L^2(0,2\pi)}$ holds,
3. for all functions $D \in \mathcal{L}(l^2, l^2)$, the $\|D\|_{l^2\leftarrow l^2} = \left\|\mathcal{F}^{-1}D\right\|_{L^2(0,2\pi)\leftarrow l^2}$ holds.

To verify these three points, we need to restrict f as a fixed function on $L^2(0,2\pi)$. Furthermore, let the Fourier coefficients of f be denoted by φ. Then,

$$\left\|\mathcal{F}^{-1}\varphi\right\|_{L^2(0,2\pi)} = \|f\|_{L^2(0,2\pi)} = \|\varphi\|_{l^2} = \|\mathcal{F}f\|_{l^2}$$

yields the first point, as $\mathcal{F}$ and $\mathcal{F}^{-1}$ do not change the norm values. Since $\mathcal{F}f = \varphi$, the norms of $M\mathcal{F}f$ and $M\varphi$ are equal and using (6.1), the equality of the norms of $M\mathcal{F}$ and M also follows. For the third point, the following equalities hold:

$$\left\|\mathcal{F}^{-1}D\varphi\right\|_{L^2(0,2\pi)} = \left\|\mathcal{F}^{-1}\gamma\right\|_{L^2(0,2\pi)} = \|g\|_{L^2(0,2\pi)} = \|\gamma\|_{l^2} = \|D\varphi\|_{l^2}$$

where we use the fact that operator D maps φ to another sequence, denoted by γ. In this manner, the norms of $\mathcal{F}^{-1}D\varphi$ and $\mathcal{F}^{-1}\gamma$ are equal. Furthermore, there exists a function g, for which γ is the set of its Fourier coefficients. The two extremes of the equalities imply the third point.

As an application of the aforesaid three points, we can state that the norms of C^μ and $\hat{C}^\mu$ are equivalent since

$$\begin{aligned} \|C^\mu\|_{l^2 \leftarrow l^2} &\overset{3.}{=} \left\|\mathcal{F}^{-1}C^\mu\right\|_{L^2(0,2\pi)\leftarrow l^2} \overset{2.}{=} \left\|\mathcal{F}^{-1}C^\mu\mathcal{F}\right\|_{L^2(0,2\pi)\leftarrow L^2(0,2\pi)} \overset{1.}{=} \\ &= \left\|[\mathcal{F}^{-1}C\mathcal{F}]^\mu\right\|_{L^2(0,2\pi)\leftarrow L^2(0,2\pi)} = \left\|\hat{C}^\mu\right\|_{L^2(0,2\pi)\leftarrow L^2(0,2\pi)}. \end{aligned} \tag{6.2}$$

Therefore, an equivalent definition of l^2 stability with the stability constant $K(\lambda)$ also holds for $\hat{C}(\lambda, \Delta t)$ as

$$\left\|\hat{C}^\mu\right\|_{L^2(0,2\pi)\leftarrow L^2(0,2\pi)} \leqslant K(\lambda) \qquad \text{for } \forall \mu \in \mathbb{N}_0, \Delta t > 0 \text{ with } \mu\Delta t \leqslant T. \tag{6.3}$$

For a concrete determination of $\hat{C}$, using (4.8) and (4.9), we consider a term $C_j = a_j E_j$ from $C = \sum_j C_j$ as $(C_j U)_v := a_j U_{v+j}$ for $U \in l^2$. The *Fourier synthesis* of $C_j U = (a_j U_{\alpha+j})_{\alpha\in\mathbb{Z}}$ yields

$$\begin{aligned} \mathcal{F}^{-1}C_jU &\overset{\text{def.}}{=} \frac{1}{\sqrt{2\pi}}\sum_{\alpha\in\mathbb{Z}}(a_jU_{\alpha+j})\exp(i\alpha\xi) = \frac{a_j}{\sqrt{2\pi}}\sum_{\alpha\in\mathbb{Z}}U_{\alpha+j}\exp(i\alpha\xi) \underset{\beta=\alpha+j}{=} \frac{a_j}{\sqrt{2\pi}}\sum_{\beta\in\mathbb{Z}}U_\beta\exp(i(\beta-j)\xi) = \\ &= \frac{a_j\exp(-ij\xi)}{\sqrt{2\pi}}\sum_{\beta\in\mathbb{Z}}U_\beta\exp(i\beta\xi) = a_j\exp(-ij\xi)\hat{U}. \end{aligned}$$

Since $\hat{C}_j\hat{U} = \mathcal{F}^{-1}C_jU$ holds with $\hat{U} = \mathcal{F}^{-1}U$, $\hat{C}_j$ can be defined as $a_j\exp(-ij\xi)$, i.e. the linear mapping $\hat{C}_j$ is the multiplication by the function $a_j\exp(-ij\xi)$. Since the equality $\hat{C} = \sum_{j\in\mathbb{Z}}\hat{C}_j$ also holds, we can obtain that the Fourier transformed operator is $\hat{C} = \sum_{j\in\mathbb{Z}} a_j\exp(-ij\xi)$. Other words, the application of $\hat{C}$ corresponds to the multiplication by the trigonometric polynomial

$$G(\xi) := \sum_{j\in\mathbb{Z}} a_j\exp(-ij\xi) \tag{6.4}$$

which is called the *characteristic function* of C. We need to note that parameters λ and Δt are omitted in $G(\xi)$, however a_j depends on λ and Δt. We state that the identity (6.2) becomes

$$\|C^\mu\|_{l^2\leftarrow l^2} = \sup\{|G(\xi)|^\mu : \xi \in \mathbb{R}\},$$

and it yields

$$\begin{aligned} \left\|\hat{C}^\mu\right\|_{L^2(0,2\pi)\leftarrow L^2(0,2\pi)} &= \left\|\left(\sum\nolimits_{j\in\mathbb{Z}} a_j\exp(-ij\xi)\right)^\mu\right\|_\infty = \sup\{|G(\xi)|^\mu : \xi\in\mathbb{R}\} = \\ &= \sup\{|G(\xi)| : \xi\in\mathbb{R}\}^\mu = \left\|\sum\nolimits_{j\in\mathbb{Z}} a_j\exp(-ij\xi)\right\|_\infty^\mu = \left\|\hat{C}\right\|^\mu_{L^2(0,2\pi)\leftarrow L^2(0,2\pi)}. \end{aligned}$$

The latter equality is valid only for scalar a_j. Together with (6.3) it leads to the fact that the boundedness of the absolute value of the characteristic function is a necessary and sufficient condition for the stability of the difference scheme (4.8).

Theorem 6.1. *The difference scheme (4.8) formed by $C(\lambda, \Delta t)$ is l^2 stable if and only if the characteristic function $G(\xi) = \sum_{j\in\mathbb{Z}} a_j \exp(-ij\xi)$ satisfies the following estimate with a suitable K_λ:*

$$|G(\xi)| \leqslant 1 + K_\lambda \Delta t \quad \textit{for all } \xi \in \mathbb{R}.$$

Proof. *It is clear that the inequality is sufficient for the stability of C, since*

$$\|C\|_{l^2\leftarrow l^2} = \left\|\hat{C}\right\|_{L^2(0,2\pi)\leftarrow L^2(0,2\pi)} = \|G\|_\infty \leqslant 1 + K_\lambda \Delta t.$$

So that, we only need to prove the necessity. Similarly to (5.3), set $c(\Delta t) := (G(\xi) - 1)/\Delta t$. If there is no constant K_λ satisfying $|G(\xi)| \leqslant 1 + K_\lambda \Delta t$, then $c(\Delta t) \to \infty$ follows for $\Delta t \to 0$. ■

Using Theorem 6.1, we are also able to prove Proposition 5.5 for l^2 instability of C (cf. Chapter 5). Since inequality $\sum_{j\in\mathbb{Z}} a_j \geqslant 1 + \Delta t c(\Delta t)$ with $\lim_{\Delta t\to\infty} c(\Delta t) = \infty$ is supposed and $\|G\|_\infty \geqslant |G(0)| = |\sum a_j|$ holds, therefore $|G(\xi)|$ is unbounded and yields the l^2 instability. This analysis refers to l^2, however the characteristic function has also consequences for l^∞ stability. The negation of this statement is: if the difference scheme is l^2 unstable, it is also l^∞ unstable.

Lemma 6.2. *Let C be of the form (4.8) with constant coefficients a_j. Then l^∞ stability implies l^2 stability:*

$$\|C^\mu\|_{l^\infty\leftarrow l^\infty} \geqslant \|G^\mu\|_\infty = \|C\mu\|_{l^2\leftarrow l^2} \qquad \textit{for all } \mu \in \mathbb{N}_0.$$

Proof. *Choose the special initial value U^0 with $U^0_v = \exp(iv\xi)$, where $\xi \in \mathbb{R}$ is characterized by the equality $|G(\xi)| = \|G\|_\infty$. Note that $U^0 \in l^\infty$ so $\|U^0\|_{l^\infty} = \|U\|^0_{l^\infty} = 1$. The application of C yields $U^1 = CU^0$ with*

$$U^1_v = \sum_{j\in\mathbb{Z}} a_j U^0_{v+j} = \sum_{j\in\mathbb{Z}} a_j \exp(i(v+j)\xi) = \exp(iv\xi) \sum_{j\in\mathbb{Z}} a_j \exp(ij\xi) = U^0 G(\xi),$$

hence $C^\mu U^0 = G^\mu(\xi) U^0$. Then,

$$\|U^0\|_{l^\infty} \|C^\mu\|_{l^\infty\leftarrow l^\infty} \geqslant \|C^\mu U^0\|_{l^\infty} = \|G^\mu(\xi)U^0\|_\infty = |G^\mu(\xi)| \|U^0\|_{l^\infty} = \|G^\mu\|_\infty \|U^0\|_{l^\infty}$$

and the assertion follows. ■

A possible application of Theorem 6.1 and Lemma 6.2 is the following. Because of

$$G(\xi) := \lambda \exp(i\xi) + 1 - 2\lambda + \lambda \exp(-i\xi) = 1 - 2\lambda(1 - cos(\xi))$$

and $|G(\xi)|$ reaches its maximum at π, $\|G(\xi)\|_\infty = |G(\pi)| = |1 - 4\lambda|$. So the difference scheme (4.10) is conditionally stable for $\lambda \in (0, 1/2]$ but unstable for $\lambda > 1/2$ (with respect to l^2 and l^∞).

Chapter 7

Implicit schemes

All the schemes using the explicit form $U^{\mu+1} = C(\lambda, \Delta t)U^{\mu}$ examined so far are conditionally stable. In order to obtain *unconditionally stable* schemes, one must admit implicit difference schemes. Now, in the case of $A = \frac{\partial^2}{\partial x^2}$, let the difference quotient at time level $t + \Delta t$ be approximated as

$$\frac{u(t + \Delta t, x) - u(t, x)}{\Delta t} \quad \text{for} \quad \frac{\partial u}{\partial t}$$

while the second difference quotient is

$$\frac{u(t + \Delta t, x - \Delta x) - 2u(t + \Delta t, x) + u(t + \Delta t, x + \Delta x)}{\Delta x^2} \quad \text{for} \quad \frac{\partial^2 u}{\partial x^2}$$

which lead together with $\lambda = \frac{\Delta t}{\Delta x^2}$ to

$$-\lambda U_{v-1}^{\mu+1} + (1 + 2\lambda)U_{v}^{\mu+1} - \lambda U_{v+1}^{\mu+1} = U_{v}^{\mu}. \tag{7.1}$$

So that, one may obtain an implicit scheme of the form

$$C_1(\lambda, \Delta t)U^{\mu+1} = C_2(\lambda, \Delta t)U^{\mu} \tag{7.2}$$

where in the present case the operators $C_1(\lambda, \Delta t) = C_1$ and $C_2(\lambda, \Delta t) = C_2$ are given by

$$(C_1 U)_v = \sum_{-1 \leqslant j \leqslant 1} a_{1,j} U_{v+j} \quad \text{with} \;\; a_{1,-1} = a_{1,1} = -\lambda, \;\; a_{1,0} = 1 + 2\lambda$$
$$(C_2 U)_v = \sum_{j \in \mathbb{Z}} a_{2,j} U_{v+j} \quad \text{with} \;\; a_{2,0} = 1,$$

and all other coefficients are equal to 0, i.e C_2 is the identity operator. Since one would like to solve (7.2) with respect to $U^{\mu+1}$, therefore the existence of the inverse C_1^{-1} is to be investigated.

Lemma 7.1. *(a) If the coefficients $a_{1,j}$ of C_1 satisfy the inequality*

$$\left(\sum_{j\in\mathbb{Z}\setminus\{0\}} |a_{1,j}|\right)/|a_{1,0}| \leqslant 1-\varepsilon < 1 \quad \text{for some } \varepsilon > 0$$

then the inverse exists and it satisfies $\left\|C_1^{-1}\right\|_{l^p\leftarrow l^p} \leqslant \frac{1}{\varepsilon|a_{1,0}|}$.
(b) The inverse C_1^{-1} exists in $\mathcal{L}(l^2,l^2)$ iff the characteristic function $G_1(\xi)$ of C_1 satisfies an inequality $|G_1| \geqslant \eta > 0$. The norm equals

$$\left\|C_1^{-1}\right\|_{l^2\leftarrow l^2} = \left(\inf_{\xi\in\mathbb{R}} |G_1(\xi)|\right)^{-1}.$$

<u>**Proof.**</u> *Case (a). The equation $C_1V = U$ is equivalent to the fixed-point equation $Q(V) = V$ with*

$$Q(V) := \frac{1}{a_{1,0}}(U - C_1'V) \quad \text{with} \quad C_1' = C_1 - a_{1,0}I.$$

The contraction constant of $Q(V) = V$ for $p \in \{2,\infty\}$ is the following:

$$\|Q(V_1) - Q(V_2)\|_{l^p} \geqslant \|1/a_{1,0}(U - C_1'V_1) - 1/a_{1,0}(U - C_1'V_2)\|_{l^p} = \frac{\|C_1'\|_{l^p}}{|a_{1,0}|} \|V_1 - V_2\|_{l^p}.$$

Using $\|C\|_{l^p\leftarrow l^p} \leqslant \sum_j |a_j|$ (cf. Chapter 5):

$$\frac{\|C_1\|_{l^p\leftarrow l^p}}{|a_{1,0}|} \leqslant \frac{\sum_{j\in\mathbb{Z}\setminus\{0\}} |a_{1,j}|}{|a_{1,0}|} \leqslant 1-\varepsilon$$

is valid, so that a unique inverse exists. Furthermore, the estimate

$$\|V\|_{l^p} \leqslant \frac{1}{|a_{1,0}|}\|U\|_{l^p} + (1-\varepsilon)\|V\|_{l^p}$$

shows that $\varepsilon\|V\|_{l^p} \leqslant \frac{1}{|a_{1,0}|}\|U\|_{l^p}$. so that $\left\|C^{-1}\right\|_{l^p\leftarrow l^p}$ has the bound $1/(\varepsilon|a_{1,0}|)$.
Case (b). We suppose that $\hat{C}_1 = \mathcal{F}^{-1}C_1\mathcal{F}$ is the Fourier transformation of C_1, and $\hat{C}_1$ is the operator $\hat{U} \mapsto G_1(\xi)\hat{U}$. Obviously, the multiplication operator $M \in \mathcal{L}(l^2,l^2)$ with $M\hat{U} := (1/G_1(\xi))\hat{U}$ is the inverse of $\hat{C}_1$. The norm of M is:

$$\|M\|_{l^2\leftarrow l^2} = \left\|\hat{C}_1^{-1}\right\|_{l^2\leftarrow l^2} = \|1/G_1\|_\infty.$$

One may verify that the norm of $\hat{C}_1^{-1}$ can not be finite if $\inf\{|G_1(\xi)|\}$ *equals zero. So that, there exists $\eta > 0$ for which $|G_1(\xi)| \geqslant \eta$ for any $\xi \in \mathbb{R}$. Moreover, since the Fourier transform does not change the l^2 norm, the equality $\left\|C_1^{-1}\right\|_{l^2\leftarrow l^2} = \left\|\hat{C}_1^{-1}\right\|_{l^2\leftarrow l^2} = 1/\inf_{\xi in\mathbb{R}}\{|G_1(\xi)|\}$ follows.* ■

Using the latter lemma, it is easy to show that the scheme (7.1) is l^2-stable for all $\lambda = \Delta t/\Delta x^2$, i.e. unconditionally stable. Since the characteristic function of C_1 is $G_1(\xi) := -\lambda\exp(i\xi) + 1 + 2\lambda - \lambda\exp(-i\xi) = 1 + 2\lambda(1 - \cos(x)) \geqslant \inf|G_1(\xi)| = 1$ and $C_2 = I$. we obtain $C = C_1^{-1}C_2 = C_1^{-1}$. Hence, stability follows by $\left\|C_1^{-1}\right\|_{l^2\leftarrow l^2} = 1 \leqslant 1 + K_\lambda \Delta t$ for a suitable K_λ. For a concrete determination, we define C_1^{-1} as an infinite operator

$$C_1^{-1} := \sum_{j\in\mathbb{Z}} a_j(\lambda)E_j \quad \text{with} \quad a_j := \frac{1}{\sqrt{1+4\lambda}}\left(\frac{2\lambda}{2\lambda+1+\sqrt{1+4\lambda}}\right)^{|j|}$$

i.e the implicit scheme is identical to $U_v^{\mu+1} = \sum_{j\in\mathbb{Z}} a_j U_{v+j}^{\mu}$. In general, it is also possible to assume that the scheme (7.2) is l^2 stable if and only if the characteristic function

$$G(\xi) := G_2(\xi)/G_1(\xi)$$

satisfies Theorem 6.1. A possible application of the latter statement is represented in the following. First, we introduce the *theta scheme*, a modification of the schemes used so far as

$$-\lambda\theta U_{v-1}^{\mu+1} + (1+2\lambda\theta)U_v^{\mu+1} - \lambda\theta U_{v+1}^{\mu+1} = \lambda(1-\theta)U_{v-1}^{\mu} + (1-2\lambda(1-\theta))U_v^{\mu} + \lambda(1-\theta)U_{v+1}^{\mu}. \tag{7.3}$$

It is easy to see that (7.3) is the presented *explicit* one for $\theta = 0$ and the *implicit* one for $\theta = 1$. The above defined scheme satisfies the following lemma:

Lemma 7.2. *The scheme (7.3) is unconditionally l^2 stable for $\theta \in [1/2, 1]$, whereas in the case of $\theta \in [0, 1/2)$ it is conditionally stable for*

$$\lambda \leqslant 1/(2(1-2\theta)).$$

In all stable cases, $\|C\|_{l^2\leftarrow l^2} = 1$ holds.

Proof. *The idea is to define a second difference operator D with the scheme $(DU)_v = -U_{v-1} + 2U_v - U_{v+1}$. The characteristic function of D using Euler's formula is $G_D(\xi) = 2 - 2\cos(\xi) = 4\sin^2(\xi/2)$. In the case of scheme (7.3), the operators C_1 and C_2 change to*

$$C_1 = I + \lambda\theta D \quad \textit{with} \quad G_1(\xi) = 1 + \lambda\theta G_D(\xi)$$
$$C_2 = I - \lambda(1-\theta)D \quad \textit{with} \quad G_2(\xi) = 1 - \lambda(1-\theta)G_D(\xi),$$

The characteristic function $G(\xi) := G_2(\xi)/G_1(\xi)$ is monotonically decreasing with respect to $G_D(\xi)$, so that the maximum of $|G(\xi)|$ is taken at $G_D(0) = 0$ or $G_D(\pi) = 4$, thus $\|G\|_\infty = \max\{G(0), -G(\pi)\}$. On the one hand, if $\theta \in [1/2, 1]$, then $-G(\pi) \leqslant 1$ for any λ and $\|G(\xi)\|_\infty = 1$ provides stability. On the other hand, if $\theta \in [0, 1/2)$, then the choice of $\lambda \leqslant 1/(2(1-2\theta))$ ensures the estimate $-G(\pi) \leqslant 1$ and provides conditional stability. ■

Chapter 8

Conclusion

The starting point of our considerations was to rewrite the heat conduction problem to an abstract Cauchy problem with differential operator A on two Banach spaces of functions in spatial variable x with respect to the associated norms. We mapped the initial function to the solution $u(t,x)$ through a uniformly bounded and continuous solution operator $S(t)$. The solution operators for all $t \geqslant 0$ formed an operator semigroup with generator operator A. For the numerical approximations, we defined a rectangular grid on the space-time domain of the problem. To map the continuous Banach space onto the linear space of grid functions, we used transfer operators r and p. After that, it was possible to introduce an explicit difference scheme with linear difference operator C. On a numerical example we observed that the scheme provided a good solution only for certain values of $\lambda = \Delta t/\Delta x^2$. We sought the answer to the question why this instability occurred and what conditions of stability could be applied. Our investigation was based on Lax's equivalence theorem, which states that if the difference scheme is consistent with the PDE, then stability is necessary and sufficient for providing convergence. Then, we saw via the perturbation lemma that a stable scheme remained stable after a perturbation when a norm of the perturbation operator was bounded. Using the lemma, we derived that it was sufficient to investigate operator A without terms of order zero. Introducing the spectral radius made it possible to derive a necessary condition for normal and a sufficient one for almost normal difference schemes. We introduced the Fourier analysis, synthesis and Fourier transform operator $\mathcal{F}$ to prove further conditions for $L^2(0,2\pi) \to L^2(0,2\pi)$ operators. One could show that the application of the transformed difference operator $\hat{C}$ corresponded to multiplication by the characteristic function of C. Using this function, it was possible to prove further sufficient and necessary criteria for stability. To obtain unconditionally stable schemes, we admitted the implicit difference schemes with their stability conditions. At the end, in order to close our arguments, we considered the general theta scheme, a modification of the used schemes, and showed that it was conditionally or unconditionally stable under some assumptions.

List of Figures

Appendix A

The source code of the programming task

```
function [x, u] = prog(dt,l)
%GENERAL INFORMATIONS
xmax = 5; % Maximum length
xo = -5;
tmax = 1; % Maximum time
to = 0;

dx = sqrt(dt/l);
x = (xo-(dx*tmax)/dt):dx:(xmax+(dx*tmax)/dt);
t = to:dt:tmax;

%INITIAL VALUE:
u(:,1) = ones(1,size(x,2));
u(round((end-1)/2),1)=1+(1e-5);


%IMPLEMENTATION:
for j = 1:size(t,2)-1
    for i = (j+1):size(u(:,j),1)-j
        u(i,j+1) = l*u(i-1,j)+(1-2*l)*u(i,j)+l*u(i+1,j);
    end
end

end
```

Appendix B

The main file

```
clear all
close all

tic
[x1,u1] = prog(.01,0.45);
[x2,u2] = prog(.01,0.7);
[x3,u3] = prog(.001,0.7);
toc

subplot(3,1,1);
plot(x1,u1(:,1),x1,u1(:,20),x1,u1(:,50),x1,u1(:,70),x1,u1(:,100))
xlabel('x','FontSize',12);
ylabel('u(x,t)','FontSize',12);
title('Test with dt=.01, lambda=0.45 ','FontSize',12);
legend('ts = 1', 'ts = 20', 'ts = 50', 'ts = 70', 'ts = 100')
axis([-5 5 0 2])

subplot(3,1,2);
plot(x2,u2(:,1),x2,u2(:,90),x2,u2(:,95),x2,u2(:,97),x2,u2(:,100))
xlabel('x','FontSize',12);
ylabel('u(x,t)','FontSize',12);
title('Test with dt=.01, lambda=0.7','FontSize',12);
legend('ts = 1', 'ts = 90', 'ts = 95', 'ts = 97', 'ts = 100')
axis([-5 5 -2e19 2e19])


subplot(3,1,3);
plot(x3,u3(:,1),x3,u3(:,990),x3,u3(:,995),x3,u3(:,997),x3,u3(:,1000))
xlabel('x','FontSize',12);
ylabel('u(x,t)','FontSize',12);
title('Test with dt=.001, lambda=0.7','FontSize',12);
legend('ts = 1', 'ts = 990', 'ts = 995', 'ts = 997', 'ts = 1000')
axis([-5 5 -2e248 2e248])
```

Bibliography

[1] Wolfgang Hackbusch, *The Concept of Stability in Numerical Mathematics.* ISBN: 978-3-642-39386-0, 2014

[2] Csatár Györgyné, Máté László, *SOROK, FÜGGVÉNYSOROK.* ID: 051365, Műegyetemi kiadó 2005.

[3] Sikolya Eszter, *Matematika Bsc Tanárszak Analízis III. előásjegyzet.* Department of Applied Analysis and Computational Mathematics at Eötvös Loránd University 11. oct. 2011 from http://www.cs.elte.hu/~seszter/oktatas/2010_11_1/BSc_mattanar_ea/analizis_III_jegyzet2010.pdf

[4] Bátkai András, *Fourier sorok.* Department of Applied Analysis and Computational Mathematics at Eötvös Loránd University from http://www.cs.elte.hu/~batka/oktatas/fouriersor.pdf

[5] Karátson János, *Numerikus funkcionálanalízis.* Department of Applied Analysis and Computational Mathematics at Eötvös Loránd University Budapest, 2010

[6] John K. Hunter, *Notes on Partial Differential Equations.* Department of Mathematics, University of California at Davis from https://www.math.ucdavis.edu/~hunter/pdes/pdes.html

[7] Kendall Atkinson, Weimin Han, *Theoretical Numerical Analysis.* Texts in Applied Mathematics 39, DOI: 10.1007/978-1-4419-0458-4-2 2009

[8] Pascal Frey, *The numerical simulation of complex PDE problems.* Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas 2008

[9] Gordon C. Everstine, *Numerical solution of Partial Differential Equations.* Department of Mechanical and Aerospace Engineering, The George Washington University 21. January 2010

[10] Charles Walkden, *Ergodic Theory lecture notes.* MATH4/61112 School of Mathematics, The University of Manchester 2015 from http://www.maths.manchester.ac.uk/~cwalkden/ergodic-theory/ergodic-theory.html

[11] Thomas G. Kurtz, *A GENERAL THEOREM ON THE CONVERGENCE OF OPERATOR SEMI-GROUPS.* TRANSACTIONS OF THE AMERICAN MATHEMATICAL SOCIETY Volume 148, March 1970

[12] Dave Gilliam, *Semigroups of Operators in Banach Space.* Department of Mathematics Texas Tech University, Lubbock

[13] P.D. Lax, R.D Richtmyer, *Survey of the stability of linear finite difference equations.* Comm. Pure Appl. Math. 9, 267-293 (1965)