

14. PARABOLISCHE DIFFERENTIALGLEICHUNGEN

14.1 Schwache Formulierung parabolischer Probleme

Als Modellproblem eines parabolischen Anfangs-Randwertproblems betrachten wir die Wärmeleitungsgleichung in einem beschränkten Gebiet $\Omega \subset \mathbb{R}^d$ mit stückweise stetig differenzierbarem Rand $\Gamma := \partial\Omega$

$$\begin{aligned} \frac{\partial u}{\partial t} &= \Delta u + f(x, t), & x \in \Omega, \quad t \in [0, T] \\ u(x, t) &= 0, & x \in \Gamma & \text{Randbedingung} \\ u(x, 0) &= u_0 & x \in \Omega & \text{Anfangsbedingung} \end{aligned}$$

Im Allgemeinen hat dieses Problem keine *klassische* Lösung u , die zweimal stetig differenzierbar nach der Ortsvariablen x und einmal stetig differenzierbar nach der Zeit t ist. Daher müssen wir uns wie bei den elliptischen Problemen mit *schwachen* Lösungen begnügen.

Zur Definition einer schwachen Lösung multiplizieren wir die partielle Differentialgleichung mit einer Testfunktion $v \in H_0^1(\Omega)$ und integrieren über das Gebiet Ω . Mit Hilfe der Green'schen Formel erhalten wir für $u, v \in V = H_0^1(\Omega)$

$$-\int_{\Omega} v \Delta u dx = \int_{\Omega} \sum_{j=1}^d \frac{\partial u}{\partial x_j} \frac{\partial v}{\partial x_j} dx, \quad (14.1)$$

denn das Randintegral verschwindet.

Wir formulieren die folgenden Resultate für einen beliebigen Hilbert-Raum V mit Skalarprodukt $(\cdot, \cdot)_V$ und denken uns etwa $V = H_0^1(\Omega)$. Zudem sei eine V -elliptische, symmetrische Bilinearform $a : V \times V \rightarrow \mathbb{R}$ gegeben, d. h.

$$|a(v, w)| \leq M \|v\|_V \|w\|_V \quad a \text{ stetig (beschränkt)}, \quad (14.2)$$

$$a(v, v) \geq \alpha \|v\|_V^2, \quad \alpha > 0 \quad a \text{ koerziv}, \quad (14.3)$$

$$a(v, w) = a(w, v) \quad a \text{ symmetrisch}. \quad (14.4)$$

Ein Beispiel wäre die oben definierte rechte Seite von (14.1). Für festes $u \in V$ betrachten wir die durch a induzierte Linearform

$$\begin{aligned} a(u, \cdot) : V &\rightarrow \mathbb{R} \\ v &\mapsto a(u, v). \end{aligned}$$

Wegen (14.2) ist $a(u, \cdot)$ stetig und liegt damit im Dualraum V' . Wir können also einen linearen und stetigen Operator $A : V \rightarrow V'$ durch

$$\begin{aligned} A : V &\rightarrow V' \\ u &\mapsto a(u, \cdot) \end{aligned}$$

definieren. Wir schreiben im Folgenden für $\varphi \in V'$ und $v \in V$

$$\langle \varphi, v \rangle := \varphi(v).$$

Es gilt dann

$$\langle Au, v \rangle = a(u, v) \quad \text{für alle } u, v \in V. \quad (14.5)$$

Lemma 14.1. *$A : V \rightarrow V'$ ist linear, stetig und bijektiv und es gilt*

$$\|Au\|_{V'} = \sup_{\|v\|_V=1} |a(u, v)| \leq M\|u\|_V.$$

(V' ist ein Hilbert-Raum mit der Norm $\|\varphi\|_{V'} = \sup_{\|v\|_V=1} |\langle \varphi, v \rangle|$.)

Beweis. Es ist nur noch die Bijektivität zu zeigen. Für $\ell \in V'$ ist $Au = \ell$ genau dann, wenn $a(u, v) = \ell(v)$ für alle $v \in V$ gilt. Nach dem Satz von Lax-Milgram hat letztere Gleichung eine eindeutige Lösung in V , also ist A bijektiv. \square

Wir wählen nun einen zweiten Hilbert-Raum H mit der Norm $\|\cdot\|_H$ und dem Skalarprodukt $(\cdot, \cdot)_H$ aus, und zwar so, dass $V \subset H$ dicht liegt und die Einbettung $V \hookrightarrow H$ stetig ist, d. h. es eine Konstante $C > 0$ gibt, mit

$$\|v\|_H \leq C\|v\|_V \quad \text{für alle } v \in V.$$

In unserem Modellproblem ist $H = L^2(\Omega)$ und $V = H_0^1(\Omega)$ und die Bedingung trivialerweise erfüllt.

Erinnern wir uns an einen weiteren wichtigen Satz aus der Funktionalanalysis.

Satz 14.2. (*Riesz'scher Darstellungssatz*)

Zu jeder stetigen Linearform $\varphi \in H'$ gibt es ein eindeutig bestimmtes Element $v \in H$, so dass

$$\varphi(u) = (u, v)_H \quad \text{für alle } u \in H$$

gilt. Ferner ist $\|v\|_H = \|\varphi\|_{H'}$. \square

Bemerkungen.

- (a) Der Riesz'sche Darstellungssatz liefert eine eindeutige Zuordnung (den Riesz-Isomorphismus)

$$\begin{aligned} j_H : H &\rightarrow H' \\ v &\mapsto \varphi(\cdot) = (v, \cdot)_H \end{aligned}$$

mit $\|j_H\|_{H' \leftarrow H} = \|j_H^{-1}\|_{H \leftarrow H'} = 1$. H ist also normisomorph zu seinem Dualraum H' und kann mit ihm identifiziert werden, d. h.

$$H = H' \quad \text{via} \quad f \stackrel{\text{id}}{=} (f, \cdot)_H.$$

- (b) H' ist ein Hilbert-Raum mit dem Skalarprodukt $(x', y')_{H'} = (j_H^{-1}x', j_H^{-1}y')_H$.

Lemma 14.3. *$V \subset H$ seien Hilbert-Räume mit stetiger und dichter Einbettung. Dann ist H' stetig und dicht in V' eingebettet.*

Beweis. Es sei $\varphi \in H'$. Wegen $V \subset H$ ist φ auf V definiert. Da V ein dichter Teilraum von H ist, gilt

$$\|\varphi\|_{H'} = \sup_{v \in H} \frac{|\varphi(v)|}{\|v\|_H} = \sup_{v \in V} \frac{|\varphi(v)|}{\|v\|_H} \geq \frac{1}{C} \sup_{v \in V} \frac{|\varphi(v)|}{\|v\|_V} = \frac{1}{C} \|\varphi\|_{V'},$$

d. h. H' ist stetig in V' eingebettet.

Um zu zeigen, dass die Einbettung dicht ist, zeigen wir, dass das orthogonale Komplement von H' in V' nur die Null enthält. Die Behauptung folgt dann aus $V' = \overline{H'} \oplus (H')^\perp$.

Es sei nun $\psi \in (H')^\perp$ beliebig gewählt,

$$(H')^\perp = \{\psi \in V' \mid (\psi, \varphi)_{V'} = 0 \ \forall \varphi \in H'\}.$$

Wir definieren $u := j_V^{-1}\psi \in V \subset H$. Dann ist $\varphi := j_H u \in H' \subset V'$ durch $\varphi(x) = (u, x)_H$ für alle $x \in H$ charakterisiert. Es gilt dann

$$0 = (\psi, \varphi)_{V'} = (j_V^{-1}\psi, j_V^{-1}\varphi)_V = (u, j_V^{-1}\varphi)_V = \varphi(u) = (u, u)_H$$

und daraus folgt zunächst $u = 0$ und damit auch $\psi = 0$. Also ist $(H')^\perp = \{0\}$. \square

Gemäß obiger Bemerkung können wir H und H' identifizieren. Damit erhalten wir ein sogenanntes *Gelfand-Tripel*

$$V \subset H = H' \subset V'$$

von Hilbert-Räumen mit stetigen und dichten Inklusionen. Die zugehörigen Normen sind $\|\cdot\|_V$, $\|\cdot\|_H$ und $\|\cdot\|_{V'}$.

Aufgrund der Identifizierung gilt

$$\langle f, v \rangle_{V'V} = (f, v)_H \quad \text{für alle } f \in H, v \in V.$$

Vorsicht: Wir könnten genauso auch V mit V' identifizieren, aber es ist nicht möglich, gleichzeitig H mit H' und V mit V' zu identifizieren. Im ersten Fall interpretiert man $x(y) = \langle x, y \rangle_{H' \times H}$ für $x, y \in H$ (also insbesondere für $x, y \in V \subset H$) als $(x, y)_H$, im zweiten als $(x, y)_V$ für $x, y \in V$.

Nach diesem funktionalanalytischen Exkurs können wir nun die schwache Formulierung parabolischer Anfangs-Randwertprobleme angeben:

Zu gegebenem $u_0 \in H$, beliebig oft differenzierbarem $f : [0, T] \rightarrow H$ und einer V -elliptischen Bilinearform $a(\cdot, \cdot)$, bzw. dem zugehörigen Operator $A : V \rightarrow V'$ definiert in (14.5), suche $u : (0, T) \rightarrow V$ mit

$$\frac{d}{dt} \langle u, v \rangle + a(u, v) = \langle f(t), v \rangle \quad \text{für alle } v \in V, \ 0 < t < T \quad (14.6)$$

oder

$$\frac{d}{dt} u + Au = f(t) \quad \text{in } V' \quad (14.7)$$

und dem Anfangswert

$$u(0+) = u_0, \quad \text{d. h. } \lim_{t \rightarrow 0+} \|u(t) - u_0\|_H = 0. \quad (14.8)$$

Beispiel. Wir betrachten noch einmal die Wärmeleitungsgleichung vom Beginn dieses Kapitels. Falls eine *klassische* Lösung $u \in C^2(\Omega \times [0, T])$ existiert, multiplizieren wir die Differentialgleichung mit einer beliebigen Testfunktion $v \in C_0^\infty(\Omega)$ und integrieren über Ω . Mit Hilfe der Green'schen Formel gilt dann (14.6) mit

$$a(u, v) = \int_{\Omega} \operatorname{grad} u \cdot \operatorname{grad} v \, dx, \quad f(t) = f(\cdot, t) \in L^2(\Omega).$$

a ist V -elliptisch, denn nach der Poincaré'schen Ungleichung gilt $a(v, v) \geq c\|v\|_V^2$. \diamond

Der folgende Satz liefert die Existenz und Eindeutigkeit einer schwachen Lösung.

Satz 14.4. *Das Anfangs-Randwertproblem (14.6), (14.8) hat eine eindeutig bestimmte Lösung $u \in C([0, T]; H) \cap C^1((0, T); H) \cap L^2(0, T; V)$.*

Hierbei bezeichnet $C^m([0, T]; H)$ den Raum aller stetigen Funktionen $u : [0, T] \rightarrow H$, die stetige Ableitungen bis zur Ordnung m haben, versehen mit der Norm

$$\|u\|_{C^m} := \sum_{j=0}^m \max_{0 \leq t \leq T} \|u^{(j)}(t)\|_H.$$

$L^2(0, T; V)$ ist der Raum aller Funktionen $u : (0, T) \rightarrow V$ für die gilt

$$\|u\|_{L^2} := \left(\int_0^T \|u(t)\|_V^2 \, dt \right)^{1/2} < \infty.$$

Der Beweis ist recht aufwendig. Wir teilen ihn daher in kleine Happen auf. Die wesentliche Idee besteht darin, die homogene Gleichung ($f = 0$) und die inhomogene Gleichung mit homogenen Anfangsbedingungen ($u_0 = 0$) getrennt zu betrachten. Den allgemeinen Fall erhält man dann aus der Linearität durch Addition.

Um die weitere Vorgehensweise zu motivieren, betrachten wir zunächst das homogene Anfangs-Randwertproblem im \mathbb{R}^n :

$$u' + Au = 0 \quad \text{in } \mathbb{R}^n, \quad u(0) = u_0.$$

In diesem Fall ist $A \in \mathbb{R}^{n,n}$ eine Matrix. Die Lösung ist durch

$$u(t) = e^{-tA} u_0 = \frac{1}{2\pi i} \int_{\Gamma} e^{\lambda t} (\lambda I + A)^{-1} u_0 \, d\lambda,$$

gegeben, wobei Γ eine Kurve ist, die alle Eigenwerte von A umschließt. Das Integral ist komponentenweise zu verstehen. Im allgemeinen Fall ist $A : V \rightarrow V'$ ein Operator, für den wir etwas ähnliches wie die Exponentialfunktion definieren werden. Um die Cauchy-Integralformel in diesem Rahmen anwenden zu können, müssen wir uns zunächst mit dem Problem $(\lambda + A)u = f$ beschäftigen, wobei $\lambda \in \Gamma$ ist. Ein erstes Problem ist jetzt, dass wir wegen $\lambda \in \mathbb{C}$ zu komplexen Hilbert-Räumen übergehen müssen. In einem ersten Schritt komplexifizieren wir den reellen Hilbert-Raum V :

$$V_{\mathbb{C}} = V \times V = V \oplus iV$$

mit

$$(\alpha + i\beta)(u + iv) = (\alpha u - \beta v) + i(\beta u + \alpha v), \quad u, v \in V, \quad \alpha, \beta \in \mathbb{R}$$

und

$$(u + iv, x + iy)_{V, \mathbb{C}} := (u, x)_V - i(v, x)_V + i(u, y)_V + (v, y)_V.$$

$(\cdot, \cdot)_{V, \mathbb{C}}$ ist eine Hermitesche Sesquilinearform, d. h. konjugiert linear im ersten, linear im zweiten Argument und es gilt $(v, w)_{V, \mathbb{C}} = \overline{(w, v)_{V, \mathbb{C}}}$ für alle $v, w \in V_{\mathbb{C}}$. Zudem ist $(\cdot, \cdot)_{V, \mathbb{C}}$ definit, denn es gilt

$$\|u + iv\|_{V, \mathbb{C}}^2 = (u + iv, u + iv)_{V, \mathbb{C}} = \|u\|_V^2 + \|v\|_V^2 = 0$$

genau dann, wenn $u = v = 0$ oder $u + iv = 0$ ist. $(\cdot, \cdot)_{V, \mathbb{C}}$ ist also ein Skalarprodukt auf $V_{\mathbb{C}}$. Die Vollständigkeit bleibt unter dieser Konstruktion erhalten. Ebenso definieren wir

$$a_{\mathbb{C}}(u + iv, x + iy) = a(u, x) - ia(v, x) + ia(u, y) + a(v, y).$$

Da a eine Bilinearform ist, ist $a_{\mathbb{C}}$ eine Sesquilinearform. Ist a symmetrisch, dann ist offensichtlich $a_{\mathbb{C}}$ Hermitesch und insbesondere ist $a_{\mathbb{C}}(v, v)$ reell. Allgemein gilt für a V -elliptisch

$$\operatorname{Re} a_{\mathbb{C}}(u + iv, u + iv) = a(u, u) + a(v, v) \geq \alpha(\|u\|_V^2 + \|v\|_V^2) = \alpha\|u + iv\|_{V, \mathbb{C}}^2,$$

so dass $a_{\mathbb{C}}$ nun $V_{\mathbb{C}}$ -elliptisch ist.

Ab jetzt schreiben wir der Einfachheit halber wieder V statt $V_{\mathbb{C}}$, a statt $a_{\mathbb{C}}$, usw.

Lemma 14.5. (*Komplexe Version des Satzes von Lax-Milgram*)

Sei V ein komplexer Hilbert-Raum mit Skalarprodukt (\cdot, \cdot) und Norm $\|\cdot\|$. $b : V \times V \rightarrow \mathbb{C}$ sei eine Sesquilinearform mit $|b(v, w)| \leq M\|v\| \cdot \|w\|$ für alle $v, w \in V$. Ferner sei b V -elliptisch, d. h. zusätzlich existiere ein $\alpha > 0$, so dass $\operatorname{Re} b(v, v) \geq \alpha\|v\|^2$ für alle $v \in V$. Dann gibt es zu jeder stetigen Linearform $\ell : V \rightarrow \mathbb{C}$ genau ein $u \in V$ mit $b(u, v) = \ell(v)$ für alle $v \in V$.

Beweis. Nach dem Riesz'schen Darstellungssatz gibt es zu $\ell \in V'$ ein eindeutig bestimmtes $z = j_V^{-1}\ell \in V$ mit

$$(z, w) = \ell(w) \quad \text{für alle } w \in V.$$

Ebenso existiert nach dem Riesz'schen Darstellungssatz zu beliebigem $v \in V$ ein eindeutig bestimmtes $Bv \in V$ mit $(Bv, w) = b(v, w)$ für alle $w \in V$. Die Abbildung $B : V \rightarrow V$ ist linear und stetig, denn

$$\|Bv\| = \sup_{\|w\|=1} |(Bv, w)| = \sup_{\|w\|=1} |b(v, w)| \leq M\|v\|.$$

Somit ist zum einen

$$Bu = z \iff b(u, v) = \ell(v), \quad \text{für alle } v \in V$$

und zum anderen

$$Bu = z \iff \epsilon Bu - u + u = \epsilon z, \quad \epsilon > 0 \text{ beliebig.}$$

Die Fixpunktiteration

$$u^{(k+1)} = u^{(k)} - \epsilon Bu^{(k)} + \epsilon z, \quad u^{(0)} \in V \text{ beliebig}$$

konvergiert gegen den eindeutigen Fixpunkt, falls $\|I - \epsilon B\| < 1$:

$$\begin{aligned} \|v - \epsilon Bv\|^2 &= \|v\|^2 - 2\epsilon \underbrace{\operatorname{Re} (Bv, v)}_{b(v, v)} + \epsilon^2 \|Bv\|^2 \\ &\leq (1 - 2\epsilon\alpha + \epsilon^2 M)\|v\|^2 \\ &< \|v\|^2 \end{aligned}$$

für $\epsilon > 0$ hinreichend klein. Damit hat $Bu = z$ eine eindeutige Lösung $u \in V$, also gibt es ein eindeutiges $u \in V$ mit $b(u, v) = \ell(v)$ für alle $v \in V$. \square

Statt der schwachen Formulierung (14.6), (14.8) betrachten wir jetzt für festes $\lambda \in \mathbb{C}$:

$$\lambda(u, v)_H + a(u, v) = \langle f(t), v \rangle \quad \text{für alle } v \in V, f \in V', \quad (14.9)$$

wobei λ im Sektor $|\arg \lambda| \leq \pi - \theta$ liegt, vgl. Abbildung 14.1. Da a beschränkt und V -elliptisch ist, gilt für $v \neq 0$

$$\left| \arg \frac{a(v, v)}{\|v\|_V^2} \right| \leq \phi, \quad \text{wobei} \quad \cos \phi = \frac{\alpha}{M}. \quad (14.10)$$

Der Winkel ϕ hängt also nur von M und α ab, siehe Abbildung 14.2 (beachte $\alpha \leq M$). Ist a eine Hermitesche Sesquilinearform, so ist $\phi = 0$.

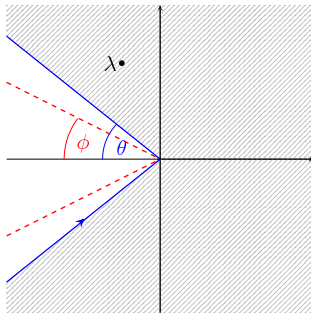


Abb. 14.1: Sektor für λ

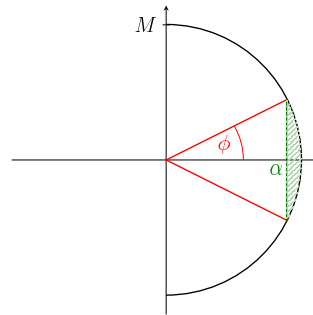


Abb. 14.2: Rayleigh-Quotienten von a

Lemma 14.6. Das Randwertproblem (14.9) hat für $f \in V'$, $|\arg \lambda| \leq \pi - \theta$, $\phi < \theta < \pi/2$, wobei ϕ der in (14.10) definierte Winkel ist, eine eindeutige Lösung $u \in V$ mit

$$\|u\|_V \leq \frac{c_\theta}{\alpha} \|f\|_{V'}, \quad c_\theta = \frac{1}{|\sin(\theta - \phi)|}$$

und, falls sogar $f \in H$ und $\lambda \neq 0$ gilt,

$$\|u\|_H \leq \frac{c_\theta}{|\lambda|} \|f\|_H.$$

Beweis. Die Sesquilinearform

$$\begin{aligned} \tilde{b} : V \times V &\rightarrow \mathbb{C} \\ (u, v) &\mapsto \lambda(u, v)_H + a(u, v) \end{aligned}$$

ist beschränkt:

$$|\tilde{b}(u, v)| \leq |\lambda| \|u\|_H \|v\|_H + M \|u\|_V \|v\|_V \leq (C^2 |\lambda| + M) \|u\|_V \|v\|_V = C(\lambda) \|u\|_V \|v\|_V.$$

Die zweite Ungleichung gilt wegen der stetigen Einbettung von V in H .

Sei $\lambda = |\lambda|e^{i\varphi}$ mit $|\varphi| \leq \pi - \theta$. Es gilt

$$\lambda(v, v)_H = \lambda \|v\|_H^2 \quad \text{und} \quad \operatorname{Re} a(v, v) \geq \alpha \|v\|_V^2.$$

Wir unterscheiden jetzt drei Fälle:

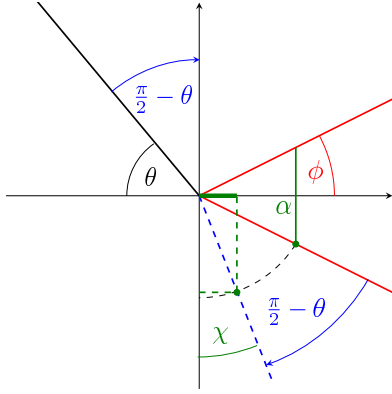
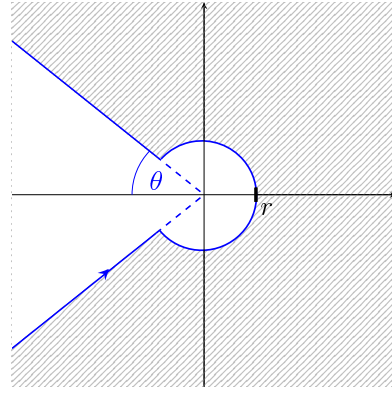


Abb. 14.3: Skizze zum Beweis von Lemma 14.6


 Abb. 14.4: Integrationskurve Γ_r

- (a) $|\varphi| \leq \pi/2$, also $\operatorname{Re} \lambda \geq 0$: Für $b(u, v) := \tilde{b}(u, v)$ gilt

$$\operatorname{Re} b(v, v) = \operatorname{Re} [\lambda(v, v)_H + a(v, v)] \geq \alpha \|v\|_V^2$$

also ist b V -elliptisch.

- (b) $\pi/2 \leq \varphi \leq \pi - \theta$: Wir definieren $b(u, v) := e^{-i(\varphi - \frac{\pi}{2})} \tilde{b}(u, v)$. Dann gilt

$$\begin{aligned} \operatorname{Re} b(v, v) &= \operatorname{Re} (e^{-i(\varphi - \frac{\pi}{2})} [\lambda(v, v)_H + a(v, v)]) \\ &= \operatorname{Re} (e^{-i(\varphi - \frac{\pi}{2})} a(v, v)) \\ &\geq \alpha |\sin(\theta - \phi)| \|v\|_V^2, \end{aligned}$$

denn für $\psi = \arg a(v, v)$ ist

$$\operatorname{Re} (e^{-i(\varphi - \frac{\pi}{2})} e^{i\psi}) = \cos(\psi - \varphi + \frac{\pi}{2}) = -\sin(\psi - \varphi) \geq |\sin(\theta - \phi)|.$$

Die letzte Abschätzung folgt aus Abbildung 14.3.

- (c) $\pi/2 \leq -\varphi \leq \pi - \theta$: Folgt aus Symmetriegründen wie in (b), indem man $b(u, v) := e^{-i(\varphi + \frac{\pi}{2})} \tilde{b}(u, v)$ setzt.

Aus der komplexen Version des Satzes von Lax-Milgram folgt in beiden Fällen, dass ein eindeutiges $u \in V$ existiert, für das

$$b(u, v) = \begin{cases} \langle f, v \rangle & \text{in Fall (a),} \\ \langle e^{-i(\varphi - \frac{\pi}{2})} f, v \rangle & \text{in Fall (b),} \\ \langle e^{-i(\varphi + \frac{\pi}{2})} f, v \rangle & \text{in Fall (c).} \end{cases}$$

Äquivalent existiert ein eindeutiges $u \in V$ mit

$$\lambda(u, v)_H + a(u, v) = \langle f, v \rangle \quad \text{für alle } v \in V.$$

Setzen wir $u = v$, so folgt

$$\alpha |\sin(\theta - \phi)| \|u\|_V^2 \leq \operatorname{Re} b(u, u) \leq |b(u, u)| \leq \|f\|_{V'} \|u\|_V$$

und damit die erste Behauptung

$$\|u\|_V \leq \frac{1}{\alpha|\sin(\theta - \phi)|} \|f\|_{V'}.$$

Für $f \in H$ gilt $\langle f, u \rangle = (f, u)_H$ wegen der Identifizierung $H = H'$. Setzen wir auch hier $u = v$, so gilt

$$\lambda \|u\|_H^2 + a(u, u) = (f, u)_H$$

und daher für $\psi = \pm \arg a(u, u)$ (Fallunterscheidung)

$$\begin{aligned} \|f\|_H \|u\|_H &\geq |(f, u)_H| = |\lambda \|u\|_H^2 + a(u, u)| \\ &= |\lambda| e^{i(\varphi \mp \psi)} \|u\|_H^2 + e^{\mp i\psi} a(u, u)| \\ &\geq |\lambda| \|u\|_H^2 |\sin(\theta - \phi)|, \end{aligned}$$

also $\|u\|_H \leq \|f\|_H / |\lambda \sin(\theta - \phi)|$. □

Bemerkung. Mit $A : V \rightarrow V'$, definiert in (14.5) durch $\langle Au, v \rangle = a(u, v)$, ist (14.9) äquivalent zu

$$\lambda u + Au = f \quad \text{in } V'. \quad (14.11)$$

Wir schreiben $(\lambda + A)^{-1} f := u$ mit der Resolvente $(\lambda + A)^{-1} : V' \rightarrow V$. Dann gilt offensichtlich

$$(\lambda + A)(\lambda + A)^{-1} f = (\lambda + A)u = f$$

und

$$(\lambda + A)^{-1}(\lambda + A)u = (\lambda + A)^{-1}f = u,$$

weil u nach Lemma 14.6 eine eindeutige Lösung von (14.11) ist.

Wir können nun den Existenzsatz für das homogene Problem angeben und beweisen:

Satz 14.7. *Das homogene Problem $u' + Au = 0$ in V' mit $u(0+) = u_0 \in H$ hat eine Lösung $u : (0, T) \rightarrow D(A) = \{v \in V \mid Av \in H\}$. u ist C^∞ bezüglich der Zeit und es gilt*

$$\begin{aligned} \|u(t)\|_H &\leq c \|u_0\|_H, \\ \|u'(t)\|_H &= \|Au(t)\|_H \leq \frac{c}{t} \|u_0\|_H, \\ \|u(t)\|_V &\leq \frac{c}{\sqrt{t}} \|u_0\|_H, \\ \|u(T)\|_H^2 &\leq \|u(0)\|_H^2 - 2\alpha \int_0^T \|u(t)\|_V^2 dt \end{aligned}$$

für alle $t > 0$.

Die Lösung wird also glatter für t weg von Null.

Beweis. Motiviert durch die Lösungsdarstellung über das Cauchy-Integral in \mathbb{R}^n definieren wir

$$u(t) := U(t)u_0 := \frac{1}{2\pi i} \int_{\Gamma} e^{\lambda t} \underbrace{(\lambda + A)^{-1} u_0}_{\in V} d\lambda \in V, \quad t > 0 \quad (14.12)$$

mit dem Operator $U(t) : H \rightarrow V$. Als Kurve Γ wählen wir den positiv orientierten Rand des in Abb. 14.1 dargestellten Sektors. Das unbestimmte Riemann-Integral existiert dann nach Lemma 14.6.

Es gilt dann wegen $\lambda = (\lambda + A) - A$

$$\begin{aligned} u'(t) &= \frac{1}{2\pi i} \int_{\Gamma} \lambda e^{\lambda t} (\lambda + A)^{-1} u_0 d\lambda \\ &= \frac{1}{2\pi i} \int_{\Gamma} e^{\lambda t} (I - A(\lambda + A)^{-1}) u_0 d\lambda \\ &= \frac{1}{2\pi i} \int_{\Gamma} e^{\lambda t} u_0 d\lambda - A \frac{1}{2\pi i} \int_{\Gamma} e^{\lambda t} (\lambda + A)^{-1} u_0 d\lambda \\ &= -Au(t). \end{aligned}$$

Das erste Integral verschwindet nach dem Cauchy'schen Integralsatz, da der Integrand analytisch ist. Es gilt also $u'(t) = -Au(t)$ für alle $t > 0$ und $u'(t) \in V$, da $(\lambda + A)^{-1}u_0 \in V$. Analog zeigen wir durch Differenzieren von u' , dass $u''(t) \in V$, $u'''(t) \in V$, \dots , so dass wir $u \in C^\infty((0, T); V)$ erhalten.

Da die Abschätzung von $\|(\lambda + A)^{-1}u_0\|_H$ aus Lemma 14.6 für $\lambda = 0$ eine Singularität hat, modifizieren wir die Integrationskurve und integrieren über Γ_r wie in Abb. 14.4 dargestellt.

Nach Lemma 14.6 gilt $\|(\lambda + A)^{-1}u_0\|_H \leq c_\theta \|u_0\|_H / |\lambda|$. Daraus folgt für $r = t^{-1}$

$$\begin{aligned} \|u(t)\|_H &\leq \frac{c_\theta}{2\pi} \int_{\Gamma_{t^{-1}}} \frac{|e^{\lambda t}|}{|\lambda|} \|u_0\|_H \frac{|d\lambda t|}{t}, & \mu = \lambda t, \quad t\Gamma_{t^{-1}} = \Gamma_1 \\ &\leq \frac{c_\theta}{2\pi} \int_{\Gamma_1} \frac{|e^\mu|}{|\mu|} \|u_0\|_H |d\mu| \\ &\leq c \|u_0\|_H \end{aligned}$$

mit c unabhängig von t .

Die zweite Behauptung schließen wir ebenso:

$$\begin{aligned} \|u'(t)\|_H = \|Au(t)\|_H &\leq \frac{c_\theta}{2\pi} \int_{\Gamma_{t^{-1}}} |e^{\lambda t}| \|u_0\|_H \frac{|d\lambda t|}{t}, & \mu = \lambda t, \quad t\Gamma_{t^{-1}} = \Gamma_1 \\ &\leq \frac{c_\theta}{2\pi} \int_{\Gamma_1} |e^\mu| \|u_0\|_H \frac{|d\mu|}{t} \\ &\leq \frac{c}{t} \|u_0\|_H. \end{aligned}$$

Die dritte Abschätzung folgt aus

$$\underbrace{\langle u', v \rangle}_{(u', v)_H} + \underbrace{\langle Au, v \rangle}_{a(u, v)} = 0 \quad \text{für alle } v \in V, \quad (14.13)$$

wegen $\langle w, v \rangle = (w, v)_H$ für alle $w \in H \subset V'$ und alle $v \in V \subset H$. Setzen wir speziell $v = u = u(t) \in V$, so ergibt sich mit Hilfe der oben gezeigten Abschätzungen und der V -Elliptizität

$$\alpha \|u\|_V^2 \leq \operatorname{Re} a(u, u) \leq |a(u, u)| = |(u', u)_H| \leq \|u'\|_H \|u\|_H \leq \frac{c^2}{t} \|u_0\|_H^2,$$

und daraus die Behauptung.

Um die letzte Abschätzung zu beweisen, verwenden wir $\operatorname{Re} (u'(t), u(t))_H = \frac{1}{2} \frac{d}{dt} \|u(t)\|_H^2$. Dann folgt aus (14.13)

$$0 = \operatorname{Re} [(u'(t), u(t))_H + a(u(t), u(t))] = \frac{1}{2} \frac{d}{dt} \|u(t)\|_H^2 + \operatorname{Re} a(u(t), u(t)).$$

Durch Integration von 0 bis T erhalten wir

$$\frac{1}{2}\|u(T)\|_H^2 - \frac{1}{2}\|u(0)\|_H^2 + \int_0^T \operatorname{Re} a(u(t), u(t)) dt = 0$$

und daraus wiederum mit der V -Elliptizität die Abschätzung

$$\|u(T)\|_H^2 + 2\alpha \int_0^T \|u(t)\|_V^2 dt \leq \|u(0)\|_H^2.$$

Es bleibt noch $\|u(t) - u_0\|_H \rightarrow 0$ für $t \rightarrow 0+$ zu zeigen. Wir zeigen dies zunächst für spezielle u_0 , nämlich $u_0 \in D(A) = \{v \in V \mid Av \in H\}$. Beachte, dass $A : V \rightarrow V' \supset H$ gilt. Mit Lemma 14.6 gilt

$$\begin{aligned} \|u(t) - u_0\|_H &= \left\| \frac{1}{2\pi i} \int_{\Gamma_{t-1}} e^{\lambda t} (\lambda + A)^{-1} u_0 d\lambda - \frac{1}{2\pi i} \int_{\Gamma_{t-1}} e^{\lambda t} \frac{1}{\lambda} u_0 d\lambda \right\|_H \\ &\leq \frac{1}{2\pi} \int_{\Gamma_{t-1}} |e^{\lambda t}| \left\| \underbrace{\left((\lambda + A)^{-1} - \frac{1}{\lambda} \right) u_0}_{\lambda^{-1}(\lambda + A)^{-1} \underbrace{Au_0}_{\in H}} \right\|_H |d\lambda| \\ &\leq \frac{c_\theta}{2\pi} \int_{\Gamma_{t-1}} |e^{\lambda t}| |\lambda^{-2}| \|Au_0\|_H |d\lambda| \\ &\leq \frac{c_\theta}{2\pi} \int_{\Gamma_1} |e^\mu| |\mu^{-2}| |d\mu| t \|Au_0\|_H \\ &\leq \frac{C'}{2\pi} t \|Au_0\|_H. \end{aligned}$$

Damit gilt $\|u(t) - u_0\|_H \rightarrow 0$ für $t \rightarrow 0+$ falls $u_0 \in D(A)$.

Wir zeigen jetzt, dass $D(A)$ dicht in H ist. Nach Voraussetzung ist H dicht in V' und $A^{-1} : V' \rightarrow V$ ist nach Lemma 14.1 stetig und bijektiv. Also folgt $D(A) = A^{-1}(H)$ liegt dicht in $A^{-1}(V') = V$. Somit ist $D(A)$ dicht in V . V liegt dicht in H und die Einbettung $V \hookrightarrow H$ ist stetig, also ist $D(A)$ dicht in H .

Sei jetzt $u_0^{(n)}$ eine Folge in $D(A)$ mit $\|u_0^{(n)} - u_0\|_H \leq n^{-1}$. Dann gilt

$$\begin{aligned} \|u(t) - u_0\|_H &= \|U(t)u_0 - u_0\|_H \\ &\leq \|U(t)u_0 - U(t)u_0^{(n)}\|_H + \|U(t)u_0^{(n)} - u_0^{(n)}\|_H + \|u_0^{(n)} - u_0\|_H \\ &\leq C\|u_0 - u_0^{(n)}\|_H + C't\|Au_0^{(n)}\|_H + n^{-1} \\ &\leq cn^{-1} \end{aligned}$$

für $t < n^{-1}\|Au_0^{(n)}\|_H^{-1}$. Also gilt $\|u(t) - u_0\|_H \rightarrow 0$ für $t \rightarrow 0+$ auch für beliebige $u_0 \in H$. \square

Als nächstes betrachten wir jetzt das inhomogene Problem

$$\begin{aligned} u' + Au &= f(t), \quad t > 0, \\ u(0) &= u_0. \end{aligned} \tag{14.14}$$

In \mathbb{R}^n ist die Lösung bekanntlich durch die Variation der Konstanten Formel gegeben:

$$u(t) = e^{-tA}u_0 + \int_0^t e^{-(t-\tau)A}f(\tau)d\tau.$$

Wir werden wie beim homogenen Problem diese Formel verallgemeinern.

Satz 14.8. *Es sei $f : [0, T] \rightarrow H$ stetig differenzierbar. Dann hat das Anfangs-Randwertproblem (14.14) die eindeutige Lösung*

$$u(t) = U(t)u_0 + \int_0^t U(t - \tau)f(\tau)d\tau \in C([0, T]; H) \cap L^2(0, T; V)$$

(Variation der Konstanten Formel) und es gilt

$$\|u(t)\|_H + \left(\int_0^t \|u(\tau)\|_V^2 d\tau \right)^{1/2} \leq C \left(\|u_0\|_H + \left(\int_0^t \|f(\tau)\|_{V'}^2 d\tau \right)^{1/2} \right), \quad 0 \leq t \leq T.$$

Beweis. Wir schreiben äquivalent

$$u(t) = U(t)u_0 + \int_0^t U(\tau)f(t - \tau)d\tau.$$

Wegen $U'(t) = -AU(t)$ gilt einerseits

$$u'(t) = -AU(t)u_0 + U(t)f(0) + \int_0^t U(\tau)f'(t - \tau)d\tau$$

und andererseits mit Hilfe partieller Integration

$$\begin{aligned} Au(t) &= AU(t)u_0 + \int_0^t AU(\tau)f(t - \tau)d\tau \\ &= AU(t)u_0 - U(\tau)f(t - \tau) \Big|_0^t - \int_0^t U(\tau)f'(t - \tau)d\tau \\ &= AU(t)u_0 - (U(t)f(0) - U(0)f(t)) - \int_0^t U(\tau)f'(t - \tau)d\tau. \end{aligned}$$

Hieraus folgt wegen $U(0)f(t) = f(t)$, dass $u(t)$ Lösung von (14.14) ist.

Um die Abschätzung zu zeigen, verwenden wir jetzt die Technik der “Energieabschätzung”. Wir integrieren

$$(u', u)_H + a(u, u) = \langle f, u \rangle$$

von 0 bis t . Hierbei genügt es, homogene Anfangsbedingungen $u_0 = 0$ zu betrachten, denn für das homogene Problem haben wir <https://na.math.kit.edu/research/preprints.php> wir die entsprechende Abschätzung bereits in Satz 14.7 bewiesen:

$$\begin{aligned} \frac{1}{2}\|u(t)\|_H^2 - \underbrace{\frac{1}{2}\|u(0)\|_H^2}_{=0} + \operatorname{Re} \int_0^t a(u(\tau), u(\tau))d\tau &= \operatorname{Re} \int_0^t \langle f(\tau), u(\tau) \rangle d\tau \\ &\leq \int_0^t |\langle f(\tau), u(\tau) \rangle| d\tau \\ &\leq \int_0^t \|f(\tau)\|_{V'} \|u(\tau)\|_V d\tau \\ &\leq \left(\int_0^t \|f(\tau)\|_{V'}^2 d\tau \right)^{1/2} \left(\int_0^t \|u(\tau)\|_V^2 d\tau \right)^{1/2}, \end{aligned}$$

wobei wir für die letzte Abschätzung die Cauchy-Schwarz'sche Ungleichung im $L^2(0, t)$ verwendet haben. Aus der V -Elliptizität folgt daraus

$$\left(\int_0^t \|u(\tau)\|_V^2 d\tau \right)^{1/2} \leq \frac{1}{\alpha} \left(\int_0^t \|f(\tau)\|_V^2 d\tau \right)^{1/2},$$

und

$$\|u(t)\|_H^2 \leq \frac{2}{\alpha} \int_0^t \|f(\tau)\|_V^2 d\tau.$$

Die Eindeutigkeit der Lösung ist klar, weil die Differenz zweier Lösungen das homogene Anfangs-Randwertproblem mit homogenen Anfangsbedingungen erfüllt. Diese besitzt wegen der Energieabschätzung nur die Nulllösung. \square

Der Beweis von Satz 14.4 folgt nun aus den Sätzen 14.7 und 14.8 mit Hilfe der Linearität des Problems durch Addition der Lösungen.

14.2 Finite Element-Diskretisierung im Raum

Wie in Abschnitt 14.1 sei $a : V \times V \rightarrow \mathbb{R}$ eine V -elliptische Bilinearform und es sei $f : [0, T] \rightarrow H$ stetig differenzierbar. Die schwache Formulierung des parabolischen Problems lautet

$$\begin{aligned} (u'(t), v)_H + a(u(t), v) &= (f(t), v)_H & \forall v \in V, \quad 0 < t \leq T, \\ (u(0+), w)_H &= (u_0, w)_H & \forall w \in H. \end{aligned} \quad (14.15)$$

Ab jetzt beschränken wir uns auf symmetrische Bilinearformen a und damit wieder auf reelle Hilbert-Räume, denn man kann sich leicht überlegen, dass die im letzten Abschnitt konstruierte Lösung reell ist, wenn u_0 und f reell sind. Dazu muss man nur in der Integralformel zur Darstellung der Lösung die Kurve Γ symmetrisch zur reellen Achse wählen (Übung).

Wir ersetzen jetzt V durch einen endlichdimensionalen Unterraum $V_h \subset V$. Typisch wäre etwa $V = H_0^1(\Omega)$, $H = L^2(\Omega)$ und V_h ein finiter Element-Raum, z. B. die Menge der stetigen, stückweise linearen Funktionen auf $\bar{\Omega}$ mit Randwerten 0. Dabei sei h der maximale Durchmesser der finiten Elemente.

Gesucht ist dann $u_h : [0, T] \rightarrow V_h$ mit

$$\begin{aligned} (u_h'(t), v_h)_H + a(u_h(t), v_h) &= (f(t), v_h)_H & \forall v_h \in V_h, \quad 0 < t \leq T, \\ (u_h(0+), w_h)_H &= (u_0, w_h)_H & \forall w_h \in V_h. \end{aligned} \quad (14.16)$$

Sei $(\varphi_1, \dots, \varphi_N)$ eine Basis von V_h ,

$$u_h(t) = \sum_{j=1}^N \mu_j(t) \varphi_j, \quad \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_N \end{bmatrix}.$$

Es genügt, in (14.16) für v_h die Basisfunktionen φ_i , $i = 1, \dots, N$, einzusetzen:

$$\sum_{j=1}^N \mu_j'(t) (\varphi_j, \varphi_i)_H + \sum_{j=1}^N \mu_j(t) a(\varphi_j, \varphi_i) = (f(t), \varphi_i)_H, \quad i = 1, \dots, N.$$

Mit

$$\begin{aligned} M &= (m_{ij})_{i,j=1}^N, & m_{ij} &= (\varphi_j, \varphi_i)_H && \text{Massematrix, sym. pos. def.} \\ K &= (k_{ij})_{i,j=1}^N, & k_{ij} &= a(\varphi_j, \varphi_i) && \text{Steifigkeitsmatrix} \\ \ell &= (\ell_i)_{i=1}^N, & \ell_i &= (f, \varphi_i)_H && \text{Lastvektor} \end{aligned}$$

erhalten wir ein System gewöhnlicher Differentialgleichungen

$$\begin{aligned} M\mu'(t) + K\mu(t) &= \ell(t), & t &> 0, \\ \mu(0) &= \mu_0, & \mu_{0,i} &= (u_0, \varphi_i)_H. \end{aligned}$$

Die Steifigkeitsmatrix K ist symmetrisch, wenn die Bilinearform a symmetrisch ist. Aus der V -Elliptizität folgt ferner, dass K auch positiv definit ist.

Die symmetrisch positiv definite Massematrix M können wir als $M = CC^T$ zum Beispiel mit der Choleski-Zerlegung faktorisieren. Setzen wir

$$\begin{aligned} y &= C^T \mu, & A &= C^{-1}KC^{-T}, & g &= C^{-1}\ell, \\ y_0 &= C^T \mu_0, \end{aligned}$$

so ist A ebenfalls symmetrisch, wenn K symmetrisch ist. In jedem Fall ist A positiv definit (der Wertebereich liegt in der offenen rechten Halbebene). Die neuen Variablen erfüllen ein System von Differentialgleichungen in Standardform:

$$\begin{aligned} y' + Ay &= g(t), & \text{in } \mathbb{R}^N, \\ y(0) &= y_0. \end{aligned}$$

Diese ist für theoretische Zwecke günstiger. Bei der Implementierung ist es häufig besser, das System in der Ausgangsform zu belassen und die Verfahren entsprechend umzuschreiben. Dies führt bei Mehrschritt- oder Runge-Kutta-Verfahren auf lineare Gleichungssysteme der Form $M - \gamma hJ$ statt $I - \gamma hJ$.

Die Vorgehensweise, die allen Konvergenzbeweisen in diesem Kapitel zugrundeliegt, ist folgende: Zunächst zeigt man die Stabilität der Diskretisierung. Das bedeutet grob gesprochen, dass sich die Lösung durch die rechte Seite und die Anfangsbedingung abschätzen lässt. Im zweiten Schritt zeigt man, dass der Fehler schwache Lösung eines modifizierten Problems mit "kleiner" rechter Seite und "kleiner" Anfangsbedingung ist. Beide Resultate zusammen liefern die Konvergenz.

Wir beginnen mit der Stabilitätsaussage.

Lemma 14.9. (*Stabilität der Semidiskretisierung*)

Sei $u_h : [0, T] \rightarrow V_h$ Lösung von (14.16), dann ist

$$\|u_h(t)\|_H + \left(\int_0^t \|u_h(\tau)\|_V^2 d\tau \right)^{1/2} \leq C \left(\|u_0\|_H + \left(\int_0^t \|f(\tau)\|_V^2 d\tau \right)^{1/2} \right).$$

Beweis. Der Beweis von Satz 14.8 kann direkt auf diesen Fall übertragen werden. Hierbei ist wichtig, dass $\operatorname{Re} a(v_h, v_h) \geq \alpha \|v_h\|_V^2$ für alle $v_h \in V_h$ mit α unabhängig von h gilt. \square

Das nächste Ziel ist jetzt, die Konvergenz der Semidiskretisierung zu zeigen. Dazu benötigen wir die sogenannte *Ritz-Projektion*: Für $u \in V$ sei $R_h u \in V_h$ die Lösung von

$$a(R_h u, v_h) = a(u, v_h) \quad \forall v_h \in V_h. \quad (14.17)$$

Aus Satz 12.7 und Lemma 13.12 wissen wir, dass für stückweise lineare finite Elemente ($V = H_0^1(\Omega)$, $H = L^2(\Omega)$)

$$\begin{aligned} \|R_h u - u\|_V &\leq Ch|u|_2, \\ \|R_h u - u\|_H &\leq Ch^2|u|_2 \end{aligned} \quad (14.18)$$

gilt, wenn das Problem H^2 -regulär ist. Zur Erinnerung:

$$|u|_2^2 = \sum_{|\beta|=2} \int_{\Omega} (D^{\beta} u)^2 dx, \quad \|u\|_2^2 = \sum_{|\beta| \leq 2} \int_{\Omega} (D^{\beta} u)^2 dx.$$

Die Orthogonalprojektion $P_h : H \rightarrow V_h$ bezüglich des Skalarprodukts $(\cdot, \cdot)_H$ in H , ist definiert durch

$$(P_h u, v_h)_H = (u, v_h)_H \quad \text{für alle} \quad v_h \in V_h.$$

Es gilt dann

$$\|P_h u - u\|_H = \min_{v_h \in V_h} \|v_h - u\|_H \leq Ch^2|u|_2. \quad (14.19)$$

Satz 14.10. Die Lösung $u : [0, T] \rightarrow V$ von (14.15) sei genügend oft differenzierbar und es gelte (14.18). Dann gilt für die Lösung $u_h : [0, T] \rightarrow V_h$ von (14.16)

$$\begin{aligned} \left(\int_0^T \|u_h(t) - u(t)\|_V^2 dt \right)^{1/2} &= O(h), \\ \max_{0 \leq t \leq T} \|u_h(t) - u(t)\|_H &= O(h^2). \end{aligned}$$

Beweis. Nach Voraussetzung gilt

$$\begin{aligned} (u'_h, v_h)_H + a(u_h, v_h) &= (f, v_h)_H & \forall v_h \in V_h, \\ (u', v_h)_H + a(u, v_h) &= (f, v_h)_H & \forall v_h \in V_h \subset V. \end{aligned}$$

Nach Definition (14.17) der Ritz-Projektion folgt aus der zweiten Gleichung

$$(R_h u', v_h)_H + a(R_h u, v_h) = (f, v_h)_H + (R_h u' - u', v_h)_H.$$

Wir definieren $e_h = u_h - R_h u$ und subtrahieren:

$$(e'_h, v_h)_H + a(e_h, v_h) = -(R_h u' - u', v_h)_H \quad \forall v_h \in V_h.$$

Aus Lemma 14.9 folgt daraus

$$\|e_h(t)\|_H + \left(\int_0^t \|e_h(\tau)\|_V^2 d\tau \right)^{1/2} \leq C \left(\|e_h(0)\|_H + \left(\int_0^t \|R_h u'(\tau) - u'(\tau)\|_V^2 d\tau \right)^{1/2} \right)$$

Da die Einbettung $H \hookrightarrow V'$ stetig ist, folgt

$$\|R_h u' - u'\|_{V'} \leq C \|R_h u' - u'\|_H \leq Ch^2|u'|_2.$$

Ferner ist wegen $u_h(0) = P_h u_0$

$$e_h(0) = P_h u_0 - R_h u_0 = (P_h u_0 - u_0) + (u_0 - R_h u_0),$$

woraus mit Voraussetzung (14.18) und (14.19) $\|e_h(0)\|_H \leq Ch^2|u_0|_2$ folgt. Es gilt also

$$\|e_h(t)\|_H + \left(\int_0^t \|e_h(\tau)\|_V^2 d\tau \right)^{1/2} \leq Ch^2 \left(|u_0|_2 + \left(\int_0^t |u'(\tau)|_2^2 d\tau \right)^{1/2} \right).$$

Aus

$$u_h - u = (u_h - R_h u) + (R_h u - u) = e_h + (R_h u - u)$$

folgt die zweite Abschätzung aus der Dreiecksungleichung. Die erste Abschätzung ergibt sich mit Hilfe der Minkowski'schen Ungleichung

$$\left(\int_0^t \|u_h(\tau) - u(\tau)\|_V^2 d\tau \right)^{1/2} \leq \left(\int_0^t \|e_h(\tau)\|_V^2 d\tau \right)^{1/2} + \left(\int_0^t \|R_h u(\tau) - u(\tau)\|_V^2 d\tau \right)^{1/2}$$

und Voraussetzung (14.18). \square

14.3 Vollständige Diskretisierung mit dem impliziten Euler-Verfahren

Im letzten Abschnitt haben wir das Anfangs-Randwertproblem im Ort diskretisiert. Jetzt wollen wir auch die Zeit diskretisieren. Da explizite Verfahren ungeeignet sind, beginnen wir mit dem einfachsten impliziten Verfahren, dem impliziten Euler-Verfahren. Die Zeitschrittweite bezeichnen wir mit τ , den maximalen Durchmesser der finiten Elemente der Ortsdiskretisierung wie bisher mit h .

Wir bezeichnen mit $u_n \in V_h$, $u_n \approx u_h(t_n)$ die Approximation zur Zeit $t_n = n\tau$. Die Aufgabe lautet: Zu gegebenem Anfangswert u_0 , suche

$$\left(\frac{u_n - u_{n-1}}{\tau}, v_h \right)_H + a(u_n, v_h) = (f(t_n), v_h)_H \quad \forall v_h \in V_h, \quad n = 1, 2, \dots \quad (14.20)$$

Mit Hilfe der Basis $\{\varphi_j\}_{j=1}^N$ von V_h schreiben wir

$$u_n = \sum_{j=1}^N U_{n,j} \varphi_j, \quad U_n = (U_{n,j})_{j=1}^N \in \mathbb{R}^N.$$

Dann ist (14.20) äquivalent zum linearen Gleichungssystem

$$\frac{1}{\tau} M(U_n - U_{n-1}) + K U_n = \ell(t_n).$$

In jedem Schritt muss also ein lineares Gleichungssystem mit der Matrix $(\frac{1}{\tau}M + K)$ gelöst werden. Dazu kann man im symmetrischen Fall zum Beispiel das vorkonditionierte Verfahren der konjugierten Gradienten oder Mehrgitterverfahren verwenden.

Wir gehen wie im letzten Abschnitt vor und zeigen zuerst die Stabilität.

Lemma 14.11. (*Stabilität*)

Sei $u_n \in V_h$, $n = 0, 1, 2, \dots$ Lösung von (14.20). Dann gilt

$$\|u_n\|_H + \left(\tau \sum_{j=1}^n \|u_j\|_V^2 \right)^{1/2} \leq C \left(\|u_0\|_H + \left(\tau \sum_{j=1}^n \|f(t_j)\|_{V'}^2 \right)^{1/2} \right).$$

Beweis. Wir verwenden wieder die Technik der Energieabschätzung und wählen $v_h = u_n$ in (14.20) und summieren von $n = 1, \dots, m$:

$$\sum_{n=1}^m ((u_n - u_{n-1}, u_n)_H + \tau a(u_n, u_n)) = \tau \sum_{n=1}^m (f(t_n), u_n)_H.$$

Es gilt

$$\begin{aligned} \sum_{n=1}^m (u_n - u_{n-1}, u_n)_H &= \frac{1}{2} \|u_m\|_H^2 + \frac{1}{2} \|u_m\|_H^2 - (u_{m-1}, u_m)_H + \frac{1}{2} \|u_{m-1}\|_H^2 \\ &\quad + \frac{1}{2} \|u_{m-1}\|_H^2 - (u_{m-2}, u_{m-1})_H + \frac{1}{2} \|u_{m-2}\|_H^2 \\ &\quad + \dots \\ &\quad + \frac{1}{2} \|u_1\|_H^2 - (u_0, u_1)_H + \frac{1}{2} \|u_0\|_H^2 \\ &\quad - \frac{1}{2} \|u_0\|_H^2 \\ &\geq \frac{1}{2} (\|u_m\|_H^2 - \|u_0\|_H^2), \end{aligned}$$

da aus der Cauchy-Schwarz'schen Ungleichung $(u_{n-1}, u_n)_H \leq \|u_n\|_H \|u_{n-1}\|_H$ folgt. Somit erhalten wir aus $2xy \leq \frac{1}{\alpha} x^2 + \alpha y^2$ ($\alpha > 0$)

$$\begin{aligned} \frac{1}{2} (\|u_m\|_H^2 - \|u_0\|_H^2) + \tau \sum_{n=1}^m a(u_n, u_n) &\leq \tau \sum_{n=1}^m \|f(t_n)\|_{V'} \|u_n\|_V \\ &\leq \tau \frac{1}{2\alpha} \sum_{n=1}^m \|f(t_n)\|_{V'}^2 + \frac{\alpha}{2} \tau \sum_{n=1}^m \|u_n\|_V^2 \end{aligned}$$

und daraus mit Hilfe der V -Elliptizität

$$\|u_m\|_H^2 - \|u_0\|_H^2 + \alpha \tau \sum_{n=1}^m \|u_n\|_V^2 \leq \frac{\tau}{\alpha} \sum_{n=1}^m \|f(t_n)\|_{V'}^2$$

und daraus die Behauptung, da für $x, y > 0$ die Ungleichung $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ gilt. \square

Satz 14.12. (Konvergenz)

Die Lösung u von (14.15) sei genügend regulär und es gelte (14.18) für die Ritz-Projektion. Dann gilt

$$\begin{aligned} \left(\tau \sum_{j=1}^n \|u_j - u(t_j)\|_V^2 \right)^{1/2} &\leq C(h + \tau) \\ \|u_n - u(t_n)\|_H &\leq C(h^2 + \tau), \end{aligned}$$

wobei die Konstante C unabhängig von n, h und τ ist.

Beweis. Wir gehen analog zum Beweis von Satz 14.10 vor und setzen die exakte Lösung in die Gleichung (14.20) für das numerische Verfahren ein. Die exakte Lösung erfüllt

$$(u'(t), v)_H + a(u(t), v) = \langle f(t), v \rangle \quad \forall v \in V \supset V_h.$$

Daraus folgt nach Definition der Ritz-Projektion

$$(R_h u'(t), v_h)_H + a(R_h u(t), v_h) = \langle f(t), v_h \rangle + (R_h u'(t) - u'(t), v_h)_H \quad \forall v_h \in V_h.$$

Eingesetzt in (14.20) ergibt sich

$$\begin{aligned} & \frac{1}{\tau} (R_h u(t_n) - R_h u(t_{n-1}), v_h)_H + a(R_h u(t_n), v_h) \\ &= \langle f(t_n), v_h \rangle + (R_h u'(t_n) - u'(t_n), v_h)_H + \left(R_h \left(\frac{u(t_n) - u(t_{n-1})}{\tau} - u'(t_n) \right), v_h \right)_H. \end{aligned}$$

Wir setzen $e_n = u_n - R_h u(t_n)$. Subtraktion von (14.20) von obiger Gleichung liefert

$$\frac{1}{\tau} (e_n - e_{n-1}, v_h)_H + a(e_n, v_h) = (d_n, v_h)_H,$$

wobei

$$d_n = R_h u'(t_n) - u'(t_n) + \frac{u(t_n) - u(t_{n-1})}{\tau} - u'(t_n)$$

Voraussetzung (14.18) liefert $\|R_h u'(t_n) - u'(t_n)\|_H \leq Ch^2$ und aus der Konvergenz des impliziten Euler-Verfahrens folgt

$$\left\| \frac{u(t_n) - u(t_{n-1})}{\tau} - u'(t_n) \right\|_V \leq C\tau.$$

Schließlich ist R_h ein beschränkter Operator, also ist wegen der stetigen Einbettung $H \subset V'$

$$\|d_n\|_{V'} \leq C\|d_n\|_H \leq C(h^2 + \tau).$$

Außerdem ist $\|e_0\|_H = \|u(0) - R_h u(0)\|_H = O(h^2)$, woraus mit der Stabilitätsungleichung in Lemma 14.11

$$\|e_n\|_H + \left(\tau \sum_{j=0}^n \|e_j\|_V^2 \right)^{1/2} \leq C(h^2 + \tau)$$

folgt. Die Behauptung ergibt sich aus

$$u_n - u(t_n) = u_n - R_h u(t_n) + R_h u(t_n) - u(t_n) = e_n + R_h u(t_n) - u(t_n)$$

und $\|R_h u(t_n) - u(t_n)\|_V = O(h)$ sowie $\|R_h u(t_n) - u(t_n)\|_H = O(h^2)$ nach (14.18) wie in Satz 14.10 mit der Minkowski-Ungleichung. \square

Die hier angewendete Technik lässt sich auf algebraisch stabile Runge-Kutta-Verfahren, deren Stabilitätsfunktion $|R(\infty)| < 1$ erfüllt, z. B. auf Radau-Kollokationsverfahren, verallgemeinern. Man kann für solche Verfahren

$$\|u_n - u(t_n)\|_H \leq C(h^2 + \tau^{q+1})$$

zeigen, wobei q die Stufenordnung ist.

14.4 Zeitdiskretisierung mit BDF-Verfahren

Bevor wir jetzt die Zeitdiskretisierung mit BDF-Verfahren diskutieren, wiederholen wir einige Resultate aus Abschnitt 10.3 über BDF-Verfahren für gewöhnliche Differentialgleichungen. Abweichend von Kapitel 10 bezeichnen wir die Zeitschrittweite mit τ . Die Resultate in diesem und den folgenden Abschnitten stammen aus den Arbeiten von Lubich (Lubich 1988, Lubich 1991, Lubich 1992).

Für $y' = f(y)$, $y(t_0) = y_0$ hat das BDF(k)-Verfahren die Form

$$\sum_{j=0}^k \alpha_{k-j} y_{n-j} = \tau f(y_n), \quad n \geq k,$$

wobei $y_0 = y(t_0)$ und y_1, \dots, y_{k-1} geeignet zu wählende Startwerte sind (etwa durch BDF-Verfahren niedrigerer Ordnung). BDF(k)-Verfahren sind implizite Verfahren der Ordnung k . Für $k \leq 6$ ist BDF(k) $A(\theta)$ -stabil mit $\theta > 0$, der Stabilitätsbereich enthält also einen Sektor in der linken Halbebene. Da BDF-Verfahren für $k > 6$ instabil sind, setzen wir im Folgenden $k \leq 6$ voraus. Eine wichtige Rolle spielen die charakteristischen Polynome ρ und σ eines Mehrschrittverfahrens, vgl. (10.7). Definieren wir

$$\delta(\zeta) = \frac{\rho(\zeta^{-1})}{\sigma(\zeta^{-1})} = \sum_{j=1}^k \frac{1}{j} (1 - \zeta)^j = \sum_{j=0}^k \delta_j \zeta^j,$$

wobei $\delta_j = \alpha_{k-j}$, $j = 0, \dots, k$, dann ist der Stabilitätsbereich nach Lemma 10.6 durch

$$\mathcal{S} = \mathbb{C} \setminus \{\delta(\zeta) \mid |\zeta| < 1\}$$

gegeben. Die Ordnungsbedingung kann man nach (10.8) als

$$\frac{1}{\tau} \delta(e^{-\tau}) = 1 + O(\tau^k), \quad \text{für} \quad \tau \rightarrow 0,$$

schreiben, insbesondere gilt $\delta(1) = 0$. Nach Definition 10.10 ist das Verfahren $A(\theta)$ -stabil genau dann, wenn

$$|\arg(\delta(\zeta))| \leq \pi - \theta \quad \text{für} \quad |\zeta| \leq 1. \quad (14.21)$$

Man kann explizit nachrechnen, dass Eins die einzige Nullstelle von δ vom Betrag eins ist. Wegen der Stabilität ist Eins dann auch die einzige Nullstelle von δ vom Betrag ≤ 1 . Es gilt also

$$\delta(\zeta) = (1 - \zeta)\mu(\zeta), \quad \mu(\zeta) \neq 0 \quad \text{für} \quad |\zeta| \leq 1. \quad (14.22)$$

Bevor wir uns parabolischen Anfangs-Randwertproblemen widmen, betrachten wir wieder die Testgleichung $u' = \lambda u$, $u(0) = u_0$. Die Näherungen des BDF(k)-Verfahrens sind durch

$$\sum_{j=0}^k \delta_j u_{n-j} = \tau \lambda u_n, \quad n \geq k, \quad (14.23)$$

gegeben. Multiplikation mit ζ^n und Summation liefert

$$\sum_{n=k}^{\infty} \sum_{j=0}^k \delta_j u_{n-j} \zeta^n = \tau \lambda \sum_{n=k}^{\infty} u_n \zeta^n =: \tau \lambda u(\zeta).$$

Andererseits ist

$$\begin{aligned} \delta(\zeta) u(\zeta) - \sum_{n=k}^{\infty} \sum_{j=0}^k \delta_j u_{n-j} \zeta^n &= \sum_{n=k}^{\infty} \left(\sum_{j=0}^{\min\{n-k, k\}} \delta_j u_{n-j} - \sum_{j=0}^k \delta_j u_{n-j} \right) \zeta^n \\ &= \sum_{n=k}^{\infty} \left(\sum_{j=n-k+1}^k \delta_j u_{n-j} \right) \zeta^n \\ &=: u^0(\zeta) = \sum_{n=k}^{2k-1} u_n^0 \zeta^n, \end{aligned}$$

da die innere Summe in der zweiten Zeile für $n > 2k - 1$ leer ist. Damit ist u^0 ein Polynom in ζ welches nur von den Startwerten u_0, \dots, u_{k-1} abhängt (daher auch die Bezeichnung u^0). Es folgt zunächst formal

$$[\delta(\zeta) - \tau\lambda]u(\zeta) = u^0(\zeta). \quad (14.24)$$

Wie in Lemma 14.6 gilt wegen (14.21) für $|\arg(-\lambda)| \leq \phi$, $0 < \phi < \theta$ und $\operatorname{Re}(-\tau\lambda) \geq \alpha$, die Abschätzung

$$\|u(\zeta)\|_H \leq C\|u^0(\zeta)\|_H, \quad C = \frac{1}{\alpha \sin(\theta - \phi)}.$$

Damit ist die Abbildung $\zeta \mapsto u(\zeta)$ analytisch in $|\zeta| < 1$ und stetig auf $|\zeta| \leq 1$. Die Koeffizienten u_n erfüllen (14.23). Die Koeffizienten einer analytischen Funktion können mit Hilfe der Cauchy-Integralformel durch

$$u_n = \frac{1}{2\pi i} \int_{|\zeta|=1} \zeta^{-(n+1)} u(\zeta) d\zeta = \frac{1}{2\pi} \int_0^{2\pi} e^{-in\omega} \widehat{u}(\omega) d\omega \quad (14.25)$$

ausgedrückt werden. Hierbei haben wir die Notation $\widehat{u}(\omega) = u(e^{i\omega})$ verwendet. u_n ist also der n -te Fourier-Koeffizient von \widehat{u} .

Wir definieren jetzt Koeffizienten $r_n(z)$ durch

$$(\delta(\zeta) - z)^{-1} =: \sum_{n \geq 0} r_n(z) \zeta^n. \quad (14.26)$$

Damit ist

$$u_n = \sum_{j=k}^n r_{n-j}(\tau\lambda) u_j^0, \quad n \geq k.$$

Wählen wir die Anfangswerte so, dass $u^0(\zeta) = \zeta^k$, dann ist $u_n = r_{n-k}(\tau\lambda)$. Zur Abschätzung von u_n benötigen wir also Abschätzungen für die Koeffizienten $r_n(z)$.

Lemma 14.13. *Ist θ der Winkel, für den das BDF-Verfahren $A(\theta)$ -stabil ist, so gibt es für ϕ mit $0 < \phi < \theta$ ein $\rho < 1$ und ein r hinreichend klein, so dass gilt*

$$|r_n(z)| \leq \begin{cases} C \frac{\rho^n}{1+|z|}, & |\arg z| = \pi - \phi, |z| \geq r, \\ C |e^{nz/2}|, & |\arg z| = \pi - \phi, |z| \leq r, \\ C e^{2n|z|}, & |z| \leq r, \end{cases} \quad (14.27)$$

wobei C unabhängig von z und n ist.

Beweis. Nach Voraussetzung ist $\delta(\zeta)$ analytisch (Polynom) und hat in einer Umgebung des abgeschlossenen Einheitskreises keine Nullstellen außer $\zeta = 1$. Wegen (14.21) ist die Funktion $(\delta(\zeta) - z)^{-1}$ für $|z| \geq r$ und $|\arg(-z)| \leq \phi < \theta$ analytisch in $|\zeta| \leq \rho^{-1}$, wobei ρ unabhängig ist von z aber abhängig von r . Schwierigkeiten treten für z nahe 0 auf, denn aus der Konsistenz des Verfahrens folgt $\delta(1) = 0$. Wir unterscheiden daher im Wesentlichen zwei Fälle, nämlich z weg von 0 auf der Kurve Γ_r in Abbildung 14.5 und z nahe 0. Aus technischen Gründen teilen wir den zweiten Fall noch einmal auf.

(i) $|\arg z| = \pi - \phi$, $|z| \geq r > 0$:

Mit der Cauchy-Integralformel gilt

$$r_n(z) = \frac{1}{2\pi i} \int_{|\zeta|=\rho^{-1}} \frac{1}{\delta(\zeta) - z} \zeta^{-(n+1)} d\zeta. \quad (14.28)$$

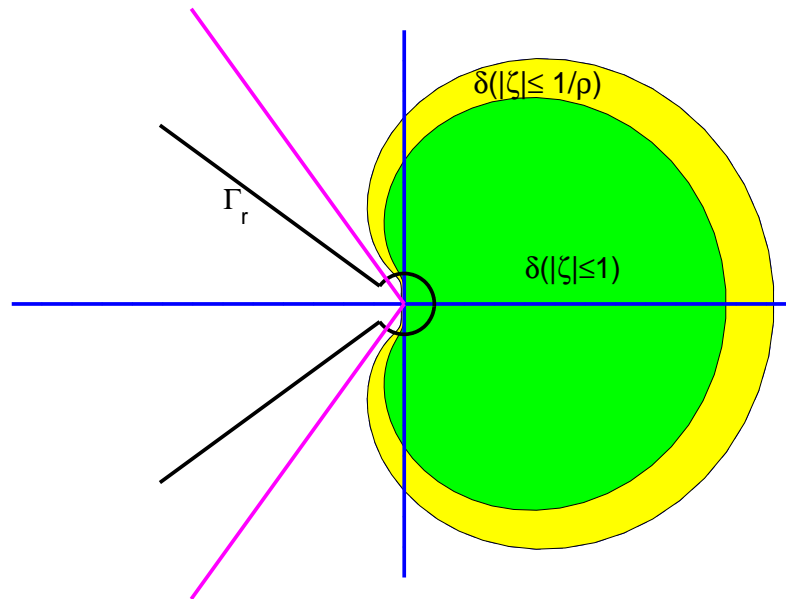


Abb. 14.5: Kurve Γ_r , $\mathbb{C} \setminus \mathcal{S} = \{\delta(\zeta) \mid |\zeta| \leq 1\}$ und $\{\delta(\zeta) \mid |\zeta| \leq \rho^{-1}\}$ für $\rho < 1$.

Da δ ein Polynom ist, ist die Menge $\{\delta(\zeta) \mid |\zeta| \leq \rho^{-1}\}$ beschränkt und es gilt

$$|\delta(\zeta) - z| > C|z| \quad \text{für } |\zeta| \leq \rho^{-1}, \quad C > 0.$$

Verwenden wir diese Abschätzung in der Cauchy-Integralformel, so ergibt sich

$$|r_n(z)| \leq \frac{1}{2\pi} \int_{|\zeta|=\rho^{-1}} \frac{1}{C|z|} \rho^{n+1} |d\zeta| = \frac{1}{2\pi} \frac{1}{C|z|} 2\pi \rho^n = \frac{1}{C|z|} \rho^n \leq C' \frac{\rho^n}{1+|z|}.$$

(ii) $|\arg z| = \pi - \phi$, z nahe 0 ($\operatorname{Re} z \leq 0$):

Dieser Fall ist ein wenig komplizierter, da hier die Singularität im Integranden auftritt. Wie oben erwähnt ist für konsistente Verfahren $\delta(1) = 0$. Ferner ist 1 eine einfache Nullstelle von δ , also $\delta'(1) \neq 0$. Die Funktion δ ist also in einer Umgebung von $\zeta = 1$ lokal umkehrbar, d. h. es gibt genau ein $\zeta^*(z)$ mit $\delta(\zeta^*(z)) = z$, wobei $\zeta^*(0) = 1$. Zudem hat BDF(k) die Ordnung k , was äquivalent ist zu

$$\delta(e^{-z}) = z + O(z^{k+1}), \quad z \rightarrow 0.$$

Fassen wir beide Relationen zusammen, so ergibt sich

$$\zeta^*(z) = e^{-z} + O(z^{k+1}), \quad z \rightarrow 0$$

(Übung). Wir schreiben jetzt in Analogie zur Stabilitätsfunktion bei Runge-Kutta-Verfahren $R(z) = \frac{1}{\zeta^*(z)}$. Dann ist $R(z) = e^z + O(z^{k+1})$. Mit diesen Bezeichnungen gilt

$$\frac{1}{\delta(\zeta) - z} = \frac{\zeta - \zeta^*(z)}{\delta(\zeta) - z} \frac{1}{\zeta - \zeta^*(z)} = \sum_{n=0}^{\infty} \beta_n(z) \zeta^n \frac{1}{\zeta - \zeta^*(z)}. \quad (14.29)$$

Der erste Faktor ist analytisch in $|\zeta| \leq \rho^{-1}$ und in z gleichmäßig beschränkt. Also muss nach dem Majorantenkriterium $|\beta_n(z)| \leq C\rho^n$ gelten. Den zweiten Faktor schreiben wir als

Neumann'sche Reihe:

$$\frac{1}{\zeta - \zeta^*(z)} = -\frac{R(z)}{1 - R(z)\zeta} = -\sum_{n=0}^{\infty} R(z)^{n+1} \zeta^n.$$

Wegen $R(z) = e^z + O(z^{k+1})$ gilt $|R(z)| \leq |e^{z/2}|$ für $|z|$ hinreichend klein und $\operatorname{Re} z \leq 0$. Wir wählen nun r so klein, dass $\rho e^{r/2} \leq \tilde{\rho} < 1$. Dann ist nach (14.29) und (14.26)

$$\begin{aligned} |r_n(z)| &= \left| \sum_{j=0}^n \beta_j(z) R(z)^{n-j+1} \right| \leq C \sum_{j=0}^n \rho^j |e^{(n-j)z/2}| \\ &\leq C |e^{nz/2}| \sum_{j=0}^{\infty} \underbrace{|\rho e^{-z/2}|^j}_{\leq \tilde{\rho} < 1} \\ &\leq C' |e^{nz/2}|. \end{aligned}$$

(iii) z nahe 0, $|z| \leq r$:

Diesen Fall benötigen wir, da Γ_r nicht genau einen Sektor berandet, sondern in der Nähe von Null auch in der rechten Halbebene verläuft. Ähnlich wie in (ii) zeigt man hier $|r_n(z)| \leq e^{2n|z|}$ für $|z|$ hinreichend klein (Übung). \square

Wir wenden nun BDF-Verfahren auf das parabolische Anfangs-Randwertproblem

$$(u', v)_H + a(u, v) = \langle f(t), v \rangle \quad \text{für alle } v \in V,$$

wobei V im Folgenden als separabel vorausgesetzt wird, beziehungsweise

$$u' + Au = f \quad \text{in } V'$$

an:

$$\frac{1}{\tau} \sum_{j=0}^k \delta_j u_{n-j} + Au_n = f(t_n), \quad n \geq k \quad (14.30)$$

mit gegebenen Startwerten u_0, \dots, u_{k-1} . Das Ziel ist, Stabilität und Konvergenz dieser Zeitdiskretisierung zu zeigen.

Im Folgenden setzen wir für in (14.10) definierten Winkel $0 < \phi < \theta$ voraus, wobei θ der Winkel ist, für den das BDF(k)-Verfahren $A(\theta)$ -stabil ist.

Wie in den letzten Abschnitten beginnen wir mit Stabilitätsungleichungen.

Lemma 14.14. (Stabilitätsungleichung bezüglich der V -Norm)

Es seien $u_0, \dots, u_{k-1} \in V$ gegebene Startwerte und $f \in V'$. Dann gilt für die Lösung von (14.30)

$$\left(\tau \sum_{j=k}^n \|u_j\|_V^2 \right)^{1/2} \leq C \left(\sum_{j=0}^{k-1} \|u_j\|_V + \left(\tau \sum_{j=k}^n \|f(t_j)\|_{V'}^2 \right)^{1/2} \right),$$

wobei C unabhängig von n und τ ist mit $n\tau \leq T$.

Beweis. Die Technik der Energieabschätzung lässt sich leider nicht auf BDF-Verfahren anwenden, daher verwenden wir wie bei der Testgleichung erzeugende Funktionen. Wir zeigen die Ungleichung für $n = N$ beliebig aber fest gewählt mit $N\tau \leq T$. Da die Werte $f(t_j)$ für

$j > N$ keinen Einfluss auf die Lösung haben, ersetzen wir die rechte Seite von (14.30) durch f_j mit

$$f_j = \begin{cases} f(t_j), & j \leq N, \\ 0 & j > N. \end{cases}$$

Um die erzeugenden Funktionen zu definieren, multiplizieren wir (14.30) mit ζ^n , wobei $\zeta \in \mathbb{C}$ mit $|\zeta| < 1$ und summieren von $n = k$ bis unendlich. Mit der obigen Notation erhalten wir zunächst formal

$$\frac{1}{\tau} \delta(\zeta) u(\zeta) + A u(\zeta) = f(\zeta) + \frac{1}{\tau} u^0(\zeta) = \sum_{n=k}^{2k-1} u_n^0 \zeta^n, \quad (14.31)$$

wobei u_n^0 wieder nur von den Startwerten abhängt.

Da BDF-Verfahren $A(\theta)$ -stabil sind, liegt $\delta(\zeta)$ für $|\zeta| \leq 1$ in einem Sektor. Nach Lemma 14.6 gibt es daher für $|\zeta| \leq 1$ eine eindeutige Lösung $u(\zeta)$ von (14.31):

$$u(\zeta) = \left(\frac{\delta(\zeta)}{\tau} + A \right)^{-1} \left(f(\zeta) + \frac{1}{\tau} u^0(\zeta) \right).$$

Wie beim Beweis der Existenz schwacher Lösungen in Abschnitt 14.1 betrachten wir getrennt das homogene Problem mit den vorgegebenen Startwerten und das inhomogene Problem mit homogenen Startwerten. Wir beginnen mit letzterem und setzen $u_0 = \dots = u_{k-1} = 0$, also $u^0(\zeta) = 0$. Lemma 14.6 liefert die Abschätzung

$$\|u(\zeta)\|_V \leq C \|f(\zeta)\|_{V'}, \quad C = \frac{1}{\alpha \sin(\theta - \phi)},$$

wobei ϕ der durch $|\arg a(v, v)| \leq \phi$ für $\|v\|_V = 1$ bestimmte Winkel ist (vgl. Abb. 14.2). Damit ist die Abbildung $\zeta \mapsto u(\zeta) = \sum_{n=k}^{\infty} u_n \zeta^n \in V$ analytisch in $|\zeta| < 1$ und stetig auf $|\zeta| \leq 1$. Die Koeffizienten u_n erfüllen (14.30). Die Koeffizienten einer analytischen Funktion können mit Hilfe der Cauchy-Integralformel durch (14.25) ausgedrückt werden. Da $\zeta^{-(n+1)} u(\zeta)$ für $n < k$ analytisch ist, folgt $u_n = 0$ für $n < k$ in Übereinstimmung mit der in diesem Teilbeweis getroffenen Annahme. Aus der Parseval'schen Gleichung (Übung) folgt für $\hat{u}(\omega) = u(e^{i\omega})$ und $\hat{f}(\omega) = f(e^{i\omega})$

$$\sum_{n=k}^{\infty} \|u_n\|_V^2 = \frac{1}{2\pi} \int_0^{2\pi} \|\hat{u}(\omega)\|_V^2 d\omega \leq \frac{C}{2\pi} \int_0^{2\pi} \|\hat{f}(\omega)\|_{V'}^2 d\omega = C \sum_{n=k}^N \|f_n\|_{V'}^2,$$

und daraus

$$\sum_{n=k}^N \|u_n\|_V^2 \leq C \sum_{n=k}^N \|f_n\|_{V'}^2.$$

Im zweiten Schritt betrachten wir jetzt das homogene Problem, also $f_n = 0$ für alle n , mit beliebigen Startwerten $u_0, \dots, u_{k-1} \in V$. Für die Lösung

$$u(\zeta) = \left(\frac{\delta(\zeta)}{\tau} + A \right)^{-1} \frac{1}{\tau} u^0(\zeta)$$

gilt wegen (14.22)

$$u(\zeta) = \frac{\delta(\zeta)}{\tau} \left(\frac{\delta(\zeta)}{\tau} + A \right)^{-1} \frac{1}{\mu(\zeta)} \frac{u^0(\zeta)}{1 - \zeta}, \quad (14.32)$$

wobei $\mu(\zeta) \neq 0$ für $|\zeta| \leq 1$ gilt. Wir betrachten jetzt das formale Cauchy-Produkt der Reihen von u^0 und $(1 - \zeta)^{-1}$:

$$\frac{u^0(\zeta)}{1 - \zeta} = \sum_{n=k}^{\infty} \sum_{j=k}^{\min\{n, 2k-1\}} u_j^0 \zeta^n =: \sum_{n=k}^{\infty} c_n \zeta^n. \quad (14.33)$$

Für die Koeffizienten c_n gilt

$$\|c_n\|_V \leq C_0 \sum_{j=0}^{k-1} \|u_j\|_V, \quad n = k, k+1, \dots$$

Setzen wir (14.33) in (14.32) ein, so kann u durch die formale Reihe

$$u(\zeta) = \frac{\delta(\zeta)}{\tau} \left(\frac{\delta(\zeta)}{\tau} + A \right)^{-1} \frac{1}{\mu(\zeta)} \sum_{n=k}^{\infty} c_n \zeta^n$$

dargestellt werden. Für die Abschätzung ersetzen wir die unendliche Reihe durch eine endliche, indem wir die Hilfsfunktion

$$w(\zeta) = \frac{\delta(\zeta)}{\tau} \left(\frac{\delta(\zeta)}{\tau} + A \right)^{-1} \frac{1}{\mu(\zeta)} \sum_{n=k}^N c_n \zeta^n =: \sum_{n=k}^{\infty} w_n \zeta^n \quad (14.34)$$

definieren. Durch Koeffizientenvergleich sieht man, dass

$$u_n = w_n, \quad n = k, \dots, N$$

gilt.

Ferner gilt für $\lambda = \delta(\zeta)/\tau$ die Abschätzung

$$\|(\lambda + A)^{-1}v\|_V \leq \frac{C}{|\lambda|} \|v\|_V,$$

denn aus der Stetigkeit von $A : V \rightarrow V'$ und Lemma 14.6 folgt

$$\begin{aligned} \|\lambda(\lambda + A)^{-1}v\|_V &= \|[(\lambda + A)^{-1}(\lambda + A) - (\lambda + A)^{-1}A]v\|_V \\ &\leq \|v\|_V + C\|Av\|_{V'} \\ &\leq C'\|v\|_V. \end{aligned}$$

Mit $\mu(\zeta) \geq C_\mu$ für $|\zeta| \leq 1$ erhalten wir damit aus (14.34)

$$\|w(\zeta)\|_V = \frac{C'}{C_\mu} \left\| \sum_{n=k}^N c_n \zeta^n \right\|_V.$$

Mit Hilfe der Parseval'schen Gleichung ergibt sich schließlich

$$\begin{aligned} \sum_{n=k}^N \|u_n\|_V^2 &\leq \sum_{n=k}^{\infty} \|w_n\|_V^2 = \frac{1}{2\pi} \int_0^{2\pi} \|w(e^{i\theta})\|_V^2 d\omega \\ &\leq \frac{1}{2\pi} \frac{C'}{C_\mu} \int_0^{2\pi} \left\| \sum_{n=k}^N c_n e^{in\omega} \right\|_V^2 d\theta \\ &= \frac{1}{2\pi} \frac{C'}{C_\mu} \sum_{n=k}^N \|c_n\|_V^2 \\ &\leq \frac{N}{2\pi} \frac{C' C_0}{C_\mu} \left(\sum_{j=0}^{k-1} \|u_j\|_V \right)^2. \end{aligned}$$

Damit haben wir

$$\tau \sum_{n=k}^N \|u_n\|_V^2 \leq C(N\tau) \left(\sum_{j=0}^{k-1} \|u_j\|_V \right)^2$$

gezeigt. Hierbei beachte man $N\tau \leq T$, die Konstante hängt damit nur von T ab.

Der allgemeine Fall folgt nun durch Superposition aus den beiden Spezialfällen. \square

Lemma 14.15. (*Stabilitätsungleichung bezüglich der H -Norm*)

Es seien $u_0, \dots, u_{k-1} \in V$ gegebene Startwerte und $f \in H'$. Dann gilt für die Lösung von (14.30)

$$\|u_n\|_H \leq C \left(\sum_{j=0}^{k-1} \|u_j\|_H + \tau \sum_{j=k}^n \|f(t_j)\|_H \right),$$

wobei C unabhängig von n und $\tau \leq \tau_0$ ist.

Beweis. In der Notation des Beweises von Lemma 14.14 ist

$$u(\zeta) = \left(\frac{\delta(\zeta)}{\tau} + A \right)^{-1} \left(f(\zeta) + \frac{1}{\tau} u^0(\zeta) \right).$$

Mit den Koeffizienten $r_n(-\tau A) : H \rightarrow H$ definiert in (14.26) ist

$$u_n = \tau \sum_{j=k}^n r_{n-j}(-\tau A) \left(f(t_j) + \frac{1}{\tau} u_j^0 \right), \quad n \geq k.$$

Um u_n mit der H -Norm abzuschätzen, benötigen wir also Schranken für die Operatornorm von $r_n(-\tau A)$:

$$\|r_n(-\tau A)\|_{H \leftarrow H} = \sup_{\|v\|_H=1} \|r_n(-\tau A)v\|_H.$$

Im folgenden Lemma 14.16 zeigen wir $\|r_n(-\tau A)\|_{H \leftarrow H} \leq C$. Damit gilt

$$\begin{aligned} \|u_n\|_H &\leq \tau \sum_{j=k}^n C (\|f(t_j)\|_H + \frac{1}{\tau} \|u_j^0\|_H) \\ &\leq C\tau \sum_{j=k}^n \|f(t_j)\|_H + C \sum_{j=k}^{2k-1} \|u_j^0\|_H \\ &\leq C\tau \sum_{j=k}^n \|f(t_j)\|_H + C' \sum_{j=0}^{k-1} \|u_j\|_H, \end{aligned}$$

denn u_j^0 hängt nur von u_0, \dots, u_{k-1} ab. \square

Im Beweis haben wir folgendes Lemma verwendet:

Lemma 14.16. $\|r_n(-\tau A)\|_{H \leftarrow H} \leq C$ mit C unabhängig von n und τ .

Beweis. Mit Hilfe der Cauchy-Integralformel führen wir das Problem auf den in Lemma 14.13 behandelten skalaren Fall zurück.

(a) Durch zweimalige Anwendung der Cauchy-Integralformel erhält man aus (14.28)

$$\begin{aligned}
 \tau r_n(-\tau A) &= \frac{1}{2\pi i} \int_{|\zeta|=1} \zeta^{-(n+1)} \left(\frac{\delta(\zeta)}{\tau} + A \right)^{-1} d\zeta \\
 &= \frac{1}{2\pi i} \int_{|\zeta|=1} \zeta^{-(n+1)} \left[\frac{1}{2\pi i} \int_{\Gamma} (\lambda + A)^{-1} \frac{1}{\frac{\delta(\zeta)}{\tau} - \lambda} d\lambda \right] d\zeta \\
 &= \frac{1}{2\pi i} \int_{\Gamma} (\lambda + A)^{-1} \left[\frac{1}{2\pi i} \int_{|\zeta|=1} \zeta^{-(n+1)} \frac{1}{\frac{\delta(\zeta)}{\tau} - \lambda} d\zeta \right] d\lambda \\
 &= \frac{1}{2\pi i} \int_{\Gamma} (\lambda + A)^{-1} \tau r_n(\tau \lambda) d\lambda.
 \end{aligned}$$

Hierbei ist Γ eine negativ orientierte Kurve um $\frac{\delta(\zeta)}{\tau}$ (man beachte das Vorzeichen in der Integralformel). Der Beitrag des Integrals für λ weit in der positiven rechten Halbebene geht im Grenzwert gegen Null wegen des Abfalls wie $|\lambda|^{-2}$ (nach Lemma 14.6 und da $\delta(\zeta)$ als Polynom für $|\zeta| \leq 1$ beschränkt ist). Daher wählen wir die Kurve $\Gamma = \Gamma_r$ wie in Abbildung 14.5, jedoch dieses Mal negativ orientiert, denn die Kurve soll in der rechten Halbebene geschlossen werden.

(b) Wir setzen die Resultate aus (a) und Lemma 14.13 zusammen. Mit Lemma 14.6 gilt für $\rho^n \leq C'/n$.

$$\begin{aligned}
 \|r_n(-\tau A)\|_{H \leftarrow H} &\leq \frac{1}{2\pi} \int_{\Gamma_{1/(n\tau)}} \|(\lambda + A)^{-1}\|_{H \leftarrow H} \cdot |r_n(\tau \lambda)| |d\lambda| \\
 &\leq \frac{C}{2\pi} \int_{\Gamma_{1/(n\tau)}} \frac{1}{|\lambda|} \cdot \left(|e^{n\tau\lambda/2}| + \frac{\rho^n}{1 + |\tau\lambda|} \right) |d\lambda| + \frac{C}{2\pi} \int_{|\zeta|=1/n\tau} \frac{e^{2|n\tau\lambda|}}{|\lambda|} |d\lambda| \\
 &\leq \frac{C}{2\pi} \int_{\Gamma_{1/(n\tau)}} \frac{|e^{n\tau\lambda/2}|}{|n\tau\lambda|} |d\lambda n\tau| + \frac{C}{2\pi} \int_{\Gamma_{1/(n\tau)}} \frac{1}{|n\tau\lambda|} \cdot \underbrace{\frac{\rho^n}{1 + \left| \frac{n\tau\lambda}{n} \right|}}_{\frac{C'}{n + |n\tau\lambda|}} |d\lambda n\tau| \\
 &\quad + \frac{C}{2\pi} \int_{|\zeta|=1/n\tau} \frac{e^{2|n\tau\lambda|}}{|n\tau\lambda|} |d\lambda n\tau| \\
 &\leq \frac{C}{2\pi} \int_{\Gamma_1} \frac{|e^{\mu/2}|}{|\mu|} |d\mu| + \frac{C}{2\pi} \int_{\Gamma_1} \frac{1}{\mu} \cdot \frac{C'}{1 + |\mu|} |d\mu| + \frac{C}{2\pi} \int_{|\mu|=1} \frac{e^{2|\mu|}}{|\mu|} |d\mu| \\
 &\leq \text{const.}
 \end{aligned}$$

Hierbei haben wir für die zweite Abschätzung die Kurve Γ_r in die beiden Geradenstücke und den Kreisbogen zerlegt und den Kreisbogen dann zu einem Kreis erweitert. \square

Satz 14.17. (Konvergenz der BDF-Zeitdiskretisierung)

Es seien u_n die Näherungen des BDF(k)-Verfahrens für $k \leq 6$, deren Startwerte $\|u_j - u(t_j)\|_V = O(\tau^k)$ für $j = 0, \dots, k-1$ erfüllen. Es sei θ der Winkel, für den das BDF(k)-Verfahren $A(\theta)$ -stabil ist und es gelte für den in (14.10) definierten Winkel $0 < \phi < \theta$. Die Lösung des parabolischen Anfangs-Randwertproblems sei genügend oft differenzierbar. Dann gilt

$$\|u_n - u(t_n)\|_H + \left(\tau \sum_{j=k}^n \|u_j - u(t_j)\|_V^2 \right)^{1/2} \leq C\tau^k.$$

Beweis. Eingesetzt in die Iterationsvorschrift des BDF-Verfahrens gilt für die exakte Lösung für $n \geq k$.

$$\frac{1}{\tau} \sum_{j=0}^k \delta_j u(t_{n-j}) + Au(t_n) = f(t_n) + \left[\frac{1}{\tau} \sum_{j=0}^k \delta_j u(t_{n-j}) - u'(t_n) \right] =: f(t_n) + d_n.$$

Der Defekt d_n erfüllt $\|d_n\|_V = O(\tau^k)$.

Subtraktion von Gleichung (14.30) für die BDF-Näherung liefert für den Fehler $e_n = u_n - u(t_n)$:

$$\frac{1}{\tau} \sum_{j=0}^k \delta_j e_{n-j} + Ae_n = -d_n, \quad n \geq k.$$

Das Konvergenzresultat folgt jetzt aus den Lemmas 14.14 und 14.15. \square

14.5 Die Burgers-Gleichung

Bisher haben wir uns auf lineare parabolische Differentialgleichungen beschränkt. In diesem Abschnitt betrachten wir die Burgers-Gleichung als Beispiel einer nichtlinearen parabolischen Differentialgleichung. Dieses Beispiel dient als Motivation für allgemeinere nichtlineare parabolische Differentialgleichungen.

Gesucht ist eine Funktion $u = u(x, t)$, $u : [0, 1] \times [0, T] \rightarrow \mathbb{R}$ mit

$$\begin{array}{lll} u_t = u_{xx} - uu_x & 0 < x < 1, \quad 0 < t \leq T & \\ u(0, t) = u(1, t) = 0 & 0 < t \leq T & \text{Randbedingungen} \\ u(x, 0) = u_0(x) & 0 < x < 1 & \text{Anfangsbedingungen} \end{array}$$

Diese eindimensionale Burgers-Gleichung diskretisieren wir im Ort mit finiten Differenzen:

$$\begin{aligned} u_{xx}(x, t) &\approx \frac{1}{h^2} (u(x+h, t) - 2u(x, t) + u(x-h, t)) \\ (uu_x)(x, t) &\approx \frac{1}{2h} u(x, t) (u(x+h, t) - u(x-h, t)). \end{aligned}$$

Alternativ kann man auch

$$(uu_x)(x, t) = \frac{1}{2} \frac{\partial}{\partial x} (u^2(x, t)) \approx \frac{1}{4h} (u^2(x+h, t) - u^2(x-h, t)).$$

verwenden. Wir setzen jetzt

$$y_n(t) = u(x_n, t), \quad x_n = nh, \quad h = \frac{1}{N+1}, \quad n = 0, 1, \dots, N+1,$$

wobei wegen der Randbedingungen $y_0(t) = y_{N+1}(t) = 0$ gesetzt wird. Da bezüglich des Ortes diskretisiert wird und die Zeit t kontinuierlich behandelt wird, spricht man von einer *Linienmethode*, denn in der (x, t) -Ebene wird das Ausgangsproblem auf Zeitlinien gelöst. Die Linienmethode führt auf ein System gewöhnlicher Differentialgleichungen

$$y'_n = \frac{1}{h^2} (y_{n+1} - 2y_n + y_{n-1}) - \frac{1}{2h} y_n (y_{n+1} - y_{n-1}),$$

welches wir später mit BDF-Verfahren lösen werden.

Fassen wir die einzelnen Funktionen y_n in einem Vektor $y = [y_1 \ \cdots \ y_N]^T$ und entsprechend $g(y) = [g(y)_1 \ \cdots \ g(y)_N]^T$ mit $g(y)_n = -\frac{1}{2h}y_n(y_{n+1} - y_{n-1})$ zusammen, so können wir das obige System kompakt als

$$y' + Ay = g(y), \quad A = \frac{1}{h^2} \text{tridiag}(-1, 2, -1), \quad (14.35)$$

schreiben. Die Matrix A kennen wir schon aus früheren Vorlesungen. Sie ist symmetrisch positiv definit und hat Eigenwerte $0 < \lambda_1 < \cdots < \lambda_N \approx 4h^{-2}$.

Analog zur L^2 -Norm definieren wir die diskrete L^2 -Norm als

$$\|y\|_{H,h}^2 = h \sum_{n=1}^N y_n^2 \quad \left(\approx \int_0^1 y(x)^2 dx = \|y\|_0^2 \right).$$

Die diskrete L^2 -Norm wird durch das diskrete L^2 -Skalarprodukt

$$(y, z)_{H,h} = h \sum_{n=1}^N y_n z_n$$

induziert.

Die Schwierigkeit besteht darin, dass die Nichtlinearität g bezüglich dieser Norm nicht gleichmäßig Lipschitz-stetig ist. Definieren wir

$$Dy = \left(\frac{1}{2h}(y_{n+1} - y_{n-1}) \right)_{n=1}^N,$$

so ist $g(y)_n = -y_n(Dy)_n$. Wir betrachten zunächst den kontinuierlichen Fall. Hier gilt

$$\begin{aligned} \|uu_x - vv_x\|_0 &= \|u(u_x - v_x) + (u - v)v_x\|_0 \\ &\leq \|u(u_x - v_x)\|_0 + \|(u - v)v_x\|_0 \\ &\leq \max_x |u(x)| \cdot \|u_x - v_x\|_0 + \max_x |u(x) - v(x)| \cdot \|v_x\|_0. \end{aligned}$$

Aus der Cauchy-Schwarz'schen Ungleichung folgt außerdem

$$|u(x)| = \left| \int_0^x u_x(y) \cdot 1 dy \right| \leq \|u_x\|_0, \quad 0 \leq x \leq 1.$$

(Dies ist der Sobolev'sche Einbettungssatz in einer Dimension.) Oben eingesetzt erhalten wir die Abschätzung

$$\|uu_x - vv_x\|_0 \leq \|u_x - v_x\|_0 (\|u_x\|_0 + \|v_x\|_0).$$

Das diskrete Analogon dazu ist

Lemma 14.18. *Ist N gerade, so gilt für $g(u)_n = -u_n(Du)_n$*

$$\|g(u) - g(v)\|_{H,h} \leq \sqrt{2} \|Du - Dv\|_{H,h} (\|Du\|_{H,h} + \|Dv\|_{H,h}).$$

Beweis. Es gilt

$$\begin{aligned} \|g(u) - g(v)\|_{H,h} &\leq \|(u_n(Du - Dv)_n)_{n=1}^N\|_{H,h} + \|(u_n - v_n)(Dv)_n\|_{H,h} \\ &\leq \max_n |u_n| \|Du - Dv\|_{H,h} + \max_n |u_n - v_n| \|Dv\|_{H,h}. \end{aligned}$$

Ferner ist wegen $u_0 = u_{N+1} = 0$ (Randbedingung) $u_n = u_n - u_0 = -(u_{N+1} - u_n)$ oder, als Teleskopsumme geschrieben,

$$u_n = \begin{cases} 2h \sum_{\substack{j \text{ ungerade} \\ 1 \leq j \leq n-1}} \frac{u_{j+1} - u_{j-1}}{2h} \cdot 1 & n \text{ gerade,} \\ -2h \sum_{\substack{j \text{ gerade} \\ n+1 \leq j \leq N}} \frac{u_{j+1} - u_{j-1}}{2h} \cdot 1 & n \text{ ungerade,} \end{cases}$$

wobei wir im zweiten Fall die Voraussetzung N gerade benötigt haben. Die Cauchy-Schwarz'sche Ungleichung liefert auch hier $|u_n| \leq \sqrt{2} \|Du\|_{H,h}$ für alle n und daraus folgt die Behauptung. \square

Wir möchten jetzt $\|Du\|_{H,h}$ durch eine mit A zusammenhängende Norm abschätzen. Da A symmetrisch und positiv definit ist, gibt es eine orthogonale Matrix Q , so dass $A = Q\Lambda Q^T$, wobei $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$ positive Diagonaleinträge hat. Daher existiert $A^\alpha = Q\Lambda^\alpha Q^T$ für alle $\alpha \in \mathbb{R}$ und ist ebenfalls symmetrisch positiv definit. Insbesondere existiert $A^{\frac{1}{2}}$ und es gilt

$$\|u\|_{V,h}^2 := \|A^{\frac{1}{2}}u\|_{H,h}^2 = (A^{\frac{1}{2}}u, A^{\frac{1}{2}}u)_{H,h} = (Au, u)_{H,h}.$$

Die Norm $\|\cdot\|_{V,h}$ wird auch als *Energienorm* bezeichnet. Sie entspricht der H_0^1 -Norm im kontinuierlichen Fall.

Lemma 14.19. *Es gilt $\|Du\|_{H,h} \leq \|u\|_{V,h} = \|A^{\frac{1}{2}}u\|_{H,h}$.*

Beweis. Nach Definition von D und A ist

$$\begin{aligned} \|Du\|_{H,h}^2 &= h \sum_{n=1}^N \left(\frac{u_{n+1} - u_{n-1}}{2h} \right)^2, \\ (Au, u)_{H,h} &= h \sum_{n=1}^N u_n \left(\frac{-u_{n+1} + 2u_n - u_{n-1}}{h^2} \right). \end{aligned}$$

Außerdem folgt aus

$$\begin{aligned} 2u_n(-u_{n+1} + 2u_n - u_{n-1}) &= -2u_n u_{n+1} + 4u_n^2 - 2u_n u_{n-1} \\ &= (u_{n+1} - u_n)^2 - u_{n+1}^2 + 2u_n^2 - u_{n-1}^2 + (u_n - u_{n-1})^2 \end{aligned}$$

durch Aufsummieren

$$\begin{aligned} \sum_{n=1}^N (u_{n+1} - u_{n-1})^2 &= \sum_{n=1}^N ((u_{n+1} - u_n) + (u_n - u_{n-1}))^2 \\ &\leq 2 \sum_{n=1}^N ((u_{n+1} - u_n)^2 + (u_n - u_{n-1})^2) \\ &\leq 2 \sum_{n=1}^N ((u_{n+1} - u_n)^2 + (u_n - u_{n-1})^2 - u_{n+1}^2 + 2u_n^2 - u_{n-1}^2) \\ &= 4 \sum_{n=1}^N u_n(-u_{n+1} + 2u_n - u_{n-1}), \end{aligned}$$

wobei wir für die dritte Ungleichung

$$\sum_{n=1}^N (-u_{n+1}^2 + 2u_n^2 - u_{n-1}^2) = \sum_{n=1}^N ((u_n^2 - u_{n+1}^2) + (u_n^2 - u_{n-1}^2)) = u_1^2 + u_N^2 \geq 0$$

verwendet haben. \square

Als Folgerung aus diesen Lemmas gilt

$$\|g(u) - g(v)\|_{H,h} \leq L(r) \|u - v\|_{V,h} \quad \text{für} \quad \|u\|_{V,h} + \|v\|_{V,h} \leq r. \quad (14.36)$$

Wir betrachten jetzt die gestörte Gleichung

$$\tilde{y}' + A\tilde{y} = g(\tilde{y}) + d(t), \quad \tilde{y}(0) = \tilde{y}_0. \quad (14.37)$$

Ein Beispiel wäre $\tilde{y}(t) = (u(nh, t))_{n=1}^N$. Falls $\tilde{y} \in C^2$, so wäre der Defekt d durch $\|d(t)\|_{H,h} = O(h^2)$ beschränkt.

Bevor wir jedoch ein Stabilitätsresultat für die Raumdiskretisierung angeben, zeigen wir zwei Lemmas, die später mehrfach verwendet werden.

Lemma 14.20. (*Lemma von Gronwall*)

Erfüllt die Funktion $f : [0, T] \rightarrow \mathbb{R}$ für $M \geq 0$, $L > 0$ die Ungleichung

$$0 \leq f(t) \leq M + L \int_0^t f(s) ds, \quad 0 \leq t \leq T,$$

so gilt $f(t) \leq Me^{Lt}$.

Beweis. Wir nehmen ohne Einschränkung $M > 0$ an, denn falls die Ungleichung für $M = 0$ gilt, dann erst recht für jedes $M > 0$. Die Behauptung folgt dann durch Grenzübergang $M \rightarrow 0$.

Die Funktion $\varphi(t) = M + L \int_0^t f(s) ds$ erfüllt $\varphi'(t) = Lf(t)$. Nach Voraussetzung gilt dann

$$\frac{d}{dt} \log \varphi(t) = \frac{\varphi'(t)}{\varphi(t)} = \frac{Lf(t)}{M + L \int_0^t f(s) ds} \leq L$$

und es folgt durch Integration

$$\log \varphi(t) - \log \varphi(0) \leq Lt,$$

also $f(t) \leq \varphi(t) \leq \varphi(0)e^{Lt} = Me^{Lt}$. \square

Lemma 14.21. (*Verallgemeinertes Lemma von Gronwall*)

Erfüllt die Funktion $f : [0, T] \rightarrow \mathbb{R}$ für $M \geq 0$, $L > 0$ und $\mu > 0$ die Ungleichung

$$0 \leq f(t) \leq M + L \int_0^t (t-s)^{\mu-1} f(s) ds, \quad 0 \leq t \leq T,$$

so gilt $f(t) \leq C(L, T, \mu)M$.

Beweis. Um diese Verallgemeinerung auf das Gronwall-Lemma zurückzuführen, zeigen wir zunächst mit Hilfe der Laplace-Transformation folgende Gleichung für die m -fache Faltung mit der Funktion $t^{\mu-1}$:

$$\frac{t^{\mu-1}}{\Gamma(\mu)} * \frac{t^{\mu-1}}{\Gamma(\mu)} * \dots * \frac{t^{\mu-1}}{\Gamma(\mu)} = \frac{t^{m\mu-1}}{\Gamma(m\mu)}.$$

Für die Laplace-Transformation

$$\mathcal{L}(f) = \int_0^\infty e^{-st} f(t) dt$$

gilt bekanntlich die Faltungsregel

$$\mathcal{L}(f * g) = \int_0^\infty e^{-st} \int_0^t f(\sigma) g(t - \sigma) d\sigma dt = \mathcal{L}(f) \mathcal{L}(g).$$

Wegen

$$\begin{aligned} \mathcal{L}\left(\frac{t^{\mu-1}}{\Gamma(\mu)}\right) &= \int_0^\infty e^{-st} \frac{t^{\mu-1}}{\Gamma(\mu)} dt & r = st \\ &= \int_0^\infty e^{-r} \left(\frac{r}{s}\right)^{\mu-1} \frac{1}{s\Gamma(\mu)} dr \\ &= \frac{1}{s^\mu \Gamma(\mu)} \int_0^\infty e^{-r} r^{\mu-1} dr \\ &= \frac{1}{s^\mu} \end{aligned}$$

ist dann

$$\left(\mathcal{L}\left(\frac{t^{\mu-1}}{\Gamma(\mu)}\right)\right)^m = \frac{1}{s^{m\mu}}$$

Anwendung der inversen Laplace-Transformation liefert dann

$$\mathcal{L}^{-1}\left(\mathcal{L}\left(\frac{t^{\mu-1}}{\Gamma(\mu)}\right)\right)^m = \frac{t^{\mu-1}}{\Gamma(\mu)} * \frac{t^{\mu-1}}{\Gamma(\mu)} * \dots * \frac{t^{\mu-1}}{\Gamma(\mu)} = \mathcal{L}^{-1}(s^{-m\mu}) = \frac{t^{m\mu-1}}{\Gamma(m\mu)}.$$

Zum Beweis der eigentlichen Aussage schreiben wir die Voraussetzung als Faltung:

$$0 \leq f(t) \leq M + L t^{\mu-1} * f.$$

Durch m -faches Einsetzen dieser Voraussetzung erhalten wir

$$\begin{aligned} f(t) &\leq M + L t^{\mu-1} * (M + L t^{\mu-1} * f) \\ &\leq \dots \\ &\leq C_1(T, L, \mu) M + L^m t^{\mu-1} * \dots * t^{\mu-1} * f. \end{aligned}$$

Wir wählen m so groß, dass $m\mu - 1 \geq 0$, denn dann gilt

$$t^{m\mu-1} * f \leq T^{m\mu-1} * f = T^{m\mu-1} \int_0^t f(s) ds.$$

Insgesamt gilt

$$0 \leq f(t) \leq C_1(T, L, \mu) M + C_2(T, L, \mu) \int_0^t f(s) ds$$

und das Gronwall-Lemma liefert $f(t) \leq C_1 M e^{C_2 t} \leq C M$. □

Für das Konvergenzresultat benötigen wir noch weitere Abschätzungen für die BDF-Koeffizienten r_n .

Lemma 14.22. *Es sei $A \in \mathbb{R}^{N,N}$ eine symmetrische und positiv definite Matrix und $0 \leq \alpha < 1$. Dann gilt für die zur Euklid-Norm passende Matrixnorm $\|\cdot\|$:*

- (a) Für $|\arg \lambda| \leq \pi - \theta$, $0 < \theta < \pi/2$ gilt $\|(\lambda + A)^{-1} A^\alpha\| \leq M |\lambda|^{\alpha-1}$.
- (b) Es gilt $\|e^{-\tau A} A^\alpha\| \leq C_0 \tau^{-\alpha}$ für $\tau > 0$.
- (c) Es gilt $\|A^\alpha r_n(-\tau A)\| \leq \frac{C}{((n+1)\tau)^\alpha}$ mit C unabhängig von A , n und τ .

Beweis. Da A symmetrisch ist, genügt es die Aussagen für alle Eigenwerte σ von A zu zeigen. Wegen der positiven Definitheit können wir uns auf $\sigma > 0$ beschränken.

(a) Für

$$\Psi(\lambda) = \frac{\sigma^\alpha}{\lambda + \sigma}$$

gilt

$$|\Psi(\lambda)| = |\lambda|^{\alpha-1} \left| \frac{\left(\frac{\lambda}{\sigma}\right)^{1-\alpha}}{\frac{\lambda}{\sigma} + 1} \right| \leq M |\lambda|^{\alpha-1},$$

denn wegen $|\arg(\lambda/\sigma)| \leq \pi - \theta$ ist der zweite Faktor beschränkt.

(b) Nach der Cauchy-Integralformel und Definition (14.26) der Koeffizienten r_n ist

$$\begin{aligned} \Psi\left(\frac{\delta(\zeta)}{\tau}\right) &= \frac{1}{2\pi i} \int_{\Gamma} \left(\frac{\delta(\zeta)}{\tau} - \lambda\right)^{-1} \Psi(\lambda) d\lambda \\ &= \frac{1}{2\pi i} \int_{\Gamma} \tau \sum_{n \geq 0} r_n(\tau \lambda) \zeta^n \Psi(\lambda) d\lambda \\ &= \tau \sum_{n \geq 0} \left(\frac{1}{2\pi i} \int_{\Gamma} r_n(\tau \lambda) \Psi(\lambda) d\lambda \right) \zeta^n \\ &=: \tau \sum_{n \geq 0} \psi_n \zeta^n. \end{aligned}$$

Hierbei ist Γ die Integrationskurve Γ_r aus Abb. 14.5. Definieren wir ψ durch

$$\psi(\tau) = \frac{1}{2\pi i} \int_{\Gamma} e^{\lambda \tau} \Psi(\lambda) d\lambda = \sigma^\alpha e^{-\sigma \tau},$$

so gilt nach (a)

$$\begin{aligned} |\psi(\tau)| &\leq \frac{1}{2\pi} \int_{\Gamma_{\tau^{-1}}} |e^{\lambda \tau}| |\Psi(\lambda)| |d\lambda| \\ &\leq \frac{1}{2\pi} \int_{\Gamma_{\tau^{-1}}} |e^{\lambda \tau}| M |\lambda|^{\alpha-1} |d\lambda| & s = \lambda \tau \\ &= \frac{M}{2\pi} \int_{\Gamma_1} |e^s| |s|^{\alpha-1} |ds| \tau^{-\alpha} \\ &\leq C_0 \tau^{-\alpha}. \end{aligned}$$

(c) Mit Hilfe von (b) und Lemma 14.13 zeigt man $|\psi_n| \leq C/[(n+1)\tau]^\alpha$ (Übung). Daraus folgt wegen

$$\psi_n = \frac{1}{2\pi i} \int_{\Gamma} r_n(\tau\lambda)(\lambda + \sigma)^{-1} \sigma^\alpha d\lambda = \sigma^\alpha r_n(-\tau\sigma)$$

die Behauptung. \square

Bemerkung. $\|A\|$ ist unabhängig von h , denn für die Euklid-Norm $\|\cdot\|$ gilt wegen $\|x\|_{H,h} = \sqrt{h}\|x\|$

$$\|A\|_{H,h} = \sup_{x \neq 0} \frac{\|Ax\|_{H,h}}{\|x\|_{H,h}} = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \|A\|.$$

Satz 14.23. (Stabilität der Raumdiskretisierung)

Ist y eine Lösung von (14.35) und \tilde{y} eine Lösung von (14.37), so gilt

$$\|y(t) - \tilde{y}(t)\|_{V,h} \leq C(\|y_0 - \tilde{y}_0\|_{V,h} + \max_{0 \leq \tau \leq t} \|d(\tau)\|_{H,h}), \quad 0 \leq t \leq T,$$

mit C unabhängig von h .

Beweis. Die Lösungen y und \tilde{y} sind durch die Variation der Konstanten-Formel gegeben:

$$\begin{aligned} y(t) &= e^{-tA}y_0 + \int_0^t e^{-(t-\tau)A}g(y(\tau))d\tau, \\ \tilde{y}(t) &= e^{-tA}\tilde{y}_0 + \int_0^t e^{-(t-\tau)A}(g(\tilde{y}(\tau)) + d(\tau))d\tau. \end{aligned}$$

Subtraktion liefert wegen $\|\cdot\|_{V,h} = \|A^{\frac{1}{2}}\cdot\|_{H,h}$ und (14.36)

$$\begin{aligned} \|y(t) - \tilde{y}(t)\|_{V,h} &\leq \|A^{\frac{1}{2}}e^{-tA}(y_0 - \tilde{y}_0)\|_{H,h} \\ &\quad + \int_0^t \|A^{\frac{1}{2}}e^{-(t-\tau)A}\| L(r) \|y(\tau) - \tilde{y}(\tau)\|_{V,h} d\tau \\ &\quad + \int_0^t \|A^{\frac{1}{2}}e^{-(t-\tau)A}\| \|d(\tau)\|_{H,h} d\tau, \end{aligned}$$

wobei r so groß gewählt wird, dass $\|\tilde{y}(t)\|_{V,h} + \|y(t)\|_{V,h} \leq r$ für $0 \leq t \leq T$.

Nach Lemma 14.22(b) gilt $\|A^{\frac{1}{2}}e^{-tA}\|_{H,h} \leq c/\sqrt{t}$ mit einer Konstanten c unabhängig von h . Daraus ergibt sich wegen $\|e^{-tA}\|_{H,h} \leq 1$

$$\|y(t) - \tilde{y}(t)\|_{V,h} \leq \|y_0 - \tilde{y}_0\|_{V,h} + \int_0^t \frac{C}{\sqrt{t-\tau}} \|y(\tau) - \tilde{y}(\tau)\|_{V,h} d\tau + \max_{0 \leq \tau \leq t} \|d(\tau)\|_{H,h} \int_0^t \frac{C}{\sqrt{t-\tau}} d\tau.$$

Die Stabilitätsaussage folgt jetzt aus dem verallgemeinerten Gronwall-Lemma 14.21. \square

14.6 BDF-Verfahren bei nichtlinearen Problemen

Wir betrachten zunächst die Differentialgleichung

$$y' + Ay = g(y), \quad y(0) = y_0 \quad \text{im } \mathbb{R}^N.$$

Hierbei sei A eine symmetrische und positiv definite Matrix bezüglich eines gewichteten Euklidischen Skalarprodukts $(\cdot, \cdot)_{H,h}$ mit der zugehörigen Norm $\|\cdot\|_{H,h}$. Ferner sei $\|v\|_{V,h} := \|A^\alpha v\|_{H,h}$ mit $0 \leq \alpha < 1$. Für die rechte Seite g fordern wir

$$\|g(y) - g(z)\|_{H,h} \leq L(r)\|y - z\|_{V,h} \quad \text{für} \quad \|y\|_{V,h}, \|z\|_{V,h} \leq r. \quad (14.38)$$

Differentialgleichungen der obigen Form mit (14.38) heißen *semilineare parabolische Differentialgleichungen*.

Ein Beispiel für diese Situation mit $\alpha = 1/2$ haben wir im letzten Abschnitt mit der Burgers-Gleichung schon kennengelernt. Weitere wichtige Beispiele sind Reaktions-Diffusionsgleichungen sowie die Navier-Stokes-Gleichungen.

Die Zeitdiskretisierung mit BDF(k)-Verfahren ($k \leq 6$) liefert

$$\frac{1}{\tau} \sum_{j=0}^k \delta_j y_{n-j} + A y_n = g(y_n), \quad n \geq k$$

mit gegebenen Startwerten y_0, \dots, y_{k-1} . Hierbei resultiert $A = A_h$ von einer Ortsdiskretisierung.

Wir möchten y_n mit der Lösung $\tilde{y}_n(t_n) = u(\cdot, t_n)$ der partiellen Differentialgleichung vergleichen. Diese erfüllt

$$\tilde{y}' + A\tilde{y} = g(\tilde{y}) + \tilde{d}(t), \quad \tilde{y}(0) = \tilde{y}_0,$$

wobei $\tilde{d}(t)$ die aus der Raumdiskretisierung resultierende Störung ist. Die Lösung existiert auf $[0, T]$.

Satz 14.24. *Für die Näherungslösungen des BDF(k)-Verfahrens mit $k \leq 6$ gilt unter obigen Voraussetzungen*

$$\|y_n - \tilde{y}_n\|_{V,h} \leq C \left(\max_{j=0, \dots, k-1} \|y_j - \tilde{y}_j\|_{V,h} + \max_{0 \leq s \leq t_n} \|\tilde{d}(s)\|_{H,h} + \tau^k \max_{0 \leq s \leq t_n} \|\tilde{y}^{(k+1)}(s)\|_{V,h} \right)$$

gleichmäßig für alle τ, n mit $0 \leq n\tau \leq T$, sofern die rechte Seite hinreichend klein ist (wegen der Existenz der numerischen Lösung).

Beweis. Nach Voraussetzung gilt

$$\begin{aligned} \frac{1}{\tau} \sum_{j=0}^k \delta_j y_{n-j} + A y_n &= g(y_n), \\ \frac{1}{\tau} \sum_{j=0}^k \delta_j \tilde{y}(t_{n-j}) + A \tilde{y}(t_n) &= g(\tilde{y}(t_n)) + d_n, \end{aligned}$$

wobei $\|d_n\|_{H,h} \leq C(\|\tilde{d}(t_n)\|_{H,h} + \tau^k \max_{0 \leq s \leq t_n} \|\tilde{y}^{(k+1)}(s)\|_{V,h})$.

Wir setzen $\Delta y_n = y_n - \tilde{y}(t_n)$, $\Delta g_n = g(y_n) - g(\tilde{y}(t_n))$ und subtrahieren:

$$\frac{1}{\tau} \sum_{j=0}^k \delta_j \Delta y_{n-j} + A \Delta y_n = \Delta g_n - d_n.$$

Diese Gleichung multiplizieren wir mit ζ^n und summieren über n :

$$\left(\frac{\delta(\zeta)}{\tau} + A \right) \Delta y(\zeta) = \Delta g(\zeta) - d(\zeta) - \frac{1}{\tau} \Delta y^0(\zeta),$$

wobei $\Delta y^0(\zeta)$ die Anfangsdaten enthält und die Bezeichnungen denen aus Abschnitt 14.4 entsprechen. Mit

$$\left(\frac{\delta(\zeta)}{\tau} + A\right)^{-1} = \tau \sum_{n=0}^{\infty} r_n(-\tau A) \zeta^n$$

erhalten wir

$$\Delta y_n = \tau \sum_{j=k}^n r_{n-j}(-\tau A) (\Delta g_j - d_j - \frac{1}{\tau} \Delta y_j^0),$$

wobei $\Delta y_j^0 = 0$ für $j \geq 2k$. Nach Voraussetzung (14.38) gilt $\|\Delta g_j\|_{H,h} \leq L(r) \|\Delta y_j\|_{V,h}$ sofern r so groß gewählt wird, dass $\|y_j\|_{V,h}, \|\tilde{y}(t_j)\|_{V,h} \leq r$. Damit ist zunächst nach Definition von $\|\cdot\|_{V,h}$ und dann mit Lemma 14.22

$$\begin{aligned} \|\Delta y_n\|_{V,h} &\leq \tau \sum_{j=k}^n \|r_{n-j}(-\tau A) (\Delta g_j + d_j)\|_{V,h} + \sum_{j=k}^{2k-1} \|r_{n-j}(-\tau A) \Delta y_j^0\|_{V,h} \\ &= \tau \sum_{j=k}^n \|r_{n-j}(-\tau A) A^\alpha (\Delta g_j + d_j)\|_{H,h} + \sum_{j=k}^{2k-1} \|r_{n-j}(-\tau A) A^\alpha \Delta y_j^0\|_{H,h} \\ &\leq \tau \sum_{j=k}^n \|A^\alpha r_{n-j}(-\tau A)\| (L(r) \|\Delta y_j\|_{V,h} + \|d_j\|_{H,h}) + \sum_{j=k}^{2k-1} \|r_{n-j}(-\tau A)\| \|\Delta y_j^0\|_{V,h} \\ &\leq C\tau \sum_{j=k}^n \frac{L(r) \|\Delta y_j\|_{V,h} + \|d_j\|_{H,h}}{((n+1-j)\tau)^\alpha} + C \sum_{j=k}^{2k-1} \|\Delta y_j^0\|_{V,h}. \end{aligned} \quad (14.39)$$

Außerdem gilt für $n\tau \leq T$

$$\tau \sum_{j=k}^n \frac{1}{((n+1-j)\tau)^\alpha} \leq \tau \sum_{j=1}^n \frac{1}{(\tau j)^\alpha} \leq \int_0^T \frac{1}{x^\alpha} dx = \frac{1}{1-\alpha} T^{1-\alpha},$$

also erhalten wir insgesamt

$$\|\Delta y_n\|_{V,h} \leq CL(r)\tau \sum_{j=k}^n \frac{1}{((n+1-j)\tau)^\alpha} \|\Delta y_j\|_{V,h} + M,$$

wobei

$$\begin{aligned} M &= C' \left(\max_{j=0,\dots,n} \|d_j\|_{H,h} + \max_{j=k,\dots,2k-1} \|\Delta y_j^0\|_{V,h} \right) \\ &\leq C \left(\max_{j=0,\dots,n} \|d_j\|_{H,h} + \max_{j=0,\dots,k-1} \|y_j\|_{V,h} \right). \end{aligned}$$

Setzen wir $f(t) := \|\Delta y_n\|_{V,h}$ für $t_n \leq t < t_{n+1}$, so folgt

$$f(t) \leq M + L \int_0^T \frac{1}{(t-s)^\alpha} f(s) ds$$

und daraus mit dem verallgemeinerten Gronwall-Lemma 14.21 die Behauptung. \square