8 FEM für parabolische Probleme

Sei $\Omega \subset \mathbb{R}^d$. Betrachte das Modellproblem

$$\left.\begin{array}{ll} u_t - \Delta u = f & \textit{auf } \Omega \times (0, \mathcal{T}) \\ u(x,t) = 0 & \forall\, x \in \partial\Omega,\ t > 0 \\ u(\cdot, 0) = u_0 & \textit{auf } \Omega \end{array}\right\} \tag{8.1}$$

Die numerische Behandlung dieser Aufgabenstellung erfolgt in 2 Schritten:

1. Schritt: Durch Diskretisierung im Ort erhält man ein System von ODEs, dieses Vorgehen nennt man *Semidiskretisierung.*

2. Schritt: Lösen des ODE-Systems mit Techniken aus dem 1. Teil der Vorlesung.

8.1 Variationsformulierung im Ort

Wie im 7. Kapitel wird der Lösungsbegriff erweitert. Betrachte dazu folgendes Beispiel:

Beispiel 8.1 Eine Variationsformulierung für

a) $-\Delta u = f \quad \textit{auf } \Omega$

b) $u = 0 \quad \textit{auf } \partial\Omega$

ergibt sich aus dem „Rezept“: Multiplizieren mit einer Testfunktion, Integrieren, partiell Integrieren. Eine klassische Lösung von a) erfüllt für $v \in C_0^\infty(\Omega)$

$$\int_\Omega \nabla u \cdot \nabla v = \int_\Omega -\Delta u v = \int_\Omega f v.$$

Deswegen erfüllen klassische Lösungen von a)

$$a(u,v) = \int_\Omega \nabla u \cdot \nabla v = l(v) = \int_\Omega f v \qquad \forall\, v \in C_0^\infty(\Omega)$$

$$\overset{\substack{C_0^\infty \textit{ ist dicht} \\ \textit{in } H_0^1}}{\Rightarrow} \quad a(u,v) = l(v) \qquad \forall\, v \in H_0^1(\Omega)$$

Wir erkennen, dass dieser Ausdruck bereits sinnvoll definiert ist, wenn u lediglich in $H^1(\Omega)$ ist, d.h. wir nennen $u \in H^1(\Omega)$ eine schwache Lösung von a), falls

$$\int_\Omega \nabla u \cdot \nabla v = \int_\Omega f v \qquad \forall\, v \in H_0^1(\Omega)$$

Die Randbedingung b) muss nun separat gefordert werden, d.h. eine schwache Lösung von a), b) ist ein $u \in H_0^1(\Omega)$, welches $\int_\Omega \nabla u \cdot \nabla v = \int_\Omega f v$ für alle $v \in H_0^1(\Omega)$ erfüllt.

Analog zu Beispiel 8.1 wird eine Variationsformulierung für (8.1) erzeugt. Für $v \in C_0^\infty(\Omega)$ und eine klassische Lösung u von (8.1) gilt (Im folgenden wird auch L^2 für $L^2(\Omega)$ und H^1 für $H^1(\Omega)$ geschrieben.):

$$\begin{aligned}
&u_t - \Delta u = f \\
\Rightarrow &\int_\Omega u_t v - \int_\Omega \Delta u v = \int_\Omega f v \\
\Rightarrow &\langle u_t(\cdot,t), v\rangle_{L^2} + a(u(\cdot,t), v) = \langle f(\cdot,v), v\rangle_{L^2},
\end{aligned}$$

wobei $a(w,v) = \int_\Omega \nabla w \cdot \nabla v$ für $w, v \in H^1(\Omega)$. Weil $C_0^\infty(\Omega)$ dicht in $H_0^1(\Omega)$ ist, folgt, dass u

$$\langle u_t(\cdot,t), v\rangle_{L^2} + a(u(\cdot,t), v) = \langle f(\cdot,t), v\rangle_{L^2} \qquad \forall\, v \in H_0^1(\Omega)$$

erfüllt. Wir sehen, dass der Lösungsbegriff abgeschwächt werden kann. So reicht es z.B., dass für jedes feste t die Funktion $u(\cdot,t) \in H_0^1(\Omega)$ ist. Dann ergibt sich

$$\left.\begin{aligned}
&\textit{Finde } u \in C^1([0,T], H_0^1(\Omega)), \textit{ sodass} \\
&\langle u'(t), v\rangle_{L^2} + a(u(t), v) = \langle f(t), v\rangle_{L^2} \qquad \forall\, v \in H_0^1(\Omega) \\
&u(0) = u_0 \ \textit{(in } H_0^1(\Omega))
\end{aligned}\right\} \tag{8.2}$$

Dabei wird gefordert

- $f \in C([0,T]; L^2(\Omega))$
- $u_0 \in H_0^1(\Omega)$

Bemerkung: Die Formulierung (8.2) ist nicht die schwächstmögliche, aber bequem, um die ODE-Theorie des ersten Teils der Vorlesung einzusetzen. Eine Verallgemeinerung der Formulierung wird in der Vorlesung PDE betrachtet. Insbesondere kann der Lösungsbegriff so erweitert werden, dass auch $u_0 \in L^2(\Omega)$ sinnvoll behandelt werden kann.

Bemerkung: (Ableitungsbegriff) Sei $u : (0,T) \to (\mathcal{X}; \|\cdot\|_{\mathcal{X}})$ eine Funktion wobei $t \in (0,T)$. Ein Element $u'(t) \in \mathcal{X}$ heißt Ableitung von u an der Stelle t, falls

$$\lim_{h\to 0} \left\| \frac{u(t+h) - u(t)}{h} - u'(t) \right\|_{\mathcal{X}} = 0.$$

Übung 8.2 Sei $u \in C^1((0,T); H_0^1(\Omega))$. Dann ist $u \in C^1((0,T); L^2(\Omega))$

Übung 8.3 Für $u \in C^1([0,T]; H_0^1(\Omega))$ gilt:

- $t \mapsto \|u(t)\|_{L^2}^2$ ist stetig differenzierbar und
- $\frac{d}{dt}\|u(t)\|_{L^2}^2 = 2\langle u'(t), u(t)\rangle_{L^2}$

Die Idee hinter dem 2. Punkt ist: $\frac{d}{dt}\langle u(t), u(t)\rangle_{L^2} = \langle u'(t), u(t)\rangle_{L^2} + \langle u(t), u'(t)\rangle_{L^2} = 2\langle u'(t), u(t)\rangle_{L^2}$.

Satz 8.4 *(Energieungleichung) Es gelte für ein $\gamma > 0$*

- $\gamma\|v\|_{H^1(\Omega)}^2 \le a(v,v) \qquad \forall\, v \in H_0^1(\Omega)$
- *u löst (8.2)*

Dann gilt:

$$\|u\|_{L^2(\Omega)} \leq e^{-\gamma t}\|u_0\|_{L^2(\Omega)} + \int_0^t e^{-\gamma(t-s)}\|f(s)\|_{L^2(\Omega)}ds$$

Beweis:
1. Schritt: Wir nehmen zunächst an, dass $\|u(s)\|_{L^2} > 0$ für alle $0 < s < t$ ist. Dann gilt:

$$\frac{1}{2}\frac{d}{dt}\|u(t)\|_{L^2}^2 \overset{\text{Übung 8.3}}{=} \langle u'(t), u(t)\rangle_{L^2} \qquad \textit{und}$$
$$\frac{d}{dt}\|u(t)\|_{L^2} = \frac{d}{dt}\sqrt{\|u(t)\|_{L^2}^2} = \frac{1}{2\sqrt{\|u(t)\|_{L^2}^2}}\frac{d}{dt}\|u(t)\|_{L^2}^2 = \frac{\langle u'(t), u(t)\rangle_{L^2}}{\|u(t)\|_{L^2}}$$

Mit einer Testfunktion $v = u(s)$ für festes s folgt aus (8.2)

$$\langle f(s), u(s)\rangle_{L^2} = \langle f(s), v\rangle_{L^2} \overset{(8.2)}{=} \langle u'(s), v\rangle_{L^2} + a(u(s), v) =$$
$$= \underbrace{\langle u'(s), u(s)\rangle_{L^2}}_{\|u(s)\|_{L^2}\frac{d}{dt}\|u(t)\|_{L^2}\Big|_{t=s}} + a(u(s), u(s))$$
$$\Rightarrow a(u(s), u(s)) + \|u(s)\|_{L^2}\frac{d}{dt}\|u(t)\|_{L^2}\Big|_{t=s} \overset{\substack{\textit{Cauchy-}\\ \textit{Schwarz}}}{\leq} \|f(s)\|_{L^2}\|u(s)\|_{L^2}$$

Da $a(u(s), u(s)) \geq \gamma\|u(s)\|_{H^1}^2 \geq \gamma\|u(s)\|_{L^2}^2$

$$\Rightarrow \gamma\|u(s)\|_{L^2} + \frac{d}{dt}\|u(t)\|_{L^2}\Big|_{t=s} \leq \|f(s)\|_{L^2} \qquad \forall\ 0 \leq s \leq t$$

Ein integrierender Faktor für die linke Seite ist $e^{\gamma t}$

$$\Rightarrow \frac{d}{dt}\left(e^{\gamma t}\|u(t)\|_{L^2}\right)\Big|_{t=s} = e^{\gamma s}\left(\gamma\|u(s)\|_{L^2} + \frac{d}{dt}\|u(t)\|_{L^2}\Big|_{t=s}\right) \leq e^{\gamma s}\|f(s)\|_{L^2}$$

Durch Integration von 0 bis t ergibt sich

$$e^{\gamma t}\|u(t)\|_{L^2} - e^{\gamma 0}\|u(0)\|_{L^2} \leq \int_0^t e^{\gamma s}\|f(s)\|_{L^2}ds$$
$$\Rightarrow \|u(t)\|_{L^2} \leq e^{-\gamma t}\|u_0\|_{L^2} + \int_0^t e^{-\gamma(t-s)}\|f(s)\|_{L^2}ds$$

2.Schritt: Wir betrachten den Fall $\|u(s)\|_{L^2} \geq 0$ auf $[0, T]$. Aus (8.2) ergibt sich mit $v = u(s)$

$$\underbrace{a(u(s), u(s))}_{\geq\gamma\|u\|_{H^1}^2 \geq \gamma\|u\|_{L^2}^2} + \underbrace{\langle u'(s), u(s)\rangle_{L^2}}_{\frac{1}{2}\frac{d}{dt}\|u(t)\|_{L^2}^2\Big|_{t=s}} = \langle f(s), u(s)\rangle_{L^2} \overset{\substack{\textit{Cauchy-}\\ \textit{Schwarz}}}{\leq} \|f(s)\|_{L^2}\|u(s)\|_{L^2}$$

Um das Problem zu umgehen, dass $\|u(s)\|_{L^2} = 0$ sein kann, weil man ja durch $\|u(s)\|_{L^2}$ dividieren will, definiert man für ein $\varepsilon > 0$ die Funktion $h_\varepsilon(t) := \sqrt{\|u(t)\|_{L^2}^2 + \varepsilon^2}$. Dann ist

- $\frac{d}{dt} h_\varepsilon(t)\Big|_{t=s} = \frac{d}{dt}\sqrt{h_\varepsilon^2(t)}\Big|_{t=s} = \frac{1}{2h_\varepsilon(s)}\frac{d}{dt}h_\varepsilon^2(t)\Big|_{t=s} = \frac{1}{h_\varepsilon(s)}\langle u'(t), u(t)\rangle_{L^2}\Big|_{t=s}$
- $\frac{d}{dt}\|u(t)\|_{L^2}^2\Big|_{t=s} = \frac{d}{dt}h_\varepsilon^2(t)\Big|_{t=s} = 2h_\varepsilon(s)\frac{d}{dt}h_\varepsilon(t)\Big|_{t=s}$

Also folgt:

$$\gamma h_\varepsilon^2(s) + h_\varepsilon(s)\frac{d}{dt}h_\varepsilon(t)\Big|_{t=s} \le \|f(s)\|_{L^2}\underbrace{\|u(s)\|_{L^2}}_{\le h_\varepsilon(s)} + \gamma\varepsilon^2$$

Da $\frac{\varepsilon^2}{h_\varepsilon(s)} = \frac{\varepsilon^2}{\sqrt{\|u\|_{L^2}^2+\varepsilon^2}} \le \frac{\varepsilon^2}{\sqrt{\varepsilon^2}} = \varepsilon$, ergibt sich bei Division der Ungleichung durch $h_\varepsilon(s)$

$$\gamma h_\varepsilon(s) + \frac{d}{dt}h_\varepsilon(t)\Big|_{t=s} \le \|f(s)\|_{L^2} + \gamma\frac{\varepsilon^2}{h_\varepsilon(s)} \le \|f(s)\|_{L^2} + \gamma\varepsilon$$

Wie im 1. Schritt ergibt sich durch Multiplikation mit $e^{\gamma s}$ und Integration

$$e^{\gamma s}h_\varepsilon(t) - h_\varepsilon(0) \le \int_0^t e^{\gamma s}\left[\|f(s)\|_{L^2} + \gamma\varepsilon\right] ds$$

$$\Rightarrow h_\varepsilon(t) \le e^{-\gamma t}h_\varepsilon(0) + \int_0^t e^{-\gamma(t-s)}\left[\|f(s)\|_{L^2} + \gamma\varepsilon\right] ds \qquad \forall\, \varepsilon > 0$$

Für $\varepsilon \to 0$ folgt die Behauptung. □

Übung: Satz 8.4 liefert die Eindeutigkeit der Lösung von (8.2).

Bemerkung: Für $f \equiv 0$ ist die Wärmeleitungsleichung *dissipativ* (in $L^2(\Omega)$), d.h. $\|u(t)\|_{L^2} \le e^{-\gamma t}\|u(0)\|_{L^2}$. Dann folgt für 2 verschiedene Anfangsbedingungen u_0, $\tilde{u}_0$, dass die Lösungen $u(t)$, $\tilde{u}(t)$ die Bedingung $\|u(t) - \tilde{u}(t)\|_{L^2} \le e^{-\gamma t}\|u_0 - \tilde{u}_0\|_{L^2}$ erfüllen. Gute numerische Verfahren sollten dieses qualitative Verhalten widerspiegeln.

8.2 Semidiskretisierung im Ort („Linienmethode“)

Das *Ziel* dabei ist die Approximotion von (8.2) durch ein (endliches) System von ODEs. Sei $V_N \subset H_0^1(\Omega)$ mit $dim(V_N) = N < \infty$ und Basis $\{\varphi_i \mid i = 1, \dots, N\}$. Sei $u_{0,N} \in V_N$ eine Approximation an u_0. Dann ist die *semidiskrete Approximation* u_N an die Lösung u gegeben durch

$$\left.\begin{array}{ll} \textit{Finde } u_N \in C^1([0,T]; V_N) \textit{ sodass} & \\ (8.3a) \quad \langle u_N'(t), v\rangle_{L^2} + a(u_N(t), v) = \langle f(t), v\rangle_{L^2} & \forall\, v \in V_N \\ (8.3b) \quad u_N(0) = u_{0,N} & \end{array}\right\} \tag{8.3}$$

(8.3) stellt ein ODE-System dar. Definiert man die Steifigkeitsmatrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ und die Massematrix $\mathbf{M} \in \mathbb{R}^{N\times N}$ durch

$$\mathbf{A}_{ij} = a(\varphi_j, \varphi_i), \qquad \mathbf{M}_{ij} = \langle \varphi_j, \varphi_i\rangle_{L^2}, \quad i, j = 1, \dots, N \tag{8.4}$$

und $\mathbf{F}(t)$ durch

$$\mathbf{F}_i(t) = \langle f(t), \varphi_i \rangle_{L^2} \tag{8.5}$$

so ergibt sich aus dem Ansatz $u_N(t) = \sum_{i=1}^{N} \mathbf{u}_i(t)\varphi_i$, dass (8.3) zum ODE-System

$$\left.\begin{aligned} &\mathbf{M}\mathbf{u}'(t) + \mathbf{A}\mathbf{u}(t) = \mathbf{F}(t) \qquad t > 0 \\ &\mathbf{u}(0) = \mathbf{u}_0 \end{aligned}\right\} \tag{8.6}$$

äquivalent ist, wobei $u_{0,N} = \sum_{i=1}^{N} \mathbf{u}_{0,i}\,\varphi_i$ die Darstellung in der Basis $\{\varphi_i \mid i = 1, \ldots, N\}$ von $u_{0,N} \in V_N$ ist.

Um die Äquivalenz von (8.3) mit (8.6) zu sehen, schreiben wir $u_N(t) = \sum_{i=1}^{N} \mathbf{u}_i(t)\varphi_i$. Damit gilt:

$$\begin{aligned}
&u_N \textit{ löst (8.3a)} \\
&\Leftrightarrow \left\langle \sum_{j=1}^{N} \mathbf{u}_j'(t)\varphi_j, \sum_{i=1}^{N} \mathbf{v}_i\varphi_i \right\rangle_{L^2} + a\left(\sum_{j=1}^{N} \mathbf{u}_j(t)\varphi_j, \sum_{i=1}^{N} \mathbf{v}_i\varphi_i \right) = \left\langle f(t), \sum_{i=1}^{N} \mathbf{v}_i\varphi_i \right\rangle_{L^2} \qquad \forall\, \mathbf{v} \in \mathbb{R}^N \\
&\Leftrightarrow \sum_{i,j=1}^{N} \mathbf{u}_j'(t)\mathbf{v}_i\langle \varphi_j, \varphi_i \rangle_{L^2} + \sum_{i,j=1}^{N} \mathbf{u}_j(t)\mathbf{v}_i\, a(\varphi_j, \varphi_i) = \sum_{i=1}^{N} \mathbf{v}_i\langle f(t), \varphi_i \rangle_{L^2} \qquad \forall\, v \in \mathbb{R}^N \\
&\Leftrightarrow \mathbf{v}^T\mathbf{M}\mathbf{u}'(t) + \mathbf{v}^T\mathbf{A}\mathbf{u}(t) = \mathbf{v}^T\mathbf{F}(t) \qquad \forall\, \mathbf{v} \in \mathbb{R}^N \\
&\Leftrightarrow \mathbf{M}\mathbf{u}'(t) + \mathbf{A}\mathbf{u}(t) = \mathbf{F}(t)
\end{aligned}$$

Übung 8.5 Die Matrizen $\mathbf{A}$ und $\mathbf{M}$ sind SPD. Überdies gilt für alle $\mathbf{v}, \mathbf{w} \in \mathbb{R}^N$ mit $\mathbf{v} = \sum_{i=1}^{N} \mathbf{v}_i\varphi_i$, $\mathbf{w} = \sum_{i=1}^{N} \mathbf{w}_i\varphi_i$, dass $\mathbf{v}^T\mathbf{M}\mathbf{w} = \langle w, v \rangle_{L^2}$ und $\mathbf{v}^T\mathbf{A}\mathbf{w} = a(w, v)$.

Bemerkung: Aus Übung 8.5 folgt, dass (8.6) äquivalent ist zu

$$\begin{aligned} &\mathbf{u}' = \mathbf{M}^{-1}\mathbf{F}(t) - \mathbf{M}^{-1}\mathbf{A}\mathbf{u}(t) \\ &\mathbf{u}(0) = u_0 \end{aligned}$$

D.h. die Existenz und Eindeutigkeit von (8.3) ist gegeben.

Analog zu Satz 8.4 gilt

Lemma 8.6 *Seien $r \in C^0([0,T]; L^2(\Omega))$ und $w \in C^1([0,T]; V_N)$ und erfüllen*

$$\langle w', v \rangle_{L^2} + a(w, v) = \langle r(t), v \rangle_{L^2} \textit{ für alle } v \in V_N.$$

Dann gilt:

$$\|w(t)\|_{L^2} \le e^{-\gamma t}\|w(0)\|_{L^2} + \int_0^t e^{-\gamma(t-s)}\|r(s)\|_{L^2} ds$$

Beweis: Der Beweis erfolgt analog zu jenem von Satz 8.4. □

Im folgenden wollen wir $\|u(t)-u_N(t)\|_{L^2}$ abschätzen, indem wir $\|u(t)-R_N u(t)\|_{L^2}$ abschätzen. R_N bezeichnet dabei wieder den Ritzprojektor $R_N : H_0^1(\Omega) \to V_N$. Dieser ist lt. Definition 7.27 für alle $v \in V_N$ durch $a(w,v) = a(R_N w, v)$ definiert. Wir wissen bereits, dass R_N linear und beschränkt ist. Die Beschränktheit folgt aus der Existenz einer Konstante $C > 0$, sodass $\|R_N w\|_{H^1} \leq C\|w\|_{H^1}$ für alle $w \in H_0^1(\Omega)$ gilt.

Satz 8.7 *Sei $u \in C^1([0,T]; H_0^1(\Omega))$ eine Lösung von (8.2) und u_N eine Lösung von (8.3). Dann gilt*

$$\|u(t) - u_N(t)\|_{L^2} \leq \|u(t) - R_N u(t)\|_{L^2} + \|u_N(0) - R_N u(0)\|_{L^2} e^{-\gamma t} + \\ + \int_0^t e^{-\gamma(t-s)} \|u'(s) - R_N u'(s)\|_{L^2} ds$$

Beweis: Wird $u_N(t) - u(t) = \underbrace{u_N(t) - R_N u(t)}_{=:\theta(t)} + \underbrace{R_N u(t) - u(t)}_{=:\varrho(t)}$ geschrieben, dann ist

$$\|u_N(t) - u(t)\|_{L^2} \leq \|\theta(t)\|_{L^2} + \|\varrho(t)\|_{L^2}.$$

1. Schritt: Aus der Linearität und der Beschränktheit von R_N und $u \in C^1([0,T]; H_0^1(\Omega))$ folgt, dass $(R_N u)' = R_N u'$, weil

$$\begin{aligned}
&\overline{\lim_{h\to 0}} \left\| \frac{1}{h}\left(R_N u(t+h) - R_N u(t)\right) - R_N u'(t) \right\|_{H^1} \overset{R_N \text{ linear}}{=} \\
&= \overline{\lim_{h\to 0}} \left\| R_N \left(\frac{u(t+h) - u(t)}{h} - u'(t) \right) \right\|_{H^1} \overset{R_N \text{ beschränkt}}{=} \\
&= \overline{\lim_{h\to 0}} C \left\| \frac{u(t+h) - u(t)}{h} - u'(t) \right\|_{H^1} = 0 \qquad \textit{nach Definition von } u'.
\end{aligned}$$

2. Schritt: Aus dem 1. Schritt folgt $\theta \in C^1([0,T]; V_N)$. Weiters ist für jedes $v \in V_N$

$$\begin{aligned}
&\langle \theta'(t), v\rangle_{L^2} + a(\theta(t), v) = \langle u_N', v\rangle_{L^2} + a(u_N, v) - \langle R_N u', v\rangle_{L^2} - a(R_N u, v) = \\
&\overset{(8.3)}{=} \langle f(t), v\rangle_{L^2} - \langle R_N u', v\rangle_{L^2} - a(R_N u, v) = \\
&\overset{\text{Def. v. } R_N}{=} \langle f(t), v\rangle_{L^2} - \langle R_N u', v\rangle_{L^2} - a(u, v) = \\
&\overset{(8.2)}{=} \langle u', v\rangle_{L^2} - \langle R_N u', v\rangle_{L^2} = \langle u' - R_N u'(t), v\rangle_{L^2}
\end{aligned}$$

3. Schritt: Aus Lemma 8.6 folgt für θ

$$\|\theta\|_{L^2} \leq e^{-\gamma t} \underbrace{\|\theta(0)\|_{L^2}}_{=\|u_N(0) - R_N u(0)\|_{L^2}} + \int_0^t e^{-\gamma(t-s)} \|u'(s) - R_N u'(s)\|_{L^2} ds.$$

□

Bemerkung: Satz 8.7 zeigt, dass $\|u(t) - u_N(t)\|_{L^2}$ durch den Fehler $\|u(t) - R_N u(t)\|_{L^2}$ + 2 *Terme* abgeschätzt werden kann, die eine Fehlerakkumulation für die Zeiten $0 \leq s < t$ darstellen. Man spricht vom „Gedächtnis“ von parabolischen Gleichungen.

Regularitätsannahmen an u erlauben Abschätzungen, die explizit in h sind.

Korollar 8.8 *Sei $V_N = S_0^1(\mathcal{T})$, wobei $\mathcal{T}$ die Bedingungen aus Satz 7.25 erfüllt. Erfülle die Lösung u von (8.2) die Regularitätsvorraussetzung $u \in C^3(\bar{\Omega} \times [0,T])$. Sei $u_{0,N} \in V_N$ entweder $u_{0,N} = R_N u_0$ oder $u_{0,N} = I u_0$, wobei $I u_0$ den stückweisen linearen Interpolanden bezeichne. Dann gilt:*

$$\|u(t) - u_N(t)\|_{L^2(\Omega)} \leq Ch \max_{0 \leq s < t} \Big(|u(\cdot, s)|_{C^2(\bar{\Omega})} + |u_t(\cdot, s)|_{C^2(\bar{\Omega})} \Big)$$

Ist Ω sogar konvex, dann ist

$$\|u(t) - u_N(t)\|_{L^2(\Omega)} \leq Ch^2 \max_{0 \leq s < t} \Big(|u(\cdot, s)|_{C^2(\bar{\Omega})} + |u_t(\cdot, s)|_{C^2(\bar{\Omega})} \Big)$$

Beweis: Nach Satz 7.29 gilt für jedes $v \in C^2(\bar{\Omega})$

- $\|v - R_N v\|_{L^2(\Omega)} \leq \|v - R_N v\|_{H^1(\Omega)} \leq C\|v - Iv\|_{H^1(\Omega)} \leq Ch|v|_{C^2(\bar{\Omega})}$
- falls Ω konvex ist, gilt sogar: $\|v - R_N v\|_{L^2(\Omega)} \leq Ch^2 |v|_{C^2(\bar{\Omega})}$

Damit folgt die Behauptung aus Satz 8.7 □

Übung 8.9 Sei $\theta \in C^1([0,T]; V_N)$ und gelte für alle $v \in V_N$ und für ein $r \in C^0([0,T]; L^2(\Omega))$, dass $\langle \theta', v \rangle_{L^2(\Omega)} + a(\theta, v) = \langle r(t), v \rangle_{L^2(\Omega)}$. Zeigen Sie:

$$|\theta(t)|^2_{H^1(\Omega)} \leq |\theta(0)|^2_{H^1(\Omega)} + \int_0^t \|r(s)\|^2_{L^2(\Omega)} ds$$

Hinweis: Betrachte $v = \theta'$.
Zeigen Sie, dass

$$|u(t) - u_N(t)|^2_{H^1} \leq 2\, |u(t) - R_N u(t)|^2_{H^1} + 2\, |u_{0,N} - R_N u_0|^2_{H^1} + 2 \int_0^t \|u'(s) - R_N u'(s)\|^2_{L^2} ds$$

8.3 Volldiskrete Verfahren

Die Semidiskretisierung führt auf das ODE-System

$$\mathbf{M}\mathbf{u}' + \mathbf{A}\mathbf{u} = \mathbf{F}, \qquad \mathbf{u}(0) = \mathbf{u}_0 \tag{8.7}$$

wobei $\mathbf{M}$ und $\mathbf{A}$ SPD sind.

Um zu verstehen, wie sich die Lösungen von (8.7) verhalten, versuchen wir, (8.7) in ein entkoppeltes ODE-System umzuwandeln.

Satz 8.10 *Seien $\mathbf{A}$ und $\mathbf{M} \in \mathbb{R}^{N \times N}$ SPD. Dann gelten für das verallgemeinerte Eigenwertproblem*

$$\textit{Finde } (\mathbf{v}, \lambda) \in \mathbb{R}^N \setminus \{0\} \times \mathbb{C}, \ \textit{sodass } \mathbf{A}\mathbf{v} = \lambda \mathbf{M}\mathbf{v} \tag{8.8}$$

folgende Aussagen:

(i) Der Eigenwert λ erfülle $\lambda > 0$.

(ii) Es gibt N Eigenpaare $(\mathbf{v}_i, \lambda_i)$, $i = 1, \ldots, N$, die orthogonal bzgl. $(\cdot,\cdot)_\mathbf{A}$ und $(\cdot,\cdot)_\mathbf{M}$ sind, d.h.

$$(\mathbf{v}_i, \mathbf{v}_j)_\mathbf{M} = \langle \mathbf{M}\mathbf{v}_i, \mathbf{v}_j\rangle_2 = 0 \qquad \forall i \neq j$$
$$(\mathbf{v}_i, \mathbf{v}_j)_\mathbf{A} = \langle \mathbf{A}\mathbf{v}_i, \mathbf{v}_j\rangle_2 = 0 \qquad \forall i \neq j$$

(iii) Die Matrix $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_N) \in \mathbb{R}^{N\times N}$ diagonalisiert $\mathbf{M}$ und $\mathbf{A}$ simultan, d.h.

$$\mathbf{V}^T\mathbf{M}\mathbf{V} = \textit{Diagonalmatrix}$$
$$\mathbf{V}^T\mathbf{A}\mathbf{V} = \textit{Diagonalmatrix}$$

(iv) Falls die $\mathbf{v}_i$ so normiert werden, dass $(\mathbf{v}_i, \mathbf{v}_j)_\mathbf{M} = \delta_{ij}$, dann gilt

$$\mathbf{V}^T\mathbf{M}\mathbf{V} = Id, \quad \mathbf{V}^T\mathbf{A}\mathbf{V} = \mathbf{D} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{pmatrix}$$

Beweis: Übung. *Hinweis:* Betrachte das EWP $\mathbf{M}^{-\frac{1}{2}}\mathbf{A}\mathbf{M}^{-\frac{1}{2}}\mathbf{x} = \lambda\mathbf{x}$ □

Definiert man nun $\widetilde{\mathbf{u}} = \mathbf{V}^{-1}\mathbf{u}$, $\widetilde{\mathbf{f}} = \mathbf{V}^T\mathbf{F}$, $\widetilde{u}_0 = \mathbf{V}^{-1}u_0$, dann ist (8.7) äquivalent zu

$$\widetilde{\mathbf{u}}' + \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{pmatrix} \tilde{\mathbf{u}} = \widetilde{\mathbf{f}}, \quad \widetilde{\mathbf{u}}(0) = \tilde{u}_0 \tag{8.9}$$

(8.9) stellt ein im Sinne des 1. Teils der Vorlesung steifes ODE-System dar, falls einige EW $\lambda_i > 0$ groß sind. Das ist bei parabolischen Problemen der Fall, wie der folgende Satz zeigt.

Satz 8.11 *Sei $\mathcal{T}$ eine Triangulierung, die die Bedingungen aus Satz 7.25 erfüllt. Seien die Steifungkeitsmatrix $\mathbf{A}$ und die Massematrix $\mathbf{M}$ durch (8.4) definiert, wobei $\{\varphi_i \mid i = 1, \ldots, N\}$ die Basis aus Hutfunktionen von $V_N = S_0^1(\mathcal{T})$ ist. Sei $h_{min} = \min\limits_{K\in\mathcal{T}} h_K$. Dann existiert eine Konstante $C > 0$, die nur von $\varepsilon > 0$ abhängt (siehe Satz 7.25), sodass*

$$C^{-1}\|u\|^2_{L^2(\Omega)} \leq |u|^2_{H^1(\Omega)} \leq \frac{C}{h^2_{min}}\|u\|^2_{L^2(\Omega)} \qquad \forall\ u \in S_0^1(\mathcal{T})$$

Insbesondere folgt für die EW λ_i von (8.8)

$$C^{-1} \leq \min_{i=1,\ldots,N} \lambda_i \leq \max_{i=1,\ldots,N} \lambda_i \leq \frac{C}{h^2_{min}} \tag{8.10}$$

Bemerkung: (Interpretation von λ_{min} und λ_{max})

$$\lambda_{min} = \min_{0\neq u\in S_0^1(\mathcal{T})} \frac{|u|^2_{H^1}}{\|u\|^2_{L^2}}$$
$$\lambda_{max} = \max_{0\neq u\in S_0^1(\mathcal{T})} \frac{|u|^2_{H^1}}{\|u\|^2_{L^2}}$$

Beweis: (von Satz 8.11)
1. Schritt: Nach Satz 7.20 gilt $C^{-1}\|u\|_{L^2(\Omega)} \leq |u|_{H^1(\Omega)}$ für alle $u \in H_0^1(\Omega)$. Damit folgt die erste Ungleichung.

2. Schritt: (inverse Ungleichung)
Sei $K \in \mathcal{T}$ fest. Wir behaupten, dass es eine Konstante $C > 0$ gibt, welche nur von $\varepsilon > 0$ abhängt, sodass

$$\|\nabla w\|_{L^2(K)} \leq \frac{C}{h_K}\|w\|_{L^2(K)} \qquad \forall\, w \in \mathcal{P}_1.$$

Dazu seien $w \in \mathcal{P}_1$ und $x^* \in \overline{K}$ mit $|w(x^*)| = \|w\|_{C(\overline{K})}$. Dann gilt (Das zweite Ungleichheitszeichen folgt aus dem 2. Schritt des Beweises von Satz 7.25.):

$$\|\nabla w\|^2_{L^2(K)} \leq h_K^2\|\nabla w\|^2_{C(\overline{K})} \leq C^*\|w\|^2_{C(\overline{K})} = C^*|w(x^*)|^2$$

Weil $w \in \mathcal{P}_1$ ist, kann es Taylorentwickelt werden (sinnvollerweise um x^*):

$$\begin{aligned} &w(x) = w(x^*) + \nabla w(x^*)(x - x^*) \\ &\Rightarrow |w(x)| \geq |w(x^*)| - |\nabla w(x^*)|\ |x - x^*| \geq \\ &\qquad \geq |w(x^*)| - \frac{\sqrt{C^*}}{h_K}|w(x^*)|\ |x - x^*| = |w(x^*)|\left[1 - \frac{\sqrt{C^*}}{h_K}|x - x^*|\right] \end{aligned}$$

für $B := K \cap B_{\frac{1}{2\sqrt{C^*}}h_K}(x^*)$ gilt:

- $|w(x)| \geq \frac{1}{2}|w(x^*)|$
- $area(B) \geq Ch_K^2$ für geeignetes $C > 0$, das nur von ε abhängt

Damit gilt

$$\|w\|^2_{L^2(K)} \geq \|w\|^2_{L^2(B)} \geq \frac{1}{4}area(B)\,|w(x^*)|^2 \geq Ch_K^2\,|w(x^*)|^2$$

Damit folgt die Behauptung.

3. Schritt: Aus dem 2. Schritt folgt für $w \in V_N$:

$$\begin{aligned} \|\nabla w\|^2_{L^2(\Omega)} &= \sum_{K\in\mathcal{T}} \|\nabla w\|^2_{L^2(K)} \leq \sum_{K\in\mathcal{T}} \frac{C}{h_K^2}\|w\|^2_{L^2(K)} \\ &\leq \frac{C}{h_{min}^2}\sum_{K\in\mathcal{T}} \|w\|^2_{L^2(K)} = \frac{C}{h_{min}^2}\|w\|^2_{L^2(\Omega)} \end{aligned}$$

4. Schritt: Für jedes Eigenpaar $(\mathbf{v}, \lambda)$ des Eigenwertproblems (EWP) $\lambda\mathbf{M}\mathbf{v} = \mathbf{A}\mathbf{v}$ gilt $\lambda\mathbf{v}^T\mathbf{M}\mathbf{v} = \mathbf{v}^T\mathbf{A}\mathbf{v}$, d.h. $\lambda\|v\|^2_{L^2(\Omega)} = |v|^2_{H^1(\Omega)}$, falls $v = \sum_{i=1}^{N} \mathbf{v}_i\varphi_i$. Also folgt $C^{-1} \leq \lambda \leq \frac{C}{h_{min}^2}$ für jeden EW λ. □

Bemerkung: Die h-Abhängigkeit von λ_{max} ist scharf. In 1D auf regelmäßigen Gittern ist $\lambda_{max} \sim \frac{1}{h_{min}^2}$

Satz 8.11 zeigt, dass wir ein *implizites* Verfahren zum Lösen von (8.5) verwenden sollten. Das einfachste Verfahren ist das *implizite Eulerverfahren*:

$$\left\langle \frac{u_N^{n+1} - u_N^n}{k}, v \right\rangle_{L^2(\Omega)} + a(u_N^{n+1}, v) = \langle f(t_{n+1}), v \rangle_{L^2(\Omega)} \qquad \forall\, v \in V_N \tag{8.11}$$

wobei $k > 0$ der Zeitschritt, $t_n = nk$ und $u_N^n \approx u_N(t_n)$ ist. In Matrixschreibweise ist (8.11)

$$\frac{1}{k}\mathbf{M}(\mathbf{u}^{n+1} - \mathbf{u}^n) + \mathbf{A}\mathbf{u}^{n+1} = \mathbf{f}^{n+1} \tag{8.12}$$

d.h., in jedem Schritt muss das LGS

$$(\mathbf{M} + k\mathbf{A})\mathbf{u}^{n+1} = \mathbf{M}\mathbf{u}^n + k\mathbf{f}^{n+1}$$

gelöst werden.

Bemerkung: Es bietet sich an, $\mathbf{M} + k\mathbf{A}$ einmal zu zerlegen (Choleskyzerlegung), und dann in jedem Zeitschritt eine Vorwärts- und Rückwärtssubstitution zu machen.

Satz 8.12 *Sei $u \in C^1([0,T]; H_0^1(\Omega))$ Lösung von (8.2) und erfülle die Regularitätsvoraussetzung $u \in C^2([0,T]; L^2(\Omega))$. Sei $k_0 > 0$ fest gewählt. Sei λ_{min} der kleinste EW des verallgemeinerten EWP*

$$\lambda \mathbf{M} x = \mathbf{A} x,$$

wobei $\mathbf{M}$ und $\mathbf{A}$ die Massematrix und die Steifigkeitsmatrix der Ortsdiskretisierung sind. Dann gilt: Es existiert $b > 0$, welches nur von k_0 und λ_{min} abhängt, so dass für jeden Zeitschritt $k \in (0, k_0]$ gilt:

$$\begin{aligned}\|u_N^n - u(t_n)\|_{L^2} \le \|u(t_n) - R_N u(t_n)\|_{L^2} + e^{-bt_n}\|u_{0,N} - R_N u_0\|_{L^2} + \\ + \int_0^{t_n} e^{-b(t_n - t)} \left[\|u'(t) - R_N u'(t)\|_{L^2} + k\, \|u''(t)\|_{L^2}\right] dt\end{aligned}$$

Beweis:
1. Schritt:

$$u_N^n - u(t_n) = \underbrace{u_N^n - R_N u(t_n)}_{=:\theta^n} + \underbrace{R_N u(t_n) - u(t_n)}_{=:\varrho^n}$$

2. Schritt: (Rekurrenzrelation für θ^n)

$$\begin{aligned}&\tfrac{1}{k}\langle u_N^{n+1} - u_N^n, v\rangle_{L^2} + a(u_N^{n+1}, v) = \langle f(t_{n+1}), v\rangle_{L^2} \quad \forall\, v \in V_N \\ &\langle u'(t_{n+1}), v\rangle_{L^2} + a(u(t_{n+1}), v) = \langle f(t_{n+1}), v\rangle_{L^2} \quad \forall\, v \in H_0^1(\Omega)\end{aligned}$$

$$\overset{Taylor}{\Rightarrow} u(t_n) = u(t_{n+1}) + \underbrace{(t_n - t_{n+1})}_{=-k} u'(t_{n+1}) + \int\limits_{t_{n+1}}^{t_n} (t - t_{n+1}) u''(t) dt$$

$$\Rightarrow u'(t_{n+1}) = \frac{u(t_{n+1}) - u(t_n)}{k} - \frac{1}{k} \int\limits_{t_n}^{t_{n+1}} (t - t_{n+1}) u''(t) dt =$$

$$= \frac{R_N u(t_{n+1}) - R_N u(t_n)}{k} + \frac{u(t_{n+1}) - u(t_n)}{k} - \frac{R_N u(t_{n+1}) - R_N u(t_n)}{k} -$$

$$- \frac{1}{k} \int\limits_{t_n}^{t_{n+1}} (t - t_{n+1}) u''(t) dt =$$

$$= \frac{R_N u(t_{n+1}) - R_N u(t_n)}{k} + \underbrace{\frac{1}{k} \int\limits_{t_n}^{t_{n+1}} u'(t) - R_N u'(t) dt}_{=:w_1^{n+1}} - \underbrace{\frac{1}{k} \int\limits_{t_n}^{t_{n+1}} (t - t_{n+1}) u''(t) dt}_{=:w_2^{n+1}}$$

Es gilt:

$$a(u(t_{n+1}), v) = a(R_N u(t_{n+1}), v) \qquad \forall\ v \in V_N$$

$$\tfrac{1}{k} \langle u_N^{n+1} - u_N^n, v \rangle_{L^2} + a(u_N^{n+1}, v) = \langle f(t_{n+1}), v \rangle_{L^2} \qquad \forall\ v \in V_N$$

$$\tfrac{1}{k} \langle R_N u(t_{n+1}) - R_N u(t_n), v \rangle_{L^2(\Omega)} + a(R_N u(t_{n+1}), v) =$$

$$= \langle f(t_{n+1}), v \rangle_{L^2} - \langle w_1^{n+1}, v \rangle_{L^2} - \langle w_2^{n+1}, v \rangle_{L^2} \qquad \forall v \in V_N$$

Differenzenbildung führt auf

$$\frac{1}{k} \langle \theta^{n+1} - \theta^n, v \rangle_{L^2} + a(\theta^{n+1}, v) = \langle w_1^{n+1}, v \rangle_{L^2} + \langle w_2^{n+1}, v \rangle_{L^2} \qquad \forall\ v \in V_N$$

mit $v = \theta^{n+1}$ ergibt sich

$$\|\theta^{n+1}\|_{L^2}^2 + k \underbrace{a(\theta^{n+1}, \theta^{n+1})}_{=|\theta^{n+1}|_{H^1}^2 \geq \lambda_{min} \|\theta^{n+1}\|_{L^2}^2} = \langle \theta^n, \theta^{n+1} \rangle_{L^2} + k \langle w_1^{n+1}, v \rangle_{L^2} + k \langle w_2^{n+1}, v \rangle_{L^2}$$

$$\Rightarrow (1 + k\lambda_{min}) \|\theta^{n+1}\|_{L^2}^2 \overset{\substack{Cauchy-\\Schwarz}}{\leq}$$

$$\leq \|\theta^n\|_{L^2} \|\theta^{n+1}\|_{L^2} + k \|w_1^{n+1}\|_{L^2} \|\theta^{n+1}\|_{L^2} + k \|w_2^{n+1}\|_{L^2} \|\theta^{n+1}\|_{L^2}$$

$$\Rightarrow (1 + k\lambda_{min}) \|\theta^{n+1}\|_{L^2} \leq \|\theta^n\|_{L^2} + k \|w_1^{n+1}\|_{L^2} + k \|w_2^{n+1}\|_{L^2}$$

3. Schritt: (Auflösen der Rekurrenz)

$$\|\theta^{n+1}\|_{L^2} \leq \frac{1}{1 + k\lambda_{min}} \|\theta^n\|_{L^2} + \frac{k}{1 + k\lambda_{min}} \left(\|w_1^{n+1}\|_{L^2} + \|w_2^{n+1}\|_{L^2} \right)$$

$$\Rightarrow \|\theta^n\|_{L^2} \leq (1 + \lambda_{min} k)^{-n} \|\theta^0\|_{L^2} + \frac{k}{1 + \lambda_{min} k} \sum_{j=1}^{n} (1 + k\lambda_{min})^{-(n-j)} \left(\|w_1^j\|_{L^2} + \|w_2^j\|_{L^2} \right) =$$

$$= (1 + \lambda_{min} k)^{-n} \|\theta^0\|_{L^2} + k \sum_{j=1}^{n} (1 + \lambda_{min} k)^{-(n-(j-1))} \left(\|w_1^j\|_{L^2} + \|w_2^j\|_{L^2} \right).$$

Mit $t_n = nk$ und $t_{n-(j-1)} = k(n-(j-1))$ ergibt sich, dass man

$$(1+\lambda_{min}k)^{-n} = (1+\lambda_{min}k)^{-t_n\lambda_{min}/(\lambda_{min}k)}$$
$$(1+\lambda_{min}k)^{-(n-(j-1))} = (1+\lambda_{min}k)^{-t_{n-j+1}\lambda_{min}/(\lambda_{min}k)}$$

abschätzen muss. Elementare Überlegungen zeigen, dass die Funktion

$$x \mapsto (1+x)^{-1/x}$$

monoton wachsend ist. Aus $k \le k_0$ folgt somit, dass $\lambda_{min}k \le \lambda_{min}k_0$ ist, d.h.

$$(1+\lambda_{min}k)^{-n} = (1+\lambda_{min}k_0)^{-t_n\lambda_{min}/(\lambda_{min}k_0)} \le e^{-bt_n}$$
$$(1+\lambda_{min}k)^{-(n-(j-1))} = (1+\lambda_{min}k_0)^{-t_{n-j+1}\lambda_{min}/(\lambda_{min}k_0)} \le e^{-b(t_n-t_{j-1})},$$

wobei $b > 0$ durch die Beziehung $(1+\lambda_{min}k_0)^{-\lambda_{min}/(\lambda_{min}k_0)} = e^{-b}$ definiert ist. Damit erhalten wir

$$\|\theta^n\|_{L^2} \le e^{-bt_n}\|\theta^0\|_{L^2} + k\sum_{j=1}^{n} e^{-b(t_n-t_{j-1})}\left(\|w_1^j\|_{L^2} + \|w_2^j\|_{L^2}\right)$$

Wegen

$$\|w_1^j\|_{L^2} \le \frac{1}{k}\int_{t_{j-1}}^{t_j} \|u'(t) - R_N u'(t)\|_{L^2} dt, \qquad \|w_2^j\|_{L^2} \le \frac{k}{k}\int_{t_{j-1}}^{t_j} \|u''(t)\|_{L^2} dt,$$

folgt damit

$$\|\theta^n\|_{L^2} \le e^{-bt_n}\|\theta^0\|_{L^2} + \sum_{j=1}^{n} e^{-b(t_n-t_{j-1})}\int_{t_{j-1}}^{t_j} \|u'(t) - R_N u'(t)\|_{L^2} + k\|u''(t)\|_{L^2} dt \le$$
$$\le e^{-bt_n}\|\theta^0\|_{L^2} + \sum_{j=1}^{n}\int_0^{t_n} e^{-b(t_n-t)}\|u'(t) - R_N u'(t)\|_{L^2} + k\|u''(t)\|_{L^2} dt$$

□

Korollar 8.13 *Sei $u \in C^3([0,T]\times\bar{\Omega})$. Dann gilt:*

(i) $\|u_N^n - u(t_n)\|_{L^2} \le C[h+k]$

(ii) falls Ω konvex ist, dann ist $\|u_N^n - u(t_n)\|_{L^2} \le C[h^2+k]$

Beweis: Der Beweis verbleibt als Übung. □

Wir betrachten nun das *explizite Eulerverfahren* und stellen uns die Frage, wie groß k (relativ zur Ortsdiskretisierung h) sein muss, damit die u_N^n vernünftige Approximationen an $u(t_n)$ sind.

Das explizite Eulerverfahren ist von der Form

$$\frac{1}{k}\langle u_N^{n+1} - u_N^n, v\rangle_{L^2} + a(u_N^n, v) = \langle f(t_n), v\rangle_{L^2} \qquad \forall\ v \in V_N \tag{8.13}$$

oder in Matrixschreibweise

$$\mathbf{M}\mathbf{u}^{n+1} = (\mathbf{M} - k\mathbf{A})\mathbf{u}^n + k\mathbf{f}^n \tag{8.14}$$

$$\Rightarrow \mathbf{u}^{n+1} = \mathbf{u}^n - k\mathbf{M}^{-1}\mathbf{A}\mathbf{u}^n + k\mathbf{M}^{-1}\mathbf{f}^n$$

Wir erinnern an den Beweis der Konvergenz von Einschrittverfahren. Diese haben die Form $y_{i+1} = \psi(t_i, y_i, k_i)$, wobei k_i die Schrittweite im i-ten Schritt bezeichnet. Der Beweis beruhte auf 2 Komponenten

(i) *Konsistenz*, d.h. kleiner Fehler in jedem Schritt

(ii) *Stabilität* durch Lipschitzstetigkeit der Inkrementsfunktion. Dazu wurde benötigt (L bezeichnet die Lipschitzkonstante):

$$\left.\begin{array}{l} y_{i+1} = \psi(t_i, y_i, k_i) \\ \tilde{y}_{i+1} = \psi(t_i, \tilde{y}_i, k_i) \end{array}\right\} \textit{ impliziert } |y_{i+1} - \tilde{y}_{i+1}| \leq (1 + Lk_i)\,|y_i - \tilde{y}_i|$$

Bemerkung: Stabilität ist ein Maß für die Fehlerfortpflanzung. Ein Fehler im i-ten Schritt wird höchstens um den Faktor $(Lk_i + 1)$ verstärkt.

Auf (8.14) übertragen heißt das, $\|\mathbf{I} - k\mathbf{M}^{-1}\mathbf{A}\|$ zu untersuchen. Eine sinnvolle Norm ist diejenige, die zur L^2-Norm gehört, d.h. $\|\cdot\|_{\mathbf{M}}$ (da $\mathbf{u} = \sum_i \mathbf{u}_i\varphi_i = \|u\|_{L^2}^2 = \mathbf{u}^T\mathbf{M}\mathbf{u} = \langle \mathbf{u}, \mathbf{u}\rangle_{\mathbf{M}}$).

Lemma 8.14

Sei $\mathbf{V}$*, sodass* $\mathbf{V}^T\mathbf{M}\mathbf{V} = \mathbf{I}$*,* $\mathbf{V}^T\mathbf{A}\mathbf{V} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{pmatrix}$ *Dann gilt:*

$$\|\mathbf{I} - k\mathbf{M}^{-1}\mathbf{A}\|_{\mathbf{M}} = \max_{i=1,\dots,N} |1 - \lambda_i k_i|$$

Beweis: Für $x = \mathbf{V}y$ gilt

$$\|x\|_{\mathbf{M}}^2 = \langle \mathbf{M}\mathbf{V}y, \mathbf{V}y\rangle_2 = \langle \mathbf{V}^T\mathbf{M}\mathbf{V}y, y\rangle_2 = \|y\|_2^2 = \|\mathbf{V}^{-1}x\|_2.$$

Weiters gilt $\mathbf{M}^{-1} = \mathbf{V}\mathbf{V}^T$ und $\mathbf{V}^T\mathbf{A} = \mathbf{D}\mathbf{V}^{-1}$. Damit ergibt sich:

$$(\mathbf{I} - k\mathbf{M}^{-1}\mathbf{A}) = \mathbf{I} - k\mathbf{V}\mathbf{V}^T\mathbf{A} = \mathbf{I} - k\mathbf{V}\mathbf{D}\mathbf{V}^{-1} = \mathbf{V}\mathbf{V}^{-1} - k\mathbf{V}\mathbf{D}\mathbf{V}^{-1} = \mathbf{V}\,(\mathbf{I} - k\mathbf{D})\,\mathbf{V}^{-1}$$

woraus wir folgern:

$$\begin{aligned} \|(\mathbf{I} - k\mathbf{M}^{-1}\mathbf{A})\|_{\mathbf{M}} &= \sup_{x\neq 0} \frac{\|(\mathbf{I} - k\mathbf{M}^{-1}\mathbf{A})x\|_{\mathbf{M}}}{\|x\|_{\mathbf{M}}} \sup_{x\neq 0} \frac{\|\mathbf{V}^{-1}(\mathbf{I} - k\mathbf{M}^{-1}\mathbf{A})x\|_2}{\|\mathbf{V}^{-1}x\|_2} \\ &= \sup_{x\neq 0} \frac{\|\mathbf{V}^{-1}\mathbf{V}(\mathbf{I} - k\mathbf{D})\mathbf{V}^{-1}x\|_2}{\|\mathbf{V}^{-1}x\|_2} = \sup_{y\neq 0} \frac{\|(\mathbf{I} - k\mathbf{D})y\|_2}{\|y\|_2} = \|\mathbf{I} - k\mathbf{D}\|_2 = \max_{i=1,\dots,N} |1 - k\lambda_i|. \end{aligned}$$

□

Für Stabilität fordern wir $\|\mathbf{I} - k\mathbf{M}^{-1}\mathbf{A}\|_{\mathbf{M}} \leq 1 + Lk$, wobei L eine „moderate“ Konstante ist. Weil $\lambda_i > 0$ für alle $i = 1, \dots, N$ ist, folgt also

$$1 + kL \geq \max_{i=1,\dots,N} |1 - k\lambda_i| = \max\{|1 - k\lambda_{min}|, |1 - k\lambda_{max}|\}$$

$$\Rightarrow 1 + kL \geq |1 - k\lambda_{min}| \wedge 1 + kL \geq |1 - k\lambda_{max}|$$

Die Forderung $1 + kL \geq |1 - k\lambda|$ für $\lambda = \lambda_{min}$ bzw $\lambda = \lambda_{max}$ impliziert

$$-(1 + kL) \leq \underbrace{1 - k\lambda}_{\leq 1} \leq 1 + kL$$

Für λ_{max} folgt damit

$$\begin{aligned} &\Rightarrow k(L + \lambda_{max}) \leq 2 \\ &\Leftrightarrow k \leq \frac{2}{L + \lambda_{max}} \end{aligned}$$

Weil lt. Satz 8.11 $\lambda_{max} \sim \frac{C}{h_{min}^2}$ und damit sehr groß ist, ergibt sich praktisch die sogenannte Courant-Friedrichs-Lewy (CFL) Bedingung:

$$k\lambda_{max} \leq 2 \tag{8.15}$$

als Bedingung für die Schrittweite. Für regelmäßige Gitter gilt $\lambda_{max} \sim \frac{C}{h^2}$, d.h. dass das *explizite* Eulerverfahren der Schrittweitenbeschränkung $k \leq Ch^2$ unterliegt. Beim *implizitem* Eulerverfahren gab es keine Schrittweitenbeschränkung.

Das explizite und das implizite Eulerverfahren sind 1. Ordnung in der Zeit. Das *Crank-Nicholson-Verfahren* („implizite Mittelpunktregel“, θ-Schema mit $\theta = \frac{1}{2}$) ist 2. Ordnung in der Zeit. Zudem hat es genau wie das implizite Eulerverfahren keine Schrittweitenbeschränkungen für k.

Satz 8.15 *(Crank-Nicholson-Verfahren) Seien die u_N^n definiert durch*

$$\frac{1}{k}\langle u_N^{n+1} - u_N^n, v\rangle_{L^2} + a\left(\frac{u_N^{n+1} - u_N^n}{2}, v\right) = \langle f(t_n + \tfrac{k}{2}), v\rangle_{L^2} \qquad \forall\ v \in V_N$$

Dann gilt:

$$\|u_N^n - u(t_n)\|_{L^2} \leq \|u_{0,N} - R_N u_0\|_{L^2} + \|u(t_n) - R_N u(t_n)\|_{L^2} + \int_0^{t_n} \|u' - R_N u'\|_{L^2} + Ck^2\|u'''(t)\|_{L^2} dt$$

Beweis: Wie in Satz 8.12. Für die Stabilität verwendet man anstatt von $v = \theta^{n+1}$ die Testfunktion $v = \frac{1}{2}(\theta^{n+1} + \theta^n)$. □

8.3.1 Zusammenfasssendes Beispiel

Zum Abschluss fassen wir noch einige Ergebnisse bzgl. dem explizitem und implizitem Eulerverfahren und dem Crank-Nicholson-Verfahren zusammen. (Folien 17 aus der VO) Dazu betrachten wir die 1D-Wärmeleitungsgleichung

$$u_t - u_{xx} = f \qquad \textit{auf}\ \Omega = (0, 1)$$

mit der Randbedingung $u(0) = u(1) = 0$.

Die *Semidiskretisierung im Ort* mittels der FEM führt auf ein ODE-System der Form

$$\mathbf{M}\mathbf{u}' + \mathbf{A}\mathbf{u} = f.$$

Name	$\mathbf{P}$	$\sigma(\mathbf{P})$	$\varrho(\mathbf{P})$	stabil?
$\mathbf{P}_{exp}$	$\mathbf{I} - k\mathbf{M}^{-1}\mathbf{A})$	$\{1 - k\lambda \mid \lambda \in \sigma\}$	$\lvert 1 - k\lambda_{max}\rvert$	falls $k \leq \frac{2}{\lambda_{max}}$
$\mathbf{P}_{impl}$	$(\mathbf{M} + k\mathbf{A})^{-1}\mathbf{M}$	$\left\{\frac{1}{1+k\lambda} \mid \lambda \in \sigma\right\}$	$\frac{1}{\lvert 1+k\lambda_{min}\rvert} \leq 1$	für alle $k > 0$
$\mathbf{P}_{CN}$	$\left(\mathbf{M} + \frac{k}{2}\mathbf{A}\right)^{-1}\left(\mathbf{M} - \frac{k}{2}\mathbf{A}\right)$	$\left\{\frac{1-k/2\lambda}{1+k/2\lambda} \mid \lambda \in \sigma\right\}$	$\left\lvert\frac{1-k\lambda_{max}/2}{1+k\lambda_{max}/2}\right\rvert \leq 1$	für alle $k > 0$

Tabelle 8.1: Analyse von $\mathbf{P}$ in $\mathbf{u}^{n+1} = \mathbf{P}\mathbf{u}^n + \cdots$

Die *Zeitdiskretisierung* erfolgt mit einem der drei oben erwähnten Verfahren:

expliziter Euler	$\mathbf{M}(\mathbf{u}^{n+1} - \mathbf{u}^n) + k\mathbf{A}\mathbf{u}^n = k\mathbf{f}(t_n)$
impliziter Euler	$\mathbf{M}(\mathbf{u}^{n+1} - \mathbf{u}^n) + k\mathbf{A}\mathbf{u}^{n+1} = k\mathbf{f}(t_{n+1})$
Crank-Nicholson	$\mathbf{M}(\mathbf{u}^{n+1} - \mathbf{u}^n) + \frac{k}{2}(\mathbf{A}\mathbf{u}^n + \mathbf{A}\mathbf{u}^{n+1}) = k\mathbf{f}(t_{n+\frac{1}{2}})$

oder in „expliziter“ Form angeschrieben: $\mathbf{u}^{n+1} = \mathbf{P}\mathbf{u}^n + \cdots$, wobei

$$\begin{aligned}
\mathbf{P}_{expl} &= \mathbf{I} - k\mathbf{M}^{-1}\mathbf{A} \\
\mathbf{P}_{impl} &= (\mathbf{M} + k\mathbf{A})^{-1}\mathbf{M} \\
\mathbf{P}_{CN} &= \left(\mathbf{M} + \tfrac{k}{2}\mathbf{A}\right)^{-1}\left(\mathbf{M} - \tfrac{k}{2}\mathbf{A}\right)
\end{aligned}$$

Damit nun das jeweilige Verfahren stabil ist, muss $\|\mathbf{P}\| \leq 1$ in einer geeigneten Norm sein. Dazu muss der Spektralradius $\varrho(\mathbf{P}) \leq 1$ sein. Das Spektrum des verallgemeinerten EWP ist durch $\sigma = \{\lambda \mid \exists\, x \neq 0 : \mathbf{A}\mathbf{x} = \lambda\mathbf{M}\mathbf{x}\}$ gegeben.

In Tabelle 8.1 sind das Spektrum, der Spektralradius und die aus dem Spektralradius ablesbare Forderung für Stabilität an die Schrittweite k aufgelistet. Die Ergebnisse von Abschnitt 8.3 werden dadurch bestätigt: Um stabile Verfahren zu erhalten, gibt es beim explizitem Eulerverfahren eine Schrittweitenbeschränkung, beim implizitem Eulerverfahren und beim Crank-Nicholson-Verfahren nicht.

Wir wählen nun den Anfangswert für die 1D Wärmeleitungsgleichung $u_0 = 1$ und $f \equiv 0$ und betrachten die graphisch dargestellten Ergebnisse aus Abbildung 8.1. In den oberen beiden Graphiken ist zu sehen wie sich das explizite Eulerverfahren zu zwei verschiedenen Schrittweiten knapp über bzw. knapp unter der Stabilitätsschranke verhält.

Links ist $k = \frac{2.001}{\lambda_{max}} \geq \frac{2}{\lambda_{max}}$ gewählt. Die Lösung oszilliert sehr stark (die größten Werte liegen im Bereich 10^7) und geben in keinster Weise das tatsächliche Lösungsverhalten wieder. In der rechten Graphik ist $k = \frac{1.999}{\lambda_{max}} \leq \frac{2}{\lambda_{max}}$ gewählt und die numerische und die exakte Lösung sind ziemlich ähnlich.

In der unteren linken Graphik von Abbildung 8.1 ist die Wärmeleitungsgleichung mit dem implizitem Eulerverfahren und in der rechten unteren Graphik mit dem Crank-Nicholson-Verfahren gelöst worden. Hier ist die numerische Lösung in beiden Fällen eine gute Approximation an die exakte Lösung.

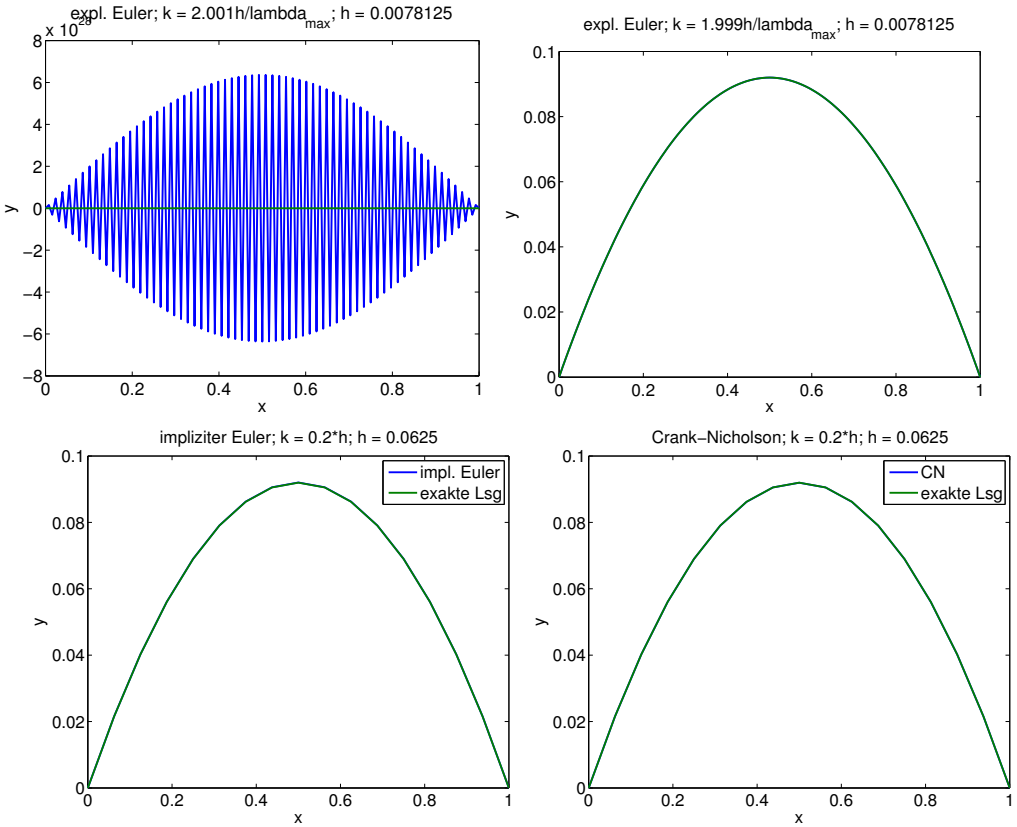

Abbildung 8.1: Vergleich zwischen exakter und numerischer Lösung

Zum Abschluss betrachten wir noch das Konvergenzverhalten für $u(x,t) = e^{-t}x(1-x)$. In der linken Graphik von Abbildung 8.2 wird das explizite Eulerverfahren betrachtet. Wieder mit den zwei Zeitschrittweiten $k = \frac{2.001}{\lambda_{max}}$ und $k = \frac{1.999}{\lambda_{max}}$ knapp über bzw. knapp unter der Stabilitätsschranke. Der Fehler für die Lösung zur Wahl von $k = \frac{1.999}{\lambda_{max}}$ verhält sich wie $\mathcal{O}(h)$. Liegt k aber nur knapp über $\frac{2}{\lambda_{max}}$ konvergiert das explizite Eulerverfahren nicht mehr.

Das Problem besteht in der Praxis darin, dass λ_{max} nicht einfach zu bestimmen ist. Würde man λ_{max} kennen, könnte man daraus k berechnen. Kann λ_{max} aber nicht (exakt) bestimmt werden, wirkt sich diese Ungenauigkeit auch auf das daraus berechnete k aus. Durch die sehr restriktive Schrittweitenbeschränkung, hängt die Konvergenz des Verfahrens aber schon von sehr kleinen Abweichungen ab. Man sagt dazu auch „hit or miss“ – also entweder hat man Glück und bestimmt k so, dass das Verfahren konvergiert, oder man liegt knapp daneben und das Verfahren konvergiert nicht.
In der rechten Graphik von Abbildung 8.2 ist zu sehen, dass sich das implizite ebenso wie das explizite Eulerverfahren wie $\mathcal{O}(h)$ verhält. und das Crank-Nicholson-Verfahren hat Konvergenzordnung 2, unabhängig von der Wahl von k. Das waren auch unsere Erwartungen, denn der Fehler des expl. sowie des implizite Eulerverfahrens ist $\mathcal{O}(k+h^2)$ und jener vom Crank-Nicholson-Verfahren ist $\mathcal{O}(k^2+h^2)$.

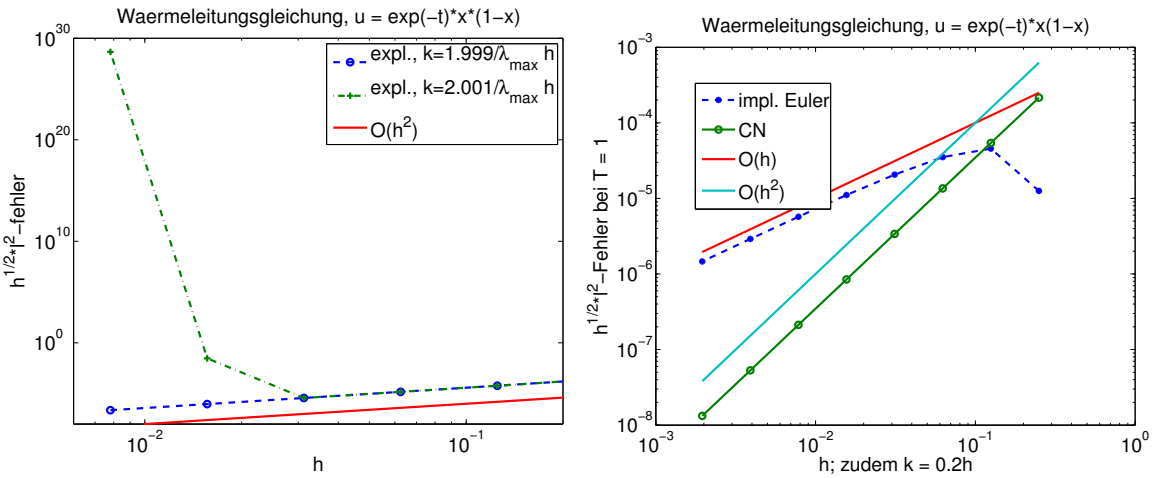

Abbildung 8.2: Konvergenzbetrachtung