Kapitel 4

Zeitdiskretisierung

In diesem Kapitel werden wir kurz ein paar Aspekte der Zeitdiskretisierung parabolischer und hyperbolischer Probleme diskutieren. Bei einer vertikalen Linienmethode hat man nach der Diskretisierung im Ort (mit Methoden analog zu Kapitel 2) ein System gewöhnlicher Differentialgleichungen in der Zeit zu diskretisieren, und es werden tatsächlich dieselben Methoden wie bei gewöhnlichen Differentialgleichungen verwendet. Allerdings ist bei partiellen Differentialgleichungen besondere Vorsicht geboten wegen der Änderung mit der Diskretisierung h. Vor allem für feine Ortsdiskretisierung erhält man sogenannte *steife gewöhnliche Differentialgleichungen*, die besondere Vorsicht bezüglich der Stabilität erforden. Wir werden uns deshalb in diesem Kapitel vor allem mit der Stabilität der Verfahren befassen.

4.1 Parabolische Probleme

Ein abstraktes parabolisches Problem ist von der Form

$$\frac{\partial u}{\partial t} - Lu = f, \qquad \text{in } \Omega \times T \tag{4.1}$$

mit Anfangswerten $u(0) = u_0$, wobei L ein elliptischer Differentialoperator (mit Randbedingungen) ist. Wir nehmen an, dass L bzw. die zugehörige Bilinearform B in $H^m(\Omega)$ stabil ist. Im Falle eines Operators zweiter Ordnung gilt dann unter den typischen Voraussetzungen $L : H^1(\Omega) \to H^{-1}(\Omega)$. Damit sollten natürlich f und $\frac{\partial u}{\partial t}$ für fast alle t sinnvollerweise Elemente in $H^{-1}(\Omega)$ sein. Ein bisschen Vorsicht ist noch bezüglich der Abhängigkeit bezüglich t geboten. Dazu verwendet man normalerweise sogenannte *vektorwertige Sobolevräume*. Wir definieren

$$H^k(0,T;H^m(\Omega)) := \{\ u \in H^m(\Omega) \mid \frac{\partial u^j}{\partial t^j} \in H^m(\Omega), j = 1,\dots,k\ \}.$$

Analog schreiben wir $L^2(0,T;H^m(\Omega))$ im Fall $k = 0$ bzw. $H^k(0,T;L^2(\Omega))$ im Fall $m = 0$. Natürlich können wir den Raum auch für m negativ in der üblichen Weise definieren.

4.1.1 Schwache Formulierung

Um eine Idee für die schwache Formulierung zu erhalten, multiplizieren wir wieder mit einer Testfunktion v und integrieren, zunächst nur bezüglich x. Dann gilt

$$\langle \frac{\partial u}{\partial t}, v\rangle + \int_\Omega (-Lu)v\ dx = \int_\Omega fv\ dx.$$

Bis auf den Teil mit der Zeitableitung können wir alle Terme analog zur schwachen Formulierung im elliptischen Fall behandeln und erhalten

$$\langle \frac{\partial u}{\partial t}, v\rangle + B(u,v) = \ell(v).$$

Für $\frac{\partial u}{\partial t} \in H^{-m}(\Omega)$ und $v \in H^m(\Omega)$. Um eine Idee über die Zeitabhängigkeit zu erhalten, setzen wir wieder $v = u$ und integrieren über die Zeit. Dann gilt

$$\begin{aligned}\frac{1}{2}\|u(t)\|_2^2 + \int_0^t B(u(s),u(s))\ ds &= \frac{1}{2}\|u_0\|_2^2 + \int_0^t \langle(\frac{\partial u}{\partial t}, u\rangle + B(u(s),u(s)))\ ds \\ &= \int_0^t fu(s)\ ds + \frac{1}{2}\|u_0\|_2^2.\end{aligned}$$

Ist nun B koerziv in $H^m(\Omega)$, dann folgt

$$\frac{1}{2}\|u(t)\|_2^2 + c\int_0^t \|u(s)\|_{m,2}^2\ ds \leq \frac{1}{2c}\int_0^t \|f\|_{-m,2}^2\ ds + \frac{c}{2}\int_0^t \|u(s)\|_{m,2}^2\ ds + \frac{1}{2}\|u_0\|_2^2,$$

und da $t \leq T$ beliebig ist, folgt

$$\sup_t \|u(t)\|_2^2 \leq \frac{1}{c}\int_0^T \|f\|_{-m,2}^2\ ds + \|u_0\|_2^2$$

und

$$\int_0^T \|u(s)\|_{m,2}^2\ ds \leq \frac{1}{c^2}\int_0^T \|f\|_{-m,2}^2\ ds + \frac{1}{c}\|u_0\|_2^2.$$

Damit folgt $u \in L^\infty(0,T;L^2(\Omega)$ (mit der offensichtlichen Definition dieses Raums) und $u \in L^2(0,T;H^m(\Omega)$ mit jeweiliger Stabilitätsabschätzung.

Eine Abschätzung für die Zeitableitung erhält man mit der Testfunktion $v = (-L)^{-1}\frac{\partial u}{\partial t}$ (man beachte dass unter den üblichen Elliptizitätsvoraussetzungen $-L$ ein regulärer linearer Operator von $H^m(\Omega)$ nach $H^{-m}(\Omega)$ ist). Es folgt dann

$$B(v,v) + \langle u, \frac{\partial u}{\partial t}\rangle = \langle\frac{\partial u}{\partial t}, (-L)^{-1}\frac{\partial u}{\partial t}\rangle + \langle(-L)u, (-L)^{-1}\frac{\partial u}{\partial t}\rangle = \langle f, (-L)^{-1}\frac{\partial u}{\partial t}\rangle$$

und mit der Koerzivität folgt nach Zeitintegration

$$c\int_0^t \|v(s)\|_{m,2}^2\ ds + \frac{1}{2}\|u(t)\|_2^2 \leq \frac{1}{2c}\int_0^t \|f\|_{-m,2}^2\ ds + \frac{c}{2}\int_0^t \|v(s)\|_{m,2}^2\ ds + \frac{1}{2}\|u_0\|_2^2.$$

Damit erhalten wir eine Stabilitätsabschätzung für $v = (-L)^{-1}\frac{\partial u}{\partial t} \in L^2(0,T;H^m(\Omega))$. Da $(-L)^{-1}$ ein stetiger Operator von $H{-}m(\Omega)$ nach $H^m(\Omega)$ ist, folgt also $\frac{\partial u}{\partial t} \in L^2(0,T;H^{-m}(\Omega))$.

Eine vernünftige schwache Lösung des parabolischen Problems mit Anfangswert $u_0 \in L^2(\Omega)$ sollte also

$$u \in L^\infty(0,T;L^2(\Omega)) \cap L^2(0,T;H^m(\Omega)) \cap H^1(0,T;H^{-1}(\Omega))$$

erfüllen. Eine analoge Stabilität sollte man dann auch vom diskretisierten Problem erwarten.

4.1.2 Maximumprinzip

Ist $u_0 \in L^\infty(\Omega)$ und der Operator L zweiter Ordnung, dann kann man auch wieder Maximumprinzipien anwenden und daraus eine Lösung $u \in L^\infty((0,T) \times \Omega)$ mit entsprechender Stabilitätsabschätzung erwarten. Das Maximumprinzip gilt analog zum elliptischen Fall, sogar noch mit einigen Erweiterungen:

Satz 4.1. *Sei L ein Operator zweiter Ordnung wie in Kapitel 2 mit $c = 0$. Weiter sei $u \in C(0,T;C^2(\Omega)) \cap C^1(0,T;C(\Omega))$ so dass*

$$\frac{\partial u}{\partial t} - Lu < 0(> 0) \qquad \text{in } \Omega \times (0,T)$$

gilt. Dann nimmt u kein Maximum (Minimum) in $\Omega \times (0,T]$ an.

Proof. Angenommen ein Maximum wird in $(x,t) \in \Omega \times (0,T)$ angenommen. Dann gilt dort $\frac{\partial u}{\partial t}(x,t) = 0$ und wie in Kapitel 2 $Lu(x,t) \leq 0$. Dies ergibt bei Einsetzen in die Gleichung einen Widerspruch. Im Fall eines Maximums (x,T) würde ein Randmaximum bezüglich der Zeit vorliegen und damit gilt zumindest $\frac{\partial u}{\partial t}(x,t) \geq 0$ und wegen der Maximalität im Ort gilt $Lu(x,t) \leq 0$. Einsetzen in die Gleichung führt auch hier zu einem Widerspruch. □

Analoge Aussagen lassen sich natürlich auch für $c \geq 0$ zeigen, einige wie z.B. die Erhaltung von Positivität sind sogar für negatives c möglich. Dazu verwendet man die Transformation $v = e^{-\alpha t} u$ mit α hinreichend gross. Es gilt nämlich

$$\frac{\partial u}{\partial t} - Lu = e^{\alpha t} \left(\frac{\partial v}{\partial t} + \alpha v - Lv \right)$$

und für $\alpha > -\max c$ ist der Koeffizient niedrigster Ordnung für v sicher positiv. Damit kann man Maximumprinzipien für v anwenden, und da u dasselbe Vorzeichen wie v hat auch wieder aus Positivität oder Negativität von u schließen.

4.1.3 Ortsdiskretisierung

Wir betrachten nun ein im Ort diskretisiertes Problem. Dazu wenden wir entweder direkt finite Differenzen (mit zeitabhängigen Werten an den Gitterpunkten) oder finite Elemente (mit zeitabhängigen Koeffizienten) für die schwache Formulierung

$$\langle \frac{\partial u}{\partial t}, v \rangle + \int_\Omega (-Lu) v \; dx = \int_\Omega f v \; dx.$$

m Fall finiter Differenzen führt dies auf ein System gewöhnlicher Differentialgleichungen

$$\frac{dU^h}{dt} + K_h U^h = F_h, \qquad U^h(0) = U_0^h. \tag{4.2}$$

Bei einer finite Elemente Diskretisierung im Ort erhält man

$$M_h \frac{dU^h}{dt} + K_h U^h = F_h, \qquad U^h(0) = U_0^h \tag{4.3}$$

mit der Massenmatrix

$$(M_h)_{ij} = \langle \varphi_i, \varphi_j \rangle = \int_\Omega \varphi_i(x) \varphi_j(x) \; dx.$$

Die Matrix K_h (und auch M_h) ist unter den üblichen Annahmen symmetrisch positiv definit. Damit kann man (für theoretische Zwecke) die Matrix K_h diagonalisieren. Es existieren eine orthogonale Matrix S_h und eine Diagonalmatrix Σ_h, sodass $S_h^T K_h S_h = \Sigma_h$ gilt. Wir definieren nun $V^h = S_h^T U^h$, dann folgt

$$\frac{dV^h}{dt} + \Sigma_h V^h = G_h, \qquad V^h(0) = V_0^h \tag{4.4}$$

mit $G_h = S_h^T F_h$ und $V_0^h = S_h^T U_h^0$. Analog zu Kapitel 1 folgt $\|U^h\| = \|V^h\|$, d.h. L^2-Stabilität kann auch bezüglich V^h untersucht werden. Bei einer finite Elemente Diskretisierung kann man analog vorgehen, wenn man zuerst von links und rechts mit $(M_h^{-1/2})$ multipliziert (so eine Matrix existiert, da M_h symmetrisch positiv definit ist) und danach die Diagonalisierung anwendet.

4.1.4 Zeitdiskretisierungen: Explizit, Implizit und Mehrschritt

Wir führen nun drei verschiedene Zeitdiskretisierungen durch Differenzenquotienten in der Zeit ein. Wir bezeichnen dazu mit $U_k^{h,\tau}$ die diskrete Lösung zum Zeitpunk $t_k = k\tau$. Die erste Wahl ist das *explizite Euler-Verfahren*

$$U_{k+1}^{h,\tau} = U_k^{h,\tau} + \tau(-K_h U_k^{h,\tau} + F_h(t_k)). \tag{4.5}$$

Die erste Alternative dazu ist das *implizite Euler-Verfahren*

$$U_{k+1}^{h,\tau} + \tau K_h U_{k+1}^{h,\tau} = U_k^{h,\tau} + \tau F_h(t_{k+1}). \tag{4.6}$$

Ein drittes interessantes Verfahren erhält man aus dem Mittelwert der beiden obigen Formeln: das sogenannte *Crank-Nicholson Verfahren*

$$U_{k+1}^{h,\tau} + \frac{\tau}{2} K_h U_{k+1}^{h,\tau} = U_k^{h,\tau} - \frac{\tau}{2} K_h U_k^{h,\tau} + \frac{\tau}{2}(F_h(t_k) + F_h(t_{k+1})). \tag{4.7}$$

Bei einer finite Elemente Diskretisierung im Ort sehen alle drei Verfahren genauso aus, es werden nur die Terme, die aus der Zeitableitung resultieren, mit M_h multipliziert.

Wir sehen sofort, dass diese Verfahren eigentlich nur drei Beispiele eines allgemeineren Verfahrens mit Parameter $\theta \in [0,1]$ sind

$$U_{k+1}^{h,\tau} + \theta\tau K_h U_{k+1}^{h,\tau} = U_k^{h,\tau} - (1-\theta)\tau K_h U_k^{h,\tau} + \tau((1-\theta)F_h(t_k) + \theta F_h(t_{k+1}))). \tag{4.8}$$

Für $\theta = 0$ erhalten wir das explizite Euler-Verfahren, für $\theta = 1$ das implizite Euler-Verfahren, und für $\theta = \frac{1}{2}$ das Crank-Nicholson Verfahren. Andererseits kann man dieses Verfahren für $\theta \in (0,1)$ als Mehrschrittverfahren, wiederum zusammengesetzt aus explizitem und implizitem Euler-Verfahren interpretieren, d.h. man berechnet eigentlich einen Zwischenschritt $U_{k+1-\theta}^{h,\tau}$ mit Zeitschrittweite $(1-\theta)\tau$ explizit und danach $U_{k+1}^{h,\tau}$ implizit mit Zeitschrittweite $\theta\tau$

$$U_{k+1-\theta}^{h,\tau} = U_k^{h,\tau} + (1-\theta)\tau(-K_h U_k^{h,\tau} + F_h(t_k)) \tag{4.9}$$

$$U_{k+1}^{h,\tau} + \theta\tau K_h U_{k+1}^{h,\tau} = U_{k+1-\theta}^{h,\tau} + \theta\tau F_h(t_{k+1}). \tag{4.10}$$

4.1.5 Konsistenz

Wir beginnen unsere Untersuchungen mit der Konsistenz der Zeitdiskretisierung. Sei U^h Lösung des semidiskreten Problems (4.2) und $U_k^{h,\tau}$ Lösung des voll diskretisierten Problems (4.8). Aus einer Taylor-Entwicklung erhalten wir (man beachte in der zweiten Gleichung die zusätzliche Approximation $\frac{d^2U^h}{dt^2}(t_{k+1}) = \frac{d^2U^h}{dt^2}(t_k) + \mathcal{O}(\tau)$)

$$U^h(t_{k+1}) - U^h(t_k) = \tau \frac{dU^h}{dt}(t_k) + \frac{\tau^2}{2}\frac{d^2U^h}{dt^2}(t_k) + \mathcal{O}(\tau^3)$$
$$U^h(t_{k+1}) - U^h(t_k) = \tau \frac{dU^h}{dt}(t_{k+1}) - \frac{\tau^2}{2}\frac{d^2U^h}{dt^2}(t_k) + \mathcal{O}(\tau^3).$$

Nun multiplizieren wir die erste Gleichung mit $\frac{1-\theta}{\tau}$, die zweite mit $\frac{\theta}{\tau}$ und addieren sie. Damit erhalten wir

$$\frac{1}{\tau}(U^h(t_{k+1}) - U^h(t_k)) - (1-\theta)\frac{dU^h}{dt}(t_k) - \theta\frac{dU^h}{dt}(t_{k+1}) = \frac{\tau}{2}(1-2\theta)\frac{d^2U^h}{dt^2}(t_k) + \mathcal{O}(\tau^2).$$

Nun setzen wir noch (4.2) ein und erhalten

$$\begin{aligned}\frac{1}{\tau}(U^h(t_{k+1}) - U^h(t_k)) &+ (1-\theta)K_hU^h(t_k) + \theta K_hU^h(t_{k+1})\\ &= (1-\theta)F_h(t_k) + \theta F_h((t_{k+1}) + \frac{\tau}{2}(1-2\theta)\frac{d^2U^h}{dt^2}(t_k) + \mathcal{O}(\tau^2).\end{aligned}$$

Also ist der Fehler bei Einsetzen von U^h in das voll diskrete Verfahren

$$\frac{\tau}{2}(1-2\theta)\frac{d^2U^h}{dt^2}(t_k) + \mathcal{O}(\tau^2).$$

Wir sehen also, dass explizites und implizites Euler-Verfahren einen Fehler erster Ordnung $\mathcal{O}(\tau)$ liefern. Das Crank-Nicholson Verfahren (und jedes Verfahren mit der Wahl $\theta = \frac{1}{2}+\mathcal{O}(\tau)$) liefert eine Fehler zweiter Ordnung $\mathcal{O}(\tau^2)$, also eine höhere Konsistenzordnung.

4.1.6 Stabilität

Im folgenden werden wir uns der Stabilitätsanalyse der obigen zeitdiskretisierten Verfahren widmen. Dies werden wir in drei verschiedenen Versionen durchführen: bezüglich L^2-Stabilität mittels Diagonalisierung der Matrix K_h, bezüglich L^∞-Stabilität mittels M-Matrix Techniken für finite Differenzen und bezüglich H^1-Stabilität mittels Variationstechniken für finite Elemente.

L^2-Stabilität

Wie auch schon im Kapitel über finite Elemente Methoden diskutiert, ist die L^2-Norm einer Funktion u^h auf einem Gitter proportional zu h^dmultipliziert mit der euklidischen Norm des Vektors U^h von Funktionswerten am Gitter. Deshalb führen wir hier die Stabilitätsanalyse bezüglich der euklidischen Norm des Vektors $U_k^{h,\tau}$ durch.

Wie oben verwenden wir die Diagonalisierung $S_h^T K_h S_h = \Sigma_h$ gilt. Wir definieren wieder $V_k = S_h^T U_k^{h,\tau}$, und erhalten das diagonalisierte Verfahren

$$V_{k+1} + \theta\tau\Sigma_h V_k = V_k + (1-\theta)\tau\Sigma_h V_k + \theta\tau G_{k+1} + (1-\theta)\tau G_k, \qquad (4.11)$$

mit $G_k = S_h^T F_h(t_k)$. Da Σ_h eine Diagonalmatrix aus den Eigenwerten λ_j. Damit können wir $(V_{k+1})_j$ aus der Iterationsformel

$$(V_{k+1})_j = \frac{1-\tau(1-\theta)\lambda_j}{1+\tau\theta\lambda_j}(V_k)_j + \frac{\tau}{1+\tau\theta\lambda_j}(\theta(G_{k+1}) + (1-\theta)(G_k)_j).$$

Entscheidend für die Stabilität oder Instabilität des Verfahrens ist der Faktor dieser geometrischen Reihe. Das Verfahren wird stabil genau dann, wenn

$$-1 \leq \frac{1-\tau(1-\theta)\lambda_j}{1+\tau\theta\lambda_j} \leq 1$$

für alle j gilt. Die rechte Ungleichung ist wegen der Positivität der Eigenwerte λ_j offensichtlich für beliebige Werte τ und $\theta \in [0,1]$ erfüllt. Die linke Ungleichung ist äquivalent zu

$$1 + \tau\theta\lambda_j \geq -1 + \tau(1-\theta)\lambda_j$$

bzw.

$$2 \geq (1-2\theta)\tau\lambda_j.$$

Wir sehen für $\theta \geq \frac{1}{2}$, dass die rechte Seite nichtpositiv und damit immer kleiner 2 ist. Für diese Werte von θ (und damit auch für das implizite Euler und Crank-Nicholson Verfahren) gilt also unbedingte Stabilität.

Für $\theta < \frac{1}{2}$ erhalten wir nur bedingte Stabilität mit der Einschränkung

$$\tau \leq \frac{2}{(1-2\theta)\max_j \lambda_j}.$$

Nun erinnern wir uns, dass für K_h resultierend aus der Ortsdiskretisierung eines elliptischen Differentialoperators zweiter Ordnung der maximale Eigenwert $\max_j \lambda_j$ von der Ordnung $\frac{1}{h^2}$ und damit ist die Zeitschrittbeschränkung von der Form $\tau = \mathcal{O}(h^2)$, eine meist zu starke Einschränkung. Damit wird auch das explizite Euler Verfahren weniger attraktiv, denn es ist zwar jeder Zeitschritt sehr einfach durchzuführen, man benötigt aber wegen der kleinen Wahl von τ sehr viele Zeitschritte im Vergleich zu einem impliziten oder zum Crank-Nicholson Verfahren.

L^∞-Stabilität

Eine speziell für Ortsdiskretisierungen mit finiten Differenzen sehr interessante Untersuchung ist die Stabilität in der Supremumsnorm, die wir ja auch im elliptischen Fall durchgeführt haben. Wir beschränken uns dabei der Einfachheit halber auf $\theta = 1$ und $\theta = 0$, d.h. explizites und implizites Euler-Verfahren.

Wir nehmen an, dass die Ortsdiskretisierung wie in Kapitel 2 eine M-Matrix K_h liefert. Damit ist natürlich auch die Matrix $A_h = I + \tau K_h$ eine M-Matrix (die Nebendiagonalterme bleiben gleich und negativ, die Diagonalterme werden sogar grösser). Also gilt für das implizite Euler Verfahren

$$U_{k+1}^{h,\tau} = A_h^{-1} U_k^{h,\tau} + \tau A_h^{-1} F_h(t_{k+1}).$$

Wegen der M-Matrix Eigenschaft gilt ein Maximumprinzip und damit Stabilität für A_h^{-1}:

$$\|U_{k+1}^{h,\tau}\|_\infty \leq \|A_h^{-1}\|(\|U_k^{h,\tau}\|_\infty + \tau\|F_h(t_{k+1})\|_\infty),$$

wobei für A_h^{-1} als passende Norm zur Supremumsnorm der Vektoren die Zeilensummennorm

$$\|B\| = \max_i \sum_j |B_{ij}|$$

verwendet wird. Die spezielle Gestalt der Matrix A_h als summe einer M-Matrix und der Einheitsmatrix liefert aber sogar noch mehr, nämlich die Konstante eins für $\|A_h^{-1}\|$ in der Stabilitätsabschätzung, wie wir im folgenden Lemma sehen werden:

Lemma 4.2. *Sei $B = (I + C)^{-1}$ für eine M-Matrix $C \in \mathbb{R}^{n\times n}$. Dann gilt*

$$\|B\| = \max_i \sum_j |B_{ij}| \leq 1.$$

Proof. Sei $G \in \mathbb{R}^n$ beliebig und $V = BG$, also $(I + C)V = G$. Wir nehmen an es gelte

$$V_\ell > G_i \quad \text{und} \quad V_\ell \geq V_i$$

für alle $i \in \{1, \ldots, n\}$. Dann folgt aus der $\ell-$ten Zeile des Gleichungssystems

$$(I + C_{\ell\ell})V_\ell = G_\ell - \sum_{j\neq\ell} C_{j\ell}V_j.$$

Wegen der Nichtpositivität der Nebendiagonalelemente in der M-Matrix $C_{j\ell}$ und der Eigenschaften der Zeilensummen einer M-Matrix folgt

$$(I + C_{\ell\ell})V_\ell \leq G_\ell - \sum_{j\neq\ell} C_{j\ell}V_\ell \leq G_\ell + C_{\ell\ell}V_\ell.$$

Also folgt $V_\ell \leq G_\ell$, ein Widerspruch zur Annahme $V_\ell > G_\ell$. Damit muss gelten

$$\max_i V_i \leq \max_i G_i, \qquad \min_i V_i \geq \min G_i,$$

wobei man die zweite Ungleichung durch Anwendung des selben Arguments auf $-G$ und $-V$ erhält. Insgesamt ist also

$$\|G\|_\infty = \max_i |G_i| \geq \max_i |V_i| = \max_i |(BG)_i| = \max_i |\sum_j B_{ij}G_j|.$$

Diese Ungleichung gilt für beliebiges $G \in \mathbb{R}^n$, also wählen wir G so, dass $G_k = \operatorname{sign}(B_{kj})$ gilt, wobei der Index k so gewählt ist, dass

$$\sum_j |B_{kj}| = \max_i \sum_j |B_{ij}|.$$

Somit gilt

$$1 \geq \|G\|_\infty \geq \max_i |\sum_j B_{ij}G_j| \geq |\sum_j B_{kj}G_j| = \sum_j |B_{kj}| = \max_i \sum_j |B_{ij}|$$

und wir erhalten die gewünschte Abschätzung. □

Aus dieser Eigenschaft der Matrix A_h folgt die Abschätzung

$$\|U_{k+1}^{h,\tau}\|_\infty \leq \|U_k^{h,\tau}\|_\infty + \tau\|F_h(t_{k+1})\|_\infty.$$

Summieren wir über k in dieser Ungleichung ($k = 0, \ldots, K-1$), dann folgt

$$\|U_K^{h,\tau}\|_\infty \leq \|U_0^{h,\tau}\|_\infty + \tau \sum_{k=1}^{K} \|F_h(t_k)\|_\infty.$$

Die Norm der Lösung lässt sich also wieder durch die Norm von der Daten (Anfangswert und rechte Seite) abschätzen. Für $h \to 0$ und $\tau \to 0$ sollte (unter geeigneten Annahmen and u_0 und f) gelten

$$\|U_0^{h,\tau}\|_\infty \to \|u_0\|\infty$$

und

$$\tau \sum_{k=1}^{K} \|F_h(t_k)\|_\infty \to \int_0^T \|f(t)\|_\infty \, dt,$$

mit $T = \lim K\tau$. Also wird die Stabilitätsabschätzung auch unabhängig von h und τ.

Für das explizite Euler Verfahren gilt

$$U_{k+1}^{h,\tau} = (I - \tau K_h)U_k^{h,\tau} + \tau F_h(t_k).$$

Die Nebendiagonalelemente der Matrix $I - \tau K_h$ sind nun nichtnegativ, die Diagonalelemente nur unter der Bedingung $\tau(K_h)_{jj} \leq 1$. In diesem Fall ist die Zeilensummennorm

$$\|I - \tau K_h\|_\infty = \max_i \sum_j |(I - \tau K_h)_{jj}| = 1 - \tau \min_i \sum_j (K_h)_{jj} \leq 1.$$

Also folgt Stabilität in der Supremumsnorm, es gilt

$$\|U_{k+1}^{h,\tau}\|_\infty \leq \|U_k^{h,\tau}\|_\infty + \tau\|F_h(t_k)\|_\infty.$$

Beachtet man, dass $(K_h)_{jj} = \mathcal{O}(\frac{1}{h^2})$ gilt, so folgt wieder bedingte Stabilität mit Zeitschritten $\tau = \mathcal{O}(h^2)$.

H^1-Stabilität

Wir erinnern noch einmal an das allgemeine Verfahren im Fall einer finite Elemente Diskretisierung im Ort.

$$M_h U_{k+1}^{h,\tau} + \theta\tau K_h U_{k+1}^{h,\tau} = M_h U_k^{h,\tau} - (1-\theta)\tau K_h U_k^{h,\tau} + \tau((1-\theta)F_h(t_k) + \theta F_h(t_{k+1}))). \quad (4.12)$$

Hier ist M_h die Massenmatrix und K_h die Steifigkeitsmatrix, d.h.

$$\begin{aligned} (M_h)_{ij} &= \langle \varphi_i, \varphi_j \rangle = \int_\Omega \varphi_i \varphi_j \, dx \\ (K_h)_{ij} &= B(\varphi_i, \varphi_j) \end{aligned}$$

mit der selben Bilinearform B resultierend aus der schwachen Formulierung wie im elliptischen Fall. Die rechte Seite berechnet sich wieder aus

$$F_h(t) = (\ell(t; \varphi))_{i=1,\ldots,n}.$$

Man beachte, dass man auch leicht zeitabhängigkeit von M_h und K_h hier einbauen könnte (etwa wegen zeitlich veränderlicher Koeffizienten in der parabolischen Differentialgleichunge), was wir der Einfachheit halber aber ignorieren.

Wie schon im Fall elliptischer Gleichungen kann man auch zeitdiskretisierte parabolische Probleme der Form (4.12) in eine schwache Formulierung überführen und Variationsprinzipien herleiten. Wir definieren dazu wieder die Funktion

$$u_k^{h,\tau} := \sum_j (U_k^{h,\tau})_j \varphi_j. \tag{4.13}$$

Dann gilt

$$(M_h U_k^{h,\tau})_i = \sum_j (M_h)_{ij}(U_k^{h,\tau})_j = \sum_j \langle \varphi_i, \varphi_j\rangle (U_k^{h,\tau})_j = \langle \varphi_i, u_k^{h,\tau}\rangle,$$

und

$$(K_h U_k^{h,\tau})_i = \sum_j (K_h)_{ij}(U_k^{h,\tau})_j = \sum_j B(\varphi_i, \varphi_j)(U_k^{h,\tau})_j = B(\varphi_i, u_k^{h,\tau}).$$

Mit der obigen Definition der rechten Seite (und unter Ausnutzung der Symmetrie des Skalarprodukts und der Bilinearform) können wir also (4.12) äquivalent als

$$\langle u_{k+1}^{h,\tau}, \varphi_i\rangle + \theta\tau B(u_{k+1}^{h,\tau}, \varphi_i) = \langle u_k^{h,\tau}, \varphi_i\rangle - (1-\theta)\tau B(u_k^{h,\tau}, \varphi_i) + \tau((1-\theta)\ell(t_k;\varphi_i) + \theta\ell(t_{k+1};\varphi_i)),$$

für $i = 1, \ldots, n$ schreiben. Da die Funktionen φ_i eine Basis von X^h bilden, gilt dann wieder

$$\langle u_{k+1}^{h,\tau}, v\rangle + \theta\tau B(u_{k+1}^{h,\tau}, v) = \langle u_k^{h,\tau}, v\rangle - (1-\theta)\tau B(u_k^{h,\tau}, v) + \tau((1-\theta)\ell(t_k;v) + \theta\ell(t_{k+1};v))$$

für alle $v \in X^h$.

Zur Vereinfachung werden wir bei der folgenden Stabilitätsuntersuchung annehmen, dass ℓ nicht von der Zeit abhängt. In diesem Fall hat die parabolische Gleichung eine dissipative Struktur, da das Energiefunktional

$$E(u) = \frac{1}{2}B(u,u) - \ell(u)$$

in der Zeit nicht zunimmt. Um dies (formal) zu sehen verwendet man in der schwachen Form $v = \frac{\partial u}{\partial t}$ und erhält

$$0 = \langle \frac{\partial u}{\partial t}, \frac{\partial u}{\partial t}\rangle + B(u, \frac{\partial u}{\partial t}) - \ell(\frac{\partial u}{\partial t}) = \|\frac{\partial u}{\partial t}\|^2 + \frac{dt}{dE}(u).$$

Mit Integration über die Zeit liefert

$$E(u(S)) \leq \int_s^S \|\frac{\partial u}{\partial t}\|^2 \, dt + E(u(S)) = E(u(s)),$$

d.h. das Energiefunktional nimmt in der Zeit ab, solange $\frac{\partial u}{\partial t} \neq 0$ gilt.

In Analogie zum kontinuierlichen Fall verwenden wir auch in der schwachen Formulierung des diskretisierten Problems die zeitliche Veränderung $v = u_{k+1}^{h,\tau} - u_k^{h,\tau}$ als Testfunktion (zur Vereinfachung der Notation lassen wir den Index h, τ im folgenden weg):

$$\langle v, v\rangle + \theta\tau B(u_{k+1}, v) = -(1-\theta)\tau B(u_k, v) + \tau\ell(v).$$

Einsetzen der Definition von v liefert

$$\|v\|^2 + \theta\tau B(u_{k+1}, u_{k+1}) - (1-\theta)\tau B(u_k, u_k)) + (2\theta - 1)B(u_{k+1}, u_k) - \tau\ell(u_{k+1}) + \tau\ell(u_k) = 0.$$

Eine einfache Umstellung der Terme und Division durch τ liefert

$$\begin{aligned} \frac{1}{\tau}\|v\|^2 + E(u_{k+1}) - E(u_k) &= -(\theta - \frac{1}{2})(B(u_{k+1}, u_{k+1}) - 2B(u_{k+1}, u_k) + B(u_k, u_k)) + 2(\theta - 1)B(u_k, u_k) \\ &= -(\theta - \frac{1}{2})B(v, v) + 2(\theta - 1)B(u_k, u_k). \end{aligned}$$

Wegen der Koerzivität folgt $B(v, v) \geq c\|v\|_H^2 \geq 0$ und $B(u_k, u_k) \geq c\|u_k\|_H^2 \geq 0$, wobei $\|.\|_H$ die Norm des entsprechenden Sobolev-Raumes bezeichnet (etwa $\|.\|_{1,2}$ für einen Differentialoperator zweiter Ordnung). Wegen $\theta \leq 1$ folgt

$$\frac{1}{\tau}\|v\|^2 + E(u_{k+1}) \leq E(u_k) - (\theta - \frac{1}{2})B(v, v)$$

und für $\theta \geq \frac{1}{2}$ erhalten wir direkt die "Energiestabilität"

$$E(u_{k+1}) \leq \|v\|^2 + E(u_{k+1}) \leq E(u_k).$$

Für $\theta < \frac{1}{2}$ sind wieder zusätzliche Bedingungen nötig, um Stabilität zu erhalten. In diesem Fall kann man die Stetigkeit der Bilinearform für die Abschätzung (mit $\eta := \frac{1}{2} - \theta > 0$)

$$\frac{1}{\tau}\|v\|^2 + E(u_{k+1}) \leq E(u_k) + \eta B(v, v) \leq E(u_k) + \eta C\|v\|_H^2.$$

Energiestabilität folgt, falls

$$\|v\|_H^2 \frac{1}{\eta C\tau}\|v\|^2,$$

also wieder eine inverse Ungleichung gilt. In dem in Kapitel 2 untersuchten Fall (Differentialoperator zweiter Ordnung, Sobolevraum $H_0^1(\Omega)$ und stückweise lineare Ansatzfunktionen) gilt eine inverse Ungleichung der Form

$$\|v\|_H^2 \frac{\beta^2}{h^2}\|v\|^2,$$

also folgt die Energiestabilität mit der schon bekannten Bedingung

$$\tau \leq \frac{1}{\eta C\beta^2}h^2 = \mathcal{O}(h^2).$$

Abschließend betrachten wir die Energiestabilität noch etwas näher. Aus dem Nichtanstieg des Zielfunktionals folgt

$$E(u_k) \leq E(u_{k-1}) \leq \ldots E(u_0),$$

und damit

$$c\|u_k\|_H^2 \leq B(u_k, u_k) \leq 2E(u_0) + 2\ell(u_k) \leq 2E(u_0) + 2\|\ell\|\|u_k\|_H$$

bzw. äquivalent

$$c(\|u_k\|_H - \frac{1}{c}\|\ell\|)^2 \leq 2E(u_0) + \frac{1}{c}\|\ell\|^2.$$

Damit erhalten wir wieder eine Stabilitätsabschätzung

$$\|u_k\|_H \leq \frac{1}{c}\|\ell\| + \sqrt{\frac{2c}{E}(u_0) + \frac{1}{c^2}\|\ell\|^2},$$

d.h. wieder eine uniforme Abschätzung der Norm von u_k abhängig von den Daten (Anfangswert und Linearform).

4.1.7 Hyperbolische Probleme

In diesem Abschnitt diskutieren wir noch kurz die Zeitdiskretisierung hyperbolischer Probleme. Der Einfachheit halber beschränken wir uns auf den örtlich eindimensionalen Fall. Wir untersuchen hyperbolische Systeme erster Ordnung in der Form

$$\frac{\partial u}{\partial t} = C \frac{\partial u}{\partial x},$$

wobei nun $u : \Omega \times \mathbb{R}_+ \to \mathbb{R}^p$ und $C : \Omega \times \mathbb{R}_+ \to \mathbb{R}^{p \times p}$ symmetrisch ist. Eine Transformation hyperbolischer Gleichungen auf eine solche Form ist in grosser Allgemeinheit möglich. Beispiele dafür sind die in Kapitel 1 untersuchte Transportgleichung, die einfach dem skalaren Fall des obigen Systems entspricht, oder die Wellengleichung

$$\frac{\partial^2 v}{\partial t^2} = c^2 \frac{\partial^2 v}{\partial x^2},$$

mit einer positiven Wellenzahl c. Im letzeren Fall erhält man die Transformation auf ein System erster Ordnung mit den Variablen $u^1 = \frac{\partial v}{\partial t}$ und $u^2 = c\frac{\partial v}{\partial x}$, es gilt dann

$$\begin{aligned} \frac{\partial u^1}{\partial t} &= \frac{\partial^2 v}{\partial t^2} = c^2 \frac{\partial^2 v}{\partial x^2} = c \frac{\partial u^2}{\partial x} \\ \frac{\partial u^2}{\partial t} &= c \frac{\partial^2 v}{\partial x t} = c \frac{\partial u^1}{\partial x}. \end{aligned}$$

Die Matrix C hat für die Wellengleichung also die Form

$$C = \begin{pmatrix} 0 & c \\ c & 0 \end{pmatrix}.$$

C ist eine symmetrische Matrix mit den Eigenwerten $\pm c$.

Durch geeignete Wahl einseitiger Differenzenquotienten im Ort (siehe auch die Diskussion zur Transportgleichung in Kapitel 1) kann man eine Ortsdiskretisierung der Form

$$\frac{dU^h}{dt} = L_h U^h$$

mit Knotenwerten $U^h \in \mathbb{R}^{np}$ und einer schief-symmetrischen Matrix $L_h \in \mathbb{R}^{np \times np}$ erreichen, d.h. $(L_h)_{ij} = -(L_h)_{ji}$ bzw. $L_h^T = -L_h$. Wir illustrieren dies für den Fall der Wellengleichung und nehmen o.B.d.A. an, der Vektor U^h bestehe zuerst aus den n Gitterwerten von u^1 und danach den n Gitterwerten von u^2. Sei $v_j(t) = (U^h(t))_j$ und $w_j(t) = (U^h(t))_j$. Wir wählen für die Ortsableitung von v einen Vorwärts- und für jene von w einen Rückwärtsdifferenzenquotienten. Dann gilt

$$\frac{dv_j}{dt} = c(w_j - w_{j-1}), \qquad \frac{dw_j}{dt} = c(v_{j+1} - v_j).$$

Die Matrix L_h hat dann die Form

$$L_h = \begin{pmatrix} \mathbf{0} & c - D_+^T \\ cD_+ & \mathbf{0} \end{pmatrix},$$

wobei D_+ die Matrix für den Vorwärtsdifferenzenquotienten (siehe Kapitel 1) ist.

Schiefsymmetrische Matrizen liefern ein Gegenstück zu symmetrischen Matrizen: ihre Eigenwerte sind nicht reell, sondern rein imaginär. Dies sieht man aus der Identität

$$\lambda u^H u = u^H L_h u = -u^H L^H u = -(Lu)^H u = -\overline{\lambda} u^H u,$$

für einen Eigenwert λ mit zugehörigem Eigenvektor u. Eine schiefsymmetrische Matrix L_h kann komplex diagonalisiert werden. Mit einer Matrix Q (sodass $Q^H Q = QQ^H = I$) und einer Diagonalmatrix $i\Sigma$ bestehend aus den Imaginärwerten Eigenwerten λ_j gilt:

$$L_h = Q\Sigma Q^H.$$

Also können wir mit der Transformation $V^h = Q^H U^h$ das ortsdiskrete hyperbolische Problem zu

$$\frac{dV^h}{dt} = \Sigma V^h$$

diagonalisieren, und damit erfüllt jede Komponente die skalare Differentialgleichung

$$\frac{dV_j^h}{dt} = i\sigma_j V_j^h,$$

deren Lösung durch

$$V_j^h(t) = e^{i\sigma_j t})V_j^h(0)$$

gegeben ist. Wegen $|e^{i\sigma_j t}| = 1$ (dies folgt aus der Euler-Formel $e^{ia} = \cos a + i \sin a$) gilt $|V_j^h(t)| = |V_j^h(0)|$ und damit auch

$$|U^h(t)| = |V^h(t)| = |V^h(0)| = |U^h(0)|,$$

d.h. Erhaltung der Euklidischen Norm von U^h in der Zeit.

Zur Zeitdiskretisierung betrachten wir nun explizites und implizites Euler-Verfahren. Beim impliziten Euler-Verfahren berechnen wir einen Zeitschritt aus

$$U_{k+1}^h - \tau L_h U_{k+1}^h = U_k^h.$$

Die Diagonalisierung liefert dann

$$V_{k+1}^h - \tau i\Sigma V_{k+1}^h = V_k^h.$$

Komponentenweise erhalten wir

$$(V_{k+1}^h)_j = \frac{1}{1 - i\tau\sigma_j}(V_k^h)_j.$$

Wegen $|1 - i\tau\sigma_j| = \sqrt{1 + \tau^2\sigma_j^2} \geq 1$ folgt $|(V_{k+1}^h)_j| \leq |(V_k^h)_j|$ und daraus resultiert wieder die undbedingte Stabilität des Verfahrens.

Beim expliziten Euler-Verfahren berechnen wir einen Zeitschritt aus

$$U_{k+1}^h = U_k^h + \tau L_h U_k^h.$$

In diesem Fall ist nach der Diagonalisierung

$$V_{k+1}^h = V_k^h + \tau i\Sigma V_k^h,$$

komponentenweise erhalten wir

$$(V_{k+1}^h)_j = (1 + i\tau\sigma_j)(V_k^h)_j.$$

Der Verstärkungsfaktor ist in diesem Fall $1+i\tau\sigma_j$ und für $\sigma_j \neq 0$ gilt $|1+i\tau\sigma_j| = \sqrt{1+\sigma_j^2} > 1$ und damit wird dieses explizite Verfahren instabil. Bei einem System aus mehreren Gleichungen haben wir aber auch noch andere Möglichkeiten ein explizites Verfahren zu konstruieren. Wir schreiben dazu die Matrix L_h als Summe einer rechten oberen und einer linken unteren Dreiecksmatrix, d.h. (wegen der Schiefsymmetrie)

$$L_h = R_h - R_h^T.$$

Nun können wir ein explizites Verfahren auch als

$$U_{k+1}^h + \tau R_h^T U_{k+1}^h = U_k^h + \tau R_h U_k^h.$$

Man beachte, dass man in diesem Fall keine Matrix zu invertieren muss, denn R^T ist eine untere Dreiecksmatrix. Man kann also schrittweise die Komponenten von U_{k+1}^h explizit berechnen. Die Update-Formel kann als

$$U_{k+1}^h = (I + \tau R_h^T)^{-1}(I + \tau R_h)U_k^h$$

geschrieben werden, und für die Stabilität untersuchen wir wieder die Eigenwerte der Matrix $(I + \tau R_h^T)^{-1}(I + \tau R_h)$. Sei λ Eigenwert dieser Matrix mit Eigenvektor V, dann gilt

$$\lambda(V + \tau R_h^T V) = V + \tau R_h V.$$

Wir untersuchen nun der Einfachheit halber wieder den Fall der Wellengleichung, in dem wir $V = (V_1, V_2)$ mit $V_i \in \mathbb{R}^n$ schreiben können und

$$R_h = \begin{pmatrix} 0 & cD \\ 0 & 0 \end{pmatrix},$$

mit der Wellenzahl c und der Matrix D für einen einseitigen Differenzenquotienten. Zur einfacheren Notation schreiben wir $C = \tau c D$. Es gilt

$$\begin{pmatrix} I & 0 \\ C^T & I \end{pmatrix}^{-1} = \begin{pmatrix} I & 0 \\ -C^T & I \end{pmatrix}$$

und damit erfüllt die Lösung des Eigenwertproblems die Gleichung

$$\lambda V = \begin{pmatrix} I & C \\ -C^T & I - C^T C \end{pmatrix} V$$

bzw. mit $\mu = \lambda - 1$ folgt $\mu V_1 = C V_2$ und damit aus der zweiten Gleichung

$$\mu^2 V_2 = -(1+\mu) C^T C V_2 = -\lambda C^T C V_2.$$

Damit ist $\frac{\mu^2}{\lambda}$ Eigenwert der negativ definiten Matrix $-C^T C$ und es gilt

$$(\lambda - 1)^2 = -\gamma^2 \lambda$$

für ein $\gamma \in \mathbb{R}$ (Eigenwert von $C^T C$). Diese quadratische Gleichung hat die Lösungen

$$\lambda = \frac{\gamma^2}{2} - 1 \pm \frac{1}{2}\sqrt{\gamma^4 - 4\gamma^2}.$$

Für $\gamma^2 \leq 4$ ist die Wurzel imaginär und damit gilt

$$|\lambda|^2 = (1 - \frac{\gamma^2}{2}) + \gamma^2 - \frac{\gamma^4}{4} = 1,$$

d.h. das Verfahren wird stabil. Die Bedingung $\gamma \leq 2$ kann man wieder zu h und τ in Beziehung setzen, da ja γ^2 Eigenwert von $\tau^2 c^2 D^T D$ ist. Wegen der Skalierung der Matrix für einen Differenzenquotienten ist damit höchstens $\gamma \sim \frac{\tau c}{h}$. Die Stabilitätsbedingung lautet also

$$\tau = \mathcal{O}(\frac{h}{c}),$$

die sogenannte CFL-Bedingung (nach Courant, Friedrichs, Levy). Man erkennt dass diese Bedingung wesentlich weniger restriktiv ist als im parabolischen Fall, weshalb im hyperbolischen Fall bevorzugt explizite Verfahren der obigen Form benutzt werden.