13. Hochgenaue Methoden

13.1. Künstliche Viskosität

Wir suchen Methoden mit der TVD-Eigenschaft, die höhere Konsistenzordnung besitzen. Die Grundidee besteht darin, eine Verfahren höherer Ordnung zu modifizieren, indem man die numerische Dissipation erhöht, sodaß Oszillationen verschwinden.

> **Beispiel :** Das Lax-Wendroff-Verfahren für die Gleichung $u_t + au_x = 0$ mit einer künstlichen Viskosität, die klein genug ist, daß das entstehende Verfahren noch Konsistenzordnung 2 hat:
>
> $$U_j^{n+1} = U_j^n - \nu(U_{j+1}^n - U_{j-1}^n) + \left(\frac{1}{2}\nu^2 + kQ\right)(U_{j+1}^n - 2U_j^n + U_{j-1}^n)$$
>
> mit der Courantzahl $\nu = ak/h$ und der künstlichen Viskosität Q. Reicht dieses Ausmaß an künstlicher Viskosität? Die Antwort ist negativ.

Satz 13.1. *Die Konsistenzordnung einer linearen monotonieerhaltenden Methode ist höchstens 1.*

Beweis. Wir werden zeigen, daß jede lineare monotonieerhaltende Methode monoton ist. Daraus folgt, daß für lineare Methoden die 4 Stabilitätseigenschaften des vorigen Abschnittes äquivalent sind. Satz 13.1 folgt dann aus Satz 12.6.

Sei $V_j^n \geq U_j^n$ für alle j. Weiters sei e^j die Gitterfunktion mit $e_i^j = \delta_{ij}$. Dann gilt

$$V^n - U^n = \sum_{j=-\infty}^{\infty} (V_j^n - U_j^n)e^j$$

und für eine lineare Methode mit Lösungsoperator $\mathcal{H}$ daher

$$V^{n+1} - U^{n+1} = \mathcal{H}V^n - \mathcal{H}U^n = \sum_{j=-\infty}^{\infty} (V_j^n - U_j^n)\mathcal{H}e^j\,. \tag{13.1}$$

Nun definieren wir die monotone Gitterfunktion

$$W_j^n = \begin{cases} 0 & j < J, \\ 1 & j \geq J, \end{cases}$$

sowie $\tilde{W}^n$ durch $\tilde{W}_j^n = W_{j-1}^n$. Offensichtlich gilt $W^n = \tilde{W}^n + e^J$ und daher $W^{n+1} = \tilde{W}^{n+1} + \mathcal{H}e^J$. Da die Methode monotonieerhaltend ist, folgt daraus

$$\mathcal{H}(e^J; j) = W_j^{n+1} - W_{j-1}^{n+1} \geq 0\,.$$

Zusammen mit (13.1) impliziert das $V_j^{n+1} - U_j^{n+1} \geq 0$, und damit die Monotonie der Methode. ∎

Die im obigen Beispiel angegebene Methode ist linear und zweiter Ordnung, kann daher nicht monotonieerhaltend (und damit nicht TVD) sein.

Satz 13.1 zwingt uns, zur Erreichung unseres Zieles nichtlineare Methoden zu verwenden. Eine Möglichkeit besteht darin, eine lösungsabhängige künstliche Viskosität zu verwenden. Eine solche Methode hat den numerischen Fluß

$$F(U; j) = F_H(U; j) - hQ(U; j)(U_{j+1} - U_j)\,,$$

wobei F_H der numerische Fluß einer Methode höherer Ordnung ist. Die künstliche Viskosität Q sollte dort am größten sein, wo die Lösung am stärksten variiert. Es ist allerdings sehr schwer, gerade das richtige Ausmaß an numerischer Dissipation zu finden. Deshalb werden wir uns mit zwei alternativen Zugängen beschäftigen.

13.2. Methoden mit Flußbegrenzung

Hier geht man aus von einer Methode höherer Ordnung mit numerischem Fluß $F_H(U;j)$ und einer TVD-Methode niederer Ordnung mit numerischem Fluß $F_L(U;j)$. Nun konstruiert man eine Methode, deren numerischer Fluß eine Konvexkombination von F_H und F_L ist:

$$F(U;j) = F_L(U;j) + \Phi(U;j)[F_H(U;j) - F_L(U;j)]$$

Die Flußbegrenzung Φ wird so gewählt, daß sie nahe bei 1 ist, wenn die Lösung in der Nähe von x_j glatt ist, und daß sie in der Nähe von Unstetigkeiten nahe bei 0 ist.

Als Beispiel betrachten wir die lineare skalare Gleichung $u_t + au_x = 0$ mit $a > 0$ und wählen für F_H den Lax-Wendroff-Fluß und für F_L den Upwind-Fluß. Dann gilt

$$F_H(U;j) = F_L(U;j) + \frac{1}{2}a(1-\nu)(U_{j+1} - U_j)\,,$$

mit dem Upwind-Fluß $F_L(U;j) = aU_j$ und der Courantzahl $\nu = ak/h$. Die Methode mit Flußbegrenzung ist nun gegeben durch

$$F(U;j) = aU_j + \frac{1}{2}a(1-\nu)(U_{j+1} - U_j)\phi_j\,, \tag{13.2}$$

mit $\phi_j = \Phi(U;j)$. Die Glattheit der Lösung wird gemessen am Wert von

$$\theta_j = \frac{U_j - U_{j-1}}{U_{j+1} - U_j}\,.$$

Wir wählen ϕ_j als Funktion von θ_j: $\phi_j = \phi(\theta_j)$. Gemäß unseren obigen Überlegungen ergibt sich für die Wahl der Funktion ϕ die Forderung $\phi(1) = 1$. Ein Problem entsteht in der Nähe von Extremalpunkten der Lösung, wo θ_j trotz Glattheit der Lösung weit weg von 1 sein kann. Man kann zeigen, daß TVD-Methoden in der Nähe von Extremalpunkten nicht zweiter Ordnung sein können.

Satz 13.2. *Die Methode mit dem numerischen Fluß (13.2) ist konsistent, wenn $\phi(\theta)$ eine beschränkte Funktion ist. Sie ist für glatte Lösungen mit $u_x \neq 0$ von zweiter Ordnung, wenn ϕ in einer Umgebung von 1 Lipschitzstetig ist und $\phi(1) = 1$ gilt.*

Die Methode kann in der Form

$$U_j^{n+1} = U_j^n - \left(\nu - \frac{1}{2}\nu(1-\nu)\phi_{j-1}\right)(U_j^n - U_{j-1}^n) - \frac{1}{2}\nu(1-\nu)\phi_j(U_{j+1}^n - U_j^n) \tag{13.3}$$

geschrieben werden. Für Methoden der Form

$$U_j^{n+1} = U_j^n - C_{j-1}(U_j^n - U_{j-1}^n) + D_j(U_{j+1}^n - U_j^n) \tag{13.4}$$

gilt das folgende Resultat:

Satz 13.3. *Eine Methode der Form (13.4) ist TVD, wenn die folgenden Bedingungen erfüllt sind:*

$$C_j \geq 0\,, \quad D_j \geq 0\,, \quad C_j + D_j \leq 1 \quad \forall j$$

Beweis. Es gilt

$$U_{j+1}^{n+1} - U_j^{n+1} = (1 - C_j - D_j)(U_{j+1}^n - U_j^n) + D_{j+1}(U_{j+2}^n - U_{j+1}^n) + C_{j-1}(U_j^n - U_{j-1}^n)\,.$$

Aus der Nichtnegativität der Koeffizienten auf der rechten Seite läßt sich leicht die Abschätzung $TV(U^{n+1}) \leq TV(U^n)$ herleiten. ∎

Für eine bestimmte Methode hat man große Freiheit bei der Wahl von C_j und D_j. Für (13.3) stellt sich die Wahl

$$\begin{aligned} C_{j-1} &= \nu + \frac{1}{2}(1-\nu)\nu \frac{\phi_j(U_{j+1}-U_j) - \phi_{j-1}(U_j - U_{j-1})}{U_j - U_{j-1}}\,, \\ D_j &= 0\,, \end{aligned}$$

als zweckmäßig heraus. Die Voraussetzungen von Satz 13.3 sind in diesem Fall für

$$0 \le C_j \le 1 \tag{13.5}$$

erfüllt. Schreiben wir C_{j-1} in der Form

$$C_{j-1} = \nu \left[1 + \frac{1}{2}(1-\nu)\left(\frac{\phi(\theta_j)}{\theta_j} - \phi(\theta_{j-1})\right)\right]\,,$$

so sieht man leicht, daß unter der Annahme der CFL-Bedingung $0 \le \nu \le 1$ die Ungleichung

$$\left|\frac{\phi(\theta_j)}{\theta_j} - \phi(\theta_{j-1})\right| \le 2 \tag{13.6}$$

hinreichend für (13.5) ist.

Ein negativer Wert von θ_j deutet ein Extremum in der Nähe von x_j an. Da die Methode in diesem Bereich ohnehin nur erster Ordnung sein kann, wenn sie TVD sein soll, wählen wir

$$\phi(\theta) = 0 \quad \text{für } \theta \le 0\,.$$

In diesem Fall sind die Ungleichungen

$$0 \le \frac{\phi(\theta)}{\theta} \le 2 \quad \text{und} \quad 0 \le \phi(\theta) \le 2$$

hinreichend für (13.6). Zusätzlich fordern wir $\phi(1) = 1$, damit wir eine Methode zweiter Ordnung erhalten. Numerische Experimente von Sweby (1984) haben außerdem ergeben, daß es günstig ist, eine Methode zwischen dem Lax-Wendroff-Verfahren ($\phi(\theta) = 1$) und dem Beam-Warming-Verfahren ($\phi(\theta) = \theta$) zu wählen. Zwei Versionen, die all diesen Forderungen entsprechen, sind der "Superbee limiter" von Roe (1985)

$$\phi(\theta) = \max(0, \min(1, 2\theta), \min(\theta, 2))$$

und die Wahl von van Leer (1974)

$$\phi(\theta) = \frac{|\theta| + \theta}{1 + |\theta|}\,.$$

Analog kann man für $a < 0$ vorgehen. Für die Behandlung von Systemen ist es von Vorteil, die Methoden für beide Fälle mit einer Formel angeben zu können. Dafür schreibt man den Lax-Wendroff-Fluß in der Form

$$F_H(U;j) = \frac{1}{2}a(U_j + U_{j+1}) - \frac{1}{2}\nu a(U_{j+1} - U_j)$$

und den Upwind-Fluß in der Form

$$F_L(U;j) = \frac{1}{2}a(U_j + U_{j+1}) - \frac{1}{2}|a|(U_{j+1} - U_j)\,.$$

Der numerische Fluß für die Methode mit Flußbegrenzung ist dann

$$F(U;j) = F_L(U;j) + \frac{1}{2}\phi_j(\operatorname{sign}(\nu) - \nu)a(U_{j+1} - U_j)\,.$$

Für die Berechnung von ϕ_j wird θ_j aus dem Differenzenquotienten in Upwind-Richtung bestimmt:

$$\theta_j = \frac{U_{j'+1} - U_{j'}}{U_{j+1} - U_j} \qquad \text{mit } j' = j - \operatorname{sign}(\nu)$$

Für lineare Systeme wendet man die obige Vorgangsweise auf die entkoppelten Gleichungen in charakteristischen Variablen an. Der Lax-Wendroff-Fluß und der Upwind-Fluß sind gegeben durch

$$\mathbf{F}_H(\mathbf{U}; j) = \frac{1}{2}\mathbf{A}(\mathbf{U}_j + \mathbf{U}_{j+1}) - \frac{k}{2h}\mathbf{A}^2(\mathbf{U}_{j+1} - \mathbf{U}_j)$$

und

$$\mathbf{F}_L(\mathbf{U}; j) = \frac{1}{2}\mathbf{A}(\mathbf{U}_j + \mathbf{U}_{j+1}) - \frac{1}{2}|\mathbf{A}|(\mathbf{U}_{j+1} - \mathbf{U}_j)\,.$$

Für die Methode mit Flußbegrenzung gilt

$$\mathbf{F}(\mathbf{U}; j) = \mathbf{F}_L(\mathbf{U}; j) + \frac{1}{2}\sum_{p=1}^{m} \phi(\theta_{jp})(\operatorname{sign}(\nu_p) - \nu_p)\lambda_p \alpha_{jp}\mathbf{r}_p$$

mit $\boldsymbol{\alpha}_j = \mathbf{R}^{-1}(\mathbf{U}_{j+1} - \mathbf{U}_j)$, $\nu_p = \lambda_p k/h$, $\theta_{jp} = \alpha_{j'_p p}/\alpha_{jp}$ und $j'_p = j - \operatorname{sign}(\nu_p)$.

Eine geeignete Verallgemeinerung auf nichtlineare Probleme führt auf ähnliche Methoden wie der unten besprochene Zugang über Anstiegsbegrenzung. Wir verzichten daher an dieser Stelle auf die Behandlung des nichtlinearen Falles.

13.3. Methoden mit Anstiegsbegrenzung

Das Lax-Wendroff-Verfahren für die skalare Gleichung $u_t + au_x = 0$ mit $a > 0$ kann folgendermaßen interpretiert werden: Es werden 3 Schritte durchgeführt.

1) Aus den Daten zum Zeitpunkt t_n wird durch

$$\tilde{u}^n(x, t_n) = U_j^n + \sigma_j^n(x - x_j) \quad \text{in } [x_{j-1/2}, x_{j+1/2})$$

eine stückweise lineare Funktion mit den Anstiegen

$$\sigma_j^n = \frac{U_{j+1}^n - U_j^n}{h} \tag{13.7}$$

konstruiert.

2) Die Gleichung wird mit den Anfangsdaten $\tilde{u}^n$ exakt gelöst.

3) U^{n+1} wird durch Mittelung der Lösung zum Zeitpunkt $t = t_{n+1}$ berechnet.

Dazu sind zwei Dinge anzumerken. Zunächst ist die höhere Konvergenzordnung des Verfahrens auf die stückweise lineare Rekonstruktion statt der sonst üblichen stückweise konstanten zurückzuführen. Andererseits ist der obige Schritt 1 der einzige, der zu einer Erhöhung der Totalvariation führen kann. Die bekannten Oszillationen in der Lösung entstehen durch die manchmal ungünstige Wahl (13.7) für die Anstiege. Gefragt ist daher eine Modifikation von (13.7), die der Forderung

$$TV(\tilde{u}^n(., t_n)) \leq TV(U^n) \tag{13.8}$$

genügt. Für allgemeine σ_j^n hat die entstehende Methode den numerischen Fluß

$$F(U; j) = aU_j + \frac{1}{2}a(1 - \nu)h\sigma_j$$

und ist leicht als Methode mit Flußbegrenzung mit

$$\sigma_j = \frac{U_{j+1} - U_j}{h}\phi_j$$

zu erkennen. Im gegenwärtigen Zusammenhang kann die Flußbegrenzung ϕ_j als Anstiegsbegrenzung interpretiert werden. Für beliebiges Vorzeichen von a erhält man

$$F(U;j) = aU_{j_1} + \frac{1}{2}a(\operatorname{sign}(\nu) - \nu)h\sigma_{j_1}$$

mit

$$j_1 = \begin{cases} j & \text{für } a > 0, \\ j+1 & \text{für } a < 0. \end{cases}$$

Eine Wahl für σ_j, die die Bedingung (13.8) erfüllt, ist der sogenannte **minmod-Anstieg**

$$\sigma_j = \frac{1}{h}\operatorname{minmod}(U_{j+1} - U_j, U_j - U_{j-1})$$

mit der minmod-Funktion

$$\begin{aligned} \operatorname{minmod}(a,b) &= \begin{cases} a & \text{für } |a| < |b| \text{ und } ab > 0, \\ b & \text{für } |b| < |a| \text{ und } ab > 0, \\ 0 & \text{für } ab \le 0 \end{cases} \\ &= \frac{1}{2}(\operatorname{sign}(a) + \operatorname{sign}(b)) \min(|a|, |b|)\,. \end{aligned}$$

Die entsprechende Flußbegrenzung $\phi(\theta) = \max(0, \min(1, \theta))$ erfüllt die im vorigen Abschnitt formulierten Bedingungen.

Für lineare Systeme transformieren wir auf charakteristische Variable $\mathbf{V}_j = \mathbf{R}^{-1}\mathbf{U}_j$ und verwenden die obige Vorgangsweise für die entkoppelten Gleichungen. Nach Rücktransformation ergibt sich der numerische Fluß

$$\mathbf{F}(\mathbf{U};j) = \sum_{p=1}^{m} V_{j_p p}\lambda_p \mathbf{r}_p + \frac{1}{2}\sum_{p=1}^{m} \lambda_p(\operatorname{sign}(\nu_p) - \nu_p)h\boldsymbol{\sigma}_{j_p p} \tag{13.9}$$

mit

$$\boldsymbol{\sigma}_{jp} = \frac{1}{h}\operatorname{minmod}(V_{j+1,p} - V_{jp}, V_{jp} - V_{j-1,p})\mathbf{r}_p \tag{13.10}$$

und

$$j_p = \begin{cases} j & \text{für } \lambda_p > 0, \\ j+1 & \text{für } \lambda_p < 0. \end{cases}$$

Man beachte, daß der erste Summand in (13.9) eine Schreibweise für den Upwind-Fluß ist.

Für nichtlineare Systeme verallgemeinern wir (13.9), indem wir die Matrix $\mathbf{A}$ durch eine Roe-Matrix $\mathbf{A}_j = \mathbf{A}(\mathbf{U}_j, \mathbf{U}_{j+1})$ ersetzen. Der entsprechende Fluß niederer Ordnung ist dann der des Godunov-Verfahrens mit näherungsweiser Lösung des Riemann-Problems

$$\mathbf{F}_L(\mathbf{U};j) = \mathbf{f}(\mathbf{U}_j) + \sum_{p=1}^{m} \lambda_{pj}^{-}\alpha_{jp}\mathbf{r}_{pj}$$

mit den Eigenwerten λ_{pj} und den Eigenvektoren $\mathbf{r}_{pj}$ von $\mathbf{A}_j$ und $\mathbf{U}_{j+1} - \mathbf{U}_j = \mathbf{R}_j\boldsymbol{\alpha}_j$. Für den numerischen Fluß des Verfahrens mit Anstiegsbegrenzung ergibt sich die Schreibweise

$$\mathbf{F}(\mathbf{U};j) = \mathbf{F}_L(\mathbf{U};j) + \frac{1}{2}\sum_{p=1}^{m} \lambda_{pj}(\operatorname{sign}(\nu_{pj}) - \nu_{pj})h\boldsymbol{\sigma}_{j_p p}\,. \tag{13.11}$$

Da die Eigenvektoren nun von j abhängen, muß die Formel (13.10) für den Anstiegsvektor im p-ten Feld modifiziert werden. Eine natürliche Verallgemeinerung ist

$$\boldsymbol{\sigma}_{jp} = \frac{1}{h}\operatorname{minmod}(\alpha_{jp}\mathbf{r}_{pj}, \alpha_{j-1,p}\mathbf{r}_{p,j-1})\,, \tag{13.12}$$

wobei die minmod-Funktion komponentenweise angewendet wird.

Damit haben wir eine hochgenaue Methode für nichtlineare Systeme vorgestellt. In der Praxis müßte sie noch durch einen "sonic entropy fix" modifiziert werden. Auf die Details wollen wir allerdings hier nicht eingehen.