

Die Methode der Finiten Differenzen

1. Finite Differenzen für skalare Erhaltungsgleichungen

In diesem Abschnitt entwickeln wir Finite-Differenz Verfahren oder Differenzenverfahren für skalare Erhaltungsgleichungen formuliert als Cauchyproblem

$$(1.1.1) \quad \begin{cases} u_t + f(u)_x &= 0 & \text{in } \mathbb{R} \times (0, \infty) \\ u &= u_0 & \text{auf } \mathbb{R} \times \{t = 0\} \end{cases}$$

wobei $f : \mathbb{R} \rightarrow \mathbb{R}$ eine vorgegebene Flußfunktion ist. Weiter sei $a(u) = f'(u)$ die charakteristische Geschwindigkeit.

Wir wollen uns zunächst auf explizite Einschrittverfahren beschränken. Dazu verwendet man eine Raum- und Zeitdiskretisierung, d.h. man überzieht den Raum $\mathbb{R} \times (0, \infty)$ mit einem regulären Gitter mit Gitterpunkten $(x_i, t_j) = (ih, jk)$ mit $i \in \mathbb{Z}, j \in \mathbb{N}_0$ und Schrittweiten $h, k > 0$.

Das Verhältnis der Schrittweiten in Raum und Zeit spielt eine wichtige Rolle bei der Untersuchung von Differenzenverfahren. Wir definieren daher das Gitterverhältnis λ als

$$\lambda = \frac{k}{h}$$

An jedem Gitterpunkt (x_i, t_j) wollen wir approximative Werte u_i^j für die exakte Lösung $u(x_i, t_j)$ an dieser Stelle berechnen. Dazu verwenden wir die Rekursion

$$(1.1.2) \quad u_i^{j+1} = H(u_{i-m}^j, \dots, u_{i+m}^j)$$

mit $m \in \mathbb{N}$ fest und der Gittertransferfunktion

$$H : \mathbb{R}^{2m+1} \rightarrow \mathbb{R}$$

die als stetig vorausgesetzt wird. Die Rekursion (1.1.2) bezeichnet man als ein Finite-Differenz Verfahren.

Das Verfahren nennt man

- **explizit**, da die neuen Werte zur Zeit t_{j+1} allein durch Vorgabe der Werte zur Zeit t_j berechnet werden können.
- Es heißt **Einschrittverfahren**, da nur die Werte u_i^j vom letzten Zeitschritt verwendet werden.

Das Verfahren (1.1.2) ist offensichtlich vollständig durch die Gittertransferfunktion H beschrieben. Diese Funktion erhält man durch geeignete Approximation der in der Ausgangsgleichung $u_t + f(u)_x = 0$ auftretenden Ableitungen durch Differenzenquotienten, also

diskrete Approximationen der Ableitungen auf dem vorgegebenen Gitter. Wird das Gitter immer weiter verfeinert, $h, k \rightarrow 0$, so möchten wir erreichen, dass die approximativen Werte u_i^j gegen die Werte $u(x_i, t_j)$ der exakten Lösung konvergieren. Im weiteren Verlauf werden wir sehen, welche speziellen Eigenschaften die Funktion H zu erfüllen hat, um diese Konvergenz zu gewährleisten.

Wir werden das Verhalten von (1.1.2) zunächst anhand von zwei klassischen Verfahren studieren: das naive Verfahren erhält man, in dem man einen Vorwärtsdifferenzenquotient für die Zeitableitung u_t verwendet, d.h.

$$u_t(x_i, t_j) \approx \frac{u_i^{j+1} - u_i^j}{k}$$

und dies mit einer zentralen Differenz für die Ortsableitung $f(u)_x$ kombiniert, also

$$f(u(x_i, t_j))_x \approx \frac{f(u_{i+1}^j) - f(u_{i-1}^j)}{2h}$$

Damit erhält man zunächst

$$\frac{u_i^{j+1} - u_i^j}{k} + \frac{f(u_{i+1}^j) - f(u_{i-1}^j)}{2h} = 0$$

und daraus

$$\begin{aligned} u_i^{j+1} &= H(u_{i-1}^j, u_i^j, u_{i+1}^j) \\ (1.1.3) \quad H(u_{i-1}^j, u_i^j, u_{i+1}^j) &= u_i^j - \frac{k}{2h} [f(u_{i+1}^j) - f(u_{i-1}^j)] \end{aligned}$$

Insbesondere verwendet man in diesem Fall $m = 1$.

Ersetzt man in (1.1.3) den Term u_i^j durch den Mittelwert der beiden Nachbarpunkte u_{i-1}^j und u_{i+1}^j , erhält man das Verfahren von Lax und Friedrichs,

$$\begin{aligned} u_i^{j+1} &= H(u_{i-1}^j, u_{i+1}^j) \\ H(u_{i-1}^j, u_{i+1}^j) &= \frac{1}{2}(u_{i-1}^j + \frac{k}{h}f(u_{i-1}^j)) + \frac{1}{2}(u_{i+1}^j - \frac{k}{h}f(u_{i+1}^j)) \end{aligned}$$

Im weiteren Verlauf werden wir die folgenden Notationen verwenden:

- 1) für eine fest vorgegebene Zeitschicht j schreiben wir auch abkürzend $u^j = (u_i^j)_{i \in \mathbb{Z}}$,
- 2) damit schreibt sich ein Finite-Differenz Verfahren auch in der Form $u^{j+1} = H(u^j)$ und $u_i^{j+1} = (Hu^j)_i$,
- 3) wir nehmen im Allgemeinen an, dass das Gitterverhältnis $\lambda = k/h$ fest ist und eine Gitterverfeinerung allein über die Schrittweite k gesteuert wird,
- 4) mit Hilfe der approximativen Werte u_i^j an den Gitterpunkten definieren wir eine stückweise konstante Funktion $u_k^j(x)$ durch

$$u_k^j(x) = u_i^j, \quad \text{für } x \in [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$$

wobei $x_{i+\frac{1}{2}} = (i+\frac{1}{2})h$ und schreiben auch $u_k^{j+1}(x) = H(u_k^j)(x)$. Weiterhin definieren wir eine stückweise konstante Funktion $u_k(x, t)$, die auf dem ganzen Raum $\mathbb{R} \times (0, \infty)$

definiert ist,

$$u_k(x, t) = u_i^j$$

für $(t, x) \in [t_j, t_{j+1}) \times [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$

- 5) um dann eine konsistente Definition für die Anfangsdaten u^0 zu erhalten, setzen wir die Anfangsbedingung als die Gittermittelwerte, d.h.

$$u_i^0 = \frac{1}{h} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u^0(x) dx$$

- 6) der Differenzenstern von H gibt eine geometrische Darstellung der Abhängigkeit von u^{j+1} von den Werten u^j der vorangegangenen Zeitschicht.

1.1. Konservative Finite-Differenz Verfahren. Ein fundamentales Prinzip bei der Herleitung von numerischen Methoden für partielle Differentialgleichungen ist es, Eigenschaften des kontinuierlichen Problems auf das diskrete Problem zu übertragen. Bei skalaren Erhaltungsgleichungen erhält man mit Hilfe der Konservativität der Gleichung Verfahren in Erhaltungssform oder auch konservative Verfahren. Wir wissen bereits, dass die eindeutig bestimmte Entropielösung der skalaren Erhaltungsgleichung

$$(1.1.4) \quad u_t + f(u)_x = 0$$

folgende Konservativitätseigenschaft besitzt: ist $u_0 \in L^1(\mathbb{R})$, so gilt

$$\int_{\mathbb{R}} u(x, t) dx = \int_{\mathbb{R}} u_0(x) dx$$

Wir nehmen nun (1.1.4) und integrieren die Gleichung über den Bereich $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}) \times [t_j, t_{j+1})$, den wir bereits zur Definition der stückweise konstanten Funktion $u_k(x, t)$ verwendet haben,

$$(1.1.5) \quad \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} (u(x, t_{j+1}) - u(x, t_j)) dx = - \int_{t_j}^{t_{j+1}} (f(u(x_{i+\frac{1}{2}}, t)) - f(u(x_{i-\frac{1}{2}}, t))) dt$$

Mit Hilfe der Diskretisierungsparameter h, k und λ läßt sich (1.1.5) auch in der Form

$$(1.1.6) \quad \frac{1}{h} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u(x, t_{j+1}) dx = \frac{1}{h} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u(x, t_j) dx - \frac{k}{h} \left(\frac{1}{k} \int_{t_j}^{t_{j+1}} f(u(x_{i+\frac{1}{2}}, t)) dt - \frac{1}{k} \int_{t_j}^{t_{j+1}} f(u(x_{i-\frac{1}{2}}, t)) dt \right)$$

schreiben. Auf der linken Seite von (1.1.6) steht der Zellmittelwert von u auf der Zeitschicht $j+1$. Dieser Zellmittelwert berechnet sich aus dem Zellmittelwert der Schicht j , korrigiert

um die zeitlichen Mittelwerte der Flüsse durch die Randpunkte des Intervalls $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ gewichtet mit dem Gitterverhältnis λ . Diese Darstellung legt folgende Definition für ein konservatives Verfahren nahe.

DEFINITION 1.1. *Ein Differenzenverfahren nennt man konservativ, falls sich die Funktion H in der Erhaltungsform*

$$H(u_{-m}, \dots, u_0, \dots, u_m) = u_0 - \lambda (F(u_{-m+1}, \dots, u_m) - F(u_{-m}, \dots, u_{m-1}))$$

schreiben läßt, wobei $F : \mathbb{R}^{2m} \rightarrow \mathbb{R}$ eine stetige Funktion ist. Das Verfahren lautet dann

$$u_i^{j+1} = H(u_{i-m}^j, \dots, u_{i+m}^j) = u_i^j - \lambda \left(f_{i+\frac{1}{2}} - f_{i-\frac{1}{2}} \right)$$

wobei wir die Notation

$$f_{i+\frac{1}{2}} = F(u_{i-m+1}^j, \dots, u_{i+m}^j)$$

verwendet haben. Die Funktion F nennt man auch numerische Flussfunktion.

Konservative Verfahren lassen sich auf folgende Weise charakterisieren.

LEMMA 1.2. *Die Funktion H kann genau dann in Erhaltungsform dargestellt werden, falls für alle $(u_i)_{i \in \mathbb{Z}} \in l_1(\mathbb{Z})$ gilt*

$$\sum_{i \in \mathbb{Z}} (Hu)_i = \sum_{i \in \mathbb{Z}} u_i$$

Der Beweis des Lemmas ist eine Übungsaufgabe.

BEMERKUNG 1.3. *Ist ein Verfahren konservativ, so folgt direkt*

$$\begin{aligned} \int_{\mathbb{R}} u_k^j(x) dx &= h \sum_i u_i^j = h \sum_i (Hu^j)_i \\ &= \int_{\mathbb{R}} (Hu^j)_k(x) dx = \int_{\mathbb{R}} u_k^{j+1}(x) dx \end{aligned}$$

Schreibt man ein konservatives Verfahren in der Form

$$\underbrace{\frac{u_i^{j+1} - u_i^j}{k}}_{\sim \partial_t u(t_j, x_i)} + \underbrace{\frac{f_{i+\frac{1}{2}} - f_{i-\frac{1}{2}}}{h}}_{\sim \partial_x f(u)(t_j, x_i)} = 0$$

so liefert der zweite Term gerade ein Approximation für den Term $\partial_x f(u)$ und kann daher nicht beliebig gewählt werden.

DEFINITION 1.4. *Eine numerische Flussfunktion heißt konsistent, falls gilt*

$$F(\underbrace{u, \dots, u}_{2m}) = f(u)$$

BEISPIEL 1.5. Für das naive Verfahren erhält man

$$\begin{aligned} u_i^{j+1} &= u_i^j - \frac{\lambda}{2}(f(u_{i+1}^j) - f(u_{i-1}^j)) \\ &= u_i^j - \lambda \left[\frac{1}{2} \left(f(u_i^j) + f(u_{i+1}^j) \right) - \frac{1}{2} \left(f(u_{i-1}^j) + f(u_i^j) \right) \right] \end{aligned}$$

Damit ist das Verfahren konservativ und die numerische Flussfunktion von der Form

$$F(u, v) = \frac{1}{2} (f(u) + f(v))$$

und konsistent, da $F(u, u) = f(u)$.

Beim Verfahren von Lax und Friedrichs schreibt man

$$\begin{aligned} u_i^{j+1} &= \frac{1}{2} \left(u_{i-1}^j + u_{i+1}^j \right) - \frac{\lambda}{2} (f(u_{i+1}^j) - f(u_{i-1}^j)) \\ &= u_i^j - \lambda \left[\frac{1}{\lambda} u_i^j - \frac{1}{2\lambda} (u_{i-1}^j + u_{i+1}^j) + \frac{1}{2} (f(u_{i+1}^j) - f(u_{i-1}^j)) \right] \\ &= u_i^j - \lambda \left[\frac{1}{2\lambda} (u_i^j - u_{i+1}^j) + \frac{1}{2} (f(u_i^j) + f(u_{i+1}^j)) \right. \\ &\quad \left. - \frac{1}{2\lambda} (u_{i-1}^j - u_i^j) - \frac{1}{2} (f(u_{i-1}^j) + f(u_i^j)) \right] \end{aligned}$$

Die numerische Flussfunktion ist also gegeben durch

$$F(u, v) = \frac{1}{2\lambda} (u - v) + \frac{1}{2} (f(u) + f(v))$$

und das Verfahren ist konsistent und konservativ.

Ein erstes Resultat zur Konvergenz von Finite-Differenz Verfahren ist das Lax-Wendroff Theorem:

SATZ 1.6. Sei durch die Gittertransferfunktion H ein konservatives und konsistentes Verfahren definiert und betrachte den Grenzwert $k \rightarrow 0$ mit $\lambda = \frac{k}{h}$ fest. Gilt

$$\|u_k\|_{L^\infty(\mathbb{R} \times \mathbb{R}_+)} \leq C$$

und $u_k \rightarrow u$ in $L_{loc}^1(\mathbb{R} \times \mathbb{R}_+)$ and $f.$ ü. in $\mathbb{R} \times \mathbb{R}_+$, dann ist u eine schwache Lösung des zugehörigen Cauchyproblems.

BEWEIS. Sei $\varphi \in C_0^1(\mathbb{R} \times \mathbb{R}_+)$ eine Testfunktion und definiere $\varphi_i^j = \varphi(t_j, x_i)$. Wir multiplizieren die Gleichung

$$\frac{u_i^{j+1} - u_i^j}{k} + \frac{f_{i+\frac{1}{2}} - f_{i-\frac{1}{2}}}{h} = 0$$

mit $h k \varphi_i^j$ und summieren über $i \in \mathbb{Z}$, $j \in \mathbb{N}_0$,

$$(1.1.7) \quad h \sum_{i,j} (u_i^{j+1} - u_i^j) \varphi_i^j + k \sum_{i,j} (f_{i+\frac{1}{2}} - f_{i-\frac{1}{2}}) \varphi_i^j = 0$$

Eine Indexverschiebung in (1.1.7) liefert

$$(1.1.8) \quad h \sum_{i,j} w_i^{j+1} (\varphi_i^{j+1} - \varphi_i^j) + k \sum_{i,j} f_{i+\frac{1}{2}} (\varphi_{i+1}^j - \varphi_i^j) + h \sum_i u_i^0 \varphi_i^0 = 0$$

Wir definieren nun stückweise konstante Funktionen f_k und φ_k durch

$$\begin{aligned} f_k(x, t) &= f_{i+\frac{1}{2}}^j, & (x, t) &\in [x_i, x_{i+1}) \times [t_j, t_{j+1}) \\ \varphi_k(x, t) &= \varphi_i^j, & (x, t) &\in [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [t_j, t_{j+1}] \end{aligned}$$

Dann läßt sich Gleichung (1.1.8) in Integralform schreiben,

$$\begin{aligned} & \int_k^\infty \int_{\mathbb{R}} u_k(x, t) \frac{\varphi_k(x, t) - \varphi_k(x, t-k)}{k} dx dt \\ & + \int_0^\infty \int_{\mathbb{R}} f_k(x, t) \frac{\varphi_k(x+h, t) - \varphi_k(x, t)}{h} dx dt \\ & + \int_{\mathbb{R}} u_k(x, 0) \varphi_k(x, 0) dx = 0 \end{aligned}$$

Dies ist gerade das diskrete Analogon zur Definition einer schwachen Lösung. Wir haben also zu zeigen, dass im Grenzfall $k \rightarrow 0$ die auftretenden Integrale gegen die entsprechenden Terme in der Definition einer schwachen Lösung konvergieren.

Wir untersuchen zunächst den Term

$$\int_{\mathbb{R}} u_k(x, 0) \varphi_k(x, 0) dx$$

Da

$$\begin{aligned} u_k(x, 0)|_{[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})} &= u_i^0 = \frac{1}{h} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u_0(x) dx \\ \varphi_k(x, 0)|_{[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})} &= \varphi_i^0 = \varphi(x_i, 0), \end{aligned}$$

erhalten wir

$$\int_{\mathbb{R}} u_k(x, 0) \varphi_k(x, 0) dx = h \sum_i \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u_0(x) \varphi_i^0 dx = \int_{\mathbb{R}} u_0(x) \varphi_k(x, 0) dx$$

Nun gilt $\varphi_k \xrightarrow{k \rightarrow 0} \varphi$ gleichmäßig und daher

$$\int_{\mathbb{R}} u_0(x) \varphi_k(x, 0) dx \longrightarrow \int_{\mathbb{R}} u_0(x) \varphi(x, 0) dx$$

für $k \rightarrow 0$.

Als nächstes untersuchen wir den Term

$$(1.1.9) \quad \int_k^\infty \int_{\mathbb{R}} u_k(x, t) \frac{\varphi_k(x, t) - \varphi_k(x, t - k)}{k} dx dt$$

Aus

$$\mathcal{X}_{[k, \infty)} \frac{\varphi_k(x, t) - \varphi_k(x, t - k)}{k} \xrightarrow{k \rightarrow 0} \partial_t \varphi$$

in L^1 , folgt

$$\int_k^\infty \int_{\mathbb{R}} u_k(x, t) [\chi_{(k, \infty)} \frac{\varphi_k(x, t) - \varphi_k(x, t - k)}{k} - \partial_t \varphi] \rightarrow 0$$

da u_k nach Voraussetzung in L^∞ beschränkt ist.

Da zusätzlich $u_k \rightarrow u$ in L^1_{loc} und φ einen kompakten Träger besitzt, erhalten wir

$$\int_k^\infty \int_{\mathbb{R}} u_k(x, t) \partial_t \varphi(x, t) dx dt \rightarrow \int_0^\infty \int_{\mathbb{R}} u(x, t) \partial_t \varphi(x, t) dx dt$$

für $k \rightarrow 0$. Fassen wir die Ergebnisse zusammen, so haben wir gezeigt, dass (1.1.9) für

$k \rightarrow 0$ gegen $\int_0^\infty \int_{\mathbb{R}} u(x, t) \partial_t \varphi(x, t) dt dx$ konvergiert.

Wir kommen nun zum letzten Term

$$\int_0^\infty \int_{\mathbb{R}} f_k(t, x) \frac{\varphi_k(t, x + h) - \varphi_k(t, x)}{h} dx dt$$

Zunächst schreiben wir

$$f_k(x, t) = F(v_k^{-m+1}(x, t), \dots, v_k^m(x, t))$$

wobei

$$v_k^l(x, t) = u_k(x + (l - \frac{1}{2})h, t)$$

d.h. v_k^l bezeichnet eine Verschiebung von u_k . Sei $K \subset \mathbb{R} \times \mathbb{R}_+$ kompakt, dann gilt

$$\begin{aligned} \int_K |v_k^l(x, t) - u(x, t)| dx dt &\leq \int_K |u_k\left(x + (l - \frac{1}{2})h, t\right) - u\left(x + (l - \frac{1}{2})h, t\right)| dx dt \\ &\quad + \int_K |u\left(x + (l - \frac{1}{2})h, t\right) - u(x, t)| dx dt \end{aligned}$$

Da $u_k \rightarrow u$ in L^1_{loc} , $u(t, x + ch) \rightarrow u(t, x)$ punktweise f.ü. und $u(x + ch, t), u(x, t) \in L^1$, folgt $v_k^l \rightarrow u$ in $L^1(K)$.

Wir können dann eine Teilfolge v_k^l finden, die f.ü. in K gegen u konvergiert und zwar für $-m+1 \leq l \leq m$. Da F eine stetige Funktion und konsistent ist, folgt

$$f_k(x, t) = F(v_k^{-m+1}(x, t), \dots, v_k^m(x, t)) \rightarrow F(u(x, t), \dots, u(x, t)) = f(u(x, t))$$

f.ü. in K . Nach dem Satz von der majorisierten Konvergenz von Lebesgue (die beiden Funktionen φ und $\partial_x \varphi$ haben einen kompakten Träger), folgt dann, dass

$$\int_0^\infty \int_{\mathbb{R}} f_k(x, t) \frac{\varphi_k(x+h, t) - \varphi_k(x, t)}{h} dx dt \rightarrow \int_0^\infty \int_{\mathbb{R}} f(u(x, t)) \partial_x \varphi(x, t) dx dt$$

□

BEMERKUNG 1.7. Für nicht-konservative Verfahren existieren Beispiele, bei denen die numerischen Approximationen gegen eine kontinuierliche Funktion konvergieren, die keine schwache Lösung des zugrundeliegenden Problems ist (siehe Übungen).

Der Satz von Lax und Wendroff ist kein Konvergenzsatz, denn wir müssen die Konvergenz der diskreten Approximationen gegen eine kontinuierliche Funktion voraussetzen. Nach dem Resultat ist ebenfalls nicht gewährleistet, dass man gegen die eindeutig bestimmte Entropielösung des Problems konvergiert. Schliesslich erhalten wir keine Aussage über die Konvergenzgeschwindigkeit des Verfahrens, also die Genauigkeit, um etwa unterschiedliche numerische Verfahren miteinander vergleichen zu können.

1.2. Genauigkeit, Modifizierte Gleichung und CFL-Bedingung.

DEFINITION 1.8. Die natürliche Zahl $p > 0$ heisst Genauigkeitsordnung eines Verfahrens H , falls p die größte Zahl ist, sodass

$$(1.1.10) \quad u(x, t+k) - H(u(x-mh, t), \dots, u(x+mh, t)) = O(k^{p+1})$$

für $k \rightarrow 0$ und jede genügend glatte Lösung einer skalaren Erhaltungsgleichung $u + f(u)_x = 0$.

Dividiert man die linke Seite von (1.1.10) durch die Zeitschrittweite k , so erhält man den lokalen Abschneidefehler oder Konsistenzfehler $(Lu)(x, t)$ in der Form

$$(Lu)(x, t) := \frac{1}{k} (u(t+k, x) - H(u(x-mh, t), \dots, u(x+mh, t)))$$

Dementsprechend nennt man die natürliche Zahl p aus Definition 1.8 auch die **Konsistenzordnung**. Allgemein heisst ein Finite-Differenz Verfahren **konsistent**, falls $|(Lu)(x, t)| \rightarrow 0$ für $k \rightarrow 0$. Verwendet man ein konservatives Verfahren,

$$H = u(x, t) - \lambda(f_{i+\frac{1}{2}}(u) - f_{i-\frac{1}{2}}(u))$$

so ist Konsistenzfehler gegeben durch

$$(1.1.11) \quad (Lu)(x, t) = \frac{u(t+k, x) - u(t, x)}{k} + \frac{f_{i+\frac{1}{2}}(u) - f_{i-\frac{1}{2}}(u)}{h}$$

Weiterhin gilt

LEMMA 1.9. *Sei durch die Funktion H ein konsistentes und konservatives Finite-Differenz Verfahren definiert.*

- 1) *Ist $F \in \mathcal{C}^2(\mathbb{R}^{2m})$, dann ist das Verfahren für jede hinreichend glatte Lösung mindestens erster Ordnung,*
- 2) *Ist $F \in \mathcal{C}^3(\mathbb{R}^{2m})$, dann ergibt sich für jede hinreichend glatte Lösung eines Cauchy-problems die Beziehung*

$$(Lu)(x, t) = -k \partial_x (\beta(u, \lambda)(x, t) \partial_x u(x, t)) + O(k^2)$$

mit

$$\beta(u, \lambda) = \frac{1}{2} \left(\frac{1}{\lambda^2} \sum_{l=-m}^m l^2 \partial_l H(u, \dots, u) - a^2(u) \right)$$

und $a(u) = f'(u)$.

BEWEIS. Für die erste Aussage müssen wir zeigen, dass der Konsistenzfehler nach (1.1.11) von der Ordnung $O(k)$ ist. Dazu verwenden wir geeignete Taylorentwicklungen: Es gilt

$$u(t + k, x) = u(t, x) + k u_t(t, x) + O(k^2)$$

sodass man für den ersten Term des Konsistenzfehler die Beziehung

$$\frac{u(t + k, x) - u(t, x)}{k} = u_t(t, x) + O(k)$$

erhält. Für die numerische Flussfunktion ergibt sich

$$\begin{aligned} f_{i+\frac{1}{2}}(u) - f_{i-\frac{1}{2}}(u) &= F(u(t, x - (m-1)h), \dots, u(t, x + mh)) \\ &\quad - F(u(t, x - mh), \dots, u(t, x + (m-1)h)) \end{aligned}$$

und daher

$$\begin{aligned} f_{i+\frac{1}{2}}(u) - f_{i-\frac{1}{2}}(u) &= - \sum_{l=1}^{m-1} (\partial_l F(u(t, x), \dots, u(t, x))) (m-l) h u_x(t, x) \\ &\quad + \sum_{l=m+1}^{2m} (\partial_l F(u(t, x), \dots, u(t, x))) (l-m) h u_x(t, x) \\ &\quad + \sum_{l=1}^m (\partial_l F(u(t, x), \dots, u(t, x))) (m+1-l) h u_x(t, x) \\ &\quad - \sum_{l=m+2}^{2m} (\partial_l F(u(t, x), \dots, u(t, x))) (l-(m+1)) h u_x(t, x) \\ &= h u_x(t, x) \sum_{l=1}^{2m} (\partial_l F(u(t, x), \dots, u(t, x))) + O(h^2) \end{aligned}$$

Damit können wir den zweiten Term des Konsistenzfehlers in der Form

$$(1.1.12) \quad \frac{f_{i+\frac{1}{2}}(u) - f_{i-\frac{1}{2}}(u)}{h} = u_x(t, x) \sum_{l=1}^{2m} (\partial_l F(u(t, x), \dots, u(t, x))) + O(h)$$

darstellen. Gleichzeitig gilt

$$f(u(t, x + h)) - f(u(t, x - h)) = 2 (\partial_u f(u(t, x))) h u_x(t, x) + O(h^2)$$

und aufgrund der Konsistenz der numerischen Flussfunktion

$$\begin{aligned} f(u(t, x + h)) - f(u(t, x - h)) &= F(u(t, x + h), \dots, u(t, x + h)) \\ &\quad - F(u(t, x - h), \dots, u(t, x - h)) \\ &= 2h u_x(t, x) \sum_{l=1}^{2m} (\partial_l F(u(t, x), \dots, u(t, x))) \end{aligned}$$

Damit liefert (1.1.12)

$$\frac{f_{i+\frac{1}{2}}(u) - f_{i-\frac{1}{2}}(u)}{h} = (\partial_u f(u(t, x))) u_x(t, x) + O(h)$$

und für den lokalen Abschneidefehler gilt

$$(Lu)(t, x) = u_t(t, x) + f(u(t, x))_x + O(k) + O(h)$$

Da wir angenommen haben, dass $u(x, t)$ eine Lösung der Erhaltungsgleichung ist und $h = k/\lambda$, λ fest, erhalten wir $(Lu)(x, t) = O(k)$. Der zweite Teil des Lemmas ist eine Übungsaufgabe. $\square$

Allgemein verwendet man den Konsistenzfehler $(Lu)(x, t)$ zur Definition der **modifizierten Gleichung**

DEFINITION 1.10. Der Konsistenzfehler $(Lu)(x, t)$ eines Verfahrens H der Ordnung p sei gegeben durch

$$(Lu)(x, t) = k^p \gamma_p + O(k^{p+1})$$

Dann nennt man die partielle Differentialgleichung

$$v_t + f(v)_x = k^p \gamma_p$$

die **modifizierte Gleichung zum Verfahren H** .

Das Verfahren H löst also die gegebene Gleichung bis zur Ordnung $O(k^p)$, die modifizierte Gleichung allerdings zur Ordnung $O(k^{p+1})$. Für den Konvergenzfehler aus Lemma 1.9 ergibt sich demnach die modifizierte Gleichung als

$$v_t + f(u)_x = k(\beta(u, \lambda) u_x)_x$$

Der Korrekturterm ist also ein nichtlinearer Diffusionsterm mit Diffusionskoeffizienten $k\beta$, aus dem man eine Idee über das Verhalten eines Verfahrens erhält:

- Ist der Diffusionskoeffizient $k\beta > 0$ groß, so wird das Verfahren die Lösung sehr stark glätten, zum Beispiel werden Stoßwellen stark verschmiert. Dadurch gewährleistet man aber auch die Konvergenz gegen die eindeutige Entropielösung.

- Ist $k\beta > 0$ klein, so wird die Lösung nur wenig geglättet, Stoßwellen behalten ihre scharfe Struktur. Allerdings reicht die Glättung unter Umständen nicht aus, um gegen die Entropielösung zu konvergieren.
- Ist $k\beta < 0$, so ist das Verfahren instabil, da wir versuchen, die Lösung einer Rückwärts-Wärmeleitungsgleichung zu berechnen.

DEFINITION 1.11.

Den Ausdruck $k\beta(u, \lambda)(x, t)$ nennt man die **numerische Viskosität** des *Finite-Differenz Verfahrens* gegeben durch H .

BEISPIEL 1.12. Für das naive Verfahren

$$u_i^{j+1} = u_i^j - \frac{k}{2h}(f(u_{i+1}^j) - f(u_{i-1}^j))$$

und die Funktion H

$$H(u_{-1}, u_0, u_1) = u_0 - \frac{\lambda}{2}(f(u_1) - f(u_{-1}))$$

berechnet man

$$\begin{aligned} \beta(u, \lambda) &= \frac{1}{2} \left[\frac{1}{\lambda^2} \sum_{l=-m}^m l^2 \partial_l H(u, \dots, u) - a^2(u) \right] \\ &= \frac{1}{2} \left[\frac{1}{\lambda} \left(\frac{a(u)\lambda}{2} - \frac{a(u)\lambda}{2} \right) - a^2(u) \right] \\ &= -\frac{a^2(u)}{2} \end{aligned}$$

und die modifizierte Gleichung ergibt sich als

$$v_t + f(v)_x = -\frac{ka^2(v)}{2}v_{xx}$$

Das bedeutet, dass das Verfahren eine negative numerische Viskosität besitzt und daher nicht funktionieren wird.

Beim Lax-Friedrichs Verfahren

$$u_i^{j+1} = \frac{u_{i+1}^j + u_{i-1}^j}{2} - \frac{k}{2h}(f(u_{i+1}^j) - f(u_{i-1}^j))$$

erhalten wir

$$H(u_{-1}, u_0, u_1) = \frac{1}{2}(u_1 + u_{-1}) - \frac{\lambda}{2}(f(u_1) - f(u_{-1}))$$

und für β ergibt sich der Ausdruck

$$\begin{aligned} \beta(u, \lambda) &= \frac{1}{2} \left(\frac{1}{\lambda^2} \left(\frac{1}{2} + \frac{a(u)\lambda}{2} + \frac{1}{2} - \frac{a(u)\lambda}{2} \right) - a^2(u) \right) \\ &= \frac{1}{2} \left(\frac{1}{\lambda^2} - a^2(u) \right) \\ &= \frac{1}{2\lambda^2} (1 - (a(u)\lambda)^2) \end{aligned}$$

Damit lautet die modifizierte Gleichung

$$v_t + f(v)_x = \frac{k}{2\lambda^2}(1 - (a(v)\lambda)^2)v_{xx}$$

und der Diffusionskoeffizient ist positiv, falls die Bedingung

$$(1.1.13) \quad |a\lambda| < 1$$

erfüllt ist. Wir erhalten also eine Restriktion an das Gitterverhältnis λ in Abhängigkeit von der lokalen Ausbreitungsgeschwindigkeit $a(u) = f'(u)$.

Die Bedingung (1.1.13) läßt sich sehr anschaulich zum Beispiel im Fall der linearen Advektionsgleichung mit konstanter Geschwindigkeit a , mit der Information transportiert wird, interpretieren.

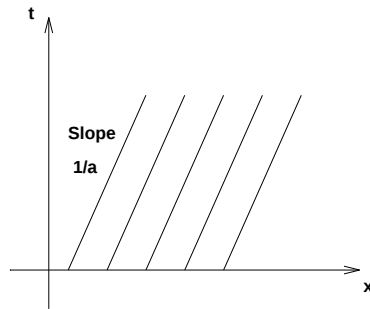


Bild 1.1 Charakteristiken der linearen Advektionsgleichung

Im Verfahren von Lax und Friedrichs werden Information von den Gitterpunkten $u_{i\pm 1}^j$ nach u_i^{j+1} mit der Geschwindigkeit $\pm \frac{h}{k} = \pm 1/\lambda$ transportiert.

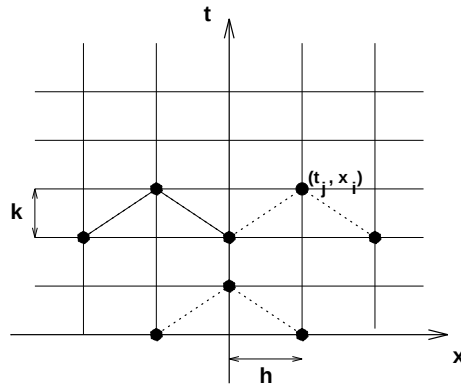


Bild 1.2 Informationstransport im Verfahren von Lax und Friedrichs

Die Bedingung $|a\lambda| < 1$ ergibt $|a| < 1/\lambda$, d.h. das Verfahren muss in der Lage sein, Informationen mindestens so schnell zu transportieren, wie es durch die Advektionsgeschwindigkeit a vorgegeben ist.

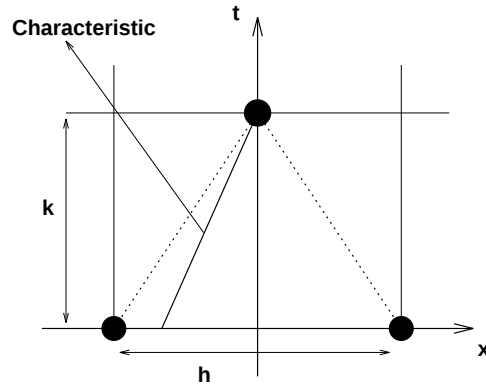


Bild 1.3 Analytischer und numerischer Abhängigkeitsbereich

Das bedeutet ebenfalls, dass der analytische Abhängigkeitsbereich im numerischen Abhängigkeitsbereich enthalten sein muss. Ist diese Bedingung verletzt, so gilt dies auch im Grenzfall $k \rightarrow 0$, da das Gitterverhältnis λ fest ist, und man kann nicht erwarten, dass das Verfahren konvergieren wird.

DEFINITION 1.13. *Die Bedingung*

$$(1.1.14) \quad |\lambda a(v)| < 1$$

nennt man **CFL-Bedingung**. Bei nichtlinearen Problemen muss (1.1.14) für alle $a(v) = f'(v)$ mit $v \in [-\|u\|_{L^\infty}, \|u\|_{L^\infty}]$ erfüllt sein.

Wie der Name schon sagt, wurde die Bedingung von Courant, Friedrichs und Levy beim Versuch aufgestellt, mit Hilfe von Finite-Differenz Verfahren die Existenz von Lösungen für Cauchyprobleme zu beweisen. Die CFL-Bedingung ist notwendig, aber nicht hinreichend für die Konvergenz des Verfahrens.

1.3. Lineare Gleichungen, Äquivalenzsatz, v. Neumann Stabilität. Das Lax-Wendroff Theorem gibt eine erste Aussage über die Konvergenz von Finite-Differenz Verfahren, allerdings muss dabei die Konvergenz der numerischen Approximationen vorausgesetzt werden. Die modifizierte Gleichung gibt einen Eindruck über das Verhalten eines Verfahrens. Für den Fall einer linearen Gleichung kann man den Äquivalenzsatz von Lax zeigen.

SATZ 1.14. Gegeben sei die lineare Advektionsgleichung $u_t + au_x = 0$ und ein lineares Finite-Differenz Verfahren definiert durch H . Dann ist H konvergent genau dann, wenn H konsistent und stabil ist. Die Konvergenzordnung ist zudem identisch mit der Konsistenzordnung.

DEFINITION 1.15. Das Differenzenverfahren definiert durch H ist linear, falls

$$H(u_{-m}, \dots, u_m) = \sum_{l=-m}^m c_l u_l$$

Zur Konvergenz definieren wir zunächst den globalen Fehler $e(x, t)$ zwischen der exakten Lösung und der numerischen Approximation: seien $\bar{u}_k(x, t)$ definiert als die stückweise konstanten Funktionen

$$\bar{u}_k(x, t) = u(x_i, t_j) \quad \text{für } (x, t) \in [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}) \times [t_j, t_{j+1})$$

wobei $u(x_i, t_j)$ die exakte Lösung am Punkt (x_i, t_j) ist. Dann ist der globale Fehler gegeben durch

$$e(x, t) = \bar{u}_k(x, t) - u_k(x, t)$$

Zusätzlich erweitern wir die Gittertransferfunktion H auf Funktionen $v(x)$ mittels

$$(Hv)(x) = \sum_{l=-m}^m c_l v(x + lh)$$

Damit erhalten wir die Beziehung

$$u_k(t_j, x) = (Hu_k(\cdot, t_{j-1}))(x)$$

und der globale Fehler zur Zeit $t = t_j$ ist gegeben durch

$$\begin{aligned} e(x, t_j) &= \bar{u}_k(x, t_j) - u_k(x, t_j) \\ &= \bar{u}_k(x, t_j) - (Hu_k(\cdot, t_{j-1}))(x) \\ &= \bar{u}_k(x, t_j) - (H\bar{u}(\cdot, t_{j-1}))(x) + H(\bar{u}_k(\cdot, t_{j-1}) - u_k(\cdot, t_{j-1}))(x) \\ &= kL_k(x, t_{j-1}) + (He(\cdot, t_{j-1}))(x), \end{aligned}$$

wobei wir die Linearität des Verfahrens ausgenutzt haben. Weiter sei L_k die stückweise konstante Rekonstruktion des Konsistenzfehlers (Lu) . Insbesondere gilt

$$e(x, t_1) = kL_k(x, t_0) + (He(\cdot, t_0))(x)$$

und iteriert man diese Gleichung bis zur Zeitschicht j , ergibt sich

$$e(x, t_j) = (H^j e(\cdot, t_0))(x) + k \sum_{l=0}^{j-1} (H^l L_k(\cdot, t_{j-1-l}))(x)$$

Halten wir nun $T > 0$ fest, setzen $t_0 = 0$ und führen m_k Zeitschritte durch, wobei $T \leq m_k k$, so erhalten wir die Abschätzung

$$(1.1.15) \quad \sup_{t \in [0, T]} \|e(\cdot, t)\| \leq \sup_{j \in \{0, \dots, m_k\}} \left\{ \|H^j e(\cdot, 0)\| + k \sum_{l=0}^{j-1} \|H^l L_k(\cdot, t_{j-1-l})\| \right\}$$

Wir bezeichnen daher ein Differenzenverfahren als **konvergent**, falls die linke Seite von (1.1.15) für $k \rightarrow 0$ gegen Null geht.

Da die Anfangswerte bekannt sind, können wir ohne Einschränkung annehmen, dass

$$e(0, \cdot) = 0$$

gilt. Es genügt daher den zweiten Term auf der rechten Seite von (1.1.15) zu untersuchen. Dies ist nicht-trivial, da für $k \rightarrow 0$ folgt $m_k \rightarrow \infty$ und damit in (1.1.15) eine unendliche Summe entsteht. Insbesondere benötigt man die Beschränktheit von $\|H^j\|$ für alle $j \in \{0, \dots, m_k\}$. Mit Hilfe der Konsistenz des Verfahrens kann man dann die Konvergenz für

$k \rightarrow 0$ zeigen.

Die Beschränktheit von $\|H^j\|$ liefert eine geeignete Definition eines **stabilen Verfahrens**.

DEFINITION 1.16. *Das Differenzenverfahren definiert durch H ist $\|\cdot\|$ -stabil, falls*

$$\|H^j\| = \sup_{\|u\| \neq 0} \frac{\|H^j u\|}{\|u\|} \leq C$$

für alle $k \in (0, k_0]$ und $j \in \{0, \dots, m_k\}$.

Ist das Differenzenverfahren bezüglich der Norm $\|\cdot\|$ **gleichmässig konsistent** von der Ordnung p , d.h.

$$\|L_k(t, \cdot)\| \leq Ck^p$$

für alle $t \in [0, T]$ and $k \in (0, k_0]$, so erhalten wir die Abschätzung

$$\begin{aligned} \sup_{t \in [0, T]} \|e(t, \cdot)\| &\leq \sup_{j \in \{0, \dots, m_k\}} \left\{ k \sum_{l=0}^{j-1} \|H^l\| \|L_k(t_{j-1-l}, \cdot)\| \right\} \\ &\leq Ck^p(T + k_0) \end{aligned}$$

und der globale Fehler geht für $k \rightarrow 0$ gegen Null.

Damit haben wir die Aussage des Äquivalenzsatzes erklärt und gleichzeitig eine Richtung der Aussage bewiesen: ist das Verfahren H konsistent und stabil, so ist es gleichzeitig konvergent. Weiterhin ist nach der Abschätzung (1.1.16) die Konvergenzordnung gleich der Konsistenzordnung. Die umgekehrte Richtung des Satzes ist weitaus schwieriger zu zeigen und wir verzichten daher hier auf den Beweis.

Satz 1.14 liefert die Konvergenz des Verfahrens in der ∞ -Norm bezüglich der Zeitvariablen sowie einer beliebigen Norm im Ortsraum, die die Stabilität des Verfahrens definiert. Für ein gegebenes Verfahren ist es insbesondere schwierig die $\|\cdot\|$ -Stabilität von H zu untersuchen und die Wahl der zugrundeliegenden Norm in von entscheidender Bedeutung:

- Da wir in diesem Abschnitt Erhaltungsgleichungen untersuchen, wäre eine natürliche Norm im Ortsraum die L^1 -norm, da eine Erhaltungsgleichung als partielle Differentialgleichung aus einer Integraldarstellung abgeleitet wird.
- Eine ideale Norm ist im Ortsraum ebenfalls die L^∞ -Norm. Dies ist aber in Zusammenhang mit Erhaltungsgleichungen unrealistisch, da auch nichtstetige Lösungen auftreten können. Es kann etwa der Fall auftreten, dass das Verfahren aufgrund einer nichtstetigen Lösung in der L^∞ -Norm nicht, aber in der L^1 -Norm konvergiert.
- Für lineare Gleichungen verwendet man häufig die L^2 -norm,

$$\|v\|_2 = \int_{\mathbb{R}} |v(x)|^2 dx$$

da man unter Verwendung dieser Norm auf die Fourieranalyse zurückgreifen kann. Aus der Fourieranalyse kennt man etwa die Beziehung

$$\|\widehat{v}\|_2 = \|v\|_2$$

wobei $\widehat{v}$ die Fouriertransformierte von v bezeichnet.

Im Folgenden wollen wir für den Fall der linearen Advektionsgleichung die Anwendung der L^2 -Norm näher untersuchen. Die Fouriertransformierte $\widehat{v}(\xi)$ von $v(x)$ ist definiert durch

$$\widehat{v}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} v(x) e^{-ix\xi} dx$$

Wie oben erwähnt, gilt $\|\widehat{v}\|_2 = \|v\|_2$. Weiterhin erhält man für die verschobene Funktion $v_y(\cdot) = v(\cdot + y)$ mit y fest

$$\widehat{v}_y(\xi) = e^{iy\xi} \widehat{v}(\xi)$$

Um die $\|\cdot\|_2$ -Stabilität eines Differenzenverfahrens zu untersuchen, müssen wir die L^2 -Norm von $H^j v$ berechnen, die gleich der L^2 -Norm der Fouriertransformierten $\widehat{H^j v}$ ist: für $j = 1$ erhalten wir

$$\widehat{Hv}(\xi) = \sum_{l=-m}^m c_l \widehat{v_{lh}}(\xi) = \left(\sum_{l=-m}^m c_l e^{ilh\xi} \right) \widehat{v}(\xi) = \rho(\xi) \widehat{v}(\xi)$$

Damit gilt für $j > 1$ die Beziehung

$$\widehat{H^j v}(\xi) = H(\widehat{H^{j-1} v})(\xi) = \rho(\xi) \widehat{H^{j-1} v}(\xi) = \rho^j(\xi) \widehat{v}(\xi)$$

und die L^2 -Norm von H^j ist gegeben durch

$$\|H^j v\|_2 = \|\rho^j \widehat{v}\|_2$$

DEFINITION 1.17. Der Ausdruck $\rho(\xi)$ definiert durch

$$\rho(\xi) = \sum_{l=-m}^m c_l e^{ilh\xi}$$

heißt Verstärkungsfaktor oder Symbol von H .

SATZ 1.18. Ein lineares Verfahren definiert durch H ist L^2 -stabil, falls

$$\sup_{\xi \in \mathbb{R}} |\rho(\xi)| \leq 1$$

Der Beweis ist eine Übungsaufgabe.

BEISPIEL 1.19. Für das naive Verfahren

$$u_i^{j+1} = u_i^j - \frac{\lambda}{2} a(u_{i+1}^j - u_{i-1}^j)$$

berechnet man

$$\begin{aligned} \rho(\xi) &= \sum_{l=-m}^m c_l e^{ilh\xi} = c_{-1} e^{-ih\xi} + c_1 e^{ih\xi} + c_0 = 1 - \frac{\lambda}{2} a \left(e^{ih\xi} - e^{-ih\xi} \right) \\ &= 1 - i\lambda a \sin(h\xi) \end{aligned}$$

und

$$|\rho(\xi)|^2 = 1 + \lambda^2 a^2 \sin^2(h\xi)$$

Daraus folgt

$$\sup |\rho(\xi)| = \sqrt{1 + \lambda^2 a^2} > 1$$

und das naive Verfahren ist instabil für $a \neq 0$.

Für das Verfahren von Lax und Friedrichs

$$u_i^{j+1} = \frac{u_{i+1}^j + u_{i-1}^j}{2} - \frac{a\lambda}{2} (u_{i+1}^j - u_{i-1}^j)$$

erhalten wir

$$\rho(\xi) = \frac{1 - a\lambda}{2} e^{ih\xi} + \frac{1 + a\lambda}{2} e^{-ih\xi} = \cos(h\xi) - i\lambda a \sin(h\xi)$$

und

$$|\rho(\xi)|^2 = \cos^2(h\xi) + (a\lambda)^2 \sin^2(h\xi)$$

Daraus folgt

$$\sup |\rho(\xi)| \leq 1 \quad \Leftrightarrow \quad a^2 \lambda^2 \leq 1 \quad \Leftrightarrow \quad |a\lambda| \leq 1$$

und dies ist gerade die CFL-Bedingung.

1.4. Lineare 3-Punkt Verfahren, das Lax-Wendroff Verfahren. In diesem Abschnitt untersuchen wir lineare 3-Punkt Verfahren der Form

$$(1.1.16) \quad u_i^{j+1} = c_{-1} u_{i-1}^j + c_0 u_i^j + c_1 u_{i+1}^j$$

für eine lineare Advektionsgleichung $u_t + au_x = 0$. Dabei möchten wir allgemeine Bedingungen für die L^2 -Stabilität eines linearen Verfahrens, sowie ein Verfahren zweiter Ordnung herleiten.

Ein allgemeines 3-Punkt Verfahren (1.1.16) ist konservativ, falls $\sum u_i = \sum (Hu)_i$, also

$$\sum_i u_i = c_{-1} \sum_i u_{i-1} + c_0 \sum_i u_i + c_1 \sum_i u_{i+1} = (c_{-1} + c_0 + c_1) \sum_i u_i$$

Daraus erhält man zunächst die Bedingung

$$c_{-1} + c_0 + c_1 = 1$$

Mit $c_0 = 1 - c_1 - c_{-1}$ läßt sich das Verfahren auch in der Form

$$(1.1.17) \quad u_i^{j+1} = u_i^j - \lambda \left(\frac{c_{-1} u_i^j - c_1 u_{i+1}^j}{\lambda} - \frac{c_{-1} u_{i-1}^j - c_1 u_i^j}{\lambda} \right)$$

darstellen; die numerische Flussfunktion ist demnach

$$F(u, v) = \frac{c_{-1} u - c_1 v}{\lambda}$$

Das Verfahren soll weiterhin konsistent sein, d.h. es muss gelten

$$F(u, u) = f(u) = au \quad \forall u \in \mathbb{R}$$

und wir erhalten wir daraus eine weitere Bedingung an die Koeffizienten c_{-1} und c_1 in der Form

$$\frac{c_{-1} - c_1}{\lambda} u = au \quad \Leftrightarrow \quad c_{-1} - c_1 = a\lambda$$

Definieren wir $q = c_{-1} + c_1 = 1 - c_0$, so ergibt sich für das allgemeine 3-Punkt Verfahren

$$(1.1.18) \quad u_i^{j+1} = u_i^j - \underbrace{\lambda a}_{c_{-1}-c_1} \frac{u_{i+1}^j - u_{i-1}^j}{2} + \frac{\overbrace{q}^{c_{-1}+c_1}}{2} (u_{i+1}^j - 2u_i^j + u_{i-1}^j)$$

wobei q ein freier Parameter des Verfahrens ist.

Das Verfahren (1.1.18) ist nichts anderes als das naive Verfahren korrigiert um einen Diffusionsterm,

$$\frac{qh^2}{2} \frac{u_{i+1}^j - 2u_i^j + u_{i-1}^j}{h^2} = \frac{qh^2}{2} \partial_x^2 u(x_i, t_j) + O(h^4)$$

Man beachte hier wiederum die Bedeutung der modifizierten Gleichung: das naive Verfahren besitzt einen negativen Diffusionskoeffizienten und ist daher stets instabil. In (1.1.18) kann zu diesem Koeffizienten so viel positive Diffusion addiert werden, sodass das resultierende Verfahren stabil wird.

BEISPIEL 1.20. Wählen wir in (1.1.18) den Parameter $q = 1$, so erhalten wir

$$c_0 = 0, \quad c_1 = \frac{1 - a\lambda}{2}, \quad c_{-1} = \frac{1 + a\lambda}{2}$$

Dies liefert offensichtlich gerade das Verfahren von Lax und Friedrichs,

$$u_i^{j+1} = u_i^j - \lambda a \frac{u_{i+1}^j - u_{i-1}^j}{2} + \frac{1}{2} (u_{i+1}^j + u_{i-1}^j) - u_i^j$$

von dem wir bereits wissen, dass es stabil ist.

LEMMA 1.21. Ein lineares 3-Punkt Verfahren ist genau dann L^2 -stabil, wenn die Bedingung

$$(\lambda a)^2 \leq q \leq 1$$

erfüllt ist.

BEWEIS. Wir berechnen den Verstärkungsfaktor des 3-Punkt Verfahrens:

$$\begin{aligned} \rho(\xi) &= 1 - \frac{\lambda a}{2} (e^{ih\xi} - e^{-ih\xi}) + \frac{q}{2} (e^{ih\xi} - 2 + e^{-ih\xi}) \\ &= 1 - i\lambda a \sin(h\xi) + q(\cos(h\xi) - 1) \\ &= 1 - q + q \cos(h\xi) - i\lambda a \sin(h\xi) \\ &= 1 - q + q(c^2 - s^2) - i\lambda 2acs \end{aligned}$$

mit

$$c = \cos \frac{h\xi}{2} \quad s = \sin \frac{h\xi}{2}$$

Daraus folgt

$$\rho(\xi) = 1 - q + q(1 - 2s^2) - i\lambda a 2cs = 1 - 2qs^2 - i\lambda a 2cs$$

und

$$\begin{aligned} |\rho(\xi)|^2 &= (1 - 2qs^2)^2 + (2\lambda a)^2 c^2 s^2 = (1 - 2qs^2)^2 + 4(\lambda a)^2 (1 - s^2) s^2 \\ &= (1 - 2qt)^2 + 4(\lambda a)^2 t(1 - t) \end{aligned}$$

wobei $t = s^2 = \sin^2\left(\frac{h\xi}{2}\right)$ und $t \in [0, 1]$. Damit lässt sich das Quadrat des Verstärkungsfaktor als ein Polynom $f(t)$ darstellen,

$$|\rho(\xi)|^2 = f(t) = (1 - \alpha t)^2 + \beta t(1 - t), \quad \alpha = 2q, \quad \beta = 4(\lambda a)^2$$

mit $f(0) = 1$, $f \geq 0$ on $[0, 1]$. Insbesondere gilt $f(t) \leq 1$ für alle $t \in [0, 1]$ genau dann, wenn $f'(0) \leq 0$ und $f(1) \leq 1$. Daraus folgt

$$\begin{aligned} f(1) &= (1 - \lambda)^2 \leq 1 \Leftrightarrow 0 \leq \alpha \leq 2 \stackrel{\alpha=2q}{\Leftrightarrow} 0 \leq q \leq 1 \\ f'(0) &= \beta - 2\alpha \leq 0 \Leftrightarrow \underbrace{\beta}_{4(\lambda a)^2} \leq 2\alpha = 4q(\beta = 4(\lambda a)^2) \Leftrightarrow q \geq (\lambda a)^2 \end{aligned}$$

Dies ist gerade die Bedingung, die zu zeigen war. $\square$

Man kann nun den Parameter q so wählen, dass man ein Verfahren zweiter Ordnung erhält. Dazu muss q so gewählt werden, dass die numerische Viskosität identisch verschwindet, d.h. $\beta(u, \lambda) = 0$, wobei

$$\beta(u, \lambda) = \frac{1}{2} \left(\frac{1}{\lambda^2} \sum_{l=-m}^m l^2 \partial_l H(u, \dots, u) - a^2(u) \right)$$

Der numerische Viskosität von (1.1.18) ist gegeben durch

$$\begin{aligned} k\beta(u, \lambda) &= \frac{k}{2} \left(\frac{1}{\lambda^2} \left(\overbrace{\frac{\lambda a}{2} + \frac{q}{2}}^{l=-1} - \overbrace{\frac{\lambda a}{2} + \frac{q}{2}}^{l=1} \right) - a^2 \right) \\ &= \frac{k}{2} \left(\frac{q}{\lambda^2} - a^2 \right) \end{aligned}$$

Damit ist die einzige Möglichkeit, ein Verfahren zweiter Ordnung zu erhalten, die Wahl $q = (a\lambda)^2$ und dies liefert gerade das **Verfahren von Lax und Wendroff**.

SATZ 1.22. *Das Verfahren von Lax und Wendroff,*

$$w_i^{j+1} = w_i^j - \lambda a \frac{u_{i+1}^j - u_{i-1}^j}{2} + (\lambda a)^2 \frac{u_{i+1}^j - 2u_i^j + u_{i-1}^j}{2}$$

ist das einzige lineare 3-Punkt Verfahren zweiter Ordnung, das gleichzeitig konservativ, konsistent und L^2 -stabil ist.

Wir haben gesehen, dass 3-Punkt Verfahren L^2 -stabil, falls $(\lambda a)^2 \leq q \leq 1$. Hier existieren zwei Grenzfälle: $q = 1$ ergibt das Verfahren von Lax und Friedrichs, $q = (a\lambda)^2$ liefert das Verfahren von Lax und Wendroff. Beide Verfahren sind stabil unter der Bedingung $|\lambda a| \leq 1$, was gleichzeitig die CFL-Bedingung ist.

KOROLLAR 1.23. *Besitzt das zugrundeliegende Cauchyproblem eine genügend glatte Lösung, so ist das Verfahren von Lax und Wendroff zweiter Ordnung konvergent, sofern die CFL-Bedingung erfüllt ist.*

Als Übungsaufgabe berechne man die modifizierte Gleichung des Lax–Wendroff Verfahrens. Da die numerische Viskosität des Verfahrens identisch verschwindet, ist die modifizierte Gleichung keine Diffusionsgleichung, sondern eine **dispersive Gleichung**, d.h. die Gleichung beschreibt Oszillationen im Ortsraum.

1.5. Upwind–Verfahren und das Verfahren von Godunov. Der Grund, wieso das Verfahren von Lax und Friedrichs Stosslösungen stark verschmiert, ist folgender: die lineare Advektionsgleichung

$$u_t + au_x = 0, \quad a > 0$$

mit Anfangsbedingung $u(x, 0) = u_0(x)$ hat die exakte Lösung $u(x, t) = u_0(x - at)$. Dies bedeutet, dass die Lösung an einem Punkt (x_i, t_{j+1}) nur von Daten an den Punkten (y, t_j) mit $y < x_i$ abhängt. Im Verfahren von Lax und Friedrichs dagegen werden zur Berechnung der numerische Approximation am Punkt (x_i, t_{j+1}) Daten links und rechts vom Ortspunkt x_i verwendet, nämlich gerade eine Kombination (Interpolation) der Werte in den Punkten (x_{i-1}, t_j) und (x_{i+1}, t_j) .

Für die lineare Advektionsgleichung mit positiver Ausbreitungsgeschwindigkeit wäre es tatsächlich besser, nur die Werte an den Gitterpunkten x_{i-1} und x_i zu verwenden, etwa durch Verwendung eines Rückwärts–Differenzenquotienten für die Ortsableitung. Dies führt auf das Upwind–Verfahren, das durch die Gleichung

$$(1.1.19) \quad u_i^{j+1} = u_i^j - \lambda a(u_i^j - u_{i-1}^j), \quad a > 0$$

definiert ist. In diesem numerischen Verfahren werden Informationen nur von links nach rechts transportiert, genauso so wie bei der exakten Lösung.

Für negative Geschwindigkeiten $a < 0$ verwendet man dementsprechend ein Upwind–Verfahren in der Form

$$(1.1.20) \quad u_i^{j+1} = u_i^j - a\lambda(u_{i+1}^j - u_i^j)$$

Das Upwind–Verfahren läßt sich in der allgemeinen Form eines linearen Verfahrens aus Abschnitt 1.4 schreiben, wobei $q = |\lambda a|$ gewählt werden muss. Dies folgt aus den beiden Darstellungen (1.1.19) und (1.1.20), die sich einheitlich für $a \in \mathbb{R}$ auch als

$$(1.1.21) \quad u_i^{j+1} = u_i^j - \lambda a \frac{u_{i+1}^j - u_{i-1}^j}{2} + \frac{|\lambda a|}{2}(u_{i+1}^j - 2u_i^j + u_{i-1}^j)$$

schreiben lassen.

Unter der CFL–Bedingung $|\lambda a| \leq 1$ ist das Upwind–Verfahren (1.1.21) L^2 –stabil, da dann die Abschätzung $(\lambda a)^2 \leq |\lambda a| = q \leq 1$ gilt. Die numerische Viskosität des Verfahrens ist für $a > 0$ gegeben durch

$$\begin{aligned} \beta(u, \lambda) &= \frac{1}{2} \left[\frac{1}{\lambda^2} [\lambda a] - a^2 \right] = \frac{1}{2} \left(\frac{a}{\lambda} - a^2 \right) \\ &= \frac{1}{2\lambda^2} (\lambda a - (a\lambda)^2) = \frac{1}{2\lambda^2} \lambda a (1 - a\lambda) < 1 \end{aligned}$$

und es gilt $\beta(u, \lambda) > 0$ für $0 < a\lambda < 1$. Ist die Ausbreitungsgeschwindigkeit a negativ, so erhält man

$$\beta(u, \lambda) = -\frac{1}{2\lambda^2} \lambda a (1 + a\lambda)$$

und die numerische Viskosität ist positiv für $-1 < a\lambda < 0$.

BEMERKUNG 1.24. *Die numerische Viskosität des Verfahrens von Lax und Friedrichs war gegeben durch $\frac{1}{2\lambda^2}(1 - (\lambda a)^2)$ und ist damit größer als die des Upwind-Verfahrens, denn es gilt*

$$x(1-x) < 1-x^2 = (1+x)(1-x), \quad 0 < x < 1$$

Es ist komplizierter ein Upwind-Verfahren für allgemeine nichtlineare Gleichungen zu definieren: ist die Flussfunktion f monoton, so ist eine natürliche Erweiterung des Upwind-Verfahrens (1.1.21) gegeben durch

$$u_i^j = \begin{cases} u_i^j - \lambda(f(u_i^j) - f(u_{i-1}^j)), & f' > 0 \\ u_i^j - \lambda(f(u_{i+1}^j) - f(u_i^j)), & f' < 0 \end{cases}$$

Die numerische Flussfunktion F_{up} ist dabei gegeben durch

$$F_{up}(u, v) = \begin{cases} f(u), & f' > 0 \\ f(v), & f' < 0 \end{cases}$$

Ist der Fluss f nicht monoton, d.h. die charakteristische Richtung des Informationstransports kann wechseln, so ist eine Möglichkeit, eine numerische Flussfunktion in der Form

$$(1.1.22) \quad F_{up}(u, v) = \begin{cases} f(u), & \frac{f(u) - f(v)}{u - v} > 0 \\ f(v), & \frac{f(u) - f(v)}{u - v} < 0 \end{cases}$$

zu verwenden. Mit $a(u, v) = \frac{f(u) - f(v)}{u - v}$ schreibt man (1.1.22) auch in der Form

$$F_{up} = \frac{1}{2}(f(u) + f(v)) - |a(u, v)| \frac{(v - u)}{2}$$

Damit erhält man das **Verfahren von Roe**,

$$u_i^{j+1} = u_i^j - \lambda[F_{up}(u_i, u_{i+1}) - F_{up}(u_{i-1}, u_i)]$$

Ein allgemeineres Prinzip die Idee von Upwind-Verfahren für lineare Gleichungen auf nichtlineare zu übertragen, führt auf die sogenannten **Godunov-Verfahren**: die Idee hinter diesen Verfahren ist, in jedem Zeitschritt die exakte Lösung von restringierten Anfangswertproblemen definiert auf Gitterzellen zu berechnen. Dabei sind die Anfangsbedingungen jeweils durch stückweise-konstante Rekonstruktionen bezüglich der vorgegebenen Gitterstruktur gegeben, d.h. man löst jeweils die dadurch definierten **Riemannprobleme**.

Wir nehmen für den Moment an, dass die Flussfunktion f eine strikt-konvexe Funktion ist, d.h. die Ableitung f' ist streng monoton wachsend. Weiter sei eine stückweise-konstante Rekonstruktion auf der vorgegebenen Gitterstruktur gegeben, die wir aus den diskreten Werten der numerischen Approximation erhalten, siehe Bild 1.4.

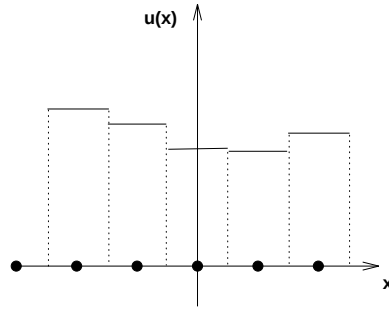


Bild 1.4 Anfangsbedingung in Godunov-Verfahren

Die zeitliche Evolution der stückweise-konstanten Rekonstruktion ist nun vollständig durch das Verhalten der Unstetigkeitsstellen an den Punkten $x_{i+1/2}$ bestimmt, die selbst als Lösungen von Riemannproblemen auf den Intervallen $[x_i, x_{i+1})$ bestimmt werden können. Die eindeutigen Entropielösungen dieser Riemannprobleme sind entweder

- Stosswellen, die mit einer Stossgeschwindigkeit $\dot{s} = \frac{f(u_r) - f(u_l)}{u_r - u_l}$ wandern

oder

- Verdünnungswellen, deren Grenzen mit den charakteristischen Geschwindigkeiten $f'(u_r)$ und $f'(u_l)$ laufen.

Die Ausbreitungsgeschwindigkeit ist damit offensichtlich beschränkt durch den Term

$$(1.1.23) \quad v = \sup_{|\xi| \leq \|u\|_{L^\infty}} |f'(\xi)|$$

Da benachbarte Unstetigkeitsstellen gerade im Gitterabstand h liegen, werden benachbarte Unstetigkeitsstellen nicht interagieren, sofern die Bedingung $k \cdot v \leq \frac{h}{2}$ erfüllt ist. Dies liefert gerade

$$\lambda f'(u_i^j) \leq \frac{1}{2}$$

In diesem Fall lassen sich benachbarte Riemannprobleme unabhängig voneinander betrachten und man erhält die exakte Lösung des Gesamtproblems einfach durch Zusammenfassung der einzelnen Riemannprobleme.

Um diese Idee der Godunov-Verfahren verwenden zu können, müssen wir die eindeutige Entropielösung des Riemannproblems

$$u_t + f(u)_x = 0, \quad u_0(x) = \begin{cases} u_l & : x < x_{i+\frac{1}{2}} \\ u_r & : x > x_{i+\frac{1}{2}} \end{cases}$$

auf dem Intervall $[x_i, x_{i+1})$ bestimmen. Da wir nur an Gittermittelwerten interessiert sind, können wir die gegebene Erhaltungsgleichung über dem Intervall $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}) \times [t_j, t_{j+1})$

integrieren und erhalten

$$\begin{aligned} \frac{1}{h} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u(t_{j+1}, x) dx &= \frac{1}{h} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u(t_j, x) dx \\ &\quad - \frac{1}{h} \left[\frac{1}{k} \int_{t_j}^{t_{j+1}} f(u(t, x_{i+\frac{1}{2}})) dt - \frac{1}{k} \int_{t_j}^{t_{j+1}} f(u(t, x_{i-\frac{1}{2}})) dt \right] \end{aligned}$$

Nach Definition der Zellmittelwerte ergibt sich damit

$$u_i^{j+1} = u_i^j - \lambda \left[\frac{1}{k} \int_{t_j}^{t_{j+1}} f(u(t, x_{i+\frac{1}{2}})) dt - \frac{1}{k} \int_{t_j}^{t_{j+1}} f(u(t, x_{i-\frac{1}{2}})) dt \right]$$

und man erkennt, dass es genügt, die Lösung der jeweiligen Riemannprobleme genau an den Punkten $x = x_{i+\frac{1}{2}}$ zu kennen, allerdings über dem kompletten Zeitintervall $t \in [t_j, t_{j+1}]$.

Für den Fall einer strikt-konvexen Flussfunktion, d.h. f' ist streng monoton wachsend, läßt sich die Lösung am Punkt $x = x_{i+\frac{1}{2}}$ mittels Fallunterscheidung genau angeben:

1) Für $f'(u_l) \geq 0$ und $f'(u_r) \geq 0$ gilt:

- a) Ist $f'(u_l) < f'(u_r)$, so folgt $u_l < u_r$ und die Entropielösung ist eine nach rechts wandernde Verdünnungswelle. solution will be a rarefaction wave moving to the right.
- b) Ist $f'(u_l) > f'(u_r)$, so folgt $u_l > u_r$ und die Entropielösung ist eine Stosswelle. Aus der Rankine-Hugoniot Bedingung erhält man die Stossgeschwindigkeit $\dot{s}$,

$$\dot{s} = \frac{f(u_r) - f(u_l)}{u_r - u_l} = f'(\xi)$$

wobei $\xi \in [u_l, u_r]$. Da $f'(u_l), f'(u_r) \geq 0$ und f strikt konvex ist, folgt $\dot{s} = f'(\xi) \geq 0$, d.h. die Stosswelle wandert nach rechts.

In beiden Fällen a) und b) gilt demnach $u(x_{i+\frac{1}{2}}, t) = u_l$ für alle $t \in [t_j, t_{j+1})$.

- 2) Für $f'(u_l) \geq 0$ and $f'(u_r) < 0$, erhält man $u_l > u_r$ und die Entropielösung ist wieder eine Stosswelle. Man ist allerdings nicht in der Lage, die Richtung der Welle anzugeben und daher gilt

$$u(x_{i+\frac{1}{2}}, t) = \begin{cases} u_l, & \text{if } \dot{s} \geq 0 \\ u_r, & \text{if } \dot{s} < 0 \end{cases}$$

- 3) Für $f'(u_l) < 0$ and $f'(u_r) \leq 0$, so sind wiederum Verdünnungs- und Stosswelle möglich:

- a) Ist $f'(u_l) < f'(u_r)$, erhalten wir $u_l < u_r$ und die Entropielösung ist eine nach links laufende Verdünnungswelle.

b) Ist $f'(u_l) > f'(u_r)$, so ist die Lösung eine links laufende Stosswelle.

Ähnlich zum Fall 1) erhalten wir in beiden Fällen $u(x_{i+\frac{1}{2}}, t) = u_r$ für alle $t \in [t_j, t_{j+1})$.

- 4) Für $f'(u_l) < 0$ and $f'(u_r) > 0$ folgt $u_l < u_r$. Die Lösung des Riemannproblems ist dann eine Verdünnungswelle, wobei die linke Grenze nach links, die rechte Grenze nach rechts wandert. Um die Lösung am Punkt $x = x_{i+\frac{1}{2}}$ zu bestimmen, müssen wir den Wert u mit $f'(u) = 0$ bestimmen, da dieser Punkt gerade den Wert auf der t -Achse liefert. Dazu schreiben wir

$$u(t, x_{i+\frac{1}{2}}) = u_s$$

wobei u_s die Lösung von $f' = 0$ bezeichnet. Diese Nullstelle ist eindeutig, da f als strikt konvex vorausgesetzt wurde, und nennt man auch **sonischen Punkt**.

In allen möglichen Fällen ist die Lösung am Punkt $x = x_{i+\frac{1}{2}}$ konstant über dem Zeitintervall $[t_j, t_{j+1})$, i.e.

$$u(x_{i+\frac{1}{2}}, t) = u^*(u_l, u_r) = u^*(u_i^j, u_{i+1}^j)$$

und man kann das Godunov-Verfahren daher in der Form

$$(1.1.24) \quad u_i^{j+1} = u_i^j - \lambda \left[f(u^*(u_i^j, u_{i+1}^j)) - f(u^*(u_{i-1}^j, u_i^j)) \right]$$

schreiben, wobei also $u^* = u^*(u, v)$ die eindeutige Entropielösung des Riemannproblems

$$w_t + f(w)_x = 0, \quad w_0(x) = \begin{cases} u & : x \geq 0 \\ v & : x < 0 \end{cases}$$

am Punkt $x = 0$ bezeichnet. Insbesondere gilt $u^*(u, u) = u$.

LEMMA 1.25. Für strikt konvexe Flussfunktionen f ist das Godunov-Verfahren (1.1.15) konsistent und konservativ. Eine einfachere Darstellung der numerischen Flussfunktion in (1.1.24) ist gegeben durch

$$(1.1.25) \quad F_G(u, v) = \begin{cases} \min_{w \in [u, v]} f(w) & : u \leq v \\ \max_{w \in [v, u]} f(w) & : u \geq v \end{cases}$$

wobei die Darstellung für den Fall $u \leq v$ die Fälle 1 a), 3 a) und 4) von oben abdeckt.

BEWEIS. Das Godunov-Verfahren ist nach Konstruktion konsistent und konservativ. Wir müssen also nur noch die Darstellung (1.1.25) aus den Fallunterscheidungen 1) – 4) ableiten:

- 1) Ist $f'(u_l), f'(u_r) \geq 0$, so liefert die Lösung des zugehörigen Riemannproblems

$$u(t, x_{i+\frac{1}{2}}) = u_l$$

und wir müssen die beiden Fälle $u_l > u_r$ und $u_l < u_r$ mit der Formel (1.1.25) vergleichen:

a) für $u_l < u_r$ folgt

$$\min_{w \in [u_l, u_r]} f(w) = f(u_l)$$

da f monoton wachsend ist. In diesem Fall gilt also

$$F_G(u, v) = f(u)$$

was mit der Darstellung in (1.1.25) übereinstimmt.

b) für $u_l > u_r$ folgt wieder

$$\max_{w \in [u_l, u_r]} f(w) = f(u_l)$$

da f monoton wachsend ist und

$$F_G(u, v) = f(u)$$

2) ist $f'(u_l) \geq 0$ und $f'(u_r) < 0$, so folgt $u_l > u_r$ und wir erhalten

$$\max_{w \in [u_r, u_l]} f(w) = \begin{cases} f(u_l) & : \text{ if } \overbrace{f(u_l) - f(u_r)}^{\sim \dot{s} > 0} > 0 \\ f(u_r) & : \text{ if } \underbrace{f(u_l) - f(u_r)}_{\dot{s} < 0} < 0 \end{cases}$$

Die Flussfunktion f kann wachsen oder fallen und daher

$$F_G(u, v) = \begin{cases} f(u) & : a(u, v) > 0 \\ f(v) & : a(u, v) < 0 \end{cases}$$

$$\text{mit } a(u, v) = \frac{f(u) - f(v)}{u - v}$$

3) Dieser Fall läuft analog zu 1) und man erhält

$$F_G(u, v) = f(u)$$

4) schliesslich gilt für $f'(u_l) < 0$ und $f'(u_r) > 0$ die Beziehung $u_l < u_r$ und es folgt

$$\min_{w \in [u_l, u_r]} f(w) = f(u_s)$$

sowie

$$F_G(u, v) = f(u_s)$$

□

Kommen wir zurück zur CFL-Bedingung: da die lokalen Riemannprobleme im Godunov-Verfahren nicht interagieren sollen, hatten wir die Bedingung

$$|\lambda f'| \leq \frac{1}{2}$$

gefordert. Wir wissen aber nun, dass die Lösung am Punkt $x = x_{i+1/2}$ über dem Zeitintervall $[t_j, t_{j+1}]$ konstant ist. Daraus kann man schliessen, dass das Verfahren funktioniert, solange der Wert $u^*(u_i^j, u_{i+1}^j)$ nicht durch ein benachbartes Riemannprobleme gestört wird. Da der Abstand zweier benachbarter Riemannprobleme zur Zeit $t = t_j$ gerade gleich h ist, erhalten wir wieder die gewöhnliche CFL-Bedingung in der Form

$$|\lambda f'| \leq 1$$

Wir wollen nun den Zusammenhang zwischen dem Godunov–Verfahren (1.1.25) und dem nichtlinearen Upwind–Verfahren (1.1.21) untersuchen: die numerische Flussfunktion im Upwind–Verfahren ist gegeben durch

$$F_{\text{up}}(u, v) = \begin{cases} f(u) & \text{if } a(u, v) > 0 \\ f(v) & \text{if } a(u, v) < 0 \end{cases}$$

mit $a(u, v) = \frac{f(u)-f(v)}{u-v}$; die Flussfunktion im Godunov–Verfahren ist von der Form

$$F_G(u, v) = f(u^*(u, v))$$

Wir betrachten nun wieder die 4 Fälle von oben:

- 1) Ist $f'(u), f'(v) \geq 0$, dann folgt auch $a(u, v) \geq 0$, da $a(u, v) = f'(\xi)$ mit $\xi \in [u, v]$ or $[v, u]$. Damit sind die numerischen Flussfunktionen in beiden Fällen identisch

$$F_{\text{up}}(u, v) = f(u) = F_G(u, v)$$

- 2) Ist $f'u \geq 0$ und $f'v < 0$, so ist die zugehörige Entropielösung eine Stosswelle, die nach links oder rechts läuft, also

$$\dot{s} \geq 0 \implies F_{\text{up}}(u, v) = f(u) = F_G(u, v)$$

$$\dot{s} < 0 \implies F_{\text{up}}(u, v) = f(v) = F_G(u, v)$$

- 3) In Analogie zum Fall 1) gilt

$$F_{\text{up}} = F_G(u, v) = f(v)$$

- 4) Im vierten Fall verwenden wir beim Godunov–Verfahren den sonischen Punkt,

$$F_G(u, v) = f(u_s)$$

Das Upwindverfahren lautet allerdings

$$F_{\text{up}}(u, v) = \begin{cases} f(u) & : a(u, v) > 0 \\ f(v) & : a(u, v) < 0 \end{cases}$$

und im Allgemeinen gilt $f(u_s) \neq f(u)$ und $f(u_s) \neq f(v)$.

Insgesamt gilt also $F_G(u, v) = F_{\text{up}}(u, v)$ bis auf den Fall, in dem eine Verdünnungswelle die Zellränder durchläuft. Weiter kann man zeigen, dass das Upwind–Verfahren gegen eine Stosswellenlösung konvergieren kann, die die Entropiebedingung verletzt. Dagegen werden wir später sehen, dass das Godunov–Verfahren stets gegen die korrekte und eindeutige Entropielösung des Problems konvergiert. Dies liegt allein an der Verwendung des sonischen Punkts im Godunov–Verfahren.

BEMERKUNG 1.26. *Das Upwind–Verfahren läßt sich auch als ein Godunov–Verfahren interpretieren, das lokal linear Riemannprobleme verwendet: im normalen Godunov–Verfahren löst man die lokalen Riemannprobleme*

$$u_t + f'(u)u_x = 0, \quad u(x, 0) = \begin{cases} u_l & : x < x_{i+\frac{1}{2}} \\ u_r & : x > x_{i+\frac{1}{2}} \end{cases}$$

Wir betrachten nun die Linearisierung

$$(1.1.26) \quad v_t + a(u_l, u_r)v_x = 0$$

mit der selben Anfangsbedingung wie oben. Die Lösung von (1.1.26) ist gegeben durch

$$v(t, x) = \begin{cases} u_l & x < x_{i+\frac{1}{2}} + ta(u_l, u_r) \\ u_r & x > x_{i+\frac{1}{2}} + ta(u_l, u_r) \end{cases}$$

d.h. die Unstetigkeit läuft n Abhängigkeit vom Vorzeichen von a nach links oder rechts. Bezeichnen wir die Lösung am Punkt $x = x_{i+\frac{1}{2}}$ mit $v^*(u_l, u_r)$ und verwenden diesen Wert an Stelle von u^* aus dem Godunov-Verfahren, so erhalten wir ein Verfahren mit numerischer Flussfunktion

$$F(u, v) = \begin{cases} f(u) & \text{if } a(u, v) > 0 \\ f(v) & \text{if } a(u, v) < 0 \end{cases}$$

und dies ist exakt die numerische Flussfunktion aus dem Upwind-Verfahren.

1.6. Konvergenz im nichtlinearen Fall, L^∞ -Stabilität, TVD-Eigenschaft.

In diesem Abschnitt untersuchen wir, wie die Konvergenz eines Verfahrens im Fall einer nichtlinearen Gleichung untersucht werden kann. Für lineare Gleichungen hatten wir den Äquivalenzsatz von Lax, da wir eine aus Taylorentwicklungen abgeleitete Rekursionsformel für den globalen Fehler des Verfahren zur Verfügung hatten. Dies ist nur für glatte Lösungen möglich, also bei nichtlinearen Gleichungen nicht anwendbar, da die eindeutige Entropielösung Sprungstellen haben wird.

In diesem Abschnitt werden wir allerdings noch nicht die Konvergenz gegen die Entropielösung zeigen, sondern nur die Konvergenz gegen eine schwache Lösung. Desweiteren benötigen wir neben der Konservativität und Konsistenz einige weiteren Eigenschaften von numerischen Verfahren, die diese Konvergenz sichern. Diese Zusatzeigenschaften werden wieder aus Eigenschaften von Lösungen des kontinuierlichen Problems abgeleitet.

Die eindeutige Entropielösung einer nichtlinearen Gleichung $u_t + f(u)_x = 0$ erfüllt die beiden folgenden Eigenschaft

- Es gilt folgende Abschätzung für die L^∞ -Norm der Lösung

$$(1.1.27) \quad \|u(t, \cdot)\|_{L^\infty} \leq \|u_0\|_{L^\infty}$$

für fast alle $t \in \mathbb{R}_+$.

- Es gilt folgende Abschätzung für die totale Variation der Lösung

$$(1.1.28) \quad \text{TV}(u(t, \cdot)) \leq \text{TV}(u_0)$$

für fast alle $t \in \mathbb{R}_+$. Dabei ist die totale Variation definiert durch

$$\begin{aligned} \text{TV}(u) &= \sup_{|\varphi| \leq 1} \int_{\mathbb{R}} \varphi' u \, dx \\ &= \limsup_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_{\mathbb{R}} |u(x + \epsilon) - u(x)| \, dx \end{aligned}$$

und $\text{BV}(\mathbb{R}) = \{u : u \in L^1_{loc}(\mathbb{R}), \text{TV}(u) < \infty\}$.

Die beiden Eigenschaften (1.1.27) und (1.1.28) lassen sich natürlich direkt auf ein Differenzenverfahren übertragen: ist $u_k(x) \in BV(\mathbb{R})$ eine stückweise konstante Funktion,

$$u_k(x) = u_i \quad x \in [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$$

dann definieren wir

$$TV(u_k) = \sum_{i \in \mathbb{Z}} |u_{i+1} - u_i|$$

Weiter setzen wir

$$\|u_k\|_{L^\infty} = \max_{i \in \mathbb{Z}} |u_i|$$

DEFINITION 1.27. Ein Differenzenverfahren definiert durch H ist L^∞ -stabil, falls für alle stückweise konstanten Funktionen $v_k \in BV(\mathbb{R})$ und $k \in [0, k_0]$ gilt

$$\|Hv_k\|_{L^\infty} \leq \|v_k\|_{L^\infty}$$

DEFINITION 1.28. Ein Differenzenverfahren definiert durch H besitzt die TVD-Eigenschaft (Total Variation Diminishing), falls für alle stückweise konstanten Funktionen $v_k \in BV(\mathbb{R})$ und $k \in [0, k_0]$ gilt

$$TV(Hv_k) \leq TV(v_k)$$

Let us check this additional properties for the Lax–Friedrichs scheme.

BEISPIEL 1.29. Das Verfahren von Lax und Friedrichs lautet

$$u_i^{j+1} = \frac{u_{i+1}^j + u_{i-1}^j}{2} - \lambda \frac{f(u_{i+1}^j) - f(u_{i-1}^j)}{2}$$

Um die L^∞ -Stabilität zu überprüfen, berechnet man unter der Annahme $u_{i+1}^j \neq u_{i-1}^j$

$$\begin{aligned} u_i^{j+1} &= \frac{u_{i+1}^j + u_{i-1}^j}{2} - \frac{u_{i+1}^j - u_{i-1}^j}{2} \lambda \frac{f(u_{i+1}^j) - f(u_{i-1}^j)}{u_{i+1}^j - u_{i-1}^j} \\ &= \frac{u_{i+1}^j + u_{i-1}^j}{2} - \frac{u_{i+1}^j - u_{i-1}^j}{2} \lambda f'(\xi_i^j), \end{aligned}$$

wobei $\xi_i^j \in (u_{i-1}^j, u_{i+1}^j)$. Aus der CFL-Bedingung folgt $|\lambda f'(\xi_i^j)| \leq 1$ und damit

$$u_i^{j+1} = \frac{u_{i+1}^j + u_{i-1}^j}{2} + \alpha \frac{u_{i+1}^j - u_{i-1}^j}{2}$$

für ein $\alpha \in [-1, 1]$. Also gilt $u_i^{j+1} \in (u_{i-1}^j, u_{i+1}^j)$ und

$$\|u_k^{j+1}\|_{L^\infty} \leq \|u_k^j\|_{L^\infty}$$

Zur TVD-Eigenschaft betrachtet man

$$\begin{aligned} u_i^{j+1} &= u_i^j + \frac{u_{i+1}^j - u_i^j}{2} - \lambda \frac{f(u_{i+1}^j) - f(u_i^j)}{2} - \left(\frac{u_i^j - u_{i-1}^j}{2} + \lambda \frac{f(u_i^j) - f(u_{i-1}^j)}{2} \right) \\ &= u_i^j + \frac{1}{2}(u_{i+1}^j - u_i^j)(1 - \lambda f'(\xi_i^j)) - \frac{1}{2}(u_i^j - u_{i-1}^j)(1 + \lambda f'(\xi_{i-1}^j)) \end{aligned}$$

wobei $\xi_i^j \in (u_i^j, u_{i+1}^j)$. Da $\text{TV}(u_k^{j+1}) = \sum_i |u_{i+1}^{j+1} - u_i^{j+1}|$ setzen wir

$$\begin{aligned} u_{i+1}^{j+1} - u_i^{j+1} &= u_{i+1}^j + \frac{1}{2}(u_{i+2}^j - u_{i+1}^j)(1 - \lambda f'(\xi_{i+1}^j)) - \frac{1}{2}(u_{i+1}^j - u_i^j)(1 + \lambda f'(\xi_i^j)) \\ &\quad - \left(u_i^j + \frac{1}{2}(u_{i+1}^j - u_i^j)(1 - \lambda f'(\xi_i^j)) - \frac{1}{2}(u_i^j - u_{i-1}^j)(1 + \lambda f'(\xi_{i-1}^j)) \right) \\ &= \frac{1}{2}(u_{i+2}^j - u_{i+1}^j)(1 - \lambda f'(\xi_{i+1}^j)) + \frac{1}{2}(u_i^j - u_{i-1}^j)(1 + \lambda f'(\xi_{i-1}^j)) \end{aligned}$$

Unter Zuhilfenahme der CFL-Bedingung folgt $1 \mp \lambda f'(\xi_{i\pm 1}^j) \geq 0$ für alle $i \in \mathbb{Z}$ und daher

$$\begin{aligned} \text{TV}(u_k^{j+1}) &= \sum_i |u_{i+1}^{j+1} - u_i^{j+1}| \\ &\leq \sum_i \left(\frac{1}{2} |u_{i+2}^j - u_{i+1}^j| \underbrace{(1 - \lambda f'(\xi_{i+1}^j))}_{>0} + \frac{1}{2} |u_i^j - u_{i-1}^j| \underbrace{(1 + \lambda f'(\xi_{i-1}^j))}_{>0} \right) \\ &= \sum_i |u_{i+1}^j - u_i^j| = \text{TV}(u_k^j) \end{aligned}$$

wobei der Index um 2 verschoben wurde. Damit besitzt das Verfahren von Lax und Friedrichs unter der CFL-Bedingung die TVD-Eigenschaft und ist L^∞ -stabil.

Die zentrale Aussage ist nun, das Verfahren, die L^∞ -stabil sind und die TVD-Eigenschaft besitzen, gegen eine schwache Lösung konvergieren.

SATZ 1.30. Sei durch H mit $(Hv)_i = H(v_{i-m}, \dots, v_{i+m})$ ein Differenzenverfahren definiert. Das Verfahren sei konservativ, konsistent mit lokal Lipschitz-stetiger Flussfunktion F , L^∞ -stabil und besitze die TVD-Eigenschaft. Weiter sei $u_0 \in BV(\mathbb{R}) \cap L^\infty$. Dann existiert eine schwache Lösung u der Erhaltungsgleichung $u_t + f(u)_x = 0$ und eine Folge $k_l \rightarrow 0, l \rightarrow \infty$, sodass für $k_l = \lambda h_l$, λ konstant, eine mittels H generierte Teilfolge $u_{k_l}(x, t)$ in $L^\infty((0, T), L_{loc}^1(\mathbb{R}))$ für jedes $T > 0$ gegen u konvergiert.

BEMERKUNG 1.31. Der Satz liefert nur die Konvergenz gegen eine schwache Lösung, nicht notwendigerweise gegen die eindeutige Entropielösung. Für das nichtlineare Upwind-Verfahren nach Roe kann man etwa ein Beispiel angeben, für das eine Teilfolge des Verfahrens gegen die korrekte Verdünnungswelle, eine andere Teilfolge aber gegen eine falsche Stosswelle konvergiert.

Zum Beweis des Satzes benötigen wir noch zwei Hilfsmittel:

LEMMA 1.32. Gegeben sei ein TVD-Verfahren mit einer lokal Lipschitz-stetigen numerischen Flussfunktion. Weiter sei das Verfahren L^∞ -stabil. Dann existiert eine Konstante $C > 0$, sodass für alle $n \geq m \geq 0$

$$\|H^n(u_k) - H^m(u_k)\|_{L^1} \leq C(n - m)k\text{TV}(u_k)$$

wobei $u_k(x)$ eine stückweise konstante Funktion ist.

BEWEIS. Wir schreiben

$$(H(u_k))_i - u_i = -\lambda(f_{i+\frac{1}{2}} - f_{i-\frac{1}{2}})$$

mit

$$f_{i+\frac{1}{2}} = f_{i+\frac{1}{2}}(u_{i-m+1}, \dots, u_{i+m}), \quad f_{i-\frac{1}{2}} = f_{i-\frac{1}{2}}(u_{i-m}, \dots, u_{i+m-1})$$

Da f Lipschitz-stetig ist, erhalten wir mit $c_1 = c_1(f, \|u_k\|_{L^\infty})$

$$|(H(u_k))_i - u_i| \leq \lambda c_1 \sum_{l=-m+1}^m |u_{i+l} - u_{i+l-1}|$$

Summation über $i \in \mathbb{Z}$ und Multiplikation mit $1/h$ ergibt dann

$$\|H(u_k) - u_k\|_{L^1} \leq 2mc_1 k \text{TV}(u_k)$$

Mit der Beziehung $H^{l+1}(u) = H(H^l(u))$ erhält man auf die gleiche Weise

$$\|H^{l+1}(u_k) - H^l(u_k)\| \leq 2mc_2 k \text{TV}(H^l(u_k)) \leq 2mc_2 k \text{TV}(u_k)$$

aufgrund der TVD-Eigenschaft und

$$c_2 = c_2(f, \|H^l u_k\|_{L^\infty}) = c_2(f, \|u_k\|_{L^\infty}), \quad l \in \mathbb{N}$$

da das Verfahren L^∞ -stabil ist. Für $n > m$ gilt

$$H^n(u_k) - H^m(u_k) = \sum_{l=m}^{n-1} (H^{l+1}(u_k) - H^l(u_k))$$

und dies ergibt

$$\|H^n(u_k) - H^m(u_k)\|_{L^1} \leq \sum_{l=m}^{n-1} C k \text{TV}(u_k) = (n - m) C k \text{TV}(u_k)$$

□

LEMMA 1.33. *Gegeben sei ein konservatives, L^∞ -stabiles TVD-Verfahren mit lokal Lipschitz-stetiger Flussfunktion. Dann existiert eine Konstante $C > 0$ (abhängig von f und $\|u_0\|_{L^\infty}$), sodass für alle $T > 0$ gilt*

- 1) $\|u_k(t, \cdot)\|_{L^1} \leq \|u_k(0, \cdot)\|_{L^1} + C \cdot T \cdot \text{TV}(u_k(0, \cdot)), \quad 0 \leq t \leq T$
- 2) $\|u_k(t, \cdot) - u_k(s, \cdot)\|_{L^1} \leq C(|t - s| + k) \text{TV}(u_k(0, \cdot)), \quad 0 \leq s, t \leq T$
- 3) $\text{TV}(u_k(\cdot, t)) \leq \text{TV}(u_k(\cdot, 0))$

BEWEIS. Die dritte Aussage folgt offensichtlich aus der TVD-Eigenschaft; Teil 1) folgt direkt aus dem vorangegangenen Lemma. Sei also $t_m \leq s \leq t_{m+1}$ und $t_n \leq t \leq t_{n+1}$. Dann gilt

$$|t_m - t_n| \leq |t - s| + k$$

und

$$u_k(t, \cdot) - u_k(s, \cdot) = u_k(t_n, \cdot) - u_k(t_m, \cdot)$$

Damit folgt ebenfalls die zweite Aussage aus obigem Lemma. □

BEWEIS. (zu Satz 1.30) Wir geben hier nur die wesentlichen Schritte an, ohne alle Einzelheiten genau auszuführen. Die Idee ist, das Ergebnis aus den Satz von Lax und Wendroff aus Abschnitt 1.1 zurückzuführen. Dazu müssen wir überprüfen, dass

- 1) die stückweise konstanten Funktionen u_k in $L^\infty((0, T), \mathbb{R})$ beschränkt sind,
- 2) die Folge $u_k(t, \cdot)$ in $L^1_{loc}((0, T), \mathbb{R})$ und f.ü. in $(0, T) \times \mathbb{R}$ konvergiert.

Wir zeigen zunächst, dass $u_k(t, \cdot)$ in $BV(\mathbb{R})$ und für $t \in [0, T]$ gleichmäßig beschränkt sind: es gilt

$$\begin{aligned} \text{TV}(u_k(0, \cdot)) &= \sum_i |u_{i+1}^0 - u_i^0| = \sum_i \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \frac{|u_0(x+h) - u_0(x)|}{h} dx \\ &\leq \int_{\mathbb{R}} \frac{|u_0(x+h) - u_0(x)|}{h} dx \end{aligned}$$

Daraus folgt $\text{TV}(u_k(0, \cdot)) \leq \text{TV}(u_0)$ und mit Teil 3) aus Lemma 1.33

$$\text{TV}(u_k(t, \cdot)) \leq \text{TV}(u_0)$$

Also sind die $u_k(t, \cdot)$ gleichmäßig beschränkt in $BV(\mathbb{R})$.

Gleichzeitig ist aber $u_k(t, \cdot)$ gleichmäßig beschränkt in $L^\infty(\mathbb{R})$, da das Verfahren L^∞ -stabil ist.

Danach zeigen wir, dass die Folge $u_k(t, \cdot)$ für $k \rightarrow 0$ konvergiert. Zunächst beweist man, dass die kanonische Einbettung von $BV(\Omega)$ in $L^1(\Omega)$, $\Omega \subset \mathbb{R}^d$ mit Lipschitz-stetigem Rand, kompakt ist. Sei dann $(s_n)_{n \in \mathbb{N}}$ eine abzählbare Teilmenge in $[0, T]$. Damit existiert für jedes s_l eine in $BV(\mathbb{R})$ beschränkte Teilfolge $u_{k_l}(s_l, \cdot)$, die in $L^1_{loc}(\mathbb{R})$ konvergiert, d.h.

$$\int_{\Omega} |u_{k_l}(s_l, \cdot) - u(s_l, \cdot)| dx \rightarrow 0$$

für $k_l \rightarrow 0$ und jede beschränkte Menge $\Omega \subset \mathbb{R}$.

$\Omega \subset \mathbb{R}$.

Gleichzeitig zeigt ein Diagonalisierungsargument, dass die gleiche Teilfolge $u_{k_l}(s_l, \cdot)$ für jedes s_j gegen $u(s_j, \cdot)$ konvergiert. Um dieses Ergebnis für alle $t \in [0, T]$ zu erhalten, betrachten wir eine Zerlegung des Intervalls $[0, T]$ in Teilintervalle (t_N, t_{N+1}) , $t_N \in (s_n)_{n \in \mathbb{N}}$, wobei die Länge der Teilintervalle kleiner als ein ε ist.

Dann existiert für jedes $t \in [0, T]$ ein t_N mit $|t_N - t| < \varepsilon$ und für $k_l > k_{\bar{l}}$ gilt die Abschätzung

$$\begin{aligned} \|u_{k_l}(\cdot, t) - u_{k_{\bar{l}}}(\cdot, t)\|_{L^1} &\leq \|u_{k_l}(\cdot, t) - u_{k_l}(\cdot, t_N)\|_{L^1} \\ &+ \|u_{k_l}(\cdot, t_N) - u_{k_{\bar{l}}}(\cdot, t_N)\|_{L^1} \\ &+ \|u_{k_{\bar{l}}}(\cdot, t_N) - u_{k_{\bar{l}}}(\cdot, t)\|_{L^1} \end{aligned}$$

Aus Teil 2) von Lemma 1.33 wissen wir, dass der erste und dritte Term auf der rechten Seite durch $C \cdot (\varepsilon + k_l)$ beschränkt sind. Der mittlere Term entsteht durch eine konvergente Folge aus dem zweiten Schritt. Daher ist $u_{k_l}(t, \cdot)$ eine Cauchyfolge in $L^1_{loc}(\mathbb{R})$ und gleichmäßig in t und wir können eine Teilfolge auswählen, die in $L^\infty((0, T), L^1_{loc}(\mathbb{R}))$ und f.ü. konvergiert

und gleichzeitig in $L^\infty(\mathbb{R} \times (0, T))$ beschränkt ist. Damit haben wir die Voraussetzungen des Satzes von Lax und Wendroff und können damit auf die Konvergenz gegen eine schwache Lösung schliessen. $\square$

Erinnern wir uns an das Beispiel 1.29, so sehen wir, dass die Überprüfung, ob ein Verfahren L^∞ -stabil ist oder die TVD-Eigenschaft besitzt, aufwendig sein kann. Man sucht daher nach Möglichkeiten, die Überprüfung einfacher zu machen.

DEFINITION 1.34. *Ein Verfahren definiert durch H lässt sich in inkrementeller Form schreiben, falls lokal beschränkte Funktionen C und D existieren, sodass*

$$u_i^{j+1} = u_i^j + C_{i+\frac{1}{2}}^j (u_{i+1}^j - u_i^j) - D_{i-\frac{1}{2}}^j (u_i^j - u_{i-1}^j)$$

wobei

$$\begin{aligned} C_{i+\frac{1}{2}}^j &= C(u_{i-p+1}^j, \dots, u_{i+p}^j) \\ D_{i+\frac{1}{2}}^j &= D(u_{i-p+1}^j, \dots, u_{i+p}^j) \end{aligned}$$

und $p \in \mathbb{N}_0$.

BEISPIEL 1.35. *Wir betrachten wieder das Verfahren von Lax und Friedrichs:*

$$\begin{aligned} u_i^{j+1} &= u_i^j + \frac{1}{2} \left(1 - \lambda \frac{f(u_{i+1}^j) - f(u_i^j)}{u_{i+1}^j - u_i^j} \right) (u_{i+1}^j - u_i^j) \\ &\quad - \frac{1}{2} \left(1 + \lambda \frac{f(u_i^j) - f(u_{i-1}^j)}{u_i^j - u_{i-1}^j} \right) (u_i^j - u_{i-1}^j) \end{aligned}$$

Damit sind die beiden Funktionen C und D gegeben durch

$$\begin{aligned} C(u, v) &= \frac{1}{2} \left(1 - \lambda \frac{f(v) - f(u)}{v - u} \right) \\ D(u, v) &= \frac{1}{2} \left(1 + \lambda \frac{f(v) - f(u)}{v - u} \right) \end{aligned}$$

und, da $f \in C^1$ ist, auch lokal beschränkt.

SATZ 1.36. *Das Differenzenverfahren definiert durch H sei in inkrementeller Form gegeben und es gelte $C_{i+\frac{1}{2}}^j, D_{i+\frac{1}{2}}^j \geq 0$ für alle $i \in \mathbb{Z}$.*

- 1) *Gilt $C_{i+\frac{1}{2}}^j + D_{i-\frac{1}{2}}^j \leq 1$, so ist das Verfahren L^∞ -stabil.*
- 2) *Gilt $C_{i+\frac{1}{2}}^j + D_{i+\frac{1}{2}}^j \leq 1$, so besitzt das Verfahren die TVD-Eigenschaft.*

BEMERKUNG 1.37. *Ist ein Verfahren konservativ und konsistent, so haben wir*

$$\begin{aligned}
 u_i^{j+1} &= u_i^j - \lambda(f_{i+\frac{1}{2}} - f_{i-\frac{1}{2}}) \\
 &= u_i^j + \lambda(f(u_i^j) - f_{i+\frac{1}{2}} - [f(u_i^j) - f_{i-\frac{1}{2}}]) \\
 &= u_i^j + \lambda \frac{F(u_i^j, \dots, u_i^j) - F(u_{i-m+1}^j, \dots, u_{i+m}^j)}{u_{i+1}^j - u_i^j} \cdot (u_{i+1}^j - u_i^j) \\
 &\quad - \lambda \frac{F(u_i^j, \dots, u_i^j) - F(u_{i-m}^j, \dots, u_{i+m-1}^j)}{u_i^j - u_{i-1}^j} (u_i^j - u_{i-1}^j)
 \end{aligned}$$

Daher muss $F(u_{i-m+1}^j, \dots, u_{i+m}^j)$ für $u_{i+1}^j \rightarrow u_i^j$ gegen $F(u_i^j, \dots, u_i^j)$ konvergieren, um ein Verfahren in inkrementeller Form schreiben zu können. Anderenfalls wären die Koeffizienten nicht lokal beschränkt. Dies gilt natürlich für 3-Punkt Verfahren.

Aufgrund der letzten Bemerkung definiert man

DEFINITION 1.38. *Ein konservatives Differenzenverfahren definiert durch H nennt man essentiell 3-Punkt, falls die numerische Flussfunktion F für $u \in \mathbb{R}$ die strengere Konsistenzbedingung*

$$F(u_{-m+1}, \dots, u_{-1}, u, u, u_2, \dots, u_m) = f(u)$$

erfüllt.

BEMERKUNG 1.39. *Jedes konsistente und konservative 3-Punkt Verfahren ist essentiell 3-Punkt.*

LEMMA 1.40. *Ist ein Verfahren essentiell 3-Punkt mit lokal Lipschitz-stetiger Flussfunktion F , so lässt sich das Verfahren in inkrementeller Form schreiben mit*

$$C_{i+\frac{1}{2}}^j = \lambda \frac{f(u_i^j) - f_{i+\frac{1}{2}}}{u_{i+1}^j - u_i^j}, \quad D_{i+\frac{1}{2}}^j = \lambda \frac{f(u_{i+1}^j) - f_{i+\frac{1}{2}}}{u_{i+1}^j - u_i^j}$$

Eine andere hilfreiche Darstellung von Differenzenverfahren ist die viskose Form.

DEFINITION 1.41. *Ein Differenzenverfahren lässt sich in viskoser Form darstellen, falls eine lokal beschränkte Funktion $Q = Q(u_{-m+1}, \dots, u_m)$ existiert, sodass*

$$u_i^{j+1} = u_i^j - \lambda \frac{f(u_{i+1}^j) - f(u_{i-1}^j)}{2} + \frac{Q_{i+\frac{1}{2}}^j (u_{i+1}^j - u_i^j) - Q_{i-\frac{1}{2}}^j (u_i^j - u_{i-1}^j)}{2}$$

mit $Q_{i+\frac{1}{2}} = Q(u_{i-m+1}, \dots, u_{i+m})$.

BEISPIEL 1.42. *Wir betrachten wieder das Verfahren von Lax und Friedrichs*

$$\begin{aligned}
 u_i^{j+1} &= \frac{u_{i+1}^j + u_{i-1}^j}{2} - \lambda \frac{f(u_{i+1}^j) - f(u_{i-1}^j)}{2} \\
 &= u_i^j - \lambda \frac{f(u_{i+1}^j) - f(u_{i-1}^j)}{2} + \frac{u_{i+1}^j - u_i^j}{2} - \frac{u_i^j - u_{i-1}^j}{2}
 \end{aligned}$$

Daher gilt $Q(u, v) = 1$.

Bleiben wir kurz beim Verfahren von Lax und Friedrichs und definieren

$$\begin{aligned} C_{i+\frac{1}{2}}^j &= \frac{1}{2}(Q_{i+\frac{1}{2}}^j - \lambda \frac{f(u_{i+1}^j) - f(u_i^j)}{u_{i+1}^j - u_i^j}) \\ D_{i+\frac{1}{2}}^j &= \frac{1}{2}(Q_{i+\frac{1}{2}}^j + \lambda \frac{f(u_{i+1}^j) - f(u_i^j)}{u_{i+1}^j - u_i^j}), \end{aligned}$$

so lautet die inkrementelle Form

$$\begin{aligned} & u_i^j + C_{i+\frac{1}{2}}^j(u_{i+1}^j - u_i^j) - D_{i-\frac{1}{2}}^j(u_i^j - u_{i-1}^j) \\ &= u_i^j - \frac{\lambda}{2}(f(u_{i+1}^j) - f(u_i^j)) - \frac{\lambda}{2}(f(u_i^j) - f(u_{i-1}^j)) \\ & \quad + \frac{1}{2}Q_{i+\frac{1}{2}}^j(u_{i+1}^j - u_i^j) - \frac{1}{2}Q_{i-\frac{1}{2}}^j(u_i^j - u_{i-1}^j) \\ &= u_i^j - \lambda \frac{f(u_{i+1}^j) - f(u_{i-1}^j)}{2} + \frac{Q_{i+\frac{1}{2}}^j(u_{i+1}^j - u_i^j) - Q_{i-\frac{1}{2}}^j(u_i^j - u_{i-1}^j)}{2} \end{aligned}$$

Im allgemeinen Fall gilt folgende Aussage:

- LEMMA 1.43. 1) Ist ein Verfahren in viskoser Form gegeben, so ist das Verfahren essentiell 3-Punkt und kann in inkrementeller Form mit Koeffizienten $C_{i+\frac{1}{2}}^j$ und $D_{i+\frac{1}{2}}^j$ wie oben angegeben geschrieben werden.
- 2) Ist ein Verfahren essentiell 3-Punkt mit lokal Lipschitz-stetiger Flussfunktion F , dann kann das Verfahren in viskoser Form geschrieben werden mit

$$Q_{i+\frac{1}{2}} = C_{i+\frac{1}{2}} + D_{i+\frac{1}{2}} = \lambda \frac{f(u_i) + f(u_{i+1}) - 2f_{i+\frac{1}{2}}}{u_{i+1} - u_i}$$

Eine direkte Folgerung ist dann

SATZ 1.44. Sei ein Verfahren in viskoser Form gegeben. Dann gilt:

- a) Gilt $\lambda \left| \frac{f(u_{i+1}^j) - f(u_i^j)}{u_{i+1}^j - u_i^j} \right| \leq Q_{i+\frac{1}{2}}^j \leq 1$, so ist das Verfahren TVD.
- b) Gilt $\lambda \left| \frac{f(u_{i+1}^j) - f(u_i^j)}{u_{i+1}^j - u_i^j} \right| \leq Q_{i+\frac{1}{2}}^j \leq \frac{1}{2}$, so ist das Verfahren L^∞ -stabil.

BEWEIS. a) Wir nehmen an, dass

$$\lambda \left| \frac{f(u_{i+1}^j) - f(u_i^j)}{u_{i+1}^j - u_i^j} \right| \leq Q_{i+\frac{1}{2}}^j \leq 1$$

und die beiden Terme $C_{i+\frac{1}{2}}^j$ und $D_{i+\frac{1}{2}}^j$ wie oben. Dann gilt

$$C_{i+\frac{1}{2}}^j \geq 0, \quad D_{i+\frac{1}{2}}^j \geq 0$$

and

$$C_{i+\frac{1}{2}}^j + D_{i+\frac{1}{2}}^j = Q_{i+\frac{1}{2}}^j \leq 1$$

Damit besitzt das Verfahren die TVD-Eigenschaft.

b) Wir nehmen an, dass

$$\lambda \left| \frac{f(u_{i+1}^j) - f(u_i^j)}{u_{i+1}^j - u_i^j} \right| \leq Q_{i+\frac{1}{2}}^j \leq \frac{1}{2}$$

Dann gilt $C_{i+\frac{1}{2}}^j, D_{i+\frac{1}{2}}^j \in [0, \frac{1}{2}]$ und daher

$$C_{i+\frac{1}{2}}^j + D_{i+\frac{1}{2}}^j \in [0, 1]$$

Damit ist das Verfahren L^∞ -stabil. □

Für ein konservatives und konsistentes 3-Punkt Verfahren, das in viskoser Form geschrieben werden kann, gilt also

$$\begin{aligned} C(u, v) &= \frac{1}{2} \left(Q(u, v) - \lambda \frac{f(v) - f(u)}{v - u} \right) = \lambda \frac{f(u) - F(u, v)}{v - u} \\ D(u, v) &= \frac{1}{2} \left(Q(u, v) + \lambda \frac{f(v) - f(u)}{v - u} \right) = \lambda \frac{f(v) - F(u, v)}{v - u} \\ Q(u, v) &= \lambda \frac{f(u) + f(v) - 2F(u, v)}{v - u} \end{aligned}$$

und $Q = C + D$.

BEISPIEL 1.45. Beim Verfahren von Lax und Friedrichs gilt $Q_{i+\frac{1}{2}}^j = 1$. Die Bedingung aus dem letzten Satz lautet dann

$$\lambda \max_i \left| \frac{f(u_{i+1}) - f(u_i)}{u_{i+1} - u_i} \right| \leq 1$$

was gerade wieder die CFL-Bedingung ergibt. Dies liefert einen einfachen Beweis, dass das Lax-Friedrichs Verfahren die TVD-Eigenschaft besitzt.

BEISPIEL 1.46. Die numerische Flussfunktion im Verfahren von Roe ist gegeben durch

$$F_R = \begin{cases} f(u), & a(u, v) > 0 \\ f(v), & a(u, v) < 0 \end{cases}$$

beziehungsweise

$$F_R = \frac{1}{2} (f(u) + f(v)) - |a(u, v)| \frac{(v - u)}{2}, \quad Q(u, v) = \lambda |a(u, v)|$$

Damit folgt

$$Q_{i+\frac{1}{2}}^j = \lambda |a(u_i, u_{i+1})| = \lambda \left| \frac{f(u_{i+1}) - f(u_i)}{u_{i+1} - u_i} \right|$$

und das Verfahren besitzt unter der CFL-Bedingung die TVD-Eigenschaft.

BEISPIEL 1.47. Die Verallgemeinerung des Verfahrens von Lax und Wendroff auf den nichtlinearen Fall ist gegeben durch

$$u_i^{j+1} = u_i^j - \lambda \frac{f(u_{i+1}) - f(u_{i-1}))}{2} + \lambda^2 \left[a(u_i, u_{i+1}) \frac{f(u_{i+1}) - f(u_i)}{2} - a(u_{i-1}, u_i) \frac{f(u_i) - f(u_{i-1}))}{2} \right]$$

Damit ist die numerische Flussfunktion gegeben durch

$$F(u, v) = \frac{f(u) + f(v)}{2} - \lambda a(u, v) \frac{f(v) - f(u)}{2}$$

Man erhält also für Q den Ausdruck

$$Q = \lambda \frac{f(u) + f(v) - 2F(u, v)}{v - u} = (\lambda a)^2$$

und

$$Q_{i+\frac{1}{2}}^j = (\lambda a_{i+\frac{1}{2}})^2 \quad a_{i+\frac{1}{2}} = a(u_i, u_{i+1})$$

Demnach ist die Bedingung aus dem Satz nicht erfüllt, d.h. das Verfahren ist nicht notwendigerweise TVD. In der Tat lassen sich Beispiele angeben, in denen die totale Variation anwächst.

1.7. Monotone Verfahren. In diesem Abschnitt übertragen wir eine weitere Eigenschaft des kontinuierlichen Problems auf ein numerisches Verfahren, nämlich die **Monotonieeigenschaft**: sind u_0 und v_0 zwei gegebene Anfangsbedingungen zur Zeit $t = 0$ für die Gleichung $u_t + f(u)_x = 0$ mit $u_0 \geq v_0$, so erfüllen die zugehörigen Entropielösungen die Ungleichung $u(\cdot, t) \geq v(\cdot, t)$ fast überall in $\mathbb{R}$ und für fast alle $t \in \mathbb{R}_+$.

DEFINITION 1.48. Ein Verfahren definiert durch H ist **monoton**, falls für alle $u_i \leq v_i$ folgt: $(Hu)_i \geq (Hv)_i$ für alle $i \in \mathbb{Z}$.

Ein Verfahren ist offensichtlich **monoton**, falls die Gittertransferfunktion H in jedem Argument **monoton wachsend** ist.

BEISPIEL 1.49. Für das Verfahren von Lax und Friedrichs gilt

$$H(u_{-1}, u_0, u_1) = \frac{u_1 + u_{-1}}{2} - \lambda \frac{f(u_1) - f(u_{-1}))}{2}$$

und man berechnet direkt

$$\partial_{-1}H = \frac{1 + \lambda f'(u_{-1}))}{2}, \quad \partial_0H = 0, \quad \partial_1H = \frac{1 - \lambda f'(u_1))}{2}$$

Ist also die CFL-Bedingung erfüllt, so sind alle Ableitungen **nicht-negativ** und daher H in jedem Argument **monoton wachsend**. Also ist das Lax-Friedrichs Verfahren **monoton**.

Beim Godunov-Verfahren erinnern wir uns an die beiden Schritte des Verfahrens: die exakte Berechnung der lokalen Riemannprobleme mit stückweise konstanten Funktionen sowie die sich daran anschließende Berechnung der neuen Zellmittelwerte. Da wir im ersten

Schritt die exakte Entropielösung berechnen, erfüllt dieser Teil des Verfahrens offensichtlich die Monotonieeigenschaft. Im zweiten Schritt berechnet man für $u \geq v$

$$\frac{1}{h} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u dx \geq \frac{1}{h} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} v dx$$

Die erste Aussage bezüglich monotoner Verfahren ist

SATZ 1.50. *Ist ein Verfahren definiert durch H monoton und konservativ, so ist es L^∞ -stabil und besitzt die TVD-Eigenschaft.*

Wir zeigen zunächst

LEMMA 1.51. *(Crandall–Tartar)*

Sei $M \in L^1(\mathbb{R})$ mit $0 \in M$ und $\max(u, v) \in M$, falls $u, v \in M$. Weiter sei $T : M \rightarrow L^1(\mathbb{R})$, sodass

$$\int_{\mathbb{R}} T u dx = \int_{\mathbb{R}} u dx$$

Dann sind die beiden folgenden Aussagen äquivalent:

- 1) *T ist monoton auf M , d.h. gilt $u \leq v$ f.ü., so folgt $Tu \leq Tv$ f.ü.*
- 2) *T ist eine Kontraktion, d.h. $\|Tu - Tv\|_{L^1} \leq \|u - v\|_{L^1}$ für alle $u, v \in M$.*

BEWEIS. Wir zeigen zunächst die Richtung 1) $\Rightarrow$ 2): Sei $\omega_t = \max(\omega, 0)$. Dann folgt mit $\omega \in M$ ebenfalls $\omega_t \in M$. Nun gilt $\max(u, v) \geq u, v$ und, da T monoton ist, $T(\max(u, v)) \geq T(u), T(v)$. Wir erhalten also

$$T(\max(u, v)) - Tv \geq \max(Tu - Tv, 0) = (Tu - Tv)_t$$

und

$$\begin{aligned} \int_{\mathbb{R}} (Tu - Tv)_t dx &\leq \int_{\mathbb{R}} (T(\max(u, v)) - Tv) dx = \int_{\mathbb{R}} (\max(u, v) - v) dx \\ &\leq \int_{\mathbb{R}} (u - v)_t dx \end{aligned}$$

Da aber $|u - v| = (u - v)_t + (v - u)_t$, folgt

$$\int_{\mathbb{R}} |Tu - Tv| dx \leq \int_{\mathbb{R}} |u - v| dx$$

d.h. die Abbildung T ist eine Kontraktion.

Für die umgekehrte Richtung beobachten wir zunächst, dass

$$(u - v)_t = \frac{1}{2}(|u - v| + u - v)$$

und daher

$$\begin{aligned} \int_{\mathbb{R}} (Tu - Tv)_t dx &= \frac{1}{2} \int_{\mathbb{R}} (|Tu - Tv| + Tu - Tv) dx \\ &\leq \frac{1}{2} \int_{\mathbb{R}} (|u - v| + u - v) dx = \int_{\mathbb{R}} (u - v)_t dx \end{aligned}$$

Gilt $u \leq v$ f.ü., so folgt $(u - v)_t = 0$ f.ü. und

$$(1.1.29) \quad \int_{\mathbb{R}} (Tu - Tv)_t dx \leq 0$$

Da aber $\omega_t \geq 0$ f.ü. für alle $\omega \in M$ gilt, folgt $(Tu - Tv)_t \geq 0$ f.ü. Zusammen mit (1.1.29) erhalten wir also $(Tu - Tv)_t = 0$ f.ü. und nach Definition von ω_t gilt $Tu \leq Tv$ f.ü., d.h. die Abbildung T ist monoton. $\square$

LEMMA 1.52. *Ein monotones und konservatives Verfahren definiert durch H ist L^1 -kontrahierend, d.h. die Funktion H erfüllt für alle stückweise konstanten Funktionen $u_k, v_k \in L^1(\mathbb{R})$ und alle $k \in [0, k_0]$ die Bedingung*

$$\|Hu_k - Hv_k\|_{L^1} \leq \|u_k - v_k\|_{L^1}$$

BEWEIS. Sei M die Menge aller stückweise konstanten Funktionen in $L^1(\mathbb{R})$. Dann gilt $\max(u, v) \in M$, falls $u, v \in M$ und wir können das Lemma von Crandall und Tartar anwenden: Nach Definition gilt

$$\|Hu\|_{L^1} = h \sum_{i \in Z} |(Hu)_i|$$

Gleichzeitig haben wir

$$|(Hu)_i| = |H(u_{i-m}, \dots, u_{i+m})| \leq \max_{l \in \{-m, \dots, m\}} |u_{i+l}| \leq \sum_{l=-m}^m |u_{i+l}|$$

wobei wir verwendet haben, dass das Verfahren L^∞ -stabil ist. Damit folgt

$$\|Hu\|_{L^1} \leq h \sum_{i \in Z} \sum_{l=-m}^m |u_{i+l}| = \sum_{l=-m}^m h \sum_{i \in Z} |u_{i+l}| = (2m+1) \|u\|_{L^1}$$

und $H : M \rightarrow L^1(\mathbb{R})$.

Da wir angenommen haben, dass das Verfahren konservativ ist, gilt

$$(1.1.30) \quad \int_{\mathbb{R}} Hu dx = h \sum_{i \in Z} (Hu)_i = h \sum_{i \in Z} u_i = \int_{\mathbb{R}} u dx$$

Nun ist H monoton und erfüllt (1.1.30). Nach dem Lemma von Crandall und Tartar folgt daraus, dass H als Abbildung in L^1 kontrahierend ist. $\square$

LEMMA 1.53. *Ist ein Verfahren definiert durch H kontrahierend in L^1 , so besitzt das Verfahren die TVD-Eigenschaft.*

BEWEIS. Man berechnet

$$\mathrm{TV}(Hu) = \sum_{i \in \mathbb{Z}} |(H(u))_{i+1} - (H(u))_i| = \frac{1}{h} \|Hv - Hu\|_{L^1}$$

mit $u = u_i, v = u_{i+1}$ auf $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$. Da das Verfahren L^1 -kontrahierend ist, folgt

$$\mathrm{TV}(Hu) \leq \frac{1}{h} \|v - u\|_{L^1} = \sum_{i \in \mathbb{Z}} |u_{i+1} - u_i| = \mathrm{TV}(u)$$

□

Es bleibt zu zeigen, dass monotone Verfahren auch L^∞ -stabil sind.

LEMMA 1.54. *Ist ein Verfahren definiert durch H monoton und konservativ, so ist das Verfahren auch L^∞ -stabil.*

BEWEIS. Wir zeigen, dass die Gittertransferfunktion H folgendes Maximumprinzip erfüllt:

$$(1.1.31) \quad \min_{l=-m, \dots, m} u_{i+l} \leq H(u_{i-m}, \dots, u_{i+m}) \leq \max_{l=-m, \dots, m} u_{i+l}$$

aus dem bereits die L^∞ -Stabilität des Verfahrens folgt.

Für eine gegebene Folge $(u_i)_{i \in \mathbb{Z}}$, definieren wir die konstante Folge mit $w_i = w$ für alle $i \in \mathbb{Z}$ und $w = \max_{i \in \mathbb{Z}} u_i$. Da das Verfahren konservativ ist, gilt

$$H(w_{i-m}, \dots, w_{i+m}) = w_i - \lambda \left(f_{i+\frac{1}{2}} - f_{i-\frac{1}{2}} \right) = w$$

Wegen der Monotonie folgt mit $u_i \leq w_i$

$$H(u_{i-m}, \dots, u_{i+m}) \leq H(w_{i-m}, \dots, w_{i+m}) = w$$

Mit einem ähnlichen Argument zeigt man den linken Teil der Abschätzung (1.1.31) und damit ist das Verfahren L^∞ -stabil. □

Monotone Verfahren scheinen also sehr attraktiv zu sein, aber solche Verfahren besitzen einen entscheidenden Nachteil:

SATZ 1.55. *Sei durch $H \in \mathcal{C}^3(\mathbb{R}^{2m+1})$ ein konservatives, konsistentes und monotones Verfahren definiert. Dann ist das Verfahren maximal ein Verfahren erster Ordnung.*

BEWEIS. Damit das Verfahren kein Verfahren zweiter Ordnung ist, müssen wir überprüfen, dass der lokale Abschneiderfehler,

$$(Lu)(x, t) = -k \partial_x (\beta(u, \lambda) \partial_x u) + O(k^2)$$

mit

$$\beta(u, \lambda) = \frac{1}{2} \left[\frac{1}{\lambda^2} \sum_{l=-m}^m l^2 \partial_l H(u, \dots, u) - a^2(u) \right], \quad a(u) = f'(u)$$

nicht verschwindet. Zunächst berechnet man

$$\partial_l H(u_{-m}, \dots, u_m) = \delta_{l0} - \lambda (\partial_{l-1} F(u_{-m+1}, \dots, u_m) - \partial_l F(u_{-m}, \dots, u_{m-1}))$$

und Summation über $l = -m, \dots, m$ ergibt

$$\begin{aligned}
\sum_{l=-m}^m l \partial_l H(u, \dots, u) &= -\lambda \sum_{l=-m}^m l (\partial_{l-1} F(u, \dots, u) - \partial_l F(u, \dots, u)) \\
&= -\lambda \sum_{l=-m}^m ((l+1) \partial_l F(u, \dots, u) - l \partial_l F(u, \dots, u)) \\
&= -\lambda \sum_{l=-m}^m \partial_l F(u, \dots, u)
\end{aligned}$$

Da das Verfahren konsistent ist, $f(u) = F(u, \dots, u)$, folgt

$$a(u) = f'(u) = \sum_{l=-m}^m \partial_l F(u, \dots, u)$$

Aus der Monotonie folgt $\partial_l H \geq 0$ und damit

$$-\lambda f'(u) = \sum_{l=-m}^m l \partial_l H(u, \dots, u) = \sum_{l=-m}^m l \sqrt{\partial_l H} \sqrt{\partial_l H}$$

Mit der Cauchy–Schwarzschen Ungleichung folgt

$$(\lambda f'(u))^2 \leq \left(\sum_{l=-m}^m l^2 \partial_l H \right) \left(\sum_{l=-m}^m \partial_l H \right)$$

Da H konservativ ist, haben wir $\sum_{l=-m}^m \partial_l H(u, \dots, u) = 1$ und

$$(\lambda f'(u))^2 \leq \sum_{l=-m}^m l^2 \partial_l H(u, \dots, u)$$

Daraus folgt, dass

$$(1.1.32) \quad \beta(u, \lambda) = \frac{1}{2\lambda^2} \left(\sum_{l=-m}^m l^2 \partial_l H(u, \dots, u) - (\lambda f'(u))^2 \right) \geq 0$$

Das Gleichheitszeichen in (1.1.32) gilt nur, falls in der Cauchy–Schwarzschen Ungleichung das Gleichheitszeichen gilt. Dies ist aber genau dann der Fall, wenn die beiden Vektoren $(-m \partial_{-m} H, \dots, m \partial_m H)$ und $(\partial_{-m} H, \dots, \partial_m H)$ in $\mathbb{R}^{2m+1}$ linear abhängig sind, d.h. es existiert ein $c \neq 0$, sodass $l \partial_l H = c \partial_l H$ für alle $l = -m, \dots, m$. Das kann aber nur erfüllt sein, falls maximal eine Ableitung ungleich null ist, d.h.

$$\partial_l H = 0, \quad \forall l \neq l_0$$

Wegen der Konservativität folgt dann aber $\sum_l \partial_l H = \partial_{l_0} H = 1$ und

$$-\lambda f'(u) = l_0 \partial_{l_0} H = l_0$$

Das bedeutet aber, dass $f'(u) = l_0/\lambda = a$ konstant ist und die Flussfunktion demnach eine lineare Funktion ist, $f = a \cdot u$. In diesem Fall kann man sogar zeigen, dass das Verfahren

die exakte Lösung liefert. Abgesehen von diesem trivialen Fall, ist das Verfahren allerdings maximal von erster Ordnung. $\square$

Eine Abschwächung des Begriffs *monotone Verfahren* sind die sogenannten *Monotonie-erhaltenden Verfahren*:

DEFINITION 1.56. *Ein Verfahren definiert durch H heißt Monotonie-erhaltend, falls für jede stückweise konstante Funktion $u_k \in L^\infty$, $k \in (0, k_0)$, die entweder monoton nicht-wachsend oder nichtfallend ist, dieselbe Aussage für $(Hu)_k$ gilt.*

Eine Verfahren, das die TVD-Eigenschaft besitzt, ist gleichzeitig Monotonie-erhaltend und wir haben die Folgerungen

$$\text{Monoton} \implies L^1\text{-kontrahierend} \implies \text{TVD} \implies \text{Monotonie-erhaltend}$$

Weiter gilt, dass ein lineares, Monotonie-erhaltendes Verfahren auch gleichzeitig monoton ist. Damit ist aber auch jedes lineare, konservative, konsistente TVD-Verfahren maximal erster Ordnung.

Eine ähnliche Aussage gilt auch für nichtlineare Verfahren: ein nichtlineares, konservatives, konsistentes 3-Punkt Verfahren, das zusätzlich die TVD-Eigenschaft besitzt, ist maximal erster Ordnung. Möchte man also TVD-Verfahren zweiter Ordnung konstruieren, so müssen diese mindestens nichtlinear und keine 3-Punkt Verfahren sein. Das Verfahren von Lax und Wendroff ist zwar ein Verfahren zweiter Ordnung, erfüllt aber nicht die TVD-Eigenschaft.

1.8. Konvergenz zur Entropielösung. In Abschnitt 1.6 hatten wir die Konvergenz eines Verfahrens für den Fall einer nichtlinearen, skalaren Erhaltungsgleichung untersucht: ist das Verfahren konservativ, konsistent, L^∞ -stabil und besitzt es die TVD-Eigenschaft, dann konvergieren die numerischen Approximationen für $k \rightarrow 0$ gegen eine schwache Lösung der Gleichung, allerdings nicht notwendigerweise gegen die eindeutige Entropielösung.

Um die Konvergenz gegen die Entropielösung zu gewährleisten, müssen wir noch das Konzept der *Entropie* vom kontinuierlichen auf den diskreten Fall übertragen: ist $u(x, t)$ die Entropielösung des Problems, so erfüllt u im schwachen Sinne die Ungleichung

$$(1.1.33) \quad \eta(u)_t + \sigma(u)_x \leq 0$$

für alle Entropie-Entropiefluss-Paare η, σ , wobei $\eta(u)$ eine konvexe Funktion ist und $\sigma' = \eta' f'$.

Diese Eigenschaft setzen wir wie folgt in ein numerisches Verfahren um:

DEFINITION 1.57. *Ein Verfahren definiert durch H mit $(Hu)_i = H(u_{i-m}, \dots, u_{i+m})$ nennt man e-konsistent für ein festes Entropie-Entropiefluss Paar (η, σ) , falls eine Funktion $\Sigma : \mathbb{R}^{2m} \rightarrow \mathbb{R}$ existiert, die mit σ konsistent ist, d.h.*

$$\Sigma(u, \dots, u) = \sigma(u), \quad \forall u \in \mathbb{R}$$

und zusätzlich die folgende diskrete Entropiebedingung erfüllt,

$$(1.1.34) \quad \eta(u_i^{j+1}) \leq \eta(u_i^j) - \lambda \left(\Sigma_{i+\frac{1}{2}} - \Sigma_{i-\frac{1}{2}} \right)$$

wobei $\Sigma_{i+\frac{1}{2}} = \Sigma(u_{i-m+1}^j, \dots, u_{i+m}^j)$.

Die Entropiebedingung (1.1.34) ist also eine diskrete Approximation der Entropieungleichung (1.1.33), denn es gilt

$$\frac{\eta(u_i^{j+1}) - \eta(u_i^j)}{k} + \frac{\Sigma_{i+\frac{1}{2}} - \Sigma_{i-\frac{1}{2}}}{h} \leq 0$$

SATZ 1.58. *Ist ein Verfahren konservativ, konsistent und monoton, so ist das Verfahren e-konsistent für alle Entropie-Entropiefluss Paare (η, σ) .*

Beispiele für e-konsistente Verfahren sind das Lax-Friedrichs und das Godunov-Verfahren. Den Beweis des Satzes führen wir zurück auf folgende Aussage

SATZ 1.59. *Ist ein Verfahren konservativ, konsistent, L^∞ -stabil, TVD und e-konsistent mit einer lokal Lipschitz-stetigen numerischen Flussfunktion F sowie $TV(u_0) < \infty$, so konvergiert die Folge $(u_k)_{k \in \mathbb{R}}$ für $k \rightarrow 0$ in $L^\infty((0, T), L_{loc}^1(\mathbb{R}))$ für alle $T > 0$ gegen die eindeutige Entropielösung des Problems.*

BEWEIS. Wir nehmen an, dass die Folge nicht gegen die eindeutige Entropielösung konvergiert. Dann existiert eine Teilfolge u_{k_l} und ein $\Omega \subset \mathbb{R}$, sodass

$$\|u_{k_l}(\cdot, t) - u(\cdot, t)\|_{L^1(\Omega)} \geq c > 0$$

wobei u die eindeutige Entropielösung bezeichnet. Aus dem Konvergenzsatz 1.30 wissen wir aber, da das Verfahren L^∞ -stabil und TVD ist, dass die Folge u_{k_l} eine Teilfolge besitzt, die in $L^\infty((0, T), L_{loc}^1(\mathbb{R}))$ konvergiert. Da das Verfahren e-konsistent ist, erfüllt diese Teilfolge im Grenzwert die Entropiebedingung und es muss daher, im Widerspruch zur Annahme, gelten

$$\|u_{k_{l_n}}(\cdot, t) - u(\cdot, t)\|_{L^1(\Omega)} \rightarrow 0$$

□

KOROLLAR 1.60. *Jedes monotone Verfahren konvergiert gegen die Entropielösung.*

Da die Überprüfung, ob ein Verfahren e-konsistent ist, schwierig ist, führt man den Begriff von sogenannten E-Verfahren ein.

DEFINITION 1.61. *Ein konsistentes und konservatives Verfahren heißt E-Verfahren, falls die numerische Flussfunktion für alle u zwischen u_i und u_{i+1} die Ungleichung*

$$\text{sign}(u_{i+1} - u_i)(f_{i+\frac{1}{2}} - f(u)) \leq 0$$

erfüllt.

Dann gilt

LEMMA 1.62. *Sei*

$$\lambda \max_{i \in \mathbb{Z}} \left| \frac{f(u_{i+1}) - f(u_i)}{u_{i+1} - u_i} \right| \leq 1$$

Ein Verfahren definiert durch H ist genau dann ein E-Verfahren, falls gilt

$$\begin{aligned} f_{i+\frac{1}{2}} &\leq f_{i+\frac{1}{2}}^G \quad \text{for } u_i < u_{i+1} \\ f_{i+\frac{1}{2}} &\geq f_{i+\frac{1}{2}}^G \quad \text{for } u_i > u_{i+1} \end{aligned}$$

wobei $f_{i+\frac{1}{2}}^G$ die numerische Flussfunktion des Godunov-Verfahrens bezeichnet, d.h.

$$f_{i+\frac{1}{2}}^G = F^G(u_i, u_{i+1})$$

Ein Verfahren definiert durch H ist ebenfalls genau dann ein E-Verfahren, falls gilt

$$Q_{i+\frac{1}{2}}^G \leq Q_{i+\frac{1}{2}} \quad \forall i \in Z$$

wobei $Q_{i+\frac{1}{2}}^G$ die numerische Viskosität des Godunov-Verfahrens bezeichnet, d.h.

$$Q_{i+\frac{1}{2}}^G = Q^G(u_i, u_{i+1}), \quad Q^G = \lambda \frac{f(u) + f(v) - 2F^G(u, v)}{v - u}$$

BEWEIS. Für den ersten Teil des Satzes wiederholen wir kurz die Definition der numerischen Flussfunktion im Godunov-Verfahren: es gilt

$$f_{i-\frac{1}{2}}^G = \begin{cases} \min_{u \in [u_{i-1}, u_i]} f(u) & : \text{ if } u_{i-1} < u_i \\ \max_{u \in [u_i, u_{i-1}]} f(u) & : \text{ if } u_{i-1} > u_i \end{cases}$$

Damit ist ein Verfahren unter der im Satz angegebenen Bedingung offensichtlich ein E-Verfahren. Für den zweiten Teil verwenden wir die Darstellung der numerischen Viskosität in der Form

$$Q_{i+\frac{1}{2}} = \lambda \frac{f(u_i) + f(u_{i+1}) - 2f_{i+\frac{1}{2}}}{u_{i+1} - u_i}$$

Damit folgt die zweite Aussage direkt aus dem ersten Teil. $\square$

Das Godunov-Verfahren ist dementsprechend ein Beispiel für ein E-Verfahren. Weiter kann man zeigen, dass monotone 3-Punkt Verfahren E-Verfahren sind. Wir untersuchen als nächstes das Verfahren von Roe, wobei wir annehmen, dass die Flussfunktion strikt konvex ist. Wir wissen, dass das Verfahren von Roe und das Godunov-Verfahren bis auf den Fall einer sonischen Verdünnungswelle identisch sind. In diesem Fall folgt aus $f'(u) < 0 < f'(v)$, da f konvex ist, $u < v$. Daher gilt

$$\begin{aligned} Q^G(u, v) &= \lambda \frac{f(u) + f(v) - 2f(u_s)}{v - u} \\ Q^{up}(u, v) &= \lambda \frac{f(u) + f(v) - 2f(\tilde{u})}{v - u} \end{aligned}$$

wobei $f'(u_s) = 0$ und $\tilde{u}$ abhängig von $a(u, v)$ entweder gleich u oder v ist.

Da $f'(u_s) = 0$ und f konvex ist, ist der Wert $f(u_s)$ das globale Minimum von f . Daraus folgt, dass für den Fall einer sonischen Verdünnungswelle die Beziehung $Q^{up}(u, v) < Q^G(u, v)$ gilt.

Hinsichtlich der Konvergenz von E-Verfahren gilt folgender Satz:

SATZ 1.63. *Das Verfahren definiert durch H sei ein E-Verfahren und erfülle die CFL-Bedingung mit Parameter $1/2$. Gilt $Q_{i+\frac{1}{2}} \leq \frac{1}{2}$, so ist das Verfahren L^∞ -stabil, TVD und e-konsistent für jedes Entropie-Entropiefluss Paar (η, σ) . Insbesondere konvergieren E-Verfahren und obigen Annahmen gegen die eindeutige Entropielösung.*

BEWEIS. Wir haben die Abschätzung

$$\frac{1}{2} \geq Q_{i+\frac{1}{2}} \geq Q_{i+\frac{1}{2}}^G \geq Q_{i+\frac{1}{2}}^{up} = \lambda \left| \frac{f(u_{i+1}) - f(u_i)}{u_{i+1} - u_i} \right|$$

also ist das Verfahren L^∞ -stabil und TVD. Der Beweis der e-konsistent ist etwas länglich und daher verzichten wir an dieser Stelle darauf. $\square$

Diese Konvergenzaussage bestätigt wiederum, dass das nichtlineare Upwindverfahren von Roe kein E-Verfahren ist, denn es konvergiert nicht notwendigerweise gegen die eindeutige Entropielösung. Dies liegt offensichtlich daran, dass die numerische Viskosität in einigen Fällen zu klein ist.

Man kann allerdings das Verfahren leicht abwandeln und damit die numerische Viskosität erhöhen, um ein E-Verfahren zu erhalten. Diese modifizierten Verfahren, die man als approximative Godunov-Verfahren bezeichnet, besitzen alle wichtigen Eigenschaften des exakten Godunov-Verfahrens und sind insbesondere bei Systemen von Gleichungen sinnvoll, für die die Berechnung der exakten Lösung von Riemannproblemen bereits sehr kompliziert wird.

Im folgenden geben wir ein einfaches Beispiel für ein approximatives Godunov-Verfahren. Das Upwind Verfahren war gegeben durch die Lösungen von lokalen Riemannproblemen der Form

$$w_t + a(u, v)w_x = 0, \quad a(u, v) = \frac{f(u) - f(v)}{u - v}$$

mit $u = u_i$ and $v = u_{i+1}$. Diese Gleichung läßt sich auch als eine linearisierte Gleichung interpretieren: setzen wir $\tilde{f}(w) = f(u) + (w - u)a(u, v)$ als Linearisierung der exakten Flussfunktion $f(u)$, so erhalten wir auch die lineare Gleichung

$$w_t + \tilde{f}(w)_x = 0$$

Dieser Ansatz ist aber offensichtlich zu einfach, um die wesentlichen Eigenschaften des Godunov-Verfahrens zu behalten. Eine bessere Approximation liefert zum Beispiel der Ansatz

$$(1.1.35) \quad \tilde{f}(w) = \begin{cases} f(u) + (w - u)f'(u) & : w \leq \bar{u} \\ f(v) + (w - v)f'(v) & : w > \bar{u} \end{cases}$$

wobei $\bar{u}$ die w -Koordinate des Schnittpunktes der beiden Tangenten an die Kurve $f(w)$ in den Punkten $w = u$ und $w = v$ ist. Wir lösen dann das Problem

$$w_t + \tilde{f}(w)_x = 0, \quad w_0(x) = \begin{cases} u & : x \leq 0 \\ v & : x > 0 \end{cases}$$

und erhalten die Lösung dieses lokalen Riemannproblems in der Form

$$w(x, t) = \begin{cases} u & : \frac{x}{t} \leq f'(u) \\ v & : \frac{x}{t} \geq f'(v) \\ \bar{u} & : \text{sonst} \end{cases}$$

mit

$$\bar{u} = u + \frac{f'(v) - \dot{s}(u, v)}{f'(v) - f'(u)}(u - v), \quad \dot{s} = a$$

Insbesondere bedeutet dies, dass der Verdünnungsfächer durch den konstanten Wert $\bar{u}$ ersetzt wird.

Dies definiert gerade das approximative Godunov–Verfahren nach Roe:

- a) Wir verwenden das klassische Upwind–Verfahren in allen Fällen, bis auf den Fall der sonischen Verdünnungswelle.
- b) In Fall der sonischen Verdünnungswelle ersetzen wir das gegebene Problem durch ein lineares Problem mit Flussfunktion (1.1.35). Damit wird die numerische Flussfunktion, die im Godunov–Verfahren durch $f(u_s)$ gegeben ist, durch den Wert $\tilde{f}(\bar{u})$ ersetzt. Da f als konvex angenommen wurde, gilt $\tilde{f}(\bar{u}) \leq f(u_s)$ und damit $Q \geq Q^G$.

Diese Verfahren ist ein E–Verfahren und sichert damit die Konvergenz gegen die Entropielösung.

1.9. TVD–Verfahren zweiter Ordnung. In diesem Abschnitt versuchen wir Verfahren herzuleiten, die die TVD–Eigenschaft besitzen und gleichzeitig Verfahren zweiter Ordnung sind. Als Verfahren zweiter Ordnung haben wir bis jetzt nur das Verfahren von Lax und Wendroff kennengelernt, das allerdings nicht die TVD–Eigenschaft besitzt. Insbesondere galt für die numerische Viskosität (bei linearen Gleichungen) die Abschätzung

$$Q^{LW} = (\lambda(a(u, v)))^2 < \lambda|a(u, v)| = Q^{up} < Q^G$$

Nach den Ausführungen am Ende von Abschnitt 1.7 wissen wir bereits, dass es für allgemeine nichtlineare Gleichungen unmöglich ist, ein lineares TVD–Verfahren zweiter Ordnung herzuleiten. Genauso unmöglich ist es, für nichtlineare Gleichungen ein 3–Punkt TVD–Verfahren zweiter Ordnung zu formulieren. Das bedeutet, dass wir im folgenden nichtlineare m –Punkt Verfahren mit $m > 3$ suchen.

Die einfachste Methode ein solches Verfahren zu konstruieren, ist der Ansatz eines modifizierten Flusses: Dazu betrachten wir ein 3–Punkt TVD–Verfahren in viskoser Form

$$(1.1.36) \quad u_i^{j+1} = u_i^j - \lambda \frac{f(u_{i+1}^j) - f(u_{i-1}^j)}{2} + \frac{Q_{i+\frac{1}{2}}(u_{i+1}^j - u_i^j) - Q_{i-\frac{1}{2}}(u_i^j - u_{i-1}^j)}{2}$$

und berechnen die numerische Viskosität, die allgemein in der Form

$$\beta(u; \lambda) = \frac{1}{2\lambda^2} \left[\sum_{l=-m}^m l^2 \partial_l H(u, \dots, u) - \lambda^2 a^2(u) \right]$$

gegeben ist. Es gilt

$$\begin{aligned}\partial_1 H(u, u, u) &= -\frac{\lambda}{2} f'(u) + \frac{Q(u, u)}{2} \\ \partial_{-1} H(u, u, u) &= \frac{\lambda}{2} f'(u) + \frac{Q(u, u)}{2}\end{aligned}$$

Daraus folgt

$$\beta(u, \lambda) = \frac{1}{2\lambda^2} [Q(u, u) - \lambda^2 a^2(u)]$$

Damit ist das Verfahren (1.1.36) angewendet auf die Gleichung $u_t + (f(u))_x = 0$ ein Verfahren zweiter Ordnung zur modifizierten Gleichung

$$u_t + [f(u) - \frac{h}{2\lambda} (Q(u, u) - \lambda^2 a^2(u)) u_x]_x = 0$$

Dies liefert bereits die Idee des modifizierten Flusses: wir wenden das 3-Punkt Verfahren auf die Gleichung

$$(1.1.37) \quad u_t + [f(u) \pm \frac{h}{2\lambda} (Q(u, u) - \lambda^2 a^2(u)) u_x]_x = 0$$

an und erwarten, dass das resultierende Verfahren ein Verfahren zweiter Ordnung zur Ausgangsgleichung ist.

Wir bezeichnen mit $\tilde{f}$ eine geeignete Approximation der exakten Flussfunktion in Gleichung (1.1.37) und setzen

$$(1.1.38) \quad \tilde{f}(u_i) = f(u_i) + \frac{g(u_{i-1}, u_i, u_{i+1})}{\lambda} = f(u_i) + \frac{g_i}{\lambda}$$

Die Funktion g in (1.1.38) muß natürlich so gewählt werden, dass g eine Approximation zweiter Ordnung des Terms

$$\frac{h}{2\lambda} (Q(u, u) - \lambda^2 a^2(u)) u_x$$

in (1.1.37) ist. Da die numerische Viskosität Q ebenfalls von f abhängt, verwenden wir auch eine modifizierte Viskosität $\tilde{Q}$, sodass wir als Ansatz folgendes 5-Punkt Verfahren wählen

$$(1.1.39) \quad u_i^{j+1} = u_i^j - \lambda \frac{\tilde{f}(u_{i+1}^j) - \tilde{f}(u_{i-1}^j)}{2} + \frac{\tilde{Q}_{i+\frac{1}{2}}^j (u_{i+1}^j - u_i^j) - \tilde{Q}_{i-\frac{1}{2}}^j (u_i^j - u_{i-1}^j)}{2}$$

Es bleibt zu zeigen, wie die beiden Terme g und $\tilde{Q}$ gewählt werden müssen, damit das Verfahren (1.1.39) ein TVD-Verfahren zweiter Ordnung zur Erhaltungsgleichung $u_t + f(u)_x = 0$ ist.

LEMMA 1.64. Sei $a_{i+1/2} = a(u_i, u_{i+1})$ und $g : \mathbb{R}^3 \rightarrow \mathbb{R}$ gegeben.

1) Wählt man

$$\tilde{Q}_{i+\frac{1}{2}} = \lambda^2 a_{i+\frac{1}{2}}^2 + \frac{g(u_{i-1}, u_i, u_{i+1}) + g(u_i, u_{i+1}, u_{i+2})}{u_{i+1} - u_i} + 0(|u_{i+1} - u_i|),$$

so ist das Verfahren bezüglich der Gleichung $u_t + f(u)_x = 0$ formal zweiter Ordnung.

2) Gilt

$$|\lambda a_{i+\frac{1}{2}} + \frac{g(u_i, u_{i+1}, u_{i+2}) - g(u_{i-1}, u_i, u_{i+1})}{u_{i+1} - u_i}| \leq \tilde{Q}_{i+\frac{1}{2}} \leq 1,$$

so besitzt das Verfahren die TVD-Eigenschaft.

BEWEIS. Für den ersten Teil schreiben wir das Verfahren mit Hilfe von (1.1.39) in der viskosen Form

$$\begin{aligned} u_i^{j+1} &= u_i^j - \lambda \frac{f(u_{i+1}^j) - f(u_{i-1}^j)}{2} \\ &+ \frac{(\tilde{Q}_{i+\frac{1}{2}}^j - \frac{g_i^j + g_{i+1}^j}{u_{i+1} - u_i})(u_{i+1}^j - u_i^j) - (\tilde{Q}_{i-\frac{1}{2}}^j - \frac{g_i^j + g_{i-1}^j}{u_i - u_{i-1}})(u_i^j - u_{i-1}^j)}{2} \end{aligned}$$

Damit ist die numerische Viskosität des Verfahrens gegeben durch

$$\tilde{Q}_{i+\frac{1}{2}} - \frac{g_i^j + g_{i+1}^j}{u_{i+1} - u_i} = \lambda^2 a_{i+\frac{1}{2}}^2 + 0(|u_{i+1} - u_i|)$$

und bis zur Ordnung $0(|u_{i+1} - u_i|)$ gleich der Viskosität des Verfahren von Lax und Wendroff. Also ist das Verfahren formal zweiter Ordnung.

Für die zweite Aussage wendet man das Resultat aus Satz 1.44 über die TVD-Eigenschaft von Verfahren in viskoser Form auf das Verfahren in (1.1.39) an. Wegen $a_{i+1/2} = (f(u_{i+1}) - f(u_i))/(u_{i+1} - u_i)$ gilt demnach

$$|\lambda \frac{\tilde{f}(u_{i+1}) - \tilde{f}(u_i)}{u_{i+1} - u_i}| \leq \tilde{Q}_{i+\frac{1}{2}} \leq 1$$

mit $\tilde{f}(u_i) = f(u_i) + g_i/\lambda$. □

Die Funktion g kann etwa folgendermassen gewählt werden.

LEMMA 1.65. Sei H ein 3-Punkt Verfahren in viskoser Form mit Lipschitz-stetiger numerischer Viskosität Q und

$$\lambda |\frac{f(u_{i+1}) - f(u_i)}{u_{i+1} - u_i}| \leq Q_{i+\frac{1}{2}} \leq 1$$

d.h. das Verfahren besitzt die TVD-Eigenschaft. Weiter setzt man

$$g(u_{i-1}, u_i, u_{i+1}) = \frac{s_i}{2} \min \left\{ (Q_{i+\frac{1}{2}} - \lambda^2 a_{i+\frac{1}{2}}^2) |u_{i+1} - u_i|, (Q_{i-\frac{1}{2}} - \lambda^2 a_{i-\frac{1}{2}}^2) |u_i - u_{i-1}| \right\}$$

mit

$$s_i = \begin{cases} \text{sign}(u_{i+1} - u_i) & \text{falls } (u_{i+1} - u_i)(u_i - u_{i-1}) > 0 \\ 0 & \text{sonst} \end{cases}$$

Erfüllt der Term $\tilde{Q}$ in (1.1.39) die Bedingung 2) aus Lemma 1.64 und gilt

$$\tilde{Q}_{i+\frac{1}{2}} = Q_{i+\frac{1}{2}} + 0(|u_{i+1} - u_i|)$$

so ist das Verfahren gegeben durch (1.1.39) ein TVD-Verfahren zweiter Ordnung zur Gleichung $u_t + f(u)_x = 0$.

Wir geben hier nur die Beweisidee an: Aus

$$\min(a, b) = a + \frac{b - a - |b - a|}{2}$$

leitet man die Beziehung

$$g_i = \frac{1}{2}(Q_{i+\frac{1}{2}} - \lambda^2 a_{i+\frac{1}{2}}^2)(u_{i+1} - u_i) + 0(|u_{i+1} - u_i|^2)$$

ab. Damit ist $\tilde{f}(u_i) = f(u_i) + g_i/\lambda$ eine geeignete Approximation zweiter Ordnung zur Gleichung (1.1.37). Gleichzeitig gilt auch

$$g_{i+1} = \frac{1}{2}(Q_{i+\frac{1}{2}} - \lambda^2 a_{i+\frac{1}{2}}^2)(u_{i+1} - u_i) + 0(|u_{i+1} - u_i|^2)$$

und daher

$$\tilde{Q}_{i+\frac{1}{2}} = Q_{i+\frac{1}{2}} + 0(|u_{i+1} - u_i|) = \frac{g_i + g_{i+1}}{u_{i+1} - u_i} + \lambda^2 a_{i+\frac{1}{2}} + 0(|u_{i+1} - u_i|)$$

und nach Teil 1) aus Lemma 1.64 ist das Verfahren damit zweiter Ordnung zur Gleichung $u_t + f(u)_x = 0$.

Ein Beispiel zur geeigneten Wahl der numerischen Viskosität $\tilde{Q}$ ist die Methode von Harten. Dazu nehmen wir an, dass die numerische Viskosität Q des vorgegeben 3-Punkt Verfahrens durch

$$Q_{i+\frac{1}{2}} = q(\lambda a_{i+\frac{1}{2}})$$

gegeben ist, wobei q eine Lipschitz-stetige Funktion mit

$$|x| \leq q(x) \leq 1 \quad \forall 0 \leq |x| \leq \mu \leq 1$$

ist. Beispiele wären das Verfahren von Lax und Friedrichs mit $q = 1$ oder das Upwind-Verfahren mit $q(x) = |x|$ und $\mu = 1$.

Dann definiert man

$$\tilde{Q}_{i+\frac{1}{2}} = q(\lambda a_{i+\frac{1}{2}} + \frac{g(u_i, u_{i+1}, u_{i+2}) - g(u_{i-1}, u_i, u_{i+1})}{u_{i+1} - u_i})$$

und g wie in Lemma 1.65. Dann sind unter der CFL-Bedingung gegeben durch

$$|\lambda a_{i+\frac{1}{2}}| \leq 1 - \sqrt{2(1 - \mu)}$$

die Bedingungen von Lemma 1.65 erfüllt, d.h. das Verfahren ist ein Verfahren zweiter Ordnung und besitzt die TVD-Eigenschaft.

1.10. Flusslimitierer. Ein anderer Ansatz Verfahren höherer Ordnung zu konstruieren, die gleichzeitig in der Lage sind, Unstetigkeitsstellen aufzulösen, ist der sogenannte Flusslimitierer. Dies sind Verfahren, die in Bereichen, in denen die Lösung glatt ist, möglichst hoher Ordnung sind, allerdings bei Unstetigkeitsstellen als Verfahren erster Ordnung die TVD-Eigenschaft besitzen. Dazu kombinieren wir ein

- ein Verfahren erster Ordnung mit numerischer Flussfunktion F^L

mit

- einem Verfahren zweiter Ordnung mit numerischer Flussfunktion F^H

Die Kombination beider ergebe eine numerische Flussfunktion in der Form

$$F = F^L + \Phi(F^H - F^L)$$

wobei Φ eine geeignete Schalterfunktion sei, d.h. es soll gelten $\Phi \approx 1$ in Bereichen, in denen die Lösung glatt ist, und $\Phi \approx 0$ in der Umgebung von Unstetigkeitsstellen.

Wir entscheiden, ob der diskrete Wert u_i^j in einem glatten oder nicht-glatten Bereich der Lösung $u(x, t)$ liegt, in dem wir die diskreten Ableitungen nach links und rechts vergleichen. Dazu setzen wir

$$\Theta_i^j = \frac{\frac{u_i^j - u_{i-1}^j}{h}}{\frac{u_{i+1}^j - u_i^j}{h}} = \frac{u_i^j - u_{i-1}^j}{u_{i+1}^j - u_i^j}$$

und definieren die Schalterfunktion Φ als Funktion von Θ .

Das Verhalten der Lösung lässt sich mit Hilfe der diskreten Werte Θ_i^j etwa folgendermassen klassifizieren:

$$\begin{aligned} \Theta_i^j \approx 1 & \Rightarrow \text{glatte Lösung} \\ \Theta_i^j = \epsilon > 0 & \Rightarrow \text{Sprung} \\ \Theta_i^j = \infty & \Rightarrow \text{Sprung} \\ \Theta_i^j < 0 & \Rightarrow \text{lokaler Extremwert} \end{aligned}$$

Daher sollte die Schalterfunktion die folgenden Eigenschaften besitzen:

$$\begin{aligned} \Phi(1) &= 1 \\ \Phi(0) &= \Phi(x \rightarrow \infty) = 0 \\ \Phi(x < 0) &= 0 \end{aligned}$$

SATZ 1.66. *Es seien F^L und F^H die numerischen Flussfunktionen von zwei konservativen und konsistenten Verfahren erster und zweiter Ordnung. Wir kombinieren beide Verfahren zum einem Verfahren H mit numerischer Flussfunktion F in der Form*

$$F = F^L + \Phi(\Theta)(F^H - F^L) = (1 - \Phi)F^L + \Phi F^H$$

Das Verfahren H ist konsistent, falls Φ für alle $\Theta \in \mathbb{R}$ beschränkt ist. Gilt $\Phi(1) = 1$ und ist Φ Lipschitz-stetig bei $\Theta = 1$, so ist das Verfahren (für glatte Lösungen mit u_x beschränkt weg von der Null) ein Verfahren zweiter Ordnung.

BEWEIS. Wir berechnen den lokalen Abschneidefehler und verwenden

$$\begin{aligned} |1 - \Phi| &\leq L|1 - \Theta| \approx L \left| 1 - \frac{u(x) - u(x-h)}{u(x+h) - u(x)} \right| \\ &= L \left| \frac{u(x+h) - 2u(x) + u(x-h)}{u(x+h) - u(x)} \right| = L \left| \frac{h^2 u''(x) + O(h^4)}{hu'(x) + O(h^2)} \right| = O(h) \end{aligned}$$

Daher lässt sich der lokale Abschneidefehler des Verfahrens H in der Form

$$(Lu)(x, t) = (1 + O(h))(Lu)^H(x, t) + O(h)(Lu)^L(x, t)$$

schreiben. Da F^H ein Verfahren zweiter Ordnung und F^L ein Verfahren zweiter Ordnung ist, folgt $(Lu)(x, t) = O(h^2)$. $\square$

Wir untersuchen nun, wie die Schalterfunktion Φ gewählt werden muss, damit das Verfahren die TVD-Eigenschaft besitzt. Dazu betrachten wir zunächst den Fall einer linearen Advektionsgleichung

$$u_t + au_x = 0, \quad a \geq 0$$

und wählen als Verfahren erster Ordnung das lineare Upwind-Verfahren,

$$u_i^{j+1} = u_i^j - \lambda a(u_i^j - u_{i-1}^j)$$

als Verfahren zweiter Ordnung das (lineare) Verfahren von Lax und Wendroff,

$$u_i^{j+1} = u_i^j - \lambda a \frac{u_{i+1}^j - u_{i-1}^j}{2} + \lambda^2 a^2 \frac{u_{i+1}^j - 2u_i^j + u_{i-1}^j}{2}$$

Damit ergibt sich

$$F^H - F^L = \frac{1}{2}a(1 - \lambda a)(v - u)$$

und

$$F^L + \Phi(\Theta)(F^H - F^L) = au_i + \underbrace{\Phi\left(\frac{u_i - u_{i-1}}{u_{i+1} - u_i}\right)}_{\Theta_i} \frac{1}{2}a(1 - \lambda a)(u_{i+1} - u_i)$$

Das kombinierte Verfahren H ist also gegeben durch

$$(1.1.40) \quad \begin{aligned} u_i^{j+1} &= u_i^j - \lambda a(u_i^j - u_{i-1}^j) \\ &\quad - \frac{1}{2}\lambda a(1 - \lambda a)(\Phi(\Theta_i^j)(u_{i+1}^j - u_i^j) - \Phi(\Theta_{i-1}^j)(u_i^j - u_{i-1}^j)) \end{aligned}$$

und offensichtlich ein 4-Punkt Verfahren, da Θ_i^j von u_k^j , $k = i - 1, i$ und $i + 1$ und Θ_{i-1}^j von u_k^j , $k = i - 2, i - 1$ und i abhängt.

Mit Hilfe der inkrementellen Form aus Abschnitt 1.6 und den Bedingungen aus Satz 1.36 zur L^∞ -Stabilität und TVD-Eigenschaft lässt sich direkt folgender Satz ableiten.

SATZ 1.67. *Gilt $\Phi(\Theta) = 0$ für $\Theta < 0$ und $0 \leq \Phi(\Theta) \leq \min(2, 2\Theta)$ für $\Theta \geq 0$, so ist das Verfahren (1.1.40) für $0 \leq \lambda a \leq 1$ L^∞ -stabil und besitzt die TVD-Eigenschaft.*

Die TVD-Region für die Schalterfunktion ist damit wie in Bild 1.5 dargestellt.

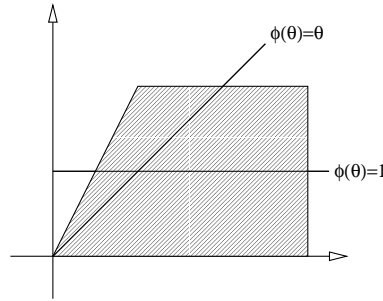


Bild 1.5 Bereich, in dem die Schalterfunktion Φ ein TVD-Verfahren liefert.

Im Bild sind ebenfalls die Schalterfunktionen des Verfahrens Lax und Wendroff mit $\Phi(\Theta) = 1$ und des Beam–Warming Verfahrens mit $\Phi(\Theta) = \Theta$ dargestellt, die beide nicht die TVD–Eigenschaft besitzen. Sweby hat 1984 gezeigt, dass es sinnvoll ist, eine Konvexkombination der beiden Schalterfunktionen $\Phi(\Theta) = 1$ und $\Phi(\Theta) = \Theta$ zu verwenden. Dieser sogenannte TVD–Bereich zweiter Ordnung nach Sweby ist in Bild 1.6 angegeben. Damit die Verfahren tatsächlich zweiter Ordnung sind, müssen die Kurven der Schalterfunktion durch den Punkt $(1, 1)$ gehen.

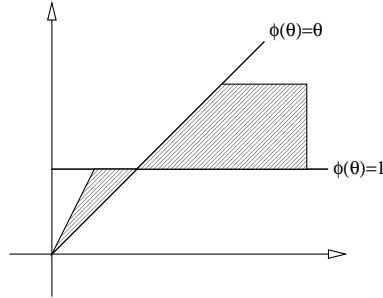


Bild 1.6 TVD–Bereich zweiter Ordnung nach Sweby.

Beispiele für geeignete Flusslimitierer sind

- der Superbee–Limitierer von Roe (1985)

$$\Phi(\Theta) = \max(0, \min(1, 2\Theta), \min(\Theta, 2)),$$

der gerade entlang der oberen Kanten des TVD–Bereiches in Bild 1.6 verläuft.

- der van Leer–Limitierer (1974)

$$\Phi_0(\Theta) = \frac{\Theta + |\Theta|}{1 + |\Theta|}$$

- sowie der Chakravarthy–Osher Limitierer (1984)

$$\Phi_\alpha(\Theta) = \max(0, \min(\Theta, \alpha)) \quad \alpha \in [1, 2]$$

der für $\alpha = 1$ gerade entlang der unteren Kanten des TVD–Bereiches in Bild 1.6 verläuft (Minmod–Limitierer von Roe).

Für die lineare Advektionsgleichung mit konstanter Geschwindigkeit $a > 0$ verwendet man ein 4–Punkt Verfahren. Betrachtet man Probleme, bei denen die lokale Ausbreitungsgeschwindigkeit wechselt, etwa durch $a = a(x)$, sodass $a < 0$ und $a \geq 0$ betrachtet werden muss, so kombiniert man den angegebenen Ansatz und erhält ein 5–Punkt Verfahren.

Im nichtlinearen Fall startet man mit einem 3–Punkt E–Verfahren als Verfahren erster Ordnung und kombiniert dies mit dem nichtlinearen Lax–Wendroff Verfahren. Dann läßt sich folgender Satz zeigen.

SATZ 1.68. Gilt $\Phi(\Theta) = 0$ für $\Theta < 0$ und $0 \leq \Phi(\Theta) \leq \min(\alpha, \alpha\Theta)$ für $\Theta \geq 0$ und $\alpha \in [0, 2]$, so besitzt das nach obigem Schema kombinierte Verfahren die TVD–Eigenschaft, falls die Viskosität des verwendeten E–Verfahrens die Bedingung

$$Q^E(u_i^j, u_{i+1}^j) \leq \frac{1}{1 + \frac{\alpha}{2}}$$

erfüllt.

BEMERKUNG 1.69.

- 1) Unter den Bedingungen des obigen Satzes läßt sich ebenfalls beweisen, dass das resultierende Verfahren L^∞ -stabil ist.
- 2) Das Verfahren konvergiert gegen eine schwache Lösung, aber nicht notwendigerweise gegen die eindeutige Entropielösung.
- 3) Mit Hilfe zusätzlicher Bedingungen kann man zeigen, dass das Verfahren e -konsistent ist und damit gegen die eindeutige Entropielösung konvergiert (siehe Artikel von Chakravarthy und Osher, *SIAM Journal of Numerical Analysis*, 1984, Seite 955).
- 4) Mehr Details findet man im Textbuch von Godlewski und Raviart, *Hyperbolic systems of conservation laws, Ellipses*, 1991.

1.11. Godunov–Verfahren zweiter Ordnung, MUSCL–Ansatz. Wir hatten gesehen, dass das Godunov–Verfahren als monotones Verfahren nur ein Verfahren erster Ordnung ist. Da man aber die exakte Lösung lokaler Riemannprobleme verwendet, muss die erste Ordnung eine Konsequenz aus der Projektion auf stückweise konstante Funktionen sein.

Eine Idee für ein Verfahren zweiter Ordnung, die auf van Leer zurückgeht, ist, statt der Projektion auf stückweise konstante eine Projektion auf stückweise lineare Funktionen zu verwenden. Demnach muss auch die Lösung der lokalen Riemannprobleme neu untersucht werden. Dieser Ansatz wird in der Literatur gewöhnlich als MUSCL–Verfahren bezeichnet (Monotone Upstream Schemes for Conservation Laws) und auch mit Runge–Kutta Verfahren hoher Ordnung bei der Zeitintegration verknüpft.

Wir ersetzen also die stückweise konstanten Funktionen im Godunov–Verfahren durch stückweise lineare Funktionen der Form

$$u_k(x, t_j) = u_i^j + \frac{x - x_i}{h} \sigma_i^j, \quad x \in [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$$

Dabei bezeichnet der Wert u_i^j wiederum den Zellmittelwert von $u_k(x, t_j)$. Integriert man die skalare Erhaltungsgleichung, so erhalten wir für die zeitliche Änderung des Zellmittelwertes wie zuvor die Gleichung

$$u_i^{j+1} = u_i^j - \lambda \left[\frac{1}{k} \int_{t_j}^{t_{j+1}} f(u(x_{i+\frac{1}{2}}, t)) dt - \frac{1}{k} \int_{t_j}^{t_{j+1}} f(u(x_{i-\frac{1}{2}}, t)) dt \right]$$

Die Neigungen σ_i^j sollen dabei so gewählt werden, dass das resultierende Verfahren die TVD–Eigenschaft besitzt.

Die Lösung der verallgemeinerten Riemannprobleme mit stückweise linearen Funktionen läßt sich nicht in exakter Form angeben. Wir werden daher später ein Verfahren zur approximativen Lösung angeben. Zunächst überlegen wir uns, wie die Koeffizienten σ_i^j gewählt werden müssen, um die TVD–Eigenschaft zu gewährleisten.

Mit Hilfe von Taylor-Entwicklungen kann man zeigen, dass man mit der Wahl

$$\frac{\sigma_i^j}{h} = u_x(x_i, t_j) + O(h)$$

ein Verfahren zweiter Ordnung erhält. Ein möglicher Ansatz für die Koeffizienten σ_i^j wäre daher

$$\sigma_i^j = \frac{1}{2}(u_{i+1}^j - u_{i-1}^j)$$

Allerdings besitzt das resultierende Verfahren nicht die TVD-Eigenschaft.

Daher muss man die σ_i^j auf geeignete Weise beschränken, zum Beispiel in der Form

$$\sigma_i^j = \begin{cases} \text{sign}(u_{i+1}^j - u_i^j) M & \text{falls (1.1.41) gilt} \\ 0 & \text{sonst} \end{cases}$$

$$M := \min \left\{ |u_{i+1}^j - u_i^j|, \frac{1}{2}|u_{i+1}^j - u_{i-1}^j|, |u_i^j - u_{i-1}^j| \right\}$$

und

$$(1.1.41) \quad \text{sign}(u_{i+1}^j - u_i^j) = \text{sign}(u_{i+1}^j - u_{i-1}^j) = \text{sign}(u_i^j - u_{i-1}^j)$$

Dies läßt sich auch in der Form

$$\sigma_i^j = \min \text{ mod } (u_{i+1}^j - u_i^j, u_i^j - u_{i-1}^j)$$

schreiben, wobei

$$\min \text{ mod } (a, b) = \begin{cases} a & \text{falls } |a| < |b| \text{ und } ab \geq 0 \\ b & \text{falls } |b| > |a| \text{ und } ab \geq 0 \\ 0 & \text{sonst} \end{cases}$$

Bei lokalen Extremwerten gilt offensichtlich $\sigma_i^j = 0$ und das Verfahren ist an dieser Stelle erster Ordnung.

Die neue Methode lautet daher:

- 1) berechne die Entropielösung der verallgemeinerten Riemannprobleme auf der Basis stückweise linearer Funktionen
- 2) berechne die neuen Zellmittelwerte
- 3) berechne die Neigungen aus den Zellmittelwerten mit Hilfe des minimalen Limitierers

SATZ 1.70.

Das Godunov-Verfahren auf der Basis stückweise linearer Rekonstruktionen ist L^∞ -stabil und besitzt die TVD-Eigenschaft.

BEMERKUNG 1.71. *Für den Fall einer linearen Flussfunktion $f = au$, kann man zeigen, dass dieses Verfahren genau einem Verfahren mit Flusslimitierer nach Chakravarty und Osher und $\alpha = 1$ entspricht, d.h. dem Flusslimitierer, der entlang der unteren Kante der TVD-Bereichs zweiter Ordnung in Bild 1.6 verläuft.*

Eine Möglichkeit, eine approximative Lösung der verallgemeinerten Riemannprobleme mit stückweise linearen Funktionen zu verwenden, ist das **MUSCL–Hancock Verfahren**. Hier verwendet man wieder die Lösung von Riemannproblemen mit stückweise konstanten Funktionen, wobei die linken und rechten Werte u_i^{--} und u_i^{++} in einer lokalen Riemannzelle durch

$$\begin{aligned} u_i^{--} &= u_i^+ - \lambda \frac{f(u_i^+) - f(u_i^-)}{2} \\ u_{i+1}^{++} &= u_{i+1}^- + \lambda \frac{f(u_{i+1}^-) - f(u_{i+1}^+)}{2} \end{aligned}$$

gegeben sind. Danach setzt man, wie beim gewöhnlichen Godunov–Verfahren, die numerische Flussfunktion als

$$f_{i+1/2} = f(u^*(u_i^{--}, u_{i+1}^{++}))$$

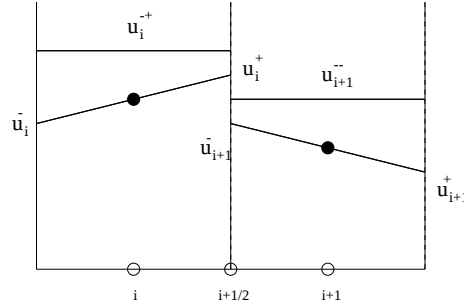


Bild 1.7 Diskrete Werte im MUSCL–Hancock Verfahren.

Das MUSCL–Hancock Verfahren ist dann unter einer geeigneten CFL–Bedingung ein TVD–Verfahren zweiter Ordnung. Wählt man einen restriktiveren Limitierer, so lässt sich auch ein e–konsistentes Verfahren konstruieren. Zum Abschluss präsentieren wir noch ein Verfahren, das sehr ähnlich zu den verallgemeinerten Godunov–Verfahren zweiter Ordnung ist, das allerdings ohne die Berechnung von Riemannproblemen auskommt und daher sehr effizient zu implementieren ist: das **Verfahren von Nessyahu und Tadmor**, siehe gleichnamige Autoren, J. Comp. Physics 87, Seite 408 (1990).

Die Idee bei diesem Verfahren ist, den Integrationsbereich im gewöhnlichen Godunov–Verfahren zur Berechnung der Evolution der Zellmittelwerte durch den Bereich $[x_i, x_{i+1}) \times [t_j, t_{j+1})$ zu ersetzen. Damit müssen Riemannprobleme auf den Zellen $[x_i, x_{i+1})$ berechnet werden, deren Lösung aber gerade durch $u(x_i, t) = u_i^j$ für $t \in [t_j, t_{j+1}]$ gegeben sind. Da die alten Zellmittelwerte auf $[x_i, x_{i+1})$ aber gerade durch den Mittelwert $(u_i + u_{i+1})/2$ gegeben sind, ergibt sich das Verfahren

$$(1.1.42) \quad u_{i+1/2}^{j+1} = \frac{u_i + u_{i+1}}{2} - \lambda(f(u_{i+1}^j) - f(u_i^j))$$

Dies ist gerade das Verfahren von Lax und Friedrichs, allerdings auf einem versetzten Gitter (staggered grid). Die neuen Werte u_i^{j+1} und u_{i+1}^{j+1} werden anschliessend wieder durch Zellmittelung berechnet. Insbesondere benötigt man beim Verfahren (1.1.42) keinerlei Informationen über die Struktur der lokalen Riemannprobleme.

Eine Erweiterung auf Verfahren zweiter Ordnung wird nun folgendermassen erreicht: man

verwendet stückweise linearen Approximationen und verwendet lokale Riemannprobleme auf dem versetzten Gitter. Ist u die Lösung der verallgemeinerten Riemannprobleme und

$$u_k(x, t_j) = u_i^j + \frac{x - x_i}{h} \sigma_i^j, \quad x \in [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$$

eine stückweise lineare Rekonstruktion, so erhält man durch Integration über $[x_i, x_{i+1}]$

$$u_{i+\frac{1}{2}}^{j+1} = \frac{1}{h} \int_{x_i}^{x_{i+1}} u_k(x, t_j) dx - \lambda \left[\frac{1}{h} \int_{t_j}^{t_{j+1}} f(u(x_{i+1}, t)) - f(u(x_i, t)) dt \right]$$

Für das erste Integral ergibt sich

$$\frac{1}{h} \int_{x_i}^{x_{i+1}} u_k(x, t_j) dx = \frac{u_{i+1}^j + u_i^j}{2} - \frac{\sigma_{i+1}^j - \sigma_i^j}{8}$$

Für die Integrale der Form $\frac{1}{k} \int_{t_j}^{t_{j+1}} f(u(x_i, t)) dt$ ergibt sich das Problem, dass die Funktion u nicht mehr konstant ist, sondern stückweise linear. Man muss daher geeignete Approximationen berechnen. Verwendet man etwa die Mittelpunktsregel, so ergibt sich

$$\frac{1}{k} \int_{t_j}^{t_{j+1}} f(u(x_i, t)) dt = f(\overbrace{u(x_i, t_{j+\frac{1}{2}})}^{u_i^{j+\frac{1}{2}}}) + O(k^2)$$

und, da u als Lösung der Riemannprobleme glatt ist,

$$u_i^{j+\frac{1}{2}} = u(x_i, t_{j+\frac{1}{2}}) = u_i^j + \frac{k}{2} u_t(x_i, t_j) + O(k^2)$$

Weiter löst u die Erhaltungsgleichung, also gilt $u_t(x_i, t_j) = -(f(u))_x(x_i, t_j)$ Schliesslich folgt aus $\frac{1}{h} f_i'^j = (f(u))_x(x_i, t_j) + O(k)$ die Gleichung

$$u_i^{j+\frac{1}{2}} = u_i^j - \frac{\lambda}{2} f_i'^j + O(k^2) + k \cdot O(k)$$

Fassen wir das Ergebnis zusammen, erhalten wir als Nessyahu–Tadmor Verfahren auf der Basis stückweise linearer Rekonstruktionen in der Form

$$(1.1.43) \quad \begin{cases} u_i^{j+\frac{1}{2}} &= u_i^j - \frac{\lambda}{2} f_i'^j \\ u_{i+\frac{1}{2}}^{j+1} &= \frac{u_{i+1}^j + u_i^j}{2} - \frac{\sigma_{i+1}^j - \sigma_i^j}{8} - \lambda(f(u_{i+\frac{1}{2}}^{j+\frac{1}{2}}) - f(u_i^{j+\frac{1}{2}})) \end{cases}$$

Eine mögliche Wahl für die Koeffizienten σ_i^j und $f_i'^j$ ist dabei

$$\begin{aligned} \sigma_i^j &= \min \text{ mod } (u_{i+1}^j - u_i^j, u_i^j - u_{i-1}^j) \\ f_i'^j &= \min \text{ mod } (f(u_{i+1}^j) - f(u_i^j), f(u_i^j) - f(u_{i-1}^j)) \end{aligned}$$

Das Verfahren (1.1.43) ist konservativ, konsistent und ein Verfahren zweiter Ordnung. Weiter ist das Verfahren L^∞ -stabil und besitzt die TVD-Eigenschaft unter der CFL-Bedingung mit Parameter $(\sqrt{7} - 2)/2 \sim 0.32$. Unter weiteren Restriktionen kann man noch zeigen, dass das Verfahren stets gegen die eindeutige Entropielösung konvergiert (siehe dazu den oben angegebenen Artikel).

In der Literatur findet man eine Reihe von ähnlichen Ansätzen, die auf Verfahren zweiter Ordnung führen, siehe etwa

- Goodman und LeVeque, SIAM J. Numerical Analysis 25, Seite 268 (1988),
- die stückweise parabolische Methode nach Colella und Woodward, J. Comp. Physics 54, Seite 174 (1984)
- oder sogenannte ENO-Verfahren (Essentially Non Oscillating), die eine gleichmässig hohe Ordnung besitzen und die TVD-Eigenschaft nur wenig verletzen (Harten et al., J. Comp. Physics 71, Seite 231 (1987)).