

Erhaltungsgleichungen: Theorie und Numerik

Roland Pulch

Skript zur Vorlesung im Wintersemester 2024/25

Universität Greifswald

Institut für Mathematik und Informatik

20. Januar 2025

Inhaltsverzeichnis

1	Definitionen und Beispiele	4
1.1	Einführungsbeispiel	4
1.2	Formulierung von Erhaltungsgleichungen	5
1.3	Beispiele	8
1.4	Hyperbolische Systeme	10
2	Analytische Eigenschaften	12
2.1	Charakteristiken	12
2.2	Lineare Advektionsgleichung	13
2.3	Burgers Gleichung	15
2.4	Lösungstheorie	18
2.5	Entropiebedingungen	25
2.6	Beispiel für nichtlineare Gleichung: Verkehrsfluss	28
2.7	Lineare Systeme	31
2.8	Nichtlineare Systeme	36
3	Numerische Methoden für lineare Probleme	42
3.1	Diskretisierung auf Gittern	42
3.2	Herleitung von Verfahren	43
3.3	Konvergenzuntersuchung	45
3.4	Einseitige Verfahren	54
3.5	Verhalten bei unstetigen Lösungen	56
4	Konservative Methoden für nichtlineare Probleme	61
4.1	Konservative Form	61

4.2	Herleitung konservativer Verfahren	65
4.3	Lax-Wendroff Theorem	68
4.4	Nichtlineare Stabilitätskonzepte	72
5	Hochauflösende Verfahren	77
5.1	Motivation	77
5.2	Fluss-begrenzende Methoden	78
5.3	Steigungs-begrenzende Methoden	84

1 Definitionen und Beispiele

Wir betrachten zeitabhängige Prozesse, welche nur in einer Raumdimension stattfinden. Verallgemeinerungen auf zwei oder drei Raumdimensionen sind analog möglich.

1.1 Einführungsbeispiel

Betrachtet wird ein (unendlich langes) Rohr mit einem monomolekularen Gas darin. Es bezeichnet t die Zeit und x die eindimensionale Raumkoordinate. Sei $x_1 \leq x_2$ und $0 \leq t_1 \leq t_2$. Die Massenerhaltung impliziert

$$\begin{aligned} \text{Masse in } [x_1, x_2] \text{ zur Zeit } t_2 &= \text{Masse in } [x_1, x_2] \text{ zur Zeit } t_1 \\ &\quad + \text{Einfluss bei } x_1 \text{ während } [t_1, t_2] \\ &\quad - \text{Ausfluss bei } x_2 \text{ während } [t_1, t_2]. \end{aligned}$$

Die Dichte ist gegeben durch $u = \text{Masse/Volumen}$. In einer Raumdimension gilt $u = \text{Masse/Länge}$. Es gilt $u(x, t) \geq 0$. Damit erhalten wir

$$\text{Masse in } [x_1, x_2] \text{ zur Zeit } t = \int_{x_1}^{x_2} u(x, t) \, dx.$$

Es bezeichne $v(x, t) \in \mathbb{R}$ die Geschwindigkeit des Gases. Der Einfluss des Gases ist dann gegeben durch $u(x, t)v(x, t)$. Die Massenerhaltung führt auf die Bedingung

$$\begin{aligned} \int_{x_1}^{x_2} u(x, t_2) \, dx &= \int_{x_1}^{x_2} u(x, t_1) \, dx \\ &\quad + \int_{t_1}^{t_2} u(x_1, t)v(x_1, t) \, dt - \int_{t_1}^{t_2} u(x_2, t)v(x_2, t) \, dt \end{aligned} \tag{1.1}$$

für alle $x_1 \leq x_2$ und $0 \leq t_1 \leq t_2$. Wir nehmen nun an, dass u und v stetig differenzierbar sind. Es gilt mit dem Hauptsatz der Differential- und Integralrechnung

$$\begin{aligned} u(x, t_2) - u(x, t_1) &= \int_{t_1}^{t_2} \frac{\partial}{\partial t} u(x, t) \, dt \\ u(x_2, t)v(x_2, t) - u(x_1, t)v(x_1, t) &= \int_{x_1}^{x_2} \frac{\partial}{\partial x} u(x, t)v(x, t) \, dx. \end{aligned}$$

Mit (1.1) folgt

$$\int_{t_1}^{t_2} \int_{x_1}^{x_2} \frac{\partial}{\partial t} u(x, t) + \frac{\partial}{\partial x} u(x, t)v(x, t) \, dx \, dt = 0$$

für beliebige $x_1 \leq x_2$ und $0 \leq t_1 \leq t_2$. Da der Integrand stetig ist, folgt aus dem Hauptsatz der Variationsrechnung, dass der Integrand identisch null ist. Wir erhalten die Differentialgleichung

$$\frac{\partial}{\partial t} u(x, t) + \frac{\partial}{\partial x} u(x, t)v(x, t) = 0. \quad (1.2)$$

Ist v vorgegeben, dann stellt dies eine Differentialgleichung für die Unbekannte u dar. Da v im allgemeinen unbekannt ist, wird eine weitere Gleichung zur Bestimmung von v benötigt. Dazu verwenden wir die Erhaltung des Impulses. Die Impulsdichte ist uv . Wir nehmen an, dass es keine inneren oder äußeren Kräfte auf die Gasmoleküle gibt. Es gilt $F = ma$ mit Kraft F , Masse m und Beschleunigung a . Aus $F = 0$ folgt $a = 0$. Die Beschleunigung ist die totale Ableitung der Geschwindigkeit. Differentiation mit der Kettenregel liefert

$$0 = a = \frac{dv}{dt} = \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} \frac{\partial x}{\partial t} = \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} v.$$

Zusammen mit (1.2) leiten wir eine Differentialgleichung für die Impulsdichte her

$$\frac{\partial(uv)}{\partial t} = u \frac{\partial v}{\partial t} + \frac{\partial u}{\partial t} v = -u \frac{\partial v}{\partial x} v - \frac{\partial(uv)}{\partial x} v = -\frac{\partial((uv)v)}{\partial x} = -\frac{\partial(uv^2)}{\partial x}.$$

Die Unbekannten sind die Massendichte $u_1 := u$ und die Impulsdichte $u_2 := uv$. Es gilt $uv^2 = \frac{(uv)^2}{u}$ unter der Annahme $u > 0$. Wir erhalten das Differentialgleichungssystem

$$\begin{aligned} \frac{\partial u_1}{\partial t} + \frac{\partial}{\partial x} u_2 &= 0 \\ \frac{\partial u_2}{\partial t} + \frac{\partial}{\partial x} \frac{u_2^2}{u_1} &= 0. \end{aligned}$$

Diese Darstellung enthält die Flussfunktion $f(u_1, u_2) = (u_2, \frac{u_2^2}{u_1})^\top$.

1.2 Formulierung von Erhaltungsgleichungen

Jetzt befassen wir uns mit der allgemeinen Definition der Erhaltungsgleichungen.

Definition 1.1 Die Integralform eines Systems aus Erhaltungsgleichungen lautet

$$\int_{x_1}^{x_2} u(x, t_2) \, dx = \int_{x_1}^{x_2} u(x, t_1) \, dx + \int_{t_1}^{t_2} f(u(x_1, t)) \, dt - \int_{t_1}^{t_2} f(u(x_2, t)) \, dt \quad (1.3)$$

für alle $x_1 \leq x_2$ und $0 \leq t_1 \leq t_2$ mit einer Lösung $u : \mathbb{R} \times \mathbb{R}_0^+ \rightarrow \mathbb{R}^k$ und einer stetig differenzierbaren Flussfunktion $f : D \rightarrow \mathbb{R}^k$ ($D \subseteq \mathbb{R}^k$).

Die Integration der vektorwertigen Funktionen erfolgt hier komponentenweise. Es muss $u \in \mathcal{F}$ mit einem geeigneten Funktionenraum $\mathcal{F}$ gelten, um die Existenz der Integrale sicherzustellen. Beispielsweise können global stetige Funktionen vorausgesetzt werden. Jedoch gibt es auch unstetige Lösungen, die physikalisch relevant sind.

Definition 1.2 Die Differentialform eines Systems aus Erhaltungsgleichungen lautet

$$\frac{\partial}{\partial t} u(x, t) + \frac{\partial}{\partial x} f(u(x, t)) = 0 \quad (1.4)$$

mit einer stetig differenzierbaren Lösung $u : \mathbb{R} \times \mathbb{R}_0^+ \rightarrow \mathbb{R}^k$ und einer stetig differenzierbaren Flussfunktion $f : D \rightarrow \mathbb{R}^k$ ($D \subseteq \mathbb{R}^k$).

Wie zu erwarten erweist sich die Differentialform aus Definition 1.2 als stärker als die Integralform aus Definition 1.1.

Satz 1.3 Ist $u : \mathbb{R} \times \mathbb{R}_0^+ \rightarrow \mathbb{R}^k$ eine stetig differenzierbare Lösung der Differentialform (1.4), dann ist u auch Lösung der Integralform (1.3).

Beweis:

Wir erhalten mit dem Hauptsatz der Differential- und Integralrechnung

$$\begin{aligned} \int_{x_1}^{x_2} u(x, t_2) \, dx - \int_{x_1}^{x_2} u(x, t_1) \, dx &= \int_{x_1}^{x_2} u(x, t_2) - u(x, t_1) \, dx \\ &= \int_{x_1}^{x_2} \int_{t_1}^{t_2} \frac{\partial}{\partial t} u(x, t) \, dt \, dx \\ &= \int_{t_1}^{t_2} \int_{x_1}^{x_2} \frac{\partial}{\partial t} u(x, t) \, dx \, dt \\ &\stackrel{(1.4)}{=} - \int_{t_1}^{t_2} \int_{x_1}^{x_2} \frac{\partial}{\partial x} f(u(x, t)) \, dx \, dt \\ &= - \int_{t_1}^{t_2} f(u(x_2, t)) - f(u(x_1, t)) \, dt \\ &= \int_{t_1}^{t_2} f(u(x_1, t)) \, dt - \int_{t_1}^{t_2} f(u(x_2, t)) \, dt. \end{aligned}$$

Die Integrationsreihenfolge kann vertauscht werden, da der Integrand stetig ist. Also ist die Integralform erfüllt. $\square$

O.E.d.A. betrachten wir eine einzelne Differentialgleichung ($k = 1$) mit Lösung $u \in C^1(\mathbb{R} \times \mathbb{R}_0^+)$. Sei $\phi \in C_0^1(\mathbb{R} \times \mathbb{R})$, d.h. stetig differenzierbar und mit kompaktem Träger. Aus (1.4) folgt

$$\phi \frac{\partial u}{\partial t} + \phi \frac{\partial f(u)}{\partial x} = 0$$

für alle $x \in \mathbb{R}$ und alle $t \geq 0$. Also gilt auch

$$\int_0^\infty \int_{-\infty}^{+\infty} \left[\phi \frac{\partial u}{\partial t} + \phi \frac{\partial f(u)}{\partial x} \right] dx dt = 0.$$

Partielle Integration liefert

$$\int_0^\infty \int_{-\infty}^{+\infty} \left[\frac{\partial \phi}{\partial t} u + \frac{\partial \phi}{\partial x} f(u) \right] dx dt = - \int_{-\infty}^{+\infty} \phi(x, 0) u(x, 0) dx. \quad (1.5)$$

Diese Bedingung kann nun auch für nur stetiges oder sogar unstetiges u erfüllt sein. Daher wird (1.5) auch als schwache Formulierung bezeichnet. Es bezeichnet $L_{\text{loc}}^1(\mathbb{R} \times \mathbb{R}_0^+)$ die Menge der lokal Lebesgue-integrierbaren Funktionen. Falls $u \in L_{\text{loc}}^1(\mathbb{R} \times \mathbb{R}_0^+)$ gilt, dann ist die linke Seite in (1.5) wohldefiniert, jedoch nicht notwendigerweise die rechte Seite, da die Gerade $\{(x, t) : t = 0\}$ eine Nullmenge der Ebene ist.

Definition 1.4 Eine Funktion $u \in \mathcal{F}$ ist eine schwache Lösung der Differentialgleichung (1.4) mit $k = 1$, falls die Bedingung (1.5) für alle $\phi \in C_0^1(\mathbb{R} \times \mathbb{R})$ erfüllt ist.

Die Integralform (1.3) der Erhaltungsgleichung und die Bedingung (1.5) für eine schwache Lösung sind äquivalent.

Der Funktionenraum $\mathcal{F}$ ist noch zu definieren, so dass

- i) Die Integrale in sowohl der Integralform (1.3) als auch der schwachen Formulierung (1.5) existieren.
- ii) Physikalisch relevante Zustände sind enthalten.

Wir betrachten endlich viele Halbgeraden $\{(x, t) : x = at + b, t \geq 0\}$ mit $a, b \in \mathbb{R}$. In der (x, t) -Ebene sind damit vertikale Geraden (x konstant) gestattet, jedoch keine horizontalen Geraden (t konstant). Diese Halbgeraden partitionieren die Halbebene mit $t \geq 0$ in endlich viele Teile $X_1, \dots, X_m$. Die Funktionenmenge $\mathcal{F}$ wird festgelegt durch eine stückweise Stetigkeit. Wir betrachten eine vektorwertige Funktion $u = (u_1, \dots, u_k)^\top$. Die Funktionenmenge $\mathcal{F}$ bestehe aus den Funktionen u , zu denen eine Partition $X_1, \dots, X_m$ existiert mit den Bedingungen:

1. u ist stetig im Inneren von X_j und die Grenzwerte zum Rand hin existieren für $j \in \{1, \dots, m\}$.
2. Die Grenzwerte stimmen am Rand $t = 0$ mit $u(x, 0)$ überein außer ggf. bei den Halbgeraden.
3. Ist u unstetig entlang eines Teils einer Geraden $x = \hat{x}$ mit einem festen $\hat{x} \in \mathbb{R}$, dann soll dort $f(u(\hat{x}-, t)) = f(u(\hat{x}+, t))$ gelten.

Diese Bedingungen stellen sicher, dass die Integranden in der Integralform (1.3) als Funktionen einer Veränderlichen jeweils stetig oder stückweise stetig sind. Desweiteren gilt $\mathcal{F} \subset L^1_{\text{loc}}(\mathbb{R} \times \mathbb{R}_0^+)$. Zudem ist in der schwachen Formulierung (1.5) die rechte Seite durch die Bedingung 2. wohldefiniert. Ein Beispiel für den Fall in Bedingung 3. ist $f(u) = u^2$ mit $u(\hat{x}-, t) = 1$ und $u(\hat{x}+, t) = -1$.

Wir untersuchen Anfangswertprobleme (AWP) zur Integralform oder Differentialform.

Definition 1.5 *Ein Cauchy-Problem ist ein AWP eines Systems von Erhaltungsgleichungen mit einer Bedingung*

$$u(x, 0) = u_0(x) \quad \text{für } x \in \mathbb{R}, \quad (1.6)$$

wobei die Funktion $u_0 : \mathbb{R} \rightarrow \mathbb{R}^k$ vorgegeben ist. Ein Riemann-Problem ist der Spezialfall einer Vorgabe

$$u(x, 0) = \begin{cases} u_l & \text{für } x < 0 \\ u_r & \text{für } x > 0 \end{cases} \quad (1.7)$$

mit Konstanten $u_l, u_r \in \mathbb{R}^k$ ($u_l \neq u_r$).

1.3 Beispiele

Wir stellen zwei Beispiele für Systeme aus Erhaltungsgleichungen vor. Die Darstellung erfolgt in Differentialform. Jedoch existiert dazu auch die entsprechende Integralform.

Eulersche Gasgleichungen

Aus den Navier-Stokes-Gleichungen entstehen durch Idealisierung (Annahme eines idealen Gases, Vernachlässigung der Diffusion) die Eulerschen Gasgleichungen, welche ein Musterbeispiel für Erhaltungsgleichungen darstellen. Die eindimensionalen Eulerschen Gasgleichungen besitzen die Gestalt

$$\frac{\partial}{\partial t} \begin{pmatrix} \rho \\ \rho v \\ E \end{pmatrix} + \frac{\partial}{\partial x} \begin{pmatrix} \rho v \\ \rho v^2 + p(\rho, v, E) \\ v(E + p(\rho, v, E)) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad (1.8)$$

mit der Dichte ρ , der Geschwindigkeit v , der Energie E und dem Druck p . Diese Größen hängen jeweils vom eindimensionalen Ort x und der Zeit t ab. Das System (1.8) spiegelt die Erhaltung von Masse, Impuls und Energie wider. Erhaltungsgrößen sind daher die Massendichte ρ , die Impulsdichte ρv und die Energie E . Der Druck hängt von diesen Variablen in einer vordefinierten Weise ab. Die eindimensionalen Eulerschen Gasgleichungen beschreiben z.B. die Strömung eines Gases durch ein Rohr. Hierbei sind auch unstetige Lösungen physikalisch sinnvoll. Dann ist von der Differentialform (1.8) zur zugehörigen Integralform überzugehen.

Die Eulerschen Gasgleichungen (1.8) besitzen die Differentialform (1.4) mit

$$\text{Erhaltungsgrößen: } u = \begin{pmatrix} \rho \\ \rho v \\ E \end{pmatrix} \quad \text{und} \quad \text{Flussfkt.: } f(u) = \begin{pmatrix} \rho v \\ \rho v^2 + p \\ v(E + p) \end{pmatrix}.$$

Der Druck p wird als Funktion der Erhaltungsgrößen physikalisch festgelegt, z.B. über ein Stoffgesetz

$$E = (\gamma - 1)p + \frac{1}{2}\rho v^2 \quad (1.9)$$

mit $\gamma = \frac{5}{3}, \frac{7}{5}$ für ein- bzw. zweiatomige Gasmoleküle. Für die Funktionalmatrix der Flussfunktion folgt dann

$$\frac{\partial f}{\partial u} = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2}(\gamma - 3)v^2 & (3 - \gamma)v & (\gamma - 1) \\ \frac{1}{2}(\gamma - 1)v^3 - \frac{v(E+p)}{\rho} & \frac{E+p}{\rho} - (\gamma - 1)v^2 & \gamma v \end{pmatrix}. \quad (1.10)$$

Die Eigenwerte dieser Matrix sind

$$\lambda_1(u) = v - c, \quad \lambda_2(u) = v, \quad \lambda_3(u) = v + c \quad (1.11)$$

mit der Schallgeschwindigkeit $c = \sqrt{\gamma p / \rho}$. Es gilt $\lambda_1(u) < \lambda_2(u) < \lambda_3(u)$ ($p > 0$ und $\rho > 0$ vorausgesetzt).

Flachwassergleichungen

Die Beschreibung von Strömungen in flachen Gewässern wie z.B. Flüssen erfolgt über ein System von Erhaltungsgleichungen, die sogenannten Flachwassergleichungen. Wir betrachten eine Raumdimension. Die stetig differenzierbare Funktion $S(x)$ beschreibe das Profil des Grunds (Bodens) eines Flusses. Dann sei $h(x, t) > 0$ die Wasserhöhe über dem Grund und $v(x, t) \in \mathbb{R}$ die Wassergeschwindigkeit. Es folgt ein System aus zwei Erhaltungsgleichungen in Differentialform

$$\frac{\partial}{\partial t} \begin{pmatrix} h \\ hv \end{pmatrix} + \frac{\partial}{\partial x} \begin{pmatrix} hv \\ hv^2 + \frac{1}{2}gh^2 \end{pmatrix} = \begin{pmatrix} 0 \\ -gh \frac{\partial S}{\partial x} \end{pmatrix}, \quad (1.12)$$

wobei $g > 0$ die Gravitationskonstante bezeichnet. Die Erhaltungsgrößen sind $u_1 := h$ und $u_2 := hv$. Damit kann das System (1.12) umgeschrieben werden zu

$$\frac{\partial}{\partial t} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \frac{\partial}{\partial x} \begin{pmatrix} u_2 \\ \frac{u_2^2}{u_1} + \frac{1}{2}gu_1^2 \end{pmatrix} = \begin{pmatrix} 0 \\ -gu_1 \frac{\partial S}{\partial x} \end{pmatrix},$$

wo wir die Flussfunktion f ablesen können.

Ist S konstant, dann folgt $\frac{\partial S}{\partial x} \equiv 0$ und die rechte Seite in (1.12) ist null, wodurch Erhaltungsgleichungen vorliegen. Ist S nicht konstant, dann ist die rechte Seite in (1.12) ungleich null und es liegt eine Erhaltungsgleichung mit Quellterm vor.

Die Funktionalmatrix der Flussfunktion ergibt sich zu

$$\frac{\partial f}{\partial u} = \begin{pmatrix} 0 & 1 \\ -\frac{u_2^2}{u_1^2} + gu_1 & 2\frac{u_2}{u_1} \end{pmatrix}.$$

Wir erhalten als Eigenwerte

$$\lambda_1(u) = \frac{u_2}{u_1} - \sqrt{gu_1} = v - \sqrt{gh} \quad \text{und} \quad \lambda_2(u) = \frac{u_2}{u_1} + \sqrt{gu_1} = v + \sqrt{gh}. \quad (1.13)$$

Wegen $h > 0$ gilt stets $\lambda_1(u) < \lambda_2(u)$.

1.4 Hyperbolische Systeme

Ein hyperbolisches System ist ein Spezialfall eines Systems aus Erhaltungsgleichungen. Jedoch ist dieser Fall bei den meisten Anwendungsbeispielen gegeben.

Definition 1.6 *Ein System aus Erhaltungsgleichungen (in Integral- oder Differentialform) mit Flussfunktion f heißt hyperbolisch, wenn die Funktionalmatrix $\frac{\partial f}{\partial u}$ für alle relevanten Lösungswerte u reell diagonalisierbar ist. Das System heißt strikt hyperbolisch, wenn zusätzlich die Eigenwerte der Funktionalmatrix stets paarweise verschieden sind.*

Eine Matrix $A \in \mathbb{R}^{k \times k}$ ist reell diagonalisierbar genau dann, wenn eine reguläre Matrix $T \in \mathbb{R}^{k \times k}$ und eine Diagonalmatrix $D \in \mathbb{R}^{k \times k}$ existieren, so dass $A = TDT^{-1}$ gilt. Mit Definition 1.6 ist eine einzelne Erhaltungsgleichung ($k = 1$) trivialerweise strikt hyperbolisch.

Ein lineares System aus Erhaltungsgleichungen ist in Differentialform gegeben durch

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} Au = \frac{\partial u}{\partial t} + A \frac{\partial u}{\partial x} = 0. \quad (1.14)$$

Hier liegt eine Flussfunktion $f(u) = Au$ mit einer konstanten Matrix $A \in \mathbb{R}^{k \times k}$. Laut Definition 1.6 ist das lineare System (1.14) hyperbolisch, wenn die Matrix A reell diagonalisierbar ist.

Betrachtet werden nun die Eulerschen Gasgleichungen (1.8) und die Flachwassergleichungen (1.12) als nichtlineare Systeme. Die Eigenwerte (1.11) bzw. (1.13) der Funktionalmatrix der Flussfunktion sind reell und paarweise verschieden. Also sind beide Systeme strikt hyperbolisch.

Das Musterbeispiel einer hyperbolischen Dgl. zweiter Ordnung ist die Wellengleichung

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} \quad (1.15)$$

mit der Wellengeschwindigkeit $c > 0$. Wir setzen $u \in C^2$ voraus. Es sei $v = u_t$ und $w = u_x$. Die Wellengleichung (1.15) ergibt dann $v_t = c^2 w_x$. Der Satz von Schwarz besagt $u_{tx} = u_{xt}$. Dadurch folgt $v_x = (u_t)_x = (u_x)_t = w_t$. Wir erhalten ein lineares System aus zwei partiellen Dgln. erster Ordnung

$$\frac{\partial}{\partial t} \begin{pmatrix} v \\ w \end{pmatrix} + \underbrace{\begin{pmatrix} 0 & -c^2 \\ -1 & 0 \end{pmatrix}}_{=:A} \frac{\partial}{\partial x} \begin{pmatrix} v \\ w \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (1.16)$$

Dieses System hat die Gestalt einer linearen Erhaltungsgleichung (1.14). Die Matrix A besitzt die Eigenwerte $\lambda_1 = c$ und $\lambda_2 = -c$. Folglich ist die Matrix A reell diagonalisierbar. Das System (1.16) ist (strikt) hyperbolisch laut Definition 1.6.

Wir verwenden für die partiellen Ableitungen die Kurzschreibweise:

$$u_t = \frac{\partial u}{\partial t}, u_x = \frac{\partial u}{\partial x}, u_{xx} = \frac{\partial^2 u}{\partial x^2}, u_{tx} = \frac{\partial^2 u}{\partial t \partial x}, \text{ etc.}$$

2 Analytische Eigenschaften

In diesem Kapitel wird die Theorie von Erhaltungsgleichungen und die Eigenschaften ihrer Lösungen behandelt.

2.1 Charakteristiken

Wir nehmen an, dass $u : \mathbb{R} \times \mathbb{R}_0^+ \rightarrow \mathbb{R}$ eine stetig differenzierbare Lösung der Differentialform (1.4) einer einzelnen Erhaltungsgleichung ist. Differentiation mit der Kettenregel liefert die *quasilineare Form*

$$u_t + f'(u) u_x = 0. \quad (2.1)$$

Dies kann als ein Skalarprodukt geschrieben werden

$$\begin{pmatrix} 1 \\ f'(u) \end{pmatrix} \cdot \begin{pmatrix} u_t \\ u_x \end{pmatrix} = 0 \quad \text{bzw.} \quad \begin{pmatrix} 1 \\ f'(u) \end{pmatrix} \cdot \nabla u = 0.$$

Auf der linken Seite liegt somit eine Richtungsableitung vor. Da die rechte Seite null ist, ändert sich die Lösung in diese Richtung nicht. Wir bestimmen Kurven, deren Tangenten in diese Richtungen verlaufen. Es ergibt sich ein System gew. Dgl. mit dem Kurvenparameter τ als unabhängige Veränderliche

$$\begin{aligned} \dot{t}(\tau) &= 1 \\ \dot{x}(\tau) &= f'(u(x(\tau), t(\tau))) \\ \dot{u}(\tau) &= 0. \end{aligned}$$

Die Ableitung f' ist durch die Flussfunktion gegeben. Wegen der ersten Dgl. setzen wir $t = \tau$. Es ergibt sich eine Dgl. für $x(t)$.

Definition 2.1 Es sei $u : D \rightarrow \mathbb{R}$ ($D \subseteq \mathbb{R} \times \mathbb{R}_0^+$) eine Lösung der Erhaltungsgleichung (1.4). Die Charakteristiken sind Kurven $(x(t), t)$, welche die gew. Dgl.

$$x'(t) = f'(u(x(t), t)) \quad (2.2)$$

erfüllen.

Die Charakteristiken sind durch folgende Eigenschaft gekennzeichnet.

Satz 2.2 Es sei $(x(t), t)$ für $t \geq 0$ oder $t \in [t_0, t_{\text{end}}]$ eine Charakteristik einer Erhaltungsgleichung mit Lösung u . Dann ist die Charakteristik eine Gerade bzw. ein Geradenstück und die Lösung ist konstant entlang der Charakteristik.

Beweis:

Wir rechnen nach

$$\begin{aligned}\frac{d}{dt}u(x(t), t) &= u_x(x(t), t)x'(t) + u_t(x(t), t) \\ &= u_x(x(t), t)f'(u(x(t), t)) + u_t(x(t), t) \\ &= f(u)_x + u_t|_{x=x(t), t=t} = 0.\end{aligned}$$

Somit ist die Lösung $u = u^*$ konstant auf der Charakteristik. Mit der Dgl. (2.2) gilt $x'(t) = f'(u^*)$. Also ist x' konstant und damit $x(t)$ ein Geradenstück. $\square$

Bei Cauchy-Problemen geben wir Anfangswerte auf der Geraden $\{(x, 0) \in \mathbb{R}^2 : x \in \mathbb{R}\}$ vor. Um die Lösbarkeit sicherzustellen dürfen die Charakteristiken nie tangential zu dieser Geraden liegen. Da x' stets endlich ist, ist diese Eigenschaft gegeben.

Bemerkungen:

- Bei einer stetig differenzierbaren Lösung lässt sich mittels der Charakteristiken die (lokale) Eindeutigkeit der Lösung von Cauchy-Problemen zeigen. Dies folgt dann aus der eindeutigen Lösbarkeit der gew. Dgln.
- Bei einer Erhaltungsgleichung mit Quellterm $u_t + f(u)_x = g(u)$ ist eine Lösung nicht konstant entlang einer Charakteristik. Man hat aber mit $\dot{u} = g(u)$ eine gew. Dgl. zur Bestimmung von u .

Die allgemeine Theorie zu Charakteristiken bei einer einzelnen partiellen Dgl. erster Ordnung wird beschrieben in F. Erwe, E. Peschl: Partielle Differentialgleichungen erster Ordnung. BI Hochschultaschenbücher, 1973.

2.2 Lineare Advektionsgleichung

Die *lineare Advektionsgleichung* (auch: Konvektionsgleichung) ist eine Erhaltungsgleichung mit der linearen Flussfunktion $f(u) = au$, wobei $a \in \mathbb{R}$ eine Konstante ist. Die Differentialform der linearen Advektionsgleichung lautet daher

$$u_t + au_x = 0. \quad (2.3)$$

Bei Anfangswertproblemen wird das Profil in der Zeit mit konstanter Geschwindigkeit a weitertransportiert.

Satz 2.3 Für Anfangswerte u_0 sei

$$u(x, t) = u_0(x - at) \quad (2.4)$$

für $x \in \mathbb{R}$, $t \geq 0$. Ist u_0 stetig differenzierbar, dann ist (2.4) eine Lösung der Differentialform der linearen Advektionsgleichung. Ist u_0 lokal integrierbar, dann ist (2.4) eine Lösung der Integralform der linearen Advektionsgleichung.

Beweis:

i) Differentialform:

$$u_t(x, t) + au_x(x, t) = u'_0(x - at)(-a) + au'_0(x - at) = 0$$

ii) Integralform:

Für die Terme in den Integralen berechnen wir mittels Variablentransformationen

$$\begin{aligned} \int_{x_1}^{x_2} u(x, t_2) \, dx &= \int_{x_1}^{x_2} u_0(x - at_2) \, dx = \int_{x_1 - at_2}^{x_2 - at_2} u_0(s) \, ds \\ \int_{x_1}^{x_2} u(x, t_1) \, dx &= \int_{x_1}^{x_2} u_0(x - at_1) \, dx = \int_{x_1 - at_1}^{x_2 - at_1} u_0(s) \, ds \\ \int_{t_1}^{t_2} au(x_1, t) \, dt &= \int_{t_1}^{t_2} au_0(x_1 - at) \, dt = - \int_{x_1 - at_1}^{x_1 - at_2} u_0(s) \, ds \\ - \int_{t_1}^{t_2} au(x_2, t) \, dt &= - \int_{t_1}^{t_2} au_0(x_2 - at) \, dt = \int_{x_2 - at_1}^{x_2 - at_2} u_0(s) \, ds. \end{aligned}$$

Die Integration über $[x_1 - at_2, x_2 - at_2]$ kann zerlegt werden in

$$\int_{x_1 - at_2}^{x_2 - at_2} u_0(s) \, ds = \int_{x_1 - at_2}^{x_1 - at_1} u_0(s) \, ds + \int_{x_1 - at_1}^{x_2 - at_1} u_0(s) \, ds + \int_{x_2 - at_1}^{x_2 - at_2} u_0(s) \, ds.$$

Ein Vergleich der Terme liefert die Behauptung. $\square$

Wir bestimmen noch die Charakteristiken zur Advektionsgleichung (2.3). Die Dgl. (2.2) wird hier einfach zu

$$x'(t) = a.$$

Also sind die Charakteristiken die Geraden

$$x(t) = at + c$$

mit Konstanten $c \in \mathbb{R}$. Mit (2.4) gilt $u(x(t), t) = u_0(at + c - at) = u_0(c)$.

2.3 Burgers Gleichung

Die bei Erhaltungsgleichungen auftretenden Effekte sollen anhand einer einfachen nichtlinearen Gleichung analysiert werden. Das bekannteste Modellproblem ist die *Burgers Gleichung*, welche die Flussfunktion $f(u) = \frac{1}{2}u^2$ besitzt. Diese Erhaltungsgleichung lautet daher

$$u_t + \left(\frac{1}{2}u^2\right)_x = 0 \quad (2.5)$$

bzw. in quasilinearer Form

$$u_t + uu_x = 0. \quad (2.6)$$

Wir untersuchen nun Cauchy-Probleme (1.6) zur Burgers Gleichung (2.5).

Glatte Anfangswerte

Es seien stetig differenzierbare Anfangswerte $u_0(x)$ vorgegeben. Die Charakteristiken sind gekennzeichnet durch die gew. Dgl. (2.2), d.h. wegen $f'(u) = u$ hier

$$x'(t) = u(x(t), t). \quad (2.7)$$

Nach Satz 2.2 sind die Charakteristiken Geraden und die Lösung u konstant entlang einer Charakteristik. Wir betrachten die Charakteristiken, die von der Geraden $t = 0$ ausgehen. Es folgt mit (2.7) die Geradenschar

$$x(t) = \xi + u_0(\xi)t \quad (2.8)$$

mit Scharparameter $\xi \in \mathbb{R}$. (Es gilt $x(0) = \xi$ und $x'(t) = u_0(\xi)$.)

Wir betrachten nun zwei Anfangspunkte $\xi_1 < \xi_2$ und die korrespondierenden Charakteristiken

$$x_1(t) = \xi_1 + u_0(\xi_1)t, \quad x_2(t) = \xi_2 + u_0(\xi_2)t. \quad (2.9)$$

Gilt $u_0(\xi_1) = u_0(\xi_2)$, dann sind die Geraden parallel und schneiden sich nicht. Gilt $u_0(\xi_1) \neq u_0(\xi_2)$, dann sind die Geraden nicht parallel und schneiden sich in der (x, t) -Ebene bei der Zeit

$$T = \frac{\xi_1 - \xi_2}{u_0(\xi_2) - u_0(\xi_1)}. \quad (2.10)$$

Es ist $T > 0$ für $u_0(\xi_1) > u_0(\xi_2)$ und $T < 0$ für $u_0(\xi_1) < u_0(\xi_2)$.

1. Fall: $u'_0(x) \geq 0$ für alle x

Hier ist u_0 monoton steigend. Es folgt $u_0(\xi_1) \leq u_0(\xi_2)$. Also schneiden sich die

Charakteristiken nie bei Zeiten (2.10) mit $T \geq 0$. Eine klassische Lösung des Cauchy-Problems existiert somit in ganz $\mathbb{R} \times \mathbb{R}_0^+$.

2. Fall: $u'_0(x) < 0$ für (mindestens) ein x

Da u'_0 stetig ist, gibt es ein offenes Intervall I mit $u_0 < 0$. Jetzt existieren Zeiten (2.10) mit $T > 0$. In diesem Fall kann spätestens ab einer solchen Zeit T keine klassische Lösung des Anfangswertproblems existieren. Da die Lösung nämlich auf Charakteristiken konstant ist, müsste im Schnittpunkt $x_1(T) = x_2(T)$ sowohl $u(x_1(T), T) = u_0(\xi_1)$ als auch $u(x_2(T), T) = u_0(\xi_2)$ gelten im Widerspruch zu $u_0(\xi_1) \neq u_0(\xi_2)$ für $\xi_1, \xi_2 \in I$. Die klassische Lösung existiert somit nicht für alle $t > 0$. Die Lösung hört auf zu existieren bei

$$\bar{T} := \inf \left\{ \frac{\xi_1 - \xi_2}{u_0(\xi_2) - u_0(\xi_1)} : \xi_1 < \xi_2 \text{ und } u_0(\xi_1) > u_0(\xi_2) \right\}.$$

Ist u'_0 nach unten beschränkt, dann gilt mit dem Mittelwertsatz der Differentialrechnung

$$\frac{\xi_1 - \xi_2}{u_0(\xi_2) - u_0(\xi_1)} = -\frac{\xi_2 - \xi_1}{u_0(\xi_2) - u_0(\xi_1)} = -\frac{1}{u'_0(\eta)} \geq -\frac{1}{\inf\{u'_0(x) : x \in \mathbb{R}\}} =: \hat{T}$$

mit $\xi_1 < \eta < \xi_2$. Es folgt $\bar{T} \geq \hat{T}$, d.h. eine klassische Lösung existiert für $t \in [0, \hat{T}]$. Bemerkenswert ist nun, dass schwache Lösungen des AWP's zur Burgers Gleichung für alle $t > 0$ existieren können.

Durch Einbezug eines Diffusionsterms entsteht die *viskose Burgers Gleichung*

$$u_t + \left(\frac{1}{2}u^2\right)_x = \varepsilon u_{xx} \quad \Leftrightarrow \quad u_t + uu_x = \varepsilon u_{xx} \quad (2.11)$$

mit $\varepsilon > 0$. Diese Gleichung dient zum Vergleich von physikalischen Effekten. Insbesondere existieren zu Anfangswertproblemen hier glatte Lösungen für alle Zeiten $t > 0$. In Grenzfall $\varepsilon \rightarrow 0$ entsteht die Burgers Gleichung (2.5)

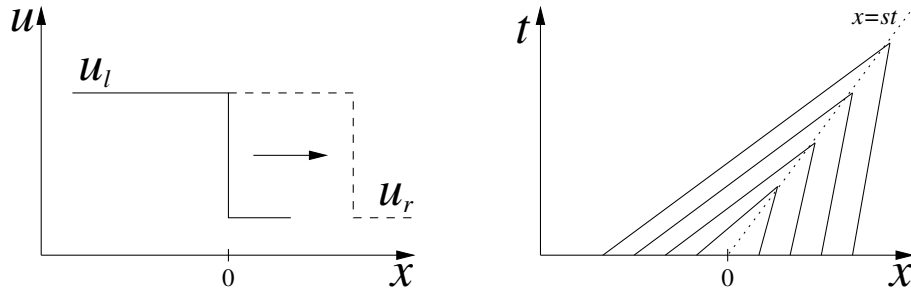
Riemann-Problem

Wir untersuchen das Riemann-Problem (1.7) für die Burgers Gleichung. Dazu sind zwei Fälle zu unterscheiden.

1. Fall: $u_l > u_r$

Hier ergibt sich die eindeutige schwache Lösung

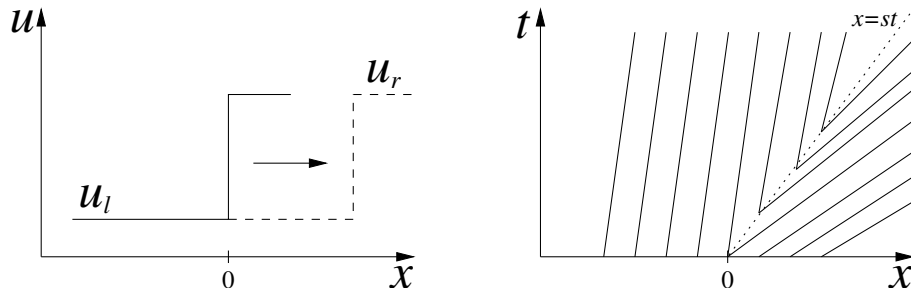
$$u(x, t) = \begin{cases} u_l & \text{für } x < st \\ u_r & \text{für } x > st \end{cases} \quad \text{mit} \quad s = \frac{u_l + u_r}{2}. \quad (2.12)$$



Diese Form der Lösung nennt man *Stoßwelle* oder *Schockwelle*. Die Unstetigkeit bzw. der Schock wird einfach mit der *Schockgeschwindigkeit* s weitertransportiert. Eine Herleitung dieser Geschwindigkeit folgt im nächsten Abschnitt. Die Charakteristiken laufen dabei mit der Zeit in den Schock hinein. Werden die Anfangswerte leicht zu einer glatten Funktion abgeändert, so laufen auch die neuen Charakteristiken in den Schock hinein und nach kurzer Zeit entsteht die gleiche Lösung. Somit ist dieser Verdichtungsschock stabil.

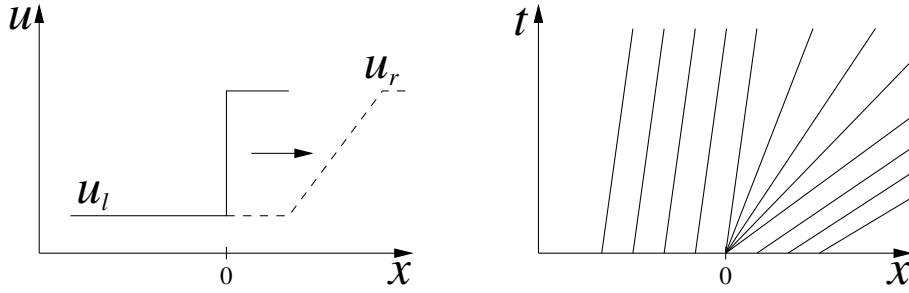
2. Fall: $u_l < u_r$

Auch hier ist die Schockwelle (2.12) eine schwache Lösung des Problems, die nun einen Verdünnungsschock darstellt.



Hier laufen die Charakteristiken jedoch aus dem Schock heraus, wodurch eine Instabilität gekennzeichnet ist. Werden die Anfangswerte leicht zu einer glatten Funktion modifiziert, so ändert sich die Lösung erheblich. Im 2. Fall existieren jedoch unendlich viele schwache Lösungen. Eine davon ist die *Verdünnungswelle*

$$u(x, t) = \begin{cases} u_l & \text{für } x < u_l t \\ \frac{x}{t} & \text{für } u_l t \leq x \leq u_r t \\ u_r & \text{für } x > u_r t. \end{cases} \quad (2.13)$$



Diese Lösung entsteht auch näherungsweise, wenn die Anfangswerte leicht zu glatten Daten geändert werden. Zudem ist (2.13) die Lösung bei verschwindender Diffusion in Gleichung (2.11) und somit der physikalisch sinnvolle Zustand.

2.4 Lösungstheorie

In diesem Abschnitt wird die Lösungstheorie für eine einzelne nichtlineare Erhaltungsgleichung dargestellt.

Glatte Anfangswerte

Zunächst zeigen wir eine hinreichende Bedingung dafür, dass die Charakteristiken den Definitionsbereich der gewünschten Lösung ausfüllen.

Lemma 2.4 *Sind die Anfangswerte $u_0 : \mathbb{R} \rightarrow \mathbb{R}$ stetig und beschränkt, dann geht durch jeden Punkt (x, t) mit $x, t \in \mathbb{R}$ (mindestens) eine Charakteristik.*

Beweis:

Da die Lösung auf den Charakteristiken konstant ist, sind die Charakteristiken beim AWP gegeben durch

$$x(t) = \xi + f'(u_0(\xi))t = \xi + f'(u_0(\xi))t.$$

Seien $x^*, t^* \in \mathbb{R}$ fest gewählt. Es ergibt sich die Gleichung

$$x^* = \xi + f'(u_0(\xi))t^*$$

für die Unbekannte ξ . In Nullstellenform erhalten wir die Funktion

$$g(\xi) = \xi + f'(u_0(\xi))t^* - x^*. \quad (2.14)$$

Da u_0 beschränkt ist, gilt $u_0(\xi) \in U = [\underline{u}, \bar{u}]$. Da f' stetig ist, ist auch $f'(u_0)$ auf U beschränkt. Ebenso ist $f'(u_0)$ und damit g stetig in ξ . Für hinreichend hohes ξ

gilt $g(\xi) > 0$. Für hinreichend kleines ξ gilt $g(\xi) < 0$. Mit dem Zwischenwertsatz gibt es (mindestens) ein $\hat{\xi}$ mit $g(\hat{\xi}) = 0$. $\square$

Lemma 2.4 gilt nicht falls u_0 einen Sprung beinhaltet, da dann der Zwischenwertsatz nicht mehr angewendet werden kann.

Satz 2.5 *In einer Erhaltungsgleichung sei die Flussfunktion f zweimal stetig differenzierbar und konvex. Die Anfangswerte $u_0 : \mathbb{R} \rightarrow \mathbb{R}$ seien stetig differenzierbar.*

- i) *Falls $u'_0 \geq 0$ gilt, dann existiert eine eindeutige klassische Lösung in ganz $\mathbb{R} \times \mathbb{R}_0^+$.*
- ii) *Falls $u'_0 < 0$ auftritt, dann existiert eine eindeutige klassische Lösung nur für $t \in [0, T)$ mit*

$$T = -\frac{1}{\inf\{(f' \circ u_0)'(x) : x \in \mathbb{R}\}}.$$

Existiert das Infimum nicht, dann gibt es keinen Bereich $\mathbb{R} \times [0, T)$, $T > 0$ mit einer klassischen Lösung.

Beweis:

i) Die Konvexität der Flussfunktion bedeutet $f'' > 0$. Folglich ist f' streng monoton steigend. Wegen $u'_0 \geq 0$ ist u_0 monoton steigend. Also ist auch $f'(u_0)$ monoton steigend. Wir betrachten wieder die Funktion (2.14) für feste $x^* \in \mathbb{R}$, $t^* \geq 0$. Wegen $t^* \geq 0$ ist g streng monoton steigend. Für hinreichend hohes ξ gilt $g(\xi) > 0$. Für hinreichend niedriges ξ gilt $g(\xi) < 0$. Mit dem Zwischenwertsatz und der Monotonie gibt es genau ein $\hat{\xi}$ mit $g(\hat{\xi}) = 0$. Somit geht genau eine Charakteristik durch (x^*, t^*) .

Die Charakteristiken sind Geraden und die Lösung ist konstant entlang einer Charakteristik. Dadurch ist die Lösung implizit gegeben durch

$$u(x, t) = u_0(x - f'(u(x, t))t).$$

Differentiation liefert

$$u_x = u'_0 \cdot (1 - f''(u)u_x t)$$

$$u_t = u'_0 \cdot (-f'(u) - f''(u)u_t t).$$

Wir lösen diese Gleichungen jeweils nach u_x und u_t auf

$$u_x = \frac{u'_0}{1 + u'_0 f''(u)t}$$

$$u_t = -\frac{f'(u)u'_0}{1 + u'_0 f''(u)t}.$$

Wegen $f'' > 0$, $u'_0 \geq 0$ sind die Nenner stets positiv für $t \geq 0$. Es folgt

$$u_t + f'(u)u_x = -\frac{f'(u)u'_0}{1 + u'_0 f''(u)t} + \frac{f'(u)u'_0}{1 + u'_0 f''(u)t} = 0.$$

Also ist die Dgl. erfüllt.

ii) Wenn $u'_0 < 0$ auftritt, dann gibt es $\xi_1 < \xi_2$ mit $u_0(\xi_1) > u_0(\xi_2)$. Es folgt $f'(u_0(\xi_1)) > f'(u_0(\xi_2))$. Die Lösung hört auf zu existieren bei

$$\bar{T} := \inf \left\{ \frac{\xi_1 - \xi_2}{f'(u_0(\xi_2)) - f'(u_0(\xi_1))} : \xi_1 < \xi_2 \text{ und } f'(u_0(\xi_1)) > f'(u_0(\xi_2)) \right\}.$$

Aus dem Mittelwertsatz der Differentialrechnung ergibt sich mit einer Zwischenstelle $\xi_1 < \eta < \xi_2$

$$\begin{aligned} \frac{\xi_1 - \xi_2}{f'(u_0(\xi_2)) - f'(u_0(\xi_1))} &= -\frac{\xi_2 - \xi_1}{f'(u_0(\xi_2)) - f'(u_0(\xi_1))} \\ &= -\frac{1}{(f' \circ u_0)'(\eta)} \geq -\frac{1}{\inf\{(f' \circ u_0)'(x) : x \in \mathbb{R}\}} =: \hat{T}. \end{aligned}$$

Dadurch gilt $\bar{T} \geq \hat{T}$.

Angenommen es wäre $\bar{T} > \hat{T}$. Dann würde gelten $\frac{1}{\bar{T}} < \frac{1}{\hat{T}}$ und $-\frac{1}{\bar{T}} > -\frac{1}{\hat{T}}$. Zu $\varepsilon > 0$ wählen wir ein $\hat{x}$ mit

$$\left| (f' \circ u_0)'(\hat{x}) - \left(-\frac{1}{\hat{T}} \right) \right| < \frac{\varepsilon}{2}$$

Für $h > 0$ hinreichend klein gilt

$$\left| \frac{f'(u_0(\hat{x} + h)) - f'(u_0(\hat{x}))}{h} - \left(-\frac{1}{\hat{T}} \right) \right| < \varepsilon.$$

Mit $\xi_1 = \hat{x}$ und $\xi_2 = \hat{x} + h$ folgt aber

$$\frac{f'(u_0(\xi_2)) - f'(u_0(\xi_1))}{\xi_2 - \xi_1} \geq -\frac{1}{\bar{T}} > -\frac{1}{\hat{T}}.$$

Für ε hinreichend klein entsteht ein Widerspruch. □

Bemerkungen:

- Analoge Aussagen gelten für eine Erhaltungsgleichung mit konkaver Flussfunktion ($f'' < 0$).
- Die Burgers Gleichung (2.5) besitzt eine konvexe Flussfunktion ($f'' = 1$).

Schockgeschwindigkeit

Für ein allgemeines System aus Erhaltungsgleichungen soll eine Bedingung für die Schockgeschwindigkeit s , d.h. die Geschwindigkeit mit der Unstetigkeiten sich bewegen, hergeleitet werden. Dazu dient wieder eine Diskussion des Riemann-Problems (1.7). Wir betrachten die Schockwelle

$$u(x, t) = \begin{cases} u_l & \text{für } x < st \\ u_r & \text{für } x > st \end{cases} \quad (2.15)$$

mit $u_l, u_r \in \mathbb{R}^k$, $u_l \neq u_r$ als schwache Lösung, wobei s noch nicht spezifiziert sei.

Satz 2.6 *Falls die Schockwelle (2.15) eine schwache Lösung ist, dann erfüllt die Schockgeschwindigkeit notwendigerweise die Rankine-Hugoniot Sprungbedingung*

$$f(u_l) - f(u_r) = s(u_l - u_r). \quad (2.16)$$

Beweis:

Zur Zeit $t \geq 0$ sei $M > 0$ so groß im Vergleich zu st gewählt, dass klar $u(x, t) = u_l$ für $x \leq -M$ und $u(x, t) = u_r$ für $x \geq M$ gilt. Aus der Integralform (1.3) folgt

$$\begin{aligned} \int_{-M}^M u(x, t + \Delta t) \, dx - \int_{-M}^M u(x, t) \, dx &= \int_t^{t+\Delta t} f(u(-M, \tau)) \, d\tau \\ &\quad - \int_t^{t+\Delta t} f(u(M, \tau)) \, d\tau. \end{aligned}$$

Division durch Δt und Grenzübergang $\Delta t \rightarrow 0$ liefert die Aussage

$$\frac{d}{dt} \int_{-M}^M u(x, t) \, dx = f(u(-M, t)) - f(u(M, t)) = f(u_l) - f(u_r).$$

Aus (2.15) erhalten wir andererseits

$$\int_{-M}^M u(x, t) \, dx = (M + st)u_l + (M - st)u_r$$

und damit

$$\frac{d}{dt} \int_{-M}^M u(x, t) \, dx = s(u_l - u_r).$$

Zusammen folgt die Rankine-Hugoniot Sprungbedingung (2.16). $\square$

Bei einem linearen System gilt $f(u) = Au$ mit einer Matrix $A \in \mathbb{R}^{k \times k}$. Dann lautet die Rankine-Hugoniot Sprungbedingung

$$A(u_l - u_r) = Au_l - Au_r = s(u_l - u_r). \quad (2.17)$$

Damit ist $u_l - u_r \neq 0$ ein Eigenvektor der Matrix A zum Eigenwert s . Für beliebig vorgegebene Werte u_l, u_r ist die Bedingung (2.17) daher im allgemeinen nicht erfüllt.

Im Fall einer einzelnen Erhaltungsgleichung vereinfacht sich (2.16) zu

$$s = \frac{f(u_l) - f(u_r)}{u_l - u_r} \quad (2.18)$$

mit $u_l \neq u_r$. Diese Bedingung kann für gegebene u_l, u_r stets erfüllt werden. Bei der linearen Advektionsgleichung (2.3) folgt wegen $f(u) = au$ hier $s = a$.

Bemerkungen:

- Man kann zeigen, dass bei nichtkonstanter Lösung jeweils zur linken und rechten Seite einer Unstetigkeit dann die Rankine-Hugoniot Sprungbedingung lokal mit im allgemeinen nichtkonstanter Geschwindigkeit erfüllt ist.
- Die Rankine-Hugoniot Sprungbedingung ist auch hinreichend dafür, dass eine Schockwelle (2.15) schwache Lösung ist. Dadurch ist aber nicht garantiert, dass die zulässige Unstetigkeit auch physikalisch sinnvoll ist (siehe instabile Schockwelle (2.12) bei Burgers Gleichung für $u_l < u_r$).
- Die Sprungbedingung (2.16) bedeutet die lineare Abhängigkeit zweier Vektoren und ist im allgemeinen nicht erfüllt. Eine Lösung des Riemann-Problems bei Systemen besteht daher nicht nur aus einer einzelnen Unstetigkeit wie in (2.15) sondern im allgemeinen aus k verschiedenen Unstetigkeiten oder Verdünnungen.
- In (2.18) folgt aus $s = 0$ wegen $u_l \neq u_r$ notwendigerweise $f(u_l) = f(u_r)$. Diese Bedingung wird auch gefordert bei der Definition des Funktionsraums $\mathcal{F}$ in Abschnitt 1.2.
- Bei der Burgers Gleichung folgt aus (2.18) $s = \frac{\frac{1}{2}u_l^2 - \frac{1}{2}u_r^2}{u_l - u_r} = \frac{1}{2}(u_l + u_r)$.

Verdünnungswellen

Wir betrachten ein Riemann-Problem (1.7) mit $u_l < u_r$ für eine einzelne nichtlineare Erhaltungsgleichung.

Satz 2.7 *Gegeben sei ein Riemann-Problem mit $u_l < u_r$ zu einer Erhaltungsgleichung mit konvexer Flussfunktion. Dann stellt die Verdünnungswelle*

$$u(x, t) = \begin{cases} u_l & \text{für } x < f'(u_l)t \\ v(\frac{x}{t}) & \text{für } f'(u_l)t \leq x \leq f'(u_r)t \\ u_r & \text{für } x > f'(u_r)t, \end{cases} \quad (2.19)$$

wobei $v(\xi)$ die Gleichung $f'(v(\xi)) = \xi$ erfüllt, eine schwache Lösung dar.

Beweis:

Konstante Funktionen erfüllen trivialerweise die Differentialform der Erhaltungsgleichung. Auch $v(\frac{x}{t})$ ist eine Lösung der Differentialform wegen

$$v(\frac{x}{t})_t + f(v(\frac{x}{t}))_x = v'(\frac{x}{t})(-\frac{x}{t^2}) + f'(v(\frac{x}{t}))v'(\frac{x}{t})\frac{1}{t} = -v'(\frac{x}{t})\frac{x}{t^2} + \frac{x}{t}v'(\frac{x}{t})\frac{1}{t} = 0.$$

Die Funktion (2.19) ist in $\mathbb{R} \times \mathbb{R}_0^+$ in den drei Bereichen

$$\begin{aligned} &\{(x, t) \in \mathbb{R} \times \mathbb{R}_0^+ : x < f'(u_l)t\} \\ &\{(x, t) \in \mathbb{R} \times \mathbb{R}_0^+ : f'(u_l)t \leq x \leq f'(u_r)t\} \\ &\{(x, t) \in \mathbb{R} \times \mathbb{R}_0^+ : x > f'(u_r)t\} \end{aligned}$$

jeweils verschieden definiert. Diese drei Bereiche sind durch zwei Geraden abgegrenzt. Im jeweiligen Bereich ist die Funktion u glatt. Mit $f'' > 0$ ist f' streng monoton steigend und damit injektiv. Die Bedingung $f'(v(\xi)) = \xi$ bedeutet gerade, dass v die Umkehrfunktion von f' ist. Es folgt

$$v\left(\frac{f'(u_{l/r})t}{t}\right) = v(f'(u_{l/r})) = u_{l/r}$$

und dadurch ist u an den zwei Grenzen der drei Bereiche stetig. Insgesamt ist u stetig in $\mathbb{R} \times \mathbb{R}_0^+$.

Für die Integralform (1.3) muss gelten

$$I := \int_{x_1}^{x_2} u(x, t_2) - u(x, t_1) \, dx + \int_{t_1}^{t_2} f(u(x_2, t)) - f(u(x_1, t)) \, dt = 0$$

für alle $x_1 < x_2$, $0 \leq t_1 < t_2$. Durch Aufteilen der Integrale in kleine Intervalle erhalten wir

$$I = \sum_{i=0}^n \int_{x_1+ih}^{x_1+(i+1)h} u(x, t_2) - u(x, t_1) \, dx + \sum_{j=0}^n \int_{t_1+jk}^{t_1+(j+1)k} f(u(x_2, t)) - f(u(x_1, t)) \, dt$$

mit Schrittweiten $h = \frac{x_2 - x_1}{n+1}$ und $k = \frac{t_2 - t_1}{n+1}$. Durch bilden von Teleskopsummen folgt

$$I = \sum_{j=0}^n \sum_{i=0}^n \int_{x_1 + ih}^{x_1 + (i+1)h} u(x, t_1 + (j+1)k) - u(x, t_1 + jk) \, dx \\ + \sum_{i=0}^n \sum_{j=0}^n \int_{t_1 + jk}^{t_1 + (j+1)k} f(u(x_1 + (i+1)h, t)) - f(u(x_1 + ih, t)) \, dt,$$

wodurch eine Zerlegung in kleine Rechtecke entsteht. Betrachtet wird der Grenzfall $n \rightarrow \infty$. Durch den Bereich $[x_1, x_2] \times [t_1, t_2]$ laufen höchstens zwei Geradenstücke, auf denen u stetig aber nicht glatt ist. Auf allen Rechtecken, die keines der Geradenstücke enthalten, ist daher die Differentialform erfüllt und damit auch die Integralform über das kleine Rechteck.

Es bleiben höchstens $4n$ Rechtecke übrig. Die Integrale bestehen auf Intervallen der Länge h, k antiproportional zu n . Da u und f stetig sind, sind sie in dem betrachteten kompakten Bereich auch gleichmäßig stetig. Zu $\varepsilon > 0$ gibt es ein $\delta > 0$ mit

$$|u(x, t_1 + (j+1)k) - u(x, t_1 + jk)| < \varepsilon \\ |f(u(x_1 + (i+1)h, t)) - f(u(x_1 + ih, t))| < \varepsilon$$

für $h, k < \delta$. Es folgt die Abschätzung $|I| < ((x_2 - x_1) + (t_2 - t_1))4\varepsilon$. Da ε beliebig klein sein kann, erhalten wir $|I| = 0$. $\square$

Bemerkungen:

- Im Beweis von Satz 2.7 ist wesentlich, dass die Lösung (2.19) global stetig ist. Bei einer Schockwelle kann das Beweisprinzip nicht angewendet werden, da ein Sprung vorliegt.
- Bei der Burgers Gleichung gilt $f'(u) = u$. Aus Satz 2.7 ergibt sich die Bedingung $v(\xi) = \xi$. Daher liefert die Formel (2.19) die bekannte Verdünnungswelle (2.13).

2.5 Entropiebedingungen

Schwache Lösungen von AWPen zu Erhaltungsgleichungen sind im allgemeinen nicht eindeutig. Dagegen besitzt die zugehörige parabolische Differentialgleichung

$$u_t + f(u)_x = \varepsilon u_{xx} \quad (2.20)$$

zu AWPen eine eindeutige Lösung. Daher versuchen wir, durch zusätzliche Bedingungen aus der Menge der schwachen Lösungen die Lösung bei verschwindender Diffusion ($\varepsilon \rightarrow 0$) von (2.20) als physikalisch sinnvollen Zustand zu erhalten. Die entstehenden Forderungen nennt man *Entropiebedingungen* in Anlehnung an die Gasdynamik. Ein physikalisch realer Zustand kann beim Durchlaufen einer Unstetigkeit seine Entropie nur erhöhen aber nicht verringern. Die folgenden Konzepte sind aber eher mathematischer Natur und lassen sich nicht direkt über die Entropie von Gasen interpretieren.

Bedingung über Charakteristiken

Wir betrachten eine skalare Erhaltungsgleichung $u_t + f(u)_x = 0$ mit konvexer Flussfunktion ($f'' > 0$). Das Riemann-Problem für die Burgers Gleichung zeigt, dass eine Schockwelle stabil ist, wenn alle Charakteristiken in die Unstetigkeit hineinlaufen, und instabil, wenn Charakteristiken aus der Unstetigkeit herauslaufen. Daher stellen wir als eine erste Entropiebedingung einfach

$$f'(u_l) > s > f'(u_r), \quad (2.21)$$

wobei s die Geschwindigkeit der Unstetigkeit von u_l nach u_r bezeichnet. Dazu muss natürlich $f'(u_l) > f'(u_r)$ gelten. Für konvexes f ist daher $u_l > u_r$ notwendig. Nach der Rankine-Hugoniot Sprungbedingung (2.18) gilt

$$s = \frac{f(u_l) - f(u_r)}{u_l - u_r} = f'(\theta u_l + (1 - \theta)u_r) \quad \text{mit } \theta \in (0, 1),$$

wodurch wegen f konvex dann (2.21) stets erfüllt ist. Glättet man die Anfangsdaten aus dem Riemann-Problem, so treten alle Zwischenwerte im Intervall $[u_r, u_l]$ auf. Mit der Konvexität von f laufen auch alle Charakteristiken, die in den Zwischenwerten starten, in den Schock hinein und er bleibt stabil. Entsprechende ist für konkave Flussfunktionen ($f'' < 0$) dann (2.21) für $u_r > u_l$ immer gegeben.

Ist f weder konvex noch konkav, so wird stattdessen verlangt

$$\frac{f(u) - f(u_l)}{u - u_l} \geq s \geq \frac{f(u) - f(u_r)}{u - u_r} \quad (2.22)$$

für alle $u \in (u_l, u_r)$ bzw. $u \in (u_r, u_l)$. Die beiden Terme in (2.22) können als die Schockgeschwindigkeiten zu den Unstetigkeiten u_l, u bzw. u, u_r angesehen werden. Also garantiert (2.22), dass eine Abänderung des Riemann-Problems in zwei

Unstetigkeiten dann die einzelnen Sprünge in einem einzigen Schock verschwinden.

Beispiel: Wir betrachten eine Erhaltungsgleichung mit Flussfunktion $f(u) = u^3$. Wegen $f''(u) = 6u$ ist f weder konvex noch konkav. Für $u_l = -1$ und $u_r = 1$ ist (2.22) verletzt. Für $u_l = -2$ und $u_r = 1$ ist (2.22) erfüllt.

Bedingung über Entropiefunktionen

Ein weiteres Konzept erhält man durch die sogenannten *Entropiefunktionen*. Sei u eine glatte Lösung der Erhaltungsgleichung. Eine zweimal stetig differenzierbare, konvexe Funktion $\eta : \mathbb{R} \rightarrow \mathbb{R}$ sei vorgegeben. Wir nehmen an, dass die Entropiefunktion $\eta(u)$ die Erhaltungsgleichung

$$\eta(u)_t + \psi(u)_x = 0 \quad \Leftrightarrow \quad \eta'(u)u_t + \psi'(u)u_x = 0 \quad (2.23)$$

mit dem Entropiefluss $\psi(u)$ erfüllt. Aus der Erhaltungsgleichung folgt der Fluss über

$$\eta'(u)u_t + \eta'(u)f'(u)u_x = 0 \quad \Rightarrow \quad \psi'(u) = \eta'(u)f'(u). \quad (2.24)$$

Für eine glatte Lösung u sind alle auftretenden Funktionen glatt und die Erhaltung der Entropie $\eta(u)$ ist nach Definition erfüllt. Für unstetiges u sind die obigen Herleitungen nicht zulässig. Daher betrachten wir die Diffusionsgleichung (2.20), die auch zu unstetigen Anfangswerten eine glatte Lösung u^ε besitzt. Aus Gründen der Übersichtlichkeit schreiben wir nur u für die von ε abhängige Lösung. Durch Multiplikation mit $\eta'(u)$ entsteht aus (2.20)

$$\begin{aligned} \eta(u)_t + \psi(u)_x &= \varepsilon \eta'(u)u_{xx} \\ &= \varepsilon (\eta'(u)u_x)_x - \varepsilon \eta''(u)u_x^2. \end{aligned}$$

Integration über einen Bereich $[x_1, x_2] \times [t_1, t_2]$ ergibt

$$\begin{aligned} & \int_{t_1}^{t_2} \int_{x_1}^{x_2} \eta(u)_t + \psi(u)_x \, dx \, dt \\ &= \varepsilon \int_{t_1}^{t_2} [\eta'(u(x_2, t))u_x(x_2, t) - \eta'(u(x_1, t))u_x(x_1, t)] \, dt \\ & \quad - \varepsilon \int_{t_1}^{t_2} \int_{x_1}^{x_2} \eta''(u)u_x^2 \, dx \, dt. \end{aligned} \quad (2.25)$$

Wir sind an der Lösung bei verschwindender Diffusion als physikalischen Zustand interessiert. Für $\varepsilon \rightarrow 0$ fällt der erste Term mit ε in (2.25) weg. Besitzt die Grenzlösung u eine Unstetigkeit im Integrationsbereich, dann wird der zweite

Term im allgemeinen nicht verschwinden bei $\varepsilon \rightarrow 0$. Wegen $\varepsilon > 0$, $u_x^2 \geq 0$, $\eta'' > 0$ folgt

$$\int_{t_1}^{t_2} \int_{x_1}^{x_2} \eta(u)_t + \psi(u)_x \, dx \, dt \leq 0 \quad \text{für alle } t_1 < t_2, \, x_1 < x_2. \quad (2.26)$$

Beim Grenzübergang müssen die Ableitungen hier nicht existieren. Daher gehen wir schon vorher zur korrespondierenden Integralform über und erhalten dann

$$\begin{aligned} \int_{x_1}^{x_2} \eta(u(x, t_2)) \, dx &\leq \int_{x_1}^{x_2} \eta(u(x, t_1)) \, dx \\ &+ \int_{t_1}^{t_2} \psi(u(x_1, t)) \, dt - \int_{t_1}^{t_2} \psi(u(x_2, t)) \, dt. \end{aligned} \quad (2.27)$$

Als Abkürzung für diese Integralbedingung wird häufig $\eta(u)_t + \psi(u)_x \leq 0$ verwendet.

Definition 2.8 *Eine schwache Lösung zu einer Erhaltungsgleichung heißt Entropielösung, wenn die Ungleichung (2.27) für alle $x_1 \leq x_2$, $0 \leq t_1 \leq t_2$ und alle konvexen Entropiefunktionen η mit zugehörigen Flüssen ψ erfüllt ist.*

Die Lösung bei verschwindender Diffusion ist über diese Konstruktion eine Entropielösung.

Analog zu Erhaltungsgleichungen kann auch hier eine äquivalente schwache Formulierung mittels Testfunktionen hergeleitet werden. Eine schwache Lösung u ist dann Entropielösung, wenn

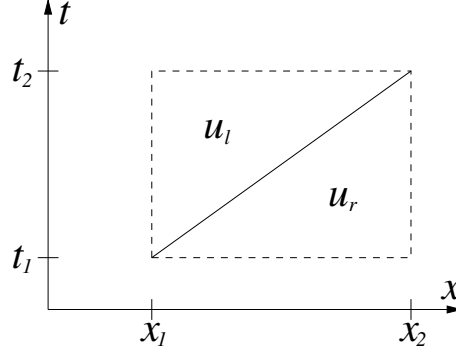
$$\int_0^\infty \int_{-\infty}^\infty \phi_t \eta(u) + \phi_x \psi(u) \, dx \, dt \leq - \int_{-\infty}^\infty \phi(x, 0) \eta(u(x, 0)) \, dx \quad (2.28)$$

für alle $\phi \in C_0^1(\mathbb{R} \times \mathbb{R})$ mit $\phi(x, t) \geq 0$ gilt. Diese Bedingung ist häufig leichter zu verifizieren als (2.27). Diese Entropiebedingung ist jedoch nicht dazu gedacht, sie für konkrete Lösungen nachzuprüfen, was extrem aufwändig wäre. Stattdessen dienen sie als Kriterien für numerische Verfahren. Unter gewissen Voraussetzungen lässt sich nachweisen, dass ein konvergentes Verfahren nur Entropielösungen liefert. Dies ist die Eigenschaft, an der wir in den Anwendungen interessiert sind.

Bemerkung: Die Forderung einer Entropiebedingung führt nicht immer auf eine eindeutige schwache Lösung. In manchen Fällen können mehrere Entropielösungen existieren.

Beispiel

Wir betrachten die Burgers Gleichung (2.5). Eine Entropiefunktion ist $\eta(u) = u^2$ und der zugehörige Entropiefluss lautet $\psi(u) = \frac{2}{3}u^3$. Wir verwenden den folgenden Integrationsbereich mit einer Unstetigkeit bzw. Schock.



Sei $\Delta x = x_2 - x_1$ und $\Delta t = t_2 - t_1$. Nach der Rankine-Hugoniot Sprungbedingung (2.18) gilt für die Schockgeschwindigkeit $s = \frac{\Delta x}{\Delta t} = \frac{1}{2}(u_l + u_r)$. Elementare Rechnungen führen dann auf

$$\begin{aligned} & \int_{x_1}^{x_2} \eta(u(x, t_2)) - \eta(u(x, t_1)) \, dx + \int_{t_1}^{t_2} \psi(u(x_2, t)) - \psi(u(x_1, t)) \, dt \\ &= \Delta t \frac{1}{6} (u_r - u_l)^3. \end{aligned}$$

Für eine Entropielösung gilt somit notwendigerweise $\Delta t \frac{1}{6} (u_r - u_l)^3 \leq 0$, d.h. $u_l \geq u_r$. Mit $u_l \neq u_r$ ist dies äquivalent dazu, dass die Charakteristiken in den Schock hineinlaufen.

2.6 Beispiel für nichtlineare Gleichung: Verkehrsfluss

In diesem Abschnitt soll am Beispiel von einfachen *Verkehrsflüssen* die hergeleitete Theorie angewendet werden und physikalische Interpretationen gegeben werden.

Es sei ρ die Dichte von etwa gleichgroßen Fahrzeugen auf einer geraden Autobahn in der Einheit Anzahl pro Kilometer. Damit hängt $\rho(x, t)$ vom eindimensionalen Ort x und der Zeit t ab. Es gilt $0 \leq \rho \leq \rho_{\max}$, wobei $\rho_{\max}$ die maximale Dichte ist, bei der die Autos Stoßstange an Stoßstange stehen. Die Anzahl der Fahrzeuge auf der Autobahn bleibt erhalten. Ist $v(x, t)$ die Geschwindigkeit der Autos zur Zeit t am Ort x , so gilt vollkommen analog zu (1.2) die Erhaltungsgleichung

$$\rho_t + (\rho v)_x = 0. \quad (2.29)$$

Es ist nun sinnvoll, die Geschwindigkeit als eine Funktion der Dichte anzunehmen. Jeder Fahrer würde mit der höchsten zulässigen Geschwindigkeit $v_{\max}$ fahren, falls die Straße leer ($\rho = 0$) ist. Durch Zunahme des Verkehrs, d.h. Anstieg von ρ , muss jedoch die Geschwindigkeit gesenkt werden. Im Stau ($\rho = \rho_{\max}$) gilt $v = 0$. Daher machen wir für die Geschwindigkeit den einfachen Ansatz mit linearer Interpolation

$$v(\rho) = v_{\max} \left(1 - \frac{\rho}{\rho_{\max}} \right).$$

Die zugehörige Flussfunktion resultiert damit zu

$$f(\rho) = \rho v_{\max} \left(1 - \frac{\rho}{\rho_{\max}} \right).$$

Für die Geschwindigkeit des Informationstransports (Steigung der Charakteristiken) folgt

$$f'(\rho) = v_{\max} \left(1 - 2 \frac{\rho}{\rho_{\max}} \right).$$

Diese Geschwindigkeit kann auch negativ sein, d.h. obwohl alle Fahrzeuge in die gleiche Richtung fahren transportiert sich Information dann in entgegengesetzter Richtung. Wegen $f''(\rho) = -2v_{\max}/\rho_{\max} < 0$ ist die Flussfunktion konkav.

Die Trajektorien $x(t)$ der einzelnen Fahrzeuge, d.h. die Wegbahnen mit der Zeit, ergeben sich aus der Gleichung

$$\frac{dx}{dt} = v(\rho(x, t)) \quad \text{bzw.} \quad x(t) = x_0 + \int_{t_0}^t v(\rho(x(\tau), \tau)) \, d\tau \quad (2.30)$$

mit Startwerten t_0, x_0 . Ist eine Lösung ρ von (2.29) bestimmt, so können die Trajektorien über (2.30) berechnet werden. Man beachte, dass die Lösung ρ unstetig sein kann, die Wegbahnen jedoch stetig sind.

Wir betrachten das Riemann-Problem

$$\rho(x, 0) = \begin{cases} \rho_l & \text{für } x < 0 \\ \rho_r & \text{für } x > 0 \end{cases} \quad \text{mit } 0 \leq \rho_l, \rho_r \leq \rho_{\max}.$$

Nach der Rankine-Hugoniot Sprungbedingung (2.18) muss eine Unstetigkeit die Schockgeschwindigkeit

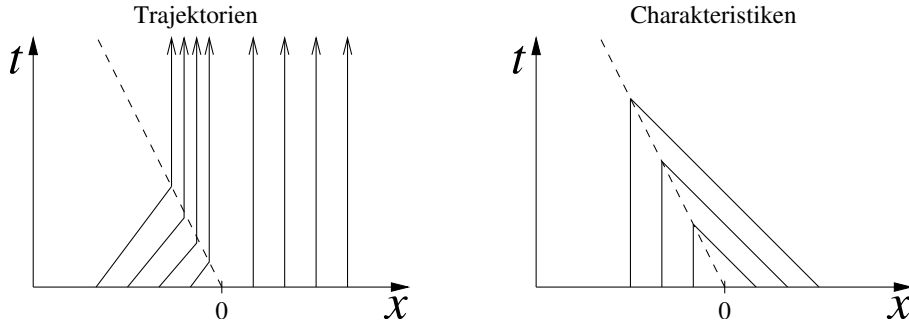
$$s = \frac{f(\rho_l) - f(\rho_r)}{\rho_l - \rho_r} = v_{\max} \left(1 - \frac{\rho_l + \rho_r}{\rho_{\max}} \right) \quad (2.31)$$

besitzen. Die Entropiebedingung (2.21) verlangt $f'(\rho_l) > f'(\rho_r)$, wodurch wegen $f'' < 0$ dann $\rho_l < \rho_r$ gelten muss, damit eine Schockwelle hier physikalisch sinnvoll ist.

1. Fallbeispiel: $\rho_r = \rho_{\max}, \rho_l < \rho_{\max}$

Hier liegt auf der rechten Seite ein Stau vor. Wegen $\rho_l < \rho_r$ ist hier eine Schockwelle eine schwache Lösung, welche die Entropiebedingung erfüllt, d.h. physikalisch sinnvoll. Die Schockgeschwindigkeit ist nach (2.31) negativ. Fahrer auf der linken Seite, die auf den Stau zukommen, müssen sofort ihre Geschwindigkeit auf null verringern. In der Realität dauert das Abbremsen eine kurze Zeit. Dieser Effekt ist im Modell hier nicht vorhanden, was einer mathematischen Idealisierung entspricht. Die Schockwelle, die die Ausbreitung des Staus beschreibt, ist jedoch stabil. Die genaue Struktur, wie die Fahrzeuge die Schockwelle erreichen, stellt hier nur einen kleinen Teil der Lösung dar und ist daher uninteressant.

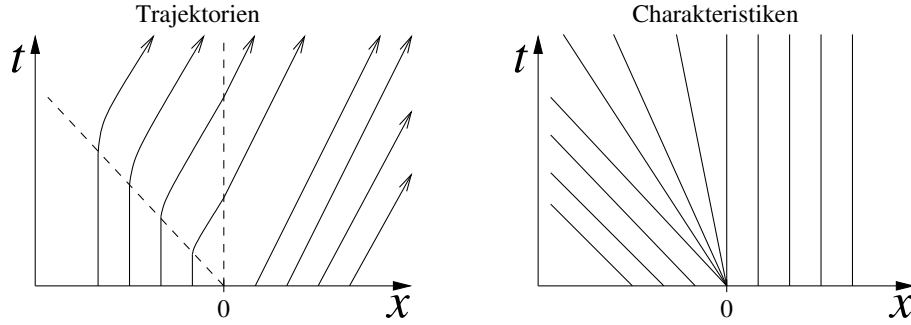
Die folgende Abbildung zeigt die Trajektorien der Fahrzeuge und die Charakteristiken im Spezialfall $\rho_l = \frac{1}{2}\rho_{\max}$.



2. Fallbeispiel: $\rho_l = \rho_{\max}, \rho_r < \rho_{\max}$

Diese Situation beschreibt die Auflösung eines Staus. Auf der linken Seite der Straße stauen sich die Autos bei $t = 0$, während auf der rechten Seite noch Freiraum ist. Nach der obigen Entropiebedingung ist hier ein Verdünnungsschock nicht sinnvoll. Die Autos beschleunigen nicht sofort von Geschwindigkeit null auf den nun möglichen Wert. Erst wenn Autos vor einem Fahrer sich langsam in Bewegung setzen, fährt dieser selbst los. Die Phase der Beschleunigung stellt einen wesentlichen Teil der Lösung dar und kann deshalb nicht vernachlässigt werden. Die korrekte schwache Lösung ist daher eine Verdünnungswelle.

Die Abbildung zeigt wieder die Trajektorien der Fahrzeuge und die Charakteristiken, nun im Spezialfall $\rho_r = \frac{1}{2}\rho_{\max}$.



Die beiden Fallbeispiele zeigen, dass die Geschwindigkeit der Fahrzeuge nicht mit der Geschwindigkeit des Informationstransports über die Charakteristiken identisch ist. Analog gilt dies in der Gasdynamik für die Geschwindigkeit der Moleküle. Die Charakteristiken beschreiben den Transport einer Dichteverteilung. Das einzelne Teilchen kann dabei durch die Dichteverteilung laufen und transportiert selbst keine Information.

2.7 Lineare Systeme

In diesem Abschnitt diskutieren wir das Verhalten von linearen Systemen aus Erhaltungsgleichungen. Insbesondere kann zu einem Anfangswertproblem stets eine analytische Lösung bestimmt werden. Dies erleichtert dann die Behandlung von Riemann-Problemen für lineare Systeme. Lineare Systeme spielen ferner eine Rolle bei der Linearisierung von nichtlinearen Systemen aus Erhaltungsgleichungen.

Analytische Lösung linearer Systeme

Wir betrachten zu einem linearen System das Cauchy-Problem

$$\begin{aligned} u_t + Au_x &= 0 \\ u(x, 0) &= u_0(x) \end{aligned} \quad (2.32)$$

mit der Lösung $u : \mathbb{R} \times \mathbb{R}_0^+ \rightarrow \mathbb{R}^k$ und der konstanten Matrix $A \in \mathbb{R}^{k \times k}$. Die Anfangswerte u_0 seien (zunächst) stetig vorgegeben. Es ist (2.32) ein System aus Erhaltungsgleichungen mit Flussfunktion $f(u) = Au$. Das System sei hyperbolisch gemäß Definition 1.6, d.h. die Matrix A sei reell diagonalisierbar. Damit haben wir die Zerlegung

$$A = RDR^{-1} \quad \text{mit } D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k) \quad \text{und } R = (r_1 | r_2 | \dots | r_k), \quad (2.33)$$

wobei $\lambda_j \in \mathbb{R}$ die Eigenwerte und $r_j \in \mathbb{R}^k$ die korrespondierenden Eigenvektoren sind. Durch lineare Transformation auf charakteristische Variablen

$$v(x, t) = R^{-1}u(x, t) \quad (2.34)$$

kann das System umgeformt werden zu

$$v_t + Dv_x = 0.$$

Da D Diagonalmatrix ist, wird dadurch das System entkoppelt in die skalaren Dgln.

$$(v_j)_t + \lambda_j(v_j)_x = 0 \quad \text{für } j = 1, \dots, k$$

mit den Komponenten $v = (v_1, \dots, v_k)^\top$. Jede dieser Gleichungen für sich ist eine lineare Advektionsgleichung (2.3) mit der Lösung

$$v_j(x, t) = v_j(x - \lambda_j t, 0).$$

Laut (2.34) transformieren sich die Anfangswerte über

$$v(x, 0) = R^{-1}u_0(x), \quad (2.35)$$

so dass die Gleichungen sofort gelöst werden können. Die gesuchte Lösung erhalten wir dann aus

$$u(x, t) = Rv(x, t) = \sum_{j=1}^k v_j(x, t)r_j = \sum_{j=1}^k v_j(x - \lambda_j t, 0)r_j. \quad (2.36)$$

Die Lösung u in einem Punkt (x^*, t^*) hängt damit von den Anfangswerten bei $t = 0$ in höchstens k Ortspunkten ab. Der zugehörige Abhängigkeitsbereich lautet

$$\mathcal{D}(x^*, t^*) = \{x = x^* - \lambda_j t^* : j = 1, \dots, k\}. \quad (2.37)$$

Die Kurven $x(t) = \xi + \lambda_j t$ bzw. $x'(t) = \lambda_j$ nennt man *Charakteristiken der j -ten Familie* oder *j -Charakteristiken*. Ist das System strikt hyperbolisch gemäß Definition 1.6 (Eigenwerte λ_j paarweise verschieden), dann liegen genau k unterschiedliche Familien vor. Die j -Charakteristiken repräsentieren jeweils Scharen aus parallelen Geraden im Definitionsbereich. Im Gegensatz zum skalaren Fall ist die Lösung u nun im allgemeinen nicht mehr konstant entlang irgendeiner Charakteristik. Konstant entlang einer j -Charakteristik ist lediglich der Koeffizient v_j , d.h. ein gewisser Lösungsanteil.

Aus den Beziehungen (2.35) und (2.36) folgt, dass zu glatten Anfangswerten u_0 dann die Lösung u ebenfalls glatt ist. Sind die Anfangswerte zumindest in Umgebungen der Ortspunkte $x^* - \lambda_j t^*$ für alle j glatt, dann ist auch die Lösung in einer Umgebung von (x^*, t^*) glatt. Folglich können Singularitäten in den Anfangswerten nur entlang der Charakteristiken transportiert werden. Dies entspricht vollkommen dem skalaren linearen Fall.

Riemann-Problem für lineare Systeme

Insbesondere kann auch das Riemann-Problem (1.7) mit $u_l, u_r \in \mathbb{R}^k$ für ein lineares hyperbolisches System (2.32) analytisch gelöst werden. Die reellen Eigenwerte der Matrix A seien ihrer Größe nach geordnet

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{k-1} \leq \lambda_k.$$

Die Anfangswerte können in die korrespondierende Eigenvektorbasis entwickelt werden

$$u_l = \sum_{j=1}^k \alpha_j r_j \quad \text{und} \quad u_r = \sum_{j=1}^k \beta_j r_j.$$

Somit folgt mit $v(x, 0) = R^{-1}u_l$ für $x < 0$ und $v(x, 0) = R^{-1}u_r$ für $x > 0$

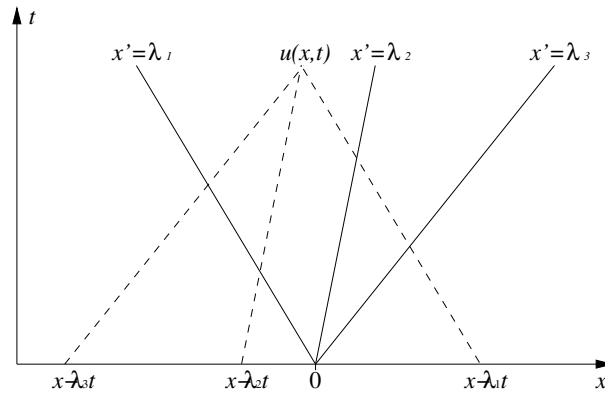
$$v_j(x, 0) = \begin{cases} \alpha_j & \text{für } x < 0 \\ \beta_j & \text{für } x > 0 \end{cases} \Rightarrow v_j(x, t) = \begin{cases} \alpha_j & \text{für } x - \lambda_j t < 0 \\ \beta_j & \text{für } x - \lambda_j t > 0. \end{cases}$$

Es sei $J(x, t)$ die größte ganze Zahl j mit $x - \lambda_j t > 0$, insbesondere $J(x, t) = 0$ falls die Bedingung für kein λ_j zutrifft. Dann folgt für die Lösung

$$u(x, t) = \sum_{j=1}^{J(x,t)} \beta_j r_j + \sum_{j=J(x,t)+1}^k \alpha_j r_j. \quad (2.38)$$

Alternativ haben wir die Darstellungen

$$u(x, t) = u_l + \sum_{\lambda_j < x/t} (\beta_j - \alpha_j) r_j = u_r - \sum_{\lambda_j > x/t} (\beta_j - \alpha_j) r_j. \quad (2.39)$$



Die Rankine-Hugoniot Sprungbedingung (2.16) fordert für ein lineares System bei einer Unstetigkeit von $\tilde{u}_l$ nach $\tilde{u}_r$

$$A(\tilde{u}_l - \tilde{u}_r) = s(\tilde{u}_l - \tilde{u}_r),$$

d.h. der Sprung ist ein Eigenvektor von A und die Sprunggeschwindigkeit der zugehörige Eigenwert. Ist das System strikt hyperbolisch, dann liegen in (2.38) genau k Sprünge mit den Differenzen $(\beta_j - \alpha_j)r_j$ vor und es gilt

$$A(\beta_j - \alpha_j)r_j = \lambda_j(\beta_j - \alpha_j)r_j \quad \text{für } j = 1, \dots, k.$$

Bei einem mehrfachen Eigenwert fallen mehrere Sprünge zusammen. O.E.d.A. sei $\lambda_1 = \lambda_2 = \dots = \lambda_L$ ($L \geq 2$). Es folgt auch hier für die Differenz zwischen der Lösung links und rechts der Unstetigkeit

$$A \left(\sum_{j=1}^L (\beta_j - \alpha_j)r_j \right) = \sum_{j=1}^L \lambda_j(\beta_j - \alpha_j)r_j = \lambda_1 \left(\sum_{j=1}^L (\beta_j - \alpha_j)r_j \right).$$

Die Rankine-Hugoniot Sprungbedingung ist also durch die Konstruktion (2.38) stets erfüllt. Die Lösung des Riemann-Problems kann man interpretieren als eine Zerlegung des Anfangssprungs in eine Summe von Sprüngen

$$u_r - u_l = (\beta_1 - \alpha_1)r_1 + \dots + (\beta_k - \alpha_k)r_k,$$

so dass jeder Teilsprung mit der korrekten Geschwindigkeit λ_j nach der Rankine-Hugoniot Sprungbedingung weitertransportiert wird. Mit der Definition

$$w_j(x, t) = \begin{cases} 0 & \text{für } x - \lambda_j t < 0 \\ (\beta_j - \alpha_j)r_j & \text{für } x - \lambda_j t > 0 \end{cases} \quad \text{für } j = 1, \dots, k$$

erfüllen diese einzelnen Schockwellen die Rankine-Hugoniot Sprungbedingung, sind also schwache Lösungen. Mit (2.39) folgt

$$u(x, t) = u_l + \sum_{j=1}^k w_j(x, t).$$

Diese Summe ist ebenfalls eine schwache Lösung, da für die zu (2.32) korrespondierende Integralform wegen der Linearität das Superpositionsprinzip gilt.

Ein Spezialfall entsteht, wenn die Unstetigkeit $u_r - u_l$ bereits ein Eigenvektor von A ist. Dann gilt $u_r - u_l = (\beta_i - \alpha_i)r_i$ für ein i und $\alpha_j = \beta_j$ für $j \neq i$. Dadurch wird die Unstetigkeit ausschließlich mit der Geschwindigkeit λ_i weitertransportiert, während die anderen Charakteristikenfamilien nur Sprünge der Höhe null befördern.

Linearisierung von nichtlinearen Systemen

Zu dem nichtlinearen System aus Erhaltungsgleichungen $u_t + f(u)_x = 0$ mit der Lösung $u : \mathbb{R} \times \mathbb{R}_0^+ \rightarrow \mathbb{R}^k$ und $f \in C^1(\mathbb{R}^k, \mathbb{R}^k)$ lautet die quasilineare Form

$$u_t + A(u)u_x = 0, \quad (2.40)$$

wobei $A(u)$ die Funktionalmatrix von $f(u)$ bezeichnet. Das nichtlineare System sei hyperbolisch gemäß Definition 1.6, d.h. die Matrix $A(u)$ ist stets reell diagonalisierbar.

Im strikt hyperbolischen Fall sind die Eigenwerte $\lambda_j(u)$ von $A(u)$ paarweise verschieden und man kann wie im linearen Fall j Familien aus Charakteristiken definieren über

$$x'(t) = \lambda_j(u(x(t), t)) \quad \text{für } j = 1, \dots, k.$$

Da die Eigenwerte nun aber von der Lösung abhängen, können wir die Charakteristiken nicht a priori bestimmen. Stattdessen entsteht ein kompliziertes gekoppeltes System, da Charakteristiken unterschiedlicher Familien über die Lösung u zusammenhängen. Dadurch wird ein globaler Lösungsansatz über die Charakteristiken im nichtlinearen Fall ineffizient.

Dennoch liefern die Charakteristiken wertvolle Informationen, wie sich das System lokal verhält. Dafür betrachten wir einen konstanten Zustand $\hat{u} \in \mathbb{R}^k$, etwa einen gewissen Mittelwert. Funktionen in der Umgebung dieses Zustands sind gegeben durch

$$u(x, t) = \hat{u} + \varepsilon \tilde{u}(x, t)$$

mit fester Funktion $\tilde{u}$ und kleinem $\varepsilon \in \mathbb{R} \setminus \{0\}$. Gilt $f \in C^2(\mathbb{R}^k, \mathbb{R}^k)$, dann erhalten wir die Aussagen

$$\begin{aligned} u_t &= \varepsilon \tilde{u}_t \\ u_x &= \varepsilon \tilde{u}_x \\ A(u) &= A(\hat{u}) + \mathcal{O}(\|u - \hat{u}\|) = A(\hat{u}) + \mathcal{O}(\varepsilon). \end{aligned}$$

Somit folgt aus (2.40) die Differentialgleichung

$$\begin{aligned} \varepsilon \tilde{u}_t + A(\hat{u})\varepsilon \tilde{u}_x &= \mathcal{O}(\varepsilon^2) \\ \tilde{u}_t + A(\hat{u})\tilde{u}_x &= \mathcal{O}(\varepsilon). \end{aligned}$$

Im Grenzfall $\varepsilon \rightarrow 0$ entsteht ein lineares hyperbolisches System für die Funktion $\tilde{u}$. Kleine Störungen um den Zustand $\hat{u}$ breiten sich demnach näherungsweise entlang der Charakteristiken dieses linearen Systems aus.

Diese Eigenschaft spiegelt sich auch in der Ausbreitung von Schallwellen in der Luft gemäß den Eulerschen Gasgleichungen (1.8) wider. Unter gewisser Vorgabe

für den Druck besitzt die Funktionalmatrix der Flussfunktion die drei Eigenwerte $\hat{v}, \hat{v} \pm \hat{c}$. Der Parameter $\hat{c}$ hängt vom Zustand $\hat{u} = (\hat{\rho}, \hat{\rho}\hat{v}, \hat{E})^\top$ ab. Der Schall stellt kleine Dichteschwankungen um einen Mittelwert $\hat{u}$ dar. Die Schallwellen breiten sich dann mit der Geschwindigkeit $\hat{c}$ relativ zur Geschwindigkeit $\hat{v}$ des Gases aus.

Bei linearen Systemen können sich Singularitäten nur entlang der Charakteristiken ausbreiten. Für nichtlineare Systeme gilt dies im allgemeinen nicht. Stattdessen muss in Unstetigkeiten die Rankine-Hugoniot Sprungbedingung (2.16) lokal erfüllt sein. Betrachtet man aber einen kleinen Sprung $\|u_r - u_l\| \equiv \varepsilon$, so liefert Taylor-Entwicklung

$$f(u_r) = f(u_l) + A(u_l)(u_r - u_l) + \mathcal{O}(\varepsilon^2)$$

und somit nach (2.16)

$$A(u_l) \frac{1}{\varepsilon}(u_r - u_l) = s \frac{1}{\varepsilon}(u_r - u_l) + \mathcal{O}(\varepsilon).$$

Für $\varepsilon \rightarrow 0$ geht damit der normierte Sprung in einen Eigenvektor der Matrix $A(u_l)$ über und s erreicht den zugehörigen Eigenwert.

Unstetigkeiten in den Anfangswerten bei linearen Systemen führen ausschließlich zu Schockwellen. Demgegenüber entstehen bei nichtlinearen Systemen auch zum Teil Verdünnungswellen in der physikalisch sinnvollen schwachen Lösung. Wird ein nichtlineares System über eine Linearisierung gelöst, so sind daher gegebenenfalls Nachbesserung notwendig, um eine gute Näherung für eine Entropielösung zu erhalten.

2.8 Nichtlineare Systeme

In diesem Abschnitt wird ein Teil der Lösungstheorie für nichtlineare Systeme aus Erhaltungsgleichungen besprochen. Dabei wird ein strikt hyperbolisches System gemäß Definition 1.6 vorausgesetzt. Es seien $\lambda_1(u) < \lambda_2(u) < \dots < \lambda_k(u)$ die Eigenwerte und $r_1(u), r_2(u), \dots, r_k(u)$ die zugehörigen Eigenvektoren der Funktionalmatrix der Flussfunktion.

Schockwellen und Hugoniot-Locus

Für einen Sprung zwischen konstanten Zuständen $\hat{u}, \tilde{u} \in \mathbb{R}^k$ lautet die Rankine-Hugoniot-Sprungbedingung

$$f(\tilde{u}) - f(\hat{u}) = s(\tilde{u} - \hat{u})$$

mit der Schockgeschwindigkeit $s \in \mathbb{R}$. Im linearen Fall ist der Sprung $\tilde{u} - \hat{u}$ ein Vielfaches eines Eigenvektors r_i zum Eigenwert $s = \lambda_i$. Ist $\hat{u}$ fest gewählt, dann

sind Sprünge möglich zu den Zuständen

$$\begin{aligned}\tilde{u}_i(\xi) &= \hat{u} + \xi r_i \\ s_i(\xi) &= \lambda_i\end{aligned}$$

mit Parameter $\xi \in \mathbb{R}$ und einem $i \in \{1, \dots, k\}$.

In nichtlinearen Fall ergibt sich analog

$$f(\tilde{u}_i(\xi)) - f(\hat{u}) = s_i(\xi)(\tilde{u}_i(\xi) - \hat{u}) \quad (2.41)$$

für ein $i \in \{1, \dots, k\}$. Differentiation nach dem Parameter ξ zeigt

$$\frac{\partial f}{\partial u}(\tilde{u}_i(\xi))\tilde{u}'_i(\xi) = s'_i(\xi)(\tilde{u}_i(\xi) - \hat{u}) + s_i(\xi)\tilde{u}'_i(\xi).$$

Die Auswertung bei $\xi = 0$ ergibt

$$\frac{\partial f}{\partial u}(\hat{u})\tilde{u}'_i(0) = s_i(0)\tilde{u}'_i(0).$$

Also ist $\tilde{u}'_i(0) \neq 0$ ein Vielfaches eines Eigenvektors $r_i(\hat{u})$ und $s_i(0) = \lambda_i(\hat{u})$. Die Kurve $\tilde{u}_i(\xi)$ ist somit tangential zu $r_i(\hat{u})$ bei $\hat{u}$.

Der Satz über implizite Funktionen liefert die lokale Existenz von k derartigen Kurven in einer Umgebung von $\hat{u}$. Zudem sind die Funktionen $\tilde{u}_i$ und s_i stetig differenzierbar für $i = 1, \dots, k$.

Definition 2.9 *Sei ein strikt hyperbolisches, nichtlineares System gegeben. Zu $\hat{u} \in \mathbb{R}^k$ lautet die Menge aller Punkte $\tilde{u} \in \mathbb{R}^k$, welche über eine Kurve $\tilde{u}_i(\xi)$ mit der Bedingung (2.41) erreicht werden können, der Hugoniot-Locus zu $\hat{u}$.*

Beispiel: Isothermische Gasgleichungen

Die isothermischen Gasgleichungen lauten

$$\begin{aligned}\rho_t + m_x &= 0 \\ m_t + \left(\frac{m^2}{\rho} + a^2 \rho\right)_x &= 0\end{aligned} \quad (2.42)$$

mit der Lösung $u = (\rho, m)^\top$ und der Schallgeschwindigkeit $a > 0$. Die Funktionalmatrix der Flussfunktion lautet

$$\frac{\partial f}{\partial u} = \begin{pmatrix} 0 & 1 \\ a^2 - \frac{m^2}{\rho^2} & 2\frac{m}{\rho} \end{pmatrix}.$$

Die Eigenwerte der Matrix sind

$$\lambda_1(u) = \frac{m}{\rho} - a \quad \text{und} \quad \lambda_2(u) = \frac{m}{\rho} + a \quad (2.43)$$

mit den Eigenvektoren

$$r_1(u) = \begin{pmatrix} 1 \\ \frac{m}{\rho} - a \end{pmatrix} \quad \text{und} \quad r_2(u) = \begin{pmatrix} 1 \\ \frac{m}{\rho} + a \end{pmatrix}. \quad (2.44)$$

Für zwei Zustände $\hat{u}, \tilde{u}$ liefert die Rankine-Hugoniot-Sprungbedingung

$$\begin{aligned} \tilde{m} - \hat{m} &= s(\tilde{\rho} - \hat{\rho}) \\ \left(\frac{\tilde{m}^2}{\tilde{\rho}} + a^2 \tilde{\rho}\right) - \left(\frac{\hat{m}^2}{\hat{\rho}} + a^2 \hat{\rho}\right) &= s(\tilde{m} - \hat{m}). \end{aligned}$$

Für $\hat{\rho}, \hat{m}$ gegeben sind die zwei Gleichungen für die Unbekannten $\tilde{\rho}, \tilde{m}, s$. Mit $\tilde{\rho}$ als Freiheitsgrad ergeben sich die beiden Lösungen

$$\begin{aligned} \tilde{m} &= \frac{\hat{m}\tilde{\rho}}{\hat{\rho}} \pm a\sqrt{\frac{\tilde{\rho}}{\hat{\rho}}}(\tilde{\rho} - \hat{\rho}) \\ s &= \frac{\hat{m}}{\hat{\rho}} \pm a\sqrt{\frac{\tilde{\rho}}{\hat{\rho}}}. \end{aligned}$$

Wir parametrisieren $\tilde{\rho}$ als ein Vielfaches von $\hat{\rho}$ durch

$$\tilde{\rho}_i(\xi) = \hat{\rho}(1 + \xi) \quad \text{für } i = 1, 2$$

mit Parameter $\xi \in \mathbb{R}$. Dadurch folgt

$$\tilde{u}_1(\xi) = \hat{u} + \xi \begin{pmatrix} \hat{\rho} \\ \hat{m} - a\hat{\rho}\sqrt{1 + \xi} \end{pmatrix}, \quad s_1(\xi) = \frac{\hat{m}}{\hat{\rho}} - a\sqrt{1 + \xi}$$

und

$$\tilde{u}_2(\xi) = \hat{u} + \xi \begin{pmatrix} \hat{\rho} \\ \hat{m} + a\hat{\rho}\sqrt{1 + \xi} \end{pmatrix}, \quad s_2(\xi) = \frac{\hat{m}}{\hat{\rho}} + a\sqrt{1 + \xi}.$$

Im Fall $\xi = 0$ erhalten wir

$$\begin{aligned} \tilde{u}'_i(0) &= \hat{\rho} r_i(\hat{u}) \\ s_i(0) &= \lambda_i(\hat{u}) \end{aligned}$$

für $i = 1, 2$ mit den Eigenwerten (2.43) und den Eigenvektoren (2.44).

Verdünnungswellen und Integralkurven

Als schwache Lösungen eines strikt hyperbolischen, nichtlinearen Systems sind wir nun an Verdünnungswellen der Form

$$u(x, t) = \begin{cases} u_l & \text{für } x < \xi_1 t \\ v(\frac{x}{t}) & \text{für } \xi_1 t \leq x \leq \xi_2 t \\ u_r & \text{für } x > \xi_2 t \end{cases} \quad (2.45)$$

interessiert. Diese Funktion soll global stetig sein. Dadurch muss $v(\xi_1) = u_l$ und $v(\xi_2) = u_r$ gelten.

Definition 2.10 Sei ein (stetiges) Vektorfeld $r : \mathbb{R}^k \rightarrow \mathbb{R}^k$ gegeben. Eine zugehörige Integralkurve ist eine stetig differenzierbare Abbildung $u : (\xi_{\min}, \xi_{\max}) \rightarrow \mathbb{R}^k$, so dass $u'(\xi) \neq 0$ und $u'(\xi) = \alpha(\xi)r(u(\xi))$ mit einem $\alpha(\xi) \in \mathbb{R}$ für alle ξ gilt.

Eine Integralkurve liegt also tangential am Vektorfeld an. Wir betrachten nun die Eigenvektoren $r_1(u), \dots, r_k(u)$ jeweils als Vektorfeld. Da das System strikt hyperbolisch ist und die Flussfunktion stetig differenzierbar vorliegt, sind die Eigenvektoren jeweils stetig in u . Für den i -ten Eigenvektor erhalten wir ein System aus gewöhnlichen Dgln.

$$u'_i(\xi) = \alpha(\xi)r_i(u_i(\xi))$$

mit skalaren Vielfachen $\alpha(\xi)$.

Nun betrachten wir die Funktion $u(x, t) = v(\frac{x}{t})$ aus (2.45). Mit

$$u_t(x, t) = -\frac{x}{t^2}v'(\frac{x}{t}) \quad \text{und} \quad u_x(x, t) = \frac{1}{t}v'(\frac{x}{t})$$

liefert die quasilineare Form $u_t + \frac{\partial f}{\partial u}u_x = 0$ hier

$$-\frac{x}{t^2}v'(\frac{x}{t}) + \frac{1}{t}\frac{\partial f}{\partial u}(v(\frac{x}{t}))v'(\frac{x}{t}) = 0.$$

Die Substitution $\xi = \frac{x}{t}$ führt auf

$$\frac{\partial f}{\partial u}(v(\xi))v'(\xi) = \xi v'(\xi). \quad (2.46)$$

Eine konstante Funktion $v \equiv v_0$ wäre eine Lösung. Jedoch benötigen wir an einen stetigen Übergang von u_l zu u_r . Die Bedingung (2.46) bedeutet, dass v' ein Eigenvektor der Funktionalmatrix zum Eigenwert ξ ist. Also muss mit einem $i \in \{1, \dots, k\}$

$$v'(\xi) = \alpha(\xi)r_i(v(\xi)) \quad (2.47)$$

mit einem Vielfachen $\alpha(\xi)$ und

$$\xi = \lambda_i(v(\xi)) \quad (2.48)$$

gelten. Somit stellt v eine Integralkurve zum Vektorfeld r_i dar. Jedoch ist die Parametrisierung dieser Kurve nun von Bedeutung. Differentiation der Gleichung (2.48) mit der Kettenregel liefert

$$1 = \nabla \lambda_i(v(\xi))v'(\xi) = \alpha(\xi) \nabla \lambda_i(v(\xi))r_i(v(\xi))$$

und somit

$$\alpha(\xi) = \frac{1}{\nabla \lambda_i(v(\xi))r_i(v(\xi))} \quad (2.49)$$

unter der Annahme, dass der Nenner ungleich null ist. Wir erhalten ein System aus gewöhnlichen Dgln.

$$v'(\xi) = \frac{1}{\nabla \lambda_i(v(\xi)) r_i(v(\xi))} r_i(v(\xi)) \quad (2.50)$$

für $\xi_1 \leq \xi \leq \xi_2$ mit den Anfangswerten

$$v(\xi_1) = u_l.$$

Aus (2.48) und der Stetigkeit von u folgt $\xi_1 = \lambda_i(u_l)$ und $\xi_2 = \lambda_i(u_r)$. Notwendiger für eine Verdünnungswelle ist daher $\lambda_i(u_l) < \lambda_i(u_r)$.

Analog zum Hugoniot-Locus gibt es eine Menge aus Zuständen $\tilde{u}$, die von $\hat{u} = u_l$ aus durch eine Verdünnungswelle erreicht werden können.

Beispiel: Isothermische Gasgleichungen

Wir betrachten wieder das nichtlineare System (2.42). Sei $\hat{\rho}, \hat{m}$ gegeben. Wir verwenden den ersten Eigenwert aus (2.43). Sei $\xi_1 = \lambda_1(\hat{u}) = \frac{\hat{m}}{\hat{\rho}} - a$. Es folgt mit dem Eigenvektor aus (2.44)

$$\nabla \lambda_1(u) r_1(u) = \left(-\frac{m}{\rho^2} \quad \frac{1}{\rho} \right) \left(\frac{m}{\rho} - a \right) = -\frac{a}{\rho}.$$

Wir erhalten die gewöhnlichen Dgln. (2.50)

$$\begin{aligned} \rho'(\xi) &= -\frac{\rho(\xi)}{a}, & \rho(\xi_1) &= \hat{\rho}, \\ m'(\xi) &= \rho(\xi) - \frac{m(\xi)}{a}, & m(\xi_1) &= \hat{m}. \end{aligned}$$

Die erste Gleichung liefert direkt

$$\rho(\xi) = \hat{\rho} e^{-(\xi - \xi_1)/a}.$$

Einsetzen in die zweite Gleichung ergibt

$$m'(\xi) = \hat{\rho} e^{-(\xi - \xi_1)/a} - \frac{m(\xi)}{a}, \quad m(\xi_1) = \hat{m}.$$

Mit Variation der Konstanten folgt die Lösung

$$m(\xi) = (\hat{\rho}(\xi - \xi_1) + \hat{m}) e^{-(\xi - \xi_1)/a} = \hat{\rho}(\xi + a) e^{-(\xi - \xi_1)/a}.$$

Damit sind ist die Integralkurve im Phasenraum bestimmt. Desweiteren kann m als Funktion von ρ dargestellt werden (Elimination der Terme mit ξ, ξ_1)

$$m(\rho) = \rho \frac{\hat{m}}{\hat{\rho}} - a \rho \log \left(\frac{\rho}{\hat{\rho}} \right).$$

Analog kann eine Integralkurve bezüglich $\lambda_2(u), r_2(u)$ hergeleitet werden.

Lineare Degeneriertheit

Bei einem (strikt) hyperbolischen, nichtlinearen System ist folgender Spezialfall möglich.

Definition 2.11 *Bei einem hyperbolischen, nichtlinearen System heißt die i -te Charakteristikenfamilie linear degeneriert (nicht echt nichtlinear), falls für den zugehörigen Eigenwert und Eigenvektor*

$$\nabla \lambda_i(u) r_i(u) = 0 \quad (2.51)$$

für alle (relevanten) u gilt.

Im Falle der linearen Degeneriertheit ist somit der Eigenwert jeweils konstant entlang einer Integralkurve. Der Nenner in (2.49) ist dann null. Also kommen Verdünnungswellen bei einer solchen Charakteristikenfamilie nicht in Frage. Eine Unstetigkeit bzw. ein Sprung bezüglich einer linear degenerierten Charakteristikenfamilie heißt *Kontaktunstetigkeit*. Falls die Eigenschaft (2.51) nicht vorliegt, dann spricht man von einer *echt nichtlinearen* Charakteristikenfamilie.

Bei einer einzelnen Erhaltungsgleichung ist $\nabla \lambda(u) = f'(u)$ und $r(u) = 1$. Für nichtlineares f und $f'(u) \neq 0$ ist die Erhaltungsgleichung somit stets echt nichtlinear. Hinreichend dafür ist $f'' > 0$ oder $f'' < 0$. Bei einem hyperbolischen, linearen System wäre $\nabla \lambda_i = 0$ für alle $i = 1, \dots, k$, d.h. jede Charakteristikenfamilie würde das Kriterium (2.51) aus Definition 2.11 erfüllen.

Beispiel: Eulersche Gasgleichungen

Wir betrachten die Eulerschen Gasgleichungen (1.8) aus Abschnitt 1.3. Der zweite Eigenwert ist $\lambda_2(u) = v = \frac{u_2}{u_1}$ aus (1.11) mit dem zugehörigen Eigenvektor $r_2(u) = (1, v, \frac{1}{2}v^2)^\top$. Dadurch folgt

$$\nabla \lambda_2(u) r_2(u) = \left(-\frac{u_2}{u_1^2} \quad \frac{1}{u_1} \quad 0 \right) \begin{pmatrix} 1 \\ \frac{u_2}{u_1} \\ \frac{u_2^2}{2u_1^2} \end{pmatrix} = 0,$$

d.h. die zweite Charakteristikenfamilie ist linear degeneriert.

3 Numerische Methoden für lineare Probleme

In diesem Kapitel befassen wir uns mit der numerischen Lösung von Anfangswertproblemen zu linearen hyperbolischen Systemen. Die Dimension eines Systems wird fortan mit m bezeichnet.

3.1 Diskretisierung auf Gittern

Betrachtet wird ein Cauchy-Problem (2.32) eines linearen hyperbolischen Systems aus m Gleichungen. Wir diskretisieren den Definitionsbereich in der x - t Ebene $\mathbb{R} \times \mathbb{R}^+$ mit einem Gitter zu den konstanten Schrittweiten $h = \Delta x$ und $k = \Delta t$. Hierbei entstehen die Gitterpunkte

$$(x_j, t_n) = (jh, nk) \quad \text{für } j \in \mathbb{Z}, \quad n \in \mathbb{N}_0.$$

Von Bedeutung sind auch die Zwischenstellen

$$x_{j+\frac{1}{2}} = x_j + \frac{h}{2} = \left(j + \frac{1}{2}\right) h.$$

Es sei $u_j^n = u(x_j, t_n)$ die exakte Lösung in den Gitterpunkten. Ziel ist nun, punktweise Näherungen $U_j^n \in \mathbb{R}^m$ für u_j^n zu bestimmen. Eine solche Näherung können wir auch als Approximation für das *Zellenmittel*

$$\bar{u}_j^n = \frac{1}{h} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x, t_n) \, dx \quad (3.1)$$

interpretieren, was bei schwachen Lösungen geeignet ist. Die zur Differentialgleichung korrespondierende Integralgleichung beschreibt nämlich genau die Entwicklung der Zellenmittel (3.1) mit der Zeit. Die Anfangsdaten U_j^0 kann man daher entweder punktweise über $u_0(x_j)$ oder durch die Zellenmittel $\bar{u}_j^0$ definieren. Zum Vergleich mit der exakten Lösung wird aus den diskreten Näherungen eine Funktion definiert über

$$U_k(x, t) = U_j^n \quad \text{für } (x, t) \in [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}) \times [t_n, t_{n+1}). \quad (3.2)$$

Diese Funktion trägt nur den Index k , da häufig das Schrittweitenverhältnis $\frac{k}{h}$ als konstant vorausgesetzt wird. Dadurch ist $k \rightarrow 0$ äquivalent mit $h \rightarrow 0$.

Ein numerisches Verfahren bestimmt nun aus den Anfangsdaten U_j^0 die Näherungen der ersten Zeitschicht U_j^1 . Sukzessive konstruiert ein Einschrittverfahren aus den Daten U_j^n eine Approximation in der nächsten Zeitschicht U_j^{n+1} . Bei einem Mehrschrittverfahren mit $\ell + 1$ Schritten wird aus mehreren alten Zeitschichten

$U_j^{n-\ell}, \dots, U_j^{n-1}, U_j^n$ die neue Näherung U_j^{n+1} gebildet. Im folgenden betrachten wir jedoch nur Einschrittverfahren, da Mehrschrittverfahren spätestens im Falle mehrerer Raumdimensionen zu unhandlich werden.

Beim Cauchy-Problem (2.32) wird die Lösung im gesamten Definitionsbereich $\mathbb{R} \times \mathbb{R}^+$ betrachtet. In einem numerischen Verfahren muss ein begrenzter Bereich ausgewählt werden. Im Ort wird hierzu das Intervall $a \leq x \leq b$ verwendet. An den Rändern dieses Intervalls sind nun geeignete Randbedingungen vorzugeben. Bei periodischen Anfangsdaten mit Periode $b - a$ sind die Randbedingungen

$$u(a, t) = u(b, t) \quad \text{für alle } t \geq 0$$

sinnvoll. Andernfalls können auch Randwerte explizit vorgeschrieben werden. Bei einem Riemann-Problem (1.7) etwa ist die Vorgabe ($a < 0 < b$)

$$u(a, t) = u_l, \quad u(b, t) = u_r \quad \text{für } 0 \leq t \leq T$$

zulässig, falls die Ränder weit genug von der Unstetigkeit entfernt sind, da sich Information nur mit endlicher Geschwindigkeit ausbreitet. Eine korrekte Vorgabe von Randwerten ist jedoch oft a priori nicht möglich. Ein Ausweg besteht darin, gemäß einer Differenzenformel aus den Anfangswerten in $[a, b]$ alle Gitterpunkte der nächsten Zeitschichten zu bestimmen, die ohne Randbedingungen berechnet werden können. Dadurch wird im allgemeinen der berechnete Ortsbereich mit fortschreitender Zeit kleiner, so dass die Lösung nur über eine begrenzte Zeit $0 \leq t \leq T$ möglich ist.

3.2 Herleitung von Verfahren

Es existiert eine hohe Anzahl von anwendbaren Differenzenverfahren für lineare Systeme $u_t + Au_x = 0$. Viele entstehen einfach dadurch, dass die partiellen Ableitungen durch Differenzenformeln ersetzt werden. Beispielsweise resultiert aus der expliziten Euler-Methode in der Zeit mit einer symmetrischen Diskretisierung im Ort die Vorschrift

$$\frac{1}{k} (U_j^{n+1} - U_j^n) + A \frac{1}{2h} (U_{j+1}^n - U_{j-1}^n) = 0 \quad (3.3)$$

und damit die Approximation

$$U_j^{n+1} = U_j^n - \frac{k}{2h} A (U_{j+1}^n - U_{j-1}^n). \quad (3.4)$$

Aus den Näherungen in einer Zeitschicht t_n kann so die nächste Schicht t_{n+1} direkt berechnet werden. Leider ist diese Formel instabil und damit in den Anwendungen nutzlos. Dagegen führt die implizite Euler-Methode auf eine stabile Rekursion, nämlich

$$\frac{1}{k} (U_j^{n+1} - U_j^n) + A \frac{1}{2h} (U_{j+1}^{n+1} - U_{j-1}^{n+1}) = 0. \quad (3.5)$$

Hier entsteht bei N Gitterpunkten im Ort ein lineares Gleichungssystem der Dimension mN für die Näherungen in der nächsten Zeitschicht. Dadurch resultiert ein deutlich höherer Rechenaufwand im Gegensatz zu (3.3).

Bei parabolischen Dgln. (z.B. Wärmeleitungsgleichung) zeigt sich, dass explizite Differenzenverfahren aus Stabilitätsgründen starke Restriktionen an die Größe der Schrittweiten verursachen. Bei impliziten Verfahren liegen diese Einschränkungen nicht mehr vor. Für parabolische Dgln. besteht ein beliebig schneller Informationstransport und implizite Verfahren spiegeln diese Struktur wider. Im Gegensatz dazu sind hyperbolische Dgln. durch einen Informationstransport mit endlichen Geschwindigkeiten gekennzeichnet. Diese Struktur wird qualitativ durch explizite Differenzenverfahren wiedergegeben. Dementsprechend sind die Restriktionen an die Schrittweiten bei expliziten im Vergleich zu impliziten Methoden für Erhaltungsgleichungen nicht extrem unterschiedlich. Folglich erweisen sich vom gesamten Rechenaufwand her explizite Differenzenverfahren im allgemeinen günstiger bei hyperbolischen Dgln. Implizite Verfahren werden daher im folgenden nicht mehr betrachtet.

Ferner kann das Verfahren (3.4) zu einer stabilen Methode modifiziert werden, indem die Auswertung U_j^n in der Mitte durch ein arithmetisches Mittel ersetzt wird. Damit entsteht das *Lax-Friedrichs Verfahren*

$$U_j^{n+1} = \frac{1}{2} (U_{j-1}^n + U_{j+1}^n) - \frac{k}{2h} A (U_{j+1}^n - U_{j-1}^n). \quad (3.6)$$

Desweiteren existieren auch Methoden, welche auf Taylor-Entwicklung beruhen. Die Expansion einer exakten Lösung $u \in C^3$ in der Zeit liefert

$$u(x, t + k) = u(x, t) + k u_t(x, t) + \frac{1}{2} k^2 u_{tt}(x, t) + \mathcal{O}(k^3). \quad (3.7)$$

Die lineare Differentialgleichung ermöglicht dann die Ersetzungen

$$u_t = -A u_x, \quad u_{tt} = -A u_{xt} = -A u_{tx} = -A(-A u_x)_x = A^2 u_{xx} \quad (3.8)$$

und es folgt

$$u(x, t + k) = u(x, t) - k A u_x(x, t) + \frac{1}{2} k^2 A^2 u_{xx}(x, t) + \mathcal{O}(k^3). \quad (3.9)$$

Werden nun die Ortsableitungen durch zentrierte Differenzenformeln ersetzt, so entsteht das häufig verwendete *Lax-Wendroff Verfahren*

$$U_j^{n+1} = U_j^n - \frac{k}{2h} A (U_{j+1}^n - U_{j-1}^n) + \frac{k^2}{2h^2} A^2 (U_{j+1}^n - 2U_j^n + U_{j-1}^n). \quad (3.10)$$

Eine einseitigen Diskretisierung der Ortsableitungen führt dagegen auf das *Beam-Warming Verfahren*

$$U_j^{n+1} = U_j^n - \frac{k}{2h} A (3U_j^n - 4U_{j-1}^n + U_{j-2}^n) + \frac{k^2}{2h^2} A^2 (U_j^n - 2U_{j-1}^n + U_{j-2}^n). \quad (3.11)$$

Differenzenverfahren für lineare Systeme $u_t + Au_x = 0$:

Name	Differenzenformel
expl. Euler	$U_j^{n+1} = U_j^n - \frac{k}{2h} A(U_{j+1}^n - U_{j-1}^n)$
impl. Euler	$U_j^{n+1} = U_j^n - \frac{k}{2h} A(U_{j+1}^{n+1} - U_{j-1}^{n+1})$
Upwind (links)	$U_j^{n+1} = U_j^n - \frac{k}{h} A(U_j^n - U_{j-1}^n)$
Upwind (rechts)	$U_j^{n+1} = U_j^n - \frac{k}{h} A(U_{j+1}^n - U_j^n)$
Lax-Friedrichs	$U_j^{n+1} = \frac{1}{2}(U_{j-1}^n + U_{j+1}^n) - \frac{k}{2h} A(U_{j+1}^n - U_{j-1}^n)$
Lax-Wendroff	$U_j^{n+1} = U_j^n - \frac{k}{2h} A(U_{j+1}^n - U_{j-1}^n) + \frac{k^2}{2h^2} A^2(U_{j+1}^n - 2U_j^n + U_{j-1}^n)$
Beam-Warming	$U_j^{n+1} = U_j^n - \frac{k}{2h} A(3U_j^n - 4U_{j-1}^n + U_{j-2}^n) + \frac{k^2}{2h^2} A^2(U_j^n - 2U_{j-1}^n + U_{j-2}^n)$

Analog können auch Verfahren höherer Ordnung hergeleitet werden.

Die obigen Verfahren sind für die lineare Differentialgleichung konstruiert. Wir sind jedoch auch an schwachen Lösungen, d.h. Lösungen der korrespondierenden Integralgleichung interessiert. Hierzu erfolgt später noch eine eigene Konsistenzbedingung. Es ist erstaunlich, dass sich Verfahren für die Differentialgleichung meistens direkt auf den Fall unstetiger/unglatter Lösungen anwenden lassen und zu sinnvollen Näherungen führen.

3.3 Konvergenzuntersuchung

Im diesem Abschnitt wird die Konvergenz von Differenzenverfahren für lineare Systeme diskutiert. Die Untersuchungen und Resultate erfolgen analog zu Konvergenzaussagen bei Methoden für gewöhnliche Differentialgleichungen.

Operatornotation

Für die Konvergenzuntersuchung sind noch geeignete Notationen erforderlich. Ein Einschrittverfahren, das aus einer Zeitschicht t_n die nächste Schicht $t_{n+1} = t_n + k$ bestimmt, sei beschrieben durch den Operator $\mathcal{H}_k$, d.h.

$$U^{n+1} = \mathcal{H}_k(U^n), \quad U_j^{n+1} = \mathcal{H}_k(U^n; j),$$

wobei U^n den Vektor aller Approximationen $U_j^n \in \mathbb{R}^m$ mit $j \in \mathbb{Z}$ bezeichnet. Die einzelne Näherung U_j^{n+1} hängt formal vom vollen Vektor U^n ab, da sie aus mehreren Näherungen der alten Schicht berechnet wird. Beispielsweise ergibt sich für das Verfahren (3.4) der Operator

$$\mathcal{H}_k(U^n; j) = U_j^n - \frac{k}{2h} A (U_{j+1}^n - U_{j-1}^n).$$

Der Differenzenoperator $\mathcal{H}_k$ kann in natürlicher Weise auf eine kontinuierliche Funktion $v : \mathbb{R} \rightarrow \mathbb{R}^m$ angewendet werden. Dabei wird zu jedem $x \in \mathbb{R}$ die Formel um diesen Punkt zentriert. Im Beispiel (3.4) ist $\mathcal{H}_k(v)$ definiert durch

$$\mathcal{H}_k(v; x) = (\mathcal{H}_k(v))(x) = v(x) - \frac{k}{2h} A (v(x+h) - v(x-h)).$$

Insbesondere kann so $\mathcal{H}_k$ auf der in (3.2) definierten stückweise konstanten Funktion U_k operieren. Es gilt dann

$$U_k(x, t+k) = \mathcal{H}_k(U_k(\cdot, t); x), \quad (3.12)$$

wodurch die Anwendung des Operators auf die diskreten Näherungen und die kontinuierliche Näherungsfunktion übereinstimmt. In der Notation wird wegen dieser Identität das gleiche Symbol verwendet.

In diesem Kapitel betrachten wir nur lineare Verfahren, wodurch der zugehörige Operator $\mathcal{H}_k$ ebenfalls linear ist, d.h.

$$\mathcal{H}_k(\alpha U^n + \beta V^n) = \alpha \mathcal{H}_k(U^n) + \beta \mathcal{H}_k(V^n). \quad (3.13)$$

für $\alpha, \beta \in \mathbb{R}$

Der Operator ist auch für abzählbar unendliche Gitterwerte $U^n \in \mathbb{R}^\infty$ sinnvoll definiert. Auf einem Gitter mit N Punkten ist $U^n \in \mathbb{R}^{mN}$ und der Operator kann durch eine $mN \times mN$ -Matrix beschrieben werden, was wir durch

$$U^{n+1} = \mathcal{H}_k U^n$$

mit gleichem Symbol für Operator und Matrix notieren.

Globaler Fehler und Konvergenz

Wir sind daran interessiert, wie gut die Ergebnisse U_j^n aus einem numerischen Verfahren die exakte Lösung u approximieren. Der *globale Fehler* eines Verfahrens wird daher als die Differenz zwischen exakter und berechneter Lösung definiert. Dies kann bei glatten Lösungen punktweise geschehen über

$$E_j^n = U_j^n - u_j^n.$$

Bei Erhaltungsgleichungen ist es wegen der Integralform mit schwachen Lösungen oft günstiger, die Differenz zum Zellenmittel (3.1) zu betrachten

$$\bar{E}_j^n = U_j^n - \bar{u}_j^n.$$

Definiert man die Fehlerfunktion

$$E_k(x, t) = U_k(x, t) - u(x, t),$$

dann ist E_j^n die punktweise Auswertung dieser Funktion in (x_j, t_n) und $\bar{E}_j^n$ das entsprechende Zellenmittel. Hiermit kann nun die *Konvergenz* einer Methode definiert werden.

Definition 3.1 Ein Verfahren $U^{n+1} = \mathcal{H}_k(U^n)$ mit dem globalen Fehler

$$E_k(x, t) = U_k(x, t) - u(x, t),$$

wobei sich $U_k(x, t)$ rekursiv mittels $\mathcal{H}_k$ aus den Anfangsdaten u_0 berechnet, wird konvergent bezüglich einer Norm $\|\cdot\|$ genannt, wenn

$$\lim_{k \rightarrow 0} \|E_k(\cdot, t)\| = 0 \quad (3.14)$$

für jedes $t \geq 0$ und alle Anfangswerte u_0 (aus einer bestimmten Menge) gilt.

Die Konvergenz eines Verfahrens ist von der betrachteten Norm auf dem Funktionenraum abhängig. Betrachten wir zunächst den Fall einer skalaren Funktion $v : \mathbb{R} \rightarrow \mathbb{R}$. Für stetige Funktionen wäre eine Konvergenz in der Maximumnorm

$$\|v\|_\infty = \sup \{|v(x)| : x \in \mathbb{R}\}$$

optimal. Für unstetige Lösungen ist dies jedoch nicht zu erwarten. Die Integralform der Erhaltungsgleichung zeigt, dass die Integralnorm

$$\|v\|_1 = \int_{-\infty}^{+\infty} |v(x)| \, dx$$

der Problemstellung entspricht.

Im weiteren benutzen wir stets die Integralnorm $\|\cdot\|_1$ (und lassen daher den Index 1 meist weg). Diese Integralnorm lässt sich auf die diskreten Gitterfunktionen übertragen mittels

$$\|U^n\|_1 = \|U_k(\cdot, t_n)\|_1 = h \sum_{j=-\infty}^{+\infty} |U_j^n|. \quad (3.15)$$

Die Definition der Konvergenz (3.14) erfolgt für eine vektorwertige Fehlerfunktion. Das Verfahren soll konvergent heißen, wenn der Fehler in jeder Komponente gegen null geht. Die Normen für die Funktionen werden daher in jeder Komponente einzeln angewendet. Die m Komponenten können dann durch eine Vektornorm erfasst werden, z.B. der Maximumnorm

$$\lim_{k \rightarrow 0} \|E_k(\cdot, t)\| = 0 \quad :\Leftrightarrow \quad \lim_{k \rightarrow 0} \max \{ \|(E_k(\cdot, t))_i\|_1 : i = 1, \dots, m \} = 0.$$

Da wir für den globalen Fehler meist nicht direkt eine Abschätzung angeben können, sind weitere Konzepte zur Untersuchung der Konvergenz nötig.

Lokaler Fehler und Konsistenz

Wir interessieren uns dafür, wie gut eine Differenzengleichung eine Differentialgleichung approximiert. Die diskreten Lösungen U_j^n oder die unstetige Funktion U_k aus dem Differenzenverfahren können jedoch nicht in die Differentialgleichung eingesetzt werden. Umgekehrt kann aber die Differenzenformel an einer Lösung der Differentialgleichung in den Gitterpunkten ausgewertet werden. So erhält man ein Maß für die Approximationsgüte.

Beispielsweise kann das Lax-Friedrichs Verfahren (3.6) in der Form

$$\frac{1}{k} (U_j^{n+1} - \frac{1}{2} (U_{j-1}^n + U_{j+1}^n)) + \frac{1}{2h} A (U_{j+1}^n - U_{j-1}^n) = 0.$$

geschrieben werden. Der *lokale Fehler* entsteht nun durch Auswerten dieser Formel an der exakten Lösung der Differentialgleichung

$$\begin{aligned} L_k(x, t) &= \frac{1}{k} (u(x, t+k) - \frac{1}{2} (u(x-h, t) + u(x+h, t))) \\ &\quad + \frac{1}{2h} A (u(x+h, t) - u(x-h, t)). \end{aligned}$$

Ist die Lösung hinreichend glatt, so folgt mittels Taylor-Entwicklung bei der Voraussetzung $r = \frac{k}{h}$ ist konstant ($u \equiv u(x, t)$)

$$\begin{aligned} L_k(x, t) &= \frac{1}{k} \left((u + ku_t + \frac{1}{2}k^2u_{tt} + \mathcal{O}(k^3)) - (u + \frac{1}{2}h^2u_{xx} + \mathcal{O}(h^4)) \right) \\ &\quad + \frac{1}{2h} A (2hu_x + \mathcal{O}(h^3)) \\ &= u_t + Au_x + \frac{1}{2} \left(ku_{tt} - \frac{h^2}{k} u_{xx} \right) + \mathcal{O}(k^2). \end{aligned} \tag{3.16}$$

Da u exakte Lösung der Differentialgleichung ist, kann man die Ersetzungen (3.8) verwenden und erhält

$$L_k(x, t) = \frac{1}{2}k \left(A^2 - \frac{h^2}{k^2} I \right) u_{xx}(x, t) + \mathcal{O}(k^2), \tag{3.17}$$

also insbesondere $L_k \rightarrow 0$ für $k \rightarrow 0$ punktweise. Die Ableitungen der Lösung lassen sich durch die Ableitungen der Anfangswerte u_0 beschränken. Wir erhalten so für die einzelnen Komponenten $i = 1, \dots, m$ die Abschätzung für beliebiges x

$$|(L_k(x, t))_i| \leq C_i k \quad \text{für alle } k < k_0,$$

d.h. auch in der Maximumnorm. Die Konstanten C_i hängen dabei nur von den Anfangswerten u_0 der Lösung ab. Ist der Träger der Anfangsdaten kompakt, dann ist die Integralnorm für jedes $t \geq 0$ endlich und es folgt

$$\|(L_k(\cdot, t))_i\|_1 \leq \tilde{C}_i k \quad \text{für alle } k < k_0$$

mit von den Anfangswerten abhängigen Konstanten $\tilde{C}_i$.

Analog kann nun die Definition der Konsistenz für ein allgemeines Einschrittverfahren erfolgen.

Definition 3.2 Ein Einschrittverfahren $U^{n+1} = \mathcal{H}_k(U^n)$ mit dem lokalen Fehler

$$L_k(x, t) = \frac{1}{k} [u(x, t+k) - \mathcal{H}_k(u(\cdot, t); x)] \quad (3.18)$$

ist konsistent, wenn für jedes feste $t \geq 0$ und alle Anfangsdaten (aus einer bestimmten Menge) gilt

$$\lim_{k \rightarrow 0} \|L_k(\cdot, t)\| = 0.$$

Das Einschrittverfahren ist konsistent mit Ordnung q , wenn für alle hinreichend glatten Anfangsdaten mit kompaktem Träger und alle $T > 0$ jeweils eine Konstante C_L und ein $k_0 > 0$ existiert, so dass

$$\|L_k(\cdot, t)\| \leq C_L k^q \quad \text{für alle } k < k_0, \quad t \leq T. \quad (3.19)$$

Die Definition der Konsistenz ist wieder von der Norm abhängig, wobei wir die Integralnorm $\|\cdot\|_1$ voraussetzen. Die Anwendung dieser Norm auf vektorwertige Funktionen erfolgt wie beim globalen Fehler.

Bei einem konsistenten Verfahren kann also der lokale Fehler durch Reduzierung der Schrittweiten beliebig klein gemacht werden. Die Konsistenz allein ist jedoch nicht hinreichend für die Konvergenz, d.h. die Verkleinerung des globalen Fehlers.

Stabilität

Um die Konvergenz eines Verfahrens zu garantieren wird noch die *Stabilität* eines Verfahrens benötigt. Der lokale Fehler (3.18) kann umgeformt werden zu

$$u(x, t+k) = \mathcal{H}_k(u(\cdot, t); x) + kL_k(x, t).$$

Mit (3.12) folgt aus der Linearität (3.13) des Operators für den globalen Fehler die Rekursion

$$E_k(x, t+k) = \mathcal{H}_k(E_k(\cdot, t); x) - kL_k(x, t).$$

Der globale Fehler zur Zeit $t + k$ besteht damit aus zwei Anteilen: einem Beitrag aus dem globalen Fehler der vorhergehenden Zeitschicht und einem neuen Anteil aus dem aktuellen lokalen Fehler. Auflösen dieser Rekursion liefert zur Zeit t_n

$$E_k(\cdot, t_n) = \mathcal{H}_k^n(E_k(\cdot, 0)) - k \sum_{i=1}^n \mathcal{H}_k^{n-i}(L_k(\cdot, t_{i-1})), \quad (3.20)$$

wobei $\mathcal{H}_k^i$ die i -fache Verkettung des Operators $\mathcal{H}_k$ bezeichnet. In (3.20) tritt der Fehler in den Anfangswerten auf, der dadurch entsteht, dass die kontinuierlichen Daten u_0 durch diskrete Werte approximiert werden. Dieser Fehler verschwindet jedoch für $h \rightarrow 0$. Damit der globale Fehler zur Zeit t_n beschränkt bleibt, muss sichergestellt werden, dass die lokalen Fehler in (3.20) nicht durch die Verkettungen des Operators $\mathcal{H}_k$ verstärkt werden. Dies führt auf die folgende Definition.

Definition 3.3 *Das lineare Einschrittverfahren $U^{n+1} = \mathcal{H}_k(U^n)$ heißt stabil (nach Lax-Richtmyer), wenn für jede Zeit $T \geq 0$ eine Konstante C_S und ein $k_0 > 0$ existiert, so dass*

$$\|\mathcal{H}_k^n\| \leq C_S \quad \text{für alle } nk \leq T, \quad k < k_0. \quad (3.21)$$

Die Bedingung für die Stabilität benutzt die Operatornorm der Verkettungen von $\mathcal{H}_k$. Für einen linearen Operator $\mathcal{G} : \mathbb{R}^\infty \rightarrow \mathbb{R}^\infty$ ist die Operatornorm gegeben durch

$$\|\mathcal{G}\| = \sup_{V \neq 0} \frac{\|\mathcal{G}(V)\|}{\|V\|} \quad (V \in \mathbb{R}^\infty),$$

wobei wir auf $\mathbb{R}^\infty$ die Norm (3.15) betrachten. Eine Verallgemeinerung dieser Norm auf ein System erfolgt wieder durch Bilden des Maximums. Im Fall von N Gitterpunkten im Ort kann eine beliebige Vektornorm auf $\mathbb{R}^{mN}$ verwendet werden.

Wegen $\|\mathcal{H}_k^n\| \leq \|\mathcal{H}_k\|^n$ ist die Bedingung $\|\mathcal{H}_k\| \leq 1$ hinreichend für die Stabilität. Ein kleines Wachstum ist jedoch erlaubt, denn aus

$$\|\mathcal{H}_k\| \leq 1 + \alpha k \quad \text{für alle } k < k_0$$

folgt

$$\|\mathcal{H}_k^n\| \leq \|\mathcal{H}_k\|^n \leq (1 + \alpha k)^n \leq e^{\alpha kn} \leq e^{\alpha T} \quad \text{für alle } nk \leq T.$$

Als Beispiel betrachten wir das Lax-Friedrichs Verfahren (3.6) für die lineare Advektionsgleichung

$$U_j^{n+1} = \frac{1}{2} (U_{j-1}^n + U_{j+1}^n) - \frac{ak}{2h} (U_{j+1}^n - U_{j-1}^n).$$

Mit der Norm (3.15) auf $\mathbb{R}^\infty$ folgt dann

$$\begin{aligned}\|U^{n+1}\| &= h \sum_{j=-\infty}^{+\infty} |U_j^{n+1}| \\ &\leq \frac{h}{2} \left[\sum_{j=-\infty}^{+\infty} \left|1 - \frac{ak}{h}\right| \cdot |U_{j+1}^n| + \sum_{j=-\infty}^{+\infty} \left|1 + \frac{ak}{h}\right| \cdot |U_{j-1}^n| \right].\end{aligned}$$

Mit der Stabilitätsbedingung

$$\left| \frac{ak}{h} \right| \leq 1 \quad (3.22)$$

werden die Koeffizienten positiv, so dass wir abschätzen können

$$\begin{aligned}\|U^{n+1}\| &\leq \frac{h}{2} \left[\left(1 - \frac{ak}{h}\right) \sum_j |U_{j+1}^n| + \left(1 + \frac{ak}{h}\right) \sum_j |U_{j-1}^n| \right] \\ &= \frac{1}{2} \left[\left(1 - \frac{ak}{h}\right) \|U^n\| + \left(1 + \frac{ak}{h}\right) \|U^n\| \right] = \|U^n\|.\end{aligned}$$

Somit folgt aus (3.22) in diesem Verfahren $\|\mathcal{H}_k\| \leq 1$ und damit die Stabilität.

Für ein lineares hyperbolisches System $u_t + Au_x = 0$ können wir mit der linearen Transformation (2.33),(2.34) das System in skalare Gleichungen zerlegen. Das Lax-Friedrichs Verfahren ist dementsprechend stabil, wenn

$$\left| \frac{\lambda_i k}{h} \right| \leq 1 \quad \text{für alle } i = 1, \dots, m \quad (3.23)$$

gilt, wobei λ_i die Eigenwerte der Matrix A sind.

Das Konzept nach Lax-Richtmyer stimmt auch mit der üblichen Lipschitz-stetigen Abhängigkeit der numerischen Lösung von Störungen in den Anfangswerten überein. Wird aus den Anfangsdaten U_0 die Näherung U^n und aus V^0 dann V^n berechnet, so gilt

$$\begin{aligned}\|U^n - V^n\| &= \|\mathcal{H}_k^n(U^0) - \mathcal{H}_k^n(V^0)\| = \|\mathcal{H}_k^n(U^0 - V^0)\| \\ &\leq \|\mathcal{H}_k^n\| \cdot \|U^0 - V^0\| \leq C_S \|U^0 - V^0\|.\end{aligned}$$

Ein anderes Konzept liefert die *Stabilität nach von-Neumann*, welche nur bei linearen Gleichungen Bedeutung hat und Bezüge zur Fourieranalysis enthält. Die skalare Gleichung $u_t + au_x = 0$ besitzt die triviale Lösung $u \equiv 0$. Diese Lösung wird nun gestört, indem man den Separationsansatz $u(x, t) = v(t) \cdot w(x)$ verwendet. Für die Funktion v wird eine Exponentialfunktion eingesetzt, während w eine oszillatorische Störung in komplexer Schreibweise darstellt, d.h.

$$v(t) = \exp(\alpha t), \quad w(x) = \exp(i\mu x) \quad (\alpha \in \mathbb{C}, \mu \in \mathbb{R}, i = \sqrt{-1}).$$

Wegen $v(0) = 1$ stellt $w(x)$ eine Störung der exakten Anfangsdaten $u(x, 0) \equiv 0$ dar. Entscheidend ist nun, ob diese Störung von einem numerischen Verfahren mit der Zeit verstärkt wird oder zur korrekten Lösung $u \equiv 0$ gedämpft wird. In eine Differenzenformel für die Gitterpunkte $U_j^{n+1} = \mathcal{H}_k(U^n; j)$ wird daher der Separationsansatz eingesetzt. Daraus leitet man eine Bedingung für die Gitterweiten h, k her, so dass $|e^{\alpha k}| \leq 1$ für alle $\mu \in \mathbb{R}$ gilt. Dies bedeutet, dass beliebige oszillatorische Störungen in den Anfangswerten mit der Zeit nicht verstärkt werden. Erfüllt eine spezielle Wahl der Gitterweiten die Bedingung, so heißt das zugehörige Verfahren stabil nach von-Neumann.

Als Beispiel betrachten wir wieder das Lax-Friedrichs Verfahren (3.6). Im skalaren Fall folgt nach der Formel mit dem Separationsansatz

$$\begin{aligned} e^{\alpha(n+1)k} \cdot e^{i\mu jh} &= \frac{1}{2} [e^{\alpha nk} \cdot e^{i\mu(j-1)h} + e^{\alpha nk} \cdot e^{i\mu(j+1)h}] \\ &\quad - \frac{k}{2h} a [e^{\alpha nk} \cdot e^{i\mu(j+1)h} - e^{\alpha nk} \cdot e^{i\mu(j-1)h}]. \end{aligned}$$

Kürzen der Terme liefert

$$\begin{aligned} e^{\alpha k} &= \frac{1}{2} [e^{-i\mu h} + e^{i\mu h}] - \frac{ka}{2h} [e^{i\mu h} - e^{-i\mu h}] \\ &= \cos(\mu h) - i \frac{ka}{h} \sin(\mu h). \end{aligned}$$

Somit ist $|e^{\alpha k}| \leq 1$ äquivalent zu

$$\cos^2(\mu h) + \left(\frac{ka}{h}\right)^2 \sin^2(\mu h) \leq 1 \quad \Leftrightarrow \quad \left(\left(\frac{ka}{h}\right)^2 - 1\right) \sin^2(\mu h) \leq 0$$

und damit folgt als Stabilitätsbedingung nach von-Neumann genau (3.22). Diese Bedingung lässt sich für festes $a \in \mathbb{R}$ erreichen, indem die Zeitschrittweite k hinreichend kleiner als die Ortsschrittweite h gewählt wird. Ferner kann mit diesem Konzept auch die Stabilität des expliziten und impliziten Euler-Verfahrens (3.4) bzw. (3.5) entschieden werden.

Äquivalenzsatz von Lax

Mit den Aussagen der vorhergehenden Abschnitte kann nun das entscheidende Resultat über die Konvergenz von linearen Verfahren gegeben werden, der sogenannte *Äquivalenzsatz von Lax*.

Satz 3.4 (Lax) *Das lineare Verfahren $U^{n+1} = \mathcal{H}_k(U^n)$ für die lineare Differentialgleichung $u_t + Au_x = 0$ sei konsistent. Dann sind Stabilität (nach Lax-Richtmyer) und Konvergenz des Verfahrens äquivalent.*

Beweis: siehe z.B. J.C. Strikwerda: Finite Difference Schemes and Partial Differential Equations. Wadsworth & Brooks/Cole, 1989.

Wir zeigen hier die für die Numerik wichtige Aussage, dass aus der Stabilität die Konvergenz folgt. Aus der Formel (3.20) folgt nach den Regeln für die Norm

$$\|E_k(\cdot, t_n)\| \leq \|\mathcal{H}_k^n\| \cdot \|E_k(\cdot, 0)\| + k \sum_{i=1}^n \|\mathcal{H}_k^{n-i}\| \cdot \|L_k(\cdot, t_{i-1})\|.$$

Mit der Stabilität (3.21) können wir sofort abschätzen

$$\|E_k(\cdot, t_n)\| \leq C_S \left(\|E_k(\cdot, 0)\| + k \sum_{i=1}^n \|L_k(\cdot, t_{i-1})\| \right).$$

Ist das Verfahren konsistent mit Ordnung q , dann gilt nach (3.19) für alle Zeitpunkte $nk = t_n \leq T$

$$\|E_k(\cdot, t_n)\| \leq C_S (\|E_k(\cdot, 0)\| + TC_L k^q) \quad \text{falls } k < k_0.$$

Ist der Fehler in den Anfangsdaten null, dann liegt sofort Konvergenz vor. Andernfalls fordert man, dass dieser Anfangsfehler die Größenordnung $\mathcal{O}(h^q)$ besitzt. Es sei dabei wieder $r = \frac{k}{h}$ konstant. Dann gilt mit $t = t_n$

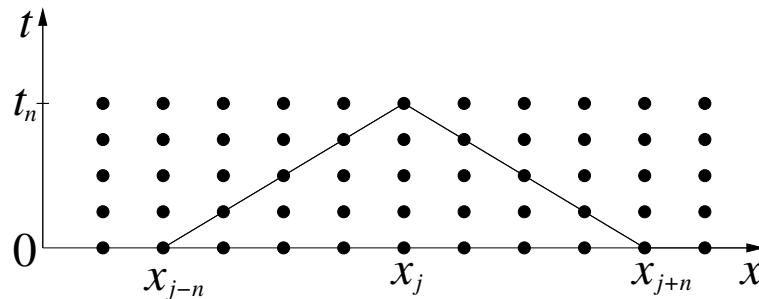
$$\|E_k(\cdot, t)\| \leq C_E k^q \quad \text{für alle } t \leq T \text{ und } k < k_0.$$

Insbesondere ist dann die Konsistenzordnung identisch mit der Konvergenzordnung des Verfahrens.

CFL-Bedingung

Bei der Theorie für lineare hyperbolische Systeme haben wir aus (2.37) erkannt, dass die Lösung in einem Punkt (x^*, t^*) nur von den Anfangswerten in höchstens m verschiedenen Ortspunkten zur Zeit $t = 0$ abhängt. Eine notwendige Forderung an ein numerisches Verfahren ist daher, dass zur Berechnung der Näherung in einem Gitterpunkt (x_j, t_n) die Anfangsdaten nahe der jeweiligen Ortspunkte aus seinem Abhängigkeitsbereich eingehen müssen. Anderenfalls könnte man die Anfangsdaten so abändern, dass sich die exakte Lösung in (x_j, t_n) deutlich verändert, jedoch die Näherung aus dem Verfahren gleich bleibt.

Als Beispiel diskutieren wir ein Verfahren, bei dem die Näherung U_j^{n+1} von den drei Vektoren U_ℓ^n mit $\ell = j-1, j, j+1$ in der alten Zeitschicht abhängt.



Rekursiv folgt dann für den *numerischen Abhängigkeitsbereich* der Lösung zur Zeit $t = 0$

$$\mathcal{D}_k(x_j, t_n) \subset \{x : |x - x_j| \leq nh\}$$

oder allgemein für den Punkt (x^*, t^*) mit $t^* = nk$

$$\mathcal{D}_k(x^*, t^*) \subset \{x : |x - x^*| \leq \frac{h}{k} t^*\}.$$

Mit konstantem Schrittweitenverhältnis $r = \frac{k}{h}$ gilt im Grenzfalle $k \rightarrow 0$ dann

$$\mathcal{D}_0(x^*, t^*) = \{x : |x - x^*| \leq \frac{t^*}{r}\}.$$

Die Gleichheit hier folgt daraus, dass für $k \rightarrow 0$ dann die Gitterpunkte dicht liegen. Für lokal integrable und insbesondere für glatte Lösungen werden so alle Anfangswerte in dem Ortsintervall erfasst. Dieser numerische Abhängigkeitsbereich muss nun den *analytischen Abhängigkeitsbereich* $\mathcal{D}$ aus (2.37) enthalten

$$\mathcal{D}(x^*, t^*) \subset \mathcal{D}_0(x^*, t^*). \quad (3.24)$$

Diese Forderung heißt *CFL-Bedingung* benannt nach Courant, Friedrichs und Lewy. Sie ist notwendig für die Konvergenz des Verfahrens und damit nach dem Äquivalenzsatz von Lax eine notwendige Stabilitätsbedingung.

Für das obige Beispiel eines numerischen Abhängigkeitsbereichs folgt die spezielle CFL-Bedingung

$$|(x^* - \lambda_i t^*) - x^*| \leq \frac{t^*}{r} \quad \Leftrightarrow \quad \left| \frac{\lambda_i k}{h} \right| \leq 1 \quad (3.25)$$

für alle Eigenwerte λ_i der Matrix A aus der Differentialgleichung. Diese notwendige Forderung entspricht genau der hinreichenden Stabilitätsbedingung (3.23) für das Lax-Friedrichs Verfahren, welches obigen numerischen Abhängigkeitsbereich aufweist.

Das explizite Euler-Verfahren (3.4) erfüllt ebenfalls unter der Voraussetzung (3.25) die CFL-Bedingung. Diese Methode ist jedoch instabil. Daraus erkennt man, dass die CFL-Bedingung nur ein notwendiges Konzept darstellt und kein hinreichendes.

3.4 Einseitige Verfahren

Differenzenschemata wie das Lax-Friedrichs und das Lax-Wendroff Verfahren benutzen zur Berechnung von U_j^{n+1} die alten Daten U_ℓ^n für $\ell = j-1, j, j+1$. Damit verwenden sie Information zu beiden Seiten der neuen Näherung in der x - t -Ebene. Ebenso können auch einseitige Methoden konstruiert werden. Betrachten wir die

lineare Advektionsgleichung $u_t + au_x = 0$, so kann die Ortsableitung mit dem gewöhnlichen Differenzenquotienten in zwei Weisen diskretisiert werden, nämlich

$$U_j^{n+1} = U_j^n - \frac{ak}{h} (U_j^n - U_{j-1}^n) \quad (3.26)$$

oder

$$U_j^{n+1} = U_j^n - \frac{ak}{h} (U_{j+1}^n - U_j^n). \quad (3.27)$$

Die Diskussion der CFL-Bedingung zeigt: Eine notwendige Stabilitätsbedingung für (3.26) bzw. (3.27) ist

$$0 \leq \frac{ak}{h} \leq 1 \quad \text{bzw.} \quad -1 \leq \frac{ak}{h} \leq 0. \quad (3.28)$$

Neben einer Forderung an die Schrittweiten folgt, dass (3.26) nur für $a \geq 0$ und (3.27) nur für $a \leq 0$ verwendbar ist. Die Bedingungen (3.28) erweisen sich auch als hinreichend für die Stabilität. Man nennt die Methoden (3.26), (3.27) dann *Upwind Verfahren* in Anlehnung an die Gasdynamik. Beide Verfahren berücksichtigen jeweils die Richtung des Informationstransports entlang der Charakteristiken. Durch diese Spezialisierung sind sie im allgemeinen effizienter als beidseitige Verfahren, die weniger signifikante Information aus alten Schichten benutzen.

Die Verwendung einer einzelnen Upwind Methode für ein lineares hyperbolisches System $u_t + Au_x = 0$ ist nur möglich, falls alle Eigenwerte von A das gleiche Vorzeichen haben. Dies folgt wieder aus einer Diskussion der CFL-Bedingung. Für Systeme in den Anwendungen ist dies im allgemeinen nicht gegeben. Beispielsweise besitzen die Eulerschen Gasgleichungen die Eigenwerte $v - c, v, v + c$, so dass ein gleiches Vorzeichen nur besteht, wenn die Gasgeschwindigkeit v größer als die Schallgeschwindigkeit c ist.

Wir können jedoch ein lineares hyperbolisches System in einzelne skalare Gleichungen entkoppeln über (2.33), (2.34). Bei jeder einzelnen Gleichung ist es dann möglich, das entsprechende Upwind Verfahren anzuwenden. Wir erhalten mit den Bezeichnungen

$$\begin{aligned} \lambda_i^+ &= \max\{\lambda_i, 0\}, & D^+ &= \text{diag}(\lambda_1^+, \dots, \lambda_m^+) \\ \lambda_i^- &= \min\{\lambda_i, 0\}, & D^- &= \text{diag}(\lambda_1^-, \dots, \lambda_m^-) \end{aligned}$$

und der Transformation $V = R^{-1}U$ das Schema

$$V_j^{n+1} = V_j^n - \frac{k}{h} D^+ (V_j^n - V_{j-1}^n) - \frac{k}{h} D^- (V_{j+1}^n - V_j^n)$$

und mittels der Rücktransformation $U = RV$

$$U_j^{n+1} = U_j^n - \frac{k}{h} A^+ (U_j^n - U_{j-1}^n) - \frac{k}{h} A^- (U_{j+1}^n - U_j^n) \quad (3.29)$$

mit

$$A^+ = RD^+R^{-1} \quad \text{und} \quad A^- = RD^-R^{-1}.$$

Es gilt $A^+ + A^- = A$. Besitzen alle Eigenwerte gleiches Vorzeichen, so gilt $A^+ = 0$ oder $A^- = 0$ und das Schema reduziert sich zu einer einseitigen Methode. Im Kontext der CFL-Bedingung folgt hier analog zu (3.28) noch die Einschränkung (3.23) an die Schrittweiten in Abhängigkeit von den Beträgen der Eigenwerte.

Analoge Aussagen lassen sich auch für das Beam-Warming Verfahren (3.11), welches eine einseitige Methode zweiter Ordnung darstellt, herleiten.

3.5 Verhalten bei unstetigen Lösungen

Die bisherigen Konsistenz- und Konvergenzuntersuchungen haben hinreichend glatte Lösungen vorausgesetzt. Bei unstetigen Lösungen erwarten wir daher, dass Probleme bei für glatte Funktionen konstruierten Differenzenverfahren auftreten. Um dies zu untersuchen, betrachten wir das folgende Riemann-Problem der linearen Advektionsgleichung

$$\begin{aligned} u_t + au_x &= 0, \\ u_0(x) &= \begin{cases} 1 & \text{für } x < 0, \\ 0 & \text{für } x > 0. \end{cases} \end{aligned} \tag{3.30}$$

Die exakte Lösung ist gerade $u(x, t) = u_0(x - at)$. Aufgrund der Unstetigkeit wird jedoch eine Differenzenformel für u_x im Fall $h \rightarrow 0$ unbeschränkt. Der lokale Fehler einer Methode konvergiert somit nicht gegen null. Dadurch lässt sich der Satz von Lax nicht anwenden. Ein Ausweg besteht darin, die Anfangsdaten u_0 durch glatte Funktionen u_0^ε zu approximieren, wobei im Grenzfall $\varepsilon \rightarrow 0$ dann u_0 erreicht wird. Für ein bei glatten Funktionen konsistentes und stabiles Verfahren folgt somit wieder die Konvergenz. Jedoch kann sich die Konvergenzordnung im Grenzfall deutlich reduzieren. Insbesondere wird die numerische Näherung für ein bestimmtes endliches Gitter meist zu fehlerhaft sein.

An der numerischen Simulation des Riemann-Problems (3.30) zu $a = 1$ erkennen wir die auftretenden Probleme. Bei Verfahren erster Ordnung, nämlich Lax-Friedrichs und Upwind, wird die Unstetigkeit fälschlicherweise stark abgeglättet. Methoden zweiter Ordnung, also Lax-Wendroff und Beam-Warming, lösen die Unstetigkeit relativ scharf auf, produzieren aber inkorrekte Oszillationen. Dieses Verhalten der numerischen Verfahren ist typisch.

Um das qualitative Verhalten der Differenzenverfahren zu verstehen, werden sogenannte *modifizierte Gleichungen* betrachtet. Bei einer konsistenten Methode approximiert die Differenzenformel die Differentialgleichung von einer gewissen

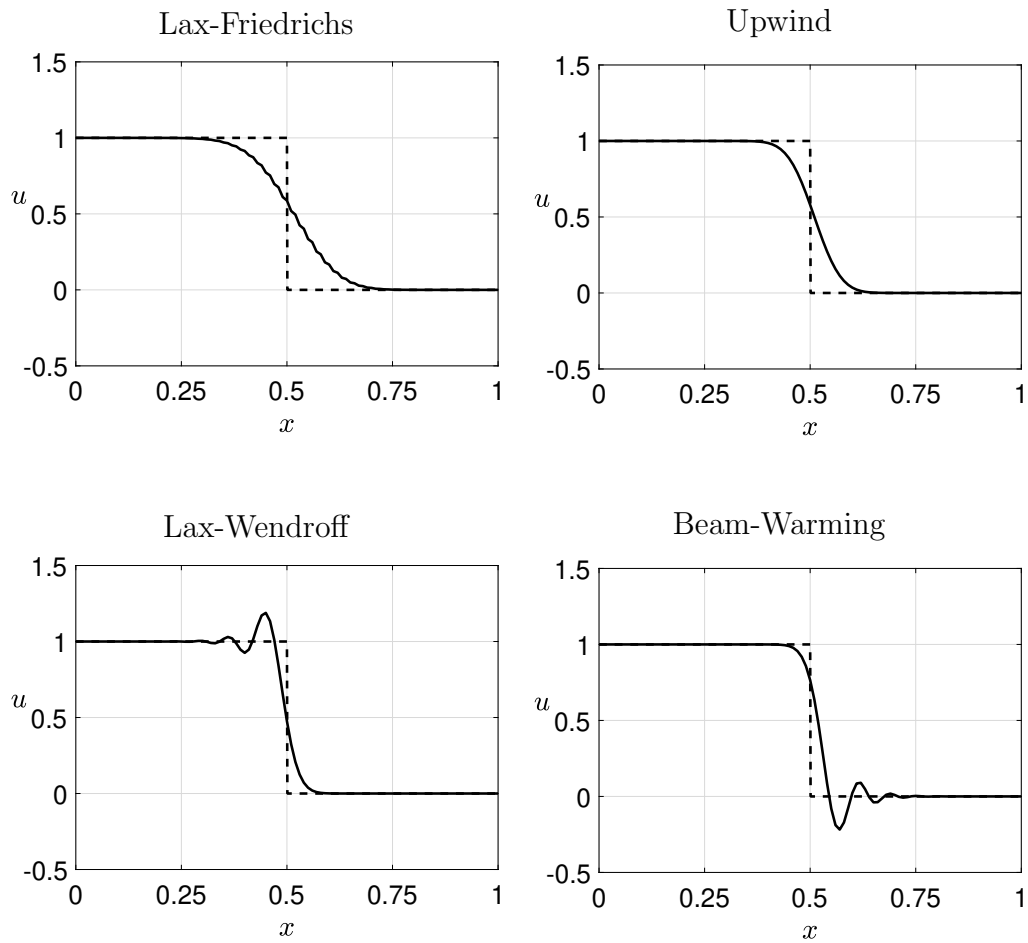


Abbildung: Numerische Lösungen für Riemann-Problem der Advektionsgleichung mit $a = 1$ zur Zeit $t = 0.5$ berechnet mit Schrittweiten $h = 0.01$ und $k = 0.005$.

Ordnung. Man kann jedoch eine neue modifizierte Differentialgleichung aufstellen, welche dann von dem Differenzenverfahren mit höherer Ordnung gelöst wird.

Als Beispiel betrachten wir das Lax-Friedrichs Verfahren (3.6), welches für ein lineares hyperbolisches System nach (3.17) konsistent von Ordnung 1 ist. Aus der Herleitung (3.16) folgt nun, dass das Verfahren angewendet auf die Differentialgleichung

$$u_t + Au_x + \frac{1}{2} \left(ku_{tt} - \frac{h^2}{k} u_{xx} \right) = 0 \quad (3.31)$$

konsistent von Ordnung 2 ist. Zur Vereinfachung soll u_{tt} durch Ortsableitungen dargestellt werden. Eine direkte Anwendung der Ersetzungen (3.8) ist nicht möglich, da u nicht mehr die ursprüngliche Differentialgleichung erfüllt. Stattdessen kann man analog herleiten

$$\begin{aligned} u_{tt} &= -Au_{tx} - \frac{1}{2} \left(ku_{ttt} - \frac{h^2}{k} u_{xxt} \right) \\ &= -A(-Au_{xx} + \mathcal{O}(k)) + \mathcal{O}(k) \\ &= A^2 u_{xx} + \mathcal{O}(k). \end{aligned}$$

Dies in (3.31) eingesetzt liefert einen Fehlerterm $\mathcal{O}(k^2)$, d.h. in der Größenordnung der nicht berücksichtigten Konsistenzfehler. Zusammen folgt somit die modifizierte Gleichung für das Lax-Friedrichs Verfahren

$$u_t + Au_x = \frac{h^2}{2k} \left(I - \frac{k^2}{h^2} A^2 \right) u_{xx}. \quad (3.32)$$

Das Lax-Friedrichs Verfahren löst diese Gleichung mit einer Genauigkeit von zweiter Ordnung. Für $\frac{k}{h}$ konstant geht dieser Ausdruck bei $k \rightarrow 0$ in die ursprüngliche Erhaltungsgleichung über. Die Gleichung (3.32) besitzt die Form

$$u_t + Au_x = Bu_{xx} \quad \text{mit} \quad B = \frac{h^2}{2k} \left(I - \frac{k^2}{h^2} A^2 \right)$$

und entspricht daher einer Advektions-Diffusionsgleichung. Mit der CFL-Bedingung sind alle Eigenwerte von B nichtnegativ. Dies erklärt das Verhalten des Lax-Friedrichs Verfahrens bei dem Riemann-Problem (3.30). Durch die Diffusion wird die Unstetigkeit mit der Zeit stark verschmiert. Für $k \rightarrow 0$ geht der Diffusionsterm ebenfalls gegen null. Ein Vorteil der Lax-Friedrichs Methode ist daher, dass im Grenzfall $k \rightarrow 0$ automatisch die Lösung bei verschwindender Diffusion, also der physikalisch sinnvolle Zustand, berechnet wird.

Für das Upwind Verfahren kann man als modifizierte Gleichung

$$u_t + Au_x = \frac{h}{2} A \left(I - \frac{k}{h} A \right) u_{xx}$$

herleiten, was ebenfalls einer Advektions-Diffusionsgleichung entspricht. Mit der zugehörigen CFL-Bedingung folgt, dass alle Eigenwerte der Diffusionsmatrix nicht-negativ sind. Diese Eigenwerte sind hier häufig kleiner als bei der Lax-Friedrichs Methode, wodurch die Unstetigkeit weniger stark abgeglättet wird, d.h. die Genauigkeit ist höher.

Für das Lax-Wendroff Verfahren, welches konsistent von Ordnung 2 ist, folgt, dass die modifizierte Gleichung

$$u_t + Au_x = \frac{h^2}{6} A \left(\frac{k^2}{h^2} A^2 - I \right) u_{xxx}$$

von dritter Ordnung approximiert wird. Für das Beam-Warming Verfahren erhalten wir die ähnliche modifizierte Gleichung

$$u_t + Au_x = \frac{h^2}{6} A \left(2I - \frac{3k}{h} A + \frac{k^2}{h^2} A^2 \right) u_{xxx}.$$

Im eindimensionalen Fall besitzen beide Gleichungen die Gestalt

$$u_t + au_x = \mu u_{xxx} \tag{3.33}$$

mit einer Konstante μ abhängig von a, h, k . Dies stellt eine Dispersionsgleichung dar. Wird als Ansatz für die Lösung die Wellenfunktion

$$u(x, t) = e^{i(\kappa x - \omega(\kappa)t)} \tag{3.34}$$

in (3.33) benutzt, dann folgt die Dispersionsrelation

$$\omega(\kappa) = a\kappa + \mu\kappa^3.$$

Die Phasengeschwindigkeit der jeweiligen Welle ist daher

$$v_{\text{ph}}(\kappa) = \frac{\omega(\kappa)}{\kappa} = a + \mu\kappa^2$$

und die entsprechende Gruppengeschwindigkeit lautet

$$v_{\text{gr}}(\kappa) = \omega'(\kappa) = a + 3\mu\kappa^2. \tag{3.35}$$

Wellen mit unterschiedlicher Wellennummer κ bewegen sich daher mit verschiedenen Geschwindigkeiten fort. Wird der Ansatz (3.34) dagegen in die Wellengleichung $u_{tt} = c^2 u_{xx}$ eingesetzt, dann folgt $v_{\text{ph}} = v_{\text{gr}} = \pm c$.

Der Parameter μ in (3.33) berechnet sich mit $\nu = \frac{ak}{h}$ für das Lax-Wendroff zu

$$\mu = \frac{1}{6} h^2 a (\nu^2 - 1)$$

und für das Beam-Warming Verfahren zu

$$\mu = \frac{1}{6}h^2a(2 - 3\nu + \nu^2) = \frac{1}{6}h^2a\left(\left(\nu - \frac{3}{2}\right)^2 - \frac{1}{4}\right)$$

Aufgrund der CFL-Bedingung sei bei der Lax-Wendroff Methode $|\nu| < 1$. Dadurch folgt bei $a > 0$ hier $\mu < 0$ und somit $v_{\text{gr}}(\kappa) < a$ für alle $\kappa \neq 0$ nach (3.35). Beim Beam-Warming Verfahren muss nach der CFL-Bedingung jetzt $a > 0$ gelten (damit $\nu > 0$). Dann ist $\mu \leq 0$ genau dann, wenn $\nu \in [1, 2]$ gilt. Aus $0 < \nu < 1$ folgt somit $\mu > 0$ und dadurch $v_{\text{gr}}(\kappa) > a$ für alle $\kappa \neq 0$.

Setzt man die Anfangsdaten im Riemann-Problem (3.30) für große $|x|$ auf null, dann können wir die Lösung über die Fourier-Transformierte darstellen

$$u(x, t) = \int_{-\infty}^{+\infty} \tilde{u}(\kappa, t) e^{i\kappa x} d\kappa.$$

Für beliebig oft differenzierbare Anfangsdaten u_0 fällt die Fouriertransformierte $\tilde{u}(\kappa, 0)$ für $|\kappa| \rightarrow \infty$ exponentiell ab. Aufgrund der Unstetigkeit in (3.30) gilt jedoch $|\tilde{u}(\kappa, 0)| \sim \frac{1}{|\kappa|}$ für $|\kappa| \rightarrow \infty$, also eine viel langsamere Verringerung. Daher ist das Verhalten der hochoszillatorischen Komponenten der Lösung bei den Methoden zweiter Ordnung noch von Bedeutung. Wegen der Aussagen für die Gruppengeschwindigkeiten werden beim Lax-Wendroff Verfahren Oszillationen mit der Zeit langsamer als die Unstetigkeit transportiert und beim Beam-Warming Verfahren schneller. Die modifizierten Gleichungen erklären also genau die beobachteten Effekte beim Riemann Problem.

Desweiteren kann für das Riemann Problem (3.30) der globale Fehler der Methoden in der Integralnorm abgeschätzt werden. Bei $h \rightarrow 0$ und $\frac{k}{h}$ konstant zeigt sich für das Lax-Friedrichs Verfahren

$$\|u(\cdot, t) - U_k(\cdot, t)\|_1 \approx C\sqrt{ht} \quad \text{für alle } t \geq 0. \quad (3.36)$$

Die Konvergenzordnung 1 bei glatten Lösungen reduziert sich daher zur Ordnung $\frac{1}{2}$ bei Unstetigkeiten. Analog folgt für das Lax-Wendroff Verfahren eine Reduktion der Konvergenzordnung von 2 auf $\frac{2}{3}$.

4 Konservative Methoden für nichtlineare Probleme

Nun wollen wir geeignete numerische Verfahren für Anfangswertprobleme zu nichtlinearen Systemen aus Erhaltungsgleichungen konstruieren.

4.1 Konservative Form

Gegeben sei ein nichtlineares System aus Erhaltungsgleichungen

$$u_t + f(u)_x = 0 \quad (4.1)$$

mit $u : \mathbb{R} \times \mathbb{R}_0^+ \rightarrow \mathbb{R}^m$, $f : \mathbb{R}^m \rightarrow \mathbb{R}^m$. Durch Differentiation erhalten wir die quasilineare Form

$$u_t + A(u)u_x = 0 \quad \text{mit} \quad A(u) = \frac{\partial f}{\partial u}, \quad (4.2)$$

die jedoch nur für glatte Lösungen äquivalent zu (4.1) ist. Numerische Verfahren für lineare Systeme lassen sich häufig direkt auf die quasilineare Gleichung (4.2) durch eine lokale Linearisierung übertragen. Dadurch ist jedoch nicht garantiert, dass die Verfahren auch schwache Lösungen des Systems (4.1) in korrekter Weise liefern, da die Äquivalenz nur für glatte Funktionen gilt. Zudem kann ein numerisches Verfahren bei linearen Problemen stabil sein, jedoch gewisse nichtlineare Stabilitätskonzepte verletzen, wodurch bei nichtlinearen Systemen keine Konvergenz mehr vorliegt. Ein weiteres Problem entsteht dadurch, dass im nichtlinearen Fall schwache Lösungen im allgemeinen nicht eindeutig sind. Ein numerisches Verfahren wird daher nur gegen die physikalisch sinnvolle Lösung konvergieren, wenn es geeignete Entropiebedingungen erfüllt.

Um entsprechende Verfahren zur Berechnung schwacher Lösungen zu erhalten ist daher von der Integralform der Erhaltungsgleichung auszugehen. Für das nichtlineare System (4.1) liefert die Integralform (1.3) eine Formel für die Zellenmittel (3.1), nämlich

$$\begin{aligned} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x, t_{n+1}) \, dx &= \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x, t_n) \, dx \\ &- \left[\int_{t_n}^{t_{n+1}} f(u(x_{j+\frac{1}{2}}, t)) \, dt - \int_{t_n}^{t_{n+1}} f(u(x_{j-\frac{1}{2}}, t)) \, dt \right]. \end{aligned} \quad (4.3)$$

Eine geeignete numerische Methode für schwache Lösungen soll daher diese Integralgleichung approximieren.

Definition 4.1 Ein numerisches Verfahren heißt konservativ, wenn es in der Form

$$U_j^{n+1} = U_j^n - \frac{k}{h} [F(U_{j-p}^n, U_{j-p+1}^n, \dots, U_{j+q}^n) - F(U_{j-p-1}^n, U_{j-p}^n, \dots, U_{j+q-1}^n)] \quad (4.4)$$

mit einer (festen) Funktion $F : \mathbb{R}^{(p+q+1)m} \rightarrow \mathbb{R}^m$ dargestellt werden kann.

Eine (4.4) entsprechende Kurzschreibweise ist

$$U_j^{n+1} = U_j^n - \frac{k}{h} [F(U^n; j) - F(U^n; j-1)], \quad (4.5)$$

wobei U^n wieder den Vektor mit allen Näherungen aus der n -ten Zeitschicht bezeichnet. Somit soll F den mittleren Fluss durch die Orte $x_{j+\frac{1}{2}}$ über die Zeit $[t_n, t_{n+1}]$ approximieren

$$F(U^n; j) \approx \frac{1}{k} \int_{t_n}^{t_{n+1}} f(u(x_{j+\frac{1}{2}}, t)) dt. \quad (4.6)$$

Daher wird F in diesem Zusammenhang als *numerische Flussfunktion* des Verfahrens bezeichnet. Im einfachsten aber häufigen Fall hängt F nur von zwei Variablen ab, d.h. 4.4 reduziert sich zu ($p = 0, q = 1$)

$$U_j^{n+1} = U_j^n - \frac{k}{h} [F(U_j^n, U_{j+1}^n) - F(U_{j-1}^n, U_j^n)]. \quad (4.7)$$

Bei numerischer Integration ist die elementarste Konsistenzbedingung, dass eine konstante Funktion exakt integriert wird. Für konstante Eingangsdaten soll daher die Approximation in (4.6) exakt sein, d.h.

$$F(\hat{u}, \hat{u}, \dots, \hat{u}) = f(\hat{u}) \quad \text{für alle (relevanten) } \hat{u} \in \mathbb{R}^m. \quad (4.8)$$

Für die Konvergenz der Verfahren wird hier noch benötigt, dass F den Wert $f(\hat{u})$ gleichförmig erreicht. Daher fordern wir eine lokale Lipschitz-Bedingung.

Definition 4.2 Ein konservatives Verfahren (4.4) ist konsistent, wenn die lokale Lipschitz-Bedingung

$$\|F(U_{j-p}, U_{j-p+1}, \dots, U_{j+q}) - f(\hat{u})\| \leq C \max_{-p \leq i \leq q} \|U_{j+i} - \hat{u}\| \quad (4.9)$$

für alle U_{j+i} in einer Umgebung eines (relevanten) $\hat{u}$ in einer beliebigen Vektornorm erfüllt ist.

Wir bemerken, dass hier keine Definition einer Konsistenzordnung erfolgt. Für Konsistenz höherer Ordnung müsste nämlich wieder die Glattheit der Lösung vorausgesetzt werden und dann wäre der übliche Konsistenzbegriff für Lösungen der Differentialgleichung anwendbar.

Die Lipschitz-Bedingung (4.9) ist hinreichend für die elementare Bedingung (4.8). Hinreichend für die Lipschitz-Bedingung (4.9) ist die Bedingung (4.8) zusammen mit $F \in C^1$. Da F meistens nur von f und eventuell Ableitungen dazu abhängt, impliziert eine hinreichende Glattheit von f dann $F \in C^1$.

Satz 4.3 *Gegeben sei ein Anfangswertproblem einer Erhaltungsgleichung, wobei u_0 außerhalb eines endlichen Intervalls konstant ist mit Werten $u_{-\infty}, u_{+\infty}$. Ein konsistentes, konservatives Verfahren erfüllt das diskrete Erhaltungsprinzip*

$$h \sum_{j=J}^K U_j^N = h \sum_{j=J}^K U_j^L - (t_N - t_L) [f(u_{+\infty}) - f(u_{-\infty})] \quad (4.10)$$

für $0 \leq L < N$ jeweils mit $J \ll 0 \ll K$ hinreichend groß.

Beweis:

Ein konservatives Verfahren besitzt die Gestalt (4.4). Aufsummieren der Formeln über beliebige Zellenindizes $J, J+1, \dots, K$ liefert

$$h \sum_{j=J}^K U_j^{n+1} = h \sum_{j=J}^K U_j^n - k \sum_{j=J}^K [F(U^n; j) - F(U^n; j-1)],$$

wobei die letzte Summe sich vereinfacht zu

$$h \sum_{j=J}^K U_j^{n+1} = h \sum_{j=J}^K U_j^n - k [F(U^n; K) - F(U^n; J-1)].$$

Damit verbleiben nur noch die numerischen Flüsse durch die Ortspunkte $x_{J-\frac{1}{2}}$ und $x_{K+\frac{1}{2}}$. Mit der Annahme, dass J, K weit genug gewählt werden, damit u in einer Umgebung dieser Orte über eine gewisse Zeit konstant ist, folgt mit der Konsistenz (4.8) des Verfahrens

$$h \sum_{j=J}^K U_j^{n+1} = h \sum_{j=J}^K U_j^n - k [f(u_{+\infty}) - f(u_{-\infty})].$$

Rekursiv impliziert dies dann für die Näherungen zu den Zeitschichten $L < N$, d.h. $t_L = Lk$, $t_N = Nk$, nun die diskrete Erhaltung (4.10). $\square$

Hat die Anfangsfunktion u_0 einen kompakten Träger, dann folgt $u_{-\infty} = u_{+\infty} = 0$. In diesem Fall folgt

$$h \sum_{j=J}^K U_j^N = h \sum_{j=J}^K U_j^L$$

für alle $0 \leq L < N$.

Satz 4.4 *Sind die Voraussetzungen aus Satz 4.3 erfüllt und werden als Anfangswerte im Verfahren die exakten Zellenmittel verwendet, dann folgt*

$$h \sum_{j=J}^K U_j^n = \int_{x_{J-\frac{1}{2}}}^{x_{K+\frac{1}{2}}} u(x, t_n) \, dx \quad (4.11)$$

für alle $n \in \mathbb{N}$ mit einer (beliebigen) schwachen Lösung u .

Beweis:

Im Verfahren werden als Anfangswerte die exakten Zellenmittel verwendet, d.h. $U_j^0 = \bar{u}_j^0$ für alle $j \in \mathbb{Z}$. Damit folgt

$$h \sum_{j=J}^K U_j^0 = \int_{x_{J-\frac{1}{2}}}^{x_{K+\frac{1}{2}}} u_0(x) \, dx.$$

Die Integralform (1.3) der Erhaltungsgleichung liefert für eine beliebige schwache Lösung

$$\begin{aligned} \int_a^b u(x, d) \, dx &= \int_a^b u(x, c) \, dx \\ &\quad - \left[\int_c^d f(u(b, t)) \, dt - \int_c^d f(u(a, t)) \, dt \right] \end{aligned}$$

für beliebige $a < b$, $c < d$. Die Anfangswerte u_0 besitzen außerhalb eines endlichen Intervalls die konstanten Zustände $u_{-\infty}, u_{+\infty}$. Zu festem d seien daher a, b hinreichend weit gewählt, dass u an diesen Orten die konstanten Zustände über die gesamte Zeit $[c, d]$ annimmt. Dadurch folgt

$$\int_a^b u(x, d) \, dx = \int_a^b u(x, c) \, dx - (d - c) [f(u_{+\infty}) - f(u_{-\infty})]. \quad (4.12)$$

Zusammen mit den Integrationsgrenzen $a = x_{J-\frac{1}{2}}$, $b = x_{K+\frac{1}{2}}$, $c = 0$, $d = t_n$

in (4.12) und $t_L = 0$, $t_N = t_n$ in (4.10) folgt somit

$$\begin{aligned} \int_{x_{J-\frac{1}{2}}}^{x_{K+\frac{1}{2}}} u(x, t_n) \, dx &= \int_{x_{J-\frac{1}{2}}}^{x_{K+\frac{1}{2}}} u_0(x) \, dx - t_n [f(u_{+\infty}) - f(u_{-\infty})] \\ &= h \sum_{j=J}^K U_j^0 - t_n [f(u_{+\infty}) - f(u_{-\infty})] = h \sum_{j=J}^K U_j^n, \end{aligned}$$

was zu zeigen war. $\square$

Mit der Definition der Näherungsfunktion (3.2) ist (4.11) äquivalent zu

$$\int_{x_{J-\frac{1}{2}}}^{x_{K+\frac{1}{2}}} U_k(x, t_n) \, dx = \int_{x_{J-\frac{1}{2}}}^{x_{K+\frac{1}{2}}} u(x, t_n) \, dx. \quad (4.13)$$

Diese Relation bedeutet, dass das Integral über die Näherungsfunktion dem exakten Integral über die Erhaltungsgrößen entspricht. Das numerische Verfahren erfüllt also genau die globale Erhaltung der Größen in der Diskretisierung und verdient daher die Bezeichnung konservativ.

Aus der diskreten Erhaltung folgt insbesondere, dass ein konservatives Verfahren eine Schockwelle an der korrekten Stelle berechnet. Würde sich die Unstetigkeit in der numerischen Näherung mit falscher Geschwindigkeit bewegen, dann wäre die diskrete Erhaltung verletzt. Die Unstetigkeit kann möglicherweise durch ein konservatives Verfahren abgeglättet werden, jedoch geschieht dies nur um die korrekte Position in symmetrischer Weise.

4.2 Herleitung konservativer Verfahren

In diesem Abschnitt sollen bereits eingeführte Methoden für lineare Systeme auf den nichtlinearen Fall übertragen werden. Eine offensichtliche Verallgemeinerung des (expliziten) *Euler-Verfahrens* (3.4) ist

$$U_j^{n+1} = U_j^n - \frac{k}{2h} [f(U_{j+1}^n) - f(U_{j-1}^n)]. \quad (4.14)$$

Diese Methode kann in konservativer Form geschrieben werden mit der numerischen Flussfunktion

$$F(U_j, U_{j+1}) = \frac{1}{2} [f(U_j) + f(U_{j+1})].$$

Die numerische Flussfunktion approximiert den exakten Fluss an der Stelle $x_{j+\frac{1}{2}}$ durch das arithmetische Mittel der Flüsse in den benachbarten Gitterpunkten. Die Konsistenz (4.8) ist sofort wegen $F(\hat{u}, \hat{u}) = f(\hat{u})$ gegeben. Leider ist das

Euler-Verfahren bereits im linearen Fall instabil und damit auch im nichtlinearen Fall unbrauchbar.

Analog modifiziert sich das *Lax-Friedrichs Verfahrens* (3.6) bei nichtlinearen Problemen zu

$$U_j^{n+1} = \frac{1}{2} (U_{j-1}^n + U_{j+1}^n) - \frac{k}{2h} [f(U_{j+1}^n) - f(U_{j-1}^n)]. \quad (4.15)$$

Eine konservative Form dieser Methode liegt mit der numerischen Flussfunktion

$$F(U_j, U_{j+1}) = \frac{h}{2k} (U_j - U_{j+1}) + \frac{1}{2} [f(U_j) + f(U_{j+1})]$$

vor. Die Approximation des Flusses durch das arithmetische Mittel wird hier durch eine Differenz von Lösungswerten ungünstig verändert. Die Konsistenz wird jedoch beibehalten, da sich diese Differenz dann weghebt.

Zur Anwendung eines *Upwind Verfahrens* erscheint zunächst eine Linearisierung des nichtlinearen Systems (4.1) naheliegend, d.h. Verwendung von (4.2). Zur Bestimmung der Näherung U_j^{n+1} wird die Funktionalmatrix $A(U_j^n)$ herangezogen. Mit dieser konstanten Matrix kann lokal genau das Upwind Verfahren (3.29) aus dem linearen Fall angewendet werden. Man nennt dies die *Courant-Isaacson-Rees Methode*. Leider ist dieser Ansatz nicht konservativ und daher nur bei glatten Lösungen sinnvoll.

Zur Berechnung von schwachen Lösungen bei nichtlinearen Systemen ist eine direkte Verwendung des Upwind Verfahrens möglich, wenn die Eigenwerte der Funktionalmatrix stets gleiches Vorzeichen besitzen. Bei ausschließlich nichtnegativen oder nichtpositiven Eigenwerten lautet die numerische Flussfunktion

$$F(U_j, U_{j+1}) = f(U_j) \quad \text{bzw.} \quad F(U_j, U_{j+1}) = f(U_{j+1}).$$

Dadurch wird der Informationstransport entlang der Charakteristiken in korrekter Weise berücksichtigt. Die Konsistenz des Upwind Verfahrens ist trivial.

Für eine skalare Gleichung $u_t + f(u)_x = 0$ ist die Anwendung eines Upwind Verfahrens unproblematisch. Eine natürliche Verallgemeinerung des linearen Falls ist gegeben durch die numerische Flussfunktion

$$F(U_j, U_{j+1}) = \begin{cases} f(U_j) & , Df \geq 0 \\ f(U_{j+1}) & , Df < 0 \end{cases} \quad \text{mit} \quad Df = \frac{f(U_{j+1}) - f(U_j)}{U_{j+1} - U_j}.$$

Dabei wird durch numerische Differentiation eine Näherung für die Ableitung von f an der Zellengrenze ermittelt. Anhand des Vorzeichens wird dann die korrekte Richtung des Informationstransports verwendet. Im Fall $U_j = U_{j+1}$ ist die Wahl der Richtung beliebig.

Auch Verfahren zweiter Ordnung lassen sich für glatte Lösungen herleiten. Ist u eine exakte Lösung des nichtlinearen Systems (4.1), dann gilt analog zu (3.8)

$$u_{tt} = -f(u)_{xt} = -f(u)_{tx} = -(A(u)u_t)_x = (A(u)f(u)_x)_x, \quad (4.16)$$

wobei $A(u)$ die Funktionalmatrix von f bezeichnet. Eine Verallgemeinerung des *Lax-Wendroff Verfahrens* auf nichtlineare Systeme ist daher über (3.7) gegeben durch

$$\begin{aligned} U_j^{n+1} = & U_j^n - \frac{k}{2h} (f(U_{j+1}^n) - f(U_{j-1}^n)) \\ & + \frac{k^2}{2h^2} \left[A\left(\frac{1}{2}(U_j^n + U_{j+1}^n)\right) \cdot (f(U_{j+1}^n) - f(U_j^n)) \right. \\ & \left. - A\left(\frac{1}{2}(U_{j-1}^n + U_j^n)\right) \cdot (f(U_j^n) - f(U_{j-1}^n)) \right]. \end{aligned} \quad (4.17)$$

Dieses Verfahren ist zur nichtlinearen Gleichung bei hinreichend glatten Lösungen konsistent von zweiter Ordnung. Für eine lineare Flussfunktion $f(u) = Au$ mit A konstant reduziert sich die Methode auf das gewöhnliche Lax-Wendroff Verfahren. Zudem ist die Vorschrift konservativ und konsistent gemäß (4.8), lässt sich also auch zur Bestimmung von schwachen Lösungen verwenden. Ein deutlicher Nachteil von (4.17) ist jedoch, dass die Funktionalmatrix $A(\cdot)$ ausgewertet werden muss. Dies erfordert im allgemeinen numerische Differentiation und erhöht somit den Rechenaufwand deutlich. Die Auswertungen von $A(\cdot)$ sollen daher durch Anwendung geeigneter Differenzenformeln vermieden werden, woraus Verfahren mit einem Zwischenschritt entstehen. Das *Richtmyer Zwei-Schritt Lax-Wendroff Verfahren* ist gegeben durch die Vorschrift

$$\begin{aligned} U_{j+\frac{1}{2}}^{n+\frac{1}{2}} &= \frac{1}{2} (U_j^n + U_{j+1}^n) - \frac{k}{2h} [f(U_{j+1}^n) - f(U_j^n)] \\ U_j^{n+1} &= U_j^n - \frac{k}{h} \left[f\left(U_{j+\frac{1}{2}}^{n+\frac{1}{2}}\right) - f\left(U_{j-\frac{1}{2}}^{n+\frac{1}{2}}\right) \right]. \end{aligned}$$

Eine weitere Methode dieser Art ist das *MacCormack Verfahren*, bei dem zuerst Vorwärtsdifferenzen und dann Rückwärtsdifferenzen die Formel

$$\begin{aligned} U_j^* &= U_j^n - \frac{k}{h} [f(U_{j+1}^n) - f(U_j^n)] \\ U_j^{n+1} &= \frac{1}{2} (U_j^n + U_j^*) - \frac{k}{2h} [f(U_j^*) - f(U_{j-1}^*)] \end{aligned}$$

liefern. Alternativ führen hier erst Rückwärtsdifferenzen und dann Vorwärtsdifferenzen auf

$$\begin{aligned} U_j^* &= U_j^n - \frac{k}{h} [f(U_j^n) - f(U_{j-1}^n)] \\ U_j^{n+1} &= \frac{1}{2} (U_j^n + U_j^*) - \frac{k}{2h} [f(U_{j+1}^*) - f(U_j^*)]. \end{aligned} \quad (4.18)$$

Alle diese Verfahren sind konsistent von zweiter Ordnung bei glatten Lösungen. Sie reduzieren sich bei linearen Problemen alle auf das gewöhnliche Lax-Wendroff Verfahren. Zudem sind sie konservativ und konsistent bzgl. der numerischen Flussfunktion.

4.3 Lax-Wendroff Theorem

Das diskrete Erhaltungsprinzip legt nahe, dass sich konservative Verfahren zur Bestimmung von schwachen Lösungen bei Erhaltungsgleichungen eignen. Das *Lax-Wendroff Theorem* gibt diesbezüglich eine Konvergenzaussage, dass wenn die Näherungen aus einem konservativen Verfahren für bestimmte Folgen von Schrittweiten $k, h \rightarrow 0$ gegen eine Funktion konvergieren, dann diese Grenzfunktion eine schwache Lösung der Erhaltungsgleichung repräsentiert. Das Theorem garantiert nicht, dass die Näherungen aus dem Verfahren überhaupt konvergieren, denn dazu sind noch weitere nichtlineare Stabilitätskonzepte notwendig. Dennoch liefert das Theorem von Lax-Wendroff eine wichtige Aussage.

Im folgenden betrachten wir den skalaren Fall. Alle Aussagen lassen sich jedoch direkt auf Systeme übertragen, indem man komponentenweise arbeitet. Die Konvergenz der Näherungen ist noch zu spezifizieren. Es muss hier nicht notwendigerweise $\frac{k}{h}$ konstant sein. Desweiteren braucht die Konvergenz nicht für alle $k, h \rightarrow 0$ vorzuliegen, sondern nur für eine Teilfolge. Zu einer Folge von Gittern mit ℓ numeriert seien die Näherungsfunktionen mit U_ℓ bezeichnet

$$U_\ell(x, t) = U_{j,\ell}^n \quad \text{für } (x, t) \in [(j - \frac{1}{2})h_\ell, (j + \frac{1}{2})h_\ell] \times [nk_\ell, (n + 1)k_\ell] \quad (4.19)$$

mit $U_{j,\ell}^n$ jeweils auf dem ℓ -ten Gitter berechnet. Zudem soll sich die totale Variation der Näherungsfunktionen gleichmäßig beschränken lassen. Die totale Variation einer Funktion $v : \mathbb{R} \rightarrow \mathbb{R}$ ist dabei definiert durch

$$\text{TV}(v) = \sup \left\{ \sum_{j=1}^N |v(\xi_j) - v(\xi_{j-1})| : \xi_0 < \xi_1 < \dots < \xi_N, N \in \mathbb{N} \right\}. \quad (4.20)$$

Damit dieses Supremum existiert, muss notwendigerweise $v(x)$ für $x \rightarrow \pm\infty$ jeweils konvergieren. Alternativ ist eine Definition möglich über

$$\text{TV}(v) = \limsup_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_{-\infty}^{+\infty} |v(x) - v(x - \varepsilon)| \, dx. \quad (4.21)$$

Mit $v \in C^1$ ergibt sich daraus für die totale Variation

$$\text{TV}(v) = \int_{-\infty}^{+\infty} |v'(x)| \, dx. \quad (4.22)$$

Für die stückweise konstante Näherungsfunktion (4.19) folgt nach (4.20) die einfachere Gestalt

$$\text{TV}(U_\ell(\cdot, t)) = \sum_{j=-\infty}^{+\infty} |U_{j+1,\ell}^n - U_{j,\ell}^n| \quad \text{für } t \in [t_n, t_{n+1}). \quad (4.23)$$

Für einen Vektor $U^n \in \mathbb{R}^\infty$ mit den Näherungen aus der n -ten Zeitschicht ist daher eine sinnvolle Definition der totalen Variation gegeben durch

$$\text{TV}(U^n) = \sum_{j=-\infty}^{+\infty} |U_{j+1}^n - U_j^n|. \quad (4.24)$$

Nun können wir das Lax-Wendroff Theorem formulieren.

Satz 4.5 (Lax-Wendroff) *Gegeben sei eine Folge von Gittern durchnummeriert mit $\ell \in \mathbb{N}$, wobei die Gitterweiten $k_\ell, h_\ell \rightarrow 0$ für $\ell \rightarrow \infty$ erfüllen. Es bezeichne $(U_\ell)_{\ell \in \mathbb{N}}$ die Folge der Näherungsfunktionen (4.19), welche mit einem konservativen und konsistenten Verfahren entsprechend (4.4) bzw. (4.9) berechnet werden. Die Folge $(U_\ell)_{\ell \in \mathbb{N}}$ konvergiere gegen eine lokal integrierbare Funktion u in folgender Weise:*

1. *Über jeden endlichen Bereich $\Omega = [a, b] \times [0, T]$ in der x - t -Ebene sei*

$$\lim_{\ell \rightarrow \infty} \int_0^T \int_a^b |U_\ell(x, t) - u(x, t)| \, dx \, dt = 0 \quad (4.25)$$

(Konvergenz in der Integralnorm auf der Menge Ω).

2. *Zu jedem $T > 0$ existiere ein $R > 0$, so dass*

$$\text{TV}(U_\ell(\cdot, t)) < R \quad \text{für alle } 0 \leq t \leq T \text{ und alle } \ell \in \mathbb{N} \quad (4.26)$$

(Gleichmäßige Beschränkung der totalen Variation).

Dann ist die Grenzfunktion u eine schwache Lösung der Erhaltungsgleichung.

Beweis:

Wir werden zeigen, dass die Grenzfunktion u die schwache Form (1.5) für alle Testfunktionen $\phi \in C_0^1$ erfüllt. Im folgenden vermeiden wir häufig den Gitterindex ℓ aus Gründen der Übersichtlichkeit.

Multiplikation der Formel (4.4) des konservativen Verfahrens mit einer Testfunktion liefert bei der abkürzenden Schreibweise (4.5)

$$\phi(x_j, t_n) U_j^{n+1} = \phi(x_j, t_n) U_j^n - \frac{k}{h} \phi(x_j, t_n) [F(U^n; j) - F(U^n; j-1)]$$

für alle j und n auf dem ℓ -ten Gitter. Summation über alle $j \in \mathbb{Z}$ und $n \in \mathbb{N}$ ergibt

$$\begin{aligned} & \sum_{n=0}^{\infty} \sum_{j=-\infty}^{+\infty} \phi(x_j, t_n) (U_j^{n+1} - U_j^n) \\ &= -\frac{k}{h} \sum_{n=0}^{\infty} \sum_{j=-\infty}^{+\infty} \phi(x_j, t_n) [F(U^n; j) - F(U^n; j-1)]. \end{aligned} \quad (4.27)$$

Die Funktion ϕ besitzt kompakten Träger, d.h. es ist $\phi \neq 0$ lediglich in einem Bereich $[-M, M] \times [-T, T]$. Dadurch sind nur endlich viele Summanden hier ungleich null. Partielle Summation gemäß der Formel

$$\sum_{i=1}^I a_i (b_i - b_{i-1}) = a_I b_I - a_1 b_0 - \sum_{i=1}^{I-1} (a_{i+1} - a_i) b_i \quad (4.28)$$

angewendet auf jeweils die Summe bzgl. n links und j rechts in (4.27) liefert

$$\begin{aligned} & - \sum_{j=-\infty}^{+\infty} \left[\phi(x_j, t_0) U_j^0 + \sum_{n=1}^{\infty} (\phi(x_j, t_n) - \phi(x_j, t_{n-1})) U_j^n \right] \\ &= \frac{k}{h} \sum_{n=0}^{\infty} \sum_{j=-\infty}^{+\infty} (\phi(x_{j+1}, t_n) - \phi(x_j, t_n)) F(U^n; j). \end{aligned}$$

Dabei wurden die Summationsgrenzen gemäß (4.28) hinreichend weit gewählt, so dass aufgrund des kompakten Trägers von ϕ alle Randterme bis auf $n = 0$ wegfallen. Dieser Ausdruck kann umgeformt werden zu

$$\begin{aligned} & hk \sum_{n=1}^{\infty} \sum_{j=-\infty}^{+\infty} \frac{1}{k} (\phi(x_j, t_n) - \phi(x_j, t_{n-1})) U_j^n \\ &+ hk \sum_{n=0}^{\infty} \sum_{j=-\infty}^{+\infty} + \frac{1}{h} (\phi(x_{j+1}, t_n) - \phi(x_j, t_n)) F(U^n; j) = -h \sum_{j=-\infty}^{+\infty} \phi(x_j, 0) U_j^0. \end{aligned}$$

Unter Beachtung, dass $U_j^n = U_\ell(x_j, t_n)$ ist, folgt sofort

$$\begin{aligned} & \lim_{\ell \rightarrow \infty} h_\ell k_\ell \sum_{n=1}^{\infty} \sum_{j=-\infty}^{+\infty} \frac{1}{k_\ell} (\phi(x_j, t_n) - \phi(x_j, t_{n-1})) U_\ell(x_j, t_n) \\ &= \int_0^\infty \int_{-\infty}^{+\infty} \phi_t(x, t) u(x, t) \, dx \, dt \end{aligned}$$

als Grenzwert reeller Zahlen. Werden als Anfangswerte die exakten Zellenmittel verwendet $U_j^0 = \bar{u}_j^0$, dann gilt

$$\lim_{\ell \rightarrow \infty} -h_\ell \sum_{j=-\infty}^{+\infty} \phi(x_j, 0) U_j^0 = - \int_{-\infty}^{+\infty} \phi(x, 0) u(x, 0) \, dx.$$

Kritisch ist nur der verbleibende Term mit F . Über die Konsistenz der numerischen Flussfunktion haben wir jedoch nach (4.9)

$$\begin{aligned} & |F(U_\ell(x_j - ph_\ell, t_n), \dots, U_\ell(x_j + qh_\ell, t_n)) - f(U_\ell(x_j, t_n))| \\ & \leq C \max_{-p \leq i \leq q} |U_\ell(x + ih_\ell, t_n) - U_\ell(x_j, t_n)|. \end{aligned}$$

Wegen der Beschränkung der totalen Variation nach (4.26) muss jedoch für jedes $t \in [0, T]$

$$\lim_{\ell \rightarrow 0} \max_{-p \leq i \leq q} |U_\ell(x + ih_\ell, t) - U_\ell(x, t)| = 0$$

punktweise für fast alle x gelten. Andernfalls würde nämlich die totale Variation beliebig hoch ausfallen. Damit kann die numerische Flussfunktion durch die tatsächliche Flussfunktion approximiert werden, wobei die Fehler gleichmäßig fast überall verschwinden. Dadurch gilt

$$\begin{aligned} & \lim_{\ell \rightarrow 0} h_\ell k_\ell \sum_{n=0}^{\infty} \sum_{j=-\infty}^{+\infty} \frac{1}{h_\ell} (\phi(x_{j+1}, t_n) - \phi(x_j, t_n)) F(U^n; j) \\ & = \int_0^\infty \int_{-\infty}^{+\infty} \phi_x(x, t) f(u(x, t)) \, dx \, dt. \end{aligned}$$

Also ist (1.5) für eine beliebige Testfunktion erfüllt und somit u eine schwache Lösung der Erhaltungsgleichung. $\square$

Eine naheliegende Frage ist nun, unter welchen Umständen garantiert werden kann, dass bei Konvergenz des Verfahrens die Grenzfunktion physikalisch sinnvoll ist, d.h. eine Entropielösung der Erhaltungsgleichung darstellt. Eine Entropielösung liegt genau dann vor, wenn die schwache Form (2.28) für beliebige konvexe Entropiefunktionen η mit korrespondierendem Entropiefluss ψ und alle nichtnegativen Testfunktionen erfüllt ist. Als hinreichend erweist sich die folgende Bedingung: Für ein numerisches Verfahren (4.4) sei bei beliebig vorgegebener konvexer Entropiefunktion η die *diskrete Entropieungleichung*

$$\eta(U_j^{n+1}) \leq \eta(U_j^n) - \frac{k}{h} [\Psi(U^n; j) - \Psi(U^n; j-1)] \quad (4.29)$$

mit einer numerischen Entropieflussfunktion Ψ , die konsistent zu ψ in gleicher Weise wie F konsistent zu f ist, stets erfüllt. Mit dieser Voraussetzung kann genau analog zum Beweis des Lax-Wendroff Theorems gezeigt werden, dass die Grenzfunktion der Entropiebedingung (2.28) genügt, also eine Entropielösung darstellt. Die Bedingung (4.29) allein gewährleistet jedoch nicht, dass eine Teilfolge der Gitterfunktionen gegen eine Entropielösung konvergiert oder dass sie überhaupt konvergiert.

4.4 Nichtlineare Stabilitätskonzepte

Das Theorem von Lax-Wendroff (Satz 4.5) gibt keine hinreichenden Kriterien für die Konvergenz eines konsistenten Verfahrens. Der Äquivalenzsatz von Lax (Satz 3.4) lässt sich leider nicht auf nichtlineare Probleme verallgemeinern, da er die Linearität der Erhaltungsgleichung wesentlich verwendet. Daher sind schärfere Stabilitätskonzepte als im linearen Fall notwendig, um die Konvergenz eines konsistenten Verfahrens im nichtlinearen Fall zu garantieren.

Der Lax-Wendroff Theorem gilt auch für nichtlineare Systeme. Demgegenüber gelten die folgenden Untersuchungen nur für skalare nichtlineare Gleichungen. Dies ist typisch für kompliziertere Aufgabenstellungen bei Erhaltungsgleichungen. Vollständige Analysen und zufriedenstellende Aussagen lassen sich meist nur für lineare Systeme oder skalare nichtlineare Gleichungen herleiten. Dies gilt sowohl für Existenz- und Eindeutigkeitsfragen von Lösungen der Erhaltungsgleichungen als auch für die Untersuchung von numerischen Methoden. In der Numerik ist man bereits zufrieden, wenn ein Verfahren für lineare Systeme oder skalare nichtlineare Gleichungen als konvergent nachgewiesen wird und dann in direkter Weise auf nichtlineare Systeme verallgemeinert werden kann, auch ohne dass die Konvergenz dort gesichert ist.

Zur Untersuchung der Konvergenz wird wieder jeweils nur ein endliches Zeitintervall $[0, T]$ betrachtet. Da wir von schwachen Lösungen ausgehen, wird die Konvergenz in der Integralnorm

$$\|v\|_{1,T} = \int_0^T \|v(\cdot, t)\|_1 dt = \int_0^T \int_{-\infty}^{+\infty} |v(x, t)| dx dt \quad (4.30)$$

betrachtet. Zugrunde gelegt sei daher der Funktionenraum

$$\mathcal{L}_{1,T} = \{v : \mathbb{R} \times [0, T] \rightarrow \mathbb{R} : v \text{ messbar und } \|v\|_{1,T} < \infty\}.$$

Ein wesentliches Problem gegenüber dem linearen Fall besteht darin, dass für die nichtlineare Erhaltungsgleichung $u_t + f(u)_x = 0$ die Menge $\mathcal{W}$ aller schwachen Lösungen $u : \mathbb{R} \times \mathbb{R}_0^+ \rightarrow \mathbb{R}$, die gleiche Anfangswerte $u(x, 0) = u_0(x)$ besitzen, im allgemeinen mehrere Funktionen enthält. Man kann daher nicht erwarten, dass eine Folge von Näherungsfunktionen U_k definiert durch (3.2) bei Schrittweiten $k \rightarrow 0$ mit $\frac{k}{h}$ konstant gegen eine bestimmte schwache Lösung konvergiert. Stattdessen ist der globale Fehler zu ersetzen durch die *Distanz*

$$\text{dist}(U_k, \mathcal{W}) = \inf_{u \in \mathcal{W}} \|U_k - u\|_{1,T},$$

d.h. den Abstand der Näherung zur Menge aller schwachen Lösungen.

Im Lax-Wendroff Theorem wurde neben der Konvergenz der Folge von Näherungsfunktionen auch gefordert, dass sich die totale Variation im Zeitintervall

$[0, T]$ gleichmäßig beschränken lässt. Dies erweist sich als die entscheidende Forderung für die Stabilität eines konservativen Verfahrens im nichtlinearen Fall. Für diskrete Näherungswerte ist die totale Variation über (4.24) gegeben.

Definition 4.6 *Ein konservatives Verfahren (4.4) mit glatter numerischer Flussfunktion heißt TV-stabil, wenn für $T > 0$ und integrable Anfangswerte u_0 mit kompaktem Träger jeweils Konstanten $k_0, R > 0$ existieren, so dass*

$$\text{TV}(U^n) \leq R \quad \text{für alle } k, n \text{ mit } k < k_0 \text{ und } nk \leq T.$$

Nun kann der entscheidende Satz über die Konvergenz im nichtlinearen Fall formuliert werden.

Satz 4.7 *Sei U_k die von einem konservativen Verfahren (4.4) mit glatter numerischer Flussfunktion, welche konsistent gemäß Bedingung (4.9) ist, berechnete Näherungsfunktion. Ist die Methode TV-stabil, dann konvergiert die Methode für alle integrablen Anfangswerte u_0 mit kompaktem Träger, d.h. es gilt $\lim_{k \rightarrow 0} \text{dist}(U_k, \mathcal{W}) = 0$ in jedem festen Zeitintervall $[0, T]$.*

Beweisskizze:

Der kompakte Träger der Anfangswerte impliziert, dass alle Näherungsfunktionen $U_k(\cdot, t)$ kompakten Träger besitzen, da eine explizite Methode verwendet wird. Wie üblich sei $\frac{k}{h}$ konstant, so dass für alle $t \in [0, T]$ die Funktionen $U_k(\cdot, t)$ nur in einem festen Ortsintervall $[-M, M]$ ungleich null sind. Zusammen mit der TV-Stabilität folgt $U_k \in \mathcal{L}_{1,T}$ für alle $k < k_0$ mit einem hinreichend kleinem $k_0 > 0$. Darüberhinaus kann gezeigt werden, dass diese Eigenschaften sogar $U_k \in \mathcal{K}$ für alle $k < k_0$ mit einer kompakten Menge $\mathcal{K} \subset \mathcal{L}_{1,T}$ garantieren. Dabei geht auch die Glattheit der numerischen Flussfunktion ein.

Wir verwenden nun einen indirekten Beweis. Angenommen die Näherungsfunktionen würden nicht konvergieren. Dann existiert eine Folge $(U_{k_\ell})_{\ell \in \mathbb{N}}$ mit $k_\ell \rightarrow 0$ und ein $\varepsilon > 0$, so dass

$$\text{dist}(U_{k_\ell}, \mathcal{W}) = \inf_{u \in \mathcal{W}} \|U_{k_\ell} - u\|_{1,T} > \varepsilon \quad \text{für alle } \ell \in \mathbb{N}. \quad (4.31)$$

Da jedoch $U_{k_\ell} \in \mathcal{K}$ für alle $\ell \in \mathbb{N}$ gilt und $\mathcal{K}$ eine kompakte Menge des vollständigen Raums $\mathcal{L}_{1,T}$ ist, existiert eine konvergente Teilfolge. O.B.d.A. sei die Teilfolge $(U_{k_\ell})_{\ell \in \mathbb{N}}$ selbst. Die Folge konvergiert dann gegen eine Funktion $v \in \mathcal{K}$, d.h.

$$\lim_{\ell \rightarrow \infty} \|U_{k_\ell} - v\|_{1,T} = 0.$$

Die Voraussetzungen des Theorems von Lax-Wendroff (Satz 4.5) sind jedoch hier erfüllt. Die Grenzfunktion ist daher eine schwache Lösung der Erhaltungsgleichung, d.h. $v \in \mathcal{W}$. Damit erhalten wir

$$\lim_{\ell \rightarrow \infty} \text{dist}(U_{k_\ell}, \mathcal{W}) = 0$$

im Widerspruch zu (4.31). Also müssen alle Folgen von Näherungsfunktionen konvergieren. $\square$

Trotz der Mehrdeutigkeit der schwachen Lösung können wir daher erwarten, dass die Näherungen, die einmal auf einem feinen Gitter berechnet sind, bereits eine bestimmte schwache Lösung gut approximieren. Existiert eine eindeutige Lösung, welche die Entropiebedingung (2.28) erfüllt, dann können wir hoffen, durch geeignete Verfahren doch Konvergenz gegen genau diese Lösung zu erhalten.

Satz 4.7 wirft die Frage auf, wann die entscheidende Eigenschaft der TV-Stabilität von einem numerischen Verfahren erfüllt wird. Für eine exakte schwache Lösung u der Erhaltungsgleichung $u_t + f(u)_x = 0$ gilt

$$\text{TV}(u(\cdot, t_2)) \leq \text{TV}(u(\cdot, t_1)) \quad \text{für alle } t_1 \leq t_2.$$

Dies motiviert die folgende Klasse von numerischen Methoden.

Definition 4.8 *Ein (lineares oder nichtlineares) Verfahren der Form $U_j^{n+1} = \mathcal{H}_k(U^n; j)$ heißt TV-vermindernd, wenn für alle Gitterfunktionen U^n gilt*

$$\text{TV}(U^{n+1}) \leq \text{TV}(U^n).$$

Ein TV-verminderndes Verfahren laut Definition 4.8 ist offensichtlich auch TV-stabil laut Definition 4.6 und damit konvergent im konsistenten Fall. Eine weitere günstige Eigenschaft TV-vermindernder Methoden besteht darin, dass sie *Monotonie-erhaltend* sind. Näherungswerte $U^n \in \mathbb{R}^\infty$ sind *monoton* fallend bzw. steigend, wenn

$$U_j^n \geq U_{j+1}^n \quad \text{bzw.} \quad U_j^n \leq U_{j+1}^n \quad \text{für alle } j \in \mathbb{Z}$$

gilt.

Definition 4.9 *Ein (lineares oder nichtlineares) Verfahren der Form $U_j^{n+1} = \mathcal{H}_k(U^n; j)$ heißt Monotonie-erhaltend, wenn bei monotonen Anfangswerten U^0 die resultierenden Näherungen U^n für alle $n \in \mathbb{N}$ ebenfalls in gleicher Weise monoton sind.*

Ein TV-verminderndes Verfahren ist Monotonie-erhaltend, da eine Verletzung der Monotonie in späteren Zeitschichten sofort mit einem Anstieg der totalen Variation einhergeht.

Bei einem Monotonie-erhaltenden Verfahren kann insbesondere ausgeschlossen werden, dass sich an Unstetigkeitsstellen der Lösung inkorrekte Oszillationen bilden. Solche Oszillationen wurden für Verfahren zweiter Ordnung bereits bei der linearen Advektionsgleichung festgestellt. Die Monotonie der Anfangswerte aus dem Riemann-Problem bleibt dort nicht bestehen. Die Verfahren Lax-Wendroff und Beam-Warming sind daher nicht Monotonie-erhaltend.

Eine weitere wichtige Klasse von Methoden ist in der folgenden Definition gegeben.

Definition 4.10 *Ein (lineares oder nichtlineares) Verfahren der Form $U_j^{n+1} = \mathcal{H}_k(U^n; j)$ heißt monoton, wenn mit Vorgaben V^n, W^n und den resultierenden Näherungen $V_j^{n+1} = \mathcal{H}_k(V^n; j)$, $W_j^{n+1} = \mathcal{H}_k(W^n; j)$ stets die Implikation*

$$V_j^n \geq W_j^n \text{ für alle } j \in \mathbb{Z} \quad \Rightarrow \quad V_j^{n+1} \geq W_j^{n+1} \text{ für alle } j \in \mathbb{Z}$$

gilt.

Es kann gezeigt werden, dass ein monotones Verfahren auch TV-vermindernd und damit stabil ist. Jedoch können monotone Methoden bei glatten Lösungen nur konvergent von höchstens erster Ordnung sein. Beispiele für monotone Formeln sind das Lax-Friedrichs Verfahren und die Upwind Methoden. Die modifizierten Gleichungen von monotonen Verfahren besitzen die Gestalt

$$u_t + f(u)_x = k (g(u)u_x)_x \tag{4.32}$$

mit einer von f und dem Verfahren abhängigen Funktion g . Es gilt dabei im allgemeinen $g(u) > 0$. Diese modifizierten Gleichungen werden dann von höherer Ordnung von den Verfahren gelöst. Die rechte Seite von (4.32) repräsentiert einen Diffusionsterm. Für $k \rightarrow 0$ verschwindet diese numerische Diffusion. Daher konvergieren monotone Verfahren gegen die physikalisch sinnvolle Lösung bei verschwindender Diffusion, siehe (2.20). Dies ist ein großer Vorteil von monotonen Methoden.

Insgesamt gelten die Implikationen

$$\text{monoton} \quad \Rightarrow \quad \text{TV-vermindernd} \quad \Rightarrow \quad \text{Monotonie-erhaltend}$$

bei numerischen Verfahren. Da ein monotones Verfahren höchstens Konvergenzordnung 1 besitzt, können wir versuchen, genauere Methoden zu konstruieren, die nur TV-vermindernd oder Monotonie-erhaltend aber auch TV-stabil sind. Dabei liegt allerdings die folgende Einschränkung vor.

Satz 4.11 (Godunov) *Ein konvergentes lineares Verfahren $U_j^{n+1} = \mathcal{H}_k(U^n; j)$, welches Monotonie-erhaltend ist, besitzt bei glatten Lösungen höchstens die Konvergenzordnung 1.*

Der Satz folgt daraus, dass jede lineare Monotonie-erhaltende Methode sogar monoton ist. Dies impliziert leider, dass ein Monotonie-erhaltendes Verfahren höherer Ordnung bereits für die lineare Advektionsgleichung in nichtlinearer Weise auf den Anfangsdaten operieren muss. Dadurch entsteht eine Vielzahl an komplizierten numerischen Methoden.

5 Hochauflösende Verfahren

Schließlich möchten wir fortgeschrittene numerische Verfahren erhalten, welche einerseits in glatten Bereichen der Lösung von höherer Ordnung sind und andererseits Unstetigkeitsstellen korrekt auflösen.

5.1 Motivation

Bei der Diskussion von Riemann-Problemen für die lineare Advektionsgleichung haben wir festgestellt, dass Verfahren erster Ordnung die Unstetigkeit abglätten. Dagegen lösen Verfahren zweiter Ordnung die Unstetigkeit relativ scharf auf, produzieren jedoch inkorrekte Oszillationen. Dieses Verhalten ist unbefriedigend. Wir möchten Methoden höherer Ordnung konstruieren, die Unstetigkeiten scharf auflösen, jedoch dort keine Oszillationen hervorrufen.

Ein *hochauflösendes Verfahren* für Systeme aus Erhaltungsgleichungen soll die folgenden Forderungen erfüllen:

1. In Bereichen, wo die Lösung glatt ist, soll das Verfahren konvergent von (mindestens) zweiter Ordnung sein.
2. Unstetigkeitsstellen sollen scharf aufgelöst werden und dabei keine Oszillationen entstehen.
3. Das Verfahren soll konservativ sein.

Um die erste Bedingung zu berücksichtigen, müssen Approximationen zweiter Ordnung einbezogen werden. Das Vermeiden von Oszillationen aus der zweiten Forderung erreichen wir, indem das Verfahren TV-vermindernd konstruiert wird. Die dritte Bedingung garantiert das diskrete Erhaltungsprinzip für die Methode.

Ein TV-verminderndes Verfahren ist Monotonie-erhaltend. Nach dem Satz von Godunov (Satz 4.11) ist jedoch ein lineares, Monotonie-erhaltendes Verfahren bei glatten Lösungen höchstens von erster Ordnung konvergent. Um eine Methode zweiter Ordnung zu erreichen, die zudem TV-vermindernd wirkt, muss daher eine Vorschrift konstruiert werden, die bereits für die lineare Advektionsgleichung nichtlinear auf den Anfangswerten operiert. Daher entstehen komplizierte Ansätze für hochauflösende Verfahren. Die Forderung einer konservativen Methode bedeutet jedoch, dass auch diese Verfahren von der Form (4.5) mit einer bestimmten numerischen Flussfunktion sind.

Bei volldiskreten Schemata existieren die beiden Klassen der *Fluss-begrenzenden Methoden* und *Steigungs-begrenzenden Methoden*. Im folgenden werden die Verfahren für die lineare Advektionsgleichung hergeleitet. Danach wird jeweils kurz skizziert, wie die Schemata für lineare Systeme und dann auch nichtlineare Systeme verallgemeinert werden. Desweiteren gibt es auch hochauflösende Methoden, die auf Semi-Diskretisierung beruhen, nämlich die sogenannten *ENO Verfahren* (Essentially Non-Oscillatory).

5.2 Fluss-begrenzende Methoden

Die Idee der Fluss-begrenzenden Verfahren besteht darin, dass wir in Bereichen mit glatter Lösung einen numerischen Fluss F_H höherer Ordnung verwenden möchten, während nahe von Unstetigkeiten ein numerischer Fluss F_N gewählt werden soll, der die Unstetigkeit ohne Oszillationen relativ gut erfasst. Für F_N wird im allgemeinen eine monotone Methode benutzt, die daher nur von erster Ordnung, d.h. niedriger Ordnung, ist. Beispielsweise kann für F_H der Fluss aus dem Lax-Wendroff Verfahren und für F_N der Fluss aus dem Upwind Verfahren angesetzt werden.

Den Fluss höherer Ordnung kann man zusammen mit dem Fluss niedriger Ordnung in der Erweiterung

$$F_H(U; j) = F_N(U; j) + [F_H(U; j) - F_N(U; j)] \quad (5.1)$$

darstellen. Der Term $F_H - F_N$ repräsentiert eine Korrektur vom Fluss niedriger Ordnung hin zum Fluss höherer Ordnung. Die Größe dieser Korrektur wird nun mittels einer Funktion $\chi(U; j)$ begrenzt, d.h. es entsteht der neue numerische Fluss

$$F(U; j) = F_N(U; j) + \chi(U; j) [F_H(U; j) - F_N(U; j)]. \quad (5.2)$$

Man nennt in diesem Zusammenhang die Funktion χ den Fluss-Begrenzer. Ist die Lösung glatt im Bereich zwischen U_j und U_{j+1} , dann soll $\chi(U; j)$ Werte nahe 1 annehmen, während bei Unstetigkeiten $\chi(U; j)$ nahe null gewünscht ist. Jedoch erweist sich ein Wertebereich $[0, 1]$ für die Funktion χ als zu restriktiv. Auch Werte $\chi > 1$ in bestimmten Bereichen können zu geeigneten Verfahren führen.

Als Beispiel betrachten wir die lineare Advektionsgleichung $u_t + au_x = 0$ mit Transportgeschwindigkeit $a > 0$. Für F_H verwenden wir den Lax-Wendroff Fluss und für F_N den linksseitigen Upwind Fluss

$$\begin{aligned} F_H(U_j, U_{j+1}) &= \frac{1}{2}a(U_j + U_{j+1}) + \frac{k}{2h}a^2(U_j - U_{j+1}) \\ F_N(U_j, U_{j+1}) &= aU_j \end{aligned} \quad (5.3)$$

und erhalten daher über (5.2) den modifizierten Fluss

$$F(U; j) = aU_j + \frac{a}{2} \left(1 - \frac{ak}{h}\right) (U_{j+1} - U_j) \chi(U; j). \quad (5.4)$$

Zur Wahl des Begrenzers χ muss die Glattheit der Daten U_j quantitativ gemessen werden. Eine Möglichkeit hierzu liefert das Verhältnis

$$\theta_j = \frac{U_j - U_{j-1}}{U_{j+1} - U_j} \quad (5.5)$$

aus aufeinanderfolgenden Differenzen. Sind die Daten glatt, dann gilt auf einem feinen Gitter näherungsweise $U_{j+1} \approx U_j + \eta_j h$ und $U_{j-1} \approx U_j - \eta_j h$ mit einer Steigung η_j , und folglich ist θ_j nahe 1. Ist dagegen θ_j weit entfernt vom Wert 1, so deutet dies auf eine Unstetigkeit hin. Der Begrenzer soll daher als feste Funktion χ jeweils von den Größen θ_j abhängen, d.h.

$$\chi_j = \chi(U; j) = \chi(\theta_j). \quad (5.6)$$

Dabei muss die Funktion χ nichtlinear von den Daten U abhängen, da sonst ein lineares Verfahren vorliegt. Ein lineares und TV-verminderndes Schema kann nach dem Satz von Godunov (Satz 4.11) höchstens von erster Ordnung konvergent sein und wir erreichen so keine höhere Ordnung bei glatten Lösungen. Die Nichtlinearität von $\chi(U; j)$ resultiert hier jedoch direkt, da die Werte θ_j nach (5.5) nichtlinear von den Daten U abhängen.

Der obige Ansatz bricht jedoch in glatten Bereichen an lokalen Extrema der exakten Lösung u zusammen. Dort können nämlich Zähler oder Nenner in (5.5) nahe null sein und damit θ_j weit weg vom Wert 1, obwohl die Lösung glatt ist. Zudem sind bei Extremstellen negative Werte θ_j möglich. Daher geben wir uns mit Verfahren zufrieden, die nur in glatten Bereichen ohne Extremstellen von höherer Ordnung konvergent sind.

Die konservative Methode mit der numerischen Flussfunktion (5.4), wobei der Begrenzer über (5.6) festgelegt ist, besitzt nun die beiden günstigen Eigenschaften:

1. Sofern $\chi(\theta)$ eine beschränkte Funktion ist, erweist sich das Verfahren bereits konsistent zur Advektionsgleichung im Sinne von Definition 4.2.
2. Die Methode ist konsistent von zweiter Ordnung für glatte Lösungen u mit $u_x \geq \gamma > 0$, falls $\chi(1) = 1$ gilt und χ Lipschitz-stetig bei $\theta = 1$ ist.

Das gesamte konservative Verfahren besitzt hier die Formel

$$U_j^{n+1} = U_j^n - \left(\nu - \frac{1}{2}\nu(1 - \nu)\chi_{j-1}\right) (U_j^n - U_{j-1}^n) - \frac{1}{2}\nu(1 - \nu)\chi_j (U_{j+1}^n - U_j^n)$$

mit $\nu = a \frac{k}{h}$. Um inkorrekte Oszillationen an Unstetigkeitsstellen zu verhindern soll das gesamte Verfahren TV-vermindernd nach Definition 4.8 wirken. Eine Diskussion der obigen Formel zeigt, dass hierfür die CFL-Bedingung $|\nu| \leq 1$ zusammen mit

$$\left| \frac{\chi(\theta_j)}{\theta_j} - \chi(\theta_{j-1}) \right| \leq 2 \quad \text{für alle } \theta_{j-1}, \theta_j \in \mathbb{R} \quad (5.7)$$

hinreichend ist. Gilt $\theta_j \leq 0$ für ein j , so haben aufeinanderfolgende Steigungen unterschiedliches Vorzeichen und die Daten U besitzen dort eine Extremstelle. Wird dieses Extremum durch eine Oszillation aus dem Fluss höherer Ordnung verstärkt, so steigt sofort die totale Variation an. Um dies zu vermeiden, wird zur Sicherheit in solchen Bereichen nur der Fluss niedriger Ordnung benutzt. Im folgenden sei daher stets

$$\chi(\theta) = 0 \quad \text{für alle } \theta \leq 0$$

gesetzt. Es ist leicht zu erkennen, dass die Bedingungen

$$0 \leq \frac{\chi(\theta)}{\theta} \leq 2 \quad \text{und} \quad 0 \leq \chi(\theta) \leq 2 \quad \text{für alle } \theta \in \mathbb{R}$$

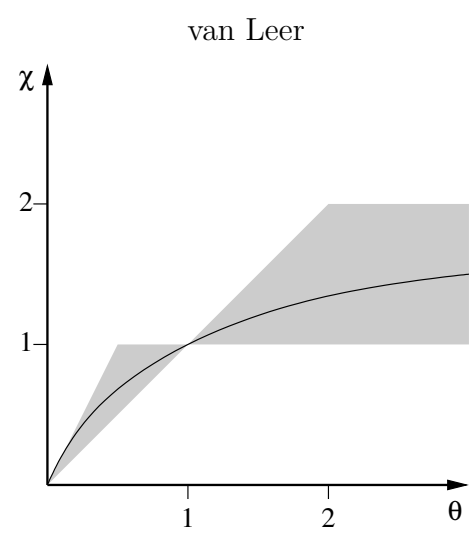
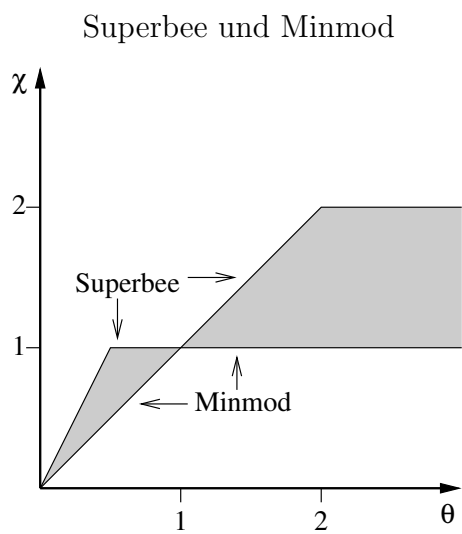
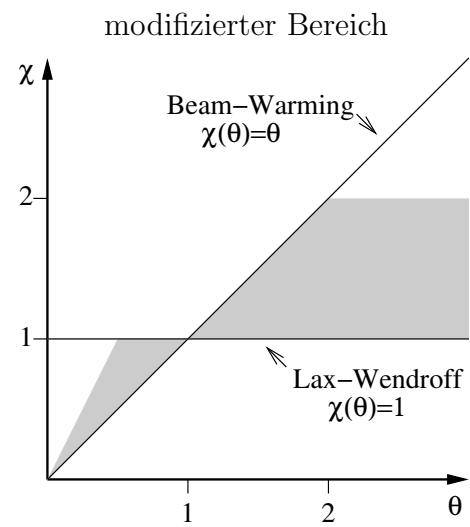
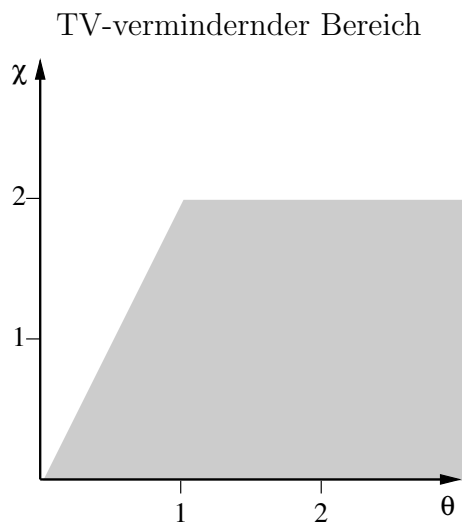
hinreichend für die Forderung (5.7) sind. Die beiden Bedingungen legen einen Wertebereich für χ mit Argumenten $\theta > 0$ fest, in dem die TV-Verminderung der Daten garantiert ist. Für die Konvergenz von zweiter Ordnung bei glatten Lösungen ist noch $\chi(1) = 1$ und die Lipschitz-Stetigkeit in $\theta = 1$ notwendig. Man beachte, dass die Wahl $\chi(\theta) \equiv 1$ die Lax-Wendroff Methode liefert und die Wahl $\chi(\theta) \equiv \theta$ auf das bekannte Beam-Warming Verfahren führt. Beides sind Verfahren zweiter Ordnung. Es hat sich als günstig erwiesen, als Werte für χ nur Konvexkombinationen aus diesen beiden Fällen zuzulassen, da ansonsten zuviel Kompression in der Methode entsteht. Dadurch resultiert ein modifizierter TV-vermindernder Wertebereich für den Fluss-Begrenzer.

In diesem Bereich sind nun viele Ansätze für die Funktion $\chi(\theta)$ denkbar. Es kann einfach der obere Rand des zulässigen Bereichs verwendet werden, wodurch der sogenannte *Superbee* Begrenzer

$$\chi(\theta) = \max\{0, \min\{1, 2\theta\}, \min\{\theta, 2\}\} = \begin{cases} 0 & , \theta \leq 0 \\ 2\theta & , 0 < \theta \leq \frac{1}{2} \\ 1 & , \frac{1}{2} < \theta \leq 1 \\ \theta & , 1 < \theta \leq 2 \\ 2 & , \theta > 2 \end{cases} \quad (5.8)$$

entsteht. Analog kann der untere Rand des Bereichs benutzt werden, was auf den *Minmod* Begrenzer

$$\chi(\theta) = \max\{0, \min\{1, \theta\}\} = \begin{cases} 0 & , \theta \leq 0 \\ \theta & , 0 < \theta \leq 1 \\ 1 & , \theta > 1 \end{cases} \quad (5.9)$$



Abbildungen: Konzepte für Fluss-begrenzende Verfahren.

führt. Beide Varianten sind stückweise linear und verlaufen Lipschitz-stetig durch $\chi(1) = 1$. Eine für $\theta > 0$ glatte Funktion χ ist durch den *van Leer* Begrenzer

$$\chi(\theta) = \frac{|\theta| + \theta}{1 + |\theta|} \quad (5.10)$$

gegeben. Trotz der Glattheit ist dieser nicht notwendig besser als die beiden vorherigen Ansätze. Alle drei Varianten haben sich in der Praxis bewährt.

Als Testbeispiel betrachten wir die Advektionsgleichung $u_t + au_x = 0$ mit $a = 1$. Die Anfangswerte bei $t = 0$ sind mit

$$u(x, 0) = \begin{cases} 1 - 0.1x & \text{für } x < 0 \\ -0.1x & \text{für } x > 0 \end{cases}$$

stückweise linear vorgegeben, um den trivialen Fall konstanter Daten im Riemann-Problem zu vermeiden. In den numerischen Verfahren verwenden wir stets die Zeitschrittweite $k = 0.02$ und die Ortsschrittweite $h = 0.04$. Wir betrachten dann die Näherungen zur Zeit $t = 1$. Das Upwind Verfahren glättet wie üblich die Unstetigkeit deutlich ab. Lax-Wendroff und Beam-Warming lösen die Unstetigkeit schärfer auf, produzieren aber inkorrekte Oszillationen. Die drei Flussbegrenzenden Methoden, also Superbee (5.8), Minmod (5.9) und van Leer (5.10), erfassen die Unstetigkeit relativ scharf und ohne Oszillationen. Insbesondere fällt die Abglättung in allen drei Verfahren signifikant geringer aus als im Upwind Verfahren.

Die Vorgehensweise läßt sich leicht auf eine Advektionsgleichung $u_t + au_x = 0$ mit beliebigem $a \in \mathbb{R}$ verallgemeinern. Je nach Vorzeichen von a wird für den Fluss niedriger Ordnung das entsprechende Upwind Verfahren benutzt. Der Fluss kann einheitlich geschrieben werden mit

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

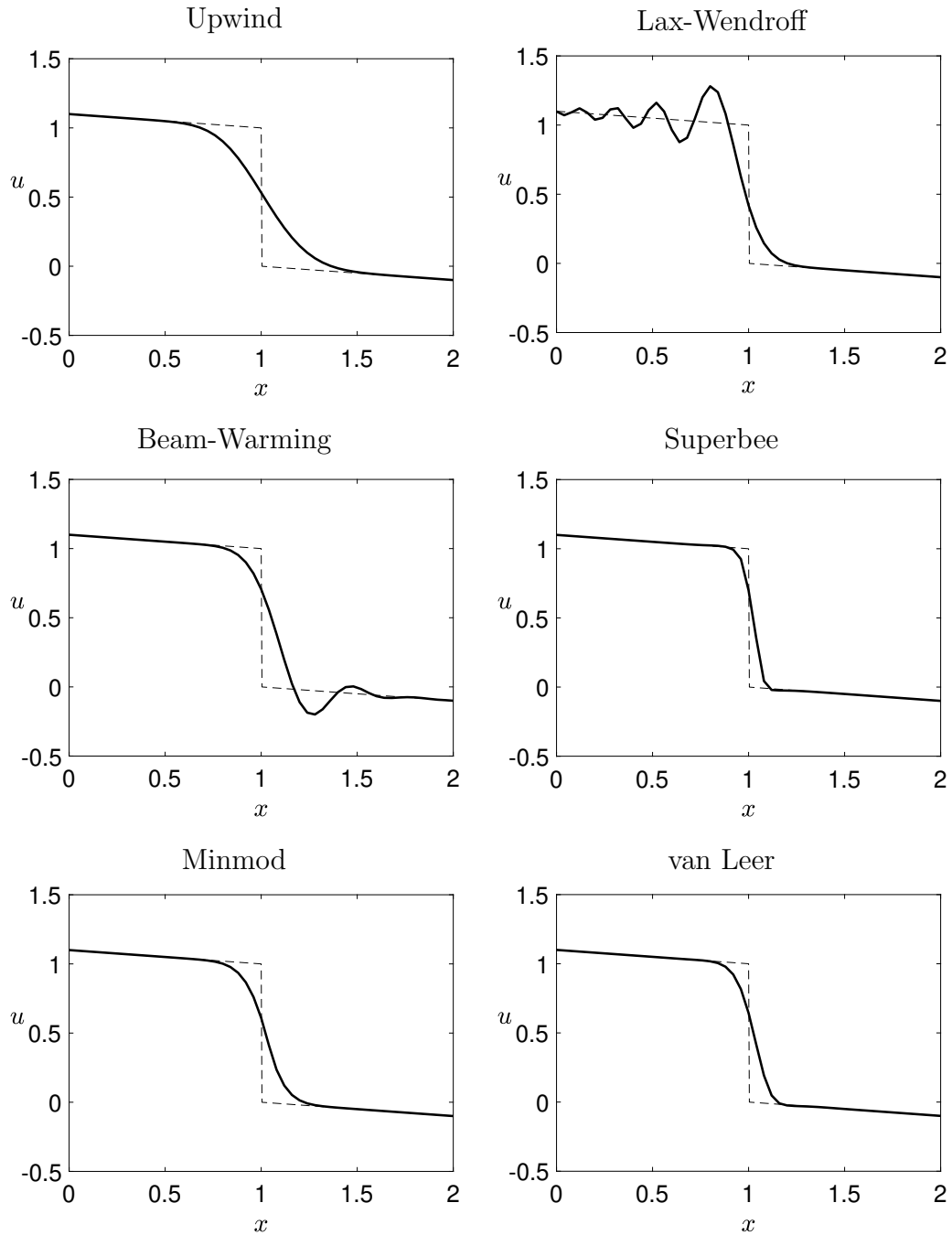
Wird für F_H wieder der Lax-Wendroff Fluss aus (5.3) verwendet, dann entsteht der modifizierte Fluss

$$F(U; j) = \frac{a}{2}(U_j + U_{j+1}) - \frac{|a|}{2}(U_{j+1} - U_j) + \frac{a}{2}(\text{sign}(a) - \nu)(U_{j+1} - U_j)\chi(U; j)$$

Für $\chi(U; j)$ wird (5.6) angesetzt, jedoch wird das Verhältnis der aufeinanderfolgenden Steigungen jeweils in Richtung des Informationstransports gebildet, d.h.

$$\theta_j = \frac{U_{j'+1} - U_{j'}}{U_{j+1} - U_j} \quad \text{mit } j' = j - \text{sign}(a).$$

Damit gilt $j' = j \pm 1$. Der triviale Fall $a = 0$ ist hier auch eingeschlossen.



Abbildungen: Lösungen des Testproblems zur Zeit $t = 1$.

Die Verallgemeinerung dieser Technik auf ein lineares hyperbolisches System der Form $u_t + Au_x = 0$ ist offensichtlich. Mittels der Diagonalisierung $A = RDR^{-1}$ kann das System in einzelne skalare lineare Advektionsgleichungen entkoppelt werden. Auf jede einzelne Gleichung ist jetzt die obige Vorgehensweise anwendbar. Dadurch lassen sich geschlossene Formeln für den entsprechenden numerischen Fluss herleiten, in dem die Eigenwerte und Eigenvektoren der Matrix A auftreten.

Für ein nichtlineares hyperbolisches System $u_t + f(u)_x = 0$ wird zur Bestimmung des Flusses $F(U; j)$ eine Linearisierung um die entsprechende Zellengrenze durchgeführt. Somit entsteht ein lineares System $u_t + \hat{A}_j u_x = 0$. Der verwendete Ansatz für $\hat{A}_j = \hat{A}(U_j^n, U_{j+1}^n)$ soll dabei bestimmte Eigenschaften einer Matrix aus einem Näherungsverfahren für Riemann-Probleme erfüllen. Leider gibt es keinen allgemeingültigen Ansatz für die Matrizen $\hat{A}_j$. Es kann nur bei konkreten Systemen (z.B. Eulersche Gasgleichungen) jeweils eine effiziente Konstruktion erfolgen. Auf das lineare System kann nun das fluss-begrenzende Verfahren angewendet werden. Dabei ist jedoch zu beachten, dass zur Bestimmung von Entropielösungen noch geeignete Modifikationen notwendig sind. Auch erhöht sich der Rechenaufwand hier deutlich, da die Linearisierung $\hat{A}_j$ an jeder Zellengrenze verschieden sein kann, jedoch stets alle Eigenwerte und Eigenvektoren der Matrix zu bestimmen sind.

5.3 Steigungs-begrenzende Methoden

Das Godunov Verfahren liefert einen Ansatz zur Konstruktion von Verfahren höherer Ordnung. Anstatt im Rekonstruktionsschritt die Daten jeweils konstant fortzusetzen, kann man auch stückweise polynomiale Anfangsfunktionen bilden. Der Evolutions- und Projektionsschritt erfolgt dann wie in der gewöhnlichen Godunov Methode. Wird die Rekonstruktion stückweise linear durchgeführt, so hat die Anfangsfunktion im skalaren Fall die Gestalt

$$\tilde{u}(x, t_n) = U_j^n + \sigma_j^n (x - x_j) \quad \text{für } x \in (x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}), \quad (5.11)$$

wobei $\sigma_j^n \in \mathbb{R}$ die Steigung in der j -ten Zelle repräsentiert. Für beliebige Steigungen sind die Zellenmittel jeweils gleich U_j^n , d.h. der Rekonstruktionsschritt bleibt konservativ. Im gewöhnlichen Godunov Verfahren gilt $\sigma_j^n \equiv 0$. Die Steigungen müssen nun in Abhängigkeit von den Daten U^n geeignet gewählt werden.

Ein naheliegendes Schema für die Steigungen ist gegeben durch

$$\sigma_j^n = \frac{1}{h} (U_{j+1}^n - U_j^n). \quad (5.12)$$

Im Fall der Advektionsgleichung $u_t + au_x = 0$ mit $a > 0$ entsteht bei exaktem Lösungsprozess im Evolutionsschritt aus dieser Verallgemeinerung der Godunov

Methode genau das Lax-Wendroff Verfahren, d.h. ein Schema höherer Ordnung. Die Wahl (5.12) kann jedoch die totale Variation der Anfangsfunktion gegenüber der stückweise konstanten Konstruktion erhöhen. An Unstetigkeitsstellen wird diese Rekonstruktion große Steigungen erzeugen und damit die totale Variation deutlich vergrößern. Dieses Überspringverhalten erklärt auch die im Lax-Wendroff Verfahren resultierenden inkorrekten Oszillationen.

Deshalb wählt man die Steigungen im Rekonstruktionsschritt derart, dass für die Anfangsfunktion die Bedingung

$$\mathrm{TV}(\tilde{u}(\cdot, t_n)) \leq \mathrm{TV}(U^n) \quad (5.13)$$

erfüllt ist. Genau dann verhält sich die gesamte Methode TV-vermindernd, denn Evolutions- und Projektionsschritt erhöhen nie die totale Variation. Eine Möglichkeit, die Forderung (5.13) zu garantieren, besteht in der Wahl der Steigungen

$$\sigma_j^n = \frac{1}{h} \cdot \minmod(U_{j+1}^n - U_j^n, U_j^n - U_{j-1}^n) \quad (5.14)$$

mit der sogenannten *minmod-Funktion*, die definiert ist durch

$$\begin{aligned} \minmod(a, b) &:= \frac{1}{2}(\mathrm{sign}(a) + \mathrm{sign}(b)) \min(|a|, |b|) \\ &= \begin{cases} a, & \text{wenn } |a| \leq |b| \text{ und } ab > 0 \\ b, & \text{wenn } |a| > |b| \text{ und } ab > 0 \\ 0, & \text{wenn } ab \leq 0. \end{cases} \end{aligned}$$

Die minmod-Funktion liefert hier aus der linksseitigen und rechtsseitigen Steigung jeweils die betragsmäßig kleinere. An lokalen Extrema der Daten wird die Steigung auf null gesetzt. Dieses Verhalten erweist sich günstig an Unstetigkeitsstellen und lokalen Extrema der exakten Lösung. Wegen dieser Eigenschaft nennt man die Vorschrift (5.14) einen Steigungs-Begrenzer. Weitere derartige Ansätze für die Wahl der Steigungen sind möglich.

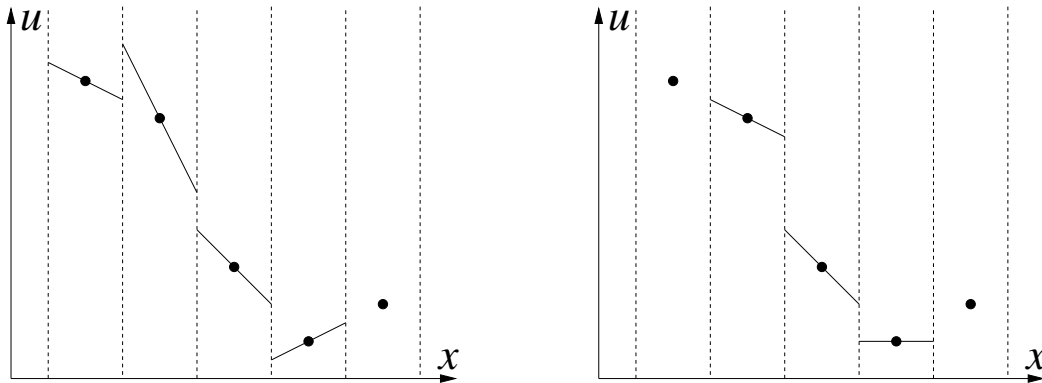


Abbildung: Stückweise lineare Rekonstruktion der Daten über Lax-Wendroff Ansatz (links) und mit Minmod Begrenzer (rechts).

Für die lineare Advektionsgleichung $u_t + au_x = 0$ ist die Lösung explizit gegeben durch $\tilde{u}(x, t_{n+1}) = \tilde{u}(x - ak, t_n)$. Mit den Anfangsdaten (5.11) hat man über den Projektionsschritt bei $a > 0$ für die Näherungen die Formel

$$U_j^{n+1} = U_j^n - \nu(U_j^n - U_{j-1}^n) - \frac{1}{2}\nu(1 - \nu)h(\sigma_j^n - \sigma_{j-1}^n) \quad (5.15)$$

wieder mit $\nu = a\frac{k}{h}$. Für $\sigma_j^n \equiv 0$ entsteht das Godunov Verfahren, d.h. hier das Upwind Schema, während die Wahl (5.12) das Lax-Wendroff Verfahren liefert. Der zu (5.15) korrespondierende numerische Fluss lautet

$$F(U; j) = aU_j + \frac{a}{2} \left(1 - \frac{ak}{h}\right) h\sigma_j. \quad (5.16)$$

Ein Vergleich mit dem numerischen Fluss (5.4) aus dem Fluss-begrenzenden Ansatz zeigt, dass die beiden Funktionen identisch sind, wenn man

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

setzt. Damit kann ein Fluss-Begrenzer χ_j sofort als Steigungs-Begrenzer σ_j interpretiert werden und umgekehrt. Die Methode (5.14) liefert über diese Äquivalenz genau den Minmod Begrenzer (5.9).

Die Verallgemeinerung dieser Technik auf Advektionsgleichungen $u_t + au_x = 0$ mit beliebigem $a \in \mathbb{R}$ liefert die numerische Flussfunktion

$$F(U_j, U_{j+1}) = aU_{j'} + \frac{a}{2} \left(\text{sign}(a) - \frac{ak}{h}\right) h\sigma_{j'} \quad \text{mit } j' = \begin{cases} j & \text{für } a > 0 \\ j + 1 & \text{für } a < 0, \end{cases}$$

wobei im trivialen Fall $a = 0$ dann j' beliebig ist.

Die Anwendung der Steigungs-begrenzenden Verfahren auf lineare sowie nicht-lineare hyperbolische Systeme erfolgt analog zur Verallgemeinerung der Fluss-begrenzenden Ansätze, wie sie im vorhergehenden Abschnitt skizziert ist.