

# **Mathematische Modellbildung und numerische Simulation von Gas-Flüssigkeits-Blasenströmungen**

Der Fakultät Maschinenbau  
der Universität Stuttgart  
eingereichte

**Habilitationsschrift**

vorgelegt von  
**Alexander Sokolichin**  
geboren in Moskau

Hauptberichter : Prof. Dr.-Ing. G. Eigenberger  
Mitberichter : Prof. Dr.-Ing. M. Reuß  
Mitberichter : Prof. Dr. S. Turek

Tag des Habilitationskolloquiums: 19.12.2003

2004



# Vorwort

Die vorliegende Arbeit entstand im Rahmen meiner Tätigkeit als Wissenschaftlicher Mitarbeiter am Institut für Chemische Verfahrenstechnik der Universität Stuttgart.

Dem Institutsleiter, Herrn Prof. Dr.-Ing. G. Eigenberger, gilt mein herzlicher Dank für die stete und großzügige wissenschaftliche und menschliche Unterstützung.

Recht herzlich möchte ich mich bei Herrn Dr. A. Lapin bedanken für seine großzügige Unterstützung bei der Entwicklung des Programms zur Berechnung von Blasenströmungen und für die fruchtbare Zusammenarbeit bei der Entwicklung der in diesem Programm eingesetzten numerischen Verfahren.

Herrn Prof. Dr.-Ing. M. Reuß und Prof. Dr. S. Turek gilt mein besonderer Dank für die Übernahme der Mitberichte und die damit verbundene Mühe, die umfangreiche Habilitationsschrift zu lesen und zu beurteilen.

Allen momentanen und ehemaligen Mitarbeitern danke ich für die einmalige Arbeitsatmosphäre am Institut, die sich durch Hilfsbereitschaft und gegenseitige Unterstützung auch über die Arbeit hinaus auszeichnet. Ganz besonders möchte ich mich bei den Mitarbeitern der Gas/Liquid-Gruppe Stefan Becker, Claudia Schopf, Oliver Borchers und Robert Fettig bedanken für die äußerst produktive und sehr angenehme Zusammenarbeit und für die Bereitstellung der zahlreichen Meßergebnisse, ohne die eine detaillierte Validierung des mathematischen Modells nicht möglich wäre. Wolfgang Lengerer möchte ich sehr für seine menschliche Unterstützung danken, die er mir während einer schwierigen Phase meines Lebens geleistet hat.

Der BASF und der Deutschen Forschungsgemeinschaft sei für ihre großzügige Förderung gedankt.

Stuttgart, im Januar 2004

Meinen Eltern

# Inhaltsverzeichnis

<b>Formelzeichen und Abkürzungen</b>	<b>IX</b>
<b>Einleitung</b>	<b>1</b>
<b>I Diskretisierung hyperbolischer Gleichungen</b>	<b>11</b>
<b>1 Lineare Konvektionsgleichung mit konstanter Konvektionsgeschwindigkeit</b>	<b>13</b>
1.1 Herleitung der Erhaltungsgleichung . . . . .	13
1.2 Analytische Lösung der linearen Konvektionsgleichung . . . . .	17
1.3 Finite-Volumen-Formulierung . . . . .	24
1.4 Upwind- und Downwind-Diskretisierung . . . . .	32
1.5 Zentral-Differenz . . . . .	50
1.6 Upwind-Verfahren zweiter Ordnung . . . . .	57
1.7 Agarwal-Verfahren und QUICK-Verfahren . . . . .	65
1.8 Monotonie-Erhaltung und TVD-Eigenschaft . . . . .	74
1.9 Konstruktion eines TVD-Verfahrens höherer Ordnung . . . . .	79
1.10 TVD-Verfahren: Untersuchung der Testfälle . . . . .	91
1.11 Testfall No.3: nichtmonotone glatte Lösungsfunktion . . . . .	99
<b>2 Lineare Konvektionsgleichung mit variabler Konvektionsgeschwindigkeit</b>	<b>110</b>
2.1 Analytische Lösung der Gleichung . . . . .	110

2.2	Fall einer positiven variablen Geschwindigkeit . . . . .	114
2.3	Fall einer beliebigen variablen Geschwindigkeit . . . . .	127
<b>3</b>	<b>Burger-Gleichung</b>	<b>132</b>
3.1	Theoretische Aspekte und 2 Testfälle . . . . .	132
3.2	Flux-Limiter-Verfahren für die Burger-Gleichung . . . . .	143
3.3	Slope-Limiter-Verfahren für die Burger-Gleichung . . . . .	152
3.4	Splitting-Verfahren und QTVD-Diskretisierung . . . . .	156
<b>4</b>	<b>Mehrdimensionale Probleme</b>	<b>161</b>
4.1	Lineare Konvektionsgleichung in zwei Dimensionen . . . . .	161
4.2	Fall einer konstanten Konvektionsgeschwindigkeit . . . . .	164
4.3	Fall einer variablen Konvektionsgeschwindigkeit . . . . .	172
4.4	Diskretisierung der Navier-Stokes-Gleichungen . . . . .	175
<b>5</b>	<b>Zeitintegrationsverfahren</b>	<b>181</b>
5.1	Implizite und explizite Verfahren erster Ordnung . . . . .	181
5.2	Das Crank-Nicolson-TVD-Verfahren . . . . .	185
<b>II</b>	<b>Modellbildung und Simulation von Blasenströmungen</b>	<b>189</b>
<b>6</b>	<b>Das mathematische Modell</b>	<b>191</b>
6.1	Grundgleichungen des Two-Fluid Modells . . . . .	191
6.2	Kopplung zwischen den Phasen im Two-Fluid-Modell . . . . .	194
6.3	Das Drift-Flux-Modell . . . . .	197
6.4	Boussinesq-Approximation für Drift-Flux-Modell . . . . .	201
6.5	Erste Zwischenbilanz . . . . .	209
6.6	Die Impulsbilanz für die Gasphase . . . . .	210

6.6.1	Impulsbilanz für die Einzelblase . . . . .	210
6.6.2	Druckkraft . . . . .	211
6.6.3	Schwerkraft . . . . .	211
6.6.4	Weitere Kräfte: Vorbemerkungen . . . . .	211
6.6.5	Widerstandskraft . . . . .	212
6.6.6	Die virtuelle Masse („added mass force“) . . . . .	225
6.6.7	Radiale Kräfte („lift forces“) . . . . .	230
6.6.8	Zur Relevanz der einzelnen Kräfte . . . . .	235
6.6.9	Die Impulsbilanz für die Gasphase: Zusammenfassung . . . . .	236
6.7	Das Euler-Lagrange Model . . . . .	237
6.8	Modellierung der Turbulenz in einer Blasenströmung . . . . .	239
6.8.1	Einleitende Bemerkungen . . . . .	239
6.8.2	Modellierung der blaseninduzierten Turbulenz . . . . .	242
6.8.3	Modellierung der turbulenten Dispersion in der Gasphase . . . . .	246
6.9	Das mathematische Modell: Zusammenfassung und numerische Aspekte . . . . .	253
<b>7</b>	<b>Lokal begaste Blasensäule</b>	<b>258</b>
7.1	Außer mittig begaste Blasensäule . . . . .	261
7.1.1	Beschreibung des Testfalls . . . . .	261
7.1.2	Ergebnisse einer laminaren Rechnung . . . . .	262
7.1.3	Einsatz eines Turbulenzmodells . . . . .	268
7.1.4	Einfluß des Diskretisierungsverfahrens auf die Genauigkeit der Lösung	280
7.2	Mittig begaste Blasensäule . . . . .	283
7.2.1	Beschreibung der Testfälle . . . . .	283
7.2.2	Stationäre Strömung . . . . .	285
7.2.3	Instationäre Strömung . . . . .	286

7.2.4	Vergleich der zeitabhängigen Daten, Phasentrajektorien . . . . .	290
7.2.5	Gitterabhängigkeit der Lösung . . . . .	294
7.2.6	Einfluß der Schlupfgeschwindigkeit auf die Ergebnisse . . . . .	295
<b>8</b>	<b>Schlaufenreaktor</b>	<b>298</b>
8.1	Schlaufenapparat mit einem breiten Einbau . . . . .	299
8.2	Schlaufenapparat mit einem schmalen Einbau . . . . .	304
<b>9</b>	<b>Gleichmäßig begaste Blasensäule</b>	<b>314</b>
9.1	Flache Blasensäule . . . . .	317
9.2	Zylindrische Blasensäule . . . . .	324
<b>10</b>	<b>Zusammenfassung</b>	<b>336</b>
<b>A</b>	<b>Ausgelagerte Beweise zu einzelnen Theoremen</b>	<b>340</b>
A.1	Beweis des Theorems 3 . . . . .	340
A.2	Beweis des Theorems 4 . . . . .	341
A.3	Beweis des Theorems 6 . . . . .	342
A.4	Beweis des Theorems 7 . . . . .	343
A.5	Beweis des Theorems 8 . . . . .	344
A.6	Beweis des Theorems 9 . . . . .	347
	<b>Literaturverzeichnis</b>	<b>350</b>

# Formelzeichen und Abkürzungen

## Teil I

### Abkürzungen

CDS	central differencing scheme
(FL)	flux limiter
LUDS	linear upwind differencing scheme
L-AGARWAL	limitierte Version des Agarwal-Verfahrens
L-CDS	limitierte Version von CDS
L-LUDS	limitierte Version von LUDS
L-QUICK	limitierte Version von QUICK
MC	monotonized centered
MM	minmod
QTVD	Quasi-TVD
QUICK	quadratic upstream interpolation for convective kinematics
SB	superbee
(SL)	slope limiter
TVD	total variation diminishing
VL	van Leer

### Lateinische Buchstaben

$a, a(x)$	konstante oder variable Konvektionsgeschwindigkeit
$D$	numerische Diffusion
$E(x, t)$	totale Energie
$E_n(\Delta x, \Delta t)$	numerischer Gesamtfehler
$E(\Delta x)$	numerischer Ortsfehler
$erf(x)$	error function, $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-x^2} dx$
$f(u)$	Stromfunktion
$F(U; i - 1/2)$	numerischer Konvektionsstrom an der Stelle $x = x_{i-1/2}$

$F(U; i - 1/2, j)$	numerischer Konvektionsstrom an der Stelle $x = x_{i-1/2}, y = y_j$
$F(U; l)$	Abkürzung für $F(U; i - 1/2), F(U; i - 1/2, j)$
$F(U; r)$	Abkürzung für $F(U; i + 1/2)$
$G(U; i, j - 1/2)$	numerischer Konvektionsstrom an der Stelle $x = x_i, y = y_{j-1/2}$
$G(U; u)$	Abkürzung für $G(U; i, j - 1/2)$
$h(x; x_0)$	Heavyside Sprungfunktion, $h(x; x_0) = \begin{cases} 0, & \text{wenn } x \leq x_0 \\ 1, & \text{wenn } x_0 < x \end{cases}$
$k$	Wellenzahl
$L_j$	Lipschitz-Konstanten
$L(\Delta x, \Delta t)$	lokaler Abbruchfehler
$L(\Delta x)$	örtlicher Abbruchfehler
$N$	Anzahl der Kontrollvolumina
$N(x; x_0, \sigma)$	Wahrscheinlichkeitsintegral der Normalverteilung, $N(x; x_0, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{1}{2} \frac{(x-x_0)^2}{\sigma^2}} dx$
$p$	Approximationsordnung eines Diskretisierungsverfahrens
$p$	Konvergenzordnung eines Diskretisierungsverfahrens
$p(x, t)$	Druck
$t$	Zeit
$TV(U)$	totale Variation der Gitterfunktion $U$ , $TV(U) = \sum_{i=-\infty}^{i=+\infty}  U_{i+1} - U_i $
$u(x, t)$	unbekannte Funktion
$u(x, t)$	genaue analytische Lösung
$u(\cdot, t)$	Abhängigkeit $x \mapsto u(x, t)$ bei konstant gehaltenem $t$
$u_0(x)$	Anfangsbedingung
$u^K(x, t)$	Lösung der Konvektionsgleichung $u_t + au_x = 0$
$u^{KD}(x, t)$	Lösung der Konvektions-Diffusions-Gleichung $u_t + au_x = Du_{xx}$
$u^{KDi}(x, t)$	Lösung der Konvektions-Dispersions-Gleichung $u_t + au_x = Bu_{xxx}$
$u_i^n$	exakte Lösung an der Stelle $(x_i, t_n)$
$U_i^n$	numerische Approximation der exakten Lösung an der Stelle $(x_i, t_n)$
$U_i$	Abkürzung für $U_i^{n+1}$
$\bar{U}_i$	Abkürzung für $U_i^n$
$U^n$	Gitterfunktion, $U^n = \{U_i^n\} = \{U_i^n, i = \dots, -1, 0, 1, 2, \dots\}$
$U$	Abkürzung für $U^{n+1}$
$\bar{U}$	Abkürzung für $U^n$
$v(x, t)$	Geschwindigkeit
$x, y$	Koordinaten
$X_1$	Einströmrand
$X_2$	Ausströmrand

## Griechische Buchstaben

$\alpha, \beta, \gamma$	unbestimmte Koeffizienten
$\Delta x, \Delta y$	Maschenweite des numerischen Gitters
$\Delta t$	Zeitschritt
$\theta$	Verhältnis von aufeinanderfolgenden Gradienten der Gitterfunktion $U$ bzw. der Stromfunktion $f(U)$
$\lambda$	freier Gewichtungparameter
$\mu(t)$	Randbedingung
$\varrho(x, t)$	Dichte
$\sigma$	Parameter der Normalverteilung
$\phi, \psi$	Gewichtungsfaktoren im TVD-Konvektionsstrom
$\phi(\theta), \psi(\theta)$	Limiter-Funktionen

## Indizes

<i>agar</i>	Agarwal
<i>cds</i>	central differencing scheme
<i>cn</i>	Crank-Nicolson
<i>dc</i>	deferred correction
<i>down</i>	downwind
<i>fl</i>	flux limiter
<i>i, j</i>	Index für Kontrollvolumina
<i>l</i>	Abkürzung für $i - 1/2$
<i>(l)</i>	Laufvariable für Iteration
<i>luds</i>	linear upwind differencing scheme
<i>mc</i>	monotonized centered
<i>mm</i>	minmod
<i>n</i>	Laufvariable für Zeitschritt
<i>r</i>	Abkürzung für $i + 1/2$
<i>qtvd</i>	quasi-tvd
<i>sb</i>	superbee
<i>sl</i>	slope limiter
<i>tvd</i>	total variation diminishing
<i>up</i>	upwind
<i>vl</i>	van Leer

# Teil II

## Lateinische Buchstaben

$BIT$		blaseninduzierte Turbulenz
$C_{am}$	—	Koeffizient der virtuellen Trägheitskraft („added mass coefficient“)
$C_d$	—	Widerstandsbeiwert („drag coefficient“)
$C_\epsilon$	—	Parameter im Modell für blaseninduzierte Turbulenz
$C_{\epsilon 1}$	—	Parameter im $k$ - $\epsilon$ -Modell
$C_{\epsilon 2}$	—	Parameter im $k$ - $\epsilon$ -Modell
$C_k$	—	Parameter im Modell für blaseninduzierte Turbulenz
$C_l$	—	Koeffizient der radialen Kraft („lift coefficient“)
$C_\mu$	—	Parameter im $k$ - $\epsilon$ -Modell
$C_w$	$kg/(m^3 \cdot s)$	Schwarz-und-Turner-Konstante, $C_w = 5 \cdot 10^4 kg/(m^3 \cdot s)$
$D_{lg}^t$	$m^2/s$	turbulenter Diffusionskoeffizient
$d_b$	$m$	Blasendurchmesser
$e_x$	—	vertikal nach oben gerichteter Einheitsvektor
$E\ddot{o}$	—	Eötövöszahl, $E\ddot{o} = \frac{g \rho_l d_b^2}{\sigma}$
$F_{am}$	$N$	virtuelle Trägheitskraft („added mass force“)
$F_d$	$N$	Widerstandskraft („drag force“)
$F_g$	$N$	Schwerkraft
$F_l$	$N$	radiale Kraft („lift force“)
$F_p$	$N$	Druckkraft
$F_{total}$	$N$	Summe aller auf eine Einzelblase wirkenden Kräfte
$F_W$	$N/m^3$	Wechselwirkungskraft
$g$	$m/s^2$	Erdbeschleunigungsvektor
$I$	—	Einheitstensor
$k$	$m^2/s^2$	turbulente kinetische Energie
$m_b$	$kg$	Blasenmasse
$Mo$	—	Mortonzahl, $Mo = \frac{g \mu_l^4}{\rho_l \sigma^3}$
$p$	$N/m^2$	Druck
$P_k$	$N/(m^2 \cdot s)$	Produktionsrate der turbulenten kinetischen Energie
$Re$	—	Blasen-Reynoldszahl, $Re = \frac{\rho_l d_b  \mathbf{u}_b - \mathbf{u}_l }{\mu_l}$
$S_\epsilon$	$N/(m^2 \cdot s^2)$	Zusatzquellterm in der $\epsilon$ -Gleichung, beschreibt den Einfluß der Blasen auf die Turbulenz
$S_k$	$N/(m^2 \cdot s)$	Zusatzquellterm in der $k$ -Gleichung, beschreibt den Einfluß der Blasen auf die Turbulenz
$Sc$	—	Schmidtzahl für den turbulenten Transport

$t$	$s$	Zeit
$\mathbf{T}$	$N/m^2$	Schubspannungstensor
$\mathbf{T}^{Re}$	$N/m^2$	Tensor der Reynolds'schen Schubspannungen
$\mathbf{u}$	$m/s$	Geschwindigkeitsvektor
$\mathbf{u}'$	$m/s$	turbulente Geschwindigkeitsschwankungen
$\mathbf{u}_b$	$m/s$	Blasengeschwindigkeit
$\mathbf{u}_{drift}$	$m/s$	drifting velocity
$\mathbf{u}^{mom}$	$m/s$	momentane Geschwindigkeit
$\mathbf{u}_{rise}$	$m/s$	Blasenaufstiegsgeschwindigkeit
$\mathbf{u}_{slip}$	$m/s$	Schlupfgeschwindigkeit, $\mathbf{u}_{slip} = \mathbf{u}_b - \mathbf{u}_l$
$V_b$	$m^3$	Blasenvolumen
$\dot{V}_G$	$l/min$	Gasdurchsatz
$We$	—	Weberzahl, $We = \frac{\rho_l  \mathbf{u}_b - \mathbf{u}_l ^2 d_b}{\sigma}$
$\mathbf{x}$	$m$	Ortskoordinatenvektor
$\mathbf{x}_b$	$m$	Blasenposition
$X_d$	—	Phasenindikatorfunktion der dispersen Phase

## Griechische Buchstaben

$\epsilon$	$m^2/s^3$	Energiedissipationsrate
$\varepsilon$	—	Volumenanteil
$\delta$	—	Kroneckersymbol
$\mu$	$kg/(m \cdot s)$	laminare Viskosität
$\mu^t$	$kg/(m \cdot s)$	turbulente Wirbelviskosität
$\nabla$	$m^{-1}$	Gradientenvektor $\left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right)^T$
$\rho$	$kg/m^3$	Dichte
$\sigma$	$kg/s^2$	Oberflächenspannung
$\sigma_\epsilon$	—	Parameter im $k$ - $\epsilon$ -Modell
$\sigma_k$	—	Parameter im $k$ - $\epsilon$ -Modell

## Tiefgestellte Indizes

$b$	zur Einzelblase gehörige Größe
$BI$	blaseninduziert
$d$	zur dispersen Phase gehörige Größe
$g$	zur Gasphase gehörige Größe
$l$	zur Flüssigphase gehörige Größe
$SI$	scherinduziert

# Zusammenfassung

Ziel dieser Arbeit ist die Entwicklung eines mathematischen Modells, das eine zutreffende und effiziente numerische Simulation von Gas-Flüssigkeits-Reaktoren mit Blasenströmungen ermöglicht. Die Qualität eines mathematischen Modells kann nur anhand detaillierter Vergleiche zwischen Simulation und Experiment erfolgen. Da ein mathematisches Modell aus einem System von nichtlinearen partiellen Differentialgleichungen besteht, die auf numerischen Wege gelöst werden, ist die Validierung des Modells nur dann möglich, wenn die Simulationsergebnisse nicht zu sehr durch numerische Fehler beeinflusst werden.

Bei der numerischen Behandlung der partiellen Differentialgleichungen stellt die Diskretisierung der konvektiven Terme ein besonders schwieriges Problem dar. Während die linearen Verfahren erster Ordnung an einem hohen numerischen Fehler infolge der numerischen Diffusion leiden, führen die linearen Verfahren höherer Ordnung zu einem unphysikalischen Verlauf der Lösungsprofile. Wegen der auftretenden Oszillationen im Lösungsprofil können diese Verfahren zum Lösen von konvektionsdominanten Problemen nicht eingesetzt werden, wenn die zu transportierende Größe physikalisch keine negativen Werte annehmen darf (wie z.B. Konzentration, Gasgehalt, turbulente kinetische Energie usw.).

In der letzten Zeit wurde eine Reihe von neuen *nichtlinearen* Diskretisierungsverfahren entwickelt, die auf dem sog. *TVD-Konzept* (Total Variation Diminishing) basieren. Sie liefern oszillationsfreie Profile und sind auf glatten Lösungen bis zur dritten Ordnung genau. Eine ausführliche Auseinandersetzung mit dem TVD-Konzept ist ein wichtiger Bestandteil dieser Arbeit.

Da die TVD-Verfahren nicht nur zur Berechnung von Zweiphasenströmungen sondern auch allgemein zur Diskretisierung von beliebigen konvektionsdominanten Gleichungen eingesetzt werden können, und damit für ein breites Spektrum verfahrenstechnischer Anwendungen von Interesse sind, werden sie im **ersten Teil** der Arbeit behandelt. Zunächst wird am Beispiel einer eindimensionalen linearen Konvektionsgleichung mit einer konstanten Konvektionsgeschwindigkeit die Herleitung der TVD-Diskretisierung ausführlich erläutert. Anschließend wird die Verallgemeinerung des TVD-Konzepts auf komplexere lineare und nichtlineare ein- und mehrdimensionale hyperbolische Gleichungen diskutiert. Dabei wer-

den die Vorteile der TVD-Verfahren gegenüber den linearen Diskretisierungsmethoden an 11 Testfällen veranschaulicht.

Die Fragen der Modellbildung und numerischen Simulation von blaseninduzierten Strömungen werden dann im **zweiten Teil** dieser Arbeit behandelt. Zur Diskretisierung der Modellgleichungen werden die im ersten Teil entwickelten TVD-Verfahren eingesetzt. Ausgehend von einem sogenannten „Basismodell“ des Euler-Euler-Typs, das nur solche Terme im Modell berücksichtigt, deren Existenz und mathematische Darstellung weitgehend akzeptiert ist, wird das Modell im engen Wechselspiel mit Experimenten validiert und weiterentwickelt. Es werden dazu unterschiedliche Konfigurationen von lokal begasten Blasensäulen, Schlaufenapparaten und gleichmäßig begasten Blasensäulen simuliert. Zu den ausführlich untersuchten Fragen gehört der Einfluß der unterschiedlichen Kräfte, insbesondere der Widerstandskraft, der „added mass force“ und der „lift force“ auf die Blasenbewegung, sowie die Frage der Erfassung und der Auswirkung von Turbulenz in der Gas-Flüssigkeitsströmung. Die Turbulenz äußert sich dabei zum einen in der Erhöhung der effektiven Viskosität der Flüssigphase und zum anderen in der Dispersion der Gasblasen, wobei die Gasblasen selber maßgeblich zur Turbulenz in der Flüssigphase beitragen.

Als Ergebnis folgt, daß das typisch instationäre Strömungsverhalten in *lokal* begasten Blasensäulen mit *niedrigem* Gasgehalt gut vorhergesagt werden kann, wenn die instationären, dreidimensionalen Modellgleichungen mit den entwickelten numerischen Verfahren gelöst werden. Dazu reicht bei lokaler Begasung ein Turbulenzansatz nach dem Standard- $k$ - $\epsilon$ -Modell aus. Bei vollständiger Begasung und höherem Gasgehalt muß dagegen die blaseninduzierte Turbulenz mitberücksichtigt werden. Bisherige Modellansätze dafür erlauben zwar eine zutreffende Beschreibung nach vorheriger Anpassung der entsprechenden Modellparameter aber noch keine sichere Vorausberechnung.



# Einleitung

Die vorliegende Arbeit entstand als Ergebnis einer 5-jährigen Forschungstätigkeit im Rahmen des DFG-Schwerpunktprogrammes „Analyse, Modellierung und Berechnung mehrphasiger Strömungen“ (Vorhaben Ei 131/10 mit dem Titel „Modellierung und effiziente numerische Simulation von Gas-Flüssigkeits-Reaktoren mit Blasenströmungen nach dem Euler-Euler-Konzept“).

Es war das Ziel des Vorhabens, ein Simulationsprogramm zu entwickeln, mit dem das stationäre und instationäre Verhalten von Gas-Flüssigkeits-Blasenströmungen in Apparaten unterschiedlicher Geometrie modellmäßig zutreffend beschrieben und numerisch effizient simuliert werden kann.

Zutreffende **Modellierung** bedeutet, daß alle wichtigen physikalischen Effekte durch das mathematische Modell richtig wiedergegeben werden. Die Effizienz der **Numerik** setzt voraus, daß die Modellgleichungen zum einen mit einem klar eingrenzbaeren numerischen Fehler und zum anderen mit möglichst niedrigem Speicherplatz- und vor allem Rechenzeitbedarf gelöst werden. Die Effektivität des Rechenverfahrens ist sehr wichtig, da die Modellgleichungen in mehreren Hunderttausenden von Gitterpunkten und über Zehntausende von Zeitschritten aufgelöst werden müssen, um das dreidimensionale und zeitabhängige Verhalten der Strömung wiederzugeben.

Es gibt heute viele unterschiedliche Ansätze zur Modellierung von Zweiphasenströmungen. Bis auf den Fall der Direkten Numerischen Simulation von Gas-Flüssigströmungen, bei welcher die genauen einphasigen Gleichungen innerhalb der jeweiligen Phase gelöst werden, sind überwiegend sog. statistische Modelle (das Euler-Euler-Modell bzw. das Euler-Lagrange-Modell) im Einsatz. Diese statistischen Modelle ergeben sich aus den lokalen einphasigen Gleichungen durch eine geeignete Mittelung. Ihr Einsatz ist allerdings mit Unsicherheit verbunden, die in der Natur der statistischen Modelle liegt. Da bei einem Mittelungsprozeß genaue Informationen über die tatsächliche Strömung benötigt werden, die eben so nicht vorliegen, werden die Terme in den gemittelten Gleichungen von verschiedenen Autoren auf unterschiedliche Weise modelliert. Die Unterschiede betreffen vor allem die Modellierung der Wechselwirkungsterme zwischen beiden Phasen und die Modellierung

der Zweiphasenturbulenz. Ein breiter Einsatz von statistischen Modellen zur Simulation von Gas-Flüssigkeits-Blasenströmungen wird z.Zt. vor allem dadurch erschwert, daß diese Modellterme häufig nur an das jeweilige Problem angepaßt werden und somit keine allgemeine Gültigkeit besitzen.

Diese Unsicherheit bezüglich des mathematischen Modells ist ein Grund dafür, daß das **Verhältnis zwischen der Modellierung und der numerischen Lösung** bei zweiphasigen Strömungen ein anderes ist als im einphasigen Bereich. Die Grundgleichungen der **einphasigen Strömungen** sind im Prinzip bekannt. Die Unsicherheiten betreffen vor allem die Modellierung der einphasigen Turbulenz, aber selbst dafür gibt es eine Reihe von erprobten Modellen, so daß der Einsatz der numerischen Strömungsmechanik zur Vorhersage der Hydrodynamik von einphasigen Systemen heute als ein weitverbreitetes Standardverfahren angesehen werden kann. Der Schwerpunkt liegt hier eindeutig bei der Entwicklung der geeigneten numerischen Methoden, die es ermöglichen, Apparate mit komplizierten Geometrien auszulegen und die resultierenden Gleichungssysteme effektiv zu lösen. Es werden daher ständig neue flexible Gittergenerierungsverfahren sowie Methoden zur Verkürzung des Rechenganges entwickelt. Parallelisierung, Gebietszerlegung sowie den Einsatz von Mehrgitterverfahren zur Konvergenzbeschleunigung sind typische Beispiele.

Bei der Modellierung und Berechnung von **Zweiphasenströmungen** liegt der Schwerpunkt dagegen eindeutig bei der Entwicklung der mathematischen Modelle und deren Validierung. Man konzentriert sich deshalb überwiegend auf Apparate mit einfachen Geometrien (flach bzw. zylindrisch, mit und ohne Einbauten), die den Einsatz von strukturierten und zum großen Teil äquidistanten kartesischen bzw. zylindrischen Gittern ermöglichen, für die effektive numerische Standardverfahren eingesetzt werden können. Die wichtigste Anforderung an die Numerik ist daher nicht die Behandlung komplexer Berechnungsgebiete, sondern Gewährleistung solcher Eigenschaften des Rechenverfahrens, die eine zuverlässige **Validierung** der mathematischen Modelle ermöglichen.

Die Validierung der Modelle kann dabei nur durch detaillierte Vergleiche zwischen Simulationen und Experiment erfolgen. Ein Maß für die Güte des Modells ist dabei die Höhe der **Modellfehler**, die die Differenz zwischen der tatsächlichen Strömung und der exakten Lösung des zugrundegelegten mathematischen Modells quantitativ beschreibt. Das mathematische Modell besteht aus einem System von nichtlinearen partiellen Differentialgleichungen, bei deren numerischer Lösung **numerische Fehler** entstehen, die die Differenz zwischen der exakten Lösung des Differentialgleichungssystems und der numerischen Lösung beschreiben. Die Modellfehler werden somit von den numerischen Fehlern überlagert. Die Validierung des Modells ist daher nur dann möglich, wenn die Simulationsergebnisse nicht zu sehr durch numerische Fehler beeinflußt werden. Somit steht sie im direkten Zusammenhang mit der Validierung des verwendeten numerischen Verfahrens.

Die numerischen Fehler unterteilen sich in Diskretisierungsfehler und Lösungsfehler. Die Diskretisierungsfehler beschreiben die Differenz zwischen der exakten Lösung des Differentialgleichungssystems und der exakten Lösung des durch die Diskretisierung entstandenen algebraischen Gleichungssystems. Die Lösungsfehler beschreiben die Differenz zwischen der exakten Lösung des algebraischen Gleichungssystems und der iterativ berechneten Näherungslösung.

Die **Diskretisierungsfehler** hängen im wesentlichen von der Ordnung des Diskretisierungsverfahrens, der Feinheit des numerischen Ortsgitters und der Größe des Zeitschritts ab. Im Prinzip sind sie vom Anwender kontrollierbar: mit feiner werdenden Gittern und kleiner werdenden Zeitschritten wird jedes konsistente Berechnungsverfahren zu einer gitterunabhängigen Lösung konvergieren. Die Diskretisierungsfehler lassen sich daher anhand von Lösungen auf mehreren systematisch verfeinerten Gittern bzw. mit verkleinerten Zeitschritten abschätzen. Im Bereich der monotonen Konvergenz (d.h. bei hinreichend feiner Diskretisierung) ist die Differenz zwischen den Lösungen auf zwei aufeinanderfolgenden Gittern proportional zum Fehler auf dem feinsten Gitter (analog für den zeitlichen Diskretisierungsfehler), wobei der Proportionalitätsfaktor von der Ordnung des Diskretisierungsschemas abhängt. Für eine verlässliche Abschätzung sind Lösungen auf mindestens drei Gittern bzw. mit drei Zeitschritten notwendig.

Leider ist es bis heute noch nicht üblich, Veröffentlichungen zur Simulation von Zweiphasenströmungen mit einer **detaillierten Fehleranalyse** zu verbinden. So wurden bis vor kurzem für die Diskretisierung der Konvektionsterme überwiegend Upwind-Diskretisierungsverfahren erster Ordnung eingesetzt, die große Fehler in Form von numerischer Diffusion verursachen können. Wie in [111] gezeigt wurde, können diese Fehler die Qualität der numerischen Lösung sehr stark beeinträchtigen. Deshalb sind Behauptungen, dieses oder jenes Modell sei durch einen Vergleich mit experimentellen Daten ausreichend validiert, als spekulativ einzuschätzen, wenn die Veröffentlichung keine Angaben zu der Genauigkeit der numerischen Lösung enthält.

Diese Meinung setzt sich in den letzten Jahren immer stärker durch, und hat dazu geführt, daß einige bekannte Zeitschriften, wie das *ASME Journal of Fluids Engineering*, seit 1993 grundsätzlich keine Veröffentlichungen akzeptieren, wenn sie keine detaillierte Angaben zur Höhe der numerischen Fehler enthalten [31]. Gleichzeitig wird gefordert, daß die eingesetzten Diskretisierungsverfahren formal mindestens zweiter Ordnung genau sein sollen, da bei Verfahren erster Ordnung aufgrund der numerischen Diffusion eine starke Verfälschung der Lösung auftritt. In diesem Zusammenhang wird darauf hingewiesen, daß Strategien, die eine lineare Umschaltung zwischen den Verfahren erster und zweiter Ordnung verwenden (insbesondere die sehr bekannten „hybrid-“ und „power-law-“ Verfahren, s. [93]) als Diskretisierungsverfahren erster Ordnung angesehen werden müssen, da ihre Ergebnisse

beim Lösen von konvektionsdominanten Problemen auf groben Gittern mit dem Upwind-Verfahren erster Ordnung weitgehend übereinstimmen.

Wenn trotzdem in bekannten Zeitschriften des Chemical Engineering immer noch Simulationsergebnisse mit Verfahren erster Ordnung zur Diskretisierung von Konvektionstermen eingesetzt werden, so gibt es dafür gute Gründe. Es existieren zwar viele unterschiedliche lineare Diskretisierungsschemata, die numerisch stabil sind, zweiter oder sogar dritter Ordnung genau sind, und keinerlei numerische Diffusion aufweisen. Sie führen jedoch zu Lösungsprofilen, die an mehr oder weniger starken Oszillationen leiden, wobei unphysikalische lokale Maxima und Minima entstehen. Setzt man solche Verfahren z.B. zur Berechnung des Gasgehalts oder der turbulenten kinetischen Energie ein, so können sich lokal negative Lösungswerte ergeben, die zum Zusammenbruch des gesamten Berechnungsverfahrens führen. Entsprechende Diskretisierungsverfahren sind daher ebenfalls nicht sinnvoll.

In der letzten Zeit wurde aber eine Reihe von neuen *nichtlinearen* Diskretisierungsverfahren entwickelt, die auf dem sog. **TVD-Konzept** (Total Variation Diminishing) basieren [77]. Sie liefern **oszillationsfreie Profile** und sind bei glattem Lösungsverlauf bis zur **dritten Ordnung** genau. Die mathematische Theorie der TVD-Schemata ist allerdings nur für den eindimensionalen Fall relativ gut entwickelt. Die direkte Übertragung von eindimensionalen Diskretisierungsschemata auf den mehrdimensionalen Fall führt zu konservativen TVD-Modellen, die höchstens erster Ordnung genau sein können [32]. Deshalb erfolgt die Erweiterung des TVD-Konzepts auf den mehrdimensionalen Fall auf *ad hoc* Basis durch lokal eindimensionale Dekomposition der mehrdimensionalen Modellgleichungen. Die resultierenden Schemata überzeugen im mehrdimensionalen Fall durch sehr genaue und vor allem physikalisch sinnvolle (oszillationsfreie) Lösungen. Wegen der hohen Komplexität werden diese TVD-Schemata bisher in der angewandten numerischen Strömungsmechanik allerdings noch relativ selten benutzt.

Im Rahmen der vorliegenden Arbeit wurden TVD-Verfahren seit 1996 eingesetzt und an zahlreichen ein- und mehrdimensionalen linearen und nichtlinearen Beispielen eingehend getestet. Trotz der komplexeren Handhabung zahlt sich der Einsatz der TVD-Verfahren mehrfach aus, da die gitterunabhängige Lösung auf viel gröberen (im Vergleich zum Upwind-Verfahren) Gittern erzielt werden kann. Wir sind daher zu der Überzeugung gekommen, daß es momentan keine andere Alternative gibt, wenn man auf handhabbaren Gittern eine weitgehend gitterunabhängige Lösung berechnen möchte.

In zahlreichen Veröffentlichungen sowie in Monographien [46, 77] findet man eine ausführliche Beschreibung von unterschiedlichen TVD-Verfahren. Für einen unvorbereiteten Leser wird die Lektüre dieser Literaturquellen zum einen dadurch erschwert, daß sie eine fundierte mathematische Ausbildung voraussetzen, und zum anderen sehr allgemein konzipiert sind.

Im ersten Teil der vorliegenden Arbeit wird daher die Problematik der TVD-Diskretisierung aus einer anderen Perspektive behandelt. Der einfachste Fall einer eindimensionalen linearen Konvektionsgleichung mit konstanter Konvektionsgeschwindigkeit wird nicht als ein Spezialfall sondern als Einstiegsbeispiel und daher sehr ausführlich dargestellt. Die Entwicklung von TVD-Verfahren für zunehmend komplexere Fälle baut dann konsequent darauf auf.

Da der Einsatz von TVD-Verfahren einerseits eine zwingende Voraussetzung für eine genaue Lösung der Zweiphasengleichungen ist, und diese Verfahren andererseits nicht nur zur Berechnung von Zweiphasenströmungen sondern auch allgemein zur Diskretisierung von beliebigen konvektionsdominanten Gleichungen eingesetzt werden können, sind sie für ein breites Spektrum verfahrenstechnischer Anwendungen von Interesse. Daher werden sie im ersten Teil der Arbeit in allgemeiner Form behandelt, bevor sie im Teil II für den Fall der Zweiphasenströmung konkretisiert werden.

Der erste Teil der Arbeit besteht aus 5 Kapiteln. Im **Kapitel 1** werden wir uns mit der linearen Konvektionsgleichung mit konstanter Konvektionsgeschwindigkeit befassen. Nach der Herleitung der Erhaltungsgleichung (Abschn. 1.1) und deren analytischer Lösung (Abschn. 1.2) leiten wir im Abschn. 1.3 die Finite-Volumen-Formulierung ab, die uns als Ausgangspunkt für unterschiedliche Diskretisierungsverfahren dienen wird. Wir beginnen mit der Untersuchung des Upwind-Verfahrens erster Ordnung und demonstrieren an einigen Beispielen den sehr starken Einfluß der numerischen Diffusion (Abschn. 1.4). Danach werden unterschiedliche lineare Verfahren zweiter und dritter Ordnung untersucht, und die Gründe für das Auftreten der Oszillationen in den Lösungsprofilen besprochen (Abschn. 1.5–1.7). Im Abschn. 1.8 wird eine allgemeine Definition eines TVD-Verfahrens gegeben und im Abschn. 1.9 wird eine Reihe von TVD-Verfahren für die lineare Konvektionsgleichung mit konstanter Konvektionsgeschwindigkeit hergeleitet. Die beiden Abschnitte gehören zu den „formelintensivsten“ Abschnitten der ganzen Arbeit, die meisten Umformungen sind jedoch rein arithmetischer Natur und sind bei etwas Geduld einfach nachzuvollziehen. Das Material dieser beiden Abschnitte ist Voraussetzung für das Verständnis der restlichen Kapitel. In den letzten beiden Abschnitten des ersten Kapitels werden die entwickelten TVD-Verfahren an einer Reihe von Beispielen getestet.

In den **Kapiteln 2–4** wird der Aufbau eines TVD-Verfahrens zunächst am Beispiel einer linearen Konvektionsgleichung mit *variabler* Konvektionsgeschwindigkeit erläutert, und dann auf den Fall einer eindimensionalen nichtlinearen Burger-Gleichung und schließlich auf den Fall mehrerer Dimensionen übertragen. Zur Illustration der entwickelten Verfahren werden acht weitere Testbeispiele herangezogen.

In den ersten vier Kapiteln werden wir uns ausschließlich mit der Diskretisierung der Kon-

vektionsterme beschäftigen. Als Zeitintegrationsverfahren wird dabei die implizite Zeitdiskretisierung erster Ordnung eingesetzt. Im **Kapitel 5** gehen wir daher kurz auf andere Möglichkeiten der Zeitdiskretisierung unter dem Blickwinkel des TVD-Konzeptes ein.

Der mathematischen Modellbildung und numerischen Simulation von Gas-Flüssigkeits-Blasenströmungen ist der **zweite Teil** dieser Arbeit gewidmet. Erste Arbeiten auf dem Gebiet der Simulation von Gas-Flüssig-Strömungen kamen zu Anfang der 60-er Jahre aus dem Bereich der Kerntechnik. Ziel dieser Arbeiten war die Untersuchung des sog. „lost-of-coolant-accident“. Die Simulationen wurden **eindimensional** anhand einfacher Modelle durchgeführt. Das Gas-Flüssig-Gemisch wurde als eine quasi-homogene Phase betrachtet, die Gasgeschwindigkeit aus einer algebraischen Schlupfbeziehung berechnet, und die Wandreibung durch den turbulenten Rohrreibungsansatz mit konstantem Reibungsbeiwert berücksichtigt. Seit Ende der 60-er Jahre werden eindimensionale Zweiphasenmodelle auch zur Berechnung von Blasensäulen eingesetzt. Die Rückvermischung in beiden Phasen wird mit Hilfe von axialen Dispersionskoeffizienten beschrieben, deren Werte aus empirischen Korrelationen ermittelt werden und die von mehreren Größen wie z.B. Reaktordurchmesser und Gasleerrohrgeschwindigkeit abhängen. Die große Zahl unterschiedlicher empirischer Korrelationen für axiale Dispersionskoeffizienten belegt, daß die Vorhersage dieses Modellparameters mit einer großen Unsicherheit verbunden ist. In ersten Ansätzen zur zweidimensionalen Modellierung von Blasensäulen wurde das radiale Gasgehaltsprofil postuliert und die Flüssigkeitsgeschwindigkeit aus einer vereinfachten Impulsbilanz berechnet.

Die ersten detaillierteren Zweiphasenmodelle zur mehrdimensionalen Berechnung von Gas-Flüssig-Strömungen in unterschiedlichen Reaktoren wurden zu Anfang der 80-er Jahre veröffentlicht. Zur Modellierung der Zweiphasenströmung wurde dabei meist das sog. Euler-Euler-Modell verwendet, bei welchem für beide Phasen eigene Bewegungsgleichungen aufgestellt werden, die durch einen Austauschterm gekoppelt sind. Die Gasphase wird dabei als Pseudokontinuum behandelt. Typisch für diese Arbeiten war eine **zweidimensionale stationäre** Betrachtungsweise der Zweiphasenströmung. Die Auflösung beschränkte sich in den meisten Fällen auf nicht mehr als 500 Kontrollvolumina und die überwiegende Zahl der Veröffentlichungen kam aus dem akademischen Bereich.

Ein fast lawineartiges Anwachsen der Forschungsaktivitäten auf dem Gebiet der Simulation der Zweiphasenströmungen kann seit Mitte der 90-er Jahre beobachtet werden, bedingt durch die Weiterentwicklung der numerischen Methoden der Strömungsmechanik und die drastische Erhöhung der Rechenleistung. Immer öfter werden Simulationsergebnisse veröffentlicht, die eine **dreidimensionale instationäre** Berechnung mit mehreren Hunderttausenden von Stützstellen verwenden. Auch das Interesse seitens der Industrie nimmt ständig zu, zumal immer mehr kommerzielle Softwarepakete eine dynamische, mehrdimensionale Simulation von Zweiphasenströmungen zulassen.

Bereits in einem Übersichtsartikel aus dem Jahre 1997 von *Jakobsen et al.* [53] werden 180 Veröffentlichungen zitiert, die die Problematik der numerischen Simulation von Gas-Flüssig-Strömungen betreffen. Mittlerweile ist die Anzahl der Artikel weiter gestiegen. Darunter ist kaum eine Arbeit zu finden, in welcher von einer mangelhaften Übereinstimmung mit Experimenten berichtet wird, obwohl sich die Modelle häufig deutlich unterscheiden. Nicht selten verwenden unterschiedliche Forschungsgruppen sich gegenseitig ausschließende Modellannahmen, und erzielen trotzdem für den gleichen Testfall hervorragende Übereinstimmung zwischen Simulation und Experiment.

Diese Lage macht es einem Neueinsteiger nicht gerade leicht, die numerische Strömungsmechanik zur Berechnung der Zweiphasenhydrodynamik sinnvoll einzusetzen. Hinzu kommt, daß die meisten kommerziellen Pakete zwar eine große Vielfalt von Modellvarianten enthalten und durch die Verwendung von User-Subroutinen fast uneingeschränkt erweiterbar sind. Die Paketbeschreibungen geben allerdings kaum Hinweise darauf, welche Modellvarianten in einem konkreten Fall eingesetzt werden sollten und welche Modellannahmen als besonders kritisch angesehen werden müssen.

Wir sehen die **Aufgabe des zweiten Teils dieser Arbeit** daher darin, dem Leser einen kritischen Überblick über die vorhandenen Modelle zu verschaffen und dies mit einigen Beispielen zu illustrieren. Die Behandlung der unsicheren Aspekte wird dabei genau so hoch gewichtet, wie die Darstellung von Modellen, die an einigen ausgewählten Beispielen sehr gute Ergebnisse erzielt haben. Wegen dem ständig wachsenden Volumen an neuen Veröffentlichungen können allerdings nicht alle Aspekte der Modellierung von Zweiphasenströmungen behandelt werden. Der Rahmen des untersuchten Gebietes soll deshalb im Folgenden genau definiert werden.

Der Gegenstand unserer Untersuchungen sind die sog. **statistischen Euler-Euler-Modelle**. Die Grundgleichungen dieser Modelle ergeben sich aus den genauen lokalen einphasigen Gleichungen durch eine geeignete Mittelung. Der dafür erforderliche Formalismus wird in der oft zitierten Arbeit von *Ishii* [50] ausführlich dargestellt. Die resultierenden Gleichungen haben eine Form, die für alle zweiphasigen Systeme gleich ist, unabhängig davon, ob es sich z.B. um eine Gas-Feststoff- oder eine Gas-Flüssig-Strömung handelt. Sie enthalten allerdings einige Terme, deren Form postuliert werden muß, wie z.B. die Wechselwirkungsterme zwischen beiden Phasen. Es ist klar, daß die Gestalt dieser Terme entscheidend davon abhängt, um was für ein System es sich handelt. So wird z.B. die Widerstandskraft für ein sich im Gas bewegendes Feststoffpartikel durch andere Korrelationen beschrieben als für eine relativ zur Flüssigkeit aufsteigende Gasblase. Auch der Einfluß der einzelnen Terme auf die Lösung des Gleichungssystems hängt von der Art des Zweiphasensystems ab. So kann man in den meisten Fällen den Einfluß der Trägheit und der Schwerkraft bei der Berechnung der Blasengeschwindigkeit in einer Gas-Liquid-Strömung vollständig vernachlässi-

gen, während in einer Gas-Feststoff-Strömung diese Kräfte einen entscheidenden Einfluß auf die Bewegung der Feststoffpartikel haben. Wir werden uns deshalb im Rahmen dieser Arbeit auf ein konkretes Zweiphasensystem festlegen, und zwar auf die **Gas-Flüssigkeits-Strömungen**, in denen die Gasphase in Form von **einzelnen Gasblasen** vorliegt, und alle im System befindlichen Gasblasen die **gleiche Blasenmasse** haben. Dabei werden sowohl die Blasenkoaleszenz als auch der Blasenzerfall ausgeschlossen. Es wird sich zeigen, daß diese Annahme für Luftblasen im Wasser zulässig ist, solange sich die Blasengröße im Bereich von 1 bis 10 *mm* Durchmesser bewegen. Wir werden uns ausschließlich auf die Betrachtung der **Hydrodynamik** konzentrieren und schließen daher sowohl den Stoffaustausch zwischen beiden Phasen als auch die chemische Reaktion aus. Außerdem werden wir von **isothermen** Verhältnissen ausgehen.

Abschließend sei vermerkt, daß das hier vorgestellte Modell auf der Voraussetzung beruht, daß der **Anteil der Gasphase am Gesamtvolumen moderat ist** und lokal die 5%-Grenze nicht wesentlich bzw. nur an wenigen Stellen im Reaktor übersteigt. Der Grund für diese Einschränkung liegt darin, daß ab einem Gasgehalt von 5% die Blasenwechselwirkungen sehr stark zunehmen. Das betrifft nicht nur die erhöhte Koaleszenzwahrscheinlichkeit und Bildung von größeren Blasen, sondern auch eine stark nichtlineare Wechselwirkung der blaseninduzierten und scherinduzierten Turbulenz, sowie eine Veränderung des Widerstandsverhaltens von Blasenclustern im Vergleich zu Einzelblasen. Die bisher vorhandenen experimentellen Daten sind zum Teil widersprüchlich und reichen nicht aus, um die oben genannte Effekte zuverlässig abzuschätzen. Veröffentlichungen, die Euler-Euler-Modelle trotzdem zur Berechnung von Blasenströmungen einsetzen, bei denen der Gasgehalt in großen Teilen des Reaktors die Marke von 20%, 30% und zum Teil sogar 40% übersteigt, müssen bisher als spekulativ angesehen werden. Der starke Fortschritt, der in den letzten Jahren bei der Berechnung von Blasenströmungen mit einem niedrigen Gasgehalt erzielt wurde, läßt jedoch hoffen, daß auch in praxisrelevanten Strömungen mit hoher Gasbelastung genaue Berechnungen möglich sind. Voraussetzungen dafür sind experimentell überprüfte Korrelationen zur Beschreibung der bis jetzt noch nicht genau verstandenen Phänomene oder entsprechende Ergebnisse aus der Direkten Numerischen Simulation.

Die Mehrzahl der Veröffentlichungen zu Modellierung von Blasenströmungen läßt sich in sechs Gruppen aufteilen: lokal begaste Blasensäulen, gleichmäßig begaste Blasensäulen, Schlaufenapparate, begaste Rührkessel, Blasenströmungen in vertikal durchströmten Röhren und Blasenströmungen in durchströmten horizontalen Krümmern. Außerdem existiert eine breite Reihe von Veröffentlichungen aus dem Bereich der metallurgischen Produktion, wo die Injektion der Gasblasen in die Metallschmelze die turbulente Durchmischung der Schmelze anregen soll ( die sog. „gas-agitated tanks“ bzw. „gas-stirred ladles“). Da die dabei eingesetzten Modelle meistens an isothermen Luft-Wasser-Systemen getestet werden, kann

man diese Arbeiten in die Gruppe der lokal begasten Blasensäulen einordnen.

Die oben genannten sechs Gruppen kann man zwei größeren Klassen zuordnen. Die lokal oder gleichmäßig begasten Blasensäulen sowie die Blasenströmungen in Schlaufenapparaten bilden die Klasse der sogenannten **blaseninduzierten Strömungen** („bubble-driven flows“). In diesen Apparaten befindet sich die Flüssigphase in einem Ruhezustand, solange der Begaser ausgeschaltet ist. Erst nach dem Einschalten des Begasers stellt sich eine Auftriebsströmung ein, die aus der Druckdifferenz infolge der ungleichmäßigen Verteilung der Gasphase im Apparat resultiert. Die blaseninduzierten Strömungen sind durch eine **starke zweiseitige Kopplung** zwischen den beiden Phasen gekennzeichnet. Während die Bewegungen in der Flüssigphase eine Umverteilung der Gasphase zu Folge haben, beeinflusst die sich ständig ändernde Verteilung der Gasphase unmittelbar die Dynamik der Flüssigphase. Eine möglichst genaue Berechnung der Verteilung der Gasphase ist bei den blaseninduzierten Strömungen somit eine zwingende Voraussetzung für eine zutreffenden Modellierung der Hydrodynamik des gesamten Zweiphasensystems.

Die restlichen drei Gruppen kann man zu der Klasse der **mechanisch induzierten** Strömungen zusammenfassen. Auch wenn der Begaser komplett ausgeschaltet ist, kommt hier die Strömung nicht zum Erliegen, sondern wird durch den Impuls des Rührers bzw. durch die Druckdifferenz zwischen dem Ein- und Ausströmrand aufrechterhalten. Auch in den Blasenströmungen dieser Klasse ist eine zweiseitige Kopplung zwischen den Phasen vorhanden. Der Einfluß der Gasphase auf die Dynamik der Flüssigphase ist allerdings wesentlich schwächer als im Falle von blaseninduzierten Strömungen, da die ohne die Blasenpräsenz bereits vorhandene Strömungsstruktur weitgehend beibehalten wird. Entsprechend werden an die Berechnung der Verteilung der Gasphase keine so hohe Ansprüche wie im Falle einer blaseninduzierten Strömung gestellt. Die Untersuchung von mathematischen Modellen zur Beschreibung der Klasse der mechanisch induzierten Blasenströmungen ist *nicht* das Ziel dieser Arbeit, d.h. wir werden uns **ausschließlich mit den blaseninduzierten Strömungen** befassen.

Die Fragen der Modellbildung und numerischen Simulation von blaseninduzierten Strömungen werden in **5 Kapiteln im zweiten Teil** dieser Arbeit behandelt. Im **Kapitel 6** werden die Grundgleichungen des Euler-Euler-Modells sowie unterschiedliche Varianten zur Modellierung der Wechselwirkungsterme und der Zweiphasenturbulenz vorgestellt. Gleichzeitig werden einige numerische Aspekte besprochen. Während im ersten Teil der Arbeit die Diskretisierung von *einzelnen* Modellgleichungen behandelt wird, wird im **Kapitel 6** ein Verfahren zur Lösung vom *gesamten* System vorgestellt. Die hohe Effektivität des resultierenden Verfahrens wird durch eine Reihe von vereinfachenden Modellvoraussetzungen ermöglicht, deren Zulässigkeit ausführlich diskutiert wird.

In **Kapiteln 7 bis 9** wird der Einsatz von unterschiedlichen Modellvarianten an den Beispielen von lokal begasten Blasensäulen, Schlaufenapparaten und gleichmäßig begasten Blasensäulen diskutiert. Bei allen drei Reaktortypen wird die gleiche Vorgehensweise verwendet. Ausgehend von einem sogenannten „Basismodell“ des Euler-Euler-Typs werden zunächst nur solche Terme im Modell berücksichtigt, deren Existenz und mathematische Darstellung weitgehend gesichert ist. So wird z.B. bei der Modellierung der Wechselwirkungskraft zwischen beiden Phasen nur die Widerstandskraft berücksichtigt; der Einfluß der virtuellen Masse und der sog. „lift force“ wird vernachlässigt. Ausgehend von dieser „Basisversion“ wird das Modell im engen Wechselspiel mit Experimenten validiert und weiterentwickelt. Dabei wird besonderer Wert darauf gelegt, daß die Modellgleichungen nicht an einen speziellen Testfall angepaßt, sondern an einer Reihe von unterschiedlichen Experimenten getestet werden.

Da das Ziel des zweiten Teils der Arbeit in einer kritischen Diskussion und Auseinandersetzung mit den Modellen und Lösungsverfahren anderer Autoren liegt, werden die unterschiedlichen Modellvarianten zunächst ausführlich beschrieben. Dies ist Inhalt des gesamten 6. Kapitels. Die Hinweise auf fremde Literaturquellen sind dabei auf das gesamte Kapitel verteilt. Auch in den darauffolgenden Kapiteln 7 bis 9 wird auf einige fremde Simulationsergebnisse ausführlich eingegangen, diesmal nach den einzelnen Reaktortypen gegliedert. Die so gewonnenen Erkenntnisse werden anschließend im **Kapitel 10**, zusammengefaßt.

# **Teil I**

## **Diskretisierung hyperbolischer Gleichungen**



## Kapitel 1

# Lineare Konvektionsgleichung mit konstanter Konvektionsgeschwindigkeit

### 1.1 Herleitung der Erhaltungsgleichung

Den Ausgangspunkt unserer Untersuchungen bildet die eindimensionale lineare Konvektionsgleichung mit konstanter Konvektionsgeschwindigkeit (in der englischsprachigen Literatur „convection equation“, oft aber auch „advection equation“ genannt):

$$u_t + au_x = 0 \tag{1.1}$$

Hier bezeichnet  $x$  die Ortskoordinate,  $t$  die Zeit,  $u(x, t)$  ist die gesuchte Funktion und  $a$  ist eine Konstante. Ohne Beschränkung der Allgemeinheit werden wir stets annehmen, daß  $a$  positiv ist.

Diese Gleichung ist ein Spezialfall einer allgemeineren (in der Regel nichtlinearen) hyperbolischen Erhaltungsgleichung

$$u_t + f(u)_x = 0 \tag{1.2}$$

für den Fall einer linearen Stromfunktion

$$f(u) = au. \tag{1.3}$$

In diesem Abschnitt wollen wir zeigen, wie eine hyperbolische Gleichung aus den physikalischen Erhaltungssätzen hergeleitet werden kann. Da wir uns in einem weiteren Kapitel auch mit *nichtlinearen* Gleichungen befassen werden, wird sich die Herleitung nicht auf den linearen Spezialfall beschränken.

Wir beginnen (in Anlehnung an *LeVeque* [77]) mit der Herleitung der Erhaltungsgleichung für die Masse für ein eindimensionales gasdynamisches Problem. Man stelle sich eine Gasströmung in einem Rohr mit einem kleinen (relativ zur Rohrlänge) konstanten Querschnitt

vor, so daß die Dichte und die Geschwindigkeit des Gases als konstant über den Querschnitt angenommen werden können. Man bezeichne mit  $x$  die Position entlang des Rohres und mit  $\varrho(x, t)$  die Dichte des Gases an der Stelle  $x$  zum Zeitpunkt  $t$ . Die Dichte  $\varrho(x, t)$  sei dabei so definiert, daß die Gesamtmasse des Gases in einem beliebigen Rohrabschnitt zwischen den Positionen  $x_1$  und  $x_2$  als Integral der Dichte dargestellt werden kann:

$$\text{Masse in } [x_1, x_2] \text{ zum Zeitpunkt } t = \int_{x_1}^{x_2} \varrho(x, t) dx. \quad (1.4)$$

Nehmen wir weiter an, daß die Rohrwand kein Gas durchläßt, so kann sich die Gasmasse im Rohrabschnitt  $[x_1, x_2]$  nur durch Zu- bzw. Abfluß der Masse durch die Querschnitte  $x_1$  und  $x_2$  ändern.

Wir bezeichnen mit  $v(x, t)$  die Geschwindigkeit des Gases im Querschnitt  $x$  zum Zeitpunkt  $t$ . Der Massenstrom des Gases durch diesen Querschnitt ist dann durch

$$\text{Massenstrom in } (x, t) = \varrho(x, t)v(x, t) \quad (1.5)$$

gegeben.

Die Änderung der Gasmasse in Abschnitt  $[x_1, x_2]$  ist somit durch die Differenz der Massenströme durch die Querschnitte  $x_1$  und  $x_2$  bestimmt:

$$\frac{d}{dt} \int_{x_1}^{x_2} \varrho(x, t) dx = \varrho(x_1, t)v(x_1, t) - \varrho(x_2, t)v(x_2, t). \quad (1.6)$$

Integriert man diese Gleichung über das Zeitintervall  $[t_1, t_2]$ ,  $t_2 > t_1$ , so erhält man die Erhaltungsgleichung für die Masse in der sogenannten *Integralform*:

$$\begin{aligned} \int_{x_1}^{x_2} \varrho(x, t_2) dx &= \int_{x_1}^{x_2} \varrho(x, t_1) dx \\ &+ \int_{t_1}^{t_2} \varrho(x_1, t)v(x_1, t) dt - \int_{t_1}^{t_2} \varrho(x_2, t)v(x_2, t) dt. \end{aligned} \quad (1.7)$$

Die Integralform drückt die Gesamtmasse des Gases im Abschnitt  $[x_1, x_2]$  zum Zeitpunkt  $t_2$  durch die Gesamtmasse des Gases im Abschnitt  $[x_1, x_2]$  zum Zeitpunkt  $t_1$  und die totalen (aufintegrierten) Massenströme des Gases durch die Querschnitte  $x_1$  und  $x_2$  aus.

Setzt man voraus, daß sowohl die Gasdichte  $\varrho(x, t)$  als auch die Gasgeschwindigkeit  $v(x, t)$  stetig differenzierbare Funktionen von  $x$  und  $t$  sind, kann man Gleichung (1.7) auch in der folgenden Form schreiben:

$$\int_{t_1}^{t_2} \int_{x_1}^{x_2} \left( \frac{\partial}{\partial t} \varrho(x, t) + \frac{\partial}{\partial x} (\varrho(x, t)v(x, t)) \right) dx dt = 0. \quad (1.8)$$

Da diese Gleichung für jeden Rohrabschnitt  $[x_1, x_2]$  und über jedes Zeitintervall  $[t_1, t_2]$  gilt, folgt daraus, daß der Integrand von (1.8) an jeder Stelle  $(x, t)$  gleich Null sein muß, d.h.

$$\rho_t + (\rho v)_x = 0. \quad (1.9)$$

Das ist die sogenannte *Differentialform* der Massenerhaltungsgleichung.

In zwei Fällen kann die Massenbilanz (1.9) als eigenständige Gleichung gelöst werden. Der erste Fall liegt vor, wenn die Geschwindigkeit  $v(x, t)$  eine bekannte Funktion von  $\rho(x, t)$  ist. Ein typisches Beispiel ist das sog. „traffic flow“ Problem. Hier beschreibt die Funktion  $\rho(x, t)$  nicht die Gasdichte in einem Rohr, sondern die Fahrzeugdichte auf einer Autobahn (Anzahl der Fahrzeuge pro Kilometer).  $v(x, t)$  ist entsprechend die Geschwindigkeit des Fahrzeugs, das sich zum Zeitpunkt  $t$  an der Stelle  $x$  befindet. Bei der Herleitung der Traffic-Flow-Gleichung geht man davon aus, daß die Geschwindigkeit der Fahrzeuge allein von der lokalen Fahrzeugdichte abhängt, und zwar so, daß die Geschwindigkeit mit der zunehmenden Fahrzeugdichte abnimmt.

Ist die Geschwindigkeit  $v$  als Funktion von  $\rho$  bekannt, so ist das Produkt  $\rho v$  ebenfalls eine Funktion von  $\rho$ , d.h.  $\rho v = f(\rho)$ . Die Massenerhaltungsgleichung (1.9) läßt sich dann folgendermaßen schreiben:

$$\rho_t + f(\rho)_x = 0. \quad (1.10)$$

Das entspricht genau der *hyperbolischen* Erhaltungsgleichung (1.2) bis auf die Bezeichnung der unbekanntenen Funktion.

Die Massenbilanz (1.9) kann auch dann als eigenständige Gleichung gelöst werden, wenn die Geschwindigkeit  $v(x, t)$  als Funktion von  $x$  und  $t$  *a priori* bekannt ist. Ein Spezialfall, wenn die Geschwindigkeit  $v$  nur von  $x$  abhängt, d.h.  $v(x, t) \equiv v(x)$ , wird im Kapitel 2 untersucht. Man beachte, daß sich in diesem Fall die Massenerhaltungsgleichung (1.9) nicht in der Form (1.2) schreiben läßt, da die Stromfunktion  $f$  nicht nur von der gesuchten Funktion  $\rho$ , sondern auch von der unabhängigen Variable  $x$  abhängt:

$$\rho_t + f(\rho, x)_x = 0, \quad \text{mit } f(\rho, x) = \rho v(x). \quad (1.11)$$

Nur wenn die Geschwindigkeit  $v(x, t)$  weder von  $x$  noch von  $t$  abhängt, d.h.

$$v(x, t) \equiv a = \text{const},$$

hängt die Stromfunktion  $f$  nur von  $\rho$  ab:

$$\rho_t + f(\rho)_x = 0, \quad \text{mit } f(\rho) = a\rho. \quad (1.12)$$

Das entspricht genau der *linearen* Konvektionsgleichung (1.1) bis auf die Bezeichnung der unbekanntenen Funktion.

Vollständigkeitshalber soll auch der Fall erwähnt werden, daß die Geschwindigkeit  $v$  weder a priori bekannt noch eine Funktion von  $\varrho$  ist, wie das im Falle einer eindimensionalen Gasströmung meistens der Fall ist. Die Massenbilanz (1.9) kann dann nicht als eigenständige Gleichung gelöst werden, sondern nur als Teil eines Gleichungssystems.

Neben der Massenbilanz soll auch die Impulsbilanz und die Energiebilanz der eindimensionalen Gasströmung aufgestellt werden. Vernachlässigt man dabei den Einfluß der Reibung und der Wärmeleitung, so erhält man die eindimensionalen Eulerschen Gleichungen der Gasdynamik:

$$(\varrho v)_t + (\varrho v^2 + p)_x = 0 \quad (1.13)$$

$$E_t + (v(E + p))_x = 0. \quad (1.14)$$

Hier bezeichnet  $E$  die totale Energie und  $p$  den Druck. Das Gleichungssystem (1.9, 1.13, 1.14) enthält 3 Gleichungen und 4 Unbekannte. Damit das System gelöst werden kann, muß es um die Zustandsgleichung erweitert werden, die den Druck  $p$  als Funktion von  $\varrho$ ,  $\varrho v$  und  $E$  darstellt.

Es ist interessant zu bemerken, daß das System (1.9, 1.13, 1.14) ebenfalls in der Form (1.2) dargestellt werden kann, wenn man die Funktionen  $u(x, t)$  und  $f(u)$  als Vektorfunktionen mit je 3 Komponenten versteht:

$$u^T = (\varrho, \varrho v, E) \equiv (u_1, u_2, u_3) \quad (1.15)$$

$$f(u)^T = (\varrho v, \varrho v^2 + p, v(E + p)) \equiv (u_2, u_2^2/u_1 + p(u), u_2(u_3 + p(u))/u_1) \quad (1.16)$$

Bei der Herleitung der Differentialform (1.9) der Massenerhaltungsgleichung haben wir angenommen, daß die Funktionen  $\varrho(x, t)$  und  $v(x, t)$  stetig differenzierbar sind. Ist das nicht der Fall, so gilt weiterhin nur die allgemeinere Integralform (1.7). Auch im Falle der allgemeinen hyperbolischen Erhaltungsgleichung gilt die Differentialform

$$u_t + f(u)_x = 0 \quad (1.17)$$

nur, wenn die Funktionen  $u(x, t)$  und  $f(u)$  stetig differenzierbar sind. Die stetige Differenzierbarkeit der Stromfunktion  $f(u)$  wird stets angenommen. Die gesuchte Funktion  $u(x, t)$  braucht aber nicht stetig differenzierbar zu sein, selbst wenn die Anfangsbedingungen diese Eigenschaft besitzen. In einem solchen Fall sind eine oder mehrere Ableitungen in (1.17) nicht definiert und man muß auf die entsprechende Integralform zurückgreifen, die für die hyperbolische Gleichung (1.17) folgendermaßen aussieht:

$$\int_{x_1}^{x_2} u(x, t_2) dx = \int_{x_1}^{x_2} u(x, t_1) dx \quad (1.18)$$

$$+ \int_{t_1}^{t_2} f(u(x_1, t)) dt - \int_{t_1}^{t_2} f(u(x_2, t)) dt.$$

Es können somit Lösungen  $u(x, t)$  existieren, die nur der Integralform (1.18), nicht aber der Differentialform (1.17) der Erhaltungsgleichung genügen. Man nennt solche Lösungen „schwache Lösungen“ der Differentialgleichung (1.17). Im Falle einer linearen Konvektionsgleichung (1.1) existiert für das Cauchy-Problem mit nicht glatten Anfangsbedingungen nur eine schwache Lösung (s. nächsten Abschnitt). Im Falle einer nichtlinearen hyperbolischen Gleichung können selbst bei stetig differenzierbaren Anfangsbedingungen mehrere schwache Lösungen existieren, wovon nur eine „physikalisch sinnvoll“ ist. Diese Problematik wird im Kapitel 3 noch ausführlich besprochen.

## 1.2 Analytische Lösung der linearen Konvektionsgleichung

In diesem Abschnitt wollen wir die analytische Lösung der linearen Konvektionsgleichung

$$u_t + au_x = 0, \quad a > 0 \quad (1.19)$$

bestimmen.

Das Lösungsgebiet sei zunächst durch eine Halbebene  $(x, t) \in (-\infty, +\infty) \times [0, +\infty)$  definiert. Zur Berechnung der Lösung benötigen wir die Anfangs- und Randbedingungen für die gesuchte Funktion. Im Falle einer Halbebene besitzt das Lösungsgebiet allerdings keine Ränder in der  $x$ -Richtung, so daß wir uns auf die Anfangsbedingung beschränken können.

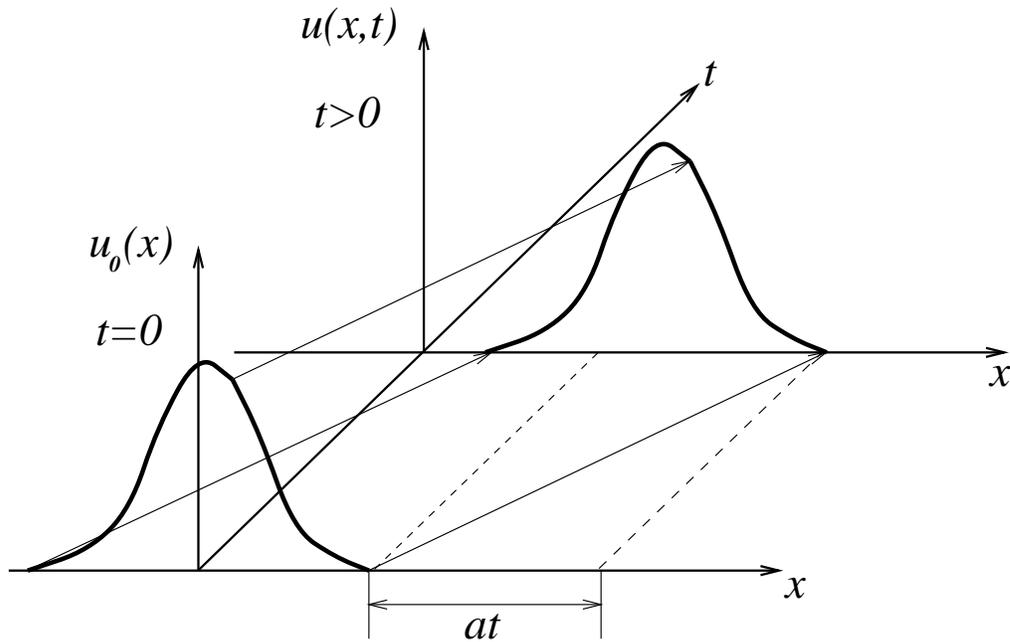
Die Anfangsbedingung sei durch die folgende Gleichung definiert:

$$u(x, t = 0) = u_0(x), \quad x \in (-\infty, +\infty). \quad (1.20)$$

Das System (1.19, 1.20) wird auch als *Cauchy-Problem* für die lineare Konvektionsgleichung bezeichnet.

Wir nehmen zunächst an, daß die Funktion  $u_0(x)$  stetig differenzierbar ist. In diesem Fall lautet die analytische Lösung des Cauchy-Problems

$$u(x, t) = u_0(x - at). \quad (1.21)$$



**Abbildung 1.1:** Lösung der linearen Konvektionsgleichung.

Man kann das leicht überprüfen, indem man die Ausdrücke für die partiellen Ableitungen von  $u(x, t)$

$$u_t(x, t) = u'_0(x - at) \cdot (x - at)'_t = -au'_0(x - at) \quad (1.22)$$

$$u_x(x, t) = u'_0(x - at) \cdot (x - at)'_x = u'_0(x - at) \quad (1.23)$$

in die Gleichung (1.19) einsetzt. Auch die Anfangsbedingung (1.20) ist offensichtlich erfüllt.

Die Lösung (1.21) läßt sich leicht geometrisch interpretieren. Sie besagt, daß das Profil der gesuchten Funktion  $u(x, t)$  zum Zeitpunkt  $t$  gleich ist dem Profil der Lösung zum Zeitpunkt  $t = 0$ , das um  $at$  entlang der  $x$ -Achse nach rechts verschoben wurde (s. Abb. 1.1). Daraus folgt, daß die Lösung  $u(x, t)$  entlang jeder Geraden  $x - at = x_0$ ,  $x_0 \in (-\infty, +\infty)$  konstant bleibt. Diese geraden Linien nennt man Charakteristiken der Differentialgleichung.

Man beachte, daß Charakteristiken Kurven in der  $x$ - $t$ -Ebene sind, die den gewöhnlichen Differentialgleichungen  $x'(t) = a$ ,  $x(0) = x_0$  genügen. Wenn man die Ableitung der Lösung  $u$  entlang einer Charakteristik berechnet, so folgt aus (1.19), daß diese Ableitung verschwindet:

$$\begin{aligned} \frac{d}{dt}u(x(t), t) &= \frac{\partial}{\partial t}u(x(t), t) + \frac{\partial}{\partial x}u(x(t), t) \cdot x'(t) \\ &= u_t + au_x = 0, \end{aligned} \quad (1.24)$$

was wiederum bestätigt, daß der Wert der Funktion  $u$  entlang jeder Charakteristik konstant sein muß. Daraus folgt gleichzeitig, daß die Lösung  $u$  eindeutig durch die Anfangsbedingung (1.20) bestimmt wird.

Ist nun die Anfangsbedingung  $u_0(x)$  nicht differenzierbar, so kann die Funktion  $u(x, t) = u_0(x - at)$  nicht mehr als klassische Lösung der Differenzialgleichung interpretiert werden. Sie genügt allerdings der Integralform der linearen Konvektionsgleichung (s. Abschnitt 1.1), falls  $u_0(x)$  eine beliebige integrierbare Funktion ist. Um das zu zeigen, setzen wir die Funktionen  $u(x, t) = u_0(x - at)$  und  $f(u) = au$  in die Integralgleichung (1.18) ein:

$$\begin{aligned} \int_{x_1}^{x_2} u_0(x - at_2) dx &= \int_{x_1}^{x_2} u_0(x - at_1) dx \\ &+ \int_{t_1}^{t_2} au_0(x_1 - at) dt - \int_{t_1}^{t_2} au_0(x_2 - at) dt. \end{aligned} \quad (1.25)$$

bzw.

$$\begin{aligned} \int_{x_1 - at_2}^{x_2 - at_2} u_0(x - at_2) d(x - at_2) &= \int_{x_1 - at_1}^{x_2 - at_1} u_0(x - at_1) d(x - at_1) \\ - \int_{x_1 - at_1}^{x_1 - at_2} u_0(x_1 - at) d(x_1 - at) &+ \int_{x_2 - at_1}^{x_2 - at_2} u_0(x_2 - at) d(x_2 - at) \end{aligned} \quad (1.26)$$

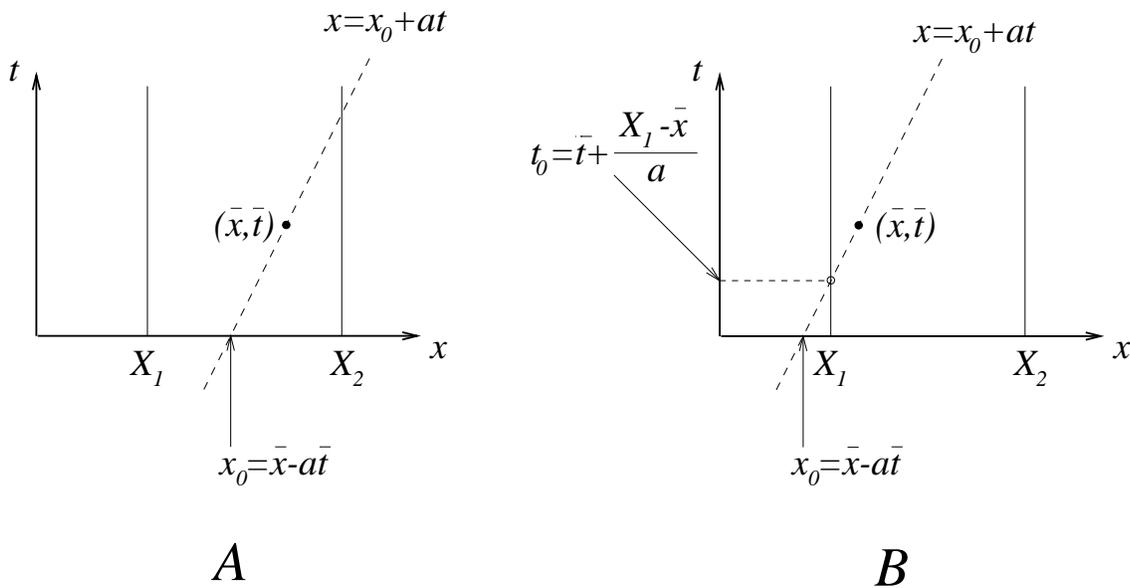
Die Umbenennung der Variablen und die Umgruppierung der Terme auf der rechten Seite führt zu:

$$\int_{x_1 - at_2}^{x_2 - at_2} u_0(y) d(y) = \int_{x_1 - at_2}^{x_1 - at_1} u_0(y) d(y) + \int_{x_1 - at_1}^{x_2 - at_1} u_0(y) d(y) + \int_{x_2 - at_1}^{x_2 - at_2} u_0(y) d(y) \quad (1.27)$$

Es ist nun offensichtlich, daß die letzte Gleichung für jede integrierbare Funktion  $u_0(x)$  und für alle Wertepaare  $(x_1, x_2), (t_1, t_2)$  erfüllt ist. Die Funktion (1.21) kann somit als *schwache Lösung* der linearen Konvektionsgleichung (1.19) interpretiert werden (s. Abschnitt 1.1).

Obwohl die analytische Lösung der linearen Konvektionsgleichung trivial ist, ist es bekanntlich nicht einfach, diese Gleichung numerisch mit hoher Genauigkeit zu lösen. In weiteren Abschnitten dieses Kapitels werden unterschiedliche numerische Diskretisierungsverfahren diskutiert. Unser Ziel ist es, ein Verfahren zu finden, das eine möglichst hohe Genauigkeit besitzt, numerisch stabil ist und physikalisch sinnvolle Lösungen liefert. Das Verfahren soll sowohl für glatte als auch für diskontinuierliche Anfangsbedingungen einsetzbar sein. Um die Güte der numerischen Verfahren zu testen, benötigen wir einige Referenzlösungen.

Bis jetzt haben wir die analytische Lösung der linearen Konvektionsgleichung nur für das Cauchy-Problem gefunden. Numerisch kann eine Gleichung allerdings nicht auf einem räumlich unbegrenzten Gebiet gelöst werden. Wir werden uns daher auf eine endliche



**Abbildung 1.2:** Berechnung der Lösung im Punkt  $(\bar{x}, \bar{t})$ : **A** - die Charakteristik kreuzt die  $x$ -Achse rechts von  $X_1$ ; **B** - die Charakteristik kreuzt die  $x$ -Achse links von  $X_1$ .

Strecke  $[X_1, X_2]$ , normalerweise auf die Strecke  $[0, 2]$ , beschränken. Die Anfangsbedingung lautet dann

$$u(x, t = 0) = u_0(x), \quad x \in [X_1, X_2]. \quad (1.28)$$

Damit die Lösung der Konvektionsgleichung eindeutig bestimmt ist, müssen neben der Anfangsbedingung noch die Randbedingungen spezifiziert werden. Da die Lösung sich entlang der Charakteristiken stromabwärts fortpflanzt, können die Werte der Funktion  $u(x, t)$  an der Stelle  $(X_2, t)$  keinen Einfluß auf die Lösung im Inneren des Intervalls  $[X_1, X_2]$  haben. Am Rand  $x = X_2$  braucht die Randbedingung also nicht angegeben zu werden. Die Randbedingung am Rand  $x = X_1$  sei gegeben durch:

$$u(x = X_1, t) = \mu(t), \quad t > 0. \quad (1.29)$$

Wir werden die Lösung der Gleichung (1.19) mit der Anfangsbedingung (1.28) und der Randbedingung (1.29) zunächst auf geometrischem Weg konstruieren. Das Lösungsgebiet ist ein Halbstreifen  $(x, t) \in [X_1, X_2] \times [0, +\infty)$ . Wir gehen davon aus, daß wie im Falle einer Halbebene die Lösung entlang jeder Charakteristik konstant ist. Für jedes Wertepaar  $(\bar{x}, \bar{t})$  aus dem Halbstreifen werden wir die Charakteristik, die durch den Punkt  $(\bar{x}, \bar{t})$  läuft, zurückverfolgen, bis sie einen der Ränder des Streifens kreuzt.

Der Punkt  $(\bar{x}, \bar{t})$  liegt auf der Charakteristik  $x = x_0 + at$  mit  $x_0 = \bar{x} - a\bar{t}$ . Diese Charakteristik kreuzt die  $x$ -Achse im Punkt  $x_0$ . Liegt der Punkt  $x_0$  innerhalb der Strecke  $[X_1, X_2]$

(s. Abb. 1.2, A), so ist die Lösung in  $(\bar{x}, \bar{t})$  durch die Anfangsbedingung in  $x_0$  bestimmt:  $u(\bar{x}, \bar{t}) = u_0(x_0) = u_0(\bar{x} - a\bar{t})$ . Liegt der Punkt  $x_0$  dagegen außerhalb der Strecke  $[X_1, X_2]$  (s. Abb. 1.2, B), so kreuzt die Charakteristik die Gerade  $x = X_1$  an der Stelle  $t_0 = \bar{t} + (X_1 - \bar{x})/a$ , und die Lösung  $u(\bar{x}, \bar{t})$  wird durch die Randbedingung in  $t_0$  bestimmt. Zusammengefaßt kann die so berechnete Lösung folgendermaßen dargestellt werden:

$$u(x, t) = \begin{cases} u_0(x - at), & \text{wenn } x - at \geq X_1 \\ \mu(t + \frac{X_1 - x}{a}), & \text{wenn } x - at < X_1 \end{cases} \quad (1.30)$$

Analog zu (1.25 - 1.27) kann man zeigen, daß die Funktion (1.30) der Integralform der linearen Konvektionsgleichung genügt. Ist die Funktion  $u(x, t)$  nicht stetig differenzierbar, so handelt es sich um eine schwache Lösung, ansonsten ist die konstruierte Funktion gleichzeitig eine klassische Lösung der Differentialgleichung (1.19), da in diesem Fall die Integralform und die Differentialform äquivalent sind.

Nun wollen wir 2 Referenzlösungen vorstellen, die wir als Testfälle in weiteren Abschnitten dieses Kapitels verwenden werden. In beiden Fällen wird der Lösungstreifen durch die Anfangsstrecke  $[X_1, X_2] = [0, 2]$  definiert und wir nehmen an, daß die Konvektionsgeschwindigkeit  $a$  gleich 1 ist. Die Formel (1.30) vereinfacht sich dann zu

$$u(x, t) = \begin{cases} u_0(x - t), & \text{wenn } x - t \geq 0 \\ \mu(t - x), & \text{wenn } x - t < 0 \end{cases} \quad (1.31)$$

**Testbeispiel No.1.** Im ersten Testbeispiel wird eine *stetig differenzierbare* Lösung dargestellt. Die Anfangsbedingung sei durch das Wahrscheinlichkeitsintegral der Normalverteilung mit Parametern  $x_0$  und  $\sigma$  gegeben:

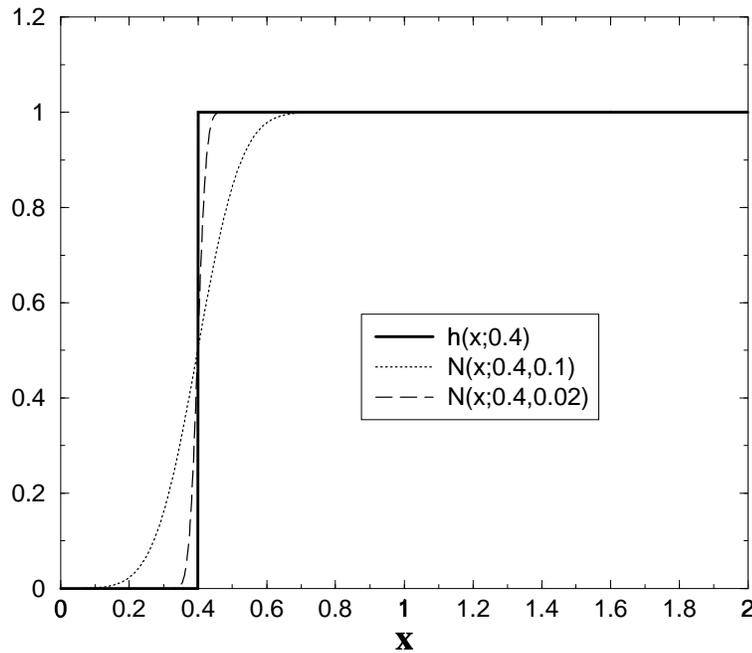
$$u_0(x) = N(x; x_0, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{1}{2} \frac{(x-x_0)^2}{\sigma^2}} dx, \quad x \in [0, 2]. \quad (1.32)$$

Diese Anfangsbedingung kann als eine stetige Approximation der Heavyside'schen Sprungfunktion

$$h(x; x_0) = \begin{cases} 0, & \text{wenn } x \leq x_0 \\ 1, & \text{wenn } x_0 < x \end{cases} \quad (1.33)$$

betrachtet werden. Dabei definiert der Wert von  $x_0$  die Position der Sprungstelle, und  $\sigma$  - die Steilheit des Wahrscheinlichkeitsintegrals an der Stelle  $x_0$ :

$$N'_x(x; x_0, \sigma)|_{x=x_0} = \frac{1}{\sqrt{2\pi}\sigma}. \quad (1.34)$$



**Abbildung 1.3:** Heavyside'sche Sprungfunktion sowie ihre Approximation durch das Wahrscheinlichkeitsintegral für  $\sigma = 0.1$  und  $\sigma = 0.02$ .

In Abbildung 1.3 sind die Heavyside'sche Sprungfunktion für  $x_0 = 0.4$  sowie die Funktion  $N(x; x_0, \sigma)$  für 2 unterschiedliche  $\sigma$ -Werte dargestellt.

Die Randbedingung sei nun durch

$$\mu(t) = N(-t; x_0, \sigma), \quad t > 0 \quad (1.35)$$

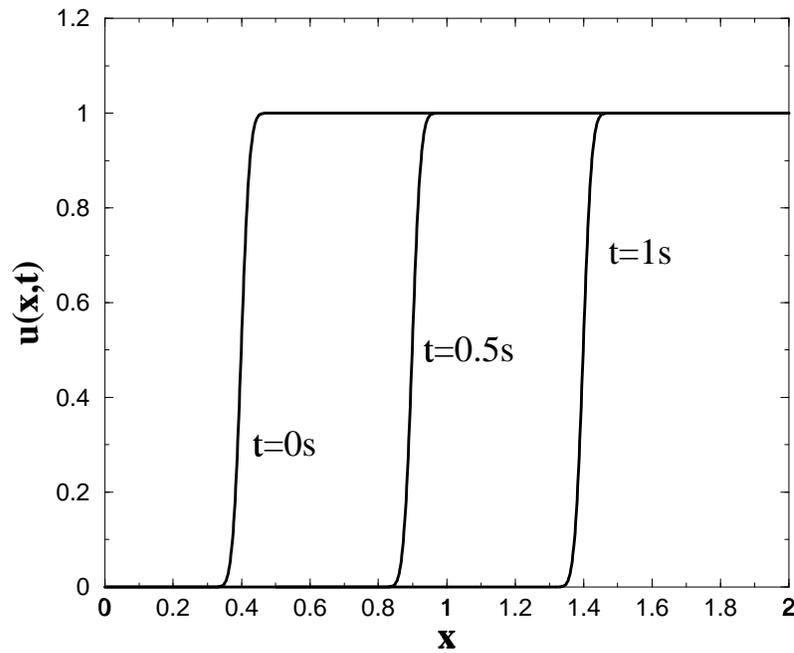
gegeben. Wegen  $u_0(x-t) = \mu(t-x)$  folgt aus (1.31), daß die Lösung der linearen Konvektionsgleichung mit der Anfangsbedingung (1.32) und der Randbedingung (1.35) durch die Formel

$$u(x, t) = N(x-t; x_0, \sigma), \quad x \in [0, 2], \quad t \geq 0 \quad (1.36)$$

beschrieben wird. Das Startprofil der Lösung sowie ihr Verlauf zu den Zeitpunkten  $t = 0.5$  und  $t = 1$  sind in der Abb. 1.4 für den Fall  $x_0 = 0.4$  und  $\sigma = 0.02$  abgebildet.

Für die programmtechnische Umsetzung des Testbeispiels ist es wichtig zu wissen, daß das Wahrscheinlichkeitsintegral  $N(x; x_0, \sigma)$  durch folgende Beziehung dargestellt werden kann:

$$N(x; x_0, \sigma) = \frac{1}{2} + \frac{1}{2} \cdot \operatorname{erf} \left( \frac{x - x_0}{\sqrt{2}\sigma} \right). \quad (1.37)$$



**Abbildung 1.4:** Test No.1 ( $\sigma = 0.02$ ): Anfangsprofil und Lösungsverlauf zu 2 verschiedenen Zeitpunkten.

Hier bezeichnet  $erf(x)$  die „error function“, die folgendermaßen definiert ist:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-x^2} dx. \quad (1.38)$$

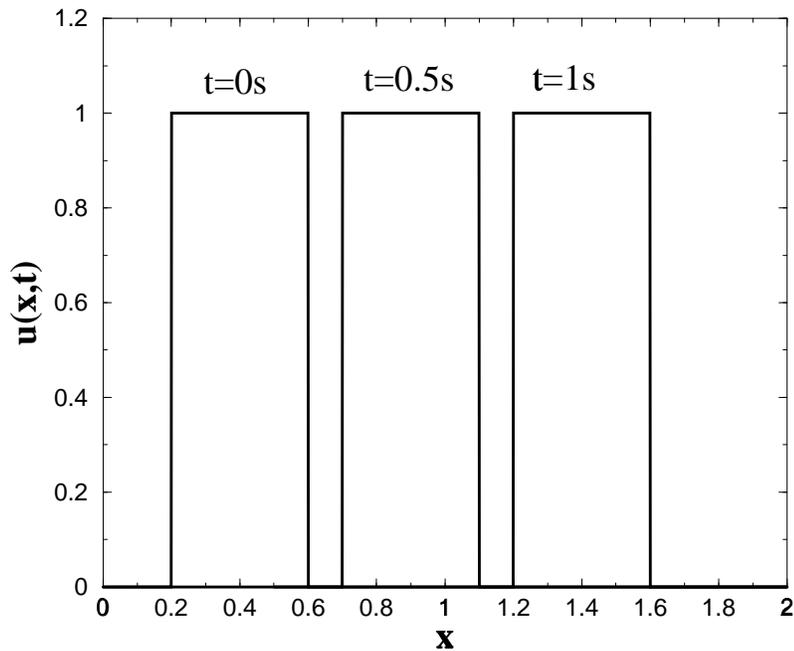
Die Funktion  $erf(x)$  ist in den meisten FORTRAN-Compilern als „intrinsic function“ enthalten. Außerdem kann die Randbedingung (1.35) für den Fall  $x_0 = 0.4$  und  $\sigma = 0.02$  mit einer sehr hohen Genauigkeit durch Null approximiert werden.

**Testbeispiel No.2.** Im zweiten Testbeispiel wird eine *diskontinuierliche* Lösung dargestellt. Die Anfangs- und Randbedingung seien durch folgende Funktionen gegeben

$$u_0(x) = \begin{cases} 0, & \text{wenn } 0 \leq x \leq 0.2 \\ 1, & \text{wenn } 0.2 < x < 0.6 \\ 0, & \text{wenn } 0.6 \leq x \leq 2 \end{cases} \quad (1.39)$$

$$\mu(t) = 0, \quad t > 0 \quad (1.40)$$

Wir sind an der Entwicklung der Lösung im Zeitintervall  $t \in [0, 1]$  interessiert. Während



**Abbildung 1.5:** Testbeispiel No.2: Anfangsprofil und Lösungsverlauf zu 2 verschiedenen Zeitpunkten.

dieses Zeitintervalls kann die Lösung folgendermaßen dargestellt werden:

$$u(x, t) = \begin{cases} 0, & \text{wenn } x \leq 0.2 + t \\ 1, & \text{wenn } 0.2 + t < x < 0.6 + t \\ 0, & \text{wenn } 0.6 + t \leq x \end{cases} \quad (1.41)$$

Das Startprofil der Lösung sowie ihr Verlauf zu den Zeitpunkten  $t = 0.5$  und  $t = 1$  sind in der Abb. 1.5 dargestellt.

### 1.3 Finite-Volumen-Formulierung

Im letzten Abschnitt dieses Kapitels haben wir die *analytische* Lösung der linearen Konvektionsgleichung hergeleitet. Der Rest des Kapitels wird sich mit der Entwicklung der *numerischen* Methoden zur Lösung dieser Gleichung befassen. Alle Diskretisierungsverfahren, die wir betrachten werden, basieren auf einer einheitlichen Finite-Volumen-Formulierung, die in diesem Abschnitt hergeleitet wird. Die Finite-Volumen-Formulierung enthält Konvektionsflüsse über die Seiten der Kontrollvolumina, die sich nicht eindeutig aus den Werten der diskretisierten Lösung bestimmen lassen. Diese Konvektionsflüsse müssen daher

durch die Werte der Gitterfunktion approximiert werden. Je nach verwendeter Approximationsformel erhält man entsprechend unterschiedliche Diskretisierungsverfahren. Die Finite-Volumen-Formulierung kann somit als ein Halbfabrikat angesehen werden, das erst durch die Spezifizierung der Berechnung der Konvektionsflüsse zu einem ausformulierten numerischen Algorithmus wird. In diesem Abschnitt wird nur die Herleitung der Finite-Volumen-Formulierung durchgeführt, die einzelnen Diskretisierungsverfahren werden in den nachfolgenden Abschnitten diskutiert.

Die unterschiedlichen Approximationen der unbekanntenen Massenströme über die Seiten der Kontrollvolumina haben zur Folge, daß auch die Aufstellung der *numerischen* Randbedingungen für jedes konkrete Diskretisierungsverfahren gesondert betrachtet werden muß. Diese werden daher jeweils zusammen mit dem konkreten Diskretisierungsverfahren im entsprechenden Abschnitt diskutiert. In diesem Abschnitt wollen wir der Problematik der numerischen Randbedingungen zunächst aus dem Wege gehen, und einfachheitshalber von einem unendlichen Berechnungsgebiet ausgehen, auch wenn die Lösung einer Gleichung auf einem räumlich unbegrenzten Gebiet numerisch gar nicht realisierbar ist.

Wir betrachten also das Cauchy-Problem für die eindimensionale lineare Konvektionsgleichung:

$$u_t + au_x = 0, \quad x \in (-\infty, +\infty), t \geq 0 \quad (1.42)$$

$$u(x, t = 0) = u_0(x), \quad x \in (-\infty, +\infty). \quad (1.43)$$

Es wird zunächst ein numerisches Netz über das Berechnungsgebiet gelegt. Wir werden dabei von einem äquidistanten Gitter ausgehen, obwohl die meisten Verfahren auf den Fall nichtäquidistanter Gitter verallgemeinert werden können. Die Maschenweite in der  $x$ -Richtung sei durch  $\Delta x$  bezeichnet, und der Zeitschritt durch  $\Delta t$ . Das numerische Gitter wird durch die Wertepaare  $(x_i, t_n)$  definiert, wobei

$$x_i = i \cdot \Delta x, \quad i = \dots, -1, 0, 1, 2, \dots \quad (1.44)$$

$$t_n = n \cdot \Delta t, \quad n = 0, 1, 2, \dots$$

gilt.

Der Wert  $x_i$  definiert den Mittelpunkt des  $i$ -ten Kontrollvolumens. Die Randpunkte des  $i$ -ten Volumenelements sind entsprechend durch

$$x_{i-1/2} = x_i - \Delta x/2 = (i - 1/2) \cdot \Delta x$$

$$x_{i+1/2} = x_i + \Delta x/2 = (i + 1/2) \cdot \Delta x$$

definiert (s. Abb. 1.6).

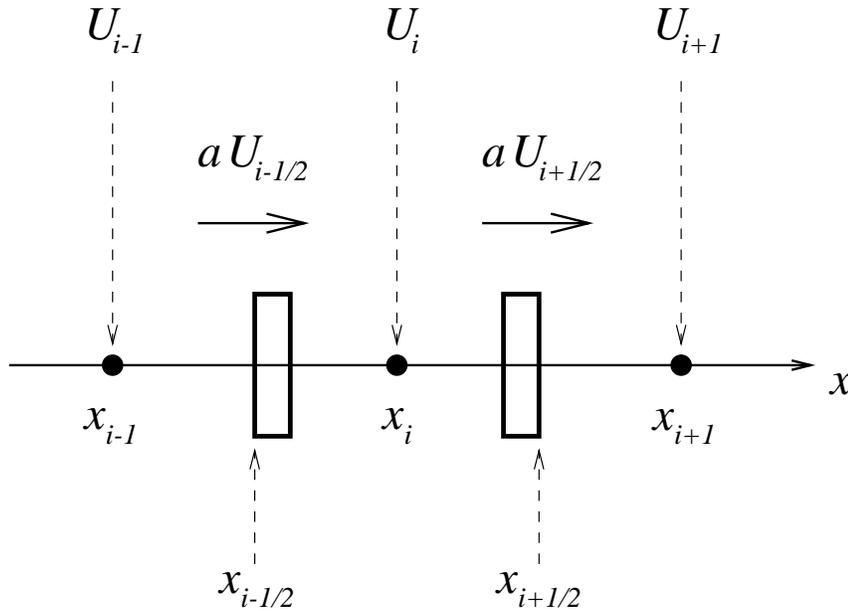


Abbildung 1.6: Skizze des  $i$ -ten Kontrollvolumens.

Jedem Wertepaar  $(x_i, t_n)$  wird eine Größe  $U_i^n$  zugeordnet, die als numerische Approximation der exakten Lösung der Konvektionsgleichung an der Stelle  $(x_i, t_n)$  interpretiert wird. Der zu approximierende Wert wird mit  $u_i^n$  bezeichnet:

$$U_i^n \approx u_i^n \equiv u(x_i, t_n). \quad (1.45)$$

Die Startwerte der gesuchten Gitterfunktion  $U_i^n$  können direkt aus der Anfangsbedingung (1.43) ermittelt werden:

$$U_i^0 := u_i^0 \equiv u(x_i, 0) = u_0(x_i). \quad (1.46)$$

Das numerische Lösungsverfahren soll nun ermöglichen, die Werte der Gitterfunktion zum Zeitpunkt  $t_{n+1}$  aus den Werten der Gitterfunktion zum Zeitpunkt  $t_n$  rekursiv zu berechnen. Dazu wird die Gleichung (1.42) über das Kontrollvolumen  $[x_{i-1/2}, x_{i+1/2}]$  sowie über das Zeitintervall  $[t_n, t_{n+1}]$  integriert und durch  $\Delta x \Delta t$  dividiert:

$$\frac{1}{\Delta x \Delta t} \int_{t_n}^{t_{n+1}} \int_{x_{i-1/2}}^{x_{i+1/2}} (u_t + au_x) dx dt = 0. \quad (1.47)$$

Anschließend werden die einzelnen Terme durch die Werte  $\{U_i^n\}$  und  $\{U_i^{n+1}\}$  approximiert.

Wir behandeln die beiden Summanden des zu integrierenden Ausdrucks separat. Für den ersten Summanden gilt:

$$I_1 = \frac{1}{\Delta x \Delta t} \int_{t_n}^{t_{n+1}} \int_{x_{i-1/2}}^{x_{i+1/2}} u_t dx dt = \frac{1}{\Delta t} \left\{ \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} \left( \int_{t_n}^{t_{n+1}} u_t dt \right) dx \right\} =$$

$$= \frac{1}{\Delta t} \left\{ \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} u(x, t_{n+1}) dx - \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} u(x, t_n) dx \right\}$$

Nun wird angenommen, daß bei kleinen Werten von  $\Delta x$  die Funktionen  $u(\cdot, t_n)$  und  $u(\cdot, t_{n+1})$  im Intervall  $[x_{i-1/2}, x_{i+1/2}]$  durch den Funktionswert in der Mitte des Intervalls hinreichend genau approximiert werden. Unter Berücksichtigung von (1.45) ergibt sich folgende Darstellung für den Ausdruck  $I_1$ :

$$\begin{aligned} I_1 &\approx \frac{1}{\Delta t} \left\{ \frac{1}{\Delta x} (u(x_i, t_{n+1}) \cdot \Delta x) - \frac{1}{\Delta x} (u(x_i, t_n) \cdot \Delta x) \right\} = \\ &= \frac{u_i^{n+1} - u_i^n}{\Delta t} \approx \frac{U_i^{n+1} - U_i^n}{\Delta t}. \end{aligned} \quad (1.48)$$

Für den zweiten Summanden aus (1.47) gilt:

$$\begin{aligned} I_2 &= \frac{1}{\Delta x \Delta t} \int_{t_n}^{t_{n+1}} \int_{x_{i-1/2}}^{x_{i+1/2}} a u_x dx dt = \frac{1}{\Delta x} \left\{ \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \left( \int_{x_{i-1/2}}^{x_{i+1/2}} a u_x dx \right) dt \right\} = \\ &= \frac{1}{\Delta x} \left\{ \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} a u(x_{i+1/2}, t) dt - \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} a u(x_{i-1/2}, t) dt \right\}. \end{aligned} \quad (1.49)$$

Zur Auswertung der beiden Integrale in (1.49) können die Werte der unbekannt Funktion in den Zeitpunkten  $t = t_n$  und  $t = t_{n+1}$  eingesetzt werden. Werden ausschließlich die Werte der Zeitebene  $t = t_n$  verwendet, so führt das zu einem *expliziten* Diskretisierungsverfahren. Werden dagegen nur die Werte aus der Zeitebene  $t = t_{n+1}$  eingesetzt, ergibt sich eine *vollimplizite* Zeitdiskretisierung. Beide Verfahren sind erster Ordnung genau. Die zweite Ordnung der Zeitdiskretisierung kann erreicht werden, wenn der Beitrag der Funktionswerte aus beiden Zeitebenen  $t = t_n$  und  $t = t_{n+1}$  je zu Hälfte berücksichtigt wird (die sog. *Crank-Nicolson-Methode*). Wir werden aus Stabilitätsgründen ausschließlich die vollimplizite Diskretisierung verwenden. Auf andere Zeitdiskretisierungsverfahren gehen wir später im Kapitel 5 ein. Das Integral  $I_2$  wird somit wie folgt approximiert:

$$\begin{aligned} I_2 &\approx \frac{1}{\Delta x} \left\{ \frac{1}{\Delta t} (a u(x_{i+1/2}, t_{n+1}) \cdot \Delta t) - \frac{1}{\Delta t} (a u(x_{i-1/2}, t_{n+1}) \cdot \Delta t) \right\} = \\ &= \frac{a u(x_{i+1/2}, t_{n+1}) - a u(x_{i-1/2}, t_{n+1})}{\Delta x} \approx \frac{a U_{i+1/2}^{n+1} - a U_{i-1/2}^{n+1}}{\Delta x}. \end{aligned} \quad (1.50)$$

Die neuen Größen  $U_{i\pm 1/2}^{n+1}$  bezeichnen dabei die numerischen Approximationen der gesuchten Funktion  $u(x, t)$  an den Stellen  $(x_{i\pm 1/2}, t_{n+1})$ . An diesen Stellen ist die Gitterfunktion  $\{U^{n+1}\}$  jedoch nicht definiert. Die Werte  $U_{i\pm 1/2}^{n+1}$  müssen deshalb aus den Werten der Gitterfunktion in den Knoten  $(x_i, t_{n+1})$ ,  $(x_{i\pm 1}, t_{n+1})$ ,  $(x_{i\pm 2}, t_{n+1})$  usw. interpoliert werden. Jede konkrete Interpolationsformel für  $U_{i-1/2}^{n+1}$  kann dabei von einer unterschiedlichen (endlichen) Anzahl der Elemente des Vektors  $U^{n+1} = \{U_i^{n+1}\}$  abhängen, die um das Element  $U_i^{n+1}$  „zentriert“ sind. Wir werden daher folgende einheitliche Formulierung verwenden, in der die genaue Anzahl der in die Interpolation miteinbezogenen Nachbarknoten nicht explizit angegeben wird:

$$aU_{i-1/2}^{n+1} = F(U^{n+1}; i - 1/2). \quad (1.51)$$

Die Funktion  $F(U^{n+1}; i - 1/2)$  wird als „numerischer Konvektionsstrom“ bezeichnet.

Setzt man nun die Approximationen (1.48) und (1.50) für  $I_1$  und  $I_2$  in die Gleichung

$$I_1 + I_2 = 0$$

ein, erhält man die *Finite-Volumen-Formulierung* für die lineare Konvektionsgleichung

$$\boxed{\frac{U_i^{n+1} - U_i^n}{\Delta t} + a \frac{U_{i+1/2}^{n+1} - U_{i-1/2}^{n+1}}{\Delta x} = 0.} \quad (1.52)$$

Unter Berücksichtigung von (1.51) läßt sich die Gleichung (1.52) auch folgendermaßen schreiben:

$$\boxed{\frac{U_i^{n+1} - U_i^n}{\Delta t} + \frac{F(U^{n+1}; i + 1/2) - F(U^{n+1}; i - 1/2)}{\Delta x} = 0.} \quad (1.53)$$

Die letzte Schreibweise der Finite-Volumen-Formulierung hat gegenüber der Schreibweise (1.52) den Vorteil, daß sie auch für den Fall einer nichtlinearen hyperbolischen Gleichung

$$u_t + f(u)_x = 0$$

verwendet werden kann, wenn man mit  $F(U^{n+1}; i - 1/2)$  die numerische Approximation des Konvektionsstroms

$$\frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} f(u(x_{i-1/2}, t)) dt \approx f(U^{n+1}) \Big|_{x=x_{i-1/2}} = F(U^{n+1}; i - 1/2) \quad (1.54)$$

bezeichnet.

Ist die Funktion  $F(U^{n+1}; i - 1/2)$  genau spezifiziert, so liefert die **Finite-Volumen-Formulierung** (1.53) eine implizite Gleichung, aus der  $U^{n+1}$  als Funktional von  $U^n$  berechnet werden kann:

$$U^{n+1} = FV(U^n). \quad (1.55)$$

Ist der numerische Konvektionsstrom  $F(U; i - 1/2)$  eine lineare Funktion von  $U$ , so ist auch  $FV$  ein linearer Operator, d.h.

$$FV(a_1 U^1 + a_2 U^2) = a_1 FV(U^1) + a_2 FV(U^2), \quad \forall a_1, a_2 \in \mathbf{R}^1, \quad \forall U^1, U^2 \in \mathbf{R}^\infty. \quad (1.56)$$

Mit Hilfe des Funktionals (1.55) kann nun die numerische Lösung  $U^n$  zum Zeitpunkt  $t = t_n$  aus den Startwerten  $U^0$  (1.46) rekursiv berechnet werden:

$$U^n = (FV)^n(U^0). \quad (1.57)$$

**Konvergenz.** Uns wird interessieren, wie gut die aus (1.57) berechnete Gitterfunktion  $U^n$  die exakten Lösungswerte  $u^n = \{u_i^n\}$  approximiert. Um die Güte der Approximation zu quantifizieren, definieren wir den numerischen Fehler  $E_n(\Delta x, \Delta t)$  als 1-Norm der Differenz zwischen den Vektoren  $U^n$  und  $u^n$ :

$$E_n(\Delta x, \Delta t) := \|U^n - u^n\| = \Delta x \sum_i |U_i^n - u_i^n|. \quad (1.58)$$

Eine numerische Lösungsmethode heißt *konvergent*, wenn für jeden Zeitpunkt  $t = T > 0$  die Differenz zwischen der numerischen und der exakten Lösung der Gleichung bei feiner werdenden Gittern gegen Null geht. Die Anzahl der Zeitschritte  $n$ , die gemacht werden müssen, um die Lösung im fest vorgegebenen Zeitpunkt  $T$  zu berechnen, hängt folgendermaßen mit der Zeitschrittweite  $\Delta t$  zusammen:  $T = n\Delta t$ . Konvergenz eines Lösungsverfahrens bedeutet somit, daß für  $\forall T > 0$

$$\lim_{\substack{\Delta x \rightarrow 0 \\ n \rightarrow \infty}} E_n(\Delta x, \Delta t = T/n) = 0 \quad (1.59)$$

gelten muß.

Die Konvergenz eines Lösungsverfahrens besagt nur, daß die numerische Lösung gegen die exakte Lösung konvergiert. Sie gibt jedoch keinen Aufschluß darüber, *wie schnell* die Gitterfunktion gegen die exakte Lösung konvergiert. Die Geschwindigkeit der Konvergenz wird erst durch die *Konvergenzordnung* beschrieben.

Der numerische Fehler  $E_n(\Delta x, \Delta t)$  hängt sowohl von der Zeitschrittweite  $\Delta t$  als auch von der Maschenweite  $\Delta x$  ab. Bei der zeitlichen Diskretisierung haben wir uns zunächst ausschließlich auf die vollimplizite Methode festgelegt. Die örtliche Diskretisierung hängt jedoch von der konkreten Darstellung des numerischen Konvektionsstroms  $F(U; i - 1/2)$

ab. Unterschiedliche Methoden der örtlichen Diskretisierung werden das Thema folgender Abschnitte sein. Für uns steht deshalb zunächst der Vergleich dieser Methoden miteinander und somit die Abhängigkeit des numerischen Fehlers von der Maschenweite  $\Delta x$  im Vordergrund. Wir sondern daher den numerischen *Ortsfehler*  $E(\Delta x)$  aus, indem wir den Zeitschrittweitenanteil aus dem Gesamtfehler  $E_n(\Delta x, \Delta t)$  durch einen Grenzübergang eliminieren:

$$E(\Delta x) := \lim_{\substack{n \rightarrow \infty \\ \Delta t = T/n}} E_n(\Delta x, \Delta t). \quad (1.60)$$

Hinter dieser Definition steht eine Annahme, daß die Zeitschrittweite so klein ist, daß sie keinen Einfluß auf die Höhe des Gesamtfehlers mehr ausübt. Man beachte, daß der numerische Ortsfehler außer von  $\Delta x$  auch noch vom fixierten Zeitpunkt  $T$  abhängt. Diese Abhängigkeit wird jedoch nicht explizit angegeben, da wir in unseren numerischen Vergleichsexperimenten immer denselben Zeitpunkt für alle Diskretisierungsverfahren verwenden werden.

Man sagt, daß ein Diskretisierungsverfahren von der *Konvergenzordnung*  $p$  ist, wenn  $p$  die größte Zahl ist, für die

$$E(\Delta x) = O(\Delta x^p) \quad (1.61)$$

gilt. Je größer die Konvergenzordnung  $p$  eines Verfahrens ist, um so schneller konvergiert die numerische Lösung gegen die exakte Lösung der Differentialgleichung, wenn  $\Delta x$  reduziert wird.

**Konsistenz.** Eine *notwendige* Bedingung der Konvergenz ist die *Konsistenz* des Diskretisierungsverfahrens. Konsistenz eines Diskretisierungsverfahrens bedeutet, daß beim Übergang  $\Delta x \rightarrow 0, \Delta t \rightarrow 0$  die diskretisierte Gleichung in die ursprüngliche Differentialgleichung übergeht. Die Konsistenz definiert also einen Zusammenhang zwischen der Differentialgleichung (1.42) und der diskretisierten Gleichung (1.53).

Um die Definition der Konsistenz zu formalisieren, führen wir den Begriff des *lokalen Abbruchfehlers* ein. Der lokale Abbruchfehler  $L(\Delta x, \Delta t)$  ist die Differenz zwischen der diskretisierten Gleichung und der Differentialgleichung, wenn die exakten Lösungswerte eingesetzt werden:

$$\begin{aligned} L(\Delta x, \Delta t) &:= \\ &= \left( \frac{u_i^{n+1} - u_i^n}{\Delta t} + \frac{F(u^{n+1}; i + 1/2) - F(u^{n+1}; i - 1/2)}{\Delta x} \right) \Bigg|_{\substack{x_i = x \\ t_n = t}} - \underbrace{(u_t + a u_x)}_{\equiv 0} \end{aligned}$$

$$= \left( \frac{u_i^{n+1} - u_i^n}{\Delta t} + \frac{F(u^{n+1}; i + 1/2) - F(u^{n+1}; i - 1/2)}{\Delta x} \right) \Big|_{\substack{x_i = x \\ t_n = t}}. \quad (1.62)$$

Der Ausdruck in den runden Klammern entspricht der Finite-Volumen-Formulierung an der Stelle  $(x_i, t_n)$ . Wegen  $x_i = i \cdot \Delta x$ ,  $t_n = n \cdot \Delta t$  würde sich beim Grenzübergang  $\Delta x \rightarrow 0, \Delta t \rightarrow 0$  die Stelle  $(x_i, t_n)$  dem Koordinatenanfang  $(0,0)$  nähern. Um diese Abhängigkeit zu eliminieren, wird daher das Wertepaar  $(x_i, t_n)$  vor dem Grenzübergang durch ein beliebiges aber fixiertes Wertepaar  $(x, t)$  ersetzt. Die Konsistenz eines Diskretisierungsverfahrens bedeutet somit, daß

$$\lim_{\substack{\Delta x \rightarrow 0 \\ \Delta t \rightarrow 0}} L(\Delta x, \Delta t) = 0 \quad (1.63)$$

gilt.

Analog zum numerischen Ortsfehler kann man den Begriff des örtlichen Abbruchfehlers einführen:

$$L(\Delta x) := \lim_{\Delta t \rightarrow 0} L(\Delta x, \Delta t). \quad (1.64)$$

Die größte Zahl  $p$ , für die

$$L(\Delta x) = O(\Delta x^p) \quad (1.65)$$

gilt, wird als (örtliche) *Approximationsordnung* des Diskretisierungsverfahrens bezeichnet. Bei einem konvergenten Diskretisierungsverfahren ist die Konvergenzordnung in der Regel um so höher, je höher die Approximationsordnung des Verfahrens ist.

**Stabilität.** Die Konsistenz des Diskretisierungsverfahrens ist nur *eine* notwendige Bedingung für dessen Konvergenz. Eine weitere notwendige Bedingung der Konvergenz ist die sog. *Stabilität* des Verfahrens. Ein numerisches Lösungsverfahren heißt stabil, wenn der Fehler  $E_n(\Delta x, \Delta t)$  gleichmäßig beschränkt ist bei  $n \rightarrow \infty$  für alle fixierten Zeitschrittweiten  $\Delta t < \Delta t_0$ . Das Äquivalenz-Theorem von Lax [99] besagt, daß für ein konsistentes *lineares* Verfahren (s. (1.56)) Stabilität eine notwendige *und* hinreichende Bedingung der Konvergenz ist. Dabei stimmen die Konvergenz- und Approximationsordnung des Verfahrens überein. Für ein *nichtlineares* Verfahren bedarf es allerdings einer anderen Definition der Stabilität, damit ein konsistentes und stabiles Verfahren konvergent ist.

**Bezeichnungen.** Zum Abschluß des Abschnittes wollen wir noch eine Konvention bezüglich der Bezeichnungen treffen. In weiteren Kapiteln wird die Finite-Volumen-

Formulierung (1.53) mit vielen unterschiedlichen numerischen Konvektionsstromfunktionen  $F(U^{n+1}; i - 1/2)$  verwendet. Diese Gleichung enthält die Werte der Gitterfunktion auf zwei Zeitebenen:  $U^n$  und  $U^{n+1}$ . Der genaue Wert des Zeitindex  $n$  ist in den meistens Fällen allerdings nicht relevant, man könnte z.B. genau so gut die Bezeichnungen  $U^{alt}$  und  $U^{neu}$  verwenden. Um die Lesbarkeit der Formeln nicht unnötig zu erschweren, wird daher vereinbart, daß mit  $U$  immer die Werte der „neuen“ Zeitebene  $U^{n+1}$  bezeichnet werden, und mit  $\bar{U}$  – die Werte der „alten“ Zeitebene  $U^n$ . Außerdem werden die Indizes  $i - 1/2$  und  $i + 1/2$  oft durch  $l$  und  $r$  („links“ bzw. „rechts“ von  $i$ ) ersetzt, dort wo es dem besseren Verständnis dienen kann. Die Finite-Volumen-Formulierung (1.53), zum Beispiel, könnte in der neuen kompakten Schreibweise folgendermaßen aussehen:

$$\frac{U_i - \bar{U}_i}{\Delta t} + \frac{F(U; r) - F(U; l)}{\Delta x} = 0.$$

## 1.4 Upwind- und Downwind-Diskretisierung

Damit die Finite-Volumen-Formulierung (1.53) zur Berechnung der Gitterfunktion  $U$  eingesetzt werden kann, müssen zunächst die numerischen Konvektionsströme an den Rändern des Volumenelements spezifiziert werden. Da die Gitterfunktion  $U$  nur in den Zentren der Kontrollvolumina definiert ist, müssen die Werte der Gitterfunktion an den Rändern der Kontrollvolumina interpoliert werden. Man muß also eine Annahme über den Verlauf der Lösung *zwischen* den Gitterpunkten treffen.

Die einfachste Möglichkeit besteht darin, daß man einen konstanten Verlauf der gesuchten Funktion innerhalb jedes Volumenelements annimmt:

$$U(x) = U_i, \quad x \in [x_{i-1/2}, x_{i+1/2}] \quad (1.66)$$

Diese Formulierung ist jedoch nicht konsistent, da jedem Rand eines Volumenelements gleichzeitig zwei Funktionswerte zugeordnet werden. Der Randpunkt  $x_{i-1/2}$ , zum Beispiel, gehört sowohl zum Volumenelement  $[x_{i-3/2}, x_{i-1/2}]$  als auch zum Volumenelement  $[x_{i-1/2}, x_{i+1/2}]$ . Die Funktion  $U(x)$  sollte daher laut (1.66) an der Stelle  $x_{i-1/2}$  gleichzeitig zwei Werte, nämlich  $U_{i-1}$  und  $U_i$ , annehmen. Dies steht mit der Definition der Funktion als einer eindeutigen Abbildung im Widerspruch. Um diesen Widerspruch zu beseitigen, muß man sich für einen der zwei möglichen Werte entscheiden. Entscheidet man sich konsequent (d.h. für Ränder *aller* Kontrollvolumina) für den Wert der Gitterfunktion mit dem kleineren Index, so erhält man folgende stückweise konstante Interpolation (s. Abb. 1.7, oben):

$$U(x) = U_i, \quad x \in (x_{i-1/2}, x_{i+1/2}]. \quad (1.67)$$

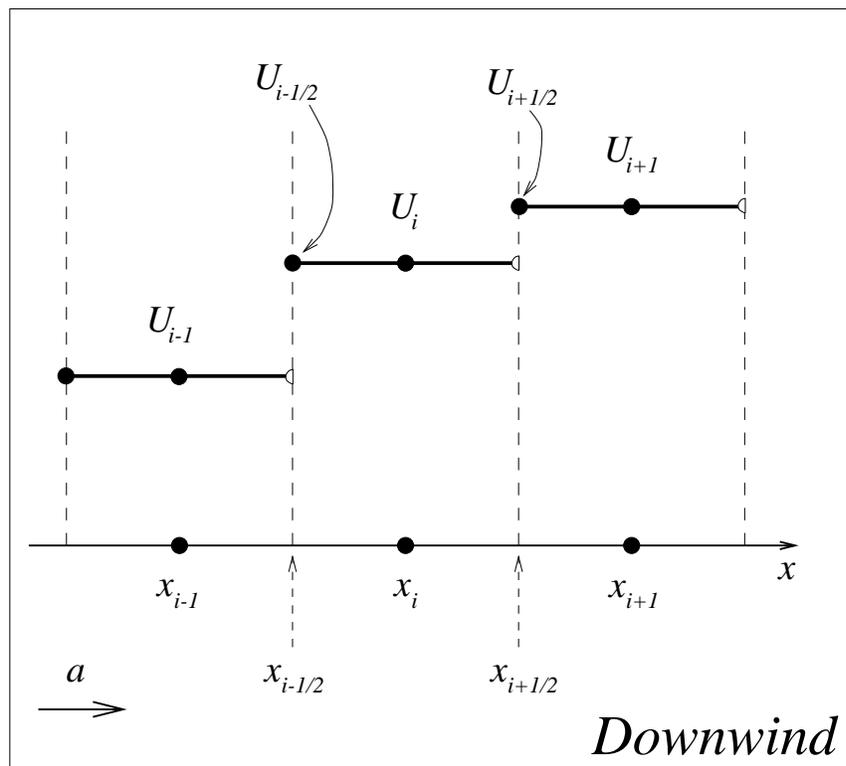
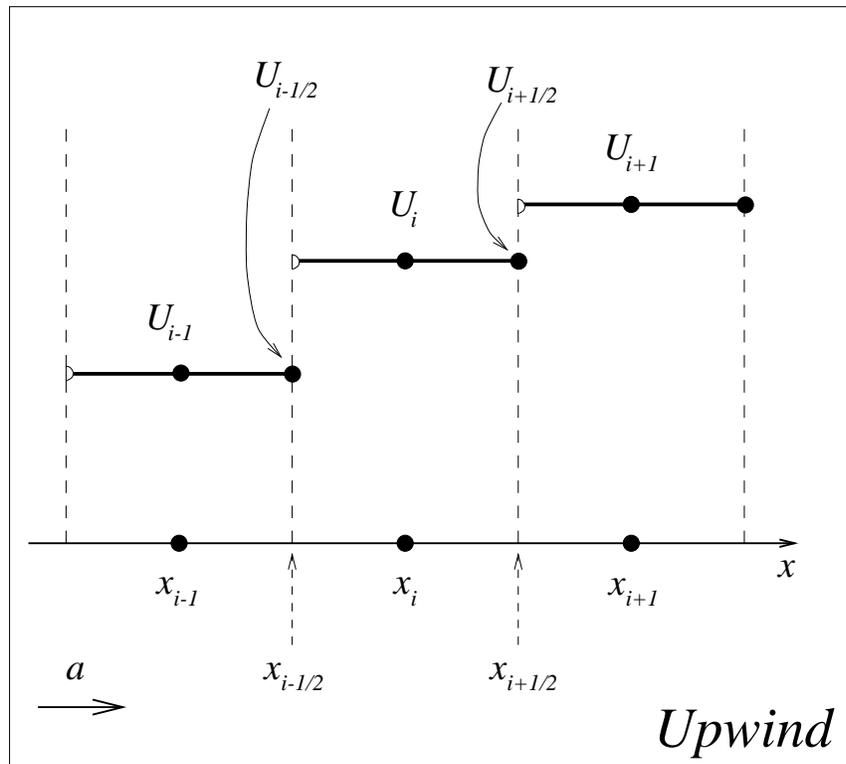


Abbildung 1.7: Upwind- und Downwind-Interpolation der unbekanntnen Funktion.

Entscheidet man sich dagegen konsequent für den Wert der Gitterfunktion mit dem größeren Index, so erhält man eine andere, ebenfalls stückweise konstante Interpolation (s. Abb. 1.7, unten):

$$U(x) = U_i, \quad x \in [x_{i-1/2}, x_{i+1/2}). \quad (1.68)$$

Im ersten Fall erhält man folgenden Ausdruck für den numerischen Konvektionsstrom:

$$aU_{i-1/2} = F^{up}(U; i - 1/2) = aU_{i-1}. \quad (1.69)$$

Das hochgestellte „up“ ist eine Abkürzung für das englische Wort „Upwind“. Sie kommt daher, daß der Wert der gesuchten Funktion an der Stelle  $x_{i-1/2}$  durch den Wert der Gitterfunktion in der Stelle  $x_{i-1}$  approximiert wird, die sich stromaufwärts von der Stelle  $x_{i-1/2}$  befindet.

Im zweiten Fall erhält man den numerischen Konvektionsstrom, der entsprechend als „Downwind“-Approximation bezeichnet wird:

$$aU_{i-1/2} = F^{down}(U; i - 1/2) = aU_i. \quad (1.70)$$

Die Bezeichnungen „Upwind“ und „Downwind“ sind nur für den Fall einer positiven Konvektionsgeschwindigkeit  $a$  sinnvoll. Ist  $a$  negativ, dann müssen diese Bezeichnungen offensichtlich ausgetauscht werden.

Setzt man den Upwind-Konvektionsstrom in die Finite-Volumen-Gleichung (1.53) ein, so ergibt sich die sog. *Upwind-Diskretisierung* der Konvektionsgleichung:

$$\boxed{\text{UPWIND : } \frac{U_i - \bar{U}_i}{\Delta t} + a \frac{U_i - U_{i-1}}{\Delta x} = 0.} \quad (1.71)$$

Diese Formel kann man auch direkt aus der linearen Konvektionsgleichung mittels eines Finite-Differenzen-Verfahrens ableiten, wenn man für die Approximation der örtlichen Ableitung die *Rückwärtsdifferenz* verwendet.

Der Einsatz des Downwind-Konvektionsstroms in die Finite-Volumen-Gleichung (1.53) führt entsprechend zu der *Downwind-Diskretisierung*, die einer Approximation der örtlichen Ableitung durch die *Vorwärtsdifferenz* entspricht:

$$\boxed{\text{DOWNWIND : } \frac{U_i - \bar{U}_i}{\Delta t} + a \frac{U_{i+1} - U_i}{\Delta x} = 0.} \quad (1.72)$$

**Upwind-Verfahren.** Wir werden zunächst die Eigenschaften der Upwind-Diskretisierung untersuchen und zwar als erstes die Approximationsordnung. Wir nehmen an, daß die Funktion  $u(x, t)$  mindestens dreifach differenzierbar ist. Der lokale Abbruchfehler ergibt sich aus dem Einsatz des Konvektionsstromes (1.69) in die Definition (1.62):

$$L(\Delta x, \Delta t) = \frac{u(x, t + \Delta t) - u(x, t)}{\Delta t} + a \frac{u(x, t + \Delta t) - u(x - \Delta x, t + \Delta t)}{\Delta x}. \quad (1.73)$$

Der Grenzübergang  $\Delta t \rightarrow 0$  führt zu

$$L(\Delta x) = \lim_{\Delta t \rightarrow 0} L(\Delta x, \Delta t) = u_t + a \frac{u(x, t) - u(x - \Delta x, t)}{\Delta x}. \quad (1.74)$$

Wir setzen die Taylorentwicklung der Funktion  $u(\cdot, t)$  um den Punkt  $x$

$$u(x - \Delta x, t) = u(x, t) - \Delta x u_x(x, t) + \frac{\Delta x^2}{2} u_{xx}(x, t) + O(\Delta x^3) \quad (1.75)$$

in die Gleichung (1.74) ein und erhalten

$$L(\Delta x) = u_t + a u_x - \frac{a \Delta x}{2} u_{xx} + O(\Delta x^2). \quad (1.76)$$

Da man annimmt, daß die Funktion  $u(x, t)$  eine Lösung der linearen Konvektionsgleichung ist, ist die Summe der ersten beiden Terme gleich Null, und es gilt

$$L(\Delta x) = -\frac{a \Delta x}{2} u_{xx} + O(\Delta x^2) = O(\Delta x). \quad (1.77)$$

Wir haben damit gezeigt, daß das Upwind-Verfahren konsistent ist und die Approximationsordnung 1 hat. Da dieses Verfahren linear ist, ist es nach dem Äquivalenz-Theorem von Lax für die Konvergenz des Verfahrens hinreichend und notwendig, daß das Verfahren auch stabil ist. Die Stabilität eines linearen Verfahrens kann mit vielen unterschiedlichen Methoden untersucht werden, die in einschlägigen Büchern über die numerische Strömungsmechanik ausführlich behandelt werden (s. z.B. [45]). Auf die Darstellung dieser Methoden wird in dieser Arbeit daher verzichtet. Für das Upwind-Verfahren erster Ordnung gilt, daß es *absolut* (d.h. unabhängig von der Zeitschrittweite  $\Delta t$ ) stabil ist („unconditionally stable“). Daraus folgt, daß das Upwind-Verfahren konvergent ist und die Konvergenzordnung 1 hat.

Im Unterschied zu linearen Verfahren ist es bei nichtlinearen Diskretisierungsverfahren oft nur mit erheblichem Aufwand bzw. gar nicht möglich, ihre Konvergenzordnung auf theoretischem Wege zu bestimmen. Man ist somit gezwungen, die Konvergenzordnung empirisch anhand von einzelnen Testbeispielen zu ermitteln. Wir werden daher die Vorgehensweise zur empirischen Ermittlung der Konvergenzordnung eines Verfahrens anhand der Testbeispiele No.1 und No.2 (s. Abschnitt 1.2) ausführlich erläutern. Es wird uns nicht überraschen,

daß für das Testbeispiel No.1 die empirisch ermittelte Konvergenzordnung des Upwind-Verfahrens mit dem theoretisch festgestellten Wert von 1 übereinstimmt. Für das Testbeispiel No.2 werden wir allerdings nur die Konvergenzordnung von 0.5 feststellen. Der vermeintliche Widerspruch mit der Theorie läßt sich leicht erklären. Der theoretisch ermittelte Wert von 1 ergibt sich aus der Übereinstimmung zwischen der Konvergenzordnung und der Approximationsordnung eines stabilen linearen Verfahrens. Bei der Ermittlung der Approximationsordnung des Upwind-Verfahrens haben wir die mehrfache Differenzierbarkeit der Lösung vorausgesetzt und für die Taylorentwicklung der Lösung (s. (1.75)) benutzt. Im Testbeispiel No.2 handelt es sich allerdings um eine diskontinuierliche Lösung, so daß die Annahme der Differenzierbarkeit verletzt ist. Daher stimmt auch die Konvergenzordnung im Testbeispiel No.2 nicht mit dem für differenzierbare Funktionen theoretisch ermittelten Wert überein.

Bisher haben wir die Upwind-Diskretisierung nur für das Cauchy-Problem (1.42, 1.43), d.h. für ein räumlich unbegrenztes Lösungsgebiet diskutiert. In unseren Testbeispielen wird die Lösung allerdings in einem örtlich begrenzten Gebiet und zwar für  $x \in [0, 2]$  gesucht. Die an die Ränder grenzenden Kontrollvolumina bedürfen einer besonderen Behandlung, die wir jetzt diskutieren werden.

Es wird zunächst ein numerisches Gitter über das Berechnungsgebiet gelegt. Für eine beliebige (aber feste) natürliche Zahl  $N$  unterteilen wir das Berechnungsintervall  $[0, 2]$  in  $N$  Teilintervalle gleicher Länge. Die Gitterpunkte seien nun durch die Endpunkte der Teilintervalle so definiert, daß dem ersten Punkt der Index 0 und dem letzten Punkt der Index  $N$  zugeordnet wird:

$$0 = x_0 < x_1 < \dots < x_i < \dots < x_N = 2$$

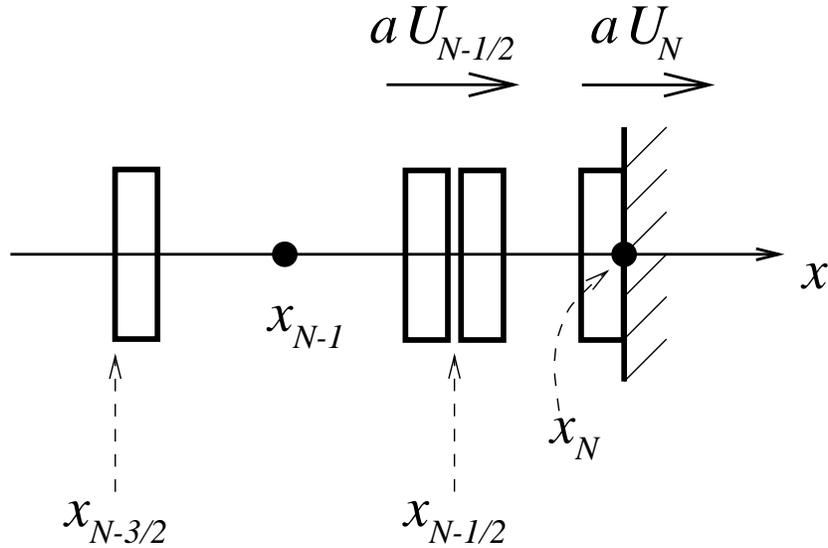
Bezeichnet man mit  $\Delta x = 2/N$  die Länge eines Teilintervalls, so gilt in Übereinstimmung mit (1.44)

$$x_i = i \cdot \Delta x, \quad i = 0, 1, 2, \dots, N. \quad (1.78)$$

Zu jedem Gitterpunkt mit dem Index  $i = 1, 2, \dots, N - 1$  gehört ein Kontrollvolumen  $[x_{i-1/2}, x_{i+1/2}]$  der Länge  $\Delta x$ , das im Inneren des Intervalls  $[0, 2]$  liegt. Wir bezeichnen diese Kontrollvolumina deshalb als *innere Kontrollvolumina*. Für jedes innere Kontrollvolumen kann die Upwind-Diskretisierung (1.71) direkt angewendet werden. Man erhält dadurch  $N - 1$  lineare Gleichungen.

Der Punkt  $x_0$  liegt direkt am Einströmrand des Berechnungsgebietes. Der Wert der Gitterfunktion  $U_0$  kann daher direkt aus der physikalischen Randbedingung ermittelt werden:

$$U_0 = \mu(t_{n+1}). \quad (1.79)$$



**Abbildung 1.8:** Skizze des  $N$ -ten Kontrollvolumens.

Am Ausströmrand ( $x_N = 2$ ) liegt keine physikalische Randbedingung vor. Zu diesem Gitterpunkt gehört auch kein inneres Kontrollvolumen, so daß die Formel (1.71) nicht angewendet werden kann. Man kann diesem Punkt aber ein „halbes“ Kontrollvolumen  $[x_{N-1/2}, x_N]$  der Länge  $\Delta x/2$  zuordnen (s. Abb. 1.8) und die Finite-Volumen-Formulierung wie für ein inneres Kontrollvolumen durch zweifache Integration der Konvektionsgleichung über  $[x_{N-1/2}, x_N] \times [t_n, t_{n+1}]$  herleiten. Analog zu (1.52) erhält man dann

$$\frac{U_N - \bar{U}_N}{\Delta t} + \frac{aU_N - aU_{N-1/2}}{\Delta x/2} = 0. \quad (1.80)$$

Hier braucht nur der numerische Konvektionsstrom  $aU_{N-1/2}$  approximiert werden, wofür der Upwind-Konvektionsstrom  $aU_{N-1}$  (s. (1.69)) verwendet wird. Die resultierende Gleichung für das Randvolumen

$$\frac{U_N - \bar{U}_N}{\Delta t} + a \frac{U_N - U_{N-1}}{\Delta x/2} = 0 \quad (1.81)$$

ergibt zusammen mit den  $N - 1$  Gleichungen für die inneren Kontrollvolumina und der Randbedingung (1.79) ein System, daß aus  $N + 1$  linearen Gleichungen für  $N + 1$  unbekannte Größen  $\{U_i, i = \overline{0, N}\}$  besteht. Die zugehörige Koeffizientenmatrix ist tridiagonal, das Gleichungssystem kann also durch die  $LU$ -Zerlegung für tridiagonale Systeme, bekannt als *Thomas-Algorithmus* (s. z.B. [45]), gelöst werden.

Die Definition des numerischen Gesamtfehlers  $E_n(\Delta x, \Delta t)$  (1.58) soll nun ebenfalls für den Fall eines endlichen Berechnungsgebietes angepaßt werden. Da die beiden Randpunkte,  $x_0$  und  $x_N$ , jeweils mit einem „halben“ Kontrollvolumen in das Berechnungsgebiet eingehen, werden die lokalen Fehler in diesen Punkten nur mit einem Faktor  $\Delta x/2$  gewichtet:

$$E_n(\Delta x, \Delta t) := \frac{\Delta x}{2}|U_0^n - u_0^n| + \Delta x \cdot \sum_{i=1}^{i=N-1} |U_i^n - u_i^n| + \frac{\Delta x}{2}|U_N^n - u_N^n|. \quad (1.82)$$

**Testbeispiel No.1.** Wir werden nun das Upwind-Verfahren zur Berechnung der numerischen Lösung für den Testfall No.1 mit  $x_0 = 0.4$  und  $\sigma = 0.1$  einsetzen. Unser Ziel ist, die Konvergenzordnung des Verfahrens zu bestimmen. Dazu müssen wir das Verhalten des Ortsfehlers  $E(\Delta x)$  bei  $\Delta x \rightarrow 0$  zu einem beliebigen aber festen Zeitpunkt untersuchen. Wir wählen dafür den Zeitpunkt  $T = 1$ . Der Ortsfehler ist als Grenzwert des Gesamtfehlers  $E_n(\Delta x, \Delta t)$  bei  $\Delta t \rightarrow 0$ ,  $n\Delta t = T$  definiert. Numerisch kann man diesen Grenzwert nur näherungsweise bestimmen, mit einer beliebig hohen aber dennoch fest im voraus vorgegebenen Genauigkeit. Wir werden uns dabei auf die ersten vier signifikantesten Stellen beschränken.

Wir beginnen mit der Maschenweite  $\Delta x = 0.1$ . Wir müssen nun die Zeitschrittweite  $\Delta t$  solange sukzessiv verkleinern, bis sich die ersten 4 signifikantesten Stellen des Gesamtfehlers (zum Zeitpunkt  $T = 1$ ) nicht mehr ändern. In der Tabelle 1.1 sind die Werte des Gesamtfehlers für verschiedene Zeitschrittweiten  $\Delta t$  aufgelistet. Als Näherungswert des Ortsfehlers  $E(\Delta x)$  für  $\Delta x = 0.1$  erhält man die Zahl 0.1754. In der Abbildung 1.9 ist der Verlauf der Lösung sowie des lokalen Fehlers  $|U_i - u_i|$  exemplarisch für  $\Delta t = 0.2$  und  $\Delta t = 0.00001$  dargestellt. Bei der Zeitschrittweite  $\Delta t = 0.00001$  ist der Zeitfehleranteil aus dem Gesamtfehler fast vollständig eliminiert (s. Tabelle 1.1). Bei der Zeitschrittweite  $\Delta t = 0.2$  liegt der Zeitfehleranteil allerdings noch in derselben Größenordnung wie der Ortsfehleranteil. **Die durchgeführte Elimination des Zeitfehlers aus dem Gesamtfehler ist daher eine notwendige Voraussetzung für die korrekte Bestimmung des Konvergenzverhaltens eines Verfahrens.**

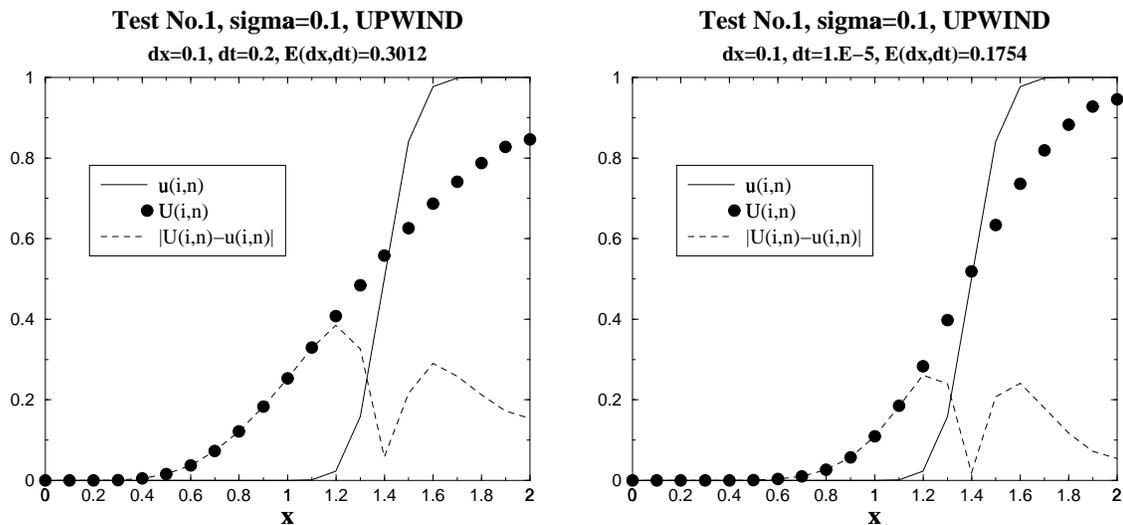
Wir wiederholen die Prozedur der Elimination des Zeitfehlers für unterschiedliche Maschenweiten und erhalten so den Ortsfehler als Funktion von  $\Delta x$  in einer tabellarischen Form (s. Tabelle 1.2). Wir werden eine solche Tabelle als **E- $\Delta x$ -Tabelle** bezeichnen.

Wir gehen nun davon aus, daß bei hinreichend kleinen Werten von  $\Delta x$  der Ortsfehler sich folgendermaßen verhält:

$$E(\Delta x) \approx C \Delta x^p. \quad (1.83)$$

$\Delta t$	$E_n(\Delta x, \Delta t)$	$\Delta t$	$E_n(\Delta x, \Delta t)$
.2	0.3012529491	.001	0.1764255058
.1	0.2517855533	.0005	0.1759179599
.05	0.2187887219	.0002	0.1756125256
.02	0.1944360214	.0001	0.1755105623
.01	0.1852530811	.00005	0.1754595522
.005	0.1804194249	.00002	0.1754289369
.002	0.1774349795	.00001	0.1754187303

**Tabelle 1.1:** Test No.1 ( $\sigma = 0.1$ ): Gesamtfehler als Funktion von  $\Delta t$  bei  $\Delta x = 0.1$ .



**Abbildung 1.9:** Test No.1 ( $\sigma = 0.1$ ): Verlauf der Lösung sowie der lokalen Fehler  $|U_i - u_i|$  für zwei unterschiedliche Zeitschrittweiten.

Für zwei unterschiedliche Maschenweiten  $\Delta x_1$  und  $\Delta x_2$  gilt dann

$$\frac{E(\Delta x_1)}{E(\Delta x_2)} \approx \left( \frac{\Delta x_1}{\Delta x_2} \right)^p, \quad (1.84)$$

woraus man folgende Approximationsformel für die Konvergenzordnung  $p$  erhält

$$p \approx \log \left( \frac{\Delta x_1}{\Delta x_2} \right) \left( \frac{E(\Delta x_1)}{E(\Delta x_2)} \right). \quad (1.85)$$

Wir setzen nun jeweils zwei nachfolgende Maschenweiten aus der Tabelle 1.2 in die Formel

$\Delta x$	$E(\Delta x)$	$\Delta x$	$E(\Delta x)$
1.000e-1	1.754e-1	5.000e-3	1.793e-2
5.000e-2	1.142e-1	2.500e-3	9.417e-3
2.500e-2	6.936e-2	1.000e-3	3.894e-3
1.000e-2	3.304e-2	5.000e-4	1.970e-3

**Tabelle 1.2:** Test No.1 ( $\sigma = 0.1$ ): Ortsfehler als Funktion von  $\Delta x$ .

$\Delta x$	1.000e-1	5.000e-2	2.500e-2	1.000e-2	5.000e-3	2.500e-3	1.000e-3
$p$	0.6190	0.7194	0.8093	0.8818	0.9290	0.9637	0.9830

**Tabelle 1.3:** Test No.1 ( $\sigma = 0.1$ ): Approximationswerte für Konvergenzordnung  $p$  für unterschiedliche Maschenweiten  $\Delta x$ .

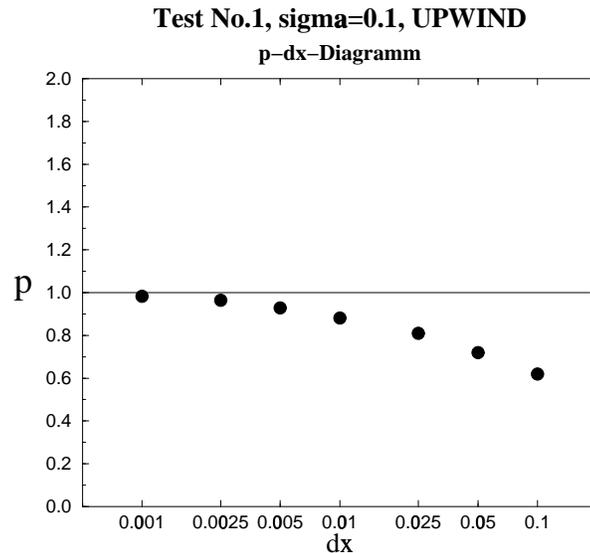
(1.85) ein und erhalten Approximationswerte für  $p$ , die in der Tabelle 1.3 zusammengetragen sind. Eine solche Tabelle werden wir entsprechend als **p- $\Delta x$ -Tabelle** bezeichnen.

Aus der  $p$ - $\Delta x$ -Tabelle wird ersichtlich, daß mit zunehmender Gitterfeinheit der empirisch ermittelte Wert der Approximationsordnung sich immer weniger von dem theoretisch hergeleiteten Wert von 1 unterscheidet.

Die hier vorgestellte Methode zur empirischen Ermittlung der Konvergenzordnung eines Diskretisierungsverfahrens wird in diesem wie auch in den folgenden Kapiteln noch mehrmals angewandt. Es wird dabei allerdings nur das Endergebnis präsentiert, entweder in Form einer  $p$ - $\Delta x$ -Tabelle oder in Form eines **p- $\Delta x$ -Diagramms**. Ein  $p$ - $\Delta x$ -Diagramm ist eine graphische Darstellung der Wertepaare aus der  $p$ - $\Delta x$ -Tabelle, wobei die  $p$ -Achse linear und die  $\Delta x$ -Achse logarithmisch skaliert wird. Eine logarithmische Skalierung ist sinnvoll, da auf einer linear-skalierten Achse der Abstand zwischen zwei benachbarten Werten von  $\Delta x$  immer kleiner wird, wenn  $\Delta x$  gegen Null geht. Ein der  $p$ - $\Delta x$ -Tabelle 1.3 entsprechendes  $p$ - $\Delta x$ -Diagramm ist in der Abb. 1.10 dargestellt.

Analog zum  $p$ - $\Delta x$ -Diagramm können die Wertepaare aus einer  $E$ - $\Delta x$ -Tabelle in einem **E- $\Delta x$ -Diagramm** graphisch dargestellt werden. Hier werden allerdings beide Achsen logarithmisch skaliert, da sowohl die  $\Delta x$ -Werte als auch die  $E$ -Werte immer dichter beieinander liegen, wenn  $\Delta x$  gegen Null geht. Eine doppel-logarithmische Skalierung hat auch einen anderen Vorteil. Benimmt sich der Ortsfehler bei hinreichend kleinen Werten von  $\Delta x$  wie (1.83), so gilt auch

$$\log E(\Delta x) \approx \log C + p \cdot \log \Delta x. \quad (1.86)$$

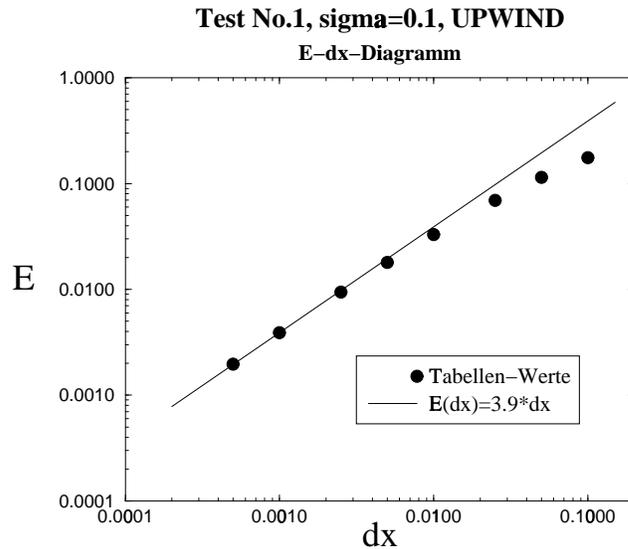


**Abbildung 1.10:** Test No.1 ( $\sigma = 0.1$ ): Das  $p$ - $\Delta x$ -Diagramm für das Upwind-Verfahren.

Das bedeutet, daß bei einer log-log-Skalierung der Ortsfehler einen linearen Verlauf hat, mit einer Steigung, die proportional zur Konvergenzordnung  $p$  ist. Ein  $E$ - $\Delta x$ -Diagramm ist insbesondere dann nützlich, wenn man die Effektivität mehrerer Diskretisierungsverfahren an einem Testbeispiel vergleicht. Man kann damit gleichzeitig sowohl die Konvergenzordnung als auch die absoluten Werte des Ortsfehlers beim Einsatz verschiedener Verfahren miteinander vergleichen. Für das hier diskutierte Testbeispiel ist das  $E$ - $\Delta x$ -Diagramm für das Upwind-Verfahren in der Abb. 1.11 dargestellt.

Zum Abschluß schauen wir uns die Lösungsprofile auf vier unterschiedlichen numerischen Gittern an (Abb. 1.12). Man sieht, daß die Annäherung der numerischen Lösung an die exakten Lösungsprofile nur sehr langsam erfolgt, und daß selbst bei einer Maschenweite  $\Delta x = 0.005$  (das entspricht 400 Gitterpunkten) der Unterschied zwischen beiden Profilen sehr deutlich zu sehen ist. Interessant ist auch der Charakter des numerischen Ortsfehlers. Qualitativ wird das Verhalten der Lösung korrekt wiedergegeben, die berechneten Profile sind aber durch einen im Vergleich zur exakten Lösung flacheren Verlauf gekennzeichnet. Der Grund für dieses diffusive Verhalten des Upwind-Verfahrens wird im weiteren Verlauf dieses Abschnittes erklärt.

**Testbeispiel No.2.** Wir wollen nun die Konvergenzordnung des Upwind-Verfahrens für den Testfall No.2 bestimmen. Wir lassen eine ausführliche Beschreibung aller Zwischenschritte aus, und präsentieren nur das Endergebnis in Form eines  $p$ - $\Delta x$ -Diagramms (Abb. 1.13). Wie bereits erwähnt, beträgt die empirisch ermittelte Konvergenzordnung in diesem



**Abbildung 1.11:** Test No.1 ( $\sigma = 0.1$ ): Das  $E$ - $\Delta x$ -Diagramm für das Upwind-Verfahren.

Fall nur 0.5, was auf die Diskontinuität der Lösung zurückzuführen ist. In Abb. 1.14 sind die Lösungsprofile für 2 unterschiedliche Maschenweiten aufgetragen. Während im Testfall No.1 auf einem Gitter mit 400 Kontrollvolumina eine durchaus brauchbare Übereinstimmung mit der exakten Lösung festgestellt werden konnte, kann im Testfall No.2 auf dem gleichen Gitter kein zufriedenstellendes Ergebnis erzielt werden (Abb. 1.14, links). Selbst auf einem Gitter mit der Maschenweite  $\Delta x = 0.0005$  (entspricht 4000 (!) Kontrollvolumina) kann man deutlich die Unterschiede zwischen der exakten und der numerischen Lösung erkennen (Abb. 1.14, rechts).

**Steile Gradienten.** Das Problem mit dem schlechten Konvergenzverhalten des Upwind-Verfahrens tritt nicht nur im Falle einer diskontinuierlichen Lösung auf, sondern auch dann, wenn die Lösung steile Gradienten aufweist, wie das z.B. im Testbeispiel No.1 für  $\sigma = 0.02$  der Fall ist (s. Abb. 1.4). Aus dem  $p$ - $\Delta x$ -Diagramm (Abb. 1.15) erkennt man, daß die empirisch ermittelte Konvergenzordnung sich zwar dem theoretischen Wert von 1 nähert, aber selbst auf einem sehr feinen Gitter mit der Maschenweite  $\Delta x = 0.001$  noch weit davon entfernt ist. Das kommt daher, daß der Anstieg der Lösung vom Wert 0 auf den Wert 1 sich auf einem sehr engen Intervall vollzieht, und obwohl dieser Anstieg auf eine kontinuierliche Weise erfolgt, wird diese Kontinuität nur auf sehr feinen Gittern vom numerischen Verfahren erfaßt. Auf den gröbereren Gittern dagegen, „fällt“ dieser stetige Übergang durch das numerische Gitter, und die numerische Lösung benimmt sich wie im Falle einer diskontinuierlichen Lösungsfunktion.

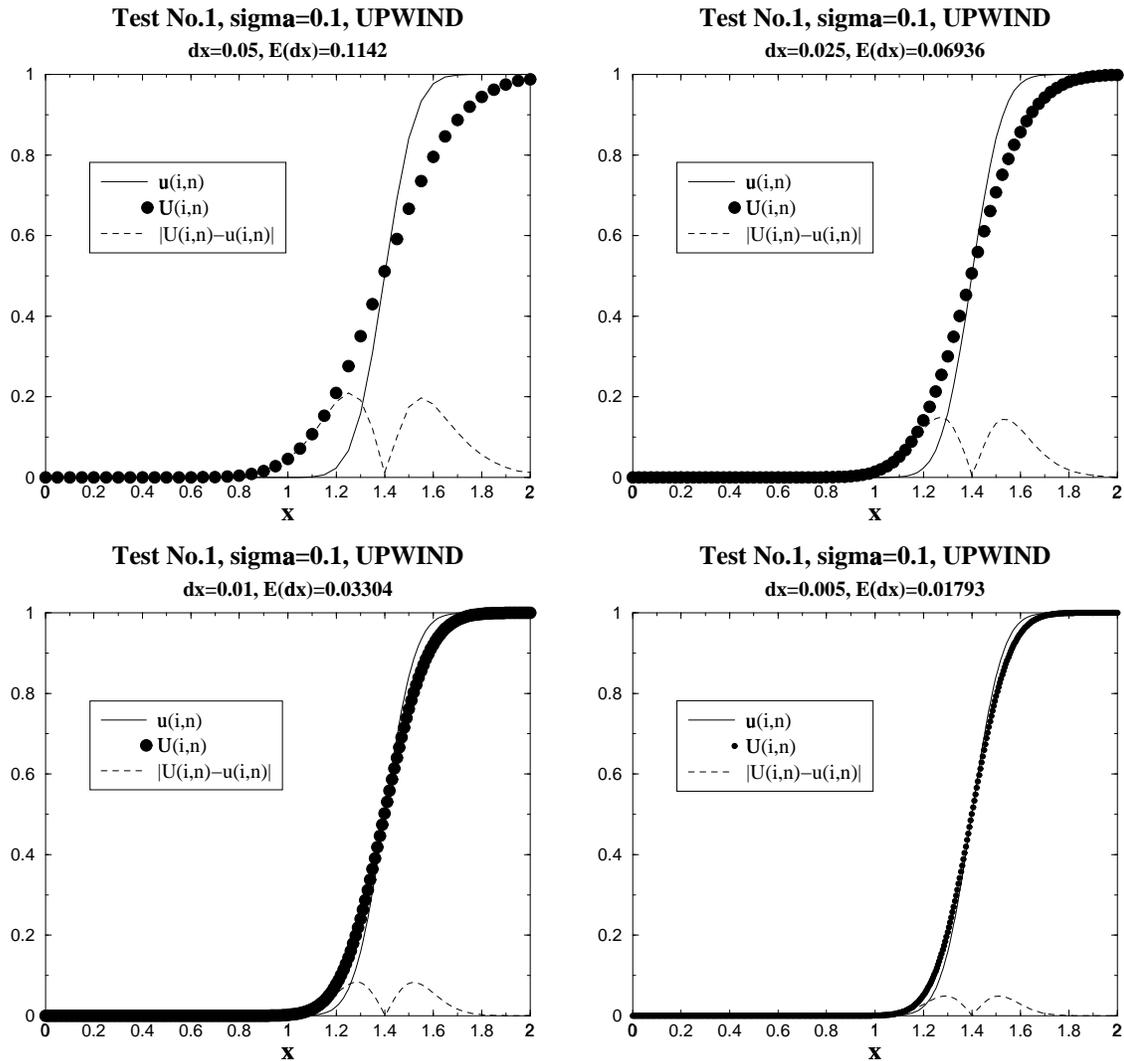
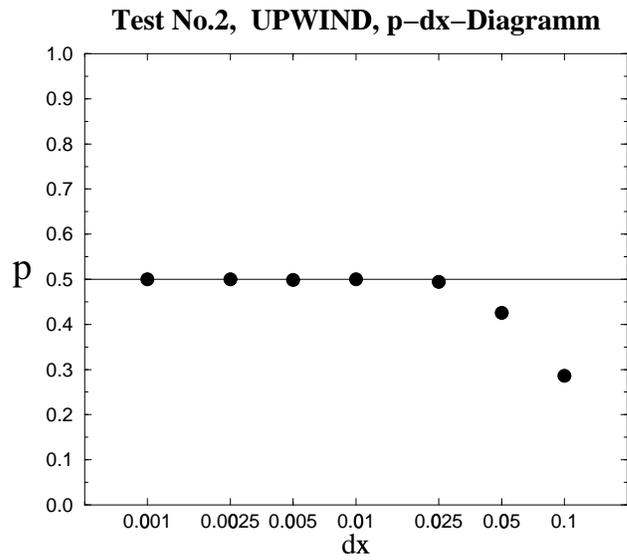


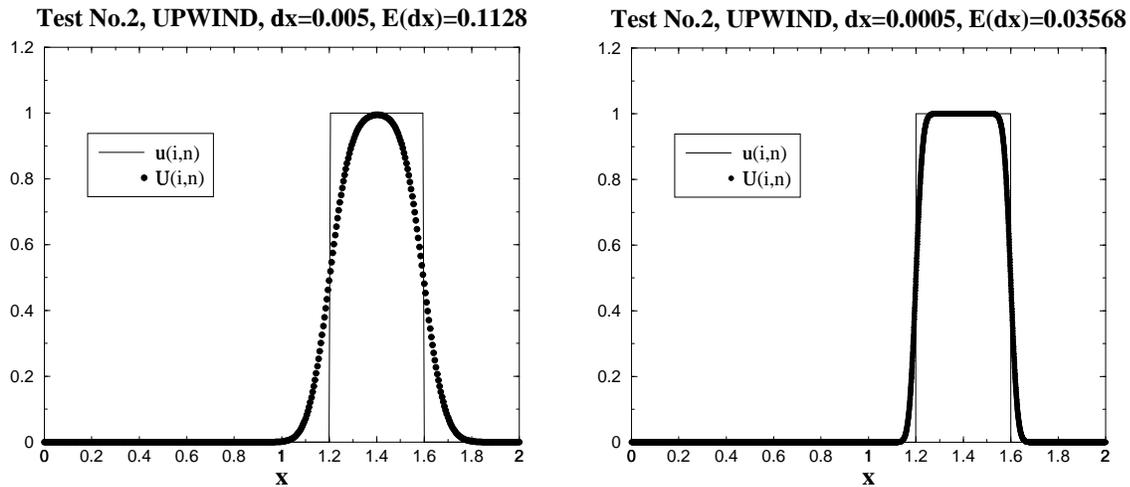
Abbildung 1.12: Test No.1 ( $\sigma = 0.1$ ): Verlauf der Lösung sowie der lokalen Fehler  $|U_i - u_i|$  für vier unterschiedliche Maschenweiten.

**Numerische Diffusion.** Wir haben bereits gesehen, daß die Ortsfehler beim Einsatz des Upwind-Verfahrens vor allem dadurch verursacht werden, daß die berechnete Lösung einen flacheren Verlauf hat, als die exakte Lösung der linearen Konvektionsgleichung. Das gilt sowohl für den Testfall No.1 als auch für den Testfall No.2. Dieses stark diffusive Verhalten der Upwind-Diskretisierung läßt sich auch theoretisch begründen. Schaut man sich den örtlichen Abbruchfehler (1.76) an, so stellt man fest, daß die Differenzenformel (1.71) die ursprüngliche Konvektionsgleichung (1.42) zwar mit erster Ordnung approximiert, aber noch besser, und zwar mit zweiter Ordnung, approximiert sie die Konvektions-Diffusions-Gleichung

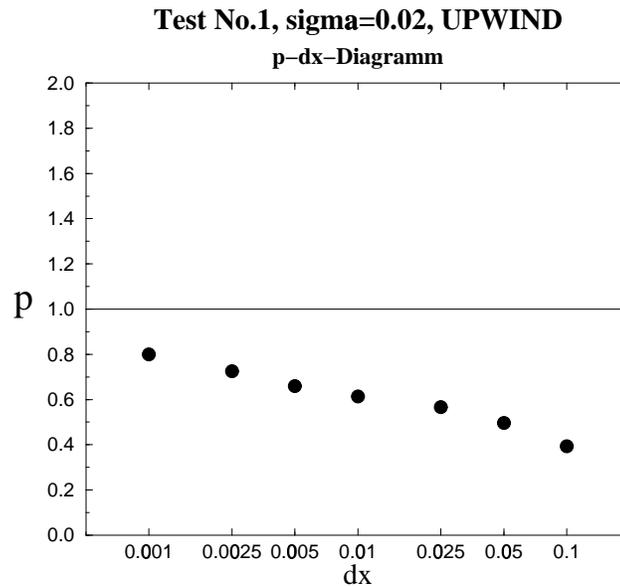
$$u_t + au_x = Du_{xx} \quad (1.87)$$



**Abbildung 1.13:** Testbeispiel No.2: Das  $p$ - $\Delta x$ -Diagramm für das Upwind-Verfahren.



**Abbildung 1.14:** Testbeispiel No.2: Verlauf der Lösung für 2 unterschiedliche Maschenweiten.



**Abbildung 1.15:** Test No.1 ( $\sigma = 0.02$ ): Das  $p$ - $\Delta x$ -Diagramm für das Upwind-Verfahren.

mit

$$D = \frac{a\Delta x}{2}. \quad (1.88)$$

Würde man die Gleichung (1.87) bei einem fest vorgegebenen Wert von  $\Delta x$  (als Parameter) exakt lösen, und diese exakte Lösung zusammen mit der numerischen Lösung der Konvektionsgleichung mittels Upwind-Verfahrens (auf einem Gitter mit derselben Maschenweite  $\Delta x$ ) graphisch darstellen, so würde man praktisch keinen Unterschied zwischen den beiden Profilen sehen können.

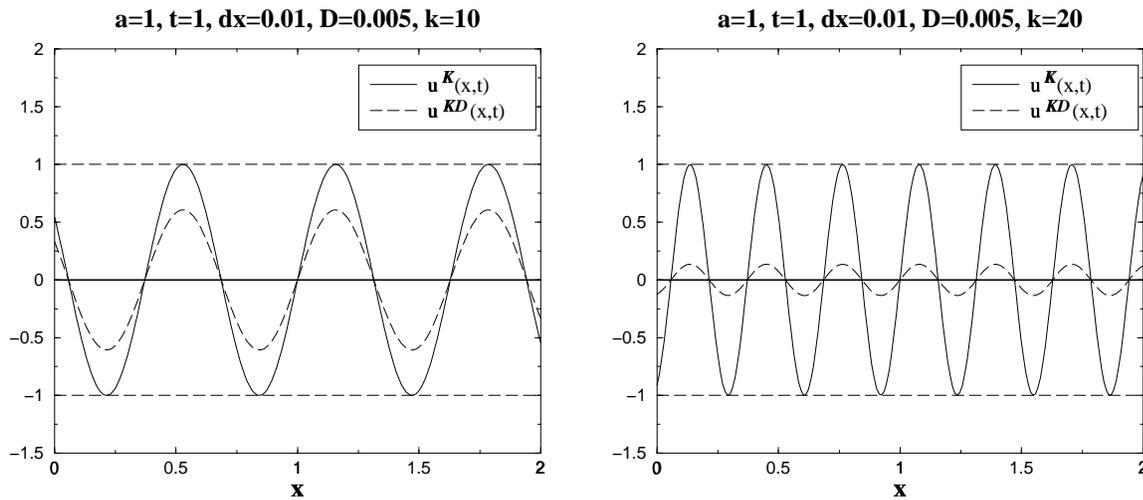
Die Größe  $D$  wird als *numerische Diffusion* bezeichnet. Wir sehen, daß die Größe der numerischen Diffusion linear von  $\Delta x$  abhängt und mit zunehmender Gitterverfeinerung gegen Null geht. Da man allerdings immer nur mit einer endlich großen Maschenweite arbeitet, ist der Wert von  $D$  stets größer als Null. Das erklärt sowohl das diffusive Verhalten des Upwind-Verfahrens als auch die Tatsache, daß die Glättung der Lösung mit zunehmender Verfeinerung des Gitters abnimmt (s. Abb. 1.12).

Den Einfluß der numerischen Diffusion auf die Lösung der linearen Konvektionsgleichung kann man qualitativ untersuchen, wenn man als Anfangsbedingung für das Cauchy-Problem (1.19, 1.20) die Funktion

$$u_0(x) = \sin(kx), \quad x \in (-\infty, +\infty). \quad (1.89)$$

wählt. Die Lösung der Konvektionsgleichung lautet in diesem Fall

$$u^K(x, t) = \sin(k(x - at)), \quad (1.90)$$



**Abbildung 1.16:** Vergleich der Funktionen  $u^K(x, t)$  und  $u^{KD}(x, t)$  für zwei Parametersätze.

während die Lösung der Konvektions-Diffusions-Gleichung (1.87) folgendermaßen aussieht:

$$u^{KD}(x, t) = e^{-Dk^2t} \sin(k(x - at)) = e^{-Dk^2t} \cdot u^K(x, t). \quad (1.91)$$

Für zwei Werte von  $k$  sind die Verläufe der Funktionen  $u^K(x, t)$  und  $u^{KD}(x, t)$  zum Zeitpunkt  $t = 1$  in der Abb. 1.16 exemplarisch dargestellt. Man sieht, daß das Lösungsprofil mit der richtigen Geschwindigkeit entlang der  $x$ -Achse transportiert wird, die Amplitude der Lösung nimmt aber mit der Zeit exponential ab. Die Abnahmerate nimmt sowohl mit dem steigenden Diffusionskoeffizienten  $D$  als auch mit der steigenden Wellenzahl  $k$  zu.

Löst man nun die lineare Konvektionsgleichung für eine beliebige Funktion  $u_0(x)$  auf einem Gitter mit der Maschenweite  $\Delta x$ , so kann auf diesem Gitter die Anfangsbedingung in eine endliche Fourier-Reihe mit der maximalen Wellenzahl („wave number“)  $k_{max} = \pi / \Delta x$  zerlegt werden. Bei einer Gitterverfeinerung nimmt der numerische Diffusionskoeffizient einerseits linear ab. Andererseits nimmt die maximale Wellenzahl und die Anzahl der Summanden in der Fourier-Reihe zu. Treten in der Anfangsbedingung steile Gradienten auf, so sind die neuen hochfrequenten Summanden übergewichtet, solange das Gitter nicht fein genug ist. In diesem Bereich wird die theoretische Konvergenzordnung von 1 empirisch nicht bestätigt. Erst wenn das numerische Gitter so fein ist, daß der Beitrag der hochfrequenten Summanden exponentiell abnimmt, wird die Konvergenzordnung von 1 tatsächlich erreicht. Im Testfall No.1 mit  $\sigma = 0.02$  passiert das bei einer kleineren Maschenweite, als im Fall mit  $\sigma = 0.1$ , was erklärt, warum bei derselben Auflösung das Upwind-Verfahren für  $\sigma = 0.02$  viel schlechter abschneidet als für  $\sigma = 0.1$ .

Diese Vorgehensweise zur Untersuchung des qualitativen Verhaltens der numerischen Lö-

sung mittels einer Differentialgleichung, die von einem Diskretisierungsverfahren mit einer höheren Ordnung approximiert wird, als die ursprüngliche lineare Konvektionsgleichung, werden wir noch mehrmals verwenden. Die entsprechende Differentialgleichung hat in der Literatur einen eigenen Namen erhalten. Sie wird als „modifizierte Gleichung“ („modified equation“ [77]), manchmal auch als „äquivalente Differentialgleichung“ („equivalent differential equation“ [45]) eines Verfahrens bezeichnet.

**Upwind-Verfahren: Zusammenfassung.** Wir wollen nun die gewonnenen Erkenntnisse aus der Untersuchung des Upwind-Verfahrens zusammenfassen. Die Upwind-Diskretisierung hat die Approximationsordnung von 1 und ist ein absolut stabiles Verfahren. Sie hat damit die Konvergenzordnung 1. Diese Aussage gilt allerdings nur für hinreichend glatte Lösungen. Im Falle einer diskontinuierlichen (schwachen) Lösung der linearen Konvektionsgleichung hat das Upwind-Verfahren eine niedrigere Konvergenzordnung von 0.5. Ist die Lösung glatt, weist aber steile Gradienten auf, so ist die theoretische Konvergenzordnung nur auf sehr feinen Gittern erreichbar. Bei den in der Praxis verwendbaren gröberen Gittern ist die (empirisch ermittelte) Konvergenzordnung wesentlich niedriger. Das Upwind-Verfahren liefert physikalisch sinnvolle Lösungen, die allerdings wesentlich flacher verlaufen, als die exakte Lösung der Differentialgleichung. Der Grund dafür liegt in der numerischen Diffusion, die mit der Upwind-Diskretisierung verbunden ist, da das Upwind-Verfahren mit zweiter Ordnung eine Konvektions-Diffusions-Gleichung approximiert. Das Ausmaß der numerisch bedingten Diffusion ist dabei linear proportional zu der Konvektionsgeschwindigkeit  $a$  und zu der Maschenweite  $\Delta x$ .

**Downwind-Verfahren.** In einer ähnlichen Weise, wie wir das Upwind-Verfahren untersucht haben, wollen wir nun auch das Downwind-Verfahren behandeln. Für den lokalen Abbruchfehler der Downwind-Diskretisierung gilt:

$$L(\Delta x, \Delta t) = \frac{u(x, t + \Delta t) - u(x, t)}{\Delta t} + a \frac{u(x + \Delta x, t + \Delta t) - u(x, t + \Delta t)}{\Delta x}. \quad (1.92)$$

Der Grenzübergang  $\Delta t \rightarrow 0$  führt zu

$$L(\Delta x) = \lim_{\Delta t \rightarrow 0} L(\Delta x, \Delta t) = u_t + a \frac{u(x + \Delta x, t) - u(x, t)}{\Delta x}. \quad (1.93)$$

Einsatz der Taylorentwicklung der Funktion  $u(\cdot, t)$  um den Punkt  $x$

$$u(x + \Delta x, t) = u(x, t) + \Delta x u_x(x, t) + \frac{\Delta x^2}{2} u_{xx}(x, t) + O(\Delta x^3) \quad (1.94)$$

in die Gleichung (1.93) führt nun zu

$$L(\Delta x) = u_t + a u_x + \frac{a \Delta x}{2} u_{xx} + O(\Delta x^2) = \frac{a \Delta x}{2} u_{xx} + O(\Delta x^2) = O(\Delta x), \quad (1.95)$$

und wir stellen fest, daß das Downwind-Verfahren die Approximationsordnung 1 hat.

Im Unterschied zur Upwind-Diskretisierung ist das Downwind-Verfahren allerdings numerisch instabil, und zwar unabhängig von der Zeitschrittweite  $\Delta t$ . Der Grund für diese Instabilität liegt darin, daß dieses Verfahren wichtige Eigenschaften der Konvektionsgleichung verletzt. Während sich die Lösung der linearen Konvektionsgleichung mit der positiven Geschwindigkeit  $a$  entlang der  $x$ -Achse fortpflanzt, die Information also von links nach rechts „propagiert“, wird bei der Downwind-Approximation für  $U_{i-1/2}$  der Wert der Gitterfunktion verwendet, der stromabwärts liegt. Man approximiert damit sozusagen *gegen* die Stromrichtung. Wir wollen nun das Verhalten des Downwind-Verfahrens anhand des Testfalls No.1 veranschaulichen, und dabei gleichzeitig die etwas vage Aussage über die Fehlapproximation verdeutlichen.

Die Behandlung der Randelemente erfolgt auf die gleiche Weise wie bei Upwind-Diskretisierung. Am Einströmrand gilt die Randbedingung (1.79), und am Ausströmrand verwendet man die Gleichung (1.80) für das „halbe“ Kontrollvolumen, in die man die Downwind-Approximation  $aU_N$  (s. (1.70)) für den numerischen Konvektionsstrom  $aU_{N-1/2}$  einsetzt. Zur besseren Übersichtlichkeit fassen wir das resultierende Gleichungssystem hier noch einmal zusammen:

$$\begin{aligned}
 U_0 &= \mu(t_{n+1}) & (1.96) \\
 \frac{U_1 - \bar{U}_1}{\Delta t} + a \frac{U_2 - U_1}{\Delta x} &= 0 \\
 \frac{U_2 - \bar{U}_2}{\Delta t} + a \frac{U_3 - U_2}{\Delta x} &= 0 \\
 &\dots \\
 \frac{U_{N-1} - \bar{U}_{N-1}}{\Delta t} + a \frac{U_N - U_{N-1}}{\Delta x} &= 0 \\
 \frac{U_N - \bar{U}_N}{\Delta t} + a \frac{U_N - U_N}{\Delta x/2} &= 0. & (1.97)
 \end{aligned}$$

Bereits anhand dieses Gleichungssystems kann man feststellen, daß das Downwind-Verfahren nicht gegen die exakte Lösung der Differentialgleichung konvergieren kann. Man sieht z.B., daß die Gleichung (1.96) von den anderen Gleichungen des Systems vollständig entkoppelt ist, da die unbekannte Größe  $U_0$  in keine weitere Gleichung eingeht. Die physikalische Randbedingung am Einströmrand bestimmt zwar den Wert  $U_0$ , hat aber überhaupt keinen Einfluß auf die Lösung im Inneren des Berechnungsgebietes. Am rechten Rand wird

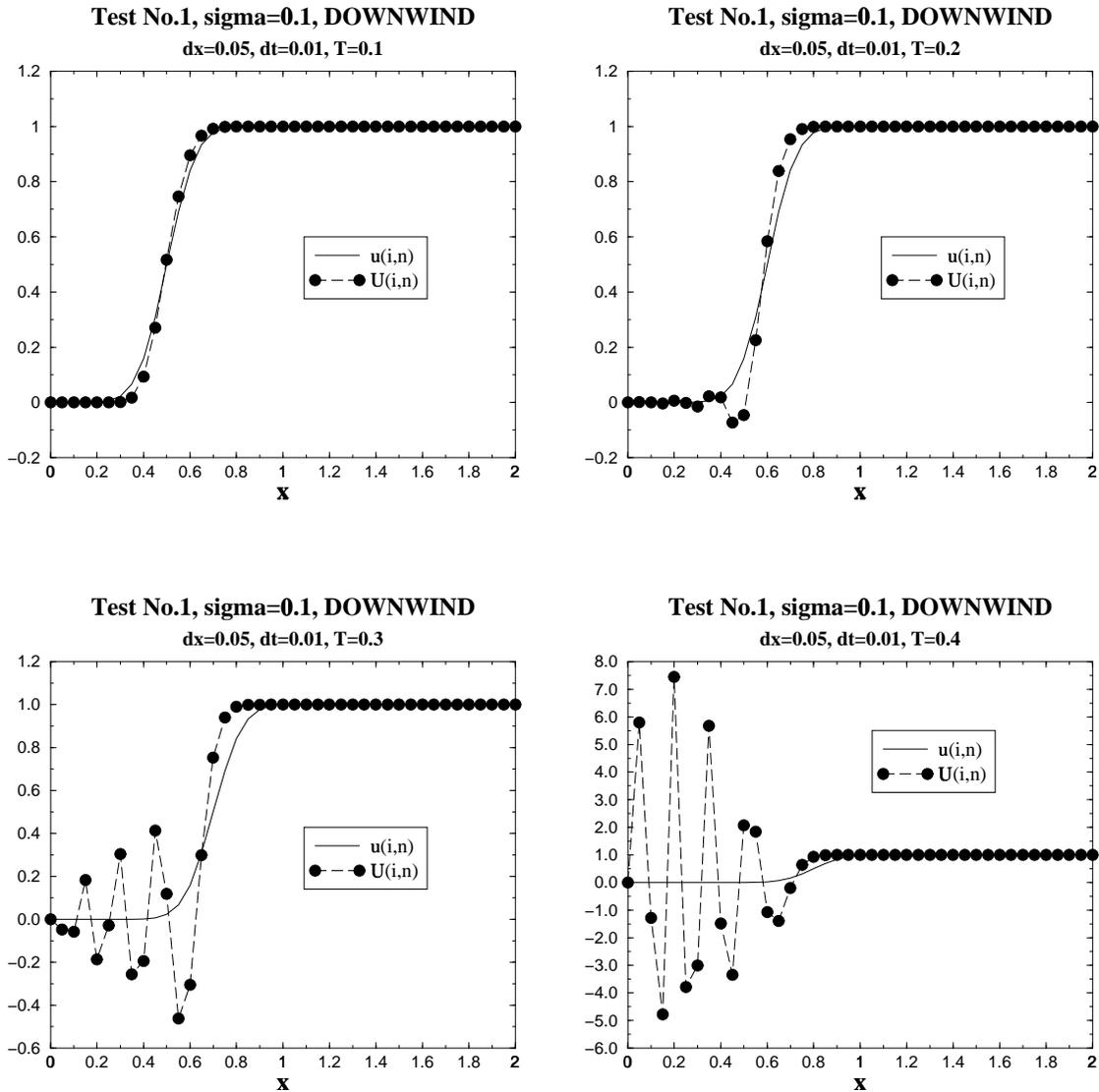


Abbildung 1.17: Test No.1 ( $\sigma = 0.1$ ): Verlauf der Lösung zu 4 unterschiedlichen Zeitpunkten.

ebenfalls keine sinnvolle Lösung berechnet, da aus Gleichung (1.97) folgt, daß  $U_N = \bar{U}_N$  gilt, die numerische Lösung am rechten Rand ändert sich also mit der Zeit nicht. Die Instabilität des Verfahrens sorgt außerdem dafür, daß im Inneren des Berechnungsgebietes Oszillationen entstehen, die mit der Zeit anwachsen, wie aus der Abb.1.17 für die Lösung auf einem Gitter mit der Maschenweite  $\Delta x = 0.05$  ersichtlich ist. Reduziert man die Maschenweite, fallen die Oszillationen noch stärker aus.

Das qualitative Verhalten der mit dem Downwind-Verfahren errechneten Lösung kann man ebenfalls mit Hilfe der modifizierten Gleichung erklären. Schaut man sich den örtlichen Abbruchfehler (1.95) an, so stellt man fest, daß die Differenzenformel (1.72) mit zweiter

Ordnung wiederum die Konvektions-Diffusions-Gleichung

$$u_t + au_x = Du_{xx} \quad (1.98)$$

approximiert, wobei diesmal — im Gegensatz zum Upwind-Verfahren — der Diffusionskoeffizient  $D$  negativ ist:

$$D = -\frac{a\Delta x}{2}. \quad (1.99)$$

Während bei einem positiven Diffusionskoeffizienten die Gradienten in der Lösung mit der Zeit abgeflacht werden, passiert bei einem negativen Diffusionskoeffizienten genau das Gegenteil: die kleinsten Unterschiede im Lösungsverlauf wachsen mit der Zeit an. Aufgrund von numerischen Rundungsfehlern können kleine Unebenheiten im Lösungsverlauf selbst dort entstehen, wo die Lösung zunächst konstant war. Mit der Gitterverfeinerung nimmt die Anzahl solcher Unebenheiten sogar noch zu, so daß die Frequenz der Oszillationen ebenfalls größer wird.

## 1.5 Zentral-Differenz

Im letzten Abschnitt haben wir zwei Diskretisierungsverfahren untersucht, die auf einer stückweise konstanten Interpolation der numerischen Lösung zwischen den Gitterpunkten basieren. Eines von diesen beiden Verfahren, die Upwind-Diskretisierung, hat sich als stabiles Verfahren erwiesen, das auch physikalisch sinnvolle Lösungen liefert, ist allerdings nur erster Ordnung genau, was eine viel zu langsame Konvergenzordnung zu Folge hat. Wir wollen nun versuchen, ein Verfahren mit einer höheren Konvergenzordnung zu erzeugen, indem wir anstatt der stückweise konstanten Interpolation eine lineare Interpolation der Lösung zwischen den Gitterpunkten verwenden:

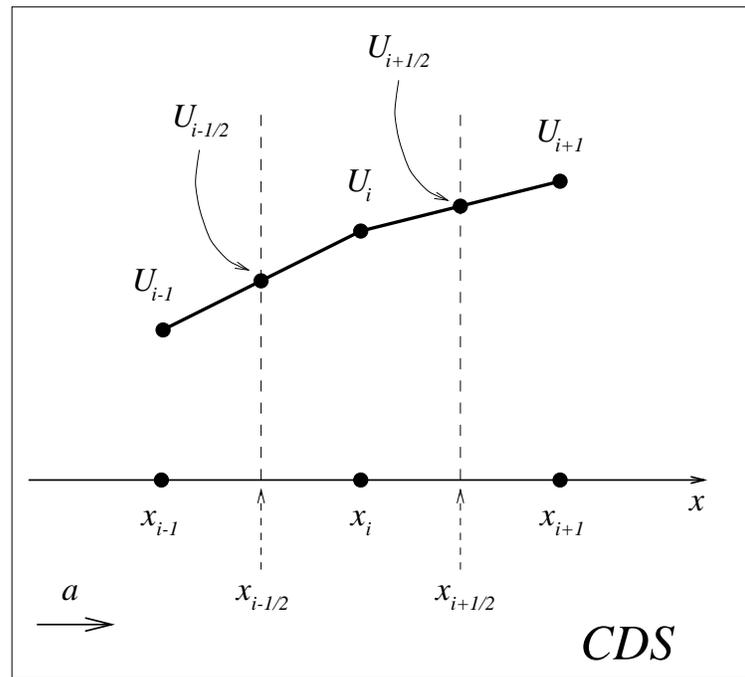
$$U(x) = U_{i-1} + (x - x_{i-1}) \cdot \frac{U_i - U_{i-1}}{\Delta x}, \quad x \in [x_{i-1}, x_i] \quad (1.100)$$

Der Verlauf der so definierten Funktion  $U(x)$  ist in der Abb. 1.18 skizziert.

Für eine so definierte lineare Interpolation ist der Wert der Funktion  $U(x)$  an der Stelle  $x_{i-1/2}$  gleich dem Mittelwert von  $U_{i-1}$  und  $U_i$ . Für den numerischen Konvektionsstrom gilt daher

$$aU_{i-1/2} = F^{cds}(U; i - 1/2) = a\frac{U_{i-1} + U_i}{2}. \quad (1.101)$$

Setzt man diesen Konvektionsstrom in die Finite-Volumen-Gleichung (1.53) ein, so ergibt sich eine Diskretisierung der Konvektionsgleichung, die der Approximation der örtlichen Ableitung durch die *zentrale Differenz* entspricht:



**Abbildung 1.18:** Lineare Interpolation der Lösung zwischen den Gitterpunkten.

$$\boxed{\text{CDS : } \frac{U_i - \bar{U}_i}{\Delta t} + a \frac{U_{i+1} - U_{i-1}}{2\Delta x} = 0.} \quad (1.102)$$

Für das resultierende Diskretisierungsverfahren werden wir die in der Literatur übliche Abkürzung *CDS* („central differencing scheme“) benutzen.

Analog zum vorigen Abschnitt kann man den örtlichen Abbruchfehler für das CDS-Verfahren bestimmen:

$$L(\Delta x) = u_t + a \frac{u(x + \Delta x, t) - u(x - \Delta x, t)}{2\Delta x}. \quad (1.103)$$

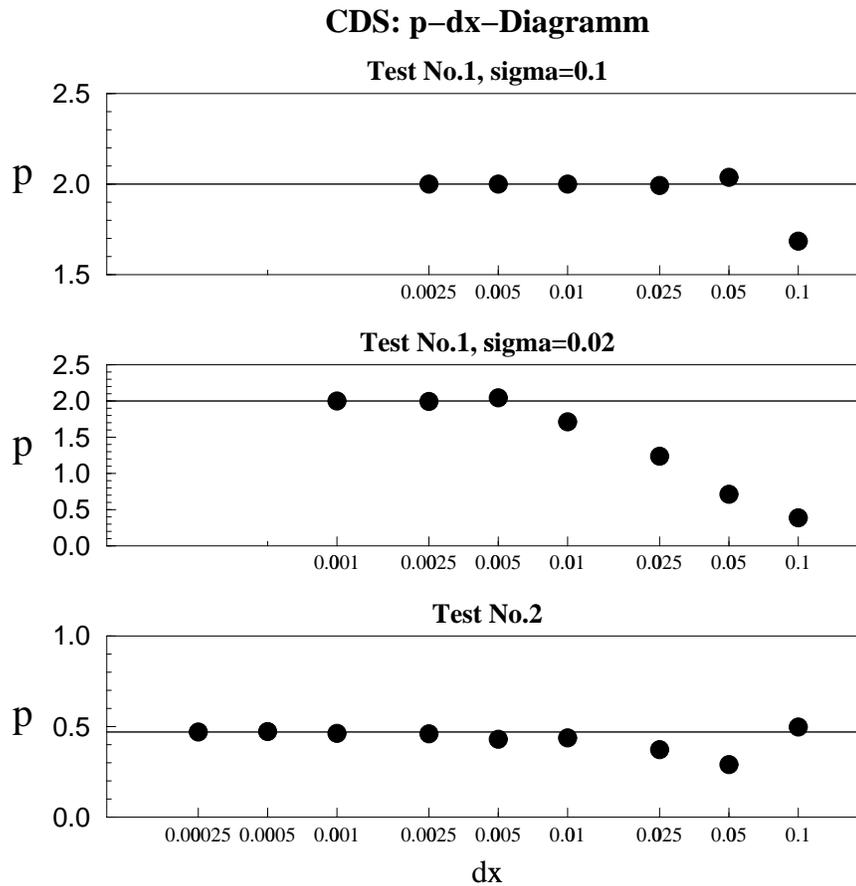
Einsatz der Taylorentwicklung der Funktion  $u(\cdot, t)$  um den Punkt  $x$  in die rechte Seite führt zu:

$$L(\Delta x) = u_t + au_x + \frac{a\Delta x^2}{6}u_{xxx} + O(\Delta x^4). \quad (1.104)$$

bzw., da die Summe der ersten beiden Terme gleich Null ist,

$$L(\Delta x) = \frac{a\Delta x^2}{6}u_{xxx} + O(\Delta x^4) = O(\Delta x^2). \quad (1.105)$$

Das CDS-Verfahren besitzt somit die Approximationsordnung von 2. Wie das Upwind-Verfahren ist das CDS-Verfahren ebenfalls absolut stabil. Daraus folgt, daß es konvergent ist und auf glatten Lösungen die Konvergenzordnung 2 hat.



**Abbildung 1.19:** Testfälle 1 und 2: Das  $p$ - $\Delta x$ -Diagramm für das CDS-Verfahren.

Wir werden nun das Verhalten der CDS-Diskretisierung an unseren Testfällen 1 und 2 untersuchen. Die Behandlung der Randelemente erfolgt auf die gleiche Weise wie bei Upwind-Diskretisierung. Am Einströmrand gilt die Randbedingung (1.79), und am Ausströmrand verwendet man die Gleichung (1.80) für das „halbe“ Kontrollvolumen, in die man die CDS-Approximation  $a(U_{N-1} + U_N)/2$  (s. (1.101)) für den numerischen Konvektionsstrom  $aU_{N-1/2}$  einsetzt. Die resultierende Formel lautet:

$$\frac{U_N - \bar{U}_N}{\Delta t} + a \frac{U_N - U_{N-1}}{\Delta x} = 0. \quad (1.106)$$

In Abb. 1.19 sind die  $p$ - $\Delta x$ -Diagramme für beide Testfälle dargestellt. Man sieht, daß die empirische Konvergenzordnung im Testfall 1 für beide  $\sigma$ -Werte 2 beträgt. Im Testfall 2 dagegen wird lediglich eine Konvergenzordnung von 0.47 erreicht. Wir wollen nun das Verhalten der numerischen Lösung für beide Testfälle detailliert untersuchen.

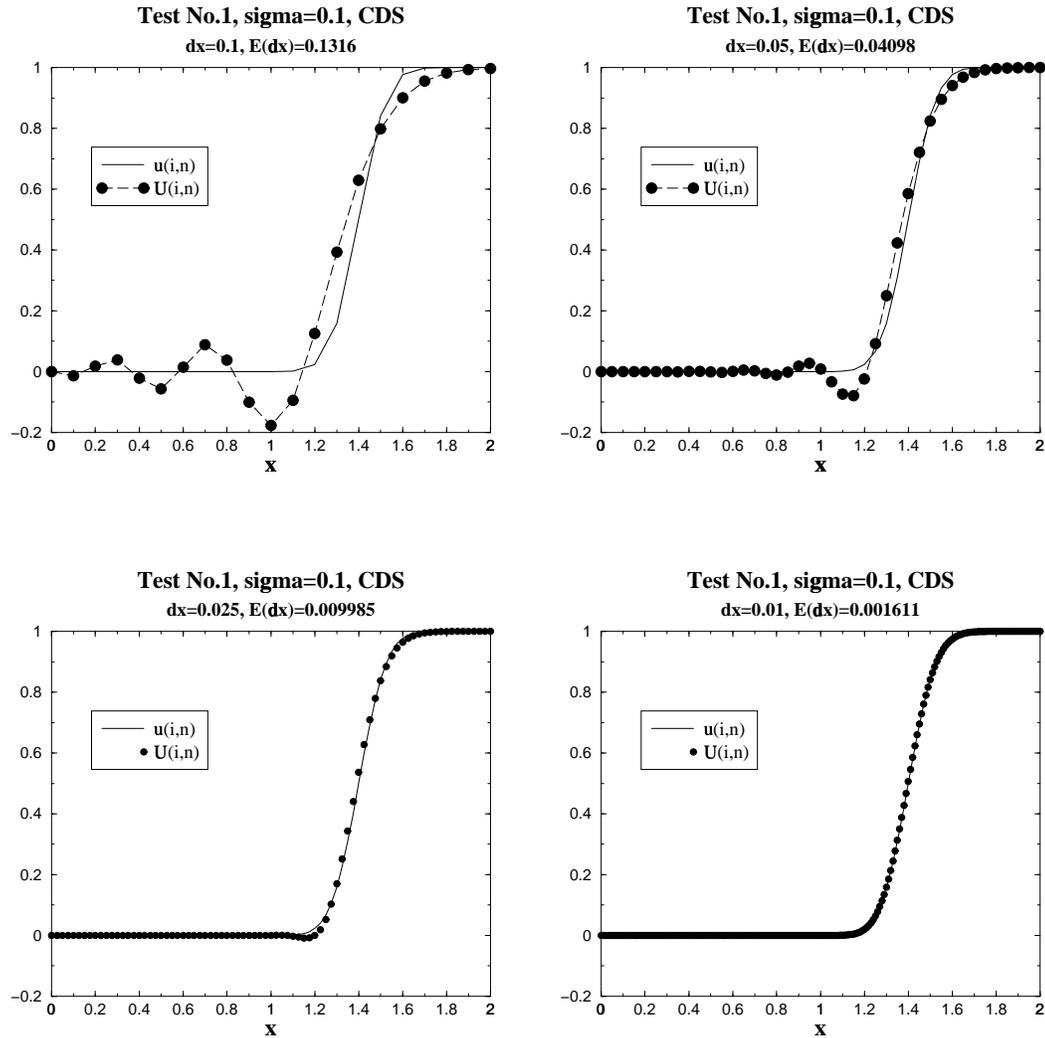


Abbildung 1.20: Test No.1 ( $\sigma = 0.1$ ): Verlauf der Lösung für 4 unterschiedliche Maschenweiten.

**Testbeispiel No.1.** In der Abb. 1.20 sind die Profile der Lösung auf 4 unterschiedlichen Gittern für den Fall  $\sigma = 0.1$  dargestellt. Man sieht, daß bereits bei einer Auflösung von 40 Gitterpunkten ( $\Delta x = 0.05$ ) eine gute Übereinstimmung mit der exakten Lösung im Frontbereich erzielt wird. Der Großanteil des numerischen Fehlers resultiert aus dem Verlauf der Lösung *hinter* der Front. Hier weisen die Lösungsprofile starke Oszillationen auf. Im Gegensatz zu Downwind-Diskretisierung nehmen diese Oszillationen mit der Gitterverfeinerung jedoch ab. Bei einer Auflösung von 200 Gitterpunkten ( $\Delta x = 0.01$ ) sind sie (zumindest optisch) nicht mehr erkennbar. Im Fall von  $\sigma = 0.02$  sind die Oszillationen allerdings viel stärker ausgeprägt (Abb. 1.21) und sind selbst bei einer Auflösung von 200 Gitterpunkten noch deutlich zu sehen.

Da in der exakten Lösung keine solche Schwankungen vorhanden sind, sind diese Oszillatio-

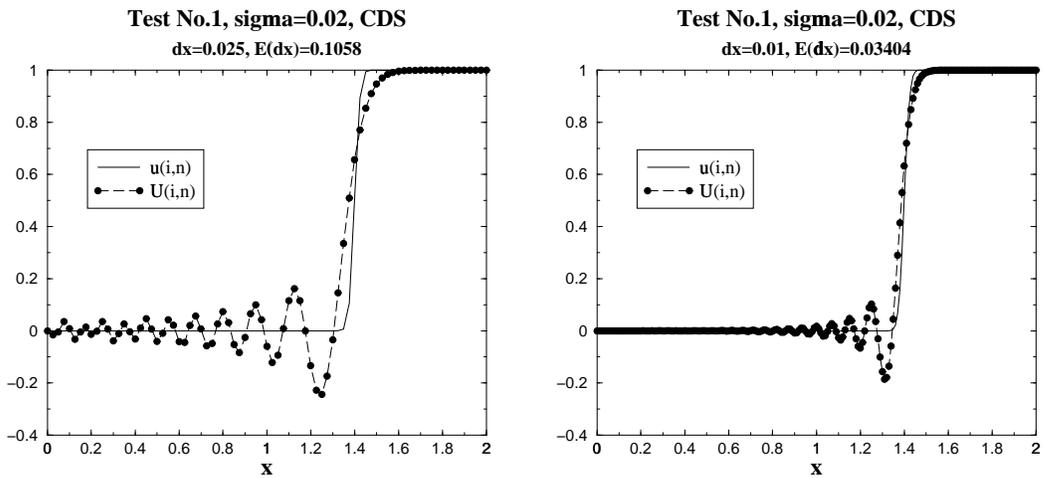


Abbildung 1.21: Test No.1 ( $\sigma = 0.02$ ): Verlauf der Lösung für 2 unterschiedliche Maschenweiten.

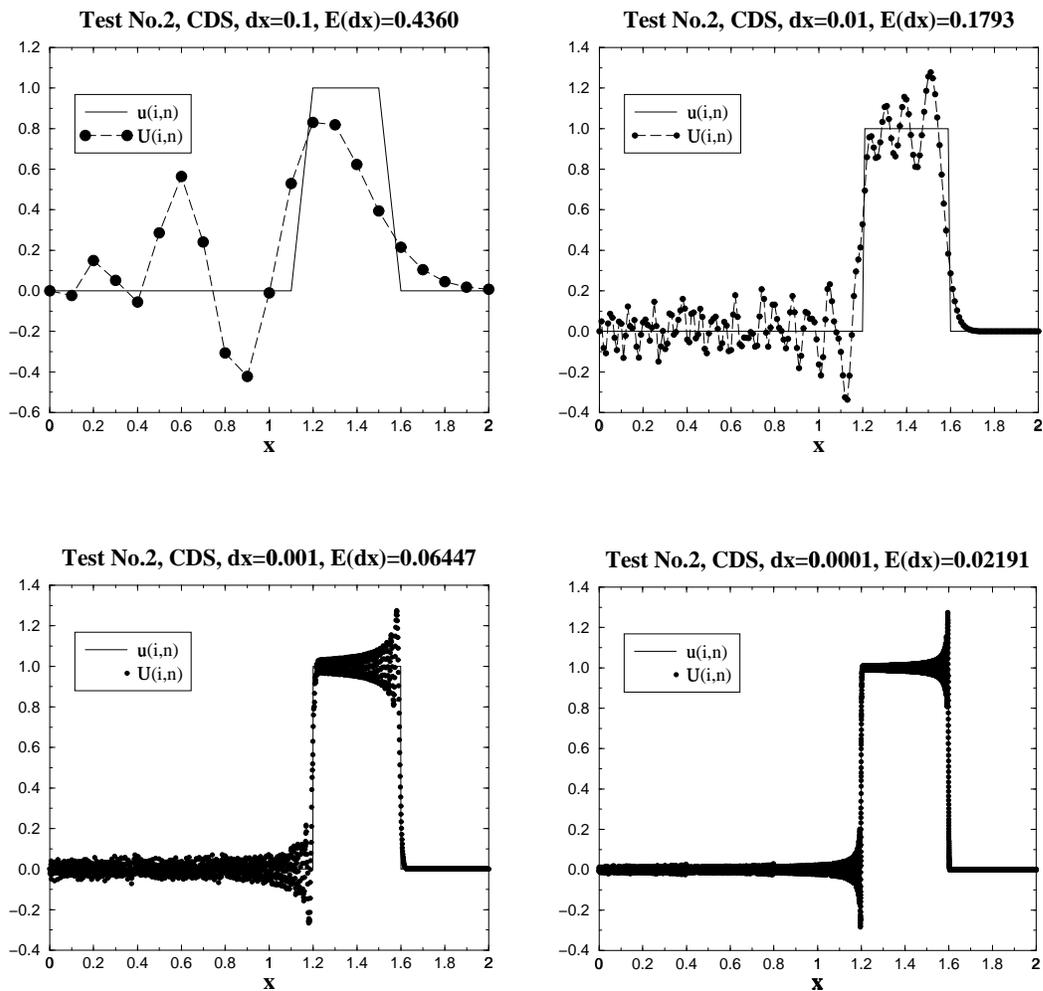


Abbildung 1.22: Test No.2 : Verlauf der Lösung für 4 unterschiedliche Maschenweiten.

nen rein numerischer Natur. Man spricht in diesem Zusammenhang von einer „*numerischen Dispersion*“. Der Grund für das dispersive Verhalten des CDS-Verfahrens wird weiter unten erklärt.

**Testbeispiel No.2.** In der Abb. 1.22 sind die Profile der Lösung für den Testfall No.2 auf 4 unterschiedlichen Gittern dargestellt. Man sieht, daß selbst bei einer Auflösung von 20000 (!) Gitterpunkten ( $\Delta x = 0.0001$ ) starke Oszillationen vorhanden sind. Im großen Teil der Lösungsstrecke nimmt die Amplitude der Schwankungen mit der Gitterverfeinerung jedoch ab, auch wenn das sehr langsam passiert. Obwohl die beiden Diskontinuitäten im Vergleich zu Abb. 1.14 viel besser aufgelöst werden, sorgt der große numerische Fehler im Bereich der Fluktuationen dafür, daß das CDS-Verfahren in diesem Testfall mit einer empirischen Konvergenzordnung von 0.47 sogar schlechter abschneidet als das Upwind-Verfahren. Die hier vorgestellten Lösungsprofile sind insbesondere für den Vergleich mit anderen Diskretisierungsverfahren interessant, die später vorgestellt werden.

**Numerische Dispersion.** Zur Untersuchung des qualitativen Verhaltens des CDS-Verfahrens wenden wir uns wieder der modifizierten Gleichung zu. Aus der Darstellung (1.104) für den örtlichen Abbruchfehler des CDS-Verfahrens sieht man, daß die Differenzenformel (1.102) die ursprüngliche Konvektionsgleichung (1.42) zwar mit zweiter Ordnung approximiert, aber noch besser, und zwar mit vierter Ordnung, approximiert sie die *Konvektions-Dispersions-Gleichung* (auch als Korteweg-de Vries-Gleichung bekannt):

$$u_t + au_x = Bu_{xxx} \quad (1.107)$$

mit

$$B = -\frac{a\Delta x^2}{6}. \quad (1.108)$$

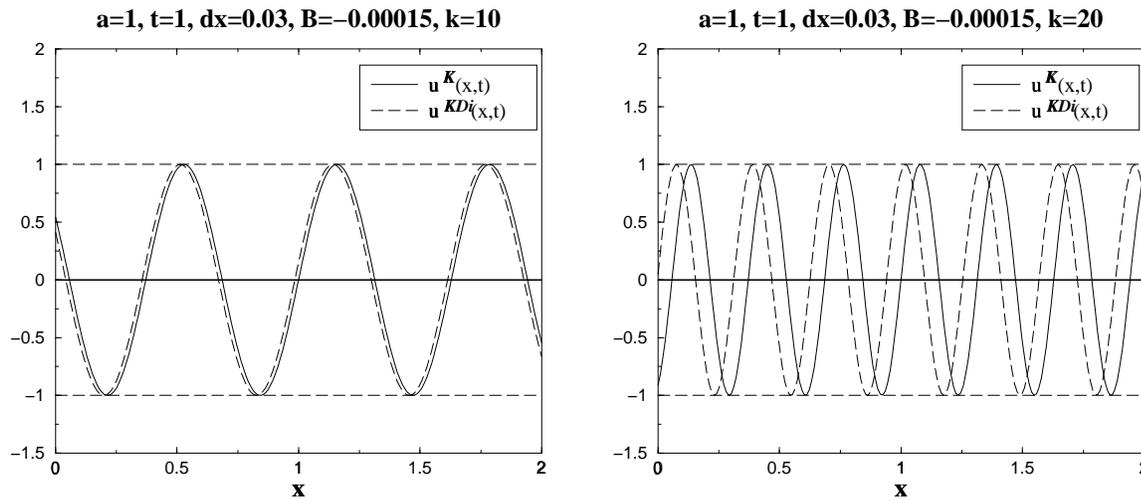
Die Größe  $B$  wird als *numerische Dispersion* bezeichnet. Sie hängt quadratisch von  $\Delta x$  ab und geht mit zunehmender Gitterverfeinerung gegen Null.

Der Einfluß der numerischen Dispersion auf die Lösung der linearen Konvektionsgleichung kann man qualitativ untersuchen, wenn man als Anfangsbedingung für das Cauchy-Problem (1.19, 1.20) wiederum die Funktion

$$u_0(x) = \sin(kx), \quad x \in (-\infty, +\infty). \quad (1.109)$$

wählt. Die Lösung der Konvektionsgleichung lautet in diesem Fall

$$u^K(x, t) = \sin(k(x - at)), \quad (1.110)$$



**Abbildung 1.23:** Vergleich der Funktionen  $u^K(x,t)$  und  $u^{KDi}(x,t)$  für zwei Parametersätze.

während die Lösung der Konvektions-Dispersions-Gleichung (1.107) folgendermaßen aussieht:

$$u^{KDi}(x,t) = \sin(k(x - at - Bk^2t)) = u^K(x - Bk^2t, t). \quad (1.111)$$

Man sieht, daß die Lösung der Konvektions-Dispersions-Gleichung um  $Bk^2t$  der Lösung der Konvektionsgleichung vorausseilt (bzw. – bei negativem  $B$  – hinterherbleibt). Wichtig ist, daß die Stärke der Verschiebung zwischen beiden Lösungen nicht nur von der numerischen Dispersion  $B$ , sondern auch von der Wellenzahl  $k$  abhängt. Für zwei Werte von  $k$  sind die Verläufe der Funktionen  $u^K(x,t)$  und  $u^{KDi}(x,t)$  zum Zeitpunkt  $t = 1$  in Abb. 1.23 exemplarisch dargestellt.

Löst man die lineare Konvektionsgleichung für eine beliebige Funktion  $u_0(x)$  auf einem Gitter mit der Maschenweite  $\Delta x$ , so kann auf diesem Gitter die Anfangsbedingung in eine endliche Fourier-Reihe zerlegt werden. Da die einzelnen Komponenten dieser Zerlegung sich mit unterschiedlichen Geschwindigkeiten  $(a + Bk^2)$  entlang der  $x$ -Achse fortpflanzen, „driften“ sie auseinander und verursachen damit die bereits beobachteten Oszillationen in der numerischen Lösung. Beim CDS-Verfahren ist der numerische Dispersionskoeffizient  $B$  negativ, so daß alle Komponenten sich zu langsam relativ zur exakten Lösung bewegen. Die Oszillationen treten deshalb im Testfall No.1 hinter der Front auf.

**CDS-Verfahren: Zusammenfassung.** Wir wollen nun die gewonnenen Erkenntnisse aus der Untersuchung des CDS-Verfahrens kurz zusammenfassen. Die CDS-Diskretisierung hat die Approximationsordnung von 2 und ist ein absolut stabiles Verfahren. Sie hat damit die Konvergenzordnung 2. Diese Aussage gilt allerdings nur für hinreichend glatte Lösungen.

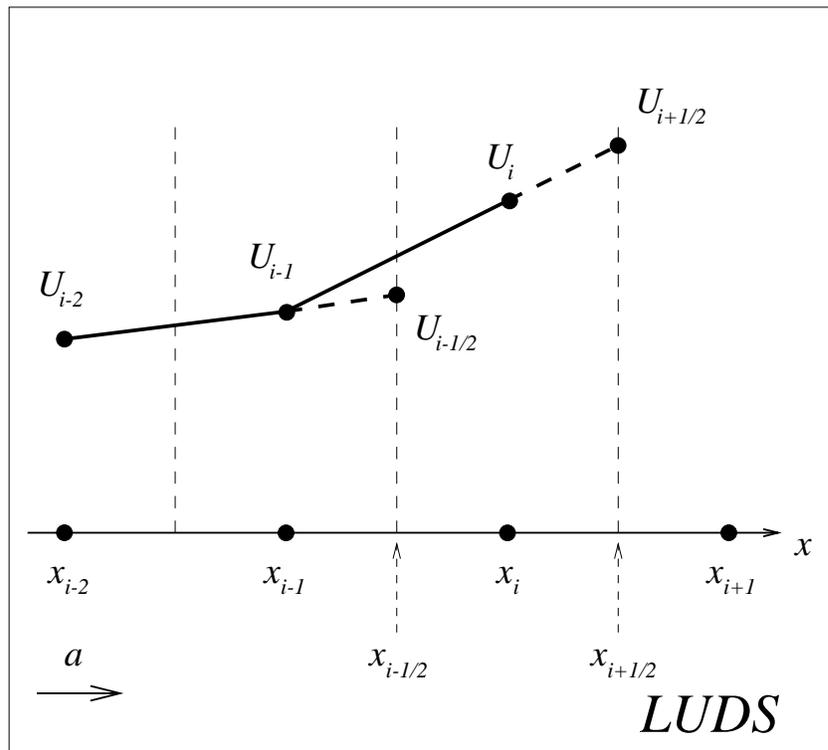


Abbildung 1.24: Lineares Upwind Verfahren (LUDS).

Im Falle einer diskontinuierlichen Lösung der linearen Konvektionsgleichung hat das CDS-Verfahren eine wesentlich niedrigere Konvergenzordnung.

Obwohl das CDS-Verfahren eine im Vergleich zum Upwind-Verfahren höhere Konvergenzrate hat, kann es zum Lösen von konvektionsdominanten Problemen nicht eingesetzt werden. Der Grund dafür ist die numerische Dispersion, die starke unphysikalische Oszillationen in der numerischen Lösung verursacht. Solche Oszillationen führen insbesondere dann zu Schwierigkeiten, wenn die zu transportierende Größe physikalisch keine negativen Werte annehmen darf (wie z.B. Konzentration, Gasgehalt, turbulente kinetische Energie usw.).

Im weiteren Verlauf des Kapitels werden wir daher versuchen, ein Verfahren zu konstruieren, das gleichzeitig eine hohe Konvergenzordnung hat und keine unphysikalischen Lösungen produziert.

## 1.6 Upwind-Verfahren zweiter Ordnung

In vorigen Abschnitten haben wir zwei stabile Diskretisierungsverfahren kennengelernt. Das Upwind-Verfahren benutzt die stückweise konstante Interpolation der Gitterfunktion und ist

daher nur erster Ordnung genau, verwendet bei der Interpolation aber ausschließlich den Wert der Gitterfunktion stromaufwärts von der Approximationsstelle und liefert deshalb physikalisch sinnvolle Lösungen ohne Oszillationen. Das CDS-Verfahren benutzt dagegen die lineare Interpolation der unbekanntem Funktion zwischen den Gitterpunkten, woraus die Approximationsordnung von 2 resultiert, gewichtet dabei beide Werte der Gitterfunktion, die sich links und rechts von der Approximationsstelle befinden, mit einem Faktor 0.5, was in einem dispersiven Charakter des Verfahrens resultiert.

Da wir auf der Suche nach einem Verfahren sind, das die Vorteile dieser beiden Diskretisierungsmethoden verkoppelt, liegt es nahe, eine „konzeptionelle Mischung“ aus beiden zu bilden, indem man die lineare Interpolation mit der Übergewichtung der stromaufwärts liegenden Punkte verbindet. Ein entsprechendes Verfahren wurde von *Price et al.* [96] bereits im Jahre 1966 vorgeschlagen und ist unter dem Namen „Upwind-Verfahren zweiter Ordnung“ bekannt. Wir benutzen für dieses Verfahren die in der englischsprachigen Literatur übliche Abkürzung *LUDS* („linear upwind differencing scheme“).

Beim *LUDS*-Verfahren verwendet man zur Berechnung des Wertes der Funktion  $U(x)$  an der Stelle  $x_{i-1/2}$  eine Gerade, die nicht wie im CDS-Verfahren durch die Punkte  $(x_{i-1}, U_{i-1})$  und  $(x_i, U_i)$  geht, sondern durch die Punkte  $(x_{i-2}, U_{i-2})$  und  $(x_{i-1}, U_{i-1})$  (s. Abb. 1.24). Für den numerischen Konvektionsstrom gilt dann

$$aU_{i-1/2} = F^{luds}(U; i - 1/2) = a \frac{3U_{i-1} - U_{i-2}}{2}. \quad (1.112)$$

Setzt man diesen Konvektionsstrom in die Finite-Volumen-Gleichung (1.53) ein, so ergibt sich eine Diskretisierung der Konvektionsgleichung, die der Approximation der örtlichen Ableitung durch die Rückwärtsdifferenz zweiter Ordnung entspricht:

$$\boxed{\text{LUDS : } \frac{U_i - \bar{U}_i}{\Delta t} + a \frac{3U_i - 4U_{i-1} + U_{i-2}}{2\Delta x} = 0.} \quad (1.113)$$

Für den örtlichen Abbruchfehler des *LUDS*-Verfahrens gilt

$$L(\Delta x) = u_t + au_x - \frac{a\Delta x^2}{3}u_{xxx} + \frac{a\Delta x^3}{4}u_{xxxx} + O(\Delta x^4), \quad (1.114)$$

das Verfahren hat also die Approximationsordnung von 2. Es ist außerdem wie das CDS- und Upwind-Verfahren absolut stabil. Daraus folgt, daß es konvergent ist und auf glatten Lösungen die Konvergenzordnung 2 hat.

Die Behandlung von Randbedingungen erfolgt auf die gleiche Weise wie bei anderen Verfahren. Die resultierende Formel für das letzte Volumenelement lautet:

$$\frac{U_N - \bar{U}_N}{\Delta t} + a \frac{2U_N - 3U_{N-1} + U_{N-2}}{\Delta x} = 0. \quad (1.115)$$

Einer besonderen Behandlung bedarf das Volumenelement mit der Nummer 1. Aus der Differenzenformel (1.113) sieht man, daß für dieses Kontrollvolumen der Wert der Gitterfunktion an der Stelle  $x_{-1}$  benötigt wird, die außerhalb des Berechnungsgebietes liegt. Dieses Problem kann man dadurch umgehen, daß man den Wert  $U_{-1}$  gleich dem Wert der Gitterfunktion am Rand setzt:  $U_{-1} = U_0$ . Diese Vorgehensweise ist nur dann zulässig, wenn die Randbedingung sich zeitlich nicht bzw. sehr wenig ändert. Testfälle 1 und 2 genügen dieser Anforderung. Bei *instationären* Randbedingungen dagegen kann diese Approximation die Konvergenzordnung des Verfahrens im gesamten Lösungsgebiet negativ beeinflussen. Diese Problematik wird im Abschnitt 1.11 noch ausführlich besprochen.

Die numerische Behandlung des resultierenden Gleichungssystems ist um einiges aufwendiger als beim CDS-Verfahren, da infolge des Upwindings die Koeffizientenmatrix keine tridiagonale Struktur mehr hat und das Thomas-Algorithmus nicht direkt eingesetzt werden kann. Man kann allerdings das System in eine tridiagonale Form bringen, indem man einige Terme auf die rechte Seite schiebt und als Quellterme quasi-explizit behandelt. „Quasi-explizit“ bedeutet in diesem Zusammenhang, daß man das resultierende Gleichungssystem mehrfach löst und auf der rechten Seite die Werte aus der vorigen Iteration (und nicht aus der alten Zeitebene wie bei zeit-expliziten Verfahren) rekursiv einsetzt.

Die Aufteilung der Gleichung (1.113) in Terme, die implizit, und Terme, die quasi-explizit behandelt werden, kann sehr bequem auf der Ebene der numerischen Konvektionsströme durchgeführt werden. Der LUDS-Konvektionsstrom kann folgendermaßen dargestellt werden:

$$\begin{aligned} F^{luds}(U; i - 1/2) &= aU_{i-1} + a \frac{U_{i-1} - U_{i-2}}{2} \\ &=: F^{up}(U; i - 1/2) + F^{dc}(U; i - 1/2). \end{aligned} \quad (1.116)$$

Anschließend führt man die Initialisierung durch:

$$U^{(0)} = \bar{U}$$

und löst rekursiv das Gleichungssystem

$$\begin{aligned} \frac{U_i^{(l+1)} - \bar{U}_i}{\Delta t} + \frac{F^{up}(U^{(l+1)}; i + 1/2) - F^{up}(U^{(l+1)}; i - 1/2)}{\Delta x} = \\ - \frac{F^{dc}(U^{(l)}; i + 1/2) - F^{dc}(U^{(l)}; i - 1/2)}{\Delta x}, \quad l = 1, 2, \dots \end{aligned} \quad (1.117)$$

solange, bis die Differenz zwischen zwei nachfolgenden Lösungsvektoren  $U^{(l)}$  und  $U^{(l+1)}$  eine vorgegebene Toleranzgrenze unterschreitet.

Die Idee zur Aufteilung eines numerischen Konvektionsstroms höherer Ordnung in einen Konvektionsstrom erster Ordnung  $F^{up}(U; i - 1/2)$  und einen Korrekturstrom  $F^{dc}(U; i - 1/2)$

wurde von *Khosla und Rubin* [59] vorgeschlagen. Der Korrekturstrom wird in der Literatur als „deferred correction term“ bezeichnet, daher das hochgestellte „ $dc$ “ in  $F^{dc}(U; i - 1/2)$ .

Die quasi-explizite Behandlung des Korrekturstroms bringt für das Lösen des Gleichungssystems gleich drei Vorteile. Erstens, wird dadurch eine einfache Struktur der Koeffizientenmatrix erreicht. Bei eindimensionalen Gleichungen erhält man eine tridiagonale Koeffizientenmatrix, was den Einsatz des direkten Thomas-Algorithmus ermöglicht. Im 2-dimensionalen Fall entsteht eine Fünf-Diagonalmatrix. Das resultierende Gleichungssystem kann sehr effektiv mit dem SIP-Algorithmus („strongly implicit procedure“) von *Stone* [118] (s. auch [30]) iterativ gelöst werden. Im 3-dimensionalen Fall hat die Matrix eine sieben-diagonale Struktur, eine (vektorierte) Verallgemeinerung des SIP-Algorithmus für diesen Fall wurde von *Leister und Peric* [75] vorgeschlagen.

Der zweite Vorteil besteht darin, daß die resultierende Matrix mit der Koeffizientenmatrix des Upwind-Verfahrens erster Ordnung übereinstimmt. Sie enthält positive Koeffizienten auf der Hauptdiagonale, alle anderen Koeffizienten sind dagegen negativ. Diese Eigenschaft der Koeffizientenmatrix sorgt für das gute Konvergenzverhalten des Stone-Algorithmus. Man beachte, daß beim CDS-Verfahren (1.102) dagegen nicht alle Koeffizienten außerhalb der Hauptdiagonale negativ sind. Im eindimensionalen Fall spielt das keine Rolle, da der Thomas-Algorithmus ein direktes Lösungsverfahren ist. Im 2- und 3-D Fall führt diese Eigenschaft zu Konvergenzproblemen beim Stone-Algorithmus, so daß auch hier eine quasi-explizite Behandlung des Korrekturstroms sinnvoll ist, obwohl das Originalgleichungssystem bereits eine fünf- (bzw. sieben-) diagonale Koeffizientenmatrix besitzt.

Im weiteren Verlauf dieses Kapitels werden wir Diskretisierungsverfahren kennenlernen, die *nichtlineare* numerische Konvektionsströme verwenden. In diesem Fall kann die quasi-explizite Behandlung des Korrekturstroms gleichzeitig als eine Art Linearisierung des Gleichungssystems betrachtet werden, was ein weiterer Vorteil dieser Vorgehensweise ist.

Wir wenden uns nun der Untersuchung unserer Testbeispiele zu.

**Testbeispiel No.1.** In den Abbildungen 1.25 und 1.26 sind die Profile der Lösung auf 2 unterschiedlichen Gittern für die Fälle  $\sigma = 0.1$  und  $\sigma = 0.02$  dargestellt. Man sieht, daß die numerische Lösung auch beim LUDS-Verfahren an unphysikalischen Oszillationen leidet. Wie beim CDS-Verfahren kann man dieses Verhalten mit Hilfe der modifizierten Differenzialgleichung erklären. Aus der Darstellung (1.114) für den örtlichen Abbruchfehler des LUDS-Verfahrens sieht man, daß die modifizierte Differenzialgleichung wiederum eine Konvektions-Dispersions-Gleichung ist. Es gibt allerdings einige Unterschiede zum Dispersionsverhalten des CDS-Verfahrens. Erstens, ist der numerische Dispersionskoeffizient  $B$

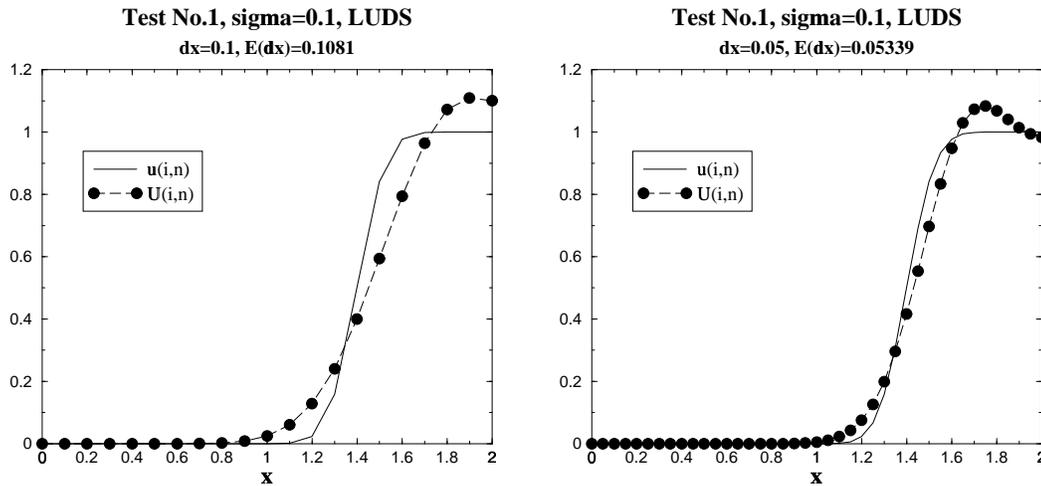


Abbildung 1.25: Test No.1 ( $\sigma = 0.1$ ): Verlauf der Lösung für 2 unterschiedliche Maschenweiten.

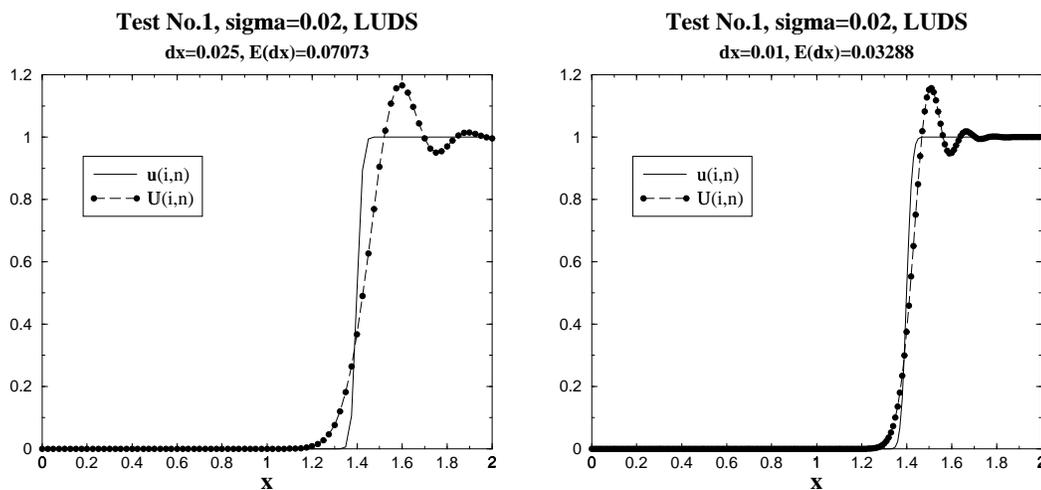
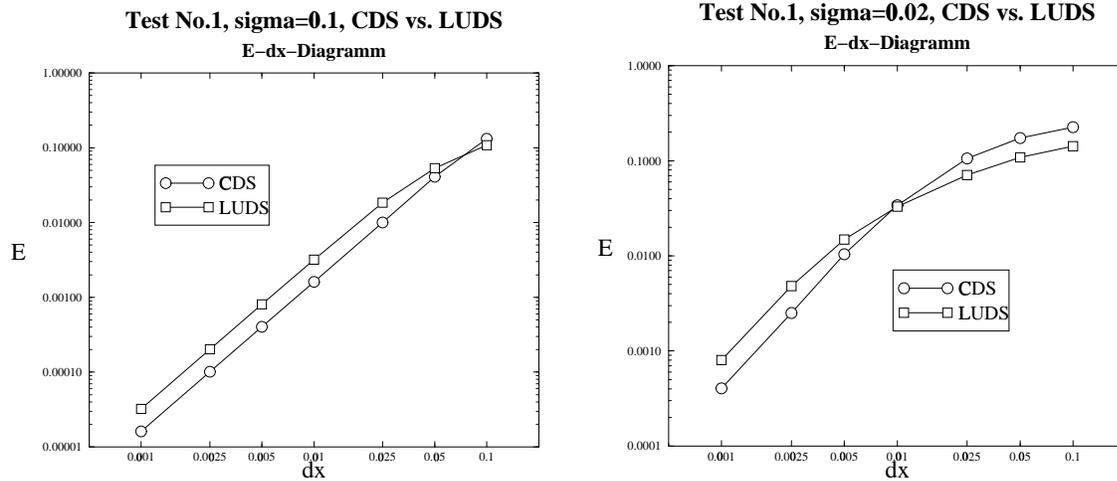


Abbildung 1.26: Test No.1 ( $\sigma = 0.02$ ): Verlauf der Lösung für 2 unterschiedliche Maschenweiten.

diesmal positiv und beträgt

$$B = \frac{a \Delta x^2}{3}. \quad (1.118)$$

Die Oszillationen treten deshalb (im Gegensatz zum CDS-Verfahren) relativ zu der Front stromabwärts auf. Zweitens ist beim LUDS-Verfahren der *absolute* Betrag des numerischen Dispersionskoeffizienten doppelt so hoch wie beim CDS-Verfahren. Obwohl beide Verfahren dieselbe Konvergenzordnung von 2 haben, sollte der numerische Fehler  $E(\Delta x)$  beim CDS-Verfahren deshalb kleiner sein als beim LUDS-Verfahren, zumindest bei hinreichend kleinen Werten von  $\Delta x$ . In Abb. 1.27 (links) sind die numerischen Fehler beider Verfahren für den Fall  $\sigma = 0.1$  in einem  $E$ - $\Delta x$ -Diagramm gegenübergestellt. Man sieht, daß



**Abbildung 1.27:** Test No.1: Vergleich von CDS- und LUDS-Verfahren für  $\sigma = 0.1$  (links) und  $\sigma = 0.02$ ) (rechts).

das CDS-Verfahren tatsächlich wesentlich besser abschneidet, als das LUDS-Verfahren. Im Fall  $\sigma = 0.02$  ist der numerische Fehler beim CDS-Verfahren erst ab einer Auflösung von  $\Delta x = 0.005$  wesentlich kleiner als beim LUDS-Verfahren (Abb. 1.27, rechts). Auf Gittern mit einer Maschenweite oberhalb von  $\Delta x = 0.01$  liefert das Upwind-Verfahren zweiter Ordnung sogar wesentlich genauere Lösungen als die Zentral-Differenz. Dafür gibt es auch eine logische Erklärung. Der Abbruchfehler (1.114) des LUDS-Verfahrens enthält neben dem dispersiven Term noch einen diffusiven Term vierter Ordnung, nämlich  $\frac{a\Delta x^3}{4}u_{xxxx}$ , während beim Abbruchfehler des CDS-Verfahrens (1.104) ein solcher Term nicht vorhanden ist. Auf hinreichend feinen Gittern ist dieser Term im Vergleich zum dispersiven Term vernachlässigbar klein, und die Genauigkeit des Verfahrens wird durch die Höhe des numerischen Dispersionskoeffizienten bestimmt. Da dieser beim CDS-Verfahren kleiner als beim LUDS-Verfahren ist, liefert das CDS-Verfahren auch genauere Lösungen. Auf gröberen Gittern dagegen, sorgt dieser Term dafür, daß die Oszillationen in der numerischen Lösung (die auf feinen Gittern ohnehin verschwinden) stark gedämpft werden. Man sieht das z.B. sehr deutlich, wenn man die Abbildungen 1.21 und 1.26 miteinander vergleicht. In den CDS-Lösungen sind die Oszillationen viel stärker ausgeprägt, als im LUDS-Fall. Deshalb ist auch der numerische Fehler beim LUDS-Verfahren auf groben Gittern kleiner. Ab welcher Auflösung der diffusive Term keine Rolle mehr spielt, hängt von der Glattheit der Lösung ab. Im Falle  $\sigma = 0.02$  passiert das bei einem wesentlich feineren Gitter, als im Falle  $\sigma = 0.1$ . Das zeigt, daß ein besseres Konvergenzverhalten eines Verfahrens bei  $\Delta x \rightarrow 0$  noch nichts darüber aussagt, wie gut dieses Verfahren auf *einem konkreten Gitter* im Vergleich zu anderen Verfahren ist.

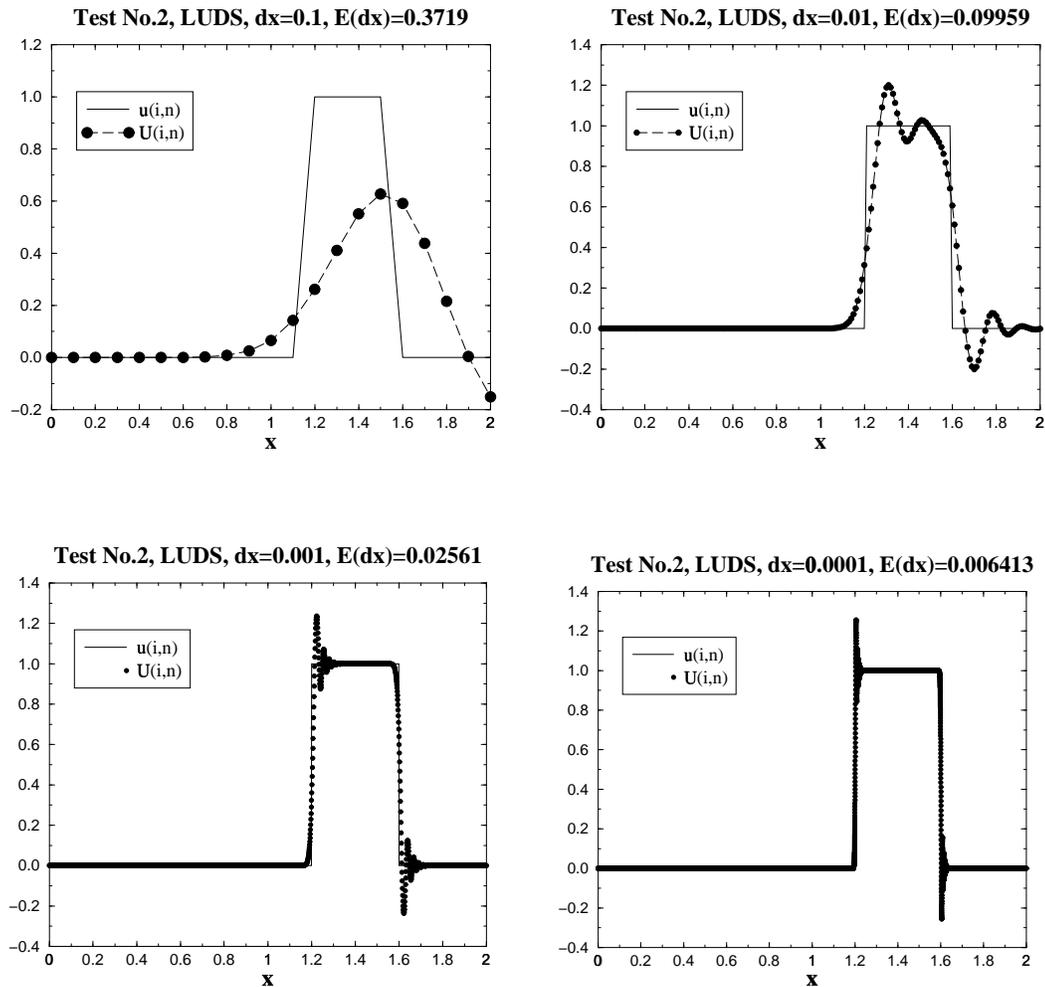
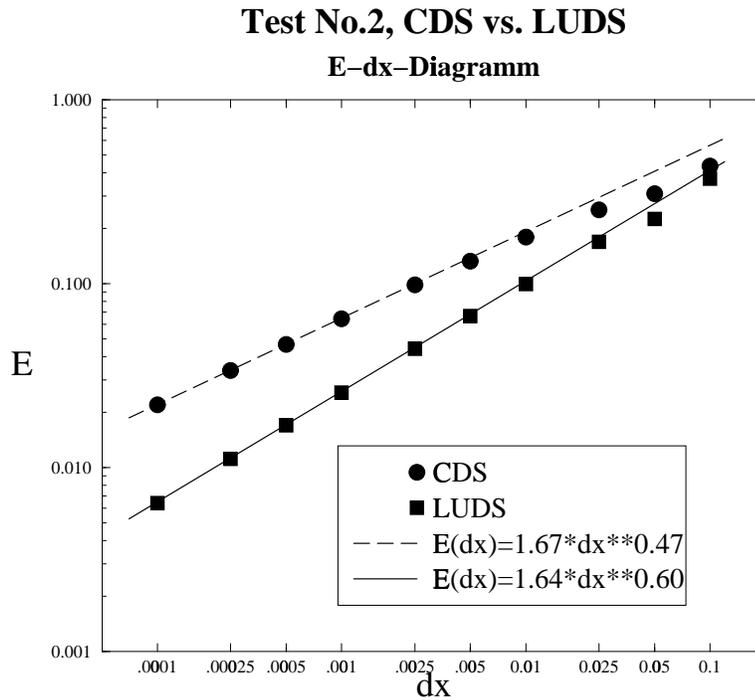


Abbildung 1.28: Test No.2 : Verlauf der Lösung für 4 unterschiedliche Maschenweiten.

**Testbeispiel No.2.** Die Aussage über den sinkenden Einfluß des diffusiven Termes auf die Genauigkeit des LUDS-Verfahrens mit zunehmender Gitterverfeinerung ist nur für hinreichend oft differenzierbare Lösungsfunktionen gültig. Im Falle einer diskontinuierlichen Lösung nimmt die Frequenz der Oszillationen mit der höheren Auflösung immer weiter zu, und der dämpfende Einfluß des diffusiven Termes sorgt dafür, daß die LUDS-Lösung unabhängig von der Gitterfeinheit immer genauer ist, als die CDS-Lösung.

In der Abb. 1.28 sind die Profile der Lösung für den Testfall No.2 auf 4 unterschiedlichen Gittern dargestellt. Auf allen Gittern sind die Oszillationen viel weniger ausgeprägt, als im CDS-Fall (vgl. Abb. 1.22). Nicht nur der absolute Wert des numerischen Fehlers ist beim LUDS-Verfahren auf allen Gittern kleiner als im CDS-Fall (Abb. 1.29). Man sieht auch deutlich, daß die  $E-\Delta x$ -Kurve für das LUDS-Verfahren wesentlich steiler verläuft. Die empirisch ermittelte Konvergenzordnung beträgt 0.60 und ist tatsächlich höher als beim



**Abbildung 1.29:** Test No.2: Vergleich von CDS- und LUDS-Verfahren.

CDS-Verfahren.

**LUDS-Verfahren: Zusammenfassung.** Das Upwind-Verfahren 2. Ordnung verwendet eine lineare Approximation des numerischen Konvektionsstromes, wobei zwei Werte der Gitterfunktion verwendet werden, die stromaufwärts von der Approximationsstelle liegen. Das Verfahren hat einen Abbruchfehler 2. Ordnung und ist absolut stabil. Es hat damit die Konvergenzordnung 2 auf glatten Lösungen.

Wie beim CDS-Verfahren leiden die numerischen Lösungen beim Einsatz des LUDS-Verfahrens an unphysikalischem Verhalten. Auf groben Gittern fallen die Oszillationen allerdings weniger stark aus, als im CDS-Fall. Das führt auf groben Gittern zu einer höheren Genauigkeit der numerischen Lösung als beim Einsatz des CDS-Verfahrens. Auf hinreichend feinen Gittern ist dagegen die CDS-Lösung genauer, weil dann der Abbruchfehler des CDS-Verfahrens um Faktor 2 kleiner ist, als im LUDS-Fall.

Diese Aussage gilt allerdings nur für hinreichend glatte Lösungen. Im Falle einer diskontinuierlichen Lösung hat das LUDS-Verfahren auf allen Gittern eine bessere Genauigkeit und eine höhere empirische Konvergenzordnung als die Zentralkdifferenz.

Wegen der auftretenden Oszillationen im Lösungsprofil kann das LUDS-Verfahren zum

Lösen von konvektionsdominanten Problemen ebenfalls nicht eingesetzt werden, wenn die zu transportierende Größe physikalisch keine negativen Werte annehmen darf.

## 1.7 Agarwal-Verfahren und QUICK-Verfahren

**Agarwal-Verfahren.** Das im letzten Abschnitt untersuchte Upwind-Verfahren 2.Ordnung stellt nur eine der vielen Möglichkeiten dar, bei der Approximation des numerischen Konvektionsstroms die stromaufwärts von der Approximationsstelle liegenden Punkte überzugewichten. Man könnte daher annehmen, daß andere Verfahren mit der Upwinding-Eigenschaft existieren, die im Gegensatz zu LUDS keine Oszillationen in der numerischen Lösung verursachen. Wir wollen nun diese Annahme genauer untersuchen.

Ein linearer numerischer Konvektionsstrom, der zur Approximation der Funktion  $U(x)$  an der Stelle  $x_{i-1/2}$  die Werte  $U_{i-2}$ ,  $U_{i-1}$  und  $U_i$  verwendet, sieht in der *allgemeinen Form* folgendermaßen aus:

$$aU_{i-1/2} = F^{allg}(U; i - 1/2) = a \cdot (\alpha U_i + \beta U_{i-1} + \gamma U_{i-2}). \quad (1.119)$$

Setzt man diesen Konvektionsstrom in die Finite-Volumen-Gleichung (1.53) ein, so ergibt sich folgende Diskretisierung der Konvektionsgleichung:

$$\boxed{\text{ALLG : } \frac{U_i - \bar{U}_i}{\Delta t} + a \frac{\alpha U_{i+1} + (\beta - \alpha)U_i + (\gamma - \beta)U_{i-1} - \gamma U_{i-2}}{\Delta x} = 0.} \quad (1.120)$$

Für den örtlichen Abbruchfehler dieses Verfahrens gilt

$$\begin{aligned} L(\Delta x) &= u_t + au_x(\alpha + \beta + \gamma) + \\ &\quad \frac{a\Delta x}{2}u_{xx}(\alpha - \beta - 3\gamma) + \frac{a\Delta x^2}{6}u_{xxx}(\alpha + \beta + 7\gamma) + \\ &\quad \frac{a\Delta x^3}{24}u_{xxxx}(\alpha - \beta - 15\gamma) + O(\Delta x^4). \end{aligned} \quad (1.121)$$

Damit das Verfahren die Approximationsordnung von 2 hat, müssen folgende Bedingungen erfüllt werden:

$$\alpha + \beta + \gamma = 1 \quad (1.122)$$

$$\alpha - \beta - 3\gamma = 0 \quad (1.123)$$

Das Gleichungssystem (1.122, 1.123) enthält 2 Gleichungen und drei Unbekannte, wir wählen daher  $\alpha$  als freie Variable und erhalten die allgemeine Lösung des Gleichungssystems in der Form:

$$\beta = \frac{3}{2} - 2\alpha \quad (1.124)$$

$$\gamma = \alpha - \frac{1}{2} \quad (1.125)$$

Definiert man nun eine neue Variable  $\lambda = 2\alpha$ , so kann die allgemeine Lösung auch folgendermaßen geschrieben werden:

$$\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} \alpha \\ \frac{3}{2} - 2\alpha \\ \alpha - \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{\lambda}{2} \\ \frac{3}{2} - \lambda \\ \frac{\lambda}{2} - \frac{1}{2} \end{pmatrix} = \lambda \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{pmatrix} + (1 - \lambda) \begin{pmatrix} 0 \\ \frac{3}{2} \\ -\frac{1}{2} \end{pmatrix} \quad (1.126)$$

Die zwei Vektorspalten auf der rechten Seite dieser Darstellung entsprechen genau den Koeffizienten der numerischen Konvektionsströme für das CDS- und das LUDS-Verfahren. Wir haben somit festgestellt, daß jeder lineare Konvektionsstrom der Form (1.119), der zu einem konsistenten Diskretisierungsverfahren mit Approximationsordnung nicht kleiner als 2 führt, notwendigerweise ein gewichtetes Mittel aus dem CDS- und dem LUDS-Konvektionsstrom ist:

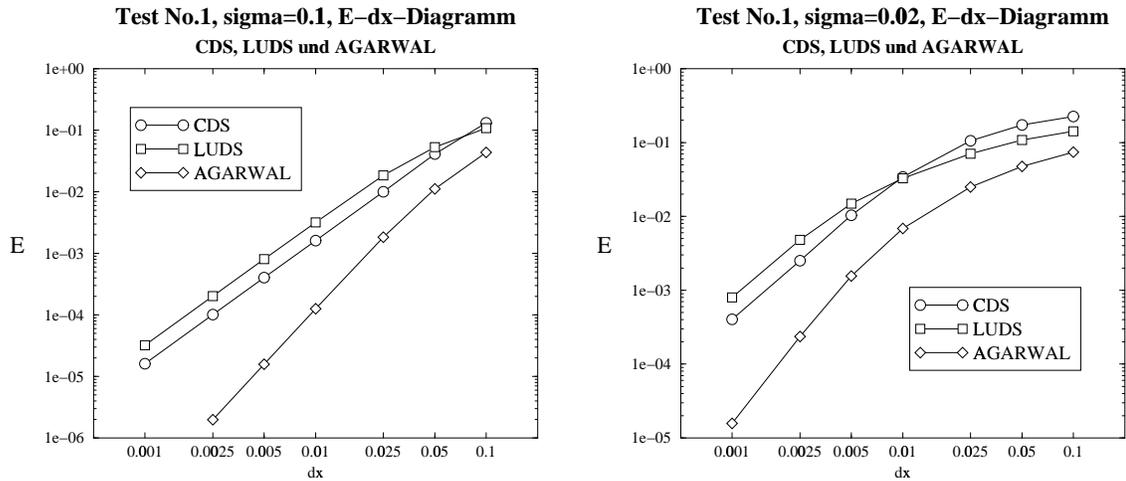
$$\boxed{F^{allg}(U; i - 1/2) = \lambda F^{cds}(U; i - 1/2) + (1 - \lambda) F^{luds}(U; i - 1/2)}. \quad (1.127)$$

Da sowohl die CDS- als auch die LUDS-Diskretisierung an numerischer Dispersion leiden, kann man erwarten, daß auch der numerische Konvektionsstrom (1.127) für alle  $\lambda$ -Werte zu einem Diskretisierungsverfahren führt, das ebenfalls unphysikalische Oszillationen in der numerischen Lösung verursacht. Es sei denn, der freie Parameter  $\lambda$  wird so gewählt, daß der Dispersionsterm in der Darstellung des lokalen Abbruchfehlers verschwindet.

Die dem Konvektionsstrom (1.127) entsprechende Diskretisierung sieht nun folgendermaßen aus:

$$\boxed{\frac{U_i - \bar{U}_i}{\Delta t} + a \frac{\frac{\lambda}{2} U_{i+1} + \left(\frac{3}{2} - \frac{3\lambda}{2}\right) U_i + \left(\frac{3\lambda}{2} - 2\right) U_{i-1} + \left(\frac{1}{2} - \frac{\lambda}{2}\right) U_{i-2}}{\Delta x} = 0}. \quad (1.128)$$

und für den örtlichen Abbruchfehler (1.121) gilt:



**Abbildung 1.30:** Test No.1: Vergleich von 3 Verfahren für  $\sigma = 0.1$  (links) und  $\sigma = 0.02$  (rechts).

$$L(\Delta x) = \frac{a\Delta x^2}{6}u_{xxx}(3\lambda - 2) + \frac{a\Delta x^3}{24}u_{xxxx}(6 - 6\lambda) + O(\Delta x^4). \quad (1.129)$$

Man sieht, daß der dispersive Term bei einem Wert von  $\lambda = \frac{2}{3}$  verschwindet. Das resultierende Diskretisierungsverfahren hat in diesem Fall sogar eine Approximationsordnung von 3. Dieses Upwind-Verfahren 3. Ordnung wurde im Jahre 1981 von *Agarwal* [1] vorgeschlagen. Das Agarwal-Verfahren ist absolut stabil und hat für glatte Lösungen Konvergenzordnung von 3.

Vollständigkeitshalber seien hier noch der numerische Konvektionsstrom und die Finite-Volumen-Formel für das Agarwal-Verfahren dargestellt:

$$aU_{i-1/2} = F^{agar}(U; i - 1/2) = a \cdot \left( \frac{1}{3}U_i + \frac{5}{6}U_{i-1} - \frac{1}{6}U_{i-2} \right), \quad (1.130)$$

$$\boxed{\text{AGARWAL : } \frac{U_i - \bar{U}_i}{\Delta t} + a \frac{2U_{i+1} + 3U_i - 6U_{i-1} + U_{i-2}}{6\Delta x} = 0.} \quad (1.131)$$

Wir wenden uns nun der Untersuchung unserer Testfälle für das Agarwal-Verfahren zu.

**Testbeispiel No.1.** In der Abb. 1.30 sind die numerischen Fehler von CDS-, LUDS- und Agarwal-Verfahren für den Testfall No.1 für beide  $\sigma$ -Werte in einem  $E$ - $\Delta x$ -Diagramm gegenübergestellt. Man sieht, daß das Agarwal-Verfahren tatsächlich wesentlich genauer als

die beiden anderen ist, was sowohl die Konvergenzrate als auch den absoluten Betrag des numerischen Fehlers auf allen Gittern betrifft. Unglücklicherweise leiden auch die mit dem Agarwal-Verfahren berechneten Lösungen an Oszillationen (Abb. 1.31, 1.32), auch wenn diese viel weniger ausgeprägt sind, als im CDS- (Abb. 1.20, 1.21) und LUDS-Fall (Abb. 1.25, 1.26). Diese Oszillationen sind von anderer Natur als bei beiden vorigen Verfahren. Man sieht z.B. deutlich, daß sie diesmal sowohl vor der Front *als auch* hinter der Front auftreten. Die Ursachen für diese Fluktuationen können nicht mehr so einfach mit Hilfe der modifizierten Differentialgleichung erklärt werden. Eine andere (indirekte) Begründung wird aber im nächsten Abschnitt gegeben.

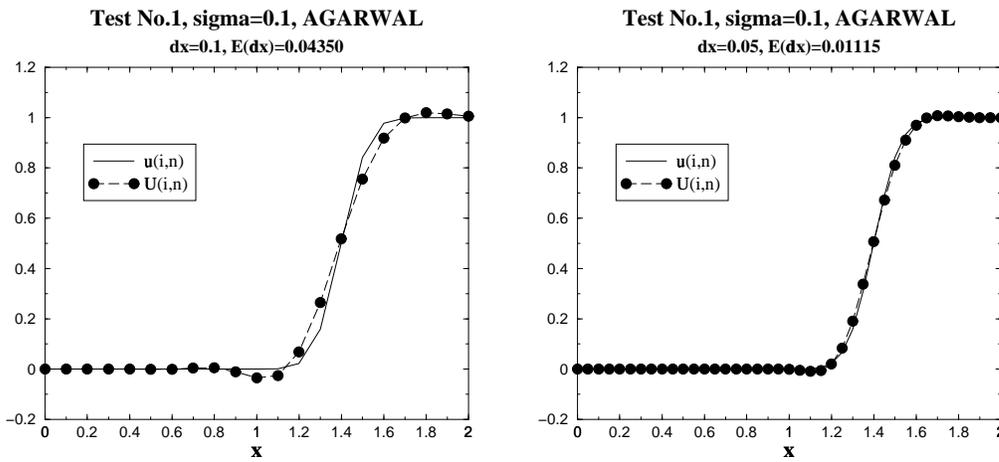


Abbildung 1.31: Test No.1 ( $\sigma = 0.1$ ): Verlauf der Lösung für 2 unterschiedliche Maschenweiten.

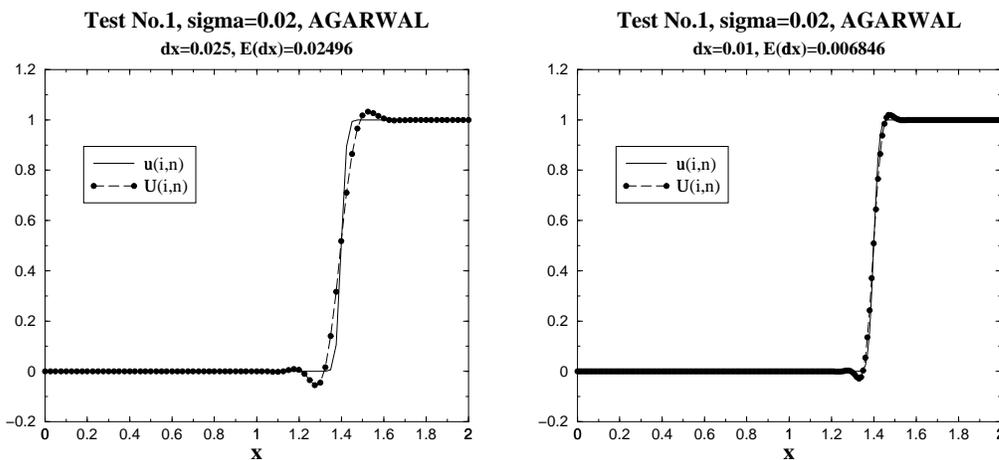


Abbildung 1.32: Test No.1 ( $\sigma = 0.02$ ): Verlauf der Lösung für 2 unterschiedliche Maschenweiten.

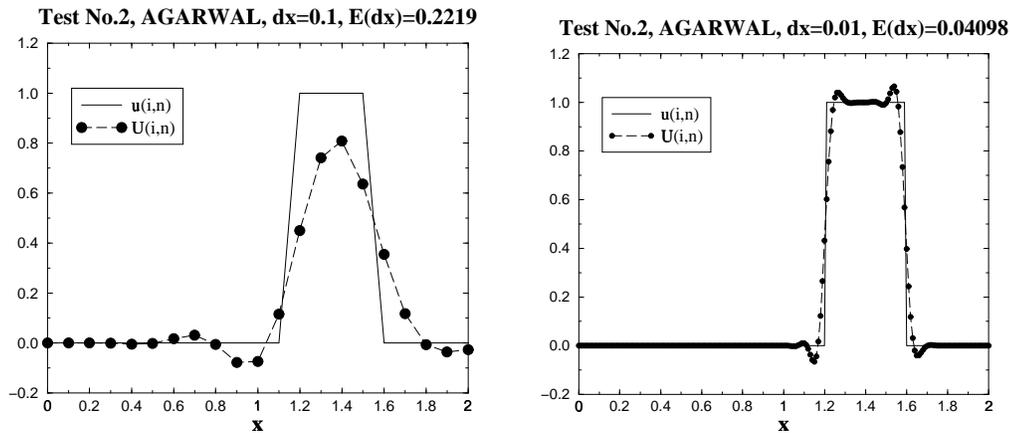
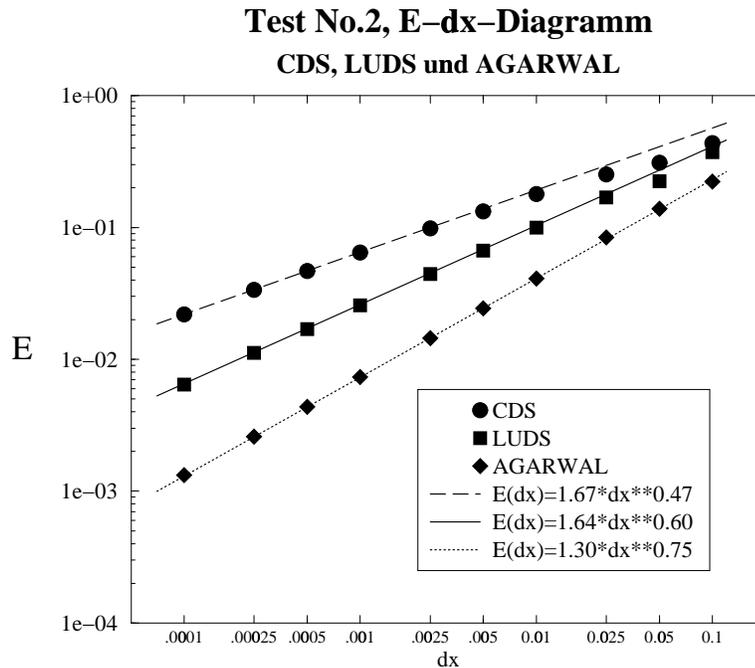


Abbildung 1.33: Test No.2 : Verlauf der Lösung für 2 unterschiedliche Maschenweiten.

**Testbeispiel No.2.** Auch im Testfall No.2 (Abb. 1.33) gehen die Oszillationen im Vergleich zu beiden anderen Verfahren (Abb. 1.22, 1.28) sehr stark zurück. Das sorgt wiederum für die höhere Genauigkeit des Agarwal-Verfahrens auf allen Gittern (Abb. 1.34). Die empirische Konvergenzordnung des Verfahrens beträgt 0.75, was die Konvergenzordnung alle bisher diskutierten Diskretisierungsmethoden übertrifft.

**Agarwal-Verfahren: Zusammenfassung.** Ausgehend von einer allgemeinen Formulierung eines linearen Upwind-Konvektionsstroms wurde gezeigt, daß jedes darauf basierende Verfahren, das mindestens 2. Ordnung genau ist, als ein gewichtetes Mittel aus Zentral-Differenz und Upwind-Verfahren 2. Ordnung dargestellt werden kann. Da diese beiden Verfahren an numerischer Dispersion leiden, gilt dasselbe auch für dessen Mittel, es sei denn, der Gewichtungsfaktor ist so gewählt, daß der Dispersionskoeffizient des resultierenden Verfahrens zu Null wird. Bei diesem speziellen Wert des Gewichtungsfaktors erhält man das Agarwal-Verfahren, das 3. Ordnung genau ist. Das Verfahren ist absolut stabil, und hat in allen untesteten Testfällen und auf allen numerischen Gittern eine höhere Genauigkeit als CDS- und LUDS-Verfahren erzielt. Im Falle einer diskontinuierlichen Lösung wurde dabei eine empirische Konvergenzordnung von 0.75 festgestellt. Unglücklicherweise leiden auch die mit dem Agarwal-Verfahren berechneten Lösungen an Oszillationen, so daß auch dieses Verfahren zum Lösen von konvektionsdominanten Problemen nicht eingesetzt werden kann, wenn die zu transportierende Größe physikalisch keine negativen Werte annehmen darf.

**QUICK-Verfahren.** Das am weitesten verbreitete Upwind-Verfahren höherer Ordnung ist das von *Leonard* [76] entwickelte „**Quadratic Upstream Interpolation for Convective Kinematics**“-Verfahren (*QUICK*). Beim QUICK-Verfahren wird zur Berechnung des Wertes der



**Abbildung 1.34:** Test No.2: Vergleich von CDS-, LUDS- und Agarwal-Verfahren.

Funktion  $U(x)$  an der Stelle  $x_{i-1/2}$  eine *Parabel* durch die Punkte  $(x_{i-2}, U_{i-2})$ ,  $(x_{i-1}, U_{i-1})$  und  $(x_i, U_i)$  gelegt. Der Wert, den diese Parabel an der Stelle  $x_{i-1/2}$  annimmt, wird dann als Approximationswert für  $U(x_{i-1/2})$  benutzt (s. Abb. 1.35). Für den numerischen Konvektionsstrom gilt dann

$$aU_{i-1/2} = F^{quick}(U; i - 1/2) = a \cdot \left( \frac{3}{8}U_i + \frac{3}{4}U_{i-1} - \frac{1}{8}U_{i-2} \right). \quad (1.132)$$

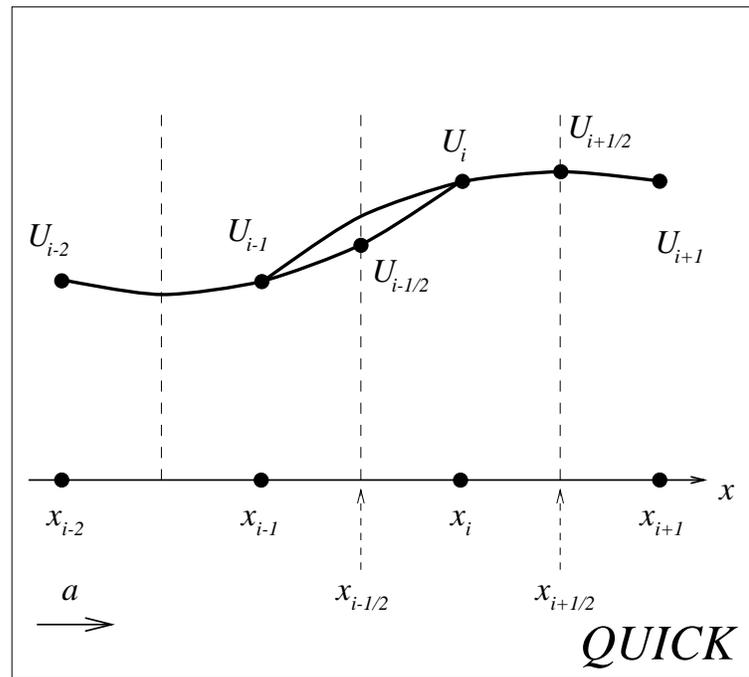
Setzt man diesen Konvektionsstrom in die Finite-Volumen-Gleichung (1.53) ein, so ergibt sich folgende Diskretisierung der Konvektionsgleichung:

**QUICK** :  $\frac{U_i - \bar{U}_i}{\Delta t} + a \frac{3U_{i+1} + 3U_i - 7U_{i-1} + U_{i-2}}{8\Delta x} = 0.$

(1.133)

Man sieht, daß diese Differenzenformel auch aus der allgemeinen Darstellung (1.128) abgeleitet werden kann, wenn man für den Gewichtungsfaktor  $\lambda$  den Wert  $\frac{3}{4}$  einsetzt. Der Abbruchfehler errechnet sich dann aus (1.129) zu

$$L(\Delta x) = \frac{a\Delta x^2}{24}u_{xxx} + \frac{a\Delta x^3}{16}u_{xxxx} + O(\Delta x^4). \quad (1.134)$$

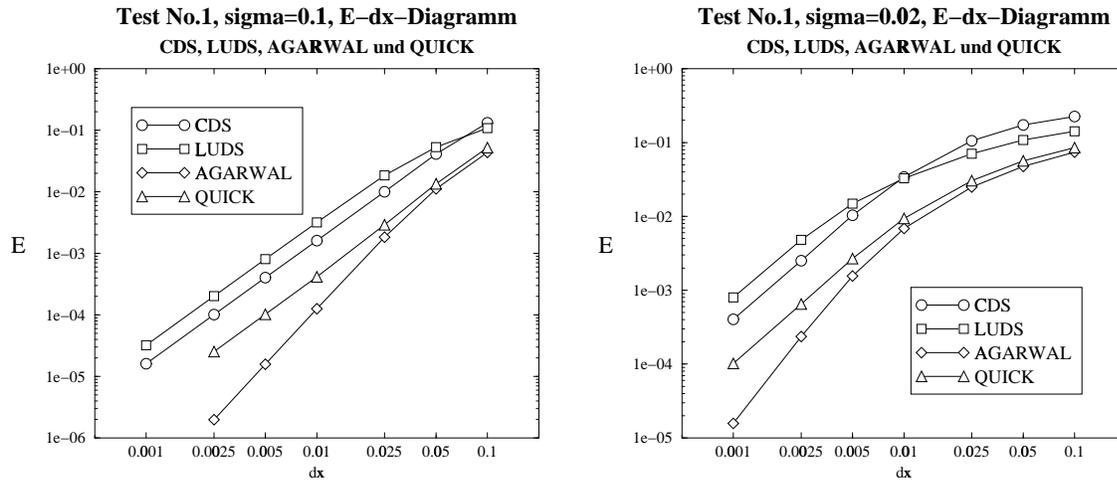


**Abbildung 1.35:** Quadratische Interpolation der Lösung zwischen den Gitterpunkten.

Man sieht, daß das QUICK-Verfahren lediglich 2. Ordnung genau ist, entgegen der weit verbreiteten Überzeugung, daß die Approximationsordnung dieses Verfahrens 3 ist. Der numerische Dispersionskoeffizient des QUICK-Verfahrens ist allerdings wesentlich kleiner als im CDS- bzw. LUDS-Verfahren (um Faktor 6 bzw. 8). Außerdem enthält der Abbruchfehler einen Diffusionsterm vierter Ordnung, der wie beim LUDS-Verfahren für die Dämpfung der Oszillationen auf den groben Gittern sorgt. Das führt dann auf allen Gittern zu einer wesentlich höheren Genauigkeit des QUICK-Verfahrens im Vergleich zu beiden anderen, wie man aus der Abb. 1.36 sieht. Auf groben Gittern ist die Genauigkeit des QUICK-Verfahrens sogar vergleichbar mit der Genauigkeit des Agarwal-Verfahrens, was ein Grund dafür sein kann, warum man allgemein von der 3. Ordnung des QUICK-Verfahrens überzeugt ist. Gleichzeitig sieht man deutlich, daß bei kleineren  $\Delta x$ -Werten die  $E$ - $\Delta x$ -Kurven von CDS, LUDS und QUICK parallel verlaufen, was ein Beweis dafür ist, daß alle drei Verfahren die gleiche Konvergenzordnung haben, und zwar die Ordnung 2.

Auch im Testfall No.2 ist die Genauigkeit des QUICK-Verfahrens wesentlich höher als von CDS und LUDS (Abb. 1.37), obwohl die empirisch ermittelte Konvergenzordnung von 0.6 mit der des LUDS-Verfahrens übereinstimmt. Auf groben Gittern ist die Genauigkeit wiederum vergleichbar mit der des Agarwal-Verfahrens.

Auf die ausführliche Darstellung der Lösungsprofile für unsere Testfälle werden wir diesmal verzichten, da sich daraus keine neuen Erkenntnisse gewinnen lassen. Im beiden Testfällen



**Abbildung 1.36:** Test No.1: Vergleich von 4 Verfahren für  $\sigma = 0.1$  (links) und  $\sigma = 0.02$  (rechts).

weisen diese Profile Oszillationen auf (wie dies auch bei allen anderen untersuchten Verfahren höherer Ordnung der Fall war), was den Einsatz des QUICK-Verfahrens für konvektionsdominante Probleme stark einschränkt.

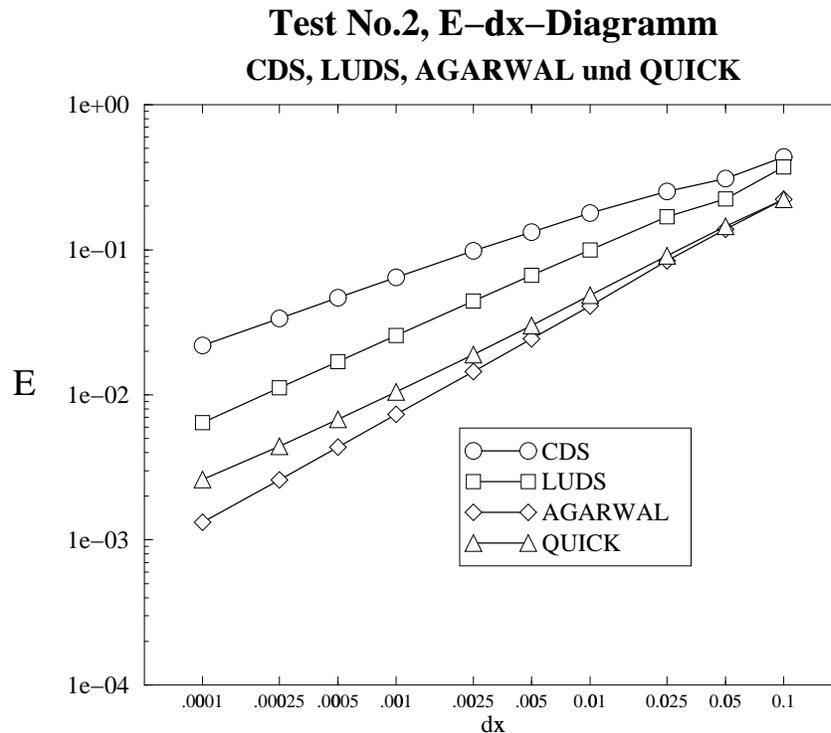
**Der vermeintliche Widerspruch.** Die einzelnen bis jetzt untersuchten Verfahren unterscheiden sich lediglich in der Approximation des Konvektionsstroms  $aU_{i-1/2}$ . Beim Upwind-Verfahren wurde ein stückweise *konstanter* Verlauf der unbekanntes Funktion angenommen, und das resultierende Verfahren hatte die Approximationsordnung von 1. Bei CDS und LUDS wurde eine *lineare* Interpolation verwendet und das entsprechende Diskretisierungsverfahren hatte eine Approximationsordnung von 2. Da im QUICK-Verfahren eine *quadratische* Interpolation verwendet wird, erwartet man auch eine entsprechend höhere Approximationsordnung dieses Verfahrens. Die Tatsache, daß die Approximationsordnung von QUICK-Verfahren trotzdem nicht höher als 2 ist, ist daher sehr überraschend.

Für dieses Phänomen gibt es allerdings eine einfache Erklärung. Wir erinnern uns an die Herleitung der Finite-Volumen-Formulierung (1.53). Man integrierte beide Summanden der ursprünglichen Differentialgleichung (1.42) und erhielt die Integralform

$$I_1 + I_2 = 0,$$

in die für die Ausdrücke  $I_1$  und  $I_2$  Approximationsformeln eingesetzt wurden. Der Term  $I_1$  wurde einheitlich für alle Verfahren approximiert, wobei

$$\frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} u(x, t) dx \approx u(x_i, t) \quad (1.135)$$



**Abbildung 1.37:** Test No.2: Vergleich von 4 Verfahren.

angenommen wurde, was einen Approximationsfehler von  $O(\Delta x^2)$  für  $I_1$  zu Folge hat. Der Approximationsfehler für  $I_2$  hängt dagegen von der Darstellung des numerischen Konvektionsstromes ab. Beim Upwind-Verfahren wird der Ausdruck  $I_2$  mit erster Ordnung approximiert, so daß die Summe beider Ausdrücke die Approximationsordnung von 1 hat. Beim CDS und LUDS werden beide Terme,  $I_1$  und  $I_2$ , mit 2. Ordnung angenähert, die Summe hat somit einen Fehler von Ordnung 2. Beim QUICK-Verfahren wird lediglich der Term  $I_2$  mit 3. Ordnung approximiert, der Term  $I_1$  wird nach wie vor mit 2. Ordnung angenähert, so daß die Summe beider Ausdrücke nur die Approximationsordnung von 2 hat.

Es gibt daher nur 2 Möglichkeiten, ein Verfahren 3. Ordnung zu konstruieren. Die erste Möglichkeit besteht darin, daß das Integral (1.135) ebenfalls mit der 3. Ordnung approximiert wird. Diese Vorgehensweise ist beim Finite-Volumen-Verfahren allerdings nicht üblich. Die andere Möglichkeit besteht darin, einen Approximationsausdruck 2. Ordnung für  $I_2$  so zu wählen, daß die Koeffizienten bei  $\Delta x^2$  in den Approximationsfehlern für  $I_1$  und  $I_2$  sich gegenseitig eliminieren. Genau dieser Anforderung entspricht das Agarwal-Verfahren, das 3. Ordnung genau ist, obwohl der Term  $I_2$  beim Einsatz des Agarwal-Konvektionsstroms nur mit 2. Ordnung approximiert wird.

Aus der Darstellung (1.62) des lokalen Abbruchfehlers für die Finite-Volumen-Formulie-

nung kann man durch den Grenzübergang  $\Delta t \rightarrow 0$  ableiten, daß

$$L(\Delta x) = \frac{F(u; i + 1/2) - F(u; i - 1/2)}{\Delta x} \Big|_{x_i=x} - au_x$$

gilt. Für die Approximationsordnung eines Verfahrens ist daher nicht entscheidend, wie genau der Konvektionsstrom  $au(x_{i-1/2})$  durch  $F(u; i - 1/2)$  approximiert wird, sondern wie genau die Ableitung der Stromfunktion  $-(au)_x$  durch den Ausdruck

$$\frac{F(u; i + 1/2) - F(u; i - 1/2)}{\Delta x} \Big|_{x_i=x}$$

approximiert wird. Diese Erkenntnis wird in den folgenden Kapiteln bei der Herleitung von Verfahren 3. Ordnung für nichtlineare Konvektionsgleichungen nützlich sein.

## 1.8 Monotonie-Erhaltung und TVD-Eigenschaft

Die wichtigsten Anforderungen an ein numerisches Verfahren sind, daß es möglichst hohe Genauigkeit besitzt, numerisch stabil ist und physikalisch sinnvolle Lösungen liefert. Keines von den bis jetzt untersuchten Verfahren genügt allen drei Anforderungen gleichzeitig. Das einzige Verfahren, das oszillationsfreie Lösungen liefert – nämlich das Upwind-Verfahren – ist lediglich 1. Ordnung genau. Alle Verfahren, die eine höhere Ordnung haben, produzierten Lösungen mit unphysikalischen Oszillationen.

Die Vorgehensweise, die wir bis jetzt verwendet haben, um ein Verfahren mit den gewünschten drei Eigenschaften zu bilden, kann man als empirisch bezeichnen. Zunächst wurde ein Verfahren konstruiert, danach wurden seine Eigenschaften untersucht. Waren diese nicht zufriedenstellend, haben wir versucht, die Ursachen dafür zu bestimmen und das Verfahren in die richtige Richtung zu modifizieren. Anschließend wurden die Eigenschaften des neuen Verfahrens getestet usw.

So haben wir beim Upwind-Verfahren festgestellt, daß der einzige Mangel des Verfahrens seine niedrige Konvergenzordnung ist. Wir haben die Approximationsordnung des Verfahrens durch den Einsatz der linearen Interpolation erhöht, und kamen so zum CDS-Verfahren, das eine höhere Konvergenzordnung hatte, aber oszillierende Lösungen produzierte. Um diese zu vermeiden, haben wir die lineare Interpolation mit der Übergewichtung der stromaufwärts liegenden Punkte kombiniert, und erhielten so das LUDS-Verfahren. Das Austesten des LUDS-Verfahrens hat gezeigt, daß dieses ebenfalls an unphysikalischem Verhalten der Lösung leidet. Nachdem wir festgestellt haben, daß der Grund für die Oszillationen die numerische Dispersion ist, haben wir ein Verfahren gefunden, bei dem kein dispersiver Term

in der modifizierten Gleichung enthalten ist (Agarwal-Verfahren). Anschließend mußten wir leider feststellen, daß auch dieses Verfahren numerische Lösungen mit Fluktuationen erzeugt.

Nachdem wir gesehen haben, daß dieser empirische Weg nicht zum Erfolg führt, wollen wir nun versuchen, das Problem vom anderen Ende anzugreifen. Wir wollen die Verfahren auf die Weise konstruieren, daß es von *vornherein* sichergestellt ist, daß sie keine unphysikalischen Lösungen erzeugen *können*. Um dieses Ziel zu erreichen, muß diese Anforderung an das Verfahren zunächst genau formalisiert werden. Dazu müssen wir als erstes die Frage beantworten, was wir unter physikalisch sinnvollen numerischen Lösungen verstehen. Das sind für uns solche numerischen Lösungen, die wichtige Eigenschaften der analytischen Lösung nicht verletzen. Beim Cauchy-Problem für die lineare Konvektionsgleichung bedeutet das, daß

- keine neuen lokalen Extrema in der Lösung auftreten können,
- der Wert des lokalen Minimums nicht kleiner werden kann,
- der Wert des lokalen Maximums nicht wachsen kann.

Wichtig ist, daß die Lösung der allgemeinen hyperbolischen Gleichung  $u_t + f(u)_x = 0$  diese Eigenschaften ebenfalls besitzt.

Diese Bedingungen sind allerdings schwer formalisierbar, da die Stellen der lokalen Extrema sich mit der Zeit entlang der  $x$ -Achse bewegen. Aus diesen 3 Forderungen an die Lösungsfunktion folgt allerdings, daß wenn die Lösung zum Zeitpunkt  $t = 0$  *monoton* (fallend oder wachsend) ist, so muß sie es auch für alle  $t > 0$  bleiben.

Genügt auch ein numerisches Schema dieser Anforderung, so sagt man, daß es *monotonie erhaltend* („monotonicity preserving“) ist. Formal bedeutet das, wenn die Gitterfunktion  $\bar{U} = \{\bar{U}_i\}$  monoton bezüglich  $i$  ist, so muß das auch für die Gitterfunktion  $U = \{U_i\}$  mit  $U = FV(\bar{U})$  gelten (s. 1.55). Es ist klar, daß ein monotonie-erhaltendes Verfahren keine oszillierenden Lösungen bei der Berechnung der linearen Konvektion von wandernden Fronten produzieren kann.

Eine hinreichende Bedingung für die Monotonie-Erhaltung ist die sogenannte *Monotonie-Eigenschaft* („monotonicity“) eines Verfahrens. Ein numerisches Verfahren heißt *monoton*, wenn für zwei Gitterfunktionen  $U$  und  $V$  aus  $U \leq V$  folgt:  $FV(U) \leq FV(V)$ . Es ist bekannt, daß jedes monotone Diskretisierungsverfahren stabil ist. Leider kann es höchstens von erster Ordnung genau sein [45].

Weiterhin gilt, daß jedes *lineare* Finite-Volumen-Verfahren, das monotonie erhaltend ist,

gleichzeitig monoton ist. Daraus folgt, daß jedes monotonie erhaltende lineare Verfahren ebenfalls höchstens von erster Ordnung genau sein kann. **Das erklärt, warum alle bisher untersuchten Verfahren 2. und 3. Ordnung (inkl. Agarwal-Verfahren) oszillierende Lösungen produziert haben.**

Wir kommen somit zu der Schlußfolgerung, daß wir ein monotonie-erhaltendes Verfahren höherer Ordnung nur finden können, wenn wir auf die Klasse der *nichtlinearen* Diskretisierungsverfahren ausweichen. Damit entstehen für uns gleichzeitig zwei neue Probleme. Erstens ist es nicht einfach, die Monotonie-Erhaltung eines nichtlinearen Verfahrens zu überprüfen. Zweitens, ist es für ein nichtlineares Verfahren wesentlich schwieriger, eine hinreichende Bedingung für die Stabilität zu finden, die für ein konsistentes Verfahren dessen Konvergenz sichert.

Glücklicherweise gibt es einen sehr eleganten Weg, diese beiden Probleme zu lösen. Wir haben uns vorher darauf geeinigt, daß ein numerisches Verfahren wichtige Eigenschaften der analytischen Lösung nicht verletzen darf. Eine dieser Eigenschaften betrifft die *totale Variation* der Lösung. Für eine Funktion  $u = u(x)$  ist die totale Variation  $TV(u)$  folgendermaßen definiert:

$$TV(u) = \int \left| \frac{\partial u}{\partial x} \right| dx. \quad (1.136)$$

Eine wichtige Eigenschaft der Lösung der linearen Konvektionsgleichung ist, daß die totale Variation der Lösung mit der Zeit nicht zunimmt (vorausgesetzt  $TV(u_0) < \infty$ , was wir stets annehmen werden):

$$TV(u(\cdot, t_2)) \leq TV(u(\cdot, t_1)), \quad \text{für } t_2 > t_1 \geq 0. \quad (1.137)$$

Für die analytische Lösung der linearen Konvektionsgleichung ist diese Bedingung offensichtlich erfüllt, da sich die Profile der Lösung zu 2 verschiedenen Zeitpunkten nur durch eine Verschiebung entlang der  $x$ -Achse unterscheiden, deshalb muß die totale Variation der beiden Profile gleich groß sein. Viel wichtiger ist es, daß die Bedingung (1.137) *auch für die Lösung der allgemeinen hyperbolischen Gleichung  $u_t + f(u)_x = 0$  erfüllt ist*, wie Lax in 1973 gezeigt hat [74].

Analog zu (1.136) kann die totale Variation einer Gitterfunktion  $U$  folgendermaßen definiert werden:

$$TV(U) = \sum_{i=-\infty}^{i=+\infty} |U_{i+1} - U_i|. \quad (1.138)$$

Wir nennen ein Verfahren *TV-stabil*, wenn  $TV(U^n)$  gleichmäßig beschränkt ist für alle  $\Delta x$ ,  $\Delta t$  und  $n$  mit  $\Delta x < \Delta x_0$ ,  $\Delta t < \Delta t_0$  und  $n\Delta t \leq T$ . Für ein TV-stabiles Verfahren gilt folgendes

**Theorem 1** [77] *Ein konsistentes und TV-stabiles Diskretisierungsverfahren mit einer Lipschitz-stetigen Konvektionsstromfunktion  $F(U; i - 1/2)$  ist konvergent.*

Wir sagen, daß ein Diskretisierungsverfahren die *TVD-Eigenschaft* hat („**T**otal **V**ariation **D**iminishing“), wenn für ein beliebiges  $\bar{U}$  mit  $TV(\bar{U}) < \infty$  gilt:

$$TV(U) \leq TV(\bar{U}), \quad \text{mit } U = FV(\bar{U}). \quad (1.139)$$

Hat ein Verfahren die TVD-Eigenschaft, so werden wir dieses als *TVD-Verfahren* bezeichnen. Das Konzept eines TVD-Verfahrens wurde von *Harten* [40] im Jahre 1983 eingeführt. In dieser Arbeit wurden diese Verfahren noch als „**T**V**N**I-Verfahren“ bezeichnet („**T**otal **V**ariation **N**on**I**ncreasing“). Ab 1984 [41] setzt sich die Abkürzung „**TVD**“ durch.

**Theorem 2** *Ein TVD-Verfahren ist TV-stabil.*

**Beweis.** Aus der Definition eines TVD-Verfahrens folgt, daß

$$TV(U^n) \leq TV(U^{n-1}) \leq \dots \leq TV(U^0) \quad (1.140)$$

gilt. Damit ist  $TV(U^0)$  die gesuchte Obergrenze für alle  $TV(U^n)$ , die unabhängig von  $\Delta x$ ,  $\Delta t$  und  $n$  ist.  $\square$

**Theorem 3** *Ein TVD-Verfahren mit einer Lipschitz-stetigen Konvektionsstromfunktion ist monotonie-erhaltend.*

**Beweis.** Die Monotonie-Erhaltung eines TVD-Verfahrens folgt daraus, daß die Entstehung von Oszillationen die totale Variation der Lösung erhöht. Den vollständigen Beweis findet man im Anhang A.1.  $\square$

Aus diesen drei Sätzen folgt, daß **ein konsistentes TVD-Verfahren mit einem Lipschitz-stetigen Konvektionsstrom konvergent ist und oszillationsfreie Lösungen liefert**. Offen bleibt noch die Frage, wie kann man feststellen, ob ein Verfahren die TVD-Eigenschaft besitzt oder nicht. Folgendes Theorem von *Harten* [40, 41] gibt Antwort auf diese Frage.

**Theorem 4** *Kann ein Finite-Volumen-Verfahren in folgender Form dargestellt werden*

$$U_i = \bar{U}_i - C_{i-1}(U_i - U_{i-1}) + D_i(U_{i+1} - U_i), \quad (1.141)$$

*dann hat dieses Verfahren die TVD-Eigenschaft, wenn folgende Bedingungen erfüllt sind:*

$$0 \leq C_i \leq C < \infty, \quad \forall i \quad (1.142)$$

$$0 \leq D_i \leq D < \infty, \quad \forall i \quad (1.143)$$

Der Ausdruck (1.141) erweckt den Eindruck, daß es sich hier um ein lineares Verfahren handelt. Das muß allerdings nicht notwendigerweise der Fall sein. Die Koeffizienten  $C_i$  und  $D_i$  können durchaus selbst von den Werten der Gitterfunktion  $U$  abhängen.

**Beweis.** In [40] hat Harten eine Variante dieses Theorems für explizite Finite-Volumen-Verfahren angegeben. In [41] wurde dieser Satz für implizite Verfahren mit periodischer Lösung in einem endlichen Intervall bewiesen. Wir werden stattdessen verlangen, daß

$$\lim_{i \rightarrow \pm\infty} |U_i - U_{i-1}| = 0 \quad (1.144)$$

gilt. Den Beweis des Theorems (unter der getroffenen Annahme (1.144)) findet man im Anhang A.2.□

Als Illustration betrachten wir das Upwind-Verfahren 1.Ordnung. Die Differenzenformel

$$\frac{U_i - \bar{U}_i}{\Delta t} + a \frac{U_i - U_{i-1}}{\Delta x} = 0. \quad (1.145)$$

läßt sich in der Form

$$U_i = \bar{U}_i - \frac{a\Delta t}{\Delta x} (U_i - U_{i-1}) \quad (1.146)$$

schreiben. Das entspricht (1.141) mit

$$C_{i-1} = \frac{a\Delta t}{\Delta x} \quad (1.147)$$

$$D_i = 0 \quad (1.148)$$

und die Bedingungen (1.142, 1.143) sind offensichtlich erfüllt.

Man beachte, daß für ein unendliches Gebiet die Gleichung (1.146) keine eindeutige Lösung besitzt. Man kann einen beliebigen Index  $i_0$  wählen, den Wert  $U_{i_0}$  frei setzen, und alle anderen Werte der Gitterfunktion  $U$  aus (1.146) rekursiv berechnen. Man müsste dazu lediglich die Gleichung (1.146) nach  $U_{i-1}$  (für  $i < i_0$ ) bzw. nach  $U_i$  (für  $i > i_0$ ) auflösen.

Um Eindeutigkeit der Lösung zu gewährleisten, könnte man z.B. verlangen, daß die Bedingung

$$\lim_{i \rightarrow -\infty} U_i = \lim_{x \rightarrow -\infty} u(x, t_{n+1}) \equiv \lim_{x \rightarrow -\infty} u_0(x) \quad (1.149)$$

erfüllt wird. Der letzte Grenzwert existiert, da wir angenommen haben, daß die totale Variation  $TV(u_0) < \infty$  ist. Aus (1.149) folgt automatisch, daß die Voraussetzung (1.144) für  $i \rightarrow -\infty$  erfüllt ist.

Auf der anderen Seite, folgt aus (A.9) für das Upwind-Verfahren

$$S(i_1, i_2) := \sum_{i=i_1}^{i=i_2} |U_{i+1} - U_i| \leq TV(\bar{U}) + \frac{a\Delta t}{\Delta x} |U_{i_1} - U_{i_1-1}|. \quad (1.150)$$

Beim fixierten Wert von  $i_1$  bildet die Reihe  $S(i_1, i_2)$  eine monoton wachsende Folge, die von oben begrenzt ist. Deshalb hat diese Folge einen endlichen Grenzwert bei  $i_2 \rightarrow +\infty$ , woraus folgt, daß die Voraussetzung (1.144) auch für  $i \rightarrow +\infty$  erfüllt ist.

## 1.9 Konstruktion eines TVD-Verfahrens höherer Ordnung

In diesem Abschnitt werden wir das Theorem von Harten dazu verwenden, um TVD-Verfahren 2. und 3. Ordnung zu konstruieren. Im Abschnitt 1.6 haben wir bei der Herleitung des LUDS-Verfahrens eine „konzeptionelle Mischung“ aus Upwind-Verfahren und CDS benutzt. Beim Aufbau eines TVD-Verfahrens werden wir ähnlich vorgehen, indem wir *direkt* ein gewichtetes Mittel aus dem Upwind- und CDS-Konvektionsstrom bilden.

Der numerische Konvektionsstrom für das CDS-Verfahren läßt sich analog zu (1.116) als Upwind-Konvektionsstrom plus Korrekturterm darstellen:

$$\begin{aligned} F^{cds}(U; l) &= a \frac{U_{i-1} + U_i}{2} \\ &= F^{up}(U; l) + (F^{cds}(U; l) - F^{up}(U; l)) \\ &= aU_{i-1} + \frac{a}{2}(U_i - U_{i-1}) \end{aligned} \quad (1.151)$$

Hier und weiter verwenden wir die Abkürzung „ $l$ “ für den Ausdruck „ $i - 1/2$ “. Der Ausdruck „ $i + 1/2$ “ wird entsprechend durch „ $r$ “ abgekürzt.

Für ein gewichtetes Mittel aus  $F^{up}(U; l)$  und  $F^{cds}(U; l)$  gilt daher

$$\begin{aligned} F^{tvd}(U; l) &:= (1 - \phi_l)F^{up}(U; l) + \phi_l F^{cds}(U; l) \\ &= F^{up}(U; l) + \phi_l (F^{cds}(U; l) - F^{up}(U; l)) \\ &= aU_{i-1} + \phi_l \frac{a}{2}(U_i - U_{i-1}). \end{aligned} \quad (1.152)$$

Setzt man diesen Konvektionsstrom in die Finite-Volumen-Gleichung (1.53) ein, so ergibt sich folgende Diskretisierung der Konvektionsgleichung:

$$\boxed{\text{TVD} : \frac{U_i - \bar{U}_i}{\Delta t} + a \frac{U_i + \frac{\phi_r}{2}(U_{i+1} - U_i) - U_{i-1} - \frac{\phi_l}{2}(U_i - U_{i-1})}{\Delta x} = 0.} \quad (1.153)$$

Würde man den Gewichtungsfaktor  $\phi_l$  für alle Volumina gleich nehmen, so wäre das resultierende Diskretisierungsverfahren linear. Für ein lineares Verfahren gilt aber, daß es entweder höchstens 1. Ordnung genau ist, oder nicht monotonie-erhaltend ist (s. Abschnitt 1.8). Es wird deshalb für jedes Kontrollvolumen ein *eigener* Gewichtungsfaktor gesucht, der vom Lösungsverlauf in der unmittelbaren Umgebung des betrachteten Volumenelements abhängt. Ist die Lösung an der Stelle  $x_i$  glatt genug, so kann man die Zentralfdifferenz einsetzen, an kritischen Stellen dagegen sollte man den Upwind-Anteil übergewichten. Zu den kritischen Stellen gehören insbesondere diejenigen, wo die zweite Ableitung der Lösung hoch ist, sowie die Stellen der lokalen Extrema der Lösungsfunktion, da dort das CDS-Verfahren die meisten Oszillationen verursacht. Zur Beschreibung der Glattheit der Lösungsfunktion benutzen wir das Verhältnis von aufeinanderfolgenden Gradienten

$$\theta_l := \frac{U_{i-1} - U_{i-2}}{U_i - U_{i-1}} \quad (1.154)$$

und suchen den Gewichtungsfaktor  $\phi_l$  als eine Funktion von  $\theta_l$ :

$$\boxed{\phi_l = \phi(\theta_l)}. \quad (1.155)$$

Da die Funktion  $\phi(\theta)$  den Beitrag des CDS-Konvektionsstromes in Abhängigkeit vom lokalen Lösungsverhalten limitiert, wird sie als „*Limitier-Funktion*“ oder kürzer als „*Limitier*“ bezeichnet.

Man sieht, daß in die Definition der Größe  $\theta_l$  zwei Werte der Gitterfunktion eingehen, die stromaufwärts von  $x_i$  liegen ( $U_{i-2}$  und  $U_{i-1}$ ), hier wird also die Stromrichtung mitberücksichtigt. Ist der Wert von  $\theta_l$  negativ, so bedeutet das, daß  $x_i$  in der Nähe eines lokalen Extremums der Gitterfunktion  $U$  liegt. Wir werden an solchen Stellen den Konvektionsstrom  $F^{tvd}(U; l)$  gleich dem Upwind-Strom setzen, das entspricht

$$\boxed{\phi(\theta_l) = 0, \text{ wenn } \theta_l \leq 0.} \quad (1.156)$$

Für positive Werte von  $\theta_l$  verlangen wir dagegen, daß der Gewichtungsfaktor positiv ist:

$$\boxed{\phi(\theta_l) > 0, \text{ wenn } \theta_l > 0.} \quad (1.157)$$

Ist der Nenner in (1.154) gleich Null, so bedeutet das, daß die Werte  $U_{i-1}$  und  $U_i$  gleich sind. In diesem Fall stimmen auch die Größen  $F^{up}(U; l)$  und  $F^{cds}(U; l)$  überein und der Wert des Gewichtungsfaktors  $\phi_l$  beeinflußt den Konvektionsstrom  $F^{tvd}(U; l)$  nicht. Einfachheitshalber wird der Wert von  $\phi_l$  zu Null gesetzt:

$$\boxed{\phi_l = 0, \text{ wenn } U_i - U_{i-1} = 0.} \quad (1.158)$$

Wir wollen nun untersuchen, welche Eigenschaften die Funktion  $\phi(\theta)$  besitzen soll, damit das Diskretisierungsverfahren (1.153) die TVD-Eigenschaft hat. Wir nehmen zunächst an, daß  $(U_i - U_{i-1}) \neq 0$  und  $(U_{i+1} - U_i) \neq 0$  gilt, setzen  $\phi_l = \phi(\theta_l)$  und  $\phi_r = \phi(\theta_r)$  in die Gleichung (1.153) ein, und schreiben die resultierende Gleichung in folgender Form:

$$U_i = \bar{U}_i - \frac{a\Delta t}{\Delta x} \left( U_i + \frac{\phi(\theta_r)}{2}(U_{i+1} - U_i) - U_{i-1} - \frac{\phi(\theta_l)}{2}(U_i - U_{i-1}) \right). \quad (1.159)$$

Hier bezeichnet  $\theta_r$  folgenden Ausdruck

$$\theta_r := \frac{U_i - U_{i-1}}{U_{i+1} - U_i} \quad (1.160)$$

Es gibt **drei** Möglichkeiten, die Formel (1.159) in Form (1.141) darzustellen. Die einfachste Möglichkeit

$$U_i = \bar{U}_i - \frac{a\Delta t}{2\Delta x} (2 - \phi(\theta_l))(U_i - U_{i-1}) - \frac{a\Delta t}{2\Delta x} \phi(\theta_r)(U_{i+1} - U_i) \quad (1.161)$$

führt zu

$$C_{i-1} = \frac{a\Delta t}{2\Delta x} (2 - \phi(\theta_l)) \quad (1.162)$$

$$D_i = -\frac{a\Delta t}{2\Delta x} \phi(\theta_r) \quad (1.163)$$

Wegen (1.157) ist  $D_i$  für positive Werte von  $\theta_r$  negativ und die Bedingung (1.143) ist offensichtlich verletzt.

Die zweite Möglichkeit besteht darin, daß wir den Koeffizienten  $C_{i-1}$  zu Null machen, indem wir die Gleichung (1.159) folgendermaßen umschreiben:

$$U_i = \bar{U}_i - \frac{a\Delta t}{2\Delta x} \left( (2 - \phi(\theta_l)) \frac{U_i - U_{i-1}}{U_{i+1} - U_i} + \phi(\theta_r) \right) (U_{i+1} - U_i) \quad (1.164)$$

Das entspricht (1.141) mit

$$C_{i-1} = 0 \quad (1.165)$$

$$\begin{aligned} D_i &= -\frac{a\Delta t}{2\Delta x} \left( (2 - \phi(\theta_l)) \frac{U_i - U_{i-1}}{U_{i+1} - U_i} + \phi(\theta_r) \right) \\ &= -\frac{a\Delta t}{2\Delta x} ((2 - \phi(\theta_l))\theta_r + \phi(\theta_r)) \end{aligned} \quad (1.166)$$

Ist  $\theta_l < 0$  und  $\theta_r > 0$ , so gilt wegen (1.156, 1.157)

$$D_i = -\frac{a\Delta t}{2\Delta x} (2\theta_r + \phi(\theta_r)) < 0 \quad (1.167)$$

und die Bedingung (1.143) ist wieder verletzt.

Schließlich kann man die Gleichung (1.159) analog zu (1.164) so umschreiben, daß der Koeffizient  $D_i$  zu Null wird:

$$U_i = \bar{U}_i - \frac{a\Delta t}{2\Delta x} \left( 2 - \phi(\theta_l) + \phi(\theta_r) \frac{U_{i+1} - U_i}{U_i - U_{i-1}} \right) (U_i - U_{i-1}) \quad (1.168)$$

Das entspricht (1.141) mit

$$\begin{aligned} C_{i-1} &= \frac{a\Delta t}{2\Delta x} \left( 2 - \phi(\theta_l) + \phi(\theta_r) \frac{U_{i+1} - U_i}{U_i - U_{i-1}} \right) \\ &= \frac{a\Delta t}{2\Delta x} \left( 2 - \phi(\theta_l) + \frac{\phi(\theta_r)}{\theta_r} \right) \end{aligned} \quad (1.169)$$

$$D_i = 0 \quad (1.170)$$

Damit die Bedingung (1.142) erfüllt ist, muß folgende Ungleichung gelten

$$\boxed{0 \leq 2 - \phi(\theta_l) + \frac{\phi(\theta_r)}{\theta_r} \leq C} \quad (1.171)$$

Es gilt folgendes

**Theorem 5** *Damit das Diskretisierungsverfahren (1.153) die TVD-Eigenschaft hat, ist es hinreichend, daß der Limiter  $\phi(\theta)$  neben (1.156, 1.157, 1.158) folgende Bedingungen erfüllt*

$$\phi(\theta) \leq 2, \quad \forall \theta > 0, \quad (1.172)$$

$$\phi(\theta) \leq 2\theta, \quad \forall \theta > 0. \quad (1.173)$$

**Beweis.** Wir zeigen zunächst, daß die Ungleichungen (1.171) erfüllt sind. Aus (1.156, 1.157, 1.172) folgt:

$$2 - \phi(\theta_l) + \frac{\phi(\theta_r)}{\theta_r} \geq 2 - \phi(\theta_l) \geq 0. \quad (1.174)$$

Auf der anderen Seite gilt wegen (1.156, 1.157, 1.173):

$$2 - \phi(\theta_l) + \frac{\phi(\theta_r)}{\theta_r} \leq 2 + \frac{\phi(\theta_r)}{\theta_r} \leq 4 \quad (1.175)$$

Nun sollen wir noch 3 Sonderfälle betrachten, wenn eins von  $\theta_l$ ,  $\theta_r$  oder beide gleichzeitig nicht definiert sind. Ist  $(U_i - U_{i-1}) = 0$ , so ist auch  $\phi_l = 0$  wegen (1.158). Gleichzeitig gilt

entweder  $\theta_r = 0$  oder  $(U_{i+1} - U_i) = 0$ , in beiden Fällen ist auch  $\phi_r = 0$  (1.156, 1.158). Die resultierende Diskretisierungsformel stimmt dann mit dem Upwind-Verfahren überein, das die Harten-Bedingungen (1.142, 1.143) erfüllt, wie am Schluß des vorigen Abschnittes gezeigt wurde.

Ist nun  $(U_i - U_{i-1}) \neq 0$  und  $(U_{i+1} - U_i) = 0$ , führt (1.152) zu

$$U_i = \bar{U}_i - \frac{a\Delta t}{2\Delta x}(2 - \phi(\theta_i))(U_i - U_{i-1}) \quad (1.176)$$

und die Harten-Bedingungen (1.142, 1.143) sind wegen (1.156, 1.157, 1.172) ebenfalls erfüllt.  $\square$

Theorem 5 gibt hinreichende Bedingungen dafür an, daß der Limiter  $\phi$  zu einem TVD-Verfahren führt. Über die Approximationsordnung des resultierenden Verfahrens sagt es allerdings nichts aus. Im Abschnitt 1.7 wurde gezeigt, daß jedes lineare Upwind-Verfahren, das mindestens 2. Ordnung genau ist, als ein gewichtetes Mittel aus Zentral-Differenz und LUDS-Verfahren dargestellt werden kann. Würde man als Limiter-Funktion  $\phi(\theta) \equiv 1$  wählen, so würde der Konvektionsstrom (1.152) mit dem CDS-Konvektionsstrom übereinstimmen. Die Limiter-Funktion  $\phi(\theta) = \theta$  entspricht dagegen dem LUDS-Verfahren. Es wäre daher sinnvoll zu verlangen, daß ein TVD-Limiter nicht nur den Bedingungen des Theorems 5 genügt, sondern auch für  $\theta > 0$  zwischen den Geraden  $\phi(\theta) \equiv 1$  und  $\phi(\theta) = \theta$  liegt. Der Bereich, in welchem die Werte der Limiter-Funktion unter den gemachten Voraussetzungen liegen dürfen, ist in der Abb. 1.38 dargestellt.

Man sieht, daß die Limiter-Funktion, deren Werte im gekennzeichneten Bereich liegen, zwangsläufig für  $\theta = 1$  den Wert 1 annehmen muß.

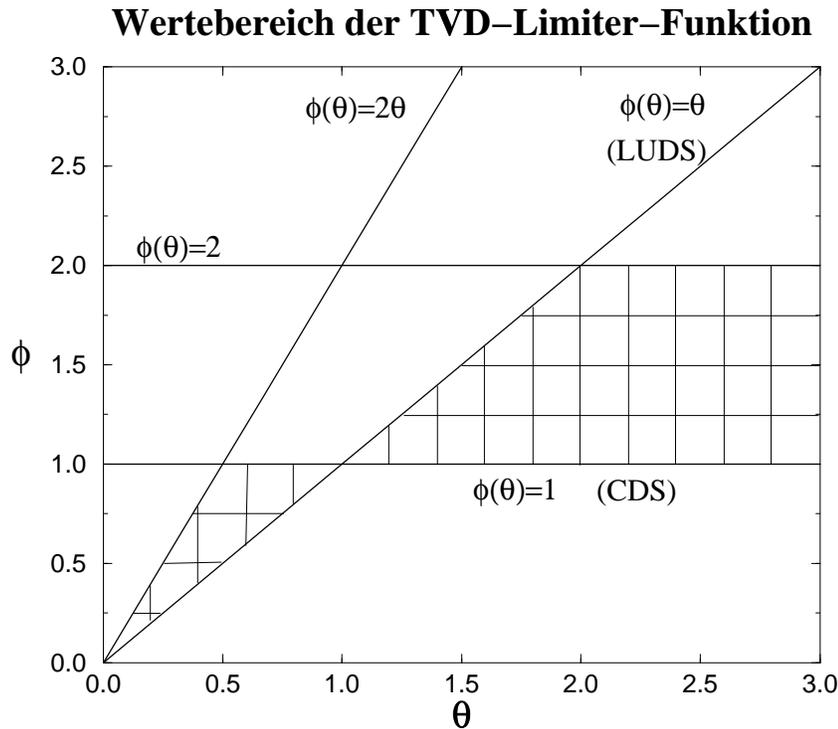
**Theorem 6** *Damit das Diskretisierungsverfahren (1.153) für hinreichend glatte Lösungen an den Stellen  $(x, t)$  mit  $u_x \neq 0$  die Approximationsordnung 2 hat, ist es hinreichend, daß der Limiter  $\phi(\theta)$  folgende Bedingungen erfüllt*

1.  $\phi(1) = 1.$
2. *Die Limiter-Funktion  $\phi$  ist Lipschitz-stetig in der Nähe von  $\theta = 1.$*

**Beweis.** Siehe Anhang A.3.  $\square$

Wir wollen die bisherigen Ergebnisse zusammenfassen.

Das Diskretisierungsverfahren (1.153) mit einer Lipschitz-stetigen Limiter-Funktion, die Werte im kariert dargestellten Bereich der Abb. 1.38 annimmt, **ist TV-stabil, produziert oszillationsfreie Lösungen und hat bis auf die kritischen Stellen der Lösungsfunktion Approximationsordnung 2.**



**Abbildung 1.38:** Wertebereich der TVD-Limiter-Funktion. Der zulässige Bereich ist kariert dargestellt.

Obwohl aus TV-Stabilität und Konsistenz eines Verfahrens dessen Konvergenz folgt, kann man im allgemeinen nicht davon ausgehen, daß die Konvergenzordnung und die Approximationsordnung des Verfahrens übereinstimmen. Wir werden daher die Konvergenzordnung der TVD-Verfahren in folgenden Abschnitten an unseren Testfällen empirisch ermitteln.

Zunächst müssen wir allerdings die Limiter-Funktion genau spezifizieren. Die am weitesten verbreiteten Limiter sind folgende:

$$\text{minmod} : \quad \phi^{mm}(\theta) = \max(0, \min(1, \theta)) \quad (1.177)$$

$$\text{superbee} : \quad \phi^{sb}(\theta) = \max(0, \min(1, 2\theta), \min(2, \theta)) \quad (1.178)$$

$$\text{vanLeer} : \quad \phi^{vl}(\theta) = \frac{\theta + |\theta|}{1 + |\theta|} \quad (1.179)$$

$$\text{monotonized centered} : \quad \phi^{mc}(\theta) = \max \left\{ 0, \min \left( \frac{1 + \theta}{2}, 2, 2\theta \right) \right\} \quad (1.180)$$

Der Verlauf dieser Limiter-Funktionen ist in der Abb. 1.39 skizziert. Der *Minmod-Limiter*

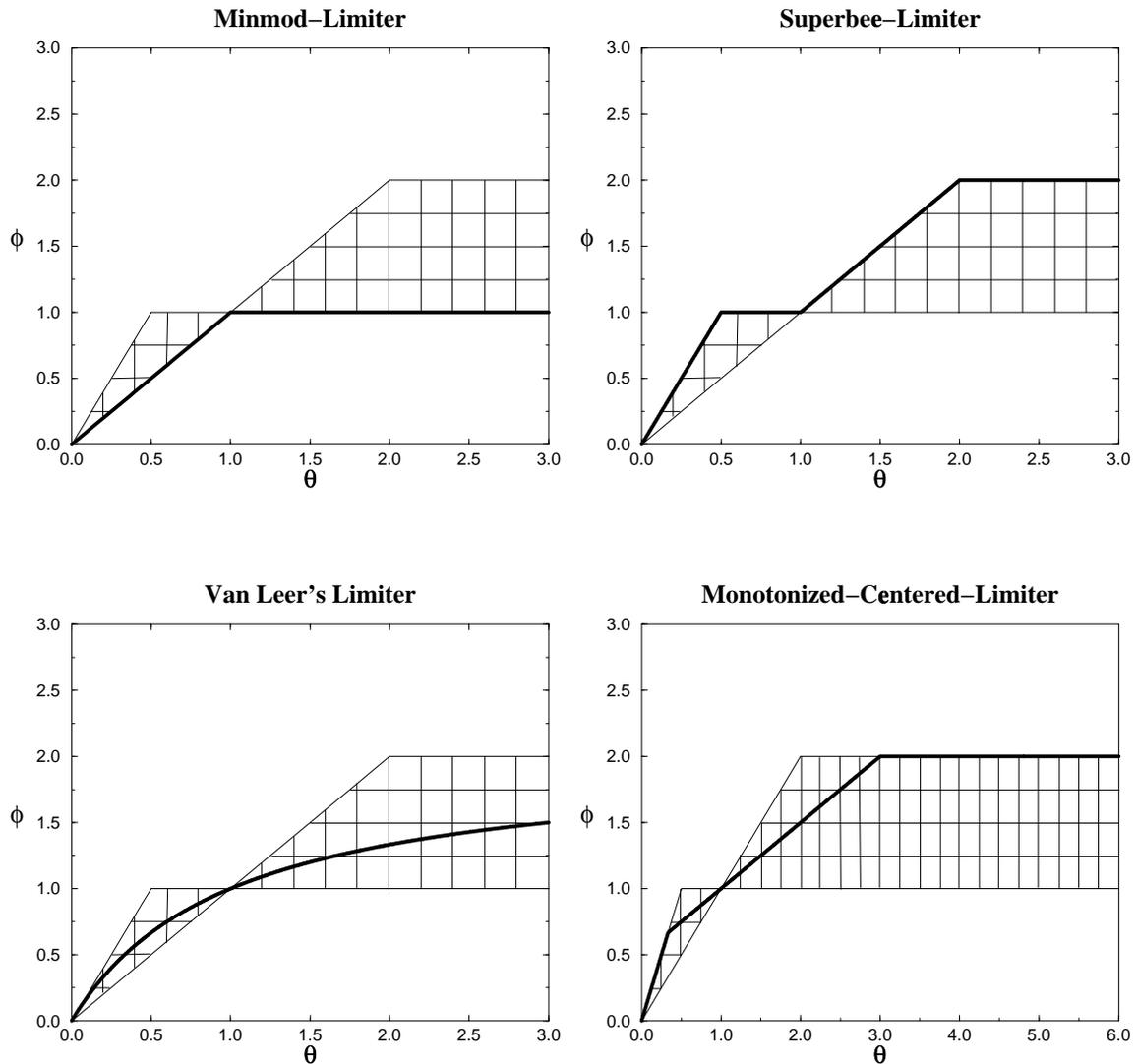


Abbildung 1.39: Skizzen der Limiter-Funktionen.

(kurz: MM) verläuft entlang der unteren Grenze des zulässigen Bereichs. Der Name des Limiters kommt daher, daß für den Limiter  $\phi^{mm}$  gilt:

$$\phi^{mm}(\theta) = \min\text{mod}(1, \theta) \quad (1.181)$$

Dabei ist die Funktion  $\min\text{mod}(x, y)$  so definiert, daß sie den Wert 0 annimmt, wenn ihre Argumente unterschiedliche Vorzeichen haben, ansonsten ist sie gleich dem Maximum aus den Beträgen beider Argumente. Sie kann auch durch eine einzige Formel definiert werden:

$$\min\text{mod}(x, y) = \text{sign}(x) \cdot \max[0, \min(|x|, \text{sign}(x) \cdot y)]. \quad (1.182)$$

Der *Superbee-Limiter* (kurz: SB) verläuft dagegen entlang der oberen Grenze des zulässigen

Bereichs. Er wurde von *Roe* [101] im Jahr 1985 zum ersten Mal eingesetzt und verdankt ihm seinen Namen.

Der *van Leer's Limiter* (kurz: VL) verläuft zwischen den ersten beiden Limiter-Funktionen und ist im Bereich  $\theta > 0$  überall differenzierbar. Er wurde bereits im Jahre 1974 eingesetzt [128], und trägt den Namen seines „Erfinders“. Im Bereich  $\theta > 2$  nähert sich der VL-Limiter nur sehr langsam dem Wert 2. Der *Monotonized-Centered-Limiter* (kurz: MC) verläuft im Bereich  $\theta < 2$  ähnlich wie VL-Limiter, erreicht danach aber ziemlich schnell die obere Grenze des zulässigen Bereichs. Er wurde ebenfalls von *van Leer* eingeführt [129].

Alle vier vorgestellten Limiter haben die sogenannte *Symmetrie-Eigenschaft*:

$$\frac{\phi(\theta)}{\theta} = \phi\left(\frac{1}{\theta}\right). \quad (1.183)$$

Wir werden diese Eigenschaft noch später in diesem Abschnitt benötigen.

**Alternative Vorgehensweise.** Bei der Konstruktion eines TVD-Verfahrens sind wir bisher vom numerischen Konvektionsstrom ausgegangen, der als gewichtetes Mittel aus Upwind- und CDS-Konvektionsstrom dargestellt wurde. Man könnte aber anstatt von CDS-Strom  $F^{cds}(U; l)$  auch einen anderen Konvektionsstrom 2.Ordnung nehmen, z.B. den LUDS-Konvektionsstrom  $F^{luds}(U; l)$ . Wir wollen nun untersuchen, ob man auf diesem Wege zu prinzipiell anderen TVD-Verfahren gelangen kann.

Für ein gewichtetes Mittel aus  $F^{up}(U; l)$  und  $F^{luds}(U; l)$  gilt (s. auch (1.116)):

$$\begin{aligned} \overline{F}^{tvd}(U; l) &= (1 - \psi_l)F^{up}(U; l) + \psi_l F^{luds}(U; l) \\ &= F^{up}(U; l) + \psi_l(F^{luds}(U; l) - F^{up}(U; l)) \\ &= aU_{i-1} + \psi_l \frac{a}{2}(U_{i-1} - U_{i-2}). \end{aligned} \quad (1.184)$$

Wir bezeichnen den neuen Konvektionsstrom mit  $\overline{F}^{tvd}(U; l)$  und den Gewichtungsfaktor mit  $\psi_l$ , um sie von  $F^{tvd}(U; l)$  und  $\phi_l$  zu unterscheiden. Analog zu (1.159) gilt:

$$U_i = \overline{U}_i - \frac{a\Delta t}{\Delta x} \left( U_i + \frac{\psi(\theta_r)}{2}(U_i - U_{i-1}) - U_{i-1} - \frac{\psi(\theta_l)}{2}(U_{i-1} - U_{i-2}) \right), \quad (1.185)$$

was man folgendermaßen in der Form (1.141) darstellen kann (mit  $D_i = 0$ ):

$$U_i = \overline{U}_i - \frac{a\Delta t}{2\Delta x} \left( 2 + \psi(\theta_r) - \psi(\theta_l) \frac{U_{i-1} - U_{i-2}}{U_i - U_{i-1}} \right) (U_i - U_{i-1}) \quad (1.186)$$

Das entspricht (1.141) mit

$$\begin{aligned} C_{i-1} &= \frac{a\Delta t}{2\Delta x} \left( 2 + \psi(\theta_r) - \psi(\theta_l) \frac{U_{i-1} - U_{i-2}}{U_i - U_{i-1}} \right) \\ &= \frac{a\Delta t}{2\Delta x} (2 + \psi(\theta_r) - \psi(\theta_l)\theta_l) \end{aligned} \quad (1.187)$$

$$D_i = 0 \quad (1.188)$$

Damit die Bedingung (1.142) erfüllt ist, müssen folgende Ungleichungen gelten:

$$\boxed{0 \leq 2 + \psi(\theta_r) - \psi(\theta_l)\theta_l \leq C} \quad (1.189)$$

Anstatt den Limiter  $\psi(\theta)$  neu zu erfinden, kann man mit Hilfe des folgenden Theorems die bereits nachgewiesene Eigenschaften der Limiter-Funktionen  $\phi^{mm}(\theta)$ ,  $\phi^{sb}(\theta)$ ,  $\phi^{vl}(\theta)$  und  $\phi^{mc}(\theta)$  sinnvoll ausnutzen:

**Theorem 7** *Genügt die Limiter-Funktion  $\phi(\theta)$  den Voraussetzungen des Theorems 5, und ist die Funktion  $\psi(\theta)$  entweder als*

$$1. \quad \psi(\theta) := \frac{\phi(\theta)}{\theta}$$

oder als

$$2. \quad \psi(\theta) := \phi\left(\frac{1}{\theta}\right)$$

definiert, so erfüllt die Funktion  $\psi(\theta)$  die Bedingungen (1.189)

**Beweis.** Siehe Anhang A.4.□

Wir wollen nun untersuchen, was passiert, wenn wir die auf diese Weise konstruierten Limiter in die Definition des Konvektionsstromes (1.184) einsetzen. Im Fall  $\psi(\theta) = \phi(\theta)/\theta$  erhält man:

$$\begin{aligned} \overline{F}^{tvd}(U; l) &= aU_{i-1} + \psi(\theta_l) \cdot \frac{a}{2} \cdot (U_{i-1} - U_{i-2}) \\ &= aU_{i-1} + \frac{\phi(\theta_l)}{\theta_l} \cdot \frac{a}{2} \cdot (U_{i-1} - U_{i-2}) \\ &= aU_{i-1} + \phi(\theta_l) \cdot \frac{U_i - U_{i-1}}{U_{i-1} - U_{i-2}} \cdot \frac{a}{2} \cdot (U_{i-1} - U_{i-2}) \\ &= aU_{i-1} + \phi(\theta_l) \cdot \frac{a}{2} \cdot (U_i - U_{i-1}) \equiv F^{tvd}(U; l), \end{aligned} \quad (1.190)$$

d.h. die Konvektionsströme  $\overline{F}^{tvd}(U; l)$  und  $F^{tvd}(U; l)$  stimmen in diesem Fall überein!

Würde man stattdessen den Ausdruck  $\psi(\theta) = \phi(1/\theta)$  in (1.184) einsetzen, so würde das im allgemeinen zu einem anderen Diskretisierungsverfahren führen. Für die vorgestellten 4 Limiter-Funktionen  $\phi^{mm}(\theta)$ ,  $\phi^{sb}(\theta)$ ,  $\phi^{vl}(\theta)$  und  $\phi^{mc}(\theta)$  allerdings, stimmen wegen deren Symmetrie-Eigenschaft (1.183) die Funktionen  $\psi(\theta) = \phi(1/\theta)$  und  $\psi(\theta) = \phi(\theta)/\theta$  überein, so daß auch in diesem Fall gilt  $\overline{F}^{tvd}(U; l) \equiv F^{tvd}(U; l)$ . Wir haben somit gezeigt, daß

unabhängig davon, ob man den TVD-Strom als gewichtetes Mittel aus Upwind und CDS mit einem symmetrischen Limiter  $\phi(\theta)$ , oder als gewichtetes Mittel aus Upwind und LUDS mit einem Limiter  $\psi(\theta) = \phi(\theta)/\theta$  oder  $\psi(\theta) = \phi(1/\theta)$  konstruiert, **sind die resultierenden Diskretisierungsverfahren in allen drei Fällen identisch.**

**Alternative Limiter.** Beim Aufbau der 4 vorgestellten Limiter-Funktionen (1.177-1.180) ging man von geometrischen Überlegungen aus: entscheidend war, daß deren Werte im vorgeschriebenen Bereich lagen. Man kann aber als *Ausgangspunkt* auch eine beliebige *Funktion* nehmen, und *anschließend* ihre Werte wenn notwendig *korrigieren*.

Wir haben vorher schon gesehen, daß die Funktion  $\phi(\theta) \equiv 1$  dem CDS-Verfahren und die Funktion  $\phi(\theta) = \theta$  dem LUDS-Verfahren entsprechen. Das Agarwal-Verfahren und das QUICK-Verfahren erhält man durch  $\phi(\theta) = 2/3 + \theta/3$  bzw.  $\phi(\theta) = 3/4 + \theta/4$ , da diese beiden Verfahren durch lineare Kombination aus CDS und LUDS gebildet werden können. Offensichtlich liegen diese vier Funktionen nicht im zulässigen Wertebereich der Abbildung 1.38. Man kann aber aus diesen Funktionen die entsprechenden TVD-Limiter dadurch konstruieren, daß man die Funktionswerte dort verändert, wo sie den zulässigen Bereich verlassen, und ansonsten unverändert läßt. Die resultierenden Limiter sind in der Abb. 1.40 skizziert, und werden folgendermaßen definiert:

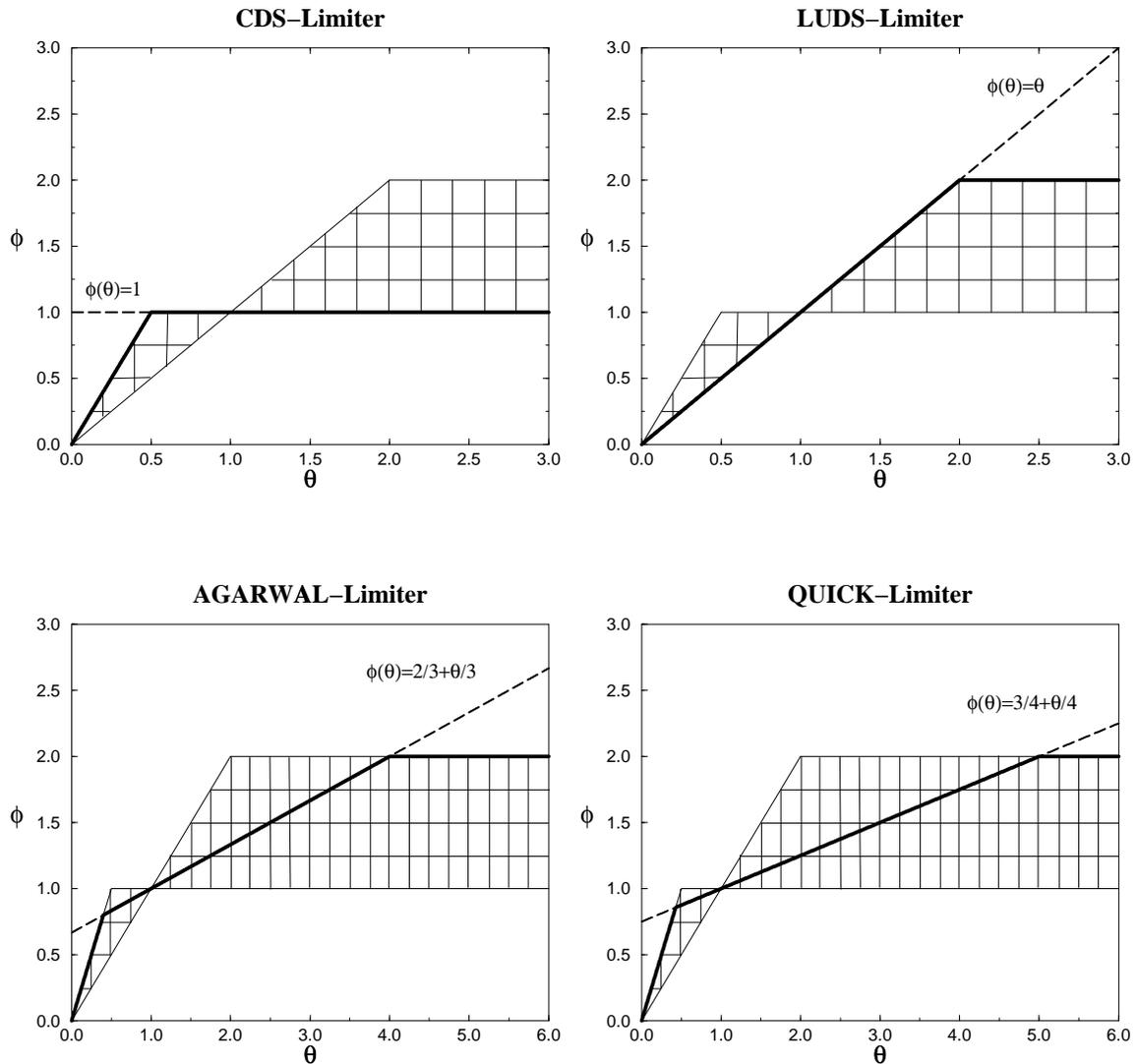
$$CDS : \quad \phi^{cds}(\theta) = \max \{0, \min(1, 2\theta)\} \quad (1.191)$$

$$LUDS : \quad \phi^{luds}(\theta) = \max \{0, \min(2, \theta)\} \quad (1.192)$$

$$AGARWAL : \quad \phi^{agar}(\theta) = \max \left\{ 0, \min \left( \frac{2+\theta}{3}, 2, 2\theta \right) \right\} \quad (1.193)$$

$$QUICK : \quad \phi^{quick}(\theta) = \max \left\{ 0, \min \left( \frac{3+\theta}{4}, 2, 2\theta \right) \right\} \quad (1.194)$$

Man beachte, daß diese vier Limiter-Funktionen die Symmetrie-Eigenschaft (1.183) nicht besitzen. Es gilt jedoch  $\phi^{cds}(1/\theta) = \phi^{luds}(\theta)/\theta$ , so daß der Konvektionsstrom  $\overline{F}^{tvd}(U; l)$  mit  $\psi(\theta) = \phi^{cds}(1/\theta)$  zu demselben Verfahren führt wie Konvektionsstrom  $F^{tvd}(U; l)$  mit



**Abbildung 1.40:** Skizzen der Limiter-Funktionen  $\phi^{cds}(\theta)$ ,  $\phi^{luds}(\theta)$ ,  $\phi^{agar}(\theta)$  und  $\phi^{quick}(\theta)$ .

$\phi^{luds}(\theta)$  und umgekehrt, der Konvektionsstrom  $\overline{F}^{tvd}(U; l)$  mit  $\psi(\theta) = \phi^{luds}(1/\theta)$  führt zu demselben Verfahren wie Konvektionsstrom  $F^{tvd}(U; l)$  mit  $\phi^{cds}(\theta)$ .

Alle vier Limiter genügen den Voraussetzungen des Theorems 6, so daß resultierende Verfahren an den unkritischen Stellen die Approximationsordnung von mindestens 2 haben. Der AGARWAL-Limiter  $\phi^{agar}(\theta)$  führt an solchen Stellen sogar zu einer Approximationsordnung von 3. Das folgt daraus, daß in der Nähe von  $\theta = 1$  der AGARWAL-Limiter mit der Funktion  $\phi(\theta) = 2/3 + \theta/3$  übereinstimmt (Abb. 1.40, unten links), so daß der entsprechende TVD-Verfahren die gleiche Approximationsordnung hat wie das Agarwal-Verfahren, und dessen Approximationsordnung ist bekanntlich 3.

Wir werden die limitierten Versionen der linearen Verfahren CDS, LUDS, AGARWAL und

QUICK entsprechend mit *L-CDS*, *L-LUDS*, *L-AGARWAL* und *L-QUICK* bezeichnen.

**Numerische Aspekte.** Bevor wir in den nächsten Abschnitten mit der Untersuchung des Einflusses verschiedener Limiter-Funktionen auf die numerische Lösung unserer Testfälle beginnen, sollen hier noch einige numerische Aspekte besprochen werden. Bisher sind wir bei der Herleitung eines TVD-Verfahrens von einem räumlich unbegrenzten Lösungsgebiet ausgegangen. In den Testfällen wollen wir das Verfahren allerdings für das Intervall  $[0, 2]$  einsetzen. Die Diskretisierung des Berechnungsgebietes sowie die Behandlung der Randelemente erfolgt analog zum Upwind-Verfahren und ist im Abschnitt 1.4 ausführlich beschrieben.

Zur Berechnung des numerischen Konvektionsstromes  $F^{tvd}(U; 1/2)$  braucht man den Wert der Gitterfunktion an der Stelle  $x_{-1}$ , die außerhalb des Berechnungsgebietes liegt. Wie beim LUDS-Verfahren kann man bei *stationären* Randbedingungen einen fiktiven Gitterpunkt einfügen und  $U_{-1} = U_0$  setzen. Alternativ kann man den Gewichtungsfaktor  $\phi_{1/2}$  gleich Null setzen, was im Endergebnis auf dasselbe hinausläuft. Die erste Variante ist programmtechnisch effektiver, da dann das erste Volumenelement auf die gleiche Weise wie alle anderen inneren Kontrollvolumina behandelt werden kann. Der *instationäre* Fall wird im Abschnitt 1.11 behandelt.

Das resultierende Gleichungssystem ist nichtlinear. Die Linearisierung kann durch den „deferred correction approach“ durchgeführt werden, der im Abschnitt 1.6 beschrieben wurde (s. (1.116, 1.117)). Für den Korrekturstrom  $F^{dc}(U; l)$  gilt (s. 1.155, 1.158):

$$\begin{aligned} F^{dc}(U; l) &= \phi_l \cdot \frac{a}{2} \cdot (U_i - U_{i-1}) = \\ &= \begin{cases} 0, & \text{wenn } U_i - U_{i-1} = 0 \\ \frac{a}{2} \cdot (U_i - U_{i-1}) \cdot \phi\left(\frac{U_{i-1} - U_{i-2}}{U_i - U_{i-1}}\right), & \text{wenn } U_i - U_{i-1} \neq 0 \end{cases} \end{aligned} \quad (1.195)$$

Wenn für einen Limiter  $\phi(\cdot)$  eine Funktion  $\bar{\phi}(\cdot, \cdot)$  existiert mit der Eigenschaft

$$\bar{\phi}(x, y) = \begin{cases} 0, & \text{wenn } y = 0 \\ y \cdot \phi\left(\frac{x}{y}\right), & \text{wenn } y \neq 0 \end{cases}, \quad (1.196)$$

dann kann man den Korrekturstrom  $F^{dc}(U; l)$  wesentlich einfacher berechnen durch:

$$F^{dc}(U; l) = \frac{a}{2} \cdot \bar{\phi}(U_{i-1} - U_{i-2}, U_i - U_{i-1}). \quad (1.197)$$

So eine Funktion  $\bar{\phi}$  existiert für 7 der 8 betrachteten Limiter. Wir zeigen exemplarisch die

Herleitung einer solchen Funktion für den MM-Limiter  $\phi^{mm}$ . Es gilt für  $y \neq 0$ :

$$\begin{aligned}
 y \cdot \phi^{mm} \left( \frac{x}{y} \right) &= y \cdot \max \left\{ 0, \min \left( 1, \frac{x}{y} \right) \right\} \\
 &= \text{sign}(y) \cdot |y| \cdot \max \left\{ 0, \min \left( 1, \frac{x}{y} \right) \right\} \\
 &= \text{sign}(y) \cdot \max \left\{ 0, |y| \cdot \min \left( 1, \frac{x}{y} \right) \right\} \\
 &= \text{sign}(y) \cdot \max \left\{ 0, \min \left( |y|, \frac{x|y|}{y} \right) \right\} \\
 &= \text{sign}(y) \cdot \max \{ 0, \min (|y|, x \cdot \text{sign}(y)) \} =: \bar{\phi}^{mm}(x, y).
 \end{aligned}$$

Wie man leicht überprüfen kann, ist die Bedingung  $\bar{\phi}^{mm}(x, 0) = 0$  ebenfalls erfüllt. Bis auf den VL-Limiter kann die Funktion  $\bar{\phi}$  analog hergeleitet werden. Wir präsentieren hier nur das Endergebnis:

$$\begin{aligned}
 \bar{\phi}^{sb}(x, y) &= \text{sign}(y) \cdot \max \{ 0, \min[|y|, 2x \cdot \text{sign}(y)], \min[2|y|, x \cdot \text{sign}(y)] \} \\
 \bar{\phi}^{mc}(x, y) &= \text{sign}(y) \cdot \max \left\{ 0, \min \left[ \frac{|y| + x \cdot \text{sign}(y)}{2}, 2|y|, 2x \cdot \text{sign}(y) \right] \right\} \\
 \bar{\phi}^{cds}(x, y) &= \text{sign}(y) \cdot \max \{ 0, \min[|y|, 2x \cdot \text{sign}(y)] \} \\
 \bar{\phi}^{luds}(x, y) &= \text{sign}(y) \cdot \max \{ 0, \min[2|y|, x \cdot \text{sign}(y)] \} \\
 \bar{\phi}^{agar}(x, y) &= \text{sign}(y) \cdot \max \left\{ 0, \min \left[ \frac{2|y| + x \cdot \text{sign}(y)}{3}, 2|y|, 2x \cdot \text{sign}(y) \right] \right\} \\
 \bar{\phi}^{quick}(x, y) &= \text{sign}(y) \cdot \max \left\{ 0, \min \left[ \frac{3|y| + x \cdot \text{sign}(y)}{4}, 2|y|, 2x \cdot \text{sign}(y) \right] \right\}
 \end{aligned}$$

Diese Idee zur Berechnung des Korrekturstromes mit Hilfe der Funktion  $\bar{\phi}$  stammt von Yee [135].

## 1.10 TVD-Verfahren: Untersuchung der Testfälle

In diesem Abschnitt wollen wir den Einfluß der Limiter-Funktionen auf die numerischen Lösungen der Testfälle 1 und 2 untersuchen. Durch den Aufbau der TVD-Verfahren ist sichergestellt, daß die Lösungen keine Oszillationen enthalten können. Die Unterdrückung der Oszillationen geschieht allerdings durch starke Übergewichtung des Upwind-Stromes in

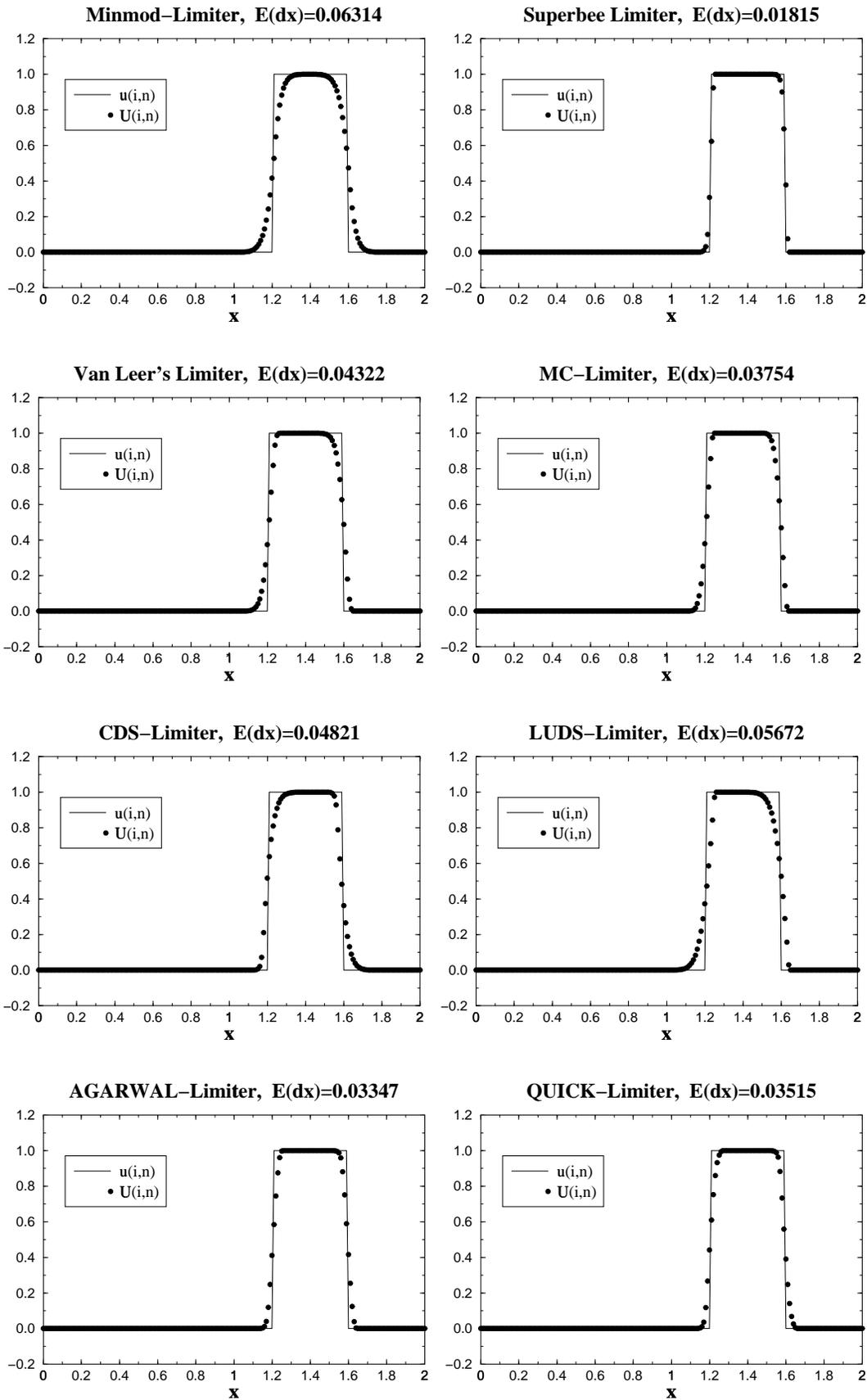


Abbildung 1.41: Test No.2: Verlauf der Lösung auf einem Gitter  $\Delta x = 0.01$  für 8 Limiter.

der Nähe der kritischen Stellen. Dieser ist allerdings nur erster Ordnung genau, so daß man erwarten kann, daß TVD-Verfahren zwar physikalisch sinnvolle Lösungen liefern, möglicherweise aber auf Kosten der Genauigkeit.

Wir beginnen mit der Untersuchung des Testbeispiels No.2, da in diesem Fall beim Einsatz von linearen Verfahren höherer Ordnung die Oszillationen am stärksten ausgeprägt waren (Abb. 1.22, 1.28, 1.33). In Abb. 1.41 ist der Verlauf der Lösung auf einem Gitter mit der Maschenweite  $\Delta x = 0.01$  für alle 8 Limiter-Funktionen dargestellt. Man sieht, daß die Lösungen in allen 8 Fällen tatsächlich keine Oszillationen aufweisen.

VERFAHREN	$p$	VERFAHREN	$p$	VERFAHREN	$p$
CDS	0.47	L-CDS	0.67	MM	0.66
LUDS	0.60	L-LUDS	0.66	SB	1.00
AGARWAL	0.75	L-AGARWAL	0.76	VL	0.69
QUICK	0.60	L-QUICK	0.71	MC	0.67

**Tabelle 1.4:** Testbeispiel No.2: Empirische Konvergenzordnung von 12 Verfahren

In der Tabelle 1.4 sind die empirischen Konvergenzordnungen für 4 lineare Verfahren sowie für 8 TVD-Verfahren zusammengetragen. Man sieht, daß die limitierten Versionen von allen vier linearen Verfahren (CDS, LUDS, AGARWAL und QUICK) eine höhere Konvergenzordnung haben, als die nichtlimitierten Originalverfahren. Auch die *absoluten Fehler* der limitierten Versionen sind auf fast allen Gittern niedriger, wie man auf den  $E-\Delta x$ -Diagrammen (Abb. 1.42) sehen kann. Besonders stark fällt die Fehlerreduktion beim CDS-Verfahren aus. Das lineare CDS-Verfahren leidete am meisten von den starken Oszillationen und profitiert daher am meisten von deren Eliminierung. Beim Agarwal-Verfahren fällt die Fehlerreduktion am kleinsten aus. Entscheidend ist aber, daß in allen vier Fällen der Einsatz der entsprechenden Limiter-Funktion die Oszillationen in der Lösung beseitigt und *gleichzeitig* die Genauigkeit der Lösung erhöht.

Sehr interessant ist auch der Vergleich der vier alternativen Limiter *untereinander* (Abb. 1.43, links). Im Abschnitt 1.6 haben wir gesehen (s. Abb. 1.29), daß das LUDS-Verfahren im Testfall No.2 auf allen Gittern genauere Lösungen erzielte als das CDS-Verfahren, obwohl der absolute Wert des Dispersionskoeffizienten des CDS-Verfahrens 2 mal kleiner ist. Wir haben dieses Phänomen damit erklärt, daß der Abbruchfehler des LUDS-Verfahrens einen zusätzlichen diffusiven Term enthält, der für starke Reduktion der Oszillationen sorgt im Vergleich zum CDS-Verfahren, was eine höhere Genauigkeit auf allen Gittern zu Folge hatte. Bei den *limitierten* Versionen beider Verfahren werden diese Oszillationen bereits durch den Einsatz des Limiters unterdrückt, so daß nun das L-CDS-Verfahren besser als das

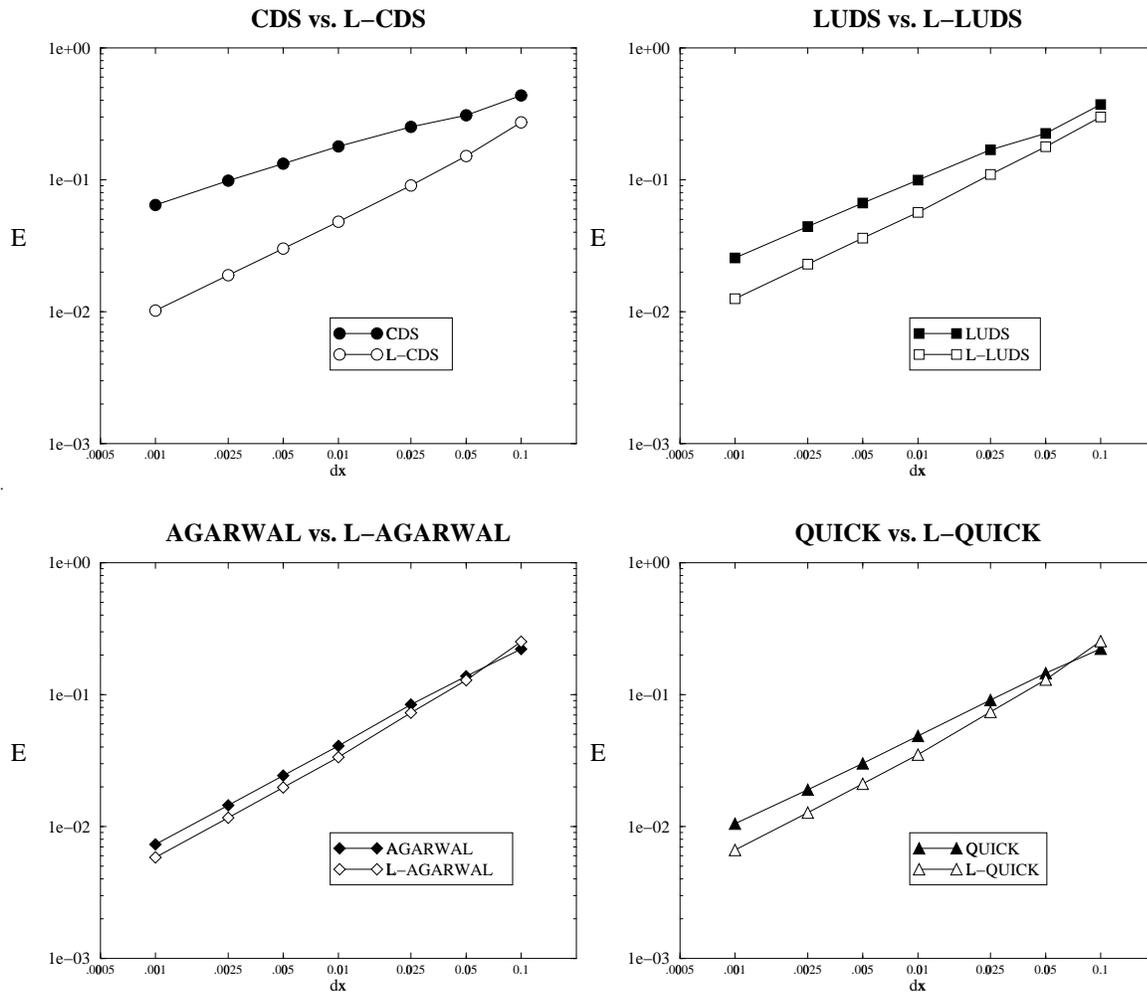


Abbildung 1.42: Test No.2: Vergleich von limitierten und nichtlimitierten Verfahren.

L-LUDS abschneidet. Die limitierten Versionen von AGARWAL und QUICK haben wiederum eine viel höhere Genauigkeit als L-CDS- und L-LUDS-Verfahren. Der Unterschied zwischen L-AGARWAL und L-QUICK fällt diesmal allerdings viel schwächer aus als im nichtlimitierten Fall (vgl. Abb. 1.37).

In der Abb. 1.43 (rechts) ist der Fehlerverlauf in Abhängigkeit von der Gitterauflösung für vier Standard-Limiter dargestellt. Zu Vergleichszwecken ist auch die  $E-\Delta x$ -Kurve für das Upwind-Verfahren erster Ordnung aufgetragen. Am schlechtesten schneidet der MM-Limiter ab. Man sieht in der Abb. 1.41 (ganz oben, links), daß beim Einsatz des MM-Limiters die Lösung am stärksten abgeflacht wird. Trotzdem ist die Genauigkeit und die Konvergenzordnung des MM-Verfahrens wesentlich höher als beim Upwind-Verfahren. Die  $E-\Delta x$ -Kurven von MC- und VL-Verfahren verlaufen sehr eng nebeneinander. Obwohl das VL-Verfahren eine etwas höhere Konvergenzordnung hat (s. Tab. 1.4), ist der absolute Feh-

ler des MC-Verfahrens auf allen Gittern niedriger. Die mit Abstand besten Ergebnisse liefert aber der SB-Limiter. Das SB-Verfahren hat sowohl die höchste Konvergenzordnung von allen untersuchten Verfahren als auch den kleinsten numerischen Fehler auf allen Gittern. In der Abb. 1.41 (ganz oben, rechts) sieht man, daß beim Einsatz des SB-Limiters beide Diskontinuitäten sehr genau aufgelöst werden. Die hervorragende Genauigkeit des SB-Verfahrens im Testfall No.2 läßt sich durch sein stark antidiffusives Verhalten erklären. Je größer der Wert der Limiter-Funktion ist, um so stärker wird der CDS-Konvektionsstromanteil in  $F^{tvd}(U; l)$  gewichtet. Bei einem Gewichtungsfaktor von 2 stimmt der TVD-Konvektionsstrom sogar mit dem Downwind-Strom überein, der (wie im Abschnitt 1.4 gezeigt wurde) antidiffusiv ist. Im Unterschied zur reinen Downwind-Diskretisierung verursacht das SB-Verfahren allerdings keine Oszillationen, da der Limiter den Wert von 2 nicht überall annimmt. Er verläuft aber entlang der oberen Grenze des zulässigen Wertebereichs für ein TVD-Verfahren 2.Ordnung und erlaubt somit soviel „Antidiffusion“ wie nur möglich. Das SB-Verfahren ist daher sehr effektiv bei der Auflösung von diskontinuierlichen Fronten, weil es den vertikalen Verlauf der Lösung durch Antidiffusion aufrechterhält und dabei nicht über die exakte Lösung hinauschießt (infolge der TVD-Eigenschaft). Gleichzeitig ist damit der größte Nachteil des SB-Verfahrens bei Auflösung von *geneigten* Fronten verbunden, da der SB-Limiter dafür sorgt, daß diese Fronten über das zulässige Maß hinaus „aufgestellt“ werden, wie weiter unten gezeigt wird.

Im **Testfall No.1** mit  $\sigma = 0.1$  wurde für alle TVD-Verfahren bis auf MM und L-AGARWAL die empirische Konvergenzordnung von 2 festgestellt. Das MM-Verfahren erreichte auf einem Gitter mit der Maschenweite  $\Delta x = 0.005$  eine etwas niedrigere Konvergenzrate von 1.9, während das L-AGARWAL-Verfahren eine Konvergenzordnung von 3 hat, was mit der

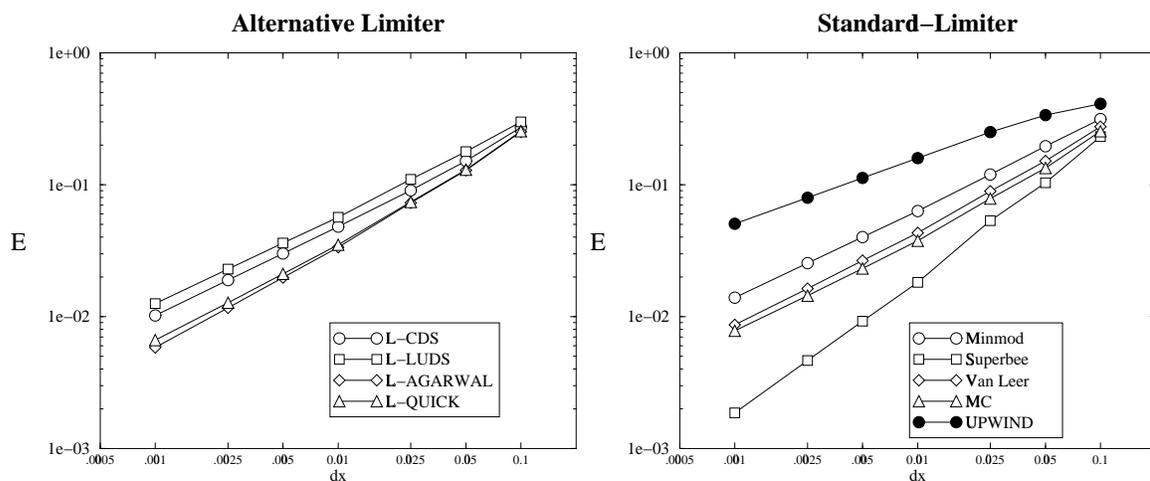
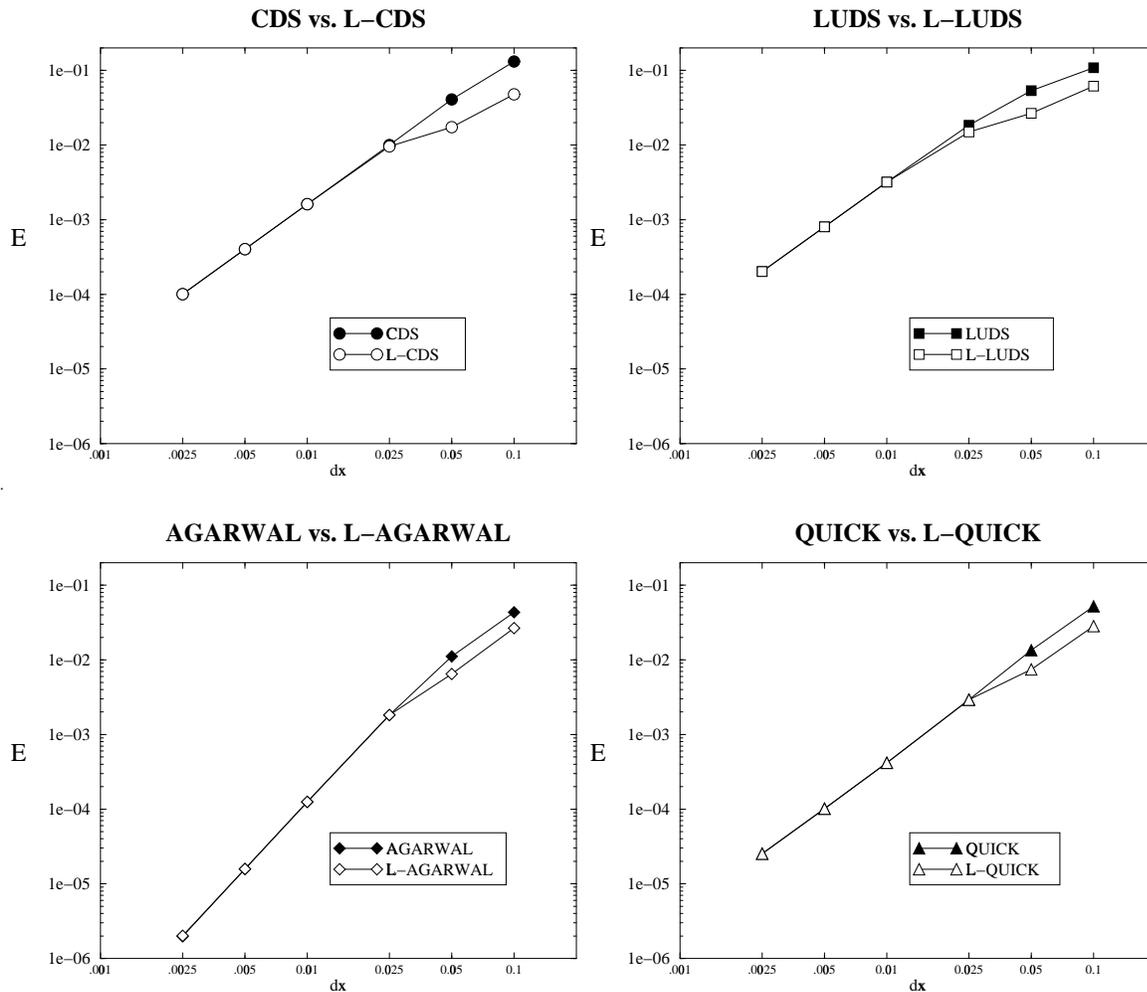


Abbildung 1.43: Test No.2: Vergleich von alternativen und Standard-Limitern untereinander.



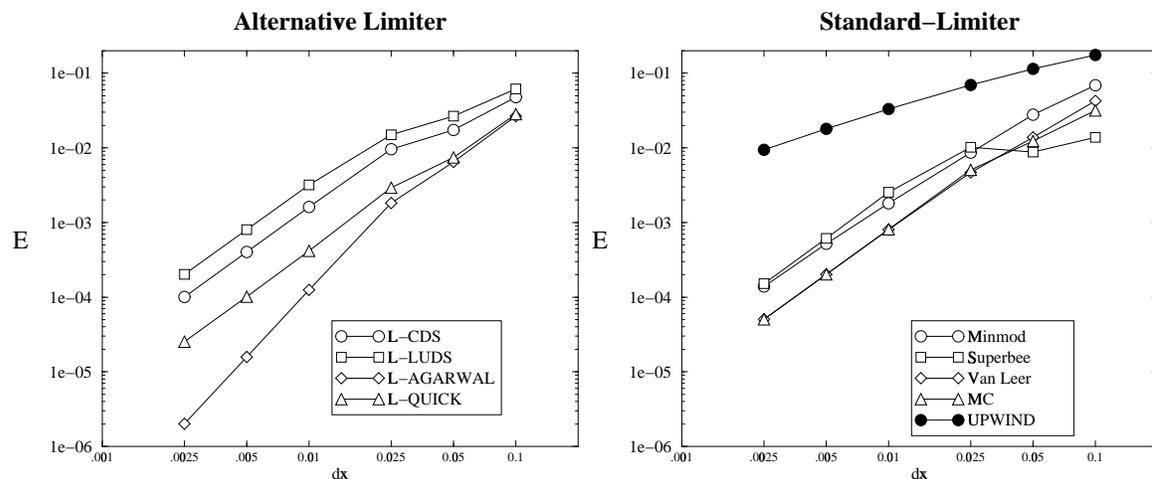
**Abbildung 1.44:** Test No.1 ( $\sigma = 0.1$ ): Vergleich von limitierten und nichtlimitierten Verfahren.

Ordnung des nichtlimitierten AGARWAL-Verfahrens übereinstimmt. In der Abb. 1.44 ist der Fehlerverlauf der limitierten Versionen von vier linearen Verfahren mit dem der nichtlimitierten Originalverfahren verglichen. Auf den groben Gittern schneiden die limitierten Verfahren immer deutlich besser ab, während mit zunehmender Gitterverfeinerung der Unterschied in der Genauigkeit zwischen den limitierten und nichtlimitierten Versionen fast vollständig verschwindet. Das liegt daran, daß sich mit feiner werdenden Gittern die Werte von  $\theta_l$  und  $\theta_r$  dem Wert von 1 nähern, und in der Umgebung von  $\theta = 1$  stimmen die limitierten Versionen mit den nichtlimitierten überein (s. Abb. 1.40).

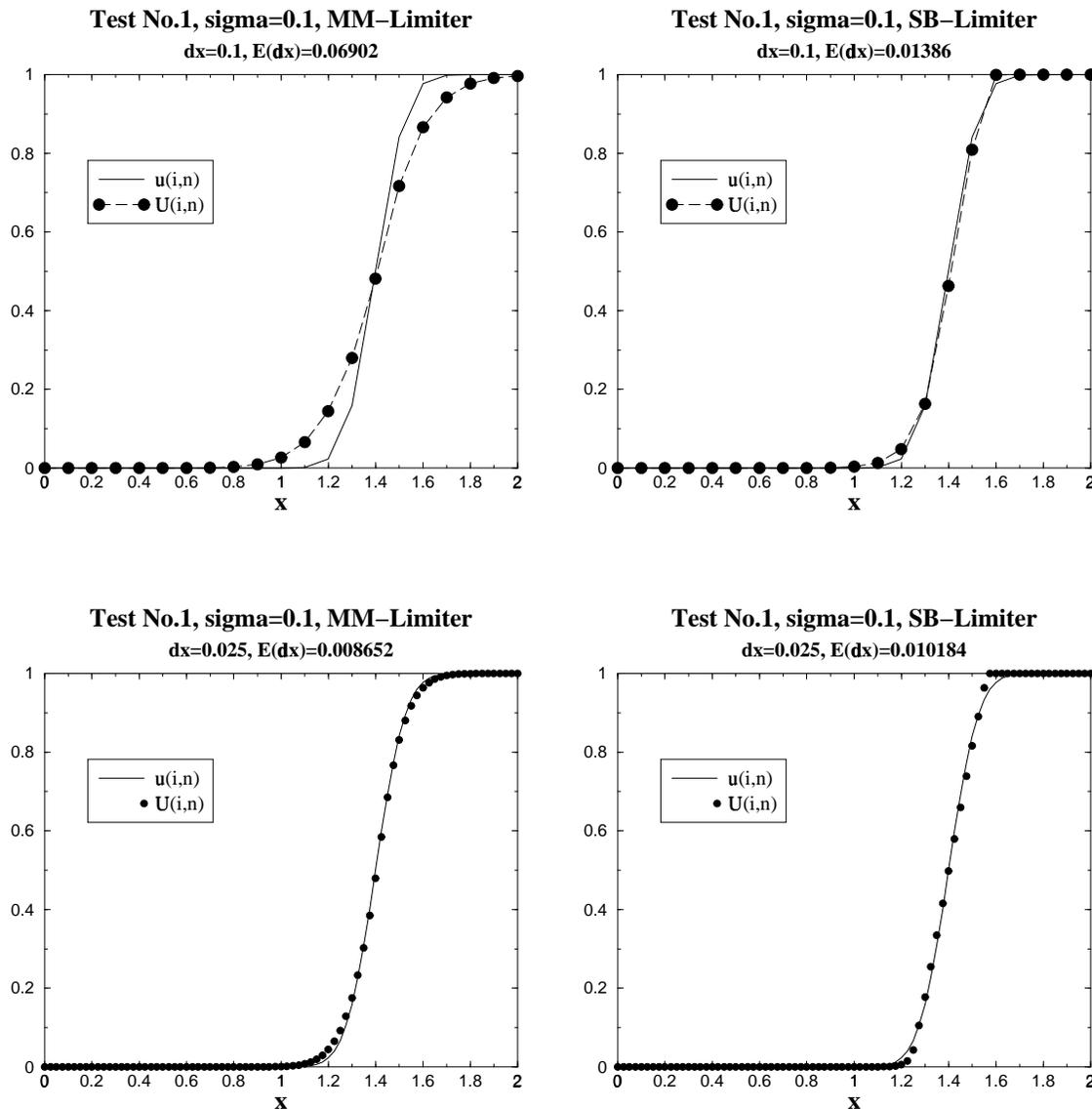
Der Vergleich der vier alternativen Limiter *untereinander* (Abb. 1.45, links) zeigt, daß die mit Abstand beste Genauigkeit von dem AGARWAL-Limiter erzielt wird, gefolgt vom QUICK-Limiter, wobei der Vorsprung des AGARWAL-Limiters bei groben Gittern nicht sehr stark ausfällt. Sehr interessant ist der Verlauf der  $E-\Delta x$ -Kurven der 4 Standard-Limiter

(Abb. 1.45, rechts). Abgesehen vom SB-Verfahren ist das Ergebnis vergleichbar mit dem für das Testbeispiel No. 1. Bei kleinen Werten von  $\Delta x$  liegen die Kurven für VL- und MC-Limiter sehr eng nebeneinander, wobei auf groben Gittern das MC-Verfahren etwas genauer ist. Wesentlich schlechter schneidet der MM-Limiter ab, wobei seine Genauigkeit immer noch viel höher als die des Upwind-Verfahrens erster Ordnung ist. Sehr merkwürdig ist dagegen der Verlauf der  $E$ - $\Delta x$ -Kurve für das SB-Verfahren. Während auf einem Gitter mit der Maschenweite  $\Delta x = 0.1$  der SB-Limiter die beste Lösung von allen Verfahren produziert, nimmt der Fehler bei Verdoppelung der Stützstellenanzahl nur ganz leicht ab, und wächst bei einer weiteren Gitterverfeinerung sogar an.

Die Erklärung für dieses Phänomen liegt im bereits erwähnten stark antidiffusiven Verhalten des SB-Verfahrens. In der Abb. 1.46 ist die mit MM- und SB-Verfahren berechnete Lösung für zwei unterschiedliche Maschenweiten dargestellt. Auf einem Gitter mit der Maschenweite  $\Delta x = 0.1$  verläuft die MM-Lösung wesentlich flacher als die exakte Lösung, während das SB-Profil viel steiler als das MM-Profil ist, was in einer 5fach höheren Genauigkeit resultiert. Die Antidiffusion des SB-Verfahrens hebt somit die Diffusion des Upwind-Verfahrens auf. Auf einem Gitter mit der Maschenweite  $\Delta x = 0.025$  geht die numerische Diffusion des MM-Verfahrens stark zurück, die Lösung ist nun wesentlich genauer, als auf dem groben Gitter. Beim Einsatz des SB-Verfahrens dagegen sorgt dessen starke Antidiffusion dafür, daß die Front in der numerischen Lösung nun steiler verläuft, als die exakte Front (Abb. 1.46, unten rechts). Die Genauigkeit der Lösung des SB-Verfahrens ist sogar schlechter als im MM-Fall. Bei weiteren Gitterverfeinerung geht der Einfluß der Antidiffusion zwar deutlich zurück, die Genauigkeit der SB-Lösung bleibt aber weiterhin niedriger, als die der MM-Lösung (s. Abb. 1.45, rechts).



**Abbildung 1.45:** Test No.1 ( $\sigma = 0.1$ ): Vergleich von alternativen und Standard-Limitern untereinander.



**Abbildung 1.46:** Test No.1 ( $\sigma = 0.1$ ): Vergleich von MM- und SB-Verfahren für zwei unterschiedliche Maschenweiten.

**Testfälle 1 und 2: Zusammenfassung.** Wir haben gezeigt, daß der Einsatz der TVD-Limiter-Funktionen bei der Berechnung des numerischen Konvektionsstroms ermöglicht, oszillationsfreie Lösungen der linearen Konvektionsgleichung zu erhalten. Ein Vergleich der linearen Verfahren höherer Ordnung mit ihren limitierten Versionen hat gezeigt, daß Unterdrückung der Oszillationen durch den Limiter gleichzeitig die Genauigkeit der Lösung auf fast allen untersuchten Gittern verbessert. Die Konvergenzordnung der Verfahren wurde durch die Übergewichtung des Upwind-Stromes in der Nähe der kritischen Stellen somit **nicht negativ beeinflußt**. Im Falle der diskontinuierlichen Lösungsfunktion haben die limitierten Versionen sogar eine höhere Konvergenzrate erzielt, als die linearen Originalverfahren.

ren.

Die mit Abstand besten Ergebnisse wurden beim Einsatz des limitierten AGARWAL-Verfahrens erzielt, das im Falle der glatten Lösungsfunktion eine Konvergenzordnung von 3 hat. Das zweitbeste Ergebnis unter den alternativen Limitern hat das limitierte QUICK-Verfahren gezeigt, wobei im Testfall No.2 seine Genauigkeit fast genauso gut wie die des AGARWAL-Verfahrens war.

Unter den Standard-Limiter-Funktionen haben für beide Testfälle der VL- und der MC-Limiter sehr gut abgeschnitten. Das MM-Verfahren leidet stärker als diese beiden an der numerischen Diffusion und zeigte eine viel schlechtere Auflösung der Fronten auf groben Gittern. Das stark antidiffusive Verhalten des SB-Limiters sorgt dafür, daß er im Falle der glatten Lösungsfunktion auf feinen Gittern die Front zu steil auflöst und dadurch von allen Verfahren die schlechteste Genauigkeit erzielt, obwohl im Testfall No.2 dieser Limiter die besten Lösungen produzierte.

## 1.11 Testfall No.3: nichtmonotone glatte Lösungsfunktion

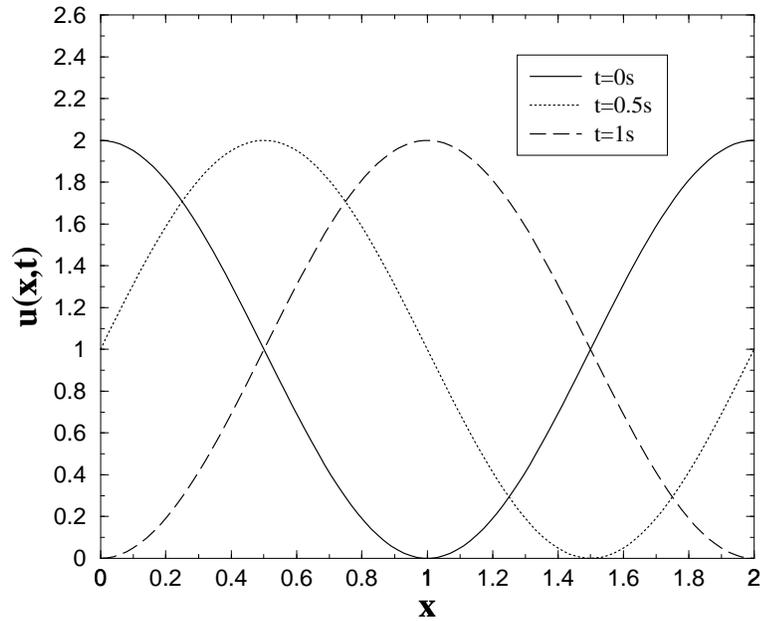
Im letzten Abschnitt haben wir unterschiedliche TVD-Verfahren an 2 Beispielen getestet und festgestellt, daß die limitierten Versionen von linearen Verfahren höherer Ordnung die Oszillationen aus der Lösung eliminieren und gleichzeitig die Genauigkeit der Lösung erhöhen. Im Testfall No. 1 wurde außerdem festgestellt, daß die limitierten Verfahren die Konvergenzordnung der Originalverfahren beibehalten und auf *allen untersuchten Gittern* den Lösungsfehler reduzieren.

Diese Aussagen sind jedoch noch nicht verallgemeinerbar. Sowohl im Testfall No.1 als auch im Testfall No.2 handelte es sich um einen Lösungsprofil mit einer scharfen Front. In diesem Fall sind die Oszillationen in den Lösungen beim Einsatz von linearen Verfahren besonders stark ausgeprägt, so daß der Einsatz von Limitern besonders große Vorteile bringt. Bei realen mehrdimensionalen strömungsmechanischen Problemen können die Lösungen zu verschiedenen Zeitpunkten bzw. an verschiedenen Orten unterschiedlich glatten Verlauf aufweisen. Für ein numerisches Verfahren ist es wichtig, daß es in *allen* möglichen Fällen das Problem effizient löst. In diesem Abschnitt wollen wir daher einen weiteren Testfall mit einer Lösung ohne steile Fronten untersuchen. Wir wählen als Anfangsbedingung

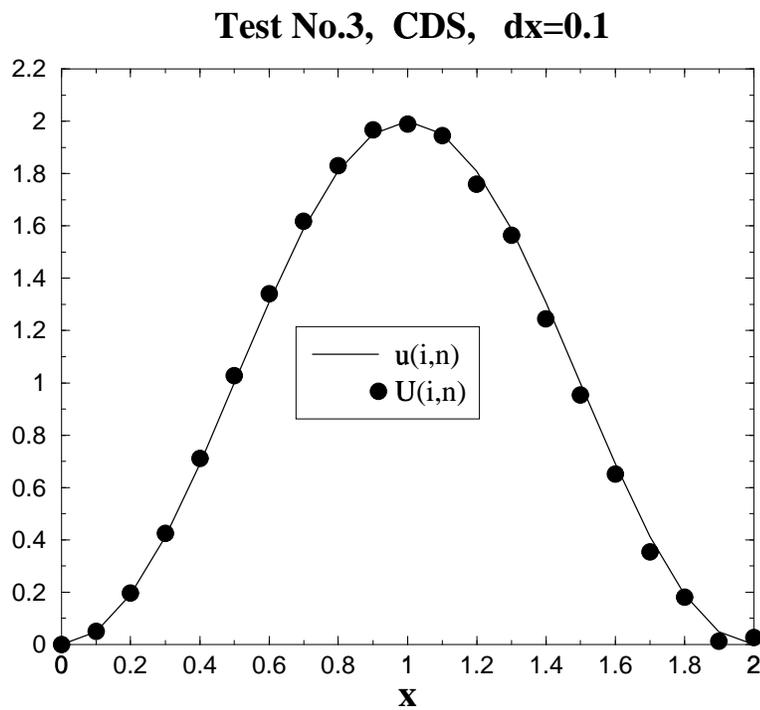
$$u_0(x) = 1 + \cos(\pi x) \quad x \in [0, 2]. \quad (1.198)$$

und als Randbedingung

$$\mu(t) = 1 + \cos(\pi t), \quad t > 0. \quad (1.199)$$



**Abbildung 1.47:** Testbeispiel No.3: Anfangsprofil und Lösungsverlauf zu 2 verschiedenen Zeitpunkten.



**Abbildung 1.48:** Test No.3: CDS-Lösung auf einem Gitter mit 20 Stützstellen.

Die Lösung der linearen Konvektionsgleichung (1.19) mit  $a = 1$  lautet in diesem Fall

$$u(x, t) = 1 + \cos(\pi(x - t)), \quad x \in [0, 2], \quad t \geq 0. \quad (1.200)$$

Das Startprofil der Lösung sowie deren Verlauf zu den Zeitpunkten  $t = 0.5$  und  $t = 1$  sind in der Abb. 1.47 abgebildet. In Abb. 1.48 ist die numerische Lösung mit dem CDS-Verfahren auf einem Gitter mit nur 20 Stützstellen dargestellt. Man sieht, daß die numerische Lösung das exakte Lösungsprofil sehr gut wiedergibt, und keine starken Oszillationen auftreten. Wir wollen nun untersuchen, wie gut die TVD-Verfahren in diesem Testfall im Vergleich zu nichtlimitierten Verfahren abschneiden.

Der neue Testfall ist außerdem aus zwei weiteren Gesichtspunkten interessant. Im Testfall No.1 handelte es sich um eine glatte monotone Lösungsfunktion, die keine lokalen Extrema hatte. Kritische Stellen, die beim Einsatz von linearen Verfahren Oszillationen verursachten, waren diejenigen mit einer starken Krümmung der Lösung. Auf groben Gittern wurde an solchen Stellen bei der Berechnung des TVD-Konvektionsstroms der Upwind-Anteil übergewichtet. Mit zunehmender Gitterverfeinerung nähern sich  $\theta_l$  und  $\theta_r$  dem Wert 1, die limitierten Konvektionsströme stimmen mit den nichtlimitierten überein und alle TVD-Verfahren erreichen eine Approximationsordnung von mindestens 2. Im Testfall No.3 handelt es sich dagegen um eine Lösung mit lokalen Extrema. Unabhängig davon, wie fein das numerische Gitter ist, ist in der Nähe einer lokalen Extremumstelle entweder  $\theta_l$  oder  $\theta_r$  negativ. Und das bedeutet, daß an solchen Stellen ein TVD-Verfahren maximal erster Ordnung genau ist. Der Testfall No.3 gibt uns somit die Möglichkeit zu untersuchen, inwieweit die niedrigere Approximationsordnung an einzelnen Stellen die Konvergenzordnung des Verfahrens im gesamten Lösungsgebiet beeinflusst.

Ein weiterer wichtiger Aspekt betrifft die Behandlung des Volumenelements mit der Nummer 1. Alle TVD-Verfahren benötigen zur Berechnung des Konvektionsstromes  $F(U; 1/2)$  den Wert der Gitterfunktion an der Stelle  $x_{-1}$ , die außerhalb des Berechnungsgebietes liegt. Bisher haben wir zwei Testfälle mit quasi-stationären Randbedingungen untersucht (im Test No.2 ist  $\mu(t) \equiv 0$  und für Test No.1 gilt mit einer sehr hohen Genauigkeit  $\mu(t) \approx 0$ ). Wir konnten daher das entstandene Problem dadurch umgehen, daß wir  $U_{-1} = U_0$  gesetzt haben. Im Test No.3 handelt es sich dagegen um eine *instationäre* Randbedingung. In diesem Fall kann die Annahme  $U_{-1} = U_0$  das Konvergenzverhalten des Verfahrens im gesamten Lösungsgebiet negativ beeinflussen.

Als erstes untersuchen wir das Problem mit der Behandlung des fiktiven Wertes  $U_{-1}$ , da dieses nicht nur bei allen TVD-Verfahren, sondern auch bei der Berechnung des Konvektionsstromes  $F(U; 1/2)$  mit LUDS-, AGARWAL- und QUICK-Verfahren entsteht. Wir wissen, daß die analytische Lösung mit der Konvektionsgeschwindigkeit  $a$  entlang der  $x$ -Achse propagiert. Wäre das Lösungsgebiet räumlich unbegrenzt, so würde der Lösungs-

wert  $u(x_{-1}, t_{n+1})$  mit dem Wert  $u(0, t_{n+1} + \Delta x/a)$  übereinstimmen. Tatsächlich gilt wegen  $u(x, t) = u_0(x - at)$ :

$$\begin{aligned} u(x_{-1}, t_{n+1}) &= u(-\Delta x, t_{n+1}) = u_0(-\Delta x - at_{n+1}), \\ u(0, t_{n+1} + \Delta x/a) &= u_0(0 - a \cdot (t_{n+1} + \Delta x/a)) = u_0(-\Delta x - at_{n+1}). \end{aligned}$$

Wir können deshalb den „genauen“ Wert  $U_{-1}$  folgendermaßen berechnen:

$$U_{-1} = u(0, t_{n+1} + \Delta x/a) = u_0(t_{n+1} + \Delta x/a).$$

Wir werden diese Formel zur Berechnung von  $U_{-1}$  „Variante A“ nennen. Diese Vorgehensweise setzt allerdings voraus, daß wir die analytische Lösung bereits kennen. Sie kann daher nicht auf andere Gleichungen verallgemeinert werden. Wir werden diese Variante deshalb nur als Referenzlösung betrachten.

Ist die analytische Lösung der Differentialgleichung nicht bekannt, so muß der Wert  $U_{-1}$  aus den Werten der Gitterfunktion approximiert werden, die innerhalb des Berechnungsgebietes liegen. Die einfachste Möglichkeit wurde bereits angesprochen. Wir nennen diese Approximation „Variante 1“:

$$U_{-1} = U_0. \quad (1.201)$$

Unter der Annahme, daß die Lösungsfunktion eine glatte Fortsetzung außerhalb des Berechnungsgebietes hat, hat Variante 1 die Approximationsordnung von 1. Die Approximationsordnung 2 kann erreicht werden, wenn man davon ausgeht, daß der Punkt  $(x_{-1}, U_{-1})$  auf einer Geraden liegt, die durch die Stellen  $(x_0, U_0)$  und  $(x_1, U_1)$  geht. Die entsprechende Formel („Variante 2“) lautet:

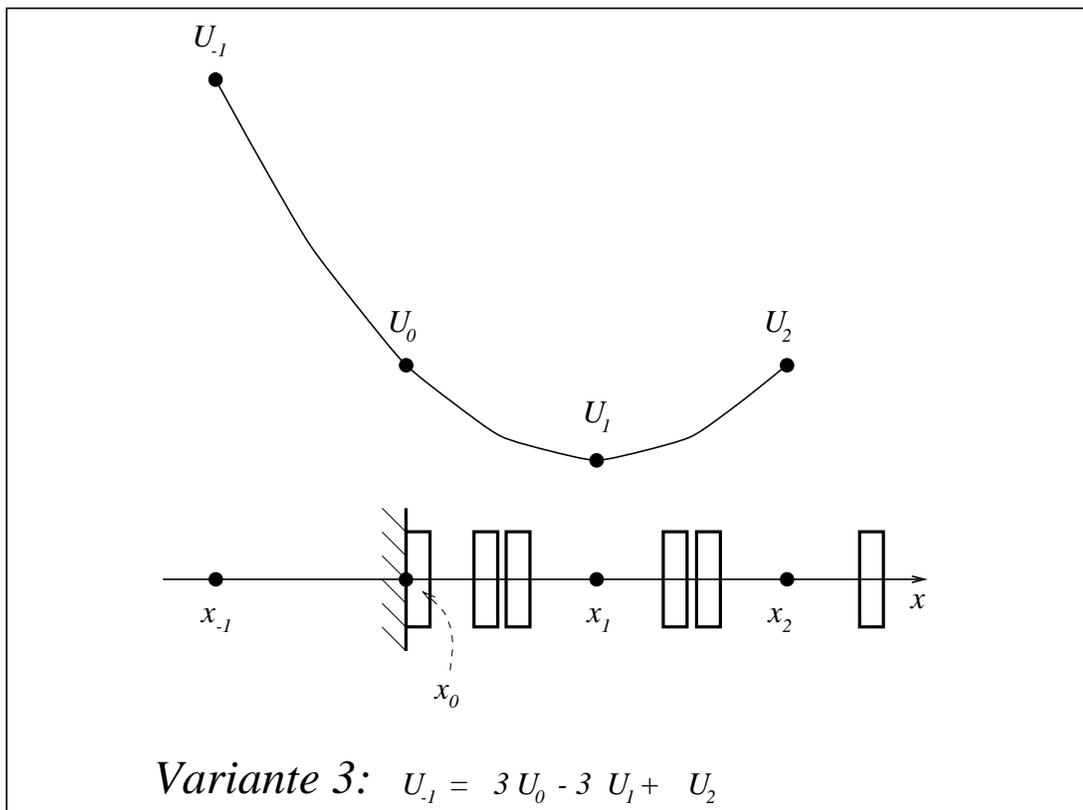
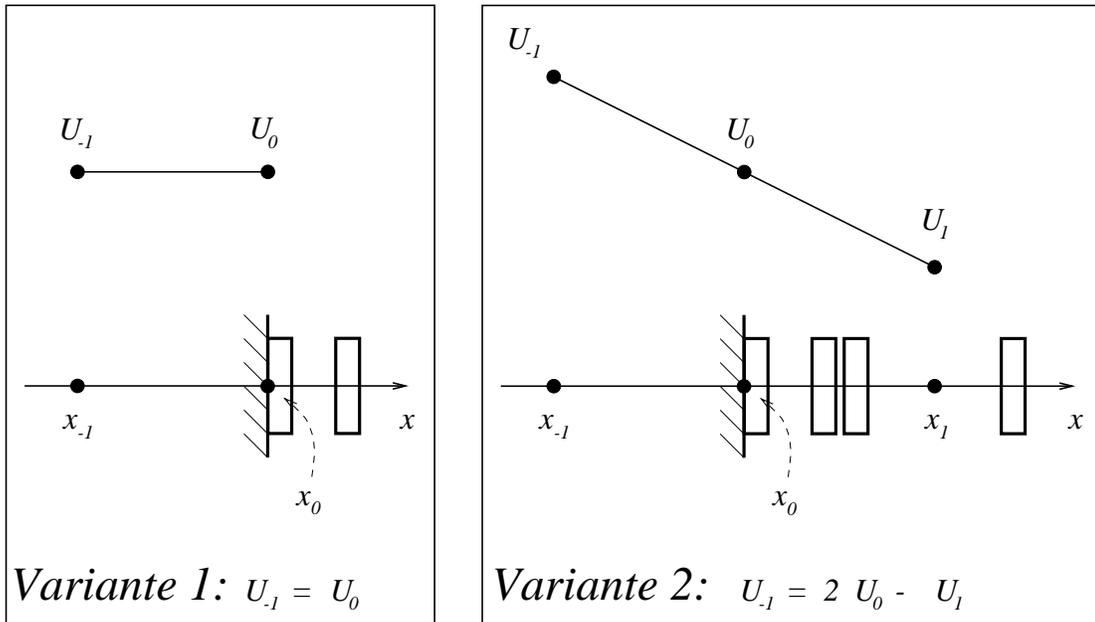
$$U_{-1} = 2U_0 - U_1. \quad (1.202)$$

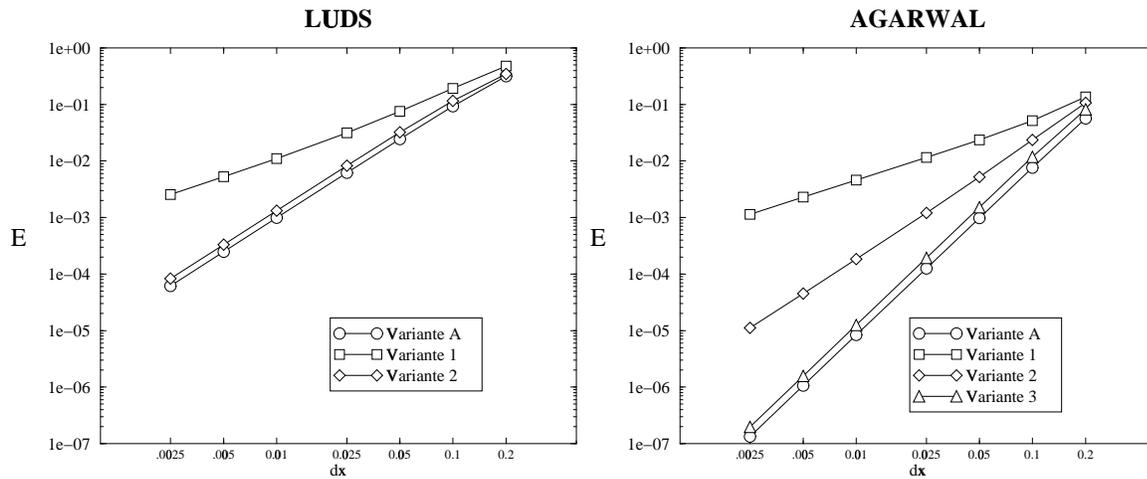
Schließlich können wir eine Parabel durch drei Punkte  $(x_0, U_0)$ ,  $(x_1, U_1)$  und  $(x_2, U_2)$  legen und den Wert ausrechnen, den diese Parabel an der Stelle  $x_{-1}$  annehmen wird („Variante 3“):

$$U_{-1} = 3U_0 - 3U_1 + U_2. \quad (1.203)$$

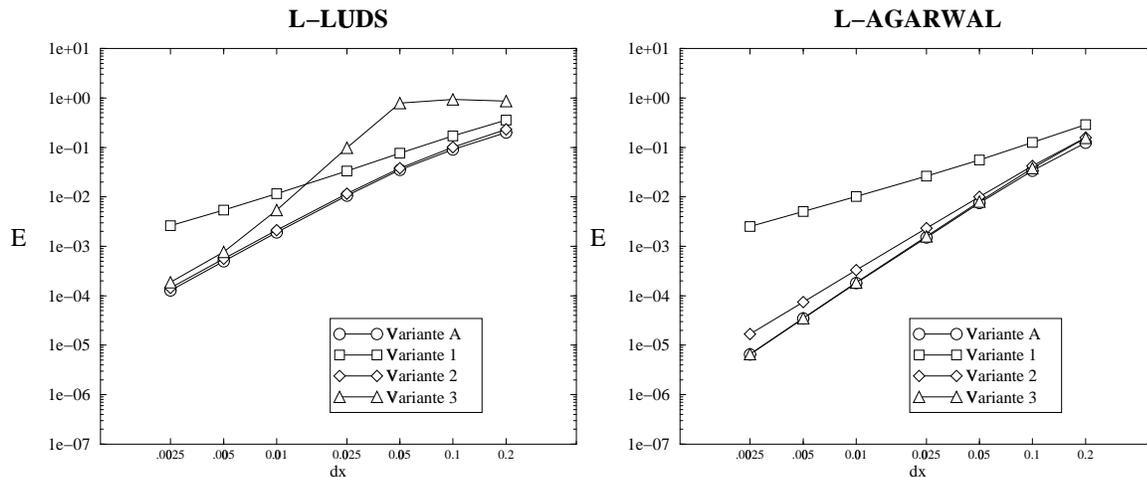
Die Variante 3 hat entsprechend die Approximationsordnung von 3. In Abb. 1.49 sind alle 3 Varianten der Approximation von  $U_{-1}$  graphisch veranschaulicht.

Wir werden zunächst untersuchen, welchen Einfluß die Berechnung des fiktiven Wertes  $U_{-1}$  auf das Konvergenzverhalten von linearen Diskretisierungsverfahren hat. Wir wählen dazu

Abbildung 1.49: Drei Varianten zur Berechnung von  $U_{-1}$ .

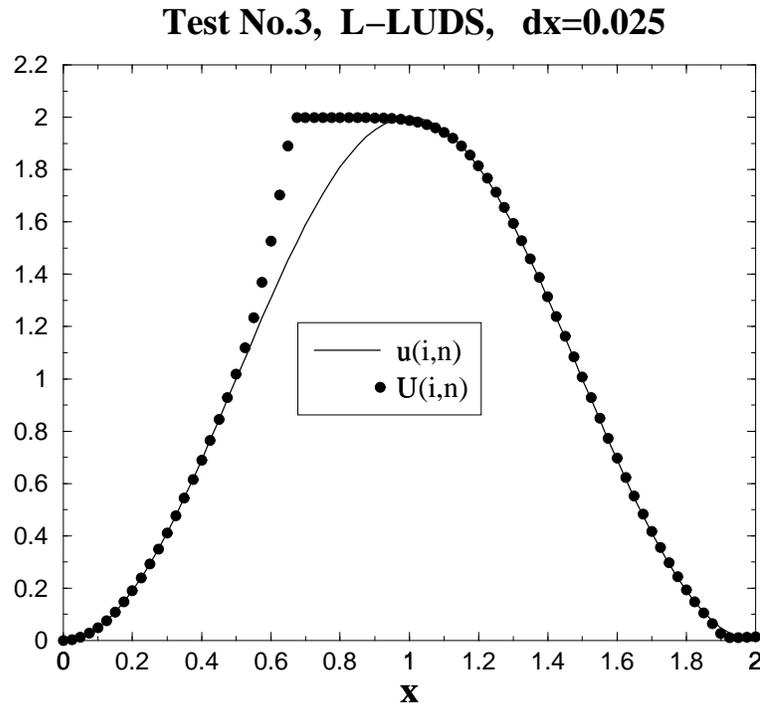


**Abbildung 1.50:** Test No.3: Einfluß der Approximation von  $U_{-1}$  auf die Konvergenzordnung von LUDS und AGARWAL.



**Abbildung 1.51:** Test No.3: Einfluß der Approximation von  $U_{-1}$  auf die Konvergenzordnung von L-LUDS und L-AGARWAL.

ein Verfahren 2. Ordnung (LUDS) und ein Verfahren 3. Ordnung (AGARWAL). In der Abb. 1.50 (links) sind die  $E$ - $dx$ -Kurven für das LUDS-Verfahren mit unterschiedlichen Approximationsvarianten der Größe  $U_{-1}$  dargestellt. Während bei der Variante A und Variante 2 die Konvergenzordnung von 2 erreicht wird, führt die Variante 1 nur zu einer Konvergenzordnung von 1, was in einer viel schlechteren Genauigkeit im gesamten Lösungsbereich resultiert. Einsatz der Variante 3 (ohne Grafik) erhöht nicht die Konvergenzrate des Verfahrens, verkleinert allerdings den absoluten Betrag des numerischen Fehlers im Vergleich zur Variante 2.



**Abbildung 1.52:** Test No.3: L-LUDS-Lösung (Var.3) auf einem Gitter mit 80 Stützstellen.

Auch beim Einsatz des AGARWAL-Verfahrens hat die Approximation des Wertes  $U_{-1}$  einen entscheidenden Einfluß auf die Konvergenzordnung des Verfahrens (Abb. 1.50, rechts). Während bei der Variante 3 die Konvergenzordnung von 3 erreicht wird, führen die Varianten 1 und 2 entsprechend zu Konvergenzordnungen von 1 und 2. Der absolute Betrag des numerischen Fehlers ist bei der Variante 3 nur unwesentlich höher als bei Variante A.

In der Abb. 1.51 (links) sind die  $E-dx$ -Kurven für das *limitierte* LUDS-Verfahren dargestellt. Für die Varianten A und 2 ist der Verlauf der  $E-dx$ -Kurve fast identisch. In beiden Fällen wird die Konvergenzordnung von 2 erzielt. Die niedrigere Approximationsordnung des limitierten Verfahrens an den Extremalstellen der Lösung beeinflusst also *nicht* die Konvergenzordnung des Verfahrens im gesamten Bereich. Bei der Variante 1 wird wie im nicht-limitierten Fall lediglich die Konvergenzordnung von 1 erreicht. Sehr merkwürdig sieht der Verlauf der  $E-dx$ -Kurve bei der Variante 3 aus. In der Abb. 1.52 ist die numerische Lösung mit dem L-LUDS-Verfahren und Variante 3 auf einem Gitter mit 80 Stützstellen dargestellt. Man sieht daß die numerische Lösung innerhalb eines größeren Intervalls von der exakten Lösung sehr stark abweicht. Um den Grund für dieses Verhalten der Lösung zu verstehen, müssen wir untersuchen, welchen Einfluß die Approximation des Wertes  $U_{-1}$  auf die Berechnung des numerischen Konvektionsstromes  $F^{tvd}(U; 1/2)$  hat. Die Variante 1 (1.201) führt wegen  $U_0 - U_{-1} = 0$  zu  $\phi_{1/2} = 0$  und es gilt  $F^{tvd}(U; 1/2) \equiv F^{up}(U; 1/2)$ . An der

Stelle  $x_{1/2}$  wird bei der Variante 1 immer der Upwind-Strom eingesetzt. Bei der Variante 2 (1.202) gilt  $U_0 - U_{-1} = U_1 - U_0$ , daraus folgt  $F^{tvd}(U; 1/2) \equiv F^{cds}(U; 1/2)$ . An der Stelle  $x_{1/2}$  wird somit bei der Variante 2 immer der CDS-Strom eingesetzt. Beide Varianten sind physikalisch sinnvoll.

Der Einsatz der Approximationsvariante 3 (1.203) beim LUDS-Verfahren kann dagegen zu unerwarteten Ergebnissen führen. Es gilt

$$\begin{aligned}\theta_{1/2} &= \frac{U_0 - U_{-1}}{U_1 - U_0} = \frac{-2U_0 + 3U_1 - U_2}{U_1 - U_0} \\ &= 2 + \frac{U_1 - U_2}{U_1 - U_0}.\end{aligned}\quad (1.204)$$

Ist  $U_1$  die Stelle des lokalen Maximums der Lösung, so gilt  $\theta_{1/2} > 2$  und  $\phi^{luds}(\theta_{1/2}) = 2$ ; der L-LUDS-Konvektionsstrom an der Stelle  $x_{i-1/2}$  stimmt dann mit dem Downwindstrom  $F^{down}(U; 1/2)$  überein. Gleichzeitig gilt  $\theta_{3/2} < 0$  und somit  $\phi^{luds}(\theta_{3/2}) = 0$ ; der L-LUDS-Konvektionsstrom an der Stelle  $x_{i+1/2}$  stimmt dann mit dem Upwindstrom  $F^{up}(U; 3/2)$  überein. Wegen  $F^{down}(U; 1/2) = F^{up}(U; 3/2)$  nimmt die Diskretisierung für das erste Volumenelement folgende Form ein:

$$U_1 = \bar{U}_1$$

In diesem Fall hat die Randbedingung keinen Einfluß auf die Lösung im Inneren des Berechnungsgebietes. Im Testfall No. 3 befindet sich die Maximumstelle der Lösung zum Zeitpunkt  $t_0 = 0$  in  $x = x_0$ . Nach einigen Zeitschritten verlagert sie sich nach  $x_1$ , so daß genau der eben besprochene Fall eintritt:  $U_1$  wird zum lokalen Maximum der Lösung. Solange  $\phi_{1/2} = 2$  und  $\phi_{3/2} = 0$  gilt, bleibt die numerische Lösung von der analytischen abgekoppelt - ein Effekt, der in der Abb. 1.52 deutlich zu sehen ist.

In der Abb. 1.51 (rechts) sind die  $E-dx$ -Kurven für das limitierte AGARWAL-Verfahren dargestellt. Analog zum nichtlimitierten AGARWAL-Verfahren führen die Varianten 1 und 2 zu Konvergenzordnungen von 1 und 2. Die  $E-dx$ -Kurven für Varianten A und 3 stimmen überein. Bei der Variante 3 des L-AGARWAL Verfahrens findet keine Entkoppelung der numerischen Lösung von der Randbedingung statt, da  $\phi^{agar}(\theta_{1/2}) \equiv 2$  erst für  $\theta_{1/2} \geq 5$  gilt. Daß dieser Wert im Test No.3 nicht erreicht wird, bedeutet nicht, daß die Variante 3 mit dem L-AGARWAL-Verfahren uneingeschränkt eingesetzt werden darf. Man sollte auch beim L-AGARWAL-Verfahren grundsätzlich die Variante 2 verwenden, zumal der Unterschied im Fehlerverlauf zwischen den Varianten 2 und 3 nicht so dramatisch ausfällt. Das ist auf den ersten Blick überraschend, da beim nichtlimitierten AGARWAL-Verfahren (Abb. 1.50, rechts) die Variante 3 zu wesentlich niedrigeren Fehlern als Variante 2 geführt hat. Daß der Unterschied zwischen diesen beiden Varianten beim limitierten Verfahren nicht so stark

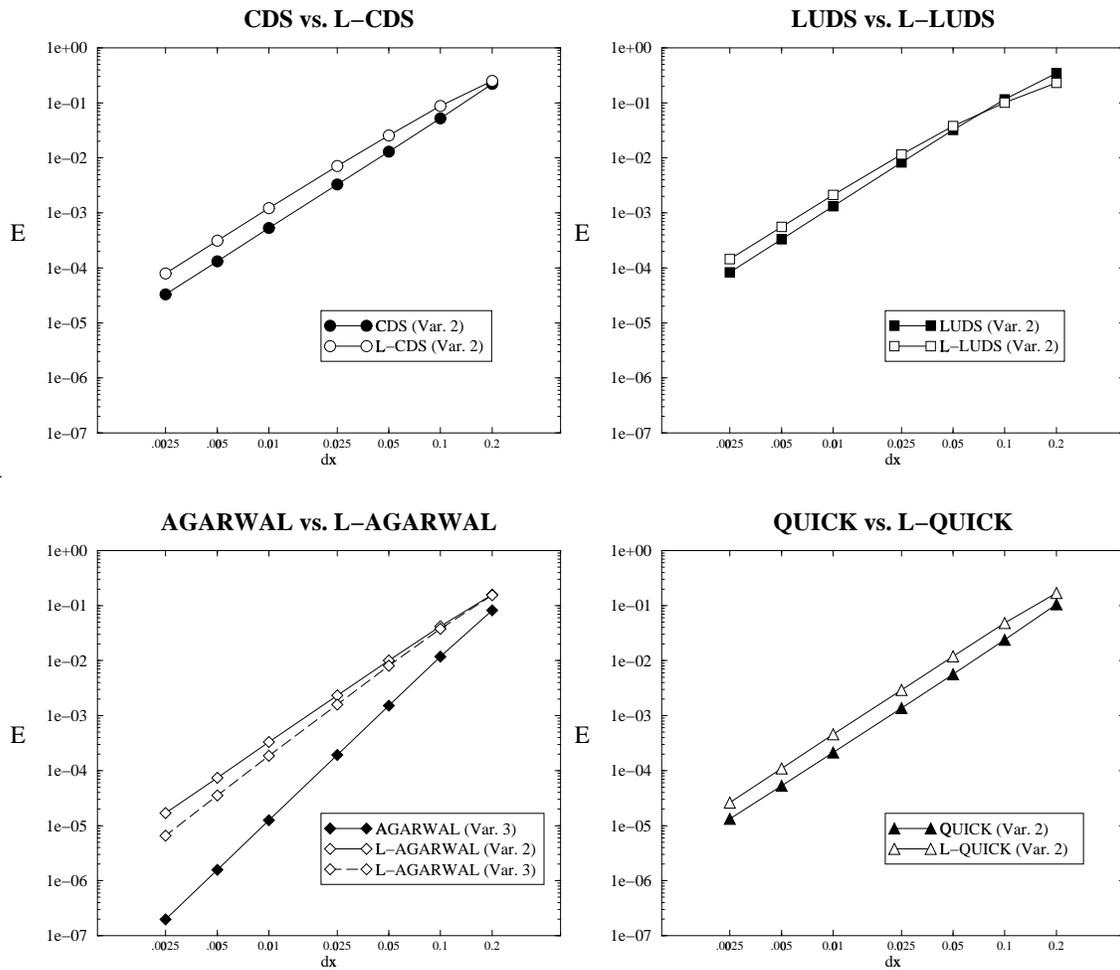


Abbildung 1.53: Test No.3: Vergleich von limitierten und nichtlimitierten Verfahren.

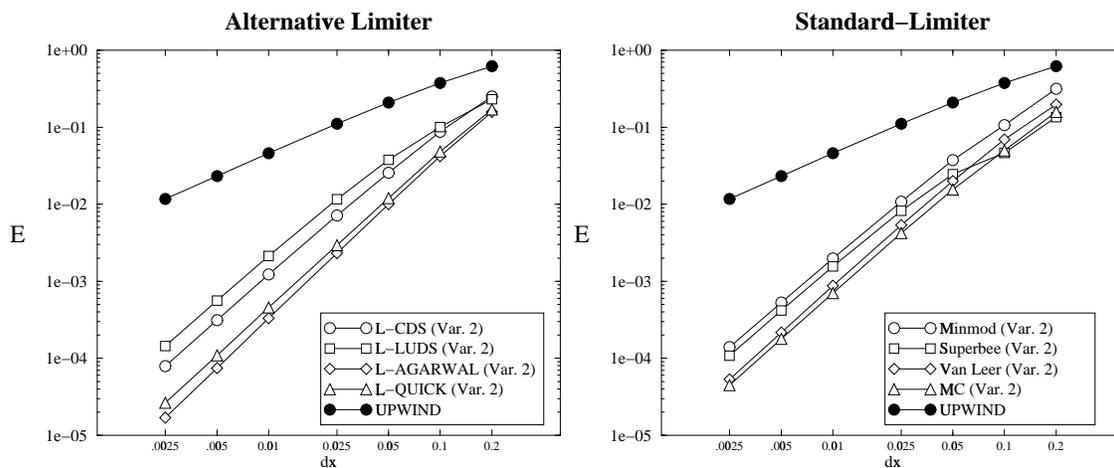


Abbildung 1.54: Test No.3: Vergleich von alternativen und Standard-Limitern untereinander.

ausfällt, liegt daran, daß beim L-AGARWAL mit Variante 3 nur eine Konvergenzordnung von 2.4 erreicht wird. Das bedeutet, daß beim L-AGRAWAL-Verfahren (im Unterschied zu L-LUDS) die niedrigere Approximationsordnung an den Extremalstellen die Konvergenzordnung des Verfahrens im gesamten Bereich negativ beeinflusst; sie geht von 3 auf 2.4 zurück.

In der Abb. 1.53 ist der Fehlerverlauf der limitierten Versionen von vier linearen Verfahren mit dem der nichtlimitierten Originalverfahren verglichen. Im Unterschied zu Testbeispielen No. 1 und 2 haben die limitierten Versionen von allen 4 Verfahren eine niedrigere Genauigkeit als die Originalverfahren. Bis auf L-AGARWAL fällt dieser Unterschied jedoch nicht sehr stark aus. In allen 4 Fällen konnte die Konvergenzordnung von 2 erreicht werden, so daß alle limitierten Verfahren immer noch wesentlich genauer als die UPWIND-Diskretisierung erster Ordnung sind (Abb. 1.54, links). Der Vergleich der alternativen Limiter *untereinander* zeigt, daß wie in Testfällen 1 und 2, die besten Ergebnisse mit L-AGARWAL erzielt werden, gefolgt von L-QUICK. Unter den Standard-Limitern (Abb. 1.54, rechts) schneidet der MC-Limiter am besten ab. Die schlechtesten Ergebnisse produziert der Minmod-Limiter. Das Superbee-Verfahren verhält sich ähnlich wie im Test No.1. Auf dem groben Gitter gehört seine Lösung zu den besten, danach folgt die Übergangsphase mit einer niedrigen Konvergenzrate infolge der starken Kompression der Lösung. Im Grenzwert  $\Delta x \rightarrow 0$  wird die Konvergenzordnung von 2 jedoch erreicht.

**Zusammenfassung.** In diesem Abschnitt haben wir 3 Fragen untersucht. Wir haben gesehen, daß bei instationären Randbedingungen die Approximation des Konvektionsstromes  $F(U; 1/2)$  einen entscheidenden Einfluß auf die Genauigkeit der Lösung im gesamten Lösungsbereich hat. Bei allen Verfahren, die den Wert von  $U_{-1}$  zur Berechnung von  $F(U; 1/2)$  benötigen, kann mit der bisher verwendeten Approximation  $U_{-1} := U_0$  lediglich die Konvergenzordnung von 1 erreicht werden. Es wurden alternative Varianten vorgeschlagen und deren Einfluß auf die Konvergenzordnung der Verfahren untersucht.

Ein Vergleich zwischen den limitierten Versionen der linearen Verfahren und den nichtlimitierten Originalverfahren für eine nichtmonotone glatte Lösungsfunktion hat gezeigt, daß die limitierten Verfahren auf fast allen Gittern eine schlechtere Genauigkeit der Lösung erzielen. Sie sind aber trotzdem wesentlich genauer als das UPWIND-Verfahren erster Ordnung, was zusammen mit der Garantie einer physikalisch sinnvollen Lösung für alle Lösungsfunktionen und dem hervorragenden Konvergenzverhalten in Fällen mit steilen Lösungsprofilen **für den universalen Einsatz** der limitierten Verfahren zur Lösung von konvektionsdominanten Problemen spricht.

Bis auf das L-AGARWAL-Verfahren haben die limitierten Verfahren die Konvergenzordnung der Originalverfahren beibehalten, obwohl die Lösungsfunktion nicht-monoton gewe-

sen ist. Das L-AGARWAL-Verfahren erzielt lediglich eine Konvergenzordnung von 2.4, die niedriger ist als die Ordnung des Original-Verfahrens. Selbst diese Konvergenzordnung kann nur bei einem Einsatz zur Berechnung von  $U_{-1}$  erzielt werden (Variante 3), der für andere Testfälle nichtkontrollierbare Ergebnisse produzieren kann. Es wird deshalb vorgeschlagen, das L-AGARWAL-Verfahren ausschließlich mit der Variante 2 der Berechnung von  $U_{-1}$  einzusetzen. Obwohl in diesem Fall die Konvergenzordnung des Verfahrens weiter auf 2 absinkt, erzielt dieses Verfahren die genauesten Lösungen von allen 8 TVD-Diskretisierungen (s. Abb. 1.54).

**Vereinbarung.** In diesem Kapitel haben wir einen Weg zur Konstruktion von TVD-Verfahren für die lineare Konvektionsgleichung vorgestellt und 8 Varianten dieser Verfahren untersucht. In einem weiteren Kapitel werden wir die hier vorgestellte Vorgehensweise auch für nichtlineare Gleichungen anwenden. Es wird sich zeigen, daß auch in diesem Fall die hier diskutierten Limiter-Funktionen eingesetzt werden können. Da in allen 3 bisher diskutierten Testfällen die CDS und LUDS-Limiter wesentlich schlechter als QUICK- und AGARWAL-Limiter, und MM- und VL-Limiter schlechter als MC-Limiter abgeschnitten haben, werden die CDS-, LUDS-, MM- und VL-Limiter nicht mehr berücksichtigt.

## Kapitel 2

# Lineare Konvektionsgleichung mit variabler Konvektionsgeschwindigkeit

## 2.1 Analytische Lösung der Gleichung

Im letzten Kapitel haben wir die Herleitung von TVD-Verfahren für die eindimensionale lineare Konvektionsgleichung mit einer konstanten positiven Konvektionsgeschwindigkeit in aller Ausführlichkeit präsentiert. In den meisten praxisrelevanten Fällen handelt es sich allerdings um mehrdimensionale Probleme. Die Konvektionsgeschwindigkeit ist dabei eine Funktion der Zeit und der Ortskoordinaten und ist nicht nur variabel, sondern wechselt oft auch das Vorzeichen. Schließlich kann die Geschwindigkeit selbst als eine unbekannte Größe in einer Gleichung auftreten (man denke an die Navier-Stokes-Gleichungen). In diesem Falle ist die zu lösende Gleichung nicht linear.

Unser Ziel ist es daher, die im letzten Kapitel vorgestellte Vorgehensweise zur Diskretisierung der eindimensionalen linearen Konvektionsgleichung mit konstanter Konvektionsgeschwindigkeit auf mehrdimensionale lineare und nichtlineare Probleme mit variabler Konvektionsgeschwindigkeit zu übertragen. Als erster Zwischenschritt wollen wir in diesem Kapitel auf die Annahme einer konstanten Konvektionsgeschwindigkeit verzichten.

Ob die Konvektionsgeschwindigkeit nur eine Funktion der Ortskoordinate  $x$  ist, oder von beiden Variablen  $x$  und  $t$  abhängt, ist nicht entscheidend. Sowohl die Herleitung der analytischen Lösung als auch der Aufbau der numerischen Verfahren unterscheidet sich in beiden Fällen nur unwesentlich. Wir werden daher annehmen, daß die Konvektionsgeschwindigkeit nur eine Funktion der Ortskoordinate  $x$  ist:  $a = a(x)$ . Die resultierende Gleichung lautet:

$$\frac{\partial}{\partial t}u(x, t) + \frac{\partial}{\partial x}(a(x)u(x, t)) = 0 \quad (2.1)$$

Die Anfangsbedingung sei durch die folgende Gleichung definiert:

$$u(x, t = 0) = u_0(x), \quad x \in (-\infty, +\infty). \quad (2.2)$$

Wie im Falle einer konstanten Geschwindigkeit können nun Charakteristiken der Gleichung (2.1) als solche Kurven definiert werden, die der gewöhnlichen Differentialgleichung

$$x'(t) = a(x) \quad (2.3)$$

mit der Anfangsbedingung

$$x(0) = x_0 \quad (2.4)$$

genügen. Dabei kann  $x_0$  ein beliebiger Wert aus  $(-\infty, +\infty)$  sein. Er definiert die Startposition der charakteristischen Kurve auf der  $x$ -Achse.

Wir können die Konvektionsgleichung (2.1) folgendermaßen umschreiben:

$$u_t + a(x)u_x = -a'(x)u. \quad (2.5)$$

Für die Ableitung der Lösung  $u$  entlang einer Charakteristik gilt daher

$$\begin{aligned} \frac{d}{dt}u(x(t), t) &= \frac{\partial}{\partial t}u(x(t), t) + \frac{\partial}{\partial x}u(x(t), t) \cdot x'(t) \\ &= u_t + a(x)u_x = -a'(x)u. \end{aligned} \quad (2.6)$$

Entlang einer Charakteristik  $x = x(t)$ , die vom Punkt  $x_0$  startet, kann die Lösungsfunktion also aus der gewöhnlichen Differentialgleichung

$$\frac{d}{dt}u(x(t), t) = -a'(x(t))u(x(t), t) \quad (2.7)$$

mit der Anfangsbedingung

$$u(x(0), 0) \equiv u(x_0, 0) = u_0(x_0) \quad (2.8)$$

berechnet werden.

Ist die Funktion  $a(x)$  im gesamten Lösungsbereich stetig differenzierbar, so können die gewöhnlichen Differentialgleichungen (2.3) und (2.7) eindeutig gelöst werden, und man kann daraus die Funktion  $u(x, t)$  für alle Wertepaare  $(x, t)$  bestimmen. Ist das Anfangsprofil  $u_0(x)$  ebenfalls eine stetig differenzierbare Funktion von  $x$ , so stellt die Funktion  $u(x, t)$  eine klassische Lösung der Differentialgleichung (2.1) dar, ansonsten kann sie als *schwache Lösung* dieser Gleichung interpretiert werden (s. Abschnitt 1.1).

Wir wollen nun das Charakteristiken-Verfahren zur Bestimmung der analytischen Lösung der Konvektionsgleichung an einem Beispiel veranschaulichen, und dabei gleichzeitig zwei neue Testfälle für die numerischen Verfahren definieren. Wir gehen von einer linearen Konvektionsgeschwindigkeit  $a(x) = x$  aus, und suchen die Lösung im Bereich  $(x, t) \in [0, 2] \times [0, +\infty)$ . Die Anfangsbedingung sei entsprechend durch

$$u(x, t = 0) = u_0(x), \quad x \in [0, 2] \quad (2.9)$$

definiert. Eine genaue Spezifikation der Funktion  $u_0(x)$  ist an dieser Stelle noch nicht notwendig. Im Unterschied zur linearen Konvektionsgleichung mit konstanter Geschwindigkeit werden diesmal keine Randbedingungen benötigt. Das liegt daran, daß die Konvektionsgeschwindigkeit  $a(x)$  am linken Rand gleich Null ist, und die Lösung am linken Rand eindeutig durch die *Anfangsbedingung* am Rand definiert ist.

Nun sei  $x_0$  ein beliebiger Punkt aus dem Intervall  $[0, 2]$ . Die allgemeine Lösung der charakteristischen Gleichung

$$x'(t) = x \quad (2.10)$$

lautet

$$x(t) = Ce^t. \quad (2.11)$$

Aus der Anfangsbedingung (2.4) folgt  $C = x_0$ , und somit gilt für die Charakteristik, die vom Punkt  $x_0$  startet:

$$x(t) = x_0 e^t. \quad (2.12)$$

Entlang einer Charakteristik gilt für die Lösungsfunktion

$$\frac{d}{dt}u(x(t), t) = -a'(x(t))u(x(t), t) = -u(x(t), t) \quad (2.13)$$

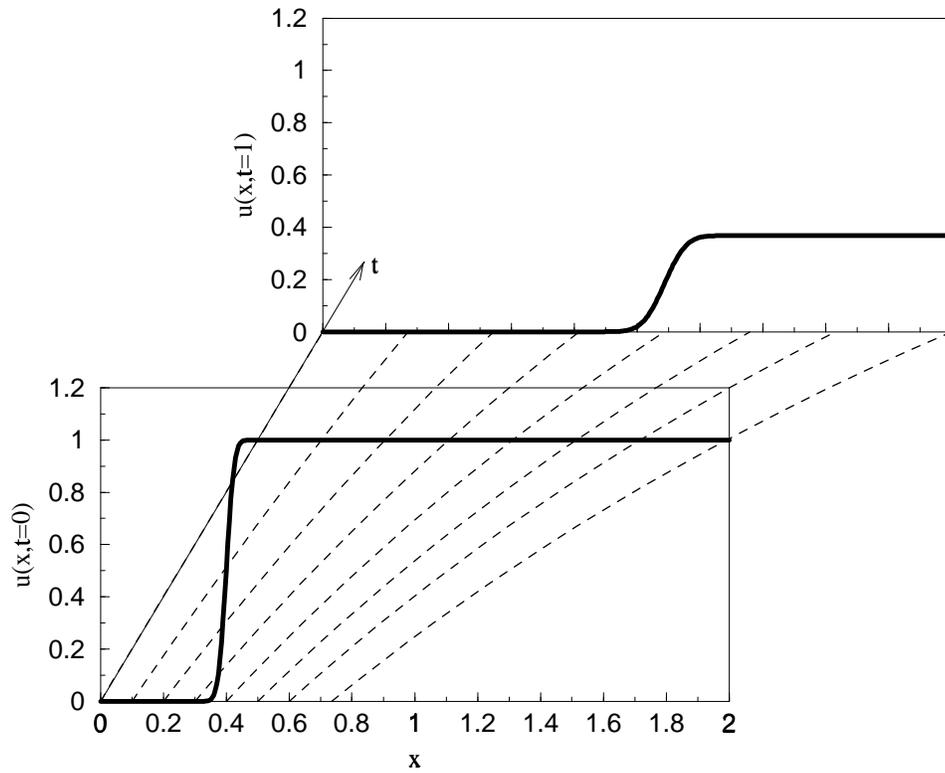
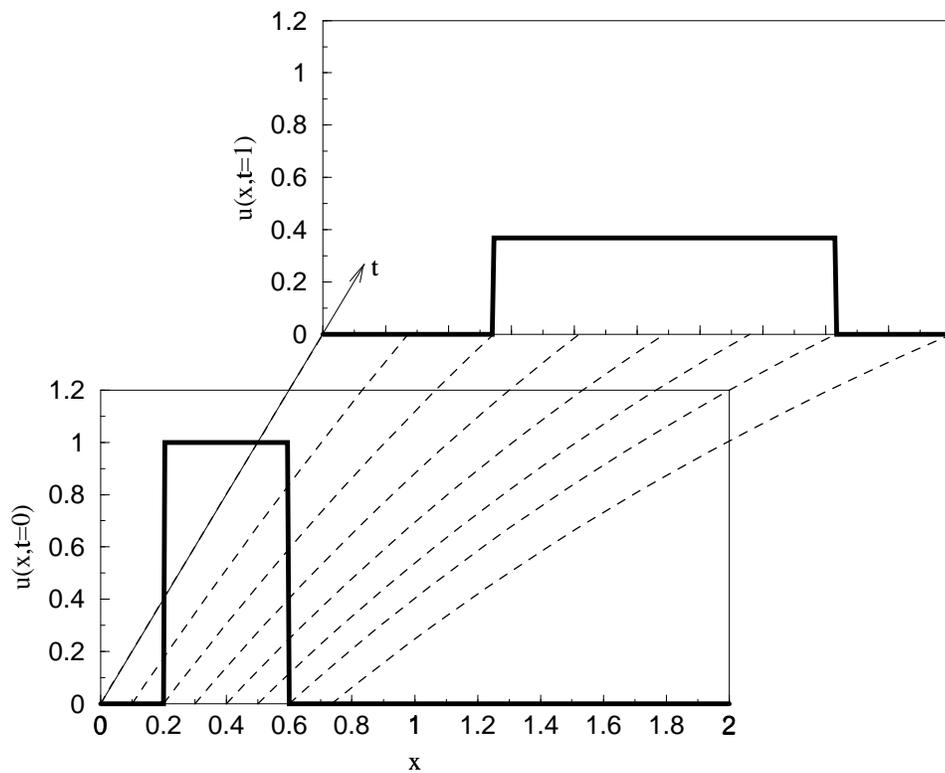
Daraus folgt wegen (2.8):

$$u(x(t), t) = u_0(x_0)e^{-t}, \quad \text{für } x(t) = x_0 e^t. \quad (2.14)$$

Wir nehmen einen beliebigen Punkt  $(x, t) \in [0, 2] \times [0, +\infty)$ . Dieser liegt auf einer Charakteristik, die von der Stelle  $(x_0, 0)$  startet mit  $x_0 = xe^{-t}$ . Einsatz des Wertes  $x_0$  in (2.14) führt zu:

$$u(x, t) = u_0(xe^{-t}) \cdot e^{-t}. \quad (2.15)$$

Das ist die gesuchte Lösung der Gleichung (2.1) mit der Anfangsbedingung (2.9), was man durch Einsatz der Funktion (2.15) in die Gleichung (2.1) überprüfen kann.

**TEST No. 4, sigma=0.02****TEST No. 5****Abbildung 2.1:** Exakte Lösung für Test No. 4 mit  $\sigma = 0.02$  und für Test No.5

Die Charakteristik, die von der Stelle  $x_0 = 0$  startet, lautet  $x = 0$ . Das heißt, daß der linke Rand des Berechnungsgebietes mit dieser Charakteristik übereinstimmt, und es gilt daher

$$u(0, t) = u_0(0) \cdot e^{-t}. \quad (2.16)$$

Nun wollen wir die Anfangsbedingung genauer spezifizieren. Wie im letzten Kapitel werden wir einen Fall mit einer stetigen und einen Fall mit einer diskontinuierlichen Lösungsfunktion untersuchen. Wir übernehmen daher die Anfangsbedingungen aus den Testfällen 1 und 2 (s. Abschnitt 1.2, Glngn. (1.32) und (1.39)). Die resultierenden Fälle werden als Testbeispiele No.4 und No.5 bezeichnet. Das Anfangsprofil sowie der Lösungverlauf für Test No.4 mit  $\sigma = 0.02$  und für Test No.5 ist zusammen mit den Charakteristiken in der Abb. 2.1 dargestellt.

Im nächsten Abschnitt werden wir verschiedene Diskretisierungen für die Konvektionsgleichung (2.1) vorstellen, und deren Effizienz an diesen Testbeispielen ausprobieren. Dabei werden wir uns zunächst auf den Fall einer variablen aber *positiven* Konvektionsgeschwindigkeit beschränken (so wie das bei  $a(x) = x$  für  $x \in (0, 2)$  auch der Fall ist). Die Verallgemeinerung der Verfahren auf den Fall einer beliebigen Konvektionsgeschwindigkeit wird in einem separaten Abschnitt besprochen.

## 2.2 Fall einer positiven variablen Geschwindigkeit

Ziel dieses Abschnittes ist es, die Vorgehensweise zum Aufbau eines Verfahrens höherer Ordnung, das oszillationsfreie Lösungen produziert, auf den Fall einer variablen Konvektionsgeschwindigkeit zu verallgemeinern. Eine direkte Übertragung der Herleitung eines TVD-Verfahrens ist **nicht** möglich, da im Falle einer variablen Konvektionsgeschwindigkeit die Konvektionsgleichung nicht in der Form (1.2) dargestellt werden kann und die totale Variation der analytischen Lösung je nach Verlauf der Funktion  $a(x)$  sowohl ab- als auch zunehmen kann. Würde ein numerisches Verfahren die TVD-Eigenschaft besitzen, so wären dadurch die physikalischen Eigenschaften der exakten Lösung unter Umständen verletzt, was wir gerade vermeiden wollen.

Anstatt ein Verfahren mit der TVD-Eigenschaft aufzubauen, können wir dagegen wie im Falle einer konstanten Geschwindigkeit den Konvektionsstrom als ein gewichtetes Mittel aus einem Verfahren erster Ordnung und einem Verfahren zweiter Ordnung bilden, und den Gewichtungsfaktor so wählen, daß das resultierende Verfahren **im Falle einer konstanten Konvektionsgeschwindigkeit** die TVD-Eigenschaft besitzen würde. Obwohl solche Verfahren für variable Geschwindigkeiten die TVD-Eigenschaft verletzen können, werden wir sie trotzdem als TVD-Verfahren bezeichnen.

Der Ausgangspunkt für unterschiedliche Diskretisierungsverfahren ist die einheitliche Finite-Volumen-Formulierung:

$$\frac{U_i - \bar{U}_i}{\Delta t} + \frac{F(U; i + 1/2) - F(U; i - 1/2)}{\Delta x} = 0 \quad (2.17)$$

mit

$$F(U; i - 1/2) = (aU)_{i-1/2} \approx \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} a(x_{i-1/2}) u(x_{i-1/2}, t) dt, \quad (2.18)$$

die analog zu (1.53) hergeleitet werden kann. Als erstes wollen wir den Upwind-Konvektionsstrom für den Fall einer variablen positiven Geschwindigkeit definieren. Im Falle einer konstanten positiven Geschwindigkeit wurde der Wert von  $U_{i-1/2}$  durch den nächsten stromaufwärts liegenden Wert  $U_{i-1}$  approximiert. Gleiche Vorgehensweise wird auch im Falle variabler Geschwindigkeit verwendet. Gleichzeitig gibt es zwei Möglichkeiten für die Darstellung der Geschwindigkeit  $a$  an der Stelle  $x_{i-1/2}$ . Vorausgesetzt, daß die Funktion  $a(x)$  bekannt ist, kann man sowohl den Wert der Funktion  $a(x)$  im stromaufwärts liegenden Gitterpunkt nehmen als auch direkt den Wert  $x_{i-1/2}$  in die Funktion  $a(x)$  einsetzen. Daraus ergeben sich zwei unterschiedliche Formulierungen für das Upwind-Verfahren. Im ersten Fall wird der gesamte Konvektionsstrom an der stromaufwärts von  $x_{i-1/2}$  liegenden Stelle ausgewertet:

$$(aU)_{i-1/2} = F^{up,fl}(U; i - 1/2) = a_{i-1} U_{i-1}. \quad (2.19)$$

Im zweiten Fall wird nur der Wert der unbekannt Funktion  $U$  an der Stelle  $x_{i-1}$  ausgewertet, und für die Konvektionsgeschwindigkeit  $a$  wird der genaue Wert an der Stelle  $x_{i-1/2}$  genommen:

$$(aU)_{i-1/2} = F^{up,sl}(U; i - 1/2) = a_{i-1/2} U_{i-1} \quad (2.20)$$

$$\text{mit } a_{i-1/2} = a(x_{i-1/2}).$$

Um die beiden Upwind-Ströme voneinander zu unterscheiden, verwenden wir zusätzliche Bezeichnung „FL“ im ersten und „SL“ im zweiten Fall. Die Bedeutung dieser Abkürzungen wird im weiteren Verlauf des Abschnittes erklärt. Diese zwei Definitionen von Upwind-Strömen haben entsprechend unterschiedliche Differenzenformeln zu Folge. Im ersten Fall erhält man folgende Diskretisierung der Konvektionsgleichung:

$$\boxed{\text{UPWIND(FL)} : \frac{U_i - \bar{U}_i}{\Delta t} + \frac{a_i U_i - a_{i-1} U_{i-1}}{\Delta x} = 0.} \quad (2.21)$$

Diese Formel kann man auch direkt aus der Konvektionsgleichung (2.1) mittels eines Finite-Differenzen-Verfahrens ableiten, wenn man für die Approximation der örtlichen Ableitung die Rückwärtsdifferenz verwendet. Für den Abbruchfehler des Verfahrens gilt:

$$L^{up,fl}(\Delta x) = -\frac{\Delta x}{2}(au)_{xx} + O(\Delta x^2) = O(\Delta x). \quad (2.22)$$

Im zweiten Fall sieht die Diskretisierungsformel folgendermaßen aus:

$$\boxed{\text{UPWIND(SL)} : \frac{U_i - \bar{U}_i}{\Delta t} + \frac{a_{i+1/2}U_i - a_{i-1/2}U_{i-1}}{\Delta x} = 0.} \quad (2.23)$$

Für den Abbruchfehler des Verfahrens gilt:

$$L^{up,sl}(\Delta x) = -\frac{\Delta x}{2}((au)_{xx} - (a_x u)_x) + O(\Delta x^2) = O(\Delta x). \quad (2.24)$$

Man sieht, daß beide Verfahren die Approximationsordnung von 1 haben, und daß im Falle  $a = const$  beide Diskretisierungen übereinstimmen. In Abb. 2.2 (links) sind die mit beiden Verfahren berechneten Lösungsprofile für den Test No.4 und Maschenweite  $\Delta x = 0.025$  gegenübergestellt. Beide Profile haben einen ähnlichen, für das UPWIND-Verfahren charakteristischen diffusiven Verlauf, wie man ihn aus dem vorigen Kapitel bereits kennt. Obwohl man deutlich den Unterschied zwischen zwei Lösungsprofilen sehen kann, ist der absolute Wert des numerischen Fehlers in beiden Fällen bis auf die dritte signifikante Stelle gleich groß. Der Vergleich von  $E-\Delta x$ -Kurven für beide Verfahren (Abb. 2.2, rechts) zeigt, daß dies auch für andere Maschenweiten der Fall ist. Das ist etwas überraschend, da man annehmen könnte, daß der UPWIND(SL)-Strom genauere Lösungen produziert, zumal bei der Approximation von  $(aU)_{i-1/2}$  der *exakte* Wert der Funktion  $a(x)$  an der Stelle  $x_{i-1/2}$  verwendet wurde. Daß man keine Verbesserung der Genauigkeit im Vergleich zu UPWIND(FL)-Lösung beobachtet, liegt daran, daß nicht die Genauigkeit der Approximation von  $(aU)_{i-1/2}$  durch  $F(U; i - 1/2)$  entscheidend ist, sondern wie gut die Ableitung  $(au)_x$  durch

$$\frac{F(u; i + 1/2) - F(u; i - 1/2)}{\Delta x}$$

approximiert wird, und das erfolgt in beiden Verfahren gleich gut (s. in diesem Zusammenhang die Diskussion am Ende des Abschnittes 1.7).

Analog zu beiden Upwind-Konvektionsströmen kann man zwei CDS-Ströme definieren. Der CDS(FL)-Strom wird als Mittelwert von Konvektionsflüssen  $(aU)_{i-1}$  und  $(aU)_i$  gebildet:

$$(aU)_{i-1/2} = F^{cd,sl}(U; i - 1/2) = \frac{a_{i-1}U_{i-1} + a_iU_i}{2}. \quad (2.25)$$

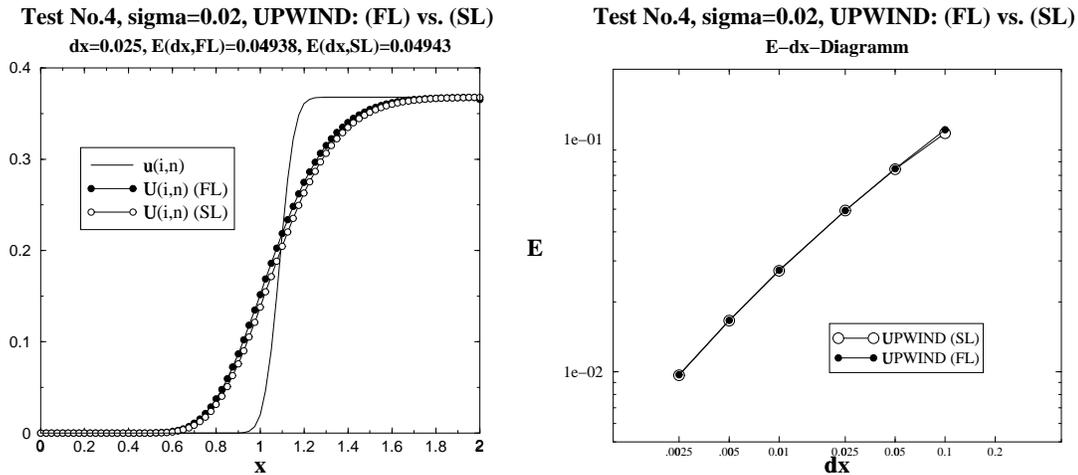


Abbildung 2.2: Test No.4 ( $\sigma = 0.02$ ): Vergleich von UPWIND (FL) und UPWIND (SL).

Beim CDS(SL)-Strom wird dagegen der genauere Wert der Konvektionsgeschwindigkeit  $a$  an der Stelle  $x_{i-1/2}$  mit dem Mittelwert aus  $U_{i-1}$  und  $U_i$  multipliziert:

$$(aU)_{i-1/2} = F^{cds,sl}(U; i-1/2) = a_{i-1/2} \frac{U_{i-1} + U_i}{2}. \quad (2.26)$$

Setzt man den Konvektionsstrom  $F^{cds,fl}(U; i-1/2)$  in die Finite-Volumen-Gleichung (2.17) ein, so ergibt sich eine Diskretisierung der Konvektionsgleichung, die der Approximation der örtlichen Ableitung durch die zentrale Differenz entspricht:

$$\boxed{\text{CDS(FL)} : \frac{U_i - \bar{U}_i}{\Delta t} + \frac{a_{i+1}U_{i+1} - a_{i-1}U_{i-1}}{2\Delta x} = 0.} \quad (2.27)$$

Für den Abbruchfehler des Verfahrens gilt

$$L^{cds,fl}(\Delta x) = \frac{\Delta x^2}{6} (au)_{xxx} + O(\Delta x^4) = O(\Delta x^2). \quad (2.28)$$

Das Verfahren hat also die Approximationsordnung von 2. Setzt man dagegen den Konvektionsstrom  $F^{cds,sl}(U; i-1/2)$  in die Finite-Volumen-Gleichung (2.17) ein, so führt das zu einer Diskretisierung, die auf den ersten Blick wenig mit der zentralen Differenz zu tun hat:

$$\boxed{\text{CDS(SL)} : \frac{U_i - \bar{U}_i}{\Delta t} + \frac{a_{i+1/2} \frac{U_i + U_{i+1}}{2} - a_{i-1/2} \frac{U_{i-1} + U_i}{2}}{\Delta x} = 0.} \quad (2.29)$$

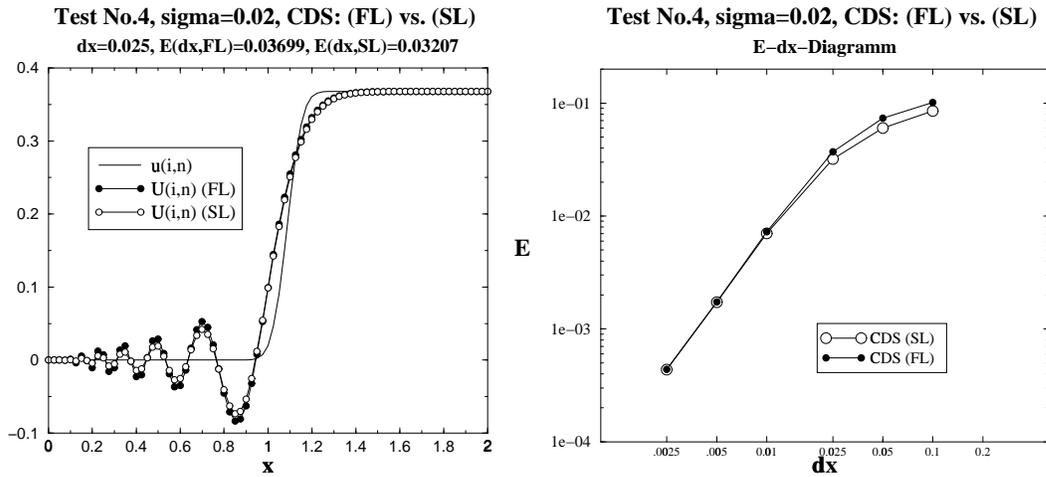


Abbildung 2.3: Test No.4 ( $\sigma = 0.02$ ): Vergleich von CDS (FL) und CDS (SL).

Im Falle einer konstanten Konvektionsgeschwindigkeit  $a = const$  stimmt dieses Verfahren mit dem CDS(FL) jedoch überein.

Für den Abbruchfehler des Verfahrens gilt

$$L^{cds,sl}(\Delta x) = \frac{\Delta x^2}{6} \left( (au)_{xxx} - \frac{3}{4}(a_{xx}u)_x - \frac{3}{2}(a_x u_x)_x \right) + O(\Delta x^4) = O(\Delta x^2). \quad (2.30)$$

Damit hat auch das CDS(SL)-Verfahren die Approximationsordnung von 2. In Abb. 2.3 (links) sind die mit beiden Verfahren berechneten Lösungsprofile für den Test No.4 und Maschenweite  $\Delta x = 0.025$  gegenübergestellt. Beide Profile haben einen ähnlichen, für das CDS-Verfahren charakteristischen dispersiven Verlauf. Dabei sind beim CDS(SL)-Profil die Fluktuationen etwas schwächer ausgeprägt, als beim CDS(FL)-Profil. Das kann man dadurch erklären, daß in unserem Testfall wegen  $a(x) = x$  für den Abbruchfehler des CDS(SL)-Verfahrens gilt:

$$L^{cds,sl}(\Delta x) = \frac{\Delta x^2}{6} \left( (au)_{xxx} - \frac{3}{2}u_{xx} \right) + O(\Delta x^4) = O(\Delta x^2). \quad (2.31)$$

Man sieht, daß dieser Abbruchfehler im Vergleich zu (2.28) einen Diffusionsterm enthält, der dämpfend auf die Oszillationen wirkt. Das resultiert in einer höheren Genauigkeit des CDS(SL)-Verfahrens auf groben Gittern gegenüber der CDS(FL)-Diskretisierung. Mit zunehmender Gitterverfeinerung verschwindet der Genauigkeitsvorsprung des CDS(SL)-Verfahrens jedoch vollständig, wie ein Vergleich von  $E$ - $\Delta x$ -Kurven für beide Verfahren in Abb. 2.3 (rechts) deutlich zeigt.

Wir haben somit alle notwendigen Vorbereitungen getroffen, um einen TVD-Konvektionsstrom zu definieren. Auch hier gibt es zwei Möglichkeiten, je nach dem, ob man ein gewich-

tetes Mittel aus den beiden FL- oder aus den beiden SL-Verfahren bildet. Wir beginnen mit der Definition eines TVD(FL)-Konvektionsstromes. Analog zu (1.152) definieren wir

$$\begin{aligned}
 F^{tvd,fl}(U; l) &:= (1 - \phi_l^{fl}) F^{up,fl}(U; l) + \phi_l^{fl} F^{cds,fl}(U; l) \\
 &= F^{up,fl}(U; l) + \phi_l^{fl} (F^{cds,fl}(U; l) - F^{up,fl}(U; l)) \\
 &= a_{i-1} U_{i-1} + \frac{\phi_l^{fl}}{2} (a_i U_i - a_{i-1} U_{i-1}).
 \end{aligned} \tag{2.32}$$

Setzt man diesen Konvektionsstrom in die Finite-Volumen-Gleichung (2.17) ein, so ergibt sich folgende Diskretisierung der Konvektionsgleichung:

**TVD(FL) :**

$$\frac{U_i - \bar{U}_i}{\Delta t} + \frac{a_i U_i + \frac{\phi_r^{fl}}{2} (a_{i+1} U_{i+1} - a_i U_i) - a_{i-1} U_{i-1} - \frac{\phi_l^{fl}}{2} (a_i U_i - a_{i-1} U_{i-1})}{\Delta x} = 0. \tag{2.33}$$

Die zentrale Frage ist die Definition des Gewichtungsfaktors  $\phi_l^{fl}$ . Wie oben bereits erwähnt, darf man nicht verlangen, daß das TVD(FL)-Verfahren für *beliebige* Konvektionsgeschwindigkeiten die TVD-Eigenschaft besitzt, sondern nur für den Fall einer konstanten Geschwindigkeit. Da im Falle  $a \equiv const$  die Ströme  $F^{up,fl}(U; l)$  und  $F^{cds,fl}(U; l)$  mit den im letzten Kapitel definierten Konvektionsströmen  $F^{up}(U; l)$  und  $F^{cds}(U; l)$  übereinstimmen, reicht es zu gewährleisten, daß auch der Gewichtungsfaktor  $\phi_l^{fl}$  im Falle  $a \equiv const$  dem Wert  $\phi_l$  aus (1.153) gleich ist. Wir setzen dazu

$$\boxed{\phi_l^{fl} = 0, \text{ wenn } a_i U_i - a_{i-1} U_{i-1} = 0} \tag{2.34}$$

und ansonsten

$$\boxed{\phi_l^{fl} = \phi(\theta_l^{fl})}, \tag{2.35}$$

wobei  $\phi(\theta)$  ein beliebiger TVD-Limiter ist, und  $\theta_l^{fl}$  folgendermaßen definiert ist:

$$\theta_l^{fl} := \frac{a_{i-1} U_{i-1} - a_{i-2} U_{i-2}}{a_i U_i - a_{i-1} U_{i-1}}. \tag{2.36}$$

Im Falle  $U_i = U_{i-1}$  und  $a \equiv const$  gilt auch  $a_i U_i = a_{i-1} U_{i-1}$  und beide Gewichtungsfaktoren  $\phi_l$  und  $\phi_l^{fl}$  sind gleich Null (1.158, 2.34). Gilt dagegen  $U_i \neq U_{i-1}$ , so ist bei  $a \equiv const$

auch  $a_i U_i \neq a_{i-1} U_{i-1}$  und wir erhalten

$$\begin{aligned}\phi_l^{fl} &= \phi(\theta_l^{fl}) = \phi\left(\frac{a_{i-1}U_{i-1} - a_{i-2}U_{i-2}}{a_i U_i - a_{i-1}U_{i-1}}\right) = \\ &= \phi\left(\frac{U_{i-1} - U_{i-2}}{U_i - U_{i-1}}\right) = \phi(\theta_l) = \phi_l.\end{aligned}\quad (2.37)$$

Wir haben damit gezeigt, daß für  $a \equiv const$  die Diskretisierungsformeln (1.153) und (2.33) beim Einsatz derselben TVD-Limiter-Funktion  $\phi(\theta)$  übereinstimmen. Man beachte, daß nicht nur der TVD(FL)-Konvektionsstrom als ein gewichtetes Mittel aus zwei *Strömen* gebildet wurde, sondern auch bei der Definition von  $\theta_l^{fl}$  die Differenz von zwei benachbarten *Strömen* im Zähler und Nenner steht. Da die Konvektionsströme (auf englisch „fluxes“) im Mittelpunkt des Aufbaus eines TVD(FL)-Verfahrens stehen, werden diese Verfahren als „*Flux-Limiter*“-Verfahren bezeichnet. Das erklärt auch die bisher verwendete Abkürzung „FL“.

Analog zu (2.32) können wir den TVD(SL)-Konvektionsstrom als ein gewichtetes Mittel aus UPWIND(SL)- und CDS(SL)-Strom bilden:

$$\begin{aligned}F^{tvd,sl}(U;l) &:= (1 - \phi_l^{sl})F^{up,sl}(U;l) + \phi_l^{sl}F^{cds,sl}(U;l) \\ &= F^{up,sl}(U;l) + \phi_l^{sl}(F^{cds,sl}(U;l) - F^{up,sl}(U;l)) \\ &= a_{i-1/2} \left( U_{i-1} + \frac{\phi_l^{sl}}{2}(U_i - U_{i-1}) \right)\end{aligned}\quad (2.38)$$

was zu folgendem TVD(SL)-Verfahren führt:

**TVD(SL) :**

$$\frac{U_i - \bar{U}_i}{\Delta t} + \frac{a_{i+1/2} \left( U_i + \frac{\phi_l^{sl}}{2}(U_{i+1} - U_i) \right) - a_{i-1/2} \left( U_{i-1} + \frac{\phi_l^{sl}}{2}(U_i - U_{i-1}) \right)}{\Delta x} = 0. \quad (2.39)$$

Man beachte, daß für alle 3 SL-Ströme gilt

$$F^{\dots,sl}(U;l) = a_{i-1/2} U_{i-1/2}, \quad (2.40)$$

mit

$$U_{i-1/2}^{up} := U_{i-1} \quad (2.41)$$

$$U_{i-1/2}^{cds} := (U_{i-1} + U_i)/2 \quad (2.42)$$

$$U_{i-1/2}^{tvd} := (1 - \phi_l^{sl})U_{i-1/2}^{up} + \phi_l^{sl}U_{i-1/2}^{cds} \quad (2.43)$$

d.h. sie unterscheiden sich nur in der Approximation des Wertes  $U_{i-1/2}$ . Obwohl wir den TVD(SL)-Konvektionsstrom als ein gewichtetes Mittel aus zwei *Strömen* gebildet haben, könnte man die Definition (2.38) auch folgendermaßen interpretieren: Man definiere die TVD-Approximation für die unbekannte *Variable*  $U_{i-1/2}$  als ein gewichtetes Mittel aus der Upwind-Approximation  $U_{i-1/2}^{up}$  und der CDS-Approximation  $U_{i-1/2}^{cds}$  und setze diese anschließend in die allgemeine Definition (2.40) ein. Bei dieser Interpretation steht nicht der Konvektionsstrom als Ganzes, sondern die Approximation der unbekanntes Funktion im Vordergrund. Man kann weiterhin den Gewichtungsfaktor  $\phi_l^{sl}$  genauso wie im Falle einer konstanten Geschwindigkeit definieren:

$$\boxed{\phi_l^{sl} = 0, \text{ wenn } U_i - U_{i-1} = 0.} \quad (2.44)$$

und ansonsten

$$\boxed{\phi_l^{sl} = \phi(\theta_l^{sl}),} \quad (2.45)$$

wobei für  $\theta_l^{sl}$  gilt:

$$\theta_l^{sl} := \frac{U_{i-1} - U_{i-2}}{U_i - U_{i-1}}. \quad (2.46)$$

In diesem Fall gilt  $\phi_l^{sl} \equiv \phi_l$  und die Diskretisierungsformeln (1.153) und (2.39) stimmen für  $a \equiv \text{const}$  überein. Man beachte, daß die Gleichung  $\phi_l^{sl} = \phi_l$  auch im Falle einer variablen Konvektionsgeschwindigkeit erfüllt ist, da die Funktion  $a(x)$  in die Definition von  $\theta_l^{sl}$  gar nicht eingeht. Wir sehen, daß nicht nur beim Aufbau des TVD(SL)-Konvektionsstroms die Approximation der unbekanntes Variablen im Vordergrund steht, sondern auch bei der Definition von  $\theta_l^{sl}$  der Verlauf des *Lösungsprofils* (und nicht die Glattheit des Konvektionsstromes  $aU$ ) entscheidend ist. Schreibt man die Definition (2.43) folgendermaßen um

$$U_{i-1/2}^{tvd} = U_{i-1} + \phi_l^{sl}(U_i - U_{i-1})/2, \quad (2.47)$$

so kann man auch sagen, daß der Faktor  $\phi_l^{sl}$  die zu berücksichtigende *Steigung* (auf englisch „slope“) des Lösungsprofils bei der Approximation von  $U_{i-1/2}$  limitiert. Daher werden die TVD(SL)-Verfahren als „*Slope-Limiter*“-Verfahren bezeichnet, was auch die Bedeutung der bisher verwendeten Abkürzung „SL“ erklärt.

Bevor wir die entwickelten TVD-Verfahren für unsere Testfälle einsetzen, soll noch eine wichtige Bemerkung gemacht werden. Wir haben gesehen, daß die beiden Versionen von UPWIND-Verfahren die Approximationsordnung von 1 haben und daß sowohl CDS(FL) als auch CDS(SL) zweiter Ordnung genau sind. Das gleiche gilt auch für die QUICK(FL)- und QUICK(SL)-Verfahren, die man durch Einsatz des Limiters  $\phi(\theta) = 3/4 + \theta/4$  (der kein

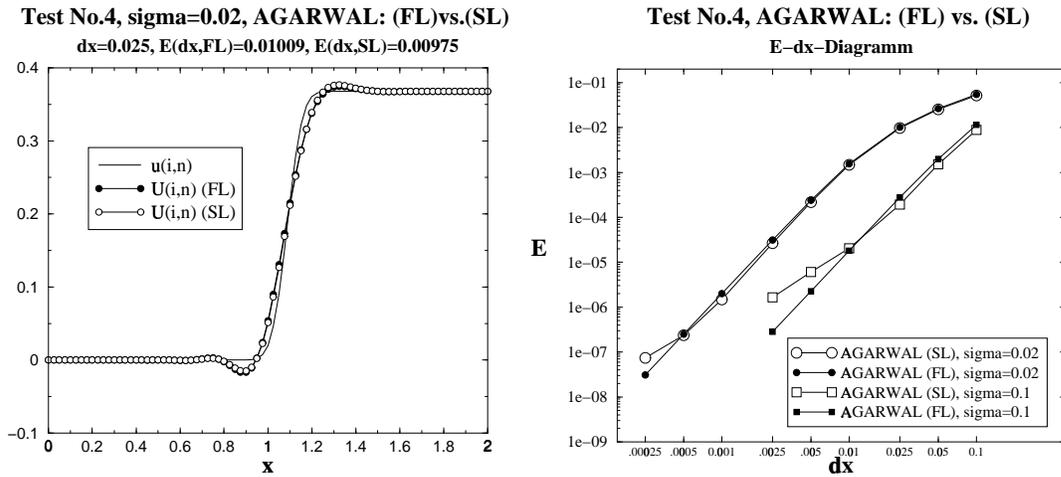


Abbildung 2.4: Test No.4 ( $\sigma = 0.02$ ): Vergleich von AGARWAL (FL) und AGARWAL (SL).

TVD-Limiter ist !) in das TVD(FL)- bzw. TVD(SL)-Verfahren erhält. Setzt man stattdessen die Funktion  $\phi(\theta) = 2/3 + \theta/3$  in beide Diskretisierungsformeln ein, so erhält man die AGARWAL(FL)- und AGARWAL(SL)-Verfahren, die *nicht* mehr die gleiche Approximationsordnung haben. Während das AGARWAL(FL)-Verfahren mit einem Abbruchfehler

$$L^{agar,fl}(\Delta x) = \frac{\Delta x^3}{12}(au)_{xxxx} + O(\Delta x^4) = O(\Delta x^3) \quad (2.48)$$

die Approximationsordnung von 3 hat, gilt für den Abbruchfehler von AGARWAL(SL):

$$L^{agar,sl}(\Delta x) = \frac{\Delta x^2}{24}((a_{xx}u)_x + 2(a_x u_x)_x) + O(\Delta x^3) = O(\Delta x^2). \quad (2.49)$$

Das bedeutet, daß das AGARWAL(SL)-Verfahren nur 2. Ordnung genau ist. In Abb. 2.4 (links) sind die mit beiden Verfahren berechneten Lösungsprofile für den Test No.4 und Maschenweite  $\Delta x = 0.025$  gegenübergestellt. Man sieht praktisch keinen Unterschied zwischen beiden Lösungen. Ein Vergleich von  $E$ - $\Delta x$ -Kurven für beide Verfahren (Abb. 2.4, rechts) bestätigt zwar, daß bei  $\Delta x \rightarrow 0$  die Konvergenzordnung von AGARWAL(SL) niedriger als die Konvergenzordnung von AGARWAL(FL) ist, dennoch stimmen die absoluten Werte vom numerischen Fehler bis auf sehr feine Gitterauflösungen fast genau überein. Das bedeutet, daß (zumindest im Testfall No.4) das AGARWAL(SL)-Verfahren die empirische Konvergenzrate von 3 hat, obwohl das Verfahren formal nur 2. Ordnung genau ist.

In der Abb. 2.5 sind die Ergebnisse für das Testbeispiel No.4 ( $\sigma = 0.02$ ) und eine Maschenweite  $\Delta x = 0.025$  für SB-, MC-, QUICK- und AGARWAL-Limiter, integriert in das TVD(FL)- sowie in das TVD(SL)-Verfahren dargestellt. Man sieht, daß in allen vier FL-Verfahren die numerische Lösung leichte Oszillationen (man würde in diesem Fall eher sagen „Überschwinger“) aufweist. An dieser Stelle sei daran erinnert, daß die TVD-Verfahren

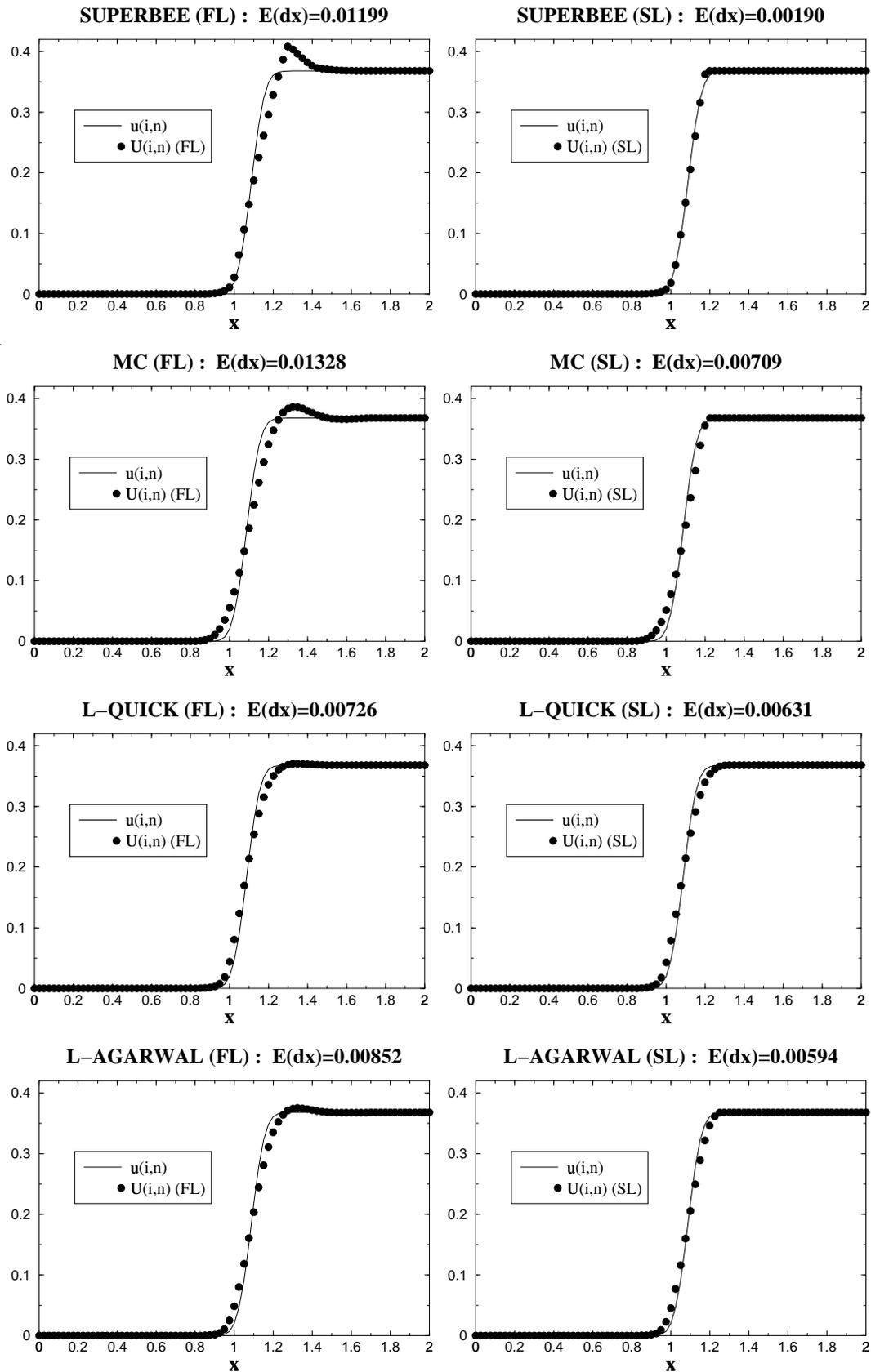


Abbildung 2.5: Test No.4 ( $\sigma = 0.02$ ): Verlauf der Lösung auf einem Gitter  $\Delta x = 0.025$  für 8 Verfahren.

für die lineare Konvektionsgleichung mit variabler Geschwindigkeit auf rein empirischer Basis konstruiert wurden und es besteht daher keine Garantie, daß die damit produzierten Lösungen oszillationsfrei sind. Trotzdem ergeben alle vier TVD(SL)-Verfahren Lösungsprofile, die keinerlei Fluktuationen aufweisen. Die genaueste Lösung auf dem hier betrachteten Gitter wurde vom TVD(SL)-Verfahren mit dem Superbee-Limiter produziert, wobei man sieht, daß diese Lösung (wie im Falle einer konstanten Konvektionsgeschwindigkeit) an zu starker Antidiffusion leidet. Das führt dazu, daß dieses Verfahren auf feineren Gittern schlechter als andere TVD(SL)-Verfahren abschneidet (s. Abb. 2.6).

### Test No.4 ( $\sigma=0.02$ ): E-dx-Diagramm

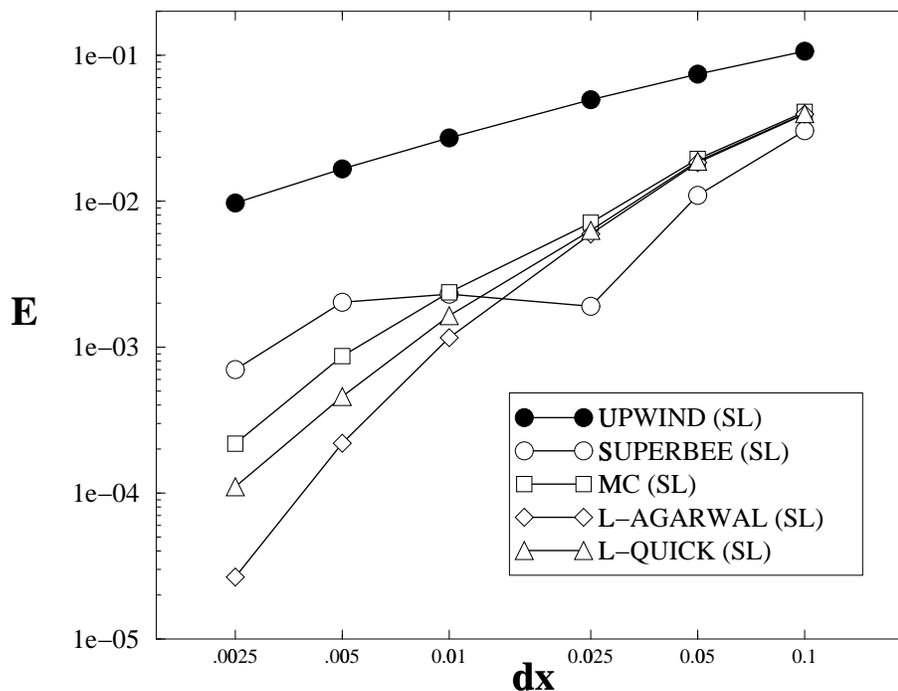


Abbildung 2.6: Test No.4 ( $\sigma = 0.02$ ): Vergleich von 4 Limitern für TVD(SL) untereinander.

Von den restlichen drei untersuchten Limitern werden die besten Ergebnisse vom AGARWAL-Limiter produziert, gefolgt von QUICK-Limiter und MC-Limiter, wobei der Unterschied in der Genauigkeit zwischen diesen drei Verfahren erst auf relativ feinen Gittern deutlich wird.

In der Abb. 2.7 sind die Ergebnisse für das Testbeispiel No.5 für FL- und SL-Versionen von SB-, MC-, QUICK- und AGARWAL-Verfahren auf einem Gitter mit der Maschenweite  $\Delta x = 0.025$  dargestellt. Wie im Test No.4 (Abb. 2.5), treten bei allen vier FL-Verfahren leichte Oszillationen im Lösungsprofil auf. Bei allen vier SL-Verfahren ist die numerische Lösung dagegen oszillationsfrei. Obwohl beide Diskontinuitäten nicht sehr gut aufgelöst

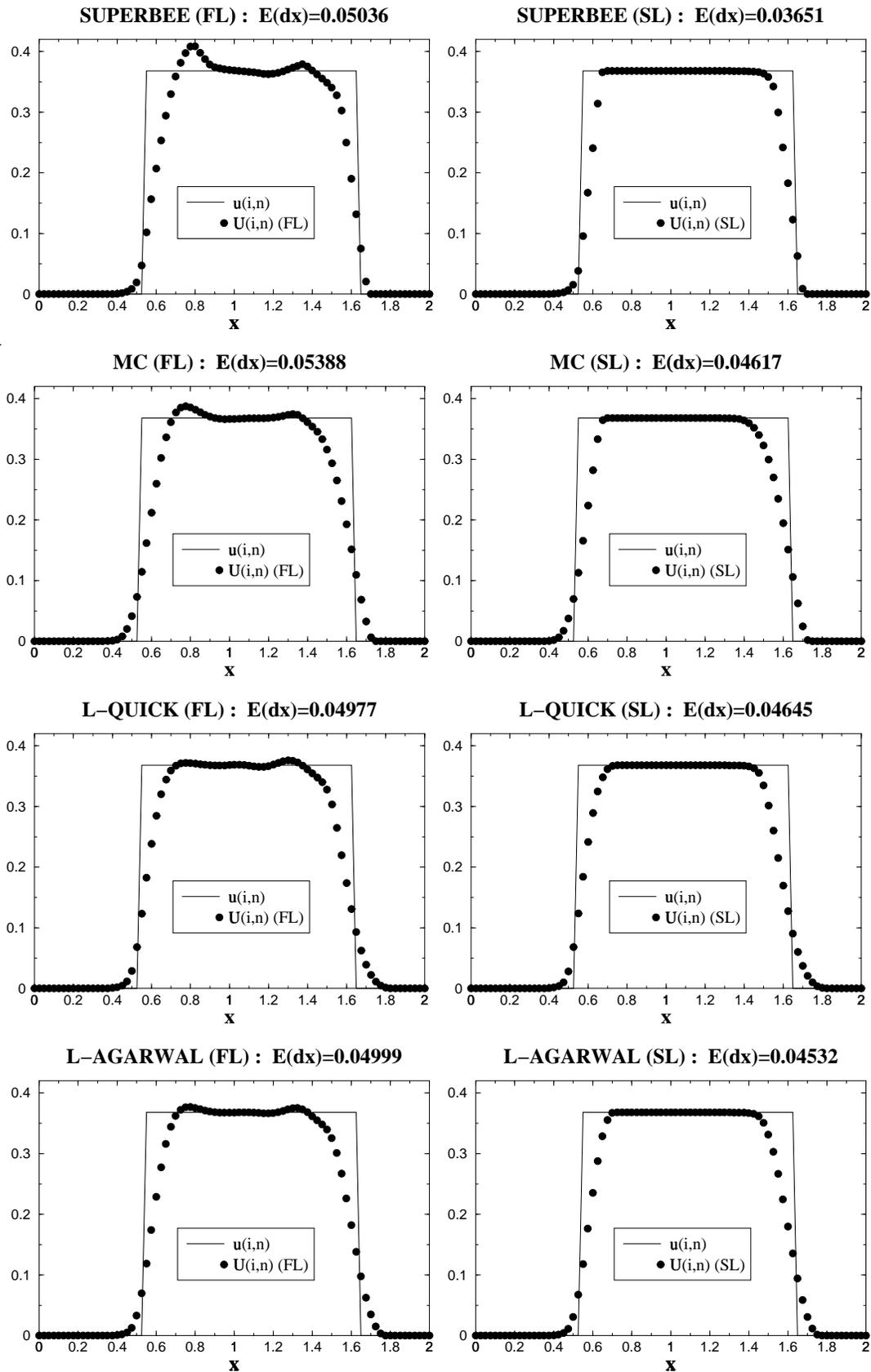
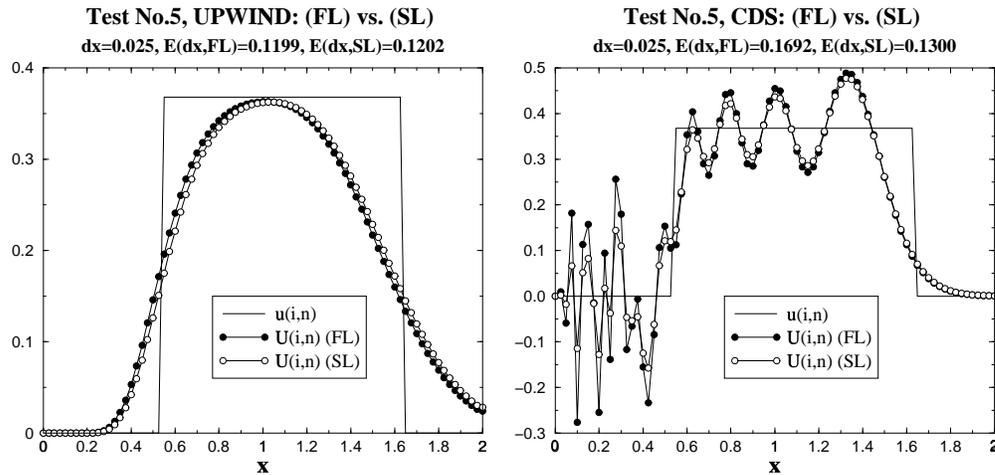


Abbildung 2.7: Test No.5: Verlauf der Lösung auf einem Gitter  $\Delta x = 0.025$  für 8 Verfahren.



**Abbildung 2.8:** Test No.5 : Lineare UPWIND- und CDS-Verfahren.

werden, ist der Lösungsverlauf wesentlich steiler als beim UPWIND-Verfahren bei der gleichen Auflösung (s. Abb. 2.8, links). Zu Vergleichszwecken wird auch die CDS-Lösung (FL- und SL-Versionen) in der Abbildung 2.8 (rechts) dargestellt. Beide Lösungen leiden stark an dispersivem Verhalten des Diskretisierungsverfahrens, und die Genauigkeit des CDS-Verfahrens ist sogar niedriger als im UPWIND-Fall.

Ein Vergleich von 4 limitierten SL-Verfahren untereinander (Abb. 2.9) zeigt, daß wie im Test No.2 der SB-Limiter zu den mit Abstand genauesten Ergebnissen führt. Die anderen drei Limiter unterscheiden sich nur unwesentlich in der Genauigkeit, wobei das L-AGARWAL(SL)-Verfahren die kleinsten Fehler produziert.

**Zusammenfassung.** In diesem Abschnitt wurde ein Versuch unternommen, die Idee zur Konstruktion eines oszillationfreien Verfahrens für die lineare Konvektionsgleichung von dem Fall einer konstanten Konvektionsgeschwindigkeit auf den Fall einer variablen positiven Geschwindigkeit zu übertragen. Es wurden zwei mögliche Erweiterungen vorgestellt: Flux-Limiter-Verfahren und Slope-Limiter-Verfahren. Bei den Flux-Limiter-Verfahren wird der Beitrag des CDS-Anteils bei der Berechnung des numerischen Konvektionsstroms limitiert, während bei den Slope-Limiter-Verfahren die Limitierung den Verlauf der zwischen den Gitterpunkten interpolierten Lösungsfunktion betrifft. Beide Verfahren wurden auf einer rein empirischen Basis konstruiert, so daß man von vornherein nicht garantieren konnte, daß die numerische Lösung tatsächlich oszillationfrei sein wird. Die Eigenschaften der resultierenden Verfahren wurden an 2 Testfällen untersucht. Während alle Flux-Limiter-Verfahren in beiden Testbeispielen leichte Oszillationen produzierten, lieferten die Slope-Limiter-Verfahren in beiden Fällen physikalisch sinnvolle Lösungen.

### Test No.5: E-dx-Diagramm

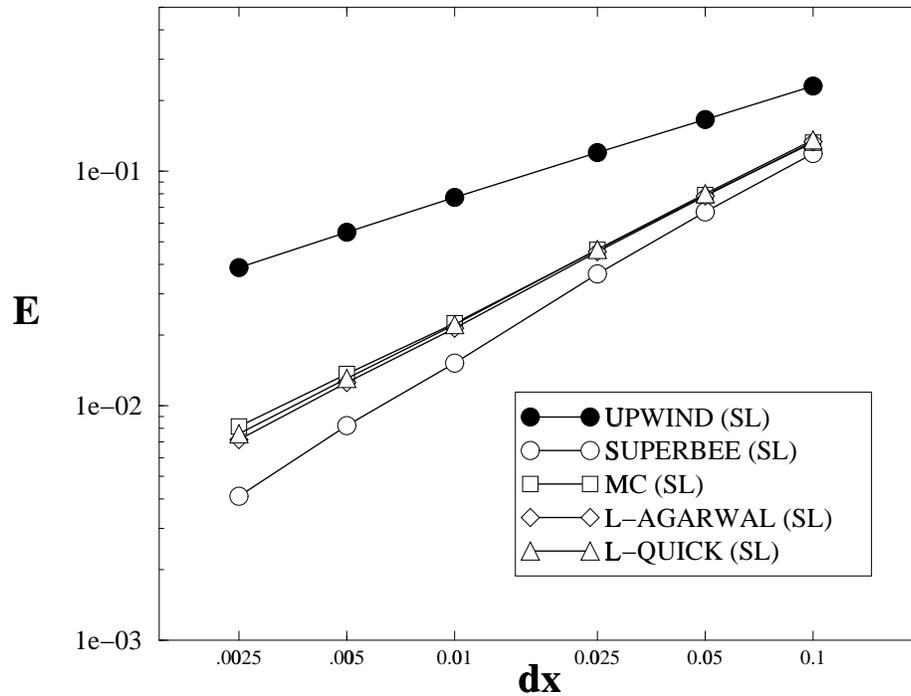


Abbildung 2.9: Test No.5: Vergleich von 4 Limitern für TVD(SL) untereinander.

Ein Vergleich von Slope-Limiter-Verfahren mit unterschiedlichen Limiter-Funktionen untereinander führte zu folgendem Ergebnis. Wie im Falle einer konstanten Konvektionsgeschwindigkeit hatte der SB-Verfahren die beste Konvergenzrate von allen Verfahren im Falle von diskontinuierlichen Lösungsfunktion, während im Falle einer glatten Lösung das stark antidiffusive Verhalten des Verfahrens seine Genauigkeit beeinträchtigte. Da wir an einem Verfahren interessiert sind, daß sowohl für glatte als auch für diskontinuierliche Lösungen gute Konvergenzeigenschaften aufweist, wird der SB-Limiter im weiteren Verlauf der Arbeit nicht mehr eingesetzt. Von den restlichen 3 untersuchten Limitern wurde das beste Konvergenzverhalten vom L-AGARWAL-Verfahren gezeigt, gefolgt von L-QUICK- und MC-Limiter.

## 2.3 Fall einer beliebigen variablen Geschwindigkeit

Im letzten Abschnitt haben wir uns auf den Fall einer variablen aber überall nichtnegativen Konvektionsgeschwindigkeit beschränkt. Nun wollen wir die Annahme, daß die Konvektionsgeschwindigkeit nicht negativ sein darf, aufheben, und die Konstruktion eines TVD-

Verfahrens entsprechend verallgemeinern. Wir werden uns dabei auf die SL-Verfahren beschränken, da die TVD(FL)-Verfahren bereits im Falle einer positiven Konvektionsgeschwindigkeit nicht zufriedenstellende Ergebnisse geliefert haben. Das hochgestellte „sl“ wird diesmal bei den Bezeichnungen einfachheitshalber weggelassen.

Wir gehen wiederum von der Finite-Volumen-Formulierung (2.17) aus, und approximieren den numerischen Konvektionsstrom  $(aU)_{i-1/2}$  durch

$$(aU)_{i-1/2} = F^{tvd}(U; i - 1/2) = a_{i-1/2} U_{i-1/2}^{tvd}, \quad (2.50)$$

wobei für die Konvektionsgeschwindigkeit  $a_{i-1/2}$  der genaue Wert  $a(x_{i-1/2})$  eingesetzt wird (s. auch 2.40).

Bei einer UPWIND-Approximation der Größe  $U_{i-1/2}$  wird der Wert der Lösungsfunktion im nächsten stromaufwärts von  $x_{i-1/2}$  liegenden Gitterpunkt genommen. Dieser ist gleich  $x_{i-1}$ , wenn die Konvektionsgeschwindigkeit  $a_{i-1/2}$  positiv ist, und gleich  $x_i$ , wenn die Konvektionsgeschwindigkeit negativ ist. Das führt zu:

$$U_{i-1/2}^{up} = \begin{cases} U_{i-1}, & \text{wenn } a_{i-1/2} \geq 0 \\ U_i, & \text{wenn } a_{i-1/2} < 0 \end{cases} \quad (2.51)$$

Die CDS-Approximation für  $U_{i-1/2}$  ist dagegen unabhängig von der Stromrichtung, und wird wie im Falle einer positiven Geschwindigkeit definiert:

$$U_{i-1/2}^{cds} := (U_{i-1} + U_i)/2. \quad (2.52)$$

Man bildet nun die TVD-Approximation für  $U_{i-1/2}$  als

$$U_{i-1/2}^{tvd} := (1 - \phi_l) U_{i-1/2}^{up} + \phi_l U_{i-1/2}^{cds} \quad (2.53)$$

mit

$$\phi_l = \begin{cases} \phi(\theta_l), & \text{wenn } U_i - U_{i-1} \neq 0 \\ 0, & \text{wenn } U_i - U_{i-1} = 0 \end{cases} \quad (2.54)$$

Bei der Definition der Größe  $\theta_l$  muß wiederum die Stromrichtung berücksichtigt werden:

$$\theta_l = \frac{U_I - U_{I-1}}{U_i - U_{i-1}}, \quad (2.55)$$

wobei Index  $I$  wie folgt definiert wird:

$$I = \begin{cases} i - 1, & \text{wenn } a_{i-1/2} \geq 0, \\ i + 1, & \text{wenn } a_{i-1/2} < 0. \end{cases} \quad (2.56)$$

Damit ist der Aufbau des TVD(SL)-Verfahrens bereits vollständig definiert. Man berechne zunächst die UPWIND- und CDS-Approximationen für  $U_{i-1/2}$  nach (2.51, 2.52). Man bestimme danach den Gewichtungsfaktor  $\phi_l$  nach (2.54) und berechne die TVD-Approximation  $U_{i-1/2}^{tvd}$  für  $U_{i-1/2}$  nach (2.53). Schließlich setze man den so ermittelten Wert  $U_{i-1/2}^{tvd}$  in die Formel (2.50) zur Berechnung des numerischen Konvektionsstromes  $F^{tvd}(U; i - 1/2)$  ein.

Obwohl der TVD(SL)-Konvektionsstrom auf diese Weise eindeutig definiert wird, wird eine direkte Umsetzung der beschriebenen Berechnungsmethode dadurch erschwert, daß die Werte der Gitterfunktion  $\{U_i\}$  nicht bekannt sind und der numerische Konvektionsstrom  $F^{tvd}(U; i - 1/2)$  in diesem Fall eine komplexe nichtlineare Funktion von  $\{U_i\}$  ist. Die Linearisierung kann durch den „deferred correction approach“ durchgeführt werden, s. Abschnitt 1.6 (1.116, 1.117). Die Aufteilung des Konvektionsstromes  $F^{tvd}(U; i - 1/2)$  in einen UPWIND-Strom  $F^{up}(U; l)$  und einen Korrekturstrom  $F^{cds}(U; l)$  wird folgendermaßen durchgeführt. Der Einsatz von (2.53) in (2.50) führt zu:

$$\begin{aligned} F^{tvd}(U; l) &= a_{i-1/2} U_{i-1/2}^{tvd} \\ &= a_{i-1/2} \left( (1 - \phi_l) U_{i-1/2}^{up} + \phi_l U_{i-1/2}^{cds} \right) \\ &= a_{i-1/2} U_{i-1/2}^{up} + \phi_l a_{i-1/2} \left( U_{i-1/2}^{cds} - U_{i-1/2}^{up} \right) \\ &=: F^{up}(U; l) + F^{dc}(U; l). \end{aligned} \quad (2.57)$$

Für den UPWIND-Strom  $F^{up}(U; l) := a_{i-1/2} U_{i-1/2}^{up}$  gilt:

$$\begin{aligned} F^{up}(U; l) &= \begin{cases} a_{i-1/2} U_{i-1}, & \text{wenn } a_{i-1/2} \geq 0 \\ a_{i-1/2} U_i, & \text{wenn } a_{i-1/2} < 0 \end{cases} \\ &= \max(a_{i-1/2}, 0) \cdot U_{i-1} - \max(-a_{i-1/2}, 0) \cdot U_i \end{aligned} \quad (2.58)$$

Damit lassen sich die Koeffizienten des Tridiagonal-Systems (1.117) direkt berechnen.

Für den Korrekturstrom  $F^{dc}(U; l) := \phi_l a_{i-1/2} \left( U_{i-1/2}^{cds} - U_{i-1/2}^{up} \right)$  gilt:

$$F^{dc}(U; l) = \phi_l \cdot \left( a_{i-1/2} U_{i-1/2}^{cds} - a_{i-1/2} U_{i-1/2}^{up} \right)$$

$$\begin{aligned}
&= \phi_l \cdot \left( a_{i-1/2} \frac{U_{i-1} + U_i}{2} - \max(a_{i-1/2}, 0) \cdot U_{i-1} + \max(-a_{i-1/2}, 0) \cdot U_i \right) \\
&= \phi_l \cdot \frac{1}{2} \cdot |a_{i-1/2}| \cdot (U_i - U_{i-1}).
\end{aligned} \tag{2.59}$$

Durch den Einsatz einer Funktion  $\bar{\phi}(\cdot, \cdot)$  (s. (1.196)) läßt sich der Korrekturstrom  $F^{dc}(U; l)$  analog zu (1.197) berechnen:

$$F^{dc}(U; l) = \frac{|a_{i-1/2}|}{2} \cdot \bar{\phi}(U_I - U_{I-1}, U_i - U_{i-1}). \tag{2.60}$$

**Behandlung der Randbedingungen.** Zum Schluß wollen wir auf die Formulierung der numerischen Randbedingungen eingehen. Löst man die lineare Konvektionsgleichung (2.1) auf einem endlichen Intervall  $[X_1, X_2]$ , so kann — in Abhängigkeit von der Richtung der Konvektionsgeschwindigkeit  $a(x)$  am Rand — jeder der Randpunkte  $X_1$  und  $X_2$  sowohl dem Einströmrand als auch dem Ausströmrand entsprechen. Wir werden uns deshalb auf die Betrachtung des Randpunktes  $X_1$  beschränken.

Gilt  $a(X_1) > 0$ , so ist  $X_1$  ein Einströmrand und es soll für diesen Randpunkt eine physikalische Randbedingung in der Form

$$u(x = X_1, t) = \mu(t), \quad t > 0 \tag{2.61}$$

vorliegen, damit die analytische Lösung der Konvektionsgleichung eindeutig bestimmt ist. Aus dieser physikalischen Randbedingung läßt sich der Wert der Gitterfunktion im ersten Gitterpunkt direkt berechnen durch:

$$U_0 = \mu(t_{n+1}). \tag{2.62}$$

Gilt  $a(X_1) < 0$ , so ist  $X_1$  ein Ausströmrand und es liegt keine physikalische Randbedingung vor. In diesem Fall wird eine Finite-Volumen-Formulierung für ein „halbes“ Kontrollvolumen  $[x_0, x_{1/2}]$  verwendet:

$$\frac{U_0 - \bar{U}_0}{\Delta t} + \frac{F(U; 1/2) - F(U; 0)}{\Delta x/2} = 0. \tag{2.63}$$

Hier braucht nur der Strom  $F(U; 1/2)$  approximiert werden, da für  $F(U; 0)$  gilt:

$$F(U; 0) = (aU)_0 = a(x_0) \cdot U_0. \tag{2.64}$$

Für  $F(U; 1/2)$  wird der TVD-Konvektionsstrom  $F^{tvd}(U; 1/2)$  nach (2.50) eingesetzt. Gilt  $a_{1/2} \geq 0$ , so benötigt man zur Berechnung des Konvektionsstromes  $F(U; 1/2)$  den Wert der

Gitterfunktion an der Stelle  $x_{-1}$ , die außerhalb des Berechnungsgebietes liegt. Im Abschnitt 1.11 wurden drei Varianten zur Berechnung des fiktiven Wertes  $U_{-1}$  angegeben. Aus den dort dargelegten Gründen sollte nur die Variante 2 eingesetzt werden.

Ist die Konvektionsgeschwindigkeit am Rand  $a(X_1)$  gleich 0 (das ist übrigens in beiden Testbeispielen der Fall), so stimmt der Rand  $\{(x, t) : x = X_1, t > 0\}$  mit der Charakteristik überein, die von der Stelle  $(x, t) = (X_1, 0)$  startet. In diesem Fall gilt wegen (2.7):

$$\frac{\partial}{\partial t} u(X_1, t) = -a'(X_1)u(X_1, t), \quad (2.65)$$

woraus folgt

$$u(X_1, t) = u_0(X_1) \cdot e^{-a'(X_1)t}. \quad (2.66)$$

Dieser Wert kann analog zu (2.61, 2.62) für die numerische Randbedingung am Rand  $x = X_1$  verwendet werden.

## Kapitel 3

# Burger-Gleichung

### 3.1 Theoretische Aspekte und 2 Testfälle

Bisher haben wir uns ausschließlich mit linearen Gleichungen beschäftigt. In diesem Kapitel wollen wir uns mit einer eindimensionalen *nichtlinearen* hyperbolischen Gleichung befassen. Dazu betrachten wir die eindimensionalen Eulerschen Gleichungen der Gasdynamik (s. Abschnitt 1.1):

$$\varrho_t + (\varrho v)_x = 0 \quad (3.1)$$

$$(\varrho v)_t + (\varrho v^2 + p)_x = 0 \quad (3.2)$$

$$E_t + (v(E + p))_x = 0. \quad (3.3)$$

Im Falle einer *isothermen* Strömung ist der Druck  $p$  eine Funktion von  $\rho$  und die Gleichungen (3.1, 3.2) bilden ein geschlossenes System. Die Diskretisierungsverfahren für die Kontinuitätsgleichung (3.1) wurden bereits im letzten Kapitel behandelt. Wir konzentrieren uns deshalb auf die Impulsbilanz (3.2). Der Druckgradient stellt in dieser Gleichung einen Quellterm dar, dessen numerische Behandlung keinerlei Schwierigkeiten bereitet. Wir werden den Einfluß dieses Termes deshalb zunächst vernachlässigen. Die partielle Differentiation der beiden restlichen Terme führt unter Berücksichtigung von (3.1) zu:

$$\varrho_t \cdot v + \varrho \cdot v_t + (\varrho v)_x \cdot v + \varrho v \cdot v_x = 0 \iff \quad (3.4)$$

$$\underbrace{(\varrho_t + (\varrho v)_x)}_{\equiv 0} \cdot v + \varrho \cdot (v_t + v v_x) = 0 \iff \quad (3.5)$$

$$v_t + v v_x = 0. \quad (3.6)$$

Die letzte Gleichung entspricht der hyperbolischen Erhaltungsgleichung

$$u_t + f(u)_x = 0 \quad (3.7)$$

für den Fall einer *quadratischen* Stromfunktion:

$$f(u) = \frac{1}{2}u^2 \quad (3.8)$$

bis auf die Bezeichnung der unbekanntenen Funktion  $u(x, t)$ . Die Gleichung

$$\boxed{\text{BURGER – GLEICHUNG : } u_t + uu_x = 0} \quad (3.9)$$

wird in der Literatur als *Burger-Gleichung* bezeichnet und wird sehr oft zum Testen von Diskretisierungsverfahren eingesetzt, da eine Reihe von exakten Lösungen bekannt ist (s. *Whitham*, [133]). Die quadratische Stromfunktion stellt die einfachste Form der Nichtlinearität dar und läßt sich relativ leicht untersuchen. Gleichzeitig ist man bei der analytischen und numerischen Behandlung der Burger-Gleichung mit der ganzen Fülle von Problemen konfrontiert, die auch im Falle einer allgemeinen nichtlinearen Stromfunktion vorhanden sind. Wir werden uns deshalb bei der Herleitung der numerischen Methoden für die nichtlinearen hyperbolischen Gleichungen nur auf die Burger-Gleichung beschränken, zumal diese Gleichung die gleiche Art von Nichtlinearität hat, wie die Navier-Stokes-Gleichung, mit der wir uns im nächsten Kapitel beschäftigen werden.

Wir beginnen wie immer mit der analytischen Behandlung der zu untersuchenden Differentialgleichung. Die Spezifikation der Anfangsbedingung erfolgt wie üblich durch:

$$u(x, t = 0) = u_0(x), \quad x \in (-\infty, +\infty). \quad (3.10)$$

Wie im Falle einer linearen Konvektionsgleichung können Charakteristiken der Burger-Gleichung (3.9) als solche Kurven definiert werden, die der gewöhnlichen Differentialgleichung

$$x'(t) = u(x(t), t) \quad (3.11)$$

mit der Anfangsbedingung

$$x(0) = x_0 \quad (3.12)$$

genügen. Dabei kann  $x_0$  ein beliebiger Wert aus  $(-\infty, +\infty)$  sein, und definiert die Startposition der charakteristischen Kurve. Auf den ersten Blick scheint es nicht möglich zu sein, das Charakteristiken-Verfahren zur Bestimmung der Lösung der Burger-Gleichung einzusetzen, da zur Berechnung der Charakteristik durch (3.11) die exakte Lösung  $u(x, t)$  bereits vorliegen soll. Wir werden aber gleich sehen, daß es nicht notwendig ist, diese Lösung im *gesamten* Lösungsbereich zu kennen, um die Charakteristik berechnen zu können.

Für die Ableitung der Lösung  $u(x, t)$  entlang einer Charakteristik gilt:

$$\begin{aligned} \frac{d}{dt}u(x(t), t) &= \frac{\partial}{\partial t}u(x(t), t) + \frac{\partial}{\partial x}u(x(t), t) \cdot x'(t) \\ &= u_t + uu_x = 0. \end{aligned} \quad (3.13)$$

Das bedeutet, daß die Lösungsfunktion  $u(x, t)$  entlang der Charakteristik konstant ist. Dabei gilt:

$$u(x(t), t) \equiv \text{const} \equiv u(x(0), 0) = u(x_0, 0) = u_0(x_0). \quad (3.14)$$

Aus (3.11) folgt nun, daß die Charakteristik eine gerade Linie ist, deren Ableitung durch die Anfangsbedingung  $u_0(x_0)$  definiert ist.

Als Beispiel (das wir zugleich als **Testfall No.6** bezeichnen werden) betrachten wir die Burger-Gleichung für  $x \in [0, 2]$  mit der Anfangsbedingung

$$u_0(x) = x^2, \quad x \in [0, 2]. \quad (3.15)$$

Der linke Rand des Berechnungsgebietes stimmt mit der Charakteristik überein, die durch den Punkt  $x_0 = 0$  geht, so daß die Spezifikation der Randbedingung nicht notwendig ist. Auch am Rand  $x = 2$  wird keine Randbedingung vorgegeben, da dies ein Ausströmrand ist.

Es sei  $x_0$  ein beliebiger Punkt aus dem Intervall  $[0, 2]$ . Für eine Charakteristik, die vom Punkt  $x_0$  aus startet, gilt

$$x(t) = x_0 + u_0(x_0)t = x_0 + x_0^2 t. \quad (3.16)$$

Für die Lösung entlang dieser Charakteristik gilt wegen (3.14):

$$u(x(t), t) = u_0(x_0) = x_0^2. \quad (3.17)$$

Wir nehmen nun einen beliebigen Punkt  $(x, t) \in [0, 2] \times [0, +\infty)$ . Löst man die Gleichung (3.16) nach  $x_0$  auf, so stellt man fest, daß der Punkt  $(x, t)$  auf einer Charakteristik liegt, die von der Stelle  $(x_0, 0)$  startet mit

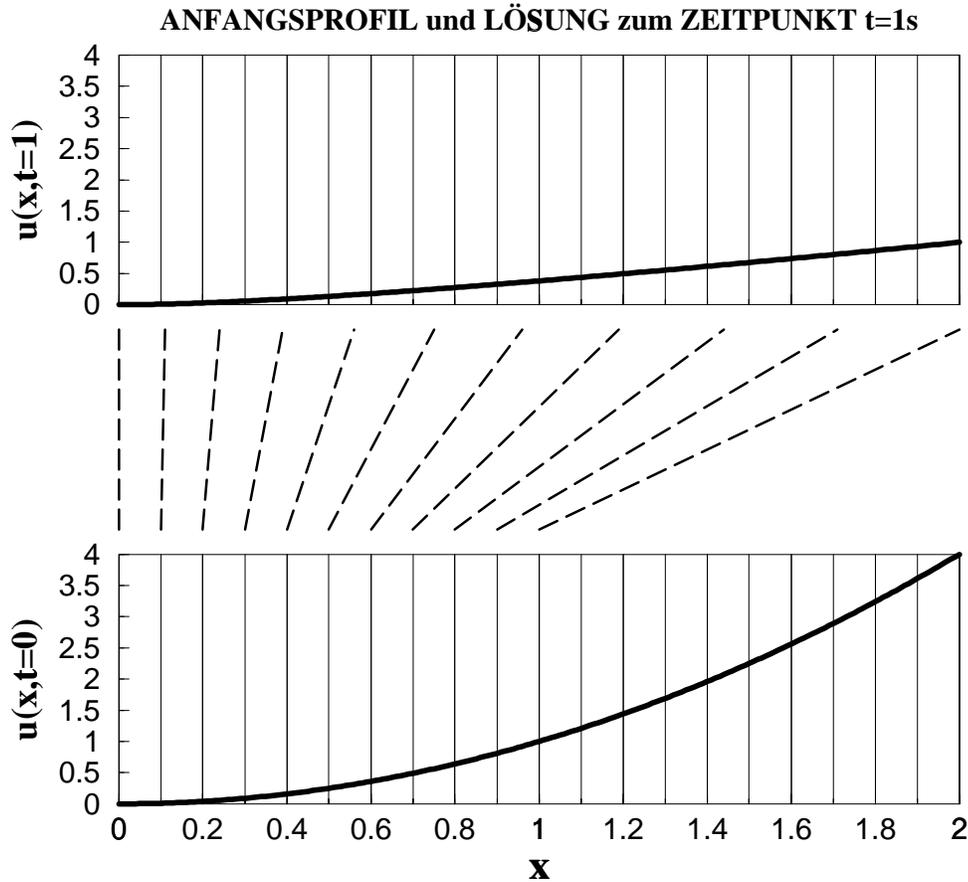
$$x_0 = \frac{-1 + \sqrt{1 + 4xt}}{2t}. \quad (3.18)$$

Einsatz des Wertes  $x_0$  in (3.17) führt zu:

$$u(x, t) = \left( \frac{-1 + \sqrt{1 + 4xt}}{2t} \right)^2 \equiv \frac{2x^2}{1 + 2xt + \sqrt{1 + 4xt}}. \quad (3.19)$$

In Abb.3.1 ist das Anfangsprofil sowie der Lösungverlauf zum Zeitpunkt  $t = 1$  für Test No.6 zusammen mit den Charakteristiken dargestellt.

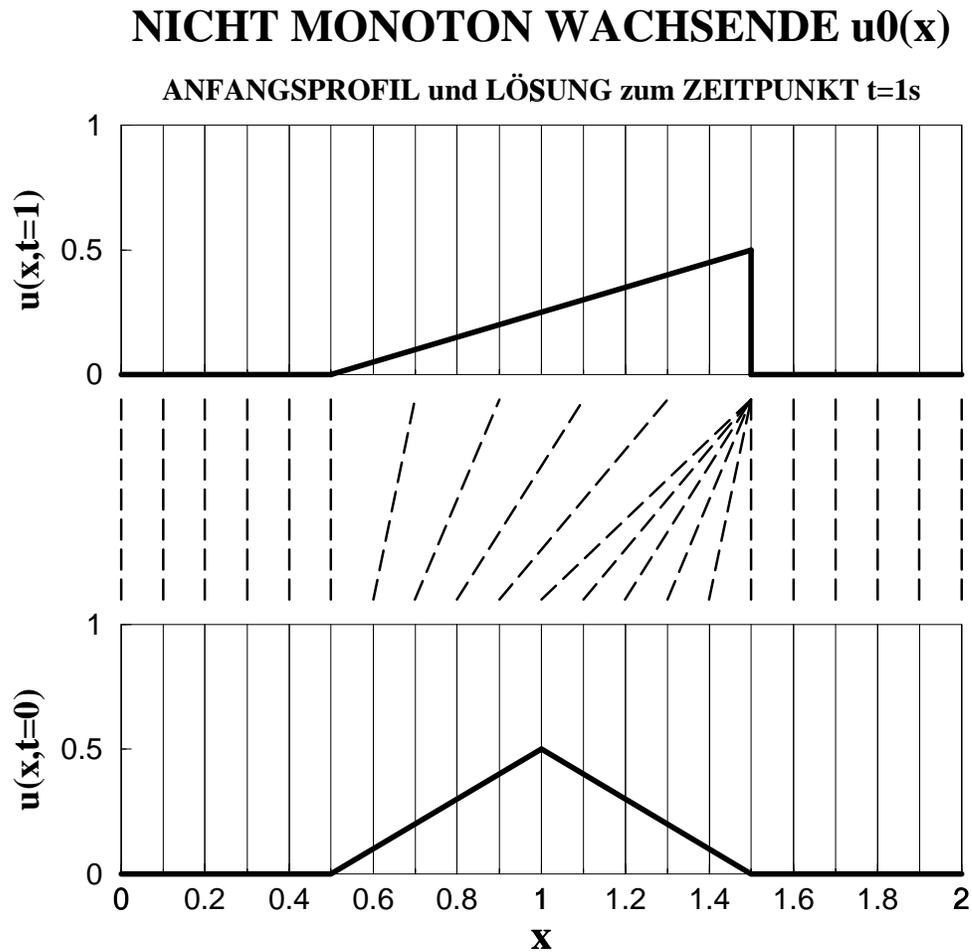
Der entscheidende Punkt beim Aufbau der Lösung mit Hilfe des Charakteristiken-Verfahrens ist die Voraussetzung, daß jeder Punkt  $(x, t)$  aus dem Lösungsbereich auf genau einer Charakteristik liegt. Diese Voraussetzung ist erfüllt, wenn die Funktion  $u_0(x)$  stetig und

**TESTFALL No.6:**

**Abbildung 3.1:** Test No. 6: Anfangsprofil, Charakteristiken und exakte Lösung zum Zeitpunkt  $t=1s$ .

monoton wachsend ist. In diesem Fall decken die Charakteristiken, die von der  $x$ -Achse starten, das gesamte Gebiet komplett ab, und zwar ohne Überschneidung. Ist die Funktion  $u_0(x)$  nicht stetig bzw. nicht monoton wachsend, so kann es passieren, daß entweder nicht das gesamte Lösungsgebiet durch die Charakteristiken abgedeckt wird, oder daß zwei oder mehr Charakteristiken sich an verschiedenen Stellen kreuzen.

In der Abb. 3.2 ist ein Beispiel für eine nicht monoton wachsende Anfangsbedingung gezeigt. Wir sehen, daß sich im Punkt  $(x, t) = (1.5, 1)$  mehrere Charakteristiken kreuzen, so daß im Lösungsprofil eine Diskontinuität („*Schock*“) entsteht. Nach dem Zeitpunkt  $t = 1$  kann das Charakteristiken-Verfahren nicht mehr eingesetzt werden. Es ist interessant, daß die Lösung zum Zeitpunkt  $t = 1$  unstetig ist, obwohl das Anfangsprofil eine stetige Funktion von  $x$  ist. Selbst wenn das Anfangsprofil eine stetig *differenzierbare* Funktion ist (was im hier gezeigten Beispiel nicht der Fall ist), deren Ableitung zumindest an einem Punkt



**Abbildung 3.2:** Nicht monoton wachsende Anfangsbedingung: Anfangsprofil, Charakteristiken und exakte Lösung zum Zeitpunkt  $t=1s$ .

negativ ist, bildet sich zum Zeitpunkt  $T_{shock} = -1 / \min u_0'(x)$  ein Schock, und die klassische Lösung der Burger-Gleichung hört auf zu existieren, da die örtliche Ableitung der Lösungsfunktion nicht mehr definiert ist ([133]). Das ist der wesentliche Unterschied zwischen den linearen und nichtlinearen hyperbolischen Gleichungen: während sich bei einer linearen Gleichung die Diskontinuität nur entlang einer Charakteristik fortpflanzen kann und die Stetigkeit der Anfangsbedingung automatisch zur Stetigkeit der Lösung im gesamten Lösungsbereich führt, kann bei einer nichtlinearen Gleichung eine Diskontinuität in der Lösung selbst dann entstehen, wenn das Anfangsprofil stetig oder sogar stetig differenzierbar ist.

Obwohl die Burger-Gleichung nach der Entstehung eines Schocks keine klassische Lösung mehr besitzt, besitzt sie eine *schwache* Lösung im Sinne von Glng. (1.18), Abschnitt 1.1. Um das Verhalten dieser schwachen Lösung qualitativ zu untersuchen, werden wir den Schock

isoliert betrachten.

Dazu untersuchen wir das Cauchy-Problem für die Burger-Gleichung mit einer stückweise konstanten Anfangsbedingung, die eine einzige Diskontinuität hat:

$$u_0(x) = \begin{cases} u_l, & \text{wenn } x < 0 \\ u_r, & \text{wenn } x > 0 \end{cases} \quad (3.20)$$

Diese Aufgabe wird in der Literatur als *Riemann-Problem* für die Burger-Gleichung bezeichnet. Die Lösung des Riemann-Problems hängt vom Verhältnis zwischen den Werten von  $u_l$  und  $u_r$  ab.

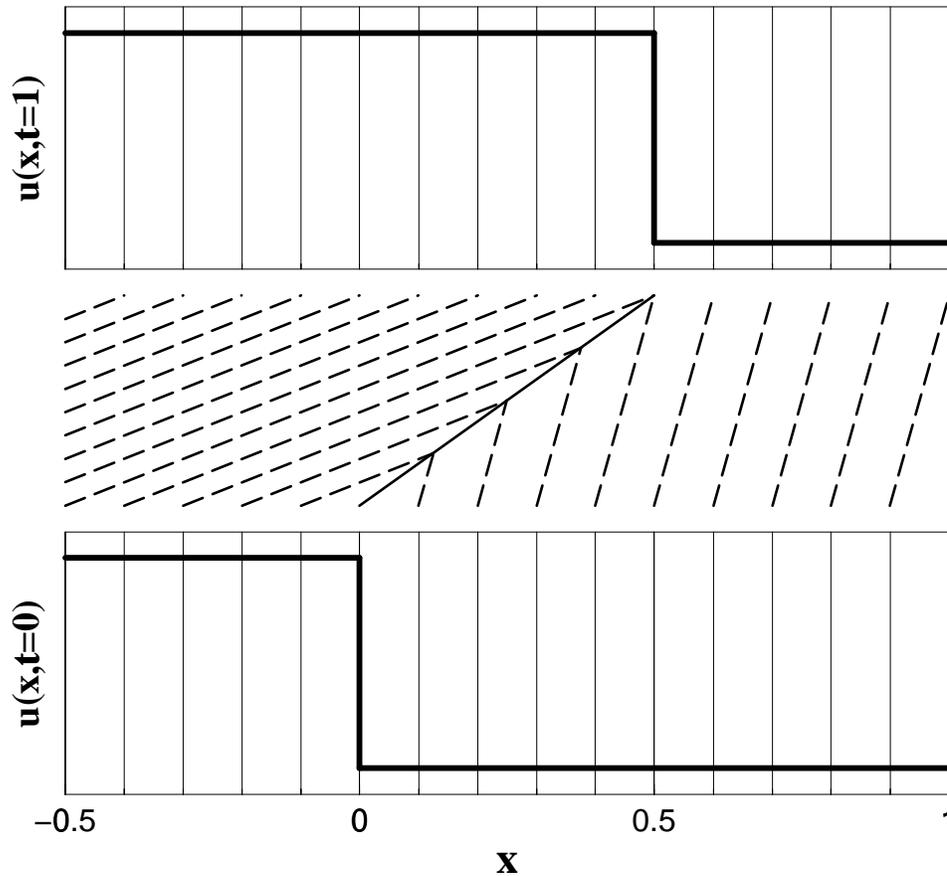
In der Abbildung 3.2 wurde die Bildung eines Schocks durch Aufsteilen der Lösungsfront skizziert. Diesem Fall entspricht das Riemann-Problem mit  $u_l > u_r$ . Wie könnte in einem solchen Fall die schwache Lösung aussehen? Da der Wert der Lösung *vor* dem Schock *höher* ist als der Wert *nach* dem Schock, „laufen die Charakteristiken in den Schock hinein“. Damit ist ein Abflachen der Front im weiteren Zeitverlauf ausgeschlossen. Auf der anderen Seite ist ein Anwachsen der Diskontinuität ebenfalls nicht möglich, denn das würde eine Zunahme der totalen Variation der Lösung bedeuten, was für hyperbolische Gleichungen der Form (3.7) ausgeschlossen ist [74]. Das bedeutet, daß die schwache Lösung die Form eines Schocks behalten muß, mit den Werten  $u_l$  und  $u_r$  vor bzw. nach dem Schock. Bezeichnen wir mit  $s$  die Geschwindigkeit, mit der sich der Schock entlang der  $x$ -Achse fortpflanzt, so sieht die Lösung des Riemann-Problems für den Fall  $u_l > u_r$  folgendermaßen aus (s. Abb. 3.3):

$$u_0(x) = \begin{cases} u_l, & \text{wenn } x < st \\ u_r, & \text{wenn } x > st \end{cases} \quad (3.21)$$

Um die Schock-Geschwindigkeit  $s$  zu bestimmen, setzen wir die Lösungsfunktion (3.21) in die Integralgleichung (1.18) ein. Wir wählen dazu  $t_1 = 0$ ,  $t_2 = 1$ . Die Größen  $x_1$  und  $x_2$  sollten so gewählt werden, daß sich der Schock während des Zeitintervalls  $[t_1, t_2]$  innerhalb von  $[x_1, x_2]$  befindet. Ohne Beschränkung der Allgemeinheit nehmen wir an, daß  $0 < s < 1$  gilt und nehmen daher  $x_1 = 0$  und  $x_2 = 1$ . Wir schreiben die Gleichung (1.18) für ausgewählte Werte von  $[x_1, x_2]$  und  $[t_1, t_2]$  sowie für den Fall einer quadratischen Stromfunktion  $f(u) = \frac{u^2}{2}$  um:

$$\begin{aligned} \int_0^1 u(x, 1) dx &= \int_0^1 u(x, 0) dx \\ &+ \int_0^1 \frac{u^2(0, t)}{2} dt - \int_0^1 \frac{u^2(1, t)}{2} dt. \end{aligned} \quad (3.22)$$

## RIEMANN-PROBLEM: $u_l > u_r$



**Abbildung 3.3:** Riemann-Problem mit  $u_l > u_r$ : Anfangsprofil, Charakteristiken und die schwache Lösung zum Zeitpunkt  $t = 1$ s für den Fall  $u_l = 1.8$ ,  $u_r = 0.2$ .

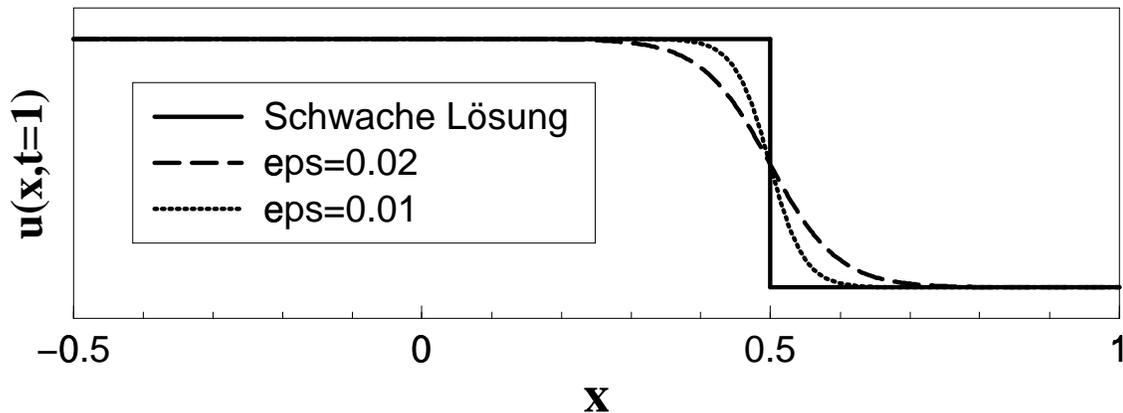
Das ist eine Erhaltungsgleichung für die Größe  $u$ . Links steht das Integral der Lösungsfunktion zum neuen Zeitpunkt  $t = 1$  über das Intervall  $[0, 1]$ . Aus (3.21) folgt, daß dieser Wert gleich  $u_l \cdot s + u_r \cdot (1 - s)$  ist. Erster Term auf der rechten Seite entspricht dem Integral der Lösungsfunktion über das Intervall  $[0, 1]$  zum alten Zeitpunkt  $t = 0$  und ist gleich  $u_r$ . Der zweite und der dritte Term beschreiben die aufintegrierten Konvektionsströme durch die Ränder des Intervalls,  $0.5u_l^2$  bzw.  $0.5u_r^2$ . Für die Schock-Geschwindigkeit  $s$  gilt daher:

$$s = \frac{0.5u_l^2 - 0.5u_r^2}{u_l - u_r} = \frac{u_l + u_r}{2}. \quad (3.23)$$

Das ist die sog. *Rankine-Hugoniot-Bedingung* für die Burger-Gleichung.

Es gibt allerdings auch eine andere Möglichkeit, an die schwache Lösung (3.21) des Riemann-Problems zu gelangen. Wir erinnern uns daran, daß wir die Gleichung (3.6) über die Eulerschen Gleichungen hergeleitet haben. Im allgemeinen mehrdimensionalen Fall

## RIEMANN–PROBLEM: $u_l > u_r$



**Abbildung 3.4:** Riemann-Problem mit  $u_l > u_r$ : Schwache Lösung von (3.9) sowie Lösungen von (3.24) für  $\epsilon = 0.02$  und  $\epsilon = 0.01$  zum Zeitpunkt  $t = 1s$ .

stellen die Eulerschen Gleichungen der Gasdynamik einen Grenzfall der Navier-Stokes-Gleichungen dar, für den Fall einer vernachlässigbar niedrigen Reibung. Analog dazu kann die Burger-Gleichung (3.9) als ein Grenzfall der *viskosen Burger-Gleichung*

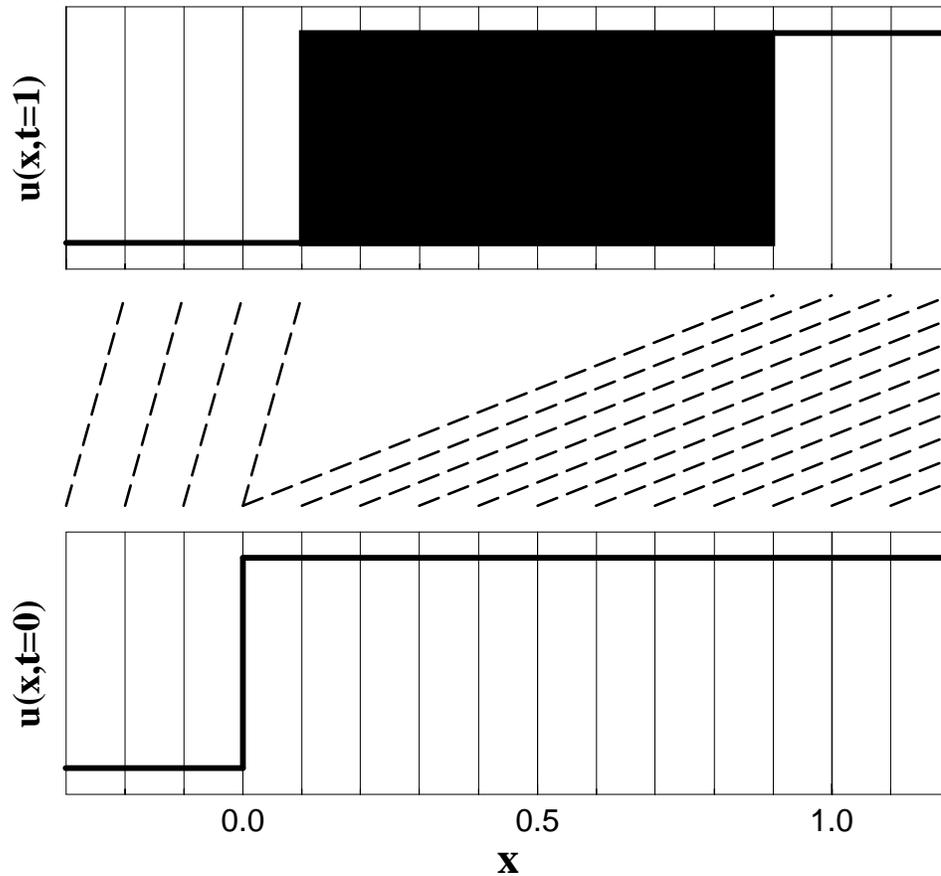
$$u_t + uu_x = \epsilon u_{xx} \quad (3.24)$$

betrachtet werden, für  $\epsilon \rightarrow 0$ . Die „nicht-viskose“ Burger-Gleichung, die wir bisher untersucht haben, ist daher nur eine Vereinfachung der physikalisch korrekteren Gleichung (3.24). Wenn der Wert von  $\epsilon$  klein und die Lösungsfunktion  $u(x, t)$  glatt ist, dann ist der Term  $\epsilon u_{xx}$  vernachlässigbar klein im Vergleich zu anderen Termen der Gleichung (3.24), und die nicht-viskose Burger-Gleichung (3.9) ist eine gute Approximation für (3.24). So würde sich die Lösung des Tests No.6 kaum verändern, wenn man anstatt von (3.9) die Gleichung (3.24) mit derselben Anfangsbedingung für ein kleines  $\epsilon$  gelöst hätte.

Bildet sich im Lösungsverlauf ein Schock, so wächst die zweite Ableitung  $u_{xx}$  viel schneller als  $u_x$  und der Term  $\epsilon u_{xx}$  kann selbst bei kleinen Werten von  $\epsilon$  nicht mehr vernachlässigt werden. Dieser Term bewahrt die Lösung vor der Entstehung einer Diskontinuität und sorgt dafür, daß selbst im Falle einer unstetigen Anfangsbedingung die Lösung für alle  $t > 0$  glatt wird. Für sehr kleine Werte von  $\epsilon$  wäre dann die Lösung (3.21) des Riemann-Problems durch die glatte Lösung ersetzt, die in der Abb. 3.4 für  $\epsilon = 0.02$  und  $\epsilon = 0.01$  exemplarisch skizziert ist. Läßt man  $\epsilon$  gegen Null gehen, so werden die Lösungsprofile steiler, und die Lösung von (3.24) nähert sich der schwachen Lösung von (3.9).

Während das Charakteristiken-Verfahren beim Riemann-Problem für den Fall  $u_l > u_r$  deswegen nicht eingesetzt werden kann, da einige Punkte des Lösungsbereichs auf mehreren

## RIEMANN-PROBLEM: $u_l < u_r$

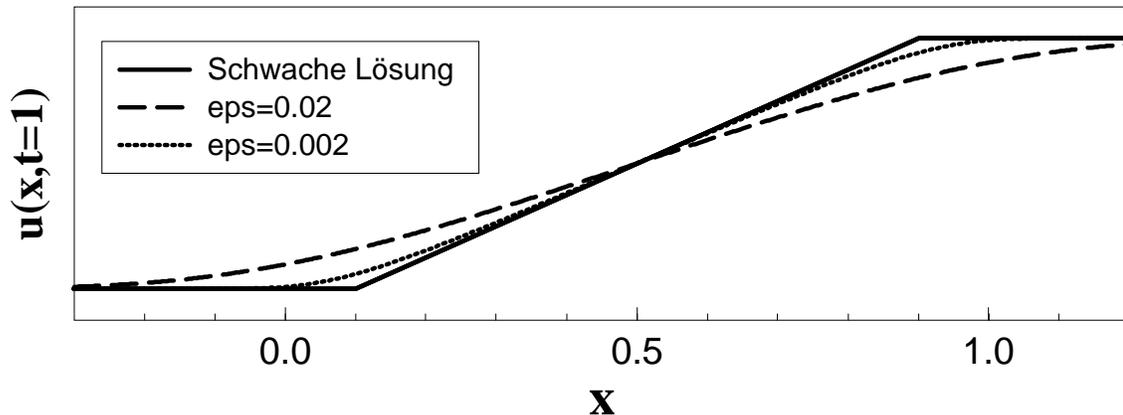


**Abbildung 3.5:** Riemann-Problem mit  $u_l < u_r$ : Anfangsprofil, Charakteristiken und die schwache Lösung zum Zeitpunkt  $t = 1s$  mit einem undefinierten Bereich (als Black-Box dargestellt). Es wurde angenommen, daß  $u_l = 0.2$ ,  $u_r = 1.8$ .

Charakteristiken gleichzeitig liegen, tritt im Falle  $u_l < u_r$  eine andere Situation ein, die den Einsatz des Charakteristiken-Verfahrens ebenfalls verhindert. Wie man aus der Abb. 3.5 sehen kann, haben die Charakteristiken, die vor dem Schock starten, eine kleinere Neigung als die, die hinter dem Schock starten. Das führt dazu, daß für jedes  $t > 0$  ein Bereich  $x \in (u_l t, u_r t)$  existiert, in welchem die Lösung nicht definiert ist, da durch solche Punkte  $(x, t)$  keine einzige Charakteristik läuft, die von der  $x$ -Achse aus startet. Selbst wenn eine schwache Lösung des Riemann-Problems existiert, kann deren Verlauf innerhalb der in Abb. 3.5 dargestellten „Black-Box“ mittels des Charakteristiken-Verfahrens *nicht* bestimmt werden.

Nimmt man an, daß die schwache Lösung wieder die Form (3.21) hat, so kann man durch Einsatz dieser Funktion in die Integralgleichung (3.22) zeigen, daß die Schock-Geschwin-

## RIEMANN–PROBLEM: $u_l < u_r$



**Abbildung 3.6:** Riemann-Problem mit  $u_l < u_r$ : Schwache Lösung (3.25) von (3.9) sowie Lösungen von (3.24) für  $\epsilon = 0.02$  und  $\epsilon = 0.002$  zum Zeitpunkt  $t = 1$ s.

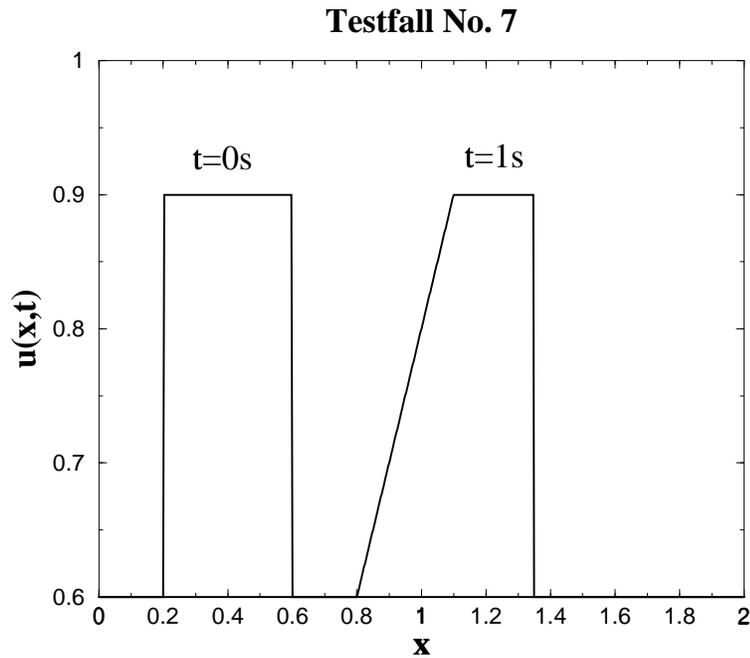
digkeit  $s$  wie im Falle  $u_l > u_r$  aus (3.23) berechnet werden kann.

Andererseits kann man wiederum die Lösung der viskosen Burger-Gleichung (3.24) für kleine  $\epsilon$ -Werte nehmen, und den Grenzwert der Lösungsfunktion bei  $\epsilon \rightarrow 0$  bilden. Man erhält in diesem Fall folgende Funktion (s. Abb. 3.6):

$$u(x, t) = \begin{cases} u_l, & \text{wenn } x < u_l t \\ x/t, & \text{wenn } u_l t < x < u_r t \\ u_r, & \text{wenn } x > u_r t \end{cases} \quad (3.25)$$

Setzt man diese Funktion in die Integralgleichung (1.18) ein, so stellt man fest, daß die Gleichung erfüllt ist. Es handelt sich hier also um eine *weitere* schwache Lösung des Riemann-Problems mit  $u_l < u_r$ . Man kann sogar zeigen, daß im Falle  $u_l < u_r$  *unendlich viele schwache Lösungen* der Burger-Gleichung existieren. Betrachtet man allerdings die „nicht-viskose“ Burger-Gleichung (3.9) lediglich als einen Grenzfall der physikalisch korrekteren *viskosen Burger-Gleichung* (3.24), so muß man nur die Lösung (3.25) als eine physikalisch sinnvolle Lösung anerkennen.

Wir wollen nun ein Testbeispiel (Test No. 7) definieren, das beide Fälle des Riemann-Problems gleichzeitig berücksichtigt. Wir beschränken uns wie immer auf das Intervall  $[0, 2]$



**Abbildung 3.7:** Testbeispiel No.7: Anfangsprofil und Lösungsverlauf zum Zeitpunkt  $t = 1s$ .

und definieren die Anfangsbedingung durch:

$$u_0(x) = \begin{cases} 0.6, & \text{wenn } 0 \leq x \leq 0.2 \\ 0.9, & \text{wenn } 0.2 < x < 0.6 \\ 0.6, & \text{wenn } 0.6 \leq x \leq 2 \end{cases} \quad (3.26)$$

und die Randbedingung durch:

$$\mu(t) = 0.6, \quad t > 0. \quad (3.27)$$

Wir sind an der Entwicklung der Lösung im Zeitintervall  $t \in [0, 1]$  interessiert. Während dieses Zeitintervalls kann die „physikalisch sinnvolle“ schwache Lösung der Burger-Gleichung folgendermaßen dargestellt werden (vgl.(3.21), (3.25)) :

$$u(x, t) = \begin{cases} 0.6, & \text{wenn } x < 0.2 + 0.6t \\ (x - 0.2)/t, & \text{wenn } 0.2 + 0.6t \leq x \leq 0.2 + 0.9t \\ 0.9, & \text{wenn } 0.2 + 0.9t < x < 0.6 + 0.75t \\ 0.6, & \text{wenn } 0.6 + 0.75t \leq x \end{cases} \quad (3.28)$$

Das Startprofil der Lösung sowie deren Verlauf zum Zeitpunkt  $t = 1$  ist in der Abb. 3.7 dargestellt.

## 3.2 Flux-Limiter-Verfahren für die Burger-Gleichung

Wie im Falle einer linearen Konvektionsgleichung mit variabler Geschwindigkeit, kann ein TVD-Verfahren für die Burger-Gleichung sowohl als ein Flux-Limiter- als auch als ein Slope-Limiter-Verfahren aufgebaut werden. Wir werden diese beiden Möglichkeiten in zwei separaten Abschnitten betrachten. Da eine Verwechslung von beiden Verfahren dadurch ausgeschlossen ist, werden wir die im Abschnitt 2.2 zu Diskriminierungszwecken eingeführten Indizes „FL“ und „SL“ diesmal weglassen.

Ein wesentlicher Unterschied zur linearen Konvektionsgleichung mit *variabler* Geschwindigkeit besteht darin, daß die totale Variation der analytischen Lösung der Burger-Gleichung mit der Zeit tatsächlich *nicht* zunehmen kann. Man kann daher beim Aufbau der Diskretisierungsverfahren für die Burger-Gleichung verlangen, daß diese die TVD-Eigenschaft (1.139) besitzen, ohne daß dadurch die physikalischen Eigenschaften der Gleichung verletzt wären.

Während von den empirisch aufgebauten TVD-Verfahren für die lineare Konvektionsgleichung nur die SL-Verfahren oszillationsfreie Lösungen produzierten, kann man bei der Burger-Gleichung davon ausgehen, daß sowohl die Flux-Limiter- als auch die Slope-Limiter-Verfahren keine Oszillationen in der numerischen Lösung verursachen werden, vorausgesetzt natürlich, daß sie die TVD-Eigenschaft besitzen.

Im Abschnitt 1.8 wurde gezeigt, daß ein konsistentes TVD-Verfahren mit einem Lipschitzstetigen Konvektionsstrom konvergent ist. Konvergenz eines Verfahrens bedeutet, daß die Differenz zwischen der numerischen und der exakten Lösung der Differentialgleichung bei feiner werdenden Gittern gegen Null geht. Dieser Begriff bedarf im Falle einer nichtlinearen Gleichung einer weiteren Präzisierung. Wie im vorigen Abschnitt gezeigt wurde, kann die Burger-Gleichung je nach Anfangsbedingung entweder eine oder gleichzeitig mehrere schwache Lösungen haben. Hat die Gleichung mehrere schwache Lösungen, so bedeutet die Konvergenz eines TVD-Verfahrens, daß der *Abstand* zwischen der numerischen Lösung und der *Menge aller schwachen Lösungen* gegen Null geht [77]. Hat die Burger-Gleichung aber mehrere schwache Lösungen, so kann davon nur eine physikalisch sinnvoll sein. Sie kann durch einen Grenzübergang  $\epsilon \rightarrow 0$  aus der Lösung  $u^\epsilon(x, t)$  der *viskosen Burger-Gleichung* (3.24) bestimmt werden. Damit ein konsistentes TVD-Verfahren nicht nur gegen die Menge *aller* schwachen Lösungen der Burger-Gleichung sondern konkret gegen diese einzige physikalisch sinnvolle Lösung konvergiert, muß dieses Verfahren noch *weitere* Bedingungen erfüllen. Eine Reihe von hinreichenden Bedingungen für die Konvergenz eines TVD-Verfahrens gegen die korrekte Lösung wird in der Literatur in Form einer sogenannten „*Entropie-Ungleichung*“ formuliert (s. dazu mehr in [46, 77]). Wir werden dieses Thema jedoch nicht weiter vertiefen. Die Untersuchung von Diskretisierungsverfahren für die Burger-Gleichung ist für uns nur ein Zwischenschritt auf dem Weg zur Behandlung

der Navier-Stokes-Gleichungen. Da diese allerdings immer einen diffusiven Term enthalten, können sie auch keine nicht-physikalischen schwachen Lösungen haben. Auch für den Testfall 7 kann gezeigt werden, daß die weiter unten vorgestellten Verfahren nur gegen die *korrekte* schwache Lösung konvergieren können. Einige dieser Verfahren dürfen jedoch *nicht allgemein* zur Lösung der *Burger-Gleichung* eingesetzt werden und müssen deshalb etwas modifiziert werden. Die entsprechenden kritischen Momente dieser Verfahren werden im Folgenden im Text explizit erwähnt.

Der Ausgangspunkt für unterschiedliche Diskretisierungsverfahren ist wie immer die einheitliche Finite-Volumen-Formulierung (1.53):

$$\frac{U_i - \bar{U}_i}{\Delta t} + \frac{F(U; i + 1/2) - F(U; i - 1/2)}{\Delta x} = 0. \quad (3.29)$$

Für die Burger-Gleichung gilt dabei:

$$F(U; i - 1/2) = \frac{1}{2}(U_{i-1/2})^2 \approx \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \frac{1}{2} u^2(x_{i-1/2}, t) dt. \quad (3.30)$$

Wir skizzieren nun das Ablauf-Diagramm zum Aufbau eines Flux-Limiter-Verfahrens:

1. Man definiere den Upwind-Konvektionsstrom  $F^{up}(U; i - 1/2)$
2. Man definiere den CDS-Konvektionsstrom  $F^{cds}(U; i - 1/2)$
3. Man definiere den TVD-Konvektionsstrom  $F^{tvd}(U; l)$  als  
 $(1 - \phi_l)F^{up}(U; l) + \phi_l F^{cds}(U; l)$
4. Man setze den TVD-Konvektionsstrom  $F^{tvd}(U; l)$  in die Finite-Volumen-Formulierung ein, und bestimme den Gewichtungsfaktor  $\phi_l$  so, daß das resultierende Verfahren die TVD-Eigenschaft hat

Einfachheitshalber gehen wir zunächst von der Annahme aus, daß die Lösungsfunktion überall positiv ist. (Die Behandlung des allgemeineren Falls wird am Ende des Abschnittes erfolgen.) Bei der Formulierung des Upwind-Konvektionsstromes wird der gesamte Strom an der stromaufwärts von  $x_{i-1/2}$  liegenden Stelle ausgewertet. Ist die Lösungsfunktion positiv, so ist auch die Stromrichtung positiv (da die Konvektionsgeschwindigkeit mit der Lösungsfunktion übereinstimmt). Wir definieren daher:

$$\frac{1}{2}(U_{i-1/2})^2 = F^{up}(U; i - 1/2) = \frac{1}{2}U_{i-1}^2. \quad (3.31)$$

Der CDS-Strom wird als Mittelwert von Konvektionsflüssen  $\frac{1}{2}U_{i-1}^2$  und  $\frac{1}{2}U_i^2$  gebildet:

$$\frac{1}{2}(U_{i-1/2})^2 = F^{cds}(U; i - 1/2) = \frac{1}{2} \cdot \left( \frac{1}{2}U_{i-1}^2 + \frac{1}{2}U_i^2 \right). \quad (3.32)$$

Für den TVD-Strom gilt daher

$$\begin{aligned} F^{tvd}(U; l) &:= (1 - \phi_l)F^{up}(U; l) + \phi_l F^{cds}(U; l) \\ &= F^{up}(U; l) + \phi_l (F^{cds}(U; l) - F^{up}(U; l)) \\ &= \frac{1}{2}U_{i-1}^2 + \frac{\phi_l}{2} \cdot \frac{U_i^2 - U_{i-1}^2}{2}. \end{aligned} \quad (3.33)$$

Setzt man diesen Konvektionsstrom in die Finite-Volumen-Gleichung (3.29) ein, so ergibt sich folgende Diskretisierung der Burger-Gleichung:

**TVD (Flux – Limiter) :**

$$\frac{U_i - \bar{U}_i}{\Delta t} + \frac{\frac{1}{2}U_i^2 + \frac{\phi_r}{2} \cdot \frac{U_{i+1}^2 - U_i^2}{2} - \frac{1}{2}U_{i-1}^2 - \frac{\phi_l}{2} \cdot \frac{U_i^2 - U_{i-1}^2}{2}}{\Delta x} = 0. \quad (3.34)$$

Wir definieren

$$\boxed{\phi_l = 0, \text{ wenn } U_i^2 - U_{i-1}^2 = 0} \quad (3.35)$$

und ansonsten

$$\boxed{\phi_l = \phi(\theta_l)}, \quad (3.36)$$

wobei  $\theta_l$  folgendermaßen definiert ist:

$$\theta_l := \frac{U_{i-1}^2 - U_{i-2}^2}{U_i^2 - U_{i-1}^2}. \quad (3.37)$$

Wir suchen nun eine solche Limiter-Funktion  $\phi(\theta)$ , daß das Verfahren (3.34) die TVD-Eigenschaft besitzt. Wir nehmen an, daß sowohl  $\theta_l$  als auch  $\theta_r$  wohldefiniert sind, und schreiben die Gleichung (3.34) unter Berücksichtigung von (3.36) folgendermaßen um:

$$U_i = \bar{U}_i - \frac{\Delta t}{2\Delta x} \left( U_i^2 + \frac{\phi(\theta_r)}{2} (U_{i+1}^2 - U_i^2) - U_{i-1}^2 - \frac{\phi(\theta_l)}{2} (U_i^2 - U_{i-1}^2) \right) \quad (3.38)$$

bzw.

$$\begin{aligned}
 U_i &= \bar{U}_i - \frac{\Delta t}{4\Delta x} \left( 2 - \phi(\theta_l) + \phi(\theta_r) \frac{U_{i+1}^2 - U_i^2}{U_i^2 - U_{i-1}^2} \right) (U_i^2 - U_{i-1}^2) \\
 &= \bar{U}_i - \frac{(U_i + U_{i-1})\Delta t}{4\Delta x} \left( 2 - \phi(\theta_l) + \frac{\phi(\theta_r)}{\theta_r} \right) (U_i - U_{i-1}) \quad (3.39)
 \end{aligned}$$

Das entspricht (1.141) mit

$$\begin{aligned}
 C_{i-1} &= \frac{(U_i + U_{i-1})\Delta t}{4\Delta x} \left( 2 - \phi(\theta_l) + \frac{\phi(\theta_r)}{\theta_r} \right) \\
 D_i &= 0
 \end{aligned}$$

Damit die Bedingung (1.142) erfüllt ist, müssen folgende Ungleichungen gelten

$$\boxed{0 \leq 2 - \phi(\theta_l) + \frac{\phi(\theta_r)}{\theta_r} \leq C} \quad (3.40)$$

Diese stimmen jedoch mit (1.171) komplett überein. Daher können alle TVD-Limiter, die für die lineare Konvektionsgleichung mit konstanter Geschwindigkeit eingesetzt wurden, auch für das Flux-Limiter-Verfahren für die Burger-Gleichung verwendet werden.

**Behandlung der Nichtlinearität.** Das Flux-Limiter-Verfahren (3.34) ist nichtlinear. Diese Nichtlinearität hat allerdings einen anderen Charakter als im Falle von TVD-Verfahren für lineare Konvektionsgleichung, die im Abschnitt 1.9 behandelt wurden. Damals waren beide Bausteine eines TVD-Verfahrens – die Upwind- und CDS-Konvektionsströme – lineare Funktionen der Gitterfunktion. Die Nichtlinearität betraf nur die Berechnung des Gewichtungsfaktors  $\phi_l$ . Durch die Aufteilung des TVD-Stroms in einen linearen Upwind-Strom und einen nichtlinearen Korrekturstrom konnte das Problem der Nichtlinearität durch quasiexplizite Behandlung des Korrekturstromes gelöst werden (s. auch Abschnitt 1.6).

Im Falle der Burger-Gleichung sind infolge der Nichtlinearität der zugrundeliegenden Stromfunktion —  $f(u) = \frac{1}{2}u^2$  — nicht nur der Korrekturstrom, sondern auch der Upwind- und der CDS-Strom nichtlinear. Eine quasi-explizite Behandlung des Korrekturstromes (s. (1.117)) löst das Problem der Nichtlinearität nicht, da durch die vollimplizite Behandlung des Upwind-Stromes die resultierende algebraische Gleichung nach wie vor nichtlinear bleibt.

Die Linearisierung des Upwind-Stromes kann entweder mit dem *Newton*-Verfahren oder mit Hilfe der *Picard*-Iteration durchgeführt werden. Die erste Methode ist z.B. in [30] ausführlich beschrieben. Wir ziehen jedoch die Picard-Iteration vor. Bei dieser Methode wird der

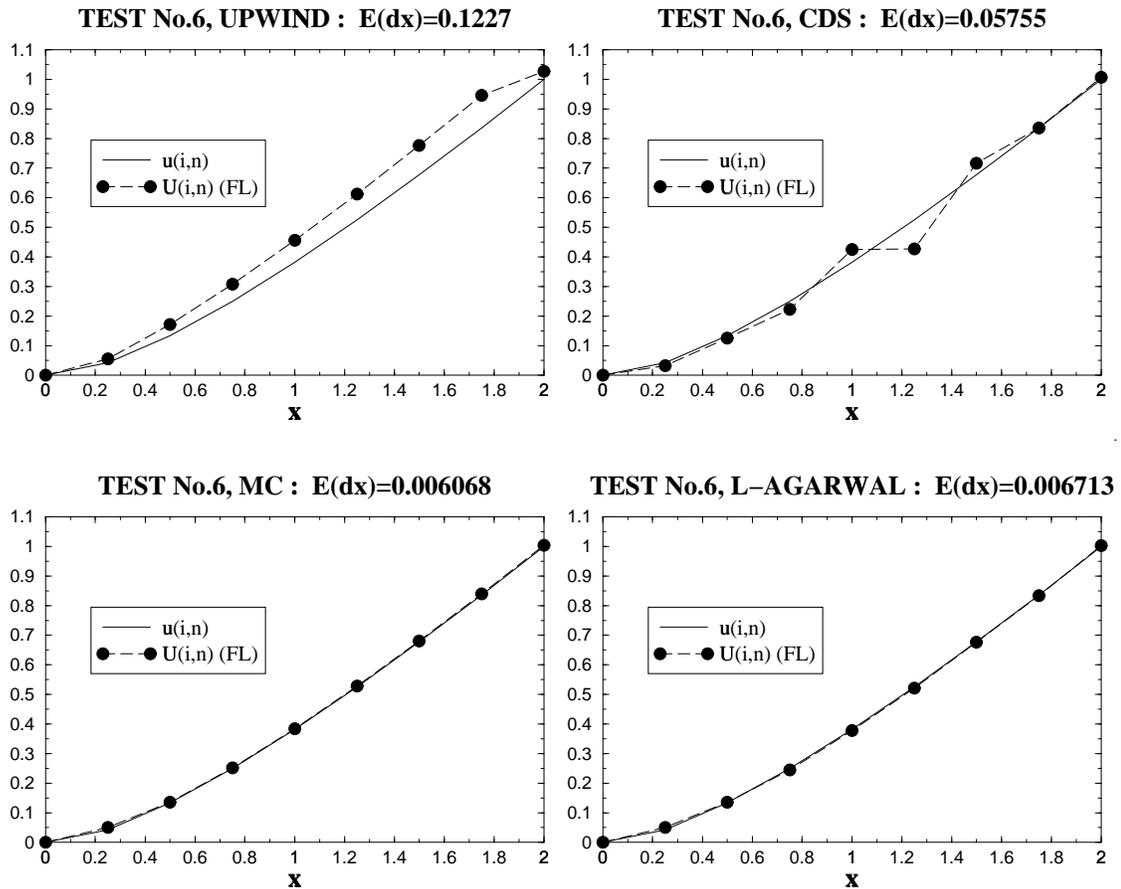


Abbildung 3.8: Test No.6: Verlauf der Lösung auf einem Gitter  $\Delta x = 0.25$  für 4 Verfahren.

quadratische Upwind-Strom  $F^{up}(U; i - 1/2) = \frac{1}{2}U_{i-1}^2$  in zwei Multiplikatoren aufgeteilt, von denen einer quasi-explizit und der anderer voll-implizit behandelt wird:

$$F^{up}(U; i - 1/2) = \frac{1}{2}U_{i-1}^{(l+1)}U_{i-1}^{(l)} \quad (3.41)$$

(vgl. die Bezeichnungen in (1.117)). Die quasi-explizite Behandlung der Konvektionsgeschwindigkeit erfordert zwar mehr Iterationen als das Newton-Verfahren, sie ist aber viel einfacher zu programmieren und steht methodisch im Einklang mit der quasi-expliziten Behandlung des Korrekturstromes.

**Testbeispiel No.6.** In der Abb. 3.8 sind die Ergebnisse für den Testfall No.6 und eine Maschenweite  $\Delta x = 0.25$  für UPWIND- und CDS-Verfahren, sowie für das TVD-Verfahren (3.34) mit MC- und AGARWAL-Limiter dargestellt. Die niedrigste Genauigkeit wird vom UPWIND-Verfahren erzielt, gefolgt von CDS-Verfahren, dessen Lösung die aus früheren Beispielen bekannten Oszillationen aufweist. Beide TVD-Verfahren erzielen dagegen hervorragende Ergebnisse. In Abb. 3.9 sind die  $E-\Delta x$ -Kurven für diese 4 Verfahren gegenüber-

### Test No.6: E-dx-Diagramm

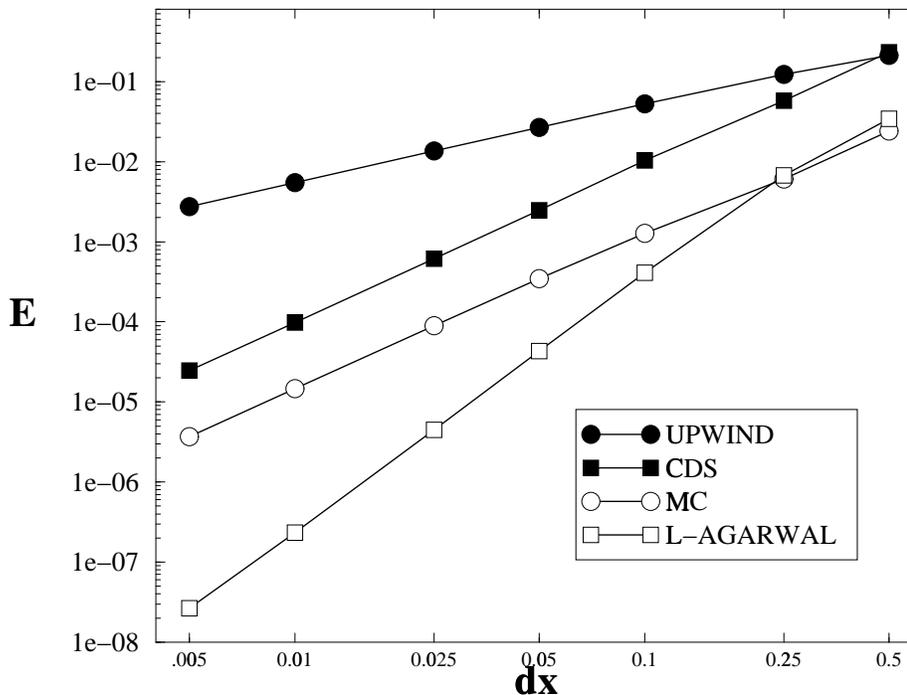


Abbildung 3.9: Test No.6: Vergleich von 4 Flux-Limiter-Verfahren untereinander.

gestellt. Man sieht, daß das MC-Verfahren die gleiche Konvergenzordnung wie das CDS-Verfahren hat, wobei der absolute Wert des Fehlers auf allen Gittern beim MC-Verfahren wesentlich niedriger ist. Der AGARWAL-Limiter erzielt wie erwartet die noch höhere Konvergenzordnung von 3.

**Testbeispiel No.7.** Richtig deutlich sind die Vorteile der TVD-Verfahren gegenüber UPWIND und CDS erst im Test No.7 zu sehen. In der Abb. 3.10 sind die Ergebnisse für 4 Verfahren auf zwei unterschiedlichen Gittern ( $\Delta x = 0.025$  und  $\Delta x = 0.01$ ) dargestellt. Man sieht, daß sowohl der MC- als auch der AGARWAL-Limiter sehr genaue Lösungen produzieren. Das UPWIND-Verfahren leidet dagegen wie im Falle der linearen Gleichung an zu starker numerischen Diffusion, während die numerische Dispersion des CDS-Verfahrens dafür sorgt, daß die Gitterverfeinerung sich kaum auf die Genauigkeit der Lösung auswirkt. In Abb. 3.11 sind die  $E$ - $\Delta x$ -Kurven für diese 4 Verfahren gegenübergestellt. Man sieht, daß die Fehlerkurve für das CDS-Verfahren nur ganz langsam abnimmt. Das UPWIND-Verfahren erreicht die Konvergenzordnung von 0.8. Das L-AGARWAL- und das MC-Verfahren haben eine etwas höhere Konvergenzordnung von 1, die absoluten Werte des numerischen Fehlers sind aber wesentlich niedriger, als bei der UPWIND-Lösung. Während im Test No.6 der AGARWAL-Limiter zu einer wesentlich genaueren Lösung als der MC-

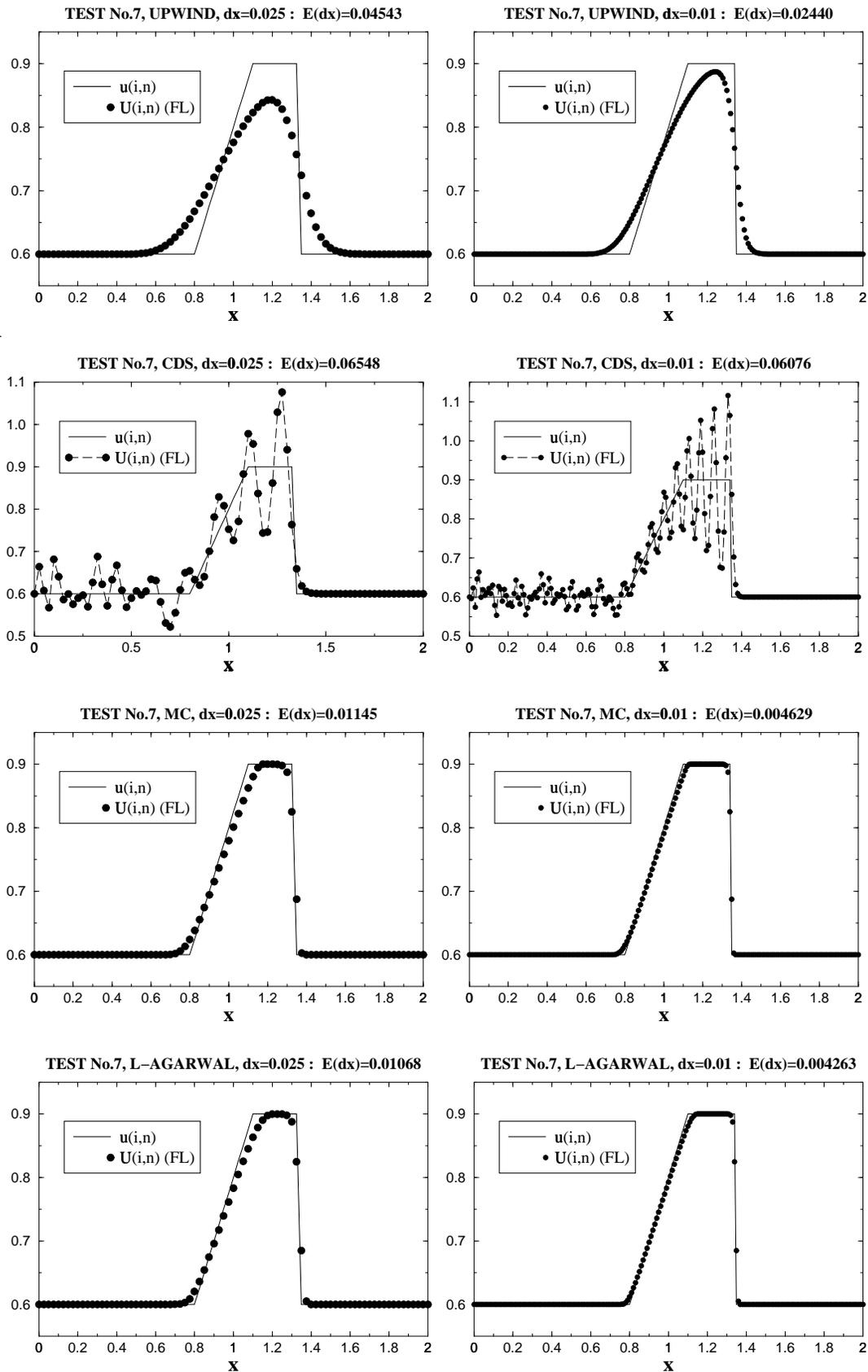


Abbildung 3.10: Test No.7: Verlauf der Lösung auf zwei unterschiedlichen Gittern für 4 Verfahren.

### Test No.7: E-dx-Diagramm

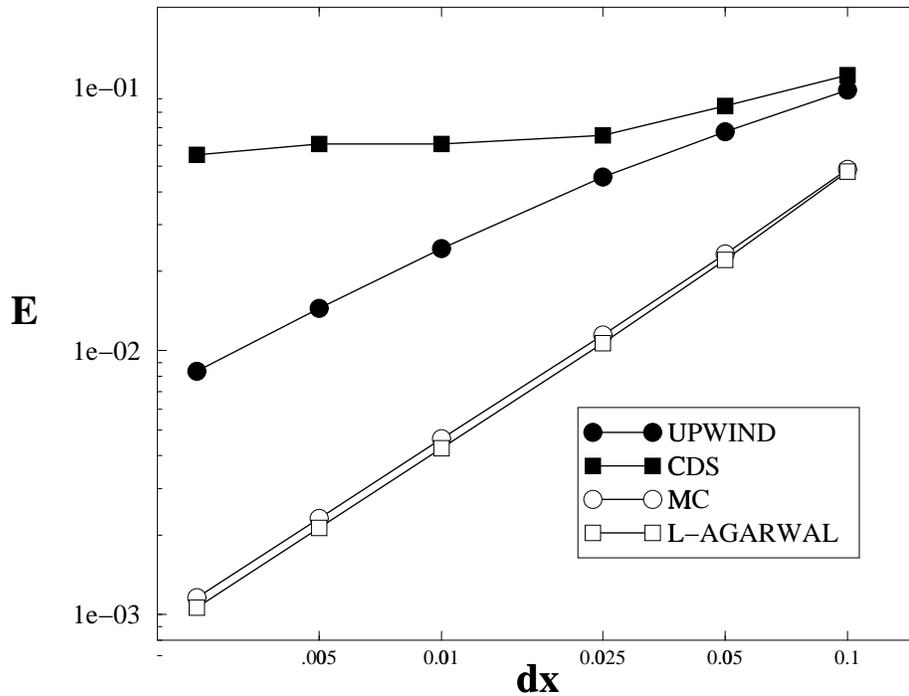


Abbildung 3.11: Test No.7: Vergleich von 4 Flux-Limiter-Verfahren untereinander.

Limiter geführt hat, ist im Testfall No.7 die Genauigkeit beider Verfahren in etwa gleich gut.

**Beliebige Lösungsfunktion.** Wir haben bisher bei der Herleitung des Flux-Limiter-Verfahrens für die Burger-Gleichung angenommen, daß die Lösungsfunktion  $u(x, t)$  überall positiv ist. Wir wollen nun auf diese Annahme verzichten und die Herleitung des TVD-Verfahrens auf den Fall einer beliebigen (was das Vorzeichen betrifft) Lösungsfunktion verallgemeinern.

Wir beginnen mit der Definition des Upwind-Stromes  $F^{up}(U; i - 1/2)$ . Dabei wird der gesamte Strom an der stromaufwärts von  $x_{i-1/2}$  liegenden Stelle ausgewertet. Entscheidend dabei ist die Stromrichtung an der Stelle  $x_{i-1/2}$ , hier ist die Gitterfunktion jedoch nicht definiert. Um die Stromrichtung an der Stelle  $x_{i-1/2}$  zu bestimmen, verwenden wir daher eine lineare Interpolation für die Lösungsfunktion zwischen den Gitterpunkten  $x_{i-1}$  und  $x_i$  und definieren zu diesem Zweck die Geschwindigkeit  $U_l$  an der Stelle  $x_{i-1/2}$  als

$$U_l := \frac{1}{2}(U_{i-1} + U_i). \quad (3.42)$$

Das Vorzeichen von  $U_l$  ist nun entscheidend, ob der gesamte Strom an der Stelle  $x_{i-1}$  oder

an der Stelle  $x_i$  ausgewertet wird:

$$\frac{1}{2}(U_{i-1/2})^2 = F^{up}(U; i - 1/2) = \begin{cases} \frac{1}{2}U_{i-1}^2, & \text{wenn } U_l \geq 0 \\ \frac{1}{2}U_i^2, & \text{wenn } U_l < 0 \end{cases} \quad (3.43)$$

Der so definierte Upwind-Strom, sowie das darauf basierende TVD-Verfahren, genügt keiner der bekannten Entropie-Ungleichungen (s. die Diskussion am Anfang des Abschnittes) und kann dazu führen, daß die numerische Lösung gegen eine nichtphysikalische schwache Lösung konvergiert. Ein entsprechendes Beispiel wird in [77] angegeben. Durch eine leichte Modifikation der Definition (3.43) kann die Erfüllung einer Entropie-Ungleichung erzwungen werden, das resultierende TVD-Verfahren wäre dadurch allerdings wesentlich komplizierter geworden. Ein interessierter Leser sei auf das Buch von *Hirsch* [46] verwiesen, in dem die Problematik der Konvergenz eines TVD-Verfahrens gegen eine „falsche“ schwache Lösung ausführlich diskutiert wird. Für die Navier-Stokes-Gleichungen, die neben den konvektiven auch die diffusiven Terme enthalten, ist dieses Problem nicht relevant.

Wir setzen die Konstruktion des TVD-Verfahrens fort. An der Definition des CDS-Konvektionsstromes (3.32) ändert sich gegenüber dem Fall einer positiven Lösungsfunktion nichts, da dieser stromrichtungsunabhängig ist. Der TVD-Strom wird wie in (3.33) durch

$$\begin{aligned} F^{tvd}(U; l) &:= (1 - \phi_l)F^{up}(U; l) + \phi_l F^{cds}(U; l) \\ &= F^{up}(U; l) + \phi_l(F^{cds}(U; l) - F^{up}(U; l)) \end{aligned}$$

definiert, wobei die konkrete Darstellung diesmal vom Vorzeichen der Geschwindigkeit  $U_l$  abhängt:

$$F^{tvd}(U; l) = \begin{cases} \frac{1}{2}U_{i-1}^2 + \frac{\phi_l}{2} \cdot \frac{U_i^2 - U_{i-1}^2}{2}, & \text{wenn } U_l \geq 0 \\ \frac{1}{2}U_i^2 - \frac{\phi_l}{2} \cdot \frac{U_i^2 - U_{i-1}^2}{2}, & \text{wenn } U_l < 0 \end{cases} \quad (3.44)$$

Auch bei der Definition der Größe  $\theta_l$  wird die Stromrichtung an der Stelle  $x_{i-1/2}$  berücksichtigt:

$$\theta_l := \frac{U_I^2 - U_{I-1}^2}{U_i^2 - U_{i-1}^2}. \quad (3.45)$$

wobei Index  $I$  wie folgt definiert wird:

$$I = \begin{cases} i - 1, & \text{wenn } U_l \geq 0, \\ i + 1, & \text{wenn } U_l < 0. \end{cases} \quad (3.46)$$

Es gilt folgendes

**Theorem 8** *Genügt die Limiter-Funktion den Voraussetzungen des Theorems 5 (Abschnitt 1.9), so besitzt das resultierende Verfahren die TVD-Eigenschaft.*

**Beweis.** Siehe Anhang A.5.  $\square$

### 3.3 Slope-Limiter-Verfahren für die Burger-Gleichung

Ein Slope-Limiter-Verfahren unterscheidet sich von einem Flux-Limiter-Verfahren dadurch, daß man den Limiter zur Berechnung der Approximation der unbekannt Variable an beiden Rändern der Kontrollvolumina verwendet, und diese Approximation anschließend in die Stromfunktion  $f(u)$  einsetzt.

Wir skizzieren zunächst das Ablauf-Diagramm zum Aufbau eines Slope-Limiter-Verfahrens:

1. Man definiere die Upwind-Approximation  $U_{i-1/2}^{up}$
2. Man definiere die CDS-Approximation  $U_{i-1/2}^{cds}$
3. Man definiere die TVD-Approximation  $U^{tvd}(U; l)$  als
 
$$(1 - \phi_l)U_{i-1/2}^{up} + \phi_l U_{i-1/2}^{cds}$$
4. Man setze die TVD-Approximation  $U^{tvd}(U; l)$  in die Stromfunktion  $f(u)$  ein und erhalte so den TVD-Konvektionsstrom  $F^{tvd}(U; l) = f(U^{tvd}(U; l))$

Die einzelnen Schritte zur Definition der TVD-Approximation  $U^{tvd}(U; l)$  unterscheiden sich nur wenig von dem Fall einer linearen Konvektionsgleichung mit variabler Geschwindigkeit und wurden in Abschnitt 2.3 ausführlich dokumentiert. Der einzige Unterschied liegt darin, daß die Stromrichtung an der Stelle  $x_{i-1/2}$  nicht durch die vorgegebene Geschwindigkeit  $a_{i-1/2}$  sondern durch die Lösungsfunktion selbst ( $U_{i-1/2} = U_l = \frac{1}{2}(U_{i-1} + U_i)$ ) bestimmt wird.

Im Falle der linearen Konvektionsgleichung wurde die Approximation  $U^{tvd}(U; l)$  anschließend mit der Geschwindigkeit  $a_{i-1/2}$  multipliziert, was eine kompakte einheitliche Darstellung des Upwind- und des Korrekturstroms ermöglichte (2.58, 2.60). Bei der Burger-Gleichung müssen dagegen die Fälle  $U_l \geq 0$  und  $U_l < 0$  separat behandelt werden. Exem-

plarisch betrachten wir den Fall  $U_i \geq 0$ . Es gilt

$$F^{tvd}(U; l) = f(U^{tvd}(U; l)) = \frac{1}{2} \left( U_{i-1} + \phi_l \frac{U_i - U_{i-1}}{2} \right)^2 \quad (3.47)$$

Dieser Konvektionsstrom wird folgendermaßen aufgespalten:

$$\begin{aligned} F^{tvd}(U; l) &= \frac{1}{2} U_{i-1}^2 + \frac{\phi_l}{2} \frac{U_i - U_{i-1}}{2} \left( 2U_{i-1} + \phi_l \frac{U_i - U_{i-1}}{2} \right) \\ &=: F^{up}(U; l) + F^{dc}(U; l). \end{aligned} \quad (3.48)$$

Der Korrekturstrom kann nun quasi-explizit behandelt und der Upwind-Strom mit Hilfe der Picard-Iteration linearisiert werden.

**Theorem 9** *Unter den Voraussetzungen des Theorems 5 (Abschnitt 1.9) besitzt das resultierende Verfahren die TVD-Eigenschaft.*

**Beweis.** Siehe Anhang A.6.□

Nun wenden wir uns den Testbeispielen zu.

**Untersuchung der Testfälle.** In der Abb. 3.12 sind die  $E$ - $\Delta x$ -Kurven für das *Testbeispiel No.6* für UPWIND- und CDS-Verfahren, sowie für das TVD-Verfahren (A.36) mit MC- und AGARWAL-Limiter dargestellt. Wie im Flux-Limiter-Fall (vgl. Abb. 3.9), erzielen beide TVD-Verfahren eine höhere Genauigkeit als UPWIND- und CDS-Verfahren.

Vergleicht man die  $E$ - $\Delta x$ -Kurven der Flux-Limiter- und Slope-Limiter-Versionen des L-AGARWAL-Verfahrens, so stellt man fest, daß im Gegensatz zu L-AGRAWAL(FL)-Verfahren, das eine Konvergenzordnung von 3 für diesen Testfall erreicht hat, die Slope-Limiter-Version des Verfahrens lediglich eine Konvergenzordnung von 2 hat. Der Grund für die niedrigere Konvergenzordnung der Slope-Limiter-Variante ist der gleiche wie im Falle einer linearen Konvektionsgleichung mit variabler Geschwindigkeit und wurde im Abschnitt 2.2 ausführlich erläutert.

Bei den CDS- und MC-Verfahren kann dagegen ein gegenläufiger Effekt beobachtet werden. Die Konvergenzordnung von beiden Versionen dieser Verfahren ist zwar dieselbe (2), die absoluten Werte des numerischen Fehlers sind bei Slope-Limiter-Versionen um Faktor 3 (CDS) bzw. 8 (MC) niedriger. Dadurch erreicht die Slope-Limiter-Version des MC-Verfahrens eine sogar bessere Genauigkeit als das L-AGRAWAL(SL)-Verfahren.

In der Abb. 3.13 sind die  $E$ - $\Delta x$ -Kurven von 4 Slope-Limiter-Verfahren für das *Testbeispiel No.7* gegenübergestellt. Der Verlauf der Kurven ist dabei fast identisch mit dem der Flux-Limiter-Versionen dieser Verfahren (vgl. Abb. 3.11), so daß sich keine eindeutige Präferenzen für die eine oder andere Version der Verfahren ableiten lassen.

### Test No.6: E-dx-Diagramm

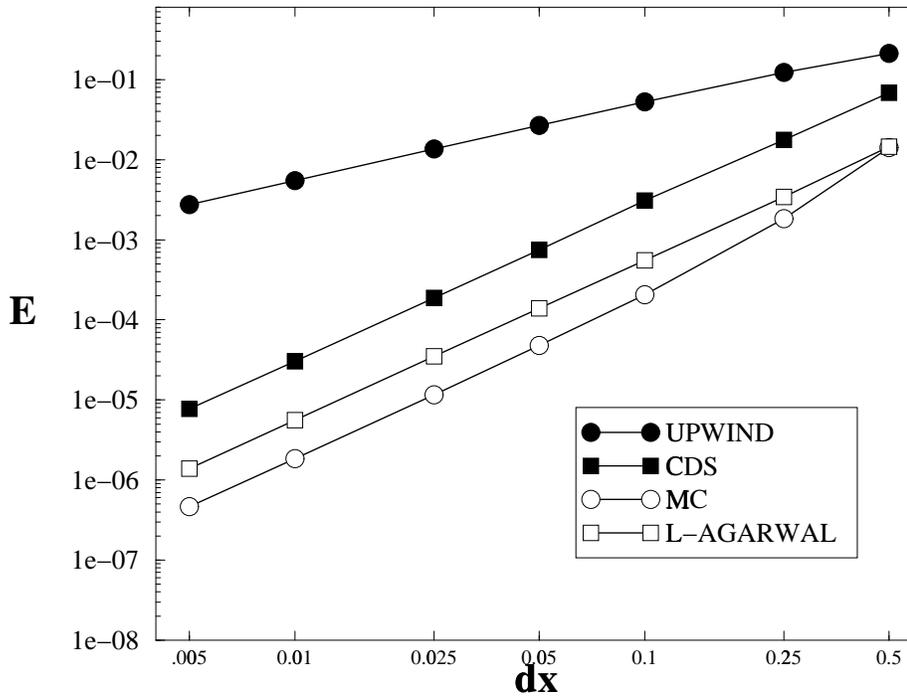


Abbildung 3.12: Test No.6: Vergleich von 4 Slope-Limiter-Verfahren untereinander.

**Zusammenfassung.** In den letzten beiden Abschnitten wurden zwei Möglichkeiten zum Aufbau eines TVD-Verfahrens für die nichtlineare Burger-Gleichung vorgestellt. Es wurde gezeigt, daß es durch eine geeignete Wahl des Limiters erreicht werden kann, daß sowohl die Flux-Limiter- als auch die Slope-Limiter-Verfahren die TVD-Eigenschaft besitzen. Die mit diesen Verfahren berechneten Lösungen weisen keine Oszillationen auf, welche beim Einsatz von nichtlimitierten Verfahren höherer Ordnung entstehen können. Gleichzeitig erreichen diese TVD-Verfahren eine Konvergenzordnung von mindestens 2 für glatte Lösungsfunktionen.

Obwohl die dargestellten TVD-Verfahren eine Reihe von Vorteilen gegenüber den herkömmlichen nichtlimitierten Verfahren besitzen, werden sie in der hier vorgestellten Form zur Lösung von nichtlinearen Gleichungen der Strömungsmechanik äußerst selten eingesetzt. Dafür gibt es eine Reihe von Gründen.

Der wichtigste Grund ist wohl der, daß die Behandlung von TVD-Verfahren numerisch sehr aufwendig ist. Während im Falle von Slope-Limiter-Verfahren für *lineare* Konvektionsgleichung mit variabler Geschwindigkeit eine kompakte Darstellung des UPWIND- sowie des Korrekturstromes (s. (2.58, 2.60)) eine einfache und übersichtliche Programmierung der resultierenden Diskretisierungsformel ermöglichte, ist eine ähnliche Vereinfachung im Falle

### Test No.7: E-dx-Diagramm

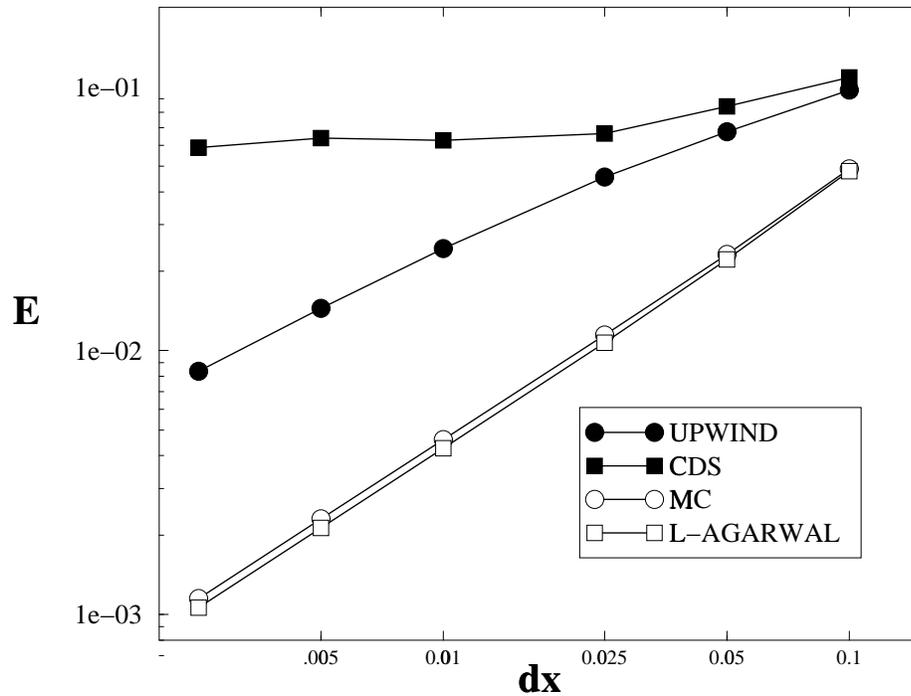


Abbildung 3.13: Test No.7: Vergleich von 4 Slope-Limiter-Verfahren untereinander.

der nichtlinearen Gleichung nicht möglich. Hier müssen die Fälle  $U_l > 0$ ,  $U_l < 0$ ,  $U_r > 0$ ,  $U_r < 0$  separat behandelt werden, wobei jedes Mal eine andere Darstellung sowohl für den Upwind-Anteil als auch für den Korrekturstrom (insbesondere bei Slope-Limiter-Verfahren, vgl. (3.48)) resultiert. Das führt sowohl zu einem hohen Programmieraufwand als auch zu einem rechenzeitintensiven und nicht sehr übersichtlichen Programm.

Auf der anderen Seite weisen einige nichtlimitierten Verfahren höherer Ordnung (wie z.B. QUICK, AGARWAL) ein Konvergenzverhalten auf, das mit dem der TVD-Verfahren vergleichbar ist, wobei der Programmieraufwand wesentlich niedriger ist. Der Einsatz von nichtlimitierten Verfahren hat zwar einen negativen Einfluß auf die Optik der Lösung, zu einem Zusammenbruch des Programms führt dies jedoch nicht: eine leicht oszillierende Geschwindigkeitsverteilung ist nicht so kritisch wie eine negative Konzentration oder eine negative turbulente kinetische Energie. Daher ist der Einsatz von QUICK-Verfahren zur Diskretisierung von Navier-Stokes-Gleichungen immer noch sehr verbreitet.

Wir wollen im nächsten Abschnitt allerdings ein weiteres Verfahren vorstellen, das zwar kein reines TVD-Verfahren ist (d.h. die TVD-Eigenschaft *nicht* besitzt), dafür aber wesentlich einfacher zu handhaben ist und gleichzeitig weniger Oszillationen als die herkömmlichen nichtlimitierten Verfahren in der numerischen Lösung verursacht. Wir werden die-

ses Verfahren daher als „*QTVD-Verfahren*“ (Quasi-TVD-Verfahren) bezeichnen. Zu den wesentlichen Vorteilen des QTVD-Verfahrens zählt unter anderem, daß die linearen und nicht-linearen Gleichungen auf die gleiche Weise behandelt werden, und daß die meisten Programme, in denen eine Upwind-Diskretisierung bereits implementiert ist, durch eine leichte Modifikation „QTVD-fähig“ gemacht werden können.

### 3.4 Splitting-Verfahren und QTVD-Diskretisierung

Der einzige Unterschied zwischen allen bisher diskutierten Diskretisierungsverfahren für die Burger-Gleichung lag in der unterschiedlichen Darstellung des numerischen Konvektionsstromes:

$$F(U; i - 1/2) = \frac{1}{2}(U_{i-1/2})^2 \approx \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \frac{1}{2} u^2(x_{i-1/2}, t) dt. \quad (3.49)$$

Gleichzeitig haben alle Flux-Limiter- und Slope-Limiter-Verfahren eines gemeinsam – der quadratische Ausdruck  $\frac{1}{2}(U_{i-1/2})^2$  wurde in allen diesen Verfahren als Ganzes approximiert, d.h. beide Faktoren des Produktes  $(U_{i-1/2})^2 = U_{i-1/2} \cdot U_{i-1/2}$  wurden gleich behandelt.

Man betrachte z.B. die CDS-Approximation. Beim Flux-Limiter-CDS-Strom wird der Mittelwert von Konvektionsflüssen  $\frac{1}{2}U_{i-1} \cdot U_{i-1}$  und  $\frac{1}{2}U_i \cdot U_i$  gebildet (s. (3.32)), während beim Slope-Limiter-CDS-Strom zunächst die CDS-Approximation  $U_{i-1/2}^{cds} = \frac{1}{2}(U_{i-1} + U_i)$  definiert und anschließend mit sich selbst multipliziert wird :

$$F^{cds,sl}(U; i - 1/2) = \frac{1}{2}U_{i-1/2}^{cds} \cdot U_{i-1/2}^{cds}. \quad (3.50)$$

Diese Vorgehensweise ist zwar notwendig, um die TVD-Eigenschaft des aus UPWIND- und CDS-Strom gebildeten TVD-Verfahrens zu gewährleisten, bei den meisten Programmen, die nur die Implementierung von den herkömmlichen nichtlimitierten Verfahren höherer Ordnung (CDS, LUDS, QUICK) enthalten, geht man allerdings anders vor.

Die unbekannte Funktion  $u$  hat in der Burger-Gleichung eine doppelte Bedeutung. Einerseits, bestimmt sie die *Geschwindigkeit*, mit welcher der Konvektionstransport stattfindet. Gleichzeitig ist sie diejenige *Größe*, welche durch die Konvektion *transportiert wird*. Dementsprechend kann man auch bei der Approximation des numerischen Konvektionsstromes  $\frac{1}{2}U_{i-1/2} \cdot U_{i-1/2}$  einen Faktor als die Konvektionsgeschwindigkeit ( $U_{i-1/2}^{vel}$ ) und den anderen Faktor als die Erhaltungsgröße selbst auffassen, und für beide Faktoren unterschiedliche Interpolationen verwenden:

$$F(U; i - 1/2) = \frac{1}{2}U_{i-1/2}^{vel} \cdot U_{i-1/2}. \quad (3.51)$$

Da dabei der numerische Konvektionsstrom in zwei Teile aufgespalten wird, kann man alle Verfahren, die die Darstellung (3.51) verwenden, auch als „*Splitting-Verfahren*“ bezeichnen. Üblicherweise wird bei einem Splitting-Verfahren für diejenige Komponente, die die Konvektions*geschwindigkeit* beschreibt, einheitlich die lineare Interpolation eingesetzt  $U_{i-1/2}^{vel} = \frac{1}{2}(U_{i-1} + U_i) = U_l$ :

$$F(U; i - 1/2) = \frac{1}{2}U_l \cdot U_{i-1/2}. \quad (3.52)$$

Die Approximation der Größe  $U_{i-1/2}$  kann je nach Verfahren unterschiedlich erfolgen. Einsatz von  $U_{i-1/2} = U_{i-1/2}^{up}$  führt zu der UPWIND-Variante des Splitting-Verfahrens, die CDS-Variante wird durch Einsatz von  $U_{i-1/2} = U_{i-1/2}^{cds}$  erzeugt. Die resultierende Diskretisierungsformel sieht in diesem Fall folgendermaßen aus:

$$\boxed{\text{CDS(SPLITTING)} : \frac{U_i - \bar{U}_i}{\Delta t} + \frac{0.5U_r \frac{U_i + U_{i+1}}{2} - 0.5U_l \frac{U_{i-1} + U_i}{2}}{\Delta x} = 0.} \quad (3.53)$$

Man beachte, daß sich diese Formel vom CDS(SL)-Verfahren (2.29) für die lineare Gleichung nur dadurch unterscheidet, daß anstatt von  $a_{i+1/2}$  und  $a_{i-1/2}$  diesmal Terme  $0.5U_r$  und  $0.5U_l$  verwendet werden. Setzt man die Approximation  $U_{i-1/2} = U_{i-1/2}^{tvd}$  in den Ausdruck (3.52) ein, so führt das zu einem Verfahren, das zwar keine TVD-Eigenschaft besitzt, aber dennoch ein TVD-ähnliches Verhalten aufweist. Wir bezeichnen den resultierenden Konvektionsstrom als Quasi-TVD (QTVD):

$$F^{qtvd}(U; i - 1/2) = \frac{1}{2}U_l \cdot U_{i-1/2}^{tvd}. \quad (3.54)$$

Im Gegensatz zu Flux-Limiter- und Slope-Limiter-TVD-Strömen, die in den vorigen 2 Abschnitten definiert wurden, gilt für den QTVD-Strom unabhängig vom Vorzeichen von  $U_l$  folgende einheitliche Darstellung:

$$F^{qtvd}(U; i - 1/2) = F^{up}(U; i - 1/2) + F^{dc}(U; i - 1/2) \quad (3.55)$$

mit

$$F^{up}(U; i - 1/2) = \max(0.5U_l, 0) \cdot U_{i-1} - \max(-0.5U_l, 0) \cdot U_i \quad (3.56)$$

$$F^{dc}(U; i - 1/2) = \phi_l \cdot \frac{1}{2} \cdot |0.5U_l| \cdot (U_i - U_{i-1}). \quad (3.57)$$

Auch diese Darstellung unterscheidet sich vom TVD(SL)-Strom für die lineare Gleichung lediglich durch die Verwendung des Termes von  $0.5U_l$  anstelle von  $a_{i-1/2}$  (vgl. (2.57, 2.58, 2.59)).

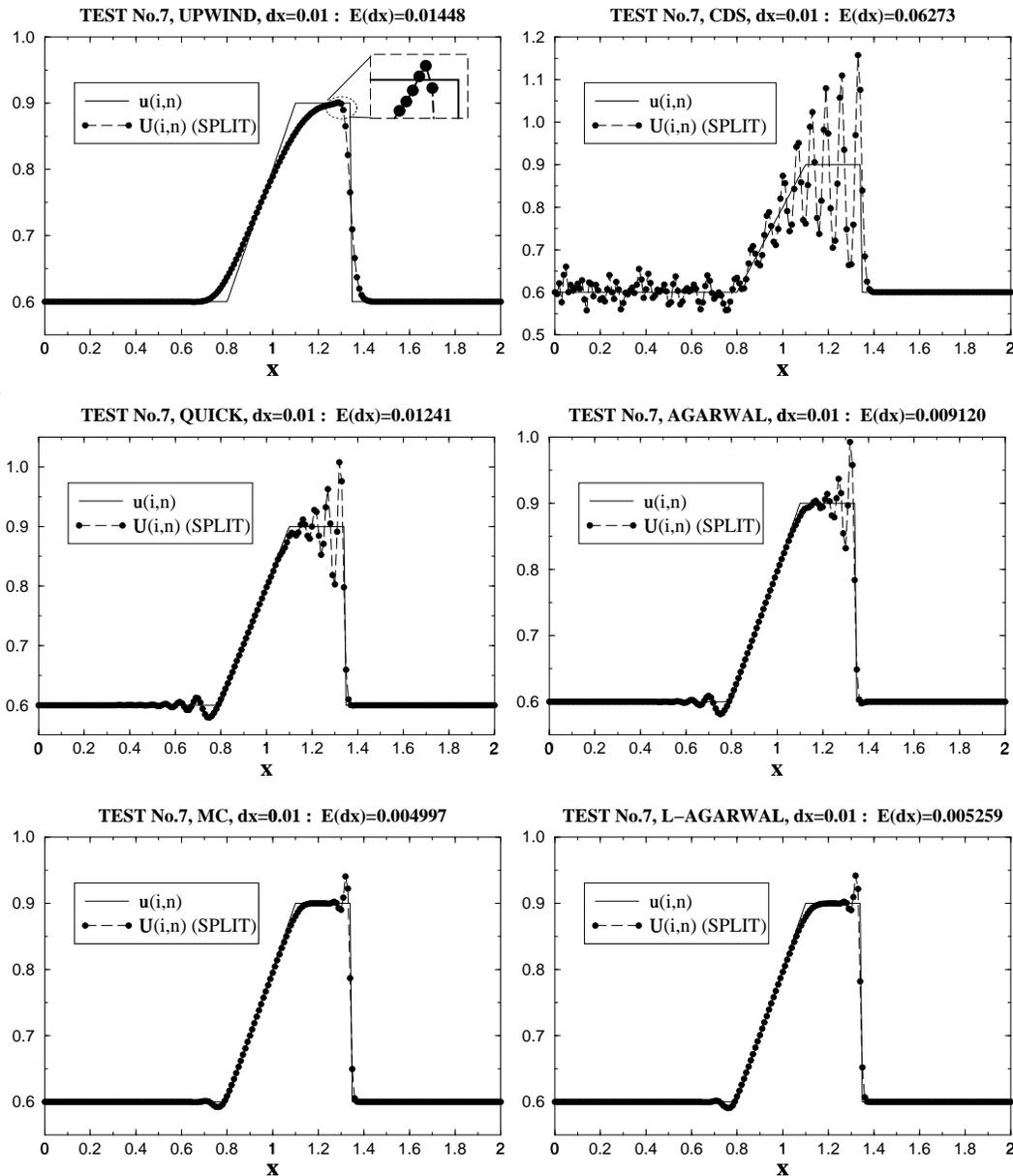


Abbildung 3.14: Test No.7: Verlauf der Lösung auf einem Gitter  $\Delta x = 0.01$  für 6 Splitting-Verfahren.

Setzt man in einem Programm das TVD(SL)-Verfahren für alle linearen Gleichungen und das QTVD-Verfahren für die Gleichungen zu Bestimmung der Geschwindigkeitskomponenten ein, so erfolgt die Berechnung der Koeffizienten für alle Modellgleichungen nach einem einheitlichen Schema, was sowohl die Übersichtlichkeit des Codes erhöht, als auch die Wahrscheinlichkeit eines Programmierfehlers reduziert. Liegt außerdem eine Programmvariante vor, in der eine Upwind(Splitting)-Diskretisierung bereits implementiert ist, so kann dieses Programm auf einfache Weise auf das QTVD-Verfahren umgestellt werden, indem die Korrekturströme als Quellterme der

### Test No.7: E-dx-Diagramm

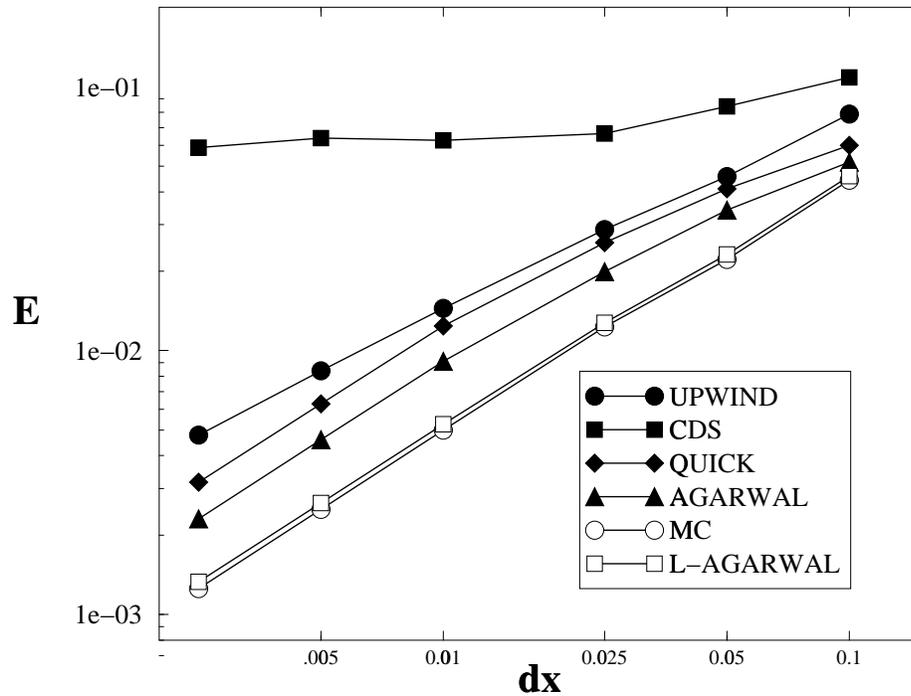


Abbildung 3.15: Test No.7: Vergleich von 6 Splitting-Verfahren untereinander.

#### Upwind-Diskretisierung hinzugefügt werden.

Das ist der Grund, warum in den meisten kommerziellen Softwarepaketen das QTVD-Verfahren implementiert wird, obwohl in den Paketbeschreibungen von einer TVD-Diskretisierung gesprochen wird, die oszillationsfreie Lösungen produziert.

Daß tatsächlich aber Oszillationen auftreten, können wir den Lösungsprofilen für den Testfall No.7 entnehmen, die mit QTVD-Verfahren mit MC- und AGARWAL-Limitern berechnet wurden (s. Abb. 3.14). Interessant ist, daß selbst die UPWIND-Version des Splitting-Verfahrens in diesem Testfall leichte Oszillationen im Lösungsprofil produziert (Abb. 3.14, ganz oben, links). Gleichzeitig sieht man, daß die Oszillationen in den QTVD-Lösungen weniger stark ausgeprägt sind als in den mit CDS-, QUICK- und AGARVAL-Verfahren berechneten Profilen. Die weitgehende Unterdrückung von Oszillationen beim Einsatz von QTVD-Verfahren führt gleichzeitig zu einem besseren Konvergenzverhalten im Vergleich zu nichtlimitierten Verfahren höherer Ordnung (Abb. 3.15). Vergleicht man den Verlauf von  $E-\Delta x$ -Kurven der QTVD-Versionen von MC- und L-AGARWAL-Verfahren mit dem der Flux- und Slope-Limiter-Varianten dieser Verfahren (Abbildungen 3.11 und 3.13) so kann man feststellen, daß alle drei Varianten ein sehr ähnliches Konvergenzverhalten aufweisen, wobei die QTVD-Versionen am einfachsten zu implementieren sind. Wir werden daher

für alle Simulationen, die im zweiten Teil dieser Arbeit vorgestellt werden, ausschließlich die QTVD-Diskretisierung für die nichtlinearen Navier-Stokes-Gleichungen verwenden, die im Abschnitt 4.4 noch ausführlich beschrieben wird.

Wir beenden hiermit die Betrachtung von eindimensionalen hyperbolischen Gleichungen und fahren im nächsten Kapitel mit der Untersuchung der linearen und nichtlinearen Gleichungen in *zwei* Dimensionen fort.