

5. Numerik

5.1 Einleitung

Das hier beschriebene Verfahren habe ich entwickelt, um solche Anfangs-Randwertprobleme zu lösen, die einen Knick beinhalten, d.h. in einigen Stellen mögen die vorgegebenen Daten aus $C^0 \setminus C^1$ sein. Eine derartige Störung breitet sich, solange es nicht zum Schock kommt, entlang einer Charakteristik aus. Daher scheint Integration in charakteristischer Richtung sinnvoll zu sein. Andererseits möchte man zu festen Zeiten die Lösung gewinnen und deshalb nicht auf einem charakteristischen Gitter arbeiten. Ein solches **Hybrid-Verfahren**, das Lösungen in einem Rechteckgitter ermittelt, hat Hartree in [Hart58] angegeben und Titze [Titz89] um eine Nachiteration erweitert. Dabei werden Daten in den Zielpunkten ermittelt, indem man in näherungsweise charakteristischer Richtung auf die alte Zeitschicht zurückblickt und von dem dort interpolierten Wert ausgehend vorwärts integriert. Dadurch gibt es aber ein Verrunden von Sprüngen, wenn man z.B. $\partial_t u + \frac{1}{2} \partial_x u = 0$ mit einem Sprung am Anfang numerisch löst.

Dagegen bleibt für dieses Beispiel bei meinem hier vorzustellenden Verfahren das Profil der Anfangsfunktion exakt erhalten, weil das Gitter in charakteristischer Richtung mitbewegt wird. Bei mehreren Unbekannten hat man dabei mehrere Charakteristikenscharen zu berücksichtigen, so daß **jede**

Unbekannte in jeder Zeitschicht ihre eigenen Gitterpunkte hat. In der neuen Zeitschicht wird jede Unbekannte interpoliert, wobei ein Verrunden entfällt, wenn man nur über Teilintervalle integriert, in denen die Lösung glatt ist.

Das Verfahren ist ein Hybridverfahren für diagonale Systeme der speziellen Form

$$\begin{aligned} \partial_t u_j(t, x) &= f_j(t, x, (\mathbf{u}, \mathbf{v})(t, x)) \\ \partial_t v_k(t, x) + \lambda_k(t, x, \mathbf{u}(t, x)) \partial_x v_k(t, x) &= g_j(t, x, (\mathbf{u}, \mathbf{v})(t, x)) \end{aligned} \quad (5.1)$$

mit $\mathbf{u} = (u_1, \dots, u_{K_{\mathbf{u}}})$, $\mathbf{v} = (v_1, \dots, v_{K_{\mathbf{v}}})$. Für jedes k sei λ_k entweder strikt positiv oder strikt negativ.

Gesucht ist eine Lösung (\mathbf{u}, \mathbf{v}) für das Anfangsrandwertproblem (5.1) mit (5.2) und (5.3) auf $[t_0, T] \times [x_{\min}, x_{\max}]$, wobei $T > t_0$ eine Konstante ist. Gegebene Anfangswerte mögen also ebenso angenommen werden

$$(\mathbf{u}, \mathbf{v})(t_0, \cdot) = (\mathbf{u}_{\text{AW}}, \mathbf{v}_{\text{AW}}) \text{ auf } [x_{\min}, x_{\max}] \quad (5.2)$$

wie gegebene Randwerte auf $[t_0, T]$

$$v_k(\cdot, x_{\min}) = (\mathbf{v}_{\text{RW}})_k \quad (\text{bzw. } v_k(\cdot, x_{\max}) = (\mathbf{v}_{\text{RW}})_k) \quad (5.3)$$

für jedes k mit $\lambda_k > 0$ (bzw. < 0).

Man beachte, daß λ nur von dem Teil \mathbf{u} der Lösung (\mathbf{u}, \mathbf{v}) abhängt, der trivial propagiert wird. Sofern \mathbf{u} stetig ist, gibt es deshalb auch dann keine Schocks, wenn \mathbf{v} unstetig ist. Sprünge in einem v_k treten im allgemeinen genau dort auf, wo \mathbf{u} knickt, also entlang der sogenannten **Knick-Charakteristik**. Das ist eine charakteristische Kurve mit der Eigenschaft, daß mindestens ein u_j entlang jeder die Charakteristik transversal schneidenden Kurve links- und rechtsseitig verschiedene Steigungen hat.

Wie man ein derartiges System 5.1 aus gewissen hyperbolischen Systemen der Form

$$\partial_t \mathbf{u}(t, x) + \mathbf{A}(t, x, \mathbf{u}(t, x)) \cdot \partial_x \mathbf{u}(t, x) = \mathbf{b}(t, x, \mathbf{u}(t, x)) \quad (5.4)$$

durch Diagonalisieren erzeugen kann, wurde in Kapitel 4 dargestellt.

Das Verfahren ist in zeitlicher Richtung zunächst explizit und von erster Ordnung. Durch eine Korrektur wird es implizit und durch eine Nachiteration läßt sich der Fehler deutlich reduzieren.

Zunächst wird nun die Notation eingeführt und das Verfahren in seiner einfachsten Form dargestellt.

5.2 Notation und Gitterbewegung

In jeder Zeitschicht $t = t^n$ wird die Funktion $v_k(t, \cdot)$ durch ihre Werte in $X_k^n = \{x_{k1}^n, x_{k2}^n, \dots, x_{kN^n}^n\}$ repräsentiert und gespeichert. Dabei ist n als oberer Index zu lesen, welcher die Zeitrichtung diskretisiert, während unten im Index einerseits ggf. die Komponente der Unbekannten und andererseits der Ortspunkt stehen. Diese Knoten x_{ki}^n wandern im Verlaufe der zeitlichen Entwicklung in (näherungsweise) charakteristischer Richtung. Zwischenwerte erhält man durch eine **Interpolation**, die Knicke/Sprünge respektiert, sofern sie vorgegeben sind oder detektiert¹ wurden. Hat etwa $v_k(t, \cdot)$ genau in $x = x_{ki}^n$ einen Sprung, so erfolgen Interpolationen höchstens auf den Teilintervallen $[x_{k1}^n; x_{ki}^n]$ und $[x_{ki}^n; x_{kN^n}^n]$.

¹Aus glatten Anfangsdaten können nach endlicher Zeit nicht nur Knicke, sondern auch Schocks entstehen. Letztere würde mein Programm daran erkennen, daß sich Charakteristiken derselben Unbekannten v_k schneiden, was zum Programmabbruch führt. Neu entstehende Knicke könnte ein weiter entwickeltes Programm aufgrund des Verlaufes der neuen Näherung zu erkennen versuchen. Ich habe aber eine derartige Knick-Detektion nicht weiter betrachtet

Die Näherung von $\mathbf{v}(t^n, \cdot)$ im ortsfesten Gitterpunkt x_i sei durch \mathbf{v}_i^n symbolisiert. Die Gitterpunkte x_{ki}^n der Unbekannten v_k wandern beim Zeitschritt von $t = t^n$ zu $t^{n+1} = t^n + \Delta t$ zunächst durch

$$x_{ki}^{n+1} = x_{ki}^n + \Delta t \cdot \lambda_{ki}^n \quad (5.5)$$

mit $\lambda_{ki}^n = \lambda(t^n, x_{ki}^n, \mathbf{u}_{ki}^n)$, wobei \mathbf{u}_{ki}^n die Approximation von $\mathbf{u}(t^n, \cdot)$ im Gitterpunkt x_{ki}^n darstellt. Numerische Näherungswerte für $\mathbf{u}(t^n, \cdot)$ liegen aber nur im (ortsfesten) Gitter $X_0 = \{x_0, x_1, \dots, x_{N_0}\}$ vor. Zur Ermittlung der \mathbf{u}_{ki}^n bedarf es daher einer Interpolation, die im folgenden beschrieben wird.

5.3 Interpolation

Wesentlich ist, daß niemals über Knicke hinwegintegriert werden soll, um ein Verrunden derselben zu verhindern. Sofern am Anfang bereits Knicke in $\mathbf{u}(0, \cdot)$ vorliegen, muß das Gitter so gebaut werden, daß für jedes unstetige v_k in dieser Stelle ein Gitterpunkt vorliegt. Dort ist dann ein oberer (und ein unterer) v_k -Wert zu speichern. Oben (bzw. unten) heißt dabei der für die zeitliche Vorwärts- (bzw. Rückwärts-)integration zu verwendende Wert, also etwa der linke (bzw. rechte) Grenzwert im Falle $\lambda_k > 0$.

Grundsätzlich wird mit kubischen Polynomen interpoliert, und zwar zwischen den mittleren von je vier Knotenpunkten, sofern kein Knick beteiligt ist. Zwischen Knick und dem Nachbarn des jeweiligen Nachbarknotens wird nur mit 2. Ordnung inter- bzw. extrapoliert, um Oszillationen zu vermeiden (siehe Abbildung 5.1).

Ähnlich wird vorgegangen, wenn für jedes k aus im jeweiligen Gitter X_k^n vorliegenden Näherungswerten für $v_k(t^n, \cdot)$ die Werte \mathbf{v}_i^n im \mathbf{u} -Gitter X_0 zu ermitteln sind.

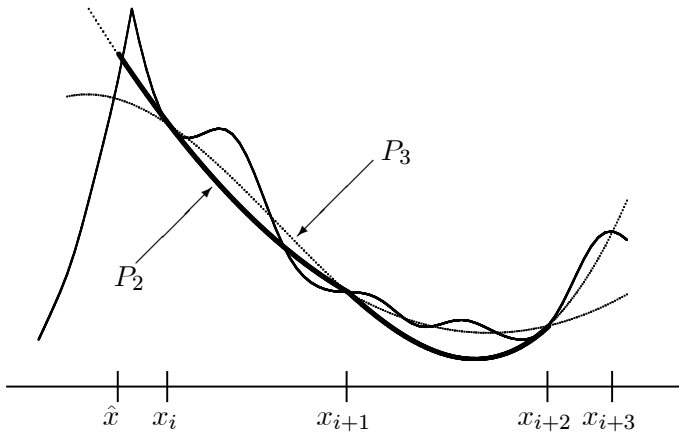


Abb. 5.1: In $x = \hat{x}$ liege (näherungsweise) ein Knick vor. Sei P_2 (bzw. P_3) das quadratische (bzw. kubische) Interpolations-Polynom zu den Stützstellen $\{x_i, x_{i+1}, x_{i+2}\}$ (bzw. $\{x_i, x_{i+1}, x_{i+2}, x_{i+3}\}$). Dann werden P_2 auf $[\hat{x}; x_{i+1}]$ und P_3 auf $[x_{i+1}; x_{i+2}]$ benutzt. Die beiden Polynome sind im jeweils wirksamen Bereich fett und sonst gepunktet dargestellt. Durchgezogen erscheint ein möglicher, hier aber frei erfundener Verlauf der echten Lösung

5.4 Der Zeitschritt

Für den Zeitschritt ist zunächst ein einfacher Euler-Schritt vorgesehen, der aber optional (siehe Abschnitt 5.5 um eine Korrektur und eine Nachiteration erweitert wird.

0. **Voraussetzung:** In x_{ki}^n seien u_{ki}^n und v_{ki}^n gegeben, wobei $(v_{\hat{k}}^n)_{ki}$ für $\hat{k} \neq k$ aus dem Gitter X_k^n interpoliert wurde. Es gelte $\{x_{\min}, x_{\max}\} \subset X_k^n$ für jedes k .
1. **λ -Ermittlung:** In jedem v_k -Gitterpunkt aus X_k^n wird die charakteristische Richtung

$$\lambda_{ki}^n = \lambda(t^n, x_{ki}^n, u_{ki}^n)$$

aus den dort vorliegenden Funktionswerten ermittelt.

2. **bewegtes Gitter X_k^{n+1} :** Sei \hat{X}_k^{n+1} die Menge aller

$$\hat{x}_{ki}^{n+1} := x_{ki}^n + \Delta t \cdot \lambda_{ki}^n \quad \text{mit} \quad \hat{x}_{ki}^{n+1} \in]x_{\min}, x_{\max}[.$$

Sei m_k deren Mächtigkeit $|\hat{X}_k^{n+1}|$. Dann sei

$$X_k^{n+1} := \hat{X}_k^{n+1} \cup \{x_{\min}, x_{\max}\},$$

wobei die beiden Randpunkte \hat{x}_{k0}^{n+1} und $\hat{x}_{km_k+1}^{n+1}$ heißen mögen (siehe Abbildung 5.2). Unterstellt² wird, daß für k mit $\lambda_k^0 > 0$ (bzw. < 0) auch alle λ_{ki}^n echt positiv (bzw. negativ) sind. Außerdem gebe es keine Überschneidungen der Charakteristiken, d.h. es gelte $\hat{x}_{ki}^{n+1} < \hat{x}_{ki+1}^{n+1}$ für alle $i = 0, 1, \dots, m_k$.

3. **u-Euler-Schritt:** Neue Funktionswerte

$$u_{ji}^{n+1} = u_{ji}^n + \Delta t \cdot f_j(t^n, x_i, u_i^n, v_i^n)$$

²Eventuelle Verstöße dagegen werden vom Programm erkannt und zur Beendigung desselben benutzt.

ermitteln, wobei $\mathbf{u}_i^n = (u_1, u_2, \dots, u_{N_0})_i^n$ die Näherung der Lösung \mathbf{u} von (5.1) der alten Zeitschicht $t = t^n$ im ortsfesten Gitter X_0 sei.

Sonderfall: Falls der Integrationsweg $[t^n, t^{n+1}] \times \{x_i\}$ eine (genäherte) **Knick**charakteristik **schneidet**, so wird zunächst nur bis zum Knick integriert. Die Fortsetzung erfolgt dann mit einer gesonderten Integration (siehe Abbildung 5.3)

4. **v-Euler-Schritt:** Analog wird ein neues \mathbf{v}^{n+1} durch

$$v_{ki}^{n+1} = v_{ki}^n + \Delta t \cdot g_k(t^n, x_{ki}^n, \mathbf{u}_{ki}^n, \mathbf{v}_{ki}^n)$$

für jedes k und $i = 1, 2, \dots, m_k$ errechnet, wobei \mathbf{u}_{ki}^n die interpolierte Näherung der Lösung in der alten Zeitschicht $t = t^n$ im i -ten Knoten des zu v_k gehörigen Gitters X_k^n ist.

5. **Randwertkorrektur:**

- a) Randauslauf: Entsprechend Abbildung 5.2 wird ins Gitter X_k^n ein Knoten eingefügt, der in der neuen Zeitschicht $t = t^{n+1}$ auf dem rechten (bzw. linken) Rand für $\lambda_k > 0$ (bzw. < 0) endet. Dessen Fußpunkt x_K^* ist zu ermitteln und mit Näherungswerten durch Interpolation seiner Nachbarwerte zu versehen.
- b) Randeinlauf: Im Falle $\lambda_k > 0$ (bzw. < 0) wird in der neuen Zeitschicht $t = t^{n+1}$ der Knoten $\hat{x}_{k0}^{n+1} = x_{\min}$ (bzw. $\hat{x}_{k, m_k+1}^{n+1} = x_{\max}$) neu eingefügt. Dort ist des Anfangsrandwertproblems Randwertvorgabe

$$v_k(t^{n+1}, x_{\min}) = (\mathbf{v}_{\text{RW}})_k$$

(bzw. $v_k(t^{n+1}, x_{\max}) = (\mathbf{v}_{\text{RW}})_k$) aus (5.3) einzusetzen.

6. **Interpolieren der v_k -Werte:** Mit der in Abschnitt 5.3 beschriebenen Vorgehensweise werden die v_k -Werte vom zugehörigen bewegten Gitter X_k^{n+1} ins ortsfeste u-Gitter X_0 übertragen³.
7. **Korrektur der u-Werte:** Eine optionale Korrektur der u-Werte muß ggf. an dieser Stelle erfolgen, weil hier in beiden Zeitschichten im Gitter X_0 sowohl u als auch v näherungsweise bekannt sind.
8. **Interpolieren der u-Werte:** Mit dem in Abschnitt 5.3 beschriebenen Verfahren werden nun für jedes v_k die Werte u_{ki}^{n+1} aus den Näherungswerten u_{ji}^{n+1} des festen Gitters X_0 interpoliert und in das zugehörige mitschwimmende Gitter X_k^{n+1} übertragen.

Nach Abschluß von Schritt 8. sind die in Schritt 0. für Zeitschicht t^n genannten Voraussetzungen wieder für den nächsten Schritt erfüllt, wenn man von n zu $n + 1$ übergeht.

Hinweis beim Auftreten mehrerer v_k -Unbekannter

Die bei Abbildung 5.3 beschriebene stückweise Integration ist analog in Schritt 4. für v_k einzusetzen, wenn der Integrationsweg eine (genäherte) Knickcharakteristik zu $v_{\hat{k}}$ mit $\hat{k} \neq k$ schneidet⁴. Analog ist vorzugehen bei jeder anderen Integration, sofern eine Knickcharakteristik den Integrationsweg schneidet, sei es in Abbildung 5.3 bei der Ermittlung von $v_{\hat{k}}^*$ oder bei Integrationen im Zuge der im folgenden Abschnitt beschriebenen Verbesserungen des Zeitschritts.

³Sollte zufällig im Ziel-Ort ein Sprung vorliegen, wird wegen der evtl. Nachiteration der untere Wert übertragen. Die Vorwärts-Iteration im nächsten Zeitschritt muß wieder auf den oberen Wert zugreifen, es handelt sich somit um den Grenzfall $\tau = 0$ des in Abbildung 5.3 skizzierten Vorgehens.

⁴Dabei können übrigens die Vorzeichen von λ_k und $\lambda_{\hat{k}}$ durchaus gleich sein

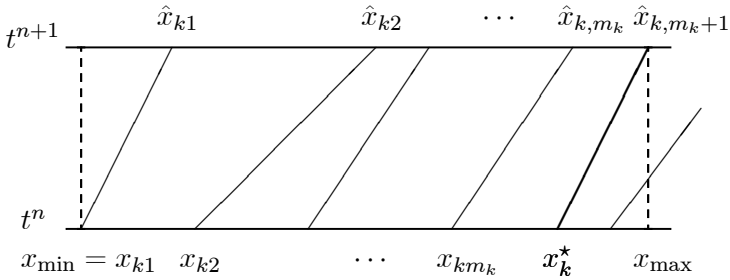


Abb. 5.2: Skizze für den Fall $\lambda_k^0 > 0$:

Der zur Zeitschicht $t = t^n$ (bzw. t^{n+1}) gehörige Index wird hier unterdrückt. Der Fußpunkt x_k^* kann erst nach dem 3. Schritt ermittelt werden, weil man zunächst $u(t^{n+1}, x_{\max})$ approximieren muß, um die charakteristische Richtung λ_{k,m_k+1}^{n+1} zu erhalten (siehe 5. Schritt Randauslauf).

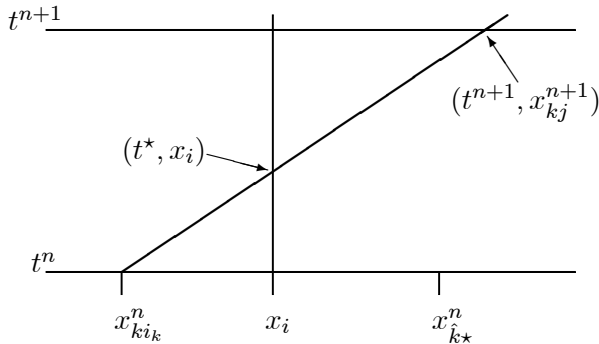


Abb. 5.3: Knick im Integrationsweg: Es sei $\tau := \frac{t^* - t^n}{\Delta t}$ der Anteil des zunächst zu integrierenden Weges. Dann ermittelt man

$$u_j^* := u_j(t^*, x_i) = u_j(t^n, x_i) + \tau \cdot \Delta t \cdot f_j(t^n, x_i, \mathbf{u}_i^n, \mathbf{v}_i^n).$$

Für jedes $\hat{k} \neq k$ läßt sich anschließend der Fußpunkt $x_{\hat{k}*}^n = x_i - \lambda(t^*, x_i, \mathbf{u}^*)$ ermitteln. Man erhält dann analog zum \mathbf{v} -Eulerschritt

$$v_{\hat{k}}^* := v_{\hat{k}*}^n + \tau \cdot \Delta t \cdot g_{\hat{k}}(t^n, x_{\hat{k}*}^n, \mathbf{u}_{\hat{k}*}^n, \mathbf{v}_{\hat{k}*}^n),$$

wobei innerhalb der alten Zeitschicht die Funktionswerte $\mathbf{u}_{\hat{k}*}^n$ und $\mathbf{v}_{\hat{k}*}^n$ aus den benachbarten Werten interpoliert werden. Nach Voraussetzung liegt in (t^*, x_i) ein Sprung in v_k vor, so daß für $t > t^*$ (bzw. $t \leq t^*$) der obere (bzw. untere) Wert verwandt werden muß. Aus dem oberen $v_k(t^*, x_i)$ und den $v_{\hat{k}}^*$ bilde man \mathbf{v}^* , um durch

$$u_j(t^{n+1}, x_i) = u_j^* + (1 - \tau) \cdot \Delta t \cdot f_j(t^*, x_i, \mathbf{u}^*, \mathbf{v}^*)$$

die Integration fortzusetzen.

5.5 Verbessern des Zeitschritts

5.5.1 Korrektur der u-Werte

Im festen Gitter X_0 liegen nun sowohl in der alten Zeitschicht $t = t^n$ als auch in der neuen Zeitschicht $t = t^{n+1}$ Näherungswerte für u und v vor. Dabei sind u_{ki}^{n+1} und v_{ki}^{n+1} aus einem einfachen Eulerschritt (und für v_{ki}^{n+1} zusätzlich einer Interpolation in $t = t^{n+1}$) hervorgegangen.

Eine Korrekturformel für u_{ki}^{n+1} scheint schon deshalb sinnvoll, weil die in charakteristischer Richtung über die v -Unbekannten zugeflossene Information gar nicht in u_{ki}^{n+1} berücksichtigt wurde⁵.

Der Ansatz

$$\begin{aligned} +u_{ki}^{n+1} &= u_{ki}^n + \Delta t \cdot \frac{1}{2} [f_k(t^n, x_i, u_{ki}^n, v_{ki}^n) \\ &\quad + f_k(t^{n+1}, x_i, u_{ki}^{n+1}, v_{ki}^{n+1})] \\ &= u_{ki}^{n+1} - \Delta t \cdot \frac{1}{2} [f_k(t^n, x_i, u_{ki}^n, v_{ki}^n) \\ &\quad - f_k(t^{n+1}, x_i, u_{ki}^{n+1}, v_{ki}^{n+1})] \end{aligned}$$

verhält sich zum ursprünglichen u -Eulerschritt so, wie die Trapezregel bei der gewöhnlichen Integration zur Summenregel (Treppenfunktion), wenn man davon absieht, daß die Größen u_{ki}^{n+1} , v_{ki}^{n+1} , u_{ki}^n und v_{ki}^n bereits Approximationen sind.

5.5.2 Nachiteration

Bei der Konstruktion einer ähnlichen Korrekturformel für die v -Unbekannten muß darüber hinaus beachtet werden, daß sich die Lage der Charakteristiken ändert, wenn man für den Zeitschritt von $t = t^n$ zu t^{n+1} statt

$$\lambda_{ki}^0 = \lambda(t^n, x_{ki}^n, u_{ki}^n)$$

⁵Außer im Sonderfall aus Abbildung 5.3

den **nachiterierten** Wert

$$\begin{aligned} \lambda_{ki}^1 &= \frac{1}{2} [\lambda_{ki}^0 + \lambda(t^{n+1}, x_{ki}^{n+1}, u_{ki}^{n+1})] \\ \text{oder} \quad +\lambda_{ki}^1 &= \frac{1}{2} [\lambda_{ki}^0 + \lambda(t^{n+1}, x_{ki}^{n+1}, +u_{ki}^{n+1})] \end{aligned}$$

verwendet. Wie dabei ein zunächst ermitteltes $x_{ki}^{n+1,0}$ durch Nachiteration zu $x_{ki}^{n+1,1}$ verändert werden kann, zeigt Abbildung 5.4. Dabei wird auch deutlich, daß eine weitere Nachiteration keine nennenswerte Qualitätssteigerung erwarten läßt.

5.5.3 Sich kreuzende Knicke

Wenn sich eine Knickcharakteristik zu v_k und eine zu $v_{\hat{k}}$ mit $k \neq \hat{k}$ schneiden, so ergibt sich kein prinzipielles Problem, da die charakteristischen Richtungen λ_k und $\lambda_{\hat{k}}$ stetig bleiben und deshalb keine Schocks entstehen. Es ist aber programmtechnisch recht aufwendig, sicherzustellen, daß in jedem Fall der richtige Wert (d.h. oberer bzw. untere Wert) für die jeweilige Näherung und ggf. die Vorwärts- (bzw. Rückwärts-)Integration verwandt wird.

Neben der im vorigen Abschnitt ausgeführten Verschiebung der Charakteristiken, also auch der Schnittstellen sich kreuzender Knickcharakteristiken ist für die Nachiteration noch zu bemerken, daß das Ergebnis auch von der Reihenfolge abhängen kann, in der man die einzelnen v_k bearbeitet.

Außerdem ergibt sich aus der im Hinweis auf Seite 154 angedeuteten Notwendigkeit einer stückweisen Integrationen bei Knicken im Integrationsweg ein erheblicher Programmieraufwand.

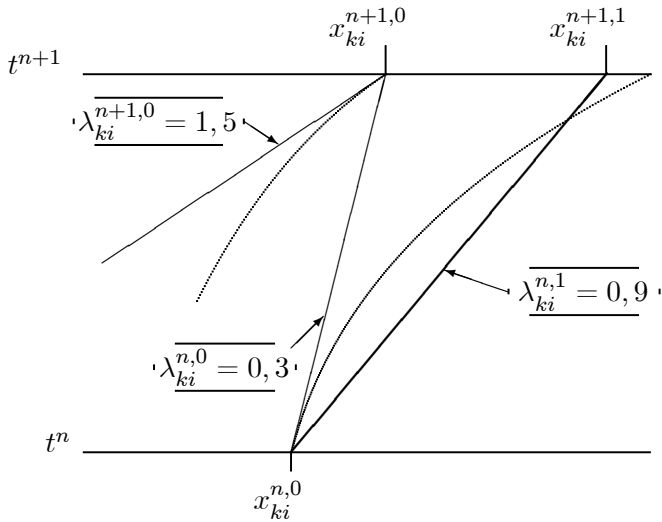


Abb. 5.4: Die Nachiteration kann das Gitter ändern:
 Die erste Näherung an die Charakteristik ergibt sich durch $\lambda_k^{n,0}$ allein aus den Daten der alten Zeitschicht $t = t^n$. Mit Werten der neuen Zeitschicht erhält man $\lambda_k^{n+1,0}$. Aus beiden ergibt sich $\lambda_k^{n,1}$ als Mittelwert.

5.6 Test des Verfahrens am Beispiel

5.6.1 Wahl des Modellproblems

Ein Modellproblem zum Testen des numerischen Verfahrens sollte so gewählt werden, daß erstens die interessanten Effekte sichtbar werden, zweitens ein zumindest qualitativer Vergleich mit der Literatur möglich ist und drittens eine explizite Lösung bekannt ist. Ferner sollte ein Bezug zur ingenieurwissenschaftlichen Praxis herstellbar sein. Schließlich muß das betrachtete quasilineare hyperbolische Differentialgleichungssystem auf die spezielle diagonale Form gebracht werden können, die auf Seite 148 in (5.1) beschrieben wurde.

Um interessante Effekte zu zeigen, muß die Lösung mindestens eine v -Unbekannte haben, deren Anfangswertvorgabe einen Sprung hat, damit in einer u -Unbekannten ein Knick vorliegt, der dann im zeitlichen Verlauf entlang einer Charakteristik⁶ wandert. Das System muß echt quasilinear sein, d.h. die charakteristische Richtung muß von der Lösung u abhängen. Bei aller anzustrebenden Einfachheit dürfen die Lösung (bzw. die Charakteristiken-Scharen) höchstens auf einer Seite einer Knickcharakteristik konstant oder linear sein, weil sonst die Interpolation trivial wird.

In der Literatur wird als nichtlineares System mit Störungen häufig das Riemann-Problem für die Euler-Gleichungen (*Shock Tube Problem*) betrachtet (etwa [Sod85]). Dabei werden zwei anfänglich durch eine Membran getrennte in Ruhe befindliche Flüssigkeiten betrachtet, deren jeweils konstante Drücke und Dichten sich unterscheiden. Beim (infinitesimal schnellen) Entfernen der Membran breiten sich eine Ausdünnungswelle (*rarefaction wave* oder *centered simple wave*), eine Kontaktunstetigkeit und ein Schock mit jeweils konstanter Geschwindigkeit

⁶**Knickcharakteristik** nenne ich eine solche Charakteristik, entlang derer sich ein Knick ausbreitet. Diese kann durchaus glatt sein

aus.

Sprünge (wie etwa Kontaktunstetigkeiten oder echte Schockwellen) in den u -Unbekannten sind in meinem Verfahren nicht vorgesehen⁷. Deshalb kann aus dem Riemann-Problem nur der stetige Teil der Lösung, d.h. eine Umgebung der Ausdünnungswelle benutzt werden. Diese Welle tritt qualitativ genauso auf bei den Flachwasser-Gleichungen

$$\begin{aligned}\partial_t u + u \cdot \partial_x u + \partial_x h &= 0 \\ \partial_t h + h \cdot \partial_x u + u \cdot \partial_x h &= 0,\end{aligned}$$

wobei h die Wasserstandshöhe und u die Geschwindigkeit bezeichne, die man aus den Anfangswerten

$$u(0, x) = 0 \quad \text{und} \quad h(0, x) = \begin{cases} 1 & \text{für } x < 0 \\ 0 & \text{für } x \geq 0 \end{cases}$$

sich entwickeln lasse (Dammbruch-Problem). Dieses Problem hat gegenüber den vollen Eulergleichungen den Vorteil, wegen der nur zwei Unbekannten leichter diagonalisierbar zu sein. Sofern man den Gleichungen hinreichende ingenieurwissenschaftliche Relevanz zugesteht, so haftet dem System noch der Mangel an, daß die Charakteristiken Geradenstücke sind. Das läßt sich durch eine Transformation der unabhängigen Unbekannten $t = \tau^2$ beseitigen.

Vorher ist aber noch zu bemerken, daß sich das Dammbruch-Problem mit Hilfe der Anfangswerte und der Riemann-Invarianten (siehe [Kevo90, Seite 421f.]) auf die eine Gleichung

⁷Grundsätzlich wäre es möglich, das Verfahren entsprechend zu verallgemeinern, weil jede \mathbf{v}_k -Unbekannten ein eigenes Gitter benutzt und in den Nachbarknoten die im Schock zu berücksichtigenden Werte vorliegen. Man müßte die Funktion zur Ermittlung von λ_k so intelligent machen, daß **die richtige** Schockgeschwindigkeit heraus kommt. Dabei sind aber auch diverse Spezialfälle abzufangen, so daß der Aufwand gegenüber einem evtl. Nutzen sorgsam abzuwägen ist

$\partial_t h + (2 - 3\sqrt{h})\partial_x h = 0$ reduzieren läßt, welche ihrerseits durch die Variablentransformation $v = (2 - 3\sqrt{h})$ in Burgers Gleichung $\partial_t v + v \cdot \partial_x v = 0$ übergeht. Insofern ist die Burger-Gleichung kein rein akademisches Beispiel, sondern prinzipiell für einen Praxistest meines Verfahrens geeignet. Allerdings ist die Funktion $\lambda(t, x, u) = u$ für die charakteristische Richtung zu einfach, um Interpolationseffekte bzgl. λ zu zeigen. Ich wähle daher ein im folgenden Abschnitt beschriebenes Modellproblem, das im wesentlichen den Wasserstand beim Dammbruch beschreibt.

Die Beschränkung auf nur eine \mathfrak{v} -Unbekannte ist dabei keine unverhältnismäßige Beschränkung der Allgemeinheit, weil die einzelnen \mathfrak{v} -Unbekannten ohnehin nahezu unabhängig voneinander behandelt werden können. Allerdings entfallen bei Beschränkung auf nur eine \mathfrak{v} -Unbekannte die in Abbildung 5.3 angedeutete Berechnung des Fußpunktes $x_{k^*}^n$ und die im Hinweis auf Seite 154 angedeuteten Mühen stückweiser Integrationen bei Knicken im Integrationsweg.

5.6.2 Explizite Lösung des Modellproblems

Als Modellsystem wird die einzelne Gleichung

$$\partial_t u(t, x) + \lambda(t, x, u(t, x)) \cdot \partial_x u(t, x) = 0 \quad (5.6)$$

mit $\lambda(t, x, u) = 2t \cdot (3 - \sqrt{u})$ gewählt. Gelöst wird das Anfangsrandwertproblem (5.6) mit den Anfangswerten

$$u(t_0, x) = \max \left\{ \min \left\{ 4, \left(3 - \frac{x}{t_0^2} \right)^2 \right\}, 1 \right\} \quad \text{für } x \in [0, 4] \quad (5.7)$$

und stetig anschließenden konstanten Randwerten. Die explizite Lösung lautet $u : [1, \infty[\times [0, 4] \rightarrow [1, 4]$, $(t, x) \mapsto u(t, x)$ mit

$$u(t, x) = \begin{cases} 4 & x \leq t^2 \\ \left(3 - \frac{x}{t^2} \right)^2 & \text{für } t^2 < x \leq 2t^2 \\ 1 & 2t^2 < x \end{cases}$$

und hat je einen Knick entlang der beiden Knick-Charakteristiken $t = \sqrt{x}$ und $t = \sqrt{x/2}$. Nach dem in Kapitel 4 beschriebenen Diagonalisierungsverfahren liefert Formel (8.2) von Seite 238 explizit das diagonale System

$$\begin{aligned} \partial_t v &= v \\ \partial_t v + 2t \cdot (3 - \sqrt{u}) \partial_x v &= v \frac{2 \cdot (3 - \sqrt{u}) - t \frac{v}{\sqrt{u}}}{2 \cdot t \cdot (3 - \sqrt{u})} \end{aligned}$$

mit den Anfangswerten (5.7) und

$$v(t_0, x) = \begin{cases} \frac{4x}{t_0^3} \cdot \left(3 - \frac{x}{t_0^2} \right) & \text{für } t_0^2 < x \leq 2t_0^2 \\ 0 & \text{sonst} \end{cases} \quad \text{für } x \in [0, 4]$$

und der Randwertvorgabe $v(0, t) = 0$ für alle $t \geq t_0$.

Diese Anfangswertvorgabe für v folgt aus (5.6) in der Form⁸ $v = -\lambda \partial_x u$ und (5.7), das $\partial_x u(t_0, x)$ zu ermitteln ermöglicht.

5.6.3 Numerische Ergebnisse

Bei einem Modellproblem mit explizit bekannter Lösung läßt sich die Qualität der numerischen Approximation in jedem Auswertungs-Zeitpunkt t messen erstens durch den absoluten Fehler je approximiertem Punkt, zweitens durch die L_1 -Norm der Abweichung einer diese Punkte interpolierenden Lösung von der exakten Lösung und drittens durch die Abweichung der numerisch ermittelten Knickpositionen von den jeweils echten Knickpositionen⁹. Dabei sind die folgenden Parameter als Einflußgrößen zu betrachten:

- Auswertungs-Zeitpunkt t
- anfänglicher Abstand der Ortspunkte Δx
- Wahl des Verfahrens
 - Eulerschritt
 - u -Korrektur ohne Nachiteration
 - Nachiteration ohne u -Korrektur
 - Kombination beider Verbesserungen
- Zeitschrittlänge Δt

⁸Das ist die konkrete Gestalt von (4.23) auf Seite 108

⁹Wenn es eine propagierte Zwangsbedingung gibt, was hier nicht der Fall ist, mag auch deren Norm ein geeignetes Fehlermaß sein

Das **Fazit der Untersuchungen** lautet:

Bei vernünftigem Δx verkleinert die Wahl des aufwendigsten Verfahrens gegenüber dem einfachen Eulerschritt praktisch alle Fehlerkategorien auf ein Zehntel, während der Einfluß der Zeitschrittlänge Δt nur proportional wirkt. Das Verfahren ist also ein implizites Verfahren erster Ordnung.

Im folgenden sollen die Wirkungen der verschiedenen Einflußgrößen mit Hilfe von Abbildungen erläutert werden. Diese zeigen jeweils die numerischen Approximationswerte und die explizite Lösung oder deren Differenz.

Im Laufe der zeitlichen Entwicklung vergrößert sich der Fehler mit dem **Auswertungszeitpunkt** t erwartungsgemäß, die Welle wandert von links nach rechts. Fehler entstehen zwischen den Knicken (in der Welle) und links vom Knick (hinter der Welle). Punkte, die hinter der Welle zu hoch angesetzt wurden, werden nicht mehr nach unten korrigiert, sondern bei großem Δx sogar noch größer (siehe Abbildung 5.5ff).

In Abbildung 5.5 und 5.6 wird dieser Effekt für das Euler-Verfahren gezeigt. Prinzipiell genauso sieht es beim verbesserten Verfahren aus, wie man in Abbildung 5.7 sieht. Dort wird allerdings die Differenz gezeigt, weil diese Effekte in der anderen Darstellung nicht sichtbar wären.

Ein sehr großer **Abstand der Ortspunkte** Δx führt zu einer Dominanz der Interpolationsfehler mit der Folge, daß andere Effekte nicht sichtbar werden. Insbesondere blieb der Fehler bei 10-fach vergrößerter Zeitschrittanzahl fast unverändert bei anfänglich 17 Ortspunkten. Im folgenden arbeite ich daher überwiegend mit anfänglich 51 Ortspunkten. Eine Ausnahme ist Abbildung 5.8, welches gerade zeigt, daß der Fehler bei hinreichend kleinem Δx praktisch unabhängig ist von Δx , weil der Interpolationsfehler, der allein von Δx abhängt, dann unerheblich wird.

Entwicklung beim Euler-Schritt (17 Ortspunkte)

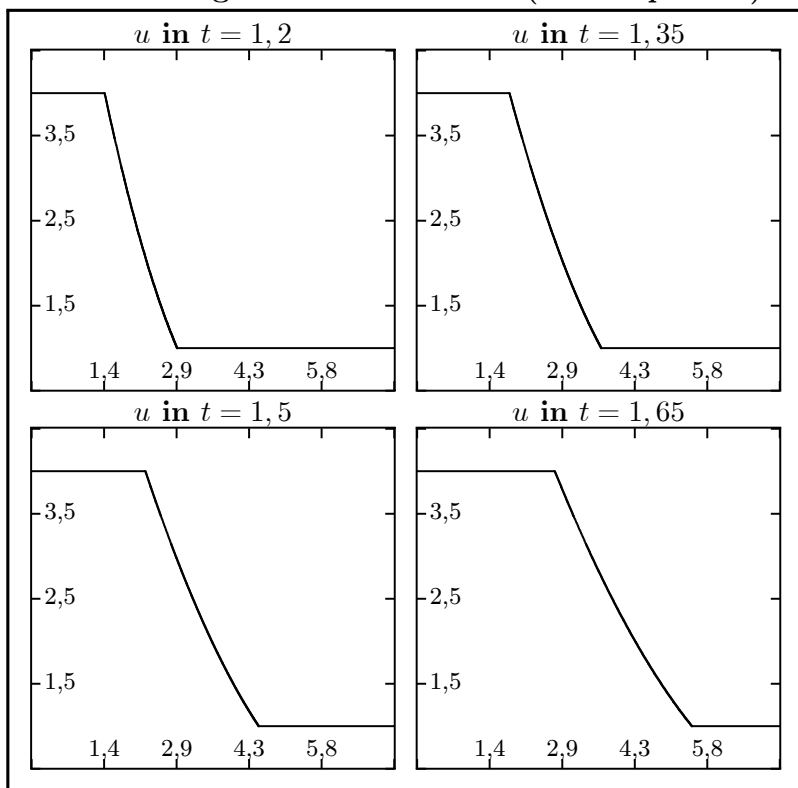


Abb. 5.5: Die Welle wandert im Laufe der Zeit von links nach rechts. Dargestellt ist als durchgezogene Linie die exakte Lösung u und als Einzelpunkte die Näherung zum Anfangsabstand $\Delta x = 0.4235$

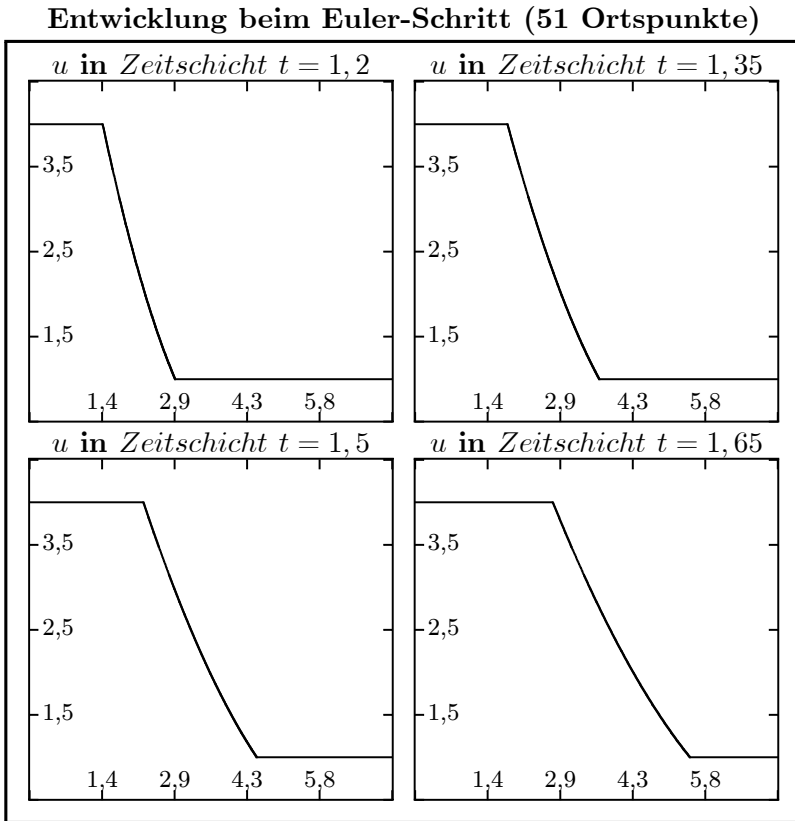


Abb. 5.6: Dargestellt ist wieder als durchgezogene Linie die exakte Lösung u und als Einzelpunkte die Näherung zum kleineren Anfangsabstand etwa $\Delta x = 0.1$

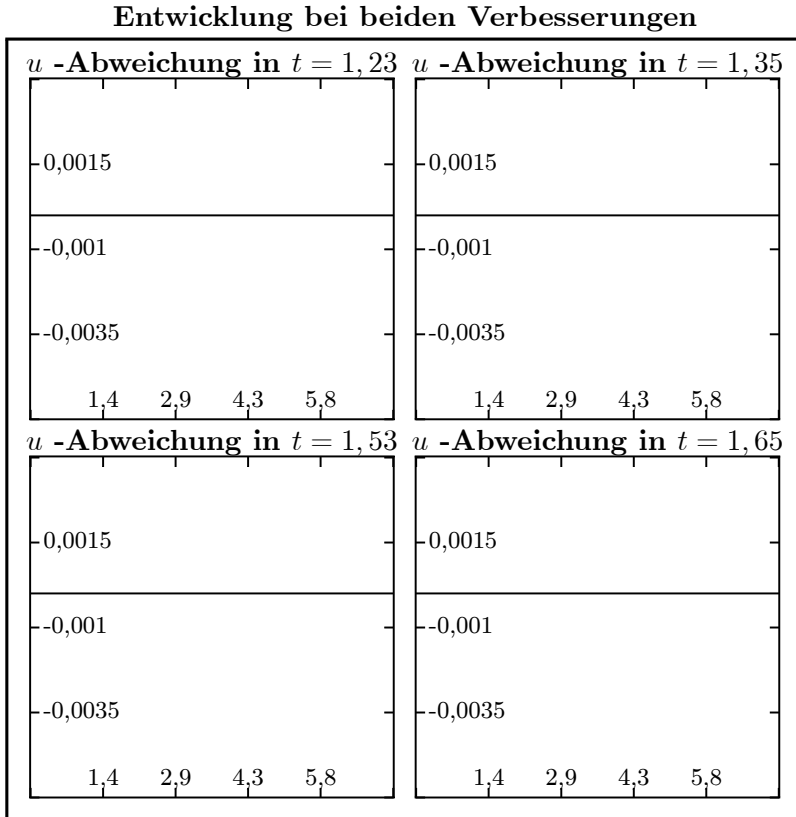


Abb. 5.7: Bei u -Korrektur in Verbindung mit Nachiteration ist die Abweichung so gering, daß die bisherige Darstellungsform ungeeignet ist. Gezeigt wird daher die Differenz $h_{\text{numerisch}} - h_{\text{exakt}}$ selbst.

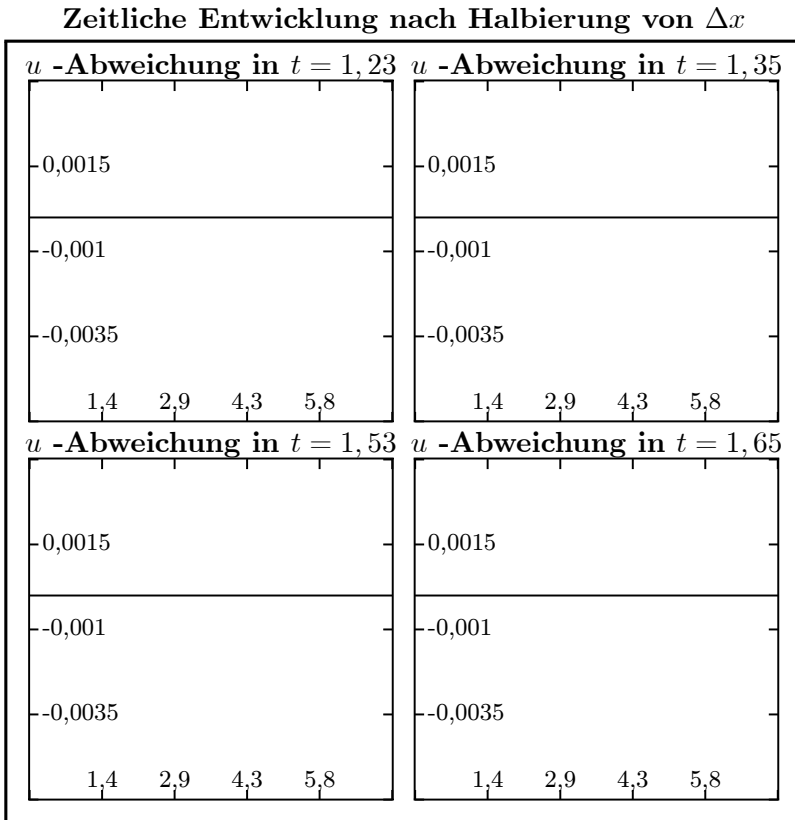


Abb. 5.8: Bei Halbierung des räumlichen Abstands Δx verringert sich der Fehler nicht auf die Hälfte. Gezeigt wird wieder die Differenz $h_{\text{numerisch}} - h_{\text{exakt}}$.

Bei 10 Zeitschritten der Länge $\Delta t = 0.03$ läßt sich die Wirkung der **Wahl des Verfahrens** noch mit bloßem Auge erkennen. Die u-Korrektur scheint deutlich mehr als die Nachiteration zu bringen (siehe Abbildung 5.9f ab Seite 171). In Abbildung 5.9 erkennt man, daß die Kombination der beiden Verbesserungen u-Korrektur und Nachiteration etwa den Faktor 10 bewirkt.

Zusätzlich wird der v -Fehler in Abbildung 5.11 gezeigt. Dieser bleibt erwartungsgemäß von einer u-Korrektur unbeeinflusst.

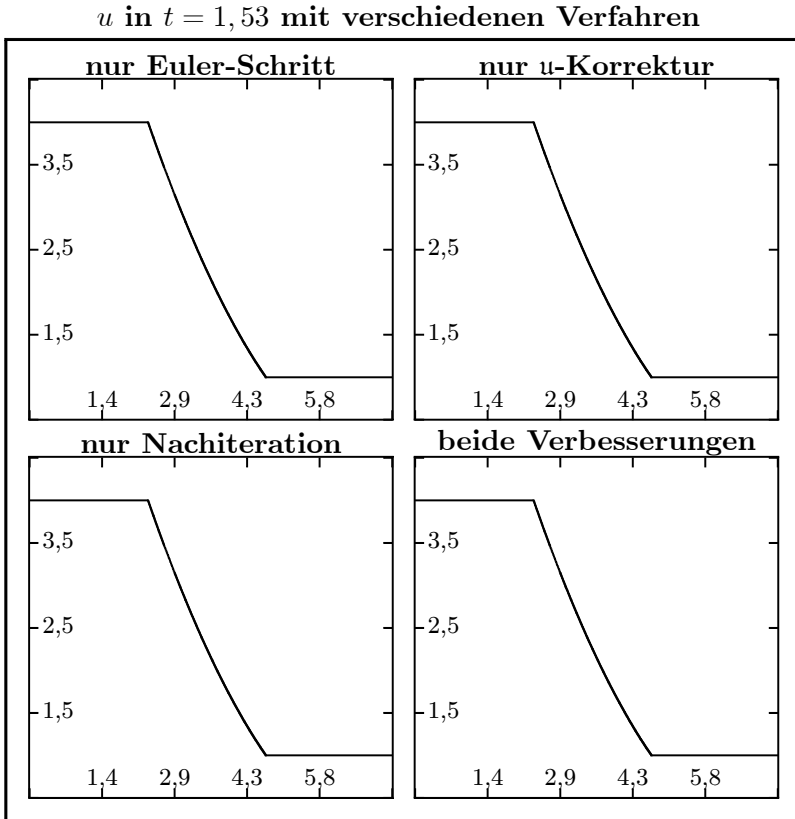


Abb. 5.9: Mit Schrittweite $\Delta t = 0.03$ wurden 10 Zeitschritte gerechnet. Die Wirkung der u -Korrektur ist dabei deutlich größer als die der Nachiteration.

u -Fehler in $t = 1.53$ mit verschiedenen Verfahren

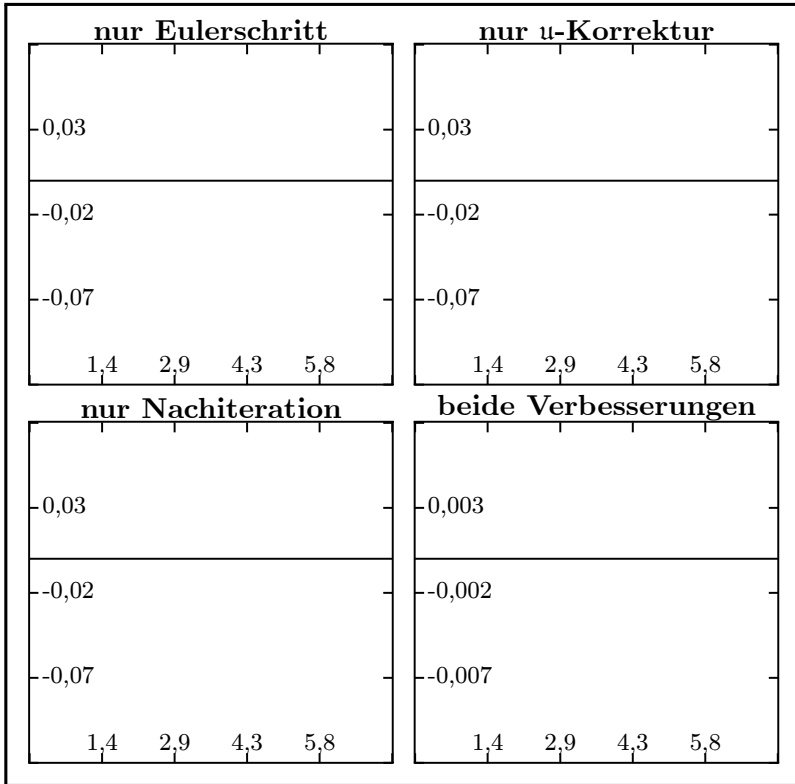


Abb. 5.10: Auswertung in $t = 1,53$ mit Zeitschrittlänge $\Delta t = 0,03$. Man sieht, daß die Kombination beider Verfahren sehr wirksam ist, man beachte den Faktor 10 in der Skalierung.

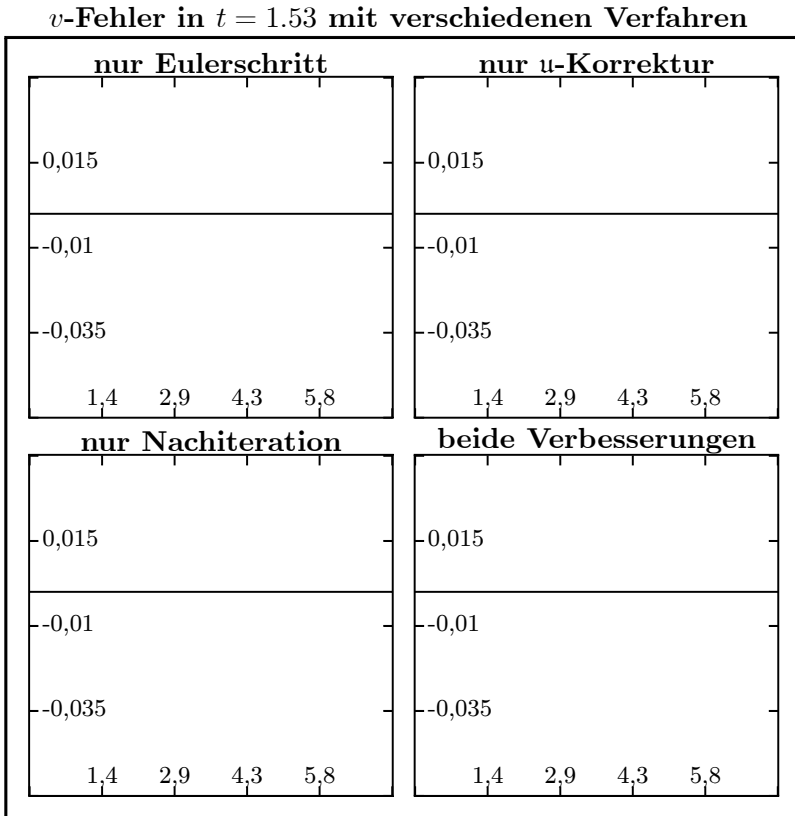


Abb. 5.11: Auswertung in $t = 1,53$ mit Zeitschrittlänge $\Delta t = 0,03$. Hier hat die reine u -Korrektur keine Wirkung, weil ja gerade der v -Fehler betrachtet wird.

Zur **Zeitschrittlänge** Δt verhalten sich alle Fehler proportional, wie Tabelle 5.1 zeigt. Dort wird die L_1 -Norm von

$$u_{\text{numerisch}} - u_{\text{exakt}}$$

ermittelt, indem $u_{\text{numerisch}}$ mit der Simpson-Regel und 800 Stützstellen integriert wurde. Der Integrand wurde dabei aus den vorliegenden Approximationsstellen, wie in Abschnitt 5.3 beschrieben, interpoliert.

Im Ergebnis ist der Faktor $1/10$, auf den jeder Fehler beim Verkleinern des Zeitschritts Δt schrumpft, deutlich zu erkennen. Lediglich die Lage von Knick_2 im Falle beider Verbesserungen ließ sich nicht weiter verkleinern, weil dort wohl Rundungseffekte dominieren.

Die Tabelle zeigt auch, daß bei durchgeführter u -Korrektur die Nachiteration bei der L_1 -Norm etwa den Faktor 8 bringt. Dieser Effekt wird in Abbildung 5.12 visualisiert. Bezüglich der v -Abweichung ist die Wirkung der Nachiteration geringer, wie Abbildung 5.13 zeigt. In diesem Bild erkennt man außerdem, daß sich auch der v -Fehler proportional zu Δt verhält.

Δt	u- Korr.	Nach- iter.	u -Abw. L_1 -Norm	Knick ₁ - Fehler	Knick ₂ - Fehler
0,03			0,064 566	0,065 496	-0,017 098
0,03	×		0,049 399	0,049 656	-0,012 043
0,03		×	0,037 991	0,038 203	-0,005 677
0,03	×	×	0,005 896	0,005 897	-0,000 021
0,003			0,006 710	0,006 720	-0,001 787
0,003	×		0,003 777	0,003 781	-0,001 229
0,003		×	0,003 914	0,003 916	-0,000 543
0,003	×	×	0,000 519	0,000 519	0,000 022

Tabelle 5.1: Gezeigt wird in $t = 1,53$ die u -Abweichung in der L_1 -Norm, wobei die exakte Lösung mit der interpolierten numerischen Lösung verglichen wurde. Dabei wurde die Simpsonregel mit 800 Teilstücken angewandt

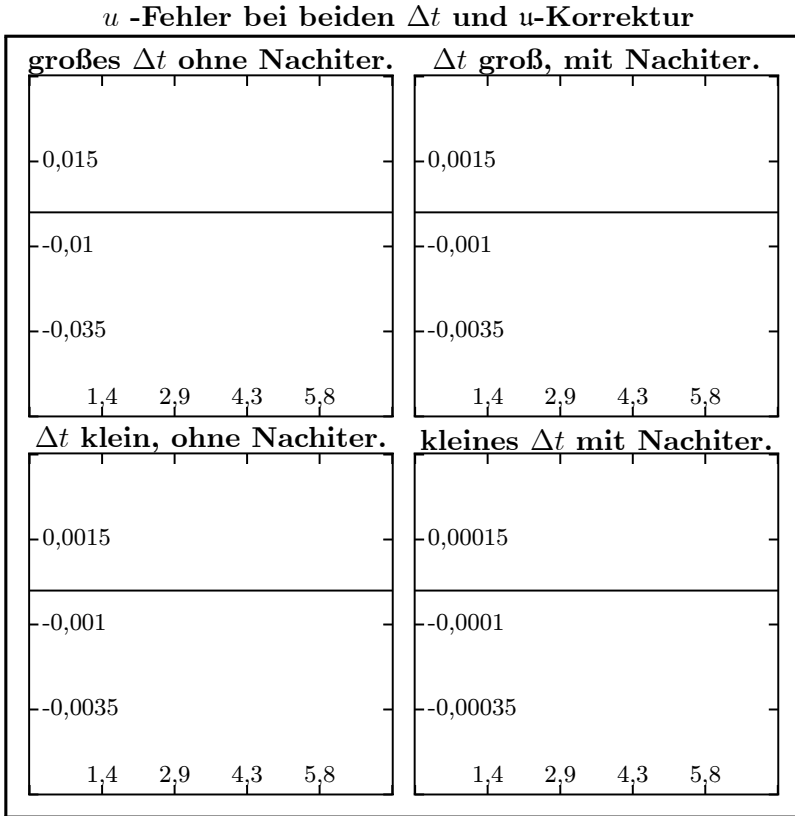


Abb. 5.12: Diese Darstellung der Abweichung des numerischen u vom exakten Wert in $t = 1,53$ zeigt den Nutzen der Nachiteration bei durchgeführter u -Korrektur und verschiedenen Δt . Zwischen der links ohne und der rechts mit Nachiteration entstehenden Abweichung liegt der Skalierungsfaktor 10. Derselbe Faktor wird von oben nach unten durch Schrittweitenverkleinerung erzielt.

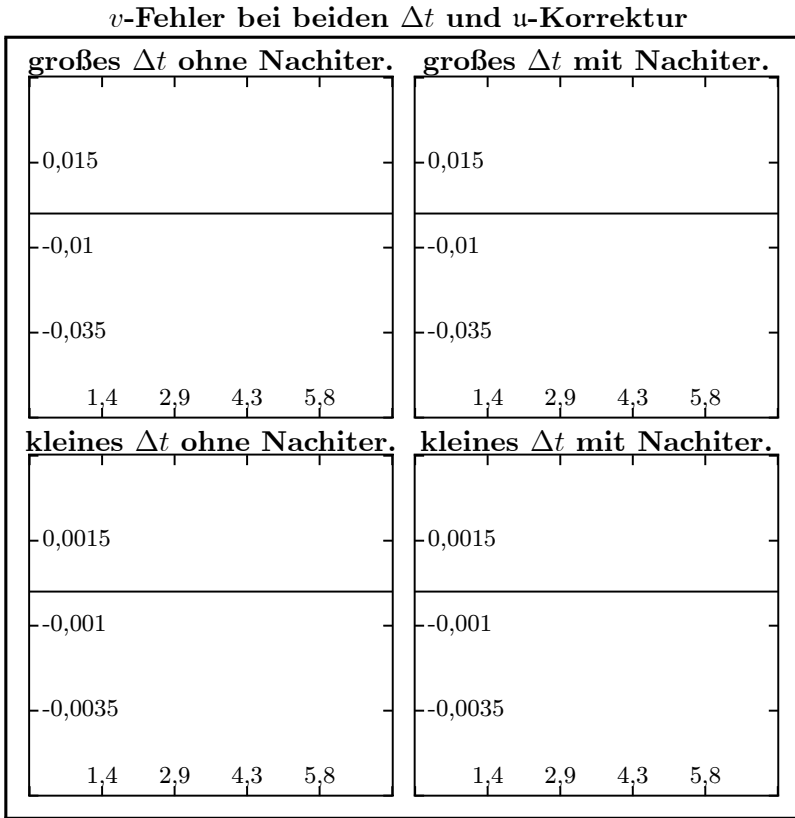


Abb. 5.13: Diese Darstellung der Abweichung des numerischen v vom exakten Wert in $t = 1,53$ zeigt, daß die Nachiteration bezüglich v einen geringen Nutzen hat. Außerdem wird wieder durch 10-fachen Rechenaufwand ($\Delta t_{\text{klein}} = \Delta t_{\text{groß}}/10$) die 10-fache Genauigkeit erzielt.

Diese Ergebnisse gelten nur für hinreichend kleine Δx . Bei anfänglich 17 Ortspunkten dominiert offenbar der Interpolationsfehler so sehr, daß die Wahl des Verfahrens selbst dann keinerlei Wirkung zeigt, wenn das kleine $\Delta t = 0.003$ gewählt wird (siehe Abbildung 5.14), was gegenüber 3-facher Ortspunkteanzahl (siehe Abbildung 5.12) zu einem mehr als 500-fachen Fehler führt.

5.6.4 Bewertung der Ergebnisse

Vergleich zur Literatur

Ein qualitativer Vergleich zu den Abbildungen in der Arbeit von Sod [Sod78] läßt das Euler-Verfahren aus Abbildung 5.5 vergleichbar mit den meisten der dort vorgestellten Verfahren erscheinen. Die Nachbesserungen führen zu einem sehr guten Abschneiden meines Verfahrens, wobei mir nicht klar ist, ob eine Vergleichbarkeit im Hinblick auf den Auswertungszeitpunkt t überhaupt gegeben ist, da dieser bei Sod nicht genannt wurde.

Effektivität des Verfahrens

Einerseits ist für den betrachteten 1D-Fall die Struktur des diagonalen Systems 5.1 sehr übersichtlich, so daß ein derartiges numerisches Verfahren recht untersuchungswürdig erscheint. Andererseits sind doch recht viele Spezialfälle auszuprogrammieren, so daß man sich fragen muß, ob diese Lösung, die sich ja nicht auf 3D-Probleme verallgemeinern läßt, Ausgangspunkt weiterer Untersuchungen sein kann. Kritisch zu bewerten ist dabei auch, daß es diesem Ansatz schwierig zu klären ist, wie er in die theoretische Numerik einzuordnen ist. Ist das Verfahren stabil? Ist es ein TVD¹⁰-Verfahren? Wie ließe sich die

¹⁰**total variation diminishing** bezeichnet Verfahren, bei denen die totale Variation nicht größer wird mit fortschreitender Zeit

Ordnung steigern?

Positiv zu bewerten wäre die Möglichkeit der Adaption der Gitterpunkte an die lokalen Bedürfnisse, wie es etwa Hern in [Hern99] durchführt und Massó et.al. in [Mass94] untersuchen. Weil ohnehin jeder v_k -Unbekannten ein eigenes Gitter zugeordnet ist, könnte man Auswertungspunkte hinzufügen (bzw. weglassen), wo eine höhere (bzw. niedrigere) Genauigkeit erforderlich (bzw. vertretbar) ist. Dabei entsteht aber weiterer Programmieraufwand, insbesondere bei den Verbesserungen, der sich vermutlich nur für sehr spezielle Probleme lohnen dürfte. Außerdem ist zu bedenken, daß der räumliche Abstand Δx bei dem betrachteten Beispiel in Bezug auf den Gesamtfehler fast keine Rolle spielte, so daß ggf. lieber Δt zu verkleinern wäre.

Alternative Verfahren

Seit Mitte der 1990er Jahre ist es möglich, die Einsteinschen Gleichungen in konservativer Form zu schreiben (siehe etwa [Bona95], [Bona97]), um den Einsatz fortschrittlicher numerischer Methoden der klassischen Hydrodynamik auch in 3 Raumdimensionen zu ermöglichen. Ob ein derartiger Ansatz auch für das hier behandelte Problem geeignet ist, habe ich nicht untersucht.

Ich hielt es jedenfalls für wichtig, die theoretischen Überlegungen, insbesondere zur Diagonalisierung, durch ein passendes numerisches Verfahren zu ergänzen.

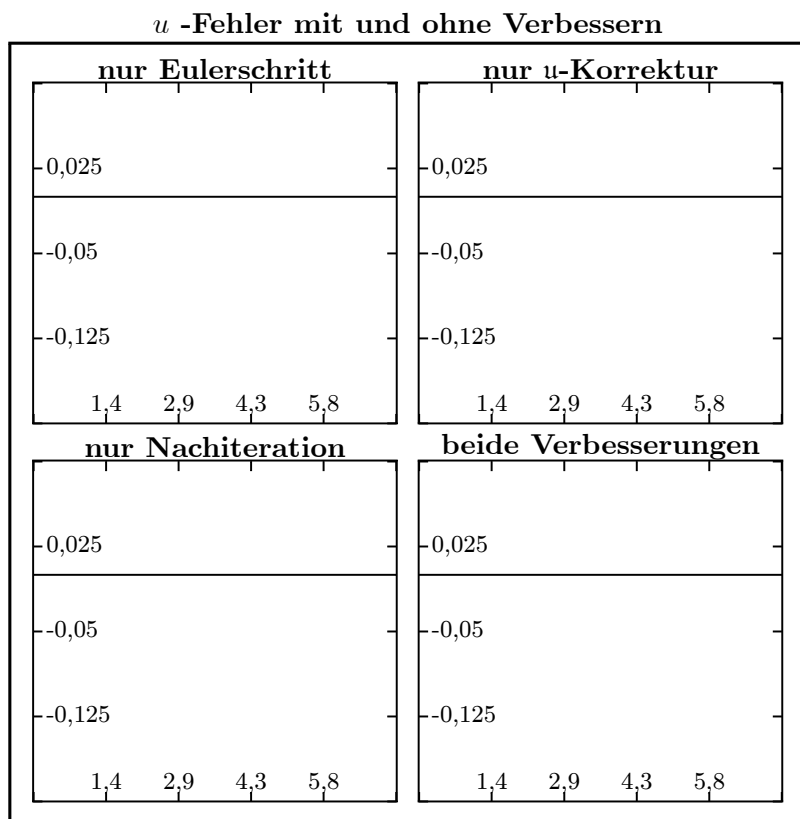


Abb. 5.14: Es wurden wieder 110 Zeitschritte mit Schrittweite $\Delta t = 0.003$ bis in $t = 1,53$ gerechnet. Dabei wird erneut deutlich, daß der Interpolationsfehler so groß ist, daß die relativ kleine Zeitschrittweite nichts nützt.