

# Numerische Mathematik (Numerik der ODEs)

Prof. Ralf Hiptmair, Dr. Vasile Gradinaru

Seminar for Applied Mathematics, ETH Zürich

Entwurf Februar 2011, Subversion rev 35327

<http://www.sam.math.ethz.ch/~hiptmair/tmp/NUMODE11.pdf>

# Inhaltsverzeichnis

0.1	Danksagung . . . . .	10
<b>1</b>	<b>Einleitung</b>	<b>11</b>
1.1	Anfangswertprobleme (AWP) . . . . .	12
1.2	Beispiele und Grundbegriffe . . . . .	22
1.2.1	Ökologie . . . . .	23
1.2.2	Chemische Reaktionskinetik . . . . .	29
1.2.3	Physiologie . . . . .	32
1.2.4	Mechanik . . . . .	37
1.3	Theorie [8, Sect. 2], [1, Ch. II] . . . . .	43
1.3.1	Existenz und Eindeutigkeit von Lösungen . . . . .	44
1.3.2	Lineare AWPe [3, Sect. 8.2] . . . . .	52
1.3.3	Sensitivität [8, Sect. 3.1] . . . . .	57

1.3.3.1	Grundbegriffe . . . . .	57
1.3.3.2	Unser Problem: das Anfangswertproblem . . . . .	59
1.3.3.3	Wohlgestelltheit . . . . .	61
1.3.3.4	Asymptotische Kondition . . . . .	63
1.3.3.5	Schlecht konditionierte AWPes . . . . .	66
1.4	Polygonzugverfahren . . . . .	72
1.4.1	Das explizite Eulerverfahren . . . . .	73
1.4.2	Das implizite Euler-Verfahren . . . . .	85
1.4.3	Implizite Mittelpunktsregel . . . . .	99
1.4.4	Störmer-Verlet-Verfahren [15] . . . . .	104
<b>2</b>	<b>Einschrittverfahren</b>	<b>117</b>
2.1	Grundlagen . . . . .	117
2.1.1	Abstrakte Einschrittverfahren [8, Sect. 4.1] . . . . .	118
2.1.2	Konsistenz [8, Sect. 4.1.1] . . . . .	124
2.1.3	Konvergenz . . . . .	130
2.1.4	Das Äquivalenzprinzip (Dahlquist, Lax) . . . . .	137
2.1.5	Reversibilität . . . . .	140
2.2	Kollokationsverfahren[8, Sect. 6.3], [16, Sect. II.1.2] . . . . .	144
2.2.1	Konstruktion . . . . .	144
2.2.2	Abstrakte Projektionsverfahren . . . . .	163
2.2.3	Konvergenz von Kollokationsverfahren . . . . .	172

2.2.3.1	Konsistenzordnung . . . . .	172
2.2.3.2	Spektrale Konvergenz . . . . .	185
2.3	Runge-Kutta-Verfahren . . . . .	222
2.3.1	Konstruktion . . . . .	222
2.3.2	Konvergenz . . . . .	237
2.4	Extrapolationsverfahren [8, Sect. 4.3] . . . . .	249
2.4.1	Der Kombinationstrick . . . . .	249
2.4.2	Extrapolationsidee . . . . .	252
2.4.3	Extrapolation von Einschrittverfahren . . . . .	259
2.4.4	Lokale Extrapolations-Einschrittverfahren . . . . .	265
2.4.5	Ordnungssteuerung . . . . .	271
2.4.6	Extrapolation reversibler Einschrittverfahren . . . . .	274
2.5	Splittingverfahren [16, Sect. 2.5] . . . . .	278
2.6	Schrittweitensteuerung [8, Kap. 5], [19, Sect. 2.8] . . . . .	287
<b>3</b>	<b>Stabilität [8, Kap. 6]</b>	<b>320</b>
3.1	Modellproblemanalyse . . . . .	322
3.2	Vererbung asymptotischer Stabilität . . . . .	339
3.2.1	Attraktive Fixpunkte . . . . .	339
3.2.2	Attraktive Fixpunkte von Einschrittverfahren . . . . .	345
3.3	Nichtexpansivität [8, Abschn. 6.3.3] . . . . .	352
3.4	Gleichmässige Stabilität . . . . .	363
3.5	Steifheit . . . . .	375
3.6	Linear-implizite Runge-Kutta-Verfahren [8, Sect. 6.4] . . . . .	386
3.7	Exponentielle Integratoren [24, 28, 25] . . . . .	396
3.8	Differentiell-Algebraische Anfangswertprobleme . . . . .	402
3.8.1	Grundbegriffe . . . . .	402
3.8.2	Runge-Kutta-Verfahren für Index-1-DAEs . . . . .	408
3.8.3	DAEs mit höherem Index . . . . .	419

<b>4</b>	<b>Strukturerhaltende numerische Integration</b>	<b>433</b>	
4.1	Polynomiale Invarianten . . . . .	434	Numerische Mathematik
4.2	Volumenerhaltung . . . . .	446	
4.3	Verallgemeinerte Reversibilität . . . . .	456	
4.4	Symplektizität . . . . .	466	
4.4.1	Symplektische Evolutionen Hamiltonscher Differentialgleichungen . . . . .	466	
4.4.2	Symplektische Integratoren . . . . .	483	
4.4.3	Rückwärtsanalyse . . . . .	507	
4.4.4	Modifizierte Gleichungen: Fehleranalyse . . . . .	519	
4.4.5	Strukturerhaltende modifizierte Gleichungen . . . . .	545	
4.5	Methoden für oszillatorische Differentialgleichungen [23] . . . . .	556	R. Hiptmair
	<b>Verzeichnisse</b>	<b>570</b>	rev 35327, 13. Mai 2011
	Stichwortverzeichnis . . . . .	570	
	Verzeichnis der Beispiele und Bemerkungen . . . . .	581	
	Verzeichnis der Definitionen und Konzepte . . . . .	585	
	Verzeichnis der MATLAB-CODE-Fragmente . . . . .	587	
	Symbolverzeichnis . . . . .	588	

# Allgemeine Informationen

Dozent: Prof. Ralf Hiptmair, SAM, D-MATH, Büro: HG G 58.2 Tel.: 044 632 3404,  
hiptmair@sam.math.ethz.ch  
Assistent: Cedric Effenberger, SAM, D-MATH Büro: HG G 55, Tel.: 044 632 0392,  
ece@sam.math.ethz.ch  
Tutoren: Dr. Jingzhi Li, SAM, D-MATH Büro: HG G 55 Tel.: 044 632 0392,  
jingzhi.li@math.ethz.ch  
Jan Ernest jernest@student.ethz.ch

## Website:

<http://www.math.ethz.ch/education/bachelor/lectures/fs2011/math/nm2>

**Prüfung:** schriftliche Prüfung *am Rechner* (teilweise MATLAB-basierte Programmieraufgaben),  
*keine* (mitgebrachten) Hilfsmittel

Vorlesungsunterlagen werden als PDF-Datei zur Verfügung gestellt

## Übungen:

Einschreibung: [http://www.math.ethz.ch/~grsam/NumMath2\\_MATH\\_FS11/i/](http://www.math.ethz.ch/~grsam/NumMath2_MATH_FS11/i/)

- wöchentliches Übungsblatt zum Download (Bearbeitungszeit: 1 Woche)

- MATLAB-basierte Programmieraufgaben
- Abgabe: Dienstag bis 8:00 Uhr in den Fächern im Vorraum zum HG G 53
- Abgabe der Codes via Webupload: <http://www.math.ethz.ch/~grsam/submit/>
- Pflicht: 2× Vorlösen im Semester (mit Voranmeldung)
- Korrektur nur auf Anfrage (jeder Teilnehmer erhält 5 Korrekturvouchers)
- Selbstkorrektur mit stichprobenartigen Kontrollen
- bei Betrug: Aberkennung der Punkte, +10% zur Testatbedingung
- Testatbedingung: 50% der Übungen rechtzeitig abgegeben und akzeptiert

### Sprechstunde der Assistenten:

- Dr. Jingzhi Li, montags 08:15 – 09:00 Uhr, HG G 53 (Vorraum)
- Jan Ernest, montags 08:15 – 09:00 Uhr, HG G 53 (Vorraum)

### 👉 Literatur:

P. DEUFLHARD AND F. BORNEMANN, *Numerische Mathematik II*, DeGruyter, Berlin, 2 ed., 2002.

Kapitel 4: <http://www.sam.math.ethz.ch/~hiptmair/tmp/Literatur1.pdf>

Kapitel 6: <http://www.sam.math.ethz.ch/~hiptmair/tmp/Literatur2.pdf>

(Von Springer auch in Englisch erhältlich → Link)

## Enzyklopädische Präsentation klassischer numerischer Integratoren:

E. HAIRER, S. NORSETT, AND G. WANNER, *Solving Ordinary Differential Equations I. Nonstiff Problems*, Springer, Heidelberg, 2nd ed., 1993.

E. HAIRER AND G. WANNER, *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*, Springer, Heidelberg, 1991.

## Umfassende Darstellung “strukturhaltender” Integratoren:

E. HAIRER, C. LUBICH, AND G. WANNER, *Geometric numerical integration*, vol. 31 of Springer Series in Computational Mathematics, Springer, Heidelberg, 2002.

R. Hiptmair  
rev 35327,  
20. Februar  
2011

## Gute Einführung in die Numerik Hamiltonscher Differentialgleichungen:

B. LEIMKÜHLER AND S. REICH, *Simulating Hamiltonian Dynamics*, vol. 14 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, UK, 2004.



# Hinweise auf Fehler in den Vorlesungsunterlagen

Bitte melden Sie Fehler in den Vorlesungsunterlagen via folgende [Wikiseite!](#)

```
http://elbanet.ethz.ch/wikifarm/rhiptmair/index.php?n=Main.NumOde
```

(Passwort: NUMODE, bitte das **EDIT**-Menu wählen, um Text einzugeben.)

Bitte versehen Sie eine Fehlermeldung mit folgenden Angaben:

- Abschnitt, in dem der Fehler auftritt.
- Genau Ortsangabe (a.B. nach Gleichung (4), in Bsp. 2.4.5, vor Satz 2.3.3, etc. ). Bitte vermeiden Sie die Angabe von Seitennummern.
- Kurze Fehlerbeschreibung

Alternative: E-mail an C. Effenberger [ece@sam.math.ethz.ch](mailto:ece@sam.math.ethz.ch), Subject: NUMODE Error

# 0.1 Danksagung

Dank geht an Frau Evgenia Ageeva für die Aufarbeitung der MATLAB-Codes zu numerischen Beispielen.

# 1

## Einleitung

Vertrautheit mit Grundbegriffen der Theorie der Anfangswertprobleme für gewöhnliche Differentialgleichungen wird für diesen Kurs vorausgesetzt. Diese Grundbegriffe werden in den Vorlesungen Analysis I & II im Basisjahr des Bachelorstudiums Mathematik vermittelt und sind in [3, Kap. 8 & Sekt. 11.6]. Zur Wiederholung wird ein Studium dieser Abschnitte empfohlen.

Das erste Kapitel der Vorlesung frischt die theoretischen Grundlagen für Anfangswertprobleme für gewöhnliche Differentialgleichungen nochmals auf und stellt wichtige Beispiele vor. Das Verhalten von einfachen numerischen Verfahren wird anhand dieser Beispiele diskutiert.


# 1.1 Anfangswertprobleme (AWP)

Eine **gewöhnliche Differentialgleichung** erster Ordnung (engl. *first-order ordinary differential equation* (ODE)):

$$\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}) \quad (1.1.1)$$

In dieser Vorlesung verwendete Terminologie, siehe [8]:

- $\mathbf{f} : I \times D \mapsto \mathbb{R}^d \hat{=} \text{rechte Seite}$  ( $d \in \mathbb{N}$ )
- $I \subset \mathbb{R} \hat{=} \text{(Zeit)intervall} \leftrightarrow \text{“Zeitvariable” } t$
- offene Teilmenge  $D \subset \mathbb{R}^d \hat{=} \text{Zustandsraum/Phasenraum}$  (engl. *state space/phase space*)  
 $\leftrightarrow \text{“Zustandsvariable” } \mathbf{y}$  (Beschreibt “Zustand” eines Systems durch  $d$  reelle Zahlen)
- $\Omega := I \times D \hat{=} \text{erweiterter Zustandsraum}$  (enthält Tupel  $(t, \mathbf{y})$ )

 Notation (Newton): Punkt  $\dot{\cdot}$   $\hat{=}$  (totale) Ableitung nach der Zeit  $t$

 Notation: Fettdruck für Spaltenvektoren (Komponenten selektiert durch Subscript-Indices, z.B.  $\mathbf{y} = (y_1, \dots, y_d)^T \in \mathbb{R}^d$ )

- Für  $d = 1$  handelt es sich bei (1.1.1) um eine **skalare** gewöhnliche Differentialgleichung.
- Für  $d > 1$  heisst (1.1.1) auch **System gewöhnlicher Differentialgleichungen**:

$$(1.1.1) \quad \iff \quad \frac{d}{dt} \begin{pmatrix} y_1 \\ \vdots \\ y_d \end{pmatrix} = \begin{pmatrix} f_1(t, y_1, \dots, y_d) \\ \vdots \\ f_d(t, y_1, \dots, y_d) \end{pmatrix} .$$

Grundannahme: **stetige** rechte Seite  $\mathbf{f} : I \times D \mapsto \mathbb{R}^d$

**Definition 1.1.2** (Lösung einer gewöhnlichen Differentialgleichung).

Eine Funktion  $\mathbf{y} \in C^1(J, D)$ ,  $J \subset I$  Intervall positiver Länge, heisst **Lösung** der gewöhnlichen Differentialgleichung (1.1.1), falls

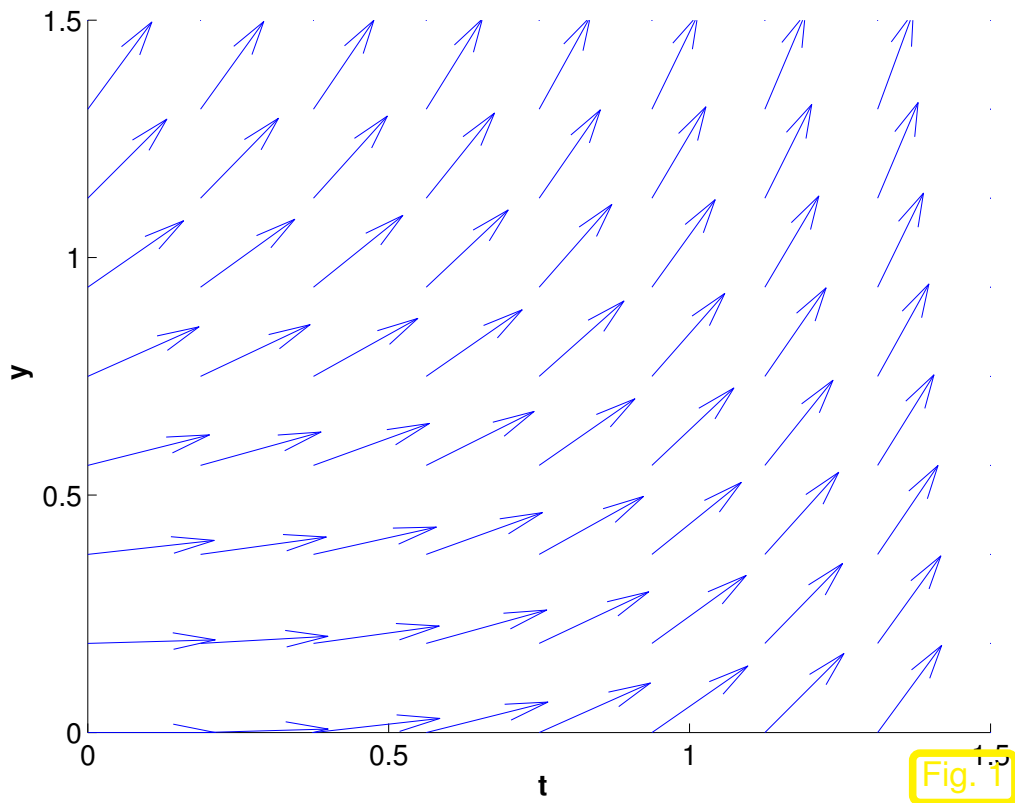
$$\dot{\mathbf{y}}(t) = \mathbf{f}(t, \mathbf{y}(t)) \quad \text{für alle } t \in J .$$

*Beispiel 1.1.3* (Richtungsfeld und Lösungskurven).

**Riccati-Differentialgleichung**

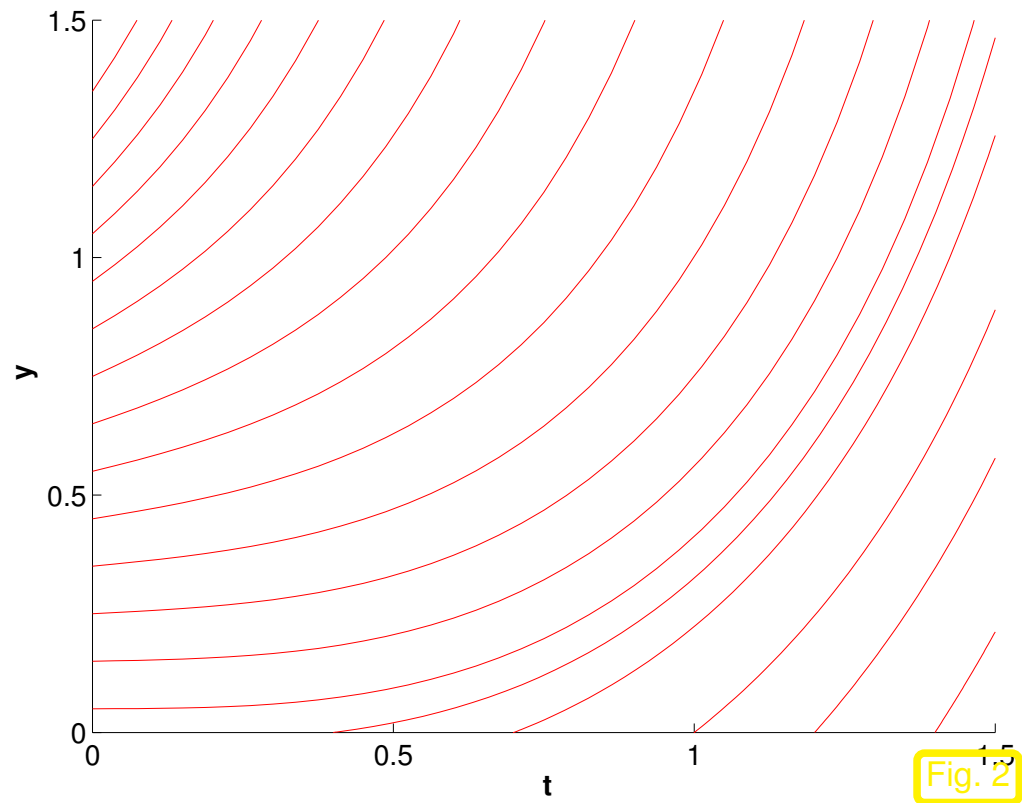
$$\dot{y} = y^2 + t^2 \quad \blacktriangleright \quad d = 1, \quad I, D = \mathbb{R}^+ . \quad (1.1.4)$$

skalare ODE



Richtungsfeld, [3, Fig. 8.1.4]

Fig. 1



Lösungskurven

Fig. 2

Lösungskurven tangential zum Richtungsfeld in jedem Punkt des erweiterten Zustandsraumes.

Alternative Interpretation des *Richtungsfeldes als Geschwindigkeitsfeld* einer Flüssigkeit: die Lösungskurven sind die Trajektorien von Partikeln, die in der Flüssigkeit treiben.

Listing 1.1: Erzeugen der Grafiken zu Beispiel 1.1.3

```
1 % MATLAB function plotting the vector field and the solution curves for the
```

```
2 % Ricatti differential equation for Example 1.1.3
3 function Ricatti
4
5 % define the right hand side of the differential equation as function handle
6 fn = @(t,x) x.^2+t^2;
7
8 % plot solution curves
9 figure ('Name','Ricatti'); hold on;
10 % run ode45 and plot results for different starting values on y-axis
11 for v = 0.05:0.1:1.4
12     [t,y] = ode45(fn,[0 1.5],v);
13     plot(t,y,'r-');
14 end
15 % run ode45 and plot results for different starting values on x-axis
16 for v = [0.4 0.7 1.0 1.2 1.4]
17     [t,y] = ode45(fn,[v 1.5],0);
18     plot(t,y,'r-');
19 end
20 % set axes, labels, ...
21 set(gca, 'fontsize',14); axis ([0 1.5 0 1.5]);
22 xlabel ('{\bf t}'); ylabel ('{\bf y}');
23 % Create EPS output file
24 print -depsc2 'riccatt1.eps'
```



```
25
26 % plot tangent field
27 figure('Name','LV field'); hold on;
28 % set up the sampling grid
29 N = 8; [X,Y] = meshgrid(0:1.5/N:1.5,0:1.5/N:1.5);
30 U = zeros(size(X)); V = zeros(size(Y));
31 % get velocity vectors
32 for i=0:N-1
33     for j=0:N-1
34         x = [1;fn(X(i+1,j+1),Y(i+1,j+1))];
35         x = 0.3*x/norm(x);
36         U(i+1,j+1) = x(1); V(i+1,j+1) = x(2);
37     end
38 end
39
40 % plot velocity vectors
41 quiver(X,Y,U,V,'b-');
42 % set axes, labels, ...
43 set(gca,'fontsize',14); axis([0 1.5 0 1.5]);
44 xlabel('\bf t'); ylabel('\bf y');
45 % Create EPS output file
46 print -depsc2 'riccatti2.eps'
```



Spezialfall:  $\mathbf{f}(t, \mathbf{y}) = \mathbf{f}(\mathbf{y}) \Rightarrow$  **autonome** Differentialgleichung (hier  $I = \mathbb{R}$ )

$$\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}) . \quad (1.1.5)$$

Hier:  $\mathbf{f} : D \subset \mathbb{R}^d \mapsto \mathbb{R}^d$  is ein (stetiges) **Vektorfeld** (“Geschwindigkeitsfeld”, siehe Bsp. 1.1.3).

*Bemerkung 1.1.6* (Translationsinvarianz von Lösungen autonomer Dgl.).

$$t \mapsto \mathbf{y}(t) \text{ Lösung von (1.1.5)} \Rightarrow t \mapsto \mathbf{y}(t + \tau) \text{ Lösung von (1.1.5)} \quad \forall \tau \in \mathbb{R}$$

R. Hiptmair  
rev 35327,  
25. April  
2011




*Bemerkung 1.1.7* (Autonomisierung).

$$\mathbf{z}(t) := \begin{pmatrix} \mathbf{y}(t) \\ t \end{pmatrix} = \begin{pmatrix} \mathbf{z}' \\ z_{d+1} \end{pmatrix} : (1.1.1) \Leftrightarrow \dot{\mathbf{z}} = \mathbf{g}(\mathbf{z}), \quad \mathbf{g}(\mathbf{z}) := \begin{pmatrix} \mathbf{f}(z_{d+1}, \mathbf{z}') \\ 1 \end{pmatrix} . \quad (1.1.8)$$



Verallgemeinerung: Eine **gewöhnliche Differentialgleichung**  $n$ -ter Ordnung,  $n \in \mathbb{N}$ :

$$\mathbf{y}^{(n)} = \mathbf{f}(t, \mathbf{y}, \dot{\mathbf{y}}, \dots, \mathbf{y}^{(n-1)}) \quad (1.1.9)$$

 Notation: Superscript  $^{(n)} \hat{=} n$ . Ableitung nach der Zeit  $t$

► Umwandlung in ODE (System !) erster Ordnung ( $d \leftarrow n \cdot d$ ):

$$\mathbf{z}(t) := \begin{pmatrix} \mathbf{y}(t) \\ \mathbf{y}^{(1)}(t) \\ \vdots \\ \mathbf{y}^{(n-1)}(t) \end{pmatrix} = \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_n \end{pmatrix} \in \mathbb{R}^{dn}: (1.1.9) \Leftrightarrow \dot{\mathbf{z}} = \mathbf{g}(\mathbf{z}), \quad \mathbf{g}(t, \mathbf{z}) := \begin{pmatrix} \mathbf{z}_2 \\ \mathbf{z}_3 \\ \vdots \\ \mathbf{z}_n \\ \mathbf{f}(t, \mathbf{z}_1, \dots, \mathbf{z}_n) \end{pmatrix}. \quad (1.1.10)$$

R. Hiptmair  
rev 35327,  
25. April  
2011

Theorie Numerik für ODEs 1. Ordnung  $\Rightarrow$  Theorie Numerik für ODEs  $n$ . Ordnung !

Vorsicht: (1.1.10) weist *spezielle Struktur* auf, die ein generisches Verfahren für ODEs 1. Ordnung vielleicht nicht zur Verbesserung der Genauigkeit/Verringerung des Rechenaufwandes auszunutzen vermag ( $\rightarrow$  Diskussion in späteren Kapiteln).

*Bemerkung* 1.1.11. Die Transformation (1.1.10) ist nur eine von (unendlich) vielen Möglichkeiten der Transformation von (1.1.9) in eine ODE 1. Ordnung.



Analysis: *symbolisches Rechnen* (Trennung der Variablen, Variation der Konstanten) liefert **allgemeine Lösung** einer ODE als parameterabhängige Funktionenschar, z.B. für eine skalare autonome ODE erhält man formal (mit Kettenregel) das unbestimmte Integral

$$\dot{y} = f(y) \Rightarrow \frac{d}{dt}G(y) = 1 \Rightarrow G(y) = t + C \Rightarrow y(t) = G^{-1}(t + C), \quad (1.1.12)$$

$$\text{mit } G(\eta) = \int_{\eta_0}^{\eta} \frac{1}{f(\xi)} d\xi,$$

wobei  $f(y) \neq 0$  anzunehmen ist.

Allerdings ist eine symbolische Darstellung von  $G$ , geschweige denn von  $G^{-1}$  meist nicht verfügbar.

Daher sind Darstellungsformeln wie (1.1.12) nur von beschränktem Nutzen und wir sind angewiesen auf *numerische Lösung*. Eine solche kann aber nur eine Approximation einer konkreten Funktion sein, so dass die numerischen Betrachtungen sich auf Probleme konzentrieren, für die Existenz und Eindeutigkeit von Lösungen gewährleistet ist. Das ist nur der Fall, wenn die gewöhnliche Differentialgleichung noch durch Anfangswerte komplettiert wird.

ODE + Anfangsbedingungen = **Anfangswertproblem (AWP)**

$$\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}) \quad , \quad \mathbf{y}(t_0) = \mathbf{y}_0 \quad \text{für ein } (t_0, \mathbf{y}_0) \in \Omega . \quad (1.1.13)$$

**Definition 1.1.14** (Lösung eines Anfangswertproblems).

Eine Lösung  $\mathbf{y} : J \mapsto D$ ,  $t_0 \in J$ , von (1.1.1), die  $\mathbf{y}(t_0) = \mathbf{y}_0$  erfüllt, heisst **Lösung** des Anfangswertproblems (1.1.13).

Bem. 1.1.6

*Bemerkung* 1.1.15.      Wenn ODE autonom       $\triangleright$       O.B.d.A. setze  $t_0 = 0$  in (1.1.13)



*Bemerkung* 1.1.16 (Anfangswerte für Dgl. höherer Ordnung).

Anfangswertproblem für gewöhnliche Differentialgleichung  $n$ -ter Ordnung (1.1.9):

$$\mathbf{y}^{(n)} = f(t, \mathbf{y}) \quad , \quad \mathbf{y}(t_0) = \mathbf{y}_0 \quad , \quad \dot{\mathbf{y}}(t_0) = \mathbf{y}_1, \dots, \mathbf{y}^{(n-1)}(t_0) = \mathbf{y}_{n-1} \quad .$$

➔  $n$  unabhängige Anfangswerte sind vorzugeben.



## 1.2 Beispiele und Grundbegriffe

Modellierung: Anfangswertprobleme (1.1.13) beschreiben **deterministische Evolutionsen**

## 1.2.1 Ökologie

*Beispiel* 1.2.1 (Ressourcenbegrenztetes Wachstum). [1, Sect. 1.1]

Autonome **logistische Differentialgleichung**: ( $d = 1$ ,  $D = \mathbb{R}^+$ ,  $I = \mathbb{R}$ )

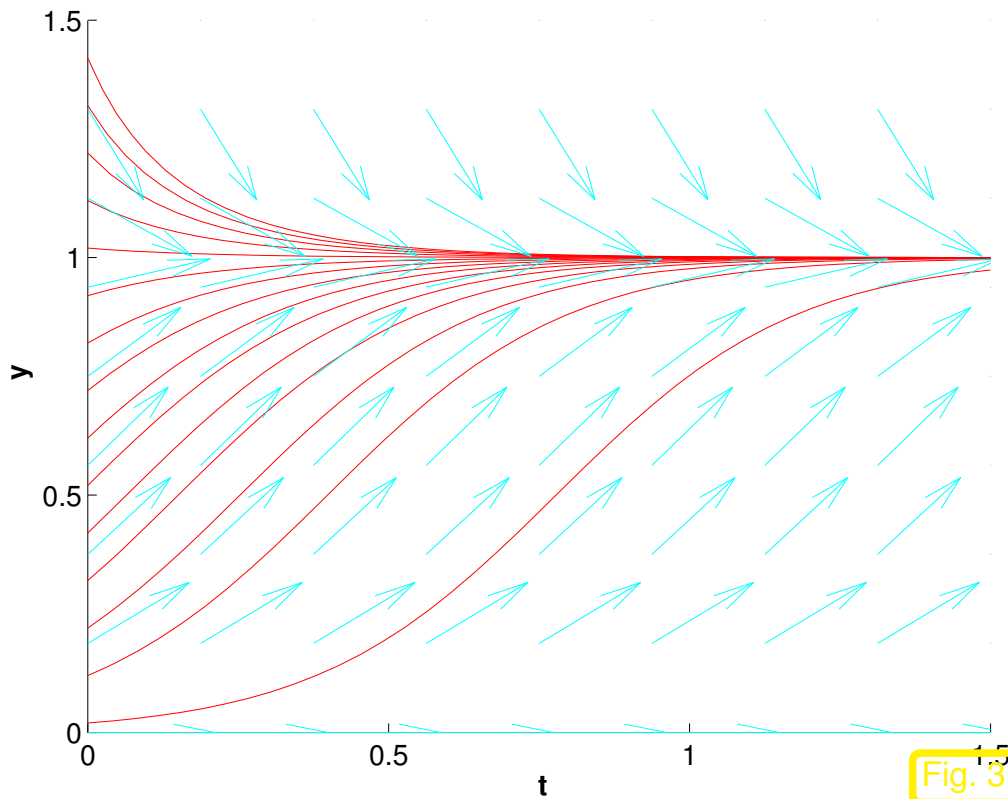
$$\dot{y} = (\alpha - \beta y) y \quad (1.2.2)$$

- $y \hat{=}$  Populationsdichte,  $[y] = \frac{1}{\text{m}^2}$
- Wachstumsrate  $\alpha - \beta y$  mit Wachstumskoeffizienten  $\alpha, \beta > 0$ ,  $[\alpha] = \frac{1}{\text{s}}$ ,  $[\beta] = \frac{\text{m}^2}{\text{s}}$

R. Hiptmair  
rev 35327,  
25. April  
2011

Allgemein für ODE (1.1.1):

$\mathbf{y}^* \in D$ ,  $\mathbf{f}(t, \mathbf{y}^*) = 0 \quad \forall t \quad \triangleright \quad \mathbf{y}^*$  ist **Fixpunkt/stationärer Punkt** für die ODE  $\rightarrow$  Sect. 3.2.

Richtungsfeld und Lösungskurven ( $\alpha, \beta = 5$ )

- Attraktiver Fixpunkt  $y = \alpha/\beta$
- Repulsiver Fixpunkt  $y = 0$

Separation der Variablen (1.1.12)

➔ Lösung des AWP für (1.2.2)

mit  $y(0) = y_0 > 0$ 

$$y(t) = \frac{\alpha y_0}{\beta y_0 + (\alpha - \beta y_0) \exp(-\alpha t)}, \quad (1.2.3)$$

für alle  $t \in \mathbb{R}$ 

MATLAB-CODE: ODE-Integration

```
fn = @(t,y) 5*y*(1-y);
[t,y] = ode45(fn, [0 1.5], y0);
plot(t,y,'r-');
```

Numerische Integration der logistischen Differentialgleichung in MATLAB mittels Funktion `ode45()`:

- Funktions-Handle zur Übergabe der rechten Seite
- Zeitintervall  $[t_0, T]$
- Anfangswert  $y_0$

Rückgabewerte:  $t \hat{=}$  (Spalten)vektor von Zeitpunkten,  $y \hat{=}$  (Spalten)vektor von Lösungswerten  $\diamond$



Bemerkung 1.2.4 (AWP-Löser in MATLAB). → [31]

Aufrufsyntax:

```
[t, y] = solver(odefun, tspan, y0);
```

Funktionsargumente:

`solver` :  $\in \{ \text{ode23}, \text{ode45}, \text{ode113}, \text{ode15s}, \text{ode23s}, \text{ode23t}, \text{ode23tb} \}$   
`odefun` : Funktions-Handle vom Typ  $@(t, y) \leftrightarrow$  rechte Seite  $\mathbf{f}(t, \mathbf{y})$   
`tspan` : 2-Vektor  $(t_0, T)^T$ : Anfangs- und Endzeitpunkt für numerische Integration  
`y0` : Anfangswert  $\mathbf{y}_0 \in \mathbb{R}^d$

Rückgabewerte: `t` : (Spalten)vektor von Zeitpunkten  $t_0 < t_1 < t_2 < \dots < t_N = T$   
`y` :  $(N + 1) \times d$  Lösungsmatrix,  $i$ . Zeile  $\sim \mathbf{y}(t_i)$

Warum bietet MATLAB so viele verschiedene Löser für AWPe an?

Die Antwort auf diese Frage und die Auswahl des “richtigen” Löser wird eines der Kernthemen der Vorlesung sein.

Beispiel 1.2.5 (Räuber-Beute-Modelle). [1, Sect. 1.1] & [16, Sect. 1.1.1]

Autonome Lotka-Volterra-Dgl.: ( $d = 2$ )

$$\begin{aligned} \dot{u} &= (\alpha - \beta v)u \\ \dot{v} &= (\delta u - \gamma)v \end{aligned}, \quad I = \mathbb{R}, \quad D = (\mathbb{R}^+)^2, \quad \alpha, \beta, \gamma, \delta > 0. \tag{1.2.6}$$

Populationsdichten:

$u \rightarrow$  Beute,  
 $v \rightarrow$  Räuber

Vektorfeld  $f$  für Lotka-Volterra-Dgl. ▷

Lösungskurven sind **Trajektorien** von Partikeln, die vom Geschwindigkeitsfeld  $f$  mitgetragen werden.

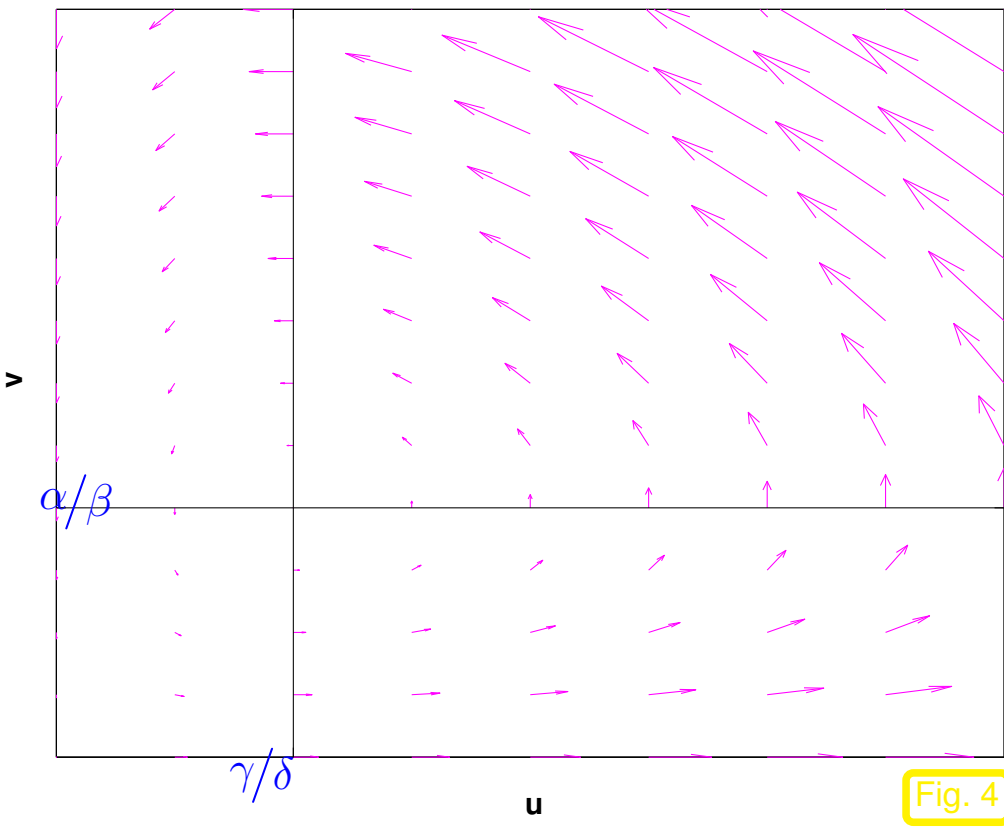
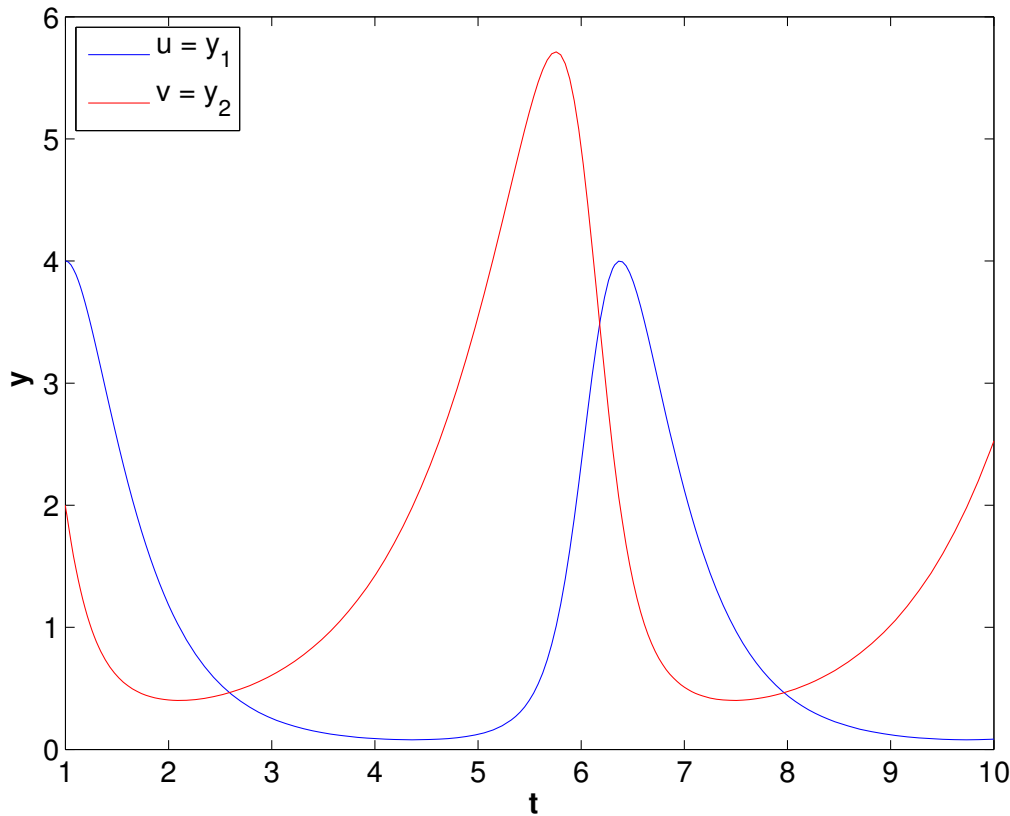


Fig. 4

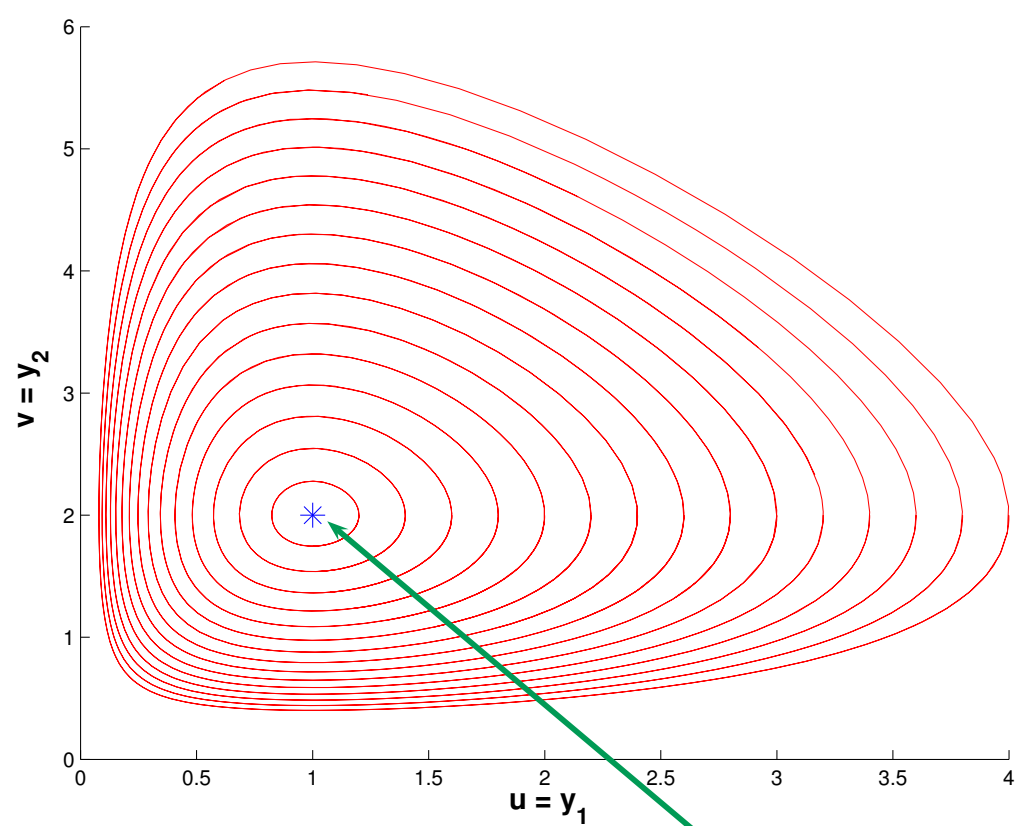
$$(1.2.6) \Rightarrow 0 = \left(\delta - \frac{\gamma}{u}\right)\dot{u} - \left(\frac{\alpha}{v} - \beta\right)\dot{v} = \frac{d}{dt} \underbrace{(\delta u - \gamma \log u - \alpha \log v + \beta v)}_{=: I(u,v)} = 0 .$$

▶ Falls  $(u(t), v(t))$  Lösung von (1.2.6)  $\Rightarrow I(u(t), v(t)) \equiv \text{const}$

▶ Lösungen von (1.2.6) sind **Niveaulinien** von  $I$



Zeitabhängige Lösung,  $\mathbf{y}_0 := \begin{pmatrix} u(0) \\ v(0) \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$



Lösungskurven von (1.2.6)  $\leftrightarrow$  Niveaulinien von  $I$   
Fixpunkt

Geschlossene Lösungskurven  $\leftrightarrow$  (1.2.6) hat ausschliesslich **periodische Lösungen**  
 (für  $u(0), v(0) > 0$ )



**Definition 1.2.7** (Erstes Integral).

Ein Funktional  $I : D \mapsto \mathbb{R}$  heisst **erstes Integral/Invariante** (engl. invariant) der ODE (1.1.1), wenn

$$I(\mathbf{y}(t)) \equiv \text{const}$$

für jede Lösung  $\mathbf{y} = \mathbf{y}(t)$  von (1.1.1).

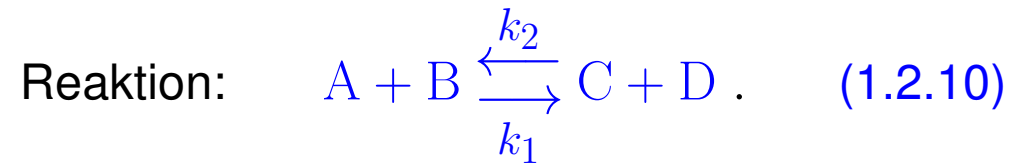
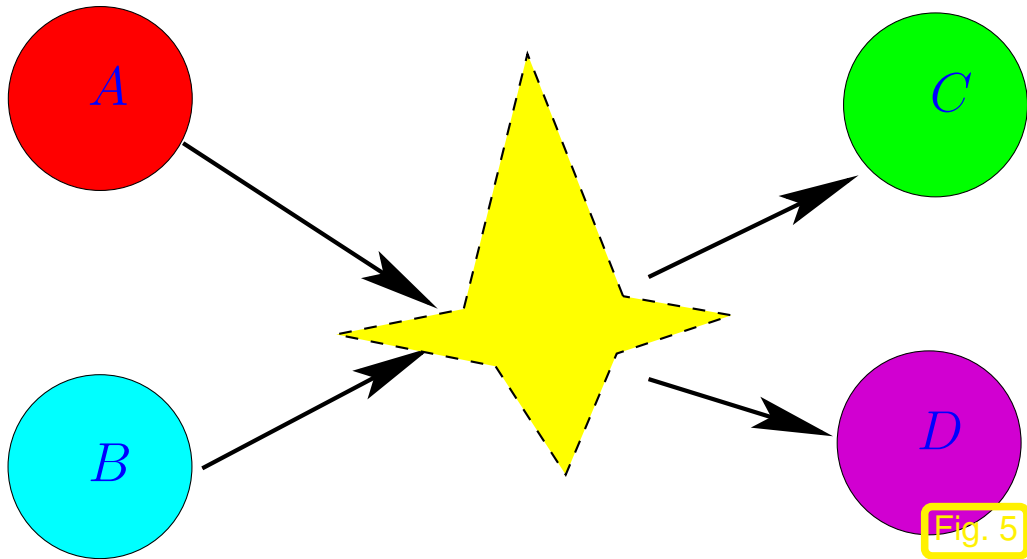
Notwendige und hinreichende Bedingung für differenzierbares erstes Integral

$$I \text{ erstes Integral von (1.1.1)} \Leftrightarrow \text{grad } I(\mathbf{y}) \cdot \mathbf{f}(t, \mathbf{y}) = 0 \quad \forall (t, \mathbf{y}) \in \Omega . \quad (1.2.8)$$

Euklidisches Skalarprodukt

## 1.2.2 Chemische Reaktionskinetik [8, Sect. 1.3]

*Beispiel* 1.2.9 (Bimolekulare Reaktion).



mit Reaktionskonstanten  $k_1$  ("Hinreaktion"),  $k_2$  ("Rückreaktion"),  $[k_1] = [k_2] = \frac{\text{cm}^3}{\text{mol s}}$ .

Fig. 5

Faustregel: Geschwindigkeit einer bimolekularen Reaktion proportional zum Produkt der Konzentrationen der Reaktionspartner:

► für (1.2.10):  $\dot{c}_A = \dot{c}_B = -\dot{c}_C = -\dot{c}_D = -k_1 c_A c_B + k_2 c_C c_D$ . (1.2.11)

$c_A, c_B, c_C, c_D \hat{=}$  (zeitabhängige) Konzentrationen der Reaktanden,  $[c_X] = \frac{\text{mol}}{\text{cm}^3} \rightarrow c_X(t) > 0; \forall t$

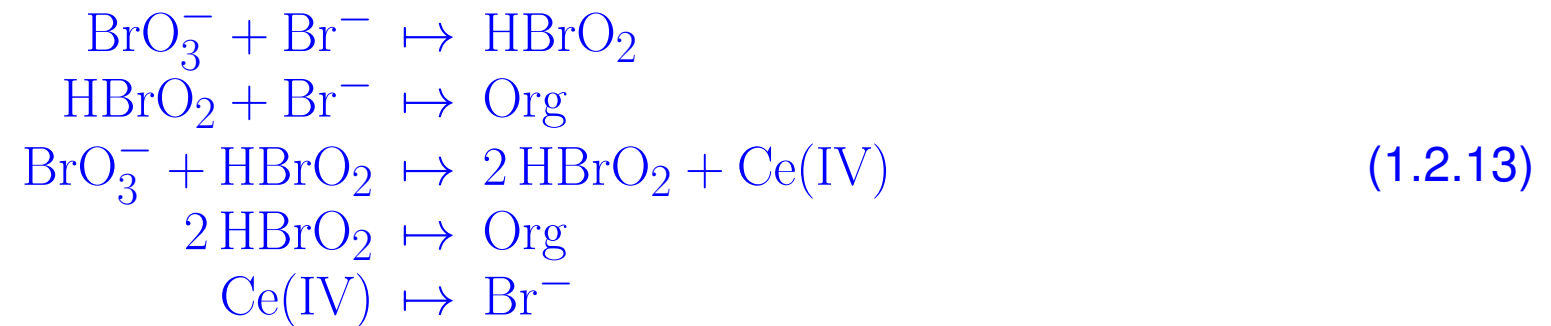
(1.2.11) = autonome gewöhnliche Dgl. (1.1.5) mit

$$\mathbf{y}(t) = \begin{pmatrix} c_A(t) \\ c_B(t) \\ c_C(t) \\ c_D(t) \end{pmatrix}, \quad \mathbf{f}(t, \mathbf{y}) = (-k_1 y_1 y_2 + k_2 y_3 y_4) \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}.$$

► **Massenerhaltung:**  $\frac{d}{dt} (c_A(t) + c_B(t) + c_C(t) + c_D(t)) = 0$

**Beispiel 1.2.12 (Oregonator-Reaktion).**

Spezialfall einer zeitlich oszillierenden Zhabotinski-Belousov-Reaktion [11]:



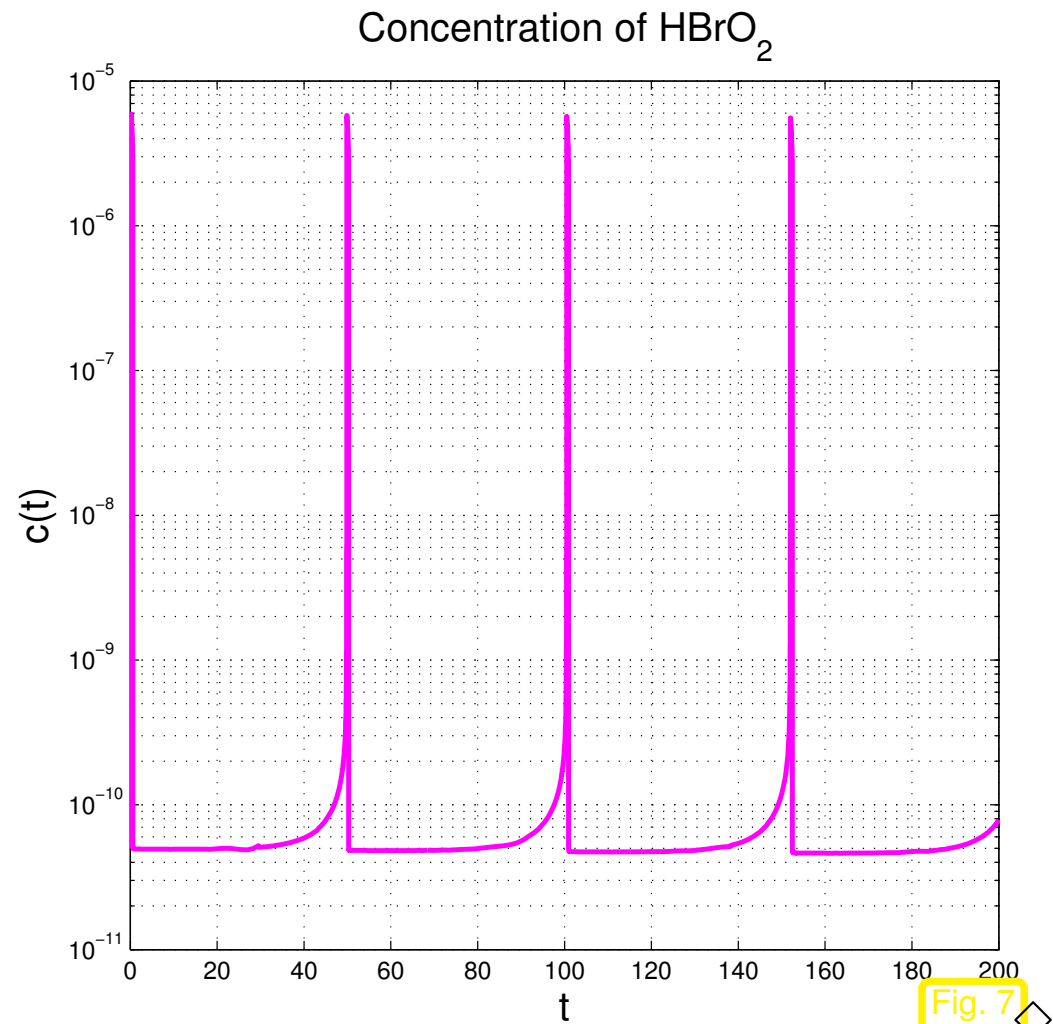
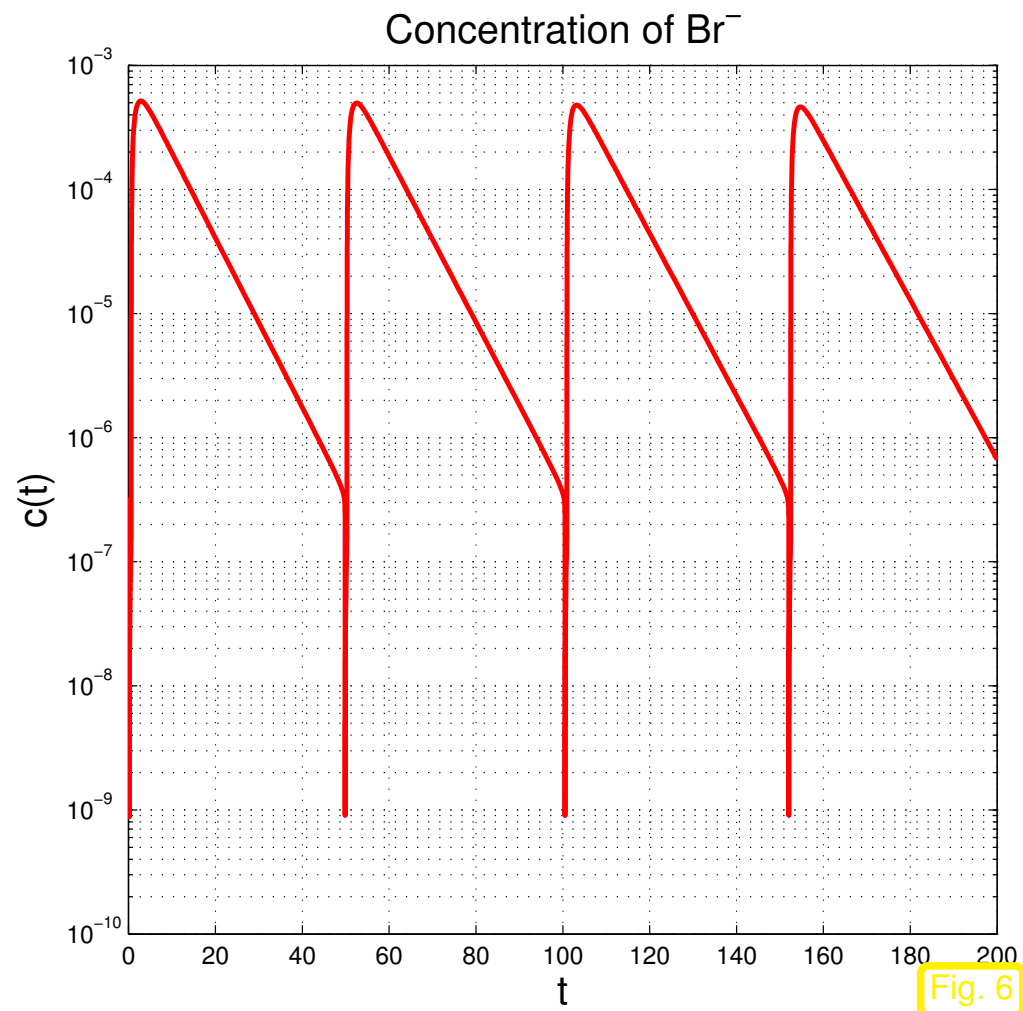
$$\begin{aligned}
 y_1 := c(\text{BrO}_3^-): \quad \dot{y}_1 &= -k_1 y_1 y_2 - k_3 y_1 y_3, \\
 y_2 := c(\text{Br}^-): \quad \dot{y}_2 &= -k_1 y_1 y_2 - k_2 y_2 y_3 + k_5 y_5, \\
 y_3 := c(\text{HBrO}_2): \quad \dot{y}_3 &= k_1 y_1 y_2 - k_2 y_2 y_3 + k_3 y_1 y_3 - 2k_4 y_3^2, \\
 y_4 := c(\text{Org}): \quad \dot{y}_4 &= k_2 y_2 y_3 + k_4 y_3^2, \\
 y_5 := c(\text{Ce(IV)}): \quad \dot{y}_5 &= k_3 y_1 y_3 - k_5 y_5,
 \end{aligned}
 \tag{1.2.14}$$

mit (dimensionslosen) Reaktionskonstanten:

$$k_1 = 1.34, \quad k_2 = 1.6 \cdot 10^9, \quad k_3 = 8.0 \cdot 10^3, \quad k_4 = 4.0 \cdot 10^7, \quad k_5 = 1.0.$$

**Periodische** chemische Reaktion  Video 1, Video 2

MATLAB-Simulation mit Anfangswerten  $y_1(0) = 0.06$ ,  $y_2(0) = 0.33 \cdot 10^{-6}$ ,  $y_3(0) = 0.501 \cdot 10^{-10}$ ,  
 $y_4(0) = 0.03$ ,  $y_5(0) = 0.24 \cdot 10^{-7}$ :



### 1.2.3 Physiologie

Beispiel 1.2.15 (Zeemans Herzschlagmodell). → [6, p. 655]

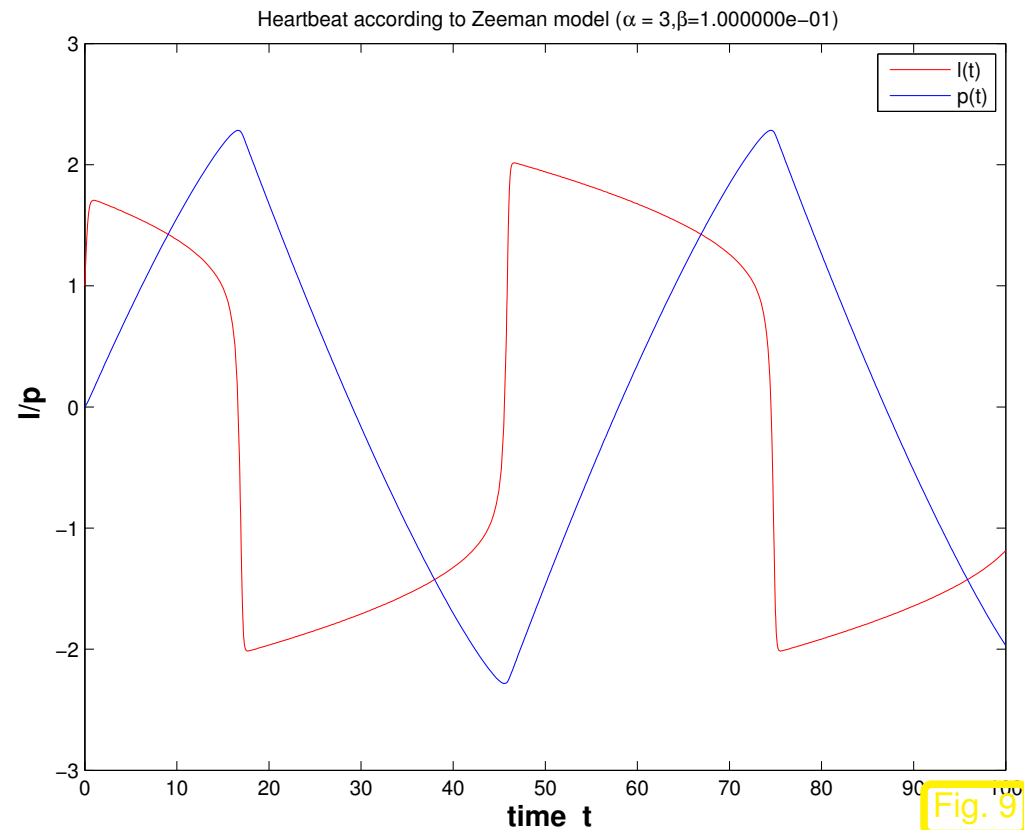
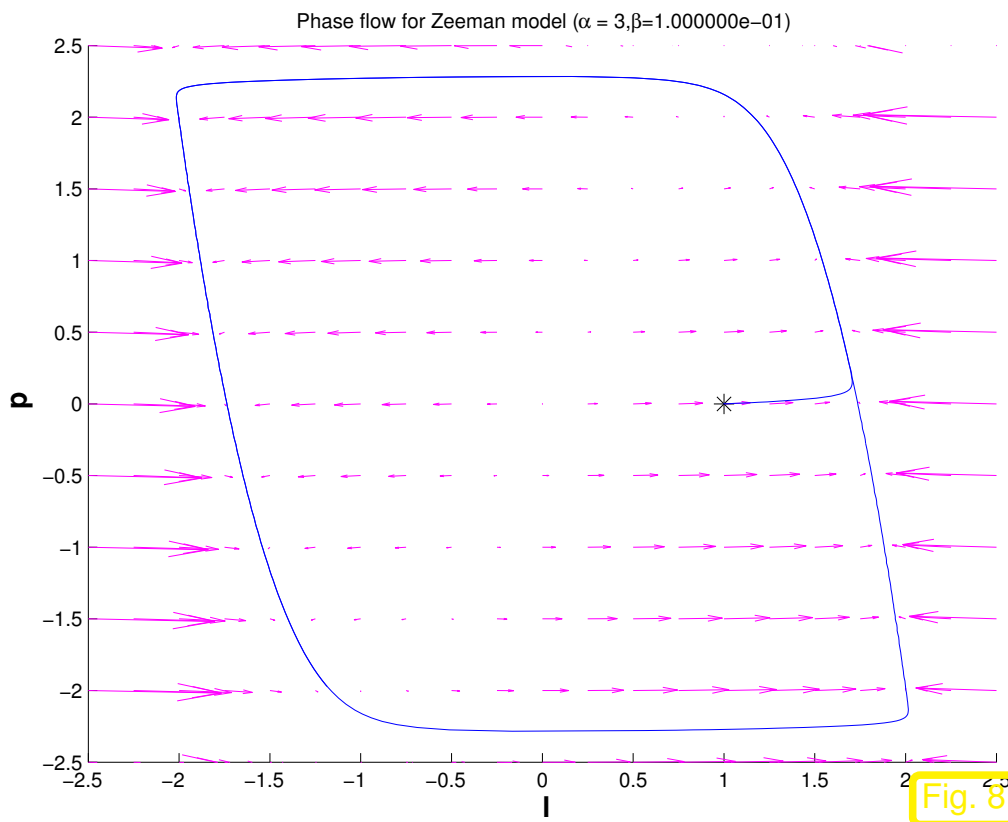


Größen:  $l = l(t) \hat{=}$  Länge der Herzmuskelfaser  
 $p = p(t) \hat{=}$  elektrochemisches Potential

Dimensionsloses **phänomenologisches** Modell: 
$$\begin{aligned} \dot{l} &= -(l^3 - \alpha l + p), \\ \dot{p} &= \beta l, \end{aligned} \tag{1.2.16}$$

mit Parametern:  $\alpha \hat{=}$  Vorspannung der Muskelfaser  
 $\beta \hat{=}$  (phänomenologischer) Rückkopplungsparameter

Vektorfelder und numerische Lösungen für verschiedene Parameter:



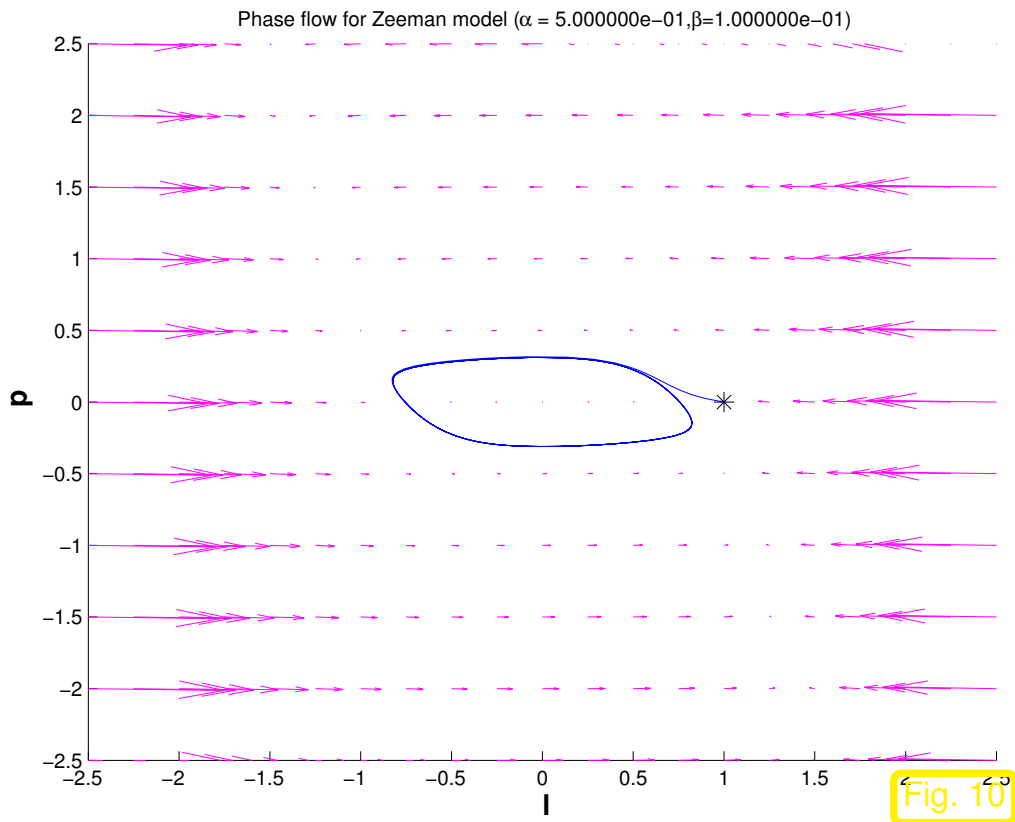


Fig. 10

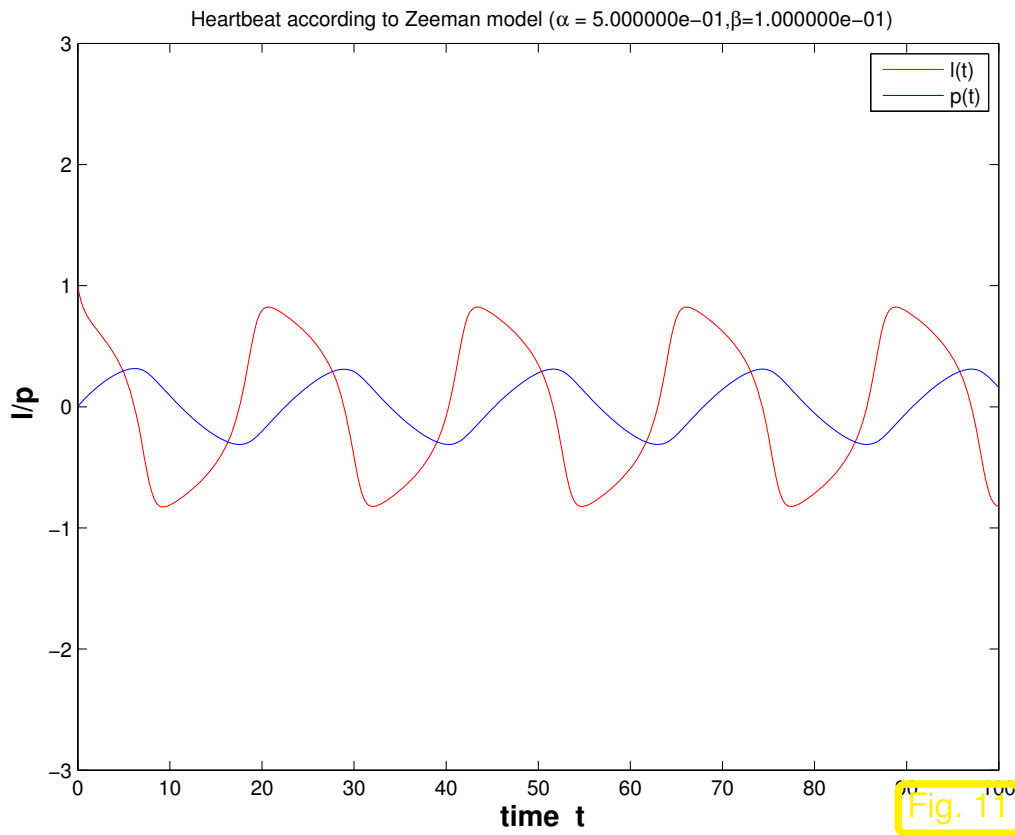


Fig. 11

Beobachtung:  $\alpha \ll 1 \rightarrow$  "Kammerflimmern"

Listing 1.2: Erzeugen der Grafiken zu Beispiel 1.2.15

```

1 function heartbeat
2 % MATLAB function for creating plots for Example 1.2.15
3 beat(3, 'normalbeat'); beat(0.5, 'crazybeat');
4 return
    
```

```
5
6 function beat(alpha,filename)
7 % MATLAB function for numerical simulation of the Zeeman model
8 % (1.2.16) of heartbeat
9
10 if (nargin < 2), filename = 'Heartbeat'; end
11 if (nargin < 1), alpha = 2; end
12
13 % Model equations (right hand side)
14
15 beta = 0.1; % feedback parameter
16 l0 = 0; % length of relaxed muscle fibre
17
18 % Function handle to right hand side vector field
19 f_l = @(l,p) -(l.^3-alpha*l+p);
20 f_p = @(l,p) beta*(l-l0);
21 odefun = @(t,y) [f_l(y(1),y(2)); f_p(y(1),y(2))];
22
23 % Create plot of vector field
24 figure ('name','heartbeat field'); hold on;
25 [L,P] = meshgrid ((-2.5:0.25:2.5), (-2.5:0.5:2.5));
26 quiver (L,P, f_l(L,P), f_p(L,P), 1.5, 'm-');
27 axis ([-2.5 2.5 -2.5 2.5]);
```

```
28 xlabel ('{\bf l}', 'fontsize', 14);
29 ylabel ('{\bf p}', 'fontsize', 14);
30 title (sprintf ('Phase flow for Zeeman model (\alpha =
    %d, \beta=%d)', ...
31         alpha, beta));
32
33 % Compute evolution of l (length) and p (potential), see Rem. 1.2.4
34 tspan = [0 100]; % Duration of simulation
35 [t, y] = ode45 (odefun, tspan, [1; 0], odeset ('abstol', 1E-12));
36
37 % Plot trajectory of solution
38 plot (1, 0, 'k*', 'markersize', 10);
39 plot (y(:, 1), y(:, 2), 'b-');
40 hold off;
41 print ('-depsc2', sprintf ('%s1.eps', filename));
42
43 % Plot time-dependent solution
44 figure ('name', 'heartbeat');
45 plot (t, y(:, 1), 'r-', t, y(:, 2), 'b-');
46 title (sprintf ('heartbeat according to Zeeman model (\alpha =
    %d, \beta=%d)', alpha, beta));
47 xlabel ('{\bf time t}', 'fontsize', 14);
48 ylabel ('{\bf l/p}', 'fontsize', 14);
```

```

49 axis ([tspan -3 3]); legend ('l(t)', 'p(t)');
50
51 print ('-depsc2', sprintf ('%s2.eps', filename));

```



## 1.2.4 Mechanik

*Beispiel* 1.2.17 (Mathematisches Pendel). [1, I.3. Bsp. (3.4c)]

Zustandsraum  $D$  = Konfigurationsraum für **Mini-**  
**malkoordinaten** (= Auslenkungswinkel)

➤  $d = 1$ ,  $D = \mathbb{T}$  (Kreislinie = "1D Torus")

Auslenkungswinkel  $\alpha \in [-\pi, \pi]$

Newtonsche Bewegungsgleichungen:

$$ml \ddot{\alpha}(t) = -mg \sin \alpha(t). \quad (1.2.18)$$

▶ autonome ODE 2. Ordnung, siehe (1.1.9)

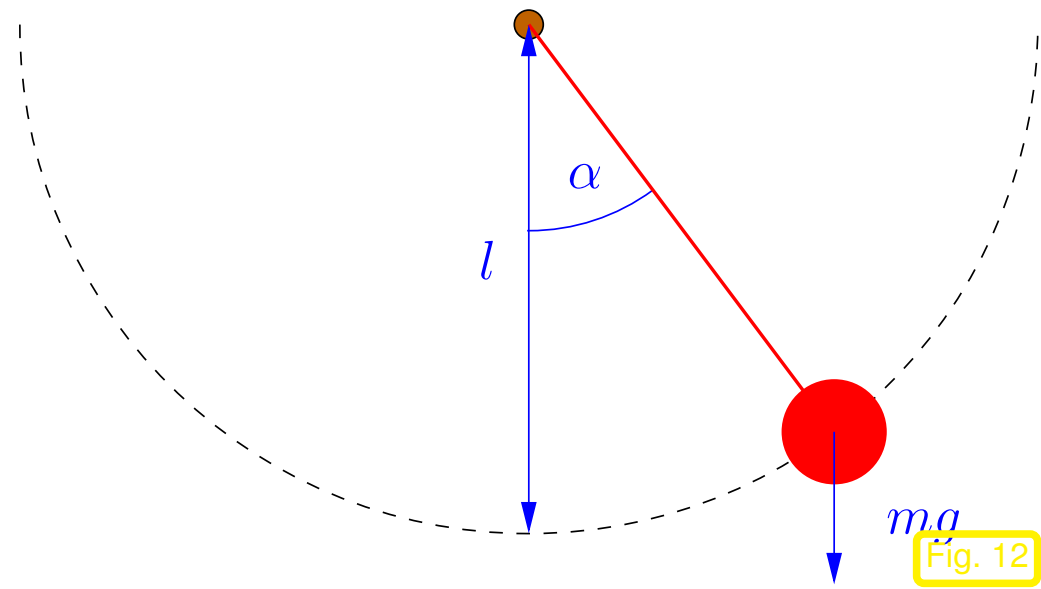


Fig. 12

# Formale Umwandlung in gewöhnliche Differentialgleichungen 1. Ordnung:

Winkelgeschwindigkeit  $p := \dot{\alpha} \Rightarrow \frac{d}{dt} \begin{pmatrix} \alpha \\ p \end{pmatrix} = \begin{pmatrix} p \\ -\frac{g}{l} \sin \alpha \end{pmatrix} . \quad (1.2.19)$

Energieerhaltung:  $E(t) = \frac{1}{2}ml^2 p(t)^2 - mgl \cos \alpha(t) \Rightarrow E(t) \equiv \text{const.}$  (1. Integral  $\rightarrow$  Def. 1.2.)

↑  
kinetische Energie  $T(t)$

←  
potentielle Energie  $U(t)$

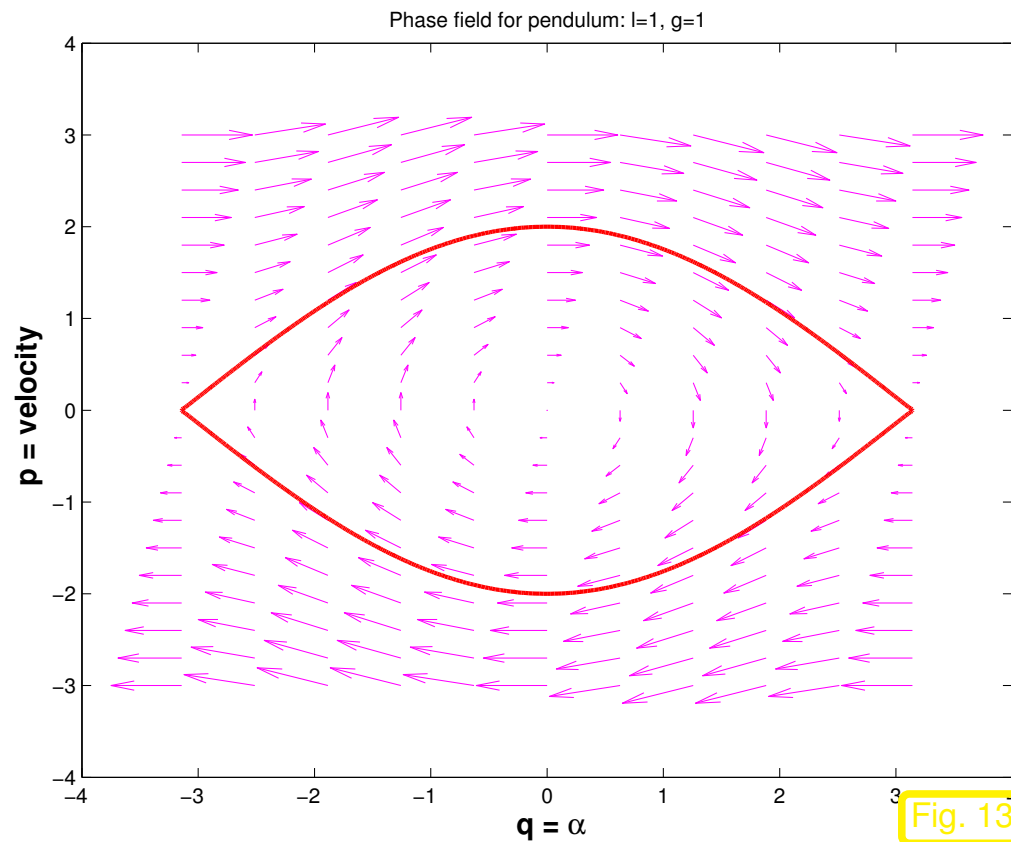


Fig. 13

Vektorfeld für (1.2.19)

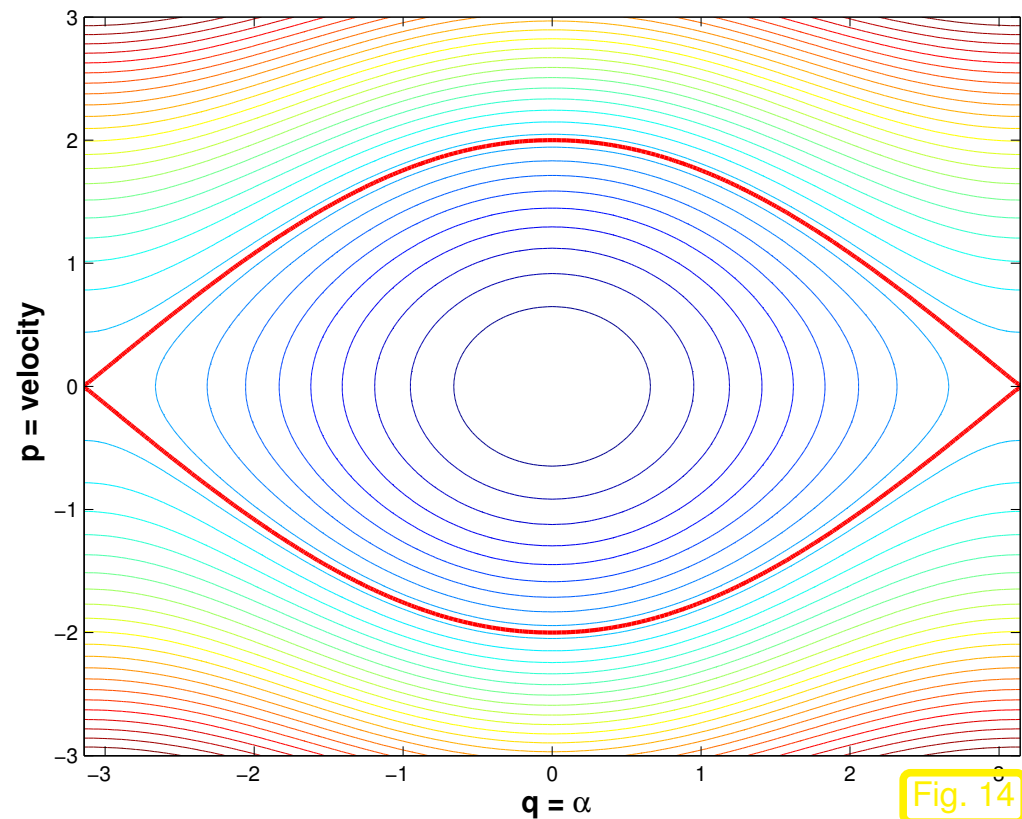


Fig. 14

Isolinien der Energie  $\leftrightarrow$  Lösungskurven

**Definition 1.2.20** (Hamiltonsche Differentialgleichung).  $\rightarrow$  [16, Sect. VI.1.2]

Es sei  $n \in \mathbb{N}$ ,  $M \subset \mathbb{R}^n$  offen,  $H : \mathbb{R}^n \times M \mapsto \mathbb{R}$ ,  $H = H(\mathbf{p}, \mathbf{q})$ , stetig differenzierbar. Dann heisst die gewöhnliche Differentialgleichung erster Ordnung

$$\dot{\mathbf{p}}(t) = -\frac{\partial H}{\partial \mathbf{q}}(\mathbf{p}(t), \mathbf{q}(t)) \quad , \quad \dot{\mathbf{q}}(t) = \frac{\partial H}{\partial \mathbf{p}}(\mathbf{p}(t), \mathbf{q}(t)) \quad , \quad (1.2.21)$$

ein *autonomes Hamiltonsches System* mit *Hamilton-Funktion* (engl. Hamiltonian)  $H$ .

*Bemerkung 1.2.22.* Bewegungsgleichungen der klassischen Mechanik führen auf autonome Hamiltonsches Systeme, siehe [1, Sect. I.3] für eine Einführung, [2] für eine umfassende Darstellung.




**Lemma 1.2.23** (“Energieerhaltungssatz”).

Die Hamilton-Funktion  $H$  ist ein erstes Integral des autonomen Hamiltonschen Systems.

Hamiltonsches System in der Form (1.1.1):

$$\mathbf{y} = \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} \Rightarrow (1.2.21) \Leftrightarrow \dot{\mathbf{y}} = \mathbf{J}^{-1} \cdot \mathbf{grad} H(\mathbf{y}), \quad \mathbf{J} := \begin{pmatrix} 0 & \mathbf{I}_n \\ -\mathbf{I}_n & 0 \end{pmatrix} \in \mathbb{R}^{2n,2n}. \quad (1.2.24)$$

 Notation:  $\mathbf{I}_n \hat{=} n \times n$  Einheitsmatrix

Zusammen mit (1.2.8) folgt sofort Lemma 1.2.23, denn  $\mathbf{J}$  ist *schiefsymmetrisch* ( $\mathbf{J}^T = -\mathbf{J}$ ) und für jede schiefsymmetrische Matrix  $\mathbf{A} \in \mathbb{R}^{n,n}$  gilt  $\mathbf{x} \cdot \mathbf{A}\mathbf{x} = 0 \quad \forall \mathbf{x} \in \mathbb{R}^n$ .


*Beispiel* 1.2.25 (Massenpunkt im Zentralfeld).

Newtonsche Bewegungsgleichungen eines Körpers (Ortskoordinate  $\mathbf{r} = \mathbf{r}(t)$ ) mit Masse  $m > 0$  im Kraftfeld  $\mathbf{f} : \mathbb{R}^n \mapsto \mathbb{R}^n, n \in \mathbb{N}$ :

$$m \ddot{\mathbf{r}}(t) = \mathbf{f}(\mathbf{r}(t)). \quad (1.2.26)$$

Spezialfall: radialsymmetrisches **konservatives** Kraftfeld

$$\mathbf{f}(\mathbf{x}) = -\mathbf{grad} U(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^n, \quad U(\mathbf{x}) = G(\|\mathbf{x}\|). \quad (1.2.27)$$

 Notation:  $\|\mathbf{x}\| := \sqrt{x_1^2 + \cdots + x_n^2} \hat{=}$  Euklidische Norm eines Vektors

Speziell  $G(r) = -\frac{G_0}{r}$  : **Keplerproblem**: [16, Sect. I.2], [8, Sect. 1.1]



$\longleftrightarrow$  Hamiltonsches System ( $\rightarrow$  Def. 1.2.20) mit Konfigurationsraum  $M := \mathbb{R}^n \setminus \{0\}$ ,  $\mathbf{q} := \mathbf{r}$ , und

Hamilton-Funktion (Energie)  $H(\mathbf{p}, \mathbf{q}) := \frac{1}{2m} \|\mathbf{p}\|^2 + G(\|\mathbf{q}\|)$  (1.2.28)

$\mathbf{p} := m\dot{\mathbf{r}} \hat{=} \text{Impuls,}$

kinetische Energie

potentielle Energie

$$\dot{\mathbf{p}} = -G'(\|\mathbf{q}\|) \frac{\mathbf{q}}{\|\mathbf{q}\|}, \quad \dot{\mathbf{q}} = m^{-1} \mathbf{p}. \quad (1.2.29)$$

*Lemma 1.2.30 (Bahnebene).*

Jede Lösung  $\begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} : J \subset \mathbb{R} \mapsto \mathbb{R}^{2n}$  von (1.2.29) erfüllt

$$\mathbf{p}(t), \mathbf{q}(t) \in \text{Span} \{ \mathbf{p}(t_0), \mathbf{q}(t_0) \} \quad \forall t_0, t \in J.$$

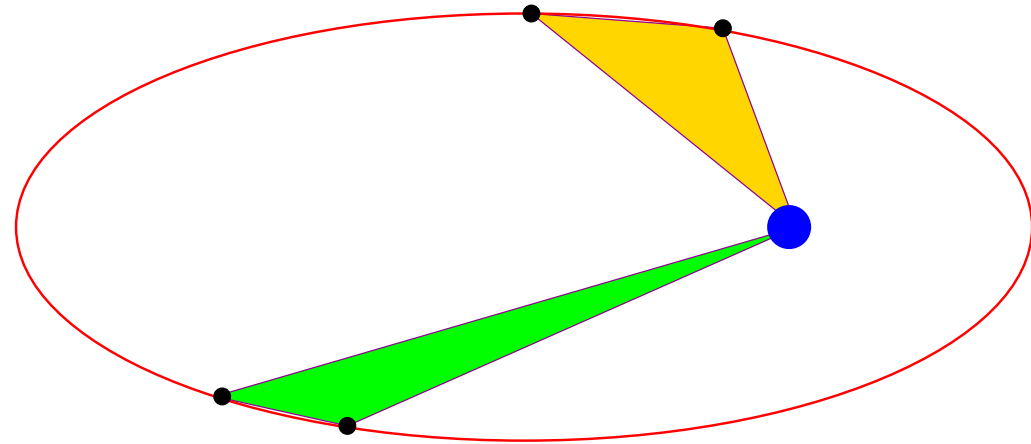
*Lemma 1.2.32 (Drehimpulserhaltung).*

Für  $n = 3$  ist der Drehimpuls **Drehimpuls** (bzgl. 0)  $M := \mathbf{p} \times \mathbf{q}$  (engl. angular momentum) ein erstes Integral ( $\rightarrow$  Def. 1.2.7) von (1.2.29).

Notation:  $\times \hat{=} \text{Vektorprodukt im } \mathbb{R}^3$ :

$$\mathbf{a} \times \mathbf{b} = \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix}.$$

*Theorem 1.2.33 (2. Keplersches Gesetz).  
Löst  $t \mapsto \mathbf{q}(t)$  die Differentialgleichung  
(1.2.29), so überstreicht der Vektor  $\mathbf{q}(t)$  in  
gleichen Zeitspannen gleiche Flächen in der  
Bahnebene.*



R. Hiptmair  
rev 35327,  
25. April  
2011

1. Keplersches Gesetz (für Gravitationspotential):

Für  $G(r) = -\frac{G_0}{r}$  sind die Lösungskurven von (1.2.26) Ellipsen mit Brennpunkt  $0$ .

Ausblick: Zur Simulation der Planetenbewegung in unserem Sonnensystem müsste man (1.2.29) über einen langen Zeitraum integrieren. Leider ist eine solche Langzeitintegration nicht unproblematisch. Zwar erhalten sowohl das implizite Euler- als auch das Störmer-Verlet-Verfahren den Dre-

himpuls exakt, jedoch nicht die Energie des Systems (Hamilton-Funktion) [16, Table I.2.1]. Diese Problematik wird in Abschnitt 4.4 eingehend behandelt.



## 1.3 Theorie [8, Sect. 2], [1, Ch. II]

Zu gegebener rechter Seite  $f : \Omega := I \times D \mapsto \mathbb{R}^d$ ,  $d \in \mathbb{N}$ ,  $I \subset \mathbb{R}$  offenes Intervall,  $D \subset \mathbb{R}^d$  offene Menge, betrachte das Anfangswertproblem

$$\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}) \quad , \quad \mathbf{y}(t_0) = \mathbf{y}_0 \quad \text{für gegebenes } (t_0, \mathbf{y}_0) \in \Omega . \quad (1.1.13)$$

# 1.3.1 Existenz und Eindeutigkeit von Lösungen

Zwei wichtige Begriffe:

**Definition 1.3.1** (Maximale Fortsetzbarkeit einer Lösung).

Eine Lösung  $\mathbf{y} \in C^1([t_0, t_+[ , D)$  ( $\rightarrow$  Def. 1.1.2) des AWP (1.1.13) heisst *maximal (in die Zukunft) fortgesetzt*, wenn genau einer der drei folgenden Fälle zutrifft

(i)  $t_+ = \infty$  (Lösung existiert für alle Zeiten)

(ii)  $t_+ < \infty$ , („Blow-up“)

$$\lim_{t \rightarrow t_+} \|\tilde{\mathbf{y}}(t)\| = \infty$$

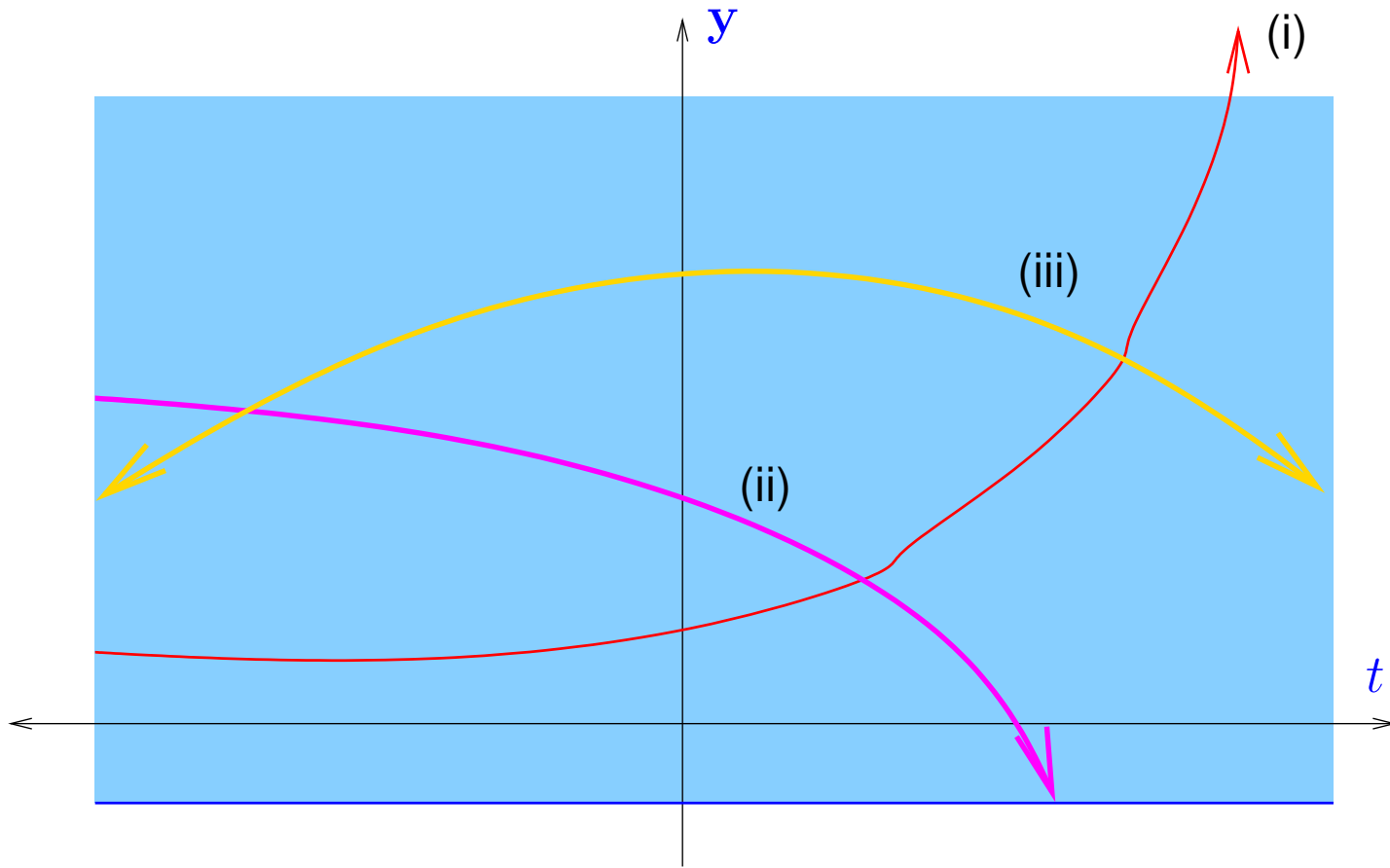
(iii)  $t_+ < \infty$ , („Kollaps“)

$$\lim_{t \rightarrow t_+} \text{dist}((t, \tilde{\mathbf{y}}(t)), \partial\Omega) = 0 .$$

(Analog: Maximal fortgesetzt in die Vergangenheit auf  $]t_-, t_0]$ )

Erinnerung:  $\Omega := I \times D$  ist der erweiterte Zustandsraum

➤ Kollaps  $\leftrightarrow$  Lösung läuft zum Rand des erweiterten Zustandsraumes!



■: „Blow-Up”

■: „Kollaps”

■:  $J(t_0, y_0) = \mathbb{R}$

Notation:  $J(t_0, \mathbf{y}_0) = ]t_-, t_+[$  = maximales Existenzintervall für Lösung von AWP (1.1.13).

**Definition 1.3.2** (Lokale Lipschitz-Stetigkeit).

$\mathbf{f} : \Omega \mapsto \mathbb{R}^d$  heisst *lokal Lipschitz-stetig*

$$\forall (t, \mathbf{y}) \in \Omega: \exists \delta > 0, L > 0:$$

$$\Leftrightarrow: \quad \|\mathbf{f}(\tau, \mathbf{z}) - \mathbf{f}(\tau, \mathbf{w})\| \leq L \|\mathbf{z} - \mathbf{w}\|$$

$$\forall \mathbf{z}, \mathbf{w} \in D: \|\mathbf{z} - \mathbf{y}\| < \delta, \|\mathbf{w} - \mathbf{y}\| < \delta, \forall \tau \in I: |t - \tau| < \delta .$$

Lokale Lipschitz-Stetigkeit impliziert globale Lipschitz-Stetigkeit auf jeder *kompakten* Teilmenge  $K$  von  $\Omega$ :

$$\exists L = L(K) > 0: \quad \|\mathbf{f}(\tau, \mathbf{z}) - \mathbf{f}(\tau, \mathbf{w})\| \leq L \|\mathbf{z} - \mathbf{w}\| \quad \forall (\tau, \mathbf{z}), (\tau, \mathbf{w}) \in K .$$

 Notation:  $D_{\mathbf{y}}\mathbf{f} \hat{=}$  Ableitung von  $\mathbf{f}$  nach Zustandsvariablen (= Jacobimatrix  $\in \mathbb{R}^{d,d}$  !)

**Lemma 1.3.3** (Kriterium für lokale Lipschitz-Stetigkeit).

Sind  $\mathbf{f}$  und  $D_{\mathbf{y}}\mathbf{f}$  stetig auf  $\Omega$ , so ist  $\mathbf{f}$  lokal Lipschitz-stetig ( $\rightarrow$  Def. 1.3.2).

**Theorem 1.3.4** (Satz von Peano & Picard-Lindelöf). [1, Satz II(7.6)]

Falls  $\mathbf{f} : \hat{\Omega} \mapsto \mathbb{R}^d$  lokal Lipschitz-stetig in der Variablen  $\mathbf{y}$  ( $\rightarrow$  Def. 1.3.2), so hat das AWP (1.1.13) für beliebige Anfangsbedingungen  $(t_0, \mathbf{y}_0) \in \Omega$  eine eindeutige maximal fortgesetzte ( $\rightarrow$  Def. 1.3.1) Lösung  $\mathbf{y} : J(t_0, \mathbf{y}_0) \mapsto D$ .

*Beweisidee:* ( $\rightarrow$  [32, I.§6], [3, Sekt. 11.6]) Integration von (1.1.13) liefert

$$\mathbf{y}(t) = \mathbf{y}_0 + \int_{t_0}^t \mathbf{f}(s, \mathbf{y}(s)) \, ds, \quad t \geq t_0. \quad (1.3.5)$$

Definiere Raum

$$\mathcal{F} = \{\mathbf{y} \in C([t_0, t_1[), \mathbf{y}(t_0) = \mathbf{y}_0\}$$

für ein  $t_1 > t_0$  und den Operator

$$T : \mathcal{F} \rightarrow \mathcal{F}, \quad T : \mathbf{y} \mapsto \mathbf{z}(t) = \mathbf{y}_0 + \int_{t_0}^t \mathbf{f}(s, \mathbf{y}(s)) \, ds.$$

Damit kann (1.3.5) auf dem Intervall  $[t_0, t_1]$  als Fixpunktgleichung  $T(\mathbf{y}) = \mathbf{y}$  in  $\mathcal{F}$  geschrieben werden. Aus der lokalen Lipschitz-Stetigkeit folgt für genügend kleines  $t_1 > t_0$ , dass  $T$  eine Kontraktion ist. Mit dem Banachschen Fixpunktsatz folgt die Behauptung für das Zeitintervall  $(t_0, t_1)$ . Das maximale Existenzintervall erhält man über Fortsetzung.  $\square$

*Bemerkung* 1.3.6 (Definitionsintervalle von Lösungen von AWPen).

„Die Lösung eines Anfangswertproblems sucht sich Ihren Definitionsbereich selbst“

**!** Definitionsbereich  $J(t_0, \mathbf{y}_0)$  hängt (meist) von  $(t_0, \mathbf{y}_0)$  ab !

Terminologie: Falls  $J(t_0, \mathbf{y}_0) = I \rightarrow$  Lösung  $\mathbf{y} : I \mapsto \mathbb{R}^d$  ist **global**.





**Definition 1.3.7** (Evolutionsoperator).

Die zweiparametrische Familie  $\Phi^{s,t}$  von Abbildungen  $\Phi^{s,t} : D \mapsto D$  heisst **Evolutionsoperator** zur Dgl.  $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$ , wenn

$$t \in J(s, \mathbf{z}) \mapsto \Phi^{s,t} \mathbf{z} \quad \text{Lösung des AWP} \quad \begin{cases} \dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}) \\ \mathbf{y}(s) = \mathbf{z} \end{cases}, \quad \text{für alle } (s, \mathbf{z}) \in \Omega .$$

Definitionsbereich:  $\Phi : \begin{cases} \tilde{\Omega} \mapsto D \\ (t, s, \mathbf{y}) \mapsto \Phi^{s,t} \mathbf{y} \end{cases}, \quad \tilde{\Omega} := \bigcup_{(s, \mathbf{y}) \in \Omega} J(s, \mathbf{y}) \times \{(s, \mathbf{y})\}$

Satz 1.3.4  $\Rightarrow$   $\Phi^{t,t} = \text{Id}$  ,  $\Phi^{s,t} \mathbf{y} = (\Phi^{r,t} \circ \Phi^{s,r}) \mathbf{y}$  ,  $t, r \in J(s, \mathbf{y})$ ,  $(s, \mathbf{y}) \in \Omega$  . (1.3.8)

Konvention: Für autonome Differentialgleichungen (1.1.5) ( $\rightarrow$  Bem. 1.1.15):  $\Phi^t := \Phi^{0,t}$

➤ Falls  $J(0, \mathbf{y}) = \mathbb{R} \quad \forall \mathbf{y} \in D$  aus (1.3.8):

**Gruppe** von Abbildungen von  $D$ :  $\Phi^s \circ \Phi^t = \Phi^{s+t}$  ,  $\Phi^{-t} \circ \Phi^t = \text{Id} \quad \forall t \in \mathbb{R}$  . (1.3.9)

*Bemerkung* 1.3.10 (Numerische Integratoren als approximative Evolutionsoperatoren).

MATLAB-Lösung eines , vgl. Bem. 1.2.4

```
[t, y] = solver(odefun, [t0 T], y0)
```

$\Phi^{s,t}_y$

The diagram shows four purple arrows pointing from the MATLAB code above to the symbol  $\Phi^{s,t}_y$  below. The arrows originate from the arguments 'odefun', '[t0 T]', 'y0', and the 'solver' function name, all pointing towards the evolution operator symbol.

☞ Numerische Lösungsverfahren für Anfangswertprobleme für *eine* gewöhnliche Differentialgleichung realisieren *Approximationen* von Evolutionsoperatoren  $\rightarrow$  Def. 2.1.2. △

*Beispiel* 1.3.11 (Autonome skalare Differentialgleichungen).  $\triangleright d = 1$

- $f(t, y) = -\lambda y, \lambda \in \mathbb{R}$   $\blacktriangleright$  Lösung des AWP  $y(t) = y_0 e^{-\lambda t}, t \in \mathbb{R}$ 
  - existiert für alle Zeiten, d.h.  $]t_-, t_+[ = \mathbb{R}$  für jedes  $y_0$ : globale Lösung.
  - Zugehöriger Evolutionsoperator:

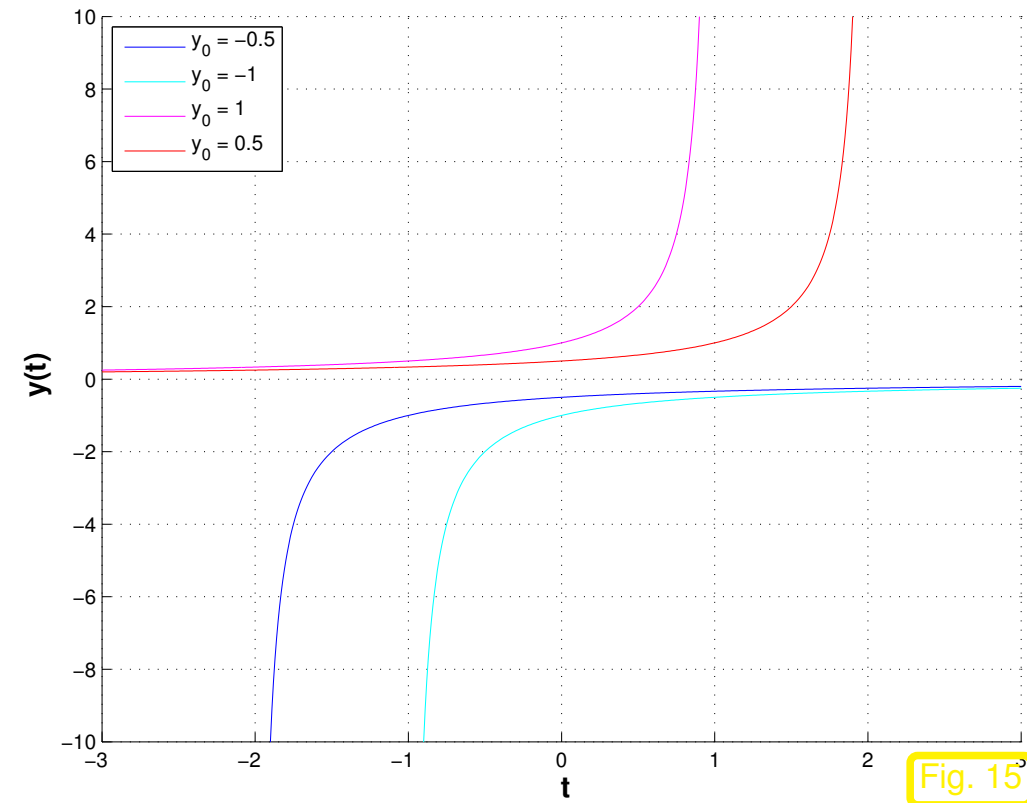
$$\Phi^t : \mathbb{R} \mapsto \mathbb{R} \quad , \quad \Phi^t(y_0) = e^{-\lambda t} y_0 .$$

- $f(t, y) = \lambda y^2, \lambda \in \mathbb{R}$ :  $\dot{y} = \lambda y^2, y(0) = y_0 \in \mathbb{R}$

Lösung:

$$y(t) = \begin{cases} \frac{1}{y_0^{-1} - \lambda t} & , \text{ falls } y_0 \neq 0, \text{ (Blow-up)} \\ 0 & , \text{ falls } y_0 = 0. \end{cases}$$

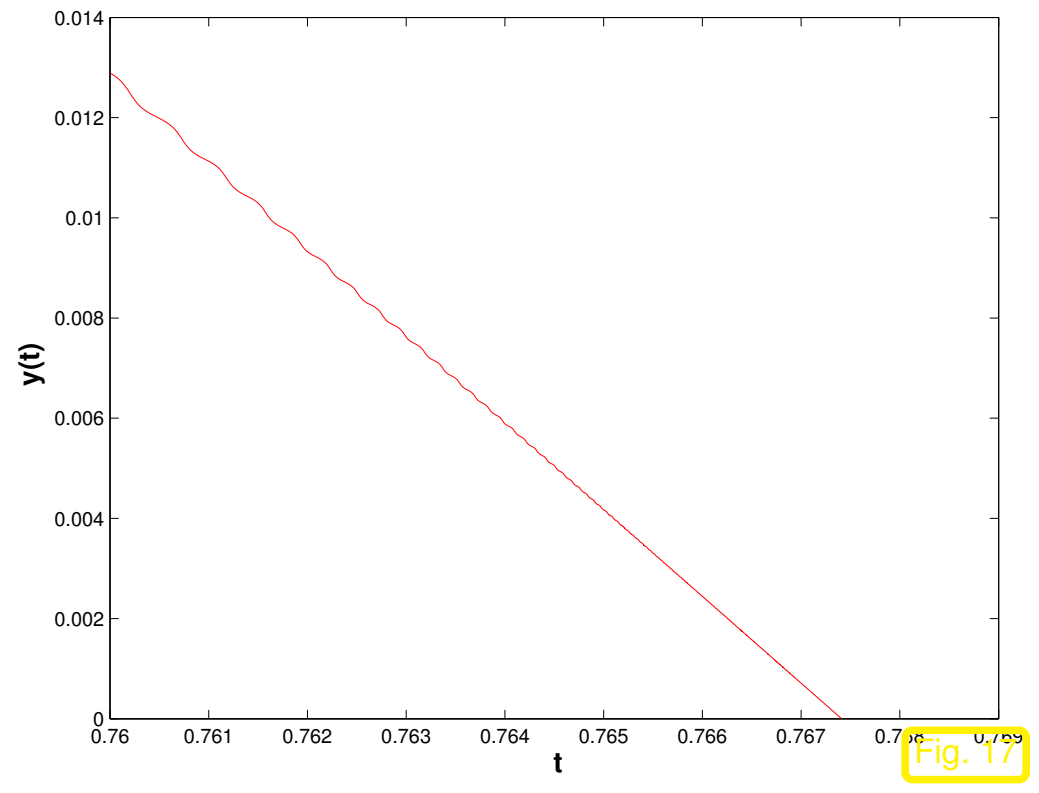
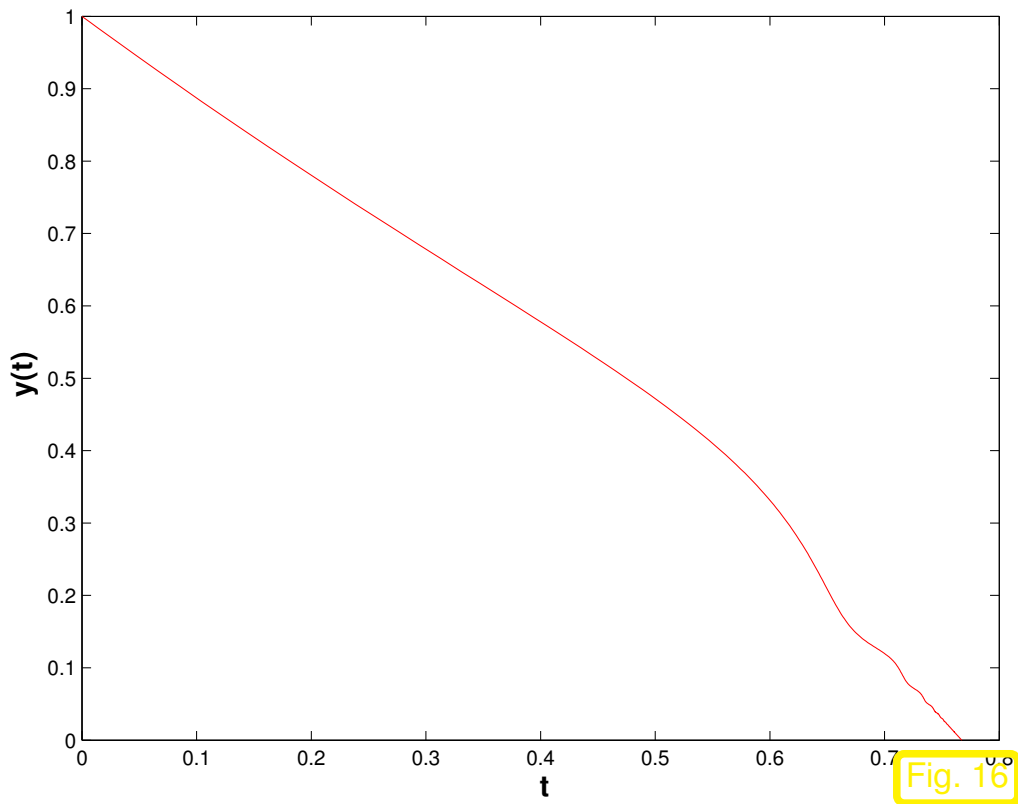
$\lambda, y_0 > 0 \Rightarrow J(0, y_0) = ] -\infty, 1/\lambda y_0[$ .



Lösungskurven für  $\lambda = 1$  ▷

- $f(t, y) = -\frac{1}{\sqrt{y}}$ ,  $D = \mathbb{R}^+$ , Anfangswert  $y(0) = 1$ 
  - $y(t) = (1 - 3t/2)^{2/3}$ ,  $t_- = -\infty$ ,  $t_+ = 2/3$
  - (Lösung läuft zum Rand  $y = 0$  des erweiterten Zustandsraums: **Kollaps**)
- $f(t, y) = \sin(1/y) - 2$ ,  $D = \mathbb{R}^+$ , Anfangswert  $y(0) = 1$  [8, Bsp. 2.14]
  - Lösung  $y(t)$  erfüllt  $\dot{y} \leq -1 \Rightarrow y(t) \leq 1 - t \Rightarrow$  Kollaps für  $t^* < 1$ .

Fig. 15



### 1.3.2 Lineare AWPe [3, Sekt. 8.2]

Vorbereitung: Basiswechsel im Zustandsraum (**kovariante Transformation**):

$$\hat{\mathbf{y}} = \mathbf{S}^{-1}\mathbf{y} \quad , \quad \mathbf{S} \in \mathbb{R}^{d,d} \quad \text{reguläre Matrix (zeitunabhängig).}$$

$$\mathbf{y} \text{ löst } \begin{cases} \dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}) , \\ \mathbf{y}(t_0) = \mathbf{y}_0 \end{cases} \Leftrightarrow \hat{\mathbf{y}} := \mathbf{S}^{-1}\mathbf{y} \text{ löst } \begin{cases} \dot{\hat{\mathbf{y}}} = \hat{\mathbf{f}}(t, \hat{\mathbf{y}}) , \\ \hat{\mathbf{y}}(t_0) = \mathbf{S}^{-1}\mathbf{y}_0 \end{cases} \text{ mit } \hat{\mathbf{f}}(t, \hat{\mathbf{y}}) = \mathbf{S}^{-1}\mathbf{f}(t, \mathbf{S}\hat{\mathbf{y}}) .$$

(1.3.12)

Betrachte Lineare Differentialgleichung mit konstanten Koeffizienten im  $\mathbb{R}^d$ :

- $D = \mathbb{R}^d, \Omega = I \times \mathbb{R}^d$
  - Koeffizientenmatrix  $\mathbf{A} \in \mathbb{R}^{d,d}$
  - „Quellterm“: stetige Funktion  $\mathbf{g} : I \mapsto \mathbb{R}^d$
- $$\dot{\mathbf{y}} = \mathbf{A}\mathbf{y} + \mathbf{g}(t) . \quad (1.3.13)$$

Annahme:  $\mathbf{A}$  diagonalisierbar,  $\exists \mathbf{S} \in \mathbb{R}^{d,d}$  regulär:  $\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \text{diag}(\lambda_1, \dots, \lambda_d), \lambda_i \in \mathbb{C}$

- $\mathbf{g} \equiv 0$ :  $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$  (autonome homogene lineare Dgl., allgemeinere Diskussion [8, Sect. 3.2.2])

$$\hat{\mathbf{y}} := \mathbf{S}^{-1}\mathbf{y} \text{ löst } \begin{array}{l} \dot{\hat{y}}_1 = \lambda_1 \hat{y}_1 , \\ \quad \quad \quad \vdots \\ \dot{\hat{y}}_d = \lambda_d \hat{y}_d \end{array} \Rightarrow \hat{y}_i(t) = (\mathbf{S}^{-1}\mathbf{y}_0)_i e^{\lambda_i t} , \quad t \in \mathbb{R} .$$

$$\blacktriangleright \quad \mathbf{y}(t) = \mathbf{S} \underbrace{\begin{pmatrix} e^{\lambda_1 t} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & e^{\lambda_d t} \end{pmatrix}}_{\text{Matrixexponentialfunktion } \exp(\mathbf{A}t)} \mathbf{S}^{-1} \mathbf{y}_0 .$$

Allgemeine Definition der Matrixexponentialfunktion durch

“Matrixexponentialreihe”:

$$\exp(\mathbf{M}) = \sum_{k=0}^{\infty} \frac{1}{k!} \mathbf{M}^k . \quad (1.3.14)$$

Wichtige Eigenschaft:

Matrixexponentialfunktion kommutiert mit Ähnlichkeitstransformationen

$$\mathbf{M} = \mathbf{S}^{-1} \mathbf{A} \mathbf{S} \quad \Rightarrow \quad \exp(\mathbf{M}) = \mathbf{S}^{-1} \exp(\mathbf{A}) \mathbf{S} \quad \forall \mathbf{A}, \mathbf{M}, \mathbf{S} \in \mathbb{C}^{d,d}, \mathbf{S} \text{ regulär.} \quad (1.3.15)$$

- Inhomogener Fall  $\dot{\mathbf{y}}(t) = \mathbf{A} \mathbf{y}(t) + \mathbf{g}(t)$  **partikuläre Lösung** durch „**Variation der Konstanten**“:

Ansatz:  $\mathbf{y}(t) = \exp(\mathbf{A}t)\mathbf{z}(t)$  mit  $\mathbf{z} \in C^1(\mathbb{R}, \mathbb{R}^d) \rightarrow [1, \text{Thm I(5.14)}]$

$$\dot{\mathbf{y}}(t) = \mathbf{A} \exp(\mathbf{A}t)\mathbf{z}(t) + \exp(\mathbf{A}t)\dot{\mathbf{z}}(t) = \mathbf{A}\mathbf{y}(t) + \mathbf{g}(t) = \mathbf{A} \exp(\mathbf{A}t)\mathbf{z}(t) + \mathbf{g}(t)$$

$$\blacktriangleright \dot{\mathbf{z}}(t) = \exp(-\mathbf{A}t)\mathbf{g}(t) \Rightarrow \mathbf{z}(t) = \mathbf{y}_0 + \int_{t_0}^t \exp(-\mathbf{A}\tau)\mathbf{g}(\tau) d\tau$$

$$\blacktriangleright \mathbf{y}(t) = \boxed{\exp(\mathbf{A}(t - t_0))\mathbf{y}_0} + \boxed{\int_{t_0}^t \exp(\mathbf{A}(t - \tau))\mathbf{g}(\tau) d\tau} =: \Phi^{t_0,t}\mathbf{y}_0 .$$

Lsg. des homogenen Problems

Faltung mit Inhomogenität

Allgemeinere Betrachtungen  $\rightarrow [1, \text{Kap. III}]$ :

*Bemerkung 1.3.16* (Allgemeine Variation-der-Konstanten-Formel).  $\rightarrow [1, \text{Thm. (11.13)}]$

- $\mathbf{A} : J \subset \mathbb{R} \mapsto \mathbb{R}^{d,d}$  stetige Matrixfunktion,  $J \subset \mathbb{R}$  Intervall
- $\mathbf{g} : J \mapsto \mathbb{R}^d$  stetig
- $(s, t) \mapsto \mathbf{E}(s, t) \in \mathbb{R}^{d,d}$  beschreibt Evolutionsoperator, definiert durch

$$\frac{\partial \mathbf{E}}{\partial t}(s, t) = \mathbf{A}(t)\mathbf{E}(s, t) \quad \forall (s, t) \in J \times J, \quad \mathbf{E}(s, s) = \mathbf{I} . \quad (1.3.17)$$

Dann ist die (eindeutige  $\rightarrow$  Thm. 1.3.4) Lösung des nicht-autonomen *linearen* Anfangswertproblems

$$\dot{\mathbf{y}} = \mathbf{A}(t)\mathbf{y} + \mathbf{g}(t) \quad , \quad \mathbf{y}(t_0) = \mathbf{y}_0 \quad ,$$

gegeben durch

$$\mathbf{y}(t) = \mathbf{E}(t, t_0)\mathbf{y}_0 + \int_{t_0}^t \mathbf{E}(t, s)\mathbf{g}(s) ds \quad , \quad t \in J \quad . \quad (1.3.18)$$

*Bemerkung* 1.3.19 (Bedeutung linearer AWPe).

**Linearisierung** um einen stationären Punkt  $\mathbf{f}(\mathbf{y}^*) = 0$ :

$$\mathbf{y} \approx \mathbf{y}^* : \quad \mathbf{f}(\mathbf{y}) = D_{\mathbf{y}}\mathbf{f}(\mathbf{y}^*)(\mathbf{y} - \mathbf{y}^*) + O(|\mathbf{y} - \mathbf{y}^*|^2) \quad ,$$

falls  $\mathbf{f} \in C^2$ .

➤ Lösungen von  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$  verhalten sich *in der Umgebung von  $\mathbf{y}^*$*  (qualitativ) wie Lösungen der linearen ODE  $\dot{\mathbf{y}} = D_{\mathbf{y}}\mathbf{f}(\mathbf{y}^*)\mathbf{y}$ .





# 1.3.3 Sensitivität [8, Sect. 3.1]

## 1.3.3.1 Grundbegriffe

Erinnerung (→ Vorlesung „Numerische Methoden“):

Problem = Abbildung  $\Pi : X \mapsto Y$  von Datenraum  $X$  in den Ergebnisraum  $Y$   
(beide versehen mit Metriken  $d_X, d_Y$ )

Problem ist **wohlgestellt** (engl. *well-posed*), wenn  $\Pi$  stetig.

Kondition/Sensitivität eines Problems:

Mass für Einfluss von Störungen in den Daten auf das Ergebnis

**Absolute Kondition:** 
$$\kappa_{\text{abs}} := \sup_{x, x' \in X, x \neq x'} \frac{d_Y(\Pi(x), \Pi(x'))}{d_X(x, x')} . \quad (1.3.20)$$

Sprachgebrauch: Problem ist „**gut konditioniert**“, wenn „ $\kappa_{\text{abs}} \approx 1$ “

- Wohlgestelltheit und Gutkonditioniertheit eines Problems hängen entscheidend von den gewählten Metriken ab. Diese sind bei praktischen Problemen dadurch bestimmt, “woran der Anwender interessiert ist”.
- Als Folge von unvermeidlichen **Eingabefehlern**, macht nur die numerische Lösung von wohlgestellten Problemen Sinn.
- Absolute Konditionszahl ist die globale Lipschitz-Konstante der Problemabbildung (bzgl. der gewählten Metriken).

**Asymptotische (absolute) Kondition** (linearisierte Störungstheorie):

$$\kappa_{\text{abs}}^{\infty} := \lim_{\delta \rightarrow 0} \sup_{0 < d_X(x, x') < \delta} \frac{d_Y(\Pi(x), \Pi(x'))}{d_X(x, x')}$$

➔  $\kappa_{\text{abs}}^{\infty}$  misst Einfluss “kleiner Störungen”

Technik: Differentielle Konditionsanalyse für differenzierbares  $\Pi : X \subset \mathbb{R}^m \mapsto Y \subset \mathbb{R}^n$ ,

$$\kappa_{\text{abs}}^{\infty} = \sup_{\mathbf{x} \in X} \|D\Pi(\mathbf{x})\|, \quad (1.3.21)$$

wobei  $\|\cdot\| \hat{=}$  Matrixnorm induziert durch Vektornormen auf  $\mathbb{R}^m, \mathbb{R}^n$ .

### 1.3.3.2 Unser Problem: das Anfangswertproblem

Anwendung (der abstrakten Konzepte) auf Anfangswertproblem (1.1.13):

Szenario ❶:

- Eingabedatum  $\mathbf{y}_0$   $\mapsto$  Datenraum  $\mathbb{R}^d$ , Metrik: Euklidische Vektornorm
- Ausgabe  $\mathbf{y}(T)$  zu Endzeitpunkt  $T > t_0$   $\mapsto$  Ergebnisraum  $\mathbb{R}^d$ , Metrik: Euklidische Vektornorm

Szenario ❷:

- Eingabedatum  $\mathbf{y}_0$   $\mapsto$  Datenraum  $\mathbb{R}^d$ , Metrik: Euklidische Vektornorm
- Ausgabe: Lösungsfunktion  $t \in J \subset I \mapsto \mathbf{y}(T)$   
 $\mapsto$  Ergebnisraum  $C^0(J, \mathbb{R}^d)$ , Metrik: Maximumnorm  $\|\cdot\|_{L^\infty(J)}$

Szenario ❸:

- Eingabedaten: Anfangswert  $\mathbf{y}_0$  und rechte Seite  $\mathbf{f}$ 
  - ↳ Datenraum  $\mathbb{R}^d \times C^1(I \times \mathbb{R}^d, \mathbb{R}^d)$ , Metrik: Euklidische Vektornorm & Maximumnorm
- Ausgabe: Lösungsfunktion  $t \in J \subset I \mapsto \mathbf{y}(T)$ 
  - ↳ Ergebnisraum  $C^0(J, \mathbb{R}^d)$ , Metrik: Maximumnorm  $\|\cdot\|_{L^\infty(J)}$

Terminologie: Szenario ❶ :  $\kappa_{\text{abs}}, \kappa_{\text{abs}}^\infty \sim$  **punktweise Kondition**,  
 Szenario ❷ :  $\kappa_{\text{abs}}, \kappa_{\text{abs}}^\infty \sim$  **intervallweise Kondition**

*Beispiel* 1.3.22 (Kondition skalarer linearer Anfangswertprobleme).

$$\dot{y} = \lambda y, \quad \lambda \in \mathbb{R}, \quad y(0) = y_0 \in \mathbb{R}, \quad (1.3.23)$$

$$\Rightarrow y(t) = y_0 \exp(\lambda t), \quad t \in \mathbb{R}. \quad (1.3.24)$$

(Untersuchung für Szenarios ❶ and ❷)

Punktweise Kondition: für Endzeitpunkt  $T$ :

$$\kappa_{\text{abs}} = \exp(\lambda T) \quad \begin{cases} \gg 1 & \text{für } \lambda > 0, \\ \ll 1 & \text{für } \lambda < 0. \end{cases} \quad (1.3.25)$$

Intervallweise Kondition: in  $[0, T]$ :

$$\kappa_{\text{abs}} = \max\{1, \exp(\lambda T)\} \quad \begin{cases} \gg 1 & \text{für } \lambda > 0, \\ 1 & \text{für } \lambda < 0. \end{cases} \quad (1.3.26)$$



### 1.3.3.3 Wohlgestelltheit

Annahme: Rechte Seite  $\mathbf{f} : I \times D \mapsto \mathbb{R}^d$  von (1.1.13) erfüllt **globale Lipschitzbedingung**  
(vgl. *lokale* Lipschitzbedingung aus Def. 1.3.2)

$$\forall t \in I: \exists L(t) > 0: \quad \|\mathbf{f}(t, \mathbf{x}) - \mathbf{f}(t, \mathbf{y})\| \leq L(t) \|\mathbf{x} - \mathbf{y}\| \quad \forall \mathbf{x}, \mathbf{y} \in D \subset \mathbb{R}^d, \quad (1.3.27)$$

für geeignete Vektornorm  $\|\cdot\|$  auf  $\mathbb{R}^d$ .

**Theorem 1.3.28** (Lipschitz-stetige Abhängigkeit vom Anfangswert).

Es seien  $\mathbf{y}, \tilde{\mathbf{y}}$  Lösungen des AWP (1.1.13) zu Anfangswerten  $\mathbf{y}_0 \in D$  bzw.  $\tilde{\mathbf{y}}_0 \in D$ . Unter der Annahme (1.3.27) mit stetigem  $L(t)$  gilt

$$\|\mathbf{y}(t) - \tilde{\mathbf{y}}(t)\| \leq \|\mathbf{y}_0 - \tilde{\mathbf{y}}_0\| \cdot \exp\left(\int_{t_0}^t L(\tau) d\tau\right) \quad \forall t \in I.$$

Hilfsmittel beim Beweis:

**Lemma 1.3.29** (Gronwalls Lemma).  $\rightarrow [1, \text{Sect. II.6}], [8, \text{Lemma 3.9}]$

Sei  $J \subset \mathbb{R}$  Intervall,  $t_0 \in J$ ,  $u, a, \beta \in C^0(J, \mathbb{R}^+)$ ,  $a$  monoton wachsend. Dann gilt

$$u(t) \leq a(|t - t_0|) + \int_{t_0}^t \beta(\tau)u(\tau) d\tau \quad \Rightarrow \quad u(t) \leq a(|t - t_0|) \exp\left|\int_{t_0}^t \beta(\tau) d\tau\right|.$$

- AWP (1.1.13) is **wohlgestellt** unter Annahme (1.3.27) !
- *Schranke* für absolute punktweise Kondition (für Endzeitpunkt  $T$ )

$$\kappa_{\text{abs}} \leq \exp\left(\int_{t_0}^T L(\tau) d\tau\right). \quad (1.3.31)$$

*Bemerkung* 1.3.32 („Gronwall-Schranke“ für Kondition).

Schranke aus (1.3.31) oft extrem pessimistisch !

Beispiel (siehe Bsp. 1.3.22): für skalares lineares AWP mit  $\lambda < 0$ , punktweise Kondition, Endzeitpunkt  $T > 0$  (1.3.23)

$$(1.3.31) \quad \triangleright \quad \kappa_{\text{abs}} \leq e^{|\lambda|T} \xrightarrow{T \rightarrow \infty} \infty \quad \longleftrightarrow \quad \kappa_{\text{abs}} = e^{\lambda T} \xrightarrow{T \rightarrow \infty} 0 .$$



### 1.3.3.4 Asymptotische Kondition

Szenario ❶: Wie wirken sich *kleine Störungen im Anfangswert*  $\mathbf{y}_0$  in (1.1.13) auf die Lösung  $\mathbf{y}(t)$  aus ?

Asymptotische absolute Konditionszahl durch **differentielle Konditionsanalyse**, siehe (1.3.21):

Erforderlich: „Differenzieren der Lösung eines Anfangswertproblems nach dem Anfangswert  $\mathbf{y}_0$ “  
(Dabei wird die Zeit  $t$  als fester „Parameter“ behandelt.)

➤ Zu betrachten ist, mit Evolution  $\Phi^{s,t} \mathbf{y} = \Phi(s, t, \mathbf{y})$

$$\left. \frac{d\mathbf{y}(t)}{d\mathbf{y}_0} \right|_{t \text{ fest}} \longleftrightarrow \frac{\partial \Phi}{\partial \mathbf{y}}(t_0, t, \mathbf{y}_0) .$$

Formales Vorgehen unter der Annahme der Vertauschbarkeit partieller Ableitungen:

$$\frac{d}{dt} \Phi^{t_0, t} \mathbf{y} = \mathbf{f}(t, \Phi^{t_0, t} \mathbf{y}) \quad \text{für festes } \mathbf{y} .$$

➤  $\frac{d}{d\mathbf{y}}$

$$\frac{d}{d\mathbf{y}} \left( \frac{\partial}{\partial t} \Phi^{t_0, t} \mathbf{y} \right) = \frac{d}{d\mathbf{y}} \mathbf{f}(t, \Phi^{t_0, t} \mathbf{y})$$

$$\frac{\partial}{\partial t} \left( \frac{\partial \Phi}{\partial \mathbf{y}}(t_0, t, \mathbf{y}) \right) = \frac{d}{d\mathbf{y}} \mathbf{f}(t, \Phi^{t_0, t} \mathbf{y}) = \frac{\partial \mathbf{f}}{\partial \mathbf{y}}(t, \Phi^{t_0, t} \mathbf{y}) \frac{d}{d\mathbf{y}} \Phi^{t_0, t} \mathbf{y} .$$

(Annahme:  $\mathbf{f}$  nach  $\mathbf{y}$  stetig differenzierbar)

Die **Propagationsmatrix** (Wronski-Matrix),

$$\mathbf{W}(t; t_0, \mathbf{z}) := \left. \frac{d}{d\mathbf{y}} \Phi^{t_0, t} \mathbf{y} \right|_{\mathbf{y}=\mathbf{z}} \in \mathbb{R}^{d,d}, \quad (1.3.33)$$



zum AWP (1.1.13) erfüllt Anfangswertproblem für

$$\text{Variationsgleichung } \frac{d}{dt} \mathbf{W}(t; t_0, \mathbf{y}_0) = \frac{\partial \mathbf{f}}{\partial \mathbf{y}}(t, \Phi^{t_0, t} \mathbf{y}_0) \mathbf{W}(t; t_0, \mathbf{y}_0), \quad (1.3.34)$$

$$\mathbf{W}(t_0; t_0, \mathbf{y}_0) = \mathbf{I}.$$

Beachte: Variationsgleichung = lineare Differentialgleichung auf Zustandsraum  $D = \mathbb{R}^{d,d}$   
 ➔ Matrix-Differentialgleichung der Form

$$\dot{\mathbf{W}} = \mathbf{A}(t) \mathbf{W} \quad \text{mit} \quad \mathbf{A}(t) = \frac{\partial \mathbf{f}}{\partial \mathbf{y}}(t, \mathbf{y}(t)),$$

wobei  $\frac{\partial \mathbf{f}}{\partial \mathbf{y}} \hat{=}$  Jacobi-Matrix, abhängig von  $(t, \mathbf{y})$ ,  
 $\mathbf{y}(t) \hat{=}$  Lösung des AWP  $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}), \mathbf{y}(t_0) = \mathbf{y}_0$

➤ Um die Variationsgleichung zu lösen muss auch das zugehörige Anfangswertproblem gelöst werden!

$$\mathbf{y}_0 \leftarrow \mathbf{y}_0 + \delta \mathbf{y}_0 \quad \Rightarrow \quad \delta \mathbf{y}(t) \approx \mathbf{W}(t; t_0, \mathbf{y}_0) \delta \mathbf{y}_0 \quad \text{für „kleine } \delta \mathbf{y}_0 \text{“}$$



Intervallweise asymptotische Kondition des AWP (1.1.13) auf  $[t_0, T]$  (bzgl. Norm  $\|\cdot\|$  auf  $\mathbb{R}^d$ ):

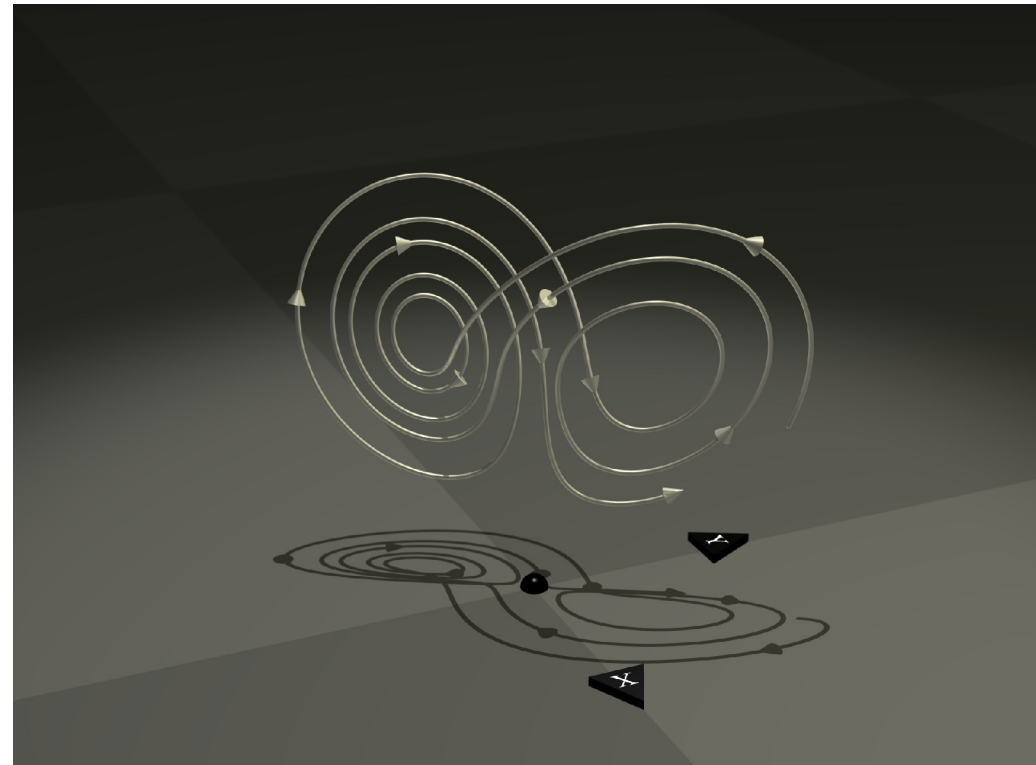
$$\kappa_{\text{abs}}^{\infty} := \max\{\|\mathbf{W}(t; t_0, \mathbf{y}_0)\| : t_0 \leq t \leq T\} . \quad (1.3.35)$$

### 1.3.3.5 Schlecht konditionierte AWPe

*Beispiel* 1.3.36 (Lorenz-System).  $\rightarrow$  [27, 21]

Autonome Differentialgleichung,  $D = \mathbb{R}^3$ ,  
 $\sigma, \rho, \beta \in \mathbb{R}^+$ :

$$\begin{aligned} \dot{x} &= \sigma(y - x) , \\ \dot{y} &= x(\rho - z) - y , \\ \dot{z} &= xy - \beta z . \end{aligned} \quad (1.3.37)$$



## Listing 1.3: Numerische Integration des Lorenz-Systems

```
1 function lorenzplot(rho,sigma,beta)
2
3 % MATLAB script for plotting 3D trajectories of the Lorenz system for
4 Ex. 1.3.36
5
6 % Arguments: parameters of the Lorenz system (1.3.37):  $\rho$ ,  $\sigma$ ,  $\beta$ .
7
8
9 % Default paramters
10 if (nargin < 3), rho=28; sigma = 10; beta = 8/3; end
11
12 % function handle for right hand side of the Lorez system (1.3.37)
13 f = @(t,y) ([sigma*(y(2)-y(1));rho*y(1) - y(2) -
14             y(1)*y(3);y(1)*y(2) - beta*y(3)]);
15
16
17 y0 = [8 9 9.5]; ystart = y0; % initial conditions
18 ts = [0 20]; % Time for simulation
19
20 % Numerical integration of Lorenz system using MATLAB standard integrator,
21 % see Rem. 1.2.4
22
23 opts = odeset('reltol',1E-10,'abstol',1E-10,'stats','on');
24 [t,y] = ode45(f,ts,y0,opts);
25
26 y0(3) = y0(3) + 1.0E-5; % Slight perturbation of initial value
27 [tt,yt] = ode45(f,ts,y0,opts);
```

```
22 % 3D plot of trajectories
23 figure('name','Lorenz'); hold on;
24 plot3(y(:,1),y(:,2),y(:,3),'r-');
25 plot3(yt(:,1),yt(:,2),yt(:,3),'b-');
26 xlabel('\bf x','fontsize',14);
27 ylabel('\bf y','fontsize',14);
28 zlabel('\bf z','fontsize',14);
29 title(sprintf('\sigma = %d, \rho = %d, \beta =
   %d',sigma,rho,beta));
30 view(45,15); grid on;
31 print -depsc2 'lorenz.eps';
```

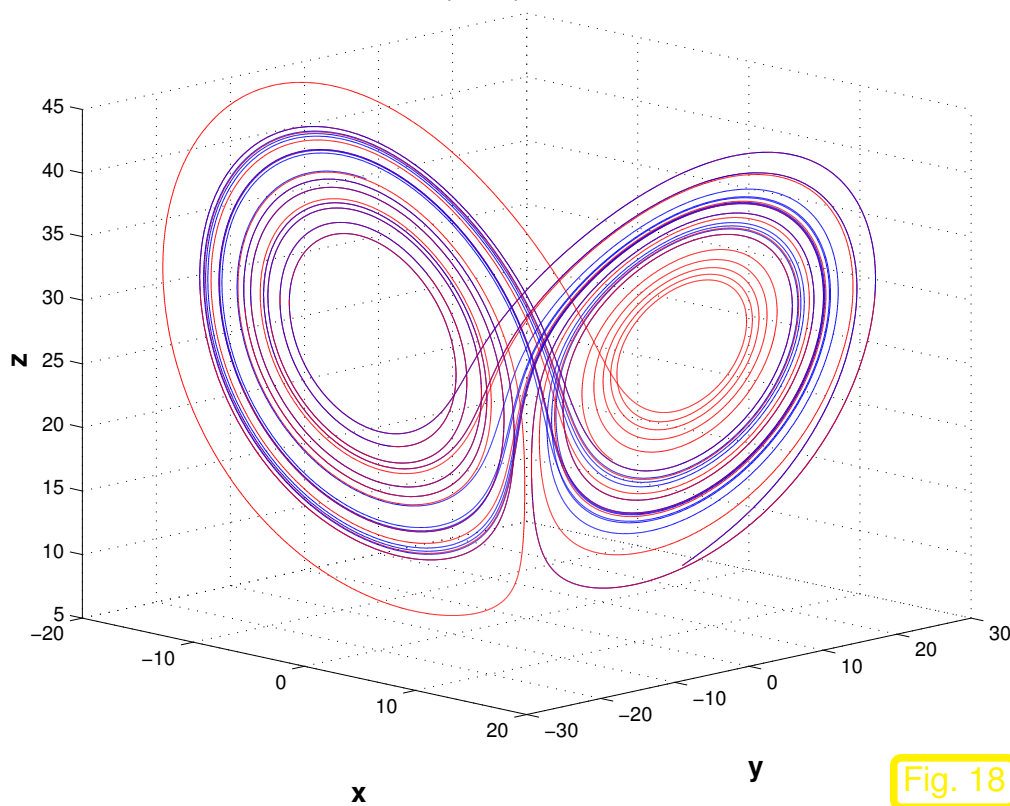
$\sigma = 10, \rho = 28, \beta = 2.666667e+00$ 

Fig. 18

— : Anfangswert  $y_0 := (8, 9, 9.5)^T$   
 — : Anfangswert  $y_0 := (8, 9, 9.5 + 10^{-5})^T$

Aus der Theorie der  
**reellen dynamischen Systeme** [21]:

Lorenz-System ist **chaotisches System**

Anfängliche exponentielle Divergenz der  
 beschränkten Trajektorien  
 (Schranke (1.3.31) gilt für kleine  $T$  !)

Winzige Störungen der Anfangswerte

➤ völlig verschiedene Zustände nach „expo-  
 nentiell kurzer Zeit“.



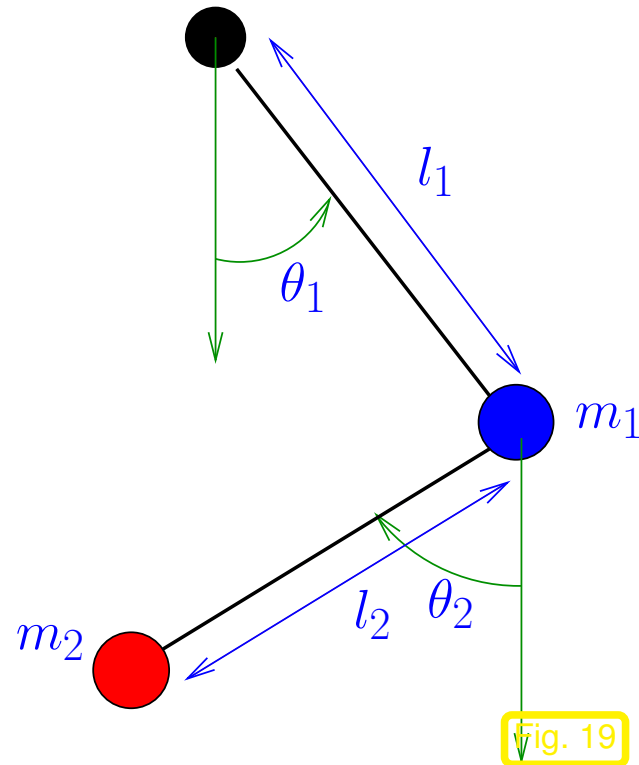
Ziel numerischer Simulation chaotischer dynamischer Systeme:

Identifikation des „**typischen**“ Verhaltens von Trajektorien

Essentiell:

Korrekte Behandlung von Erhaltungsgrößen, z.B. Gesamtenergie  
 (→ Erste Integrale, Def. 1.2.7)

## Beispiel 1.3.38 (Doppelpendel).



◁ (Mathematisches) Doppelpendel mit fester Aufhängung und masselosen Stäben.

Minimalkoordinaten: Auslenkungswinkel  $\theta_1, \theta_2$

Konfigurationsraum  $D = [0, 2\pi]^2$  (Torus)

Hamilton-Funktion ( $\rightarrow$  Def. 1.2.20) = Summe kinetischer und potentieller Energie:

$$H(p_1, p_2, \theta_1, \theta_2) = \frac{l_2^2 m_2 p_1^2 + l_1^2 (m_1 + m_2) p_2^2 - 2m_2 l_1 l_2 p_1 p_2 \cos(\theta_1 - \theta_2)}{2l_1^2 l_2^2 m_2 (m_1 + \sin^2(\theta_1 - \theta_2) m_2)} - m_2 g l_2 \cos \theta_2 - (m_1 + m_2) g l_1 \cos \theta_1 .$$

R. Hiptmair  
rev 35327,  
25. April  
2011

Beobachtung (in Experiment und Simulation):

Extrem sensitive Abhängigkeit der Pendelbewegung von Anfangsbedingungen

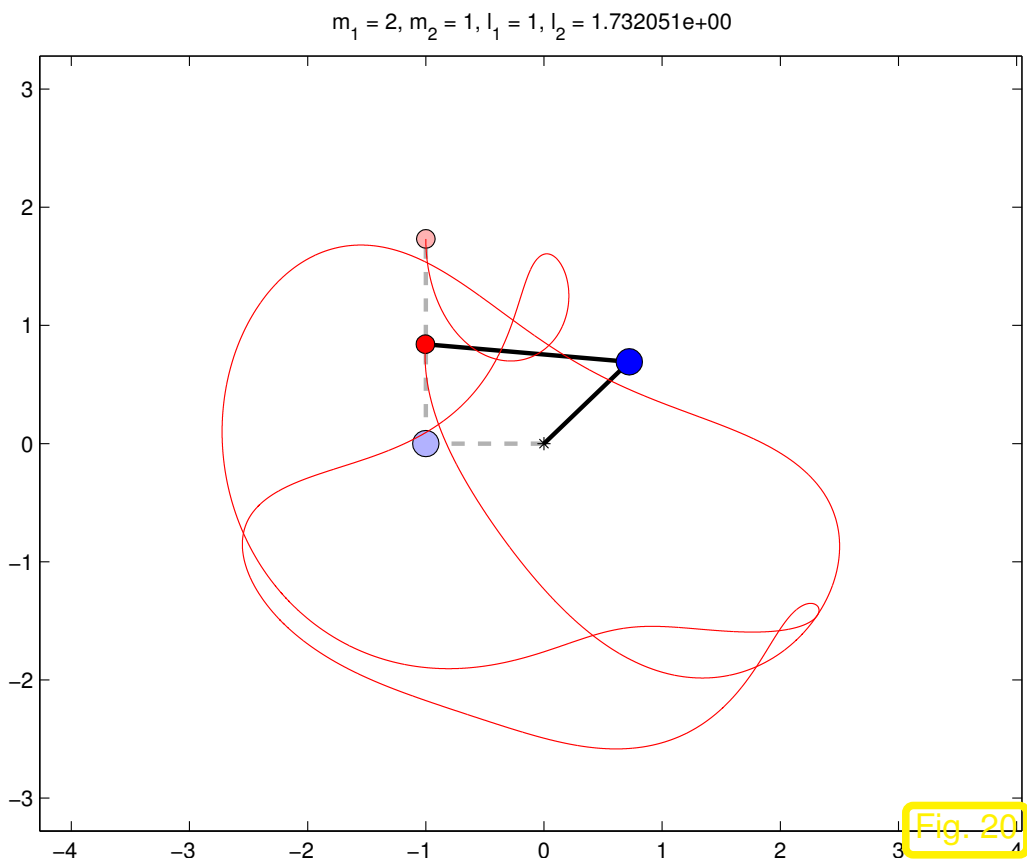


Fig. 20

$$\theta_1(0) = -\frac{\pi}{2}, \theta_2(0) = \pi, p_1(0) = p_2(0) = 0$$

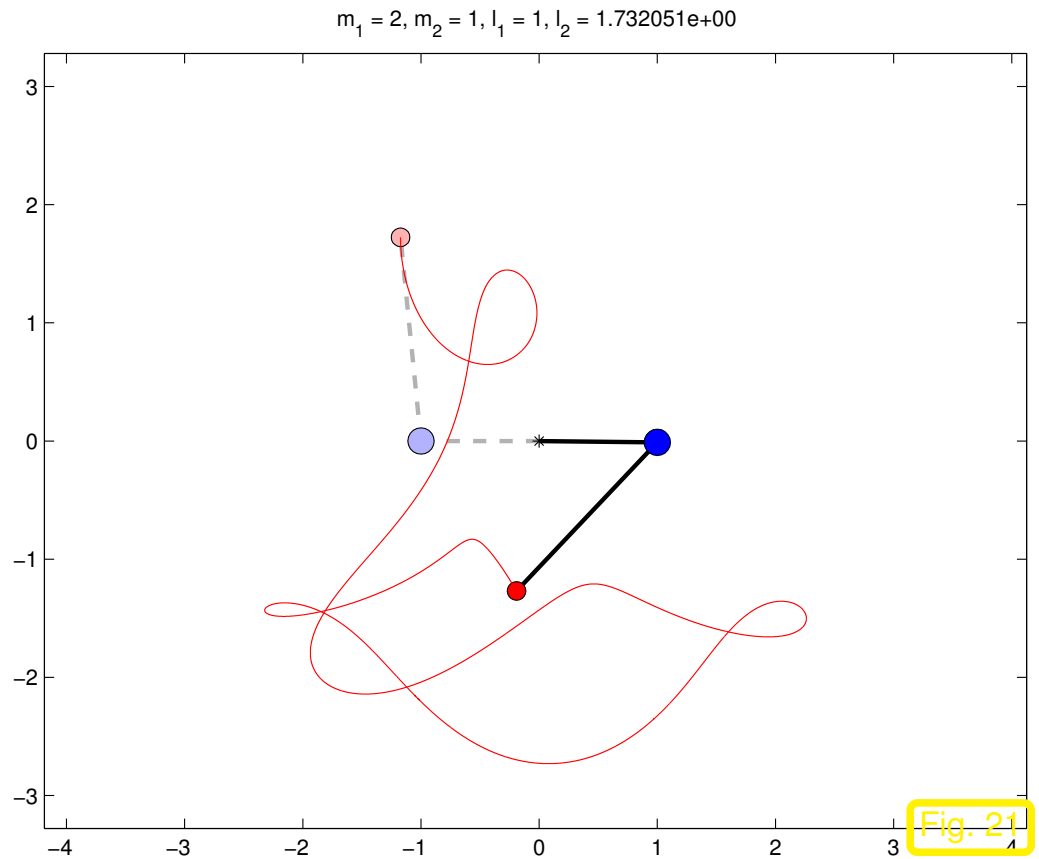


Fig. 21

$$\theta_1(0) = -\frac{\pi}{2}, \theta_2(0) = \pi + 10^{-2},$$

$$p_1(0) = p_2(0) = 0$$

[Simulation, MATLAB, ode45, Zeit [0, 20], Schrittweite  $10^{-3}$ ]



- Gegeben:
- Rechte Seite  $\mathbf{f} : \Omega \mapsto \mathbb{R}^d$ , lokal Lipschitz-stetig ( $\rightarrow$  Def. 1.3.2) auf erweitertem Zustandsraum  $\Omega := I \times D \subset \mathbb{R}^{d+1}$
  - Anfangsbedingungen  $\mathbf{y}_0 \in D$  zum Anfangszeitpunkt  $t_0$

Thm. 1.3.4 (Peano & Picard-Lindelöf)  $\blacktriangleright$  Existenz & Eindeutigkeit von Lösungen ( $\rightarrow$  Def. 1.1.14) des AWP

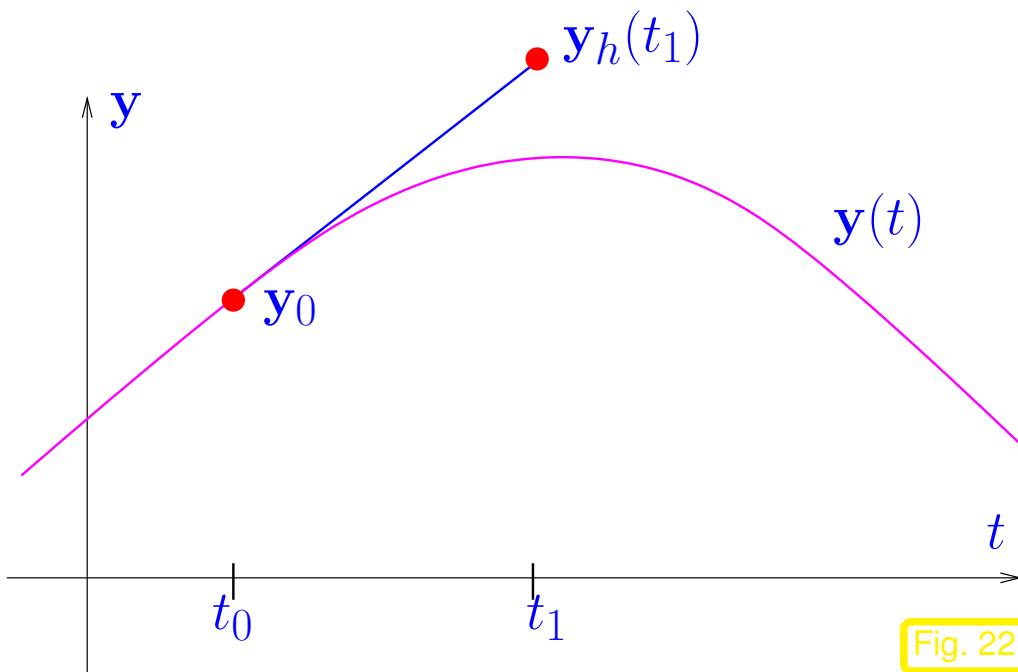
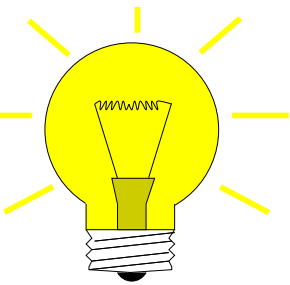
$$\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}) \quad , \quad \mathbf{y}(t_0) = \mathbf{y}_0 . \quad (1.4.1)$$

- Ziel:
- ☞ Approximation von  $\mathbf{y}(T)$  für Endzeitpunkt  $T \in J(t_0, \mathbf{y}_0) \triangleright \mathbf{y}_h(T)$ .
  - ☞ Approximation der Funktion  $t \mapsto \mathbf{y}(t)$ ,  $t \in [t_0, T]$ ,  $T \in J(t_0, \mathbf{y}_0) \triangleright t \mapsto \mathbf{y}_h(t)$ .



# 1.4.1 Das explizite Euler-Verfahren (Euler 1768)

- Idee: ❶ „Vorantasten“ in der Zeit: (1.4.1) = Komposition von AWPe zu *gegebenen kleinen* Zeitintervallen  $[t_{k-1}, t_k]$ ,  $k = 1, \dots, N$ ,  $t_N := T$
- ❷ Approximation der zeitlokalen Lösungskurven durch **Tangente** im aktuellen Anfangszeitpunkt.



Explicites Euler-Verfahren  
(Eulersches Polygonzugverfahren)

◁ Erster Schritt des expliziten Euler-Verfahrens ( $d = 1$ ):

Steigung der Tangente =  $f(t_0, \mathbf{y}_0)$

$\mathbf{y}_h(t_1)$  ist Startwert für nächsten Schritt !

In Formeln: durch explizites Eulerverfahren erzeugte Näherungen für  $\mathbf{y}(t_k)$  erfüllen die Rekursion

$$\mathbf{y}_{k+1} := \mathbf{y}_h(t_{k+1}) = \mathbf{y}_h(t_k) + h_k \mathbf{f}(t_k, \mathbf{y}_h(t_k)), \quad k = 0, \dots, N-1, \quad (1.4.2)$$

mit lokaler (Zeit)schrittweite  $h_k := t_{k+1} - t_k$ .

Alternative Notation:

$$\mathbf{y}_k := \mathbf{y}_h(t_k)$$

Veranschaulichung: Explizites Eulerverfahren

Anfangswertproblem

Riccati-Differentialgleichung, siehe Bsp. 1.1.3

$$\dot{y} = y^2 + t^2. \quad (1.1.4)$$

$$y_0 = \frac{1}{2}, t_0 = 0, T = 1,$$

gleichgrosse Zeitschritte  $h = 0.2$

$\rightarrow \hat{=}$  Richtungsfeld der Riccati-Dgl.

für

Bsp. 1.1.3

▷

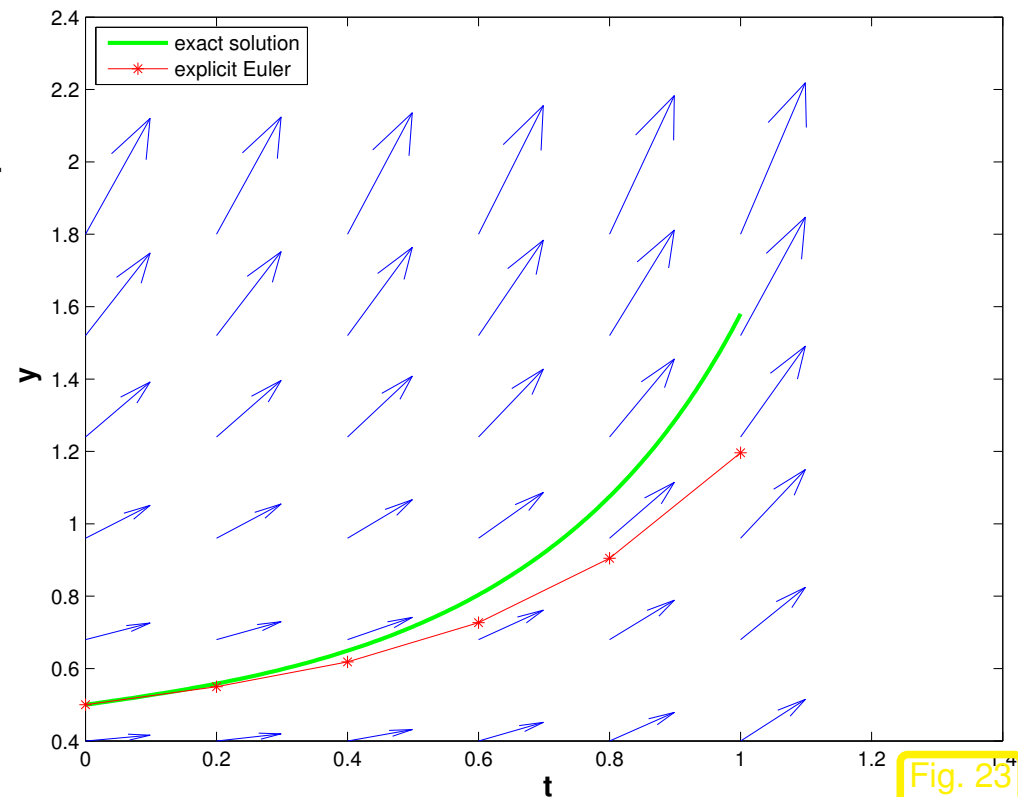


Fig. 23

**Bemerkung 1.4.3** (Explizites Eulerverfahren als Differenzenverfahren).

(1.4.2) aus Approximation von Ableitung  $\frac{d}{dt}$  durch **Vorwärtsdifferenzenquotienten** auf **Zeitgitter**  $\mathcal{G} := \{t_0, t_1, \dots, t_N\}$ :

$$\dot{\mathbf{y}} = f(t, \mathbf{y}) \quad \longleftrightarrow \quad \frac{\mathbf{y}_h(t_{k+1}) - \mathbf{y}_h(t_k)}{h_k} = f(t_k, \mathbf{y}_h(t_k)), \quad k = 0, \dots, N-1.$$

△

Frage: Wie genau ist die Näherungslösung ?

**Beispiel 1.4.4** (**Konvergenz**(-geschwindigkeit) des expliziten Euler-Verfahrens).

#### Listing 1.4: Erzeugen von Fehlerkurven für explizites Eulerverfahren

```

1 function err = eulerConvergence(odefun, tspan, y0v, N)
2 % MATLAB function computing the error (at final time) of the explicit Euler
  % method (1.4.2)
3 % Arguments:
4 % odefun = @(t,y): handle to function returning a vector

```

```
5 % tspan = [t0 T]: initial and final time
6 % y0v  $\hat{=}$  array of initial values
7 % N  $\hat{=}$  vector containing numbers of steps. For each the error is returned
8
9 err = []; l = 1; % Initialize error array
10
11 for y0 = y0v
12 % Compute 'exact' solution
13 [t,y] =
14     ode45(odefun,tspan,y0,odeset('reltol',1E-11,'abstol',1E-11));
15 % Compute Euler solutions
16 erri = [];
17 for n=N
18     h = (tspan(2)-tspan(1))/n; % uniform timestep size
19     t_eul = tspan(1); % initial time
20     y_eul = y0; % initialize iteration
21     for k=1:n
22         y_eul = y_eul + h*odefun(t_eul,y_eul); % see (1.4.2)
23         t_eul = t_eul + h; % increment time
24     end
25     erri = [erri, norm(y(end,:) - y_eul)]; % record error
26 end
```

```
27 err = [err;erri]; % assemble matrix of error values
28 leg{1} = sprintf ('y0 = %3.2f',y0);
29 l = l+1;
30 end
31
32 % Plot error curves in log-log scale to discern algebraic convergence →
   Def. 1.4.5
33 figure ('name', 'erreul');
34 ts = (tspan(2)-tspan(1))./N;
35 loglog (ts,err, '-+'); hold on;
36 loglog (ts,10*ts, 'k-');
37 xlabel ('{\bf timestep h}', 'fontsize', 14);
38 ylabel ('{\bf error (Euclidean norm)}', 'fontsize', 14);
39 leg{1} = 'O(h)'; legend (leg, 'location', 'southeast');
```

- AWP für Riccati-Dgl. (1.1.4) auf  $[0, 1]$
- Explizites Euler-Verfahren (1.4.2) mit uniformem Zeitschritt  $h = 1/n$ ,  
 $n \in \{5, 10, 20, 40, 80, 160, 320, 640\}$ .
- Fehler  $\text{err}_h := |y(1) - y_h(1)|$

Beobachtung:

Algebraische Konvergenz  $\text{err}_h = O(h)$

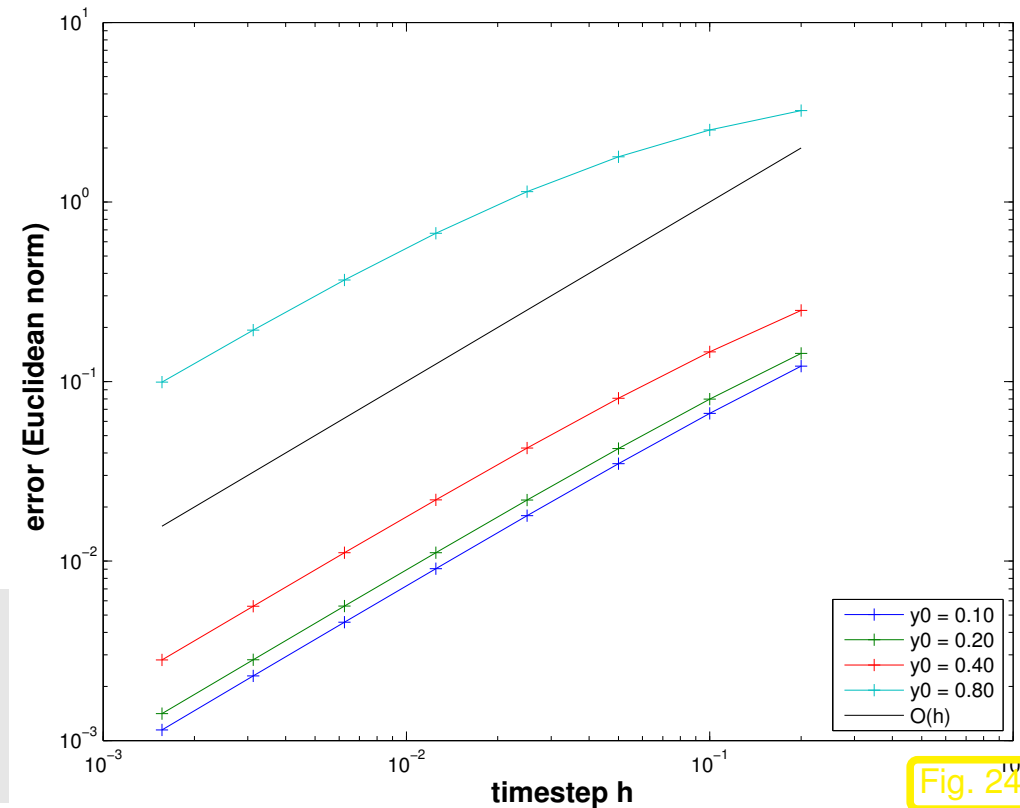


Fig. 24

R. Hiptmair

rev 35327,  
25. April  
2011

Notation “Landau- $O$ ”:

$$e(h) = O(g(h)) \quad \text{für } h \rightarrow 0 \quad \Leftrightarrow \quad \exists h_0 > 0, C > 0: |e(h)| \leq Cg(h) \quad \forall 0 \leq h \leq h_0.$$

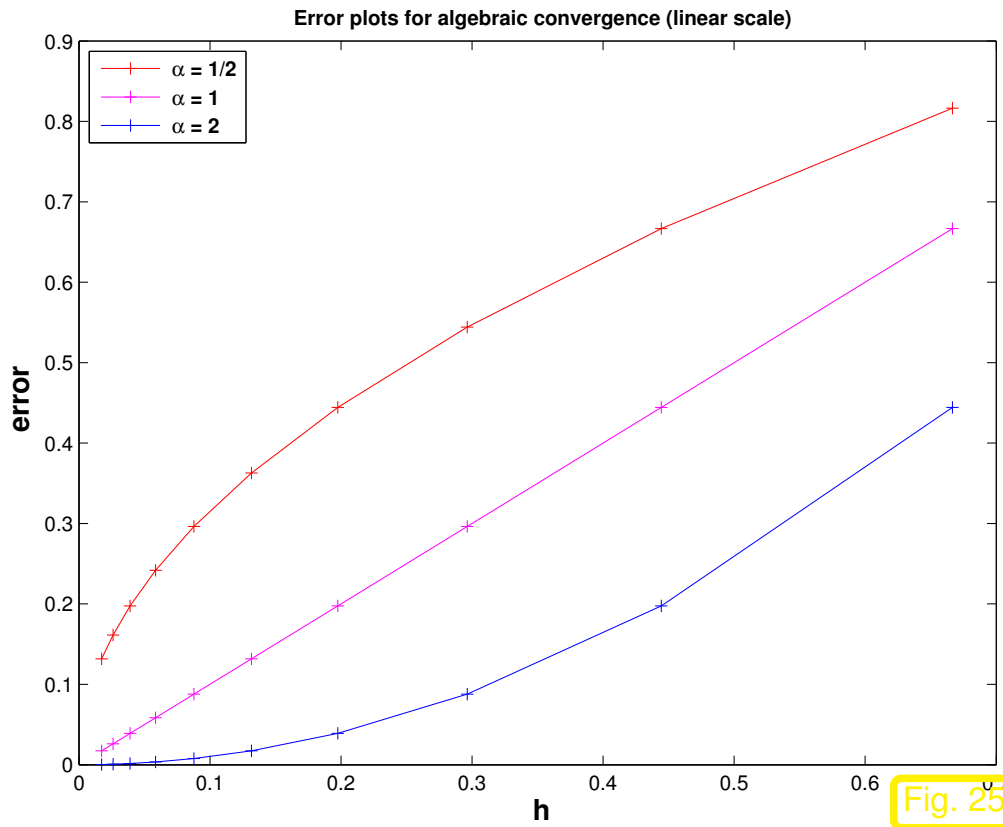
**Definition 1.4.5** (Arten der Konvergenz).

Sei  $\text{err}_h$  der *Diskretisierungsfehler* eines Verfahrens zum Diskretisierungsparameter/Schrittweite  $h$ ,  $h > 0$ .

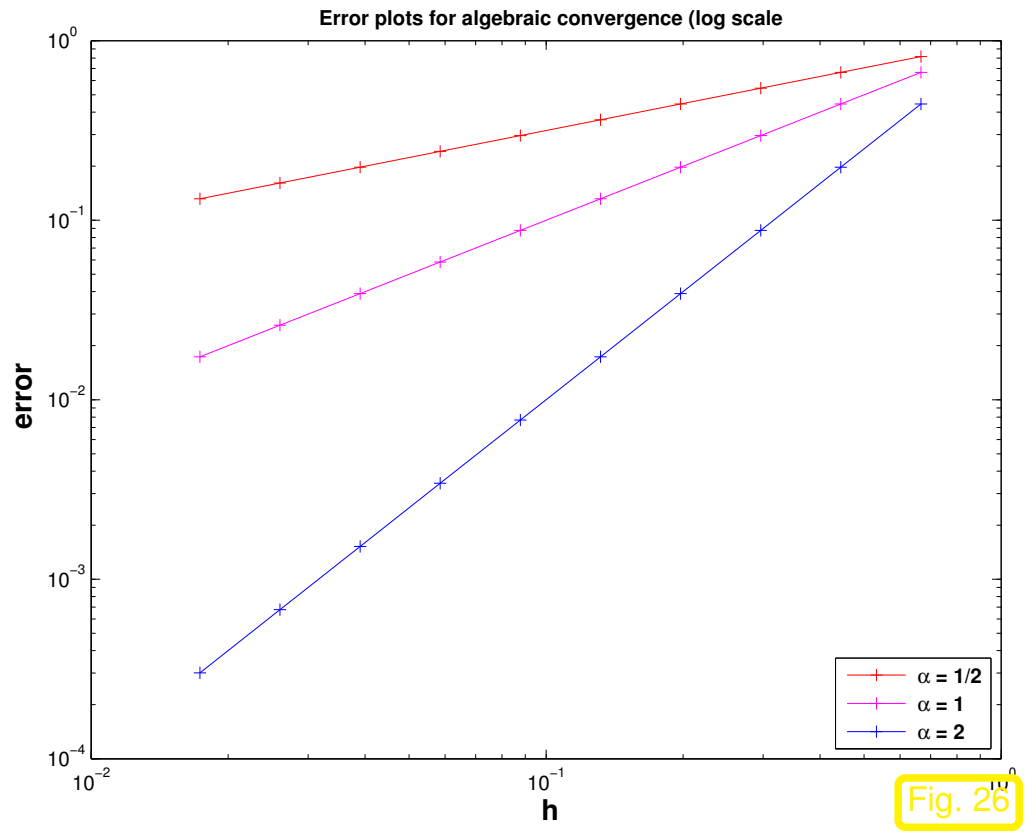
$$\text{err}_h = O(h^\alpha) \quad :\Leftrightarrow \text{ *Algebraische Konvergenz* der Ordnung } \alpha > 0$$

$$\text{err}_h = O(\exp(-\beta h^{-\gamma})), \quad :\Leftrightarrow \text{ *exponentielle Konvergenz*, falls } \beta, \gamma > 0$$

Fehlerplots bei algebraischer Konvergenz ( $h_i = (3/2)^{-i}$ ,  $i = 1, \dots, 10$ )



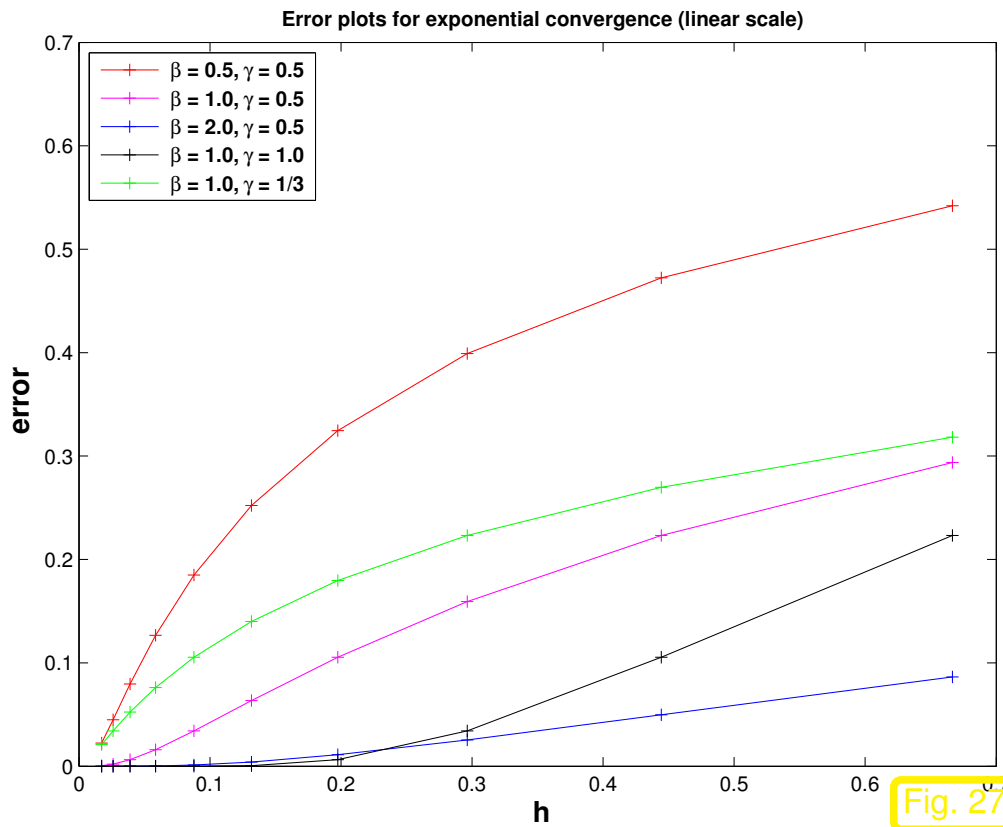
lineare Skalen



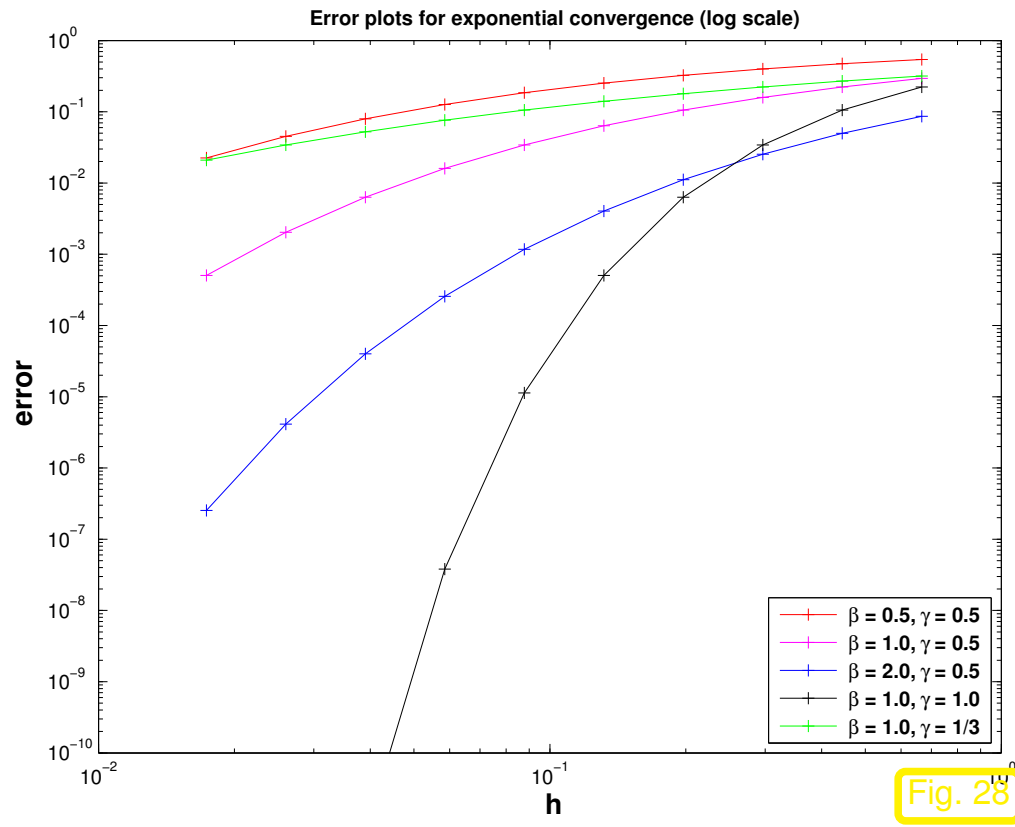
logarithmische Skalen

Fehlerplots bei exponentieller Konvergenz ( $h_i = (3/2)^{-i}, i = 1, \dots, 10$ )

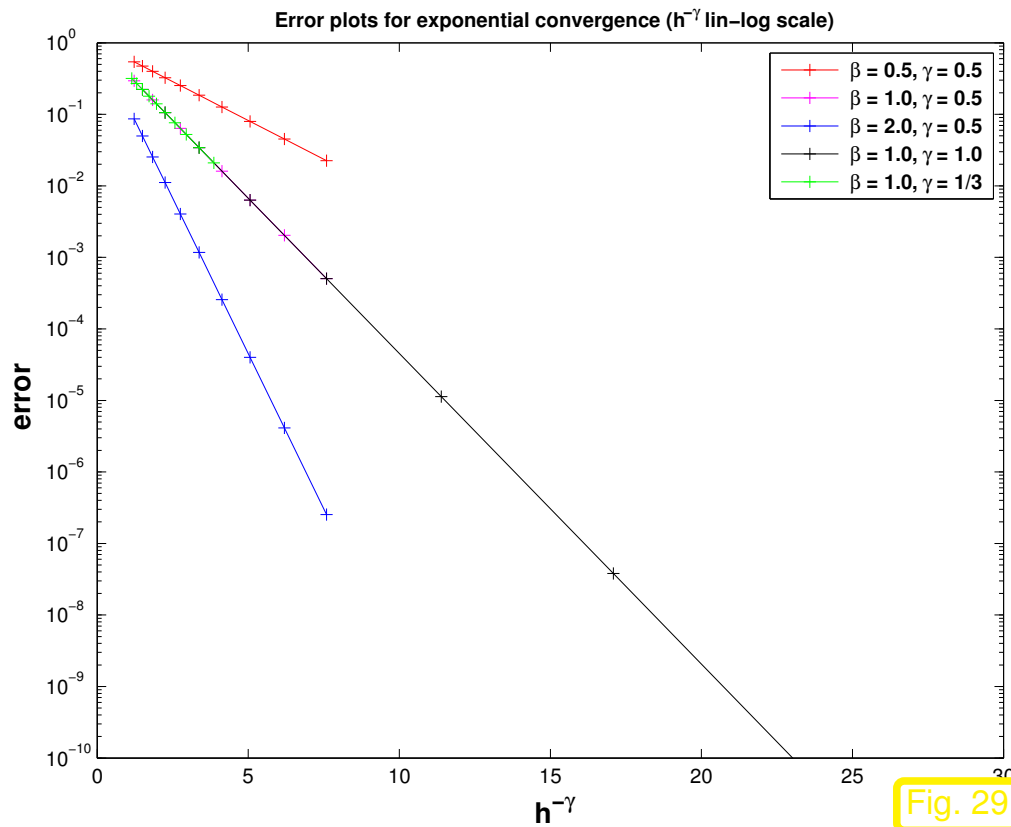




lineare Skalen



logarithmische Skalen



◁  $(h_i^{-\gamma}, \epsilon_i)$  ( $h_i \hat{=}$  Schrittweiten,  $\epsilon_i \hat{=}$  zugehörige Diskretisierungsfehler) liegen auf Geraden mit Steigung  $-\beta$

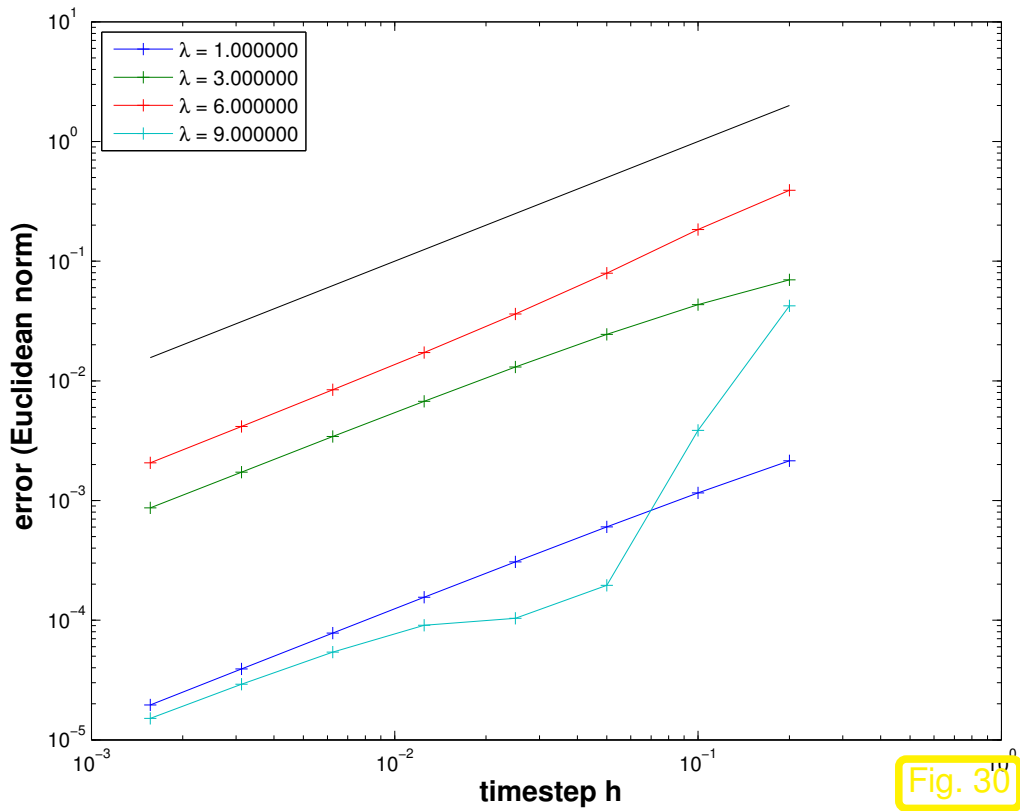
### Beispiel 1.4.9 (Explizites Euler-Verfahren für logistische Dgl.)

- Anfangswertproblem für logistische Differentialgleichung, siehe Bsp. 1.2.1

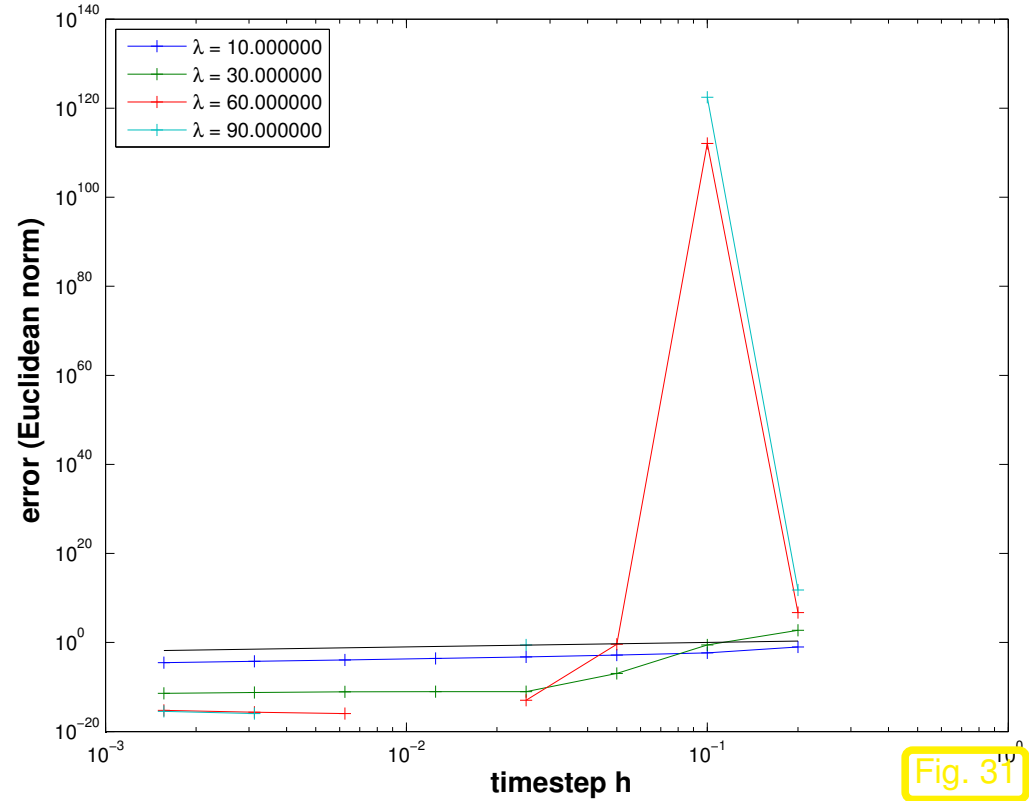
$$\dot{y} = \lambda y(1 - y) \quad , \quad y(0) = 0.01 \quad .$$

- Explizites Euler-Verfahren (1.4.2) mit uniformem Zeitschritt  $h = 1/n$ ,  $n \in \{5, 10, 20, 40, 80, 160, 320, 640\}$ .

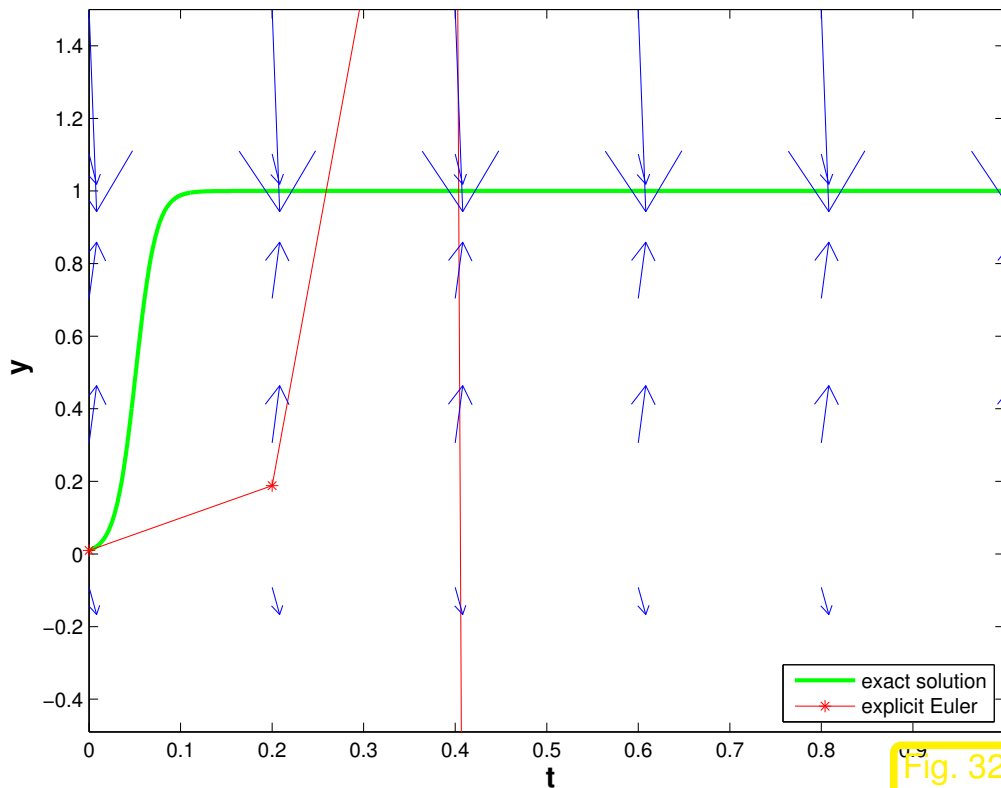
- Fehler zum Endzeitpunkt  $T = 1$



$\lambda$  klein:  $O(h)$ -Konvergenz (asymptotisch)



$\lambda$  gross: Explosion von  $y_k$  für grosse Zeitschrittweiten  $h$



◁  $\lambda = 90$ , —  $\hat{=}$  exakte Lösung, —  $\hat{=}$  Eulerpolygon

$y_k$  schießen über den stark attraktiven Fixpunkt  $y = 1$  hinaus.

➔ Beobachtung: exponentiell anwachsende Oszillationen der  $y_k$



Einsicht durch Modellproblemanalyse: einfachste Dgl. mit stark attraktivem Fixpunkt  $y = 0$

$$\text{Homogene lineare skalare Dgl., Sect. 1.3.2 : } \dot{y} = f(y) := \lambda y, \quad \lambda < 0. \quad (1.4.10)$$

$$\dot{y} = \lambda y, \quad y(0) = y_0 \Rightarrow y(t) = y_0 \exp(\lambda t) \rightarrow 0 \quad \text{für } t \rightarrow \infty. \quad (1.4.11)$$

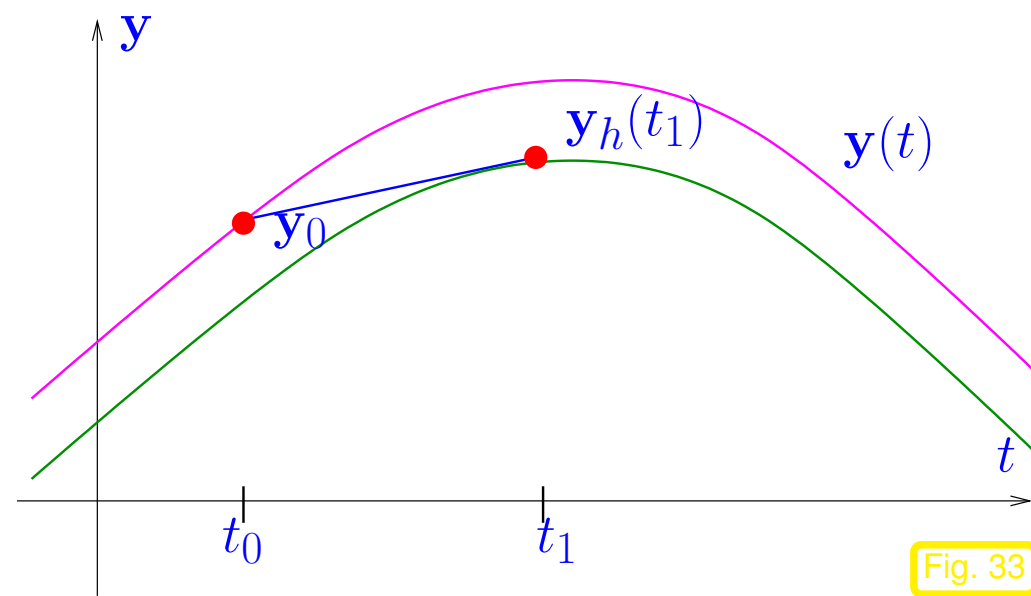
Rekursion des expliziten Eulerverfahrens für (1.4.10) (uniforme Zeitschrittweite  $h > 0$ )

$$(1.4.2) \text{ for } f(y) = \lambda y: \quad y_{k+1} = y_k(1 + \lambda h) . \quad (1.4.12)$$

$$\blacktriangleright \quad y_k = y_0(1 + \lambda h)^k \Rightarrow |y_k| \rightarrow \begin{cases} 0 & , \text{ wenn } \lambda h > -2 \quad (\text{qualitativ richtig}) , \\ \infty & , \text{ wenn } \lambda h < -2 \quad (\text{qualitativ falsch}) . \end{cases}$$

## 1.4.2 Das implizite Euler-Verfahren

Wie vermeidet man das Überschiessen des expliziten Eulerverfahrens bei stark attraktiven Fixpunkten und grossen Zeitschrittweiten ?



Idee: Approximiere Lösung durch  $(t_0, y_0)$  auf  $[t_0, t_1]$  durch

- Strecke durch  $(t_0, y_0)$
- mit Steigung  $f(t_1, y_1)$

◁ —  $\hat{=}$  Lösungskurve durch  $(t_0, y_0)$ ,  
 —  $\hat{=}$  Lösungskurve durch  $(t_1, y_1)$ ,  
 —  $\hat{=}$  Tangente an — in  $(t_1, y_1)$ .

Anwendung auf kleine Zeitintervalle  $[t_0, t_1], [t_1, t_2], \dots, [t_{N-1}, t_N]$  ➤ **implizites Euler-Verfahren**

▶ durch implizites Eulerverfahren erzeugte Näherung für  $\mathbf{y}(t_k)$  erfüllt

$$\mathbf{y}_{k+1} := \mathbf{y}_h(t_{k+1}) = \mathbf{y}_h(t_k) + h_k \mathbf{f}(t_{k+1}, \mathbf{y}_{k+1}), \quad k = 0, \dots, N-1, \quad (1.4.13)$$

mit lokaler **(Zeit)schrittweite**  $h_k := t_{k+1} - t_k$ .

Beachte: (1.4.13) erfordert Auflösen einer (evtl. nichtlinearen) Gleichung nach  $\mathbf{y}_{k+1}$  !

(▶ Terminologie „implizit“)

*Bemerkung* 1.4.14 (Implizites Eulerverfahren als Differenzenverfahren).

(1.4.13) aus Approximation der Zeitableitung  $\frac{d}{dt}$  durch Rückwärtsdifferenzenquotienten auf Zeitgitter  $\mathcal{G} := \{t_0, t_1, \dots, t_N\}$ :

$$\dot{\mathbf{y}} = f(t, \mathbf{y}) \quad \longleftrightarrow \quad \frac{\mathbf{y}_h(t_{k+1}) - \mathbf{y}_h(t_k)}{h_k} = f(t_{k+1}, \mathbf{y}_h(t_{k+1})), \quad k = 0, \dots, N-1.$$



*Beispiel* 1.4.15 (Implizites Eulerverfahren für logistische Differentialgleichung).  $\rightarrow$  Bps. 1.4.9

Wiederholung der numerischen Experimente aus Beispiel 1.4.9 für implizites Eulerverfahren (1.4.13):

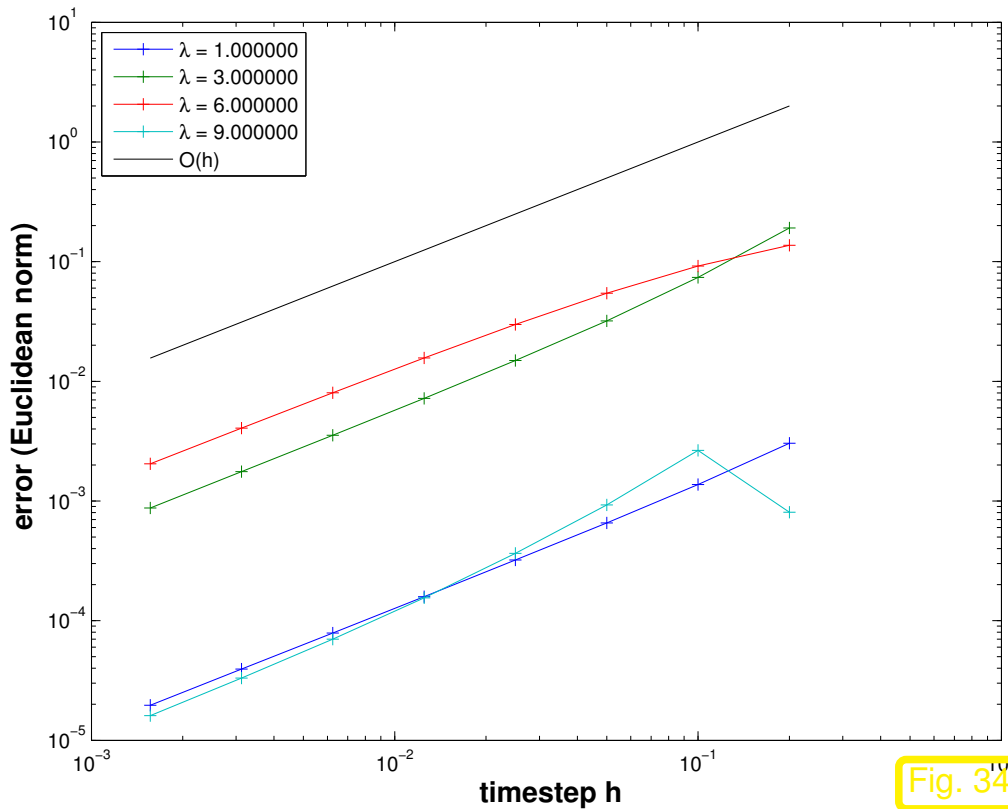


Fig. 34

$\lambda$  klein:  $O(h)$ -Konvergenz (asymptotisch)

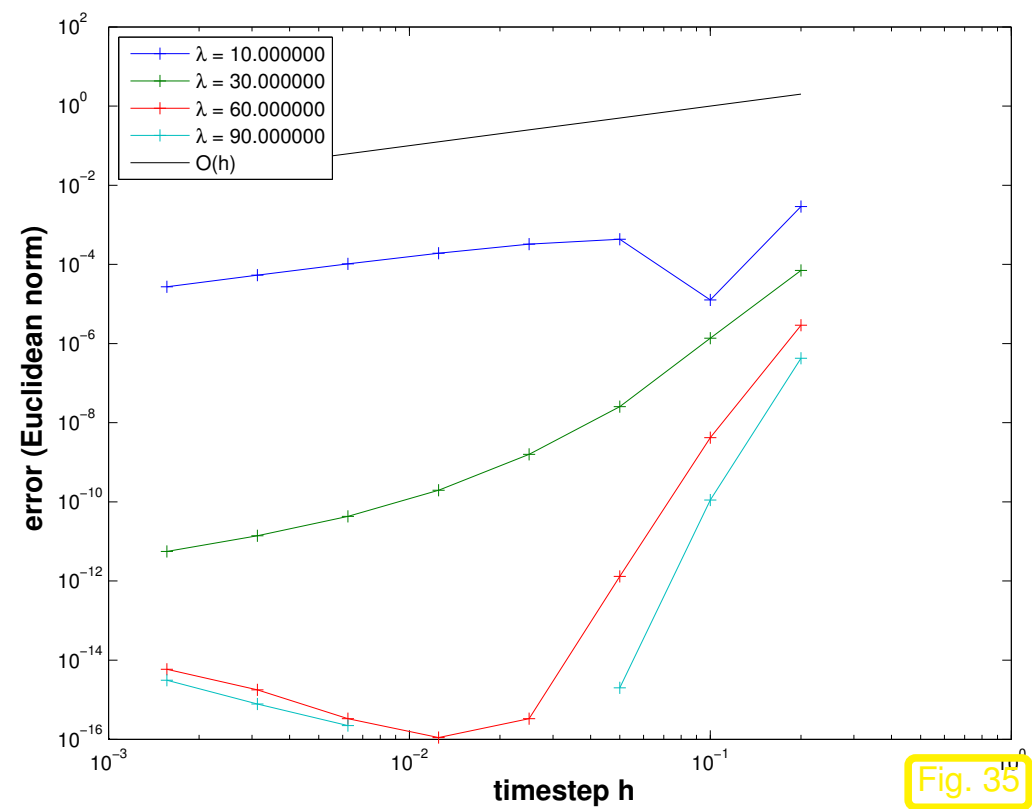


Fig. 35

$\lambda$  gross: stabil für alle Zeitschrittweiten  $h$  !

Modellproblemanalyse (wie in Abschnitt 1.4.1):

$$(1.4.13) \text{ for } f(y) = \lambda y: \quad y_{k+1} = y_k \frac{1}{1 - \lambda h}. \quad (1.4.16)$$



$$\blacktriangleright \quad y_k = y_0 \left( \frac{1}{1 - \lambda h} \right)^k \Rightarrow |y_k| \rightarrow \begin{cases} 0 & , \text{wenn } \lambda h < 0 \quad (\text{qualitativ richtig}) , \\ \infty & , \text{wenn } 0 < \lambda h < 1 \quad (\text{qualitativ richtig}) , \\ \infty & , \text{wenn } \lambda h > 1 \quad (\text{Oszillationen, qualitativ falsch}) . \end{cases}$$

*Beispiel 1.4.17* (Euler-Verfahren für Pendelgleichung).

Mathematisches Pendel  $\rightarrow$  Bsp. 1.2.17: Hamiltonsche Form (1.2.19) der Bewegungsgleichungen

$$\text{Winkelgeschwindigkeit } p := \dot{\alpha} \Rightarrow \frac{d}{dt} \begin{pmatrix} \alpha \\ p \end{pmatrix} = \begin{pmatrix} p \\ -\frac{g}{l} \sin \alpha \end{pmatrix}, \quad g = 9.8, l = 1. \quad (1.2.19)$$

- Approximative numerische Lösung mit explizitem/implizitem Eulerverfahren (1.4.2)/(1.4.13),
- Konstante Zeitschrittweite  $h = T/N$ ,  $T = 5$  Endzeitpunkt,  $N \in \{50, 100, 200\}$ ,
- Startwert:  $\alpha(0) = \pi/4$ ,  $p(0) = 0$ .

R. Hiptmair  
rev 35327,  
25. April  
2011

Listing 1.5: Simulation des mathematischen Pendels mit einfachen Polygonzugverfahren

```
1 function pendeul(y0,T,N,filename)
2 % MATLAB function applying explicit and implicit Euler methods and implicit
```

```
3 % midpoint rule of Sect. 1.4.3 to mathematical pendulum equation in
4 % minimal coordinates and Hamiltonian form for Ex. 1.4.17
5 % Arguments: y0  $\hat{=}$  initial position, T  $\hat{=}$  final time N  $\hat{=}$ 
6 % number of equidistant timesteps
7
8 g = 9.8; % constant of gravity
9 l = 1; % length of pendulum
10
11 % Compute 'exact' solution by means of high-order single step method with tight
12 % error control
13 odefun = @(t,y) [y(2); -g/l*sin(y(1))];
14 [t,s] =
15     ode45(odefun, [0,T], y0, odeset('abstol', 1E-10, 'reltol', 1E-10));
16 h = T/N; % timestep
17
18 % Explicit Euler (1.4.2)
19 y_expl = y0; y = y0;
20 for k=1:N
21     y = y + h*[y(2); -g/l*sin(y(1))];
22     y_expl = [y_expl, y];
23 end
24
```

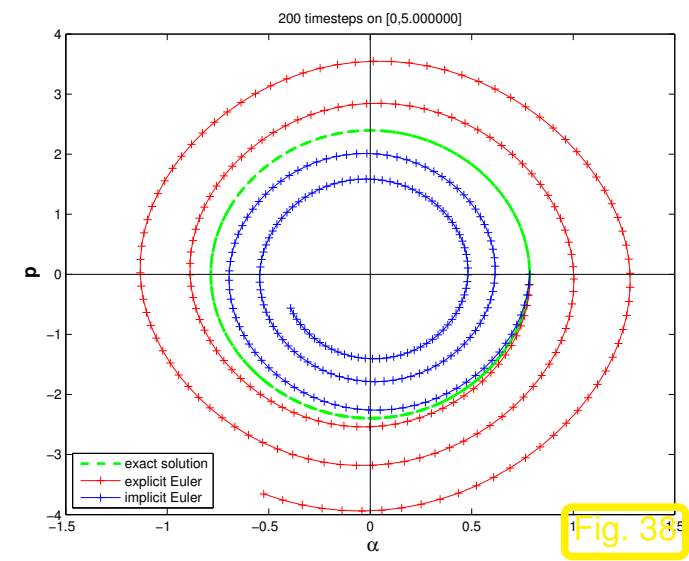
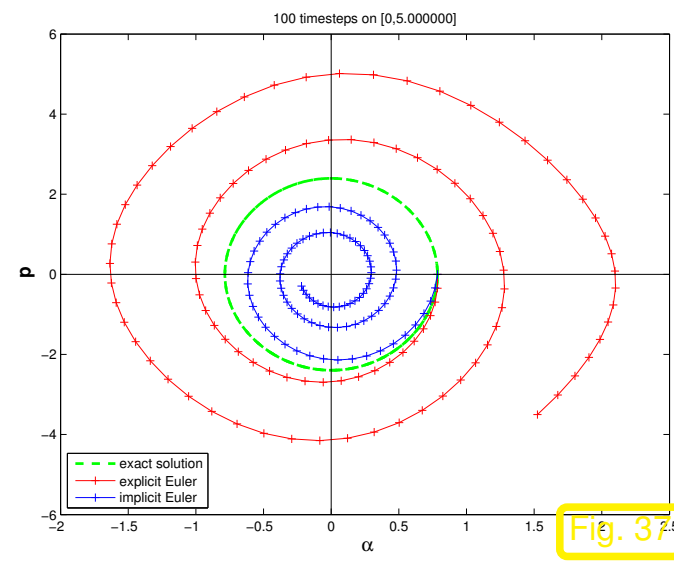
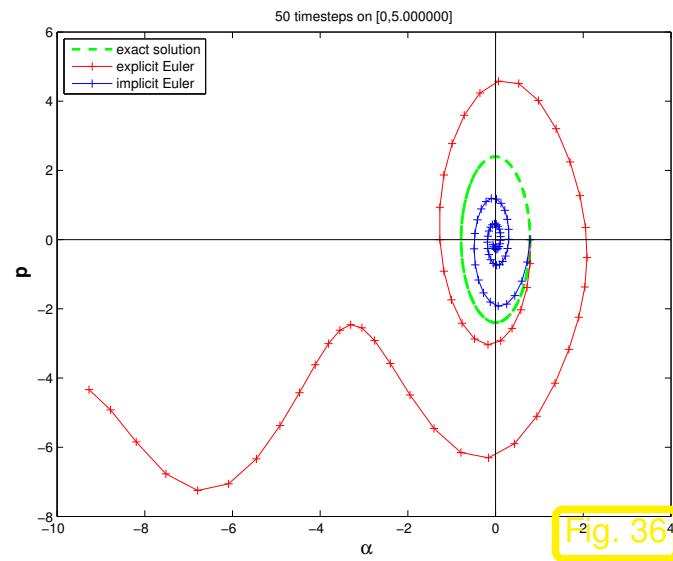
```
25 % Implicit Euler
26 y_imp = y0; y = y0;
27 for k=1:N
28     % Implicit Euler equation for next angle
29     F = @(x) x+h*h*g/l*sin(x) - y(1) - h*y(2);
30     [y(1),Fval] = fsolve(F,y(1)+h*y(2)); % solve non-linear system of
        equations
31     fprintf('Impl Euler step %d: residual %f\n',k,Fval);
32     y(2) = y(2) - h*g/l*sin(y(1));
33     y_imp = [y_imp,y];
34 end
35
36 % Implicit midpoint rule
37 y_mid = y0; y = y0;
38 rhs = @(y) [y(2);-g/l*sin(y(1))];
39
40 for k=1:N
41     % Implicit equation (1.4.19) for implicit midpoint rule
42     F = @(x) (x - h*rhs(y+0.5*x));
43     [dy,Fval] = fsolve(F,h*rhs(y)); y = y+dy;
44     fprintf('Impl midp step %d: residual %f\n',k,norm(Fval));
45     y_mid = [y_mid,y];
46 end
47
```

```
48 tg = h*(0:N);
49
50 % Plotting of trajectories in phase space
51 figure('name','pendeul');
52 ph = plot(s(:,1),s(:,2),'g--',...
53          y_expl(1,:),y_expl(2,),'r-+',...
54          y_imp(1,:),y_imp(2,),'b-+',...
55          y_mid(1,:),y_mid(2,),'m-*'); hold on;
56 set(ph(1),'linewidth',2);
57 ax = axis;
58 plot([ax(1) ax(2)], [0 0], 'k-');
59 plot([0 0], [ax(3) ax(4)], 'k-');
60 xlabel('\bf \alpha','fontsize',14);
61 ylabel('\bf p','fontsize',14);
62 legend('exact solution','explicit Euler','implicit Euler',...
63        'implicit midpoint','location','southwest');
64 title(sprintf('%d timesteps on [0,%f]',N,T));
65
66 if (nargin > 3)
67     print('-depsc2', sprintf('%s.eps',filename));
68 end
69
70 % Tracking energies for explicit Euler
```

```
71 y = y_expl;
72
73 E_kin = 0.5*(y(2,:).^2);
74 E_pot = -g/l*cos(y(1,:));
75 E_pot = E_pot - min(E_pot) + min(E_kin);
76 E_tot = E_kin + E_pot;
77
78 % Plot of evolution of energies
79 figure('name','Pendulum: energy');
80 plot(tg,E_kin,'b-',...
81      tg,E_pot,'c-',...
82      tg,E_tot,'r-');
83 xlabel('\bf time t','fontsize',14);
84 ylabel('\bf energy','fontsize',14);
85 legend('kinetic energy','potential energy','total energy');
86 title('Energies for {\bf explicit} Euler discrete evolution');
87
88 if (nargin > 3),
89     print('-depsc2',sprintf('%s_EnExpl.eps',filename)); end
89
90 % Tracking energies for implicit Euler
91 y = y_imp;
92 E_kin = 0.5*(y(2,:).^2);
```

```
93 E_pot = -g/l*cos(y(1,:));
94 E_pot = E_pot - min(E_pot) + min(E_kin);
95 E_tot = E_kin + E_pot;
96
97 figure('name','Pendulum: energy');
98 plot(tg,E_kin,'b-',...
99      tg,E_pot,'c-',...
00      tg,E_tot,'r-');
01 xlabel('\bf time t','fontsize',14);
02 ylabel('\bf energy','fontsize',14);
03 legend('kinetic energy','potential energy','total energy');
04 title('Energies for {\bf implicit} Euler discrete evolution');
05
06 if (nargin > 3)
07     print('-depsc2',sprintf('%s_EnImpl.eps',filename));
08 end
09
10 % Tracking energies for implicit midpoint rule
11 y = y_mid;
12
13 E_kin = 0.5*(y(2,:).^2);
14 E_pot = -g/l*cos(y(1,:));
15 E_pot = E_pot - min(E_pot) + min(E_kin);
```

```
16 E_tot = E_kin + E_pot;
17
18 figure ('name', 'Pendulum: energy');
19 plot (tg, E_kin, 'b-', ...
20       tg, E_pot, 'c-', ...
21       tg, E_tot, 'r-');
22 xlabel ('\bf time t', 'fontsize', 14);
23 ylabel ('\bf energy', 'fontsize', 14);
24 legend ('kinetic energy', 'potential energy', 'total
25         energy', 'location', 'southwest');
26
27 title ('Energies for {\bf implicit midpoint} discrete evolution');
28
29 if (nargin > 3)
30     print ('-depsc2', sprintf ('%s_EnImid.eps', filename));
31 end
```



Verhalten der approximativen Energien: kinetische Energie :  $E_{\text{kin}}(t) = \frac{1}{2}p(t)^2$   
 potentielle Energie :  $E_{\text{pot}}(t) = -\frac{g}{l} \cos \alpha(t)$



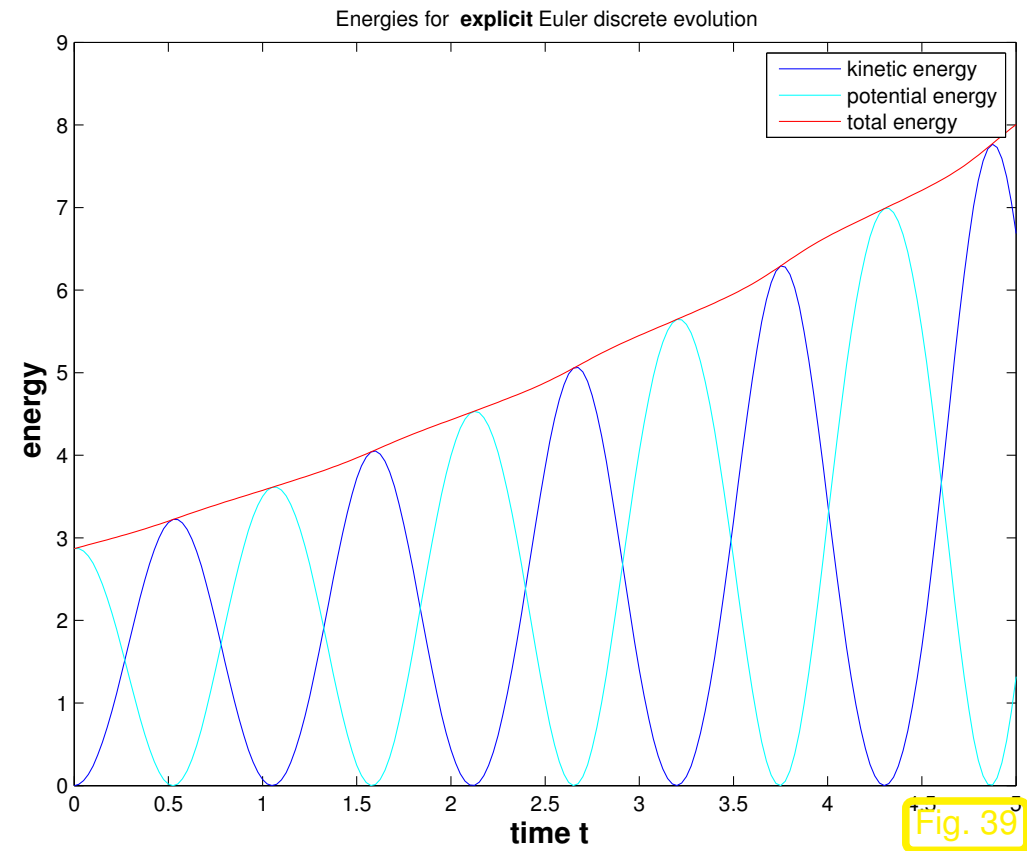


Fig. 39

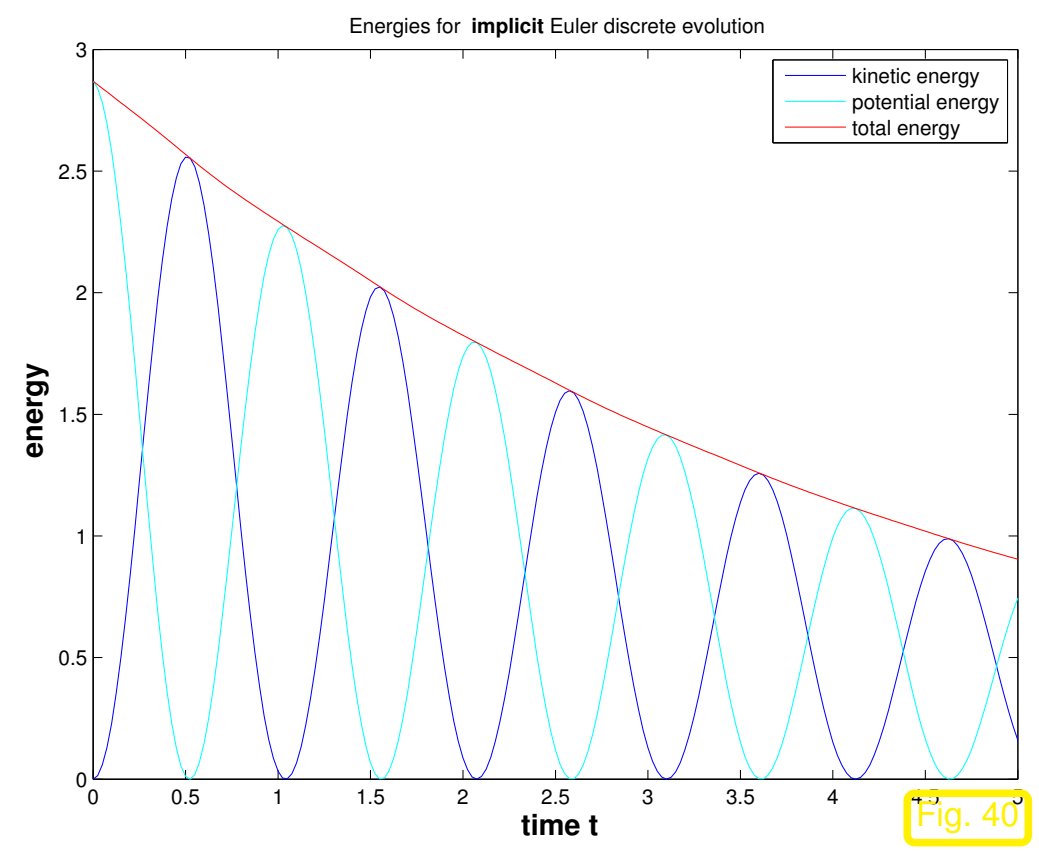


Fig. 40

☞ Expliziter Euler: Anwachsen der Gesamtenergie des Pendels

☞ Impliziter Euler: Pendel kommt zur Ruhe („numerische Reibung“)



Beispiel 1.4.18 (Eulerverfahren für längenerhaltende Evolution).

Anfangswertproblem für ,  $D = \mathbb{R}^2$ :

$$\dot{\mathbf{y}} = \begin{pmatrix} y_2 \\ -y_1 \end{pmatrix}, \quad \mathbf{y}(0) = \mathbf{y}_0 \quad \blacktriangleright \quad \mathbf{y}(t) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \mathbf{y}_0 .$$



Erstes Integral ( $\rightarrow$  Def. 1.2.7):

$$I(\mathbf{y}) = \|\mathbf{y}\|$$

(Bewegung mit konstanter Geschwindigkeit auf Kreisbahn)

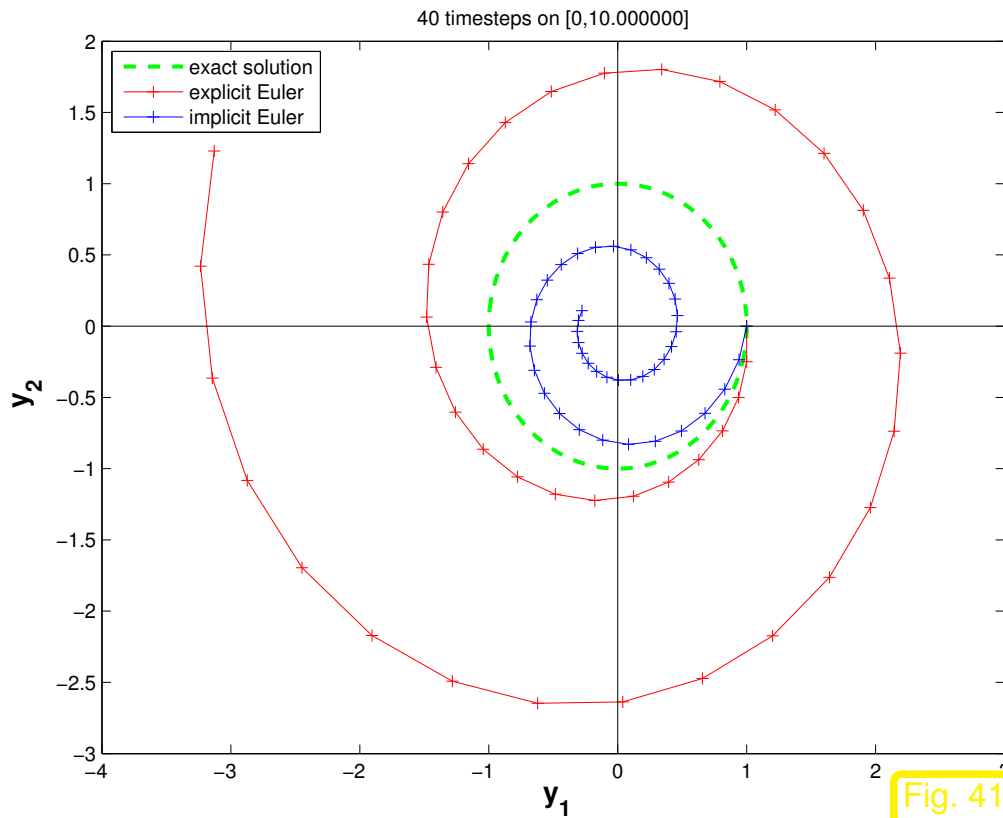


Fig. 41

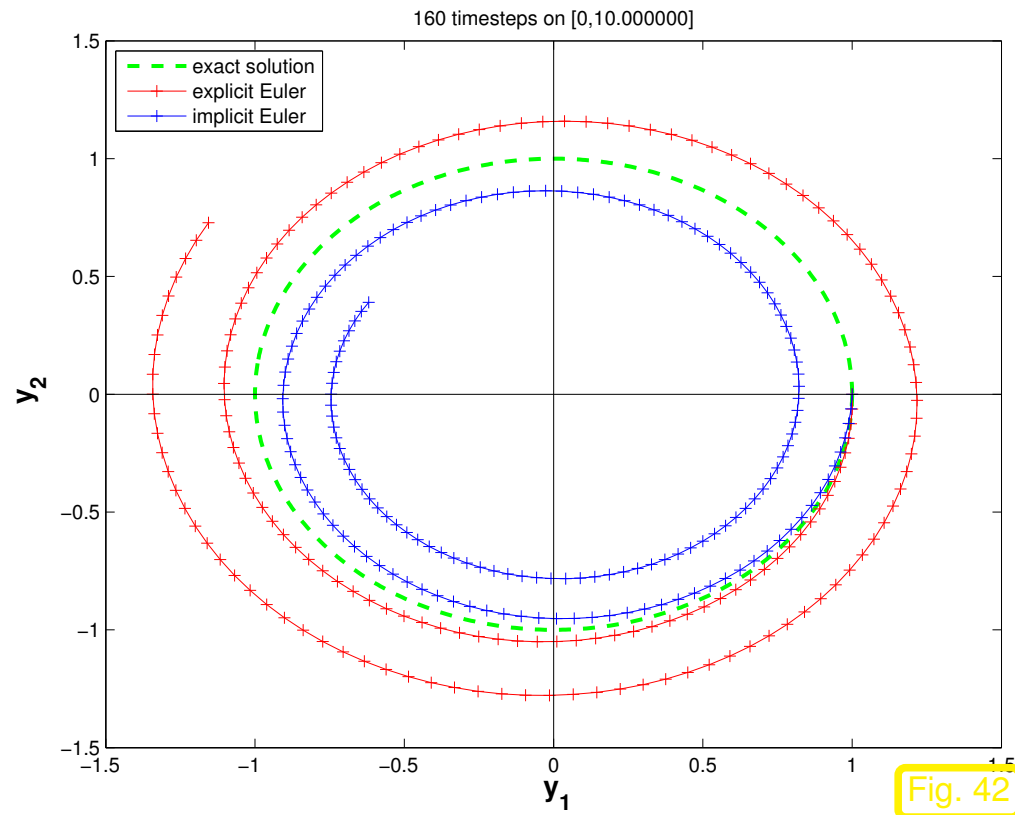


Fig. 42



Expliziter Euler: Numerische Lösung wird „aus der Kurve getragen“

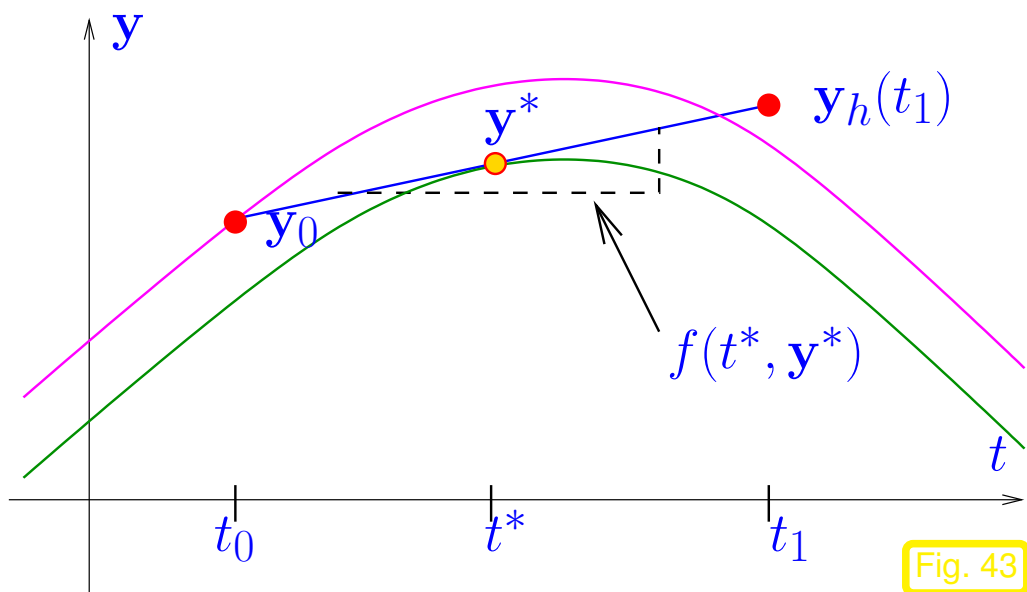


Impliziter Euler: Numerische Lösung „stürzt ins Zentrum“



### 1.4.3 Implizite Mittelpunktsregel

Wie vermeidet man die Energiedrift für explizites/implizites Euler-Verfahren angewandt auf konservative Systeme ?



Idee: Approximiere Lösung durch  $(t_0, \mathbf{y}_0)$  auf  $[t_0, t_1]$  durch

- lineares Polynom durch  $(t_0, \mathbf{y}_0)$
- mit Steigung  $f(t^*, \mathbf{y}^*)$ ,  

$$t^* := \frac{1}{2}(t_0 + t_1), \mathbf{y}^* = \frac{1}{2}(\mathbf{y}_0 + \mathbf{y}_1)$$

- ◁ —  $\hat{=}$  Lösungskurve durch  $(t_0, \mathbf{y}_0)$ ,
- $\hat{=}$  Lösungskurve durch  $(t^*, \mathbf{y}^*)$ ,
- $\hat{=}$  Tangente an — in  $(t^*, \mathbf{y}^*)$ .

Anwendung auf kleine Zeitintervalle  $[t_0, t_1], [t_1, t_2], \dots, [t_{N-1}, t_N]$  ➤ **implizite Mittelpunktsregel**

▶ durch implizite Mittelpunktsregel erzeugte Näherung  $\mathbf{y}_{k+1}$  für  $\mathbf{y}(t_k)$  erfüllt

$$\mathbf{y}_{k+1} := \mathbf{y}_h(t_{k+1}) = \mathbf{y}_k + h_k \mathbf{f}\left(\frac{1}{2}(t_k + t_{k+1}), \frac{1}{2}(\mathbf{y}_k + \mathbf{y}_{k+1})\right), \quad k = 0, \dots, N-1, \quad (1.4.19)$$

mit lokaler **(Zeit)schrittweite**  $h_k := t_{k+1} - t_k$ .

Beachte: (1.4.19) erfordert Auflösen einer (evtl. nichtlinearen) Gleichung nach  $\mathbf{y}_{k+1}$  !

▶ Terminologie „implizit“)

*Bemerkung 1.4.20* (Implizite Mittelpunktsregel als Differenzenverfahren).

(1.4.19) aus Approximation der Zeitableitung  $\frac{d}{dt}$  durch **zentralen Differenzenquotienten** auf **Zeitgitter**  $\mathcal{G} := \{t_0, t_1, \dots, t_N\}$ :

$$\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}) \quad \longleftrightarrow \quad \frac{\mathbf{y}_h(t_{k+1}) - \mathbf{y}_h(t_k)}{h_k} = \mathbf{f}\left(\frac{1}{2}(t_k + t_{k+1}), \frac{1}{2}(\mathbf{y}_h(t_k) + \mathbf{y}_h(t_{k+1}))\right), \quad k = 0, \dots, N-1.$$

△

Beispiel 1.4.21 (Implizite Mittelpunktsregel für logistische Dgl.).

Wiederholung der numerischen Experimente aus Beispiel 1.4.9 für implizite Mittelpunktsregel (1.4.19):

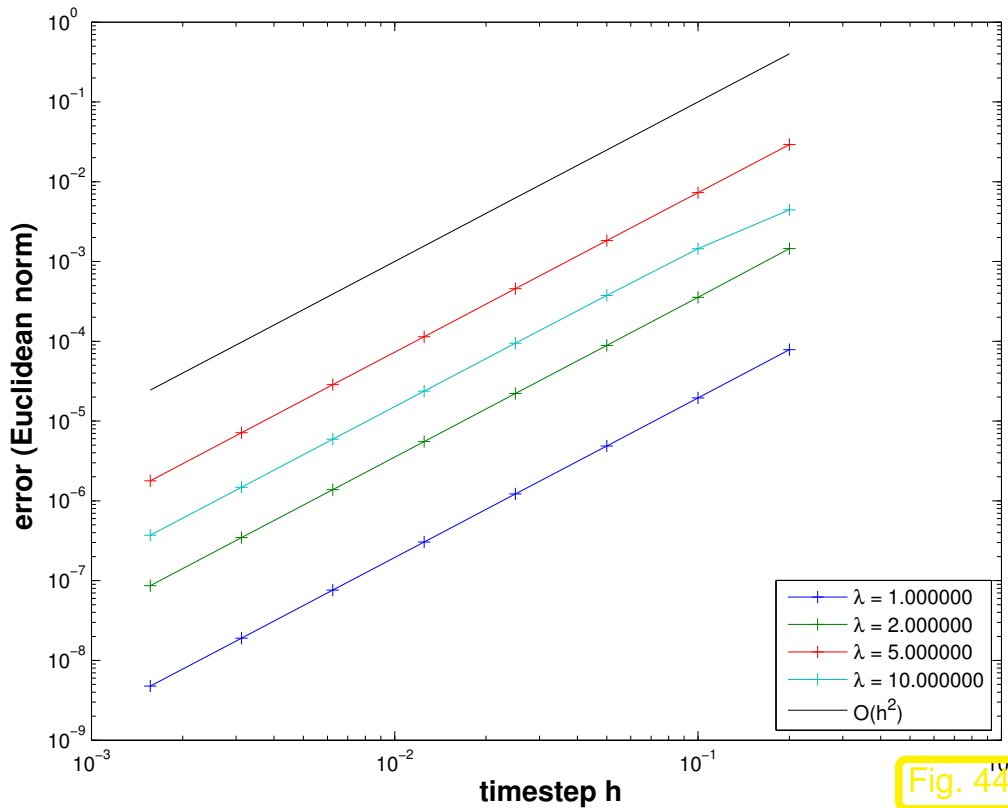


Fig. 44

$\lambda$  klein:  $O(h^2)$ -Konvergenz (asymptotisch)

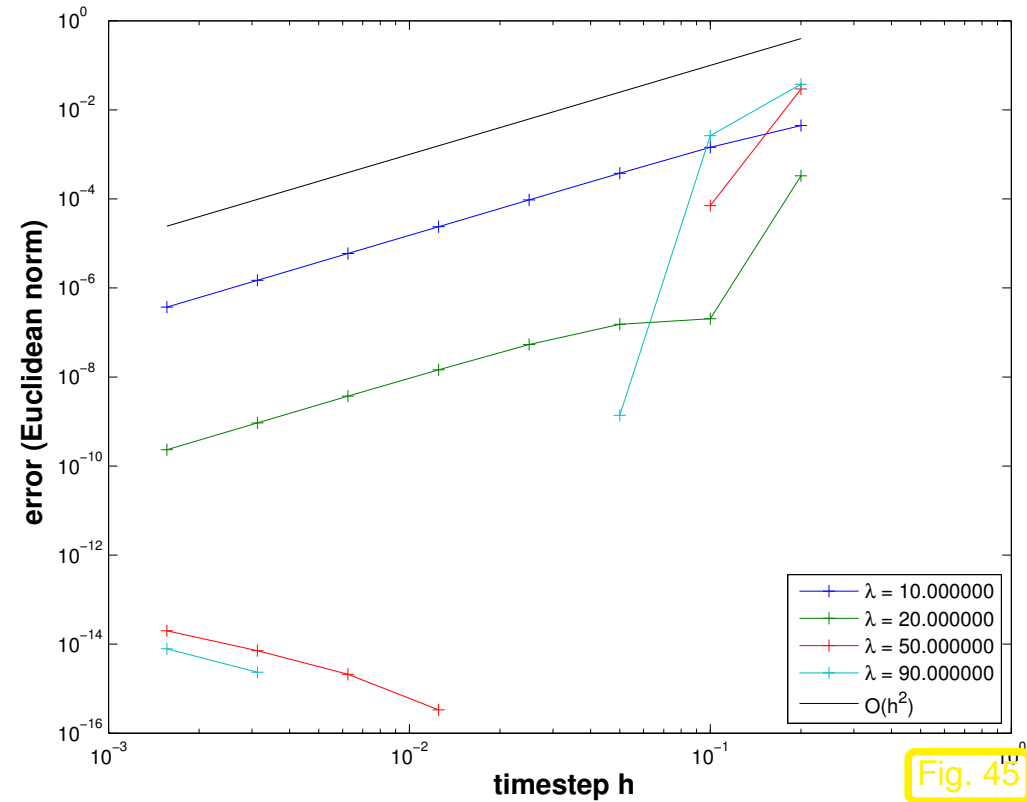
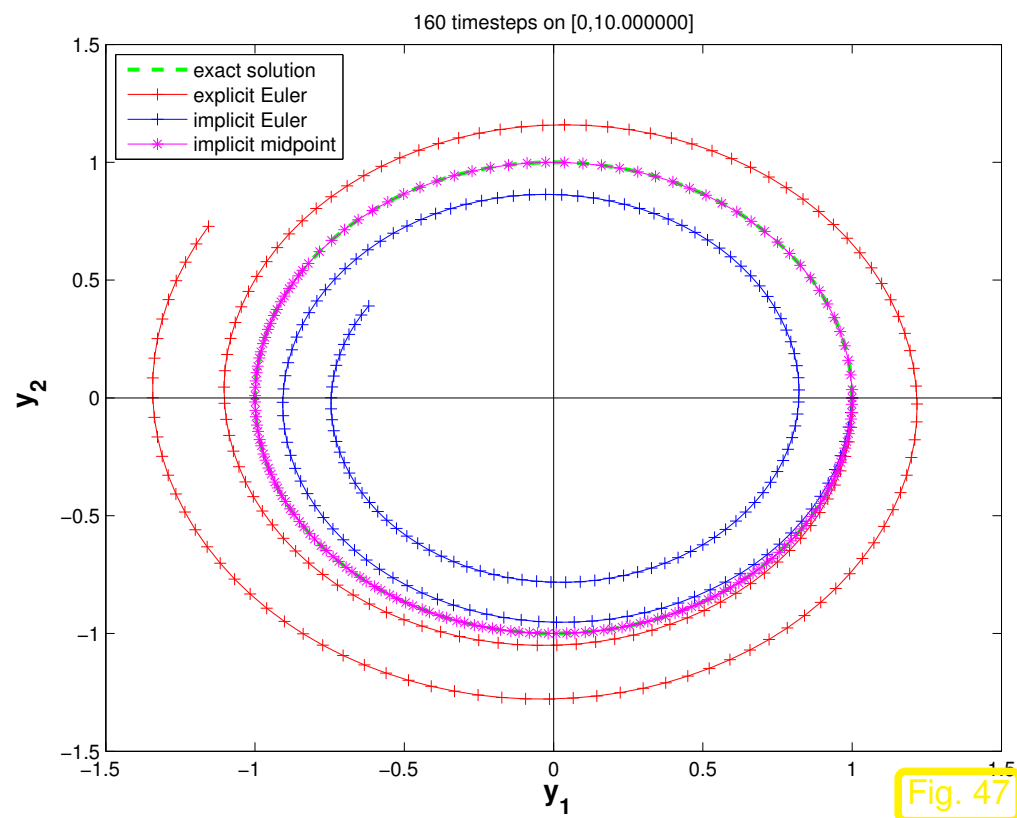
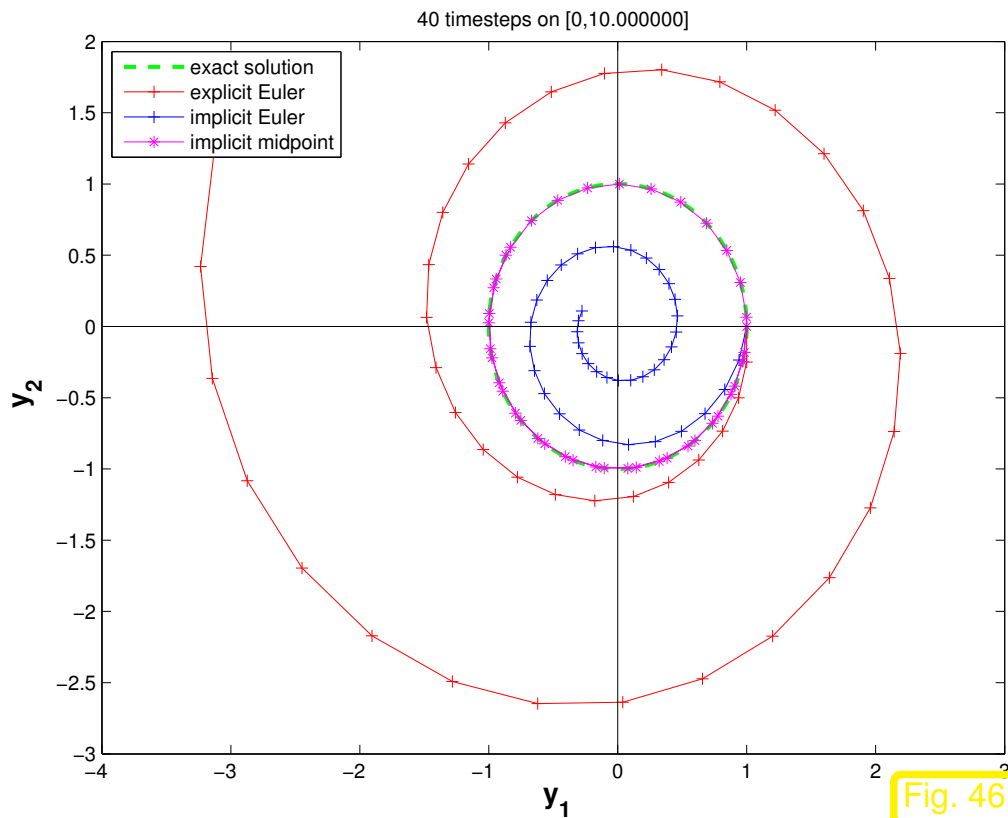


Fig. 45

$\lambda$  gross: stabil für alle Zeitschrittweiten  $h$  !  $\diamond$

Beispiel 1.4.22 (Implizite Mittelpunktsregel für Kreisbewegung).



☞ Implizite Mittelpunktsregel: Perfekte Längenerhaltung !

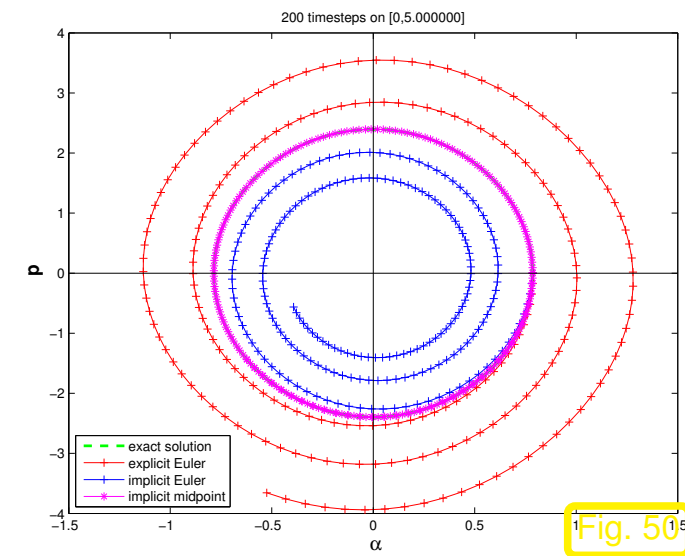
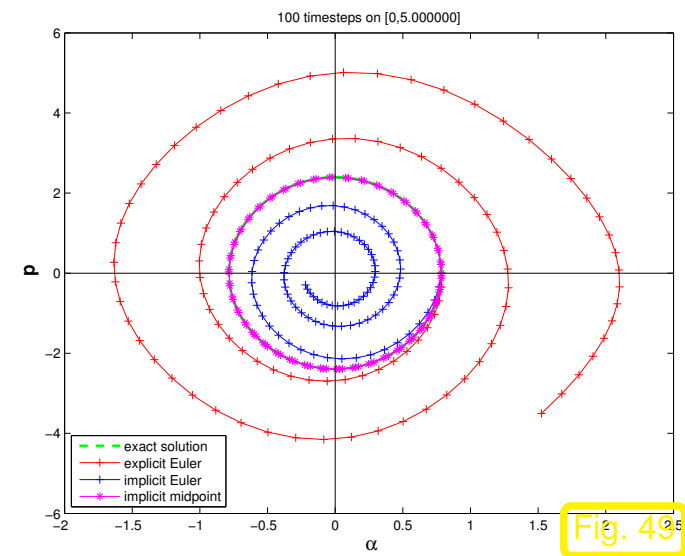
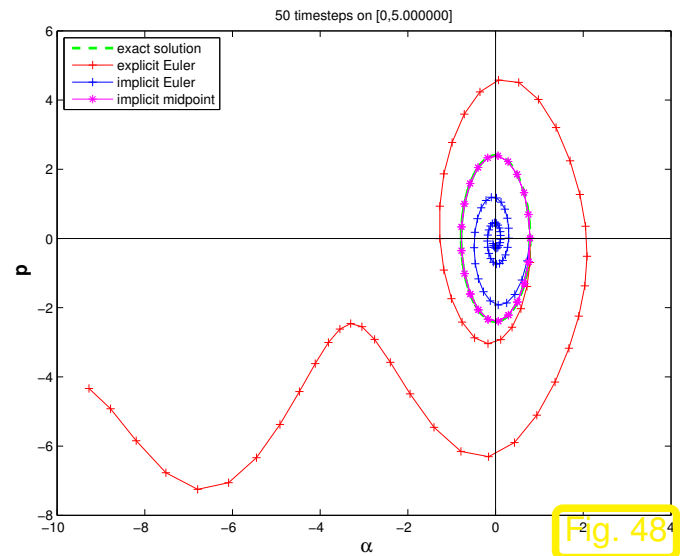
**Lemma 1.4.23** (Erhaltung quadratischer erster Integrale durch implizite Mittelpunktsregel).

Falls  $I : D \subset \mathbb{R}^d \mapsto \mathbb{R}$ ,  $I(\mathbf{y}) := \frac{1}{2}\mathbf{y}^T \mathbf{A} \mathbf{y}$ ,  $\mathbf{A} \in \mathbb{R}^{d,d}$ , erstes Integral ( $\rightarrow$  Def. 1.2.7) der autonomen Dgl.  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$  mit global differenzierbarer rechter Seite  $\mathbf{f} : D \mapsto \mathbb{R}^d$ , dann gilt

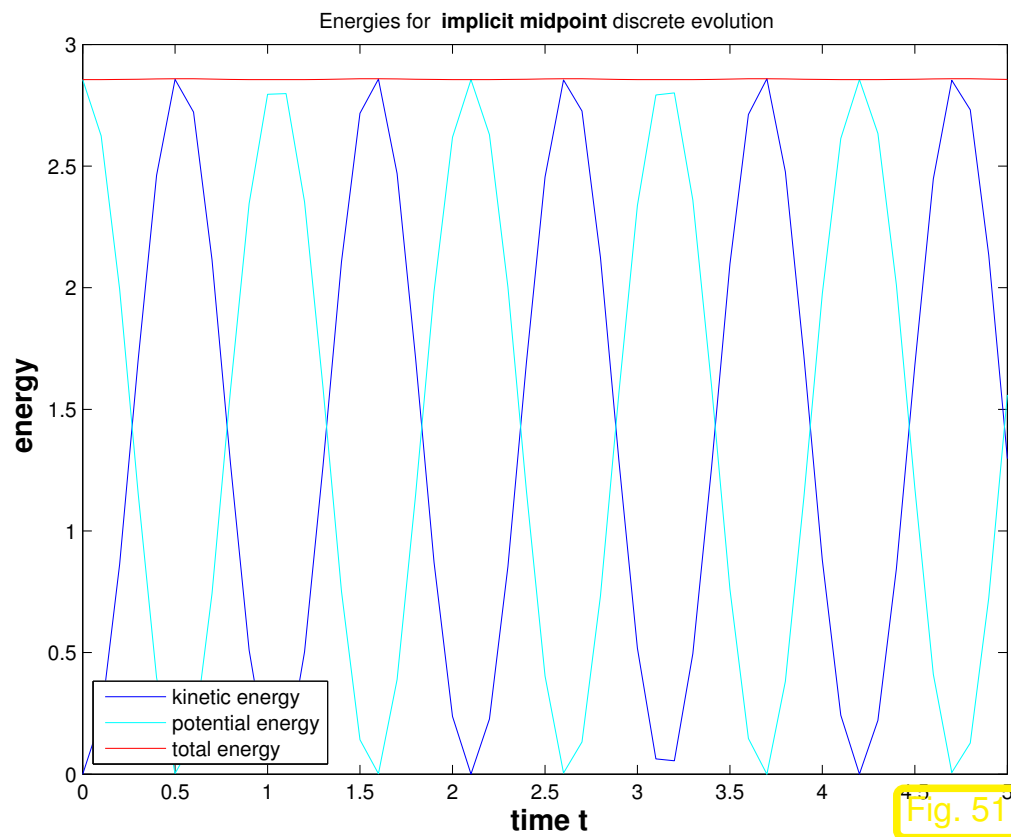
$$I(\mathbf{y}_k) = I(\mathbf{y}_0) \quad \forall k \in \mathbb{Z} \quad \text{für } \mathbf{y}_k \text{ gemäss (1.4.19)}$$

# Beispiel 1.4.24 (Implizite Mittelpunktsregel für Pendelgleichung).

Anfangswertproblem und numerische Experimente wie in Bsp. 1.4.17



R. Hiptmair  
rev 35327,  
25. April  
2011



◁ Verhalten der Energien bei numerischer Integration mit impliziter Mittelpunktsregel (1.4.19),  $N = 50$ .

Keine (sichtbare) **Energiedrift** im Vergleich zu Euler-Verfahren (trotz grosser Zeitschritte)

◇ R. Hiptmair  
rev 35327,  
25. April  
2011

## 1.4.4 Störmer-Verlet-Verfahren [15]

Übertragung der Idee der Euler-Verfahren ( $\rightarrow$  Sect. 1.4.1, 1.4.2) auf Differentialgleichungen 2. Ordnung

$$\ddot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}) . \quad (1.4.25)$$



Gegeben  $\mathbf{y}_{k-1} \approx \mathbf{y}(t_{k-1})$ ,  $\mathbf{y}_k \approx \mathbf{y}(t_k)$  approximiere  $\mathbf{y}(t)$  auf  $[t_{k-1}, t_{k+1}]$  durch

- **Parabel**  $\mathbf{p}(t)$  durch  $(t_{k-1}, \mathbf{y}_{k-1})$ ,  $(t_k, \mathbf{y}_k)$  (\*),
- mit  $\ddot{\mathbf{p}}(t_k) = \mathbf{f}(\mathbf{y}_k)$  (\*).

(\*)  $\rightarrow$  Parabel eindeutig bestimmt.

$$\mathbf{y}_{k+1} := \mathbf{p}(t_{k+1}) \approx \mathbf{y}(t_{k+1})$$

Störmer-Verlet-Verfahren für (1.4.25) (Zeitgitter  $\mathcal{G} := \{t_0, t_1, \dots, t_N\}$ ):

$$\mathbf{y}_{k+1} = -\frac{h_k}{h_{k-1}}\mathbf{y}_{k-1} + \left(1 + \frac{h_k}{h_{k-1}}\right)\mathbf{y}_k + \frac{1}{2}(h_k^2 + h_k h_{k-1})\mathbf{f}(t_k, \mathbf{y}_k), \quad k = 1, \dots, N-1. \quad (1.4.26)$$

R. Hiptmair  
rev 35327,  
25. April  
2011

Für uniforme Zeitschrittweite  $h$ :

$$\mathbf{y}_{k+1} = -\mathbf{y}_{k-1} + 2\mathbf{y}_k + h^2\mathbf{f}(t_k, \mathbf{y}_k), \quad k = 1, \dots, N-1. \quad (1.4.27)$$

Beachte: (1.4.26) erfordert nicht das Lösen einer Gleichung (► explizites Verfahren)

Terminologie:  $\mathbf{y}_{k+1} = \mathbf{y}_{k+1}(\mathbf{y}_k, \mathbf{y}_{k-1})$  ► (1.4.26) ist ein **Zweischrittverfahren**  
(Explizites/implizites Euler-Verfahren, Mittelpunktsregel = **Einschrittverfahren**)

**Bemerkung 1.4.28** (Störmer-Verlet-Verfahren als Differenzenverfahren).

(1.4.27) aus Approximation der zweiten Zeitableitung durch **zweiten zentralen Differenzenquotienten** auf Zeitgitter  $\mathcal{G} := \{t_0, t_1, \dots, t_N\}$ : für uniforme Zeitschrittweite  $h > 0$

$$\ddot{\mathbf{y}} = \mathbf{f}(\mathbf{y}) \quad \longleftrightarrow \quad \frac{\frac{\mathbf{y}_h(t_{k+1}) - \mathbf{y}_h(t_k)}{h} - \frac{\mathbf{y}_h(t_k) - \mathbf{y}_h(t_{k-1})}{h}}{h} = \frac{\mathbf{y}_h(t_{k+1}) - 2\mathbf{y}_h(t_k) + \mathbf{y}_h(t_{k-1}))}{h^2} = \mathbf{f}(\mathbf{y}_h(t_k)) .$$

△

**Bemerkung 1.4.29** (Startschritt für Störmer-Verlet-Verfahren).

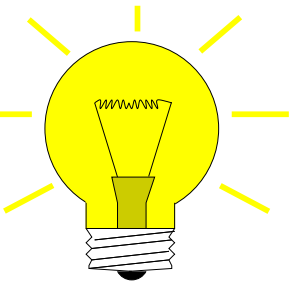
Anfangswerte für (1.4.25), siehe Bem. 1.1.16:  $\mathbf{y}(0) = \mathbf{y}_0, \dot{\mathbf{y}}(0) = \mathbf{v}_0$

- Benutze virtuellen Zeitpunkt  $t_{-1} := t_0 - h_0$
- Wende (1.4.27) an auf  $[t_{-1}, t_1]$ :

$$\mathbf{y}_1 = -\mathbf{y}_{-1} + 2\mathbf{y}_0 + h_0^2 \mathbf{f}(t_0, \mathbf{y}_0) . \quad (1.4.30)$$

- Zentraler Differenzenquotient auf  $[t_{-1}, t_1]$ :

$$\frac{\mathbf{y}_1 - \mathbf{y}_{-1}}{2h_0} = \mathbf{v}_0 . \quad (1.4.31)$$



Beispiel 1.4.32 (Störmer-Verlet-Verfahren für Pendelgleichung).

Listing 1.6: Störmer-Verlet-Verfahren für Pendelgleichung

```
1 function sverletpend(y0,v0,T,N,filename)
2 % MATLAB function for Stoermer-Verlet simulation of movement of mathematical
3 % pendulum, Example 1.4.32
4 %
5 % y0,v0: initial values for angle and its temporal derivative
6 % T : final time
7 % N : number of timesteps
8
9 g = 9.8; l = 1;
10
11 % right hand side for (1.4.25) (Newton's equations of motion)
12 f = @(y) -g/l*sin(y);
13
14 h = T/N; % uniform timestep
15 y_old = y0; y_new = h*v0+y0+ 0.5*h*h*f(y0); % initial step, see
    Rem. 1.4.29
```

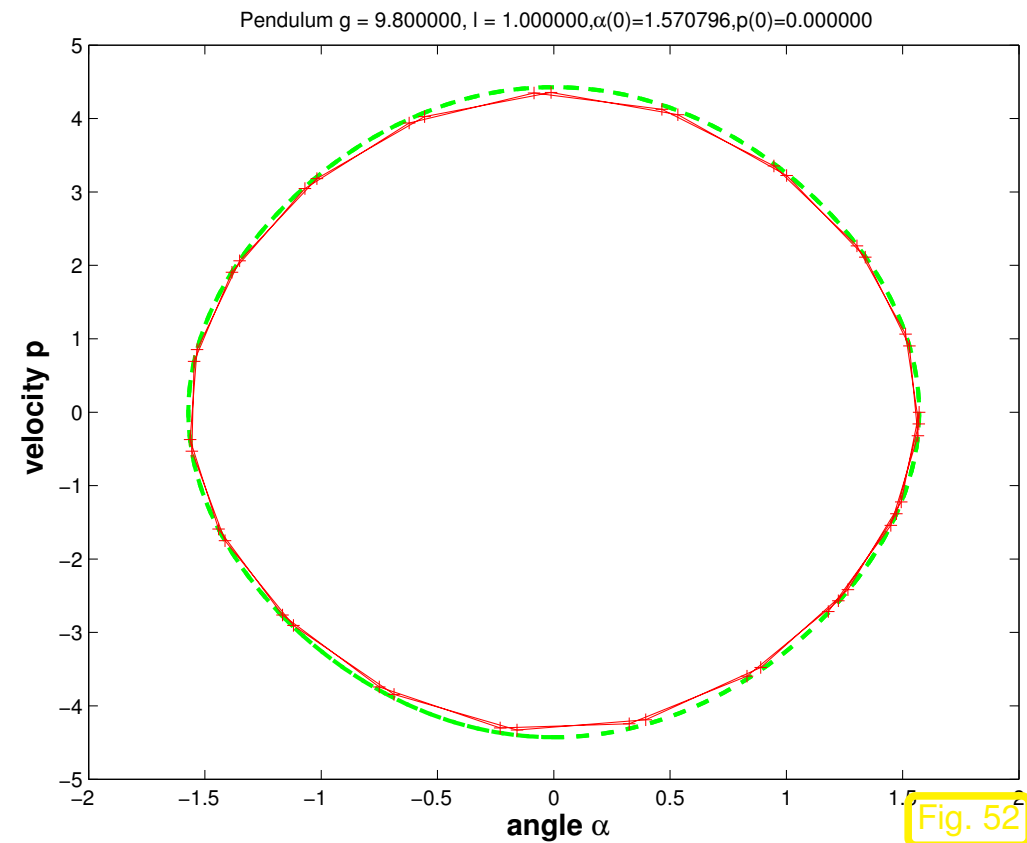
R. Hiptmair  
rev 35327,  
25. April  
2011

```
16
17 % Stoermer-Verlet iteration
18 y_sv = y0; y_sv = [y_sv, y_new]; y_p = v0;
19 for k=2:N+1
20     y = -y_old + 2*y_new + h*h*f(y_new);
21     y_p = [y_p, (y-y_old)/(2*h)];
22     y_old = y_new; y_new = y;
23     y_sv = [y_sv, y];
24 end
25 y_sv = y_sv(1:end-1);
26
27 % right hand side (Hamiltonian form) for computation of reference solution
28 % with high-order single step method with tight tolerances
29 odefun = @(t,y) [y(2); -g/l*sin(y(1))];
30 [t,y] = ode45(odefun, [0,T], [y0;v0], ...
31             odeset('abstol', 1E-11, 'reltol', 1E-11, 'stats', 'on'));
32
33 % Plot of angle vs. time
34 figure ('name', 'Pendulum alpha');
35 plot(t, y(:,1), 'g-', h*(0:N), y_sv, 'r-+');
36 xlabel ('{\bf time t}', 'fontsize', 14);
37 ylabel ('{\bf angle \alpha}', 'fontsize', 14);
38 title (sprintf ('Pendulum g = %f, l =
```

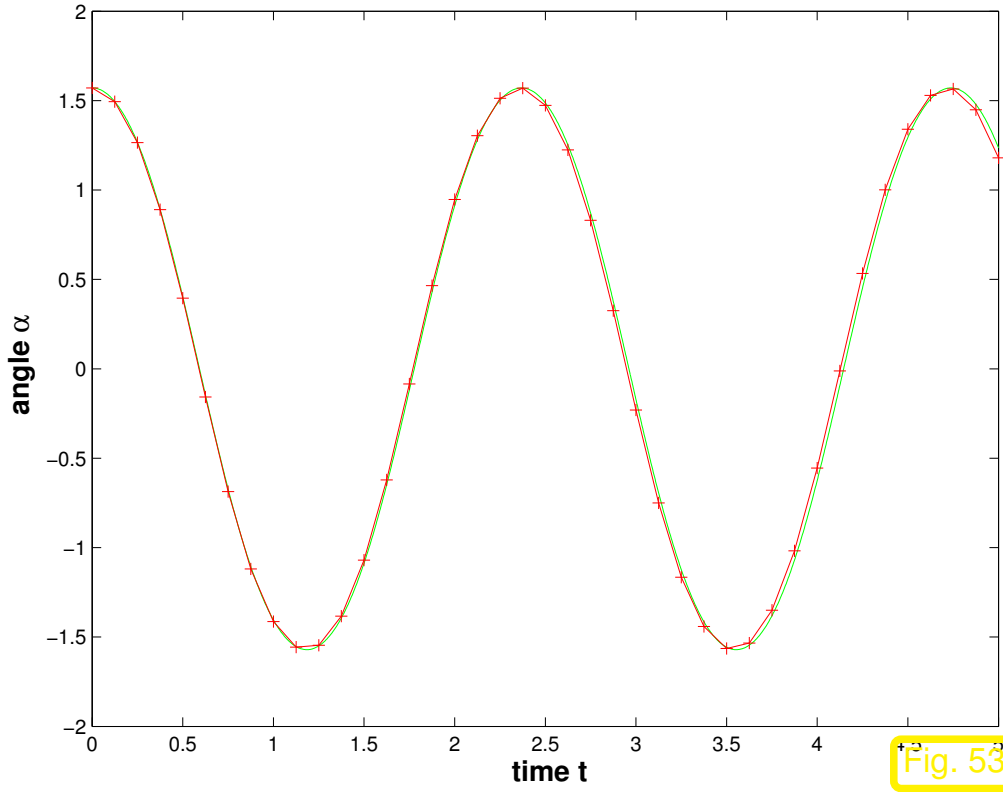
```
39 | %f, \\alpha(0)=%f,p(0)=%f'
    if nargin
        print '-depsc2' sprintf '%s_alpha.eps' end
40
41 % Plot of velocity vs. time
42
43 figure 'name' 'Pendulum velocity'
44 plot          'g-'          'r-+'
45 xlabel '{\bf time t}' 'fontsize'
46 ylabel '{\bf velocity p}' 'fontsize'
47 title sprintf 'Pendulum g = %f, l =
    | %f, \\alpha(0)=%f,p(0)=%f'
48 if nargin          print '-depsc2' sprintf '%s_p.eps'
    end
49
50 % PLOT of trajectory in phase space
51 figure 'name' 'Pendulum trajectory'
52      plot          'g--'          'r-+'
53 set          'linewidth'
54 xlabel '{\bf angle \alpha}' 'fontsize'
55 ylabel '{\bf velocity p}' 'fontsize'
56 title sprintf 'Pendulum g = %f, l =
    | %f, \\alpha(0)=%f,p(0)=%f'
```

```
57 if (nargin > 3),
58     print ('-depsc2', sprintf ('%s_orbit.eps', filename)); end
59
60 % Tracking of energies
61 E_kin = 0.5*(y_p.^2);
62 E_pot = -g/l*cos(y_sv);
63 E_pot = E_pot - min(E_pot) + min(E_kin);
64 E_tot = E_kin + E_pot;
65
66 figure ('name', 'Pendulum: energy');
67 plot (tg, E_kin, 'b-', tg, E_pot, 'c-', tg, E_tot, 'r-');
68 xlabel ('\bf time t', 'fontsize', 14);
69 ylabel ('\bf energy', 'fontsize', 14);
70 legend ('kinetic energy', 'potential energy', 'total
    energy', 'location', 'southeast');
71 title ('Energies for {\bf Stoermer-Verlet} discrete evolution');
72 if (nargin > 3),
    print ('-depsc2', sprintf ('%s_EnSV.eps', filename)); end
```

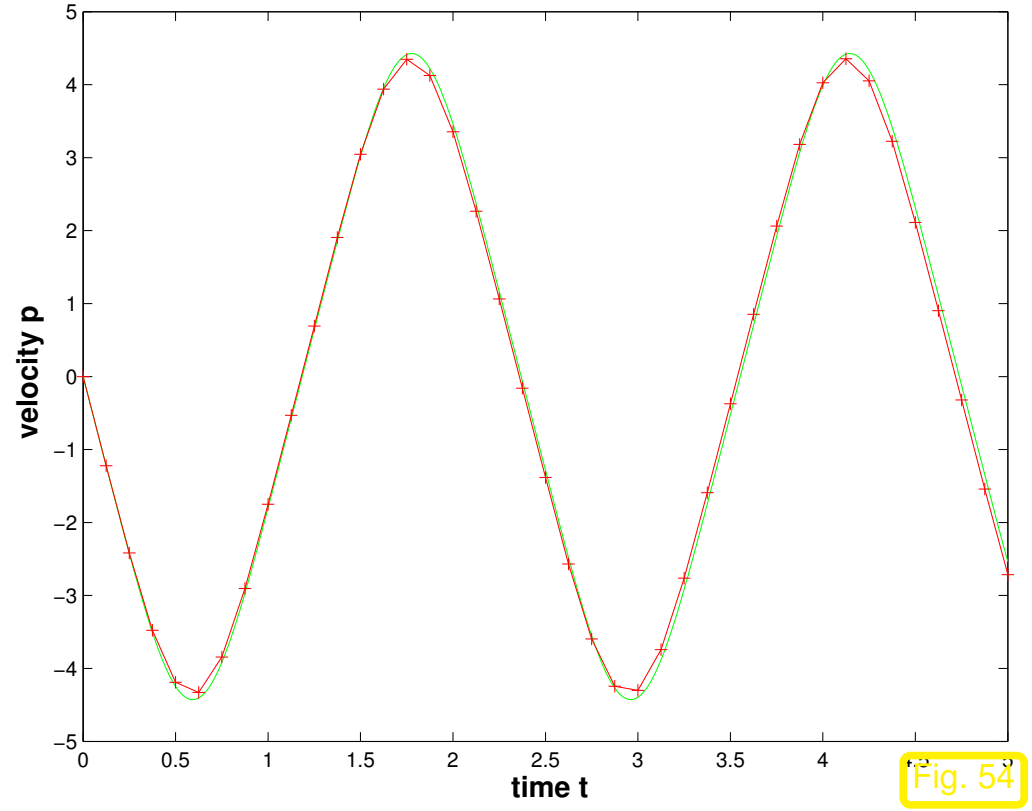
- (1.4.27) angewandt auf (1.2.18)
- Startschritt gemäss Bem. 1.4.29
- Uniforme Zeitschrittweite  $h := T/N$ ,  $N \in \mathbb{N}$   
Zeitschritte
- Referenzlösung durch MATLAB-Funktion  
`ode45()` (extrem kleine Toleranzen)
- $\alpha_0 = \pi/2$ ,  $p_0 = 0$ ,  $T = 5$ , vgl. Bsp. 1.4.17
- Anzahl Zeitschritte:  $N = 40$



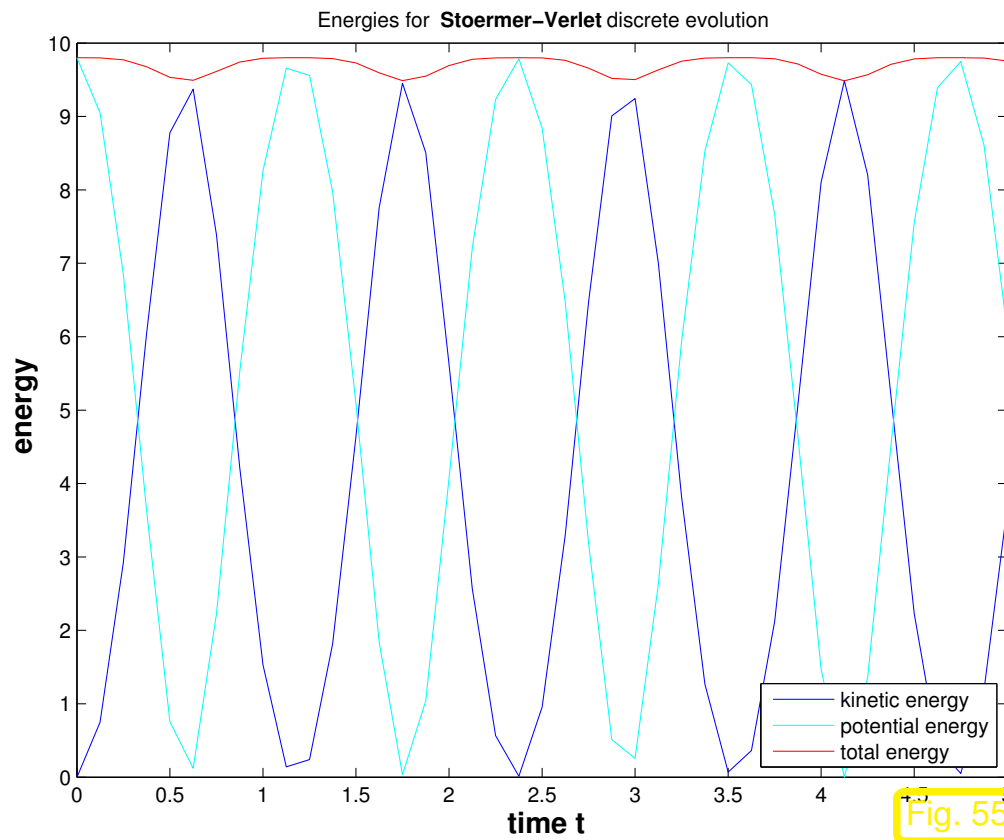
Pendulum  $g = 9.800000$ ,  $l = 1.000000$ ,  $\alpha(0)=1.570796$ ,  $p(0)=0.000000$



Pendulum  $g = 9.800000$ ,  $l = 1.000000$ ,  $\alpha(0)=1.570796$ ,  $p(0)=0.000000$







☞ Keine Energiedrift trotz grosser Zeitschrittweite

Perfekt periodische Orbits !

Kontrast: Bsp. 1.4.17

*Bemerkung* 1.4.33 (Einschrittformulierung des Störmer-Verlet-Verfahrens).

Für uniforme Zeitschrittweite, vgl. (1.4.27), analog zur Umwandlung einer Dgl. 2. Ordnung  $\rightarrow$  Dgl. 1.

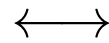
Ordnung, siehe (1.1.10): mit  $\mathbf{v}_{k+\frac{1}{2}} := \frac{\mathbf{y}_{k+1} - \mathbf{y}_k}{h}$

$$\ddot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$$



$$\mathbf{y}_{k+1} - 2\mathbf{y}_k + \mathbf{y}_{k-1} = h^2 \mathbf{f}(\mathbf{y}_k)$$

Zweischrittverfahren



$$\begin{aligned} \dot{\mathbf{y}} &= \mathbf{v}, \\ \dot{\mathbf{v}} &= \mathbf{f}(\mathbf{y}). \end{aligned}$$



$$\begin{aligned} \mathbf{v}_{k+\frac{1}{2}} &= \mathbf{v}_k + \frac{h}{2} \mathbf{f}(\mathbf{y}_k), \\ \mathbf{y}_{k+1} &= \mathbf{y}_k + h \mathbf{v}_{k+\frac{1}{2}}, \\ \mathbf{v}_{k+1} &= \mathbf{v}_{k+\frac{1}{2}} + \frac{h}{2} \mathbf{f}(\mathbf{y}_{k+1}). \end{aligned}$$

Einschrittverfahren

Startschritt ( $\rightarrow$  Bem. 1.4.29) ist implizit in der Einschrittformulierung enthalten.

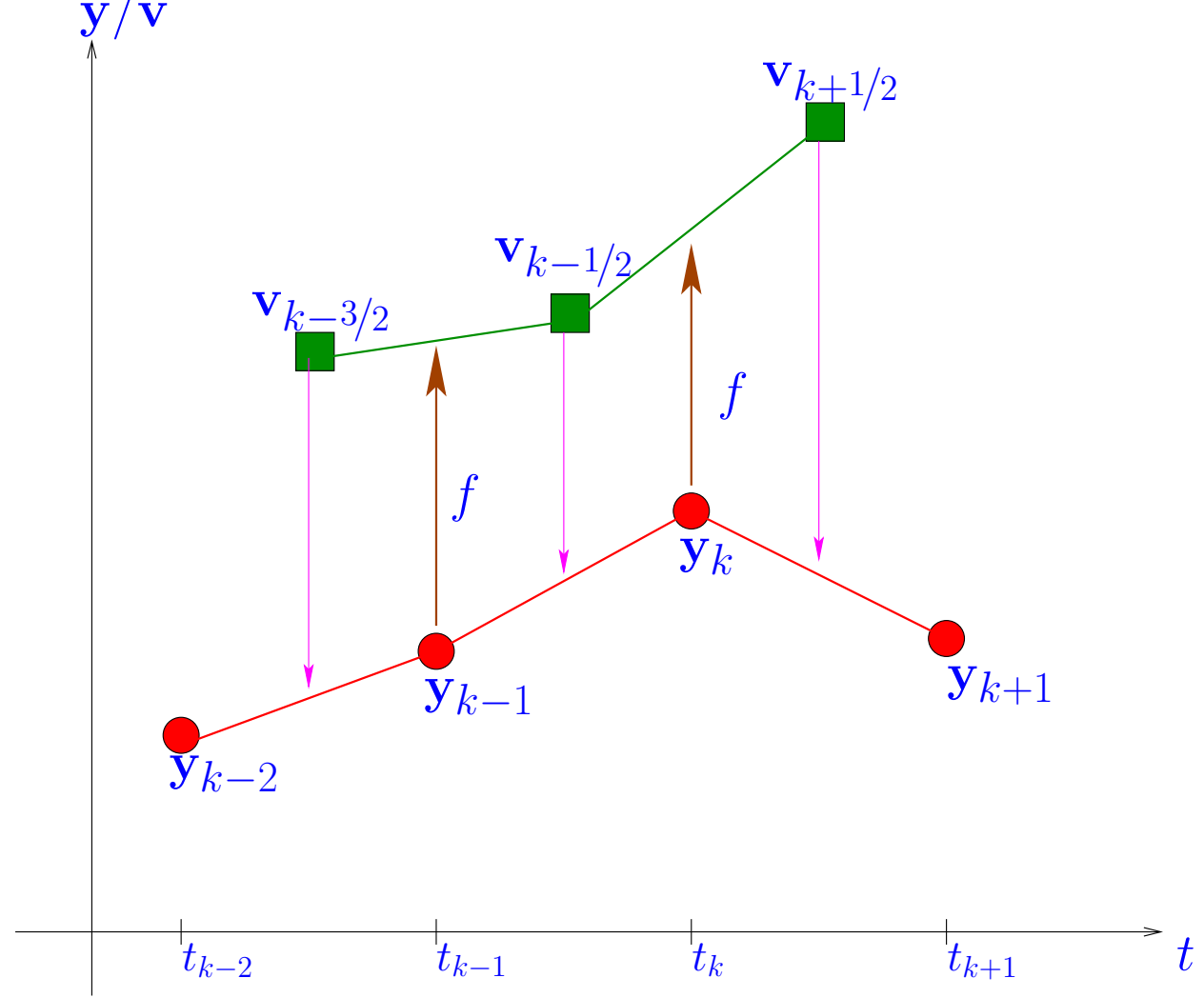


*Bemerkung 1.4.34* (Störmer-Verlet-Verfahren als Polygonzugmethode).

Perspektive: Störmer-Verlet-Verfahren  
als Einschrittverfahren  
(siehe Bem. 1.4.33)

$$\mathbf{v}_{k+\frac{1}{2}} = \mathbf{v}_{k-\frac{1}{2}} + h\mathbf{f}(\mathbf{y}_k),$$

$$\mathbf{y}_{k+1} = \mathbf{y}_k + h\mathbf{v}_{k+\frac{1}{2}}.$$



Erinnerung (Bem. 1.2.4) an die Frage „Warum viele verschiedene numerischer Lösungsverfahren für ODEs?“

Antwort: Jeder numerische Integrator hat spezielle Eigenschaften  
↳ besonders geeignet/ungeeignet für bestimmte Klassen von AWP

## 2

## Einschrittverfahren

## 2.1 Grundlagen

Gegeben:  $\mathbf{f} : \Omega \mapsto \mathbb{R}^d$  lokal Lipschitz-stetig ( $\rightarrow$  Def. 1.3.2) auf erweitertem Zustandsraum  $\Omega \subset \mathbb{R} \times D$

➤ Definiert ODE  $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$  ( $\rightarrow$  Sect. 1.1)

▶ Zugehöriger Evolutionsoperator:  $\Phi^{s,t} : D \mapsto D$  ( $\rightarrow$  Def. 1.3.7)

Gegeben: Anfangsdaten  $(t_0, \mathbf{y}_0) \in \Omega$  ➤ Konkretes Anfangswertproblem (1.1.13)

Ziel: ➤ Approximation von  $\mathbf{y}(T)$  für Endzeitpunkt  $T \in J(t_0, \mathbf{y}_0)$ .

➤ Approximation der Funktion  $t \mapsto \mathbf{y}(t)$ ,  $t \in [t_0, T]$ ,  $T \in J(t_0, \mathbf{y}_0)$  ▷  $\mathbf{y}_h(t)$ .

*Bemerkung 2.1.1* (Glattheitsannahmen an rechte Seite  $\mathbf{f}$ ).

Für die Konvergenztheorie von Einschrittverfahren:

**Annahme:**  $\mathbf{f}$  „hinreichend“ glatt  $\Rightarrow t \mapsto \mathbf{y}(t)$  „hinreichend“ glatt

- Beweise werden zunächst für beliebig glattes  $\mathbf{f}$  konzipiert.
- Im Nachhinein werden minimale Glattheitsanforderungen an  $\mathbf{f}$  für die jeweiligen Aussagen spezifiziert.  
(Dies wird im diesem Kurs in der Regel übersprungen werden)



## 2.1.1 Abstrakte Einschrittverfahren [8, Sect. 4.1]

- Baustein: **Verfahrensfunktion** (diskrete Evolution)

$$\Psi : \tilde{\Omega}_h \subset I \times I \times D \mapsto \mathbb{R}^d$$

✎ Notation:  $\Psi^{s,t} \mathbf{y} := \Psi(s, t; \mathbf{y})$

- Baustein: **Zeitgitter**  $\mathcal{G} := \{t_0, t_1, \dots, t_N = T\}$ ,  $t_0 < t_1 < \dots < t_N$ .

(Terminologie:  $t_k \hat{=}$  Gitterpunkte, lokale **(Zeit)schrittweite**  $h_k := t_{k+1} - t_k$ )

✎ Notation: globale Zeitschrittweite  $h = h_{\mathcal{G}} = \max_{0 \leq i < N} h_k$

### Definition 2.1.2 (Einschrittverfahren).

Gegeben: diskrete Evolution  $\Psi$  und Zeitgitter  $\mathcal{G} := \{t_0 < t_1 < \dots < t_N = T\}$ . Die Rekursion

$$\mathbf{y}_{k+1} := \Psi(t_k, t_{k+1}; \mathbf{y}_k), \quad k = 0, \dots, N-1, \quad (2.1.3)$$

definiert ein **Einschrittverfahren** (ESV, engl. single step method) zum Anfangswertproblem (1.1.13).

**Bemerkung 2.1.4** (Notation fuer Einschrittverfahren).

Oft spezifiziert man Einschrittverfahren durch Angabe des ersten Schritts

$$\mathbf{y}_1 = \text{Ausdruck in } \mathbf{y}_0 \text{ und } \mathbf{f} .$$

Dieser Gepflogenheit wird sich auch diese Vorlesung manchmal anschliessen.

**Einschrittverfahren**

```
function [T,Y] = esv(Psi,tspan,y0)
t = tspan(1); y = y0; Y = y; T = t;
while (t < tspan(2))
    h = Aktuelle Zeitschrittweite
    y = Psi(t,t+h,y); t = t+h;
    Y = [Y,y]; T = [T,t];
end
```

**Funktionshandle**

Psi = @(t0,t1,y) ...

Beachte: Die aktuelle Zeitschrittweite wird jeweils aus den Genauigkeitsanforderungen und  $\mathbf{y}_k$  berechnet(→ Sect. 2.6).



**Definition 2.1.5** (Explizite und implizite Einschrittverfahren).

Ein Einschrittverfahren zur approximativen Lösung eines AWP heisst **explizit**, falls die zugrundeliegende diskrete Evolution durch endlich viele  $f$ -Auswertungen zu realisieren ist.

Die diskrete Evolution eines **impliziten** Einschrittverfahrens erfordert die Lösung eines Gleichungssystems.

▶ ESV + Anfangswert + Zeitgitter erzeugt Gitterfunktion  $\mathbf{y}_{\mathcal{G}} : \mathcal{G} \mapsto \mathbb{R}^d$ ,  $\mathbf{y}_{\mathcal{G}}(t_k) = \mathbf{y}_k$

Bei „geschickter Wahl“ von  $\Psi$ :  $\mathbf{y}_k \approx \mathbf{y}(t_k)$  ( $\mathbf{y} \hat{=}$  exakte Lösung)

**Definition 2.1.6** (Diskretisierungsfehler).

- Für gegebenes  $T \in J(t_0, \mathbf{y}_0)$ , sei  $\mathbf{y} : [t_0, T] \mapsto \mathbb{R}^d$  Lösung des AWP (1.1.13)
- $\mathbf{y}_{\mathcal{G}}$  eine Näherungslösung auf dem Gitter  $\mathcal{G} = \{t_0 < t_1 < \dots < t_N = T\}$ .

▶ **Diskretisierungsfehler**  $\epsilon_{\mathcal{G}} := \max_{0 \leq k \leq N} \|\mathbf{y}(t_k) - \mathbf{y}_k\|$  .

Hier ist  $\|\cdot\|$  irgendeine Vektornorm auf dem Zustandsraum  $D \subset \mathbb{R}^d$ . Wegen der Äquivalenz aller Normen auf endlichdimensionalen Vektorräumen, gelten alle im folgenden abgeleiteten Aussagen für beliebige Normen.

**Definition 2.1.7** (Konvergenz und Konvergenzordnung).  $\rightarrow [8, \text{Def. 4.6}]$

Das ESV (2.1.3) zum AWP (1.1.13) **konvergiert**, falls

$$\forall \epsilon > 0: \exists \delta > 0: \forall \text{Zeitgitter } \mathcal{G} \subset [0, T]: h_{\mathcal{G}} \leq \delta \Rightarrow \begin{array}{l} \text{ESV wohldefiniert,} \\ \epsilon_{\mathcal{G}} \leq \epsilon. \end{array}$$

(Kurz:  $\epsilon_{\mathcal{G}} \rightarrow 0$ , falls  $h_{\mathcal{G}} \rightarrow 0$ )

Das ESV heisst (algebraisch  $\rightarrow$  Def. 1.4.5) **konvergent von der Ordnung**  $p \in \mathbb{N}$ , falls

$$\exists h_0 > 0, C > 0: \begin{array}{l} \text{ESV wohldefiniert,} \\ \epsilon_{\mathcal{G}} \leq Ch_{\mathcal{G}}^p \end{array} \quad \forall \text{Zeitgitter } \mathcal{G}, h_{\mathcal{G}} \leq h_0.$$

(Kurzschreibweise mit Landau-Symbol  $\epsilon_{\mathcal{G}} = O(h^p)$ )

Erweiterung: Konvergenz für alle  $(t_0, \mathbf{y}_0) \in \Omega \triangleright$  **globale Konvergenz**

Beachte: Konvergenz gemäss Def. 2.1.7 ist ein **asymptotischer Begriff** ( $h_G \rightarrow 0$ )



Die Aussage, dass ein Verfahren mit einer gewissen Ordnung konvergiert, sagt uns in der Regel *nichts* über die tatsächliche Grösse (einer Norm) des Fehlers. Solche stärkeren Aussagen gelingen der numerischen Analysis von Einschrittverfahren in der Regel nicht.

Was nützt denn dann das Wissen über Konvergenz der Ordnung  $p$  überhaupt ?

Wenn wir annehmen, dass die Aussage scharf ist, also  $\epsilon_G \approx Ch_G^p$ , dann können wir schliessen, um welchen Faktor wir die globale Zeitschrittweite verringern müssen, um den Fehler um einen vorgegebenen Faktor zu reduzieren.

## 2.1.2 Konsistenz [8, Sect. 4.1.1]

Kontinuierliche Evolution ( $\rightarrow$  Def. 1.3.7)  $\longleftrightarrow$

Diskrete Evolution

$$\Phi^{s,t}$$

$$\Psi^{s,t}$$

erfüllt für alle  $(t, \mathbf{y}) \in \Omega$

sollte erfüllen:

- (i)  $\Phi^{t,t} \mathbf{y} = \mathbf{y}$
- (ii)  $\left. \frac{d}{ds} \Phi^{t,t+s} \mathbf{y} \right|_{s=0} = \mathbf{f}(t, \mathbf{y})$
- (iii)  $\Phi^{r,s} \Phi^{t,r} \mathbf{y} = \Phi^{t,s} \mathbf{y} \quad \forall r, s \in J(t, \mathbf{y})$

- (i)  $\Psi^{t,t} \mathbf{y} = \mathbf{y}$  klar!
- (ii)  $\left. \frac{d}{ds} \Psi^{t,t+s} \mathbf{y} \right|_{s=0} = \mathbf{f}(t, \mathbf{y})$  unbedingt!
- (iii)  $\Psi^{r,s} \Psi^{t,r} \mathbf{y} = \Psi^{t,s} \mathbf{y} \quad \forall r, s \in J(t, \mathbf{y})$  utopisch!

R. Hiptmair  
rev 35327,  
25. April  
2011

Unter der Annahme, dass  $t \mapsto \Psi^{s,t} \mathbf{y}$  differenzierbar:

Falls  $\Psi$  (i)–(iii) erfüllt, dann gilt  $\Psi = \Phi$  !

$$\frac{d}{dt} \left( \Psi^{s,t} \mathbf{y} \right) = \lim_{\tau \rightarrow 0} \frac{\Psi^{s,t+\tau} \mathbf{y} - \Psi^{s,t} \mathbf{y}}{\tau} \stackrel{\text{(iii)}}{=} \lim_{\tau \rightarrow 0} \frac{\Psi^{t,t+\tau}(\Psi^{s,t} \mathbf{y}) - \Psi^{t,t}(\Psi^{s,t} \mathbf{y})}{\tau} \stackrel{\text{(ii)}}{=} \mathbf{f}(t, \Psi^{s,t} \mathbf{y}) .$$

$t \mapsto \Psi^{s,t}$  löst das gleiche Anfangswertproblem für  $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$  wie  $t \mapsto \Phi^{s,t} \mathbf{y}$ . Mit Satz von Picard-Lindelöf ( $\rightarrow$  Thm. 1.3.4) folgt  $\Psi = \Phi$ .

**Definition 2.1.8** (Konsistenz einer diskreten Evolution).

Diskrete Evolution  $\Psi$  ist *konsistent* mit der ODE  $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$ , falls für alle  $(t, \mathbf{y}) \in \Omega$

$$\Psi^{t,t} \mathbf{y} = \mathbf{y} \quad \text{und} \quad \left. \frac{d}{ds} \Psi^{t,t+s} \mathbf{y} \right|_{s=0} = \mathbf{f}(t, \mathbf{y}) .$$

**Lemma 2.1.9** (Darstellung konsistenter diskreter Evolutionen).  $\rightarrow$  [8, Lemma 4.4]

Sei  $(t, \mathbf{y}) \in \Omega$  und  $s \mapsto \Psi^{t,t+s} \mathbf{y}$  stetig differenzierbar in Umgebung von 0.

$\Psi$  ist genau dann konsistent mit  $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$  ( $\rightarrow$  Def. 2.1.8), wenn eine auf dieser Nullumgebung stetige *Inkrementfunktion*  $h \mapsto \psi(t, \mathbf{y}, h)$  existiert mit

$$\Psi^{t,t+h} \mathbf{y} = \mathbf{y} + h\psi(t, \mathbf{y}, h) \quad , \quad \psi(t, \mathbf{y}, 0) = \mathbf{f}(t, \mathbf{y}) . \quad (2.1.10)$$


**Definition 2.1.11** (Konsistenzfehler einer diskreten Evolution).  $\rightarrow$  [8, Def. 4.3]

*Konsistenzfehler:*  $\tau(t, \mathbf{y}, h) := \Phi^{t,t+h} \mathbf{y} - \Psi^{t,t+h} \mathbf{y}$  ( $h$  hinreichend klein); .

**Lemma 2.1.12** (Konsistenz und Konsistenzfehler).

Sei  $(t, \mathbf{y}) \in \Omega$ ,  $s \mapsto \Psi^{t, t+s} \mathbf{y}$  stetig differenzierbar in einer Umgebung von 0.  $\Psi$  ist genau dann konsistent mit  $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$  ( $\rightarrow$  Def. 2.1.8), wenn für den Konsistenzfehler gilt

$$\|\boldsymbol{\tau}(t, \mathbf{y}, h)\| = o(h) \quad \text{für } h \rightarrow 0 \quad \text{lokal gleichmässig in } (t, \mathbf{y}) \in \Omega .$$

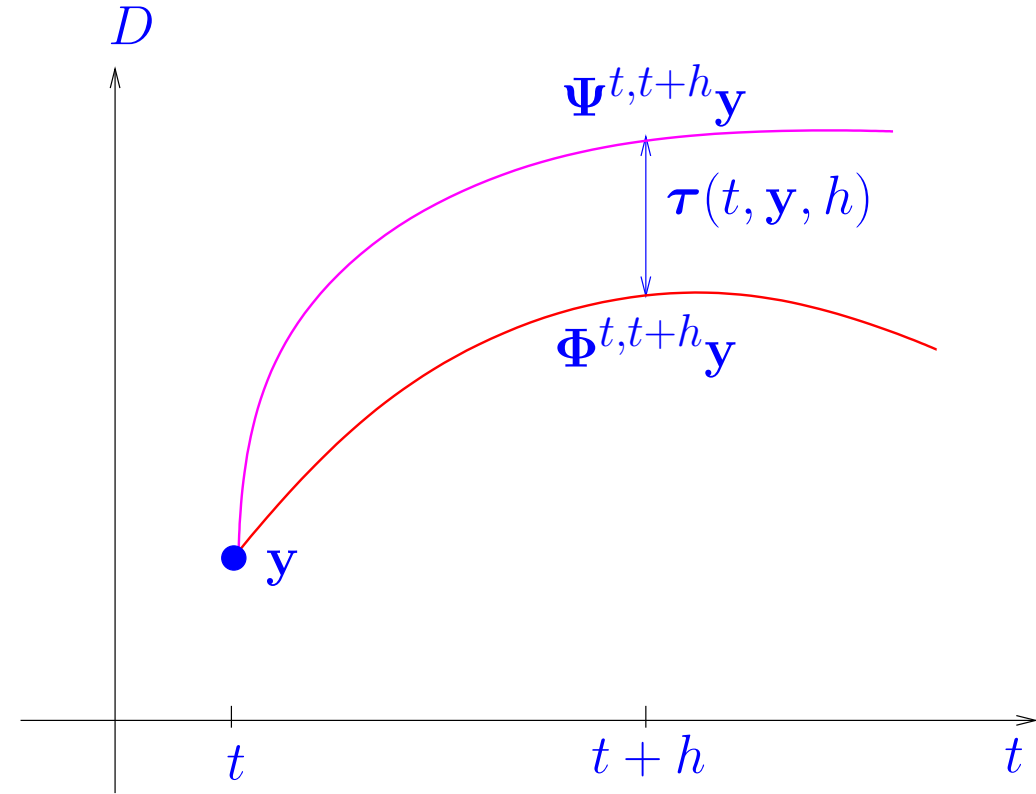
 Notation: „Landau-o“:  $g(h) = o(h) \iff \frac{g(h)}{h} \rightarrow 0$  für  $h \rightarrow 0$

Interpretation:

Konsistenzfehler = Einschrittfehler

—  $\hat{=}$  exakte Lösung durch  $(y, y)$

—  $\hat{=}$  Näherungslösung aus diskreter Evolution



**Definition 2.1.13** (Konsistenzordnung einer diskreten Evolution). [8, Def. 4.7]

Eine diskrete Evolution hat **Konsistenzordnung**  $p \in \mathbb{N}$ , falls für den Konsistenzfehler **lokal gleichmässig** in  $\Omega$  gilt

$$\|\tau(t, \mathbf{y}, h)\| = O(h^{p+1}) \quad \text{für } h \rightarrow 0. \quad (2.1.14)$$

Dass (2.1.14) *lokal gleichmässig* gilt bedeutet

$$\forall (t, \mathbf{y}) \in \Omega: \exists h_0, \delta, C > 0: \tau(\tilde{t}, \tilde{\mathbf{y}}, h) \leq Ch^{p+1} \quad \forall \tilde{t}, \tilde{\mathbf{y}}, h: \quad |\tilde{t} - t| \leq \delta, \|\tilde{\mathbf{y}} - \mathbf{y}\| \leq \delta, \\ 0 \leq h \leq h_0 .$$

Wegen der Äquivalenz aller Normen auf dem endlichdimensionalen Raum  $\mathbb{R}^d$  ist die Wahl der Norm in den Definitionen 2.1.11 und 2.1.13 belanglos.

Technik zur Bestimmung der Konsistenzordnung:

*Taylor-Entwicklung*

*Beispiel 2.1.15* (Konsistenzordnung einfacher Einschrittverfahren).

Implizite Mittelpunktsregel (1.4.19):

$$\mathbf{y}_1 = \mathbf{y}_0 + hf\left(\frac{1}{2}(t_0 + t_1), \frac{1}{2}(\mathbf{y}_0 + \mathbf{y}_1)\right)$$

Beachte: Keine explizite Formel für  $\Psi$  ! (“implicit” method)

Für die „faulen“ Leute: Computeralgebra (MAPLE) !



$$D(y) := x \rightarrow f(y(x));$$

$$D(y) := x \mapsto f(y(x))$$

$$y0 := y(0);$$

$$y0 := y(0)$$

$$\text{solve}(y0+h*f((y0+y1)/2)=y1, \{y1\});$$

$$\{y1 = \text{RootOf}(-y(0) - hf(1/2 y(0) + 1/2 _Z) + _Z)\}$$

$$\text{assign}(\%);$$

$$\text{taylor}(y1-y(h), h=0, 4);$$

$$\text{series}\left(\left(-1/6 \left(D^{(2)}\right)(f)(y(0)) (f(y(0)))^2 - 1/6 (D(f)(y(0)))^2 f(y(0))\right.\right. \\ \left.\left.+ 1/8 f(y(0)) \left(\left(D^{(2)}\right)(f)(y(0)) f(y(0)) + 2 (D(f)(y(0)))^2\right)\right) h^3 + O(h^4), h, 4\right)$$

► Implizite Mittelpunktsregel hat Konsistenzordnung 2 !



Bestimmung der *formalen* Konsistenzordnung eines ESV immer unter der Annahme  
hinreichender (\*) Glattheit der exakten Lösung !

(\*) : „hinreichend“  $\hat{=}$  so glatt, wie für Taylorentwicklung erforderlich

Falls exakte Lösung nicht hinreichend glatt  $\rightarrow$  eingeschränkte Bedeutung der Konsistenzordnung für das tatsächliche Verhalten eines Verfahrens.

In dieser Vorlesung:

(Oft) stillschweigende Annahme „hinreichender Glattheit“ !

## 2.1.3 Konvergenz

Betrachte wird Anfangswertproblem

$$\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}) \quad , \quad \mathbf{y}(t_0) = \mathbf{y}_0 \quad \text{für ein} \quad (t_0, \mathbf{y}_0) \in \Omega := I \times D .$$

1.1.13

$$f : \Omega \mapsto \mathbb{R}^d$$

1.3.2

1.3.4

$$t \mapsto \mathbf{y}(t)$$

Betrachte Einschrittverfahren ( $\rightarrow$  Def. 2.1.2) mit Verfahrensfunktion  $\Psi : \Omega_h \subset I \times I \times D \mapsto \mathbb{R}^d$

$$\Psi^{t,t+h} \mathbf{y} := \mathbf{y} + h\psi(t, \mathbf{y}, h) . \quad (2.1.17)$$

Inkrementfunktion

Annahme: **Lokale** Abschätzung für Konsistenzfehler ( $\rightarrow$  Def. 2.1.11): für ein  $p \in \mathbb{N}$

$$\forall (\bar{t}, \bar{\mathbf{y}}) \in \Omega: \quad \exists C_c > 0, \delta > 0: \quad \left\| \Phi^{t,t+h} \mathbf{y} - \Psi^{t,t+h} \mathbf{y} \right\| \leq C_c h^{p+1} \quad \forall |h| \text{ hinreichend klein ,}$$
$$\forall t \mathbf{y}: |t - \bar{t}| < \delta, \quad \|\mathbf{y} - \bar{\mathbf{y}}\| < \delta .$$

(2.1.18)

Beachte: Konsistenzordnung  $p$  für diskrete Evolution  $\Psi$  bzgl.  $\dot{\mathbf{y}} = f(t, \mathbf{y}) \quad \Rightarrow \quad (2.1.18)$

$\mathbf{y}_{\mathcal{G}} = (\mathbf{y}_k)_{k=0}^N$ : Gitterfunktion erzeugt durch ESV  $\Psi$  auf Zeitgitter  $(T > t_0 \hat{=} \text{Endzeitpunkt})$

$$\mathcal{G} := \{t_0 < t_1 < \dots < t_N = T\} \subset J(t_0, \mathbf{y}_0),$$

vgl. Def. 2.1.2.

**Theorem 2.1.19** (Kovergenztheorem für Einschrittverfahren). [8, Thm. 4.10]

Es gelte Annahme (2.1.18) und die Darstellung (2.1.17). Ist die Inkrementfunktion  $\psi$  lokal Lipschitz-stetig ( $\rightarrow$  Def. 1.3.2) in der Zustandsvariablen  $\mathbf{y}$ , dann

- (i) liefert die Verfahrensfunktion  $\Psi$  für alle Zeitgitter  $\mathcal{G}$  mit hinreichend kleinem  $h_{\mathcal{G}}$  eine Gitterfunktion  $\mathbf{y}_{\mathcal{G}}$  zum Anfangswert  $\mathbf{y}_0$ ,
- (ii) konvergiert diese Familie  $\{\mathbf{y}_{\mathcal{G}}\}_{\mathcal{G}}$  von Gitterfunktionen von der Ordnung  $p$  gegen  $t \mapsto \mathbf{y}(t)$ , siehe Def. 2.1.7

Hilfsmittel beim Beweis:

**Lemma 2.1.20** (Diskretes Gronwall-Lemma, siehe Lemma 1.3.29).

Erfüllt die Folge  $(\xi_k)_{k \in \mathbb{N}_0}$ ,  $\xi_k \geq 0$ , die Differenzungleichung

$$\xi_{k+1} \leq Ch_k^{p+1} + (1 + Lh_k)\xi_k, \quad k \in \mathbb{N}_0, \quad L, C, h_k \geq 0, \quad (2.1.21)$$

so gilt

$$\xi_N \leq C \left( \max_{k=0, \dots, N-1} h_k^p \right) \frac{1}{L} \left( \exp\left(L \sum_{k=0}^{N-1} h_k\right) - 1 \right) + \exp\left(L \sum_{k=0}^{N-1} h_k\right) \cdot \xi_0, \quad N \in \mathbb{N}_0.$$

*Beweis.* (durch Induktion nach  $N$ )

Mit der Konvention, dass leere Summen verschwinden, gilt die Behauptung für  $N = 0$  (Induktionsbeginn)

Induktionsschluss:

$$\begin{aligned} \xi_{N+1} &\stackrel{(2.1.21)}{\leq} Ch_N^{p+1} + (1 + Lh_N)\xi_N \\ &\stackrel{*}{\leq} Ch_N^{p+1} + (1 + Lh_N) \left( C \left( \max_{k=0}^{N-1} h_k^p \right) \frac{1}{L} \left( \exp\left(L \sum_{k=0}^{N-1} h_k\right) - 1 \right) + \exp\left(L \sum_{k=0}^{N-1} h_k\right) \xi_0 \right) \end{aligned}$$

$$\leq C \left( \max_{k=0}^N h_k^p \right) \left( h_N + \frac{1}{L} \left( \exp \left( L \sum_{k=0}^N h_k \right) - 1 - L h_N \right) \right) + \exp \left( L \sum_{k=0}^N h_k \right) \xi_0$$

Das ist die Behauptung des Lemmas für  $N + 1$ .

\*: Benutzt die Induktionsannahme, d.h., die Behauptung des Lemmas für  $\xi_N$ .

■ Benutzt die elementare Abschätzung  $1 + x \leq \exp(x)$ ,  $x \in \mathbb{R}$  (Konvexität der Exponentialfunktion).

*Beweis* von Thm. 2.1.19; Verallgemeinerung des Beweises der algebraischen Konvergenz des expliziten Eulerverfahrens aus Abschnitt 1.4.1. Das vorbereitende Studium jenes Beweises wird empfohlen.

① Kompakte Umgebung der Lösungstrajektorie  $t \mapsto \mathbf{y}(t)$  zum Anfangswert  $\mathbf{y}_0$ :

$$K_\delta := \{(t, \mathbf{y}) \in I \times \mathbb{R}^d: t_0 \leq t \leq T, \|\mathbf{y} - \mathbf{y}(t)\| \leq \delta\}, \quad \delta > 0.$$

Für hinreichend kleines  $\delta > 0$ :  $K_\delta \subset \Omega$

Infolge der lokalen Lipschitz-Bedingung an  $\psi$  und der lokalen Konsistenzfehlerabschätzung (2.1.18)

② Annahme **A1**:  $(\mathbf{y}_k)_{k=0}^N$  existiert und  $\mathbf{y}_k \in K_\delta \subset \Omega$  für ein  $\delta > 0$ . Diese Annahme wird a posteriori (durch Induktion nach  $N$ ) für hinreichend kleines  $h_G$  besätigt.

③ Beachte  $\mathbf{y}_{k+1} = \Psi^{t_k, t_{k+1}} \mathbf{y}_k \quad \triangleright$  **Rekursion** für Fehler  $\mathbf{e}_k := \mathbf{y}(t_k) - \mathbf{y}_k$

$$\mathbf{e}_{k+1} = \underbrace{\left( \mathbf{y}(t_{k+1}) - \Psi^{t_k, t_{k+1}} \mathbf{y}(t_k) \right)}_{\text{Einschrittfehler}} + \underbrace{\left( \Psi^{t_k, t_{k+1}} \mathbf{y}(t_k) - \Psi^{t_k, t_{k+1}} \mathbf{y}_k \right)}_{\text{propagierter Fehler}} \quad (2.1.22)$$

$$\stackrel{(2.1.17)}{=} \tau(t_k, \mathbf{y}(t_k), h_k) + \mathbf{e}_k + h_k (\boldsymbol{\psi}(t_k, \mathbf{y}(t_k), h_k) - \boldsymbol{\psi}(t_k, \mathbf{y}_k, h_k)) ,$$

wobei die Def. 2.1.11 des Konsistenzfehlers  $\tau$  benutzt worden ist.

④ Kompaktheitsargumente:

- Konsequenz der lokalen Konsistenzfehlerabschätzung (2.1.18): für  $|h|$  hinreichend klein

$$\exists C > 0: \quad \left\| \Phi^{t, t+h} \mathbf{y} - \Psi^{t, t+h} \mathbf{y} \right\| \leq C_c h^{p+1} \quad \forall (t, \mathbf{y}), (t+h, \mathbf{y}) \in K_\delta . \quad (2.1.23)$$

- Konsequenz der lokalen Lipschitz-Stetigkeit ( $\rightarrow$  Def. 1.3.2) der Inkrementfunktion  $\boldsymbol{\psi}$ : für  $|h|$  hinreichend klein

$$\exists L > 0: \quad \left\| \boldsymbol{\psi}(t, \mathbf{z}, h) - \boldsymbol{\psi}(t, \mathbf{w}, h) \right\| \leq L \left\| \mathbf{z} - \mathbf{w} \right\| \quad \forall (t, \mathbf{z}), (t, \mathbf{w}) \in K_\delta . \quad (2.1.24)$$

⑤ (2.1.22) &  $\triangle$ -Ungleichung  $\Rightarrow$  Rekursion für Fehlernorm:

$$\begin{aligned} \|\mathbf{e}_{k+1}\| &\leq \|\mathbf{e}_k\| + \|\boldsymbol{\tau}(t_k, \mathbf{y}(t_k), h_k)\| + h_k \|\boldsymbol{\psi}(t_k, \mathbf{y}(t_k), h_k) - \boldsymbol{\psi}(t_k, \mathbf{y}_k, h_k)\| \\ (2.1.24) \quad &\leq \|\mathbf{e}_k\| + \|\boldsymbol{\tau}(t_k, \mathbf{y}(t_k), h_k)\| + h_k L \|\mathbf{y}(t_k) - \mathbf{y}_k\| \\ (2.1.23) \quad &\leq Ch_k^{p+1} + (1 + Lh_k) \|\mathbf{e}_k\| . \end{aligned}$$

Anwendung des diskreten Gronwall-Lemmas mit  $\xi_k := \|\mathbf{e}_k\|$ ,  $\xi_0 = 0$ :

$$\text{Lemma 2.1.20} \Rightarrow \|\mathbf{e}_k\| \leq Ch_{\mathcal{G}}^p \frac{\exp(L(T - t_0)) - 1}{L} .$$

⑥ Die Abschätzung zeigt, dass  $\mathbf{y}_k - \mathbf{y}(t_k) \rightarrow 0$  für  $h_{\mathcal{G}} \rightarrow 0$ . Damit kann durch Induktion bewiesen werden

$$\forall \delta > 0: \exists h^* = h^*(\delta) > 0: h_{\mathcal{G}} < h^* \Rightarrow (t_k, \mathbf{y}_k) \in K_{\delta} \quad \forall k .$$

Damit ist Annahme **A1** gerechtfertigt. □

Beachte: Der Beweis benutzt nur den *Konsistenzfehler entlang der Lösungstrajektorie*  $\boldsymbol{\tau}(t, \mathbf{y}(t), h)$ .  
Man kann also die Voraussetzung der Konsistenzordnung  $p$  ( $\rightarrow$  Def. 2.1.13) schwächer formulieren als

$$\|\boldsymbol{\tau}(t, \mathbf{y}(t), h)\| \leq C_c h^{p+1} \quad \forall t \in [t_0, T] , \quad \text{für } h \text{ hinreichend klein.}$$



Merkregel: (Nur) für Einschrittverfahren:

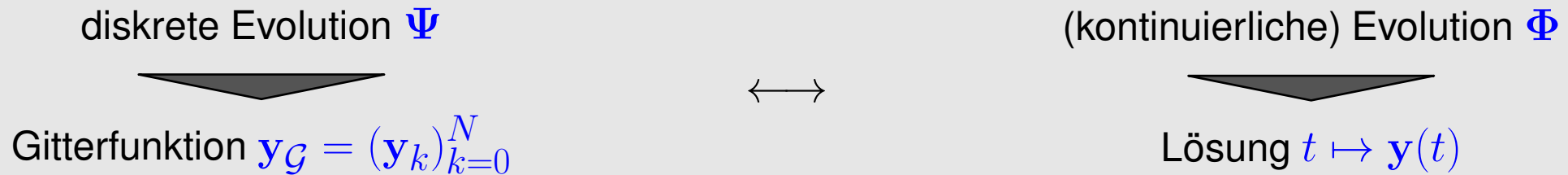
Konsistenzordnung  $p \implies$  Konvergenzordnung  $p$

Aus dem diskreten Gronwall-Lemma ergibt sich, dass die Konstante in der asymptotischen Fehlerabschätzung von Thm. 2.1.19 *exponentiell* von  $T - t_0$  abhängt. Dies macht die Abschätzung des Theorems u.U. wertlos für *Langzeitintegration*, vgl. Lemma 4.4.82.

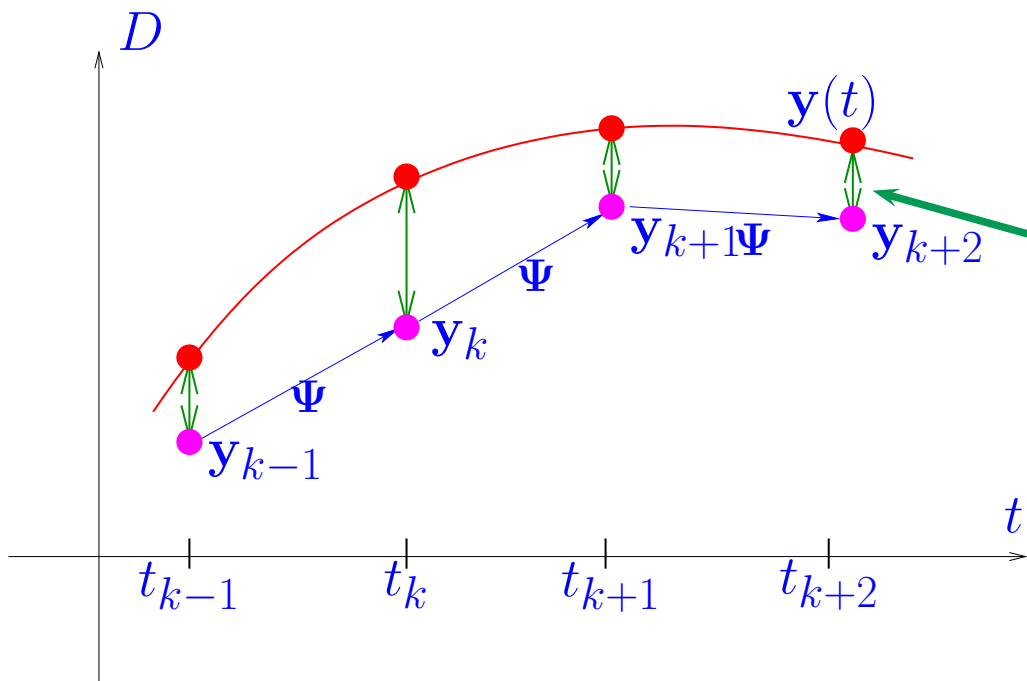
## 2.1.4 Das Äquivalenzprinzip (Dahlquist, Lax)

Ziel: Abstraktion des Beweises von Thm. 2.1.19

Betrachte: Äquidistante Zeitgitter  $\mathcal{G} = \{t_k\}_{k=0}^N, t_k := t_0 + hk, h := (T - t_0)/N, N \in \mathbb{N}$



(Annahme:  $\mathcal{G} \subset J(t_0, \mathbf{y}_0), \mathbf{y}_{\mathcal{G}}$  wohldefiniert)



**Konsistenzfehler**, vgl. Def. 2.1.11:

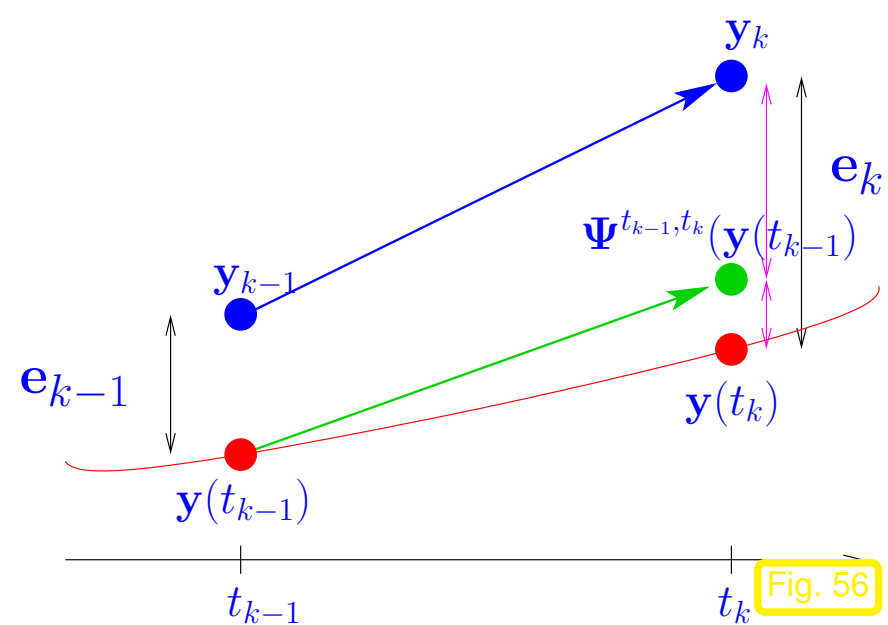
$$\tau(t, \mathbf{y}, h) := (\Phi^{t, t+h} \mathbf{y} - \Psi^{t, t+h} \mathbf{y}) .$$

**Fehlerfunktion:**  $\mathbf{e}_k := \mathbf{y}(t_k) - \mathbf{y}_k .$

- ◁ —  $\hat{=}$   $t \mapsto \mathbf{y}(t)$
- $\hat{=}$   $\mathbf{y}(t_k)$
- $\hat{=}$   $\mathbf{y}_k$
- $\hat{=}$   $\Psi^{t_k, t_{k+1}}$

## Fehlerrekursion

$$\begin{aligned}
 \mathbf{e}_k &= \mathbf{y}(t_k) - \mathbf{y}_k \\
 &= \mathbf{y}(t_k) - \Psi^{t_{k-1}, t_k}(\mathbf{y}(t_{k-1}) - \mathbf{e}_{k-1}) \\
 &= \underbrace{\Psi^{t_{k-1}, t_k}(\mathbf{y}(t_{k-1})) - \Psi^{t_{k-1}, t_k}(\mathbf{y}(t_{k-1}) - \mathbf{e}_{k-1})}_{\text{fortgeplanter Fehler}} \\
 &\quad + \underbrace{\mathbf{y}(t_k) - \Psi^{t_{k-1}, t_k}(\mathbf{y}(t_{k-1}))}_{\text{Einschrittfehler = Konsistenzfehler}}
 \end{aligned}$$



## Alternative Fehlerrekursion

$$\begin{aligned}
 \mathbf{e}_k &= \mathbf{y}(t_k) - \mathbf{y}_k = \Phi^{t_{k-1}, t_k} \mathbf{y}(t_{k-1}) - \Psi^{t_{k-1}, t_k}(\mathbf{y}_{k-1}) \\
 &= \underbrace{\Phi^{t_{k-1}, t_k}(\mathbf{y}_{k-1} + \mathbf{e}_{k-1}) - \Phi^{t_{k-1}, t_k}(\mathbf{y}_{k-1})}_{\text{fortgeplanter Fehler}} + \underbrace{\Phi^{t_{k-1}, t_k}(\mathbf{y}_{k-1}) - \Psi^{t_{k-1}, t_k}(\mathbf{y}_{k-1})}_{\text{Einschrittfehler = Konsistenzfehler}}
 \end{aligned}$$

1. Fehlerrekursion: Abschätzung des Konsistenzfehlers in einer Umgebung der exakten Lösung ausreichend
2. Fehlerrekursion: Konsistenzfehler abzuschätzen in einer Umgebung der Lösung des ESV

**Definition 2.1.25** (Nichtlineare Stabilität).

Eine diskrete Evolution  $\Psi$  ist *(nichtlinear) stabil*

$$:\Leftrightarrow \exists c > 0: \left\| \Psi^{t,t+h} \mathbf{y} - \Psi^{t,t+h} \mathbf{z} \right\| \leq (1 + ch) \|\mathbf{y} - \mathbf{z}\|$$

lokal gleichmässig in  $(t, \mathbf{y})$  für hinreichend kleine  $\|\mathbf{y} - \mathbf{z}\|$ ,  $h > 0$ .

Für ESV (2.1.17): Lokale Lipschitz-Stetigkeit von  $\psi$   $\triangleright$  nichtlineare Stabilität

**Theorem 2.1.26** ( Konsistenz & (nichtlineare) Stabilität  $\Rightarrow$  Konvergenz ).

Falls  $\Psi$  konsistent mit  $\Phi$  (von Ordnung  $p$ ) und (nichtlinear) stabil, so konvergiert das Einschrittverfahren global (von Ordnung  $p$ ).

## 2.1.5 Reversibilität

Wir haben gesehen: Eine approximative diskrete Evolution  $\Psi$  ( $\rightarrow$  Sect. 2.1.1) kann *im Allgemeinen nicht* erfüllen:  $\Psi^{r,s} \Psi^{t,r} = \Psi^{t,s}$

Jedoch: für  $s = t$  ist diese Forderung realisierbar !

**Definition 2.1.27** (Reversible diskrete Evolutionen).  $\rightarrow$  [8, Def. 4.40]

Eine diskrete Evolution  $\Psi : \tilde{\Omega}_h \subset I \times I \times D \mapsto \mathbb{R}^d$  (und das zugehörige Einschrittverfahren) heisst *reversibel*, falls

$$\Psi^{t,s} \Psi^{s,t} \mathbf{y} = \mathbf{y} \quad \forall (t, \mathbf{y}) \in \Omega, \quad \forall |t - s| \text{ hinreichend klein.}$$

*Beispiel 2.1.28* (Einfache reversible Einschrittverfahren).

- implizite Mittelpunktsregel (1.4.19)

$$\begin{aligned} \Psi^{t,t+h} \mathbf{y} &= \mathbf{y} + h \mathbf{f}\left(t + \frac{1}{2}h, \frac{1}{2}(\mathbf{y} + \Psi^{t,t+h} \mathbf{y})\right) \\ &\Downarrow \\ \mathbf{y} &= \Psi^{t,t+h} \mathbf{y} - h \mathbf{f}\left(t + h - \frac{1}{2}h, \frac{1}{2}(\mathbf{y} + \Psi^{t,t+h} \mathbf{y})\right). \end{aligned}$$

Unter der Annahme der Eindeutigen Auflösbarkeit der Definitionsgleichung (1.4.19) nach  $\mathbf{y}_{k+1}$ :

$$\Rightarrow \mathbf{y} = \Psi^{t+h,t} \Psi^{t,t+h} \mathbf{y}.$$

- Störmer-Verlet-Verfahren ( $\rightarrow$  Sect. 1.4.4) in Einschrittformulierung von Bem. 1.4.33

$$\begin{aligned} \mathbf{v}_{k+\frac{1}{2}} &= \mathbf{v}_k + \frac{h}{2}\mathbf{f}(\mathbf{y}_k), & \mathbf{v}_{k+\frac{1}{2}} &= \mathbf{v}_{k+1} - \frac{h}{2}\mathbf{f}(\mathbf{y}_{k+1}), \\ \mathbf{y}_{k+1} &= \mathbf{y}_k + h\mathbf{v}_{k+\frac{1}{2}}, & \Rightarrow \quad \mathbf{y}_k &= \mathbf{y}_{k+1} - h\mathbf{v}_{k+\frac{1}{2}}, \\ \mathbf{v}_{k+1} &= \mathbf{v}_{k+\frac{1}{2}} + \frac{h}{2}\mathbf{f}(\mathbf{y}_{k+1}). & \mathbf{v}_k &= \mathbf{v}_{k+\frac{1}{2}} - \frac{h}{2}\mathbf{f}(\mathbf{y}_k). \end{aligned}$$

Man erkennt Reversibilität an der Verfahrensvorschrift, wenn der Austausch  $\mathbf{y}_k \leftrightarrow \mathbf{y}_{k+1}$  und  $h \leftrightarrow -h$  Gleichungen liefert, die mit der ursprünglichen Verfahrensvorschrift identisch sind.



**Theorem 2.1.29** (Konsistenzordnung reversibler ESV).  $\rightarrow$  [8, Satz 4.42]

Die maximale Konsistenzordnung ( $\rightarrow$  Def. 2.1.13) eines reversiblen Einschrittverfahrens ( $\rightarrow$  Def. 2.1.27) ist gerade.

Der Beweis verwendet folgendes Hilfsresultat ([8, Lemma 4.38]):

**Lemma 2.1.30** (Störungslemma für diskrete Evolutionen).

Sei  $\mathbf{f}$  zweimal stetig differenzierbar und  $\Psi$  eine zur ODE  $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$  konsistente ( $\rightarrow$  Def. 2.1.8) diskrete Evolution, stetig differenzierbar in  $h$  und  $\mathbf{y}$ . Dann gilt für  $(t, \mathbf{y}) \in \Omega$  und hinreichend kleine  $\mathbf{z} \in \mathbb{R}^d$

$$\Psi^{t,t+h}(\mathbf{y} + \mathbf{z}) = \Psi^{t,t+h}\mathbf{y} + \mathbf{z} + h \frac{\partial \mathbf{f}}{\partial \mathbf{y}}(t, \mathbf{y})\mathbf{z} + \mathbf{r}(h, \mathbf{z}), \quad \|\mathbf{r}(h, \mathbf{z})\| \leq C(h^2 \|\mathbf{z}\| + h \|\mathbf{z}\|^2),$$

mit  $C > 0$  unabhängig von  $h$  und  $\mathbf{z}$ .

R. Hiptmair

rev 35327,  
25. April  
2011

☞ Thm. 2.1.29 erklärt in Bsp. 1.4.21 beobachtete  $O(h^2)$ -Kovergenz der impliziten Mittelpunktsregel.

## 2.2 Kollokationsverfahren[8, Sect. 6.3], [16, Sect. II.1.2]

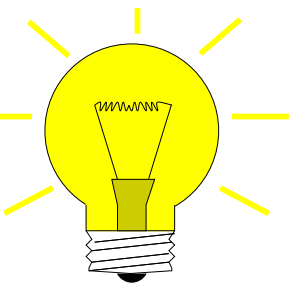
### 2.2.1 Konstruktion

Zunächst: Fokus auf ersten ( $\leftrightarrow$  allgemeinem) Schritt  
(dies genügt bei Einschrittverfahren  $\rightarrow$  Def. 2.1.2)

- Idee: ❶ Approximiere  $\mathbf{y}(t)$ ,  $t \in [t_0, t_1]$ , in  $s + 1$ -dimensionalem **Ansatzraum**  $V$  von Funktionen  $[t_0, t_1] \mapsto \mathbb{R}^d \triangleright \mathbf{y}_h$ .
- ❷ Festlegung von  $\mathbf{y}_h \in V$  durch **Kollokationsbedingungen**

$$\mathbf{y}_h(t_0) = \mathbf{y}_0 \quad , \quad \dot{\mathbf{y}}_h(\tau_j) = \mathbf{f}(\tau_j, \mathbf{y}_h(\tau_j)) \quad , \quad j = 1, \dots, s \quad , \quad (2.2.1)$$

für **Kollokationspunkte**  $t_0 \leq \tau_1 < \dots < \tau_s \leq t_1$ .



„Standardoption“:

Polynomialer Ansatzraum  $V = \mathcal{P}_s$



 Notation:  $\mathcal{P}_s \hat{=} \text{Raum der univariaten Polynome vom Grad } \leq s, s \in \mathbb{N}_0$

Bekannt:  $\dim \mathcal{P}_s = s + 1$

- Ein Polynom  $p \in \mathcal{P}_s$  ist durch  $s + 1$  **Interpolationsbedingungen** für Werte/Ableitungen eindeutig festgelegt.
- Kollokationsbedingungen (2.2.1) legen Polynomgrad  $s$  nahe (im Sinne von Existenz/Eindeutigkeit von  $\mathbf{y}_h$ )

R. Hiptmair

rev 35327,  
25. April  
2011

Herleitung: Formel für  $\mathbf{y}_h(t_1)$  ( $h := t_1 - t_0, \tau_j := t_0 + c_j h, 0 \leq c_1 < c_2 < \dots < c_s \leq 1$ )

Hilfsmittel:  $\{L_j\}_{j=1}^s \subset \mathcal{P}_{s-1} \hat{=} \text{Lagrange-Polynome}$  zu Stützstellen  $c_i, i = 1, \dots, s$ , in  $[0, 1]$ :

$$L_i(\tau) = \prod_{j=1, j \neq i}^s \frac{\tau - c_j}{c_i - c_j}, \quad i = 1, \dots, s \quad \Rightarrow \quad L_j(c_i) = \delta_{ij}, \quad i, j = 1, \dots, s. \quad (2.2.2)$$

$$(2.2.1) \Rightarrow \dot{\mathbf{y}}_h(t_0 + \tau h) = \sum_{j=1}^s \mathbf{k}_j L_j(\tau) \quad , \quad \mathbf{k}_j := \mathbf{f}(t_0 + c_j h, \mathbf{y}_h(t_0 + c_j h)) .$$

$$\Rightarrow \mathbf{y}_h(t_0 + \tau h) = \mathbf{y}_0 + h \sum_{j=1}^s \mathbf{k}_j \int_0^\tau L_j(\zeta) d\zeta .$$



Definierende Gleichungen des Kollokations-Einschrittverfahrens  
(zu Kollokationspunkten  $0 \leq c_1 < c_2 < \dots < c_s \leq 1$ ):

$$\mathbf{y}_h(t_1) = \mathbf{y}_0 + h \sum_{i=1}^s b_i \mathbf{k}_i , \quad \text{mit} \quad a_{ij} = \int_0^{c_i} L_j(\tau) d\tau , \quad (2.2.3)$$

$$\mathbf{k}_i = \mathbf{f}(t_0 + c_i h, \mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{k}_j) . \quad b_i = \int_0^1 L_i(\tau) d\tau .$$



**Diskrete Evolution**  $\Psi^{t_0, t_1} : \Omega \mapsto \Omega$  ,  $\Psi^{t_0, t_1} \mathbf{y}_0 := \mathbf{y}_1 := \mathbf{y}_h(t_1)$

➤ (2.2.3)  $\hat{=}$  (Nichtlineares) Gleichungssystem für **Inkrement**  $\mathbf{k}_i$  ( $\approx \dot{\mathbf{y}}(t_0 + c_i h)$ )

▷ (Generische) Kollokationsverfahren = *implizites* Einschrittverfahren ( $\rightarrow$  Def. 2.1.2)

Kollokations-Einschrittverfahren in der Form von Lemma 2.1.9:

$$\Psi^{t_0, t_0+h} \mathbf{y}_0 = \mathbf{y}_0 + h \psi(t_0, \mathbf{y}_0, h) \quad \text{mit Inkrementfunktion} \quad \psi(t_0, \mathbf{y}_0, h) = \sum_{i=1}^s b_i \mathbf{k}_i . \quad (2.2.4)$$

*Bemerkung 2.2.5* (Umformulierung der Inkrementgleichungen (2.2.3)).

Äquivalente Form der Inkrementgleichungen (2.2.3):

Ersetze Inkremente  $\mathbf{k}_i$  durch

$$\mathbf{g}_i := \mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{k}_j, \quad i = 1, \dots, s \quad \Leftrightarrow \quad \mathbf{k}_i = \mathbf{f}(t_0 + c_i h, \mathbf{g}_i).$$

$$(2.2.3) \Leftrightarrow \begin{aligned} \mathbf{g}_i &= \mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{f}(t_0 + c_j h, \mathbf{g}_j) \\ \mathbf{y}_1 &= \mathbf{y}_0 + h \sum_{i=1}^s b_i \mathbf{f}(t_0 + c_i h, \mathbf{g}_i) . \end{aligned} \quad (2.2.6)$$



**Lemma 2.2.7** (Lösbarkeit der Inkrementgleichungen).

Ist  $\mathbf{f}$  lokal Lipschitz-stetig ( $\rightarrow$  Def. 1.3.2) auf dem erweiterten Zustandsraum  $\Omega$ , so gibt es zu jedem  $(t_0, \mathbf{y}_0) \in \Omega$  ein  $h_0 > 0$  so, dass (2.2.3) für jedes  $h < h_0$  eindeutig nach den **Inkrementen**  $\mathbf{k}_i$  auflösbar ist, und diese sind stetige Funktionen in  $h$ .

Ist  $f \in C^m(\Omega, \mathbb{R}^d)$ ,  $m \in \mathbb{N}$ , dann sind auch die Inkremente  $m$ -fach stetig differenzierbare Funktionen von  $\mathbf{y}_0, t_0, h$ .

„Beweis“ (von Lemma 2.2.7, unter stärkeren Glattheitsvoraussetzungen, hier nur ausgeführt für autonomen Fall  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ )

**Annahme:**

$\mathbf{f}$  ist stetig differenzierbar auf  $\Omega$

$$\mathbf{k}_i = \mathbf{f}\left(\mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{k}_j\right) \Leftrightarrow G(h, \mathbf{k}) = 0, \quad G(h, \mathbf{k}) := \mathbf{k} - \begin{pmatrix} \mathbf{f}\left(\mathbf{y}_0 + h \sum_{j=1}^s a_{1j} \mathbf{k}_j\right) \\ \vdots \\ \mathbf{f}\left(\mathbf{y}_0 + h \sum_{j=1}^s a_{sj} \mathbf{k}_j\right) \end{pmatrix},$$

mit  $\mathbf{k} = (\mathbf{k}_1, \dots, \mathbf{k}_s)^T \in \mathbb{R}^{s \cdot d}$ .

Idee: Anwendung des **Satzes über implizite Funktionen** auf  $G : \mathbb{R} \times D \mapsto D$

R. Hiptmair

rev 35327,  
25. April  
2011

**Theorem 2.2.8** (Satz über implizite Funktionen).  $\rightarrow$  Analysis-Vorlesung

Seien  $I \subset \mathbb{R}^q$ ,  $U \subset \mathbb{R}^n$  offen und  $G = G(\xi, y) : I \times U \mapsto \mathbb{R}^n$  sei stetig differenzierbar. Für ein  $(\xi_0, y_0) \in I \times U$  gelte  $G(\xi_0, y_0) = 0$ .

Ist die Jacobi-Matrix  $\frac{\partial G}{\partial y}(\xi_0, y_0)$  invertierbar, dann gibt es eine Umgebung  $V \subset U$  von  $y_0$  und eine eindeutige stetig differenzierbare Funktion  $\xi \mapsto z(\xi)$  so, dass

- $\xi_0 := (\mathbf{f}(\mathbf{y}_0), \dots, \mathbf{f}(\mathbf{y}_0))^T$  erfüllt  $G(0, \xi_0) = 0$
- Ableitung ( $\hat{=}$  Jacobi-Matrix) von  $G$  in  $(0, \xi_0)$  (aus Kettenregel)

$$D_{\xi}G(0, \xi_0) = \mathbf{I}$$

ist die Einheitsmatrix und damit offensichtlich invertierbar. □

Ein alternativer, technisch aufwändigerer, Beweis erfordert bloss lokale Lipschitz-Stetigkeit von  $\mathbf{f}$  und gibt zusätzliche Schrittweitschranke für die Existenz einer Lösung der Inkrementgleichungen:

Hilfsmittel bei alternativem Beweis ( $\rightarrow$  Analysis-Vorlesung):

**Theorem 2.2.9** (Banachscher Fixpunktsatz, parameterabhängige Version).

$V \subset \mathbb{R}^d$  abgeschlossen,  $U \subset \mathbb{R}^n$  offen,  $F : U \times V \mapsto V$  sei total  $m$ -mal stetig differenzierbar,  $m \in \mathbb{N}_0$ , und besitze die **gleichmässige Kontraktionseigenschaft**

$$\exists 0 \leq q < 1: \quad \|F(\mathbf{u}, \mathbf{z}) - F(\mathbf{u}, \mathbf{w})\| \leq q \|\mathbf{z} - \mathbf{w}\| \quad \forall \mathbf{z}, \mathbf{w} \in V, \quad \forall \mathbf{u} \in U.$$

Dann gibt es eine  $m$ -mal stetig differenzierbare Funktion  $G : U \mapsto V$  so dass

$$F(\mathbf{u}, G(\mathbf{u})) = G(\mathbf{u}) \quad \forall \mathbf{u} \in U$$

✎ Übliche Notation für Koeffizientenmatrix  $\mathfrak{A} := (a_{ij})_{i,j=1}^s \in \mathbb{R}^{s,s}$

✎ Zeilensummennorm  $\|\mathfrak{A}\|_\infty := \max_{i=1,\dots,s} \sum_{j=1}^s |a_{ij}|$  ( $\hat{=}$  Matrixnorm zur Maximumnorm)

*Beweis.* (von Lemma 2.2.7 für autonomen Fall  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ )

Vorbereitung: Wie im Beweis von Thm. 2.1.19 betrachten wir  $\mathbf{f}$  wieder auf einer kompakten Umgebung  $K_\delta$  der Lösungskurve  $t \mapsto \mathbf{y}(t)$  im erweiterten Phasenraum  $\Omega$ . Daher (zunächst) ohne Beschränkung der Allgemeinheit die *Annahme*:

$\mathbf{f}$  global Lipschitz-stetig, vgl. Def. 1.3.2:

$$\exists L > 0: \quad \|\mathbf{f}(\mathbf{z}) - \mathbf{f}(\mathbf{w})\| \leq L \|\mathbf{z} - \mathbf{w}\| \quad \forall \mathbf{z}, \mathbf{w} \in D. \quad (2.2.10)$$

Wir nehmen auch an, dass sich eine  $r$ -Umgebung von  $\mathbf{y}_0$  in  $D$  befindet:

$$\exists r > 0: \quad \|\mathbf{z} - \mathbf{y}_0\| \leq r \Rightarrow \mathbf{z} \in D.$$

Idee: Anwendung des Banachschen Fixpunktsatzes Thm. 2.2.9 auf die (äquivalenten) Inkrementgleichungen (2.2.6) für die  $\mathbf{g}_i$ : mit  $\mathbf{g} := (\mathbf{g}_1, \dots, \mathbf{g}_s) \in \mathbb{R}^{s \cdot d}$

$$(2.2.6) \Leftrightarrow \mathbf{g} = F(h, \mathbf{g}), \quad F(h, \mathbf{g}) := \begin{pmatrix} \mathbf{y}_0 + h \sum_{j=1}^s a_{1j} \mathbf{f}(\mathbf{g}_j) \\ \vdots \\ \mathbf{y}_0 + h \sum_{j=1}^s a_{sj} \mathbf{f}(\mathbf{g}_j) \end{pmatrix}. \quad (2.2.11)$$

Auf  $\mathbb{R}^{s \cdot d}$  verwende Norm  $\|\mathbf{g}\| := \max_{i=1, \dots, s} \|\mathbf{g}_i\|$ .

Zu zeigen: für hinreichend kleines  $h$  bleiben alle  $\mathbf{g}_i$  in der *abgeschlossenen*  $r$ -Umgebung von  $\mathbf{y}_0$ : mit  $\boldsymbol{\eta}_0 = (\mathbf{y}_0, \dots, \mathbf{y}_0)$

$$\begin{aligned} \|F(h, \mathbf{g}) - \boldsymbol{\eta}_0\| &= \max_{i=1, \dots, s} |h| \left\| \sum_{j=1}^s a_{ij} \mathbf{f}(\mathbf{g}_j) \right\| \leq |h| \|\mathfrak{A}\|_\infty \max_{j=1, \dots, s} \|f(\mathbf{y}_0) + f(\mathbf{g}_j) - f(\mathbf{y}_0)\| \\ (2.2.10) \quad &\leq |h| \|\mathfrak{A}\|_\infty (\|\mathbf{f}(\mathbf{y}_0)\| + L \|\mathbf{g} - \boldsymbol{\eta}_0\|). \end{aligned}$$

$$\Rightarrow \left\{ |h| < \frac{r}{\|\mathfrak{A}\|_\infty (\|\mathbf{f}(\mathbf{y}_0)\| + Lr)} \Rightarrow \|F(h, \mathbf{g}) - \boldsymbol{\eta}_0\| \leq r, \text{ if } \|\mathbf{g} - \boldsymbol{\eta}_0\| \leq r \right\}.$$

Zu zeigen:  $\mathbf{g} \mapsto F(h, \mathbf{g})$  ist  $h$ -gleichmässige Kontraktion

$$\|F(h, \mathbf{g}) - F(h, \mathbf{p})\| \leq |h| \cdot \max_{i=1, \dots, s} \left\| \sum_{j=1}^s a_{ij} (\mathbf{f}(\mathbf{g}_j) - \mathbf{f}(\mathbf{p}_j)) \right\|$$



$$\begin{aligned} &\leq |h| \cdot \|\mathfrak{A}\|_\infty \max_{j=1,\dots,s} \|\mathbf{f}(\mathbf{g}_j) - \mathbf{f}(\mathbf{p}_j)\| \\ &\leq |h|L \cdot \|\mathfrak{A}\|_\infty \|\mathbf{g} - \mathbf{p}\| , \end{aligned}$$

wobei im letzten Schritt die globale Lipschitz-Bedingung für  $\mathbf{f}$  benutzt wurde.

$$|h| < \frac{1}{L \|\mathfrak{A}\|_\infty} \Rightarrow \mathbf{g} \mapsto F(h, \mathbf{g}) \text{ ist } h\text{-gleichmässige Kontraktion.}$$

Wähle

$$h_0 = \frac{r}{\|\mathfrak{A}\|_\infty (\|\mathbf{f}(\mathbf{y}_0)\| + Lr)}$$

Damit erfüllt  $F$  mit  $U = ]-h_0, h_0[$  und  $V = \{\mathbf{g} : \|\mathbf{g} - \mathbf{y}_0\| \leq r\}$  die Voraussetzungen des Banachschen Fixpunktsatzes Thm. 2.2.9.

$$\Rightarrow \left\{ \mathbf{f} \in C^m(D) \Rightarrow \exists \mathbf{g} : ]-h_0, h_0[ \mapsto \mathbb{R}^{d \cdot s} : F(h, \mathbf{g}(h)) = \mathbf{g}(h) \right\} .$$

Wegen der Äquivalenz (2.2.11) ist damit der Beweis abgeschlossen.  $\square$

*Bemerkung 2.2.12* (Schrittweitenbeschränkung aus Lemma 2.2.7).

Aus dem Beweis von Lemma 2.2.7 mit Hilfe des Fixpunktarguments, Thm 2.2.9: Lösbarkeit der Inkre-

mentgleichungen nur garantiert, wenn

$$|h| \leq \frac{1}{L \|\mathfrak{A}\|_\infty},$$

wobei  $L > 0$  eine (lokale) Lipschitz-Konstante ( $\rightarrow$  Def. 1.3.2) für den Quellterm  $\mathbf{f}$  ist.

Dies ist eine **Schrittweitenbeschränkung** analog der Schrittweitenbeschränkung für das explizite Euler-Verfahren in der Nähe stark attraktiver Fixpunkte, vgl. Bsp. 1.4.9.



Es bleibt noch die Verifikation einer Voraussetzung von Thm. 2.1.19:

**Lemma 2.2.13** (Lipschitz-Stetigkeit der Inkrementfunktion).

*Unter den Voraussetzungen von Lemma 2.2.7 existiert zu jedem  $(t_0, \mathbf{y}_0) \in \Omega$  ein  $h_0 > 0$  so, dass  $\psi$  aus (2.2.4) lokal Lipschitz-stetig im Zustandsargument ist.*

*Beweis* auf der Grundlage des Satzes über implizite Funktionen, Thm. 2.2.8, unter Annahme von hinreichender Glattheit von  $\mathbf{f}$ :

Wie im ersten Beweis zu Lemma 2.2.7 Umformulierung der Inkrementgleichungen als parameterabhängiges Nullstellenproblem:

$$\mathbf{k}_i = \mathbf{f}\left(\mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{k}_j\right) \Leftrightarrow G(h, \mathbf{y}_0, \boldsymbol{\xi}) = 0, \quad G(h, \mathbf{y}_0, \boldsymbol{\xi}) := \boldsymbol{\xi} - \begin{pmatrix} \mathbf{f}\left(\mathbf{y}_0 + h \sum_{j=1}^s a_{1j} \mathbf{k}_j\right) \\ \vdots \\ \mathbf{f}\left(\mathbf{y}_0 + h \sum_{j=1}^s a_{sj} \mathbf{k}_j\right) \end{pmatrix},$$

Wir haben

- $G$  ist stetig differenzierbar in allen Argumenten, falls  $\mathbf{f}$  hinreichend glatt.
- $G(0, \mathbf{y}_0, \boldsymbol{\xi}) = 0$  für  $\boldsymbol{\xi} = (\mathbf{f}(\mathbf{y}_0), \dots, \mathbf{f}(\mathbf{y}_0)) \in \mathbb{R}^{s \cdot d}$
- $D_{\boldsymbol{\xi}} G(0, \mathbf{y}_0, \boldsymbol{\xi}) = \mathbf{I}$  für beliebiges  $\boldsymbol{\xi} \in \mathbb{R}^{s \cdot d}$ ,  $\mathbf{y}_0 \in D$ .

Nach dem Satz über implizite Funktionen gibt es also eine lokal stetig differenzierbare Lösungskurve  $\boldsymbol{\xi} = \boldsymbol{\xi}(\mathbf{y}, h)$ , definiert in einer Umgebung von  $(0, \mathbf{y}_0)$ . Daher folgt die Behauptung aus einem Analogon von Lemma 1.3.3. □

*Beweis* von Lemma 2.2.13 mit **Fixpunktargument**, ohne Glattheitsanforderungen an  $\mathbf{f}$  (für autonome ODE):

Ohne Beschränkung der Allgemeinheit wird  $\mathbf{f}$  als **global Lipschitz-stetig** angenommen, vgl. Beweis zu Lemma 2.2.7:

$$\exists L > 0: \quad \|\mathbf{f}(\mathbf{z}) - \mathbf{f}(\mathbf{w})\| \leq L \|\mathbf{z} - \mathbf{w}\| \quad \forall \mathbf{z}, \mathbf{w} \in D. \quad (2.2.14)$$

Mit  $\mathbf{g}_i$  aus den äquivalenten Inkrementgleichungen (2.2.6):

$$\text{Inkrementfunktion: } \psi(t, \mathbf{y}, h) = \mathbf{y} + h \sum_{j=1}^s b_j \mathbf{f}(\mathbf{g}_j) \quad , \quad \mathbf{g}_i = \mathbf{y} + h \sum_{j=1}^s a_{ij} \mathbf{f}(\mathbf{g}_j). \quad (2.2.15)$$

Wähle  $\mathbf{y}, \mathbf{z} \in D$  und definiere (für hinreichend kleines  $h$ , siehe Lemma 2.2.7)  $\mathbf{g}_i^y, \mathbf{g}_i^z \in \mathbb{R}^d$  als Lösungen von

$$\begin{aligned} \mathbf{g}_i^y &= \mathbf{y} + \sum_{j=1}^s a_{ij} \mathbf{f}(\mathbf{g}_j^y) , \\ \mathbf{g}_i^z &= \mathbf{z} + \sum_{j=1}^s a_{ij} \mathbf{f}(\mathbf{g}_j^z) . \end{aligned} \quad , \quad i = 1, \dots, s .$$

Mit  $\mathbf{g}^y := (\mathbf{g}_1^y, \dots, \mathbf{g}_s^y)$ ,  $\mathbf{g}^z := (\mathbf{g}_1^z, \dots, \mathbf{g}_s^z)$ :

$$\|\mathbf{g}^y - \mathbf{g}^z\| \leq \|\mathbf{y} - \mathbf{z}\| + h \max_{i=1, \dots, s} \sum_{j=1}^s |a_{ij}| \left( \|\mathbf{f}(\mathbf{g}_j^y)\| - \|\mathbf{f}(\mathbf{g}_j^z)\| \right)$$

$$(2.2.14) \quad \leq \|\mathbf{y} - \mathbf{z}\| + hL \cdot \|\mathfrak{A}\|_\infty \|\mathbf{g}^y - \mathbf{g}^z\| .$$

$$h \|\mathfrak{A}\|_\infty L < 1 \quad \Rightarrow \quad \|\mathbf{g}^y - \mathbf{g}^z\| \leq \frac{1}{1 - h \|\mathfrak{A}\|_\infty L} \|\mathbf{y} - \mathbf{z}\| .$$

vgl. die Schrittweitschranke aus Bem. 2.2.12

Aus dieser Abschätzung und wieder mit (2.2.14) folgt (falls  $h \|\mathfrak{A}\|_\infty L < 1$ )

$$\|\psi(t, \mathbf{y}, h) - \psi(t, \mathbf{z}, h)\| \leq h \sum_{i=1}^s |b_i| \|\mathbf{f}(\mathbf{g}_i^y) - \mathbf{f}(\mathbf{g}_i^z)\| \leq \frac{L}{1 - Lh \|\mathfrak{A}\|_\infty} \sum_{i=1}^s |b_i| \cdot \|\mathbf{y} - \mathbf{z}\| .$$

(2.2.16)

Da  $\mathbf{y}$ ,  $\mathbf{z}$  beliebig, folgt die Behauptung. □

*Beispiel 2.2.17* (Konvergenz von einfachen Kollokations-Einschrittverfahren).

- Skalare logistische Differentialgleichung (1.2.2),  $\lambda = 10$ ,  $y(0) = 0.01$ ,  $T = 1$
- Kollokations-Einschrittverfahren (2.2.3) für  $s = 1, \dots, 4$ , uniforme Zeitschrittweite  $h$

## Äquidistante Kollokationspunkte:

$$s = 1 : \mathbf{c} = \left(\frac{1}{2}\right),$$

$$s = 2 : \mathbf{c} = \left(\frac{1}{3}, \frac{2}{3}\right)^T,$$

$$s = 3 : \mathbf{c} = \left(\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\right)^T;$$

$$s = 4 : \mathbf{c} = \left(\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}\right)^T.$$

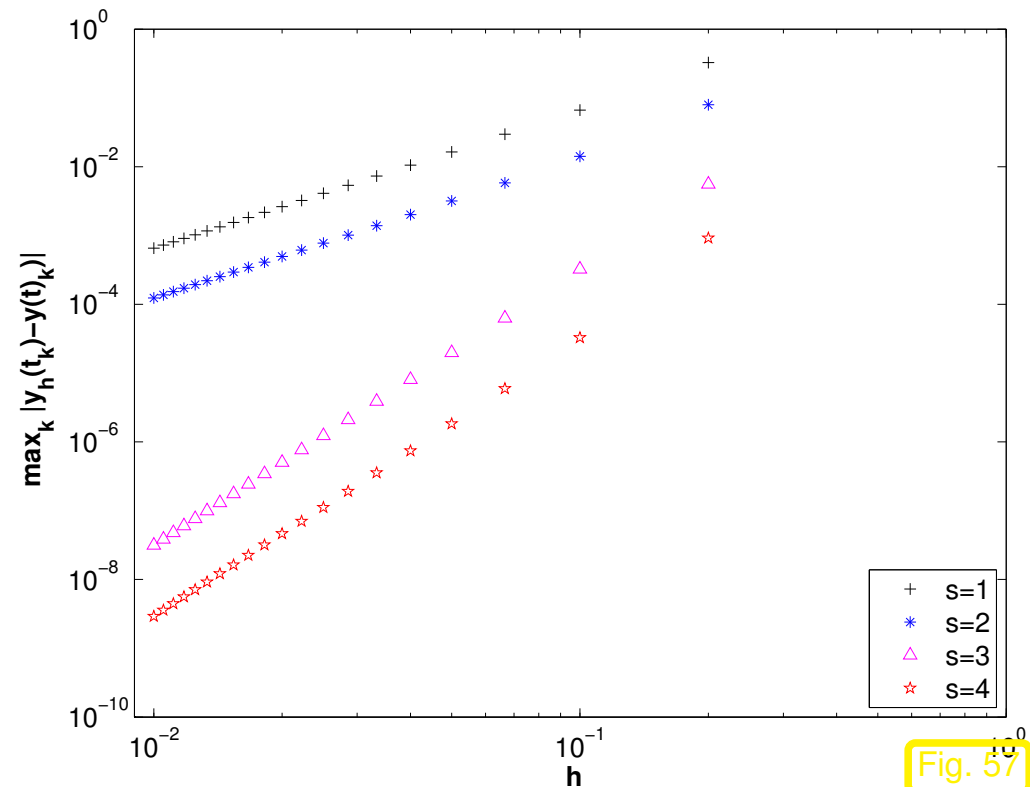
Numerische Konvergenzraten :  
(berechnet durch lineare Regression)

$$s = 1 : p = 1.96$$

$$s = 2 : p = 2.03$$

$$s = 3 : p = 4.00$$

$$s = 4 : p = 4.04$$

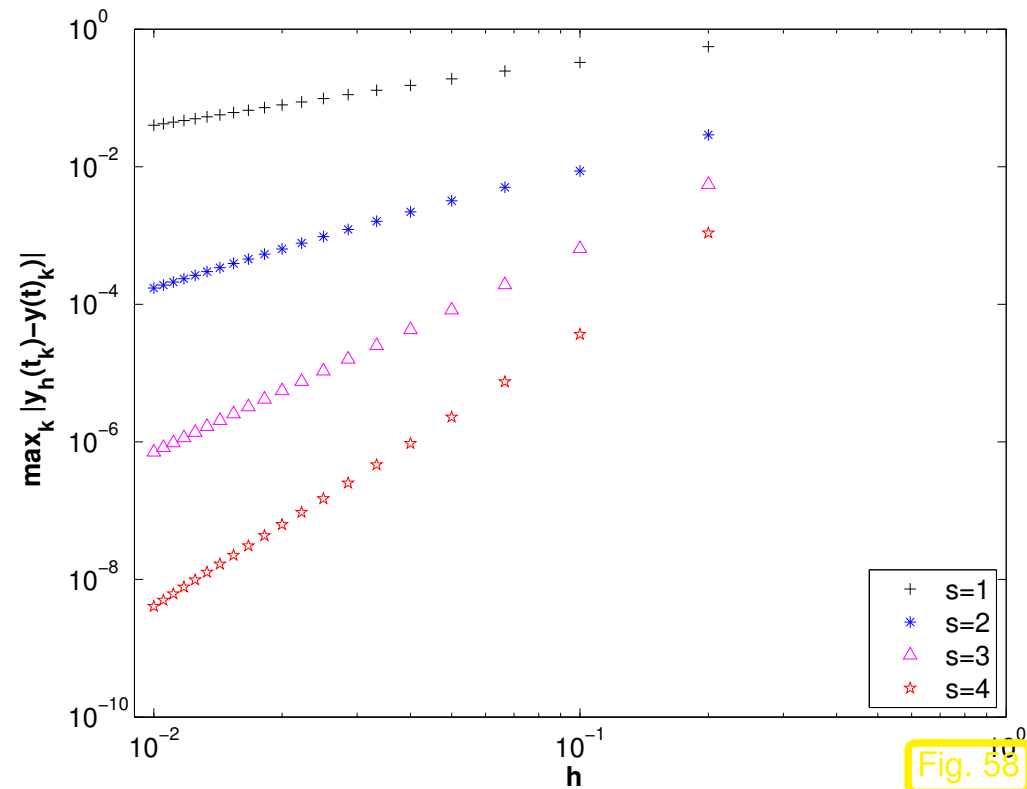


# Verschobene äquidistante Kollokationspunkte:

$$\begin{aligned}
 s = 1 & : \mathbf{c} = (0) , \\
 s = 2 & : \mathbf{c} = (0, \frac{1}{2})^T , \\
 s = 3 & : \mathbf{c} = (0, \frac{1}{3}, \frac{2}{3})^T ; , \\
 s = 4 & : \mathbf{c} = (0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4})^T .
 \end{aligned}$$

Numerische Konvergenzraten :  
(berechnet durch lineare Regression)

$$\begin{aligned}
 s = 1 & : p = 0.95 \\
 s = 2 & : p = 1.77 \\
 s = 3 & : p = 2.95 \\
 s = 4 & : p = 3.92
 \end{aligned}$$



Beobachtung bei *symmetrisch* gelegenen äquidistanten Kollokationspunkten:  
 Algebraische Konvergenz der Ordnung  $\begin{cases} p = s + 1 & , \text{ falls } s \text{ ungerade} , \\ p = s & , \text{ falls } s \text{ gerade.} \end{cases}$   
 Erklärung → Sect. 2.2.3 & Thm. 2.1.29



**Bemerkung 2.2.18** (Kollokationsverfahren und numerische Quadratur).

$\mathbf{f}(t, \mathbf{y}) = \mathbf{f}(t)$  &  $\mathbf{y}_0 = 0$  ➤ Numerische Quadratur (→ Vorlesung „Numerische Methoden“)

$$\mathbf{y}(t_1) = \int_{t_0}^{t_1} \mathbf{f}(t) dt \approx h \sum_{i=1}^s b_j \mathbf{f}(t_0 + c_j h) = \text{Quadraturformel}$$

→  $c_1, \dots, c_s \leftrightarrow$  **Knoten** (engl. *nodes*) einer Quadraturformel (z.B. Gauss-Punkte auf  $[0, 1]$ )

$b_1, \dots, b_s \leftrightarrow$  **Gewichte** (engl. *weights*) einer Quadraturformel



Aus Zusammenhang zwischen Kollokationsverfahren und numerische Quadratur

➤ Wahl der Kollokationspunkte  $c_i$  als Knoten bewährter Quadraturformeln auf  $[0, 1]$

Die folgenden Beispiele zeigen, dass sich sinnvolle Verfahren ergeben:



- Fall  $s = 1$  &  $c_1 = 1/2$  ( $\Leftrightarrow$  einfachste Gauss-Legendre-Quadraturformel)

$$L_1 \equiv 1 \Rightarrow a_{11} = 1/2, \quad b_1 = 1.$$

$$\mathbf{k}_1 = \mathbf{f}(t_0 + 1/2h, \mathbf{y}_0 + 1/2h\mathbf{k}_1) \quad , \quad \mathbf{y}_h(t_1) = \mathbf{y}_0 + h\mathbf{k}_1. \quad (2.2.19)$$

(2.2.19) = Implizite Mittelpunktsregel (1.4.19)

- Fall  $s = 1$  &  $c_1 = 0$  ( $\Leftrightarrow$  linksseitige Ein-Punkt-Quadraturformel)

$$L_1 \equiv 1 \Rightarrow a_{11} = 0, \quad b_1 = 1.$$

$$\mathbf{k}_1 = \mathbf{f}(t_0, \mathbf{y}_0) \quad , \quad \mathbf{y}_h(t_1) = \mathbf{y}_0 + h\mathbf{k}_1 = \mathbf{y}_0 + h\mathbf{f}(t_0, \mathbf{y}_0).$$

(2.2.1) = Explizites Eulerverfahren (1.4.2) (kein Lösen einer Gleichung erforderlich !)

- Fall  $s = 1$  &  $c_1 = 1$  ( $\Leftrightarrow$  rechtsseitige Ein-Punkt-Quadraturformel)

$$L_1 \equiv 1 \Rightarrow a_{11} = 1, \quad b_1 = 1.$$

$$\mathbf{k}_1 = \mathbf{f}(t_1, \mathbf{y}_0 + h\mathbf{k}_1) \quad , \quad \mathbf{y}_h(t_1) = \mathbf{y}_0 + h\mathbf{k}_1 = \mathbf{y}_0 + h\mathbf{f}(t_1, \mathbf{y}_h(t_1)).$$

(2.2.1) = Implizites Eulerverfahren

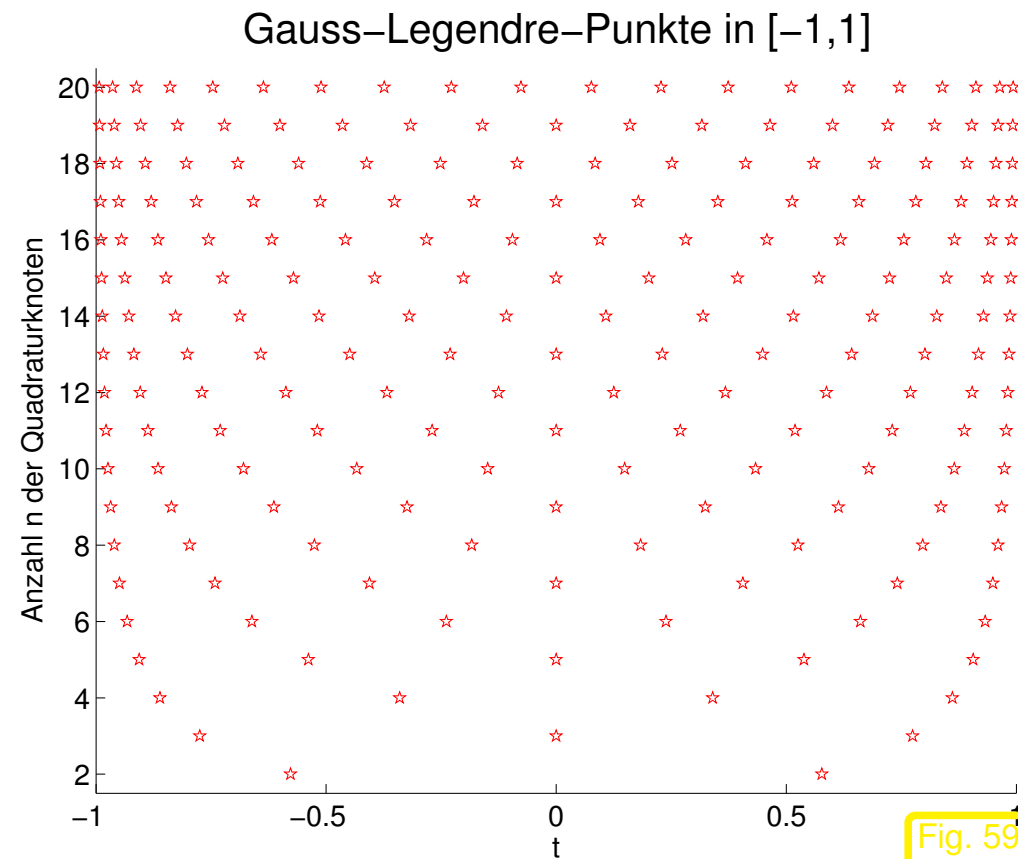
Erinnerung:

„Optimale Quadraturverfahren“: **Gaussquadratur**  
( $\rightarrow$  Vorlesung „Numerische Methoden“ [9, Sect. 9.3])

☞ Die  $n$  **Knoten** der  $n$ . Gaussischen Quadraturformel auf  $[-1, 1]$ ,  $n \in \mathbb{N}$ , sind die Nullstellen des **Legendre-Polynoms** vom Grad  $n$ . ▷

☞ Die  $n$ . Gaussische Quadraturformel hat **Ordnung  $2n$** .

☞ Die **Gewichte** der  $n$ . Gaussischen Quadraturformel sind positiv.



## Gauss-Kollokations-Einschrittverfahren

*Bemerkung 2.2.20* (Lösungsfunktion aus Kollokationsverfahren).

▷ Kollokationsverfahren liefert (per constructionem) sogar *stückweise polynomiale* approximative Lösung  $\mathbf{y}_h \in C^0([t_0, T])$

## 2.2.2 Abstrakte Projektionsverfahren

Ziel dieses Abschnitts ist es, die Kollokationsverfahren in eine grössere Klasse von Einschrittverfahren einzuordnen, die eine elegante abstrakte Konvergenztheorie zulässt.

*Bemerkung 2.2.21* (Kollokationsverfahren als Projektionsverfahren).

$P_s : C^0([t_0, t_1]) \mapsto \mathcal{P}_{s-1} \hat{=}$       Polynominterpolationsoperator zu Knoten  $\tau_1 \leq \dots \leq \tau_s$   
(vgl. Sect. 2.2.1, „Kollokationspunkte“)

Damit lassen sich die Kollokationsbedingungen kompakt umformulieren:

$$\Rightarrow \left( (2.2.1) \Leftrightarrow \dot{\mathbf{y}}_h = P_s \mathbf{f}(\cdot, \mathbf{y}_h(\cdot)) \quad , \quad \mathbf{y}_h(t_0) = \mathbf{y}_0 \cdot \right)$$

Beachte:

**Projektoreigenschaft**  $P_s^2 = P_s$

Bekannt aus der linearen Algebra:

**Definition 2.2.22** (Projektionsoperator).

Seien  $X$  ein Vektorraum. Eine lineare Abbildung  $P : X \mapsto X$  ist ein **Projektionsoperator**, falls  $P^2 = P$ .

Bekannt aus der Analysis:

**Definition 2.2.23** (Stetiger linearer Operator).

Seien  $X, Y$  normierte Vektorräume. Ein linearer Operator  $T : X \mapsto Y$  heisst **stetig/beschränkt**, falls

$$\|T\| := \sup_{x \in X \setminus \{0\}} \frac{\|Tx\|_Y}{\|x\|_X} < \infty .$$

$\|T\|$  heisst die **Norm** des stetigen Operators  $T$ .

Betrachte: ODE  $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$ ,  $\mathbf{f} : I \times D \mapsto \mathbb{R}^d$  lokale Lipschitz-stetig, siehe Sect. 1.1  
 Zugehörige AWPe  $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$ ,  $\mathbf{y}(t_0) = \mathbf{y}_0$ , auf  $[t_0, T] \in J(t_0, \mathbf{y}_0)$

► Verallgemeinerung : **Projektions-Einschrittverfahren**  $\Psi^{t,t+h}$  zu ODE  $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$   
 von Kollokations-ESV  
 diskrete Evolution des ESV

$\Psi^{t,t+h} \mathbf{y}_0$  definiert durch

- endlichdimensionalen **Ansatzraum**

$$V \subset (C^1([t, t+h]))^d \quad \mapsto \quad W := \left\{ \frac{d}{dt} \mathbf{v} : \mathbf{v} \in V \right\}$$

- stetigen Projektionsoperator  $\mathbf{P} : (C^0([t, t+h]))^d \mapsto W$

$$\Psi^{t,t+h} \mathbf{y}_0 := \mathbf{y}_h(t+h) \quad \text{mit} \quad \mathbf{y}_h \in V \quad \wedge \quad \begin{cases} \dot{\mathbf{y}}_h = \mathbf{P}(\mathbf{f}(\cdot, \mathbf{y}_h(\cdot))) \\ \mathbf{y}_h(t) = \mathbf{y}_0 \in D \end{cases}, \quad (2.2.24)$$

interpretiert als Funktion  $\in (C^0([t, t+h]))^d$

*Bemerkung 2.2.25* (Fixpunktform von Projektions-Einschrittverfahren).

Verallgemeinerung: (2.2.24)  $\longrightarrow$  **Fixpunktgleichung**,

$$(2.2.24) \Rightarrow \mathbf{y}_h(\tau) = \mathbf{y}_0 + \int_t^\tau \mathbf{P}(\mathbf{f}(\cdot, \mathbf{y}_h(\cdot))) (\xi) d\xi, \quad t \leq \tau \leq t+h. \quad (2.2.26)$$

! geringere Glattheitsanforderungen an  $V$ : (2.2.26) sinnvoll für  $\mathbf{y}_h \in (C^0([t, t+h]))^d$ .



**Lemma 2.2.27** (Wohldefiniertheit der diskreten Evolution für Projektions-Einschrittverfahren).

Erfüllt  $\mathbf{f}$  eine lokale Lipschitz-Bedingung ( $\rightarrow$  Def. 1.3.2), dann ist  $\Psi^{t,t+h} \mathbf{y}_0$  für hinreichend kleines  $h$  wohldefiniert.



Notation: Maximumnorm  $\|\mathbf{y}(\cdot)\|_{\infty, I} := \max_{\tau \in I} \|\mathbf{y}(\tau)\|,$

speziell im Folgenden:  $\|\mathbf{y}(\cdot)\|_{\infty} := \max_{t \leq \tau \leq t+h} \|\mathbf{y}(\tau)\|$

*Beweis.* (Analog zum Beweis von Lemma 2.2.7, Kontraktionsargument)

Wir müssen die eindeutige Lösbarkeit der Definitionsgleichung (2.2.24) für die Funktion  $\mathbf{y}_h$  zeigen.

Technik: Banachscher Fixpunktsatz Thm. 2.2.9 angewandt auf **Fixpunktgleichung** (2.2.26), vgl. Beweis von Thm. 1.3.4.

$$\mathbf{y}_h(\tau) = F(\mathbf{y}_h), \quad F(\mathbf{y}_h)(\tau) := \mathbf{y}_0 + \int_t^\tau \mathbf{P}(\mathbf{f}(\cdot, \mathbf{y}_h(\cdot)))(\xi) \, d\xi, \quad t \leq \tau \leq t + h. \quad (2.2.28)$$

Beachte: Abbildungseigenschaft  $F : (C^0([t, t + h]))^d \mapsto (C^0([t, t + h]))^d$

Erinnerung an Analysis:  $\triangleright (C^0([t, t + h]), \|\cdot\|_\infty)$  ist **Banachraum** !

Lokale Lipschitz-Bedingung & Kompaktheitsargument, vgl. Beweis von Thm. 2.1.19

$$\exists L > 0: \quad \|\mathbf{f}(\tau, \mathbf{z}) - \mathbf{f}(\tau, \mathbf{w})\| \leq L \|\mathbf{z} - \mathbf{w}\| \quad \forall t \leq \tau \leq t + h, \quad \forall \mathbf{z}, \mathbf{w} \in K \subset D, \quad (2.2.29)$$

mit Kompaktum  $K \subset D$ , für das (rückblickend) angenommen werden kann, dass  $\mathbf{y}_h(\tau) \in K$  für alle  $t \leq \tau \leq t + h$ . Dann für alle  $\mathbf{y}, \mathbf{z} \in (C^0([t, t + h]))^d$ ,  $\mathbf{y}(\tau), \mathbf{z}(\tau) \in K \forall t \leq \tau \leq t + h$ ,

$$\|F(\mathbf{y}(\cdot)) - F(\mathbf{z}(\cdot))\|_\infty \leq h \|\mathbf{P}(\mathbf{f}(\cdot, \mathbf{y}(\cdot)) - \mathbf{f}(\cdot, \mathbf{z}(\cdot)))\|_\infty \stackrel{(2.2.29)}{\leq} h \|\mathbf{P}\|_L \|\mathbf{y}(\cdot) - \mathbf{z}(\cdot)\| .$$

$$\boxed{|h| < \frac{1}{\|\mathbf{P}\|_L}} \Rightarrow F \text{ ist Kontraktion.}$$

Für hinreichend kleines  $|h|$  bleibt  $F(\mathbf{y}(\cdot))$  in einer Umgebung der konstanten Funktion  $\mathbf{y}_0$ , wenn  $\mathbf{y}(\cdot)$  daraus gewählt wird.

Damit sind die Voraussetzungen des Fixpunktsatzes Thm. 2.2.9 erfüllt. □

 Notation:  $\tau \mapsto \mathbf{y}(\tau) \hat{=}$  Lösung des AWP  $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}), \mathbf{y}(t) = \mathbf{y}_0 \in D$ ,

$\tau \mapsto \mathbf{y}_h(\tau) \hat{=}$  Lösung von (2.2.24) (zu Anfangswert  $\mathbf{y}_0$ )



**Theorem 2.2.30** (Einschritt-Fehlerabschätzung für Projektions-Einschrittverfahren).

Erfüllt  $\mathbf{f}$  eine lokale Lipschitz-Bedingung ( $\rightarrow$  Def. 1.3.2), dann gibt es  $h_0 > 0$ , so dass

$$\exists C > 0: \quad \|\mathbf{y} - \mathbf{y}_h\|_\infty \leq Ch \|(Id - P)(\mathbf{f}(\cdot, \mathbf{y}(\cdot)))\|_\infty \quad \forall |h| \leq h_0 .$$

Projektionsfehler !

Thm. 2.2.30  $\Rightarrow$  **Konsistenzfehlerabschätzung** für Projektions-Einschrittverfahren:

$$\|\boldsymbol{\tau}(t, \mathbf{y}, h)\| := \left\| \Phi^{t, t+h} \mathbf{y} - \Psi^{t, t+h} \mathbf{y} \right\| \leq Ch \|(Id - P)(\mathbf{f}(\cdot, \mathbf{y}(\cdot)))\|_\infty ,$$

mit  $C > 0$  unabhängig von

- $\mathbf{y} \in K$ ,  $K \hat{=}$  kompakte Umgebung der Lösungstrajektorie  $t \mapsto \mathbf{y}(t)$ ,  $t_0 \leq t \leq T$ ,
- hinreichend kleiner Zeitschrittweite  $h > 0$ .

Verallgemeinerung von Lemma 2.2.13:

**Lemma 2.2.31** (Lipschitz-Stetigkeit der Inkrementfunktion von Projektions-Einschrittverfahren).

Erfüllt  $\mathbf{f}$  eine lokale Lipschitz-Bedingung ( $\rightarrow$  Def. 1.3.2), dann existiert zu jedem  $(t, \mathbf{y}) \in \Omega$  ein  $h_0$  so, dass, für  $|h| < h_0$ ,

$$\Psi^{t,t+h} \mathbf{y} = \mathbf{y} + h\psi(t, \mathbf{y}, h) ,$$

mit einer in der Zustandsvariablen  $\mathbf{y}$  lokal Lipschitz-stetigen ( $\rightarrow$  Def. 1.3.2) Inkrementfunktion  $\psi$  ( $\rightarrow$  Lemma 2.1.9).

*Beweis.* Unter Berufung auf Kompaktheitsargumente, vgl. Beweis von Thm. 2.1.19, o.B.d.A. Annahme einer globalen Lipschitz-Bedingung

$$\exists L > 0: \quad \|\mathbf{f}(\tau, \mathbf{z}) - \mathbf{f}(\tau, \mathbf{w})\| \leq L \|\mathbf{z} - \mathbf{w}\| \quad \forall \mathbf{z}, \mathbf{w} \in D, t \leq \tau \leq t + h .$$

Wie im Beweis von Lemma 2.2.27: sind  $\mathbf{y}_h, \mathbf{z}_h \in (C^0([t, t + h]))^d$  Lösungen von (2.2.26) zu „Anfangswerten“  $\mathbf{y}_0, \mathbf{z}_0 \in D$ , dann

$$|h| < \frac{1}{hL \|\mathbf{P}\|} \Rightarrow \|\mathbf{y}_h - \mathbf{z}_h\|_\infty \leq \frac{1}{1 - hL \|\mathbf{P}\|} \|\mathbf{y}_0 - \mathbf{z}_0\| . \quad (2.2.32)$$

Garantiert Existenz von  $\uparrow$  Lösungen von (2.2.26)

$$(2.2.26) \Rightarrow \Psi^{t,t+h} \mathbf{y}_0 = \mathbf{y}_0 + \underbrace{\int_t^{t+h} \mathbf{P}(\mathbf{f}(\cdot, \mathbf{y}_h(\cdot))) (\xi) d\xi}_{=: \mathbf{y}_0 + h\psi(t, \mathbf{y}_0, h)} .$$

$$\begin{aligned} \blacktriangleright \quad |\psi(t, \mathbf{y}_0, h) - \psi(t, \mathbf{z}_0, h)| &\leq \frac{1}{h} \int_0^h \mathbf{P}(\mathbf{f}(\cdot, \mathbf{y}_h(\cdot)) - \mathbf{f}(\cdot, \mathbf{z}_h(\cdot)))(\xi) \, d\xi \\ &\leq \|\mathbf{P}\| L \|\mathbf{y}_h - \mathbf{z}_h\|_\infty \stackrel{(2.2.32)}{\leq} \frac{\|\mathbf{P}\| L}{1 - hL \|\mathbf{P}\|} \|\mathbf{y}_0 - \mathbf{z}_0\| \quad . \quad \square \end{aligned}$$

$(\mathbf{y}_k)_{k=0}^N \hat{=}$  Gitterfunktion, erzeugt durch Projektions-Einschrittverfahren für  $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$  auf Zeitgitter  $\{t_0 < t_1 < \dots < t_N = T\} \subset J(t_0, \mathbf{y}_0)$ :  $\mathbf{y}_{k+1} = \Psi^{t_k, t_{k+1}} \mathbf{y}_k$

Mit Lemma 2.2.31 & Beweis von Thm. 2.1.19 (diskretes Gronwall-Lemma 2.1.20):

$\blacktriangleright$  Globale Fehlerabschätzung für Projektions-Einschrittverfahren (auf Zeitgitter)

$$\|\mathbf{y}_k - \mathbf{y}(t_k)\| \leq C \max_{j=1, \dots, N} \|(\text{Id} - \mathbf{P})(\mathbf{f}(\cdot, \mathbf{y}(\cdot)))\|_{\infty, [t_{j-1}, t_j]} \frac{\exp(L(h_1 + \dots + h_k)) - 1}{L}, \quad (2.2.33)$$

für  $k = 1, \dots, N$ ,  $h_j$  hinreichend klein,  $C > 0$  unabhängig von  $h_j$ ,  $k$ .

## 2.2.3 Konvergenz von Kollokationsverfahren

### 2.2.3.1 Konsistenzordnung

Frage: Konsistenz(ordnung) ( $\rightarrow$  Konvergenz, Sect. 2.1.3) von Kollokationsverfahren ?


Bem. 2.2.21  $\triangleright$  Konsequenz von Lemma 2.2.31:

**Lemma 2.2.34** (Konsistenz von Kollokationsverfahren).

*Unter den Voraussetzungen von Lemma 2.2.7 ist jedes Kollokations-Einschrittverfahren konsistent ( $\rightarrow$  Def. 2.1.8).*

Erinnerung: Verfahrensgleichungen eines Kollokations-Einschrittverfahrens shadowfullframeblack

$$\begin{aligned}
 \mathbf{y}_h(t_1) &= \mathbf{y}_0 + h \sum_{i=1}^s b_i \mathbf{k}_i, \\
 \mathbf{k}_i &= \mathbf{f}(t_0 + c_i h, \mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{k}_j).
 \end{aligned}
 \quad \text{mit} \quad
 \begin{aligned}
 a_{ij} &= \int_0^{c_i} L_j(\tau) d\tau, \\
 b_i &= \int_0^1 L_i(\tau) d\tau.
 \end{aligned}
 \quad (2.2.3)$$

 Notation:  $t \in [t_0, t_0 + h] \mapsto \mathbf{y}_h(t) \hat{=}$  durch Kollokationsverfahren erzeugte Näherungslösung, Polynom vom Grad  $s$ , siehe (2.2.1)

➤ (Einschritt-)Fehlerfunktion  $\mathbf{e}(t) := \mathbf{y}(t) - \mathbf{y}_h(t), \quad t_0 \leq t \leq t_1$

**Lemma 2.2.36** ((Suboptimale) Konsistenzordnung von Kollokationsverfahren).

Für hinreichend glatte rechte Seite  $\mathbf{f}$  ist das Kollokations-Einschrittverfahren (2.2.3) zu  $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$  konsistent von der Ordnung  $s$  ( $\rightarrow$  Def. 2.1.13).

**Lemma 2.2.37** (Fehlerabschätzung für Polynominterpolation). → [9, Satz 7.16]

Sei  $f \in C^{n+1}([x_0, x_n])$ ,  $x_0 < x_1 < \dots < x_n$ , und  $p \in \mathcal{P}_n$  das Interpolationspolynom von  $f$  zu den Stützstellen  $x_i$  (d.h.  $p(x_i) = f(x_i)$ ), dann gilt

$$|f^{(k)}(x) - p^{(k)}(x)| \leq \frac{|x_n - x_0|^{n+1-k}}{(n+1-k)!} \max_{x_0 < \xi < x_n} |f^{(n+1)}(\xi)| \quad \forall x_0 \leq x \leq x_n, k = 0, \dots, n+1.$$

*Beweis* von Lemma 2.2.36. Aus Lemma 2.2.37 folgern wird die konkrete Interpolationsfehlerabschätzung für  $P_{s-1}$  auf  $[t, t+h]$ :

$$\|(Id - P_{s-1})\mathbf{f}(\cdot, \mathbf{y}(\cdot))\|_\infty \leq h^s \left\| \frac{d^s}{dt^s} \mathbf{f}(\cdot, \mathbf{y}(\cdot)) \right\|_\infty. \quad (2.2.38)$$

Nach Annahme hinreichender Glattheit von  $\mathbf{f}$ , die sich auf die exakte Lösung  $\mathbf{y}$  des AWP überträgt, ist die rechte Seite in (2.2.38) asymptotisch  $O(h^s)$  für  $h \rightarrow 0$ .

Zusammen mit Thm. 2.2.30 gibt dies eine Schranke  $O(h^{s+1})$  für den Einschrittfehler (= Konsistenzfehler). □

“Direkter” Beweis von Lemma 2.2.36. (für autonome Dgl.  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}), t_0 = 0$ )

- $t \mapsto \mathbf{y}(t) \hat{=}$  exakte Lösung für AWP  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}), \mathbf{y}(0) = \mathbf{y}_0$
- $t \mapsto \mathbf{y}_h(t), 0 \leq t \leq h$ , polynomiale approximative Lösung aus einem Schritt des Kollokationsverfahrens,  $\mathbf{y}_h(0) = \mathbf{y}_0$ .

(Annahme:  $h$  hinreichend klein, siehe Lemma 2.2.7,  $h \in J(\mathbf{y}_0)$ )

Wiederum vereinfachende Annahme:  $\mathbf{f}$  global Lipschitz-stetig

$$\exists L > 0: \quad \|\mathbf{f}(\mathbf{z}) - \mathbf{f}(\mathbf{w})\| \leq L \|\mathbf{z} - \mathbf{w}\| \quad \forall \mathbf{z}, \mathbf{w} \in D. \quad (2.2.39)$$

Zur Konsistenzuntersuchung betrachte die **Einschrittfehler**funktion

$$\mathbf{e}(t) := \mathbf{y}(t) - \mathbf{y}_h(t) \quad \triangleright \quad \boldsymbol{\tau}(\mathbf{y}_0, h) = \mathbf{e}(h). \quad (2.2.40)$$

Aus den Kollokationsbedingungen (2.2.1):

$$\dot{\mathbf{y}}_h(t) = \sum_{i=1}^s \mathbf{f}(\mathbf{y}_h(c_i h)) \cdot L_i(\tau), \quad \tau := \frac{t}{h}. \quad (2.2.41)$$

Aus der Lösungseigenschaft von  $t \mapsto \mathbf{y}(t)$

$$\dot{\mathbf{y}}(t) = \mathbf{f}(\mathbf{y}(t)) = \sum_{i=1}^s \mathbf{f}(\mathbf{y}(c_i h)) \cdot L_i(\tau) + \mathbf{r}(t), \quad \tau := \frac{t}{h}. \quad (2.2.42)$$

Interpolationspolynom  $\in \mathcal{P}_{s-1}$  zur Funktion  $t \mapsto \mathbf{f}(\mathbf{y}(t))$  und Knoten  $c_i h$  auf  $[0, h]$

$\mathbf{r} \hat{=}$  Restglied für Polynominterpolation, siehe Lemma 2.2.37, erfüllt

$$\left\| \mathbf{r}^{(k)}(t) \right\| \leq \underbrace{\frac{1}{(s-k)!} \max_{0 < t < T} \left\| \mathbf{y}^{(s+1)}(t) \right\|}_{\text{unabhängig von } h} h^{s-k} \leq C h^{s-k}, \quad k = 0, \dots, s. \quad (2.2.43)$$

Konvention: Alle Konstanten  $C$  unabhängig von (hinreichend kleinem)  $h$ , dürfen abhängen von  $\mathbf{y}(t)$ ,  $\mathbf{f}$ , Parametern des Kollokationsverfahrens, etc.

Aus (2.2.41) & (2.2.42) & Integration  $\triangleright$  Ausdruck für Einschrittfehlerfunktion

$$\mathbf{e}(t) = h \sum_{i=1}^s \Delta \mathbf{f}(c_i h) \cdot \int_0^\tau L_i(\sigma) d\sigma + \int_0^t \mathbf{r}(\sigma) d\sigma, \quad 0 \leq \tau \leq 1, \quad (2.2.44)$$

wobei  $\Delta \mathbf{f}(t) = \mathbf{f}(\mathbf{y}(t)) - \mathbf{f}(\mathbf{y}_h(t))$ .

$$(2.2.39) \Rightarrow \|\Delta \mathbf{f}(t)\| \leq L \|\mathbf{e}(t)\|. \quad (2.2.45)$$

$$(2.2.44) \& (2.2.43) \implies \|\mathbf{e}\|_\infty := \max_{0 \leq t \leq h} \|\mathbf{e}(t)\| \leq C_1 L h \|\mathbf{e}\|_\infty + C_2 h^{s+1}, \quad (2.2.46)$$



mit von  $h$  unabhängigen Konstanten  $C_1, C_2 > 0$  (, die von den Kollokationspunkten  $c_i$  und  $\mathbf{y}$  abhängen.)

$$C_1 L h_0 < 1 \quad \Rightarrow \quad \|\boldsymbol{\tau}(\mathbf{y}, h)\| \leq \|\mathbf{e}\|_\infty \leq \frac{C_2}{1 - C_1 h_0 L} h^{s+1} \quad \forall |h| \leq h_0. \quad (2.2.47)$$

☞ Wir folgern: Das Kollokationsverfahren hat mindestens Konsistenzordnung  $s$ . □

Aus (2.2.44) und (2.2.43) lässt sich sogar folgern, für  $k = 0, \dots, s$ ,

$$\begin{aligned} \max_{0 \leq t \leq h} \left\| \mathbf{e}^{(k)}(t) \right\| &\leq C_1(k) L h \|\mathbf{e}\|_\infty + C_2(k) h^{s+1-k} \\ \Rightarrow \max_{0 \leq t \leq h} \left\| \mathbf{e}^{(k)}(t) \right\| &\leq C(k) h^{s+1-k}, \end{aligned} \quad (2.2.48)$$

mit von  $h$  unabhängigen Konstanten  $C_1(k), C_2(k), C(k) > 0$ .

*Beispiel 2.2.49* (Konvergenz von Gauss-Kollokations-Einschrittverfahren).  $\rightarrow$  Bsp. 2.2.17

- Skalare logistische Differentialgleichung (1.2.2),  $\lambda = 10$ ,  $y(0) = 0.01$ ,  $T = 1$
- Gauss-Kollokations-Einschrittverfahren (2.2.3) für  $s = 1, \dots, 4$ , uniforme Zeitschrittweite  $h$

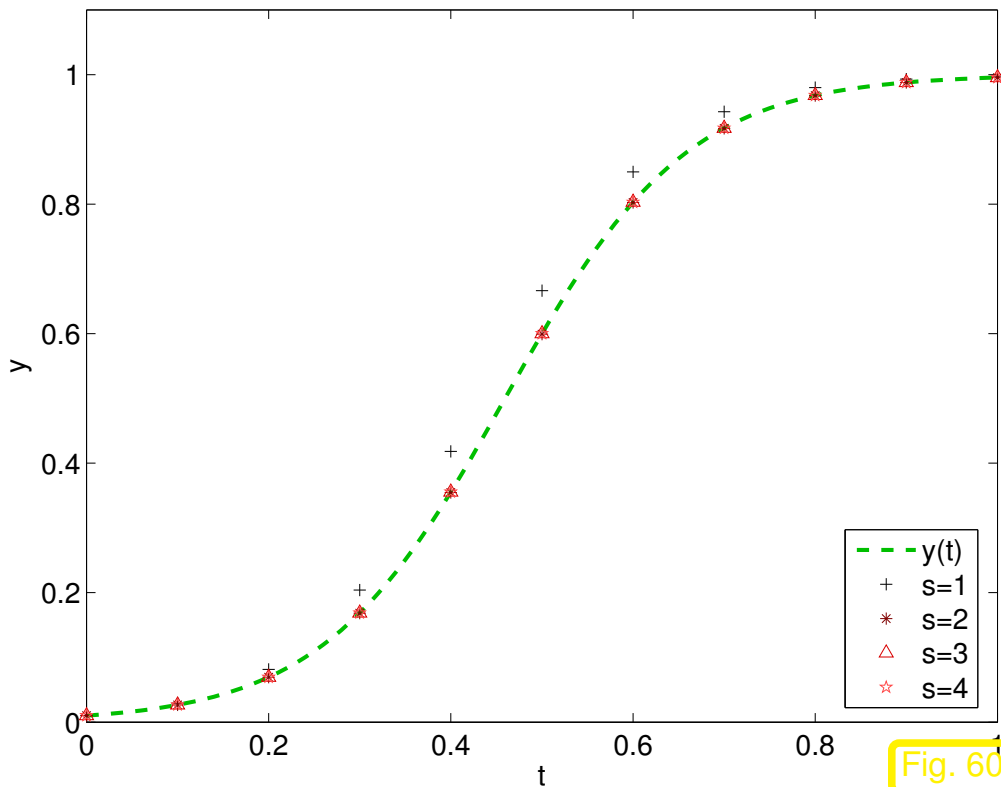


Fig. 60

$y_h(j/10)$ : Gauss-Kollokationsverfahren

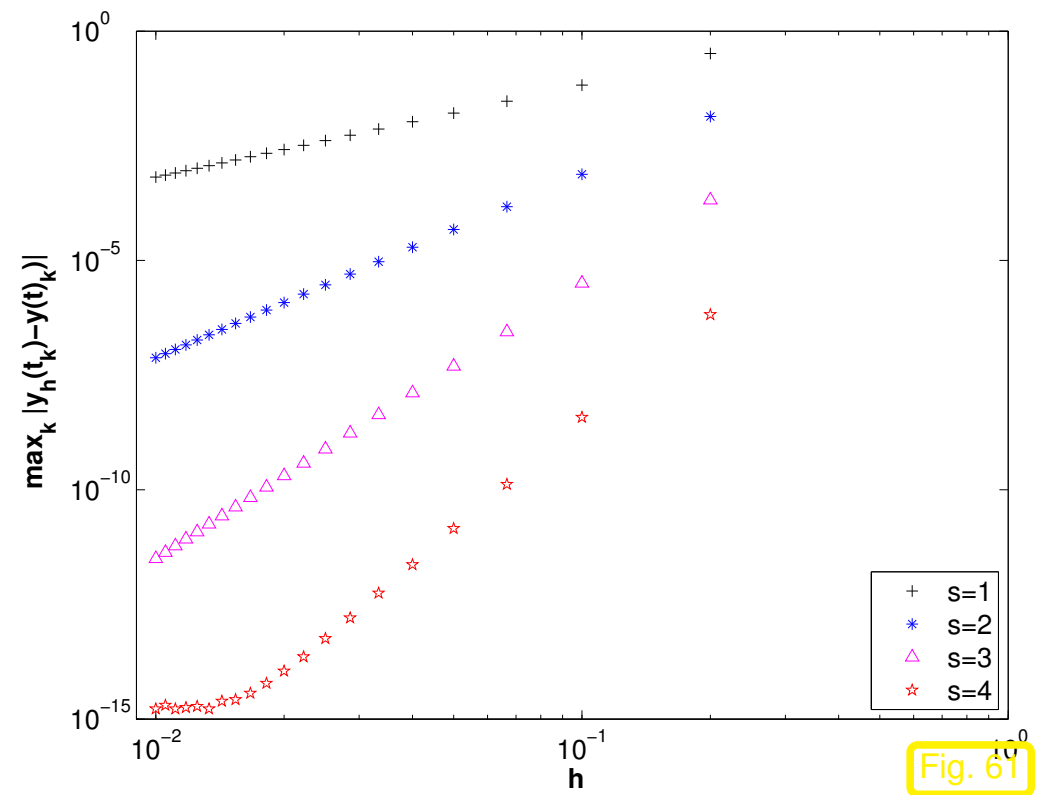


Fig. 61

Konvergenz des Fehlers  $\max_k |y_k - y(t_k)|$

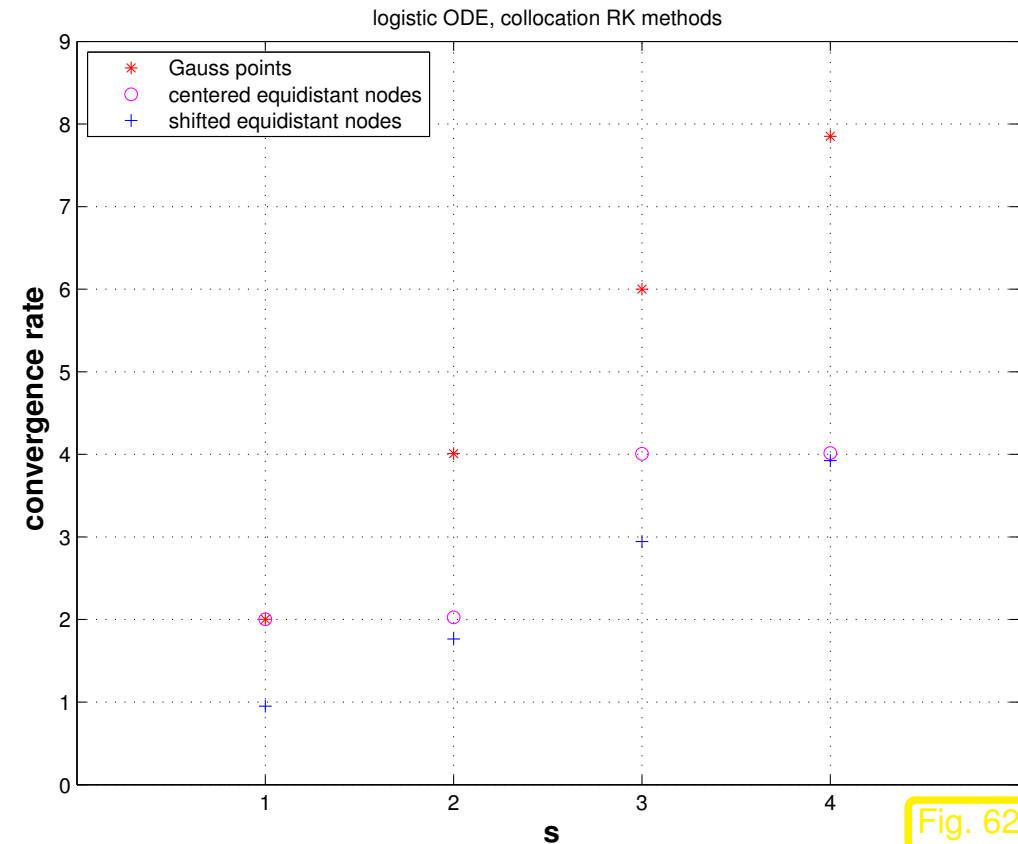
Numerische Konvergenzraten :  
 (berechnet durch lineare Regression)

$s = 1$	:	$p = 1.96$
$s = 2$	:	$p = 4.01$
$s = 3$	:	$p = 6.00$
$s = 4$	:	$p = 8.02$

Student Version of MATLAB

Beobachtung: Konvergenzraten sind doppelt so hoch wie die untere Schranke aus Lemma 2.2.36!

Vergleich der (empirischen) Konvergenzraten  $\triangleright$



R. Hiptmair  
rev 35327,  
25. April  
2011

Betrachte: Kollokationsverfahren zu  $\dot{\mathbf{y}} = f(t, \mathbf{y})$  mit (relativen) Kollokationspunkten  $c_i \in [0, 1]$ ,  $i = 1, \dots, s$ ,  $s \in \mathbb{N}$   $\triangleright$  Koeffizienten  $b_i$ ,  $a_{ij}$  in (2.2.3).

$\blacktriangleright$  Zugeordnete Quadraturformel, Bem. 2.2.18:  $Q(f) = h \sum_{i=1}^s b_j f(t_0 + c_j h)$ . (2.2.50)

Bsp. 2.2.49 legt die Vermutung nahe, dass die Konsistenzordnung eines Kollokationsverfahrens mit der Ordnung der gemäss (2.2.50) zugeordneten Quadraturformel übereinstimmt.

**Theorem 2.2.51** (Konsistenzordnung von Kollokationsverfahren).

*Die Konsistenzordnung ( $\rightarrow$  Def. 2.1.13) eines Kollokations-Einschrittverfahrens stimmt mit der Ordnung der zugeordneten Quadraturformel überein.*

Hilfsmittel beim Beweis: Fehlerabschätzung für numerische Quadratur ( $\rightarrow$  Vorlesung “Numerische Methoden”)

$$s\text{-Punkt-Quadraturformel auf } [a, b]: \quad Q(f) := (b - a) \sum_{i=1}^s b_i f(a + c_i(b - a)) \approx \int_a^b f(x) dx . \quad (2.2.52)$$

Annahme: innere Knoten  $0 \leq c_i \leq 1, i = 1, \dots, s$

→ Quadraturformel von der **Ordnung**  $n + 1$ **Lemma 2.2.53** (Quadraturfehlerabschätzung).*Ist eine Quadraturformel (2.2.52) exakt für Polynome von Grad  $\leq n$ , so gilt*

$$f \in C^{n+1}([a, b]) \Rightarrow \left| Q(f) - \int_a^b f(x) dx \right| \leq C \frac{(b-a)^{n+2}}{(n+1)!} \max_{a < x < b} |f^{(n+1)}(x)| ,$$

mit  $C = 1 + \sum_{i=1}^s |b_i|$ .Annahme:  $\mathbf{f}$  „hinreichend“ glatt, lokal Lipschitz-stetig (→ Def. 1.3.2)*Beweis von Thm. 2.2.51* (für autonome Dgl.  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ )Idee: Interpretiere  $t \mapsto \mathbf{y}_h(t)$  als Lösung eines **gestörten Anfangswertproblems** !

$$\dot{\mathbf{y}}_h(t) = \mathbf{f}(\mathbf{y}_h(t)) + \underbrace{\dot{\mathbf{y}}_h(t) - \mathbf{f}(\mathbf{y}_h(t))}_{:=\delta(t)} , \quad 0 \leq t \leq h . \quad (2.2.54)$$

Wegen  $\dot{\mathbf{y}}(t) = \mathbf{f}(\mathbf{y}(t))$  folgt für die Einschrittfehlerfunktion  $\mathbf{e}(t) = \mathbf{y}(t) - \mathbf{y}_h(t)$

$$\dot{\mathbf{e}}(t) = \mathbf{f}(\mathbf{y}(t)) - \mathbf{f}(\mathbf{y}_h(t)) - \delta(t), \quad 0 \leq t \leq h.$$

Hilfsmittel: Taylorformel

$$\varphi(1) - \varphi(0) = \varphi'(0) + \int_0^1 (1 - \tau) \varphi''(\tau) d\tau,$$

für  $\varphi(\xi) := \mathbf{f}(\mathbf{y}(t) + \xi(\mathbf{y}_h(t) - \mathbf{y}(t)))$  mit der Kettenregel:

$$\varphi'(\xi) = D\mathbf{f}(\mathbf{y}(t) + \xi(\mathbf{y}_h(t) - \mathbf{y}(t))) \cdot (\mathbf{y}_h(t) - \mathbf{y}(t)),$$

$$\varphi''(\xi) = D^2\mathbf{f}(\mathbf{y}(t) + \xi(\mathbf{y}_h(t) - \mathbf{y}(t))) (\mathbf{y}_h(t) - \mathbf{y}(t), \mathbf{y}_h(t) - \mathbf{y}(t)).$$

Einsetzen in die Formel für die Einschrittfehlerfunktion:

$$\dot{\mathbf{e}}(t) = \varphi(0) - \varphi(1) - \delta(t) = D\mathbf{f}(\mathbf{y}(t))\mathbf{e}(t) - \underbrace{\int_0^1 (1 - \tau) D^2\mathbf{f}(\mathbf{y}(t) + \tau\mathbf{e}(t)) (\mathbf{e}(t), \mathbf{e}(t)) d\tau}_{=:\rho(t)} - \delta(t).$$

Dabei gilt die offensichtliche Abschätzung:

$$\|\rho(t)\| \leq \max_{\mathbf{y} \in K} \|D^2\mathbf{f}(\mathbf{y})\| \cdot \|\mathbf{e}(t)\|^2 \stackrel{\text{Lemma 2.2.36}}{\leq} Ch^{2s+2}, \quad (2.2.55)$$

wobei  $K = \{\mathbf{z} \in D: \|\mathbf{z} - \mathbf{y}(t)\| \leq R\}$  mit von  $t$  unabhängigem  $R > 0$  und  $C > 0$  unabhängig von  $h$ .

➤ Einschrittfehlerfunktion löst das Anfangswertproblem

$$\dot{\mathbf{e}} = D\mathbf{f}(\mathbf{y}(t))\mathbf{e} - \rho(t) - \delta(t) \quad , \quad \mathbf{e}(0) = 0 \quad . \quad (2.2.56)$$

Betrachtet man  $\rho(t), \delta(t)$  als bloße Funktionen von  $t$ , dann ist (2.2.56) eine *nichtautonome lineare* Differentialgleichung.

➤ Lösung durch allgemeine Variation-der-Konstanten-Formel (1.3.18):

$$\mathbf{e}(t) = - \int_0^t \mathbf{W}(t; \mathbf{y}_0) \mathbf{W}(\tau; \mathbf{y}_0)^{-1} (\rho(\tau) + \delta(\tau)) \, d\tau \quad , \quad 0 \leq t \leq h \quad .$$

mit der **Propagationsmatrix**  $t \mapsto \mathbf{W}(t; \mathbf{y}_0)$ , vgl. (1.3.33), Sect. 1.3.3.4. Sie löst das Anfangswertproblem für die **Variationsgleichung** (1.3.34)

$$\dot{\mathbf{W}}(t; \mathbf{y}_0) = D\mathbf{f}(\mathbf{y}(t))\mathbf{W}(t; \mathbf{y}_0) \quad , \quad \mathbf{W}(0; \mathbf{y}_0) = \mathbf{I} \quad .$$

Die Propagationsmatrix ist natürlich unabhängig von  $h$ , also

$$\begin{aligned} \exists C > 0 \quad \text{unabhängig von } h: \quad & \left\| \mathbf{W}(t; \mathbf{y}_0) \mathbf{W}(\tau; \mathbf{y}_0)^{-1} \right\| \leq C \quad \forall 0 \leq t, \tau \leq h \quad . \\ & \stackrel{(2.2.55)}{\Rightarrow} \left\| \int_0^t \mathbf{W}(t; \mathbf{y}_0) \mathbf{W}(\tau; \mathbf{y}_0)^{-1} \rho(\tau) \, d\tau \right\| \leq Ch^{2s+3} \quad , \end{aligned}$$

mit  $C > 0$  unabhängig von  $h$ .

Beachte: (2.2.40) ➤ Konsistenzfehler bestimmt durch  $\mathbf{e}(h)$  !

Geniale Idee: Abschätzung von  $\int_0^h \mathbf{W}(t; \mathbf{y}_0) \mathbf{W}(\tau; \mathbf{y}_0)^{-1} \delta(\tau) d\tau$  als Quadraturfehler !

Wegen  $\delta(c_i h) = 0$  (Kollokationsbedingung (2.2.1) !):

$$\underbrace{\sum_{i=1}^s b_i \mathbf{W}(t; \mathbf{y}_0) \mathbf{W}(c_i h; \mathbf{y}_0)^{-1} \underbrace{\delta(c_i h)}_{=0}}_{=0} = 0 \quad \forall 0 \leq t \leq h .$$

Quadraturformel auf  $[0, h]$  für  $\tau \mapsto \mathbf{W}(t; \mathbf{y}_0) \mathbf{W}(\tau; \mathbf{y}_0)^{-1} \delta(\tau)$

➤ Mit der Quadraturfehlerabschätzung aus Lemma 2.2.53

$$\left\| \int_0^h \mathbf{W}(t; \mathbf{y}_0) \mathbf{W}(\tau; \mathbf{y}_0)^{-1} \delta(\tau) d\tau - \sum_{i=1}^s b_i \mathbf{W}(t; \mathbf{y}_0) \mathbf{W}(c_i h; \mathbf{y}_0)^{-1} \delta(c_i h) \right\| \leq C_3 h^{p+1} ,$$

mit

$$C_3 := \frac{1}{p!} \max_{0 \leq \tau \leq h} \left\| \frac{d^p}{d\tau^p} \left\{ \tau \mapsto \mathbf{W}(h; \mathbf{y}_0) \mathbf{W}(\tau; \mathbf{y}_0)^{-1} \delta(\tau) \right\} \right\| .$$

$\delta(t) := \dot{\mathbf{y}}_h(t) - \mathbf{f}(\mathbf{y}_h(t))$  hängt natürlich im Gegensatz zu  $\mathbf{W}(t; \mathbf{y})$  von  $h$  ab, doch dank der Schranken aus (2.2.48) sind alle Ableitungen von  $\mathbf{y}_h$  gleichmässig in  $h$  beschränkt !. Also lässt sich auch  $C_3$  unabhängig von  $h$  beschränken.

$$\blacktriangleright \left\| \int_0^h \mathbf{W}(h; \mathbf{y}_0) \mathbf{W}(\tau; \mathbf{y}_0)^{-1} \delta(\tau) d\tau \right\| \leq C h^{p+1} .$$



Zusammen mit der Abschätzung für den  $\rho$ -Term ergibt sich die Behauptung des Theorems, vgl. Def. 2.1.13 □

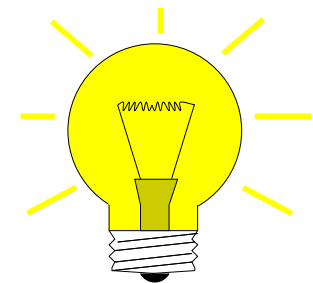
▶ s-stufige implizite Gauss-Kollokations-Einschrittverfahren haben Ordnung  $2s$

### 2.2.3.2 Spektrale Konvergenz

Erinnerung: Polynominterpolation: „Grad  $\rightarrow \infty \Rightarrow$  Fehler  $\rightarrow 0$ “  
(für geeignete Knoten, z.B. Tschebyscheff-Knoten [9, Sect. 7.1.4])

▶ Bei Gauss-Kollokations-Einschrittverfahren:

Konvergenz für  $s \rightarrow \infty$  ?



*Beispiel 2.2.57* (Konvergenz von globalen Gauss-Kollokationsverfahren).

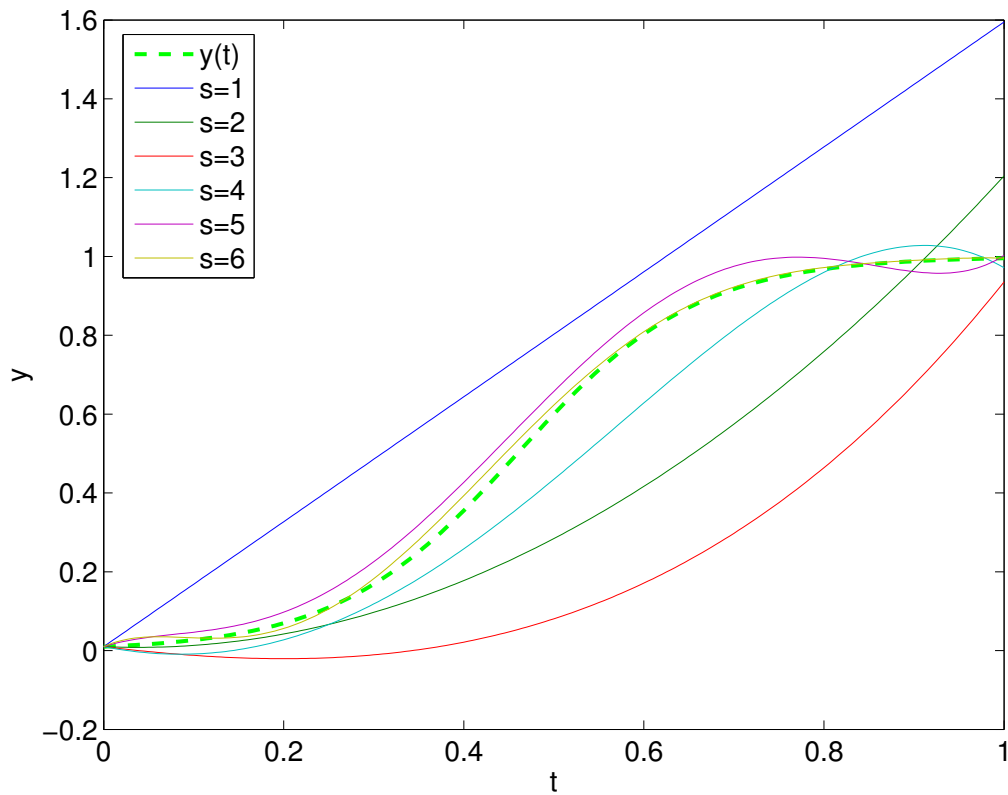
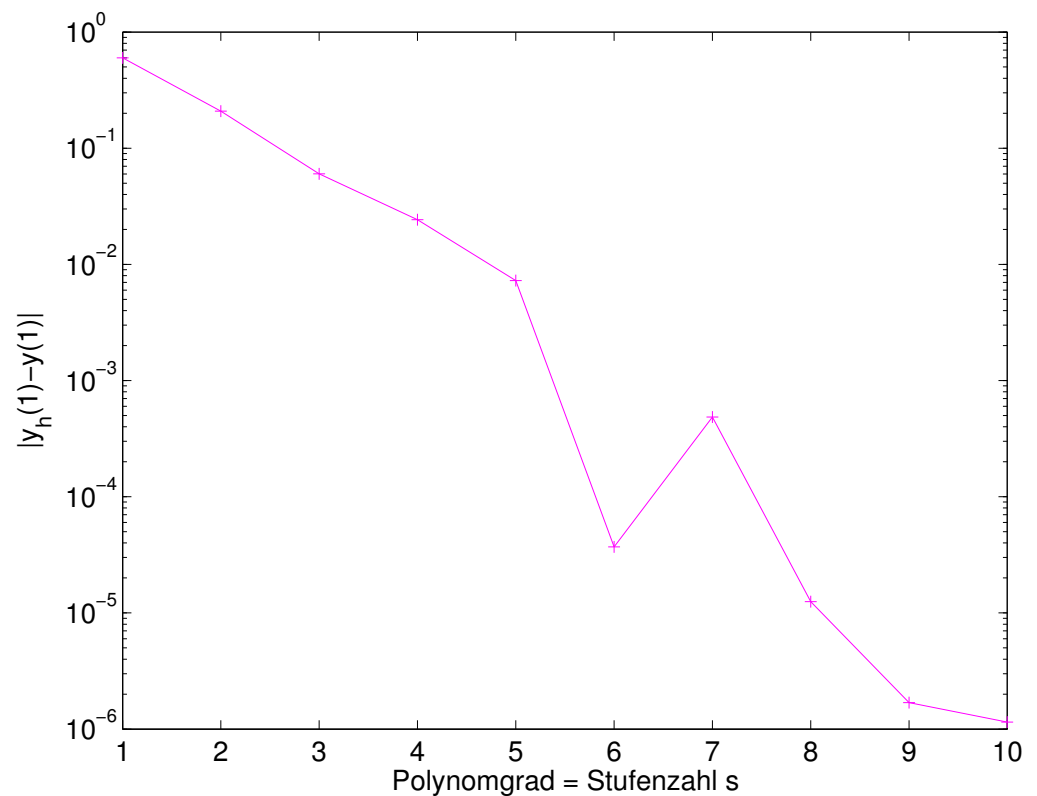
Dieses Beispiel studiert den *Einschrittfehler* von Kollokationsverfahren in Abhängigkeit von der Anzahl der Kollokationspunkte  $\leftrightarrow$  Polynomgrad  $s$ . Bisher haben wir nur die Strategie betrachtet, durch Verfeinerung des Zeitgitters eine genauere Lösung zu erhalten.

Logistische Differentialgleichung ( $\rightarrow$  Bsp. 1.2.1)

$$\dot{y} = \lambda y(1 - y), \quad y_0 \in ]0, 1[ \Rightarrow y(t) = \frac{1}{1 + (y_0^{-1} - 1)e^{-\lambda t}}, \quad t \in \mathbb{R}. \quad (2.2.58)$$

Numerische Experimente mit Gauss-Kollokationsverfahren auf  $[0, 1]$ ,  $y_0 = 0.01$ ,  $\lambda = 10$ :  
(Lösung der Gleichungen für Inkremente  $k_j$ : MATLAB `fsolve`, Toleranz  $10^{-9}$ )

Hier: Kollokationsverfahren als **globales** Integrationsverfahren

Näherungslösungen  $y_h(t)$ 

▷ Exponentielle Konvergenz in  $s$  (Warum ?)



Naheliegend: Konvergenzanalyse auf der Grundlage von Thm. 2.2.30

☞ Benötigt: **Spektrale Interpolationsfehlerabschätzungen** für Polynominterpolation in Gauss-Knoten, siehe Abb. 59 (spektral: Fehlerabschätzungen in Abhängigkeit vom Polynomgrad, ein neuer Aspekt im Vergleich zur Vorlesung “numerische Methoden”).

# Polynominterpolationsfehlerabschätzungen für analytische Funktionen

*Beispiel 2.2.59* (Interpolationsfehler bei Polynominterpolation in Gauss-Knoten).

Interpoland: Lösung der logistischen Dgl. (1.2.2)  
auf  $[-1, 1]$ , vgl. Bsp. 1.2.1:

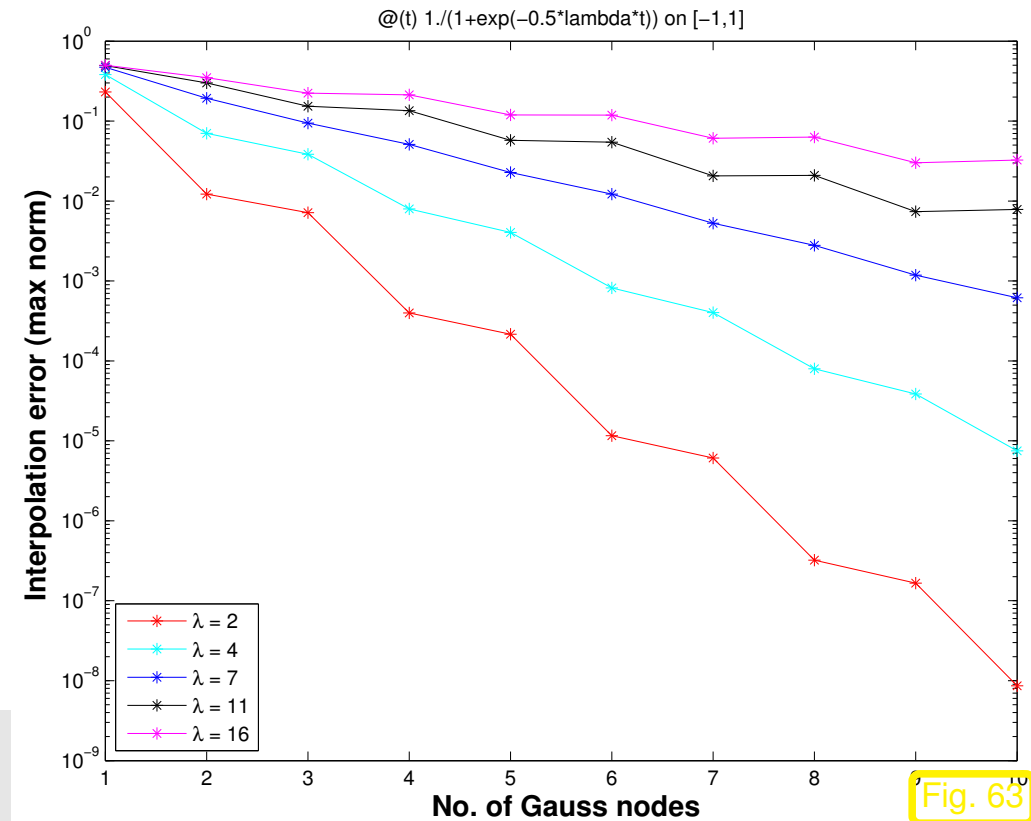
$$y(t) = \frac{1}{1 + \exp(-\frac{1}{2}\lambda t)} .$$

Fehler:

$$\text{err} := \max_{-1 \leq t \leq 1} |y(t) - p_n(t)| ,$$

$p_n(t) \hat{=}$  Interpolationspolynom von  $y(t)$ , Grad  $n - 1$ , zu  $n$  Gauss-Knoten.

Näherungsweise Auswertung von  $\text{err}$  durch Abtasten auf sehr feinem Gitter



## Listing 2.1: Berechnung approximativer Maximumnorm des Fehlers bei Polynominterpolation

```
1 function errinf = polyintperr(fun,nodes,span)
2 % Error of polynomial interpolation in maximum norm
3 % fun : handle to function to be interpolated
4 % nodes : interpolation nodes
5 % span : evaluation interval (default [-1,1])
6
7 if (nargin < 3), span = [-1,1]; end
8 n = length(nodes);
9
10 fval = zeros(1,n);
11 for j=1:n, fval(j) = fun(nodes(j)); end
12
13 p = polyfit(nodes,fval,n-1); % built-in polynomial interpolation
14
15 % Compute maximum norm by sampling on fine mesh
16 N = 1000*n;
17 t = span(1) + (0:N)*(span(2)-span(1))/N;
18 pval = polyval(p,t);
19 fval = zeros(1,N+1);
20 for j=1:N+1, fval(j) = fun(t(j)); end
21 errinf = max(abs(pval-fval));
```

```
1 function errinf = gaussintperr(fun,n)
2 % Error of polynomial interpolation in Gauss points in maximum norm
3 % fun : handle to function to be interpolated
4 % n : number of interpolation nodes
5
6 path (path , '../SupportScripts');
7 [nodes,weights] = GaussQuad(n);
8 errinf = polyintperr(fun,nodes');
```

Listing 2.3: Erzeugen der Plots für Bsp. 2.2.59

```
1 function plotgaussintperr
2 % Error of Gaussian interpolation for solution of logistic differentialequation
3
4 rec = []; % Array for recording errors
5 k = 1;
6 for lambda=[2,4,7,11,16]
7     sol = @(t) 1./(1+exp(-0.5*lambda*t));
8     errs = [];
9     for n=1:10, errs = [errs,gaussintperr(sol,n)]; end
10    rec = [rec;errs];
11    leg{k} = sprintf ('\\lambda = %d',lambda);
12    k = k+1;
13 end
```

```
15 figure ('name', 'Gaussian interpolation error');
16 semilogy (1:10, rec(1,:), 'r*-', ...
17           1:10, rec(2,:), 'c*-', ...
18           1:10, rec(3,:), 'b*-', ...
19           1:10, rec(4,:), 'k*-', ...
20           1:10, rec(5,:), 'm*-');
21 xlabel ('\bf No. of Gauss nodes', 'fontsize', 14);
22 ylabel ('\bf Interpolation error (max norm)', 'fontsize', 14);
23 title ('@ (t) 1./ (1+exp(-0.5*lambda*t)) on [-1,1]');
24 legend (leg, 'location', 'southwest');
25
26 print -depsc2 'gaussintperr.eps';
```

Beobachtung: **exponentielle Konvergenz** ( $\rightarrow$  Def. 1.4.5) des Interpolationsfehlers, schneller bei kleinerem  $\lambda$ .

Bekannt sein sollte aus der Funktionentheorie: Konzept einer

- *holomorphen* Funktion,
- Cauchy Integralsatz , und
- Laurent-Entwicklung.

Erinnerung, siehe etwa [30, Ch. 13]:

**Theorem 2.2.60** (Residuensatz).

Sei  $D \subset \mathbb{C}$  eine offene Menge,  $\Gamma \subset D$  ein einfach geschlossener Integrationsweg und  $\Pi \subset D$  eine endliche Menge.

Für jede in  $D \setminus \Pi$  holomorphe (analytische) Funktion  $f : D \setminus \Pi \mapsto \mathbb{C}$  gilt

$$\frac{1}{2\pi i} \int_{\gamma} f(z) dz = \sum_{p \in \Pi} \operatorname{res}_p f ,$$

wobei  $\operatorname{res}_p f$  das **Residuum** von  $f$  im Punkt  $p$  ist.



$\Pi$  wird oft als die Menge der **Pole** von  $f$  bezeichnet.

**Definition 2.2.61** (Residuum einer komplexwertigen Funktion).

Ist  $f$  holomorph in einer punktierten Umgebung von  $p \in \mathbb{C}$ , so ist das Residuum  $\operatorname{res}_p f$  von  $f$  im Punkt  $p$  der Koeffizient  $a_{-1}$  der **Laurent-Entwicklung** von  $f$  in  $p$ .

*Beweisskizze.* (von Thm. 2.2.60)

Hat  $f$  in einer punktierten Umgebung von  $p \in \mathbb{C}$  die konvergente Laurent-Entwicklung

$$f(z) = \sum_{k=-\infty}^{\infty} a_k (z - p)^k$$

so gilt für einen (hinreichend kleinen) Kreis  $\gamma$  um  $p$

$$\frac{1}{2\pi i} \int_{\gamma} f(z) dz = a_{-1} .$$

Dann zerlege  $\int_{\Gamma}$  wie in der Skizze angedeutet und verwende den Cauchy-Integralsatz.

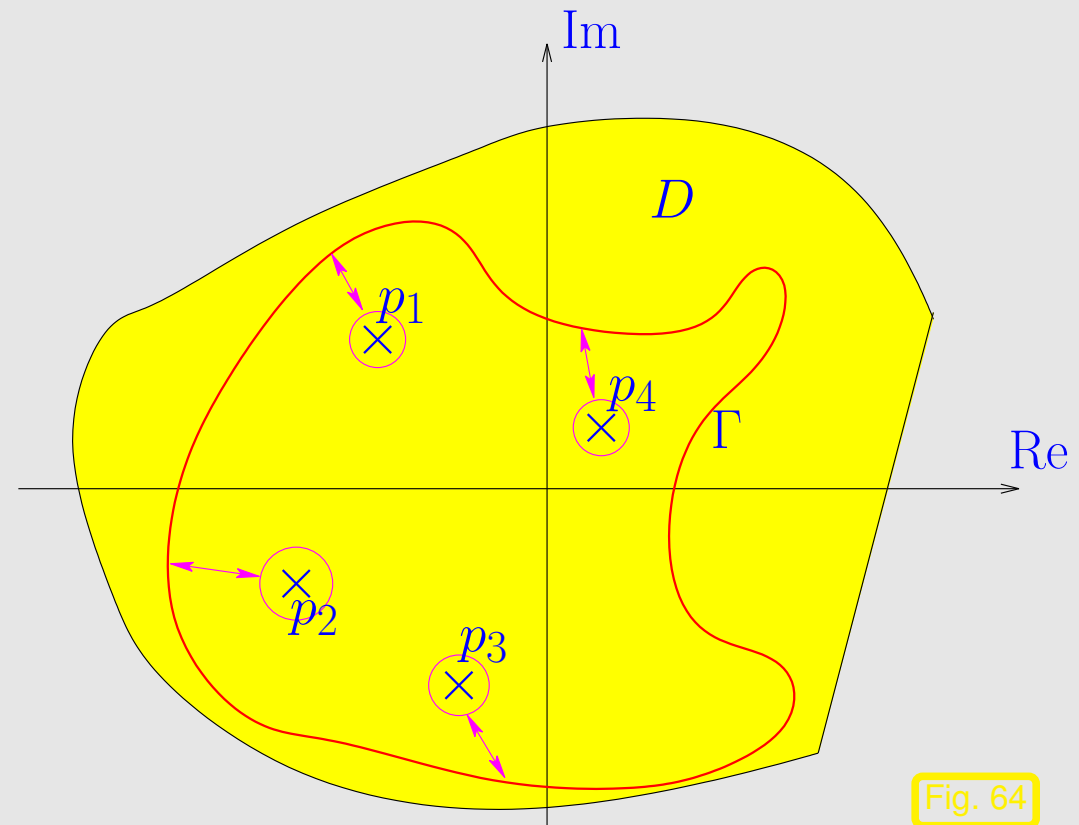


Fig. 64

**Lemma 2.2.62** (Residuenformel für einfachen Pol).

Ist  $f$  holomorph in einer punktierten Umgebung von  $p \in \mathbb{C}$  und  $(z - p)f(z)$  holomorph in  $p$ , so gilt

$$\operatorname{res}_p f = \lim_{z \rightarrow p} (z - p)f(z) . \quad (2.2.63)$$

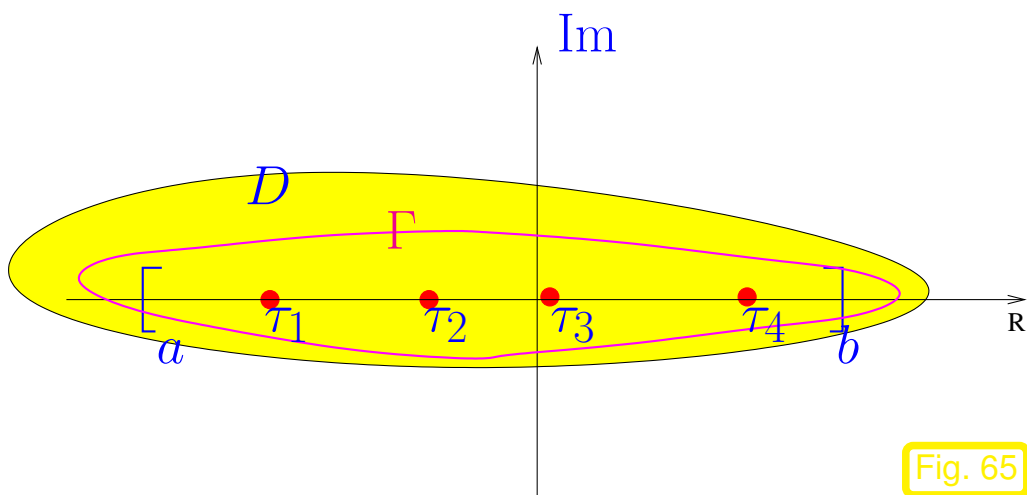
Daraus folgt sofort:

**Lemma 2.2.64** (Residuenformel für Quotienten).

Sind  $g, h$  holomorph in einer Umgebung von  $p \in \mathbb{C}$  und  $h(p) = 0, h'(p) \neq 0$ , so gilt

$$\operatorname{res}_p \frac{g}{h} = \frac{g(p)}{h'(p)} .$$

Betrachte: Polynominterpolation von  $f \in C^0([a, b])$  in Knoten  $\tau_1 < \tau_2 < \dots < \tau_s, s \in \mathbb{N}$



**Annahme 2.2.65** (Analytizität des Interpolanden).

*f* holomorph (analytisch) in eine komplexen Umgebung  $D \subset \mathbb{C}$  von  $[a, b]$  fortsetzbar.

Fig. 65

◁ Analytizitätsgebiet (gelb)

Wir betrachten folgende Funktion mit Polmenge  $\Pi = \{t, \tau_1, \dots, \tau_s\}$

$$g_t(z) := \frac{f(z)}{(z - t)P(z)}, \quad t \in [a, b] \setminus \{\tau_1, \dots, \tau_s\}, \quad P(z) := \alpha(z - \tau_1) \cdots (z - \tau_s), \quad \alpha \in \mathbb{R}.$$

➤  $g_t$  ist holomorph auf  $D \setminus \{t, \tau_1, \dots, \tau_s\}$  ( $t$  ist Parameter!).)

▶ Anwendung des Residuensatzes Thm. 2.2.60 auf  $g_t$  mit einfach geschlossenem Integrationsweg  $\Gamma \subset D$ , der  $[a, b]$  umschliesst, siehe die magenta Kurve in Abb. 65:

$$\frac{1}{2\pi i} \int_{\Gamma} g_t(z) dz = \text{res}_t g_t + \sum_{j=1}^s \text{res}_{\tau_j} g_t \stackrel{\text{Lemma 2.2.64}}{=} \frac{f(t)}{P(t)} + \sum_{j=1}^s \frac{f(\tau_j)}{(\tau_j - t)P'(\tau_j)}$$

Möglich, da  $P$  ausschliesslich einfache Nullstellen hat !

$$\begin{aligned}
 f(t) = & \underbrace{-\sum_{j=1}^s f(\tau_j) \frac{P(t)}{(\tau_j - t)P'(\tau_j)}}_{\text{Interpolationspolynom !}} + \underbrace{\frac{P(t)}{2\pi i} \int_{\Gamma} g_t(z) dz}_{\text{Interpolationsfehler !}}. \quad (2.2.66)
 \end{aligned}$$

➤ Nun abzuschätzen: rechte Seite von

$$|f(t) - \text{Interpolationspolynom}(t)| \leq \left| \frac{P(t)}{2\pi i} \int_{\Gamma} \frac{f(z)}{(z-t)P(z)} dz \right|, \quad a \leq t \leq b. \quad (2.2.67)$$

Bausteine der Abschätzung:

- Obere Schranke für  $|P(t)|$ ,  $a \leq t \leq b$
- Untere Schranke für  $|P(z)|$ ,  $z \in \Gamma$  für einen geschickt gewählten Integrationsweg  $\Gamma \subset D$

Offensichtlich hängt die Schranke in (2.2.67) nicht von  $\alpha$  ab.

# Abschätzungen für Legendre-Polynome

Erinnerung: Für  $\{\tau_j\}_{j=1}^s \hat{=}$  Gauss-Knoten in  $[-1, 1]$   $\blacktriangleright$   $P(t) \hat{=}$   $s$ . **Legendre-Polynom** (Grad  $s$ )

 Notation:  $P_n \hat{=}$  Legendre-Polynom vom Grad  $n \in \mathbb{N}_0$

$$\text{Rekursionsformel: } (n+1)P_{n+1}(t) - (2n+1)tP_n(t) + nP_{n-1}(t) = 0, \quad (2.2.68)$$

$$\text{Rodrigues-Formel: } P_n(t) = \frac{1}{2^n n!} \frac{d^n}{dt^n} (t^2 - 1)^n. \quad (2.2.69)$$

(Start der Rekursion mit  $P_0 \equiv 1, P_1(t) = t$ )

Legendre-Polynome auf  $[-1, 1]$

$$P_0(x) = 1,$$

$$P_1(x) = x,$$

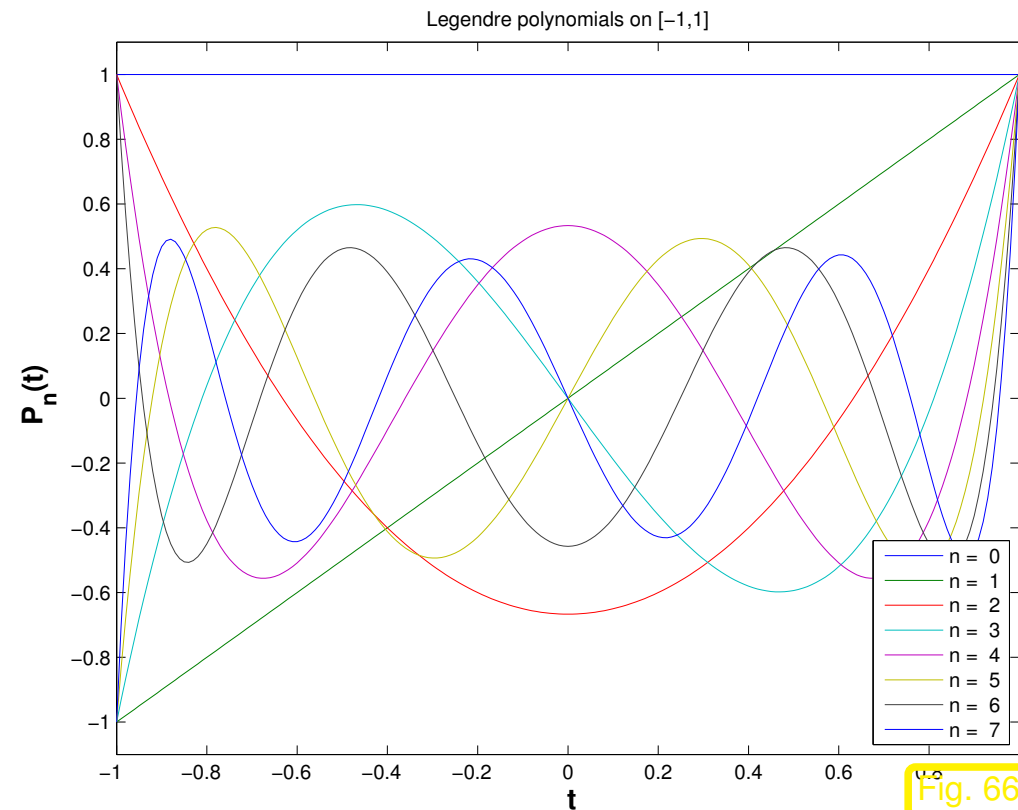
$$P_2(x) = \frac{1}{2}(3x^2 - 1),$$

$$P_3(x) = \frac{1}{2}(5x^3 - 3x),$$

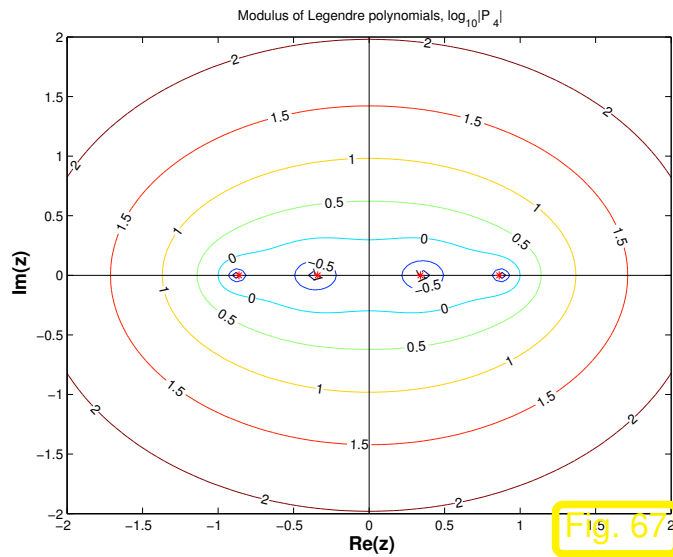
$$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3),$$

$$P_5(x) = \frac{1}{8}(63x^5 - 70x^3 + 15x),$$

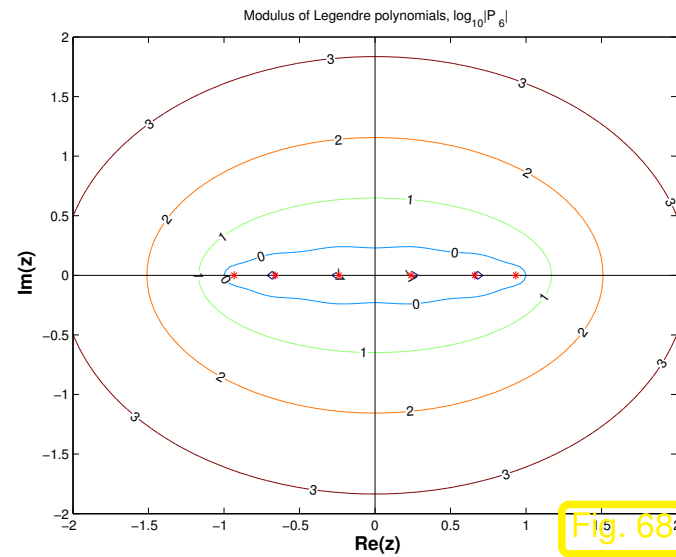
$$P_6(x) = \frac{1}{16}(231x^6 - 315x^4 + 105x^2 - 5).$$



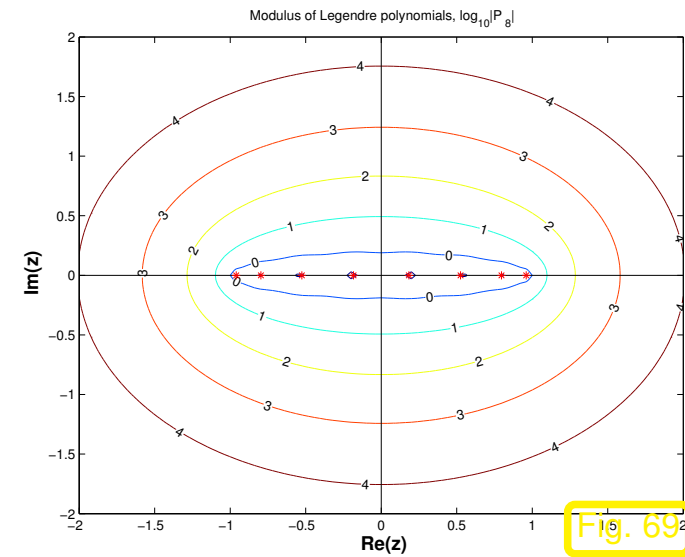
► Vermutung:  $|P_n(t)| \leq 1$  für alle  $-1 \leq t \leq 1$



Niveaus von  $|P_4(z)|$



Niveaus von  $|P_6(z)|$



Niveaus von  $|P_8(z)|$

Beobachtung/Vermutung:

- Niveaulinien von  $|P_n(z)|$  sind näherungsweise *Ellipsen* mit Brennpunkten  $-1, 1$ .
- Exponentielles Anwachsen von  $|P_n(z)|$  auf sich ausweitenden Ellipsen mit Brennpunkten  $\{-1, 1\}$ .

Hilfsmittel für Abschätzung der Legendre-Polynome nach oben und nach unten:

**Erzeugende Funktion** der Legendre-Polynome

$$\text{formale Reihe: } F_w(z) = \sum_{n=0}^{\infty} P_n(w)z^n, \quad z, w \in \mathbb{C}. \quad (2.2.70)$$

Durch gliedweise Differentiation:

$$\frac{dF_w}{dz}(z) = \sum_{n=0}^{\infty} (n+1)P_{n+1}(w)z^n, \quad \frac{d}{dz}(zF_w(z)) = \sum_{n=0}^{\infty} (n+1)P_n(w)z^n, \quad z\frac{d}{dz}(zF_w(z)) = \sum_{n=0}^{\infty} nF_w(z)z^n$$

Aus der Rekursionsformel (2.2.68) folgt daher

$$\begin{aligned} \frac{dF_w}{dz}(z) - \left(2z\frac{d}{dz}(zF_w(z)) - zF_w(z)\right) + z\frac{d}{dz}(zF_w(z)) &= 0, \\ \frac{dF_w}{dz}(z) &= \frac{w-z}{z^2 - 2wz + 1}F_w(z). \end{aligned} \quad (2.2.71)$$

Nach Ersetzung  $z \leftarrow t$ : ODE für  $t \mapsto F_w(t)$ .

Zugehöriges Anfangswertproblem mit  $F_w(0) = P_0(w) = 1$  hat eindeutige Lösung

$$F_w(z) = \left(z^2 - 2wz + 1\right)^{-1/2}.$$

Dabei wurde im Sinne der komplexen Fortsetzung wieder ersetzt  $t \leftarrow z$ . Also gilt für festes  $w \in \mathbb{C}$  und  $|z|$  hinreichend klein

$$\left(z^2 - 2wz + 1\right)^{-1/2} = \sum_{n=0}^{\infty} P_n(w)z^n. \quad (2.2.72)$$

**Erzeugende Funktion** der Legendre-Polynome

Faktorisierung von  $F_w(z)$ : mit

$$w := \frac{1}{2}(\zeta + \zeta^{-1}), \quad \zeta \in \mathbb{C} \setminus \{0\} \Rightarrow z^2 - 2wz + 1 = (1 - z\zeta)(1 - z/\zeta). \quad (2.2.73)$$

Aus Taylorreihe für  $(1 - z)^{-1/2}$ : ( $\rightarrow$  Analysis)

$$(1 - z)^{-1/2} = \sum_{n=0}^{\infty} a_n z^n, \quad a_n = \frac{(2n)!}{(n!)^2 2^{2n}} = \frac{1 \cdot 2 \cdot \dots \cdot 2n}{(2 \cdot 4 \cdot \dots \cdot 2n)^2} > 0.$$

Aus dem Multiplikationssatz für Potenzreihen mit Transformation (2.2.73)

$$\left(z^2 - 2wz + 1\right)^{-1/2} = (1 - z\zeta)^{-1/2}(1 - z/\zeta)^{-1/2} = \sum_{n=0}^{\infty} \underbrace{\left(\sum_{j=0}^n a_j a_{n-j} \zeta^{n-2j}\right)}_{=P_n(w)} z^n.$$



**Lemma 2.2.74** (Obere Schranke für Legendre-Polynome).

Es gilt  $|P_n(t)| \leq 1$  für alle  $-1 \leq t \leq 1$ .

*Beweis.* Wir verwenden die Darstellung des  $n$ . Legendre-Polynoms aus der Reihenentwicklung der erzeugenden Funktion: mit der **Joukowski-Transformation**  $T(\zeta) = \frac{1}{2}(\zeta + \zeta^{-1})$  haben wir

$$P_n(T(\zeta)) = \sum_{j=0}^n a_j a_{n-j} \zeta^{n-2j}, \quad |\zeta| \geq 1. \quad (2.2.75)$$

Beachte:  $\frac{1}{2}(e^{i\varphi} + e^{-i\varphi}) = \cos \varphi =: t$

$\Rightarrow$  Für  $T(\zeta) := \frac{1}{2}(\zeta + \zeta^{-1})$  gilt:  $T(\{|z| = 1\}) = [-1, 1]$ .

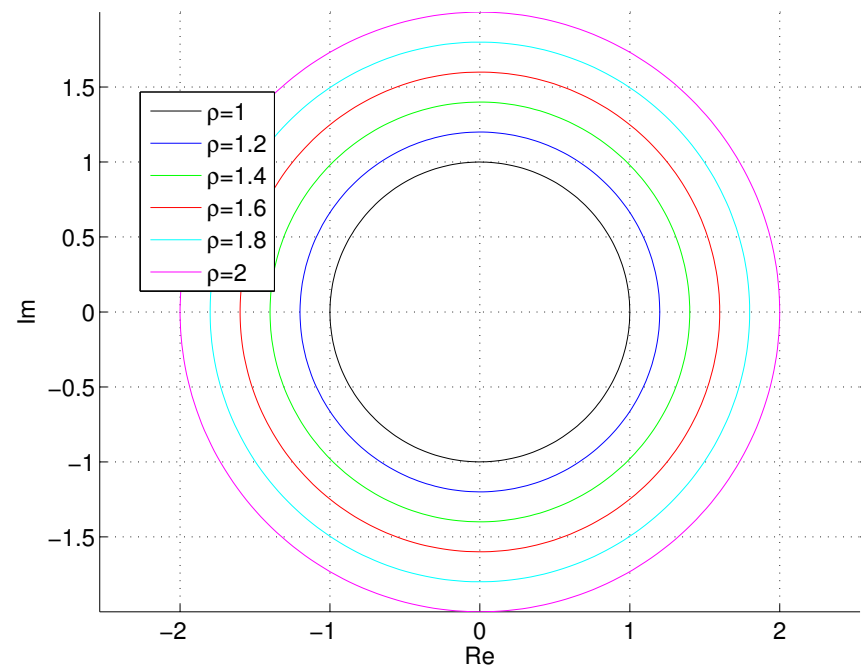
$$\Rightarrow |P_n(t)| \stackrel{(2.2.75)}{=} \left| \sum_{j=0}^n a_j a_{n-j} \exp(i(n-2j)\varphi) \right| \stackrel{a_j > 0}{\leq} \sum_{j=0}^n a_j a_{n-j} = P_n(1) = 1,$$

für ein  $\varphi \in [0, 2\pi]$  so, dass  $T(\exp(i\varphi)) = t \in [-1, 1]$ .

Wir betrachten nun die **Joukowski-Transformation**  $T(z) = \frac{1}{2}(z + z^{-1})$  in Ihrer Wirkung auf Kreise um 0 etwas näher

$$T(\rho e^{i\varphi}) = \frac{1}{2} (\rho e^{i\varphi} + \rho^{-1} e^{-i\varphi}) = \underbrace{\frac{1}{2}(\rho + \rho^{-1})}_{\text{grosse Halbachse}} \cos(\varphi) + i \cdot \underbrace{\frac{1}{2}(\rho - \rho^{-1})}_{\text{kleine Halbachse}} \sin(\varphi).$$

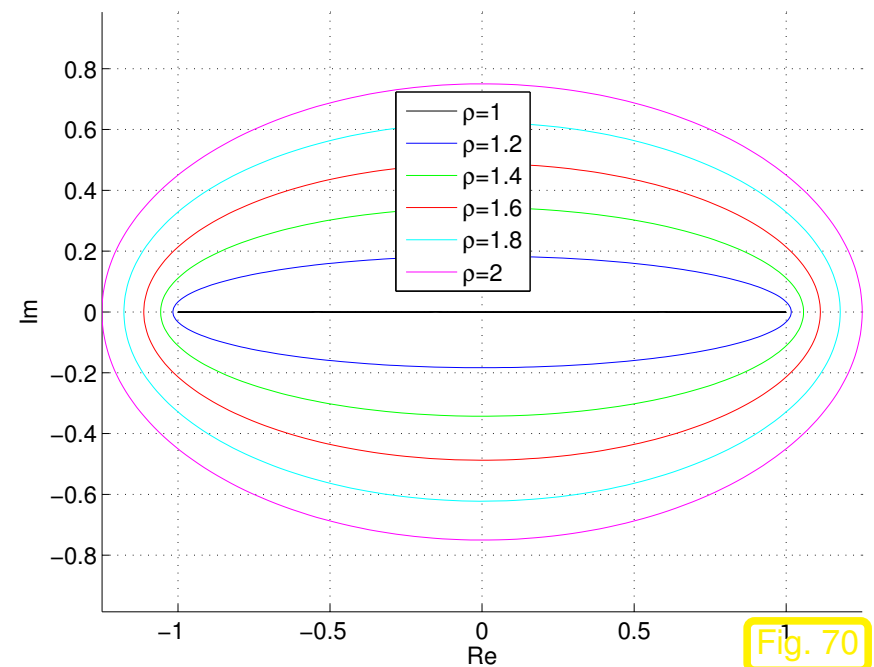
➤  $T$  bildet einen Kreis mit Radius  $\rho > 1$  auf eine Ellipse mit Brennpunkten  $\{-1, 1\}$ , kleiner Halbachse  $\frac{1}{2}(\rho - \rho^{-1})$  und grosser Halbachse  $\frac{1}{2}(\rho + \rho^{-1})$  ab.



Kreise  $\{z \in \mathbb{C} : |z| = \rho\}$

Transformation  $T$

$$z \rightarrow \frac{1}{2}(z + 1/z)$$



Ellipsen  $E_\rho$  gemäss (2.2.76)

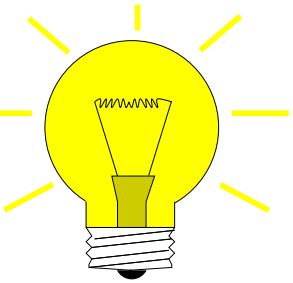
Fig. 70

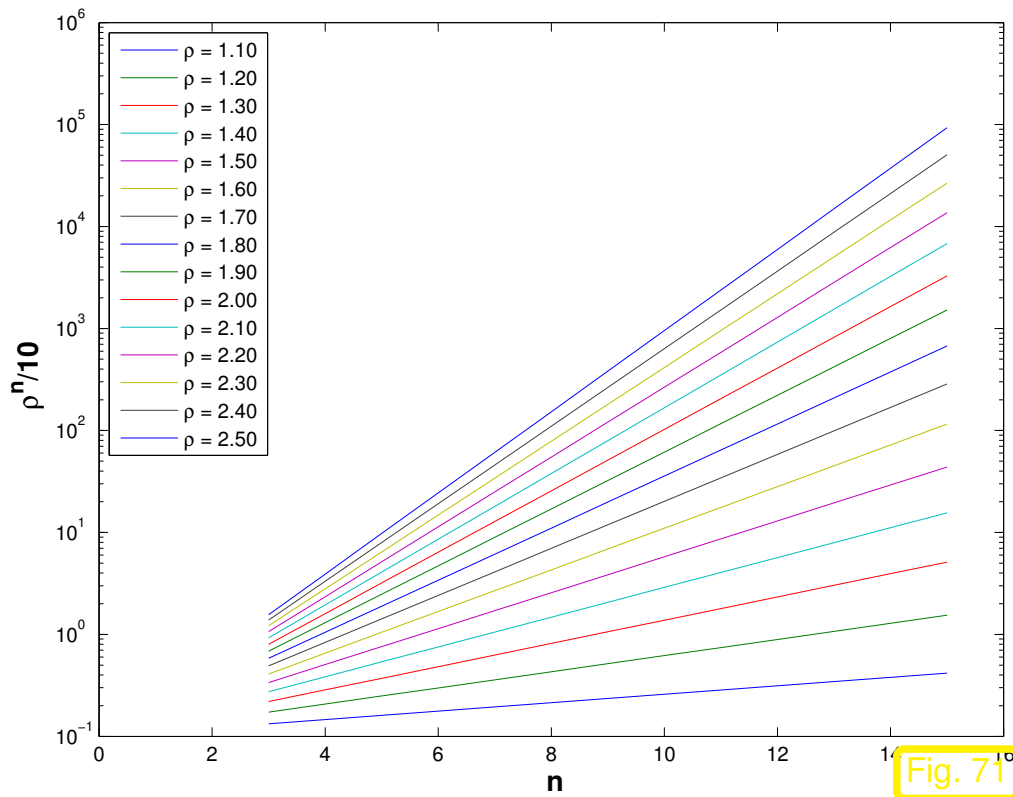
Abbildungen 67-69: Niveaulinien von  $|P_n(z)|$  sind näherungsweise Ellipsen mit Brennpunkten  $\{-1, 1\}$ .

Idee: Benutze **elliptische Integrationswege**, siehe Abb. 70

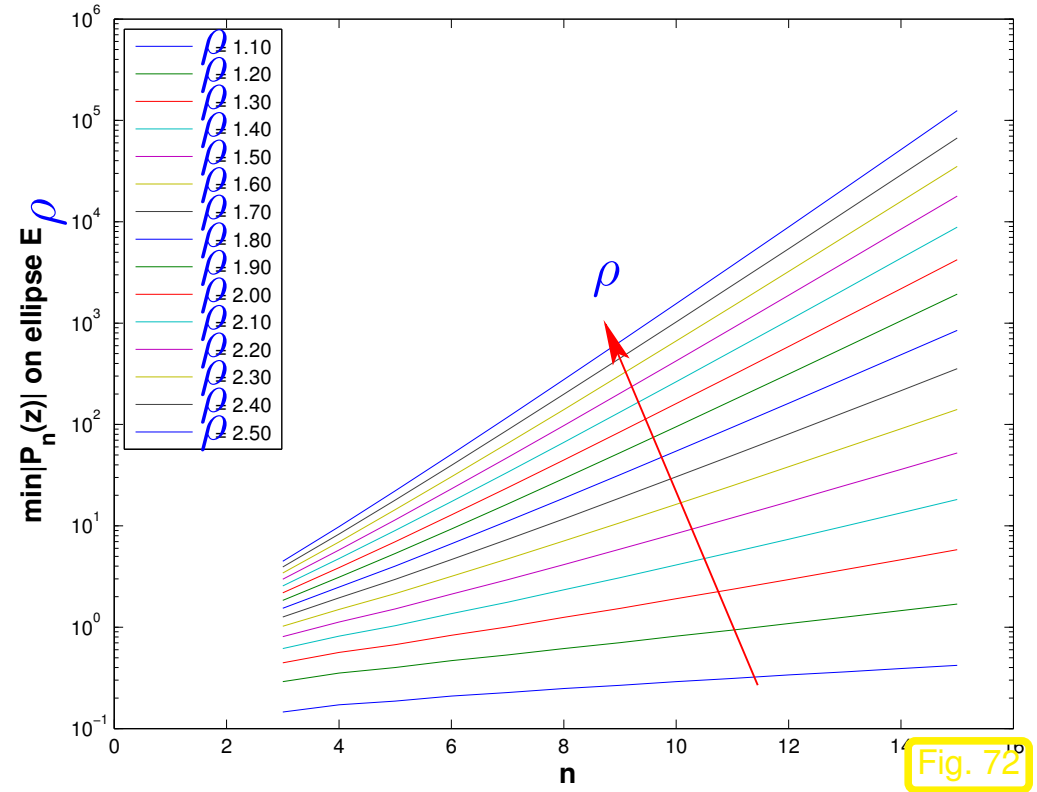
$$E_\rho := T(\{z \in \mathbb{C}, |z| = \rho\}), \quad \rho > 1, \quad (2.2.76)$$

mit **Joukowski-Transformation**  $T(z) := \frac{1}{2}(z + 1/z)$ . (2.2.77)

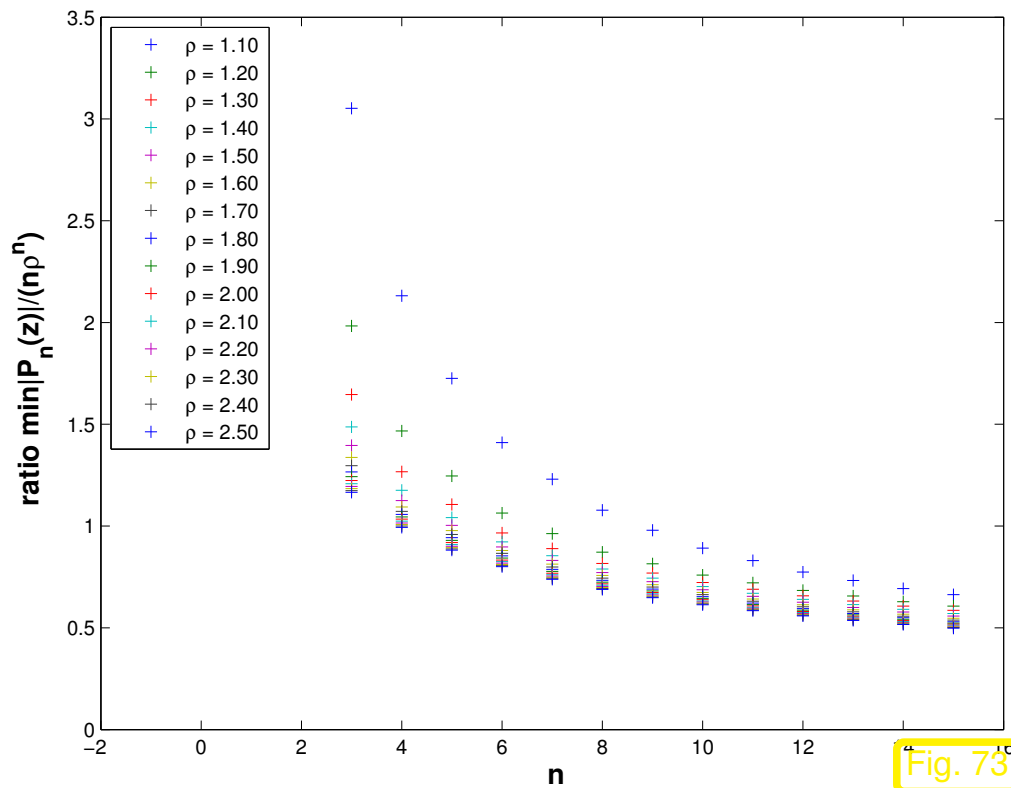




$0.1 \cdot \rho^n$  als Funktion von  $n$  für verschiedene  $\rho > 1$



Approximation von  $\min_{z \in E_\rho} |P_n(z)|$



Vermutung:

Für grosse  $n$ :

$$\min_{z \in E_\rho} |P_n(z)| \sim n\rho^n. \quad (2.2.78)$$

R. Hiptmair

rev 35327,  
25. April  
2011

Nach bestem Wissen des Autors gibt es keinen Beweis von (2.2.78). In [5, Sect. 12.4], [4] wird die schwächere Behauptung

$$\forall \epsilon > 0: \quad \exists N = N(\epsilon): \min_{z \in E_\rho} |P_n(z)| \geq (\rho - \epsilon)^n \quad \forall n > N(\epsilon)$$

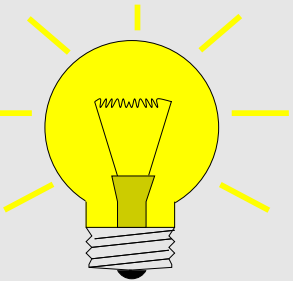
gezeigt.

Hier nur *Heuristik*:

Erinnerung: Formel von Cauchy-Hadamard für den **Konvergenzradius** einer **Potenzreihe** [30, Sect. 4.1.3]

$$R := \frac{1}{\limsup_{n \rightarrow \infty} |a_n|^{1/n}} \Rightarrow \sum_{n=0}^{\infty} a_n z^n \text{ konvergiert für } |z| < R. \quad (2.2.79)$$

Idee: (2.2.79) liefert asymptotische untere Schranke für  $|a_n|$ , wenn  $R$  bekannt!  
(Voraussetzung ist  $\limsup_{n \rightarrow \infty} |a_n|^{1/n} = \lim_{n \rightarrow \infty} |a_n|^{1/n}$ , gegeben z.B. bei Monotonie von  $(a_n)_n$ .)



Aus dem Entwicklungssatz von Cauchy schliesst man, siehe [30, Sect. 8.1.5]:

**Lemma 2.2.80** (Konvergenzradius von Potenzreihenentwicklungen).

Sei  $D \subset \mathbb{C}$  offen,  $f : D \mapsto \mathbb{C}$  holomorph und  $0 \in D$ . Dann hat die Taylorreihe von  $f$  um  $0$  den Konvergenzradius  $R = \text{dist}(0, \partial D)$ .

► Zu untersuchen mit Hilfe von Lemma 2.2.80: Konvergenzradius der Taylorreihe um  $0$  der erzeugenden Funktion (2.2.72), also  $f(z) = (z^2 - 2wz + 1)^{-1/2}$ , der Legendre-Polynome aus (2.2.72).

Uns interessiert:  $\min_{z \in E_\rho} |P_n(z)|$   $\rightsquigarrow$  Gesucht: Konvergenzradius für  $w \in E_\rho$

$$\text{Lemma 2.2.80} \Rightarrow R = \min\{\text{dist}(0, \zeta), \text{dist}(0, \zeta^{-1})\},$$

denn  $\zeta, \zeta^{-1}$  sind die beiden Nullstellen von  $z \mapsto z^2 - 2wz + 1$  für  $w = \frac{1}{2}(\zeta + \zeta^{-1})$ , vgl. (2.2.73).

$$\blacktriangleright \quad \zeta = \rho \exp(i\varphi), \quad \rho > 1 \Rightarrow R = \rho^{-1}.$$

**Annahme.** Existenz des Limes:  $\limsup_{n \rightarrow \infty} |P_n(w)|^{1/n} = \lim_{n \rightarrow \infty} |P_n(w)|^{1/n}$

Damit aus (2.2.72):

$$\forall \epsilon > 0: \exists N = N(\epsilon) \in \mathbb{N}: \left( |P_n(w)|^{1/n} > \rho - \epsilon \Leftrightarrow |P_n(w)| > (\rho - \epsilon)^n \right) \quad \forall n > N(\epsilon).$$

**Theorem 2.2.81** (Fehlerabschätzung für Interpolation in Gauss-Knoten).

Es sei  $f : [-1, 1] \mapsto \mathbb{C}$  nach  $D \subset \mathbb{C}$  analytisch fortsetzbar und  $E_\rho \subset D$  für ein  $\rho > 1$ . Unter der Annahme (2.2.78) finden wir  $N = N(\rho) \in \mathbb{N}$  und  $C = C(\rho) > 0$ , so dass

$$\max_{-1 \leq t \leq 1} \left| f(t) - \sum_{j=1}^s f(\tau_j) L_j(t) \right| \leq C \frac{\rho^{-s}}{s} \max_{z \in E_\rho} |f(z)| \frac{\text{length}(E_\rho)}{2\pi} \quad \forall s > N(\rho).$$

Aussage von Theorem 2.2.81: bzgl. der Maximumnorm **exponentielle Konvergenz** ( $\rightarrow$  Def. 1.4.5) des Interpolationspolynoms in den Gausspunkten einer in einer “geeigneten” Umgebung von  $[-1, 1]$  analytischen Funktion.

*Beispiel 2.2.82* (Fehler bei Polynominterpolation in Gauss-Knoten).  $\rightarrow$  Fortsetzung Bsp. 2.2.59

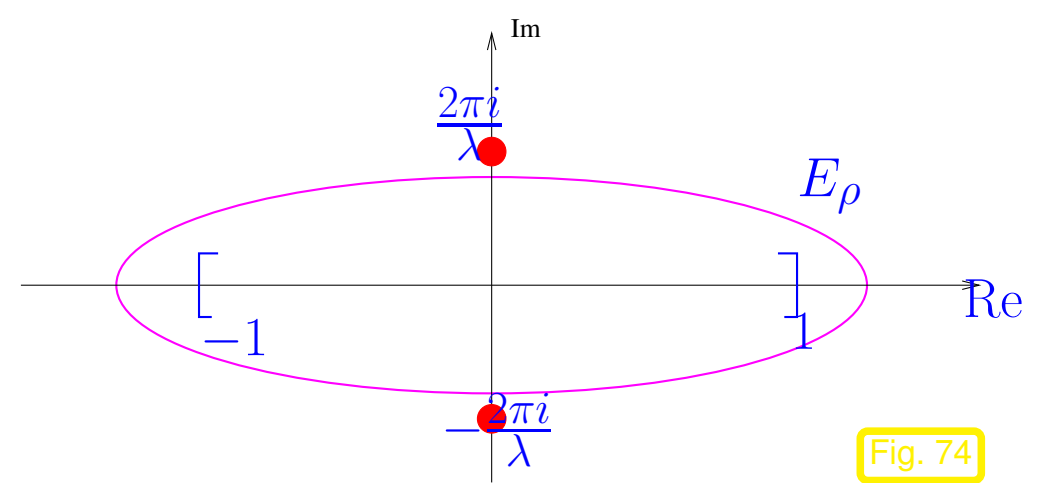
$$f(z) = \frac{1}{1 + \exp(-\frac{1}{2}\lambda z)} \Rightarrow f \text{ analytisch in } E_\rho \text{ für } \rho < \frac{\pi}{\lambda} + \sqrt{\left(\frac{\pi}{\lambda}\right)^2 + 1}.$$



Pole von  $f$ :

$$\pm(2k + 1)\frac{2\pi i}{\lambda}, \quad k \in \mathbb{Z}.$$

$$\Rightarrow \rho - \rho^{-1} < \frac{2\pi}{\lambda}.$$



Dieses Beispiel demonstriert die allgemeine Strategie zum Auffinden zulässiger Analytizitätsellipsen für “einfache” Funktionen: Man bestimmt die Pole der zu untersuchenden Funktion in  $\mathbb{C}$  und dadurch das Gebiet, auf dem die Funktion holomorph ist.

R. Hiptmair  
rev 35327,  
25. April  
2011



*Beispiel 2.2.83* (Analytizitätsgebiet für Lösung der logistischen Dgl.).

Logistische Differentialgleichung ( $\rightarrow$  Bsp. 1.2.1)

$$\dot{y} = \lambda y(1 - y), \quad y_0 > 0[ \Rightarrow y(t) = \frac{1}{1 + (y_0^{-1} - 1)e^{-\lambda t}}, \quad t \in \mathbb{R}. \quad (2.2.84)$$

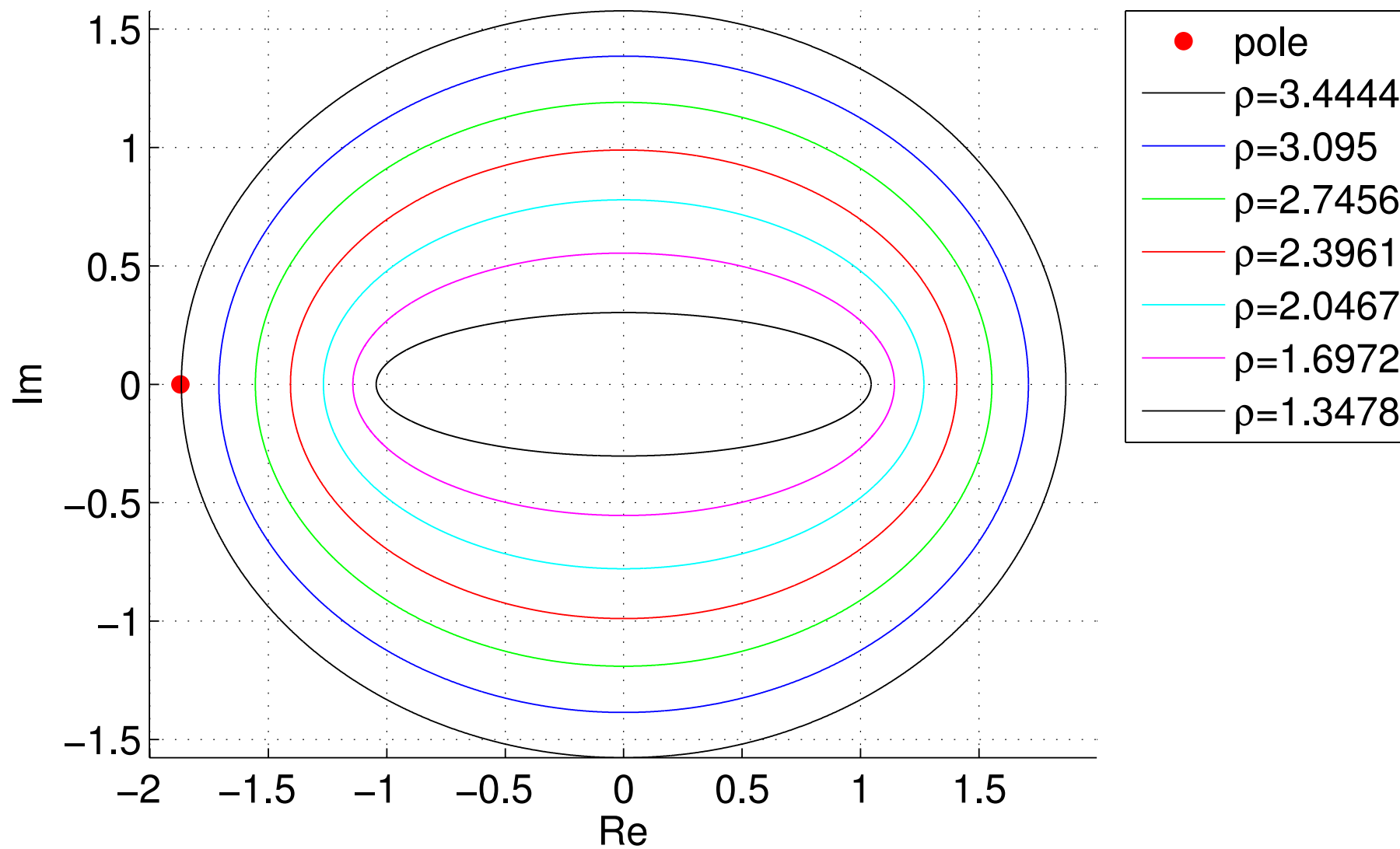
Wenden wir ein Gauss-Kollokations-Einschrittverfahren, so reicht es

$$\|(Id - P)\mathbf{f}(\cdot, \mathbf{y}(\cdot))\|_{\infty, [0,1]} = \left\| (Id - P)\mathbf{f}\left(\frac{\cdot+1}{2}, \mathbf{y}\left(\frac{\cdot+1}{2}\right)\right) \right\|_{\infty, [-1,1]}$$

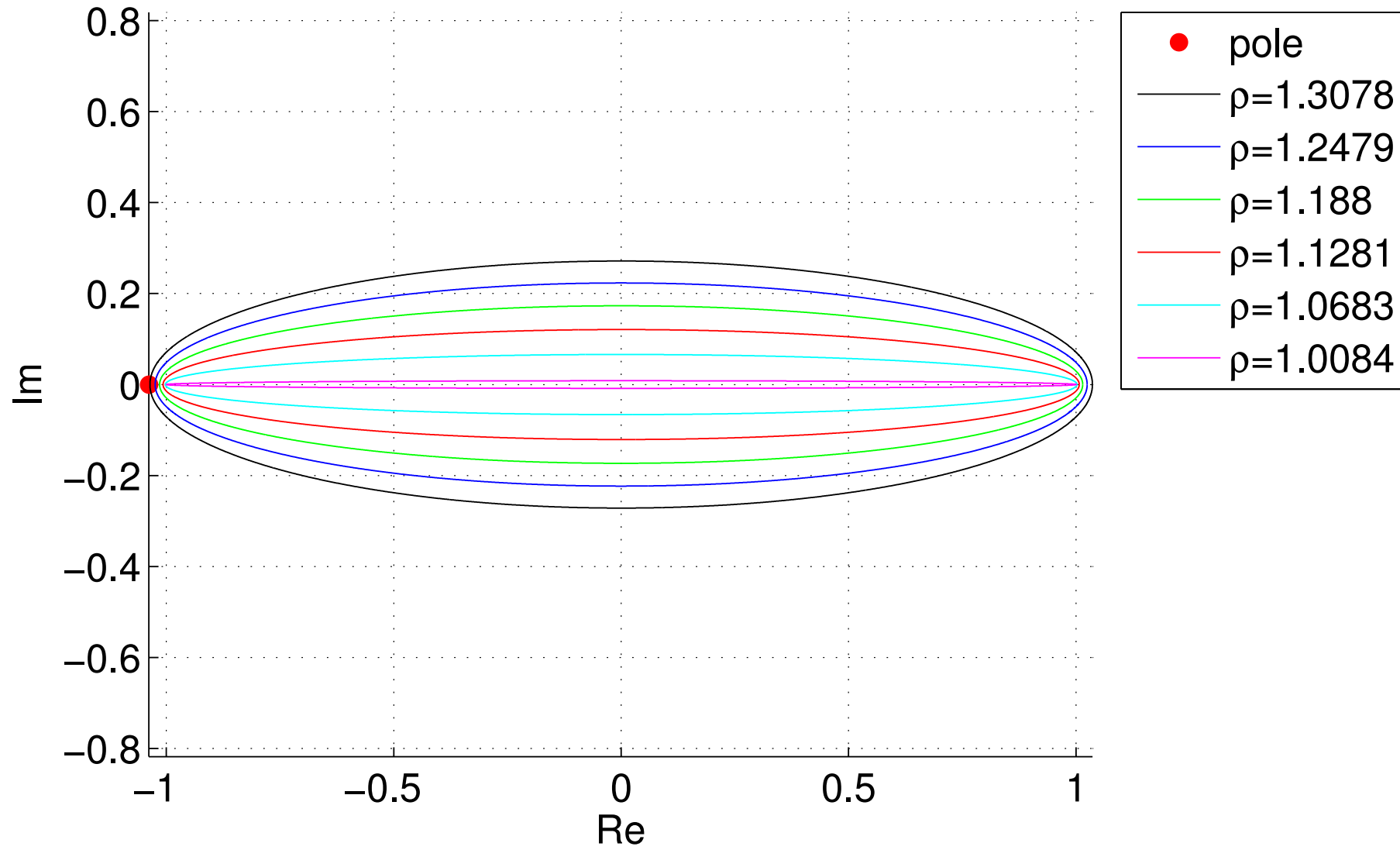
zu untersuchen, siehe Thm. 2.2.30, wobei  $P$  den Operator der Polynominterpolation in den Gausspunkten bezeichnet.

Für  $y_0 > 1$ : ein Pol in  $-1 - \frac{2}{\lambda} \ln(-a)$  mit  $1/a = 1/y_0 - 1$ .

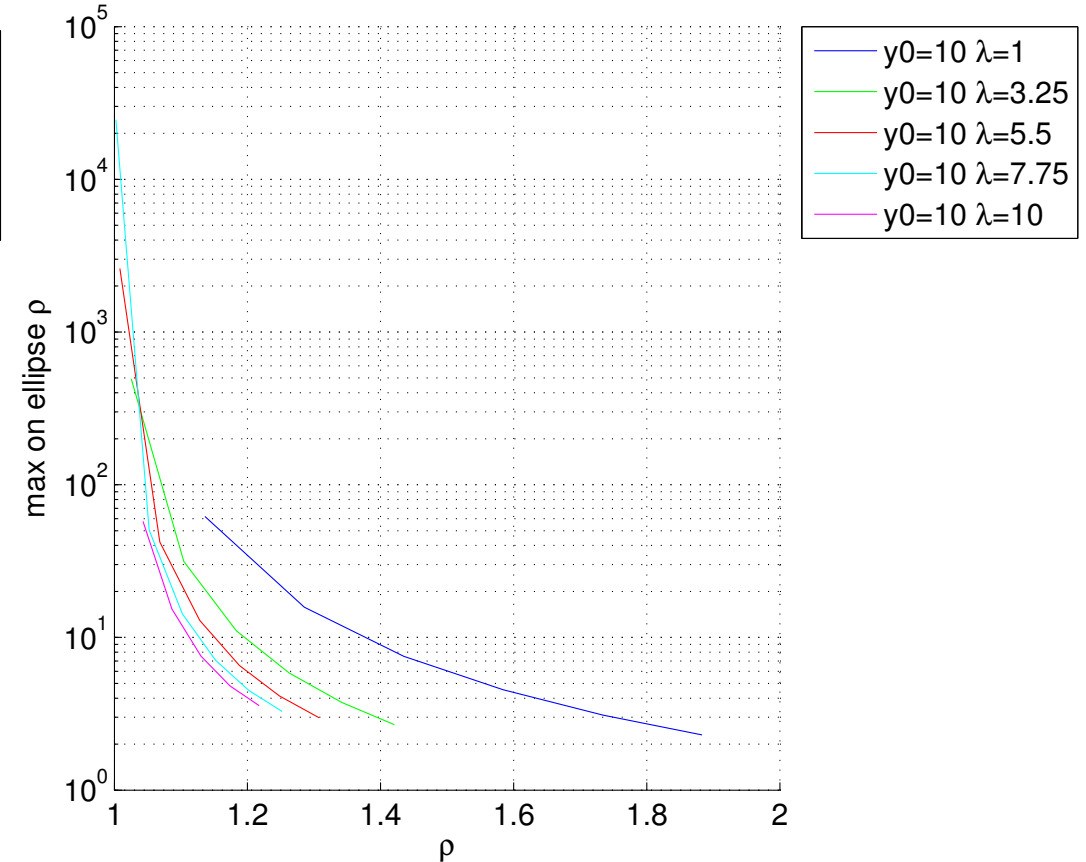
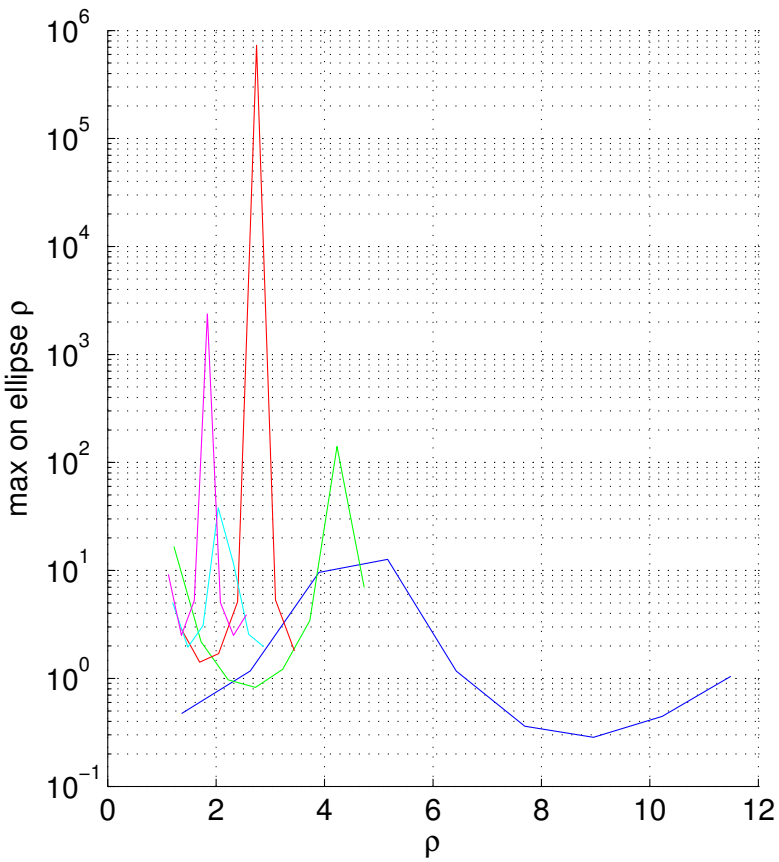
$y_0 = 1.1, \lambda = 5.5:$



$$y_0 = 10, \lambda = 5.5:$$



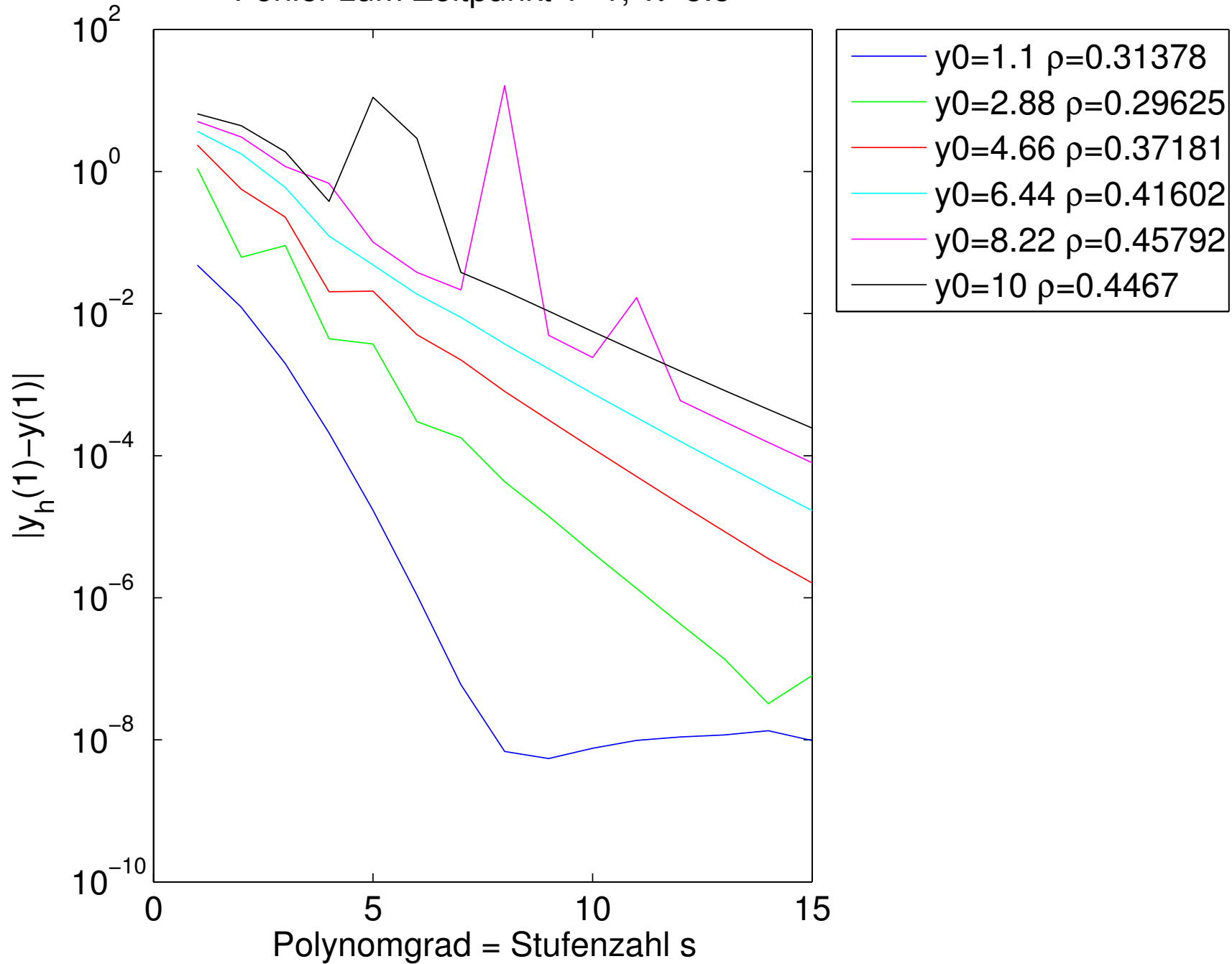
Maximum auf Ellipsen:



Fehler im Endzeitpunkt für globalen Schritt des Gauss-Kollokations-Einschrittverfahrens:

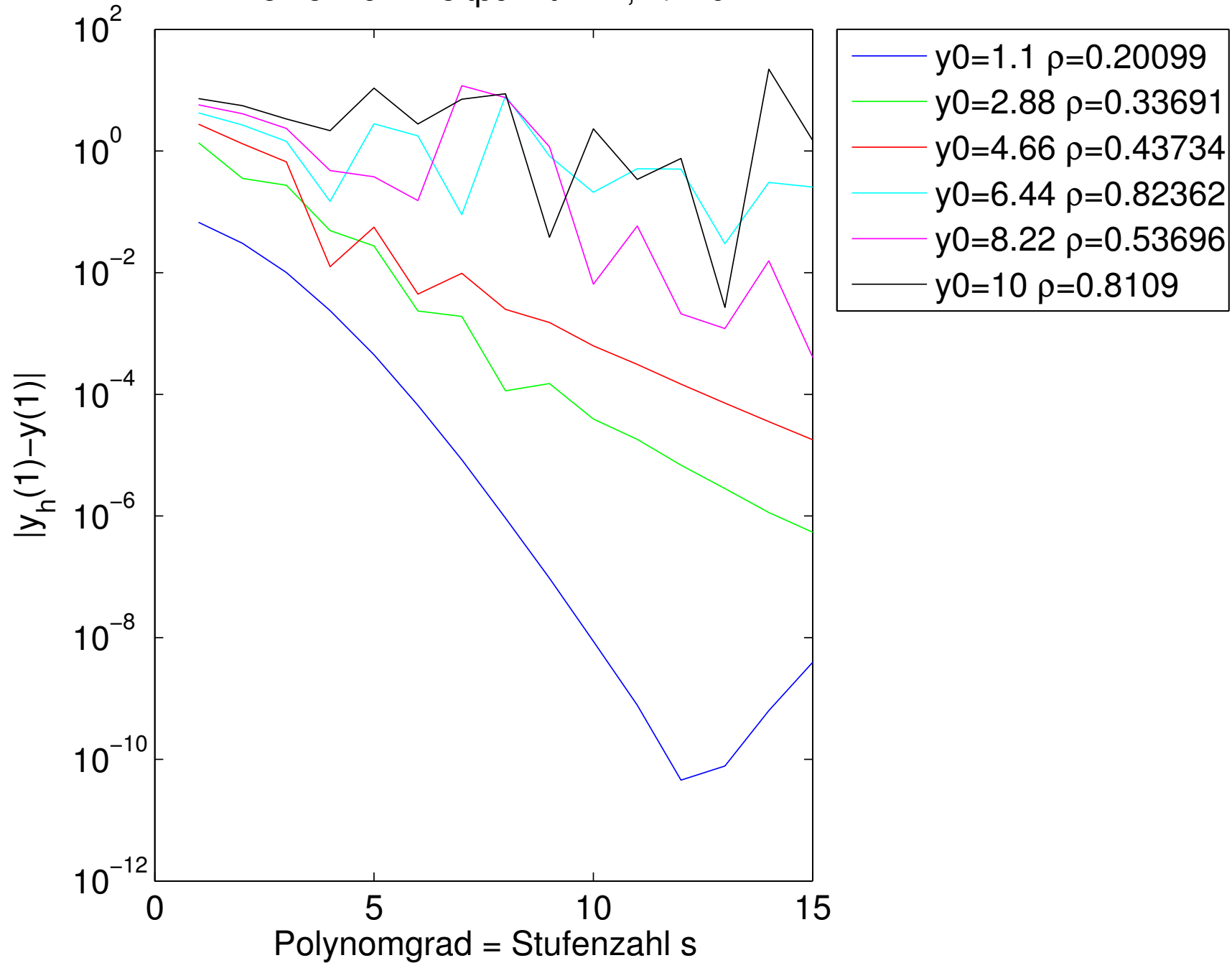
$\lambda = 5.5:$

Fehler zum Zeitpunkt T=1,  $\lambda=5.5$



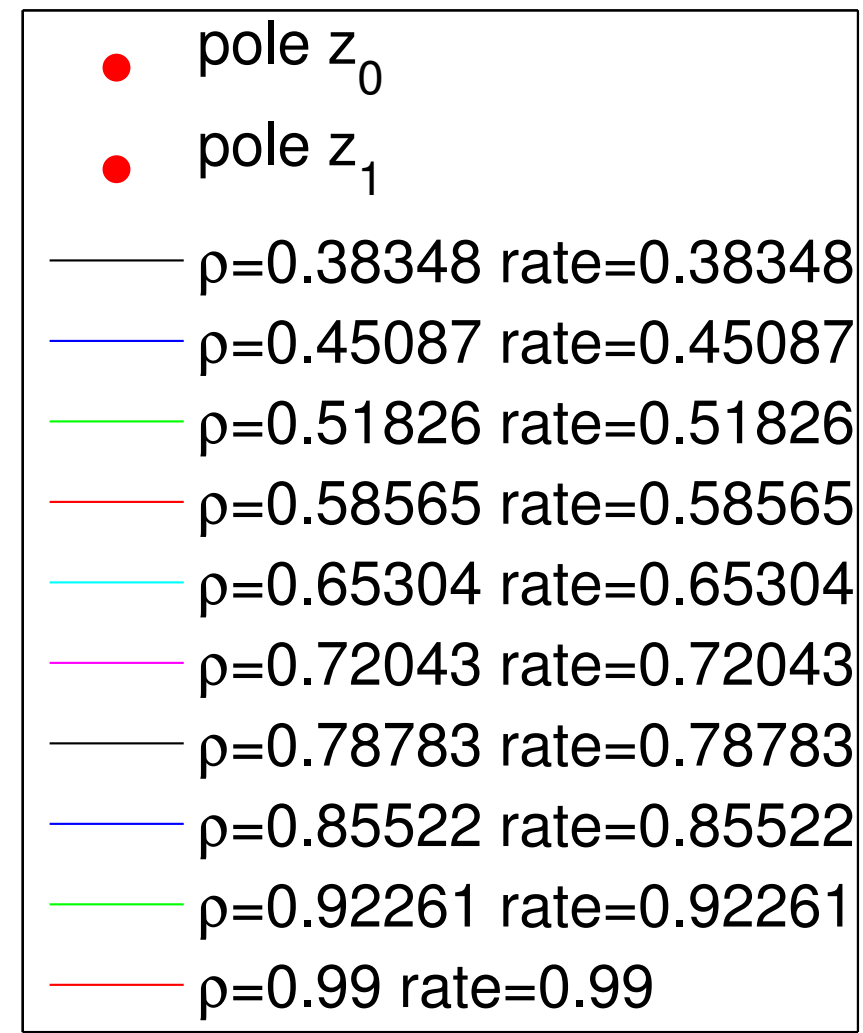
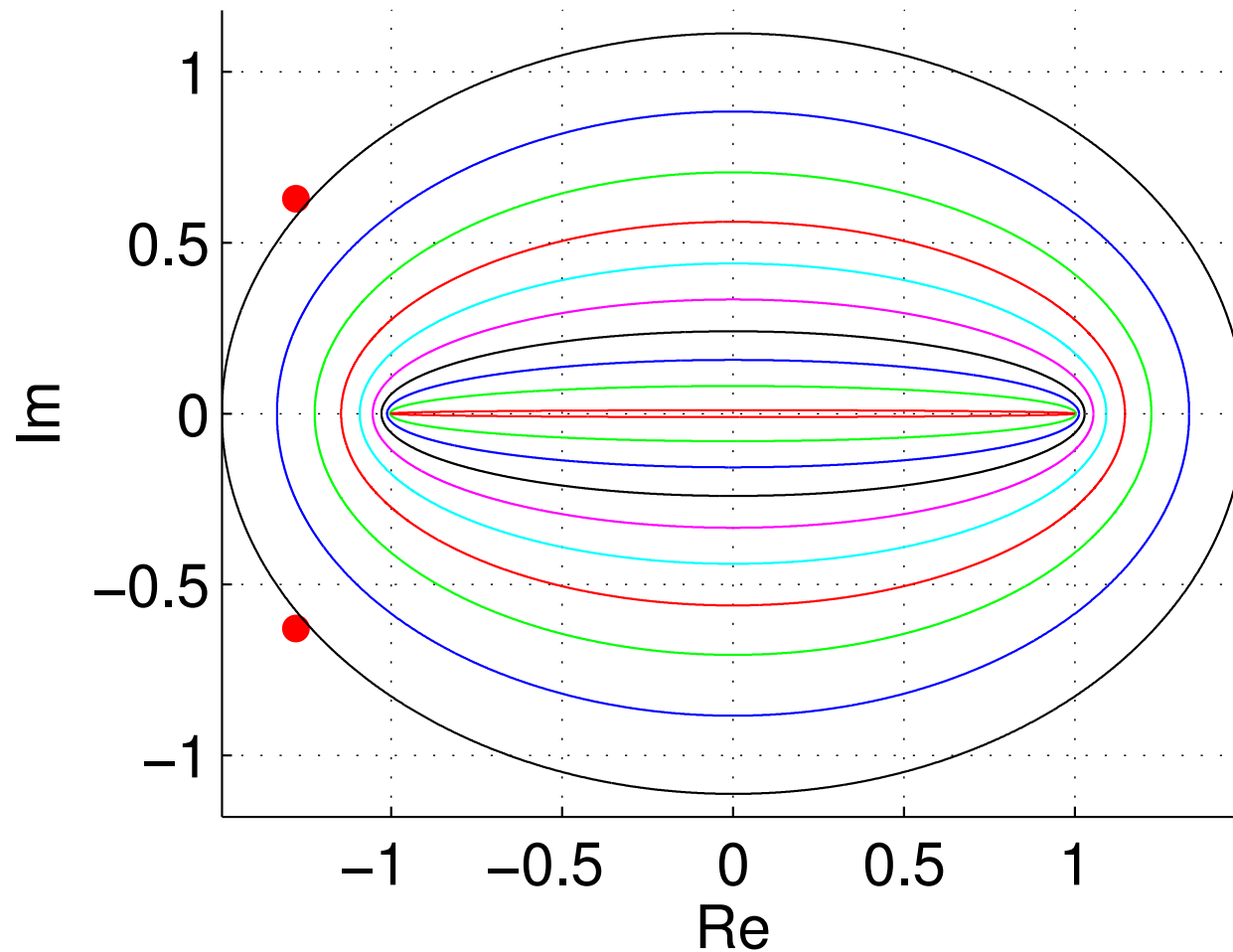
$\lambda = 10:$

Fehler zum Zeitpunkt  $T=1$ ,  $\lambda=10$



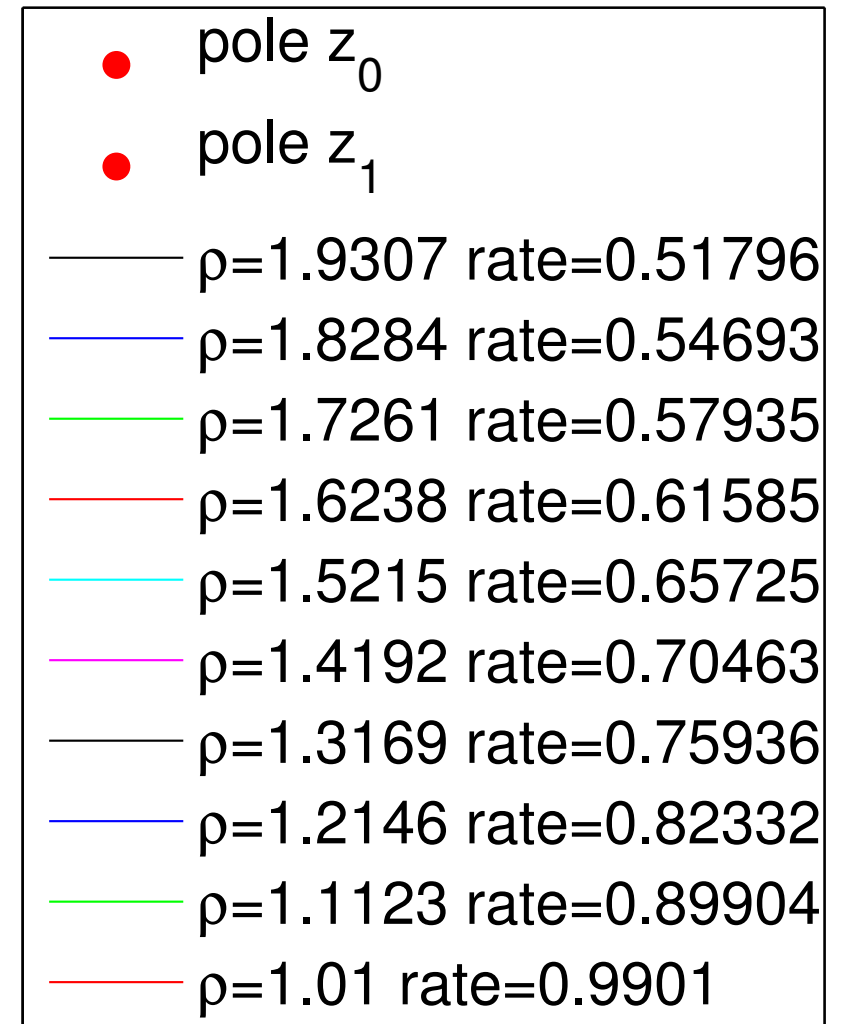
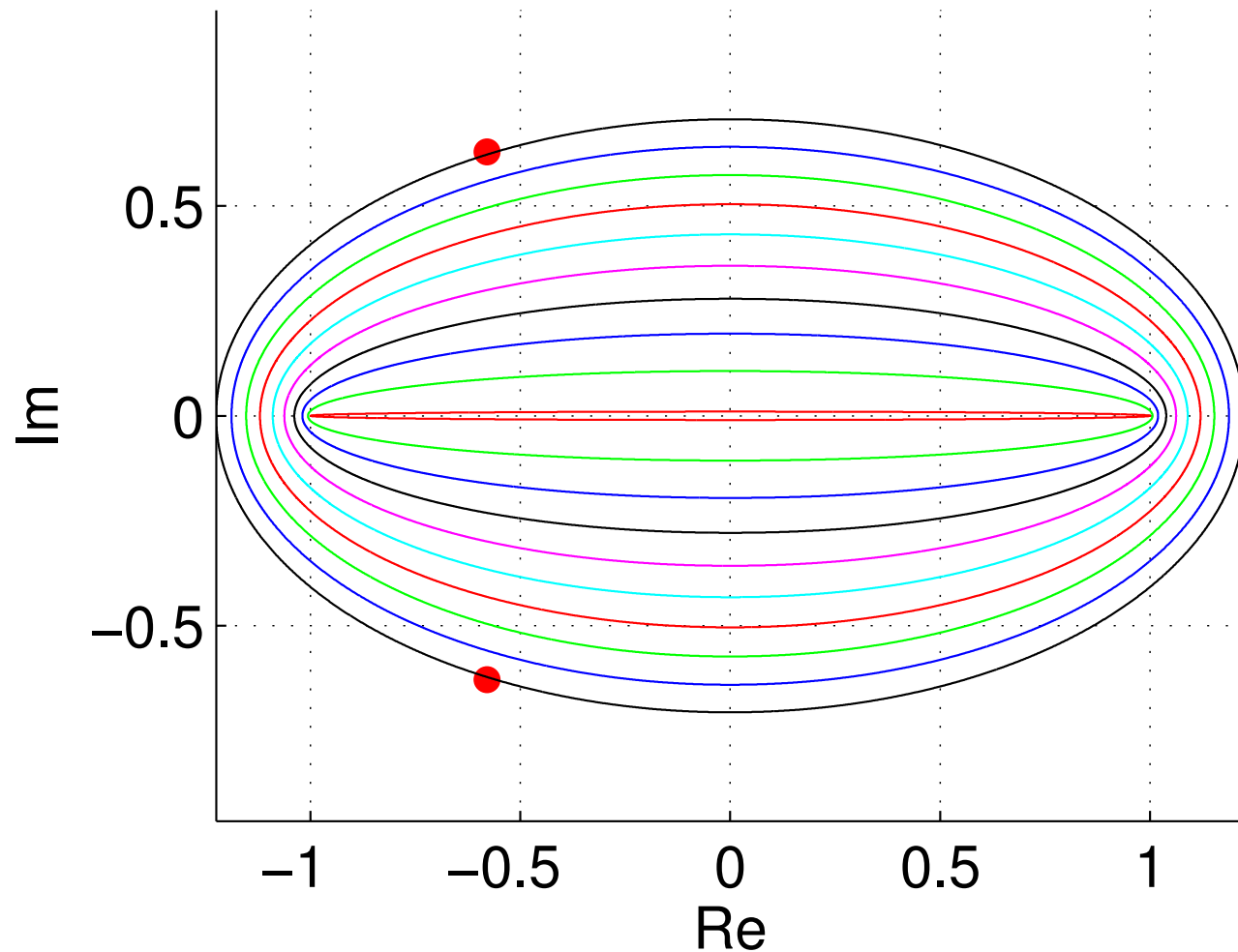
Für  $y_0 < 1$ : Pole in  $-1 - \frac{2}{\lambda} \ln(a) - \frac{2}{\lambda}(2k + 1)\pi i$  mit  $1/a = 1/y_0 - 1$ .

$y_0 = 0.80111, \lambda = 10$ :

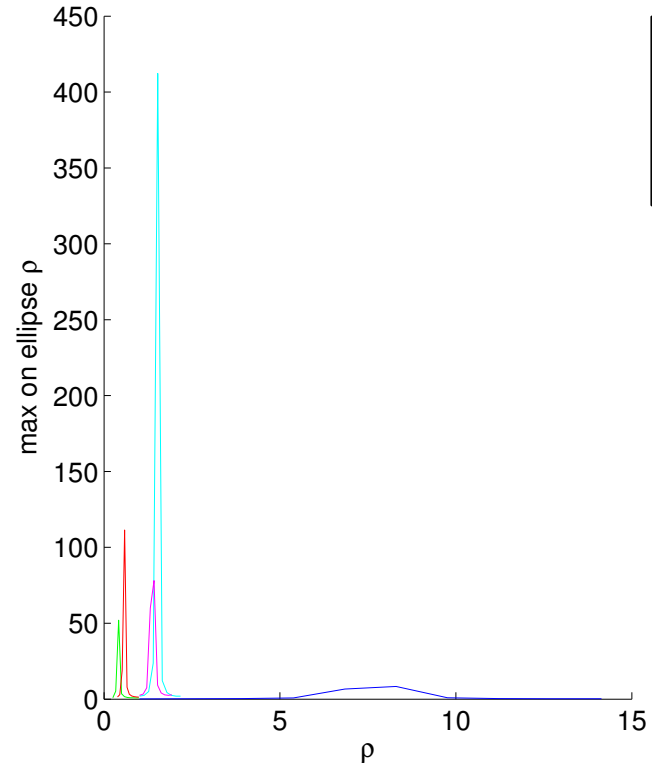
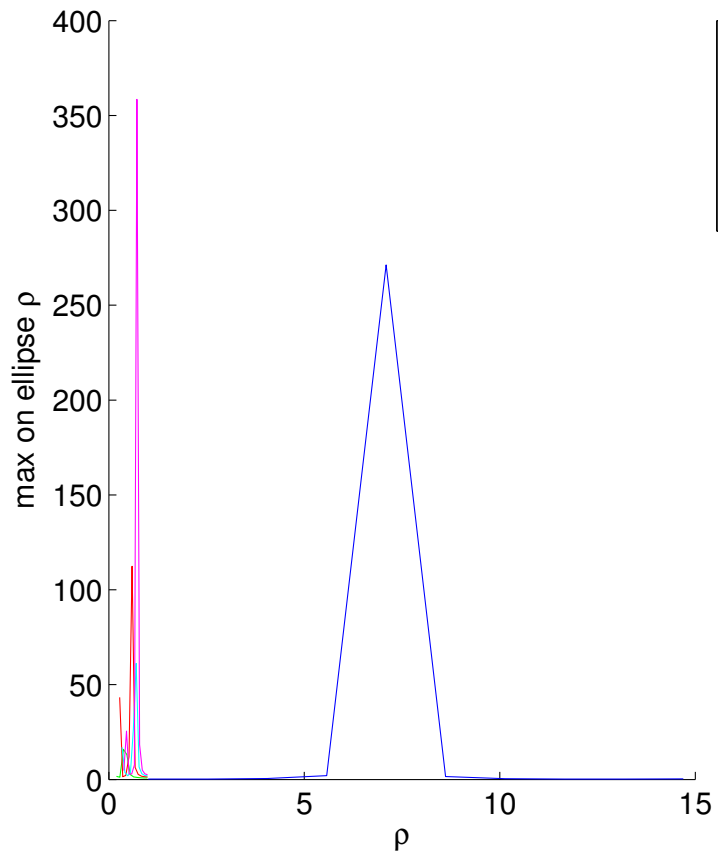




$y_0 = 0.10889, \lambda = 10:$



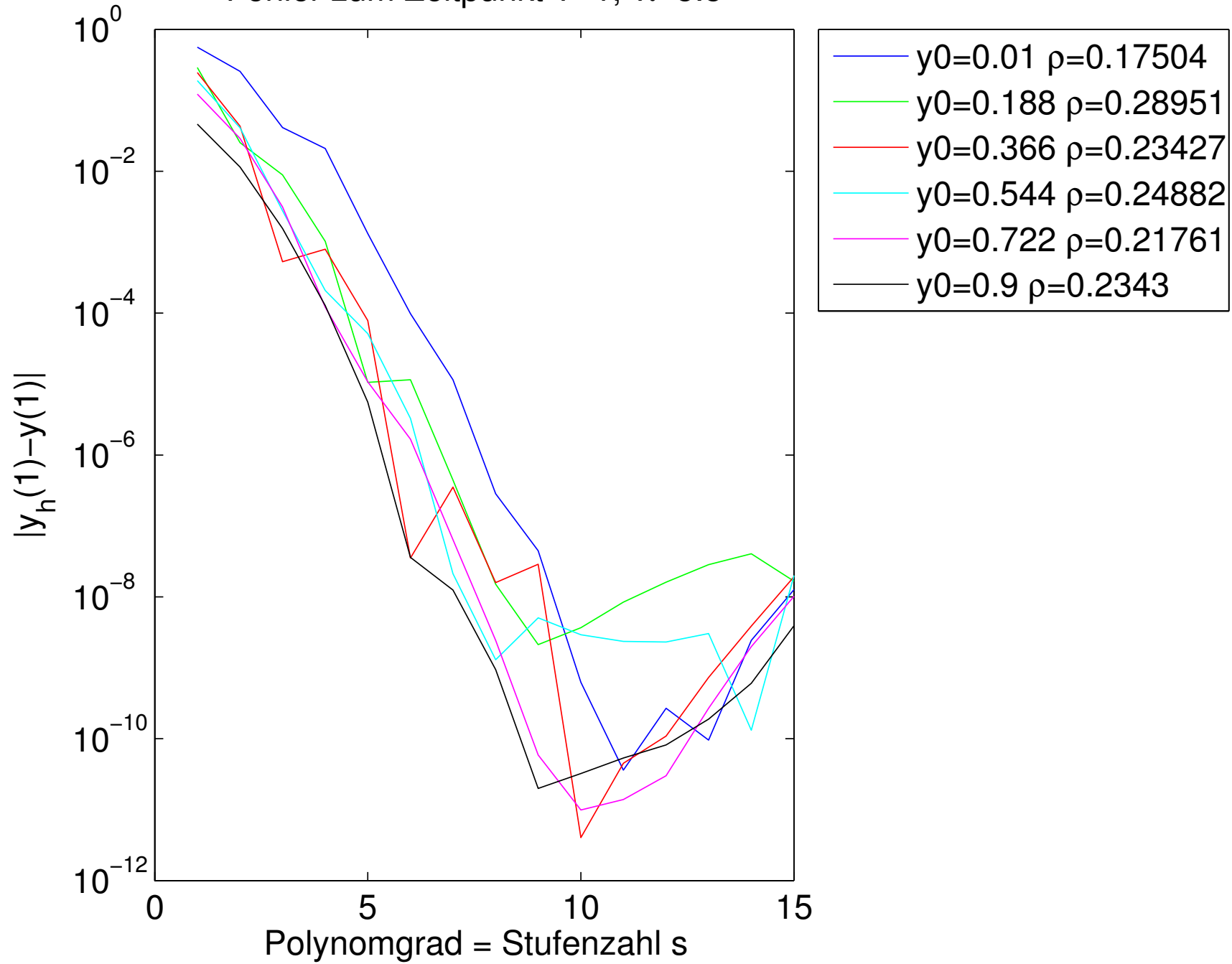
Maximum auf Ellipsen:



Fehler im Endzeitpunkt für globalen Schritt des Gauss-Kollokations-Einschrittverfahrens:

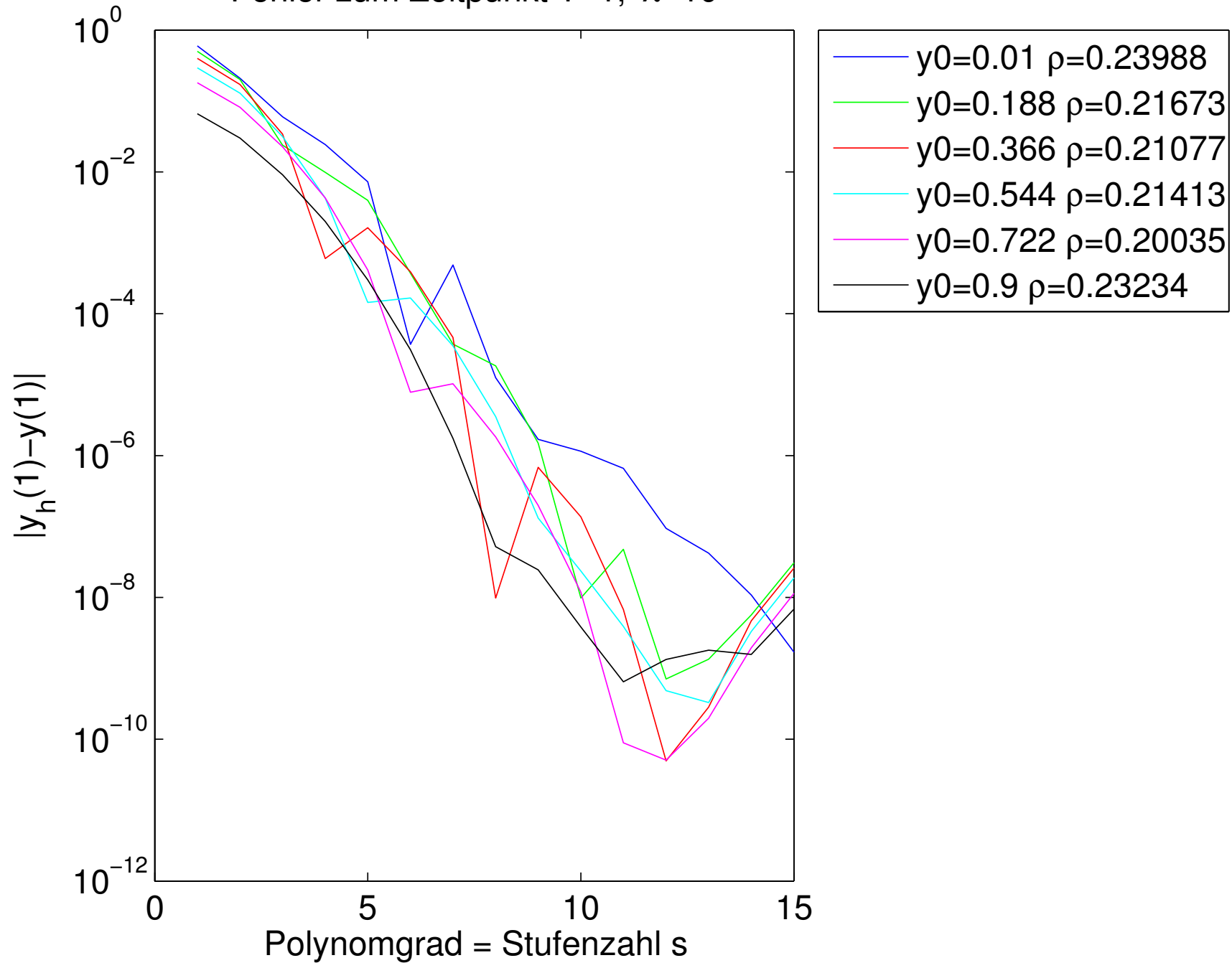
$\lambda = 5.5:$

Fehler zum Zeitpunkt T=1,  $\lambda=5.5$



$\lambda = 10:$

Fehler zum Zeitpunkt T=1,  $\lambda=10$



In Beispiel 2.2.83 konnten wir uns für die Bestimmung des Analytizitätsgebiets auf die explizit gegebene Lösung des AWP stützen, um die exponentielle Konvergenz des globalen Gauss-Kollokationsverfahrens zu bestätigen.

Im Allgemeinen fehlt diese Information. Dennoch sind Aussagen über das Analytizitätsgebiet der Lösungen von AWP für **Differentialgleichungen in  $\mathbb{C}$**  mit lokal holomorpher rechter Seite möglich:

**Theorem 2.2.85** (Existenz- und Eindeigkeitssatz für Dgl. in  $\mathbb{C}$ ).  $\rightarrow [32, \text{Kap. I, §8}]$

Ist  $f : D \subset \mathbb{C} \mapsto \mathbb{C}$  in einer Umgebung  $B_\rho(z_0) := \{z \in \mathbb{C} : |z - z_0| < \rho\} \subset D$  von  $z_0 \in D$  holomorph und  $|f(z)| \leq M$  für alle  $z \in B_\rho(z_0)$ , dann existiert genau eine auf  $B_{\rho/M}(0)$  holomorphe Lösung  $y$  des Anfangswertproblems

$$y'(z) = f(y(z)) \quad \forall z \in B_{\rho/M}(0) \quad , \quad y(0) = z_0 .$$

Notation:  $' \hat{=}$  komplexe Differentiation

Wenn  $f(z) \in \mathbb{R}$  für  $z \in \mathbb{R}$  und  $y_0 \in \mathbb{R}$  dann stimmt die vom Theorem postulierte lokal holomorphe Lösung des komplexen AWP für reelle Argumente natürlich mit der Lösung gemäss Theorem 1.3.4 überein.

Nachteil der Kollokationseinschrittverfahren: Alle (mit Ausnahme des expliziten Euler-Verfahrens) sind *implizit* ( $\rightarrow$  Def. 2.1.5)

Gibt es explizite Einschrittverfahren höherer Ordnung? Wenn ja, wie findet man diese?

## 2.3.1 Konstruktion

$$\text{AWP: } \begin{array}{l} \dot{\mathbf{y}}(t) = \mathbf{f}(t, \mathbf{y}(t)) \\ \mathbf{y}(t_0) = \mathbf{y}_0 \end{array} \Rightarrow \mathbf{y}(t_1) = \mathbf{y}_0 + \int_{t_0}^{t_1} \mathbf{f}(\tau, \mathbf{y}(t_0 + \tau)) \, d\tau$$

Approximation durch Quadraturformel (auf  $[0, 1]$ ) mit  $s$  Knoten  $c_1, \dots, c_s$ :

$$\mathbf{y}(t_1) \approx \mathbf{y}_1 (= \mathbf{y}_h(t_1)) = \mathbf{y}_0 + h \sum_{i=1}^s b_i \mathbf{f}(t_0 + c_i h, \mathbf{y}(t_0 + c_i h)), \quad h := t_1 - t_0.$$

Wie bekommt man diese Werte ?

$\triangleright$  **Bootstrapping**

**Beispiel 2.3.1** (Konstruktion einfacher Runge-Kutta-Verfahren).

Quadraturformel  $\rightarrow$  Trapezregel: auf Intervall  $[a, b]$

$$Q(f) = \frac{1}{2}(b-a)(f(a) + f(b)) \quad \leftrightarrow \quad s = 2: \quad c_1 = 0, c_2 = 1, \quad b_1 = b_2 = \frac{1}{2}, \quad (2.3.2)$$

und  $\mathbf{y}_h(T)$  aus explizitem Eulerschritt (1.4.2)

$$\mathbf{k}_1 = \mathbf{f}(t_0, \mathbf{y}_0), \quad \mathbf{k}_2 = \mathbf{f}(t_0 + h, \mathbf{y}_0 + h\mathbf{k}_1), \quad \mathbf{y}_1 = \mathbf{y}_0 + \frac{h}{2}(\mathbf{k}_1 + \mathbf{k}_2). \quad (2.3.3)$$

(2.3.3) = **explizite Trapezregel**

Quadraturformel  $\rightarrow$  einfachste Gauss-Quadraturformel (**Mittelpunktsregel**) &  $\mathbf{y}_h(\frac{1}{2}(t_1 + t_0))$  aus explizitem Eulerschritt (1.4.2)

$$\mathbf{k}_1 = \mathbf{f}(t_0, \mathbf{y}_0), \quad \mathbf{k}_2 = \mathbf{f}(t_0 + \frac{h}{2}, \mathbf{y}_0 + \frac{h}{2}\mathbf{k}_1), \quad \mathbf{y}_1 = \mathbf{y}_0 + h\mathbf{k}_2. \quad (2.3.4)$$

(2.3.4) = **explizite Mittelpunktsregel**



Diskrete Evolutionen der Form (2.2.3)

**Definition 2.3.5** (Runge-Kutta-Verfahren).

Für  $b_i, a_{ij} \in \mathbb{R}$ ,  $c_i := \sum_{j=1}^s a_{ij}$ ,  $i, j = 1, \dots, s$ ,  $s \in \mathbb{N}$ , definiert

$$\mathbf{k}_i := \mathbf{f}(t_0 + c_i h, \mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{k}_j), \quad i = 1, \dots, s, \quad \Psi^{t_0, t_0+h} \mathbf{y}_0 := \mathbf{y}_0 + h \sum_{i=1}^s b_i \mathbf{k}_i,$$

ein *s*-stufiges Runge-Kutta-Einschrittverfahren (RK-ESV) für AWP (1.1.13) mit Inkrementen  $\mathbf{k}_i \in \mathbb{R}^d$ .

- Verallgemeinerung der Kollokationsverfahren → Sect. 2.2  
(doch keine konkrete Konstruktionsvorschrift !)

Falls  $a_{ij} = 0$  für  $i \leq j$  ► Explizites Runge-Kutta-Verfahren → Def. 2.1.5

Kurznotation für Runge-Kutta-Verfahren:

**Butcher-Schema**

$$\triangleright \begin{array}{c|c} \mathbf{c} & \mathfrak{A} \\ \hline & \mathbf{b}^T \end{array} := \begin{array}{c|cc} c_1 & a_{11} & \cdots & a_{1s} \\ \vdots & \vdots & & \vdots \\ c_s & a_{s1} & \cdots & a_{ss} \\ \hline & b_1 & \cdots & b_s \end{array}. \quad (2.3.6)$$



$\mathcal{U}$  echte untere Dreiecksmatrix

➤ explizites Runge-Kutta-Verfahren

$\mathcal{L}$  untere Dreiecksmatrix

➤ diagonal-implizites Runge-Kutta-Verfahren (DIRK)

*Bemerkung 2.3.7* (Stufenform der Inkrementgleichungen).

Für  $s$ -stufiges Runge-Kutta-Einschrittverfahren (RK-ESV),  $\rightarrow$  Def. 2.3.5, definiere (Annahme: eindeutige Lösbarkeit der Inkrementgleichungen)

**Stufen** (*engl. stages:*)  $\mathbf{g}_i = \mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{k}_j$ ,  $i = 1, \dots, s \Rightarrow \mathbf{k}_i = \mathbf{f}(t_0 + c_i h, \mathbf{g}_i)$ .

(2.3.8)

R. Hiptmair  
rev 35327,  
25. April  
2011

► Stufengleichungen

$$\mathbf{g}_i = \mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{f}(t_0 + c_j h, \mathbf{g}_j) \quad , \quad i = 1, \dots, s .$$

(2.3.9)



Interpretation: Runge-Kutta-Verfahren  $\leftrightarrow$  Polygonzugapproximation der Lösungskurve

→ Sect. 1.4

- Anzahl  $b_i \neq 0 \hat{=}$  Anzahl der Teilstrecken im Polygonzug
- $b_i, i = 1, \dots, s - 1 \hat{=}$  relative Länge des  $i$ . Teilintervalls
- $k_i \hat{=}$  „Steigung“ der  $i$ . Teilstrecke
- $c_i \hat{=}$  relativer Zeitpunkt (in  $[t_k, t_{k+1}]$ ) für Auswertung der  $i$ . Abschnittsteigung

*Beispiel 2.3.10* (Explizite Runge-Kutta-Polygonzugapproximation für Ricatti-Differentialgleichung).

→ Bsp 1.1.3

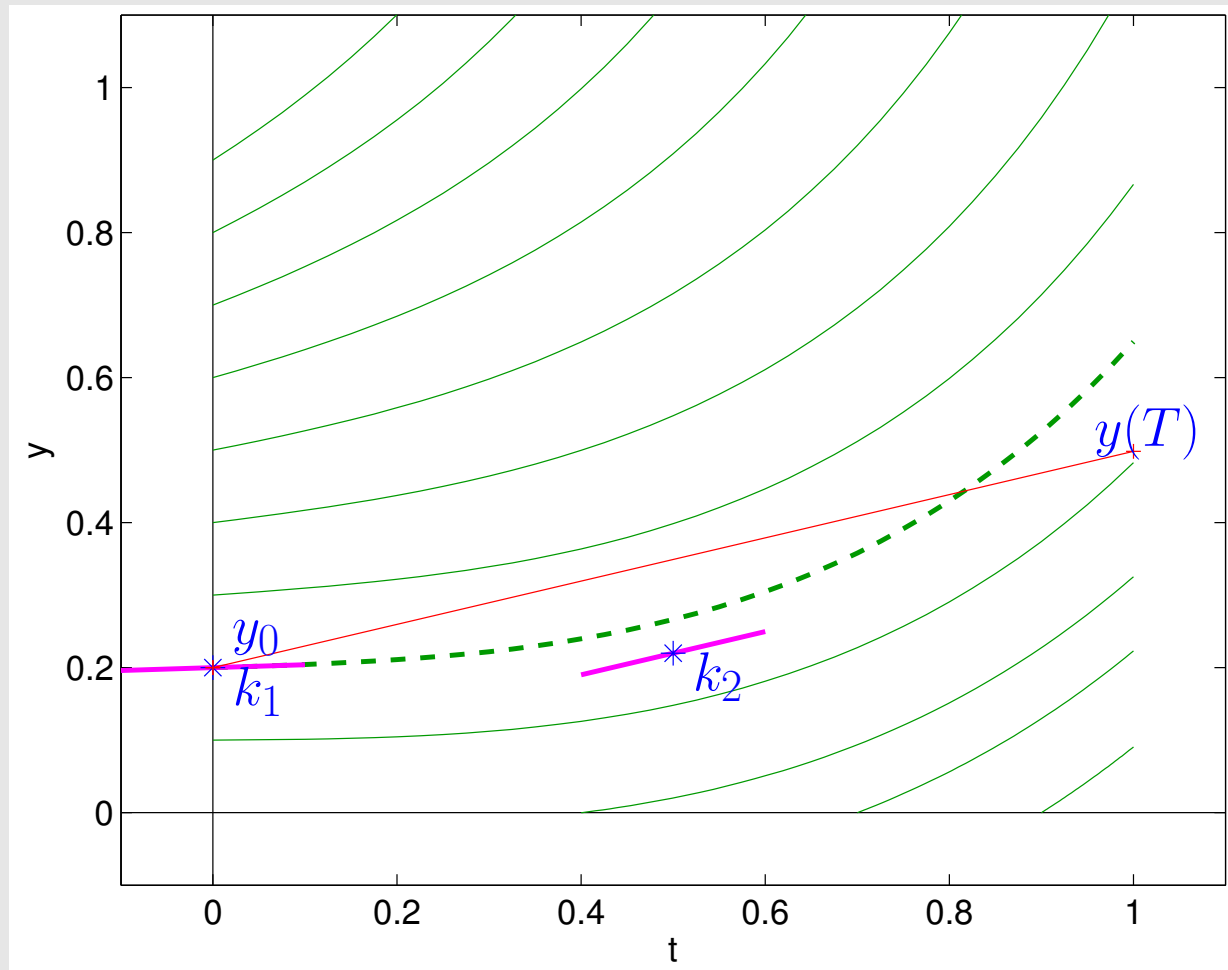
Anfangswertproblem:  $\dot{y} = t^2 + y^2, y(0) = 0.2$ .

Geometrische Interpretation von expliziten RK-ESV als Polygonzugverfahren → Verallgemeinerung des expliziten Euler-Verfahrens, siehe Sect. 1.4.1, Fig. ??.

Explizite Mittelpunktsregel:

$$\begin{array}{c|cc} 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \hline & 0 & 1 \end{array}$$

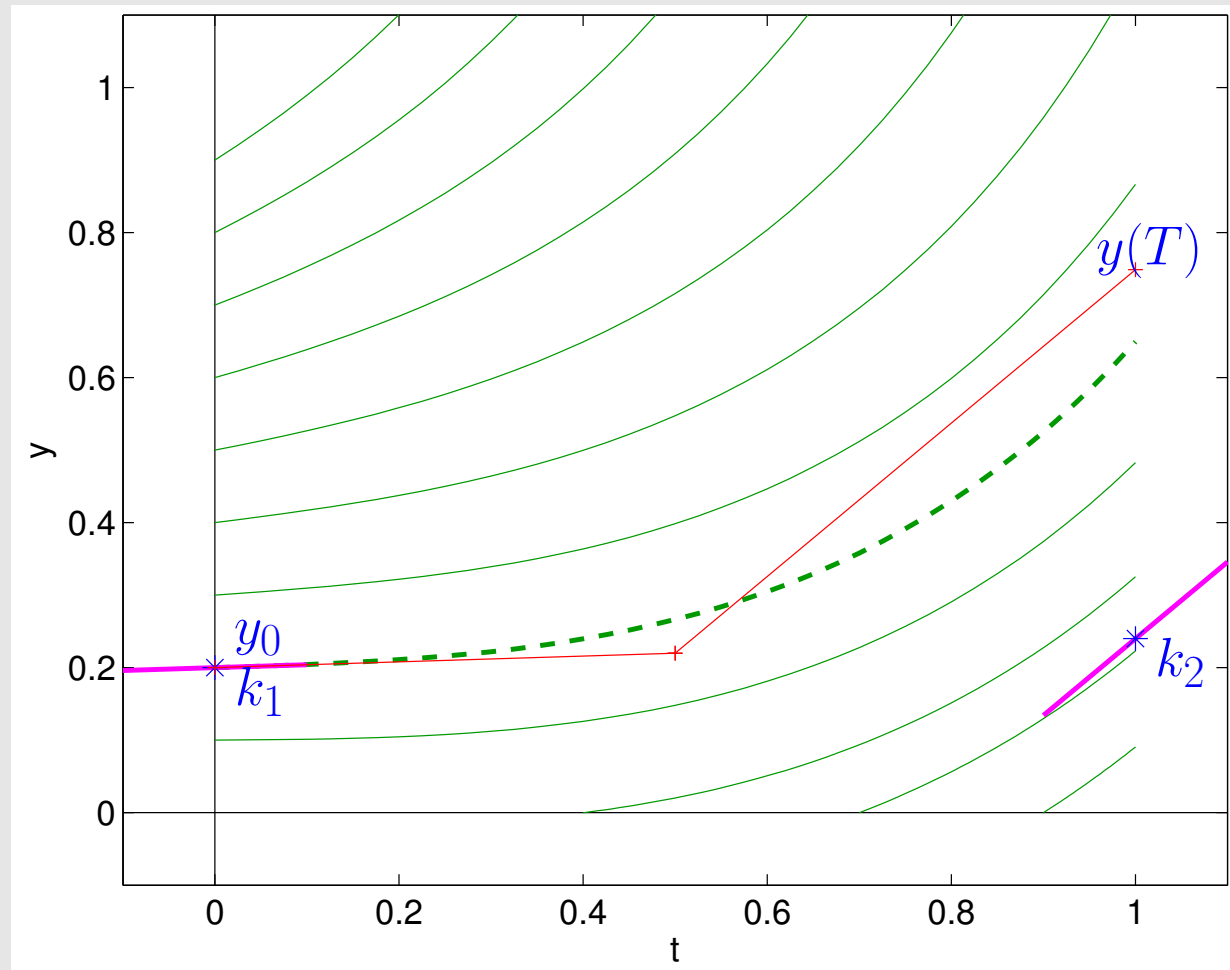
- grün Lösungskurven
- magenta Abschnittsteigungen  $k_i$
- \* Punkte  $f$ -Auswertung
- rot: Polygonzug



### Explizite Trapezregel

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

- grün: Lösungskurven
- magenta: Abschnittsteigungen  $k_i$
- \*: Punkte  $f$ -Auswertung
- rot: Polygonzug

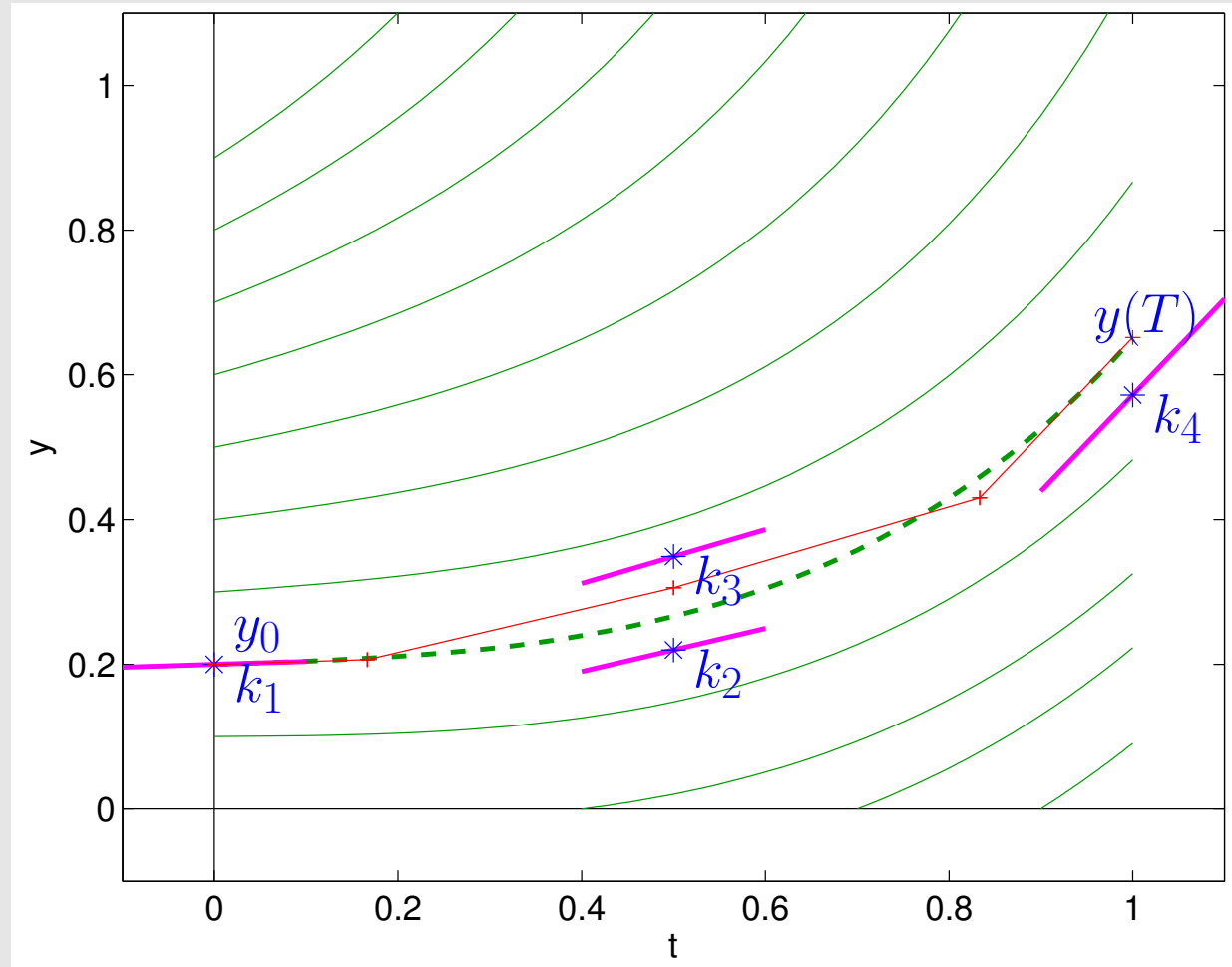


# Klassisches Runge-Kutta-Verfahren (RK4)

$0$	$0$	$0$	$0$	$0$
$\frac{1}{2}$	$\frac{1}{2}$	$0$	$0$	$0$
$\frac{1}{2}$	$0$	$\frac{1}{2}$	$0$	$0$
$1$	$0$	$0$	$1$	$0$
$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$0$

(2.3.11)

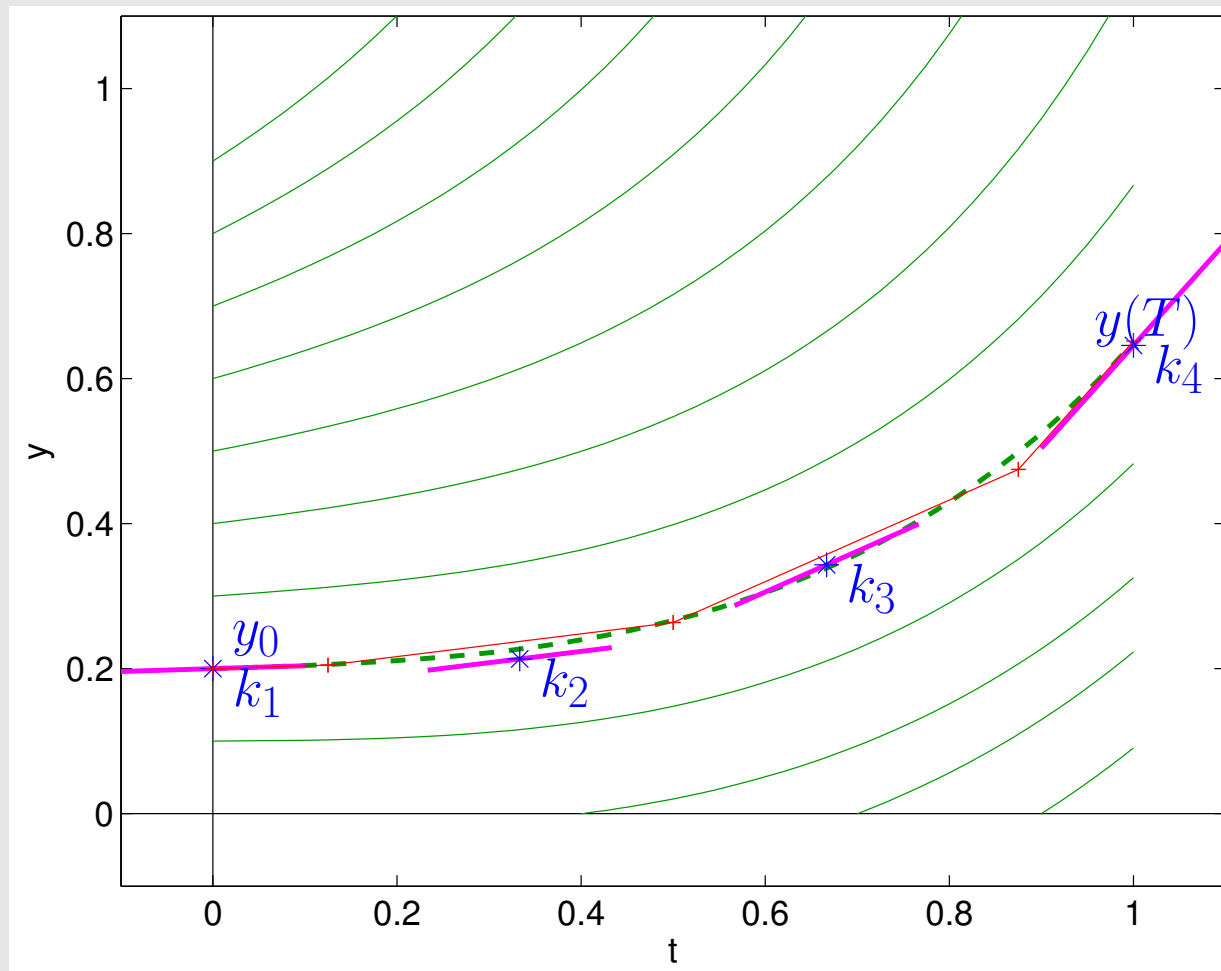
- grün Lösungskurven
- magenta Abschnittsteigungen  $k_i$
- \* Punkte  $f$ -Auswertung
- rot: Polygonzug



### Kuttas 3/8-Regel

$$\begin{array}{c|cccc}
 0 & 0 & 0 & 0 & 0 \\
 1 & \frac{1}{3} & 0 & 0 & 0 \\
 2 & -\frac{1}{3} & 1 & 0 & 0 \\
 3 & 1 & -1 & 1 & 0 \\
 \hline
 & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8}
 \end{array} \quad (2.3.12)$$

- grün Lösungskurven
- magenta Abschnittsteigungen  $k_i$
- \* Punkte  $f$ -Auswertung
- rot: Polygonzug



Bemerkung 2.3.13 (Affin-Kovarianz der Runge-Kutta-Verfahren).

Wie reagiert ein Runge-Kutta -ESV auf einen Basiswechsel im Zustandsraum ( $\rightarrow$  Sect. 1.3.2, (1.3.12))?

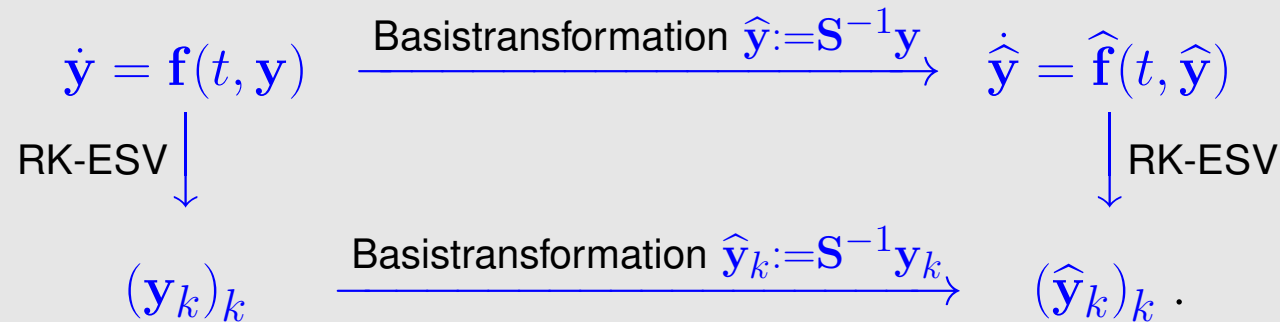
Für  $S \in \mathbb{R}^{d,d}$  regulär,  $\hat{y} := S^{-1}y$   $\Psi, \hat{\Psi}$  aus RK-Verfahren ( $\rightarrow$  Def. 2.3.5)


$$\begin{aligned} \Psi_h^{s,t} &= \text{Diskrete Evolution zu } \dot{y} = f(t, y), \\ \hat{\Psi}_h^{s,t} &= \text{Diskrete Evolution zu } \dot{\hat{y}} = \hat{f}(t, \hat{y}) \rightarrow (1.3.12) \end{aligned} \quad \blacktriangleright \quad \boxed{S \hat{\Psi}_h^{s,t} S^{-1} y = \Psi_h^{s,t} y} \quad (2.3.14)$$

Dazu zeigt man, dass die Inkremente  $k_i$  und transformierten Inkremente  $S k_i S^{-1}$  eines Runge-Kutta-ESV angewandt auf die ODEs  $\dot{y} = f(t, y)$ , bzw.  $\dot{\hat{y}} = \hat{f}(t, \hat{y}) = S^{-1}f(t, S\hat{y})$  die gleichen Gleichungen erfüllen. Wegen deren eindeutiger Lösbarkeit für hinreichend kleines  $h > 0$  folgt  $k_i = S k_i S^{-1}, i = 1, \dots, s$ .

R. Hiptmair  
rev 35327,  
25. April  
2011

Die obige Aussage lässt sich auch durch ein *kommutierendes Diagramm ausdrücken*



 Affin-Kovarianz drückt die *Erhaltung* einer einfachen algebraischen *Struktur* der Lösungsmenge eines AWP aus.

*Bemerkung 2.3.15* (Autonomisierungsinvarianz von Runge-Kutta-Verfahren).

Eine weitere Transformation einer ODE: **Autonomisierung** → Bem. 1.1.7

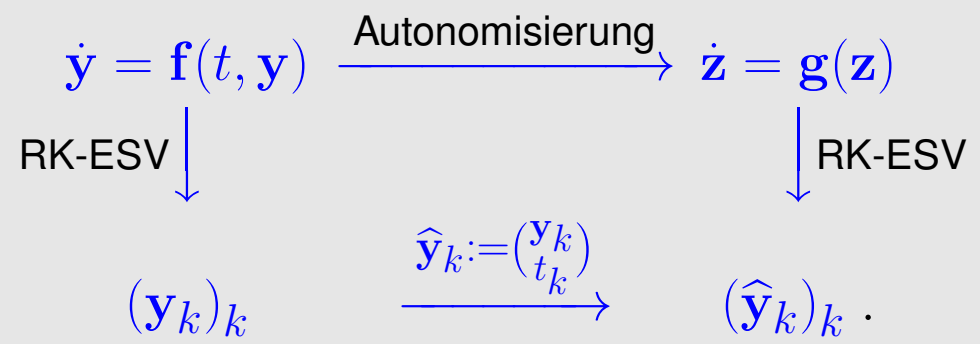
Auch hier stellt sich die Frage, wann “RK-ESV mit Autonomisierung kommutieren”, vgl. Bem. 2.3.13.

Autonomisierung:  $\begin{matrix} \dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}), \\ \mathbf{y}(t_0) = \mathbf{y}_0 \end{matrix} \Rightarrow \dot{\mathbf{z}} := \begin{pmatrix} \dot{\mathbf{y}} \\ \dot{s} \end{pmatrix} = \begin{pmatrix} \mathbf{f}(s, \mathbf{y}) \\ 1 \end{pmatrix} =: \mathbf{g} \begin{pmatrix} \mathbf{y} \\ s \end{pmatrix}, \begin{pmatrix} \mathbf{y}(0) \\ s(0) \end{pmatrix} = \begin{pmatrix} \mathbf{y}_0 \\ t_0 \end{pmatrix}$

Evolutionen:  $\Phi^{t,t+h} \Leftrightarrow \hat{\Phi}^h$   
 Diskrete Evl.:  $\Psi_h^{t,t+h} \Leftrightarrow \hat{\Psi}_h^h$

Wunsch:  $\begin{pmatrix} \Phi^{t,t+h} \mathbf{y} \\ t+h \end{pmatrix} = \hat{\Phi}^h \begin{pmatrix} \mathbf{y} \\ t \end{pmatrix} \blacktriangleright \begin{pmatrix} \Psi_h^{t,t+h} \mathbf{y} \\ t+h \end{pmatrix} = \hat{\Psi}_h^h \begin{pmatrix} \mathbf{y} \\ t \end{pmatrix} . \tag{2.3.16}$

(2.3.16) kann wieder durch durch ein *kommutierendes Diagramm* ausgedrückt werden:





Nun wollen wir Bedingungsgleichungen für die Koeffizienten  $a_{ij}$  und  $b_i$  des RK-ESV ( $\rightarrow$  Def. 2.3.5) herleiten, so dass (2.3.16) gilt.

$$\widehat{\Psi}_h^h \begin{pmatrix} \mathbf{y} \\ t \end{pmatrix} = \begin{pmatrix} \mathbf{y} + h \sum_{i=1}^s b_i \widehat{\mathbf{k}}_i \\ t + h \sum_{i=1}^s b_i \widehat{\kappa}_i \end{pmatrix}, \quad \begin{pmatrix} \widehat{\mathbf{k}}_i \\ \widehat{\kappa}_i \end{pmatrix} = \begin{pmatrix} \mathbf{f}(t + h \sum_{j=1}^s a_{ij} \widehat{\kappa}_j, \mathbf{y} + h \sum_{j=1}^s a_{ij} \widehat{\mathbf{k}}_j) \\ 1 \end{pmatrix}.$$

$$\boxed{c_i = \sum_{j=1}^s a_{ij}} \quad \& \quad \boxed{\sum_{i=1}^s b_i = 1} \quad \blacktriangleright \quad \widehat{\mathbf{k}}_i = \mathbf{k}_i. \quad (2.3.17)$$

= Hinreichende + notwendige Bedingungen für **Autonomisierungsinvarianz** eines RK-Verfahrens

Darum  $c_i = \sum_{j=1}^s a_{ij}$  in Def. 2.3.5 !

▷ Analyse von autonomisierungsinvarianten RK-Verfahren kann sich auf autonome Probleme beschränken.



*Bemerkung 2.3.18* (“Dense output”). [17, Sect. II.5]

Runge-Kutta-Einschrittverfahren ( $\rightarrow$  Def. 2.3.5) liefern Gitterfunktionen  $\mathcal{G} \mapsto \mathbb{R}^d$  als Näherung von  $t \mapsto \mathbf{y}(t)$  in diskreten Zeitpunkten.

Was, wenn Näherungen für  $\mathbf{y}(t)$  zu anderen Zeitpunkten/überall auf  $[0, T]$  gebraucht werden ?

- Ziel:
- Stückweise polynomiale Definition von  $t \mapsto \mathbf{y}_h(t)$
  - Interpolationseigenschaft  $\mathbf{y}_h(t_k) = \mathbf{y}_k, k = 0, \dots, N$
  - $\mathbf{y}_h|_{[t_k, t_{k+1}]}$  berechenbar aus  $\mathbf{y}_k, \mathbf{y}_{k+1}$  und Inkrementen im  $k$ . Schritt

$$\blacktriangleright \quad \mathbf{y}_h(t_k + \xi h_k) = p_0(\xi) \mathbf{y}_k + p_1(\xi) \mathbf{y}_{k+1} + \sum_{i=1}^s q(\xi) \mathbf{k}_i, \quad 0 \leq \xi \leq 1,$$

mit Polynomen  $p_0, p_1, q_i : \mathbb{R} \mapsto \mathbb{R}$ .

Wunsch: Für RK-ESV der Ordnung  $p \quad \triangleright \quad \max_{0 \leq t \leq T} \|\mathbf{y}(t) - \mathbf{y}_h(t)\| = O(h^p)$



*Bemerkung 2.3.19* (Lösung der Inkrementgleichungen).  $\rightarrow$  [8, Sect. 6.2.2]

Inkrementgleichungen für implizites RW-ESV ( $\rightarrow$  Def. 2.3.5) = (i.a. *nichtlineares*) Gleichungssystem mit  $s \cdot d$  Unbekannten

Im autonomen Fall (vgl. Beweis von Lemma 2.2.7)

$$\mathbf{k}_i := \mathbf{f}\left(\mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{k}_j\right) \iff \mathbf{k}_i = \mathbf{f}(\mathbf{y}_0 + \mathbf{g}_i) \iff \mathbf{g}_i = h \sum_{j=1}^s a_{ij} \mathbf{f}(\mathbf{y}_0 + \mathbf{g}_j), \quad i = 1, \dots, s. \quad (2.3.20)$$

Die Grössen  $\mathbf{g}_i + \mathbf{y}_0$  heissen auch **Stufen** (*engl.* stages) des Runge-Kutta-Verfahrens, siehe Bem. 2.3.7. Daher heisst die Formulierung der Inkrementgleichungen mit Hilfe der  $\mathbf{g}_i$  wie in (2.3.20) auch deren **Stufenform**.

- iterative Lösung mit *vereinfachtem Newton-Verfahren* („eingefrorene“ Jacobi-Matrix)  
! Effizienz: Minimiere Anzahl von  $\mathbf{f}$ ,  $D\mathbf{f}$ -Auswertungen

Mit  $\mathbf{g} = (\mathbf{g}_1, \dots, \mathbf{g}_s)^T \in \mathbb{R}^{s \cdot d}$  definiere

$$F(\mathbf{g}) := \mathbf{g} - \begin{pmatrix} h \sum_{j=1}^s a_{1j} \mathbf{f}(\mathbf{y}_0 + \mathbf{g}_j) \\ \vdots \\ h \sum_{j=1}^s a_{sj} \mathbf{f}(\mathbf{y}_0 + \mathbf{g}_j) \end{pmatrix} \Rightarrow \{(2.3.20) \Leftrightarrow F(\mathbf{g}) = 0\}. \quad (2.3.21)$$

$h$  „klein“  $\triangleright$  Natürliche Anfangsnäherung für vereinfachte Newton-Iteration:  $\mathbf{g}^{(0)} = \mathbf{0}$

$$\triangleright DF(\mathbf{g}^{(0)}) = \begin{pmatrix} \mathbf{I} - ha_{11}D\mathbf{f}(\mathbf{y}_0) & -ha_{12}D\mathbf{f}(\mathbf{y}_0) & \cdots & -ha_{1s}D\mathbf{f}(\mathbf{y}_0) \\ -ha_{21}D\mathbf{f}(\mathbf{y}_0) & \mathbf{I} - ha_{22}D\mathbf{f}(\mathbf{y}_0) & & \vdots \\ \vdots & & \ddots & \vdots \\ -ha_{s1}D\mathbf{f}(\mathbf{y}_0) & \cdots & -ha_{s,s-1}D\mathbf{f}(\mathbf{y}_0) & \mathbf{I} - ha_{ss}D\mathbf{f}(\mathbf{y}_0) \end{pmatrix} .$$

$\triangleright$  Vereinfachte Newton-Iteration

$$\mathbf{g}^{(0)} = \mathbf{0} \quad , \quad \mathbf{g}^{(k+1)} = \mathbf{g}^{(k)} - DF(\mathbf{g}^{(k)})^{-1}F(\mathbf{g}^{(k)}) \quad , \quad k = 0, 1, 2, \dots .$$

Wiedergewinnung der Inkremente  $\mathbf{k}_i$  aus  $\mathbf{g}_i$ : betrachte  $l$ . Komponente,  $l = 1, \dots, d$ . Mit  $\mathbf{g}_i = (g_{i,1}, \dots, g_{i,d})^T \in \mathbb{R}^d$

$$g_{i,l} = h \sum_{j=1}^s a_{ij}k_{j,l} \quad \iff \quad \left( g_{i,l} \right)_{l=1}^d = h\mathfrak{A} \left( k_{i,l} \right)_{l=1}^d .$$

$\mathfrak{A}$  regulär  $\triangleright$   $\mathbf{k}_i$  durch Lösen von  $s$  linearen Gleichungssystemen mit Koeffizientenmatrix  $\mathfrak{A}$ .

Natürlich kann das vereinfachte Newton-Verfahren auch auf die Standardform der Inkrementgleichungen aus Def. 2.3.5 angewandt werden.

## 2.3.2 Konvergenz

*Beispiel 2.3.22* (Konvergenz expliziter Runge-Kutta-Verfahren).

- Skalare logistische Differentialgleichung (1.2.2),  $\lambda = 10$ ,  $y(0) = 0.01$ ,  $T = 1$
- Explizite Runge-Kutta-Einschrittverfahren, uniforme Zeitschrittweite  $h$

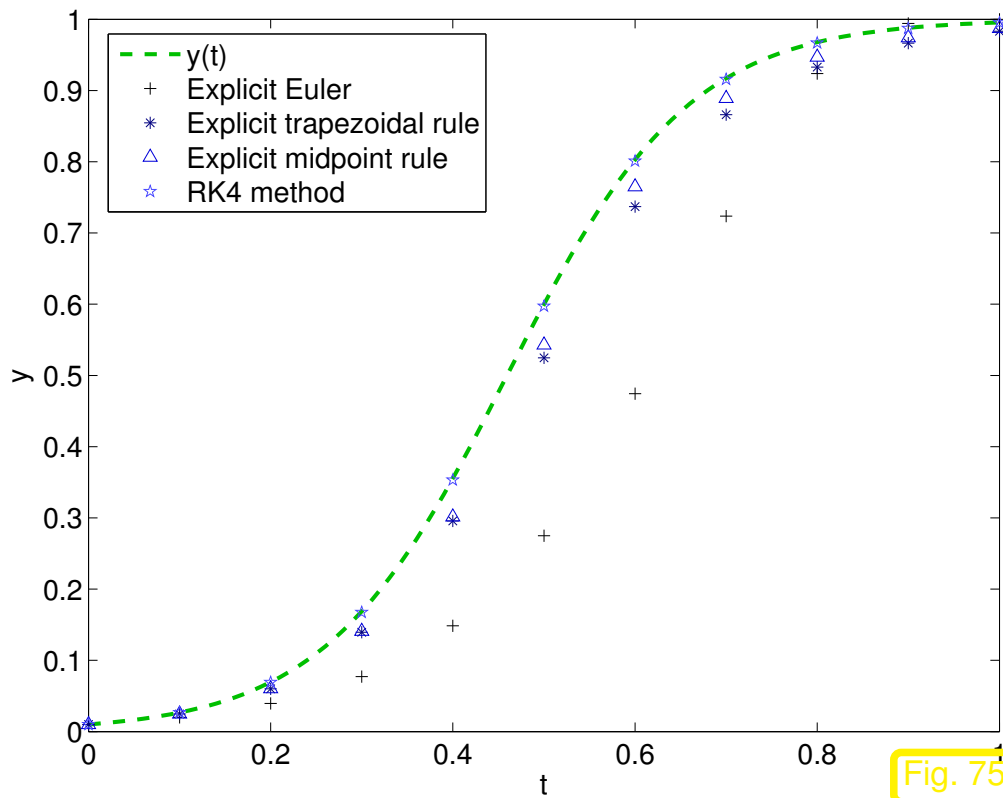
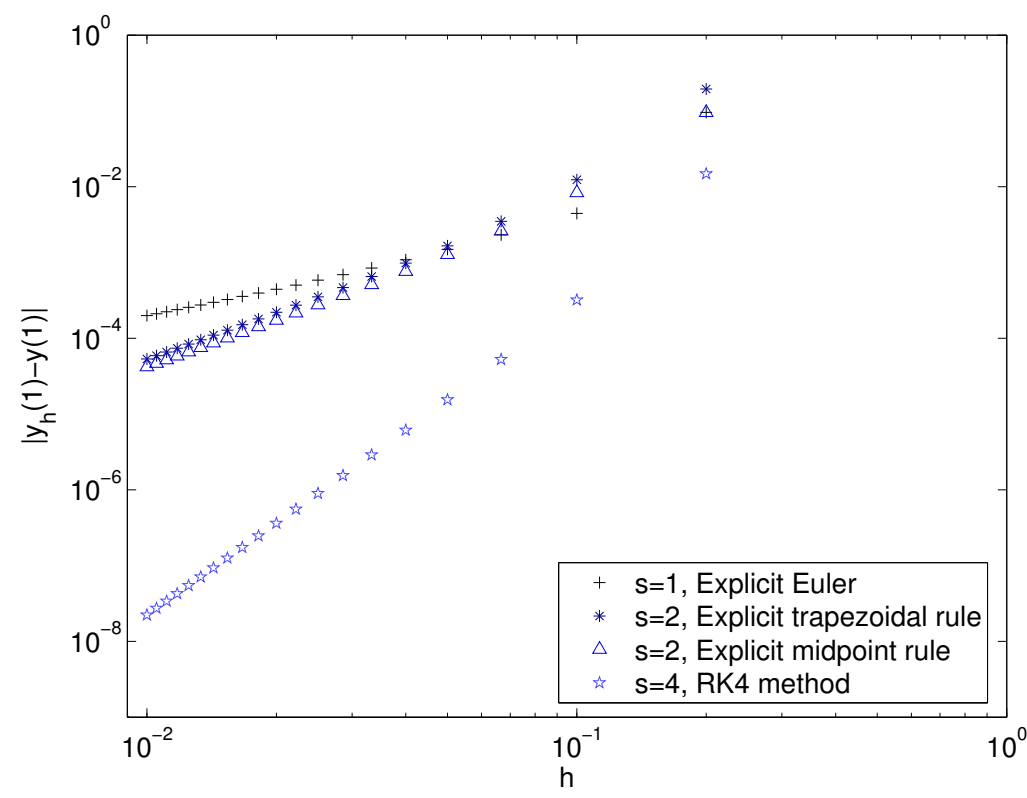


Fig. 75



Konvergenz des Fehlers  $y_h(1) - y(1)$

$y_h(j/10), j = 1, \dots, 10$  für explizite RK-Verfahren

R. Hiptmair  
rev 35327,  
25. April  
2011

Viele unserer Resultate über Kollokationsverfahren ( $\rightarrow$  Sect. 2.2) bleiben gültig für die allgemeinere Klasse der Runge-Kutta-Einschrittverfahren (mit im wesentlichen den gleichen Beweisen):

Lemmas 2.2.7, 2.2.13 bleiben gültig für Runge-Kutta-Einschrittverfahren aus Def. 2.3.5

**Lemma 2.3.23** (Konsistenz von Runge-Kutta-Einschrittverfahren).

Unter den Voraussetzungen von Lemma 2.2.7 ist ein Runge-Kutta-Einschrittverfahren ( $\rightarrow$  Def. 2.3.5) konsistent ( $\rightarrow$  Def. 2.1.8) genau dann, wenn  $\sum_{i=1}^s b_i = 1$ .

Frage: Wann ist ein Runge-Kutta-Verfahren konsistent ( $\Rightarrow$  konvergent, Thm. 2.1.19) von der Ordnung  $p$  ?



(Konsistenz-)Bedingungsgleichungen für Koeffizienten  $a_{ij}$ ,  $b_i$

Hilfsmittel zum Aufstellen der Bedingungsgleichungen : *Taylor-Entwicklung* (des Konsistenzfehlers  $\tau(t, \mathbf{y}, h) \rightarrow$  Dff. 2.1.11)

Annahme:  $\mathbf{f}$  "hinreichend glatt"

*Beispiel 2.3.24* (RK-Bedingungsgleichungen für Konsistenzordnung).  $p = 3$  [8, Sect. 4.2.2]

**Konsistenzfehler:**  $\tau(t, \mathbf{y}, h) := (\Phi^{t,t+h} - \Psi^{t,t+h})\mathbf{y}$  ( $h$  hinreichend klein); .

Fokus: autonome Differentialgleichung  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ ,  $\mathbf{f}$  "hinreichend glatt"

Fixiere Anfangswert  $\mathbf{y}_0 \in D$ , O.B.d.A.  $t_0 = 0$  (vgl. Bem. 1.1.15)

① Taylorentwicklung der kontinuierlichen Evolution in  $h$  um  $h = 0$ :

$$\Phi^h \mathbf{y}_0 = \mathbf{y}(h) = \mathbf{y}_0 + \dot{\mathbf{y}}(0)h + \frac{1}{2}\ddot{\mathbf{y}}(0)h^2 + \frac{1}{6}\mathbf{y}^{(3)}(0)h^3 + O(h^4), \quad (2.3.25)$$

mit

$$\dot{\mathbf{y}}(0) = \mathbf{f}(\mathbf{y}_0),$$

$$\ddot{\mathbf{y}}(0) = D\mathbf{f}(\mathbf{y}_0)\dot{\mathbf{y}}(0) = D\mathbf{f}(\mathbf{y}_0)\mathbf{f}(\mathbf{y}_0),$$

$$\mathbf{y}^{(3)}(0) = D^2\mathbf{f}(\mathbf{y}_0)(\dot{\mathbf{y}}(0), \dot{\mathbf{y}}(0)) + D\mathbf{f}(\mathbf{y}_0)\ddot{\mathbf{y}}(0) = D^2\mathbf{f}(\mathbf{y}_0)(\mathbf{f}(\mathbf{y}_0), \mathbf{f}(\mathbf{y}_0)) + D\mathbf{f}(\mathbf{y}_0)D\mathbf{f}(\mathbf{y}_0)\mathbf{f}(\mathbf{y}_0).$$

Für analoge Überlegungen siehe auch Bsp. 2.1.15.



$$\Phi^h \mathbf{y}_0 = \mathbf{y}_0 + h\mathbf{f}(\mathbf{y}_0) + \frac{1}{2}h^2 D\mathbf{f}(\mathbf{y}_0)\mathbf{f}(\mathbf{y}_0) + \frac{1}{6}h^3 (D\mathbf{f}(\mathbf{y}_0)D\mathbf{f}(\mathbf{y}_0)\mathbf{f}(\mathbf{y}_0) + D^2\mathbf{f}(\mathbf{y}_0)(\mathbf{f}(\mathbf{y}_0), \mathbf{f}(\mathbf{y}_0))) + O(h^4). \quad (2.3.26)$$

② Taylorentwicklung der diskreten Evolution in  $h$  um  $h = 0$



Taylorentwicklung der Inkremente  $\mathbf{k}_i$  in  $h$  um  $h = 0$

$$\mathbf{k}_i = \mathbf{f}(\mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{k}_j) = \quad (2.3.27)$$

$$= \mathbf{f}(\mathbf{y}_0) + h D\mathbf{f}(\mathbf{y}_0) \left( \sum_{j=1}^s a_{ij} \mathbf{k}_j \right) + \frac{1}{2}h^2 D^2\mathbf{f}(\mathbf{y}_0) \left( \sum_{j=1}^s a_{ij} \mathbf{k}_j, \sum_{j=1}^s a_{ij} \mathbf{k}_j \right) + O(h^3)$$

Einsetzen „kürzerer Taylorentwicklungen“ anstelle der Inkremente

Inkremente werden mit  $h$  multipliziert !

$$\begin{aligned}
\mathbf{k}_i &= \mathbf{f}(\mathbf{y}_0) + h D\mathbf{f}(\mathbf{y}_0) \sum_{j=1}^s a_{ij} \left( f(\mathbf{y}_0 + h D\mathbf{f}(\mathbf{y}_0) \sum_{l=1}^s a_{il} \mathbf{k}_l + O(h^2)) \right) + \\
&\quad \frac{1}{2} h^2 D^2 \mathbf{f}(\mathbf{y}_0) \left( \sum_{j=1}^s a_{ij} (\mathbf{f}(\mathbf{y}_0) + O(h)), \sum_{j=1}^s a_{ij} (\mathbf{f}(\mathbf{y}_0) + O(h)) \right) + O(h^3) \\
&= \mathbf{f}(\mathbf{y}_0) + h \underbrace{\sum_{j=1}^s a_{ij}}_{=c_i, \text{ siehe Bem. 2.3.15}} D\mathbf{f}(\mathbf{y}_0) \mathbf{f}(\mathbf{y}_0) + \\
&\quad h^2 \left( \sum_{l=1}^s a_{il} c_l \right) D\mathbf{f}(\mathbf{y}_0) D\mathbf{f}(\mathbf{y}_0) \mathbf{f}(\mathbf{y}_0) + h^2 \frac{1}{2} c_i^2 D^2 \mathbf{f}(\mathbf{y}_0) (\mathbf{f}(\mathbf{y}_0), \mathbf{f}(\mathbf{y}_0)) + O(h^3) .
\end{aligned}$$

Beachte:  $\Psi^h \mathbf{y}_0 = \mathbf{y}_0 + h \sum_{i=1}^s b_i \mathbf{k}_i \blacktriangleright$  Entwicklung bis  $O(h^3)$  ausreichend

$$\blacktriangleright \Psi^h \mathbf{y}_0 = \mathbf{y}_0 + \left( h \sum_{i=1}^s b_i \right) \mathbf{f}(\mathbf{y}_0) + \left( h^2 \sum_{i=1}^s b_i c_i \right) D\mathbf{f}(\mathbf{y}_0) \mathbf{f}(\mathbf{y}_0) + \quad (2.3.28)$$

$$\left( h^3 \sum_{i=1}^s b_i \sum_{j=1}^s a_{ij} c_j \right) D\mathbf{f}(\mathbf{y}_0) D\mathbf{f}(\mathbf{y}_0) \mathbf{f}(\mathbf{y}_0) +$$

$$\left( \frac{1}{2} h^3 \sum_{i=1}^s b_i c_i^2 \right) D^2 \mathbf{f}(\mathbf{f}(\mathbf{y}_0), \mathbf{f}(\mathbf{y}_0)) + O(h^4) .$$

③ Gleichsetzen der Koeffizienten der *linear unabhängigen elementaren Differentiale*

$$1, \mathbf{f}(\mathbf{y}_0), D\mathbf{f}(\mathbf{y}_0) \mathbf{f}(\mathbf{y}_0), D\mathbf{f}(\mathbf{y}_0) D\mathbf{f}(\mathbf{y}_0) \mathbf{f}(\mathbf{y}_0), D^2 \mathbf{f}(\mathbf{y}_0)(\mathbf{f}(\mathbf{y}_0), \mathbf{f}(\mathbf{y}_0))$$

in (2.3.28) und (2.3.26)

$\blacktriangleright$  *Hinreichende & notwendige Bedingungsgleichungen* für Konsistenzordnung  $p = 3$  eines (autonomisierungsinvarianten  $\rightarrow$  Rem. 2.3.15) RK-Verfahrens:

$$\sum_{i=1}^s b_i = 1, \quad (2.3.29)$$

$$\sum_{i=1}^s b_i c_i = \frac{1}{2}, \quad (2.3.30)$$

$$\sum_{i=1}^s b_i c_i^2 = \frac{1}{3}, \quad (2.3.31)$$

$$\sum_{i=1}^s b_i \sum_{j=1}^s a_{ij} c_j = \frac{1}{6}.$$

☞ (2.3.29) hinreichend & notwendig für Konsistenzordnung  $p = 1$ , siehe Lemma 2.3.23

☞ (2.3.29) + (2.3.30) hinreichend & notwendig für Konsistenzordnung  $p = 2$



*Bemerkung 2.3.32* (Butcher-Bäume).

Allgemeiner kombinatorischer Algorithmus zum Aufstellen der RK-Bedingungsgleichungen: **Butcher-Bäume** [8, Sect. 4.2.3], [16, Ch. III]



→ Konstruktion von RK-Verfahren vorgegebener Konvergenzordnung durch Lösen der (nichtlinearen) Bedingungsgleichungen (vom Typ (2.3.29)-(2.3.31)):

$p$	1	2	3	4	5	6	7	8	9	10	20
#B.G.	1	2	4	8	17	37	85	200	486	1205	20247374

Einige Konvergenzordnungen von Runge-Kutta-Verfahren:

Explizite Verfahren		Implizite Verfahren	
Explizites Eulerverfahren (2.2.1)	$p = 1$	Implizites Eulerverfahren (2.2.1)	$p = 1$
Explizite Trapezregel (2.3.3)	$p = 2$	Implizite Mittelpunktsregel (2.2.19)	$p = 2$
Explizite Mittelpunktsregel (2.3.4)	$p = 2$	Gauss-Kollokationsverfahren	$p = 2s$
Klassisches Runge-Kutta-V. (2.3.11)	$p = 4$		
Kuttas $3/8$ -Regel (2.3.12)	$p = 4$		

R. Hiptmair  
rev 35327,  
25. April  
2011

Viele weitere RK-Verfahren ▷ [17, 18]

Ordnungsschranken:

Für explizite Runge-Kutta-Verfahren  $p \leq s$

Für allgemeine Runge-Kutta-Verfahren  $p \leq 2s$

➤ Gauss-Kollokationsverfahren realisieren maximale Ordnung

*Bemerkung 2.3.33* (“Butcher barriers” für explizite RK-ESV).

Ordnung $p$	1	2	3	4	5	6	7	8	$\geq 9$
Minimale Stufenzahl $s$	1	2	3	4	6	7	9	11	$\geq p + 3$

Eine allgemeine Formel für die minimale Stufenzahl konnte bisher nicht hergeleitet werden.



*Bemerkung 2.3.34* (Warum Einschrittverfahren hoher Ordnung?).

Die allgemeine Konvergenztheorie von Einschrittverfahren aus Abschnitt 2.1.3 liefert uns bei hinreichender Glattheit der Lösung eines Anfangswertproblems die **asymptotische** Fehlerabschätzung

$$\text{err}(\mathcal{G}) := \max_{k=1,\dots,N} \|\mathbf{y}_k - \mathbf{y}(t_k)\| \leq Ch_{\mathcal{G}}^p \quad \text{für } h_{\mathcal{G}} \text{ hinreichend klein,} \quad (2.3.35)$$

siehe Thm. 2.1.19, wobei die Konstante  $C > 0$  nicht von der maximalen Zeitschrittweite  $h_{\mathcal{G}} > 0$  abhängt, aber in der Regel nicht bekannt ist.

Daher können wir aus (2.3.35) in der Regel keine Aussage über den Integrationsfehler auf einem konkreten Zeitgitter machen ( $\rightarrow$  Diskussion am Ende von Abschnitt 2.1.1) und auch nicht die Zeitschrittweite vorhersagen, die erforderlich ist, um eine gewünschte Genauigkeit zu erreichen.

Wie bereits bemerkt erlaubt die Abschätzung (2.3.35) unter der Annahme, dass sie scharf ist, nur die Vorhersage

welche Reduktion des Integrationsfehlers bei Verringerung der Zeitschrittweite zu erwarten ist.

**Annahme:** Abschätzung (2.3.35) ist scharf

Dann lässt sich vorhersagen, welcher Gewinn an Genauigkeit durch zusätzlichen Rechenaufwand für die numerische Integration eines AWP zu erzielen ist.

Konvention:

Rechenaufwand  $\sim$  Gesamtzahl der **f**-Auswertungen

➤ Für  $s$ -stufiges Runge-Kutta-Einschrittverfahren ( $\rightarrow$  Def. 2.3.5):


Rechenaufwand  $\sim s \cdot \text{Anzahl(Schritte)} \sim Csh^{-1}$  mit einer Konstanten  $C > 0$


für uniformes Zeitgitter, Zeitschrittweite  $h > 0$

Ziel: vorgegebene Reduktion des Fehlers: für  $0 < \rho < 1$

$$\frac{\text{err}(\mathcal{G}_{\text{neu}})}{\text{err}(\mathcal{G}_{\text{alt}})} \stackrel{!}{=} \rho \quad (2.3.35) \quad \implies \quad \frac{h_{\text{neu}}^p}{h_{\text{alt}}^p} \stackrel{!}{=} \rho \quad \Leftrightarrow \quad \boxed{h_{\text{neu}} = \rho^{1/p} h_{\text{alt}}} .$$

Faustregel für ein RK-ESV der (Konsistenz- = Konvergenz-)Ordnung  $p \in \mathbb{N}$  (uniformer Zeitschritt)

Erhöhung des Rechenaufwands um Faktor  $\rho^{-1/p}$   Erwarte Fehlerreduktion im Faktor  $\rho$

 Je höher die Ordnung, desto weniger relativer Zusatzaufwand ist für eine Reduktion des Fehler um einen vorgegebenen Faktor erforderlich.



### 2.4.1 Der Kombinationstrick

Einschrittverfahren für Anfangswertproblem

$$\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}) \quad , \quad \mathbf{y}(t_0) = \mathbf{y}_0 \quad , \quad (t_0, \mathbf{y}_0) \in \Omega \quad , \quad ((1.1.13))$$

betrachtet auf  $[t_0, T]$ , liefert Gitterfunktion  $\mathbf{y}_{\mathcal{G}} = (\mathbf{y}_k)_{k=1}^N$  als Lösung:  $\mathbf{y}_k \approx \mathbf{y}(t_k)$ ,  $t_N = T$ .

Bei äquidistanter Zeitschrittweite  $h > 0$ , schreiben wir auch  $\mathbf{y}_h(t_k) := \mathbf{y}_k$ , also  $\mathbf{y}_h(T) = \mathbf{y}_N$ .

Sect. 2.1.3: Einschrittverfahren für  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ , konvergent von Ordnung  $p$

$$\exists C > 0: \quad \|\mathbf{y}_h(T) - \mathbf{y}(T)\| \leq Ch^p \quad \text{für } h \rightarrow 0, N = T/h \in \mathbb{N} .$$

(Annahme: Äquidistante Zeitschritte der Länge  $h > 0$ )

Spekulation:  $\exists \mathbf{c} \in \mathbb{R}^d$ :  $\mathbf{y}_h(T) - \mathbf{y}(T) = \mathbf{c}h^p + O(h^{p+1})$  für  $h \rightarrow 0$  . (2.4.1)

$$\mathbf{y}_h(T) - \mathbf{y}(T) = \mathbf{c}h^p + O(h^{p+1}) \quad (\text{I})$$

$$\mathbf{y}_{h/2}(T) - \mathbf{y}(T) = \mathbf{c}2^{-p}h^p + O(h^{p+1}) \quad (\text{II})$$

$$(\text{I}) - 2^p \cdot (\text{II}): \quad y_h(T) - 2^p y_{h/2}(T) - (1 - 2^p)\mathbf{y}(T) = O(h^{p+1}),$$

$$\Rightarrow \quad \frac{y_h(T) - 2^p y_{h/2}(T)}{1 - 2^p} - \mathbf{y}(T) = O(h^{p+1}).$$

kombiniertes Verfahren, konvergent von Ordnung  $p + 1$  !

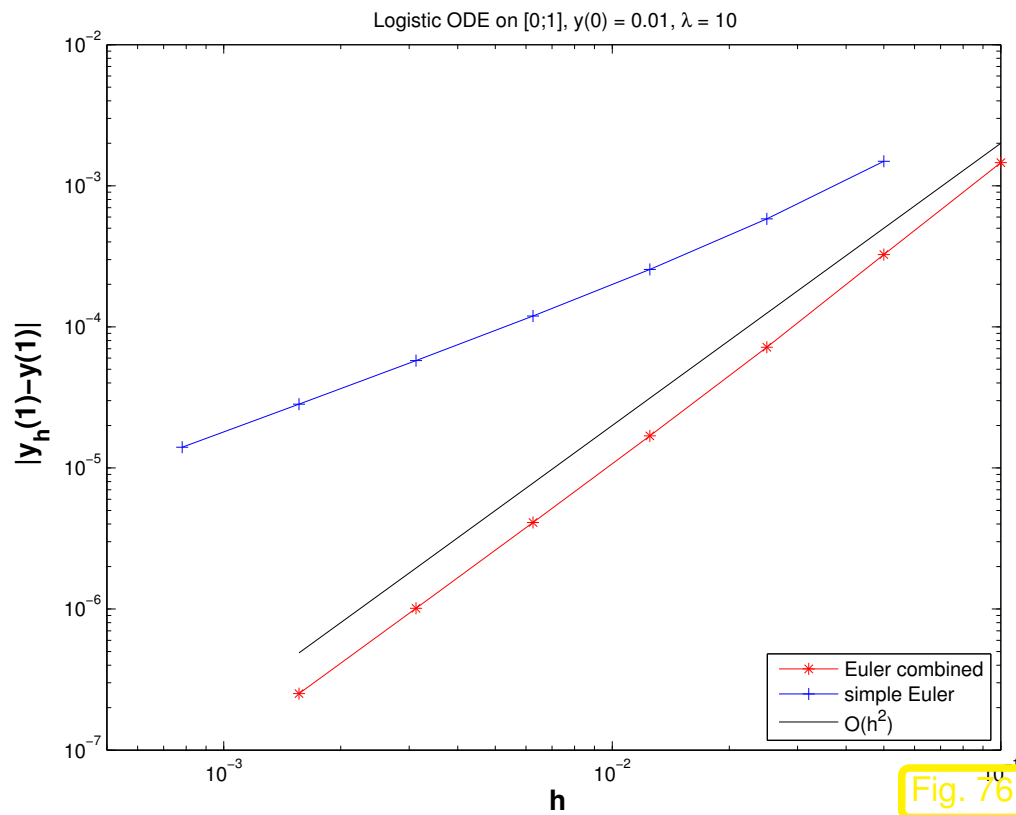
*Beispiel 2.4.2* (Konvergenz kombinierter Verfahren).

AWP für logistische Differentialgleichung (2.2.84): ( $\rightarrow$  Bsp. 1.2.1)

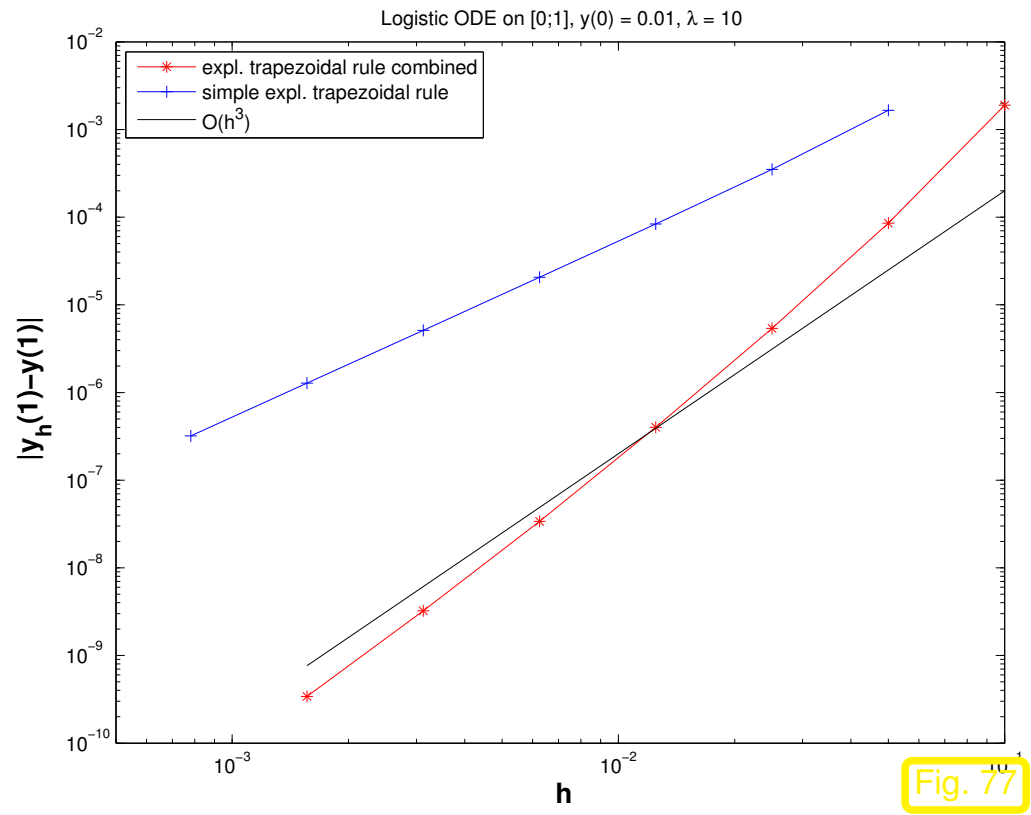
$$\dot{y} = \lambda y(1 - y), \quad y_0 = 0.01 \quad \Rightarrow \quad y(t) = \frac{1}{1 + 99 \cdot e^{-\lambda t}}, \quad t \in \mathbb{R}.$$

Basisverfahren: Explizites Euler-Verfahren (1.4.2),  $p = 1$

Explizite Trapezregel (2.3.3),  $p = 2$



Euler-Verfahren



Explizite Trapezregel



Beachte: Falls  $h \mapsto \mathbf{y}_h(T)$  glatt, Verfahren konvergent von Ordnung  $p$ , dann ist (2.4.1) naheliegend:  
**Taylorentwicklung !**

Dies wird in Abschnitt 2.4.3 vertieft.

Erinnerung ( $\rightarrow$  Vorlesung „Numerische Methoden“): **Romberg-Quadratur** ( $\rightarrow$  [9, Sect. 9.4])

Abstrakter Rahmen:

Problem:  $\Pi : X \mapsto \mathbb{R}^d$ , gesucht  $\Pi(x_0)$  für festes  $x_0 \in X$ ,  $X \hat{=}$  Datenraum

**Familie** numerischer Näherungsverfahren  $\left\{ \Pi_h : X \mapsto \mathbb{R}^d \right\}_h \blacktriangleright$  Näherungen  $\Pi_h(x_0) \approx \Pi(x_0)$

$\Pi_h$  abhängig von **skalarem Diskretisierungsparameter**  $h > 0$  (z.B. Zeitschrittweite)

- Berechne  $\Pi_h(x_0)$  für  $h \in \{h_1, \dots, h_k\}$  („Schrittweitenfolge“,  $h_i > h_{i+1}$ )

- Berechne **Interpolationspolynom**  $\mathbf{p} \in (\mathcal{P}_{k-1})^d$  mit

$$\mathbf{p}(h_i) = \Pi_{h_i}(x_0), \quad i = 1, \dots, k$$

- Bessere (?) Näherung

$$\Pi(x_0) \approx \mathbf{p}(0)$$

R. Hiptmair  
rev 35327,  
25. April  
2011

**Beispiel 2.4.3** (Romberg-Quadratur).

Interpretation des abstrakten Rahmens für die Romberg-Quadratur:  $X = C^0([a, b])$ ,  $a, b \in \mathbb{R}$ ,  
 $a < b$

$$\mathbf{\Pi}(f) := \int_a^b f(x) \, dx \quad , \quad \mathbf{\Pi}_h := \frac{h}{2}f(a) + h \sum_{j=1}^{N-1} f\left(a + j\frac{b-a}{N}\right) + \frac{h}{2}f(b) \quad , \quad h := \frac{1}{N} \quad ,$$

$\mathbf{\Pi}_h \hat{=}$  **Trapezregel** zur numerischen Quadratur

Diskretisierungsparameter  $h = \frac{1}{N}$ ,  $N \in \mathbb{N}$  ("Maschenweite" der Trapezregel): kann nur diskrete  
 Werte annehmen !

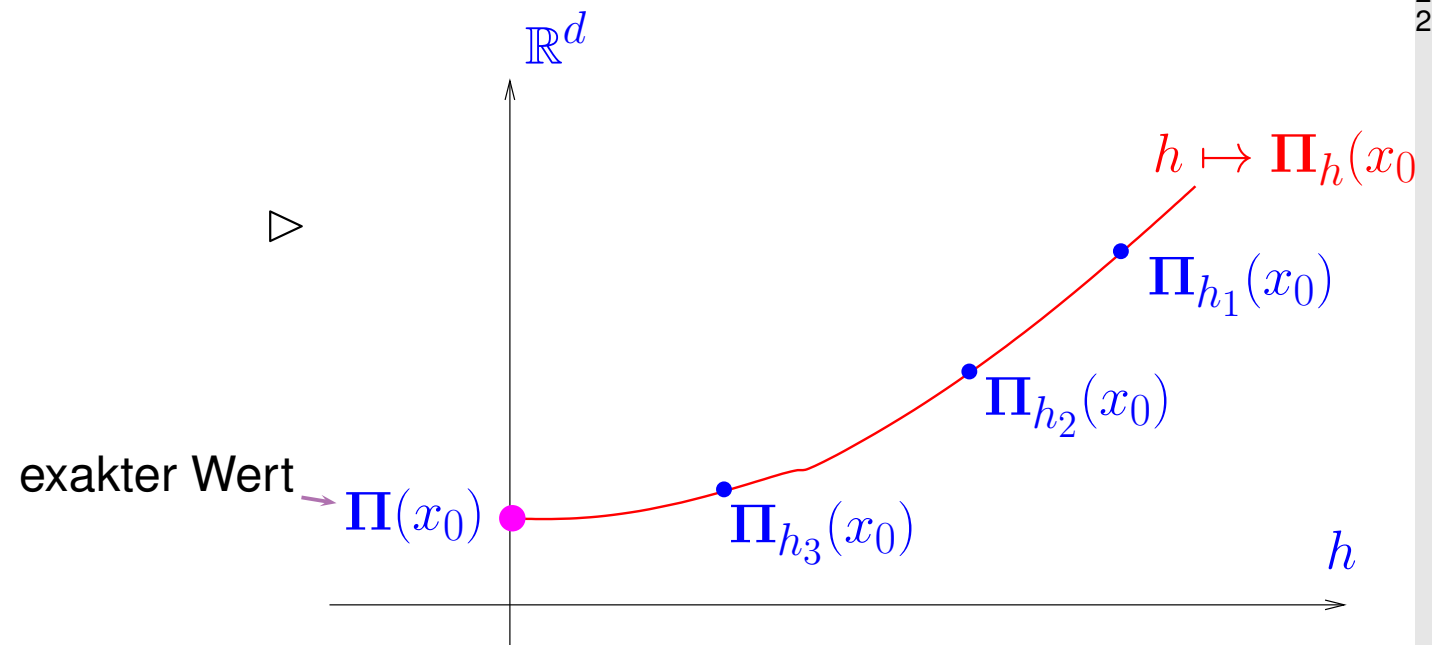


R. Hiptmair

rev 35327,  
25. April  
2011

Visualisierung:

Idee von Extrapolationsverfahren



*Bemerkung 2.4.4* (Skalierungsinvarianz der Extrapolation).

$p(t) \in \mathcal{P}_{k-1} \hat{=}$  Interpolationspolynom zu  $(t_1, y_1), \dots, (t_k, y_k)$

$\tilde{p}(t) \in \mathcal{P}_{k-1} \hat{=}$  Interpolationspolynom zu  $(\xi t_1, y_1), \dots, (\xi t_k, y_k)$  für ein  $\xi \in \mathbb{R}$



$$p(0) = \tilde{p}(0)$$

(Wenn  $p(t) = \sum_{j=0}^s a_j t^j$ , dann haben alle Polynome  $p_\xi(t) = \sum_{j=0}^s a_j (\xi t)^j$  offensichtlich den gleichen Wert für  $t = 0$ .)



Es genügt, die Verhältnisse  $\eta_i := \frac{h_1}{h_i}$  zu spezifizieren !

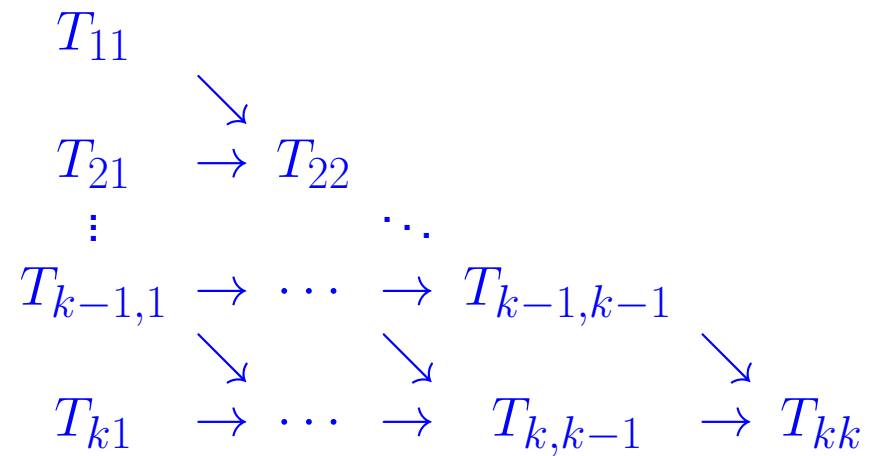


Algorithmus: **Aitken-Neville-Schema** [9, Sect. 9.4]  $\rightarrow$  Vorlesung „Numerische Methoden“

# Rekursive Berechnung der Werte von Interpolationspolynomen für $h = 0, p = 1$ :

$$T_{i1} := \Pi_{h_i}(x_0), \quad i = 1, \dots, k, \quad (2.4.5)$$

$$T_{il} := T_{i,l-1} + \frac{T_{i,l-1} - T_{i-1,l-1}}{\frac{h_{i-l+1}}{h_i} - 1}, \quad 2 \leq l \leq k. \quad (2.4.6)$$

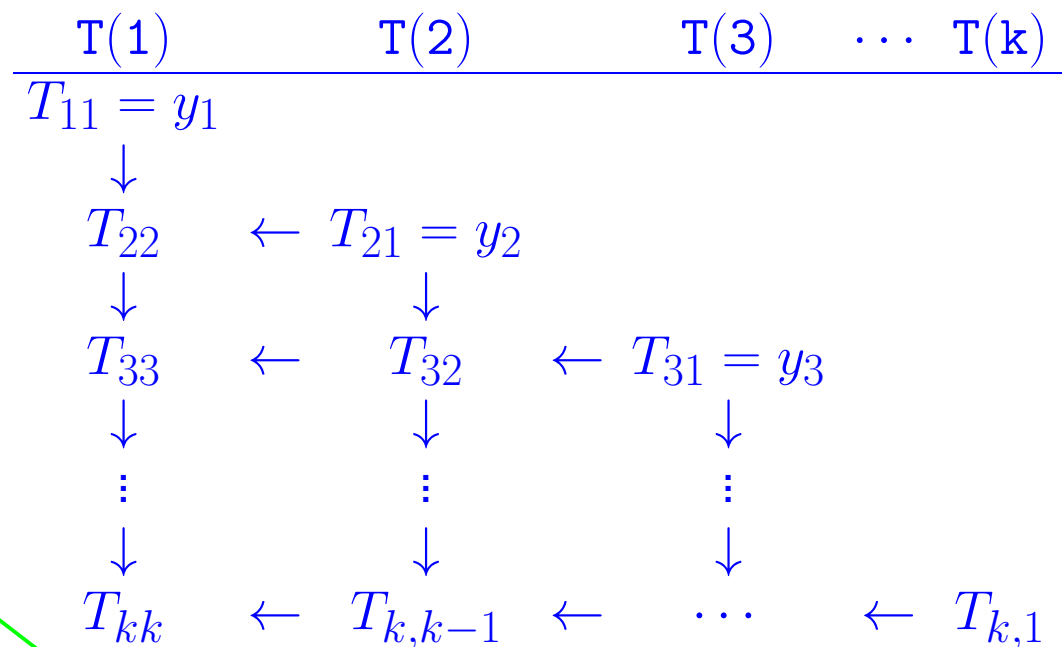


## Extrapolationstableau ▷

MATLAB-CODE : Aitken-Neville-Extrapolation

```
function T = anexpol(y,h)
k = length(h);
T(1) = y(1);
for i=2:k
    T(i) = y(i);
    for l=i-1:-1:1
        T(l) = T(l+1) + (T(l+1) - T(l)) / ...
            (h(l)/h(i) - 1);
    end
end
```

$\eta : \eta_i$



Ausgabe: unterste Tableauzeile absteigend

☞ Extrapolation „funktioniert“, wenn

- $\lim_{h \rightarrow 0} \mathbf{\Pi}_h(x_0) = \mathbf{\Pi}(x_0) \hat{=} \text{Konvergenz,}$
- $h \mapsto \mathbf{\Pi}_h(x_0)$  „sich für kleine  $h$  wie ein Polynom verhält.“

**Definition 2.4.7** ((Abgeschnittene) asymptotische Entwicklung).

$h \mapsto \mathbf{\Pi}_h(x_0)$  ( $x_0 \in X$  fest) besitzt eine (abgeschnittene) **asymptotische Entwicklung** in  $h$  bis zur Ordnung  $k$ , falls es Konstanten<sup>(\*)</sup>  $\alpha_0, \alpha_1, \dots, \alpha_k \in \mathbb{R}^d$  und eine für hinreichend kleine  $h$  **gleichmässig beschränkte** Funktion  $h \mapsto R_k(h)$  gibt, so dass

$$\mathbf{\Pi}_h(x_0) = \alpha_0 + \alpha_1 h + \alpha_2 h^2 + \dots + \alpha_k h^k + R_k(h) h^{k+1} \quad \text{für kleine } h > 0 .$$

(\*)  $\alpha_i$  Konstanten  $\hat{=} \alpha_i$  unabhängig von  $h$  !

Klar: Hinreichend & notwendig für Konvergenz:  $\alpha_0 = \mathbf{\Pi}(x_0)$



**Theorem 2.4.8** (Konvergenz extrapolierte Werte).

$\Pi_h(x_0)$  besitze asymptotische Entwicklung in  $h$  bis zur Ordnung  $k$  gemäss Def. 2.4.7. Dann erfüllen die Werte aus dem Extrapolationstableau, siehe (2.4.5), (2.4.6), für hinreichend kleine  $h_j > 0$

$$\blacktriangleright \quad \left\| T_{i,l} - \alpha_0 \right\| \leq \|\alpha_l\| h_{i-l+1} \cdots h_i + C \cdot \sum_{j=i-l+1}^i \|R_k(h_j)\| h_j^{l+1}, \quad 1 \leq i, l \leq k,$$

wobei  $C > 0$  nur von den Verhältnissen  $h_i : h_j$  abhängt.

*Beweis:* Jedes  $T_{i,k}$  aus dem Extrapolationstableau lässt sich als “Endwert”  $T_{kk}$  eines Teiltableaus interpretieren  $\blackrightarrow$  Es genügt, den Beweis für  $i = l = k$  zu führen

Voraussetzung: Existenz einer (abgeschnittenen) asymptotischen Entwicklung  $\rightarrow$  Def. 2.4.7

$$T_{i,1} = \Pi_{h_i}(x_0) = \alpha_0 + \alpha_1 h_i + \alpha_2 h_i^2 + \cdots + \alpha_k h_i^k + R_k(h) h_i^{k+1} \quad \text{für kleines } h_i > 0.$$

Extrapolationspolynom zu  $(h_i, T_{i,1})$ ,  $i = 1, \dots, k$ :  $q \in \mathcal{P}_{k-1}$ , dargestellt durch Lagrange-Polynome, siehe (2.2.2)

$$q(t) = \sum_{i=1}^k T_{i,1} L_i(t), \quad L_i \in \mathcal{P}_{k-1}, \quad L_i(h_j) = \delta_{ij}, \quad i, j = 1, \dots, k.$$

$$\sum_{i=1}^k L_i(0)h_i^j = \begin{cases} 1 & \text{für } j = 0, \\ 0 & \text{für } 1 \leq j \leq k-1, \\ (-1)^{k-1}h_1 \cdots h_k & \text{für } j = k. \end{cases} \quad (2.4.9)$$

Nachweis von (2.4.9): für  $0 \leq j \leq k-1$  stimmt  $r_j(t) := \sum_{i=1}^k h_i^j L_i(t) \in \mathcal{P}_{k-1}$  mit  $t \mapsto t^j$  überein. Für  $j = k$  hat  $t^k - r_k(t) \in \mathcal{P}_k$  die Nullstellen  $h_i, i = 1, \dots, k$  und führenden Koeffizienten 1:

$$\blacktriangleright \quad t^k - r_k(t) = (t - h_1) \cdots (t - h_k).$$

Damit folgt (2.4.9).

$$\begin{aligned} T_{k,k} = q(0) &= \sum_{i=1}^k L_i(0)T_{i,1} = \sum_{i=1}^k L_i(0) \left( \sum_{j=0}^k \alpha_j h_i^j + R_k(h_i)h_i^{k+1} \right) \\ &= \sum_{j=0}^k \alpha_j \sum_{i=1}^k h_i^j L_i(0) + \sum_{i=1}^k L_i(0)R_k(h_i)h_i^{k+1} \\ &\stackrel{(2.4.9)}{=} \alpha_0 + \alpha_k \cdot (-1)^{k-1}h_1 \cdots h_k + \sum_{i=1}^k L_i(0)R_k(h_i)h_i^{k+1}. \end{aligned}$$

Also gilt die Behauptung mit  $C := \max_{i=1, \dots, l} |L_i(0)|$ . Beachte, dass  $L_i(0)$  nur von den Verhältnissen  $h_i : h_j$  abhängt, siehe Bem. 2.4.4. □

## 2.4.3 Extrapolation von Einschrittverfahren

Anfangswertproblem (1.1.13):  $\dot{\mathbf{y}} = f(t, \mathbf{y}), \mathbf{y}(t_0) = \mathbf{y}_0 \succ$  Lösung  $t \mapsto \mathbf{y}(t)$

Das Anfangswertproblem wird betrachtet auf festem Zeitintervall  $[t_0, T]$

Annahme:  $f$  „hinreichend“ glatt  $\Leftrightarrow \mathbf{y}(t)$  „hinreichend“ glatt

Gegeben: Konsistentes Einschrittverfahren  $\Leftrightarrow$  diskrete Evolution ( $\rightarrow$  Lemma 2.1.9)

$$\Psi^{t,t+h} \mathbf{y} = \mathbf{y} + h\psi(t, \mathbf{y}, h), \quad \psi(t, \mathbf{y}, 0) = f(t, \mathbf{y}), \quad (t, \mathbf{y}) \in \Omega, \quad h \text{ klein.} \quad (2.4.10)$$

Annahmen: 

- Inkrementfunktion  $\psi$  stetig differenzierbar in  $(t, \mathbf{y})$

- ESV hat Konsistenzordnung = Konvergenzordnung  $p \in \mathbb{N}$  ( $\rightarrow$  Thm. 2.1.19)

Gegeben: Endzeitpunkt  $T \in J(t_0, \mathbf{y}_0) \Rightarrow$  uniforme Zeitschrittweite  $h = (T-t_0)/N, N \in \mathbb{N}$

Einschrittverfahren  $\Rightarrow$  Gitterfunktion  $\{\mathbf{y}_k\}_{k=0}^N, \mathbf{y}_N \approx \mathbf{y}(T)$

**Theorem 2.4.11** (Asymptotische Entwicklung des Diskretisierungsfehlers von ESV).

Es existieren ein  $K \in \mathbb{N}$  (abhängig von der Glattheit von  $\mathbf{f}$ ) und glatte Funktionen  $\mathbf{e}_i : J(t_0, \mathbf{y}_0) \mapsto \mathbb{R}^d$ ,  $i = p, p+1, \dots, p+K$ , mit  $\mathbf{e}_i(0) = 0$  und (für hinreichend kleine  $h$ ) gleichmässig beschränkte Funktionen  $(T, h) \mapsto \mathbf{r}_{k+p+1}(T, h)$ ,  $0 \leq k \leq K$ , so dass

$$\mathbf{y}_N - \mathbf{y}(T) = \sum_{l=0}^k \mathbf{e}_{l+p}(T) h^{l+p} + \mathbf{r}_{k+p+1}(T, h) h^{k+p+1} \quad \text{für kleines } h .$$

Dabei gilt

$$\begin{aligned} \|\mathbf{r}_{k+p+1}(T, h)\| &= O(T - t_0) \quad \text{für } T - t_0 \rightarrow 0 \text{ gleichmässig in } h < T, \\ \|\mathbf{e}(T)\| &= O(T - t_0) \quad \text{für } T - t_0 \rightarrow 0. \end{aligned}$$

*Beweis.* Annahme:  $\mathbf{f}, \mathbf{y}(t)$  „hinreichend“ glatt

Weiter nehmen wir *globale Lipschitz-Stetigkeit* der Inkrementfunktion  $\psi$  des ESV aus (2.4.10) an:

$$\exists L > 0: \quad \|\psi(t, \mathbf{z}, h) - \psi(t, \mathbf{w}, h)\| \leq L \|\mathbf{z} - \mathbf{w}\| \quad \text{gleichmässig in } t_0 \leq t \leq T, h. \quad (2.4.12)$$

(Kompaktheitsargumente, vgl. Beweis von Thm. 2.1.19, machen Verzicht auf diese Annahme möglich.)

Konsequenz der Konsistenzordnung  $p$  ( $\rightarrow$  Def. 2.1.13) und Glattheit von  $\mathbf{f}$ : für Konsistenzfehler ( $\rightarrow$

Def. 2.1.11) entlang der Lösungstrajektorie (nur dort wird die Konsistenzfehlerabschätzung im Beweis von Thm. 2.1.19 gebraucht !) gilt

$$\tau(t, \mathbf{y}(t), h) := \mathbf{y}(t+h) - \Psi^{t, t+h} \mathbf{y}(t) = \mathbf{d}(t)h^{p+1} + O(h^{p+2}) \quad \text{für } h \rightarrow 0, \quad (2.4.13)$$

mit stetiger Funktion  $\mathbf{d} : [t_0, T] \mapsto \mathbb{R}^d$ . Dies ergibt sich mit **Taylorentwicklung**, siehe Bsp. 2.3.24:

RK-ESV:  $\mathbf{d}$  hängt nur von Ableitungen von  $\mathbf{f}$  ab  $\Rightarrow$   $\mathbf{d}$  “hinreichend glatt”

Idee: Betrachte ESV mit **modifizierter Inkrementfunktion**

$$\hat{\psi}(t, \mathbf{u}, h) := \psi(t, \mathbf{u} + \mathbf{e}(t)h^p, h) - (\mathbf{e}(t+h) - \mathbf{e}(t))h^{p-1}, \quad (2.4.14)$$

mit “hinreichend glatter” Funktion  $\mathbf{e} : [t_0, T] \mapsto \mathbb{R}^d$ .

Beachte: Auch  $\hat{\psi}$  erfüllt (2.4.12) mit dem gleichen  $L > 0$ .

Warum betrachten wir dieses modifizierte ESV ?

$\mathbf{y}_j / \hat{\mathbf{y}}_j$ ,  $j = 0, \dots, N \hat{=}$  Gitterfunktionen erzeugt durch ursprüngliches/modifiziertes ESV mit Zeitschrittweite  $h := \frac{(T-t_0)}{N}$ . Setze  $\hat{\mathbf{y}}_0 = \mathbf{y}_0$

$$\blacktriangleright \quad \hat{\mathbf{y}}_j = \mathbf{y}_j - \mathbf{e}(t_j)h^p, \quad t_j := t_0 + jh, \quad j = 0, \dots, N. \quad (2.4.15)$$

Beweis von (2.4.15) durch Induktion:

$$\begin{aligned}
 \widehat{\mathbf{y}}_{j+1} &= \widehat{\mathbf{y}}_j + h\widehat{\boldsymbol{\psi}}(t_j, \widehat{\mathbf{y}}_j, h) \\
 &\stackrel{(2.4.14)}{=} \widehat{\mathbf{y}}_j + h\boldsymbol{\psi}(t_j, \widehat{\mathbf{y}}_j + \mathbf{e}(t_j)h^p, h) - h^p(\mathbf{e}(t_{j+1}) - \mathbf{e}(t_j)) \\
 &\stackrel{(*)}{=} \underbrace{\mathbf{y}_j + h\boldsymbol{\psi}(t_j, \mathbf{y}_j, h)}_{=\mathbf{y}_{j+1}} - \mathbf{e}(t_{j+1})h^p .
 \end{aligned}$$

(\*) ← Induktionsannahme.

**Annahme:** Das modifizierte Einschrittverfahren ist konsistent mit  $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$  zur Ordnung  $p + 1$

$$\text{Thm. 2.1.19} \quad \widehat{\mathbf{y}}_N - \mathbf{y}(T) = \mathbf{r}_{p+1}(T, h)h^{p+1}, \quad \|\mathbf{r}_{p+1}(T, h)\| \leq C \underbrace{\frac{\exp(L(T - t_0)) - 1}{L}}_{=O(T-t_0) \text{ für } T-t_0 \rightarrow 0},$$

mit  $C > 0$  unabhängig von  $h, T$ .  $L \hat{=}$  gemeinsame Lipschitz-Konstante von  $\boldsymbol{\psi}, \widehat{\boldsymbol{\psi}}$  (bzgl.  $\mathbf{y}$ ) aus (2.4.12)

$$\stackrel{(2.4.15)}{\Rightarrow} \mathbf{y}_N - \mathbf{y}(T) = \mathbf{e}(T)h^p + \mathbf{r}_{p+1}(T, h)h^{p+1} .$$

Damit haben wir das erste Glied der asymptotischen Entwicklung des Theorems erhalten.

Induktive Anwendung des Arguments  $\triangleright$  Modifizierte ESVen  $\widehat{\Psi}_1 := \widehat{\Psi}, \widehat{\Psi}_2, \dots, \widehat{\Psi}_{k+1}$  konsistent zu  $\dot{\mathbf{y}} = f(t, \mathbf{y})$  mit Ordnungen  $p+1, p+2, \dots, p+k+1$  erzeugen Näherungslösungen  $\widehat{\mathbf{y}}_j^1 := \widehat{\mathbf{y}}_j, \widehat{\mathbf{y}}_j^2, \dots, \widehat{\mathbf{y}}_j^{k+1}, j = 1, \dots, N$ .

Mit  $\widehat{\mathbf{y}}_k^0 = \mathbf{y}_k$  (Teleskopsumme)

$$\widehat{\mathbf{y}}_j^{l+1} = \widehat{\mathbf{y}}_j^l - \mathbf{e}_l(t_j)h^{p+l}, \quad l = 0, \dots, k.$$

$$\begin{aligned} \blacktriangleright \quad \mathbf{y}_N - \mathbf{y}(T) &= \sum_{l=0}^k \widehat{\mathbf{y}}_N^l - \widehat{\mathbf{y}}_N^{l+1} + \mathbf{r}_{p+k+1}(T, h)h^{p+k+1} \\ &= \sum_{l=0}^k \mathbf{e}_l(T)h^{p+l} + \mathbf{r}_{p+k+1}(T, h)h^{p+k+1}. \end{aligned}$$

Daraus folgt die Behauptung des Theorems.

**?** Existenz von  $\mathbf{e}(t)$  so dass das modifizierte ESV Konsistenzordnung  $p+1$  besitzt.

**☞** Betrachte den Konsistenzfehler des modifizierten Verfahrens & Taylorentwicklung(en)

$$\mathbf{y}(t+h) - \widehat{\Psi}^{t,t+h} \mathbf{y}(t) = \mathbf{y}(t+h) - \mathbf{y}(t) - h\widehat{\psi}(t, \mathbf{y}(t), h)$$

$$\begin{aligned}
&= \mathbf{y}(t+h) - \mathbf{y}(t) - h\boldsymbol{\psi}(t, \mathbf{y}(t) + \mathbf{e}(t)h^p, h) + (\mathbf{e}(t+h) - \mathbf{e}(t))h^p \\
&= \mathbf{y}(t+h) - \mathbf{y}(t) - h \left( \boldsymbol{\psi}(t, \mathbf{y}(t), h) + \frac{\partial \boldsymbol{\psi}}{\partial \mathbf{y}}(t, \mathbf{y}(t), h)\mathbf{e}(t)h^p + O(h^{2p}) \right) + \dot{\mathbf{e}}(t)h^{p+1} + O(h^{p+2}) \\
&\stackrel{(2.4.13)}{=} \mathbf{d}(t)h^{p+1} + O(h^{p+2}) - \left( \frac{\partial \boldsymbol{\psi}}{\partial \mathbf{y}}(t, \mathbf{y}(t), 0) + \frac{\partial^2 \boldsymbol{\psi}}{\partial \mathbf{y} \partial h}(t, \mathbf{y}(t), 0)h \right) \mathbf{e}(t)h^{p+1} \\
&\quad + \dot{\mathbf{e}}(t)h^{p+1} + O(h^{p+2}) \\
&= (\mathbf{d}(t) - \frac{\partial \mathbf{f}}{\partial \mathbf{y}}(t, \mathbf{y}(t))\mathbf{e}(t) + \dot{\mathbf{e}}(t))h^{p+1} + O(h^{p+2}) .
\end{aligned}$$

► Löst  $\mathbf{e}$  folgendes AWP für eine inhomogene linear Variationsgleichung

$$\dot{\mathbf{e}}(t) = \frac{\partial \mathbf{f}}{\partial \mathbf{y}}(t, \mathbf{y}(t))\mathbf{e}(t) - \mathbf{d}(t) \quad , \quad \mathbf{e}(0) = 0 \quad \Rightarrow \quad \mathbf{e} \text{ glatt} \quad , \quad (2.4.16)$$

dann ist die Annahme über das modifizierte ESV erfüllt. □

Logisch:  $K$  hängt von der Glattheit von  $\mathbf{f}$  ab.



Idee: **Ordnungserhöhung** durch Extrapolation ( $\rightarrow$  Sect. 2.4.2)

- Wähle  $N_1 < N_2 < \dots < N_{k+1}$ ,  $N_i \in \mathbb{N}$
- ESV (Schrittweite  $h_i = (T-t_0)/N_i$ ) liefert  $\mathbf{y}_{N_i}$ ,  $i = 1, \dots, k+1$
- Polynomextrapolation (\*) aus  $(h_i, \mathbf{y}_{h_i, N_i})$ 
  - ↳ Näherung  $\tilde{\mathbf{y}}$  mit  $\|\tilde{\mathbf{y}} - \mathbf{y}(T)\| = O(h_1^{p+k})$  (vgl. Thm. 2.4.8)

(\*): Thm. 2.4.11  $\blacktriangleright$  Extrapolation basierend auf Polynom der Form

$$p(t) = \alpha_0 + \alpha_p h^p + \alpha_{p+1} h^{p+1} + \dots + \alpha_{p+k-1} h^{p+k-1} !$$

- Thm. 2.4.11 erfordert „hinreichend kleines“  $h$
- Nicht nur  $\mathbf{y}(T)$  von Interesse, sondern (genäherte) Lösung  $t \mapsto \mathbf{y}(t)$

## 2.4.4 Lokale Extrapolations-Einschrittverfahren

$\blacktriangleright$  Anwendung der Extrapolationsidee auf Intervallen eines Zeitgitters  $\mathcal{G} := \{t_0 < t_1 < \dots < t_N = T\} \Leftrightarrow$  **Makroschritte**: auf  $[t_j, t_{j+1}]$ , Makroschrittweite  $H_j := t_{j+1} - t_j$

- Fixiere Sequenz  $(n_l)_{l=1}^{k+1}$ ,  $n_l \in \mathbb{N}$ , z.B.  $(1, 2, 3, 4, 5, 6, \dots)$   $\leftrightarrow$  Anzahl **Mikroschritte**
- $n_l$  Schritte des ESV mit Startwert  $\mathbf{y}_j$ , Schrittweite  $h = \frac{t_{j+1} - t_j}{n_l} \mapsto \mathbf{y}_{j+1}^l$ ,  $l = 1, \dots, k+1$   
(ESV = **Basisverfahren**, Ordnung  $p$ )
- Polynomextrapolation (\*) aus  $(n_l^{-1}, \mathbf{y}_{j+1}^l) \mapsto \mathbf{y}_{j+1}$  vgl. Bem. 2.4.4

  
 Extrapolations-Einschrittverfahren der Ordnung  $p + k$

MATLAB-CODE : Einzelschritt, lokales Extrapolations-ESV, skalare ODE

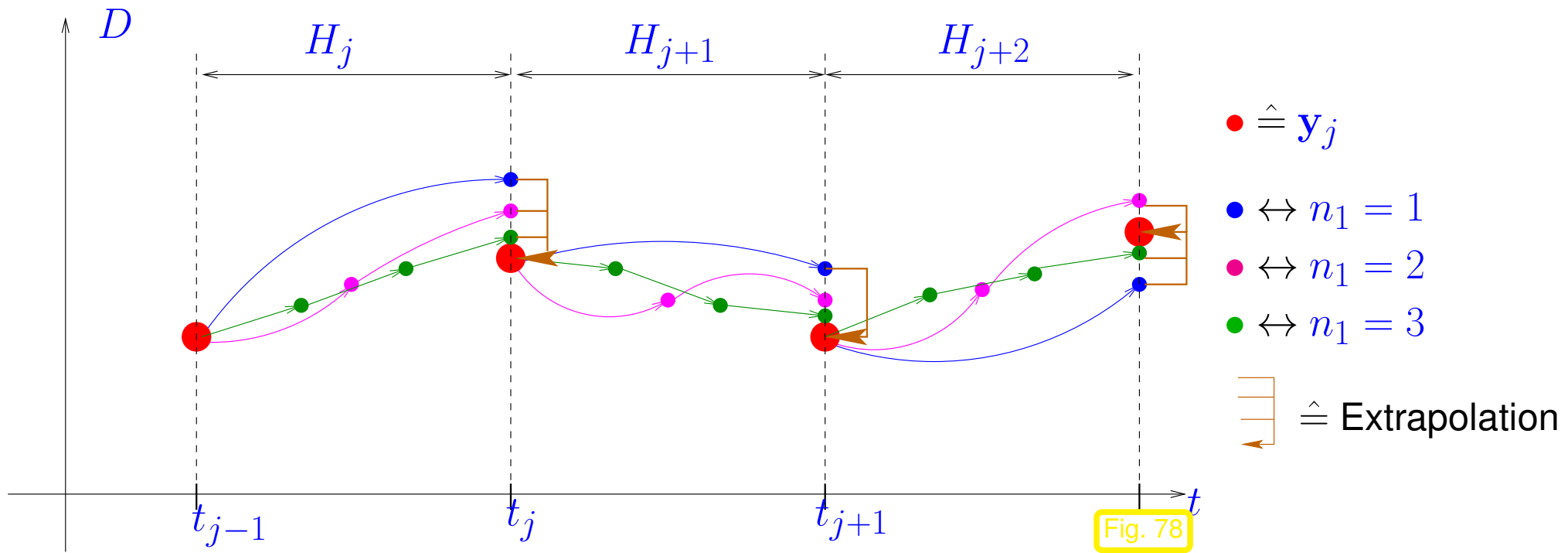
```
function y = expesvstep(esvstep, y, t, h, n)
for i=1:length(n)
    yt(i) = y;
    ht = h/n(i); tt = t;
    for j=1:n(i)
        yt(i) = esvstep(yt(i), tt, ht);
        tt = tt + ht;
    end
T = anexpol(yt, 1./n, p);
return(T(1));
```

$\text{esvstep}(y, t, h) \hat{=}$  ein Schritt  
des Basisverfahrens, Schrittweite  
 $h$ , ausgehend vom Zustand  $(t, \mathbf{y})$ :  
 $\text{esvstep}(y, t, h) := \Psi^{t, t+h} \mathbf{y}$

$n \hat{=}$  Vektor  $(n_l)_{l=1}^{k+1}$

$\text{anexpol} \hat{=}$  verallgemeinerte  
Version für Extrapolationspolynom  
 $p(t) = \alpha_0 + \alpha_p h^p + \alpha_{p+1} h^{p+1} + \dots + \alpha_{p+k-1} h^{p+k-1}$

Ablauf: Lokales Extrapolations-Einschrittverfahren ( $n_1 = 1, n_2 = 2, n_3 = 3$ )



Beispiel 2.4.17 (Extrapoliertes Euler-Verfahren).

AWP für logistische Dgl. ( $\rightarrow$  Bsp. 1.2.1)

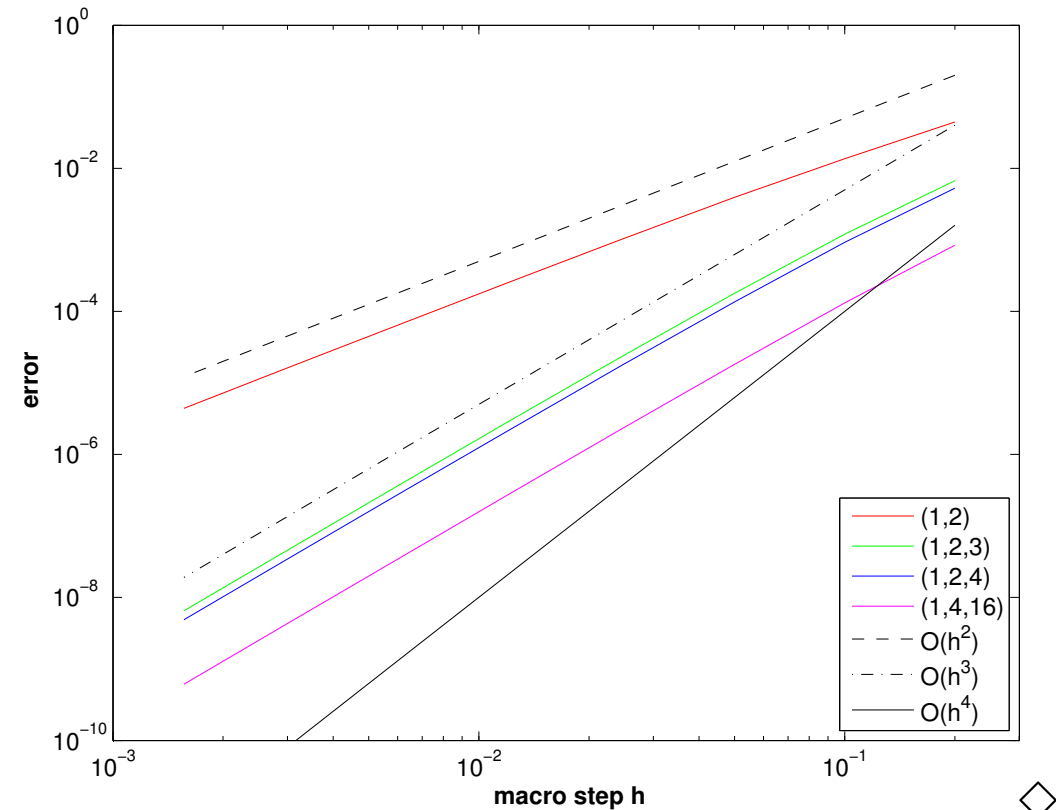
$$\dot{y} = 5y(1 - y), \quad y(0) = 0.02.$$

Endzeit:  $T = 1$  Basis-ESV: explizites Euler-Verfahren (1.4.2)

Extrapolation: verschiedene  $k, n_l, l = 1, \dots, k+1$ ,  
uniforme Makroschrittweite  $h$

$$\text{Fehler } \max_j |y(t_j) - y_j|.$$

Algebraische Konvergenz der Ordnung  $k + 1$   $\triangleright$



R. Hiptmair

rev 35327,  
25. April  
2011

Theoretische Analyse  $\leftrightarrow$  Verifikation der Voraussetzungen von Thm. 2.1.19

Überlegungen für Spezialfall  $p = 1$   $\leftrightarrow$  Euler-Verfahren, vgl. Bsp. 2.4.17

Notationen:

$\bullet$   $t \mapsto \mathbf{y}(t) \hat{=}$  exakte Lösung durch  $(t, \mathbf{y}) \in \Omega$

- $H > 0 \hat{=}$  Schrittweite des Makroschritts
- $n_1, \dots, n_{k+1} \hat{=}$  Anzahl von Mikroschritten in  $[t, t + H]$
- $\mathbf{y}_N \hat{=}$  Resultat der Anwendung von  $N$  Schritten des Basis-Einschrittverfahrens auf  $[t, t + H]$  mit uniformer Schrittweite  $h := H/N$  und Startwert  $\mathbf{y}$
- $\hat{\mathbf{y}} \hat{=}$  durch Extrapolation aus  $\mathbf{y}_{n_1}, \mathbf{y}_{n_2}, \dots, \mathbf{y}_{n_{k+1}}$  gewonnener Näherungswert für  $\mathbf{y}(t + H)$

► Konsistenzfehler ( $\rightarrow$  Def. 2.1.11):  $\boldsymbol{\tau}(t, \mathbf{y}, H) = \mathbf{y}(t + H) - \hat{\mathbf{y}}$ .

Zu zeigen ist: Konsistenzordnung  $k + 1 \Leftrightarrow \|\boldsymbol{\tau}(t, \mathbf{y}, H)\| = O(H^{k+2})$

① Lokale Anwendung von Thm. 2.4.11 mit  $t_0 = t, T = t + H$ : für hinreichend grosses  $K \in \mathbb{N}$

$$\Rightarrow \mathbf{y}_N - \mathbf{y}(t + H) = \sum_{l=1}^K \mathbf{e}_l(t + H) h^l + \mathbf{r}_K(t + H, h) h^{K+1}, \quad h = H/N,$$

wobei  $\Rightarrow \|\mathbf{r}_K(t + H, h)\| \leq CH$

$\Rightarrow \|\mathbf{e}(t + H)\| \leq CH$

mit  $C > 0$  unabhängig von  $t$  und (hinreichend kleinem)  $h$ .

② Damit aus Thm. 2.4.8

$$\|\hat{\mathbf{y}} - \mathbf{y}(t + H)\| \leq \|\mathbf{e}_{k+1}(t + H)\| h_1 \cdots h_{k+1} + C \sum_{j=1}^k \|\mathbf{r}_j(t + H, h_j)\| h_j^{k+2},$$

wobei  $C > 0$  nur von den Verhältnissen  $n_j : n_l$  abhängt.

$$\Rightarrow \|\hat{\mathbf{y}} - \mathbf{y}(t + H)\| \leq CH^{k+2},$$

mit  $C > 0$  unabhängig von  $H$ .

*Bemerkung 2.4.18* (Extrapolationsverfahren als Runge-Kutta-Verfahren).

Basisverfahren: Explizites Euler-Verfahren (1.4.2), Ordnung  $p = 1$

→ Polynomextrapolation zur Sequenz  $(1, 2, 3, 4, \dots, k)$  liefert explizites (→ Def. 2.1.5)  
Runge-Kutta-Verfahren (→ Def. 2.3.5) der Ordnung  $k$  mit  $s = k(k - 1)/2 + 1$  Stufen.



## 2.4.5 Ordnungssteuerung

Für Extrapolations-Einschrittverfahren:  $k = k(j)$  einfach zu realisieren

Idee: [(lokale) **Ordnungssteuerung**]

„Erhöhe lokale Ordnung, bis es sich nicht mehr lohnt (\*)“

(\*) Heuristisches Beurteilungskriterium (basierend auf Aitken-Neville-Extrapolationstableau (2.4.6)):

$$\left\| T_{k,k-1} - T_{k,k} \right\| \leq \text{TOL} \cdot \left\| T_{k,k} \right\| \quad \text{für Toleranz TOL} > 0 .$$

Zweitbeste Näherung beste Näherung

```
function [y,k] = eulexstep(f,y,H,TOL)
kmax = 1000;
T{1} = y + H*f(y);
for i=2:kmax
    T{i} = y; h = H/i;
    for k=1:i, T{i} = T{i} + h*f(T{i}); end
    for l=i-1:-1:1
        T{l} = T{l+1} + (T{l+1}-T{l})/(i/l-1);
    end
    if (norm(T{1}-T{2}) < TOL*norm(T{1}))
        y = T{1}; k = i; return;
    end
end
```

## Adaptives

### Euler-Extrapolationsverfahren:

(für autonomes AWP)

Makroschritt der Länge  $H$

Argumente:

$f$  : Funktionshandle  $f=@(y)$   
auf rechte Seite

$y$  : Anfangswert zu  $t = 0$

TOL : Toleranz

Rückgabewerte:

$y$  : Näherung zu  $t = H$

$k$  : Verwendete Extrapolationstiefe

Beachte: Einfache Erweiterung des Extrapolationstableaus um eine weitere Zeile

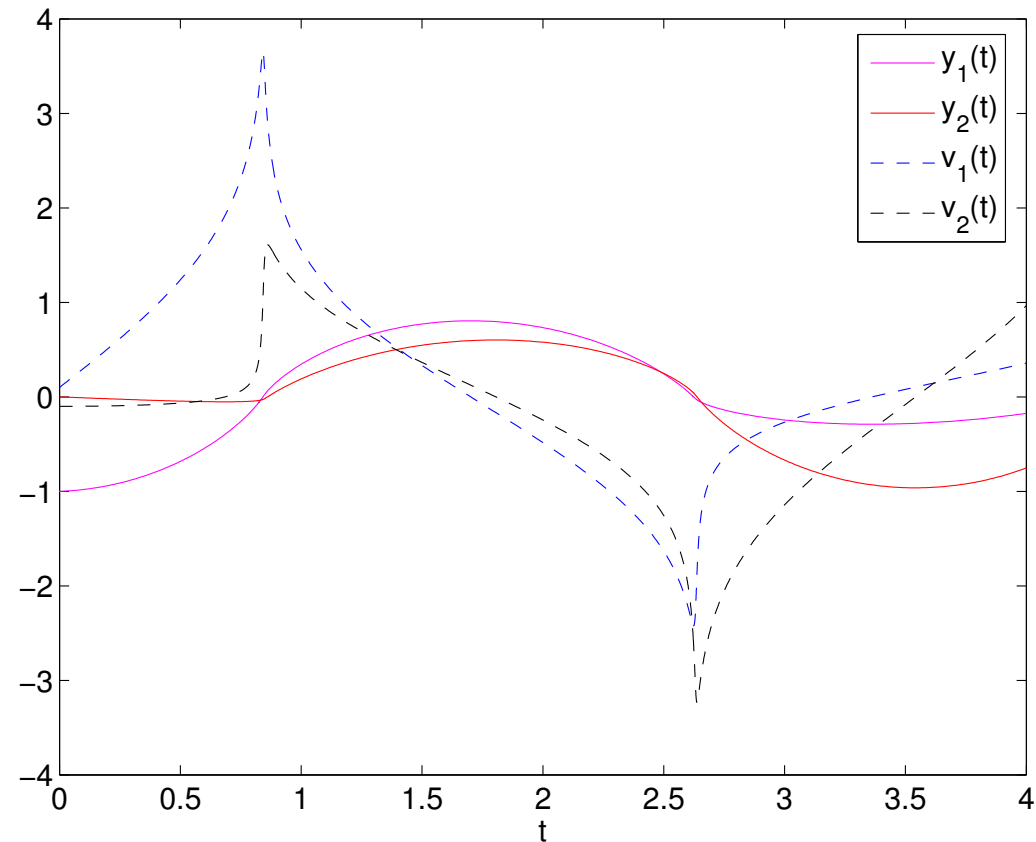
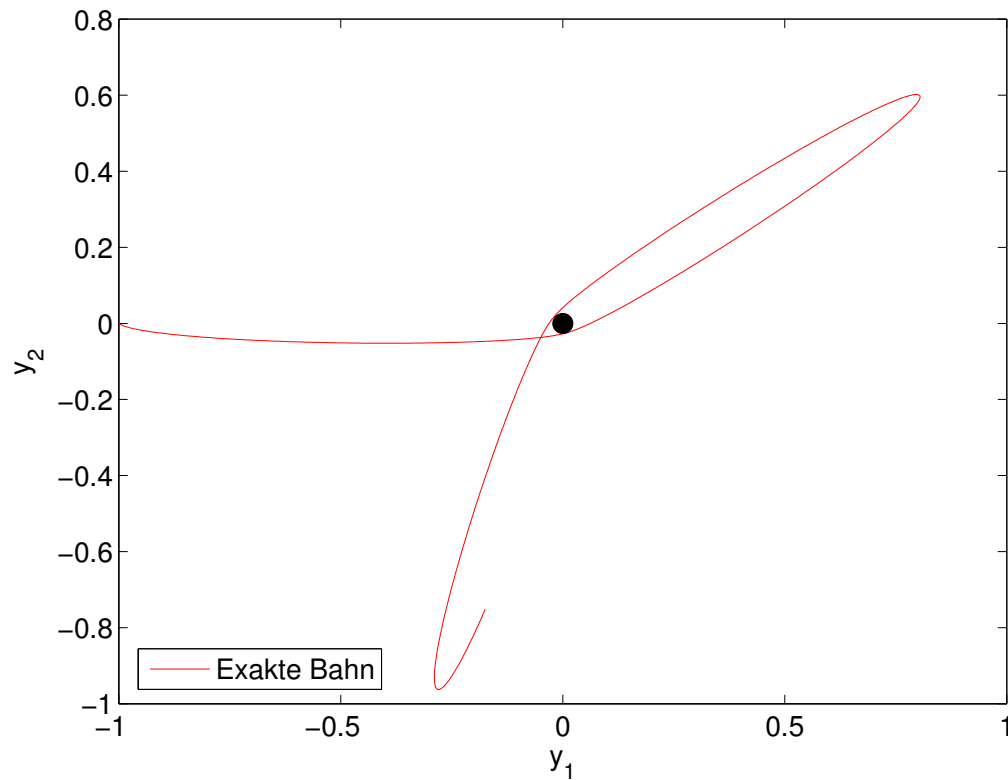
*Beispiel 2.4.19* (Euler-Extrapolationsverfahren mit Ordnungssteuerung).

Bewegung eines geladenen Teilchens im Feld eines geraden Drahtes = Linienladung (konservatives  
Zentralfeld, Zentrum  $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ , Potential  $U(\mathbf{x}) := -2 \log \|\mathbf{x}\|$ ):  $\rightarrow$  Bsp. 1.2.25

$$\ddot{\mathbf{y}} = -\frac{2\mathbf{y}}{\|\mathbf{y}\|^2} \Rightarrow \begin{pmatrix} \dot{\mathbf{y}} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{v} \\ -\frac{2\mathbf{y}}{\|\mathbf{y}\|^2} \end{pmatrix}, \quad \mathbf{y}(0) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad \mathbf{v}(0) = \begin{pmatrix} 0.1 \\ -0.1 \end{pmatrix}.$$

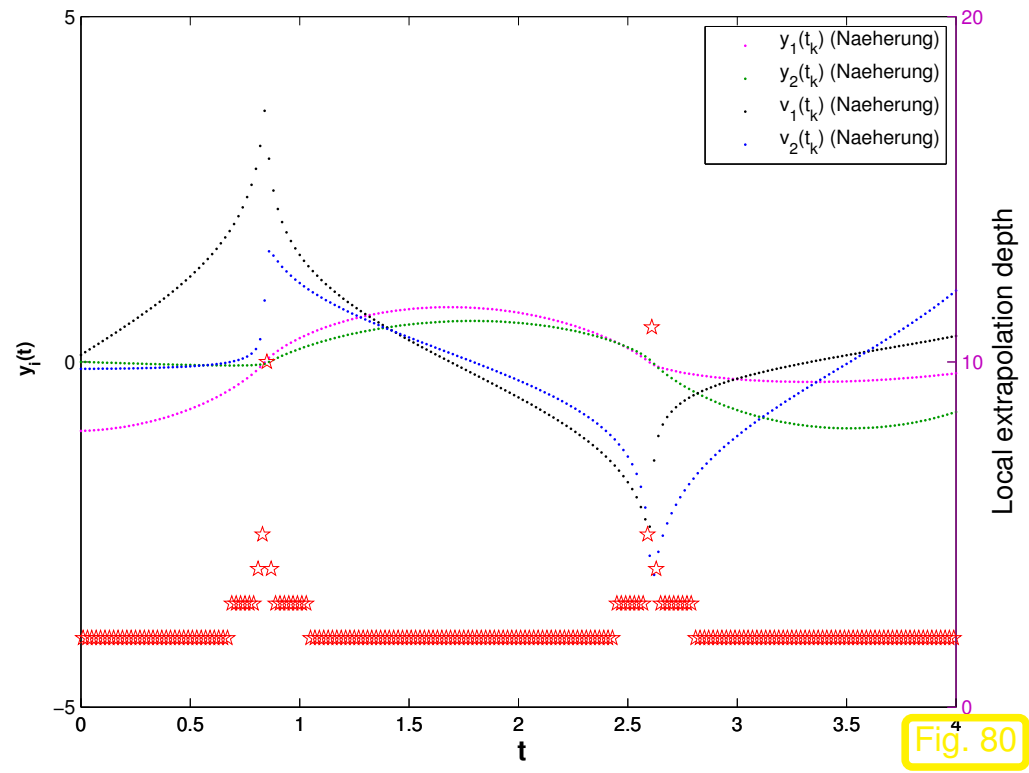
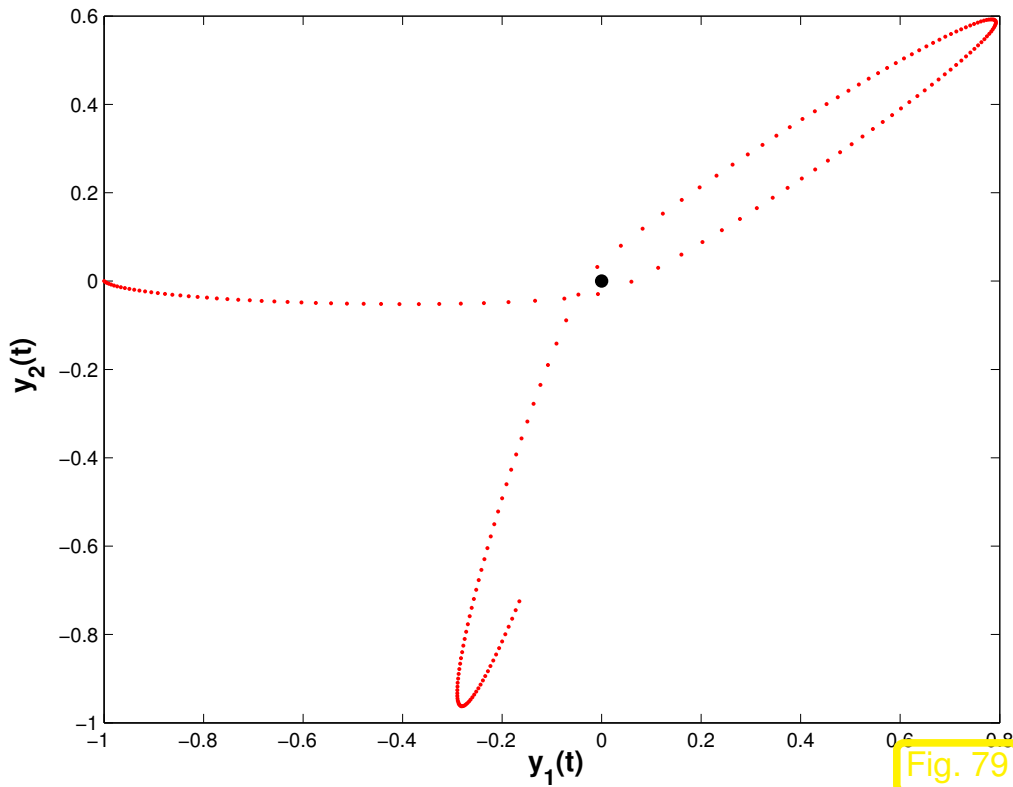


Anfangswert:  $\mathbf{y}(0) = (-1, 0, 0.1, -0.1)$ , Endzeitpunkt:  $T = 4$



Beobachtung: „Peaks“ in der Lösungskomponente  $\mathbf{v}(t)$  (= zeitlokale Charakteristika)

Ordnungsadaptives Euler-Extrapolationsverfahren ( $\text{TOL} = 0.01$ ), uniforme Makrozeitschrittweite  $H = 0.02$



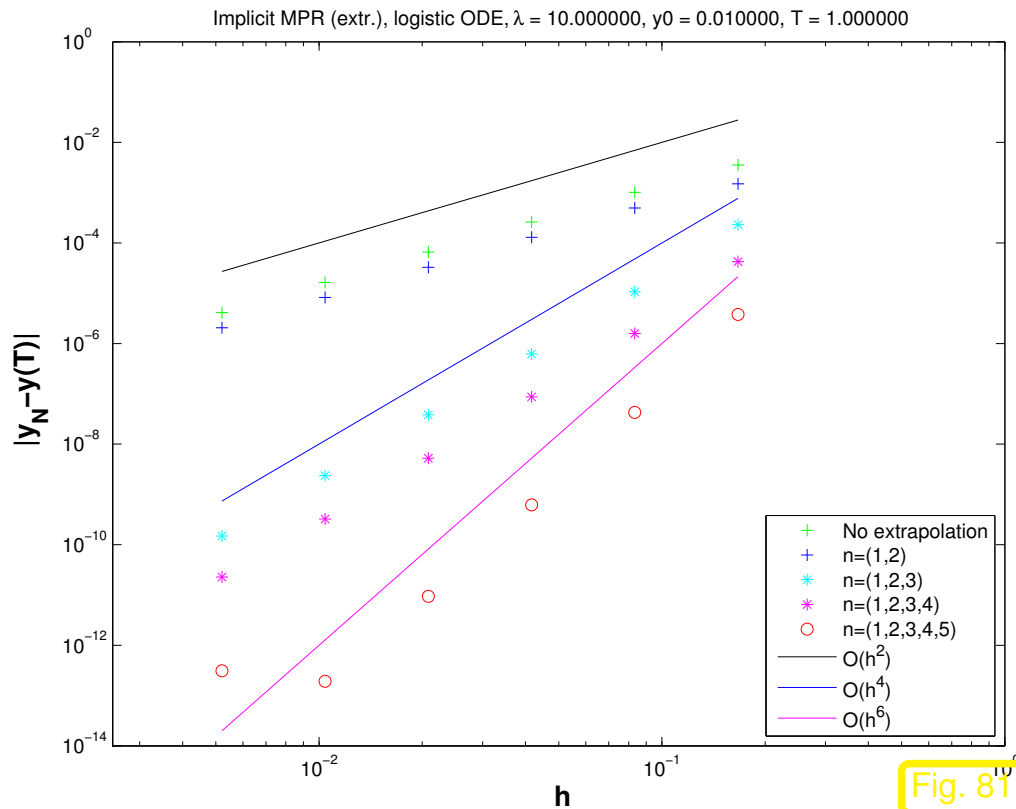
▷ Automatische Erhöhung der Ordnung an „kritischen Stellen“



## 2.4.6 Extrapolation reversibler Einschrittverfahren

Beispiel 2.4.20 (Extrapolierte implizite Mittelpunktsregel).

- Anfangswertproblem aus Bsp. 1.4.9 (logistische Dgl., siehe Bsp. 1.2.1),  $\lambda = 10$ ,  $y_0 = 0.01$ , aus  $[0, 1]$  ( $T = 1$ )
  - Einschrittverfahren: implizite Mittelpunktsregel (1.4.19)
  - Globale Extrapolation ( $\rightarrow$  Abschnitt 2.4.3) von  $y_h(T)$  aus Lösungen erhalten durch uniforme Schrittweiten  $h/n_i$
- Beachte: Extrapolation auf der Grundlage des Standard-Tableaus (2.4.5)



◁ Ordnungserhöhung nur in jedem zweite Extrapolationsschritt

Ordnungserhöhung um jeweils zwei



Beobachtung aus Bsp. 2.4.20 einfach zu erklären, falls

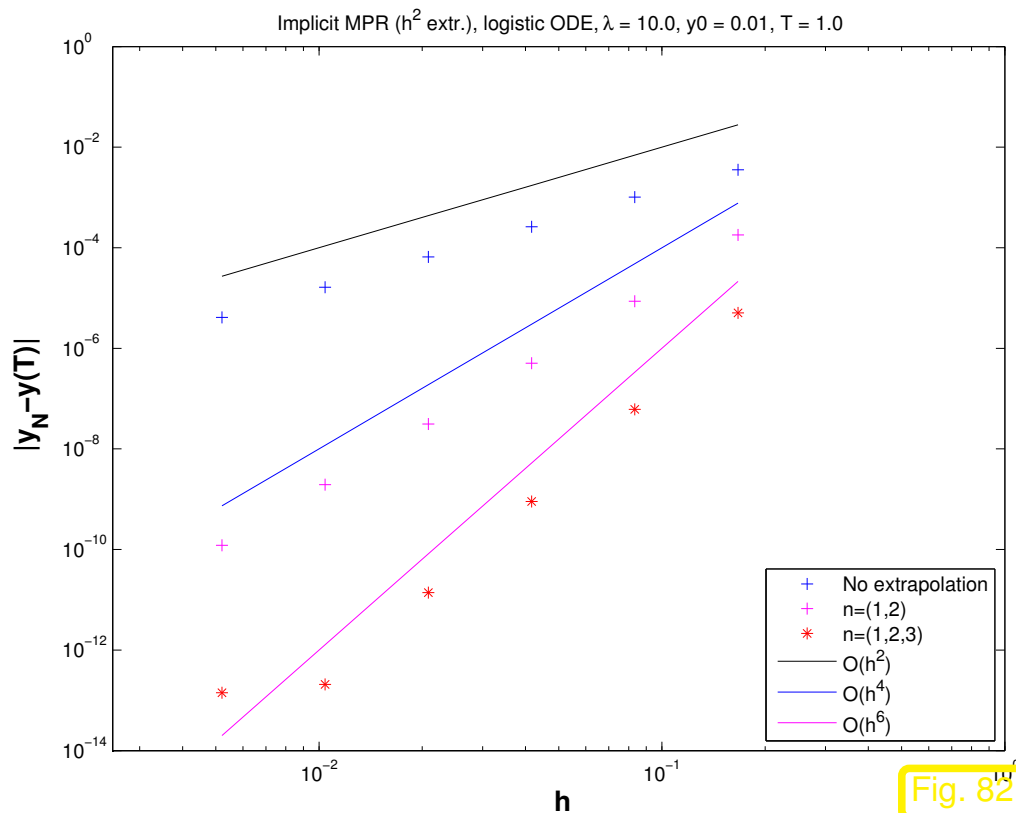
$$y_h(T) = y(T) + \alpha_1 h^2 + \alpha_2 h^4 + \alpha_6 h^6 + \dots .$$

*Beispiel 2.4.21* (Globale  $h^2$ -Extrapolation für implizite Mittelpunktsregel).

- (Fast) wie Bsp. 2.4.20

- NEU:

$y_N$  aus Extrapolation in  $h^2$



◁ Ordnungserhöhung um **zwei** in jedem Extrapolationsschritt !

**Theorem 2.4.22** (Asymptotische Entwicklung des Diskretisierungsfehlers in  $h^2$ ).

Bezeichne  $\mathbf{y}_h(t)$ ,  $t \in$  äquidistantes Zeitgitter mit Schrittweite  $h > 0$  auf  $[t_0, T]$ , die durch ein reversibles Einschrittverfahren ( $\rightarrow$  Def. 2.1.27) erzeugte Näherungslösung eines Anfangswertproblems  $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y})$ ,  $\mathbf{y}(t_0) = \mathbf{y}_0$ , mit exakter Lösung  $t \mapsto \mathbf{y}(t)$ .

Dann existieren ein  $K \in \mathbb{N}$  (abhängig von der Glattheit von  $\mathbf{f}$ ) und glatte Funktionen  $\mathbf{e}_i : J(t_0, \mathbf{y}_0) \mapsto \mathbb{R}^d$ ,  $i = 1, \dots, K$ , mit  $\mathbf{e}_i(0) = 0$  und (für hinreichend kleine  $h$ ) gleichmässig beschränkte Funktionen  $(T, h) \mapsto \mathbf{r}_k(T, h)$ ,  $0 \leq k \leq K$ , so dass

$$\mathbf{y}_h(T) - \mathbf{y}(T) = \sum_{l=1}^k \mathbf{e}_l(T) h^{2l} + \mathbf{r}_k(T, h) h^{2k+2} \quad \text{für kleines } h .$$

Dabei gilt  $\|\mathbf{r}_k(T, h)\| = O(T - t_0)$  für  $T - t_0 \rightarrow 0$  gleichmässig in  $h < T$ .

*Beweis.* Siehe [8, Satz 4.42] □

Bemerkung 2.4.23 (DIFEX).

Praktische Extrapolationsverfahren stützen sich auf *explizite* Verfahren, deren Fehler eine **asymptotische Entwicklung in  $h^2$**  besitzt (eine spezielle Trapezregel)



DIFEX-Algorithmus [8, Sect. 4.3.3]



## 2.5 Splittingverfahren [16, Sect. 2.5]

R. Hiptmair  
rev 35327,  
25. April  
2011

Autonomes AWP mit additiv zerlegter rechter Seite:

$$\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}) + \mathbf{g}(\mathbf{y}) \quad , \quad \mathbf{y}(0) = \mathbf{y}_0 \quad , \quad (2.5.1)$$

mit  $\mathbf{f} : D \subset \mathbb{R}^d \mapsto \mathbb{R}^d$ ,  $\mathbf{g} : D \subset \mathbb{R}^d \mapsto \mathbb{R}^d$  “hinreichend glatt”, lokal Lipschitz-stetig ( $\rightarrow$  Def. 1.3.2)

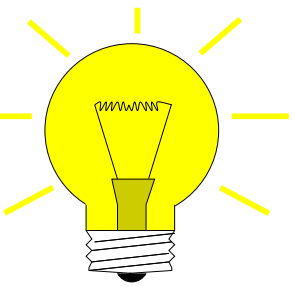
(Kontinuierliche) Evolutionen:

$$\Phi_f^t \leftrightarrow \text{Dgl. } \dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}) \quad ,$$

$$\Phi_g^t \leftrightarrow \text{Dgl. } \dot{\mathbf{y}} = \mathbf{g}(\mathbf{y}) \quad .$$

Annahme:  $\Phi_f^t, \Phi_g^t$  (analytisch) bekannt

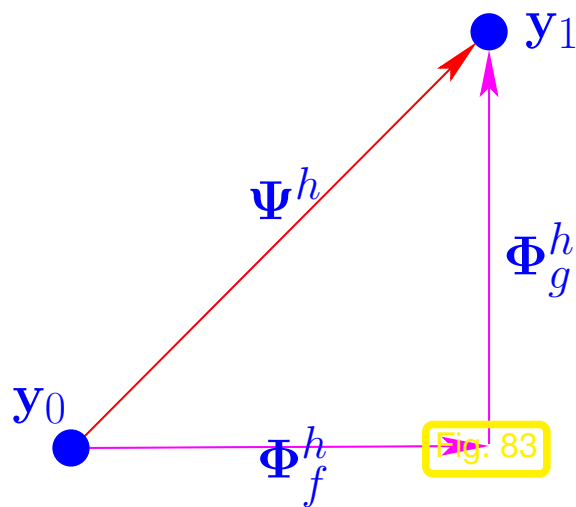
Idee: Konstruiere Einschrittverfahren mit diskreten Evolutionsen



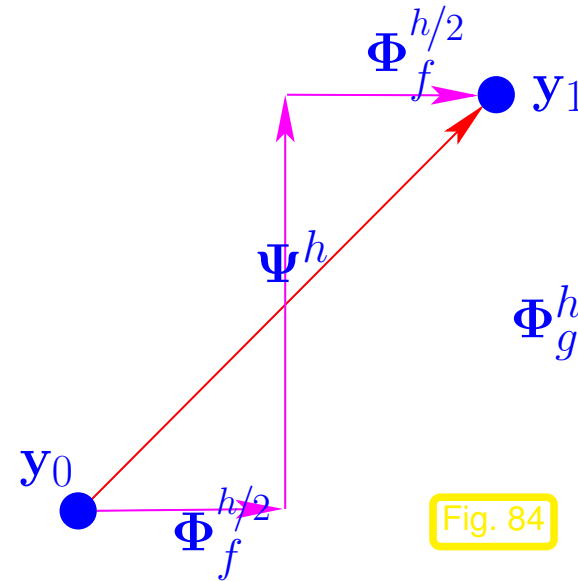
**Lie-Trotter-Splitting:**  $\Psi^h = \Phi_g^h \circ \Phi_f^h, \tag{2.5.2}$

**Strang-Splitting:**  $\Psi^h = \Phi_f^{h/2} \circ \Phi_g^h \circ \Phi_f^{h/2}. \tag{2.5.3}$

(2.5.2)  $\leftrightarrow$



(2.5.3)  $\leftrightarrow$



R. Hiptmair  
rev 35327,  
25. April  
2011

Beispiel 2.5.4 (Konvergenz einfacher Splittingverfahren).

$$\dot{y} = \underbrace{\lambda y(1 - y)}_{=:f(y)} + \underbrace{\sqrt{1 - y^2}}_{=:g(y)}, \quad y(0) = 0.$$

▶  $\Phi_{fy}^t = \frac{1}{1 + (y^{-1} - 1)e^{-\lambda t}}$ ,  $t > 0, y \in ]0, 1]$  (Logistische Differentialgleichung (2.2.84))

▶  $\Phi_{gy}^t = \begin{cases} \sin(t + \arcsin(y)) & , \text{ falls } t + \arcsin(y) < \frac{\pi}{2} \\ 1 & , \text{ sonst,} \end{cases}$   $t > 0, y \in [0, 1]$ .

Numerisches Experiment:

$T = 1, \lambda = 1$ , Vergleich von Splittingverfahren (konstante Schrittweite) mit hochgenauer numerischer Lösung erhalten durch

```
f=@(t,x) lambda*x*(1-x)+sqrt(1-x^2);
options=odeset('reltol',1.0e-10,...
               'abstol',1.0e-12);
[t,yex]=ode45(f,[0,1],y0,options);
```

R. Hiptmair  
rev 35327,  
25. April  
2011

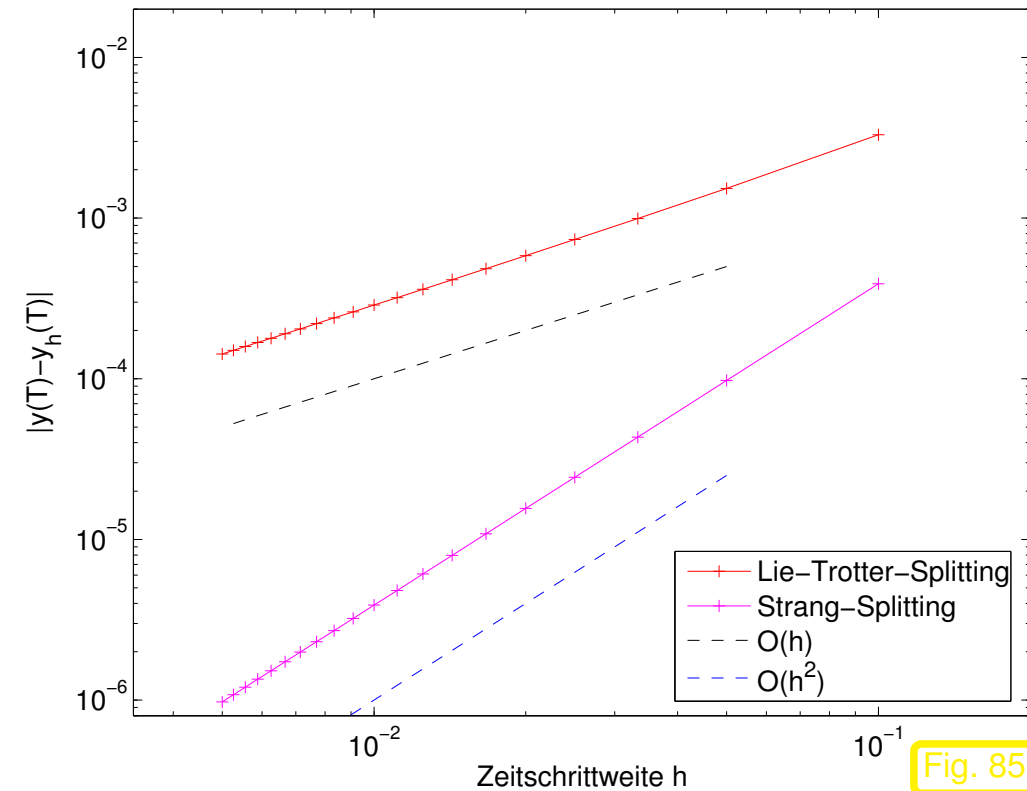


Fig. 85

◁ Fehlerverhalten zum Endzeitpunkt  $T = 1$  ◊

**Theorem 2.5.5** (Konsistenzordnung einfacher Splittingverfahren). Die ESV (2.5.2) und (2.5.3) haben die Konsistenzordnungen ( $\rightarrow$  Def. 2.1.13) 1 bzw. 2.



*Beweis.* Für den Konsistenzfehler ( $\rightarrow$  Def. 2.1.11) haben wir nach Def. 2.1.13 zu zeigen (wir betrachten autonome ODEs!)

$$\|\tau(t, \mathbf{y}, h)\| = \|\Phi^h \mathbf{y} - \Psi^h \mathbf{y}\| = \begin{cases} O(h^2) & \text{für } \Psi \text{ aus (2.5.2) ,} \\ O(h^3) & \text{für } \Psi \text{ aus (2.5.3) .} \end{cases}$$

Der Beweis wird hier für das Strang-Splitting geführt.

Übliche Annahme:  $\mathbf{f}, \mathbf{g}$  hinreichend glatt.

Technik: **Taylorentwicklung**

Taylorentwicklung der exakten Evolution nach  $h$ , vgl. (2.3.25):

$$\begin{aligned} \Phi^h \mathbf{y} &= \mathbf{y} + \dot{\mathbf{y}}(0)h + \frac{1}{2}\ddot{\mathbf{y}}(0)h^2 + O(h^3) \\ &= \mathbf{y} + h(\mathbf{f}(\mathbf{y}) + \mathbf{g}(\mathbf{y})) + \frac{1}{2}h^2(D\mathbf{f}(\mathbf{y}) + D\mathbf{g}(\mathbf{y}))(\mathbf{f}(\mathbf{y}) + \mathbf{g}(\mathbf{y})) + O(h^3) . \end{aligned} \quad (2.5.6)$$

Taylorentwicklung der partiellen Evolutionen  $\Phi_f^h, \Phi_g^h$  nach  $h$ , vgl. (2.3.25)

$$\Phi_f^h \mathbf{y} = \mathbf{y} + h\mathbf{f}(\mathbf{y}) + \frac{1}{2}h^2 D\mathbf{f}(\mathbf{y})\mathbf{f}(\mathbf{y}) + O(h^3) , \quad (2.5.7)$$

$$\Phi_g^h \mathbf{y} = \mathbf{y} + h\mathbf{g}(\mathbf{y}) + \frac{1}{2}h^2 D\mathbf{g}(\mathbf{y})\mathbf{g}(\mathbf{y}) + O(h^3) . \quad (2.5.8)$$

Sukzessives Einsetzen von Taylorentwicklungen, wobei multiplikative Faktoren  $h^k$  ein frühzeitiges Abbrechen der eingesetzten Taylorentwicklung ermöglichen, vgl. die Taylorentwicklung der Runge-

Kutta-Inkrementen in Bsp. 2.3.24, (2.3.27).

$$\begin{aligned}
 \Psi^h \mathbf{y} &= \Phi_f^{h/2}(\Phi_g^h(\Phi_f^{h/2} \mathbf{y})) \\
 &\stackrel{\textcircled{1}}{=} \Phi_f^{h/2}(\Phi_g^h(\mathbf{y} + h/2\mathbf{f}(\mathbf{y}) + \frac{1}{8}h^2 D\mathbf{f}(\mathbf{y})\mathbf{f}(\mathbf{y}) + O(h^3))) \\
 &\stackrel{\textcircled{2}}{=} \Phi_f^{h/2}(\mathbf{y} + h/2\mathbf{f}(\mathbf{y}) + \frac{1}{8}h^2 D\mathbf{f}(\mathbf{y})\mathbf{f}(\mathbf{y}) + O(h^3) + h\mathbf{g}(\mathbf{y} + h/2\mathbf{f}(\mathbf{y}) + O(h^2)) \\
 &\quad + \frac{1}{2}h^2 D\mathbf{g}(\mathbf{y} + O(h))\mathbf{g}(\mathbf{y} + O(h)) + O(h^3)) \\
 &\stackrel{\textcircled{3}}{=} \Phi_f^{h/2}(\mathbf{y} + h/2\mathbf{f}(\mathbf{y}) + h\mathbf{g}(\mathbf{y}) + \frac{1}{8}h^2 D\mathbf{f}(\mathbf{y})\mathbf{f}(\mathbf{y}) + \frac{1}{2}h^2 D\mathbf{g}(\mathbf{y})\mathbf{f}(\mathbf{y}) + \\
 &\quad \frac{1}{2}h^2 D\mathbf{g}(\mathbf{y})\mathbf{g}(\mathbf{y}) + O(h^3)) \\
 &\stackrel{\textcircled{4}}{=} \mathbf{y} + \frac{h}{2}\mathbf{f}(\mathbf{y}) + h\mathbf{g}(\mathbf{y}) + \frac{1}{8}h^2 D\mathbf{f}(\mathbf{y})\mathbf{f}(\mathbf{y}) + \frac{1}{2}h^2 D\mathbf{g}(\mathbf{y})\mathbf{f}(\mathbf{y}) + \frac{1}{2}h^2 D\mathbf{g}(\mathbf{y})\mathbf{g}(\mathbf{y}) + O(h^3) + \\
 &\quad h/2\mathbf{f}(\mathbf{y} + \frac{h}{2}\mathbf{f}(\mathbf{y}) + h\mathbf{g}(\mathbf{y})) + O(h^2)) + \frac{1}{8}h^2 D\mathbf{f}(\mathbf{y} + O(h))\mathbf{f}(\mathbf{y} + O(h)) \\
 &\stackrel{\textcircled{5}}{=} \mathbf{y} + \frac{h}{2}\mathbf{f}(\mathbf{y}) + h\mathbf{g}(\mathbf{y}) + \frac{1}{8}h^2 D\mathbf{f}(\mathbf{y})\mathbf{f}(\mathbf{y}) + \frac{1}{2}h^2 D\mathbf{g}(\mathbf{y})\mathbf{f}(\mathbf{y}) + \frac{1}{2}h^2 D\mathbf{g}(\mathbf{y})\mathbf{g}(\mathbf{y}) \\
 &\quad + \frac{1}{2}h\mathbf{f}(\mathbf{y}) + \frac{1}{4}h^2 D\mathbf{f}(\mathbf{y})\mathbf{f}(\mathbf{y}) + \frac{1}{2}h^2 D\mathbf{f}(\mathbf{y})\mathbf{g}(\mathbf{y}) + \frac{1}{8}h^2 D\mathbf{f}(\mathbf{y})\mathbf{f}(\mathbf{y}) + O(h^3) \\
 &= \mathbf{y} + h(\mathbf{f}(\mathbf{y}) + \mathbf{g}(\mathbf{y})) + \frac{1}{2}h^2(D\mathbf{f}(\mathbf{y})\mathbf{f}(\mathbf{y}) + D\mathbf{g}(\mathbf{y})\mathbf{f}(\mathbf{y}) + D\mathbf{f}(\mathbf{y})\mathbf{g}(\mathbf{y}) + D\mathbf{g}(\mathbf{y})\mathbf{g}(\mathbf{y})) \\
 &\quad + O(h^3) .
 \end{aligned}$$

R. Hiptmair  
rev 35327,  
25. April  
2011

① Wende (2.5.7) auf  $\Phi_f^{h/2} \mathbf{y}$  an.

- ② Benutze (2.5.8) mit  $\mathbf{y} \leftarrow \mathbf{y} + h/2\mathbf{f}(\mathbf{y}) + \frac{1}{8}h^2 D\mathbf{f}(\mathbf{y})\mathbf{f}(\mathbf{y}) + O(h^3)$ , um  $(\Phi_g^h(\mathbf{y} + h/2\mathbf{f}(\mathbf{y}) + \frac{1}{8}h^2 D\mathbf{f}(\mathbf{y})\mathbf{f}(\mathbf{y}) + O(h^3)))$  zu entwickeln. Vernachlässige Terme  $O(h^3)$ .
- ③ Taylorentwicklung (in  $h$ ) von  $\mathbf{g}$  und  $D\mathbf{g}$  um  $\mathbf{y}$ .
- ④ Benutze (2.5.7) mit  $\mathbf{y} \leftarrow \mathbf{y} + h/2\mathbf{f}(\mathbf{y}) + h\mathbf{g}(\mathbf{y}) + \frac{1}{8}h^2 D\mathbf{f}(\mathbf{y})\mathbf{f}(\mathbf{y}) + \frac{1}{2}h^2 D\mathbf{g}(\mathbf{y})\mathbf{f}(\mathbf{y}) + \frac{1}{2}h^2 D\mathbf{g}(\mathbf{y})\mathbf{g}(\mathbf{y}) + O(h^3)$  und vernachlässige Terme in  $O(h^3)$ .
- ⑤ Taylorentwicklung (in  $h$ ) von  $\mathbf{f}$  und  $D\mathbf{f}$  um  $\mathbf{y}$ .

Vergleich mit (2.5.6) liefert die Behauptung für das Strang-Splitting. □

*Bemerkung 2.5.9* ( reversible Strang-Splitting-Einschrittverfahren).

R. Hiptmair  
rev 35327,  
25. April  
2011



*Beispiel 2.5.10* (Splittingverfahren für mechanische Systeme).

Newton'sche Bewegungsgleichung  $\ddot{\mathbf{r}} = a(\mathbf{r}) \stackrel{(1.1.10)}{\iff} \dot{\mathbf{y}} := \begin{pmatrix} \dot{\mathbf{r}} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{v} \\ a(\mathbf{r}) \end{pmatrix} =: \mathbf{F}(\mathbf{y}) .$

Splitting: 
$$F(\mathbf{y}) = \underbrace{\begin{pmatrix} 0 \\ a(\mathbf{r}) \end{pmatrix}}_{=:f(\mathbf{y})} + \underbrace{\begin{pmatrix} \mathbf{v} \\ 0 \end{pmatrix}}_{=:g(\mathbf{y})} .$$

► 
$$\Phi_f^t \begin{pmatrix} \mathbf{r}_0 \\ \mathbf{v}_0 \end{pmatrix} = \begin{pmatrix} \mathbf{r}_0 \\ \mathbf{v}_0 + ta(\mathbf{r}_0) \end{pmatrix} , \quad \Phi_g^t \begin{pmatrix} \mathbf{r}_0 \\ \mathbf{v}_0 \end{pmatrix} = \begin{pmatrix} \mathbf{r}_0 + t\mathbf{v}_0 \\ \mathbf{v}_0 \end{pmatrix} .$$

► Lie-Trotter-Splitting 2.5.2

Symplektisches Eulerverfahren

$$\Psi^h \begin{pmatrix} \mathbf{r} \\ \mathbf{v} \end{pmatrix} = \left( \Phi_g^h \circ \Phi_f^h \right) \begin{pmatrix} \mathbf{r} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{r} + h(\mathbf{v} + ha(\mathbf{r})) \\ \mathbf{v} + ha(\mathbf{r}) \end{pmatrix} . \quad (2.5.11)$$

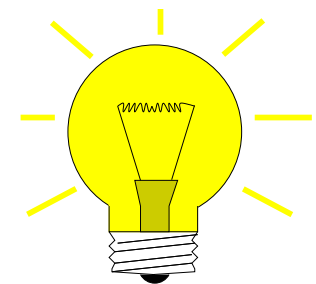
► Strang-Splitting 2.5.3

$$\Psi^h \begin{pmatrix} \mathbf{r} \\ \mathbf{v} \end{pmatrix} = \left( \Phi_g^{h/2} \circ \Phi_f^h \circ \Phi_g^{h/2} \right) \begin{pmatrix} \mathbf{r} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{r} + h\mathbf{v} + \frac{1}{2}h^2a(\mathbf{r} + \frac{1}{2}h\mathbf{v}) \\ \mathbf{v} + ha(\mathbf{r} + \frac{1}{2}h\mathbf{v}) \end{pmatrix} . \quad (2.5.12)$$

= Einschrittformulierung des Störmer-Verlet-Verfahrens (1.4.27), siehe Bem. 1.4.33 !

$$(2.5.12) \quad \longleftrightarrow \quad \begin{aligned} \mathbf{r}_{k+\frac{1}{2}} &= \mathbf{r}_k + \frac{1}{2}h\mathbf{v}_k , \\ \mathbf{v}_{k+1} &= \mathbf{v}_k + ha(\mathbf{r}_{k+\frac{1}{2}}) , \\ \mathbf{r}_{k+1} &= \mathbf{r}_{k+\frac{1}{2}} + \frac{1}{2}h\mathbf{v}_{k+1} . \end{aligned} \quad (2.5.13)$$

Einwand: Splittingverfahren sind nur in Spezialfällen zu gebrauchen, denn die exakten Evolutionen  $\Phi_f$  und  $\Phi_g$  werden oft nicht analytisch auswertbar sein.

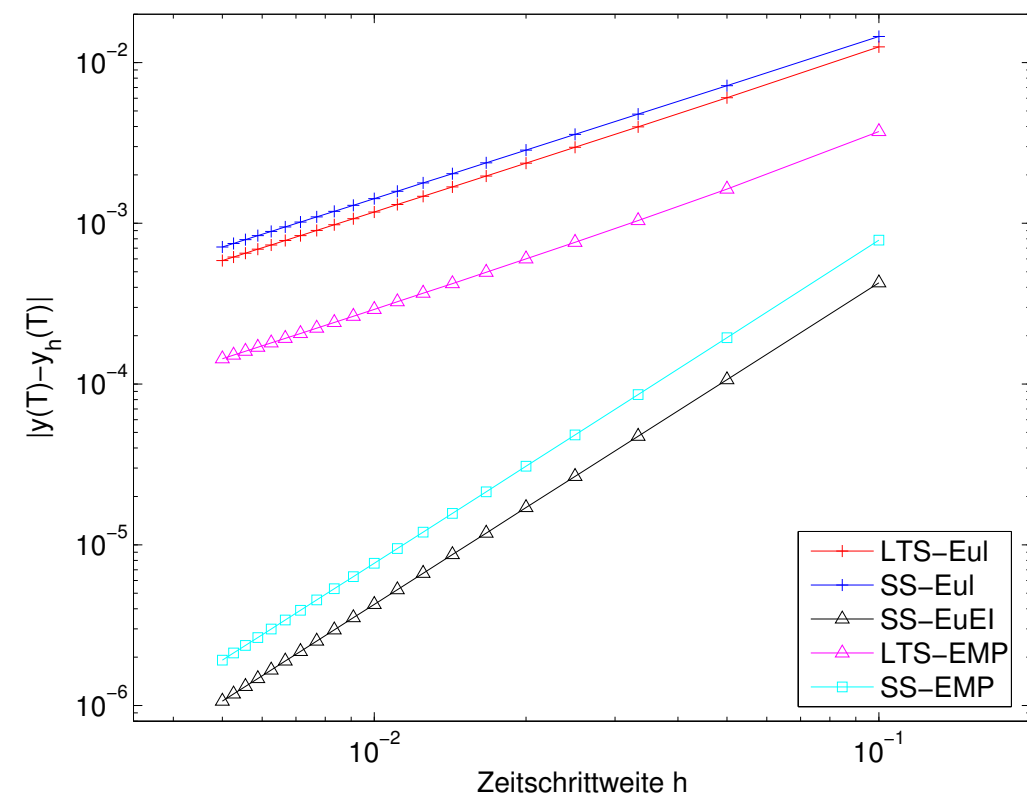


Idee: Ersetze

Exakte Evolutionen  $\longrightarrow$  diskrete Evolutionen  
 $\Phi_g^h, \Phi_f^h \longrightarrow \Psi_g^h, \Psi_f^h$

*Beispiel* 2.5.14 (Inexakte Splittingverfahren). Forsetzung Bsp. 2.5.4

AWP von Bsp. 2.5.4, Inexakte Splittingverfahren auf der Grundlage verschiedener inexakter Basisverfahren:



- LTS-Eul Explizites Eulerverfahren (2.2.1) →  $\Psi_{h,g}^h, \Psi_{h,f}^h$  + Lie-Trotter-Splitting (2.5.2)
- SS-Eul Explizites Eulerverfahren (2.2.1) →  $\Psi_{h,g}^h, \Psi_{h,f}^h$  + Strang-Splitting (2.5.3)
- SS-EuEI Strang-Splitting (2.5.3): Explizites Eulerverfahren (2.2.1) ○ exakte Evolution  $\Phi_g^h$   
○ implizites Eulerverfahren (2.2.1)
- LTS-EMP Explizite Mittelpunktsregel (2.3.4) →  $\Psi_{h,g}^h, \Psi_{h,f}^h$  + Lie-Trotter-Splitting (2.5.2)
- SS-EMP Explizite Mittelpunktsregel (2.3.4) →  $\Psi_{h,g}^h, \Psi_{h,f}^h$  + Strang-Splitting (2.5.3)



Ordnung der Splittingverfahren wird durch Konsistenzordnung von  $\Phi_f^h, \Phi_g^h$  begrenzt.

Ausnahme: SS-EuEI: **reversibles** Verfahren ➤ Konsistenzordnung  $\geq 2$  nach Thm. 2.1.29



## 2.6 Schrittweitensteuerung [8, Kap. 5], [19, Sect. 2.8]

Beispiel 2.6.1 (Numerische Integration bei Blow-up).

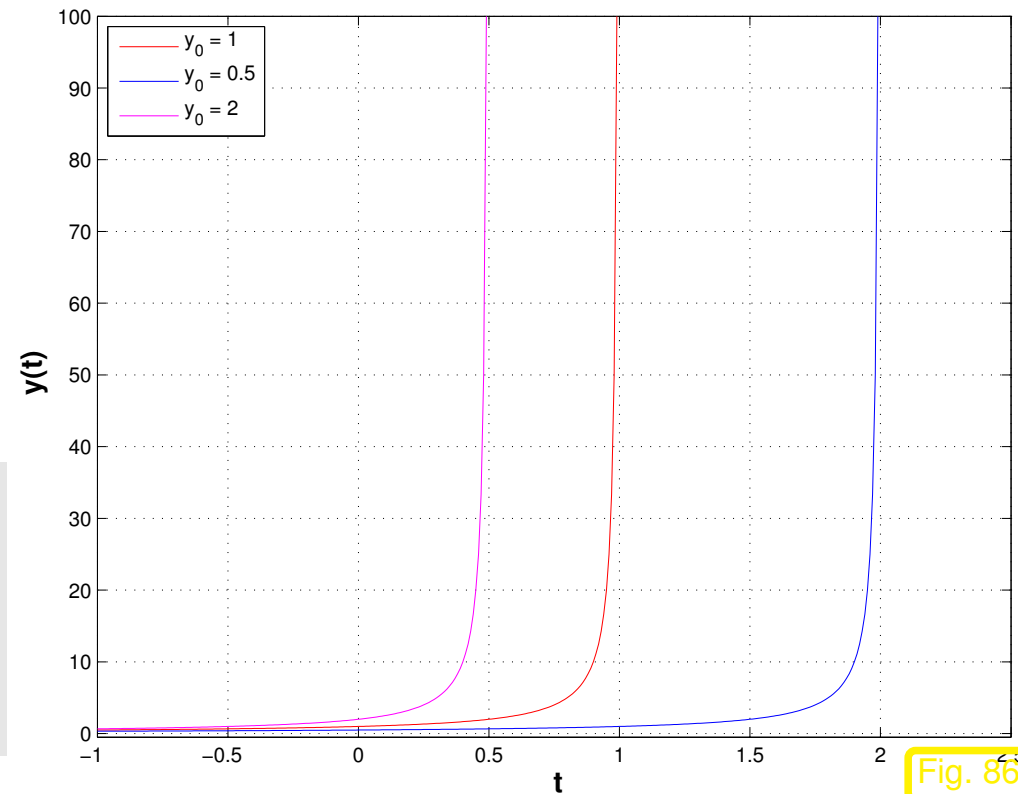
Skalares autonomes AWP  $\rightarrow$  Bsp. 1.3.11

$$\dot{y} = y^2, \quad y(0) = y_0 > 0.$$

$$\blacktriangleright \quad y(t) = \frac{y_0}{1 - y_0 t}, \quad t < 1/y_0.$$

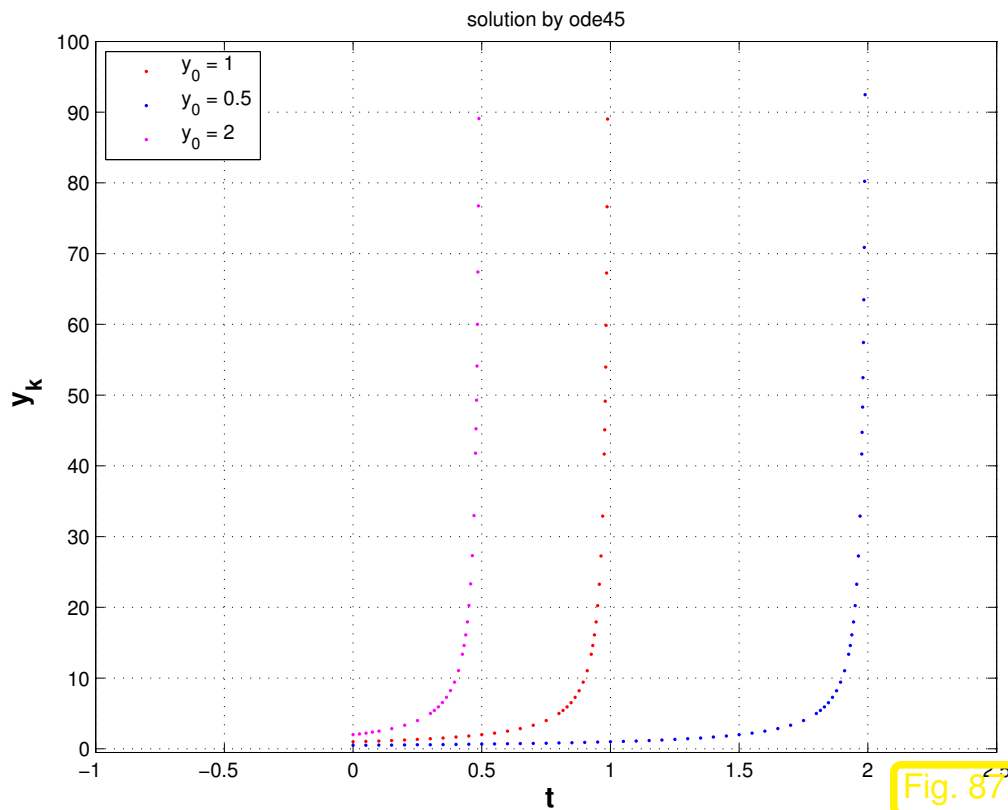
Die Lösung existiert nur für endliche Zeit und erleidet dann einen **Blow-up**, siehe Def. 1.3.1:

$$\lim_{t \rightarrow 1/y_0} y(t) = \infty : J(y_0) = ] - \infty, 1/y_0 ]!$$



Herausforderung: Wie sollte das Zeitgitter  $\{t_0 < t_1 < \dots < t_{N-1} < t_N\}$  für ein ESV gewählt werden, wenn  $J(y_0)$  nicht a priori bekannt ist und nicht klar ist, ob sich ein Blow-up ereignen wird?

Gedankenexperiment: wie wird sich wohl ein Runge-Kutta-Einschrittverfahren ( $\rightarrow$  Def. 2.3.5) bei Verwendung *uniformer* (*äquidistanter*) Zeitschritte verhalten, wenn es auf das obige AWP angewendet wird.



```

1 fun = @(t,y) y.^2;
2 [t1,y1] = ode45(fun,[0 2],1);
3 [t2,y2] = ode45(fun,[0 2],0.5);
4 [t3,y3] = ode45(fun,[0 2],2);

```

R. Hiptmair  
rev 35327,  
25. April  
2011

## MATLAB Warnungsmeldungen:

```

Warning: Failure at t=9.999694e-01. Unable to meet integration
tolerances without reducing the step size below the smallest
value allowed (1.776357e-15) at time t.

```

```
> In ode45 at 371
```



In simpleblowup at 22

Warning: Failure at t=1.999970e+00. Unable to meet integration tolerances without reducing the step size below the smallest value allowed (3.552714e-15) at time t.

> In ode45 at 371

In simpleblowup at 23

Warning: Failure at t=4.999660e-01. Unable to meet integration tolerances without reducing the step size below the smallest value allowed (8.881784e-16) at time t.

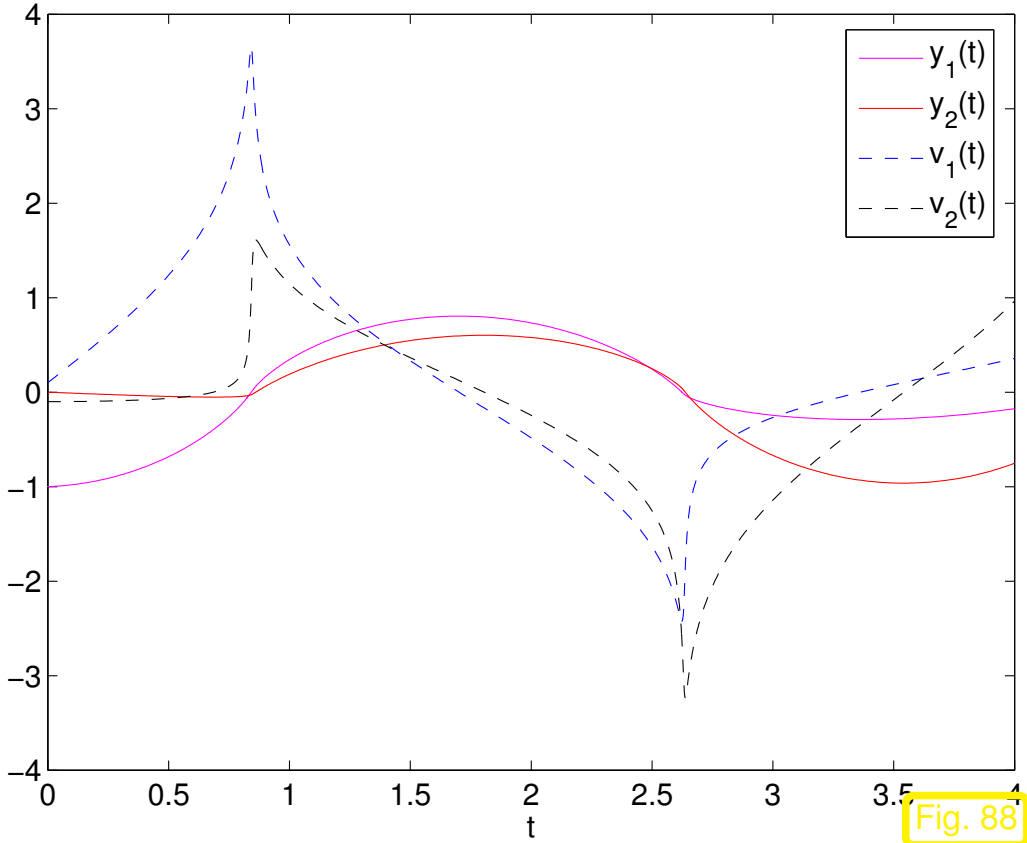
> In ode45 at 371

In simpleblowup at 24

We stellen fest, dass es `ode45` gelingt, die Schrittweite immer weiter zu reduzieren, wenn es sich dem Pol der Lösung nähert.

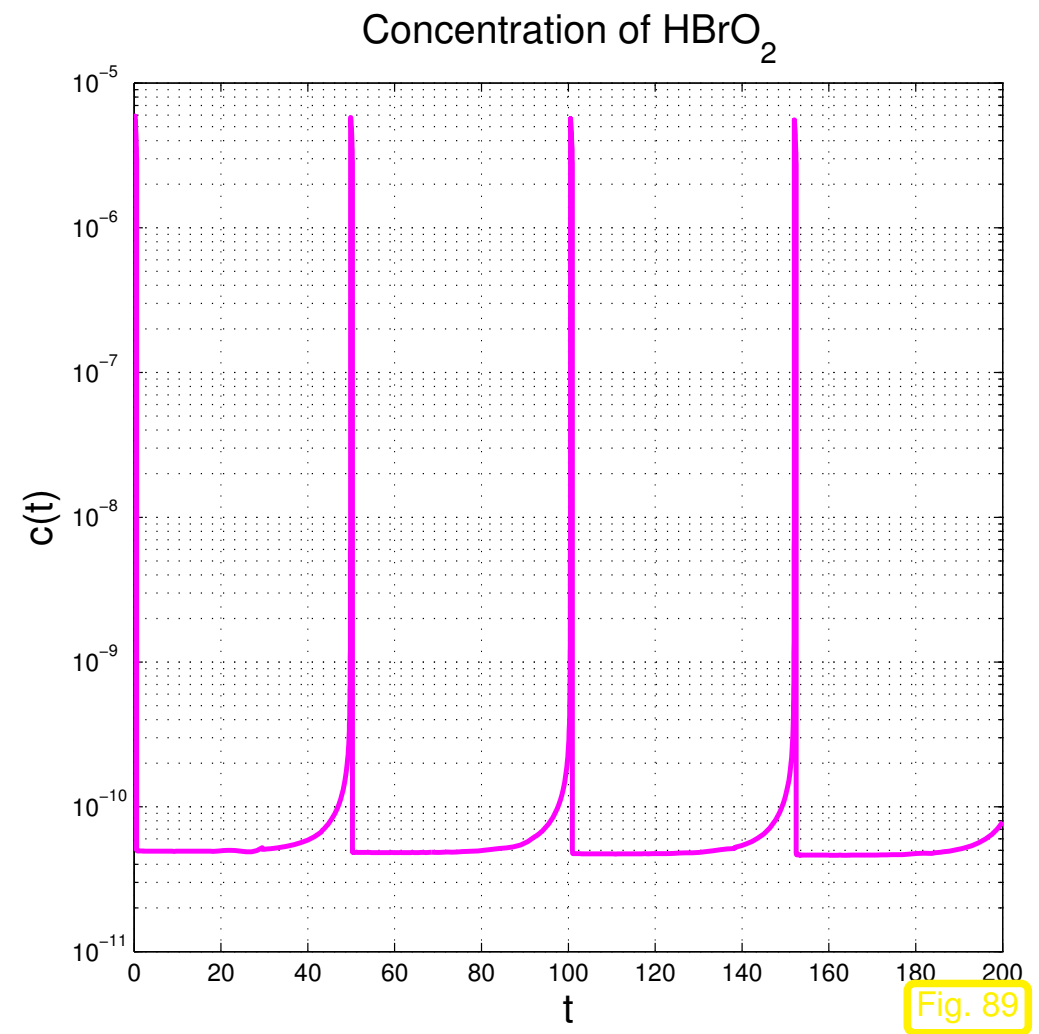


*Bemerkung 2.6.2* (Zeitlich ungleichmässiges Verhalten von Lösungen).



Keplerproblem von Bsp. 2.4.19

Fig. 88



Oregonator-Reaktion von Bsp. 1.2.12

Fig. 89

Häufig: Lösungen von AWP's zeigen stark ungleichmässiges Verhalten in der Zeit.

Eine Möglichkeit, Einschrittverfahren an das zeitlokale Verhalten der Lösung anzupassen haben wir bereits kennengelernt:

→ **Ordnungssteuerung** bei Extrapolationsverfahren, Sect. 2.4.5

Doch im Fall eines Blow-up nützt uns das gar nichts!



Grundlegende Fragestellung

Wie wählt man ein *geeignetes* Zeitgitter  $\mathcal{G} = \{t_0 < t_1 < \dots < t_N = T\}$   
für ein gegebenes Einschrittverfahren und Anfangswertproblem?

Was heisst *geeignet*?

Effizienz

Genauigkeit

Ziel:  $N$  so klein wie möglich &  $\max_{k=1,\dots,N} \|\mathbf{y}(t_k) - \mathbf{y}_k\| < \text{TOL}$ ,  $\text{TOL} = \text{Toleranz}$   
 oder  $\|\mathbf{y}(T) - \mathbf{y}_N\| < \text{TOL}$

Strategie: Kontrolliere **Einschrittfehler** durch

- Anpassung der *aktuellen* Schrittweite  $h_k$ ,
- Vorhersage der *nächsten* Schrittweite  $h_{k+1}$

} **Zeitlokale  
Schrittweitensteuerung**

Vorgehen: **Zeitlokale** Schätzung des Einschrittfehlers (Konsistenzfehlers, Def. 2.1.11)  
 (**a posteriori**, da  $\mathbf{y}_k, h_{k-1}$  verwendet wird)

Warum entscheidet man sich für *zeitlokale* Schrittweitensteuerung, die sich nur auf Schätzung des Einschrittfehlers stützt?

Einwand: Wenn ein kleiner Fehler in einem Zeitschritt zu einem späteren Zeitpunkt zu grösseren Fehlern  $\|\mathbf{y}_k - \mathbf{y}(t_k)\|$  führt, dann kann zeitlokale Schrittweitensteuerung nichts dagegen ausrichten!

➤ Bsp. 2.6.9

Trotzdem scheint zeitlokale Schrittweitensteuerung das einzige praktikable Verfahren zu sein,

- ☞ weil man nicht während der Rechnung viele Zeitschritte zurückgehen will, was viel Rechenzeit kosten kann,
- ☞ weil sie einfach zu implementieren ist und mit wenig zusätzlichem Rechenaufwand auskommt,
- ☞ weil man prinzipiell keine Methode finden wird, die eine garantierte Genauigkeit liefert.

Idee:

*Schätzung des Konsistenzfehlers*

Vergleich zweier diskreter Evolutions  $\Psi^{t,t+h}$ ,  $\tilde{\Psi}^{t,t+h}$  **verschiedener Ordnung**

( $\rightarrow$  Def. 2.1.13) für eine *aktuelle Zeitschrittweite*  $h$ :

Falls Ordnung( $\tilde{\Psi}$ ) > Ordnung( $\Psi$ )

$$\Rightarrow \underbrace{\Phi^{t,t+h} \mathbf{y}(t_k) - \Psi^{t,t+h} \mathbf{y}(t_k)}_{\text{Konsistenzfehler}} \approx \text{EST}_k := \tilde{\Psi}^{t,t+h} \mathbf{y}(t_k) - \Psi^{t,t+h} \mathbf{y}(t_k). \quad (2.6.3)$$

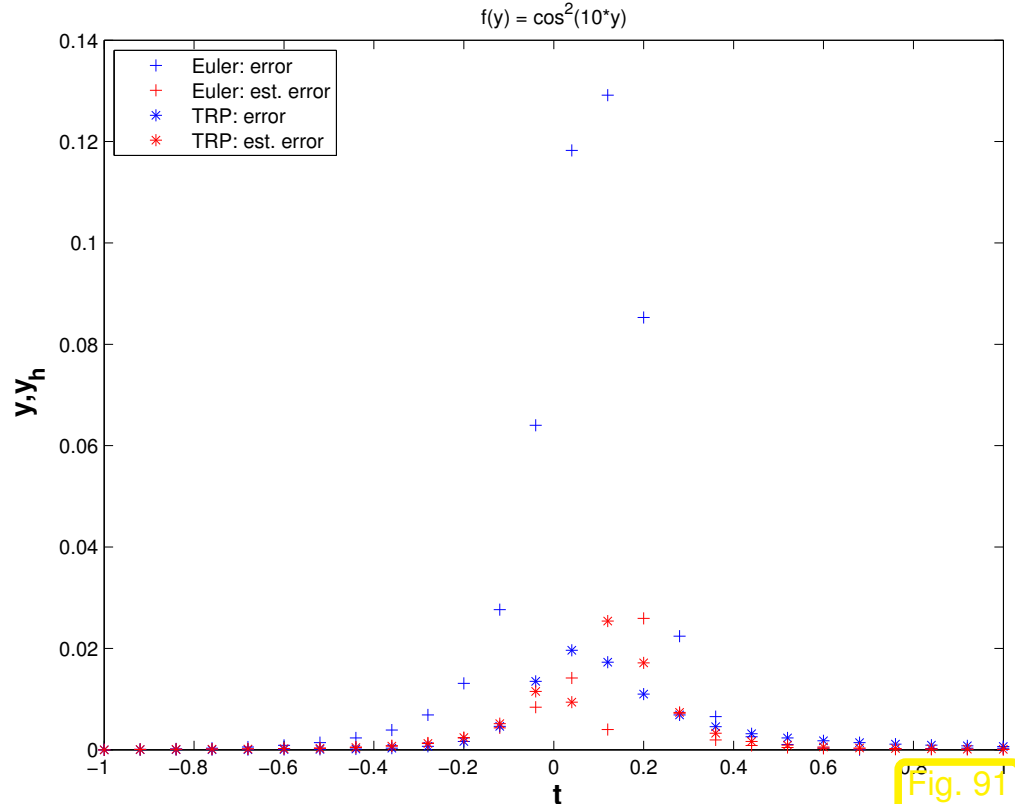
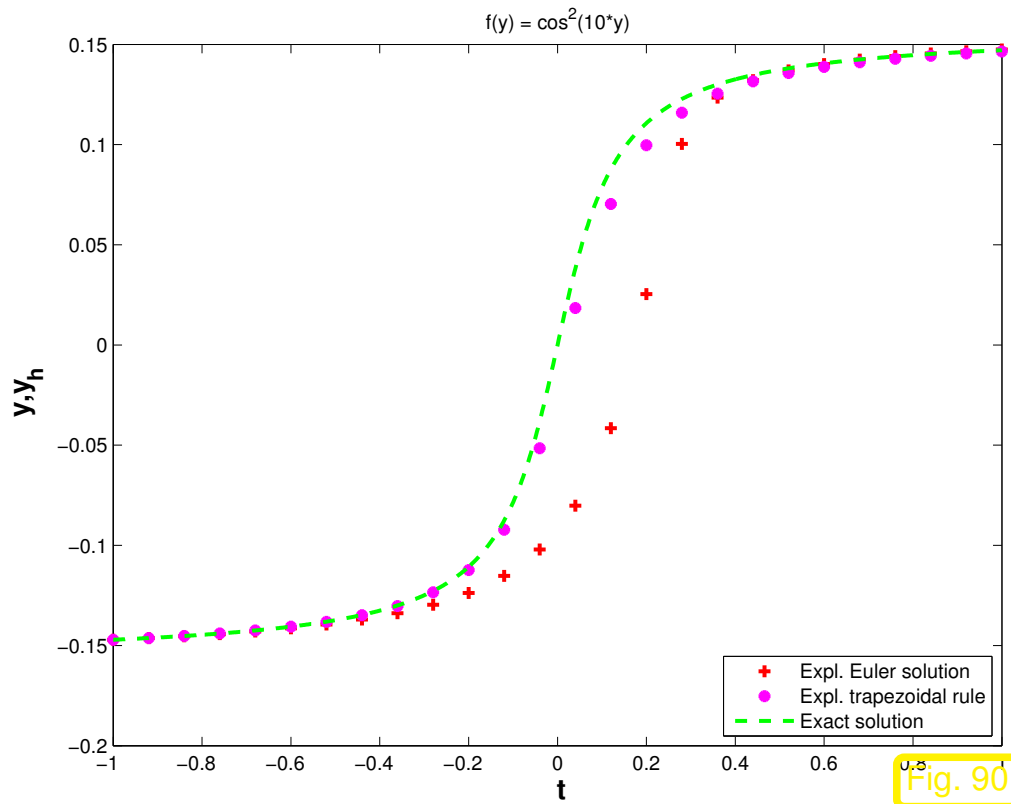
**Heuristik** für konkretes  $h$

*Beispiel 2.6.4* (Qualität der Fehlerschätzung).

- Skalares AWP:  $\dot{y} = \cos^2(ay)$ , Lösung  $y(t) = 1/a \arctan(at)$  auf  $[-1, 1]$ ,  $a = 10$

$\Psi \leftrightarrow$  Explizites Euler-Verfahren (1.4.2), Ordnung  $p = 1$

$\tilde{\Psi} \leftrightarrow$  Explizite Trapezregel (2.3.3), Ordnung  $p = 2$



Beobachtung:  $\left[ \begin{array}{l} \text{☞ Grosse Unterschiede zwischen geschätztem und wahrem Fehler möglich} \\ \text{☞ Jedoch: Fehlerschätzung für } \tilde{\Psi} \text{ durch } \Psi \text{ sinnvoll, da "richtig in der Tendenz"} \end{array} \right.$

Vergleich

$$EST_k \leftrightarrow TOL$$

$$EST_k \leftrightarrow TOL \|y_k\|$$

 $\geq$ 

Verwerfen/Akzeptieren des aktuellen Schritts

Absolute Toleranz

Relative Toleranz

Führt zu einem sehr einfachen Algorithmus:

$EST_k < TOL$ : Ausführen des aktuellen Schritts (Schrittweite  $h$ )

Nächster Schritt mit Schrittweite  $\alpha h$ , mit einem  $\alpha > 1$  (\*)

$EST_k > TOL$ : Wiederholung des aktuellen Schritts mit Schrittweite  $< h$ , z.B.  $\frac{1}{2}h$

Begründung für (\*): Wenn die aktuelle Schrittweite bereits einen hinreichend kleinen Einschrittfehler sicherstellt, dann ist es unter Umständen möglich, auch mit einer etwas grösseren Schrittweite einen Einschrittfehler zu erhalten, der immer noch genügend klein ist. Dadurch kann die Gesamtzahl der Zeitschritte reduziert werden, was die Effizienz des Verfahrens erhöht. Das Risiko eines Genauigkeitsverlusts wird durch die Fehlerschätzung im nächsten Schritt in Grenzen gehalten.

R. Hiptmair  
rev 35327,  
25. April  
2011

Listing 2.4: Einfache zeitlokale Schrittweitensteuerung für Einschrittverfahren (autonome ODE)

```

1 function [t, y] =
   odeintadapt(Psilow, Psihigh, T, y0, h0, reltol, abstol, hmin)
2 t = 0; y = y0; h = h0; %

```

2.6

p. 295

```
3 while ((t(end) < T) (h > hmin)) %  
4   yh = Psihigh(h, y0); % ESV hoher Ordnung  
5   yH = Psilow(h, y0); % ESV niedriger Ordnung  
6   est = norm(yH-yh); %  
7  
8   if (est < max(reltol*norm(y0), abstol)) %  
9     y0 = yh; y = [y, y0]; h = 1.1*h; % Schritt akzeptiert  
10    h = min(h, T-t(end)); t=[t, t+h]; %  
11    else, h = h/2; end % Schritt verworfen  
12 end
```

## Kommentare zu Code 2.4:

- Argumente von `odeintadapt`:

- `Psilow`, `Psihigh`: Funktionshandles auf diskrete Evolutionen (für autonome ODE) unterschiedlicher Ordnungen, Typ  $@(y, h)$ , Zustandsvektor als erstes Argument, Schrittweite als zweites,
- `T`: Endzeitpunkt  $T > 0$ ,



- $y_0$ : Anfangszustand  $y_0$ ,
  - $h_0$ : Schrittweite  $h_0$  für den ersten Zeitschritt
  - `reltol`, `abstol`: Relative and absolute Toleranzen, siehe oben,
  - `hmin`: minimale Zeitschrittweite, Verfahren bricht ab, falls  $h_k < h_{\min}$ , was für das Erkennen von Blow-ups und Kollaps wichtig ist.
- Zeile 3: Überprüfe, ob der Endzeitpunkt erreicht ist oder das Verfahren steckengeblieben ist ( $h_k < h_{\min}$ ).
  - Zeile 4, 5: Propagiere den aktuellen Zustand mit Hilfe beider Einschrittverfahren.
  - Zeile 6: Berechne die Norm des geschätzten Fehlers, siehe (2.6.3).
  - Zeile 8: Vergleich, um zu entscheiden, ob der aktuelle Schritt akzeptiert oder verworfen werden sollte.
  - Zeile 9, 10: **Schritt akzeptiert**: Aktualisiere den Zustand und schlage 1.1 mal die aktuelle Schrittweite für den nächsten Schritt vor.
  - Zeile 11 **Schritt verworfen**: Versuche es nochmal mit der halben Zeitschrittweite.
  - Rückgabewerte
    - $\tau$ : Zeitgitter  $t_0 < t_1 < t_2 < \dots < t_N < T$ , wobei  $t_N < T$  auf vorzeitigem Abbruch hinweist (Kollaps, Blow-up),
    - $y$ : Folge von Zuständen  $(y_k)_{k=0}^N$ .

! Gemäss unserer Heuristik, siehe (2.6.3), scheint es, dass  $EST_k$  den Einschrittfehler des Einschrittverfahrens niedrigerer Ordnung  $\Psi$  misst, und dass wir  $\mathbf{y}_{k+1} = \Psi^{h_k} \mathbf{y}_k$ , setzen sollten, wenn der Zeitschritt akzeptiert wird.

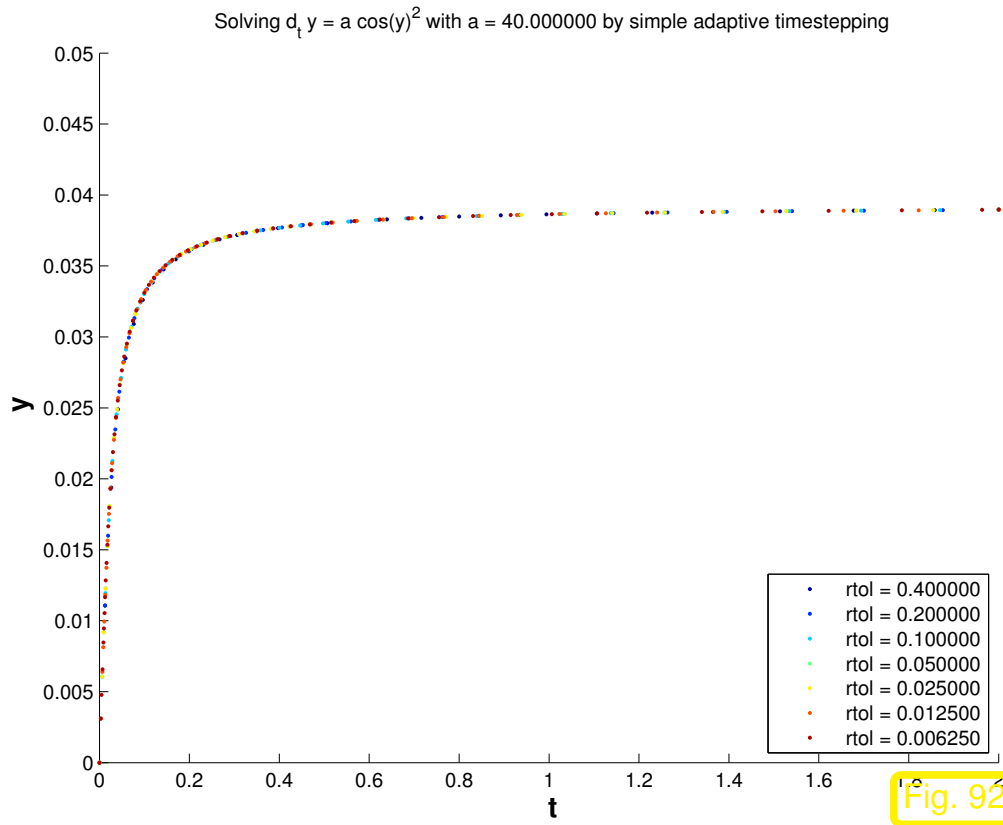
Jedoch wäre es ungeschickt, nicht den vermutlich besseren Wert  $\mathbf{y}_{k+1} = \tilde{\Psi}^{h_k} \mathbf{y}_k$  zu nehmen, zumal er ohne Zusatzaufwand verfügbar ist. Jede Implementierung zeitlokaler Schrittweitensteuerung folgt dieser Idee, also auch Code 2.4, und dieses Vorgehen kann durch steuerungstheoretische Argumente begründet werden [8, Sect. 5.2], siehe auch die folgende Bem. 2.6.7.

*Beispiel 2.6.5* (Effizienzgewinn durch Adaptivität). → Ex. 2.6.4

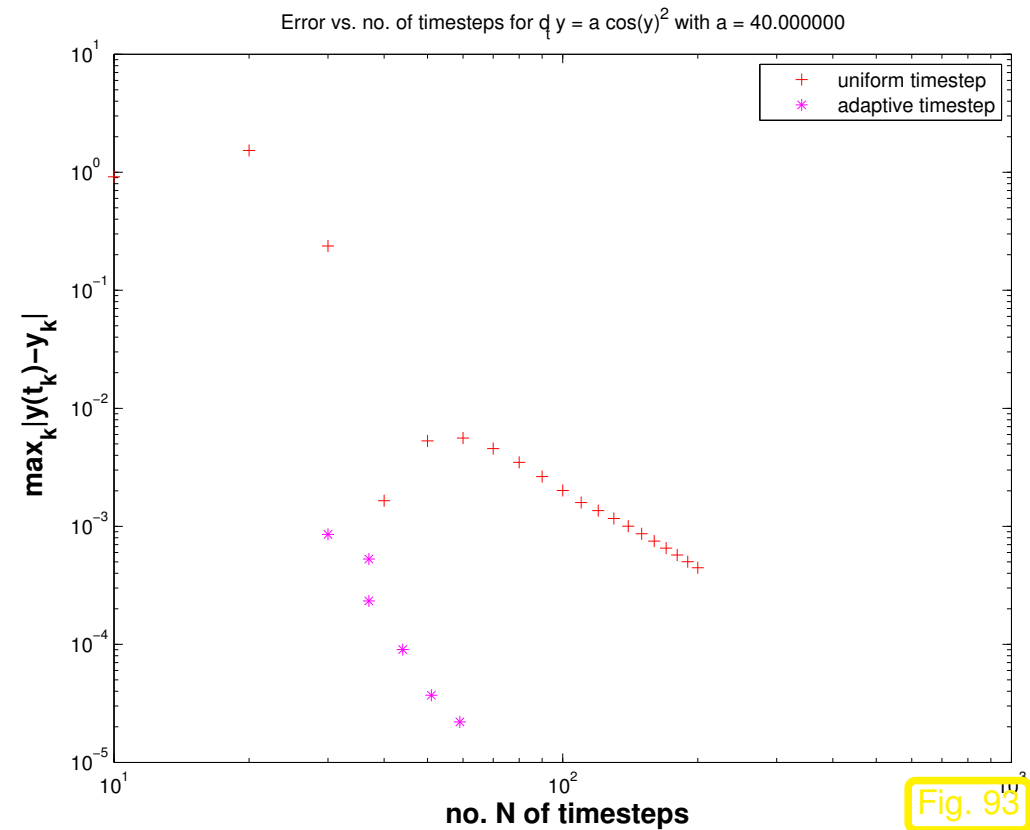
- AWP für ODE  $\dot{y} = \cos(\alpha y)^2$ ,  $\alpha > 0$ ,
- Analytische Lösung  $y(t) = \arctan(\alpha(t - c))/\alpha$  für  $y(0) \in ] -\pi/2, \pi/2[$
- Integrationsintervall  $[0, 2]$ , Anfangswert  $y(0) = 0$

# Einfache adaptive Strategie aus Code 2.4 mit der lokalen Fehlerschätzung aus Bsp. 2.6.4.

Nun untersuchen wir die Abhängigkeit des Diskretisierungsfehlers vom Rechenaufwand, der proportional zu der Anzahl der Zeitschritte ist.



Lösungen  $(y_k)_k$  für verschiedene `rtol`



Fehler als Funktion des Rechenaufwandes

- ☞ Adaptive Zeitschrittweitensteuerung erzielt bei vergleichbarem Rechenaufwand wesentlich bessere Genauigkeit als das gleiche Einschrittverfahren mit uniformer Zeitschrittweite.



Nachteil von Code 2.4: Pauschale Vergrößerung/Verringerung der Schrittweite in Zeilen 9, 11 “verschwendet” Information enthalten in  $EST_k : TOL$ .

Wir wollen mehr !

Wenn  $EST_k > TOL$  : Schrittweitenkorrektur  $t_{k+1} = ?$

Wenn  $EST_k < TOL$  : Schrittweitevorschlag  $t_{k+2} = ?$

**Schrittweitenkorrektur** bezieht sich auf verworfenen zu wiederholenden *aktuellen* Schritt  
**Schrittweitevorschlag** wird benutzt für den *nächsten* Schritt  
 (vgl. Code 2.4)

Falls Ordnung( $\Psi$ ) =  $p$ , Ordnung( $\tilde{\Psi}$ ) >  $p$ ,  $p \in \mathbb{N}$ ,

$$\Psi^{t,t+h} \mathbf{y}(t_k) - \Phi^{t,t+h} \mathbf{y}(t_k) = ch^{p+1} + O(h^{p+2}),$$

$$\tilde{\Psi}^{t,t+h} \mathbf{y}(t_k) - \Phi^{t,t+h} \mathbf{y}(t_k) = O(h^{p+2})$$

$$h \ll 1 \Rightarrow$$

Ziel: Effizienz  
 $EST_k \approx ch^{p+1} \stackrel{!}{=} TOL.$

Heuristik!



„Optimale Schrittweite“:  
 (Schrittweitevorschlag)

$$h^* = h^{p+1} \sqrt{\frac{TOL}{EST_k}}$$

(2.6.6)

Korrigierte Schrittweite  
 Schrittweitevorschlag

*Bemerkung 2.6.7* (Steuerung von  $\tilde{\Psi}$  durch  $\Psi$ ). Code 2.4

Bisherige Überlegung: „Schätzung des Fehlers“ von  $\tilde{\Psi}$  ➤ Steuerung von  $\tilde{\Psi}$

Effizient ? Genaueres (teureres) Verfahren  $\tilde{\Psi}$  wird nur zur Steuerung des ungenaueren Verfahrens verwendet.

Mit gleichem Aufwand: Integration des AWP mit  $\tilde{\Psi}$  gesteuert durch  $\Psi$  !

So wird es in der Praxis auch gemacht !

Erinnerung an Bsp. 2.6.4: Euler-Verfahren (Ordnung  $p = 1$ ) lieferte gute Fehlerschätzung für explizite Trapezregel (Ordnung  $p = 2$ )

Noch eine Heuristik:

$EST_k > TOL$  ist ein Hinweis darauf, dass eines der beiden Verfahren  $\Psi$ ,  $\tilde{\Psi}$  Probleme mit der (lokalen) Approximation der Lösung hat. Eine Verringerung von  $h_k$  ist daher angezeigt.

Mathematische Rechtfertigung: Steuerungstheorie  $\rightarrow$  [8, Sect. 5.2]



- MATLAB-Implementierung:
- $\Psi, \tilde{\Psi} \hat{=}$  diskrete Evolutionen, Konsistenzordnung  $p/p + 1$
  - $t_0 \hat{=}$  Anfangszeitpunkt,  $T \hat{=}$  Endzeitpunkt
  - $y_0 \hat{=}$  Anfangswert (Spaltenvektor)
  - $reltol, abstol \hat{=}$  absolute/relative Toleranzen
  - $h_0, h_{min} \hat{=}$  Schrittweite für 1. Schritt/minimale Schrittweite

ESV mit Schrittweitensteuerung

```
function [t,y] = ssctrl(Ψ,Ψ̃,t0,T,y0,h0,reltol,abstol,hmin)
t = t0; y = y0; h = h0;
while ((t(end) < T) && (h > hmin))
    yh = Ψ̃(t(end),y(:,end),h);
    yH = Ψ(t(end),y(:,end),h);
    est = norm(yH-yh);
    tol = max(reltol*norm(y(:,end)),abstol);
    h = h*max(0.5,min(2,(tol/est)^(1/(p+1)))));
    if (est < tol)
        y = [y,yh]; h = min(h,T-t(end)); t = [t,t(end)+h];
    end
end
end
```

*Beispiel 2.6.8* (Schrittweitensteuerung für explizite Trapezregel/Euler-Verfahren).

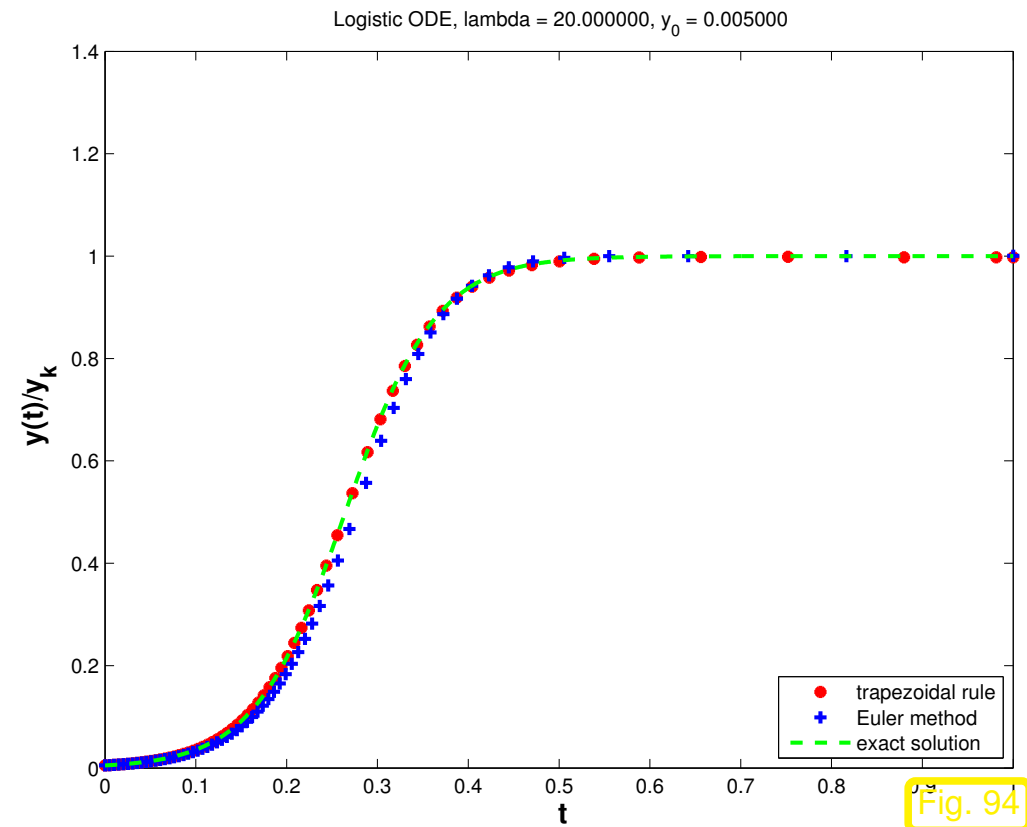
## Anfangswertproblem für skalare logistische Dgl, siehe Bsp. 1.2.1

$$\dot{y} = \lambda y(1 - y), \quad \lambda = 20 \quad \Rightarrow \quad y(t) = \frac{y_0}{y_0 + (1 - y_0) \exp(-\lambda t)}.$$

Einschrittverfahren aus Bsp. 2.6.4, Schrittweitanpassung gemäss (2.6.6)

- ❶ Integration mit explizitem Euler-Verfahren (1.4.2), Fehlerschätzung (2.6.3) mit expliziter Trapezregel (2.3.3)
- ❷ Integration mit expliziter Trapezregel (2.3.3), Schrittweitensteuerung mit explizitem Euler-Verfahren gemäss Bem. 2.6.7

Absolute/relative Toleranz = 0.005,  $y_0 = 0.1/\lambda$



Trapezregel/Euler: 63/62 Schritte, 12 verworfen



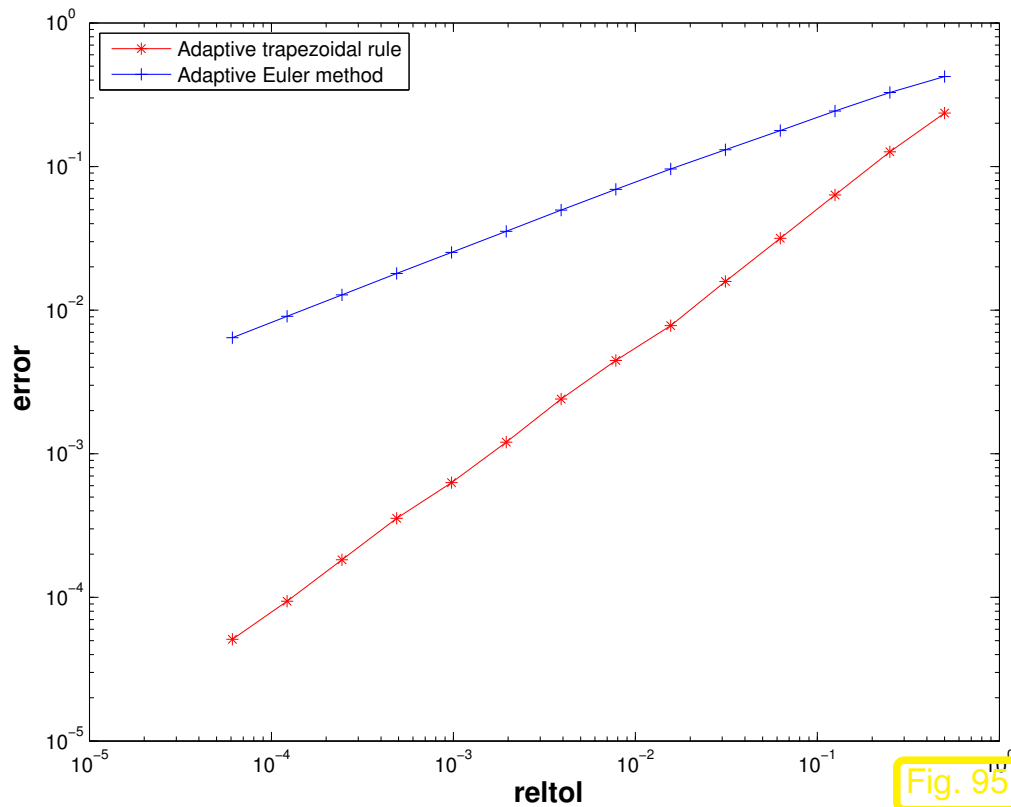


Fig. 95

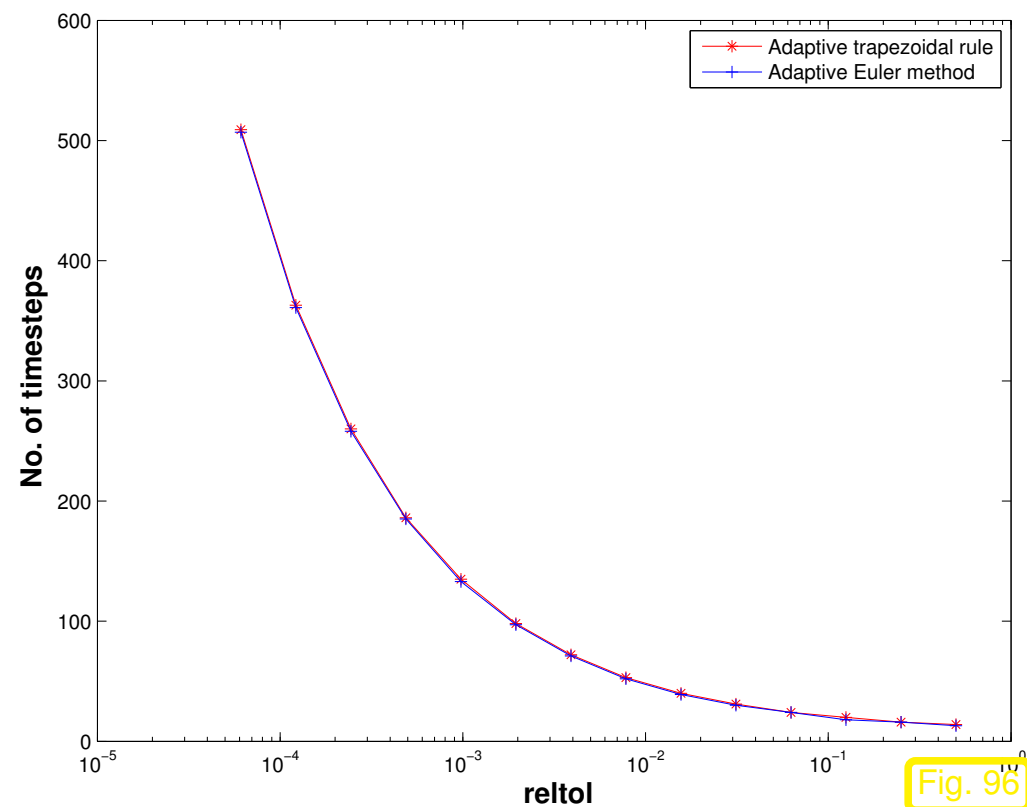


Fig. 96

In diesem Beispiel liefert die Fortführung der Rechnung mit der Näherung aus dem Verfahren höherer Ordnung  $\tilde{\Psi}$  ganz klar die bessere Genauigkeit.

Beobachtung: Fehler  $\max_j |y(t_j) - y_j|$  ist in diesem Beispiel gut mit Toleranz  $TOL$  korreliert.

Beispiel 2.6.9 (“Versagen” adaptive Zeitschrittsteuerung). → Ex. 2.6.5

Gleiche ODE und einfache adaptive Schrittweitensteuerung wie in Bsp. 2.6.5. Ebenfalls gleiche Auswertungen.

Nun: Anfangswert  $y(0) = -0.0386$ , vgl. Bsp. 2.6.4.

Solving  $d_t y = a \cos(y)^2$  with  $a = 40.000000$  by simple adaptive timestepping

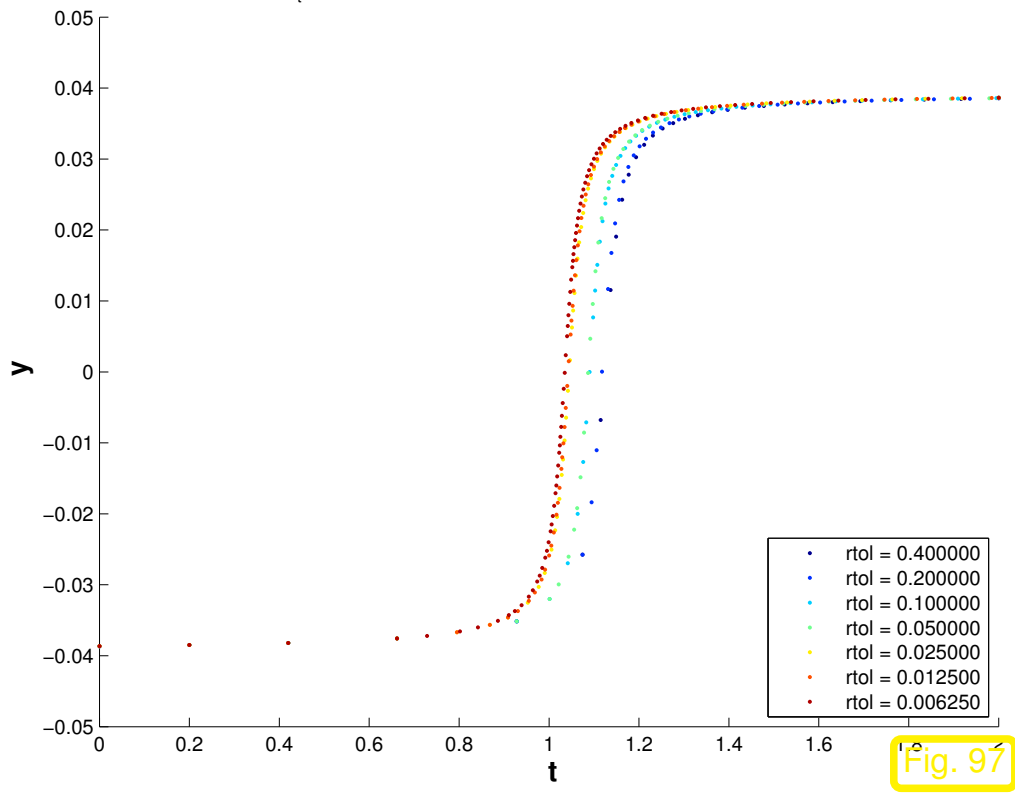


Fig. 97

Lösungen  $(y_k)_k$  für verschiedene `rtol`

Error vs. no. of timesteps for  $d_t y = a \cos(y)^2$  with  $a = 40.000000$

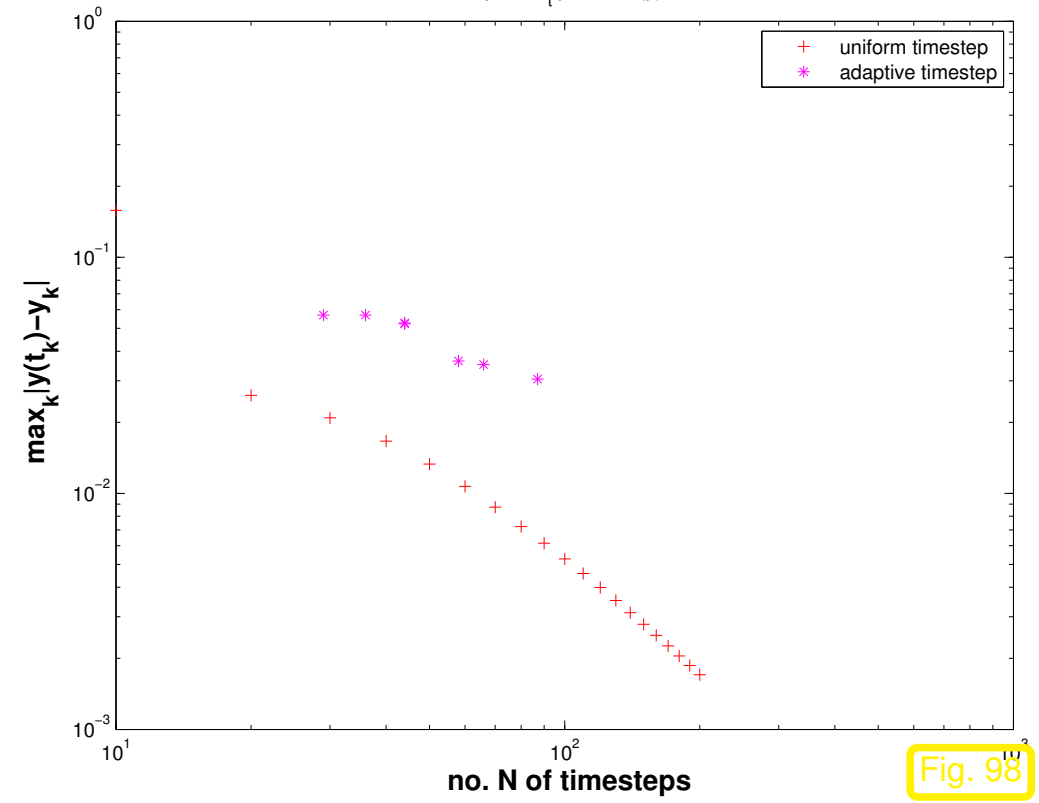


Fig. 98

Fehler als Funktion des Rechenaufwandes

☞ Grösserer Fehler bei adaptiver Schrittweitensteuerung im Vergleich zu uniformer Schrittweite

Erklärung: die Lage der steilen Flanke der Lösung hängt *sensitiv* vom Anfangswert ab. Daher werden kleine Einschrittfehler in den ersten Zeitschritten zu grossen Fehlern zur Zeit  $t \approx 1$  führen. Die lokale Schrittweitensteuerung hält diese kleinen Einschrittfehler für harmlos und kann daher nichts gegen die durch sie hervorgerufenen beträchtlichen Diskretisierungsfehler zur späteren Zeiten ausrichten.

Allgemeiner Kontext: Im Falle von *schlecht konditionierten* Anfangswertproblemen (d.h., die Lösung hängt sensitiv vom Anfangswert ab, vgl. Sect. 1.3.3.5, “chaotische Systeme”) kann selbst ein winziger Einschrittfehler, der nur im ersten Schritt passiert, zu einer von der exakten Lösung völlig abweichenden diskreten Lösung führen. Für solche Probleme ist allerdings der auf dem Konzept des Diskretisierungsfehlers aufbauende Genauigkeitsbegriff nicht mehr angemessen, siehe die Diskussion in Sect. 1.3.3.5.

Beispiel 2.6.10 (Schrittweitensteuerung und Instabilität).

- Anfangswertproblem für skalare logistische Dgl, siehe Bsp. 2.6.8, nun  $\lambda = 100$
- explizites Euler-Verfahren (1.4.2), explizite Trapezregel (2.3.3) mit Schrittweitensteuerung wie Bsp. 2.6.8

Absolute/relative Toleranz = 0.05, Anfangszeitritt (für adaptive ESV)  $h = 0.05$

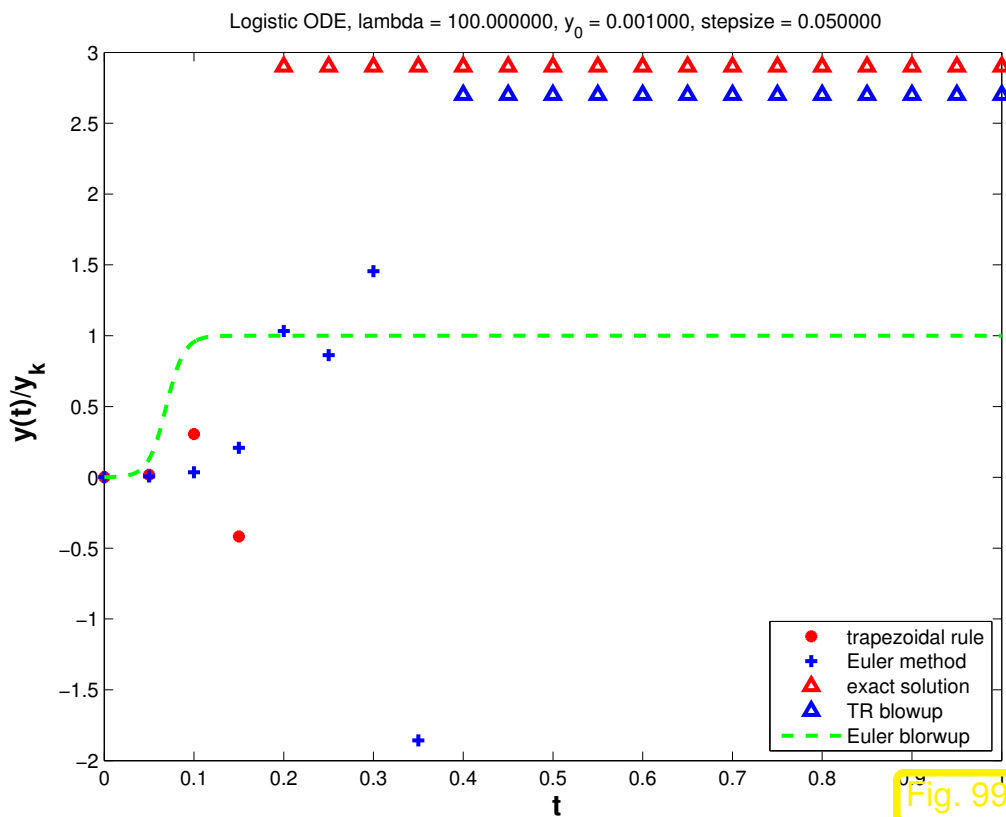


Fig. 99

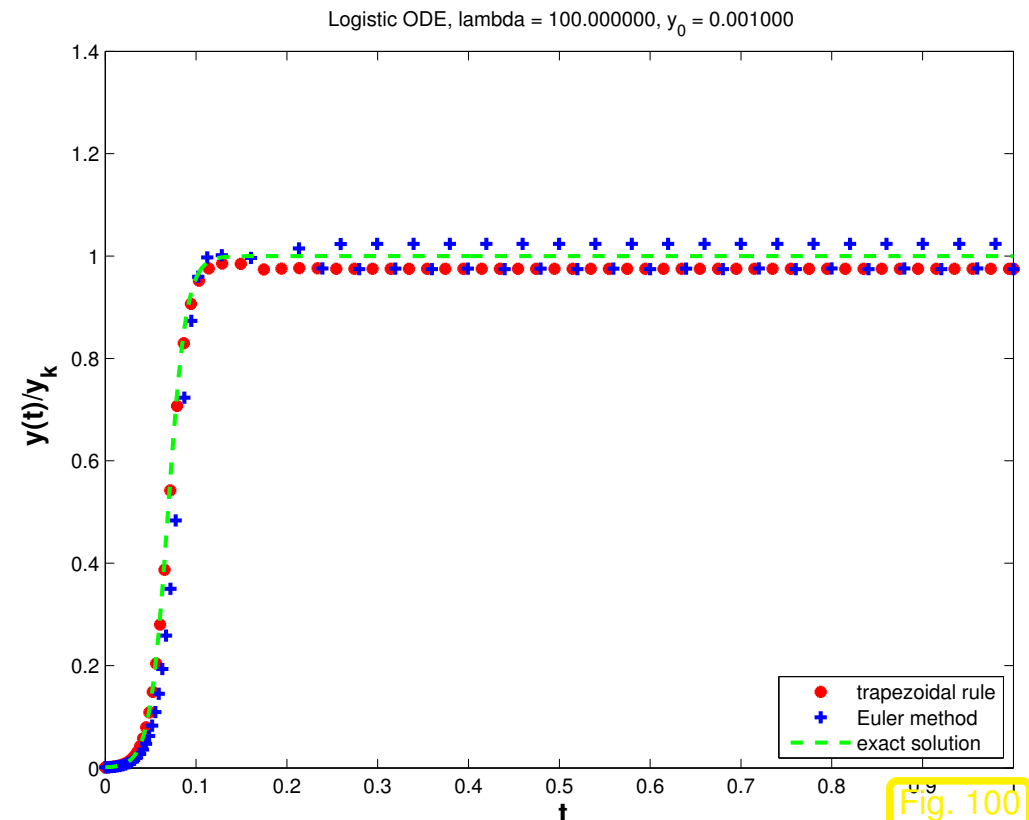


Fig. 100

R. Hiptmair  
rev 35327,  
25. April  
2011

Trapez/Euler: uniforme Zeitschrittweite  $h = 0.05$

Trapez/Euler: 119/114 Schritte, 28/42 verworfen



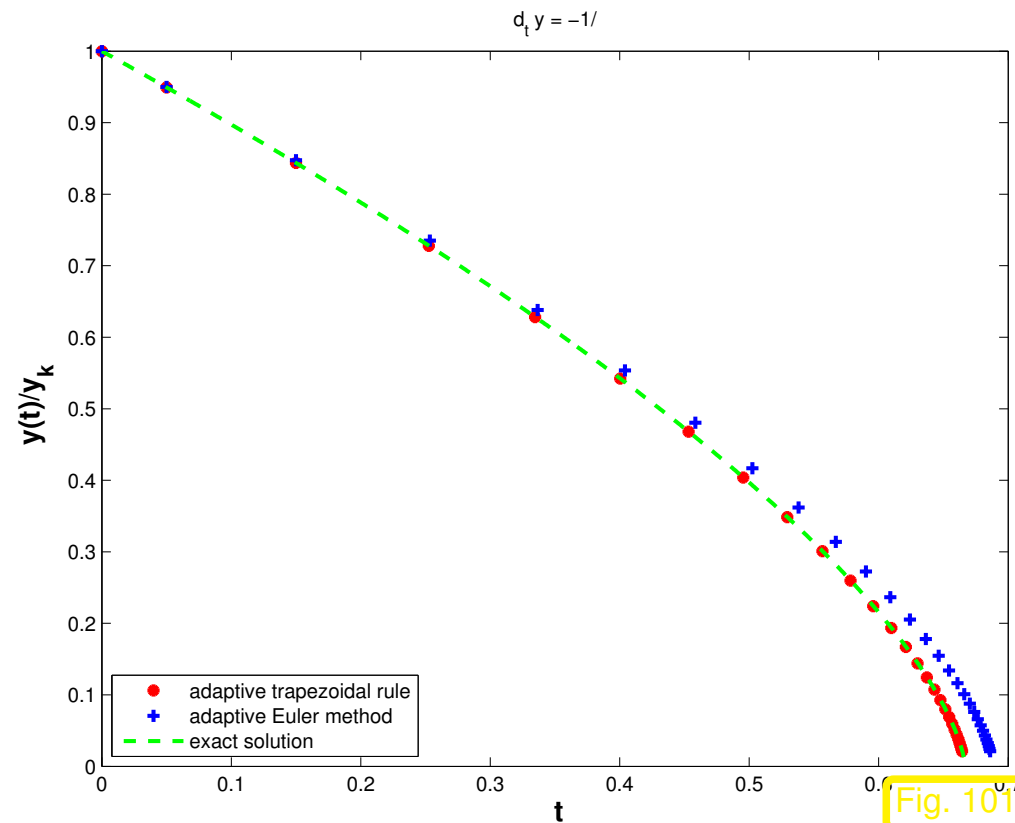
Beispiel 2.6.11 (Schrittweitensteuerung und Kollaps).

Skalares Anfangswertproblem mit Kollaps, vgl.  
Bsp. 1.3.11

$$\dot{y} = -\frac{1}{\sqrt{y}}, \quad y(0) = 1$$

$$\Rightarrow y(t) = (1 - 3t/2)^{2/3}.$$

Schrittweitensteuerung wie Bsp. 2.6.8 , absolute/relative Toleranz = 0.005



Schrittweitensteuerung ➤ Verfahren "erkennt" Kollaps der Lösung



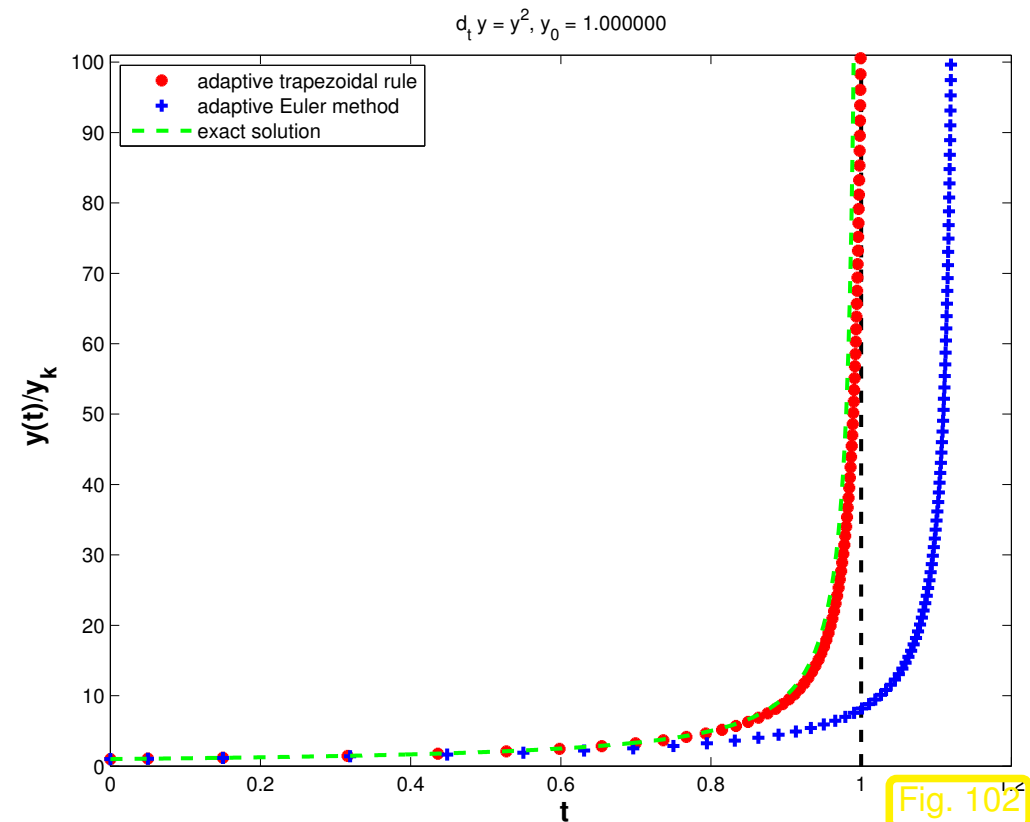
*Beispiel 2.6.12* (Schrittweitensteuerung und Blow-up).

Skalares Anfangswertproblem mit Blow-up, vgl.  
Bsp. 1.3.11

$$\dot{y} = y^2, \quad y(0) = 1$$

$$\Rightarrow y(t) = \frac{1}{1-t}.$$

Schrittweitensteuerung wie Bsp. 2.6.8, absolute/relative Toleranz = 0.05



Schrittweitensteuerung ➤ Verfahren “erkennt” Blow-up der Lösung



*Bemerkung 2.6.13* (Eingebettete RK-ESV).

Algorithmische Realisierung (ESV):

## Eingebettete Runge-Kutta-Verfahren

Gleiche Inkremente  $\mathbf{k}_i$ , verschiedene Gewichte  $b_i$   
( $\rightarrow$  Def 2.3.5) realisieren RK-Evolutionen  $\Psi_h, \tilde{\Psi}_h$   
der Ordnungen  $p$  und  $p + 1$ .

$$\begin{array}{c|c} \mathbf{c} & \mathfrak{A} \\ \hline & \mathbf{b}^T \\ \hline & \widehat{\mathbf{b}}^T \end{array} := \begin{array}{c|ccc} c_1 & a_{11} & \cdots & a_{1s} \\ \vdots & \vdots & & \vdots \\ c_s & a_{s1} & \cdots & a_{ss} \\ \hline & b_1 & \cdots & b_s \\ \hline & \widehat{b}_1 & \cdots & \widehat{b}_s \end{array} .$$

Eingebettetes RK-ESV: Butcher-Schema

$$\Psi_h \mathbf{y} = \mathbf{y} + h \sum_{i=1}^s b_i \mathbf{k}_i, \quad \tilde{\Psi}_h \mathbf{y} = \mathbf{y} + h \sum_{i=1}^s \widehat{b}_i \mathbf{k}_i .$$

Motivation: Effizienz (Inkremente  $\mathbf{k}_i$  nur einmal zu berechnen, siehe Def. 2.3.5)

Gebräuchlich:  $p = 4, p = 7$



*Beispiel* 2.6.14 (Eingebettete Runge-Kutta-Verfahren).  $\rightarrow$  [17, Sect. II.4]

0					
$\frac{1}{3}$	$\frac{1}{3}$				
$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$			
$\frac{1}{2}$	$\frac{1}{8}$	0	$\frac{3}{8}$		
1	$\frac{1}{2}$	0	$-\frac{3}{2}$	2	
$y_1$	$\frac{1}{6}$	0	0	$\frac{2}{3}$	$\frac{1}{6}$
$\hat{y}_1$	$\frac{1}{10}$	0	$\frac{3}{10}$	$\frac{2}{5}$	$\frac{1}{5}$

Eingebettes RK-Verfahren der Ordnung 3(„4“)  
von Merson

0					
$\frac{1}{2}$	$\frac{1}{2}$				
$\frac{1}{2}$	0	$\frac{1}{2}$			
1	0	0	1		
$\frac{3}{4}$	$\frac{5}{32}$	$\frac{7}{32}$	$\frac{13}{32}$	$-\frac{1}{32}$	
$y_1$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$	
$\hat{y}_1$	$-\frac{1}{2}$	$\frac{7}{3}$	$\frac{7}{3}$	$\frac{13}{6}$	$-\frac{16}{3}$

Eingebettes RK-Verfahren der Ordnung 3(4) von  
Zonneveld



0							
$\frac{1}{5}$	$\frac{1}{5}$						
$\frac{3}{10}$	$\frac{3}{40}$	$\frac{9}{40}$					
$\frac{4}{5}$	$\frac{44}{45}$	$-\frac{56}{15}$	$\frac{32}{9}$				
$\frac{8}{9}$	$\frac{19372}{6561}$	$-\frac{25360}{2187}$	$\frac{64448}{6561}$	$-\frac{212}{729}$			
1	$\frac{9017}{3168}$	$-\frac{355}{33}$	$\frac{46732}{5247}$	$\frac{49}{176}$	$-\frac{5103}{18656}$		
1	$\frac{35}{384}$	0	$\frac{500}{1113}$	$\frac{125}{192}$	$-\frac{2187}{6784}$	$\frac{11}{84}$	0
$y_1$	$\frac{35}{384}$	0	$\frac{500}{1113}$	$\frac{125}{192}$	$-\frac{2187}{6784}$	$\frac{11}{84}$	0
$\hat{y}_1$	$\frac{5179}{57600}$	0	$\frac{7571}{16695}$	$\frac{393}{640}$	$-\frac{92097}{339200}$	$\frac{187}{2100}$	$\frac{1}{40}$

DOPRI5: Eingebettes RK-  
Verfahren der Ordnung 4(5)  
von Dormand & Prince  
(MATLAB `ode45`)



Adaptive Integratoren für Anfangswertprobleme in MATLAB:

```
options = odeset('abstol', atol, 'reltol', rtol, 'stats', 'on');
[t, y] = ode45/ode23(@ (t, x) f(t, x), tspan, y0, options);
(f = function handle, tspan  $\hat{=}$   $[t_0, T]$ ,  $y_0 \hat{=}$   $\mathbf{y}_0$ ,  $t \hat{=}$   $t_k$ ,  $y \hat{=}$   $\mathbf{y}_k$ )
```

Beispiel 2.6.15 (Adaptive RK-ESV zur Teilchenbahnberechnung). → Bsp. 2.4.19

Bewegung eines geladenen Teilchens im Feld eines geraden Drahtes = Linienladung (konservatives  
Zentralfeld, Zentrum  $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ , Potential  $U(\mathbf{x}) := -2 \log \|\mathbf{x}\|$ ): → Bsp. 1.2.25

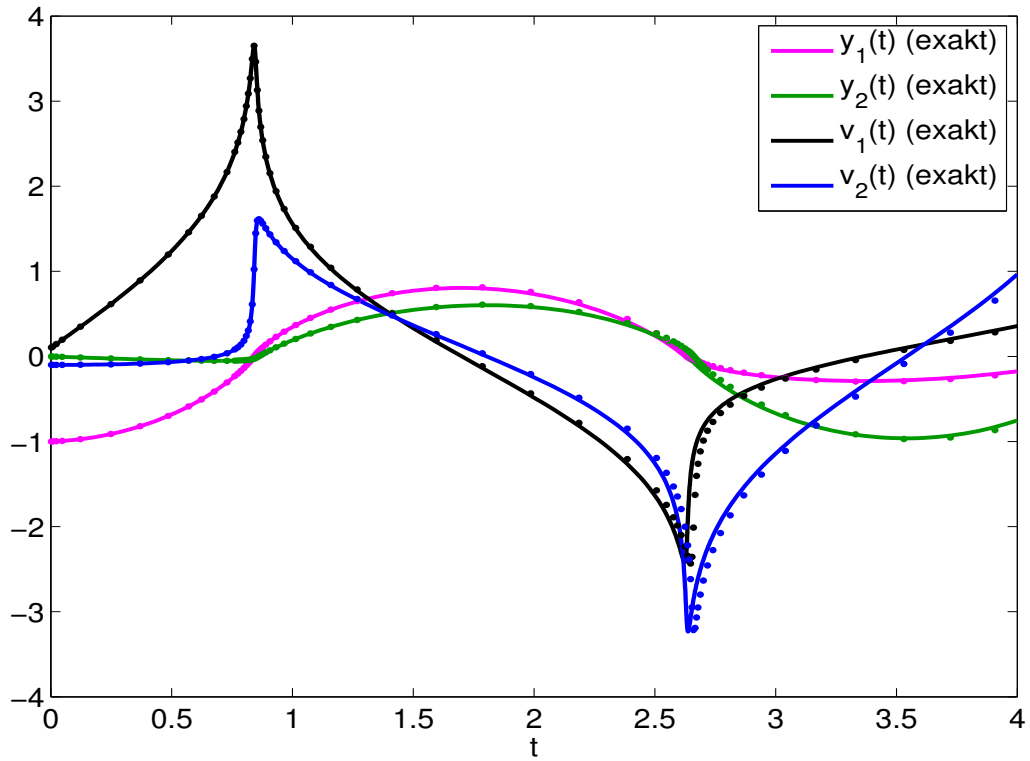
$$\ddot{\mathbf{y}} = -\frac{2\mathbf{y}}{\|\mathbf{y}\|^2} \Rightarrow \begin{pmatrix} \dot{\mathbf{y}} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{v} \\ -\frac{2\mathbf{y}}{\|\mathbf{y}\|^2} \end{pmatrix}, \quad \mathbf{y}(0) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad \mathbf{v}(0) = \begin{pmatrix} 0.1 \\ -0.1 \end{pmatrix}.$$

Anfangswert:  $\mathbf{y}(0) = (-1, 0, 0.1, -0.1)$ , Endzeitpunkt:  $T = 4$

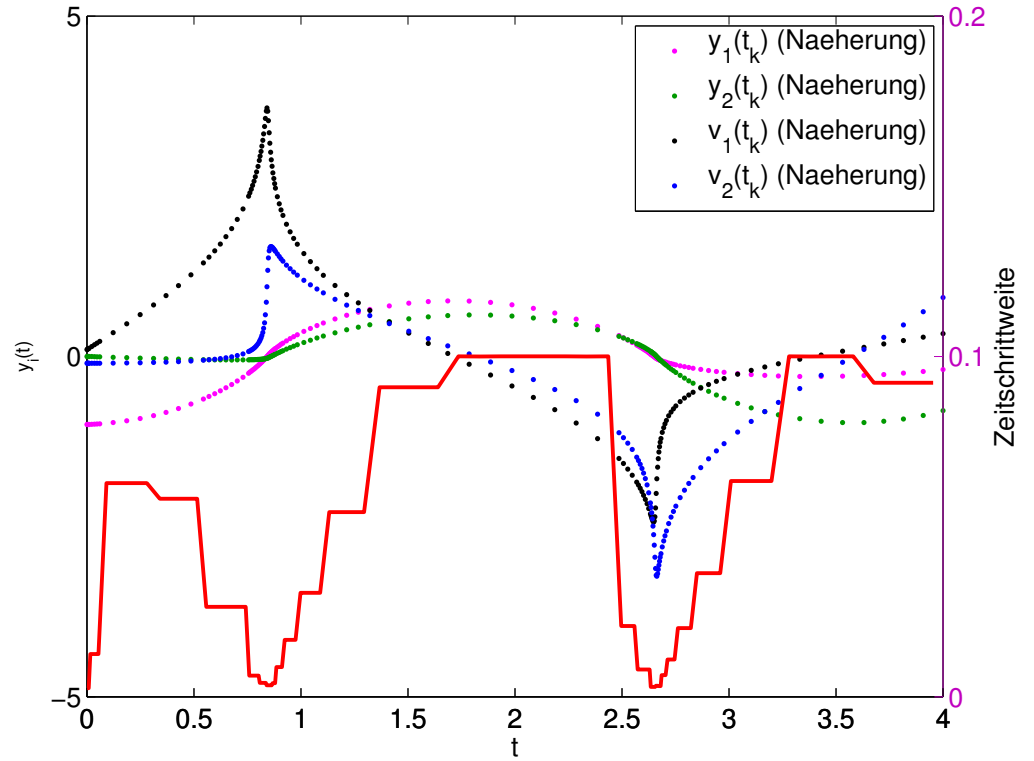
Adaptiver Integrator: `ode45(@(t,x) satf, [0 4], [-1;0;0.1;-0.1], options):`

- ❶ `options = odeset('reltol', 0.001, 'abstol', 1e-5);`
- ❷ `options = odeset('reltol', 0.01, 'abstol', 1e-3);`

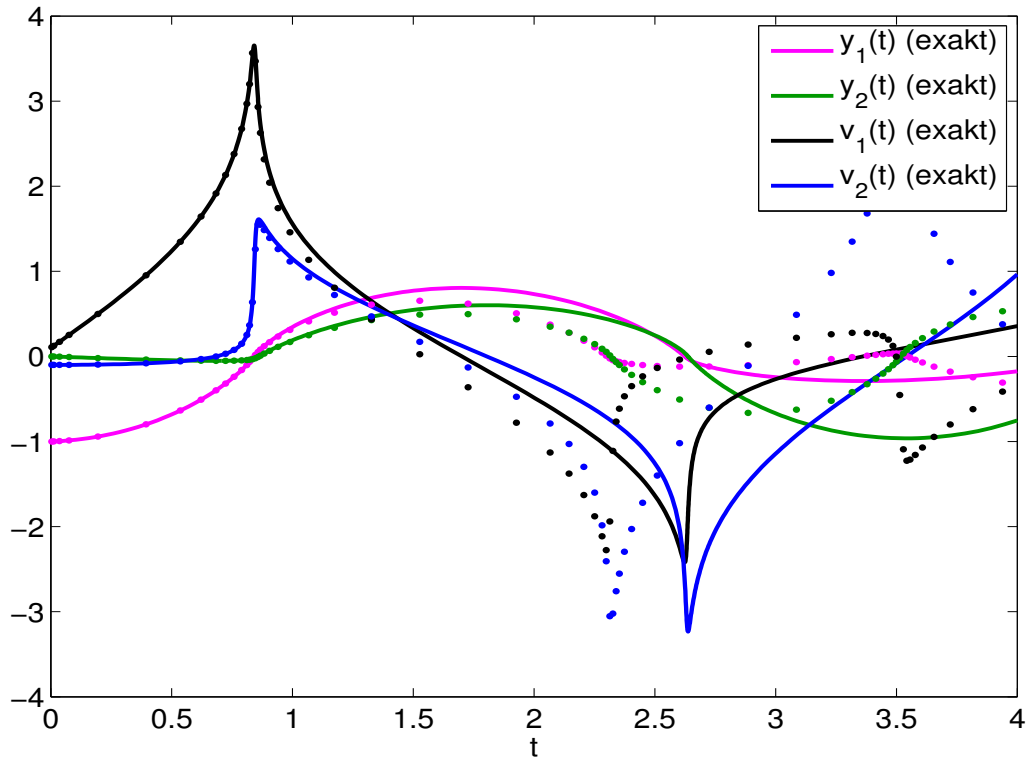
abstol = 0.000010, reltol = 0.001000



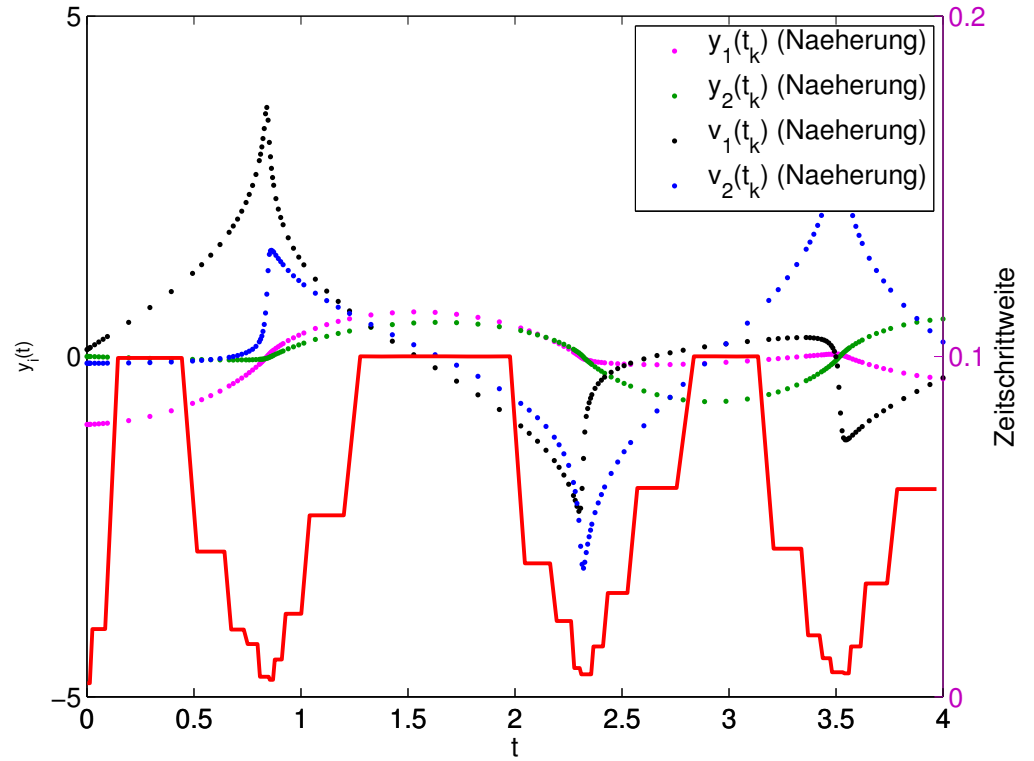
abstol = 0.000010, reltol = 0.001000

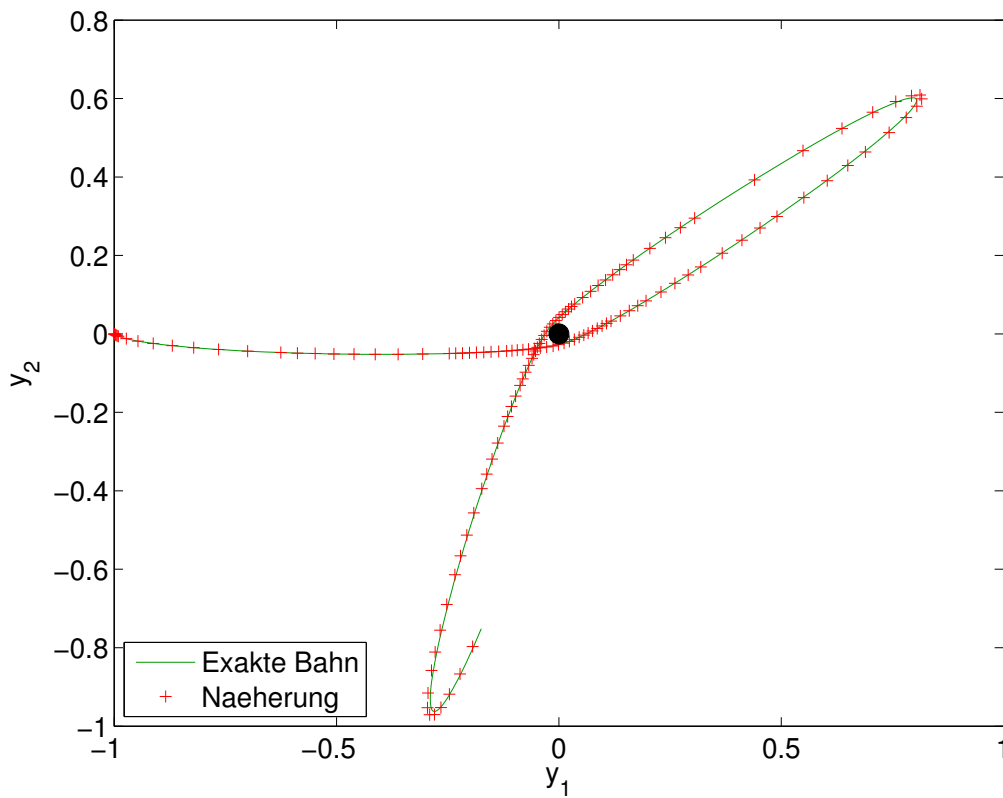


abstol = 0.001000, reltol = 0.010000

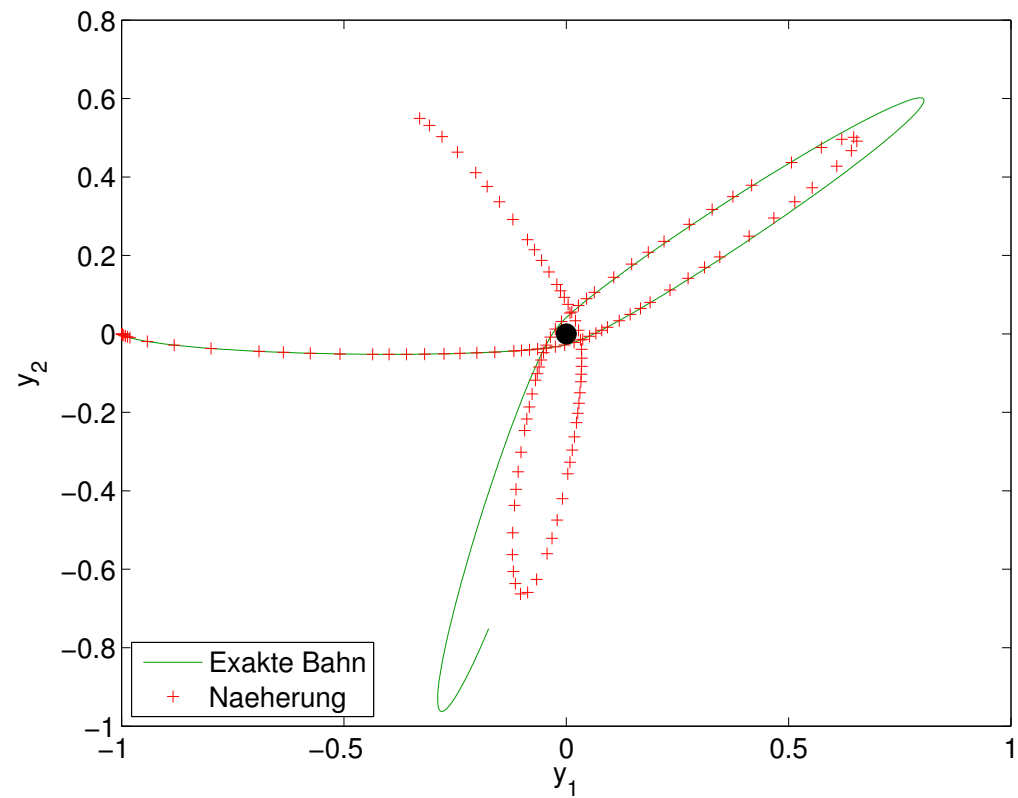


abstol = 0.001000, reltol = 0.010000





reltol=0.001, abstol=1e-5



reltol=0.01, abstol=1e-3

Qualitativ falsche Lösung bei geringfügig erniedrigter Toleranz !



R. Hiptmair  
rev 35327,  
25. April  
2011

Beispiel 2.6.16 (Schrittweitensteuerung für Bewegungsgleichungen). → Bsp. 2.4.19

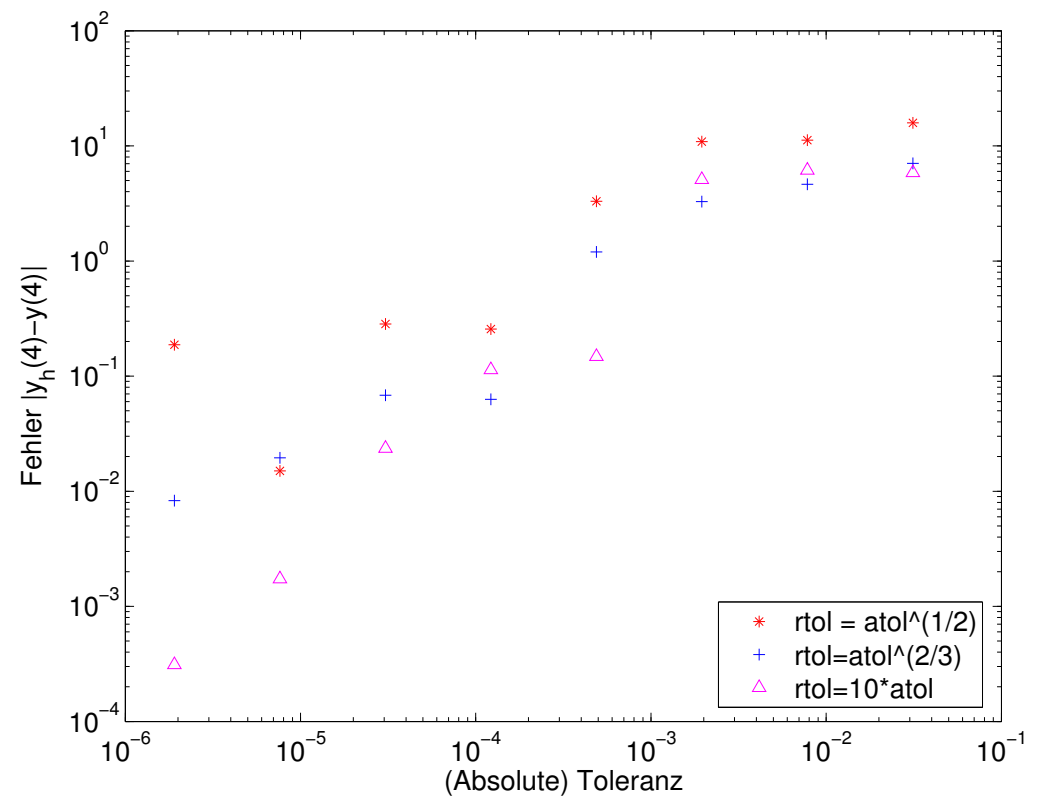
## AWP aus Bsp. 2.4.19

ode45 mit verschiedenen absoluten/relativen Toleranzen

Im Gegensatz zu Bsp. 2.6.8:

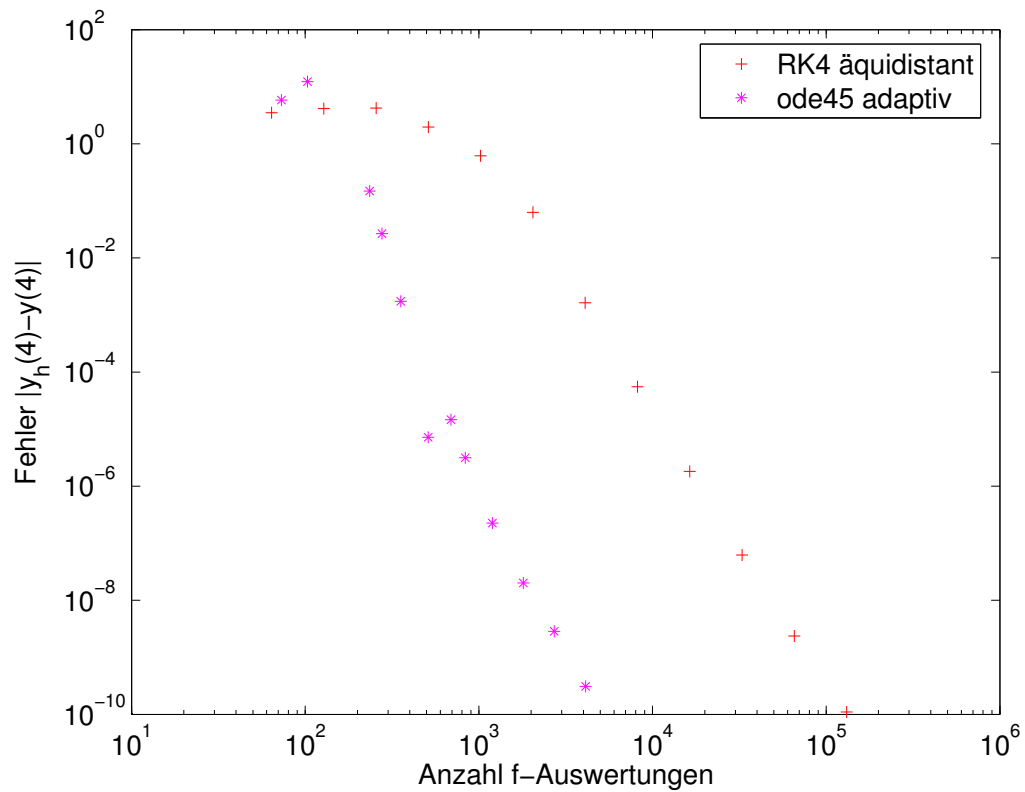
! Toleranzen sagen nichts über globalen Fehler

Erklärung: wie in Bsp. 2.6.9 liegt ein schlecht konditioniertes AWP vor, was den Einfluss von Einzschrittfehlern auf den Diskretisierungsfehler unkalulierbar macht.



R. Hiptmair

rev 35327,  
25. April  
2011



## Effizienz von Schrittweitensteuerung:

### Vergleich:

- Klassisches Runge-Kutta-Verfahren (2.3.11)
- Eingebettetes Runge-Kutta-Verfahren mit Schrittweitensteuerung: `ode45`

Aufwandsmass:  $\#f$ -Auswertungen

Adaptivität zahlt sich aus !



R. Hiptmair

rev 35327,  
25. April  
2011

# 3

## Stabilität [8, Kap. 6]

*Beispiel* 3.0.1 (Ineffizienz expliziter Runge-Kutta-Verfahren). → Bsp. 1.4.9, 1.4.15

Logistische Differentialgleichung  $\dot{y} = f(y)$ ,  $f(y) = \lambda y(1 - y) \rightarrow (2.2.84)$ ,  $\lambda = 50$ , Anfangswert  $y_0 = 0.1$ , Zeitintervall  $[0, 1]$ :

- Integratoren: Implizites Euler-Verfahren (1.4.13), klassisches Runge-Kutta-Verfahren (2.3.11)
- uniforme Zeitschrittweite  $h = 1/N$ ,  $N \in \mathbb{N}$
- Fehlermass:  $\text{err} = \max_k |y_k - y(t_k)|$ ,  $k = 1, \dots, N$



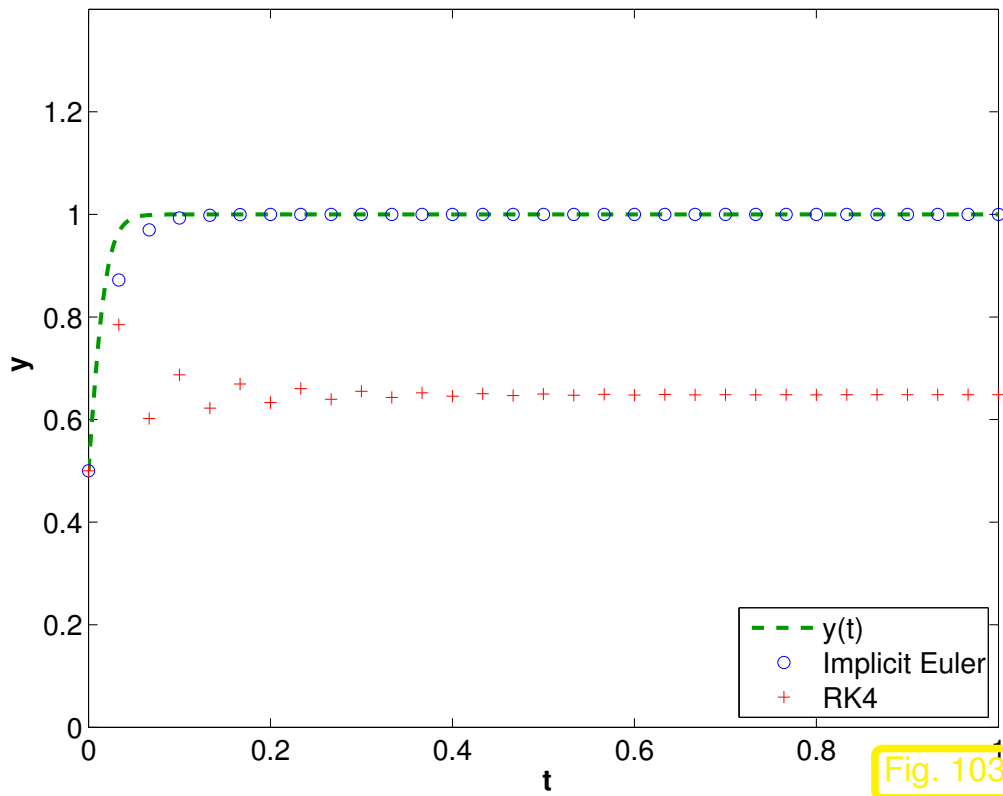


Fig. 103

(Approximative) Lösungen für  $N = 30$

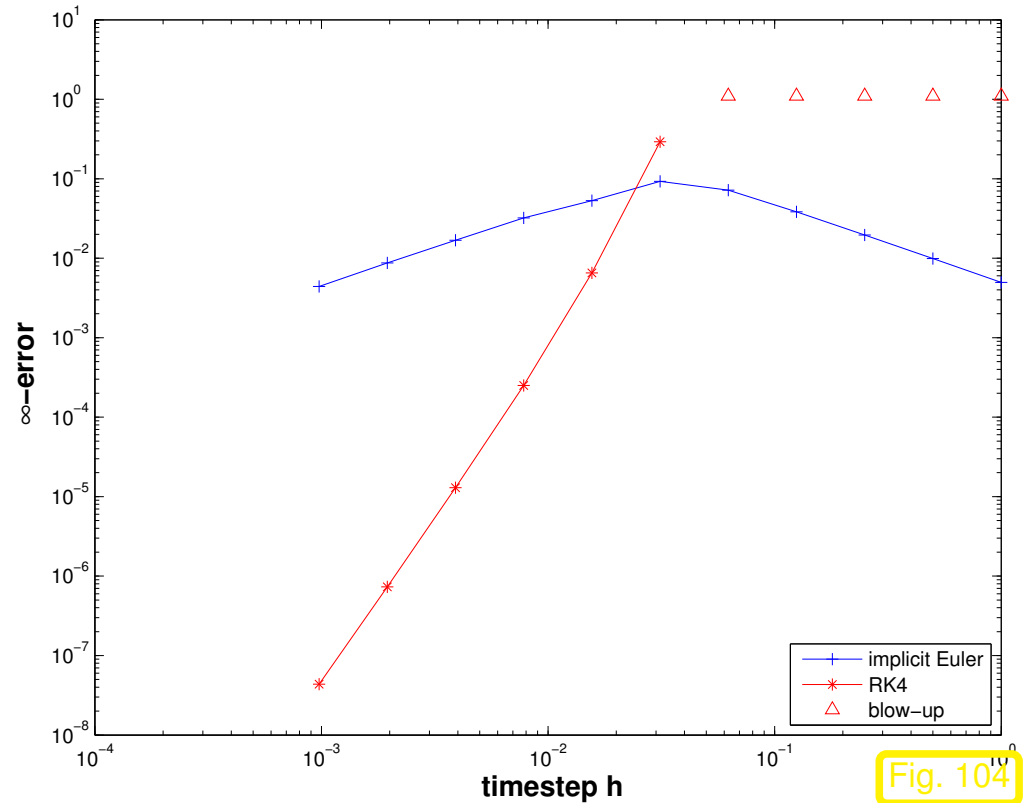


Fig. 104

Fehler gegen Schrittweite (doppeltlogarithmisch)

Beobachtung:

- RK4 *asymptotisch* genauer als implizites Euler-Verfahren
- RK4 *präasymptotisch* (für  $h > 0.02$ ) unbrauchbar (*Instabilität*)



Überlegung: **Linearisierung** um Fixpunkt, siehe Bem. 1.3.19

## 3.1 Modellproblemanalyse

Überlegung zu Bsp. 3.0.1: In der Umgebung eines Fixpunktes verhalten sich die Lösungen einer (vorläufig skalaren) ODE wie die ihrer Linearisierung.

➤ Relevanz der (um einen Fixpunkt) linearisierten ODE: Numerischer Integrator ist nur dann für die Lösung der ODE in der Nähe des Fixpunktes geeignet, wenn er sich zumindest für die linearisierte ODE bewährt.

Lineare autonome skalare ODE sind einfach:  $\dot{y} = \lambda y$  (bis auf Translation)

In diesem Abschnitt untersuchen wir das Verhalten numerischer Integratoren für solche einfachen ODEs

Nichts Neues! Erinnerung an Abschnitt 1.4.1: Einsichten in das Verhalten des expliziten Euler-Verfahrens (1.4.2) durch *Modellproblemanalyse*, d.h., analytische Untersuchung der diskreten Evolution für die skalare lineare ODE  $\dot{y} = \lambda y$ ,  $\lambda \in \mathbb{C}$ .

Autonomes skalares lineares AWP:  $\dot{y} = \lambda y$ ,  $y(0) = 1$ ,  $\operatorname{Re} \lambda < 0$  auf  $[0, \infty[$  (3.1.1)

$y(t) = e^{\lambda t} \rightarrow 0$  für  $t \rightarrow \infty$  (sog. **Asymptotische Stabilität** von  $y = 0$ ).

R. Hiptmair  
rev 35327,  
24. Juni  
2011

Beachte: komplexes  $\lambda \in \mathbb{C}$  im Modellproblem zugelassen  $\triangleright$  komplexer Zustandsraum  $\mathbb{C}$   
(Grund: „Diagonalisierungstechnik“ für lineare, autonome AWP, Sect. 1.3.2, vgl. Bem. 3.1.13)

Frage: Wann „erbt“ Lösung  $\{y_k\}_{k=0}^{\infty}$ ,  $y_{k+1} = \Psi_{\lambda}^h y_k$  ( $\Psi_{\lambda}^h \hat{=}$  diskrete Evolution) aus RK-ESV auf (unendlichem) äquidistantem Gitter (Maschenweite  $h$ ) asymptotische Stabilität ?

Dies ist eine Frage nach **Strukturerhaltung**: Übereinstimmung von qualitativen Eigenschaften der kontinuierlichen und diskreten Evolution.

*Bemerkung 3.1.2* (Reskalierung des Modellproblems).

Beachte: Anwendung eines linearen Operators auf  $\mathbb{R} \leftrightarrow$  Multiplikation mit reeller Zahl

$$L(\mathbb{R}, \mathbb{R}) \cong \mathbb{R}$$

 Notation:  $L(\mathbb{R}, \mathbb{R}) \hat{=}$  Raum linearer Operatoren auf  $\mathbb{R}$

Da AWP (3.1.1) autonom & skalar

- $\Phi_\lambda^h \in L(\mathbb{R}, \mathbb{R})$
- Anwendung von  $\Phi_\lambda^h$  auf Zustand  $y \in \mathbb{C} \sim$  Multiplikation
- $(h, \lambda) \mapsto \Phi_\lambda^h$  beschreibbar durch Funktion  $\mathbb{R} \times \mathbb{C} \mapsto \mathbb{C}$

Welche Funktion ist das?

$$\Phi_\lambda^h(y) = e^{\lambda h} y \quad \forall y \in \mathbb{R} \quad \Rightarrow \quad \text{Funktion } (h, \lambda) \mapsto e^{\lambda h} .$$

▶  $\boxed{\Phi_\lambda^h = \Phi_1^{\lambda h}} \Rightarrow$  Funktion hängt nur von Produkt  $\lambda h$  ab. (3.1.3)

Auch für die diskrete Evolution eines Runge-Kutta-Einschrittverfahrens gilt  $\Psi_\lambda^h \in L(\mathbb{R}, \mathbb{R})$

▶  $(h, \lambda) \mapsto \Psi_\lambda^h$  ebenfalls beschreibbar durch Funktion  $\mathbb{R} \times \mathbb{C} \mapsto \mathbb{C}$

Naheliegende Frage: Gilt die **Zeitskalierungsinvarianz** (3.1.3) auch für  $\Psi_\lambda^h$ , d.h. gilt

$$\Psi_\lambda^h = \Psi_1^{\lambda h} \quad \forall \lambda \in \mathbb{C}, h \text{ hinreichend klein ?}$$

Die Zeitskalierungsinvarianz (3.1.3) ist für Runge-Kutta-Einschrittverfahren ( $\rightarrow$  Def. 2.3.5) erfüllt, wie durch einfaches Nachrechnen bestätigt werden kann! ( $h$  und  $\lambda$  gehen in die Inkrementgleichung des RK-ESV für (3.1.1) nur in Form des Produkts  $h\lambda$  ein, siehe Beweis zu Thm. 3.1.6.)

▶  $\boxed{\Psi_\lambda^h = \Psi_1^{\lambda h}}$  hängt nur von  $z := \lambda h$  ab:  $S(z) := \Psi_\lambda^h$

Stabilitätsfunktion ↗ ↖ interpretiert als Zahl

Was sagt uns diese Stabilitätsfunktion über die qualitative Asymptotik der diskreten Lösung ?

▶ Diskrete Lösung:  $y_k = S(z)^k y_0$ ,  $k \in \mathbb{N}_0$ ,  $z := \lambda h$ .

$$\begin{aligned} \rightarrow |S(z)| < 1 &\Leftrightarrow \lim_{k \rightarrow \infty} y_k = 0 \quad \forall y_0 \in \mathbb{R} \\ &\Leftrightarrow y = 0 \text{ asymptotisch stabil } (\rightarrow \text{Def. 3.2.2}) \text{ für diskrete Evolution } \Psi_\lambda^h. \end{aligned}$$

**Definition 3.1.4** (Stabilitätsgebiet eines Einschrittverfahrens). [8, Sect. 6.1.2]

Das **Stabilitätsgebiet** eines ESV für das AWP (3.1.1) auf der Grundlage der diskreten Evolution  $\Psi_\lambda^h y =: S(z)y$ ,  $y \in \mathbb{C}$ ,  $z := \lambda h$ ,  $S : D_S \subset \mathbb{C} \mapsto \mathbb{C}$ , ist

$$\mathcal{S}_\Psi := \{z \in D_S : |S(z)| < 1\} \subset \mathbb{C}.$$

☞ Für von RK-ESV zu AWP (3.1.1) erzeugte Gitterfunktion  $\{y_k\}_{k \in \mathbb{N}}$  auf äquidistantem Zeitgitter mit Maschenweite  $h > 0$  gilt

$$y_0 \neq 0: \quad \lim_{k \rightarrow \infty} y_k = 0 \quad \Leftrightarrow \quad \lim_{k \rightarrow \infty} S(h\lambda)^k = 0 \quad \Leftrightarrow \quad h\lambda \in \mathcal{S}_\Psi. \quad (3.1.5)$$

**Theorem 3.1.6** (Stabilitätsfunktion von Runge-Kutta-Verfahren).

Die diskrete Evolution  $\Psi_\lambda^h$  zu einem  $s$ -stufigen Runge-Kutta-Einschrittverfahren ( $\rightarrow$  Def. 2.3.5) mit Butcher-Schema  $\begin{array}{c|c} \mathbf{c} & \mathfrak{A} \\ \hline & \mathbf{b}^T \end{array}$  (siehe (2.3.6)) für die ODE  $\dot{y} = \lambda y$  ist ein Multiplikationsoperator der Form

$$\Psi_\lambda^h = \underbrace{1 + z\mathbf{b}^T (\mathbf{I} - z\mathfrak{A})^{-1} \mathbf{1}}_{\text{Stabilitätsfunktion } S(z)} = \frac{\det(\mathbf{I} - z\mathfrak{A} + z\mathbf{1}\mathbf{b}^T)}{\det(\mathbf{I} - z\mathfrak{A})}, \quad z := \lambda h, \quad \mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^s.$$

**Bemerkung 3.1.7** (Interpretation der Stabilitätsfunktion).

$$\Psi_\lambda^h y = S(z)y = (1 + \lambda h \mathbf{b}^T (\mathbf{I} - \lambda h \mathfrak{A})^{-1} \mathbf{1})y$$

Diskrete Evolution

$$\Phi_\lambda^h = e^{\lambda h}$$

(Kontinuierliche) Evolution

➤  $S(z) \approx \exp(z)$ : Stabilitätsfunktion = Approximation der Exponentialfunktion (um 0).



**Korollar 3.1.8.***Explizite Runge-Kutta-Verfahren*

➤  $S(z) \in \mathcal{P}_s,$

*Allgemeine Runge-Kutta-Verfahren*

➤  $S(z) = \frac{P(z)}{Q(z)}, P, Q \in \mathcal{P}_s.$

*Beweis.* Aus der Determinantenformel von Thm. 3.1.6:

$$\text{Explizite Runge-Kutta-Verfahren} \Rightarrow \mathfrak{A} \text{ echte untere Dreiecksmatrix} \Rightarrow \det(\mathbf{I} - z\mathfrak{A}) = 1$$

Allgemein ist  $z \mapsto \det(\mathbf{I} - z\mathbf{M}), \mathbf{M} \in \mathbb{R}^{s,s}$ , ein Polynom vom Grad  $s$ , wie aus der kombinatorischen Definition der Determinante folgt. □

Korollar (3.1.8) ➤ Kein RK-ESV kann  $\dot{y} = \lambda y$  bei vorgegebener Schrittweite  $h$  für alle  $\lambda \in \mathbb{R}$  ohne Fehler lösen, denn die Exponentialfunktion ( $\rightarrow$  Bem. 3.1.7) lässt sich natürlich durch keine rationale Funktion modellieren.

Korollar (3.1.8) ➤ **Ordnungsschranken** für explizite/implizite RK-ESV, siehe Sect. 2.3.2



**Lemma 3.1.9** (Rationale Approximation der Exponentialfunktion).

Ist  $S(z) = \frac{P(z)}{Q(z)}$ ,  $P, Q \in \mathcal{P}_s$ ,  $s \in \mathbb{N}$ , so gilt

$$S(z) - \exp(z) = O(|z|^m) \quad \text{für } z \rightarrow 0 \quad \Rightarrow \quad m \leq 2s + 1 .$$

*Beweis:* ( $\rightarrow$  [8, Lemma 6.4], doch der dortige Beweis ist falsch!)

Indirekte Beweisführung, Annahme  $S(z) - \exp(z) = O(|z|^{2s+2})$  für  $z \rightarrow 0$ :

Ansatz:

$$P(z) = p_0 + p_1 z + \cdots + p_s z^s ,$$

$$Q(z) = q_0 + q_1 z + \cdots + q_s z^s , \quad q_0 = 1, \text{ da O.B.d.A } Q(0) = 1 .$$

►  $Q(z) \exp(z) - P(z) = \alpha_{2s+2} z^{2s+2} + \alpha_{2s+3} z^{2s+3} + \dots$  (global konvergente Potenzreihe) .

Einsetzen der Exponentialreihe und Multiplikation, dann Koeffizientenvergleich ➤ lineares Gleichungssystem

$$\sum_{j=0}^s q_j \frac{1}{(i-j)!} = 0 , \quad i = s+1, \dots, 2s+1 .$$

$$\sum_{j=0}^s q_j \frac{1}{(i-j)!} - p_i = 0 , \quad i = 0, \dots, s .$$



Dieses hat nur die triviale Lösung, was auf einen Widerspruch zu  $q_0 = 1$  führt.

*Beispiel 3.1.10* (Stabilitätsfunktionen einiger RK-ESV).

• Explizites Euler-Verfahren (1.4.2): 
$$\frac{0 \mid 0}{1} \quad \Rightarrow \quad S(z) = 1 + z .$$

• Implizites Eulerverfahren (1.4.13): 
$$\frac{1 \mid 1}{1} \quad \Rightarrow \quad S(z) = \frac{1}{1 - z} .$$

• Explizite Trapezregel (2.3.3): 
$$\frac{0 \mid 0 \ 0}{1 \mid 1 \ 0} \quad \Rightarrow \quad S(z) = 1 + z + \frac{1}{2}z^2 .$$

• Implizite Mittelpunktsregel (2.2.19): 
$$\frac{\frac{1}{2} \mid \frac{1}{2}}{1} \quad \Rightarrow \quad S(z) = \frac{1 + \frac{1}{2}z}{1 - \frac{1}{2}z} .$$

- RK4-Verfahren (2.3.11):

$$\begin{array}{c|cccc}
 0 & 0 & 0 & 0 & 0 \\
 \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\
 \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\
 1 & 0 & 0 & 1 & 0 \\
 \hline
 & \frac{1}{6} & \frac{2}{6} & \frac{2}{6} & \frac{1}{6}
 \end{array}$$

$$\triangleright S(z) = 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 + \frac{1}{24}z^4.$$



*Beispiel 3.1.11* (Verhalten von Stabilitätsfunktionen).

Verhalten von Stabilitätsfunktionen (für reelles Argument  $z$ ):

Bem. 3.1.7  $\triangleright$  wir erwarten, dass sich die Stabilitätsfunktionen in  $z = 0$  an  $\exp(z)$  “anschmiegen”, d.h., beiden Funktionen stimmen im Werte und einigen niedrigsten Ableitungen überein. Die *Mindestzahl* der übereinstimmenden Ableitungen ist gegeben durch die Konsistenzordnung des Einschrittverfahrens.

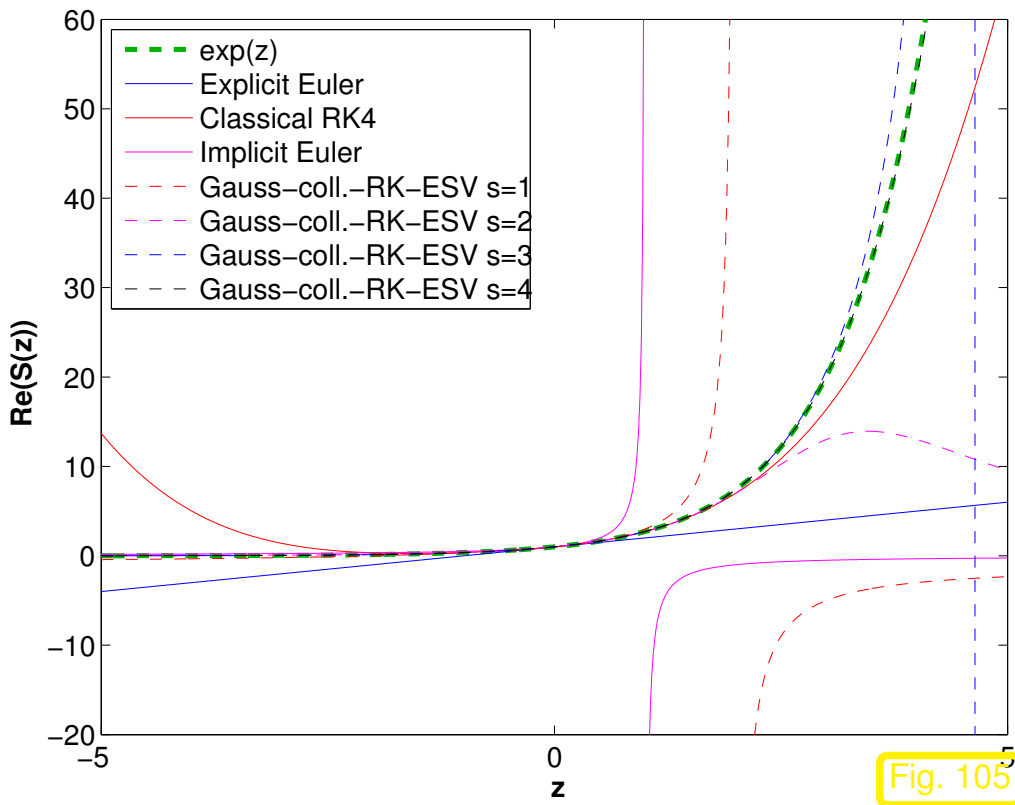


Fig. 105

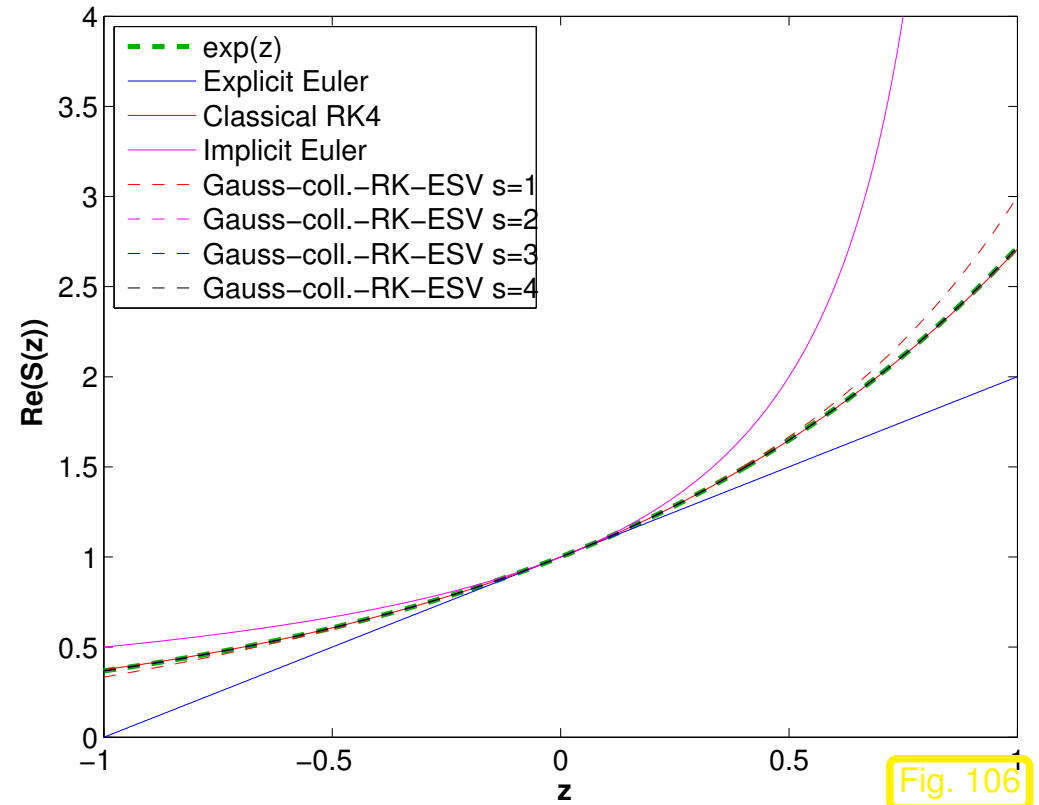
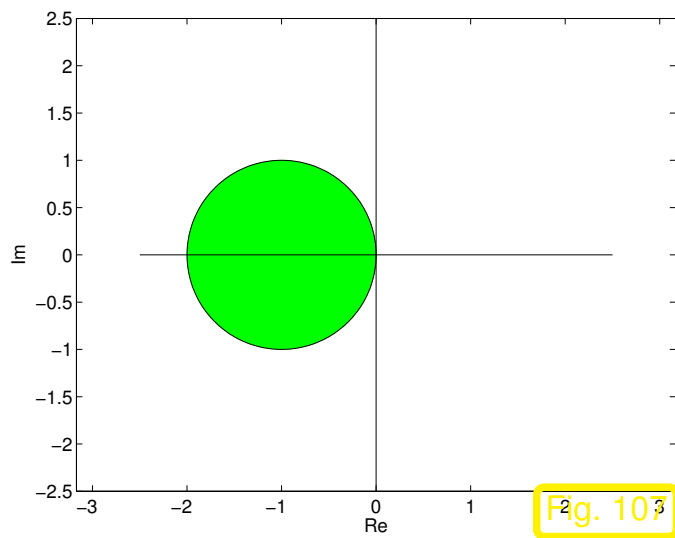


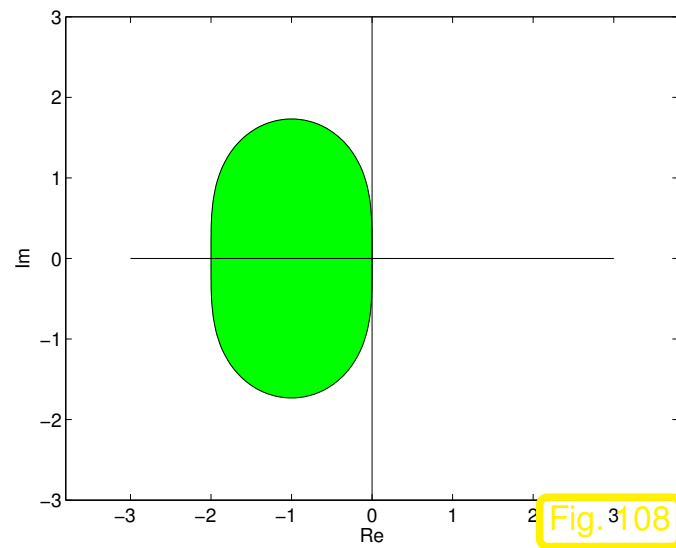
Fig. 106



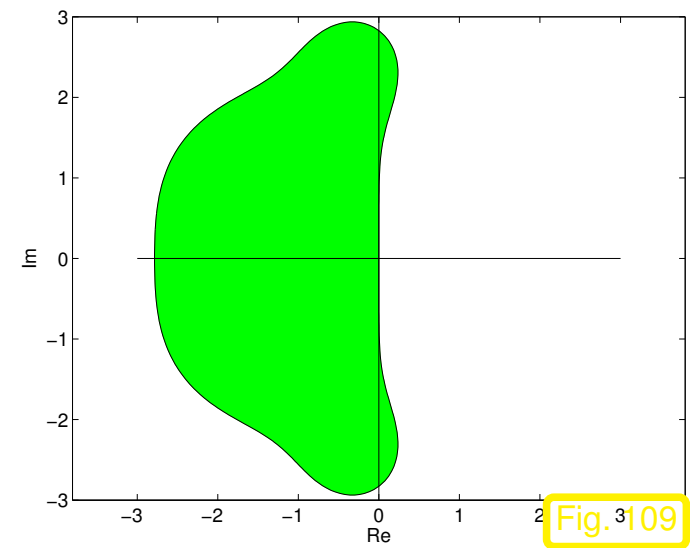
Beispiele: Stabilitätsgebiete  $\mathcal{S}$  expliziter RK-ESV:



$\mathcal{S}_\Psi$ : expliziter Euler (2.2.1)

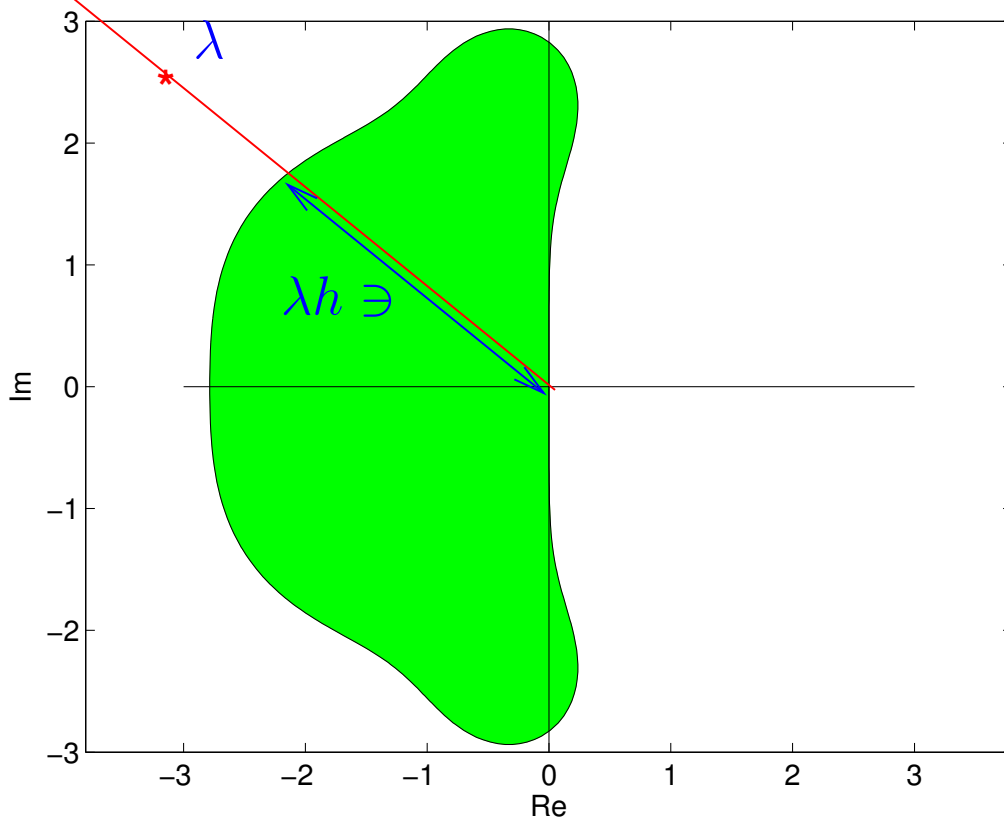


$\mathcal{S}_\Psi$ : explizite Trapezregel  
Mittelpunktsregel



$\mathcal{S}_\Psi$ : RK4-Verfahren (2.3.11)  
Kuttas 3/8-Regel (2.3.12)

Das Stabilitätsgebiet expliziter RK-Verfahren ist beschränkt !



► **Stabilitätsbedingte Schrittweitenbeschränkung** für explizite RK-ESV angewandt auf (3.1.1):

$$h < \sup\{t > 0: t\lambda \in \mathcal{S}_\Psi\}. \quad (3.1.12)$$

Für eine detailliertere Betrachtung siehe [8, Lemma 6.6]

*Bemerkung 3.1.13* (RK-ESV für autonome homogene lineare ODE). ← siehe Sect. 1.3.2

Was ist die Diskrete Evolution eines RK-ESV angewandt auf

homogene, autonome, lineare ODE  $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$ ,  $\mathbf{A} \in \mathbb{C}^{d,d}$  ?

① Annahme: **A diagonalisierbar**  $\Leftrightarrow \exists \mathbf{S} \in \mathbb{C}^{d,d}$  regulär:  $\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \mathbf{D} := \text{diag}(\lambda_1, \dots, \lambda_d)$

Folgerung aus **Affin-Kovarianz** von RK-ESV ( $\rightarrow$  Bem. 2.3.13):

Ist  $\widehat{\Psi}$  die diskrete Evolution zu  $\frac{d}{dt}\widehat{\mathbf{y}} = \mathbf{D}\widehat{\mathbf{y}}$  (entkoppelte skalare lineare ODE!), dann, mit  $\widehat{\mathbf{y}} := \mathbf{S}^{-1}\mathbf{y}$ ,

$$\begin{aligned} \Psi^h \mathbf{y} &\stackrel{(2.3.14)}{=} \mathbf{S} \widehat{\Psi}^h \mathbf{S}^{-1} \mathbf{y} = \mathbf{S} \begin{pmatrix} \widehat{\Psi}_{\lambda_1}^h \widehat{y}_1 \\ \vdots \\ \widehat{\Psi}_{\lambda_d}^h \widehat{y}_d \end{pmatrix} = \mathbf{S} \begin{pmatrix} S(h\lambda_1) & & \\ & \cdots & \\ & & S(h\lambda_d) \end{pmatrix} \widehat{\mathbf{y}} \\ &= \mathbf{S} \begin{pmatrix} P(h\lambda_1) & & \\ & \cdots & \\ & & P(h\lambda_d) \end{pmatrix} \mathbf{S}^{-1} \left( \mathbf{S} \begin{pmatrix} Q(h\lambda_1) & & \\ & \cdots & \\ & & Q(h\lambda_d) \end{pmatrix} \mathbf{S}^{-1} \right)^{-1} \mathbf{y} \\ &= \mathbf{S} P(h\mathbf{D}) \mathbf{S}^{-1} \left( \mathbf{S} Q(h\mathbf{D}) \mathbf{S}^{-1} \right)^{-1} \mathbf{y} = P(h\mathbf{A}) Q(h\mathbf{A})^{-1} \mathbf{y} = S(h\mathbf{A}) \mathbf{y} . \end{aligned}$$

wobei  $S(z) = P(z)/Q(z) \hat{=}$  Stabilitätsfunktion ( $\rightarrow$  Thm. 3.1.6) des RK-ESV.

② Allgemeine Matrix  $\mathbf{A} \in \mathbb{C}^{d,d}$  mit Eigenwerten (mit Vielfachheit gezählt)  $\lambda_1, \dots, \lambda_d \in \mathbb{C}$

## Hilfsmittel: Schur-Zerlegung

*Lemma 3.1.14 (Schur-Zerlegung). Zu jeder Matrix  $\mathbf{A} \in \mathbb{C}^{d,d}$  existiert eine unitäre Matrix  $\mathbf{U} \in \mathbb{C}^{d,d}$  und eine obere Dreiecksmatrix  $\mathbf{T} \in \mathbb{C}^{d,d}$  so, dass*

$$\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^H .$$

Aus der Schur-Zerlegung für Matrizen folgt, dass die diagonalisierbaren Matrizen in  $\mathbb{C}^{d,d}$  dicht liegen (bzgl. der von der Euklidischen Norm induzierten Matrixnorm): addiere zu  $\mathbf{T}$  eine Diagonalmatrix  $\mathbf{D}$  mit beliebig kleinen Diagonaleinträgen so, dass  $\mathbf{T} + \mathbf{D}$  paarweise verschiedene Diagonaleinträge hat. Dann ist  $\mathbf{U}(\mathbf{T} + \mathbf{D})\mathbf{U}^H$  diagonalisierbar, denn auch diese Matrix hat paarweise verschiedene Eigenwerte.

➤ Ist  $\sigma(h\mathbf{A}) \subset D_S$ , dann gibt es also eine Folge diagonalisierbarer Matrizen  $\mathbf{A}_n \rightarrow \mathbf{A}$ ,  $n \in \mathbb{N}$ ,  $\sigma(h\mathbf{A}_n) \subset D_S$ , für die offensichtlich infolge der Stetigkeit der Matrixmultiplikation gilt

$$P(h\mathbf{A}_n) \rightarrow P(h\mathbf{A}) \quad , \quad Q(h\mathbf{A}_n) \rightarrow Q(h\mathbf{A}) .$$



Wegen der Stetigkeit der Matrixinversion  $\mathbf{A} \mapsto \mathbf{A}^{-1}$  auf  $GL(d) := \{\mathbf{M} \in \mathbb{C}^{d,d} : \mathbf{M} \text{ regulär}\}$  folgt damit

$$S(h\mathbf{A}_n) = P(h\mathbf{A}_n)Q(h\mathbf{A}_n)^{-1} \rightarrow P(h\mathbf{A})Q(h\mathbf{A})^{-1} = S(h\mathbf{A}) . \quad (3.1.15)$$

Ist nun  $\Psi_n^h$  die diskrete Evolution zu  $\dot{\mathbf{y}} = \mathbf{A}_n \mathbf{y}$ , so gilt

$$\Psi^h \mathbf{y} = \lim_{n \rightarrow \infty} \Psi_n^h \mathbf{y} \stackrel{\textcircled{1}}{=} \lim_{n \rightarrow \infty} S(h\mathbf{A}_n) \mathbf{y} \stackrel{(3.1.15)}{=} S(h\mathbf{A}) .$$

Stabilitätsfunktion

$$\blacktriangleright \quad \Psi^h \mathbf{y} = S(h\mathbf{A}) \mathbf{y} \quad \forall \mathbf{y} \in \mathbb{C}^d . \quad (3.1.16)$$

△

*Bemerkung 3.1.17* (Funktionenkalkül für Matrizen).

Für  $\mathbf{A} \in \mathbb{R}^{d,d}$ :

- Klar ist  $p(\mathbf{A}) = \sum_{j=1}^s c_j \mathbf{A}^j$  für Polynom  $p \in \mathcal{P}_s$ ,  $p(z) = \sum_{j=1}^s c_j z^j$ .

- Für rationale Funktion

$$R(z) = \frac{\sum_{j=1}^s p_j z^j}{\sum_{j=1}^s q_j z^j} \quad \blacktriangleright \quad R(\mathbf{A}) = \left( \sum_{j=1}^s q_j \mathbf{A}^j \right)^{-1} \left( \sum_{j=1}^s p_j \mathbf{A}^j \right), \quad (3.1.18)$$

falls  $\sum_{j=1}^s q_j \mathbf{A}^j$  invertierbar.

- Ist  $f(z) = \sum_{i=0}^{\infty} a_i z^i$  eine **Potenzreihe** mit Konvergenzradius  $\rho > 0$ , so ist

$$f(\mathbf{A}) := \sum_{i=0}^{\infty} a_i \mathbf{A}^i \quad \text{wohldefiniert für } \|\mathbf{A}\| < \rho.$$

So lassen sich transzendente Funktionen von Matrizen, wie etwa die Matrixexponentialfunktion (1.3.14) definieren.

Für alle oben eingeführten Matrixfunktionen gilt, vgl. 1.3.15,

$$\boxed{\mathbf{A} = \mathbf{S}^{-1} \mathbf{B} \mathbf{S} \Rightarrow f(\mathbf{A}) = \mathbf{S}^{-1} f(\mathbf{B}) \mathbf{S}} \quad \forall \mathbf{A}, \mathbf{B} \in \mathbb{C}^{d,d}, \quad \mathbf{S} \in \mathbb{C}^{d,d} \text{ regulär.} \quad (3.1.19)$$

Für das Spektrum gilt

$$\sigma(f(\mathbf{A})) = f(\sigma(\mathbf{A})) := \{f(\lambda) : \lambda \in \sigma(\mathbf{A})\} . \quad (3.1.20)$$

## 3.2 Vererbung asymptotischer Stabilität

### 3.2.1 Attraktive Fixpunkte

Betrachte: Autonomes AWP  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ ,  $\mathbf{f} \in C^1(D, \mathbb{R}^d)$ ,  $D \subset \mathbb{R}^d$  offen.

**Definition 3.2.1** (Fixpunkt).

$\mathbf{y}^*$  ist **Fixpunkt** (stationärer Punkt) von  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ , falls  $\mathbf{f}(\mathbf{y}^*) = 0$ .

Die Begriffsbildung ist klar: ein Fixpunkt repräsentiert einen Zustand, der sich während der Evolution nicht ändert:

$$\mathbf{y}(0) = \mathbf{y}^* \Rightarrow \mathbf{y}(t) = \mathbf{y}^* \quad \forall t \in \mathbb{R}.$$

**Definition 3.2.2** (Asymptotische Stabilität eines Fixpunkts).  $\rightarrow [8, \text{Def. 3.19}]$

Fixpunkt  $\mathbf{y}^* \in D$  *asymptotisch stabil* (attraktiv)

$$:\Leftrightarrow \exists \delta > 0: \|\mathbf{y}_0 - \mathbf{y}^*\| < \delta \Rightarrow \mathbb{R}_0^+ \subset J(\mathbf{y}_0) \quad \wedge \quad \lim_{t \rightarrow \infty} \mathbf{y}(t) = \mathbf{y}^*,$$

wobei  $\mathbf{y}(t)$  Lösung des AWP  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}), \mathbf{y}(0) = \mathbf{y}_0$ .

Erinnerung an Def. 1.3.1:  $J(\mathbf{y}_0) \hat{=}$  maximales Existenzintervall der Lösung einer autonomen Differentialgleichung zum Anfangswert  $\mathbf{y}_0$ .

“Asymptotische Stabilität” in Worten: Ein Fixpunktzustand  $\mathbf{y}^*$  ist asymptotisch stabil/attraktiv, wenn alle Lösungskurven, die hinreichend nahe bei ihm starten gegen  $\mathbf{y}^*$  konvergieren.

Das folgende Beispiel vermittelt eine bildliche Vorstellung:

*Beispiel 3.2.3* (Attraktive und repulsive Fixpunkte einer skalaren ODE).

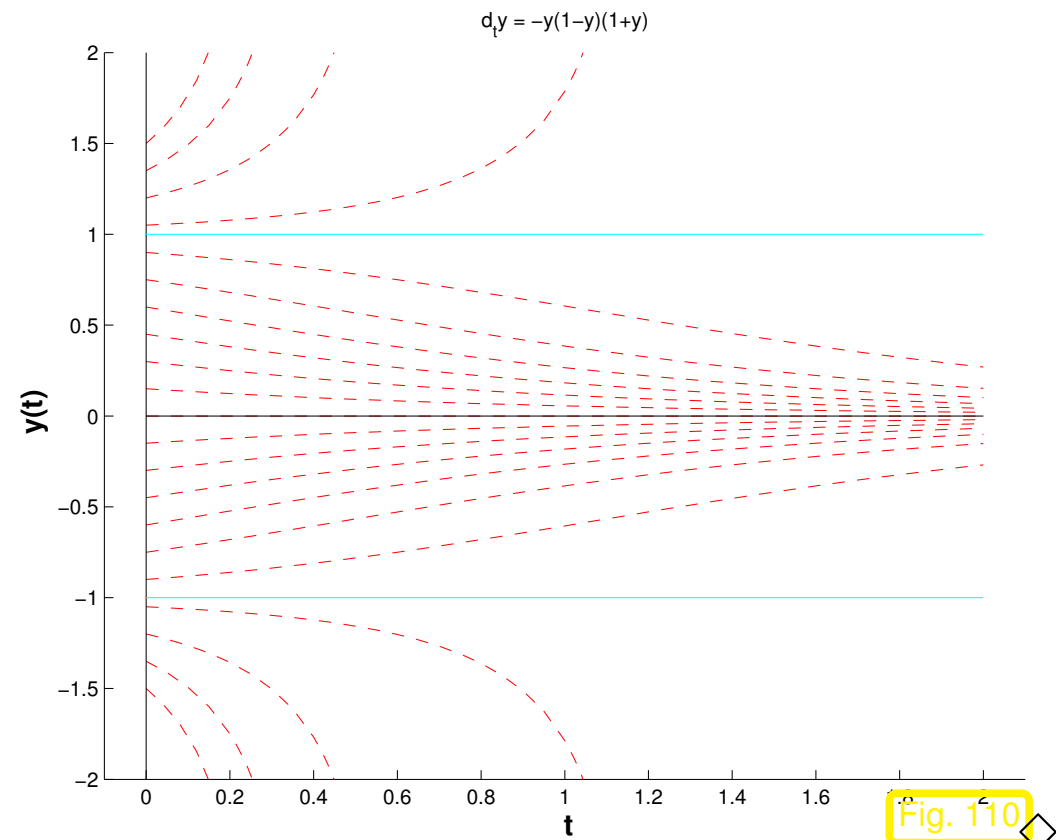
Lösungskurven der ODE

$$\dot{y} = -y(1-y)(1+y)$$

$y^* = 0$  ist asymptotisch stabiler (attraktiver) Fixpunkt,

$y^* = \pm 1$  sind instabile (repulsive) Fixpunkte

( $\rightarrow$  Bsp. 1.2.1)



**Theorem 3.2.4** (Hinreichende Bedingung für asymptotische Stabilität).

Fixpunkt  $\mathbf{y}^* \in D$  ist asymptotisch stabil, falls

$$\sigma(D\mathbf{f}(\mathbf{y}^*)) \subset \mathbb{C}^- := \{z \in \mathbb{C} : \operatorname{Re} z < 0\} .$$

 Notation:  $\sigma(\mathbf{A}) := \{\lambda : \lambda \text{ ist Eigenwert von } \mathbf{A}\} \hat{=} \text{Spektrum einer Matrix}$

Hilfsmittel: Matrixexponentialfunktion (1.3.14), **Jordan-Normalform** von  $\mathbf{A} \in \mathbb{C}^{d,d}$ :

$$\exists \mathbf{S} \in \mathbb{C}^{d,d} \text{ regulär: } \mathbf{S}^{-1} \mathbf{A} \mathbf{S} = \operatorname{diag}(\mathbf{J}_1, \dots, \mathbf{J}_m) ,$$

mit **Jordan-Blöcken** der Form ( $\lambda \in \sigma(\mathbf{A})$ )

$$\mathbf{J}_k = \begin{pmatrix} \lambda & 1 & 0 & \dots & \dots & 0 \\ 0 & \lambda & 1 & 0 & & \vdots \\ & & \ddots & \ddots & & \\ & & & & \lambda & 1 \\ & & & & & \lambda \end{pmatrix} = \lambda \mathbf{I} + \mathbf{N}_k \in \mathbb{C}^{d_k, d_k}, \quad d_k \in \{1, \dots, d\} .$$

$\mathbf{N}_k \in \mathbb{C}^{d_k, d_k}$  sind **nilpotente Matrizen**:  $\mathbf{N}_k^{d_k} = 0$

Wegen (1.3.15) genügt es  $\exp(\mathbf{J})$  für einen generischen Jordan-Block  $\mathbf{J} \in \mathbb{C}^{n, n}$  zu betrachten.

Verwende:  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n, n}$ :  $\mathbf{AB} = \mathbf{BA} \Rightarrow \exp(\mathbf{A} + \mathbf{B}) = \exp(\mathbf{A}) \cdot \exp(\mathbf{B})$ . (3.2.5)

$$\blacktriangleright \exp(t\mathbf{J}) = \exp(t\lambda\mathbf{I} + \lambda\mathbf{N}_k) = \exp(t\lambda\mathbf{I}) \exp(t\mathbf{N}_k) = e^{\lambda t} \exp(t\mathbf{N}_k).$$

Beachte:  $\exp(t\mathbf{N})$  ist ein *Polynom* in  $t$ , wenn  $\mathbf{N}$  nilpotent (Exponentialreihe bricht nach endlich vielen Gliedern ab). Also finden wir

$$\exp(\mathbf{A}t) = \mathbf{S} \exp(\mathbf{D}t) \mathbf{P}(t) \mathbf{S}^{-1}, \quad \mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_d),$$

mit einem Matrixpolynom  $\mathbf{P}$  vom Grad  $< d$ ,  $\lambda_1, \dots, \lambda_d \hat{=}$  Eigenwerte von  $\mathbf{A}$  (mit Vielfachheit gezählt).

$$\blacktriangleright \|\exp(\mathbf{A}t)\| \leq \|\mathbf{S}\| \|\mathbf{S}^{-1}\| \|\exp(\mathbf{D}t)\| \cdot \|\text{Matrixpolynom in } t\|,$$

wobei  $\mathbf{D}$  die Diagonalmatrix der Eigenwerte von  $\mathbf{A}$  ist. Wegen

$$\forall \lambda \in \mathbb{R}: \forall \beta > \lambda: \forall p \in \mathcal{P}_n: \exists C = C(\lambda, \beta, p): e^{\lambda t} p(t) \leq C e^{\beta t} \quad \forall t \in \mathbb{R}.$$

schliessen wir:  $\|\exp(\mathbf{A}t)\| \leq C e^{\beta t}$  für jedes  $\beta > \max\{\text{Re } \sigma(\mathbf{A})\}$ .

*Beweis.* (von Thm.3.2.4)  $\rightarrow$  [8, Satz 3.30], O.B.d.A.  $\mathbf{y}^* = 0$

Linearisierung, siehe Bem. 1.3.19:

$$\mathbf{f}(\mathbf{y}) = D\mathbf{f}(0)\mathbf{y} + r(\mathbf{y}), \quad \|r(\mathbf{y})\| = o(\|\mathbf{y}\|) \quad \text{für } \mathbf{y} \rightarrow 0.$$

$\mathbf{y}(t) \hat{=}$  Lösung des AWP zu Anfangswert  $\mathbf{y}_0$ ,  $t \in J(\mathbf{y}_0)$ . Variation-der-Konstanten-Formel, siehe Sect. 1.3.2:

$$\mathbf{y}(t) = \exp(D\mathbf{f}(0)t)\mathbf{y}_0 + \int_0^t \exp(D\mathbf{f}(0)(t-\tau))r(\mathbf{y}(\tau)) \, d\tau. \quad (3.2.6)$$

Mit Hilfe der Jordan-Normalform, siehe oben:

$$\forall \beta \in ] \underbrace{\max\{\operatorname{Re} \lambda : \lambda \in \sigma(D\mathbf{f}(0))\}}_{<0!}, 0[: \quad \exists C = C(\beta) > 0: \quad \|\exp(D\mathbf{f}(0)t)\| \leq C e^{\beta t} \quad \forall t \in \mathbb{R}.$$

Fixiere ein geeignetes  $\beta < 0$  und  $C > 0$ . Dazu gibt es  $\epsilon > 0$ :  $\|r(\mathbf{y})\| \leq \frac{|\beta|}{2C} \|\mathbf{y}\|$ , wenn  $\|\mathbf{y}\| < \epsilon$

Annahme:  $\|\mathbf{y}(t)\| < \epsilon$  für  $0 < t < \delta$ . Damit für  $0 \leq t < \delta$  aus (3.2.6)

$$\|\mathbf{y}(t)\| \leq C e^{\beta t} \|\mathbf{y}_0\| + \frac{|\beta|}{2} \int_0^t e^{\beta(t-\tau)} \|\mathbf{y}(\tau)\| \, d\tau,$$

$$\blacktriangleright e^{|\beta|t} \|\mathbf{y}(t)\| \leq C \|\mathbf{y}_0\| + \frac{|\beta|}{2} \int_0^t e^{|\beta|\tau} \|\mathbf{y}(\tau)\| \, d\tau$$

Benutze: **Gronwalls Lemma** (Lemma 1.3.29) für  $u(t) := e^{|\beta|t} \|\mathbf{y}(t)\|$

$$\blacktriangleright \|\mathbf{y}(t)\| \leq C \|\mathbf{y}_0\| \exp\left(-\frac{|\beta|}{2}t\right) \quad \forall 0 \leq t < \delta. \quad (3.2.7)$$



Nun sieht man, dass die Annahme  $\|\mathbf{y}(t)\| < \epsilon$  erfüllt ist, wenn  $\|\mathbf{y}_0\| < \frac{\epsilon}{\max\{C, 1\}}$ .

Unter dieser Bedingung gilt (3.2.7) für alle  $t \geq 0$  und auch  $\mathbb{R}_0^+ \subset J(\mathbf{y}_0)$  mit Thm. 1.3.4.  $\square$

Asymptotische Stabilität eines Fixpunktes  $\mathbf{y}^*$  folgt aus der asymptotischen Stabilität des Fixpunktes  $\mathbf{y}^*$  der **um  $\mathbf{y}^*$  linearisierten ODE**

$$\dot{\mathbf{y}} = D\mathbf{f}(\mathbf{y}^*)(\mathbf{y} - \mathbf{y}^*) . \quad (3.2.8)$$

Dies bestätigt die die Modellproblemanalyse von Sect. 3.1 motivierende Intuition, dass das Verhalten von Lösungen einer ODE in einer Umgebung eines Fixpunktes durch das Verhalten der Lösungen der um den Fixpunkt linearisierten ODE qualitativ richtig beschrieben wird.

### 3.2.2 Attraktive Fixpunkte von Einschrittverfahren

Wir betrachten weiterhin ein autonomes AWP  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ ,  $\mathbf{f} \in C^1(D, \mathbb{R}^d)$ ,  $D \subset \mathbb{R}^d$  offen.

Ferner sei  $\mathbf{y}^*$  ein Fixpunkt ( $\rightarrow$  Def. 3.2.1):  $\mathbf{f}(\mathbf{y}^*) = 0$ .

Betrachte: (Konsistentes) RK-ESV für autonome ODE  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$  ( $\rightarrow$  Def. 2.3.5)

$$\mathbf{k}_i := \mathbf{f}\left(\mathbf{y} + h \sum_{j=1}^s a_{ij} \mathbf{k}_j\right), \quad i = 1, \dots, s, \quad \Psi^h \mathbf{y} := \mathbf{y} + h \sum_{i=1}^s b_i \mathbf{k}_i. \quad (3.2.9)$$

Hinreichend:  $\sum_{i=1}^s b_i = 1$ , Lemma 2.3.23

Annahme:  $h$  hinreichend klein für Wohldefiniertheit des ESV für  $\mathbf{y}$  „nahe bei“  $\mathbf{y}^*$ ,  $\rightarrow$  Lemma. 2.2.7.

$$\blacktriangleright \quad \Psi^h \mathbf{y}^* = \mathbf{y}^* \quad \forall h \text{ hinreichend klein.} \quad (3.2.10)$$

Betrachte eine mit Hilfe der Abbildung  $\Pi : D \subset \mathbb{R}^d \mapsto \mathbb{R}^d$  rekursiv definierte Folge  $\mathbf{y}_{k+1} = \Pi(\mathbf{y}_k)$ .

(Man sagt auch, dass  $\Pi$  ein diskretes dynamisches System definiert. Offensichtlich stellen alle Einschrittverfahren diskrete dynamische Systeme dar.)

Klar:  $\mathbf{y}^* \in D$  heisst **Fixpunkt** des diskreten dynamischen Systems, falls  $\Pi(\mathbf{y}^*) = \mathbf{y}^*$ .

Auch klar: Definition der asymptotischen Stabilität eines Fixpunkts eines diskreten dynamischen Systems analog zu Def. 3.2.2

R. Hiptmair  
rev 35327,  
24. Juni  
2011

**Theorem 3.2.12** (Asymptotische Stabilität von Fixpunkten diskreter dynamischer Systeme).

Sei  $\Pi : D \subset \mathbb{R}^d \mapsto \mathbb{R}^d$  stetig differenzierbar und  $\Pi(\mathbf{y}^*) = \mathbf{y}^*$  für ein  $\mathbf{y}^* \in D$ . Dann gilt

$$\rho(D\Pi(\mathbf{y}^*)) < 1 \quad \Rightarrow \quad \mathbf{y}^* \text{ ist asymptotisch stabiler Fixpunkt von } \mathbf{y}_{k+1} := \Pi(\mathbf{y}_k) .$$

**Lemma 3.2.13** (Den Spektralradius approximierende Matrixnorm).  $\rightarrow [12, \text{Sect. 2.9.3}]$

Zu jeder Matrix  $\mathbf{A} \in \mathbb{C}^{d,d}$  und jedem  $\epsilon > 0$  gibt es eine Vektornorm  $\|\cdot\|_{A,\epsilon}$  auf  $\mathbb{R}^d$  so, dass für die induzierte Matrixnorm gilt

$$\rho(\mathbf{A}) \leq \|\mathbf{A}\|_{A,\epsilon} \leq \rho(\mathbf{A}) + \epsilon .$$

Der Beweis stützt sich auf die **Schur-Normalform** von  $\mathbf{A}$ .

☞ Für diskrete Evolution  $\Psi^h$ :      Untersuche die **Jacobi-Matrix**  $D_{\mathbf{y}}(\Psi^h \mathbf{y})$  für  $\mathbf{y} = \mathbf{y}^*$  !

Für RK-ESV: 
$$D_{\mathbf{y}}(\Psi^h)(\mathbf{y}^*) = S(hD\mathbf{f}(\mathbf{y}^*)) . \quad (3.2.14)$$

Erinnerung:  $S$  ist die (rationale) Stabilitätsfunktion ( $\rightarrow$  Thm. 3.1.6) des RK-ESV, in (3.2.14) benutzt im Sinne von Bem. 3.1.17.

$$\{\mathbf{f}(\mathbf{y}^*) = 0 \Leftrightarrow \Phi^h \mathbf{y}^* = \mathbf{y}^*\} \Rightarrow \Psi^h \mathbf{y}^* = \mathbf{y}^*$$



$$\mathbf{y}(h) \approx (\mathbf{y}_0 - \mathbf{y}^*) \exp(D\mathbf{f}(\mathbf{y}^*)h) + \mathbf{y}^* \Leftrightarrow \Psi^h \mathbf{y} \approx (\mathbf{y}_0 - \mathbf{y}^*) S(hD\mathbf{f}(\mathbf{y}^*)) + \mathbf{y}^*$$



**Theorem 3.2.15** (Vererbung asymptotischer Stabilität).

Ein Fixpunkt  $\mathbf{y}^* \in D$  der diskreten Evolution eines RK-ESV mit Stabilitätsgebiet  $\mathcal{S}_\Psi$  ist asymptotisch stabil ( $\rightarrow$  Def. 3.2.2), wenn

$$h\sigma(D\mathbf{f}(\mathbf{y}^*)) \subset \mathcal{S}_\Psi .$$



Explizite RK-ESV : **Schrittweisenbeschränkung** für Vererbung  
von Stabilität eines Fixpunktes ( $\rightarrow$  (3.1.12))

Für welche Verfahren entfällt Schrittweitenbeschränkung ?

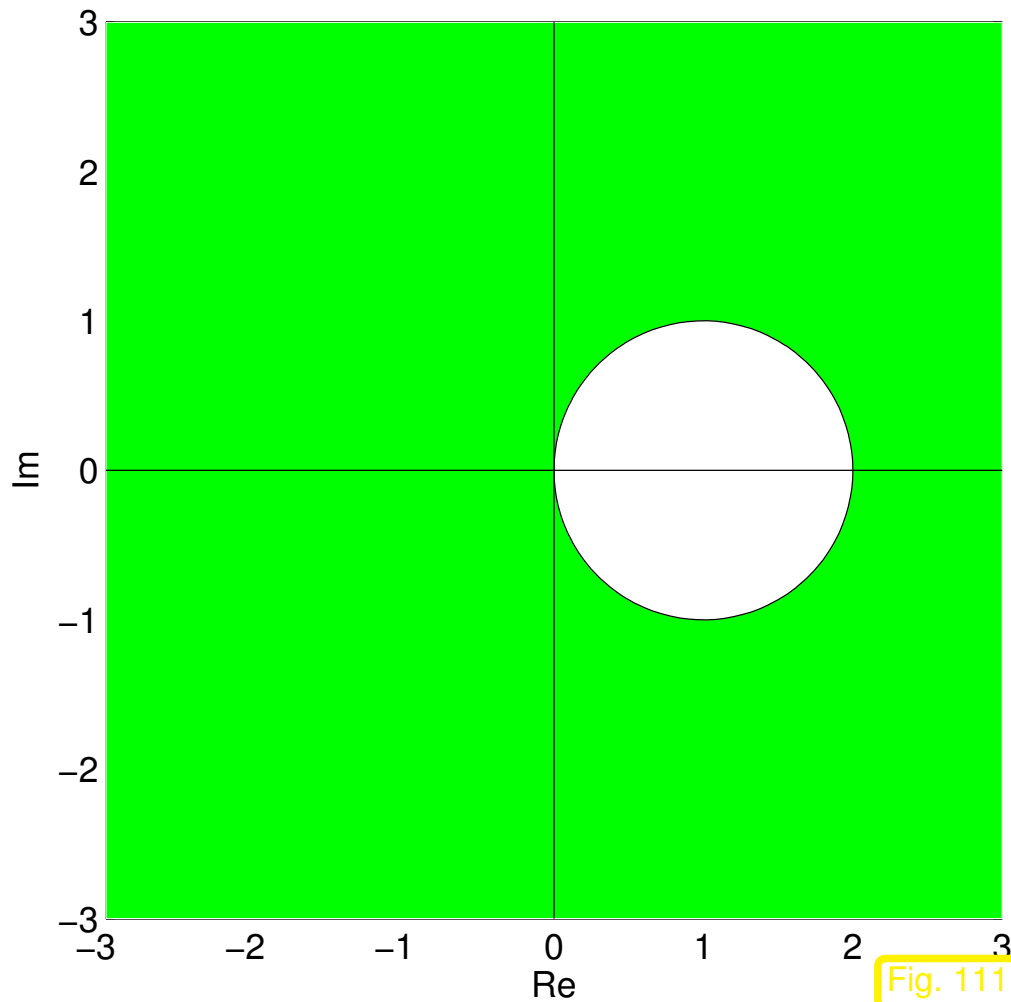
**Definition 3.2.16** (A-stabiles Einschrittverfahren).  $\rightarrow [8, \text{Sect. 6.1.3}]$

$$\text{ESV } \mathbf{A}\text{-stabil} \quad :\Leftrightarrow \quad \mathbb{C}^- := \{z \in \mathbb{C} : \operatorname{Re} z < 0\} \subset \mathcal{S}_\Psi \quad (\mathcal{S}_\Psi = \text{Stabilitätsgebiet,} \\ \rightarrow \text{Def. 3.1.4})$$

Aus Thm. 3.2.15: für A-stabile ESV (mit diskreter Evolution  $\Psi^h$ )

$$\begin{array}{l} \text{Autonome Dgl. } \dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}), \\ \text{Fixpunkt } \mathbf{f}(\mathbf{y}^*) = 0 \end{array} \wedge \begin{array}{l} \sigma(D\mathbf{f}(\mathbf{y}^*)) \subset \mathbb{C}^- \\ (\hat{=} \text{ asymp. Stabilität von } \mathbf{y}^*) \end{array} \Rightarrow \begin{array}{l} \text{Falls } \|\mathbf{y}_0 - \mathbf{y}^*\| < \delta, \\ \text{dann } \lim_{k \rightarrow \infty} (\Psi^h)^k \mathbf{y}_0 = \mathbf{y}^* . \end{array}$$

*Beispiel 3.2.17* (Einfache A-stabile RK-ESV).

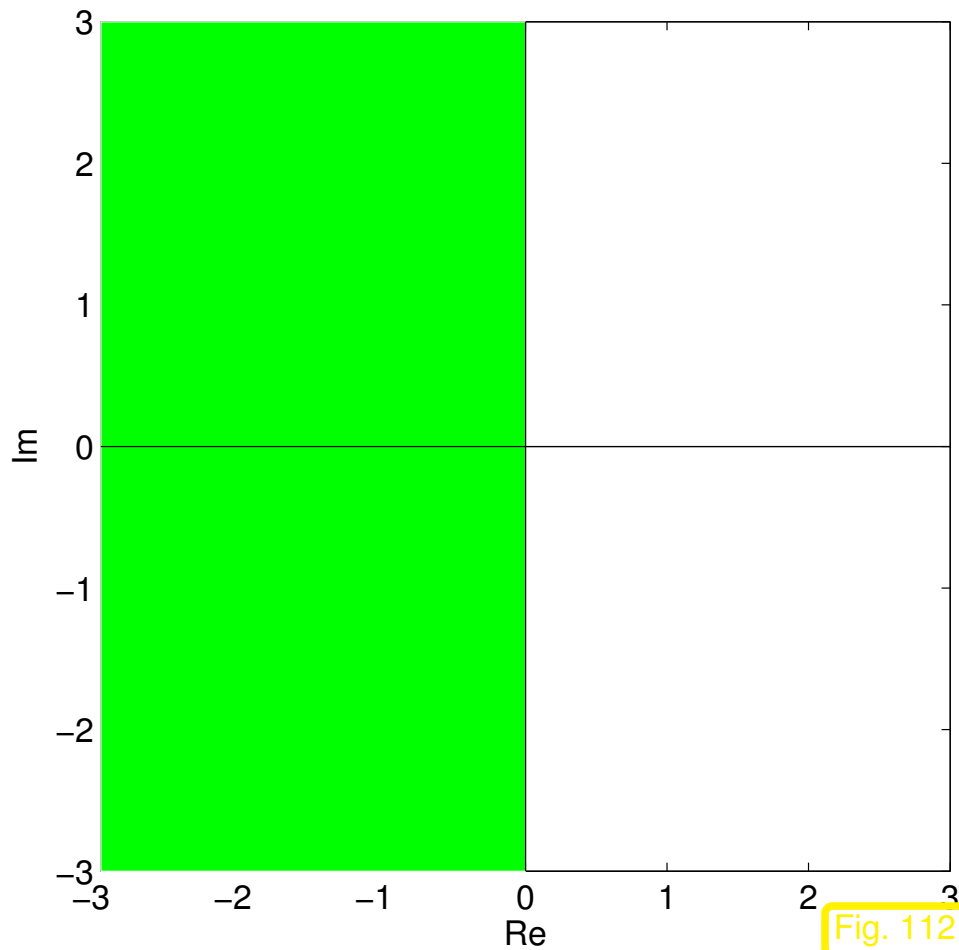


Implizites Eulerverfahren (1.4.13)

Stabilitätsfunktion ( $\rightarrow$  Thm. 3.1.6)

$$S(z) = \frac{1}{1-z}.$$

$\triangleleft$  Stabilitätsgebiet  $\mathcal{S}_\Psi$  ( $\rightarrow$  Def. 3.1.4)



implizite Mittelpunktsregel (1.4.19)

Stabilitätsfunktion ( $\rightarrow$  Thm. 3.1.6)

$$S(z) = \frac{1 + \frac{1}{2}z}{1 - \frac{1}{2}z}.$$

$\triangleleft$  Stabilitätsgebiet  $\mathcal{S}_\Psi$  ( $\rightarrow$  Def. 3.1.4)

Dies ist das “ideale Stabilitätsgebiet”!



### 3.3 Nichtexpansivität [8, Abschn. 6.3.3]

Betrachte: Autonomes AWP  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ ,  $\mathbf{f} \in C^1(D, \mathbb{R}^d)$ ,  $D \subset \mathbb{R}^d$  offen.



Wir fixieren  $\mathbf{M} \in \mathbb{R}^{d,d}$  s.p.d.  $\triangleright$  Norm  $\|\mathbf{y}\|_M := (\mathbf{y}^T \mathbf{M} \mathbf{y})^{1/2}$  auf  $\mathbb{R}^d$ .

**Definition 3.3.1** (Nichtexpansivität).

Eine Evolution  $\Phi^t$  zu einer autonomen Dgl. bzw. eine diskrete Evolution  $\Psi^h$  zu einem zugehörigen Einschrittverfahren heisst *nichtexpansiv*, falls

$$\begin{aligned} \|\Phi^t \mathbf{y} - \Phi^t \tilde{\mathbf{y}}\|_M &\leq \|\mathbf{y} - \tilde{\mathbf{y}}\|_M, \\ \|\Psi^h \mathbf{y} - \Psi^h \tilde{\mathbf{y}}\|_M &\leq \|\mathbf{y} - \tilde{\mathbf{y}}\|_M \end{aligned} \quad \forall \mathbf{y}, \tilde{\mathbf{y}} \in D,$$

und für alle  $t \in J(\mathbf{y}) \cap J(\tilde{\mathbf{y}}) \cap \mathbb{R}_0^+$  und alle „hinreichend kleinen“  $h > 0$ .

*Beispiel 3.3.2* (Gradientenfluss  $\rightarrow$  „Kriechvorgänge“).

Gegeben:  $C^1$ -Potential  $V : \mathbb{R}^d \mapsto \mathbb{R}$  **konvex**

Erinnerung: Eine Abbildung  $V : \mathbb{R}^d \mapsto \mathbb{R}$  heisst **konvex**, falls

$$V(\xi \mathbf{x} + (1 - \xi) \mathbf{y}) \leq \xi V(\mathbf{x}) + (1 - \xi) V(\mathbf{y}) \quad \forall 0 \leq \xi \leq 1. \quad (3.3.3)$$

Erinnerung: Eine  $C^1$ -Funktion  $\varphi : \mathbb{R} \mapsto \mathbb{R}$  is genau dann konvex, wenn  $\varphi'$  monoton steigt.

Offensichtliche Konsequenz aus (3.3.3): Ist  $V : \mathbb{R}^d \mapsto \mathbb{R}$  konvex, so gilt das für jeden "Schnitt"  
 $\tau \mapsto V(\mathbf{y} + \tau(\mathbf{x} - \mathbf{y}))$ ,  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ .



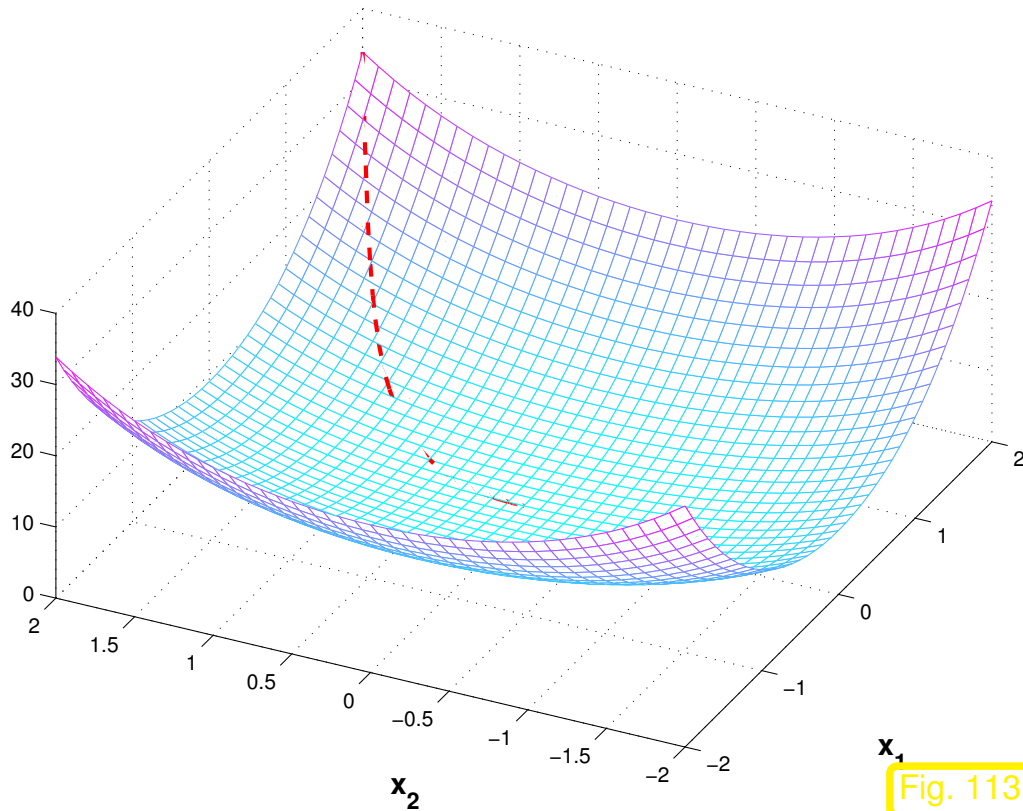
$C^1$ -Potential  $V : \mathbb{R}^d \mapsto \mathbb{R}$  konvex



$\frac{d}{d\tau} \varphi$  monoton steigend für  $\varphi(\tau) := V(\mathbf{y} + \tau(\mathbf{x} - \mathbf{y})) \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$



$(\mathbf{grad} V(\mathbf{x}) - \mathbf{grad} V(\mathbf{y}))^T (\mathbf{x} - \mathbf{y}) \geq 0 \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ .



Gradientenfluss-AWP:

$$\begin{aligned} \dot{\mathbf{y}}(t) &= -\mathbf{grad} V(\mathbf{y}(t)) , \\ \mathbf{y}(0) &= \mathbf{y}_0 \in \mathbb{R}^d . \end{aligned} \quad (3.3.4)$$

Fig. 113

Die Evolution zu (3.3.4) ist nichtexpansiv bzgl. der Euklidischen Norm:

$$\chi(t) := \left\| \Phi^t \mathbf{y} - \Phi^t \tilde{\mathbf{y}} \right\|_2^2 \Rightarrow \dot{\chi}(t) = -2 \underbrace{(\mathbf{grad} V(\Phi^t \mathbf{y}) - \mathbf{grad} V(\Phi^t \tilde{\mathbf{y}}))^T (\Phi^t \mathbf{y} - \Phi^t \tilde{\mathbf{y}})}_{\geq 0} \leq 0 .$$

➤ Nichtexpansivität (→ Def. 3.3.1) mit  $\mathbf{M} = \mathbf{I}$ .

**Definition 3.3.5** (Dissipatives Vektorfeld).

$$\mathbf{f} : D \subset \mathbb{R}^d \mapsto \mathbb{R}^d \text{ dissipativ} \quad :\Leftrightarrow \quad \mathbf{M}(\mathbf{f}(\mathbf{y}) - \mathbf{f}(\tilde{\mathbf{y}})) \cdot (\mathbf{y} - \tilde{\mathbf{y}}) \leq 0 \quad \forall \mathbf{y}, \tilde{\mathbf{y}} \in D .$$

Dies ist eine Verallgemeinerung der Eigenschaft „monoton fallend“ von skalarwertigen Funktionen.

**Lemma 3.3.6** (Bedingung für Nichtexpansivität einer Evolution).

$$\text{Rechte Seite } \mathbf{f} \text{ dissipativ} \quad \Leftrightarrow \quad \text{Nichtexpansivität der Evolution} \\ \text{zur autonomen ODE } \dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$$

**Theorem 3.3.7** (Gauss-Kollokations-RK-ESV nichtexpansiv).

Die diskreten Evolutionen zu Gauss-Kollokations-RK-ESV ( $\rightarrow$  Sect. 2.2.3) erben die Nichtexpansivität der (exakten) Evolution.

*Beweis von Thm. 3.3.7.* Betrachte Gauss-Kollokations-ESV mit  $s$  Knoten:

$\mathbf{y}_h(t), \widehat{\mathbf{y}}_h(t) \in \mathcal{P}_s \hat{=}$  Kollokationspolynome zu Anfangswerten  $\mathbf{y}_0$  bzw.  $\widetilde{\mathbf{y}}_0$ , siehe Sect. 2.2.1

$$\blacktriangleright \quad \Psi^h \mathbf{y}_0 = \mathbf{y}_h(h) \quad , \quad \Psi^h \widetilde{\mathbf{y}}_0 = \widetilde{\mathbf{y}}_h(h) .$$

$d(\tau) := \|\mathbf{y}_h(\tau h) - \widetilde{\mathbf{y}}_h(\tau h)\|_M^2$  ist Polynom in  $\tau$  vom Grad  $\leq 2s$  .

Nichtexpansivität von  $\Psi^h$  is äquivalent zu

$$\left\| \Psi^h \mathbf{y}_0 - \Psi^h \widetilde{\mathbf{y}}_0 \right\|_M^2 = d(1) = d(0) + \underbrace{\int_0^1 d'(\tau) d\tau}_{\text{Ziel } \leq 0} = \|\mathbf{y}_0 - \widetilde{\mathbf{y}}_0\|_M^2 + \int_0^1 d'(\tau) d\tau . \quad (3.3.8)$$

Gauss-Quadratur (mit  $s$  Knoten) ist exakt für Polynome  $\in \mathcal{P}_{2s-1}$

$$\blacktriangleright \quad \int_0^1 d'(\tau) d\tau = \sum_{j=1}^s b_j d'(c_j) . \quad (3.3.9)$$

Ableitung aus der Kettenregel:

$$d'(\tau) = 2h\mathbf{M}(\mathbf{y}_h(\tau h) - \widetilde{\mathbf{y}}_h(\tau h)) \cdot (\dot{\mathbf{y}}_h(\tau h) - \dot{\widetilde{\mathbf{y}}}_h(\tau h)) . \quad (3.3.10)$$

Aus Kollokationsbedingungen (2.2.1):

$$\dot{\mathbf{y}}_h(c_j h) = \mathbf{f}(\mathbf{y}_h(c_j h)) \quad , \quad \dot{\tilde{\mathbf{y}}}_h(c_j h) = \mathbf{f}(\tilde{\mathbf{y}}_h(c_j h)) \quad , \quad j = 1, \dots, s .$$

(3.3.10)



$$d'(c_j) = 2h\mathbf{M}(\mathbf{y}_h(c_j h) - \tilde{\mathbf{y}}_h(c_j h)) \cdot (\mathbf{f}(\mathbf{y}_h(c_j h)) - \mathbf{f}(\tilde{\mathbf{y}}_h(c_j h))) \leq 0 \quad , \quad (3.3.11)$$

da  $\mathbf{f}$  dissipativ  $\Leftrightarrow$  Nichtexpansivität von  $\Phi^t$ , vgl Lemma 3.3.6.

(3.3.8), (3.3.9), (3.3.11)  $\Rightarrow$  Behauptung, da Gewichte  $b_j$  der Gauss-Quadraturformeln positiv !.  $\square$

**Lemma 3.3.12** (Diskrete Nichtexpansivität  $\Rightarrow$  A-Stabilität).

*Nichtexpansivität ( $\rightarrow$  Def. 3.3.1) erbende RK-ESV ( $\rightarrow$  Def. 2.3.5) sind A-stabil ( $\rightarrow$  Def. 3.2.16).*

*Beweis.* Skalare komplexe Dlg.  $\Leftrightarrow$  reelle Dlg. in  $D = \mathbb{R}^2$ : für beliebiges  $\lambda = \alpha + i\beta \in \mathbb{C}$

$$\dot{y} = \lambda y \quad \stackrel{y=u+iv}{\Leftrightarrow} \quad \begin{pmatrix} \dot{u} \\ \dot{v} \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}}_{=: \mathbf{A}} \begin{pmatrix} u \\ v \end{pmatrix} \quad \Leftrightarrow \quad \dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}) \quad \text{mit } \mathbf{y} := \begin{pmatrix} u \\ v \end{pmatrix} .$$

$\operatorname{Re} \lambda < 0 \Rightarrow \alpha < 0 \Rightarrow \mathbf{x}^T \mathbf{A} \mathbf{x} = \alpha \|\mathbf{x}\|_2 \leq 0 \quad \forall \mathbf{x} \in \mathbb{R}^2$   
 $\Rightarrow$  rechte Seite  $\mathbf{f}(\mathbf{y}) = \mathbf{A} \mathbf{y}$  ist dissipativ ( $\rightarrow$  Def. 3.3.5)  
 $\Rightarrow$  Evolution nichtexpansiv, siehe Lemma 3.3.6.

(“Vererbung”)  $\blacktriangleright$  Diskrete Evolution  $\Psi^h$  nichtexpansiv

$$\Rightarrow \left\| \Psi^h \mathbf{y} \right\|_2 = |S(h\lambda)| \|y\| \leq \|\mathbf{y}\|_2 = |y| \Rightarrow |S(z)| \leq 1 \quad \forall z \in \overline{\mathbb{C}^-}.$$

$$\stackrel{*}{\Rightarrow} |S(z)| < 1 \quad \forall z \in \mathbb{C}^-.$$

$$\Rightarrow \mathbb{C}^- \subset \mathcal{S}_\Psi \quad (\text{Stabilitätsgebiet} \rightarrow \text{Def. 3.1.4}). \quad \square$$

\*:  $S(z)$  is a meromorphic function so that  $|S(z)|$  can attain its maximal value on  $\overline{\mathbb{C}^-}$  only on the boundary  $\partial\mathbb{C} = i\mathbb{R}$ .

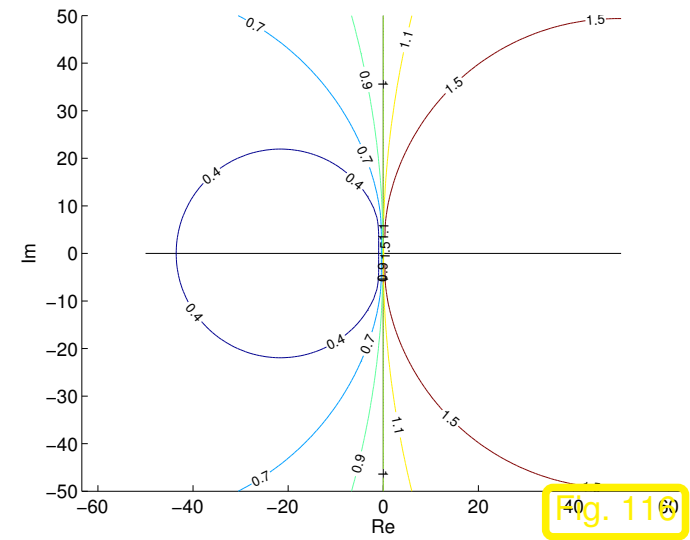
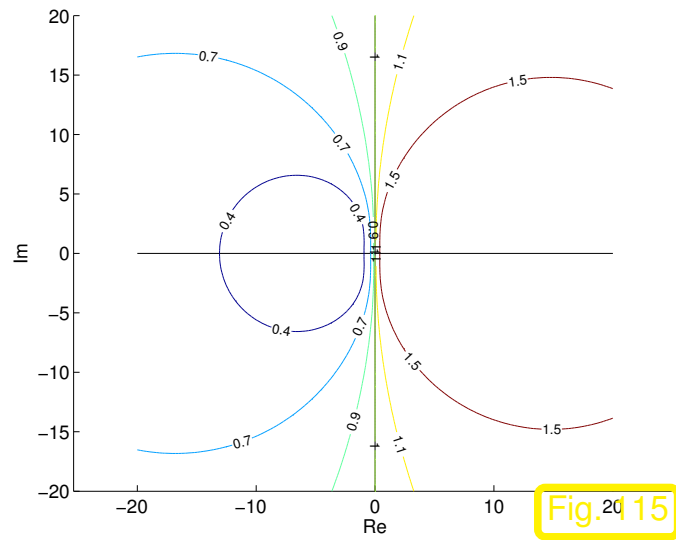
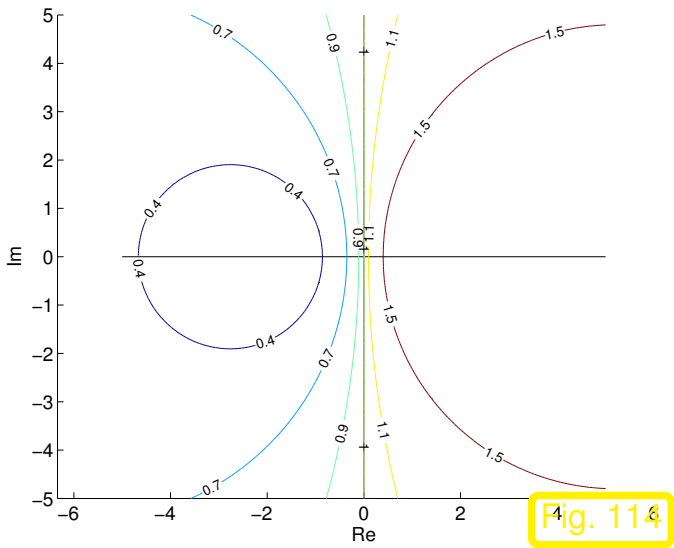
Alle Gaus-Kollokations-ESV sind A-stabil

*Bemerkung 3.3.13* (Lösbarkeit der Inkrementgleichungen für Gauss-Kollokations-ESV).

Die Inkrementgleichungen eines Gauss-Kollokations-RK-ESV für eine nichtexpansive autonome ODE sind für jedes  $h > 0$  eindeutig lösbar  $\rightarrow$ [18, Sect. IV.14].



# Stabilitätsgebiete von Gauss-Kollokations-Einschrittverfahren:



Implizite Mittelpunktsregel

$s = 2$  (Ordnung 4)

$s = 4$  (Ordnung 8)

Niveaulinien von  $|S(z)|$  für Gauss-Kollokations-Einschrittverfahren

Vermutung (Beweis später):

$$\mathcal{S}_\Psi = \mathbb{C}^-$$

Beispiel 3.3.14 (Gauss-Kollokationsverfahren für logistische Differentialgleichung). → Bsp. 3.0.1, 1.4.21



Logistische Differentialgleichung  $\dot{y} = f(y)$ ,  
 $f(y) = \lambda y(1 - y) \rightarrow (2.2.84)$ ,  $\lambda = 50$ , Anfangs-  
wert  $y_0 = 10^{-4}$ , Zeitintervall  $[0, 1]$ .

Kollokations-Einschrittverfahren ( $\rightarrow$  Ab-  
schnitt 2.2) auf äquidistantem Gitter,  $h = \frac{1}{20}$ .

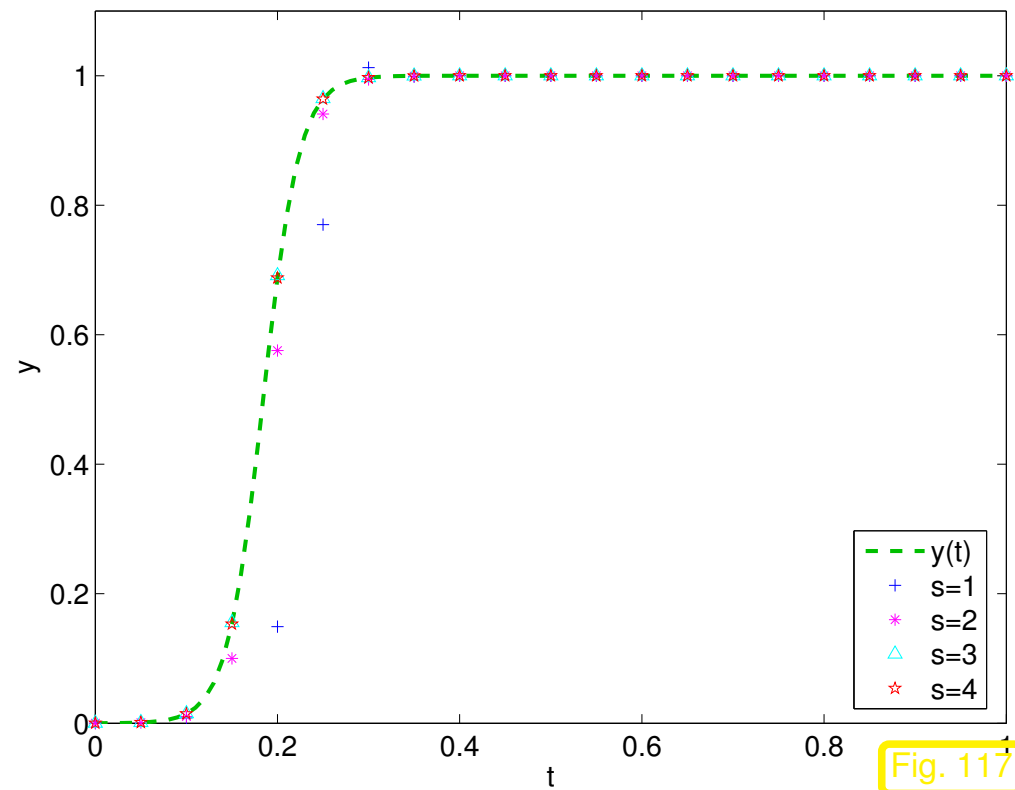


Fig. 117



Bemerkung 3.3.15 (A-Stabilität  $\not\Rightarrow$  Diskrete Nichtexpansivität).

Gegenbeispiel: **implizite Trapezregel**, Einschrittverfahren für  $\dot{y} = f(t, y)$  definiert durch

$$y_1 = y_0 + \frac{1}{2}h(f(t, y_0) + f(t + h, y_1)) \leftrightarrow \text{Butcher-Schema } \begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1/2 & 1/2 \\ \hline & 1/2 & 1/2 \end{array}$$

angewandt auf skalare autonome ODE  $\dot{y} = \begin{cases} -y^3 & \text{für } y < 0, \\ -y^2 & \text{für } y \geq 0. \end{cases} \rightarrow \text{Übungsaufgabe}$

*Bemerkung 3.3.16 (B-Stabilität).*

Einschrittverfahren, die die Nichtexpansivität der Evolution zu einer ODE erben, heissen auch **B-stabil** [18, Sect. IV.12].



Ein algebraisches Kriterium für B-Stabilität:

**Definition 3.3.17** (Algebraische Stabilität).

Ein Runge-Kutta-Einschrittverfahren ( $\rightarrow$  Def. 2.3.5) mit Butcher-Schema  $\begin{array}{c|c} \mathbf{c} & \mathfrak{A} \\ \hline & \mathbf{b}^T \end{array}$ , siehe (2.3.6), ist **algebraisch stabil**, falls

(i)  $b_i \geq 0, i = 1, \dots, s,$

(ii) und die Matrix  $\mathbf{M} := \text{diag}(b_1, \dots, b_s)\mathfrak{A} - \mathfrak{A}^T \text{diag}(b_1, \dots, b_s) - \mathbf{b}\mathbf{b}^T$  positiv semi-definit ist.

**Theorem 3.3.18** (Kriterium für B-Stabilität).

$$\textit{Algebraische Stabilität} \Rightarrow \textit{B-Stabilität}$$

## 3.4 Gleichmässige Stabilität

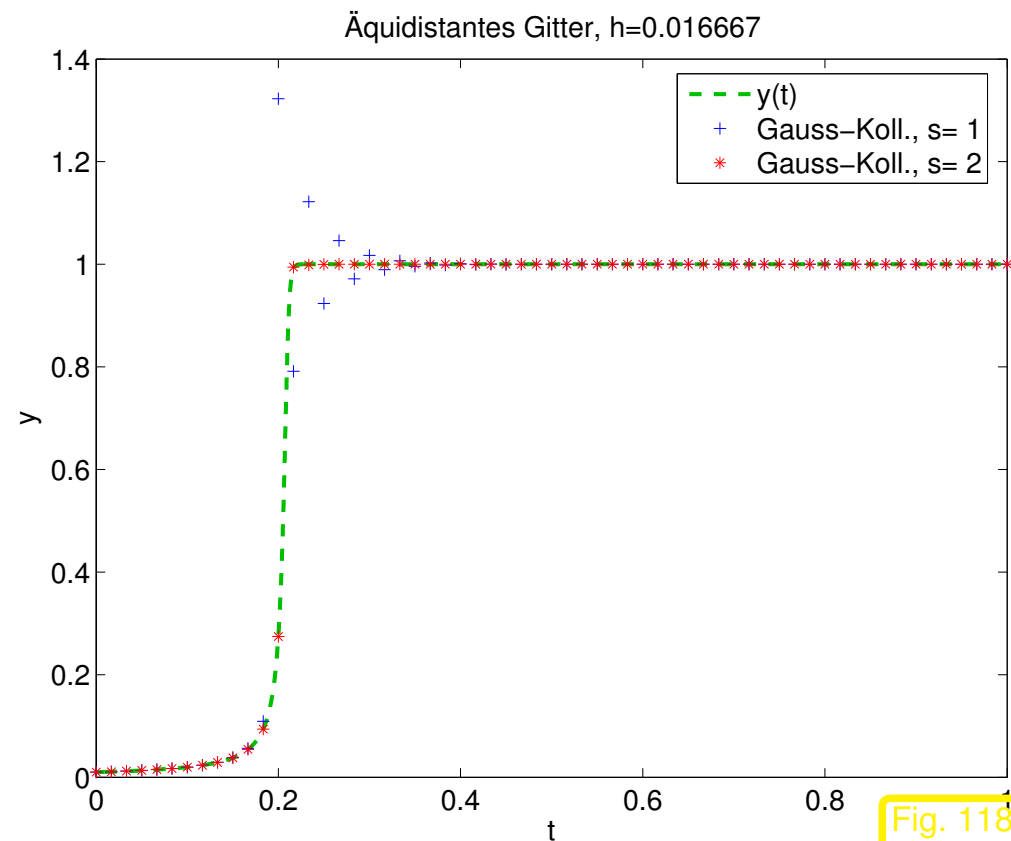
*Beispiel 3.4.1* (Gauss-Kollokations-ESV bei stark attraktiven Fixpunkten). → Bsp. 3.5.2

ODE mit stark attraktivem Fixpunkt  $y = 1$ :

$$\dot{y} = \lambda y^2(1 - y),$$

$$\lambda = 500, \quad y(0) = \frac{1}{100}.$$

Qualitatives Verhalten von  
Gauss-Kollokations-ESV ( $\rightarrow$  Abschnitt 2.2)  
auf äquidistantem Gitter  $\triangleright$



➤ Falsche Oszillationen bei Gauss-Kollokations-ESV niedriger Ordnung

Erklärung: Für Gauss-Kollokations-ESV gilt  $S(z) \approx \pm 1$  für  $|z| \rightarrow \infty$ , so dass der Fixpunkt der diskreten Evolution zwar anziehend bleibt, aber die diskrete Lösung (im Gegensatz zur kontinuierlichen) nur noch langsam (und oszillatorisch für ungerades  $s$ ) gegen ihn konvergiert.

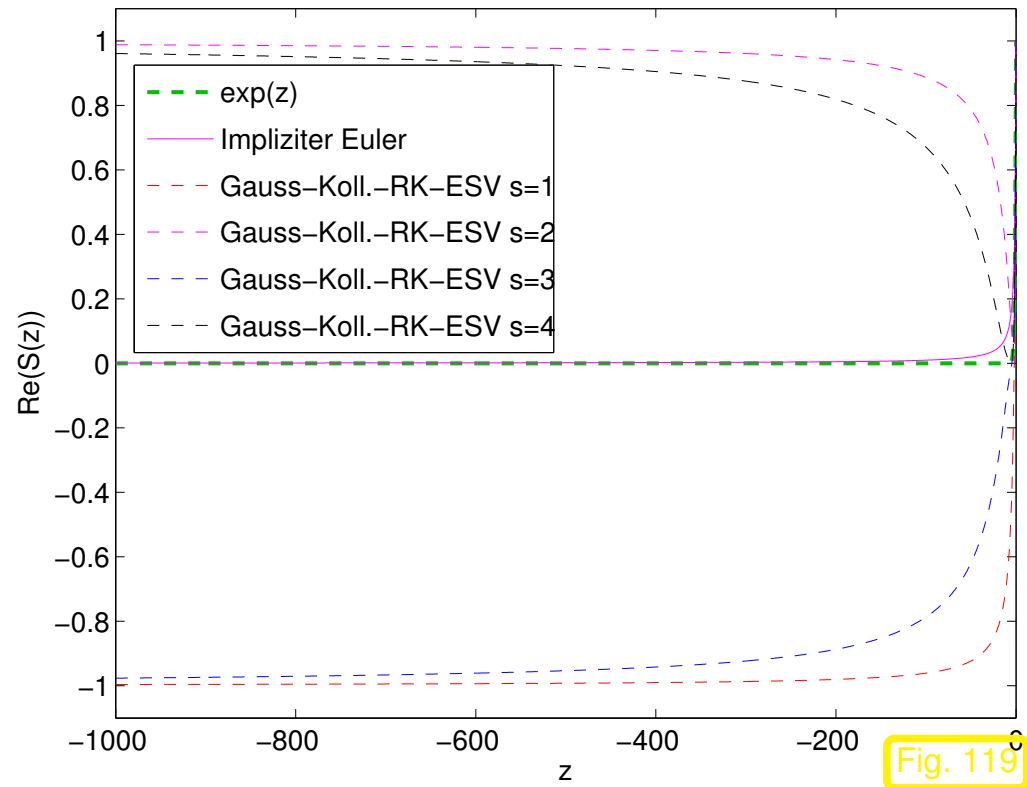
Weitere Demonstration  $\rightarrow$  Bsp. 3.4.2



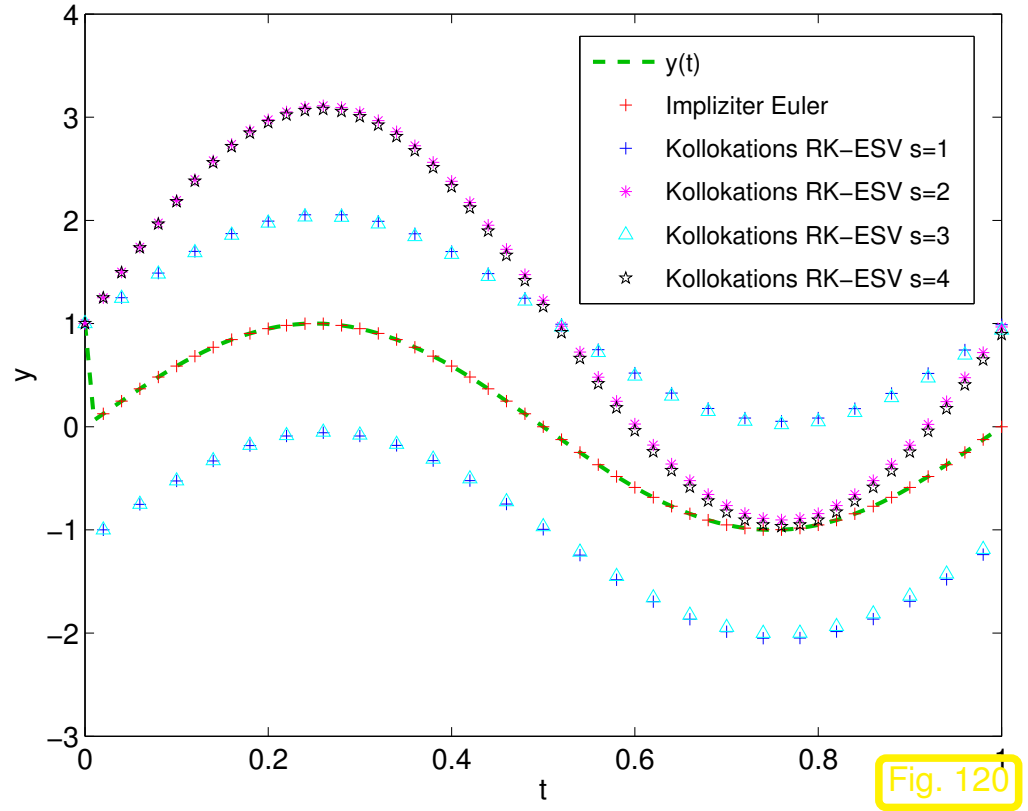
Beispiel 3.4.2 (Implizite RK-ESV bei schnellen Transienten). → Bsp. 3.5.5

AWP:  $\dot{y} = -\lambda y + \beta \sin(2\pi t)$ ,  $\lambda = 10^6$ ,  $\beta = 10^6$ ,  $y(0) = 1$ .

RK-ESV, äquidistantes Gitter auf  $[0, 1]$ ,  $h = \frac{1}{40}$ :



Stabilitätsfunktionen für  $Re z \ll 1$



Diskrete Evolutionen (Zeitverlauf)

➤ Ungenügende Dämpfung der Anfangsstörung bei Kollokations-RK-ESV !

(Oszillationen für ungerades  $s$  → vgl. Stabilitätsfunktionen,  $\lim_{\operatorname{Re} z \rightarrow -\infty} S(z) = (-1)^s$ )

➤ Implizites Euler-Verfahren (1.4.13): *sofortige* Relaxation der diskreten Lösung !

Klar, denn  $\lim_{\operatorname{Re} z \rightarrow -\infty} S(z) = \lim_{\operatorname{Re} z \rightarrow -\infty} \frac{1}{1-z} = 0$  für implizites Euler-Verfahren.



Sect. 3.1:

Stabilitätsfunktion  $S(z) \longleftrightarrow \exp(z)$

Utopie (für RK-ESV):

$$S(-\infty) = 0 \quad , \quad S(\infty) = \infty$$

(von keiner rationalen Funktion erfüllbar !)

Bescheidener Wunsch (bei stark attraktiven Fixpunkten, schnellen Relaxationen):  $S(-\infty) = 0$

**Definition 3.4.3** (L-Stabilität).

$$ESV \text{ L-stabil} \quad :\Leftrightarrow \quad \{z \in \mathbb{C} : \operatorname{Re} z < 0\} \subset \mathcal{S}_\Psi \quad \& \quad \lim_{\operatorname{Re} z \rightarrow -\infty} |S(z)| = 0$$

Kurz: L-stabil  $\Leftrightarrow$  A-stabil & „ $S(-\infty) = 0$ “

Wie findet man L-stabile RK-ESV ? Existenz ist zu fordern

$$\text{Thm. 3.1.6} \Rightarrow S(-\infty) = 1 - \mathbf{b}^T \mathfrak{A}^{-1} \mathbf{1}. \quad (3.4.4)$$

► Falls  $\mathbf{b}^T = \mathbf{a}_j^T$ . (Zeile von  $\mathfrak{A}$ )  $\wedge$   $\mathfrak{A}$  regulär  $\Rightarrow S(-\infty) = 0$ . (3.4.5)

charakterisiert **steif-genaue** (engl. *stiffly accurate*) RK-ESV [8, Lemma 6.32]

*Bemerkung 3.4.6* (Invertierbarkeit der Koeffizientenmatrix von RK-ESV).

Für jedes  $s$ -stufige Kollokationsverfahren ( $\rightarrow$  Sect. 2.2.1) mit  $c_s > 0$  (.d.h, für jedes Kollokationsverfahren mit Ausnahme des expliziten Eulerverfahrens (1.4.2)) ist die Koeffizientenmatrix (Butcher-Matrix)  $\mathfrak{A}$  nichtsingulär

*Beweis.* Es sei  $\mathbf{x} \in \mathbb{R}^s$  mit  $\mathfrak{A}\mathbf{x} = 0$

$$(2.2.3) \quad \Rightarrow \quad \sum_{j=1}^s a_{ij} x_j = \sum_{j=1}^s \int_0^{c_i} x_j L_j(\tau) d\tau = 0, \quad i = 1, \dots, s,$$

mit den Lagrange-Polynomen  $L_i \in \mathcal{P}_{s-1}$  aus (2.2.2).

$$\blacktriangleright \quad q := \sum_{j=1}^s x_j L_j \quad \text{erfüllt:} \quad \int_{c_{i-1}}^{c_i} q(\tau) d\tau = 0, \quad i = 1, \dots, s \quad (c_0 := 0).$$

$\Rightarrow q \in \mathcal{P}_{s-1}$  hat  $s$  Nullstellen in  $[0, c_s] \Rightarrow q = 0 \Rightarrow \mathbf{x} = 0.$

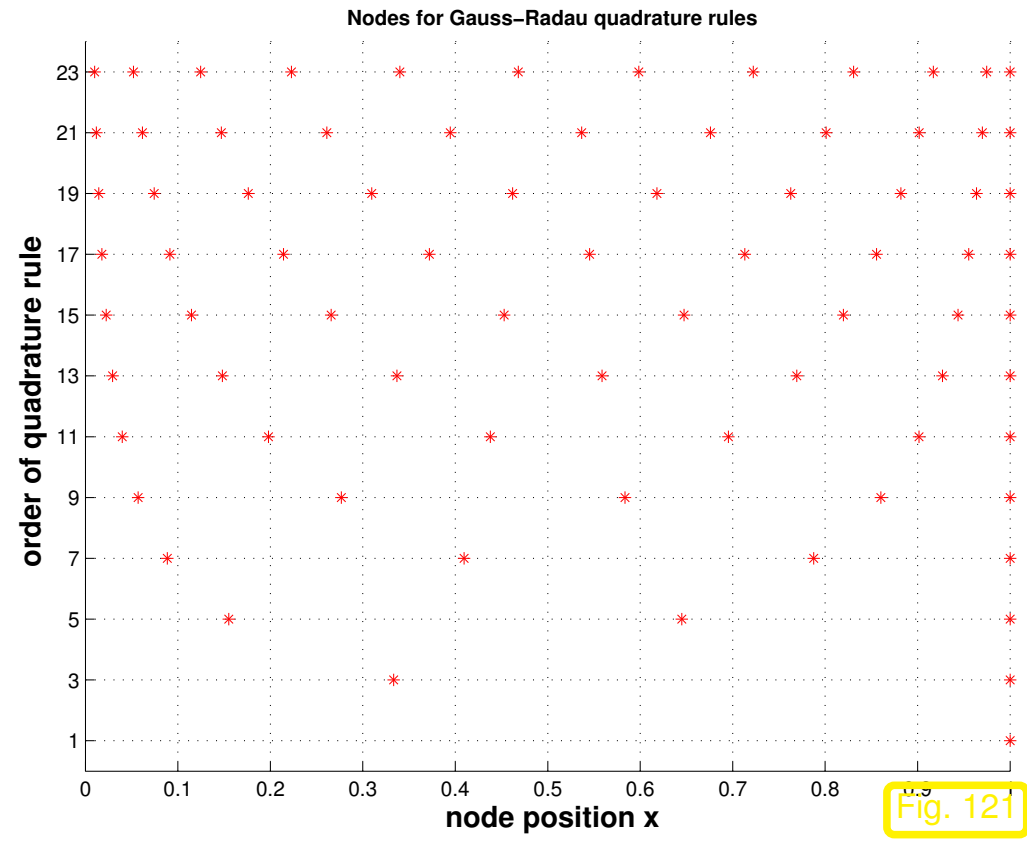




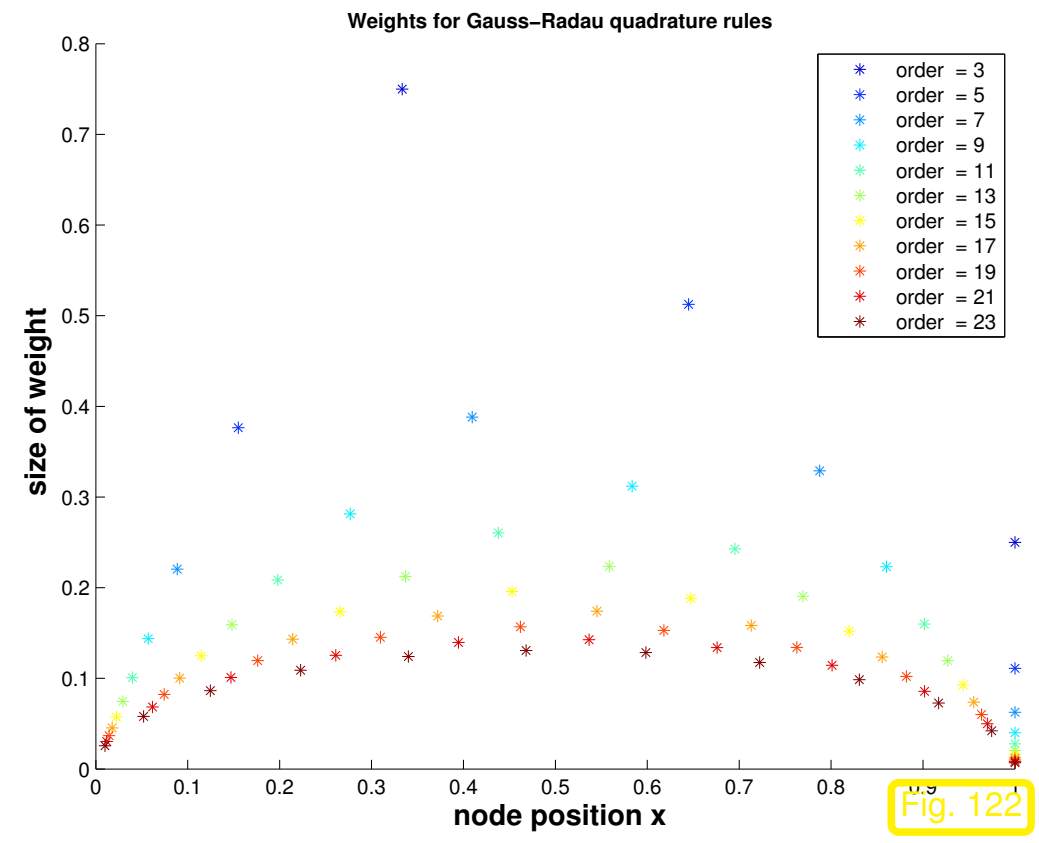
Butcher-Schema (2.3.6) für konsistente  
( $\rightarrow$  Lemma 2.3.23), L-stabile RK-ESV,  
siehe Def. 3.4.3

$$\triangleright \begin{array}{c|c} \mathbf{c} & \mathcal{A} \\ \hline \mathbf{b}^T & \end{array} := \begin{array}{c|ccc} c_1 & a_{11} & \cdots & a_{1s} \\ \vdots & \vdots & & \vdots \\ c_{s-1} & a_{s-1,1} & \cdots & a_{s-1,s} \\ \hline 1 & b_1 & \cdots & b_s \\ \hline & b_1 & \cdots & b_s \end{array} .$$

- Idee:
- Wähle  $c_s = 1$  im Kollokations-RK-ESV (2.2.3)
  - Wähle  $c_1, \dots, c_{s-1}$  als Knoten einer Quadraturformel maximaler Ordnung.  
( $\rightarrow$  **Gauss-Radau-Quadratur**, Ordnung  $2s - 1$ )



Knoten: Radau-Quadraturformeln



Gewichte: Radau-Quadraturformeln

Implizite  $s$ -stufige L-stabile **Radau-ESV**, Konvergenzordnung  $2s - 1$   
 ( $\rightarrow$  Thm. 2.2.51, [8, Sect. 6.3.2])

$$\frac{1}{1} \Big| \frac{1}{1}$$

Implizites Euler-ESV

$$\begin{array}{c|cc} \frac{1}{3} & \frac{5}{12} & -\frac{1}{12} \\ 1 & \frac{3}{4} & \frac{1}{4} \\ \hline & \frac{3}{4} & \frac{1}{4} \end{array}$$

Radau-ESV, Ordnung 3

$$\begin{array}{c|ccc} \frac{4-\sqrt{6}}{10} & \frac{88-7\sqrt{6}}{360} & \frac{296-169\sqrt{6}}{1800} & \frac{-2+3\sqrt{6}}{225} \\ \frac{4+\sqrt{6}}{10} & \frac{296+169\sqrt{6}}{1800} & \frac{88+7\sqrt{6}}{360} & \frac{-2-3\sqrt{6}}{225} \\ 1 & \frac{16-\sqrt{6}}{36} & \frac{16+\sqrt{6}}{36} & \frac{1}{9} \\ \hline & \frac{16-\sqrt{6}}{36} & \frac{16+\sqrt{6}}{36} & \frac{1}{9} \end{array}$$

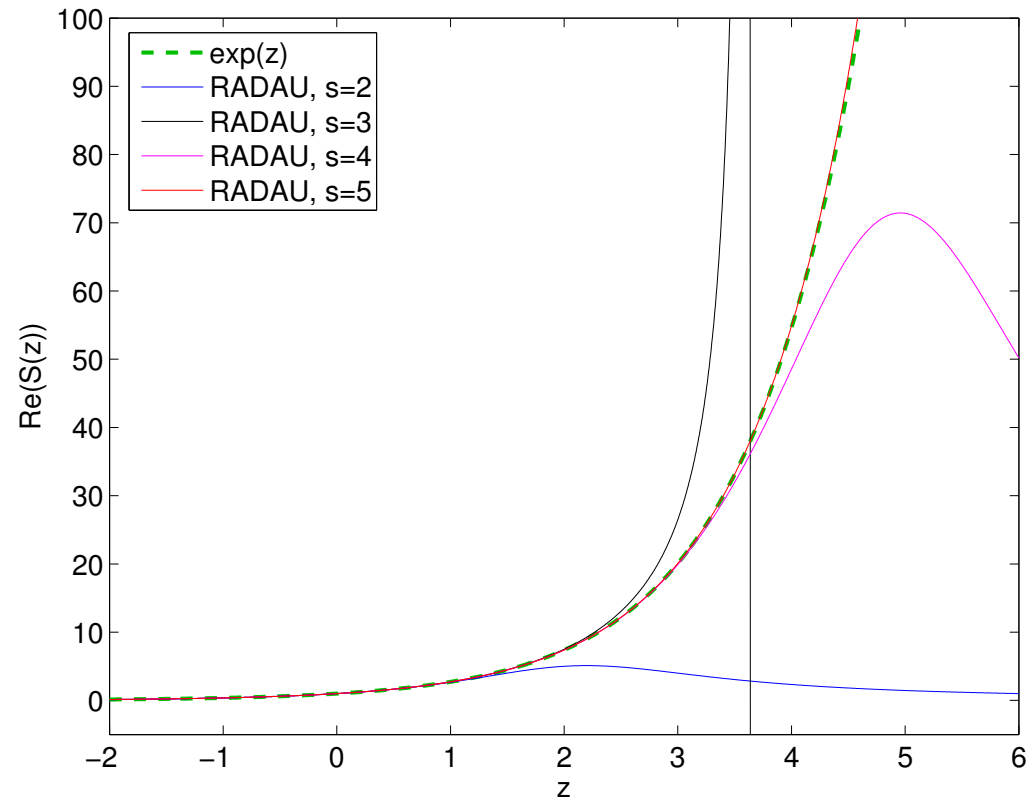
Radau-ESV, Ordnung 5

Stabilitätsfunktion  $s$ -stufiger Radau-Kollokations-RK-ESVs:

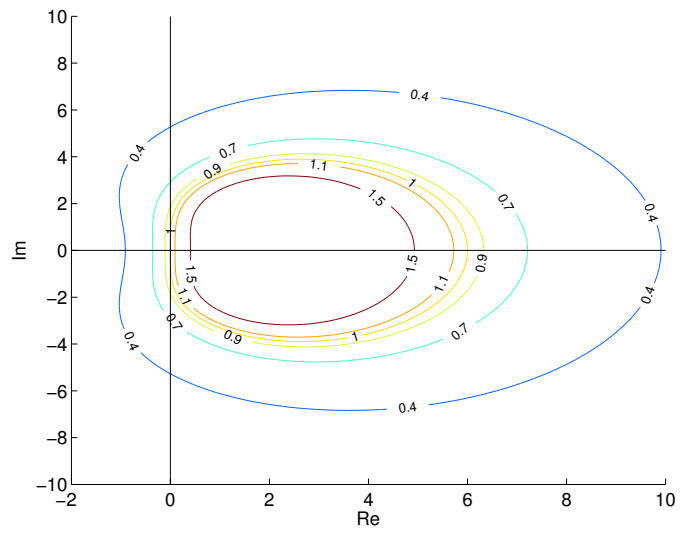
$$S(z) = \frac{P(z)}{Q(z)}, \quad P \in \mathcal{P}_{s-1}, Q \in \mathcal{P}_s.$$

Vorsicht:

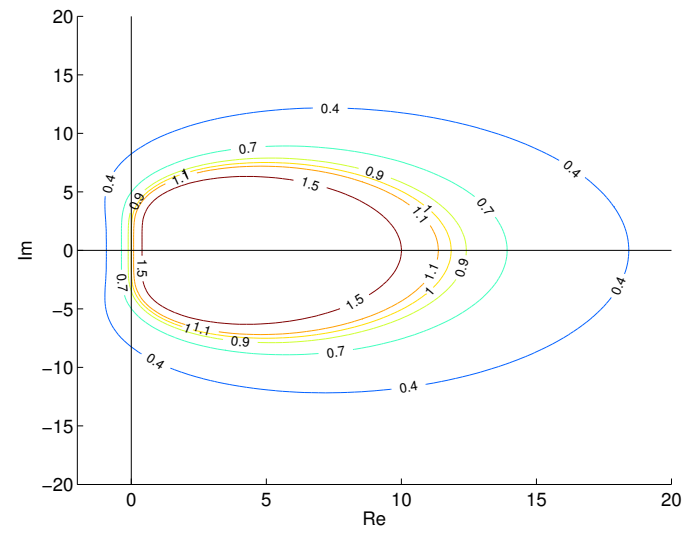
Auch „ $S(\infty) = 0$ “



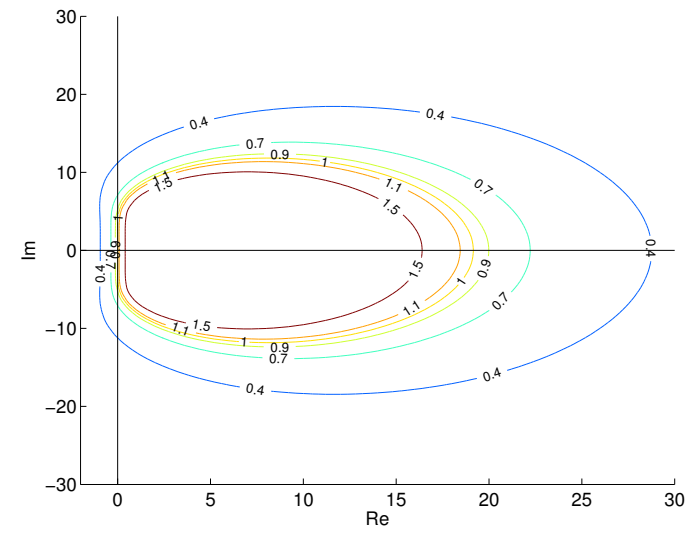
Niveaus der Stabilitätsfunktionen von  $s$ -stufigen Radau-Kollokations-RK-ESVs:



$$s = 2$$



$$s = 3$$



$$s = 4$$

Beispiel 3.4.7 (Radau-ESV bei stark attraktiven Fixpunkten). → Bsp. 3.4.1

AWP für ODE mit stark attraktivem Fixpunkt  $y = 1$ ,  
 Bsp. 3.4.1

$$\dot{y} = \lambda y^2(1 - y),$$

$$\lambda = 500, \quad y(0) = \frac{1}{100}.$$

Qualitatives Verhalten von Radau-ESV auf äquidistantem Gitter  $\triangleright$

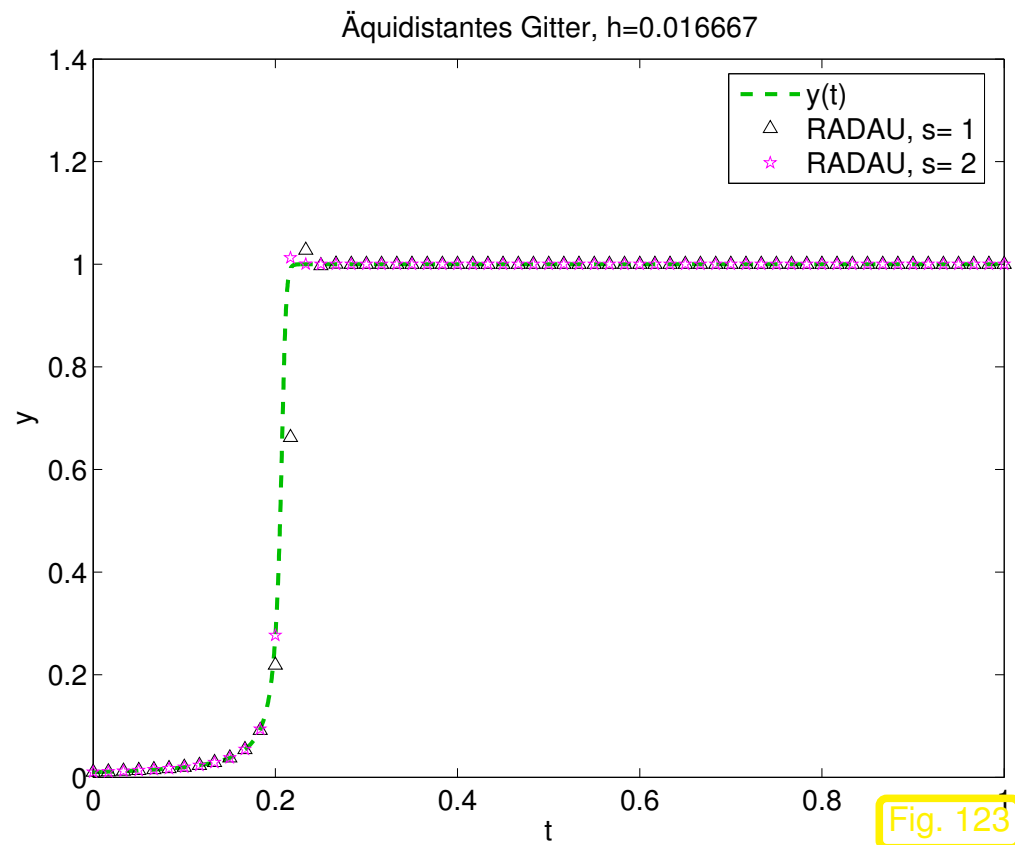


Fig. 123

$\triangleright$  Diskrete Evolution strebt schnell dem stabilen Zustand zu



Wie für Gauss-Kollokations-RK-ESV, siehe Thm. 3.3.7:

**Theorem 3.4.8** (Radau-ESV nichtexpansiv).

*Die diskreten Evolutionen zu Radau-ESV erben die Nichtexpansivität der (exakten) Evolution.*

*Beweis.* Erweiterung des Beweises zu Thm. 3.3.7. Mit den dortigen Notationen und Fehlerdarstellungsformel für Gauss-Radau-Quadraturformeln

$$\int_0^1 d'(\tau) d\tau = \sum_{j=1}^s b_j d'(c_j) - R \quad \text{mit} \quad R = c(s) d^{(2s)}(\xi), \quad 0 \leq \xi \leq 1,$$

wobei  $c(s) > 0$ . Formeln (3.3.10) und (3.3.11) bleiben gültig, so dass die Behauptung von Thm. 3.4.8 gezeigt ist, sobald  $R \geq 0$  sichergestellt ist:

$$d(\tau) = \sum_{j=0}^{2s} \delta_j \tau^j \quad \Rightarrow \quad d^{(2s)}(\tau) = (2s)! \delta_{2s},$$

$$d(\tau) \geq 0 \quad \Rightarrow \quad \lim_{|\tau| \rightarrow \infty} d(\tau) \geq 0 \quad \Rightarrow \quad \delta_{2s} \geq 0.$$

Beachte: Auch die Gewichte von Gauss-Radau-Quadraturformeln sind positiv, siehe Fig. 122.  $\square$

Das Folgende ist keine Definition, sondern ein durch die Beobachtungen von Anwendern numerischer Integratoren motivierter Begriff. Es ist nicht möglich, diesen Begriff in einer strengen mathematischen Definition zu erfassen. Dennoch ist er zentral für die Auswahl geeigneter numerischer Integratoren und Teilnehmer der Vorlesung sollten schliesslich ein “Gerfühl” dafür haben, wann ein Anfangswertproblem “steif” ist. Dieses wird in diesem Abschnitt anhand von Beispielen geschult.

Aus [25, Sect. 1]:

The usual definition of stiffness applies which states that a differential equation is stiff whenever the implicit Euler method works (tremendously) better than the explicit Euler method.

**Konzept 3.5.1** (Steifes Anfangswertproblem).

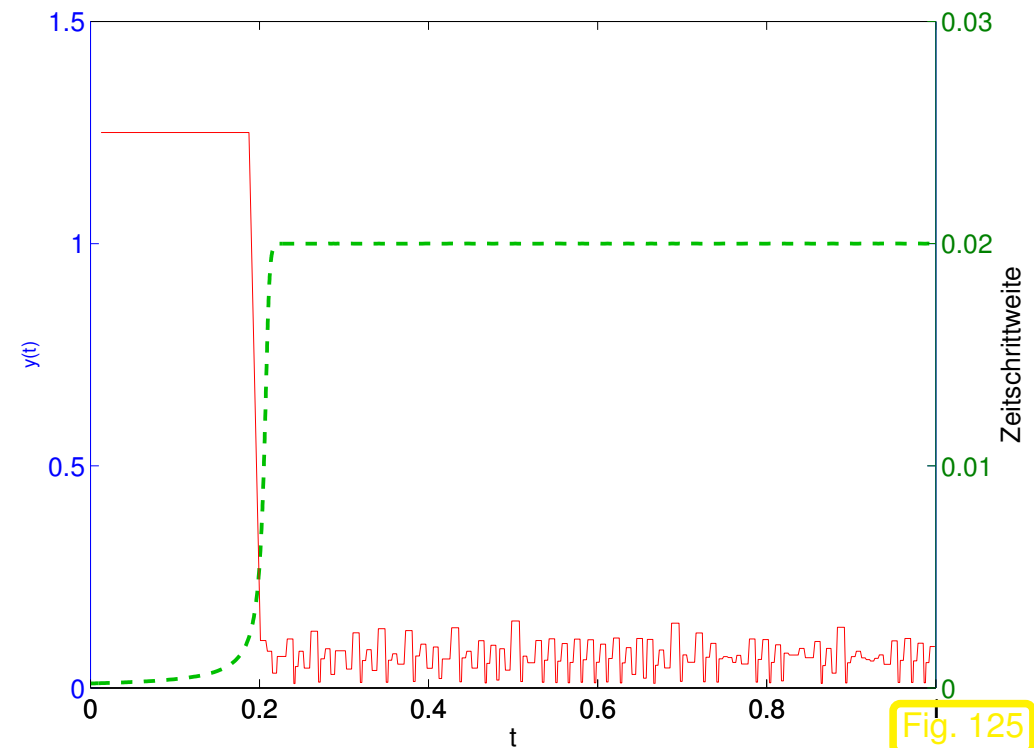
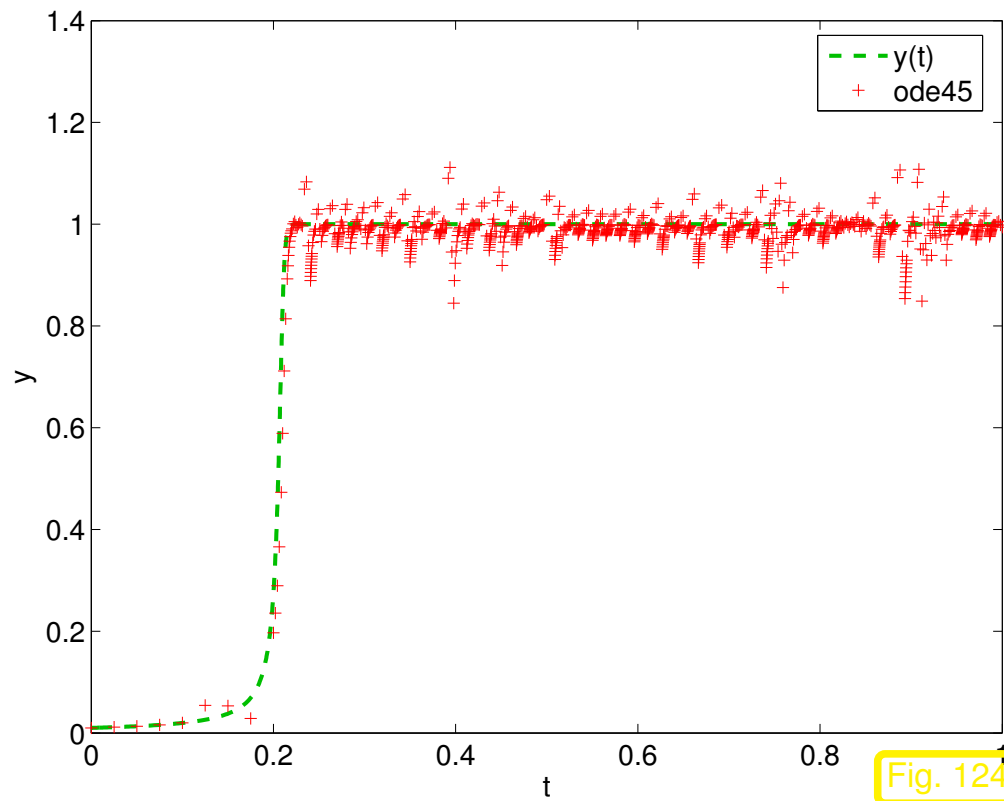
*Ein AWP heisst **steif** (engl. stiff), falls für explizite RK-ESV ( $\rightarrow$  Def. 2.1.5) Stabilität eine wesentlich kleinere Schrittweite verlangt als die Genauigkeitsanforderungen.*

Beispiel 3.5.2 (Adaptive explizite RK-ESV für steifes Problem). → Sect. 2.6

$$\dot{y}(t) = \lambda y^2(1 - y), \quad \lambda = 500, \quad y(0) = \frac{1}{100}.$$

MATLAB-CODE : Adaptives ESV für steifes Problem

```
fun = @(t,x) 500*x^2*(1-x); tspan = [0 1]; y0 = 0.01;
options = odeset('reltol',0.1,'abstol',0.001,'stats','on');
[t,y] = ode45(fun,tspan,y0,options);
plot(t,y,'r+');
```



➤ Schrittweitensteuerung realisiert Schrittweitenbeschränkung ! → Bsp. 2.6.10  
(186 successful steps, 55 failed attempts, 1447 function evaluations)



$y = 1$  stark attraktiver Fixpunkt



Extreme Schrittweitenbeschränkung für  
expliziten Integrator ode45

Beachte: die Schrittweitensteuerung erkennt Stabilitätsprobleme und reduziert die Schrittweite entsprechend! → Bsp. 2.6.10

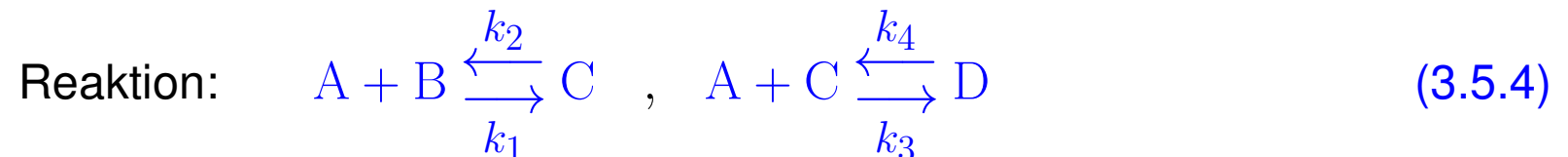


Welche Anfangswertprobleme sind steif ?

ODE-Modelle für Systeme mit schnell relaxierenden Komponenten  
(mit stark unterschiedlichen Zeitkonstanten)



Beispiel 3.5.3 (Steife Probleme in der chemischen Reaktionskinetik). → Sect. 1.2.2



Stark unterschiedliche Reaktionsgeschwindigkeiten:

$$k_1, k_2 \gg k_3, k_4$$

Numerisches Experiment, MATLAB,  $t_0 = 0$ ,  $T = 1$ ,  $k_1 = 10^4$ ,  $k_2 = 10^3$ ,  $k_3 = 10$ ,  $k_4 = 1$

MATLAB-CODE : Explizite Integration steifer chemischer Reaktionsgleichungen

```
fun = @(t,y) ([-k1*y(1)*y(2) + k2*y(3) - k3*y(1)*y(3) + k4*y(4);
              -k1*y(1)*y(2) + k2*y(3);
              k1*y(1)*y(2) - k2*y(3) - k3*y(1)*y(3) + k4*y(4);
              k3*y(1)*y(3) - k4*y(4)]);

tspan = [0 1];
y0 = [1;1;10;0];
options = odeset('reltol',0.1,'abstol',0.001,'stats','on');
[t,y] = ode45(fun,[0 1],y0,options);
```

Chemical reaction: concentrations

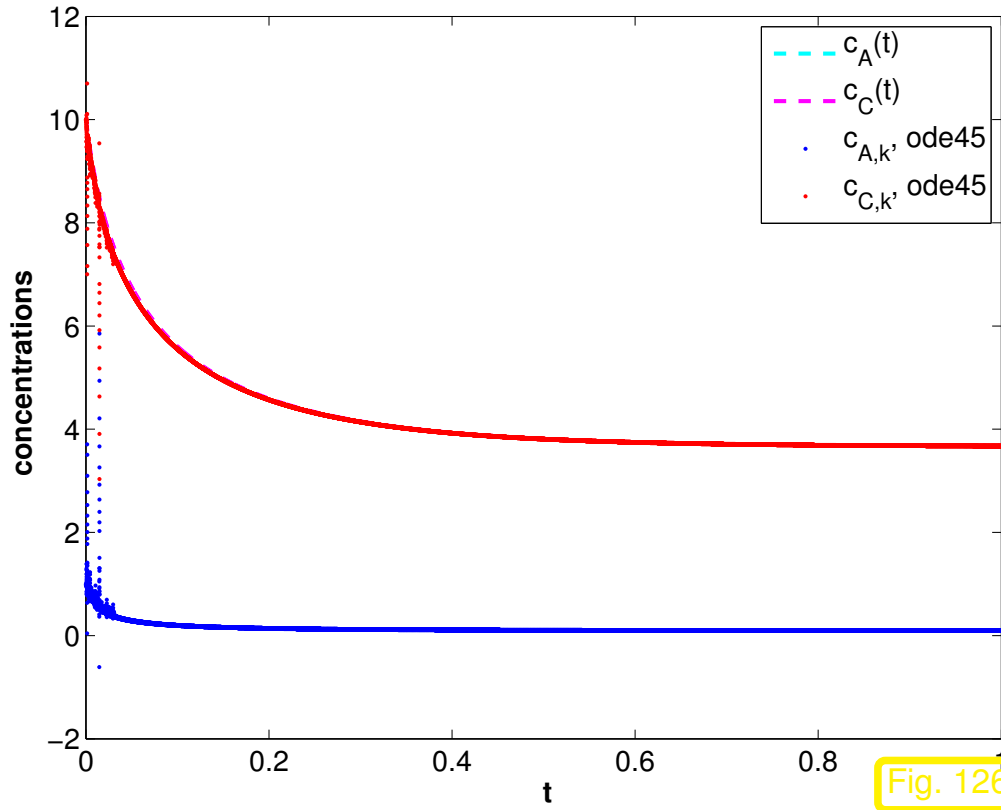


Fig. 126

Chemical reaction: stepsize

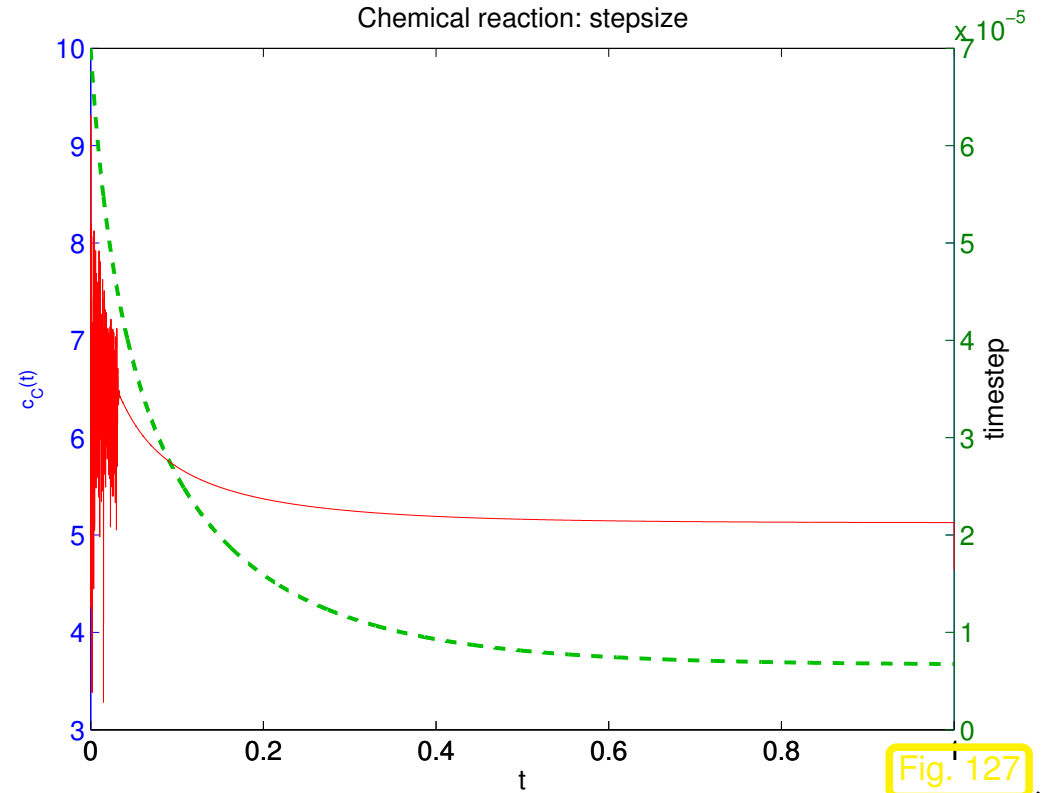


Fig. 127

Beispiel 3.5.5 (Steife Schaltkreisgleichungen im Zeitbereich).

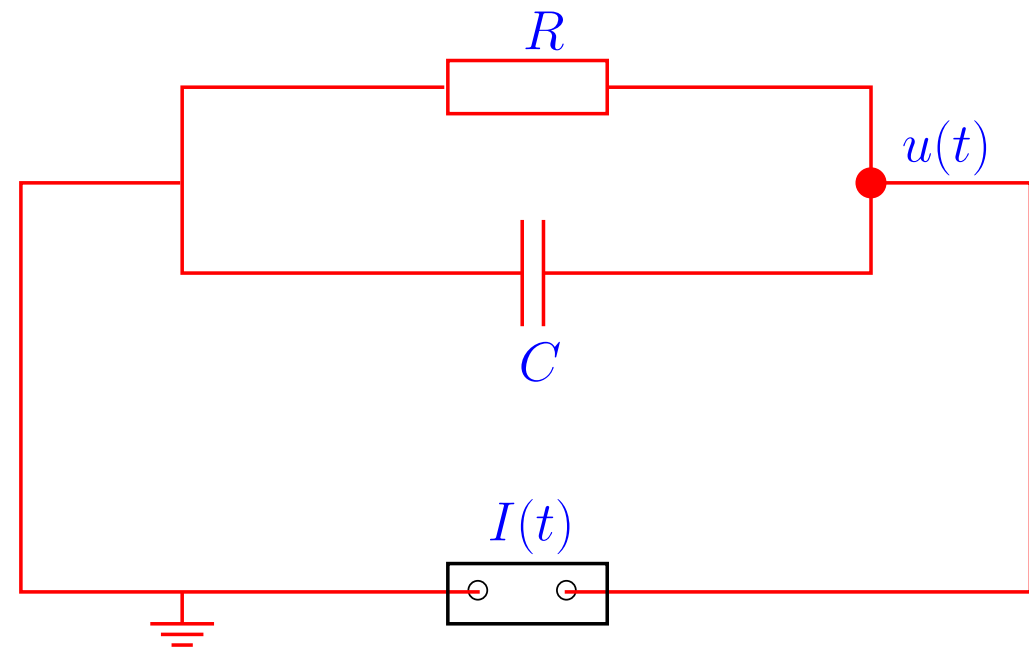
## Schaltkreisanalyse im Zeitbereich:

Bauelementgleichungen ( $i \hat{=}$  Strom,  $u \hat{=}$  Spannung):

- Widerstand:  $i(t) = R^{-1}u(t)$
- Kondensator:  $i(t) = C\dot{u}(t)$

Dgl. aus Knotenanalyse:

$$C\dot{u}(t) = -R^{-1}u(t) + I(t) .$$



Konkret:  $C = 1\text{pF}$ ,  $R = 1\text{k}\Omega$ ,  $I(t) = \sin(2\pi \text{1Hz } t)\text{mA}$ ,  $u(0) = 0\text{V}$

► Skalierte (dimensionslose) Dgl. :  $\dot{u}(t) = -10^9 u(t) + 10^9 \sin(2\pi t) \Rightarrow u(t) \approx \sin(2\pi t) .$   
 $\hat{=}$  ODE aus Bsp. 3.4.2

Im Fall der nichtautonomen Dgl.  $\dot{y} = -\lambda y + g(t)$ ,  $\lambda \gg 1$ , sind wird mit einem “zeitlich variierenden” stark attraktiven Fixpunkt  $y^*(t) = \lambda^{-1}g(t)$  konfrontiert. Auch dieser führt zu Steifheit gemäss Konzept 3.5.1.

Steife AWP ➤ (Geeignete) implizite RK-ESV (→ Def. 2.1.5) sind zu verwenden !

d.h. L-stabil (→ Def. 3.4.3)

Beispiel 3.5.6 (Attraktiver Grenzyklus). vgl. Bsp. 1.2.15

Autonome Dgl.  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$

$$\mathbf{f}(\mathbf{y}) := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \mathbf{y} + \lambda(1 - \|\mathbf{y}\|^2) \mathbf{y},$$

auf Zustandsraum  $D = \mathbb{R}^2 \setminus \{0\}$ .

Lösungstrajektorien ( $\lambda = 10$ ) ▷

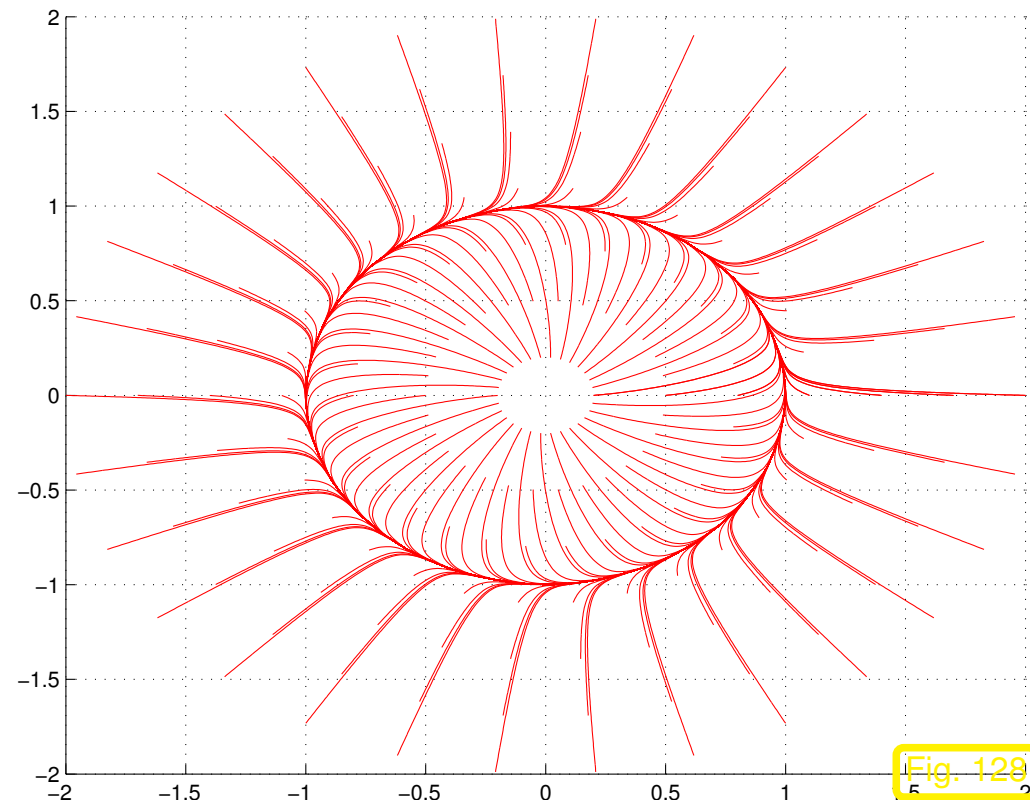


Fig. 129

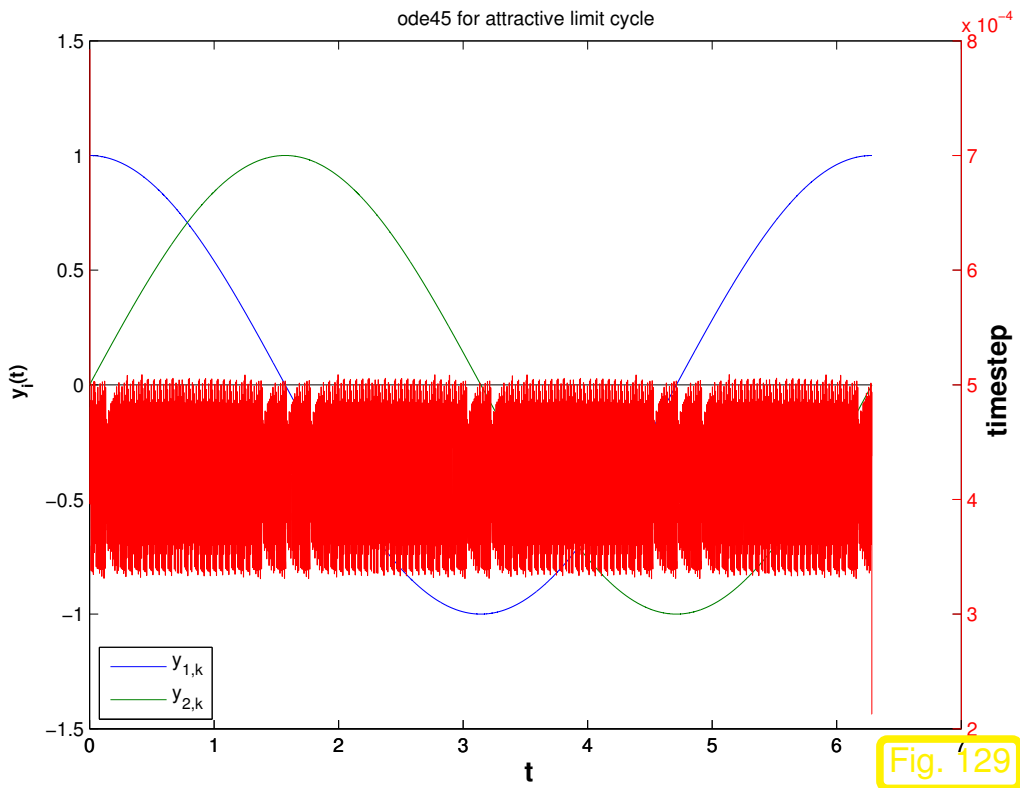
- ☞ Falls  $\|\mathbf{y}_0\| = 1 \Rightarrow \|\mathbf{y}(t)\| = 1 \quad \forall t$
- ☞ “Grenzyklus auf Einheitskreis”:  $\|\mathbf{y}(t)\| \rightarrow 1$  für  $t \rightarrow \infty$ .

In diesem Beispiel liegt kein asymptotisch stabiler Fixpunkt vor, sondern eine asymptotisch stabile **invariante Mannigfaltigkeit**, also eine echte Teilmenge  $M \subset D$  des Zustandsraums, für die gilt  $\Phi^t M \subset M$  für alle zulässigen  $t$  (“Fixmenge”) und

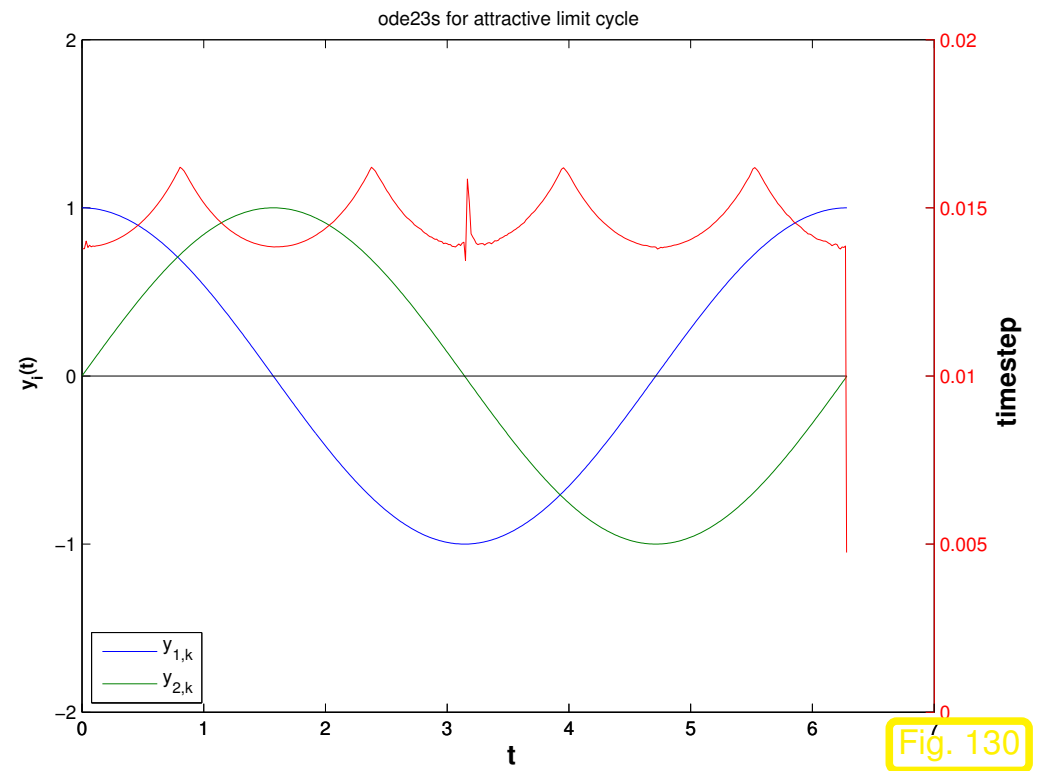
$$\exists \text{Umgebung } U \text{ von } M: : \quad \mathbf{y}(0) \in U \quad \Rightarrow \quad \lim_{t \rightarrow \infty} \text{dist}(\mathbf{y}(t), M) = 0 .$$

MATLAB-CODE Integration von Evolution mit Grenzyklus

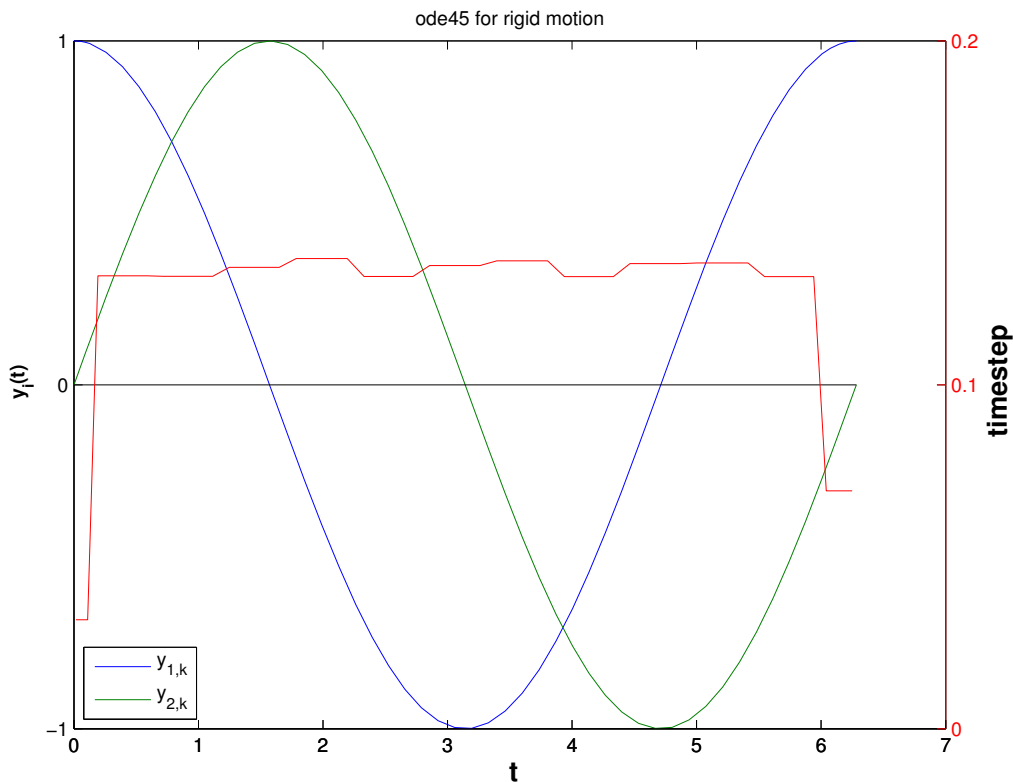
```
fun = @(t,y) ([-y(2);y(1)] + lambda*(1-y(1)^2-y(2)^2)*y);
tspan = [0,2*pi]; y0 = [1,0];
opts = odeset('stats','on','reltol',1E-4,'abstol',1E-4);
[t45,y45] = ode45(fun,tspan,y0,opts);
[t23,y23] = ode23s(fun,tspan,y0,opts);
```



ode45: 3794 Schritte ( $\lambda = 1000$ )



ode23s: 432 Schritte ( $\lambda = 1000$ )



◁  $\lambda = 0 \hat{=}$  Drehbewegung, siehe Bsp. 1.4.18

ode45 erzielt gute Genauigkeit trotz (relativ) grosser Schrittweite.

Nichtsteifes Problem!



Adaptive MATLAB-Integratoren für steife Probleme: (Schrittweitensteuerung wie in Abschnitt 2.6)

```
opts = odeset('abstol', atol, 'reltol', rtol, 'Jacobian', @J)
[t, y] = ode15s/ode23s(odefun, tspan, y0, opts);
```

Beispiel 3.5.7 (Adaptives semi-implizites RK-ESV für steifes Problem). → Bsp. 3.5.2, 3.4.1

$$\dot{y}(t) = \lambda y^2(1 - y), \quad \lambda = 500, \quad y(0) = \frac{1}{100}.$$

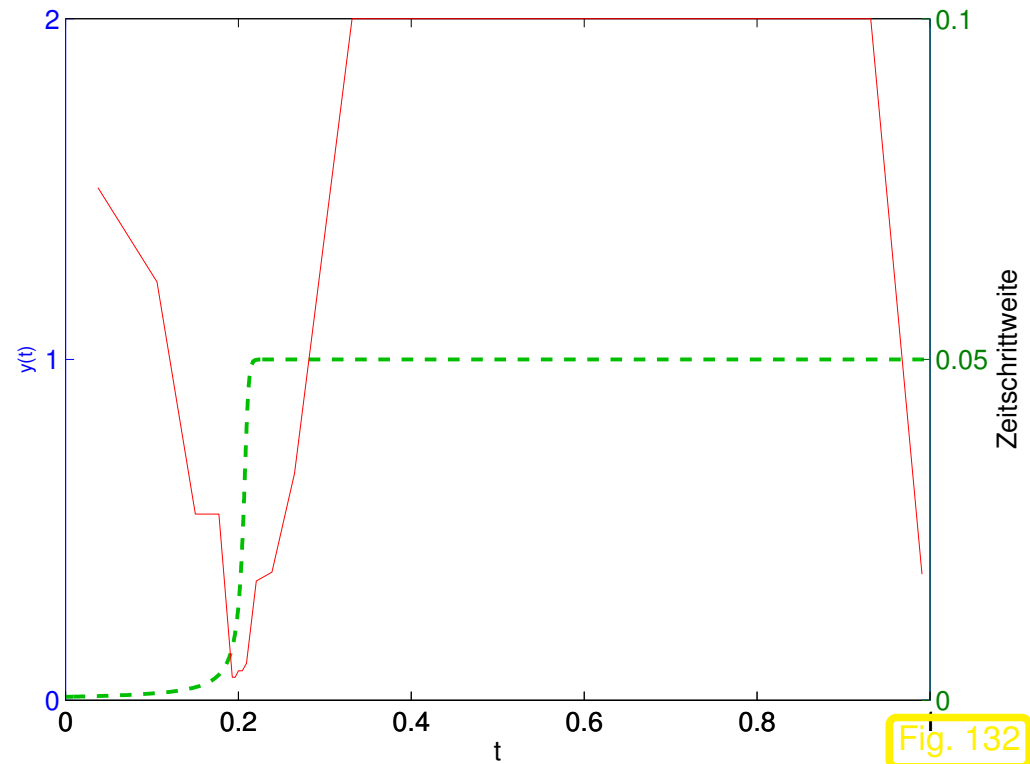
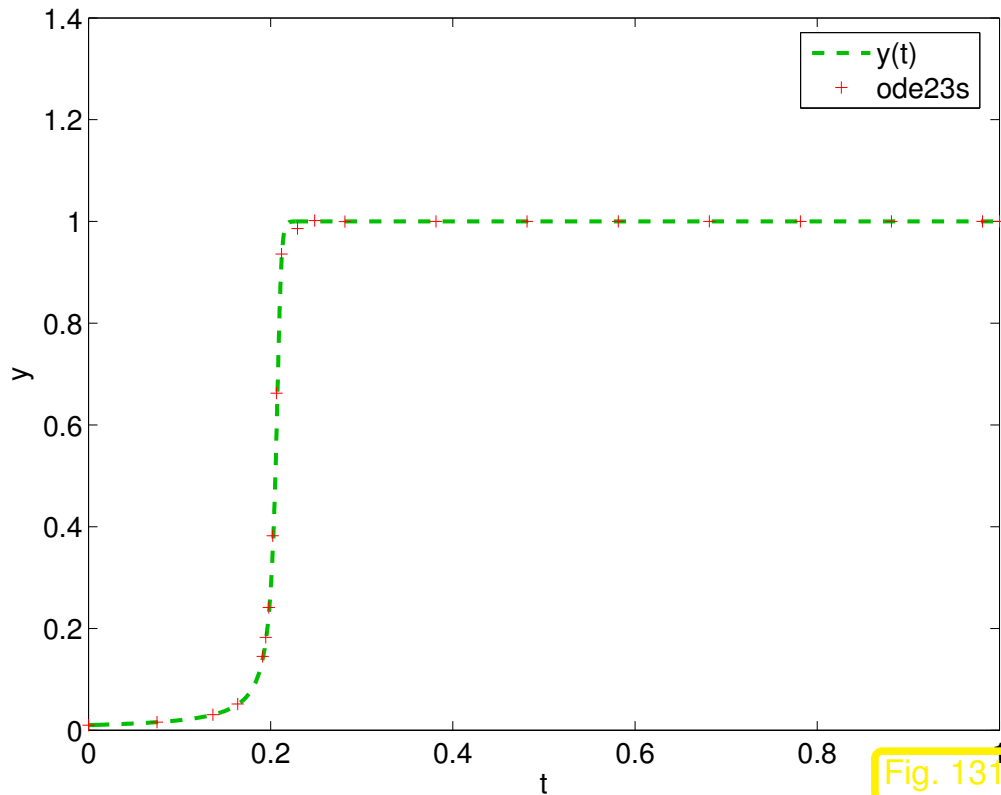


MATLAB-CODE : Semi-Implizites ESV für steifes Problem

```
lambda = 500; tspan = [0 1]; y0 = 0.01;
fun = @(t,x) lambda*x^2*(1-x);
Jac = @(t,x) lambda*(2*x*(1-x)-x^2);
o = odeset('reltol',0.1,'abstol',0.001,'stats','on','Jacobian',Jac);
[t,y] = ode23s(fun,[0 1],y0,o);
```

Statistik:

20 successful steps 4 failed attempts 70 function  
evaluations

R. Hiptmair  
rev 35327,  
24. Juni  
2011

➤ Effizientes Verfahren (vgl. Bsp. 3.5.2): Keine Schrittweitenbeschränkung für  $y \approx 1$

# 3.6 Linear-implizite Runge-Kutta-Verfahren [8, Sect. 6.4]



Inkrementgleichungen (2.2.3) für  $s$ -stufige implizite RK-ESV

=

Nichtlineares Gleichungssystem der Dimension  $s \cdot d$

*Beispiel 3.6.1* (Linearisierung der Inkrementgleichungen).

- Anfangswertproblem für logistische Differentialgleichung, siehe Bsp. 1.2.1

$$\dot{y} = \lambda y(1 - y) \quad , \quad y(0) = 0.1 \quad , \quad \lambda = 5 .$$

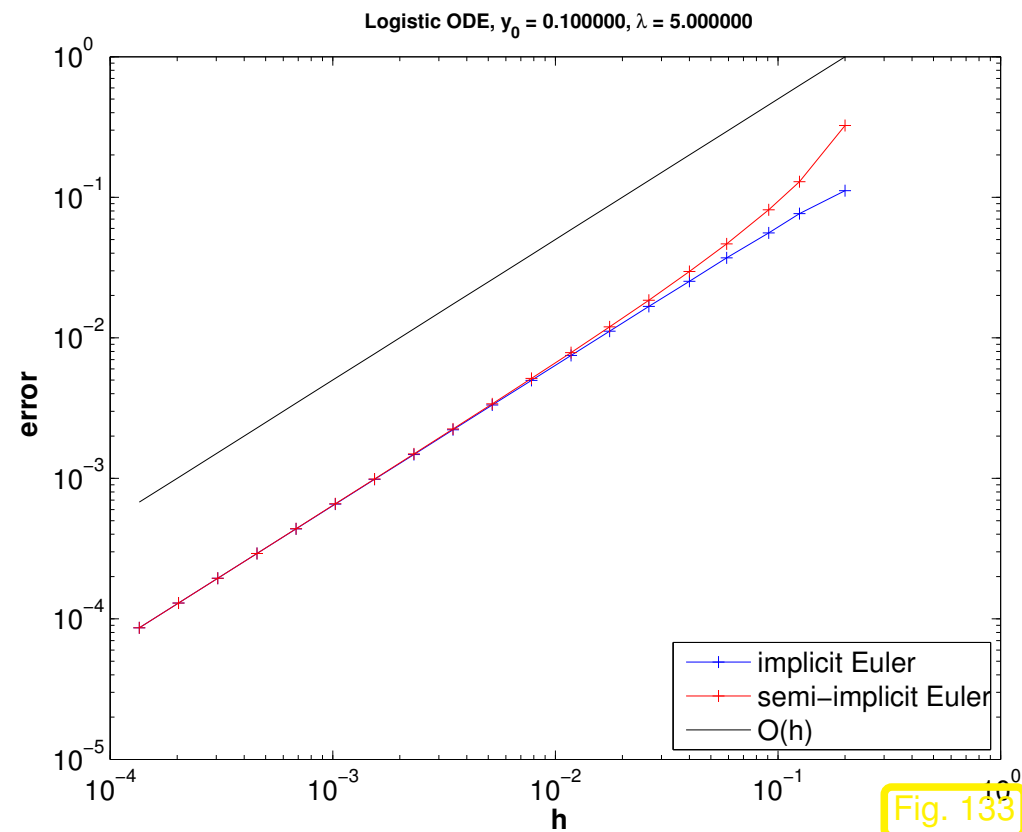
• Implizites Euler-Verfahren (1.4.13) mit uniformem Zeitschritt  $h = 1/n$ ,

$n \in \{5, 8, 11, 17, 25, 38, 57, 85, 128, 192, 288, 432, 649, 973, 1460, 2189, 3284, 4926, 7389\}$ .

& näherungsweise Berechnung von  $y_{k+1}$   
durch **1 Newton-Schritt** mit Startwert  $y_k$

= semi-implizites Euler-Verfahren

• Fehlermass  $\text{err} = \max_{j=1, \dots, n} |y_j - y(t_j)|$



- Implizite Mittelpunktsregel (1.4.19) mit uniformem Zeitschritt  $h = 1/n$  (wie oben)

& näherungsweise Berechnung von  $y_{k+1}$  durch 1 Newton-Schritt mit Startwert  $y_k$

- Fehlermass  $\text{err} = \max_{j=1, \dots, n} |y_j - y(t_j)|$

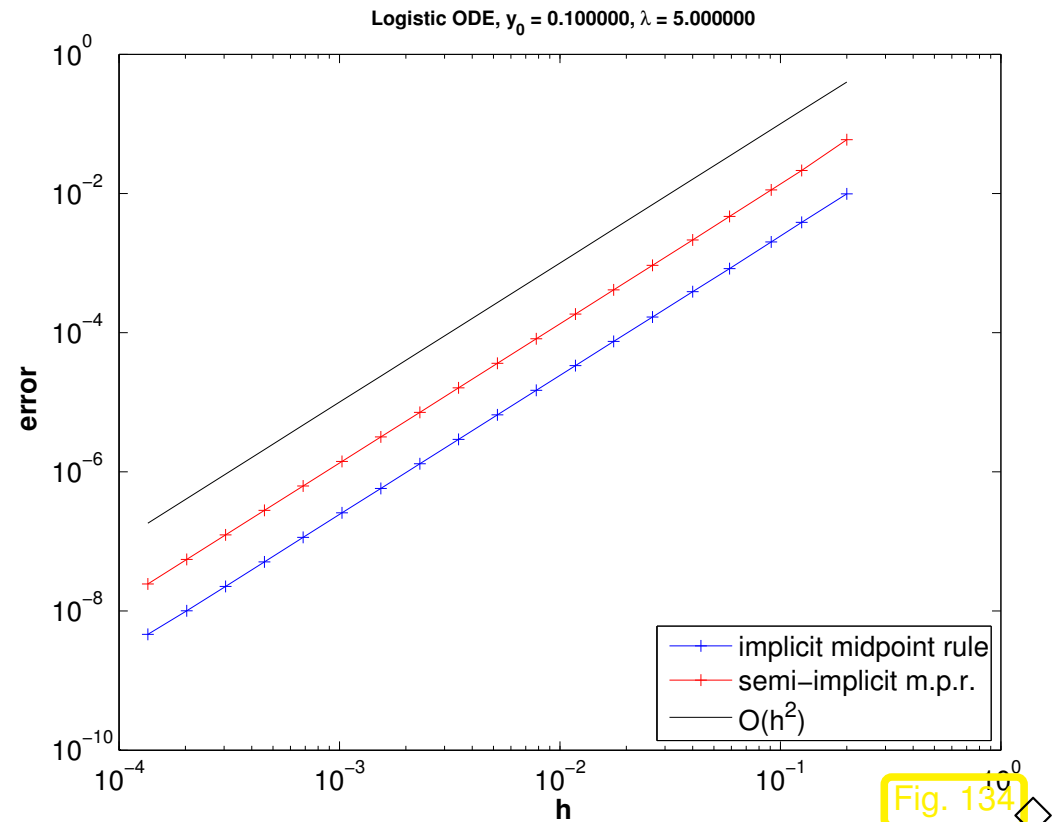


Fig. 134

Idee: Implizite RK-ESV mit **linearisierten Inkrementgleichungen** (2.2.3)

$$\mathbf{k}_i = \mathbf{f}(\mathbf{y}_0) + h D\mathbf{f}(\mathbf{y}_0) \left( \sum_{j=1}^s a_{ij} \mathbf{k}_j \right), \quad i = 1, \dots, s. \quad (3.6.2)$$

(3.6.2)  $\hat{=}$  LGS der Dimension  $s \cdot d$ : ( $s \hat{=}$  Anzahl der Stufen,  $\mathfrak{A} \in \mathbb{R}^{s,s} \hat{=}$  Koeffizientenmatrix aus Butcher-Schema (2.3.6))

$$(\mathbf{I}_{s \cdot d} - h\mathfrak{A} \otimes D\mathbf{f}(\mathbf{y}_0)) \begin{pmatrix} \mathbf{k}_1 \\ \vdots \\ \mathbf{k}_s \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \otimes \mathbf{f}(\mathbf{y}_0),$$

mit Kronecker-Produkt: für  $\mathbf{A} \in \mathbb{R}^{m,n}$ ,  $\mathbf{B} \in \mathbb{R}^{k,l}$

$$\mathbf{A} \otimes \mathbf{B} := \begin{pmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & & \vdots \\ a_{m,1}\mathbf{B} & \cdots & a_{m,n}\mathbf{B} \end{pmatrix} \in \mathbb{R}^{m \cdot k, n \cdot l}.$$

MATLAB-Kommando `kron(A,B)`.

Linearisierung folgenlos bei linearen ODE  $\rightarrow$  Stabilitätsfunktion ( $\rightarrow$  Def. 3.1.6) unverändert

*Beispiel 3.6.3* (Implizite RK-ESV mit linearisierten Inkrementgleichungen).

- Anfangswertproblem für logistische Differentialgleichung, siehe Bsp. 1.2.1

$$\dot{y} = \lambda y(1 - y) \quad , \quad y(0) = 0.1 \quad , \quad \lambda = 5.$$

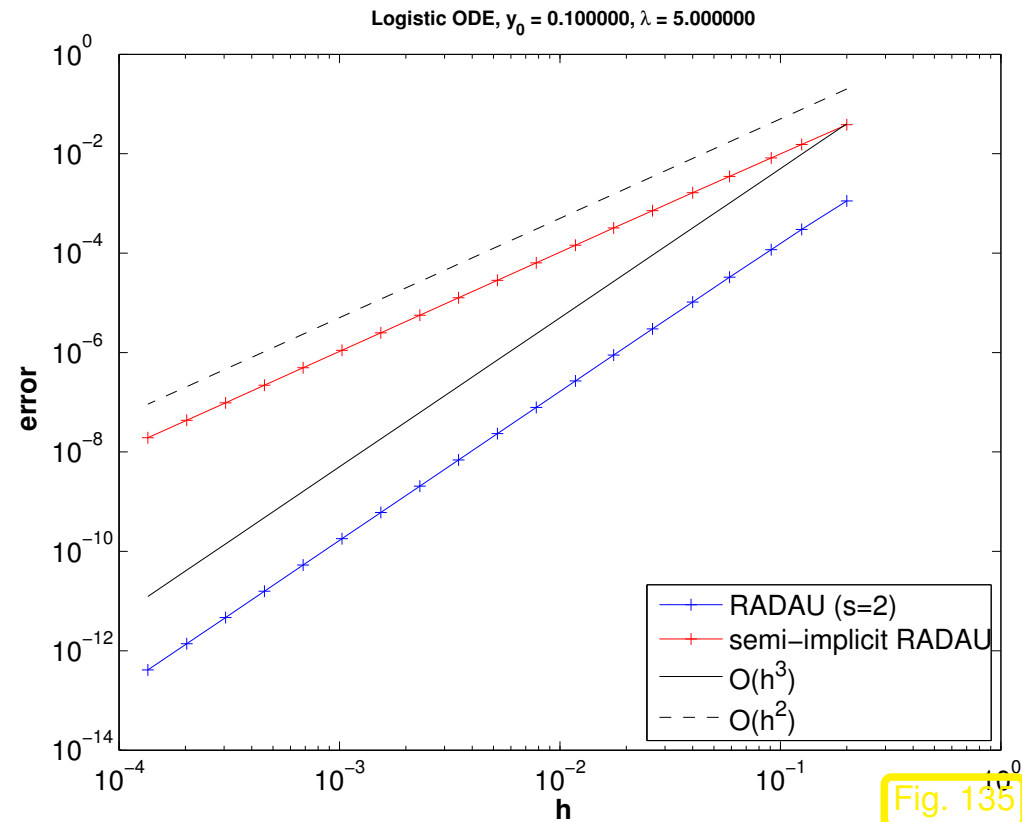
2-stufiges Radau-ESV, Butcher Schema

$$\begin{array}{c|cc} \frac{1}{3} & \frac{5}{12} & -\frac{1}{12} \\ 1 & \frac{3}{4} & \frac{1}{4} \\ \hline & \frac{3}{4} & \frac{1}{4} \end{array}, \quad (3.6.4)$$

Ordnung 3, siehe Sect. 3.4.

Inkrementale aus linearisierten Gleichungen  
(3.6.2)

Fehlermass  $err = \max_{j=1, \dots, n} |y_j - y(t_j)|$



Ordnungsverlust durch Linearisierung !

Die einfach Linearisierung (3.6.2) führt bei impliziten RK-ESV zu einem erheblichen Verlust an Genauigkeit und ist daher keine Option.



Idee: „Rettung“ der Ordnung durch bessere Startnäherung für (einen) Newtonschritt ?

Wir betrachten:

diagonal-implizite RK-ESV (DIRK)



$\mathcal{A}$  untere Dreiecksmatrix

( $\mathcal{A}$  regulär  $\Leftrightarrow a_{jj} \neq 0$ )

$$\frac{\mathbf{c} \mid \mathcal{A}}{\mathbf{b}^T} :=$$

$$\begin{array}{c|cccccc} c_1 & a_{11} & 0 & & \cdots & 0 \\ c_2 & a_{21} & a_{22} & 0 & & 0 \\ \vdots & \vdots & & \ddots & \ddots & \vdots \\ \vdots & \vdots & & & \ddots & \vdots \\ \vdots & \vdots & & & & 0 \\ c_s & a_{s1} & & \cdots & & a_{ss} \\ \hline & b_1 & & \cdots & \cdots & b_s \end{array} \quad (3.6.5)$$

► Gestaffeltes (nichtlineares) Gleichungssystem für Inkremente (autonomer Fall)

Allgemeine Inkrementgleichungen für  $s$ -stufiges DIRK-Verfahren:

$$\mathbf{k}_i = \mathbf{f}\left(\mathbf{y}_0 + h \sum_{j=1}^i a_{ij} \mathbf{k}_j\right), \quad \mathbf{y}_1 = \mathbf{y}_0 + h \sum_{i=1}^s b_i \mathbf{k}_i.$$

$i$ . Inkrementgleichung: Umformulierung als Problem der Nullstellensuche von

$$F(\mathbf{k}) := \mathbf{k} - \mathbf{f}\left(\mathbf{y}_0 + \mathbf{z} + h a_{ii} \mathbf{k}\right) = 0, \quad \mathbf{z} = h \sum_{j=1}^{i-1} a_{ij} \mathbf{k}_j.$$

$$\blacktriangleright \quad DF(\mathbf{k}) = \mathbf{I} - Df(\mathbf{y}_0 + \mathbf{z} + a_{ii}\mathbf{k})ha_{ii} .$$

Ein Newton-Schritt mit Startwert  $\mathbf{k}_i^{(0)}$ :

$$\mathbf{k}_i^{(1)} = \mathbf{k}_i^{(0)} - (\mathbf{I} - Df(\mathbf{y}_0 + \mathbf{z} + ha_{ii}\mathbf{k})ha_{ii})^{-1} \cdot \left( \mathbf{k}_i^{(0)} - \mathbf{f}(\mathbf{y}_0 + \mathbf{z} + ha_{ii}\mathbf{k}_i^{(0)}) \right)$$

Vereinfachung, vgl. Bem. 2.3.19: Benutze Jacobi-Matrix an der Stelle  $\mathbf{y}_0$

Newton-Verfahren:      Allgemeiner Ansatz für Startnäherung:

Ansatz Startnäherung (für  $\mathbf{k}_i$ ):

$$\mathbf{k}_i^{(0)} = \sum_{j=1}^{i-1} \frac{d_{ij}}{a_{ii}} \mathbf{k}_j . \quad (3.6.6)$$

$$\blacktriangleright \quad (\mathbf{I} - ha_{ii}\mathbf{J})\mathbf{k}_i = \mathbf{f}(\mathbf{y}_0 + h \sum_{j=1}^{i-1} (a_{ij} + d_{ij})\mathbf{k}_j) - h\mathbf{J} \sum_{j=1}^{i-1} d_{ij}\mathbf{k}_j , \quad (3.6.7)$$

$$\mathbf{J} := Df(\mathbf{y}_0 + \underline{h \sum_{j=1}^{i-1} (a_{ij} + d_{ij})\mathbf{k}_j}) . \quad (3.6.8)$$

Vereinfachtes Newton-Verfahren („eingefrorene“ Jacobi-Matrix)

Wie bei Standard-RK-ESV ( $\rightarrow$  Def. 2.3.5):

$$\mathbf{y}_1 = \mathbf{y}_0 + h \sum_{i=1}^s b_i \mathbf{k}_i . \quad (3.6.9)$$



Nächster Schritt: Bestimme Koeffizienten  $\mathbf{a}_{ij}$ ,  $\mathbf{d}_{ij}$  in (3.6.7) (und  $\mathbf{b}_i$ ), so dass sie *Ordnungsgleichungen* genügen

(analog zur Konstruktion von Runge-Kutta-Verfahren in Sect. 2.3)

► **Linear-implizite Runge-Kutta-Verfahren** (Rosenbrock-Wanner (ROW)-Methoden)

*Beispiel* 3.6.10 (Bedingungsgleichungen für Linear-implizite Runge-Kutta-Verfahren 2. Ordnung).

Aus der Neumannschen Reihe für Matrizen: für  $h > 0$  "hinreichend klein"

$$(\mathbf{I} - ha_{ii}\mathbf{J})^{-1} = \sum_{k=0}^{\infty} (ha_{ii}\mathbf{J})^k = \mathbf{I} + ha_{ii}\mathbf{J} + O(h^2). \quad (3.6.11)$$

Einsetzen in (3.6.7) + Taylor-Entwicklung von  $\mathbf{f}$  um  $\mathbf{y}_0$  + rekursives Einsetzen, vgl. Bsp. 2.3.24:

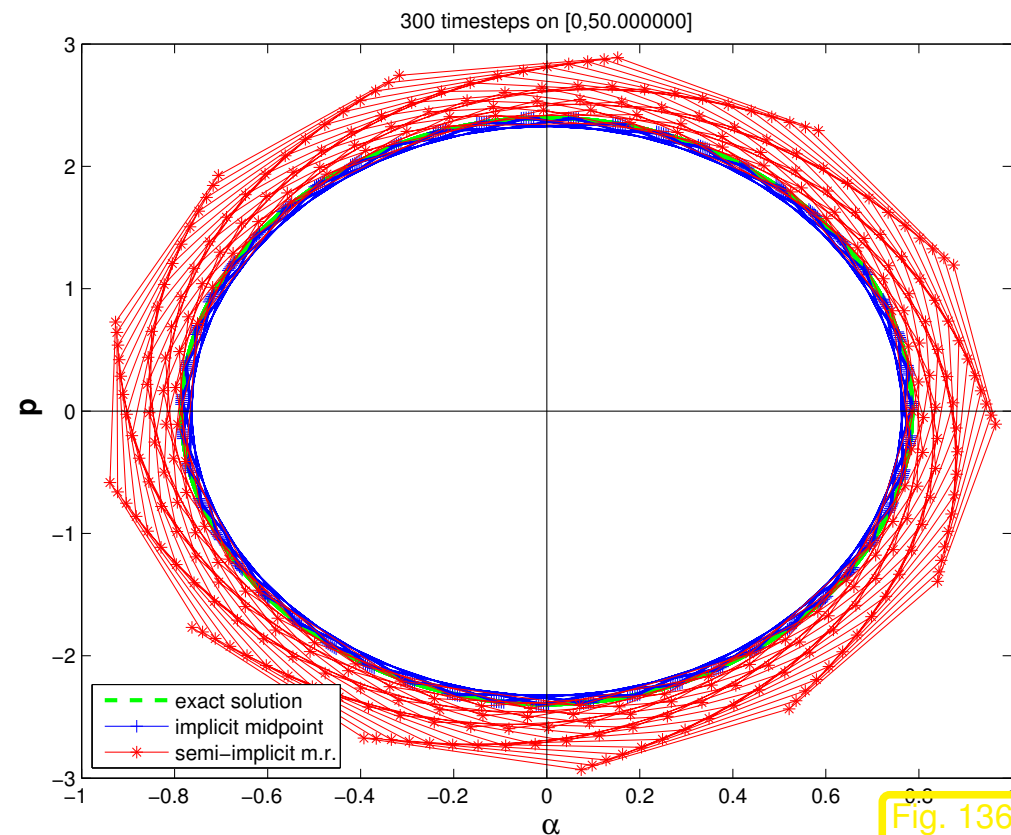
$$\begin{aligned} \mathbf{k}_i &= \left( \mathbf{I} + ha_{ii}\mathbf{J} + O(h^2) \right) \left( \mathbf{f}(\mathbf{y}_0) + h\mathbf{J} \sum_{j=1}^{i-1} (a_{ij} + d_{ij})\mathbf{k}_j + O(h^2) - h\mathbf{J} \sum_{j=1}^{i-1} d_{ij}\mathbf{k}_j \right) \\ &= \mathbf{f}(\mathbf{y}_0) + ha_{ii}\mathbf{J}\mathbf{f}(\mathbf{y}_0) + h\mathbf{J}\mathbf{f}(\mathbf{y}_0) \left( \sum_{j=1}^{i-1} a_{ij} \right) + O(h^2). \end{aligned}$$

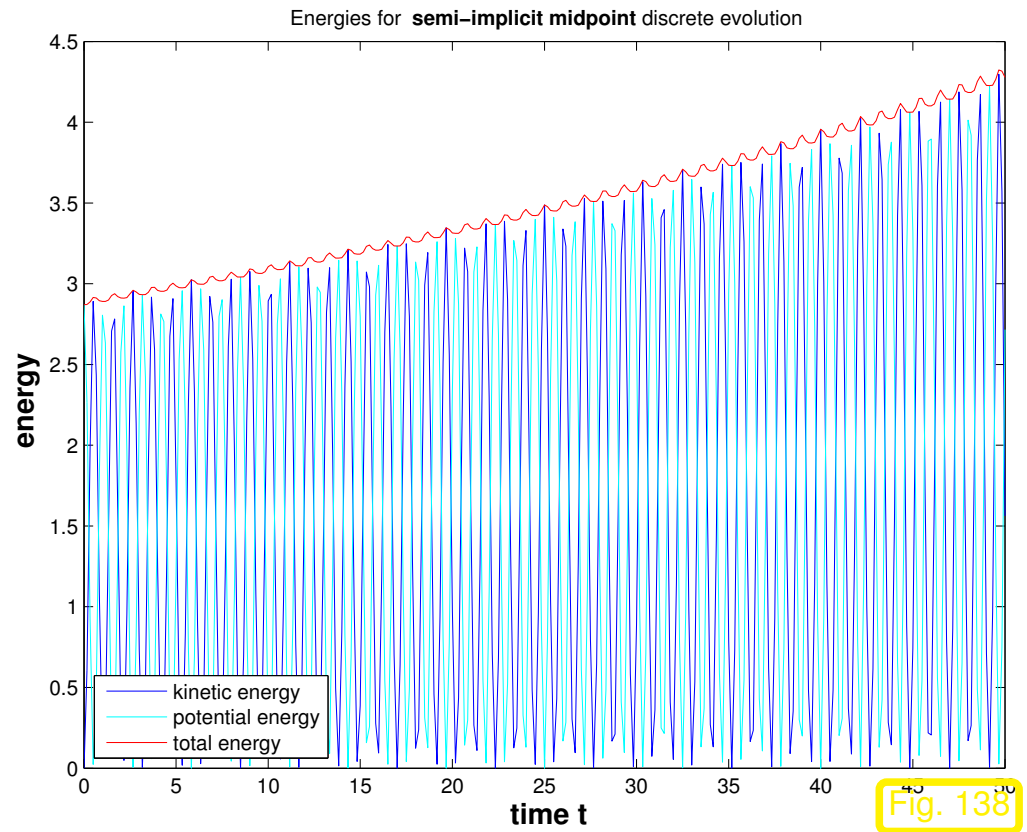
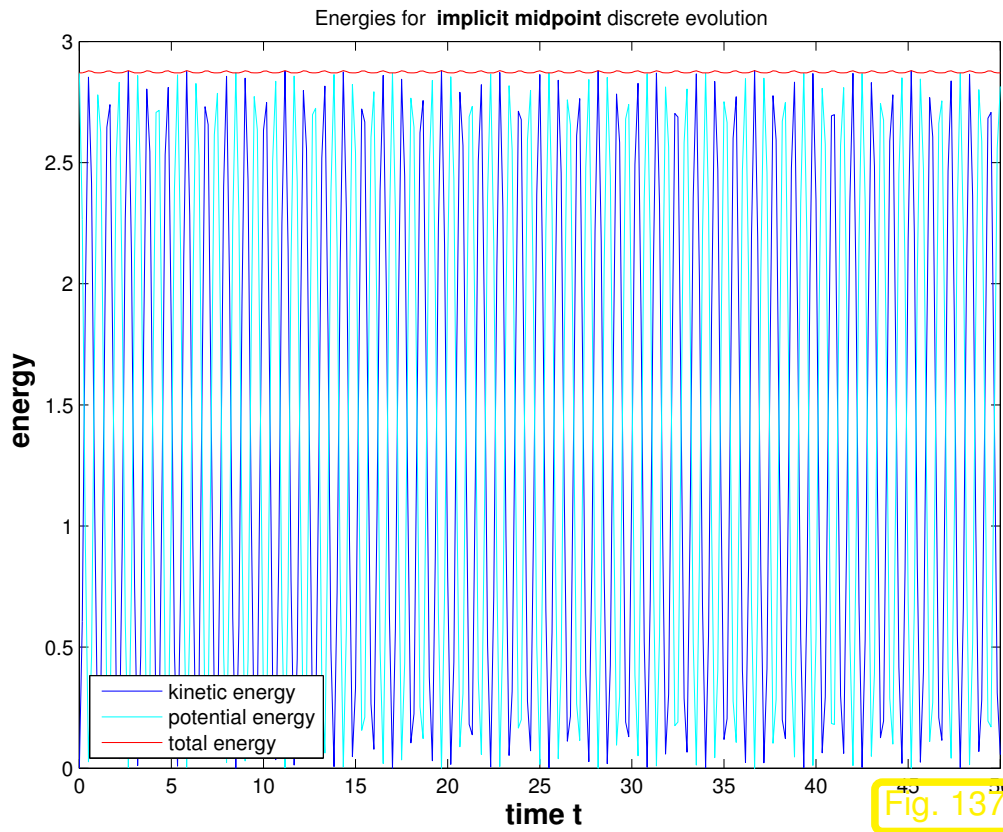
$$(3.6.9) \quad \Rightarrow \quad \mathbf{y}_1 = \mathbf{y}_0 + h \left( \sum_{i=1}^s b_i \right) \mathbf{f}(\mathbf{y}_0) + h^2 \left( \sum_{i=1}^s b_i \sum_{j=1}^{i-1} a_{ij} \right) \mathbf{Jf}(\mathbf{y}_0) + O(h^3).$$

Dabei wurde benutzt:  $\mathbf{J} = D\mathbf{f}(\mathbf{y}_0)$ . Dann Vergleich mit Taylorentwicklung (2.3.26)  $\triangleright$  Bedingungsgleichungen (2.3.29), (2.3.30) (gleich wie für Standard-Rk-ESV!).

*Beispiel 3.6.12* (Energieerhaltung bei semi-impliziter Mittelpunktsregel).

- Hamiltonsche ODE (1.2.19) für mathematisches Pendel für  $0 \leq t \leq T := 50$ , Anfangswerte  $\alpha(0) = \pi/4$ ,  $p(0) = 0$
- Implizite Mittelpunktsregel (1.4.19)/semi-implizite Mittelpunktsregel ( $\rightarrow$  Bsp. 3.6.1) auf uniformem Zeitgitter  $h = T/300$ ,
- Beobachtet: Zeitverhalten der Energien  $\rightarrow$  Bsp. 1.4.17





Energiedrift bei semi-impliziter Mittelpunktsregel



# 3.7 Exponentielle Integratoren [24, 28, 25]

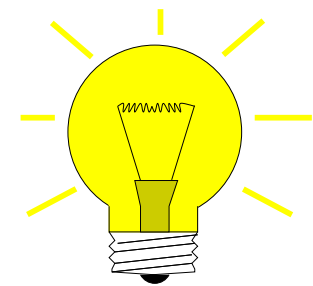
Betrachte: Autonomes AWP  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ ,  $\mathbf{f} : D \subset \mathbb{R}^d \mapsto \mathbb{R}^d$  stetig differenzierbar

Idee: „**Absubtrahieren**“ der Lösung der um  $\mathbf{y}_0$  linearisierten ODE

$$\dot{\mathbf{y}} = \mathbf{J}\mathbf{y} + \mathbf{g}(\mathbf{y}) \quad , \quad \mathbf{g}(\mathbf{y}) := \mathbf{f}(\mathbf{y}) - \mathbf{J}\mathbf{y} \quad , \quad (3.7.1)$$

mit

$$\mathbf{J} := D\mathbf{f}(\mathbf{y}_0)$$



Variation der Konstanten ( $\rightarrow$  Sect. 1.3.2) angewandt auf (3.7.1)

$$\mathbf{y}(h) = \exp(\mathbf{J}h)\mathbf{y}_0 + \int_0^h \exp(\mathbf{J}(h - \tau))\mathbf{g}(\mathbf{y}(\tau)) d\tau \quad . \quad (3.7.2)$$

Faltungsintegral

$\exp \hat{=} \mathbf{Matrixexponentialfunktion}$ , definiert durch, vgl. (1.3.14),

“Matrixexponentialreihe”:

$$\exp(\mathbf{M}) = \sum_{k=0}^{\infty} \frac{1}{k!} \mathbf{M}^k \quad .$$

! Auswertung der Matrixexponentialreihe ist kein stabiler numerischer Algorithmus (Auslöschung !)

## Alternativen:

- ☞ Pade-Approximation von  $t \mapsto e^t$  (nach Skalieren der Matrix, „scaling and squaring“)
- ☞ Schur-Zerlegung (siehe MATLAB-Kommando `schur`)  $\mathbf{M} = \mathbf{Q}^T(\mathbf{D} + \mathbf{U})\mathbf{Q}$  mit Diagonalmatrix  $\mathbf{D}$ , echter oberer Dreiecksmatrix  $\mathbf{U}$ , orthogonaler Matrix  $\mathbf{Q}$ . Anschliessend Auswertung der abgeschnittenen Matrixexponentialreihe für  $\mathbf{D} + \mathbf{U}$  & (1.3.15)

MATLAB-Funktion `expm`, Algorithmus  $\rightarrow$  [20]

Numerische Quadratur des Faltungsintegrals  $\triangleright$  Diskretisierung von (3.7.2)  $\triangleright$  ESV

R. Hiptmair  
rev 35327,  
24. Juni  
2011

Einfachste Wahl:

$$\int_0^h \exp(\mathbf{J}(h - \tau))g(\mathbf{y}(\tau)) d\tau \approx \int_0^h \exp(\mathbf{J}(h - \tau)) d\tau \cdot \mathbf{g}(\mathbf{y}_0) = h\varphi(\mathbf{J}h) \cdot \mathbf{g}(\mathbf{y}_0)$$

mit  $\varphi(z) = \frac{\exp(z) - 1}{z}$ .

▶ **exponentielles Euler-Verfahren** (auf Zeitgitter  $\{t_k\}$ ,  $h_k := t_{k+1} - t_k$ )

$$\mathbf{y}_{k+1} = \mathbf{y}_k + h_k \varphi(h_k \mathbf{J}) \mathbf{f}(\mathbf{y}_k), \quad k = 0, \dots, N, \quad \mathbf{J} := Df(\mathbf{y}_k). \quad (3.7.3)$$

*Bemerkung 3.7.4* (Stabilitätsgebiet des exponentiellen Euler-Verfahrens).

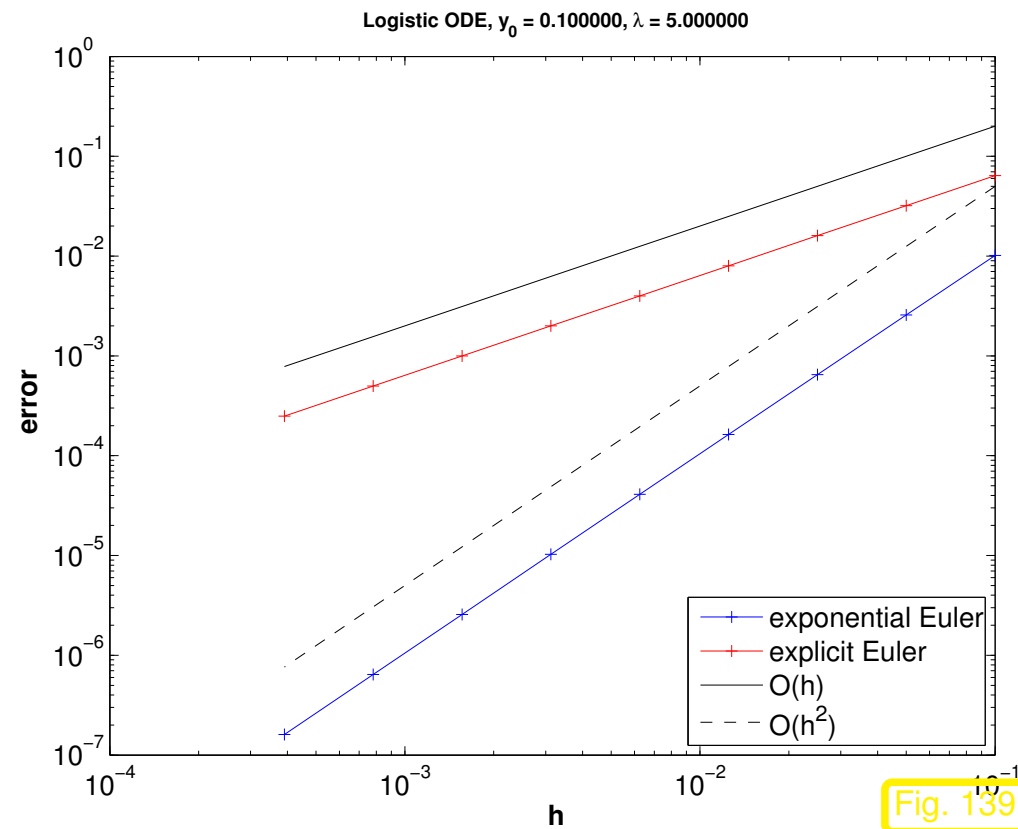
Erinnerung: Analyse des *linearen* Modellproblems, Sect. 3.1 ➤ Stabilitätsgebiet  $\mathcal{S}_\Psi \subset \mathbb{C} \rightarrow$   
Def. 3.1.4

Beachte: exponentielles Euler-Verfahren ist exakt für AWP zur ODE  $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y} + \mathbf{g}$  mit konstantem  $\mathbf{A} \in \mathbb{R}^{d,d}$ ,  $\mathbf{g} \in \mathbb{R}^d$ .

▶ (Ideale !) Stabilitätsfunktion:  $S(z) = \exp(z)$  ➤ (ideales) Stabilitätsgebiet  $\mathcal{S}_\Psi = \mathbb{C}^-$



*Beispiel 3.7.5* (Exponentielles Euler-Verfahren).



- Anfangswertproblem für  
logistische Differentialgleichung,  $\lambda = 5$ ,  $T = 1$ ,  
siehe Bsp. 1.2.1
- Exponentielles Euler-Verfahren (3.7.3) mit  
uniformen Zeitschrittweiten  $h$
- Fehlermass  $\text{err} = \max_{j=1, \dots, n} |y_j - y(t_j)|$

► Algebraische Konvergenz der **Ordnung 2** !



Beispiel 3.7.6 (Exponentielles Euler-Verfahren für steifes AWP).

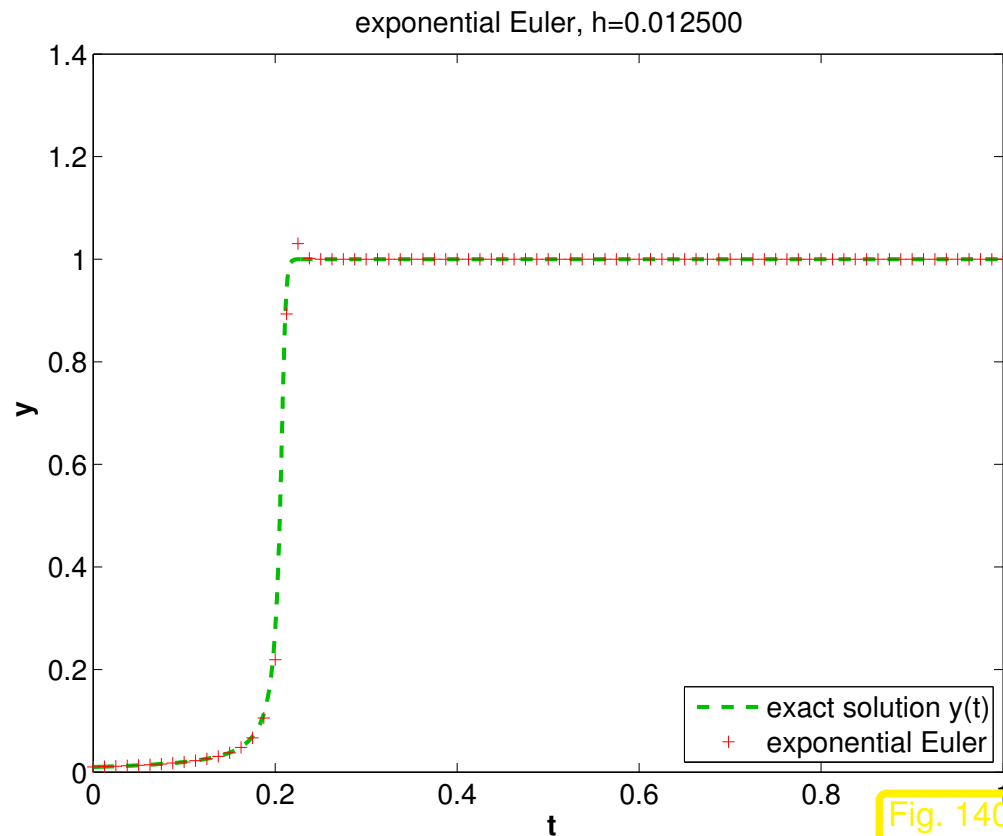
- Steifes AWP → Bsp. 3.5.2, 3.4.1, 3.5.7:

$$\dot{y}(t) = \lambda y^2(1 - y),$$

$$\lambda = 500, \quad y(0) = \frac{1}{100}.$$

- Exponentielles Euler-Verfahren (3.7.3) mit uniformen Zeitschrittweiten  $h = \frac{1}{80}$

► Qualitativ richtiges Verhalten



Verallgemeinerung: **Exponentielle Runge-Kutta-Verfahren:** mit  $\mathbf{J} := Df(\mathbf{y}_k)$

Semi-implizites Euler-Verfahren

Exponentielles Euler-Verfahren

$$\mathbf{y}_{k+1} = \mathbf{y}_k + (\mathbf{I} - h\mathbf{J})^{-1} h\mathbf{f}(\mathbf{y}_k)$$

$$\mathbf{y}_{k+1} = \mathbf{y}_k + \varphi(h\mathbf{J}) h\mathbf{f}(\mathbf{y}_k)$$



Ersetze:  $(\mathbf{I} - \gamma h\mathbf{J})^{-1} \rightarrow \varphi(\gamma h\mathbf{J})$  in linear-impliziten RK-ESV (3.6.7)



**Definition 3.7.7** (Exponentielle Runge-Kutta-Verfahren).

Für  $b_i, a_{ij}, d_{ij} \in \mathbb{R}$ ,  $i, j = 1, \dots, s$ ,  $s \in \mathbb{N}$ , definiert

$$\mathbf{k}_i := \varphi(a_{ii}h\mathbf{J}) \left( f(\mathbf{u}_i) + h\mathbf{J} \sum_{j=1}^{i-1} d_{ij}\mathbf{k}_j \right), \quad i = 1, \dots, s,$$

$$\mathbf{u}_i := \mathbf{y}_0 + h \sum_{j=1}^{i-1} (a_{ij} + d_{ij})\mathbf{k}_j, \quad i = 1, \dots, s,$$

$$\Psi^h \mathbf{y}_0 := \mathbf{y}_0 + h \sum_{i=1}^s b_i \mathbf{k}_i.$$

ein  $s$ -stufiges **exponentielles Runge-Kutta-Einschrittverfahren** für die autonome ODE  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ .

Dabei ist  $\mathbf{J} := D\mathbf{f}(\mathbf{y}_0)$ .

Wie in Sect. 2.3: Gewünschte Konsistenzordnung

- Bestimmungsgleichungen [24]
- Koeffizienten  $b_i, a_{ij}, c_{ij}, d_{ij}$

Zusatzbedingung: Exakte Integration von linearen Dgl.  $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y} + \mathbf{g}$ ,  $\mathbf{A} \in \mathbb{R}^{d,d}$ ,  $\mathbf{g} \in \mathbb{R}^d$

*Bemerkung* 3.7.8. Herausforderung: effiziente/genauere Berechnung von  $\exp(c_i h \mathbf{J})$

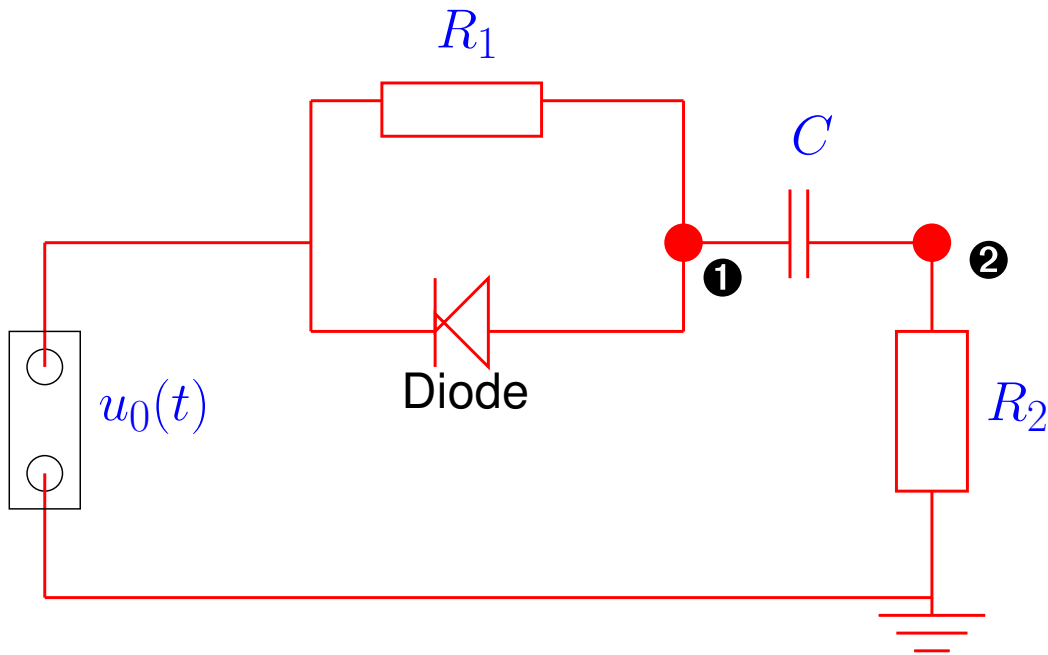
Krylov-Unterraummethoden für grosse, dünnbesetzte  $\mathbf{J} \rightarrow$  [22]



## 3.8 Differentiell-Algebraische Anfangswertprobleme

### 3.8.1 Grundbegriffe

*Beispiel* 3.8.1 (Knotenanalyse eines Schaltkreises).  $\rightarrow$  Bsp. 3.5.5



$$\textcircled{1}: 0 = I_D + I_{R_1} - I_C ,$$

$$\textcircled{2}: 0 = I_C - I_{R_2} .$$

Bauelementgleichungen:

$$I_D = I_D(u_1) = I_0(\exp(K(u_0 - u_1)) - 1) ,$$

$$I_C = C(\dot{u}_1 - \dot{u}_2) ,$$

$$I_{R_1} = R_1^{-1}(u_0 - u_1) ,$$

$$I_{R_2} = R_2^{-1}u_2 .$$

Vorgegeben: Zeitabhängige Eingangsspannung  $u_0 = u_0(t)$

$$\textcircled{1} \triangleright 0 = I_D(u_1) + R_1^{-1}(u_0 - u_1) - C(\dot{u}_1 - \dot{u}_2) ,$$

$$\textcircled{2} \triangleright 0 = C(\dot{u}_1 - \dot{u}_2) - R_2^{-1}(u_2 - u_0) .$$

$$\triangleright \begin{pmatrix} C & -C \\ -C & C \end{pmatrix} \begin{pmatrix} \dot{u}_1 \\ \dot{u}_2 \end{pmatrix} = \begin{pmatrix} I_D(u_1) + R_1^{-1}(u_0 - u_1) \\ -R_2^{-1}u_2 \end{pmatrix} \quad (3.8.2)$$

Singuläre Matrix !  $\rightarrow$  (3.8.2) ist **Differentiell-Algebraische Gleichung (DAE)**

Beachte:

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} C & -C \\ -C & C \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} .$$

► Transformation von (3.8.2):  $y_1 := u_1 - u_2, y_2 := u_2$

$$\begin{aligned} \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \dot{u}_1 \\ \dot{u}_2 \end{pmatrix} &= \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} = \begin{pmatrix} I_D(u_1) + R_1^{-1}(u_0 - u_1) \\ I_D(u_1) + R_1^{-1}(u_0 - u_1) - R_2^{-1}u_2 \end{pmatrix} \\ &= \begin{pmatrix} I_D(y_1 + y_2) + R_1^{-1}(u_0 - y_1 - y_2) \\ I_D(y_1 + y_2) + R_1^{-1}(u_0 - y_1 - y_2) - R_2^{-1}y_2 \end{pmatrix} . \end{aligned}$$

► *Algebraische Nebenbedingung*

$$c(y_1, y_2) := I_D(y_1 + y_2) + R_1^{-1}(u_0 - y_1 - y_2) - R_2^{-1}y_2 = 0 .$$

Beachte:  $\forall y_1: y_2 \mapsto c(y_1, y_2)$  monoton fallend,  $\lim_{y_2 \rightarrow \infty} c(y_1, y_2) = -\infty$ ,  $\lim_{y_2 \rightarrow -\infty} c(y_1, y_2) = \infty$

⇒ Nebenbedingung ist **auflösbar** nach  $y_2 = u_2$ :  $\exists$  Funktion  $G : \mathbb{R} \mapsto \mathbb{R}$  so, dass  $y_2 = G(y_1)$

Einsetzen

► **ODE für  $y_1$  !**

$$C\dot{y}_1 = I_D(y_1 + G(y_1)) + R_1^{-1}(u_0 - y_1 - G(y_1)) .$$



Gegeben: Rechte Seite  $\mathbf{f} : D \subset \mathbb{R}^d \mapsto \mathbb{R}^d$ ,  
singuläre Matrix  $\mathbf{M} \in \mathbb{R}^{d,d}$

☞ Autonomes (lineares) **differentiell-algebraisches Anfangswertproblem** (DAE):

$$\mathbf{M}\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}) \quad , \quad \mathbf{y}(0) = \mathbf{y}_0 \quad . \quad (3.8.3)$$

Beachte: (3.8.3) impliziert **algebraische Nebenbedingung**  $\mathbf{f}(\mathbf{y}(t)) \in \text{Im}(\mathbf{M})$

(**Konsistente** Anfangswerte erforderlich:  $\mathbf{f}(\mathbf{y}_0) \in \text{Im}(\mathbf{M})$  !)

☞ Notation:  $\text{Im}(\mathbf{M}) := \{\mathbf{M}\mathbf{x} : \mathbf{x} \in \mathbb{R}^d\} \hat{=} \text{Bild der Matrix } \mathbf{M}$

In Bsp. 3.8.1: Transformationen ➤ Reduktion auf spezielle Form:

Erweiterung von Def. 1.1.2 ➤ Lösungsbegriff für (3.8.3)

(Diffizil: Allgemeine Existenz & Eindeutigkeit von Lösungen, siehe [8, Sect. 2.6])

Gegeben: „Rechte Seiten“  $\mathbf{d} : D_1 \times D_2 \subset \mathbb{R}^p \times \mathbb{R}^q \mapsto \mathbb{R}^p$ ,  
 $\mathbf{c} : D_1 \times D_2 \subset \mathbb{R}^p \times \mathbb{R}^q \mapsto \mathbb{R}^q$  (hinreichend glatt),  $p, q \in \mathbb{N}$ ,

Anfangswerte  $\mathbf{u}_0 \in D_1, \mathbf{v}_0 \in D_2$ .

☞ Separiertes **differentiell-algebraisches Anfangswertproblem** (DAE):

Algebraische Nebenbedingung (engl. *constraint*)

$$\begin{aligned} \dot{\mathbf{u}} &= \mathbf{d}(\mathbf{u}, \mathbf{v}), \\ 0 &= \mathbf{c}(\mathbf{u}, \mathbf{v}), \end{aligned} \quad , \quad \begin{aligned} \mathbf{u}(0) &= \mathbf{u}_0, \\ \mathbf{v}(0) &= \mathbf{v}_0, \end{aligned} \quad , \quad \mathbf{c}(\mathbf{u}_0, \mathbf{v}_0) = 0. \quad (3.8.4)$$

Konsistente Anfangswerte erforderlich !

*Bemerkung 3.8.5* (DAE: Transformation auf separierte Form).

Die Form (3.8.3) einer DAE ist immer in (3.8.4) transformierbar:

$$\text{rank}(\mathbf{M}) = r \quad \Rightarrow \quad \exists \mathbf{T}, \mathbf{S} \in \mathbb{R}^{d,d} \text{ regulär: } \mathbf{TMS} = \begin{pmatrix} \mathbf{I} & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{I} \in \mathbb{R}^{r,r}.$$

Anwendung auf (3.8.3):

$$\mathbf{M}\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}) \quad \Rightarrow \quad \mathbf{TMS}\mathbf{S}^{-1}\dot{\mathbf{y}} = \mathbf{Tf}(\mathbf{y}) \quad \stackrel{\mathbf{z}:=\mathbf{S}^{-1}\mathbf{y}}{\Rightarrow} \quad \begin{pmatrix} \mathbf{I} & 0 \\ 0 & 0 \end{pmatrix} \dot{\mathbf{z}} = \mathbf{Tf}(\mathbf{S}\mathbf{z}).$$

Also definieren die ersten  $r$  Gleichungen des transformierten Systems die Differentialgleichung, während die restlichen  $d - r$  die Rolle der algebraischen Nebenbedingungen spielen.



**Annahme 3.8.6.** Partielle Ableitung (Jacobi-Matrix)  $D_{\mathbf{v}}\mathbf{c}(\mathbf{u}, \mathbf{v})$  der Nebenbedingungen ist regulär entlang von Lösungskurven  $t \mapsto (\mathbf{u}(t), \mathbf{v}(t))^T$ .

► Lokale Auflösbarkeit:  $\mathbf{v} = G(\mathbf{u}): (3.8.4) \Rightarrow \dot{\mathbf{u}} = \mathbf{d}(\mathbf{u}, G(\mathbf{u})) \hat{=} (1.1.13).$

**Definition 3.8.7** (DAE vom Index 1).

*Annahme (3.8.6) erfüllt  $\Rightarrow$  DAE-AWP (3.8.4) hat (Differenzierbarkeits)index 1*

*Bemerkung 3.8.8.* Allgemeine Diskussion des Indexbegriffes (Index  $> 1$ , Störungsindex, etc.) bei DAE: [18, Kap. VII]



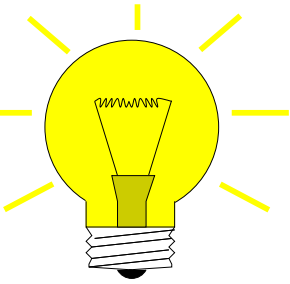
## 3.8.2 Runge-Kutta-Verfahren für Index-1-DAEs

Betrachte: differentiell-algebraisches Anfangswertproblem (3.8.4) unter Annahme 3.8.6



Idee:

# Singuläre Störungstechnik (engl. $\epsilon$ -embedding)



- ① Betrachte AWP's für Dgl.  $\begin{matrix} \dot{\mathbf{u}} = \mathbf{d}(\mathbf{u}, \mathbf{v}) \\ \epsilon \dot{\mathbf{v}} = \mathbf{c}(\mathbf{u}, \mathbf{v}) \end{matrix}, \quad \epsilon > 0.$
- ② Formuliere RK-ESV für  $\begin{matrix} \dot{\mathbf{u}} = \mathbf{d}(\mathbf{u}, \mathbf{v}) \\ \dot{\mathbf{v}} = \frac{1}{\epsilon} \mathbf{c}(\mathbf{u}, \mathbf{v}) \end{matrix}, \quad \epsilon > 0.$
- ③ Macht Verfahren noch Sinn  $\epsilon = 0$  ? Wenn ja  $\rightarrow$  😊

Beispiel 3.8.9 (Singular gestörte Schaltkreisgleichungen).

Schaltkreis aus Bsp. 3.8.1 mit **parasitärer Kapazität** (durchflossen vom Strom  $I_p$ )

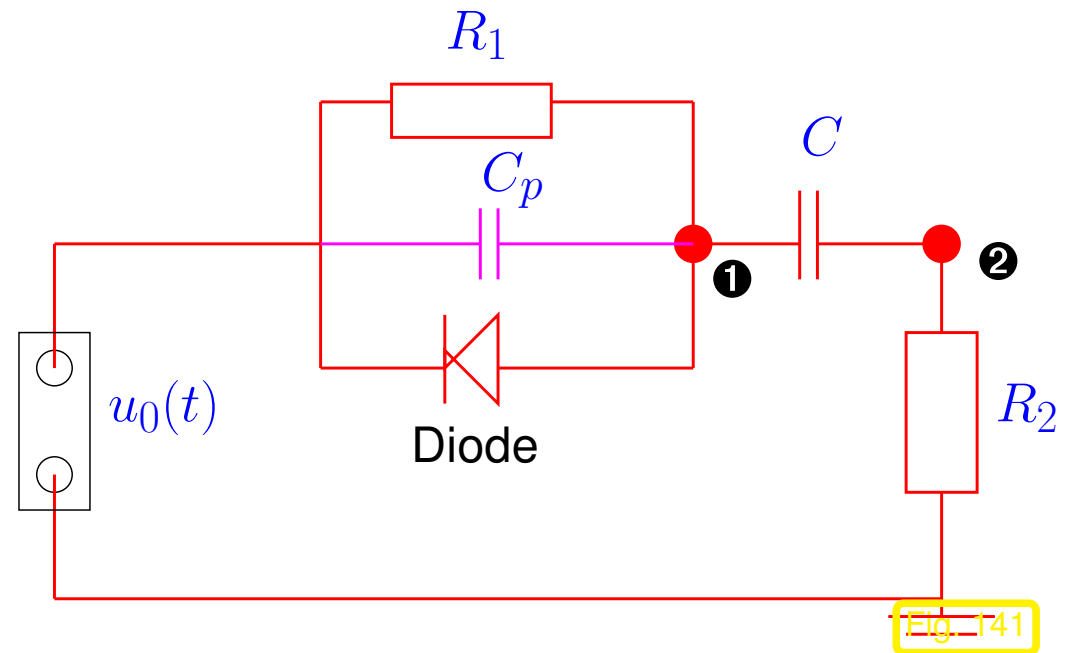
Knotengleichungen (Kirchhoffsche Regel):

①:  $0 = I_D + I_{R_1} + I_p - I_C$ ,

②:  $0 = I_C - I_{R_2}$ .

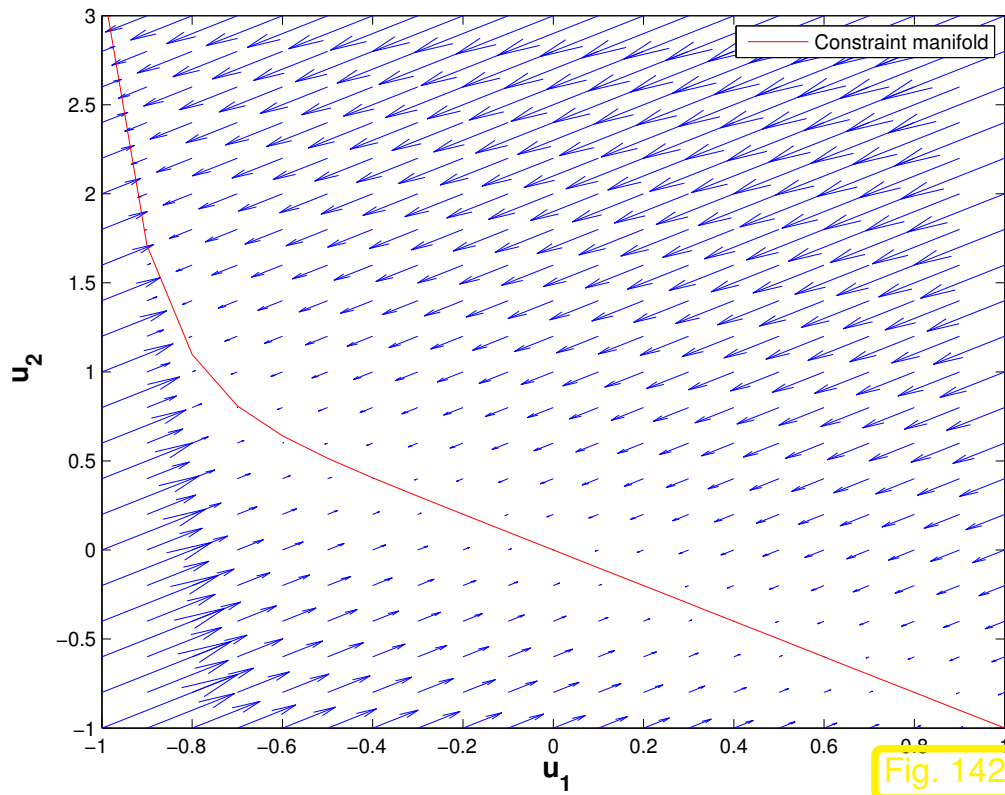
Zusätzliche Bauelementgleichung:

$$I_p = C_p(\dot{u}_0 - \dot{u}_1).$$



$$\blacktriangleright \begin{pmatrix} C+C_p & -C \\ -C & C \end{pmatrix} \begin{pmatrix} \dot{u}_1 \\ \dot{u}_2 \end{pmatrix} = \begin{pmatrix} I_D(u_1) + R_1^{-1}(u_0 - u_1) + C_p \dot{u}_0 \\ -R_2^{-1}u_2 \end{pmatrix} \quad (3.8.10)$$

Reguläre Matrix für  $C_p > 0$



◁ Richtungsfeld der singular gestörten Schaltkreisgleichung für  $u_0(t) \equiv 0$

(Skalierte Größen  $R = 1$ ,  $C = 1$ ,  $I_0 = 10^{-4}$ ,  
 $K = 10$ ,  $C_p = 10^{-3}$ )

Aus dem Richtungsfeld liest man ab: schnelle Relaxation in Richtung auf die Mannigfaltigkeit beschrieben durch die algebraische Nebenbedingung der DAE (3.8.2)

$$u_2 = R_2(I_D(u_1) + R_1^{-1})(u_0 - u_1) .$$

☞ **Steifheit** des singular gestörten Problems, siehe Bsp. 3.5.6.

Quantitative Analyse: Betrachte den Fall  $u_0(t) \equiv 0$  ➤ Stationärer Punkt:  $u_1 = 0, u_2 = 0$

Jacobi-Matrix im stationären Punkt

$$Df(0) = C_p^{-1} \begin{pmatrix} 1 & 1 \\ 1 & 1 + \frac{C_p}{C} \end{pmatrix} \begin{pmatrix} -I_0 K - R_1^{-1} & 0 \\ 0 & -R_2^{-1} \end{pmatrix}$$

➤  $C_p \rightarrow 0 \Rightarrow \lambda_{\min}(Df(0)) \rightarrow -\infty$

DAEs = „ $\infty$ -steife Anfangswertprobleme“



Butcher-Schema  $\frac{\mathbf{c} \mid \mathfrak{A}}{\mathbf{b}^T}$

Def. 2.3.5:  $s$ -stufiger Runge-Kutta-Schritt für  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ , Stufenform, siehe Bem. 2.3.7:

$$\begin{aligned} \mathbf{k}_i &= \mathbf{f}\left(\mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{k}_j\right) & \mathbf{g}_i &= \mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{f}(\mathbf{g}_j) \\ & \Leftrightarrow \mathbf{k}_i = \mathbf{f}(\mathbf{g}_i) & & , \quad i = 1, \dots, s . \\ \mathbf{y}_1 &= \mathbf{y}_0 + h \sum_{i=1}^s b_i \mathbf{k}_i & \mathbf{y}_1 &= \mathbf{y}_0 + h \sum_{i=1}^s b_i \mathbf{f}(\mathbf{g}_i) \end{aligned}$$

$$\begin{cases} \dot{\mathbf{u}} = \mathbf{d}(\mathbf{u}, \mathbf{v}) \\ \dot{\mathbf{v}} = \frac{1}{\epsilon} \mathbf{c}(\mathbf{u}, \mathbf{v}) \end{cases} \quad \blacktriangleright \quad \begin{cases} \mathbf{g}_i^u = \mathbf{u}_0 + h \sum_{j=1}^s a_{ij} \mathbf{d}(\mathbf{g}_j^u, \mathbf{g}_j^v) \\ \epsilon \mathbf{g}_i^v = \epsilon \mathbf{v}_0 + h \sum_{j=1}^s a_{ij} \mathbf{c}(\mathbf{g}_j^u, \mathbf{g}_j^v) \\ \mathbf{u}_1 = \mathbf{u}_0 + h \sum_{i=1}^s b_i \mathbf{d}(\mathbf{g}_i^u, \mathbf{g}_i^v) \\ \mathbf{v}_1 = \mathbf{v}_0 + \frac{1}{\epsilon} h \sum_{i=1}^s b_i \mathbf{c}(\mathbf{g}_i^u, \mathbf{g}_i^v) \end{cases} , \quad i = 1, \dots, s$$

$$\epsilon \rightarrow 0 \quad \Rightarrow \quad \sum_{j=1}^s a_{ij} \mathbf{c}(\mathbf{g}_j^u, \mathbf{g}_j^v) = 0, \quad i = 1, \dots, s \quad \Rightarrow \quad \boxed{\mathbf{c}(\mathbf{g}_j^u, \mathbf{g}_j^v) = 0}$$

Annahme: Koeffizientenmatrix  $\mathfrak{A} := (a_{ij})^s$  regulär

① Falls Nebenbedingung nach  $\mathbf{v}$  auflösbar ( $\rightarrow$  Index 1, Def. 3.8.7)

$$\mathbf{c}(\mathbf{u}_1, \mathbf{v}_1) = 0 \quad \Rightarrow \quad \mathbf{v}_1 = G(\mathbf{u}_1) .$$

Dies kann den Schritt  $\mathbf{v}_0 \mapsto \mathbf{v}_1$  ersetzen.

② Im allgemeinen Fall, formal, mit  $\mathfrak{A}^{-1} = (\check{a}_{ij})_{i,j=1}^s$

$$\frac{1}{\epsilon} h \mathbf{c}(\mathbf{g}_i^u, \mathbf{g}_i^v) = \sum_{j=1}^s \check{a}_{ij} (\mathbf{g}_j^v - \mathbf{v}_0)$$

$$\Rightarrow \quad \mathbf{v}_1 = \mathbf{v}_0 + \sum_{i=1}^s b_i \sum_{j=1}^s \check{a}_{ij} (\mathbf{g}_j^v - \mathbf{v}_0) = \underbrace{(1 - \mathbf{b}^T \mathfrak{A}^{-1} \mathbf{1})}_{= S(-\infty)} \mathbf{v}_0 + \sum_{j=1}^s \left( \sum_{i=1}^s b_i \check{a}_{ij} \right) \mathbf{g}_j^v .$$

Aus Formel (3.4.4) für Stabilitätsfunktion  $S$

Beachte:  $\mathbf{c}(\mathbf{u}_1, \mathbf{v}_1) = 0$  ist hier nicht garantiert !

Spezialfall: Wenn RK-ESV steif-genau, also  $\mathbf{b}^T = \mathbf{a}_{\cdot, s}^T$  (Zeile von  $\mathfrak{A}$ ), vgl. hinreichende Bedingung (3.4.5) für L-Stabilität

$$\Rightarrow \mathbf{v}_1 = \mathbf{g}_s^v \Rightarrow \mathbf{c}(\mathbf{u}_1, \mathbf{v}_1) = 0$$

*Bemerkung 3.8.11* (RK-ESV und Elimination der DAE-Nebenbedingungen).

**Index-1 DAE** ( $\rightarrow$  Def. 3.8.7)  $\leftrightarrow$  ODE  $\dot{\mathbf{u}} = \mathbf{d}(\mathbf{u}, G(\mathbf{u}))$

steif-genaues RK-ESV gemäss (3.4.5) für diese ODE:

$$\mathbf{g}_i^u = \mathbf{u}_0 + h \sum_{j=1}^s a_{ij} \mathbf{d}(\mathbf{g}_j^u, G(\mathbf{g}_j^u)), \quad i = 1, \dots, s, \quad \mathbf{u}_1 = \mathbf{g}_s^u$$



$$\begin{cases} \mathbf{g}_i^u = \mathbf{u}_0 + h \sum_{j=1}^s a_{ij} \mathbf{d}(\mathbf{g}_j^u, \mathbf{g}_j^v) \\ 0 = \mathbf{c}(\mathbf{g}_i^u, \mathbf{g}_i^v) \end{cases}, \quad i = 1, \dots, s, \quad \mathbf{u}_1 = \mathbf{g}_s^u.$$

☞ Ein “kommutierendes Diagramm”

Steif-genaues RK-ESV für  $\begin{cases} \dot{\mathbf{u}} = \mathbf{d}(\mathbf{u}, \mathbf{v}) \\ 0 = \mathbf{c}(\mathbf{u}, \mathbf{v}) \end{cases} =$  Steif-genaues RK-ESV für  $\dot{\mathbf{u}} = \mathbf{d}(\mathbf{u}, G(\mathbf{u}))$



Welche Runge-Kutta-Einschrittverfahren ( $\rightarrow$  Sect. 2.3)  
eignen sich für die singuläre Störungstechnik ?

- Notwendig (für Lösbarkeit der Inkrementgleichungen, Def. 2.3.5): implizites Verfahren
- DAE „ $\infty$ -steif“  $\triangleright$  Notwendig: A-Stabilität  $\rightarrow$  Def. 3.2.16,  
Wünschenswert: L-stabilität  $\rightarrow$  Def. 3.4.3

Wiederum: Einsichten durch Modellproblemanalyse, vgl. Sect. 3.1:

Modellproblem: 
$$\begin{aligned} \dot{u} &= v f(u), \\ 0 &= 1 - v \end{aligned} \quad \blacktriangleright \quad \begin{aligned} \dot{u} &= v f(u), \\ \epsilon \dot{v} &= 1 - v \end{aligned}, \quad 0 < \epsilon \ll 1.$$

**Singulär gestörte Dgl.**

Heuristik: ESV tauglich für Modellproblem  $\leftrightarrow$  ESV tauglich für singulär gestörtes Problem  $\forall \epsilon \approx 0$

$\blacktriangleright$  ESV muss für AWP  $\dot{v} = \epsilon^{-1}(1 - v), v(0) = 1, 0 < \epsilon \ll 1$ , Folge  $\{v_k\}_{k=0}^{\infty}$  mit  $\lim_{k \rightarrow \infty} v_k = 1$   
liefern !

Erwünscht:

$$S(-\infty) = 0$$

für Stabilitätsfunktion ( $\rightarrow$  Thm. 3.1.6)  $S(z)$  des ESV.

**Theorem 3.8.12** (Konvergenz impliziter RK-ESV für Index-1-DAEs).  $\rightarrow$  [18, Thm. 1.1, Sect. VI.1]

Es seien  $\mathbf{d}$ ,  $\mathbf{c}$  hinreichend glatt, das Anfangswertproblem (3.8.4) eindeutig lösbar und Annahme 3.8.6 erfüllt ( $\rightarrow$  Index-1-DAE, siehe Def. 3.8.7). Das  $s$ -stufige

Runge-Kutte-Einschrittverfahren mit Butcher-Tableau  $\begin{array}{c|c} \mathbf{c} & \mathfrak{A} \\ \hline & \mathbf{b}^T \end{array}$ , siehe (2.3.6), sei steif-genau, d.h.

$\mathfrak{A}$  ist regulär und  $b_i = a_{s,i}$ ,  $i = 1, \dots, s$ , und sei konsistent von der Ordnung  $p$ .

Dann ist das Verfahren angewandt auf (3.8.4) für hinreichend kleine Zeitschrittweite  $h$  wohldefiniert und es gilt

$$\|\mathbf{u}_k - \mathbf{u}(t_k)\| = O(h^p) \quad , \quad \|\mathbf{v}_k - \mathbf{v}(t_k)\| = O(h^p) \quad ,$$

auf jedem endlichen Integrationszeitintervall  $[0, T]$ .




*Beweisskizze:* Auflösen von (3.8.4) nach der algebraischen Variablen  $\mathbf{v}$  und Einsetzen  $\triangleright$  ODE

Anwendung des RK-ESV auf die resultierende ODE liefert genau die gleiche diskrete Evolution fuer  $\mathbf{u}$  wie das Verfahren für die DAE (“kommutierendes Diagramm”), vgl. Bem. 3.8.11.



Numerische Integration von Index-1-DAEs mit Radau-ESV  $\rightarrow$  Sect. 3.4

*Bemerkung 3.8.13.* Radau-ESV auch geeignet für Index-1-DAEs (!) in der Form (3.8.3)  $\mathbf{M}\dot{\mathbf{y}} = f(\mathbf{y})$  

R. Hiptmair  
rev 35327,  
24. Juni  
2011

*Bemerkung 3.8.14* (MATLAB-Integratoren für Index-1-DAEs).

MATLAB-Code zur Lösung allgemeiner  
DAEs

$$\mathbf{M}(t, \mathbf{y})\dot{\mathbf{y}} = f(t, \mathbf{y}) .$$

Funktion  $f$ Jacobi-Matrix  $D_{\mathbf{y}}f$ Matrix(funktion)  $\mathbf{M}(t, \mathbf{y})$ 

MATLAB-CODE : Lösung einer Index-1-DAE

```
f = @(t,y) [ ... ];
J = @(t,y) [ ... ];
M = @(t,y) [ ... ];
opts = odeset('Mass',M,'Jacobian',J);
[t,y] = ode15s(f,tspan,y0,opts);
```

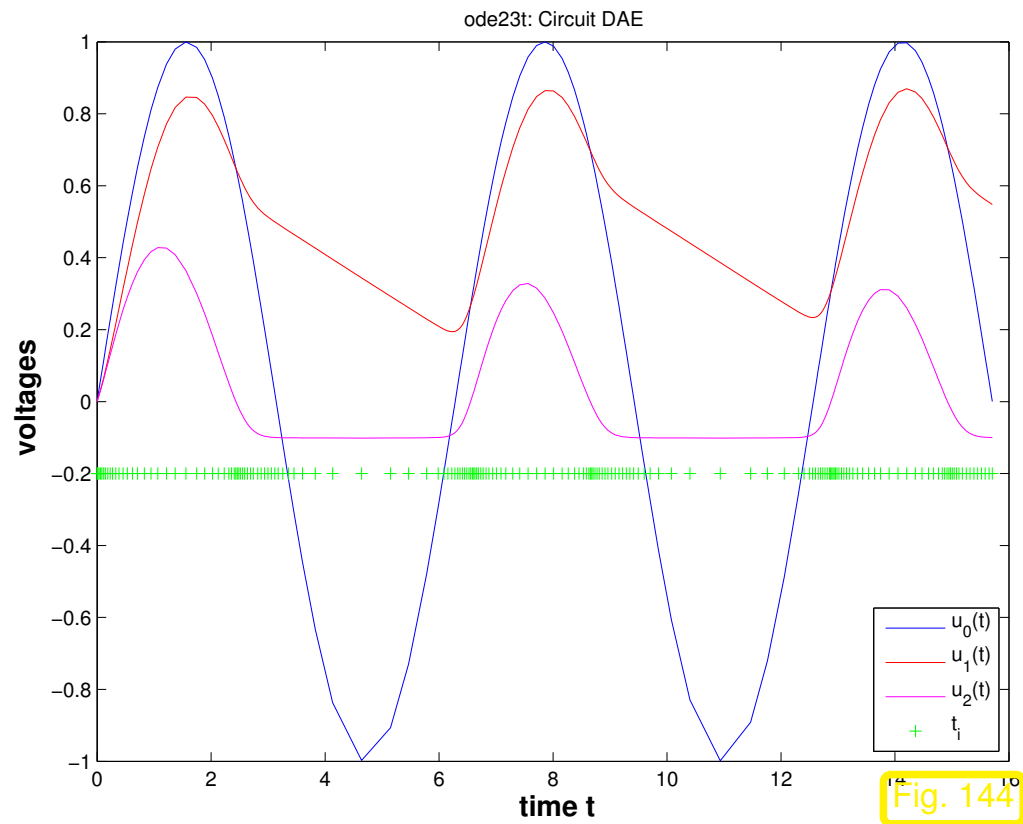
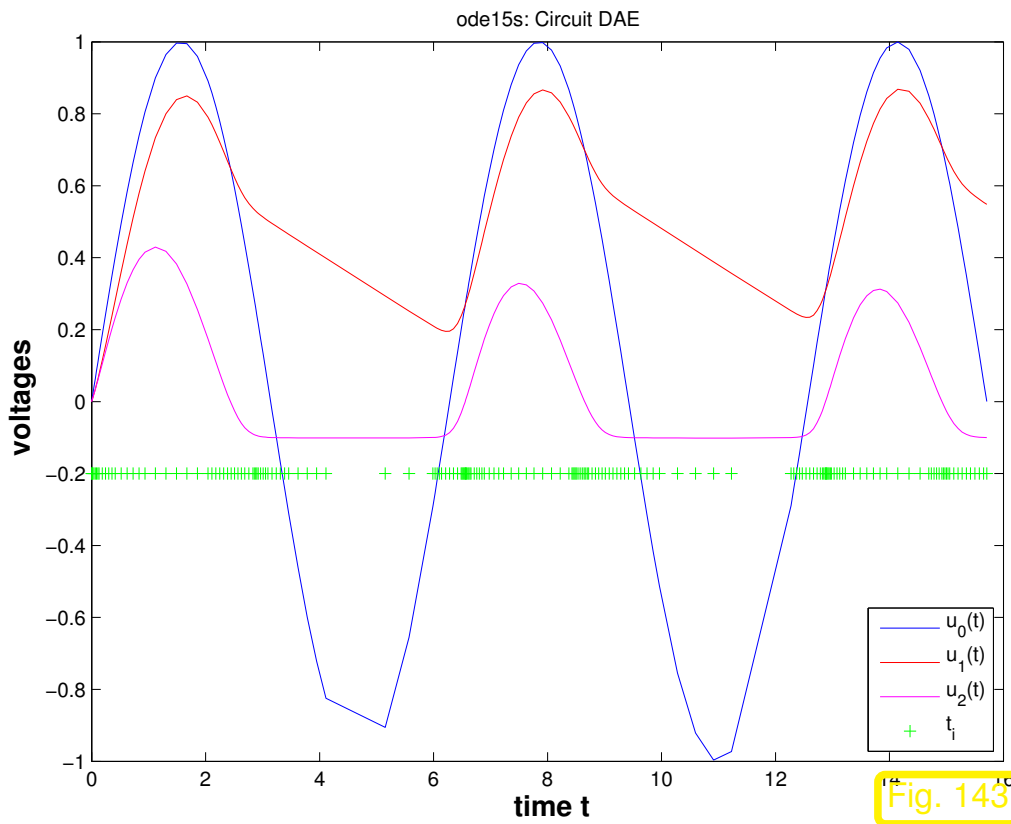
Alternativer Integrator: `ode23t` (gleicher Aufruf)

(Beachte: Alle MATLAB DAE-Integratoren benutzen adaptive Schrittweitensteuerung, vgl. Sect. 2.6)

R. Hiptmair

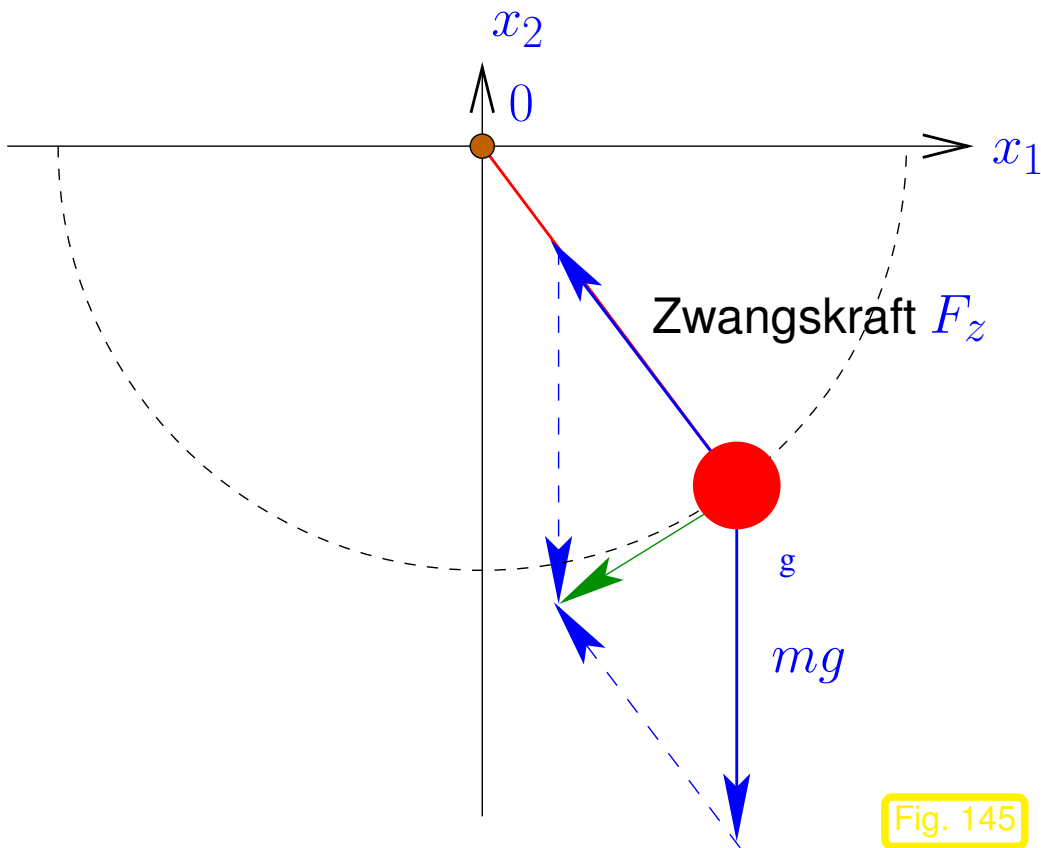
rev 35327,  
24. Juni  
2011*Beispiel 3.8.15* (Lösung der Schaltkreis-DAEs mit MATLAB).

- DAE (3.8.2) mit  $C = 1$ ,  $R_2 = 1$ ,  $R_1 = 1000$ ,  $K = 10$ ,  $I_0 = 10^{-4}$
- MATLAB-Integratoren `ode23t`, `ode15s`, default Toleranzen



### 3.8.3 DAEs mit höherem Index

Beispiel 3.8.16 (Pendelgleichung in Deskriptorform). → Bsp. 1.2.17



Mathematisches Pendel (Aufhängung in 0)

Zwangskraft (Zug des Stabs) hält Pendelmasse auf der Kreisbahn

$$F_z(\mathbf{x}) - mg \begin{pmatrix} 0 \\ 1 \end{pmatrix} \perp \mathbf{x} . \quad (3.8.17)$$

Zwangskraft in Richtung des Stabes:

$$F_z(\mathbf{x}) = -\lambda m \mathbf{x} . \quad (3.8.18)$$

$(\mathbf{x} = (x_1, x_2)^T \hat{=} \text{Position der Masse})$

Fig. 145

► Bewegungsgleichungen in Deskriptorform: ( $\leftrightarrow$  Minimalkoordinaten in Bsp. 1.2.17)

$$\ddot{\mathbf{x}} = -\lambda \mathbf{x} - g \begin{pmatrix} 0 \\ 1 \end{pmatrix} , \quad \boxed{x_1^2 + x_2^2 = l^2} . \quad (3.8.19)$$

Lagrange-Multiplikator  $\triangleright$  Zwangskraft

Zwangsbedingung

(3.8.19)  $\hat{=}$  2. Ordnung  $\triangleright$  Umwandlung in separierte DAE (3.8.4) 1. Ordnung:

$$(3.8.19) \triangleright m \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{p}_1 \\ \dot{p}_2 \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ -\lambda x_1 \\ -\lambda x_2 - g \end{pmatrix}, \quad x_1^2 + x_2^2 = l^2. \quad (3.8.20)$$

$(3.8.20)$  ist DAE vom Index  $> 1$  !

Differenzieren der Zwangsbedingung  $\blacktriangleright$   $x_1 p_1 + x_2 p_2 = 0$ , (3.8.21)

Nochmaliges Differenzieren  $\blacktriangleright$   $(p_1^2 + p_2^2) - \lambda(x_1^2 + x_2^2) - g x_2 = 0$ . (3.8.22)

- (3.8.21), (3.8.22)  $\hat{=}$  **versteckte Nebenbedingungen** (von den Anfangswerten zu erfüllen !)
- Erst nach zweimaligem Differenzieren der Nebenbedingung können wir die daraus resultierende Nebenbedingung (3.8.22) nach  $\lambda$  auflösen  $\blackrightarrow$  (3.8.20) hat Index 3

**Bemerkung 3.8.23** (Hamiltonsche Bewegungsgleichungen mit Nebenbedingungen).

Betrachte: Mechanisches System mit *Hamilton-Funktion* ( $\rightarrow$  Def. 1.2.20)  $H = H(\mathbf{p}, \mathbf{q})$   
 ( $\mathbf{q} \in \mathbb{R}^n \hat{=}$  Konfigurationsvariable,  $\mathbf{p} \in \mathbb{R}^n \hat{=}$  Impulsvariable, siehe Sect. 1.2.4)

Zwangsbedingungen für Konfigurationen:

$$\mathbf{c}(\mathbf{q}(t)) = 0 \quad \forall t \geq 0, \quad \mathbf{c} : \mathbb{R}^n \mapsto \mathbb{R}^m, \quad m < n$$

► Hamiltonsche Bewegungsgleichungen mit Zwangsbedingungen

$$\begin{aligned} \dot{\mathbf{q}} &= \frac{\partial H}{\partial \mathbf{p}}(\mathbf{p}, \mathbf{q}) \\ \dot{\mathbf{p}} &= -\frac{\partial H}{\partial \mathbf{q}}(\mathbf{p}, \mathbf{q}) - D\mathbf{c}(\mathbf{q})^T \boldsymbol{\lambda} \\ 0 &= \mathbf{c}(\mathbf{q}) . \end{aligned} \tag{3.8.24}$$

Lagrange-Multiplikator  $\boldsymbol{\lambda} : \mathbb{R} \mapsto \mathbb{R}^m$

Falls  $m = 1 \quad \triangleright \quad c(\mathbf{q}) = 0$  beschreibt  $n - 1$ -dimensionale Mannigfaltigkeit im  $\mathbb{R}^n$

$\mathbf{grad} c(\mathbf{q}) \hat{=}$  Normalenvektor auf diese Mannigfaltigkeit

$\lambda \mathbf{grad} c(\mathbf{q}) \hat{=}$  Zwangskraft *orthogonal* zur Mannigfaltigkeit, vgl. (3.8.17)

Häufiger Spezialfall: mit  $\mathbf{M} \hat{=} \mathbf{s.p.d.}$  Massenmatrix,  $U \hat{=} \text{Potential}$ ,

$$H(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \mathbf{p}^T \mathbf{M}^{-1} \mathbf{p} + U(\mathbf{q}) \quad (3.8.25)$$

kinetische Energie
potentielle Energie

(3.8.24) in diesem Spezialfall:

$$\dot{\mathbf{q}} = \mathbf{M}^{-1} \mathbf{p}, \quad \dot{\mathbf{q}} = -\text{grad } U(\mathbf{q}) - D\mathbf{c}(\mathbf{q})^T \boldsymbol{\lambda}, \quad \mathbf{c}(\mathbf{q}) = 0.$$

Nun: Differentiation der Zwangsbedingung  $\mathbf{c}(\mathbf{q}) = 0$  nach der Zeit + Anwenden der Produktregel & Kettenregel + Einsetzen der Differentialgleichungen aus (3.8.24).

➤ **Versteckte Nebenbedingungen**  $\leftrightarrow$  (3.8.21), (3.8.22)

$$0 = D\mathbf{c}(\mathbf{q}) \frac{\partial H}{\partial \mathbf{p}}(\mathbf{p}, \mathbf{q}), \quad (3.8.26)$$

$$0 = D^2\mathbf{c}(\mathbf{q}) \left( \frac{\partial H}{\partial \mathbf{p}}(\mathbf{p}, \mathbf{q}), \frac{\partial H}{\partial \mathbf{p}}(\mathbf{p}, \mathbf{q}) \right) + D\mathbf{c}(\mathbf{q}) \frac{\partial^2 H}{\partial \mathbf{p} \partial \mathbf{q}}(\mathbf{p}, \mathbf{q}) \frac{\partial H}{\partial \mathbf{p}}(\mathbf{p}, \mathbf{q}) \\ - D\mathbf{c}(\mathbf{q}) \frac{\partial^2 H}{\partial^2 \mathbf{p}}(\mathbf{p}, \mathbf{q}) \left( \frac{\partial H}{\partial \mathbf{q}}(\mathbf{p}, \mathbf{q}) + D\mathbf{c}(\mathbf{q})^T \boldsymbol{\lambda} \right). \quad (3.8.27)$$

Für  $H$  der Form (3.8.25):  $\frac{\partial^2 H}{\partial \mathbf{p}^2} = \mathbf{M}^{-1}$  (invertierbare Matrix)

➤ (3.8.27) nach  $\lambda$  auflösbar, wenn  $D\mathbf{c}(\mathbf{q})^T$  injektiv  $\Leftrightarrow D\mathbf{c}(\mathbf{q})$  hat vollen Rang (entlang der Lösungstrajektorie): Zwangsbedingungen unabhängig.



*Beispiel 3.8.28* (MATLAB-Integratoren für Pendelgleichung in Deskriptorform).

MATLAB `ode15s/ode23t` angewandt auf (3.8.20):

```
??? Error using ==> funfun/private/daeic12 at 77
```

```
This DAE appears to be of index greater than 1.
```

```
Error in ==> ode15s at 394
```

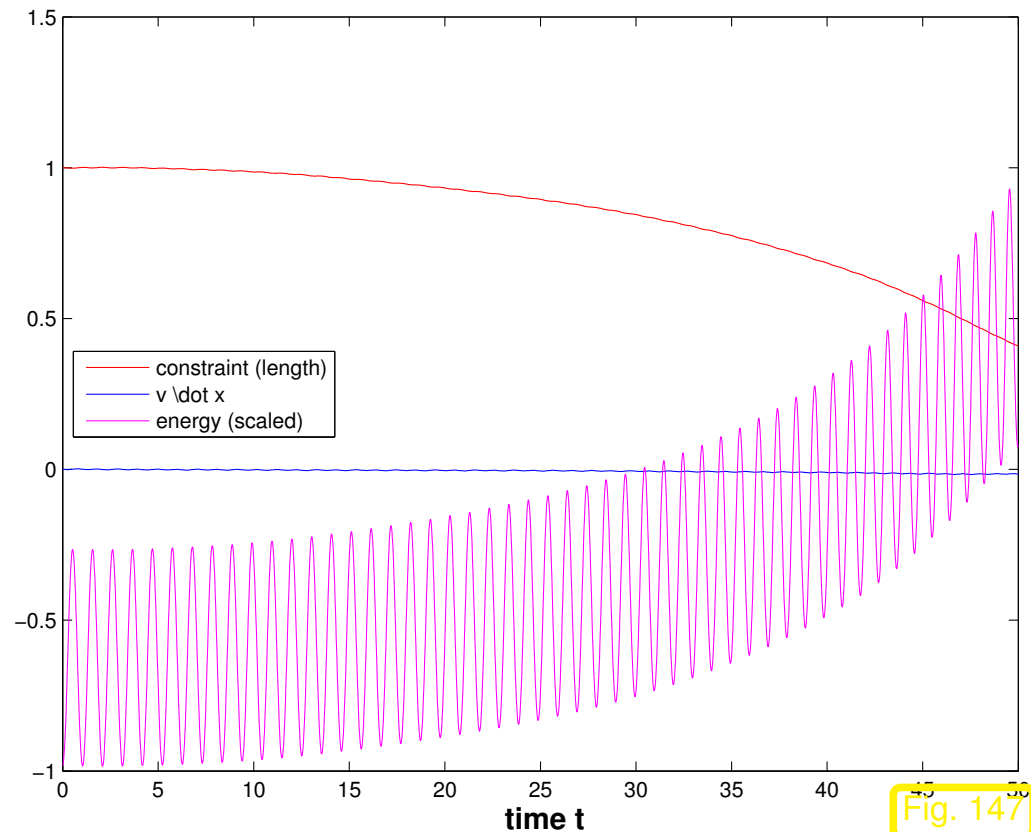
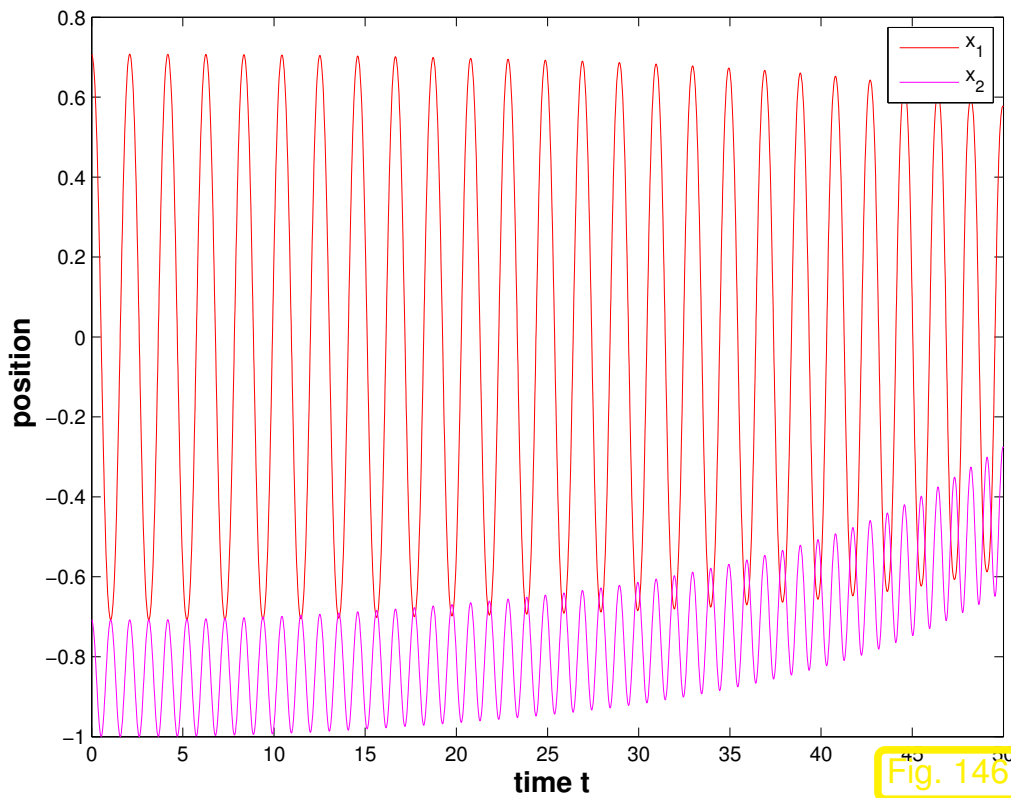
```
[y, yp, f0, dfdy, nFE, nPD, Jfac] = daeic12(odeFcn, odeArgs, ...)
```



Idee: „Überliste MATLAB“, probiere Nebenbedingung (3.8.22)

$$M\dot{y} = f(y): \quad M = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad f(y) = \begin{pmatrix} y_3 \\ y_4 \\ -y_5 y_1 \\ -y_5 y_2 - g \\ -y_5(y_1^2 + y_2^2) - g y_2 + (y_3^2 + y_4^2) \end{pmatrix}.$$

- $l = 1, g = 9.8$ , Zeitspanne  $[0, 50]$ , Löser: ode15s mit default-Toleranzen
- Konsistente Anfangswerte  $x_1(0) = -x_2(0) = \frac{1}{2}\sqrt{2}, p_1(0) = p_2(0) = 0$  ( $\rightarrow \lambda(0)$ )



*Beispiel 3.8.29* (Implizites Euler-Verfahren für Pendelgleichung in Deskriptorform).

- AWP für Pendel-DAE (3.8.20) wie in Bsp. 3.8.28, Endzeitpunkt  $T = 5$
- Implizites Eulerverfahren (1-stufiges Radau-ESV)

Formale Anwendung eines Rückwärtsdifferenzenquotienten ( $\rightarrow$  Bem. 1.4.14) auf die Hamiltonsche Bewegungsgleichung mit Zwangsbedingungen (3.8.24): berechne  $(\mathbf{q}_1, \mathbf{p}_1, \lambda_1)$  aus  $(\mathbf{q}_0, \mathbf{p}_0, \lambda_0)$  gemäss

$$\mathbf{q}_1 = \mathbf{q}_0 + h \frac{\partial H}{\partial \mathbf{p}}(\mathbf{p}_1, \mathbf{q}_1)$$

$$\mathbf{p}_1 = \mathbf{p}_0 - h \frac{\partial H}{\partial \mathbf{q}}(\mathbf{p}_1, \mathbf{q}_1) - h D\mathbf{c}(\mathbf{q}_1)^T \boldsymbol{\lambda}_1$$

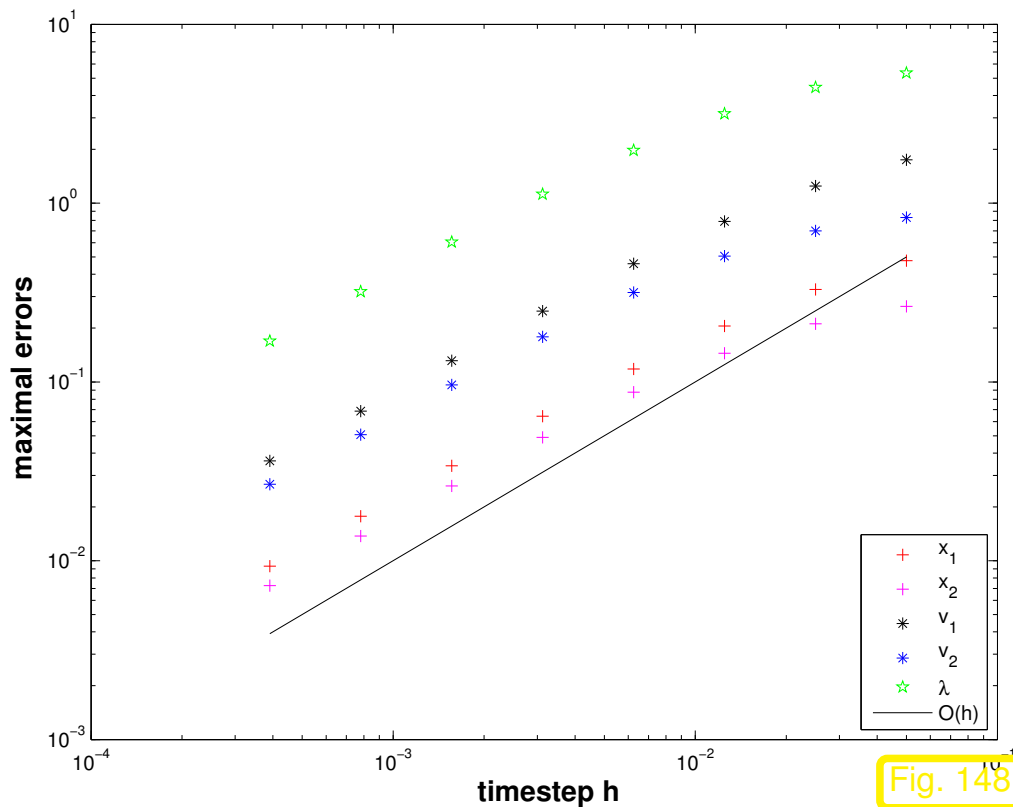
$$0 = \mathbf{c}(\mathbf{q}_1) .$$

Konkret für Pendelgleichung in Deskriptorform (3.8.20),  $\mathbf{q} \leftrightarrow \mathbf{x}$ ,  $H(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \|\mathbf{p}\|^2 + gx_2$ :

$$\mathbf{x}_1 = \mathbf{x}_0 + h\mathbf{p}_1,$$

$$\mathbf{p}_1 = \mathbf{p}_0 - h \left( \lambda \mathbf{x}_1 + \begin{pmatrix} 0 \\ g \end{pmatrix} \right),$$

$$0 = \|\mathbf{x}_1\|_2^2 - l^2.$$



◁ Fehler in Lösungskomponenten (diskrete Maximumnorm) für 100, 200, 400, 1600, 3200, 6400, 12800 implizite Euler-Schritte

(„Exakte Lösung berechnet mit impliziter Mittelpunktsregel angewandt auf Minimalkoordinatenform, siehe Bsp. 1.4.24)



Sect. 3.8.2: **Singuläre Störungstechnik** für Runge-Kutte-Einschrittverfahren

Anwendung auf autonome DAE (Index  $> 1!$ )

$$\begin{aligned}\dot{\mathbf{y}} &= \mathbf{f}(\mathbf{y}, \boldsymbol{\lambda}) , \\ 0 &= \mathbf{c}(\mathbf{y}) .\end{aligned}\tag{3.8.30}$$

( $\mathbf{f} : D \times \mathbb{R}^q \mapsto \mathbb{R}^d$ ,  $D \subset \mathbb{R}^d$ ,  $\mathbf{c} : D \mapsto \mathbb{R}^q$ , Annahme: Konsistente Anfangswerte  $\mathbf{y}_0, \boldsymbol{\lambda}_0$  zur Zeit  $t = 0$ )

R. Hiptmair  
rev 35327,  
24. Juni  
2011

Zu (3.8.30) gehörendes singular gestörtes Problem

$$\begin{aligned}\dot{\mathbf{y}} &= \mathbf{f}(\mathbf{y}, \boldsymbol{\lambda}) , \\ \epsilon \dot{\boldsymbol{\lambda}} &= \mathbf{c}(\mathbf{y}) ,\end{aligned}$$

für  $\epsilon \rightarrow 0$ .

## Sect. 3.8.2 ➤ Betrachte steif-genaue RK-ESV

Formale Rechnung: Singuläre Störungstechnik für Runge-Kutta-Einschrittverfahren mit

Butcher-Schema  $\frac{\mathbf{c}}{\mathbf{b}^T}$ ,  $b_i = a_{s,i}$ ,  $i = 1, \dots, s$ :

$$\begin{cases} \dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}, \boldsymbol{\lambda}) \\ \dot{\boldsymbol{\lambda}} = \frac{1}{\epsilon} \mathbf{c}(\mathbf{y}) \end{cases} \quad \blacktriangleright \quad \begin{cases} \mathbf{g}_i = \mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{f}(\mathbf{g}_j, \mathbf{g}_j^\lambda) \\ \epsilon \mathbf{g}_i^\lambda = \epsilon \boldsymbol{\lambda}_0 + h \sum_{j=1}^s a_{ij} \mathbf{c}(\mathbf{g}_j) \end{cases}, \quad i = 1, \dots, s$$

$$\begin{cases} \mathbf{y}_1 := \mathbf{y}_0 + h \sum_{i=1}^s a_{s,i} \mathbf{f}(\mathbf{g}_i, \mathbf{g}_i^\lambda) = \mathbf{g}_s \\ \mathbf{v}_1 := \mathbf{v}_0 + \frac{1}{\epsilon} h \sum_{i=1}^s a_{s,i} \mathbf{c}(\mathbf{g}_i) = \mathbf{g}_s^\lambda \end{cases}$$

$$\epsilon \rightarrow 0 \Rightarrow \sum_{j=1}^s a_{ij} \mathbf{c}(\mathbf{g}_j) = 0, \quad i = 1, \dots, s \Rightarrow \mathbf{c}(\mathbf{g}_j) = 0$$

RK-ESV steif-genau  $\Rightarrow$  Koeffizientenmatrix  $\mathfrak{A} := (a_{ij})_{i,j=1}^s$  regulär

➤ **Stufengleichungen** (→ Bem. 2.3.7) für steif-genaues RK-ESV für (3.8.30)

$$\begin{cases} \mathbf{g}_i = \mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{f}(\mathbf{g}_j, \mathbf{g}_j^\lambda) \\ 0 = \mathbf{c}(\mathbf{g}_i) \end{cases}, \quad i = 1, \dots, s, \quad \blacktriangleright \quad \mathbf{y}_1 = \mathbf{g}_s. \quad (3.8.31)$$

- Beachte:
- (3.8.31)  $\hat{=}$  implizite Gleichung für Stufen  $\mathbf{g}_i, \mathbf{g}_i^\lambda$  ( $s(d+q)$  Unbekannte)
  - $\lambda_0$  wird nicht benötigt!

*Bemerkung 3.8.32* (Implementierung steif-genauer RK-ESV für DAE).

Formal: Stufengleichungen eines RK-ESV für *allgemeine* autonome DAE  $\mathbf{M}\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ ,  $\mathbf{M} \in \mathbb{R}^{d,d}$  singulär

$$\mathbf{M}\mathbf{g}_i = \mathbf{M}\mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{f}(\mathbf{g}_j), \quad i = 1, \dots, s.$$

$$\Leftrightarrow$$

$$F(\mathbf{g}) = 0, \quad F(\mathbf{g}) = \begin{pmatrix} \mathbf{M}(\mathbf{g}_1 - \mathbf{y}_0) - h \sum_{j=1}^s a_{1j} \mathbf{f}(\mathbf{g}_j) \\ \vdots \\ \mathbf{M}(\mathbf{g}_s - \mathbf{y}_0) - h \sum_{j=1}^s a_{sj} \mathbf{f}(\mathbf{g}_j) \end{pmatrix}, \quad \mathbf{g} = \begin{pmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_s \end{pmatrix}. \quad (3.8.33)$$

► (steif-genau !)  $y_1 = g_s$

MATLAB-CODE : steif-genaues RK-ESV für DAE

```
function y1 = rksadaestep(rhs,M,y0,h,A)
    d = length(y0);
    s = size(A,1);
    F = @(gv) stagefn(gv,y,h,fun,A,M);
    [dgv,r,flg] = fsolve(F,zeros(s*d,1));
    if (flg <= 0), error('fsolve'); end
    y1 = dgv((s-1)*d+1:s*d);
end
```

◁ Einzelschritt: steifd-genaues RK-ESV für  $M\dot{y} = f(y)$

Koeffizientenmatrix  $\mathfrak{A} \in \mathbb{R}^{s,s}$

fsolve aus MATLAB "optimization toolbox"

Extrahiere  $s$ . Stufe  $g_s$

MATLAB-CODE : Stufen für steif-genaues RK-ESV

```
function dgv1 = stagefn(dgv0,y,h,fun,A,M);
    s = size(A,1); d = length(y);
    GV = reshape(dgv0,d,s);
    dgv1 = kron(eye(s),M)*dgv0;
    for j=1:s
        fg = feval(fun,y+GV(:,j));
        dgv1 = dgv1-h*kron(A(:,j),eye(d))*fg;
    end
```

◁ Auswertung von  $F(g - \eta_0)$ , siehe (3.8.33)



Konvergenztheorie für (3.8.31) im Fall von DAEs *mit Index 2*: [18, Sect. VII.4]

Für  $s$ -stufig Radau-ESV:  $\mathbf{y}_h(t)$  konvergiert mit Ordnung  $2s - 1$ ,  $\boldsymbol{\lambda}_h(t)$  mit Ordnung  $s$ .



## 4

# Strukturerhaltende numerische Integration

Struktur = **essentielle** *qualitative* Eigenschaften einer Evolution (☞ Sect. 1.3.3.5)

- Erste Integrale/Invarianten ( $\rightarrow$  Def. 1.2.7), z.B. Gesamtenergie, Drehmoment, siehe Sect. 1.2.4
- Anziehende und abstossende Fixpunkte, siehe 3.2
- Nichtexpansivität, siehe 3.3
- NEU: spezielle Abbildungseigenschaften des Flusses zu einer autonomen Dgl.:
  - ☞ Volumenerhaltung, Symplektizität, etc.

Ziel: „Vererbung“ von Struktur an diskrete Evolution  $\Psi^h$

Wichtig für **Langzeitintegration** zur Berechnung einer „qualitativ richtigen Lösung“

Perspektive: **Rückwärtsanalyse**

Numerische Lösung ist exakte Lösung zu einem „strukturell gleichen Problem“ mit leicht gestörten Anfangsdaten/Parametern

Beachte: In diesem Kapitel beschränken wir uns auf autonome Differentialgleichungen.

## 4.1 Polynomiale Invarianten

Erinnerung an Def. 1.2.7 & (1.2.8): Konzept und Eigenschaften von Invarianten/ersten Integralen

Beispiele: Massenerhaltung  $\rightarrow$  Sect. 1.2.2, Energieerhaltung  $\rightarrow$  Lemma 1.2.23, Längenerhaltung  
Bsp. 1.4.18

R. Hiptmair  
rev 35327,  
24. Juni  
2011

Betrachte: AWP für autonome ODE  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$  auf Zustandsraum  $D \subset \mathbb{R}^d$

$t \mapsto \mathbf{y}(t) \hat{=}$  Lösung zum Anfangswert  $\mathbf{y}_0 \in D$

Erinnerung: **erstes Integral**  $I : D \mapsto \mathbb{R}$  erfüllt  $I(\mathbf{y}(t)) = \text{const}$  für *jede* Lösung  $\mathbf{y}(t)$  ( $\rightarrow$  Def. 1.2.7)

(1.2.8):  $I$  ist erstes Integral von  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}) \Leftrightarrow \text{grad } I(\mathbf{y}) \cdot \mathbf{f}(\mathbf{y}) = 0$  für alle  $\mathbf{y} \in D$ .

lineares erstes Integral :  $I(\mathbf{y}) = \mathbf{b}^T \mathbf{y} + c$  mit  $\mathbf{b} \in \mathbb{R}^d, c \in \mathbb{R}$


quadratisches erstes Integral :  $I(\mathbf{y}) = \frac{1}{2} \mathbf{y}^T \mathbf{M} \mathbf{y} + \mathbf{b}^T \mathbf{y} + c$  mit  $\mathbf{M} \in \mathbb{R}^{d,d}, \mathbf{b} \in \mathbb{R}^d, c \in \mathbb{R}$

**Definition 4.1.1** (Polynomiale Invarianten).

Ein erstes Integral  $I(\mathbf{y})$  ist *polynomial* vom Grad  $n, n \in \mathbb{N}$ , wenn

$$I(\mathbf{y}) = \sum_{\alpha \in \mathbb{N}_0^d, |\alpha| \leq n} \beta_{\alpha} \mathbf{y}^{\alpha}, \quad \beta_{\alpha} \in \mathbb{R} \quad (\text{Multivariates Polynom}).$$

R. Hiptmair  
rev 35327,  
24. Juni  
2011

 **Multiindexnotation:**  $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d, |\alpha| = \sum_i \alpha_i, \mathbf{y}^{\alpha} := y_1^{\alpha_1} \cdots y_d^{\alpha_d}$

**Theorem 4.1.2** (Erhaltung linearer Invarianten).

Alle Runge-Kutta-Einschrittverfahren ( $\rightarrow$  Def. 2.3.5) erhalten lineare erste Integrale.

*Beweis.* (für den autonomen Fall)

Lineare Invariante  $I(\mathbf{y}) = \mathbf{b}^T \mathbf{y} + c, \mathbf{b} \in \mathbb{R}^d, c \in \mathbb{R} \quad \triangleright \quad \text{grad } I(\mathbf{y}) = \mathbf{b} \quad \forall \mathbf{y} \in D$

$$(1.2.8) \Rightarrow \mathbf{b} \cdot \mathbf{f}(\mathbf{y}) = 0,$$

$$\Rightarrow \mathbf{b} \cdot \mathbf{f}\left(\mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{k}_j\right) = \mathbf{b} \cdot \mathbf{k}_i = 0, \quad i = 1, \dots, s \quad (\text{für Inkremente}),$$

$$\Rightarrow I(\mathbf{y}_1) = \mathbf{b} \cdot \mathbf{y}_1 + c = \mathbf{b} \cdot \left(\mathbf{y}_0 + \sum_{i=1}^s b_i \mathbf{k}_i\right) + c = \mathbf{b} \cdot \mathbf{y}_0 + c = I(\mathbf{y}_0). \quad \square$$

*Beispiel 4.1.3* (Präzession einer Magnetnadel).

$\mathbf{y} : \mathbb{R} \mapsto \mathbb{R}^3$  = Trajektorie der Spitze einer Magnetnadel (im äusseren Feld  $\mathbf{h}$ , fixiert in  $0$ )

$\triangleright$  Bewegungsgleichung

$$\dot{\mathbf{y}} = \mathbf{y} \times \mathbf{h}, \quad \text{Kreuzprodukt} \quad \mathbf{y} \times \mathbf{h} = \begin{pmatrix} y_2 h_3 - y_3 h_2 \\ y_3 h_1 - y_1 h_3 \\ y_1 h_2 - y_2 h_1 \end{pmatrix} \perp \mathbf{y}$$

Quadratische Invarianten:

$$\|\mathbf{y}^{(m)}(t)\| = \text{const}, \quad m \in \mathbb{N}_0$$

Anfangswert:  $\mathbf{y}_0 = \left(\frac{1}{2}\sqrt{2}, 0, 1, \frac{1}{2}\sqrt{2}\right)^T$

MATLAB-CODE : Berechnung des Präzession einer Magnetnadel

```
h = [-1;-1;-1]; tspan = [0 10000]; y0 = [0.5*sqrt(2);0;0.5*sqrt(2)];
fun = @(t,x) cross(x,h);
Jac = @(t,x) [0 h(3) -h(2); -h(3) 0 h(1); h(2) -h(1) 0];
options = odeset('reltol',0.001,'abstol',1e-4,'stats','on');
[t45,y45] = ode45(fun,tspan,y0,options);
options = odeset('reltol',0.001,'abstol',1e-4,'stats','on','Jacobian',Jac);
[t23,y23] = ode23s(fun,tspan,y0,options);
```

ode45: 24537 successful steps, 7432 failed attempts, 191815 function evaluations  
ode23s: 93447 successful steps, 4632 failed attempts, 289607 function evaluations

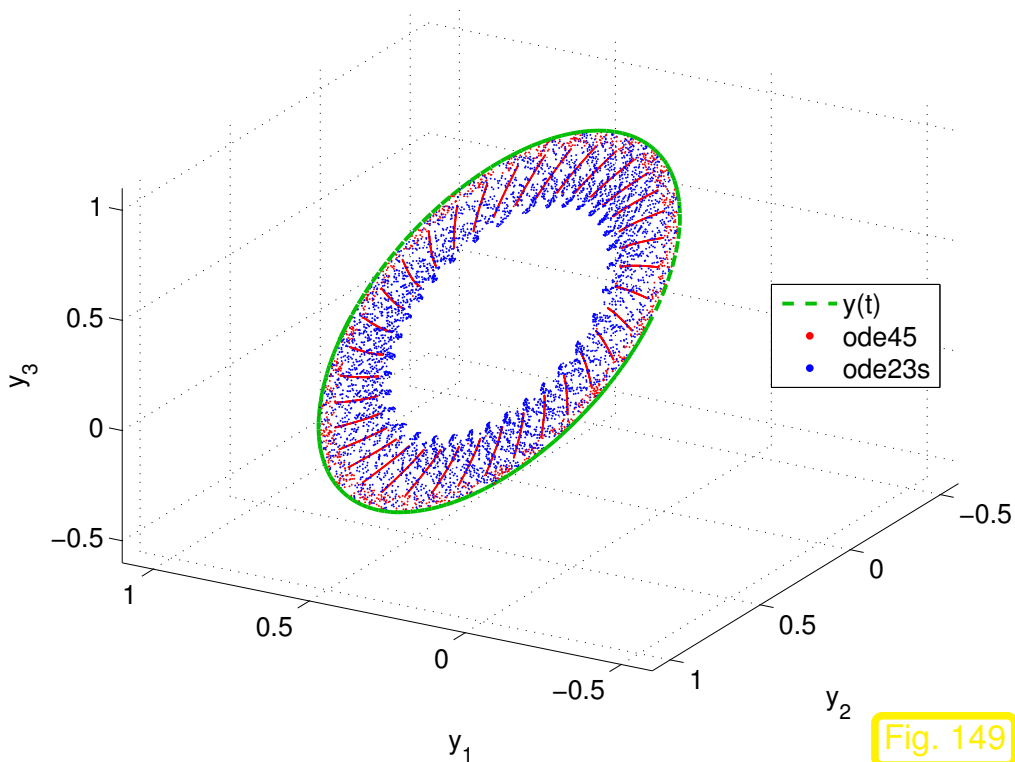


Fig. 149

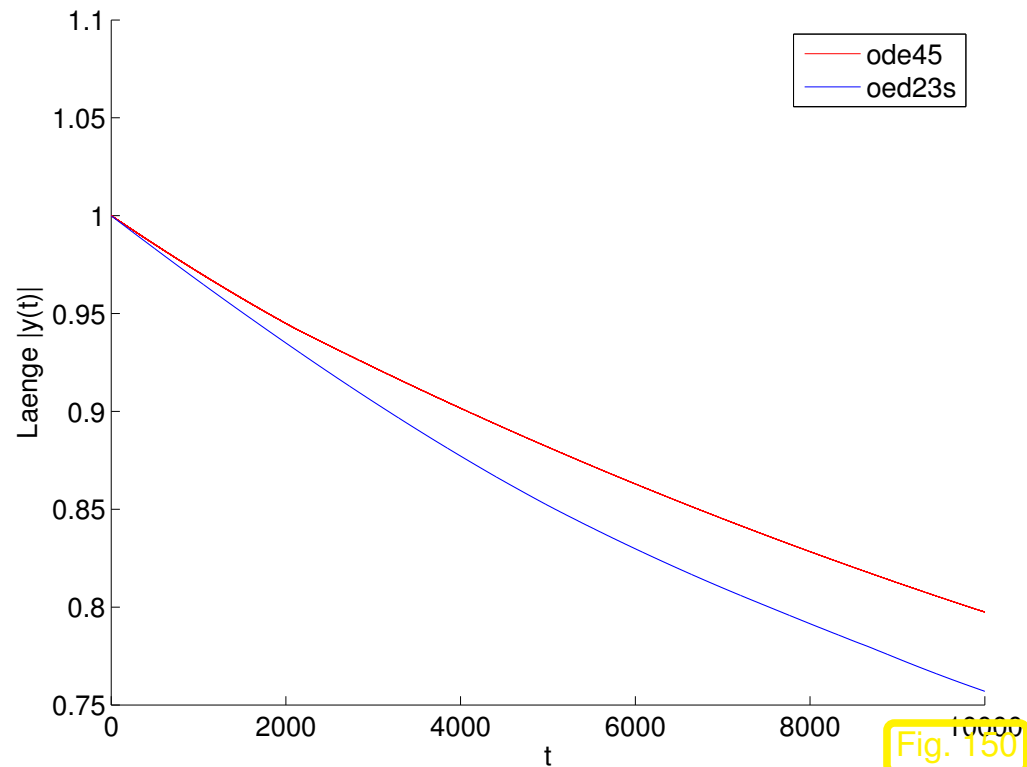
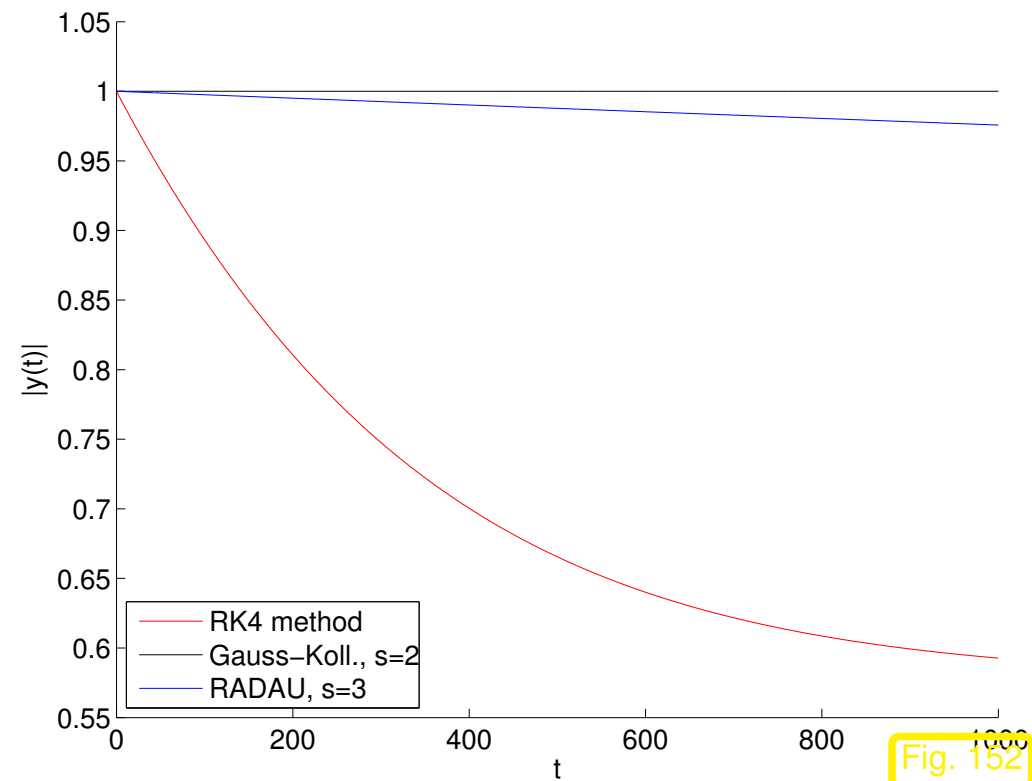
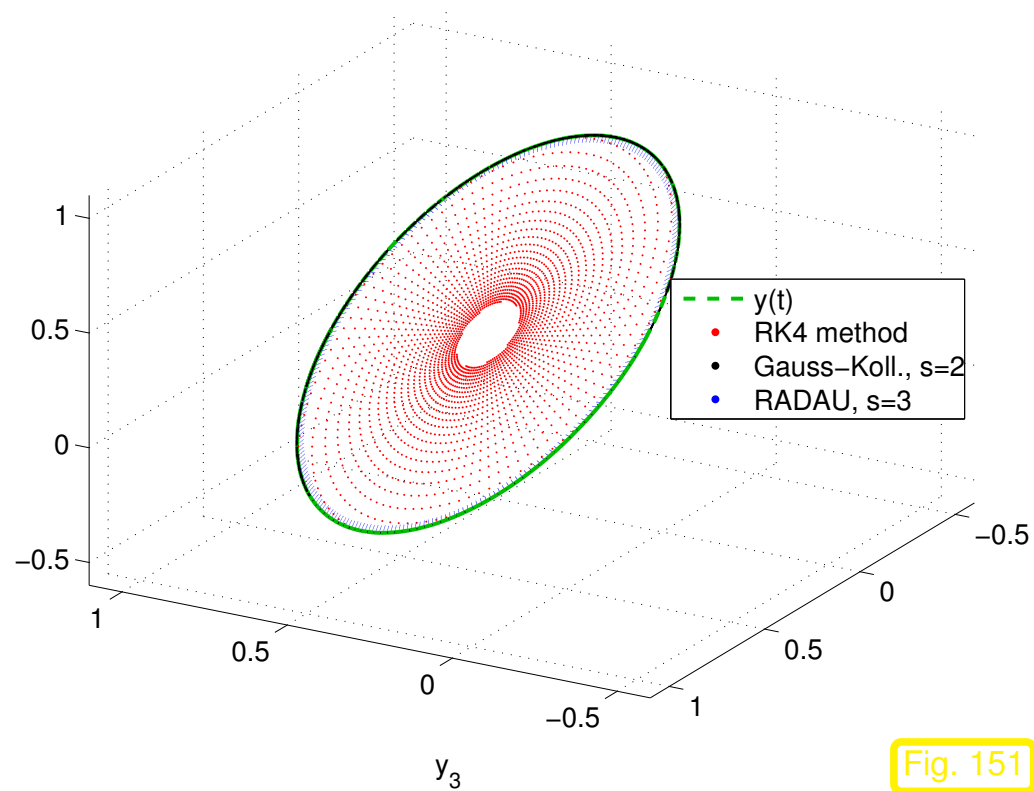
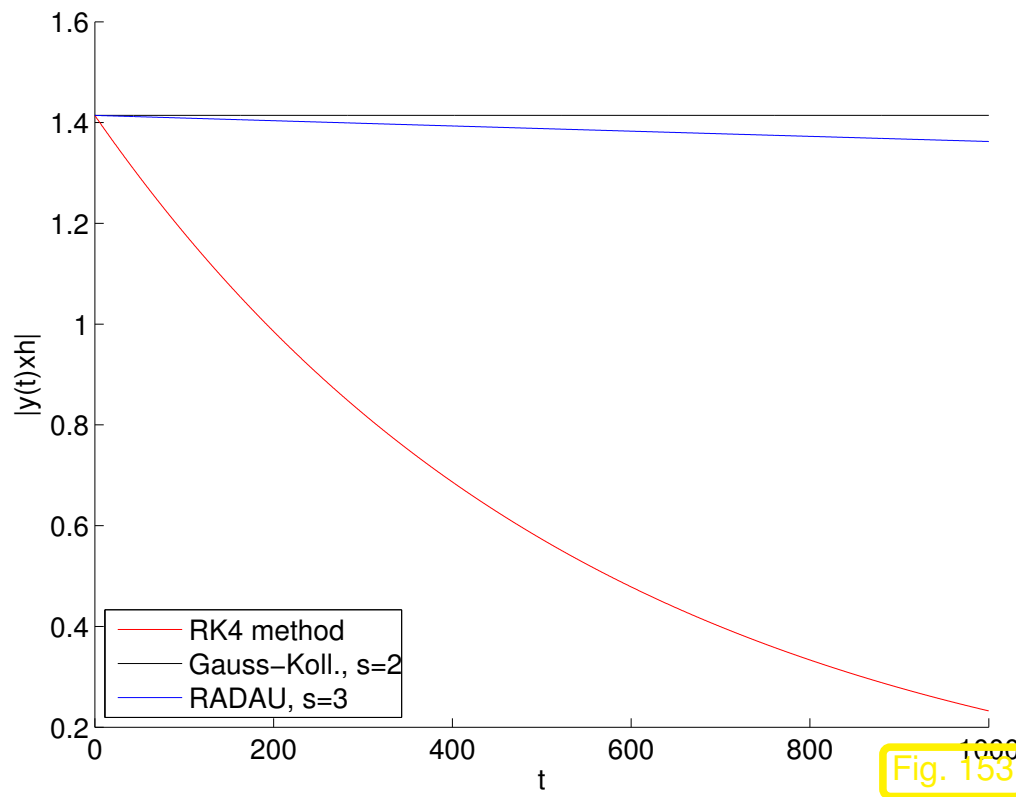


Fig. 150

➤ Keine Erhaltung von  $\|y(t)\|$  über *lange Zeiten* ( $\leftrightarrow$  viele Schwingungsperioden des Pendels)

Einschrittverfahren auf äquidistanten Zeitgittern (qualitatives Verhalten):





△  
 ◁ Erhaltung *aller* quadratischer Invarianten nur durch das Gauss-Kollokationsverfahren.

Erinnerung an Lemma 1.4.23: Erhalt quadratischer erster Integrale durch die implizite Mittelpunktsregel, das einfachste Gauss-Kollokationsverfahren.



**Theorem 4.1.4** (Erhaltung quadratischer Invarianten).

*Gauss-Kollokations-ESV (→ Sect. 2.2.3) erhalten quadratische erste Integrale.*

*Beweis.* (für autonome ODE  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ , vgl. Beweis von Thm. 3.3.7)

$\mathbf{y}_h(t) \in \mathcal{P}_s \hat{=}$  Gauss-Kollokationspolynom zum Anfangswert  $\mathbf{y}_0$  ( $t_0 = 0$ ):

$$\blacktriangleright \quad \dot{\mathbf{y}}_h(c_i h) = \mathbf{f}(\mathbf{y}_h(c_i h)) \quad , \quad h \hat{=} \text{Schrittweite, vgl. (2.2.1)} \quad . \quad (4.1.5)$$

Quadratische Invariante:  $I(\mathbf{y}) = \frac{1}{2} \mathbf{y}^T \mathbf{M} \mathbf{y} + \mathbf{b}^T \mathbf{y} + c$  mit  $\mathbf{M} = \mathbf{M}^T \in \mathbb{R}^{d,d}$ ,  $\mathbf{b} \in \mathbb{R}^d$ ,  $c \in \mathbb{R}$

$$\blacktriangleright \quad d(\tau) := I(\mathbf{y}_h(\tau h)) \quad \text{ist Polynom vom Grad } \leq 2s \quad .$$

Da  $s$ -Punkt Gauss-Quadraturformel exakt für Polynome vom Grad  $\leq 2s - 1$  und  $d' \in \mathcal{P}_{2s-1}$

$$d(1) = d(0) + \int_0^1 d'(\tau) d\tau = d(0) + \underbrace{\sum_{i=1}^s b_i d'(c_i)}_{\text{Ziel } \stackrel{!}{=} 0} \quad .$$

Aus Kollokationsbedingungen (4.1.5) und (1.2.8)

$$d'(\tau) = h \mathbf{grad} I(\mathbf{y}_h(\tau h)) \cdot \dot{\mathbf{y}}_h(\tau h) \quad \Rightarrow \quad d'(c_i) = h \underbrace{\mathbf{grad} I(\mathbf{y}_h(c_i h)) \cdot \mathbf{f}(\mathbf{y}_h(c_i h))}_{=0} = 0 \quad .$$

Da  $d(0) = I(\mathbf{y}_0)$ ,  $d(1) = I(\mathbf{y}_1)$  folgt die Behauptung. □



**Lemma 4.1.6** (Erhaltung quadratischer Invarianten durch RK-ESV).

Erfüllen die Koeffizienten eines  $s$ -stufigen (konsistenten) Runge-Kutta-Einschrittverfahrens ( $\rightarrow$  Def. 2.3.5)

$$b_i a_{ij} + b_j a_{ji} = b_i b_j \quad \text{für alle } i, j = 1, \dots, s, \quad (4.1.7)$$

dann erhält dessen diskrete Evolution quadratische erste Integrale.

*Beweis:* (für vereinfachte quadratische Invariante  $I(\mathbf{y}) := \frac{1}{2} \mathbf{y}^T \mathbf{M} \mathbf{y}$ ,  $\mathbf{M} \in \mathbb{R}^{d,d}$ ,  $\mathbf{M} = \mathbf{M}^T$ )

$$(1.2.8) \quad \triangleright \quad \mathbf{grad} I(\mathbf{y}) = \mathbf{M} \mathbf{y} \quad \Rightarrow \quad \mathbf{y}^T \mathbf{M} \mathbf{f}(\mathbf{y}) = 0 \quad \forall \mathbf{y} \in D. \quad (4.1.8)$$

Ein Schritt des RK-ESV mit Inkrementen  $\mathbf{k}_i$ , vgl. Def. 2.3.5:

$$\mathbf{y}_1 = \mathbf{y}_0 + h \sum_{i=1}^s b_i \mathbf{k}_i,$$

$$\blacktriangleright \quad \mathbf{y}_1^T \mathbf{M} \mathbf{y}_1 - \mathbf{y}_0^T \mathbf{M} \mathbf{y}_0 = 2h \sum_{i=1}^s b_i \mathbf{y}_0^T \mathbf{M} \mathbf{k}_i + h^2 \sum_{i=1}^s \sum_{j=1}^s b_i b_j \mathbf{k}_i^T \mathbf{M} \mathbf{k}_j. \quad (4.1.9)$$

Benutze Stufenform, vgl. Bem. 2.2.5:

$$\mathbf{g}_i = \mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{f}(\mathbf{g}_j) \quad \triangleright \quad \mathbf{k}_i = \mathbf{f}(\mathbf{g}_i), \quad i = 1, \dots, s, \quad \boxed{\mathbf{y}_0 = \mathbf{g}_i - h \sum_{j=1}^s a_{ij} \mathbf{f}(\mathbf{g}_j).}$$

Einsetzen in (4.1.9), da aus (4.1.8) folgt  $\mathbf{g}_i \mathbf{M} \mathbf{f}(\mathbf{g}_i) = 0$ :

$$\begin{aligned} \mathbf{y}_1^T \mathbf{M} \mathbf{y}_1 - \mathbf{y}_0^T \mathbf{M} \mathbf{y}_0 &= 2h \sum_{i=1}^s b_i \left( \mathbf{g}_i - h \sum_{j=1}^s a_{ij} \mathbf{f}(\mathbf{g}_j) \right)^T \mathbf{M} \mathbf{f}(\mathbf{g}_i) + h^2 \sum_{i=1}^s \sum_{j=1}^s b_i b_j \mathbf{f}(\mathbf{g}_i)^T \mathbf{M} \mathbf{f}(\mathbf{g}_j) \\ &= -2h^2 \sum_{i=1}^s b_i \sum_{j=1}^s a_{ij} \mathbf{f}(\mathbf{g}_j)^T \mathbf{M} \mathbf{f}(\mathbf{g}_i) + h^2 \sum_{i=1}^s \sum_{j=1}^s b_i b_j \mathbf{f}(\mathbf{g}_i)^T \mathbf{M} \mathbf{f}(\mathbf{g}_j) \\ &= h^2 \sum_{i=1}^s \sum_{j=1}^s (-2b_i a_{ij} + b_i b_j) \mathbf{f}(\mathbf{g}_i)^T \mathbf{M} \mathbf{f}(\mathbf{g}_j) . \end{aligned}$$

$\mathbf{M} = \mathbf{M}^T \quad \triangleright \quad$  Indexvertauschung in der Doppelsumme  $\quad \triangleright \quad$  Behauptung. □

**Theorem 4.1.10** (Nichterhaltung allgemeiner polynomialer Invarianten).


Für  $n \geq 3$  gibt es kein konsistentes Runge-Kutta-Einschrittverfahren ( $\rightarrow$  Def. 2.3.5) das für alle autonomen Differentialgleichungen  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$  alle ihre polynomialen Invarianten ( $\rightarrow$  Def. 4.1.1) vom Grad  $n$  erhält.


Hilfssatz für den Beweis:

**Lemma 4.1.11** (Ableitung der Determinantenfunktion).

Für die Determinantenfunktion  $\det : \mathbb{R}^{d,d} \mapsto \mathbb{R}$  gilt

$$(D \det(\mathbf{X}))(\mathbf{H}) = \text{trace}(\text{adj}(\mathbf{X})\mathbf{H}), \quad \mathbf{X}, \mathbf{H} \in \mathbb{R}^{d,d}.$$

 Notation: **Spur** einer Matrix  $\mathbf{A} = (a_{ij})_{i,j=1}^d \in \mathbb{R}^{d,d}$ :  $\text{trace}(\mathbf{A}) = \sum_{j=1}^d a_{jj}$

 Notation: **adjungierte Matrix**  $(\text{adj}(\mathbf{X}))_{ij} = (-1)^{i+j} \det(\check{\mathbf{X}}_{ij})$ ,  $\mathbf{X} \in \mathbb{R}^{d,d}$ ,  $1 \leq i, j \leq d$ ,  $\check{\mathbf{X}}_{ij} \hat{=}$  Matrix, die aus  $\mathbf{X}$  durch Streichen der  $i$ . Zeile und  $j$ . Spalte entsteht (Minor).

☞ Bekannt aus der linearen Algebra [10, Lemma 4.3.4]:

$$\mathbf{A} \cdot \text{adj}(\mathbf{A}) = \det(\mathbf{A}) \cdot \mathbf{I}$$

*Beweis.* Als Polynom in den Matrixelementen ist  $\mathbf{A} \mapsto \det \mathbf{A}$  eine  $C^\infty$ -Funktion:

$$\det \mathbf{A} := \sum_{\sigma \in \Pi_d} \text{sgn}(\sigma) \prod_{i=1}^d a_{i, \sigma(i)} .$$

$$\begin{aligned} \blacktriangleright \det(\mathbf{I} + \epsilon \mathbf{H}) &= \sum_{\sigma \in \Pi_d} \text{sgn}(\sigma) \prod_{i=1}^d (\delta_{i, \sigma(i)} + \epsilon h_{i, \sigma(i)}) \\ &= \prod_{i=1}^d (1 + \epsilon h_{ii}) + O(\epsilon^2) = 1 + \epsilon \sum_{i=1}^d h_{ii} + O(\epsilon^2) . , \end{aligned}$$

für  $\mathbf{H} = (h_{ij})_{i,j=1}^d$ , denn jede Permutation  $\neq Id$  erzeugt ein Produkt der Grösse  $O(\epsilon^2)$ . Daher für reguläres  $\mathbf{X} \in \mathbb{R}^{d,d}$ :

$$\det(\mathbf{X} + \epsilon \mathbf{H}) - \det(\mathbf{X}) = \epsilon \underbrace{\text{trace}(\det(\mathbf{X}) \mathbf{X}^{-1} \mathbf{H})}_{\text{adj}(\mathbf{X})} + O(\epsilon^2) .$$

Da die regulären Matrizen in  $\mathbb{R}^{d,d}$  dicht liegen,  $\mathbf{X} \mapsto \text{adj} \mathbf{X}$  stetig  $\rightarrow$

□

*Beweis von Thm. 4.1.10 (Widerspruchsbeweis)*

$t \mapsto \mathbf{Y}(t)$  löse lineare Matrix-Differentialgleichung

$$\dot{\mathbf{Y}} = \mathbf{A}\mathbf{Y}, \quad \mathbf{A} \in \mathbb{R}^d$$

▶ Mit Lemma 4.1.11 folgt für  $I(\mathbf{Y}) = \det \mathbf{Y}$

$$D_{\mathbf{Y}}I(\mathbf{Y})\mathbf{H} = \det \mathbf{Y} \cdot \text{trace}(\mathbf{Y}^{-1}\mathbf{H}) \Rightarrow \frac{d}{dt} \det \mathbf{Y}(t) = \mathbf{Y} \text{trace}(\dot{\mathbf{Y}}\mathbf{Y}^{-1}) = \mathbf{Y} \text{trace}(\mathbf{A}) .$$

(4.1.12)

⇒

Falls  $\text{trace}(\mathbf{A}) = 0$  ist  $I(\mathbf{Y}) := \det \mathbf{Y}$  eine polynomiale Invariante vom Grad  $d$  der Matrix-Differentialgleichung  $\dot{\mathbf{Y}} = \mathbf{A}\mathbf{Y}$ .

**Annahme:** RK-ESV erhält Polynomiale Invarianten vom Grad  $d > 2$ . Wende das Verfahren an auf  $\dot{\mathbf{Y}} = \mathbf{A}\mathbf{Y}$ ,  $\text{trace}(\mathbf{A}) = 0$ ,  $\mathbf{A} \in \mathbb{R}^{d,d}$ . Nach Bem. 3.1.13, (3.1.16)

$$\mathbf{Y}_1 = S(h\mathbf{A})\mathbf{Y}_0 \quad \text{mit Stabilitätsfunktion } S(z), \quad h > 0 .$$

▶  $\det \mathbf{Y}_1 = \det \mathbf{Y}_0 \quad \forall \mathbf{Y}_0 \Rightarrow \det S(h\mathbf{A}) = 1$

Wähle spezielle (diagonale !) Matrix mit  $\text{trace}(\mathbf{A}) = 0$  und Zeitschrittweite  $h = 1$

$$\mathbf{A} = \text{diag}(\mu, \nu, -(\mu + \nu), 0, \dots, 0) \in \mathbb{R}^{d,d}, \quad \mu, \nu \in \mathbb{R} .$$

$$\blacktriangleright \quad S(\mathbf{A}) = \text{diag}(S(\mu), S(\nu), S(-(\mu + \nu)), 0, \dots, 0)$$

Aus  $\det S(\mathbf{A}) = 1$  folgt, dass  $S$  die Funktionalgleichung  $S(\mu)S(\nu)S(-(\mu + \nu)) = 1$  erfüllt.

$$\Rightarrow S(0) = 1 \quad \Rightarrow S(-\mu) = S(\mu)^{-1} \quad \Rightarrow S(\mu)S(\nu) = S(\mu + \nu) \quad \forall \mu, \nu \in \mathbb{R} .$$

$z \mapsto S(z)$  erfüllt die Funktionalgleichung der Exponentialfunktion, ist stetig in Umgebung von 0  $\Rightarrow S(z) = \exp(z)$ .

Andererseits muss  $S(z)$  eine rationale Funktion sein, siehe Thm. 3.1.6, ein **Widerspruch**  $\square$

## 4.2 Volumenerhaltung

Physik: inkompressible Strömung  $\leftrightarrow$  volumenerhaltender Fluss

**Definition 4.2.1** (Volumenerhaltung).

Eine Abbildung  $\Phi : D \subset \mathbb{R}^d \mapsto \mathbb{R}^d$  heisst volumenerhaltend

$$\forall V \subset D \text{ messbar: } \text{Vol}(\Phi(V)) = \text{Vol}(V) .$$

**Lemma 4.2.2** (Volumenerhaltende Abbildungen).

Eine stetig differenzierbare Abbildung  $\Phi : D \subset \mathbb{R}^d \mapsto \mathbb{R}^d$  ist genau dann volumenerhaltend, wenn  $|\det D\Phi(\mathbf{y})| = 1$  für alle  $\mathbf{y} \in D$ .


*Beweis.* Nach dem Transformationssatz für Integrale:

$$\text{Vol}(\Phi(V)) = \int_{\Phi(V)} 1 \, d\mathbf{x} = \int_V |\det D\Phi(\mathbf{y})| \, d\mathbf{y} . \quad \square$$

**Theorem 4.2.3** (Satz von Liouville).

Sei  $\mathbf{f} : D \subset \mathbb{R}^d \mapsto \mathbb{R}^d$  stetig differenzierbar. Genau dann wenn  $\text{div } \mathbf{f}(\mathbf{y}) = 0$  für jedes  $\mathbf{y} \in D$ , ist die zu  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$  gehörige Evolution  $\Phi^t$  **volumenerhaltend**, d.h.

$$\forall V \subset D \text{ kompakt: } \exists \delta > 0: \quad \text{Vol}(\Phi^t(V)) = \text{Vol}(V) \quad \forall 0 \leq t < \delta .$$

 Notation: **Divergenz**  $\text{div } \mathbf{f}(\mathbf{y}) = \sum_{j=1}^d \frac{\partial f_j}{\partial y_j}(\mathbf{y}) = \text{trace } D\mathbf{f}(\mathbf{y})$ , mit  $\mathbf{f} = (f_1, \dots, f_d)^T$

*Beweis.* (basierend auf Lemma 4.2.2, vgl. [16, Lemma 9.1])

Sei  $\Phi : \tilde{\Omega} \mapsto D$  der Evolutionsoperator zur autonomen ODE  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ .

Jacobi-Matrix (Propagationsmatrix)  $\mathbf{W}(t, \mathbf{y}) = D_{\mathbf{y}}\Phi^t(\mathbf{y})$ ,  $\mathbf{y} \in D$ , erfüllt die **Variationsgleichung** (1.3.34)

$$\dot{\mathbf{W}}(t, \mathbf{y}) := \frac{d}{dt}\mathbf{W}(t, \mathbf{y}) = D\mathbf{f}(\Phi^t\mathbf{y})\mathbf{W}(t, \mathbf{y}), \quad t \in J(\mathbf{y}), \quad \mathbf{W}(0, \mathbf{y}) = \mathbf{I}. \quad (4.2.4)$$

Wie im Beweis zu Thm. 4.1.10, aus Lemma 4.1.11, vgl. (4.1.12):

$$\begin{aligned} \frac{d}{dt} \det \mathbf{W}(t, \mathbf{y}) &= \det \mathbf{W}(t, \mathbf{y}) \operatorname{trace}(\dot{\mathbf{W}}(t, \mathbf{y})\mathbf{W}^{-1}(t, \mathbf{y})) \\ &\stackrel{(\text{??})}{=} \det \mathbf{W}(t, \mathbf{y}) \operatorname{trace}(D\mathbf{f}(\Phi^t\mathbf{y})) \\ &= \det \mathbf{W}(t, \mathbf{y}) \operatorname{div} \mathbf{f}(\Phi^t\mathbf{y}). \end{aligned} \quad (4.2.5)$$

“ $\Rightarrow$ ” aus (4.2.5), da  $\det \mathbf{W}(0, \mathbf{y}) = 1$

“ $\Leftarrow$ ”: Wenn  $\operatorname{div} \mathbf{f} \neq 0$ , dann gibt es  $\delta > 0$ ,  $V \subset D$  so dass  $|\operatorname{div} \mathbf{f}(\mathbf{y})| > \delta$  für alle  $\mathbf{y} \in V$ . Daher, für  $\mathbf{y} \in V$ ,

$$\frac{d}{dt} \det \mathbf{W}(t, \mathbf{y}) \geq \delta \det \mathbf{W}(t, \mathbf{y}) \quad \text{oder} \quad \frac{d}{dt} \det \mathbf{W}(t, \mathbf{y}) \leq -\delta \det \mathbf{W}(t, \mathbf{y}).$$



Inkompressible Strömung  $\leftrightarrow$  **divergenzfreie** Geschwindigkeitsfelder

Beispiel 4.2.6 (Strömungsvisualisierung).

Anwendung numerischer ODE-Löser in der *Computergraphik*:

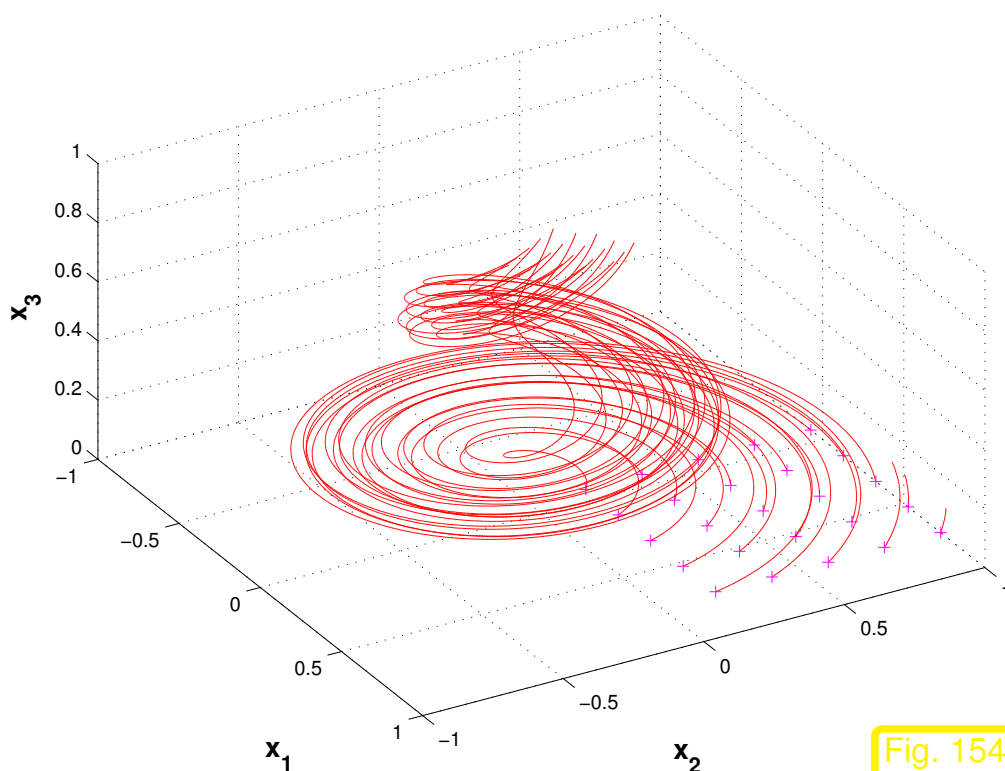


Fig. 154

Divergenzfreies Vektorfeld:

$$\mathbf{f}(\mathbf{y}) = \begin{pmatrix} -y_2 - \frac{y_1}{a^2 + y_3^2} \\ y_1 - \frac{y_2}{a^2 + y_3^2} \\ 2/a \arctan(y_3/a) \end{pmatrix}, \quad \mathbf{y} \in \mathbb{R}^3.$$

MATLAB-funktion

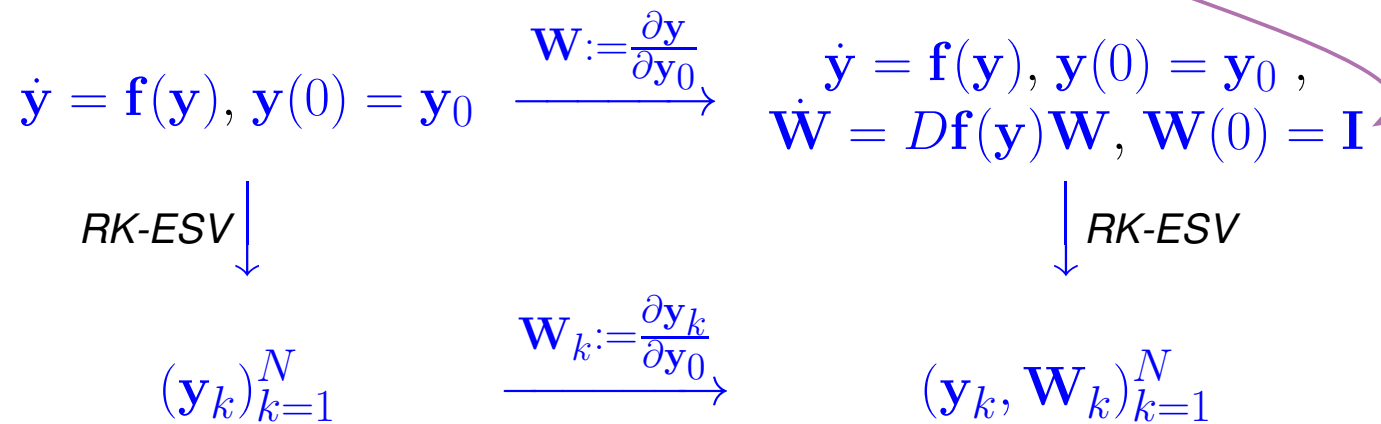
`streamline(X, Y, Z, U, V, W, ...)`

**Stromlinien** von  $\mathbf{f} \hat{=}$  Lösungen von AWPe zur autonomen ODE  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ .

**Lemma 4.2.7** (Variationsgleichung und Runge-Kutta-Einschrittverfahren).

Für Runge-Kutta-Einschrittverfahren ( $\rightarrow$  Def. 2.3.5) kommutiert das folgende Diagramm

Variationsgleichung, siehe Sect. 1.3.3.4.



*Beweis:* (nur für explizites Euler-Verfahren (1.4.2), vgl. [16, Lemma 4.1])

Rekursion des expliziten Euler-Verfahrens für  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$

$$\mathbf{y}_{k+1} = \mathbf{y}_k + h\mathbf{f}(\mathbf{y}_k) \xrightarrow{\frac{d}{d\mathbf{y}_0}} \frac{d\mathbf{y}_{k+1}}{d\mathbf{y}_0} = \frac{d\mathbf{y}_k}{d\mathbf{y}_0} + hD\mathbf{f}(\mathbf{y}_k) \frac{d\mathbf{y}_k}{d\mathbf{y}_0}.$$

Explizites Euler-Verfahren für (erweiterte) Variationsgleichung  $\dot{\mathbf{W}} = D\mathbf{f}(\mathbf{y})\mathbf{W}$ :

$$\mathbf{W}_{k+1} = \mathbf{W}_k + hD\mathbf{f}(\mathbf{y}_k)\mathbf{W}_k .$$

►  $\frac{d\mathbf{y}_k}{d\mathbf{y}_0}$  und  $\mathbf{W}_k$  erfüllen die gleiche Rekursion. □

Der Beweis im allgemeinen Fall stützt sich auf implizites Differenzieren der Runge-Kutta-Inkrementgleichungen, siehe Def. 2.3.5.

*Bemerkung 4.2.8* (Volumenerhaltende Integratoren für  $d = 2$ ).

Für RK-ESV im Fall  $d = 2$ :

Erhalt quadratischer Invarianten  $\implies$  Volumenerhaltung

Für  $d = 2$ :  $\det \mathbf{W} = w_{11}w_{22} - w_{12}w_{21}$  ist quadratische Funktion ( $\mathbf{W} = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix} \hat{=}$  Propagationsmatrix/Wronsk-Matrix, siehe (1.3.33)). Ist die Evolution zu  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$  volumenerhaltend,

so gilt  $\det \mathbf{W}(t) \equiv 1$ , also ist  $\det \mathbf{W}$  eine quadratische Invariante der Variationsgleichung und wird vom RK-ESV erhalten.

Nach Lemma 4.2.7 gilt dann

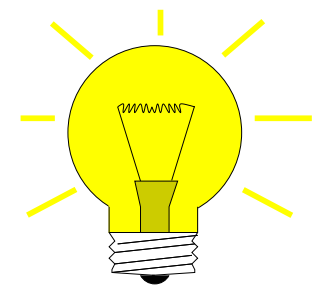
$$\forall h > 0: \det \left( D_{\mathbf{y}_0} \Psi^h \right) = 1 .$$

Nach Lemma 4.2.2 ist die diskrete Evolution daher volumenerhaltend.



$d > 2$ : (Beweis von) Thm. 4.1.10 ➤ Allgemeine Runge-Kutta-Einschrittverfahren kommen nicht in Betracht !

(Notwendig: Integratoren mit “eingebauter Zusatzinformation” über  $\mathbf{f}$ )



Idee: ❶ Additive Zerlegung von  $\mathbf{f}$  in *wesentlich zweidimensionale* Vektorfelder

❷ **Splittingverfahren** (→ Sect. 2.5) basierend auf RK-ESV, die quadratische Invarianten erhalten, siehe Sect. 4.1.

Zu ❶: 
$$\mathbf{f}(\mathbf{y}) = \sum_{i=1}^{d-1} \mathbf{g}_{i,i+1}(\mathbf{y}), \quad \mathbf{g}_{i,i+1}(\mathbf{y}) = (0 \ \cdots \ 0 \ \underset{i}{*} \ \underset{i+1}{*} \ 0 \ \cdots \ 0)^T, \quad \operatorname{div} \mathbf{g}_{i,i+1} = 0.$$

Zu ❷: 
$$\mathbf{y}_{k+1} = (\Psi_{d-1}^h \circ \cdots \circ \Psi_1^h) \mathbf{y}_k,$$

wobei  $\Psi_i^h \hat{=}$  diskrete Evolution des RK-Basisverfahrens für  $\dot{\mathbf{y}} = \mathbf{g}_{i,i+1}(\mathbf{y})$ .

Verallgemeinertes Lie-Trotter-Splitting (2.5.2)

Existenz der Vektorfelder  $\mathbf{g}_{i,i+1}$  ?

**Theorem 4.2.9** (Zerlegung in divergenzfreie Vektorfelder).

Jedes stetige divergenzfreie  $\mathbf{f} : \mathbb{R}^d \mapsto \mathbb{R}^d$  lässt sich darstellen als Summe von  $d - 1$  divergenzfreien Vektorfeldern  $\mathbf{g}_{i,i+1} : \mathbb{R}^d \mapsto \mathbb{R}^d$  der Form

$$\mathbf{g}_{i,i+1}(\mathbf{y}) = (0 \ \cdots \ 0 \ \underset{i}{p_i(\mathbf{y})} \ \underset{i+1}{q_i(\mathbf{y})} \ 0 \ \cdots \ 0)^T, \quad i = 1, \dots, d - 1,$$

mit Funktionen  $p_i, q_i : \mathbb{R}^d \mapsto \mathbb{R}$ .

$$\mathbf{f}(\mathbf{y}) = \begin{pmatrix} f_1(\mathbf{y}) \\ f_2(\mathbf{y}) \\ f_3(\mathbf{y}) \\ f_4(\mathbf{y}) \\ \vdots \\ f_{d-1}(\mathbf{y}) \\ f_d(\mathbf{y}) \end{pmatrix} = \begin{pmatrix} p_1(\mathbf{y}) \\ q_1(\mathbf{y}) \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ p_2(\mathbf{y}) \\ q_2(\mathbf{y}) \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ p_3(\mathbf{y}) \\ q_3(\mathbf{y}) \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ p_{d-2}(\mathbf{y}) \\ q_{d-2}(\mathbf{y}) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ p_{d-1}(\mathbf{y}) \\ q_{d-1}(\mathbf{y}) \end{pmatrix} .$$

*Beweis.* (vgl. [16, Theorem 9.3]) Konstruktiv für  $\mathbf{f} = (f_1, \dots, f_d)^T$  mit beliebigen  $a_i \in \mathbb{R}$

$$p_i(\mathbf{y}) = f_i(\mathbf{y}) + r_i(\mathbf{y}) \quad , \quad q_i(\mathbf{y}) = -r_{i+1}(\mathbf{y}) \quad ,$$

$$r_i(\mathbf{y}) = \int_{a_i}^{y_i} \left( \frac{\partial f_1}{\partial y_1} + \cdots + \frac{\partial f_{i-1}}{\partial y_{i-1}} \right) (y_1, \dots, y_{i-1}, \tau, y_{i+1}, \dots, y_d) \, d\tau \quad , \quad 2 \leq i \leq d-1 \quad ,$$

$$r_1(\mathbf{y}) \equiv 0 \quad ,$$

$$\Rightarrow \quad \frac{\partial p_i}{\partial y_i} = \frac{\partial f_i}{\partial y_i} + \frac{\partial f_1}{\partial y_1} + \cdots + \frac{\partial f_{i-1}}{\partial y_{i-1}} \quad , \quad 1 \leq i \leq d-1 \quad .$$

$$\frac{\partial q_i}{\partial y_{i+1}} = -\frac{\partial f_1}{\partial y_1} - \cdots - \frac{\partial f_i}{\partial y_i} = -\frac{\partial p_i}{\partial y_i} \quad .$$

Also sind die Teilfelder alle divergenzfrei. Weiter, wegen  $\operatorname{div} \mathbf{f} = 0$ ,

$$\left( \sum_{i=1}^{d-1} \mathbf{g}_{i,i+1}(\mathbf{y}) \right)_d = q_{d-1}(\mathbf{y}) = -r_d(\mathbf{y}) = - \int_{a_d}^{y_d} \frac{\partial f_1}{\partial y_1} + \dots + \frac{\partial f_{d-1}}{\partial y_{d-1}} = \int_{a_d}^{y_d} \frac{\partial f_d}{\partial y_d} = f_d(\mathbf{y}) .$$

Beachte: Konstruktion der  $\mathbf{g}_{i,i+1}$  benötigt (symbolische) Ableitungen der  $f_i$ .

*Bemerkung 4.2.10* (Volumenerhaltendes Splittingverfahren 2. Ordnung).

Einfachstes volumenerhaltendes Splittingverfahren 2. Ordnung:

- Basis-RK-ESV: implizite Mittelpunktsregel (1.4.19),  $\Psi_i^h \hat{=}$  diskrete Evolutionen zu  $\dot{\mathbf{y}} = \mathbf{g}_{i,i+1}(\mathbf{y})$   
 ➔  $\Psi_i^h$  volumenerhaltend, siehe Bem. 4.2.8 !
- Verallgemeinertes Strang-Splitting (2.5.3)

$$\Psi^h := \Psi_1^{h/2} \circ \Psi_2^{h/2} \circ \dots \circ \Psi_{d-2}^{h/2} \circ \Psi_{d-1}^h \circ \Psi_{d-2}^{h/2} \circ \dots \circ \Psi_1^{h/2} .$$

► symmetrisches Einschrittverfahren → Def. 2.1.27  $\xRightarrow{\text{Thm. 2.1.29}}$  Konsistenzordnung  $\geq 2$

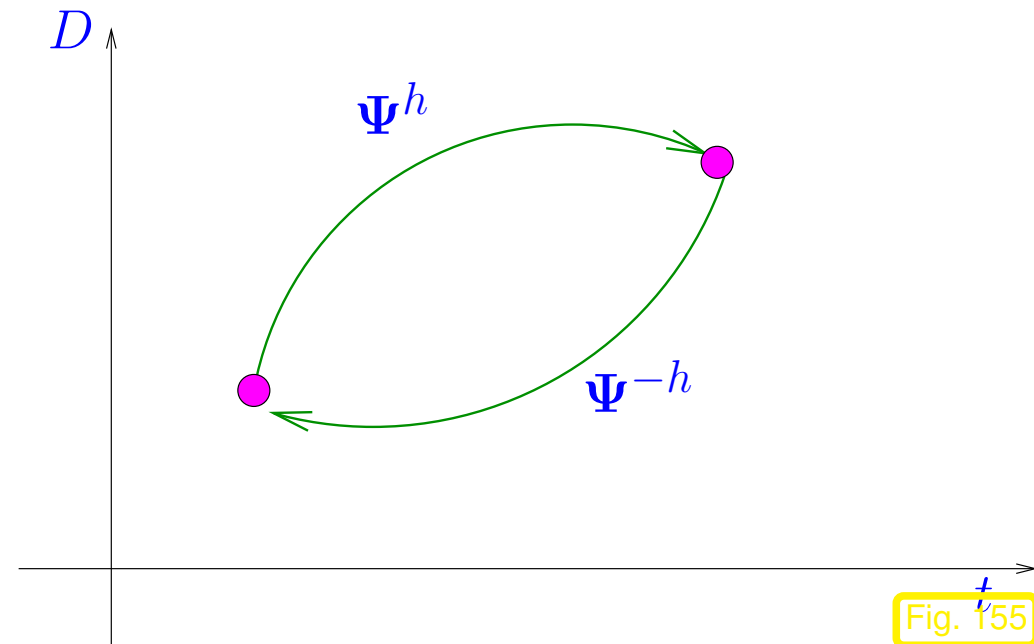


# 4.3 Verallgemeinerte Reversibilität

Sect. 2.1.5: **reversible** (symmetrische) diskrete Evolutionen/Einschrittverfahren → Def. 2.1.27

(für autonome ODE)  $\Psi^{-h} \circ \Psi^h = Id$  für  $h > 0$  hinreichend klein

Reversibilität  
↕  
Symmetrie bzgl. Zeitumkehr



R. Hiptmair  
rev 35327,  
24. Juni  
2011

Erinnerung an Thm. 2.1.29: Reversible ESV haben *gerade* Konsistenzordnung

Wir haben bereits reversible Einschrittverfahren kennengelernt: implizite Mittelpunktsregel (1.4.19) aus Abschnitt 1.4.3, das einfachste Gauss-Kollokationsverfahren, vgl. (2.2.19).



**Theorem 4.3.1** (Reversible Runge-Kutta-Einschrittverfahren).

Ein  $s$ -stufiges RK-ESV ( $\rightarrow$  Def. 2.3.5) mit Butcher-Tableau  $\frac{\mathbf{c}}{\mathbf{b}^T} \left| \begin{array}{c} \mathfrak{A} \\ \mathbf{b}^T \end{array} \right.$ , siehe (2.3.6), ist reversibel (symmetrisch,  $\rightarrow$  Def. 2.1.27), falls

$$a_{s+1-i, s+1-j} + a_{ij} = b_j \quad \forall 1 \leq i, j \leq s .$$

*Beweis* (siehe [16, Sect. V.2, Thm. 2.3])

zu zeigen:  $\mathbf{y}_0 \xrightarrow{\Psi^h} \mathbf{y}_1 \xrightarrow{\Psi^{-h}} \mathbf{y}_0 .$

Technik: Teste Invarianz der Verfahrensgleichungen bei Vertauschung  $\mathbf{y}_0 \leftrightarrow \mathbf{y}_1, h \leftrightarrow -h$  in den Verfahrensgleichungen

$$\left\{ \begin{array}{l} \mathbf{k}_i = \mathbf{f}\left(\mathbf{y}_0 + h \sum_{j=1}^s a_{ij} \mathbf{k}_j\right), \\ \mathbf{y}_1 = \mathbf{y}_0 + h \sum_{i=1}^s b_i \mathbf{k}_i . \end{array} \right. \Rightarrow \left\{ \begin{array}{l} \mathbf{k}_i = \mathbf{f}\left(\mathbf{y}_1 - h \sum_{j=1}^s a_{ij} \mathbf{k}_j\right), \\ \mathbf{y}_0 = \mathbf{y}_1 - h \sum_{i=1}^s b_i \mathbf{k}_i . \end{array} \right. \quad (4.3.2)$$

$$\Rightarrow \begin{cases} \mathbf{k}_i = \mathbf{f}\left(\mathbf{y}_0 + h \sum_{j=1}^s (b_j - a_{ij}) \mathbf{k}_j\right), \\ \mathbf{y}_1 = \mathbf{y}_0 + h \sum_{i=1}^s b_i \mathbf{k}_i. \end{cases} \quad (4.3.3)$$

$2a_{ij} = b_j \Rightarrow$  Gleichheit nach Vertauschung  $\mathbf{y}_0 \leftrightarrow \mathbf{y}_1, h \leftrightarrow -h$ , doch leider liefert das kein sinnvolles RK-ESV (Ausnahme: implizite Mittelpunktsregel (1.4.19) mit  $s = 1, a_{11} = \frac{1}{2}, b_1 = 1$ )

**!** Beachte:  $a_{s+1-i, s+1-j} + a_{ij} = b_j \Rightarrow b_{s+1-i} = b_i$

➤ Umindizieren  $i \leftarrow s + 1 - i, j \leftarrow s + 1 - j$  unter den Annahme  $b_{s+1-i} = b_i$

$$(4.3.3) \Rightarrow \begin{cases} \mathbf{k}_i = \mathbf{f}\left(\mathbf{y}_0 + h \sum_{j=1}^s (b_j - a_{s+1-i, s+1-j}) \mathbf{k}_j\right), \\ \mathbf{y}_1 = \mathbf{y}_0 + h \sum_{i=1}^s b_i \mathbf{k}_i. \end{cases}$$

= Ausgangs-RK-ESV, falls  $b_j - a_{s+1-i, s+1-j} = a_{ij}$  □

**Theorem 4.3.4** (Reversible Gauss-Kollokations-ESV).

*Gauss-Kollokations-ESV sind reversibel ( $\rightarrow$  Def. 2.1.27).*

Da Gauss-Kollokationsverfahren zur Klasse der Runge-Kutta-Einschrittverfahren gehören, genügt es, die Voraussetzungen von Thm. 4.3.1 zu verifizieren. Dazu verwende die expliziten Formeln (2.2.3) für die Runge-Kutte-Koeffizienten  $a_{ij}$  und  $b_i$ ,  $1 \leq i, j \leq s$ .

$$a_{ij} = \int_0^{c_i} L_j(\tau) d\tau \quad , \quad b_i = \int_0^1 L_i(\tau) d\tau \quad ,$$

wobei die  $c_i$  die auf  $[0, 1]$  normalisierten Kollokationspunkte (= Gausspunkte) sind, und die  $L_j$  die dazugehörigen Lagrange-Polynome, siehe (2.2.2).

Lage der Gaussknoten für die  $s$ -Punkt Gaussquadraturformel auf  $[0, 1]$  ist symmetrisch um  $\frac{1}{2}$ , siehe Fig. 59:

$$c_i = c_{s+1-i} \quad \Rightarrow \quad L_i(\tau) = L_{s+1-i}(1 - \tau) \quad , \quad 1 \leq i \leq s \quad . \quad (4.3.5)$$

$$\begin{aligned} \blacktriangleright \quad a_{s+1-i, s+1-j} + a_{ij} &= \int_0^{c_{s+1-i}} L_{s+1-j}(\tau) d\tau + \int_0^{c_i} L_j(\tau) d\tau \\ &= - \int_1^{1-c_{s+1-i}} L_{s+1-j}(1 - \tau) d\tau + \int_0^{c_i} L_j(\tau) d\tau \end{aligned}$$

Nachtrag zu Sect. 3.3:

**Theorem 4.3.6** (Stabilitätsgebiet und Reversibilität).

Für reversible und A-stabile ( $\rightarrow$  Def. 3.2.16) Runge-Kutta-Einschrittverfahren gilt  $\mathcal{S}_{\Psi} = \mathbb{C}^-$ .

Neues Konzept: **R-Reversibilität** = “verallgemeinerte Zeitumkehrsymmetrie”

*Beispiel* 4.3.7 (Reversibilität bei mechanischen Systemen).

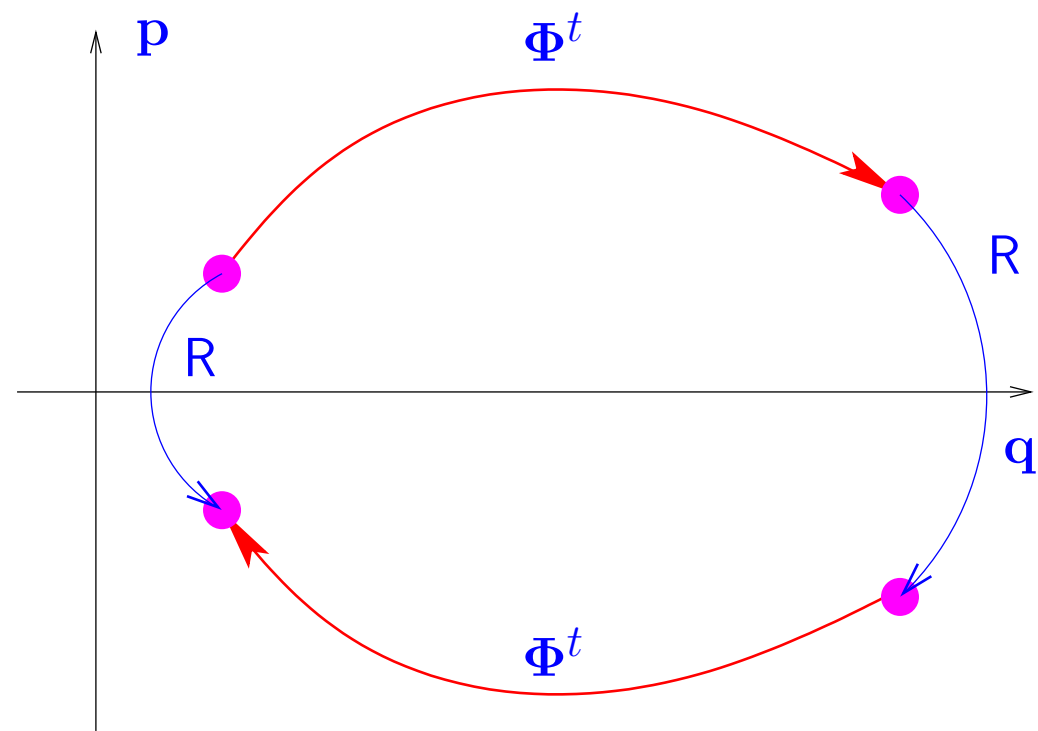
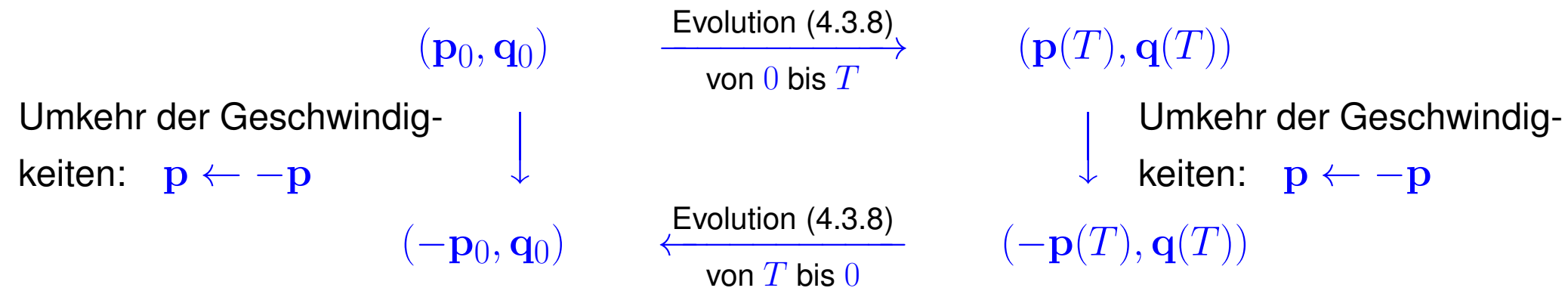
Hamiltonsche Differentialgleichung ( $\rightarrow$  Def. 1.2.20) mit Hamilton-Funktion

$$H : \begin{cases} \mathbb{R}^n \times \mathbb{R}^n & \mapsto \mathbb{R} \\ (\mathbf{p}, \mathbf{q}) & \mapsto \frac{1}{2} \mathbf{p}^T \mathbf{M}^{-1} \mathbf{p} + U(\mathbf{q}) \end{cases}, \quad \Rightarrow \quad H(\mathbf{p}, \mathbf{q}) = H(-\mathbf{p}, \mathbf{q}),$$

mit s.p.d. Massenmatrix  $\mathbf{M} \in \mathbb{R}^{d,d}$ .

$$\dot{\mathbf{p}}(t) = -\frac{\partial H}{\partial \mathbf{q}}(\mathbf{p}(t), \mathbf{q}(t)) = -\mathbf{grad} U(\mathbf{q}), \quad \dot{\mathbf{q}}(t) = \frac{\partial H}{\partial \mathbf{p}}(\mathbf{p}(t), \mathbf{q}(t)) = \mathbf{M}^{-1} \mathbf{p}. \quad (4.3.8)$$

Kommutierendes Diagramm (Zeitumkehrsymmetrie)



$\Leftrightarrow$  Evolution  $\Phi^t$  zu (1.2.21) erfüllt

$$R \circ \Phi^t = \Phi^{-t} \circ R \tag{4.3.9}$$

mit Abbildung

$$R \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} -\mathbf{p} \\ \mathbf{q} \end{pmatrix} . \tag{4.3.10}$$

(4.3.9)  $\hat{=}$  „Rückwärtsevolution“ nach Umkehr der Geschwindigkeiten



Abstraktion:

• Betrachte autonome AWP  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}), \mathbf{y}(0) = \mathbf{y}_0,$

$\mathbf{f} : D \mapsto \mathbb{R}^d$  lokal Lipschitz-stetig ( $\rightarrow$  Def. 1.3.2)

- **Annahme:** Für alle  $\mathbf{y}_0 \in D$  existiert die Lösung für alle Zeiten, vgl. Def. 1.3.1

**Definition 4.3.11** (R-reversible Abbildung).

Es sei  $R : D \mapsto D \subset \mathbb{R}^d$  eine bijektive lineare Abbildung.

Eine weitere bijektive Abbildung  $\Phi : D \mapsto D$  heisst **R-reversibel**, falls

$$R \circ \Phi = \Phi^{-1} \circ R .$$

**Lemma 4.3.12** (R-reversible Evolutionen).

Die Evolution  $\Phi^t$  zu  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$  ist R-reversibel für alle  $t \in \mathbb{R}$ , falls

$$\mathbf{f} \circ R = -R \circ \mathbf{f} \quad \text{auf } D . \tag{4.3.13}$$

*Beweis.* (siehe [16, Sect. V.1]) Zu zeigen ist, wegen  $\Phi^t \circ \Phi^{-t} = Id$  (Gruppeneigenschaft (1.3.8))

$$R \circ \Phi^t = (\Phi^t)^{-1} \circ R = \Phi^{-t} \circ R . \tag{4.3.14}$$

Idee: beide Seiten von (4.3.14) sind Lösungen des gleichen Anfangswertproblems  $\mathbf{y} \in D$

$$\frac{d}{dt}((\mathbf{R} \circ \Phi^t)(\mathbf{y})) = \mathbf{R}f(\Phi^t(\mathbf{y})) = -\mathbf{f}((\mathbf{R} \circ \Phi^t)(\mathbf{y})) , \quad (4.3.15)$$

$$\frac{d}{dt}((\Phi^{-t} \circ \mathbf{R})(\mathbf{y})) = -\mathbf{f}((\Phi^{-t} \circ \mathbf{R})(\mathbf{y})) . \quad (4.3.16)$$

$t \mapsto (\mathbf{R} \circ \Phi^t)(\mathbf{y})$  und  $t \mapsto (\Phi^{-t} \circ \mathbf{R})(\mathbf{y})$  sind beides Lösungen des Anfangswertproblems

$$\dot{\mathbf{z}} = -\mathbf{f}(\mathbf{z}) \quad , \quad \mathbf{z}(0) = \mathbf{R}\mathbf{y} .$$

Daher folgt (4.3.14) aus dem Eindeigkeitssatz Thm. 1.3.4. □

*Beispiel* 4.3.17 (Fortsetzung: Reversibilität bei mechanischen Systemen). Bsp. 4.3.7

Für Hamiltonsche Evolution (4.3.8) mit  $\mathbf{y} = (\mathbf{p}, \mathbf{q})^T$ ,  $d = 2n$ ,  $\mathbf{R}$  aus (4.3.10)

$$(\mathbf{f} \circ \mathbf{R})(\mathbf{y}) = \begin{pmatrix} -\mathbf{grad} U(\mathbf{R}\mathbf{q}(\mathbf{y})) \\ \mathbf{M}^{-1}\mathbf{R}\mathbf{p}(\mathbf{y}) \end{pmatrix} = \begin{pmatrix} -\mathbf{grad} U(\mathbf{q}) \\ -\mathbf{M}^{-1}\mathbf{p} \end{pmatrix} = -\mathbf{R} \begin{pmatrix} -\mathbf{grad} U(\mathbf{q}) \\ \mathbf{M}^{-1}\mathbf{p} \end{pmatrix} = -\mathbf{R}(\mathbf{f}(\mathbf{y})) .$$

$\hat{=}$  Voraussetzung von Lemma 4.3.12.

Alternative Perspektive: Hamiltonsche Dgl. (1.2.24)  $\dot{\mathbf{y}} = \mathbf{J}^{-1} \text{grad } H(\mathbf{y})$ ,  $\mathbf{J} = \begin{pmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{pmatrix}$ :

$$H(\mathbf{R}\mathbf{y}) = H(\mathbf{y}) \quad \Rightarrow \quad \mathbf{R} \text{grad } H(\mathbf{R}\mathbf{y}) = \text{grad } H(\mathbf{y}) . \quad (4.3.18)$$

Für  $\mathbf{R}$  aus (4.3.9):  $\mathbf{J} \circ \mathbf{R} = -\mathbf{R} \circ \mathbf{J}$ ,  $\mathbf{R}^2 = Id$

$$\begin{aligned} (4.3.18) \quad & \Rightarrow -\mathbf{R}(\mathbf{J}^{-1} \text{grad } H(\mathbf{y})) = \mathbf{J}^{-1} \mathbf{R}(\text{grad } H(\mathbf{y})) = \mathbf{J}^{-1} \mathbf{R} \mathbf{R} \text{grad } H(\mathbf{R}\mathbf{y}) = \mathbf{J}^{-1} \text{grad } H(\mathbf{R}\mathbf{y}) \\ \hat{=} \quad & (4.3.13) \text{ für } \mathbf{f}(\mathbf{y}) = \mathbf{J}^{-1} \text{grad } H(\mathbf{y}). \end{aligned}$$

**Theorem 4.3.19** (R-reversible Runge-Kutta-Evolutionen).

*Die rechte Seite  $\mathbf{f}$  der autonomen ODE  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$  erfülle (4.3.13).*

*Dann ist die von einem Runge-Kutta-Einschrittverfahren ( $\rightarrow$  Def. 2.3.5) erzeugte diskrete Evolution genau dann R-reversibel, wenn das RK-ESV reversibel/symmetrisch ( $\rightarrow$  Def. 2.1.27) ist.*



*Beweis.* (siehe [16, Sect. V.1, Thm. 1.5])

① Mit Notationen von Lemma 4.3.12 und  $\Psi^h$  als diskrete Evolution des RK-ESV zur ODE  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$  wird gezeigt (vgl. Beweis der Affin-Kovarianz von RK-ESV, Bem. 2.3.13)

$$\mathbf{f} \circ \mathbf{R} = -\mathbf{R} \circ \mathbf{f} \quad \Rightarrow \quad \mathbf{R} \circ \Psi^h = \Psi^{-h} \circ \mathbf{R} . \quad (4.3.20)$$

Gemäss Def. 2.3.5, wegen *Linearität* von  $\mathbf{R}$

$$\left\{ \begin{array}{l} \mathbf{k}_i = \mathbf{f}\left(\mathbf{y} + h \sum_{j=1}^s a_{ij} \mathbf{k}_j\right) , \\ \Psi^h \mathbf{y} = \mathbf{y} + h \sum_{i=1}^s b_i \mathbf{k}_i , \end{array} \right. \quad (4.3.13) \quad \Rightarrow \quad \left\{ \begin{array}{l} \mathbf{R}\mathbf{k}_i = -\mathbf{f}\left(\mathbf{R}\mathbf{y} + h \sum_{j=1}^s a_{ij} \mathbf{R}\mathbf{k}_j\right) , \\ \mathbf{R}\Psi^h \mathbf{y} = \mathbf{R}\mathbf{y} + h \sum_{i=1}^s b_i \mathbf{R}\mathbf{k}_i . \end{array} \right.$$

► Transformierte Inkremente  $\tilde{\mathbf{k}}_i := -\mathbf{R}\mathbf{k}_i$  erfüllen

$$\tilde{\mathbf{k}}_i = \mathbf{f}\left(\mathbf{R}\mathbf{y} - h \sum_{j=1}^s a_{ij} \tilde{\mathbf{k}}_j\right) , \quad i = 1, \dots, s .$$

►  $\tilde{\mathbf{k}}_i \hat{=}$  Inkremente des RK-ESV zur Schrittweite  $-h$ , Anfangswert  $\mathbf{R}\mathbf{y} \leftrightarrow \Psi^{-h} \mathbf{R}\mathbf{y}$

$$\mathbf{R}\Psi^h \mathbf{y} = \mathbf{R}\mathbf{y} - h \sum_{i=1}^s b_i \tilde{\mathbf{k}}_i = \Psi^{-h} \mathbf{R}\mathbf{y} \quad \Rightarrow \quad (4.3.20) .$$

② direkte Verifikation von Def. 4.3.11

RK-ESV reversibel/symmetrisch

$$\Psi^{-h} = (\Psi^h)^{-1}$$

(4.3.20)  
 $\Rightarrow$ 

$$R \circ \Psi^h = \Psi^{-h} \circ R = (\Psi^h)^{-1} \circ R. \quad \square$$

## 4.4 Symplektizität

R. Hiptmair  
rev 35327,  
25. April  
2011

### 4.4.1 Symplektische Evolutionen Hamiltonscher Differentialgleichungen

Erinnerung (Sect. 1.2.4): **Hamiltonsche Differentialgleichung**  $\rightarrow$  Def. 1.2.20

$$\dot{\mathbf{p}}(t) = -\frac{\partial H}{\partial \mathbf{q}}(\mathbf{p}(t), \mathbf{q}(t)) \quad , \quad \dot{\mathbf{q}}(t) = \frac{\partial H}{\partial \mathbf{p}}(\mathbf{p}(t), \mathbf{q}(t)) \quad , \quad (1.2.21)$$

mit (glatter) Hamilton-Funktion  $H : \mathbb{R}^n \times M \mapsto \mathbb{R}$ , Konfigurationsraum  $M \subset \mathbb{R}^n$ .

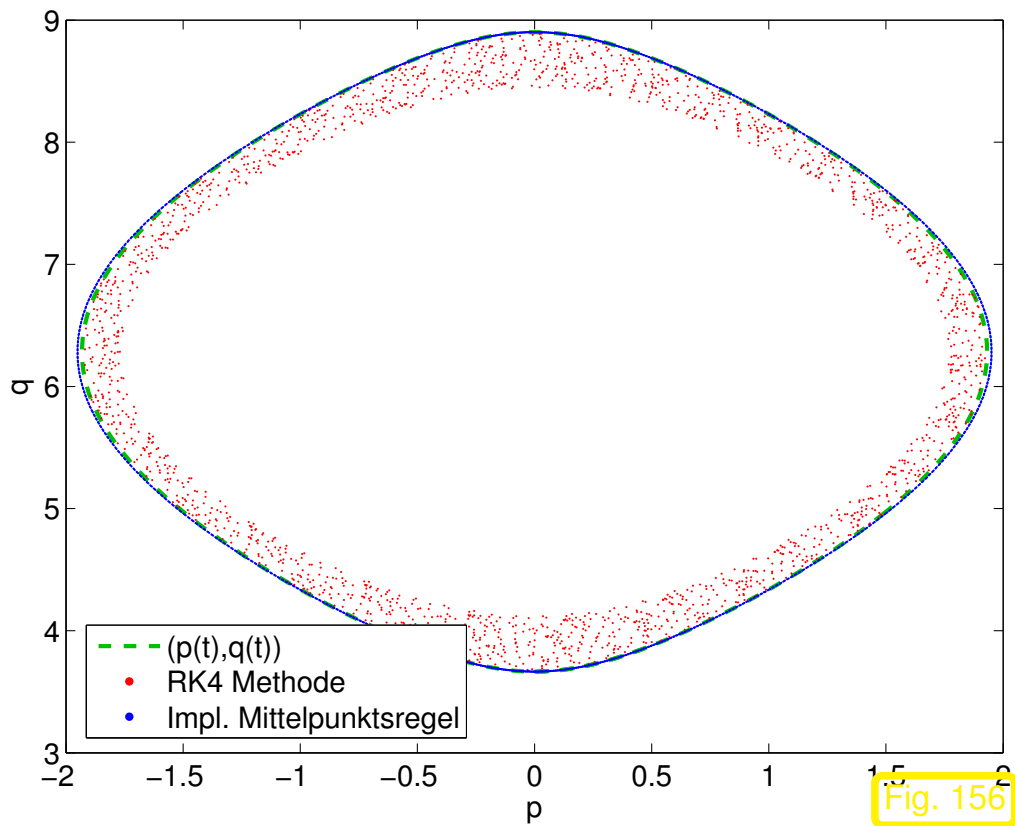
$$\mathbf{y} = \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} \quad (1.2.21) \Leftrightarrow \boxed{\dot{\mathbf{y}} = \mathbf{J}^{-1} \cdot \text{grad } H(\mathbf{y})}, \quad \mathbf{J} = \begin{pmatrix} 0 & \mathbf{I}_n \\ -\mathbf{I}_n & 0 \end{pmatrix} \in \mathbb{R}^{2n,2n}. \quad (1.2.24)$$

Lemma 1.2.23 (Energieerhaltung):  $H$  ist Invariante von (1.2.21)

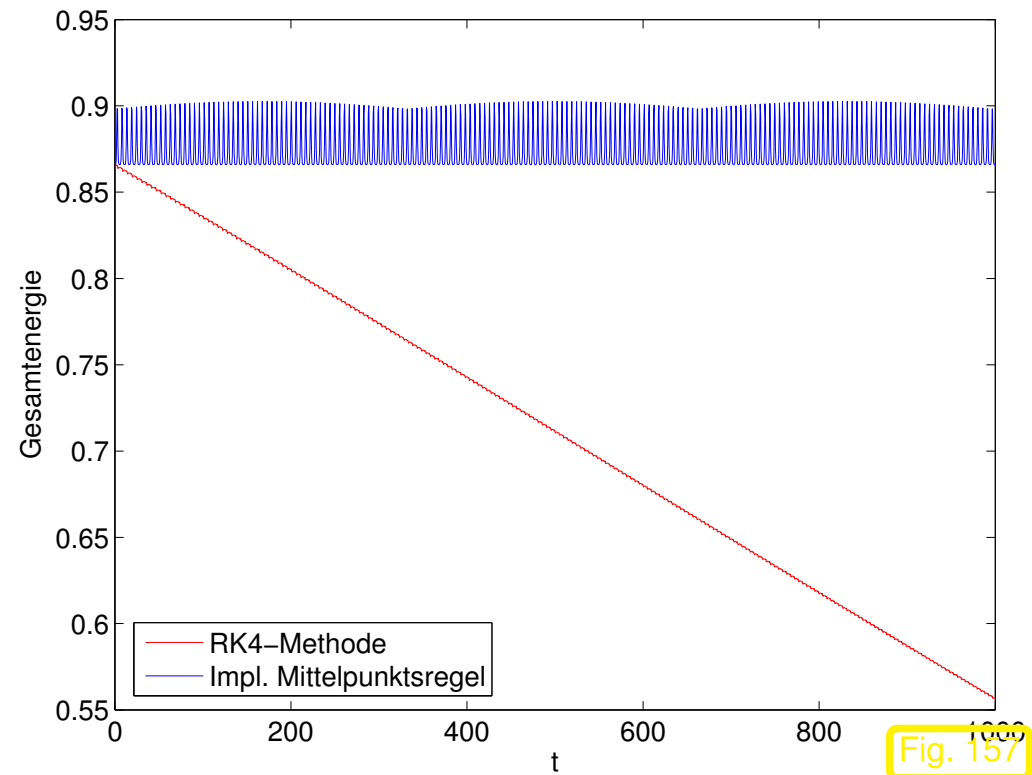
*Beispiel* 4.4.1 (Energieerhaltung bei numerischer Integration).  $\leftrightarrow$  Bsp. 1.4.24

Mathematisches Pendel Bsp. 1.2.17, AWP für (1.2.19) auf  $[0, 1000]$ ,  $p(0) = 0$ ,  $q(0) = 7\pi/6$ .

Vergleich von klassischem Runge-Kutta-Verfahren (2.3.11) (Ordnung 4) mit 1-stufigem Gauss-Kollokations-ESV (implizite Mittelpunktsregel 2.2.19), äquidistantes Gitter,  $h = \frac{1}{2}$ :



Trajektorien „exakter“ /diskreter Evolutionen



Energieerhaltung diskreter Evolutionen

➤ Keine Energiedrift bei impliziter Mittelpunktsregel



Eine rätselhafte Beobachtung:

Besonderheit mancher (\*) numerischer Integratoren:

Approximative **Langzeit-Energieerhaltung** (keine Energiedrift)

- (\*) Implizite Mittelpunktsregel (1.4.19)  $\rightarrow$  Bsp. 4.4.1, 1.4.24,  
Störmer-Verlet-Verfahren (2.5.13)  $\rightarrow$  Bsp. 1.4.32

*Bemerkung 4.4.2* (Volumenerhaltung bei zweidimensionalen Hamiltonschen ODEs).

Für Evolution  $\Phi^t : \mathbb{R}^n \times M \mapsto \mathbb{R}^n \times M$  zu einer Hamiltonschen Differentialgleichung gilt:

$$n = 1 \quad \triangleright \quad \operatorname{div}_{\mathbf{y}} \underbrace{\mathbf{J}^{-1} \operatorname{grad} H(\mathbf{y})}_{\operatorname{rot} H(\mathbf{y})} = 0 \quad \xrightarrow{\text{Thm. 4.2.3}} \quad \Phi^t \text{ volumenerhaltend (flächenerhaltend).}$$

R. Hiptmair  
rev 35327,  
25. April  
2011



*Beispiel 4.4.3* (Flächenerhaltung bei Evolution für Pendelgleichung).  $\rightarrow$  Bsp. 1.2.17

$p \leftrightarrow$  Winkelgeschwindigkeit,  $q \leftrightarrow$  Winkelvariable  $\alpha$

$$\begin{aligned} \dot{p} &= -\sin q, \\ \dot{q} &= p \end{aligned} \quad \blacktriangleright \quad \text{Hamilton-Funktion } H(p, q) = \frac{1}{2}p^2 - \cos q \quad (\text{Gesamtenergie}) \quad (4.4.4)$$

Volumenerhaltung im Zustandsraum (Phasenraum):

Evolution eines quadratischen Volumens  $\triangleleft$

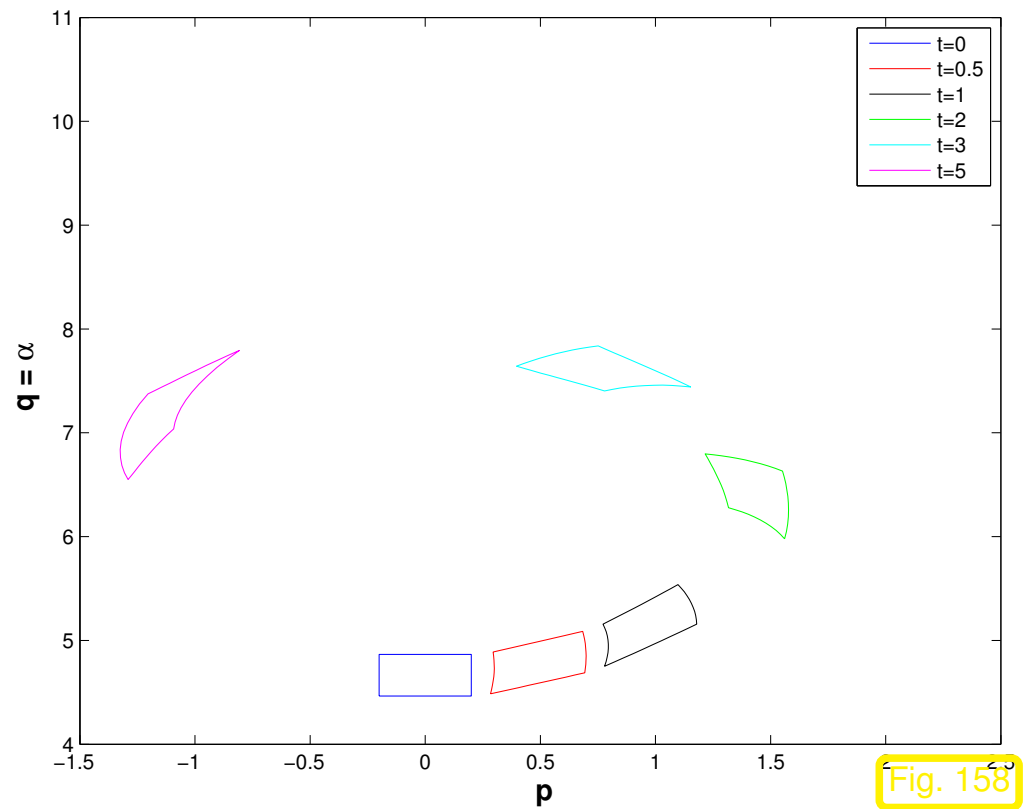
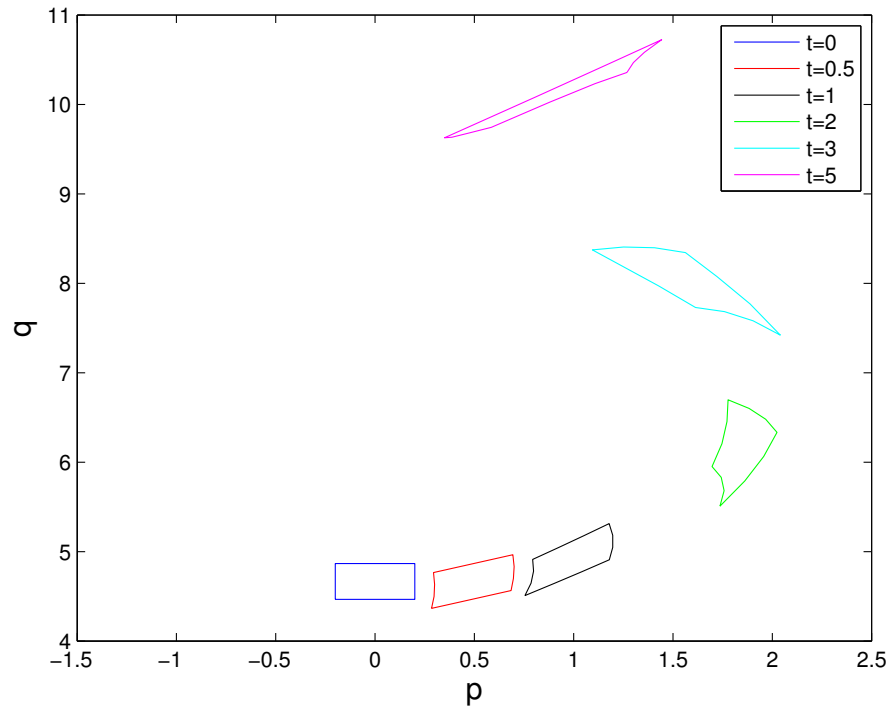
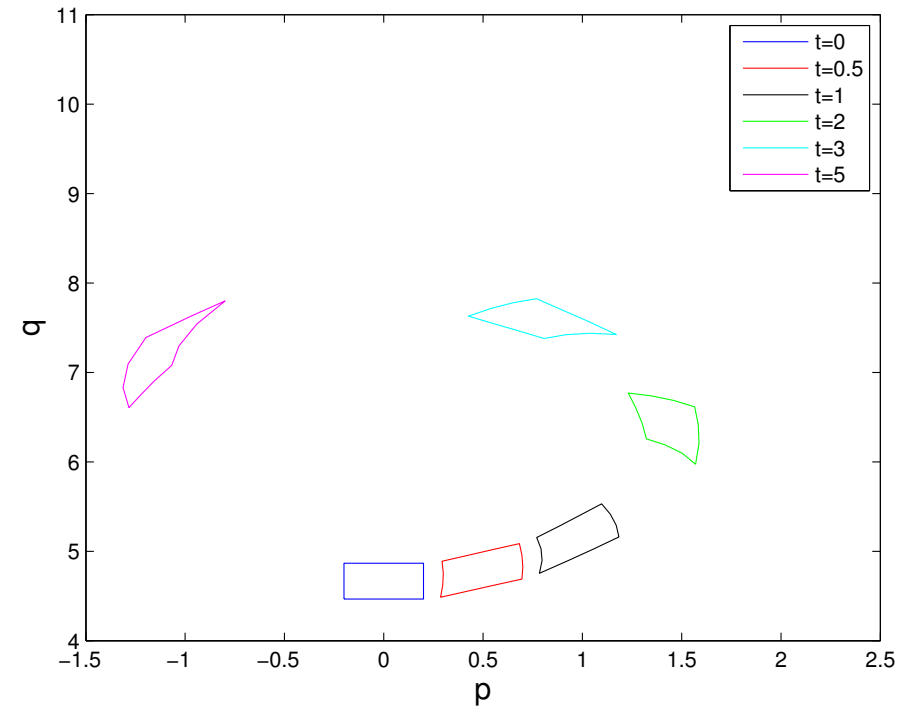


Fig. 158

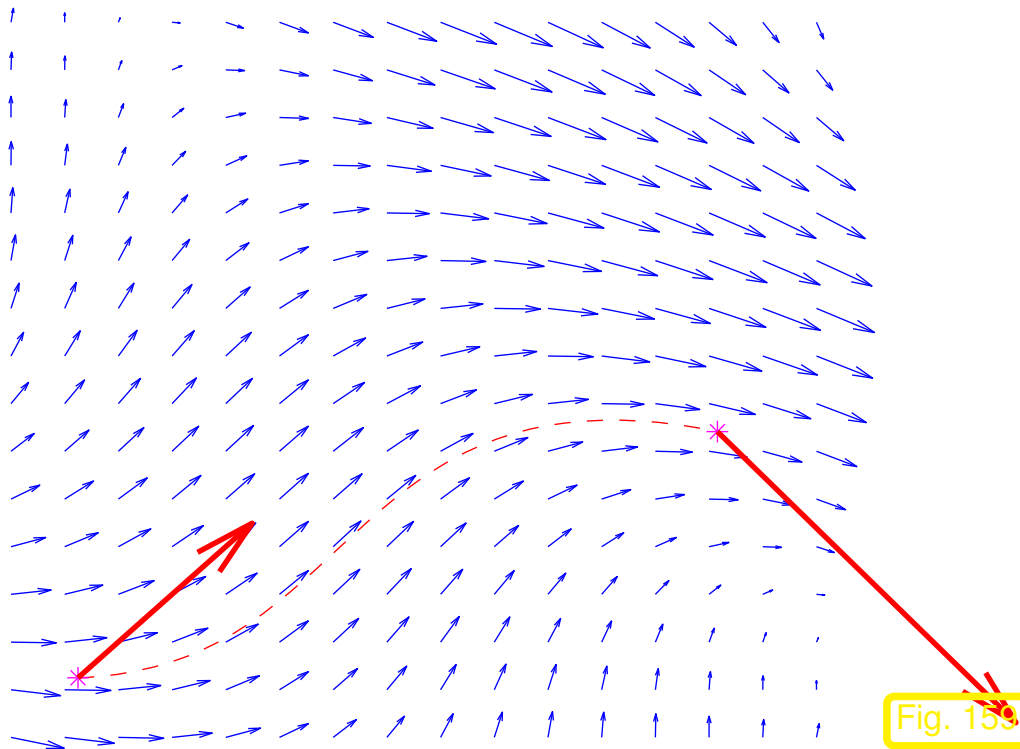


Evolution eines quadratischen Volumens  
(Explizites Eulerverfahren)



Evolution eines quadratischen Volumens  
(Implizite Mittelpunktsregel)

Bem. 4.2.8: Für  $d = 2$  ist die implizite Mittelpunktsregel volumenerhaltend (wie alle Gauss-Kollokationsverfahren nach Lemma 4.1.6)



**Push-Forward:** Wirkung einer (glatten) Abbildung auf infinitesimale Strecke = Vektor

Für  $C^1$ -Abbildung  $\Phi : D \subset \mathbb{R}^d \mapsto \mathbb{R}^d$ :

$$(\Phi_* \mathbf{v})(\mathbf{y}) = D\Phi(\mathbf{y})\mathbf{v} \quad \mathbf{y} \in D, \mathbf{v} \in \mathbb{R}^d .$$

◁ Transport eines Vektors im „Strömungsfeld“  
 $t \mapsto \Phi^t$  zu  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$

**Definition 4.4.5** (Symplektisches Produkt).

$$\omega(\mathbf{v}, \mathbf{w}) := \mathbf{v}^T \mathbf{J} \mathbf{w} , \quad \mathbf{v}, \mathbf{w} \in \mathbb{R}^{2n} \quad \text{mit} \quad \mathbf{J} = \begin{pmatrix} 0 & \mathbf{I}_n \\ -\mathbf{I}_n & 0 \end{pmatrix} .$$



Bemerkung 4.4.6 (Konstante 2-Formen).

Symplektisches Produkt  $\hat{=}$  Prototyp einer **nichtdegenerierten, alternierenden Bilinearform**:

*Definition 4.4.7* (Alternierende, nichtdegenerierte Bilinearform).

Eine Bilinearform  $\beta : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$  heisst

- **alternierend**  $:\Leftrightarrow \beta(\mathbf{x}, \mathbf{y}) = -\beta(\mathbf{y}, \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d,$
- **nichtdegeneriert**  $:\Leftrightarrow \beta(\mathbf{x}, \mathbf{y}) = 0 \quad \forall \mathbf{y} \in \mathbb{R}^d \Rightarrow \mathbf{x} = 0$

$\beta : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$  alternierende Bilinearform  $\Rightarrow \exists \mathbf{L} \in \mathbb{R}^{d,d}:$

$$\mathbf{L}^T = -\mathbf{L}$$

$$\beta(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{L} \mathbf{y} \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d$$

*Lemma 4.4.8* (Normalform schiefsymmetrischer Matrizen).

Zu jedem regulären  $\mathbf{L} \in \mathbb{R}^{2n,2n}$  mit  $\mathbf{L}^T = -\mathbf{L}$  gibt es ein reguläres  $\mathbf{U} \in \mathbb{R}^{d,d}$ , so dass

$$\mathbf{U}^T \mathbf{L} \mathbf{U} = \mathbf{J} = \begin{pmatrix} 0 & \mathbf{I}_n \\ -\mathbf{I}_n & 0 \end{pmatrix} \quad (\text{Kongruenztransformation}).$$

*Beweis.*  $\mathbf{L} = -\mathbf{L}^T \Rightarrow$  unitär diagonalisierbar (**normale Matrix** !), rein imaginäre Eigenwerte, die in konjugiert komplexen Paaren zu konjugiert komplexen Eigenvektoren auftreten:

$$\exists \mathbf{Q} \in \mathbb{C}^{2n}: \quad \mathbf{Q}^{-1} = \mathbf{Q}^H \quad \text{und} \quad \mathbf{Q}^H \mathbf{L} \mathbf{Q} = i \begin{pmatrix} \mathbf{D} & 0 \\ 0 & -\mathbf{D} \end{pmatrix},$$

mit  $\mathbf{D} = \text{diag}(\mu_1, \dots, \mu_n) \in \mathbb{R}^n$ ,  $\mu_i > 0$ . Dann setze

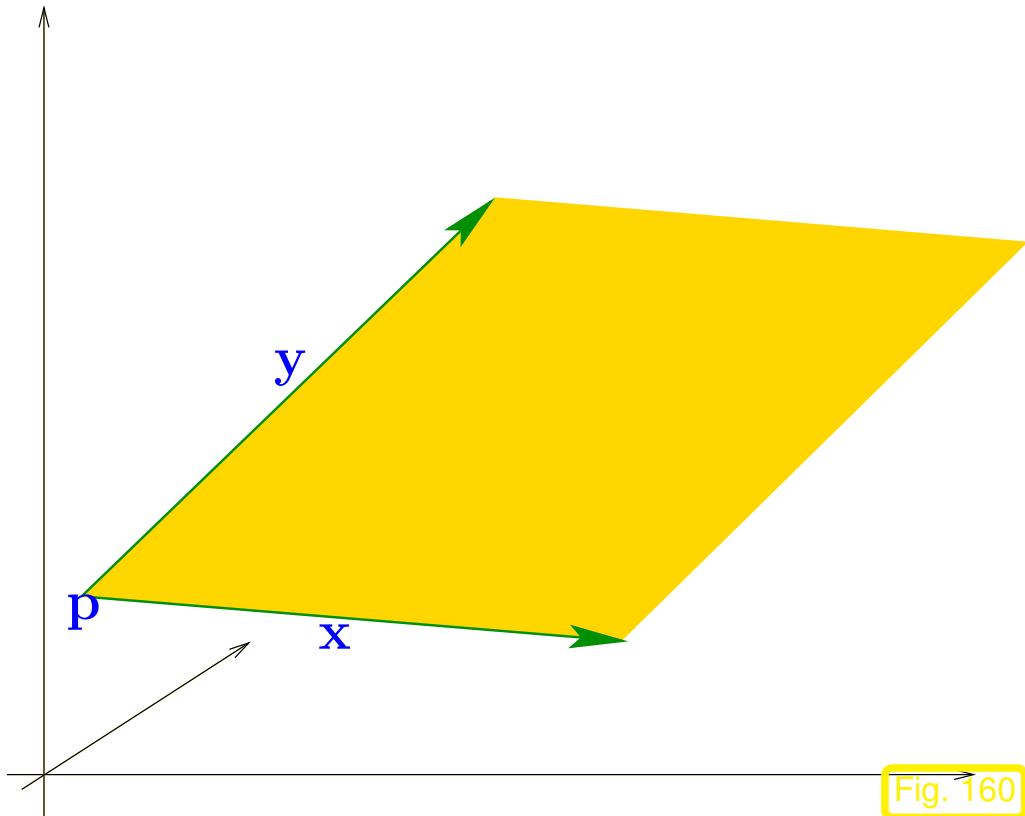
$$\mathbf{U} = \frac{1}{\sqrt{2}} \mathbf{Q} \begin{pmatrix} \mathbf{D}^{-1/2} & \mathbf{D}^{-1/2} \\ -i\mathbf{D}^{-1/2} & i\mathbf{D}^{-1/2} \end{pmatrix} .$$

□

Beachte: Die Matrix  $\mathbf{U}$  ist reell !

► Es gibt eine reelle Koordinatentransformation, die  $\beta$  in  $\omega$  ( $\rightarrow$  Def. 4.4.5) überführt.

*Bemerkung* 4.4.9 (Symplektisches Flussintegral).



◁  $\omega(\mathbf{x}, \mathbf{y}) \hat{=}$  Fluss “durch” orientiertes Parallelogramm, aufgespannt von  $\{\mathbf{p}, \mathbf{p} + \mathbf{x}, \mathbf{p} + \mathbf{y}, \mathbf{p} + \mathbf{x} + \mathbf{y}\}$  (gewichtete Fläche)

Fig. 160

“Riemann-Summation”

➤ Fluss durch beschränkte orientierte differenzierbare Fläche (= Mannigfaltigkeit der Dimension 2)

▷

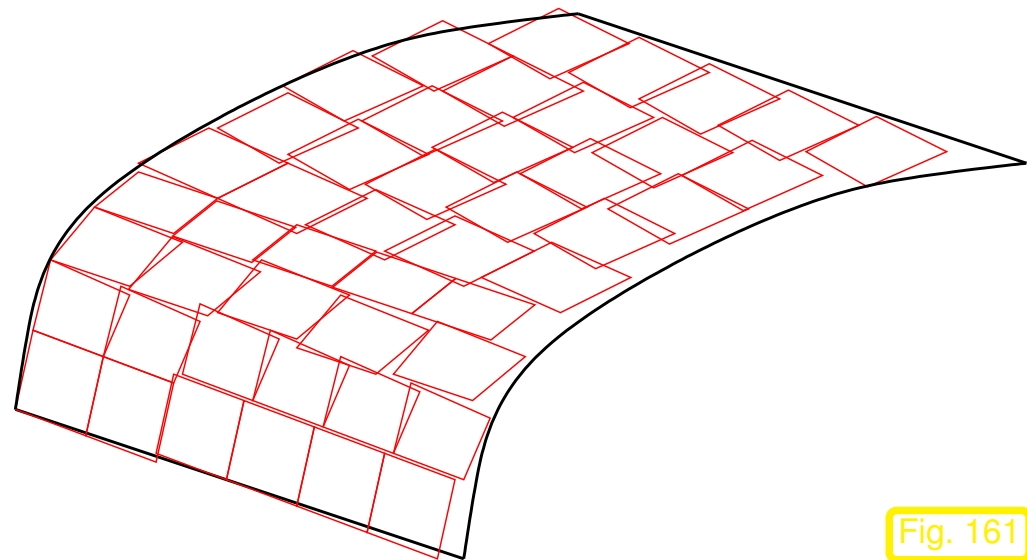


Fig. 161

Ist  $\psi : U \mapsto \mathbb{R}^d$  eine Parametrisierung (Karte) der 2-Mannigfaltigkeit  $\Sigma$ , so gilt, vgl. Push-Forward,

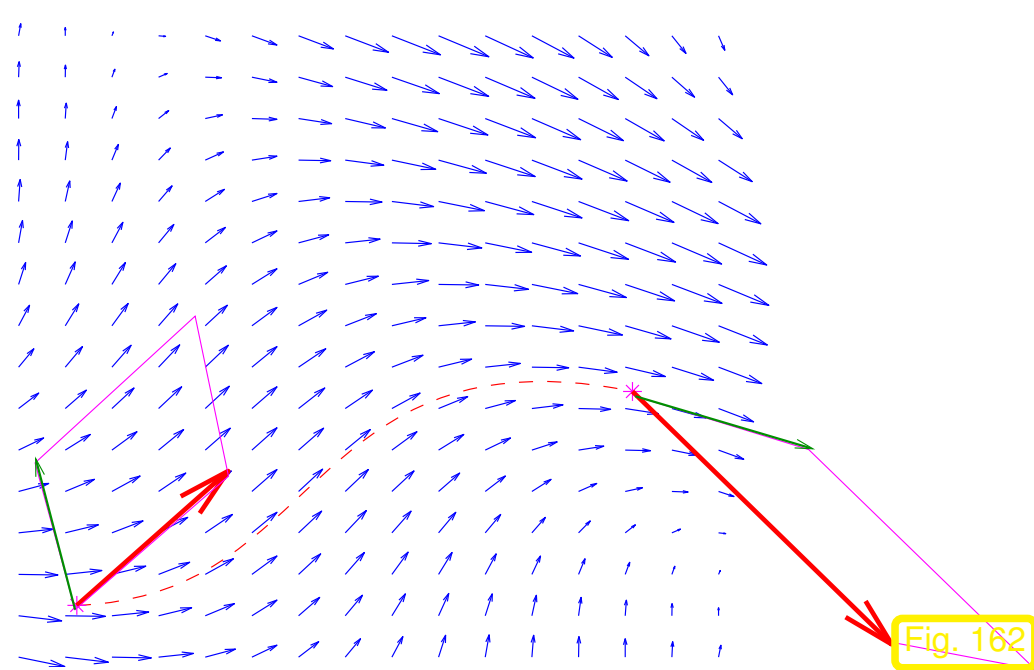
$$\text{Fluss} = \int_{\Sigma} \omega = \int_U \left( \frac{d\psi}{du_1} \right)^T \mathbf{J} \left( \frac{d\psi}{du_2} \right) d\mathbf{u} . \quad (4.4.10)$$



**Theorem 4.4.11** (Symplektischer Fluss Hamiltonscher Systeme).

Sei  $\Phi^t$  die Evolution zu einer Hamiltonschen Differentialgleichung (1.2.21) mit  $C^2$ -Hamilton-Funktion  $H : \mathbb{R}^n \times M \mapsto \mathbb{R}$ . Dann gilt

$$\forall \mathbf{y} \in D: \quad \exists \delta > 0: \quad \omega((\Phi_*^t \mathbf{v})(\mathbf{y}), (\Phi_*^t \mathbf{w})(\mathbf{y})) = \omega(\mathbf{v}, \mathbf{w}) \quad \forall \mathbf{v}, \mathbf{w} \in \mathbb{R}^{2n}, 0 \leq t < \delta .$$



◁ Veranschaulichung Push-Forward von zwei Vektoren und alternierende (Flächen)Bilinearform.

*Beweis.* (→ Beweis von [16, Thm. 2.4, Ch. VI])

$\Phi^t \hat{=}$  Evolutionsoperator zur Hamiltonschen ODE  $\dot{\mathbf{y}} = \mathbf{J}^{-1} \text{grad } H(\mathbf{y})$

Behauptung  $\iff (\Phi_*^t(\mathbf{v})\mathbf{y})^T \mathbf{J}(\Phi_*^t(\mathbf{w})\mathbf{y}) = \mathbf{v}^T \mathbf{J}\mathbf{w} \quad \forall \mathbf{v}, \mathbf{w} \in \mathbb{R}^d, \forall \mathbf{y} \in D,$   
 $\iff \left(\frac{d}{d\mathbf{y}}\Phi^t(\mathbf{y})\right)^T \mathbf{J} \left(\frac{d}{d\mathbf{y}}\Phi^t(\mathbf{y})\right) = \mathbf{J} \quad \forall \mathbf{y} \in D.$

Propagationsmatrix  $\mathbf{W}(t; \mathbf{y}) := \frac{d}{d\mathbf{y}}\Phi^t\mathbf{y}$  löst Variationsgleichung (1.3.34)

$$\dot{\mathbf{W}}(t; \mathbf{y}) = D(\mathbf{J}^{-1} \text{grad } H(\mathbf{y}))\mathbf{W}(t; \mathbf{y}) = \mathbf{J}^{-1} \nabla^2 H(\mathbf{y})\mathbf{W}(t; \mathbf{y}), \quad \mathbf{y} \in D.$$

Notation:  $\nabla^2 H \hat{=}$  (symmetrische) Hesse-Matrix der Hamilton-Funktion  $H$ .

Mit Produktregel, da  $\mathbf{J}^T = -\mathbf{J}$ ,  $\mathbf{J}^{-T} = -\mathbf{J}^{-1} = \mathbf{J}$ :

$$\begin{aligned} \frac{d}{dt} \left( \mathbf{W}(t; \mathbf{y})^T \mathbf{J} \mathbf{W}(t; \mathbf{y}) \right) &= \dot{\mathbf{W}}(t; \mathbf{y})^T \mathbf{J} \mathbf{W}(t; \mathbf{y}) + \mathbf{W}(t; \mathbf{y})^T \mathbf{J} \dot{\mathbf{W}}(t; \mathbf{y}) \\ &= \mathbf{W}(t; \mathbf{y})^T \nabla^2 H(\mathbf{y}) \underbrace{\mathbf{J}^{-T} \mathbf{J}}_{=-\mathbf{I}} \mathbf{W}(t; \mathbf{y}) + \mathbf{W}(t; \mathbf{y})^T \underbrace{\mathbf{J} \mathbf{J}^{-1}}_{=\mathbf{I}} \nabla^2 H(\mathbf{y}) \mathbf{W}(t; \mathbf{y}) = 0 . \end{aligned}$$

Da  $\mathbf{W}(0; \mathbf{y}) = \mathbf{I} \Rightarrow \mathbf{W}(t; \mathbf{y})^T \mathbf{J} \mathbf{W}(t; \mathbf{y}) = \mathbf{J} \quad \forall t$  □

**Definition 4.4.12** (Symplektische Abbildung).

Eine  $C^1$ -Abbildung  $\Phi : D \subset \mathbb{R}^{2n} \mapsto \mathbb{R}^{2n}$  heisst *symplektisch*, falls

$$D\Phi(\mathbf{y})^T \mathbf{J} D\Phi(\mathbf{y}) = \mathbf{J} \Leftrightarrow \omega \left( \underbrace{D\Phi(\mathbf{y})\mathbf{v}}_{(\Phi_*\mathbf{v})(\mathbf{y})}, \underbrace{D\Phi(\mathbf{y})\mathbf{w}}_{(\Phi_*\mathbf{w})(\mathbf{y})} \right) = \omega(\mathbf{v}, \mathbf{w}) \quad \forall \mathbf{v}, \mathbf{w} \in \mathbb{R}^{2n}, \forall \mathbf{y} \in D .$$

Thm. 4.4.11: Die Evolution zu einer Hamiltonschen Differentialgleichung ist symplektisch zu jedem Zeitpunkt.

Das Konzept der Symplektizität ist eng verbunden mit der differentialgeometrischen Betrachtung Hamiltonscher Evolutionen, siehe [2, Part III].

**Korollar 4.4.13** (Komposition symplektischer Abbildungen).

*Die Komposition symplektischer Abbildungen ist symplektisch.*

*Bemerkung 4.4.14* (Vektorräume von Vektorfeldern und Eigenschaften von Evolutionen).

Def. 1.3.7: Vektorfeld  $\mathbf{f} : D \subset \mathbb{R}^d \mapsto \mathbb{R}^d$   $\triangleright$  Evolution  $\Phi^t$  zur ODE  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$

(1.2.8):	$\mathbf{grad} I \cdot \mathbf{f} = 0$	$\Leftrightarrow$	$\Phi^t$ “ $I$ -isoflächenerhaltend” für alle $t$
Thm. 4.2.3:	$\operatorname{div} \mathbf{f} = 0$	$\Leftrightarrow$	$\Phi^t$ volumenerhaltend ( $\rightarrow$ Def. 4.2.1) $\forall t$
Lemma 4.3.12:	$\mathbf{f} \circ R = -R \circ \mathbf{f}$	$\Leftrightarrow$	$\Phi^t$ $R$ -reversibel ( $\rightarrow$ Def. 4.3.11) $\forall t$
Thm. 4.4.11:	$\mathbf{f} = \mathbf{J}^{-1} \mathbf{grad} H$	$\Rightarrow$	$\Phi^t$ symplektisch ( $\rightarrow$ Def. 4.4.12) $\forall t$

Vektorraum  $V$  von Vektorfeldern  $D \mapsto \mathbb{R}^d$   $\blacktriangleright$  Gruppe  $\mathcal{G}$  von Diffeomorphismen

Einschrittverfahren für  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$  **strukturerhaltend**  $:\Leftrightarrow \mathbf{f} \in V \Rightarrow \Psi^h \in \mathcal{G}$   
(mit diskreter Evolution  $\Psi^h$ )

**Theorem 4.4.15** (Symplektische Evolutionen und Hamiltonsche Differentialgleichungen).

Sei  $\Phi^t$  der Evolutionsoperator zu einer autonomen ODE  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ ,  $\mathbf{f} : D \subset \mathbb{R}^{2n} \mapsto \mathbb{R}^{2n}$  stetig differenzierbar, Zustandsraum  $D$  sternförmig. Dann gilt

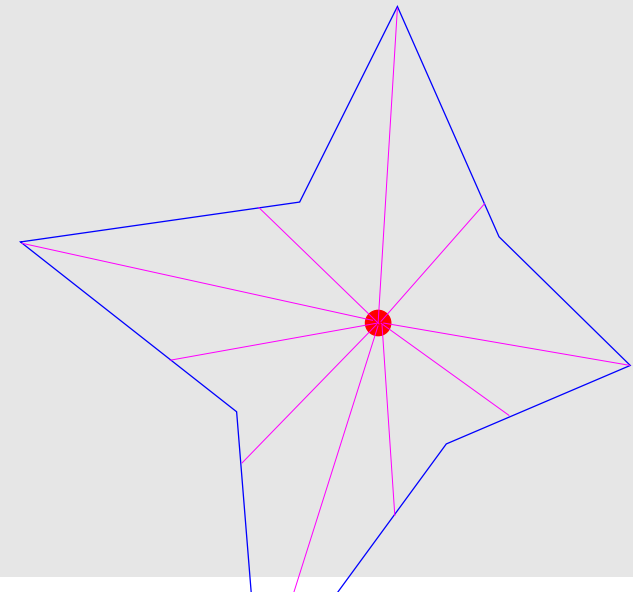
$$\Phi^t \text{ symplektisch } (\rightarrow \text{Def. 4.4.12}) \forall t \quad \Leftrightarrow \quad \exists H : D \mapsto \mathbb{R} : \mathbf{f}(\mathbf{y}) = \mathbf{J}^{-1} \text{grad } H(\mathbf{y}) .$$

**Definition 4.4.16** (Sternförmiges Gebiet).

$D \subset \mathbb{R}^d$  heisst **sternförmig**, wenn es  $\mathbf{z} \in D$  gibt, so dass

$$\{t\mathbf{z} + (1-t)\mathbf{x}, 0 \leq t \leq 1\} \subset D$$

für alle Punkte  $\mathbf{x} \in D$ .





Hilfsmittel beim Beweis:

**Lemma 4.4.17** (Integrabilitätslemma).

Es sei  $D \subset \mathbb{R}^d$  sternförmig und  $\mathbf{f} : D \mapsto \mathbb{R}^d$  stetig differenzierbar. Dann gilt

$$D\mathbf{f} = D\mathbf{f}^T \iff \exists F : D \mapsto \mathbb{R} : \mathbf{f}(\mathbf{y}) = \text{grad } F(\mathbf{y}) \quad \forall \mathbf{y} \in D .$$

*Beweis.* O.B.d.A:  $D$  sternförmig bzgl.  $0 \Rightarrow$  Wohldefiniert ist die Funktion

$$F(\mathbf{y}) := \int_0^1 \mathbf{f}(\tau \mathbf{y}) \cdot \mathbf{y} \, d\tau \quad \mathbf{y} \in D .$$

$$\Rightarrow \text{grad } F(\mathbf{y}) = DF(\mathbf{y})^T = \int_0^1 \tau D\mathbf{f}(\tau \mathbf{y})^T \cdot \mathbf{y} + \mathbf{f}(\tau \mathbf{y}) \, d\tau = \int_0^1 \frac{d}{d\tau} (\mathbf{f}(\tau \mathbf{y})\tau) (\tau) \, d\tau = \mathbf{f}(\mathbf{y}) .$$

Vorsicht: strikte Unterscheidung von Zeilen- und Spaltenvektoren wichtig; Gradient ist ein Spaltenvektor! □

*Beweis* (von Thm. 4.4.15)

“ $\Leftarrow$ ”: Siehe Thm. 4.4.11

“ $\Rightarrow$ ”: Propagationsmatrix  $\mathbf{W}(t; \mathbf{y}) := \left(\frac{d}{dy}\Phi^t\right)(\mathbf{y})$  löst Variationsgleichung (1.3.34)

$$\dot{\mathbf{W}}(t; \mathbf{y}) = D\mathbf{f}(\Phi^t \mathbf{y}) \mathbf{W}(t; \mathbf{y}) \quad , \quad \mathbf{W}(0; \mathbf{y}) = \mathbf{I} \quad , \quad \mathbf{y} \in D \quad , \quad t \in J(\mathbf{y}) .$$

$t$  fixiert, hinreichend klein:  $\mathbf{y} \mapsto \Phi^t \mathbf{y}$  ist symplektische Abbildung ( $\rightarrow$  Def. 4.4.12)

$$\blacktriangleright \quad \mathbf{W}(t; \mathbf{y})^T \mathbf{J} \mathbf{W}(t; \mathbf{y}) = \mathbf{J} \quad \stackrel{t \text{ frei}}{\implies} \quad \frac{d}{dt} \left( \mathbf{W}(t; \mathbf{y})^T \mathbf{J} \mathbf{W}(t; \mathbf{y}) \right) = 0 .$$

Mit Produktregel:

$$\begin{aligned} 0 &= \frac{d}{dt} \left( \mathbf{W}(t; \mathbf{y})^T \mathbf{J} \mathbf{W}(t; \mathbf{y}) \right) = \dot{\mathbf{W}}(t; \mathbf{y})^T \mathbf{J} \mathbf{W}(t; \mathbf{y}) + \mathbf{W}(t; \mathbf{y})^T \mathbf{J} \dot{\mathbf{W}}(t; \mathbf{y}) \\ &= (D\mathbf{f}(\Phi^t \mathbf{y}))^T \mathbf{J} \mathbf{W}(t; \mathbf{y}) + \mathbf{W}(t; \mathbf{y})^T \mathbf{J} (D\mathbf{f}(\Phi^t \mathbf{y})) \quad \forall \mathbf{y} \in D, |t| \text{ klein.} \end{aligned}$$

Setze  $t = 0$ , benutze  $\mathbf{J}^{-T} = -\mathbf{J}^{-1} = \mathbf{J} \quad \implies \quad \boxed{\mathbf{J} D\mathbf{f}(\mathbf{y}) = (\mathbf{J} D\mathbf{f}(\mathbf{y}))^T \quad \forall \mathbf{y} \in D}$

Wegen  $\mathbf{J} D\mathbf{f}(\mathbf{y}) = D(\mathbf{J}\mathbf{f})(\mathbf{y})$  Anwendung der Integrabilitätslemmas 4.4.17. □

Warum interessiert Numeriker diese „exotische“ Eigenschaft „Symplektizität“ ?

Thm. 4.4.15:  $\mathbf{f} = \mathbf{J}^{-1} \text{grad } H$   $\Leftrightarrow \Phi^t$  symplektisch ( $\rightarrow$  Def. 4.4.12)  $\forall t$   
 (“Bewegungsgleichung”)

Intuition: diskrete Evolution  $\Psi^h$  symplektisch  $\Leftrightarrow$  “Diskrete Bewegungsgleichung  
Symplektizität kann von diskreten Evolutionen geerbt werden !

R. Hiptmair  
rev 35327,  
25. April  
2011

**Definition 4.4.18** (Symplektisches Einschrittverfahren).

Ein Einschrittverfahren ( $\rightarrow$  Def. 2.1.2) heisst **symplektisch**, wenn es, angewendet auf eine Hamiltonsche Differentialgleichung ( $\rightarrow$  Def. 1.2.20)  $\dot{\mathbf{y}} = \mathbf{J}^{-1} \text{grad } H(\mathbf{y})$  eine konsistente diskrete Evolution  $\Psi^h$  erzeugt, so dass  $\Psi^h : K \subset D \mapsto \mathbb{R}^d$  für jedes Kompaktum  $K \subset D$  und festes hinreichend kleines  $h > 0$  eine symplektische Abbildung ( $\rightarrow$  Def. 4.4.12) ist.

**Bemerkung 4.4.19** (Einfache symplektische Integratoren).

Die diskreten Evolutionsen  $\Psi^h : D \subset \mathbb{R}^{2n} \mapsto \mathbb{R}^{2n}$  zur Hamiltonsche ODE (1.2.21) ( $\dot{\mathbf{y}} = \mathbf{J}^{-1} \mathbf{grad} H(\mathbf{y})$ ,  $H : D \subset \mathbb{R}^d \mapsto \mathbb{R}$ ) erzeugt durch

- implizite Mittelpunkteregel (1.4.19)
- symplektisches Eulerverfahren (2.5.11)
- Störmer-Verlet-Verfahren (2.5.13) ( $\rightarrow$  Bem. 1.4.33, Bsp. 2.5.10) für **separierte** Hamilton-Funktion der Form  $H(\mathbf{y}) = T(\mathbf{p}) + U(\mathbf{q})$ ,  $\mathbf{y} = \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix}$ ,

sind symplektisch (für hinreichend kleine Schrittweite  $h \in \mathbb{R}$ ).

Nachweise der Symplektizität:

- für implizite Mittelpunkteregel (1.4.19):

$$\Psi^h \mathbf{y}_0 := \mathbf{y}_1 = \mathbf{y}_0 + h \mathbf{J}^{-1} \mathbf{grad} H\left(\frac{1}{2}(\mathbf{y}_0 + \mathbf{y}_1)\right). \quad (4.4.20)$$

Implizites Differenzieren (Annahme:  $H$  "hinreichend glatt"):

$$\begin{aligned} D\Psi^h(\mathbf{y}_0) &= \mathbf{I} + h \mathbf{J}^{-1} \nabla^2 H\left(\frac{1}{2}(\mathbf{y}_0 + \mathbf{y}_1)\right) \frac{1}{2}(\mathbf{I} + D\Phi^h(\mathbf{y}_0)), \\ \Rightarrow D\Psi^h(\mathbf{y}_0) &= \left(\mathbf{I} - \frac{1}{2}h \mathbf{J}^{-1} \nabla^2 H(\dots)\right)^{-1} \left(\mathbf{I} + \frac{1}{2}h \mathbf{J}^{-1} \nabla^2 H(\dots)\right). \end{aligned}$$

Verwende nun

$$\mathbf{M} = \mathbf{M}^T \Rightarrow (\mathbf{I} - \mathbf{JM})^T (\mathbf{I} + \mathbf{JM})^{-T} \mathbf{J} (\mathbf{I} + V\mathbf{JM})^{-1} (\mathbf{I} - \mathbf{JM}) = \mathbf{J} . \quad (4.4.21)$$

- Störmer-Verlet-Verfahren (2.5.13) für  $H(\mathbf{p}, \mathbf{q}) = T(\mathbf{p}) + U(\mathbf{q})$ :

$$\begin{cases} \mathbf{p}_{1/2} = \mathbf{p}_0 - \frac{1}{2}h \mathbf{grad} U(\mathbf{q}_0) , \\ \mathbf{q}_1 = \mathbf{q}_0 + h \mathbf{grad} T(\mathbf{p}_{1/2}) , \\ \mathbf{p}_1 = \mathbf{p}_{1/2} - \frac{1}{2}h \mathbf{grad} U(\mathbf{q}_1) . \end{cases} \quad (4.4.22)$$

Strang-Splittingverfahren (Bem. 1.4.33): diskrete Evolution  $\Psi^h$  zu (4.4.22) erfüllt

$$\Psi^h = \Phi_U^{h/2} \circ \Phi_T^h \circ \Phi_U^{h/2} ,$$

wobei  $\Phi_T^t, \Phi_U^t$  exakte Evolutionsoperatoren zu Hamiltonschen ODE

$$\begin{aligned} \Phi_T^t &\leftrightarrow \begin{cases} \dot{\mathbf{p}} = 0 , \\ \dot{\mathbf{q}} = \mathbf{grad} T(\mathbf{p}) \end{cases} \rightarrow \text{Hamilton-Funktion } H(\mathbf{p}, \mathbf{q}) = T(\mathbf{p}) , \\ \Phi_U^t &\leftrightarrow \begin{cases} \dot{\mathbf{p}} = -\mathbf{grad} U(\mathbf{q}) , \\ \dot{\mathbf{q}} = 0 . \end{cases} \rightarrow \text{Hamilton-Funktion } H(\mathbf{p}, \mathbf{q}) = U(\mathbf{q}) . \end{aligned}$$

Korollar 4.4.13  $\Rightarrow \Psi^h$  is symplektische Abbildung ( $\rightarrow$  Def. 4.4.12).

Terminologie: implizite Mittelpunktsregel/Störmer-Verlet-Verfahren = **symplektische Integratoren**



**Theorem 4.4.23** (Symplektische Runge-Kutta-Einschrittverfahren).

*Alle Runge-Kutta-Einschrittverfahren ( $\rightarrow$  Def. 2.3.5), die quadratische Invarianten erhalten, sind symplektisch.*

*Beweis.*  $\Phi^t \hat{=}$  Evolutionsoperator zu Hamiltonschen Dgl.  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}) := \mathbf{J}^{-1} \text{grad } H(\mathbf{y})$  ist eine symplektische Abbildung für (alle zulässigen)  $t$

Def. 4.4.12  $\Rightarrow I(\mathbf{Y}) := \mathbf{Y}^T \mathbf{J} \mathbf{Y}$  ist **quadratisches erstes Integral** der Variationsgleichung

$$\dot{\mathbf{W}}(t; \mathbf{y}) = D\mathbf{f}(\Phi^t \mathbf{y}) \mathbf{W}(t; \mathbf{y}) .$$

$\Psi^h \hat{=}$  diskrete Evolution des EK-ESV für  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$

$\hat{\Psi}^h \hat{=}$  diskrete Evolution des EK-ESV für  $\begin{cases} \dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}) , \\ \dot{\mathbf{W}} = D\mathbf{f}(\mathbf{y}) \mathbf{W} \end{cases} : \quad \blacktriangleright \quad \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{W}_1 \end{pmatrix} = \hat{\Psi}^h \begin{pmatrix} \mathbf{y}_0 \\ \mathbf{I} \end{pmatrix} .$

Lemma 4.2.7  $\Rightarrow$

$$\frac{d\Psi^h}{d\mathbf{y}}(\mathbf{y}_0) = \mathbf{W}_1 = \left( \hat{\Psi}^h \begin{pmatrix} \mathbf{y}_0 \\ \mathbf{I} \end{pmatrix} \right)_{\mathbf{W}} .$$

Nach Voraussetzung erhält  $\widehat{\Psi}^h$  quadratische erste Integrale,

$$\left( \frac{d\Psi^h}{dy}(\mathbf{y}_0) \right)^T \mathbf{J} \left( \frac{d\Psi^h}{dy}(\mathbf{y}_0) \right) = \mathbf{W}_1^T \mathbf{J} \mathbf{W}_1 = \mathbf{J} \quad \forall \mathbf{y}_0 \in D. \quad \square$$

Thm. 4.1.4  $\Rightarrow$  Alle Gauss-Kollokations-Einschrittverfahren sind symplektisch.

Beispiel 4.4.24 (Symplektisches Euler-Verfahren). siehe Bsp. 2.5.10

Annahme: **Separierte** Hamilton-Funktion der Form  $H(\mathbf{p}, \mathbf{q}) = T(\mathbf{p}) + U(\mathbf{q})$ ,  $T, U : D \subset \mathbb{R}^n \mapsto \mathbb{R}$  glatt

$H(\mathbf{p}, \mathbf{q}) = T(\mathbf{p}) + U(\mathbf{q}) \iff$  Splitting der rechten Seite von (1.2.21), vgl. Bsp. 2.5.10

$$\mathbf{f}(\mathbf{y}) = \mathbf{J}^{-1} \text{grad } H(\mathbf{y}) = \begin{pmatrix} -\text{grad } U(\mathbf{q}) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \text{grad } T(\mathbf{p}) \end{pmatrix} =: \mathbf{f}_1(\mathbf{y}) + \mathbf{f}_2(\mathbf{y}) \quad (4.4.25)$$

➤ Lie-Trotter-Splitting-Einschrittverfahren (2.5.2)

$$\begin{aligned} \mathbf{p}_{k+1} &= \mathbf{p}_k - h \text{grad } U(\mathbf{q}_k) \\ \mathbf{q}_{k+1} &= \mathbf{q}_k + h \text{grad } T(\mathbf{p}_{k+1}), \end{aligned} \quad \text{bzw.} \quad \begin{aligned} \mathbf{p}_{k+1} &= \mathbf{p}_k - h \text{grad } U(\mathbf{q}_{k+1}) \\ \mathbf{q}_{k+1} &= \mathbf{q}_k + h \text{grad } T(\mathbf{p}_k). \end{aligned} \quad (4.4.26)$$

(4.4.26) = **explizite** symplektische diskrete Evolutionen (Thm. 2.5.5: Konsistenzordnung 1)Beachte: In (4.4.26) (links): Inkrement benutzt  $\mathbf{q}_k, \mathbf{p}_{k+1}$ In (4.4.26) (rechts): Inkrement benutzt  $\mathbf{q}_{k+1}, \mathbf{p}_k$ ☞ Verallgemeinerung von (4.4.26) auf  $\dot{\mathbf{y}} = \mathbf{J}^{-1} \text{grad } H(\mathbf{y})$  in der Form,  $\mathbf{y} := \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix}$ ,

$$\dot{\mathbf{p}}(t) = -\frac{\partial H}{\partial \mathbf{q}}(\mathbf{p}(t), \mathbf{q}(t)) \quad , \quad \dot{\mathbf{q}}(t) = \frac{\partial H}{\partial \mathbf{p}}(\mathbf{p}(t), \mathbf{q}(t)) : \quad (1.2.21)$$

$$\mathbf{y}_{k+1} = \mathbf{y}_k + h\mathbf{J}^{-1} \text{grad } H(\mathbf{p}_k, \mathbf{q}_{k+1}) \quad \text{bzw.} \quad \mathbf{y}_{k+1} = \mathbf{y}_k + h\mathbf{J}^{-1} \text{grad } H(\mathbf{p}_{k+1}, \mathbf{q}_k) .$$

R. Hiptmair  
rev 35327,  
25. April  
2011Für allgemeine Hamilton-Funktion  $H = H(\mathbf{p}, \mathbf{q})$  :**Symplektische Euler-Verfahren**

$$\begin{aligned} \mathbf{p}_{k+1} &= \mathbf{p}_k - h\frac{\partial H}{\partial \mathbf{q}}(\mathbf{p}_{k+1}, \mathbf{q}_k) & \text{bzw.} & \quad \mathbf{p}_{k+1} = \mathbf{p}_k - h\frac{\partial H}{\partial \mathbf{q}}(\mathbf{p}_k, \mathbf{q}_{k+1}) \\ \mathbf{q}_{k+1} &= \mathbf{q}_k + h\frac{\partial H}{\partial \mathbf{p}}H(\mathbf{p}_{k+1}, \mathbf{q}_k), & & \quad \mathbf{q}_{k+1} = \mathbf{q}_k + h\frac{\partial H}{\partial \mathbf{p}}H(\mathbf{p}_k, \mathbf{q}_{k+1}) . \end{aligned} \quad (4.4.27)$$

☞ kein Splittingverfahren mehr, trotzdem symplektisch [16, Thm. 3.3] !





**Bemerkung 4.4.28** (Partitionierte Runge-Kutta-Einschrittverfahren).

Mit lokal Lipschitz-stetigen  $\mathbf{f}_u : D_u \times D_v \mapsto \mathbb{R}^n$ ,  $\mathbf{f}_v : D_u \times D_v \mapsto \mathbb{R}^n$ ,  $D_u, D_v \subset \mathbb{R}^n$

$$\text{ODE: } \begin{aligned} \dot{\mathbf{u}} &= \mathbf{f}_u(\mathbf{u}, \mathbf{v}) , \\ \dot{\mathbf{v}} &= \mathbf{f}_v(\mathbf{u}, \mathbf{v}) . \end{aligned} \quad (4.4.29)$$

“Symplektisches Euler-Verfahren” (4.4.27) für (4.4.29):

$$\begin{aligned} \mathbf{u}_1 &= \mathbf{u}_0 + h\mathbf{f}_u(\mathbf{u}_1, \mathbf{v}_0) , \\ \mathbf{v}_1 &= \mathbf{v}_0 + h\mathbf{f}_v(\mathbf{u}_1, \mathbf{v}_0) . \end{aligned} \quad \triangleright \text{ Konsistenzordnung 1.} \quad (4.4.30)$$

Ansatz:  $s$ -stufige **partitionierte Runge-Kutta-Einschrittverfahren** (für autonome ODE)

$$\left\{ \begin{aligned} \mathbf{k}_i^u &= \mathbf{f}_u\left(\mathbf{u}_0 + h \sum_{j=1}^s a_{ij}^u \mathbf{k}_j^u, \mathbf{v}_0 + h \sum_{j=1}^s a_{ij}^v \mathbf{k}_j^v\right) , \\ \mathbf{k}_i^v &= \mathbf{f}_v\left(\mathbf{u}_0 + h \sum_{j=1}^s a_{ij}^u \mathbf{k}_j^u, \mathbf{v}_0 + h \sum_{j=1}^s a_{ij}^v \mathbf{k}_j^v\right) \end{aligned} \right. \quad i = 1, \dots, s , \quad (4.4.31)$$

$$\left\{ \begin{aligned} \mathbf{u}_1 &= \mathbf{u}_0 + \sum_{i=1}^s b_i^u \mathbf{k}_i^u , \\ \mathbf{v}_1 &= \mathbf{v}_0 + \sum_{i=1}^s b_i^v \mathbf{k}_i^v . \end{aligned} \right.$$

in Stufenform, vgl. Bem. 2.3.7:

$$\left\{ \begin{array}{l} \mathbf{g}_i^u = \mathbf{u}_0 + h \sum_{j=1}^s a_{ij}^u \mathbf{f}_u(\mathbf{g}_j^u, \mathbf{g}_j^v) , \\ \mathbf{g}_i^v = \mathbf{v}_0 + h \sum_{j=1}^s a_{ij}^v \mathbf{f}_v(\mathbf{g}_j^u, \mathbf{g}_j^v) , \end{array} \right. , \quad \left\{ \begin{array}{l} \mathbf{u}_1 = \mathbf{u}_0 + \sum_{i=1}^s b_i^u \mathbf{f}_u(\mathbf{g}_i^u, \mathbf{g}_i^v) , \\ \mathbf{v}_1 = \mathbf{v}_0 + \sum_{i=1}^s b_i^v \mathbf{f}_v(\mathbf{g}_i^u, \mathbf{g}_i^v) . \end{array} \right. \quad (4.4.32)$$

Darstellung: Zwei Butcher-Tableaus:

$$\frac{\mathbf{c}^u \mid \mathfrak{A}^u}{\mid \mathbf{b}^{u,T}} \quad \& \quad \frac{\mathbf{c}^v \mid \mathfrak{A}^v}{\mid \mathbf{b}^{v,T}}$$

Symplektisches Euler-Verfahren

$$\frac{0 \mid 0}{\mid 1} \quad \& \quad \frac{1 \mid 1}{\mid 1}$$

Störmer-Verlet-Verfahren

$$\frac{1/2 \mid 1/2 \quad 0}{\mid 1/2 \quad 1/2} \quad \& \quad \frac{0 \mid 0 \quad 0}{1 \mid 1/2 \quad 1/2} \\ \frac{\phantom{0} \mid \phantom{0} \quad \phantom{0}}{\mid 1/2 \quad 1/2}$$

In Analogie zur Theorie der konventionellen RK-ESV aus Def. 2.3.5:

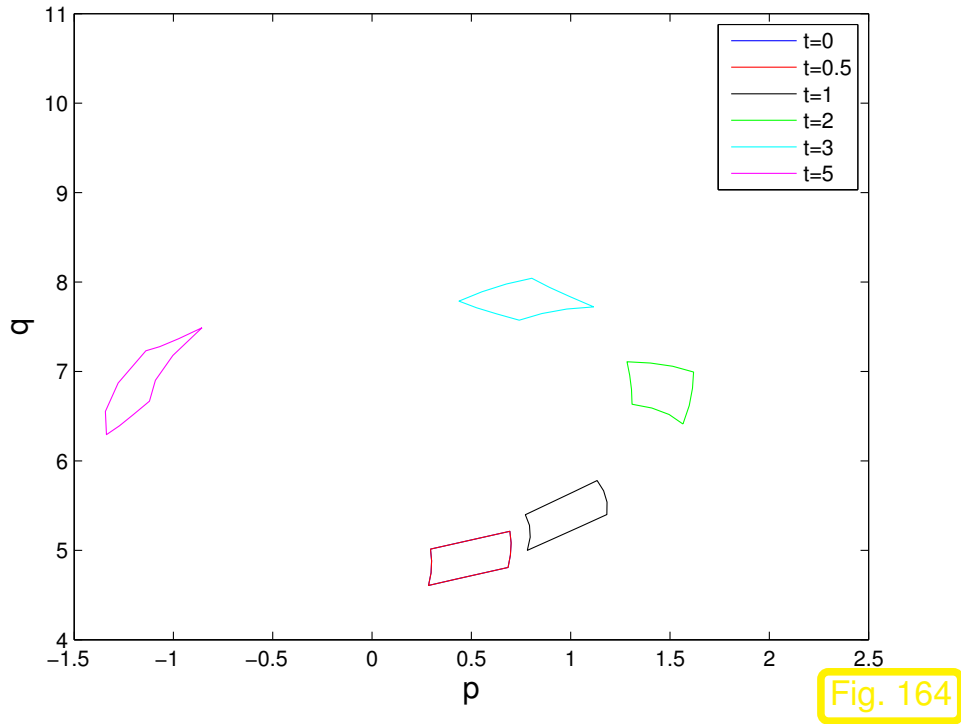
- Bedingungsgleichungen an Koeffizienten für gewünschte Konsistenzordnungen, vgl. Sect. 2.3.2 [16, Sect. II.2]

- Algebraische Bedingungen für Erhaltung quadratischer Invarianten [16, Sect. IV.2.2], vgl. Lemma 4.1.6, und Symmetrie, vgl. Thm. 4.3.1 [16, Sect. V.2.2],
- Koeffizientenbedingungen für Symplektizität, vgl. Thm. 4.4.23 [16, Sect. VI.4].

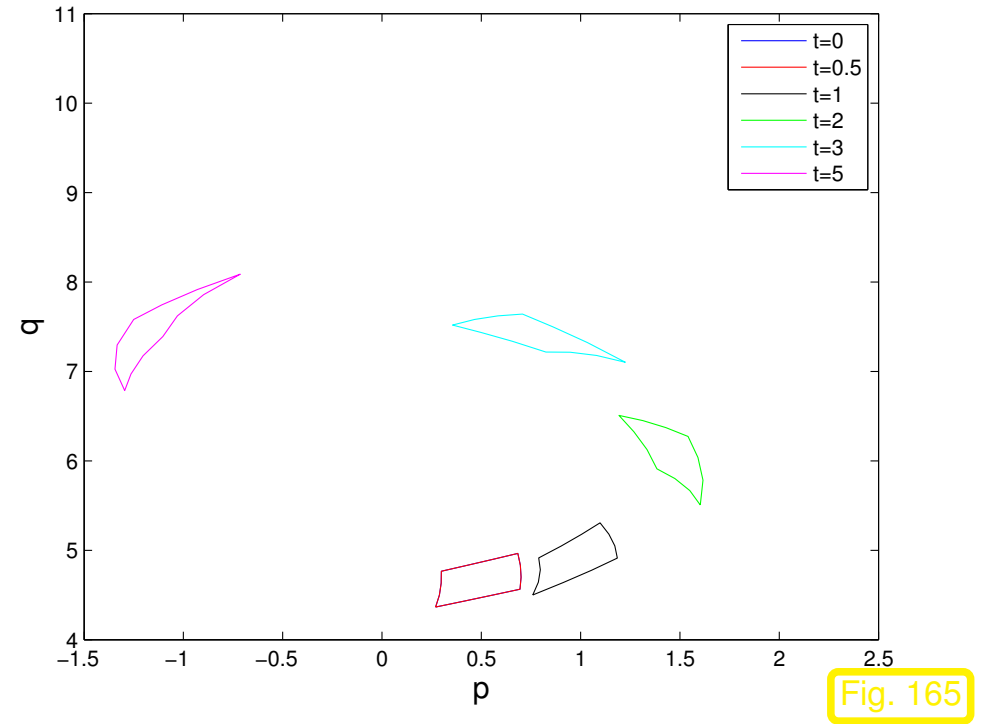


*Beispiel* 4.4.33 (Symplektisches Euler-Verfahren für Pendelgleichung).

AWP für Pendelgleichung wie in Bsp. 4.4.3.



Evolution eines quadratischen Volumens  
(Verfahren (4.4.27), links)



Evolution eines quadratischen Volumens  
(Verfahren (4.4.27), rechts)

Energieerhaltung des symplektischen partitionierten Eulerverfahrens (4.4.27) (links) ( $p(0) = 0, q(0) = \frac{7\pi}{6}$ )

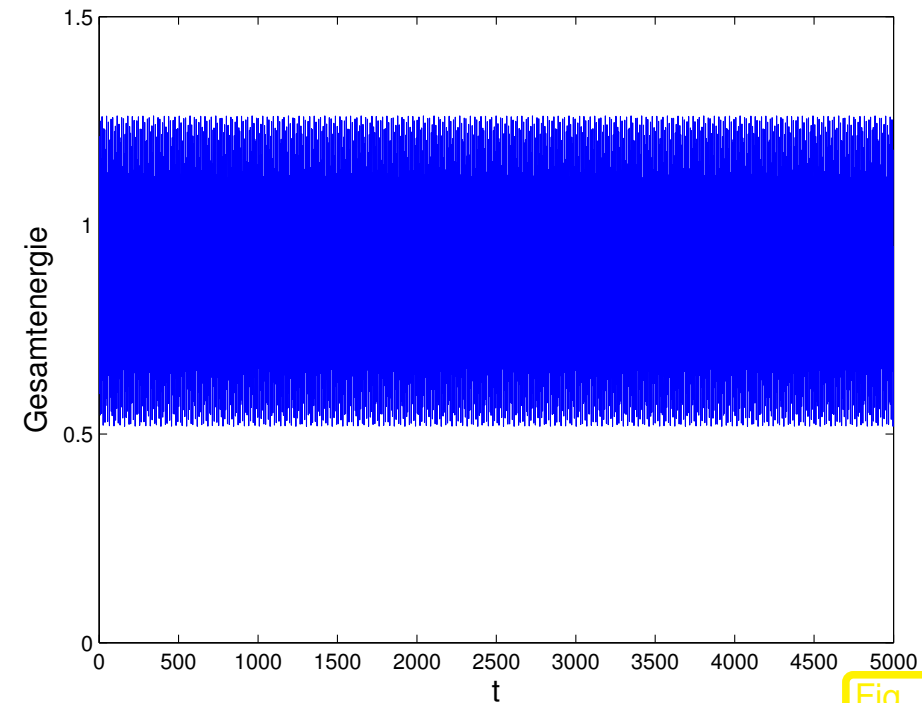


Fig. 166

R. Hiptmair  
 rev 35327,  
 25. April  
 2011

*Beispiel* 4.4.34 (Langzeit-Energieerhaltung bei symplektischer Integration). → Bsp. 4.4.1, 4.4.33, 1.4.32

Hamiltonsche Differentialgleichung (4.4.4) für mathematisches Pendel → 1.2.17 ( $p \leftrightarrow$  Winkelgeschwindigkeit,  $q \leftrightarrow$  Winkelvariable  $\alpha$ )

$$\begin{aligned} \dot{p} &= -\sin q, \\ \dot{q} &= p \end{aligned} \quad \blacktriangleright \quad \text{Hamilton-Funktion } H(p, q) = \frac{1}{2}p^2 - \cos q \quad (\text{Gesamtenergie}) \quad (4.4.4)$$

Anfangswerte:  $p(0) = 0, q(0) = 7/6\pi$ , Endzeitpunkt  $T = 5000$

- Symplektische ESV:
- Symplektisches partitioniertes Euler-Verfahren (4.4.27) (links)
  - Störmer-Verlet-Verfahren (4.4.22), siehe Bem. 4.4.19
  - Implizite Mittelpunktsregel (4.4.20), siehe Bem. 4.4.19
  - 2-stufiges Gauss-Kollokations-Einschrittverfahren, siehe Sect. 2.2.1

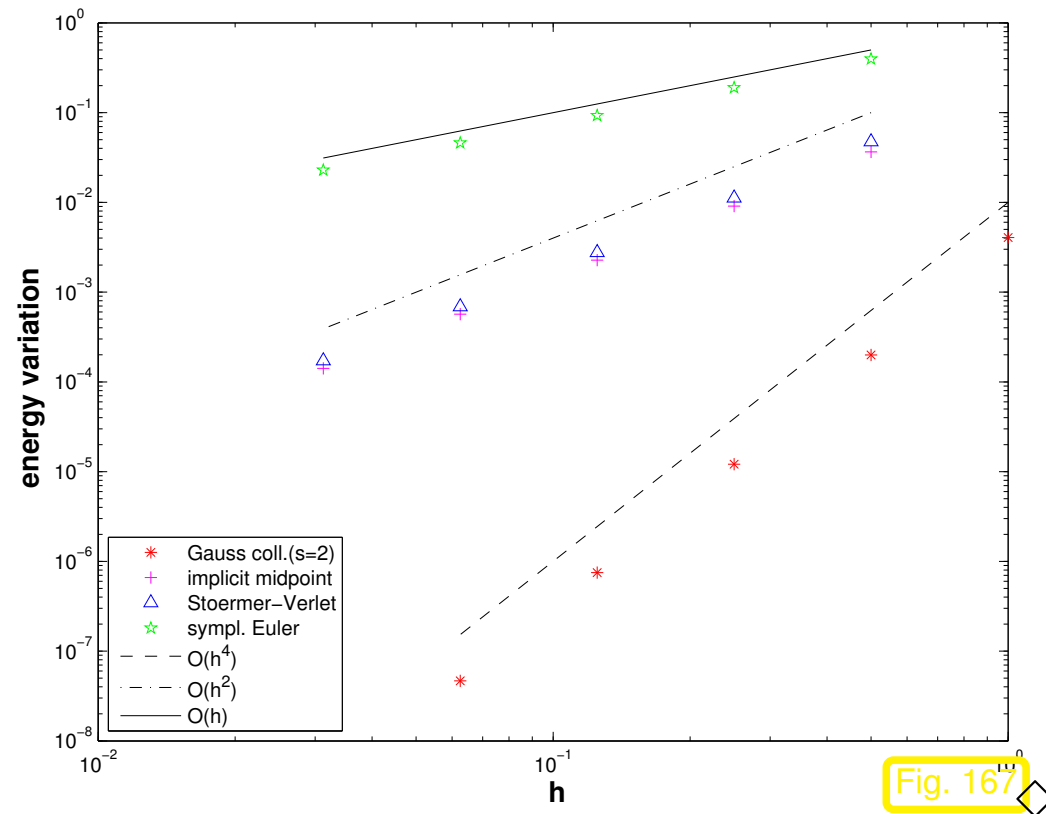
(uniforme Zeitschrittweite  $h > 0$ )

### Stärke der Energieschwankungen

$$E_{\text{var}}(h) = \max_{i=0, \dots, T/h} |E_h(ih) - E_{\text{exact}}| .$$

Sympl. Euler	$E_{\text{var}}(h) = O(h)$ ,
Störmer-Verlet	$E_{\text{var}}(h) = O(h^2)$ ,
Implizite MPR	$E_{\text{var}}(h) = O(h^2)$ ,
Gauss-Koll ( $s = 2$ )	$E_{\text{var}}(h) = O(h^4)$ .

Vermutung:  $E_{\text{var}}(h) = O(h^p)$   
 ( $p \hat{=}$  Konvergenzordnung des ESV)

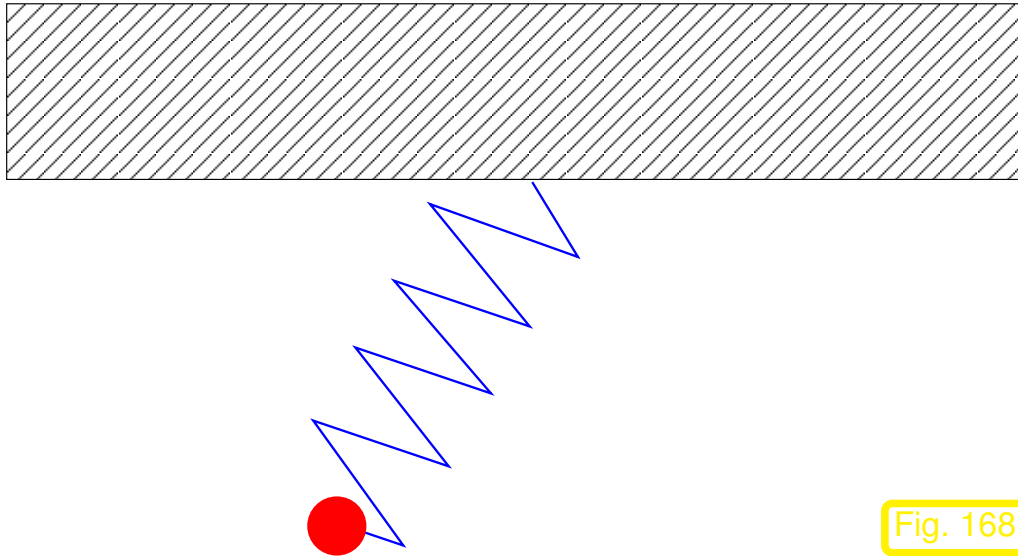


### Beispiel 4.4.35 (Federpendel).

Reibungsfreies Federpendel: Hamilton-Funktion  $H(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \|\mathbf{p}\|^2 + \frac{1}{2} (\|\mathbf{q}\| - 1)^2 + q_2$

( $\mathbf{q} \hat{=}$  Position,  $\mathbf{p} \hat{=}$  Impuls)

$$\blacktriangleright \quad \dot{\mathbf{p}} = -(\|\mathbf{q}\| - 1) \frac{\mathbf{q}}{\|\mathbf{q}\|} - \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \dot{\mathbf{q}} = \mathbf{p}. \quad (4.4.36)$$



Trajektorien bei Langzeitevolution  
(Chaotisches mechanisches System)

Fig. 168

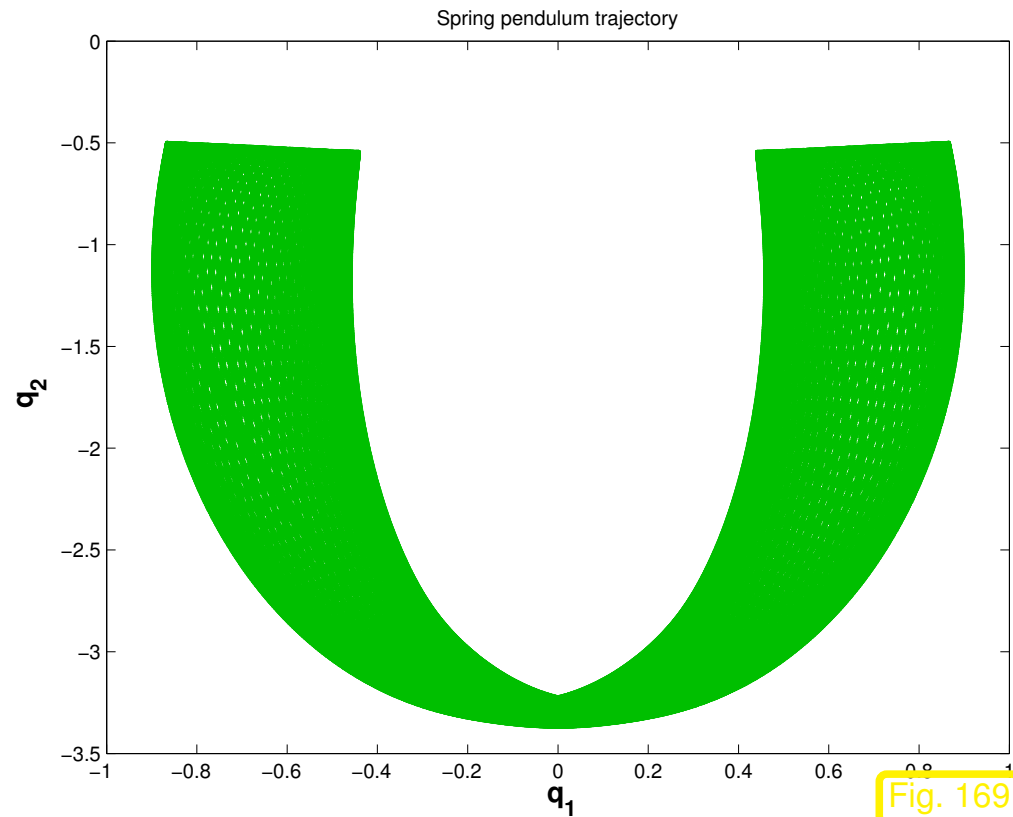
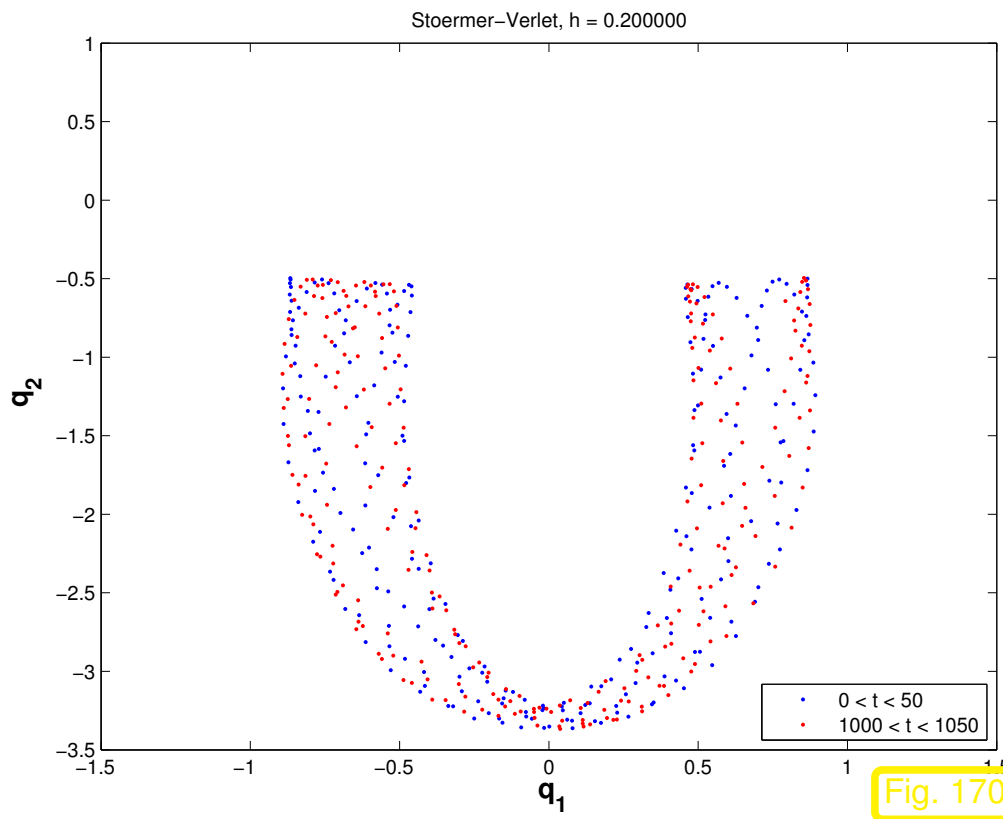


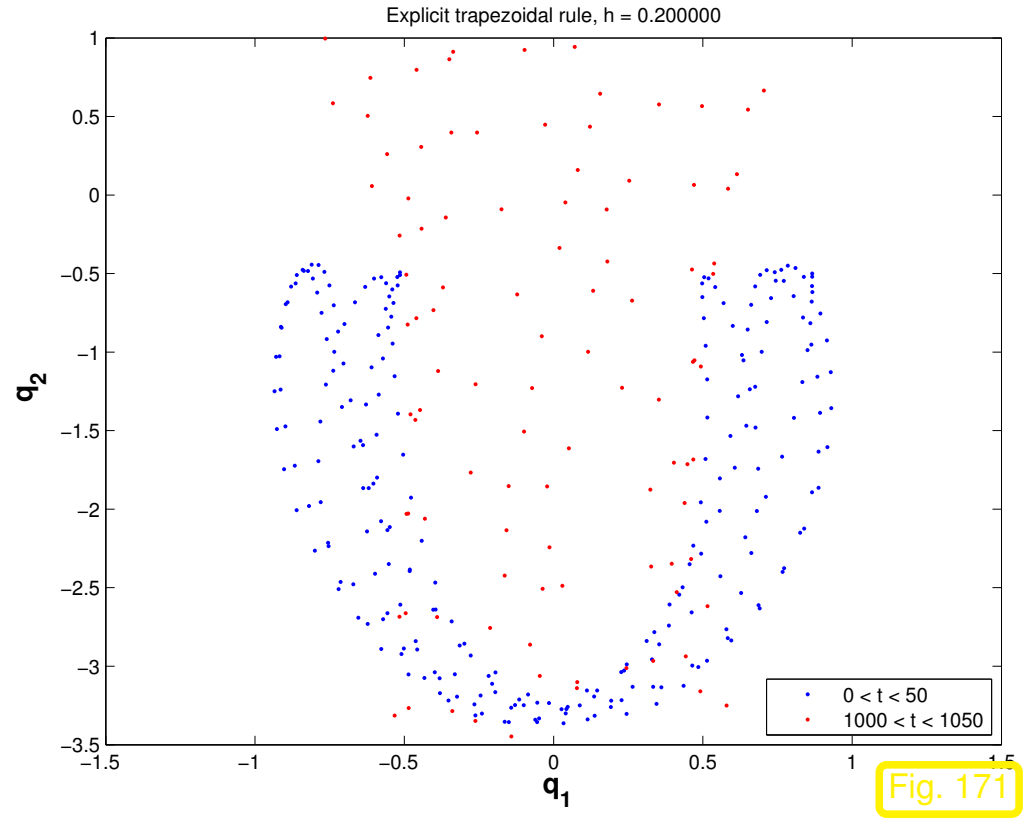
Fig. 169

R. Hiptmair  
rev 35327,  
25. April  
2011

- ESV:
- Störmer-Verlet-Verfahren (4.4.22) (Konsistenzordnung 2), siehe Bem. 4.4.19,
  - Explizite Trapezregel (2.3.3) (Konsistenzordnung 2).



Störmer-Verlet ESV



Explizite Trapezregel

Animation

Symplektischer Integrator: Positionen im “zulässigen Bereich” auch bei Langzeitintegration  
 Explizite Trapezregel: Trajektorien verlassen bei Langzeitintegration den “zulässigen Bereich” (Energiedrift !)





Beispiel 4.4.37 (Molekulardynamik). → [8, Sect. 1.2]

- Zustandsraum für  $n \in \mathbb{N}$  Atome in  $d \in \mathbb{N}$  Dimensionen:  $D = \mathbb{R}^{2dn}$   
 (Positionen  $\mathbf{q} = [\mathbf{q}^1; \dots; \mathbf{q}^n]^T \in \mathbb{R}^{dn}$ , Impulse  $\mathbf{p} = [\mathbf{p}^1, \dots, \mathbf{p}^n]^T \in \mathbb{R}^{dn}$ )

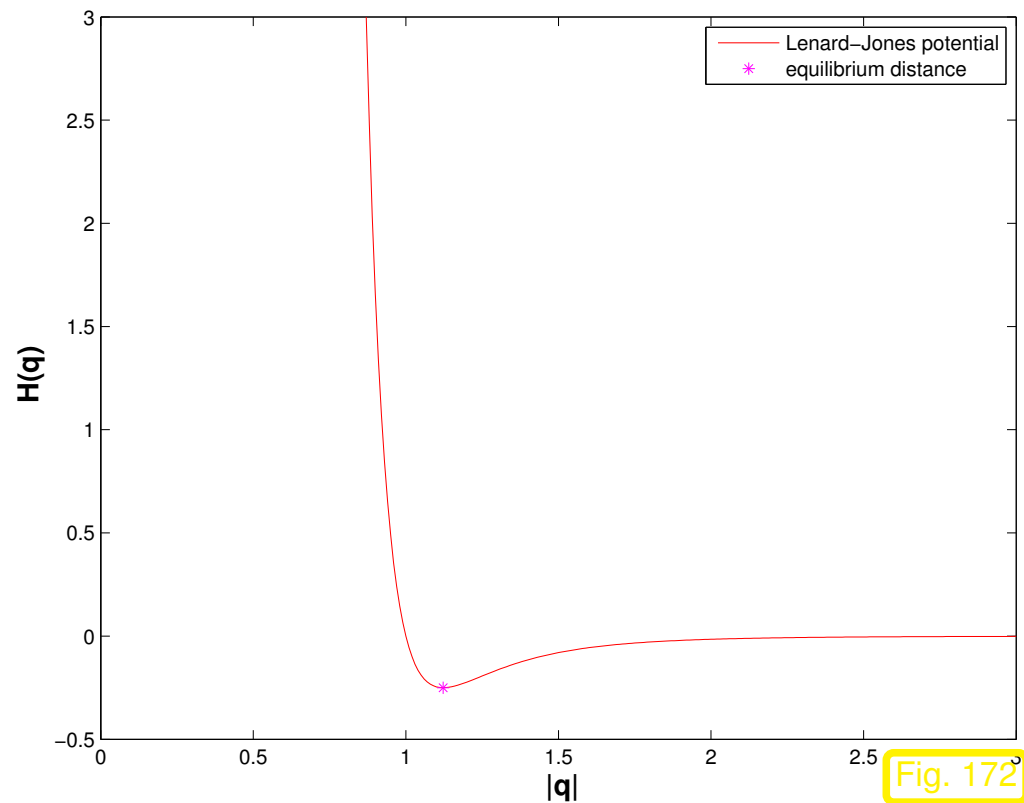
Gesamtenergie (Hamilton-Funktion):

$$H(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \|\mathbf{p}\|_2^2 + V(\mathbf{q}) .$$

Lenard-Jones-Potential:

$$V(\mathbf{q}) = \sum_{j=1}^n \sum_{i \neq j} \mathcal{V}(\|\mathbf{q}^i - \mathbf{q}^j\|_2) ,$$

$$\mathcal{V}(\xi) = \xi^{-12} - \xi^{-6} . \quad (4.4.38)$$



➔ Hamiltonsche Differentialgleichung (→ Def. 1.2.20):

$$\dot{\mathbf{p}}^j = - \sum_{i \neq j} \mathcal{V}'(\|\mathbf{q}^j - \mathbf{q}^i\|_2) \frac{\mathbf{q}^j - \mathbf{q}^i}{\|\mathbf{q}^j - \mathbf{q}^i\|_2}, \quad \dot{\mathbf{q}}^j = \mathbf{p}^j, \quad j = 1, \dots, n.$$

▶ Störmer-Verlet-Verfahren (4.4.22):

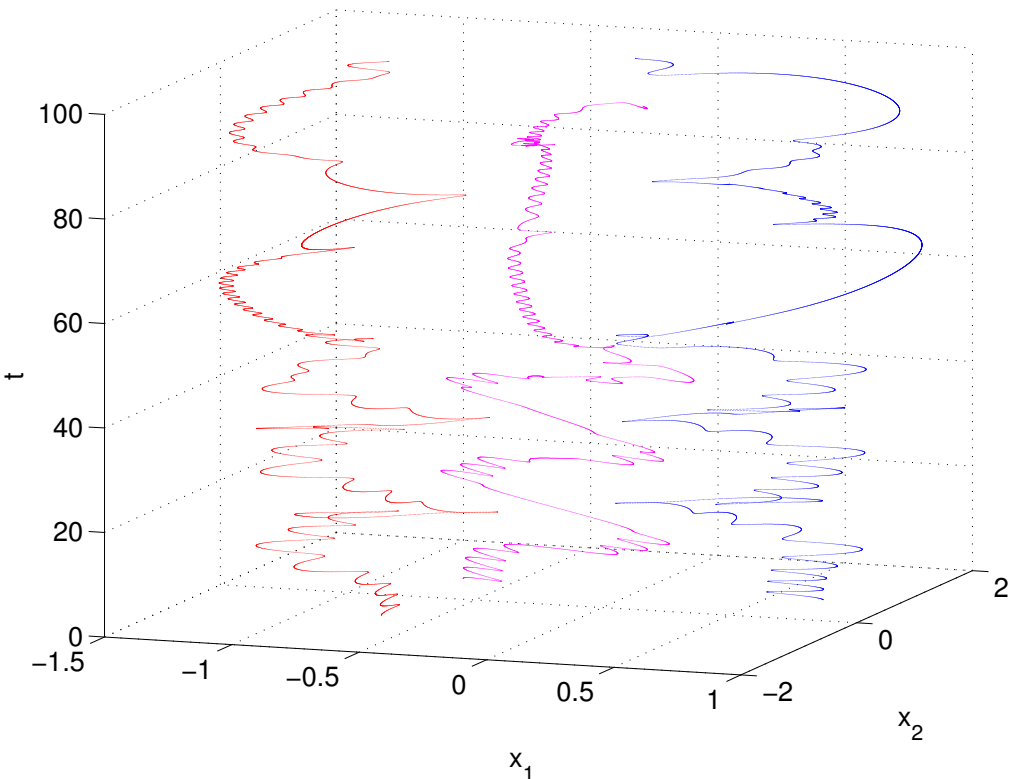
$$\mathbf{q}_h(t + \frac{1}{2}h) = \mathbf{q}_h(t) + \frac{h}{2}\mathbf{p}_h(t),$$

$$\mathbf{p}_h^j(t + h) = \mathbf{p}_h^j(t) - h \sum_{i \neq j} \mathcal{V}'(\|\mathbf{q}_h^j(t + \frac{1}{2}h) - \mathbf{q}_h^i(t + \frac{1}{2}h)\|_2) \frac{\mathbf{q}_h^j(t + \frac{1}{2}h) - \mathbf{q}_h^i(t + \frac{1}{2}h)}{\|\mathbf{q}_h^j(t + \frac{1}{2}h) - \mathbf{q}_h^i(t + \frac{1}{2}h)\|_2},$$

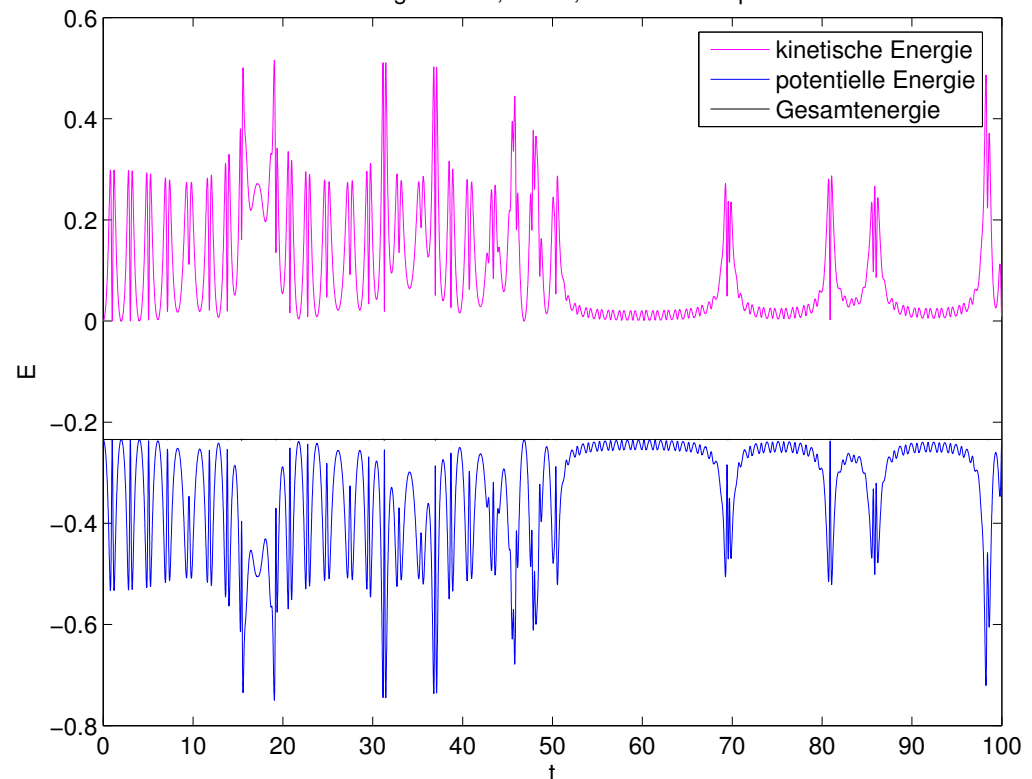
$$\mathbf{q}_h(t + h) = \mathbf{q}_h(t + \frac{1}{2}h) + \frac{h}{2}\mathbf{p}_h(t + h).$$

Simulation mit  $d = 2$ ,  $n = 3$ ,  $\mathbf{q}^1(0) = \frac{1}{2}\sqrt{2}\begin{pmatrix} -1 \\ -1 \end{pmatrix}$ ,  $\mathbf{q}^2(0) = \frac{1}{2}\sqrt{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ,  $\mathbf{q}^3(0) = \frac{1}{2}\sqrt{2}\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ ,  $\mathbf{p}(0) = \mathbf{0}$ ,  
Endzeitpunkt  $T = 100$

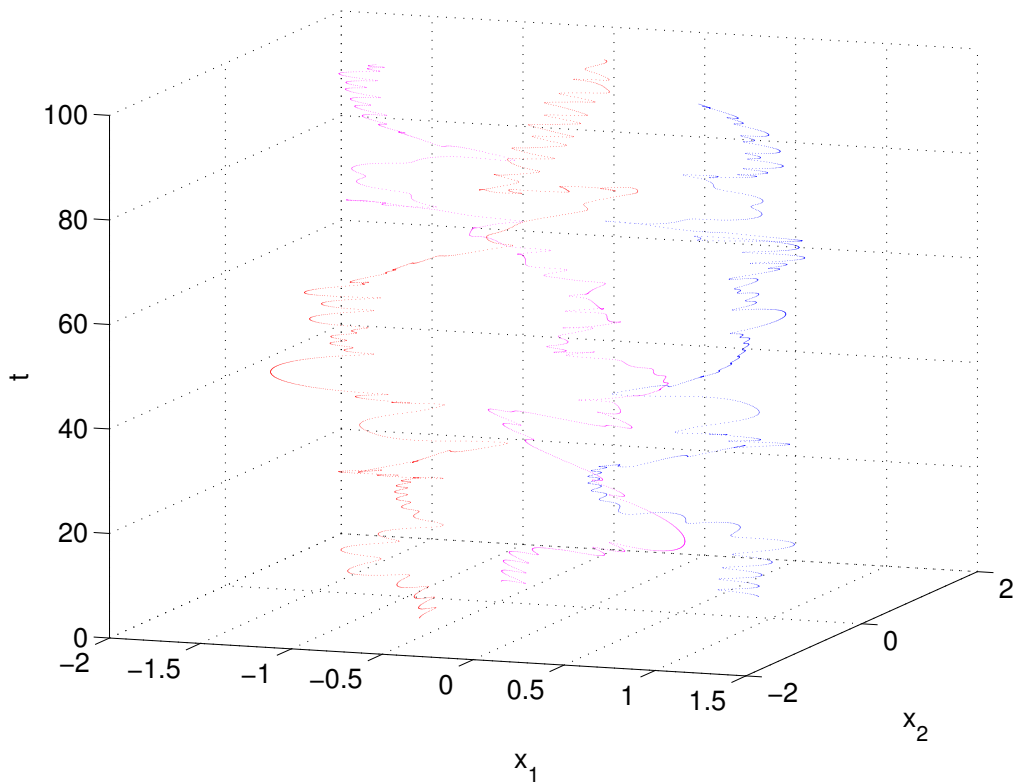
Trajektorien der Atome, Verlet, 10000 timesteps



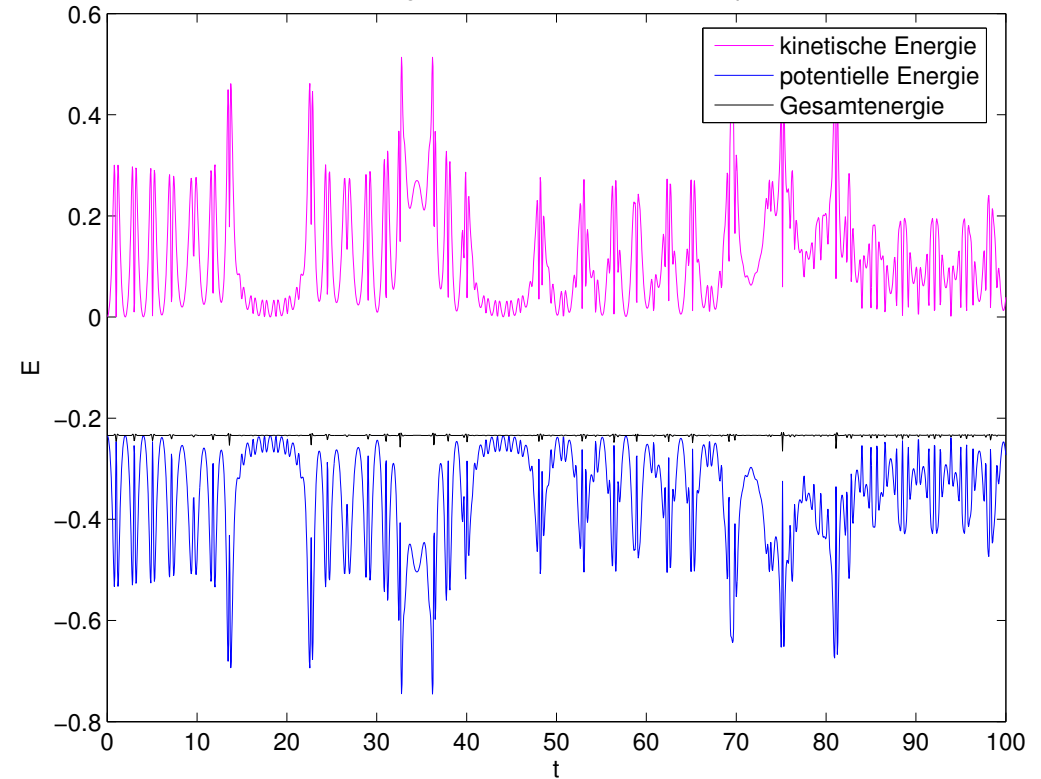
Energieanteile, Verlet, 10000 timesteps



Trajektorien der Atome, Verlet, 2000 timesteps



Energieanteile, Verlet, 2000 timesteps



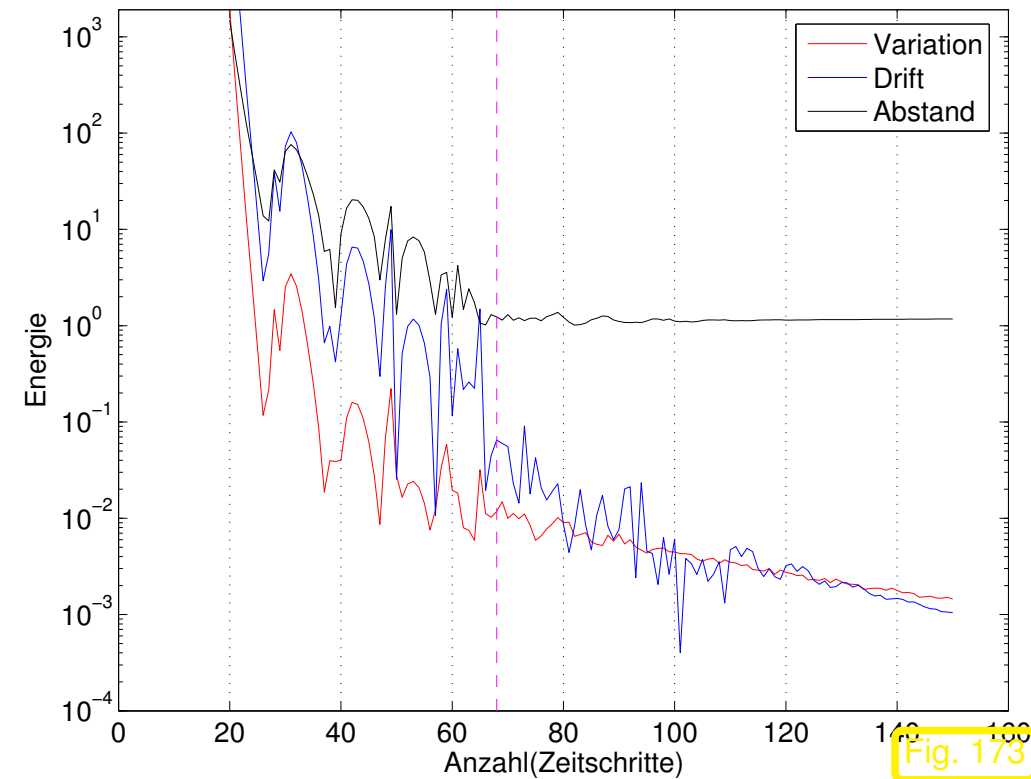
R. Hiptmair

rev 35327,  
25. April  
2011

## Beobachtungen:

- Völlig unterschiedliche Trajektorien bei Langzeitsimulation mit unterschiedlichen Zeitschrittweiten  $h$ .
- Qualitativ richtige Trajektorien in jedem Fall.

Verlet auf [0,10]: Schwankung der Gesamtenergie



$$T = 10, d = 2, n = 3, \mathbf{q}^1(0) = \frac{1}{2}\sqrt{2}\begin{pmatrix} -1 \\ -1 \end{pmatrix}, \mathbf{q}^2(0) = \frac{1}{2}\sqrt{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{q}^3(0) = \frac{1}{2}\sqrt{2}\begin{pmatrix} -1 \\ 1 \end{pmatrix}, \mathbf{p}(0) = 0.$$

$$\text{Variation} = \sum_{i=1}^{N-1} |E_{\text{tot}}((i+1)h) - E_{\text{tot}}(ih)|,$$

$$\text{Drift} = |E_{\text{tot}}(T) - E_{\text{tot}}(0)|,$$

$$\text{Abstand} = \max\left\{\left\|\mathbf{q}_h^j(T)\right\|_2, j = 1, 2, 3\right\}.$$

R. Hiptmair

◇  
rev 35327,  
25. April  
2011

Beispiel 4.4.39 (Vielteilchen-Molekulardynamik). → [26, Sect. 4.5.1]

2D konservatives Vielteilchensystem mit  
Lennard-Jones-Potential  $\rightarrow$  Bsp. 4.4.37

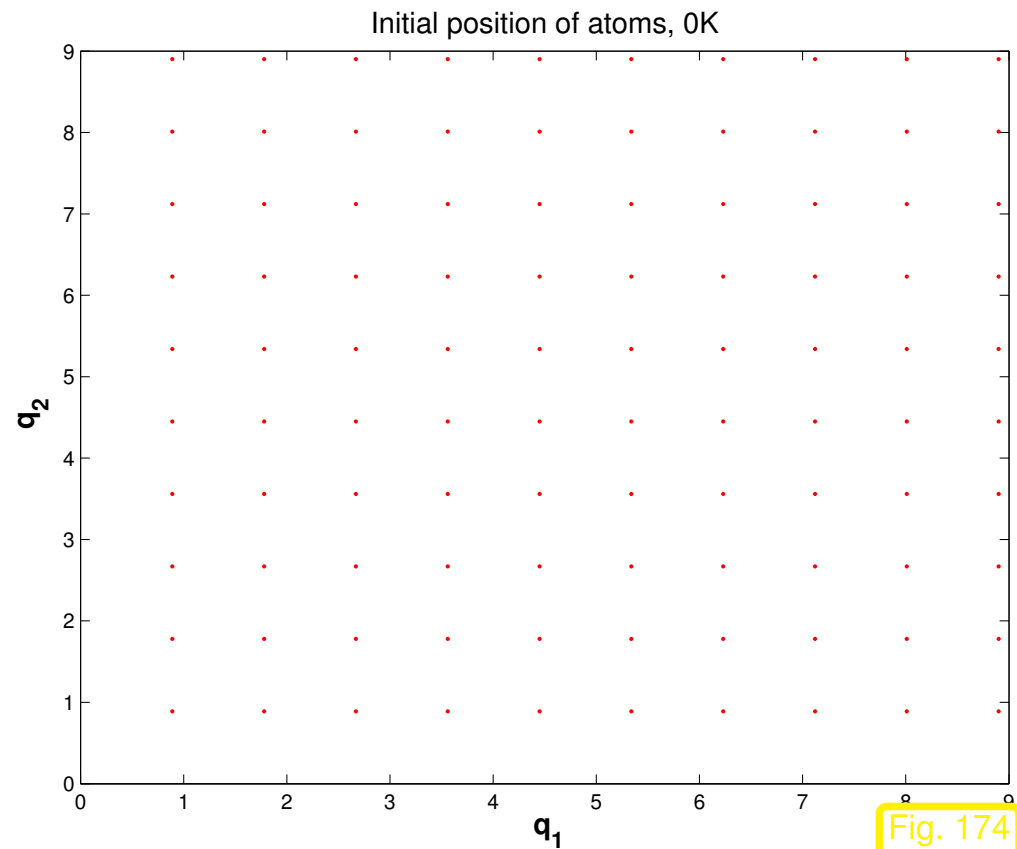
Anfangspositionen  $\triangleright$

(Anfangsimpulse = 0  $\leftrightarrow$  0K)

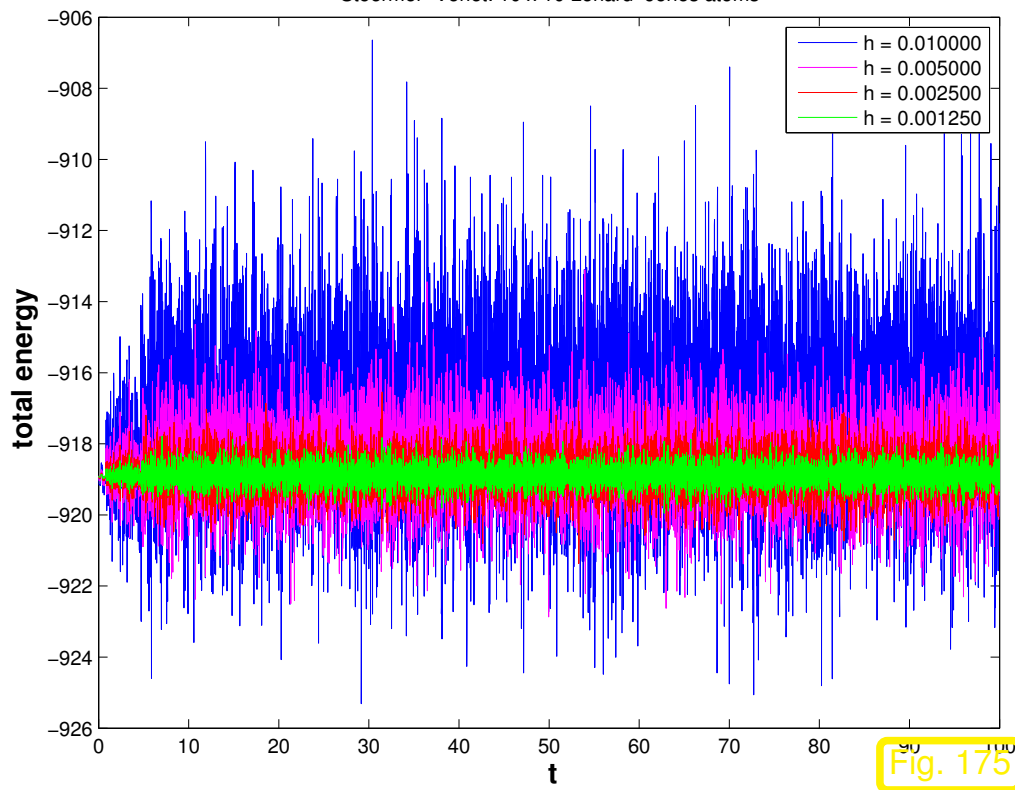
Beobachtet für explizite Trapezregel (2.3.3),  
Störmer-Verlet (4.4.22)

- Approximation der Gesamtenergie  $H(\mathbf{p}, \mathbf{q})$
- Mittlere kinetische Energie (“Temperatur”)

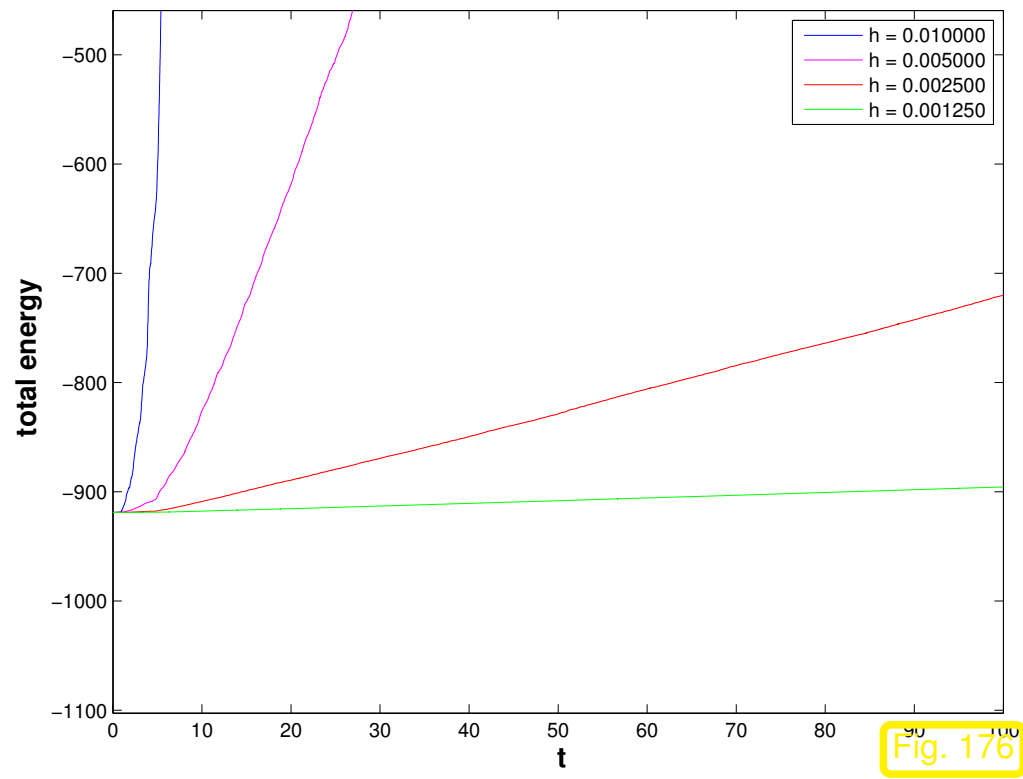
Animation  $\triangleright$

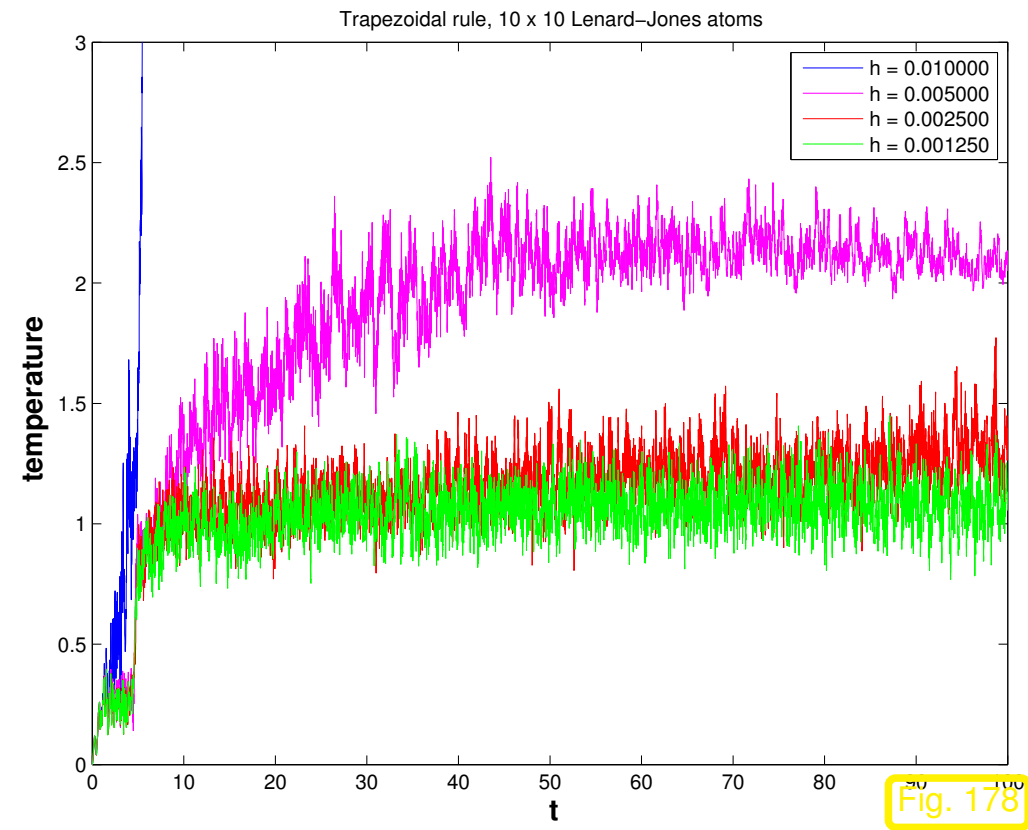
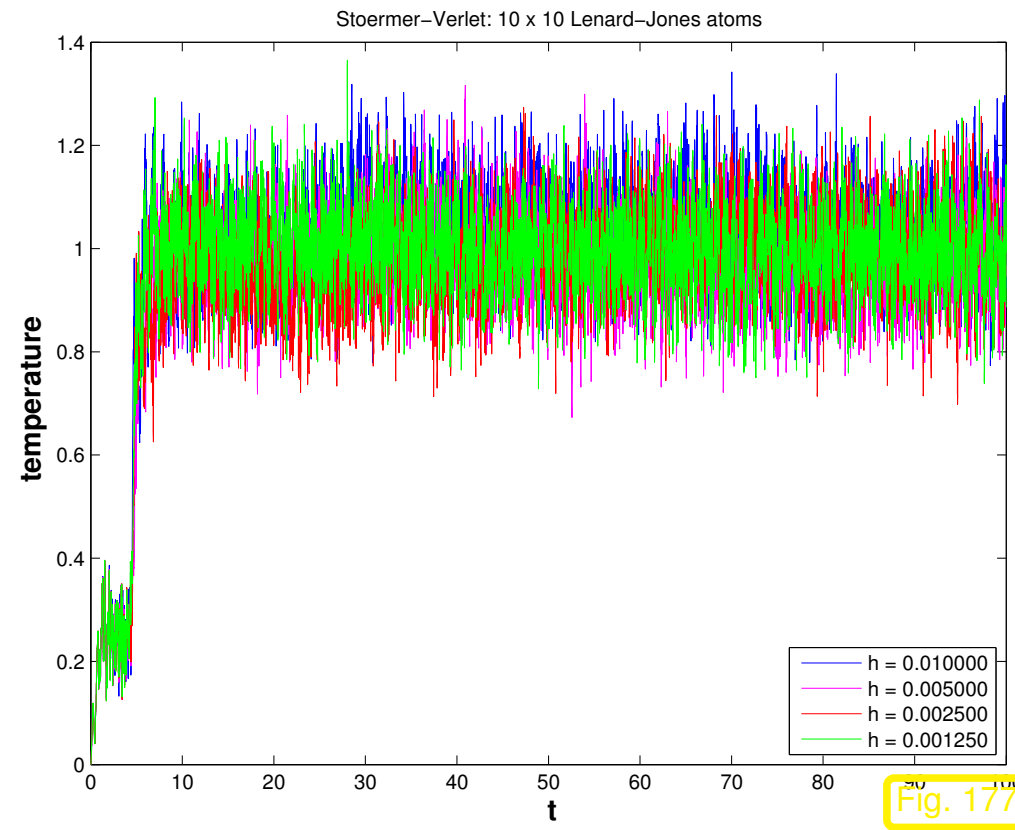


Stoermer-Verlet: 10 x 10 Lenard-Jones atoms



Trapezoidal rule, 10 x 10 Lenard-Jones atoms





Symplektischer Integrator: Qualitativ korrektes Verhalten der Temperatur



Beispiel 4.4.40 (Projektion auf Energiemannigfaltigkeit). → Bsp. 4.4.35



Idee: Korrektur der Energiedrift (bei nichtsymplektischen Integratoren) durch *Projektion* auf **Energie-**  
**mannigfaltigkeit**

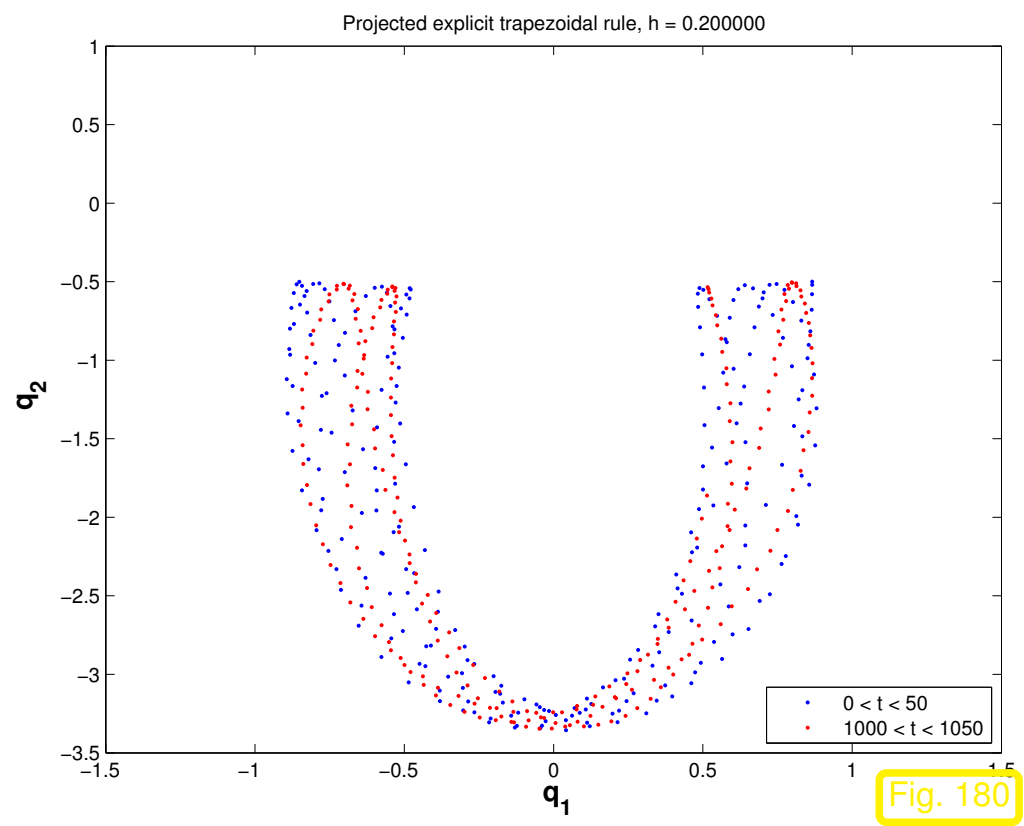
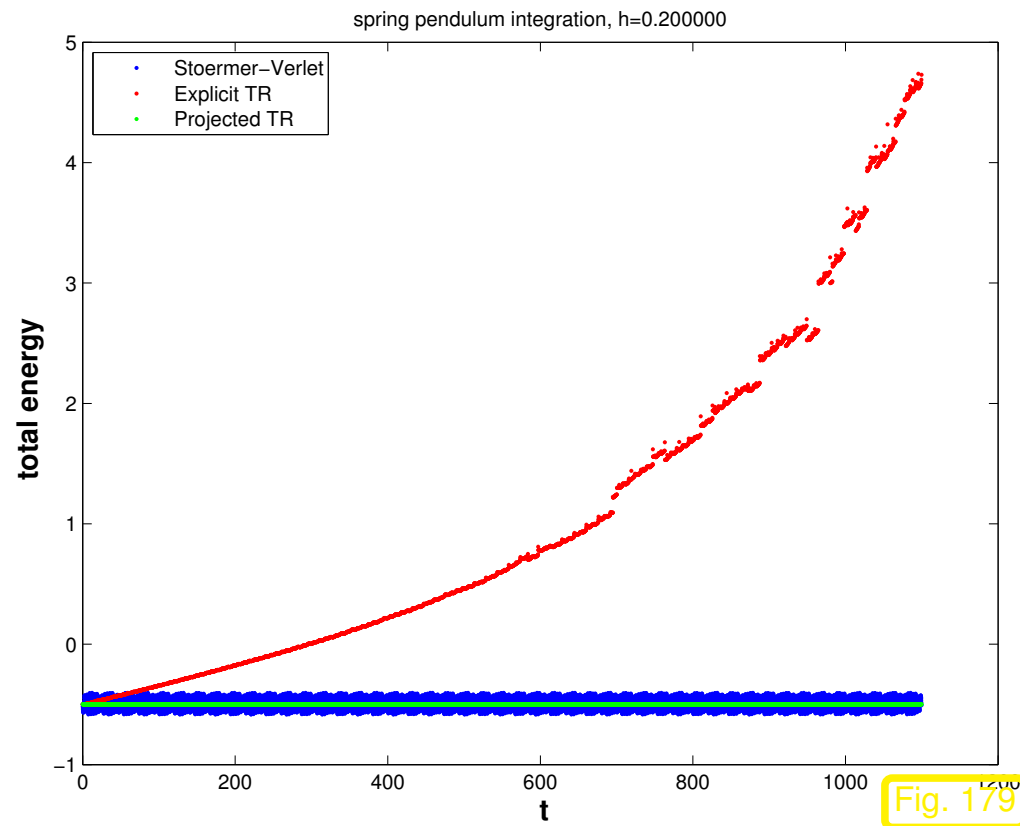
$$\{(\mathbf{p}, \mathbf{q}) \in \mathbb{R}^n \times \mathbb{R}^n : H(\mathbf{p}, \mathbf{q}) = H(\mathbf{p}_0, \mathbf{q}_0)\} . \quad (4.4.41)$$

Konkret: Orthogonalprojektion  $(\mathbf{p}, \mathbf{q}) \mapsto \mathbf{P}(\mathbf{p}, \mathbf{q}) := (\mathbf{p}^*, \mathbf{q}^*)$ : mit  $\mathbf{y} = \begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix}$ , bestimme  $\lambda \in \mathbb{R}$ ,  $\mathbf{y}^* = \begin{pmatrix} \mathbf{p}^* \\ \mathbf{q}^* \end{pmatrix} \in \mathbb{R}^{2n}$  so, dass

$$H(\mathbf{y}^*) = H_0 \quad , \quad \mathbf{y}^* = \mathbf{y} + \lambda \mathbf{grad} H(\mathbf{y}^*) . \quad (4.4.42)$$

Projiziertes ESV  $\Psi^h$ : Orthogonalprojektion nach jedem Schritt:  $\mathbf{y}_{k+1} = \mathbf{P}\Psi^h \mathbf{y}_k$

Beachte: (4.4.42) nichtlineares Gleichungssystem der Dimension  $2n + 1$ , **teuer** !



Warnung: Projektion kein Allheilmittel, siehe [16, Ch. IV, Ex. 4.3]

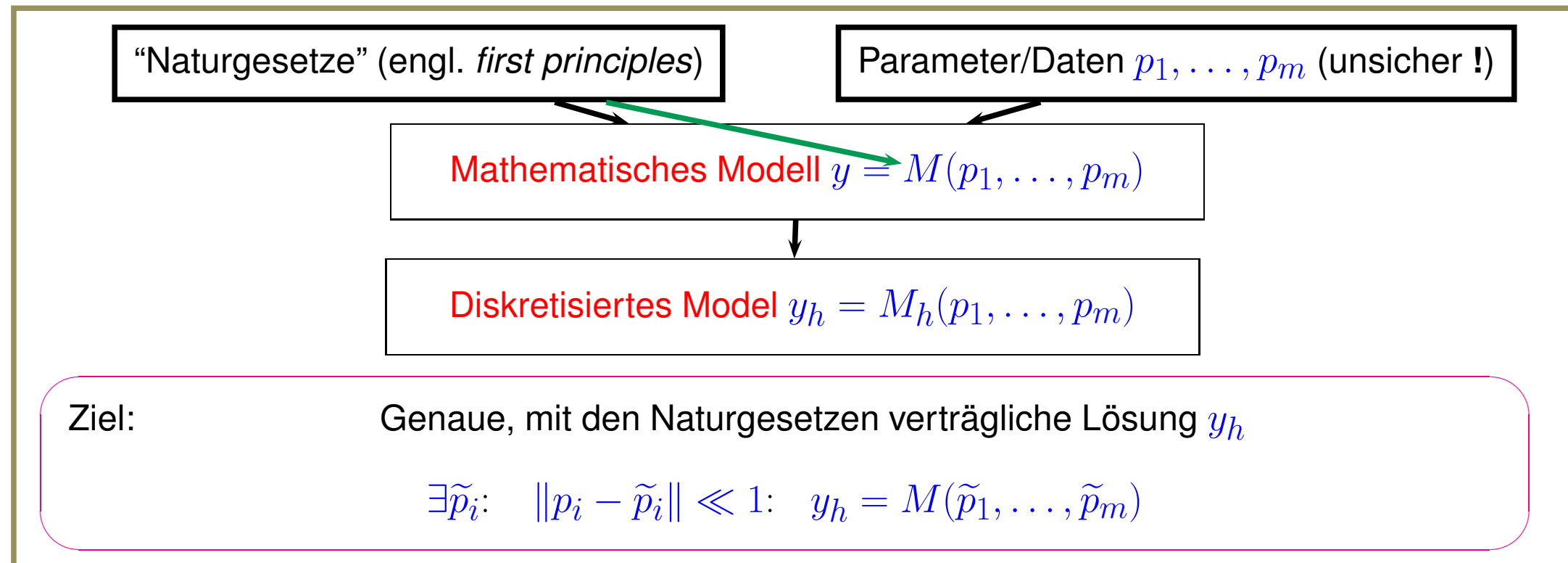


## 4.4.3 Rückwärtsanalyse

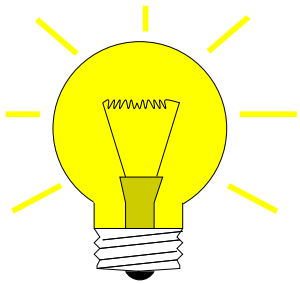
Sect. 1.3.3.5: Berechnung individueller Trajektorien sinnlos für schlecht konditionierte/chaotische Evolutionen.

Ziel: Berechnung typischer/wahrscheinlicher Trajektorien

*Bemerkung 4.4.43* (Rückwärtsanalyse (engl. *backward error analysis*): Philosophische Grundlage).



Konkrete Anwendung dieser Philosophie auf numerische Integratoren (Einschrittverfahren), siehe [26, Sect. 5.1]:



$\Psi^h \hat{=}$  diskrete Evolution eines ESV für  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y}) \quad \triangleright \quad \mathbf{y}_{k+1} = \Psi^h(\mathbf{y}_k)$   
Finde  $h$ -abhängiges Vektorfeld  $\tilde{\mathbf{f}}_h : \mathbb{R}^d \rightarrow \mathbb{R}^d$  so, dass

$$\dot{\tilde{\mathbf{y}}} = \tilde{\mathbf{f}}_h(\tilde{\mathbf{y}}), \quad \tilde{\mathbf{y}}(0) = \mathbf{y}_0 \quad \Rightarrow \quad \mathbf{y}_k = \tilde{\mathbf{y}}(hk). \quad (4.4.44)$$

Modifizierte Differentialgleichung

- $\tilde{\mathbf{f}}_h, \mathbf{f}$  gehören zur gleichen Klasse von Vektorfeldern, vgl. Bem. 4.4.14
- Kleine Störung:  $\tilde{\mathbf{f}}_h \approx \mathbf{f}$  für "kleine" Schrittweiten  $h > 0$

$\Psi^h$  strukturerhaltend & "qualitativ genau":  $(\mathbf{y}_k)$  akzeptabel

Wunsch:

Rückwärtsanalyse von auf der Grundlage modifizierter Differentialgleichung  
erfordert **uniforme Zeitschrittweite**

**Beispiel 4.4.45** (Modifizierte Gleichung für RK-ESV und lineare ODE).

- lineare ODE ( $\rightarrow$  Sect. 1.3.2):  $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y}$ ,  $\mathbf{A} \in \mathbb{R}^{d,d}$
- Runge-Kutta-Einschrittverfahren ( $\rightarrow$  Def. 2.3.5) mit Stabilitätsfunktion  $S(z)$

► Modifizierte ODE:  $\tilde{\mathbf{f}}_h(\mathbf{y}) = \tilde{\mathbf{A}}\mathbf{y}$  ,  $\tilde{\mathbf{A}} = \frac{1}{h} \log(S(h\mathbf{A}))$  , (4.4.46)

für “hinreichend kleines”  $h > 0$ .

Hier:  $\log \hat{=}$  “Matrixlogarithmus”:  $\log(\mathbf{X}) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} (\mathbf{X} - \mathbf{I})^k$  für  $\|\mathbf{X} - \mathbf{I}\| < 1$

Beweis von (4.4.46) (elementar unter Annahme, dass  $\mathbf{A}$  diagonalisierbar:  $\exists \mathbf{T} \in \mathbb{R}^{d,d}$  regulär:  $\mathbf{T}^{-1}\mathbf{A}\mathbf{T} = \mathbf{D} = \text{diag}(\mu_1, \dots, \mu_d)$ ):

Bem. 3.1.13, (3.1.16)  $\Rightarrow$

Für RK-ESV  $\mathbf{y}_1 = S(h\mathbf{A})\mathbf{y}_0$

$\xRightarrow{(4.4.44)}$   $\exp(\tilde{\mathbf{A}}h) = S(h\mathbf{A})$  mit  $\sigma(h\mathbf{A}) \cap ]-\infty, 0] = \emptyset$  für kleines  $h > 0$ .



? Modifizierte Gleichung im allgemeinen Fall

**Definition 4.4.47** (Modifizierte Gleichung der Ordnung  $q$ ).

Sei  $\Psi^h$  die diskrete Evolution eines Einschrittverfahrens der Konsistenzordnung  $p$  für die ODE  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$  mit lokal Lipschitz-stetigem  $\mathbf{f} : D \subset \mathbb{R}^d \mapsto \mathbb{R}^d$ .

Dann ist  $\tilde{\mathbf{y}} = \tilde{\mathbf{f}}_h(\tilde{\mathbf{y}})$  mit  $h$ -abhängigem, lokal Lipschitz-stetigen  $\tilde{\mathbf{f}}_h : D \mapsto \mathbb{R}^d$  eine **modifizierte Gleichung der Ordnung  $q$** ,  $q > p$ , wenn

$$\left\| \tilde{\Phi}_h^h \mathbf{y} - \Psi^h \mathbf{y} \right\| \leq C(\mathbf{y}) h^{q+1} \quad \forall \mathbf{y} \in D \quad \text{für } h \rightarrow 0,$$

wobei  $\tilde{\Phi}_h^t$  der Evolutionsoperator zu  $\tilde{\mathbf{y}} = \tilde{\mathbf{f}}_h(\tilde{\mathbf{y}})$  und  $C : D \mapsto \mathbb{R}$  lokal gleichmässig beschränkt.

Def. 4.4.47  $\hat{=}$  "Das ESV ist konsistent von der Ordnung  $q$  mit  $\tilde{\mathbf{y}} = \tilde{\mathbf{f}}_h(\tilde{\mathbf{y}})$ ."  $\rightarrow$  Def. 2.1.13

*Beispiel* 4.4.48 (Modifizierte Gleichung der Ordnung 2 zu explizitem Euler-Verfahren).

Explizites Eulerverfahren (1.4.2) für  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ :  $\mathbf{y}_1 = \mathbf{y}_1(h) = \Psi^h \mathbf{y}_0 = \mathbf{y}_0 + h\mathbf{f}(\mathbf{y}_0)$

Vergleich mit Taylorentwicklung (um 0) (2.3.25) der exakten Lösung  $\mathbf{y}(t)$ :

$$\mathbf{y}(h) = \mathbf{y}_0 + \mathbf{f}(\mathbf{y}_0)h + \frac{1}{2}D\mathbf{f}(\mathbf{y}_0)\mathbf{f}(\mathbf{y}_0)h^2 + O(h^3) \quad \text{für } h \rightarrow 0. \quad (4.4.49)$$

“störender Term”  zu “verschieben” in  $\tilde{\mathbf{f}}_h$

Modifizierte Gleichung der Ordnung 2:

$$\dot{\tilde{\mathbf{y}}} = \tilde{\mathbf{f}}_h(\tilde{\mathbf{y}}) := \mathbf{f}(\tilde{\mathbf{y}}) - \frac{1}{2}hD\mathbf{f}(\tilde{\mathbf{y}})\mathbf{f}(\tilde{\mathbf{y}}) \quad \xrightarrow{\mathbf{f} \text{ glatt}} \quad \tilde{\Phi}_h^h \mathbf{y}_0 - \mathbf{y}_1(h) = O(h^3),$$

denn aus (4.4.49), für  $h \rightarrow 0$

$$\begin{aligned} \tilde{\mathbf{y}}(h) &= \mathbf{y}_0 + \tilde{\mathbf{f}}_h(\mathbf{y}_0)h + \frac{1}{2}D\tilde{\mathbf{f}}_h(\mathbf{y}_0)\tilde{\mathbf{f}}_h(\mathbf{y}_0)h^2 + O(h^3) \\ &= \mathbf{y}_0 + h\mathbf{f}(\mathbf{y}_0) - \frac{1}{2}h^2 \frac{1}{2}D\mathbf{f}(\mathbf{y}_0)\mathbf{f}(\mathbf{y}_0) + \frac{1}{2}D\mathbf{f}(\mathbf{y}_0)\mathbf{f}(\mathbf{y}_0)h^2 + O(h^3) \\ &= \mathbf{y}_0 + h\mathbf{f}(\mathbf{y}_0) + O(h^3) = \Psi^h \mathbf{y}_0 + O(h^3). \end{aligned}$$

Durchwegs “stillschweigende Annahme”:  $\mathbf{f}$  “hinreichend glatt”  $\Rightarrow \Phi^t, \Psi^h$  “hinreichend glatt”

✎ Notationen:  $\Phi^t \hat{=}$  Evolutionsoperator zur ODE  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ ,  
 $t \mapsto \mathbf{y}(t) \hat{=}$  Lösungstrajektorien von  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$  zum Anfangswert  $\mathbf{y}_0 \in D$ .

Ziel: Formalisierung der ad-hoc-Konstruktion einer modifizierten Gleichung der Ordnung  $p + 1$  aus  
 Bsp. 4.4.48

Idee: **Rekursive Konstruktion** von  $\tilde{\mathbf{f}}_h$ :

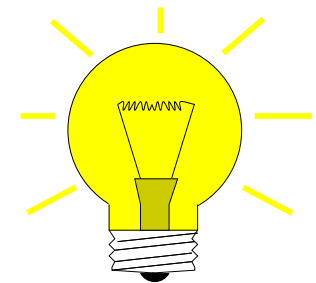
Annahme: diskrete Evolution  $\Psi^h$  konsistent von der Ordnung  $p$  mit  $\dot{\mathbf{y}} = \mathbf{f}_h(\mathbf{y})$

Ansatz:

$$\tilde{\mathbf{f}}_h = \mathbf{f}_h(\mathbf{y}) + h^p \Delta \mathbf{f}(\mathbf{y}) \quad (4.4.50)$$

Modifikatorfunktion

Ziel:  $\tilde{\Phi}_h^h \mathbf{y} - \Psi^h \mathbf{y} = O(h^{p+2})$  für  $h \rightarrow 0$  (4.4.51)





$$\tau(\mathbf{y}_0, h) := \Phi_h^h \mathbf{y}_0 - \Psi^h \mathbf{y}_0 = \mathbf{d}(\mathbf{y}_0) h^{p+1} + O(h^{p+2}) \quad \text{für } h \rightarrow 0, \quad \forall \mathbf{y}_0 \in D. \quad (4.4.52)$$

Konsistenzfehler  $\rightarrow$  Def. 2.1.11,  $(\Phi_h^t \hat{=} \text{Evolutionoperator zu } \dot{\mathbf{y}} = \mathbf{f}_h(\mathbf{y}))$

(4.4.51): Bestimme  $\Delta \mathbf{f}$  so, dass Lsg. von  $\dot{\tilde{\mathbf{y}}} = \tilde{\mathbf{f}}_h(\tilde{\mathbf{y}}), \tilde{\mathbf{y}}(0) = \mathbf{y}_0$

$$\tilde{\Phi}_h^h \mathbf{y} - \Psi^h \mathbf{y} = \tilde{\mathbf{y}}(h) - \mathbf{y}_1 = O(h^{p+2}) \quad \text{für } h \rightarrow 0. \quad (4.4.53)$$

Taylorentwicklung um  $h = 0$ , vgl. (2.3.25), benutze Dgl. und Kettenregel: Für  $h \rightarrow 0$

$$\begin{aligned} \tilde{\mathbf{y}}(h) &= \mathbf{y}_0 + \sum_{k=1}^{p+1} \frac{h^k}{k!} \tilde{\mathbf{y}}^{(k)}(0) + O(h^{p+2}) = \mathbf{y}_0 + \sum_{k=1}^{p+1} \frac{h^k}{k!} \frac{d^{k-1}}{dt^{k-1}} \tilde{\mathbf{f}}_h(\tilde{\mathbf{y}}(t)) \Big|_{t=0} + O(h^{p+2}) \\ &= \mathbf{y}_0 + h \tilde{\mathbf{f}}_h(\mathbf{y}_0) + \frac{1}{2} h^2 D \tilde{\mathbf{f}}_h(\mathbf{y}_0) \tilde{\mathbf{f}}_h(\mathbf{y}_0) \\ &\quad + \frac{1}{6} h^3 (D^2 \tilde{\mathbf{f}}_h(\mathbf{y}_0) (\tilde{\mathbf{f}}_h(\mathbf{y}_0), \tilde{\mathbf{f}}_h(\mathbf{y}_0)) + D \tilde{\mathbf{f}}_h(\mathbf{y}_0) D \tilde{\mathbf{f}}_h(\mathbf{y}_0) \tilde{\mathbf{f}}_h(\mathbf{y}_0)) + \dots + O(h^{p+2}) \\ &= \mathbf{y}_0 + h \mathbf{f}_h(\mathbf{y}_0) + h^{p+1} \Delta \mathbf{f}(\mathbf{y}_0) + \frac{1}{2} h^2 D \mathbf{f}_h(\mathbf{y}_0) \mathbf{f}_h(\mathbf{y}_0) \\ &\quad + \frac{1}{6} h^3 (D^2 \mathbf{f}_h(\mathbf{y}_0) (\mathbf{f}_h(\mathbf{y}_0), \mathbf{f}_h(\mathbf{y}_0)) + D \mathbf{f}_h(\mathbf{y}_0) D \mathbf{f}_h(\mathbf{y}_0) \mathbf{f}_h(\mathbf{y}_0)) + \dots + O(h^{p+2}), \end{aligned}$$

da " $O(h^p)$ -Modifikation" in (4.4.50), z.B.

$$\begin{aligned} h^2 D \tilde{\mathbf{f}}_h(\mathbf{y}_0) \tilde{\mathbf{f}}_h(\mathbf{y}_0) &= h^2 (D \mathbf{f}_h(\mathbf{y}_0) + h^p D \Delta \mathbf{f}(\mathbf{y}_0)) (\mathbf{f}_h(\mathbf{y}_0) + h^p \Delta \mathbf{f}(\mathbf{y}_0)) \\ &= h^2 D \mathbf{f}_h(\mathbf{y}_0) \mathbf{f}_h(\mathbf{y}_0) + O(h^{p+2}). \end{aligned}$$

➤ Beobachtung: Taylorentwicklung von  $t \mapsto \Phi_h^t V y_0$  um  $t = 0$  ist enthalten !

$$\blacktriangleright \tilde{\mathbf{y}}(h) = \Phi_h^h \mathbf{y}_0 + h^{p+1} \Delta \mathbf{f}(\mathbf{y}_0) + O(h^{p+2})$$

$$\stackrel{(4.4.52)}{=} \Psi^h \mathbf{y}_0 + h^{p+1} \mathbf{d}(\mathbf{y}_0) + h^{p+1} \Delta \mathbf{f}(\mathbf{y}_0) + O(h^{p+2}) .$$

$$(4.4.53) \text{ erfüllt durch } \boxed{\Delta f(\mathbf{y}) := -\mathbf{d}(\mathbf{y})} ! \quad (4.4.54)$$

Versuch: **Reihenansatz** für Vektorfeld der modifizierten Gleichung:

$$\tilde{\mathbf{f}}_h(\mathbf{y}) = \mathbf{f}(\mathbf{y}) + h^p \Delta \mathbf{f}_p(\mathbf{y}) + h^{p+1} \Delta \mathbf{f}_{p+1}(\mathbf{y}) + h^{p+2} \Delta \mathbf{f}_{p+2}(\mathbf{y}) + \dots \quad (4.4.55)$$

➤ Modifikatorfunktionen  $\Delta \mathbf{f}_\ell, \ell \in \mathbb{N}$ , aus rekursiver Konstruktionsvorschrift

$$(4.4.54) \Rightarrow \Delta \mathbf{f}_\ell(\mathbf{y}) = - \lim_{h \rightarrow 0} \frac{\tilde{\Phi}_{h, \ell-1}^h \mathbf{y} - \Psi^h \mathbf{y}}{h^{\ell+1}}, \quad (4.4.56)$$

mit  $\tilde{\Phi}_{h, \ell}^t \hat{=}$  Evolutionsoperator zur ODE

$$\dot{\tilde{\mathbf{y}}} = \tilde{\mathbf{f}}_{h, \ell}(\tilde{\mathbf{y}}) := \mathbf{f}(\mathbf{y}) + h^p \Delta \mathbf{f}_p(\mathbf{y}) + h^{p+1} \Delta \mathbf{f}_{p+1}(\mathbf{y}) + h^{p+2} \Delta \mathbf{f}_{p+2}(\mathbf{y}) + \dots + h^\ell \Delta \mathbf{f}_\ell(\mathbf{y}) . \quad (4.4.57)$$

**Bemerkung 4.4.58** (Berechnung der Modifikatorfunktionen  $\Delta f_j$  durch Computeralgebra).

MAPLE-Code: Berechnung der  $\Delta f_j$

```

fcn := y -> f(y) :
N := q :
fcoe [1] := fcn(y) :
for n from 2 by 1 to N do
  modeq := sum(h^j*fcoe [j+1], j=0..n-2) :
  diffy [0] := y :
  for i from 1 by 1 to n do
    diffy [i] := diff(diffy[i-1],y)+modeq :
  od :
  ytilde := sum(h^k*diffy[k]/k!, k=0..n) :
  res := ytilde-y-h*fcn(y) :
  tay := convert(series(res,h=0,n+1),polynom) :
  fcoe [n] := -coeff(tay,h,n) :
od :
simplify(sum(h^j*fcoe[j+1], j=0..N-1)) ;

```

ESV:

Explizites Euler-Verfahren

◁ MAPLE-code [13]:

Berechnung der  
Modifikatorfunktionen  $\Delta f_\ell$   
für skalare ODE  $\dot{y} = f(y)$ .

Ausgabe der Reihe (4.4.55)  
bis zum  $q$ . Term.



**Beispiel 4.4.59** (Modifikatoren für einfache ESV).

Skalare Differentialgleichung:  $\dot{y} = y^2$  → Bsp. 1.3.11

- Explizites Euler-Verfahren (1.4.2):  $y_1 = y_0 + hf(y_0)$

$$\begin{aligned} \tilde{f}(y) = & y^2 - h \underbrace{y^3}_{-\Delta f_1} + h^2 \underbrace{\frac{3}{2}y^4}_{\Delta f_2} - h^3 \underbrace{\frac{8}{3}y^5}_{-\Delta f_3} + h^4 \underbrace{\frac{31}{6}y^6}_{\Delta f_4} - h^5 \underbrace{\frac{157}{15}y^7}_{-\Delta f_5} \\ & + h^6 \underbrace{\frac{649}{30}y^8}_{\Delta f_6} - h^7 \underbrace{\frac{9427}{210}y^9}_{-\Delta f_7} + h^8 \underbrace{\frac{19423}{210}y^{10}}_{\Delta f_8} - h^9 \underbrace{\frac{6576}{35}y^{11}}_{-\Delta f_9} + O(h^{10}). \end{aligned}$$

- Implizites Euler-Verfahren (1.4.13):  $y_1 = y_0 + hf(y_1)$

(In MAPLE code: `res := ytilde-y-h*fcn(ytilde)`)

$$\begin{aligned} \tilde{f}(y) = & y^2 + hy^3 + \frac{3}{2}h^2y^4 + \frac{8}{3}h^3y^5 + \frac{31}{6}h^4y^6 + \frac{157}{15}h^5y^7 + \frac{649}{30}h^6y^8 \\ & + \frac{9427}{210}h^7y^9 + \frac{19423}{210}h^8y^{10} + \frac{6576}{35}h^9y^{11} + O(h^{10}) \end{aligned}$$

- Implizite Mittelpunktsregel (1.4.19):  $y_1 = y_0 + hf(\frac{1}{2}(y_0 + y_1))$

(In MAPLE code: `res := ytilde-y-h*fcn(0.5*(y+ytilde))`)

$$\tilde{f}(y) = y^2 + \frac{1}{4}h^2y^4 + \frac{1}{8}h^4y^6 + 0.057291667h^6y^8 + 0.02343750000h^8y^{10} + O(h^{10}).$$

- ☞ Nur *gerade* Potenzen von  $h$ , vgl. Beweis zu Thm. 2.1.29, Thm. 2.4.22





Problem: Potenzreihe (in  $h$ )  $\sum_{k=1}^{\infty} h^k \Delta f_k(\mathbf{y})$  möglicherweise divergent  
 $\forall h > 0$  ( $\leftrightarrow$  Konvergenzradius = 0)

Interpretation von (4.4.55) als **asymptotische Entwicklung** von  $\tilde{\mathbf{f}}_h$ , siehe Def. 2.4.7

*Beispiel 4.4.60* (Bedeutung der modifizierten Gleichungen niedriger Ordnung).

- Anfangswertproblem für logistische Differentialgleichung, siehe Bsp. 1.2.1

$$\dot{y} = \lambda y(1 - y) \quad , \quad y(0) = 0.01 \quad .$$

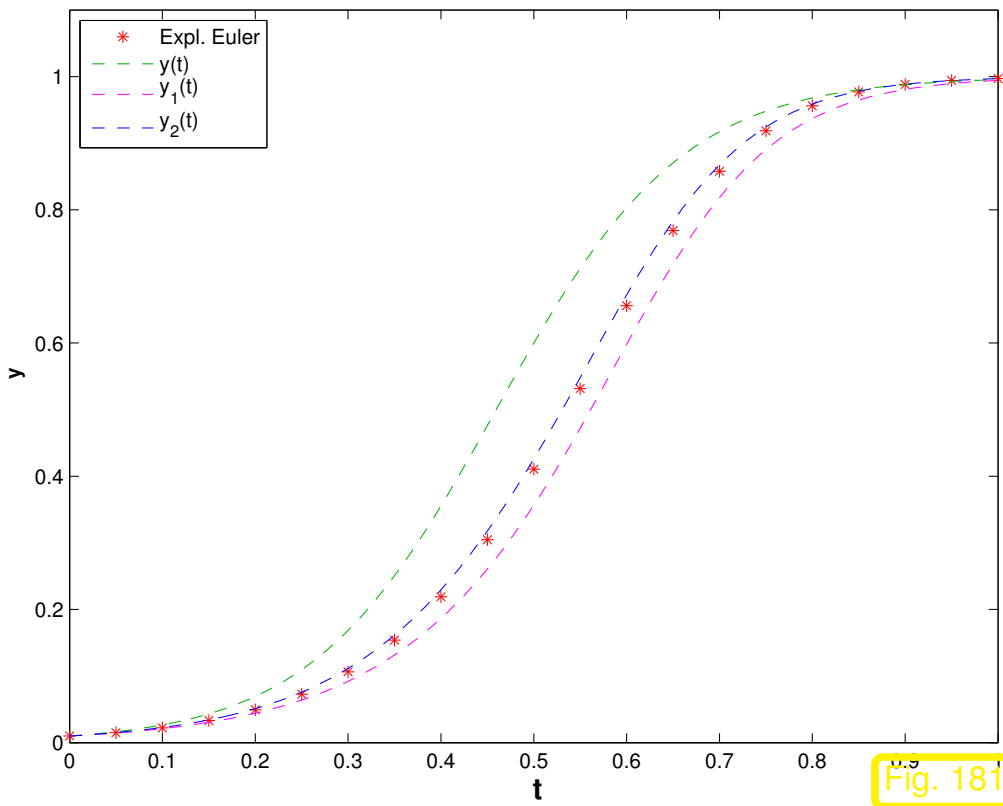
- ESV: Explizites Euler-Verfahren (1.4.2), vgl. Bsp. 1.4.9, Modifikatorfunktionen aus (4.4.55)

$$\Delta f_1(y) = \lambda^2 \left( -\frac{1}{2} y + \frac{3}{2} y^2 - y^3 \right) \quad , \quad \Delta f_2(y) = \lambda^3 \left( -\frac{11}{6} y^2 + 3 y^3 - \frac{3}{2} y^4 + \frac{1}{3} y \right) \quad .$$

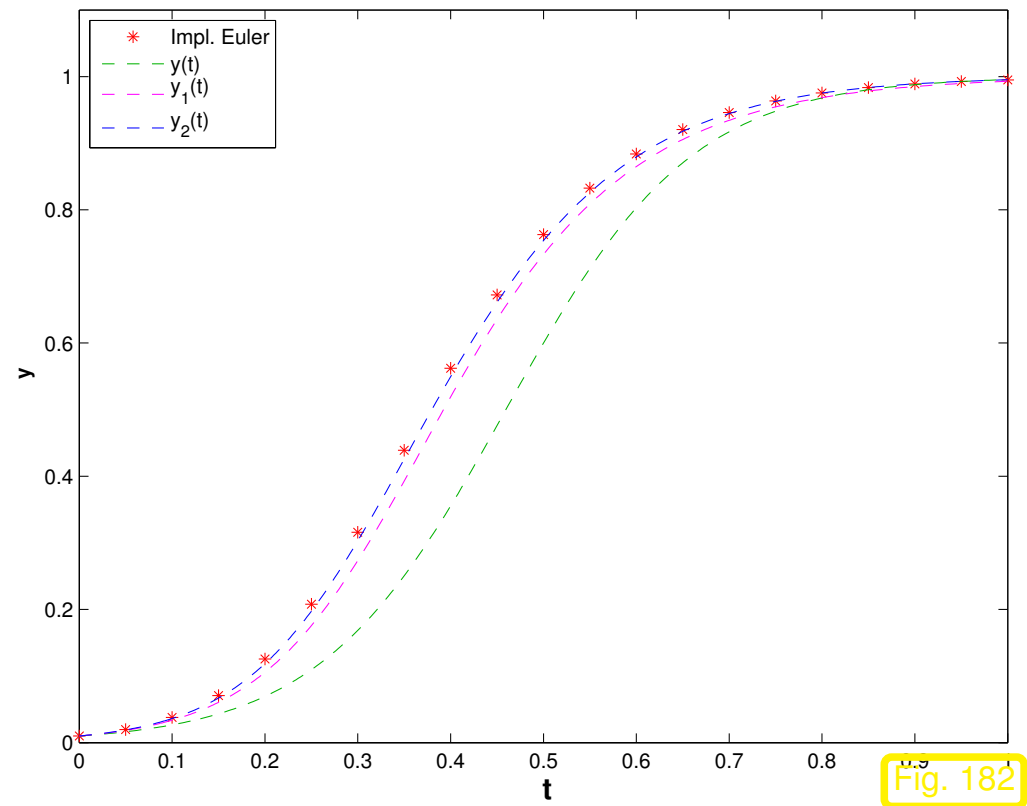
- ESV: Implizites Euler-Verfahren (1.4.13), vgl. Bsp. 1.4.15, Modifikatorfunktionen aus (4.4.55)

$$\Delta f_1(y) = \lambda^2 \left( \frac{1}{2} y - \frac{3}{2} y^2 + y^3 \right) \quad , \quad \Delta f_2(y) = \lambda^3 \left( -\frac{11}{6} y^2 + 3 y^3 - \frac{3}{2} y^4 + \frac{1}{3} y \right)$$

Explicit Euler h=0.050000, logistic ODE,  $\lambda=10.000000$



Implicit Euler h=0.050000, logistic ODE,  $\lambda=10.000000$



Die Euler-Verfahren für  $y' = f(y)$  liefern eine bessere Approximation für die Lösungen von

$$\dot{y} = \tilde{f}_{1,h}(y) = f(y) + h\Delta f_1(y) \quad \text{und} \quad \dot{y} = \tilde{f}_{2,h}(y) = f(y) + h\Delta f_1(y) + h^2\Delta f_2(y) .$$

▶ Betrachte abgeschnittene modifizierte Gleichung !

**Lemma 4.4.61** (“Abgeschnittene” modifizierte Gleichung).

Mit Modifikatorfunktionen  $\Delta \mathbf{f}_i$  gemäss (4.4.56) für die ODE  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$  und das ESV mit diskreter Evolution  $\Psi^h$  (der Konsistenzordnung  $p$ ) wie oben definiert, ist

$$\dot{\tilde{\mathbf{y}}} = \tilde{\mathbf{f}}_{h,\ell}(\tilde{\mathbf{y}}) := \mathbf{f}(\tilde{\mathbf{y}}) + h^p \Delta \mathbf{f}_p(\tilde{\mathbf{y}}) + h^{p+1} \Delta \mathbf{f}_{p+1}(\tilde{\mathbf{y}}) + \cdots + h^\ell \Delta \mathbf{f}_\ell(\tilde{\mathbf{y}}),$$

eine modifizierte Gleichung der Ordnung  $\ell + 1$ ,  $\ell > p$  ( $\rightarrow$  Def. 4.4.47)

*Beweis.* Der Beweis ergibt sich aus der rekursiven Konstruktion der  $\Delta \mathbf{f}_\ell$ , siehe (4.4.52), (4.4.53), (4.4.54).

## 4.4.4 Modifizierte Gleichungen: Fehleranalyse

Im Sinne der Rückwärtsanalyse ( $\rightarrow$  Bem. 4.4.43) des *Lanzzeitverhaltens* von Einschrittverfahren ist zu untersuchen:

- Gibt es eine (strukturerhaltende) modifizierte Gleichung, der Lösung für lange Zeiten nahe bei der numerische Lösung (Gitterfunktion  $(\mathbf{y}_k)_k$ ) bleibt.
- Wenn ja, ist die rechte Seite dieser modifizierten Gleichung nahe bei der rechten Seite der Ausgangsgleichung.

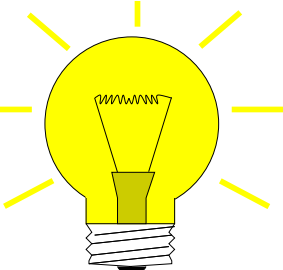
Strategie: Was wollen wir ?

☞ Lemma 4.4.61: *Familie* modifizierter Gleichungen  $\dot{\tilde{\mathbf{y}}} = \tilde{\mathbf{f}}_{h,\ell}(\tilde{\mathbf{y}})$ , „konsistent mit dem ESV“  
 $\mathbf{y}_{k+1} = \Psi^h \mathbf{y}_k$ , d.h.

Konsistenzfehler  $\tau(\mathbf{y}, h) := \Psi^h \mathbf{y} - \tilde{\Phi}_{h,\ell}^h \mathbf{y} \rightarrow 0$  für  $h \rightarrow 0$ .



Idee: Erinnerung an Beweis des *Konvergenzsatzes für ESV*, Thm. 2.1.19  
(vgl. auch Beweis von Thm. 2.1.26 und das „diskrete Gronwall-Lemma“ Lemma 2.1.20)



$$\blacktriangleright \quad \|\mathbf{y}_k - \tilde{\mathbf{y}}(kh)\| \leq \frac{1}{h} \max_{j=0, \dots, k-1} \|\boldsymbol{\tau}(\mathbf{y}_j, h)\| \frac{\exp(Lhk) - 1}{L}. \quad (4.4.62)$$

( $L > 0$ : Lipschitz-Konstante der Inkrementfunktion des ESV,  $\tilde{\mathbf{y}} \hat{=}$  Lösung der modifizierten Gleichung)

Exponentielles Wachstum der Konstanten in (4.4.62) für  $hk \rightarrow \infty$  !



( $\|\boldsymbol{\tau}(\mathbf{y}_j, h)\| = O(h^{\ell+2})$  liefert keine sinnvollen Abschätzungen bei *Langeitintegration*)

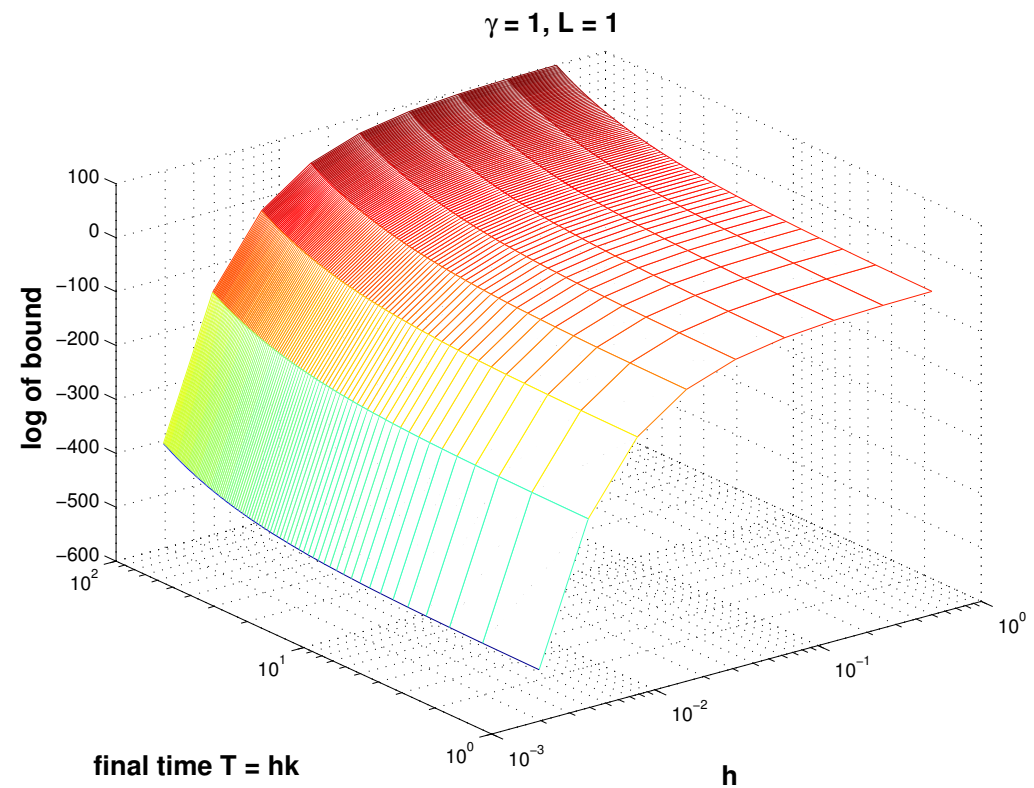
JEDOCH:

Wenn Konsistenzfehler „**exponentiell klein**“

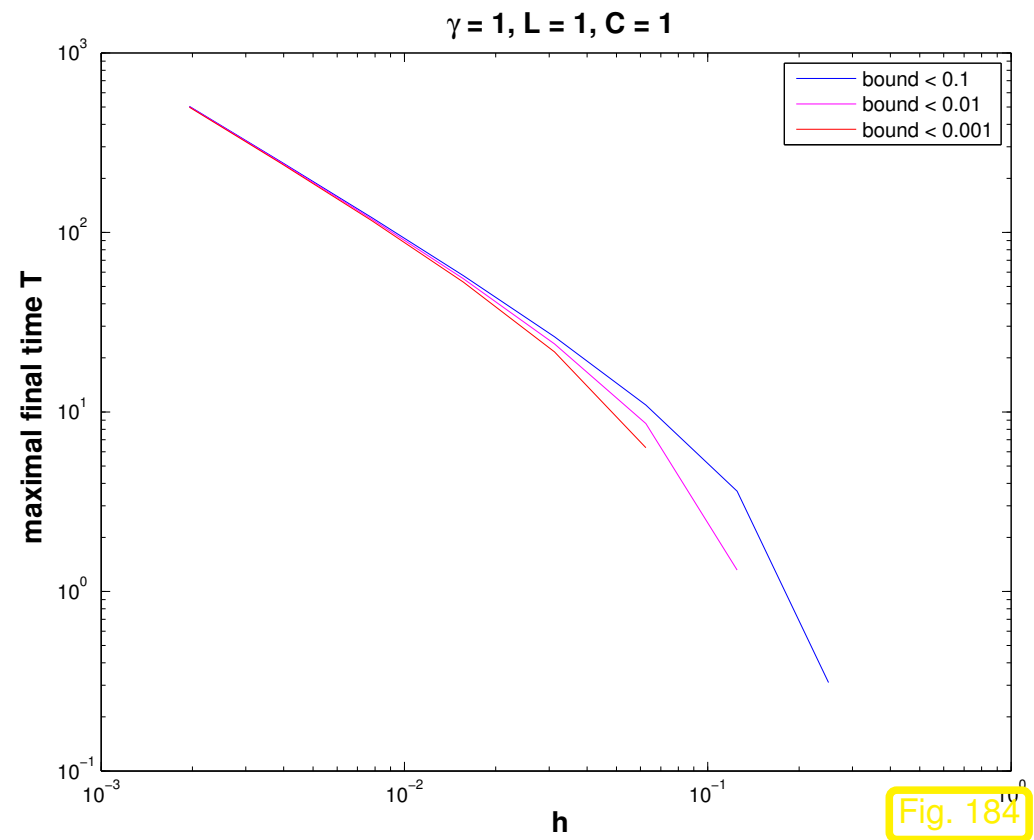
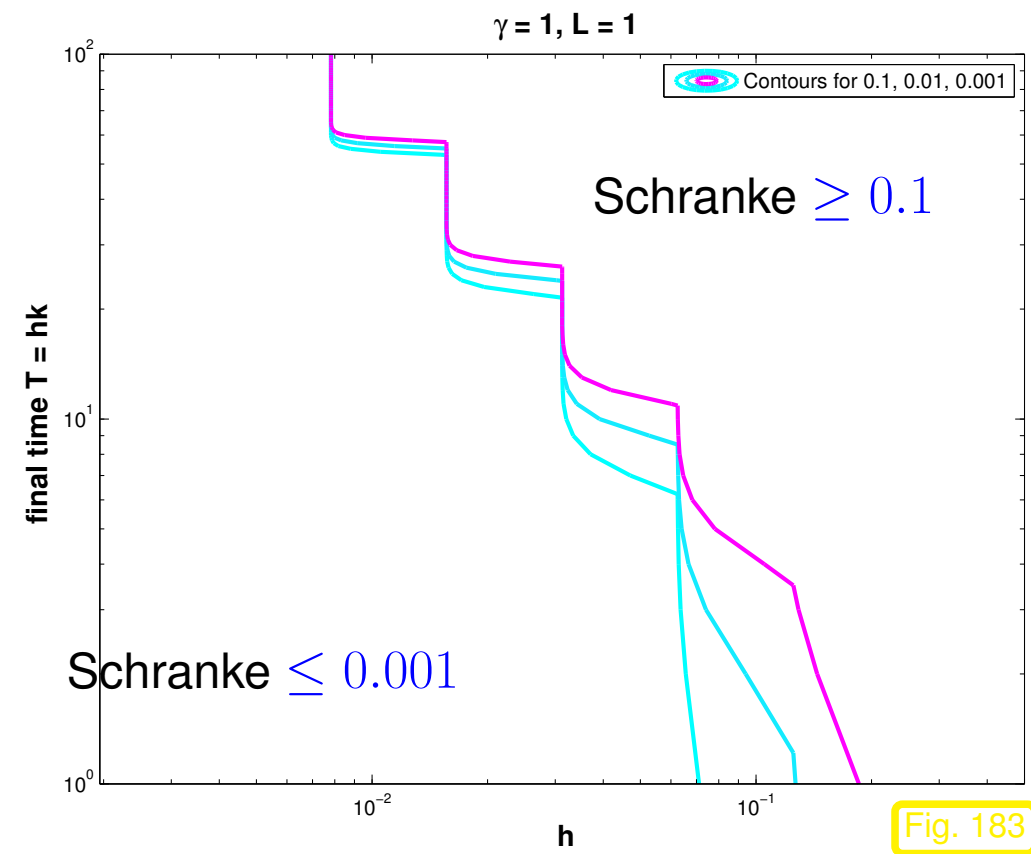
$$\|\tau\| \leq Ch \exp(-\gamma/h), \quad \gamma > 0 \quad (4.4.63)$$



$$\|y_k - \tilde{y}(kh)\| \leq C \exp(-\gamma/h + hkL) \quad (4.4.64)$$



Verhalten der Schranke aus (4.4.64)



Aus (4.4.64) lesen wir ab:

$$h < \frac{\gamma}{TL - \log(\tau/C)} \Rightarrow \|y_k - \tilde{y}(kh)\| \leq \tau \quad \text{für } 0 \leq kh \leq T.$$

Schrittweite  $h$  „klein“  $\Rightarrow$  Numerische Lösung  $y_k$  bleibt lange in der Nähe der Trajektorie  $t \mapsto \tilde{y}(t)$   
(Präzisere Diskussion in Bem. 4.4.85)

Wir sind frei in der Wahl der Abschneideindex  $\ell$  !

Frage: Was ist die beste abgeschnittene modifizierte Gleichung ?

Zur Beantwortung brauchen wir Konzepte/Hilfsmittel aus der Funktionentheorie !

### Analytizitätsvoraussetzung

für jedes Kompaktum  $K \subset D$  gibt es ein  $R = R(K) > 0$ , so dass  $\mathbf{f}(\mathbf{y})$  in jedem  $\mathbf{y} \in K$  in jeder Komponente von  $\mathbf{y}$  eine Potenzreihenentwicklung mit Konvergenzradius  $> R$  besitzt.

$\Leftrightarrow$   $\mathbf{f}$  ist holomorph in  $D$

Erklärung: Potenzreihenentwicklung um  $\mathbf{y} = (y_1, \dots, y_d)^T$  in der  $j$ . Komponente

$$\mathbf{f}(y_1, \dots, y_{j-1}, y, y_{j+1}, \dots, y_d) = \sum_{k=0}^{\infty} \mathbf{a}_k(\mathbf{y})(y - y_j)^k \quad \text{für } |y - y_j| < R .$$

**Beispiel 4.4.65** (Analytizitätsvoraussetzung für Hamiltonsche Differentialgleichungen).

$\mathbf{f}(\mathbf{y}) = \mathbf{J}^{-1} \text{grad } H(\mathbf{y})$  holomorph in  $D \iff H(\mathbf{y})$  holomorph in  $D$   
 (mit jeweils gleicher unterer Schranke  $R$  für Konvergenzradius auf Kompakta)

- Mathematisches Pendel, Bsp. 4.4.3:  $H(p, q) = \frac{1}{2}p^2 - \cos q$   
 ➤  $D = \mathbb{R}^2$ ,  $R = \infty$  (ganze Funktion !)
- Federpendel, Bsp. 4.4.35:  $H(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \|\mathbf{p}\|^2 + \frac{1}{2}(\|\mathbf{q}\| - 1)^2$   
 ➤  $D = \mathbb{R}^4$ ,  $R = \text{dist}(K, \{\mathbf{q} = 0\})$  ( $H$  nicht holomorph in  $\mathbf{q} = 0$ )

Beachte:  $H(\mathbf{p}, \mathbf{q})$  jeweils analytisch in Umgebungen physikalisch sinnvoller Trajektorien !



 Notation:  $\tilde{\Phi}_{h,\ell}^t \hat{=}$  Evolutionsoperator zu  $\dot{\tilde{\mathbf{y}}} = \tilde{\mathbf{f}}_{h,\ell}(\tilde{\mathbf{y}})$ , vgl. Lemma 4.4.61

**Theorem 4.4.66** (Konsistenzfehlerabschätzung für abgeschnittene modifizierte Gleichungen).  
Sei  $\Psi^h$  die diskrete Evolution eines zu  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$  konsistenten (partitionierten) Runge-Kutta-Einschrittverfahrens. Unter der Analytizitätsvoraussetzung gibt es für jedes Kompaktum  $K \subset D$  Konstanten  $C_1, C_2 > 0$  und ein  $h_0 \in ]0, \infty]$  so, dass

$$\left\| \Psi^h \mathbf{y} - \tilde{\Phi}_{h,\ell}^h \mathbf{y} \right\| \leq C_1 h (C_2 (\ell + 1) h)^{\ell+1} \quad \forall \mathbf{y} \in K, \quad \forall \ell \in \mathbb{N}, \quad \forall |h| \leq h_0. \quad (4.4.67)$$

Hilfsmittel bei Beweis: Differentialgleichung in  $\mathbb{C}$   $\rightarrow$  Thm. 2.2.85

**Lemma 4.4.68.** Ist  $f$  holomorph in einer Umgebung von  $B_\rho(0)$ ,  $|f(z)| \leq M$  für alle  $z \in B_\rho(0)$  und  $f(0) = \dots = f^{(p)}(0) = 0$ ,  $p \in \mathbb{N}_0$ , dann gilt

$$|f(z)| \leq M |z|^{p+1} \rho^{-(p+1)} \quad \forall z \in B_\rho(0).$$

*Beweis.* Auf  $B_\rho(0)$

$$f(z) = z^{p+1} \underbrace{\sum_{j=0}^{\infty} a_j z^j}_{=:g(z)}, \quad |g(z)| \leq \frac{M}{\rho^{p+1}} \text{ für } |z| = \rho.$$

$g$  holomorph auf  $B_\rho(0)$   $\Rightarrow$   $|g|$  nimmt Maximum auf Rand  $|z| = \rho$  an (Maximumprinzip).  $\square$

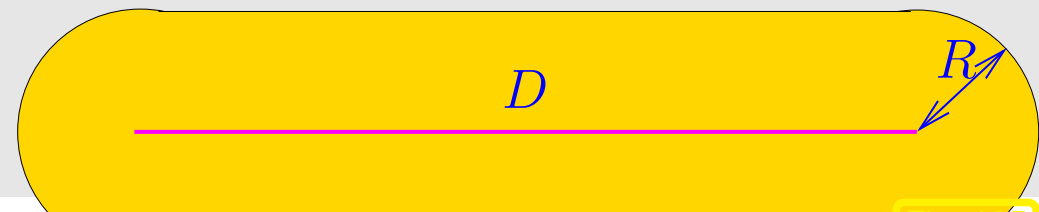
► Blosser Beschränktheit auf einer Nullumgebung einer holomorphen Funktion  $f$  mit  $|f(z)| = O(|z|^{p+1})$  genügt bereits, um das Abfallverhalten für  $z \rightarrow 0$  genau zu charakterisieren!

*Beweis* von Thm. 4.4.66  $\Rightarrow$  für skalaren Fall  $d = 1$ ,  $\dot{y} = f(y)$ ,  $D = ]a, b[ \subset \mathbb{R}$  Intervall,  
 $\Rightarrow$  für explizites Euler-Verfahren (1.4.2):  $\Psi^h y = y + hf(y)$

(Beweis nach S. Reich 1999, siehe [29, Thm. 2])

Annahme:

$f$  holomorph in Umgebung von



Ziel, vgl. (4.4.67): Abschätzung des Konsistenzfehlers  $\tilde{\Phi}_{h,\ell}^h(y) - \Psi^h(y)$  für modifizierte Gleichungen der Ordnung  $\ell + 1$  ( $\rightarrow$  Def. 4.4.47)  
 $\Updownarrow \leftarrow$  (4.4.54)  
 Abschätzung der Modifikatorfunktionen  $\Delta f_\ell$  !

**Schritt I:** Abschätzung für Modifikatorfunktion  $\Delta f_1$  (auch zur Demonstration der Technik)

! Interpretation von  $\dot{y} = f(y)$  als **Differentialgleichung in  $\mathbb{C}$**   $\rightarrow$  Thm. 2.2.85 :

$f$  holomorph  $\Rightarrow$  Lösungen  $t \mapsto y(t)$  analytisch (in Umgebung von 0)  $\Rightarrow$  fortsetzbar nach  $\mathbb{C}$   
 $\Rightarrow$  Evolution  $\Phi^t : B_R(D) \mapsto \mathbb{C}$  holomorph (für hinreichend kleines  $|t|$ )

► Im Folgenden: betrachte komplexe “Zeitschrittweiten”  $h \in \mathbb{C}$ ,  $0 < \alpha < 1$  fest gewählt.

Aus der Abschätzung für Wegintegrale im Komplexen

$$M := \max_{z \in B_R(D)} |f(z)| \Rightarrow |\Phi^h z - z| = \left| \int_0^h f(\Phi^\tau) d\tau \right| \leq M|h|, \quad \forall z \in B_{\alpha R}(D) \quad (4.4.69)$$

$$\Rightarrow |\Psi^h z - z| = |hf(z)| \leq M|h|.$$



Schranke für  $|h|$ :

Wenn  $z \in B_{\alpha R}(D)$  &  $|h| \leq (1-\alpha)\frac{R}{M}$ ,  $0 \leq \alpha < 1$   
dann bleibt die Trajektorie  $\xi \mapsto \Phi^{\xi h} z$ ,  $0 \leq \xi \leq 1$ ,  
in  $B_R(D)$  !

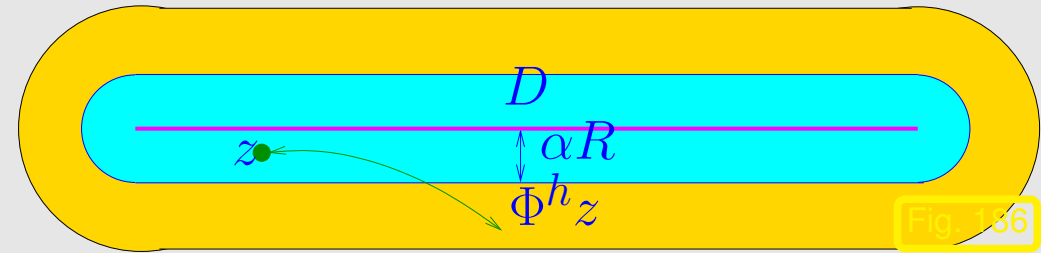


Fig. 156

$$|h| \leq h_1 := \frac{(1-\alpha)R}{M} \Rightarrow \begin{aligned} |\Phi^h y - y| &\leq (1-\alpha)R, \quad \forall y \in B_{\alpha R}(D) \\ |\Psi^h y - y| &\leq (1-\alpha)R. \end{aligned}$$

$$\Rightarrow \left( |h| \leq \frac{(1-\alpha)R}{M} \Rightarrow |\Phi^h y - \Psi^h y| \leq 2(1-\alpha)R \quad \forall y \in B_{\alpha R}(D) \right). \quad (4.4.70)$$

Anwendung von Lemma 4.4.68 auf  $g(h) = \Phi^h y - \Psi^h y$ :

- $g$  holomorph in  $B_{(1-\alpha)R/M}(0)$
- $g$  beschränkt durch  $2(1-\alpha)R$  auf  $B_{(1-\alpha)R/M}(0)$

$$\Rightarrow |\Phi^h y - \Psi^h y| \leq 2(1-\alpha)R |h|^2 \left( \frac{(1-\alpha)R}{M} \right)^{-2}$$

$$\leq 2M |h|^2 \left( \frac{M}{(1-\alpha)R} \right) \quad \forall y \in B_{\alpha R}(D), \quad (4.4.71)$$

da  $g(h) = O(h^2)$  (Euler-Verfahren Konsistenzordnung 1), so dass  $g(0) = g'(0) = 0$ .

$$\stackrel{(4.4.56)}{\Rightarrow} |\Delta f_1(y)| = \left| \lim_{h \rightarrow 0} \frac{\Phi^h y - \Psi^h y}{h^2} \right| \stackrel{(4.4.71)}{\leq} 2M \left( \frac{M}{(1-\alpha)R} \right) \quad \forall y \in B_{\alpha R}(D). \quad (4.4.72)$$

Wir haben nun gesehen, wie man unter der Analytizitätsannahme an die rechte Seite  $f$  eine Abschätzung für die erste Modifikatorfunktion erhalten kann. Benötigt wird eine Schranke für  $f$  in einer kompakten Umgebung  $B_R(D) \subset \mathbb{C}$ .

Die rekursive Konstruktion der Modifikatorfunktionen gemäss (4.4.56) legt nun folgendes Vorgehen nahe:

- ① Unter Verwendung von Abschätzungen für die Modifikatorfunktionen  $\Delta f_j$ ,  $1 \leq j \leq \ell$ , leite eine Abschätzung für die rechte Seite  $\tilde{f}_{h,\ell}$  der modifizierten Gleichung (4.4.57) aus Lemma 4.4.61 her. Ebenso wie alle Modifikatorfunktionen wird auch  $\tilde{f}_{h,\ell}$  analytisch in einer Umgebung von  $D$  sein.
- ② Benutze die Schranke für  $\tilde{f}_{h,\ell}$ , um mit gleichen Techniken wie oben für  $\Delta f_1$  die nächste Modifikatorfunktion  $\Delta f_{\ell+1}$  abzuschätzen.
- ③ Mache weiter mit ①

► Rekursive Abschätzung  $\longleftrightarrow$  Induktionsbeweis

Herausforderung: Formulierung einer geeigneten Induktionsannahme, vgl. (4.4.72).

**Schritt II. Induktionsbeweis:** Induktionsannahme: Es gibt  $\ell$ -unabhängige  $b > 0, c > 0$ , so dass

$$\forall l \in \mathbb{N}: \max_{y \in B_{\alpha R}(D)} |\Delta f_l(y)| \leq bM \left( \frac{clM}{(1-\alpha)R} \right)^\ell \quad \forall y \in B_{\alpha R}(D), \quad \forall 0 \leq \alpha < 1. \quad (4.4.73)$$

Die Konstanten  $b, c$  werden dann später geeignet festgelegt.

Induktionsbeginn “ $\ell = 0$ ”  $\Leftrightarrow$  (4.4.72)

Induktionsschritt “ $\ell \Rightarrow \ell + 1$ ”: ( $0 < \alpha < 1$  fixiert!)

$$\begin{aligned} |\tilde{f}_{h,\ell}(y)| &\leq |f(y)| + |h| |\Delta f_1(y)| + |h|^2 |\Delta f_2(y)| + \cdots + |h|^\ell |\Delta f_\ell(y)| \\ &\leq M + |h| \frac{2M}{(1-\alpha)R} + bM \sum_{j=2}^{\ell} |h|^j \left( \frac{jcM}{(1-\alpha)R} \right)^j, \quad \forall y \in B_{\alpha R}(D), \\ &\quad \forall 0 < \alpha < 1. \end{aligned}$$

aus (4.4.72)

nach Induktionsannahme (4.4.73)

?

Nötig: Schranke für  $|\tilde{f}_{h,\ell}(y)|$  in einer Umgebung von  $B_{\alpha R}(D)$

Idee: „ $\forall \alpha$ “ in (4.4.73)  $\Rightarrow$  nutze Freiheit in der Wahl von  $\alpha$  !

$$\alpha^* := \alpha + \delta(1 - \alpha) \in ]\delta, 1[ \Rightarrow 1 - \alpha^* = (1 - \alpha)(1 - \delta) \quad , \quad \alpha^* > \alpha .$$

$$\Rightarrow \max_{y \in B_{\alpha^* R}(D)} |\tilde{f}_{h,\ell}(y)| \leq M + |h| \frac{2M}{(1 - \alpha)(1 - \delta)R} + bM \sum_{j=2}^{\ell} \left( \frac{jcM|h|}{(1 - \alpha)(1 - \delta)R} \right)^j .$$

Versuch: Vereinfachung durch Beschränkung von  $|h|$ :

$$|h| \leq h_\ell := \frac{(1 - \alpha)R}{(\ell + 1)cM}$$

$$\Rightarrow \max_{y \in B_{\alpha^* R}(D)} |\tilde{f}_{h,\ell}(y)| \leq M \left( 1 + \frac{2}{c(\ell + 1)(1 - \delta)} + b \sum_{j=2}^{\ell} \left( \frac{j}{(\ell + 1)(1 - \delta)} \right)^j \right) . \quad (4.4.74)$$

Erinnerung an unser Ziel (4.4.73) für „ $\ell \leftarrow \ell + 1$ “. Wegen

$$\Delta \mathbf{f}_{\ell+1}(y) = - \lim_{h \rightarrow 0} \frac{\tilde{\Phi}_{h,\ell}^h y - \Psi^h y}{h^{\ell+2}} , \quad (4.4.56)$$

müssen wir also zeigen:

$$|\tilde{\Phi}_{h,\ell}^h y - \Psi^h y| \leq |h|^{\ell+2} bM \underbrace{\left( \frac{c(\ell + 1)M}{(1 - \alpha)R} \right)^{\ell+1}}_{=h_\ell^{-1} !} |h|^{\ell+2} bM h_\ell h_\ell^{-(\ell+2)} \quad \forall y \in B_{\alpha R}(D) \quad (4.4.75)$$

Beachte: (4.4.75)  $\Rightarrow$  Behauptung des Theorems mit  $C_1 = bM$ ,  $C_2 = \frac{cM}{(1-\alpha)R}$  !

Beachte: *per constructionem*, Lemma 4.4.61:

$$\tilde{\Phi}_{h,\ell}^h y - \Psi^h y = O(h^{\ell+2}) \text{ für } h \rightarrow 0$$

Lemma 4.4.68  $\Rightarrow$  Da  $h \mapsto \tilde{\Phi}_{h,\ell}^h y - \Psi^h y$  analytisch, genügt es zu zeigen

$$|\tilde{\Phi}_{h,\ell}^h y - \Psi^h y| \leq h_\ell b M \quad \forall h \in B_{h_\ell}(0), \quad \forall y \in B_{\alpha R}(D). \quad (4.4.76)$$

Dann Dreiecksungleichung wie in (4.4.70) & Abschätzung analog zu (4.4.69):

$$|\tilde{\Phi}_{h,\ell}^h y - y| \leq |h| \max_{y \in B_{\alpha^* R}(D)} |\tilde{f}_{h,\ell}(y)| \quad \forall y \in B_{\alpha R}(D), \quad |h| \text{ „hinreichend klein“}. \quad (4.4.77)$$

Was brauchen wir ? ( $|\Psi^h y - y| \leq M|h|$  wie oben)

- $\max_{y \in B_{\alpha^* R}(D)} |\tilde{f}_{h,\ell}(y)| \leq (b-1)M,$
- $h_\ell \cdot \max_{y \in B_{\alpha^* R}(D)} |\tilde{f}_{h,\ell}(y)| \leq \delta(1-\alpha)R,$  damit die Trajektorie  $z \mapsto \tilde{\Phi}_{h,\ell}^z y$  in  $B_{\alpha^* R}(D)$  bleibt, wenn  $|z| \leq h_\ell$  (Beachte:  $B_{\alpha R}(D) \subset B_{\alpha^* R}(D)$ ).

Dazu müssen wir die Parameter in (4.4.74) geeignet wählen!

Wir sind „frei“ in der Wahl von  $\delta \in ]0, 1[$  !  $\Rightarrow$   $\delta := \frac{b-1}{c} \cdot \frac{1}{\ell+1} \Rightarrow h_\ell(b-1)M = \delta(1-\alpha)R$

$$(4.4.74) \Rightarrow \max_{y \in B_{\alpha^* R}(D)} |\tilde{f}_{h,\ell}(y)| \leq M \underbrace{\left( 1 + \frac{2}{c(\ell+1) - b + 1} + b \sum_{j=2}^{\ell} \left( \frac{jc}{c(\ell+1) - b + 1} \right)^j \right)}_{=: \Gamma(b,c,\ell)} .$$

Frage: Gibt es  $b, c > 0$  ( $b-1 < 2c$ ) so, dass  $\max_{\ell \in \mathbb{N}} \Gamma(b, c, \ell) \leq b-1$  ?

Für Beweis von Thm. 4.4.66: Verhalten von

$$\Gamma(b, c, \ell) := 1 + \frac{2}{c(\ell+1) - b + 1} + b \sum_{j=2}^{\ell} \left( \frac{jc}{c(\ell+1) - b + 1} \right)^j :$$

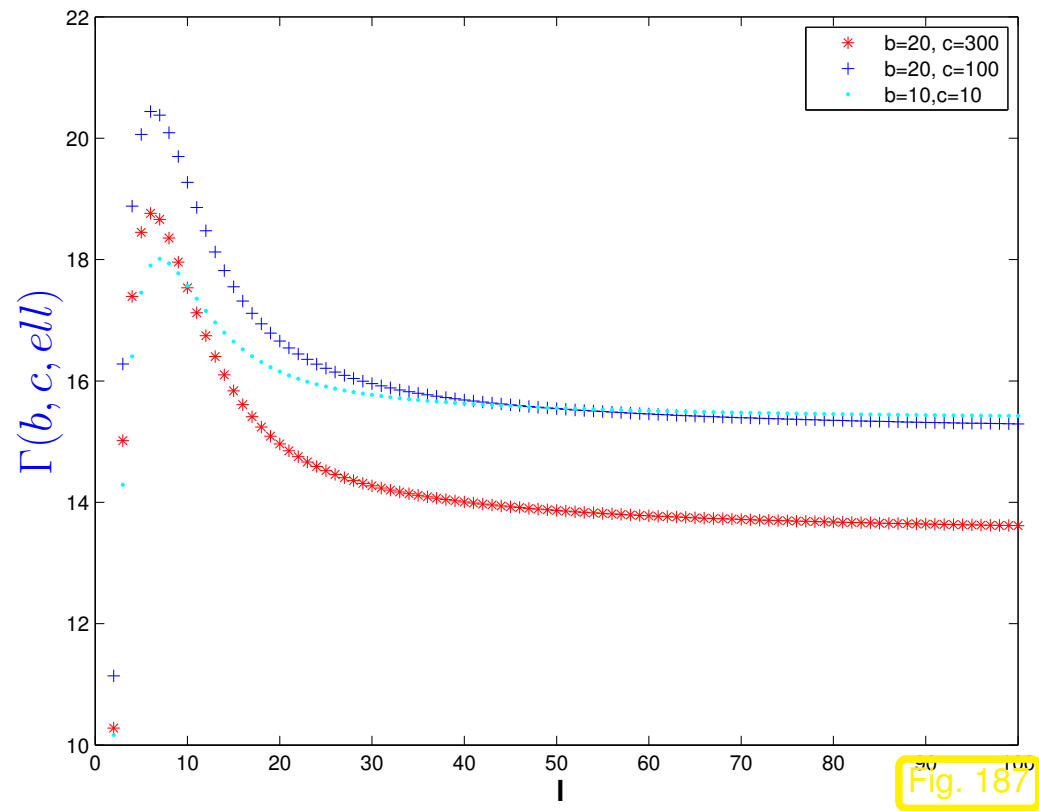


Fig. 187

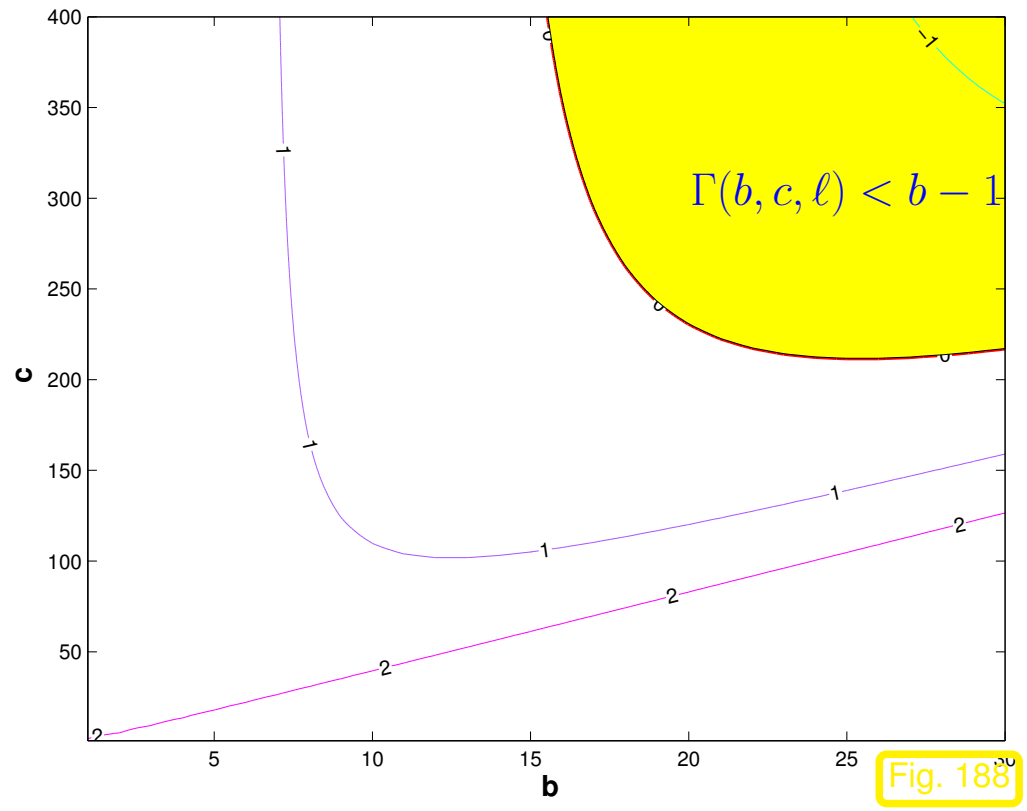


Fig. 188

Linker Plot:  $\ell \mapsto \Gamma(b, c, \ell) \succ$  eindeutiges Maximum für kleines  $\ell$

Rechter Plot: Konturen von  $(b, c) \mapsto \max_{\ell} \Gamma(b, c, \ell) - b + 1$

► Aus den Plots lesen wir ab: mögliche Wahl  $c = 300, b = 20 \quad \forall \ell.$

Dann weiter wie zuvor skizziert, siehe (4.4.76), (4.4.77):

$$\begin{aligned} \Rightarrow \quad & |\tilde{\Phi}_{h,\ell}^h y - y| \leq M(b-1)|h|, \\ \Rightarrow \quad & |\tilde{\Phi}_{h,\ell}^h y - \Psi^h y| \leq bM|h| \end{aligned} \quad \text{für } |h| \leq h_\ell, \quad \forall y \in B_{\alpha R}(D), \quad (4.4.78)$$

Beachte:  $\tilde{\Phi}_{h,\ell}^h y - \Psi^h y = O(h^{\ell+2})$  nach Konstruktion der Modifikatorfunktionen und holomorph in Umgebung von 0. Mit Formel für  $h_\ell$ , o.B.d.A.  $0 < h_\ell < 1$ ,

$$\stackrel{\text{Lemma 4.4.68}}{\Rightarrow} \quad |\tilde{\Phi}_{h,\ell}^h y - \Psi^h y| \leq bM \left( \frac{|h|}{h_\ell} \right)^{\ell+2} \leq bM |h|^{\ell+2} \left( \frac{c(\ell+1)M}{(1-\alpha)R} \right)^{\ell+1}. \quad (4.4.79)$$

Mit (4.4.56) folgt die Induktionsbehauptung für  $\ell + 1$ .

Behauptung des Theorems mit  $C_1 = bM$ ,  $C_2 = \frac{cM}{R}$  (Fall  $\alpha = 0$ ) folgt ebenfalls aus (4.4.79)  $\square$



Verhalten der Schranke aus Thm. 4.4.66  $\triangleright$

Mögliche Divergenz der asymptotischen Entwicklung (4.4.55) manifestiert sich in  $C_1 h (C_2 (\ell + 1) h)^{\ell+1} \rightarrow \infty$  für  $\ell \rightarrow \infty$ .

Optimaler Abbruchindex:

$$l_{\text{opt}} \approx \left\lceil \frac{1}{C_2 e h} \right\rceil. \quad (4.4.80)$$

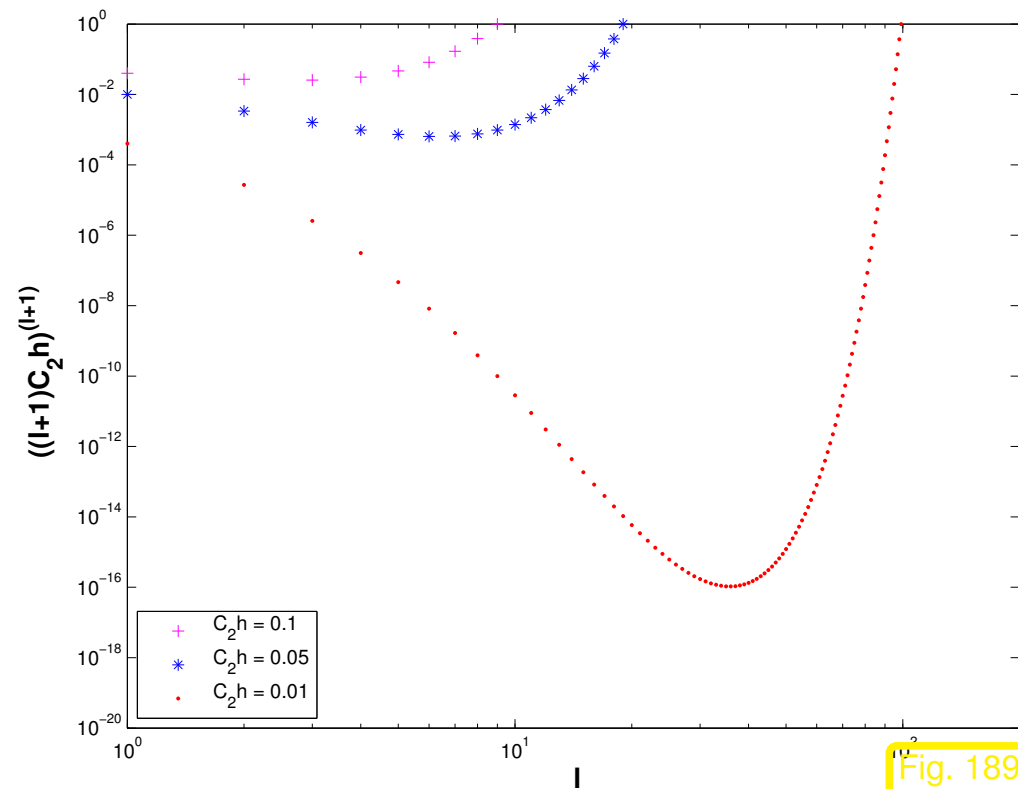
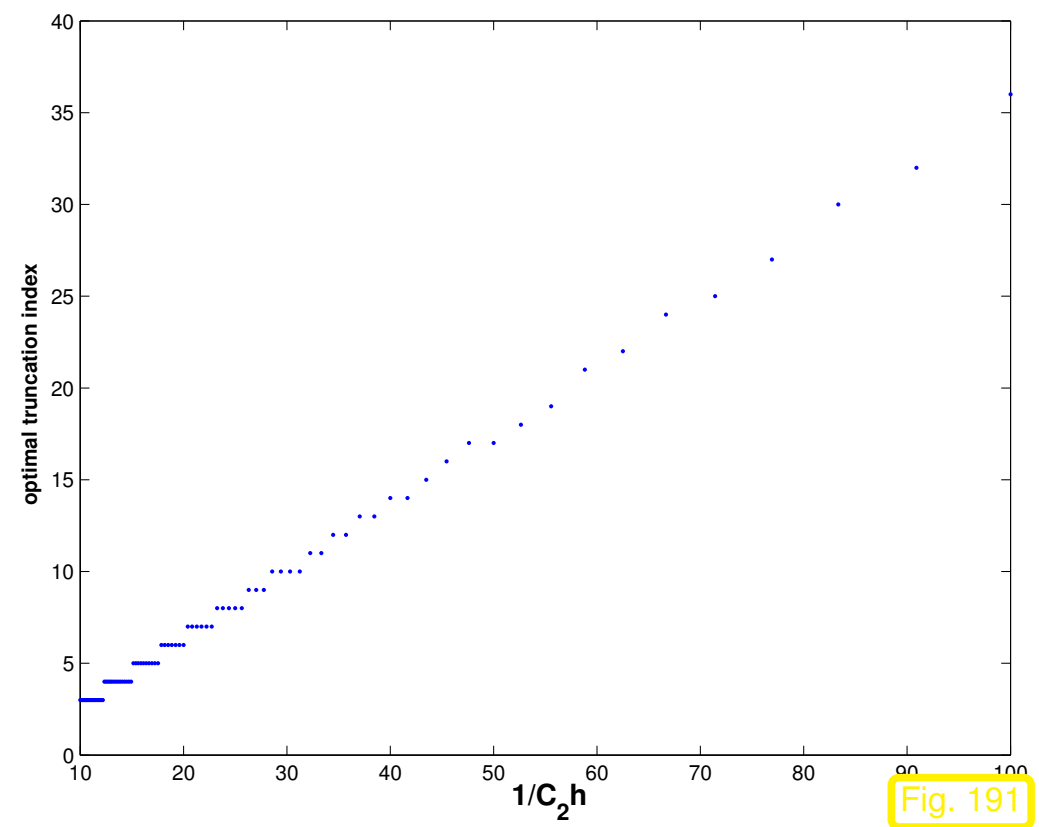
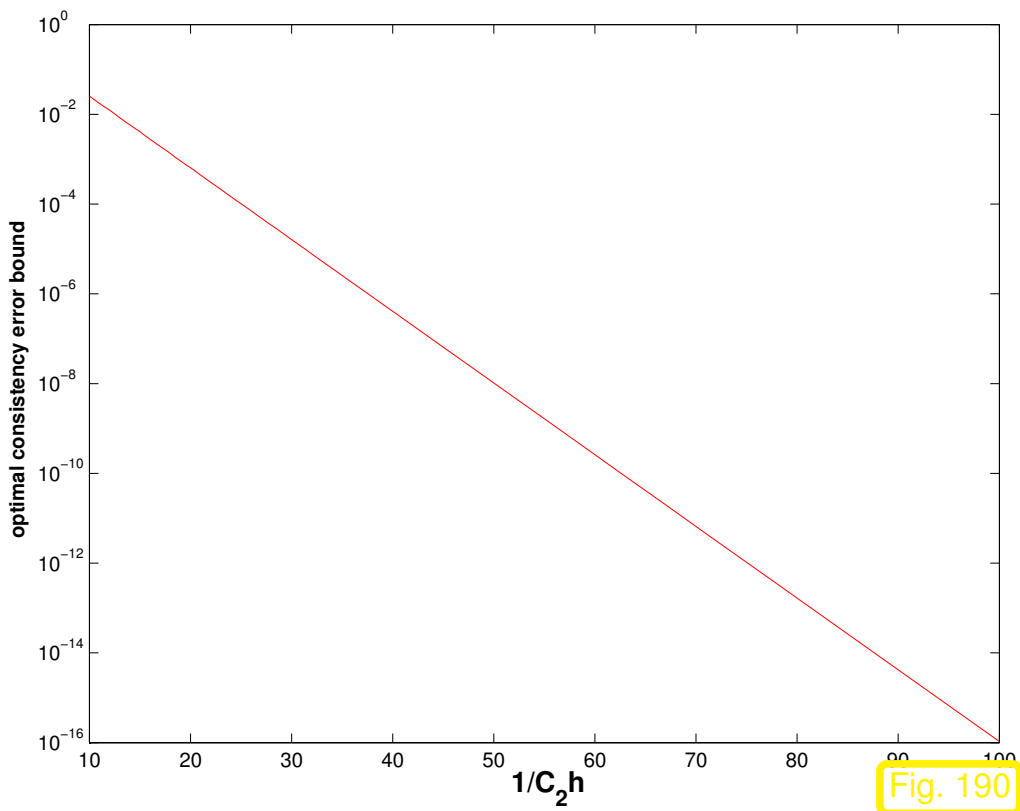



Fig. 189



 Notation:  $[x] \hat{=}$  ganzzahliger Anteil von  $x > 0$

(4.4.80) ergibt sich aus Kurvendiskussion von  $x \mapsto (ax)^x, x > 0$ .

$$\left\| \Psi^h \mathbf{y} - \tilde{\Phi}_{h, \ell_{\text{opt}}}^h \mathbf{y} \right\| \leq C_1 h \exp(-\ell_{\text{opt}}) \leq C_1 h \exp(-\gamma/h), \quad \gamma := \frac{1}{C_2 e} > 0. \quad (4.4.81)$$

Schranke **exponentiell klein** für  $h \rightarrow 0$ , vgl. (4.4.63)

Beachte: (4.4.81)  $\Leftrightarrow$  Konsistenzfehlerabschätzung (2.1.18)

Nun leiten wir eine Abschätzung für die Abweichung der numerischen Lösung von der Lösungstrajektorie der optimal abgeschnittenen modifizierten Gleichung  $\dot{\tilde{\mathbf{y}}} = \tilde{\mathbf{f}}_{h, \ell_{\text{opt}}}(\tilde{\mathbf{y}})$  her, vgl. (4.4.62).

Betrachte: AWP  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$   $\mathbf{y}(0) = \mathbf{y}_0 \in D$  auf  $[0, T]$ , Endzeitpunkt  $T \in J(\mathbf{y}_0)$

Notationen:  $\tilde{\mathbf{y}} \hat{=} \text{Lösung des AWP}$   $\dot{\tilde{\mathbf{y}}} = \tilde{\mathbf{f}}_{h, \ell_{\text{opt}}}(\tilde{\mathbf{y}}), \tilde{\mathbf{y}}(0) = \mathbf{y}_0$   
 $(\ell_{\text{opt}}$  aus (4.4.80) mit  $C_2$  aus Thm. 4.4.66 bzgl.  $K$ )

$(\mathbf{y}_k^h)_k := ((\Psi^h)^k \mathbf{y}_0)_k, k \in \{0, \dots, [T/h]\}$ : Gitterfunktion erzeugt durch das  
 Einschrittverfahren mit Schrittweite  $h > 0$  (numerische Näherungslösung)

**Lemma 4.4.82** (Konvergenz der optimal abgeschnittenen modifizierten Gleichung).

- Es gebe eine kompakte Umgebung  $K \subset D$  von  $\mathbf{y}_0$ , so dass  $\mathbf{y}_k^h \in K$  für alle  $k \in \mathbb{N}$ , wenn  $h$  hinreichend klein.
- Es gelten die Voraussetzungen von Thm. 4.4.66 (Analytizitätsannahme).
- Die diskrete Evolution zum ESV besitze die Darstellung  $\Psi^h \mathbf{y} = \mathbf{y} + h\psi(\mathbf{y}, h)$  mit einer auf  $K$  gleichmässig Lipschitz-stetigen Inkrementfunktion  $\psi$ , d.h., vgl. 2.1.24,

$$\exists L > 0: \|\psi(\mathbf{z}, h) - \psi(\mathbf{w}, h)\| \leq L \|\mathbf{z} - \mathbf{w}\| \quad \forall \mathbf{z}, \mathbf{w} \in K, |h| \text{ hinreichend klein.}$$

Dann gibt es  $h_0 > 0$  und von  $h_0$  unabhängige Konstanten  $C > 0, \gamma > 0$  so, dass

$$\|\tilde{\mathbf{y}}(hk) - \mathbf{y}_k^h\| \leq C(\exp(hkL) - 1) \exp(-\gamma/h) \quad \forall k \in \{0, \dots, [T/h]\}, \quad \forall 0 < h < h_0.$$

*Beweis.* Siehe (4.4.62) und die dortigen Bemerkungen:

Der Beweis von Thm. 2.1.19 kann fast unverändert übertragen werden, nachdem (2.1.23) durch (4.4.81) ersetzt worden ist. Siehe auch Sect. 2.1.4 für die Beweistechnik.  $\square$

$$T < \frac{\gamma}{Lh} \Rightarrow \text{“exponentiell kleiner” Fehler des Einschrittverfahrens bzgl. der optimal abgeschnittenen modifizierten Gleichung}$$

Nächster Punkt: Entsteht die optimal abgeschnittene modifizierte Gleichung wirklich durch eine „kleine“ Störung der ursprünglichen ODE?

**Lemma 4.4.83** (Störungsabschätzung für optimal abgeschnittene modifizierten Gleichung).

Neben den Voraussetzungen von Thm. 4.4.66 (Analytizitätsannahme) gibt es für jedes Kompaktum  $K \subset D$  eine von (hinreichend kleinem)  $h > 0$  unabhängige Konstante  $C > 0$  so, dass

$$\left\| \tilde{\mathbf{f}}_{h,\ell}(\mathbf{y}) - \mathbf{f}(\mathbf{y}) \right\| \leq Ch^p \quad \forall \mathbf{y} \in K, \quad \forall \ell \in \mathbb{N}.$$

*Beweis.* Ergänzung zum Beweis von Thm. 4.4.66, siehe die dort gemachten Annahmen und verwendeten Notationen. Ausführungen für das explizite Euler-Verfahren, d.h.  $p = 1$ .

Aus der Definition von  $\tilde{f}_{h,\ell}$ ,  $\rightarrow$  Lemma 4.4.61,

$$\tilde{f}_{h,\ell}(y) - f(y) = \sum_{j=1}^{\ell} h^j \Delta f_j(y).$$

Idee: Verwende Abschätzung der Modifikatorfunktionen  $\Delta f_j$  aus dem Beweis von Thm. 4.4.66

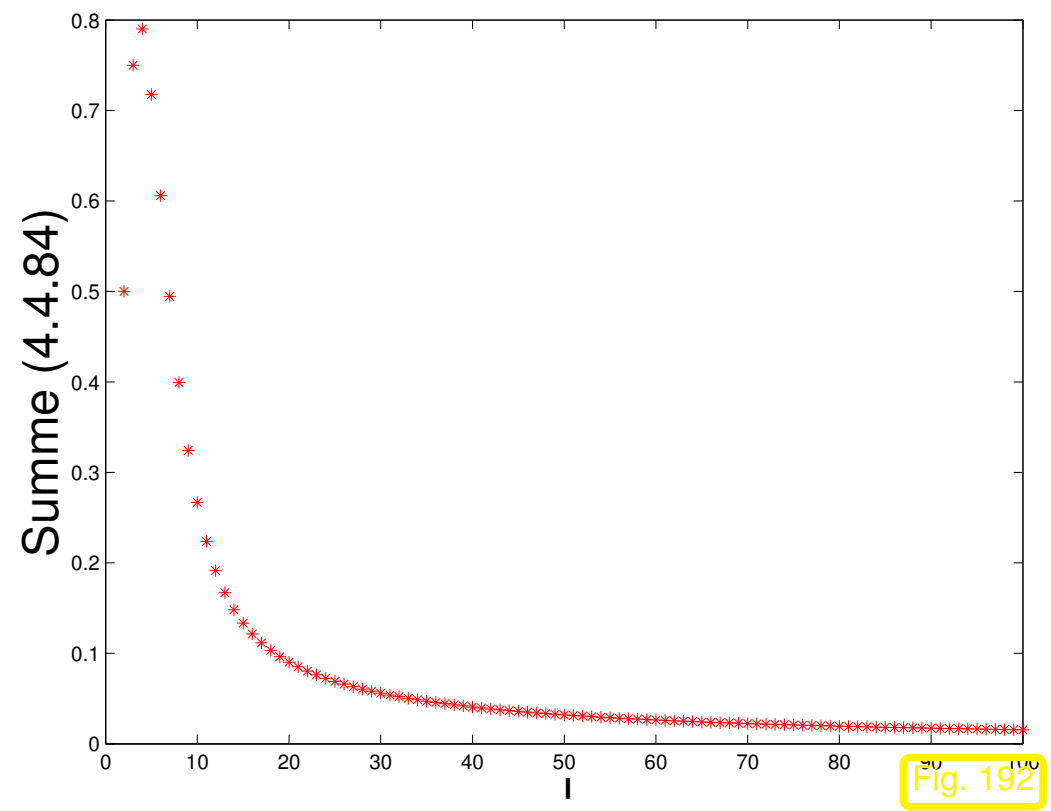
Konkret: aus (4.4.73) mit  $\alpha = 0$

$$|\tilde{f}_{h,\ell}(y) - f(y)| \leq |h| \left( \frac{2M}{R} + bM \sum_{j=2}^{\ell} |h|^{j-1} \left( \frac{jcM}{R} \right)^j \right).$$

$$|h| \leq \frac{R}{(\ell + 1)2cM} \Rightarrow |\tilde{f}_{h,\ell}(y) - f(y)| \leq |h| \left( \frac{2M}{R} + \frac{2bcM^2}{R} \underbrace{\sum_{j=2}^{\ell} 2^{-j} j \left( \frac{j+1}{\ell+1} \right)^j}_{\text{beschränkt}} \right).$$

Summe

$$\ell \mapsto \sum_{j=2}^{\ell} 2^{-j} j \left( \frac{j+1}{\ell+1} \right)^j \quad (4.4.84)$$



Also:  $|\tilde{f}_{h,\ell}(y) - f(y)| \leq Ch$  für kleines  $h$ ,  $\forall y \in K$ , mit  $C > 0$  unabhängig von  $\ell$ . □

Das Vektorfeld der optimal abgeschnittenen modifizierten Gleichung ist “ $O(h^p)$ -nah” zu  $\mathbf{f}$

*Bemerkung 4.4.85* (Schrittweitenbedingungen für „Langzeitintegration“).

- Betrachte AWP  $\dot{\mathbf{y}} = \mathbf{f}(\mathbf{y})$ ,  $\mathbf{y}(0) = \mathbf{y}_0$ , auf  $[0, T] \subset J(\mathbf{y}_0)$ ,  $\mathbf{f}$  holomorph.
- ESV  $\mathbf{y}_1 = \Psi^h \mathbf{y}_0$  der Konsistenzordnung  $p \in \mathbb{N}$ .

Schrittweitenbedingung für **genaue** numerische Lösung ( $\|\mathbf{y}(hk) - \mathbf{y}_k\|$  klein) auf  $[0, T]$

$$\text{Thm. 2.1.19} \Rightarrow h^p \exp(LT) \ll 1 \Rightarrow \boxed{h = O(\exp(-T/p))}.$$

Schrittweitenbedingung für **akzeptable** (\*) numerische Lösung ( $\|\tilde{\mathbf{y}}(hk) - \mathbf{y}_k\|$  klein) auf  $[0, T]$

$$\text{Lemmas 4.4.82, 4.4.83} \Rightarrow h < \frac{\gamma}{LT} \Rightarrow \boxed{h = O(T^{-1})}.$$





(\*) „generisch akzeptabel“ bzgl. allgemeiner additiver Störungen von  $\mathbf{f}$ . (Schärfer: strukturerhaltend akzeptabel, siehe Anfang von Sect. 4.4.3)

## 4.4.5 Strukturerhaltende modifizierte Gleichungen

Gemäss Bem. 4.4.43 müssen wir zu zeigen, dass die Vektorfelder  $\tilde{\mathbf{f}}_{h,\ell}$  der abgeschnittenen modifizierten Gleichungen strukturelle Eigenschaften ( $\mathbf{f} \in V$  von Bem. 4.4.14) von  $\mathbf{f}$  erben. Fokus ist auf Hamiltonschen Differentialgleichungen ( $\rightarrow$  Def. 1.2.20)  $\leftrightarrow$  Symplektizität ( $\rightarrow$  Def. 4.4.12)

R. Hiptmair  
rev 35327,  
25. April  
2011

*Beispiel* 4.4.86 (Modifizierte Gleichung für symplektisches Euler-Verfahren).  $\rightarrow$  [26, Sect. 5.1.2]

Separierte Hamilton-Funktion mit Potential  $U : \mathbb{R}^n \mapsto \mathbb{R}$ , vgl. (1.2.28)

$$H(\mathbf{p}, \mathbf{q}) := \frac{1}{2} \|\mathbf{p}\|^2 + U(\mathbf{q}), \quad \mathbf{p}, \mathbf{q} \in \mathbb{R}^n.$$

► Hamiltonsche Differentialgleichung:

$$\dot{\mathbf{p}} = -\text{grad } U(\mathbf{q}), \quad \dot{\mathbf{q}} = \mathbf{p}. \quad (4.4.87)$$

① Explizites Euler-Verfahren für (4.4.87), Schrittweite  $h > 0$  (Konsistenzordnung 1):

$$\mathbf{p}_1 = \mathbf{p}_0 - h \operatorname{grad} U(\mathbf{q}_0) \quad , \quad \mathbf{q}_1 = \mathbf{q}_0 + h\mathbf{p}_0 \quad .$$

Taylorentwicklung & (4.4.87) & (4.4.54)  $\triangleright$  erste Modifikatorfunktion

$$\begin{aligned} \mathbf{p}(h) &= \mathbf{p}_0 + h\dot{\mathbf{p}}(0) + \frac{1}{2}h^2\ddot{\mathbf{p}}(0) + O(h^3) \\ &= \mathbf{p}_0 - h \operatorname{grad} U(\mathbf{q}_0) - \frac{1}{2}h^2 \nabla^2 U(\mathbf{q}_0)\mathbf{p}_0 + O(h^3) \quad , \\ \mathbf{q}(h) &= \mathbf{q}_0 + h\dot{\mathbf{q}}(0) + \frac{1}{2}h^2\ddot{\mathbf{q}}(0) + O(h^3) \\ &= \mathbf{q}_0 + h\mathbf{p}_0 - \frac{1}{2}h^2 \operatorname{grad} U(\mathbf{q}_0) + O(h^3) \quad . \end{aligned}$$

► Ausdruck für den Konsistenzfehler:

$$\boldsymbol{\tau}(\mathbf{y}_0, h) = \begin{pmatrix} \mathbf{p}(h) - \mathbf{p}_1 \\ \mathbf{q}(h) - \mathbf{q}_1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}h^2 \nabla^2 U(\mathbf{q}_0)\mathbf{p}_0 \\ -\frac{1}{2}h^2 \operatorname{grad} U(\mathbf{q}_0) \end{pmatrix} + O(h^3) \quad .$$

$$(4.4.56) \quad \Rightarrow \quad \Delta \mathbf{f}_1(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \begin{pmatrix} \nabla^2 U(\mathbf{q})\mathbf{p} \\ \operatorname{grad} U(\mathbf{q}) \end{pmatrix} \neq \mathbf{J}^{-1} \operatorname{grad} \tilde{H}(\mathbf{y}) \quad .$$

② Symplektisches Euler-Verfahren (4.4.26), Schrittweite  $h > 0$  (Konsistenzordnung 1):

$$\mathbf{p}_1 = \mathbf{p}_0 - h \operatorname{grad} U(\mathbf{q}_1) \quad , \quad \mathbf{q}_1 = \mathbf{q}_0 + h\mathbf{p}_0 .$$

Taylorentwicklung & (4.4.87) & (4.4.54)  $\triangleright$  Modifikatorfunktion

Im Unterschied zu oben, unter Verwendung von (4.4.26):

$$\begin{aligned} \mathbf{q}(h) &= \mathbf{q}_0 + h\mathbf{p}_0 - \frac{1}{2}h^2 \operatorname{grad} U(\mathbf{q}_0) + O(h^3) \\ &= \mathbf{q}_0 + h\mathbf{p}_1 + \frac{1}{2}h^2 \operatorname{grad} U(\mathbf{q}_0) + O(h^3) \end{aligned}$$

R. Hiptmair

rev 35327,  
25. April  
2011

$$\Delta \mathbf{f}_1(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \begin{pmatrix} \nabla^2 U(\mathbf{q}) \mathbf{p} \\ -\operatorname{grad} U(\mathbf{q}) \end{pmatrix} = \begin{pmatrix} -\frac{\partial \tilde{H}_1}{\partial \mathbf{q}}(\mathbf{p}, \mathbf{q}) \\ \frac{\partial \tilde{H}_1}{\partial \mathbf{p}}(\mathbf{p}, \mathbf{q}) \end{pmatrix} , \quad \tilde{H}_1(\mathbf{p}, \mathbf{q}) = -\frac{1}{2} \mathbf{p} \cdot \operatorname{grad} U(\mathbf{q}) .$$

$\Rightarrow$  Modifizierte Gleichung zweiter Ordnung ist Hamiltonsche Differentialgleichung !

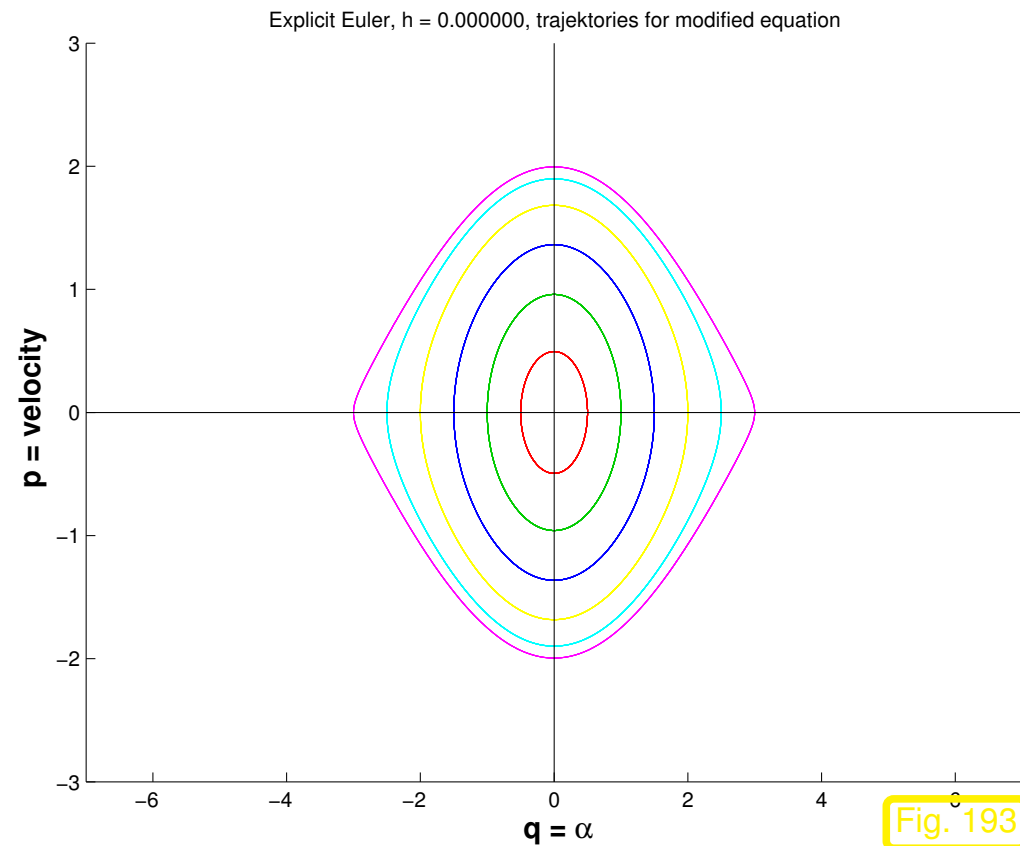
Konkret: mathematisches Pendel  $\rightarrow$  Bsp. 1.2.17

$$n = 1, \quad H(p, q) = \frac{1}{2}p^2 - \cos q.$$

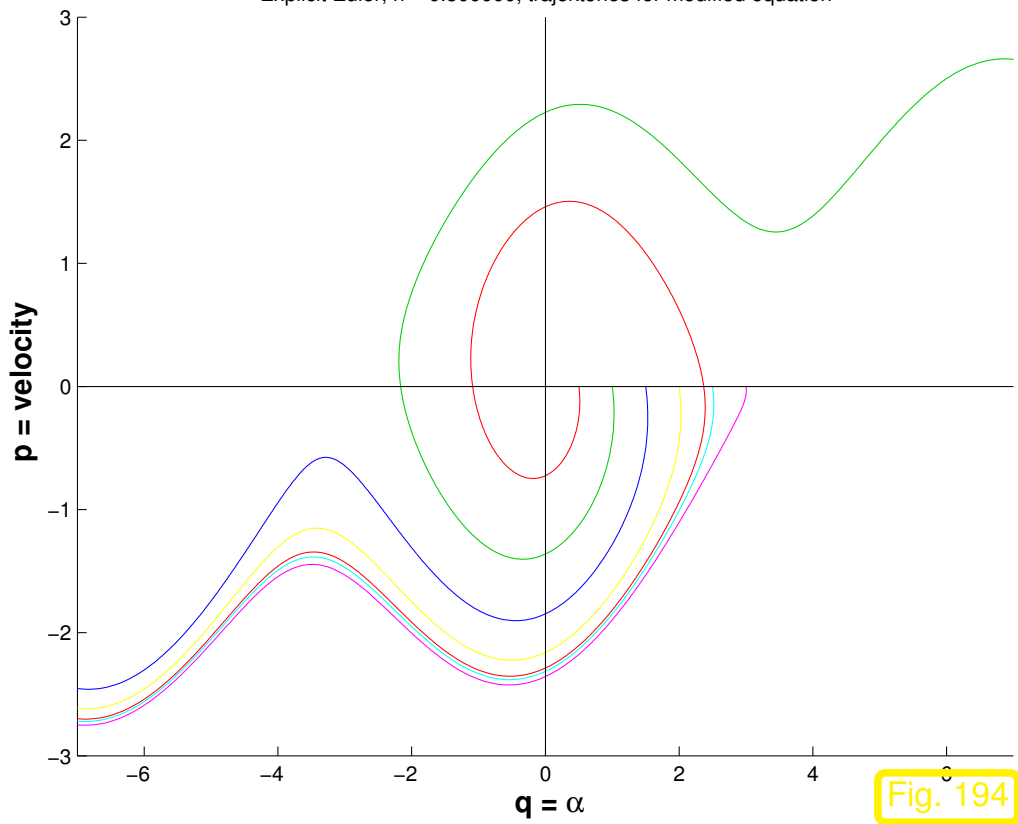
“exakte” Trajektorien



Trajektorien zu modifizierten Gleichungen 2. Ordnung

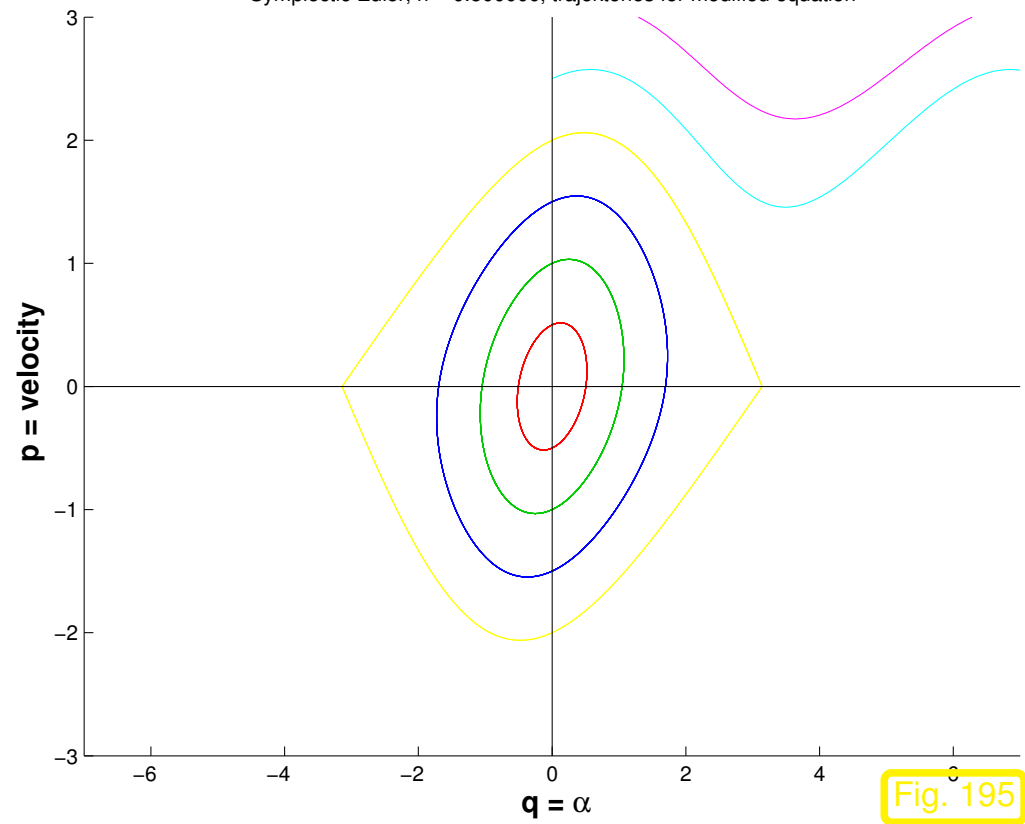


Explicit Euler,  $h = 0.500000$ , trajektorien for modified equation



Expl. Euler,  $h = 0.5$

Symplectic Euler,  $h = 0.500000$ , trajektorien for modified equation



Synplekt. Euler,  $h = 0.5$



**Theorem 4.4.88** (Symplektizität der Modifikatorfunktionen).

Sei  $\Psi^h$  die diskrete Evolution eines symplektischen Einschrittverfahrens ( $\rightarrow$  Def. 4.4.18) für die Hamiltonsche Differentialgleichung  $\dot{\mathbf{y}} = \mathbf{J}^{-1} \cdot \mathbf{grad} H(\mathbf{y})$  mit glatter Hamilton-Funktion  $H : D \subset \mathbb{R}^{2n} \mapsto \mathbb{R}$ ,  $D$  sternförmig.

Dann sind die abgeschnittenen modifizierten Gleichungen  $\dot{\tilde{\mathbf{y}}} = \tilde{\mathbf{f}}_{h,\ell}(\tilde{\mathbf{y}})$  aus Lemma 4.4.61 ebenfalls Hamiltonsch für alle  $\ell \in \mathbb{N}$  und alle (hinreichend kleinen)  $h > 0$ .

*Beweis.* (von Thm. 4.4.88) Idee: **Induktion** nach  $\ell$

Induktionsbeginn: Für  $\ell \leq p$ :  $\tilde{\mathbf{f}}_{h,\ell}(\mathbf{y}) = \mathbf{f}(\mathbf{y}) = \mathbf{J}^{-1} \cdot \mathbf{grad} H(\mathbf{y})$  ✓

“ $\ell \rightarrow \ell + 1$ ”:  $\tilde{\Phi}^t \hat{=}$  Evolutionsoperator zu  $\dot{\tilde{\mathbf{y}}} = \tilde{\mathbf{f}}_{h,\ell}(\tilde{\mathbf{y}}) = \mathbf{J}^{-1} \mathbf{grad} \tilde{H}_\ell(\mathbf{y})$  (Induktionsannahme !)  
 $\Rightarrow$  Für festes  $t$ :  $\tilde{\Phi}^t : D \mapsto \mathbb{R}^{2n}$  is symplektisch ( $\rightarrow$  Def. 4.4.12)

Nach (4.4.52) & (4.4.54), Lemma 4.4.61 für  $h \rightarrow 0$

$$\tilde{\Phi}^h \mathbf{y}_0 - \Psi^h \mathbf{y}_0 = -\Delta \mathbf{f}_{\ell+1}(\mathbf{y}_0) h^{\ell+2} + O(h^{\ell+3}) \quad \forall \mathbf{y}_0 . \quad (4.4.89)$$

$$(D_{\mathbf{y}}\tilde{\Phi}^h)(\mathbf{y}_0) - (D_{\mathbf{y}}\Psi^h)(\mathbf{y}_0) = -(D_{\mathbf{y}}\Delta\mathbf{f}_{\ell+1})(\mathbf{y})h^{\ell+2} + O(h^{\ell+3}). \quad (4.4.90)$$

$\tilde{\Phi}^h, \Psi^h$  = symplektische Abbildungen ( $\rightarrow$  Def. 4.4.12, Argument  $\mathbf{y}_0$  weggelassen)  $\Rightarrow$

$$\begin{aligned} \underbrace{(D_{\mathbf{y}}\tilde{\Phi}^h)^T \mathbf{J} D_{\mathbf{y}}\tilde{\Phi}^h}_{=\mathbf{J}} &= \underbrace{(D_{\mathbf{y}}\Psi^h)^T \mathbf{J} D_{\mathbf{y}}\Psi^h}_{=\mathbf{J}} \\ &\quad + h^{\ell+2} \left( (D_{\mathbf{y}}\Psi^h)^T \mathbf{J} D_{\mathbf{y}}\Delta\mathbf{f}_{\ell+1} + (D_{\mathbf{y}}\Delta\mathbf{f}_{\ell+1})^T \mathbf{J} D_{\mathbf{y}}\Psi^h \right) + O(h^{\ell+3}). \\ \Rightarrow 0 &= (D_{\mathbf{y}}\Psi^h)^T \mathbf{J} D_{\mathbf{y}}\Delta\mathbf{f}_{\ell+1} + (D_{\mathbf{y}}\Delta\mathbf{f}_{\ell+1})^T \mathbf{J} D_{\mathbf{y}}\Psi^h + O(h). \end{aligned}$$

Für konsistentes ESV, siehe Lemma 2.1.9:  $D_{\mathbf{y}}\Psi^h = \mathbf{I} + O(h)$  für  $h \rightarrow 0$ .

$$\xrightarrow{h \rightarrow 0} 0 = \mathbf{J} D_{\mathbf{y}}\Delta\mathbf{f}_{\ell+1} + (D_{\mathbf{y}}\Delta\mathbf{f}_{\ell+1})^T \mathbf{J} \Rightarrow D_{\mathbf{y}}(\mathbf{J}\Delta\mathbf{f}_{\ell+1}) = (D_{\mathbf{y}}(\mathbf{J}\Delta\mathbf{f}_{\ell+1}))^T.$$

► Anwendung von Lemma 4.4.17 (Integrabilitätslemma) auf  $\mathbf{J}\Delta\mathbf{f}_{\ell+1}$ . □

Symplektische Integratoren liefern strukturhaltende **akzeptable** (\*) diskrete Evolutionen für (glatte) konservative mechanische Systeme.

(\*) : (exponentiell genaue) Lösung einer Evolution mit "leicht gestörter" (nämlich  $O(h^p)$ , siehe Lemma 4.4.83) Hamilton-Funktion  $\rightarrow$  Rückwärtsanalyse  $\rightarrow$  Sect. 4.4.3

Erklärung der “Langzeitenergieerhaltung” symplektischer Integratoren durch Rückwärtsanalyse:

☞ Lösung des ESV (Konsistenzordnung  $p$ ) ist “exponentiell genau” ( $\rightarrow$  Lemma 4.4.82) Approximation der Lösung einer (optimal abgeschnittenen) modifizierten Gleichung

Diese ist eine **Hamiltonsche Differentialgleichung** ( $\rightarrow$  Def. 1.2.20) mit einer (bzgl.  $H$ ) um  $O(h^p)$  gestörten Hamilton-Funktion  $\tilde{H}(\mathbf{y})$  ( $\rightarrow$  Thm. 4.4.83).

Details:

**Annahmen:** • symplektisches ESV der Konsistenzordnung  $p$

- Schrittweite  $h < h^* \Rightarrow$  numerischen Lösungen  $(\mathbf{y}_k)_k \subset K \subset D$ ,  $K$  kompakte Teilmenge  $K \subset D$  des Zustandsraums.
- $K$  ist sternförmig (darauf kann verzichtet werden [16, Sect. XI.3.2])
- $\mathbf{f}$  erfüllt Analytizitätsannahme bzgl.  $K$

☞ Notation:  $(t, \mathbf{y}) \mapsto \tilde{\Phi}_h^t \mathbf{y} \hat{=} \text{Evolutionsoperator zur (optimal abgeschnittenen) modifizierten Gleichung } \dot{\tilde{\mathbf{y}}} = \tilde{\mathbf{f}}_{h, \ell_{\text{opt}}}(\tilde{\mathbf{y}}).$

$$\text{Abschätzung (4.4.81)} \Rightarrow \left\| \Psi^h \mathbf{y} - \tilde{\Phi}_h^h \mathbf{y} \right\| \leq Ch \exp(-\gamma/h) \quad \forall \mathbf{y} \in K, \forall h < h^* .$$



Thm. 4.4.88  $\Rightarrow \tilde{\mathbf{f}}_{h, \ell_{\text{opt}}}(\mathbf{y}) = \mathbf{J}^{-1} \text{grad } \tilde{H}_h(\mathbf{y})$  mit  $\tilde{H}_h : K \mapsto \mathbb{R}$  holomorph .

$\tilde{H}_h : K \mapsto \mathbb{R}$  holomorph  $\stackrel{K \text{ kompakt}}{\Rightarrow} \exists L > 0: |\tilde{H}_h(\mathbf{y}) - \tilde{H}_h(\mathbf{z})| \leq L \|\mathbf{y} - \mathbf{z}\| \quad \forall \mathbf{y}, \mathbf{z} \in K$  .

„Teleskopsummenargument“: da  $\tilde{H}_h(\tilde{\Phi}_h^t \mathbf{y}) = \tilde{H}_h(\mathbf{y})$  für alle  $t \in J(\mathbf{y})$ ,  $\mathbf{y} \in K$  (Hamilton-Funktion ist Invariante einer Hamiltonschen ODE, siehe Lemma 1.2.23)

$$\begin{aligned} |\tilde{H}_h(\mathbf{y}_k) - \tilde{H}_h(\mathbf{y}_0)| &\leq \sum_{j=0}^{k-1} |\tilde{H}_h(\mathbf{y}_{j+1}) - \tilde{H}_h(\mathbf{y}_j)| = \sum_{j=0}^{k-1} |\tilde{H}_h(\Psi^h \mathbf{y}_j) - \tilde{H}_h(\tilde{\Phi}_h^h \mathbf{y}_j)| \\ &\leq \sum_{j=0}^{k-1} L \left\| \Psi^h \mathbf{y}_j - \tilde{\Phi}_h^h \mathbf{y}_j \right\| \leq CL \sum_{j=0}^{k-1} h \exp(-\gamma/h) \leq CLhk \exp(-\gamma/h) . \end{aligned}$$

Thm. 4.4.88  $\Rightarrow \exists C > 0: \max_{\mathbf{y} \in K} |\tilde{H}_h(\mathbf{y}) - H(\mathbf{y})| \leq Ch^p \quad \forall h < h^*$  .

$$\Rightarrow |H(\mathbf{y}_k) - H(\mathbf{y}_0)| \leq C(Lhk \exp(-\gamma/h) + h^p) \quad \forall h < h^* .$$

$LT \lesssim h^p \exp(-\gamma/h) \triangleright$  „Energiefehler“ der numerischen Lösung von der Grösse  $O(h^p)$

für „exponentiell lange Zeit“

☞ Dies ist die Erklärung für die Vermutung aus Bsp. (4.4.34) !

**Theorem 4.4.91** (Langzeitenergieerhaltung bei symplektischer Integration).

Für die Hamiltonsche ODE  $\dot{\mathbf{y}} = \mathbf{J}^{-1} \text{grad } H(\mathbf{y})$  ( $\rightarrow$  Def. 1.2.20) und ein dazu von Ordnung  $p$  konsistentes symplektisches Einschrittverfahren ( $\rightarrow$  Def. 4.4.18) seien die Voraussetzungen von Thm. 4.4.66 erfüllt.

Für hinreichend kleine (uniforme !) Schrittweiten  $h$  gelte  $(\Psi^h)^k \mathbf{y} \in K$  für alle  $k \in \mathbb{N}_0$  und  $\mathbf{y} \in K_0$ , wobei  $K, K_0 \subset D$  kompakt. Dann gibt es  $C > 0$  mit

$$|H((\Psi^h)^k \mathbf{y}_0) - H(\mathbf{y}_0)| \leq C(hk \exp(-\gamma/h) + h^p) \quad \forall h \text{ hinreichend klein, } \forall \mathbf{y}_0 \in K_0 .$$



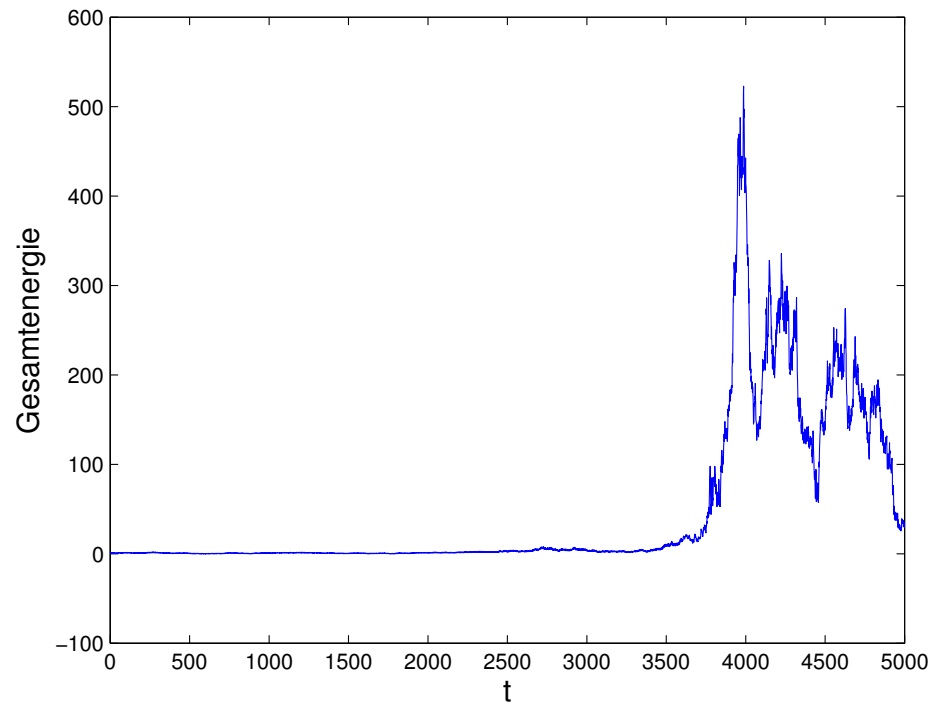
$$T \lesssim \exp(O(h^{-1})) \implies H(\mathbf{y}_k) - H(\mathbf{y}_0) = O(h^p)$$

Sect. 4.4.3": Methode der Rückwärtsanalyse erfordert *uniforme Zeitschrittweite*.

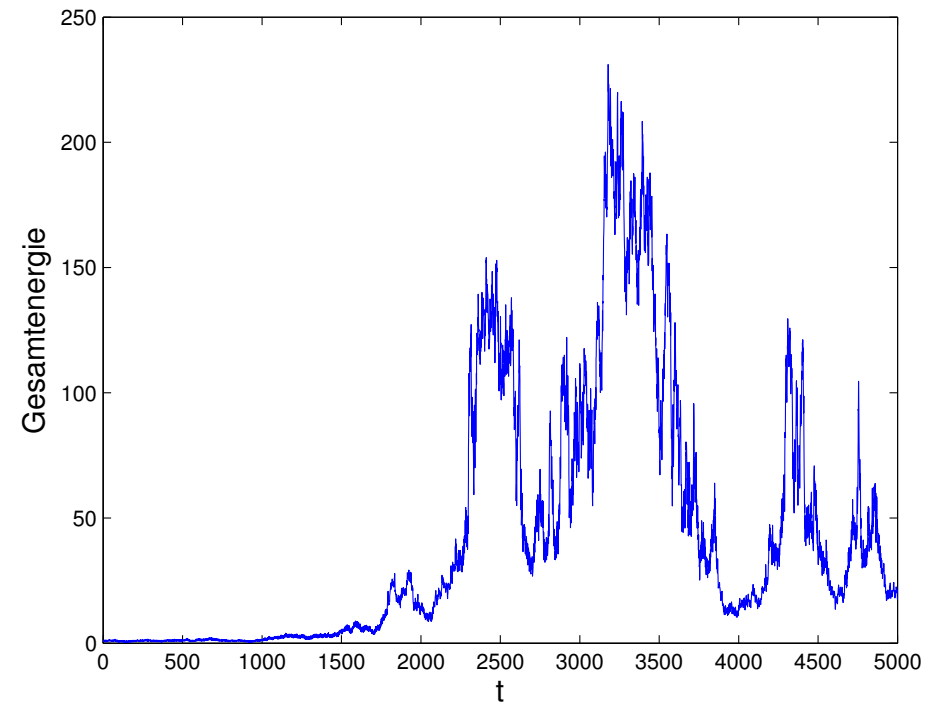
Eine bloss theoretische Einschränkung ?

*Beispiel 4.4.92* (Symplektische Integratoren und variable Schrittweite). Fortsetzung Bsp. 4.4.33

Symplektisches Eulerverfahren (4.4.26) für (4.4.4) auf  $[0, T]$ ,  $T = 5000$ . Erratische variable Schrittweite  $h_i = 0.5(1 + 0.5(\text{rand}() - 0.5))$ ,  $i = 1, \dots, 10000$ ,  $p(0) = 0$ ,  $q(0) = 7\pi/6$



Energiedrift bei variabler Schrittweite (Verfahren (4.4.26), links)



Energiedrift bei variabler Schrittweite (Verfahren (4.4.26), rechts)



# 4.5 Methoden für oszillatorische Differentialgleichungen [23]

Prototyp:

$$\dot{\ddot{y}} = -\omega^2 y \quad , \quad y(0) = y_0, \dot{y}(0) = v_0$$

$$\blacktriangleright \quad y(t) = \alpha \cos(\omega t) + \beta \sin(\omega t) \quad , \quad \alpha, \beta \in \mathbb{R}$$

Verallgemeinerung (skalar):

$$\ddot{y} = -\omega^2 y + g(y) \quad , \quad y(0) = y_0, \dot{y}(0) = v_0 \quad , \quad (4.5.1)$$

mit Lipschitz-stetiger **Störung**  $g : \mathbb{R} \mapsto \mathbb{R}$ .

Verallgemeinerung (vektoriell)

$$\ddot{\mathbf{y}} = -\mathbf{A}\mathbf{y} + g(\mathbf{y}) \quad , \quad \mathbf{y}(0) = \mathbf{y}_0, \dot{\mathbf{y}}(0) = \mathbf{v}_0 \quad , \quad (4.5.2)$$

$\mathbf{A} \in \mathbb{R}^{d,d}$  symmetrisch positiv definit,  $g : \mathbb{R}^d \mapsto \mathbb{R}^d$

*Bemerkung 4.5.3.*

$$(4.5.1) \quad \begin{matrix} v := \dot{y} \\ \iff \end{matrix} \quad \frac{d}{dt} \begin{pmatrix} y \\ v \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\omega^2 & 0 \end{pmatrix} \begin{pmatrix} y \\ v \end{pmatrix} + \begin{pmatrix} 0 \\ g(y) \end{pmatrix} \quad . \quad (4.5.4)$$

Lösung von (4.5.4) durch Variation der Konstanten:

$$\begin{pmatrix} y(t) \\ v(t) \end{pmatrix} = \begin{pmatrix} \cos t\omega & \omega^{-1} \sin t\omega \\ -\omega \sin t\omega & \cos t\omega \end{pmatrix} \begin{pmatrix} y_0 \\ v_0 \end{pmatrix} + \int_0^t \begin{pmatrix} \omega^{-1} \sin(t-s)\omega \\ \cos(t-s)\omega \end{pmatrix} g(y(s)) \, ds \quad (4.5.5)$$





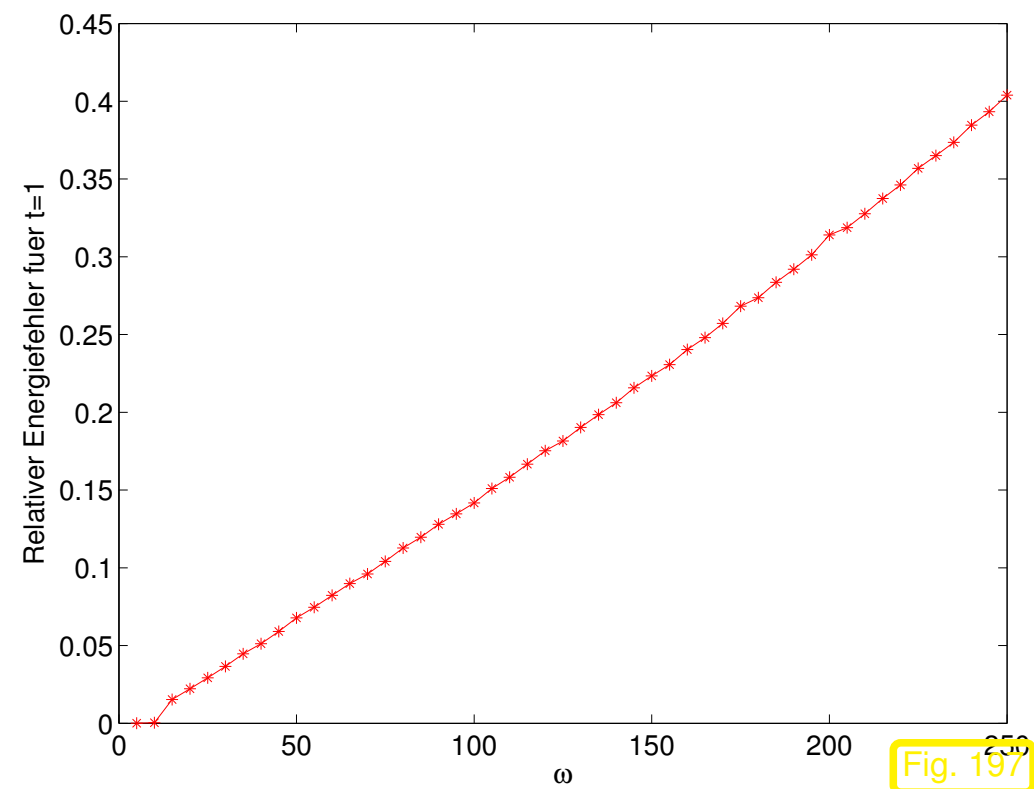
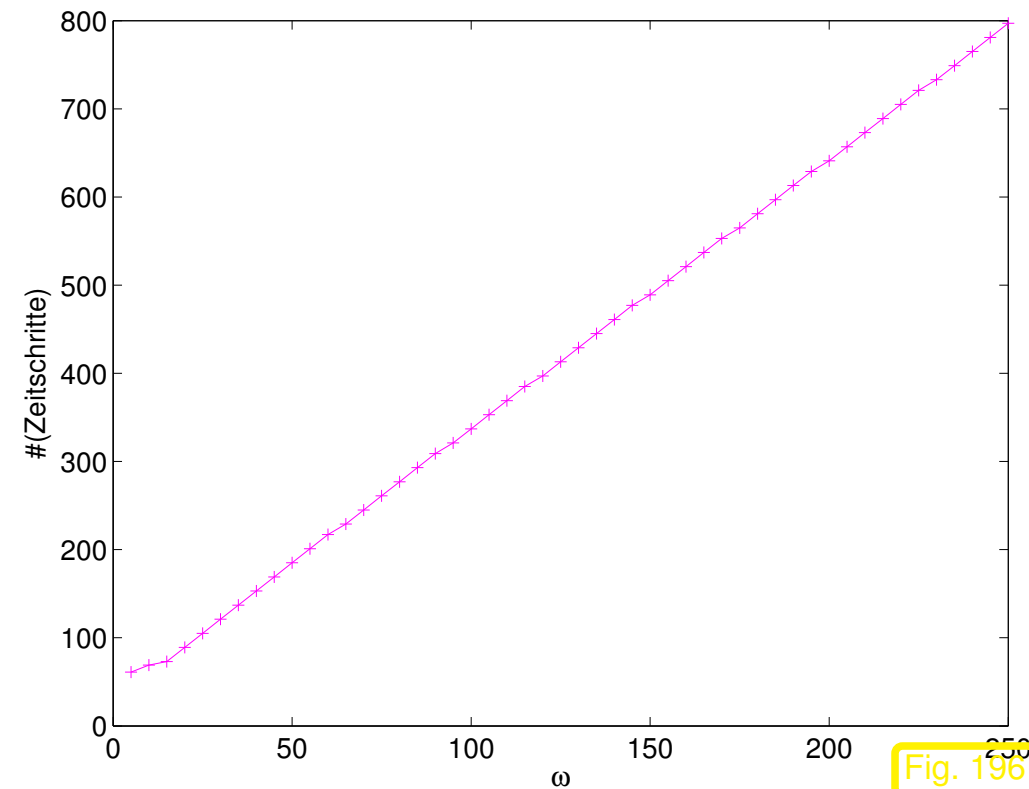
**Bemerkung 4.5.6.**  $y(t)$  löst (4.5.1) &  $G' = g \rightarrow \frac{1}{2}|\dot{y}|^2 + \frac{1}{2}\omega^2 y(t) - G(y(t)) \equiv \text{const.}$

„Energie“ für ODE (4.5.1)

**Beispiel 4.5.7** (Standardintegratoren für oszillatorische Differentialgleichung).

Adaptives explizites RK-ESV (Sect. 2.6 für (4.5.1):

```
y0=[1;0]; f=@(t,x) [0,1;-omega^2,0]*x + 20*[0;sin(x(1))];
options=odeset('reltol',1.0e-2,'abstol',1.0e-5);
[t,y]=ode45(f,[0,1],y0,options); („Energie“ ≐ Invariante aus Bem. 4.5.6)
```



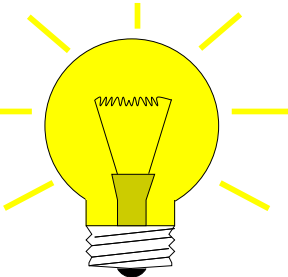
$\omega \uparrow \Rightarrow$  Oszillationen in  $y(t) \uparrow \Rightarrow$  Anzahl Zeitschritte  $\uparrow$



Ziel: Effiziente numerische Integration von (4.5.1)/(4.5.2) auch für  $\omega \gg 1$  bzw.  $\lambda_{\max}(\mathbf{A}) \gg 1$

Idee: ( wie bei exponentiellen Integratoren, siehe Sect. 3.7)

Verwende analytische Lösungsdarstellung (4.5.5) zur numerischen Integration:


$$y(t \pm h) = \cos(h\omega)y(t) \pm \frac{\sin h\omega}{\omega} \dot{y}(t) + \int_0^{\pm h} \frac{\sin(\pm h-s)\omega}{\omega} \cdot g(y(t+s)) ds \quad (4.5.8)$$

$g \equiv \text{const.}$ , (4.5.8)  $\blacktriangleright$   $y(t+h) - 2\cos(h\omega)y(t) + y(t-h) = h^2 \left( \frac{\sin(\frac{1}{2}h\omega)}{\frac{1}{2}h\omega} \right)^2 g$ . (4.5.9)

➤ **Gautschis Zweischrittverfahren** ( $y_h(t+h)$  aus  $y_h(t), y_h(t-h)$ ) für (4.5.1)

$$y_h(t+h) - 2\cos(h\omega)y_h(t) + y_h(t-h) = h^2 \left( \frac{\sin(\frac{1}{2}h\omega)}{\frac{1}{2}h\omega} \right)^2 g(y_h(t)). \quad (4.5.10)$$

Notwendig: **Startschritt** aus (4.5.5)

$$y_h(h) = \cos(h\omega)y_0 + \frac{\sin h\omega}{\omega}v_0 + \frac{1}{2}h^2 \left( \frac{\sin(\frac{1}{2}h\omega)}{\frac{1}{2}h\omega} \right)^2 g(y_0). \quad (4.5.11)$$

Ableitungsnaherung: Aus (4.5.8) für  $g \equiv \text{const}$ :

$$\blacktriangleright \quad y(t+h) - y(t-h) = 2h \frac{\sin h\omega}{h\omega} \dot{y}(t) \quad \Rightarrow \quad v_h(t) = \frac{h\omega}{\sin h\omega} \cdot \frac{y_h(t+h) - y_h(t-h)}{2h}. \quad (4.5.12)$$

*Bemerkung* 4.5.13. Gautschi-Verfahren (4.5.10), (4.5.11) für vektorielles Problem (4.5.2) ?

Ersetze  $\cos(h\omega) \mapsto \cos h\mathbf{A}$ ,  $\left( \frac{\sin(\frac{1}{2}h\omega)}{\frac{1}{2}h\omega} \right)^2 \mapsto 4(h\mathbf{A})^{-2} \sin^2(\frac{1}{2}h\mathbf{A})$ .



*Beispiel* 4.5.14 (Gautschis Zweischnittverfahren).

Anfangswertproblem vom Typ (4.5.1) auf  $[0, 1]$ :

$$\ddot{y} = -\omega^2 y + \sin y, \quad y(0) = 1, \quad \dot{y}(0) = 0.$$

Gautschi-Verfahren:  $h = 0.1000, \omega = 25$

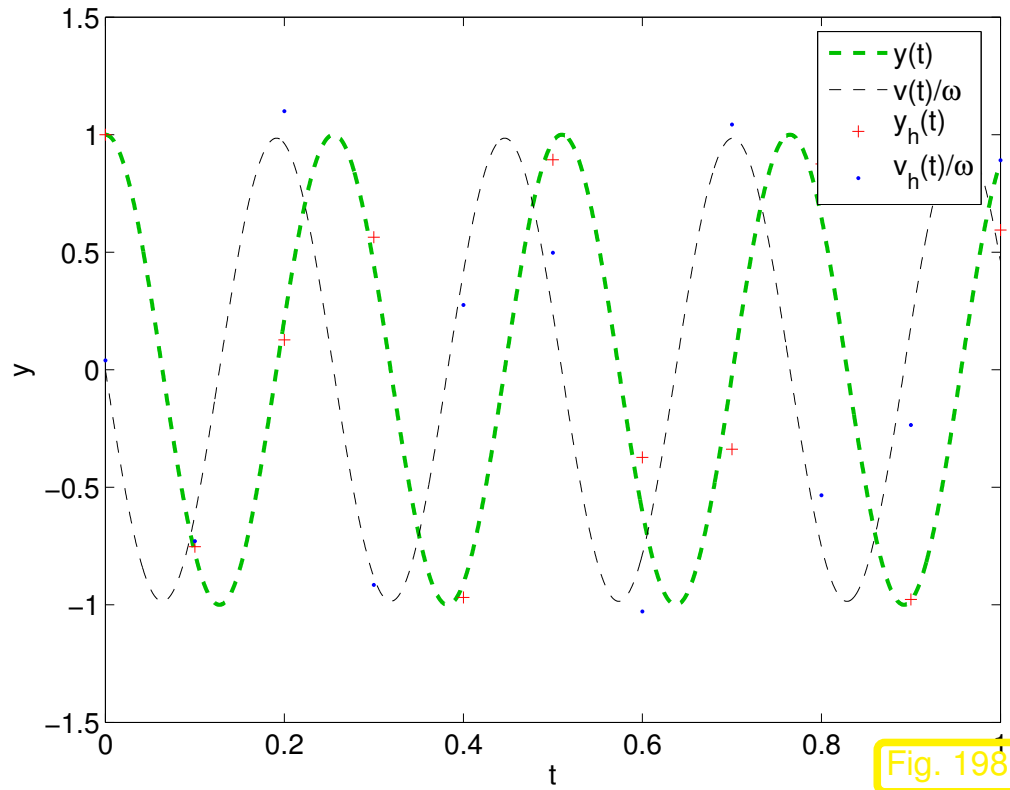


Fig. 198

Gautschi-Verfahren:  $h = 0.0333, \omega = 25$

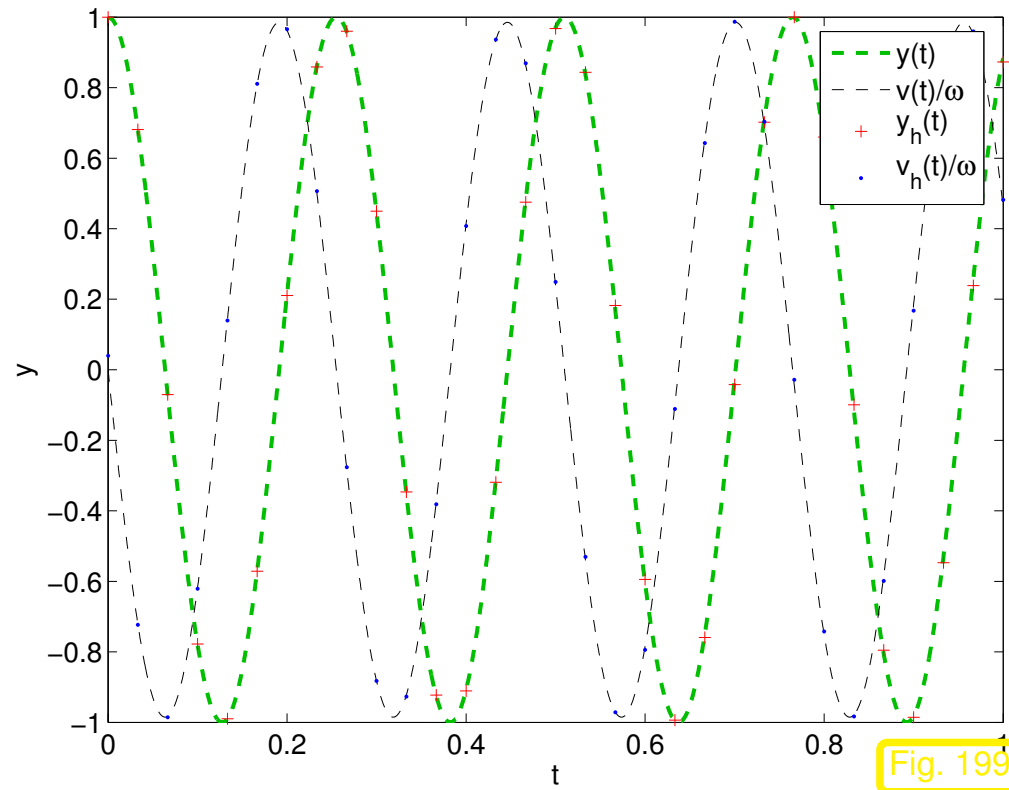


Fig. 199

$\omega = 25$ :  $y_h$  folgt (oszillatorischer Lösung), auch wenn  $h \approx \frac{2\pi}{\omega}$

Relativer Fehler in „Energie“ ( $\rightarrow$  Bem. 4.5.6) für  $t = 1$ :



Gautschi-Verfahren:  $h = 0.1000, \omega = 25$

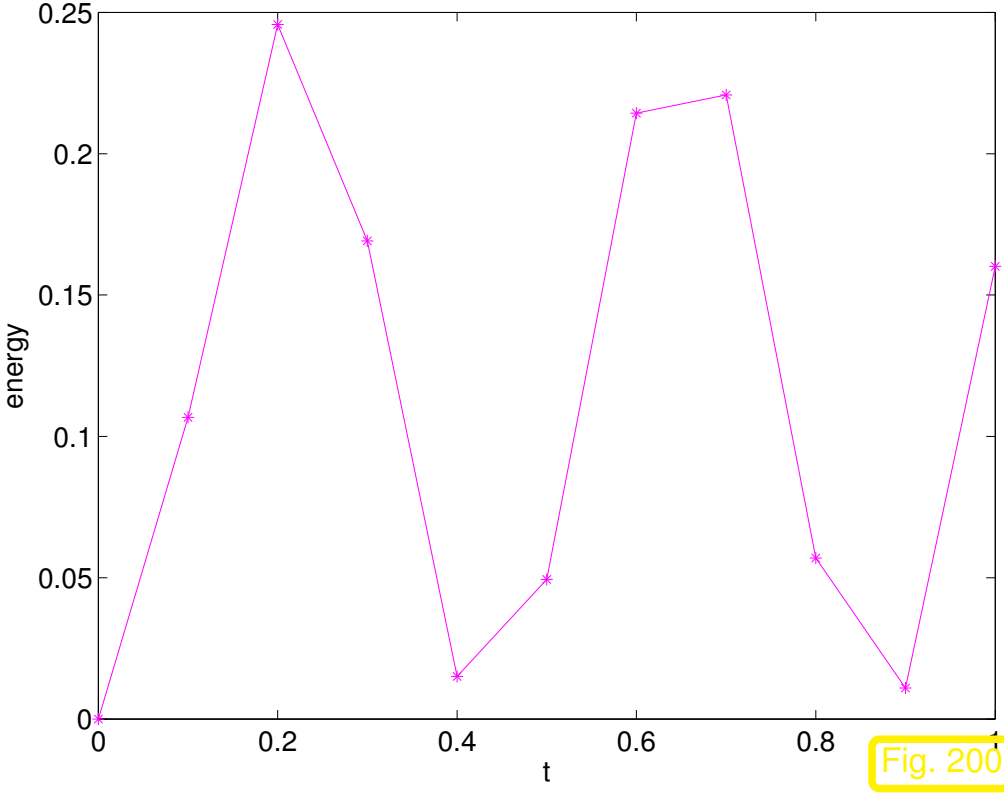


Fig. 200

Zeitschritt  $h = 0.1$

Gautschi-Verfahren:  $h = 0.0333, \omega = 25$

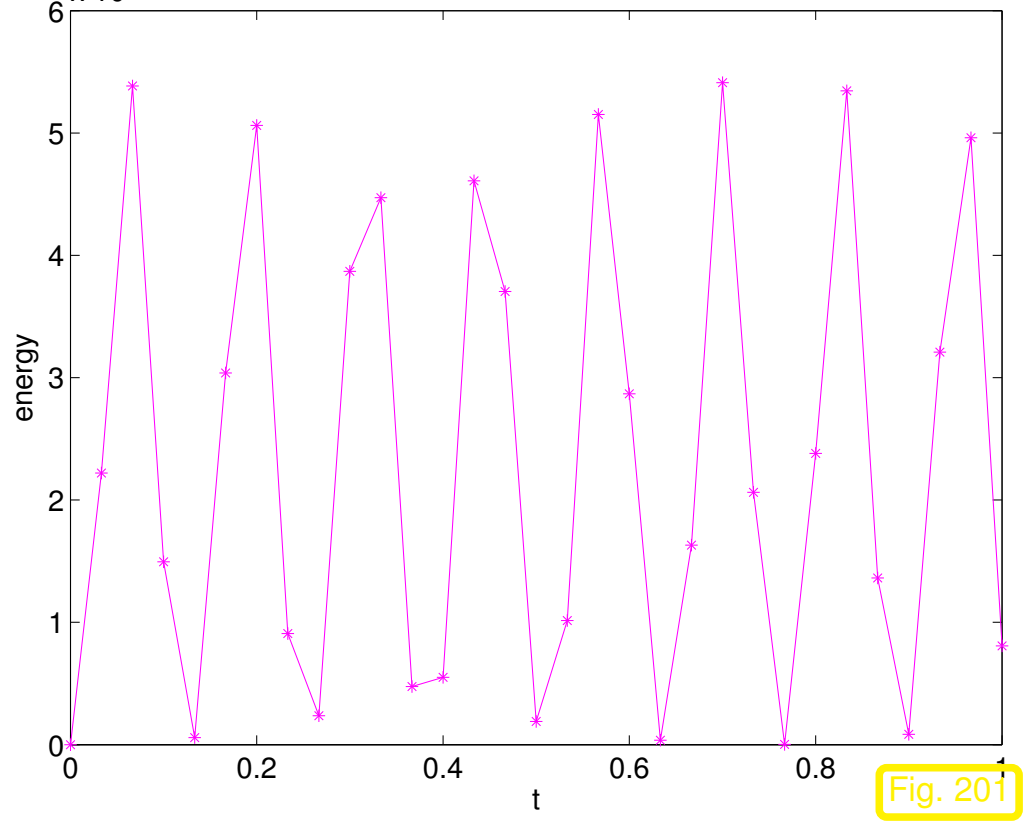
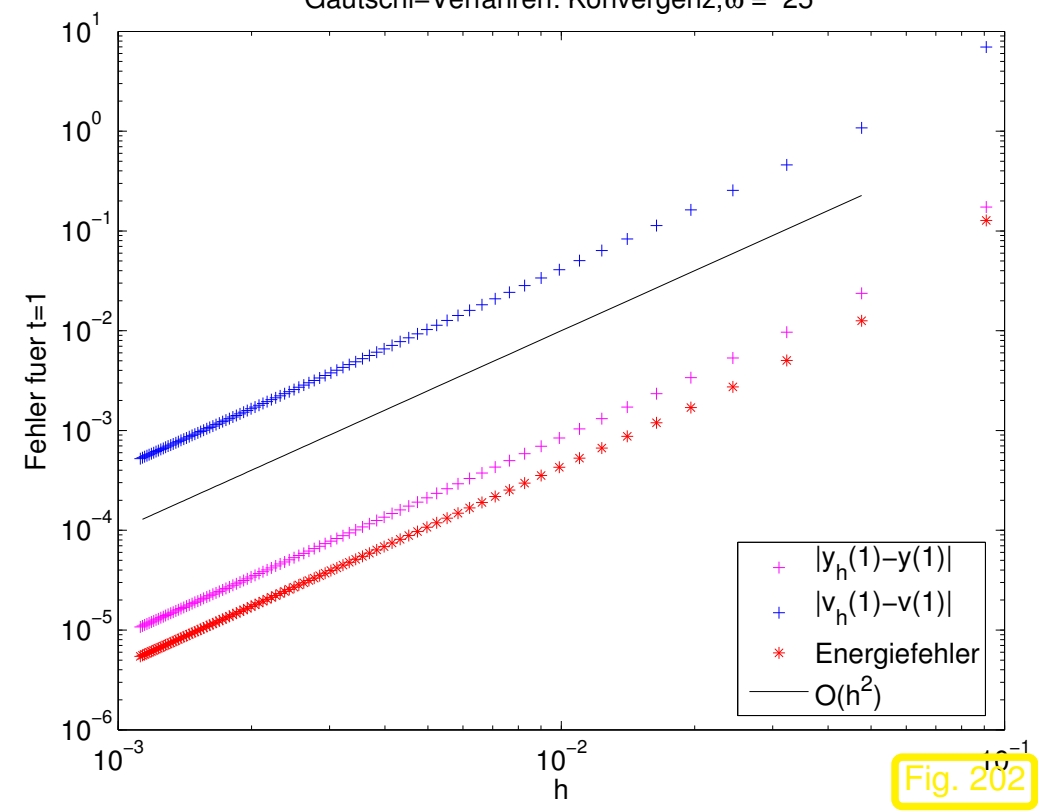


Fig. 201

Zeitschritt  $h = 0.033$

Gautschi-Verfahren: Konvergenz,  $\omega = 25$



Beobachtung:

Alle Fehler  $\approx O(h^2)$

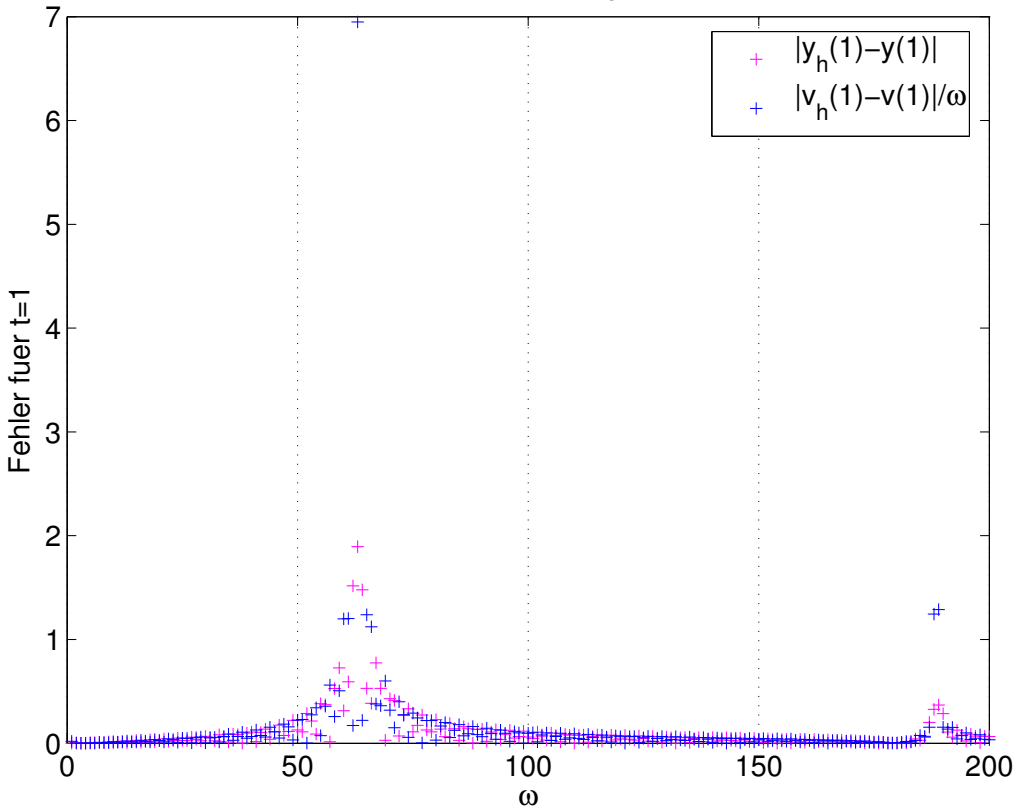


Konvergenzordnung 2

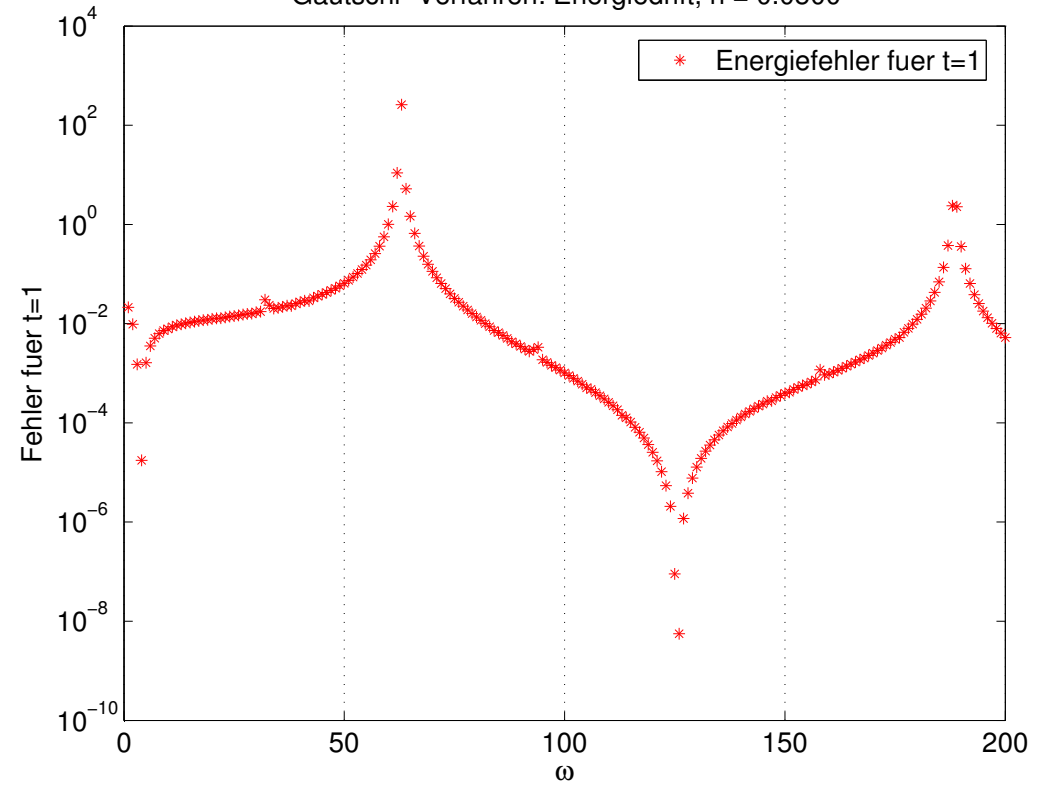
Ziel erreicht ?

Fehler für fixes  $h$  in Abhängigkeit von  $\omega$ :

Gautschi-Verfahren: Konvergenz,  $h = 0.0500$



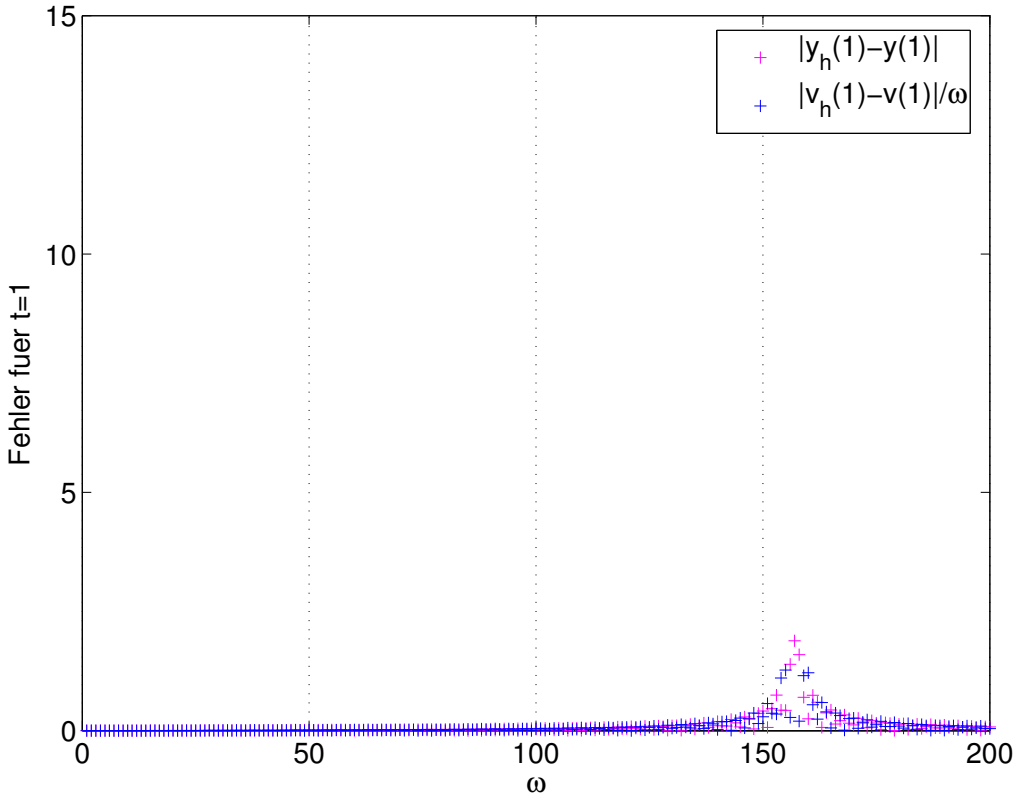
Gautschi-Verfahren: Energiedrift,  $h = 0.0500$



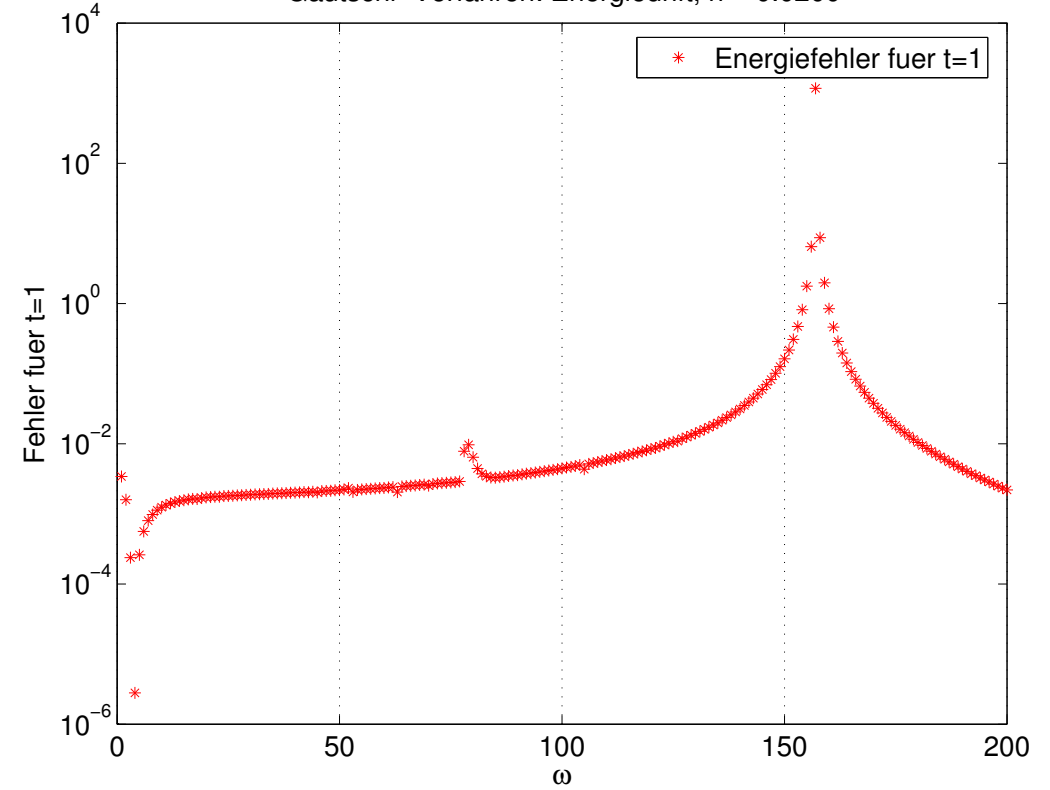
$h = 0.05$ : Was passiert für  $\omega \approx 61$ ,  $\omega \approx 123$ ,  $\omega \approx 185$  ?

Instabilität ?

Gautschi-Verfahren: Konvergenz,  $h = 0.0200$



Gautschi-Verfahren: Energiedrift,  $h = 0.0200$



$h$ -Abhängigkeit kritischer Frequenzen

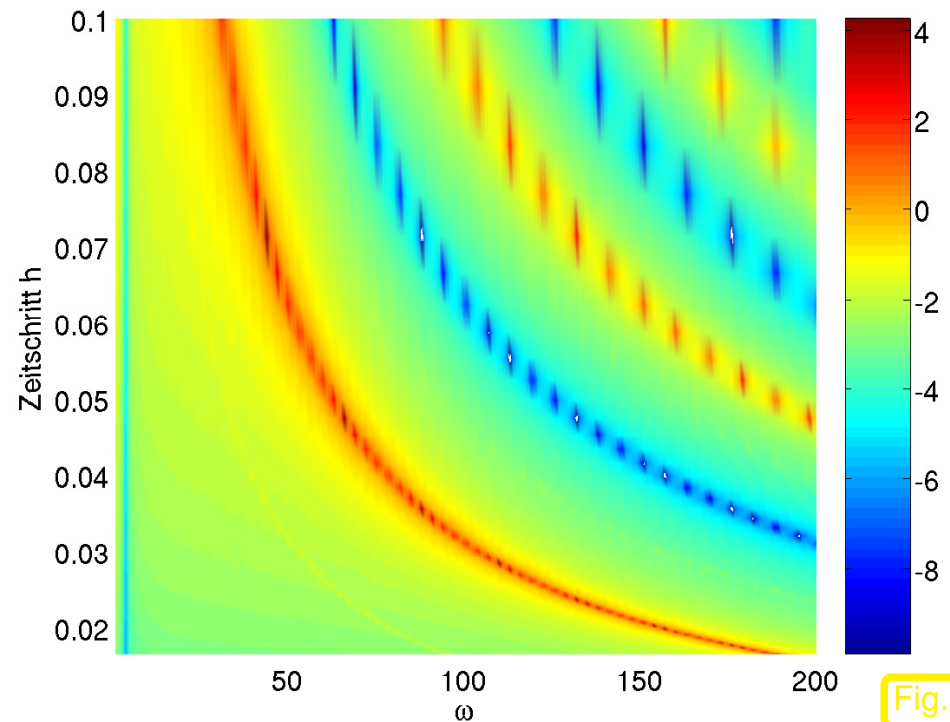
  
 $h$ -Abhängigkeit kritischer Frequenzen

---

Logarithmus  $\log_{10}$  des relativen Energiefehlers zum Endzeitpunkt  $t = 0$   $\triangleright$

Beobachtung:

$h\omega$ -Abhängigkeit kritischer Frequenzen

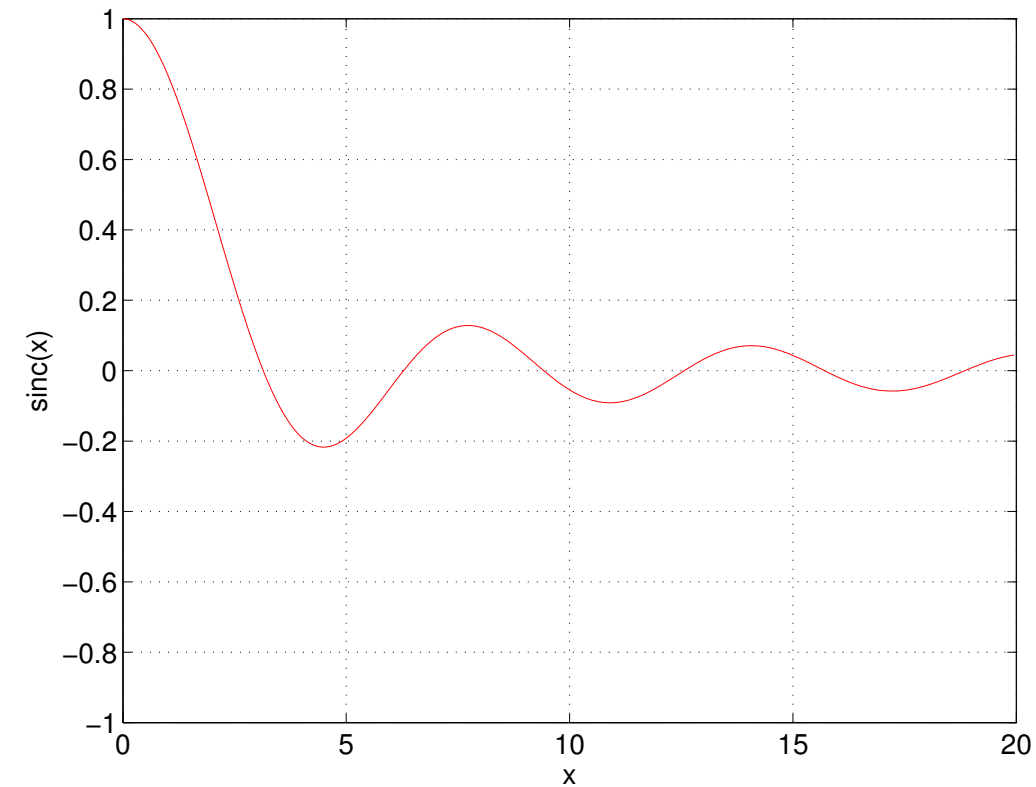


Modellproblem:

$$\ddot{y} = -\omega^2 y + \alpha y, \quad \alpha \ll \omega^2. \quad (4.5.15)$$

$\blacktriangleright$  Gautschi-Verfahren (4.5.10), Schrittweite  $h$ :  $\blacktriangleright$  **Dreitermrekursion**

$$y_h(t+h) - \left\{ 2 \cos(h\omega) + h^2 \alpha \operatorname{sinc}^2\left(\frac{1}{2}h\omega\right) \right\} y_h(t) + y_h(t-h) = 0. \quad (4.5.16)$$



Die sinc-Funktion:

$$\operatorname{sinc}(x) := \frac{\sin x}{x}$$

- ▶ Analytisch auf  $\mathbb{R}$
- ▶  $|\operatorname{sinc}(x)| \leq 1$  mit globalem Maximum in  $x = 0$

Analyse von (4.5.16): Ansatz  $y_h(kh) = \xi^k \Leftrightarrow$  Charakteristische (quadratische) Gleichung

$$\exists \text{ Lösungen } y_h(kh) \text{ von (4.5.16): } \lim_{k \rightarrow \infty} y_h(hk) = \pm\infty \Leftrightarrow \left| 2 \cos(h\omega) + h^2 \alpha \operatorname{sinc}^2\left(\frac{1}{2}h\omega\right) \right| > 2$$

▶  $(\cos h\omega \approx 1 \Leftrightarrow h\omega \approx 2\pi l, l \in \mathbb{Z}) \Rightarrow y_h(hk) \rightarrow \pm\infty$  auch für  $h \ll 1$ .

Abhilfe [23]: „Filterung“: Dämpfung von  $\alpha$ , falls  $h\omega \approx 2\pi l$ :

In (4.5.10), (4.5.11) ersetze:  $g(y_h(t)) \mapsto g(\psi(h\omega)y_h(t))$ ,  $\psi(\xi) := \text{sinc}^2 \xi (1 + \frac{1}{2}(1 - \cos \xi))$

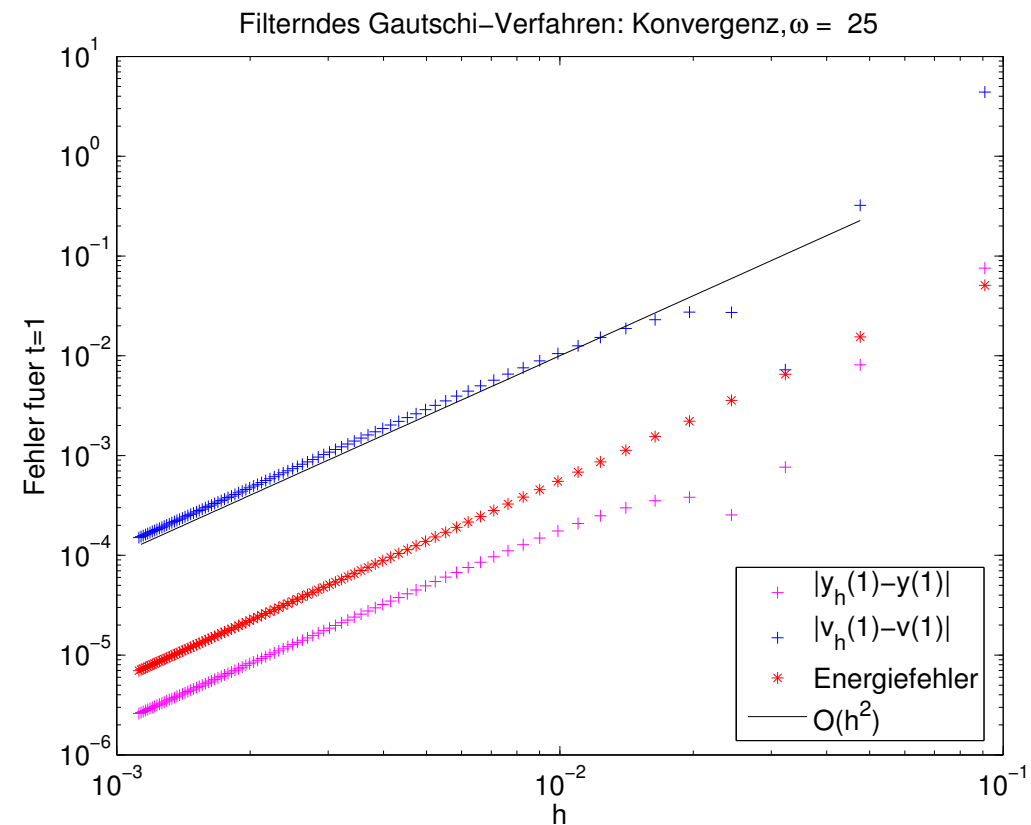
► Modifiziertes Gautschi-Verfahren:

$$y_h(t+h) - 2 \cos(h\omega)y_h(t) + y_h(t-h) = h^2 \left( \frac{\sin(\frac{1}{2}h\omega)}{\frac{1}{2}h\omega} \right)^2 g(\psi(h\omega)y_h(t)). \quad (4.5.17)$$

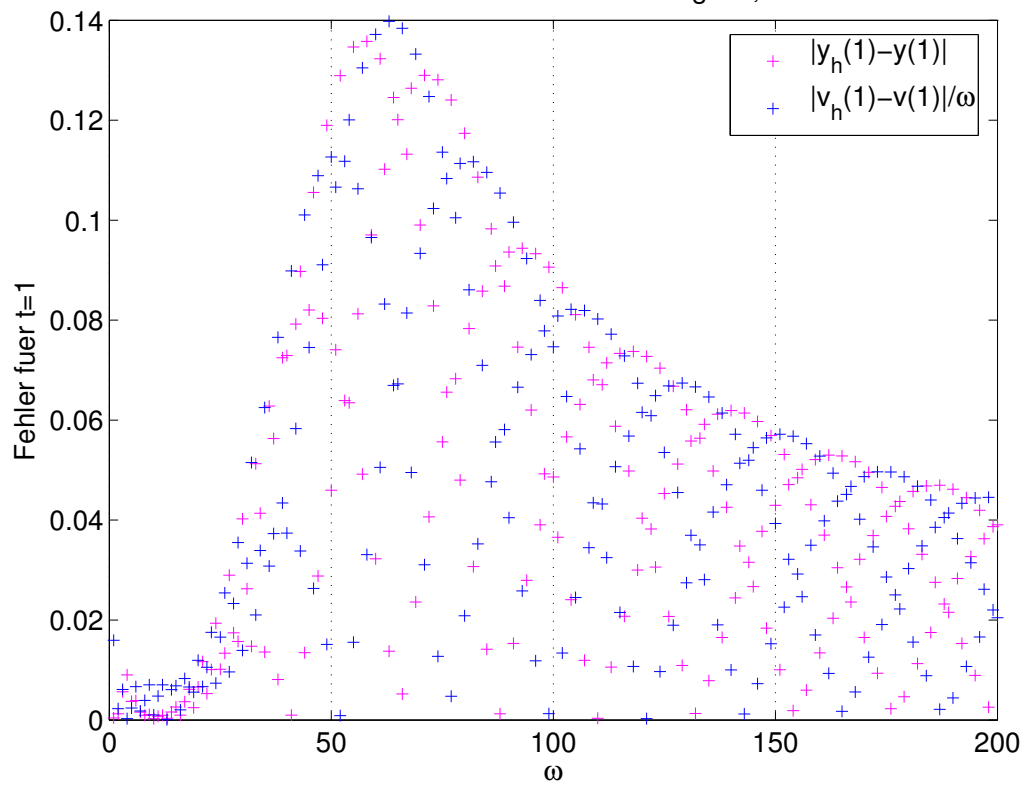
*Beispiel* 4.5.18 (Modifiziertes Gautschi-Verfahren).

AWP aus Bsp. 4.5.10, Integration gemäss (4.5.17),  
Filterfunktion  $\psi(\xi) := \text{sinc}^2 \xi (1 + \frac{1}{2}(1 - \cos \xi))$

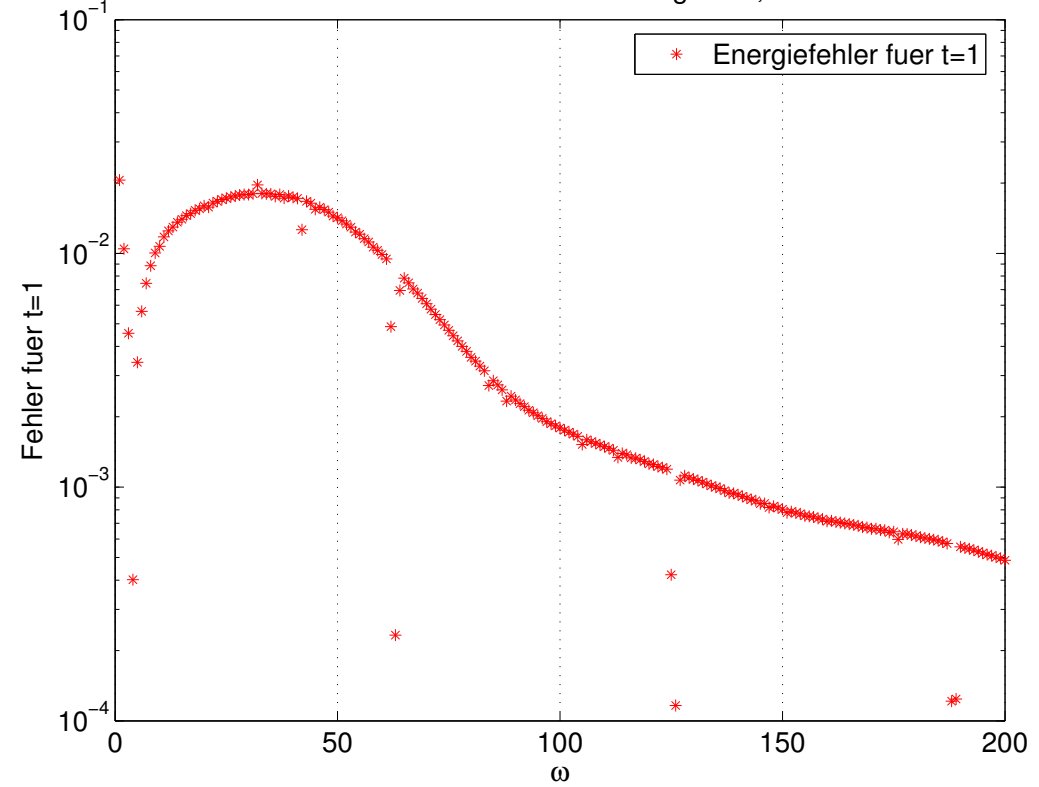
Konvergenzordnung 2



Filterndes Gautschi-Verfahren: Konvergenz,  $h = 0.0500$



Filterndes Gautschi-Verfahren: Energiedrift,  $h = 0.0500$



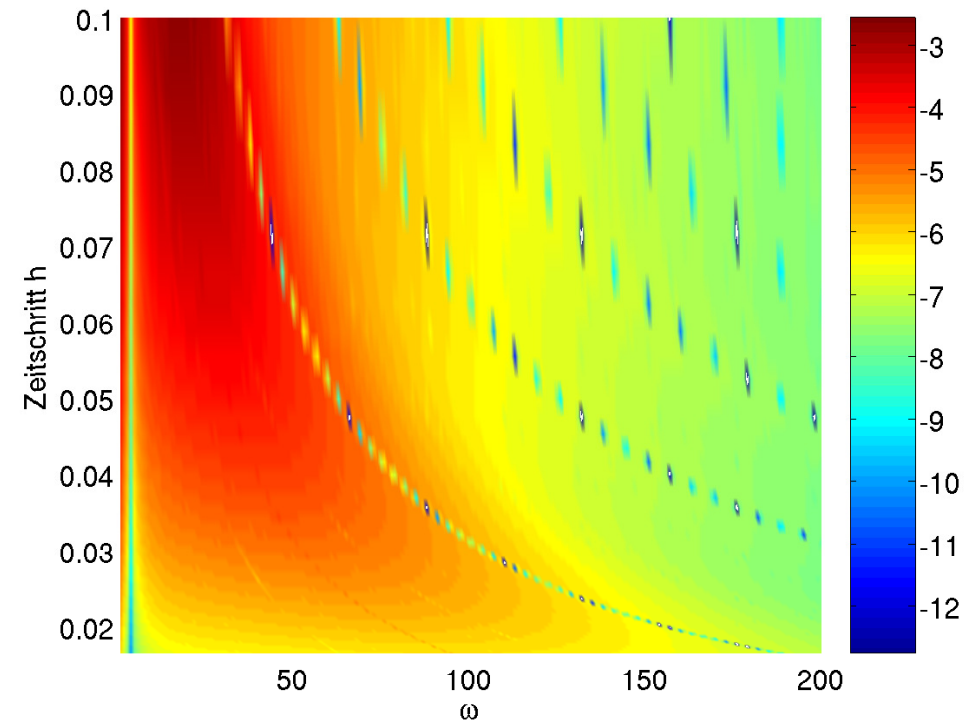


$h\omega$ -Abhängigkeit der Energiedrift

Beobachtung:

(4.5.17): • Keine Instabilität

- Im Vergleich zu (4.5.10) ( $\rightarrow$  Bsp. 4.5.14) deutlich reduzierte Energiedrift (Skala!)



# Verzeichnisse

R. Hiptmair

rev 35327,  
22. Januar  
2008

# Index

- R-reversible Abbildung, 462
- R-reversible Evolution, 462
- a posteriori
  - Fehlerschätzung, 292
- A-Stabilität, 358
- absolute Toleranz, 295
- adjungierte Matrix, 443
- Affin-Kovarianz
  - Runge-Kutta, 230
- Aitken-Neville-Schema, 254
- akzeptable Lösung, 551
- akzeptables Resultat, 508
- algebraische Konvergenz, 78, 79, 122
- algebraische Nebenbedingung
  - bei DAE, 405
- algebraische Stabilität, 362
- alternierende Bilinearform, 473
- analytisch, 195
- analytische Funktion, 192, 524
- Analytizitätsvoraussetzung, 524
- Anfangswertproblem
  - Lösung, 21
  - linear, 52
  - steifes, 375
- Asymptotische (absolute) Kondition, 58
- asymptotische Entwicklung, 256, 260, 517
- Asymptotische Stabilität
  - eines Fixpunkts, 340
- attraktiver Fixpunkt, 24, 322
- autonome Differentialgleichung, 18
- Autonomisierung, 18
- Autonomisierungsinvarianz, 232
- AWP
  - Kondition, 60
- B-Stabilität, 362
- Bahnebene, 41
- Banachscher Fixpunktsatz, 150
- Basisverfahren
  - für Extrapolation, 266
- Bewegungsgleichungen
  - Hamiltonsche Form, 38
  - Molekulardynamik, 498

Newtonsche, 37  
bimolekulare Reaktion, 29  
Blow-up, 44, 51, 310  
Bootstrapping, 222  
Butcher-Bäum, 244  
Butcher-Matrix, 368  
Butcher-Schema, 224  
Cauchy-Hadamard  
  Formel von, 205  
DAE, 404, 405  
  separiert, 406  
  vom Index 1, 408  
Deskriptorform  
  mechanischer Bewegungsgleichungen, 419  
  von Bewegungsgleichungen, 420  
Determinante  
  Ableitung, 443  
diagonal-implizite ESV, 391  
DIFEX, 278  
Differentialgleichung  
  Hamiltonsche, 39  
  linear, 53  
  logistische, 186, 209, 250  
  Variation der Konstanten, 54  
Differentialgleichungen  
  Skalare, 50  
differenziell-algebraische Gleichung (DAE), 404  
differenziell-algebraisches Anfangswertproblem (DAE), 405,  
  406  
differenzielle Konditionsanalyse, 58

Differenzenverfahren, 75, 87, 100  
DIRK-Einschrittverfahren, 391  
diskrete Evolution  
  Konsistenz, 125  
  Konsistenzfehler, 125  
  Konsistenzordnung, 127  
  reversibel, 140  
diskretes dynamisches System, 347  
Diskretisierungsfehler, 121  
Diskretisierungsparameter, 252  
dissipatives Vektorfeld, 356  
Divergenz  
  eines Vektorfeldes, 447  
Doppelpendel, 70  
Drehimpuls, 41  
Dreitermrekursion, 565  
dynamisches System  
  diskret, 347  
Eingebettete Runge-Kutta-Verfahren, 310  
eingebettete Runge-Kutta-Verfahren, 311  
Einschrittfehler, 139  
Einschrittverfahren  
  implizit, 121  
Einschrittverfahren, 105, 119  
  diagonal-implizit, 391  
  explizit, 121  
  Konvergenz, 130  
  Notation, 120  
  reversibel, 274  
  Schrittweitensteuerung, 287  
  symplektisch, 483

elementare Differentiale, 243  
 Energie  
   für oszillatorische Differentialgleichung, 557  
 Energiedrift, 104, 113, 468, 555, 569  
 Energieerhaltung, 39, 467  
   Pendel, 38  
 Energiemannigfaltigkeit, 505  
 Erstes Integral  
   Bedingung, 29  
 erstes Integral, 29, 434  
   linear, 435  
   polynomial, 435  
   quadratisch, 102, 435  
 erweiterter Zustandsraum  
   einer ODE, 12  
 erzeugende Funktion, 199  
 ESV  
   Radau, Ordnung 3, 371  
   Radau, Ordnung 5, 371  
 Euler  
   Implizit, 371  
 Euler-Verfahren, 267  
   explizites, Stabilitätsfunktion, 330  
   implizit, 85  
   semi-implizit, 387  
 Euler-Verfahrens  
   Konvergenz, 75  
 Eulersches Polygonzugverfahren, 73  
 Eulerverfahren  
   explizit, 73, 161  
   implizit, 161  
   implizites, Stabilitätsfunktion, 330  
 Evolution  
   R-reversibel, 462  
   diskrete, 146  
 Evolutionsoperator, 49  
 explizite Mittelpunktsregel, 245  
 explizite Trapezregel, 245  
 explizites Einschrittverfahren, 121  
 explizites Eulerverfahren, 73, 161  
 exponentiell klein, 522  
 exponentielle Konvergenz, 79, 187  
 exponentielle Runge-Kutta-Verfahren, 401  
 Extrapolation, 252  
 Extrapolations-Einschrittverfahren, 266  
 Extrapolationstableau, 254  
 Extrapolationsverfahren  
   global, 264  
   lokal, 265  
 Faltung, 55  
 Federpendel, 494  
 Fehlerfortpflanzung, 139  
 Fehlerfunktion, 138  
 Fixpunkt  
   asymptotisch stabil, 340  
   asymptotische Stabilität, 345  
   attraktiv, 24, 84, 322, 340  
   einer ODE, 339  
   repulsiv, 24  
 Gauss-Kollokations-Einschrittverfahren, 162, 185, 245, 356,  
   374, 486

Gauss-Radau-Quadratur, 369  
Gaussquadratur, 161  
Gautschi-Verfahren, 558  
    Filterung, 567  
    modifiziertes, 567  
Gautschis Zweischrittverfahren, 558  
gewöhnliche Differentialgleichung, 12  
    autonome, 18  
    erster Ordnung, 12  
Gewichte  
    einer Quadraturformel, 160  
Gitterfunktion, 121  
Glattheit  
    hinreichende, 130  
globale Lösung  
    eines AWP, 48  
globale Lipschitzbedingung, 61  
Gradientenfluss, 353  
Grenzyklus, 381  
Gronwalls Lemma, 62  
Hamilton-Funktion, 39, 466  
    Molekulardynamik, 497  
    separiert, 484  
Hamiltonsche Bewegungsgleichungen  
    mit Nebenbedingung, 422  
Hamiltonsche Differentialgleichung, 39, 466  
Herzschlagmodell, 32  
hinreichende Glattheit, 130  
holomorph, 192, 195  
holomorphe Funktion, 524  
homogene lineare Differentialgleichung, 54

implizite Mittelpunktsregel, 99, 128, 161, 484  
implizites Einschrittverfahren, 121  
implizites Euler-Verfahren, 85  
implizites, Euler-Verfahren, 245  
Impuls, 41  
Index  
    einer DAE, 408, 421  
inkompressible Strömung, 446  
Inkrement  
    Kollokation, 147  
    Runge-Kutta, 224  
Inkrementfunktion, 125  
Inkrementgleichungen, 147  
    linearisiert, 388  
Instabilität  
    Gautschi-Verfahren, 563  
Integrabilitätslemma, 481  
intervallweise Kondition, 60  
Invariante, 29  
invariante Mannigfaltigkeit, 382  
Jordan-Block, 342  
Jordan-Normalform, 342  
Joukowski-Transformation, 201  
Keplerproblem, 40  
Keplersches Gesetz  
    erstes, 42  
    zweites, 42  
kinetische Energie, 38  
klassisches Runge-Kutta-Verfahren, 245  
Knoten

- einer Quadraturformel, 160
- Knotenanalyse
  - von Schaltkreisen, 380, 402
- Kollaps, 44, 51, 309
- Kollokation
  - Inkremente, 147
- Kollokations
  - RK-ESV, 369
- Kollokationsbedingung, 144
- Kollokationspunkt, 144
- Kollokationsverfahren, 144
  - Inkrementfunktion, 147
  - Konsistenz, 180
- Kompaktheitsargument, 135
- Kondition, 57
  - analyse
    - differentielle, 63
    - asymptotisch, 58
    - intervallweise, 60
    - punktweise, 60
- Kongruenztransformation, 473
- Konsistenz, 125
  - Runge-Kutta-Verfahren, 239
- Konsistenzfehler, 125, 131, 138
- Konsistenzordnung, 127
  - Splittingverfahren, 280
- Kontraktion, 150
- Konvergenz, 122
  - algebraisch, 78
  - exponentiell, 187
  - Kollokationsverfahren, 185
  - von Einschrittverfahren, 130
- Konvergenzordnung, 122
- Konvergenzradius, 205
- kovariante Transformation, 53
- Kovergenz
  - global, 122
- Kraftfeld
  - konservativ, 40
- Kreuzprodukt (Vektorprodukt), 436
- Kuttas 3/8-Regel, 245
- L-Stabilität, 367
- Lösung
  - eines Anfangswertproblems, 21
- Lagrange-Multiplikator, 420, 422
- Lagrange-Polynom, 145
- Laurent-Entwicklung, 193
- Legendre-Polynom, 197
- Legendre-Polynome, 162
  - Rekursionsformel, 197
- Lenard-Jones-Potential, 497
- Lie-Trotter-Splitting, 279
- linear-implizites Runge-Kutta-Verfahren, 393
- lineare Differentialgleichung, 52, 53
- linearer Operator
  - stetig, 164
- linearisierte Störungstheorie, 58
- Linearisierung
  - um Fixpunkt, 322
- Liouville
  - Satz von, 447
- Lipschitz-Stetigkeit

lokale, 46  
Logistische Differentialgleichung, 186, 209, 250, 279  
logistische Differentialgleichung, 23, 268  
Lotka-Volterra Differentialgleichung, 26  
Makroschritt  
  bei Extrapolationsverfahren, 265  
MAPLE, 128  
mathematisches Pendel, 37  
  symplektische Integration, 467  
Matrix  
  Propagations-, 64  
  Wronski, 64  
Matrixexponentialfunktion, 54, 397  
Matrixfunktionen, 337  
maximale Fortsetzbarkeit, 44  
maximales Existenzintervall, 45  
Mikroschritt  
  bei Extrapolationsverfahren, 266  
Minimalkoordinaten, 420  
Mittelpunktsregel, 223  
  explizit, 223, 227  
  implizit, 99, 161  
  implizit, Stabilitätsfunktion, 330  
Modellprobelanalyse  
  implizites Euler-Verfahren, 88  
Modellproblem  
  für gestörte oszillatorische Differentialgleichungen, 565  
Modellproblemanalyse, 322  
  explizites Eulerverfahren, 84  
modifizierte Differentialgleichung, 508  
  für lineare AWP, 509

modifizierte Gleichung  
  abgeschnittene, 519  
  der Ordnung  $q$ , 510  
Molekulardynamik, 497  
Molekulardynamik, 501  
implizite Trapezregel, 245  
multivariates Polynom, 435  
Newton-Verfahren  
  vereinfacht, 392  
Newtonsche Bewegungsgleichungen, 37  
nichtdegenerierte Bilinearform, 473  
Nichtexpansivität, 353  
nichtlineare Stabilität, 140  
Normalform  
  bei schiefsymmetrischen Matrizen, 473  
numerische Quadratur, 160  
numerischer Integrator, 116  
ODE, 12  
  skalar, 15  
Operatornorm, 164  
ordinary differential equation (ODE), 12  
Ordnung  
  einer Quadraturformel, 181  
Ordnungsschranken  
  für Runge-Kutta-Verfahren, 245  
Ordnungssteuerung  
  bei Extrapolationsverfahren, 271  
Oregonator, 31  
oszillatorische Differentialgleichungen, 556  
parasitäre Kapazität, 409



partikuläre Lösung, 54  
partitionierte Runge-Kutta-Einschrittverfahren, 489  
Peano  
  Satz von, 47  
Pendel, 37, 419  
Pendelgleichung, 469  
Phasenraum  
  einer ODE, 12  
Picard-Lindelöf  
  Satz von, 47  
Pol  
  einer Funktion, 193  
Polygonzugverfahren, 72  
Polynom  
  multivariat, 435  
polynomiale Invariante, 435  
Polynominterpolation  
  Fehlerabschätzung, 174  
potentielle Energie, 38  
Problem  
  in der Numerik, 57  
Projektions-Einschrittverfahren, 165  
Projektionsoperator, 164  
Propagationsmatrix, 64  
Punkt  
  stationär, 28  
punktweise Kondition, 60  
Push-Forward, 472  
Quadraturformel, 160  
  Mittelpunktsregel, 223  
  Ordnung, 181

Trapezregel, 223  
Räuber-Beute-Modell, 26  
Rückwärtsanalyse, 551  
  von Integrationsverfahren, 507  
Radau-ESV, Ordnung 3, 371  
Radau-ESV, Ordnung 5, 371  
Radau-Verfahren, 417  
Rationale Approximation  
  der Exponentialfunktion, 329  
Reaktionskinetik, 29  
rechte Seite  
  einer ODE, 12  
Regel  
  Kuttas 3/8, 230  
  Mittelpunkt, explizit, 223  
  Trapez, explizit, 223, 228  
relative Toleranz, 295  
repulsiver Fixpunkt, 24  
Residuensatz, 192  
Residuum  
  einer komplexwertigen Funktion, 193  
Reversibilität, 140, 274, 460  
reversible Einschrittverfahren  
  Konsistenzordnung, 142  
Riccati-Differentialgleichung, 14, 74  
Richtungsfeld, 14  
RK4, 229  
  Stabilitätsfunktion, 331  
Rodrigues-Formel, 197  
Romberg-Quadratur, 252  
ROW-Methoden, 393

Runge-Kutta  
   3/8-Regel, 230  
   Affin-Kovarianz, 230  
   Autonomisierung, 232  
   eingebettet, 311  
   Einschritt-Verfahren, 224  
   Inkrement, 224  
   klassisch, 229  
 Runge-Kutta-Einschrittverfahren  
   R-reversibel, 464  
   steif-genaue, 367  
   symmetrisch, 457  
 Runge-Kutta-Verfahren  
   Autonomisierungsinvarianz, 232  
   Ordnungsschranken, 245  
 Runge-Kutta-Verfahren, 222, 224  
   eingebettet, 310  
   exponentiell, 401  
   Konsistenz, 239  
   Konstruktion, 223  
   Konvergenzordnung, 245  
   linear-implizit, 393  
   Stabilitätsfunktion, 327  
 Satz  
   Peano & Picard-Lindelöf, 47  
 Satz über implizite Funktionen, 149  
 Satz von Liouville, 447  
 Schaltkreis  
   Knotenanalyse, 380, 402  
 Schrittweisenbeschränkung, 153, 349  
   für explizite RK-ESV, 334  
 Schrittweitenkorrektur, 300  
 Schrittweitensteuerung, 377  
   für ESV, 287  
 Schrittweitevorschlag, 300  
 Schur-Zerlegung, 336  
 semi-implizites Euler-Verfahren, 387  
 Sensitivität, 57  
 Separation der Variablen, 24  
 sinc-Funktion, 566  
 Singuläre Störungstechnik, 409  
 Skalare Differentialgleichungen, 50  
 Skalare ODE, 15  
 Spektralradius  
   einer Matrix, 347  
 Spektrum  
   einer Matrix, 342  
 Splitting  
   Lie-Trotter, 279  
   Strang, 279  
 Splittingverfahren, 278, 487  
   inexakt, 285  
   inexakte, 285  
 symmetrische Runge-Kutta-Einschrittverfahren, 457  
 Störmer-Verlet-Verfahren, 104, 484  
   Molekulardynamik, 498  
 Stabilität  
   nichtlineare, 140  
 Stabilität, 139  
   -sfunktion, 371  
   -sgebiet, 350  
   B-, 362

L-, 367  
 Stabilitätsfunktion, 330  
   Interpretation, 327  
   von Runge-Kutta-Verfahren, 327  
 Stabilitätsgebiet, 326  
 Startschritt, 106, 559  
 steif-genau, 367, 413  
 Steifheit, 375  
 sternförmig, 481  
 stetiger linearer Operator, 164  
 Strang-Splitting, 279  
 Stromlinien, 449  
 strukturerhaltende Integratoren, 508  
 Stufen  
   eines RK-ESV, 225  
 symplektische Abbildung, 478  
 symplektische Evolution, 472  
 symplektischer Fluss, 474  
 symplektischer Integrator, 485  
 symplektisches Einschrittverfahren, 483  
 symplektisches Euler-Verfahren, 484, 487  
 symplektisches Produkt, 472  
 Taylorentwicklung, 240  
 Toleranz  
   absolute, 295  
   bei Schrittweitensteuerung, 292  
   relativ, 295  
   Relativ, 295  
 Trajektorie, 26  
 Transformation  
   kovariant, 53  
 Trapezregel, 223  
   explizit, 223, 228  
   explizit, Stabilitätsfunktion, 330  
 Variationsgleichung, 65  
 Variation der Konstanten, 54, 396, 556  
 Variationsgleichung, 450  
 Vektorfeld, 18  
 Vektorprodukt, 42  
 Verfahren  
   ESV, Runge-Kutta, 224  
   Euler, implizit, 161  
   Runge-Kutta, 222, 224  
   Runge-Kutta, klassisch, 229  
   Runge-Kutta, Konstruktion, 223  
 versteckte Nebenbedingungen, 421  
   bei DAEs, 423  
 volumenerhaltende Abbildung, 447  
 Volumenerhaltung, 447  
   bei Hamiltonschen ODEs, 469  
 wohlgestellt, 61  
   Problem, 57  
 Wronski-Matrix, 64  
 Zeeman-Modell, 32  
 Zeitgitter, 119  
 Zeitschrittweite, 119  
 Zeitskalierungsinvarianz, 325  
 Zeitumkehrsymmetrie (bei mechanischen Systemen), 461  
 Zustandsraum  
   einer ODE, 12  
   Molekulardynamik, 497

Zwangskraft, 420  
Zweischrittverfahren, 105  
Gautschis, 558

# Beispiele und Bemerkungen

- [Lösung der Schaltkreis-DAEs mit MATLAB, 418
- [Steife Probleme in der chemischen Reaktionskinetik, 377
- ‘Gronwall-Schranke’ für Kondition, 63
- “Butcher barriers” für explizite RK-ESV, 246
- “Versagen” adaptive Zeitschrittsteuerung, 306
- A-Stabilität  $\nrightarrow$  Diskrete Nichtexpansivität, 361
- Adaptive explizite RK-ESV für steifes Problem, 376
- Adaptive RK-ESV zur Teilchenbahnberechnung, 314
- Adaptives semi-implizites RK-ESV für steifes Problem, 384
- Affin-Kovarianz der Runge-Kutta-Verfahren, 230
- Allgemeine Variation-der-Konstanten-Formel, 55
- Analytizitätsgebiet für logistischen Dgl., 209
- Analytizitätsvoraussetzung für Hamiltonsche Differentialgleichungen, 525
- Anfangswerte für Dgl. höherer Ordnung, 22
- Attraktive und repulsive Fixpunkte einer skalaren ODE, 341
- Attraktiver Grenzyklus, 381
- Autonome skalare Differentialgleichungen, 50
- Autonomisierung, 18
- Autonomisierungsinvarianz von Runge-Kutta-Verfahren, 232
- AWP-Löser in MATLAB, 25
- B-Stabilität, 362
- Bedeutung der modifizierten Gleichungen niedriger Ordnung, 517
- Bedeutung linearer AWPe, 56
- Bedingungsgleichungen für Linear-implizite Runge-Kutta-Verfahren 2. Ordnung, 393
- Berechnung der Modifikatoren  $\Delta \mathbf{f}_j$  durch Computeralgebra, 515
- Bimolekulare Reaktion, 29
- Butcher-Bäume, 244
- DAE: Transformation auf separierte Form, 407
- Definitionsintervalle von Lösungen von AWPe, 48
- Dense output, 234
- DIFEX, 278
- Doppelpendel, 70
- Effizienzgewinn durch Adaptivität, 298
- Einfache A-stabile RK-ESV, 350

- Einfache reversible Einschrittverfahren, 141  
Einfache symplektische Integratoren, 484  
Eingebettete RK-ESV, 310  
Eingebettete Runge-Kutta-Verfahren, 311  
Einschrittformulierung des Störmer-Verlet-Verfahrens, 113  
Energieerhaltung bei numerischer Integration, 467  
Energieerhaltung bei semi-impliziter Mittelpunktsregel, 394  
Euler-Extrapolationsverfahren mit Ordnungssteuerung, 272  
Euler-Verfahren für Pendelgleichung, 89  
Eulerverfahren für längenerhaltende Evolution, 98  
Explizite Runge-Kutta-Schritte für Riccati-Differentialgleichung, 226  
Explizites Euler-Verfahren für logistische Dgl, 82  
Explizites Eulerverfahren als Differenzenverfahren, 75  
Exponentielles Euler-Verfahren, 398  
Exponentielles Euler-Verfahren für steifes AWP, 399  
Extrapolationsverfahren als Runge-Kutta-Verfahren, 270  
Extrapolierte implizite Mittelpunktsregel, 274  
Extrapoliertes Euler-Verfahren, 267  
Federpendel, 494  
Fehler bei Polynominterpolation in Gauss-Knoten, 208  
Fixpunktform von Projektions-Einschrittverfahren, 166  
Funktionskalkül für Matrizen, 337  
Gauss-Kollokations-ESV bei stark attraktiven Fixpunkten, 363  
Gauss-Kollokationsverfahren, 360  
Gautschis Zweischnittverfahren, 559  
Glattheitsannahmen an rechte Seite, 118  
Globale  $h^2$ -Extrapolation für implizite Mittelpunktsregel, 276  
Gradientenfluss, 353  
Hamiltonsche Bewegungsgleichungen mit Nebenbedingungen, 422  
Implementierung steif-genauer RK-ESV für DAE, 430  
Implizite Mittelpunktsregel für Kreisbewegung, 101  
Implizite Mittelpunktsregel als Differenzenverfahren, 100  
Implizite Mittelpunktsregel für logistische Dgl., 101  
Implizite Mittelpunktsregel für Pendelgleichung, 103  
Implizite RK-ESV bei schnellen Transienten, 365  
Implizite RK-ESV mit linearisierten Inkrementgleichungen, 389  
Implizites Euler-Verfahren für Pendelgleichung in Deskriptorform, 426  
Implizites Eulerverfahren als Differenzenverfahren, 87  
Implizites Eulerverfahren für logistische Differentialgleichung, 87  
Ineffizienz expliziter Runge-Kutta-Verfahren, 320  
Inexakte Splittingverfahren, 285  
Interpolationsfehler bei Polynominterpolation in Gauss-Knoten, 188  
Interpretation der Stabilitätsfunktion, 327  
Invertierbarkeit der Koeffizientenmatrix von RK-ESV, 368  
Knotenanalyse eines Schaltkreises, 402  
Kollokationsverfahren als Projektionsverfahren, 163  
Kollokationsverfahren und numerische Quadratur, 160  
Kondition skalarer linearer Anfangswertprobleme, 60  
Konsistenzordnung einfacher Einschrittverfahren, 128  
Konstante 2-Formen, 473  
Konstruktion einfacher Runge-Kutta-Verfahren, 223  
Konvergenz des expliziten Euler-Verfahrens, 75  
Konvergenz einfacher Splittingverfahren, 279

Konvergenz expliziter Runge-Kutta-Verfahren, 237	Numerische Integration bei Blow-up, 287	Numerische Mathematik
Konvergenz kombinierter Verfahren, 250	Numerische Integratoren als approximative Evolutionsoperatoren, 50	
Konvergenz von einfachen Kollokations-Einschrittverfahren, 157	Numerische Quadratur, 160	
Konvergenz von Gauss-Kollokations-Einschrittverfahren, 177	Oregonator-Reaktion, 31	
Konvergenz von Kollokationsverfahren, 185	Partitionierte Runge-Kutta-Einschrittverfahren, 489	
Lösbarkeit der Inkrementgleichungen für Gauss-Kollokations-ESV, 359	Pendelgleichung in Deskriptorform, 419	
Lösung der Inkrementgleichungen, 234	Philosophische Grundlage der Rückwärtsanalyse, 507	
Lösungsfunktion aus Extrapolationsverfahren, 162	Präzession Magnetnadel, 436	
Langzeit-Energieerhaltung bei symplektischer Integration, 493	Projektion auf Energiemannigfaltigkeit, 504	
Linearisierung der Inkrementgleichungen, 386	Qualität der Fehlerschätzung, 293	
Lorenz system, 66	Räuber-Beute-Modelle, 26	R. Hiptmair rev 35327, 22. Januar 2008
Magnetnadel Präzession, 436	Radau-ESV bei stark attraktiven Fixpunkten, 372	
Massenpunkt im Zentralfeld, 40	Reskalierung des Modellproblems, 324	
Mathematisches Pendel, 37	Ressourcenbegrenztetes Wachstum, 23	
MATLAB-Integratoren für Index-1-DAEs, 417	Reversibilität bei mechanischen Systemen, 460, 463	
MATLAB-Integratoren für Pendelgleichung in Deskriptorform, 424	Richtungsfeld und Lösungskurven, 14	
Modifikatoren für einfache ESV, 515	RK-Bedingungsgleichungen für Konsistenzordnung $p = 3$ , 239	
Modifizierte Gleichung der Ordnung 2 zu explizitem Euler-Verfahren, 511	RK-ESV für autonome homogene lineare ODE, 334	
Modifizierte Gleichung für RK-ESV und lineare AWPes, 509	RK-ESV und Elimination der DAE-Nebenbedingungen, 414	
Modifizierte Gleichung für symplektisches Euler-Verfahren, 545	Romberg-Quadratur, 253	
Modifiziertes Gautschi-Verfahren, 567	Schaltkreis steife -gleichunge Zeitbereich, 379	4.5 p. 583
Molekulardynamik, 497	Schrittweitenbedingungen für 'Langzeitintegration', 544	
Notation fuer Einschrittverfahren, 120	Schrittweitenbeschränkung aus Lemma 2.2.7, 153	
	Schrittweitensteuerung für Bewegungsgleichungen, 317	

Schrittweitensteuerung für explizite Trapezregel/Euler-Verfahren

303

Schrittweitensteuerung und Blow-up, 310

Schrittweitensteuerung und Instabilität, 308

Schrittweitensteuerung und Kollaps, 309

Singulär gestörte Schaltkreisgleichungen, 409

Skalierungsinvarianz der Extrapolation, 254

Splittingverfahren für mechanische Systeme, 283

Störmel-Verlet-Verfahren für Pendelgleichung, 107

Störmer-Verlet-Verfahren als Differenzenverfahren, 106

Störmer-Verlet-Verfahren als Polygonzugmethode, 114

Stabilitätsfunktionen einiger RK-ESV, 330

Stabilitätsgebiet des exponentiellen Euler-Verfahren, 398

Standardintegratoren für oszillatorische Differentialgleichung,

557

Steife Schaltkreisgleichungen im Zeitbereich, 379

Steuerung von  $\tilde{\Psi}$  durch  $\Psi$ , 301

Strömungsvisualisierung, 449

Strang-Splitting erzeugt reversible ESV, 283

Stufenform der Inkrementgleichungen, 225

Symplektische Integratoren und variable Schrittweite, 554

Symplektisches Euler-Verfahren, 487

Symplektisches Euler-Verfahren für Pendelgleichung, 491

Symplektisches Flussintegral, 474

Symplektizität der Pendelgleichung, 469

Translationsinvarianz von Lösungen autonomer Dgl., 18

Umformulierung der Inkrementgleichungen, 147

Vektorräume von Vektorfeldern und Eigenschaften von Evolutionen, 479

ESV adaptiv, explizit, steifes Problem, 376

Verhalten von Stabilitätsfunktionen, 331

Vielteilchen-Molekulardynamik, 501

Volumenerhaltende Integratoren für  $d = 2$ , 451

Volumenerhaltendes Splittingverfahren 2. Ordnung, 455

Volumenerhaltung bei zweidimensionalen Hamiltonschen ODEs, 469

Warum Einschrittverfahren hoher Ordnung?, 246

Zeemans Herzschlagmodell, 32

Zeitlich ungleichmässiges Verhalten von Lösungen, 289



# Definitionen

- R-reversible Abbildung, 462
- (Abgeschnittene) asymptotische Entwicklung, 256
- A-Stabilität, 350
- algebraische Stabilität, 362
- Alternierende, nichtdegenerierte Bilinearform, 473
- Arten der Konvergenz, 79
- Asymptotische Stabilität eines Fixpunkts, 340
- DAE vom Index 1, 408
- Diskretisierungsfehler, 121
- Dissipatives Vektorfeld, 356
- Einschrittverfahren, 119
- Erstes Integral, 29
- Evolutionoperator, 49
- explizite und implizite Einschrittverfahren, 121
- Exponentielle Runge-Kutta-Verfahren, 401
- Fixpunkt, 339
- Hamiltonsche Differentialgleichung, 39
- Konsistenz einer diskreten Evolution, 125
- Konsistenzfehler einer diskreten Evolution, 125
- Konsistenzordnung einer diskreten Evolution, 127
- Konvergenz und Konvergenzordnung, 122
- L-Stabilität, 367
- Lösung einer gewöhnlichen Differentialgleichung, 14
- Lösung eines Anfangswertproblems, 21
- Lokale Lipschitz-Stetigkeit, 46
- Maximale Fortsetzbarkeit einer Lösung, 44
- Modifizierte Gleichung der Ordnung  $q$ , 510
- Nichtexpansivität, 353
- Nichtlineare Stabilität, 140
- Polynomiale Invarianten, 435
- Projektionsoperator, 164
- Residuum einer komplexwertigen Funktion, 193
- Reversible diskrete Evolutionen, 141
- Runge-Kutta-Verfahren, 224
- Stabilitätsgebiet eines Einschrittverfahrens, 326
- Sternförmiges Gebiet, 480

Stetiger linearer Operator, 164  
Symplektische Abbildung, 478  
symplektisches Einschrittverfahren, 483  
Symplektisches Produkt, 472  
Volumenerhaltung, 446

# MATLAB-Codes

odeset, 314

ode45, 376

odeset, 376

ode15s, 384

ode23s, 384

ode23, 313

ode45, 280, 313

odeset, 280, 313, 384

Adaptives RK-ESV für steifes Problem, 376

Aitken-Neville-Extrapolation, 255

ode15s, 418

ode23s, 437

ode23t, 418

ode45, 437

# Notationen

$C^l(J, D) \hat{=}$   $l$ -mal stetig differenzierbare Funktionen  $J \mapsto D$ , 14

$D_y \mathbf{f} \hat{=}$  Ableitung von  $\mathbf{f}$  nach  $\mathbf{y}$  (Jacobi-Matrix), 46

$J(t_0, \mathbf{y}_0) := ]t_-, t_+[ \hat{=}$  maximales Existenzintervall, 45

$O(g(h)) \hat{=}$  "Landau- $O$ ", 78

$P_n \hat{=}$  Legendre-Polynom vom Grad  $n \in \mathbb{N}_0$ , 197

$[x] \hat{=}$  ganzzahliger Anteil von  $x > 0$ , 538

$\operatorname{div} \mathbf{f} \hat{=}$  Divergenz eines Vektorfeldes, 447

$\|\mathbf{y}\|_M := (\mathbf{y}^T \mathbf{M} \mathbf{y})^{1/2}$  induzierte Vektornorm, 353

$\mathcal{P}_s \hat{=}$  Raum der univariaten Polynome vom Grad  $\leq s$ , 145

$\Psi^{s,t} \mathbf{y} \hat{=}$  Diskrete Evolution, 119

$\mathbf{J} \hat{=}$  Matrix zum symplektischen Produkt, 472

$\mathbf{a} \times \mathbf{b} \hat{=}$  Vektorprodukt, 436

$\mathbf{y}, \mathbf{z}, \dots$  Fettdruck für Spaltenvektoren, 13

$\mathbb{C}^- := \{z \in \mathbb{C} : \operatorname{Re} z < 0\}$ , 342

$\mathbf{1} = (1, \dots, 1)^T$ , 327

$^{(n)} \hat{=}$   $n$ . Ableitung nach der Zeit  $t$ , 19

$\nabla^2 f \hat{=}$  Hesse-Matrix, 478

$\operatorname{adj}(\mathbf{X})$  adjungierte Matrix, 443

$\otimes \hat{=}$  Kronecker-Produkt von Matrizen, 389

$\rho(\mathbf{A})$ : Spektralradius einer Matrix, 347

$\mathbf{f} \hat{=}$  rechte Seite einer ODE, 12

$\sigma(\mathbf{A})$ : Spektrum einer Matrix, 342

$\dot{\cdot} \hat{=}$  Ableitung nach der Zeit  $t$ , 13

$\mathbf{y}^\alpha$  für Multiindex  $\alpha$ , 435

$\operatorname{Im}(\mathbf{M}) := \{\mathbf{M}\mathbf{x} : \mathbf{x} \in \mathbb{R}^d\}$ , Bild der Matrix  $\mathbf{M}$ , 405

TOL Toleranz, 292

Euklidische Vektornorm  $\|\cdot\|$ , 40

Hamilton-Funktion  $H$ , 39

Konsistenzfehler  $\tau(t, \mathbf{y}, h)$ , 125

Landau- $o$ , 126

Landau-Symbol  $O(h^p)$ , 122

Push-Forward  $\Phi_*$ , 472

symplektisches Produkt  $\omega(\mathbf{v}, \mathbf{w})$ , 472

Vektorprodukt  $\times$ , 42

# Appendix

## MATLAB-Files zu Beispielen

- Beispiel 1.2.5 ↔ File/Directory `ex:LV`
- Beispiel 1.2.12 ↔ File/Directory `ex:Oregonator`
- Beispiel 1.2.15 ↔ File/Directory `ex:heartbeat`
- Beispiel 1.3.36 ↔ File/Directory `ex:Lorenz`
- Beispiel 1.4.4 ↔ File/Directory `ex:expleulcvg`
- Beispiel 1.4.9 ↔ File/Directory `ex:eeullog`

- Beispiel 1.4.15 ↔ File/Directory `ex:ieullog`
- Beispiel 1.4.17 ↔ File/Directory `ex:pendeul`
- Beispiel 1.4.18 ↔ File/Directory `ex:eulspin`
- Beispiel 1.4.21 ↔ File/Directory `ex:logimid`
- Beispiel 1.4.22 ↔ File/Directory `ex:imidspin`  

```
>> eulspin([1;0],10,40,'midspin40')  
>> eulspin([1;0],10,160,'mispin160')
```
- Beispiel 1.4.24 ↔ File/Directory `ex:pendimid`  

```
>> pendmidp([pi/4;0],5,50,'pendimid50');  
>> pendmidp([pi/4;0],5,100,'pendimid100');  
>> pendmidp([pi/4;0],5,200,'pendimid200');
```
- Beispiel 1.4.32 ↔ File/Directory `ex:svpend`
- Beispiel 2.2.57 ↔ File/Directory `ex:cvgkoll`
- Beispiel 2.2.49 ↔ File/Directory `ex:GaussCollcvg`
- Beispiel 2.4.2 ↔ File/Directory `ex:kombesv`
- Beispiel 2.3.22 ↔ File/Directory `ex:rkexplcvg`
- Beispiel 2.4.17 ↔ File/Directory `ex:eulexpol`
- Beispiel 2.4.19 ↔ File/Directory `ex:eulex`

- Beispiel 2.5.4 ↔ File/Directory ex:splitcvg
- Beispiel 2.5.14 ↔ File/Directory ex:splitinex
- Beispiel 2.6.4 ↔ File/Directory ex:qualest
- Beispiel 2.6.15 ↔ File/Directory ex:adesv
- Beispiel 2.6.16 ↔ File/Directory ex:adaptsat
- Beispiel 3.0.1 ↔ File/Directory ex:logeximpl
- Beispiel 3.3.14 ↔ File/Directory ex:GaussCollLog
- Beispiel 3.4.1 ↔ File/Directory ex:iesvstiff
- Beispiel 3.5.2 ↔ File/Directory ex:ode45stiff
- Beispiel 3.5.5 ↔ File/Directory ex:odecircuit
- Beispiel 3.5.7 ↔ File/Directory ex:odes
- Beispiel 3.6.1 ↔ File/Directory ex:silog  
logsieul(0.1,round(5\*1.5.^(0:18)), 'silog');
- Beispiel 3.6.3 ↔ File/Directory ex:siradlog  
siradlog(0.1,round(5\*1.5.^(0:18)), 'siradlog')
- Beispiel 3.6.12 ↔ File/Directory ex:pendimipEnSmp
- Beispiel 3.7.5 ↔ File/Directory ex:eeul

- Beispiel 3.7.6 ↔ File/Directory `ex:eeulstiff`
- Beispiel 3.8.15 ↔ File/Directory `ex:daecircml`
- Beispiel 3.8.28 ↔ File/Directory `ex:daependmatlab`
- Beispiel 3.8.29 ↔ File/Directory `ex:daependieul`
- Beispiel 4.1.3 ↔ File/Directory `ex:magneedle`
- Beispiel 4.4.1 ↔ File/Directory `ex:enpres`
- Beispiel 4.4.33 ↔ File/Directory `ex:pendsympeul`
- Beispiel 4.4.37 ↔ File/Directory `ex:md`
- Beispiel 4.4.35 ↔ File/Directory `ex:springpend`
- Beispiel 4.4.39 ↔ File/Directory `ex:moldyn`



# Literaturverzeichnis

- [1] H. AMANN, *Gewöhnliche Differentialgleichungen*, Walter de Gruyter, Berlin, 1st ed., 1983.
- [2] V. ARNOLD, *Mathematical Methods of Classical Mechanics*, Springer, New York, 2nd ed., 1989.
- [3] C. BLATTER, *Analysis I*, vorlesungsskriptum, ETH Zürich, Zürich, Switzerland, 2003.  
<http://www.math.ethz.ch/~blatter/>.
- [4] M. CHAWLA AND M. JAIN, *Error estimates for gauss quadrature formulas for analytic functions*, *Math. Comp.*, 22 (1968), pp. 82–90.
- [5] P. DAVIS, *Interpolation and Approximation*, Dover, New York, 1975.
- [6] M. DEAKIN, *Applied catastrophe theory in the social and biological sciences*, *Bulletin of Mathematical Biology*, 42 (1980), pp. 647–679.
- [7] P. DEUFLHARD, *Numerik von anfangswertaufgaben für gewöhnliche differentialgleichungen*, Tech. Rep. TR 89-2, ZIB Berlin, Berlin, Germany, 1989.

- [8] P. DEUFLHARD AND F. BORNEMANN, *Numerische Mathematik II*, DeGruyter, Berlin, 2 ed., 2002.
- [9] P. DEUFLHARD AND A. HOHMANN, *Numerische Mathematik I*, DeGruyter, Berlin, 3 ed., 2002.
- [10] G. FISCHER, *Lineare Algebra*, Vieweg–Verlag, Braunschweig, 9th ed., 1986.
- [11] C. GRAY, *An analysis of the Belousov-Zhabotinski reaction*, Rose-Hulman Undergraduate Math Journal, 3 (2002). <http://www.rose-hulman.edu/mathjournal/archives/2002/vol3-n1/paper1/v3n1-1pd.pdf>.
- [12] W. HACKBUSCH, *Iterative Lösung großer linearer Gleichungssysteme*, B.G. Teubner–Verlag, Stuttgart, 1991.
- [13] E. HAIRER AND C. LUBICH, *Asymptotic expansions and backward analysis for numerical integrators*, in Dynamics of algorithms, vol. 118 of IMA Vol. Math. Appl., Springer, New York, 2000, pp. 91–106.
- [14] E. HAIRER, C. LUBICH, AND G. WANNER, *Geometric Numerical Integration*, vol. 31 of Springer Series in Computational Mathematics, Springer, Berlin, Germany, 2002. Table of contents.
- [15] —, *Geometric numerical integration illustrated by the Störmer-Verlet method*, Acta Numerica, 12 (2003), pp. 399–450.
- [16] —, *Geometric numerical integration*, vol. 31 of Springer Series in Computational Mathematics, Springer, Heidelberg, 2 ed., 2006.
- [17] E. HAIRER, S. NORSETT, AND G. WANNER, *Solving Ordinary Differential Equations I. Nonstiff Problems*, Springer-Verlag, Berlin, Heidelberg, New York, 2 ed., 1993.

- [18] E. HAIRER AND G. WANNER, *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*, Springer-Verlag, Berlin, Heidelberg, New York, 1991.
- [19] M. HERRMANN, *Numerik gewöhnlicher Differentialgleichungen*, Oldenbourg, München, 2004.
- [20] N. HIGHAM, *The scaling and squaring method for the matrix exponential revisited*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 1179–1193.
- [21] M. HIRSCH, S. SMALE, AND R. DEVANEY, *Differential Equations, Dynamical Systems, and an Introduction to Chaos*, vol. 60 of Pure and Applied Mathematics, Elsevier Academic Press, Amsterdam, 2 ed., 2004.
- [22] M. HOCHBRUCK AND C. LUBICH, *On Krylov subspace approximations to the matrix exponential operator*, SIAM J. Numer. Anal., 34 (1997), pp. 1911–1925.
- [23] —, *A Gautschi-type method for oscillatory second-order differential equations*, Numer. Math., 83 (1999), pp. 403–426.
- [24] M. HOCHBRUCK, C. LUBICH, AND H. SELHOFER, *Exponential integrators for large systems of differential equations*, SIAM J. Sci. Comp., 19 (1998), pp. 1552–1574.
- [25] M. HOCHBRUCK AND A. OSTERMANN, *Exponential integrators*, Acta Numerica, 19 (2010), pp. 209–286.
- [26] B. LEIMKUHLER AND S. REICH, *Simulating Hamiltonian Dynamics*, vol. 14 of Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, Cambridge, UK, 2004.

- [27] E. LORENZ, *Deterministic non-periodic flow*, J. Atmospheric Sciences, 20 (1963), pp. 130–141.
- [28] B. MINCHEV AND W. WRIGHT, *A review of exponential integrators for first order semi-linear problems*, Preprint 2/2005, NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET, Trondheim, Norway, 2005.
- [29] S. REICH, *Backward error analysis for numerical integrators*, SIAM J. Numer. Anal., 36 (1999), pp. 1549–1570.
- [30] R. REMMERT, *Funktionentheorie I*, no. 5 in Grundwissen Mathematik, Springer, Berlin, 1984.
- [31] L. SHAMPINE, M. REICHELT, AND J. KIERZENKA, *The MATLAB ODE suite*, SIAM J. Sci. Comp., 18 (1997), pp. 1–22.
- [32] W. WALTER, *Gewöhnliche Differentialgleichungen. Eine Einführung*, vol. 110 of Heidelberger Taschenbücher, Springer, Heidelberg, 3 ed., 1986.