

Lösung 12

1. Nullstellensuche mit dem Newton Verfahren

a) Für Anfangswerte im Bereich $(-\infty, 1.5)$ konvergiert das Newtonverfahren gegen die Nullstelle $x = 1$. Für Anfangswerte im Bereich $(1.5, \infty)$ konvergiert das Newtonverfahren gegen die Nullstelle $x = 2$. Für $x_0 = 1.5$ bricht das Verfahren ab, weil $F'(1.5) = 0$.

b) Siehe `newtonKonvergenz.m`. Wir beobachten, dass für die letzte 3 Iterationen die Konvergenz praktisch quadratisch ist.

Bemerkung: Die beim Newton-Verfahren beobachtete Konvergenz (quadratische Konvergenz) ist eine sehr schnelle Konvergenz. Sobald man in dem Bereich ist, in dem das Newton-Verfahren überhaupt konvergiert (also den Startwert halbwegs gut gewählt hat), braucht das Newton-Verfahren nur sehr wenige Schritte bis das Abbruchkriterium erfüllt ist.

c) Das Verfahren hat 3 deutliche Konvergenzbereiche. Für Anfangswerte in $(-\infty, 1.421)$ konvergiert das Verfahren gegen $x = 1$, für $(1.554, 2.446)$ gegen $x = 2$ und für $(2.579, \infty)$ gegen $x = 3$.

Für $(1.425, 1.534)$ konvergiert das Verfahren gegen die Nullstelle $x = 3$, weil dort die Ableitung DF klein und negativ ist. Das heisst, dass die Funktion ziemlich flach ist, und dass die Tangente eine fast horizontale Gerade mit negativer Steigung ist. Daraus folgt, dass die zweite Iterierten in $(2.579, \infty)$ liegen.

Für $(1.538, 1.55)$ konvergiert das Verfahren gegen die Nullstelle $x = 1$, weil die zweite Iterierten im $(2.467, 2.552)$ liegen, und für diese Anfangswert konvergiert das Verfahren gegen die Nullstelle $x = 1$. Um die zweite Iterierten darzustellen kann man `nMax=1` setzen.

Für $(1.55, 1.554)$ ist die Analyse noch komplizierter, weil die Iterierten zwischen den zwei Bereichen mit DF klein springen.

d) Siehe `newtonKonvergenzFD.m`. Für $h = 10^{-7}$ erhalten wir ähnlich Ergebnisse. Hingegen für $h = 10^{-16}$ konvergiert das Verfahren nur linear. In diesem Fall ist die ledigliche Reduktion der Konvergenzordnung ein glücklicher Zufall.

Bei der numerischen Berechnung von Ableitungen mit Finiten-Differenzen tritt bei kleinem h (klein relativ zur Maschinengenauigkeit `eps`) sog. Auslöschung

auf. In Abbildung 1 wird der Fehler, d.h. die Differenz zwischen der analytischen Ableitung und der Näherung mittels Finiten-Differenzen, gezeigt. Ab ca. $h \approx \sqrt{\text{eps}} \approx 10^{-8}$ vergrößert sich der Fehler mit kleiner werdendem h ! Dies kann man (sehr einfach!) theoretisch Begründen, aber dies wird nicht weiter in dieser Vorlesung erörtert.

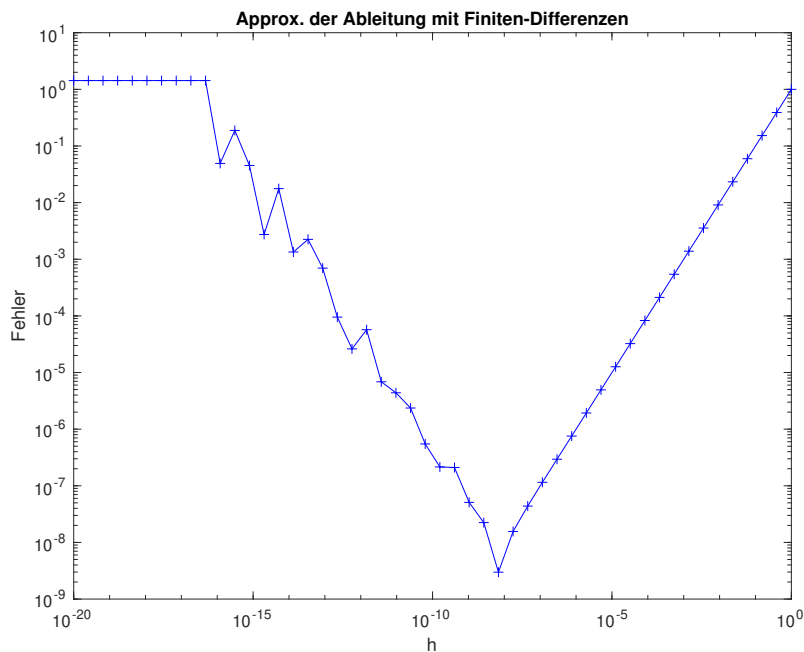


Abbildung 1 – Approximation der Ableitung mit Finiten-Differenzen.

2. Stabilitäts-funktionen und Gebiete

a) (i) Expliziter Euler

$$\begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array}$$

Wir haben

$$\begin{aligned} y_{j+1} &= y_j + hf(t_j, y_j) \\ &= y_j + h\lambda y_j \\ &= (1 + h\lambda)y_j. \end{aligned}$$

Deshalb, für $z = h\lambda$, ist die Stabilitäts funktion

$$g(z) = 1 + z.$$

Siehe nächstes Blatt!

(ii) Impliziter Euler

$$\frac{1}{1} \mid \frac{1}{1}$$

Wir haben

$$k_1 = f(t_j + h, y_j + hk_1) = \lambda(y_j + hk_1) = \lambda y_j + h\lambda k_1 \iff (1 - h\lambda)k_1 = \lambda y_j,$$

und daher

$$k_1 = y_j \left(\frac{\lambda}{1 - h\lambda} \right).$$

Dann

$$\begin{aligned} y_{j+1} &= y_j + hk_1 \\ &= y_j + hy_j \left(\frac{\lambda}{1 - h\lambda} \right) \\ &= \left(1 + \frac{h\lambda}{1 - h\lambda} \right) y_j \\ &= \left(\frac{1}{1 - h\lambda} \right) y_j. \end{aligned}$$

Deshalb, für $z = h\lambda$, ist die Stabilitätsfunktion

$$g(z) = \frac{1}{1 - z}.$$

(iii) Heun Verfahren

$$\begin{array}{c|cc} 0 & & \\ \hline 1 & 1 & \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

Wir haben

$$\begin{aligned} y_{j+1} &= y_j + \frac{h}{2}(f(t_j, y_j) + f(t_j + h, y_j + hf(t_j, y_j))) \\ &= y_j + \frac{h}{2}(\lambda y_j + \lambda(y_j + h\lambda y_j)) \\ &= y_j + \frac{h}{2}\lambda y_j + \frac{h}{2}\lambda y_j + \frac{h^2}{2}\lambda^2 y_j \\ &= \left(1 + h\lambda + \frac{h^2}{2}\lambda^2 \right) y_j. \end{aligned}$$

Deshalb, für $z = h\lambda$, ist die Stabilitätsfunktion

$$g(z) = 1 + z + \frac{1}{2}z^2.$$

Bitte wenden!

(iv) Klassisches Runge-Kutta Verfahren

| | | | | |
|---------------|---------------|---------------|---------------|---------------|
| 0 | | | | |
| $\frac{1}{2}$ | $\frac{1}{2}$ | | | |
| $\frac{1}{2}$ | 0 | $\frac{1}{2}$ | | |
| 1 | 0 | 0 | 1 | |
| | $\frac{1}{6}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{6}$ |

Wir haben

$$\begin{aligned}k_1 &= f(t_j, y_j) &&= \lambda y_j, \\k_2 &= f\left(t_j + \frac{h}{2}, y_j + \frac{h}{2}k_1\right) &&= \lambda\left(y_j + \frac{h}{2}\lambda y_j\right), \\k_3 &= f\left(t_j + \frac{h}{2}, y_j + \frac{h}{2}k_2\right) &&= \lambda\left(y_j + \frac{h}{2}\lambda\left(y_j + \frac{h}{2}\lambda y_j\right)\right), \\k_4 &= f(t_j + h, y_j + hk_3) &&= \lambda\left(y_j + h\lambda\left(y_j + \frac{h}{2}\lambda\left(y_j + \frac{h}{2}\lambda y_j\right)\right)\right).\end{aligned}$$

Dann

$$\begin{aligned}y_{j+1} &= y_j + h\left(\frac{1}{6}k_1 + \frac{2}{6}k_2 + \frac{2}{6}k_3 + \frac{1}{6}k_4\right) \\&= y_j + h\left(\frac{1}{6}\lambda y_j + \frac{2}{6}\lambda\left(y_j + \frac{h}{2}\lambda y_j\right) + \frac{2}{6}\lambda\left(y_j + \frac{h}{2}\lambda\left(y_j + \frac{h}{2}\lambda y_j\right)\right) \right. \\&\quad \left. + \frac{1}{6}\lambda\left(y_j + h\lambda\left(y_j + \frac{h}{2}\lambda\left(y_j + \frac{h}{2}\lambda y_j\right)\right)\right)\right) \\&= y_j + \frac{1}{6}h\lambda y_j + \frac{2}{6}h\lambda y_j + \frac{1}{6}h^2\lambda^2 y_j + \frac{2}{6}h\lambda y_j + \frac{1}{6}h^2\lambda^2 y_j + \frac{1}{12}h^3\lambda^3 y_j \\&\quad + \frac{1}{6}h\lambda y_j + \frac{1}{6}h^2\lambda^2 y_j + \frac{1}{12}h^3\lambda^3 y_j + \frac{1}{24}h^4\lambda^4 y_j \\&= \left(1 + \frac{1}{6}h\lambda + \frac{2}{6}h\lambda + \frac{1}{6}h^2\lambda^2 + \frac{2}{6}h\lambda + \frac{1}{6}h^2\lambda^2 + \frac{1}{12}h^3\lambda^3 \right. \\&\quad \left. + \frac{1}{6}h\lambda + \frac{1}{6}h^2\lambda^2 + \frac{1}{12}h^3\lambda^3 + \frac{1}{24}h^4\lambda^4\right)y_j \\&= \left(1 + h\lambda + \frac{1}{2}h^2\lambda^2 + \frac{1}{6}h^3\lambda^3 + \frac{1}{24}h^4\lambda^4\right)y_j.\end{aligned}$$

Deshalb, für $z = h\lambda$, ist die Stabilitätsfunktion

$$g(z) = 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3 + \frac{1}{24}z^4.$$

(v) Implicit midpoint rule

| | |
|---------------|---------------|
| $\frac{1}{2}$ | $\frac{1}{2}$ |
| | 1 |

Siehe nächstes Blatt!

Wir haben

$$k_1 = f\left(t_j + \frac{h}{2}, y_j + \frac{h}{2}k_1\right) = \lambda\left(y_j + \frac{h}{2}k_1\right) = \lambda y_j + \frac{h}{2}\lambda k_1 \iff \left(1 - \frac{h}{2}\lambda\right)k_1 = \lambda y_j,$$

and hence

$$k_1 = y_j \left(\frac{\lambda}{1 - \frac{h}{2}\lambda} \right) = y_j \left(\frac{2\lambda}{2 - h\lambda} \right).$$

Dann

$$\begin{aligned} y_{j+1} &= y_j + hk_1 \\ &= y_j + hy_j \left(\frac{2\lambda}{2 - h\lambda} \right) \\ &= \left(1 + \frac{2h\lambda}{2 - h\lambda} \right) y_j \\ &= \left(\frac{2 + h\lambda}{2 - h\lambda} \right) y_j. \end{aligned}$$

Deshalb, für $z = h\lambda$, ist die Stabilitätsfunktion

$$g(z) = \left(\frac{2 + z}{2 - z} \right).$$

b) Siehe `stabgebiete.m` und `draw_stabfunc.m`. Die Stabilitätsgebiete ist schwarz lackiert.

c) Wir benötigen:

(i) $|1 + x| < 1 \iff -1 < 1 + x < 1 \implies -2 < x < 0.$

(ii) $\left| \frac{1}{1 - x} \right| < 1 \iff 1 < |1 - x|.$

So

$$1 < 1 - x \implies x < 0 \quad \text{und} \quad 1 < x - 1 \implies x > 2.$$

(iii) $\left| 1 + x + \frac{1}{2}x^2 \right| < 1 \iff 1 + x + \frac{1}{2}x^2 < 1 \iff x(x+2) < 0 \implies -2 < x < 0.$

(iv) $\left| 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4 \right| < 1 \iff x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4 < 0.$

Siehe `stab_klassisches_RK.m`. Wir haben eine Null bei $x \approx -2.7853$ und es ist klar, dass $x = 0$ auch eine Null ist. Da $g(x)$ außerhalb dieser Region zunimmt, das gilt das Stabilitätsintervalle ist $-2.7853 < x < 0.$

(v) $\left| \frac{2 + x}{2 - x} \right| < 1 \iff |2 + x| < |2 - x|.$ Es ist klar gilt dies nur für $x < 0.$

Bitte wenden!

d) Definiere

$$\lambda = -1000 + \pi i.$$

We want to determine h such that

$$|g(z)| = |g(h\lambda)| < 1, \quad \text{or equivalently} \quad |g(h\lambda)|^2 < 1.$$

(i) We have

$$g(h\lambda) = 1 + h\lambda.$$

Therefore, h should be such that

$$1 > |g(h\lambda)|^2 = |1 + h\lambda|^2 = |1 + h(-1000 + \pi i)|^2 = (1 - 1000h)^2 + (\pi h)^2.$$

Hence we require that

$$h < \frac{2000}{1000000 + \pi^2}.$$

(ii) We have

$$g(h\lambda) = \frac{1}{1 - h\lambda}. \tag{1}$$

Therefore, h should be such that

$$\begin{aligned} 1 > |g(h\lambda)|^2 &= \left| \frac{1}{1 - h\lambda} \right|^2 = \left| \frac{1}{1 + 1000h - h\pi i} \cdot \frac{1 + 1000h + h\pi i}{1 + 1000h + h\pi i} \right|^2 \\ &= \left| \frac{1 + 1000h}{(1 + 1000h)^2 + (h\pi)^2} + i \frac{-h\pi}{(1 + 1000h)^2 + (h\pi)^2} \right|^2 \\ &= \left(\frac{1 + 1000h}{(1 + 1000h)^2 + (h\pi)^2} \right)^2 + \left(\frac{-h\pi}{(1 + 1000h)^2 + (h\pi)^2} \right)^2 \\ &= \frac{(1 + 1000h)^2 + (h\pi)^2}{((1 + 1000h)^2 + (h\pi)^2)^2} \\ &= \frac{1}{(1 + 1000h)^2 + (h\pi)^2}. \end{aligned}$$

This holds when $h < -\frac{2000}{1000000 + \pi^2}$ or $h > 0$. As the step size must be positive this means we require that

$$h > 0.$$

(iii)-(v) First we rewrite the stability condition $|g_\lambda(h)| < 1$ as

$$G(h) := 1 - |g_\lambda(h)| > 0.$$

Then we use a root finding procedure to precisely locate the points where $G(h) = 0$. Siehe `grenzen_schrittweite_h.m`.

Siehe nächstes Blatt!

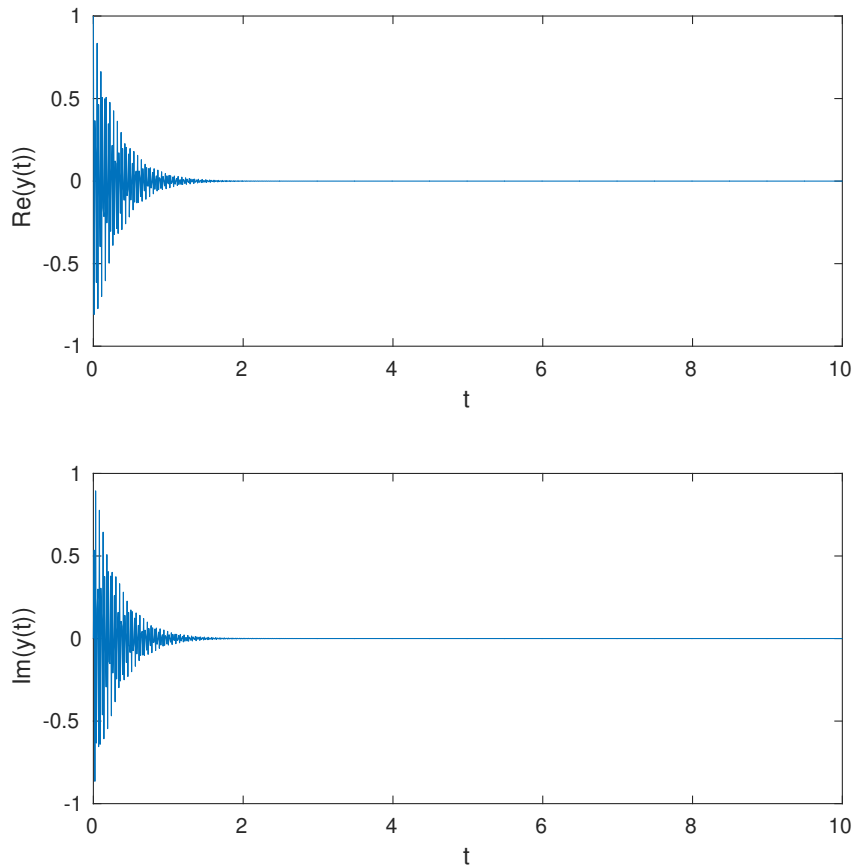


Abbildung 2 – Realteil und den Imaginärteil der Lösung $y(t)$ aus d).

- (iii) We require $h < 0.002$.
- (iv) We require $h < 0.0028$.
- (v) h can be any positive number.

e) Siehe `schnelle_oszillation_verfall.m` und Abbildung 2.

3. Lineares homogenes System

a) Diagonalisieren die Matrix A als:

$$A = PDP^{-1} \iff D = P^{-1}AP,$$

wo

$$P = \begin{bmatrix} 15 & -12 & 1 \\ 0 & 12 & 1 \\ 0 & 4 & -3 \end{bmatrix}, \quad D = \begin{bmatrix} -\frac{1}{2} & 0 & 0 \\ 0 & -15 & 0 \\ 0 & 0 & -1000 \end{bmatrix}, \quad P^{-1} = \begin{bmatrix} \frac{1}{15} & \frac{4}{75} & \frac{1}{25} \\ 0 & \frac{3}{40} & \frac{1}{40} \\ 0 & \frac{1}{10} & -\frac{3}{10} \end{bmatrix}.$$

Bitte wenden!

Sei $\mathbf{z} = P^{-1}\mathbf{y} \iff \mathbf{y} = P\mathbf{z}$ und beachten Sie das

$$\dot{\mathbf{z}} = P^{-1}\dot{\mathbf{y}} = P^{-1}A\mathbf{y} = P^{-1}AP\mathbf{z} = D\mathbf{z}.$$

Wir haben jetzt drei entkoppelte ODEs, wo die Lösung gegeben ist

$$z_i(t) = z_i(0)e^{\lambda_i t}, \quad i = 1, \dots, 3.$$

Wir haben

$$\mathbf{z}(0) = P^{-1}\mathbf{y}(0) = \begin{bmatrix} \frac{1}{15} & \frac{4}{75} & \frac{1}{25} \\ 0 & \frac{3}{40} & \frac{1}{40} \\ 0 & \frac{1}{10} & -\frac{3}{10} \end{bmatrix} \begin{bmatrix} 1 \\ 6 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{14}{30} \\ \frac{1}{2} \\ 0 \end{bmatrix}.$$

Deshalb

$$\begin{aligned} z_1(t) &= z_1(0)e^{\lambda_1 t} = \frac{14}{30}e^{-\frac{1}{2}t}. \\ z_2(t) &= z_2(0)e^{\lambda_2 t} = \frac{1}{2}e^{-15t}. \\ z_3(t) &= z_3(0)e^{\lambda_3 t} = 0e^{-1000t} = 0. \end{aligned}$$

Dann ist gilt

$$\mathbf{y}(t) = P\mathbf{z}(t) = \begin{bmatrix} 15 & -12 & 1 \\ 0 & 12 & 1 \\ 0 & 4 & -3 \end{bmatrix} \begin{bmatrix} \frac{14}{30}e^{-\frac{1}{2}t} \\ \frac{1}{2}e^{-15t} \\ 0 \end{bmatrix} = \begin{bmatrix} 7e^{-\frac{1}{2}t} - 6e^{-15t} \\ 6e^{-15t} \\ 2e^{-15t} \end{bmatrix}.$$

- b) Siehe `linsys.m`. Beachten Sie dass globale Diskretisierungsfehler divergiert, während die Schrittgröße größer ist als die maximal zulässige Schrittgröße aufgrund des Stabilitätsintervalle.

Für das verbesserte Euler-Verfahren haben wir

$$\begin{aligned} k_1 &= f(t_j, y_j) = \lambda y_j, \\ k_2 &= f\left(t_j + \frac{h}{2}, y_j + \frac{h}{2}k_1\right) = \lambda\left(y_j + \frac{h}{2}\lambda y_j\right). \end{aligned}$$

und

$$\begin{aligned} y_{j+1} &= y_j + hk_2 = y_j + h\lambda\left(y_j + \frac{h}{2}\lambda y_j\right) \\ &= \left(1 + h\lambda + \frac{1}{2}h^2\lambda^2\right)y_j \end{aligned}$$

Deshalb, für $z = h\lambda$, ist die Stabilitätsfunktion

$$g(z) = 1 + z + \frac{1}{2}z^2.$$

Siehe nächstes Blatt!

Für $z = h\lambda$, von Teil 1.c) (iii). wir haben das

$$-2 < h\lambda < 0.$$

Das Worst-Case-Szenario ist, $\lambda = \lambda_3 = -1000$. Dann wir haben

$$-2 < h\lambda < 0 \iff 0 < 1000h < 2 \iff 0 < h < \frac{1}{500}.$$

Wenn wir im Stabilitätsintervall sind, d.h. $h < \frac{1}{500}$, konvergiert die Methode, und außerdem konvergiert sie mit der erwarteten Ordnung, die $p = 2$ ist. Siehe Abbildung 3.

Beachten Sie, dass die wahre Lösung $y(t)$ nicht den $\lambda_3 = -1000$ Eigenwert enthält, was darauf hindeutet, dass wir es bei der numerischen Lösung für $y(t)$ ignorieren können. Dies ist jedoch nicht der Fall, da die numerische Lösung nicht exakt ist und somit der Beitrag aus diesem Eigenwert die anderen Modi im numerischen Fall beeinflussen wird. Es ist also notwendig, diesen Eigenwert zu berücksichtigen, um sicherzustellen, daß die numerische Approximation korrekt die wahre Lösung darstellt.

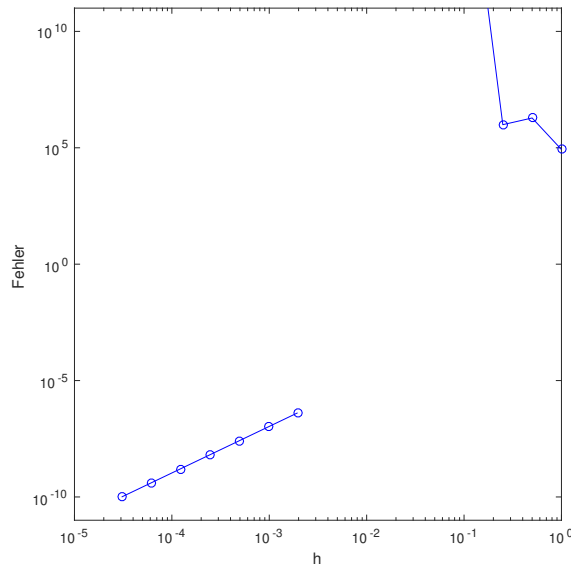


Abbildung 3 – Konvergenz der verbesserten Euler-Methode für ein System von drei linearen homogenen Differentialgleichungen 2 b).

4. Stabilitätsfunktionen für RK-ESV

Wir wenden ein explizites s -stufiges RK-ESV auf die ODE $\dot{y} = \lambda y$ an. Es gilt

$$k_i = f(t_0 + hc_i, y_0 + h \sum_{j=1}^{i-1} a_{ij} k_j) = \lambda(y_0 + h \sum_{j=1}^{i-1} a_{ij} k_j) \quad 1 \leq i \leq s.$$

Bitte wenden!

Wir behaupten, dass $k_i = p_i(h\lambda)\lambda y_0$, wobei p_i ein Polynom ist. Für $i = 1$ haben wir $k_1 = \lambda y_0$, weil das Verfahren explizit ist. Nehmen wir an, dass die Behauptung für i gilt. Dann

$$\begin{aligned}
 k_{i+1} &= f(t_0 + hc_i, y_0 + h \sum_{j=1}^i a_{ij} k_j) \\
 &= \lambda(y_0 + h \sum_{j=1}^i a_{ij} k_j) \\
 &= \lambda(y_0 + h \sum_{j=1}^i a_{ij} p_j(h\lambda)\lambda y_0) \\
 &= (1 + h\lambda \sum_{j=1}^i a_{ij} p_j(h\lambda))\lambda y_0 \\
 &=: p_{i+1}(h\lambda)\lambda y_0.
 \end{aligned}$$

Da eine endliche Kombination von Polynomen wieder ein Polynom ist, folgt die Behauptung. Nun ist

$$\begin{aligned}
 y_1 &= y_0 + h \sum_{i=1}^s b_i k_i \\
 &= y_0 + h \sum_{i=1}^s b_i p_i(h\lambda)\lambda y_0 \\
 &= (1 + h\lambda \sum_{i=1}^s b_i p_i(h\lambda))y_0,
 \end{aligned}$$

was zeigt, dass die Stabilitätsfunktion ein Polynom in $h\lambda$ ist.

5. Steifes nichtlineares AWP

- a) Zunächst müssen wir die Van der Pol-Gleichung umschreiben in ein System erster Ordnung:

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

wobei

$$\mathbf{y}(t) = \begin{pmatrix} y_0(t) \\ y_1(t) \end{pmatrix} \quad \text{und} \quad \mathbf{f}(t, \mathbf{y}(t)) = \begin{pmatrix} y_1(t) \\ \mu(1 - y_0^2(t))y_1 - y_0 \end{pmatrix}.$$

Die Anfangswerte sind dann $\mathbf{y}(0) = \mathbf{y}_0 = [2, 0]^T$.

Siehe nächstes Blatt!

Lassen wir `StiffVanDerPol.m` laufen, so beobachten wir, dass der explizite Löser `ode45` viel länger braucht als der implizite `ode23s`. In Abb. 5 sehen wir den Grund dafür: `ode45` macht $\sim 10'000$ mal kleinere Schritte als `ode23s` wo die Lösung glatt ist (d.h. wenig variiert)! Also obwohl ein Schritt mit dem impliziten Verfahren `ode23s` massiv teurer ist, gewinnt man viel an Rechenzeit verglichen mit dem expliziten Löser `ode45` der sehr viele kleine Schritte (aus Stabilitätsgründen) machen muss .

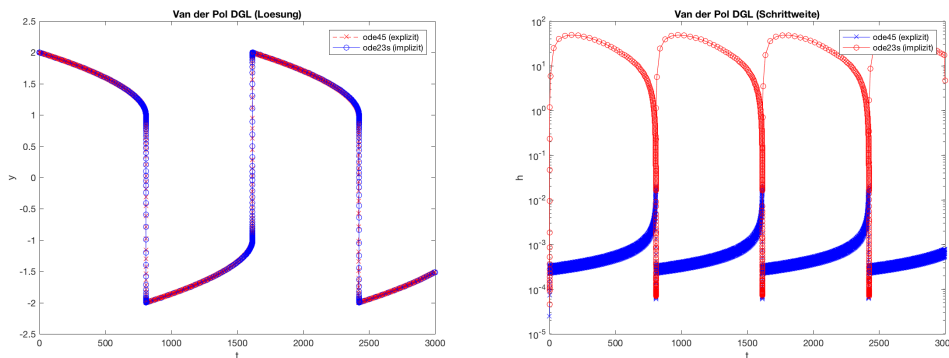


Abbildung 4 – Lösung $y(t)$ links und Schrittweite h rechts.

- b) Da die Van der Pol Diff.-Gl. nichtlinear ist, müssen wir wie in den Vorlesungsnotizen (Kap. 5 Seite 14) linearisieren. Die Jacobi Matrix von $f(t, \mathbf{y}(t))$ lautet

$$J(t, \mathbf{y}) = \begin{pmatrix} 0 & 1 \\ -2\mu y_0(t)y_1(t) - 1 & \mu(1 - y_1^2) \end{pmatrix}.$$

Setzen wir die Anfangszeit $t_0 = 0$ und die AWe ein, erhalten wir

$$J(t_0, \mathbf{y}_0) = \begin{pmatrix} 0 & 1 \\ -1 & -3000 \end{pmatrix}.$$

Die Eigenwerte dieser Matrix berechnet man sehr einfach zu

$$\lambda_{1,2} = \frac{-3000 \pm \sqrt{8999996}}{2}.$$

Damit lautet der Steifigkeitsparameter

$$S = \frac{\max_{i=1,\dots,n} |\operatorname{Re} \lambda_i|}{\min_{i=1,\dots,n} |\operatorname{Re} \lambda_i|} \approx 9 \times 10^6 \gg 1$$

Somit erwarten wir ein um den Punkt (t_0, y_0) lokal steifes Problem.

- c) In Abb. ist der Steifigkeitsparameter S als Funktion der Zeit gezeichnet. Wir beobachten, dass die Lösung hin und her oszilliert zwischen sehr steifen ($S \gg 1$) und nicht steifen Verhalten. Somit ist klar, dass bei nichtlinearen Problemen sowohl steife als auch nicht steife Abschnitte auftreten können. Weiter erklärt die Abb. wieso der explizite Löser so kleine Schrittweiten machen muss.

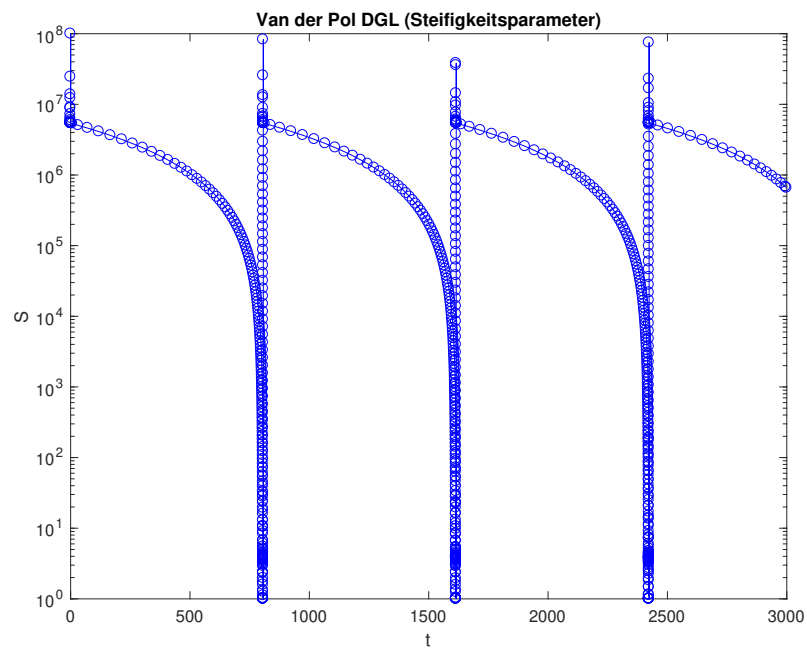


Abbildung 5 – Steifigkeitsparameter als Funktion der Zeit.