10 Gewöhnliche Differentialgleichungen

Wichtiger Hinweis: Hier im Skriptum wird das Thema kurz und überblicksmäßig behandelt. Bitte sehen Sie sich auch die Vorlesungsfolien an, sie zeigen zusätzliches Material: Bilder, Erklärungen, Zusammenfassungen. Auch die Übungsunterlagen geben vertiefend Erläuterungen.

10.1 Aufgabenstellung, Beispiele

10.1.1 Differentialgleichungen erster Ordnung

Explizite gewöhnliche Differentialgleichung 1. Ordnung mit Anfangsbedingung
Gesucht ist eine Funktion $y(x)$, welche

$$\begin{aligned} y' &= f(x,y) \\ y(x_0) &= y_0 \end{aligned}$$

erfüllt. Wenn f in x stetig ist und einer Lipschitzbedingung genügt, dann existiert eine eindeutige Lösung in der Umgebung des Anfangspunktes x_0.

Die Schreibweise $y = y(x)$ ist Standard für die typische allgemeine Funktion und Variable. Je nach Anwendungsgebiet beschreiben Differentialgleichungen auch *zeitliche* Änderungen. Gesucht ist dann eine Funktion der Zeit, also $y = y(t)$ oder auch $x = x(t)$: Weg x als Funktion der Zeit t. Sie kennen aus der Physik die Schreibweise $\dot{x}(t)$ für Ableitungen nach der Zeit.

In der Schreibweise mit $y = y(t)$ lautet die Aufgabenstellung daher

> Gesucht ist eine Funktion $y(t)$, welche
>
> $$\begin{aligned} \dot{y} &= f(t,y) \\ y(t_0) &= y_0 \end{aligned}$$
>
> erfüllt.

Lassen Sie sich also nicht verwirren: je nach Kontext heißt die gesuchte Funktion y oder x oder (siehe zweites Beispiel) s, hängt von x oder t ab, und Ableitungen werden durch Punkte oder Striche bezeichnet.

Beispiel 1: Exponentielles Wachstum einer Population Eine Bakterienpopulation befinde sich in einer Nährflüssigkeit und habe zur Zeit $t > 0$ die Größe $y(t)$. Bei $t = 0$ hatte die Population die Größe $y(0) = y_0$. Ausgehend yom Zeitpunkt t wird sich nach Ablauf der Zeitspanne Δt die Population um $\Delta y = y(t + \Delta t) - y(t)$ Mitglieder vermehrt haben. Nun ist es sinnvoll, anzunehmen, dass bei kleinen Zeitspannen Δt die Vermehrung etwa proportional zu dem Bestand $y(t)$ und zu der Zeitspanne Δt ist. Also

$$\Delta y = y(t + \Delta t) - y(t) \sim a \cdot y(t) \cdot \Delta t.$$

mit einer Proportionalitätskonstanten $a > 0$. Berechnen wir nun die Ableitung mit Hilfe des Differentialquotienten, erhalten wir

$$\dot{y}(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left(y(t + \Delta t) - y(t)\right) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} a \cdot y(t) \cdot \Delta t = a\, y(t).$$

Neben der Differentialgleichung muss die Funktion $y(t)$ noch die *Anfangsbedingung* $y(0) = y_0$ erfüllen. Dieses *Anfangswertproblem*

$$\begin{aligned} \dot{y}(t) &= ay(t) \\ y(0) &= y_0 \end{aligned}$$

beschreibt das Wachstumsgesetz der Bakterienpopulation. Sie kennen sicher die Lösung: Es handelt sich um *exponentielles Wachstum*. Die Lösung des Anfangswertproblems lautet

$$y(t) = y_0 \exp(at) \ .$$

Dieselbe Differentialgleichung ist auch das Modell eines Sparbuchs, Kredits, oder des radioaktiven Zerfalls (mit $a < 0$).

Beschränktes Wachstum einer Population:

Sind die Ressourcen nicht unbegrenzt, so spricht man vom beschränkten Wachstum. Beim beschränkten Wachstum ist die Änderungsrate proportional zum so genannten Sättigungsmanko (Differenz zwischen oberen (bzw. unteren) Schranke S und dem alten Bestand $y(t)$). Wir erhalten damit folgende Differentialgleichung

$$\dot{y}(t) = k(S - y(t)), \quad t > 0.$$

Die zugehörige Lösung lautet

$$y(t) = S - ce^{-kt}, \quad t \geq 0, \, c = S - y(0).$$

10.1.2 Höherere Ordnung und Systeme von Differentialgleichungen

Beispiel 2: Bewegung im Schwerefeld der Erde: Ein Körper der Masse m befindet sich im Schwerefeld der Erde (Masse M). Ist s der Abstand Körper—Erdmittelpunkt, dann gilt für die Anziehungskraft nach dem Gravitationsgesetz:

$$K = \gamma \frac{M\, m}{s^2}, \quad \gamma = \text{Gravitationskonstante}.$$

Zur Zeit $t = 0$ ist der Abstand des Körpers s_0. Er bewegt sich mit der Anfangsgeschwindigkeit v_0, und zwar nach oben, wenn v_0 positiv ist, und nach unten, wenn v_0 negativ ist. Nach t Zeiteinheiten hat er den Abstand $s(t)$ vom Erdmittelpunkt. Luftreibung werde vernachlässigt. Nach dem Newtonschen Kraftgesetz folgt daher die Differentialgleichung *zweiter Ordnung*

$$m\ddot{s}(t) = -\gamma \frac{M\, m}{s(t)^2}. \tag{25}$$

Diese Differentialgleichung hat viele Lösungen. Um eine festzulegen, fordert man zusätzlich folgende Anfangsbedingungen (Anfangsposition, Anfangsgeschwindigkeit):

$$s(0) = s_0, \quad \dot{s}(0) = v_0.$$

Hier lässt sich eine explizite Lösung in der Form $s(t) = \ldots$ nicht mehr angeben. *Numerische Verfahren* können aber leicht Lösungen in jeder gewünschten Genauigkeit liefern.

Nur die allereinfachsten Differentialgleichungen lassen sich analytisch lösen. In der Praxis ist man zumeist auf numerische Lösungsverfahren angewiesen.

In der Differentialgleichung 25 treten *zweite* Ableitungen auf; es handelt sich um eine *Differentialgleichung zweiter Ordnung*. Führen wir zusätzlich zum Weg $s(t)$ noch die Geschwindigkeit $v(t)$ als zweite gesuchte Funktion ein, dann lässt sich diese Gleichung äquivalent als ein *System zweier Differentialgleichungen erster Ordnung* schreiben:

$$\begin{aligned} \dot{s}(t) &= v(t) \\ \dot{v}(t) &= -\gamma \frac{M}{s(t)^2}. \end{aligned}$$

Begründung: Die erste Gleichung definiert die Geschwindigkeit v als zeitliche Ableitung des Weges s; die zweite Gleichung formuliert, weil $\dot{v}(t) = \ddot{s}(t)$, das Kraftgesetz 25.

Die allgemeine Aufgabenstellung lautet:

System von n gewöhnlichen Differentialgleichungen 1. Ordnung, Anfangswertproblem
Gesucht sind n Funktionen $y_1(x), \ldots, y_n(x)$, welche das System

$$\left.\begin{array}{rcl} y_1' &=& f_1(x, y_1, \ldots, y_n) \\ y_2' &=& f_2(x, y_1, \ldots, y_n) \\ \vdots & & \vdots \\ y_n' &=& f_n(x, y_1, \ldots, y_n) \end{array}\right\} \text{Differentialgleichungen}$$

$$\left.\begin{array}{rcl} y_1(x_0) &=& y_{10} \\ y_2(x_0) &=& y_{20} \\ \vdots & & \vdots \\ y_n(x_0) &=& y_{n0} \end{array}\right\} \text{Anfangsbedingungen}$$

erfüllen.

Für $n = 2$ ist eine Schreibweise mit zwei Funktionen $y(x)$ und $z(x)$ (oder, so wie im vorigen Beispiel, $s(t)$ und $v(t)$) einfacher als die Index-Schreibweise $y_1(x)$ und $y_2(x)$: Differentialgleichungen

$$\begin{aligned} y' &= f(x, y) \\ z' &= g(x, y) \end{aligned}$$

Bei mehr Gleichungen gehen aber rasch die Buchstaben aus, und die Formeln für Rechenverfahren werden unübersichtlich. Eine systematische Schreibweise, die auch Computerprogramme sehr vereinfacht, faßt Funktionen vektoriell zusammen:

$$\mathbf{y}(x) = \begin{bmatrix} y_1(x) \\ y_2(x) \\ \vdots \\ y_n(x) \end{bmatrix}, \quad \mathbf{y}\,'(x) = \begin{bmatrix} y_1'(x) \\ y_2'(x) \\ \vdots \\ y_n'(x) \end{bmatrix}, \quad \mathbf{f}(x, \mathbf{y}) = \begin{bmatrix} f_1(x, y_1, \ldots, y_n) \\ f_2(x, y_1, \ldots, y_n) \\ \vdots \\ f_n(x, y_1, \ldots, y_n) \end{bmatrix}$$

Dann lautet das Anfangswertproblem schlicht

$$\begin{aligned} \mathbf{y}\,' &= \mathbf{f}(x, \mathbf{y}) \\ \mathbf{y}(x_0) &= \mathbf{y}_0 \end{aligned}$$

Gewöhnliche Differentialgleichung höherer Ordnung
Eine DG höherer Ordnung lässt sich durch Einführen von Hilfsfunktionen in ein äquivalentes System von DGs 1. Ordnung transformieren.
Diese Umformung ist bei praktischen Aufgaben häufig notwendig (und auch eine beliebte Prüfungsfrage). Die Übungsunterlagen bringen eine ausführliche Anleitung und Beispiele.

10.2 Numerische Verfahren

Explizite Einschrittverfahren

Wichtige Verfahren sind das Eulersche Polygonzugverfahren (weil es das einfachste ist: Verfahren 1. Ordnung), das Verfahren von Heun und das modifizierte Euler-Verfahren (weil sie genauer sind: Verfahren 2. Ordnung) und das klassische Runge-Kutta-Verfahren (weil man damit in der Praxis oft rechnet; Verfahren 4. Ordnung).

Zur numerischen Lösung einer DG bestimmt ein explizites Einschrittverfahren ausgehend von den Anfangsbedingungen eine Folge von Wertepaaren $(x_0, y_0), (x_1, y_1), (x_2, y_2), \ldots$, die den Verlauf der gesuchten Funktion $y = y(x)$ annähern sollen. Schema:

Wähle Schrittweite h und maximale Schrittzahl N;
setze x_0 und y_0 laut Anfangsbedingung;
für $i = 0, 1, \ldots, N$
$x_{i+1} = x_i + h$;
$y_{i+1} = y_i + hF(x_i, y_i, h)$.

Die Funktion $F(x, y, h)$ heißt die *Verfahrensfunktion* des jeweiligen Verfahrens. Geometrisch interpretiert gibt $F(x, y, h)$ die *Fortschreitungsrichtung*. Beim klassischen Euler-Verfahren (Eulersche Polygonzugverfahren) ist sie gleich der Tangentensteigung im Ausgangspunkt,

$$F(x, y, h) = f(x, y),$$

beim modifizierten Euler-Verfahren

$$F(x, y, h) = f\left(x + \frac{h}{2}, y + \frac{h}{2} f(x, y)\right),$$

beim Verfahren von Heun

$$F(x, y, h) = \frac{1}{2}(k_1 + k_2)$$

mit

$$\begin{aligned} k_1 &= f(x, y) \\ k_2 &= f(x + h, y + hf(x, y)), \end{aligned}$$

beim klassischen Verfahren von Runge-Kutta

$$F(x, y, h) = \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

mit

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

10.3 Illustrationen zu einem einfachen Anfangswertproblem

Geometrische Interpretation
Eine Differentialgleichung gibt ein Richtungsfeld vor

Gegeben sei die Differentialgleichung

$$y' = xy/4 - 1\ .$$

Sie ordnet jedem Punkt (x, y) in der Ebene eine Steigung y' zu. Fassen wir diese Steigungen als Richtungsvektoren (mit einheitlicher Länge) auf, dann erhalten wir ein *Richtungsfeld*, wie in Abbildung 9 dargestellt.

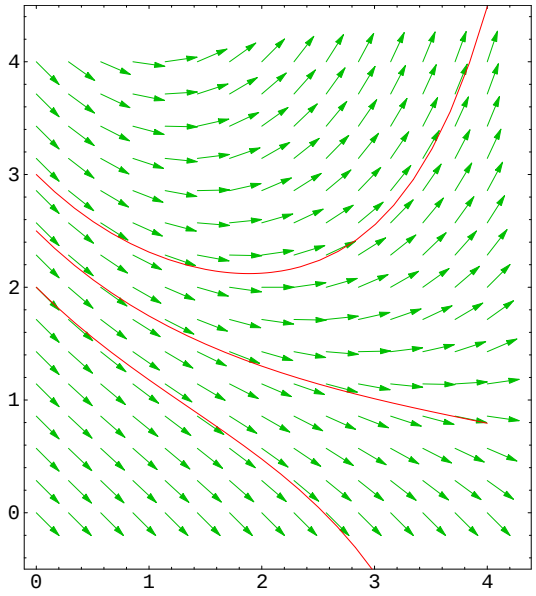


Abbildung 9: Die Differentialgleichung gibt ein Richtungsfeld vor. Drei Lösungen zu verschiedenen Anfangsbedingungen sind eingetragen

Eine Lösungskurve startet an einem Punkt (x_0, y_0), der durch die Anfangsbedingung festgelegt ist, in der dort vorgegebenen Richtung. Im weiteren Verlauf folgt die Kurve immer den Richtungspfeilen. Sie orientiert sich also in jedem Punkt (x, y), den sie erreicht, völlig opportunistisch an der dort herrschenden Richtung.

Ein Einschrittverfahren, wie beispielsweise das Eulersche Polygonzugverfahren, startet ebenfalls am Punkt (x_0, y_0). Das Eulersche Polygonzugverfahren schlägt auch genau die in (x_0, y_0) gegebene Richtung ein und steht eine Schrittweite h lang stur zu dieser Entscheidung. Es gelangt damit an einen Punkt (x_1, y_1). Dort erst überdenkt es seinen Weiterweg und orientiert sich neu, wieder genau nach der in (x_1, y_1) herrschenden Richtung. Je kleiner die Schrittweite, also um so öfter sich das Verfahren den herrschenden Gegebenheiten anpasst, um so besser sollte es der exakten Lösungskurve folgen. Die Abbildung 10 illustriert diese Gedanken. Eine Präzisierung der Idee und eine Untersuchung der Fehler bringt das nächsten Kapitel.

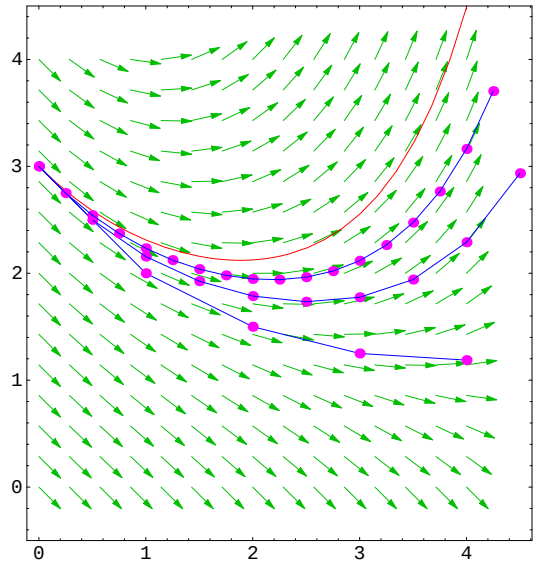


Abbildung 10: Für die Differentialgleichung $y' = xy/4 - 1$ mit Anfangsbedingung $y(0) = 3$ sind die exakte Lösung sowie drei Näherungen nach dem Eulerschen Polygonzugverfahren mit Schrittweiten $h = 1; \frac{1}{2}; \frac{1}{4}$ eingetragen

Das Eulersche Polygonzugverfahren ist allerdings kurzsichtig, in dem Sinn, dass es als Fortschreitungsrichtung (Verfahrensfunktion) einfach die am Ausgangspunkt gegebene Richtung wählt. Das modifizierte Eulerverfahren versucht, vorausschauender zu agieren: es tastet sich zuerst eine halbe Schrittweite voran, erkundet die dort gegebene Richtung und wählt diese als Fortschreiterichtung (Verfahrensfunktion). Siehe dazu die Abbildungen 11 und 12.

Ähnlich geht das Verfahren von Heun vor. Es macht versuchsweise einen Schritt in die selbe Richtung wie das gewöhnliche Polygonzugverfahren, erkundet am Zielpunkt die Richtung und wählt als endgültige Fortschreitungsrichtung den Mittelwert der Richtungen an Start- und (vorläufigem) Zielpunkt. Siehe dazu Abbildung 13

Es gibt in dieser Art noch weitere Möglichkeiten, das einfache Eulersche Polygonzugverfahren zu verbessern. In der Literatur werden dafür nicht einheitliche Namen verwendet. Auch die folgende Methode geht auf Heun zurück und wird gelegentlich als Heunsches Verfahren bezeichnet:

$$F(x, y, h) = \frac{1}{4}(k_1 + 3k_2)$$

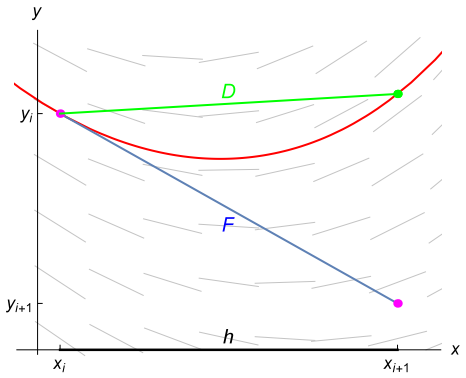


Abbildung 11: Verfahrensfunktion F des expliziten Euler-Verfahrens und exakte Richtung D.

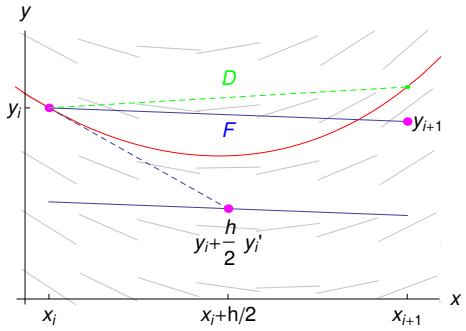


Abbildung 12: Wahl der Verfahrensfunktion F (geometrisch: Fortschreit-Richtung) beim modifizierten Eulerverfahren. F nähert den exakten Differenzenquotient D besser an als beim expliziten Euler-Verfahren, vergleiche Abbildung 11.

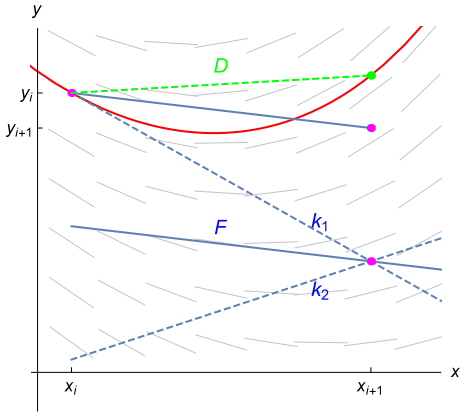


Abbildung 13: Wahl der Verfahrensfunktion F (geometrisch: Fortschreit-Richtung) beim Heun-Verfahren: Mittelwert aus Anfangsrichtung und Richtung am (genäherten) Endpunkt. F nähert den exakten Differenzenquotient ähnlich genau wie das modifizierte Euler-Verfahren D an.

mit

$$\begin{aligned} k_1 &= f(x,y) \\ k_2 &= f(x+\frac{2}{3}h, y+\frac{2}{3}hf(x,y)) \end{aligned}$$

Das klassische Runge-Kutta-Verfahren treibt das Wechselspiel aus versuchsweisem Vortasten und Erkunden der Richtung zur Perfektion: es berechnet vier verschiedene Steigungen $k_1, \ldots, k_4$, eine am Ausgangspunkt, zwei in der Mitte und eine am Ziel, und wählt als Verfahrensfunktion ein gewogenes Mittel der so berechneten Richtungen.

Allgemein bezeichnet man Methoden dieser Art als Runge-Kutta-Verfahren (deswegen oben der Zusatz „klassisch“). Moderne Runge-Kutta-Verfahren versuchen, mit möglichst geringem Rechenaufwand eine hohe Ordnung und gleichzeitig Abschätzungen über die Größe des Fehlers oder die optimale Schrittweite zu gewinnen.

10.4 Diskretisierungsfehler

Wenn ein numerisches Verfahren näherungsweise Werte y_i der Lösung y eines Anfangswertproblems berechnen soll, sind Fragen nach der Größe des Fehlers und nach der Anzahl der korrekten Stellen im Resultat wichtig.

Zwei Arten von Fehlern sind beim Einschrittverfahren zu unterscheiden:

- In jedem einzelnen Schritt wählt die Verfahrensfunktion eine Fortschreitrichtung, die normalerweise nicht exakt den Wert der Lösung trifft. Dieser Richtungsunterschied ist der lokaler Diskretisierungsfehler.
- Die Effekte der lokalen Diskretisierungsfehler in den einzelnen Schritten akkumulieren sich. Daraus resultierend weicht die numerische Lösung von der exakten Lösung ab. Letztlich ist natürlich dieser Fehler, der *globale Diskretisierungsfehler* interessant. Er lässt sich mit dem lokalen Diskretisierungsfehler abschätzen.

Die Abbildung 11, 12 und 13 zeigen unterschiedlich große lokale Diskretisierungsfehler. Eingezeichnet ist jeweils die von der Verfahrensfunktion F gewählte Richtung, sowie (für die von y_i ausgehenden exakte Lösung) die vom Differenzenquotienten D bestimmte exakte Richtung.

Sowohl F als auch D hängen von x_i, y_i und der Schrittweite h ab und nähern sich für $h \to 0$ der Tangentensteigung $f(x_i, y_i)$ an. Es gilt

$$\lim_{h \to 0} D(x_m, y_m, h) = f(x_m, y_m).$$

Die Abweichung d der von der Verfahrensfunktion F bestimmten Richtung von der exakten Richtung D

$$d(x_m, y_m, h) = |F(x_m, y_m, h) - D(x_m, y_m, h)|$$

heißt *lokaler Diskretisierungsfehler*. Natürlich soll dieser Fehler um so kleiner werden, je kleiner die Schrittweite h gewählt wird. Information darüber, wie stark d von der Schrittweite abhängt, gibt die *Ordnung des Verfahrens*.

Ordnung eines Einschrittverfahrens
Die größte natürliche Zahl p mit

$$d(x_m, y_m, h) = O(h^p)$$

heißt Ordnung des Verfahrens.

Ordnung 1: Fehler d proportional Schrittweite
Ordnung 2: Fehler d proportional dem Quadrat der Schrittweite
usw.

Die Ordnung des lokalen Diskretisierungsfehlers ist wichtig, weil sie sich durch mathematische Methoden (wie etwa Taylorreihenentwicklung) abschätzen lässt. Die Effekte der lokalen Diskretisierungsfehler in den einzelnen Schritten akkumulieren sich. Daraus resultierend weicht die numerische Lösung von der exakten Lösung ab. Letztlich ist natürlich dieser Fehler, der *globale Diskretisierungsfehler* interessant. Er lässt sich mittels dem lokalen Diskretisierungsfehler abschätzen.

Lokaler und globaler Diskretisierungsfehler
Der Fehler d zwischen Verfahrensfunktion F und exakter Richtung D

$$d(x_m, y_m, h) = |F(x_m, y_m, h) - D(x_m, y_m, h)|$$

heißt *lokaler Diskretisierungsfehler* im Punkt (x_m, y_m).
Ist Y die exakte Lösung der Anfangswertaufgabe

$$y' = f(x, y), \quad y(x_0) = y_0 \ ,$$

und y_m die Näherungslösung an der Stelle x_m, so nennt man den Fehler

$$g(x_m, h) = |y_m - Y(x_m)|$$

den *globalen Diskretisierungsfehler* .

Von jedem Einschrittverfahren verlangen wir dann, daß an einer gegebenen Stelle x der Fehler $g(x, h)$ mit h gegen Null geht und nennen die größte natürliche Zahl p mit

$$g(x, h) = |y(x, h) - Y(x)| = O(h^p)$$

die Konvergenzordnung des Verfahrens. Es gilt folgender Zusammenhang zwischen dem lokalen und dem globalen Diskretisierungsfehler:

Konvergenz des Einschrittverfahrens
Ist der lokale Diskretisierungsfehler von der Ordnung $p \geq 1$ und genügt F einer Lipschitzbedingung, so konvergiert das Einschrittverfahren mit Ordnung p.

Ordnung einzelner Verfahren
Ordnung 1: Eulersches Polygonzugverfahren
Ordnung 2: Modifiziertes Euler-Verfahren, Verfahren von Heun
Ordnung 4: Klassisches Runge-Kutta-Verfahren

10.5 Wann soll man welches Verfahren benützen

Suchen Sie z.B. `ode45` in der MATLAB-Hilfe, so listet MATLAB Ihnen folgende Gleichungslöser (solver) für Differentialgleichungen auf: `ode23`, `ode45`, `ode113`, `ode15s`, `ode23s`, `ode23t`, `ode23tb`. Welches Verfahren soll man benützen? Das erste Auswahlkriterium ist die Konvergenzordnung (je höher, desto genauer). Ist aber die zu erwartende Lösung allerdings nicht glatt (das heißt zum Beispiel, f hat Sprungstellen, Ecken, oder höhere Ableitungen existieren nicht), arbeitet ein Verfahren mit niedriger Konvergenzordnung schneller und braucht weniger Rechenleistung als ein Verfahren mit hoher Konvergenzordnung.

Auswahlkriterien des Solvers

- Konvergenzordnung - Genauigkeit der Approximation: MATLABs `ode45` ist ein Verfahren der Ordnung 5, allerdings nur, wenn die Funktion, die Sie an `ode45` übergeben, auch genügend oft differenzierbar ist. Wenn nicht, dann arbeitet `ode23` effizienter.
- Steifheit des Systems: Ist das System steif, sollte man `ode15s`, `ode23t` oder `ode23tb` verwenden.
- Komplizierte rechte Seite: explizite Solver (`ode45`, `ode23`, `ode23`, `ode113`) rechnen oft schneller, weil sie – im Gegensatz zu impliziten Verfahren – nicht in jedem Schritt Gleichungssysteme lösen müssen. Die MATLAB-Hilfe empfiehlt dazu: *"ode113 can be more efficient than ode45 [...] when the ODE function is expensive to evaluate"*.

Die folgenden Abschnitte gehen auf Probleme bei der Anwendung numerischer Lösungsverfahren ein. Wichtige Themen sind:

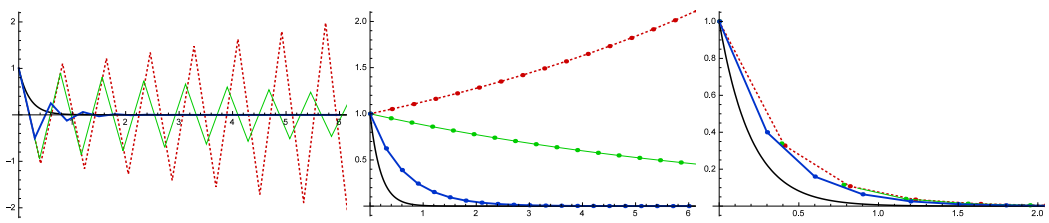
- explizite im Vergleich zu impliziten Verfahren
- Stabilität
- Steifheit

10.5.1 Stabilität

Zu diesen Punkt möchte ich als erstes den Unterschied zwischen expliziten Verfahren und impliziten Verfahren erklären. Gegeben ist folgende Differentialgleichung

$$\begin{cases} y'(x) &= -5\,y(x), \\ y(0) &= 1. \end{cases}$$

In den nachfolgenden Bildern wurde die exakte Lösung e^{-5x} (schwarz gezeichnete Kurve) mit verschiedenen Verfahren approximiert. Dabei wurden jeweils als Schrittweiten 0,3 (blau gezeichnet), 0,39 (grün), und 0,41 (rot) gewählt. Von links nach rechts: explizites Euler-Verfahren, Verfahren von Heun, implizits Euler-Verfahren.



Schaut man sich die Graphen an, sieht man, dass das explizite Euler-Verfahren für Schrittweite 0,41 osziliert und divergiert. Auch das Heun-Verfahren divergiert bei dieser Schrittweite. Man sagt, die Verfahren sind *instabil* bei dieser Schrittweite.

Bei Schrittweiten $< 0{,}4$ geben die Verfahren zumindest grob qualitativ das Verhalten der exakten Lösung wieder: sie klingen exponentiell ab. Man bezeichnet dieses Verhalten als *stabil*.

> Ein Verfahren heißt stabil, wenn es exponentiell abklingende Lösungen qualitativ richtig berechnet. Stabilität kann von der Schrittweite abhängen.

- Es gibt verschiedene Definitionen für Stabilität eines numerischen Verfahrens.
- Allgemein kann man sich merken: Implizite Verfahren – immer stabil, expliziten Verfahren – nur bei kleinen Schrittweiten stabil.
- Stabilitätsprobleme treten vor allem bei *steifen Differentialgleichungen* auf. Für solche Probleme sind implizite Verfahren besser geeignet.

Wenn man die einzelnen Schritte des expliziten Euler-Verfahrens nachrechnet, erkennt man auch, was vorgeht: die Rechenschritte sind nämlich

$$\begin{aligned}
y_1 &= y_0 + h \cdot (-5) y_0 = (1-5h) y_0, \\
y_2 &= y_1 + h \cdot (-5) y_1 = (1-5h) y_1 = (1-5h)^2 y_0, \\
y_3 &= (1-5h) y_2 = \ldots = (1-5h)^3 y_0, \\
\ldots &\quad \ldots \quad \ldots \\
y_k &= (1-5h) y_{k-1} = \ldots = (1-5h)^k y_0.
\end{aligned}$$

Das explizite Euler-Verfahren berechnet also den jeweils nächsten Wert aus dem vorhergehenden Wert durch Multiplikation mit $(1-5h)$.

Für $h = 0{,}41$ ist $(1-5h) = -1{,}05$, und die Folge $(1-5h)^k$ strebt für $k \to \infty$ mit alternierendem Vorzeichen gegen $\pm\infty$.

Andererseits, für $h = 0{,}39$ ist $(1-5h) = -0{,}95$, und die Folge $(1-5h)^k$ pendelt zwar zwischen positiven und negativen Werten, strebt aber für $k \to \infty$ nach Null.

Analysiert man das implizite Euler Verfahren, so kommt man auf folgende Rechnung:

$$\begin{aligned}
y_1 &= y_0 + h \cdot (-5) y_1, \quad \to \quad y_1 = \frac{y_0}{(1+5h)}, \\
y_2 &= y_1 + h \cdot (-5) y_2, \quad \to \quad y_2 = \frac{y_1}{(1+5h)} = \frac{y_0}{(1+5h)^2}, \\
y_3 &= y_2 + h \cdot (-5) y_3, \quad \to \quad y_3 = \frac{y_2}{(1+5h)} = \frac{y_0}{(1+5h)^3}, \\
\ldots &\quad \ldots \quad \ldots \\
y_k &= \ldots = \frac{y_{k-1}}{(1+5h)} = \frac{y_0}{(1+5h)^k}.
\end{aligned}$$

Hier gilt für $h = 0{,}41$: $\frac{1}{(1+5h)} = 0{,}327$, und und die Folge $\frac{1}{(1+5h)^k}$ strebt für $k \to \infty$ monoton fallend nach Null.

Tatsächlich gilt hier: Für alle $h > 0$ strebt $\frac{1}{(1+5h)^k}$ monoton nach Null.

Analysiert man das Heun-Verfahren, kommt man auf folgende Folge:

$$\left\{ \left(1 - 5h + \frac{(5h)^2}{2} \right)^k y_0 : k \in \mathbb{N} \right\}.$$

Für $h = 0{,}41$ ist der Ausdruck innerhalb der Klammer 1,051 und deswegen wächst die Folge monoton an und divergiert.

Für $h = 0{,}39$ ist der Ausdruck innerhalb der Klammer 0,951, die Folge konvergiert nach Null.

Stabilitätsgebiet

Für systematische Untersuchungen der Stabilität ist nützlicher, eine allgemeinere Modellgleichung $y' = \lambda y$ mit Parameter λ zu betrachten. Dabei untersucht man auch komplexe λ; die zugehörigen exakten Lösungen sind gedämpfte Sinus-Schwingungen. Diese einfache Modellgleichung beschreibt also, je nach λ-Wert, die wichtigsten Prozesse, die in praktischen Anwendungen auftreten.

Je nach Kombination von h und λ können Verfahren stabil sein oder nicht. Dabei kommt es aber nur auf den Wert des Produkts $\xi = h \cdot \lambda$ an. Man definiert daher:

Das Stabilitätsgebiet eines Verfahrens ist die Menge der komplexen Zahlen $\xi = h \cdot \lambda$, für die es bei der Lösung der Testgleichung

$$y' = \lambda y, \quad y(0) = 1$$

bei fester Schrittweite h eine beschränkte Folge von Näherungen liefert.

Im Falle des expliziten Euler Verfahrens gilt

$$\begin{aligned} \mathcal{B} &= \left\{ \lambda \in \mathbb{C} : \lim_{k\to\infty} y_k^\lambda < \infty \right\} = \left\{ \lambda \in \mathbb{C} : \lim_{k\to\infty} (1+\lambda)^k < \infty \right\} \\ &= \{ \lambda \in \mathbb{C} : |1+\lambda| \le 1 \} = \left\{ a + ib = \lambda \in \mathbb{C} : (a+1)^2 + b^2 \le 1 \right\}. \end{aligned}$$

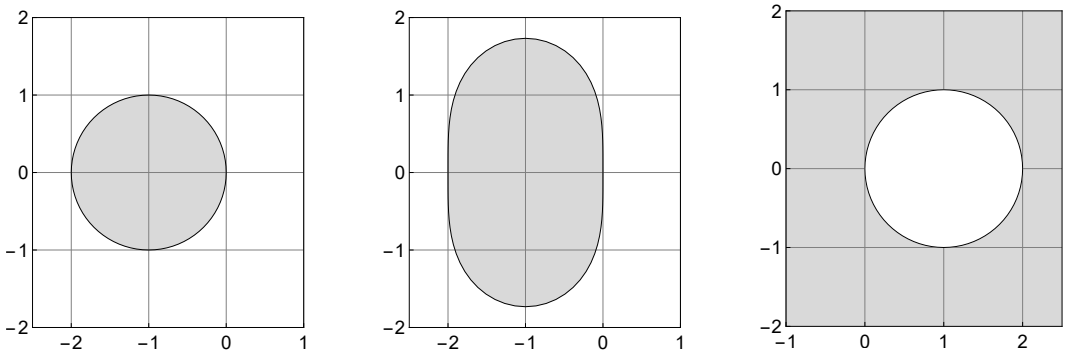
Im Falle des impliziten Euler Verfahrens gilt

$$\begin{aligned} \mathcal{B} &= \left\{ \lambda \in \mathbb{C} : \lim_{k\to\infty} y_k^\lambda < \infty \right\} = \left\{ \lambda \in \mathbb{C} : \lim_{k\to\infty} (1+\lambda)^{-k} < \infty \right\} \\ &= \left\{ a + ib = \lambda \in \mathbb{C} : (a-1)^2 + b^2 > 1 \right\}. \end{aligned}$$

Die Stabilitätsgebiete des expliziten, des modifizierten und des impliziten Eulerverfahrens sind hier (von links nach rechts) als graue Bereiche in der komplexen Zahlenebene dargestellt. Das Heun-Verfahren hat gleiches Stabilitätsgebiet wie das modifizierte Euler-Verfahren.

Praktisch relevant ist die linke Halbebene[16]; sie entspricht abklingenden Exponential- und Schwingungsprozessen. Für Stabilität muss λh in den grauen Bereichen liegen.

[16] In die rechte Halbebene fallen exponentiell anwachsende Prozesse. Dabei wachsen auch die Fehler aller numerischen Verfahren exponentiell an. Die Frage nach Stabilität ist hier nicht sinnvoll.



Bei expliziten Verfahren gibt es Einschränkungen: man muss die Schrittweite genügend klein wählen, sonst liegt λh außerhalb des Stabilitätsgebietes. Beim impliziten Verfahren liegt die gesamte linke Halbebene im Stabilitätsgebiet, man kann bei $\lambda \leq 0$ (abklingenden Exponential- und Schwingungsprozesse) die Schrittweite beliebig groß wählen.

10.5.2 Steifheit

Differentialgleichungen, die chemische oder physikalische Prozesse beschreiben, haben oft Lösungen, die sich aus sehr unterschiedlich schnell abklingenden Komponenten zusammensetzen. Das passiert, wenn Teilprozesse mir sehr unterschiedlichen Geschwindigkeiten ablaufen. Nehmen wir folgendes einfache Beispiel

$$\begin{aligned} \dot{y}_1(t) &= -y_1(t) + 50y_2(t), \\ \dot{y}_2(t) &= -70y_2(t), \end{aligned}$$

mit Anfangswerten $y_1(0) = 1$ und $y_2(0) = 10$. Da die Matrix

$$\begin{pmatrix} -1 & 50 \\ 0 & -70 \end{pmatrix}$$

Eigenwerte $\lambda_1 = -1$ und $\lambda_2 = -70$ mit Eigenvektoren $v_1 = (1, 0)^T$ und $v_2 = -(50, 96)^T$ hat, erhält man als Lösung

$$y_1(t) = 8.24638e^{-t} - 7.2464e^{-70t}, y_2(t) = 10e^{-70x}.$$

Um die am schnellsten abklingende Komponente mit einer Genauigkeit von 10^{-3} mittels einer numerischen Lösung zu berechnen, muss die Schrittweite so gewählt werden, dass e^{-70h} mit $F(-70h)$ auf fünf Stellen übereinstimmt. Aber nach einer relativen kurzen Zeit ist e^{-70t} verglichen zu e^{-t} praktisch völlig abgeklungen. Trotzdem muss der Solver auch dann noch mit der sehr kleinen Schrittweite rechnen, obwohl die e^{-70t}-Komponente eigentlich gar nicht mehr da ist.

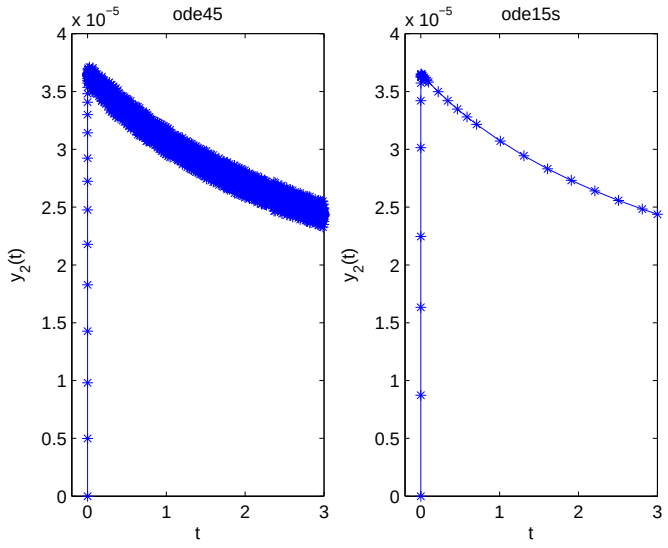
Definition 1: Ein lineares System

$$\begin{cases} \dot{x}(t) &= Ax(t), \\ x(0) &= x_0. \end{cases}$$

nennt man steif (stiff), falls der Wert

$$S := \frac{\max\{\Re(\lambda_j) : \lambda_j \text{ ist Eigenwert von } A \text{ mit negativen Realteil}\}}{\min\{\Re(\lambda_j) : \lambda_j \text{ ist Eigenwert von } A \text{ mit negativen Realteil}\}}$$

in der Größenordnung (oder größer als) 1000 ist.



Bemerkung 1: Man beachtet nur die Eigenwerte mit negativen Realteil, da diese Komponenten ergeben, die für grosse t gegen 0 laufen. Die Eigenwerte mit positiven Realteil führen zu Komponenten, die nicht konvergieren.

Ein weiteres (nicht konstruiertes) Beispiel ist folgende Differentialgleichung, die **Robertson Differentialgleichung**: Sei $y(0) = (1, 0, 0)$ und

$$\begin{aligned} \dot{y}_1(t) &= -0.04\, y_1(t) \\ &\quad +10^4 y_2(t)\, y_3(t), \\ \dot{y}_2(t) &= 0.04\, y_1(t) - 10^4 y_2(t)\, y_3(t) \\ &\quad -3 \times 10^7 y_2^2(t), \\ \dot{y}_3(t) &= 3 \times 10^7 y_2^2(t). \end{aligned}$$

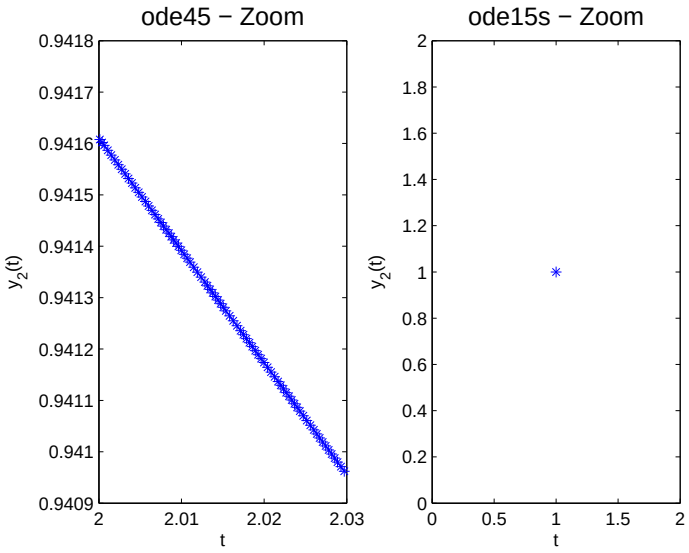
Am folgenden Bild sehen sie wie einmal die Differentialgleichung mit dem `ode45` numerisch berechnet, das andere mal mit dem `ode15s`. Obwohl der `ode45` ein Verfahren höherer Ordnung verwendet, konvergiert das Verfahren schlechter als das implizite Euler Verfahren, das der `ode15s` solver verwendet. Im zweiten Bild wurde ein Teilinterval herausgegriffen. In dem Zeitraum in dem der `ode45` Solver 81 Punkte berechnete, berechnete der `ode15s` solver einen Punkt.

Woran sieht man einen System an ob es sinnvoll wäre einen einfachen aber impliziten Solver zu verwenden.

Gegeben sei die lineare Differentialgleichung zweiter Ordnung

$$\begin{cases} \dot{x}(t) &= \begin{pmatrix} 0.5 & -0.5 \\ 10 & 10 \end{pmatrix} x(t), \\ x(0) &= \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \end{cases} \tag{26}$$

Löst man dieses mit dem `ode45` solver auf dem Zeitinterval $[0, 20]$ diskretisiert MatLab die Zeit auf 265 Punkten, der `ode15s` solver berechnet die Funktion an nur 65 Punkten bei gleicher Genauigkeit. Schaut man das System genauer an, sieht man das die Matrix zwei Eigenwerte



-0.5 und -10 hat. Da der `ode45` solver ein explizites Verfahren ist, müssen die Zeitschritte in Richtung des Eigenwertes -10 aus Stabilitätsproblemen sehr klein gewählt werden, andererseits verläuft die Lösung in dieser Richtung sehr glatt und konvergiert sehr schnell gegen Null, sodass man eigentlich mit sehr großen Schritten vorangehen könnte. Der `ode15s` solver kann eine Schrittweite wählen die dem Problem angepasst ist.

Die meisten Systeme sind nicht linear, wie kann man aber jetzt den Begriff Steifheit (stiffness) auf ein nichtlineares System übertragen. Sei

$$\begin{cases} \dot{x}(t) &= f(x(t)), \\ x(0) &= x_0. \end{cases} \tag{27}$$

Die Steifigkeit für das linearisierte System zu definieren, indem man das *lokale* Verhalten der exakten Lösung $x(t)$ in der Umgebung eines Punktes t_0 analysiert. Sei

$$x(t) \quad = \quad x(t_0) + z(t), \quad t_0 \leq t \leq t_0 + h,$$

das linearisierte System an der Stelle t_0. Entwickeln wir System (27) mittels der Taylor approximation so folgt

$$\dot{x}(t) \quad = \quad f(x(t_0)) + \nabla f(x(t_0))(x(t_0) - x(t)) + O(x(t_0) - x(t)).$$

Setzt man für $x(t)$ die Näherung $x(t_0) + z(t)$ ein, erhält man

$$\dot{x}(t) \quad = \quad f(x(t_0)) + \nabla f(x(t_0))z(t) + O(x(t_0) - x(t)).$$

Andererseits gilt $\dot{x}(t) = \dot{x}(t_0) + \dot{z}(t) = f(x(t_0)) + \dot{z}(t)$. Dies oben eingesetzt ergibt

$$f(x(t_0)) + \dot{z}(t) \quad = \quad f(x(t_0)) + \nabla f(x(t_0))z(t) + O(x(t_0) - x(t)),$$

und damit

$$\dot{z}(t) \quad = \quad \nabla f(x(t_0))z(t).$$

Die infinitesimale Änderung zum Zeitpunkt t_0 löst also das obige lineare System. Andererseits kann man für ein lineares System die Steifigkeit wie oben definieren. Diese Definition kann man durch die obige Rechnung auch auf nichtlineare System übertragen.

Definition 2: Sei

$$A(x_0) := \frac{\partial f}{\partial x}\Big|_{x=x_0}.$$

und

$$S(x_0) := \frac{\max\{\Re(\lambda_j) : \lambda_j \text{ ist Eigenwert von } A(x_0) \text{ mit negativen Realteil}\}}{\min\{\Re(\lambda_j) : \lambda_j \text{ ist Eigenwert von } A(x_0) \text{ mit negativen Realteil}\}}$$

Ein System nennen wir steif in x_0, falls $S(x_0)$ in der Größenordnung von 1000 oder größer ist.

Beispiel 3: Robertson Differentialgleichung Sei $y(0) = (1, 0, 0)$ und

$$\begin{aligned}
\dot{y}_1(t) &= -0.04\, y_1(t) + 10^4 y_2(t)\, y_3(t), \\
\dot{y}_2(t) &= 0.04\, y_1(t) - 10^4 y_2(t)\, y_3(t) - 3 \times 10^7 y_2^2(t), \\
\dot{y}_3(t) &= 3 \times 10^7 y_2^2(t).
\end{aligned}$$

Das heißt mit $\mathbf{y} = (y_1, y_2, y_3)^T$ können wir auch schreiben

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

mit

$$f(\mathbf{x}) = f\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ f_3(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} -0.04\, x_1 + 10^4 x_2\, x_3 \\ 0.04\, x_1 - 10^4 x_2\, x_3 - 3 \times 10^7 x_2^2, \\ 3 \times 10^7 x_2^2 \end{pmatrix}$$

Die Jacobi Matrix lautet jetzt

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{pmatrix} \frac{df_1}{dx_1} & \frac{df_1}{dx_2} & \frac{df_1}{dx_3} \\ \frac{df_2}{dx_1} & \frac{df_2}{dx_2} & \frac{df_3}{dx_3} \\ \frac{df_3}{dx_1} & \frac{df_3}{dx_2} & \frac{df_3}{dx_3} \end{pmatrix} = \begin{pmatrix} -0.04 & 10^4\, x_3 & 10^4\, x_2 \\ 0.04 & -10^4 x_3 - 6 \cdot 10^7\, x_2 & -10^4 x_2 \\ 0 & 6 \cdot 10^7\, x_2 & 0 \end{pmatrix}$$

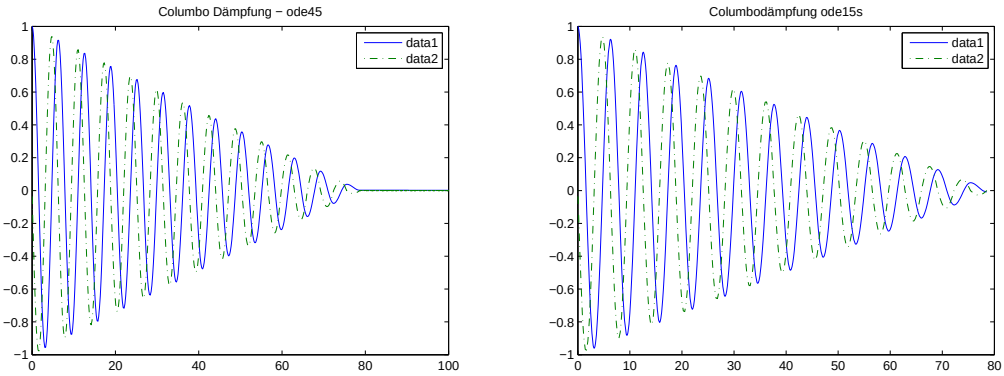
Berechnet man die Eigenwerte zum Punkt $\mathbf{x} = (1, 0, 0)^T$ so erhält man in Mathematika die Werte $-0.04, 0, 0$, zum Punkt $\mathbf{x} = (1, 0.05, 0.05)^T$, die Werte $-3 \cdot 10^6, -500, 1.4 \cdot 10^{-14}$, zum Punkt $\mathbf{x} = (1, 0.1, 0.1)^T$, die Werte $-6.0 \cdot 10^6, -1000, 5.26 \cdot 10^{-14}$). und $\mathbf{x} = (1, 1, 1)^T$ erhält man die Werte $-6 \cdot 10^7, -10000, 1.94 \cdot 10^{-12}$. Der Quotient zwischen den kleinsten und größten negativen reellen Eigenwert lautet jeweils

$$\begin{aligned}
S((1,0,0)^T) &\sim \infty, \\
S((1,0.1,0.1)^T) &\sim \frac{3 \cdot 10^6}{500} \sim 10^4, \\
S((1,0.1,0.1)^T) &\sim \frac{6.0 \cdot 10^6}{1000} \sim 10^3, \\
S((1,1,1)^T) &\sim \frac{6 \cdot 10^7}{10000} \sim 5 \cdot 10^2.
\end{aligned}$$

Beispiel 4: Die Coloumb-Dämpfung Gegeben sei folgende Differentialgleichung zweiter Ordnung:

$$m\ddot{x} + cx = -\mu m g\, \operatorname{sign}(\dot{x}). \tag{28}$$

Die Differentialgleichung wurde links mit dem ode45 solver berechnet. Die Anzahl der Stützstellen war 112221, die Differentialgleichung wurde rechts mit dem ode15s solver berechnet. Die Anzahl der Stützstellen war 945. Der Grund ist die Funktion $x \mapsto \operatorname{sign}(x)$, die in 0 nicht differenzierbar ist.



Eine Schwierigkeit bei impliziten Systemen ist die Auswertung der linken Seite. Angenommen, man berechnet das Beispiel oben mit den impliziten Euler Scheme. Es wird also zu jeden Zeitschritt folgendes Gleichungssystem gelöst:

$$\hat{\mathbf{y}}_k \quad = \quad \begin{pmatrix} \hat{y}_k^1 \\ \hat{y}_k^2 \end{pmatrix} = \hat{\mathbf{y}}_{k-1} + h \begin{pmatrix} \hat{y}_k^2 \\ -c\,\hat{y}_k^1 - \nu m g\,\mathrm{sign}(c\,\hat{y}_k^1) \end{pmatrix}.$$

Dieses System muss nach $\hat{\mathbf{y}}_k = (\hat{y}_k^1, \hat{y}_k^2)^T$ aufgelöst werden. Zumeist wird dies mit Fixpunktiteration gelöst, wobei zu berücksichtigen ist, dass die Funktion $x \mapsto \mathrm{sign}(x)$ nicht Lipschitzstetig ist.