Gewöhnliche Differentialgleichungen und dynamische Systeme

Vorlesung im Sommersemester 2016

Thomas Schmidt

Version: 6. Juli 2016

Inhaltsverzeichnis

Falls Sie in diesem Skript Fehler (jeglicher Art) finden oder sonstige Hinweise haben, bitte ich Sie, mir dies entweder persönlich oder unter thomas.schmidt@math.uni-hamburg.de mitzuteilen.

Teil I

Dynamische Systeme: Allgemeine Begriffe und Aspekte der diskreten Theorie

Kapitel 1

Grundlegende Definitionen und Beispiele

Prinzipiell lässt sich der Begriff **dynamisches System** sehr allgemein für ein **beliebiges System** verwenden, **dessen Zustand sich im Laufe der Zeit verändert** (oder zumindest potentiell verändern kann); dies kann ein physikalisches, technisches, biologisches, noch ein anderes praktisch auftretendes oder auch ein rein mathematisch-theoretisches System sein. Tatsächlich interessiert man sich in der Theorie dynamischer Systeme aber vor allem für das Verhalten eines Systems im Hinblick auf einen sehr großen Zeithorizont, **für das sogenannte Langzeitverhalten**, und spricht nur in diesem Zusammenhang von einem dynamischen System (oder der Dynamik eines Systems).

Als allgemeines mathematisches Modell für ein dynamisches System betrachtet man eine Abbildung $\Phi\colon \mathbb{T} \times \mathcal{X} \to \mathcal{X}$. Dabei bezeichnet $\mathcal{X}$ eine Menge von möglichen Zuständen des Systems, $\mathbb{T}$ eine Menge von Zeitpunkten, zu denen jeweils ein solcher Zustand eintritt, und der Funktionswert $\Phi(t,x) \in \mathcal{X}$ steht für den Zustand, der sich aus einem Initialzustand $x \in \mathcal{X}$ von der Initialzeit $0 \in \mathbb{T}$ bis zu einem Zeitpunkt $t \in \mathbb{T}$ entwickelt. Als präzise Definition wird hier, mit den für die ganze Vorlesung gültigen Konventionen

$$\mathbb{N} = \{1,2,3,\ldots\}\,, \qquad \mathbb{N}_0 = \mathbb{N} \cup \{0\}\,, \qquad \mathbb{R}^+ = (0,\infty)\,, \qquad \mathbb{R}_0^+ = [0,\infty)\,,$$

vereinbart:

Definition 1.1 (**Flüsse, dynamische Systeme**)**.** *Seien $\mathcal{X}$ ein metrischer Raum und $\mathbb{T}$ eine der vier Mengen $\mathbb{N}_0$, $\mathbb{Z}$, $\mathbb{R}_0^+$, $\mathbb{R}$. Als* ***(globalen) Fluss*** *auf dem Zustandsraum $\mathcal{X}$ (mit Zeitmenge $\mathbb{T}$) bezeichnet man eine stetige*[1] *Abbildung*

$$\Phi\colon \mathbb{T} \times \mathcal{X} \to \mathcal{X}\,,$$

die die Identitätseigenschaft

$$\Phi(0,x) = x \qquad \textit{für alle } x \in \mathcal{X} \tag{1.1}$$

und die Halbgruppeneigenschaft

$$\Phi(s{+}t,x) = \Phi(s,\Phi(t,x)) \qquad \textit{für alle } s,t \in \mathbb{T} \textit{ und } x \in \mathcal{X} \tag{1.2}$$

[1] Um von Stetigkeit von Φ überhaupt sprechen zu können, wird der Definitionsbereich $\mathbb{T} \times \mathcal{X}$ von Φ mit der Produkt-Metrik $\mathrm{d}_{\mathbb{T}\times\mathcal{X}}((t,x),(\widetilde{t},\widetilde{x})) := |\widetilde{t}{-}t| + \mathrm{d}_{\mathcal{X}}(x,\widetilde{x})$ versehen und damit als metrischer Raum betrachtet.

erfüllt. Die partiellen Abbildungen

$$\Phi_t := \Phi(t, \cdot\,) \in \mathrm{C}^0(\mathcal{X}, \mathcal{X})$$

mit $t \in \mathbb{T}$ heißen dann die ***Flussabbildungen*** *von Φ, die Funktionen $\Phi(\cdot\,, x)\colon \mathbb{T} \to \mathcal{X}$ mit $x \in \mathcal{X}$ nennt man die* ***Fluss- oder Bahnlinien*** *von Φ, und das Tripel $(\mathbb{T}, \mathcal{X}, \Phi)$ bezeichnet man als* ***(stetiges) dynamisches System*** *(auf $\mathcal{X}$ mit Zeitmenge $\mathbb{T}$).*

Bemerkungen und Erläuterungen (zur **Definition dynamischer Systeme**)**.**

(1) Notiert man den metrischen Raum als Paar $(\mathcal{X}, \mathrm{d}_{\mathcal{X}})$ mit Grundmenge $\mathcal{X}$ und Metrik $\mathrm{d}_{\mathcal{X}}$, so kann man ein dynamisches System auch als Quadrupel $(\mathbb{T}, \mathcal{X}, \mathrm{d}_{\mathcal{X}}, \Phi)$ betrachten. Stellt man sich auf den Standpunkt, dass mit der Abbildung Φ auch ihr Definitionsbereich gegeben ist, so kann man andererseits auf die explizite Angabe von $\mathbb{T}$ und $\mathcal{X}$ verzichten.

(2) Man unterscheidet bei dynamischen Systemen nach Beschaffenheit der Zeitmenge $\mathbb{T}$ zwischen **diskreter** ($\mathbb{T} \in \{\mathbb{N}_0, \mathbb{Z}\}$) und **kontinuierlicher**[2] ($\mathbb{T} \in \{\mathbb{R}_0^+, \mathbb{R}\}$) sowie zwischen **irreversibler** ($\mathbb{T} \in \{\mathbb{N}_0, \mathbb{R}_0^+\}$) und **reversibler** ($\mathbb{T} \in \{\mathbb{Z}, \mathbb{R}\}$) zeitlicher Entwicklung. Bei diskreter Zeit erfolgt die Entwicklung des Systems Schritt für Schritt (mit abzählbar vielen Schritten), bei kontinuierlicher Zeit läuft sie ununterbrochen. Bei irreversibler Zeit beginnt die Entwicklung des Systems zum festen Initialzeitpunkt 0, bei reversibler Zeit dauert sie (zumindest idealisiert) schon seit unendlich langer Zeit an.

Man spricht bei Vorliegen des diskreten Falls auch von **dynamischen Systemen in diskreter Zeit**, von **(Zeit-)diskreten dynamischen Systemen** oder kurz von **diskreter Dynamik**, und ähnlich verwendet man die anderen Adjektive. Bei Auslassung des Wortes ‚Zeit' gilt es aber darauf zu achten, dass die Eigenschaft von $\mathbb{T}$ nicht mit einer Eigenschaft von $\mathcal{X}$ verwechselt werden kann.

(3) Die Forderung der **Identitätseigenschaft** (1.1) ist in Anbetracht der eingangs gegebenen Motivation (und der Konvention, stets 0 als Initialzeit zu betrachten) plausibel: Sie besagt tatsächlich nur, dass sich aus dem Initialzustand x zur Zeit 0 bis zur immer noch gleichen Zeit 0 wieder der Zustand $\Phi(0, x) = x$ entwickelt; ohne Vergehen von Zeit erfolgt also naheliegenderweise keine Zustandsänderung.

(4) Die **Halbgruppeneigenschaft** (1.2) beinhaltet tatsächlich eine bisher noch nicht erwähnte **Einschränkung an das modellierte System**, nämlich die, dass **gleichbleibende äußere Umstände** vorliegen. Um sich dies zu veranschaulichen, denke man an eine zeitliche Entwicklung mit Initialzustand x, die sich im Laufe der Zeit von t bis $s{+}t$ von $\Phi(t, x)$ zu $\Phi(s{+}t, x)$ entwickelt. Die Gleichung (1.2) besagt dann, dass bei Vorliegen des Zustands $\Phi(t, x)$ zum Initialzeitpunkt 0 die Entwicklung zum Zustand $\Phi(s, \Phi(t, x))$ im Laufe der Zeit von 0 bis s im selben Zustand resultiert, also dass $\Phi(s, \Phi(t, x)) = \Phi(s{+}t, x)$ gilt. Es kommt bei der Evolution des Systems somit nur auf die verstrichene Zeit s an, nicht auf die Anfangs- und Endzeiten selbst, wie man es eben bei zeitlich konstanten äußeren Umständen erwartet.

Die durch (1.2) auferlegte Einschränkung wird zumindest für den ersten Teil dieser Vorlesung aufrecht erhalten, ist aber im dort relevanten Kontext auch in weiten Teilen der Literatur durchaus üblich.

[2] Gelegentlich wird auch von *stetiger* statt *kontinuierlicher* Zeit gesprochen. Im Deutschen ist es aber ratsam, diesen Sprachgebrauch zu vermeiden, um Verwechslungen mit der Bezeichnung als *stetiges* dynamisches System (bei der das Adjektiv auf die Stetigkeit von Φ verweist) vorzubeugen. Im Englischen lässt sich das Problem nicht so einfach lösen und erfordert unter Umständen eine zusätzliche Klarstellung.

(5) Die **wesentliche Gemeinsamkeit der zulässigen Zeitmengen** $\mathbb{N}_0$, $\mathbb{Z}$, $\mathbb{R}_0^+$, $\mathbb{R}$ besteht (neben der Tatsache, dass es sich um abgeschlossene Teilmengen von $\mathbb{R}$ handelt) darin, dass sie bezüglich der Addition **Halbgruppen mit neutralem Element**, auch Monoide genannt, bilden. In diesem Kontext wird die Benennung von (1.2) als Halbgruppeneigenschaft verständlich, denn diese Gleichung besagt gerade, dass durch Φ eine **Operation**[3] **der Halbgruppe** $\mathbb{T}$ **auf** $\mathcal{X}$ gegeben ist. Man kann (1.2) auch noch etwas anders auffassen und in der einprägsamen Form

$$\Phi_{s+t} = \Phi_s \circ \Phi_t \qquad \text{für } s,t \in \mathbb{T} \tag{1.3}$$

schreiben; diese Gleichung besagt dann, dass die Zuordnung $t \mapsto \Phi_t$ ein **Halbgruppenhomomorphismus** von $\mathbb{T}$ auf $\mathrm{C}^0(\mathcal{X},\mathcal{X})$ (mit der Komposition als Verknüpfung) ist. Die Identitätseigenschaft (1.1) ist in diesem Kontext ebenfalls natürlich und bedeutet, dass $t \mapsto \Phi_t$ das neutrale Element 0 von $\mathbb{T}$ auf das neutrale Element $\Phi_0 = \mathrm{id}_{\mathcal{X}}$ von $\mathrm{C}^0(\mathcal{X},\mathcal{X})$ abbildet.

(6) Bei den beiden zulässigen Zeitmengen $\mathbb{Z}$ und $\mathbb{R}$ handelt es sich nicht nur um Halbgruppen, sondern sogar um **Gruppen** (bezüglich der Addition), in denen alle Elemente invertierbar sind. Unter dem Halbgruppenhomomorphismus $t \mapsto \Phi_t$ können diese Elemente nur auf invertierbare Elemente in $\mathrm{C}^0(\mathcal{X},\mathcal{X})$ abgebildet werden, also auf Homöomorphismen[4] von $\mathcal{X}$ auf sich. Insgesamt entspricht Φ somit einer Gruppenoperation, und $t \mapsto \Phi_t$ wird zu einem Gruppenhomomorphismus von $\mathbb{T}$ auf die Gruppe der Homöomorphismen von $\mathcal{X}$ auf sich. **Insbesondere sind im Gruppenfall** $\mathbb{T} \in \{\mathbb{Z},\mathbb{R}\}$ **alle Flussabbildungen** Φ_t **Homöomorphismen**, und durch Einsetzen von $s = -t$ in (1.3) erhält man für ihre Inversen die Formel

$$(\Phi_t)^{-1} = \Phi_{-t} \qquad \text{für } t \in \mathbb{T}\,.$$

An dieser Stelle wird verständlich, warum solche Systeme als reversibel bezeichnet werden: Anders als im Fall $\mathbb{T} \in \{\mathbb{N}_0, \mathbb{R}_0^+\}$ kann die Anwendung von Φ_t im Fall $\mathbb{T} \in \{\mathbb{Z},\mathbb{R}\}$ nämlich immer rückgängig gemacht werden — einfach durch die Anwendung von Φ_{-t}.

(7) Die Bahnlinien $\Phi(\,\cdot\,,x)$ beschreiben die zeitliche Evolution der Initialzustände $x \in \mathcal{X}$. Daher stehen bei der Untersuchung eines dynamischen Systems **die Bahnlinien und besonders ihr Langzeitverhalten** (das heißt das Verhalten von $\Phi(t,x)$ bei $t \to \infty$ und für $\mathbb{T} \in \{\mathbb{Z},R\}$ auch bei $t \to -\infty$) oft **im Zentrum des Interesses.**

Bemerkungen (zu **Varianten und Verallgemeinerungen der Definition**)**.** Weitgehende Einigkeit besteht bei der Definition von dynamischen Systemen in der Literatur — jedenfalls sofern man sich nicht von vornherein auf diskrete Systeme beschränkt — einzig darin, wie oben mit Abbildungen $\Phi\colon \mathbb{T}\times\mathcal{X}\to\mathcal{X}$ sowie mit (1.1) und (1.2) zu arbeiten. Darüber hinaus finden sich bei vielen Autoren Abweichungen von Definition 1.1, die für die Vorlesung zwar nicht weiter von Bedeutung sind, aber kurz erwähnt werden sollen:

- Statt metrischen Räumen $\mathcal{X}$ werden oft auch (Klassen von) **topologischen Räume**(n) als Zustandsräume $\mathcal{X}$ zugelassen.

[3] Allgemein versteht man unter einer (Links-)Operation einer Halbgruppe $(T,\,\cdot\,)$ auf einer Menge $\mathcal{X}$ eine Verknüpfung $\otimes\colon T\times\mathcal{X}\to\mathcal{X}$ mit $(s\cdot t)\otimes x = s\otimes(t\otimes x)$ für alle $s,t\in T$, $x\in\mathcal{X}$. Schreibt man die Halbgruppenverknüpfung additiv und ersetzt $\otimes$ durch die Anwendung von Φ, so erkennt man die vorausgehende Formel als Umformulierung von (1.2).

[4] Als Homöomorphismus von $\mathcal{X}$ auf sich bezeichnet man eine bijektive, stetige Abbildung $\mathcal{X}\to\mathcal{X}$ mit stetiger Umkehrabbildung.

- Noch allgemeiner wird manchmal nur gefordert, dass **$\mathcal{X}$ eine Menge ohne jede Zusatzstruktur** ist und dementsprechend wird die (dann nicht mehr sinnvolle) **Stetigkeitsforderung an Φ fallengelassen**. Somit reduziert sich der Kern der Definition auf ein **rein algebraisches Konzept**, nämlich das einer Halbgruppenoperation (bei der das neutrale Element als Identität operiert).

- Statt den oben zugelassenen Zeitmengen $\mathbb{N}_0$, $\mathbb{Z}$, $\mathbb{R}_0^+$, $\mathbb{R}$ kann man auch **allgemeine(re) Halbgruppen** $\mathbb{T}$ mit neutralem Element erlauben (die anders als die obigen vier nicht unbedingt kommutativ sein müssen). Zur Aufrechterhaltung der Stetigkeitsforderung an Φ ist dabei zu fordern, dass die Halbgruppe $\mathbb{T}$ eine mit ihrer Verknüpfung kompatible[5] Topologie oder Metrik trägt und damit eine sogenannte **topologische Halbgruppe** ist. Verzichtet man auf Stetigkeit von Φ, so kann man die Definition aber auch für rein algebraische Halbgruppen $\mathbb{T}$ treffen.

- Statt der in Definition 1.1 verlangten Stetigkeit von Φ wird manchmal nur eine **schwächere Stetigkeitsbedingung** gefordert, die unter Umständen leichter nachzuweisen ist. Es handelt sich dabei um die Bedingung, dass für alle $t \in \mathbb{T}$ die Flussabbildungen Φ_t und gleichzeitig für alle $x \in \mathcal{X}$ die Bahnlinien $\Phi(\,\cdot\,, x)$ stetig sind. Dies macht aber nur bei kontinuierlicher Zeit überhaupt einen Unterschied.

  Andererseits bildet, gerade bei kontinuierlicher Zeit, das Studium glatter dynamischer Systeme eine der Hauptrichtungen der Theorie, und in diesem Zusammenhang sind auch die deutlichen stärkeren Forderungen, dass **Φ eine C^1- oder sogar C^∞-Abbildung** ist, durchaus üblich.

- In der sogenannten **Ergodentheorie** beschäftigt man sich mit maßtheoretischen Fragestellungen bei dynamischen Systemen und arbeitet typischerweise **mit Maßräumen $\mathcal{X}$** oder auch mit metrischen Maßräumen statt einfachen metrischen Räumen. Die Stetigkeitsforderung an Φ wird dabei unter Umständen durch eine geeignete Messbarkeitsforderung ersetzt.

Wichtige Typen dynamischer Systeme. Es werden nun (die vielleicht wichtigsten) drei Typen dynamischer Systeme genauer diskutiert. Der erste Typ wird im unmittelbar Folgenden wiederholt auftreten. Der zweite Typ wird im zweiten Teil der Vorlesung ausführlich behandelt, der dritte geht über den Vorlesungsstoff hinaus.

(A) **Diskrete Dynamiken** sind typischerweise durch **Iterationen von Selbstabbildungen** gegeben:

Ist $\varphi \in C^0(\mathcal{X}, \mathcal{X})$ eine stetige Abbildung eines metrischen Raums $\mathcal{X}$ auf sich, so erhält man durch

$$\Phi(k, x) := \varphi^k(x) = \underbrace{(\varphi \circ \varphi \circ \ldots \circ \varphi)}_{k \text{ mal}}(x) \qquad \text{für } k \in \mathbb{N}_0\,,\ x \in \mathcal{X}$$

(mit der üblichen Konvention $\varphi^0 = \mathrm{id}_{\mathcal{X}}$) ein Zeit-diskretes dynamisches System $(\mathbb{N}_0, \mathcal{X}, \Phi)$ mit Zeitmenge $\mathbb{N}_0$. Um dies einzusehen, muss man nur die Gültigkeit der Bedingungen (1.1) und (1.2) prüfen, wobei erstere per Konvention klar ist und letztere sich aus der kurzen Rechnung $\Phi(k+\ell, x) = \varphi^{k+\ell}(x) = \varphi^k(\varphi^\ell(x)) = \Phi(k, \Phi(\ell, x))$ ergibt. Die zu diesem

[5] Kompatibilität bedeutet hier, dass die Halbgruppenverknüpfung $\mathbb{T} \times \mathbb{T} \to \mathbb{T}$ und, soweit definiert, auch die Inversenbildung $\mathbb{T} \to \mathbb{T}$ stetig sind.

System gehörigen Flussabbildungen sind die Iterationen $\Phi_k = \varphi^k$ von φ, und insbesondere ist $\Phi_1 = \varphi$.

Ist φ sogar Homöomorphismus von $\mathcal{X}$ auf sich, so kann man φ^k für alle $k \in \mathbb{Z}$ erklären, indem man wie üblich φ^{-1} für die zu φ inverse Abbildung schreibt und $\varphi^{-n} := (\varphi^{-1})^n = (\varphi^n)^{-1}$ vereinbart. Durch $\Phi(k,x) := \varphi^k(x)$ für $k \in \mathbb{Z}$, $x \in \mathcal{X}$ erhält man dann sogar ein Zeit-diskretes dynamisches System $(\mathbb{Z}, \mathcal{X}, \Phi)$ mit Zeitmenge $\mathbb{Z}$ und Flussabbildungen $\Phi_k = \varphi^k$.

Tatsächlich sind **alle Zeit-diskreten Systeme von der** gerade **betrachteten Form**, das heißt genauer: Bei jedem stetigen dynamischen System $(\mathbb{N}_0, \mathcal{X}, \Phi)$ ist $\varphi := \Phi_1 \in \mathrm{C}^0(\mathcal{X}, \mathcal{X})$, und es gilt $\Phi(k,x) = \varphi^k(x)$ für alle $k \in \mathbb{N}_0$, $x \in \mathcal{X}$. Und bei jedem stetigen dynamischen System $(\mathbb{Z}, \mathcal{X}, \Phi)$ ist $\varphi := \Phi_1$ Homöomorphismus von $\mathcal{X}$ auf sich mit $\Phi(k,x) = \varphi^k(x)$ für alle $k \in \mathbb{Z}$, $x \in \mathcal{X}$.

Im Lichte dieses Zusammenhangs **notiert** man bei Systemen **in diskreter Zeit** auch $(\mathbb{T}, \mathcal{X}, \varphi)$ statt $(\mathbb{T}, \mathcal{X}, \Phi)$ **mit der Selbstabbildung φ anstelle des Flusses Φ**.

(B) Die wichtigsten **kontinuierlichen Dynamiken** erhält man aus **(Systemen von) gewöhnlichen Differentialgleichungen**:

Gemäß Sätzen des zweiten Vorlesungsteils gibt es nämlich für eine geeignete Teilmenge $\mathcal{X}$ eines vollständigen normierten Raums und unter gewissen Voraussetzungen an ein Vektorfeld $F \in \mathrm{C}^0(\mathcal{X}, \mathcal{X})$ eine eindeutige Lösung $u \in \mathrm{C}^0(\mathbb{R}, \mathcal{X})$ der gewöhnlichen Differentialgleichung

$$u'(t) = F(u(t)) \qquad \text{für alle } t \in \mathbb{R}$$

mit Anfangsbedingung $u(0) = x$, $x \in \mathcal{X}$ sowie eine eindeutige Lösung $\Phi \in \mathrm{C}^0(\mathbb{R} \times \mathcal{X}, \mathcal{X})$ der zugehörigen Flussgleichungen

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t,x) = F(\Phi(t,x)) \qquad \text{für alle } t \in \mathbb{R}\,,\ x \in \mathcal{X}$$

mit Anfangsbedingung $\Phi(0,x) = x$. Unter den entsprechenden Voraussetzungen ist dann durch $(\mathbb{R}, \mathcal{X}, \Phi)$ ein kontinuierliches dynamisches System mit Zeitmenge $\mathbb{R}$ gegeben.

Außerdem wird sich herausstellen, dass kontinuierliche dynamische Systeme mit C^1-Flüssen (auf einem normierten Raum) stets von dieser Form sind.

Für eine genauere Diskussion sei auf den späteren Abschnitt 6.6 verwiesen.

(C) Des Weiteren sind **zufällige Dynamiken** von Interesse, die beispielsweise **bei Markov-Ketten** in diskreter Zeit **und bei Markov-Prozessen** in kontinuierlicher Zeit auftreten. Eine ernsthafte Untersuchung solcher Dynamiken geht über den Rahmen der Vorlesung hinaus, daher soll hier nur vage angedeutet werden, wie homogene Markov-Ketten mit Werten in $\{1, 2, \ldots, N\}$, $N \in \mathbb{N}$ als diskrete dynamische Systeme aufgefasst werden können.

Hierzu sei daran erinnert, dass es sich bei einer solchen Markov-Kette um eine Folge von messbaren Abbildungen, genannt Zufallsvariablen, $(X_k)_{k\in\mathbb{N}_0}$ von einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathrm{P})$ nach $\{1, 2, \ldots, N\}$ handelt, bei der „X_{k+1} nur von X_k abhängt“ (d.h. präziser, dass für alle $k \in \mathbb{N}$ die bedingten Verteilungen von X_{k+1} bei Kenntnis von $(X_0, X_1, \ldots, X_{k-1}, X_k)$ mit den bedingten Verteilungen von X_{k+1} bei Kenntnis von einzig X_k übereinstimmen) und die Einträge[6] $M_{ij} := \mathrm{P}(\{X_k{=}j\}|\{X_{k-1}{=}i\})$ der Über-

[6] Per Definition ist $\{X_k{=}j\} = \{\omega \in \Omega : X_k(\omega){=}j\}$ und $\mathrm{P}(A|B) = \frac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)}$, wobei bei Verwendung der Notation $\mathrm{P}(A|B)$ automatisch $\mathrm{P}(B) > 0$ vorausgesetzt sei.

gangsmatrix $M \in [0,1]^{N\times N}$ nicht von k abhängen. Die Verteilungen aller X_k (und damit alle relevanten Informationen über die Kette) sind vollständig bestimmt durch die Verteilung von X_0 und die Matrix M, für deren Einträge übrigens $\sum_{j=1}^{N} M_{ij} = 1$ gilt. Bezeichnet nun $\mathcal{W} := \{w \in [0,1]^N : \sum_{i=1}^{N} w_i = 1\}$ den Raum der Wahrscheinlichkeits-(Zeilen-)Vektoren mit N Einträgen und beschreibt man die Verteilung von X_0 durch den Vektor $v := (\mathrm{P}(\{X_0{=}i\}))_{i=1,2,\dots,N} \in \mathcal{W}$, so ist die Verteilung von X_1 beschrieben durch den Vektor $vM \in \mathcal{W}$ und die von X_k durch $vM^k \in \mathcal{W}$. Ausgehend von dieser Gesetzmäßigkeit können obige Markov-Ketten nun als dynamische Systeme vom Typ (A) mit $\mathcal{W}$ anstelle von $\mathcal{X}$ betrachtet werden; dazu geht man bei gegebener Matrix M mit den vorausgehenden Eigenschaften von der Abbildung $\varphi(v) = vM$ aus und erhält durch

$$\Phi(k,v) := vM^k \qquad \text{für } k \in \mathbb{N}_0\,,\ v \in \mathcal{W}$$

ein diskretes dynamisches System $(\mathbb{N}_0, \mathcal{W}, \Phi)$ mit Zeitmenge $\mathbb{N}_0$. Auf diese Art und Weise entsprechen alle Markov-Ketten mit gleicher Übergangsmatrix M (aber unterschiedlichem v und damit unterschiedlich verteilten X_0 und X_k) einem dynamischen System.

Als Nächstes werden konkrete Beispiele Zeit-diskreter dynamischer Systeme des Typs (A) angegeben und verschiedene Möglichkeiten zu deren Veranschaulichung aufgezeigt.

Beispiele (für **Zeit-diskrete Dynamiken**)**.**

(1) Für jeden Parameter $\beta \in \mathbb{R}_0^+$ ist eine Selbstabbildung $\varphi\colon \mathbb{R}_0^+ \to \mathbb{R}_0^+$ des Zustandsraums $\mathbb{R}_0^+$ durch

$$\varphi(x) = \beta x \qquad \text{für } x \in \mathbb{R}_0^+$$

gegeben. Für $\beta \in \mathbb{R}^+$ handelt es sich sogar um einen Homöomorphismus, und das zugehörige System vom obigen Typ (A) ist $(\mathbb{Z}, \mathbb{R}_0^+, \Phi)$ mit $\Phi(k,x) = \beta^k x$. Es handelt sich also um ein System mit **exponentiellem Wachstum** der Zuwachsrate $\beta > 1$ (für diesen Fall veranschaulicht in Abbildung 1) oder mit **exponentiellen Abfall** der Abfallrate $\beta < 1$. Dieses System kann als sehr einfaches Modell für das Wachstum von Populationen (eventuell bis auf Rundungsproblematik), für radioaktiven Zerfall und vieles andere eingesetzt werden.

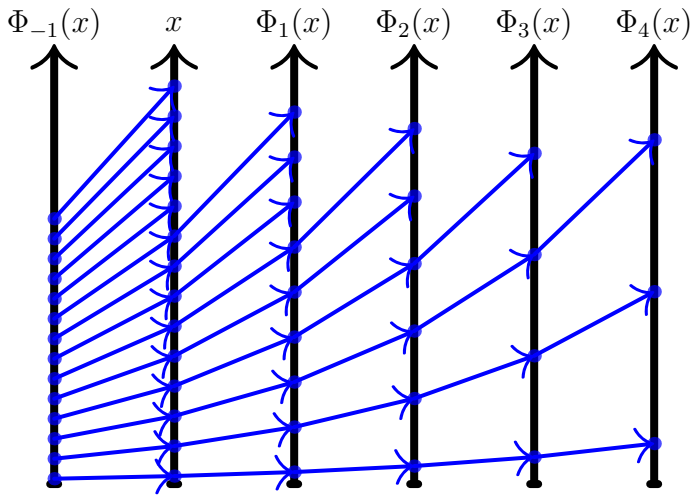


Abb. 1: Eine Veranschaulichung des diskreten exponentiellen Wachstums aus Beispiel (1) mit $\beta > 1$.

(2) Die **Collatz-Abbildung** $\varphi\colon \mathbb{N} \to \mathbb{N}$ des Zustandsraums $\mathbb{N}$ in sich ist gegeben durch

$$\varphi(n) := \begin{cases} 3n{+}1 & \text{für ungerades } n \\ \frac{1}{2}n & \text{für gerades } n \end{cases} \qquad \text{für } n \in \mathbb{N}\,.$$

Hier lässt sich für den Fluss Φ des zugehörigen Systems $(\mathbb{N}_0, \mathbb{N}, \Phi)$ vom Typ (A) keine so explizite Formel angeben. Für spezielle Initialwerte ergibt sich aber eine einfache Flusslinie,

zum Beispiel erhält man für den Initialwert 1 die 3-periodische Folge

$$\Phi(\,\cdot\,,1) = (1,4,2,1,4,2,1,4,2,1,4,2,\ldots)\,,$$

und sehr viele andere Initialwerte münden schließlich in dieselbe Folge wie beispielsweise

$$\Phi(\,\cdot\,,17) = (17,52,26,13,40,20,10,5,16,8,4,2,1,4,2,\ldots)\,.$$

Eine naheliegende und für dynamische Systeme typische Fragestellung ist nun die, ob für *alle* natürlichen Zahlen als Initialwerte die Folge schließlich in die Wiederholung von $(1,4,2)$ übergeht. Es wird vermutet, dass dies so ist, aber ein Beweis steht noch aus, und es handelt sich tatsächlich um ein berühmtes offenes Problem, bekannt als **Collatz-Problem** oder **(3x+1)-Problem**. Man kann die Collatz-Abbildung und dieses Problem teilweise durch den in Abbildung 2 gezeigten gerichteten Graphen veranschaulichen. Das offene Problem besteht dann in der Frage, ob der entsprechende abzählbar unendliche Graph, in dem alle natürlichen Zahlen auftreten, zusammenhängt (und somit nur den einzigen Zyklus $(1,4,2)$ aufweist).

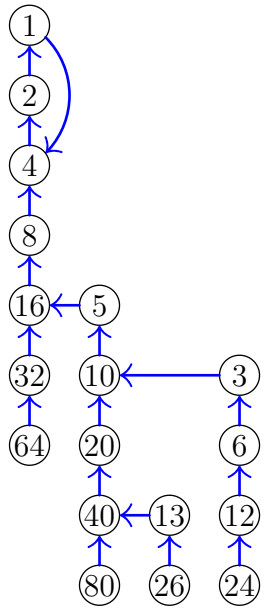


Abb. 2: Veranschaulichung von Beispiel (2) durch einen gerichteten Graphen.

(3) Für jeden Parameter $\xi \in \mathrm{S}^1$ ist ein Homöomorphismus $\varphi\colon \mathrm{S}^1 \to \mathrm{S}^1$ der Einheitskreislinie $\mathrm{S}^1 := \{z \in \mathbb{C} : |z| = 1\}$ auf sich gegeben durch die komplexe Multiplikation

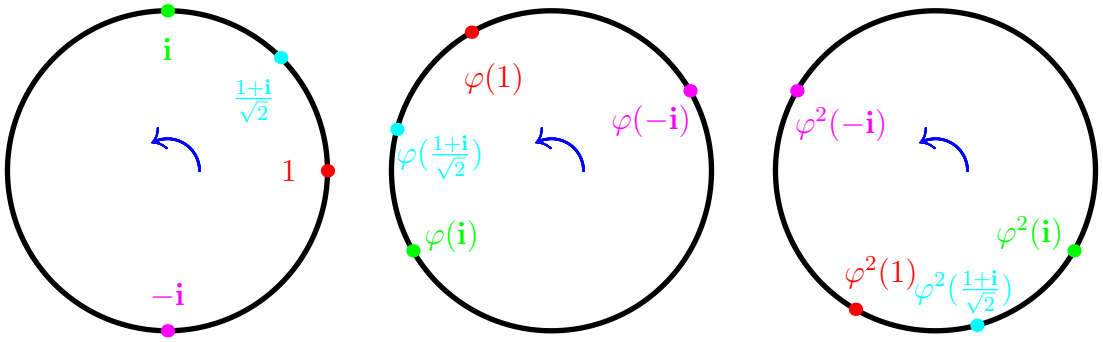
$$\varphi(z) = z{\cdot}\xi \qquad \text{für } z \in \mathrm{S}^1\,.$$

Das zugehörige System vom Typ (A) ist $(\mathbb{Z}, \mathrm{S}^1, \Phi)$ mit $\Phi(k,z) = z{\cdot}\xi^k$.

Ist nun $\xi \in \mathrm{S}^1$ eine ℓ-te Einheitswurzel (mit $\ell \in \mathbb{N}$), so erhält man wegen $\xi^\ell = 1$ für jeden beliebigen Initialzustand $z \in \mathrm{S}^1$ eine ℓ-periodische Folge, nämlich

$$\Phi(\,\cdot\,,z) = (z, z{\cdot}\xi, z{\cdot}\xi^2, z{\cdot}\xi^3, \ldots, z{\cdot}\xi^{\ell-1}, z, z{\cdot}\xi, \ldots)\,,$$

und anschaulich entspricht jeder Zeitschritt des Systems einer Drehung der Einheitskreislinie um den Winkel $\frac{2\pi\mathbf{i}}{\ell}$ — wie in Abbildung 3 für $\ell = 3$ illustriert. Damit ist das Verhalten für



$k = 0, 3, 6, \ldots$ $\qquad$ $k = 1, 4, 7, \ldots$ $\qquad$ $k = 2, 5, 8, \ldots$

Abb. 3: Veranschaulichung von Beispiel (3) für den Fall der dritten Einheitswurzel $\xi = -\frac{1}{2} + \frac{\sqrt{3}}{2}\mathbf{i}$.

alle Parameter der Form $\xi = \exp(2\pi \mathbf{i} q)$, $q \in \mathbb{Q}$ vollständig beschrieben. Für $\xi = \exp(2\pi \mathbf{i} r)$, $r \in \mathbb{R} \setminus \mathbb{Q}$ jedoch tritt in diesem Beispiel ein anderes nicht-periodisches Verhalten auf; dazu später etwas mehr.

Kontinuierliche dynamische Systeme werden im Laufe der Vorlesung noch ausführlich untersucht, weswegen an dieser Stelle nur ein Beispiel (und so ziemlich das allereinfachste) erwähnt wird.

Beispiel (für eine **kontinuierliche Dynamik**).

(4) In kontinuierlicher Zeit werden ein **exponentielles Wachstum** beziehungsweise ein **exponentieller Abfall** mit Zuwachs-/Abfallrate e^γ, $\gamma \in \mathbb{R}$ durch die (sehr einfache) gewöhnliche Differentialgleichung

$$u'(t) = \gamma u(t) \qquad \text{für alle } t \in \mathbb{R}$$

beschrieben. Man kann $u(t) = \mathrm{e}^{\gamma t} x$ als eindeutige Lösung mit Anfangsbedingung $u(0) = x$, $x \in \mathbb{R}_0^+$ erraten und erhält das dynamische System $(\mathbb{R}, \mathbb{R}_0^+, \Phi)$ mit Fluss $\Phi(t, x) = \mathrm{e}^{\gamma t} x$. Dieses System ist für $\gamma > 0$ in Abbildung 4 illustriert und sollte, wie auch die Ähnlichkeit der Abbildungen nahelegt, als kontinuierliche Version von Beispiel (1) mit $\beta = \mathrm{e}^\gamma$ betrachtet werden. Das kontinuierliche System kann ebenfalls als Modell für diverse Wachstums- oder Zerfallsprozesse eingesetzt werden.

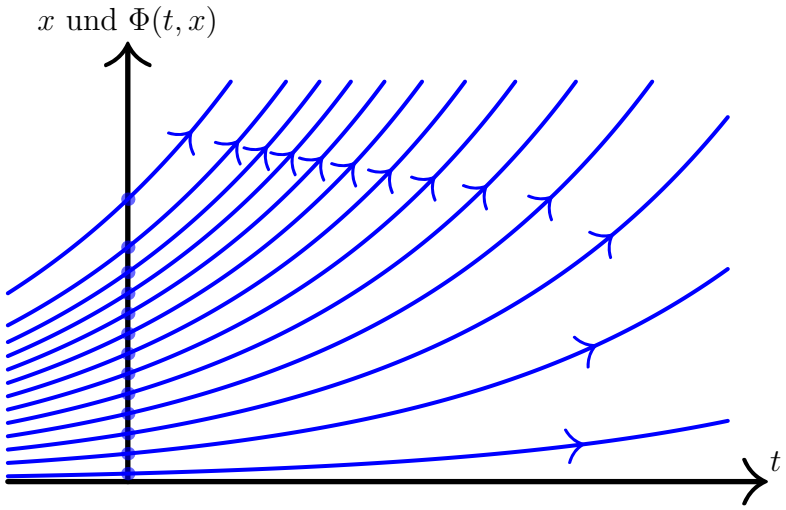


Abb. 4: Graphen einiger Bahnlinien des kontinuierlichen exponentiellen Wachstums aus Beispiel (4) mit $\gamma > 0$.

Bemerkungen (**Zusammenhänge zwischen diskreter und kontinuierlicher Dynamik**).

(1) Der Zusammenhang zwischen den Beispielen (1) und (4) ist übrigens kein Einzelfall. Auch allgemein kann man nämlich ein **kontinuierliches dynamisches System immer in** ein **Zeit-diskretes System** verwandeln, einfach indem man von $\mathbb{T} = \mathbb{R}_0^+$ auf $\mathbb{T} = \mathbb{N}_0$ beziehungsweise von $\mathbb{T} = \mathbb{R}$ auf $\mathbb{T} = \mathbb{Z}$ einschränkt. Ist Φ der Fluss des (alten) kontinuierlichen Systems, so besteht das (neue) diskrete System vom Typ (A) dann aus den **Iterationen der Zeit-1-Abbildung Φ_1**.

(2) **Umgekehrt kann man aber nicht immer von einer diskreten Zeitmenge auf eine kontinuierliche fortsetzen**, das heißt mit anderen Worten, nicht jedes $\varphi \in \mathrm{C}^0(\mathcal{X}, \mathcal{X})$ ist die Zeit-1-Abbildung Φ_1 eines Systems $(\mathbb{R}_0^+, \mathcal{X}, \Phi)$ (und nicht jeder Homöomorphismus φ von $\mathcal{X}$ auf sich die eines Systems $(\mathbb{R}, \mathcal{X}, \Phi)$). Dies liegt zum einen daran, dass wegen möglicher **topologischer Hindernisse** nicht einmal eine stetige Abbildung $\Phi\colon [0,1] \times \mathcal{X} \to \mathcal{X}$ mit $\Phi(0, \cdot) = \mathrm{id}_\mathcal{X}$, $\Phi(1, \cdot) = \varphi$ existieren muss (man sagt dann, dass φ nicht stetig in $\mathrm{id}_\mathcal{X}$ deformiert werden kann, oder auch, dass $\mathrm{id}_\mathcal{X}$ und φ nicht homotop sind). Ein extremer Fall ist der eines diskreten Zustandsraums wie beispielsweise $\mathbb{N}$; dann ist nämlich die einzig mögliche kontinuierliche Dynamik $\Phi(t, x) = x$ konstant (und damit trivial), während in Beispiel (2)

schon eine nicht-triviale diskrete Dynamik vorgestellt wurde. Ein anderer einfacher Fall ist der der Selbstabbildungen $\varphi(z) := z^k$ der Einheitskreislinie $\mathrm{S}^1 \subset \mathbb{C}$ mit Umlaufzahlen $k \in \mathbb{Z} \setminus \{1\}$ (für $k = -1$ handelt es sich sogar um einen Homöomorphismus); diese Abbildungen sind wegen der unterschiedlichen Umlaufzahlen nicht zu $\mathrm{id}_{\mathrm{S}^1}$ homotop und daher keine Zeit-1-Abbildungen eines kontinuierlichen Systems.

(3) **Selbst wenn für einen Homöomorphismus** φ von $\mathcal{X}$ auf sich **keine direkten topologischen Hindernisse vorliegen, ist er** aber immer noch **nicht unbedingt als Zeit-1-Abbildung einer kontinuierlichen Dynamik realisierbar.** Als Beispiel hierfür wird typischerweise der Fall herangezogen, dass es bis auf Vertauschung genau zwei *verschiedene* Punkte $x_0, x_1 \in \mathcal{X}$ gibt, die durch $\varphi(x_0) = x_1$, $\varphi(x_1) = x_0$ abgebildet werden (solche Abbildungen gibt es; eine explizite auf $\mathcal{X} = [0,1]$ ist $\varphi(x) = \frac{1}{2} - 4\big(x - \frac{1}{2}\big)^3$ mit $\{x_0, x_1\} = \{0,1\}$). Wäre dann $\varphi = \Phi_1$ Zeit-1-Abbildung eines kontinuierlichen Systems $(\mathbb{T}, \mathcal{X}, \Phi)$, so bekäme man gemäß der Halbgruppeneigenschaft auch $\varphi(\Phi(t, x_0)) = \Phi(t, x_1)$, $\varphi(\Phi(t, x_1)) = \Phi(t, x_0)$ für alle $t \in \mathbb{T}$, und wegen der Eindeutigkeit der Punkte x_0, x_1 könnte $\Phi(\,\cdot\,, x_0)$ nur die beiden Werte x_0 und x_1 annehmen. Bei kontinuierlicher Zeit müsste die Flusslinie $\Phi(\,\cdot\,, x_0)$ damit einerseits konstant sein und andererseits aber $\Phi(0, x_0) = x_0$ mit $\Phi(1, x_0) = \varphi(x_0) = x_1$ verbinden, ein Widerspruch! Das (im Wesentlichen) gleiche Argument erfasst übrigens auch viel allgemeinere $\varphi \in \mathrm{C}^0(\mathcal{X}, \mathcal{X})$ und zeigt, dass sie keine Zeit-1-Abbildungen eines kontinuierlichen Systems sein können: Es funktioniert, sobald es irgendeine Zahl $\ell \geq 2$ gibt, so dass — in der Terminologie des nächsten Kapitels — die Menge aller diskreten Orbits der Periode ℓ mehrere Zusammenhangskomponenten aufweist und mindestens einer dieser Orbits auch mehrere der Zusammenhangskomponenten schneidet.

(4) In spezifischen Situationen gibt es übrigens noch andere Möglichkeiten, aus kontinuierlichen Dynamiken diskrete zu erhalten, vor allem durch sogenannte Poincaré-Abbildungen. Dies wird im Verlauf der Vorlesung eventuell noch thematisiert.

(5) Als Kuriosität am Rande sei schließlich erwähnt, dass man **durch Abänderung des Zustandsraums** „schummeln“ und dann **diskrete Systeme** (in den allermeisten Fällen) **doch kontinuierlich erweitern** kann: Ist $\mathcal{X}$ metrischer Raum und $\varphi \in \mathrm{C}^0(\mathcal{X}, \mathcal{X})$, so betrachtet man dazu die Äquivalenzrelation $\sim$ auf $[0,1]\times\mathcal{X}$, die von den Relationen $(0, \varphi(x)) \sim (1, x)$ mit $x \in \mathcal{X}$ erzeugt wird. Der Faktorraum $\mathcal{X}_{\mathrm{s}} := ([0,1]\times\mathcal{X})/\!\sim$ nach der gerade erklärten Äquivalenzrelation kann, jedenfalls für Lipschitz-stetiges φ, selbst als metrischer Raum aufgefasst[7] werden. Auf diesem neuen Zustandsraum $\mathcal{X}_{\mathrm{s}}$ erhält man — nach Verifikation der Wohldefiniertheit und der entsprechenden Eigenschaften, worauf hier aber nicht im Detail eingegangen wird — einen Fluss $\Psi\colon \mathbb{R}_0^+ \times \mathcal{X}_{\mathrm{s}} \to \mathcal{X}_{\mathrm{s}}$, den sogenannten **Suspensionsfluss**, durch

$$\Psi(t, [\theta, x]_\sim) := [t+\theta - \lfloor t+\theta \rfloor, \varphi^{\lfloor t+\theta \rfloor}(x)]_\sim \qquad \text{für alle } t \in \mathbb{R}_0^+\,,\ [\theta, x]_\sim \in \mathcal{X}_{\mathrm{s}}\,,$$

wobei $[\theta, x]_\sim$ für die Äquivalenzklasse von $(\theta, x) \in [0,1]\times\mathcal{X}$ bezüglich der Relation $\sim$ steht und $\lfloor r \rfloor := \max\{z \in \mathbb{Z} : z \leq r\}$ für den ganzzahligen Anteil von $r \in \mathbb{R}$.

[7] Für den Faktorraum $\mathcal{Y}/\!\sim$ eines metrischen Raums $\mathcal{Y}$ nach einer Äquivalenzrelation $\sim$ und $[p]_\sim, [q]_\sim \in \mathcal{Y}/\!\sim$ erklärt man allgemein $\mathrm{d}_{\mathcal{Y}/\sim}([p]_\sim, [q]_\sim) := \inf \sum_{i=1}^m \mathrm{d}_{\mathcal{Y}}(p_i, q_i)$, wobei das Infimum über alle $m \in \mathbb{N}$ und alle $p_1, p_2, \ldots, p_m, q_1, q_2, \ldots, q_m \in \mathcal{Y}$ gebildet wird, für die $p \sim p_1$, $q_1 \sim p_2$, $q_2 \sim p_3$, …, $q_{m-1} \sim p_m$, $q_m \sim q$ gelten. Diese Bildung liefert im Allgemeinen nur eine Pseudo-Metrik, bezüglich der zwei verschiedene Punkte Abstand 0 aufweisen können, im Fall des oben betrachteten Faktorraums $\mathcal{X}_{\mathrm{s}}$ und unter der Lipschitz-Voraussetzung an φ ergibt sich aber sogar eine echte Metrik.

Anschaulich kann man sich das dynamische System $(\mathbb{R}_0^+, \mathcal{X}_\mathrm{s}, \Psi)$ vorstellen, indem man die Faktorisierung nach $\sim$ zunächst vergisst und die von $(0,x) \in [0,1]\times\mathcal{X}$ ausgehende zeitliche Evolution betrachtet. Diese bewegt sich für Zeiten aus $[0,1]$ bei konstantem x entlang der Linie $[0,1]\times\{x\}$ bis $(1,x)$. Dann kommt die Relation $\sim$ ins Spiel, φ wird praktisch instantan zum Zeitpunkt 1 angewendet, und der mit $(1,x)$ identifizierte Punkt $(0,\varphi(x))$ „auf der anderen Seite“ von $[0,1]\times\mathcal{X}$ ist ebenfalls erreicht. Für Zeiten aus $[1,2]$ folgt die Entwicklung der Linie $[0,1]\times\{\varphi(x)\}$ bis $(1,\varphi(x))$, und mit der zweiten instantanen Anwendung von φ wird auch $(0,\varphi^2(x))$ erreicht. Analog erfolgt der weitere Verlauf.

In gewisser Weise ist das diskrete System $(\mathbb{N}_0, \mathcal{X}, \varphi)$ hier also im kontinuierlichen System $(\mathbb{R}_0^+, \mathcal{X}_\mathrm{s}, \Psi)$ enthalten. Nichtsdestotrotz agiert der Fluss Ψ aber statt auf $\mathcal{X}$ auf dem neuen Zustandsraum $\mathcal{X}_\mathrm{s}$, der üblicherweise auch topologisch anders beschaffen ist. Ist zum Beispiel $\mathcal{X}$ ein Intervall und φ bijektiv, so ist $\mathcal{X}_\mathrm{s}$ homöomorph zu einem Zylindermantel oder einem Möbiusband, je nachdem, ob φ orientierungserhaltend oder orientierungsumkehrend ist; und ist $\mathcal{X} = \mathrm{S}^1$ die Kreislinie und φ bijektiv, so ist X_s homöomorph zu einem Torus oder einer Kleinschen Flasche, wieder je nachdem, ob φ orientierungserhaltend oder -umkehrend ist.

Kapitel 2

Geometrische Begriffe bei dynamischen Systemen

2.1 Orbits und periodische Punkte

In den vorausgehenden Beispielen (2) und (3) trat bereits ein periodisches Verhalten auf. Im Folgenden werden einige hierzu passende Begriffe formal eingeführt und der diskrete Fall genauer untersucht.

Definition 2.1 (**Orbits, Bahnen**). *Sei* $(\mathbb{T}, \mathcal{X}, \Phi)$ *ein dynamisches System und* $x \in \mathcal{X}$. *Dann nennt man*

$$\mathcal{O}^+(x) := \{\Phi(t,x) \,:\, 0 \leq t \in \mathbb{T}\} \qquad \textit{und} \qquad \mathcal{O}^-(x) := \{\Phi(t,x) \,:\, 0 \geq t \in \mathbb{T}\}$$

den positiven und den negativen ***Halborbit*** *von* x (*unter* Φ) *sowie*

$$\mathcal{O}(x) := \mathcal{O}^+(x) \cup \mathcal{O}^-(x) = \{\Phi(t,x) \,:\, t \in \mathbb{T}\}$$

den ***Orbit*** *oder die* ***Bahn*** *von* x (*unter* Φ).

Bemerkungen.

(1) Stets gilt $x \in \mathcal{O}^+(x) \cap \mathcal{O}^-(x) \subset \mathcal{O}(x)$.

(2) Für $\mathbb{T} \in \{\mathbb{N}_0, \mathbb{R}_0^+\}$ genügt es, $\mathcal{O}(x)$ zu betrachten, da $\mathcal{O}^-(x) = \{x\}$, $\mathcal{O}^+(x) = \mathcal{O}(x)$ gelten.

(3) Für $\mathbb{T} \in \{\mathbb{Z}, \mathbb{R}\}$ (also $\mathbb{T}$ Gruppe) und $x, y \in \mathcal{X}$ gelten

$$\begin{aligned} y \in \mathcal{O}(x) &\iff x \in \mathcal{O}(y)\,, \\ y \in \mathcal{O}^+(x) &\iff x \in \mathcal{O}^-(y)\,. \end{aligned}$$

(4) **Der Orbit $\mathcal{O}(x)$ ist das Bild der Bahnlinie $\Phi(\,\cdot\,, x)$** und wird bei manchen Autoren gar nicht allzu genau von dieser unterschieden. Auch bei präziser Verwendung der Terminologie spricht man im Zeit-diskreten Fall aber vorzugsweise über Orbits und im kontinuierlichen Fall vorzugsweise über Bahnlinien.

Definition 2.2 (**Ruhelagen, Fixpunkte**). *Man nennt $x \in \mathcal{X}$ eine **Ruhelage**, eine Gleichgewichtslage oder ein Equilibrium eines dynamischen Systems $(\mathbb{T}, \mathcal{X}, \Phi)$, wenn $\mathcal{O}(x) = \{x\}$ (mit anderen Worten also $\Phi(t, x) = x$ für alle $t \in \mathbb{T}$) gilt. Insbesondere in den Fällen $\mathbb{T} = \mathbb{N}_0$ und $\mathbb{T} = \mathbb{Z}$ bezeichnet man eine Ruhelage oft auch als einen **Fixpunkt** des Systems $(\mathbb{T}, \mathcal{X}, \Phi)$.*

Definition 2.3 (**periodische Punkte**). *Man nennt $x \in \mathcal{X}$ einen **Punkt der (minimalen) Periode ℓ** für ein dynamisches System $(\mathbb{T}, \mathcal{X}, \Phi)$ mit $0 < \ell \in \mathbb{T}$, wenn $\Phi(\ell, x) = x$ sowie $\Phi(t, x) \neq x$ für alle $t \in (0, \ell) \cap \mathbb{T}$ gelten.*

Bemerkungen.

(1) Bei diskreter Zeit $\mathbb{T} \in \{\mathbb{N}_0, \mathbb{Z}\}$ sind die Punkte der Periode 1 genau die Fixpunkte.

(2) Für einen Punkt x der Periode ℓ ist $\mathcal{O}(x)$ (als stetiges Bild von $[0, \ell] \cap \mathbb{T}$) stets kompakt und besteht bei diskreter Zeit $\mathbb{T} \in \{\mathbb{N}_0, \mathbb{Z}\}$ aus genau ℓ Punkten. Für $\mathbb{T} = \mathbb{Z}$ gilt auch umgekehrt, dass ein Punkt mit einem Orbit aus genau ℓ Punkten Periode ℓ besitzt.

(3) Ist x ein Punkt der Periode ℓ, so besitzt auch jeder andere Punkt im Orbit $\mathcal{O}(x)$ Periode ℓ, daher nennt man $\mathcal{O}(x)$ dann auch einen **(ℓ-)periodischen Orbit**.

Beispiele (für **Fixpunkte** und **periodische Punkte**). Das Verhalten in den Beispielen aus Kapitel 1 lässt sich wie folgt zusammenfassen:

- In Beispiel (1) ist 0 stets ein Fixpunkt. Für $\beta = 1$ sind auch alle anderen Punkte Fixpunkte, andernfalls gibt es keine weiteren Fixpunkte oder periodischen Punkte. In Beispiel (4) verhält sich dies analog (wobei der Fall $\gamma = 0$ dem Fall $\beta = 1$ entspricht).
- In Beispiel (2) gibt es keine Fixpunkte, aber 1, 4 und 2 sind Punkte der Periode 3. Teil der ungelösten Vermutung ist die Aussage, dass es keine weiteren periodischen Punkte gibt.
- In Beispiel (3) sind für $\xi = 1$ alle Punkte Fixpunkte, und für den Fall einer ℓ-ten Einheitswurzel ξ (mit $\ell \in \mathbb{N}$) weisen alle Punkte die Periode ℓ auf. Für $\xi = \exp(2\pi \mathbf{i} r)$, $r \in \mathbb{R} \setminus \mathbb{Q}$ gibt es keine periodischen Punkte.

Für **Zeit-diskrete Systeme in einer Dimension** (d.h. mit einem Intervall in $\mathbb{R}$ als Zustandsraum) gelten die folgenden beiden Sätze über die Existenz von Fixpunkten und den Zusammenhang zwischen Punkten unterschiedlicher Periode.

Satz 2.4 (Fixpunktsatz für Intervalle). *Sei I* kompaktes *Intervall in $\mathbb{R}$ und $\varphi \in \mathrm{C}^0(I, \mathbb{R})$. Gilt entweder $\varphi(I) \subset I$ oder $\varphi(I) \supset I$, so besitzt φ einen Fixpunkt $x \in I$ (d.h. es ist $\varphi(x) = x$).*

Bemerkung. Im Fall der Alternative $\varphi(I) \subset I$ ist φ eine Selbstabbildung von I, und es handelt sich bei x um einen Fixpunkt des Systems $(\mathbb{N}_0, I, \varphi)$ (wobei $(\mathbb{N}_0, I, \varphi)$, daran sei hier erinnert, im vorigen Kapitel als abkürzende Schreibweise für $(\mathbb{N}_0, I, \Phi)$ mit $\Phi_k = \varphi^k$ eingeführt wurde).

Satz 2.5 (von **Sarkovskii**, $\sim$1964). *Seien die natürlichen Zahlen gemäß der Sarkovskii-Ordnung*

$$3, 5, 7, 9, \ldots, 3{\cdot}2, 5{\cdot}2, 7{\cdot}2, 9{\cdot}2, \ldots, 3{\cdot}2^2, 5{\cdot}2^2, 7{\cdot}2^2, 9{\cdot}2^2, \ldots, 3{\cdot}2^3, 5{\cdot}2^3, 7{\cdot}2^3, 9{\cdot}2^3, \ldots, 2^4, 2^3, 2^2, 2, 1$$

sortiert. Ist I ein Intervall in $\mathbb{R}$, besitzt das System $(\mathbb{N}_0, I, \varphi)$ mit $\varphi \in \mathrm{C}^0(I, I)$ einen Punkt der minimalen Periode $m \in \mathbb{N}$, und steht $n \in \mathbb{N}$ in der Sarkovskii-Ordnung rechts von m, so besitzt $(\mathbb{N}_0, I, \varphi)$ auch einen Punkt der minimalen Periode n.

Bemerkungen.

(1) Anders als in Satz 2.4 benötigt man in Satz 2.5 keine Kompaktheits- oder andere Zusatzvoraussetzung an das Intervall I.

(2) Ein Spezialfall von Satz 2.5 ist die Aussage, dass das Auftreten eines Punktes einer beliebigen Periode schon die Existenz eines Fixpunktes erzwingt. Eine andere Teilaussage ist, dass das Auftreten eines Punktes des Periode 3 die Existenz von Punkten aller anderen Perioden impliziert. Die letztere Aussage läuft auch unter der Parole **„Three implies Chaos“**.

(3) **Satz 2.5 gilt nicht für andere Zustandsräume als Intervalle.** Insbesondere zeigt Beispiel (3) aus dem vorigen Kapitel, dass der Satz nicht einmal auf der Einheitskreislinie S^1 gültig bleibt.

Die *Beweise* von Satz 2.4 und vieler Teilaussagen von Satz 2.5 benötigen im Wesentlichen nur den Zwischenwertsatz und werden Thema der Übungen sein. Ein vollständiger Beweis von Satz 2.5 lässt sich ebenfalls nur mit diesem Hilfsmittel bewerkstelligen, ist aber kombinatorisch und bezüglich der Anzahl der Fälle etwas aufwendiger, weshalb er hier nicht ausgeführt wird. □

2.2 Limesmengen und (andere) invariante Mengen

Wie schon erwähnt ist das Langzeitverhalten bei dynamischen Systemen von besonderem Interesse. Als Nächstes werden einige Konzepte eingeführt, die dem Studium eben dieses Verhaltens dienen.

Definition 2.6 (Limesmengen). *Sei $(\mathbb{T}, \mathcal{X}, \Phi)$ ein dynamisches System. Die **ω-Limesmenge** oder **ω-Grenzmenge** $\omega(x)$ eines Punktes $x \in \mathcal{X}$ ist definiert durch*

$$\omega(x) := \bigcap_{0 \leq t \in \mathbb{T}} \overline{\mathcal{O}^+(\Phi(t,x))} \subset \mathcal{X}$$

*und die **α-Limesmenge** oder **α-Grenzmenge** $\alpha(x)$ durch*

$$\alpha(x) := \bigcap_{0 \geq t \in \mathbb{T}} \overline{\mathcal{O}^-(\Phi(t,x))} \subset \mathcal{X}\,.$$

Bemerkungen.

(1) Die ω-Limesmenge $\omega(x)$ ist die **Menge aller Häufungspunkte** von $\Phi(t,x)$ in $\mathcal{X}$ **bei $t \to \infty$**. Dies erkennt man im Wesentlichen durch explizites Ausschreiben der Definitionen.

(2) Im Fall $\mathbb{T} \in \{\mathbb{N}_0, \mathbb{R}_0^+\}$ ist $\alpha(x) = \{x\}$ keine interessante Bildung. Im Gruppenfall $\mathbb{T} \in \{\mathbb{Z}, \mathbb{R}\}$ ist $\alpha(x)$ die Menge aller Häufungspunkte von $\Phi(t,x)$ in $\mathcal{X}$ bei $t \to -\infty$.

(3) Die Limesmengen $\omega(x)$ und $\alpha(x)$ sind **stets abgeschlossen**, denn beliebige Schnitte abgeschlossener Mengen sind abgeschlossen.

(4) **Ist $\mathcal{O}^+(x)$ beziehungsweise $\mathcal{O}^-(x)$ relativ kompakt** in $\mathcal{X}$ (d.h. der Abschluss in $\mathcal{X}$ ist kompakt), **so ist $\omega(x)$ beziehungsweise $\alpha(x)$ nicht-leer und kompakt.** Dies folgt einerseits daraus, dass jede Folge im Kompaktum $\overline{\mathcal{O}^\pm(x)}$ einen Häufungspunkt besitzt, und andererseits daraus, dass jede abgeschlossene Teilmenge des Kompaktums $\overline{\mathcal{O}^\pm(x)}$ selbst kompakt ist.

(5) Für $\mathcal{X} = \mathbb{R}^N$ und jeden anderen metrischen Raum $\mathcal{X}$ mit der Heine-Borel-Eigenschaft[1] bedeutet die relative Kompaktheit von $\mathcal{O}^{\pm}(x)$ in (4) nichts anderes als Beschränktheit von $\mathcal{O}^{\pm}(x)$.

(6) Ohne die Voraussetzung der Kompaktheit (oder Beschränktheit) von $\overline{\mathcal{O}^{\pm}(x)}$ können $\omega(x)$ und $\alpha(x)$ durchaus leer oder unbeschränkt sein; dies zeigt sich an einfachen Beispielen später in diesem Abschnitt.

(7) Für einen Fixpunkt x gilt natürlich $\omega(x) = \alpha(x) = \{x\}$. Bei periodischem x ist $\omega(x) = \mathcal{O}^+(x) = \mathcal{O}(x)$, und im Fall $\mathbb{T} \in \{\mathbb{Z}, \mathbb{R}\}$ stimmt diese Menge mit $\alpha(x) = \mathcal{O}^-(x)$ überein.

Darüber hinaus haben Limesmengen noch folgende generellen Eigenschaften, die hier der Einfachheit halber nur für $\omega(x)$ formuliert werden.

Satz 2.7 (Konvergenz gegen und Zusammenhang von Limesmengen). *Sei $(\mathbb{T}, \mathcal{X}, \Phi)$ ein dynamisches System, und für $x \in \mathcal{X}$ sei $\mathcal{O}^+(x)$ relativ kompakt in $\mathcal{X}$ (oder beschränkt, wenn $\mathcal{X} = \mathbb{R}^n$ ist oder $\mathcal{X}$ die Heine-Borel-Eigenschaft hat). Dann gilt*

$$\lim_{t\to\infty} \operatorname{dist}(\Phi(t,x), \omega(x)) = 0\,, \tag{2.1}$$

*und **bei kontinuierlicher Zeit** $\mathbb{T} \in \{\mathbb{R}_0^+, \mathbb{R}\}$ ist **$\omega(x)$ zusammenhängend**.*

Beweis. Zum Nachweis von (2.1) lässt sich ein Widerspruchsargument verwenden. Wäre (2.1) falsch, so gäbe es ein $\varepsilon > 0$ und eine ∞-Folge $(t_k)_{k\in\mathbb{N}}$ in $\mathbb{T}$ mit $\operatorname{dist}(\Phi(t_k,x),\omega(x)) \geq \varepsilon$ für alle $k \in \mathbb{N}$. Wegen Kompaktheit von $\overline{\mathcal{O}^+(x)}$ müsste dann eine Teilfolge von $(\Phi(t_k,x))_{k\in\mathbb{N}}$ gegen einen Limes $y \in \mathcal{X}$ konvergieren. Einerseits müsste $y \in \omega(x)$ gelten, andererseits wegen der Stetigkeit von $\operatorname{dist}(\,\cdot\,,\omega(x))$ aber auch $\operatorname{dist}(y,\omega(x)) \geq \varepsilon$. Dieser Widerspruch vervollständigt den Beweis von (2.1).

Um den Zusammenhang von $\omega(x)$ nachzuweisen, wird eine Zerlegung $\omega(x) = \omega_1 \cup \omega_2$ in disjunkte, relativ offene Teilmengen ω_1 und ω_2 betrachtet. Dann sind $\omega_1 = \omega(x) \setminus \omega_2$ und $\omega_2 = \omega(x) \setminus \omega_1$ auch abgeschlossene Teilmengen des Kompaktums ω und damit selbst kompakt. Es wird nun der Fall $\omega_1 \neq \emptyset \neq \omega_2$ betrachtet. In diesem ist $\delta := \operatorname{dist}(\omega_1,\omega_2) \in \mathbb{R}^+$. Außerdem gibt es, da ω_1 und ω_2 jeweils mindestens einen Häufungspunkt von $\Phi(t,x)$ bei $t \to \infty$ enthalten, beliebig große $t_1 \in \mathbb{T}$ mit $\operatorname{dist}(\Phi(t_1,x),\omega_1) \leq \delta/2$ und ebenfalls beliebig große (andere) $t_2 \in \mathbb{T}$ mit $\operatorname{dist}(\Phi(t_2,x),\omega_2) \leq \delta/2$ und folglich $\operatorname{dist}(\Phi(t_2,x),\omega_1) \geq \delta/2$. Bei kontinuierlicher Zeit finden sich zwischen solchen t_1 und t_2 gemäß dem Zwischenwertsatz weitere beliebig große $t_3 \in \mathbb{T}$ mit $\operatorname{dist}(\Phi(t_3,x),\omega_1) = \delta/2$. Wegen Kompaktheit von $\mathcal{O}^+(x)$ besitzt dann $\Phi(t,x)$ bei $t \to \infty$ einen Häufungspunkt z mit $\operatorname{dist}(z,\omega_1) = \delta/2$. Nun kann $z \in \omega(x)$ weder in ω_1 noch in ω_2 liegen, aber es gilt $\omega(x) = \omega_1 \cup \omega_2$. Somit kann der betrachtete Fall $\omega_1 \neq \emptyset \neq \omega_2$ tatsächlich nicht eintreten, und es verbleibt nur die Möglichkeit, dass entweder $\omega_1 = \emptyset$ oder $\omega_2 = \emptyset$ gilt. Damit ist der Zusammenhang von $\omega(x)$ (anhand der Definition einer zusammenhängenden Menge) nachgewiesen. □

Anhand ihres Langzeitverhaltens bei $t \to +\infty$ unterscheidet man (unter anderem) folgende Typen von Zuständen beziehungsweise Punkten.

[1] Ein metrischer Raum besitzt die Heine-Borel-Eigenschaft, wenn jede beschränkte abgeschlossene Menge kompakt ist. Diese Eigenschaft kann man so interpretieren, dass der metrische Raum in einem gewissen Sinne endlichdimensional ist.

Definition 2.8 (**transitive, rekurrente, (nicht-)wandernde, fast-periodische Punkte**). *Seien* $(\mathbb{T}, \mathcal{X}, \Phi)$ *ein dynamisches System und* $x \in \mathcal{X}$. *Dann heißt der Punkt* x *...*

- ***(anhaltend) transitiv****, wenn* $\mathcal{O}^+(\Phi(t,x))$ *für alle* $t \in \mathbb{T}$ *dicht in* $\mathcal{X}$ *ist,*
- ***rekurrent****, wenn es zu jeder Umgebung* U *von* x *beliebig große* $t \in \mathbb{T}$ *mit* $\Phi_t(x) \in U$ *(oder äquivalent mit*[2] $x \in (\Phi_t)^{-1}(U)$*) gibt,*
- ***nicht-wandernd****, wenn für jede Umgebung* U *von* x *beliebig große* $t \in \mathbb{T}$ *mit* $\Phi_t(U) \cap U \neq \emptyset$ *(oder äquivalent mit* $(\Phi_t)^{-1}(U) \cap U \neq \emptyset$*) existieren,*
- ***wandernd****, wenn es eine Umgebung* U *von* x *mit* $\Phi_t(U) \cap U = \emptyset$ *(oder äquivalent mit* $(\Phi_t)^{-1}(U) \cap U = \emptyset$*) für alle hinreichend großen* $t \in \mathbb{T}$ *gibt,*
- ***fast-periodisch****, wenn für jede Umgebung* U *von* x *ein* $\ell \in \mathbb{R}^+$ *existiert, derart dass* $\Phi([T, T+\ell] \cap \mathbb{T}, x) \cap U \neq \emptyset$ *für jedes* $T \in \mathbb{T}$ *gilt. (Man spricht auch von* ***beschränkten Lücken*** *der Menge* $\{t \in \mathbb{T} : \Phi(t,x) \in U\}$*; die Lücken zwischen verschiedenen ‚Stücken' dieser Menge haben dann nämlich höchstens Länge* ℓ*.)*

Dabei bedeutet die Formulierung ‚beliebig große $t \in \mathbb{T}$*' jeweils, dass es zu jedem* $T \in \mathbb{N}$ *ein solches* $t \in \mathbb{T}$ *mit* $t \geq T$ *gibt; und der Passus ‚für alle hinreichend großen* $t \in \mathbb{T}$*' steht für die Existenz eines* $T \in \mathbb{N}$*, so dass das Gesagte alle* $t \in \mathbb{T}$ *mit* $t \geq T$ *betrifft.*

Bemerkungen.

(1) Bei den ersten drei Begriffen sind **alternative Charakterisierungen** möglich und nützlich:

- Ein Punkt $x \in \mathcal{X}$ ist genau dann transitiv, wenn es zu jedem $y \in \mathcal{X}$ eine gegen $+\infty$ konvergente Folge $(t_k)_{k\in\mathbb{N}}$ in $\mathbb{T}$ mit $\lim_{k\to\infty} \Phi(t_k, x) = y$ gibt, mit anderen Worten also genau dann, wenn $\omega(x) = \mathcal{X}$ gilt.
- Ein Punkt $x \in \mathcal{X}$ ist genau dann rekurrent, wenn es eine gegen $+\infty$ konvergente Folge $(t_k)_{k\in\mathbb{N}}$ in $\mathbb{T}$ mit $\lim_{k\to\infty} \Phi(t_k, x) = x$ gibt, mit anderen Worten also genau dann, wenn $x \in \omega(x)$ gilt.
- Ein Punkt $x \in \mathcal{X}$ ist genau dann nicht-wandernd, wenn es eine gegen x konvergente Folge $(x_k)_{k\in\mathbb{N}}$ in $\mathcal{X}$ und eine gegen $+\infty$ konvergente Folge $(t_k)_{k\in\mathbb{N}}$ in $\mathbb{T}$ mit $\lim_{k\to\infty} \Phi(t_k, x_k) = x$ gibt.

(2) **Jeder transitive Punkt und jeder fast-periodische Punkt sind rekurrent, und jeder rekurrente Punkt ist nicht-wandernd.** Die Umkehrungen gelten im Allgemeinen nicht.

(3) Ein periodischer Punkt der Periode ℓ ist fast-periodisch und damit rekurrent sowie nicht-wandernd; transitiv kann er höchstens bei ℓ-elementigem $\mathcal{X}$ (im Zeit-diskreten Fall) beziehungsweise kompaktem $\mathcal{X}$ (im kontinuierlichen Fall) sein.

(4) Aus der Definition ist klar: Die **Menge der wandernden Punkte** ist stets **offen**, und ihr Komplement, die **Menge der nicht-wandernden Punkte**, ist stets **abgeschlossen**.

[2] Hier steht $(\Phi_t)^{-1}(U)$ für das Urbild von U unter der im irreversiblen Fall nicht unbedingt invertierbaren Flussabbildung Φ_t.

Bemerkung (zur Definition der Transitivität in der Literatur)**.** In der Literatur wird ein transitiver Punkt $x \in \mathcal{X}$ oft einfach als ein Punkt mit in $\mathcal{X}$ dichtem Orbit $\mathcal{O}(x)$ definiert. In den Kontext der anderen obigen Begriffe, die nur das Verhalten bei $t \to +\infty$, nicht das für beschränkte Zeiten t oder bei $t \to -\infty$ betreffen, fügt sich (anhaltende) Transitivität im stärkeren Sinne der Definition 2.8 aber konsistenter ein, weist im allgemeinen Fall bessere Invarianzeigenschaften auf und scheint insofern etwas stringenter. Ist die Zeit diskret und besitzt $\mathcal{X}$ keinen isolierten Punkt[3], so ist anhaltende Transitivität von x im Fall $\mathbb{T} = \mathbb{Z}$ äquivalent zu Dichtheit von $\mathcal{O}^+(x)$ in $\mathcal{X}$ und im Fall $\mathbb{T} = \mathbb{N}_0$ sogar zu Dichtheit von $\mathcal{O}^+(x) = \mathcal{O}(x)$ in $\mathcal{X}$ und damit zum in der Literatur verbreiteten Begriff.

Zwei grundlegende (Existenz-)Sätze für Punkte der obigen Typen folgen.

Satz 2.9 (**Existenz fast-periodischer Punkte**)**.** *Jedes dynamische System* $(\mathbb{T}, \mathcal{X}, \Phi)$ *mit kompaktem* $\mathcal{X}$ *besitzt einen fast-periodischen Punkt* (*der nach Bemerkung* (2) *auch rekurrent und nicht-wandernd ist*)*.*

Der Beweis von Satz 2.9 benötigt noch ein weiteres Konzept und wird daher erst später in diesem Abschnitt geführt.

Satz 2.10 (**Existenz keiner oder vieler transitiver Punkte**)**.** *Für jedes dynamische System* $(\mathbb{T}, \mathcal{X}, \Phi)$ *mit separablem*[4] *metrischen Raum* $\mathcal{X}$ *ist die Menge der transitiven Punkte leer oder eine dichte* G_δ*-Menge*[5] *in* $\mathcal{X}$*.*

Bemerkung. Als direkte Konsequenz von Satz 2.10 ergibt sich, dass die **Menge der transitiven Punkte entweder leer oder residuell**[6] (und damit in einem topologischen Sinne groß) ist.

Beweis von Satz 2.10. Sei $\mathcal{T}$ die Menge der transitiven Punkte des Systems. Ist $\mathcal{T} \neq \emptyset$, so gibt es ein $\xi \in \mathcal{T}$ und bereits die Teilmenge $\mathcal{O}^+(\xi)$ von $\mathcal{T}$ liegt dicht in $\mathcal{X}$.

Um einzusehen, dass $\mathcal{T}$ eine G_δ-Menge ist, überlegt man sich

$$\mathcal{T} = \bigcap_{\substack{i, T \in \mathbb{N} \\ y \in D}} \bigcup_{T \leq t \in \mathbb{T}} (\Phi_t)^{-1}(\mathrm{B}_{1/i}(y)) \tag{2.2}$$

für jede dichte Teilmenge D von X, wobei die Abkürzung $\mathrm{B}_{1/i}(y) := \{\xi \in \mathcal{X} : \mathrm{d}_{\mathcal{X}}(\xi, y) < 1/i\}$ für Kugeln verwendet wurde. Zum Nachweis von (2.2) gilt es im Wesentlichen, die Definition beider Seiten auszuschreiben: Die Inklusion eines $x \in \mathcal{X}$ in der Menge auf der rechten Seite von (2.2) bedeutet, dass es zu jedem $y \in D$ und $i \in \mathbb{N}$ beliebig große $t \in \mathbb{T}$ mit $\Phi(t, x) \in \mathrm{B}_{1/i}(y)$ gibt.

[3] Ein isolierter Punkt eines metrischen Raumes $\mathcal{X}$ ist ein Punkt $x \in \mathcal{X}$, für den $\mathrm{dist}(x, \mathcal{X} \setminus \{x\})$ positiv und damit die Ein-Punkt-Menge $\{x\}$ offen ist.

[4] Man nennt einen metrischen (oder topologischen) Raum $\mathcal{X}$ separabel, wenn er eine abzählbare, dichte Teilmenge enthält.

[5] Als G_δ-Menge bezeichnet man einen abzählbaren Durchschnitt offener Mengen. Das Pendant hierzu sind F_σ-Mengen genannte abzählbare Vereinigungen abgeschlossener Mengen. Jede offene oder abgeschlossene Menge in einem metrischen Raum ist sowohl F_σ als auch G_δ, und jede F_σ- oder G_δ-Menge ist auch Borel-Menge.

[6] Sei A Teilmenge eines metrischen (oder topologischen) Raums $\mathcal{X}$. Dann heißt A mager oder von erster (Bairescher) Kategorie in $\mathcal{X}$, wenn A abzählbare Vereinigung nirgends dichter Mengen ist, und A heißt residuell in $\mathcal{X}$, wenn A Komplement einer mageren Menge in $\mathcal{X}$ ist. Im Kontrast dazu heißt A fett oder von zweiter (Bairescher) Kategorie in $\mathcal{X}$, sobald A nicht mager ist, und der Bairesche Kategoriensatz besagt, dass jeder nicht-leere, vollständige metrische Raum $\mathcal{X}$ (in sich) von zweiter Kategorie ist.

Dies ist gleichbedeutend damit, dass jedes $y \in D$ Häufungspunkt von $\Phi(t,x)$ bei $t \to \infty$ ist, und nach Bemerkung (1) zu Limesmengen auch damit, dass $D \subset \omega(x)$ gilt. Da $\omega(x)$ abgeschlossen und D dicht in $\mathcal{X}$ ist, ist dies weiterhin äquivalent mit $\mathcal{X} = \omega(x)$, also gemäß Bemerkung (1) zu Definition 2.8 auch mit Transitivität von x. Damit ist (2.2) verifiziert. Da die Mengen $(\Phi_t)^{-1}(\mathrm{B}_{1/i}(y))$ als Urbilder der offenen Kugeln $\mathrm{B}_{1/i}(y)$ offen sind, folgt aus (2.2) sofort, dass $\mathcal{T}$ eine G_δ-Menge ist. □

Es folgt ein weiteres zentrales Konzept dieses Abschnitts.

Definition 2.11 (invariante Mengen). *Sei $(\mathbb{T},\mathcal{X},\Phi)$ ein dynamisches System. Eine Teilmenge A von $\mathcal{X}$ heißt (unter Φ) **invariant**, wenn $\mathcal{O}(x) \subset A$ für alle $x \in A$ gilt. Analog spricht man von einer (unter Φ) **positiv invarianten** beziehungsweise **negativ invarianten Menge** A, wenn $\mathcal{O}^+(x) \subset A$ beziehungsweise $\mathcal{O}^-(x) \subset A$ für alle $x \in A$ gilt.*

Bemerkungen.

(1) Eine Menge ist genau dann invariant, wenn sie zugleich positiv und negativ invariant ist.

(2) Beliebige Vereinigungen und Schnitte von unter Φ (positiv/negativ) invarianten Mengen bleiben (positiv/negativ) invariant unter Φ.

(3) Im Fall $\mathbb{T} \in \{\mathbb{N}_0, \mathbb{R}_0^+\}$ ist negative Invarianz trivial (für jede Menge erfüllt), und positive Invarianz stimmt mit Invarianz überein. Somit verbleibt dann nur ein nicht-triviales Konzept von Invarianz.

(4) Trivial gilt: Ein Orbit $\mathcal{O}(x)$ ist stets invariant, ein Halborbit $\mathcal{O}^\pm(x)$ ist positiv/negativ invariant, und jede (positiv/negativ) invariante Menge kann als Vereinigung von (Halb-) Orbits geschrieben werden.

(5) Eine Umformulierung der Definition ergibt

$$A \text{ unter } \Phi \text{ (positiv) invariant} \iff \Phi_t(A) \subset A \text{ für alle (positiven) } t \in \mathbb{T}\,.$$

Im Zeit-diskreten Fall reicht es dabei auch schon, auf der rechten Seite nur $t = \pm 1$ zu betrachten, also nur $\varphi(A) \subset A$ und eventuell noch $\varphi^{-1}(A) \subset A$ für die Zeit-1-Abbildung $\varphi = \Phi_1$ zu verlangen. Im kontinuierlichen Fall kann man sich analog auf nur solche t mit $0 < |t| < \varepsilon$ für ein $\varepsilon > 0$ beschränkten. Im Gruppenfall $\mathbb{T} \in \{\mathbb{Z}, \mathbb{R}\}$ und bei Invarianz schließlich kann man rechts ‚$\subset$' durch ‚$=$' ersetzen.

Die nächsten beiden Sätze zeigen, dass fast alle in diesem Kapitel betrachteten Konzepte tatsächlich invariant sind. Dies ist für Konzepte des Langzeitverhaltens, für die das Verstreichen einer endlichen Zeit (tendenziell) irrelevant ist, auch sehr plausibel.

Satz 2.12 (Invarianz von Limesmengen). *Sei $(\mathbb{T},\mathcal{X},\Phi)$ ein dynamisches System. Für alle $x \in \mathcal{X}$ sind die ω-Limesmenge $\omega(x)$ und im Gruppenfall $\mathbb{T} \in \{\mathbb{Z},\mathbb{R}\}$ auch die α-Limesmenge $\alpha(x)$ unter Φ invariant.*

Beweis. Zum Beweis der Invarianz von $\omega(x)$ sei $y \in \omega(x)$. Gemäß Bemerkung (1) zu Definition 2.6 gibt es dann eine ∞-Folge $(t_k)_{k\in\mathbb{N}}$ in $\mathbb{T}$ mit $\lim_{k\to\infty} \Phi(t_k,x) = y$. Sei weiterhin z ein beliebiger Punkt im Orbit $\mathcal{O}(y)$. Dann lässt sich z als $\Phi(T,y)$ mit $0 \le T \in \mathbb{T}$ schreiben. Mit der Halbgruppeneigenschaft (1.2) und der Stetigkeit von Φ ergibt sich $\lim_{k\to\infty} \Phi(T+t_k,x) =$

$\lim_{k\to\infty} \Phi(T, \Phi(t_k, x)) = \Phi(T, y) = z$. Unter erneuter Verwendung der erwähnten Bemerkung (1) folgt $z \in \omega(x)$, also insgesamt $\mathcal{O}(y) \subset \omega(x)$. Damit ist die Invarianz von $\omega(x)$ anhand ihrer Definition nachgewiesen.

Ist $\mathbb{T} \in \{\mathbb{Z}, \mathbb{R}\}$, so weist man die Invarianz von $\alpha(x)$ durch ein völlig analoges Argument mit $(-\infty)$-Folgen nach. □

Satz 2.13 (**Invarianz verschiedener Punktmengen**). *Sei* $(\mathbb{T}, \mathcal{X}, \Phi)$ *ein dynamisches System. Dann sind stets invariant*: *die Menge der transitiven Punkte, die Menge der rekurrenten Punkte, die Menge der nicht-wandernden Punkte, die Menge der fast-periodischen Punkte sowie die Menge der ℓ-periodischen Punkte* (*letztere entweder mit freiem oder beliebig fixierten positivem* $\ell \in \mathbb{T}$). *Bei reversibler Zeit* $\mathbb{T} \in \{\mathbb{Z}, \mathbb{R}\}$ *ist auch die Menge der wandernden Punkte invariant.*

Bemerkung. Bei irreversibler Zeit muss die Menge der wandernden Punkt nicht invariant sein; dies zeigt das triviale Beispiel $(\mathbb{N}_0, \mathbb{R}, 0)$ mit der Nullabbildung $0 \in \mathrm{C}^0(\mathbb{R}, \mathbb{R})$, bei dem genau die Punkte in $\mathbb{R} \setminus \{0\}$ wandernd sind.

Beweis von Satz 2.13. Die Invarianz transitiver und periodischer Punkte ergibt sich direkt aus den jeweiligen Definitionen, daher werden in den folgenden Absätzen nur die Invarianz fast-periodischer, rekurrenter und (nicht-)wandernder Punkte behandelt:

Sei $x \in \mathcal{X}$ ein fast-periodischer Punkt und $s \in \mathbb{T}$. Ist dann U eine Umgebung von $\Phi_s(x)$ in $\mathcal{X}$, so folgt mit der Stetigkeit von Φ_s, dass das Urbild $V := (\Phi_s)^{-1}(U)$ eine Umgebung von x in $\mathcal{X}$ ist. Gemäß der Definition der Fast-Periodizität gibt es dann ein $\ell \in \mathbb{R}^+$, so dass jedes Intervall $[T, T+\ell] \cap \mathbb{T}$ mit $T \in \mathbb{T}$ ein $t \in \mathbb{T}$ mit $\Phi_t(x) \in V$ enthält. Für dieses t gilt aber auch $\Phi_t(\Phi_s(x)) = \Phi_s(\Phi_t(x)) \in \Phi_s(V) = U$, und damit ist $\Phi_s(x)$ fast-periodischer Punkt.

Die Invarianz rekurrenter Punkte erhält man mit einer völlig analogen Argumentation.

Sei $x \in \mathcal{X}$ ein nicht-wandernder Punkt und $s \in \mathbb{T}$. Ist dann U eine Umgebung von $\Phi_s(x)$ in $\mathcal{X}$, so ist $V := (\Phi_s)^{-1}(U)$ eine Umgebung von x in $\mathcal{X}$. Es gilt dann $V \cap \Phi_t(V) \neq \emptyset$ für beliebig große $t \in \mathbb{T}$, und daraus folgt $\emptyset \neq \Phi_s(V \cap \Phi_t(V)) \subset \Phi_s(V) \cap \Phi_t(\Phi_s(V)) = U \cap \Phi_t(U)$ für beliebig große $t \in \mathbb{T}$. Also ist $\Phi_s(x)$ nicht-wandernder Punkt.

Im Gruppenfall $\mathbb{T} \in \{\mathbb{Z}, \mathbb{R}\}$ wurde schon bemerkt, dass Invarianz mit Hilfe von Gleichheiten des Typs $\Phi_t(A) = A$ charakterisiert werden kann. In Anbetracht dieser Charakterisierung ergibt sich die Invarianz der wandernden Punkte durch Komplementbildung aus der der nicht-wandernden Punkte. □

Mit Hilfe des Konzepts invarianter Mengen kann nun auch der zuvor ausgelassene Beweis des Existenzsatzes für fast-periodische (und rekurrente) Punkte nachgetragen werden:

Beweis von Satz 2.9. Es gilt zunächst zu begründen, dass das System $\mathcal{S}$ der nicht-leeren, abgeschlossenen und invarianten Teilmengen von $\mathcal{X}$ ein bezüglich Mengeninklusion minimales Element K enthält. Die Existenz eines solchen Elements folgt tatsächlich direkt aus dem Zornschen Lemma[7], sobald man die hierfür benötigten Voraussetzungen verifiziert, nämlich die, dass $\mathcal{S}$ nicht-leer ist und jedes total-geordnete Teilsystem $\mathcal{P}$ von $\mathcal{S}$ eine untere Schranke besitzt. Ersteres folgt aus der Beobachtung $\mathcal{X} \in \mathcal{S}$ und für zweiteres kann man den Durchschnitt $\bigcap_{A\in\mathcal{P}} A$ aller

[7]Das Zornsche Lemma besagt, dass eine nicht-leere halb-geordnete Menge, in der jede total-geordnete Teilmenge eine untere/obere Schranke besitzt, stets ein minimales/maximales Element enthält. Die Gültigkeit dieses Lemmas ist äquivalent zum Auswahlaxiom der Mengenlehre.

Mengen des Teilsystems $\mathcal{P}$ als Schranke verwenden. Dabei ist $\bigcap_{A\in\mathcal{P}} A$ per Konstruktion abgeschlossen und invariant, dass dieser Durchschnitt auch nicht-leer ist, wird aber erst durch die Kompaktheit von $\mathcal{X}$ sichergestellt: Wäre nämlich $\bigcap_{A\in\mathcal{P}} A = \emptyset$, so würden die Komplemente der Mengen in $\mathcal{P}$ eine Überdeckung von $\mathcal{X}$ bilden, und diese besäße eine endliche Teilüberdeckung. Folglich wäre auch schon ein endlicher Durchschnitt von Mengen in $\mathcal{P}$ leer, was in Anbetracht der totalen Ordnung auf $\mathcal{P}$ unmöglich ist. Damit ist die Existenz des minimalen Elements $K \in \mathcal{S}$ begründet.

Als Nächstes wird gezeigt, dass jedes $x \in K$ fast-periodischer Punkt ist. Sei dazu U eine offene Umgebung eines solchen x. Dann ist

$$K \setminus \bigcup_{t\in\mathbb{T}} (\Phi_t)^{-1}(U)$$

abgeschlossen, invariant und Teilmenge von $K \setminus \{x\}$, also muss diese Menge wegen der Minimalität von K leer sein. Damit ist $((\Phi_t)^{-1}(U))_{t\in\mathbb{T}}$ offene Überdeckung des Kompaktums K und besitzt eine endliche Teilüberdeckung durch Mengen $(\Phi_{t_1})^{-1}(U)$, $(\Phi_{t_2})^{-1}(U)$, …, $(\Phi_{t_k})^{-1}(U)$ mit $t_1, t_2, \ldots, t_k \in \mathbb{T}$. Für $\ell/2 := 1 + \max\{|t_1|, |t_2|, \ldots, |t_k|\} \in \mathbb{T}$ und beliebiges $T \in \mathbb{T}$ gilt dann $\Phi_{T+\ell/2}(x) \in K$ (wegen Invarianz), und es gibt einen Punkt der Form $T+\ell/2+t_i \in [T, T+\ell] \cap \mathbb{T}$, so dass $\Phi_{T+\ell/2}(x) \in (\Phi_{t_i})^{-1}(U)$ ist. Letzteres bedeutet aber $\Phi(T+\ell/2+t_i, x) \in U$ und somit insbesondere $\Phi([T, T+\ell] \cap \mathbb{T}, x) \cap U \neq \emptyset$. Folglich ist Fast-Periodizität von x nachgewiesen. □

Zur (zumindest teilweisen) Illustration der in diesem Abschnitt besprochenen Konzepte sei schließlich noch einmal auf die einfachen Beispiele des ersten Vorlesungskapitels zurückgegriffen:

Beispiele (zu **Limesmengen, speziellen Punkten und invarianten Mengen**)**.** Das Verhalten in den Beispielen des Kapitels 1 lässt sich wie folgt zusammenfassen:

- Beispiel (1): In den trivialen Fällen $x = 0$ und $\beta = 1$ gilt $\alpha(x) = \omega(x) = \{x\}$. Für $\beta > 1$, $x > 0$ ist $\alpha(x) = \{0\}$, $\omega(x) = \emptyset$. Für $\beta < 1$, $x > 0$ ist umgekehrt $\alpha(x) = \emptyset$, $\omega(x) = \{0\}$. Im Fall $\beta = 1$ sind außerdem alle Punkte in $\mathbb{R}_0^+$ Fixpunkte und somit fast-periodisch, rekurrent und nicht-wandernd. Für $\beta \neq 1$ ist jede dieser Eigenschaften einzig für den Punkt 0 erfüllt. Transitive Punkte treten nicht auf.

- Beispiel (2): Die Collatz-Vermutung besagt in der inzwischen eingeführten Terminologie $\omega(n) = \{1, 4, 2\}$ für alle $n \in \mathbb{N}$. Fast-periodisch, rekurrent und nicht-wandernd bedeuten hier wegen des diskreten Zustandsraums $\mathbb{N}$ nichts anderes als periodisch, gemäß Vermutung haben also nur 1, 4, 2 diese Eigenschaften. Transitive Punkte treten nicht auf.

  Neben $\{1, 4, 2\}$ gibt es in diesem Beispiel übrigens viele weitere invariante Mengen (z.B. die Menge der Zweier-Potenzen oder die endliche Menge der in Abbildung 2 auftretenden Zahlen). Eine weitere Umformulierung der Vermutung ist die, dass jede nicht-leere invariante Menge die Elemente 1, 4, 2 enthält.

- Beispiel (3): Im Fall einer ℓ-ten Einheitswurzel ξ ist jedes $z \in \mathrm{S}^1$ ein Punkt der Periode ℓ und ist folglich fast-periodisch, rekurrent und nicht-wandernd mit $\omega(z) = \alpha(z) = \mathcal{O}(z) = \{z, z\cdot\xi, z\cdot\xi^2, z\cdot\xi^3, \ldots, z\cdot\xi^{\ell-1}\}$. Transitive Punkte treten in diesem Fall nicht auf.

  Für $\xi = \exp(2\pi\mathbf{i}r)$, $r \in \mathbb{R} \setminus \mathbb{Q}$ sind dagegen alle Punkte $z \in \mathrm{S}^1$ transitiv, fast-periodisch, rekurrent und nicht-wandernd mit $\omega(x) = \alpha(x) = \mathrm{S}^1$. Insbesondere liegen alle nicht-leeren (positiv/negativ) invarianten Mengen dicht in S^1.

- Beispiel (4): Es verhält sich alles wie bei Beispiel (1) für $\beta = \mathrm{e}^{\gamma}$ erläutert. Zudem kann man nun alle invarianten Mengen explizit angeben: Für $\gamma = 0$ sind alle Mengen invariant, für $\gamma \neq 0$ sind $\emptyset$, $\{0\}$, $\mathbb{R}^+$ und $\mathbb{R}_0^+$ die einzigen invarianten Mengen. Weitere nur positiv invariante Mengen sind im Fall $\gamma > 0$ die der Form (x, ∞), $[x, \infty)$, $\{0\} \cup (x, \infty)$, $\{0\} \cup [x, \infty)$ mit $x \in \mathbb{R}^+$ und im Fall $\gamma < 0$ die der Form $(0, x)$, $(0, x]$, $[0, x)$, $[0, x]$ mit $x \in \mathbb{R}^+$.

Feinere Beispiele, an denen auch die Unterschiede zwischen fast-periodischen, rekurrenten und nicht-wandernden Punkten ausgemacht werden können, sind Thema der Übungen.

2.3 Stabilitätsbegriffe

Ist bei einem dynamischen System $(\mathbb{T}, \mathcal{X}, \Phi)$ die Bahnlinie $\Phi(\,\cdot\,, x_0)$ eines Zustands $x_0 \in \mathcal{X}$ (explizit) bekannt, so stellt sich folgende **Grundfrage der Stabilitätstheorie**: Wenn ein anderer Zustand $x \in \mathcal{X}$ sich nur wenig von x_0 unterscheidet, weicht dann auch die (vielleicht nicht mehr explizit bekannte) Bahnlinie $\Phi(\,\cdot\,, x)$ von x nur wenig von $\Phi(\,\cdot\,, x_0)$ ab? Oder in etwas anderer Formulierung: Hängt die Bahnlinie $\Phi(\,\cdot\,, x)$ stetig von x ab?

Bei festem Zeithorizont lässt sich dies ohne Frage **bejahen**. Interessiert man sich nämlich für $\Phi(t, x)$ mit festem $t \in \mathbb{T}$ oder auch mit $t \in [-T, T] \cap \mathbb{T}$ bei Zeithorizont $T \in \mathbb{R}_0^+$, so ist durch die Stetigkeit von Φ und die daraus resultierende Stetigkeit der Flussabbildungen Φ_t sichergestellt, dass sich $\Phi(t, x)$ und $\Phi(t, x_0)$ (bei wenig von x_0 abweichendem x) nur wenig unterscheiden. In konkreten Fällen gilt es hierfür allerdings erst einmal nachzuweisen, dass ein dynamisches System im Sinne von Definition 1.1 vorliegt, und dieser Nachweis kann sich durchaus nicht-trivial gestalten; man vergleiche mit Abschnitt 6.4, in dem der Stetigkeitsbeweis für den kontinuierlichen Fall erbracht wird.

Die **Stabilitätstheorie im engeren Sinne** beschäftigt sich aber vor allem mit **Langzeit-Stabilität**, also damit, ob die Differenz zwischen $\Phi(t, x)$ und $\Phi(t, x_0)$ (bei wenig von x_0 abweichendem x) für alle Zeiten $t \geq 0$, auch für sehr große t, klein bleibt. Mit anderen Worten handelt es sich, um die Frage, ob sich aus der Langzeitentwicklung bei Initialzustand x_0 auf die Langzeitentwicklung bei einem anderen leicht gestörten Initialzustand x schließen lässt.

Mit den folgenden Begriffen lassen sich Antworten auf diese und verwandte Fragen zunächst für den Fall einer Ruhelage x_0 formulieren, der durch eine weit entwickelte Stabilitätstheorie abgedeckt wird. Zur Stabilität anderer Bahnlinien als nur Ruhelagen wird (viel) später in der Vorlesung noch etwas (aber nicht allzu viel) gesagt.

Konkret klassifiziert man Stabilitätseigenschaften von Ruhelagen und invarianten Mengen nun wie folgt.

Definition 2.14 (**Stabilitätsbegriffe für invariante Mengen und Ruhelagen**). *Gegeben sei ein dynamisches System $(\mathbb{T}, \mathcal{X}, \Phi)$. Dann heißt eine kompakte, unter Φ positiv invariante Teilmenge K von $\mathcal{X}$ ...*

- ***(Ljapunov-)stabil**, wenn es zu jeder Umgebung V von K in $\mathcal{X}$ eine weitere Umgebung U von K in $\mathcal{X}$ mit $\mathcal{O}^+(x) \subset V$ für alle $x \in U$ gibt,*
- ***instabil**, wenn sie nicht Ljapunov-stabil ist,*
- ***(lokal) attraktiv**, wenn sie nicht-leer ist und es eine Umgebung U von K in $\mathcal{X}$ mit*

$$\lim_{t \to \infty} \operatorname{dist}(\Phi(t, x), K) = 0 \tag{2.3}$$

für alle $x \in U$ gibt.

- ***global attraktiv**, wenn sie nicht-leer ist und (2.3) für alle $x \in \mathcal{X}$ gilt,*
- ***asymptotisch stabil**, wenn sie Ljapunov-stabil und lokal attraktiv ist.*

Ist $x_0 \in \mathcal{X}$ eine Ruhelage des Systems $(\mathbb{T}, \mathcal{X}, \Phi)$ und sind die obigen Bedingungen für die einelementige Menge $\{x_0\}$ erfüllt, so wendet man die entsprechenden Begriffe auch auf den Punkt x_0 an.

Bemerkungen.

(1) Statt mit beliebigen Umgebungen U und V kann man natürlich auch nur mit offenen Umgebungen oder nur mit solchen Umgebungen arbeiten, die an positive Größen ε und δ gekoppelt sind. Im letzteren Fall übernimmt die ε-Umgebung $\mathrm{U}_\varepsilon(K) := \{x \in \mathcal{X} : \operatorname{dist}(x, K) < \varepsilon\}$ die Rolle von V und die δ-Umgebung $\mathrm{U}_\delta(K)$ die von U, und speziell für $K = \{x_0\}$ sind $\mathrm{U}_\varepsilon(\{x_0\}) = \mathrm{B}_\varepsilon(x_0)$, $\mathrm{U}_\delta(\{x_0\}) = \mathrm{B}_\delta(x_0)$ nichts anderes als Kugeln.

(2) Während die Begriffe des vorausgehenden Abschnitts für Ruhelagen $x_0 \in \mathcal{X}$ trivial erfüllt (rekurrent, nicht-wandernd, fast-periodisch) oder nicht erfüllt (transitiv, wandernd) sind, machen die Stabilitätsbegriffe für einen einzelnen Punkt x_0 überhaupt erst unter der Voraussetzung Sinn, dass es sich um eine Ruhelage handelt. Ist die Voraussetzung gegeben, so erlauben die Stabilitätsbegriffe allerdings eine feinere Beschreibung der Dynamik nahe der Ruhelage.

(3) Mit etwas anderen Worten wird für die Umgebungen U und V in der Definition von **Ljapunov-Stabilität** die Implikation

$$x \in U \implies \Phi(t, x) \in V \text{ für } 0 \leq t \in \mathbb{T}$$

gefordert, und mit der Halbgruppeneigenschaft folgt dann auch

$$\Phi(s, x) \in U \text{ für ein } s \in \mathbb{T} \implies \Phi(t, x) \in V \text{ für } s \leq t \in \mathbb{T}\,.$$

Dies **bedeutet grob gesprochen, dass jede Bahnlinie, die einmal sehr nah an K herankommt, danach für immer nah an K verbleibt**.

(4) In ähnlicher Weise **bedeutet Attraktivität, dass jede Bahnlinie, die einmal sehr nah an K herankommt, letztlich gegen K konvergiert**. Die Konvergenz ist dabei im Sinne von (2.3) zu verstehen und kann so interpretiert werden, dass einerseits alle eigentlichen Häufungspunkte von $\Phi(t, x)$ bei $t \to \infty$ in K enthalten sind und es andererseits keine uneigentlichen[8] Häufungspunkte gibt. Präzise gesagt ist (2.3) (unter der Kompaktheitsvoraussetzung an K) tatsächlich äquivalent dazu, dass $\omega(x) \subset K$ gilt und $\mathcal{O}^+(x)$ relativ kompakt in $\mathcal{X}$ ist. Ein detaillierten Beweis dieser Äquivalenz ist Thema der Übungen; man vergleiche auch mit Satz 2.7.

(5) **Aus Attraktivität allein folgt** — anders als man auf Anhieb vielleicht denken könnte — im Allgemeinen **noch keine Stabilität**, denn eine Bahnlinie kann sich nach einer Annäherung an K auch wieder weit entfernen, bevor sie schließlich zurückkehrt. Ein Beispiel hierfür ist der Homöomorphismus φ der Einheitskreislinie S^1 auf sich mit $\varphi(\exp(\mathbf{i}s)) = \exp(\frac{3}{2}\mathbf{i}s)$ für

[8] In diesem Kontext soll die Nicht-Existenz uneigentlicher Häufungspunkte bedeuten, dass $\Phi(t_k, x)$ entlang *jeder* ∞-Folge $(t_k)_{k\in\mathbb{N}}$ einen eigentlichen Häufungspunkt in $\mathcal{X}$ besitzt.

$s \in [0, \pi]$ und $\varphi(\exp(\mathbf{i}s)) = \exp(\frac{1}{2}\mathbf{i}s)$ für $s \in [-\pi, 0]$ (Winkelvergrößerung auf dem oberen Halbkreis, Winkelhalbierung auf dem unteren). Bei dieser Wahl ist 1 ein global attraktiver, aber instabiler Fixpunkt des Zeit-diskreten Systems $(\mathbb{Z}, \mathrm{S}^1, \varphi)$. Ein entsprechendes Beispiel mit kontinuierlicher Zeit wird im späteren Abschnitt 8.1 (Fußnote 2) angegeben.

(6) Im speziellen Fall, dass $\mathcal{X}$ ein Intervall in $\mathbb{R}$ ist, folgt aus Attraktivität einer Ruhelage $x_0 \in \mathcal{X}$ aber doch ihre asymptotische Stabilität; auch der Beweis dieser Tatsache wird in den Übungen behandelt, er ist im Fall $\mathbb{T} = \mathbb{N}_0$ aber gar nicht einfach.

Beispiele (zur **Stabilität von Ruhelagen und invarianten Mengen**)**.** Das Verhalten in den Beispielen des Kapitels 1 lässt sich wie folgt zusammenfassen:

- In den Beispielen (1) und (4) sind für $\beta = 1$ beziehungsweise $\gamma = 0$ alle kompakten $K \subset \mathbb{R}_0^+$ Ljapunov-stabil, aber nicht attraktiv und somit nicht asymptotisch stabil; insbesondere sind auch alle Punkte in $\mathbb{R}_0^+$ Ruhelagen mit diesen Eigenschaften. Für $\beta \neq 1$ beziehungsweise $\gamma \neq 0$ ist dagegen 0 die einzige Ruhelage; diese Ruhelage und allgemeiner jedes Intervall $[0, x]$ mit $x \in \mathbb{R}_0^+$ sind für $\beta < 1$ beziehungsweise $\gamma < 0$ asymptotisch stabil und global attraktiv, für $\beta > 1$ beziehungsweise $\gamma > 0$ ist die Null-Ruhelage instabil und nicht attraktiv.

- Besteht der Zustandsraum wie in Beispiel (2) aus isolierten Punkten, so sind Stabilitätsfragen trivial, und rein formal sind alle kompakten, positiv invarianten Mengen asymptotisch stabil. Eine solche Aussage gilt allerdings nicht für den Begriff der globalen Attraktivität, der kein (lokaler) Stabilitätsbegriff (im engeren Sinne) ist und auch bei diskretem Zustandsraum einen nicht-triviales Konzept darstellt: Tatsächlich handelt es sich bei der Frage nach globaler Attraktivität der Menge $\{1, 4, 2\}$ nämlich wieder einmal um eine Umformulierung der Collatz-Vermutung.

- In Beispiel (3) sind alle periodischen Orbits Ljapunov-stabil, aber nicht asymptotisch stabil. Davon abgesehen lässt sich aber kaum Interessantes über Stabilität aussagen.

Kapitel 3

Lineare (und linearisierte) dynamische Systeme

Lineare Systeme sind eine grundlegende Klasse dynamischer Systeme, die durch eine weitgehend vollständige Theorie abgedeckt werden. In diesem Kapitel wird — nach der Einführung der entsprechenden Begriffe — der Zeit-diskrete Teil dieser Theorie behandelt und somit eine (fast) komplette Klassifikation Zeit-diskreter linearer Systeme gegeben. Für den kontinuierlichen Teil der linearen Theorie sei dagegen auf das spätere Vorlesungskapitel 7 verwiesen.

In den Kontext dieses Kapitels passen außerdem einige Aspekte Zeit-diskreter, nicht-linearer Dynamiken, die sich mit (mehr oder weniger) linearen Methoden behandeln lassen. Auf diese Aspekte wird daher gegen Endes des Kapitels kurz eingegangen.

3.1 Definitionen und Begriffe

Im Folgenden wird $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ als Platzhalter für einen der beiden Grundkörper $\mathbb{R}$, $\mathbb{C}$ verwendet.

Definition 3.1 (**lineare dynamische Systeme**).

(1) *Ein dynamisches System* $(\mathbb{T}, \mathcal{X}, \Phi)$ *mit einem normierten Raum* $\mathcal{X}$ *über* $\mathbb{K}$ *als Zustandsraum heißt* ***linear****, wenn jede Flussabbildung* Φ_t *mit* $t \in \mathbb{T}$ *eine affin lineare Abbildung von* $\mathcal{X}$ *nach* $\mathcal{X}$ *ist. Mit anderen Worten handelt es sich um ein System, dessen Fluss* Φ *die Form*

$$\Phi(t, x) = M(t)x + \Gamma(t) \qquad \text{für } (t, x) \in \mathbb{T} \times \mathcal{X}$$

mit $M \in \mathrm{C}^0(\mathbb{T}, \mathcal{L}(\mathcal{X}, \mathcal{X}))$ *und* $\Gamma \in \mathrm{C}^0(\mathbb{T}, \mathcal{X})$ *aufweist. Hierbei bezeichnet* $\mathcal{L}(\mathcal{X}, \mathcal{X})$ *den Raum der stetigen* $\mathbb{K}$*-linearen Abbildungen von* $\mathcal{X}$ *nach* $\mathcal{X}$ *(mit der Operatornorm) und* $M(t)x$ *steht für die Anwendung der linearen Abbildung* $M(t)$ *auf* x*.*

(2) *Sind die Flussabbildungen nicht nur affin linear, sondern sogar linear und ist dementsprechend* $\Gamma \equiv 0$ *und* $\Phi(t, x) = M(t)x$*, so nennt man ein lineares dynamisches System* $(\mathbb{T}, \mathcal{X}, \Phi)$ ***homogen****, andernfalls* ***inhomogen****.*

Bemerkungen.

(1) Den (für diese Vorlesung) **relevantesten Fall** bilden endlich-dimensionale Zustandsräume $\boldsymbol{\mathcal{X} = \mathbb{K}^N}$. In diesem Fall kann man $M \in \mathrm{C}^0(\mathbb{T}, \mathbb{K}^{N \times N})$ als (reelle oder komplexe) Matrix-wertige Abbildung beziehungsweise t-abhängige $(N \times N)$-Matrix und $\Gamma \in \mathrm{C}^0(\mathbb{T}, \mathbb{K}^N)$ als $\mathbb{K}^N$-wertige Abbildung beziehungsweise t-abhängigen Vektor verstehen.

(2) Die Betrachtung unendlich-dimensionaler Zustandsräume $\mathcal{X}$ besitzt zwar auch eine Berechtigung und ist für manche Anwendungen von Interesse, doch wird dieser Fall hier nur dann mitbehandelt, wenn er keinen nennenswerten Mehraufwand erfordert.

(3) Aus der Identitätseigenschaft von Φ folgt, dass $M(0)$ die Identität $\mathrm{id}_{\mathcal{X}}$ (beziehunsweise im Fall $\mathcal{X} = \mathbb{K}^N$ die $(N\times N)$-Einheitsmatrix $\mathbb{I}_N$) ist und $\Gamma(0) = 0$ gilt. Aus der Halbgruppeneigenschaft von Φ ergeben sich außerdem die analoge Eigenschaft von M

$$M(s+t) = M(s)M(t) \qquad \text{für } s,t \in \mathbb{T}$$

(wobei $M(s)$ und $M(t)$ mit der Komposition linearer Abbildungen oder als Matrizen multipliziert werden) sowie die Bedingung $\Gamma(s+t) = M(s)\Gamma(t) + \Gamma(s)$ für $s,t \in \mathbb{T}$. Insbesondere ist M damit stetiger Halbgruppenhomomorphismus von $\mathbb{T}$ nach $\mathcal{L}(\mathcal{X},\mathcal{X})$ (beziehungsweise nach $\mathbb{K}^{N\times N}$).

3.2 Die Klassifikation Zeit-diskreter linearer Systeme

Wie in Kapitel 1 erläutert, kommt jedes Zeit-diskrete dynamische System durch Iteration seiner Zeit-1-Abbildung zustande, und bei einem linearen System ist die Zeit-1-Abbildung affin linear. Es handelt sich bei einem Zeit-diskreten *und* linearen System mit Zustandsraum $\mathbb{K}^N$ also um ein System des Typs $(\mathbb{T},\mathbb{K}^N,\varphi)$ zu einer affin linearen Abbildung φ der Form

$$\varphi(x) = Mx+\gamma \qquad \text{für } x \in \mathbb{K}^N \tag{3.1}$$

mit einer festen Matrix $M \in \mathbb{K}^{N\times N}$ (die im Fall $\mathbb{T} = \mathbb{Z}$ invertierbar sein muss) und einem festen Vektor $\gamma \in \mathbb{K}^N$. Die in Definition 3.1 auftretenden Größen ergeben sich nun für $k \in \mathbb{T}$ als

$$M(k) = M^k\,, \qquad \Gamma(k) = \begin{cases} \sum_{i=0}^{k-1} M^i\gamma & \text{falls } k \geq 0 \\ -\sum_{i=-k}^{-1} M^{-i}\gamma & \text{falls } k \leq 0 \end{cases}$$

(wobei die Bezeichnung M jetzt sowohl für eine Funktion in $\mathrm{C}^0(\mathbb{T},\mathbb{K}^{N\times N})$ als auch für deren Wert an der Stelle 1 steht; diese leichte Doppeldeutigkeit wird hier als vertretbar betrachtet).

Der **Rest dieses Abschnitts beschränkt sich** der Einfachheit halber auf die **Untersuchung *homogener* linearer Systeme**, das heißt auf den Fall $\gamma = 0$ beziehungsweise $\Gamma \equiv 0$ im Vorausgehenden. Diese Annahme ist weniger einschränkend, als es scheinen mag, denn man kann **oft mit einem einfachen Argument auf den Fall $\boldsymbol{\gamma = 0}$ reduzieren**: Ist φ von der Form (3.1), und ist 1 kein Eigenwert von M, so ist die Matrix $(\mathbb{I}_N - M)$ invertierbar, und $x_0 := (\mathbb{I}_N - M)^{-1}\gamma$ mit $Mx_0+\gamma = x_0$ ist eindeutiger Fixpunkt des Systems $(\mathbb{T},\mathbb{K}^N,\varphi)$. Man kann diesen Fixpunkt dann in den Nullpunkt überführen, indem man von Φ zum neuen Fluss $(k,x) \mapsto \Phi(k,x_0+x) - x_0$ und dementsprechend von der affin linearen Zeit-1-Abbildung φ zur einfacheren linearen Abbildung $x \mapsto \varphi(x_0+x) - x_0 = Mx$ übergeht. Da sich praktisch alle qualitativen Eigenschaften problemlos übertragen, reicht es, das neue, homogene lineare System anstelle von $(\mathbb{T},\mathbb{K}^N,\varphi)$ zu untersuchen. Ist 1 ein Eigenwert von M, so bilden die Fixpunkte von $(\mathbb{T},\mathbb{K}^N,\varphi)$ entweder einen affinen Unterraum der Dimension ≥ 1 in $\mathbb{K}^N$, oder es existiert überhaupt kein Fixpunkt. Im ersten dieser beiden Fälle ist die beschriebene Reduktion nach wie vor möglich und zeigt auch, dass bei allen Fixpunkten dasselbe Verhalten vorliegt. Im zweiten, dem Fixpunkt-losen Fall ist eine solche Reduktion tatsächlich nicht möglich, und in diesem Fall

kann die Iteration von $x \mapsto Mx+\gamma$ ein etwas anderes Verhalten (im Wesentlichen mit zusätzlichen unbeschränkten Bahnlinien) ergeben als die Iteration von $x \mapsto Mx$. Der Ausschluss des Fixpunkt-losen Falls im Folgenden ist aber zumindest[1] im Hinblick auf Stabilitätsuntersuchungen bei Fixpunkten offensichtlich nicht einschränkend.

Wie angekündigt werden jetzt Iterationen von Selbstabbildungen $x \mapsto Mx$ mit $M \in \mathbb{K}^{N\times N}$ betrachtet, und für das zugehörige dynamische System mit diskreter Zeit $\mathbb{T}$ und Fluss $(k,x) \mapsto M^k x$ wird $(\mathbb{T}, \mathcal{X}, M)$ notiert. Als erste simple Beobachtung über dieses System halten wir fest:

Lemma. *Ist $M \in \mathbb{K}^{N\times N}$, so sind jeder **Eigenraum** und jeder **Hauptraum** von M **invariante Mengen** des homogenen linearen Systems $(\mathbb{T}, \mathbb{K}^N, M)$.*

Beweis. Die Behauptung ergibt sich direkt aus den Gleichungen $Mx = \lambda x$ beziehungsweise $(M-\lambda\mathbb{I}_N)^\ell x = 0$ für ein $\ell \in \mathbb{N}$, die Eigenraum- beziehungsweise Hauptraum-Vektoren $x \in \mathbb{K}^N$ zum Eigenwert $\lambda \in \mathbb{K}$ charakterisieren. Da M mit sich selbst und $\mathbb{I}_N$ kommutiert, folgt sofort, dass $M^k x$ denselben Gleichungen genügt, daher ist $\mathcal{O}(x) = \{M^k x : k \in \mathbb{T}\}$ stets im gleichen Eigen- beziehungsweise Hauptraum enthalten wie x. □

Der folgende Satz geht deutlich weiter und beschreibt für homogene lineare Systeme das Langzeitverhalten bei beliebigem Inititalzustand in $\mathbb{K}^N$ (abgesehen vom trivialen Fall der Ruhelage 0):

Satz 3.2 (über das **Langzeitverhalten bei Zeit-diskreten linearen Systemen**)**.** *Seien eine Matrix $M \in \mathbb{K}^{N\times N}$ und mit ihr das Zeit-diskrete dynamische System $(\mathbb{T}, \mathbb{K}^N, M)$ (mit $\mathbb{T} = \mathbb{N}_0$ oder bei invertierbarem M wahlweise auch mit $\mathbb{T} = \mathbb{Z}$) gegeben. Ein Punkt $x \in \mathbb{K}^N \setminus \{0\}$ besitzt eine eindeutige Zerlegung*

$$x = x_1+x_2+\ldots+x_p\,, \qquad x_i \in \mathrm{H}_{\lambda_i}(M) \setminus \{0\} \tag{3.2}$$

mit $p \in \mathbb{N}$ verschiedenen Eigenwerten $\lambda_1, \lambda_2, \ldots, \lambda_p \in \mathbb{C}$ von M über $\mathbb{C}$ (wobei je nach x aber nicht unbedingt alle *Eigenwerte von M vorkommen) und den zugehörigen Haupträumen $\mathrm{H}_{\lambda_1}(M), \mathrm{H}_{\lambda_2}(M), \ldots, \mathrm{H}_{\lambda_p}(M) \subset \mathbb{C}^N$. Es tritt dann eine der folgenden drei Alternativen ein:*

(A) $\max\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_p|\} < 1$.

In diesem Fall ist x wandernd mit $\lim_{k\to\infty} M^k x = 0$, und dementsprechend gilt $\omega(x) = \{0\}$.

(B) $\max\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_p|\} = 1$ *und für alle i mit $|\lambda_i| = 1$ ist $x_i \in \mathrm{E}_{\lambda_i}(M) \setminus \{0\}$ Eigenvektor zu λ_i.*

In diesem Fall bleibt $|M^k x|$ bei $k \to \infty$ von 0 und ∞ weg beschränkt. Falls alle λ_i Einheitswurzeln sind, ist x periodischer Punkt mit dem kleinsten gemeinsames Vielfachen der Ordnungen der Einheitswurzeln als Periode; falls alle λ_i Betrag 1 haben, ist x zumindest fastperiodischer Punkt; andernfalls ist x wandernd. Falls (nur) all diejenigen λ_i mit $|\lambda_i| = 1$ Einheitswurzeln sind, ist $\omega(x)$ endlich; andernfalls ist $\omega(x)$ kompakt und homöomorph zu einer Teilmenge des Kartesischen Produkts $\{1, 2, \ldots, \ell\} \times (\mathrm{S}^1)^q$, wobei ℓ das kleinste gemeinsame Vielfache der Ordnungen der auftretenden Einheitswurzeln λ_i ist und $q \in \{1, 2, \ldots, p\}$ die Anzahl der Nicht-Einheitswurzeln λ_i mit $|\lambda_i| = 1$.

[1] Man kann sich überlegen, dass der Ausschluss des Fixpunkt-losen Falls auch bei Stabilitätsuntersuchungen für kompakte, positiv invariante Mengen K nicht einschränkend ist. Ist nämlich $K \neq \emptyset$ solch eine Menge, so ist die kleinste konvexe Obermenge von K, die sogenannte konvexe Hülle $\mathrm{C}(K)$, ebenfalls nicht-leer, kompakt und positiv invariant. Gemäß einem Fixpunktsatz von Schauder besitzt dann φ als stetige Selbstabbildung der nicht-leeren, kompakten und konvexen Menge $\mathrm{C}(K)$ in sich einen Fixpunkt.

(C) $\max\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_p|\} > 1$ *oder es gibt ein* i *mit* $|\lambda_i| = 1$*, so dass* $x_i \in \mathrm{H}_{\lambda_i}(M) \setminus \mathrm{E}_{\lambda_i}(M)$ *kein Eigenvektor zu* λ_i *ist.*

In diesem Fall ist x *wandernd mit* $\lim_{k\to\infty} |M^k x| = \infty$*, und dementsprechend gilt* $\omega(x) = \emptyset$*.*

Insbesondere ist x *in keinem der Fälle je transitiv.*

Beweis. Die Existenz der Zerlegung (3.2) folgt aus linearer Algebra, genauer daraus, dass $\mathbb{C}^N$ bekanntlich als direkte Summe der Haupträume von M geschrieben werden kann.

Der Übersichtlichkeit halber sei für die folgende Rechnung und Argumentation vorerst angenommen, dass $x \in \mathrm{H}_\lambda(M) \setminus \{0\}$ für einen Eigenwert $\lambda \in \mathbb{C}$ von M gilt und x selbst somit der einzige Summand auf der rechten Seite von (3.2) ist. Außerdem sei $M_\lambda := M - \lambda \mathbb{I}_N$ abgekürzt. Wie in der linearen Algebra (typischerweise im Kontext der Jordan-Zerlegung) gezeigt wird, gibt es dann ein minimales $m \in \mathbb{N}$, so dass x, $M_\lambda x$, $M_\lambda^2 x$, …, $M_\lambda^{m-1} x$ über $\mathbb{C}$ linear unabhängig sind und $M_\lambda^m x = 0$ gilt. Für Punkte im Orbit von x erhält man durch Ausmultiplizieren die Formel

$$\begin{aligned} M^k x &= (\lambda \mathbb{I}_N + M_\lambda)^k x \\ &= \lambda^k x + k\lambda^{k-1} M_\lambda x + \tbinom{k}{2} \lambda^{k-2} M_\lambda^2 x + \ldots + \tbinom{k}{m-1} \lambda^{k-m+1} M_\lambda^{m-1} x \\ &= \lambda^{k-m} \Big[\lambda^m x + k\lambda^{m-1} M_\lambda x + \tbinom{k}{2} \lambda^{m-2} M_\lambda^2 x + \ldots + \tbinom{k}{m-1} \lambda M_\lambda^{m-1} x \Big] \end{aligned} \tag{3.3}$$

für alle $k \geq m$. Hierbei ist zu beachten, dass der Term in den eckigen Klammern ein Polynom in k ist, im nicht-trivialen Fall $\lambda \neq 0$ von 0 weg beschränkt bleibt (denn es handelt sich um eine Linearkombination linear unabhängiger Vektoren mit k-unabhängigem Leitkoeffizienten λ^m) und sich genau im Fall eines Eigenvektors $x \in \mathrm{E}_\lambda(M)$ mit $m = 1$ auf einen konstanten Summanden reduziert. Insgesamt entnimmt man daher aus der Formel, dass $|M^k x|$ im Fall $|\lambda| < 1$ gegen 0 konvergiert, im Fall $|\lambda| = 1$, $x \in \mathrm{E}_\lambda(M)$ von 0 und ∞ weg beschränkt bleibt, und in den restlichen Fällen gegen ∞ konvergiert. Klar ist auch, dass im Fall $x \in \mathrm{E}_\lambda(M)$ einfach $M^k x = \lambda^k x$ gilt und damit für jede Einheitswurzel λ ein periodischer Punkt x mit endlichem Orbit $\mathcal{O}(x)$ und endlicher Limesmenge $\omega(x) = \mathcal{O}(x)$ vorliegt. Ist $x \in \mathrm{E}_\lambda(M) \setminus \{0\}$ für ein λ, das Betrag 1 hat, aber keine Einheitswurzel ist, so lässt sich Fast-Periodizität von x durch eine Anwendung von Satz 2.9 auf das einfachere System $(\mathbb{Z}, \mathrm{S}^1, z \mapsto \lambda{\cdot}z)$ vom Typ des Beispiels (3) aus Kapitel 1 nachweisen. Der Satz garantiert zunächst nur, dass *irgendein* Punkt in S^1 fast-periodisch ist, doch wegen Rotationsinvarianz folgt problemlos, dass auch 1 (und jeder andere Punkt von S^1) fast-periodisch für $(\mathbb{Z}, \mathrm{S}^1, z \mapsto \lambda{\cdot}z)$ ist. Dies wiederum impliziert die behauptete Fast-Periodizität von $x \in \mathrm{E}_\lambda(M) \setminus \{0\}$ für $(\mathbb{T}, \mathbb{K}^N, M)$. Mit elementaren Argumenten lässt sich tatsächlich auch nachweisen, dass im zuletzt betrachteten Fall jeder Punkt von S^1 transitiv für $(\mathbb{Z}, \mathrm{S}^1, z \mapsto \lambda{\cdot}z)$ ist. Hieraus folgt, dass jeder Punkt zx mit $z \in \mathrm{S}^1 \subset \mathbb{C}$ Häufungspunkt von $M^k x$ bei $k \to \infty$ und deshalb $\omega(x) = \{zx : z \in \mathrm{S}^1\}$ homöomorph zu S^1 ist. Im Fall $x \in \mathrm{H}_\lambda(M) \setminus \{0\}$ sind damit alle Behauptungen des Satzes nachgewiesen.

Es verbleibt, den Fall von $p \geq 2$ Summanden x_i in der Zerlegung (3.2) zu behandeln. In dieser Situation gilt eine zu (3.3) analoge Formel für jedes x_i (mit dem zugehörigen Eigenwert λ_i und einem ebenfalls i-abhängigen $m_i \in \mathbb{N}$), und die gerade nachgewiesenen Eigenschaften gelten entsprechend für x_i und $M^k x_i$. Da gemäß dem vorausgehenden Lemma $M^k x_i \in \mathrm{H}_{\lambda_i}(M)$ im Hauptraum zu λ_i verbleibt, lässt sich nun auf das behauptete Verhalten von $M^k x = M^k x_1 + M^k x_2 + \ldots + M^k x_p$ schließen. Dieser Schluss wird hier nicht im Detail ausgeführt, sondern es werden nur die wesentlichen dafür relevanten Prinzipien erwähnt: So verwendet man naheliegenderweise, dass das Wachstum von $|M^k x|$ durch das Wachstums der $|M^k x_i|$ zu den auftretenden Eigenwerten λ_i des größten Betrags bestimmt ist. Auch benutzt

man die Beobachtung, dass x wegen der Invarianz der Haupträume wandernd ist, sobald nur ein x_i in der Zerlegung (3.2) wandernd ist (was wiederum genau dann eintritt, wenn $|\lambda_i| \neq 1$ oder $x \in \mathrm{H}_{\lambda_i}(M) \setminus \mathrm{E}_{\lambda_i}(M)$ gilt). Der Nachweis der Fast-Periodizität von x im Fall, dass alle λ_i Betrag 1 haben, gelingt wie zuvor mit Hilfe von Satz 2.9. Um schließlich den Typ der Limesmenge $\omega(x)$ im Fall (B) zu bestimmen, unterteilt man die p Vektoren $x_1, x_2, \ldots, x_p$ (in der Zerlegung des fixierten Punktes x) in drei Klassen. Eine erste Klasse enthält alle $x_i \in \mathrm{H}_{\lambda_i}(M) \setminus \{0\}$ zu Eigenwerten λ_i mit $|\lambda_i| < 1$; solche x_i spielen im Folgenden wegen $\omega(x_i) = \{0\}$ keine Rolle. Eine zweite Klasse $\mathcal{A}_1$ enthält alle $x_i \in \mathrm{E}_{\lambda_i}(M) \setminus \{0\}$ zu Einheitswurzeln λ_i. Alle $x_i \in \mathcal{A}_1$ und somit auch $\sum_{y\in\mathcal{A}_1} y$ sind periodische Punkte; die Periode ℓ von $\sum_{y\in\mathcal{A}_1} y$ und damit die Anzahl der Elemente in $\omega\big(\sum_{y\in\mathcal{A}_1} y\big)$ ergibt sich als kleinstes gemeinsames Vielfaches der Ordnungen der zugehörigen Einheitswurzeln. Die letzte Klasse $\mathcal{A}_\sharp$ enthält solche $x_i \in \mathrm{E}_{\lambda_i}(M) \setminus \{0\}$, für die das zugehörige λ_i Betrag 1 hat, doch keine Einheitswurzel ist. Für solche x_i ist $\omega(x_i)$ homöomorph zu S^1, die Bestimmung des exakten Homöomorphie-Typs von $\omega\big(\sum_{y\in\mathcal{A}_\sharp} y\big)$ ist aber schwierig, und deshalb sei nur festgehalten, dass $\omega\big(\sum_{y\in\mathcal{A}_\sharp} y\big)$ homöomorph zu *einer Teilmenge von* $(\mathrm{S}^1)^q$ mit der Anzahl q der Elemente von $\mathcal{A}_\sharp$ ist. Insgesamt ist dann $\omega(x)$ homöomorph zu einer Teilmenge des Kompaktums $\{1, 2, \ldots, \ell\} \times (\mathrm{S}^1)^q$, und damit ist die abgeschlossene Menge $\omega(x)$ auch selbst kompakt. □

Mit Hilfe des vorigen Satzes lassen sich nun folgende (notwendigen und hinreichenden) Kriterien für die Stabilität von Fixpunkten Zeit-diskreter homogen linearer Systeme aufstellen:

Korollar 3.3 (über **Stabilität von Fixpunkten Zeit-diskreter linearer Systeme**). *Durch $M \in \mathbb{K}^{N\times N}$ sei das Zeit-diskrete dynamische System $(\mathbb{T}, \mathbb{K}^N, M)$ (mit $\mathbb{T} = \mathbb{N}_0$ oder bei invertierbarem M wahlweise auch mit $\mathbb{T} = \mathbb{Z}$) gegeben, und jetzt seien $\lambda_1, \lambda_2, \ldots, \lambda_p \in \mathbb{C}$ mit* $1 \leq p \leq N$ alle *verschiedenen Eigenwerte von M über $\mathbb{C}$ mit zugehörigen Eigen- und Haupträumen $\mathrm{E}_{\lambda_i}(M) \subset \mathrm{H}_{\lambda_i}(M) \subset \mathbb{C}^N$. Dann tritt genau einer der folgenden drei Fälle ein:*

(A) $\boldsymbol{\max\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_p|\} < 1}$.

In diesem Fall ist der einzige Fixpunkt 0 ***asymptotisch stabil.***

(B) $\boldsymbol{\max\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_p|\} = 1}$ ***und für alle*** $\boldsymbol{i}$ ***mit*** $\boldsymbol{|\lambda_i| = 1}$ ***ist*** $\boldsymbol{\mathrm{H}_{\lambda_i}(M) = \mathrm{E}_{\lambda_i}(M)}$.

In diesem Fall ist der Fixpunkt 0 *(wie auch jeder andere Fixpunkt)* ***Ljapunov-stabil, aber nicht asymptotisch stabil.***

(C) $\boldsymbol{\max\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_p|\} > 1}$ ***oder es gibt ein*** $\boldsymbol{i}$ ***mit*** $\boldsymbol{|\lambda_i| = 1}$, $\boldsymbol{\mathrm{H}_{\lambda_i}(M) \neq \mathrm{E}_{\lambda_i}(M)}$.

In diesem Fall ist der Fixpunkt 0 *(wie auch jeder andere Fixpunkt)* ***instabil.***

Bemerkungen.

(1) Neben 0 treten genau dann weitere Fixpunkte auf, wenn 1 Eigenwert von M ist. Es handelt sich bei diesen Fixpunkten dann genau um die Eigenvektoren von M zum Eigenwert 1.

(2) Die Stabilitätskriterien des Korollars sind auch **bei praktischen Rechnungen gut zu handhaben** und erlauben tatsächlich eine **völlig schematische Durchführung von Stabilitätsuntersuchungen.** Ob $\mathrm{H}_{\lambda_i}(M) = E_{\lambda_i}(M)$ gilt, kann man dabei **ohne explizite Berechnung des Hauptraums** $\mathrm{H}_{\lambda_i}(M)$ entscheiden; dazu muss man gemäß linearer Algebra nämlich nur prüfen, ob die algebraische Vielfachheit des Eigenwerts λ_i (das ist seine Vielfachheit als Nullstelle des charakteristischen Polynoms von M) mit seiner geometrischen Vielfachheit (das ist die Dimension des Eigenraums $\mathrm{E}_{\lambda_i}(M)$) übereinstimmt.

Beweis. Im Fall (A) garantiert Satz 3.2 (A), dass $\lim_{k\to\infty} M^k x = 0$ für alle $x \in \mathbb{K}^N$ gilt und der Fixpunkt 0 somit global attraktiv ist. Wendet man die genannte Konvergenz zunächst nur für (kanonische) Basisvektoren x an und geht dann zu allgemeinen $x \in \mathrm{B}_1(0)$ über, so folgt, dass es eine nicht von x abhängige Konstante $C \in [1,\infty)$ mit $\sup_{k\in\mathbb{N}_0} |M^k x| \le C$ für alle $x \in \mathrm{B}_1(0)$ gibt. Für beliebiges $\varepsilon > 0$ ergibt die Wahl $\delta := \varepsilon/C$ nun, dass für jedes $x \in \mathrm{B}_\delta(0)$ der positive Halb-Orbit $\mathcal{O}^+(x)$ in $\mathrm{B}_\varepsilon(0)$ enthalten ist. Damit ist 0 stabil, also auch asymptotisch stabil.

Im Fall (B) greift für jedes $x \in \mathbb{K}^N$ entweder Teil (A) oder Teil (B) von Satz 3.2. Durch Anwendung der resultierenden Aussagen auf Basis-Vektoren von $\mathbb{K}^N$ findet man ein $C \in [1,\infty)$ mit der Eigenschaft des vorigen Beweisteils und erhält erneut die Ljapunov-Stabilität von 0.

Im Fall (C) gibt es entweder ein $x \in \mathrm{H}_{\lambda_i}(M) \setminus \{0\}$, $|\lambda_i| > 1$ oder ein $x \in \mathrm{H}_{\lambda_i}(M) \setminus \mathrm{E}_{\lambda_i}(M)$, $|\lambda_i| = 1$. Im Fall $\mathbb{K} = \mathbb{C}$ lässt sich Satz 3.2 (C) direkt auf Vielfache ax dieses x mit beliebig kleinem Vorfaktor $a \in \mathbb{C} \setminus \{0\}$ anwenden und liefert mit $\lim_{k\to\infty} |M^k(ax)| = \infty$ die Instabilität von 0. Im Fall $\mathbb{K} = \mathbb{R}$ muss nicht unbedingt $x \in \mathbb{R}^N$ gelten, aber es ist $\overline{x} \in \mathrm{H}_{\overline{\lambda_i}}(M)$, und Anwendung von Satz 3.2 (C) auf $a(x+\overline{x}) \in \mathbb{R}^N$ mit beliebig kleinem Faktor $a \in \mathbb{R} \setminus \{0\}$ liefert auch in diesem Fall die Behauptung.

Dass jeder andere Fixpunkt in den Fällen (B) und (C) dasselbe Verhalten wie 0 aufweist, folgt — wie schon diskutiert — einfach durch Verschieben solcher Fixpunkte in den Nullpunkt. □

3.3 Stabilität bei Zeit-diskreten nicht-linearen Systemen, Linearisierungskriterien

In diesem Abschnitt werden Stabilitätsuntersuchungen für Fixpunkte Zeit-diskreter Systeme auch auf den nicht-linearen (oder eigentlich nur nicht-unbedingt-linearen) Fall ausgedehnt. Eine erste Zusammenstellung hinreichender Kriterien für Stabilität und Instabilität folgt.

Satz 3.4 (Einfache Stabilitätskriterien für Fixpunkte nicht-linearer Systeme). *Seien D eine Teilmenge von $\mathbb{K}^N$ und $\varphi \in \mathrm{C}^0(D,D)$ eine stetige Abbildung, so dass das Zeit-diskrete System $(\mathbb{T}, D, \varphi)$ einen Fixpunkt $x_0 \in D$ besitzt. Dann gelten folgende Kriterien für die Stabilität von x_0:*

(A) *Gibt es ein $\delta > 0$ mit $\mathrm{B}_\delta(x_0) \subset D$ und $|\boldsymbol{\varphi(x) - x_0}| < |\boldsymbol{x - x_0}|$ für alle $x \in \mathrm{B}_\delta(x_0) \setminus \{x_0\}$, so ist x_0 **asymptotisch stabil**.*

(B) *Gibt es ein $\delta > 0$ mit $|\boldsymbol{\varphi(x) - x_0}| \le |\boldsymbol{x - x_0}|$ für alle $x \in \mathrm{B}_\delta(x_0) \cap D$, so ist x_0 **Ljapunov-stabil**.*

(C) *Gibt es ein $\delta > 0$ mit $\mathrm{B}_\delta(x_0) \subset D$ und $|\boldsymbol{\varphi(x) - x_0}| > |\boldsymbol{x - x_0}|$ für alle $x \in \mathrm{B}_\delta(x_0) \setminus \{x_0\}$, so ist x_0 **instabil**.*

Speziell im Fall $\mathbb{K} = \mathbb{R}$, $N = 1$, also für $\mathrm{D} \subset \mathbb{R}$ gilt zudem:

(D) *Gibt es ein $\delta > 0$ entweder mit $(x_0, x_0+\delta) \subset D$ und $\boldsymbol{\varphi(x) > x}$ für alle $x \in (x_0, x_0+\delta)$ oder mit $(x_0-\delta, x_0) \subset D$ und $\boldsymbol{\varphi(x) < x}$ für alle $x \in (x_0-\delta, x_0)$, so ist x_0 ebenfalls **instabil**.*

Bemerkungen.

(1) Aus dem Satz ergeben sich zumindest[2] im Fall $\mathbb{K} = \mathbb{R}$ **Ableitungskriterien für Stabilität**: Ist x_0 innerer Punkt von D und φ an der Stelle x_0 total differenzierbar, so ist für die Anwendung von (A) beziehungsweise (C) hinreichend, dass $\|\mathrm{D}\varphi(x_0)\|_{\mathbb{R}^{N\times N}} < 1$ (mit der Operatornorm auf $\mathbb{R}^{N\times N}$) gilt beziehungsweise $\mathrm{D}\varphi(x_0)$ invertierbar ist und $\left\|\mathrm{D}\varphi(x_0)^{-1}\right\|_{\mathbb{R}^{N\times N}} < 1$

[2] Im Fall $\mathbb{K} = \mathbb{C}$ gilt im Prinzip analoges, sobald man einen sinnvollen Ableitungsbegriff für $\mathbb{C}^N$-wertige

gilt. Zur Anwendung von (D) im Fall $N = 1$ reicht es, wenn φ bei x_0 nur einseitig differenzierbar ist und $\varphi'(x_0) > 1$ für die einseitig gebildete Ableitung $\varphi'(x_0)$ gilt. Insbesondere das Ableitungskriterium für Instabilität wird im weiteren Verlauf dieses Kapitels aber noch deutlich verbessert.

Für die Voraussetzung von (B) ist für $\mathbb{K} = \mathbb{R}$ und bei an der Stelle x_0 differenzierbarem φ notwendig, dass $\|\mathrm{D}\varphi(x_0)\|_{\mathbb{R}^{N\times N}} \leq 1$ gilt. Diese Bedingung ist aber nicht hinreichend und erlaubt noch keine Anwendung von (B).

(2) Statt $D \subset \mathbb{K}^N$ kann man bei obigen Kriterien auch allgemeinere metrische Räume als Zustandsräume zulassen. Tatsächlich zeigt der folgende Beweis, dass mit der Metrik formulierte Version von (A) und (C) in jedem lokal-kompakten metrischen Raum D (im Fall von (C) bei nicht-isoliertem x_0) gelten. Das Kriterium (B) und einfache Versionen von (A) und (C), bei denen stärkere quantitative Voraussetzungen wie $\mathrm{d}_D(\varphi(x), x_0) \leq \kappa \mathrm{d}(x, x_0)$, $\kappa \in [0, 1)$ beziehungsweise $\mathrm{d}_D(\varphi(x), x_0) \geq \kappa \mathrm{d}(x, x_0)$, $\kappa \in (1, \infty)$ gefordert werden, gelten sogar in einem beliebigen metrischen Raum D.

(3) Im Lichte der vorigen Bemerkung wird klar, dass Satz 3.4 nicht nur mit der Euklidischen Norm auf $\mathbb{K}^N$ (für die die oben formulierte Version des Satzes gemeint ist), sondern auch mit anderen Normen auf $\mathbb{K}^N$ gilt. Insbesondere kann (abgesehen einzig vom Fall $N = 1$) **die Gültigkeit der Voraussetzungen des Satzes von der Wahl der Norm auf $\mathbb{K}^N$ abhängen** und die Wahl einer guten Norm ein nicht-triviales Unterfangen sein. Dies gilt ebenso für die in Bemerkung (1) erwähnten Ableitungskriterien, da die Operatornorm auf $\mathbb{R}^{N\times N}$ natürlich von der auf $\mathbb{R}^N$ verwendeten Norm abhängt. Die folgenden und schon erwähnten Ableitungskriterien werden dieses Defizit (in den meisten Fällen) beseitigen.

Beweis von Satz 3.4. Im Fall (B) ist klar, dass $|\varphi^k(x) - x_0|$ für alle $x \in \mathrm{B}_\delta(x_0)$ nicht-wachsend von $k \in \mathbb{T}$ abhängt und somit $\mathcal{O}^+(x)$ für alle $x \in \mathrm{B}_\delta(x_0) \cap D$ in $\mathrm{B}_\delta(x_0)$ verbleibt. Dies zeigt Ljapunov-Stabilität von x_0.

Im Fall (A) bleiben für jedes $x \in \mathrm{B}_\delta(x_0)$ die Iterierten $\varphi^k(x)$ in einer kompakten Teilmenge von $\mathrm{B}_\delta(x_0)$, und deshalb existiert mindestens ein Häufungspunkt von $\varphi^k(x)$ bei $k \to \infty$. Da Stabilität gemäß (B) vorliegt, bleibt nur die Möglichkeit eines von x_0 verschiedenen Häufungspunktes $y \in \mathrm{B}_\delta(x_0)$ auszuschließen. Träte ein solcher Häufungspunkt $y = \lim_{\ell\to\infty} \varphi^{k_\ell}(x)$ entlang einer (strikt monotonen) Teilfolge $(k_\ell)_{\ell\in\mathbb{N}}$ auf, so ergäbe sich durch $(k_{\ell+1} - (k_\ell + 1))$-malige Anwendung der Voraussetzung in (A) die Ungleichung

$$|y - x_0| = \lim_{\ell\to\infty} \left|\varphi^{k_{\ell+1}}(x) - x_0\right| \leq \lim_{\ell\to\infty} \left|\varphi^{k_\ell+1}(x) - x_0\right| = |\varphi(y) - x_0| \,.$$

Das Ergebnis dieser Argumentation stellt aber auch einen Widerspruch zur gemachten Voraussetzung her, also ist $x_0 = \lim_{k\to\infty} \varphi^k(x)$ der einzige Häufungspunkt, und x_0 ist asymptotisch stabil.

Zum Beweis von (C) führt man für beliebiges $x \in \mathrm{B}_\delta(x_0) \setminus \{x_0\}$ zunächst die Möglichkeit, dass in $\mathrm{B}_\delta(x_0)$ ein Häufungspunkt von $\varphi^k(x)$ bei $k \to \infty$ existiert, zum Widerspruch. Dass x_0 selbst solch ein Häufungspunkt sein könnte, ist ausgeschlossen, da $|\varphi^k(x) - x_0|$ wachsend von

Funktionen auf einer Teilmenge von $\mathbb{C}^N$ zugrunde legt. Am einfachsten ist es, dazu $\mathbb{C}^N$ mit $\mathbb{R}^{2N}$ zu identifizieren und den reellen Ableitungsbegriff zu verwenden, bei dem $\mathrm{D}\varphi(x_0)$ als Matrix in $\mathbb{R}^{2N\times 2N}$ oder auch als $\mathbb{R}$-lineare Abbildung von $\mathbb{C}^N$ in sich zu verstehen ist. Als Matrix in $\mathbb{C}^{N\times N}$ kann man die Ableitung $\mathrm{D}\varphi(x_0)$ übrigens nur dann betrachten, wenn man statt einer $\mathbb{R}$-linearen sogar eine $\mathbb{C}$-lineare Abbildung erhält, was dem viel einschränkenderen, in der Funktionentheorie untersuchten komplexen Ableitungsbegriff entspricht.

$k \in \mathbb{T}$ abhängt. Für Häufungspunkte in $\mathrm{B}_\delta(x_0) \setminus \{x_0\}$ dagegen erreicht man völlig analog zum vorigen Beweisteil einen Widerspruch. Wegen Kompaktheit von $\overline{\mathrm{B}_{\delta/2}(x_0)}$ ist damit gezeigt, dass es zu (beliebig nah an x_0 gelegenen) $x \in \mathrm{B}_\delta(x_0) \setminus \{x_0\}$ stets ein $k \in \mathbb{N}$ mit $|\varphi^k(x) - x_0| > \delta/2$ gibt und $\mathcal{O}^+(x) \not\subset \mathrm{B}_{\delta/2}(x_0)$ gilt. Dies genügt zum Nachweis der Instabilität von x_0.

Der Beweis von (D) verläuft im Prinzip analog zu dem von (C), wobei man in einem Fall nur mit solchen x arbeitet, die größer als x_0 sind, im anderen Fall nur mit solchen, die kleiner als x_0 sind. □

Als Nächstes folgen die angekündigten und teils verbesserten Stabilitätskriterien; diese ähneln dem linearen Fall des Korollars 3.3 und erlauben nun aber **auch im nicht-linearen Fall eine schematische Stabilitätsuntersuchung** durch die Berechnung von Eigenwerten. Anders als im linearen Fall decken die Kriterien aber nicht *alle* möglichen Fälle ab und werden hier nur[3] für den reellen Fall $\mathbb{K} = \mathbb{R}$ angegeben.

Satz 3.5 (über **lineare Stabilitätskriterien für Fixpunkte nicht-linearer Systeme**). *Seien D eine Teilmenge von $\mathbb{R}^N$ und $\varphi \in \mathrm{C}^0(D, D)$ eine stetige Abbildung, so dass das Zeitdiskrete System $(\mathbb{T}, \mathbb{R}^N, \varphi)$ einen inneren Punkt x_0 von D als Fixpunkt besitzt. Sei φ an der Stelle x_0 total differenzierbar mit Ableitung $\mathrm{D}\varphi(x_0) \in \mathbb{R}^{N\times N}$, und seien $\lambda_1, \lambda_2, \ldots, \lambda_p \in \mathbb{C}$ mit $1 \le p \le N$ alle verschiedenen Eigenwerte von $\mathrm{D}\varphi(x_0)$ über $\mathbb{C}$. Dann gelten folgende Kriterien für Stabilität und Instabilität*:

(A) *Ist* $\mathbf{\max\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_p|\} < 1}$, *so ist der Fixpunkt* x_0 ***asymptotisch stabil***.

(B) *Ist* $\mathbf{\max\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_p|\} > 1}$, *so ist der Fixpunkt* x_0 ***instabil***.

Bemerkungen.

(1) Man bezeichnet die Kriterien des Satzes als **lineare Stabilitätskriterien**, **Linearisierungskriterien für Stabilität** oder **Erster-Ordnung-Stabilitätskriterien**, da in gewisser Weise Stabilität von der linearen Approximation, also der Erster-Ordnung-(Taylor-) Approximation $x \mapsto \varphi(x_0) + \mathrm{D}\varphi(x_0)(x - x_0)$ auf die nicht-lineare Abbildung $x \mapsto \varphi(x)$ übertragen wird; vergleiche mit dem Beweis.

(2) Im Fall $\mathbf{\max\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_p|\} = 1}$ lässt sich anhand der Eigenwerte (und auch anhand von $\mathrm{D}\varphi(x_0)$) allein **nicht über die Stabilität von x_0 entscheiden**.

(3) Die Relation zu den bereits zuvor diskutierten Stabilitätskriterien kann man wie folgt verstehen: In der Situation von Satz 3.5 (A) wird der folgende Beweis zeigen, dass $\|\mathrm{D}\varphi(x_0)\| < 1$ in einer geeigneten Operatornorm gilt; in diesem Fall handelt es sich also um eine Variante des schon nach Satz 3.4 festgehaltenen Ableitungskriteriums, wobei die wesentliche Voraussetzung in der neuen Version leichter handhabbar und nachprüfbar ist. In der Situation von Satz 3.5 (B) allerdings muss die nach Satz 3.4 formulierte hinreichende Bedingung $\left\|\mathrm{D}\varphi(x_0)^{-1}\right\| < 1$ für Instabilität bei Weitem nicht erfüllt sein. Um diese Bedingung sicherzustellen, müsste tatsächlich das Minimum statt des Maximums der Beträge der Eigenwerte größer als 1 sein. Insofern verschärft Satz 3.5 (B) das frühere Kriterium deutlich.

[3] Ist D in Satz 3.5 eine Teilmenge von $\mathbb{C}^N$ statt $\mathbb{R}^N$, so kann man $\mathbb{C}^N$ mit $\mathbb{R}^{2N}$ identifizieren und den Satz mit $2N$ anstelle von N anwenden; dies bedeutet — wie schon einmal erwähnt — dass man mit dem reellen Ableitungsbegriff arbeitet. Liegt komplexe Differenzierbarkeit vor, so lässt sich einsehen, dass die Kriterien genauso mit den Eigenwerten der komplexen Ableitung (als $\mathbb{C}$-lineare Abbildung $\mathbb{C}^N \to \mathbb{C}^N$ oder Matrix in $\mathbb{C}^{N\times N}$) gelten; dies bringt jedoch kaum einen Gewinn und wird hier nicht weiter betrachtet.

Beweis von Satz 3.5. Nach eventueller Verschiebung des Fixpunktes lässt sich $x_0 = 0$ annehmen.

Zuerst wird das Kriterium (A) behandelt. Mit den Abkürzungen

$$M := \mathrm{D}\varphi(0) \in \mathbb{R}^{N\times N} \qquad \text{und} \qquad \beta := 1 - \max\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_p|\} > 0$$

besagt die Voraussetzung dieses Falls, dass alle Eigenwerte von M Betrag $\leq (1-\beta)$ haben. Für die folgende Argumentation wird eine Jordan-Basis $v_1, v_2, \ldots, v_N$ zu M in $\mathbb{C}^N$ (deren Existenz aus der linearen Algebra bekannt ist) fixiert und eine zugehörige nicht-Euklidische Norm $\|\cdot\|_M$ auf $\mathbb{C}^N$ verwendet. Die Norm wird dabei zunächst auf den Basisvektoren v_i durch $\|v_i\|_M := (3/\beta)^{m_i}$ definiert, wobei m_i die minimale natürliche Zahl mit $(M-\lambda\mathbb{I}_N)^{m_i} v_i = 0$ für ein $\lambda \in \mathbb{C}$ bezeichnet (und die Existenz eines solchen m_i dadurch begründet ist, dass Vektoren einer Jordan-Basis in den Haupträumen liegen). Durch die Festlegung $\big\|\sum_{i=1}^N a_i v_i\big\|_M := \sum_{i=1}^N |a_i|\,\|v_i\|_M$ für $a_1, a_2, \ldots, a_N \in \mathbb{C}$ wird die Norm dann auf ganz $\mathbb{C}^N$ erklärt. Als entscheidende Eigenschaft der Norm $\|\cdot\|_M$ wird nun die Kontraktionseigenschaft

$$\|Mx\|_M \leq (1-\tfrac{2}{3}\beta)\|x\|_M \qquad \text{für alle } x \in \mathbb{C}^N \tag{3.4}$$

nachgewiesen. Dazu reicht es, (3.4) für die Vektoren der obigen Jordan-Basis zu verifizieren: Ist $v_i \in \mathrm{E}_\lambda(M)$ ein (Basis-)Vektor in einem Eigenraum, so erhält man sofort $\|Mv_i\|_M = |\lambda|\,\|v_i\|_M \leq (1-\beta)\|v_i\|_M$. Für einen Basis-Vektor $v_i \in \mathrm{H}_\lambda(M) \setminus \mathrm{E}_\lambda(M)$, der kein Eigenvektor ist, ist $m_i \geq 2$; es gelten $\|v_i\|_M = (3/\beta)^{m_i}$ und $\|(M-\lambda\mathbb{I}_N)v_i\|_M = (3/\beta)^{m_i-1}$, so dass sich $\|Mv_i\|_M \leq |\lambda|\,\|v_i\|_M + (3/\beta)^{m_i-1} \leq (1-\beta)\|v_i\|_M + \frac{1}{3}\beta\|v_i\|_M = (1-\frac{2}{3}\beta)\|v_i\|_M$ ergibt. Damit ist (3.4) nachgewiesen.

Als Nächstes sei daran erinnert, dass $x_0 = 0$ angenommen wurde, somit $\varphi(0) = 0$ gilt und $M = \mathrm{D}\varphi(0)$ gewählt wurde. Zu einem beliebigen $\varepsilon \in \mathbb{R}^+$ gibt es dann gemäß der Definition der totalen Ableitung (beziehungsweise einer einfachen Version des Satzes von Taylor) ein $\delta \in (0, \varepsilon)$, so dass für alle $x \in \mathbb{R}^N$ mit $\|x\|_M < \delta$ einerseits $x \in D$ und andererseits

$$\|\varphi(x) - Mx\|_M \leq \tfrac{1}{3}\beta\|x\|_M$$

gilt. In Kombination mit (3.4) folgt hieraus die Kontraktionseigenschaft

$$\|\varphi(x)\|_M \leq (1-\tfrac{1}{3}\beta)\|x\|_M \qquad \text{für alle } x \in \mathbb{R}^N \text{ mit } \|x\|_M < \delta$$

von φ. Ausgehend von dieser Eigenschaft lässt sich der Beweis durch eine Anwendung von Satz 3.4 (A) abschließen, doch die direkte Argumentation ist insgesamt noch einfacher: Insbesondere gilt nach obigem $\|\varphi(x)\|_M < \delta$ für alle $x \in \mathbb{R}^N$ mit $\|x\|_M < \delta$. Die Kontraktionseigenschaft von φ kann somit iteriert werden, und für $k \in \mathbb{N}_0$ und $x \in \mathbb{R}^N$ mit $\|x\|_M < \delta$ ergibt sich

$$\|\varphi^k(x)\|_M \leq (1-\tfrac{1}{3}\beta)^k\|x\|_M \leq \delta < \varepsilon\,.$$

Damit ist für x in der $\|\cdot\|_M$-δ-Kugel um 0 sowohl verifiziert, dass $\mathcal{O}^+(x)$ in der $\|\cdot\|_M$-ε-Kugel um 0 bleibt, als auch, dass $\lim_{k\to\infty} \varphi^k(x) = 0$ gilt. Also ist der Fixpunkt 0 asymptotisch stabil und der Beweis des Kriteriums (A) vollständig.

Das Kriterium (B) fußt größtenteils auf einer ähnlichen Argumentation, wird aber dadurch noch etwas subtiler, dass zusätzlich zu Eigenwerten vom Betrag > 1 auch Eigenwerte vom Betrag ≤ 1 auftreten können. Tatsächlich sei wieder $M := \mathrm{D}\varphi(0) \in \mathbb{R}^{N\times N}$ abgekürzt, doch anders als zuvor sei $\beta > 0$ jetzt derart fixiert, dass kein Eigenwert einen Betrag zwischen 1 und

$1+\beta$ aufweist, also so, dass $|\lambda_1|, |\lambda_2|, \ldots, |\lambda_p| \notin (1, 1+\beta)$ gilt. Weiterhin sei $\mathbb{C}^N = V_1 \oplus V_2$ in die $\mathbb{C}$-linearen Unterräume

$$V_1 := \bigoplus_{\substack{i=1 \\ |\lambda_i|\leq 1}}^{p} \mathrm{H}_{\lambda_i}(M) \qquad \text{und} \qquad V_2 := \bigoplus_{\substack{i=1 \\ |\lambda_i|\geq 1+\beta}}^{p} \mathrm{H}_{\lambda_i}(M)$$

zerlegt, wobei nach Annahme des Falls (B) stets $V_2 \neq \{0\}$ gilt (während $V_1 = \{0\}$ möglich ist). Mit Hilfe einer Jordan-Basis kann nun ähnlich wie zuvor eine von M abhängige Norm $\|\cdot\|_M$ auf $\mathbb{C}^N$ mit

$$\left.\begin{aligned} \|Mx_1\|_M &\leq (1+\tfrac{1}{6}\beta)\|x_1\|_M \\ \|Mx_2\|_M &\geq (1+\tfrac{5}{6}\beta)\|x_2\|_M \\ \|x_1+x_2\|_M &= \|x_1\|_M+\|x_2\|_M \end{aligned}\right\} \quad \text{für alle } x_1 \in V_1, x_2 \in V_2 \tag{3.5}$$

konstruiert werden. Ebenfalls wie zuvor gibt es ein $\varepsilon \in \mathbb{R}^+$, so dass alle $x \in \mathbb{R}^N$ mit $\|x\|_M < \varepsilon$ in D enthalten sind und

$$\|\varphi(x) - Mx\|_M \leq \tfrac{1}{6}\beta\|x\|_M$$

erfüllen. Da $\mathbb{C}^N = V_1 \oplus V_2$ gilt, kann insbesondere jedes $x \in \mathbb{R}^N$ eindeutig als $x = x_1+x_2$ mit $x_1 \in V_1$, $x_2 \in V_2$ aufgespalten werden, und insbesondere lässt sich auch $\varphi(x) = \varphi(x)_1 + \varphi(x)_2$ derart zerlegen. Es folgt

$$\left.\begin{aligned} \|\varphi(x)_1 - Mx_1\|_M &\leq \tfrac{1}{6}\beta\|x\|_M \\ \|\varphi(x)_2 - Mx_2\|_M &\leq \tfrac{1}{6}\beta\|x\|_M \end{aligned}\right\} \quad \text{für alle } x \in \mathbb{R}^N \text{ mit } \|x\|_M < \varepsilon$$

(wobei auch verwendet wurde, dass $Mx_1 \in V_1$ und $Mx_2 \in V_2$ gelten). Durch Einsetzten dieser Abschätzungen in (3.5) erhält man

$$\left.\begin{aligned} \|\varphi(x)_1\|_M &\leq (1+\tfrac{1}{3}\beta)\|x_1\|_M + \tfrac{1}{6}\beta\|x_2\|_M \\ \|\varphi(x)_2\|_M &\geq (1+\tfrac{2}{3}\beta)\|x_2\|_M - \tfrac{1}{6}\beta\|x_1\|_M \end{aligned}\right\} \quad \text{für alle } x \in \mathbb{R}^N \text{ mit } \|x\|_M < \varepsilon\,,$$

und Subtraktion der ersten von der zweiten Ungleichung ergibt

$$\|\varphi(x)_2\|_M - \|\varphi(x)_1\|_M \geq (1+\tfrac{1}{2}\beta)\big[\|x_2\|_M - \|x_1\|_M\big] \qquad \text{für alle } x \in \mathbb{R}^N \text{ mit } \|x\|_M < \varepsilon\,.$$

Gilt für $x \in \mathbb{R}^N$ mit $\|x\|_M < \varepsilon$ nun $\|x_2\|_M > \|x_1\|_M$, so folgt durch Iteration dieser letzten Ungleichung, dass es ein $k \in \mathbb{N}$ mit $\|\varphi^k(x)\|_M \geq \varepsilon$ gibt. Wäre dies nämlich nicht der Fall, so ließe sich die Ungleichung beliebig oft iterieren und man bekäme

$$\|\varphi^k(x)_2\|_M \geq \|\varphi^k(x)_2\|_M - \|\varphi^k(x)_1\|_M \geq (1+\tfrac{1}{2}\beta)^k\big[\|x_2\|_M - \|x_1\|_M\big] \overset{k\to\infty}{\longrightarrow} \infty\,,$$

was offensichtlich im Widerspruch zu $\|\varphi^k(x)_2\|_M \leq \|\varphi^k(x)\|_M < \varepsilon$ für alle $k \in \mathbb{N}$ steht. Nun gibt es beliebig nah bei 0 gelegene $x \in \mathbb{R}^N \setminus \{0\}$ mit $\|x\|_M < \varepsilon$ und $x_2 = x$, $x_1 = 0$ (denn für $0 \neq y \in V_2$ ist auch $\overline{y} \in V_2$ und damit $y+\overline{y} \in V_2 \cap \mathbb{R}^N$); und diese x erfüllen wie gerade begründet $\mathcal{O}^+(x) \not\subset \mathrm{B}_\varepsilon(0)$, daher ist der Fixpunkt 0 instabil für $(\mathbb{T}, \mathbb{R}^N, \varphi)$. Dies komplettiert den Beweis des Kriteriums (B). □

Teil II

Gewöhnliche Differentialgleichungen und kontinuierliche dynamische Systeme

Kapitel 4

Grundlagen und Terminologie, Typen von Differentialgleichungen

Eine gewöhnliche Differentialgleichung ist eine Gleichung, in der endlich viele Ableitungen einer Funktion einer reellen Variablen auftreten. Die Lösung einer derartigen Gleichung ist in der Mathematik sowie in Naturwissenschaft und Technik in vielen verschiedenen Zusammenhängen erforderlich und wird in diesem zweiten Vorlesungsteil systematisch behandelt — sowohl in rechnerischer als auch in theoretischer Hinsicht. Als Erstes werden hierzu der Gleichungs- und der Lösungsbegriff wie folgt präzisiert, wobei $\mathbb{K}$ weiterhin als Platzhalter für $\mathbb{R}$ oder $\mathbb{C}$ steht.

Terminologie. *Eine* ***gewöhnliche Differentialgleichung****, kurz* ***GDG*** *oder* **(*gewöhnliche*) *DGL****, ist eine (zunächst formale) Gleichung in der* ***impliziten Form***

$$g(\,\cdot\,, u, u', u'', \ldots, u^{(m-1)}, u^{(m)}) \equiv 0 \tag{4.1}$$

oder der nach $u^{(m)}$ aufgelösten, ***expliziten Form***

$$u^{(m)} = f(\,\cdot\,, u, u', u'', \ldots, u^{(m-1)})\,. \tag{4.2}$$

Gegeben sind dabei eine Zahl $m \in \mathbb{N} := \{1, 2, 3, \ldots\}$, genannt[1] *die* ***Ordnung der Gleichung****, normierte Räume $\mathcal{X}$ und $\mathcal{Z}$ über $\mathbb{K}$ und die* ***Strukturfunktion*** *$g\colon \widehat{D} \to \mathcal{Z}$ auf einem Definitionsbereich $\widehat{D} \subset \mathbb{R}\times\mathcal{X}^{1+m}$ beziehungsweise $f\colon D \to \mathcal{X}$ auf einem Definitionsbereich $D \subset \mathbb{R}\times\mathcal{X}^m$. Gesucht ist bei (4.1) und (4.2) stets die* ***unbekannte Funktion*** *u.*

Bemerkung. Die explizite Form ist weniger allgemein und entspricht dem Spezialfall $\widehat{D} = D \times \mathcal{X}$, $\mathcal{Z} = \mathcal{X}$, $g(t, x_0, x_1, \ldots, x_{m-1}, x_m) = x_m - f(t, x_0, x_1, \ldots, x_{m-1})$ der impliziten Form.

Die durch die Gleichungen (4.1) und (4.2) symbolisierten Forderungen an u sind wie folgt zu verstehen:

Definition 4.1 (Lösungen von DGLen). *Eine m-mal differenzierbare Funktion $u\colon I \to \mathcal{X}$ (mit Ableitung $u'\colon I \to \mathcal{X}$, zweiter Ableitung $u''\colon I \to \mathcal{X}$, …, m-ter Ableitung $u^{(m)}\colon I \to \mathcal{X}$) heißt eine Lösung der Differentialgleichung (4.1) beziehungsweise (4.2) auf einem Intervall I*

[1] Im Fall der impliziten Form (4.1) ist es tatsächlich nur dann sinnvoll, m als Ordnung der Gleichung zu bezeichnen, wenn $f(t, x_0, x_1, \ldots, x_{m-1}, x_m) \neq f(t, x_0, x_1, \ldots, x_{m-1}, \widetilde{x}_m)$ für gewisse $(t, x_0, x_1, \ldots, x_{m-1}, x_m)$ und $(t, x_0, x_1, \ldots, x_{m-1}, \widetilde{x}_m)$ in $\widehat{D}$ vorkommt und somit sichergestellt ist, dass $u^{(m)}$ tatsächlich eine Rolle spielt. Im Fall der expliziten Form (4.2) ist die Ordnung m auch ohne eine solche Zusatzbedingung sinnvoll definiert.

positiver Länge in $\mathbb{R}$*, wenn für alle* $t \in I$ *gilt:* $(t, u(t), u'(t), u''(t), \ldots, u^{(m-1)}(t), u^{(m)}(t)) \in \widehat{D}$ *erfüllt*

$$g(t, u(t), u'(t), u''(t), \ldots, u^{(m-1)}(t), u^{(m)}(t)) = 0 \qquad \text{in } \mathcal{Z}$$

beziehungsweise $(t, u(t), u'(t), u''(t), \ldots, u^{(m-1)}(t)) \in D$ *erfüllt*

$$u^{(m)}(t) = f(t, u(t), u'(t), u''(t), \ldots, u^{(m-1)}(t)) \qquad \text{in } \mathcal{X}\,.$$

(*Hierbei interpretiert man Differenzierbarkeit und Ableitungen in eventuell zu* I *gehörigen Randpunkten als einseitige Begriffsbildungen, also* $u^{(k)}(a) := \lim_{h \searrow 0} \frac{u^{(k-1)}(a+h) - u^{(k-1)}(a)}{h}$ *für einen linken Randpunkt* a *und* $u^{(k)}(b) := \lim_{h \searrow 0} \frac{u^{(k-1)}(b) - u^{(k-1)}(b-h)}{h}$ *für einen rechten Randpunkt* b*.*)

Entscheidendes Merkmal von GDGen ist, dass die unbekannte Funktion u auf einer Teilmenge von $\mathbb{R}$ definiert und somit eine **Funktion einer reellen Variablen** ist. Dies wird durch das Wort ‚gewöhnlich' zum Ausdruck gebracht und unterscheidet GDGen von den sogenannten **partiellen Differentialgleichungen**, bei denen die unbekannte Funktion von 2 oder mehr reellen Variablen abhängt und partielle Ableitungen nach diesen Variablen auftreten. Auch Differentialgleichungen für Funktionen einer komplexen Variablen gehören (wegen der Entsprechung $\mathbb{C} = \mathbb{R}^2$) eher ins Gebiet der partiellen Differentialgleichungen.

Grundlegende Fragen bei DGLen sind die nach

(1) **expliziten Formeln** für Lösungen,

(2) **Existenz** von Lösungen,

(3) **Eindeutigkeit** von Lösungen,

(4) **qualitativen Aussagen** über das Verhalten von Lösungen,

(5) **Stabilität** der Lösungen gegen Störungen, Abhängigkeit der Lösungen von den Daten,

(6) **numerischer Berechnung** der Lösungen.

Im nächsten Vorlesungskapitel werden verschiedene Methoden zur Behandlung der ersten Frage, also der Berechnung expliziter Lösungen, vorgestellt. Weitere Methoden werden sich später im Kontext der linearen Theorie ergeben, und tatsächlich gibt es auch darüber hinaus noch etliche weitere Methoden und Vorgehensweisen. Dennoch ist die Herleitung expliziter Lösungsformeln insgesamt nur für spezielle (Typen von) Gleichungen möglich. Dies verhält sich ähnlich zur Integration und der Bestimmung von Stammfunktionen: Auch dort gibt es verschiedene Verfahrensweisen (Produktintegration, Substitutionen, Partialbruchzerlegung, ...), aber dennoch können Stammfunktionen nicht in allen Fällen als elementare Funktionen durch explizite Formeln angegeben werden. Wegen dieser prinzipiellen Ähnlichkeit, aber auch, weil manche Lösungsverfahren für DGLen die Bestimmung von Stammfunktionen als Teilschritte beinhalten, spricht man bei der Bestimmung von Lösungen einer gegebenen DGL auch von der **Integration der DGL**.

Die obigen Punkte 2–6 sind insbesondere in denjenigen Fällen von Interesse, in denen nicht explizit gelöst werden kann. In dieser Vorlesung sollen hierbei erst die grundlegenden Antworten auf die Fragen 2 und 3 in Form abstrakter Existenz- und Eindeutigkeitssätze vorgestellt werden. Aufbauend auf diesen Antworten macht es dann Sinn, sich mit den Fragen 4 und 5, also qualitativen Aussagen und Stabilitätstheorie, zu beschäftigen. Auch diese beiden Themenkreise werden dabei wesentliche Themen der Vorlesung sein. Insbesondere soll dabei auf diejenigen Aspekte der

zugehörigen Theorie eingegangen werden, die einen Zusammenhang zu dynamischen Systemen herstellen und/oder die bei dynamischen Systemen typischen Fragen wie Langzeitverhalten und Langzeit-Stabilität betreffen. Sind auch die Fragen 4 und 5 (zumindest in Teilen) positiv beantwortet, so kann man prinzipiell zur numerischen Berechnung von Lösungen kommen; letzteres wird jedoch kein Thema dieser Vorlesung sein.

Terminologie. *Sind* $N := \dim_{\mathbb{K}} \mathcal{X} < \infty$ *und* $M := \dim_{\mathbb{K}} \mathcal{Z} < \infty$*, so kann man direkt*

$$\mathcal{X} = \mathbb{K}^N \qquad \text{und} \qquad \mathcal{Z} = \mathbb{K}^M$$

annehmen. Dann bildet (4.1) ein System von M (Komponenten-)Gleichungen für N unbekannte (Komponenten-)Funktionen. Man unterscheidet zwischen einer ***(Einzel-)Gleichung*** *($M = 1$) oder einem* ***System von Gleichungen*** *($M \geq 2$) für* ***eine skalare Funktion*** *($N = 1$) oder* ***mehrere Funktionen*** *($N \geq 2$). In der Regel ist nur der Fall $M = N$ mit genau so vielen Gleichungen wie unbekannten Funktionen sinnvoll, denn nur in diesem kann man allgemeine Sätze über Existenz und Eindeutigkeit von Lösungen erwarten. Für (4.2) gilt im Wesentlichen dasselbe, wobei man sinngemäß immer im Fall $\mathcal{Z} = \mathcal{X}$, $M = N$ ist.*

Bemerkung. Es ist sinnvoll, den Fall $\dim_{\mathbb{K}} \mathcal{X} = \dim_{\mathbb{K}} \mathcal{Z} = \infty$ zuzulassen, denn dies erlaubt es manchmal, partielle Differentialgleichungen formal als GDGen aufzufassen. Beispielsweise ist die Wärmeleitungsgleichung

$$\frac{\partial}{\partial t} w(t,x) = \frac{\partial^2}{\partial x^2} w(t,x) \qquad \text{für } (t,x) \in \mathbb{R}^2$$

für Funktionen $w\colon \mathbb{R}^2 \to \mathbb{R}$ formal äquivalent zur GDG

$$u'(t) = A[u(t)]$$

mit dem linearen Ableitungsoperator $A[f] := f''$, wobei $u(t)$ der partiellen Funktion $w(t,\cdot)$ entspricht und u daher Werte in einem ∞-dimensionalen Funktionenraum $\mathcal{Z} = \mathcal{X}$ annimmt.

Terminologie (Autonome DGLen, lineare DGLen).

(1) *Hängt die Strukturfunktion in (4.1) beziehungsweise (4.2) nicht explizit von der ersten Variablen, der t-Variablen, ab und kann die Gleichung somit in der Form*

$$\widetilde{g}(u, u', u'', \dots, u^{(m-1)}, u^{(m)}) \equiv 0 \qquad \text{beziehungsweise} \qquad u^{(m)} = \widetilde{f}(u, u', u'', \dots, u^{(m-1)})$$

geschrieben werden, so spricht man von einer ***autonomen DGL****.*

(2) *Ist die Strukturfunktion in (4.1) beziehungsweise (4.2)* ***bei fixierter t-Variable*** *(stetig und)* ***affin linear in allen weiteren Variablen****, und ist $\widehat{D} = I \times \mathcal{X}^{1+m}$ beziehungsweise $D = I \times \mathcal{X}^m$ mit $I \subset \mathbb{R}$, so spricht man von einer* ***linearen DGL****. Lineare DGL sind damit genau die, die in der Form*

$$A_m u^{(m)} + A_{m-1} u^{(m-1)} + \ldots + A_2 u'' + A_1 u' + A_0 u = b \qquad \text{auf } I \tag{4.3}$$

mit ***Koeffizienten(funktionen)*** *$A_k\colon I \to \mathcal{L}(\mathcal{X}, \mathcal{Z})$ und* ***Inhomogenität*** *$b\colon I \to \mathcal{Z}$ geschrieben werden können (bei expliziter Form mit $\mathcal{Z} = \mathcal{X}$ und ohne $A_m \equiv \mathrm{id}_{\mathcal{X}}$). Hierbei bezeichnet $\mathcal{L}(\mathcal{X}, \mathcal{Z})$ den Raum der (stetigen) linearen Abbildungen von $\mathcal{X}$ nach $\mathcal{Z}$. Daher kann man die Koeffizienten im Fall $\mathcal{X} = \mathbb{K}^N$, $\mathcal{Z} = \mathbb{K}^M$ als von der t-Variablen abhängige (reelle oder komplexe) $(M \times N)$-Matrizen betrachten, und analog ist die Inhomogenität dann ein t-abhängiger Vektor in $\mathbb{K}^M$. Speziell im Fall $M = N = 1$ einer einzelnen skalaren Gleichung handelt es sich sowohl bei A_k als auch bei b einfach um einzelne $\mathbb{R}$- oder $\mathbb{C}$-wertige Funktionen von t (und man schreibt dann meist a_k statt A_k).*

(3) *Die lineare DGL in* (4.3) *hat* ***konstante Koeffizienten****, wenn die Koeffizientenfunktionen* $A_0, A_1, A_2 \dots, A_{m-1}, A_m$ *alle konstant sind. Sie heißt* ***homogen****, wenn* $b \equiv 0$ *gilt, sonst heißt sie* ***inhomogen****.*

Beispiele (von einzelnen GDGen).

(1) Für jedes $m \in \mathbb{N}$ ist die Gleichung

$$u^{(m)} \equiv 0 \qquad \text{auf } \mathbb{R}$$

für $\mathbb{K}$-wertiges u ein (trivial zu lösendes) Beispiel einer einzelnen, skalaren linearen GDG m-ter Ordnung, die genau die Polynomfunktionen der Ordnung $\leq (m-1)$ als Lösungen hat. Man erkennt an diesem Beispiel bereits den **typischen Effekt, dass die Lösung nicht völlig eindeutig ist, sondern von m Parametern abhängt**, nämlich von den m reellen oder komplexen Koeffizienten eines solchen Polynoms.

(2) Ein weniger triviales Beispiel mit Parameter $\lambda \in \mathbb{K}$ ist die Gleichung

$$u' = \lambda u \qquad \text{auf } \mathbb{R}$$

für $\mathbb{K}$-wertiges u. Dies ist eine einzelne, skalare, homogene lineare GDG erster Ordnung mit konstantem Koeffizient λ. Leicht zu ratende Lösungen sind die (reellen oder komplexen) Exponentialfunktionen

$$u(t) = C\mathrm{e}^{\lambda t}$$

mit einer Konstante $C \in \mathbb{K}$ (und später wird klar, dass dies schon alle Lösungen sind).

(3) Eine einzelne, nicht-autonome, nicht-lineare GDG erster Ordnung in expliziter Form ist[2]

$$u' = \frac{u^2}{t^3}$$

für $\mathbb{K}$-wertiges u. Hier ist allerdings $D = (\mathbb{R}\setminus\{0\}) \times \mathbb{K}$ und daher kann man nur auf Teilintervallen von $(-\infty, 0)$ und $(0, \infty)$ lösen. Auf solchen Intervallen findet man die Lösungen $u \equiv 0$ und (mit etwas Herumprobieren) $u(t) = 2t^2$. Weitere Lösungen sind nicht gleichermaßen offensichtlich, später kann aber eine 1-Parameter-Schar von Lösungen leicht ausgerechnet werden; siehe Kapitel 5.1 und 5.3.

(4) Dass bei impliziter Form viel mehr schief gehen kann, sieht man schon an den zwei einfachen Beispielen

$$\mathrm{e}^{u'} \equiv 0 \qquad \text{und} \qquad u'(1-u') \equiv 0\,.$$

für $\mathbb{K}$-wertiges u. Beides sind einzelne, autonome GDGen erster Ordnung, aber die eine besitzt überhaupt keine Lösung, die andere besitzt die *zwei* 1-Parameter-Familien von Lösungen $u_0(t) = C$ und $u_1(t) = t+C$ mit einer Konstante $C \in \mathbb{K}$; somit hat man bei letzterer Gleichung einen Parameter aus $\{0, 1\} \times \mathbb{K}$ frei, und das ist auch schlecht (besonders für AWPe; vergleiche unten). Insgesamt kann man wegen solcher Beispiele **keine gute Theorie von GDGen der allgemeinen impliziten Form** (4.1) erwarten, und daher **geht man in der Theorie fast immer von der expliziten Form** (4.2) **aus**.

[2] Eigentlich vermeidet man die Vermischung von Funktionen (wie u' und u^2) und Termen (wie t^3). Da von den beiden konsequenteren Notationen (hier wären dies $u'(t) = \frac{u(t)^2}{t^3}$ und $u' = \frac{u^2}{(\cdot)^3}$) aber eine sperrig und eine schlecht lesbar ist, ist es üblich, bei konkreten GDGen eine Ausnahme zu machen und sie wie oben zu notieren.

Als Nächstes seien zwei ganz unterschiedliche allgemeine Prinzipien erwähnt:

Prinzip (Reduktion auf Ordnung 1). *Die allgemeine GDG m-ter Ordnung in (4.1) für die unbekannte Funktion u ist äquivalent zum System von m GDGen erster Ordnung*

$$\begin{aligned} u' &= u_1 \\ u_1' &= u_2 \\ u_2' &= u_3 \\ &\vdots \\ u_{m-2}' &= u_{m-1} \\ g(\,\cdot\,, u, u_1, u_2, \dots, u_{m-1}, u_{m-1}') &= 0 \end{aligned} \tag{4.4}$$

für die m unbekannten Funktionen $(u, u_1, u_2, \dots, u_{m-1})$; und auf dieselbe Art und Weise ist (4.2) äquivalent zu einem System erster Ordnung in expliziter Form. Daher kann man sich bei der Untersuchung von GDGen prinzipiell auf den einfacheren Fall der Ordnung 1 beschränken. Dies ist ***wichtig für die Theorie*** *und nicht immer, aber zumindest* ***manchmal nützlich für die explizite Berechnung*** *von Lösungen. Man zahlt aber natürlich einen Preis für die Reduktion der Ordnung, denn auch im Fall, dass (4.1) nur eine einzelne skalare Gleichung der Ordnung $m \geq 2$ ist, ist (4.4) stets ein System; und wenn (4.1) selbst schon ein System der Ordnung $m \geq 2$ mit N Gleichungen für N unbekannte Funktionen ist, so ist (4.4) ein noch größeres System mit mN Gleichungen für mN unbekannte Funktionen.*

Prinzip (Aneinandersetzen von Lösungen). *Sei eine GDG m-ter Ordnung in der expliziten Form (4.2) gegeben. Kennt man sowohl eine Lösung u_1 auf einem Intervall (t_1, τ) als auch eine Lösung u_2 auf einem Intervall (τ, t_2) (mit $-\infty \leq t_1 < \tau < t_2 \leq \infty$), gilt*

$$\lim_{t \nearrow \tau} u_1^{(k)}(t) = y_k = \lim_{t \searrow \tau} u_2^{(k)}(t) \qquad \text{für } k = 0, 1, \dots, m-1\,,$$

und ist die Strukturfunktion f stetig an der Stelle $(\tau, y_0, y_1, \dots, y_{m-1}) \in D$, so ***erhält man durch Aneinandersetzen***

$$u(t) := \begin{cases} u_1(t) & \text{für } t < \tau \\ y_0 & \text{für } t = \tau \\ u_2(t) & \text{für } t > \tau \end{cases}$$

eine ***Lösung*** *u der GDG* ***auf dem Gesamtintervall*** *(t_1, t_2). Insbesondere ist $u^{(m-1)}$ dann bei τ differenzierbar und weist dort keinen Knick auf.*

Beweis. Für $k = 0, 1, \dots, m-1$ existiert die Ableitung $u^{(k)}$ auf ganz (t_1, t_2) und nimmt den Wert $u^{(k)}(\tau) = y_k$ stetig an. Auf den Teilintervallen (t_1, τ) und (τ, t_2) ist u sogar m-mal differenzierbar und löst die GDG. Mit der Stetigkeit von f in $(\tau, y_0, y_1, \dots, y_{m-1})$ ergibt sich daher

$$\lim_{t \to \tau} u^{(m)}(t) = \lim_{t \to \tau} f(t, u(t), u'(t), \dots, u^{(m-1)}(t)) = f(\tau, u(\tau), u'(\tau), \dots, u^{(m-1)}(\tau))\,.$$

Mit dem Mittelwertsatz der Differentialrechnung folgt

$$\lim_{h \to 0} \frac{u^{(m-1)}(\tau+h) - u^{(m-1)}(\tau)}{h} = f(\tau, u(\tau), u'(\tau), \dots, u^{(m-1)}(\tau))$$

und somit die Existenz von $u^{(m)}(\tau)$ und die Lösungseigenschaft von u auf ganz (t_1, t_2). □

Schließlich werden noch einige weitere zentrale Begriffe eingeführt:

Terminologie (**Allgemeine und spezielle Lösungen, Anfangswertproblem**). *Auch bei sehr gutartigen GDGen kann man nicht erwarten, dass sie auf einem Intervall I positiver Länge in $\mathbb{R}$ eine eindeutige Lösung besitzen. Vielmehr ist es typisch, dass die Lösungsmenge, auch die* ***allgemeine Lösung*** *genannt,* ***von m Integrationskonstanten*** *aus $\mathcal{X}$* ***als Parametern abhängt*** *und man eine eindeutig bestimmte* ***spezielle Lösung*** *u erst durch die Forderung von m Nebenbedingungen erhält. Als verbreitetste Form von Nebenbedingungen stellt man m* ***Anfangsbedingungen (ABen)***

$$u(t_0) = y_0\,, \quad u'(t_0) = y_1\,, \quad u''(t_0) = y_2\,, \quad \dots \quad u^{(m-1)}(t_0) = y_{m-1} \tag{4.5}$$

mit gegebenem $t_0 \in I$ und gegebenen $y_0, y_1, y_2, \dots, y_{m-1} \in \mathcal{X}$. Koppelt man eine DGL der Form (4.1) oder (4.2) mit den Anfangsbedingungen aus (4.5), so spricht man von einem ***Anfangswertproblem (AWP)****, für das man bei gutartigen GDGen die Existenz einer* eindeutigen *Lösung erwarten kann.*

Eine andere häufige Form von Nebenbedingungen im Fall $I = [a, b]$ sind m ***Randbedingungen****, die jeweils $u(a), u'(a), u''(a), \dots, u^{(m-1)}(a)$ oder $u(b), u'(b), u''(b), \dots, u^{(m-1)}(b)$ involvieren.*

Kapitel 5

Lösungsmethoden für spezielle Typen von Gleichungen

In diesem Kapitel werden verschiedene Lösungsmethoden und Typen von explizit lösbaren DGLen vorgestellt.

5.1 Lösungsformel für die allgemeine skalare lineare GDG erster Ordnung

Die allgemeine skalare lineare GDG erster Ordnung in expliziter Form lautet

$$u' = au + b \tag{5.1}$$

mit Koeffizientenfunktion $a\colon I \to \mathbb{K}$ und Inhomogenität $b\colon I \to \mathbb{K}$ auf $I \subset \mathbb{R}$. Diese DGL kann mit höchstens zwei Quadraturen, d.h. Stammfunktionsbildungen beziehungsweise Integrationen, allgemein gelöst werden:

Satz 5.1 (**Lösung skalarer linearer GDGen erster Ordnung**). *Sei I ein Intervall positiver Länge in $\mathbb{R}$, und seien $a, b \in \mathrm{C}^0(I, \mathbb{K})$ stetig. Dann ist die allgemeine Lösung der DGL (5.1) auf I von der Form $u = \mathrm{e}^A(B+C)$ mit Stammfunktionen A zu a und B zu $\mathrm{e}^{-A}b$ auf I sowie einer Konstanten $C \in \mathbb{K}$. Fordert man zusätzliche eine AB*

$$u(t_0) = y_0 \tag{5.2}$$

mit $t_0 \in I$, $y_0 \in \mathbb{K}$, so ist die eindeutige Lösung des AWPs (5.1)–(5.2) auf I gegeben durch

$$u(t) = \mathrm{e}^{A(t)}[B(t) - B(t_0) + y_0\mathrm{e}^{-A(t_0)}] = \mathrm{e}^{A(t)}\left[\int_{t_0}^{t} \mathrm{e}^{-A(s)}b(s)\,\mathrm{d}s + y_0\mathrm{e}^{-A(t_0)}\right].$$

Bemerkung. Im **homogenen Fall** $b \equiv 0$, $B \equiv 0$ erhält man für die allgemeine Lösung zu (5.1) die Form $u = C\mathrm{e}^A$, und die Formel für die eindeutige Lösung des AWPs (5.1)–(5.2) auf I vereinfacht sich zu

$$u(t) = y_0\mathrm{e}^{A(t)-A(t_0)} = y_0 \exp\left[\int_{t_0}^{t} a(s)\,\mathrm{d}s\right].$$

Beweis von Satz 5.1. Dass $\mathrm{e}^{A}(B+C)$ Lösung von (5.1) ist, verifiziert man durch die Rechnung (mit HDI, Produkt- und Kettenregel; beachte $A' = a$, $B' = \mathrm{e}^{-A}b$)

$$\bigl[\mathrm{e}^{A}(B+C)\bigr]' = a\mathrm{e}^{A}(B+C) + \mathrm{e}^{A}\mathrm{e}^{-A}b = a\bigl[\mathrm{e}^{A}(B+C)\bigr] + b\,.$$

Um zu zeigen, dass alle Lösungen die behauptete Form haben, argumentiert man wie folgt:

(i) Ist w Lösung der homogenen DGL $w' = aw$, so folgt

$$\bigl(\mathrm{e}^{-A}w\bigr)' = \mathrm{e}^{-A}[w'-aw] \equiv 0 \qquad \text{auf } I\,.$$

Nach dem Konstanzsatz ist damit $\mathrm{e}^{-A}w \equiv C$ und $w = C\mathrm{e}^{A}$ auf I.

(ii) Ist u Lösung der inhomogenen DGL (5.1), so folgt

$$\bigl(u-\mathrm{e}^{A}B\bigr)' = u' - a\mathrm{e}^{A}B - \mathrm{e}^{A}\mathrm{e}^{-A}b = au + b - a\mathrm{e}^{A}B - b = a\bigl(u-\mathrm{e}^{A}B\bigr) \qquad \text{auf } I\,,$$

und daher löst $u-\mathrm{e}^{A}B$ die homogene DGL. Gemäß (i) ergibt sich erst $u-\mathrm{e}^{A}B = C\mathrm{e}^{A}$ und dann $u = \mathrm{e}^{A}(B+C)$ auf I.

Die Behauptung über die Anfangswerte folgt problemlos, indem man die Konstante C durch Einsetzen bestimmt. □

Beispiele.

(1) Bei

$$u' = \lambda u \qquad \text{auf } \mathbb{R}$$

mit $\lambda \in \mathbb{K}$ zeigt der Satz (mit $a \equiv -\lambda$, $A(t) = -\lambda t$), dass die zuvor geratenen Lösungen $u(t) = C\mathrm{e}^{\lambda t}$ schon *alle* Lösungen sind.

(2) Als Lösung des AWPs auf dem Intervall $\mathbb{R}^{+}$

$$u' = \tfrac{2}{t}u + t\log t\,, \qquad\qquad u(1) = 2$$

(log steht für den natürlichen Logarithmus) berechnet man mit $A(t) = 2\log t$ im Satz:

$$\begin{aligned} u(t) &= \mathrm{e}^{2\log t}\left[\int_1^t \mathrm{e}^{-2\log s} s\log s\,\mathrm{d}s + 2\mathrm{e}^{-2\log 1}\right] \\ &= t^2\left[\int_1^t \frac{\log s}{s}\,\mathrm{d}s + 2\right] \\ &= t^2\left[\tfrac{1}{2}(\log t)^2 - \frac{1}{2}(\log 1)^2 + 2\right] \\ &= \tfrac{1}{2}t^2(\log t)^2 + 2t^2\,. \end{aligned}$$

5.2 Exponentialansatz bei skalaren linearen Gleichungen mit konstanten Koeffizienten

Assoziiert mit einer skalaren linearen Gleichung

$$a_m u^{(m)} + a_{m-1} u^{(m-1)} + \ldots + a_2 u'' + a_1 u' + a_0 u = b \qquad \text{auf } I \tag{5.3}$$

ist die t-abhängige Polynomfunktion p über $\mathbb{K}$ zu

$$p(t,\lambda) := a_m(t)\lambda^m + a_{m-1}(t)\lambda^{m-1} + \ldots + a_2(t)\lambda^2 + a_1(t)\lambda + a_0(t)\,.$$

Man nennt p das **Symbol** der Gleichung (5.3) oder das **charakteristische Polynom** der Gleichung (5.3). Die Ordnung von p stimmt mit der Ordnung m der Gleichung überein, und man kann (5.3) äquivalent in der formalen Schreibweise

$$\Big[a_m \frac{\mathrm{d}^m}{\mathrm{d}t^m} + a_{m-1}\frac{\mathrm{d}^{m-1}}{\mathrm{d}t^{m-1}} + \ldots + a_2\frac{\mathrm{d}^2}{\mathrm{d}t^2} + a_1\frac{\mathrm{d}}{\mathrm{d}t} + a_0\Big]u = b \qquad \text{auf } I$$

oder kurz

$$p\Big(\,\cdot\,, \frac{\mathrm{d}}{\mathrm{d}t}\Big)u = b \qquad \text{auf } I \tag{5.4}$$

ausdrücken. Besonders nützlich ist dies *im Fall homogener GDGen mit konstanten Koeffizienten*, in dem p ein einzelnes, nicht von t abhängiges Polynom ist: Man macht dann — motiviert durch den bereits behandelten Erster-Ordnung-Fall — den **Exponentialansatz**

$$u(t) = \mathrm{e}^{\lambda t}\,,$$

und erhält wegen

$$p\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big)\mathrm{e}^{\lambda t} = p(\lambda)\mathrm{e}^{\lambda t}$$

genau dann eine Lösung von (5.4) und (5.3) mit $b \equiv 0$, wenn $\lambda \in \mathbb{K}$ eine Nullstelle von p ist. Der weitere Ausbau dieser Idee führt auf …

Satz 5.2 (**Lösung *homogener* linearer GDGen mit konstanten Koeffizienten über $\mathbb{C}$**). *Seien $m \in \mathbb{N}$ und $a_0, a_1, a_2, \ldots, a_{m-1}, a_m \in \mathbb{K}$ mit $a_m \neq 0$, und seien $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{C}$ die verschiedenen Nullstellen des charakteristischen Polynoms der Gleichung (5.3) mit zugehörigen Vielfachheiten $d_1, d_2, \ldots, d_k \in \mathbb{N}$ (folglich $d_1+d_2+\ldots+d_k = m$ nach dem Fundamentalsatz der Algebra). Dann ist die allgemeine $\mathbb{C}$-wertige Lösung der homogenen Gleichung (5.3) mit $b \equiv 0$ auf einem Intervall I positiver Länge in $\mathbb{R}$ eine $\mathbb{C}$-Linearkombination der m Funktionsterme*

$$\begin{array}{lllll}
\mathrm{e}^{\lambda_1 t}\,, & t\mathrm{e}^{\lambda_1 t}\,, & t^2\mathrm{e}^{\lambda_1 t}\,, & \ldots\,, & t^{d_1-1}\mathrm{e}^{\lambda_1 t} \qquad (d_1 \textit{ Funktionen})\,,\\
\mathrm{e}^{\lambda_2 t}\,, & t\mathrm{e}^{\lambda_2 t}\,, & t^2\mathrm{e}^{\lambda_2 t}\,, & \ldots\,, & t^{d_2-1}\mathrm{e}^{\lambda_2 t} \qquad (d_2 \textit{ Funktionen})\,,\\
\vdots & \vdots & \vdots & & \\
\mathrm{e}^{\lambda_k t}\,, & t\mathrm{e}^{\lambda_k t}\,, & t^2\mathrm{e}^{\lambda_k t}\,, & \ldots\,, & t^{d_k-1}\mathrm{e}^{\lambda_k t} \qquad (d_k \textit{ Funktionen})\,.
\end{array}$$

Mit andern Worten ist die allgemeine $\mathbb{C}$-wertige Lösung u gegeben durch

$$u(t) = \sum_{i=1}^{k}\sum_{j=0}^{d_i-1} C_{i,j}\, t^j \mathrm{e}^{\lambda_i t} \qquad \textit{mit } m \textit{ Konstanten } C_{i,j} \in \mathbb{C}\,. \tag{5.5}$$

Bemerkungen.

(1) Hat das charakteristischen Polynom $\boldsymbol{m}$ verschiedene **Nullstellen** $\lambda_1, \lambda_2, \dots, \lambda_m \in \mathbb{C}$, so haben diese alle **Vielfachheit $\boldsymbol{d_i = 1}$**, und die allgemeine Lösung ist gegeben durch

$$u(t) = \sum_{i=1}^{m} C_i \mathrm{e}^{\lambda_i t} \qquad \text{mit Konstanten } C_1, C_2, \dots, C_m \in \mathbb{C}$$

(in diesem Fall keine Potenzen t^j, sondern ausschließlich Exponentialfunktionen).

(2) Das **AWP** mit ABen

$$u(t_0) = y_0\,, \quad u'(t_0) = y_1\,, \quad u''(t_0) = y_2\,, \quad \dots \quad u^{(m-1)}(t_0) = y_{m-1}$$

(zu gegebenen Parametern $t_0 \in I$ und $y_0, y_1, \dots, y_{m-1} \in \mathbb{C}$) **kann stets eindeutig gelöst werden**, indem man die allgemeine Form (5.5) der Lösung einsetzt und ein lineares Gleichungssystem für die Koeffizienten $C_{i,j}$ löst (m Gleichungen für m Koeffizienten). Eine einfache Begründung für die eindeutige Lösbarkeit dieses Systems ergibt sich aus später in der Vorlesung behandelter Theorie; siehe Bemerkung (4) in Kapitel 7.1.

Beweis von Satz 5.2. Man argumentiert durch Induktion nach $m \in \mathbb{N}$, wobei der Induktionsanfang für $m = 1$ (nach Durchteilen durch a_m) durch Satz 5.1 abgedeckt ist. Für den Induktionsschluss von Ordnung $(m-1)$ auf Ordnung $m \geq 2$ faktorisiert man das charakteristische Polynom p in der Form $p(\lambda) = \widetilde{p}(\lambda)(\lambda - \lambda_k)$ mit einem Polynom $\widetilde{p}$ der Ordnung $(m-1)$. Dann gilt für $\mathbb{C}$-wertiges u und eine ebenfalls $\mathbb{C}$-wertige Hilfsfunktion w auf I (genauere Erklärungen folgen unten)

$$\begin{aligned}
p\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big)u \equiv 0 \;&\Longleftrightarrow\; \Big(\frac{\mathrm{d}}{\mathrm{d}t} - \lambda_k\Big)u = w \text{ und } \widetilde{p}\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big)w \equiv 0 \\
&\Longleftrightarrow\; u' - \lambda_k u = w \text{ und } w(t) = \sum_{i=1}^{k-1}\sum_{j=0}^{d_i-1} \widetilde{C}_{i,j} t^j \mathrm{e}^{\lambda_i t} + \sum_{j=0}^{d_k-2} \widetilde{C}_{k,j} t^j \mathrm{e}^{\lambda_k t} \\
&\Longleftrightarrow\; u(t) = \mathrm{e}^{\lambda_k t}\Bigg[\sum_{i=1}^{k-1}\sum_{j=0}^{d_i-1} \widetilde{C}_{i,j} \int t^j \mathrm{e}^{(\lambda_i - \lambda_k)t}\,\mathrm{d}t + \sum_{j=0}^{d_k-2} \widetilde{C}_{k,j} \int t^j\,\mathrm{d}t\Bigg] \\
&\Longleftrightarrow\; u(t) = \sum_{i=1}^{k-1}\sum_{j=0}^{d_i-1} C_{i,j} t^j \mathrm{e}^{\lambda_i t} + C_{k,0}\mathrm{e}^{\lambda_k t} + \sum_{j=0}^{d_k-2} C_{k,j+1} t^{j+1} \mathrm{e}^{\lambda_k t}\,.
\end{aligned}$$

Hierbei ergibt sich die erste Äquivalenz direkt aus der Faktorisierung von p sowie (implizit verwendeten) Rechenregeln für Ableitungen. Die zweite Äquivalenz resultiert aus der Anwendung der Induktionsannahme für die zu $\widetilde{p}$ gehörige Gleichung der Ordnung $(m-1)$, wobei zu berücksichtigen ist, dass λ_k für $\widetilde{p}$ Nullstelle der um Eins verringerten Vielfachheit $(d_k - 1)$ ist. Die dritte Äquivalenz basiert auf Satz 5.1, wobei die unbestimmten Integrale als Notation für (beliebige) Stammfunktionen verwendet wurden. Bei der vierten und letzten Äquivalenz wurden diese Stammfunktionen bestimmt (bis auf den hier irrelevanten genauen Zusammenhang zwischen $\widetilde{C}_{i,j}$ und $C_{i,j}$); beim vorderen Term geht dies mit j Produktintegrationen; $C_{k,0}$ bezeichnet die neu hinzugekommene Integrationskonstante. Da die letzte Formel für u bis auf eine Indexverschiebung mit (5.5) übereinstimmt, ist die Induktion vollständig und der Beweis komplett. □

Beispiele.

(1) Zur Gleichung
$$u'' = u$$
gehört das charakteristische Polynom $\lambda^2-1 = (\lambda-1)(\lambda+1)$ mit den einfachen Nullstellen 1 und -1. Die allgemeine Lösung ist daher gegeben durch
$$u(t) = C_1\mathrm{e}^t + C_2\mathrm{e}^{-t} = (C_1+C_2)\cosh t + (C_1-C_2)\sinh t\,.$$

(2) Zur Gleichung
$$u'' = -u$$
gehört das charakteristische Polynom $\lambda^2+1 = (\lambda-\mathring{\mathrm{i}})(\lambda+\mathring{\mathrm{i}})$ mit den einfachen Nullstellen $\mathring{\mathrm{i}}$ und $-\mathring{\mathrm{i}}$. Die allgemeine Lösung ist daher gegeben durch
$$u(t) = C_1\mathrm{e}^{\mathring{\mathrm{i}}t} + C_2\mathrm{e}^{-\mathring{\mathrm{i}}t} = (C_1+C_2)\cos t + (C_1-C_2)\mathring{\mathrm{i}}\sin t\,.$$

(3) Zur Gleichung
$$u''' - 3u' - 2u \equiv 0$$
gehört das charakteristische Polynom $\lambda^3-3\lambda-2 = (\lambda+1)^2(\lambda-2)$ mit doppelter Nullstelle -1 und einfacher Nullstelle 2. Die allgemeine Lösung ist daher gegeben durch
$$u(t) = C_1e^{-t} + C_2t\mathrm{e}^{-t} + C_3\mathrm{e}^{2t}\,.$$

Die *inhomogene* Gleichung (5.3) lässt sich für manche Inhomogenitäten $b \not\equiv 0$ durch einen speziellen, auf die gegebene Inhomogenität zugeschnittenen Ansatz lösen. Im (wahrscheinlich) wichtigsten Fall handelt es sich auch hierbei um einen Exponentialansatz, und die entsprechende Regel folgt:

Satz 5.3 (zur **Lösung *inhomogener* linearer GDGen mit konstanten Koeffizienten**)**.** *Seien $m \in \mathbb{N}$ und $a_0, a_1, a_2, \dots, a_{m-1}, a_m \in \mathbb{K}$ mit $a_m \neq 0$, und sei*
$$p(\lambda) := a_m\lambda^m + a_{m-1}\lambda^{m-1} + \ldots + a_2\lambda^2 + a_1\lambda + a_0\,.$$
das charakteristische Polynom der Gleichung (5.3). Hat die Inhomogenität b in (5.3) die spezielle Form
$$b(t) = t^\ell\mathrm{e}^{\zeta t} \qquad \text{für } t \in I$$
mit $\ell \in \mathbb{N}_0$ und $\zeta \in \mathbb{K}$, so besitzt die inhomogene Gleichung (5.3) eine spezielle Lösung u_{sp} der Form
$$u_{\mathrm{sp}}(t) = \big[c_\ell t^\ell + c_{\ell-1}t^{\ell-1} + \ldots + c_1t + c_0\big]\mathrm{e}^{\zeta t}\,, \qquad \text{falls } \zeta \text{ keine Nullstelle von } p \text{ ist,}$$
und
$$u_{\mathrm{sp}}(t) = \big[c_\ell t^\ell + c_{\ell-1}t^{\ell-1} + \ldots + c_1t + c_0\big]t^d\mathrm{e}^{\zeta t}\,, \qquad \text{falls } \zeta \text{ eine } d\text{-fache Nullstelle von } p \text{ ist,}$$
jeweils mit Konstanten $c_0, c_1, \dots, c_{\ell-1}, c_\ell \in \mathbb{K}$ und im zweiten Fall mit Vielfachheit $d \in \mathbb{N}$.

Bemerkungen.

(1) Zu Satz 5.3 gehört folgendes naheliegende **Rechenverfahren** (bei Vorliegen der obigen Voraussetzungen): Man berechnet zuerst die Nullstellen des charakteristischen Polynoms p und erhält gemäß dem früheren Satz 5.2 die allgemeine Lösung der zugehörigen homogenen GDG. Als Nächstes setzt man den jeweiligen Ansatz für u_{sp} in die inhomogene GDG (5.3) ein und bestimmt die Konstanten $c_0, c_1, \ldots, c_{\ell-1}, c_\ell$ durch Koeffizientenvergleich mit der Inhomogenität b. Man erhält dann ein konkretes u_{sp}, und die **allgemeine Lösung der inhomogenen GDG** ergibt sich als Summe von u_{sp} und der allgemeinen Lösung des homogenen GDG; vergleiche mit Teil (III) des späteren Satzes 7.1.

(2) Wegen der Linearität der GDG (5.3) ergeben sich auch Ansätze für Inhomogenitäten der Form $b(t) = \sum_{i=1}^{h} \gamma_i t^{\ell_i} \mathrm{e}^{\zeta_i t}$ mit $\gamma_i \in \mathbb{K}$. Damit können **endliche Summen und Produkte von Polynom- und Exponentialfunktionen als Inhomogenitäten** behandelt werden, und insbesondere werden auch Terme der Bauart $\cosh(\nu t)$, $\sinh(\nu t)$, $\cos(\nu t)$ oder $\sin(\nu t)$ mit $\nu \in \mathbb{K}$ erfasst.

Anwendung (auf **Schwingungsgleichungen**).

(1) Dia allgemeine Lösung der **homogenen Schwingungsgleichung** mit reellem Parameter $\omega > 0$

$$u'' + \omega^2 u \equiv 0 \qquad \text{auf } \mathbb{R}$$

ist durch $u(t) = C_1 \mathrm{e}^{\mathring{\mathrm{i}}\omega t} + C_2 \mathrm{e}^{-\mathring{\mathrm{i}}\omega t} = (C_1 + C_2)\cos(\omega t) + (C_1 - C_2)\mathring{\mathrm{i}}\sin(\omega t)$ gegeben. Diese Gleichung gehört zum (idealisierten) physikalischen Modell des sogenannten **harmonischen Oszillators**, und die Lösungen beschreiben freie Schwingungen des Oszillators, ohne Einfluss zusätzlicher äußerer Kräfte, mit Kreisfrequenz ω, Periodendauer $\frac{2\pi}{\omega}$ und Frequenz $\frac{\omega}{2\pi}$. Letztere Frequenz heißt auch die **Eigenfrequenz** des Oszillators.

(2) Die **inhomogene Schwingungsgleichung** oder Gleichung für erzwungene Schwingungen

$$u'' + \omega^2 u = b \qquad \text{auf } \mathbb{R}$$

entspricht einer Anregung des Oszillators durch eine zeitabhängige äußere Kraft, speziell im Fall[1] $b(t) = \mathrm{e}^{\mathring{\mathrm{i}}\zeta t}$ mit $\zeta > 0$ durch eine Schwingung der äußeren Frequenz $\frac{\zeta}{2\pi}$. Im Fall $\zeta \neq \omega$ bekommt man gemäß Satz 5.3 die Lösungen $u(t) = \frac{1}{\omega^2 - \zeta^2}\mathrm{e}^{\mathring{\mathrm{i}}\zeta t} + C_1 \mathrm{e}^{\mathring{\mathrm{i}}\omega t} + C_2 \mathrm{e}^{-\mathring{\mathrm{i}}\omega t}$. Es handelt sich also um Überlagerungen von harmonischen Schwingungen der Frequenzen $\frac{\omega}{2\pi}$ und $\frac{\zeta}{2\pi}$, und insbesondere bleibt mit $|u(t)|$ die Amplitude der Schwingungen bei $t \to \infty$ beschränkt. Im Fall $\zeta = \omega$, dass die äußere Frequenz mit der Eigenfrequenz übereinstimmt, sieht es anders aus: Für die Lösungen $u(t) = -\frac{\mathring{\mathrm{i}}}{2\omega} t \mathrm{e}^{\mathring{\mathrm{i}}\omega t} + C_1 \mathrm{e}^{\mathring{\mathrm{i}}\omega t} + C_2 \mathrm{e}^{-\mathring{\mathrm{i}}\omega t}$ ist dann $|u(t)|$ bei $t \to \infty$ nicht beschränkt, die Amplitude der Schwingung wird also im Laufe der Zeit beliebig groß. Man spricht daher bei der Anregung eines Oszillators mit der Eigenfrequenz auch von einer **Resonanz** und nennt den Effekt, dass sich Schwingungen dann beliebig stark, eventuell bis zur Zerstörung des Oszillators, ‚aufschaukeln' können, eine **Resonanzkatastrophe**.

[1] Physikalisch relevant ist nur der Realteil der hier betrachteten $\mathbb{C}$-wertigen Funktionen. Für die Phänomenologie der GDG macht dies aber keinen Unterschied, daher wird dieser Aspekt hier ausgeklammert.

(3) Ähnlich kann man auch die (eventuell inhomogene) **Gleichung für gedämpfte Schwingungen**

$$u'' + du' + \omega^2 u = b \qquad \text{auf } \mathbb{R}$$

mit einer Dämpfungskonstante $d \in \mathbb{R}$ behandeln. Das Lösungsverhalten hängt dann vom Vorzeichen der Diskriminante der charakteristischen Gleichung $\lambda^2+d\lambda+\omega^2 = 0$ ab, eine Diskussion der relevanten Fälle leitet aber teils in die Physik über und wird hier nicht durchgeführt.

Aufbauend auf später in der Vorlesung behandelter Theorie kann Satz 5.3 (mehr oder weniger) kurz und schematisch hergeleitet werden. An dieser Stelle wird stattdessen ein elementarer, aber technisch etwas aufwendigerer Beweis gegeben:

Beweis von Satz 5.3. Zuerst wird der Fall $p(\zeta) \neq 0$, dass ζ keine Nullstelle von p ist, behandelt. Die Argumentation basiert dann auf der Beobachtung, dass es zu $i \in \mathbb{N}$ stets ein Polynom $\widetilde{p}_{i,\zeta}$ vom Grad $\leq (i-1)$ mit

$$p\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big)t^i\mathrm{e}^{\zeta t} = t^i p\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big)\mathrm{e}^{\zeta t} + \widetilde{p}_{i,\zeta}(t)\mathrm{e}^{\zeta t} = \big[t^i p(\zeta) + \widetilde{p}_{i,\zeta}(t)\big]\mathrm{e}^{\zeta t} \qquad \text{für alle } t \in \mathbb{R}$$

gibt; dies folgt direkt aus der Produkt- beziehungsweise Leibniz-Regel, denn sobald eine einzige bei der Berechnung von $p\big(\frac{\mathrm{d}}{\mathrm{d}t}\big)$ auftretende Ableitung auf den Faktor t^i und nicht auf den Faktor $\mathrm{e}^{\zeta t}$ angewandt wird, verringert sich der Grad des Monoms t^i. Als Nächstes wird mit Induktion nach $\ell \in \mathbb{N}_0$ folgende Aussage gezeigt, die die Behauptung des Satzes im Fall $p(\zeta) \neq 0$ beinhaltet:

> Zu jedem Polynom q der Ordnung $\leq \ell$ existiert eine spezielle Lösung u_{sp} der ersten im Satz angegebenen Form zu
> $$p\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big)u(t) = q(t)\mathrm{e}^{\zeta t}\,.$$

Beim Induktionsanfang für $\ell = 0$ ist dabei $q \equiv q_0 \in \mathbb{K}$ konstant, und es reicht, $c_0 = q_0 p(\zeta)^{-1}$ zu wählen. Dann ist nämlich $p\big(\frac{\mathrm{d}}{\mathrm{d}t}\big)u_{\mathrm{sp}}(t) = c_0 p\big(\frac{\mathrm{d}}{\mathrm{d}t}\big)\mathrm{e}^{\zeta t} = c_0 p(\zeta)\mathrm{e}^{\zeta t} = q_0\mathrm{e}^{\zeta t}$. Für den Induktionsschluss von $\ell-1$ auf $\ell \in \mathbb{N}$ berechnet man zuerst

$$\begin{aligned} p\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big)u_{\mathrm{sp}}(t) &= c_\ell p\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big)t^\ell\mathrm{e}^{\zeta t} + p\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big)\big[c_{\ell-1}t^{\ell-1}+\ldots+c_1 t+c_0\big]\mathrm{e}^{\zeta t} \\ &= c_\ell\big[t^\ell p(\zeta) + \widetilde{p}_{\ell,\zeta}(t)\big]\mathrm{e}^{\zeta t} + p\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big)\big[c_{\ell-1}t^{\ell-1}+\ldots+c_1 t+c_0\big]\mathrm{e}^{\zeta t}\,. \end{aligned}$$

Jetzt bezeichne q_ℓ das Monom der Ordnung ℓ in q. Gemäß Induktionsannahme, angewandt auf das Polynom $q-q_\ell-c_\ell\widetilde{p}_{\ell,\zeta}$ vom Grad $\leq (\ell-1)$, gibt es dann $c_0, c_1, \ldots, c_{\ell-1} \in \mathbb{K}$, so dass der letzte Term der vorausgehenden Rechnung genau $\big[q(t)-q_\ell(t)-c_\ell\widetilde{p}_{\ell,\zeta}(t)\big]\mathrm{e}^{\zeta t}$ ergibt. Somit folgt

$$p\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big)u_{\mathrm{sp}}(t) = c_\ell t^\ell p(\zeta)\mathrm{e}^{\zeta t} + \big[q(t)-q_\ell(t)\big]\mathrm{e}^{\zeta t} = q(t)\mathrm{e}^{\zeta t}\,,$$

wenn noch $c_\ell t^\ell p(\zeta) = q_\ell(t)$ durch die Wahl $c_\ell = q_\ell(1)p(\zeta)^{-1}$ erreicht wird. Damit sind die Induktion und der Beweis im Fall $p(\zeta) \neq 0$ komplett.

Im anderen Fall, dass ζ Nullstelle von p der Vielfachheit $d \in \mathbb{N}$ ist, kann man d Linearfaktoren $(\lambda-\zeta)$ von $p(\lambda)$ abspalten und bekommt $p(\lambda) = (\lambda-\zeta)^d\widetilde{p}(\lambda)$ mit einem Polynom $\widetilde{p}$, das bei ζ keine

Nullstelle mehr besitzt. Nach dem bereits Gezeigten gibt es dann $c_{-d}, c_{-d+1}, \dots, c_{\ell-1}, c_\ell \in \mathbb{K}$, so dass

$$\widetilde{p}\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big)\big[c_\ell t^{\ell+d}+c_{\ell-1}t^{\ell+d-1}+\ldots+c_{-d+1}t+c_{-d}\big]\mathrm{e}^{\zeta t} = \frac{\ell!}{(\ell+d)!}t^{\ell+d}\mathrm{e}^{\zeta t}$$

gilt. Durch Anwendung von $\big(\frac{\mathrm{d}}{\mathrm{d}t}-\zeta\big)^d$ auf beide Seiten dieser Gleichung folgt

$$p\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big)\big[c_\ell t^{d+\ell}+c_{\ell-1}t^{d+\ell-1}+\ldots+c_{-d+1}t+c_{-d}\big]\mathrm{e}^{\zeta t} = \frac{\ell!}{(\ell+d)!}\Big(\frac{\mathrm{d}}{\mathrm{d}t}-\zeta\Big)^d t^{\ell+d}\mathrm{e}^{\zeta t} = t^\ell \mathrm{e}^{\zeta t}\,, \tag{5.6}$$

wobei rechts d-mal mittels $\big(\frac{\mathrm{d}}{\mathrm{d}t}-\zeta\big)t^i\mathrm{e}^{\zeta t} = it^{i-1}\mathrm{e}^{\zeta t}$ differenziert wurde. Da es sich gemäß Satz 5.2 bei $t^{d-1}\mathrm{e}^{\zeta t}$, …, $t^2\mathrm{e}^{\zeta t}$, $t\mathrm{e}^{\zeta t}$, $\mathrm{e}^{\zeta t}$ um Lösungen der homogenen Gleichung $p\big(\frac{\mathrm{d}}{\mathrm{d}t}\big)u \equiv 0$ handelt, fallen in (5.6) die Summanden $c_{-1}t^{d-1}$, …, $c_{-d+2}t^2$, $c_{-d+1}t$, c_{-d} weg, und nach Ausklammern von t^d folgt mit

$$p\Big(\frac{\mathrm{d}}{\mathrm{d}t}\Big)\big[c_\ell t^{\ell}+c_{\ell-1}t^{\ell-1}+\ldots+c_1 t+c_0\big]t^d\mathrm{e}^{\zeta t} = t^\ell \mathrm{e}^{\zeta t}$$

die Behauptung. □

5.3 Separation der Variablen

Im nächsten Satz wird ein weiterer Typ von explizit lösbaren, skalaren Gleichungen erster Ordnung beschrieben. Diesmal erhält man **nur $\mathbb{R}$-wertige Lösungen**, im Gegensatz zu den vorherigen Abschnitten werden aber erstmals auch **manche nicht-linearen Gleichungen erfasst**. Die grundlegende Herangehensweise kann als Satz wie folgt formuliert werden:

Satz 5.4 (über **GDGen mit separierten Variablen**). *Gegeben sei eine GDG*

$$g(u)u' = h \tag{5.7}$$

mit Strukturfunktionen $g \in \mathrm{C}^0(J,\mathbb{R})$ und $h \in \mathrm{C}^0(I,\mathbb{R})$ auf Intervallen I und J positiver Länge in $\mathbb{R}$. Außerdem seien Stammfunktionen G zu g auf J und H zu h auf I gegeben.

(I) *Dann sind die $\mathbb{R}$-wertigen Lösungen von (5.7) auf I genau die differenzierbaren Funktionen $u\colon I \to J$ mit $G(u) = H+C$ für ein $C \in \mathbb{R}$.*

(II) *Hat g keine Nullstelle in J, so existiert eine auf $G(J)$ definierte, differenzierbare Umkehrfunktion G^{-1} zu G, und die allgemeine $\mathbb{R}$-wertige Lösung zu (5.7) auf I ist gegeben durch*

$$u(t) = G^{-1}(H(t)+C) \qquad \text{für } t \in I \tag{5.8}$$

mit $C \in \mathbb{R}$, so dass $H(I)+C \subset G(J)$.

Beweis. Gemäß Kettenregel ist $[G(u)]' = g(u)u'$, daher ist (5.7) durch Stammfunktionsbildung äquivalent zu $G(u) = H+C$ — wie in (I) behauptet. Hat g keine Nullstelle, so ist G strikt monoton mit Ableitung $G' = g \neq 0$, der Umkehrsatz liefert Existenz und Differenzierbarkeit von G^{-1}, und Anwendung von G^{-1} auf beiden Seiten der Gleichung aus (I) ergibt (5.8). □

Bezeichnungen. Man kann auch $g\circ u$ statt $g(u)$ und $G\circ u$ statt $G(u)$ schreiben, die Notation $G(J) := \{G(x) : x \in J\}$ steht für das Bild von J unter G, und $H(I)$ ist analog zu verstehen.

Bemerkungen.

(1) Wenn g zwar Nullstellen, aber keinen Vorzeichenwechsel hat und zumindest auf keinem Teilintervall positiver Länge verschwindet, so ist G noch strikt monoton und G^{-1} existiert als stetige, aber nicht überall differenzierbare Funktion. In diesem Fall bekommt man durch (5.8) eine stetige Funktion u, die auf Intervallen mit $g(u) \neq 0$ löst, aber an den Nullstellen von $g(u)$ unendliche Steigung und somit Nichtdifferenzierbarkeitsstellen aufweisen kann.

(2) Natürlich lässt sich der **Satz auch anwenden, wenn die Form** (5.7) nicht direkt vorliegt, aber **durch Umformungen hergestellt werden kann**. Beim Herstellen der geeigneten Form handelt es sich um die eigentliche **Separation der Variablen u** (die auf die linke Seite gebracht wird) **und t** (die auf die rechte Seite gebracht wird). Aufpassen und eventuell **Fallunterscheidungen durchführen** muss man **bei Umformungen, die, genau genommen, keine Äquivalenzumformungen sind** — wie bei Division durch einen Term, der möglicherweise Null werden kann, oder beim Wurzelziehen; siehe die folgende Beispiele.

Beispiele (zur **Separation der Variablen**). Sei I Intervall positiver Länge, u stets $\mathbb{R}$-wertig.

(1) Die GDG mit separierten Variablen

$$u^2 u' = \mathrm{e}^{-t}$$

hat nach dem Satz und Bemerkung (1) (mit $J = \mathbb{R}$, $g(x) = x^2$, $G(x) = \frac{1}{3}x^3$, $G(J) = \mathbb{R}$, $G^{-1}(s) = \sqrt[3]{3s}$, $h(t) = \mathrm{e}^{-t}$, $H(t) = -\mathrm{e}^{-t}$) auf I die allgemeine Lösung[2]

$$u(t) = \sqrt[3]{3(C - \mathrm{e}^{-t})}\,.$$

mit $C \in \mathbb{R}$, so dass $C - \mathrm{e}^{-t} \neq 0$ für $t \in I$; und falls es doch ein (hier stets eindeutiges) $t \in I$ mit $C - \mathrm{e}^{-t} = 0$ gibt, so ist dieses Nichtdifferenzierbarkeitsstelle von u nach Bemerkung (1).

Die eindeutige Lösung, eventuell mit Nichtdifferenzierbarkeitsstelle, zu einer beliebigen AB $u(t_0) = y_0$ ergibt sich durch Einsetzen als

$$u(t) = \sqrt[3]{y_0^3 + 3(\mathrm{e}^{-t_0} - \mathrm{e}^{-t})}\,.$$

(2) Die GDG mit separierten Variablen

$$(1+u^2)u' = \mathrm{e}^{2t} + \tfrac{1}{4}\mathrm{e}^{6t}$$

hat nach dem Satz (mit $J = \mathbb{R}$, $g(x) = 1+x^2$, $G(x) = x + \frac{1}{3}x^3$, $G(J) = \mathbb{R}$, $h(t) = \mathrm{e}^{2t} + \frac{1}{4}\mathrm{e}^{6t}$, $H(t) = \frac{1}{2}\mathrm{e}^{2t} + \frac{1}{24}\mathrm{e}^{6t}$) die allgemeine Lösung

$$u(t) = G^{-1}\big(\tfrac{1}{2}\mathrm{e}^{2t} + \tfrac{1}{24}\mathrm{e}^{6t} + C\big)$$

mit $C \in \mathbb{R}$. Die differenzierbare Umkehrfunktion G^{-1} existiert hier nach dem Umkehrsatz global auf $\mathbb{R}$, lässt sich jedoch nicht ohne Weiteres[3] ausrechnen. Zumindest eine Lösung, nämlich die zu $C = 0$, kann man durch $u(t) = G^{-1}\big(G\big(\frac{1}{2}\mathrm{e}^{2t}\big)\big) = \frac{1}{2}\mathrm{e}^{2t}$ leicht explizit angeben, aber dies geht nur wegen der sehr glücklichen Form der rechten Seite.

Bei dieser Gleichung folgt, dass beliebige AWPe auf Intervallen stets eindeutig lösbar sind.

[2] Hier steht $\sqrt[3]{s}$ für die eindeutige reelle Kubikwurzel mit $\big(\sqrt[3]{s}\big)^3 = s$, die aus jeder reellen Zahl s, insbesondere auch aus negativem s, gezogen werden kann.

[3] Tatsächlich liefern die Cardanischen Formeln zur Lösung reduzierter kubischer Gleichungen die Wurzeldarstellung $G^{-1}(s) = \sqrt[3]{\frac{3s}{2} + \big(1 + \frac{9s^2}{4}\big)^{1/2}} + \sqrt[3]{\frac{3s}{2} - \big(1 + \frac{9s^2}{4}\big)^{1/2}}$ von G^{-1}, aber diese Formel liefert kaum Informationen, die man nicht auch ohne sie erhalten kann. Bei einem leicht modifizierten Beispiel mit linker Seite $(1+u^4)u'$ statt $(1+u^2)u'$ verhält sich alles analog, aber dann besteht wirklich keine Chance mehr, G^{-1} explizit auszurechnen.

(3) Bei der Gleichung

$$2uu' = t^2$$

lässt sich Teil (I) des Satzes (mit $J = \mathbb{R}$, $g(x) = 2x$, $G(x) = x^2$, $h(t) = t^2$, $H(t) = \frac{1}{3}t^3$) ebenfalls anwenden, und Lösungen u sind charakterisiert durch

$$u(t)^2 = \tfrac{1}{3}t^3 + C$$

mit $C \in \mathbb{R}$. Auflösen nach u ergibt die positiven und die negativen Lösungen

$$u(t) = \sqrt{\tfrac{1}{3}t^3 + C} \qquad \text{und} \qquad u(t) = -\sqrt{\tfrac{1}{3}t^3 + C}\,,$$

jeweils auf Teilintervallen I von $(-\sqrt[3]{3C}, \infty)$. Die „Verdopplung“ der Lösungen rührt hier daher, dass g bei Null einen Vorzeichenwechsel hat und G nicht global auf $\mathbb{R}$ invertierbar ist; somit können Teil (II) des Satzes und Bemerkung (1) zwar mit $J = (-\infty, 0]$, $G(J) = [0, \infty)$ und mit $J = [0, \infty)$, $G(J) = [0, \infty)$, aber eben nicht global mit $J = \mathbb{R}$ angewandt werden.

Durch Einsetzen findet man wieder Lösungen zur AB $u(t_0) = y_0$, nämlich

$$u(t) = \sqrt{\tfrac{1}{3}(t^3 - t_0^3) + y_0^2} \qquad \text{im Fall } y_0 \geq 0\,,$$
$$u(t) = -\sqrt{\tfrac{1}{3}(t^3 - t_0^3) + y_0^2} \qquad \text{im Fall } y_0 \leq 0\,.$$

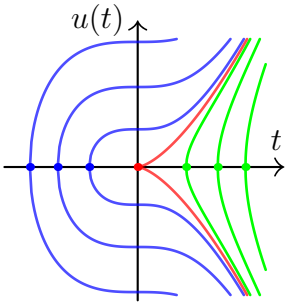


Abb. 5: Die Lösungen des Beispiels (3) mit $C < 0$, $C = 0$, $C > 0$.

Insbesondere gibt es für $y_0 = 0$ zwei verschiedene Lösungen zur selben Anfangsbedingung. Für $y_0 = 0 \neq t_0$ liegt aber Nichtdifferenzierbarkeit in t_0 selbst vor, und die Lösungseigenschaft der Definition 4.1 ist nur auf Teilintervallen von (t_0, ∞), also nicht bis hin zu t_0 selbst gegeben. Für $y_0 = 0 = t_0$ allerdings sind die beiden Lösungen $u(t) = \pm\frac{1}{\sqrt{3}}t^{3/2}$ bei Null differenzierbar, und dies ist ein echtes **Beispiel für nicht-eindeutige Lösbarkeit eines AWPs** auf $\mathbb{R}_0^+$. Ein ähnliches Beispiel für Nicht-Eindeutigkeit, bei dem t_0 sogar *im Innern* des Lösungsintervalls liegt, wird in den Übungen behandelt.

(4) Bei

$$u' = u(u-1)$$

muss man direkt bei der ersten Umformung aufpassen, denn Division durch $u(u-1)$ ist nur erlaubt, wenn u die Werte $\{0, 1\}$ nicht annimmt. Behält man die (eventuell auf Teilintervallen) konstanten Lösungen $u \equiv 0$ und $u \equiv 1$ aber im Kopf, so kann man für die weitere Rechnung erst einmal annehmen, dass u Werte in $\mathbb{R} \setminus \{0, 1\}$ hat, und die Division durchführen. Man erhält dann die GDG mit separierten Variablen $\frac{1}{u(u-1)}u' \equiv 1$. Die Lösungen u auf I sind dann nach dem Satz (mit $g(x) = \frac{1}{x(x-1)} = \frac{1}{x-1} - \frac{1}{x}$, $J \in \{(-\infty, 0), (0, 1), (1, \infty)\}$, $G(x) = \log|x-1| - \log|x| = \log\left|1 - \frac{1}{x}\right|$, $h(t) = 1$, $H(t) = t$) bestimmt durch

$$\log\left|1 - \frac{1}{u(t)}\right| = t + C\,,$$

und Auflösen nach u gibt die Lösungen

$$u(t) = \left(1 \pm \mathrm{e}^C \mathrm{e}^t\right)^{-1},$$

sofern $1 \pm \mathrm{e}^{C}\mathrm{e}^{t} \neq 0$ für alle $t \in I$ gilt. Dies lässt sich noch vereinfachend umschreiben, und zusammenfassend besteht die allgemeine Lösung der ursprünglichen GDG auf I aus der Null-Lösung und den Lösungen

$$u(t) = \left(1 + \widetilde{C}\mathrm{e}^{t}\right)^{-1}$$

mit $\widetilde{C} \in \mathbb{R}$, so dass $1 + \widetilde{C}\mathrm{e}^{t} \neq 0$ für alle $t \in I$ gilt (die Lösung $u \equiv 1$ ist für $\widetilde{C} = 0$ enthalten). Die Lösung zur AB $u(t_0) = 0$ ist hier stets $u \equiv 0$, und die Lösung zur AB $u(t_0) = y_0 \neq 0$ existiert genau dann und ist gegeben durch

$$u(t) = \left(1 + \frac{1 - y_0}{y_0}\mathrm{e}^{t - t_0}\right)^{-1},$$

wenn der Ausdruck in der Klammer für $t \in I$ nie Null wird.

Insgesamt zeigen die vorausgehenden Beispiele einige Effekte (Entwicklung von Singularitäten führt zu Lösbarkeit nur auf Teilintervallen, Nicht-Eindeutigkeit, unendliche Steigung), mit denen man bei nicht-linearen Gleichungen rechnen muss — teils aber nur dann, wenn sie nicht auf eine „gute“ explizite Form gebracht werden können; mehr dazu in Kapitel 6.

5.4 Exakte Differentialgleichungen

Wie im vorigen Abschnitt geht es auch hier um eine Klasse von skalaren, möglicherweise nicht-linearen Gleichungen erster Ordnung und um die Bestimmung der $\mathbb{R}$-wertigen Lösungen. Die Herangehensweise ist allerdings deutlich anders und erlaubt die Behandlung einer größeren Klasse von Gleichungen. Die grundlegende Beobachtung sei auch hier als Satz festgehalten:

Satz 5.5 (über **exakte Differentialgleichungen**). *Gegeben sei eine GDG*

$$f(t,u) + g(t,u)u' \equiv 0 \tag{5.9}$$

mit Strukturfunktionen $f, g \in \mathrm{C}^0(D, \mathbb{R})$ auf einem Gebiet[4] *$D \subset \mathbb{R}^2$. Sind*

$$f = \frac{\partial \Psi}{\partial t} \qquad \text{und} \qquad g = \frac{\partial \Psi}{\partial x} \qquad \text{auf } D \tag{5.10}$$

die partiellen Ableitungen einer Funktion $\Psi \in \mathrm{C}^1(D, \mathbb{R})$ der Variablen (t, x), so sind die $\mathbb{R}$-wertigen Lösungen von (5.9) auf jedem Intervall I positiver Länge genau die differenzierbaren $u\colon I \to \mathbb{R}$ mit $(t, u(t)) \in D$ und

$$\Psi(t, u(t)) = C \qquad \text{für alle } t \in I\,, \tag{5.11}$$

mit einer Konstanten $\mathrm{C} \in \mathbb{R}$.

Beweis. Gemäß Kettenregel und (5.10) gilt für differenzierbares u auf I und $t \in I$ mit $(t, u(t)) \in D$ stets

$$\frac{\mathrm{d}}{\mathrm{d}t}\Psi(t, u(t)) = \frac{\partial \Psi}{\partial t}(t, u(t)) + \frac{\partial \Psi}{\partial x}(t, u(t))u'(t) = f(t, u(t)) + g(t, u(t))u'(t)\,.$$

Somit sind Lösungen u von (5.9) auf I charakterisiert durch $(t, u(t)) \in D$ und $\frac{\mathrm{d}}{\mathrm{d}t}\Psi(t, u(t)) = 0$ für alle $t \in I$. Mit dem Konstanzsatz gelangt man dann zur Charakterisierung vermöge (5.11). □

[4] Ein Gebiet ist eine nicht-leere, offene und zusammenhängende Menge.

Bemerkungen.

(1) Bei (5.11) handelt es sich um eine *implizite* Gleichung für die Lösung u. Um daraus eine *explizite* Formel zu gewinnen, versucht man bei konkreten Rechnungen meist nach u aufzulösen (was manchmal, doch nicht immer möglich ist; dazu macht der Satz keine Aussage).

(2) Die **entscheidende Voraussetzung** des Satzes **ist die Existenz der Funktion Ψ** mit (5.10). Die Gleichungen in (5.10) bedeuten dabei, dass das Vektorfeld (f, g) auf D das Gradientenfeld von Ψ ist, also $(f, g) = \nabla\Psi$ auf D. Sind nur f und g gegeben, so muss man zur Anwendung des Satzes also zunächst eine **Stammfunktion** (oder ein **Potential**) **Ψ zum Vektorfeld (f, g) finden**.

(3) Die **Terminologie ‚exakte' DGL kommt aus der Theorie der Differentialformen**. Man kann das Vektorfeld (f, g) nämlich mit der 1-Form $f(t, x)\mathrm{d}t+g(t, x)\mathrm{d}x$ auf D identifizieren und das Problem der Stammfunktionsfindung äquivalent für diese 1-Form formulieren. In Anlehnung daran, dass eine 1-Form exakt genannt wird, wenn sie eine Stammfunktion besitzt, **nennt man die GDG** (5.9) **exakt, wenn (f, g) eine Stammfunktion besitzt**.

(4) **Stammfunktionen zu Vektorfeldern** (oder 1-Formen) **existieren nicht immer, sondern nur manchmal**. Notwendiges Kriterium für die Existenz einer Stammfunktion zu (f, g) (und damit für die Exaktheit der GDG (5.9)) ist die **Integrabilitätsbedingung**

$$\frac{\partial f}{\partial x} - \frac{\partial g}{\partial t} \equiv 0 \qquad \text{auf } D\,,$$

und im Wesentlichen ist dieses Kriterium sogar notwendig und hinreichend: Genau genommen ist die Integrabilitätsbedingung zwar nur bei einem Gebiet D „ohne Löcher"[5] hinreichend für die Existenz einer Stammfunktion auf D, aber auf geeigneten Teilgebieten von D lässt sich dies stets anwenden und reicht damit für die Rechenpraxis völlig aus. Die **Berechnung einer Stammfunktion Ψ** ist Thema einer Analysis-Vorlesung und erfolgt im Wesentlichen **durch Integration von f nach t und/oder von g nach der u-Variablen x** sowie durch geeignete Wahl der Integrationskonstanten (letztere Wahl wird in aller Regel beim Ableiten zur Probe offensichtlich).

(5) Ist bei einer GDG der Form (5.9) die **Integrabilitätsbedingung nicht erfüllt** und die Gleichung somit nicht exakt, so kann man versuchen, durch Multiplikation mit einem sogenannten **integrierenden Faktor** $h \in \mathrm{C}^0(D, \mathbb{R}^+)$ auf eine äquivalente exakte DGL

$$h(t, u)f(t, u) + h(t, u)g(t, u)u' \equiv 0$$

zu transformieren. Manchmal lässt sich h raten, andernfalls kann man versuchen, in der Integrabilitätsbedingung $\frac{\partial(hf)}{\partial x} - \frac{\partial(hg)}{\partial t} \equiv 0$ einen Ansatz wie $h(t, x) = \varphi(t)$, $h(t, x) = \varphi(x)$, $h(t, x) = \varphi(t+x)$ oder $h(t, x) = \varphi(tx)$ zu machen; tritt dabei nur dieselbe Variable t, x, $t+x$ oder tx auf, von der auch φ selbst abhängt, so kann man φ als Lösung einer linearen Erster-Ordnung-GDG erhalten. Sind φ und ein integrierender Faktor h bestimmt, so lässt sich die neue exakte DGL (hoffentlich) gemäß dem Satz und den vorigen Bemerkungen lösen.

[5] Die präzise Voraussetzung an D ist hier die, dass D einfach zusammenhängend im Sinne der Homotopietheorie ist. Im relevanten 2-dimensionalen Fall bedeutet dies, dass man keinen Punkt von $\mathbb{R}^2 \setminus D$ durch einen Weg in D vollständig umlaufen kann. Spezielle Klassen von einfach zusammenhängenden Gebieten sind konvexe Gebiete und allgemeiner sternförmige Gebiete D (letzteres bedeutet, dass ein $\omega_0 \in D$ existiert, so dass für alle $\omega \in D$ auch die Strecke von ω zum Sternpunkt ω_0 in D enthalten ist).

(6) Die (umgeformte) **GDG mit separierten Variablen** $-h(t)+g(u)u' \equiv 0$ ist offensichtlich exakt mit Potential $\Psi(t,x) = -H(t)+G(x)$ (wobei $G'=g$, $H'=h$). Die **lineare Erster-Ordnung-GDG** $-a(t)u-b(t)+u' \equiv 0$ wird durch Multiplikation mit dem integrierenden Faktor[6] $\mathrm{e}^{-A(t)}$ exakt und hat dann die Potentialfunktion $\Psi(t,x) = \mathrm{e}^{-A(t)}x-B(t)$ (wobei $A'=a$, $B'=\mathrm{e}^{-A}b$). Somit stellen sich die Lösungsformeln der Sätze 5.1 und 5.4 als **Spezialfälle der Behandlung exakter DGLen** heraus.

Beispiele (zu **exakten DGLen**).

(1) Die DGL

$$1+\frac{u}{t^2} - \frac{1}{t}u' \equiv 0$$

hat die Form aus Satz 5.5 mit $D = (-\infty,0)\times\mathbb{R}$ oder $D = (0,\infty)\times\mathbb{R}$, mit $f(t,x) = 1+\frac{x}{t^2}$ und $g(t,x) = -\frac{1}{t}$. Tatsächlich kann man eine Stammfunktion Ψ zu (f,g) hier noch raten, nämlich

$$\Psi(t,x) = t-\frac{x}{t}\,.$$

Man kann die Berechnung von Ψ auch schematisch angehen, indem man erst die Exaktheit der DGL durch die Rechnung $\frac{\partial f}{\partial x}(t,x)-\frac{\partial g}{\partial t}(t,x) = \frac{1}{t^2}-\frac{1}{t^2} = 0$ nachprüft und dann zur Bestimmung von Ψ integriert:

$$\Psi(t,x) = \int f(t,x)\,\mathrm{d}t = \int \left(1+\frac{x}{t^2}\right)\mathrm{d}t = t-\frac{x}{t} + \mathrm{const}(x)\,,$$
$$\Psi(t,x) = \int g(t,x)\,\mathrm{d}x = \int \left(-\frac{1}{t}\right)\mathrm{d}x = -\frac{x}{t} + \mathrm{const}(t)\,.$$

Nun wird klar, wie die Konstanten zu wählen sind, und man kommt auf obiges Ψ als eine Stammfunktion (wofür es auch gereicht hätte, nur die obere Integration durchführen und dann zur Probe abzuleiten). Insgesamt sind die Lösungen der GDG nach Satz 5.5 bestimmt durch

$$t - \frac{u(t)}{t} = C\,,$$

und durch Auflösen kommt man auf Teilintervallen von $(-\infty,0)$ und $(0,\infty)$ auf die allgemeine Lösung

$$u(t) = t^2-Ct \qquad \text{mit } C\in\mathbb{R}\,.$$

(2) Die DGL

$$2t+t^2+u + (1+t^2+u)u' \equiv 0$$

hat die Form aus Satz 5.5 mit $D = \mathbb{R}^2$, $f(t,x) = 2t+t^2+x$, $g(t,x) = 1+t^2+x$. Wegen $\frac{\partial f}{\partial x}(t,x)-\frac{\partial g}{\partial t}(t,x) = 1-2t \neq 0$ ist diese DGL aber nicht exakt, und (f,g) besitzt keine Stammfunktion. Zur Bestimmung eines integrierenden Faktors h macht man (eventuell erst nach anderen, vergeblichen Versuchen) den Ansatz

$$h(t,x) = \varphi(t+x)\,.$$

[6] Um diesen integrierenden Faktor gemäß Bemerkung (5) mit dem Ansatz $h(t,u) = \varphi(t)$ schematisch bestimmen zu können, muss man allerdings schon wissen, wie man *homogene* lineare Erster-Ordnung-GDGen löst. Deshalb ergibt sich hier eigentlich nur Teil (II) des Satzes 5.1 als Spezialfall, wenn man dessen Teil (I) schon kennt und benutzt.

Als Integrabilitätsbedingung für die neue DGL aus Bemerkung (5) erhält man mit Produkt- und Kettenregel

$$\begin{aligned} 0 &= \frac{\partial(hf)}{\partial x}(t,x) - \frac{\partial(hg)}{\partial t}(t,x) \\ &= \varphi'(t{+}x)f(t,x) + \varphi(t{+}x)\frac{\partial f}{\partial x}(t,x) - \varphi'(t{+}x)g(t,x) - \varphi(t{+}x)\frac{\partial g}{\partial t}(t,x) \\ &= (2t{-}1)\varphi'(t{+}x) + (1{-}2t)\varphi(t{+}x)\,. \end{aligned}$$

Wegen sehr günstiger Kürzungseffekte reduziert sich die Integrabilitätsbedingung daher auf die Erster-Ordnung-GDG $\varphi' = \varphi$ mit (spezieller) Lösung $\varphi(s) = \mathrm{e}^s$. Man erhält somit den integrierenden Faktor $h(t,x) = \mathrm{e}^{t+x}$ und die äquivalente exakte DGL

$$(2t{+}t^2{+}u)\mathrm{e}^{t+u} + (1{+}t^2{+}u)\mathrm{e}^{t+u}u' \equiv 0\,.$$

Die zugehörigen (neuen) Strukturfunktionen sind $f_*(t,x) = (2t{+}t^2{+}x)\mathrm{e}^{t+x}$ und $g_*(t,x) = (1{+}t^2{+}x)\mathrm{e}^{t+x}$, und das zugehörige Potential

$$\Psi_*(t,x) = (t^2{+}x)\mathrm{e}^{t+x}$$

lässt sich jetzt raten (oder wie in Beispiel (1) berechnen). Somit sind die Lösungen der neuen und auch der ursprünglichen DGL auf Intervallen I gemäß dem Satz bestimmt durch

$$(t^2{+}u(t))\mathrm{e}^{t+u(t)} = C \qquad \text{mit } C \in \mathbb{R}\,,$$

ein explizites Auflösen nach $u(t)$ ist aber diesmal nicht möglich.

5.5 Potenzreihenansatz

Ein allgemeiner Ansatz zur Lösung einer GDG beruht auf der (zu allermeist vernünftigen) Annahme, dass mindestens eine Lösung u auf einem Intervall I positiver Länge eine Darstellung

$$u(t) = \sum_{i=0}^{\infty} c_i(t{-}t_0)^i \qquad \text{für } t \in I \tag{5.12}$$

als **Potenzreihe mit unbekannten Koeffizienten** $c_i \in \mathbb{K}$ zum bekannten Entwicklungspunkt $t_0 \in \mathbb{R}$ besitzt (bei einem gegebenen AWP ist t_0 oft der Punkt, an dem auch die ABen gestellt werden). Zur Berechnung von Lösungen geht man dann wie folgt vor:

Verfahren (Lösung von GDGen mittels Potenzreihenansatz).

(I) *Man beginnt mit dem* ***Einsetzen des Ansatzes*** (5.12) *in die gegebene GDG,*

(II) *berechnet die Ableitungen durch* ***gliedweise Differentiation*** *und führt eventuell weitere Operationen (wie Produktbildungen) mit Potenzreihen durch.*

(III) *Dann macht man einen* ***Koeffizientenvergleich bei Potenzreihen*** *und kommt unter günstigen Umständen auf* ***Rekursionsformeln für die Koeffizienten*** c_i*.*

(IV) ***Manchmal*** *kann man die* ***Rekursion auflösen*** *und die Koeffizienten explizit bestimmen, und in seltenen Fällen stellt sich die Potenzreihe aus* (5.12) *sogar als bekannte Darstellung einer elementaren Funktion auf I heraus.*

(V) *Handelt es sich nicht gerade um eine bekannte Potenzreihe, so muss man schließlich die* ***Konvergenz der Reihe sicherstellen*** *(auf I oder zumindest auf einem Teilintervall positiver Länge), indem man den Konvergenzradius der Potenzreihe bestimmt oder abschätzt; dies kann beispielsweise mit der Euler-Formel für den Konvergenzradius oder durch Analyse des Wachstums von $|c_i|$ bei $i \to \infty$ geschehen.*

(VI) *Ist die Konvergenz gezeigt, so hat man eine* ***Lösung der GDG gefunden*** *— je nach Verlauf des Verfahrens in mehr oder weniger expliziter Form.*

Bemerkung. Die Diskussion der Konvergenz der Reihe **ist essentiell** und wird zur Rechtfertigung des gesamten Vorgehens benötigt: **Es besteht nämlich die Möglichkeit, dass das Verfahren scheitert, weil die berechnete Potenzreihe den Konvergenzradius 0 aufweist** und somit nur im Entwicklungspunkt t_0 konvergiert; siehe Beispiel (4) unten. Dann hat man Pech gehabt und keine Lösung gefunden. **Dagegen ist bei positivem Konvergenzradius r_0 das Vorgehen**, jedenfalls auf Teilintervallen des offenen Konvergenzbereichs (t_0-r_0, t_0+r_0), **berechtigt**. Insbesondere sind in letzterem Fall die gliedweise Differentiation und auch andere Operationen mit Potenzreihen, wie beispielsweise die Bildung des Cauchy-Produkts, erlaubt, und man erhält Lösungen in der zu Beginn angenommenen Form (5.12).

Beispiele (zum Lösen durch **Potenzreihenansatz** um den Entwicklungspunkt 0)**.**

(1) Die bekannte GDG

$$u'' = -u$$

wurde bereits in Kapitel 5.2 gelöst, wird jetzt aber trotzdem zur Illustration der beschriebenen Vorgehensweise herangezogen: Einsetzen des Ansatzes (5.12) mit $t_0 = 0$ und gliedweise Berechnung der zweiten Ableitung führt bei dieser Gleichung auf

$$\sum_{i=2}^{\infty} i(i-1)c_i t^{i-2} = \sum_{i=0}^{\infty} (-c_i) t^i \,,$$

wobei links die Summanden für $i = 0, 1$ beim Ableiten Null geworden sind. Durch Indexverschiebung erreicht man die für den Koeffizientenvergleich günstige Form

$$\sum_{i=0}^{\infty} (i+2)(i+1)c_{i+2} t^i = \sum_{i=0}^{\infty} (-c_i) t^i \,,$$

und Durchführung des Vergleichs gibt die Rekursionsbedingung zweiter Ordnung

$$c_{i+2} = \frac{-c_i}{(i+2)(i+1)} \qquad \text{für } i \in \mathbb{N}_0 \,.$$

Die Rekursion lässt sich in Abhängigkeit von c_0 und c_1 auflösen zu

$$c_{2j} = \frac{(-1)^j}{(2j)!} c_0 \qquad \text{und} \qquad c_{2j+1} = \frac{(-1)^j}{(2j+1)!} c_1 \qquad \text{für } j \in \mathbb{N}_0 \,,$$

und für die Lösung u ergibt sich mit trigonometrischen Reihen, die bekanntlich für alle $t \in \mathbb{R}$ konvergieren,

$$u(t) = c_0 \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j)!} t^{2j} + c_1 \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j+1)!} t^{2j+1} = c_0 \cos t + c_1 \sin t \,.$$

Die Koeffizienten $c_0, c_1 \in \mathbb{K}$ übernehmen hier also die Rolle der Integrationskonstanten, und man erhält erneut die schon früher berechnete allgemeine Lösung — beim hiesigen Verfahren allerdings erst einmal ohne die Information, dass keine weiteren Lösungen existieren und es sich wirklich um die *allgemeine* Lösung handelt.

(2) Bei der Gleichung

$$(2-t^2)u'' - 2tu' + \zeta u \equiv 0$$

mit Parameter $\zeta \in \mathbb{C}$ erhält man durch Einsetzen von (5.12) mit $t_0 = 0$ und Ableiten zuerst

$$(2-t^2)\sum_{i=2}^{\infty} i(i-1)c_i t^{i-2} - 2t\sum_{i=1}^{\infty} i c_i t^{i-1} + \zeta \sum_{i=0}^{\infty} c_i t^i = 0\,.$$

Dies lässt sich umschreiben in

$$\sum_{i=0}^{\infty} 2(i+2)(i+1)c_{i+2}t^i - \sum_{i=0}^{\infty}\big[i(i-1) + 2i - \zeta\big]c_i t^i = 0\,,$$

und Koeffizientenvergleich ergibt die Zweiter-Ordnung-Rekursionsbedingung

$$c_{i+2} = \frac{(i+1)i - \zeta}{2(i+2)(i+1)}c_i \qquad \text{für alle } i \in \mathbb{N}_0\,.$$

Im Allgemeinen[7] erhält man aus dieser Rekursion keine bekannte Reihe. Schreibt man (5.12) aber als Summe $\sum_{j=0}^{\infty} c_{2j}(t^2)^j + t\sum_{j=0}^{\infty} c_{2j+1}(t^2)^j$ zweier Reihen in t^2, so lassen sich mit der Euler-Formel und der Rekursionsvorschrift die Konvergenzradien $\lim_{j\to\infty}\left|\frac{c_{2j}}{c_{2(j+1)}}\right| = \lim_{j\to\infty}\left|\frac{2(2j+2)(2j+1)}{(2j+1)2j-\zeta}\right| = 2$ und $\lim_{j\to\infty}\left|\frac{c_{2j+1}}{c_{2(j+1)+1}}\right| = \lim_{j\to\infty}\left|\frac{2(2j+3)(2j+2)}{(2j+2)(2j+1)-\zeta}\right| = 2$ der Teilreihen bestimmen (falls $\zeta \neq (i+1)i$ für alle geraden beziehungsweise ungeraden $i \in \mathbb{N}_0$; sonst liegen abbrechende Reihen mit Konvergenzradius ∞ vor). Deshalb konvergieren die Teilreihen und folglich auch die ganze Reihe aus (5.12) jedenfalls für $t^2 < 2$, mit anderen Worten also für $|t| < \sqrt{2}$. Insgesamt ist damit für beliebiges $\zeta \in \mathbb{C}$ die Existenz von Lösungen auf $\left(-\sqrt{2}, \sqrt{2}\right)$ mit den ABen $u(0) = c_0$, $u'(0) = c_1$ gezeigt, und die Lösbarkeit des entsprechenden AWPs ist für beliebige Anfangsdaten $c_0, c_1 \in \mathbb{C}$ bewiesen.

Bemerkungen (zum Lösen durch **Potenzreihenansatz**).

(1) Ein **Nachteil** des Potenzreihenansatzes liegt in der Möglichkeit, dass man **eventuell nicht die *allgemeine* Lösung findet**. In vielen Fällen bekommt man zwar doch *alle* Lösungen, aber um sich dessen sicher sein zu können, muss man Zusatzwissen über die (Theorie der) betreffende(n) GDG zur Verfügung haben.

[7] Im Spezialfall $\zeta = 0$ allerdings kann man die Rekursion zu $c_{2j+2} = 0$ und $c_{2j+1} = \frac{c_1}{2^j(2j+1)}$ für $j \in \mathbb{N}_0$ auflösen, und man bekommt durch Zurückführung auf die Logarithmus-Reihe eine explizite Lösungsformel:

$$\begin{aligned} u(t) &= c_0 + \sum_{j=0}^{\infty}\frac{c_1}{2^j(2j+1)}t^{2j+1} = c_0 + \sum_{i=1}^{\infty}\frac{c_1}{\sqrt{2}^{i-1}\,i}t^i - \sum_{j=1}^{\infty}\frac{c_1}{\sqrt{2}^{2j-1}\,2j}t^{2j} \\ &= c_0 - c_1\sqrt{2}\sum_{i=1}^{\infty}\frac{(-1)^{i+1}}{i}\left(\frac{-t}{\sqrt{2}}\right)^i + \frac{c_1}{\sqrt{2}}\sum_{j=1}^{\infty}\frac{(-1)^{j+1}}{j}\left(\frac{-t^2}{2}\right)^j \\ &= c_0 - c_1\sqrt{2}\log\left(1-\frac{t}{\sqrt{2}}\right) + \frac{c_1}{\sqrt{2}}\log\left(1-\frac{t^2}{2}\right) = c_0 + \frac{c_1}{\sqrt{2}}\log\frac{\sqrt{2}+t}{\sqrt{2}-t} \end{aligned}$$

für $|t| < \sqrt{2}$. Die so erhaltene Formel für den Spezialfall kann man jedoch auch anders und einfacher ableiten, beispielsweise durch Separation der Variablen in der Erster-Ordnung-Gleichung $(2-t^2)(u')' - 2tu' \equiv 0$ für u'.

(2) Ein Potenzreihenansatz **kann bei ganz verschiedenen Typen von GDGen funktionieren und mehr oder weniger Information über Lösungen bringen.** Ohne die Rechnung wirklich durchzuführen, lässt sich dies oft nur schwer voraussagen.

Die (tendenziell) **größten Erfolgsaussichten** hat ein Potenzreihenansatz um $t_0 \in \mathbb{R}$ **bei linearen GDGen** $a_m u^{(m)} + a_{m-1}u^{(m-1)} + \ldots + a_2 u'' + a_1 u' + a_0 u = b$ auf I **mit rationalen Koeffizientenfunktionen** der Form

$$a_k(t) = \frac{p_k(t)}{(t-t_0)^{\ell_k}},$$

wobei p_k Polynome über $\mathbb{K}$ und $\ell_k \in \mathbb{N}_0$ sind. Man kann $a_k(t)$ dann nämlich als $\mathbb{K}$-Linearkombinationen von Potenzen $(t-t_0)^j$ mit $j \in \mathbb{Z}$ schreiben und wie im vorausgehenden Beispiel (2) „schön“ mit den Reihengliedern multiplizieren.

Bei anderen Koeffizientenfunktionen zerstört die gliedweise Multiplikation im Allgemeinen die Struktur als Potenzreihe. Um diese Struktur zu erhalten, muss man vielmehr auch die Koeffizientenfunktionen selbst in Potenzreihen entwickeln und mittels des Cauchy-Produkts multiplizieren. Dies **kann bei konkreten Rechnungen sehr kompliziert werden**, und ist oft nicht zielführend. **Ähnliches gilt für nicht-lineare Gleichungen**, bei denen die Cauchy-Multiplikation von Reihen für u und seine Ableitungen erforderlich wird.

(3) Die durch Koeffizientenvergleich erhaltene **Rekursionsbedingung ist in guten Fällen von derselben Ordnung m wie die betrachtete GDG**, d.h. für jedes $i \in \mathbb{N}_0$ liefert sie c_{i+m} als Funktion von $c_i, c_{i+1}, c_{i+2}, \ldots, c_{i+m-1}$. Ist dies der Fall, so ist man in der typischen Situation, dass die allgemeine Lösung von den m Parametern $c_0, c_1, c_2, \ldots, c_{m-1}$ abhängt und AWPe mit m ABen sinnvoll sind.

Es gibt aber keine Garantie für dieses Verhalten, und insbesondere kann es vorkommen, dass nach Koeffizientenvergleich weniger als m Parameter frei bleiben. Dies ist oft ein Anzeichen dafür ist, dass nicht alle Lösungen als Potenzreihe darstellbar sind. Ein Beispiel hierfür folgt, ein weiteres wird (ansatzweise) in den Übungen behandelt.

(4) Eine Variante des Potenzreihenansatzes ist der Ansatz $u(t) = \sum_{i=0}^{\infty} c_i (t-t_0)^{i-\ell}$, $c_0 \neq 0$ mit einer Polstelle der Ordnung $\ell \in \mathbb{N}$ bei t_0. Dieser Ansatz ist gelegentlich sinnvoll, typischerweise aber nur, wenn t_0 nicht-enthaltener Randpunkt von I ist.

Weitere Beispiele (zum **Scheitern eines Potenzreihenansatzes**).

(3) Bei der Gleichung

$$t^3 u' = 2u$$

führt der Potenzreihenansatz (5.12) mit $t_0 = 0$ durch Einsetzen und Indexverschiebung auf

$$\sum_{i=3}^{\infty} (i-2) c_{i-2} t^i = \sum_{i=0}^{\infty} 2 c_i t^i,$$

und Koeffizientenvergleich ergibt $c_0 = c_1 = c_2 = 0$ und $2c_i = (i-2)c_{i-2}$ für $i \geq 3$. Folglich erhält man durch Potenzreihenansatz um 0 nur die Null-Lösung, aber Separation der Variablen zeigt, dass die allgemeine Lösung tatsächlich durch $u(t) = C_1 \mathrm{e}^{-1/t^2}$ für $t < 0$, $u(0) = 0$ und $u(t) = C_2 \mathrm{e}^{-1/t^2}$ für $t > 0$ mit Konstanten $C_1, C_2 \in \mathbb{K}$ gegeben ist. Das Scheitern des Ansatzes erklärt sich hier daraus, dass die **C^∞-Lösung** u mit $u^{(k)}(0) = 0$ für alle $k \in \mathbb{N}_0$ im Fall $(C_1, C_2) \neq (0,0)$ bei 0 **nicht analytisch** ist, d.h. nicht in eine Potenzreihe um 0 entwickelt werden kann.

(4) Bei der Gleichung

$$t^2u' = t + u$$

führt (5.12) mit $t_0 = 0$ ähnlich wie beim vorigen Beispiel auf $c_0 = 0$, $c_1 = -1$ und $c_i = (i-1)c_{i-1}$ für $i \geq 2$. Man bekommt daher $c_i = -(i-1)!$ für $i \in \mathbb{N}$ und erhält die Potenzreihe

$$u(t) = -\sum_{i=1}^{\infty}(i-1)!\, t^i\,.$$

Ungünstigerweise hat diese Reihe den **Konvergenzradius 0** und konvergiert für kein $t \in \mathbb{R} \setminus \{0\}$, daher hat man in diesem Fall tatsächlich keine Lösung gewonnen. Mit Satz 5.1 lässt sich der Grund hierfür einsehen (auch wenn die Stammfunktion B nicht elementar ist): Lösungen existieren auf $(-\infty, 0)$ und $(0, \infty)$, werden aber bei 0 stets singulär und können daher nie in eine Potenzreihe um 0 entwickelt werden.

5.6 Reduktionsverfahren von d'Alembert

Bei der sogenannten d'Alembert-Reduktion handelt es sich um einen Lösungsansatz für skalare lineare Zweiter-Ordnung-GDGen, mit dem auch nicht-konstante Koeffizienten und Inhomogenitäten behandelt werden können. Die Verfahrensweise beruht auf folgendem Satz:

Satz 5.6 (über das **Reduktionsverfahren von d'Alembert**). *Gegeben sei die (eventuell) inhomogene lineare GDG*

$$u'' + a_1u' + a_0u = b \qquad \text{auf } I \tag{5.13}$$

mit einem Intervall I positiver Länge in $\mathbb{R}$ und $a_0, a_1, b\colon I \to \mathbb{K}$. Ist $u_0\colon I \to \mathbb{K}$ eine spezielle Lösung der zugehörigen homogenen GDG (d.h. ist $u_0'' + a_1u_0' + a_0u_0 \equiv 0$ auf I) und hat u_0 keine Nullstelle in I, so ist $u = vu_0$ mit $v\colon I \to \mathbb{K}$ genau dann Lösung von (5.13), wenn v' die Gleichung

$$(v')' + \Big(\frac{2u_0'}{u_0} + a_1\Big)v' = \frac{b}{u_0} \qquad \text{auf } I \tag{5.14}$$

löst.

Beweis. Durch Einsetzen der Gleichheit $u = vu_0$ und Anwendung der Produktregel lässt sich (5.13) in

$$v''u_0 + 2v'u_0' + \underline{vu_0''} + a_1v'u_0 + \underline{a_1vu_0'} + \underline{a_0vu_0} = b$$

umschreiben. Wegen der Lösungseigenschaft von u_0 heben sich die unterstrichenen Terme heraus, und Division durch $u_0 \neq 0$ führt auf

$$v'' + \frac{2u_0'}{u_0}v' + a_1v' = \frac{b}{u_0}\,,$$

was bis auf Ausklammern der GDG (5.14) entspricht. □

Bemerkungen.

(1) Der Satz **reduziert** die **Zweiter-Ordnung-Gleichung** (5.13) für u **auf** die **Erster-Ordnung-Gleichung** (5.14) für v'. Dies ist hilfreich, denn (5.14) kann für $a_1, b \in \mathrm{C}^0(I, \mathbb{K})$ mit der **Lösungsformel** aus Satz 5.1 behandelt werden und hat die allgemeine Lösung

$$v'(t) = \frac{1}{\mathrm{e}^{A_1(t)}u_0(t)^2}\left[\int \mathrm{e}^{A_1(t)}u_0(t)b(t)\,\mathrm{d}t + C\right] \tag{5.15}$$

mit einer Stammfunktion A_1 zu a_1 auf I und einer Konstanten $C \in \mathbb{K}$ (wobei das unbestimmte Integral für die Bildung einer beliebigen Stammfunktion steht). Hat man v' damit berechnet, so erhält man durch Integration auch v (wobei eine zweite Integrationskonstante ins Spiel kommt) und damit die allgemeine Lösung $u = vu_0$ zu (5.13).

(2) Die **entscheidende Voraussetzung** für die Anwendung des Reduktionsverfahrens ist die **Kenntnis einer speziellen Lösung $u_0 \neq 0$** zu $u'' + a_1u' + a_0u \equiv 0$. Bei konstanten Koeffizienten $a_1, a_0 \in \mathbb{K} = \mathbb{C}$ kann eine solche spezielle (und sogar die allgemeine) Lösung der homogenen Gleichung schematisch gemäß Abschnitt 5.2 berechnet werden, und daher ist die d'Alembert-Reduktion in diesem Fall stets anwendbar. Bei nicht-konstanten Koeffizienten a_0, a_1 braucht man „Glück" oder Erfahrung, um eine spezielle Lösung zu raten oder durch Ausprobieren zu finden, und das Verfahren lässt sich nur anwenden, wenn dies gelingt.

(3) Im Prinzip funktioniert ein analoges Reduktionsverfahren auch bei linearen **Gleichungen höherer Ordnung** $m \geq 3$. Ist eine (nullstellenfreie) Lösung der zugehörigen homogenen Gleichung bekannt, so kann man in diesem Fall auf eine Gleichung der Ordnung $(m-1)$ reduzieren, und das Verfahren dann eventuell iterieren. Da dieses Vorgehen in der Rechenpraxis aber selten relevant ist, wird hier nicht näher darauf eingegangen.

Beispiel (zum **Reduktionsverfahren von d'Alembert**)**.** Die homogene Gleichung

$$u'' + 2(\tan t)u' - u \equiv 0 \qquad \text{auf } I$$

mit nicht-konstantem Koeffizienten $a_1(t) = 2\tan t$ ist sinnvoll, wenn der Tangens auf dem Intervall I positiver Länge nicht singulär wird, wenn also $I \cap (2\mathbb{Z}-1)\frac{\pi}{2} = \emptyset$ gilt; dies sei für das Folgende daher unterstellt. Als spezielle Lösung lässt sich $u_0(t) = \sin t$ raten, und durch Anwendung von (5.15) (mit $\mathrm{e}^{A_1(t)} = (\cos t)^{-2}$, $b \equiv 0$) erhält man

$$v'(t) = C\frac{(\cos t)^2}{(\sin t)^2} = C(\cot t)^2\,.$$

Integration mittels $\frac{\mathrm{d}}{\mathrm{d}t}\cot t = -1-(\cot t)^2$ gibt

$$v(t) = C_1(t+\cot t) + C_2$$

mit $C_1, C_2 \in \mathbb{K}$, und durch Multiplikation mit u_0 kommt man schließlich auf die allgemeine Lösung

$$u(t) = C_1(t\sin t+\cos t) + C_2\sin t\,.$$

Eigentlich sind die obige Anwendung des Satzes und die Rechnung dieses Beispiels dabei nur erlaubt, wenn I zusätzlich zur eingangs gestellten Bedingung auch $I \cap \mathbb{Z}\pi = \emptyset$ erfüllt, denn die spezielle Lösung u_0 hat Nullstellen in den Punkten von $\mathbb{Z}\pi$. Mit einem Stetigkeitsargument beziehungsweise dem Prinzip des Aneinandersetzens von Lösungen kann man diese Zusatzbedingung an I aber im Nachhinein wieder eliminieren.

Die Erster-Ordnung-Gleichung für v' kann in diesem Beispiel übrigens als

$$v'' + 2(\cot t + \tan t)v' \equiv 0 \qquad \text{auf } I$$

geschrieben werden.

5.7 Variablentransformation bei Differentialgleichungen

Prinzipiell können GDGen m-ter Ordnung und ihre Lösungen $u\colon I \to \mathcal{X}$ (mit einem Intervall I positiver Länge in $\mathbb{R}$ und einem normierten Raum $\mathcal{X}$ über $\mathbb{K}$) auf zwei verschiedene Arten transformiert werden:

- Eine **Transformation der unabhängigen Variablen t** in eine neue unabhängige Variable $\widetilde{t}$ ist gegeben durch einen C^m-Diffeomorphismus T von $\widetilde{I}$ auf I. Man transformiert dann
$$\widetilde{u}\big(\widetilde{t}\,\big) = u\big(T\big(\widetilde{t}\,\big)\big) \qquad \text{mit Rücktransformation} \qquad u(t) = \widetilde{u}\big(T^{-1}(t)\big)\,.$$
- Eine **Transformation der abhängigen Variablen u** in eine neue abhängige Variable $\widetilde{u}$ ist gegeben durch einen C^m-Diffeomorphismus Y von (einer Teilmenge von) $\mathcal{X}$ auf (eine Teilmenge von) $\widetilde{\mathcal{X}}$. Man transformiert dann
$$\widetilde{u}(t) = Y\big(u(t)\big) \qquad \text{mit Rücktransformation} \qquad u(t) = Y^{-1}\big(\widetilde{u}(t)\big)\,.$$
Auch eine **allgemeinere Transformation der abhängigen Variablen u** mittels einer 1-Parameter-Schar $(Y_t)_{t\in I}$ von C^m-Diffeomorphismen Y_t von (Teilmengen von) $\mathcal{X}$ auf (Teilmengen von) $\widetilde{\mathcal{X}}$, mit C^m-Abhängigkeit auch vom Parameter t, ist in der Praxis oft nützlich. In diesem Fall ist die Transformation gegeben durch
$$\widetilde{u}(t) = Y_t\big(u(t)\big) \qquad \text{mit Rücktransformation} \qquad u(t) = Y_t^{-1}\big(\widetilde{u}(t)\big)\,.$$

In beiden Fällen kann man mit der Kettenregel eine neue GDG für die transformierten Funktionen $\widetilde{u}$ berechnen, und hat man eine ‚gute' Transformation T, Y oder Y_t gewählt, so kann man die Lösungen $\widetilde{u}$ der neuen GDG (hoffentlich) bestimmen. Gelingt dies, so kann man von $\widetilde{u}$ zu u rücktransformieren und erhält auch die Lösungen u der ursprünglichen GDG. Tatsächlich ist die Krux dieses Vorgehens aber das Finden geeigneter Transformationen, und nicht immer sind gute Wahlen leicht ersichtlich.

Im Folgenden werden zu diesem Themenkreis nur zwei Beispiele diskutiert (und in den Übungen wenige weitere); für eine ausführlichere Behandlung sei auf Fachliteratur verwiesen.

Beispiel (für eine **Transformation der unabhängigen Variablen**)**.** Die **Eulersche DGL**
$$t^m u^{(m)} + a_{m-1} t^{m-1} u^{(m-1)} + \ldots + a_2 t^2 u'' + a_1 t u' + a_0 u \equiv 0 \qquad \text{auf } \mathbb{R}^+$$
mit Konstanten $a_0, a_1, a_2, \ldots, a_{m-1} \in \mathbb{K}$ ist homogen, weist aber (insgesamt) keine konstanten Koeffizienten auf. Bei dieser DGL ist durch
$$\widetilde{u}\big(\widetilde{t}\,\big) = u\Big(e^{\widetilde{t}}\Big) \qquad \text{mit Rücktransformation} \qquad u(t) = \widetilde{u}(\log t)$$
(d.h. durch die Wahl $T\big(\widetilde{t}\,\big) = \mathrm{e}^{\widetilde{t}}$, $T^{-1}(t) = \log t$) die **Transformation in eine homogene lineare DGL** auf $\mathbb{R}$ **mit konstanten Koeffizienten** möglich. Gemäß Abschnitt 5.2 kann man dann schematisch lösen.

Als konkretes Beispiel wird nun der Fall $m = 3$, $a_0 = a_1 = -2$, $a_2 = 3$, also die GDG
$$t^3 u''' + 3t^2 u'' - 2tu' - 2u \equiv 0 \qquad \text{auf } \mathbb{R}^+$$

betrachtet. Aus $u(t) = \widetilde{u}(\log t)$ erhält man für die Ableitungen

$$\begin{aligned} u'(t) &= t^{-1}\widetilde{u}'(\log t)\,, \\ u''(t) &= t^{-2}\widetilde{u}''(\log t) - t^{-2}\widetilde{u}'(\log t)\,, \\ u'''(t) &= t^{-3}\widetilde{u}'''(\log t) - 3t^{-3}\widetilde{u}''(\log t) + 2t^{-3}\widetilde{u}'(\log t)\,. \end{aligned}$$

Beim Einsetzen in die GDG kürzen sich alle t-Potenzen heraus, und Zusammenfassen von Termen der gleichen Ableitungsordnung gibt die transformierte Gleichung mit konstanten Koeffizienten

$$\widetilde{u}''' - 3\widetilde{u}' - 2\widetilde{u} \equiv 0 \qquad \text{auf } \mathbb{R}\,.$$

Letztere wurde in Beispiel (3) aus Abschnitt 5.2 bereits gelöst, und aus der dort erhaltenen Formel für die Lösung (jetzt $\widetilde{u}$ genannt) ergibt sich durch Rücktransformation die allgemeine Lösung der Ausgangs-GDG

$$u(t) = \widetilde{u}(\log t) = C_1 \mathrm{e}^{-\log t} + C_2(\log t)\mathrm{e}^{-\log t} + C_3 e^{2\log t} = C_1\frac{1}{t} + C_2\frac{\log t}{t} + C_3 t^2$$

mit $C_1, C_2, C_3 \in \mathbb{C}$.

Beispiel (für eine **Transformation der abhängigen Variablen**)**.** Die nicht-autonome nichtlineare GDG

$$u' = (t+u)^2$$

kann man durch

$$\widetilde{u}(t) = t+u(t) \qquad \text{mit Rücktransformation} \qquad u(t) = \widetilde{u}(t) - t$$

(d.h. durch die Wahl $Y_t(x) = t+x$, $Y_t^{-1}(\widetilde{x}) = \widetilde{x}-t$) auf eine autonome GDG zurückführen. Berechnet man nämlich $u'(t) = \widetilde{u}'(t) - 1$ und setzt dann ein, so kommt man auf

$$\widetilde{u}' - 1 = \widetilde{u}^2\,.$$

Diese neue GDG kann durch Separation der Variablen gelöst werden und hat die allgemeine Lösung

$$\widetilde{u}(t) = \tan(t+C)\,.$$

Als Lösung der Ausgangsgleichung erhält man folglich

$$u(t) = \tan(t+C) - t$$

(auf Intervallen I positiver Länge und mit $C \in \mathbb{R}$, so dass tan auf $I+C$ nicht singulär wird).

5.8 Zur geometrischen Interpretation von GDGen und GDG-Systemen

Manchmal lässt sich eine GDG oder ein GDG-System **geometrisch veranschaulichen**, und man kann durch eine Skizze eine Vorstellung von prinzipiellen Eigenschaften der Lösungen gewinnen. Oft lässt sich damit das qualitative Lösungsverhalten, insbesondere das Langzeitverhalten,

vorhersagen, und auch wenn eine Skizze allein kein Beweis ist, so kann sie trotzdem auf die richtige Idee für einen Beweisansatz führen. Drei Fälle, in denen man tatsächlich **Skizzen in der Zeichen*ebene*** anfertigen kann, folgen:

(A) Der erste Fall sind **skalare Erster-Ordnung-Gleichungen** in expliziter Form

$$u' = f(\,\cdot\,, u)$$

mit einer Strukturfunktion $f\colon D \to \mathbb{R}$ auf einer Teilmenge D von $\mathbb{R}^2$. Hier versteht man f als **Steigungsfeld**, durch das zu jedem Punkt $(t,x) \in D$ eine Steigung $f(t,x)$ vorgegeben wird, und Lösungen der GDG sind genau die differenzierbaren Funktionen $u\colon I \to \mathbb{R}$, für die $u'(x)$ stets mit der vorgegebenen Steigung $f(x,u(x))$ am Punkt $(x,u(x)) \in D$ übereinstimmt. Die Punkte der Form $(x,u(x))$ bilden hierbei den Graph der Funktion u, daher lässt sich aus einer Skizze des Steigungsfelds f und eines Graphen u (prinzipiell) ablesen, ob u eine Lösung ist. Abbildung 6 zeigt ein Beispiel hierzu (bei dem man mit Transformation wie im vorausgehenden Abschnitt 5.7 übrigens auch eine explizite Formel für Lösungen bekommen kann).

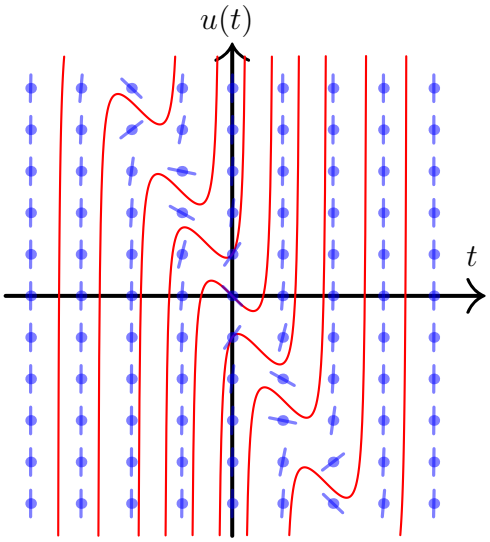


Abb. 6: Das Steigungsfeld $f(t,x) = (2t+x)^2-1$ auf $\mathbb{R}^2$ und die Graphen einiger Lösungen der zugehörigen GDG $u' = (2t+u)^2-1$.

(B) Der zweite Fall sind **autonome Systeme erster Ordnung** in expliziter Form für **$\mathbb{R}^2$-wertiges u**

$$u' = F(u) \tag{5.16}$$

mit einer Strukturfunktion $F\colon D \to \mathbb{R}^2$ auf einer Teilmenge D von $\mathbb{R}^2$. Dieser **für die Vorlesung wichtigste Fall** wird hier ausführlich beschrieben:

- Man fasst $\boldsymbol{F}$ **als Vektorfeld** auf D auf, stellt sich also vor, dass an jeden Punkt $x \in D \subset \mathbb{R}^2$ der zugehörige Vektor $F(x) \in \mathbb{R}^2$ angeheftet ist, oder, etwas präziser, dass am Punkt x ein Vektorpfeil von x nach $x+F(x)$ befestigt ist.
- Eine über einem Intervall $I \subset \mathbb{R}$ **parametrisierte Kurve** $u\colon I \to \mathbb{R}^2$ veranschaulicht man in $\mathbb{R}^2$ durch Skizzierung des **Bildes der Kurve** u und ihres Durchlaufsinns. Üblicherweise wird die Durchlauf*geschwindigkeit* $|u'|$ dabei nicht kenntlich gemacht, weshalb nicht die gesamte Information über u in die Skizze einfließt. Zu beachten ist auch, dass im Gegensatz zu anderen Darstellungen nur das *Bild*, aber nicht der *Graph* von u skizziert wird (denn letzteren kann man bei $\mathbb{R}^2$-wertigem u überhaupt nicht in der Ebene zeichnen).

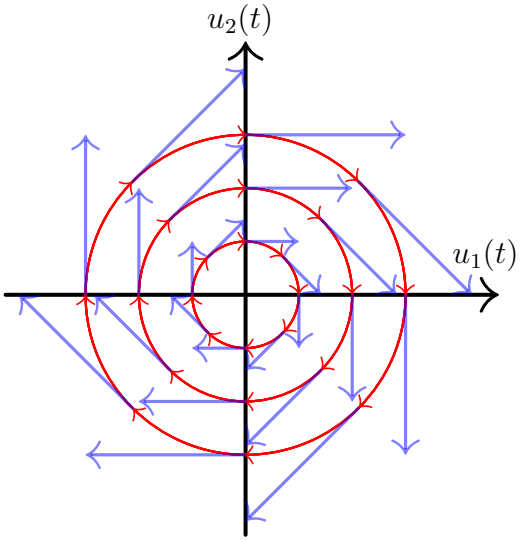


Abb. 7: Das Vektorfeld $F(x) = (x_2, -x_1)$ auf $\mathbb{R}^2$ und einige Lösungskurven des zugehörigen Systems $u_1' = u_2$, $u_2' = -u_1$.

- Dass eine Kurve u **Lösung des Systems** (5.16) ist, manifestiert sich darin, dass die Ableitung $u'(t)$ stets mit dem zum Punkt $u(t)$ gehörigen Vektor $F(u(t))$ übereinstimmt. Man nennt die Lösungen von (5.16) daher auch die **Trajektorien**, **Bahnlinien** oder **Integralkurven** des Vektorfelds F. Genau genommen lässt sich an einer Skizze der beschriebenen Art zwar nicht ablesen, ob u eine Lösung ist, aber zumindest das folgende notwendige Kriterium[8] lässt sich prinzipiell überprüfen: Für die Lösungseigenschaft von $u\colon I \to \mathbb{R}^2$ ist erforderlich, dass $u'(t)$ stets in *dieselbe Richtung* zeigt wie $F(u(t))$, oder genauer, dass $u'(t)$ und $F(u(t))$ stets positiv linear abhängig sind, oder mit noch etwas anderen Worten, dass **die Kurve u sich bei jedem Punkt $u(t)$ in die Richtung des dort befestigten Vektors $F(u(t))$ bewegt**.

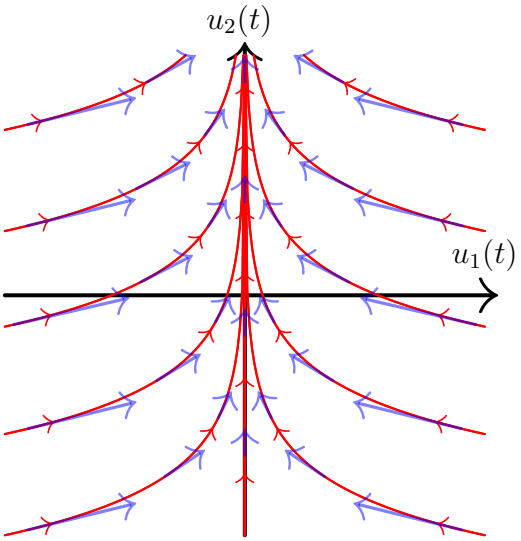


Abb. 8: Das Vektorfeld $F(x) = (-\frac{x_1}{2}, 1)$ auf $\mathbb{R}^2$ und einige Lösungskurven des zugehörigen Systems $u_1' = -\frac{u_1}{2}$, $u_2' \equiv 1$.

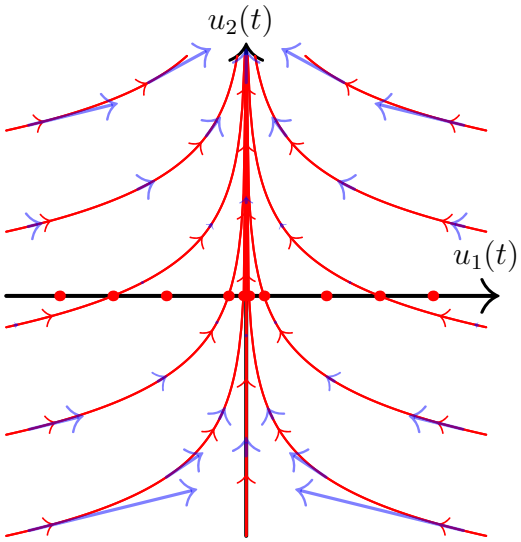


Abb. 9: Das Vektorfeld $F(x) = \frac{x_2^2}{50}(-\frac{x_1}{2}, 1)$ auf $\mathbb{R}^2$ und einige Lösungskurven des zugehörigen Systems $u_1' = -\frac{u_2^2 u_1}{100}$, $u_2' = \frac{u_2^2}{50}$.

Die Abbildungen 7, 8 und 9 zeigen Beispiele der beschriebenen Skizzen. Dabei unterscheiden sich die Vektorfelder der Abbildungen 8 und 9 nur um den skalaren Faktor $\frac{x_2^2}{50}$, durch den die Vektoren in Abbildung 9 nahe der x_1-Achse stark verkürzt werden, so dass sie teils nur noch andeutungsweise dargestellt werden können. Da die Richtung der Vektoren dennoch erhalten bleibt, gleichen sich die Bilder der Lösungskurven in den beiden Abbildungen, während die Durchlaufgeschwindigkeiten und damit die Lösungen selbst differieren. Bei den Nullstellen des Vorfaktors auf der x_1-Achse selbst kommt es allerdings zu qualitativen Unterschieden im Lösungsverhalten. In Abbildung 8 überqueren die Lösungen die x_1-Achse ohne Singularitäten, in Abbildung 9 erreichen sie die x_1-Achse nur im Limes $t \to \pm\infty$. Hinzu kommen in Abbildung 9 außerdem Lösungen, die einen Wert auf der x_1-Achse konstant annehmen, was durch die roten Punkte auf dieser Achse angedeutet wird. Insgesamt zeigt der Vergleich dieser beiden Beispiele also, **dass die** beschriebene **geometrische Interpretation zwar viele, doch nicht alle Informationen über Lösungen** des Systems (5.16) **liefern kann**.

[8] Weg von Nullstellen von u' und für $F \in \mathrm{C}^0(D, \mathbb{R}^2)$ ist das Kriterium auch hinreichend dafür, dass eine orientierungserhaltende Umparametrisierung von u eine Lösung ist; die Umparametrisierung erhält man dabei aus einem allgemeinen Existenzsatz, der erst im späteren Abschnitt 6.7 behandelt wird.

Außerdem ist zu beachten, dass die allgemeine Lösung von (5.16) typischerweise von *zwei* reellen Konstanten als Parametern abhängt, was in den Skizzen nicht direkt ersichtlich ist. Vielmehr scheinen die roten Lösungskurven nur *Ein*-Parameter-Scharen zu bilden, doch der scheinbar fehlende Parameter erklärt sich aus einer **generellen Eigenschaft *autonomer* GDGen** und GDG-Systeme: Aus einer Lösung u erhält man durch **Verschiebung der *t*-Variablen** $\widetilde{u}(t) = u(t+C)$ um jedes beliebige $C \in \mathbb{R}$ weitere Lösungen $\widetilde{u}$, und all diese Lösungen kommen zeichnerisch auf derselben Bildkurve zu liegen. Der scheinbar fehlende Parameter entspricht also dem in den Skizzen nicht auflösbaren C der Verschiebungen

Zum Abschluss dieses Punktes sei angemerkt, dass die graphische Darstellung von Trajektorien der Gleichung (5.16) selbst im homogenen linearen Fall, also für $F(x) = Ax$ mit $A \in \mathbb{R}^{2\times 2}$, sinnvoll und hilfreich sein kann. Beispiele hierzu sind Thema der Übungen.

(C) Der dritte Fall sind **autonome, skalare Zweiter-Ordnung-Gleichungen** in expliziter Form

$$u'' = \psi(u, u')$$

mit einer Strukturfunktion $\psi\colon D \to \mathbb{R}$ auf einer Teilmenge D von $\mathbb{R}^2$. Bei solchen Gleichungen kann man durch **Reduktion auf Ordnung 1** zu einem autonomen System wie in (B) übergehen, nämlich zu $u_1' = u_2$, $u_2' = \psi(u_1, u_2)$ mit dem zugehörigen Vektorfeld $F(x_1, x_2) = (x_2, \psi(x_1, x_2))$. Dieses Vektorfeld und die Trajektorien lassen sich nun genau wie in (B) beschrieben graphisch darstellen, wobei die Variable auf der x_1-Achse nun $u(t)$ und die auf der x_2-Achse der Ableitung $u'(t)$ entspricht.

Als konkretes Beispiel kann man an die Gleichung $u'' = -u$ denken. Dann erhält man das Vektorfeld F der Abbildung 7 und folglich genau das dort gezeigte Bild — einzig mit dem Unterschied, dass man sich die Achsen nun mit $u(t)$ und $u'(t)$ beschriftet vorstellen sollte.

Schließlich sei für ganz allgemeine GDGen festgehalten:

Terminologie. *Bei einer GDG oder einem GDG-System der Ordnung m für $\mathcal{X}$-wertige Funktionen u nennt man $\mathcal{X}^m$ in seiner Funktion als Wertebereich von $(u, u', u'', \dots, u^{(m-1)})$ den **Phasenraum** der Gleichung oder des Systems. Daher heißen die graphischen Darstellungen der vorausgehenden Punkte (B) und (C) auch **Phasenraumportraits** — wobei in (B) der Fall $m = 1$, $\mathcal{X} = \mathbb{R}^2$ betrachtet wurde und in (C) der Fall $m = 2$, $\mathcal{X} = \mathbb{R}$, so dass es jeweils um Portraits in $\mathcal{X}^m = \mathbb{R}^2$ ging. Als **erweiterten Phasenraum** bezeichnet man den Wertebereich $\mathbb{R} \times \mathcal{X}^m$ von $t \mapsto (t, u(t), u'(t), u''(t), \dots, u^{(m-1)}(t))$, und in diesem sind die die unter Punkt (A) beschriebenen Darstellungen angesiedelt — denn dort war $m = 1$, $\mathcal{X} = \mathbb{R}$, also $\mathbb{R} \times \mathcal{X}^m = \mathbb{R}^2$.*

Prinzipiell gibt es Phasenraumportraits natürlich auch in höheren Dimensionen, und bei autonomen Erster-Ordnung-Systemen für drei (Komponenten-)Funktionen, also im Fall $m = 1$, $\mathcal{X} = \mathbb{R}^3$, $\mathcal{X}^m = \mathbb{R}^3$, kann man sich dies noch ähnlich wie unter (B) mit Lösungskurven in $\mathbb{R}^3$ vorstellen; das Zeichnen wird in diesem räumlichen Fall aber deutlich schwieriger. Auch bei nicht-autonomen Systemen $u' = F(\,\cdot\,, u)$ für zwei oder drei Funktionen kann man sich eventuell noch eine anschauliche Vorstellung des Typs (B) machen, wenn man t als Zeitvariable auffasst und F als zeitabhängiges Vektorfeld. In zu großen Dimensionen versagt aber im Allgemeinen jede einfache Vorstellungsmöglichkeit.

Kapitel 6

Die Hauptsätze der Theorie

In diesem Kapitel sei $\mathcal{X}$ stets ein Banach-Raum (d.h. ein vollständiger normierter Raum) über $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ mit Norm $|\cdot| = \|\cdot\|_{\mathcal{X}}$. Am relevantesten sind die Fälle $\mathcal{X} = \mathbb{R}^N$ und $\mathcal{X} = \mathbb{C}^N$ mit der Euklidischen Norm, aber auch Räume unendlicher Dimension $\dim_{\mathbb{K}} \mathcal{X} = \infty$ werden nicht ausgeschlossen (sofern nicht explizit etwas anderes gesagt wird).

6.1 Der Existenz- und Eindeutigkeitssatz von Picard-Lindelöf

Es folgt der (vielleicht) grundlegendste Satz der Vorlesung über die Lösbarkeit von AWPen (∼1890, benannt nach Émile Picard und Ernst Leonard Lindelöf). Der Satz wird zuerst für GDGen erster Ordnung formuliert, die Verallgemeinerung auf GDGen beliebiger Ordnung $m \in \mathbb{N}$ ist aber nicht schwierig und wird später in diesem Abschnitt auch angegeben.

Hauptsatz 6.1 (**Satz von Picard-Lindelöf; Erster-Ordnung-Version**). *Gegeben sei das AWP*

$$u' = f(\cdot, u), \qquad u(t_0) = y_0 \tag{6.1}$$

mit Strukturfunktion $f \in \mathrm{C}^0(D, \mathcal{X})$ auf $D \subset \mathbb{R} \times \mathcal{X}$ und $(t_0, y_0) \in D$. Gibt es $\delta, r, L, M \in \mathbb{R}^+$ mit $[t_0-\delta, t_0+\delta] \times \overline{\mathrm{B}}_r(y_0) \subset D$ und

$$|f(t,x)| \le M, \tag{B}$$
$$|f(t,\widetilde{x}) - f(t,x)| \le L|\widetilde{x}-x| \tag{pLB}$$

für alle $t \in [t_0-\delta, t_0+\delta]$ und $x, \widetilde{x} \in \overline{\mathrm{B}}_r(y_0)$, so existiert für jedes $0 < \varepsilon \le \min\{\delta, r/M\}$ ***genau eine Lösung u des AWPs*** *(6.1)* ***auf $[t_0-\varepsilon, t_0+\varepsilon]$****, und diese erfüllt $u([t_0-\varepsilon, t_0+\varepsilon]) \subset \overline{\mathrm{B}}_r(y_0)$.*

Bezeichnungen. Hier stehen $\mathrm{B}_r(y_0) := \{x \in \mathcal{X} : |x-y_0| < r\}$ und $\overline{\mathrm{B}}_r(y_0) = \{x \in \mathcal{X} : |x-y_0| \le r\}$ für die offene und die abgeschlossene Kugel mit Mittelpunkt y_0 und Radius r in $\mathcal{X}$.

Bemerkungen (zum **Satz von Picard-Lindelöf**).

(1) Der Satz erfasst **sowohl einzelne skalare GDGen als auch GDG-Systeme**.

(2) Die Beschränktheits-Voraussetzung (B) folgt automatisch aus der Stetigkeit von f, im Fall $\dim_{\mathbb{K}} \mathcal{X} = \infty$ zusammen mit (pLB). Die explizite Form von (B) wurde oben aber dennoch aufgenommen, um mit Hilfe der Schranke M auch ε explizit angeben zu können.

(3) Die **entscheidenden Voraussetzungen** des Satzes sind die **explizite Form** der GDG in (6.1) und die **partielle Lipschitz-Bedingung** (pLB), wobei das Wort ‚partiell' anzeigt, dass die Lipschitz-Bedingung nur die x-Variable und nicht die t-Variable betrifft. Die Gültigkeit von (pLB) folgt aus dem Schrankensatz, wenn $x \mapsto f(t,x)$ für $t \in [t_0-\delta, t_0+\delta]$ auf $\mathrm{B}_r(y_0)$ total differenzierbar ist mit Ableitungsschranke $\|\mathrm{D}_x f(t,x)\| \le L$ (in der Operatornorm $\|\cdot\|$ auf $\mathcal{L}(\mathcal{X},\mathcal{X})$).

(4) Der Satz gilt **analog für einseitige Intervalle** mit Randpunkt t_0, d.h. unter denselben Voraussetzungen mit $[t_0, t_0+\delta]$ anstelle von $[t_0-\delta, t_0+\delta]$ bekommt man die Existenz genau einer Lösung auf $[t_0, t_0+\varepsilon]$.

(5) Der Satz wurde hier als ***lokaler*** **Existenz- und Eindeutigkeitssatz** für Lösungen des AWPs formuliert. Betrachtet man die beiden enthaltenen Fragestellungen aber separat, so stellt man fest, dass Lösungen im Allgemeinen **tatsächlich nur lokal existieren** (vergleiche die Beispiele mit dem Tangens in 5.3 und 5.7), während **Eindeutigkeit automatisch auch global** auf jedem vorgegebenen Intervall positiver Länge $I \ni t_0$ vorliegt (vorausgesetzt (pLB) ist nahe jedem $(t,x) \in D \cap (I \times \mathcal{X})$ erfüllt).

Begründung der globalen Eindeutigkeit. Man überlegt zuerst, dass zwei Lösungen u, $\widetilde{u}$ des AWPs (6.1) auf I auf $[t_0,\infty) \cap I$ übereinstimmen müssen. Wäre dies nicht der Fall, so gäbe es ein größtes τ mit $t_0 \le \tau <$ rechterRandpunkt(I) und $\widetilde{u} = u$ auf $[t_0,\tau]$ (wobei wegen Stetigkeit von u und $\widetilde{u}$ direkt ein beim Randpunkt τ abgeschlossenes Intervall betrachtet werden kann). Der Satz, mit τ anstelle von t_0 und eventuell in der einseitigen Version angewandt, gäbe dann ein $\varepsilon > 0$ mit $\widetilde{u} = u$ auch auf $[t_0, \tau+\varepsilon]$, doch dies widerspricht der Wahl von τ. Somit muss $\widetilde{u} = u$ auf $[t_0,\infty) \cap I$ gelten. Ein analoges Argument gibt $\widetilde{u} = u$ auf $(-\infty, t_0] \cap I$, und die Eindeutigkeit auf ganz I ist gezeigt. □

Gegenbeispiel („ohne explizite Form und pLB keinerlei Eindeutigkeit"). Die skalare Gleichung

$$(u')^3 = u \qquad \text{oder äquivalent} \qquad u' = \sqrt[3]{u}\,.$$

ist nicht in expliziter Form (erste Darstellung), oder ihre Strukturfunktion $\sqrt[3]{x}$ erfüllt nahe $x = 0$ keine Lipschitz-Bedingung (zweite Darstellung). Daher ist der **Satz von Picard-Lindelöf** auf das AWP mit AB $u(0) = 0$ **nicht anwendbar**, und im Folgenden wird begründet, dass bei diesem AWP keine Eindeutigkeit vorliegt. Dazu berechnet man durch Separation der Variablen die Lösungen $u(t) = \pm\big(\frac{2}{3}t - C\big)^{\frac{3}{2}}$ auf $\big[\frac{3}{2}C, \infty\big)$. Diese können — wie in Abbildung 10 skizziert — mit der Null-Lösung zusammengesetzt werden, und tatsächlich hat dann das AWP auf ganz $\mathbb{R}$ mit AB $u(0) = 0$ die **überabzählbar vielen Lösungen**

$$u(t) = \begin{cases} 0 & \text{für } t \le \frac{3}{2}C \\ \pm\big(\frac{2}{3}t - C\big)^{\frac{3}{2}} & \text{für } t \ge \frac{3}{2}C \end{cases}$$

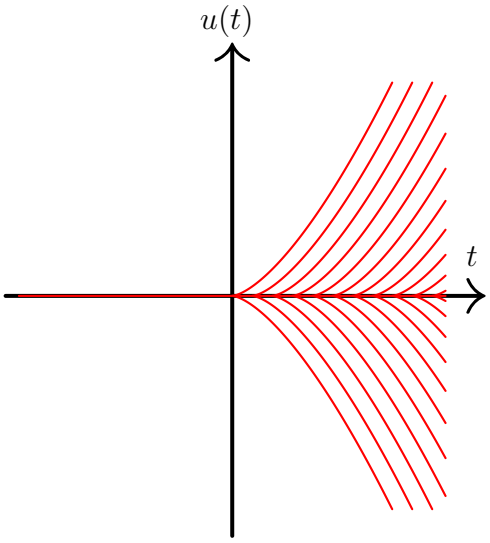


Abb. 10: Die Graphen einiger Lösungen des AWPs $(u')^3 = u$, $u(0) = 0$.

mit Parameter $C \in [0,\infty]$. Insbesondere sind auch auf jedem Intervall $[-\varepsilon,\varepsilon]$ oder $[0,\varepsilon]$ mit $\varepsilon > 0$ überabzählbar viele dieser Lösungen verschieden, und somit liegt **nicht einmal lokale Eindeutigkeit** von Lösungen vor.

Als Nächstes folgt eine Höherer-Ordnung-Version des Satzes von Picard-Lindelöf. Diese könnte völlig analog zu Hauptsatz 6.1 formuliert werden, wird hier aber bewusst in einer leicht anderen Form angegeben. Insbesondere wird obige Bemerkung (5) direkt eingebunden, wodurch der Satz zu einem **lokalen Existenz- und globalen Eindeutigkeitssatz** wird:

Korollar 6.2 (**Satz von Picard-Lindelöf; eine m-ter-Ordnung-Version**). *Gegeben sei eine Strukturfunktion $f \in \mathrm{C}^0(D, \mathcal{X})$ auf offenem $D \subset \mathbb{R} \times \mathcal{X}^m$. Ist f lokal auf D bezüglich der $\mathcal{X}^m$-Variablen Lipschitz-stetig*[1]*, so hat das AWP*

$$u^{(m)} = f(\,\cdot\,, u^{[m-1]})\,, \qquad u^{[m-1]}(t_0) = (y_0, y_1, y_2, \ldots, y_{m-1}) \tag{6.2}$$

für alle $(t_0, y_0, y_1, y_2, \ldots, y_{m-1}) \in D$ stets ...

- **(*Existenz*)** *mindestens eine Lösung auf $[t_0-\varepsilon, t_0+\varepsilon]$ mit ausreichend kleinem*[2] *$\varepsilon > 0$,*
- **(*Eindeutigkeit*)** *höchstens eine Lösung auf jedem Intervall I positiver Länge mit $t_0 \in I$.*

Bezeichnung. Für $k \in \mathbb{N}_0$ und eine k-mal differenzierbare Funktion u auf einem Intervall positiver Länge wurde hier die Abkürzung

$$u^{[k]} := \left(u, u', u'', \ldots, u^{(k-1)}, u^{(k)}\right)$$

für den sogenannten ***k*-Jet** von u verwendet.

Zur Ableitung des Korollars reduziert man das AWP (6.2), wie im einführenden Kapitel der Vorlesung erläutert, **auf Ordnung 1**. Die Behauptung ergibt sich dann problemlos durch Anwendung von Hauptsatz 6.1 und Bemerkung (5) auf das reduzierte System. □

Als Nächstes wird ein Beweis von Hauptsatz 6.1 ausgeführt. Wesentliches Hilfsmittel ist dabei folgendes, aus der Analysis-Grundvorlesung bekannte Resultat, an das hier erinnert wird:

Satz (**Banachscher Fixpunktsatz**, auch Kontraktionssatz genannt). *Ist $A \neq \emptyset$ vollständiger metrischer Raum mit Metrik d und $T\colon A \to A$ strikte Kontraktion (d.h. $d(T(\widetilde{a}), T(a)) \leq \kappa d(\widetilde{a}, a)$ für alle $a, \widetilde{a} \in A$ mit einem $\kappa < 1$), so gibt es genau einen Fixpunkt von T in A (d.h. ein $a \in A$ mit $a = T(a)$).*

Beweis von Hauptsatz 6.1. Die Argumentation basiert auf der **Formulierung des AWPs als Fixpunktproblem**

$$u = Tu \tag{6.3}$$

[1] Hier wird $\mathcal{X}^m$ als normierter Raum mit Norm $\|(x_0, x_1, x_2, \ldots, x_{m-1})\|_{\mathcal{X}^m} := \sum_{i=0}^{m-1} |x_i|$ verstanden, andere Wahlen wie $\left(\sum_{i=0}^{m-1} |x_i|^2\right)^{\frac{1}{2}}$ geben aber äquivalente Normen und könnten genauso gut verwendet werden.

[2] Dieses ε hängt — in hier nicht näher spezifizierter Weise — von f, D und $(t_0, y_0, y_1, y_2, \ldots, y_{m-1})$ ab.

mit folgendem **Integraloperator**[3] T auf stetigen Funktionen $[t_0{-}\varepsilon, t_0{+}\varepsilon] \to \overline{\mathrm{B}}_r(y_0)$:

$$T\colon \mathrm{C}^0\big([t_0{-}\varepsilon, t_0{+}\varepsilon], \overline{\mathrm{B}}_r(y_0)\big) \to \mathrm{C}^0([t_0{-}\varepsilon, t_0{+}\varepsilon], \mathcal{X})\,,\ w \mapsto Tw\,,$$
$$Tw(t) := y_0 + \int_{t_0}^{t} f(s, w(s))\,\mathrm{d}s \in \mathcal{X} \qquad \text{für } t \in [t_0{-}\varepsilon, t_0{+}\varepsilon]\,.$$

Dass dieser Operator wohldefiniert ist, folgt aus den Voraussetzungen $\varepsilon \leq \delta$, $[t_0{-}\delta, t_0{+}\delta] \times \overline{\mathrm{B}}_r(y_0) \subset D$ und $f \in \mathrm{C}^0(D, \mathcal{X})$ sowie dem HDI, wobei letzterer garantiert, dass der Integralausdruck eine stetige und tatsächlich sogar eine differenzierbare Funktion ergibt. Durch Ausschreiben der Definition von T in (6.3) sieht man außerdem, dass die Fixpunktgleichung (6.3) durch Ableiten in die im Satz betrachtete GDG übergeht und umgekehrt die GDG durch Stammfunktionsbildung in die Fixpunktgleichung. Dabei sind allerdings noch Integrationskonstanten zu berücksichtigen, und in Anbetracht dessen, dass Tu per Definition stets $Tu(t_0) = y_0$ erfüllt, kommt man für $u \in \mathrm{C}^0\big([t_0{-}\varepsilon, t_0{+}\varepsilon], \overline{\mathrm{B}}_r(y_0)\big)$ insgesamt auf die Äquivalenz

$$u \text{ löst } u' = f(\,\cdot\,, u) \text{ auf } [t_0{-}\varepsilon, t_0{+}\varepsilon] \text{ und } u(t_0) = y_0 \iff Tu = u \text{ in } \mathrm{C}^0([t_0{-}\varepsilon, t_0{+}\varepsilon], \mathcal{X}) \quad (6.4)$$

(wobei insbesondere jede stetige Lösung der Fixpunktgleichung automatisch differenzierbar ist, denn Tu ist ja nach dem HDI differenzierbar).

Zwecks Anwendung des Fixpunktsatzes wird $\mathrm{C}^0([t_0{-}\varepsilon, t_0{+}\varepsilon], \mathcal{X})$ versehen mit der **Norm der gleichmäßigen Konvergenz** (auch sup-Norm genannt)

$$\|w\|_\infty := \sup_{[t_0-\varepsilon, t_0+\varepsilon]} |w|\,.$$

Die Vollständigkeit[4] des normierten Raums $\mathrm{C}^0([t_0{-}\varepsilon, t_0{+}\varepsilon], \mathcal{X})$ und seiner abgeschlossenen Teilmenge

$$A := \mathrm{C}^0\big([t_0{-}\varepsilon, t_0{+}\varepsilon], \overline{\mathrm{B}}_r(y_0)\big)$$

ist dann im Wesentlichen aus den Analysis-Grundvorlesungen bekannt: Sie folgt direkt aus dem Cauchy-Kriterium[5] für gleichmäßige Konvergenz und dem Satz, dass Stetigkeit von Funktionen unter gleichmäßiger Konvergenz erhalten bleibt.

Zur Anwendung des Fixpunktsatzes bleiben schließlich noch zwei Eigenschaften des Operators T auf A zu prüfen:

[3] Beim Integral in der Definition von T handelt es sich um das Riemann-Integral einer stetigen, $\mathcal{X}$-wertigen Funktion. Im Fall $\dim_{\mathbb{K}} \mathcal{X} < \infty$ kann dieses Integral durch komponentenweise Riemann-Integration erklärt werden, es macht aber auch im ∞-dimensionalen Fall Sinn. Dazu geht man in der bei Kurvenintegralen üblichen Weise vor und definiert für $t > t_0$ in $\mathbb{R}$ und Banach-Raum-wertiges $g \in \mathrm{C}^0([t_0, t], \mathcal{X})$ ganz allgemein $\int_{t_0}^{t} g(s)\,\mathrm{d}s$ als Limes der Näherungssummen $\sum_{i=1}^{k} (t_i{-}t_{i-1}) g(s_i)$ mit $k \in \mathbb{N}$ und $t_0 < s_1 < t_1 < s_2 < t_2 < \ldots < s_k < t_k = t$ bei gegen Null strebender Feinheit $\max_{i\in\{1,2,\ldots,k\}} (t_i{-}t_{i-1})$ der Zerlegung $(t_0, t_1, t_2, \ldots, t_k)$ — wobei der Limes unter den gemachten Voraussetzungen stets in $\mathcal{X}$ existiert. Für $t < t_0$ setzt man $\int_{t_0}^{t} g(s)\,\mathrm{d}s := -\int_{t}^{t_0} g(s)\,\mathrm{d}s$, und für $t = t_0$ trifft man natürlich die Konvention, dass das Integral verschwindet. Der HDI und andere Basis-Resultate über „normale“ $\mathbb{K}$-wertige Integrale bleiben für diese Bildung ohne wesentliche Probleme gültig.

[4] Es sei an die Definition der Vollständigkeit im Kontext normierter oder metrischer Räume erinnert: Vollständigkeit eines solchen Raums bedeutet, dass jede Cauchy-Folge in diesem Raum konvergiert.

[5] Das Kriterium besagt, dass jede Funktionenfolge, die gleichmäßige Cauchy-Folge ist, auch gleichmäßig konvergiert. Dies gilt selbstverständlich nur bei vollständigem Zielraum, benutzt im vorliegenden Kontext also die Vollständigkeit von $\mathcal{X}$.

- Zum **Nachweis, dass T die Menge A in sich selbst abbildet**, ist zu zeigen, dass Tw mit $w \in A$ stets Werte in $\overline{\mathrm{B}}_r(y_0)$ hat. Dies gelingt unter wesentlicher Verwendung der Voraussetzungen $\varepsilon \le \min\{\delta, r/M\}$ und (B) mit der Abschätzung

$$|Tw(t) - y_0| = \left| \int_{t_0}^{t} f(s, w(s))\,\mathrm{d}s \right| \le \int_{t_0}^{t} \underbrace{|f(s, w(s))|}_{\le M \text{ weil } w(s) \in \overline{\mathrm{B}}_r(y_0)}\,\mathrm{d}s \le |t{-}t_0| M \le \varepsilon M \le r$$

  für $t \in [t_0, t_0{+}\varepsilon]$, und für $t \in [t_0{-}\varepsilon, t_0]$ folgt es ganz analog, wenn man nur die Grenzen des zweiten Integrals in der Abschätzung vertauscht.

- Zum **Nachweis einer strikten Kontraktionseigenschaft auf A** schätzt man für $w, \widetilde{w} \in A$ gemäß (pLB) wie folgt ab:

$$|T\widetilde{w}(t) - Tw(t)| = \left| \int_{t_0}^{t} \underbrace{\big[f(s, \widetilde{w}(s)) - f(s, w(s))\big]}_{|\cdot| \le L|\widetilde{w}(s) - w(s)|}\,\mathrm{d}s \right| \le |t{-}t_0| L \|\widetilde{w}{-}w\|_\infty \tag{6.5}$$

  für $t \in [t_0{-}\varepsilon, t_0{+}\varepsilon]$. Insgesamt folgt $\|T\widetilde{w}{-}Tw\|_\infty \le \varepsilon L \|\widetilde{w}{-}w\|_\infty$ und somit die strikte Kontraktionseigenschaft, *vorausgesetzt dass* $\varepsilon L < 1$ gilt. Unter der letzten Zusatzvoraussetzung könnte man den Beweis nun abschließen, aber man kann diese Voraussetzung — die im Hauptsatz ja nicht auftaucht — tatsächlich vermeiden: Dazu schließt man aus (6.5) zunächst nur im Fall $t \ge t_0$ weiter auf

$$\begin{aligned} |T^2\widetilde{w}(t) - T^2 w(t)| &= \left| \int_{t_0}^{t} \big[f(s, T\widetilde{w}(s)) - f(s, Tw(s))\big]\,\mathrm{d}s \right| \le L \int_{t_0}^{t} |T\widetilde{w}(s) - Tw(s)|\,\mathrm{d}s \\ &\le \int_{t_0}^{t} |s{-}t_0|\,\mathrm{d}s\, L^2 \|\widetilde{w}{-}w\|_\infty = \frac{1}{2}|t{-}t_0|^2 L^2 \|\widetilde{w}{-}w\|_\infty . \end{aligned}$$

  Die resultierende Abschätzung bleibt für alle $t \in [t_0{-}\varepsilon, t_0{+}\varepsilon]$ richtig, denn für $t \le t_0$ muss man in der vorigen Abschätzung nur einige Integralgrenzen vertauschen. Durch induktive Fortsetzung derselben Argumentation ergibt sich für die k-fache Iteration von T mit $k \in \mathbb{N}_0$ dann die Abschätzung

$$|T^k\widetilde{w}(t) - T^k w(t)| \le \frac{1}{k!}|t{-}t_0|^k L^k \|\widetilde{w}{-}w\|_\infty \qquad \text{für } t \in [t_0{-}\varepsilon, t_0{+}\varepsilon] .$$

  Insbesondere ist

$$\|T^k\widetilde{w}{-}T^k w\|_\infty \le \frac{(\varepsilon L)^k}{k!}\|\widetilde{w}{-}w\|_\infty ,$$

  wegen des über-exponentiellen Fakultäten-Wachstums gibt es ein $k_0 \in \mathbb{N}$ mit $\frac{(\varepsilon L)^{k_0}}{k_0!} < 1$, und zumindest **$T^{k_0}$ ist daher strikte Kontraktion** — unter nur den im Hauptsatz formulierten Voraussetzungen.

An dieser Stelle lässt sich der **Banachsche Fixpunktsatz anwenden**, und es folgt, dass T^{k_0} genau einen Fixpunkt u in A besitzt. Wegen $T^{k_0}Tu = TT^{k_0}u = Tu$ ist auch Tu ein Fixpunkt von T^{k_0} in A, und in Anbetracht der Eindeutigkeit dieses Fixpunkts folgt $u = Tu$, also die Gültigkeit von (6.3) für u. Damit ist die eindeutige Lösbarkeit der Fixpunktgleichung (6.3) in $A = \mathrm{C}^0\big([t_0{-}\varepsilon, t_0{+}\varepsilon], \overline{\mathrm{B}}_r(y_0)\big)$ gezeigt (eindeutig, da jeder Fixpunkt von T auch Fixpunkt von T^{k_0} ist), und wegen der Äquivalenz (6.4) folgt sofort die eindeutige Lösbarkeit des AWPs (6.1) durch

eine differenzierbare Funktion $u\colon [t_0{-}\varepsilon, t_0{+}\varepsilon] \to \overline{\mathrm{B}}_r(y_0)$. Da die hiermit bewiesene Existenz- und Eindeutigkeitsaussage auch bei geringfügig verkleinertem r mit entsprechend verkleinertem ε anwendbar ist, hat u auf dem offenen Intervall $(t_0{-}\varepsilon, t_0{+}\varepsilon)$ Werte in der offenen Kugel $\mathrm{B}_r(y_0)$, und daraus ergibt sich die Eindeutigkeit dieser Lösung sogar unter allen differenzierbaren Funktionen $[t_0{-}\varepsilon, t_0{+}\varepsilon] \to \mathcal{X}$ (statt nur $[t_0{-}\varepsilon, t_0{+}\varepsilon] \to \overline{\mathrm{B}}_r(y_0)$). □

Bemerkung. Als ein Nebenprodukt des Beweises ergibt sich ein **iteratives Verfahren zur Berechnung von Näherungslösungen** mit zugehöriger A-priori-Fehlerabschätzung: Beginnt man mit einem beliebigen $u_0 \in \mathrm{C}^0\big([t_0{-}\varepsilon, t_0{+}\varepsilon], \overline{\mathrm{B}}_r(y_0)\big)$ (im einfachsten Fall mit $u_0 \equiv y_0$), so zeigt die Abschätzung

$$\|u{-}T^k u_0\|_\infty = \|T^k u{-}T^k u_0\|_\infty \le \frac{(\varepsilon L)^k}{k!}\|u{-}u_0\|_\infty \le \frac{(\varepsilon L)^k}{k!}(r + \|u_0{-}y_0\|_\infty)\,,$$

dass die sogenannten **Picard-Iterationen** $T^k u_0$ bei $k \to \infty$ gleichmäßig und über-exponentiell schnell gegen die Lösung u konvergieren.

6.2 Der Satz über die maximale Lösung

Es folgt ein weiterer zentraler Satz der Theorie, der sich aus dem Satz von Picard-Lindelöf ergibt und im Wesentlichen unter denselben Voraussetzungen gilt:

Hauptsatz 6.3 (über die **maximale Lösung**). *Die Strukturfunktion $f \in \mathrm{C}^0(D, \mathcal{X})$ auf offenem $D \subset \mathbb{R} \times \mathcal{X}$ sei lokal auf D bezüglich der $\mathcal{X}$-Variablen Lipschitz-stetig. Dann existiert zu jedem Anfangsdatum $(t_0, y_0) \in D$ ein maximales Intervall $I_{\max}$ mit $t_0 \in I_{\max}$, so dass das AWP*

$$u' = f(\,\cdot\,, u)\,, \qquad u(t_0) = y_0$$

*auf $I_{\max}$ eine (eindeutige) Lösung u besitzt. Das Intervall $\boldsymbol{I_{\max} = (\alpha, \omega)}$ **ist offen** mit $-\infty \le \alpha < t_0 < \omega \le \infty$, und **bei den Grenzübergängen $\boldsymbol{t \searrow \alpha}$ und $\boldsymbol{t \nearrow \omega}$ verlässt $\boldsymbol{(t, u(t))}$ jede kompakte Teilmenge von $\boldsymbol{D}$ endgültig**.*

Bemerkungen.

(1) Dabei heißt $u\colon I_{\max} \to \mathcal{X}$ die **maximale Lösung** und $I_{\max}$ das **maximale Lösungsintervall** des betrachteten AWPs. Sinnvoll sind diese Begriffe dabei nur, wenn Eindeutigkeit (und lokale Existenz) schon garantiert sind — hier durch Picard-Lindelöf.

(2) Die **wesentliche** (und auch am schwierigsten zu beweisende) **Aussage** des Satzes ist die über das **Verhalten bei $\boldsymbol{t \searrow \alpha}$ und $\boldsymbol{t \nearrow \omega}$**. Im endlich-dimensionalen Fall $D = \mathbb{R} \times \mathbb{K}^N$ impliziert dieses Verhalten, dass

- **entweder $\boldsymbol{\omega = \infty}$ („Lösung für alle Zukunft“)**
- **oder $\boldsymbol{\omega < \infty}$, $\boldsymbol{\lim_{t \nearrow \omega} |u(t)| = \infty}$ („Lösungsexplosion zur endlichen Zeit $\boldsymbol{\omega}$“)**

sowie eine analoge Alternative für $t \searrow \alpha$ statt $t \nearrow \omega$ gelten.

(3) Durch Reduktion auf Ordnung 1 lässt sich eine **Version** des Satzes **für AWPe beliebiger Ordnung $\boldsymbol{m} \in \mathbb{N}$** ableiten. Das Verhalten für $t \searrow \alpha$ und $t \nearrow \omega$ sieht dann so aus, dass $(t, u(t), u'(t), u''(t), u^{(m-1)}(t))$ jede kompakte Teilmenge von D endgültig verlässt, und eine Alternative wie in (2) gilt im Fall $D = \mathbb{R}\times(\mathbb{K}^N)^m$ mit $|u(t)|+|u'(t)|+|u''(t)|+\ldots+|u^{(m-1)}(t)|$ anstelle von $|u(t)|$ (also mit Lösungsexplosion im Phasenraum).

Beweis von Hauptsatz 6.3. Man erhält $\boldsymbol{I_{\max}}$ **als Vereinigung aller Intervalle** I in $\mathbb{R}$ mit $t_0 \in I$, **auf denen eine Lösung** u_I des AWPs **existiert**. Gemäß dem Satz von Picard-Lindelöf ist dann ...

- $\boldsymbol{I_{\max}}$ **ein Intervall positiver Länge** um t_0, das das Picard-Lindelöf-Intervall $[t_0-\varepsilon, t_0+\varepsilon]$ enthält;
- **die Zusammensetzung** (wegen Eindeutigkeit wohldefiniert)

  $$u(x) := u_I(x) \text{ für } x \in I$$

  der Lösungen u_I **eine Lösung** u des AWPs **auf ganz** $\boldsymbol{I_{\max}}$;
- $\boldsymbol{I_{\max} = (\alpha, \omega)}$ **offen**, denn wäre etwa der rechte Randpunkt ω in $I_{\max}$ enthalten, so wäre $(\omega, u(\omega)) \in D$ und man könnte die Lösung u auf $(\omega, \omega+\varepsilon]$ fortsetzen — im Widerspruch zur Definition von $I_{\max}$.

Zum **Nachweis des behaupteten Verhaltens bei** $\boldsymbol{t \nearrow \omega}$ argumentiert man durch ‚reductio ad absurdum': Angenommen es gäbe eine Folge $t_n \nearrow \omega$ mit $(t_n, u(t_n)) \in K$ für alle $n \in \mathbb{N}$ und ein Kompaktum $K \subset D \subset \mathbb{R}\times\mathcal{X}$. Dann ist $\omega < \infty$ wegen der Beschränktheit von K, und gemäß Extremalsatz existiert $M := \max_K |f| < \infty$. Außerdem gibt es ein $\delta > 0$, so dass für die offene δ-Umgebung $\mathrm{U}_\delta(K)$ von K bezüglich der Norm $\|(t,x)\|_{\mathbb{R}\times\mathcal{X}} := |t|+|x|$ gilt:

$$\overline{\mathrm{U}_\delta(K)} \subset \{(t,x) \in D \,:\, |f(t,x)| < 1+M\} \tag{6.6}$$

(denn K ist kompakt, und die Menge auf der rechten Seite der Inklusion ist eine *offene* Obermenge von K; somit hat K notwendig positiven Abstand zum Komplement dieser Obermenge).

Die nächste Behauptung ist, dass

$$\text{ein } n_0 \in \mathbb{N} \text{ existiert, so dass } (t, u(t)) \in \mathrm{U}_\delta(K) \text{ für alle } t \in (t_{n_0}, \omega) \text{ gilt,} \tag{6.7}$$

und dies wird mit einem weiteren (Widerspruch-im-)Widerspruchsargument gezeigt. Wäre (6.7) nämlich falsch, so gäbe es für jedes $n \in \mathbb{N}$ ein t_n^* mit $t_n < t_n^* < \omega$ und $(t_n^*, u(t_n^*)) \notin \mathrm{U}_\delta(K)$. Außerdem lässt sich annehmen, dass t_n^* sogar minimal mit dieser Eigenschaft gewählt wurde — dies geht wegen Stetigkeit von u, Abgeschlossenheit des Komplements von $\mathrm{U}_\delta(K)$, und weil t_n selbst wegen $(t_n, u(t_n)) \in K$ nicht in Frage kommt. Mit diesen Wahlen und dem Schrankensatz ergäbe sich

$$\delta \le \|(t_n^*, u(t_n^*)) - (t_n, u(t_n))\|_{\mathbb{R}\times\mathcal{X}} = (t_n^*-t_n) + |u(t_n^*)-u(t_n)| \le (t_n^*-t_n)\Big[1 + \sup_{(t_n, t_n^*)} |u'|\Big],$$

wobei das Maximum auf der rechten Seite wegen der Lösungseigenschaft $u' = f(\,\cdot\,, u)$, wegen der aus der Minimalität von t_n^* resultierenden Inklusion $(t, u(t)) \in U_\delta(K)$ für $t \in (t_n, t_n^*)$ und wegen (6.6) durch $1+M$ abgeschätzt werden könnte. Daher würde man mit

$$\delta \le (\omega-t_n)[2+M] \overset{n\to\infty}{\longrightarrow} 0$$

einen Widerspruch erhalten. Somit ist (6.7) gezeigt.

Aus der Lösungseigenschaft von u, (6.7) und (6.6) liest man ab, dass $|u'(t)| = |f(\,\cdot\,, u)| < 1+M$ bei $t \nearrow \omega$ beschränkt bleibt. Nach Schrankensatz und Cauchy-Kriterium existiert dann $y_\omega := \lim_{t\nearrow\omega} u(t)$ in $\mathcal{X}$. Unter erneuter Verwendung von (6.7) und (6.6) folgt $(\omega, y_\omega) \in \overline{\mathrm{U}_\delta(K)} \subset$

D, und nun kann der Satz von Picard-Lindelöf angewandt werden, um das AWP zu $u' = f(\,\cdot\,, u)$ mit AB $u(\omega) = y_\omega$ lokal zu lösen. Eine Lösung dieses AWPs existiert insbesondere auf einem Intervall $[\omega, \omega{+}\varepsilon]$ mit $\varepsilon > 0$ und kann mit u zu einer Lösung des ursprünglichen AWPs auf $(\alpha, \omega{+}\varepsilon]$ zusammengesetzt werden. Dies widerspricht der Definition von $I_{\max}$ und bestätigt, in Anbetracht der eingangs gemachten Annahme, das behauptete Verhalten beim Grenzübergang $t \nearrow \omega$.

Die Behauptung über das Verhalten für $t \searrow \alpha$ beweist man völlig analog. □

6.3 Kriterien für globale Existenz von Lösungen

Die *globale* Lösbarkeit von AWPen ist im Zusammenhang mit dynamischen Systemen von großem Interesse (genaueres demnächst in Abschnitt 6.6), sie kann aber nur in speziellen Fällen sichergestellt werden. Ein erster Fall mit globaler Existenz folgt:

Satz 6.4 (***Globaler* Existenzsatz mit Wachstumsannahme**). *Sei I Intervall positiver Länge in $\mathbb{R}$, und die Strukturfunktion $f \in \mathrm{C}^0(I \times \mathcal{X}, \mathcal{X})$ sei lokal auf $I \times \mathcal{X}$ bezüglich der $\mathcal{X}$-Variablen Lipschitz-stetig. Erfüllt f dann eine* ***Wachstumsbedingung in der $\mathcal{X}$-Variablen***

$$|f(t,x)| \le g(t)h(|x|) \qquad \text{für alle } t \in I\,,\ x \in \mathcal{X}$$

für stetiges $g\colon I \to \mathbb{R}_0^+$ und nicht-fallendes $h\colon \mathbb{R}_0^+ \to \mathbb{R}^+$ mit $\sum_{j=1}^{\infty} \frac{1}{h(j)} = \infty$, so ist das AWP

$$u' = f(\,\cdot\,, u)\,, \qquad u(t_0) = y_0 \tag{6.8}$$

für alle $t_0 \in I$, $y_0 \in \mathcal{X}$ ***auf dem* gegebenen *Intervall I*** *(eindeutig)* ***lösbar**.*

Bemerkungen.

(1) Insbesondere beinhaltet Satz 6.4 den Fall $h(s) = 1{+}s$ einer **linearen Wachstumsbedingung**, und dieser wohl wichtigste Spezialfall des Satzes **erfasst** wiederum **alle *linearen* GDGen und GDG-Systeme mit stetigen Koeffizienten**; für die detaillierte Darstellung der linearen Theorie sei aber ein weiteres Mal auf den späteren Abschnitt 7.1 verwiesen.

(2) Die Forderung $\sum_{j=1}^{\infty} \frac{1}{h(j)} = \infty$ kann alternativ durch die äquivalente Integralbedingung[6] $\int_{s_0}^{\infty} \frac{1}{h(s)}\,\mathrm{d}s = \infty$ für ein (und damit alle) $s_0 \in \mathbb{R}_0^+$ ausgedrückt werden. Diese Bedingung besagt, dass ***h* höchstens leicht schneller als linear wachsen** darf, und ist wesentliche und recht scharfe Voraussetzung des Satzes: Bei $h(s) = 1{+}s\log(1{+}s)$ ist die Bedingung noch erfüllt, bei $h(s) = 1{+}s^p$ mit $p > 1$ gelten sowohl sie als auch die allgemeine Aussage des Satzes nicht mehr (letzteres erkennt man z.B. an expliziten Lösungen von $u' = |u|^p$).

(3) Ein **analoger Sachverhalt** gilt **bei** GDGen $u^{(m)} = f(\,\cdot\,, u^{[m-1]})$ **beliebiger Ordnung $m \in \mathbb{N}$** unter der Wachstumsbedingung $|f(t, x_0, x_1, x_2, \ldots, x_{m-1})| \le g(t)h\big(\sum_{i=0}^{m-1} |x_i|\big)$ für alle $t \in I$, $x_0, x_1, x_2 \ldots, x_{m-1} \in \mathcal{X}$ und mit den gleichen Forderungen an g und h; dies beweist man durch Reduktion auf Ordnung 1.

Der Beweis des Satzes beruht auf einer **Analyse des genauen Existenzintervalls im Satz von Picard-Lindelöf**. Im Einzelnen argumentiert man wie folgt:

[6] Das Integral ist als uneigentliches Riemann-Integral zu verstehen; es ist wohldefiniert, da h als monotone Funktion jedenfalls über kompakte Intervalle Riemann-integrierbar ist.

Beweis. Zunächst sei I als kompakt angenommen. Dann ist g auf I beschränkt, und somit gilt $\sup_I g \leq K$ für eine Schranke $K \in \mathbb{R}^+$. Es folgt

$$|f(t,x)| \leq Kh(|y_0|+1) \qquad \text{für alle } (t,x) \in I \times \overline{\mathrm{B}}_1(y_0)\,,$$

und nach Picard-Lindelöf hat das AWP (6.8)

$$\text{eine Lösung } u \text{ auf } I \cap [t_0, t_0+\varepsilon_1]\,,\ \varepsilon_1 = \frac{1}{Kh(|y_0|+1)}\,,\ \text{mit Werten in } \overline{\mathrm{B}}_1(y_0)\,.$$

Ist $t_0+\varepsilon_1 \in I$, so wendet man im nächsten Schritt Picard-Lindelöf mit dem neuen Anfangsdatum $(t_0+\varepsilon_1, u(t_0+\varepsilon_1)) \in I \times \overline{\mathrm{B}}_1(y_0)$ an und setzt die so erhaltene Lösung mit Werten in $\overline{\mathrm{B}}_1(u(t_0+\varepsilon_1)) \subset \overline{\mathrm{B}}_2(y_0)$ an u heran. Wegen

$$|f(t,x)| \leq Kh(|y_0|+2) \qquad \text{für alle } (t,x) \in I \times \overline{\mathrm{B}}_1(u(t_0+\varepsilon_1))$$

findet man für dass AWP (6.8) somit auch

$$\text{eine Lösung } u \text{ auf } I \cap [t_0, t_0+\varepsilon_1+\varepsilon_2]\,,\ \varepsilon_2 = \frac{1}{Kh(|y_0|+2)}\,,\ \text{mit Werten in } \overline{\mathrm{B}}_2(y_0)\,.$$

Ist auch $t_0+\varepsilon_1+\varepsilon_2 \in I$, so kann man Picard-Lindelöf ein weiteres Mal anwenden und bekommt

$$\text{eine Lösung } u \text{ auf } I \cap [t_0, t_0+\varepsilon_1+\varepsilon_2+\varepsilon_3]\,,\ \varepsilon_3 = \frac{1}{Kh(|y_0|+3)}\,,\ \text{mit Werten in } \overline{\mathrm{B}}_3(y_0)\,.$$

Dieses Vorgehen lässt sich induktiv fortsetzen, und wegen der Divergenz der Reihe

$$\sum_{i=1}^{\infty} \varepsilon_i = \frac{1}{K} \sum_{i=1}^{\infty} \frac{1}{h(|y_0|+i)} \geq \frac{1}{K} \sum_{j=\lfloor |y_0| \rfloor+2}^{\infty} \frac{1}{h(j)} = \infty$$

ist dadurch die Existenz einer Lösung von (6.8) auf $I \cap [t_0, \infty)$ sichergestellt. Analog bekommt man eine Lösung auf $I \cap (-\infty, t_0]$ und durch noch ein letztes Aneinandersetzen auch eine auf ganz I. Für kompaktes I ist der Satz somit bewiesen, und bei nicht-kompaktem I erhält man die Lösung auf I einfach durch Zusammensetzen der Lösungen auf allen kompakten Teilintervallen von I, die t_0 enthalten (wobei die Zusammensetzung wegen Eindeutigkeit wohldefiniert ist). □

Bemerkung. Im vorausgehenden Beweis wurde der Satz von Picard-Lindelöf stets mit $r = 1$ angewandt. Im Fall linearen Wachstums $h(s) = 1+s$ ist auch ein etwas anderes Vorgehen sinnvoll, bei dem die i-te Anwendung des Satzes mit der Wahl $r_i := |u(t_0+\varepsilon_1+\varepsilon_2+\ldots+\varepsilon_{i-1})|+1$ geschieht; dann kann man bei kompaktem I oder beschränktem g tatsächlich erreichen, dass die Intervalllängen ε_i bei $i \to \infty$ nicht gegen Null gehen, sondern konstant bleiben.

Manchmal lässt sich die globale Lösbarkeit von GDGen und GDG-Systemen auch sicherstellen, indem man mit einer Art logischer Umkehr der Satzes über die maximale Lösung arbeitet. Auf einem formalen Level verwendet man dazu folgende direkte Konsequenz dieses Satzes:

> Sind bei einem gegebenen AWP $u' = f(\,\cdot\,, u)$, $u(t_0) = y_0$ die Voraussetzungen von Hauptsatz 6.3 an D, f, t_0, y_0 erfüllt, und bleibt $\{(t, u(t)) : t \in J\}$ für jede Lösung u des AWPs auf einem beschränkten Teilintervall J mit $t_0 \in J \subset I$ in einer kompakten Teilmenge von D, so ist das AWP auf ganz I lösbar.

Von konkretem Nutzen ist diese Herangehensweise vor allem in der endlich-dimensionalen Situation $\mathcal{X} = \mathbb{R}^N$. Dann nämlich folgt globale Existenz auf I, sobald alle Lösungen u des AWPs auf (beschränkten) Teilintervallen J von I beschränkt bleiben, und dies wiederum kann manchmal durch geometrische Betrachtungen sichergestellt werden. Tatsächlich identifiziert der nächste Satz drei geometrische Situationen bei autonomen GDG-Systemen, in denen globale Existenz gewährleistet ist (wobei die Beschränkung auf den autonomen Fall nur der Vereinfachung dient; im nicht-autonomen Fall gelten prinzipiell dieselben Aussagen).

Satz 6.5 (über **geometrische Kriterien für globale Existenz**). *Sei I ein Intervall positiver Länge in $\mathbb{R}$ mit $t_0 \in I$, und sei $F \in \mathrm{C}^0(D, \mathbb{R}^N)$ ein lokal Lipschitz-stetiges Vektorfeld auf einer offenen Teilmenge D von $\mathbb{R}^N$ mit $y_0 \in D$. Außerdem sei* **eine *der folgenden drei Bedingungen*** *erfüllt*:

(A) ***F ist tangential zu einer kompakten $\mathbf{C^1}$-Untermannigfaltigkeit M ohne Rand*** *von $\mathbb{R}^N$ (d.h. $F(x)$ ist für alle $x \in M$ ein Vektor im Tangentialraum $\mathrm{T}_x M$ an M in x), und es gilt $\boldsymbol{y_0 \in M \subset D}$.*

(B) ***F ist tangential zum Rand ∂G eines beschränkten $\mathbf{C^1}$-Gebiets G*** *in $\mathbb{R}^N$ (d.h. für alle $x \in \partial G$ gilt $F(x) \cdot \nu_G(x) = 0$, wobei $\nu_G(x)$ für den äußeren Einheitsnormalenvektor an G in x steht), und es gilt $\boldsymbol{y_0 \in \overline{G} \subset D}$.*

(C) ***F zeigt auf dem Rand ∂G eines beschränkten $\mathbf{C^1}$-Gebiets G*** *in $\mathbb{R}^N$* ***nirgends nach außen*** *(d.h. für alle $x \in \partial G$ gilt $F(x) \cdot \nu_G(x) \leq 0$), und es gelten $\boldsymbol{I \subset [t_0, \infty)}$ sowie $\boldsymbol{y_0 \in \overline{G} \subset D}$.*

Dann ist das AWP

$$u' = F(u)\,, \qquad u(t_0) = y_0$$

auf dem gegebenen Intervall I (eindeutig) lösbar.

Tatsächlich lässt sich in den durch den Satz abgedeckten Situationen einsehen, dass **Lösungen des AWPs in M beziehungsweise $\overline{G}$ gefangen bleiben** und die zuvor erwähnte Argumentation anwendbar ist. Dies wird auch anschaulich sehr plausibel, wenn man sich zugehörige Phasenraumportraits vorstellt; man denke etwa an den Fall einer Sphäre M oder einer Kugel G. Formal basiert der Beweis, dass M beziehungsweise $\overline{G}$ nicht verlassen werden kann, auf einer Form des „Geradebiegens“ von M beziehungsweise G und ist technisch nicht ganz einfach:

Beweis der Fälle (A), (B) und Beweisansatz zu Fall (C) von Satz 6.5. Im Fall (A) reicht es zu zeigen, dass es zu jeder Lösung u auf $J \subset I$ (oder auch nur auf $J = I_{\max}$) und jeder Stelle $\tau \in J$ mit $u(\tau) \in M$ ein kleines Intervall um τ gibt, auf dem u in M bleibt; denn dann folgt leicht, dass u auf ganz J in M und somit beschränkt bleibt. Man betrachtet hierzu, eventuell nach einer Drehung des $\mathbb{R}^N$, eine lokale Darstellung von M nahe $u(\tau)$ als Graph $\{(x, g(x)) : x \in \mathbb{R}^n\}$ von $g \in \mathrm{C}^1(\mathbb{R}^n, \mathbb{R}^{N-n})$ mit der Dimension n von M; das bedeutet mit etwas anderen Worten, man nimmt $\mathrm{B}_\gamma(u(\tau)) \cap M = \mathrm{B}_\gamma(u(\tau)) \cap \operatorname{Graph} g$ für ein $\gamma > 0$ an. Sei nun ein neues, Lipschitz-stetiges Vektorfeld H auf einer Umgebung von $(u_1(\tau), u_2(\tau), \ldots, u_n(\tau))$ in $\mathbb{R}^n$ definiert durch

$$H(z) := \big(F_1(z, g(z)), F_2(z, g(z)), \ldots, F_n(z, g(z))\big)\,,$$

und sei w die nach Picard-Lindelöf existente und eindeutige $\mathbb{R}^n$-wertige Lösung des AWPs

$$w' = H(w)\,, \qquad w(\tau) = (u_1(\tau), u_2(\tau), \ldots, u_n(\tau))$$

auf einem kleinen Intervall um τ. Die $\mathbb{R}^N$-wertige Funktion $\widetilde{u} := (w, g(w))$ erfüllt dann $\widetilde{u}(\tau) = u(\tau)$, hat nahe τ Werte in $\operatorname{Graph} g = M$ und löst aus folgendem Grund die ursprüngliche GDG $u' = F(u)$ auf einem kleinen Intervall um τ: Die ersten n Komponenten w' von $\widetilde{u}'$ stimmen per Konstruktion mit den ersten n Komponenten $H(w)$ von $F(w, g(w)) = F(\widetilde{u})$ überein; und da $\widetilde{u}'(t)$ und $F(\widetilde{u}(t))$ beide im (nicht-vertikalen) n-dimensionalen Tangentialraum $\mathrm{T}_{\widetilde{u}(t)}(\operatorname{Graph} g) = \mathrm{T}_{\widetilde{u}(t)}M$ liegen, stimmen die restlichen $(N-n)$-ten Komponenten ebenfalls überein. Gemäß Picard-Lindelöf folgt schließlich $u = \widetilde{u} \in M$ auf einem kleinen Intervall um τ, und dies reicht, wie schon erläutert, um über Hauptsatz 6.3 auf die globale Existenzaussage zu schließen.

Im Fall (B) liefert das schon Gezeigte mit $M = \partial G$, dass Lösungen mit Anfangsdatum in ∂G in ∂G bleiben. Aufgrund der Eindeutigkeitsaussage von Picard-Lindelöf folgt dann, dass Lösungen mit Anfangsdatum in G nie den Rand ∂G berühren und folglich das Äußere $\mathbb{R}^N \setminus \overline{G}$ nicht erreichen können. Somit bleiben alle Lösungen mit Anfangsdatum in $\overline{G}$ stets in $\overline{G}$, und der Beweis des Falls (B) ist komplett.

Im Fall (C) argumentiert man durch lokales Geradebiegen von G nahe $u(\tau) \in \partial G$ mit einem C^1-Diffeomorphismus und zugehörige Transformation der abhängigen Variablen bei der GDG (wobei eine Möglichkeit des Geradebiegens, die im Prinzip für (A) verwendet wurde, auf der Verwendung von Graphenabbildungen beruht). Anders als im Fall (A) kann man im Fall (C) aber nicht unbedingt auf ein kleineres *autonomes* GDG-System reduzieren, vielmehr reduziert man auf eine einzelne, skalare, allerdings *nicht mehr unbedingt autonome* GDG für die Normalenkomponente von u. Die Details der Reduktion und die Behandlung der nicht-autonomen GDG werden hier nicht ausgeführt, sollen aber (teilweise) in den Übungen thematisiert werden. □

6.4 Stetige Abhängigkeit und der Stetigkeitssatz

Dieser Abschnitt behandelt die Stabilität von Lösungen und ihre Abhängigkeit von gegebenen Anfangs- und Strukturdaten. Es geht hierbei allerdings vor allem um Lösungen auf kompakten Intervallen, also um die **Stabilität von Lösungen gegen Störungen bei Vorliegen eines endlichen Zeithorizonts**, *nicht* um die mit den Begrifflichkeiten aus Abschnitt 2.3 einhergehende Langzeit-Stabilität. Wie die Fragestellungen der vorigen Abschnitte ist auch die Stabilitätsfrage nur dann sinnvoll, wenn (zumindest für die ungestörte Lösung) Eindeutigkeit besteht. Daher wird im Folgenden die für den Satz von Picard-Lindelöf benötigte partielle Lipschitz-Bedingung (pLB) stets angenommen.

Hauptsatz 6.6 (**Stetige Abhängigkeit von den Daten** auf kompakten Intervallen)**.** *Sei D offen in $\mathbb{R} \times \mathcal{X}$, für $f \in \mathrm{C}^0(D, \mathcal{X})$ gelte lokal auf D eine pLB bezüglich der $\mathcal{X}$-Variablen, und sei $(t_0, y_0) \in D$. Ist dann u eine Lösung des AWPs*

$$u' = f(\,\cdot\,, u)\,, \qquad u(t_0) = y_0 \tag{6.9}$$

auf einem kompakten Intervall I positiver Länge, so gibt es zu jedem $\varepsilon > 0$ und jedem $L \in \mathbb{R}^+$ ein $\delta > 0$ mit folgender Eigenschaft: Wann immer $\widetilde{f} \in \mathrm{C}^0(V, \mathcal{X})$ auf einer offenen Umgebung V von $\operatorname{Graph}(u)$ in D eine pLB bezüglich der $\mathcal{X}$-Variablen mit Konstante L erfüllt und zusammen mit den gestörten Daten $\widetilde{t}_0 \in I$, $\widetilde{y}_0 \in \mathcal{X}$ den Bedingungen

$$\bigl|\widetilde{t}_0 - t_0\bigr| < \delta\,, \qquad \bigl|\widetilde{y}_0 - y_0\bigr| < \delta\,, \qquad \max_{\operatorname{Graph}(u)} \bigl|\widetilde{f} - f\bigr| < \delta \tag{6.10}$$

genügt, so existiert die (eindeutige) Lösung $\widetilde{u}$ des gestörten AWP

$$\widetilde{u}' = \widetilde{f}\big(\,\cdot\,,\widetilde{u}\,\big)\,, \qquad \widetilde{u}\big(\widetilde{t}_0\big) = \widetilde{y}_0 \tag{6.11}$$

auf ganz I und erfüllt

$$\max_I \big|\widetilde{u}-u\big| < \varepsilon\,.$$

Bemerkungen.

(1) **Hauptsatz 6.6 besagt im Wesentlichen, dass die Lösung des AWPs (6.9) stetig von der Strukturfunktion f und den Anfangsdaten t_0 und y_0 abhängt.**

(2) Eine genauere Analyse des folgenden Beweises zeigt, dass δ proportional zu ε gewählt werden kann und damit sogar eine lokal Lipschitz-stetige Abhängigkeit vorliegt.

(3) Der Übersichtlichkeit halber wurde der Hauptsatz nur für den Erster-Ordnung-Fall angegeben. Für Systeme höherer Ordnung gilt aber ein völlig analoger Sachverhalt; dies sieht man wie üblich mit dem Verfahren der Reduktion auf Ordnung 1.

(4) Mit einer geringfügigen Modifikation des folgenden Beweises kann man auch eine Version des Hauptsatzes erhalten, bei der δ eine etwas natürlichere Abhängigkeit aufweist: Es hängt dann nicht von der Konstante in der pLB für die gestörte Strukturfunktion $\widetilde{f}$, sondern vielmehr von der in der pLB für die ursprüngliche Strukturfunktion f auf der offenen Umgebung V von $\mathrm{Graph}(u)$ ab. Um dies erreichen zu können, muss allerdings die letzte Bedingung in (6.10) zur Forderung $\sup_V |\widetilde{f}-f| < \delta$ auf der *Umgebung* V verstärkt werden, was für manche Anwendungen — beispielsweise die Ableitung des folgenden Korollars 6.8 — zu einschränkend ist.

Zentrales Hilfsmittel für den Beweis des Hauptsatzes ist folgendes **Lemma über Integral- beziehungsweise Differential*un*gleichungen**, das auch viele weitere Anwendungen besitzt:

Lemma (Gronwall-Lemma). *Gegeben seien $-\infty < t_0 < \omega_0 \le \infty$, $a, b \in \mathrm{C}^0([t_0,\omega_0),\mathbb{R}_0^+)$ und $s_0 \in \mathbb{R}_0^+$. Erfüllt $\varphi \in \mathrm{C}^0([t_0,\omega_0),\mathbb{R})$ dann die Differentialungleichung*

$$\varphi(t) \le s_0 + \int_{t_0}^{t} \big[a(s)\varphi(s)+b(s)\big]\,\mathrm{d}s \qquad \textit{für alle } t \in (t_0,\omega_0)\,, \tag{6.12}$$

so folgt die Abschätzung

$$\varphi(t) \le \mathrm{e}^{A(t)}\left[\mathrm{e}^{-A(t_0)}s_0 + \int_{t_0}^{t} \mathrm{e}^{-A(s)}b(s)\,\mathrm{d}s\right] \qquad \textit{für alle } t \in [t_0,\omega_0) \tag{6.13}$$

mit einer Stammfunktion $A \in \mathrm{C}^0([t_0,\omega_0))$ von a.

Bemerkung. Tritt in (6.12) sogar Gleichheit ein, so ist die resultierende Integralgleichung äquivalent zum skalaren AWP $\varphi' = a\varphi+b$, $\varphi(t_0) = s_0$, dessen Lösung bekanntlich durch die rechte Seite von (6.13) gegeben ist. Das Lemma besagt also im Wesentlichen, dass sich Lösungen einer Differential*ungleichung* durch Lösungen der zugehörigen Differential*gleichung* abschätzen lassen. In Anbetracht dieses Zusammenhangs scheint die Aussage des Lemmas sehr plausibel und ist leicht zu merken.

Beweis des Lemmas. Zunächst wird der Fall $b \equiv 0$ behandelt. Es genügt dabei, $s_0 > 0$ zu betrachten, denn der Fall $s_0 = 0$ folgt daraus durch einen Grenzprozess. Außerdem lässt sich $\varphi \geq 0$ annehmen, denn andernfalls kann man φ unter Erhaltung der Ungleichung (6.12) durch $\max\{\varphi, 0\}$ ersetzen. Durch die Festlegung $\Psi(t) := s_0 + \int_{t_0}^t a(s)\varphi(s)\,\mathrm{d}s$ erhält man nun eine positive stetige Funktion Ψ auf $[t_0, \omega_0)$, und gemäß dem HDI und (6.12) mit $b \equiv 0$ erfüllt diese $\Psi' = a\varphi \leq a\Psi$ auf (t_0, ω_0) sowie $\Psi(t_0) = s_0$. Unter erneuter Anwendung des HDI erhält man

$$\log \Psi(t) - \log s_0 = \int_{t_0}^t \frac{\Psi'(s)}{\Psi(s)}\,\mathrm{d}s \leq \int_{t_0}^t a(s)\,\mathrm{d}s = A(t) - A(t_0) \qquad \text{für alle } t \in [t_0, \omega_0)\,.$$

Durch Exponentieren der resultierenden Ungleichung und Multiplikation mit s_0 ergibt sich

$$\Psi(t) \leq \mathrm{e}^{A(t)-A(t_0)} s_0 \qquad \text{für alle } t \in [t_0, \omega_0)\,.$$

In Anbetracht von $\varphi \leq \Psi$ folgt hieraus die Behauptung im Fall $b \equiv 0$.

Bei beliebigem $b \in \mathrm{C}^0([t_0, \omega_0), \mathbb{R}_0^+)$ arbeitet man, ähnlich wie schon in Abschnitt 5.1, mit den Hilfsfunktionen $B(t) := \int_{t_0}^t \mathrm{e}^{-A(s)} b(s)\,\mathrm{d}s$ und $\widetilde{\varphi} := \varphi - \mathrm{e}^A B$. Mit partieller Integration, den Identitäten $B(t_0) = 0$ und $\mathrm{e}^A B' = b$ sowie der Voraussetzung (6.12) ergibt sich dann

$$\begin{aligned}
s_0 + \int_{t_0}^t a(s)\widetilde{\varphi}(s)\,\mathrm{d}s &= s_0 + \int_{t_0}^t a(s)\varphi(s)\,\mathrm{d}s - \int_{t_0}^t a(s)\mathrm{e}^{A(s)} B(s)\,\mathrm{d}s \\
&= s_0 + \int_{t_0}^t a(s)\varphi(s)\,\mathrm{d}s - \mathrm{e}^{A(t)} B(t) + \mathrm{e}^{A(t_0)} B(t_0) + \int_{t_0}^t \mathrm{e}^{A(s)} B'(s)\,\mathrm{d}s \\
&= s_0 + \int_{t_0}^t \big[a(s)\varphi(s) + b(s)\big]\,\mathrm{d}s - \mathrm{e}^{A(t)} B(t) \\
&\geq \varphi(t) - \mathrm{e}^{A(t)} B(t) = \widetilde{\varphi}(t)
\end{aligned}$$

für alle $t \in (t_0, \omega_0)$. Somit ist die Voraussetzung (6.12) für $\widetilde{\varphi}$ anstelle von φ mit $b \equiv 0$ erfüllt, und mit dem bereits Gezeigten ergibt sich erst $\widetilde{\varphi}(t) \leq \mathrm{e}^{A(t)-A(t_0)} s_0$ für alle $t \in [t_0, \omega_0)$ und dann die behauptete Ungleichung für φ. □

Beweis von Hauptsatz 6.6. Seien die Daten beider AWPe des Satzes und die Lösung u sowie ε und L gegeben. Sei außerdem V die offene Umgebung von $\mathrm{Graph}(u)$, auf der die pLB mit Konstante L für $\widetilde{f}$ gilt. Da die kompakte Teilmenge $\mathrm{Graph}(u) \subset V$ positiven Abstand zum Komplement von V hat, lässt sich, eventuell nach Verkleinerung von ε, die Inklusion $\{t\} \times \mathrm{B}_\varepsilon(u(t)) \subset V$ für alle $t \in I$ erreichen. Außerdem sei angenommen, dass Kleinheitsbedingungen (6.10) mit noch zu fixierendem $\delta > 0$ gelten, das jedenfalls ausreichend klein ist, um $\big|\widetilde{y}_0 - u\big(\widetilde{t}_0\big)\big| \leq \big|\widetilde{y}_0 - y_0\big| + \big|u(t_0) - u\big(\widetilde{t}_0\big)\big| < \varepsilon$ und damit $\big(\widetilde{t}_0, \widetilde{y}_0\big) \in V$ zu gewährleisten. Sei dann $\widetilde{u}\colon (\alpha, \omega) \to \mathcal{X}$ die eindeutige maximale Lösung des gestörten AWPs (6.11) mit $-\infty \leq \alpha < \widetilde{t}_0 < \omega \leq \infty$, und sei $(\alpha_0, \omega_0) \subset (\alpha, \omega)$ das größte Intervall um $\widetilde{t}_0$ mit $\big(t, \widetilde{u}(t)\big) \in V$ für alle $t \in (\alpha_0, \omega_0)$. Mit Hilfe der Bedingungen (6.10) ergibt sich für $t \in (\alpha_0, \omega_0) \cap I$ die Abschätzung

$$\begin{aligned}
\big|\widetilde{u}(t) - u(t)\big| &= \left|\widetilde{y}_0 + \int_{\widetilde{t}_0}^t \widetilde{f}\big(s, \widetilde{u}(s)\big)\,\mathrm{d}s - y_0 - \int_{t_0}^{\widetilde{t}_0} f(s, u(s))\,\mathrm{d}s - \int_{\widetilde{t}_0}^t f(s, u(s))\,\mathrm{d}s\right| \\
&\leq \big|\widetilde{y}_0 - y_0\big| + \big|\widetilde{t}_0 - t_0\big| \max_{\mathrm{Graph}(u)} |f| + \int_{\widetilde{t}_0}^t \left(\big|\widetilde{f}(s, \widetilde{u}(s)) - \widetilde{f}(s, u(s))\big| + \max_{\mathrm{Graph}(u)} \big|\widetilde{f} - f\big|\right) \mathrm{d}s \\
&\leq \delta M + \int_{\widetilde{t}_0}^t \big[L\big|\widetilde{u}(s) - u(s)\big| + \delta\big]\,\mathrm{d}s\,,
\end{aligned}$$

wobei $M := 1+\max_{\mathrm{Graph}(u)}|f|$ abgekürzt wurde. Auf die Hilfsfunktion

$$\varphi := |\widetilde{u}-u|$$

kann nun das Gronwall-Lemma angewandt werden. Genauer wird die vorausgehende Version des Lemmas (mit $s_0 = \delta M$, $a \equiv L$, $b \equiv \delta$) auf $\big[\widetilde{t}_0, \omega_0\big) \cap I$ und eine entsprechende Variante auf $\big(\alpha_0, \widetilde{t}_0\big] \cap I$ angewandt. Nach kurzer Rechnung ergibt sich

$$\varphi(t) \leq \delta(M+L^{-1})\mathrm{e}^{L|t-\widetilde{t}_0|} - \delta L^{-1} \leq \delta(M+L^{-1})\mathrm{e}^{L|I|} \qquad \text{für alle } t \in (\alpha_0, \omega_0) \cap I$$

mit der endlichen Länge $|I|$ des *kompakten* Intervalls I. An dieser Stelle kann δ (abhängig von M, L, $|I|$, ε) so klein fixiert werden, dass zusätzlich zur oben verwendeten Inklusion $\big(\widetilde{t}_0, \widetilde{y}_0\big) \in V$ die Ungleichung $\delta(M+L^{-1})\mathrm{e}^{L|I|} \leq \varepsilon/2$ und damit insgesamt

$$\varphi(t) \leq \varepsilon/2 \qquad \text{für alle } t \in (\alpha_0, \omega_0) \cap I$$

gilt. Wäre nun $\omega > \omega_0 \in I$, so stünde $\varphi(\omega_0) \leq \varepsilon/2$ und damit $(\omega_0, \widetilde{u}(\omega_0)) \in \{\omega_0\} \times \mathrm{B}_\varepsilon(u(\omega_0)) \subset V$ im Widerspruch zur Wahl von ω_0 mit $(\omega_0, u(\omega_0)) \notin V$. Die Möglichkeit $\omega = \omega_0 \in I$ kann man im Fall $\dim_{\mathbb{K}} \mathcal{X} < \infty$ ebenfalls schnell ausschließen, denn die Inklusion $\big(t, \widetilde{u}(t)\big) \in K_\varepsilon := \bigcup_{s\in I}\big[\{s\} \times \overline{\mathrm{B}}_{\varepsilon/2}(u(s))\big]$ für alle $t \in \big[\widetilde{t}_0, \omega\big)$ mit dem dann kompakten $K_\varepsilon \subset V$ stünde im Widerspruch zum Satz über die maximale Lösung. Um den Fall $\dim_{\mathbb{K}} \mathcal{X} = \infty$ mitzubehandeln, muss man allerdings etwas sorgfältiger argumentieren, beispielsweise wie folgt. Zunächst ist $\widetilde{f}$ beschränkt auf dem Kompaktum $\mathrm{Graph}(u) \subset V$. Wäre nun $\omega = \omega_0 \in I$, so bliebe wegen der Inklusion $\big(t, \widetilde{u}(t)\big) \in \{t\} \times \overline{\mathrm{B}}_{\varepsilon/2}(u(t)) \subset V$ für $t \in \big[\widetilde{t}_0, \omega\big)$ und der pLB für $\widetilde{f}$ auf V auch $\widetilde{f}\big(t, \widetilde{u}(t)\big)$ bei $t \nearrow \omega$ beschränkt. Gemäß der GDG für $\widetilde{u}$ ist letzteres gleichbedeutend damit, dass $\widetilde{u}'(t)$ bei $t \nearrow \omega$ beschränkt bleibt, und wie im Beweis des Satzes über die maximale Lösung könnte man nun auf die Existenz von $y_\omega := \lim_{t\nearrow\omega} \widetilde{u}(t) \in \overline{\mathrm{B}}_{\varepsilon/2}(u(\omega))$ schließen. Wegen $(\omega, y_\omega) \in \{\omega\} \times \mathrm{B}_\varepsilon(u(\omega)) \subset V$ ließe sich der Satz von Picard-Lindelöf auf das AWP mit AB $\widetilde{u}(\omega) = y_\omega$ anwenden, und man könnte $\widetilde{u}$ als Lösung über ω hinaus fortsetzen. Dies stünde im Widerspruch zur Wahl von (α, ω) als maximales Lösungsintervall. Insgesamt ist damit sowohl $\omega > \omega_0 \in I$ als auch $\omega = \omega_0 \in I$ ausgeschlossen, und es verbleibt nur die Alternative, dass $\omega_0 \notin I$ gilt. Mit analogem Vorgehen erhält man $\alpha_0 \notin I$. Folglich ist $I \subset (\alpha_0, \omega_0)$, und $\widetilde{u}$ existiert auf ganz I mit

$$\max_I |\widetilde{u}-u| = \max_I \varphi < \varepsilon\,. \qquad \square$$

Als Nächstes soll der vorausgehende Satz statt auf ein einzelnes GDG-System auf eine Parameter-abhängige Familie solcher Systeme angewandt werden. Konkret geht es dabei um GDG-Systeme (für $\mathcal{X}$-wertige u)

$$u' = f(\,\cdot\,, u\,; p) \tag{6.14}$$

mit Parameter p, der in einer Grundmenge $\mathcal{P}$ läuft, und mit p-abhängiger Strukturfunktion $f\colon D \times \mathcal{P} \to \mathcal{X}$ auf $D \subset \mathbb{R} \times \mathcal{X}$. In diesem Kontext werden nun folgende Bezeichnungen eingeführt (die insbesondere auch im Fall eines einzelnen GDG-Systems, d.h. für einelementiges $\mathcal{P}$, sinnvoll sind und dann natürlich ohne das letzte Argument p verwendet werden).

Definition 6.7 (**Lebensdauerfunktionen, charakteristische Funktionen**). *Seien $\mathcal{P}$ eine Menge und D offen in $\mathbb{R} \times \mathcal{X}$, und für jedes $p \in \mathcal{P}$ genüge $f(\,\cdot\,; p) \in \mathrm{C}^0(D, \mathcal{X})$ lokal auf D*

einer pLB in der $\mathcal{X}$-Variablem. Für beliebige Daten $(\tau, y) \in D$, $p \in \mathcal{P}$ bezeichne $u(\,\cdot\,;\tau, y\,;p)$ die maximale $\mathcal{X}$-wertige Lösung des AWPs

$$u' = f(\,\cdot\,, u\,;p)\,, \qquad u(\tau) = y$$

auf dem maximalen Lösungsintervall $(\alpha(\tau, y\,;p), \omega(\tau, y\,;p))$. Man nennt die durch diese Konvention gegebenen Abbildungen $\alpha\colon D \times \mathcal{P} \to [-\infty, \infty)$ und $\omega\colon D \times \mathcal{P} \to (-\infty, \infty]$ die Lebensdauerfunktionen des Parameter-abhängigen Systems (6.14), und die $\mathcal{X}$-wertige Abbildung

$$(t\,;\tau, y\,;p) \mapsto u(t\,;\tau, y\,;p)$$

auf $D_u := \{(t\,;\tau, y\,;p) \in \mathbb{R} \times D \times \mathcal{P} \,:\, \alpha(\tau, y\,;p) < t < \omega(\tau, y\,;p)\}$ heißt die charakteristische Funktion des Systems (6.14).

Unter minimal stärkeren Annahmen (nämlich, wenn $\mathcal{P}$ eine Metrik trägt, f auch in p stetig ist und die pLB lokal gleichmäßig in p erfüllt) lässt sich die durch Hauptsatz 6.6 garantierte stetige Abhängigkeit von der Strukturfunktion $f(\,\cdot\,;p)$ nun in eine stetige Abhängigkeit vom Parameter p selbst umwandeln. Dies führt — zusammen mit der unverändert gültigen stetigen Abhängigkeit von Anfangsdaten und der Stetigkeit von Lösungen selbst — zu folgenden Aussagen.

Korollar 6.8 (Stetigkeitssatz). *Sei $\mathcal{P}$ metrischer Raum, sei D offen in $\mathbb{R} \times \mathcal{X}$, und erfülle $f \in \mathrm{C}^0(D \times \mathcal{P}, \mathcal{X})$ lokal auf $D \times \mathcal{P}$ eine pLB in der $\mathcal{X}$-Variablen. Dann ist die untere Lebensdauerfunktion α von (6.14) oberhalbstetig auf $D \times \mathcal{P}$, die obere Lebensdauerfunktion ω ist unterhalbstetig auf $D \times \mathcal{P}$, der Definitionsbereich D_u der charakteristischen Funktion ist offen, und die charakteristische Funktion u von (6.14) ist darauf stetig.*

Bemerkungen.

(1) Das Korollar überträgt sich auf (Parameter-abhängige) GDG-Systeme höherer Ordnung.

(2) **Die Ober- und Unterhalbstetigkeit von α und ω besagt, dass sich das maximale Existenzintervall bei kleinen Störungen der Anfangsdaten oder des Parameters nur wenig verkleinern, aber eventuell stark vergrößern kann.** Bei Grenzübergängen $(\tau, y\,;p) \to (\tau_0, y_0\,;p_0)$ bedeutet dies gerade umgekehrt, dass sich das maximale Existenzintervall nicht sprunghaft vergrößern, aber eventuell plötzlich verkleinern kann.

Trotz der zuvor beschriebenen einfachen Grundidee ist ein detaillierter Beweis des Korollars nicht völlig trivial. Daher wird er nun genauer ausgeführt:

Beweis von Korollar 6.8. Seien $(\tau_0, y_0) \in D$, $p_0 \in \mathcal{P}$ fixiert, und seien die Abkürzungen $\alpha_0 := \alpha(\tau_0, y_0\,;p_0)$, $\omega_0 := \omega(\tau_0, y_0\,;p_0)$ und $u_0 := u(\,\cdot\,;\tau_0, y_0\,;p_0)$ für die Grenzen des maximalen Lösungsintervalls und die maximale Lösung des AWPs zu diesen Daten vereinbart. Seien außerdem $t_0 \in (\alpha_0, \omega_0)$, ein beliebiges kompaktes Intervall I mit $t_0 \in I \subset (\alpha_0, \omega_0)$ sowie $\varepsilon > 0$ gegeben. Mit einem Überdeckungsargument findet man eine Umgebung V des Kompaktums $\mathrm{Graph}(u_0|_I)$ in D und ein $\delta_1 > 0$, so dass f global auf $V \times \mathrm{B}_{\delta_1}(p_0)$ einer pLB mit fester Konstante $L \in \mathbb{R}^+$ genügt. Nun kann Hauptsatz 6.6 für das AWP mit Strukturfunktion $f(\,\cdot\,;p_0)$ und Daten (τ_0, y_0) auf dem kompakten Intervall I angewendet werden. Zu ε und L liefert der Hauptsatz ein $\delta_2 > 0$, so dass für $(\tau, y) \in D$, $p \in \mathcal{P}$ und das zugehörige gestörte AWP die Implikation

$$\left.\begin{array}{r} |\tau - \tau_0| < \delta_2 \\ |y - y_0| < \delta_2 \\ d(p, p_0) < \delta_1 \\ \sup\limits_{t \in I} |f(t, u_0(t)\,;p) - f(t, u_0(t)\,;p_0)| < \delta_2 \end{array}\right\} \Longrightarrow \left\{\begin{array}{c} \alpha(\tau, y\,;p) \le \text{linkerRandpunkt}(I) \\ \omega(\tau, y\,;p) \ge \text{rechterRandpunkt}(I) \\ \max\limits_{t \in I} |u(t\,;\tau, y\,;p) - u_0(t)| < \varepsilon \end{array}\right.$$

gilt. Hierbei soll auch die letzte Bedingung der Hypothese nun auf eine explizite Kleinheitsbedingung für $d(p, p_0)$ zurückgeführt werden, und dies erfordert ein weiteres Überdeckungsargument. Tatsächlich kann das kompakte Intervall I durch endliche viele Kugeln $\mathrm{B}_{\delta_3}(t_1)$, $\mathrm{B}_{\delta_3}(t_2)$, …, $\mathrm{B}_{\delta_3}(t_n)$ mit festem Radius $\delta_3 > 0$ überdeckt werden, so dass $|f(t, u_0(t), p) -$

$f(t_i, u_0(t_i), p_0)| < \frac{1}{2}\delta_2$ für alle $t \in \mathrm{B}_{\delta_3}(t_i) \cap I$ und $p \in \mathrm{B}_{\delta_3}(p_0)$ gilt (wobei Stetigkeit von f in Punkten $(t, u(t), p_0)$, $t \in I$ ausgenutzt wurde). Durch Einschieben eines Terms $f(t_i, u_0(t_i), p_0)$ und eine Anwendung der Dreiecksungleichung ergibt sich dann für alle $t \in I$ und $p \in \mathcal{P}$ der gewünschte Zusammenhang

$$d(p, p_0) < \delta_3 \implies |f(t, u_0(t), p) - f(t, u_0(t), p_0)| < \delta_2 \,.$$

Insgesamt ist damit für $(\tau, y) \in D$ und $p \in \mathcal{P}$ die Implikation

$$\left.\begin{array}{c} |\tau - \tau_0| < \delta_2 \\ |y - y_0| < \delta_2 \\ d(p, p_0) < \min\{\delta_1, \delta_3\} \end{array}\right\} \implies \left\{\begin{array}{c} \alpha(\tau, y\,; p) \le \text{linkerRandpunkt}(I) \\ \omega(\tau, y\,; p) \ge \text{rechterRandpunkt}(I) \\ \max\limits_{t \in I} |u(t\,; \tau, y\,; p) - u_0(t)| < \varepsilon \end{array}\right. \tag{6.15}$$

verifiziert, und aus (6.15) liest man nun die Behauptungen ab. Da der linke Randpunkt von I beliebig nah an α_0 gewählt werden kann, ergibt sich mit $\limsup_{(\tau,y;p)\to(\tau_0,y_0;p_0)} \alpha(\tau, y\,; p) \le \alpha_0$ die Oberhalbstetigkeit von α. Analog folgt Unterhalbstetigkeit von ω, und als direkte Konsequenz ergibt sich, dass der Definitionsbereich D_u aus Definition 6.7 offen ist. Schließlich erhält man aus (6.15) und der Stetigkeit von u_0 noch $\lim_{(t;\tau,y;p)\to(t_0;\tau_0,y_0;p_0)} u(t\,; \tau, y\,; p) = \lim_{t\to t_0} u_0(t) = u_0(t_0)$ und somit die Stetigkeit der charakteristischen Funktion u. □

6.5 Der Differenzierbarkeitssatz

Dieser Abschnitt beschäftigt sich weiterhin mit Parameter-abhängigen GDG-Systemen (für $\mathcal{X}$-wertige u)

$$u' = f(\,\cdot\,, u\,; p) \tag{6.16}$$

mit Parameter $p \in \mathcal{P}$ und p-abhängiger Strukturfunktion $f\colon D \times \mathcal{P} \to \mathcal{X}$ auf $D \subset \mathbb{R} \times \mathcal{X}$. Statt um *stetige* Abhängigkeit wie im vorigen Abschnitt 6.4 geht es nun aber um stetig *differenzierbare* Abhängigkeit oder m.a.W. um C^1-Abhängigkeit der Lösungen beziehungsweise der charakteristischen Funktion u aus Definition 6.7 von der Zeitvariablen t, von Anfangsdaten (τ, y) und vom Parameter p (wobei die Frage nach differenzierbarer Abhängigkeit von p nur Sinn macht, wenn $\mathcal{P}$ Teilmenge eines normierten Raums ist). Tatsächlich wird eine solche C^1-Abhängigkeit der charakteristischen Funktion unter recht allgemeinen, aber im Vergleich zu Abschnitt 6.4 etwas stärkeren Voraussetzungen durch den folgenden Satz sichergestellt.

Hauptsatz 6.9 (Differenzierbarkeitssatz). *Sei $\mathcal{P}$ offene Teilmenge eines normierten Raums $\mathcal{Z}$, sei D offen in $\mathbb{R}\times\mathcal{X}$, und die Parameter-abhängige Strukturfunktion $f \in \mathrm{C}^0(D\times\mathcal{P}, \mathcal{X})$ besitze stetige Ableitungen $\mathrm{D}_x f \in \mathrm{C}^0(D\times\mathcal{P}, \mathcal{L}(\mathcal{X}, \mathcal{X}))$ und $\mathrm{D}_p f \in \mathrm{C}^0(D\times\mathcal{P}, \mathcal{L}(\mathcal{Z}, \mathcal{X}))$ nach den Variablen $x \in \mathcal{X}$ und $p \in \mathcal{P} \subset \mathcal{Z}$. Dann ist die charakteristische Funktion u des Parameter-abhängigen Systems (6.16) stetig differenzierbar auf ihrem (gemäß Stetigkeitssatz offenen) Definitionsbereich D_u, es gilt also $u \in \mathrm{C}^1(D_u, \mathcal{X})$.*

Bemerkungen.

(1) Die Existenz und Stetigkeit von $\mathrm{D}_x f$ implizieren die in Abschnitt 6.4 vorausgesetzte lokale pLB bezüglich x. Somit sind Definition 6.7 und der Stetigkeitssatz aus Korollar 6.8 anwendbar, und insbesondere sind die Lebensdauerfunktionen und die charakteristische Funktion in der Situation des Hauptsatzes wohldefiniert.

(2) Der Hauptsatz gilt analog für (Parameter-abhängige) GDG-Systeme höherer Ordnung.

(3) Die im Satz enthaltene C^1-Abhängigkeit der charakteristischen Funktion von der Zeitvariablen t bedeutet nichts anderes, als dass alle Lösungen C^1-Funktionen sind, und dies gilt (fast) trivial. Zwar sind die in dieser Vorlesung betrachteten Lösungen per Definition 4.1 zunächst nur differenzierbar, nicht stetig differenzierbar, aber aus der Gültigkeit der Gleichung (6.16)

folgt direkt, dass u' stetig und jede Lösung u somit C^1 ist. **Die entscheidenden Aussagen des Hauptsatzes sind** daher **die über C^1-Abhängigkeit der Lösungen von den Anfangsdaten (τ, y) und dem Parameter p.**

Übrigens zeigt dieselbe Argumentation, dass Lösungen u schon C^1 (in t) sind, wenn f nur stetig von den Variablen $(t,x) \in D$ abhängt (aber weder eine Differenzierbarkeitsvoraussetzung an f noch überhaupt irgendeine Annahme über seine Abhängigkeit von p gemacht wird). In ähnlicher Weise folgt auch für beliebiges $k \in \mathbb{N} \cup \{\infty\}$, dass Lösungen u sogar C^k (in t) sind, wenn f nur C^{k-1} von (t,x) abhängt; vergleiche auch mit dem Beweis des folgenden Korollars 6.10.

(4) Unter den Voraussetzungen von Hauptsatz 6.9 besitzen die partiellen Ableitungen $\partial_\tau u \in \mathrm{C}^0(D_u, \mathcal{X})$, $\mathrm{D}_y u \in \mathrm{C}^0(D_u, \mathcal{L}(\mathcal{X},\mathcal{X}))$, $\mathrm{D}_p u \in \mathrm{C}^0(D_u, \mathcal{L}(\mathcal{Z},\mathcal{X}))$ der charakteristischen Funktion u selbst wieder stetige Ableitungen $\partial_\tau u' \in \mathrm{C}^0(D_u, \mathcal{X})$, $\mathrm{D}_y u' \in \mathrm{C}^0(D_u, \mathcal{L}(\mathcal{X},\mathcal{X}))$ und $\mathrm{D}_p u' \in \mathrm{C}^0(D_u, \mathcal{L}(\mathcal{Z},\mathcal{X}))$ nach der t-Variablen (wobei für Ableitungen bezüglich dieser t-Variablen der übliche Ableitungsstrich notiert wird). Dies sieht man durch Differenzieren der GDG

$$u'(t\,;\tau,y\,;p) = f(t, u(t\,;\tau,y\,;p)\,;p)$$

nach τ, y, p, Verwendung der Kettenregel und Vertauschen der Differentiations-Reihenfolge. Differenziert[7] man mit der GDG auch die AB $u(\tau\,;\tau,y\,;p) = y$, so erhält man außerdem Charakterisierungen von $\partial_\tau u$, $\mathrm{D}_y u$, $\mathrm{D}_p u$ als Lösungen der abgeleiteten linearen (!!!) AWPe

$$\begin{aligned}
\partial_\tau u'(t) &= \mathrm{D}_x f(t, u(t\,;\tau,y,p);p)\,\partial_\tau u(t)\,, & \partial_\tau u(\tau) &= -f(\tau,y,p)\,,\\
\mathrm{D}_y u'(t) &= \mathrm{D}_x f(t, u(t\,;\tau,y,p);p)\,\mathrm{D}_y u(t)\,, & \mathrm{D}_y u(\tau) &= \mathrm{id}_{\mathcal{X}}\,,\\
\mathrm{D}_p u'(t) &= \mathrm{D}_x f(t, u(t\,;\tau,y,p);p)\,\mathrm{D}_p u(t) + \mathrm{D}_p f(t, u(t\,;\tau,y,p);p)\,, & \mathrm{D}_p u(\tau) &= 0
\end{aligned}$$

(der Übersichtlichkeit halber ohne die Argumente τ, y, p von $\partial_\tau u, \mathrm{D}_y u, \mathrm{D}_p u$ notiert).

Mit den AWPen der letzten Bemerkung und iterativer Anwendung des Hauptsatzes ergibt sich bei entsprechend regulärer Strukturfunktion auch die folgende Aussage über Höherer-Ordnung-differenzierbare Abhängigkeit von Anfangsdaten und Parametern.

Korollar 6.10 (C^k-Differenzierbarkeitssatz). *Ist $k \in \mathbb{N} \cup \{\infty\}$ und ist unter den Voraussetzungen des Hauptsatzes $f \in \mathrm{C}^{k-1}(D \times \mathcal{P}, \mathcal{X})$ bezüglich (x,p) sogar von der Klasse C^k mit auf $D \times \mathcal{P}$ stetigen Ableitungen, so folgt $u \in \mathrm{C}^k(D_u, \mathcal{X})$.*

Beweis von Korollar 6.10. Man argumentiert durch Induktion nach $k \in \mathbb{N}$ (und erhält den Fall $k = \infty$ dann als direkte Konsequenz). Der Induktionsanfang für $k = 1$ ist dabei schon durch den Hauptsatz erledigt. Für den Induktionsschluss sei das Korollar für ein $k \in \mathbb{N}$ richtig, und es sei $f \in \mathrm{C}^k(D \times \mathcal{P}, \mathcal{X})$ bezüglich (x,p) sogar von der Klasse C^{k+1} mit auf $D \times \mathcal{P}$ stetigen Ableitungen. Gemäß der angenommen Version des Korollars ist dann $u \in \mathrm{C}^k(D_u, \mathcal{X})$. Außerdem erfüllt $\partial_\tau u$ das zugehörige AWP der vorausgehenden Bemerkung (4) mit der durch $\widetilde{f}(t,x\,;\tau,y,p) := \mathrm{D}_x f(t, u(t\,;\tau,y,p);p)\,x$ definierten (τ,y,p)-abhängigen Strukturfunktion $\widetilde{f} \in \mathrm{C}^{k-1}(D \times D \times \mathcal{P})$, die bezüglich den Parametern (τ,y,p) sogar C^k mit auf $D \times D \times \mathcal{P}$ stetigen Ableitungen ist. Unter erneuter Verwendung der angenommenen Version des Korollars und, weil

[7] Die Differentiation der AB nach τ ergibt gemäß folgender Rechnung unter Verwendung von (6.16) die Identität $0 = \frac{\mathrm{d}}{\mathrm{d}\tau} u(\tau\,;\tau,y\,;p) = u'(\tau\,;\tau,y\,;p) + \partial_\tau u(\tau) = f(\tau, u(\tau\,;\tau,y\,;p)\,;p) + \partial_\tau u(\tau) = f(\tau,y\,;p) + \partial_\tau u(\tau)$. Die anderen relevanten Rechnungen sind noch deutlich einfacher.

auch der Anfangswert $-f(\tau,y,p)$ eine C^k-Abhängigkeit von (τ,y,p) aufweist, erhält man also $\partial_\tau u \in \mathrm{C}^k(D_u,\mathcal{X})$. Eine analoge Argumentation mit den anderen beiden AWPen der Bemerkung (4) liefert $\mathrm{D}_y u \in \mathrm{C}^k(D_u,\mathcal{L}(\mathcal{X},\mathcal{X}))$ und $\mathrm{D}_p u \in \mathrm{C}^k(D_u,\mathcal{L}(\mathcal{Z},\mathcal{X}))$, und direkt aus der Gleichung (6.16) ergibt sich $u' \in \mathrm{C}^k(D_u,\mathcal{X})$. Also sind alle Erster-Ordnung-Ableitungen von u von der Klasse C^k, und u selbst ist somit C^{k+1} auf D_u. Dies vervollständigt die Induktion und den Beweis des Korollars. □

Es bleibt noch der Hauptsatz zu beweisen. Aus Zeitgründen wird der Beweis im Rahmen dieser Vorlesung aber nicht mehr im Detail ausgeführt, sondern nur sehr kurz angedeutet:

Beweisskizze zu Hauptsatz 6.9. Um Differenzierbarkeit der charakteristischen Funktion u in der τ-Variablen nachzuweisen, betrachtet man ihre Differenzenquotienten

$$u_h(t\,;\tau,y\,;p) := \frac{u(t\,;\tau{+}h,y,p)-u(t\,;\tau,y,p)}{h}$$

der Schrittweite $h \in \mathbb{R}\setminus\{0\}$ (die jedenfalls für $(t\,;\tau,y,p),(t\,;\tau{+}h,y,p) \in D_u$ definiert sind). Für diese Bildungen erhält man, der Übersichtlichkeit halber mit unterdrückten Argumenten (y,p) von u_h und u,

$$\begin{aligned} u_h'(t\,;\tau) &= \frac{u'(t\,;\tau{+}h)-u'(t\,;\tau)}{h} \\ &= \frac{f(t,u(t\,;\tau{+}h)\,;p)-f(t,u(t\,;\tau)\,;p)}{h} \\ &= \int_0^1 \mathrm{D}_x f(t,(1{-}s)u(t\,;\tau)+su(t\,;\tau{+}h)\,;p)\,\mathrm{d}s\;\frac{u(t\,;\tau{+}h)-u(t\,;\tau)}{h} \\ &=: A(t\,;\tau,y,p\,;h)\,u_h(t\,;\tau)\,, \end{aligned}$$

wobei $A(t\,;\tau,y,p\,;h)$ als Abkürzung für den Integralausdruck in der vorletzten Zeile vereinbart wurde. Dies bedeutet, dass man die **Differenzenquotienten u_h als Lösungen der linearen AWPe**

$$u_h' = A(\,\cdot\,;\tau,y,p\,;h)u_h\,, \qquad u_h(\tau) = u_h(\tau\,;\tau,y,p) \tag{6.17}$$

mit Parametern $(\tau,y,p\,;h)$ ansehen kann, und es gilt im Folgenden zu zeigen, dass die für $h \in \mathbb{R}\setminus\{0\}$ definierte, in allen Parametern stetige Koeffizientenfunktion A und das Anfangsdatum $u_h(\tau\,;\tau,y,p)$ auch bei $h=0$ stetig ergänzt werden können. Für A erhält man mit dem definierenden Integralausdruck und der Stetigkeit von $\mathrm{D}_x f$ und u tatsächlich

$$\lim_{h\to 0} A(t\,;\tau,y,p\,;h) = \mathrm{D}_x f(t,u(t\,;\tau,y,p)\,;p)\,,$$

wobei die Konvergenz lokal gleichmäßig in $(t\,;\tau,y,p) \in D_u$ ist. Im Hinblick auf das Anfangsdatum rechnet man, wieder mit unterdrückten Argumenten (y,p) von u_h und u und unter Verwendung von $u(\tau\,;\tau) = y = u(\tau{+}h\,;\tau{+}h)$,

$$\begin{aligned} \lim_{h\to 0} u_h(\tau\,;\tau) &= \lim_{h\to 0}\frac{u(\tau\,;\tau{+}h)-u(\tau\,;\tau)}{h} = \lim_{h\to 0}\frac{u(\tau\,;\tau{+}h)-u(\tau{+}h\,;\tau{+}h)}{h} \\ &= -\lim_{h\to 0}\int_0^1 u'(\tau{+}sh\,;\tau{+}h)\,\mathrm{d}s = -\lim_{h\to 0}\int_0^1 f(\tau{+}sh,u(\tau{+}sh\,;\tau{+}h)\,;p)\,\mathrm{d}s \\ &= -f(\tau,u(\tau\,;\tau)\,;p) = -f(\tau,y\,;p)\,, \end{aligned}$$

wobei die Konvergenz lokal gleichmäßig in $(\tau, y, p) \in D \times \mathcal{P}$ ist. Insgesamt bedeutet dies, dass die AWPe in (6.17) für $h = 0$ stetig durch das (schon in Bemerkung (4) aufgetretene) AWP

$$u_0' = \mathrm{D}_x f(t, u(t\,;\tau, y, p)\,; p) u_0\,, \qquad u_0'(\tau) = -f(\tau, y, p) \tag{6.18}$$

ergänzt werden. Mit dem Stetigkeitssatz des Korollars 6.8 und unter Berücksichtigung der Stetigkeit der Anfangsdaten in (6.17) und (4) folgt, dass die charakteristische Funktion u_0 des Grenz-AWPs (6.18) stetig von $(t\,;\tau, y, p)$ abhängt und sich als Grenzwert von u_h bei $h \to 0$ ergibt. Insbesondere existiert also der Grenzwert der Differenzenquotienten und damit die partielle Ableitung $\partial_\tau u = \lim_{h\to 0} u_h = u_0 \in \mathrm{C}^0(D_u, \mathcal{X})$. In ähnlicher Weise verifiziert man die Existenz und Stetigkeit von $\mathrm{D}_y u$ und $\mathrm{D}_p u$, Stetigkeit von u' erhält man — wie zuvor schon begründet — (sehr) einfach. Da somit alle Erster-Ordnung-Ableitungen von u existieren und stetig sind, ist u von der Klasse C^1. □

6.6 Gewöhnliche Differentialgleichungen als kontinuierliche dynamische Systeme

Zusammengenommen zeigen die Resultate der Abschnitte 6.1 bis 6.4, dass die Lösungen eines autonomen GDG-Systems (in vielen Fällen) ein kontinuierliches dynamisches System im Sinne von Definition 1.1 bilden. Um eine in großer Allgemeinheit gültige, *lokale* Version dieser Aussage ausformulieren zu können, wird folgende Variante der Begriffsbildung aus Definition 1.1 benötigt:

Definition 6.11 (lokale Flüsse). *Seien $\mathbb{T} \in \{\mathbb{R}_0^+, \mathbb{R}\}$ und D ein metrischer Raum. Eine stetige Abbildung*

$$\Phi\colon D_\Phi \to D\,,$$

*auf einer offenen Teilmenge $D_\Phi = \bigcup_{x\in D} \big(I_x \times \{x\}\big)$ von $\mathbb{T} \times D$ mit Intervallen I_x um 0 nennt man einen **lokalen Fluss** auf D, wenn die Identitätseigenschaft $\Phi(0, x) = x$ für $x \in D$ erfüllt ist, und wenn für $x \in D$, $t \in I_x$, $s \in I_{\Phi(t,x)}$ stets $s{+}t \in I_x$ und die Halbgruppeneigenschaft $\Phi(s{+}t, x) = \Phi(s, \Phi(t, x))$ gelten.*

Bemerkung. Anders als bei einem globalen Fluss muss eine Flusslinie $\Phi(\,\cdot\,, x)$ eines lokalen Flusses Φ nicht auf ganz $\mathbb{T}$, sondern nur auf dem Intervall I_x definiert sein. Sie kann also nach endlicher Zeit aufhören zu existieren; dies ist der Unterschied zu Definition 1.1.

Mit dieser Terminologie lassen sich Teile der Abschnitte 6.1, 6.3, 6.4 in folgender Weise zusammenstellen (wobei — daran sei hier erinnert — $\mathcal{X}$ nach wie vor für einen beliebigen Banach-Raum steht):

Satz 6.12 („Von autonomem GDGen zu kontinuierlichen dynamischen Systemen“). *Für ein lokal Lipschitz-stetiges Vektorfeld $F \in \mathrm{C}^0(D, \mathcal{X})$ auf offenem $D \subset \mathcal{X}$ sei $u\colon D_u \to \mathcal{X}$ die **charakteristische Funktion des autonomen GDG-Systems***

$$u' = F(u)\,.$$

Dann erhält man durch die Festlegungen

$$D_\Phi := \{(t, x) \in \mathbb{R} \times D \,:\, \alpha(0, x) < t < \omega(0, x)\} = \{(t, x) \in \mathbb{R} \times \mathcal{X} \,:\, (t\,; 0, x) \in D_u\}$$

und

$$\Phi(t,x) := u(t\,;0,x) \qquad \textit{für } (t,x) \in D_\Phi \tag{6.19}$$

einen ***lokalen Fluss*** Φ *auf* D *mit Zeitmenge* $\mathbb{T} = \mathbb{R}$. *Darüber hinaus erhält man* ***sogar einen globalen Fluss*** *zu einem dynamischen System des Typs ...*

- $(\mathbb{R}, \mathcal{X}, \Phi)$, ***vorausgesetzt dass*** $D = \mathcal{X}$ *ist und* F *höchstens linear wächst* (*oder wie in Satz 6.4 nur sehr leicht schneller*),
- $(\mathbb{R}, M, \Phi)$, ***vorausgesetzt dass*** $\mathcal{X} = \mathbb{R}^N$ *gilt und* F *tangential zu einer kompakten* C^1*-Untermannigfaltigkeit* M *ohne Rand von* $\mathbb{R}^N$ *mit* $M \subset D$ *ist,*
- $(\mathbb{R}, \overline{G}, \Phi)$, ***vorausgesetzt dass*** $\mathcal{X} = \mathbb{R}^N$ *gilt und* F *tangential zum Rand eines beschränkten* C^1*-Gebiets* G *in* $\mathbb{R}^N$ *mit* $\overline{G} \subset D$ *ist,*
- $(\mathbb{R}_0^+, \overline{G}, \Phi)$, ***vorausgesetzt dass*** $\mathcal{X} = \mathbb{R}^N$ *ist und* F *auf dem Rand eines beschränkten* C^1*-Gebiets* G *in* $\mathbb{R}^N$ *mit* $\overline{G} \subset D$ *nirgends nach außen zeigt.*

Die globalen Aussagen beinhalten hierbei, dass der Definitionsbereich D_Φ *ganz* $\mathbb{R} \times \mathcal{X}$, *eine Obermenge von* $\mathbb{R} \times M$, *eine Obermenge von* $\mathbb{R} \times \overline{G}$ *beziehungsweise eine Obermenge von* $\mathbb{R}_0^+ \times \overline{G}$ *ist* (*und in den letzten drei Fällen ist der globale Fluss, genau genommen, die* Einschränkung *von* Φ *auf* $\mathbb{R} \times M$, $\mathbb{R} \times \overline{G}$ *beziehungsweise* $\mathbb{R}_0^+ \times \overline{G}$).

Bemerkungen.

(1) Im Wesentlichen besagt der Satz, dass man einen lokalen (und unter gewissen Umständen sogar globalen) Fluss Φ durch Einschränkung der charakteristischen Funktion u aus Definition 6.7 von $D_u \subset \mathbb{R} \times \mathbb{R} \times \mathcal{X}$ auf $D_\Phi = D_u \cap (\mathbb{R} \times \{0\} \times \mathcal{X})$, also durch Beschränkung auf die Anfangszeit $\tau = 0$, **erhält**. Mit einfacheren Worten bedeutet dies gerade, dass **jede Flusslinie $\boldsymbol{\Phi(\,\cdot\,, x)}$ von $\boldsymbol{\Phi}$ die eindeutige Lösung des zu $\boldsymbol{x \in D}$ gehörigen AWPs $\boldsymbol{u' = F(u)}$, $\boldsymbol{u(0) = x}$ ist**. Der Fluss Φ ist daher charakterisiert durch die **Flussgleichungen**

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t,x) = F(\Phi(t,x)) \qquad \text{für } (t,x) \in D_\Phi$$

und die Anfangsbedingung $\Phi(0,x) = x$ für $x \in D$.

(2) In der Literatur wird eine globale Version des Satzes auf $D = \mathcal{X}$ oft mit einer *globalen Lipschitz-Bedingung*

$$|F(\widetilde{x}) - F(x)| \le L|\widetilde{x} - x| \qquad \text{für } \textit{alle } \widetilde{x}, x \in \mathcal{X}$$

als einziger Voraussetzung angegeben. Diese Bedingung impliziert einerseits lokale Lipschitz-Stetigkeit und andererseits höchstens lineares Wachstums von F, daher sind solche Versionen in der hier gemachten Aussage enthalten.

Beweis von Satz 6.12. Die aus dem Satz von Picard-Lindelöf bekannte lokale Lösbarkeit von AWPen garantiert $\alpha(0,x) < 0 < \omega(0,x)$ für alle $x \in D$, so dass D_Φ die in Definition 6.11 verlangte Form mit $I_x = (\alpha(0,x), \omega(0,x)) \ni 0$ (für $\mathbb{T} = \mathbb{R}$) hat. Gemäß Korollar 6.8 ist außerdem D_u und damit auch D_Φ offen sowie die charakteristische Funktion u und damit Φ in (6.19) stetig. Die Identitätseigenschaft $\Phi(0,x) = u(0\,;0,x) = x$ ist aus Definition 6.7 klar, und die Halbgruppeneigenschaft von Φ ergibt sich aus der Beobachtung

$$\Phi(s{+}t,x) = u(s{+}t\,;0,x) = u(s{+}t\,;t,u(t\,;0,x)) = u(s\,;0,u(t\,;0,x)) = \Phi(s,\Phi(t,x))\,,$$

wobei im zweiten Schritt die Eindeutigkeit der Lösung mit den Anfangsdaten $(t, u(t\,;0,x))$ und im dritten Schritt die Invarianz der Lösungen *autonomer* Systeme unter Zeit-Verschiebung ausgenutzt wurden. Damit ist Φ ein lokaler Fluss mit Zeitmenge $\mathbb{T} = \mathbb{R}$.

Die globalen Aussagen ergeben sich direkt aus den Sätzen 6.4 und 6.5, denn diese garantieren in den aufgelisteten Situationen für alle $x \in \mathcal{X}$, $x \in M$ beziehungsweise $x \in \overline{G}$, dass $\omega(0,x) = \infty$ und, abgesehen vom letzten gelisteten Fall, auch $\alpha(0,x) = -\infty$ eintritt. □

Tatsächlich gilt auch eine elementar zu beweisende **Umkehrung zum vorausgehenden Satz**. Diese besagt im Wesentlichen, dass alle kontinuierlichen dynamischen Systeme (auf Teilmengen eines Banach-Raums und unter einer schwachen Differenzierbarkeitsvoraussetzung) von autonomen GDG-Systemen herrühren:

Satz 6.13 („Von kontinuierlichen dynamischen Systemen zu autonomen GDGen“). *Sei $(\mathbb{T}, D, \Phi)$ mit $\mathbb{T} \in \{\mathbb{R}_0^+, \mathbb{R}\}$ ein kontinuierliches dynamisches System auf $D \subset \mathcal{X}$, so dass die partielle Ableitung $\partial_t\Phi$ von Φ nach der $\mathbb{T}$-Variablen auf $\mathbb{T} \times D$ existiert. Für das Vektorfeld*

$$F := \partial_t\Phi(0, \cdot\,)$$

gelten dann die Flussgleichungen

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t,x) = F(\Phi(t,x)) \qquad \textit{für alle } (t,x) \in \mathbb{T} \times D\,,$$

d.h. jede Bahnlinie $\Phi(\,\cdot\,, x)$ von Φ mit $x \in D$ löst auf $\mathbb{T}$ das AWP zum autonomen GDG-System $u' = F(u)$ mit AB $u(0) = x$.

Beweis. Die Halbgruppeneigenschaft des Flusses Φ besagt

$$\Phi(s{+}t, x) = \Phi(s, \Phi(t,x)) \qquad \text{für alle } s,t \in \mathbb{T} \text{ und } x \in D\,.$$

Durch Ableiten dieser Gleichung nach s an der Stelle $s = 0$ ergibt sich

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t,x) = (\partial_t\Phi)(0, \Phi(t,x)) \qquad \text{für alle } (t,x) \in \mathbb{T} \times D\,,$$

und dies ist die behauptete Lösungseigenschaft. □

Bemerkungen.

(1) Auch wenn die Existenz von $\partial_t\Phi$ nur in den Punkten $(0,x)$ mit $x \in D$ vorausgesetzt wird, zeigt die vorausgehende Argumentation, dass der Satz richtig bleibt und $\partial_t\Phi$ automatisch auf ganz $\mathbb{T} \times D$ existiert. Als Nebenprodukt sieht man dabei auch, dass Stetigkeit von $F = \partial_t\Phi(0, \cdot\,)$ auf D automatisch Stetigkeit von $\partial_t\Phi$ auf ganz $\mathbb{T} \times D$ impliziert.

(2) Mit derselben Argumentation erhält man auch folgende Version des Satzes für *lokale* Flüsse: Sei $\Phi\colon D_\Phi \to D$ lokaler Fluss auf $D \subset \mathcal{X}$ mit auf $\{0\} \times D$ existenter partieller Ableitung $\partial_t\Phi$. Mit dem Vektorfeld $F := \partial_t\Phi(0, \cdot\,)$ gilt dann $\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t,x) = F(\Phi(t,x))$ für alle $(t,x) \in D_\Phi$.

Zum Abschluss dieses Abschnitts sei erwähnt, dass man auch ***nicht-autonome* GDG-Systeme als dynamische Systeme** auffassen kann. In gewisser Hinsicht braucht man hierzu ein allgemeineres Konzept eines Flusses $\Phi\colon \mathbb{T} \times \mathbb{R} \times \mathcal{X} \to \mathcal{X}$ mit $\mathbb{T} \in \{\mathbb{R}_0^+, \mathbb{R}\}$, bei dem sich der Zustand $\Phi(t,\tau,x)$ während des Zeitintervalls $[\tau, \tau{+}t]$ der Länge $t \in \mathbb{T}$ aus dem Inititalzustand

$x \in \mathcal{X}$ zum Initialzeitpunkt $\tau \in \mathbb{R}$ entwickelt; und bei einem gegebenen GDG-System (mit globaler Existenz) ist dann $\Phi(t, \tau, x)$ nichts anderes als $u(\tau+t\,;\tau, x)$ mit der charakteristischen Funktion u des GDG-Systems. Der zuvor betrachtete Fall autonomer GDG-Systeme beziehungsweise konstanter äußerer Umstände gliedert sich hierbei als ein Spezialfall ein, in dem man durch $u(\tau+t\,;\tau, x) = u(t\,;0, x)$ beziehungsweise $\Phi(t, \tau, x) = \Phi(t, 0, x)$ auf den Intitalzeitpunkt 0 reduzieren und die zweite Variable somit vernachlässigen kann.

Auf einer formalen Ebene erweist es sich aber trotz allem als unnötig, ein allgemeineres Konzept dynamischer Systeme zu prägen, denn man kann **ein *nicht-autonomes* AWP $\boldsymbol{u' = f(\,\cdot\,, u),\ u(t_0) = y_0}$ stets als *autonomes* AWP $\boldsymbol{\overline{u}' = (1, f(\overline{u})),\ \overline{u}(t_0) = (t_0, y_0)}$ für $\boldsymbol{\overline{u} := (u_0, u)}$ auffassen**, bei dem man die zusätzliche Komponentenfunktion u_0 zur unbekannten Funktion u und die Gleichung $u_0' \equiv 1$ zum GDG-System hinzugefügt hat. Im Kontext dynamischer Systeme bedeutet dies, dass man anstelle des Phasenraums $\mathcal{X}$ den erweiterten Phasenraum $\mathbb{R} \times \mathcal{X}$ zum Zustandsraum erhebt und auf $\mathbb{R} \times \mathcal{X}$ den zugehörigen, im Rahmen von Definition 1.1 bleibenden Fluss $\overline{\Phi}$ mit $\overline{\Phi}(t, (\tau, x)) = \overline{u}(t\,;0, (\tau, x)) = (\tau+t, u(\tau+t\,;\tau, x))$ betrachtet; dieser Fluss $\overline{\Phi}$ beinhaltet dann als $\mathcal{X}$-Komponente die zuvor erwähnte Abbildung Φ der drei Variablen (t, τ, x).

Kapitel 6 wird an dieser Stelle unterbrochen und erst auf Seite 121 fortgesetzt.

Kapitel 7

Lineare GDG-Systeme

7.1 Allgemeine Theorie linearer GDG-Systeme

Hier seien $\mathcal{X}$ (mit Norm $|\cdot| = \|\cdot\|_{\mathcal{X}}$) und $\mathcal{Z}$ beliebige normierte Räume über $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, I sei ein Intervall positiver Länge in $\mathbb{R}$, und es sei erinnert an die allgemeine Form eines linearen GDG-Systems

$$A_m u^{(m)} + A_{m-1} u^{(m-1)} + \ldots + A_2 u'' + A_1 u' + A_0 u = b \qquad \text{auf } I \tag{7.1}$$

mit gegebenen Funktionen $A_0, A_1, A_2, \ldots, A_{m-1}, A_m \colon I \to \mathcal{L}(\mathcal{X}, \mathcal{Z})$ und $b \colon I \to \mathcal{Z}$. Im wichtigsten Fall $\mathcal{X} = \mathbb{K}^N$, $\mathcal{Z} = \mathbb{K}^M$ sind die Koeffizienten $A_0, A_1, A_2, \ldots, A_{m-1}, A_m$ einfach t-abhängige reelle oder komplexe $(M \times N)$-Matrizen, und die Inhomogenität b ist ein t-abhängiger reeller oder komplexer M-Vektor.

Mit dem Differentialoperator $P = \sum_{k=0}^{m} A_k \frac{\mathrm{d}^k}{\mathrm{d}t^k}$, der eine m-mal differenzierbare Funktion $u \colon I \to \mathcal{X}$ in die Funktion $Pu = \sum_{k=0}^{m} A_k u^{(k)} \colon I \to \mathcal{Z}$ überführt, kann die Notation verkürzt werden, und das GDG-System (7.1) lässt sich dann einfach durch $Pu = b$ auf I ausdrücken. Unter Verwendung dieser Schreibweise seien folgende Beobachtungen über lineare GDGen festgehalten:

Satz 7.1 (**Grundeigenschaften bei linearen GDGen**). *Für $P = \sum_{k=0}^{m} A_k \frac{\mathrm{d}^k}{\mathrm{d}t^k}$ mit gegebenen $A_0, A_1, A_2, \ldots, A_{m-1}, A_m \colon I \to \mathcal{L}(\mathcal{X}, \mathcal{Z})$ und $b \colon I \to \mathcal{Z}$ gelten:*

(I) *P ist ein $\mathbb{K}$-**linearer Operator**, d.h. für m-mal differenzierbare $u_j \colon I \to \mathcal{X}$ und $C_j \in \mathbb{K}$ gilt $P(C_1 u_1 + C_2 u_2 + \ldots + C_n u_n) = C_1 P u_1 + C_2 P u_2 + \ldots + C_n P u_n$.*

(II) *Die **Lösungen der** homogenen **GDG** $Pu \equiv 0$ auf I **bilden einen** $\mathbb{K}$**-Vektorraum**, nämlich einen Untervektorraum des Raums m-mal differenzierbarer Funktionen $I \to \mathcal{X}$. Mit anderen Worten gilt das sogenannte **Superpositionsprinzip**: Für Lösungen u_j zu $Pu \equiv 0$ und $C_j \in \mathbb{K}$ ist auch $C_1 u_1 + C_2 u_2 + \ldots + C_n u_n$ stets Lösung, und insbesondere ist die Null-Funktion stets eine Lösung zu $Pu \equiv 0$, genannt die **triviale Lösung**.*

(III) *Ist u_{sp} eine spezielle Lösung zu $Pu = b$ auf I, so hat die* **inhomogene** ***GDG*** *$Pu = b$ auf I **als Lösungsraum den** **affinen** **Unterraum***

$$u_{\mathrm{sp}} + \mathscr{L}_{\mathrm{hom}} = \{u_{\mathrm{sp}} + \widetilde{u} \,:\, \widetilde{u} \in \mathscr{L}_{\mathrm{hom}}\}$$

des Raums m-mal differenzierbarer Funktionen $I \to \mathcal{X}$, wobei $\mathscr{L}_{\mathrm{hom}}$ den Lösungsraum der homogenen Gleichung $Pu \equiv 0$ auf I bezeichnet.

(IV) *Sind die* ***Koeffizienten*** A_k **reelle** $(M\times N)$-***Matrizen*** *und ist die* ***Inhomogenität*** b *ein* **reeller** M-***Vektor****, so* ***sind die Realteile der komplexen Lösungen*** *zu* $Pu = b$ (*diese entsprechen der Auffassung* $\mathcal{X} = \mathbb{C}^N$, $\mathcal{Z} = \mathbb{C}^M$) ***genau die reellen Lösungen*** *zu* $Pu = b$ (*diese entsprechen* $\mathcal{X} = \mathbb{R}^N$, $\mathcal{Z} = \mathbb{R}^M$)*, und die Imaginärteile der komplexen Lösungen zu* $Pu = b$ *sind genau die reellen Lösungen der* homogenen *Gleichung* $Pu \equiv 0$.

Beweis. Die Behauptung (I) ergibt sich aus der Linearität der Ableitungsoperatoren $\frac{\mathrm{d}^k}{\mathrm{d}t^k}$ und Rechenregeln für die Matrix-Vektor-Multiplikation, und (II), (III) sind dann direkte Folgen von (I). Die letzte Behauptung (IV) schließlich erhält man aus der Beobachtung, dass die Gleichung $Pu = b$ durch Real- und Imaginärteilbildung auf beiden Seiten in $P(\operatorname{Re} u) = b$ und $P(\operatorname{Im} u) \equiv 0$ aufgespalten werden kann. □

Bemerkungen.

(1) Teil (III) des Satzes besagt insbesondere, **dass die allgemeine Lösung zu** $\boldsymbol{Pu = b}$ **die Summe einer speziellen Lösung zu** $\boldsymbol{Pu = b}$ **und der allgemeinen Lösung zu** $\boldsymbol{Pu \equiv 0}$ **ist**. Dies wurde auch früher (etwa in Abschnitt 5.2) schon behauptet und konnte in Beispielen beobachtet werden, die Kenntnis der allgemeinen Tatsache kann aber natürlich helfen, Rechnungen zu vereinfachen und den Überblick zu bewahren.

(2) Für A_k, b wie im Satz mit allgemeinen normierten Räumen $\mathcal{X}$, $\mathcal{Z}$ über $\mathbb{K} = \mathbb{R}$ kann (IV) ebenfalls formuliert werden. Komplexe Lösungen sind dann solche mit Werten in der Komplexifizierung $\mathcal{X}+\mathring{\mathrm{i}}\mathcal{X}$ von $\mathcal{X}$ (wobei zur Formulierung der GDG für $(\mathcal{X}+\mathring{\mathrm{i}}\mathcal{X})$-wertige u auch die Komplexifizierung $\mathcal{Z}+\mathring{\mathrm{i}}\mathcal{Z}$ von $\mathcal{Z}$ herangezogen wird).

Der **wichtigste Fall** bei linearen GDG-Systemen ist natürlich der Fall $\mathcal{Z} = \mathcal{X}$, und das bedeutet im Matrizenfall $\mathcal{Z} = \mathcal{X} = \mathbb{K}^N$, dass alle $\boldsymbol{A_k}$ ***quadratische*** $\boldsymbol{(N\times N)}$**-Matrizen** sind und genauso viele Gleichungen wie unbekannte Funktionen vorliegen. Ist $A_m(t)$ im Fall $\mathcal{X} = \mathcal{Z}$ für alle $t \in I$ in $\mathcal{L}(\mathcal{X}, \mathcal{X})$ invertierbar, so kann man durch Multiplikation von (7.1) mit der Inversen $A_m(t)^{-1}$ auf den Fall $A_m(t) = \mathrm{id}_{\mathcal{X}}$ (mit der Identität $\mathrm{id}_{\mathcal{X}} \colon \mathcal{X} \to \mathcal{X}$) reduzieren. **Im Folgenden wird nur noch dieser Fall, der einem System in expliziter Form entspricht, betrachtet.** Im Matrizenfall reicht es für Invertierbarkeit natürlich, $\det(A_m(t)) \neq 0$ für alle $t \in I$ zu prüfen, und Stetigkeit von b und allen A_k bleibt bei der Herstellung der expliziten Form dann erhalten (denn gemäß der Cramerschen Regel hängt auch $A_m(t)^{-1}$ stetig von t ab).

Die wesentlichen Sachverhalte bei linearen GDG-Systemen der beschriebenen Form werden im folgenden Satz und seinen Korollaren zusammengestellt.

Hauptsatz 7.2 (Globaler Existenz- und Eindeutigkeitssatz für lineare GDGen). *Sei* $\mathcal{X}$ *vollständig, und sei* $P = \frac{\mathrm{d}^m}{\mathrm{d}t^m} + \sum_{k=0}^{m-1} A_k \frac{\mathrm{d}^k}{\mathrm{d}t^k}$ *mit* $A_0, A_1, A_2, \ldots, A_{m-1} \in \mathrm{C}^0(I, \mathcal{L}(\mathcal{X}, \mathcal{X}))$ *und* $b \in \mathrm{C}^0(I, \mathcal{X})$*. Dann besitzt das AWP*[1] $Pu = b$, $u^{[m-1]}(t_0) = (y_0, y_1, y_2, \ldots, y_{m-1})$ *für alle* $t_0 \in I$ *und* $y_0, y_1, y_2, \ldots, y_{m-1} \in \mathcal{X}$ *stets eine eindeutige Lösung auf ganz* I*.*

Beweis. Die Strukturfunktion der auf explizite Form gebrachten GDG

$$f(t, x_0, x_1, x_2, \ldots, x_{m-1}) = -A_{m-1}(t)x_{m-1} - \ldots - A_1(t)x_1 - A_0(t)x_0 + b(t)\,,$$

[1] Die Notation $u^{[m-1]} := (u, u', u'', \ldots, u^{(m-1)})$ wurde im Zusammenhang mit Korollar 6.2 bereits eingeführt.

erfüllt eine lineare Wachstumsbedingung und eine lokale pLB auf $I \times \mathcal{X}^m$; dies sieht man aus den Abschätzungen (wobei $\|\cdot\|$ für die Operatornorm auf $\mathcal{L}(\mathcal{X},\mathcal{X})$ steht)

$$\big|f(t,x_0,x_1,x_2,\dots,x_{m-1})\big| \le \max\big\{\|A_1(t)\|,\|A_2(t)\|,\dots,\|A_{m-1}(t)\|,|b(t)|\big\}\Big(1+\sum_{i=0}^{m-1}|x_i|\Big)$$

und

$$\begin{aligned}&\big|f(t,\widetilde{x}_0,\widetilde{x}_1,\widetilde{x}_2,\dots,\widetilde{x}_{m-1})-f(t,x_0,x_1,x_2,\dots,x_{m-1})\big|\\ &\qquad\le \max\Big\{\sup_K\|A_0\|,\sup_K\|A_1\|,\dots,\sup_K\|A_{m-1}\|\Big\}\sum_{i=0}^{m-1}|\widetilde{x}_i-x_i|\,,\end{aligned}$$

letztere gültig für alle t im Kompaktum $K \subset I$. In Anbetracht von Wachstumsbedingung und pLB folgt die Aussage von Hauptsatz 7.2 nun aus der globalen Existenzaussage des Satzes 6.4 (zuzüglich der darauf folgenden Bemerkungen (1) und (3)) und dem Eindeutigkeitsteil des Satzes von Picard-Lindelöf (in der m-ter-Ordnung-Version des Korollars 6.2). □

Aus Hauptsatz 7.2 ergeben sich folgende Korollare (gültig unter denselben Voraussetzungen):

Korollar 7.3. *Für jede* nicht-triviale[2] *Lösung u zu $Pu \equiv 0$ auf I besitzt $u^{[m-1]}$ keine Nullstelle in I, d.h. u, u', u'', …, $u^{(m-1)}$ besitzen keine* gemeinsame *Nullstelle in I.*

Beweis. Ist τ eine solche Nullstelle, so ist u wegen Eindeutigkeit der Lösung des AWPs $Pu \equiv 0$, $u^{[m-1]}(\tau) = 0$ notwendigerweise die triviale Lösung. □

Korollar 7.4. *Sind A_k und b reell, so gilt für (potentiell) komplexe Lösungen u zu $Pu = b$:*

$$u \text{ ist reelle Lösung auf } I \iff u \text{ hat an einer Stelle } \tau \in I \text{ reelle AWe } u^{[m-1]}(\tau)$$

(wobei reelle und komplexe Lösungen wie in Teil (IV) des vorigen Satzes beziehungsweise wie in der zugehörigen Bemerkung (2) zu verstehen sind).

Beweis. „$\Longrightarrow$“ gilt trivial, „$\Longleftarrow$“ folgt aus Existenz der reellen und Eindeutigkeit der komplexen Lösung. □

Korollar 7.5. *Für jedes $t_0 \in I$ ist die Abbildung von $(y_0,y_1,y_2,\dots,y_{m-1})$ auf die eindeutige Lösung des AWPs $Pu \equiv 0$, $u^{[m-1]}(t_0) = (y_0,y_1,y_2,\dots,y_{m-1})$ ein Isomorphismus[3] von $\mathcal{X}^m$ auf den Lösungsraum $\mathscr{L}_{\mathrm{hom}}$ der GDG $Pu \equiv 0$ auf I. Daher ist $\dim_{\mathbb{K}}(\mathscr{L}_{\mathrm{hom}}) = m\dim_{\mathbb{K}}(\mathcal{X})$, und* ***für $\mathcal{X} = \mathbb{K}^N$ gilt $\dim_{\mathbb{K}}(\mathscr{L}_{\mathrm{hom}}) = mN =$ (Ordnung) · (Anzahl Gleichungen)***.

Beweis. Die angegebene Abbildung ist wohldefiniert (gemäß Hauptsatz), $\mathbb{K}$-linear (gemäß Superpositionsprinzip) und bijektiv (trivial). Gemäß Basis-Resultaten der linearen Algebra ist sie damit Isomorphismus und erfüllt $\dim_{\mathbb{K}}$(Definitionsbereich) = $\dim_{\mathbb{K}}$(Zielbereich). □

[2] Nicht-triviale Lösungen nennt man alle Lösungen außer der als triviale Lösung bezeichneten Null-Funktion.

[3] Hier ist ein Isomorphismus von $\mathbb{K}$-Vektorräumen, d.h. eine $\mathbb{K}$-lineare Bijektion mit $\mathbb{K}$-linearer Umkehrabbildung, gemeint und nicht — wie oft in der Funktionalanalysis — ein Isomorphismus von normierten Räumen über $\mathbb{K}$ (wobei man für letzteren Begriff auch erst sagen müsste, welche Norm man auf dem Lösungsraum verwendet).

Korollar 7.6. *Seien $u_1, u_2, \ldots, u_L$ mit $L = m \dim_{\mathbb{K}}(\mathcal{X}) \in \mathbb{N} \cup \{\infty\}$ Lösungen zu $Pu \equiv 0$ auf I. Dann sind* ***äquivalent***:

- *$u_1, u_2, \ldots, u_L$ ist Basis des Lösungsraum zu $Pu \equiv 0$ auf I;*
- *$u_1^{[m-1]}(t), u_2^{[m-1]}(t), \ldots, u_L^{[m-1]}(t)$ ist für* **ein** *$t \in I$ Basis von $\mathcal{X}^m$;*
- *$u_1^{[m-1]}(t), u_2^{[m-1]}(t), \ldots, u_L^{[m-1]}(t)$ ist für* **alle** *$t \in I$ Basis von $\mathcal{X}^m$.*

Beweis. Die Behauptung folgt unmittelbar aus Korollar 7.5 (da Isomorphismen Basen in Basen überführen). □

Definition 7.7 (**Fundamentalsysteme, Fundamentalmatrizen**)**.**

- *Eine Basis des Lösungsraums eines* homogenen *linearen GDG-Systems $Pu \equiv 0$ auf I nennt man ein* ***Fundamentalsystem (FS)*** (*oder, seltener, ein Hauptsystem*) *zu $Pu \equiv 0$ auf I.*
- *Unter den Voraussetzungen des Hauptsatzes mit $\mathcal{X} = \mathbb{K}^N$ heißt jede aus mN beliebigen Lösungen $u_1, u_2, \ldots, u_{mN}$ zu $Pu \equiv 0$ gebildete, t-abhängige* quadratische *Matrix*

$$W(t) := \left(u_1^{[m-1]}(t) \middle| u_2^{[m-1]}(t) \middle| \ldots \middle| u_{mN}^{[m-1]}(t) \right) \in \left(\mathbb{K}^N\right)^{m \times mN} = \mathbb{K}^{mN \times mN}$$

eine ***Wronski-Matrix*** *zur Stelle $t \in I$, und $\det(W(t))$ heißt die zugehörige* ***Wronski-Determinante****. Ist $u_1, u_2, \ldots, u_{mN}$ sogar Fundamentalsystem, so heißt W* ***Fundamentalmatrix (FM)*** *des GDG-Systems $Pu \equiv 0$ auf I.*

Bemerkungen (zu **Wronski- und Fundamentalmatrizen**)**.** Hier seien die Voraussetzungen des Hauptsatzes mit $\mathcal{X} = \mathbb{K}^N$ erfüllt, und es sei $\boldsymbol{L = mN}$ die Anzahl der Funktionen eines FSs.

(1) Die **Berechnung eines Fundamentalsystems** $u_1, u_2, \ldots, u_L$ zu $Pu \equiv 0$ **ist im Wesentlichen dasselbe wie die Berechnung der allgemeinen Lösung** $C_1u_1 + C_2u_2 + \ldots + C_Lu_L$ mit Konstanten $C_1, C_2, \ldots, C_L \in \mathbb{K}$. Mit der aus $u_1, u_2, \ldots, u_L$ gebildeten Fundamentalmatrix W kann man die allgemeine Lösung (im Fall $m > 1$ samt Ableitungen bis Ordnung $m-1$) auch als Matrix-Vektor-Produkt[4]

$$u^{[m-1]}(t) = W(t)C \qquad \text{für } t \in I \tag{7.2}$$

von $W(t) \in \mathbb{K}^{L \times L}$ mit dem konstanten Vektor $C = (C_1, C_2, \ldots, C_L) \in \mathbb{K}^L$ schreiben.

(2) **Fundamentalmatrizen und -systeme sind** (solange man nicht zusätzlich eine AB fordert) **nicht völlig eindeutig**. Kennt man eine Fundamentalmatrix W beziehungsweise ein Fundamentalsystem $u_1, u_2, \ldots, u_L$ zu $Pu \equiv 0$, so erhält man alle weiteren in der Form WA beziehungsweise $\sum_{i=1}^{L} a_{i1}u_i, \sum_{i=1}^{L} a_{i2}u_i, \ldots, \sum_{i=1}^{L} a_{iL}u_i$ mit einer konstanten, invertierbaren Matrix $A = (a_{ij})_{i,j=1,2,\ldots,L} \in \mathbb{K}^{L \times L}$; dies folgt aus der Definition eines Fundamentalsystems als Basis des Lösungsraums und dem Superpositionsprinzip.

[4] Hier wird $u^{[m-1]}(t)$ als *Spalten*vektor mit m Einträgen aus $\mathbb{K}^N$, also mit insgesamt mN Zeilen verstanden. Auch $(C_1, C_2, \ldots, C_L)$ und jedes mit Kommata gekennzeichnete Tupel werden als *Spalten*vektoren interpretiert.

(3) Gemäß Korollar 7.6 und linearer Algebra gilt für eine Wronski-Matrix $W\colon I \to \mathbb{K}^{L\times L}$:

$$\begin{aligned} W \text{ Fundamentalmatrix} &\iff W(t) \text{ invertierbar für } \textit{ein } t \in I \\ &\iff W(t) \text{ invertierbar für } \textit{alle } t \in I \\ &\iff \det W(t) \neq 0 \text{ für } \textit{ein } t \in I \\ &\iff \det W(t) \neq 0 \text{ für } \textit{alle } t \in I\,. \end{aligned}$$

(4) Ist die allgemeine Lösung $C_1u_1 + C_2u_2 + \ldots + C_Lu_L$ zu $Pu \equiv 0$ bestimmt, so führt das **Einsetzen einer AB** $u^{[m-1]}(t_0) = (y_0, y_1y_2, \ldots, y_{m-1})$ mit $t_0 \in I$ und $y_0, y_1, y_2, \ldots, y_{m-1} \in \mathbb{K}^N$ auf das lineare Gleichungssystem mit m Gleichungen in $\mathbb{K}^N$ (also insgesamt L Komponentengleichungen)

$$\begin{array}{llll} C_1u_1(t_0) & + C_2u_2(t_0) & + \ldots + \mathrm{C}_Lu_L(t_0) & = y_0 \\ C_1u_1'(t_0) & + C_2u_2'(t_0) & + \ldots + \mathrm{C}_Lu_L'(t_0) & = y_1 \\ \vdots & \vdots & \vdots & \vdots \\ C_1u_1^{(m-1)}(t_0) & + C_2u_2^{(m-1)}(t_0) & + \ldots + \mathrm{C}_Lu_L^{(m-1)}(t_0) & = y_{m-1} \end{array}$$

für die L Koeffizienten C_1, C_2, $\ldots C_L$. Mit der aus u_1, u_2, $\ldots$, u_L gebildeten Fundamentalmatrix W und $C = (C_1, C_2, \ldots, C_L)$ kann dies analog zu (7.2) als

$$W(t_0)C = \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{m-1} \end{pmatrix} \qquad \text{oder auch als} \qquad C = W(t_0)^{-1} \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{m-1} \end{pmatrix}$$

zusammengefasst werden. Dabei wurde natürlich auf die Invertierbarkeit von $W(t_0)$ zurückgegriffen, und es folgt, dass das Gleichungssystem stets eindeutig lösbar ist (was für den Spezialfall $N = 1$ bereits in Abschnitt 5.2 behauptet wurde, aber dort noch nicht so leicht begründet werden konnte). Insgesamt erhält man für die eindeutige Lösung u des AWPs $Pu \equiv 0$ auf I, $u^{[m-1]}(t_0) = (y_0, y_1y_2, \ldots, y_{m-1})$ die Formel

$$u^{[m-1]}(t) = W(t)W(t_0)^{-1} \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{m-1} \end{pmatrix} \qquad \text{für } t \in I\,. \tag{7.3}$$

(5) Wronski-Matrizen W zu $Pu \equiv 0$ auf I sind charakterisiert durch die Matrix-DGL

$$W' = AW \qquad \text{auf } I$$

mit der t-abhängigen Matrix

$$A(t) := \begin{pmatrix} 0 & \mathbb{I}_N & 0 & \cdots & 0 & 0 \\ 0 & 0 & \mathbb{I}_N & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \mathbb{I}_N & 0 \\ 0 & 0 & 0 & \cdots & 0 & \mathbb{I}_N \\ -A_0(t) & -A_1(t) & -A_2(t) & \cdots & -A_{m-2}(t) & -A_{m-1}(t) \end{pmatrix} \in \mathbb{K}^{L\times L}\,,$$

die neben $(N\times N)$-Einheitsblöcken $\mathbb{I}_N$ die Koeffizienten A_k von P enthält. Diese Charakterisierung folgt direkt aus der Beobachtung, dass die Spalten von W das aus $Pu\equiv 0$ durch Reduktion auf Ordnung 1 entstehende System $U'=AU$ auf I (für $\mathbb{K}^L$-wertiges U) lösen.

Auch die Wronski-Determinante genügt einer DGL, nämlich der skalaren Gleichung

$$(\det W)' = -(\operatorname{Spur} A_{m-1})\det W \qquad \text{auf } I\,,$$

und kann daher zu $\det W(t) = C\exp\big(-\int(\operatorname{Spur} A_{m-1}(t))\,\mathrm{d}t\big)$ mit $C\in\mathbb{K}$ bestimmt werden. Um die Gültigkeit der DGL einzusehen sind die Regel $(\partial_\Gamma\det)(\mathbb{I}_L)=\operatorname{Spur}\Gamma$, $\Gamma\in\mathbb{K}^{L\times L}$ für Richtungsableitungen der Determinante in der Einheitsmatrix und die daraus folgende Beobachtung $\frac{\mathrm{d}}{\mathrm{d}s}\big|_{s=0}\det(\Phi(s)) = (\partial_{\Phi'(0)}\det)(\Phi(0)) = \operatorname{Spur}(\Phi'(0))$ für $\mathbb{K}^{L\times L}$-wertiges Φ mit $\Phi(0)=\mathbb{I}_L$ nützlich. Zusammen mit weiteren Rechenregeln für Determinanten, der DGL $W'=AW$ und schließlich der Invarianz der Spur unter Ähnlichkeitstransformation verifiziert man dann

$$\begin{aligned}(\det W)'(t) &= \frac{\mathrm{d}}{ds}\Big|_{s=0}\det W(t+s) = \frac{\mathrm{d}}{ds}\Big|_{s=0}\det(W(t)^{-1}W(t+s))\det W(t)\\ &= \operatorname{Spur}\Big(\frac{\mathrm{d}}{ds}\Big|_{s=0}W(t)^{-1}W(t+s)\Big)\det W(t) = \operatorname{Spur}\big(W(t)^{-1}W'(t)\big)\det W(t)\\ &= \operatorname{Spur}\big(W(t)^{-1}A(t)W(t)\big)\det W(t) = (\operatorname{Spur}A(t))\det W(t) = -(\operatorname{Spur}A_{m-1}(t))\det W(t)\,,\end{aligned}$$

zunächst nur wo $W(t)$ invertierbar. Im Nachhinein gilt die DGL aber aus Stetigkeitsgründen allgemein auf ganz I.

(6) Mit Hilfe einer Fundamentalmatrix W des homogenen Systems $Pu\equiv 0$ auf I lassen sich **Lösungsformeln für das inhomogene System** $Pu=b$ angeben, die einerseits die Formeln aus Satz 5.1 für den skalaren Erster-Ordnung-Fall, andererseits auch (7.2) und (7.3) verallgemeinern: Die allgemeine Lösung zu $Pu=b$ auf I ist gegeben durch

$$u^{[m-1]}(t) = W(t)\left[\int W(t)^{-1}\begin{pmatrix}0\\ b(t)\end{pmatrix}\mathrm{d}t + C\right] \qquad \text{für } t\in I \tag{7.4}$$

mit Konstante $C\in\mathbb{K}^L$ und mit $L-N=(m-1)N$ Null-Einträgen im Vektor $\left(\begin{smallmatrix}0\\ b(t)\end{smallmatrix}\right)\in\mathbb{K}^L$, und die Lösung des AWPs $Pu=b$ auf I, $u^{[m-1]}(t_0)=(y_0,y_1,y_2,\dots,y_{m-1})$ mit $t_0\in I$, $y_0,y_1,y_2,\dots,y_{m-1}\in\mathbb{K}^N$ ist gegeben durch

$$u^{[m-1]}(t) = W(t)\left[\int_{t_0}^{t} W(s)^{-1}\begin{pmatrix}0\\ b(s)\end{pmatrix}\mathrm{d}s + W(t_0)^{-1}\begin{pmatrix}y_0\\ y_1\\ \vdots\\ y_{m-1}\end{pmatrix}\right] \qquad \text{für } t\in I\,. \tag{7.5}$$

Zum Nachweis dieser Formeln bleibt wegen Teil (III) von Satz 7.1 und der vorausgehenden Diskussion homogener Systeme nur nachzurechnen, dass $U_{\mathrm{sp}}(t) = W(t)\int W(t)^{-1}\left(\begin{smallmatrix}0\\ b(t)\end{smallmatrix}\right)\mathrm{d}t$ eine spezielle Lösung des auf Ordnung 1 reduzierten inhomogenen Systems $U'=AU+\left(\begin{smallmatrix}0\\ b\end{smallmatrix}\right)$ gibt. Letzteres gelingt problemlos mit der Produktregel, dem HDI und der DGL aus (5).

Für konkrete Lösungsberechnungen bei Systemen sind die Formeln (7.2), (7.3), (7.4), (7.5) aber weniger nützlich, als es zunächst scheinen mag; denn zum einen muss das homogene System bereits gelöst haben, um eine Fundamentalmatrix W zu kennen, und zum anderen verursacht das Invertieren von W oft einen nicht unerheblichen Aufwand. Ähnliches gilt auch für das als nächstes beschriebene Rechenverfahren, bei dem das Lösen eines linearen Gleichungssystems weitgehend denselben Aufwand wie das Invertieren der Matrix mit sich bringt.

(7) Ist u_1, u_2, …, u_L ein Fundamentalsystem des homogenen Systems $Pu \equiv 0$ auf I, so sind Lösungen des inhomogenen Systems $Pu = b$ auch charakterisiert durch die Gleichung

$$u(t) = K_1(t)u_1(t) + K_2(t)u_2(t) + \ldots + K_L(t)u_L(t) \qquad \text{für } t \in I \tag{7.6}$$

mit Hilfsfunktionen $K_1, K_2, \ldots, K_L \in \mathrm{C}^1(I, \mathbb{K})$, deren Ableitungen dem Gleichungssystem

$$\begin{aligned} K_1'u_1^{(j)} \quad &+ K_2'u_2^{(j)} \quad + \ldots + K_L'u_L^{(j)} \quad \equiv 0 \qquad \text{für } j = 0, 1, 2, \ldots, m-2\,, \\ K_1'u_1^{(m-1)} &+ K_2'u_2^{(m-1)} + \ldots + K_L'u_L^{(m-1)} = b \end{aligned}$$

genügen. Zu diesen Formeln gehört ein naheliegendes, als **Variation der Konstanten** bekanntes **Rechenverfahren zur Lösung des inhomogenen Systems** bei bereits gelöstem homogenen System (benannt danach, dass die Funktionen K_1, K_2, …, K_L den Platz einnehmen, an dem bei der allgemeinen Lösung des *homogenen* Systems wirklich nur Konstanten stehen): Man bestimmt zuerst die Ableitungen K_1', K_2', …, K_L' durch Lösung des linearen Gleichungssystems und danach die Stammfunktionen K_1, K_2, …, K_L durch Integrationen. Die Formel (7.6) ergibt dann die Lösung des inhomogenen Systems (je nach Behandlung der Integrationskonstanten als spezielle oder allgemeine Lösung). Von tatsächlicher praktischer Bedeutung ist dieses Verfahren aber höchstens im Erster-Ordnung-Fall $m = 1$, und in diesem muss man sich das Gleichungssystem für die Ableitungen K_i' auch nicht merken, sondern erhält es automatisch durch den Ansatz (7.6) für Lösungen des inhomogenen Systems $Pu = b$.

Um die bisher nur behauptete Charakterisierung zu verifizieren, fixiert man die aus u_1, u_2, …, u_L gebildete Fundamentalmatrix W, wählt die $K_i(t)$ als Komponenten des Ausdrucks in der eckigen Klammer von (7.4) und erhält sofort (7.6). Die Wahl der $K_i(t)$ bedeutet aber auch, dass die K_i' die Komponenten von $W^{-1}\left(\begin{smallmatrix} 0 \\ b \end{smallmatrix}\right)$ sind, und Links-Multiplikation mit W führt daher auf das angegebene Gleichungssystem.

(8) Alles Obige kann auf möglicherweise ∞-dimensionale Banach-Räume $\mathcal{X}$ verallgemeinert werden, wenn man in geeigneter Weise mit $\mathcal{L}(\mathcal{X}, \mathcal{X})$-wertigen Funktionen W arbeitet.

7.2 Matrix-Exponentialansatz bei linearen Systemen mit konstanten Koeffizienten

Um die Theorie des vorausgehenden Abschnitts bei der Berechnung von Lösungen konkreter linearer GDGen anwenden zu können, muss man zunächst ein Fundamentalsystem beziehungsweise eine Fundamentalmatrix (der zugehörigen homogenen Gleichung) bestimmen. Bei *skalaren* linearen GDGen mit konstanten Koeffizienten lassen sich hierzu die Methoden des Abschnitts 5.2 verwenden, im Folgenden sollen aber auch homogene lineare Erster-Ordnung-*Systeme* in expliziter Form

$$u' = Au \qquad \text{auf } I \tag{7.7}$$

mit konstanter $(N \times N)$-Koeffizientenmatrix $A \in \mathbb{K}^{N \times N}$ (also N Komponentengleichungen für N Komponentenfunktionen von u) behandelt werden. Homogene lineare Höherer-Ordnung-Systeme mit konstanten Koeffizienten lassen sich dann durch Reduktion auf Ordnung 1 darauf zurückführen und werden deshalb nicht mehr eigens betrachtet.

Wie in Abschnitt 5.2 bietet sich auch bei (7.7) ein Ansatz mit Exponentialfunktionen an, und tatsächlich liefert der **Exponentialansatz**

$$u(t) = \mathrm{e}^{\lambda t} v$$

mit $\lambda \in \mathbb{K}$ und $v \in \mathbb{K}^N$ **genau dann eine nicht-triviale Lösung** von (7.7) auf I, **wenn $\boldsymbol{v}$ ein Eigenvektor der Matrix $\boldsymbol{A}$ zum Eigenwert $\boldsymbol{\lambda}$ ist**. Aus solchen Eigenwert-Eigenvektor-Lösungen lässt sich zwar nicht allgemein ein Fundamentalsystem zusammensetzen (denn über $\mathbb{K} = \mathbb{R}$ muss es nicht einmal einen Eigenwert geben und über $\mathbb{K} = \mathbb{C}$ gibt es ‚genug' Eigenwerte, aber die Summe der Dimensionen der Eigenräume kann kleiner sein als die Dimension N des Lösungsraums), aber man kann die Idee dieses Ansatzes ähnlich wie in Abschnitt 5.2 ausbauen und dann ganz allgemein ein Fundamentalsystem erhalten. Statt hierauf genauer einzugehen, soll jetzt aber ein eleganterer Zugang betrachtet werden, der (zumindest vordergründig) direkt mit der Matrix A statt mit ihren Eigenwerten und -vektoren arbeitet. Die Grundidee hierbei ist, dass als Lösung der GDG $W' = AW$ für die Wronski-Matrix W heuristisch der kurze Ausdruck $W(t) = \mathrm{e}^{tA}$ in Frage kommt. Um dies zu formalisieren, muss man ‚e hoch Matrix' aber erst einmal sinnvoll erklären, was in Analogie zur bekannten Exponentialreihe $\mathrm{e}^z = \sum_{k=0}^{\infty} \frac{1}{k!} z^k$ für $z \in \mathbb{C}$ folgendermaßen geschehen kann:

Definition 7.8 (**Exponentialabbildung von Matrizen**). *Für $A \in \mathbb{K}^{N\times N}$ sei*

$$\exp(A) := \mathrm{e}^A := \sum_{k=0}^{\infty} \frac{1}{k!} A^k \in \mathbb{K}^{N\times N}, \tag{7.8}$$

wobei $A^k = A \cdot A \cdot \ldots \cdot A$ (mit k Faktoren A auf der rechten Seite) das k-fache Matrix-Produkt von A mit sich selbst bezeichnet und A^0 per Konvention für die $(N\times N)$-Einheitsmatrix $\mathbb{I}_N$ steht. Die Konvergenz der Reihe in (7.8) bezüglich der Operatornorm[5] $\|\cdot\|$ auf $\mathbb{K}^{N\times N}$ ist dabei sichergestellt, weil die Abschätzung $\|A^k\| \le \|A\|^k$ auf die konvergente Majoranten-Reihe $\sum_{k=0}^{\infty} \frac{1}{k!} \|A\|^k$ führt. Tatsächlich liegt aber auch eintragsweise sowie bezüglich jeder anderen Norm auf $\mathbb{K}^{N\times N}$ Konvergenz vor (denn alle Normen auf einem endlich-dimensionalen Raum wie $\mathbb{K}^{N\times N}$ sind bekanntlich äquivalent).

Bemerkungen (zu **Grundeigenschaften der Matrix-Exponentialabbildung**).

(1) **Für eine** vom Grad $\ell \in \mathbb{N}$ **nilpotente Matrix** $A \in \mathbb{K}^{N\times N}$ (d.h. im Fall $A^\ell = 0 \neq A^{\ell-1}$) **vereinfacht sich die Exponentialreihe zur endlichen Summe**

$$\mathrm{e}^A = \sum_{k=0}^{\ell-1} \frac{1}{k!} A^k$$

und kann (für nicht zu großes ℓ und N) problemlos berechnet werden.

(2) Für **Diagonalmatrizen** mit Einträgen $\lambda_1, \lambda_2, \ldots, \lambda_{N-1}, \lambda_N \in \mathbb{K}$ gilt die Regel

$$\exp\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_{N-1} & 0 \\ 0 & 0 & \cdots & 0 & \lambda_N \end{pmatrix} = \begin{pmatrix} \mathrm{e}^{\lambda_1} & 0 & \cdots & 0 & 0 \\ 0 & \mathrm{e}^{\lambda_2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathrm{e}^{\lambda_{N-1}} & 0 \\ 0 & 0 & \cdots & 0 & \mathrm{e}^{\lambda_N} \end{pmatrix},$$

insbesondere ist $\mathrm{e}^{\lambda \mathbb{I}_N} = \mathrm{e}^{\lambda} \mathbb{I}_N$ für $\lambda \in \mathbb{K}$ und $\mathrm{e}^{0_N} = \mathbb{I}_N$ für die $(N\times N)$-Nullmatrix 0_N.

[5] Es sei daran erinnert, dass die Operatornorm $\|\Gamma\|$ einer Matrix $\Gamma \in \mathbb{K}^{N\times N}$ die kleinste Schranke $M \in \mathbb{R}_0^+$ mit $|\Gamma v| \le M|v|$ für alle Vektoren $v \in \mathbb{K}^N$ ist.

(3) Allgemeiner gilt für **Dreiecks(block)matrizen**

$$\exp\begin{pmatrix} A_1 & * & \cdots & * & * \\ 0 & A_2 & \cdots & * & * \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & A_{\ell-1} & * \\ 0 & 0 & \cdots & 0 & A_\ell \end{pmatrix} = \begin{pmatrix} \mathrm{e}^{A_1} & \tilde{*} & \cdots & \tilde{*} & \tilde{*} \\ 0 & \mathrm{e}^{A_2} & \cdots & \tilde{*} & \tilde{*} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathrm{e}^{A_{\ell-1}} & \tilde{*} \\ 0 & 0 & \cdots & 0 & \mathrm{e}^{A_\ell} \end{pmatrix},$$

wobei A_1, A_2, …, $A_{\ell-1}$, A_ℓ sowohl einzelne Zahlen $\in \mathbb{K}$ als auch quadratische Blöcke (eventuell unterschiedlicher Größe) sein können und die Sterne für weitere, nicht spezifizierte Zahlen oder Blöcke stehen.

(4) Die Exponentialabbildung ist **mit komplexer Konjugation, Transposition und Ähnlichkeitstransformation** (letztere entspricht einem Basiswechsel und wird auch innere Konjugation genannt) **vertauschbar**, d.h. für $A \in \mathbb{K}^{N\times N}$ gelten

$$\overline{\mathrm{e}^A} = \mathrm{e}^{\overline{A}}, \qquad \left(\mathrm{e}^A\right)^{\mathrm{t}} = \mathrm{e}^{(A^{\mathrm{t}})} \qquad \text{und} \qquad \mathrm{e}^{TAT^{-1}} = T\mathrm{e}^A T^{-1} \text{ für invertierbares } T \in \mathbb{K}^{N\times N}.$$

Als gemeinsame Erklärung für diese Regeln kann man die allgemeinere Vertauschbarkeitsregel $\Phi(\mathrm{e}^A) = \mathrm{e}^{\Phi(A)}$ für Algebren(anti)endomorphismen $\Phi\colon \mathbb{K}^{N\times N} \to \mathbb{K}^{N\times N}$ heranziehen.

(5) Ist $D \in \mathbb{K}^{N\times N}$ **diagonalisierbar** mit Eigenwerten $\lambda_1, \lambda_2, \ldots, \lambda_{N-1}, \lambda_N \in \mathbb{K}$, also

$$D = T\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_{N-1} & 0 \\ 0 & 0 & \cdots & 0 & \lambda_N \end{pmatrix} T^{-1}$$

für invertierbares $T \in \mathbb{K}^{N\times N}$, so folgt aus (2) und (4) die Regel

$$\mathrm{e}^D = T\begin{pmatrix} \mathrm{e}^{\lambda_1} & 0 & \cdots & 0 & 0 \\ 0 & \mathrm{e}^{\lambda_2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathrm{e}^{\lambda_{N-1}} & 0 \\ 0 & 0 & \cdots & 0 & \mathrm{e}^{\lambda_N} \end{pmatrix} T^{-1}.$$

Symmetrische Matrizen über $\mathbb{R}$ und normale Matrizen über $\mathbb{C}$ sind stets diagonalisierbar und können daher mit dieser Regel behandelt werden (wofür es aber zunächst gilt, die Diagonaltransformation, also die Eigenwerte und -vektoren auszurechnen).

(6) Für $A \in \mathbb{K}^{N\times N}$ und $t \in \mathbb{R}$ gilt die **Ableitungsregel**

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{tA} = \mathrm{e}^{tA}A = A\mathrm{e}^{tA}.$$

(7) **Für kommutierende Matrizen** $A, B \in \mathbb{K}^{N\times N}$ (also solche mit $AB = BA$) **gilt das Exponentialgesetz**

$$\mathrm{e}^A\mathrm{e}^B = \mathrm{e}^{A+B} = \mathrm{e}^B\mathrm{e}^A.$$

Insbesondere ist e^A für jedes $A \in \mathbb{K}^{N\times N}$ invertierbar mit inverser Matrix

$$\left(\mathrm{e}^A\right)^{-1} = \mathrm{e}^{-A}.$$

Beweise. Die Regeln (1)–(4) sind direkte Folgerungen aus Definition 7.8, Regel (5) wurde bereits erläutert, und die Beweise der Regeln (6) und (7) sind Thema der Übungen. □

Der für die Anwendung auf das GDG-System (7.7) wichtigste Sachverhalt zur Matrix-Exponentialabbildung folgt:

Satz 7.9 (**Lösung von $u' = Au$ mit konstantem A durch Matrix-Exponentialansatz**). *Für jedes $A \in \mathbb{K}^{N\times N}$ ist durch*

$$W(t) = \mathrm{e}^{tA}$$

die Fundamentalmatrix W des homogenen linearen Erster-Ordnung-GDG-Systems

$$u' = Au \qquad \text{auf } \mathbb{R} \tag{7.9}$$

mit $W(0) = \mathbb{I}_N$ gegeben, und für die eindeutige Lösung u des AWPs zu (7.9) mit AB $u(t_0) = y_0$, $t_0 \in \mathbb{R}$, $y_0 \in \mathbb{K}^N$ erhält man die Formel

$$u(t) = \mathrm{e}^{(t-t_0)A} y_0 \qquad \text{für } t \in \mathbb{R}\,.$$

Beweis. Gemäß der vorausgehenden Bemerkung (6) gilt $W'(t) = A\mathrm{e}^{tA} = AW(t)$, also erfüllt W die DGL einer Wronski-Matrix zu (7.9). Nach Bemerkung (7) ist $W(t)$ außerdem für alle $t \in \mathbb{R}$ invertierbar, also ist W gemäß Abschnitt 7.1 Fundamentalmatrix zu (7.9). Insbesondere ist damit $u(t) = W(t)\mathrm{e}^{-t_0 A} y_0 = \mathrm{e}^{(t-t_0)A} y_0$ Lösung von (7.9), die die Anfangsbedingung erfüllt und deren Eindeutigkeit bereits aus Hauptsatz 7.2 bekannt ist. □

Von der theoretischen Seite sind GDG-Systeme des Typs (7.9) mit Satz 7.9 vollständig abgehandelt, zu konkreten Rechenverfahren gibt es aber noch eine ganze Menge zu sagen:

Weitere Bemerkungen (zur **Berechnung von e^{tA} und anderen FMen über $\mathbb{C}$**). Die praktische Bestimmung von Lösungen beziehungsweise Fundamentalmatrizen zu (7.9) erfordert (in den allermeisten Fällen) etwas lineare Algebra, nämlich zunächst die Berechnung der verschiedenen Eigenwerte $\lambda_1, \lambda_2, \ldots, \lambda_\ell \in \mathbb{C}$ von A und ihrer Vielfachheiten. Je nach Ergebnis bieten sich dann unterschiedliche Möglichkeiten für das weitere Vorgehen an:

(8) Ist für $A \in \mathbb{C}^{N\times N}$ und jeden Eigenwert $\lambda_i \in \mathbb{C}$ von A die Dimension des zugehörigen Eigenraums $\mathrm{E}_{\lambda_i}(A)$ gleich der algebraischen Vielfachheit d_i von λ_i als Nullstelle des charakteristischen Polynoms von A, so ist man **im diagonalisierbaren Fall** und die Berechnung von e^{tA} gelingt im Prinzip mit obiger Bemerkung (5). Genauer kann man eine Basis von $\mathbb{C}^N$ aus Eigenvektoren $v_1, v_2, \ldots, v_{d_1} \in \mathrm{E}_{\lambda_1}(A)$, $v_{d_1+1}, v_{d_1+2}, \ldots, v_{d_1+d_2} \in \mathrm{E}_{\lambda_2}(A)$, …, $v_{N-d_\ell+1}, v_{N-d_\ell+2}, \ldots, v_N \in \mathrm{E}_{\lambda_\ell}(A)$ finden, und es gilt

$$\mathrm{e}^{tA} = T \begin{pmatrix} \mathrm{e}^{t\lambda_1}\mathbb{I}_{d_1} & 0 & \cdots & 0 & 0 \\ 0 & \mathrm{e}^{t\lambda_2}\mathbb{I}_{d_2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathrm{e}^{t\lambda_{\ell-1}}\mathbb{I}_{d_{\ell-1}} & 0 \\ 0 & 0 & \cdots & 0 & \mathrm{e}^{t\lambda_\ell}\mathbb{I}_{d_\ell} \end{pmatrix} T^{-1} \qquad \text{mit } T = \left(v_1 \middle| v_2 \middle| \ldots \middle| v_N \right).$$

Möchte man nicht unbedingt e^{tA}, sondern nur ein beliebiges Fundamentalsystem oder eine beliebige Fundamentalmatrix berechnen, so kann man allerdings den Rechenaufwand zur Berechnung der Inversen T^{-1} vermeiden, denn im diagonalisierbaren Fall bilden die zu Beginn des Abschnitts erwähnten Eigenwert-Eigenvektor-Lösungen $\mathrm{e}^{\lambda_1 t} v_1, \mathrm{e}^{\lambda_1 t} v_2, \ldots, \mathrm{e}^{\lambda_1 t} v_{d_1}$,

$\mathrm{e}^{\lambda_2 t}v_{d_1+1}, \mathrm{e}^{\lambda_2 t}v_{d_1+2}, \ldots, \mathrm{e}^{\lambda_2 t}v_{d_1+d_2}, \ldots, \mathrm{e}^{\lambda_\ell t}v_{N-d_\ell+1}, \mathrm{e}^{\lambda_\ell t}v_{N-d_\ell+2}, \ldots, \mathrm{e}^{\lambda_\ell t}v_N$ bereits ein Fundamentalsystem mit zugehöriger Fundamentalmatrix

$$e^{tA}T = \left(\mathrm{e}^{\lambda_1 t}v_1 \middle| \mathrm{e}^{\lambda_1 t}v_2 \middle| \ldots \middle| \mathrm{e}^{\lambda_\ell t}v_N\right).$$

(9) Ein **allgemeiner Zugang** zur Berechnung von e^{tA} für beliebiges, möglicherweise nicht diagonalisierbares $A \in \mathbb{C}^{N\times N}$ beruht auf der sogenannten **Diagonalisierbar-plus-Nilpotent-Zerlegung**: Man kann stets $A = D+R$ mit *kommutierenden* (!) Matrizen $D, R \in \mathbb{C}^{N\times N}$ schreiben, so dass D diagonalisierbar und R nilpotent ist. Nach Bemerkung (7) erhält man dann $\mathrm{e}^{tA} = \mathrm{e}^{tR}\mathrm{e}^{tD}$, wobei e^{tR} und e^{tD} gemäß (1) und (8) berechnet werden können. Wie in (8) erfordert die Berechnung von e^{tA} selbst das Aufstellen und Invertieren einer (jetzt zu D gehörigen) Transformationsmatrix T, möchte man allerdings nur eine beliebige Fundamentalmatrix ermitteln, so kann man sich das Invertieren von T wieder ersparen, indem man $\mathrm{e}^{tA}T$ statt e^{tA} berechnet.

In der Rechenpraxis ist die Verwendung der Diagonalisierbar-plus-Nilpotent-Zerlegung **sehr empfehlenswert**, wenn man die Summanden D und R leicht sehen oder raten kann. Insbesondere ist dies immer dann der Fall, **wenn man feststellt, dass es nur einen einzigen Eigenwert λ_1 gibt**, der notwendigerweise algebraische Vielfachheit N hat; dann funktioniert die Zerlegung nämlich mit den (offensichtlich kommutierenden) Matrizen $D = \lambda_1\mathbb{I}_N$, $\mathrm{e}^{tD} = \mathrm{e}^{t\lambda_1}\mathbb{I}_N$ und $R = A-\lambda_1\mathbb{I}_N$.

Allgemein gelingen ein Beweis für die Möglichkeit der Zerlegung und auch ihre Berechnung mit denselben Methoden, die auch bei der Herstellung der eng verwandten Jordanschen Normalform eine Rolle spielen: Die diagonalisierbare Matrix D ist bestimmt durch die Festlegung $Dv = \lambda_i v$ für jeden Vektor v im Hauptraum $\mathrm{H}_{\lambda_i}(A)$ zu einem Eigenwert λ_i von A, und die nilpotente Matrix R ergibt sich dann einfach als $R = A-D$.

(10) Ein **alternativer allgemeiner Zugang** verwendet direkt die (eher zum Standard-Kanon der linearen Algebra gehörende) **Jordansche Normalform**. Man benutzt hierbei für einen Jordan-Block

$$\mathrm{J}_{\alpha,p} := \begin{pmatrix} \alpha & 1 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \alpha & 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \alpha & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \alpha & 1 & 0 & 0 \\ 0 & \cdots & \cdots & 0 & 0 & \alpha & 1 & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & \alpha & 1 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 & \alpha \end{pmatrix} \in \mathbb{C}^{p\times p}$$

der Größe $p \in \mathbb{N}$ mit Eigenwert $\alpha \in \mathbb{C}$ die Regel

$$\mathrm{e}^{t\mathrm{J}_{\alpha,p}} = \mathrm{e}^{\alpha t}\begin{pmatrix} 1 & t & \frac{1}{2}t^2 & \frac{1}{3!}t^3 & \cdots & \cdots & \cdots & \frac{1}{(p-1)!}t^{p-1} \\ 0 & 1 & t & \frac{1}{2}t^2 & \frac{1}{3!}t^3 & \cdots & \cdots & \frac{1}{(p-2)!}t^{p-2} \\ 0 & 0 & 1 & t & \frac{1}{2}t^2 & \frac{1}{3!}t^3 & \cdots & \frac{1}{(p-3)!}t^{p-3} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 1 & t & \frac{1}{2}t^2 & \frac{1}{3!}t^3 \\ 0 & \cdots & \cdots & 0 & 0 & 1 & t & \frac{1}{2}t^2 \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 1 & t \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 0 & 1 \end{pmatrix} \in \mathbb{C}^{p\times p}. \tag{7.10}$$

Begründung für (7.10). Mit Induktion verifiziert man, dass $(\mathrm{J}_{\alpha,p})^k \in \mathbb{C}^{p\times p}$ die obere Dreiecksmatrix ist, auf deren i-ter oberer Nebendiagonale lauter gleiche Einträge $\binom{k}{i}\alpha^{k-i}$ mit

Binomialkoeffizienten $\binom{k}{i}$ (per Konvention $=0$ für $i>k$) stehen. Gemäß Exponentialreihe ist dann $\mathrm{e}^{t\mathrm{J}_{\alpha,p}}$ die obere Dreiecksmatrix, auf deren i-ter oberer Nebendiagonale sich lauter gleiche Einträge $\sum_{k=i}^{\infty}\frac{1}{k!}t^k\binom{k}{i}\alpha^{k-i}=\frac{1}{i!}t^i\sum_{k=i}^{\infty}\frac{1}{(k-i)!}(\alpha t)^{k-i}=\frac{1}{i!}t^i\mathrm{e}^{\alpha t}$ befinden. □

Für beliebiges $A\in\mathbb{C}^{N\times N}$ lässt sich nun wie in der linearen Algebra eine Transformationsmatrix $T\in\mathbb{C}^{N\times N}$ bestimmen, deren Spalten geeignete Basen aller Haupträume von A enthalten und durch die A auf die als Jordansche Normalform bekannte Block-Diagonalform mit Jordan-Blöcken auf der Diagonalen gebracht werden kann. Konkret bedeutet dies

$$A=T\begin{pmatrix}\mathrm{J}_{\alpha_1,p_1} & 0 & \cdots & 0 & 0\\ 0 & \mathrm{J}_{\alpha_2,p_2} & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \cdots & \mathrm{J}_{\alpha_{h-1},p_{h-1}} & 0\\ 0 & 0 & \cdots & 0 & \mathrm{J}_{\alpha_h,p_h}\end{pmatrix}T^{-1}$$

mit $h\in\mathbb{N}$, den (jetzt nicht unbedingt verschiedenen) Eigenwerten $\alpha_1,\alpha_2,\ldots,\alpha_{h-1},\alpha_h\in\mathbb{C}$ von A und Blockgrößen $p_1,p_2,\ldots,p_{h-1},p_h\in\mathbb{N}$ mit $\sum_{i=1}^{h}p_i=N$. Mit (5) folgt

$$\mathrm{e}^{tA}=T\begin{pmatrix}\mathrm{e}^{t\mathrm{J}\alpha_1,p_1} & 0 & \cdots & 0 & 0\\ 0 & \mathrm{e}^{t\mathrm{J}\alpha_2,p_2} & \cdots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \cdots & \mathrm{e}^{t\mathrm{J}\alpha_{h-1},p_{h-1}} & 0\\ 0 & 0 & \cdots & 0 & \mathrm{e}^{t\mathrm{J}\alpha_h,p_h}\end{pmatrix}T^{-1}\,, \tag{7.11}$$

wobei die Blöcke auf der rechten Seite durch (7.10) gegeben sind. Auch hier ist die Fundamentalmatrix $\mathrm{e}^{tA}T$ im Allgemeinen leichter zu berechnen als e^{tA} selbst.

Beispiel (für eine **Anwendung der Diagonalisierbar-plus-Nilpotent-Zerlegung**)**.** Beim GDG-System

$$u'=Au\qquad\text{mit } A=\begin{pmatrix}2&1&-3\\0&2&0\\0&0&2\end{pmatrix}$$

ist 2 der einzige Eigenwert von A. Gemäß Bemerkung (9) arbeitet man daher mit der Zerlegung der Koeffizientenmatrix $A=2\mathbb{I}_3+R$ in die Diagonalmatrix $2\mathbb{I}_3$ und die nilpotente Matrix

$$R=\begin{pmatrix}0&1&-3\\0&0&0\\0&0&0\end{pmatrix}.$$

Bei der Berechnung der Potenzen von R stellt man fest, dass schon R^2 verschwindet, also R vom Grad 2 nilpotent ist. Somit erhält man die Fundamentalmatrix

$$\mathrm{e}^{tA}=\mathrm{e}^{2t\mathbb{I}_3}\mathrm{e}^{tR}=\mathrm{e}^{2t}\mathbb{I}_3\big(\mathbb{I}_3+tR\big)=\mathrm{e}^{2t}\begin{pmatrix}1&t&-3t\\0&1&0\\0&0&1\end{pmatrix},$$

und die Funktionen

$$\begin{pmatrix}\mathrm{e}^{2t}\\0\\0\end{pmatrix},\qquad\begin{pmatrix}t\mathrm{e}^{2t}\\\mathrm{e}^{2t}\\0\end{pmatrix},\qquad\begin{pmatrix}-3t\mathrm{e}^{2t}\\0\\\mathrm{e}^{2t}\end{pmatrix}$$

bilden ein Fundamentalsystem.

Beispiel (für eine **Berechnung von $\mathbf{e}^{tA}$ mittels der Jordanschen Normalform**)**.** Beim GDG-System

$$u' = Au \qquad \text{mit } A = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & -25 & 7 \\ 0 & 0 & -6 & 2 \\ 0 & 0 & -28 & 9 \end{pmatrix}$$

bestimmt man das charakteristische Polynom der Block-Dreiecksmatrix A zu

$$(\lambda{-}2)^2[(\lambda{+}6)(\lambda{-}9) + 28 \cdot 2] = (\lambda{-}2)^2[\lambda^2{-}3\lambda{+}2] = (\lambda{-}2)^3(\lambda{-}1)\,,$$

somit treten der dreifache Eigenwert 2 und der einfache Eigenwert 1 auf. Als Eigenraum und Hauptraum zum Eigenwert 2 (über $\mathbb{C}$) berechnet man

$$\mathrm{E}_2(A) = \mathrm{Kern}(2\mathbb{I}_4{-}A) = \mathrm{Kern} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 25 & -7 \\ 0 & 0 & 8 & -2 \\ 0 & 0 & 28 & -7 \end{pmatrix} = \mathbb{C} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \mathbb{C} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix},$$

$$\mathrm{H}_2(A) = \mathrm{Kern}(2\mathbb{I}_4{-}A)^2 = \mathrm{Kern} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & -1 \\ 0 & 0 & 8 & -2 \\ 0 & 0 & 28 & -7 \end{pmatrix} = \mathbb{C} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \mathbb{C} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \mathbb{C} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 4 \end{pmatrix}.$$

Zur Herstellung der Jordanschen Normalform verwendet man nun eine Basis von $\mathrm{H}_2(A)$, deren erster Vektor in $\mathrm{E}_2(A)$ liegt und deren zweiter Vektor in $\mathrm{E}_2(A)$ das $(A{-}2\mathbb{I}_4)$-Bild des dritten Vektors aus $\mathrm{H}_2(A) \setminus \mathrm{E}_2(A)$ ist. Wegen $(A{-}2\mathbb{I}_4)\begin{pmatrix} 0 \\ 0 \\ 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \\ 0 \\ 0 \end{pmatrix}$ ist eine solche Basis durch die Vektoren $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 3 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 0 \\ 1 \\ 4 \end{pmatrix}$ gegeben. Einen vierten Basisvektor erhält man durch Berechnung des 1-dimensionalen Eigen- und Hauptraums zum Eigenwert 1

$$\mathrm{E}_1(A) = \mathrm{H}_1(A) = \mathrm{Kern}(\mathbb{I}_4{-}A) = \mathrm{Kern} \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 25 & -7 \\ 0 & 0 & 7 & -2 \\ 0 & 0 & 28 & -8 \end{pmatrix} = \mathbb{C} \begin{pmatrix} 0 \\ 1 \\ 2 \\ 7 \end{pmatrix},$$

und insgesamt bekommt man die Transformationsmatrix

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 4 & 7 \end{pmatrix} \qquad \text{mit} \qquad T^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & -\frac{4}{3} & \frac{1}{3} \\ 0 & 0 & -7 & 2 \\ 0 & 0 & 4 & -1 \end{pmatrix}.$$

Diese Matrix leistet die Transformation

$$A = T \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} T^{-1}$$

und mit den Regeln (7.10), (7.11) und zwei Matrix-Multiplikationen berechnet man nun

$$\mathrm{e}^{tA} = T \begin{pmatrix} \mathrm{e}^{2t} & 0 & 0 & 0 \\ 0 & \mathrm{e}^{2t} & t\mathrm{e}^{2t} & 0 \\ 0 & 0 & \mathrm{e}^{2t} & 0 \\ 0 & 0 & 0 & \mathrm{e}^{t} \end{pmatrix} T^{-1} = \begin{pmatrix} \mathrm{e}^{2t} & 0 & 0 & 0 \\ 0 & \mathrm{e}^{2t} & (-4-21t)\mathrm{e}^{2t} + 4\mathrm{e}^{t} & (1+6t)\mathrm{e}^{2t} - \mathrm{e}^{t} \\ 0 & 0 & -7\mathrm{e}^{2t} + 8\mathrm{e}^{t} & 2\mathrm{e}^{2t} - 2\mathrm{e}^{t} \\ 0 & 0 & -28\mathrm{e}^{2t} + 28\mathrm{e}^{t} & 8\mathrm{e}^{2t} - 7\mathrm{e}^{t} \end{pmatrix}.$$

Geht es nur um die Bestimmung einer beliebigen Fundamentalmatrix W des GDG-Systems, nicht um die spezielle Matrix e^{tA}, so erhält man mit weniger Rechenaufwand (nämlich ohne explizite Berechnung von T^{-1})

$$W(t) = \mathrm{e}^{tA}T = T \begin{pmatrix} \mathrm{e}^{2t} & 0 & 0 & 0 \\ 0 & \mathrm{e}^{2t} & t\mathrm{e}^{2t} & 0 \\ 0 & 0 & \mathrm{e}^{2t} & 0 \\ 0 & 0 & 0 & \mathrm{e}^{t} \end{pmatrix} = \begin{pmatrix} \mathrm{e}^{2t} & 0 & 0 & 0 \\ 0 & 3\mathrm{e}^{2t} & 3t\mathrm{e}^{2t} & \mathrm{e}^{t} \\ 0 & 0 & \mathrm{e}^{2t} & 2\mathrm{e}^{t} \\ 0 & 0 & 4\mathrm{e}^{2t} & 7\mathrm{e}^{t} \end{pmatrix},$$

und natürlich bilden die Spalten von W ein Fundamentalsystem.

Weitere Bemerkungen (zur **Berechnung *reeller* Fundamentalsysteme und -matrizen**). Bei reeller Koeffizientenmatrix $A \in \mathbb{R}^{N\times N}$ macht auch die Frage nach der Berechnung *reeller* Lösungen, Fundamentalsysteme und -matrizen zu

$$u' = Au \qquad \text{auf } \mathbb{R}$$

Sinn. Man kann allerdings nicht hoffen, die zuvor diskutierten Rechenverfahren immer analog über $\mathbb{R}$ statt $\mathbb{C}$ durchführen zu können, denn die Eigenwerte und -vektoren von $A \in \mathbb{R}^{N\times N}$ müssen nicht notwendigerweise reell sein. Vielmehr muss man in Fällen mit nicht-reellen Eigenwerten nach wie vor über $\mathbb{C}$ rechnen und am Ende der Rechnung eventuell und wie im Folgenden erläutert vom komplexen zu einem reellen Ergebnis übergehen:

(11) Ohne solch einen Übergang am Ende lässt sich allerdings noch die **Berechnung der reellen Fundamentalmatrix $\mathbf{e^{tA}}$** mit $A \in \mathbb{R}^{N\times N}$ durchführen, denn, auch wenn eine Rechnung mit komplexen Eigenwerten und -vektoren erforderlich wird, so ist das korrekte Endergebnis automatisch reell; das ergibt sich aus der Definition über die Exponentialreihe und auch aus der Vertauschbarkeit der Exponentialabbildung mit komplexer Konjugation.

(12) Bei der **Berechnung anderer reeller Fundamentalsysteme und -matrizen** wie $\mathrm{e}^{tA}T$ können dagegen nach einer Rechnung über $\mathbb{C}$ noch weitere Überlegungen erforderlich werden. Nützlich ist hierzu zunächst die Beobachtung, dass **nicht-reelle Eigenwerte von $\mathbf{A \in \mathbb{R}^{N\times N}}$ stets in zueinander konjugierten Paaren** $(\lambda, \overline{\lambda})$ **mit** zueinander **konjugierten Eigen- und Haupträumen** $\mathrm{E}_{\overline{\lambda}}(A) = \overline{\mathrm{E}_{\lambda}(A)}$, $\mathrm{H}_{\overline{\lambda}}(A) = \overline{\mathrm{H}_{\lambda}(A)}$ **auftreten** (dies folgt direkt aus der Beobachtung $(\lambda\mathbb{I}_N - A)^k v = 0 \iff \big(\overline{\lambda}\mathbb{I}_N - A\big)^k \overline{v} = 0$). Behandelt man bei der Berechnung komplexer Fundamentalsysteme $u_1, u_2, \ldots, u_N$ (beziehungsweise der Matrix $e^{tA}T$) gemäß (8), (9), (10) nun λ und $\overline{\lambda}$ völlig analog, so erreicht man problemlos, dass mit einer Funktion u_i stets auch die konjugierte Funktion $\overline{u_i}$ im Fundamentalsystem auftritt (beziehungsweise, dass $\mathrm{e}^{tA}T$ mit einer Spalte auch die konjugierte Spalte enthält). Nach eventuellem Umsortieren der Funktionen (beziehungsweise Spalten) bedeutet dies $u_{2i} = \overline{u_{2i-1}}$ für $i = 1, 2, \ldots, \ell$, während u_i für die restlichen Indizes $i = 2\ell{+}1, 2\ell{+}2, \ldots, N$ sowieso reell

ist. Es liegt somit ein **komplexes Fundamentalsystem** (beziehungsweise eine komplexe Fundamentalmatrix mit den Spalten)

$$u_1,\ \overline{u_1},\ u_2,\ \overline{u_2},\ \dots\ u_{2\ell-1},\ \overline{u_{2\ell-1}},\ \underbrace{u_{2\ell+1},\ u_{2\ell+2},\ u_N}_{\text{reell}}$$

vor, und **daraus ergibt sich durch Real- *und* Imaginärteilbildung ein reelles Fundamentalsystem** (beziehungsweise eine reelle Fundamentalmatrix mit den Spalten)

$$\operatorname{Re} u_1,\ \operatorname{Im} u_1,\ \operatorname{Re} u_2,\ \operatorname{Im} u_2,\ \dots,\ \operatorname{Re} u_{2\ell-1},\ \operatorname{Im} u_{2\ell-1},\ u_{2\ell+1},\ u_{2\ell+2},\ \dots,\ u_N\,.$$

Zur Erklärung dieses Sachverhalts reicht es, bei den allgemeinen Lösungen gemäß Teil (IV) von Satz 7.1 zu den Realteilen überzugehen: Die allgemeine komplexe Lösung kann mit obigem Fundamentalsystem als $\sum_{i=1}^{\ell}\left[C_{2i-1}u_{2i-1}+C_{2i}\overline{u_{2i-1}}\right]+\sum_{i=2\ell+1}^{N}C_iu_i$ mit Konstanten $C_1,C_2,\dots,C_N\in\mathbb{C}$ geschrieben werden, und durch Realteilbildung erhält man die allgemeine reelle Lösung $\sum_{i=1}^{\ell}\left[\widetilde{C}_{2i-1}\operatorname{Re}u_{2i-1}-\widetilde{C}_{2i}\operatorname{Im}u_{2i-1}\right]+\sum_{i=2\ell+1}^{N}\widetilde{C}_iu_i$ mit neuen Konstanten $\widetilde{C}_{2i-1}=\operatorname{Re}C_{2i-1}+\operatorname{Re}C_{2i}$ und $\widetilde{C}_{2i}=\operatorname{Im}C_{2i-1}-\operatorname{Im}C_{2i}$ für $i=1,2,\dots,\ell$ sowie $\widetilde{C}_i=\operatorname{Re}C_i$ für $i=2\ell+1,2\ell+2,\dots,N$. Somit bilden die zuvor angegebenen Funktionen ein reelles Fundamentalsystem.

Als Folgerung aus den Resultaten und Rechenverfahren dieses Abschnitts ergibt sich auch folgende Zusammenstellung von Aussagen über Fundamentalsysteme:

Korollar 7.10 (Struktur von Fundamentalsystemen zu $u'=Au$). *Gegeben sei ein homogenes lineares GDG-System*

$$u'=Au\qquad\text{auf }\mathbb{R}\tag{7.12}$$

mit konstanter Koeffizientenmatrix $A\in\mathbb{K}^{N\times N}$.

(I) *Im **komplexen Fall** $\mathbb{K}=\mathbb{C}$ gibt es ein komplexes Fundamentalsystem zu (7.12) von folgender Struktur: Für jeden Eigenwert $\lambda\in\mathbb{C}$ von A der algebraischen Vielfachheit d gehen insgesamt d Funktionen ins FS ein, und ihre Komponentenfunktionen sind $\mathbb{C}$-Linearkombinationen von*

$$\mathrm{e}^{\lambda t},\ t\mathrm{e}^{\lambda t},\ t^2\mathrm{e}^{\lambda t},\ \dots,\ t^{\ell-1}\mathrm{e}^{\lambda t}\,.$$

(II) *Im **reellen Fall** $\mathbb{K}=\mathbb{R}$ gibt es ein reelles Fundamentalsystem zu (7.12) von folgender Struktur: Für jeden reellen Eigenwert $\lambda\in\mathbb{R}$ von A der algebraischen Vielfachheit d gehen insgesamt d Funktionen ins FS ein, und ihre Komponentenfunktionen sind $\mathbb{R}$-Linearkombinationen von*

$$\mathrm{e}^{\lambda t},\ t\mathrm{e}^{\lambda t},\ t^2\mathrm{e}^{\lambda t},\ \dots,\ t^{\ell-1}\mathrm{e}^{\lambda t}\,.$$

Nicht-reelle Eigenwerte von $A\in\mathbb{R}^{N\times N}$ treten in zueinander konjugierten Paaren $\lambda,\overline{\lambda}\in\mathbb{C}\setminus\mathbb{R}$ der gleichen algebraischen Vielfachheit d auf. Für jedes solche Paar gehen insgesamt $2d$ Funktionen ins FS ein, und mit den Bezeichnungen $\mu:=\operatorname{Re}\lambda$, $\nu:=\operatorname{Im}\lambda$ sind ihre Komponentenfunktionen $\mathbb{R}$-Linearkombinationen von

$$\begin{aligned}&\mathrm{e}^{\mu t}\cos(\nu t),\ t\mathrm{e}^{\mu t}\cos(\nu t),\ t^2\mathrm{e}^{\mu t}\cos(\nu t),\ \dots,\ t^{\ell-1}\mathrm{e}^{\mu t}\cos(\nu t)\,,\\ &\mathrm{e}^{\mu t}\sin(\nu t),\ t\mathrm{e}^{\mu t}\sin(\nu t),\ t^2\mathrm{e}^{\mu t}\sin(\nu t),\ \dots,\ t^{\ell-1}\mathrm{e}^{\mu t}\sin(\nu t)\,.\end{aligned}$$

Dabei bezeichnet $\ell \in \{1, 2, \ldots, d\}$ die Größe des größten Jordan-Blocks zu λ in der Jordan-Normalform von A und damit den Nilpotenzgrad von $(\lambda \mathbb{I}_N - A)$ auf $\mathrm{H}_\lambda(A)$. Speziell gilt $\ell{=}1$ im Fall $\mathrm{H}_\lambda(A) = \mathrm{E}_\lambda(A)$ und somit stets im diagonalisierbaren Fall.

Beweis. Teil (I) ergibt sich direkt aus den Formeln (7.10) und (7.10) zur Berechnung der Fundamentalmatrix e^{tA}. Teil (II) folgt daraus gemäß Bemerkung (12), denn Real- und Imaginärteilbildung bei $\mathrm{e}^{\lambda t} = \mathrm{e}^{(\mu + \mathring{\mathbb{i}}\nu)t} = \mathrm{e}^{\mu t}\mathrm{e}^{\mathring{\mathbb{i}}\nu t}$ ergibt $\mathrm{e}^{\mu t}\cos(\nu t)$ und $\mathrm{e}^{\mu t}\sin(\nu t)$. □

Bemerkung (zum **Zusammenhang mit skalaren linearen GDGen *m*-ter Ordnung**)**.** Die in Abschnitt 5.2 beschriebene Theorie kann als Spezialfall der allgemeineren Resultate dieses Kapitels erklärt werden. Reduktion auf Ordnung 1 überführt nämlich die in 5.2 betrachtete skalare Gleichung (ohne Einschränkung mit Leitkoeffizient 1)

$$u^{(m)} + a_{m-1}u^{(m-1)} + \ldots + a_2 u'' + a_1 u' + a_0 u \equiv 0 \tag{7.13}$$

in ein System $U' = AU$ für $\mathbb{R}^m$-wertiges U (das dem $(m{-}1)$-Jet $u^{[m-1]}$ entspricht). Die (bei konstanten a_k ebenfalls konstante) Koeffizientenmatrix A hat dabei die Form

$$A := \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{m-2} & -a_{m-1} \end{pmatrix} \in \mathbb{K}^{m\times m},$$

und als ihr charakteristisches Polynom erhält man durch Entwicklung nach der letzten Zeile $\det(\lambda\mathbb{I}_m - A) = \lambda^m + a_{m-1}\lambda^{m-1} + \ldots + a_2\lambda^2 + a_1\lambda + a_0$. Dieses Polynom stimmt also mit dem in 5.2 definierten charakteristischen Polynom der Gleichung (7.13) überein, und die Eigenwerte von A sind genau die in 5.2 betrachteten Nullstellen. Die in Teil (I) des vorigen Korollars beschriebenen komplexen Lösungen sind folglich dieselben wie die des Satzes 5.2, und aus Teil (II) des Korollars erhält man nun auch reelle Analoga.

Eine Besonderheit des skalaren Falls besteht übrigens darin, dass in obigem Korollar stets der Fall $\ell = d$ eintritt, das heißt, es treten immer Potenzfunktionen bis hin zu t^{d-1} (mit der Vielfachheit d der Nullstelle λ) auf. Dies wurde mit Satz 5.2 bereits gezeigt, aber nun lässt es sich auch abstrakt begründen — und dies sogar auf zwei Weisen: Zum einen ist $\ell \leq d$ nach dem Korollar, und $\ell < d$ ist ausgeschlossen, da man sonst zu wenige Funktionen für ein Fundamentalsystem erhielte. Zum anderen sind die ersten $(m{-}1)$ Zeilen von $(\lambda\mathbb{I}_m - A)$ linear unabhängig; somit ist der Eigenraum $\mathrm{E}_\lambda(A)$ nur 1-dimensional, zu λ gehört nur ein einziger Jordan-Block der maximalen Größe d, und damit folgt (erneut) $\ell = d$ im Korollar.

Kapitel 8

Stabilität von Ruhelagen autonomer GDG-Systeme

8.1 Stabilitätsbegriffe für Lösungen und Ruhelagen

Für ein gegebenes GDG-System erster Ordnung und eine (explizit) bekannte spezielle Lösung u_0 auf einem Intervall I lässt sich folgende allgemeine Stabilitätsfrage formulieren: Wenn eine weitere (typischerweise nicht explizit bekannte) Lösung u desselben GDG-Systems zur Anfangszeit $t_0 \in I$ nur wenig von u_0 abweicht, bleibt dann die Differenz zwischen $u(t)$ und $u_0(t)$ für alle Zeiten $t \in I$ klein? Bei kompaktem I, also festem Zeithorizont, liefern die Resultate des Abschnitts 6.4 — wie früher angedeutet — eine allgemeine positive Antwort hierauf. Auf bis nach $+\infty$ reichenden Intervallen I dagegen ist die Frage nicht so allgemein zu beantworten, und gerade die Untersuchung dieser subtileren **Langzeit-Stabilität** ist Thema des vorliegenden Kapitels. Nützlich sind hierbei folgende, aus dem Kontext dynamischer Systeme übertragene Begriffe.

Definition 8.1 (**Stabilitätsbegriffe für Lösungen von GDG-Systemen**). *Sei $\mathcal{X}$ Banach-Raum über $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, und $f \in \mathrm{C}^0(D, \mathcal{X})$ erfülle lokal auf offenem $D \subset \mathbb{R} \times \mathcal{X}$ eine pLB in der $\mathcal{X}$-Variablen. Im Zusammenhang mit dem GDG-System (für $\mathcal{X}$-wertige u)*

$$u' = f(\,\cdot\,, u) \tag{8.1}$$

und einem nach oben unbeschränkten Intervall $I \subset \mathbb{R}$ werden vereinbart:

(1) *Eine Lösung u_0 von (8.1) auf I heißt* ***(Ljapunov-)stabil****, wenn es zu jedem $t_0 \in I$ und jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so dass gilt: Für alle $y_0 \in \mathcal{X}$ mit $|y_0 - u_0(t_0)| < \delta$ ist $(t_0, y_0) \in D$, und die Lösung u des AWPs zu (8.1) mit AB $u(t_0) = y_0$ existiert auf ganz $[t_0, \infty)$ mit $|u(t) - u_0(t)| < \varepsilon$ für alle $t \in [t_0, \infty)$.*

(2) *Eine nicht Ljapunov-stabile Lösung u_0 von (8.1) auf I nennt man* ***(Ljapunov-)instabil****.*

(3) *Eine Lösung u_0 von (8.1) auf I heißt* ***(lokal) attraktiv****, wenn es zu jedem $t_0 \in I$ ein $\delta > 0$ gibt, so dass gilt: Für alle $y_0 \in \mathcal{X}$ mit $|y_0 - u_0(t_0)| < \delta$ ist $(t_0, y_0) \in D$, und die Lösung u des AWPs zu (8.1) mit AB $u(t_0) = y_0$ existiert auf ganz $[t_0, \infty)$ mit $\lim_{t\to\infty} |u(t) - u_0(t)| = 0$.*

(4) *Eine Lösung u_0 von (8.1) auf I heißt* ***asymptotisch stabil****, wenn sie Ljapunov-stabil und lokal attraktiv ist.*

Bemerkungen (zu den **Stabilitätsbegriffen für Lösungen**).

(1) Die zu Beginn der Definition gemachten Voraussetzungen, insbesondere die pLB für f, stellen die lokale Lösbarkeit der betrachteten AWPe sicher (Satz von Picard-Lindelöf) und erlauben es überhaupt erst, von „der Lösung“ des AWPs zu sprechen.

Man kann die Stabilitätsbegriffe zwar auch ohne solche Voraussetzungen einführen, wenn man die Definitionen etwas vorsichtiger formuliert, in der Praxis ist dies aber kaum relevant.

(2) Handelt es sich bei (8.1) um ein autonomes System mit globaler Existenz, so bilden seine Lösungen gemäß Abschnitt 6.6 ein dynamisches System, und in diesem Fall ist die Stabilität einer Lösung u_0 eng verknüpft mit der Stabilität kompakter Mengen im Sinn des früheren Abschnitts 2.3. Genauer ist Stabilität von u_0 im Allgemeinen eine etwas stärkere Forderung als Stabilität von $\overline{u_0([t_0,\infty))}$ für ein/alle $t_0 \in I$, jedenfalls wenn $u_0([t_0,\infty))$ relativ kompakt in $\mathcal{X}$ und somit der Stabilitätsbegriff für Mengen überhaupt anwendbar ist.

Bemerkungen und Erläuterungen (zu **Stabilitätsbegriffen für Ruhelagen**). Von besonderem Interesse ist in der Stabilitätstheorie die **Untersuchung *autonomer* GDG-Systeme**

$$u' = F(u) \tag{8.2}$$

(mit lokal Lipschitz-stetigen Vektorfeld $F \in \mathrm{C}^0(D,\mathcal{X})$ auf offenem $D \subset \mathcal{X}$) **und ihrer *konstanten* Lösungen u_0**, für die $u_0(\mathbb{R}) = \{x_0\}$ nur aus einer Nullstelle $x_0 \in D$ von F besteht. Tatsächlich handelt es sich bei x_0 dann um eine Ruhelage im Sinn dynamischer Systeme, und im vorliegenden Kontext bezeichnet man neben x_0 auch die **konstante Lösung** u_0 mit Wert x_0 **als Ruhelage des GDG-Systems** (8.2).

Im weiteren Verlauf dieses Kapitels werden nur noch Ruhelagen autonomer Systeme als grundlegendster[1] Fall untersucht. **Speziell hierfür halten wir fest:**

(1) Stabilität der konstanten Lösung u_0 (auf jedem nach oben unbeschränkten Intervall) und Stabilität der Menge $u_0(\mathbb{R}) = \{x_0\}$ sind äquivalent.

(2) In diesem Fall ist es auch äquivalent, die Definitionen (1) und (3) nur für $t_0 = 0$ statt für alle $t_0 \in \mathbb{R}$ oder mit δ unabhängig von t_0 zu treffen; dies liegt daran, dass mit u auch $\widetilde{t} \mapsto u\big(t_0+\widetilde{t}\big)$ Lösung des autonomen Systems (8.2) ist.

(3) Gemäß Abschnitt 5.8 lässt sich das GDG-System (8.2) im Fall $\mathcal{X} = \mathbb{R}^2$ durch ein **Phasenraumdiagramm** in der Zeichenebene veranschaulichen. Mit Hilfe solcher Diagramme kann man sich eine **anschauliche Vorstellung von den Stabilitätsbegriffen** machen:

- Eine **Ruhelage** x_0 entspricht im Phasenraumdiagramm einfach einem Punkt in D, an dem eine konstante Trajektorie für alle Zeiten verbleibt.

[1] Tatsächlich lässt sich durch die am Ende von Abschnitt 6.6 diskutierte Vorgehensweise stets (formal) auf autonome GDG-Systeme reduzieren, und zur Reduktion auf Ruhelagen (und sogar auf Null-Lösungen) kann man folgendermaßen vorgehen: Ist u_0 spezielle Lösung des Systems $u' = f(\cdot\,,u)$, so entsprechen beliebige Lösungen u dieses Systems den Lösungen $\widetilde{u} = u - u_0$ des transformierten Systems $\widetilde{u}' = \widetilde{f}(\cdot\,,\widetilde{u}\,)$ mit Strukturfunktion $\widetilde{f}(t,\widetilde{x}) := f(t,u_0(t)+\widetilde{x}) - f(t,u_0(t))$; daher ist Stabilität von u_0 als Lösung des ursprünglichen Systems äquivalent zur Stabilität der Null-Lösung des transformierten Systems.

Trotz dieser beiden Möglichkeiten kann man eine *simultane* Reduktion auf autonome Systeme und Ruhelagen aber nicht allgemein erreichen (denn die Reduktion auf den autonomen Fall erhält Ruhelagen nicht, und die Reduktion auf Ruhelagen erhält Autonomie nicht).

- **Ljapunov-Stabilität** einer Ruhelage x_0 bedeutet, dass es zu jedem (potenziell kleinen) $\varepsilon > 0$ ein (noch) kleineres $\delta > 0$ mit $\mathrm{B}_\delta(x_0) \subset D$ und folgender Eigenschaft gibt: Jede Trajektorie, die $\mathrm{B}_\delta(x_0)$ (zur Zeit t_0) durchläuft, bleibt danach für immer (d.h. für alle Zeiten aus $[t_0, \infty)$) in $\mathrm{B}_\varepsilon(x_0)$. Im Phasenraumdiagramm erlaubt dies (leicht) verschiedene Verhaltensweisen der Trajektorien. Typische Ljapunov-stabile Ruhelagen sind die in den Abbildungen 11 und 12 skizzierten Senken, Strudel- und Wirbelpunkte; Abbildung 12 zeigt aber auch, dass weitere, beliebig nahe gelegene Ruhelagen möglich sind.

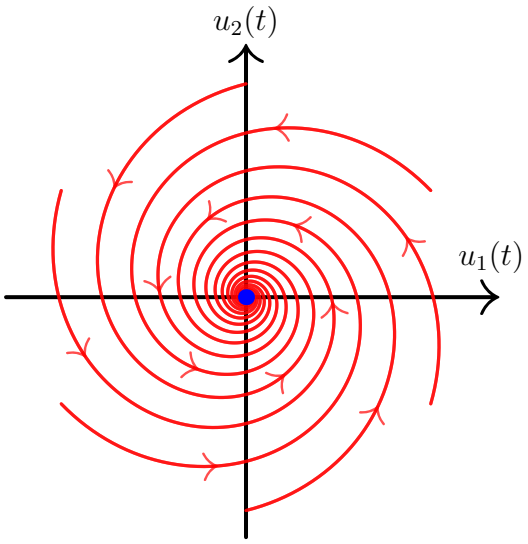


Abb. 11: Lösungskurven des linearen Systems $u_1' = -u_1 - 4u_2$, $u_2' = 4u_2 - u_1$ mit asymptotisch stabiler (und damit auch Ljapunov-stabiler) Ruhelage 0.

- Lokale **Attraktivität** einer Ruhelage x_0 bedeutet, dass für ein (eventuell sehr kleines) $\delta > 0$ mit $\mathrm{B}_\delta(x_0) \subset D$ folgende Eigenschaft vorliegt: Jede Trajektorie, die $\mathrm{B}_\delta(x_0)$ durchläuft, konvergiert schließlich (d.h. für $t \to \infty$) gegen x_0. Im Phasenraumdiagramm erscheint eine attraktive Ruhelage somit als Senke, in der alle nah herankommenden Trajektorien wie in Abbildung 11 schließlich verschwinden.

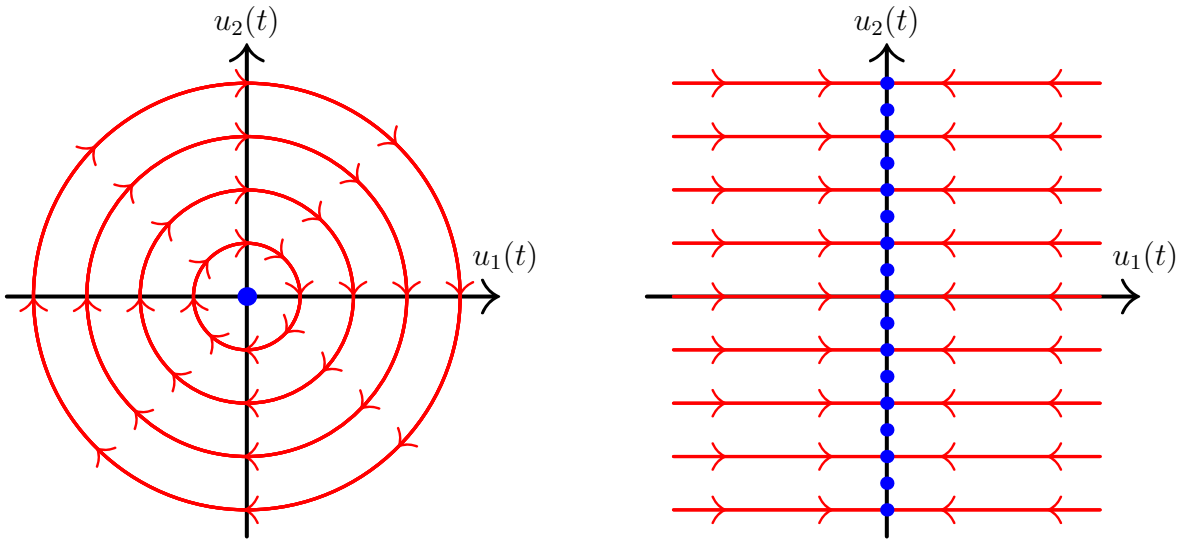


Abb. 12: Lösungskurven der linearen Systeme $u_1' = u_2$, $u_2' = -u_1$ (links) und $u_1' = -u_1$, $u_2' \equiv 0$ (rechts) mit Ljapunov-stabilen, aber nicht asymptotisch stabilen Ruhelagen.

- **Asymptotische Stabilität** beinhaltet Ljapunov-Stabilität *und* lokale Attraktivität und bedeutet somit, dass es zu jedem $\varepsilon > 0$ ein $\delta > 0$ mit $\mathrm{B}_\delta(x_0) \subset D$ und folgender Eigenschaft gibt: Jede Trajektorie, die $\mathrm{B}_\delta(x_0)$ durchläuft, bleibt danach für immer in $\mathrm{B}_\varepsilon(x_0)$ *und* konvergiert schließlich wie in Abbildung 11 gegen x_0. Im Allgemeinen ist dies stärker als Attraktivität allein, die ja erlauben würde, dass sich eine Trajektorie für gewisse $t > t_0$

weit von x_0 entfernt[2] und erst für sehr große t in die Nähe von x_0 zurückkehrt. Bei skalaren Gleichungen über $\mathbb{K} = \mathbb{R}$ wie auch bei linearen Systemen stellt sich asymptotische Stabilität aber doch als äquivalent zu Attraktivität heraus (dazu vergleiche man mit den Übungen beziehungsweise dem nächsten Abschnitt). Daher kann der Unterschied zwischen diesen Begriffen in den Abbildungen 11 und 12 nicht ausgemacht werden.

8.2 Stabilität von Ruhelagen linearer Systeme

Im Fall eines *linearen* Erster-Ordnung-GDG-Systems (für $\mathbb{K}^N$-wertige u)

$$u' = Au + b \qquad \text{auf } \mathbb{R} \tag{8.3}$$

mit konstanter Koeffizientenmatrix $A \in \mathbb{K}^{N\times N}$ und konstanter Inhomogenität $b \in \mathbb{K}^N$ sind die Ruhelagen des Systems genau die Lösungen $x \in \mathbb{K}^N$ des linearen Gleichungssystem $Ax = -b$. Gibt es also überhaupt eine Ruhelage (was jedenfalls bei invertierbarem A oder auch für $b = 0$ sichergestellt ist), so bilden die Ruhelagen einen affinen Unterraum von $\mathbb{K}^N$, im homogenen Fall $b = 0$ sogar einen Untervektorraum von $\mathbb{K}^N$.

Eine zentrale Beobachtung ist nun, dass die Stabilität der Ruhelagen von (8.3) tatsächlich nur von der Matrix A, nicht von b abhängt und alle Ruhelagen dieses Systems dasselbe Verhalten aufweisen; dies liegt daran, dass Lösungen u des (möglicherweise inhomogenen) Systems (8.3) nahe einer seiner Ruhelagen x_0 (mit $Ax_0 = -b$) den Lösungen $\widetilde{u} = u - x_0$ des homogenen Systems $\widetilde{u}' = A\widetilde{u}$ auf $\mathbb{R}$ nahe dessen Ruhelage 0 entsprechen. Folglich kann man sich darauf beschränken, die Stabilität der Null-Ruhelage von (8.3) im homogenen Fall $b = 0$ zu untersuchen.

Ausgehend von dieser Vorbemerkung ergeben sich aus Abschnitt 7.2 leicht zu prüfende, **notwendige und hinreichende Kriterien** für die Stabilität von Ruhelagen im linearen Fall; diese kann man als kontinuierliche Erweiterungen des früheren Korollars 3.3 mit $M = \mathrm{e}^A$ betrachten.

Satz 8.2 (über **Kriterien für die Stabilität von Ruhelagen linearer Systeme**). *Seien $A \in \mathbb{K}^{N\times N}$ und $b \in \mathbb{K}^N$ und seien $\lambda_1, \lambda_2, \dots, \lambda_\ell \in \mathbb{C}$ die verschiedenen Eigenwerte von A über $\mathbb{C}$ mit zugehörigen algebraischen Vielfachheiten $d_1, d_2, \dots, d_\ell \in \mathbb{N}$ und zugehörigen Eigenräumen $\mathrm{E}_{\lambda_1}(A), \mathrm{E}_{\lambda_2}(A), \dots, \mathrm{E}_{\lambda_\ell}(A) \subset \mathbb{C}^N$. Besitzt das System (8.3) mindestens eine Ruhelage, so gelten:*

(A) *Genau dann, wenn $\operatorname{Re}\lambda_i < 0$ für **alle** $i \in \{1, 2, \dots, \ell\}$ gilt, ist jede/eine Ruhelage von (8.3) **asymptotisch stabil**.*

(B) *Genau dann, wenn $\operatorname{Re}\lambda_i < 0$ oder $\operatorname{Re}\lambda_i = 0$, $\dim_{\mathbb{C}} \mathrm{E}_{\lambda_i}(A) = d_i$ für **alle** $i \in \{1, 2, \dots, \ell\}$ gilt, ist jede/eine Ruhelage von (8.3) **Ljapunov-stabil**.*

(C) *Genau dann, wenn $\operatorname{Re}\lambda_i > 0$ oder $\operatorname{Re}\lambda_i = 0$, $\dim_{\mathbb{C}} \mathrm{E}_{\lambda_i}(A) < d_i$ für **ein** $i \in \{1, 2, \dots, \ell\}$ gilt, ist jede/eine Ruhelage von (8.3) **instabil**.*

[2] Ein Beispiel für dieses Verhalten liefert das System $r' = (1-r)r$, $\varphi' = \sin\frac{\varphi}{2}$ für $\mathbb{R}^2$-wertige $u = (r\cos\varphi, r\sin\varphi)$ in Polarkoordinaten. In kartesischen Koordinaten nimmt dieses System die Form $u' = \big(1-|u|\big)u + \sqrt{\frac{|u|-u_1}{|u|}}\,u^\perp$ mit $u^\perp := (-u_2, u_1)$ an und weist in allen Punkten der Halbachse $\mathbb{R}_0^+ \times \{0\}$ instabile Ruhelagen auf. Die Ruhelage $(1, 0)$ ist allerdings (im Gegensatz zu allen anderen) attraktiv.

Die Möglichkeit eines solchen Verhaltens ist übrigens auch dafür verantwortlich, dass lokale Attraktivität einer Ruhelage im Gegensatz zu ihrer Ljapunov- oder asymptotischen Stabilität und trotz des Namens im Allgemeinen kein lokaler Begriff ist.

Bemerkung. Mit den Kriterien des Satzes 8.2 lässt sich die **Stabilitätsfrage** bei linearen Systemen des Typs (8.3) **völlig schematisch auf die Berechnung von Eigenwerten und eventuell Eigenräumen** der Koeffizientenmatrix A **zurückführen**. Konkrete Beispiele hierzu liefern Abbildung 11 (Eigenwerte $-1\pm 4\mathring{\imath}$, (A) anwendbar) und Abbildung 12 (links Eigenwerte $\pm\mathring{\imath}$, rechts Eigenwerte -1 und 0, jeweils (B) anwendbar). Weitere Beispiele sind Thema der Übungen.

Für den Beweis von Satz 8.2 ist entscheidend, dass alle Komponentenfunktionen von Lösungen u des homogenen Systems $u' = Au$ komplexe Linearkombinationen von Termen der Form $t^j \mathrm{e}^{\lambda_i t}$ sind; vergleiche Korollar 7.10 (I). Haben nun alle Eigenwerte λ_i negativen Realteil, so konvergieren die Exponentialfunktionen $|\mathrm{e}^{\lambda_i t}| = \mathrm{e}^{\operatorname{Re}\lambda_i t}$ bei $t \to \infty$ schneller gegen Null als die Potenzfunktionen t^j wachsen, und daher folgt $\lim_{t\to\infty} u(t) = 0$, also die Attraktivität der Null-Ruhelage. Hat dagegen ein Eigenwert λ_{i_0} positiven Realteil, so ist $\lim_{t\to\infty} t^j |\mathrm{e}^{\lambda_{i_0} t}| = \infty$, und es ergibt sich (jedenfalls wenn nicht alle Koeffizienten vor den $t^j \mathrm{e}^{\lambda_{i_0} t}$-Termen verschwinden) mit $\lim_{t\to\infty} |u(t)| = \infty$ die Instabilität der Null-Ruhelage. Damit sind die Kriterien (A) und (C) fast schon gezeigt. Dennoch folgt nun ein formaler und vollständiger Beweis des Satzes, der auf der vorausgehenden Argumentation fußt, diese aber — vor allem im etwas subtileren Fall von Eigenwerten mit Realteil Null — noch etwas ausbaut:

Beweis von Satz 8.2. Zuerst sei festgehalten, dass globale (eindeutige) Lösbarkeit aller AWPe durch Hauptsatz 7.2 sichergestellt wird und sich die Definitionen der Stabilitätsbegriffe in dieser Hinsicht etwas vereinfachen. Gemäß Vortext reicht es außerdem, für $b = 0$ die Null-Ruhelage des homogenen Systems $u' = Au$ zu betrachten.

Ist die Bedingung von (B) an die Eigenwerte $\lambda_1, \lambda_2, \ldots, \lambda_\ell$ erfüllt, so sind gemäß Korollar 7.10 (I) alle Einträge der Fundamentalmatrix e^{tA} Linearkombinationen von Termen der beiden Typen $t^j \mathrm{e}^{\lambda_i t}$ mit $j < d_i$, $\operatorname{Re}\lambda_i < 0$ und $\mathrm{e}^{\lambda_i t}$ mit $\operatorname{Re}\lambda_i = 0$ (wobei im Fall $\operatorname{Re}\lambda_i = 0$ durch die Bedingung $\dim_{\mathbb{C}} \mathrm{E}_{\lambda_i}(A) = d_i$ sichergestellt wird, dass $\mathrm{E}_{\lambda_i}(A) = \mathrm{H}_{\lambda_i}(A)$ gilt und keine Potenzen von t auftreten; siehe den letzten Satz von Korollar 7.10). Da die angegebenen Terme stetig sind und bei $t \to \infty$ beschränkt bleiben (beachte $|\mathrm{e}^{\lambda_i t}| = \mathrm{e}^{\operatorname{Re}\lambda_i t}$), ist $M := 1 + \sup_{t\geq 0} \|\mathrm{e}^{tA}\|$ endlich. Bei gegebenem $\varepsilon > 0$ kann man daher $\delta := \varepsilon/M > 0$ wählen und bekommt für die Lösung $u(t) = \mathrm{e}^{(t-t_0)A} y_0$ zur AB $u(t_0) = y_0$ mit $y_0 \in \mathbb{K}^N$, $|y_0| < \delta$ die Abschätzung $|u(t)| < M\delta = \varepsilon$ für alle $t \geq t_0$. Somit sind die Ljapunov-Stabilität der Null-Ruhelage und die Hin-Richtung von (B) gezeigt.

Sind sogar alle Eigenwerte negativ, so treten im vorausgehenden Argument nur Funktionen des Typs $t^j \mathrm{e}^{\lambda_i t}$ mit $\operatorname{Re}\lambda_i < 0$ auf, und daher gilt $\lim_{t\to\infty} u(t) = 0$ für Lösungen zu beliebigen Anfangsdaten. Dies reicht, um Attraktivität der Null-Ruhelage sicherzustellen, und insgesamt ist damit auch die Hin-Richtung von (A) gezeigt.

Ist die Bedingung von (C) an die Eigenwerte erfüllt, so gilt für ein $i_0 \in \{1, 2, \ldots, \ell\}$ entweder $\operatorname{Re}\lambda_{i_0} > 0$ oder $\operatorname{Re}\lambda_{i_0} = 0$, $\dim \mathrm{E}_{\lambda_{i_0}}(A) < d_{i_0}$. Im ersten Fall dieser Alternative gilt für Eigenwert-Eigenvektor-Lösungen $u(t) = \mathrm{e}^{\lambda_{i_0} t} v$ mit $v \in \mathrm{E}_{\lambda_{i_0}}(A) \setminus \{0\}$ stets $\lim_{t\to\infty} |u(t)| = \lim_{t\to\infty} \mathrm{e}^{\operatorname{Re}\lambda_{i_0} t}|v| = \infty$. Für kleine Eigenvektoren v erhält man also Lösungen mit beliebig kleinen Anfangsdaten, die bei $t \to \infty$ jede beschränkte Nullumgebung verlassen. Also ist die Null-Ruhelage instabil. Im zweiten Fall bringt man A durch eine invertierbare Transformationsmatrix $T \in \mathbb{C}^{N\times N}$ auf Jordan-Normalform $J = T^{-1}AT$. Wegen $\dim \mathrm{E}_{\lambda_{i_0}}(A) < d_{i_0} = \dim \mathrm{H}_{\lambda_{i_0}}(A)$ enthält J einen größten Jordan-Block der Größe $k \geq 2$ (wobei k mit dem Nilpotenzgrad von $(\lambda_{i_0}\mathbb{I}_N - A)$ auf $\mathrm{H}_{\lambda_{i_0}}(A)$ übereinstimmt). Gemäß der Regel (7.10) zur Berechnung der Matrix-Exponentialfunktion von Jordan-Blöcken enthält dann e^{tJ} eine Spalte, in der als von Null verschiedenen Einträge genau $\mathrm{e}^{\lambda_{i_0} t}$, $t\mathrm{e}^{\lambda_{i_0} t}$, $\frac{1}{2}t^2\mathrm{e}^{\lambda_{i_0} t}$, $\ldots$, $\frac{1}{(k-1)!}t^{k-1}\mathrm{e}^{\lambda_{i_0} t}$ auftreten. Es folgt, dass eine Spalte der Fundamentalmatrix $T\mathrm{e}^{tJ}$ aus Linearkombinationen dieser Funktionen besteht, und wegen der Invertierbarkeit von T ist in mindestens einem Eintrag dieser Spalte der Koeffizient vor dem am schnellsten wachsenden Term $\frac{1}{(k-1)!}t^{k-1}\mathrm{e}^{\lambda_{i_0} t}$ von Null verschieden. Für die durch diese Spalte gegebene Lösung u gilt daher $\lim_{t\to\infty} |u(t)| = \infty$. Durch Multiplikation von u mit kleinen Vorfaktoren kann man wieder Lösungen mit beliebig kleinen Anfangsdaten erhalten. Also ist die Null-Ruhelage auch in diesem Fall instabil, und die Hin-Richtung von (C) ist gezeigt.

Es verbleibt, die Rück-Richtungen nachzuweisen. Zur Behandlung der Rück-Richtung von (A) bemerkt man dabei, dass in der Situation des vorausgehenden Arguments die Null-Ruhelage nicht Ljapunov-stabil und damit erst recht nicht asymptotisch stabil ist. Daher ist nur noch nachzuweisen, dass bei Existenz eines $i_0 \in \{1, 2, \ldots, \ell\}$ mit $\operatorname{Re}\lambda_{i_0} = 0$ ebenfalls keine asymptotische Stabilität vorliegen kann. Hierzu betrachtet man erneut Eigenwert-Eigenvektor-Lösungen $u(t) = \mathrm{e}^{\lambda_{i_0} t} v$ mit $v \in \mathrm{E}_{\lambda_{i_0}}(A) \setminus \{0\}$ und bekommt $\lim_{t\to\infty} |u(t)| = |v| > 0$. Da durch beliebig kleine Eigenvektoren v auch beliebig kleine AWe realisiert werden können, ist die Null-Ruhelage auch in diesem Fall nicht asymptotisch stabil.

Die Rück-Richtung von (B) ergibt sich schließlich als Kontraposition der schon gezeigten Hin-Richtung von (C). Die Rück-Richtung von (C) folgt analog aus der Hin-Richtung von (B). □

8.3 Ljapunov-Funktionen und nicht-lineare Stabilität

Ein nützliches Hilfsmittel für Stabilitätsuntersuchungen bei nicht-linearen GDG-Systemen (mit $\mathbb{R}^N$-wertigen u) sind sogenannte Ljapunov-Funktionen, die wie folgt definiert werden.

Definition 8.3 (**Ljapunov-Funktionen**). *Eine Funktion $L \in \mathrm{C}^1(D, \mathbb{R})$ auf einer offenen Teilmenge D von $\mathbb{R}^N$ heißt eine Ljapunov-Funktion eines Vektorfelds $F\colon D \to \mathbb{R}^N$ beziehungsweise des GDG-Systems $u' = F(u)$, wenn die Richtungsableitung von L entlang des Vektorfelds F auf D nicht-positiv ist, d.h. wenn $\partial_F L \leq 0$ auf D gilt.*

Terminologie. *Die Richtungsableitung $\partial_F L$ von L entlang des Vektorfelds F wird dabei punktweise als ‚normale' Richtungsableitung von L entlang einzelner Vektoren $F(x)$ erklärt, genauer*

$$\partial_F L(x) := (\partial_{F(x)} L)(x) = \lim_{t \to 0} \frac{L(x+tF(x)) - L(x)}{t} \qquad \text{für alle } x \in D\,.$$

Aus dem bekannten Zusammenhang zwischen Richtungsableitungen und Gradienten folgt dann

$$\partial_F L = \nabla L \cdot F = \sum_{i=1}^{N} (\partial_i L) F_i \tag{8.4}$$

*mit dem Skalarprodukt „$\cdot$" von $\mathbb{R}^N$. Gelegentlich bietet sich auch die **alternative Notation $\dot{L}$** für $\partial_F L$ an, bei der das betrachtete Vektorfeld/System aus dem Kontext zu erschließen ist.*

Nützlich sind Ljapunov-Funktionen vor allem im Zusammenhang mit dem folgenden Satz.

Satz 8.4 (über die **direkte Methode von Ljapunov**). *Gegeben sei das autonome GDG-System (für $\mathbb{R}^N$-wertige u)*

$$u' = F(u) \tag{8.5}$$

mit lokal Lipschitz-stetiger Strukturfunktion $F \in \mathrm{C}^0(D, \mathbb{R}^N)$ auf einer offenen Teilmenge D von $\mathbb{R}^N$. Sei außerdem $x_0 \in D$ eine Ruhelage von (8.5) und L eine Ljapunov-Funktion zu (8.5).

(A) *Ist x_0 strikte lokale Minimalstelle von L, so ist die Ruhelage x_0 **Ljapunov-stabil**.*

(B) *Ist x_0 strikte lokale Minimalstelle von L und gilt für ein $\varepsilon > 0$ mit $\mathrm{B}_\varepsilon(x_0) \subset D$ die strikte Ungleichung $\partial_F L < 0$ auf $\mathrm{B}_\varepsilon(x_0) \setminus \{x_0\}$, so ist die Ruhelage x_0 **asymptotisch stabil**.*

(C) *Ist x_0 **keine** lokale Minimalstelle von L und gilt für ein $\varepsilon > 0$ mit $\mathrm{B}_\varepsilon(x_0) \subset D$ die strikte Ungleichung $\partial_F L < 0$ auf $\mathrm{B}_\varepsilon(x_0) \setminus \{x_0\}$, so ist die Ruhelage x_0 **instabil**.*

Bemerkungen.

(1) Für die Ruhelage x_0 gelten stets $F(x_0) = 0$ und $\partial_F L(x_0) = 0$. Deshalb sind die strikten Ungleichungen in (B) und (C) nicht auf ganz $\mathrm{B}_\varepsilon(x_0)$, sondern nur auf $\mathrm{B}_\varepsilon(x_0) \setminus \{x_0\}$ sinnvoll.

(2) **Mit L ist auch $L+c$ für jedes $c \in \mathbb{R}$ Ljapunov-Funktion.** Deshalb ist es keine Einschränkung und kommt in der Literatur oft vor, dass anstelle der strikten Minimumseigenschaft in (A) und (B) die Forderungen $L(x_0) = 0$ und $L > 0$ auf (einer Umgebung von x_0 in) $D \setminus \{x_0\}$ gestellt werden. Ähnliches gilt für die Situation aus (C).

(3) Jede konstante Funktion ist eine Ljapunov-Funktion im Sinne von Definition 8.3, besitzt aber weder *strikte* Minimalstellen noch Nicht-Minimalstellen und erlaubt deshalb nie die Anwendung von Satz 8.4. Die **Krux bei der Anwendung des Satzes besteht daher in der Wahl einer geeigneten** (notwendigerweise nicht-konstanten) **Ljapunov-Funktion** L, und tatsächlich ist es im Allgemeinen nicht einfach, solch ein geeignetes L zu finden oder auch nur seine Existenz sicherzustellen.

(4) **Eine Ljapunov-Funktion L kann man sich als eine Energie vorstellen.** Tatsächlich entspricht $L(x)$ bei einem physikalisch motivierten GDG-System $u' = F(u)$ häufig der Energie des Zustands $x \in D$ des zugrunde liegenden physikalischen Systems, und (nur) **mit dieser Interpretation im Hinterkopf kann man ein geeignetes L oft angeben**.

(5) Von zentraler Bedeutung ist folgende Beobachtung: Ist L Ljapunov-Funktion und u Lösung zu $u' = F(u)$, so gilt gemäß Kettenregel

$$\big[L(u)\big]' = \nabla L(u) \cdot u' = \nabla L(u) \cdot F(u) = \partial_F L(u) \leq 0\,, \tag{8.6}$$

die ‚Energie' $L(u)$ ist also nicht-wachsende Funktion der Zeit t, und **insgesamt ist L entlang allen Trajektorien des GDG-Systems nicht-wachsend**. Man kann sich daher den **Graph einer Ljapunov-Funktion L als Energie-Landschaft** vorstellen, in der sich Trajektorien nie aufwärts, sondern nur abwärts oder auf konstantem Level bewegen können, und vor diesem Hintergrund werden die Aussagen von Satz 8.4 auch anschaulich sehr plausibel.

Beweis von Teil (A) *des Satzes* 8.4. Sei x_0 strikte lokale Minimalstelle von L, und ohne Einschränkung sei $L(x_0) = 0$. Zum Nachweis der Ljapunov-Stabilität von x_0 anhand ihrer Definition reicht es, beliebige $t_0 \in \mathbb{R}$ und ausreichend kleine $\varepsilon > 0$ zu betrachten, so dass x_0 strikte Minimalstelle von L auf der abgeschlossenen Kugel $\overline{\mathrm{B}}_\varepsilon(x_0) \subset D$ ist. Insbesondere ist dann $\gamma := \min_{\partial \mathrm{B}_\varepsilon(x_0)} L > 0$. Wegen der Stetigkeit von L lässt sich nun ein (von t_0 unabhängiges) $\delta \in (0, \varepsilon]$ finden, so dass $L \leq \frac{1}{2}\gamma$ auf $\mathrm{B}_\delta(x_0)$ gilt. Für $y_0 \in \mathbb{R}^N$ mit $|y_0 - x_0| < \delta$ existiert gemäß den Hauptsätzen 6.1 und 6.3 eine eindeutige maximale Lösung u des AWPs zu (8.5) mit AB $u(t_0) = y_0$ auf einem maximalen Lösungsintervall (α, ω) mit $-\infty \leq \alpha < t_0 < \omega \leq \infty$. Gemäß Bemerkung (5) ist $L(u)$ auf (α, ω) nicht-wachsend, und es gelten

$$u(t_0) = y_0 \in \mathrm{B}_\delta(x_0) \subset \mathrm{B}_\varepsilon(x_0) \qquad \text{und} \qquad L(u) \leq L(u(t_0)) = L(y_0) \leq \frac{1}{2}\gamma \text{ auf } [t_0, \infty)\,.$$

In Anbetracht von $L \geq \gamma$ auf $\partial \mathrm{B}_\varepsilon(x_0)$ bedeutet dies, dass die Lösung u die Kugel $\mathrm{B}_\varepsilon(x_0)$ nicht verlassen kann und $|u(t) - x_0| < \varepsilon$ für alle $t \in [t_0, \omega)$ gilt. Insbesondere ist damit $\lim_{t \nearrow \omega} |u(t)| = \infty$ ausgeschlossen, also erzwingt (eine Bemerkung zu) Hauptsatz 6.3 schon $\omega = \infty$. Die eindeutige Lösung u des AWPs existiert folglich auf ganz $[t_0, \infty)$ mit $|u(t) - x_0| < \varepsilon$ für alle $t \in [t_0, \infty)$, und die Ljapunov-Stabilität von x_0 ist nachgewiesen. □

Beweis von Teil (B) *des Satzes* 8.4. Seien x_0 strikte lokale Minimalstelle von L und $t_0 \in \mathbb{R}$. Für ein eventuell verkleinertes ε und das zugehörige δ seien zudem alle Eigenschaften des vorigen Beweisteils sowie $\partial_F L < 0$ auf $\overline{\mathrm{B}}_\varepsilon(x_0) \setminus \{x_0\}$ erfüllt. Sei $y_0 \in \mathbb{R}^N$ mit $|y_0 - x_0| < \delta$, und u bezeichne die auf $[t_0, \infty)$ existente Lösung u des AWPs mit AB $u(t_0) = y_0$ und $|u(t) - x_0| < \varepsilon$ für alle $t \in [t_0, \infty)$. Da $L(u)$ auf $[t_0, \infty)$ nicht-wachsend und von unten beschränkt ist, gibt es eine Folge $(t_k)_{k \in \mathbb{N}}$ in $[t_0, \infty)$ mit $\lim_{k\to\infty} t_k = \infty$ und $\lim_{k\to\infty} \big[L(u)\big]'(t_k) = 0$. Nach Umschreiben mit (8.6) ist auch $\lim_{k\to\infty} \partial_F L(u(t_k)) = 0$, und durch Übergang zu einer Teilfolge lässt sich zudem erreichen, dass $y_\infty := \lim_{k\to\infty} u(t_k) \in \overline{\mathrm{B}}_\varepsilon(x_0)$ existiert. Es folgt $\partial_F L(y_\infty) = 0$, und dies erzwingt gemäß Annahme $y_\infty = x_0$. Als Nächstes folgt $\lim_{k\to\infty} L(u(t_k)) = L(y_\infty) = L(x_0)$ und wegen der Monotonie von $L(u)$ auch $\lim_{t\to\infty} L(u(t)) = L(x_0)$. Letzteres reicht, um auf $\lim_{t\to\infty} u(t) = x_0$ zu schließen, denn, sobald u entlang irgendeiner ∞-Folge in $[t_0, \infty)$ gegen ein $y \in \overline{\mathrm{B}}_\varepsilon(x_0) \setminus \{x_0\}$ konvergieren würde, müsste $L(u)$ entlang dieser Folge gegen $L(y) > L(x_0)$ konvergieren. Mit $\lim_{t\to\infty} u(t) = x_0$ ist die asymptotische Stabilität von x_0 nachgewiesen. □

Beweis von Teil (C) *des Satzes* 8.4. Sei x_0 keine lokale Minimalstelle von L, sei $L(x_0) = 0$, und sei ein eventuell verkleinertes $\varepsilon > 0$ mit $\partial_F L < 0$ auf $\overline{\mathrm{B}}_\varepsilon(x_0) \setminus \{x_0\}$ fixiert. Da x_0 keine lokale Minimalstelle ist, gibt es beliebig nah an x_0 gelegene $y_0 \in \overline{\mathrm{B}}_\varepsilon(x_0)$ mit $L(y_0) < 0$, und es reicht zu zeigen, dass, sobald eine Lösung u des AWPs zu (8.5) mit AB $u(0) = y_0$ auf ganz $\mathbb{R}_0^+$ existiert, diese Lösung an einer Stelle $|u| \geq \varepsilon$ erfüllt. Dazu benutzt man zuerst die Stetigkeit von L in x_0 und wählt eine (auch von y_0 abhängige) offene Umgebung V von x_0 in D mit $L > L(y_0)$ auf V. Wegen Stetigkeit von $\partial_F L$ ist $\ell := \min_{\overline{\mathrm{B}}_\varepsilon(x_0)\setminus V}(-\partial_F L) > 0$. Da $L(u)$ gemäß Bemerkung (5) nicht-wachsend ist, gelten $L(u) \leq L(u(0)) = L(y_0)$ und somit $u \notin V$ auf $\mathbb{R}_0^+$. Angenommen, es gälte nun $|u| < \varepsilon$ auf $\mathbb{R}_0^+$ und u bliebe somit in $\overline{\mathrm{B}}_\varepsilon(x_0) \setminus V$. Dann wäre $\big[L(u)\big]' = \partial_F L(u) \leq -\ell$ auf $\mathbb{R}_0^+$ und somit $\lim_{t\to\infty} L(u(t)) = -\infty$. Dies stünde im Widerspruch dazu, dass L auf dem Kompaktum $\overline{\mathrm{B}}_\varepsilon(x_0)$ von unten beschränkt ist, also muss $|u| \geq \varepsilon$ an einer Stelle in $\mathbb{R}_0^+$ gelten, und die Instabilität von x_0 ist gezeigt. □

Anwendung (auf die **Gleichung(en) des mathematischen Pendels**). Als mathematisches Pendel bezeichnet man das in Abbildung 13 skizzierte, idealisierte Modell für ein Stabpendel. Dabei ist der masselose, unendlich dünne, unflexible Pendelstab an einem Ende an einem festen Punkt befestigt, am anderen Ende trägt er eine Punktmasse, und das Pendel schwingt reibungsfrei nur unter Einfluss der Gravitation in einer vertikalen Ebene. Verwendet man den Winkel φ des Pendelstabs relativ zur unteren Ruheposition und die (skalare) Geschwindigkeit v der Punktmasse als Variablen, so wird ein solches Pendel durch das nicht-lineare Erster-Ordnung-GDG-System für $\mathbb{R}^2$-wertige Funktionen $u = \binom{\varphi}{v}$

$$\varphi' = \ell^{-1} v\,, \qquad v' = -g \sin\varphi \tag{8.7}$$

mit den festen, positiven Parametern ℓ (Länge des Pendelstabs) und g (Fallbeschleunigung) beschrieben. Äquivalent ist die skalare, nicht-lineare Zweiter-Ordnung-GDG $\varphi'' = -\frac{g}{\ell}\sin\varphi$, die Untersuchung im Kontext dieses Kapitels orientiert sich aber am System (8.7) mit der Strukturfunktion

$$F(\varphi, v) = \begin{pmatrix} \ell^{-1} v \\ -g\sin\varphi \end{pmatrix} \qquad \text{für } (\varphi, v) \in \mathbb{R}^2\,.$$

Die abzählbar vielen Ruhelagen $(z\pi, 0)$, $z \in \mathbb{Z}$ des Systems (8.7) entsprechen wegen der ‚Periodizität' des Winkels φ nur zwei physikalisch unterscheidbaren Zuständen: Bei $(2k\pi, 0)$, $k \in \mathbb{Z}$ handelt es sich um die untere, bei $((2k{+}1)\pi, 0)$, $k \in \mathbb{Z}$ um die obere Ruhelage des Pendels. Diese beiden Fälle werden nun getrennt diskutiert.

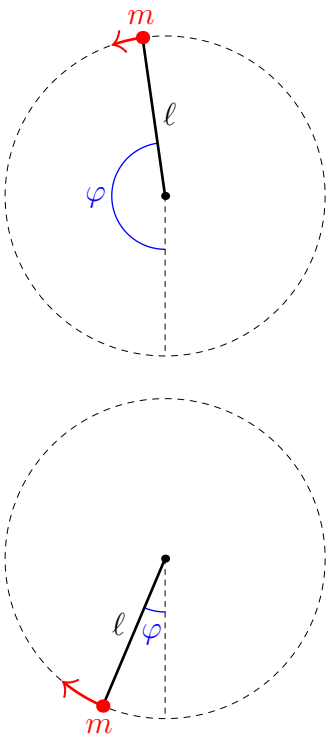


Abb. 13: Das Pendel nahe den Ruhelagen.

- Bei der Untersuchung der ‚unteren' Ruhelagen $(2k\pi, 0)$ mit $k \in \mathbb{Z}$ ist es hilfreich, die Summe der kinetischen und der potentiellen Energie

$$L(\varphi, v) := \frac{1}{2} m v^2 - mg\ell\cos\varphi \qquad \text{für } (\varphi, v) \in \mathbb{R}^2 \tag{8.8}$$

zu betrachten — mit der Masse $m > 0$ des Pendels als physikalischem Parameter, den man für die rein mathematische Betrachtung auch zu 1 normieren könnte. Wegen

$$\dot{L}(\varphi, v) = \nabla L(\varphi, v) \cdot F(\varphi, v) = \begin{pmatrix} mg\ell\sin\varphi \\ mv \end{pmatrix} \cdot \begin{pmatrix} \ell^{-1} v \\ -g\sin\varphi \end{pmatrix} = 0$$

für alle $(\varphi, v) \in \mathbb{R}^2$ ist L Ljapunov-Funktion des Systems (8.7). Da L außerdem strikte lokale Minimalstellen bei $(2k\pi, 0)$ besitzt, sind diese Ruhelagen gemäß Satz 8.4 (A) Ljapunov-stabil.

Die Frage, ob die ‚unteren' Ruhelagen sogar asymptotisch stabil sind, lässt sich ebenfalls mit Hilfe von L aus (8.8), wenn auch nicht direkt anhand des Satzes, beantworten. Tatsächlich liegt nämlich mit der Identität $\dot{L} \equiv 0$ Energie-Erhaltung vor, so dass L nicht nur nicht-wachsend, sondern sogar konstant entlang Trajektorien ist. Folglich starten und bleiben alle anderen Lösungen u auf einem höheren Energie-Niveau als die Ruhelagen $(2k\pi, 0)$. Es gilt $\lim_{t\to\infty} L(u(t)) > L(2k\pi, 0)$, womit $\lim_{t\to\infty} u(t) = (2k\pi, 0)$ ausgeschlossen ist, und daher sind die ‚unteren' Ruhelagen zwar Ljapunov-stabil, aber nicht asymptotisch stabil.

- Bei der Untersuchung der ‚oberen' Ruhelagen $((2k+1)\pi, 0)$ mit $k \in \mathbb{Z}$ lässt sich Satz 8.4 nicht (direkt) mit der Ljapunov-Funktion L aus (8.8) anwenden. Betrachtet man aber beispielsweise[3]

$$L_0(\varphi, v) := v \sin\varphi \qquad \text{für } (\varphi, v) \in \mathbb{R}^2 ,$$

so gilt

$$\dot{L}_0(\varphi, v) = \nabla L_0(\varphi, v) \cdot F(\varphi, v) = \begin{pmatrix} v\cos\varphi \\ \sin\varphi \end{pmatrix} \cdot \begin{pmatrix} \ell^{-1} v \\ -g\sin\varphi \end{pmatrix} = \ell^{-1} v^2 \cos\varphi - g(\sin\varphi)^2 < 0$$

für $(\varphi, v) \in \left[\left((2k+\frac{1}{2})\pi, (2k+\frac{3}{2})\pi\right) \times \mathbb{R}\right] \setminus \{((2k+1)\pi, 0)\}$ (denn für die hier zugelassenen φ ist der Kosinus negativ und der Sinus hat nur eine Nullstelle). Da L_0 keine Minimalstellen, sondern Sattelpunkte bei $((2k+1)\pi, 0)$ besitzt, sind diese Ruhelagen gemäß Satz 8.4 (C) instabil.

Die hiermit verifizierten Stabilitätseigenschaften (und auch das sonstige Verhalten des Pendels) lassen sich gut im Phasenraumdiagramm der Abbildung 14 erkennen.

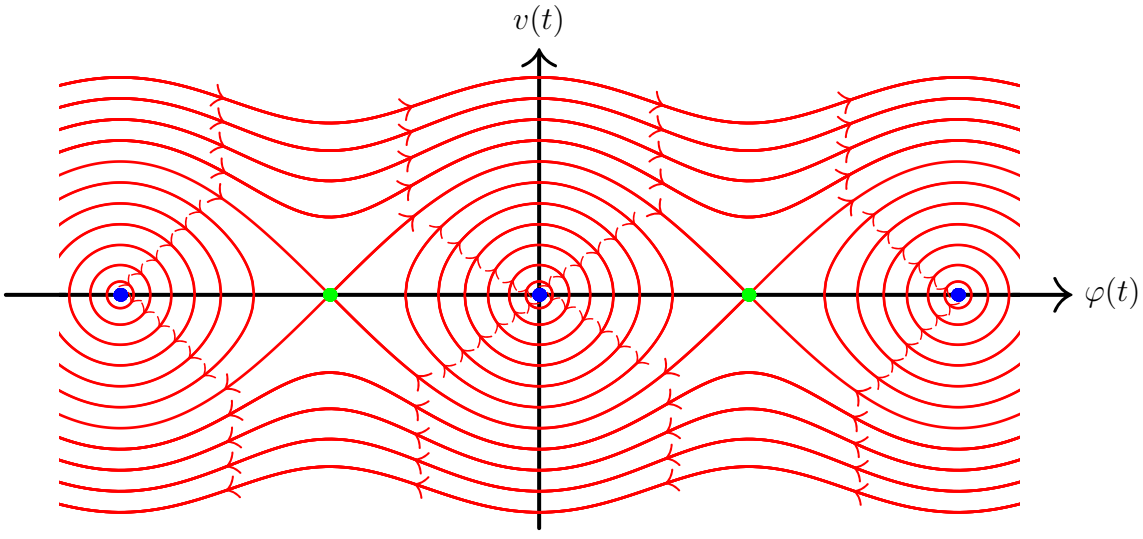


Abb. 14: Phasenraumdiagramm einiger Lösungen der Pendelgleichungen (8.7) mit den Ljapunov-stabilen Ruhelagen $(-2\pi, 0)$, $(0, 0)$, $(2\pi, 0)$ und den instabilen Ruhelagen $(-\pi, 0)$, $(\pi, 0)$.

[3] Neben dem hier gewählten L_0 führen auch etwas andere Wahlen wie $L_0(\varphi, v) = -(\varphi - (2k+1)\pi)v$ zum gleichen Ergebnis.

8.5 Linearisierungskriterien für Stabilität von Ruhelagen nicht-linearer Systeme

Eine andere naheliegende Methode zur Stabilitätsuntersuchung bei einem nicht-linearen GDG-System basiert auf der lokalen **Linearisierung** des Systems um eine Ruhelage, und tatsächlich erhält man für den nicht-linearen Fall dann ähnliche (aber weniger scharfe und vollständige) Kriterien wie in Satz 8.2 für den linearen Fall. Diese Kriterien sind die Zeit-kontinuierlichen Analoga der Linearisierungskriterien aus Satz 3.5 und werden hier nur über $\mathbb{K} = \mathbb{R}$ angegeben (für den Fall $\mathbb{K} = \mathbb{C}$ vergleiche man aber mit Fußnoten 2 und 3 in Abschnitt 3.3).

Satz 8.5 (Linearisierungskriterien für (In-)Stabilität bei nicht-linearen Systemen). *Gegeben sei das autonome GDG-System (für $\mathbb{R}^N$-wertige u)*

$$u' = F(u) \tag{8.9}$$

mit Strukturfunktion $F \in \mathrm{C}^1(D, \mathbb{R}^N)$ auf einer offenen Teilmenge D von $\mathbb{R}^N$.

(A) *Ist x_0 eine Ruhelage von (8.9) und haben alle Eigenwerte von $\mathrm{D}F(x_0) \in \mathbb{R}^{N\times N}$ über $\mathbb{C}$ negativen Realteil, so ist die Ruhelage x_0* ***asymptotisch stabil****.*

(B) *Ist x_0 eine Ruhelage von (8.9) und gibt es einen Eigenwert von $\mathrm{D}F(x_0) \in \mathbb{R}^{N\times N}$ über $\mathbb{C}$ mit positivem Realteil, so ist die Ruhelage x_0* ***instabil****.*

Bemerkung. Besitzt $\mathrm{D}F(x_0) \in \mathbb{R}^{N\times N}$ einen Eigenwert mit Realteil Null, aber keinen Eigenwert mit positivem Realteil, so kann man mit den Linearisierungskriterien nicht über die Stabilität der Ruhelage x_0 entscheiden. Tatsächlich hängt in diesem Fall die Stabilität von x_0 nicht nur vom Erster-Ordnung-Verhalten von F nahe x_0, sondern auch von Effekten höherer Ordnung ab, und insbesondere gibt es daher kein nicht-lineares Analogon zu Teil (B) des linearen Satzes 8.2.

Anwendung. Im Fall der Pendelgleichungen (8.7) mit der Strukturfunktion $F(\varphi, v) = \left(\begin{smallmatrix} \ell^{-1}v \\ -g\sin\varphi \end{smallmatrix}\right)$ lässt sich ein Teil der schon verifizierten Ergebnisse mit Satz 8.5 noch einmal bestätigen: Bei den Ruhelagen $((2k{+}1)\pi, 0)$, $k \in \mathbb{Z}$ der Pendelgleichungen weist $\mathrm{D}F((2k{+}1)\pi, 0) = \left(\begin{smallmatrix} 0 & \ell^{-1} \\ g & 0 \end{smallmatrix}\right)$ den positiven Eigenwert $\sqrt{g\ell^{-1}}$ auf, also liefert Satz 8.5 (B) erneut die Instabilität dieser ‚oberen' Ruhelagen. Bei den Ruhelagen $(2k\pi, 0)$ dagegen besitzt $\mathrm{D}F(2k\pi, 0) = \left(\begin{smallmatrix} 0 & \ell^{-1} \\ -g & 0 \end{smallmatrix}\right)$ die beiden rein imaginären Eigenwerte $\pm\mathring{\mathrm{i}}\sqrt{g\ell^{-1}}$, und mit Satz 8.5 allein lässt sich über die Stabilität dieser ‚unteren' Ruhelagen keine Aussage treffen.

Zum Abschluss dieses Kapitels wird ein Beweis von Satz 8.5 gegeben und Teil (A) sogar mit zwei verschiedenen Methoden bewiesen — einmal mit Hilfe von Ljapunov-Funktionen und ein weiteres Mal mit Hilfe eines Gronwall-Lemmas.

Erster Beweis von Satz 8.5 (A) mit Ljapunov-Funktionen. Ohne Einschränkung sei $x_0 = 0$, und es sei

$$A := \mathrm{D}F(0) \in \mathbb{R}^{N\times N}$$

abgekürzt. Da $F(0)$ verschwindet, gilt gemäß dem Satz von Taylor $F(x) = Ax + \mathrm{o}(|x|)$ bei $x \to 0$, und insbesondere gibt es zu jedem $\varepsilon > 0$ (das unten geeignet fixiert wird) ein $\delta > 0$ mit $\mathrm{B}_\delta(0) \subset D$ und

$$\big|F(x) - Ax\big| \le \varepsilon|x| \qquad \text{für alle } x \in \mathrm{B}_\delta(0)\,. \tag{8.10}$$

Obwohl als Nächstes noch ein allgemeineres Argument ausgeführt wird, sei die wesentliche Beweisstrategie zuerst für symmetrisches A illustriert. In diesem Fall sind alle Eigenwerte von A reell und negativ, die Inverse A^{-1} existiert als ebenfalls symmetrische Matrix mit lauter negativen reellen Eigenwerten, und durch

$$L(x) := -x \cdot A^{-1}x \qquad \text{für } x \in \mathbb{R}^N$$

ist eine positiv definite quadratische Form L auf $\mathbb{R}^N$ gegeben. Insbesondere besitzt $L \in \mathrm{C}^1(\mathrm{B}_\delta(0), \mathbb{R})$ eine strikte Minimalstelle bei 0 und erfüllt $\nabla L(x) = -2A^{-1}x$ für $x \in \mathrm{B}_\delta(0)$. Unter Verwendung von (8.10) mit $\varepsilon = \frac{1}{2\|A^{-1}\|}$ folgt

$$\begin{aligned}\partial_F L(x) &= -2A^{-1}x \cdot Ax - 2A^{-1}x \cdot \big[F(x) - Ax\big] \\ &\le -2x \cdot A^{-1}Ax + 2\big\|A^{-1}\big\|\,|x|\,|F(x) - Ax| \\ &\le -2|x|^2 + |x|^2 < 0\end{aligned}$$

für alle $x \in \mathrm{B}_\delta(0) \setminus \{0\}$. Damit ist L eine Ljapunov-Funktion des Systems (8.9) auf $\mathrm{B}_\delta(0)$, für die Satz 8.4 (B) anwendbar ist, und es folgt asymptotische Stabilität der Null-Ruhelage. Für symmetrisches A ist der Beweis damit vollständig.

Als Nächstes wird der allgemeine Fall ohne Zusatzvoraussetzung an A behandelt. Obige Argumentation greift dann nicht mehr, da es bei nicht-symmetrischem A keine Form L mit $\nabla L(x) \cdot y = -x \cdot A^{-1}y$ gibt. Als Ersatz reicht allerdings eine Funktion L, die dieser Gleichung nur für $y = Ax$ genügt, also $\nabla L(x) \cdot Ax = -|x|^2$ erfüllt, und eine quadratische Form L mit der letzten Eigenschaft kann tatsächlich mit Hilfe eines Integrationstricks explizit angegeben werden. Dazu argumentiert man wie folgt. Da alle Eigenwerte von A negativen Realteil haben, sind gemäß Korollar 7.10 (I) alle Einträge der Fundamentalmatrix $\mathrm{e}^{tA} \in \mathbb{R}^{N\times N}$ komplexe Linearkombinationen von Termen der Form $t^j\mathrm{e}^{\lambda t}$ mit $\lambda \in \mathbb{C}$, $\operatorname{Re}\lambda < 0$, $j \in \mathbb{N}_0$. Es folgt, dass $\big\|\mathrm{e}^{tA}\big\|$ bei $t \to \infty$ exponentiell schnell gegen 0 konvergiert, und damit ist insbesondere

$$M := \int_0^\infty \big\|\mathrm{e}^{tA}\big\|^2 \,\mathrm{d}t < \infty\,. \tag{8.11}$$

Durch

$$L(x) := \int_0^\infty \big|\mathrm{e}^{tA}x\big|^2 \,\mathrm{d}t \qquad \text{für } x \in \mathbb{R}^N$$

erhält man folglich eine wohldefinierte quadratische Form L auf $\mathbb{R}^N$ mit strikter Minimalstelle 0 (denn wegen der Invertierbarkeit von e^{tA} ist $\big|\mathrm{e}^{tA}x\big| > 0$ für $x \neq 0$) und mit

$$\nabla L(x) \cdot y = 2\int_0^\infty \mathrm{e}^{tA}x \cdot \mathrm{e}^{tA}y \,\mathrm{d}t \qquad \text{für } x, y \in \mathbb{R}^N\,.$$

Unter Verwendung der Regel $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{tA} = \mathrm{e}^{tA}A$, der Abschätzung (8.10) mit $\varepsilon = \frac{1}{4M}$, des HDI und der Wahl (8.11) ergibt sich

$$\begin{aligned}\partial_F L(x) &= 2\int_0^\infty \mathrm{e}^{tA}x \cdot \mathrm{e}^{tA}Ax \,\mathrm{d}t + 2\int_0^\infty \mathrm{e}^{tA}x \cdot \mathrm{e}^{tA}\big[F(x) - Ax\big] \,\mathrm{d}t \\ &\le \int_0^\infty \frac{\mathrm{d}}{\mathrm{d}t}\big|\mathrm{e}^{tA}x\big|^2 \,\mathrm{d}t + \frac{1}{2M}\int_0^\infty \big\|\mathrm{e}^{tA}\big\|^2 \,\mathrm{d}t\;|x|^2 \\ &= \lim_{t\to\infty} |\mathrm{e}^{tA}x|^2 - |x|^2 + \tfrac{1}{2}|x|^2 \\ &= -\tfrac{1}{2}|x|^2 < 0\end{aligned}$$

für alle $x \in \mathrm{B}_\delta(0) \setminus \{0\}$. Damit ist auch im allgemeinen Fall eine Ljapunov-Funktion L des Systems (8.9) auf $\mathrm{B}_\delta(0)$ gefunden, für die sich Satz 8.4 (B) anwenden lässt, und es folgt die Behauptung. □

Eine andere Methode zum Beweis von Satz 8.5 (A) basiert auf dem aus Abschnitt 6.4 bekannten Gronwall-Lemma.

Zweiter Beweis von Satz 8.5 (A) mit dem Gronwall-Lemma. Ohne Einschränkung sei $x_0 = 0$, und es sei erneut

$$A := \mathrm{D}F(0) \in \mathbb{R}^{N\times N}$$

abgekürzt. Wie im vorigen Beweis begründet, stellt die Bedingung an die Eigenwerte von A sicher, dass $\|\mathrm{e}^{sA}\|$ bei $s \to \infty$ exponentiell schnell gegen 0 konvergiert, und insbesondere gibt es daher ein $\gamma > 0$ und ein $M \in \mathbb{R}^+$ mit

$$\|\mathrm{e}^{sA}\| \le M\mathrm{e}^{-\gamma s} \qquad \text{für alle } s \in \mathbb{R}_0^+ . \tag{8.12}$$

Gemäß dem Satz von Taylor gibt es außerdem zu beliebigem $\varepsilon > 0$ ein $\delta \in (0, \varepsilon)$ mit $\mathrm{B}_\delta(0) \subset D$, so dass

$$|F(x) - Ax| \le \frac{\gamma}{2M}|x| \qquad \text{für alle } x \in \mathrm{B}_\delta(0) \tag{8.13}$$

erfüllt ist, und für $t_0 \in \mathbb{R}$ und $y_0 \in \mathrm{B}_\delta(0)$ bezeichne im Folgenden $u\colon (\alpha, \omega) \to \mathbb{K}^N$ die maximale Lösung des AWPs $u' = F(u)$, $u(t_0) = y_0$. Die entscheidende Beweisidee ist es nun, das nichtlineare GDG-System in der Form

$$u' = Au + \big[F(u) - Au\big]$$

zu schreiben und als lineares GDG-System mit der von u abhängigen Inhomogenität $b(s) := \big[F(u(s)) - Au(s)\big]$ aufzufassen. Bei dieser Betrachtungsweise kann die Lösungsformel (7.5) mit der Fundamentalmatrix $W(s) = \mathrm{e}^{sA}$ und ihrer Inversen $W(s)^{-1} = e^{-sA}$ angewandt werden, und nach Anwendung des Exponentialgesetzes ergibt sich

$$u(t) = \mathrm{e}^{(t-t_0)A}y_0 + \int_{t_0}^{t} \mathrm{e}^{(t-s)A}\big[F(u(s)) - Au(s)\big]\,\mathrm{d}s \qquad \text{für alle } t \in (\alpha, \omega)$$

Falls u auf $[t_0, \omega)$ nicht in $\mathrm{B}_\delta(0)$ bleibt, so bezeichne im Folgenden $\omega_0 \in (t_0, \omega)$ die kleinste Zeit mit $u(\omega_0) \notin \mathrm{B}_\delta(0)$. Falls u in $\mathrm{B}_\delta(0)$ bleibt, so garantiert der Satz über die maximale Lösung, dass $\omega = \infty$ gilt, und dann sei auch $\omega_0 := \infty$. In jedem Fall ist dann $u((t_0, \omega_0)) \subset \mathrm{B}_\delta(0)$, und mit Hilfe von (8.12) und (8.13) ergibt sich aus der vorausgehenden Lösungsformel die Abschätzung

$$|u(t)| \le M\mathrm{e}^{-\gamma(t-t_0)}|y_0| + \frac{\gamma}{2}\int_{t_0}^{t} \mathrm{e}^{-\gamma(t-s)}|u(s)|\,\mathrm{d}s \qquad \text{für alle } t \in (t_0, \omega_0) .$$

Für die Hilfsfunktion $\varphi(s) := \mathrm{e}^{\gamma s}|u(s)|$ bedeutet dies

$$\varphi(t) \le M\mathrm{e}^{\gamma t_0}|y_0| + \frac{\gamma}{2}\int_{t_0}^{t} \varphi(s)\,\mathrm{d}s \qquad \text{für alle } t \in (t_0, \omega_0) .$$

Das Gronwall-Lemma (mit $a \equiv \frac{\gamma}{2}$, $b \equiv 0$, $s_0 = M\mathrm{e}^{\gamma t_0}|y_0|$) liefert daraus

$$\varphi(t) \le \mathrm{e}^{\frac{\gamma}{2}(t-t_0)}M\mathrm{e}^{\gamma t_0}|y_0| \qquad \text{für alle } t \in [t_0, \omega_0) ,$$

und für die Lösung u erhält man

$$|u(t)| \leq \mathrm{e}^{-\frac{\gamma}{2}(t-t_0)} M |y_0| \qquad \text{für alle } t \in [t_0, \omega_0)\,.$$

Liegt dann y_0 in $\mathrm{B}_{\delta/M}(0)$ und ist $\omega_0 < \infty$, so impliziert die vorige Abschätzung $u(\omega_0) \in \mathrm{B}_\delta(0)$ und steht damit im Widerspruch zur Wahl von ω_0 mit $u(\omega_0) \notin \mathrm{B}_\delta(0)$. Für $y_0 \in \mathrm{B}_{\delta/M}(0)$ gilt also $\omega_0 = \infty$, und die Lösung u existiert auf $[t_0, \infty)$ mit $u([t_0,\infty)) \subset \mathrm{B}_\delta(0) \subset \mathrm{B}_\varepsilon(0)$ und $\lim_{t\to\infty} |u(t)| = 0$. Damit ist die asymptotische Stabilität der Null-Ruhelage gezeigt, und die Behauptung ist nachgewiesen. □

Zum Beweis von Satz 8.5 (B) gibt es ebenfalls verschiedene Vorgehensweisen, die auch technisch unterschiedlich anspruchsvoll sind. Hier wird ein vergleichsweise elementares Argument angegeben, das Elemente des vorausgehenden Beweises und des Beweises von Satz 3.5 kombiniert.

Beweis von Satz 8.5 (B). Seien wieder $x_0 = 0$ und $A := \mathrm{D}F(0) \in \mathbb{R}^{N\times N}$, und seien $\lambda_1, \lambda_2, \lambda_\ell \in \mathbb{C}$ die verschiedenen Eigenwerte von A über $\mathbb{C}$. Sei zudem $\gamma > 0$ derart fixiert, dass kein Eigenwert einen Realteil zwischen 0 und γ aufweist, also so, dass $\lambda_1, \lambda_2, \ldots, \lambda_\ell$ alle nicht im offenen Streifen $(0,\gamma)+\mathbf{i}\mathbb{R}$ der Breite γ in der komplexen Zahlenebene liegen. Dann lässt sich $\mathbb{C}^N = V_1 \oplus V_2$ in die $\mathbb{C}$-linearen Unterräume

$$V_1 := \bigoplus_{\substack{i=1 \\ \mathrm{Re}\,\lambda_i \leq 0}}^{p} \mathrm{H}_{\lambda_i}(A) \qquad \text{und} \qquad V_2 := \bigoplus_{\substack{i=1 \\ \mathrm{Re}\,\lambda_i \geq \gamma}}^{p} \mathrm{H}_{\lambda_i}(A)$$

zerlegen, wobei nach Annahme des Falls (B) stets $V_2 \neq \{0\}$ gilt (während $V_1 = \{0\}$ möglich ist). Mit Hilfe einer Jordan-Basis kann nun ähnlich wie im Beweis von Satz 3.5 ein von A abhängiges (hermitesches) komplexes Skalarprodukt $\langle \cdot\,, \cdot \rangle_A$ mit zugehöriger Norm $\| \cdot \|_A$ auf $\mathbb{C}^N$ konstruiert werden, so dass die verschiedenen Hauptträume $\mathrm{H}_{\lambda_i}(A)$ von A zueinander orthogonal sind und $\|Ax - \lambda_i x\|_A \leq \frac{1}{4}\gamma \|x\|_A$ für alle $x \in \mathrm{H}_{\lambda_i}(A)$ gilt (und tatsächlich ist das Skalarprodukt schon bestimmt, indem man die Norm auf den Elementen einer Hauptraum-Basis wie im Beweis von Satz 3.5 vorgibt und diese für orthogonal erklärt). Durch $x \odot y := \frac{1}{2}\big[\langle x, y\rangle_A + \langle y, x\rangle_A\big] \in \mathbb{R}$ für $x, y \in \mathbb{C}^N$ erhält man dann ein reelles Skalarprodukt auf $\mathbb{C}^N$, das dieselbe Norm induziert und für das man zudem

$$x \odot \lambda x = (\mathrm{Re}\,\lambda)\|x\|_A^2 \qquad \text{für alle } x \in \mathbb{C}^N,\ \lambda \in \mathbb{C}$$

sowie (mit der Invarianz der Haupträume unter $x \mapsto Ax$)

$$\left.\begin{aligned} x_1 \odot Ax_1 &\leq \tfrac{1}{4}\gamma\|x_1\|_A^2 \\ x_2 \odot Ax_2 &\geq \tfrac{3}{4}\gamma\|x_2\|_A^2 \\ \|x_1{+}x_2\|_A^2 &= \|x_1\|_A^2 {+} \|x_2\|_A^2 \end{aligned}\right\} \quad \text{für alle } x_1 \in V_1, x_2 \in V_2 \tag{8.14}$$

nachrechnet. Da $A = \mathrm{D}F(0)$ gewählt wurde, lässt sich schließlich ein $\varepsilon > 0$ mit $\mathrm{B}_\varepsilon(0) \subset D$ fixieren, so dass

$$\|F(x) - Ax\|_A \leq \tfrac{1}{6}\gamma\|x\|_A \qquad \text{für alle } x \in \mathbb{R}^N \text{ mit } \|x\|_A < \varepsilon \tag{8.15}$$

gilt.

Als Nächstes wird nun eine Lösung u zu $u' = F(u)$ mit $\|u\|_A < \varepsilon$ auf $[0,\infty)$ betrachtet. Eine solche lässt sich eindeutig als $u = u_1+u_2$ mit V_1-wertigem u_1 und V_2-wertigem u_2 schreiben, und mit analoger Konvention bei $F = F_1+F_2$ gilt $u_2' = F_2(u_1+u_2)$ auf $[0,\infty)$. Mit dieser Gleichung, (8.14) und (8.15) ergibt sich

$$\begin{aligned}\bigl(\|u_2\|_A^2\bigr)' &= 2u_2 \odot u_2' \\ &= 2u_2 \odot Au_2 + 2u_2 \odot \bigl[F_2(u_1+u_2) - Au_2\bigr] \\ &\geq \tfrac{3}{2}\gamma\|u_2\|_A^2 - \tfrac{1}{3}\gamma\|u_2\|_A\sqrt{\|u_1\|_A^2+\|u_2\|_A^2} \\ &\geq \tfrac{7}{6}\gamma\|u_2\|_A^2 - \tfrac{1}{6}\gamma\|u_1\|_A^2\,,\end{aligned}$$

wobei im letzten Schritt mit Hilfe der Youngschen Ungleichung $\frac{1}{3}\gamma\|u_2\|_A\sqrt{\|u_1\|_A^2+\|u_2\|_A^2} \leq \frac{1}{6}\gamma\|u_2\|_A^2+\frac{1}{6}\gamma\bigl(\|u_1\|_A^2+\|u_2\|_A^2\bigr)$ abgeschätzt wurde. Eine weitgehend analoge Rechnung ergibt

$$\bigl(\|u_1\|_A^2\bigr)' \leq \tfrac{5}{6}\gamma\|u_1\|_A^2 + \tfrac{1}{6}\gamma\|u_2\|_A^2\,.$$

Zusammenfassend erhält man die Differentialungleichung

$$\bigl(\|u_2\|_A^2-\|u_1\|_A^2\bigr)' \geq \gamma\bigl(\|u_2\|_A^2-\|u_1\|_A^2\bigr) \qquad \text{auf } [0,\infty)$$

sowie im Fall $\|u_2(0)\|_A > \|u_1(0)\|_A$ die zugehörige Abschätzung

$$\|u_2(t)\|_A^2-\|u_1(t)\|_A^2 \geq \mathrm{e}^{\gamma t}\bigl(\|u_2(0)\|_A^2-\|u_1(0)\|_A^2\bigr) \qquad \text{für alle } t\in[0,\infty)\,, \tag{8.16}$$

die der zuvor gemachten Annahme $\|u\|_A < \varepsilon$ auf $[0,\infty)$ widerspricht.

Zum Abschluss der Argumentation bemerkt man, dass es Lösungen u zu $u' = F(u)$ gibt, deren Anfangsdatum $u(0)$ beliebig kleinen Betrag hat, aber auch $|u_2(0)| > |u_1(0)|$ erfüllt. Solche Lösungen u können, wie gerade argumentiert, aber nicht auf ganz $[0,\infty)$ mit $\|u\|_A < \varepsilon$ existieren. Dies zeigt die Instabilität der Null-Ruhelage und vervollständigt den Beweis. □

Kapitel 6

Die Hauptsätze der Theorie (Fortsetzung)

Wie schon früher in diesem Kapitel sei $\mathcal{X}$ stets Banach-Raum über $\mathbb{K}$ mit Norm $|\cdot| = \|\cdot\|_{\mathcal{X}}$.

6.7 Der Existenzsatz von Peano

Dieser Abschnitt beschäftigt sich mit einem allgemeinen Existenzsatz für AWPe, der den aus Abschnitt 6.1 bekannten Satz von Picard-Lindelöf teilweise verbessert.

Hauptsatz 6.14 (**Lokaler Existenzsatz von Peano**). *Sei* $\dim_{\mathbb{K}} \mathcal{X} < \infty$*, sei* $f \in \mathrm{C}^0(D, \mathcal{X})$ *eine stetige Strukturfunktion auf offenem* $D \subset \mathbb{R} \times \mathcal{X}^m$*, und sei* $(t_0, y_0, y_1, y_2, \ldots, y_{m-1}) \in D$*. Dann gibt es ein* $\varepsilon > 0$*, so dass das AWP*

$$u^{(m)} = f(\,\cdot\,, u^{[m-1]})\,, \qquad u^{[m-1]}(t_0) = (y_0, y_1, y_2, \ldots, y_{m-1})$$

auf $[t_0{-}\varepsilon, t_0{+}\varepsilon]$ *mindestens eine Lösung besitzt.*

Bemerkung. Der Vorteil dieses Satzes gegenüber dem Satz von Picard-Lindelöf liegt darin, dass keine pLB als Voraussetzung benötigt wird. Der Nachteil besteht neben der Einschränkung auf endlich-dimensionale $\mathcal{X}$ vor allem **darin, dass man nur Existenz, aber keine Eindeutigkeit erhält.** In Anbetracht des Gegenbeispiels aus Abschnitt 6.1 kann man Eindeutigkeit aber auch nicht erwarten, wenn f nur als stetig vorausgesetzt wird.

Der Beweis von Hauptsatz 6.14 macht entscheidenden Gebrauch von folgendem, aus der Analysis bekannten Kompaktheitssatz im Funktionenraum $\mathrm{C}^0(I, \mathcal{X})$.

Satz (von **Arzelà-Ascoli**). *Sei* I *kompaktes Intervall in* $\mathbb{R}$*, sei* $\dim_{\mathbb{K}} \mathcal{X} < \infty$*, und sei* $(u_k)_{k\in\mathbb{N}}$ *eine Folge von Funktionen in* $\mathrm{C}^0(I, \mathcal{X})$*. Ist* $(u_k(t_0))_{k\in\mathbb{N}}$ *für ein* $t_0 \in I$ *beschränkte Folge in* $\mathcal{X}$ *und sind die* u_k *auf* I *gleichgradig stetig*[8]*, so konvergiert eine Teilfolge* $(u_{k_\ell})_{\ell\in\mathbb{N}}$ *gleichmäßig gegen eine Grenzfunktion* $u \in \mathrm{C}^0(I, \mathcal{X})$*.*

[8] Gleichgradige Stetigkeit der u_k auf I bedeutet, dass es zu jedem $x \in I$ und $\varepsilon > 0$ ein $\delta > 0$ mit folgender Eigenschaft gibt: Für alle $\widetilde{x} \in I$ mit $|\widetilde{x}{-}x| < \delta$ und alle $k \in \mathbb{N}$ gilt $|u_k(\widetilde{x}){-}u_k(x)| < \varepsilon$.
Insbesondere liegt gleichgradige Stetigkeit der u_k auf I vor, wenn alle u_k auf I eine Lipschitz-Bedingung mit derselben Konstante erfüllen.

Beweis von Hauptsatz 6.14. Ohne Einschränkung sei $m = 1$, und es seien positive Größen δ und r mit $[t_0{-}\delta, t_0{+}\delta] \times \overline{\mathrm{B}}_r(y_0) \subset D$ fixiert. Sei außerdem $M := 1{+}\max_{[t_0-\delta,t_0+\delta]\times\overline{\mathrm{B}}_r(y_0)} |f|$ und $\varepsilon := \min\{\delta, r/M\} > 0$. Der Beweis basiert nun auf der Konstruktion von stückweise affinen Näherungslösungen $u_k \in \mathrm{C}^0([t_0, t_0{+}\varepsilon])$ mit dem **Polygonzug-Verfahren von Euler und Cauchy**. Dazu zerlegt man für fixiertes $k \in \mathbb{N}$ das Intervall $(t_0, t_0{+}\varepsilon]$ durch Zwischenstellen $t_i := t_0 + \frac{i}{k}\varepsilon$ in die k Teilintervalle $(t_i, t_{i+1}]$, $i = 0, 1, 2, \ldots, k{-}1$ und definiert u_k mit $u_k(t_0) := y_0$ auf den Teilintervallen sukzessive durch

$$u_k(t) := u_k(t_i) + f(t_i, u_k(t_i))(t{-}t_i) \quad \text{für } t \in (t_i, t_{i+1}]\,.$$

Es folgt $u_k' \equiv f(t_i, u_k(t_i))$ auf (t_i, t_{i+1}), in Anbetracht der Wahl $\varepsilon = \min\{\delta, r/M\}$ bleibt $(t, u_k(t))$ für $t \in [t_0, t_0{+}\varepsilon]$ in $[t_0{-}\delta, t_0{+}\delta] \times \overline{\mathrm{B}}_r(y_0)$, und bis auf die Knickstellen t_i gilt $|u_k'| \le M$ auf $[t_0, t_0{+}\varepsilon]$ (denn $|u_k'| \le M$ gilt zunächst solange u_k in $\overline{\mathrm{B}}_r(y_0)$ bleibt, und damit kann u_k die Kugel $\overline{\mathrm{B}}_r(y_0)$ auf $[t_0, t_0{+}\varepsilon]$ überhaupt nicht verlassen). Insbesondere ist u_k auf $[t_0, t_0{+}\varepsilon]$ Lipschitz-stetig mit Konstante M. Für $t \in (t_i, t_{i+1})$ gelten nun $|t{-}t_i| \le \frac{\varepsilon}{k}$ und $|u_k(t){-}u_k(t_i)| \le \frac{M\varepsilon}{k}$, und mit der gleichmäßigen Stetigkeit von f auf dem Kompaktum $[t_0{-}\delta, t_0{+}\delta] \times \overline{\mathrm{B}}_r(y_0)$ ergibt sich daraus die näherungsweise Lösungseigenschaft

$$u_k'(t) = f(t_i, u_k(t_i)) = f(t, u_k(t)) + \mathrm{o}(1) \quad (6.20)$$

bei $k \to \infty$, wobei die resultierende Gleichung für alle $t \in [t_0, t_0{+}\varepsilon]$ bis auf die Knickstellen t_i mit gleichmäßig kontrolliertem o(1)-Störterm gültig bleibt. Wegen $u_k(t_0) = y_0$ und der gleichmäßigen Lipschitz-Stetigkeit der u_k konvergiert nach Arzelà-Ascoli eine Teilfolge $(u_{k_\ell})_{\ell\in\mathbb{N}}$ gleichmäßig auf $[t_0, t_0{+}\varepsilon]$ gegen ein $u \in \mathrm{C}^0([t_0, t_0{+}\varepsilon], \mathcal{X})$ mit $u(t_0) = y_0$. Mit dieser gleichmäßigen Konvergenz, dem HDI, (6.20) und erneut der gleichmäßigen Stetigkeit von f auf $[t_0{-}\delta, t_0{+}\delta] \times \overline{\mathrm{B}}_r(y_0)$ folgt

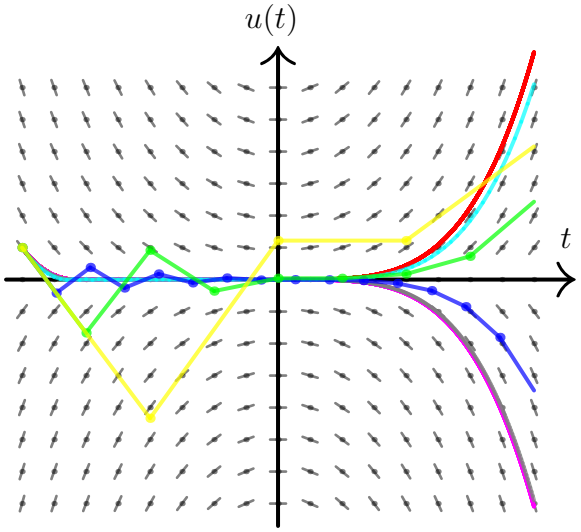


Abb. 15: Steigungsfeld und Näherungslösungen u_k mit $k = 4, 8, 15, 80, 400, 8000, 25000$ zum nicht-eindeutig lösbaren, skalaren AWP $u' = \frac{1}{6} t \operatorname{sign}(u)\sqrt{|u|}$, $u(-8) = 1$ auf $[-8, 8]$.

$$u(t) = \lim_{\ell\to\infty} u_{k_\ell}(t) = y_0 + \lim_{\ell\to\infty} \int_{t_0}^{t} u_{k_\ell}'(s)\,\mathrm{d}s = y_0 + \lim_{\ell\to\infty} \int_{t_0}^{t} f(s, u_{k_\ell}(s))\,\mathrm{d}s = y_0 + \int_{t_0}^{t} f(s, u(s))\,\mathrm{d}s$$

für alle $t \in [t_0, t_0{+}\varepsilon]$. Ableiten auf beiden Seiten der letzten Gleichung gibt $u \in \mathrm{C}^1([t_0, t_0{+}\varepsilon], \mathcal{X})$ und die erstrebte Lösungseigenschaft $u' = f(\,\cdot\,, u)$ auf $[t_0, t_0{+}\varepsilon]$. Analog beweist man die Existenz einer Lösung auf $[t_0{-}\varepsilon, t_0]$. □

Literaturverzeichnis

Die vorliegende Ausarbeitung basiert zu großen Teilen auf Vorlesungsskripten von F. Duzaar, A. Knauf, R. Lauterbach und K. Steffen. Daneben wurden auch einige der folgenden Bücher über dynamische Systeme und GDGen verwendet:

[1] B. Aulbach, *Gewöhnliche Differentialgleichungen*. Spektrum, 1997.

[2] H. Amann, *Gewöhnliche Differentialgleichungen*. De Gruyter, 1995.

[3] V.I. Arnold, *Gewöhnliche Differentialgleichungen*. Springer, 2001.

[4] E.A. Coddington, N. Levinson, *Theory of Ordinary Differential Equations*. McGraw-Hill, 1955.

[5] M. Denker, *Einführung in die Analysis dynamischer Systeme*. Springer, 2005.

[6] W. Forst, *Gewöhnliche Differentialgleichungen*. Springer, 2005.

[7] J.K. Hale, *Ordinary Differential Equations*. Wiley, 1969.

[8] B. Hasselblatt, A. Katok, *A First Course in Dynamics*. Cambridge University Press, 2003.

[9] H. Heuser, *Gewöhnliche Differentialgleichungen*. Vieweg+Teubner, 2009.

[10] A. Katok, B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*. Cambridge University Press, 1995.

[11] W. Walter, *Gewöhnliche Differentialgleichungen*. Springer, 2000.