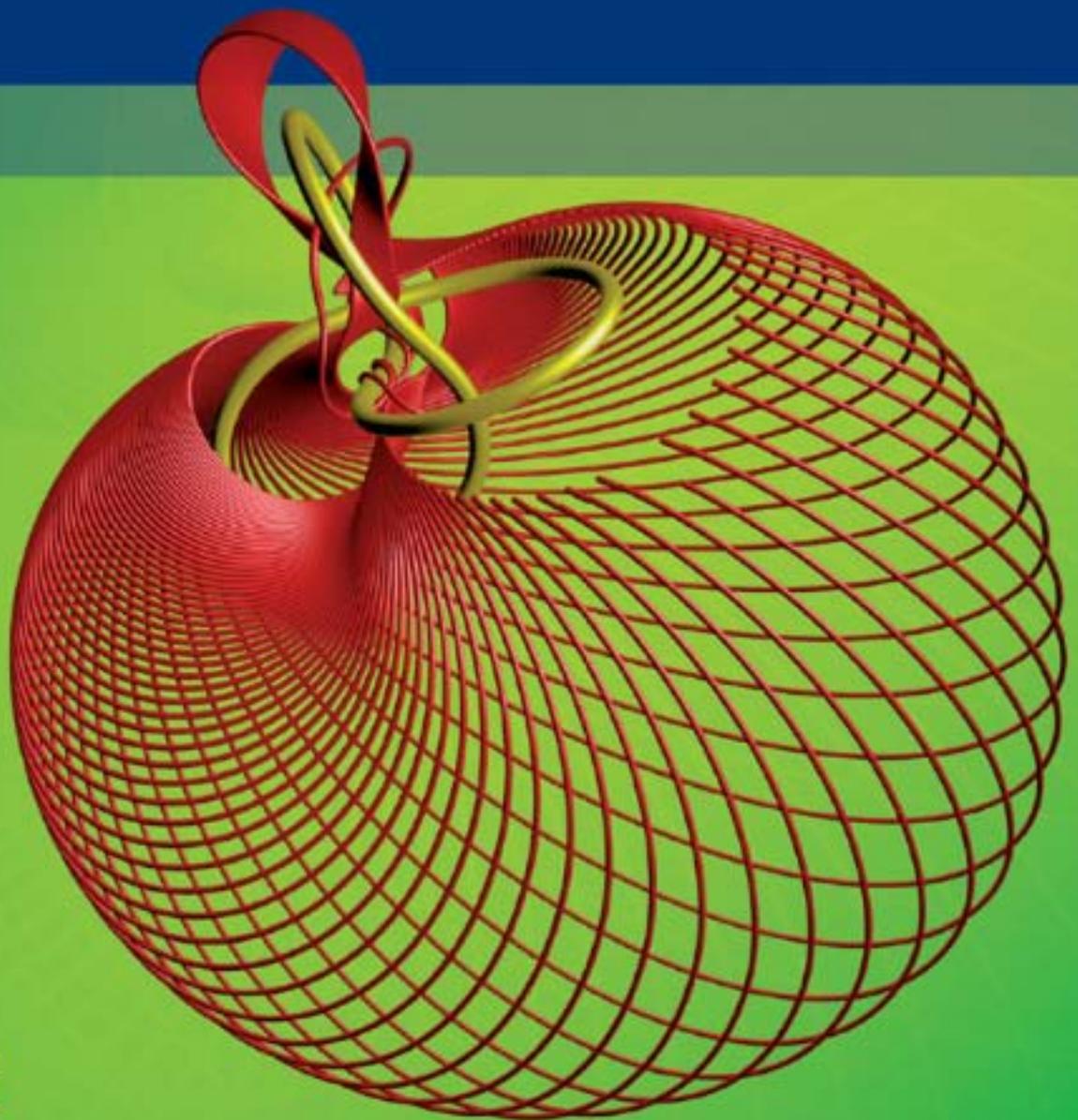


Bernd Marx Werner Vogt

Dynamische Systeme – Theorie und Numerik



Spektrum
AKADEMISCHER VERLAG

Dynamische Systeme

Bernd Marx Werner Vogt

Dynamische Systeme

Theorie und Numerik

Autoren:

Univ.-Prof. Dr. rer. nat. habil. Bernd Marx
Technische Universität Ilmenau
Institut für Mathematik
e-mail: bernd.marx@tu-ilmenau.de

Priv.-Doz. Dr. rer. nat. habil. Werner Vogt
Technische Universität Ilmenau
Institut für Mathematik
e-mail: werner.vogt@tu-ilmenau.de

Wichtiger Hinweis für den Benutzer

Der Verlag und die Autoren haben alle Sorgfalt walten lassen, um vollständige und akkurate Informationen in diesem Buch zu publizieren. Der Verlag übernimmt weder Garantie noch die juristische Verantwortung oder irgendeine Haftung für die Nutzung dieser Informationen, für deren Wirtschaftlichkeit oder fehlerfreie Funktion für einen bestimmten Zweck. Der Verlag übernimmt keine Gewähr dafür, dass die beschriebenen Verfahren, Programme usw. frei von Schutzrechten Dritter sind. Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Buch berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften. Der Verlag hat sich bemüht, sämtliche Rechteinhaber von Abbildungen zu ermitteln. Sollte dem Verlag gegenüber dennoch der Nachweis der Rechtsinhaberschaft geführt werden, wird das branchenübliche Honorar gezahlt.

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer ist ein Unternehmen von Springer Science+Business Media
springer.de

© Spektrum Akademischer Verlag Heidelberg 2011
Spektrum Akademischer Verlag ist ein Imprint von Springer

11 12 13 14 15 5 4 3 2 1

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Planung und Lektorat: Dr. Andreas Rüdinger, Sabine Bartels
Redaktion: Dr. Sonja Bernhart
Herstellung und Satz: Crest Premedia Solutions (P) Ltd, Pune, Maharashtra, India
Umschlaggestaltung: SpieszDesign, Neu-Ulm
Titelfotografie: © Jos Leys

ISBN 978-3-8274-2447-1

Vorwort

Mit diesem Buch wollen die Autoren den Studierenden in den Masterstudiengängen für Mathematik, Natur- und Ingenieurwissenschaften ein solides mathematisches Wissen zur Theorie und Numerik dynamischer Systeme vermitteln. Viele Prozesse in Physik, Chemie, Biologie, Medizin und in den Ingenieur- und Wirtschaftswissenschaften werden durch parameterabhängige nichtlineare Differenzialgleichungen beschrieben. Deshalb sind Differenzialgleichungen und dynamische Systeme als zentrale Gebiete der Mathematik auch weiterhin in ständiger Entwicklung begriffen. Die Idee zu diesem Buch entstand aus unserer langjährigen Vorlesungstätigkeit für Studierende der Mathematik, der Technischen Physik, des Maschinenbaus, der Elektrotechnik und der Informatik, sowie auf der Basis unserer wissenschaftlichen Kooperation mit Ingenieuren an der TU Ilmenau.

An die aktuelle Mathematik in den Natur- und Ingenieurwissenschaften werden Fragen von großer Komplexität gestellt, die nur durch das Ineinandergreifen sehr unterschiedlicher Gebiete erfolgreich in Angriff genommen werden können. Dies führt in wachsendem Maße zum Zusammenführen von reiner und angewandter Mathematik. Diese Entwicklungstendenzen werden sich auch auf Grund des Auftretens neuer Ideen und Methoden in Zukunft fortsetzen. Wir halten es deshalb für unabdingbar, wesentliche Begriffe und Denkweisen der Funktionalanalysis einerseits und der Numerischen Analysis andererseits einzuführen, um den Einsatz moderner mathematischer Verfahren des Wissenschaftlichen Rechnens zu unterstützen.

In der Stoffauswahl und -anordnung haben wir auf ausführliche Motivation und Erläuterung der Grundideen, auf leichte Fasslichkeit, Anschaulichkeit und Übersichtlichkeit Wert gelegt. Der mit der Thematik bereits vertraute Leser möge Verständnis aufbringen, wenn an geeigneter Stelle ihm bekannte Grundbegriffe und -aussagen definiert werden. Damit wollen wir erreichen, dass das mühevollere Zusammensuchen von Begriffen und Aussagen weitgehend entfällt und für alle Leser des Buches gleiche Bedingungen geboten werden, um den Einstieg in das umfangreiche Gebiet zu erleichtern. Wo es die Stoffauswahl zulässt, wird auch ein Praxisbezug hergestellt. Wir werden allerdings nicht in jedem Kapitel erläutern, wofür ein mathematischer Begriff oder Sachverhalt „nützlich“ ist, denn erst das Zusammenspiel der verschiedenen Begriffsbildungen und der daraus resultierenden Sätze führt zu sinnvollen Anwendungen. Von der Behandlung chaotischer Lösungen und Attraktoren fraktaler Dimension mussten wir in Anbetracht des Buchumfanges Abstand nehmen, zumal sich das erforderliche mathematische Instrumentarium teilweise stark vom übrigen Buch unterscheidet und deshalb gesondert vorgestellt werden soll.

Das Buch ist so aufgebaut, dass zunächst im *ersten Kapitel* einige Grundtatsachen aus dem Gebiet der Funktionalanalysis aufgeschrieben sind. Die gesamte moderne Analysis

basiert heute auf der Funktionalanalysis. Sie stellt als elegante mathematische Theorie allgemeine Hilfsmittel bereit, um mathematische Aufgabenstellungen, wie gewöhnliche und partielle Differenzialgleichungen, Integralgleichungen und Extremalprobleme in übersichtlicher und einheitlicher Weise zu lösen. Gemeinsame Merkmale, die bei der Lösung dieser Aufgabenstellungen auftreten, werden auf allgemein gültige Prinzipien zurückgeführt. Darüber hinaus kann man die Struktur und Konvergenz von Näherungsverfahren in einheitlicher Weise untersuchen. Der Mehraufwand zur Einarbeitung in dieses Kapitel wird dadurch abgegolten, dass scheinbar sehr unterschiedliche Aufgabenstellungen, wie sie in der Praxis auftreten, mit Hilfe der gleichen abstrakten mathematischen Methode gelöst werden können.

Das *zweite Kapitel* beschäftigt sich weniger mit Methoden der Berechnung der Lösungen von Differenzialgleichungen, sondern mehr mit strukturellen und qualitativen Aussagen. Wir haben dieser Sichtweise den Vorzug gegenüber der traditionellen, an Lösungstechniken und der linearen Theorie ausgerichteten Betrachtungsweise gegeben. Dieser Aspekt erscheint wichtig, um numerische und grafische Lösungsapproximationen beurteilen zu können.

Im *Kapitel 3* wird der Leser an den Begriff der Lösungsbifurkation bei Differenzialgleichungen behutsam herangeführt. Dieser Abschnitt beschreibt die ersten Schritte in Richtung einer zeitgemäßen Fortführung der elementaren Theorie autonomer Systeme. Es werden die drei Grundtypen Sattel-Knoten-Bifurkation, transkritische Bifurkation und die Pitchfork-Bifurkation behandelt. Den Abschluss bilden die Resultate über die Verzweigung von Ruhelagen und geschlossenen Orbits (Hopf-Bifurkation).

Das *Kapitel 4* über analytische Bifurkationstheorie kann als Brücke zwischen der elementaren Theorie gewöhnlicher Differenzialgleichungen und den Anfängen der so genannten nichtlinearen Dynamik verstanden werden. Neben verschiedenen Themen der klassischen statischen Bifurkationstheorie haben wir auf unterschiedlichem mathematischem Niveau einige grundlegende Probleme der modernen dynamischen Verzweigungstheorie vorgestellt, die in den üblichen Büchern über Differenzialgleichungen in dieser Form nicht behandelt werden. Die Gewinnung der Aussagen in diesem Abschnitt erfolgt wieder vorwiegend auf der Basis der in Kapitel 1 dargelegten Funktionalanalysis.

Die Numerik nimmt – beginnend mit *Kapitel 5* – bei der Entwicklung und Analyse leistungsfähiger Näherungsverfahren einen relativ großen Umfang ein. Wegen der Nichtlinearität der mathematischen Modelle können dynamische Systeme aus der Praxis ohne leistungsfähige numerische Verfahren zukünftig nicht erfolgreich analysiert werden. Deshalb sind Naturwissenschaftler und Ingenieure auf effiziente Algorithmen und Computer bei der Erforschung der Modelle ihres Fachgebietes angewiesen. Für umfangreiche Computersimulationen realer Prozesse werden schnelle und zuverlässige numerische Verfahren

benötigt, deren grundlegendes Verständnis und Einordnung in zunehmendem Maße Bestandteil der mathematischen Ausbildung werden sollte. Das Wissenschaftliche Rechnen (Scientific Computing) in Mathematik, Natur- und Ingenieurwissenschaften nutzt effiziente analytische und numerische Verfahren besonders für nichtlineare und großdimensionale Probleme. Deshalb werden in den *Kapiteln 5 und 6* aktuelle Näherungsverfahren zur Lösungsverfolgung von Gleichgewichtslagen und periodischen Schwingungen einschließlich ihrer programmierbaren Algorithmen vorgestellt und beispielhaft durch die Verzweigungsanalyse eines 9-dimensionalen Modells aus der Ingenieurspraxis illustriert.

Das *Kapitel 7* ist quasi-periodischen Bewegungen mit mehreren Basisfrequenzen gewidmet. Diese multifrequentiellen Schwingungen treten insbesondere bei gekoppelten DGL-Systemen, z.B. in elektrischen Dreiphasen-Systemen der Energietechnik auf und bilden ein kompliziertes und sensibles Studienobjekt. Invariante Tori als Träger dieser quasi-periodischen Bewegungen können jedoch durch die vorgestellten aktuellen numerischen Zugänge (Finite-Differenzen-Verfahren, Fourier-Galerkin-Verfahren) als Lösung der sie beschreibenden partiellen Differenzialgleichungen approximiert und untersucht werden. Konkrete Anwendungen demonstrieren den erfolgreichen Einsatz dieser Verfahren.

Die Darstellung des Stoffes ist weit davon entfernt, eine bloße Sammlung von „Kochrezepten“ zu sein. Gerade die mathematisch strenge Herleitung zentraler Ideen und durchgehend präzise Formulierungen und Darstellungen fördern entscheidend Verständnis und Durchblick und geben so erst die gewünschte Sicherheit bei den Anwendungen. Falls uns ein vollständiger Beweis nicht ratsam erschien, wird in der Regel ein Literaturverweis angegeben. Unseren Beitrag sehen wir darin, durch Klarheit, Transparenz und Konzentration auf das Wesentliche, den Leser auch an die moderneren Inhalte der Mathematik heranzuführen. Es ist uns ein Anliegen zu demonstrieren, dass die Mathematik nicht nur Eleganz und innere Schönheit besitzt, sondern auch effektive Methoden zur Lösung konkreter Fragestellungen zur Verfügung stellt. Wir wollen uns bemühen, mit diesem Buch ein solides Fundament zu legen, das dem Leser später auch bei weitergehender Beschäftigung mit Mathematik und deren Anwendung zu gute kommt.

Schließlich möchten wir die Aufmerksamkeit auch auf die Übungsaufgaben richten, die am Schluß eines jeden Kapitels zu finden sind. Die Beschäftigung mit diesen Aufgaben ist eine wichtige Voraussetzung für ein vertiefendes Verständnis des Stoffes. Die Darstellung der numerischen Verfahren als Algorithmen im Pseudocode bietet Musterlösungen für eigene Verfahrensentwicklungen der Leser und soll zu einer kritischen Bewertung der Näherungsverfahren anleiten. Sämtliche mit dem Computer gerechneten Anwendungsbeispiele benutzen automatisch die im wissenschaftlichen Bereich international übliche Notation von Gleitpunktzahlen mittels des Dezimalpunktes. Wir haben uns entschlossen, die Dezimalpunkt-Darstellung durchgehend zu verwenden, zumal damit die Lesbarkeit

von Zahlenfolgen und -vektoren verbessert wird. Für sehr häufig wiederkehrende mathematische Begriffe wie „Differenzialgleichung“ werden allgemein übliche Abkürzungen wie „DGL“ an entsprechender Stelle eingeführt und undekliniert im gesamten Text benutzt.

Zwar ist das Buch primär an Studenten im Master-Studium der Mathematik, der Natur- und Ingenieurwissenschaften und Informatik gerichtet, kann aber auch von Doktoranden mit Gewinn genutzt werden. Auch Forschungsingenieuren und Anwendern dürfte das Buch neue Elemente der Mathematik bieten. Die vorgestellten Anwendungen wurden größtenteils mit Computerprogrammen gerechnet, die von Studierenden und Doktoranden gemeinsam mit den Autoren an der TU Ilmenau entwickelt wurden. Deshalb gilt unser Dank Frau Dr. K. Bernet und den Herren Dr. F. Schilder, Dr. S. Schreiber, Dr. D. Peterseim und Dipl.-Math. M. Ernst, die mit ihrer zeitaufwändigen Leistung zum Gelingen des Buches beigetragen haben.

Als jederzeit kompetente und freundlich nachsichtige Ansprechpartner standen uns seitens des Spektrum-Verlages Herr Dr. Andreas Rüdinger und Frau Sabine Bartels stets hilfreich zur Seite. Die Autoren danken ihnen insbesondere für die wertvolle Unterstützung während der gesamten Fertigstellung des Buches.

Ilmenau, im September 2010

Bernd Marx
Werner Vogt

Inhaltsverzeichnis

1	Funktionalanalytische Grundlagen	1
1.1	Einführung	2
1.1.1	Metrischer Raum	3
1.1.2	Normierter Raum und Banach-Raum	7
1.1.3	Skalarproduktraum und Hilbert-Raum	9
1.1.4	Orthogonalreihen in Hilbert-Räumen	13
1.2	Lineare Operatoren	17
1.2.1	Bezeichnungen und Begriffe	18
1.2.2	Lineare stetige Operatoren	23
1.2.3	Adjungierte Operatoren	29
1.2.4	Direkte Summe und Projektoren	33
1.2.5	Fredholm-Operatoren	36
1.2.6	Kompakte Operatoren	41
1.3	Fréchet- und Gâteaux-Ableitung	54
1.4	Nemytski-Operator	58
1.5	Implizites Funktionentheorem	59
1.6	Aufgaben	61
2	Gewöhnliche Differenzialgleichungen (DGL)	67
2.1	Einführende Beispiele	68
2.2	Geometrische Interpretation einer DGL	72
2.3	Existenz- und Eindeutigkeitsätze	73
2.4	Lineare DGL-Systeme 1. Ordnung	83
2.5	Lineare DGL-Systeme 1. Ordnung mit konstanten Koeffizienten	89
2.6	Autonome Systeme	98
2.6.1	Allgemeine Aussagen	98
2.6.2	Kritische Punkte	104
2.7	Hilfsmittel zur Konstruktion von Phasenportraits	107
2.7.1	Stabilität linearer Systeme	107
2.7.2	Linearisierung nichtlinearer DGL-Systeme	111
2.7.3	Das Hartman-Grobman-Theorem	112
2.8	Aufgaben	122
3	Bifurkation bei gewöhnlichen DGL	127
3.1	Strukturelle Stabilität	128
3.2	Einige typische Bifurkationen	133
3.2.1	Sattel-Knoten Bifurkation	134
3.2.2	Transkritische Bifurkation	136
3.2.3	Die Pitchfork-Bifurkation	138
3.2.4	Die Hopf-Bifurkation	140

3.2.5	Zusammenfassung der Bifurkationstypen	146
3.3	Aufgaben	148
4	Analytische Bifurkationstheorie	151
4.1	Bifurkationsgleichung von Ljapunov-Schmidt	152
4.1.1	Herleitung einer Bifurkationsgleichung	153
4.1.2	Lösen der Bifurkationsgleichung	164
4.2	Anwendungsbeispiele	177
4.3	Die Hopf-Bifurkation	211
4.3.1	Abstraktes Hopf-Bifurkationstheorem	211
4.3.2	Nichtlineare Schwingungen in autonomen Systemen	214
4.4	Aufgaben	224
5	Numerik der Gleichgewichtslösungen	227
5.1	Berechnung von Gleichgewichtslösungen	228
5.1.1	Newton-Verfahren	229
5.1.2	Vereinfachte Newton-Verfahren	235
5.1.3	Newton-Verfahren mit Differenzenquotienten	238
5.1.4	Gedämpfte Newton-Verfahren und globale Konvergenz	240
5.2	Parametrisierung von Lösungskurven und Fortsetzungsmethoden	243
5.2.1	Natürliche Parametrisierung	244
5.2.2	Bogenlängen-Parametrisierung	253
5.2.3	Gauß-Newton-Fortsetzung	259
5.2.4	Fortsetzung mehrdimensionaler Mannigfaltigkeiten	264
5.2.5	Fortsetzung mit Ableitungsprädiktoren	268
5.3	Stabilitäts- und Bifurkationsanalyse	273
5.3.1	Stabilitätsanalyse	273
5.3.2	Detektierung lokaler Bifurkationen	275
5.4	Aufgaben	283
6	Numerik periodischer Lösungen	287
6.1	Periodisch erregte Systeme	289
6.1.1	Einfaches Schießverfahren	290
6.1.2	Stabilitätsanalyse	296
6.1.3	Mehrfach-Schießverfahren	301
6.2	Autonome Systeme	306
6.2.1	Einfaches Schießverfahren	306
6.2.2	Phasenbedingungen	309
6.2.3	Technische Realisierung des Schießverfahrens	312
6.2.4	Stabilitätsanalyse	315
6.3	Die Poincaré-Abbildung	318
6.3.1	Definition und Eigenschaften	318
6.3.2	Ableitung der Poincaré-Abbildung	322

6.3.3	Numerische Approximation der Poincaré-Abbildung	324
6.4	Lösungsfortsetzung und Bifurkationsanalyse	327
6.4.1	Numerische Fortsetzung und Stabilitätsanalyse	328
6.4.2	Bifurkationen periodischer Orbits	333
6.4.3	Detektierung lokaler Bifurkationen	343
6.4.4	Anwendung bei nichtlinearen energetischen Systemen	348
6.5	Aufgaben	358
7	Quasi-periodische Lösungen und invariante Tori	363
7.1	Quasi-periodische Funktionen	364
7.1.1	Torusfunktionen	364
7.1.2	Quasi-periodische Bewegungen	371
7.1.3	Transformation in Toruskoordinaten	374
7.2	Parametrisierung invarianter Tori	380
7.2.1	Eine Invarianzgleichung für invariante Tori	381
7.2.2	Basisdiskretisierungen von 2-Tori	386
7.2.3	Ein parametrisch erregtes elektrisches Netzwerk	394
7.3	Quasi-periodische invariante Tori	399
7.3.1	Eine spezielle Invarianzgleichung für p -Tori	399
7.3.2	Semidiskretisierung von 2-Tori	405
7.3.3	Volldiskretisierung mit Fourier-Galerkin-Methode	412
7.4	Aufgaben	420
	Literaturverzeichnis	423
	Symbolverzeichnis	429
	Index	431

1 Funktionalanalytische Grundlagen

Übersicht

1.1	Einführung	2
1.2	Lineare Operatoren	17
1.3	Fréchet- und Gâteaux-Ableitung	54
1.4	Nemytski-Operator	58
1.5	Implizites Funktionentheorem	59
1.6	Aufgaben	61

Alle Beziehungen zwischen Zahlen, Funktionen und Operatoren werden erst dann durchsichtig, verallgemeinerungsfähig und wirklich fruchtbar, wenn sie von ihren besonderen Objekten losgelöst und auf allgemeine begriffliche Zusammenhänge zurückgeführt werden. Diesem Anspruch versucht die Funktionalanalysis gerecht zu werden.

Wir beschäftigen uns u. a. mit Begriffsbildungen wie Banach-Raum, Hilbert-Raum und linearen, stetigen, kompakten und differenzierbaren Operatoren – definiert auf diesen Räumen – und demonstrieren, wie dadurch die Theorie der linearen und nichtlinearen Gleichungen erheblich an Klarheit und Übersichtlichkeit gewinnt.

Die behandelten Begriffe haben sich in einem Zeitraum von rund 100 Jahren herauskristallisiert. Die meisten dieser neuen Ergebnisse sind in Räumen gewonnen worden, die neben der metrischen Eigenschaft eine Vektorraumstruktur besitzen und man hat erkannt, dass sich der Abstand zweier Elemente x und y aus der Differenz $x - y$ ableitet. Dies hat zu dem wichtigen Begriff des *normierten Raumes* geführt. Er verknüpft den Abstandsbegriff (*Metrik*) und den linearen Aspekt (*Vektorraum*). Als Spezialfall ergibt sich der *Prähilbert-Raum*. Es handelt sich dabei grob gesprochen um eine Gesamtheit von (beliebigen) Objekten (z. B. Zahlen, Funktionen, Systeme von Zahlen und Funktionen), zwischen denen Relationen festgesetzt werden, die diesen oder jenen räumlichen Beziehungen entsprechen. Diese beiden Raumbegriffe haben sich für die weitere Entwicklung der Funktionalanalysis als außerordentlich fruchtbar erwiesen. Charakteristisch für einen Prähilbert-Raum ist die zusätzliche Einbeziehung der Eigenschaft der Orthogonalität zweier Elemente vermöge eines *Skalarprodukts*, wie es von einem endlichdimensionalen

euklidischen Vektorraum bekannt ist. Dadurch kann man die unendlichdimensionalen Prähilbert-Räume als eine mögliche Verallgemeinerung des endlichdimensionalen euklidischen Raumes ansehen.

*David Hilbert*¹ ist der Vater der modernen axiomatischen Methode, wonach eine mathematische Disziplin auf wenigen Grundbegriffen aufgebaut ist, die nicht weiter erklärt werden. Diese Begriffe stützen sich auf einige Grundtatsachen, die als wahr angenommen und Axiome genannt werden.

Ein vertieftes Verständnis für die Bedeutung und die Notwendigkeit dieser Begriffe kann nur aus einer Vertrautheit mit der Funktionalanalysis erwachsen. Diesem Anspruch kann dieses Kapitel nicht Rechnung tragen. Das mag auch die Lückenhaftigkeit unserer Darstellung entschuldigen. Auch auf Beweise muss vielfach verzichtet werden. Der Leser ist deshalb auf weiterführende Literatur angewiesen. Die in diesem Kapitel gelegten Grundlagen reichen jedoch für das Verständnis der weiteren Kapitel aus.

1.1 Einführung

Der Begriff des metrischen Raumes wird nur deshalb kurz aufgegriffen, da er das Wesen des Abstandes herausarbeitet. Der hier dargelegte Begriff des normierten Raumes ist von *Stefan Banach*² um 1920 eingeführt worden. Dass unter den normierten Räumen jene von besonderer Bedeutung sind, deren Norm aus einem Skalarprodukt (auch inneres Produkt genannt) entspringt, wurde schon durch David Hilberts Untersuchungen über unendlichdimensionale quadratische Formen deutlich.

Mathematische Notationen

Die Begriffe *Raum*, *Glied*, *Variable*, *Element* und *Menge* sind für ein organisiertes Denken in der Mathematik grundlegend. Dem Leser sollte die tragende Rolle des Mengenbegriffes klar sein. Ein weiterer Vorteil beim Lesen des Buches ist die Vertrautheit mit den gebräuchlichen mathematischen Schriftzeichen, mit deren Hilfe über Mengen, Elemente von Mengen oder über Relationen zwischen Elementen und Mengen gesprochen wird.

Mengen werden mit Großbuchstaben, Elemente eines (metrischen, normierten, ...) Raumes werden beispielsweise mit x, y, a, b, \dots und Folgen in solchen Räumen mit $(x_k), \dots$, manchmal auch mit $(x_k)_{k \in \mathbb{N}}, \dots$ bezeichnet. Im \mathbb{R}^n werden, um Verwechslungen vorzubeugen, für Folgen gelegentlich hochgestellte Indizes verwendet, also (x_k) mit $(x_k) = (x_1^k, \dots, x_n^k)$.

¹David Hilbert (1862–1943) wirkte in Göttingen. Er formulierte 1900 auf dem internationalen Mathematiker-Kongress in Paris die berühmten 23 Probleme, die die Mathematik bis zum heutigen Tag beeinflussen.

²Stefan Banach (1892–1945), polnischer Mathematiker, einer der Begründer der Funktionalanalysis.

Wir beginnen mit einer Zusammenstellung von Definitionen und (mathematischen) Redeweisen.

1.1.1 Metrischer Raum

Hier werden wichtige Dinge über metrische Räume genannt, die in den nächsten Abschnitten benötigt werden.

Definition 1.1 (Metrischer Raum)

Es sei X eine Menge, deren Elemente im folgenden auch Punkte genannt werden können. Ferner sei je zwei Punkten $x, y \in X$ eine reelle Zahl $d(x, y)$ zugeordnet, so dass für beliebige $x, y, z \in X$ gilt:

$$(M1) \quad d(x, y) \geq 0 \quad \text{und} \quad d(x, y) = 0 \Leftrightarrow x = y \quad (\text{Definitheit})$$

$$(M2) \quad d(x, y) = d(y, x) \quad (\text{Symmetrie})$$

$$(M3) \quad d(x, y) \leq d(x, z) + d(z, y) \quad (\text{Dreiecksungleichung})$$

Eine solche Funktion $d : X \times X \rightarrow \mathbb{R}$ wird als eine Metrik auf X , die Zahl $d(x, y)$ als Abstand zwischen den Punkten x und y und die mit dieser Metrik versehene Menge X als *metrischer Raum* (X, d) bezeichnet.

Offensichtlich stimmen die aufgezählten Bedingungen völlig mit unseren Vorstellungen über den Abstandsbegriff überein: Der Abstand ist stets nichtnegativ; nur von zusammenfallenden Punkten ist der Abstand Null und umgekehrt; beide Punkte sind bei der Abstandsbestimmung gleichberechtigt und schließlich bringt das Dreiecksaxiom grob gesprochen zum Ausdruck, dass die Strecke von x nach y nicht länger sein kann als der Streckenzug von x nach y über einen dritten Punkt z .

Beispiel 1.2

1. Im n -dimensionalen euklidischen Vektorraum \mathbb{R}^n (reell) bzw. \mathbb{C}^n (komplex) definiert man die Entfernung zweier Vektoren durch die Formel

$$d(x, y) := \|x - y\|_2 = \left(\sum_{i=1}^n |x_i - y_i|^2 \right)^{1/2}.$$

(Hier bezeichnet $\|\cdot\|_2$ den euklidischen Abstand.) Für $n = 1$ sind die Verhältnisse einfach: $\mathbb{R}^1 \equiv \mathbb{R}$ bzw. $\mathbb{C}^1 \equiv \mathbb{C}$ bezeichnen nichts anderes als die Gesamtheit aller reellen bzw. komplexen Zahlen mit dem Abstand $d(x, y) = |x - y|$.

2. Unter einem Vektor mit unendlich vielen Koordinaten versteht man eine unendliche Folge reeller Zahlen $(a_1, a_2, \dots, a_n, \dots)$. Wie üblich bezeichnet man den Vektor mit einem Buchstaben und schreibt: $a = (a_1, a_2, \dots, a_n, \dots)$. Es ist auch gerechtfertigt, dass man

die Glieder a_i , $i = 1, 2, \dots$, der Folge Koordinaten des Vektors a nennen kann. Man wählt aus dieser Menge von Vektoren, wovon jeder einzelne aus einer unendlichen Folge reeller Zahlen besteht, eine Teilmenge aus, für deren Elemente

$$\sum_{i=1}^{\infty} |a_i|^p < +\infty, \quad p \in \mathbb{R}, \quad p \geq 1$$

gilt. Die so entstehende Menge wird mit

$$l^p := \left\{ a = (a_1, a_2, \dots, a_n, \dots) \mid \sum_{i=1}^{\infty} |a_i|^p < +\infty, \quad p \geq 1 \right\}$$

bezeichnet. Führt man in dieser Menge die üblichen Operationen „+“ (bezüglich der Addition der Elemente) und „·“ (bezüglich der skalaren Multiplikation in \mathbb{R} oder \mathbb{C}) ein, dann ist l^p ein Vektorraum. Er lässt sich auch mit einer Metrik ausstatten. Für beliebige Elemente $a = (a_n)$ und $b = (b_n)$ setzt man

$$d(a, b) := \left(\sum_{i=1}^{\infty} |a_i - b_i|^p \right)^{1/p}, \quad p \geq 1$$

und weist mit Hilfe der Minkowskischen Ungleichung für unendliche Summen (siehe Walter (1995)) die Dreiecksungleichung (M3) nach. (M1) und (M2) sind trivial. ■

Wir beginnen damit, die wichtigsten metrischen Grundbegriffe einzuführen. Für das Verständnis lasse sich der Leser von dem folgenden Übertragungsprinzip leiten: Alle bisherigen, auf dem Abstand zweier Punkte in \mathbb{R} oder \mathbb{R}^n basierenden Begriffe werden für den metrischen Raum formuliert, wobei lediglich der Abstand $|x - y|$ durch $d(x, y)$ zu ersetzen ist.

Kugel, Sphäre, Umgebung

Die *offene Kugel* $K(a; r)$ mit dem Mittelpunkt $a \in X$ und dem Radius $r > 0$ ist die Menge aller Punkte $x \in X$ mit $d(x, a) < r$. Entsprechend ist die *abgeschlossene Kugel* $K[a; r]$ durch $d(x, a) \leq r$ und die *Sphäre* $S(a; r)$ (Kugeloberfläche) durch $d(x, a) = r$ definiert. Jede Menge $U = U(a) \subset X$, zu der es ein $\varepsilon > 0$ mit $K(a; \varepsilon) \subset U$ gibt, heißt eine *Umgebung* von a . Die Kugel $K(a; \varepsilon)$ ist eine spezielle Umgebung, die sogenannte ε -Umgebung von a .

Offene und abgeschlossene Mengen

Eine Teilmenge A eines metrischen Raumes (X, d) heißt *offen*, wenn sie nur aus inneren Punkten besteht, d. h. wenn es zu jedem Punkt $x \in A$ ein $\varepsilon > 0$ mit $K(x; \varepsilon) \subset A$ gibt. Die Menge A heißt *abgeschlossen*, wenn ihr Komplement $X \setminus A$ offen ist.

Achtung: Die Begriffe „offen“ und „abgeschlossen“ sind keine logischen Gegensätze. Eine Menge A kann sehr wohl weder offen als auch abgeschlossen (Beispiel: $A := \mathbb{Q}$ in $X := \mathbb{R}$) oder auch sowohl offen als auch abgeschlossenen sein (Beispiel: $A := \mathbb{R}$ in $X := \mathbb{R}$).

Beispiel 1.3

1. Jede offene Kugel ist eine offene Menge. Dies sieht man so: Es sei $x \in K(a; r)$ und $\varrho := r - d(a, x) > 0$. Dann gilt $K(x; \varrho) \subset K(a; r)$, was man mit Hilfe der Dreiecksungleichung bestätigt. Ähnlich zeigt man, dass jede abgeschlossene Kugel $K[a; r]$ abgeschlossen ist.

2. Die leere Menge \emptyset sowie der ganze Raum X sind sowohl offen als auch abgeschlossen.

3. Jede einpunktige Menge $A = \{a\}$ ist abgeschlossen. Ist $x \neq a$ und $r = d(x, a)$ so liegt $K(x; r)$ im Komplement $X \setminus A$. Also ist jedes $x \neq a$ innerer Punkt von $X \setminus A$, d. h. $X \setminus A$ ist offen, bzw. A ist abgeschlossen. ■

Folge, Konvergenz, Beschränktheit, Cauchy-Folge und Vollständigkeit

Zunächst erklären wir den Begriff der *Folge*. Sei (X, d) ein metrischer Raum. Die Abbildung

$$f : \mathbb{N} \rightarrow X, \quad n \mapsto f(n) \in X$$

heißt Folge, $x_n := f(n)$ heißt das n -te Glied der Folge. Als Symbol für diesen Sachverhalt verwenden wir

$$(x_n), \quad (x_n)_{n \in \mathbb{N}} \quad \text{oder} \quad (x_1, x_2, x_3, \dots).$$

Ist $X := \mathbb{K}$, dann heißen die Folgen Zahlenfolgen, genauer: (x_n) heißt *reelle* bzw. *komplexe Zahlenfolge*, falls $\mathbb{K} = \mathbb{R}$ bzw. $\mathbb{K} = \mathbb{C}$ gilt. Sie heißt *konvergent* mit dem Limes $a \in X$ (a heißt auch Grenzwert), falls die Abstände $d(x_n, a)$ gegen 0 streben für $n \rightarrow \infty$. Symbol:

$$a = \lim_{n \rightarrow \infty} x_n \quad \text{oder} \quad x_n \rightarrow a \quad \text{für} \quad n \rightarrow \infty.$$

Eine Folge heißt *beschränkt*, falls ihre Wertemenge beschränkt ist, d. h. falls es eine Kugel $K[x_0; r]$, $x_0 \in X$, $r > 0$ mit $(x_n) \subset K[x_0; r]$ gibt. Schließlich ist (x_n) eine *Cauchy³-Folge* (auch *Fundamental-Folge* genannt), wenn zu jedem $\varepsilon > 0$ ein Index $n_0(\varepsilon) \in \mathbb{N}$ mit $d(x_n, x_m) < \varepsilon$ für alle $n, m \geq n_0(\varepsilon)$ existiert. Es gelten dann die folgenden einfachen Aussagen:

- i. Jede konvergente Folge hat nur genau einen Grenzwert.
- ii. Jede konvergente Folge ist eine Cauchy-Folge.
- iii. Jede Cauchy-Folge ist beschränkt.
- iv. Aus $x_n \rightarrow x$, $y_n \rightarrow y$ folgt $d(x_n, y_n) \rightarrow d(x, y)$, d. h. die Metrik ist eine stetige Funktion ihrer beiden Argumente.

Aus den Aussagen ii. und iii. folgt für Folgen (x_n) die Implikation

$$(x_n) \text{ konvergent} \quad \Rightarrow \quad (x_n) \text{ Cauchy-Folge} \quad \Rightarrow \quad (x_n) \text{ beschränkt.}$$

³Augustin Louis Cauchy (1789–1857) entwickelte die von G. W. Leibniz und I. Newton aufgestellte Analysis weiter. Insbesondere in der Funktionentheorie stammen viele zentrale Sätze von ihm. Er deckte im Großen und Ganzen die komplette Bandbreite der damaligen Mathematik ab.

Im \mathbb{R}^n ist bekanntlich auch die Umkehrung von ii. richtig: Jede Cauchy-Folge ist konvergent. Dies ist der wesentliche Teil des Cauchyschen Konvergenzkriteriums im \mathbb{R}^n . Die Aussage ist nicht für alle metrischen Räume richtig. Deshalb definiert man:

Ein metrischer Raum (X, d) heißt *vollständig*, wenn jede Cauchy-Folge in X einen Grenzwert besitzt.

Beispiel 1.4

1. Der \mathbb{R}^n und die l^p -Räume mit den entsprechenden Metriken versehen, sind vollständige metrische Räume.
2. Die Räume $X := \mathbb{Q}$ (Raum der rationalen Zahlen) oder $X := [0, 1)$ (halboffenes reelles Intervall) mit dem üblichen Abstand sind nicht vollständig. ■

Für die Charakterisierung von abgeschlossenen Mengen ist der Begriff des *Häufungspunktes* erforderlich.

Definition 1.5 (Häufungspunkt)

Sei A Teilmenge eines metrischen Raumes (X, d) . Ein Punkt $x_0 \in X$ heißt *Häufungspunkt* der Menge A , falls eine Folge $(x_n) \subset A$ mit $x_n \neq x_0$, $n = 1, 2, \dots$ und $\lim_{n \rightarrow \infty} x_n = x_0$ existiert.

Die Menge \bar{A} , die aus A durch Hinzunahme aller Häufungspunkte entsteht, nennt man *Abschluss* (auch *abgeschlossene Hülle*) von A . Nun kann man zeigen, dass A genau dann abgeschlossen ist, wenn $A = \bar{A}$ ist. Der Durchschnitt beliebig vieler und die Vereinigung einer endlichen Anzahl abgeschlossener Mengen ist abgeschlossen. Mittels der *De Morganschen⁴ Formeln* ergibt sich eine entsprechende Aussage für offene Mengen: Die Vereinigung beliebig vieler und der Durchschnitt einer endlichen Anzahl offener Mengen ist offen.

Ein weiterer wichtiger Begriff für den Aufbau der Funktionalanalysis ist die *Kompaktheit* einer Menge.

Definition 1.6 (Kompaktheit)

Eine Teilmenge $A \subset X$ eines metrischen Raumes (X, d) heißt *kompakt*, wenn jede Folge (x_n) aus A eine konvergente Teilfolge mit Grenzwert in A enthält. Eine Menge A heißt *relativ kompakt*, wenn jede Folge $(x_n) \subset A$ eine konvergente Teilfolge mit Grenzwert in X enthält.

Die Bezeichnung „kompakt“ ist in der Literatur nicht einheitlich. Man verwendet für kompakt auch kompakt in sich und für relativ kompakt auch präkompakt.

Zu der Kompaktheitsdefinition ist noch die folgende Aussage äquivalent: *Jede offene Überdeckung von A enthält eine endliche Teilüberdeckung.* Ein allgemeines Kompakt-

⁴Augustus De Morgan (1806–1871), englischer Mathematiker. Er war Mitbegründer und erster Präsident der London Mathematical Society.

heitskriterium in einem metrischen Raum lautet:

Satz 1.7

Es sei $M \subset X$ und X ein vollständiger metrischer Raum. Dann sind folgende Aussagen äquivalent:

- i. M ist relativ kompakt.
- ii. M hat ein endliches ε -Netz, d. h. zu jedem $\varepsilon > 0$ existieren endlich viele Punkte $v_1, \dots, v_{k(\varepsilon)} \in M$, so dass $\min_{1 \leq j \leq k(\varepsilon)} d(u, v_j) \leq \varepsilon, \forall u \in M$.

Im euklidischen Raum \mathbb{R}^n ist Kompaktheit einer Teilmenge $A \subset \mathbb{R}^n$ äquivalent dazu, dass A beschränkt und abgeschlossen ist. Somit kann also aus jeder offenen Überdeckung von A eine endliche Teilüberdeckung ausgewählt werden. Dieses Überdeckungsprinzip wurde von *H. Heine*⁵ und *E. Borel*⁶ begründet.

1.1.2 Normierter Raum und Banach-Raum

Es seien \mathbb{R} die reellen und \mathbb{C} die komplexen Zahlen und X ein Vektorraum über dem Körper $\mathbb{K} = \mathbb{R}$ (reeller Vektorraum) oder $\mathbb{K} = \mathbb{C}$ (komplexer Vektorraum). Die Elemente des Vektorraumes heißen *Vektoren*, die Elemente des Körpers *Skalare*. Man beachte, dass sowohl der Nullvektor als auch die skalare Null mit 0 bezeichnet werden.

Definition 1.8 (Norm, normierter Raum)

Sei $\|\cdot\| : X \rightarrow \mathbb{R}$ eine Funktion mit folgenden Eigenschaften ($x, y, z \in X$ und $\lambda \in \mathbb{K}$):

- (N1) $\|x\| \geq 0$ und $\|x\| = 0 \Leftrightarrow x = 0$ (Definitheit),
- (N2) $\|\lambda x\| = |\lambda| \|x\|$ (Homogenität)
- (N3) $\|x + y\| \leq \|x\| + \|y\|$ (Dreiecksungleichung).

Eine solche Funktion $\|\cdot\|$ heißt *Norm* auf X und der damit ausgestattete Vektorraum X wird mit $(X, \|\cdot\|)$ bezeichnet und reeller oder komplexer *normierter Raum* genannt.

Die Norm auf X erzeugt eine Metrik

$$d(x, y) := \|x - y\|,$$

⁵Heinrich Eduard Heine (1821–1881), deutscher Mathematiker, führte u. a. als Erster den Begriff gleichmäßige Stetigkeit ein.

⁶Émile Borel (1871–1958), französischer Mathematiker, lieferte Beiträge zur Analysis und Wahrscheinlichkeitstheorie.

die X zu einem metrischen Raum (X, d) macht. Die eingeführten metrischen Begriffe beziehen sich immer auf diese „kanonische“ Metrik. Die Forderung der Vollständigkeit führt auf einen zentralen Raumbegriff der Funktionalanalysis.

Definition 1.9 (Banach-Raum)

Ein (bezüglich der kanonischen Metrik) vollständiger normierter Raum heißt *Banach-Raum*.

Für das Weitere sind folgende Eigenschaften nützlich:

- (i) $|\|x\| - \|y\|| \leq \|x - y\|$,
- (ii) aus $x_n \rightarrow x$ folgt $\|x_n\| \rightarrow \|x\|$.

Beispiel 1.10

1. Der Raum \mathbb{R}^n . Die Normaxiome (N1)-(N3) sind identisch mit den Eigenschaften des euklidischen Abstandes, die zugehörige kanonische Metrik ist gerade die euklidische Metrik. Wegen der Gültigkeit des Cauchyschen Konvergenzkriteriums ist der \mathbb{R}^n ein Banach-Raum.

2. Der Raum \mathbb{R}^n lässt sich auf vielfältige Weise normieren, etwa für $1 \leq p \leq \infty$ durch

$$\begin{aligned} \|x\|_p &:= (|x_1|^p + \dots + |x_n|^p)^{1/p} \quad (1 \leq p < \infty) && p\text{-Norm} \\ \|x\|_\infty &:= \max\{|x_1|, \dots, |x_n|\} && \text{Maximumnorm.} \end{aligned}$$

Der Grund für die Bezeichnungsweise $\|x\|_\infty$ ist in der Beziehung $\|x\|_p \rightarrow \|x\|_\infty$ für $p \rightarrow \infty$ zu sehen. Für $p = 1$ wird $\|x\|_1$ auch Betragssummennorm genannt.

Die Normeigenschaften (N1) und (N2) sind leicht zu verifizieren. Die Dreiecksungleichung (N3) ist im Fall der p -Norm identisch mit der Minkowskischen Ungleichung

$$\left(\sum_{i=1}^n |x_i + y_i|^p \right)^{1/p} \leq \left(\sum_{i=1}^n |x_i|^p \right)^{1/p} + \left(\sum_{i=1}^n |y_i|^p \right)^{1/p}. \quad (1.1)$$

3. Der Raum \mathbb{C}^n . Die Menge \mathbb{C}^n aller geordneten n -Tupel $z = (z_1, \dots, z_n)$ mit komplexen Koordinaten z_j bildet einen n -dimensionalen komplexen Vektorraum. Durch

$$\|z\| = \left(\sum_{i=1}^n |z_i|^2 \right)^{1/2}$$

wird auf \mathbb{C}^n eine Norm erklärt. Die Zuordnung

$$z = (z_1, \dots, z_n) \Leftrightarrow z^* = (\operatorname{Re} z_1, \operatorname{Im} z_1, \dots, \operatorname{Re} z_n, \operatorname{Im} z_n)$$

ist eine Isometrie (bijektive Abbildung, die die Abstände invariant lässt; siehe Abschnitt 1.2.1) zwischen \mathbb{C}^n und \mathbb{R}^{2n} . Es ist $\|z\| = \|z^*\|$. Konvergenzuntersuchungen im \mathbb{C}^n lassen sich also auf den \mathbb{R}^{2n} zurückführen. Der \mathbb{C}^n ist somit vollständig, also ein komplexer Banach-Raum.

4. Der Raum l^p , $1 \leq p < \infty$. Mit der Norm

$$\|x\|_p := \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{1/p} \quad (1.2)$$

ist l^p ein normierter Raum. Man erkennt, dass es sich hierbei um eine direkte Verallgemeinerung des \mathbb{R}^n mit der p -Norm auf den unendlichdimensionalen normierten Raum l^p handelt. Dem Leser sei empfohlen, den Beweis der Vollständigkeit der l^p -Räume zu führen.

5. Der Raum $C^0[a, b]$. Darunter versteht man die Menge aller stetigen Funktionen $f : [a, b] \rightarrow \mathbb{R}$ oder \mathbb{C} auf dem abgeschlossenen Intervall $[a, b]$, ausgestattet mit der Maximumnorm

$$\|f\| := \max\{|f(t)| \mid t \in [a, b]\}.$$

In dieser Norm ist die Konvergenz einer Funktionenfolge äquivalent zur gleichmäßigen Konvergenz dieser Folge. $C^0[a, b]$ ist vollständig, denn der Grenzwert einer gleichmäßig konvergenten Folge von stetigen Funktionen ist wieder stetig. Damit hat man das wichtige Ergebnis: *Der Raum $C^0[a, b]$ ist ein Banach-Raum.* ■

1.1.3 Skalarproduktraum und Hilbert-Raum

Das Skalarprodukt zweier Vektoren $x := (x_1, \dots, x_n)$, $y := (y_1, \dots, y_n)$ im \mathbb{K}^n ist dem Leser schon aus der Linearen Algebra vertraut. Es wird definiert durch

$$\langle x, y \rangle := \sum_{i=1}^n x_i y_i \quad \text{oder durch} \quad \langle x, y \rangle := \sum_{i=1}^n x_i \overline{y_i}, \quad (1.3)$$

je nach dem, ob es sich um den \mathbb{R}^n oder \mathbb{C}^n handelt. Es bezeichnet $\overline{y_i}$ die konjugiert komplexe Zahl zu y_i . Allgemein wird das Skalarprodukt durch die folgenden Eigenschaften erklärt.

Definition 1.11 (Skalarprodukt, Skalarproduktraum, Prähilbert-Raum)

Es sei X ein Vektorraum über dem Körper $\mathbb{K} = \mathbb{R}$ oder \mathbb{C} und es gebe eine Funktion $\langle \cdot, \cdot \rangle : X \times X \rightarrow \mathbb{K}$ mit den folgenden Eigenschaften ($x, y, z \in X$ und $\lambda, \mu \in \mathbb{K}$):

$$(H1) \quad \langle x, x \rangle > 0 \quad \text{für alle } x \neq 0,$$

$$(H2) \quad \langle \lambda x + \mu y, z \rangle = \lambda \langle x, z \rangle + \mu \langle y, z \rangle,$$

$$(H3) \quad \langle y, x \rangle = \overline{\langle x, y \rangle}. \quad (\overline{\langle x, y \rangle} \text{ konjugiert komplexe Zahl zu } \langle x, y \rangle.)$$

Eine solche Funktion $\langle \cdot, \cdot \rangle$ heißt ein *Skalarprodukt* (oder Innenprodukt) auf $X \times X$. Der damit versehene Vektorraum X heißt ein *Skalarproduktraum* oder *Prähilbert-Raum*.

Gemäß (H1) ist $\langle x, x \rangle$ auch im komplexen Fall reell und im reellen Fall lautet (H3) $\langle y, x \rangle = \langle x, y \rangle$. Aus (H2) und (H3) ergibt sich die Rechenregel

$$\langle x, \lambda y + \mu z \rangle = \bar{\lambda} \langle x, y \rangle + \bar{\mu} \langle x, z \rangle.$$

Wegen (H3) folgt $\langle 0, x \rangle = \overline{\langle x, 0 \rangle}$. Außerdem ergibt sich aus (H2) mit $\lambda = \mu = 0$ und $z = x : \langle 0, x \rangle = 0$. Somit hat man

$$\langle 0, x \rangle = \overline{\langle x, 0 \rangle} = 0 = \overline{\overline{\langle x, 0 \rangle}} = \langle x, 0 \rangle \quad \text{für alle } x \in X.$$

Jetzt wird gezeigt, dass jeder Prähilbert-Raum ein normierter Raum ist, wenn die Norm durch

$$(N) \quad \|x\| := \sqrt{\langle x, x \rangle}$$

definiert wird. Die Axiome (N1) und (N2) ergeben sich aus (H1)-(H3). Zum Nachweis der *Dreiecksungleichung* benötigt man eine nach *Cauchy und Schwarz*⁷ benannte Ungleichung: Für $x, y \in X$ gilt

$$|\langle x, y \rangle| \leq \|x\| \|y\| \quad (\text{Cauchy-Schwarzsche-Ungleichung}). \quad (1.4)$$

Gleichheit tritt genau dann ein, wenn x und y linear abhängig sind. Der Beweis dieser Ungleichung ist einfach: Sind x, y linear abhängig, so ist $y = 0$ oder $x = \lambda y$ und es besteht die Gleichheit. Andernfalls gilt für jedes $\lambda \in \mathbb{C}$

$$0 < \|x - \lambda y\|^2 = \langle x - \lambda y, x - \lambda y \rangle = \|x\|^2 - \alpha \bar{\lambda} - \bar{\alpha} \lambda + |\lambda|^2 \|y\|^2$$

mit $\alpha := \langle x, y \rangle$. Setzt man $\lambda := \alpha / \|y\|^2$ so ergibt sich die Behauptung $|\alpha| < \|x\| \|y\|$. Mit diesem Resultat erhält man nun die Dreiecksungleichung

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y \rangle = \|x\|^2 + \langle x, y \rangle + \langle y, x \rangle + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\| \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2 \end{aligned}$$

in quadrierter Form. Dies zeigt, dass (N) eine Norm ist.

Sind (x_n) und (y_n) Folgen in X mit $x_n \rightarrow x$, $y_n \rightarrow y$, so folgt aus der Tatsache, dass jede konvergente Folge beschränkt ist, etwa $\|y_n\| \leq K$ für alle $n \in \mathbb{N}$ und der Abschätzung

$$\begin{aligned} |\langle x_n, y_n \rangle - \langle x, y \rangle| &= |\langle x_n - x, y_n \rangle + \langle x, y_n - y \rangle| \\ &\leq \|x_n - x\| K + \|x\| \|y_n - y\| \rightarrow 0 \quad (n \rightarrow \infty) \end{aligned}$$

die Aussage:

$$\text{Aus } x_n \rightarrow x, y_n \rightarrow y \text{ folgt } \langle x_n, y_n \rangle \rightarrow \langle x, y \rangle.$$

⁷Hermann Amandus Schwarz (1843–1921), Professor u. a. in Göttingen und Berlin, Schüler von Karl Weierstraß.

Das ist die Stetigkeit des Skalarprodukts analog der Stetigkeitsaussage für die Norm. Da im Prähilbert-Raum ein Skalarprodukt zur Verfügung steht lässt sich der Begriff der Orthogonalität einführen

Definition 1.12 (Orthogonalität)

Zwei Vektoren $x, y \in X$ heißen *orthogonal*, in Zeichen $x \perp y$, wenn $\langle x, y \rangle = 0$ ist. Eine Menge $\{x_\alpha \in X \mid \alpha \in I\}$, I beliebige Indexmenge, heißt ein *Orthogonalsystem*, wenn dessen Vektoren paarweise orthogonal sind, bzw. ein *Orthonormalsystem* (ONS), wenn zudem jeder Vektor die Länge 1 hat, also $\|x_\alpha\| = 1$ für alle $\alpha \in I$ gilt.

Für orthogonale Vektoren gilt der Satz von Pythagoras:

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2, \quad \text{falls } x \perp y.$$

Die Behauptung folgt aus der Beziehung

$$\|x + y\|^2 = \langle x + y, x + y \rangle = \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle = \|x\|^2 + \|y\|^2.$$

Nun folgt die Definition des Hilbert-Raumes und im Anschluss daran Beispiele von Hilbert-Räumen.

Definition 1.13 (Hilbert-Raum)

Ein Prähilbert-Raum, der als normierter Raum vollständig ist, heißt *Hilbert-Raum*.

Beispiel 1.14

1. Der n -dimensionale euklidische Raum \mathbb{K}^n . Das in üblicher Weise in (1.3) definierte Skalarprodukt $x \cdot y \equiv \langle x, y \rangle$ ist ein Skalarprodukt auf dem \mathbb{K}^n . Die dadurch erzeugte Norm gemäß (N) ist gerade die euklidische Norm. Dieser Raum ist vollständig, also ein Hilbert-Raum.

2. Der Raum $\mathcal{C}^0[a, b]$ der auf dem abgeschlossenen Intervall $[a, b]$, $-\infty < a < b < \infty$, stetigen reellwertigen Funktionen wird durch die Festsetzung

$$\langle f, g \rangle := \int_a^b f(t)g(t) dt$$

zu einem Prähilbert-Raum, der jedoch nicht vollständig ist. Für den Nachweis, dass der Raum nicht vollständig ist, kann man auf $[a, b] = [-1, 1]$ die Funktionenfolge

$$f_n(t) := \begin{cases} 0 & t \leq 0 \\ nt & 0 < t < \frac{1}{n} \\ 1 & t \geq \frac{1}{n} \end{cases}, \quad t \in [-1, 1] \\ n = 1, 2, \dots$$

betrachten. Der Leser überlege sich, dass (f_n) eine Cauchy-Folge ist, die jedoch gegen die nicht stetige „Grenzfunktion“

$$g(t) := \begin{cases} 0, & t \leq 0 \\ 1, & t > 0 \end{cases} \quad t \in [-1, 1]$$

konvergiert und somit nicht zum Raum $C^0[a, b]$ gehört. Darin besteht der Widerspruch.

3. Der Hilbertsche Folgenraum l^2 . Gemäß der Definition der l^p -Räume (siehe Beispiel 1.10, 4.) ist jetzt $p = 2$, d. h. es wird der Raum der reellen Zahlenfolgen $x = (x_k)_{k \in \mathbb{N}}$ mit konvergenter Quadratsumme $\sum_{k=1}^{\infty} x_k^2 < \infty$ betrachtet. In diesem Folgenraum l^2 wird durch

$$\langle x, y \rangle := \sum_{i=1}^{\infty} x_i y_i$$

ein Skalarprodukt definiert, das die Norm

$$\|x\| = \sqrt{\langle x, x \rangle} = \left(\sum_{i=1}^{\infty} x_i^2 \right)^{1/2}$$

erzeugt. Wegen der Gültigkeit der Schwarzschen Ungleichung

$$\left(\sum_{i=1}^{\infty} x_i y_i \right)^2 \leq \left(\sum_{i=1}^{\infty} x_i^2 \right) \left(\sum_{i=1}^{\infty} y_i^2 \right)$$

(folgt sofort aus Ungleichung (1.4)) existiert das Skalarprodukt $\langle x, y \rangle$, da die Reihe $\sum_{i=1}^{\infty} x_i y_i$ sogar absolut konvergiert, also für $x, y \in l^2$ immer definiert ist. Mittels Minkowskischer Ungleichung

$$\left(\sum_{i=1}^{\infty} |x_i + y_i|^2 \right)^{1/2} \leq \left(\sum_{i=1}^{\infty} x_i^2 \right)^{1/2} + \left(\sum_{i=1}^{\infty} y_i^2 \right)^{1/2}$$

(siehe Walter (1995) ebenda) folgt aus $x, y \in l^2$ auch $x + y \in l^2$. Der Nachweis, dass l^2 ein Prähilbert-Raum ist, bereitet nun keine Schwierigkeiten mehr. Der Raum l^2 erweist sich sogar als ein Hilbert-Raum. Der Leser sollte den Beweis durchführen. ■

Jeder Hilbert-Raum ist ein Banach-Raum aber nicht umgekehrt. Ein Banach-Raum $(X, \|\cdot\|)$ ist genau dann ein Hilbert-Raum, wenn die Norm auf X die so genannte Parallelogrammgleichung

$$\|x - y\|^2 + \|x + y\|^2 = 2(\|x\|^2 + \|y\|^2), \quad x, y \in X \quad (\#)$$

erfüllt. Ausgehend von der in X gegebenen Norm muss daher das Skalarprodukt notwendig die Form

$$\langle x, y \rangle = \frac{1}{4} (\|x + y\|^2 - \|x - y\|^2)$$

besitzen. Ist die Gleichung $(\#)$ erfüllt, so existiert auf X ein Skalarprodukt $\langle x, y \rangle$ für das die Gleichung $\|x\| = \sqrt{\langle x, x \rangle}$, $x \in X$, gilt.

1.1.4 Orthogonalreihen in Hilbert-Räumen

Die klassischen *Fourier-Reihen* werden jetzt in ihrer Verallgemeinerung als Orthogonalreihen betrachtet. Sie sind das Ergebnis eines langen und mühevollen Erkenntnisprozesses bis sie sich als elegantes funktionalanalytisches Resultat präsentieren konnten. Eine sehr schöne Anwendung findet der Leser in dem Abschnitt über die Hopf-Bifurkation 4.3.

Sei $f : \mathbb{R} \rightarrow \mathbb{R}$ eine 2π -periodische Funktion. Dann lautet die zugehörige klassische Fourier-Reihe:

$$f(x) \sim \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx) \quad (1.5)$$

mit den so genannten *Fourier-Koeffizienten*

$$a_k := \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx \, dx, \quad b_k := \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx \, dx, \quad k = 0, 1, 2, \dots$$

Aus physikalischer Sicht vermittelt uns die Relation (1.5), dass eine 2π -periodische „Oszillation“ als Superposition einfacher „harmonischer Oszillationen“, bestehend aus $\cos kx$ und $\sin kx$ ($k \in \mathbb{N}$) der Periode $\frac{2\pi}{k}$, $k = 1, 2, \dots$, repräsentiert werden kann. Im 19. Jahrhundert studierten viele Mathematiker im Detail die Konvergenz der Fourier-Reihen. Erst im Jahr 1876 erhielt du Bois-Reymond das überraschende Resultat, dass es stetige Funktionen f gibt, deren Fourier-Reihe nicht in jedem Punkt $x \in \mathbb{R}$ gegen $f(x)$ konvergiert. Dieses Gegenbeispiel zeigt, dass die klassische Konvergenz der unendlichen Reihen nicht das richtige Konzept zur Lösung des fundamentalen Konvergenzproblems für Fourier-Reihen ist. Im Jahr 1907 bewiesen Fischer und Riesz unabhängig voneinander, dass eine natürliche Antwort auf das Konvergenzproblem (1.5) im Hilbert-Raum $L^2(-\pi, \pi)$ gegeben werden kann. Der Leser, der mit dem Raum der im Lebesgueschen Sinn quadratisch integrierbaren Funktionen L^2 nicht vertraut ist, findet einen sehr schönen Zugang in dem Buch von Heuser (1980), Teil 2. Eine kurze Darstellung und damit ein schneller Einstieg in dieses Thema ist in Hoffmann et al. (2006) nachzulesen.

In diesem Abschnitt wird gezeigt, dass dieses spezielle Konvergenzproblem (1.5) ein Spezialfall eines abstrakten Resultates für vollständige Orthonormalsysteme in Hilbert-Räumen ist. Mit anderen Worten: Es stellt sich heraus, dass für jede Funktion $f \in L^2(-\pi, \pi)$ die Fourier-Reihe (1.5) im Hilbert-Raum $L^2(-\pi, \pi)$ konvergiert, d. h. es gilt

$$\lim_{n \rightarrow \infty} \|f - s_n\| = 0 \quad \text{mit} \quad s_n(x) := \frac{a_0}{2} + \sum_{k=1}^n (a_k \cos kx + b_k \sin kx)$$

und $\|\cdot\|$ bezeichnet die Norm im $L^2(-\pi, \pi)$. Explizit heißt dies

$$\lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} (f(x) - s_n(x))^2 \, dx = 0.$$

Wir beginnen mit folgender Annahme:

(H) Sei X ein Hilbert-Raum über $\mathbb{K} = \mathbb{R}$ oder \mathbb{C} und sei $\{u_0, u_1, u_2, \dots\}$ ein endliches oder abzählbares *Orthonormalsystem* (ONS),

$$\langle u_k, u_m \rangle = \delta_{km} := \begin{cases} 1, & k = m \\ 0, & k \neq m \end{cases} \quad \text{für alle } k, m \in \mathbb{N} \cup \{0\}. \quad (1.6)$$

Ziel ist es, die Konvergenz der abstrakten Fourier-Reihe $\sum_{n=0}^{\infty} \langle f, u_n \rangle u_n$ (wegen dem ONS $\{u_n\}$ auch *Orthogonalreihe* genannt)

$$f = \sum_{n=0}^{\infty} \langle f, u_n \rangle u_n \quad (1.7)$$

zu studieren. Wir setzen $s_m := \sum_{k=0}^m \langle f, u_k \rangle u_k$. Die Zahlen $\langle f, u_k \rangle$ heißen *Fourier-Koeffizienten* von f . Bei der Entwicklung von Funktionen in Orthogonalreihen ist wesentlich, dass das vorliegende ONS umfangreich genug ist, um alle Elemente des Hilbert-Raumes approximieren zu können. Zum Beispiel bildet das System $\{u_1, u_2\}$ mit $u_1 = (1, 0, 0)$, $u_2 = (0, 1, 0)$ zwar ein ONS im (reellen) Hilbert-Raum \mathbb{R}^3 . Der Vektor $x_0 = (1, 1, 1)$ hat jedoch von allen Linearkombinationen $\sum_{j=1}^2 c_j u_j = c_1 u_1 + c_2 u_2 = (c_1, c_2, 0)$ einen Abstand

$$\|x_0 - \sum_{j=1}^2 c_j u_j\|^2 = (1 - c_1)^2 + (1 - c_2)^2 + 1 \geq 1$$

und kann somit durch das gegebene ONS $\{u_1, u_2\}$ nicht beliebig genau approximiert werden. Die entscheidende Eigenschaft eines ONS, die eine solche Approximierbarkeit gewährleistet, ist die sogenannte Vollständigkeit des ONS.

Definition 1.15 (Vollständiges Orthonormalsystem)

Es gelte (H). Das endliche ONS $\{u_0, u_1, \dots, u_n\}$ heißt vollständig in X , falls

$$f = \sum_{k=0}^n \langle f, u_k \rangle u_k \quad \text{für alle } f \in X$$

gilt. Das abzählbare ONS $\{u_0, u_1, u_2, \dots\}$ heißt vollständig in X , falls die unendliche Reihe (1.7) für alle $f \in X$ konvergiert, d. h. es gilt $f = \lim_{m \rightarrow \infty} s_m$.

Bemerkung 1.16

1. Jedes endliche ONS $\{u_0, u_1, \dots, u_n\}$ ist vollständig im Hilbert-Raum X über \mathbb{K} genau dann, wenn es eine Basis von X ist.
2. Sei $\{u_n\}$ ein abzählbares ONS im Hilbert-Raum X und es sei angenommen, dass die unendliche Reihe $\sum_{n=0}^{\infty} c_n u_n$, $c_n \in \mathbb{K}$ gegen ein festes $f \in X$ konvergiert. Dann gilt für

alle $n \in \mathbb{N} \cup \{0\} : c_n = \langle f, u_n \rangle$. Dieser Sachverhalt wird klar, wenn man (1.6) und die Stetigkeit des Skalarproduktes benutzt:

$$\begin{aligned} \langle f, u_k \rangle &= \left\langle \sum_{n=0}^{\infty} c_n u_n, u_k \right\rangle = \left\langle \lim_{m \rightarrow \infty} \sum_{n=0}^m c_n u_n, u_k \right\rangle \\ &= \lim_{m \rightarrow \infty} \left\langle \sum_{n=0}^m c_n u_n, u_k \right\rangle = \lim_{m \rightarrow \infty} \sum_{n=0}^m c_n \langle u_n, u_k \rangle = c_k. \end{aligned}$$

Eine weitere Motivation für den Ansatz (1.7) ergibt sich aus dem Problem, f durch eine Summe $\sum_{k=0}^m \gamma_k u_k$, $\gamma_k \in \mathbb{C}$, beliebig vorgegeben und festem $m \in \mathbb{N}$ möglichst gut zu approximieren. Es besteht dann die folgende Approximationsformel

$$\begin{aligned} \left\| f - \sum_{k=0}^m \gamma_k u_k \right\|^2 &= \left\langle f - \sum_{k=0}^m \gamma_k u_k, f - \sum_{k=0}^m \gamma_k u_k \right\rangle \\ &= \langle f, f \rangle - \sum_{k=0}^m \bar{\gamma}_k \langle f, u_k \rangle - \sum_{k=0}^m \gamma_k \langle u_k, f \rangle + \sum_{k=0}^m \gamma_k \bar{\gamma}_k \\ &= \|f\|^2 - \sum_{k=0}^m |\langle f, u_k \rangle|^2 + \sum_{k=0}^m |\langle f, u_k \rangle - \gamma_k|^2. \end{aligned} \quad (1.8)$$

Mit $c_k := \langle f, u_k \rangle$ und $\bar{c}_k = \overline{\langle u_k, f \rangle}$ ergibt sich die letzte Umformung, wenn man die einfache Beziehung

$$|c_k - \gamma_k|^2 = (c_k - \gamma_k) \overline{(c_k - \gamma_k)} = |c_k|^2 + |\gamma_k|^2 - c_k \bar{\gamma}_k - \bar{c}_k \gamma_k$$

verwendet. Die Funktion

$$h(\gamma_0, \gamma_1, \dots, \gamma_m) := \left\| f - \sum_{k=0}^m \gamma_k u_k \right\|^2$$

nimmt ihr Minimum für $\gamma_k = c_k = \langle f, u_k \rangle$, $k = 0, 1, \dots, m$ an. Aus der Approximationsformel (1.8) kann man nun die folgende Eigenschaft direkt ablesen.

Satz 1.17 (Approximationsatz)

Unter allen endlichen Linearkombinationen $\sum_{k=0}^m \gamma_k u_k$ stellt das Element $s_m = \sum_{k=0}^m c_k u_k$ (mit $c_k = \langle f, u_k \rangle$) und nur dieses die beste Approximation von f dar. Es ist

$$\|f - s_m\|^2 = \|f\|^2 - \sum_{k=0}^m |c_k|^2 \leq \left\| f - \sum_{k=0}^m \gamma_k u_k \right\|^2 \quad \text{für alle } m = 0, 1, 2, \dots \quad (1.9)$$

Die Koeffizienten der besten Approximation sind also die Fourier-Koeffizienten von f . Aus (1.9) folgt außerdem noch die *Besselsche Ungleichung*

$$\sum_{n=0}^m |c_n|^2 = \sum_{n=0}^m |\langle f, u_n \rangle|^2 \leq \|f\|^2, \quad \text{für alle } f \in X \text{ und alle } m \in \mathbb{N}. \quad (1.10)$$

Damit die Beziehung (1.7) überhaupt Sinn macht, muss nun die Konvergenz der Reihe $\sum_{n=0}^{\infty} c_n u_n$ untersucht werden.

Satz 1.18 (Konvergenzkriterium)

Es sei $\{u_n\}$ ein abzählbares ONS in einem Hilbert-Raum X über \mathbb{K} . Die unendliche Reihe $\sum_{n=0}^{\infty} c_n u_n$, $c_n \in \mathbb{K}$, ist genau dann konvergent, wenn die Reihe $\sum_{n=0}^{\infty} |c_n|^2$ konvergiert.

Beweis: Sei wieder $s_m := \sum_{n=0}^m c_n u_n$. Wegen (1.6) folgt

$$\begin{aligned} \|s_{m+k} - s_m\|^2 &= \left\langle \sum_{n=m+1}^{m+k} c_n u_n, \sum_{l=m+1}^{m+k} c_l u_l \right\rangle \\ &= \left\| \sum_{n=m+1}^{m+k} c_n u_n \right\|^2 = \sum_{n=m+1}^{m+k} |c_n|^2 \end{aligned} \quad (1.11)$$

für alle $m, k = 1, 2, \dots$. Setzt man voraus, dass die Reihe $\sum_{n=0}^{\infty} |c_n|^2$ konvergiert, dann ist (s_m) eine Cauchy-Folge im Hilbert-Raum X und folglich konvergent. Wird umgekehrt „ $\sum_{n=0}^{\infty} c_n u_n$ ist konvergent“ vorausgesetzt, dann ist (s_m) eine Cauchy-Folge und folglich $\sum_{n=0}^{\infty} |c_n|^2$ konvergent wegen (1.11). \square

Das Konvergenzkriterium kann nun kurz so gefasst werden:

$$\sum_{n=0}^{\infty} c_n u_n \text{ konvergent} \quad \Leftrightarrow \quad c = (c_n) \in l^2.$$

Aus der Besselschen Ungleichung (1.10) folgt wieder, dass die Reihe, gebildet mit den Fourier-Koeffizienten von f , $\sum_{n=0}^{\infty} |\langle f, u_n \rangle|^2$, konvergiert, d. h. es gilt $(\langle f, u_n \rangle) \in l^2$. Somit ist für jedes $f \in X$ die Fourier-Reihe konvergent, d. h. es gibt ein $g \in X$, so dass

$$g = \sum_{n=0}^{\infty} \langle f, u_n \rangle u_n.$$

Es ist durchaus möglich, dass $g \neq f$ gilt. Um $g = f$ für alle $f \in X$ zu sichern, benötigt man den in Definition 1.15 eingeführten Vollständigkeitsbegriff. Es sei hier noch erwähnt, dass der Begriff des vollständigen ONS auch äquivalent zu der Aussage ist: Aus $\langle f, u_n \rangle = 0$ für alle $n = 0, 1, 2, \dots$ folgt stets $f = 0$. Diese Formulierung kommt unserer anschaulichen Vorstellung, dass z. B. im euklidischen \mathbb{R}^3 Null das einzige Element ist, das zu allen Einheitsvektoren i, j, k orthogonal ist, am nächsten. Sie stellt die natürliche Verallgemeinerung auf einen unendlichdimensionalen Hilbert-Raum dar.

Mit diesem Begriff können wir nun (als Höhepunkt der allgemeinen Theorie der Fourier-Reihen in Hilbert-Räumen) den folgenden Satz beweisen.

Satz 1.19 (Darstellungssatz)

Es sei (H) erfüllt und das ONS $\{u_n\}_{n=0}^\infty$ vollständig. Dann wird jedes Element $f \in X$ durch seine Fourier-Reihe dargestellt:

$$f = \sum_{n=0}^{\infty} c_n u_n, \quad c_n = \langle f, u_n \rangle.$$

Beweis: Es sei $f \in X$ beliebig vorgegeben. Da $(\langle f, u_n \rangle) \in l^2$, konvergiert die zugehörige Fourier-Reihe:

$$f^* = \sum_{n=0}^{\infty} c_n u_n \quad \text{mit} \quad c_n = \langle f, u_n \rangle, \quad n = 0, 1, 2, \dots$$

Wir müssen zeigen, dass $f = f^*$ ist. Nun gilt jedoch $\langle f - f^*, u_m \rangle = \langle f, u_m \rangle - \langle f^*, u_m \rangle = \langle f, u_m \rangle - c_m = 0$ für alle $m = 0, 1, 2, \dots$. Wegen der Vollständigkeit des ONS $\{u_n\}$ folgt $f - f^* = 0$. \square

Eine weitere interessante Beobachtung ist die Folgende: Es sei $\{u_n\}_{n=0}^\infty$ ein vollständiges ONS in X und $f \in X$ gegeben. Dann vermittelt

$$U : X \rightarrow l^2, \quad Uf := (\langle f, u_n \rangle)_{n=0}^\infty \tag{1.12}$$

mit der Umkehrabbildung

$$U^{-1} : l^2 \rightarrow X, \quad U^{-1}c = \sum_{n=0}^{\infty} c_n u_n$$

eine lineare, bijektive und isometrische Abbildung (zur Begriffsbildung siehe Abschnitt 1.2.1) von X nach l^2 .

Vom funktionalanalytischen Standpunkt kann man nun die Räume X und l^2 als „gleich“ ansehen. Sie unterscheiden sich lediglich durch ihre Repräsentanten. Da der Raum l^2 *separabel* ist, d. h. er besitzt eine abzählbare dichte Teilmenge von Elementen aus dem l^2 , gelten alle gemachten Aussagen für separable Hilbert-Räume. Weiter kann man zeigen, dass in jedem separablen Hilbert-Raum X über \mathbb{K} mit $X \neq \{0\}$ ein vollständiges ONS existiert. Hat man eine abzählbare Menge $\{v_0, v_1, v_2, \dots\}$, die darüber hinaus dicht in X liegt, dann kann man sie nach der Schmidtschen Orthogonalisierungsmethode (siehe Zeidler (1995a), Abschnitt 3.3) zu einem ONS $\{u_0, u_1, u_2, \dots\}$ machen.

1.2 Lineare Operatoren

In diesem Abschnitt werden einige fundamentale Eigenschaften linearer Operatoren zusammengestellt. Ein Schwerpunkt liegt dabei auf *Fredholm-* und *kompakten Operatoren*.

Diese bilden die Grundlage für die Bifurkationstheorie im Kapitel 4. Wir werden zunächst einige gebräuchliche Bezeichnungen und Sprechweisen für (nichtlineare) Operatoren einführen.

1.2.1 Bezeichnungen und Begriffe

Für diesen Abschnitt werden Vektorräume bzw. dort wo es erforderlich ist auch normierte Räume X und Y zu Grunde gelegt. Zunächst folgen Schreibweisen für eine Abbildung:

$$T : X \rightarrow Y, \quad Tu = v \text{ (oder } T(u) = v), \quad u \in X, v \in Y, \quad \text{bzw.}$$

$$T : X \rightarrow Y, \quad u \mapsto Tu =: v.$$

X heißt *Definitionsbereich* (Domain) von T , geschrieben $\text{dom}(T)$. Für den *Wertebereich* (Range) von T wird geschrieben.

$$R(T) := \{v \in Y \mid Tu = v \text{ für ein } u \in X\}.$$

Das Element v heißt *Bild* von u unter der Abbildung T (siehe Abbildung 1.1). Ist $R(T) =$

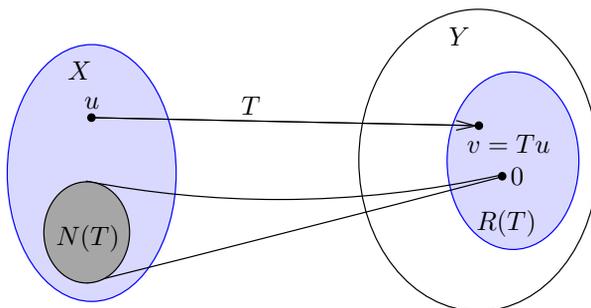


Abb. 1.1 Begriffsbildungen anhand einer Abbildung $T : X \rightarrow Y$

Y , dann heißt T *surjektiv*. Man sagt: T bildet X *auf* Y ab, ansonsten bildet T die Menge X *in* Y ab. Weiter bezeichnen wir mit

$$\ker(T) \equiv N(T) := \{u \in X \mid Tu = 0\} \quad \text{den Kern (bzw. Nullraum) von } T.$$

Das *Urbild* (auch *inverses Bild*) eines Elementes $v \in Y$ wird mit $T^{(-1)}(v)$ bezeichnet und ist die folgende Menge

$$T^{(-1)}(v) := \{u \in X \mid Tu = v\}$$

bzw. für eine Menge W (siehe Abbildung 1.2)

$$T^{(-1)}(W) := \{u \in X \mid Tu \in W\}.$$

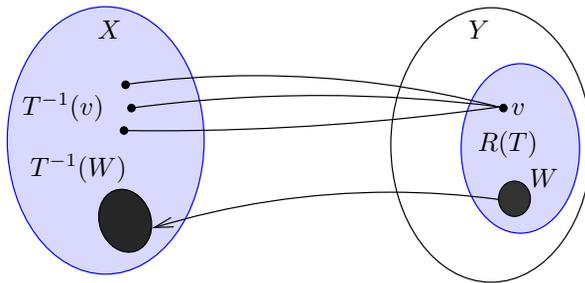


Abb. 1.2 Das inverse Bild eines Elementes oder einer Menge

Der *identische Operator* ist definiert durch $I : X \rightarrow X$, $Iu = u$ für alle $u \in X$. Der Operator $T : X \rightarrow Y$ heißt *injektiv* (bzw. *one-to-one*, bzw. 1-1), wenn Folgendes gilt (siehe Abbildung 1.3):

$$\begin{aligned} u_1 \neq u_2 &\Rightarrow Tu_1 \neq Tu_2, \quad \forall u_1, u_2 \in X, \quad \text{oder äquivalent} \\ Tu_1 = Tu_2 &\Rightarrow u_1 = u_2, \quad \forall u_1, u_2 \in X. \end{aligned}$$

Der Operator T^{-1} heißt dann *inverser Operator* bzw. *Inverse* von T , da er v nach u eindeutig (zurück) abbildet. (*Achtung*: Man unterscheide $T^{(-1)}$ und T^{-1} !) Die Inverse T^{-1} ist durch

$$T^{-1} : R(T) \rightarrow X, \quad T^{-1}(Tu) = u, \quad \forall u \in X.$$

definiert. Falls $R(T) = Y$ (d. h. T ist surjektiv) und T auch noch injektiv ist, dann heißt T *bijektiv*. Alle vorkommenden Typen einer Abbildung T findet man in den Beispielen 1.22 und 1.23.

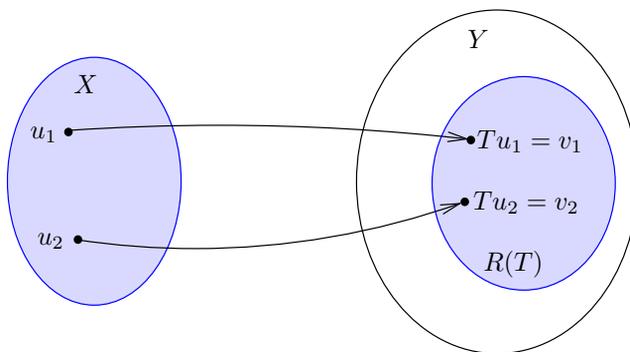


Abb. 1.3 Ein injektiver Operator

Es seien X und Y Vektorräume über $\mathbb{K} = \mathbb{R}$ oder \mathbb{C} . Eine Abbildung $T : X \rightarrow Y$ heißt *linear*, wenn für alle $\alpha, \beta \in \mathbb{K}$ und alle $x, y \in X$ Folgendes gilt:

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y).$$

Darüber hinaus gilt für lineare Operatoren der wichtige

Satz 1.20

Es seien X, Y normierte Räume und $T : X \rightarrow Y$ ein linearer Operator. Dann gilt:

$$T \text{ injektiv} \Leftrightarrow \dim \ker(T) = 0.$$

Die nachfolgenden Begriffe werden nur für normierte Räume bzw. Banach-Räume benötigt und somit auch nur in diesem Kontext formuliert, obwohl die Definitionen auch für metrische Räume (topologische Räume) zutreffen.

Es seien zwei normierte Räume $(X, \|\cdot\|_X)$ und $(Y, \|\cdot\|_Y)$ und eine Abbildung $T : X \rightarrow Y$ mit der Eigenschaft

$$\|T(x) - T(y)\|_Y = \|x - y\|_X$$

gegeben. Dann heißt T eine *Isometrie* von X nach Y . Eine solche Abbildung ist stets injektiv, da aus $T(x) = T(y)$ sofort $x = y$ folgt. Ist T sogar bijektiv, dann heißt T *isometrischer Isomorphismus* und die Räume X und Y heißen *isometrisch isomorph*. Im Fall normierter Räume spricht man auch von Norm-Isomorphismus. Die Aussage „ X und Y sind isometrisch isomorph“ wird üblicherweise als $X \cong Y$ geschrieben. Die Abbildung lässt also die Abstände zwischen den Bildern und den Urbildern von T invariant (unverändert), wie Abbildung 1.4 zeigt.

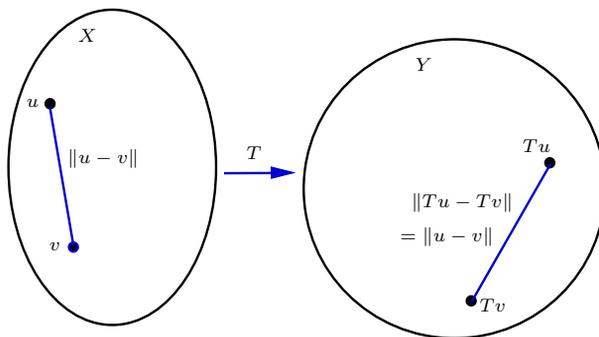


Abb. 1.4 Zwei isometrisch isomorphe Räume

Hinter dieser Begriffsbildung steckt die Vorstellung, wann zwei unterschiedliche Räume „als gleich“ angesehen werden können: Nämlich dann, wenn man von der Darstellung der Elemente der zugrunde liegenden Räume und den Namen der Relationen und Verknüpfungen absieht. Als ein Beispiel für diesen Sachverhalt kann der Operator (1.12) angeführt werden.

Ein *Homöomorphismus* T (nicht zu verwechseln mit Homomorphismus) ist ein zentraler Begriff im mathematischen Teilgebiet Topologie. In diesem Buch bezeichnet T eine bijektive, stetige Abbildung zwischen zwei normierten Räumen (bzw. Teilmengen), dessen Umkehrabbildung ebenfalls stetig ist. Die dabei zugrunde gelegte Definition der Stetig-

keit ist abhängig von den betrachteten normierten Räumen. Zwei normierte Räume heißen *homöomorph* (auch „*topologisch äquivalent*“), wenn sie durch einen Homöomorphismus ineinander überführt werden können. Sie sind unter topologischen Gesichtspunkten gleichartig. Eine solche Abbildung spielt eine wesentliche Rolle beim Satz von Hartman und Grobman (Satz 2.54).

Beispiel 1.21 (Homöomorphismen)

1. Jede Kreisscheibe $K[x_0; r]$, $r > 0$, ist homöomorph zu jedem Quadrat (mit Seitenlänge $a > 0$) in der euklidischen Ebene \mathbb{R}^2 . Eine Kreisscheibe lässt sich offenbar durch Verbiegen und Verzerren, ohne Zerschneiden, in ein Quadrat überführen, und umgekehrt.
 2. Das offene Intervall $(0, 1)$ ist homöomorph zum Raum \mathbb{R} aller reellen Zahlen. Jedes offene Intervall lässt sich ohne Weiteres ins Unendliche verzerren. Ein Homöomorphismus, der dies für $(0, 1)$ vermittelt, ist zum Beispiel

$$T : (0, 1) \rightarrow \mathbb{R}, \quad x \mapsto \tan\left(\left(x - \frac{1}{2}\right) \cdot \pi\right).$$

3. Der Produktraum $S(0; 1) \times S(0; 1)$ des Einheitskreises mit sich selbst ist homöomorph zum zweidimensionalen Torus, also zur Form eines Fahrradschlauches. ■

Ist der Homöomorphismus T darüber hinaus auch noch stetig differenzierbar, dann sprechen wir von einem *Diffeomorphismus*.

Beispiel 1.22 (Matrixabbildung)

Ein linearer Operator $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ kann als eine $m \times n$ -Matrix (T_{ij}) ($i = 1, \dots, m; j = 1, \dots, n$) aufgefasst werden. Ob T injektiv bzw. surjektiv ist, hängt von den Einträgen T_{ij} ab. Wir nehmen an, dass die Abbildung $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ durch die 2×3 -Matrix

$$Tx := \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

repräsentiert wird. Setzt man $T_{11} = T_{21} = 1$ und die restlichen $T_{ij} = 0$, so ergibt sich

$$R(T) = \{y \in \mathbb{R}^2 \mid y_1 = y_2\} \subsetneq \mathbb{R}^2,$$

d. h. T ist nicht surjektiv. Der Nullraum lautet

$$\ker(T) = \{x \in \mathbb{R}^3 \mid x_1 = 0\}.$$

Folglich ist $\dim \ker(T) = 2$. Es handelt sich hier um die x_2x_3 -Ebene.

Das gegebene T kann bei beliebiger Wahl der T_{ij} nie injektiv sein, da stets $\text{rang}(T) \leq 2$ gilt und wegen der *Dimensionsformel*

$$n = \text{rang}(T) + \dim \ker(T) \tag{1.13}$$

für die Dimension des Nullraumes $\dim \ker(T) \geq 1$ folgt. Der Rang von T ist durch $\text{rang}(T) := \dim R(T)$ gegeben.

Falls T injektiv ist, kann niemals $m < n$ sein. Dies würde wegen (1.13) sofort zu dem Widerspruch $n = \text{rang}(T) \leq \min\{n, m\} = m$ führen. Somit hat man als notwendige Bedingung für die Injektivität die Ungleichung $m \geq n$. Ein Beispiel für eine injektive Abbildung $T \neq 0$ mit $m > n$ ist

$$T : \mathbb{R} \rightarrow \mathbb{R}^2, \quad Tx := \begin{pmatrix} T_{11} \\ T_{21} \end{pmatrix} (x).$$

Wegen $\text{rang}(T) = 1$ folgt $\dim \ker(T) = 0$, was man auch direkt nachrechnet.

Speziell hat man für $n = m$ die folgende Situation: Ist die Determinante $\det(T) = 0$, dann ist $\text{rang}(T) < n$ (mindestens zwei Zeilen der Matrix T sind linear abhängig) und nach der Dimensionsformel (1.13) ist $\ker(T) \neq \{0\}$. Gilt hingegen für die Determinante $\det(T) \neq 0$, dann ist $\text{rang}(T) = n$ und $\ker(T) = \{0\}$. Damit hat man die Aussage: T ist *entweder* sowohl surjektiv als auch injektiv *oder* weder surjektiv noch injektiv. ■

Beispiel 1.23 (Differentialoperator)

Der Operator

$$T := d/dx : \mathcal{C}^1[a, b] \rightarrow \mathcal{C}^0[a, b], \quad u \mapsto u'$$

ist nicht injektiv, da aus $Tu_1 = Tu_2$ nicht $u_1 = u_2$ folgt. Jede Funktion $u_1 = u_2 + a$, $a = \text{const.}$ erfüllt ebenfalls die Bedingung $Tu_1 = Tu_2$. Aus $Tu = u' = v$ folgt durch Integration $u(x) - u(a) = \int_a^x v(\tau) d\tau$. Schränkt man jedoch den Definitionsbereich auf

$$X := \{u \in \mathcal{C}^1[a, b] \mid u(a) = 0\} \subset \mathcal{C}^1[a, b]$$

ein, dann ergibt sich der inverse Operator

$$u(x) = \int_a^x v(\tau) d\tau = (T^{-1}(v))(x).$$

Aus $Tu_1 = Tu_2$ folgt nämlich jetzt $u'_1 = u'_2$, bzw. $(u_1 - u_2)' = 0$, was wiederum bedeutet, dass $u_1 - u_2 = \text{const.}$ ist. Wegen $u_1, u_2 \in X$ folgt $u_1(a) - u_2(a) = 0 = \text{const.}$ und damit $u_1(x) = u_2(x)$ für alle $x \in [a, b]$. Dies beweist die Injektivität. ■

Die Situation, dass man durch Einschränkung des Definitionsbereiches einer Funktion f eine injektive Funktion erhalten kann, sieht man bereits ganz leicht an der Funktion $f(x) := \sin(x)$. Betrachtet man f auf \mathbb{R} , dann ist f wegen $f(x_1 + 2n\pi) = f(x_1)$ für jedes x_1 nicht injektiv, während $f : [-\pi/2, \pi/2] \rightarrow [-1, 1]$ sehr wohl injektiv ist. Die Inverse ist definiert durch $f^{-1}(y) := \arcsin y$.

1.2.2 Lineare stetige Operatoren

Es folgt eine Auswahl an Begriffen und Sätzen über lineare stetige Operatoren. Dabei wird kein Anspruch auf Vollständigkeit erhoben. Der Leser, der sich weitere Kenntnisse über lineare Operatoren verschaffen möchte, sollte die Literatur (z. B. Appell und Väh (2005), Alt (1992), Heuser (2006), Werner (1997), Kantorowitsch und Akilow (1978)) studieren. Die Auswahl an Begriffen und Sätzen ist jedoch so gewählt, dass auch ohne weiteres Literaturstudium die nachfolgenden Kapitel verstanden werden können.

Wir betrachten lineare Abbildungen $T : X \rightarrow Y$. Hierbei bezeichnen $(X, \|\cdot\|_X)$ und $(Y, \|\cdot\|_Y)$ normierte Räume über demselben Körper $\mathbb{K} (= \mathbb{R} \vee \mathbb{C})$.

$$T \text{ heißt beschränkt} := \exists C \geq 0 \forall u \in X : \|Tu\|_Y \leq C \|u\|_X. \quad (1.14)$$

Mit diesem Begriff kann die Stetigkeit linearer Operatoren charakterisiert werden.

Lemma 1.24

Für lineare Operatoren $T : X \rightarrow Y$ sind die folgenden drei Bedingungen äquivalent:

- (i) T ist stetig auf ganz X . (ii) T ist beschränkt. (iii) T ist stetig in $x_0 \in X$.

Damit haben wir ein bemerkenswertes Ergebnis: Ein linearer Operator ist *entweder* in jedem Punkt stetig *oder* in jedem Punkt unstetig.

Wir setzen

$$\mathcal{L}(X, Y) := \{ T : X \rightarrow Y \mid T \text{ ist linear und beschränkt} \}$$

und nennen $\mathcal{L}(X, Y)$ den Vektorraum der linearen stetigen Operatoren. Es lässt sich nämlich leicht zeigen, dass $\mathcal{L}(X, Y)$ mit den üblichen Operationen der Addition „+“ und der skalaren Multiplikation „·“ ein Vektorraum ist.

Wir führen nun in $\mathcal{L}(X, Y)$ eine geeignete Norm ein. Die kleinste Konstante $C \geq 0$ mit der die Ungleichung $\|Tu\| \leq C\|u\|$ gilt, wird mit $\|T\|$ bezeichnet:

$$\|T\| := \inf \{ C \geq 0 \mid \|Tu\| \leq C\|u\|, u \in X \}. \quad (1.15)$$

Die Zahl $\|T\|$ heißt *Operatornorm* von T . Sie ergibt sich auch aus

Satz 1.25 (Berechnung der Operatornorm)

Für $T \in \mathcal{L}(X, Y)$ gilt:

$$\|T\| = \sup_{\substack{u \in X \\ u \neq 0}} \frac{\|Tu\|}{\|u\|} = \sup_{\|u\| \leq 1} \|Tu\| = \sup_{\|u\|=1} \|Tu\|. \quad (1.16)$$

Dieser Satz führt zu der einfachen Folgerung:

i. Für $T \in \mathcal{L}(X, Y)$ gilt

$$\|Tu\| \leq \|T\| \|u\| \quad \text{für alle } u \in X. \quad (\text{klar, wegen (1.15)})$$

ii. Jedes $T \in \mathcal{L}(X, Y)$ ist Lipschitz-stetig, also insbesondere gleichmäßig stetig.

iii. Die Abbildung

$$\mathcal{L}(X, Y) \rightarrow \mathbb{R}^+, \quad T \mapsto \|T\|$$

ist eine Norm auf $\mathcal{L}(X, Y)$.

Resolvente und Spektrum

Aus der Linearen Algebra kennt man für quadratische Matrizen die Begriffe des *Eigenwertes* und des *Eigenvektors*. Wir verallgemeinern diese Begriffe auf normierte Räume X^8 über $\mathbb{K} = \mathbb{R}$ bzw. $\mathbb{K} = \mathbb{C}$. Wir untersuchen die Menge (Schar, Familie) von Operatoren

$$T - \lambda I \quad (T \in \mathcal{L}(X, X), \lambda \in \mathbb{K}),$$

die für jedes fest gewählte λ eine Abbildung von X in sich liefert.

Definition 1.26 (Eigenwert, Eigenvektor, charakteristischer Wert)

Eine Zahl $\lambda \in K$ heißt *Eigenwert* von $T \in \mathcal{L}(X, X)$, wenn ein Element $x \neq 0$ existiert, für das $Tx = \lambda x$ gilt. Die Lösung $x \neq 0$ heißt *Eigenvektor* von T . Die Zahl $\mu := \frac{1}{\lambda}$ ($\lambda \neq 0$) heißt *charakteristischer Wert* von T .

Aus der Definition folgt, dass die Zahl $\lambda = 0$ dann und nur dann Eigenwert ist, wenn ein $x \neq 0$ existiert, für das $Tx = 0$ gilt, d. h. wenn T nicht injektiv ist. Mit Ausnahme der Null sind also die Eigenwerte gerade die Kehrwerte der charakteristischen Werte von T (und umgekehrt).

Beispiel 1.27

Es seien $X := \mathbb{R}^n$ und $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$. Dann ist T durch eine Matrix $A := (a_{ij})_{i,j=1}^n$ darstellbar. Ist T invertierbar, so ist $\mu \in \mathbb{R}$ genau dann charakteristischer Wert von T , falls $\ker(I - \mu T) = \ker(\lambda I - T) \neq \{0\}$ ist. Die Zahlen $\lambda = 1/\mu$ sind genau die Eigenwerte von T . Insbesondere gibt es höchstens n charakteristische Werte für T . Ist T nicht invertierbar, so hat T noch den Eigenwert $\lambda = 0$ (der nicht Kehrwert eines charakteristischen Wertes ist). ■

Wir wollen nun alle Werte λ kennzeichnen, für die die lineare Inverse $(T - \lambda I)^{-1}$ auf (ganz) X existiert.

⁸Studiert man reelle (bzw. komplexe) Banach-Räume X , so muss man auch $\lambda \in \mathbb{R}$ (bzw. $\lambda \in \mathbb{C}$) wählen.

Definition 1.28 (Resolvente, Spektrum)

Es sei $T : X \rightarrow X$ ein linearer stetiger Operator. Die Menge aller Zahlen $\lambda \in \mathbb{K}$, für die die Abbildung $T - \lambda I$ surjektiv ist und für die die Umkehrabbildung

$$R(\lambda; T) \equiv R_\lambda(T) := (T - \lambda I)^{-1}$$

ein (linearer) stetiger Operator ist, heißt *Resolventenmenge* $\rho(T)$ von T . Die Elemente von $\rho(T)$ werden *reguläre Werte* von T genannt. Der Umkehroperator $R_\lambda(T) = (T - \lambda I)^{-1}$ des Operators $T - \lambda I$ heißt die *Resolvente* von T im Punkt λ . Die Komplementärmenge der Resolventenmenge $\rho(T)$, die Menge

$$\sigma(T) := \mathbb{K} \setminus \rho(T), \tag{1.17}$$

heißt *Spektrum* von T . Ein Element von $\sigma(T)$ wird *Spektralwert* von T genannt.

Ist $T \in \mathcal{L}(X, X)$, $T \neq 0$, so genügen alle charakteristischen Werte μ von T der Abschätzung $|\mu| \geq 1/\|T\|$. Das Spektrum eines linearen Operators ist also in einer Kreisscheibe um Null mit Radius $\|T\|$ enthalten.

Wenn $U := T - \lambda I$ eine lineare stetige Abbildung ist, die den normierten Raum X injektiv auf sich abbildet, so existiert die Umkehrabbildung $U^{-1} : X \rightarrow X$, d. h. die Gleichung $Ux = y$ ist äquivalent zur Gleichung $x = U^{-1}y$ für $x, y \in X$. Der Operator U^{-1} ist linear, aber nicht notwendig stetig. Ist aber zusätzlich X ein Banach-Raum, dann muss U^{-1} notwendig auch stetig sein. Dies wird durch das Open Mapping Theorem (Satz 1.34) begründet.

Somit gehört eine Zahl λ genau dann zur Resolventenmenge eines auf einem Banach-Raum X definierten linearen stetigen Operators T , wenn die Abbildung $T - \lambda I$ eine injektive Abbildung von X auf sich (also bijektiv) ist. Die Gleichung

$$Tx - \lambda x = y$$

ist dann für jedes $y \in X$ mit genau einem x lösbar und dieses x hängt stetig von y ab (falls y variiert wird).

Die Resolventenmenge $\rho(T)$ und das Spektrum $\sigma(T)$ lassen sich auch noch wie folgt charakterisieren:

Satz 1.29

Die Resolventenmenge $\rho(T)$ ist eine (in \mathbb{K}) offene Menge, das Spektrum $\sigma(T)$ ist eine abgeschlossene und beschränkte Menge.

Das Spektrum $\sigma(T)$ kann im Fall eines reellen Banach-Raumes (z. B. $X := \mathbb{R}^n$) auch leer sein. Für komplexe Banach-Räume X kann man zeigen, dass für $T \in \mathcal{L}(X, X)$ das

Spektrum $\sigma(T) \neq \emptyset$ ist. Der Nachweis beruht auf einem funktionentheoretischen Ergebnis und wird nicht erbracht (siehe hierzu Heuser (2006), Satz 96.1).

Nicht in jedem Fall lässt sich das Spektrum eines linearen Operators in so einfacher Weise wie in Beispiel 1.27 charakterisieren. Es zerfällt in mehrere qualitativ verschiedene Teile.

Definition 1.30 (Punkt-, Residual- und kontinuierliches Spektrum)

Es seien X ein Banach-Raum und $T \in \mathcal{L}(X, X)$ gegeben. Nachstehend aufgeführt zerfällt das Spektrum $\sigma(T)$ in ein *Punktspektrum* $\sigma_p(T)$, in ein *kontinuierliches Spektrum* $\sigma_c(T)$ (auch Stetigkeitsspektrum) und ein *Residualspektrum* $\sigma_r(T)$ (auch Restspektrum):

$$\sigma_p(T) := \{\lambda \in \mathbb{K} \mid \ker(\lambda I - T) \neq \{0\}\},$$

$$\sigma_c(T) := \{\lambda \in \mathbb{K} \mid \ker(\lambda I - T) = \{0\} \text{ und } \overline{R(\lambda I - T)} = X\},$$

$$\sigma_r(T) := \{\lambda \in \mathbb{K} \mid \ker(\lambda I - T) = \{0\} \text{ und } \overline{R(\lambda I - T)} \neq X\}.$$

Bemerkung 1.31

1. Es gilt die Gleichung $\sigma(T) = \sigma_p(T) \cup \sigma_c(T) \cup \sigma_r(T)$ und die einzelnen Mengen sind paarweise disjunkt.

2. $\lambda \in \rho(T) \Leftrightarrow T - \lambda I \in \mathcal{L}(X, X)$ ist bijektiv. Diese Äquivalenzaussage gilt, da aus dem Satz von Banach (Open Mapping Theorem, Satz 1.34) die Stetigkeit der Inversen $(T - \lambda I)^{-1}$ folgt.

3. Es sei $\ker(T - \lambda I) \neq \{0\}$. Nur dann gibt es ein $x \in X$, $x \neq 0$ mit $(T - \lambda I)x = 0$. Somit ist λ ein Eigenwert und sämtliche Eigenwerte liegen im Punktspektrum $\sigma_p(T)$.

4. Es sei $X := \mathbb{K}^n$. Die Eigenwerte λ können durch die Berechnung der Determinante von $(T - \lambda I)$ charakterisiert werden. Es gilt die Aussage: $\det(T - \lambda I) = 0 \Leftrightarrow \ker(T - \lambda I) \neq \{0\}$. Dies hat – wie bereits in Beispiel 1.22 ausgeführt – zur Folge: $T - \lambda I$ ist *entweder* sowohl surjektiv als auch injektiv *oder* weder surjektiv noch injektiv. Für einen endlichdimensionalen Vektorraum besteht also das kontinuierliche Spektrum und das Residualspektrum nur aus der leeren Menge: $\sigma_c(T) = \sigma_r(T) = \emptyset$. Somit kann ein solcher Operator nur reguläre Werte und Eigenwerte besitzen.

5. Die Menge $\sigma_s(T) := \sigma(T) \setminus \sigma_p(T) = \sigma_c(T) \cup \sigma_r(T)$ heißt *Streckenspektrum*. Für einen unendlichdimensionalen Banach-Raum X ist im Allgemeinen $\sigma_s(T) \neq \emptyset$. Dies ist der wesentliche Unterschied zu $\dim X < \infty$. ■

Das Beispiel 1.63 zeigt, dass das Spektrum nur aus dem Residual- bzw. aus dem kontinuierlichen Spektrum bestehen kann und dies sogar für (lineare) kompakte Operatoren (siehe Definition 1.55) zutrifft. Damit hat man auch sofort ein Beispiel dafür, dass ein linearer stetiger Operator $T \in \mathcal{L}(X, X)$ mit $\dim X = \infty$ *keine* Eigenwerte haben kann!

Das nächste (wenn auch etwas künstliche) Beispiel zeigt, dass alle Typen von Spektralpunkten tatsächlich vorkommen können.

Beispiel 1.32

Es seien die Räume

$$\mathcal{B}[0, 1] := \{x : [0, 1] \rightarrow \mathbb{R} \mid x \text{ beschränkt auf } [0, 1]\}, \quad \|x\|_{\mathcal{B}} := \sup_{t \in [0, 1]} |x(t)|,$$

$$X := \{x \in \mathcal{B}[0, 1] \mid x(0) = 0, \ x \text{ in } 0 \text{ und } 1 \text{ stetig}\}$$

gegeben. Wir betrachten den linearen stetigen Operator

$$T : X \rightarrow X, \quad (Tx)(t) := tx(t) \quad (t \in [0, 1]).$$

Man kann nun zeigen (z. B. Werner (1997), S. 211), dass für das Spektrum von T folgendes gilt: $\sigma_p(T) = (0, 1)$, $\sigma_c(T) = \{0\}$, $\sigma_r(T) = \{1\}$. ■

Wir wollen eine wichtige Aussage über lineare stetige Operatoren vorbereiten und benötigen dazu den Begriff der *offenen Abbildung*.

Definition 1.33 (Offene Abbildung)

Sind X und Y normierte Räume, so heißt $T : X \rightarrow Y$ eine *offene Abbildung*, falls gilt:

$$U \text{ offen in } X \quad \Rightarrow \quad T(U) \text{ offen in } Y.$$

Ist T bijektiv, so ist T offen genau dann, wenn T^{-1} stetig ist. Sind X, Y normierte Räume und ist $T : X \rightarrow Y$ linear, so gilt:

$$T \text{ ist offen} \quad \Leftrightarrow \quad \text{Es gibt ein } \delta > 0 \text{ mit } K_Y(0; \delta) \subset T(K_X(0; 1)).$$

Beweis: Wir wollen hier nur die Richtung „ \Leftarrow “ beweisen. Es sei U offen und $x \in U$. Man wähle ein $\varepsilon > 0$ mit $K_X(x; \varepsilon) \subset U$. Wegen $K_Y(0; \delta) \subset T(K_X(0; 1))$ folgt dann $K_Y(Tx; \varepsilon\delta) \subset T(K_X(x; \varepsilon)) \subset T(U)$, also ist $T(U)$ offen. □

Der nächste Satz, der die Begriffe surjektive und offene Abbildung in Beziehung setzt, ist in Zeidler (1995b), Theorem 3.C und Proposition 1 bzw. auch in Alt (1992), Sätze 5.7 und 5.8 bewiesen.

Satz 1.34 (Open Mapping Theorem)

Es sei $T : X \rightarrow Y$ ein linearer stetiger Operator, X und Y Banach-Räume. Dann sind die folgenden zwei Bedingungen äquivalent:

- i. T ist surjektiv.
- ii. T ist offen, d. h. T bildet offene Mengen auf offene Mengen ab.
- iii. Ist T surjektiv und auch noch injektiv, dann ist T^{-1} stetig, d. h. $T^{-1} \in \mathcal{L}(Y, X)$.

Die letzte Aussage geht auf S. Banach zurück und wird auch als *Satz über die stetige Inverse* bezeichnet. Nur diesen Teil der Aussage werden wir im Weiteren benutzen.

Für den später noch einzuführenden Begriff des adjungierten Operators ist es wichtig, zunächst den *Dualraum* eines normierten Raumes X zu erklären. Dazu benötigen wir den Begriff des *Funktional*s. Dieses ist eine Abbildung $f : X \rightarrow \mathbb{K}$. Der Dualraum X^* bezeichnet nun alle linearen stetigen Funktionale auf einem normierten Raum X : $X^* := \mathcal{L}(X, \mathbb{K})$. Der Dualraum X^* ist ein Banach-Raum und wird benötigt, um Lösbarkeitsaussagen für lineare Gleichungen allgemein zu formulieren. Grundlage hierfür sind Eigenschaften von linearen stetigen Funktionalen. Auf Beweise wird verzichtet. Sie sind ausführlich in Zeidler (1995b) dargestellt.

Satz 1.35 (Hahn-Banach Theorem)

Es sei X ein normierter Raum und L ein linearer Teilraum von X . Falls $F \in L^*$, dann existiert ein $f \in X^*$, so dass

- i. $f(y) = F(y)$ für alle $y \in L$;
- ii. $\|f\|_{X^*} = \|F\|_{L^*}$.

Aus dieser Eigenschaft kann man Folgerungen ziehen, die für das Verständnis von Aussagen, insbesondere Kapitel 4 betreffend, benötigt werden.

Korollar 1.36

Es seien X ein normierter Raum und $f \in X^*$ gegeben. Dann gilt:

- i. Zu jedem Element $x_0 \in X$, $x_0 \neq 0$, existiert ein lineares stetiges Funktional $f \in X^*$ mit der Eigenschaft

$$\|f\| = 1 \text{ und } f(x_0) = \|x_0\|.$$

- ii. Zu jedem endlichen System $\{x_1, \dots, x_n\} \subset X$ linear unabhängiger Elemente existiert ein System von Funktionalen $\{f_1, \dots, f_n\} \subset X^*$, so dass gilt:

$$f_i(x_j) = \delta_{ij}, \quad i, j = 1, \dots, n.$$

Zu der Eigenschaft ii. sagt man auch: Das System von Elementen $\{x_j\} \subset X$ ($j = 1, \dots, n$) lässt sich *biorthogonalisieren*, falls ein System von Funktionalen $\{f_i\} \subset X^*$ ($i = 1, \dots, n$) existiert, so dass die Relationen $f_i(x_j) \equiv \langle f_i, x_j \rangle = \delta_{ij}$ ($i, j = 1, \dots, n$) bestehen. Das Symbol δ_{ij} heißt *Kronecker⁹-Symbol* und ist definiert als $\delta_{ij} := 1$ (für $i = j$) und 0 (für $i \neq j$). Weitere Eigenschaften von Funktionalen sind in der Aufgabe 1.10 genannt.

⁹Leopold Kronecker (1823–1891), deutscher Mathematiker, Mitherausgeber des von Crelle begründeten Journals für Mathematik. Bekannt wurde sein Ausspruch: „Die ganzen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk.“

1.2.3 Adjungierte Operatoren

So wie einem normierten Raum X mit dem Dualraum X^* kanonisch ein zweiter normierter Raum zugeordnet wird, soll nun zu einem stetigen linearen Operator ein weiterer Operator assoziiert werden. Als ein besonders wichtiges Instrument bei der Untersuchung linearer Operatoren bzw. Operatorgleichungen hat sich der Begriff des *adjungierten Operators* erwiesen. Er wird u. a. benötigt, um die Fredholmsche Alternative in geeigneter Weise zu formulieren.

Im Folgenden seien $(X, \|\cdot\|)$, $(Y, \|\cdot\|)$ normierte Räume und $X^* = \mathcal{L}(X, \mathbb{K})$, $Y^* = \mathcal{L}(Y, \mathbb{K})$ die dazugehörigen Dualräume. Elemente aus einem Dualraum werden zur besseren Unterscheidung von jetzt ab mit einem $*$ versehen. Sie sind stets mit der Operatornorm, also $\|f^*\|_{X^*} := \sup_{\|x\|=1} |f^*(x)|$ ausgestattet und wegen der Vollständigkeit von \mathbb{K} sind X^* und Y^* Banach-Räume. Man nennt „ x, f^* “ auch ein Dualitätspaar.

Ist $X := H$ ein Hilbert-Raum und $X^* = H^*$, dann bevorzugen wir für $x^*(x)$ die Schreibweise $\langle x, x^* \rangle_H := x^*(x)$ mit $x \in H$ und $x^* \in H^*$. Der Ausdruck $\langle x, x^* \rangle_H$ bezeichnet das Skalarprodukt im Hilbert-Raum H . Die Rechtfertigung für diese Schreibweise geht auf einen Satz von Riesz zurück, der einen (schönen) Zusammenhang zwischen H und H^* liefert.

Satz 1.37 (Riesz)

Es sei H ein Hilbert-Raum und $f^* \in H^*$. Dann existiert ein eindeutig bestimmtes Element $y \in H$ mit

$$f^*(x) = \langle x, y \rangle_H, \quad \forall x \in H.$$

Die Abbildung

$$J : H \rightarrow H^*, \quad H \ni y \mapsto f^* \in H^* \quad \text{mit} \quad (Jy)(x) := \langle x, y \rangle_H$$

ist ein Norm-Isomorphismus, d. h. J ist linear, stetig, bijektiv und normerhaltend.

Mit J als Isomorphismus, $f^* = Jx_0$ und der eingeführten Schreibweise ergibt sich die Formel

$$(Jx_0)(x) = \langle x, x_0 \rangle_H.$$

Diese Überlegung erleichtert das Verständnis der folgenden

Definition 1.38 (Adjungierter Operator)

Es seien X und Y normierte Räume; X^* und Y^* ihre Dualräume und $T : X \rightarrow Y$ eine lineare stetige Abbildung. Der Operator

$$T^* : Y^* \rightarrow X^*, \quad (T^* f^*)(x) := f^*(Tx), \quad x \in X, \quad f^* \in Y^* \quad (1.18)$$

heißt der zu T adjungierte oder *duale Operator*.

Bemerkung 1.39

1. Es seien X und Y Hilbert-Räume und $T \in \mathcal{L}(X, Y)$. Dann ist der adjungierte Operator $T^* \in \mathcal{L}(Y, X)$ und charakterisiert durch die Beziehung

$$\langle Tx, y \rangle_Y = \langle x, T^*y \rangle_X \quad \text{für } x \in X, y \in Y.$$

Der Operator $T \in \mathcal{L}(X, X)$ heißt *selbstadjungiert*, falls $T^* = T$. In diesem Fall nennt man T auch *symmetrischen Operator*. Dieser Begriff macht nur für Hilbert-Räume einen Sinn, da dort aufgrund des Satzes 1.37 von Riesz zwischen den Räumen X und X^* ein Norm-Isomorphismus besteht und somit die beiden Räume aus funktionalanalytischer Sicht als „gleich“ angesehen werden können.

2. In beliebigen Banach-Räumen wird der adjungierte Operator im Allgemeinen in einem anderen Raum als dem Ausgangsraum definiert sein. In diesem Fall ist also eine Verallgemeinerung des Begriffes „selbstadjungierter Operator“ nicht möglich.

3. T^*f^* ist die Komposition (Hintereinanderausführung) zweier linearer stetiger Abbildungen und daher selbst wieder linear, stetig und es gilt $T^*f^* \in X^*$.

4. Die Abbildung $T \mapsto T^*$ von $\mathcal{L}(X, Y)$ nach $\mathcal{L}(Y^*, X^*)$ ist linear und isometrisch, d. h. $\|T\| = \|T^*\|$. Sie ist im Allgemeinen nicht surjektiv (siehe Werner (1997), Satz III.4.2). ■

Beispiel 1.40

Es seien $X := \mathbb{R}^n$ und $Y := \mathbb{R}^m$ (mit der Euklidischen Norm), $T \in \mathcal{L}(X, Y)$ eine gegebene lineare Abbildung. Bekanntlich wird T (bei gegebenen Basen) durch eine (m, n) -Matrix

$$A := (t_{ij})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$$

dargestellt. Die zu T adjungierte Abbildung T^* ist dann durch die (n, m) -Matrix

$$A^* := (t_{ij}^*)_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}} \quad \text{mit } t_{ij}^* := t_{ji}$$

gegeben, die $Y^* = Y = \mathbb{R}^m$ in $X^* = X = \mathbb{R}^n$ abbildet. A^* entsteht also aus A durch Vertauschung von Zeilen und Spalten und ist die zu A transponierte Matrix (auch mit A^T bezeichnet). ■

Es seien $T, S \in \mathcal{L}(X, Y)$ gegeben und X, Y normierte Räume. Dann gelten für den Übergang zu den adjungierten Operatoren $T^*, S^* \in \mathcal{L}(Y^*, X^*)$ die folgenden Rechenregeln:

(i) $(T + S)^* = T^* + S^*$; (ii) $(\lambda T)^* = \lambda T^*$ ($\lambda \in \mathbb{K}$); (iii) $I^* = I$;

(iv) $(T^{-1})^* = (T^*)^{-1}$ (X, Y Banach-Räume);

(v) Es seien $T \in \mathcal{L}(X, Y)$ und $S \in \mathcal{L}(Y, Z)$. Dann ist $ST \in \mathcal{L}(X, Z)$ und es gilt $(ST)^* = T^*S^*$.

(vi) Sind X und Y Banach-Räume und ist $T \in \mathcal{L}(X, Y)$, so ist T genau dann kompakt, wenn $T^* : Y^* \rightarrow X^*$ kompakt ist. (*Kompakter Operator*: siehe Abschnitt 1.2.6.)

Aus den Eigenschaften (i)-(iii) folgt speziell $(T - \lambda I)^* = T^* - \lambda I$ und für Banach-Räume X, Y aus (iv) die Beziehung

$$\rho(T) = \rho(T^*) \quad \text{und wegen (1.17)} \quad \sigma(T^*) = \sigma(T).$$

Beispiel 1.41

Es sei $k : [c, d] \times [a, b] \rightarrow \mathbb{R}$ eine stetige Funktion zweier Variabler $(s, t) \in [c, d] \times [a, b]$. Der lineare Operator

$$T : L^2[a, b] \rightarrow L^2[c, d], \quad (Tx)(s) := \int_a^b k(s, t)x(t) dt \quad (s \in [c, d], x \in L^2[a, b])$$

bildet den Hilbert-Raum $X := L^2[a, b]$ stetig in den Hilbert-Raum $Y := L^2[c, d]$ ab. Dies folgt aus

$$\begin{aligned} \|Tx\|_{L^2[c, d]}^2 &= \int_c^d [(Tx)(s)]^2 ds = \int_c^d \left(\int_a^b k(s, t)x(t) dt \right)^2 ds \\ &\stackrel{(1.4)}{\leq} \int_c^d \left[\left(\int_a^b [k(s, t)]^2 dt \right) \cdot \left(\int_a^b [x(t)]^2 dt \right) \right] ds \\ &= \left(\int_c^d \int_a^b [k(s, t)]^2 dt ds \right) \cdot \|x\|_{L^2[a, b]}^2 \\ &= C^2 \cdot \|x\|_{L^2[a, b]}^2, \quad C^2 := \left(\int_c^d \int_a^b [k(s, t)]^2 dt ds \right). \end{aligned}$$

Es wurde u. a. die Cauchy-Schwarzsche-Ungleichung (1.4) für den $L^2[a, b]$ mit der Norm $\|x\|^2 := \int_a^b [x(t)]^2 dt$ verwendet.

Für die Herleitung des adjungierten Operators T^* benötigt man den *Satz von Fubini* (siehe z. B. Hoffmann et al. (2006), S. 678), der eine Aussage über die Vertauschbarkeit der Integrationsreihenfolge macht. Der zu T adjungierte Operator T^* berechnet sich aus der Definitionsgleichung

$$\begin{aligned} \langle Tx, y \rangle_{L^2[c, d]} &= \int_c^d (Tx)(s) \cdot y(s) ds = \int_c^d \left(\int_a^b k(s, t)x(t) dt \right) \cdot y(s) ds \\ &= \int_c^d \int_a^b k(s, t)x(t)y(s) dt ds \stackrel{\text{(Fubini)}}{=} \int_a^b \int_c^d k(s, t)x(t)y(s) ds dt \\ &= \int_a^b x(t) \underbrace{\left(\int_c^d k(s, t)y(s) ds \right)}_{=: (T^*y)(t)} dt = \int_a^b x(s) \cdot (T^*y)(s) ds \\ &= \langle x, T^*y \rangle_{L^2[a, b]}. \end{aligned}$$

Der zu T adjungierte Operator T^* ist dann durch den Ausdruck

$$(T^*y)(s) = \int_c^d k(t, s)y(t) dt \quad (s \in [a, b])$$

gegeben und bildet den Raum $Y^* = L^2[c, d] = Y$ in den Raum $X^* = L^2[a, b] = X$ ab. (Man beachte, dass lineare stetige Funktionale auf L^2 und Elemente von L^2 nach dem Satz von Riesz 1.37 identifiziert werden können.) Der Integrand $k(t, s)$ heißt der zu $k(s, t)$ *transponierte Kern*.

Setzt man den Kern k auf dem Rechteck $[c, d] \times [a, b]$ als quadratisch integrable Funktion voraus, der darüber hinaus auch bzgl. der einzelnen Variablen quadratisch integrierbar ist, dann ergibt sich nach analogen Umformungen dasselbe Resultat.

Ist der Kern k *symmetrisch*, d. h. gilt $k(s, t) = k(t, s)$ auf $[a, b] \times [a, b]$, dann ist der Operator T selbstadjungiert bzw. symmetrisch. In diesem Fall gilt $T = T^*$. ■

Der Nullraum einer Abbildung T und ihr Wertevorrat stehen in einem interessanten und für spätere Anwendungen wichtigen Zusammenhang zum adjungierten Operator T^* . Dazu führen wir zuvor *zwei Notationen* ein: Falls $M \subset X$, dann ist

$$M^\perp := \{f^* \in X^* \mid x \in M \Rightarrow f^*(x) = 0\} \subset X^*.$$

Falls $N \subset X^*$, dann ist

$$N_\perp := \{x \in X \mid f^* \in N \Rightarrow f^*(x) = 0\} \subset X.$$

Der Leser möge dieses Symbol mit dem Begriff des orthogonalen Komplements, wie es im Hilbert-Raum erklärt ist, vergleichen. Für lineares T sind $\ker(T)$ und $R(T)$ stets Untervektorräume; für stetiges T ist $\ker(T) = T^{(-1)}(0)$ stets abgeschlossen. Hingegen braucht $R(T)$ nicht abgeschlossen zu sein (siehe Beispiel 1.52). Es gilt jedoch der folgende Satz, für dessen Beweis wird auf Drábek und Milota (2007), Prop. 2.1.27, bzw. auch Werner (1997), Satz III.4.5. verwiesen.

Satz 1.42 (Closed-Range-Theorem)

Es seien X, Y Banach-Räume und $T \in \mathcal{L}(X, Y)$, $T^* \in \mathcal{L}(Y^*, X^*)$ gegeben. Dann sind die folgenden Bedingungen äquivalent:

$$\begin{array}{ll} \text{Closed Range} & R(T) \text{ ist abgeschlossen.} \\ \text{Fredholmsche Alternative} & R(T) = (\ker(T^*))_\perp \Leftrightarrow R(T^*) = (\ker(T))^\perp. \end{array}$$

Mit $R(T)$ ist auch $R(T^*)$ abgeschlossen. Man beachte, dass die Bedingung $\overline{R(T)} = (\ker(T^*))_\perp$ aus dem Satz keine hinreichende Bedingung für die Lösbarkeit der Gleichung

$$Tx = b, \quad x \in X, b \in Y \tag{1.19}$$

darstellt, da nur die Abschließung von $R(T)$ charakterisiert wird.

Die Äquivalenz $R(T) = (\ker(T^*))_\perp \Leftrightarrow R(T^*) = (\ker(T))^\perp$ bedeutet für die Operatorgleichung (1.19) und ihre adjungierte Gleichung

$$T^* x^* = b^*, \quad x^* \in Y^*, b^* \in X^* : \tag{1.20}$$

- i. Für gegebenes $b \in Y$ hat die Originalgleichung (1.19) eine Lösung, genau dann wenn die Bedingung

$$x^*(b) = 0 \tag{1.21}$$

für alle Lösungen x^* der adjungierten homogenen Gleichung (1.20) erfüllt ist.

- ii. Umgekehrt gilt für gegebenes $b^* \in X^*$, dass die adjungierte Gleichung (1.20) eine Lösung genau dann hat, wenn die Bedingung

$$b^*(x) = 0 \tag{1.22}$$

für alle Lösungen x der homogenen Originalgleichung (1.19) erfüllt ist.

In der Tat, falls $Tx = b$ und $T^*x^* = 0$ gelten, dann folgt

$$x^*(b) = x^*(Tx) = (T^*x^*)(x) = 0.$$

Folglich repräsentiert (1.21) eine einfache notwendige Lösbarkeitsbedingung für (1.19). Für einen abgeschlossenen Wertebereich $R(T)$ ist diese Bedingung auch hinreichend. Des Weiteren folgt mit $T^*x^* = b^*$ und $Tx = 0$ die Bedingung

$$b^*(x) = (T^*x^*)(x) = x^*(Tx) = 0.$$

Dies ist die Lösbarkeitsbedingung (1.22) für die adjungierte Gleichung (1.20).

Falls X und Y endlichdimensionale Räume sind, dann ist $R(T)$ automatisch abgeschlossen. Die Aussagen i. und ii. stimmen mit den klassischen Resultaten für endliche lineare Gleichungssysteme überein.

Einerseits gibt es Operatoren, deren Wertevorrat nicht abgeschlossen ist. Das Beispiel 1.52 untersucht diesen Sachverhalt. Mitunter ist es nicht schwer zu entscheiden, ob ein Operator einen abgeschlossenen Wertevorrat hat oder nicht (siehe Aufgabe 2.11).

Andererseits gibt es eine (große) Klasse von Operatoren, die einen abgeschlossenen Wertebereich haben, nämlich die so genannten *Fredholm-Operatoren*. Diesen werden wir uns im Unterabschnitt 1.2.5 zuwenden.

1.2.4 Direkte Summe und Projektoren

Sind $(X, \|\cdot\|_X)$ und $(Y, \|\cdot\|_Y)$ zwei normierte Räume, so ist der Produktraum $X \times Y$ wie üblich als Menge aller Paare (x, y) mit $x \in X$ und $y \in Y$ definiert. Als Norm auf $X \times Y$ kann man etwa

$$\|(x, y)\|_{X \times Y} := \|x\|_X + \|y\|_Y \quad \text{oder} \quad \|(x, y)\|_\infty := \max\{\|x\|_X, \|y\|_Y\}$$

wählen. Sind X und Y Banach-Räume, so ist $X \times Y$ mit jeder dieser Norm wieder ein Banach-Raum.

Sind speziell X_1 und X_2 zwei Unterräume eines normierten Raumes X so betrachtet man auch statt des Produktes $X_1 \times X_2$ den Summenraum

$$X_1 + X_2 := \{x_1 + x_2 \mid x_1 \in X_1, x_2 \in X_2\}.$$

Im Fall $X_1 \cap X_2 = \{0\}$ schreibt man $X_1 \oplus X_2$ statt nur $X_1 + X_2$ und nennt dies die *direkte Summe* von X_1 und X_2 . Der Zusammenhang zu dem *Produktraum* $X_1 \times X_2$ kann durch die Zuordnung

$$\begin{aligned} X_1 \times X_2 &\rightarrow (X_1 \times \{0\}) \oplus (\{0\} \times X_2) \\ (x_1, x_2) &\mapsto (x_1, 0) + (0, x_2) \end{aligned}$$

hergestellt werden.

Darauf aufbauend wollen wir eine besonders einfache Klasse von linearen Operatoren untersuchen, die wir auch *Projektor* bzw. *Projektionsoperator* nennen. Dies sind lineare Operatoren $P : X \rightarrow X$, für die $P^2 = P$ gilt. Mit P ist dann auch $Q := I - P$ eine Projektion. Hierbei gilt $\ker(P) = R(Q)$ und $R(P) = \ker(Q)$. Die Darstellung $x = (x - Px) + (Px)$ liefert zu jeder Projektion P eine Zerlegung von X als direkte Summe $X = \ker(P) \oplus R(P)$. Dies sieht man leicht: Aus $y \in \ker(P) \cap R(P)$ folgt $P y = 0$ und $y = P x$ für ein $x \in X$, also $0 = P y = P^2 x = P x = y$.

Diese Argumentation kann man auch umkehren. Ist nämlich $X = X_1 \oplus X_2$ mit zwei Unterräumen $X_1, X_2 \subset X$, so definiert man einen Projektionsoperator P wie folgt: Ist $x = x_1 + x_2$ mit $x_1 \in X_1$ und $x_2 \in X_2$, so setzt man $P x := x_2$. Die Abbildung $P : X \rightarrow X$ ist offensichtlich linear und erfüllt $X_1 = \ker(P)$ und $X_2 = R(P)$, sowie $P^2 = P$. Man drückt diesen Sachverhalt durch die folgende Sprechweise aus: P ist eine *Projektion von X auf X_2 längs X_1* . Offensichtlich ist dann $Q := I - P$ die Projektion von X auf X_1 längs X_2 .

Beispiel 1.43

1. Es sei $X = \mathbb{R}^2 = X_1 \oplus X_2$, wobei X_1 und X_2 zwei Geraden sind (siehe Abbildung 1.5). Die Projektion $P : X \rightarrow X$ auf X_1 entlang X_2 entspricht der gewöhnlichen Parallelprojektion auf X_1 parallel zu X_2 . Da P stetig ist, repräsentiert $\mathbb{R}^2 = X_1 \oplus X_2$ eine direkte Summe. Sowohl $\ker(P)$ als auch $R(P)$ sind abgeschlossene Unterräume von \mathbb{R}^2 .
2. Wir betrachten im Folgenraum $X := l^p$ ($1 \leq p < \infty$) für jedes $n \in \mathbb{N}$ den linearen Operator

$$P_n : l^p \rightarrow l^p, \quad P_n(\xi_1, \xi_2, \xi_3, \dots) := (\xi_1, \dots, \xi_n, 0, \dots).$$

Man rechnet leicht nach, dass alle P_n Projektionsoperatoren und $\ker(P_n)$ und $R(P_n)$ abgeschlossene Unterräume von l^p sind. ■

Voran stehendes Beispiel lässt vermuten, dass es zwischen der Beschränktheit der Projektion P und den jeweiligen Unterräumen einen Zusammenhang gibt.

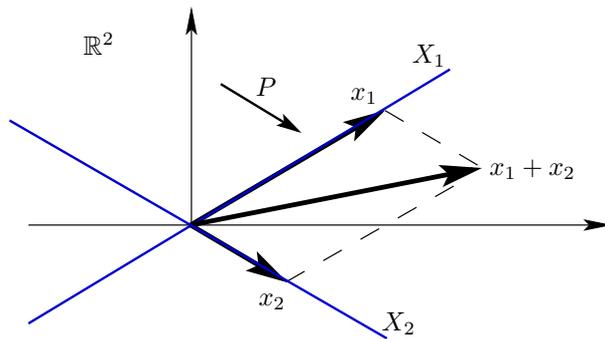


Abb. 1.5 Direkte Summe $\mathbb{R}^2 = X_1 \oplus X_2$

Satz 1.44

Es sei $X := X_1 \oplus X_2$ ein Banach-Raum und P die Projektion von X auf X_2 längs X_1 . Genau dann ist P beschränkt, wenn X_1 und X_2 beide abgeschlossen in X sind.

Nach diesem Satz gibt es also bei gegebenem Banach-Raum X und abgeschlossenem Unterraum $X_1 \subset X$ immer genau dann eine stetige Projektion von X auf X_1 , wenn ein weiterer abgeschlossener Unterraum $X_2 \subset X$ mit $X = X_1 \oplus X_2$ existiert.

Nicht jeder Unterraum eines Banach-Raumes X kann als Wertebereich eines gewissen stetigen Projektionsoperators dargestellt werden (siehe Dunford und Schwartz (1964), Bd. I). Im Hilbert-Raum kann man auf jeden abgeschlossenen Unterraum projizieren. Insbesondere kann man stets auf endlichdimensionale Unterräume $X_1 \subset X$ stetig projizieren. Somit gibt es dann auch immer einen abgeschlossenen Unterraum $X_2 \subset X$ mit $X = X_1 \oplus X_2$ (nämlich $X_2 := \ker(P)$). Einen kurzen Beweis für die letzte Behauptung findet man in Appell und Väth (2005).

Nun können wir die *Kodimension* eines linearen Teilraumes X_1 von X erklären. Der Begriff ist ein rein algebraischer Begriff, der nicht von einer Norm auf X abhängt. Wir passen die Definition dieses Begriffes an die weitere Vorgehensweise an, indem wir im Folgenden stets davon ausgehen, dass ein normierter Raum X vorliegt und dieser in eine direkte Summe von abgeschlossenen Unterräumen $X = X_1 \oplus X_2$ aufgesplittet werden kann. Die *Kodimension von X_1* ist die Dimension eines beliebigen Unterraumes X_2 von X , der die Bedingung $X = X_1 \oplus X_2$ erfüllt. Es ist dann $\text{codim } X_1 = \dim X_2$ und außerdem gilt für $\dim X < \infty$: $\text{codim } X_1 = \dim X - \dim X_1$.

Beispiel 1.45 (Kodimension)

Für einen m -dimensionalen linearen Unterraum X_1 von \mathbb{R}^n , $n \geq 1$, ergibt sich

$$\text{codim } X_1 = n - m.$$

Falls X_1 eine Ebene durch den Ursprung im \mathbb{R}^3 ist, dann gilt

$$\dim X_1 = 2 \quad \text{und} \quad \text{codim } X_1 = 1.$$



1.2.5 Fredholm-Operatoren

In Anwendung der Begriffsbildungen aus Abschnitt 1.2.4 beschäftigen wir uns mit der linearen Operatorgleichung

$$Tx = b, \quad x \in X. \quad (1.23)$$

Satz 1.46

Es seien der Operator $T : X \rightarrow Y$ linear und X, Y lineare Räume über \mathbb{K} . Es sei X_1 ein festes algebraisches Komplement des Nullraumes $\ker(T)$, d. h. X_1 ist linearer Teilraum von X , so dass

$$X = \ker(T) \oplus X_1. \quad (1.24)$$

Dann sind die folgenden Behauptungen erfüllt:

i. Die Einschränkung

$$T : X_1 \rightarrow R(T) \quad (1.25)$$

ist linear und bijektiv. Folglich gilt

$$\operatorname{codim} \ker(T) = \dim R(T). \quad (1.26)$$

ii. Zusätzlich wird vorausgesetzt, dass X und Y Banach-Räume, X_1 und $R(T)$ abgeschlossen und $T : X \rightarrow Y$ stetig sind. Dann ist der Operator (1.25) ein linearer Homöomorphismus.

Wir erinnern noch einmal daran, dass $R(T) = T(X)$ und die Zahl $\dim R(T)$ der Rang von T genannt wird. Symbol: $\operatorname{rang}(T) := \dim R(T)$.

Beweis: Zu i.: Die Abbildung 1.25 ist definitionsgemäß surjektiv. Aus $Tx = 0$ mit $x \in X_1$ folgt $x \in \ker(T) \cap X_1$ und somit $x = 0$, folglich injektiv.

Zu ii.: Dies folgt aus dem Open Mapping Theorem 1.34. \square

Eine besonders einfache Situation ergibt sich für $\dim X < \infty$ und $\dim Y < \infty$. Es sei $S : X_1 \rightarrow R(T)$ die Einschränkung des Operators $T : X \rightarrow R(T)$ auf den linearen Unterraum X_1 von X . Dann ist für gegebenes $b \in R(T)$ die Lösungsmenge der Originalgleichung (1.23) gegeben durch

$$S^{-1}b + \ker(T) := \{S^{-1}b + x \mid x \in \ker(T)\},$$

wobei $\dim \ker(T) = \dim X - \operatorname{rang}(T)$ ist.

Wie bereits in Bemerkung 1.31 erwähnt, besitzt in endlichdimensionalen Räumen der lineare Operator $T : X \rightarrow X$ nur Eigenwerte und reguläre Werte. Dies bedeutet für die Gleichung $Tx - \lambda x = y$ folgendes Lösungsverhalten:

- *Entweder*: λ ist regulärer Wert von T und die Gleichung $Tx - \lambda x = y$ besitzt damit für jedes y genau eine Lösung $x \in X$.
- *Oder*: λ ist Eigenwert von T und die Gleichung $Tx - \lambda x = y$ besitzt nur für gewisse $y \in X$ eine Lösung und dann gibt es immer mehrere (unendlich viele) Lösungen.

Diese Alternative wird *Fredholmsche Alternative* genannt. In der Regel erfolgt diese Bezeichnung meist im Zusammenhang mit einem unendlichdimensionalen Raum. Wir wollen für eine Klasse von Operatoren ein ähnliches Resultat für unendlichdimensionale Räume gewinnen. Zunächst sei an einem einfachen Beispiel gezeigt, dass in unendlichdimensionalen Räumen ohne weitere Voraussetzungen an den Operator T die Fredholmsche Alternative nicht zwingend vorliegen muss.

Beispiel 1.47

Es seien $X := l^2$ und der Operator

$$T : l^2 \rightarrow l^2, \quad T(x_1, x_2, x_3, \dots) := (x_2, x_3, x_4, \dots)$$

als die so genannte *Linksverschiebung* gegeben. Die Gleichung $Tx = y$ hat für jedes $y = (y_1, y_2, y_3, \dots) \in X$ eine Lösung, diese ist jedoch nicht eindeutig bestimmt. Jedes $x = (a, y_1, y_2, y_3, \dots)$ mit beliebigem $a \in \mathbb{R}$ ist Lösung. Darüber hinaus ist $\lambda = 0$ ein Eigenwert, da $x = (a, 0, 0, 0, \dots)$ mit $a \in \mathbb{R} \setminus \{0\}$ im Nullraum von T liegt, d. h. es ist die Gleichung $Tx = 0$ erfüllt.

Eine entsprechende Überlegung im Raum $X := l^2$ kann mit der *Rechtsverschiebung*

$$T : l^2 \rightarrow l^2, \quad T(x_1, x_2, x_3, \dots) := (0, x_1, x_2, \dots)$$

durchgeführt werden. Die Gleichung $Tx = y$ hat nur für gewisse $y = (y_1, y_2, y_3, \dots) \in X$ (nämlich solche mit $y_1 = 0$) eine Lösung $x = (y_2, y_3, y_4, \dots)$, diese ist jedoch eindeutig bestimmt. Der Wert $\lambda = 0$ gehört zu $\sigma_r(T)$, da $R(T) = \{y = (y_1, y_2, y_3, \dots) \mid y_1 = 0, y_2 = x_1, y_3 = x_2, \dots\}$ nicht dicht in X liegt. ■

Im Weiteren betrachten wir wieder lineare Operatorgleichungen

$$Tx = y, \quad x \in X \tag{1.27}$$

für gegebene $y \in Y$ und die zugehörige duale Gleichung

$$T^*x^* = y^*, \quad x^* \in Y^* \tag{1.28}$$

für gegebene $y^* \in X^*$. Wir suchen eine Klasse linearer Operatoren T , für die die bekannten Lösbarkeitseigenschaften klassischer linearer Gleichungssysteme mit $X := \mathbb{R}^n$ und $Y := \mathbb{R}^m$ soweit wie möglich erhalten bleiben. Diese Klasse wird durch die *Fredholm-Operatoren* geschaffen. Die wichtigste Aussage über die Lösung linearer Operatorgleichungen ist die Fredholmsche Alternative, die von *Fredholm*¹⁰ zuerst für eine nach ihm benannte Klasse von linearen Integralgleichungen gefunden wurde. Wir geben zunächst den Begriff des Fredholm-Operators an.

Definition 1.48 (Fredholm-Operator)

Es seien X und Y Banach-Räume über \mathbb{K} . Der Operator $T \in \mathcal{L}(X, Y)$ heißt *Fredholm-Operator* (kurz: F-Operator), falls $\dim \ker(T)$ und $\operatorname{codim} R(T)$ endlich sind. Die ganze Zahl

$$\operatorname{ind}(T) := \dim \ker(T) - \operatorname{codim} R(T)$$

heißt der Index von T .

Jeder Fredholm-Operator $T \in \mathcal{L}(X, Y)$ mit Index Null erfüllt die Fredholmsche Alternative, denn T ist genau dann injektiv ($\dim \ker(T) = 0$), wenn T surjektiv ist ($\operatorname{codim} R(T) = 0$).

Die Endlichkeit der *Kodimension* des Bildes von T bedeutet, dass der Raum Y die Darstellung $Y = R(T) \oplus Y_0$ mit einem endlichdimensionalen Unterraum $Y_0 \subset Y$ hat. Es gilt dann: $\operatorname{codim} R(T) := \dim Y_0$. Jedes Element $y \in Y$ hat die einzige Darstellung $y = y_{R(T)} + y_{Y_0}$ mit $y_{R(T)} \in R(T)$ und $y_{Y_0} \in Y_0$. Allgemein gilt die folgende Aussage: Gibt es einen abgeschlossenen Unterraum $Y_1 \subset Y$ mit $Y = R(T) \oplus Y_1$, so ist $R(T)$ abgeschlossen in Y (siehe Appell und Väth (2005), Satz 7.6 oder auch Zeidler (1986), Proposition 8.14).

Den Prototyp eines Fredholm-Operators haben wir bereits (unbewusst) in der Theorie der linearen Gleichungssysteme kennen gelernt, wie nachfolgendes Beispiel zeigt.

Beispiel 1.49

Für $X := \mathbb{K}^n$ und $Y := \mathbb{K}^m$ ist jeder lineare Operator $T : X \rightarrow Y$ ein Fredholm-Operator mit Index $\operatorname{ind}(T) = n - m$. Dies sieht man wie folgt: Es sei $r := \operatorname{rang}(T)$. Um den Index zu berechnen, schreiben wir

$$X = \ker(T) \oplus X_1, \quad Y = R(T) \oplus Y_1.$$

Der Operator $\hat{T} := T|_{X_1}$ mit $\hat{T} : X_1 \rightarrow R(T)$ ist bijektiv. Somit ist $\dim X_1 = r$ und es gilt

$$\dim \ker(T) = \dim X - r, \quad \operatorname{codim} R(T) = \dim Y_1 = \dim Y - r.$$

¹⁰Erik Ivar Fredholm (1866–1927), schwedischer Mathematiker. Er begründete 1903 die moderne Theorie der Integralgleichungen, auch Fredholm-Theorie genannt.

Hieraus folgt $\text{ind}(T) = \dim \ker(T) - \text{codim } R(T) = \dim X - \dim Y = n - m$. ■

Auch in unendlichdimensionalen Räumen gibt es Fredholm-Operatoren mit sowohl positivem als auch negativem Index.

Beispiel 1.50

Es sei wie in Beispiel 1.47

$$X := Y := l^2 \quad \text{und} \quad T(x_1, x_2, x_3, \dots) := (x_2, x_3, x_4, \dots)$$

die Linksverschiebung. Da T surjektiv ist, folgt $\text{codim } R(T) = 0$. Andererseits ist $\dim \ker T = 1$ und somit $\text{ind}(T) = 1$. Man überlegt sich leicht, dass $\text{ind}(T^n) = n$ ($n \in \mathbb{N}$) gilt. ■

Beispiel 1.51

Es sei wie in Beispiel 1.47

$$X := Y := l^2 \quad \text{und} \quad T(x_1, x_2, x_3, \dots) := (0, x_1, x_2, \dots)$$

die Rechtsverschiebung. Da T injektiv ist, folgt $\dim \ker(T) = 0$. Andererseits ist $\text{codim } R(T) = 1$ und somit $\text{ind}(T) = -1$. Man überlegt sich leicht, dass $\text{ind}(T^n) = -n$ ($n \in \mathbb{N}$) gilt. ■

Wir betrachten ein Beispiel eines linearen stetigen Operators, der kein Fredholm-Operator ist.

Beispiel 1.52

Es seien $X = Y := C^0[-1, 1]$ und beide Räume mit der Max-Norm ausgestattet. Der Operator T sei definiert durch

$$T : X \rightarrow Y, \quad (Tx)(s) := \int_{-1}^s x(t) dt \quad (s \in [-1, 1]).$$

Man rechnet leicht nach, dass $\|T\| = 2$ gilt. Aus $(Tx)(s) \equiv 0$ folgt sofort $x(s) \equiv 0$, also ist T injektiv. Nach dem *Hauptsatz der Differenzial- und Integralrechnung* bildet T den Raum $C^0[-1, 1]$ tatsächlich in den Raum $C^1[-1, 1]$ ab. Der Wertevorrat von T ist

$$R(T) = \{y \in C^1[-1, 1] \mid y(-1) = 0\}.$$

Damit ist klar, dass T nicht surjektiv ist. Wir überlegen uns, dass $R(T)$ nicht abgeschlossen ist. Dazu ist zu zeigen, dass es eine Folge (y_n) mit $(y_n) \subset R(T)$, $y_n \rightarrow y \in C^0[-1, 1]$ und $y \notin R(T)$ gibt. Wir betrachten

$$x_n(t) := \begin{cases} 0, & -1 \leq t < 0 \\ nt, & 0 \leq t < 1/n \\ 1, & 1/n \leq t \leq 1. \end{cases}$$

Damit bekommen wir eine Bildfolge (y_n) der Form

$$y_n(t) := (Tx_n)(t) = \begin{cases} 0, & -1 \leq t < 0 \\ \frac{1}{2}nt^2, & 0 \leq t < 1/n \\ t - 1/(2n), & 1/n \leq t \leq 1, \end{cases}$$

die in $R(T)$ liegt. Man sieht sofort, dass (y_n) punktweise konvergiert:

$$y(t) := \begin{cases} 0, & -1 \leq t < 0 \\ t, & 0 \leq t \leq 1. \end{cases}$$

Die Grenzfunktion y liegt in $C^0[-1, 1]$ und es gilt $\lim_{n \rightarrow \infty} \max_{t \in [-1, 1]} |y_n(t) - y(t)| = 0$. Dies folgt aus der Abschätzung

$$\begin{aligned} \|y_n - y\|_{C^0} &\leq \max_{t \in [-1, 0]} |0| + \max_{t \in [0, 1/n]} \left| \frac{1}{2}nt^2 - t \right| + \max_{t \in [1/n, 1]} \left| t - \frac{1}{2n} - t \right| \\ &\leq 0 + \frac{1}{2n} + \frac{1}{2n} = \frac{1}{n}. \end{aligned}$$

Somit konvergiert (y_n) sogar gleichmäßig gegen y , jedoch ist $y \notin R(T)$, da es in $t = 0$ nicht differenzierbar ist, d. h. $R(T)$ ist *nicht* abgeschlossen. Nun hat jeder Komplementärraum zu $R(T)$ die Form $Y_1 := \{y \in C^0[-1, 1] \mid y(-1) \neq 0\}$. Beispielsweise ist das Funktionensystem (g_n) , $g_n := \cos(2n\pi t)$ in dem Komplementärraum Y_1 enthalten und darüber hinaus linear unabhängig. Damit gilt bereits $\dim Y_1 = \infty$. Demzufolge hat $R(T)$ keine endliche Kodimension. ■

Der folgende Satz fasst wichtige Eigenschaften von Fredholm-Operatoren zusammen:

Satz 1.53 (Eigenschaften von Fredholm-Operatoren)

Es seien X und Y Banach-Räume und $T \in \mathcal{L}(X, Y)$. Für einen Fredholm-Operator T gelten die folgenden Aussagen:

- i. Falls $\text{ind}(T) = 0$ und $\ker(T) = \{0\}$, dann hat die Gleichung (1.27) genau eine Lösung für jedes $y \in Y$ und es ist $T^{-1} \in \mathcal{L}(Y, X)$.
- ii. Der Wertebereich von T ist abgeschlossen. Für festes $y \in Y$ hat die Gleichung (1.27) eine Lösung genau dann, wenn $x^*(y) = 0$ für alle $x^* \in \ker(T^*)$ gilt.
- iii. Falls $S \in \mathcal{K}(X, Y)$, dann ist $T + S$ auch ein Fredholm-Operator und es gilt

$$\text{ind}(T + S) = \text{ind}(T).$$

- iv. Der duale Operator T^* ist ebenfalls ein Fredholm-Operator und es ist

$$\begin{aligned} \dim \ker(T^*) &= \text{codim } R(T), & \text{codim } R(T^*) &= \dim \ker(T), \\ \text{ind}(T^*) &= -\text{ind}(T). \end{aligned}$$

Die duale Gleichung (1.28) hat eine Lösung für festes $y^* \in X^*$, genau dann, wenn $y^*(x) = 0$ für alle $x \in \ker(T)$.

Die Menge $\mathcal{K}(X, Y)$ in iii. bezeichnet die Klasse der linearen kompakten Operatoren. Diese werden in Abschnitt 1.2.6 ausführlich diskutiert.

Beweis: Zu i.: Der Operator $T : X \rightarrow Y$ ist wegen $\ker(T) = \{0\}$ injektiv und $\operatorname{ind}(T) = 0$ impliziert $\operatorname{codim}(R(T)) = 0$, d. h. $R(T) = Y$. Damit ist T auch surjektiv. Die Stetigkeit von T^{-1} folgt nun aus dem Open Mapping Theorem (Satz 1.34).

Zu ii.: Diese Aussage gilt bereits, wenn $R(T)$ lediglich abgeschlossen ist. Wir zeigen zuerst die Notwendigkeit der Bedingung. Es sei also $Tx = y$ lösbar (etwa durch $x = x_0$) und $T^*x^* = 0$. Aus Gründen der übersichtlicheren Darstellung setzen wir $\langle x^*, y \rangle := x^*(y)$. Dann ist $\langle x^*, y \rangle = \langle x^*, Tx_0 \rangle = \langle T^*x^*, x_0 \rangle = \langle 0, x_0 \rangle = 0$. Damit ist $\langle x^*, y \rangle = 0$ eine notwendige Lösbarkeitsbedingung.

Nun gelte $\langle x^*, y \rangle = 0$ für alle $x^* \in \ker(T^*)$. Wäre $Tx = y$ nicht auflösbar, läge also y nicht im (abgeschlossenen) Unterraum $T(X)$, so gäbe es nach Korollar 1.36 ein x^* in Y^* mit

$$\langle T^*x^*, x \rangle = \langle x^*, Tx \rangle = \langle x^*, y \rangle = 0, \quad \forall x \in X, \quad \text{aber } \langle x^*, y \rangle = 1.$$

Aus $\langle T^*x^*, x \rangle = 0$ für alle $x \in X$ folgt $T^*x^* = 0$. Wegen unserer Voraussetzung ist $\langle x^*, y \rangle = 0$ im Widerspruch zu $\langle x^*, y \rangle = 1$. Die Gleichung $Tx = y$ muss also doch auflösbar sein.

Für die weiteren Beweise konsultiere man Zeidler (1986), S. 366–367. □

Beispiel 1.54

Mit Hilfe des Index eines Fredholm-Operators kann man Eigenschaften einer Abbildung charakterisieren: Es sei $T : X \rightarrow Y$ ein linearer Fredholm-Operator, X und Y normierte Räume über \mathbb{K} . Dann gilt:

- i. T surjektiv $\Leftrightarrow \operatorname{ind}(T) = \dim \ker(T)$.
- ii. T injektiv $\Leftrightarrow \dim \ker(T) = 0$ (d. h. $\operatorname{ind}(T) = -\operatorname{codim} R(T)$).
- iii. T bijektiv $\Leftrightarrow \operatorname{ind}(T) = \dim \ker(T) = 0$.
- iv. Falls X und Y Banach-Räume sind, dann hat die Gleichung $Tx = b$, $x \in X$, für jedes $b \in Y$ genau eine Lösung, genau dann wenn $\operatorname{ind}(T) = \dim \ker(T) = 0$ gilt.

Beweis: Zu i.: T ist surjektiv, genau dann wenn $\operatorname{codim} R(T) = 0$.

Zu ii., iii.: Dies ist klar.

Zu iv.: Man wende das Open Mapping Theorem an (siehe Satz 1.34). ■

1.2.6 Kompakte Operatoren

Weitere Eigenschaften über Fredholm-Operatoren lassen sich beweisen, wenn man den Begriff des linearen *kompakten Operators* einbezieht. Im Zusammenhang mit Eigenwertgleichungen der Form $(\lambda I - T)x = y$ ist T stets aus $\mathcal{L}(X, X)$. Solche Operatoren können sich als spezielle Fredholm-Operatoren erweisen. Wenn man darüber hinaus für lineare

stetige Operatoren den Begriff Stetigkeit durch Kompaktheit ersetzt, dann lassen sich Ergebnisse über lineare stetige Operatoren von dem \mathbb{R}^n auf unendlichdimensionale Räume in geeigneter Weise übertragen.

Während lineare stetige Operatoren beschränkte Mengen in beschränkte Mengen überführen, sind kompakte Operatoren durch die folgende Eigenschaft gekennzeichnet:

Definition 1.55 (Kompakter Operator)

Es sei $T : X \rightarrow Y$ ein Operator und X sowie Y normierte Räume. T heißt kompakt, wenn für jede beschränkte Menge $M \subset X$ das Bild $T(M) \subset Y$ relativ kompakt ist. Die Menge aller *linearen kompakten Operatoren* von X in Y bezeichnen wir mit

$$\mathcal{K}(X, Y) := \{T : X \rightarrow Y \mid T \text{ linear und kompakt}\}. \quad (1.29)$$

Kompakte Operatoren überführen also beschränkte Mengen in relativ kompakte Mengen, womit stets $\mathcal{K}(X, Y) \subset \mathcal{L}(X, Y)$ gilt. Diese Mengeninklusion ist klar, da jede präkompakte Menge beschränkt ist. Darüber hinaus gilt sogar für einen endlichdimensionalen normierten Raum Y die Beziehung $\mathcal{K}(X, Y) = \mathcal{L}(X, Y)$. Insbesondere ist in Beispiel 1.49 wegen $X := \mathbb{K}^n$, $Y := \mathbb{K}^m$ der Operator T kompakt. Allerdings folgt $\mathcal{K}(X, X) \neq \mathcal{L}(X, X)$, falls X unendlichdimensional ist. Wäre $I_X \in \mathcal{K}(X, X)$, dann müsste die Einheitskugel $K_X[0; 1]$ relativ kompakt sein. Es gilt jedoch: Die Einheitskugel $K_X[0; 1]$ ist genau für endlichdimensionales X relativ kompakt. Damit ist klar: *In jedem unendlichdimensionalen Raum gibt es lineare beschränkte Operatoren, die nicht kompakt sind.* Die Beziehungen zwischen linearen kompakten bzw. linearen stetigen Operatoren T sind in folgendem Schema zusammengefasst:

$$T \text{ kompakt} \quad \begin{array}{c} \Rightarrow \\ \not\Leftarrow \end{array} \quad T \text{ beschränkt} \quad \Leftrightarrow \quad T \text{ stetig}.$$

Bemerkung 1.56

1. Da jeder lineare Operator insbesondere homogen ist, genügt es für die Kompaktheit eines Operators zu zeigen, dass er die Einheitskugel $K_X[0; 1]$ in eine präkompakte Teilmenge von Y überführt.

2. Für $T_1 \in \mathcal{L}(X, Y)$ und $T_2 \in \mathcal{L}(Y, Z)$ gilt:

$$T_1 \text{ oder } T_2 \text{ kompakt} \quad \Rightarrow \quad T_2 T_1 \text{ kompakt}.$$

3. Ist Y ein Banach-Raum, so ist $\mathcal{K}(X, Y)$ ein abgeschlossener Unterraum von $\mathcal{L}(X, Y)$ (siehe Appell und Vöth (2005), Lemma 5.2). Insbesondere gilt dann:

$$\text{Aus } (T_n)_{n \in \mathbb{N}} \subset \mathcal{K}(X, Y) \text{ und } T_n \rightarrow T \text{ folgt } T \in \mathcal{K}(X, Y). \quad \blacksquare$$

Hat man also Operatoren auf Banach-Räumen definiert, dann kann man unter Ausnutzung der Abgeschlossenheit von $\mathcal{K}(X, Y)$ in $\mathcal{L}(X, Y)$ neue kompakte Operatoren aus

bekanntem kompakten Operatoren durch „Grenzwertbildung“ erzeugen. Dabei muss die Grenzwertbildung im Sinne der Normkonvergenz erfolgen.

Beispiel 1.57

Es sei $X := Y := l^1$ und $T : X \rightarrow Y$ definiert durch

$$T(\xi_1, \xi_2, \xi_3, \dots) := \left(\xi_1, \frac{1}{2}\xi_2, \frac{1}{3}\xi_3, \dots\right). \quad (1.30)$$

Um die Kompaktheit von T nachzuweisen, betrachten wir neben T die endlichdimensionalen, beschränkten und damit kompakten Operatoren

$$T_n(\xi_1, \xi_2, \xi_3, \dots) := \left(\xi_1, \frac{1}{2}\xi_2, \frac{1}{3}\xi_3, \dots, \frac{1}{n}\xi_n, 0, \dots\right), \quad n = 1, 2, \dots$$

Wegen

$$\|T_n x - T x\|_{l^1} = \sum_{k=n+1}^{\infty} \frac{1}{k} |\xi_k| \leq \frac{1}{n+1} \sum_{k=n+1}^{\infty} |\xi_k| \leq \frac{1}{n+1} \|x\|_{l^1}$$

ist $\|T - T_n\| \leq 1/(n+1)$, d. h. $(T_n)_{n \in \mathbb{N}}$ ist normkonvergent gegen T . Somit ist nach der Bemerkung 1.56 der Operator T kompakt. ■

Wir haben bis jetzt die Kompaktheit linearer Operatoren mit Hilfe der Aussage aus Bemerkung 1.56 bewiesen. Hierzu musste man sich stets endlichdimensionale Operatoren T_n einfallen lassen, die den zu untersuchenden Operator in Bezug auf die Normkonvergenz beliebig genau approximieren. Falls man solche Operatoren nicht findet, kann man diese Untersuchungsmethode nicht anwenden.

Wenn es sich jedoch um Abbildungen aus \mathcal{C}^0 in \mathcal{C}^0 handelt, kann man das Kompaktheitskriterium von Arzelà¹¹ und Ascoli¹² heranziehen. Einen Beweis dieses Satzes findet man in Appell und Väh (2005), S. 49.

Satz 1.58 (Arzelà und Ascoli)

Eine Menge $M \subset \mathcal{C}^0[0, 1]$ ist genau dann präkompakt, wenn sie beschränkt und gleichgradig stetig ist. Dabei heißt M gleichgradig stetig, wenn es zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so dass für alle $f \in M$ und alle $x, y \in [0, 1]$ mit $|x - y| < \delta$ stets $|f(x) - f(y)| < \varepsilon$ bleibt.

Wir untersuchen nun eine Klasse von (nichtlinearen) Integraloperatoren auf Kompaktheit.

¹¹Cesare Arzelà (1847–1912) war ein italienischer Mathematiker und wurde 1880 als Professor an die Universität Bologna berufen. 1889 wurde der Satz von Ascoli von ihm zum Satz von Arzelà-Ascoli verallgemeinert.

¹²Giulio Ascoli (1843–1896) war ein italienischer Mathematiker. Auf ihn ist der Satz von Arzelà-Ascoli zurückzuführen.

Beispiel 1.59

Gegeben seien $X := C^0[a, b]$ ($-\infty < a < b < \infty$) und der Integraloperator

$$(Tx)(s) := \int_a^b F(s, t, x(t)) dt \quad \text{für alle } s \in [a, b].$$

Wir nehmen an, dass

$$F : Q \rightarrow \mathbb{R}, \quad Q := \{(s, t, x) \in \mathbb{R}^3 \mid s, t \in [a, b], |x| \leq r, r > 0, r \text{ fest}\}$$

stetig auf Q ist und behaupten: Der Operator

$$T : M \rightarrow X, \quad M := \{x \in X \mid \|x\| \leq r\}$$

ist kompakt.

Die stetige Funktion F ist auf der kompakten Menge Q sogar gleichmäßig stetig. Dies impliziert, dass es für jedes $\varepsilon > 0$ eine Zahl $\delta > 0$ gibt, so dass

$$|F(s_1, t, x_1) - F(s_2, t, x_2)| < \varepsilon \tag{1.31}$$

für alle $(s_1, t, x_1), (s_2, t, x_2) \in Q$ mit $|s_1 - s_2| + |x_1 - x_2| < \delta$. Wir zeigen zuerst die Stetigkeit von T . Falls $x \in M$, dann ist die Funktion $x : [a, b] \rightarrow \mathbb{R}$ und $|x(t)| \leq r$ für alle $t \in [a, b]$. Somit ist die Funktion $Tx : [a, b] \rightarrow \mathbb{R}$ ebenfalls stetig. Es seien $x_1, x_2 \in M$ und mit

$$\|x_1 - x_2\| = \max_{a \leq t \leq b} |x_1(t) - x_2(t)| < \delta$$

gilt die Abschätzung

$$\|Tx_1 - Tx_2\| = \max_{a \leq s \leq b} \left| \int_a^b [F(s, t, x_1(t)) - F(s, t, x_2(t))] dt \right| \leq (b-a)\varepsilon,$$

gemäß (1.31). Dies besagt, dass T stetig ist.

Nun zeigen wir die Kompaktheit von $T : M \rightarrow X$. Da M bereits beschränkt ist, verbleibt nach dem Satz von Arzelà und Ascoli noch der Nachweis:

- (i) $T(M)$ ist beschränkt, (ii) $T(M)$ ist gleichgradig stetig.

Zu (i): Mit $K := \max_{(s,t,x) \in Q} |F(s, t, x)|$ folgt für alle $x \in M$

$$\|Tx\| = \max_{a \leq s \leq b} \left| \int_a^b F(s, t, x(t)) dt \right| \leq (b-a)K.$$

Zu (ii): Sei $|s_1 - s_2| \leq \delta$ und $s_1, s_2 \in [a, b]$. Dann folgt wegen (1.31)

$$\begin{aligned} |(Tx)(s_1) - (Tx)(s_2)| &\leq \int_a^b |F(s_1, t, x(t)) - F(s_2, t, x(t))| dt \\ &\leq (b-a)\varepsilon \quad \text{für alle } x \in M. \end{aligned} \tag{1.32}$$

Dies zeigt, dass der (nichtlineare) Operator T kompakt ist. ■

Beispiel 1.60

Wir setzen $F(s, t, x(\cdot)) := k(s, t)x(\cdot)$, wobei $k : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$ stetig und $x \in \mathcal{C}^0[0, 1]$ sind und betrachten den durch

$$T : \mathcal{C}^0[0, 1] \rightarrow \mathcal{C}^0[0, 1], \quad (Tx)(s) := \int_0^1 k(s, t)x(t) dt \quad (1.33)$$

definierten Operator. T heißt *Fredholmscher Integraloperator* und k ist sein *Kern*. Dass T wieder in $\mathcal{C}^0[0, 1]$ abbildet, entnimmt man der Abschätzung (1.32). Seine Linearität ist offensichtlich, die Stetigkeit bzgl. der Maximumnorm sieht man wie folgt. Es gelten die Abschätzungen

$$\begin{aligned} \|Tx\| &= \max_{s \in [0, 1]} |(Tx)(s)| = \max_{s \in [0, 1]} \left| \int_0^1 k(s, t)u(t) dt \right| \\ &\leq \max_{s \in [0, 1]} \int_0^1 |k(s, t)| \left(\max_{t \in [0, 1]} |u(t)| \right) dt = \left(\max_{s \in [0, 1]} \int_0^1 |k(s, t)| dt \right) \|x\|. \end{aligned}$$

Dies liefert die Normabschätzung

$$\|T\| \leq \max_{s \in [0, 1]} \int_0^1 |k(s, t)| dt. \quad (1.34)$$

Darüber hinaus ist T kompakt, wie bereits in Beispiel 1.59 gezeigt. ■

Beispiel 1.61 (Inklusionsabbildung)

Es sei

$$\mathcal{C}^k[a, b] := \{y : [a, b] \rightarrow \mathbb{R} \mid y^{(j)} \text{ stetig auf } [a, b], j = 1, \dots, k\},$$

$y^{(j)}$ bezeichnet die j -te Ableitung von y und $[a, b]$ ist ein kompaktes Intervall. Dann ist $(\mathcal{C}^k[a, b], \|\cdot\|_{\mathcal{C}^k})$ mit der \mathcal{C}^k -Norm $\|y\|_{\mathcal{C}^k} := \sum_{j=0}^k \|y^{(j)}\|_{\mathcal{C}^0}$ ein Banach-Raum, wobei $\|y^{(j)}\|_{\mathcal{C}^0}$ die Max-Norm von $y^{(j)}$ in $\mathcal{C}^0[a, b]$ ist. Eine Funktion $y \in \mathcal{C}^{k+1}[a, b]$ liegt auch in $\mathcal{C}^k[a, b]$ und folglich gibt es eine sogenannte *Inklusionsabbildung* (auch *Einbettung* genannt)

$$j : \mathcal{C}^{k+1}[a, b] \rightarrow \mathcal{C}^k[a, b] \quad \text{mit} \quad y \mapsto j(y) := y. \quad (1.35)$$

Man kann zeigen, dass diese Inklusion $j : \mathcal{C}^{k+1}[a, b] \rightarrow \mathcal{C}^k[a, b]$ kompakt ist. ■

Das Spektrum eines kompakten Operators ist von besonders übersichtlicher Gestalt. Die genannten Aussagen folgen aus Beweisen, die in Riesz und Nagy Riesz und Sz-Nagy (1956) nachgelesen werden können.

Satz 1.62

Es sei $(X, \|\cdot\|)$ ein Banach-Raum über \mathbb{K} und $T \in \mathcal{K}(X, X)$, $T \neq 0$. Dann gilt:

- i. T hat nur endlich viele oder abzählbar unendlich viele Eigenwerte von endlicher Vielfachheit (d. h. $0 < \dim \ker(T - \lambda I) < \infty$), die sich bei Null häufen.
- ii. Jede Zahl $\lambda \neq 0$, die zum Spektrum von T gehört, ist ein Eigenwert von T (gehört also zum Punktspektrum $\sigma_p(T)$).

Das Eigenwertverhalten von kompakten Operatoren kann dennoch sehr vielgestaltig sein. Die nächsten Beispiele 1.63 - 1.65 zeigen verschiedene Situationen auf. Im ersten Beispiel wird an Hand eines *Volterra-Integraloperators* sichtbar, dass das Spektrum von T nur aus der Null bestehen kann und Null selbst kein Eigenwert von T ist.

Beispiel 1.63 (Volterra-Operator)

Gegeben sei der Volterra-Operator¹³

$$T : \mathcal{C}^0[0, 1] \rightarrow \mathcal{C}^0[0, 1], \quad (Tx)(s) := \int_0^s x(t) dt$$

und es sei $\mathcal{C}^0[0, 1]$ mit der Maximum-Norm ausgestattet. Wir berechnen das Spektrum $\sigma(T)$ für zwei Situationen.

a): Für die Abbildung T mit Definitionsbereich $\mathcal{C}^0[0, 1]$ gilt $\sigma(T) = \sigma_r(T) = \{0\}$. Um das einzusehen, betrachte man die Gleichung $(Tx)(s) = \int_0^s x(t) dt = 0$ ($\lambda = 0$). Nach dem Hauptsatz der Differenzial- und Integralrechnung folgt aus $(Tx)(s) = \int_0^s x(t) dt =: \tilde{x}(s) = 0$ mit $\tilde{x}(\cdot) \in \mathcal{C}^1[0, 1]$ sofort $\dot{\tilde{x}}(s) = x(s) = 0$ ($\forall s \in [0, 1]$). Das ist die Injektivität von T und T hat kein dichtes Bild, da stets $(Tx)(0) = \tilde{x}(0) = 0$ gilt. Also ist $0 \in \sigma_r(T)$.

b): Wir schränken jetzt den Definitionsbereich von T auf die Menge $X := \{x \in \mathcal{C}^0[0, 1] \mid x(0) = 0\}$ ein und berechnen nun (von der Abbildung $T : X \rightarrow X$) das Spektrum. Es gilt $\sigma(T) = \sigma_c(T) = \{0\}$. In der Tat ist der Operator T für $\lambda = 0$ wieder injektiv. Wegen

$$R(T) = \{y \in \mathcal{C}^1[0, 1] \mid y(0) = 0\} \subsetneq \mathcal{C}^1[0, 1]$$

gilt zunächst $R(T) \subsetneq X$, jedoch gilt für die Abschließung von $R(T)$ die Beziehung $\overline{R(T)} = X$, also ist diesmal $0 \in \sigma_c(T)$.

c): Die Gleichung

$$\lambda x - Tx = y \quad (\lambda \neq 0) \tag{1.36}$$

kann für $\lambda \neq 0$ eindeutig nach x für jede rechte Seite $y \in \mathcal{C}^0[0, 1]$ bzw. $y \in X$ aufgelöst werden. Es ist dann $\lambda I - T$ bijektiv, und die Stetigkeit des inversen Operators ergibt sich aus dem Satz von Banach. Somit ist $\rho(T) = \mathbb{R} \setminus \{0\}$. ■

¹³Vito Volterra (1860-1940). Studium an der Universität von Pisa. Bekannt sind vor allem seine Arbeiten zu Integralgleichungen und zur Populationsdynamik in Räuber-Beute-Beziehungen.

Wir wollen an einem weiteren Beispiel zeigen, dass ein kompakter Operator unendlich viele charakteristische Werte haben kann, die eine unbeschränkte Folge ohne Häufungspunkte bilden. Für die Eigenwerte des Operators bedeutet dies, dass sie sich höchstens bei Null häufen können.

Beispiel 1.64

Gegeben sei das Randwertproblem (RWP)

$$\ddot{x} = -\mu x, \quad x(0) = x(1) = 0, \quad \mu > 0. \quad (1.37)$$

Wir schreiben das RWP in eine Operatorgleichung um, indem wir den Operator

$$L : X \rightarrow Y, \quad Lx := \ddot{x}, \quad X := \mathcal{C}_0^2[0, 1], \quad Y := \mathcal{C}^0[0, 1], \\ \mathcal{C}_0^2[0, 1] := \{x \mid x \in \mathcal{C}^2[0, 1], x(0) = x(1) = 0\}, \quad \|x\|_{\mathcal{C}^2} := \|x\|_{\mathcal{C}^0} + \|\dot{x}\|_{\mathcal{C}^0} + \|\ddot{x}\|_{\mathcal{C}^0}$$

eingeführen. Die Gleichung (1.37) geht über in

$$Lx = -\mu x. \quad (1.38)$$

Damit der Operator L eine Abbildung in sich vermittelt, müsste man L wie folgt definieren: $L : \text{dom}(L) \subset Y \rightarrow Y$ mit $\text{dom}(L) := X$. Der Operator L ist jedoch nur als Abbildung $L : X \rightarrow Y$ stetig:

$$\|Lx\|_{\mathcal{C}^0} := \max_{t \in [0, 1]} |(Lx)(t)| = \|\ddot{x}\|_{\mathcal{C}^0} \leq \|\ddot{x}\|_{\mathcal{C}^0} + \|\dot{x}\|_{\mathcal{C}^0} + \|x\|_{\mathcal{C}^0} = \|x\|_{\mathcal{C}^2}.$$

Betrachtet man L als Abbildung $L : Y \rightarrow Y$, dann zeigt die Funktionenfolge $(x_n) \subset Y$, $x_n(t) := \sin(n\pi t)$, dass L nicht beschränkt ist:

$$\|Lx_n\|_{\mathcal{C}^0} := \max_{t \in [0, 1]} |(Lx_n)(t)| = n^2 \pi^2 \max_{t \in [0, 1]} |-\sin(n\pi t)| = n^2 \pi^2 \|x_n\|_{\mathcal{C}^0} = n^2 \pi.$$

Dies veranlasst uns, das RWP (1.37) in eine Integralgleichung zu überführen. Wir integrieren (1.37) über $[0, t]$ und erhalten

$$\int_0^t \ddot{x}(\tau) d\tau = \dot{x}(t) - \dot{x}(0) = -\mu \int_0^t x(\tau) d\tau.$$

Nochmalige Integration über $[0, s]$ liefert (man beachte $x(0) = 0$)

$$x(s) - \dot{x}(0)s = -\mu \int_0^s \left(\int_0^r x(t) dt \right) dr = -\mu \int_0^s \left(\int_t^s x(t) dr \right) dt \\ = -\mu \int_0^s (s-t)x(t) dt. \quad (1.39)$$

Der Term $\dot{x}(0)s$ folgt aus (1.39) in Verbindung mit der Randbedingung $x(1) = 0$:

$$\dot{x}(0) = \mu \int_0^1 (1-t)x(t) dt.$$

Nach Einsetzen von $\dot{x}(0)$ in (1.39) und Aufspalten der Integration über die Intervalle $[0, s]$ und $[s, 1]$, wobei $s \in [0, 1]$, ergibt sich

$$\begin{aligned} x(s) &= \dot{x}(0)s - \mu \int_0^s (s-t)x(t) dt = \mu s \int_0^1 (1-t)x(t) dt - \mu \int_0^s (s-t)x(t) dt \\ &= \mu \int_0^s \underbrace{(s(1-t) - (s-t))}_{=t(1-s)} x(t) dt + \mu \int_s^1 s(1-t)x(t) dt. \end{aligned}$$

Hierfür schreiben wir

$$x(s) = \mu \int_0^1 k(s, t)x(t) dt$$

mit der *Greenschen Funktion* (auch Kernfunktion)

$$k(s, t) := \begin{cases} t(1-s), & \text{falls } 0 \leq t \leq s, \\ s(1-t), & \text{falls } s \leq t \leq 1. \end{cases} \quad (1.40)$$

Es sei

$$K : \mathcal{C}^0[0, 1] \rightarrow \mathcal{C}^0[0, 1], \quad Kx(s) := \int_0^1 k(s, t)x(t) dt$$

mit k aus (1.40) gegeben. Nach Beispiel 1.59 ist $K \in \mathcal{K}(\mathcal{C}^0, \mathcal{C}^0)$. Wir zeigen, dass K die charakteristischen Werte $\mu_n = n^2\pi^2$ ($n = 1, 2, \dots$) hat. Zum Nachweis muss die Gleichung

$$\mu Kx = x \quad (1.41)$$

untersucht werden. Aus dem RWP (1.37) wissen wir, dass die Funktionen $x_n(t) := \sin(n\pi t)$ zusammen mit den Werten $\mu_n = n^2\pi^2$ die Gleichung (1.37) erfüllen. Die folgende Rechnung zeigt, dass auch

$$\begin{aligned} \mu_n(Kx)_n(s) &= n^2\pi^2 \int_0^1 k(s, t)x_n(t) dt \\ &= n^2\pi^2 \int_0^s t(1-s) \sin(n\pi t) dt + n^2\pi^2 \int_s^1 s(1-t) \sin(n\pi t) dt \\ &= \sin(n\pi s) = x_n(s) \end{aligned}$$

gilt. Der Operator K hat also abzählbar unendlich viele charakteristische Zahlen, die sich im Endlichen nicht häufen. ■

Wir zeigen im dritten Beispiel, dass ein nichtkompakter Operator $T \in \mathcal{L}(X, X)$ überabzählbar unendlich viele charakteristische Werte haben kann.

Beispiel 1.65

Es sei $X := l^2$ und T die Linksverschiebung, d. h.

$$T(\xi_1, \xi_2, \xi_3, \dots) := (\xi_2, \xi_3, \xi_4, \dots).$$

Die Norm ist $\|T\| = 1$. Die Nichtkompaktheit von T folgt aus der Tatsache, dass die Einheitskugel in l^2 nicht kompakt ist. Wir untersuchen die Gleichung

$$\mu T x = x. \tag{1.42}$$

Die Werte von μ , für die die Gleichung (1.42) nichttriviale Lösungen hat, liegen im Bereich $|\mu| \geq \|T\|^{-1} = 1$. Für beliebiges $\mu \in (-\infty, -1) \cup (1, \infty)$ erfüllt die Folge

$$\hat{x}_\mu := (1, \mu^{-1}, \mu^{-2}, \dots) \in X \tag{1.43}$$

die Gleichung (1.42) und es gilt

$$\|\hat{x}_\mu\|_{l^2}^2 = \sum_{n=0}^{\infty} \mu^{-2n} = \frac{\mu^2}{\mu^2 - 1}.$$

Für $\mu = \pm 1$ löst die Folge (1.43) zwar noch die Gleichung (1.42), liegt aber nicht mehr im Raum X . ■

In Bezug auf das Spektrum zeigen also kompakte Operatoren ein recht übersichtliches Verhalten. Nachfolgendes Lemma charakterisiert das Spektrum hinsichtlich endlicher- bzw. unendlichdimensionaler Räume.

Lemma 1.66

Es sei $T : X \rightarrow X$ ein linearer Operator und X ein Banach-Raum. Dann gilt:

- i.** *Ist $\dim X < \infty$, so ist $\sigma(T) = \sigma_p(T)$.*
- ii.** *Ist $\dim X = \infty$ und $T \in \mathcal{K}(X, X)$, so ist $0 \in \sigma(T)$. Im Allgemeinen ist 0 aber kein Eigenwert.*

Beweis: Zu **i.**: Ist $\lambda \in \sigma(T)$, so ist $T - \lambda I$ nicht bijektiv, also, da $\dim X < \infty$, auch nicht injektiv, d. h. $\lambda \in \sigma_p(T)$ (siehe Bemerkung 1.31).

Zu **ii.**: Es sei $T \in \mathcal{K}(X, X)$ und angenommen es sei $0 \in \rho(T)$. Dann ist $T^{-1} \in \mathcal{L}(X, X)$ und nach Bemerkung 1.56 (2) ist $I = T^{-1}T \in \mathcal{K}(X, X)$. Nun müsste die Einheitskugel kompakt sein, was aber nur in endlichdimensionalen Banach-Räumen gilt (siehe Bemerkung 1.7 und Aufgabe 1.4). Dies ist ein Widerspruch zu $\dim X = \infty$. □

Der nachfolgende Satz zeigt, dass kompakte Störungen der Identität Fredholm-Operatoren sind. Details des Beweises findet man in Alt (1992), S. 373–376.

Satz 1.67

Es sei X ein Banach-Raum und $T \in \mathcal{K}(X, X)$. Dann ist $\lambda I - T$ ($\lambda \neq 0$) ein Fredholm-Operator mit Index 0. Die linearen homogenen Gleichungen

$$Tx - \lambda x = 0, \quad x \in X \quad \text{und} \quad T^*x^* - \lambda x^* = 0, \quad x^* \in X^* \quad (\lambda \neq 0)$$

haben stets dieselbe endliche Anzahl linear unabhängiger Lösungen.

Diese Ergebnisse über Fredholm-Operatoren lassen sich in der Fredholmschen Alternative zusammenfassen.

Satz 1.68 (Fredholmsche Alternative)

Es seien $T \in \mathcal{K}(X, X)$, X ein Banach-Raum und $\lambda \neq 0$ gegeben. Dann gilt:

Entweder ist die Gleichung $Tx - \lambda x = y$ für jedes $y \in X$ eindeutig lösbar,
oder die Gleichung $Tx - \lambda x = 0$ hat nichttriviale Lösungen.

Eine knappe Formulierung der Fredholmschen Alternative lautet:

$$\begin{aligned} \text{„Entweder“:} \quad & \lambda \in \rho(T), \\ \text{„Oder“:} \quad & \lambda \in \sigma(T) \setminus \{0\} \subset \sigma_p(T). \end{aligned}$$

Die „Entweder“-Aussage folgt aus Satz 1.53, i. Da in unserem Fall $T \in \mathcal{K}(X, X)$, folgt aus Satz 1.67, dass $\text{ind}(T - \lambda I) = 0$ ist. Dies hat zur Folge, dass wegen

$$0 = \dim \ker(T - \lambda I) - \text{codim } R(T - \lambda I) \quad \text{und} \quad \dim \ker(T - \lambda I) = 0$$

der Operator $T - \lambda I$ auch surjektiv ist. Nach dem Satz von Banach über die Umkehrbarkeit linearer stetiger Operatoren existiert der Umkehroperator als lineare stetige Inverse $(T - \lambda I)^{-1}$ auf ganz X , d. h. dass die Gleichung $Tx - \lambda x = y$ für jedes $y \in X$ eindeutig lösbar ist.

Im „Oder“-Fall ist $\ker(T - \lambda I) \neq \{0\}$ und die Gleichung $Tx - \lambda x = 0$ hat nichttriviale Lösungen. Es können nach Satz 1.67 nur endlich viele linear unabhängige Lösungen auftreten. \square

Die Fredholmsche Alternative hat aufgrund der verschiedenen Sätze über kompakte Operatoren die unterschiedlichsten Möglichkeiten der Formulierung. Zieht man Satz 1.42 heran, dann erhält man

Lemma 1.69

Es seien $T \in \mathcal{K}(X, X)$, X ein Banach-Raum und $\lambda \neq 0$ gegeben. Dann gilt:

$$\text{i.} \quad \ker(\lambda I - T) = \{0\} \quad \Leftrightarrow \quad R(\lambda I - T) = X,$$

ii. $R(\lambda I - T) = [\ker(\lambda I - T^*)]_{\perp} = \{u \in X \mid \psi^*(u) = 0 \text{ für alle } \psi^* \in \ker(\lambda I - T^*)\}$.

In der analytischen Bifurkationstheorie (siehe Kapitel 4) spielt bei der Untersuchung von Eigenwertgleichungen der Form $(T - \lambda I)x = y$ der Begriff der Vielfachheit eines Eigenwertes eine besondere Rolle. Wir bringen bereits an dieser Stelle die

Definition 1.70 (Geometrische und algebraische Vielfachheit)

Mit $m(\lambda)$ bezeichnen wir die *geometrische Vielfachheit* des Eigenwertes λ des Operators T und meinen damit die Dimension von $\ker(\lambda I - T)$, während die *algebraische Vielfachheit* $\chi(\lambda)$ des Eigenwertes λ von T definiert ist als Dimension von $\cup_{n \in \mathbb{N}} \ker((\lambda I - T)^n)$.

Die algebraische Vielfachheit eines Eigenwertes λ ist im Allgemeinen *größer* als die geometrische Vielfachheit. Ein Eigenwert λ heißt *einfach*, falls seine algebraische Vielfachheit eins ist.

Man kann nun zeigen (siehe Appell und Văth (2005), S. 155-157), dass sich für $T \in \mathcal{K}(X, X)$ und $\lambda \neq 0$ eine einfache Berechnung der algebraischen Vielfachheit $\chi(\lambda)$ ergibt: Unter den gemachten Voraussetzungen existiert ein (endliches) $p \in \mathbb{N}$, so dass

$$\chi(\lambda) = \dim\{\cup_{n \in \mathbb{N}} \ker((\lambda I - T)^n)\} = \dim\{\ker((\lambda I - T)^p)\}.$$

Beispiel 1.71

Wir betrachten den linearen Operator $T : \mathbb{R}^4 \rightarrow \mathbb{R}^4$ mit

$$T := \begin{pmatrix} \lambda_1 & 1 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 \\ 0 & 0 & 0 & \lambda_2 \end{pmatrix},$$

wobei $\lambda_1, \lambda_2 \neq 0$ und $\lambda_1 \neq \lambda_2$. Dann gilt $\chi(\lambda_1) = 3$, $m(\lambda_1) = 2$, $\chi(\lambda_2) = m(\lambda_2) = 1$. ■

Bei naturwissenschaftlichen und technischen Problemstellungen kommen häufig Gleichungen der Form

$$x - Tx = y \quad \text{mit} \quad T \in \mathcal{K}(X, X)$$

vor. Nachfolgendes Theorem ist genau auf diesen Gleichungstyp zugeschnitten. Es wird in der Literatur als Satz von Riesz-Schauder bezeichnet. Alle darin enthaltenen Aussagen sind bereits aus den vorangehenden Sätzen bekannt. Somit ist nachfolgender Satz lediglich noch einmal eine *Zusammenfassung bisheriger Resultate*.

Satz 1.72 (Riesz-Schauder Theory)

Es sei X ein Banach-Raum, X^* sein Dualraum und $T \in \mathcal{K}(X, X)$, $T^* \in \mathcal{K}(X^*, X^*)$.
Dann gilt:

- i. Falls $R(T)$ abgeschlossen ist, dann ist $\dim R(T) < \infty$.
- ii. $\dim \ker(I - T) < \infty$, $R(I - T)$ ist abgeschlossen.
- iii. (Fredholmsche Alternative)

$$R(I - T) = X \text{ gilt genau dann, wenn } \ker(I - T) = \{0\}.$$

- iv. $\dim \ker(I - T) = \dim \ker(I^* - T^*) < \infty$.
- v. Falls $1 \in \sigma(T)$, dann existiert ein $k \in \mathbb{N}$, so dass

$$X = \ker(I - T)^k \oplus R((I - T)^k).$$

Darüber hinaus sind beide Räume auf der rechten Seite T -invariant und es ist $\dim \ker(I - T)^k < \infty$.

Im folgenden Beispiel werden von einem Fredholm-Operator T die charakteristischen Werte bestimmt und die Lösbarkeitsbedingung für die dazugehörige lineare inhomogene Gleichung aufgestellt.

Beispiel 1.73

Gegeben sei der Integraloperator $T : X \rightarrow X$, $X := C^0[0, 2\pi]$ durch

$$(Tx)(s) := \int_0^{2\pi} k(s, t)x(t) dt, \quad k(s, t) := \sin(s + t).$$

Der Operator T ist kompakt (siehe Beispiel 1.60) und der Kern k sogar symmetrisch. Die Voraussetzungen aus Satz 1.72 sind erfüllt. Zur Bestimmung der charakteristischen Werte von T ist eine Lösung $x(s) \neq 0$ der Gleichung

$$x(s) = \mu \int_0^{2\pi} \sin(s + t)x(t) dt \tag{1.44}$$

zu berechnen. Die Struktur der Gleichung zeigt, dass jede Lösung die Form

$$\begin{aligned} x(s) &= \mu \int_0^{2\pi} (\sin s \cdot \cos t + \cos s \cdot \sin t)x(t) dt \\ &= \mu \left[\sin s \int_0^{2\pi} \cos(t)x(t) dt + \cos s \int_0^{2\pi} \sin(t)x(t) dt \right] \\ &= \mu\xi \sin s + \mu\eta \cos s \end{aligned} \tag{1.45}$$

mit

$$\xi := \int_0^{2\pi} \cos(t) x(t) dt, \quad \eta := \int_0^{2\pi} \sin(t) x(t) dt \quad (1.46)$$

hat. Einsetzen von $x(t) = \mu\xi \sin t + \mu\eta \cos t$ in (1.46) liefert das Gleichungssystem

$$\begin{aligned} \xi &= \int_0^{2\pi} \cos t [\mu\xi \sin t + \mu\eta \cos t] dt = \mu\eta\pi, \\ \eta &= \int_0^{2\pi} \sin t [\mu\xi \sin t + \mu\eta \cos t] dt = \mu\xi\pi, \end{aligned}$$

bzw. in der Standardform

$$\begin{aligned} \xi - \pi\mu\eta &= 0 \\ \pi\mu\xi - \eta &= 0 \end{aligned} \Leftrightarrow \begin{pmatrix} 1 & -\pi\mu \\ \pi\mu & -1 \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (1.47)$$

Nichttriviale Lösungen des linearen homogenen Gleichungssystems erhalten wir für $\mu = \mu_1 := \pi^{-1}$ und $\mu = \mu_2 := -\pi^{-1}$. Eine dazugehörige Lösungsbasis lautet: $e_1 := (1, 1)^T$ und $e_2 := (1, -1)^T$. Wählt man also $\mu = \mu_1$ oder $\mu = \mu_2$, dann bekommt man nichttriviale Lösungen

$$x_1(s) = \frac{1}{\pi} \sin s + \frac{1}{\pi} \cos s, \quad x_2(s) = \frac{1}{\pi} \sin s - \frac{1}{\pi} \cos s.$$

Ist dagegen $\mu \neq \mu_1$ und $\mu \neq \mu_2$, dann muss $(\xi, \eta) = (0, 0)$ sein und (1.44) hat nur die triviale Lösung. Der Operator T besitzt genau zwei charakteristische Werte $\mu_1 = \pi^{-1}$ und $\mu_2 = -\pi^{-1}$. Damit ist auch klar, dass die lineare inhomogene Gleichung

$$x(s) - \left(\pm \frac{1}{\pi} \right) \int_0^{2\pi} \sin(s+t) x(t) dt = y(s)$$

nur für gewisse $y \in X$ eine Lösung besitzt. Die Lösbarkeitsbedingung ergibt sich aus der Lösungsstruktur

$$x(s) = \pm \frac{1}{\pi} [\xi \sin s + \eta \cos s] + y(s). \quad (1.48)$$

Einsetzen von (1.48) in (1.46) liefert (wir arbeiten mit „ $+\pi^{-1}$ “):

$$\begin{aligned} \xi &= \eta + \int_0^{2\pi} \cos(t) y(t) dt \\ \eta &= \xi + \int_0^{2\pi} \sin(t) y(t) dt \end{aligned} \Leftrightarrow \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} \int_0^{2\pi} \cos(t) y(t) dt \\ \int_0^{2\pi} \sin(t) y(t) dt \end{pmatrix}. \quad (1.49)$$

Das Gleichungssystem (1.49) besitzt nur nichttriviale Lösungen (Lösbarkeitsbedingung) für

$$\int_0^{2\pi} \cos(t) y(t) dt = 0 \quad \text{und} \quad \int_0^{2\pi} \sin(t) y(t) dt = 0. \quad \blacksquare$$

Fredholm-Operatoren kommen auch bei der Untersuchung nichtlinearer DGL-Systeme der Form $\dot{x} = A(t)x + g(t, x)$ vor. Hier bezeichnen A eine $n \times n$ -Matrix mit stetigen Einträgen und g eine Nichtlinearität. Unter gewissen Voraussetzungen ergibt sich $x(\cdot) \mapsto (Tx)(\cdot) := \frac{dx(\cdot)}{dt} - A(\cdot)x(\cdot)$ als ein Fredholm-Operator (siehe Abschnitt 4.2).

1.3 Fréchet- und Gâteaux-Ableitung

Wie bereits erwähnt, ist ein wesentliches Ziel dieses Kapitels die Verallgemeinerung des klassischen impliziten Funktionentheorems, das Aussagen zur Auflösung von Gleichungen der Form $F(x, y) = 0$ macht. Im einfachsten Fall ist F eine Abbildung $F : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$. Eine wesentliche Voraussetzung des Theorems lautet: $F_y(x_0, y_0) \neq 0$. Will man das Theorem auf Abbildungen $F : X \times Y \rightarrow Z$, wobei X, Y und Z Banach-Räume sind, übertragen, dann muss diese Forderung entsprechend ersetzt werden. Dies geschieht, indem man verlangt, dass $F_y(x_0, y_0)$ eine lineare stetige Inverse besitzt. Spätestens hier erkennt der Leser, dass das klassische Ableitungskonzept auf Abbildungen $f : X \rightarrow Y$ zwischen normierten Räumen erweitert werden muss.

Die beiden aus dem \mathbb{R}^n bekannten Begriffe „totale Ableitung“ und „Richtungsableitung“ werden hier zur *Fréchet-* und *Gâteaux-Ableitung* verallgemeinert. Dabei bleibt das bewährte Prinzip aus dem \mathbb{R}^n – *Differenzieren ist nichts weiter als Linearisieren* – erhalten.

Wir erinnern zuvor an einige bekannte Notationen (*Landau-Symbolik*). $U(x)$ bezeichnet wie immer eine (beliebige) Umgebung um den Punkt x . Für eine Abbildung $r : U(0) \subseteq X \rightarrow Y$ schreiben wir:

$$\begin{aligned} r(x) &= o(\|x\|), \quad x \rightarrow 0, \quad \text{falls } r(x)/\|x\| \rightarrow 0, \quad \text{für } x \rightarrow 0; \\ r(x) &= o(1), \quad x \rightarrow 0, \quad \text{falls } r(x) \rightarrow 0, \quad \text{für } x \rightarrow 0. \end{aligned}$$

Diese Schreibweisen ermöglichen eine prägnante Darstellung der Ableitungsbegriffe und man erkennt sofort das oben genannte Prinzip.

Im Folgenden sei die Funktion $f : \Omega \rightarrow Y$ gegeben, $\Omega \subset X$ eine offene Teilmenge des normierten Raumes X und Y ebenfalls ein normierter Raum.

Definition 1.74 (Fréchet-Differenzierbarkeit)

Die Abbildung f heißt *Fréchet-differenzierbar* (F-differenzierbar) im Punkt x , falls ein Operator $T \in \mathcal{L}(X, Y)$ existiert, so dass

$$f(x+h) - f(x) = Th + o(\|h\|), \quad h \rightarrow 0 \quad (1.50)$$

(bei festem $x \in X$) für alle h aus einer Umgebung um Null gilt. In diesem Fall heißt T die *Fréchet-Ableitung* (F-Ableitung) von f in x . Wir schreiben für $T = f'(x)$ und das *Fréchet-Differential* (F-Differential) in x ist definiert durch $df(x; h) := f'(x)h$.

In Formel (1.50) erkennt man sehr gut das bereits in der Einleitung angesprochene Konzept der Linearisierung, wie es beim klassischen Kalkül Anwendung findet. Mit anderen Worten: Der Graph von f wird approximiert durch die „Tangente“ im Punkt $(x, f(x))$. Diese hat die Darstellung $h \mapsto f(x) + f'(x) \cdot h$ und stellt exakt den linear (affinen) Teil von f dar.

Entsprechende Rechenregeln, wie für die klassische Differenzialrechnung bekannt (z. B. Summen-, Produkt- und Kettenregel, Differentiation der Umkehrfunktion), können analog hergeleitet werden.

Definition 1.75 (Gâteaux-Differenzierbarkeit)

Die Abbildung f heißt *Gâteaux-differenzierbar* (G-differenzierbar) in x , falls ein Operator $S \in \mathcal{L}(X, Y)$ existiert, so dass

$$f(x+tk) - f(x) = tSk + o(t), \quad t \rightarrow 0 \quad (1.51)$$

für alle $k \in X$ und alle reellen Zahlen t aus einer Nullumgebung gilt. Mit $S = f'(x)$ wird die *Gâteaux-Ableitung* (G-Ableitung) von f im Punkt x bezeichnet. Das *Gâteaux-Differential* (G-Differential) in x ist durch $d_G f(x; k) = Sk$ definiert.

Falls $f : X \rightarrow \mathbb{K}$ ein Gâteaux-differenzierbares Funktional auf dem normierten Raum X ist, dann gilt für jedes feste $x \in X$

$$d_G f(x; k) = \left. \frac{d}{dt} f(x+tk) \right|_{t=0}. \quad (1.52)$$

Die Beziehung ergibt sich, indem man die Funktion $s(t) := f(x+tk)$ an der Stelle $t=0$ differenziert:

$$s'(0) = \lim_{t \rightarrow 0} \frac{s(t) - s(0)}{t} = \lim_{t \rightarrow 0} \frac{f(x+tk) - f(x)}{t} = \left. \frac{d}{dt} f(x+tk) \right|_{t=0} = d_G f(x; k).$$

Partielle Ableitungen und höhere Ableitungen von Funktionen sowohl einer als auch mehrerer Variabler werden vollständig parallel zum klassischen Vorgehen gewonnen.

Definition 1.76 (Höhere Ableitungen)

Höhere Ableitungen werden gleichermaßen für die F- und die G-Ableitung sukzessive erklärt. Somit heißt die Ableitung der Abbildung

$$f' : \Omega \subseteq X \rightarrow \mathcal{L}(X, Y), \quad x \mapsto f'(x) \quad (1.53)$$

zweite Ableitung von f in $x \in \Omega$, in Zeichen $f''(x)$, usw.

Die Ableitung von f' in x , d. h. $f''(x)$, erhält man aus der Definitionsgleichung

$$f'(x+h) = f'(x) + f''(x)h + o(\|h\|), \quad h \rightarrow 0.$$

Hierbei ist $f''(x)$ ein Operator von X in $\mathcal{L}(X, Y)$, d. h. $f''(x) \in \mathcal{L}(X, \mathcal{L}(X, Y))$. Falls man $h, k \in X$ wählt, dann ist

$$f''(x)k \in \mathcal{L}(X, Y) \quad \text{und} \quad (f''(x)k)h \in Y.$$

Wir benutzen dafür die Abkürzung $f''(x)kh$ bzw. $f''kh$. Außerdem gelten die Ungleichungen

$$\|f''(x)kh\| \leq \|f''(x)k\| \|h\| \leq \|f''(x)\| \|k\| \|h\|.$$

Beispiel 1.77

Wir vergleichen die Fréchet- und Gâteaux-Differenzierbarkeit mit den Differentiationsbegriffen einer Funktion $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ in n Variablen, $x = (x_1, \dots, x_n) \in \Omega$.

1. Die F-Differenzierbarkeit von f in x ist äquivalent mit der klassischen totalen Differenzierbarkeit von f in x .
2. Die Funktion f ist G-differenzierbar in x , genau dann, wenn ein $a \in \mathbb{R}^n$ existiert, so dass für alle $k \in \mathbb{R}^n$ gilt:

$$\lim_{t \rightarrow 0} \frac{f(x+tk) - f(x)}{t} = \sum_{i=1}^n a_i k_i = \langle a, k \rangle.$$

Der Ausdruck $\langle a, k \rangle$ ist von der Form $Sk := \langle a, k \rangle$, wobei S ein linearer stetiger Operator von \mathbb{R}^n in \mathbb{R} ist. Man erkennt, dass es sich hierbei um die (klassische) Richtungsableitung handelt. ■

Ohne Beweis führen wir die folgenden Eigenschaften an:

Satz 1.78

1. Falls das Gâteaux-Differential existiert, dann ist es eindeutig bestimmt.
2. Falls eine Abbildung f Fréchet-differenzierbar in einem Punkt ist, dann ist sie dort auch Gâteaux-differenzierbar und beide Differentiale sind gleich.
3. Falls f' als Gâteaux-Ableitung in einer Umgebung von x existiert und die Abbildung $x \mapsto f'(x)$ stetig in x ist, dann ist $f'(x)$ auch die Fréchet-Ableitung im Punkt x .

Das nachfolgende Beispiel zeigt, dass es Funktionen gibt, die Gâteaux- jedoch nicht Fréchet-differenzierbar sind. Somit kann die Voraussetzung der Stetigkeit der Abbildung $x \mapsto f'(x)$ aus Satz 1.78 nicht fallen gelassen werden.

Beispiel 1.79

Es sei $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ durch

$$f(x, y) := \begin{cases} \frac{x^3 y}{x^4 + y^2}, & \text{falls } x^2 + y^2 > 0 \\ 0, & \text{falls } x = y = 0 \end{cases}$$

definiert. Es ist leicht zu prüfen, dass f im Punkt $(0, 0)$ Gâteaux-differenzierbar ist und das Gâteaux-Differential verschwindet. Gemäß (1.51) gilt mit $k = (k_1, k_2) \in \mathbb{R}^2$ und für jedes $t \in \mathbb{R}$:

$$\begin{aligned} f(tk_1, tk_2) - f(0, 0) &= \frac{t^3 k_1^3 t k_2}{t^4 k_1^4 + t^2 k_2^2} = t \cdot \underbrace{S(k_1, k_2)}_{=0} + r((0, 0), k; t), \\ \frac{r((0, 0), k; t)}{t} &:= \frac{t k_1^3 k_2}{t^2 k_1^4 + k_2^2} \xrightarrow{t \rightarrow 0} 0 \quad (k_2 \neq 0). \end{aligned}$$

Wäre f auch Fréchet-differenzierbar im Punkt $(0, 0)$, dann müssten Fréchet- und Gâteaux-Ableitung übereinstimmen. Aus (1.50) ergibt sich

$$f(h_1, h_2) - f(0, 0) = \underbrace{f'(0, 0)}_{=0}(h_1, h_2) + \frac{h_1^3 h_2}{h_1^4 + h_2^2}, \quad h := (h_1, h_2), \quad h \neq 0$$

und es müsste der Restterm

$$r((0, 0); (h_1, h_2)) := \frac{h_1^3 h_2}{h_1^4 + h_2^2} \tag{1.54}$$

nach Division von $\|h\|$ für $h \rightarrow 0$ gegen Null gehen. Nähert man sich jedoch auf der Parabel $h_2 = h_1^2$ dem Nullpunkt, dann ergibt sich für $h_1 \rightarrow 0$ der Wert

$$\frac{|r((0, 0); (h_1, h_2))|}{\|(h_1, h_2)\|} = \frac{|h_1^3 h_2|}{h_1^4 + h_2^2} \cdot \frac{1}{\sqrt{h_1^2 + h_2^2}} \stackrel{h_2 = h_1^2}{=} \frac{|h_1|^5}{2h_1^4} \cdot \frac{1}{|h_1| \sqrt{1 + h_1^2}} \xrightarrow{h_1 \rightarrow 0} \frac{1}{2}.$$

Somit ist (1.54) nicht von der Ordnung $o(\|h\|)$, $h \rightarrow 0$. ■

Obwohl die grundlegenden Definitionen der Differenzialrechnung ungeändert auf normierte Räume übertragbar sind, bedeutet das nicht, dass es in unendlichdimensionalen Räumen keine neuen Phänomene gäbe. Als Beispiel kann man die Funktion $(F(g))(t) := \sin g(t)$ verwenden, indem man $g \in X$ einmal aus $X := C^0[0, 1]$ und zum anderen Mal aus $X := L^2[0, 1]$ wählt und damit in beiden Fällen jeweils eine Abbildung $F : X \rightarrow X$ auf Differenzierbarkeit zu untersuchen hat. Der Unterschied besteht lediglich darin, dass F auf unterschiedlichen Funktionenräumen operiert. Im ersten Fall ist F Fréchet-differenzierbar mit $(F'(g))h(t) = (\cos g(t))h(t)$; im zweiten Fall ist die Funktion F Lipschitz-stetig, jedoch nicht Fréchet-differenzierbar.

1.4 Nemytski-Operator

Eine der wichtigsten nichtlinearen Abbildungen ist der so genannte *Nemytski-Operator*, manchmal auch *Substitutions-Operator* genannt. Wie der letzte Name bereits signalisiert, entsteht der Operator F durch Substitution von einer Funktion $\varphi : G \subset \mathbb{R}^n \rightarrow \mathbb{R}$ in die Funktion $f : G \times \mathbb{R} \rightarrow \mathbb{R}$. Dies führt zu einem neuen Operator

$$F : \varphi \mapsto f(\cdot, \varphi(\cdot)),$$

der erklärt ist auf einem Raum X von Funktionen φ . Nun sucht man nach Bedingungen an f , die F zu einer (differenzierbaren) Abbildung von X in X machen. In unseren späteren Betrachtungen ist der Fall $X := C^0[0, 1]$ der Wichtigste unter allen anderen. Es ist klar, dass die Stetigkeit von f auf $[0, 1] \times \mathbb{R}$ die Stetigkeit von F garantiert. Da f gleichmäßig stetig auf der kompakten Menge der Form

$$\{(x, y) \in [0, 1] \times \mathbb{R} \mid |y - \varphi(x)| \leq 1\} \quad \text{für } \varphi \in C^0[0, 1],$$

ist, folgt, dass F stetig auf X ist.

Wir nehmen nun an, dass die partielle Ableitung $\frac{\partial f}{\partial y}$ stetig auf $[0, 1] \times \mathbb{R}$ sei. Für $\varphi, h \in X$ liefert der klassische Mittelwertsatz

$$\frac{f(x, \varphi(x) + th(x)) - f(x, \varphi(x))}{t} = \frac{\partial f}{\partial y}(x, \varphi(x) + \vartheta(t, x)th(x))h(x)$$

für ein $\vartheta(t, x) \in (0, 1)$. Wegen der Stetigkeit von $\frac{\partial f}{\partial y}$ gibt es zu jedem $\varepsilon > 0$ ein $\delta > 0$, so dass für $0 < |t| < \delta$

$$\begin{aligned} & \sup_{x \in [0, 1]} \left| \frac{f(x, \varphi(x) + th(x)) - f(x, \varphi(x))}{t} - \frac{\partial f}{\partial y}(x, \varphi(x))h(x) \right| \\ & \leq \sup_{x \in [0, 1]} \sup_{0 \leq \vartheta \leq 1} \left| \frac{\partial f}{\partial y}(x, \varphi(x) + \vartheta(t, x)th(x)) - \frac{\partial f}{\partial y}(x, \varphi(x)) \right| |h(x)| \leq \varepsilon \|h\|_{C^0[0, 1]} \end{aligned}$$

bei gleichmäßiger Stetigkeit von $\frac{\partial f}{\partial y}$ auf kompakten Mengen gilt. Dies bedeutet, dass die Gâteaux-Ableitung von F existiert und es ist

$$d_G F(\varphi; h) : x \mapsto \frac{\partial f}{\partial y}(x, \varphi(x))h(x). \quad (1.55)$$

Darüber hinaus ist $d_G F$ als Abbildung $X \rightarrow \mathcal{L}(X, X)$ stetig (wieder wegen der gleichmäßigen Stetigkeit von $\frac{\partial f}{\partial y}$) und wegen Satz 1.78 existiert auch die Fréchet-Ableitung $F'(\varphi)$ für jedes $\varphi \in X$.

Warnung: Es stimmt nicht immer, dass die Existenz der partiellen Ableitung $\frac{\partial f}{\partial y}$ die Existenz der Gâteaux-Ableitung von F nach sich zieht.

Wir betrachten beispielsweise den Raum

$$X := \{\varphi \in C^0[0, \infty) \mid \varphi(x)e^{-x} \text{ ist beschränkt auf } [0, \infty)\}$$

mit der Norm

$$\|\varphi\| := \sup_{x \in [0, \infty)} |\varphi(x)e^{-x}|$$

und es sei $f(y) = \sin y$. Unser Nemytski-Operator lautet: $F : X \rightarrow X$ mit $F(\varphi) := \sin \circ \varphi$. Würde nun die Beziehung (1.55) gelten, dann wäre $d_GF(0)h = \sin'(0)h$, $h \in X$. Die Stetigkeit von F folgt unter Verwendung der Lipschitz-Stetigkeit von f aus

$$\begin{aligned} \|F(\varphi_1) - F(\varphi_2)\| &= \sup_{x \in [0, \infty)} |\sin \varphi_1(x)e^{-x} - \sin \varphi_2(x)e^{-x}| \\ &\leq \sup_{x \in [0, \infty)} |(\varphi_1(x) - \varphi_2(x))e^{-x}| = \|\varphi_1 - \varphi_2\|. \end{aligned}$$

Nun überprüfen wir, ob $d_GF(0; h) = \sin'(0)h$ gilt. Für die Stelle $\varphi = 0$ ergibt sich die Abschätzung (es muss mit obiger Norm gearbeitet werden)

$$\begin{aligned} \left\| \frac{1}{t}(F(th) - F(0)) - d_GF(0; h) \right\| &\stackrel{(h(x)=e^x)}{=} \sup_{x \in [0, \infty)} \left| e^{-x} \left[\frac{\sin(te^x) - 0}{t} - 1 \cdot e^x \right] \right| \\ &= \sup_{y \in [t, \infty)} \left| \frac{\sin y}{y} - 1 \right| \geq 1 \text{ für jedes } t > 0. \end{aligned}$$

Der Ausdruck müsste für $t \rightarrow 0$ gegen Null konvergieren, was jedoch nicht eintritt. \square

Weitere Beispiele zum Nemytski-Operator finden sich z. B. in Drábek und Milota (2007) und Ambrosetti und Prodi (1993).

1.5 Implizites Funktionentheorem

Ein weites Anwendungsfeld der nichtlinearen Funktionalanalysis ist das Lösen von nichtlinearen Gleichungen der Form

$$F(x, y) = 0, \quad F : X \times Y \rightarrow Z, \quad X, Y, Z \text{ Banach-Räume.} \quad (1.56)$$

Man sucht nach Bedingungen, die die Auflösbarkeit der Gleichung (1.56) nach y in Abhängigkeit von x sichern. Kennzeichnend für diesen zentralen Satz der Analysis ist, dass sich die Auflösung der Gleichung (1.56) nach y im Wesentlichen an der Linearisierung der Funktion F in Bezug auf die Variable y erkennen lässt. Die Bedingung $F_y(x_0, y_0) \neq 0$ besagt im einfachen Fall, dass die Funktion $F : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ in dem Punkt $(x_0, y_0) \in \mathbb{R}^2$ keine vertikale Tangente besitzt.

Diese Situation ist bereits bei der Gleichungen $F(x, y) := x - y^2 = 0$, $(x, y) \in \mathbb{R}^2$ gestört. In der Umgebung des Lösungselements $(x_0, y_0) = (0, 0)$ besitzt für $x < 0$ die Gleichung *keine* reelle Lösung, für $x > 0$ *zwei* reelle Lösungen. Hier ist $F_y(0, 0) = 0$.

Es sei nun stets (x_0, y_0) ein Lösungselement von (1.56), d. h. es gilt $F(x_0, y_0) = 0$. Gesucht ist eine Abbildung $x \mapsto y^*(x)$ auf einer Umgebung von x_0 , so dass $y^*(x_0) = y_0$ und $F(x, y^*(x)) = 0$ in $U(x_0)$ gilt. Die entscheidende Bedingung für die Existenz genau einer Lösung (neben der Stetigkeit von F) ist die folgende:

$$F_y(x_0, y_0)^{-1} : Z \rightarrow Y \text{ existiert als linearer stetiger Operator.} \quad (1.57)$$

Da Y und Z als Banach-Räume vorausgesetzt sind, ist diese Bedingung äquivalent zu der folgenden:

$$\text{Die partielle } F\text{-Ableitung } F_y(x_0, y_0) : Y \rightarrow Z \text{ ist bijektiv.} \quad (1.58)$$

Um zu sehen, wie diese Bedingung wirkt, kann man sich $F(x, y)$ als Taylor-Reihe entwickelt denken:

$$F(x, y) = F(x_0, y_0) + a(x - x_0) + b(y - y_0) + \text{Terme höherer Ordnung}$$

Man beachte: $F(x_0, y_0) = 0$ und $F_y(x_0, y_0) = b$.

Der Auflösungsatz (siehe Satz 1.80) macht nun die Aussage, dass unter zusätzlichen Regularitätsbedingungen an F die Gleichung $F(x, y) = 0$ äquivalent zu

$$y - y_0 = -b^{-1}a(x - x_0) + \text{Terme höherer Ordnung}$$

ist. Die Bedingung (1.58) garantiert die Existenz der Inversen b^{-1} .

Satz 1.80 (Implizites Funktionentheorem)

Folgende Voraussetzungen seien erfüllt:

1. Die Abbildung $F : U(x_0, y_0) \subset X \times Y \rightarrow Z$ ist auf der offenen Umgebung $U(x_0, y_0)$ definiert, und es ist $F(x_0, y_0) = 0$, wobei X, Y und Z Banach-Räume über \mathbb{K} mit $\mathbb{K} = \mathbb{R}$ oder \mathbb{C} sind.
2. F und F_y sind stetig in (x_0, y_0) .
3. F_y existiert als partielle F -Ableitung auf $U(x_0, y_0)$ und die Bedingung (1.58) sei erfüllt.

Dann gelten folgende Aussagen:

- i. Es gibt positive Zahlen r_0 und r , so dass für jedes $x \in X$ mit $\|x - x_0\| \leq r_0$ genau ein $y^*(x) \in Y$ mit $\|y^*(x) - y_0\| \leq r$ und $F(x, y^*(x)) = 0$ existiert.
- ii. Die Folge der sukzessiven Approximationen $(y_n(x)) \subset Y$ mit

$$y_n(x) := y_{n-1}(x) - F_y(x_0, y_0)^{-1}F(x, y_{n-1}(x)), \quad n = 1, 2, \dots$$

$$y_0(x) := y_0$$

konvergiert gegen $y^*(x)$ in allen Punkten x mit $\|x - x_0\| \leq r_0$.

- iii. Ist F stetig in $U(x_0, y_0)$, dann ist y^* stetig auf einer Umgebung von x_0 . Ist F eine C^m -Abbildung, $1 \leq m < \infty$, in $U(x_0, y_0)$, dann ist y^* ebenfalls eine C^m -Abbildung auf einer Umgebung von x_0 . Ist F analytisch, so ist es auch y^* .

Der Satz wird in der modernen Literatur mit dem *Banachschen Fixpunktprinzip* bewiesen und kann in Zeidler (1986) nachgelesen werden. \square

Eine interessante Aussage ist die, dass im Raum $X \times Y$ durch den Punkt (x_0, y_0) eine „glatte Kurve“ läuft. Die Abbildung $y^* : K[x_0; r] \subset X \rightarrow Y$ genügt der Bedingung $y^*(x_0) = y_0$ und ist Fréchet-differenzierbar auf $K[x_0; r]$. Lösungen der Gleichung $F(x, y) = 0$ sind die Punkte

$$\{(x, y) \in X \times Y \mid y = y^*(x), x \in K[x_0; r]\} = \text{graph}(y^*),$$

denn es gilt $F(x, y^*(x)) = 0$ für alle $x \in K[x_0; r]$.

Bemerkung 1.81

Hat man das F als C^m -Abbildung ($m \in \mathbb{N}$) gegeben, dann kann man insbesondere die Ableitung von y^* berechnen. Aus $F(x, y^*(x)) = 0$ für alle $x \in U(x_0)$ folgt durch Differentiation nach x (an der Stelle x_0):

$$F_x(x_0, y_0) + F_y(x_0, y_0) \circ (y^*)'(x_0) = 0, \quad (y^*(x_0) = y_0).$$

Löst man diese Formel nach $(y^*)'$ auf, ergibt sich:

$$(y^*)'(x_0) = -F_y^{-1}(x_0, y_0) \circ F_x(x_0, y_0) \tag{1.59}$$

Das bemerkenswerte an der Formel (1.59) ist, dass man $(y^*)'(x_0)$ ausrechnen kann, ohne y^* explizit zu kennen, weil $y^*(x_0) = y_0$ ist. \blacksquare

Im Abschnitt 4.2 wird der Satz über implizite Funktionen auf die Poincaré-Abbildung (siehe Beispiel 4.23) angewendet.

1.6 Aufgaben

Aufgabe 1.1

Es sei X der Vektorraum aller auf dem Intervall $[0, 1]$ definierten Polynome höchstens zweiten Grades. Welche der nachstehenden Abbildungen $p_i \mid X \rightarrow \mathbb{R}$ ($i = a, \dots, e$) sind Normen auf X ?

- a)** $p_a(f) := \max_{x \in [0, 1]} \{|f(x)|\}$, **b)** $p_b(f) := \max_{x \in [0, 1]} \{|f(x)|\} + \max_{x \in [0, 1]} \{|f'(x)|\}$,
c) $p_c(f) := |f(0)|$, **d)** $p_d(f) := \int_0^1 |f(x)| dx$, **e)** $p_e(f) := |f(0)| + |f'(0)| + \frac{1}{2}|f''(0)|$.

Aufgabe 1.2

In einem normierten Vektorraum kann man in Analogie zum Zahlenraum \mathbb{R} oder \mathbb{C} den Begriff der unendlichen Reihe einführen. Es sei $(x_n)_{n \in \mathbb{N}}$ eine beliebige Folge aus einem normierten Raum $(X, \|\cdot\|)$. Dann heißt die aus (x_n) gebildete unendliche Reihe $\sum_{k=1}^{\infty} x_k$ *absolut konvergent*, falls die Zahlenreihe $\sum_{k=1}^{\infty} \|x_k\|$ konvergiert. Beweisen Sie:

- a) Jede absolut konvergente Reihe in einem Banach-Raum X ist auch konvergent. Kann man auf die Voraussetzung der Vollständigkeit des normierten Raumes X verzichten?
- b) Aus der Konvergenz der Summe $\sum_{k=1}^{\infty} \|x_{k+1} - x_k\|$ folgt stets, dass $(x_n)_{n \in \mathbb{N}}$ eine Cauchy-Folge in X ist.
- c) Ein normierter Raum X ist genau dann ein Banach-Raum, wenn für jede Folge $(x_n)_{n \in \mathbb{N}} \subset X$ mit der Eigenschaft $\|x_n\| \leq 2^{-n}$, $\forall n \in \mathbb{N}$, die Reihe $\sum_{k=1}^{\infty} x_k$ in X konvergiert.

Aufgabe 1.3

Beweisen Sie das *Rieszsche Lemma*: Ist Y ein echter abgeschlossener Unterraum des normierten Raumes X , so gibt es zu jeder Zahl $\eta \in (0, 1)$ einen Vektor $x_\eta \in X$ mit

$$\|x_\eta\| = 1 \quad \text{und} \quad \|x - x_\eta\| > \eta \quad \forall x \in Y.$$

Hinweis: Heuser (2006), S. 103.

Aufgabe 1.4

Beweisen Sie die folgenden Aussagen:

- a) Im \mathbb{R}^n sind alle Normen äquivalent.
- b) In einem endlichdimensionalen normierten Raum ist jede abgeschlossene und beschränkte Menge kompakt.
- c) Jeder endlichdimensionale normierte Raum ist vollständig.
- d) Die nachfolgenden beiden Aussagen d_1) und d_2) sind äquivalent:

d_1) X ist endlichdimensional.

d_2) Die abgeschlossene Einheitskugel in X ist kompakt.

Zeigen Sie ohne Zuhilfenahme dieses Sachverhaltes, dass die abgeschlossene Einheitskugel des l^p ($p \in [1, \infty)$, siehe Beispiel 1.2, 2.) nicht kompakt ist.

Hinweis: Heuser (2006).

Aufgabe 1.5

Betrachten Sie die Funktionenräume $\mathcal{C}^0[a, b]$ und $\mathcal{C}^1[a, b]$ stetiger bzw. einmal stetig differenzierbarer Funktionen auf $[a, b]$. $\mathcal{C}^0[a, b]$ ist mit der Maximumnorm $\|f\|_\infty := \max\{|f(t)| \mid t \in [a, b]\}$ ausgestattet.

Zeigen Sie: $\mathcal{C}^1[a, b]$ ist mit der Maximumnorm kein Banach-Raum. Ist $\mathcal{C}^1[a, b]$ mit der Norm $\|f\|_1 := \|f\|_\infty + \|f'\|_\infty$ ein Banach-Raum? Sind die Normen $\|f\|_\infty$ und $\|f\|_1$ im Raum $\mathcal{C}^1[a, b]$ äquivalent?

Aufgabe 1.6

Es sei $\mathcal{C}^0[0, 1]$ mit der Maximumnorm ausgestattet. Betrachten Sie die Teilmengen

$$X := \{f \in \mathcal{C}^0[0, 1] \mid f(0) = 0\} \quad \text{und} \quad X_0 := \left\{g \in X \mid \int_0^1 g(t) dt = 0\right\}.$$

Überlegen Sie sich zunächst, dass X_0 ein abgeschlossener Teilraum von X ist und zeigen Sie dann: Es existiert kein f mit $\|f\| = 1$, so dass für alle $g \in X_0$ gilt: $\|f - g\| \geq 1$.

Aufgabe 1.7

Zeigen Sie: $M := \{x = (x_1, x_2, \dots) \in l^2 \mid 0 \leq x_i \leq \frac{1}{i}, i \in \mathbb{N}\} \subset l^2$ ist kompakt.

Hinweis: Man verwende Bemerkung 1.7.

Aufgabe 1.8

Die Voraussetzungen des Satzes von Arzelà-Ascoli (siehe Satz 1.58) können nicht mehr abgeschwächt werden.

- a) Zeigen Sie, dass die Funktionenfolge $f_n(x) = \sin(nx)$ in $\mathcal{C}^0[0, 2\pi]$ beschränkt ist, aber keine konvergente Teilfolge enthält.
- b) Betrachten Sie die Funktionenfolge $(g_n) \subset \mathcal{C}^0[0, 1]$

$$g_n(x) := \frac{x^2}{x^2 + (1 - nx)^2}$$

mit der Maximum-Norm. Zeigen Sie, dass die Folge (g_n) beschränkt ist, und dass (g_n) sogar punktweise konvergiert, aber keine konvergente Teilfolge in $\mathcal{C}^0[0, 1]$ besitzt.

Aufgabe 1.9

Prüfen Sie nach, ob die Funktion

$$f: [0, 2\pi) \rightarrow S(0; 1), \quad x \mapsto (\cos x, \sin x) \tag{1.60}$$

ein Homöomorphismus ist.

Aufgabe 1.10

Es seien X ein normierter Raum und $f \in X^*$ gegeben. Beweisen Sie:

- a) Die Zahl $\|x\|_X := \sup \{|f(x)| \mid f \in X^* \text{ mit } \|f\|_{X^*} \leq 1\}$ ist eine Norm auf X .
- b) Gilt $f(x) = 0$ für jedes Funktional $f \in X^*$, dann ist $x = 0$.
- c) Ist $X_1 \subset X$ eine lineare Menge und $x_0 \in X$ ein Element mit dem Abstand $r > 0$ von X_1 ($d(x_0, X_1) := \sup_{x_1 \in X_1} \{d(x_0, x_1)\} = r$), dann gibt es ein Funktional f mit $f(x) = 0$ ($x \in X_1$), $\|f\| = 1$, $f(x_0) = r$.

Aufgabe 1.11

Beweisen Sie folgende Aussage: Es seien X, Y Banach-Räume, $T \in \mathcal{L}(X, Y)$ und $T \neq 0$ gegeben. Wenn T injektiv und der Wertevorrat $R(T)$ abgeschlossen sind, genau dann gibt es eine positive Konstante c , so dass

$$\|Tx\| \geq c\|x\| \quad \text{für alle } x \in X.$$

Aufgabe 1.12

Es sei $X := C^0[-1, 1]$ und ausgestattet mit der Norm $\|x\| := \max_{t \in [-1, 1]} |x(t)|$. Gegeben sei

$$T : X \rightarrow X, \quad (Tx)(t) := \int_{-1}^t x(s) ds, \quad t \in [-1, 1].$$

a) Ist T kompakt?

b) Ist das Bild $T(K[0; 1])$ abgeschlossen? ($K[0; 1]$ abgeschlossene Kugel in $C^0[-1, 1]$.)

Aufgabe 1.13

Beweisen Sie, dass für einen linearen und kompakten Operator $T : X \rightarrow X$ Folgendes gilt: $\dim N(I - T) < \infty$. ($N(T)$ bezeichnet den Nullraum von T .)

Aufgabe 1.14

Es seien

$$l^\infty := \{x = (x_1, x_2, \dots) \mid |x_i| \leq \text{const}, i \in \mathbb{N}\}$$

der Raum der beschränkten Zahlenfolgen mit der Norm $\|x\| := \sup_{i \in \mathbb{N}} |x_i|$ und der Operator

$$T : l^\infty \rightarrow l^\infty, \quad T(x_1, x_2, x_3, \dots) := \left(x_1, \frac{x_2}{2}, \frac{x_3}{3}, \dots\right)$$

gegeben. Zeigen Sie: T ist ein kompakter Operator und sein Wertevorrat $R(T)$ ist nicht abgeschlossen.

Aufgabe 1.15

Die Kompaktheit eines Operators kann auch noch durch Zuhilfenahme von Folgen charakterisiert werden (siehe Definition 1.55). Beweisen Sie folgende Aussage: Ein linearer Operator $T : X \rightarrow Y$ ist genau dann kompakt, wenn folgendes gilt: Aus jeder beschränkten Folge $(x_n)_{n \in \mathbb{N}} \subset X$ kann aus der Bildfolge $(T(x_n))_{n \in \mathbb{N}}$ eine konvergente Teilfolge ausgewählt werden.

Aufgabe 1.16

Zeigen Sie, dass die Inklusion $j : C^{k+1}[a, b] \rightarrow C^k[a, b]$ eine kompakte Abbildung ist. (Zum Begriff Inklusion vgl. man das Beispiel 1.61.)

Aufgabe 1.17

Es seien X, Y, Z normierte Vektorräume. Zeigen Sie: Falls $f : X \rightarrow Y$ F-differenzierbar, und falls $T : Y \rightarrow Z$ eine lineare, stetige Abbildung ist, dann gilt:

$$(T \circ f)' = T \circ f'.$$

Aufgabe 1.18

Es sei $X := C^0[0, 1]$ mit der sup-Norm ausgestattet. Weiter seien $t_i \in [0, 1]$ und $v_i \in C^0[0, 1]$ gegeben und $f : C^0[0, 1] \rightarrow C^0[0, 1]$ durch $f(x) := \sum_{i=1}^n x^2(t_i) \cdot v_i$ definiert. Zeigen Sie: f ist F-differenzierbar in allen Punkten von $x \in C^0[0, 1]$. Berechnen Sie die F-Ableitung $f'(x)$ für eine beliebige, aber fest gewählte Stelle $x \in C^0[0, 1]$.

Aufgabe 1.19

Es sei

$$f : \mathbb{R}^2 \rightarrow \mathbb{R} \quad \text{mit} \quad f(x, y) := \begin{cases} x^2(1 + \frac{1}{y}), & y \neq 0 \\ 0, & y = 0 \end{cases}$$

gegeben. Zeigen Sie: f besitzt im Punkt $(0, 0)$ keine Gateaux-Ableitung.

Aufgabe 1.20

Es sei G eine offene Menge in einem Banach-Raum X und $F : G \rightarrow X$ eine kompakte Abbildung. Beweisen Sie: Falls die Fréchet-Ableitung $F'(x_0)$ für ein $x_0 \in G$ existiert, dann ist der (lineare) Operator $F'(x_0)$ kompakt.

Aufgabe 1.21

Beweisen Sie die Behauptung: Falls f' als Gâteaux-Ableitung in einer Umgebung von x existiert und falls die Abbildung $x \mapsto f'(x)$ stetig in x ist, dann ist $f'(x)$ auch die Fréchet-Ableitung im Punkt x .

Hinweis: Überlegen Sie sich zunächst den folgenden Sachverhalt: Die Funktion $f : G \subset X \rightarrow Y$ sei eine stetig differenzierbare Funktion auf der offenen Menge G und X und Y seien Banach-Räume. Die Punkte x_0 und $x_0 + h$ liegen mit ihrer Verbindungsstrecke S in G . Dann gilt

$$f(x_0 + h) - f(x_0) = \int_0^1 f'(x_0 + th)h \, dt.$$

Das Resultat ergibt sich, indem man die auf $[0, 1]$ definierten Funktionen

$$F_1(s) := f(x_0 + sh), \quad F_2(s) = \int_0^s f'(x_0 + th)h \, dt$$

betrachtet.

Aufgabe 1.22

Stellen Sie Bedingungen auf, unter denen die Integralgleichung

$$x(s) - \mu \int_a^b K(s, t)G(t, x(t)) \, dt = 0$$

genau eine Lösung $x(s; \mu)$ hat.

Hinweis: Betrachten Sie die Abbildung

$$F : \mathcal{C}^0[a, b] \times \mathbb{R} \rightarrow \mathcal{C}^0[a, b], \quad F(x, \mu)(s) := x(s) - \mu \int_a^b K(s, t)G(t, x(t)) dt = 0$$

und setzen Sie voraus, dass $(x_0, \mu_0) \in \mathcal{C}^0[a, b] \times \mathbb{R}$ ein Lösungselement von $F(x, \mu) = 0$ ist. Des Weiteren seien die Funktionen

$$K : [a, b] \times [a, b] \rightarrow \mathbb{R}$$

$$G, G_x : [a, b] \times \mathbb{R} \rightarrow \mathbb{R}$$

stetig. Was muss über die Linearisierung von F in Bezug auf x im Punkt (x_0, μ_0) vorausgesetzt werden?

2 Gewöhnliche Differenzialgleichungen (DGL)

Übersicht

2.1	Einführende Beispiele	68
2.2	Geometrische Interpretation einer DGL.....	72
2.3	Existenz- und Eindeutigkeitssätze	73
2.4	Lineare DGL-Systeme 1. Ordnung	83
2.5	Lineare DGL-Systeme 1. Ordnung mit konstanten Koeffizienten.....	89
2.6	Autonome Systeme	98
2.7	Hilfsmittel zur Konstruktion von Phasenportraits	107
2.8	Aufgaben	122

Die Theorie der gewöhnlichen Differenzialgleichungen (im Weiteren mit DGL¹ abgekürzt) ist ein zentrales Gebiet der Analysis. Es bestehen viele Querverbindungen zu anderen Zweigen der Mathematik. Viele Prozesse in der Physik und den Ingenieurwissenschaften werden durch Differenzialgleichungen beschrieben und sie werden inzwischen auch erfolgreich in der Chemie, Biologie, Medizin, der Umwelt und den Wirtschaftswissenschaften – um nur einige zu nennen – angewendet.

Die wichtigsten Methoden, Sätze und Beweistechniken der allgemeinen Theorie der Anfangswertprobleme bei DGL werden vorgestellt. Dabei haben wir uns nicht nur auf elementare analytische Techniken und damit auf das reine Lösen von DGL, sondern auch auf qualitative oder geometrische Verfahren konzentriert. Diese führen oftmals zu einem besseren Verständnis des globalen Verhaltens von Lösungen.

Die breite Verfügbarkeit von Computern mit enormer Rechnerleistung, einschließlich vielseitiger symbolischer Rechenpakete wirft die Frage auf, ob es überhaupt noch sinnvoll ist, analytische elementare Lösungsverfahren zu behandeln. Aus unserer Sicht erfordert das Lösen einer schwierigen Problemstellung zum einen (oftmals) den Einsatz sowohl

¹Abkürzungen werden nicht dekliniert.

analytischer als auch numerischer Mittel. Zum anderen beruht die Durchführung eines effizienten numerischen Verfahrens vor allem auf vorbereitenden analytischen Untersuchungen, um qualitative Eigenschaften der Lösung zu bestimmen (z. B. Ermittlung von Parameterbereichen, die besondere Aufmerksamkeit erfordern). Ein weiterer Grund besteht darin, dass ein komplexer Prozess in der Natur oftmals nur dadurch verständlich wird, dass einfache und grundlegende Modelle zusammengefügt oder erweitert werden. Ziel dieses Kapitels ist es, dem Leser eine kurze, zeitgemäße, anschauliche und vergleichsweise leicht verständliche Einführung in die Theorie der DGL zu geben. Nicht immer kann man dabei die mathematische Strenge einhalten. Dort, wo Beweise zu den Sätzen fehlen, werden entsprechende Literaturangaben zur Nacharbeitung gebracht. Zu einigen Ergebnissen werden Beweise gebracht, um das mathematische Verständnis zu erhöhen. Die numerischen Verfahren zur Lösungen von DGL werden in einem späteren Kapitel abgehandelt. Für die „numerischen Kapitel“ werden hier die theoretischen Grundlagen gelegt. Die Lösungsmethoden und -techniken, die mittlerweile auch vom Computer beherrscht werden, stehen *nicht* im Vordergrund dieses Kapitels. Vielmehr ist es die Vermittlung des Wissens um die theoretischen Zusammenhänge, die man kennen muss, um die vom Computer gelieferten Daten und Bilder in geeigneter Weise zu interpretieren und mögliche Fehler zu erkennen. Die hier dargestellten Methoden verstehen sich als Brücke, die Lücken zwischen einem elementaren Kurs über DGL und der inzwischen sehr umfangreichen modernen Forschungsliteratur etwas zu schließen.

2.1 Einführende Beispiele

Die Grundgleichungen der verschiedenen physikalischen Disziplinen sind Differenzialgleichungen. Bereits in der Schule wird man mit dem *Newtonschen Bewegungsgesetz der Mechanik* für die Bewegung eines Punktes (z. B. eines Himmelskörpers) der Masse m vertraut gemacht:

$$m\ddot{x} = K(x, t). \quad (2.1)$$

Gesucht wird eine Bahnkurve $t \mapsto x(t)$ die der Gleichung (2.1) bei gegebener Kraft K genügt. In unserem Fall hängt die gesuchte Funktion nur von einer reellen Variablen t ab. Deshalb wird die zugehörige Differenzialgleichung als *gewöhnlich* bezeichnet.

Die Newtonsche Grundgleichung (2.1) beschreibt alle möglichen Bewegungen eines Punktes der Masse m unter dem Einfluss der Kraft K . Möchte man jedoch die Bahn eines bestimmten Himmelskörpers berechnen, dann muss man die Differenzialgleichung um ge-

wisse Informationen zu einer fest gewählten Anfangszeit $t = t_0$ ergänzen. Man hat dann folgendes Problem (Anfangswertproblem, siehe auch Definition 2.4) zu betrachten:

$$\begin{aligned} m\ddot{x} &= K(x, t) && \text{(Bewegungsgleichung)}, \\ x(t_0) &= x_0 && \text{(Anfangslage)}, \\ \dot{x}(t_0) &= v_0 && \text{(Anfangsgeschwindigkeit)}. \end{aligned}$$

Bei der mathematischen Behandlung von Modellgleichungen aus den Naturwissenschaften sind sehr oft Funktionen für Modellgrößen aus Gleichungen zu bestimmen, in denen nicht nur die gesuchte(n) Funktion(en), sondern auch gleichzeitig einige ihrer Ableitungen vorkommen.

Beispiel 2.1

1. Die lineare Pendelbewegung eines Federschwingers führt unter Zuhilfenahme des Newtonschen Kraftgesetzes und des Hookeschen Gesetzes zu der DGL

$$\ddot{s}(t) + \omega^2 s(t) = 0.$$

Die Funktion $t \mapsto s(t)$ beschreibt die Auslenkung des Massenpunktes aus der Ruhelage zum Zeitpunkt t und ω^2 ist eine Federkonstante (siehe Abbildung 2.1).

2. Die Stromstärke $I(t)$ in einem RL -Stromkreis (R Ohmscher Widerstand, L Induktivität) wird mittels Kirchhoffscher, Ohmscher und Faradayscher Gesetze aus der Modellgleichung

$$L\dot{I}(t) + RI(t) = U(t)$$

($U(t)$ angelegte Spannung) berechnet (siehe Abbildung 2.2).

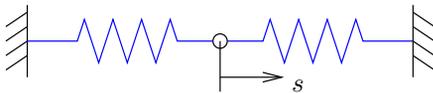


Abb. 2.1 Federschwinger

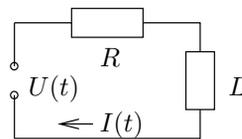


Abb. 2.2 RL -Kreis

3. Die logistische Funktion $p(t)$ beschreibt die zeitliche Entwicklung einer (biologischen oder ökonomischen) Population mit beschränktem Lebensraum:

$$\dot{p}(t) = p(t)(a - bp(t)), \quad a, b > 0 \text{ und konstant.} \quad \blacksquare$$

Anhand der Beispiele wird klar, was gesucht wird bzw. was man unter einer *Lösung einer DGL* versteht:

- Eine Lösung in Beispiel 2.1 ist eine 1- bzw. 2-mal differenzierbare Funktion, die, in die jeweilige DGL eingesetzt, die Gleichung identisch erfüllt.

- Treten in einer DGL auch höhere Ableitungen auf, etwa bis zur n -ten Ordnung, so spricht man von einer DGL n -ter Ordnung. Sie lässt sich immer in der Form

$$F(t, x, \dot{x}, \dots, x^{(n-1)}, x^{(n)}) = 0 \quad (2.2)$$

schreiben. Eine Lösung ist hier eine n -mal differenzierbare Funktion $t \mapsto \varphi(t)$, welche in F eingesetzt, die Gleichung (2.2) identisch erfüllt.

- Man nennt eine DGL n -ter Ordnung *explizit*, wenn sie nach der höchsten Ableitung aufgelöst ist, also die Form

$$x^{(n)} = f(t, x, \dot{x}, \dots, x^{(n-1)}) \quad (2.3)$$

besitzt, sonst *implizit* (siehe Gleichung (2.2)).

- Die Lösung einer DGL (2.2) bzw. (2.3) wird mit $t \mapsto \varphi(t)$, $t \in I$, $I \subset \mathbb{R}$ ein Intervall, bezeichnet und die Lösungsidentität in der Form

$$F(t, \varphi(t), \dot{\varphi}(t), \dots, \varphi^{(n-1)}(t), \varphi^{(n)}(t)) \equiv 0$$

bzw.

$$\varphi^{(n)}(t) \equiv f(t, \varphi(t), \dot{\varphi}(t), \dots, \varphi^{(n-1)}(t))$$

geschrieben. Natürlich ist stets zu beachten, dass

$$(t, \varphi(t), \dot{\varphi}(t), \dots, \varphi^{(n)}(t)) \text{ bzw. } (t, \varphi(t), \dots, \varphi^{(n-1)}(t)), \quad t \in I$$

im Definitionsbereich von F bzw. f liegt.

Bei einer Gleichung der Form

$$\dot{x} = f(t, x) \quad (2.4)$$

mit $(t, x) \in \Omega \subset \mathbb{R}^2$, $f : \Omega \rightarrow \mathbb{R}$, f stetig auf Ω , kann man sich auch geometrisch veranschaulichen, was es bedeutet, dass (2.4) eine Lösung $\varphi : I \rightarrow \mathbb{R}$ besitzt. Die Abbildung 2.3 macht auch klar (zumindest anschaulich), dass es gewiss nicht offensichtlich ist, ob es zu einer beliebig vorgegebenen Funktion $f : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ (f nicht notwendig stetig) eine auf einem Intervall I erklärte Funktion $\varphi : I \rightarrow \mathbb{R}$ gibt, die (2.4) identisch erfüllt.

Warum wir uns, wie in der Einleitung erwähnt, zunächst mehr mit den theoretischen Grundlagen beschäftigen wollen, soll das folgende Beispiel demonstrieren.

Beispiel 2.2

Für die völlig harmlos wirkende DGL

$$\dot{x} = x^2 - t$$

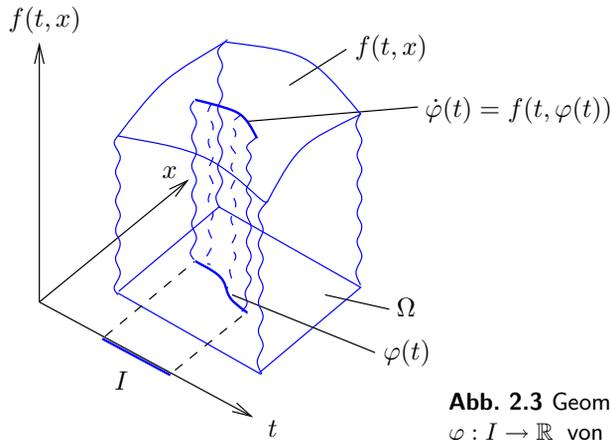


Abb. 2.3 Geometrische Interpretation der Lösung $\varphi : I \rightarrow \mathbb{R}$ von (2.4)

kann keine Lösung explizit angegeben werden, die sich als endliche Kombination von elementaren Funktionen und deren Integralen schreiben lässt. Diese etwas vage formulierte Aussage wurde von *Liouville*² gezeigt. Der Beweis ist algebraischer Natur und kann in diesem Rahmen nicht gegeben werden. Wir wenden uns in Beispiel 2.9 noch einmal dieser DGL zu. ■

Es gibt einfache Beispiele von impliziten DGL, die keine reellwertige Lösung $\varphi(\cdot)$ besitzen, z. B. $\dot{x}^2 + 1 = 0$. Auch bei expliziten DGL kommt dieser Sachverhalt vor.

Beispiel 2.3

Betrachtet wird

$$\dot{x} = d(t) \quad \text{mit} \quad d(t) := \begin{cases} 0, & t \text{ irrational} \\ 1, & t \text{ rational} \end{cases}, \quad t \in [0, 1].$$

Die Funktion $d(\cdot)$ wird *Dirichlet*³-Funktion genannt. Hätte diese DGL eine Lösung φ , dann müsste die Ableitung genau so aussehen, wie die Dirichlet-Funktion. Die Ableitung einer differenzierbaren Funktion muss zwar nicht notwendig stetig sein, aber sie besitzt die Zwischenwerteigenschaft (siehe Heuser (1980), Bd. I, S. 285). Danach muss eine differenzierbare Funktion φ mit $\varphi'(a) \neq \varphi'(b)$ ($a, b \in [0, 1]$) jeden Wert zwischen $\varphi'(a)$ und $\varphi'(b)$ annehmen, was aber bei dieser rechten Seite der DGL unmöglich ist. ■

Wir haben in den voranstehenden Beispielen gesehen, dass bei der Untersuchung von DGL die Frage nach der Existenz und Eindeutigkeit der Lösung eine wichtige Rolle spielt. In diesem Zusammenhang muss hier auf den Begriff eines *Anfangswertproblems* (AWP) eingegangen werden.

²Joseph Liouville (1809–1882), bedeutender französischer Mathematiker, lieferte Beiträge zur Algebra, Zahlentheorie, Geometrie und Analysis.

³Peter Gustav Lejeune–Dirichlet (1805–1859) war Nachfolger von Carl Friedrich Gauß in Göttingen, lieferte Beiträge zur Mathematik und mathematischen Physik.

Definition 2.4 (Anfangswertproblem (AWP))

Gegeben sei eine Menge $\Omega \subset \mathbb{R}^n$, $\Omega := \{x \in \mathbb{R}^n \mid \|x - x_0\| \leq d\}$, $I := \{t \in \mathbb{R} \mid |t - t_0| \leq a\}$ wobei d und a positive Konstanten sind, eine auf $I \times \Omega$ erklärte Funktion $f : I \times \Omega \subset \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$ und ein fester Punkt $(t_0, x_0) \in I \times \Omega$. Gesucht ist eine in einem Intervall J (mit $t_0 \in J$) differenzierbare Funktion $\varphi : J \rightarrow \mathbb{R}^n$ für die

$$\dot{\varphi}(t) = f(t, \varphi(t)), \quad t \in J \quad \text{und} \quad \varphi(t_0) = x_0 \quad (2.5)$$

gilt. Für diese Aufgabenstellung benutzen wir das Symbol

$$\dot{x} = f(t, x), \quad x(t_0) = x_0, \quad (2.6)$$

wobei $x(t_0) = x_0$ die Anfangsbedingung bezeichnet.

Natürlich ist in (2.5) wieder verlangt, dass $\text{graph } \varphi \subset I \times \Omega$ ist. Die Bezeichnung AWP stammt von Anwendungen, in denen $\varphi(t)$ und damit der Zustand des Systems (2.6) aus dem Anfangszustand $x(t_0) = x_0$ und einer DGL für die Zustandsänderung bestimmt werden soll. Für unseren Linearschwinger (Beispiel 2.1, 1.) ergibt sich das AWP

$$\ddot{x} + \omega^2 x = 0, \quad x(0) = x_{01}, \quad \dot{x}(0) = x_{02}.$$

Die eindeutige Lösung ist durch

$$\varphi(t; 0, x_0) := x_{01} \cos \omega t + \frac{x_{02}}{\omega} \sin \omega t, \quad (x_0 := (x_{01}, x_{02}))$$

gegeben. Die Lösung eines AWP (2.6) existiert nicht notwendig für alle Zeiten $t \geq t_0$ (siehe Beispiel 2.6).

2.2 Geometrische Interpretation einer DGL

Wir betrachten die explizite DGL 1. Ordnung

$$\dot{x} = f(t, x). \quad (2.7)$$

Dabei sei $f : \Omega \rightarrow \mathbb{R}$ als stetige Funktion auf $\Omega \subset \mathbb{R}^2$ erklärt. Die Funktion $\varphi : I \subset \mathbb{R} \rightarrow \mathbb{R}$ sei eine Lösung von (2.7), d. h. φ ist auf I differenzierbar, $\text{graph}(\varphi) \subset \Omega$ für alle $t \in I$ und es gilt $\dot{\varphi}(t) \equiv f(t, \varphi(t))$. Die geometrische Deutung von (2.7) ist nun einfach. Geht eine Integralkurve φ durch den Punkt $(t_0, x_0) \in \Omega$, d. h. ist $\varphi(t_0) = x_0$, so ist der Anstieg an dieser Stelle $\dot{\varphi}(t_0) = f(t_0, x_0)$. Durch die DGL (2.7) wird also die Steigung α der durch den Punkt (t_0, x_0) gehenden Lösungskurve vorgeschrieben: $\tan \alpha = f(t_0, x_0)$.

Diese Betrachtung lässt sich für jeden Punkt aus Ω anstellen. Sie führt auf die Begriffe *Linielement* und *Richtungsfeld* (siehe Abbildung 2.4).

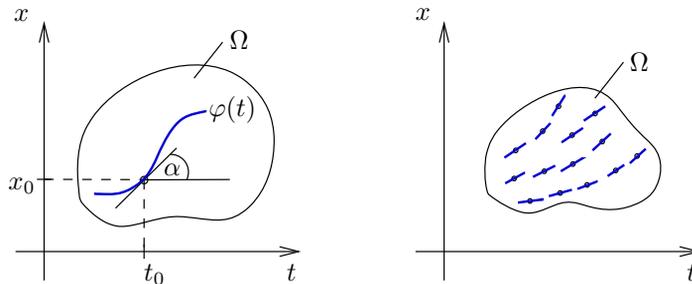


Abb. 2.4 Steigung und Linielement (links), Richtungsfeld (rechts)

Ein Zahlentripel (t, x, p) wird geometrisch so gedeutet, dass p den Anstieg einer durch den Punkt (t, x) gehenden Geraden angibt (α mit $\tan \alpha = p$ ist der Anstiegswinkel dieser Geraden). Dieses Tripel heißt *Linielement*. Gezeichnet wird dieses Linielement als ein (kleines) Geradenstück mit den genannten Eigenschaften. Die Gesamtheit aller Linielemente der Form $(t, x, f(t, x))$ ist ein *Richtungsfeld*. Der Zusammenhang zwischen dem Richtungsfeld $(t, x, f(t, x))$ und der DGL (2.7) ergibt sich wie folgt: Eine Lösung φ der DGL „passt“ auf das Richtungsfeld, d. h. in jedem Punkt des Graphen von φ stimmt die Tangentenrichtung von φ mit der Richtung des Linielementes überein.

Man kann sich also einen groben Überblick über den Verlauf der Lösungen dadurch verschaffen, dass man das Richtungsfeld zeichnet und dann versucht, Kurven auf das Richtungsfeld zu legen (siehe Abbildung 2.5). Ob durch jeden Punkt $(t_0, x_0) \in \Omega$ genau eine Lösungskurve φ geht, wird im nächsten Abschnitt erörtert.

Die soeben durchgeführte geometrische Betrachtungsweise ist umkehrbar. Ist eine Schar von differenzierbaren Kurven gegeben, deren Graphen paarweise disjunkt sind und die die Vereinigung Ω haben, so lässt sich dieser Kurvenschar eine DGL $\dot{x} = f(t, x)$ zuordnen, für die diese Kurven Lösungen der DGL sind. Dieser Zusammenhang vermittelt ein Bild von den Möglichkeiten, die bei DGL auftreten können und ist überdies nützlich für die Konstruktion von Beispielen. Ansonsten ist diese Betrachtungsweise mathematisch nicht sehr ergiebig und wird deshalb auch nicht weiter verfolgt.

2.3 Existenz- und Eindeigkeitsätze

Nach der Einführung in die grundlegenden Begriffsbildungen der Theorie gewöhnlicher Differenzialgleichungen geht es in diesem Abschnitt um die Aufgabe, die Existenz und Eindeutigkeit für Lösungen von Anfangswertproblemen zu klären. Die beiden grundle-

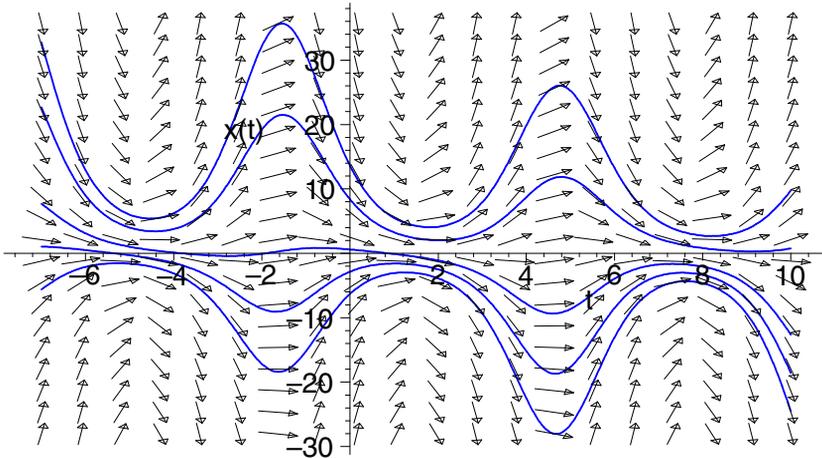


Abb. 2.5 Richtungsfeld und Kurvenschar der DGL $\dot{x} = -x \cos t - \sin t$

genden Sätze der Existenz von Lösungen sind der Satz von *G. Peano*⁴ und der Satz von *E. Picard*⁵ und *E. Lindelöf*⁶. Die beiden Sätze unterscheiden sich in ihrer Aussage lediglich in Bezug auf die Eindeutigkeit der Lösung, nicht was die Existenz der Lösung eines AWP betrifft. Die Problematik soll zunächst an Beispielen erläutert werden.

Beispiel 2.5

Wir suchen Lösungen der DGL

$$\dot{x} = \sqrt{|x|}. \quad (2.8)$$

Hier ist $\Omega = \mathbb{R}^2$. Aufgrund der Symmetrie des Richtungsfeldes ist mit φ auch die Funktion $\psi(t) := -\varphi(-t)$ eine Lösung von (2.8). In der Tat gilt

$$\dot{\psi}(t) = \dot{\varphi}(-t) = \sqrt{|\varphi(-t)|} = \sqrt{|-\varphi(-t)|} = \sqrt{|\psi(t)|}.$$

Es genügt also positive Lösungen zu betrachten. Durch die Lösungsmethode *Trennung der Variablen* (siehe z. B. Aulbach (1997)) ergibt sich

$$\int \frac{dx}{\sqrt{x}} = 2\sqrt{x} = t + C \quad \text{und damit}$$

$$x = \varphi(t; C) := \frac{(t + C)^2}{4}, \quad t \in (-C, \infty), \quad C \in \mathbb{R}. \quad (2.9)$$

⁴Giuseppe Peano (1858–1932), italienischer Mathematiker und Logiker.

⁵Emile Picard (1856–1941) arbeitete vor allem auf dem Gebiet der algebraischen Geometrie, der Analysis und der Mechanik.

⁶Ernst Leonhard Lindelöf (1870–1946), finnischer Mathematiker, lieferte wichtige Beiträge zur Funktionentheorie, Analysis und zur Theorie der Differenzialgleichungen.

Man beachte, dass \sqrt{x} positiv und somit $t > -C$ zu wählen ist. Die angegebene Funktion φ ist für $t < -C$ keine Lösung der DGL (2.8). Man überlege sich, dass mit (2.9) alle Lösungen von (2.8) gegeben sind. Ferner ist $\varphi(t) \equiv 0$ eine Lösung und $-\varphi(-t; C)$ sind die negativen Lösungen (siehe Abbildung 2.6).

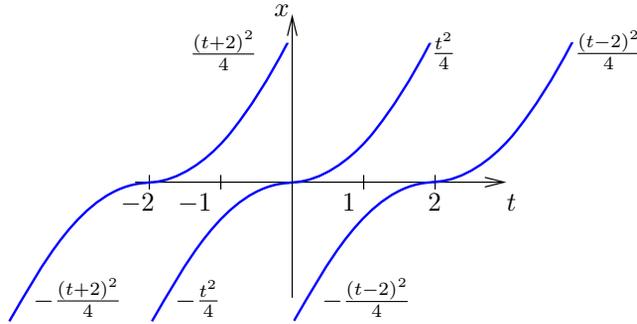


Abb. 2.6 Lösungskurven von $\dot{x} = \sqrt{|x|}$

Aus diesen Funktionen kann man die Lösungen zusammensetzen, die auf ganz \mathbb{R} erklärt sind:

$$\varphi(t) = \begin{cases} t^2/4, & \text{für } t > 0, \\ 0, & \text{für } -2 \leq t \leq 0, \\ -(t+2)^2/4, & \text{für } t < -2. \end{cases}$$

Der Leser möge sich überlegen, dass φ auch an den „Nahtstellen“ differenzierbar ist und der DGL genügt. ■

Mit diesem Beispiel begegnen wir dem Phänomen der nichteindeutigen Lösbarkeit einer DGL. Jedes AWP besitzt unendlich viele Lösungen. Beispielsweise sind Lösungen durch den Punkt $(t_0, x_0) = (2, 1)$ durch

$$\varphi_1(t; a) = \begin{cases} t^2/4, & \text{für } t > 0, \\ 0, & \text{für } a \leq t \leq 0, \\ -(t-a)^2/4, & \text{für } t < a \ (a < 0) \end{cases} \quad \text{und} \quad \varphi(t) = \begin{cases} t^2/4, & \text{für } t > 0, \\ 0, & \text{für } t \leq 0 \end{cases}$$

gegeben. Die Art der Mehrdeutigkeit ist aber verschieden, je nachdem ob der Anfangspunkt $x(t_0) = x_0$ Null oder verschieden von Null ist. Im ersten Fall verzweigen sich die Lösungen direkt an der Stelle (t_0, x_0) , im zweiten Fall hat man zunächst eine Lösung, die sich erst in einiger Entfernung von (t_0, x_0) verzweigt. Beim zuletzt vorliegenden Fall sagt man, dass das AWP lokal eindeutig lösbar sei. D. h. es gibt ein Intervall I ($t_0 \in I$) mit der Eigenschaft, dass eine auf I definierte, eindeutig bestimmte Lösung des AWP existiert. Man sagt, die AWP mit $x_0 \neq 0$ sind lokal eindeutig lösbar, die mit $x_0 = 0$ nicht.

Anhand des nächsten Beispiels soll auf ein weiteres wichtiges Phänomen aufmerksam gemacht werden.

Beispiel 2.6

Wir untersuchen die DGL

$$\dot{x} = e^x \sin t. \quad (2.10)$$

Auch hier ist wieder $\Omega = \mathbb{R}^2$. Das Richtungsfeld ist symmetrisch zur x -Achse und periodisch mit der Periode 2π , d. h. mit φ ist auch $\psi_1(t) = \varphi(-t)$ und $\psi_2(t) = \varphi(t + 2k\pi)$, $k \in \mathbb{Z}$ wieder Lösung. Durch Trennung der Variablen ergibt sich

$$\int e^{-x} dx = -e^{-x} = \int \sin t dt = -\cos t - C \quad \text{mit der Lösung}$$

$$\varphi(t; C) = -\ln(\cos t + C), \quad \cos t + C > 0 \quad (\text{siehe Abbildung 2.7}).$$

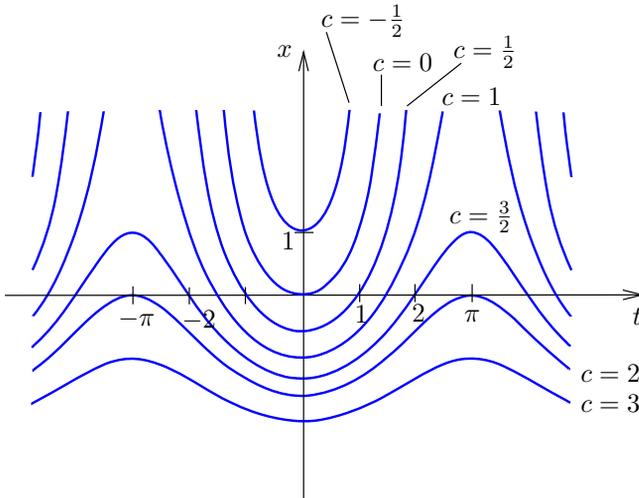


Abb. 2.7 Lösungskurven von $\dot{x} = e^x \sin t$

Man überlege sich wieder, dass damit alle Lösungen gefunden sind und dass jedes AWP eindeutig lösbar ist. ■

Dieses Beispiel zeigt ein neues Phänomen. Für $C > 1$ existieren die Lösungen für alle t und sind beschränkt. Im Fall $-1 < C \leq 1$ wachsen sie sehr rasch an und sind unbeschränkt. Wir betrachten das AWP (2.10) mit $x(0) = x_0$. Die Lösung dazu lautet

$$\varphi(t; 0, x_0) := \varphi(t; e^{-x_0} - 1) = -\ln(\cos t + e^{-x_0} - 1).$$

Speziell ergibt sich für $x_0 = -\ln 2$ die Lösung

$$\varphi(t; 0, -\ln 2) = -\ln(1 + \cos t).$$

Die Lösung existiert in $(-\pi, \pi)$ und ist nicht über dieses Intervall hinaus fortsetzbar. Weiter gilt

$$\lim_{t \rightarrow \pm\pi} \varphi(t; 0, -\ln 2) = +\infty.$$

Lösungen mit $x_0 < -\ln 2$ existieren dagegen in \mathbb{R} und sind beschränkt. Für $x_0 > -\ln 2$ existieren die Lösungen nur im Intervall $|t| < \arccos(1 - e^{-x_0})$, die Länge des Intervalles konvergiert gegen Null für $x_0 \rightarrow \infty$. Am Beispiel 2.6 sieht man, dass die Lösung eines AWP eventuell nur in einem sehr kleinen Intervall existiert und zwar auch dann, wenn die rechte Seite der DGL auf ganz \mathbb{R}^2 definiert und beliebig oft differenzierbar auf \mathbb{R}^2 ist.

Man kann allgemein zeigen, dass sich eine Lösung immer bis zum Rand von Ω (Ω Definitionsbereich der rechten Seite der DGL) fortsetzen lässt. Weiter sollte das Beispiel 2.6 zeigen, dass bei kleinen Änderungen der Anfangswerte – man betrachte etwa die Lösungen für die Anfangswerte (t_0, x_0) in der Nähe von $(0, -\ln 2)$ – sich das qualitative Verhalten von Lösungen (z. B. beschränkt, unbeschränkt) sprunghaft ändern kann. Die wichtige Frage, ob die Stetigkeit von f (rechte Seite einer DGL) bereits ausreicht, um die Existenz einer Lösung zu beweisen, wurde zuerst von G. Peano (um 1890) positiv beantwortet. Einen sehr eleganten Beweis mittels Schauderschem Fixpunktprinzips findet man z. B. in Zeidler (1995a), S. 63.

Satz 2.7 (Existenzsatz von Peano)

Ist f in einem Gebiet $\Omega \subset \mathbb{R}^{n+1}$ stetig, so geht durch jeden Punkt $(t_0, x_0) \in \Omega$ mindestens eine Lösung der DGL $\dot{x} = f(t, x)$. Jede Lösung lässt sich bis zum Rand von Ω fortsetzen.

Es stellt sich jetzt die Frage, welche Bedingungen an f zusätzlich gestellt werden müssen, um die Eindeutigkeit der Lösung von (2.6) durch den vorgegebenen Anfangspunkt (t_0, x_0) zu gewährleisten. Dass die Stetigkeit von f allein nicht ausreicht, hat Beispiel 2.5 gezeigt. Ein zentraler Satz innerhalb der Theorie der gewöhnlichen Differentialgleichungen ist der Existenz- und Eindeutigkeitsatz von E. Picard und E. Lindelöf. Er macht eine Aussage darüber, unter welchen Voraussetzungen das AWP (ein System von n gewöhnlichen Differentialgleichungen)

$$\dot{x} = f(t, x), \quad x(t_0) = x_0, \quad (t, x) \in \mathbb{R} \times \mathbb{R}^n \quad (2.11)$$

genau eine Lösung hat. Der Satz unterscheidet sich von dem Satz von Peano nur dadurch, dass zur Stetigkeit von f zusätzlich eine Lipschitz-Bedingung (2.12) hinzukommt und dafür die Eindeutigkeit der Lösung gefolgert werden kann.

Satz 2.8 (Picard und Lindelöf)

Es seien $t_0 \in \mathbb{R}$, $x_0 \in \mathbb{R}^n$ und

$$Q_{a,b} := \{(t, x) \in \mathbb{R} \times \mathbb{R}^n \mid |t - t_0| \leq a, \|x - x_0\| \leq b\}$$

mit festem $a, b > 0$ gegeben. Weiter wird vorausgesetzt, dass $f : Q_{a,b} \rightarrow \mathbb{R}^n$ stetig ist und

$$\|f(t, x) - f(t, y)\| \leq L\|x - y\| \quad \text{für alle } (t, x), (t, y) \in Q_{a,b} \quad (2.12)$$

mit konstantem $L \geq 0$ gilt. Wählt man c so, dass $0 < c \leq \min\{a, b/K\}$ mit $K := \max_{(t,x) \in Q_{a,b}} \|f(t, x)\|$, dann folgt:

- i. Existenz und Eindeutigkeit.** Das Anfangswertproblem (2.11) besitzt genau eine stetig differenzierbare Lösung $\varphi(\cdot)$ auf dem Intervall $[t_0 - c, t_0 + c]$.
- ii. Sukzessive Approximation.** Die Folge $(\varphi_n) \subset C^0[t_0 - c, t_0 + c]$ mit

$$\varphi_{n+1}(t) := x_0 + \int_{t_0}^t f(\tau, \varphi_n(\tau)) d\tau, \quad \varphi_0(t) \equiv x_0$$

konvergiert gleichmäßig auf $[t_0 - c, t_0 + c]$ für $n \rightarrow \infty$ gegen die Lösung $\varphi(\cdot)$.

- iii. Fehlerabschätzung.** Auf dem angegebenen Existenzintervall $[t_0 - c, t_0 + c]$ gilt für jedes $k \in \mathbb{N}_0$ die Fehlerabschätzung:

$$\|\varphi_k(t) - \varphi(t)\| \leq K L^k \frac{c^{k+1}}{(k+1)!}. \quad (2.13)$$

$\varphi_k(\cdot)$ bezeichnet die k -te Picard-Iterierte und $\varphi(\cdot)$ die Lösung des AWP (2.11).

- iv. Stetige Abhängigkeit vom Anfangswert.** Die Lösung $\varphi(\cdot)$ hängt stetig vom Anfangswert x_0 ab.

Beweis: Hier werden die Beweisschritte des Satzes nur grob skizziert. Dem Leser wird empfohlen, den Beweis des Satzes von Picard und Lindelöf mit Hilfe der Literatur (siehe z. B. Walter (1993) oder Aulbach (1997)) nachzuvollziehen. Der Beweis stellt gleichzeitig eine eindrucksvolle Anwendung des Banachschen Fixpunktsatzes dar.

Um den Banachschen Fixpunktsatz anwenden zu können, wird die Differenzialgleichung (2.11) in eine Integralgleichung

$$x(t) = x_0 + \int_{t_0}^t f(\tau, x(\tau)) d\tau, \quad t \in [t_0 - c, t_0 + c] \quad (2.14)$$

umgeformt. Da f stetig ist, sind (2.11) und (2.14) äquivalent: Die Gleichung (2.14) folgt durch Integration von (2.11) und (2.11) aus (2.14) durch Differentiation. Um beim Beweis die Aussage iv. gleich mit zu erhalten, fasst man (2.14) parameterabhängig in der Form

$$x(t) = p + \int_{t_0}^t f(\tau, x(\tau)) d\tau, \quad t \in [t_0 - c, t_0 + c] \quad (2.15)$$

auf, wobei (2.15) für $p = p_0 = x_0$ mit (2.14) übereinstimmt. Die zu betrachtende Fixpunktgleichung lautet nun

$$x = T_p x, \quad x \in M \subset X, \quad M := \{x \in X \mid \|x - p_0\| \leq b\},$$

$$(T_p x)(t) := p + \int_{t_0}^t f(\tau, x(\tau)) d\tau, \quad t \in [t_0 - c, t_0 + c].$$

Für X nimmt man den Banach-Raum $C^0[t_0 - c, t_0 + c]$. Mit diesem Operator T_p müssen alle Voraussetzungen des Banachschen Fixpunkttheorems (siehe Hoffmann et al. (2006), Satz 14.26 und Korollar 14.27) überprüft werden. M ist als abgeschlossene Teilmenge des Banach-Raumes X wieder vollständig. Die Lipschitz-Bedingung (2.12) sichert die Kontraktivität von T_p . Die Voraussetzungen $T_p(M) \subset M$ und T_p ist kontraktiv im Banachschen Fixpunktsatz, sind für die eventuelle Verkleinerung des Definitionsbereiches der Lösung von (2.11) verantwortlich. Das Maximum $\max_{(t,x) \in Q_{a,b}} \|f(t,x)\|$ existiert, da f auf der kompakten Menge $Q_{a,b}$ stetig ist (Satz von Weierstrass). Der Banachsche Fixpunktsatz liefert dann die Existenz und Eindeutigkeit der Lösung i., das Näherungsverfahren ii. und dessen Fehlerabschätzung iii., sowie die stetige Abhängigkeit vom Anfangswert. \square

Wir betrachten noch einmal die DGL aus Beispiel 2.2, deren Lösungen nachweislich nicht explizit angegeben werden können, für die man aber trotzdem die Existenz und Eindeutigkeit der Lösung auf einem Intervall I beweisen kann. Mit der Approximationsformel (2.13) kann der Fehler der Picard-Iterierten abgeschätzt werden. Die graphische Darstellung der Lösung und der Approximationsformeln im angegebenen Existenzintervall überlassen wir dem Leser.

Beispiel 2.9

Wir betrachten die DGL aus Beispiel 2.2 als nichtlineares AWP (DGL nach *F. Riccati*⁷ benannt)

$$\dot{x} = x^2 - t, \quad x(0) = 0. \quad (2.16)$$

Die Funktion $f(t,x) := x^2 - t$ ist für alle $(t,x) \in \mathbb{R} \times \mathbb{R}$ erklärt. Im Hinblick auf die Fehlerabschätzung (2.13) betrachten wir die Einschränkung der rechten Seite der DGL auf das Quadrat

$$Q_{\frac{1}{2}, \frac{1}{2}} := \{(t,x) \in \mathbb{R} \times \mathbb{R} \mid |t| \leq \frac{1}{2}, |x| \leq \frac{1}{2}\} = \left[-\frac{1}{2}, \frac{1}{2}\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right].$$

Dort gilt die Abschätzung

$$|f(t,x) - f(t,y)| = |(x^2 - t) - (y^2 - t)| = |(x+y)(x-y)| \leq \left(\frac{1}{2} + \frac{1}{2}\right)|x-y|$$

⁷Francesco Riccati (1676–1754). Neben seinen Untersuchungen zu der nach ihm benannten DGL $\dot{x} = a(t) + b(t)x + c(t)x^2$, widmete er sich den Anwendungen der Analysis auf Probleme der Hydrodynamik und Optik.

mit der möglichen Wahl der Lipschitz-Konstanten $L := 1$. Um die Größe des Lösungsintervalls für die DGL (2.16) zu bekommen, schätzen wir f auf der Menge $Q_{\frac{1}{2}, \frac{1}{2}}$ nach oben ab:

$$|x^2 - t| \leq |x|^2 + |t| \leq \frac{3}{4}.$$

Da die Schranke $\frac{3}{4}$ als Wert angenommen wird, gilt die Aussage (i) des Satzes 2.8 mit $a = \frac{1}{2}$, $b = \frac{1}{2}$ und $K = \frac{3}{4}$. Wegen $c := \min\{\frac{1}{2}, \frac{1}{2} \cdot \frac{4}{3}\} = \frac{1}{2}$ liefert der Satz dann die Existenz einer eindeutig bestimmten Lösung (mindestens) auf dem Intervall $I = [-\frac{1}{2}, \frac{1}{2}]$. Mit Hilfe der Picard-Iteration bestimmen wir die ersten vier Näherungslösungen. Wir starten mit $\varphi_0 = 0$ und erhalten die weiteren Näherungen gemäß sukzessiver Approximation $\varphi_{n+1}(t) = x_0 + \int_0^t f(\tau, \varphi_n(\tau)) d\tau$, $n \in \mathbb{N}_0$:

$$\varphi_1(t) = \int_0^t -\tau d\tau = -\frac{t^2}{2},$$

$$\varphi_2(t) = \int_0^t \left(\frac{\tau^4}{4} - \tau\right) d\tau = -\frac{t^2}{2} + \frac{t^5}{20},$$

$$\varphi_3(t) = \int_0^t \left[\left(-\frac{\tau^2}{2} + \frac{\tau^5}{20}\right)^2 - \tau\right] d\tau = -\frac{t^2}{2} + \frac{t^5}{20} - \frac{t^8}{160} + \frac{t^{11}}{4400}.$$

Mit Hilfe der Fehlerabschätzung (2.13) können wir die Genauigkeit angeben, mit der durch φ_k die Lösung φ approximiert wird. Für die k -te Näherung lautet die Fehler-schranke $\frac{3}{4} \frac{1}{2^{k+1}(k+1)!}$. Auf dem Intervall $[-\frac{1}{2}, \frac{1}{2}]$ ergibt sich somit für φ_3 ein Fehler, der kleiner als $\frac{3}{4} \frac{1}{2^{4!}} \approx 2 \cdot 10^{-3}$ ist. ■

Aus dem Satz von Picard und Lindelöf lassen sich nun weitere Folgerungen ziehen, die hier nur verbal beschrieben und nicht bewiesen werden. Der Leser möge die umfangreiche Literatur (z. B. Amann (1983), Coddington und Levinson (1955)) konsultieren.

Zunächst sei noch einmal festgehalten: Die lokale Unität der Lösungen impliziert sofort eine globale Unität in dem Sinn, dass sich Integralkurven in Ω nicht schneiden oder verzweigen können. Anderenfalls würden in einer Umgebung des Schnittpunktes (t_0, x_0) zwei verschiedene Lösungen existieren.

Eine globale Existenzaussage etwa, dass die Lösung zum Anfangswert (t_0, x_0) auf beliebig vorgegebenem I mit $t_0 \in I$ existiert, ist auch unter schärferen Voraussetzungen an f (etwa Analytizität) nicht zu erzielen (siehe Beispiel 2.6). Erst starke Wachstumsbeschränkungen an f sichern globale Existenzaussagen (siehe Satz 2.11). Es gibt aber einen Satz über das *Maximalintervall* eines AWP. Unter den Voraussetzungen des Satzes von Picard und Lindelöf existiert zu jedem $(t_0, x_0) \in \Omega$ ein eindeutig bestimmtes Intervall $I(t_0, x_0) = (a(t_0, x_0), b(t_0, x_0))$ mit folgenden Eigenschaften:

- i. $-\infty \leq a(t_0, x_0) < t < b(t_0, x_0) \leq +\infty$,
- ii. das AWP (2.11) besitzt genau eine Lösung φ auf $I(t_0, x_0)$,

- iii. ist φ_1 eine Lösung des AWP (2.11) auf einem Intervall I mit $t_0 \in I$, dann ist $I \subset I(t_0, x_0)$ und φ_1 ist die Einschränkung der Lösung φ auf dieses Teilintervall I : $\varphi_1(\cdot; t_0, x_0) = \varphi(\cdot; t_0, x_0)|_I$.

Beispiel 2.10

Für einen beliebigen Anfangswert $(t_0, x_0) \in \mathbb{R}^2$ soll die maximale Lösung des AWP

$$\dot{x} = tx^2, \quad x(t_0) = x_0 \tag{2.17}$$

bestimmt werden. Durch Trennung der Variablen erhält man als Lösung φ die rationale Funktion

$$\varphi(t; t_0, x_0) := \frac{2x_0}{2 + x_0(t_0^2 - t^2)}.$$

Das maximale Lösungsintervall $I(t_0, x_0)$ ergibt sich aus der Tatsache, dass sich eine rationale Funktion nicht über ihre Pole hinweg als differenzierbare Funktion und damit nicht als Lösung einer DGL fortsetzen lässt. Daraus folgt (in Abhängigkeit von t_0 und x_0) das jeweilige Maximalintervall. Dabei muss t_0 in $I(t_0, x_0)$ liegen! Bei dieser einfachen DGL treten bereits Lösungskurven mit maximalen beidseitig und einseitig unbeschränktem und beschränktem Definitionsbereich auf. Der Leser möge einige Lösungen selbst erstellen. ■

Im Zusammenhang mit Eigenschaften von maximalen Lösungen tritt die Frage auf, wie sich eine maximale Lösung verhält, wenn t dem Rand des maximalen Existenzintervalles zustrebt. Es wird auch hier wieder nur eine verbale und anschauliche Erläuterung gegeben. Legt man auf mathematisch präzise Formulierungen und Beweisführung wert, dann sei auf das Buch von Aulbach (1997) verwiesen. Die Abbildung 2.8 macht das Verhalten der maximalen Lösungskurve φ_{\max} sichtbar: Die Menge $\Omega \subset \mathbb{R}^{1+n}$ sei offen und die

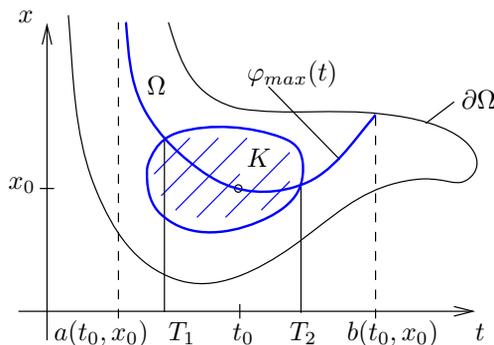


Abb. 2.8 Verhalten der Lösungskurve φ_{\max}

Funktion $f : \Omega \rightarrow \mathbb{R}^n$ sei stetig und bzgl. x Lipschitz-stetig (siehe (2.12)). Dann verlässt die Lösungskurve des AWP (2.11) einer maximalen Lösung jede kompakte Teilmenge K

von Ω in beiden Zeitrichtungen und kehrt nicht mehr in K zurück (einer der beiden Randpunkte von $I(t_0, x_0)$ ist dabei natürlich endlich). Man sagt, dass die Lösungskurve von Rand zu Rand läuft, also in beiden Zeitrichtungen gegen $\partial\Omega$ konvergiert. Den Zeitpunkt $a(t_0, x_0)$ bzw. $b(t_0, x_0)$ bezeichnet man auch als *Explosionszeit* (*endliche Entweichzeit*) der Lösung $\varphi_{\max}(t)$. In der Abbildung 2.8 ist es $t = a(t_0, x_0)$.

Der Leser möge am Beispiel der DGL

$$\dot{x} = \frac{x^2}{t}, \quad x(1) = 1, \quad \Omega := \{(t, x) \in \mathbb{R}^2 \mid t \neq 0\}$$

den anschaulich erklärten Sachverhalt nachvollziehen.

Die DGL $\dot{x} = e^x \sin t$ (siehe Beispiel 2.6) zeigt, dass selbst bei einem Definitionsbereich $\Omega = \mathbb{R}^2$ endliche Entweichzeiten auftreten können. Damit stellt sich die Frage: Gibt es (möglichst allgemeine) Bedingungen an die rechte Seite einer DGL, die das Auftreten endlicher Entweichzeiten verhindern?

Satz 2.11 (DGL mit linear beschränkter rechter Seite)

Es sei $f : \Omega \rightarrow \mathbb{R}^n$ mit $\Omega := I \times \mathbb{R}^n$, $I = (a, b)$, $-\infty \leq a < b \leq \infty$ eine stetige Funktion auf Ω und bzgl. $x \in \mathbb{R}^n$ Lipschitz-stetig, d. h. (2.12) gilt mit einem $L > 0$. Des Weiteren sei f linear beschränkt, d. h. es gilt eine Abschätzung der Form

$$\|f(t, x)\| \leq r(t)\|x\| + s(t) \quad \text{für alle } t \in I \text{ und } x \in \mathbb{R}^n \quad (2.18)$$

mit stetigen Funktionen $r, s : I \rightarrow \mathbb{R}^+$. Dann existiert die maximale Lösung jedes AWP

$$\dot{x} = f(t, x), \quad x(t_0) = x_0, \quad (t_0, x_0) \in I \times \mathbb{R}^n$$

auf dem ganzen Intervall I , d. h. es gilt $I(t_0, x_0) = I$ für jedes $t_0 \in I$ und $x_0 \in \mathbb{R}^n$.

Im Hinblick auf den letzten Satz kann man feststellen, dass die t -Abhängigkeit von f gegenüber der x -Abhängigkeit eine untergeordnete Rolle spielt. In der Abschätzung (2.18) für die Funktion f sind beliebige (lediglich stetige) Funktionen r und s zugelassen.

Beispiel 2.12

Die DGL

$$\dot{x} = t^4 x \sin x + e^{t^4}, \quad (t, x) \in \mathbb{R} \times \mathbb{R} \quad (2.19)$$

hat eine stetig differenzierbare rechte Seite, die darüber hinaus auch noch linear beschränkt ist. Wählt man nämlich $r(t) := t^4$ und $s(t) := e^{t^4}$ dann genügt die rechte Seite von (2.19) den Voraussetzungen des Satzes 2.11 (insbesondere der Abschätzung (2.18)) und man hat somit eine eindeutig bestimmte maximale Lösung. Wegen des starken Wachstums des Terms e^{t^4} verlaufen die Lösungskurven (von kleinen $|t|$ abgesehen) nahezu senkrecht. Die Abbildung 2.9 (rechts) suggeriert das Vorliegen einer endlichen

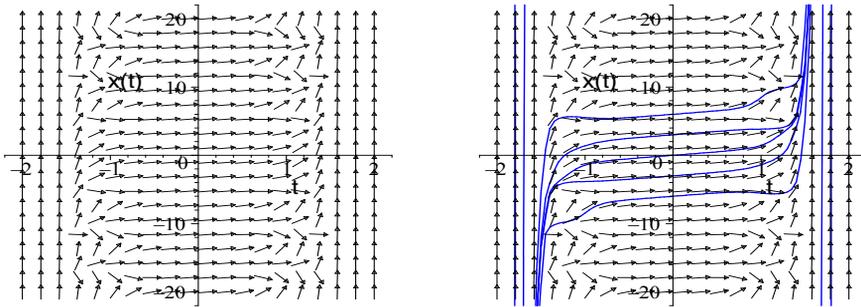


Abb. 2.9 Richtungsfeld der DGL $\dot{x} = t^4 x \sin x + e^{t^4}$ (links). Verschiedene Lösungskurven der DGL $\dot{x} = t^4 x \sin x + e^{t^4}$; keine dieser Lösungskurven hat eine endliche Entweichzeit (rechts)

Entweichzeit, was jedoch aufgrund des Satzes 2.11 nicht möglich ist, da die Lösung für alle $t \in \mathbb{R}$ existiert. ■

Beispiel 2.13

Wir betrachten die DGL

$$\dot{x} = A(t)x + b(t) \tag{2.20}$$

wobei $A(t)$ eine $n \times n$ -Matrix mit stetigen Einträgen $a_{ij} : I \rightarrow \mathbb{R}$, $i, j = 1, \dots, n$ und $b : I \rightarrow \mathbb{R}^n$ stetig auf $I := (a, b)$ sind. Für $f(t, x) := A(t)x + b(t)$ gilt die Abschätzung

$$\|f(t, x)\| \leq \|A(t)\|_{\mathbb{R}^n \times n} \|x\|_{\mathbb{R}^n} + \|b(t)\|_{\mathbb{R}^n} \tag{2.21}$$

mit irgend einer Norm im \mathbb{R}^n und einer Matrixnorm im $\mathbb{R}^{n \times n}$. Somit ist die Abschätzung (2.18) mit $r(t) := \|A(t)\|$ und $s(t) := \|b(t)\|$ erfüllt und nach Satz 2.11 existiert die maximale Lösung φ stets auf dem gemeinsamen Definitionsintervall von $A(\cdot)$ und $b(\cdot)$. Es treten also, sofern $A(\cdot)$ und $b(\cdot)$ stetig sind, bei linearen Systemen (2.20) keine Explosionszeiten auf. ■

2.4 Lineare DGL-Systeme 1. Ordnung

Eine Reihe von naturwissenschaftlichen Problemstellungen, wie z. B. Feder-Dämpfer-Systeme der Mechanik oder Kirchhoffsche Netzwerke der Elektrotechnik, lassen sich durch lineare DGL-Systeme beschreiben. D. h. sie lassen sich auf die Form

$$\dot{x} = A(t)x + b(t), \quad x \in \mathbb{R}^n, \quad A \in \mathcal{C}^0(I, \mathbb{R}^{n \times n}), \quad b \in \mathcal{C}^0(I, \mathbb{R}^n), \quad n \in \mathbb{N} \tag{2.22}$$

bringen. Man nennt $A(t) := (a_{ij}(t)) \in \mathbb{R}^{n \times n}$ die Koeffizientenmatrix und den Funktionsvektor $b(t) := (b_1(t), \dots, b_n(t))^T$ die Störfunktion (auch Inhomogenität). Die Funktionen $a_{ij}, b_i : I \rightarrow \mathbb{R}$ sind über einem gemeinsamen Intervall $I \subset \mathbb{R}$ definiert und stetig.

Das System (2.22) heißt homogen, wenn $b(t) = 0$ für alle $t \in I$, sonst inhomogen. Das DGL-System

$$\dot{x} = A(t)x \quad (2.23)$$

ist das dem DGL-System (2.22) zugeordnete lineare homogene DGL-System. Da nur Ableitungen 1. Ordnung in (2.22) bzw. (2.23) vorkommen, handelt es sich um ein DGL-System 1. Ordnung.

Das wesentliche Merkmal linearer Systeme ist die algebraische Struktur der Gleichungen, sowie der dazugehörige Lösungsraum. Aufgrund von Satz 2.11 wissen wir bereits, dass für jedes $(t_0, x_0) \in I \times \mathbb{R}^n$ das AWP

$$\dot{x} = A(t)x + b(t), \quad x(t_0) = x_0 \quad (2.24)$$

eine eindeutig bestimmte globale Lösung

$$\varphi(\cdot; t_0, x_0) : I \rightarrow \mathbb{R}^n$$

besitzt. Somit können also im Inneren des Intervalls I keine Explosionszeiten von Lösungen von (2.22) auftreten.

Wir stellen nun einige weitere wohl bekannte Tatsachen über lineare DGL-Systeme zusammen.

(1) Superpositionsprinzip

Es seien die $\varphi_i : I \rightarrow \mathbb{R}^n$ Lösungen des DGL-Systems (2.22) mit den Inhomogenitäten $b_i \in \mathcal{C}^0(I, \mathbb{R}^n)$, $i = 1, 2$. Dann ist für beliebige reelle c_1, c_2 die auf I erklärte Funktion $c_1\varphi_1(t) + c_2\varphi_2(t)$ eine Lösung von

$$\dot{x} = A(t)x + c_1b_1(t) + c_2b_2(t).$$

Insbesondere ist mit je zwei Lösungen von (2.23) auch jede Linearkombination wieder eine Lösung dieser Gleichung. Diese Überlagerung von Lösungen nennt man *Superposition*.

(2) Lösungsraum

Für ein DGL-System 1. Ordnung soll nach der Menge aller Lösungen gefragt werden. Wir führen dazu einen linearen Operator L ein, der den reellen Vektorraum $\mathcal{C}^1(I, \mathbb{R}^n)$ in den Vektorraum $\mathcal{C}^0(I, \mathbb{R}^n)$ nach der Vorschrift

$$L : \mathcal{C}^1(I, \mathbb{R}^n) \rightarrow \mathcal{C}^0(I, \mathbb{R}^n), \quad x \mapsto Lx := \left(\frac{d}{dt} - A(\cdot) \right) x$$

abbildet. Damit kann (2.22) als lineare Operatorgleichung

$$Lx = b \quad \text{mit} \quad x \in \mathcal{C}^1(I, \mathbb{R}^n) \quad (2.25)$$

und gegebenem $b \in \mathcal{C}^0(I, \mathbb{R}^n)$ geschrieben werden. Wir führen zunächst noch die folgenden Lösungsmengen

$$\mathcal{L}_0 := \{x \in \mathcal{C}^1(I, \mathbb{R}^n) \mid Lx = 0\} = \ker(L)$$

(Lösungsmenge des linearen homogenen DGL-Systems) und

$$\mathcal{L} := \{x \in \mathcal{C}^1(I, \mathbb{R}^n) \mid Lx = b, \quad b \in \mathcal{C}^0(I, \mathbb{R}^n) \text{ gegeben}\}$$

(Lösungsmenge des linearen inhomogenen DGL-Systems) ein, um die Aussage präzise formulieren zu können. Dann ist $\mathcal{L}_0 \subset \mathcal{C}^1(I, \mathbb{R}^n)$ ein reeller Vektorraum und

$$\mathcal{L} = x_p + \mathcal{L}_0 := \{x = x_p + x_0 \mid \text{für jede Lösung } x_p \text{ von } Lx = b, x_0 \in \mathcal{L}_0\}.$$

Die Linearität von \mathcal{L}_0 rechnet man einfach nach. Die nächste Feststellung ist die, dass mit zwei (beliebigen) Lösungen $x_1, x_2 \in \mathcal{L}$ zunächst $Lx_1 = b$ und $Lx_2 = b$ und daraus $L(x_1 - x_2) = 0$ folgt. Mit einer partikulären fest gewählten Lösung $x_p \in \mathcal{L}$ folgt, dass für jedes $x \in \mathcal{L}$ die Differenz $x - x_p =: x_0$ zu \mathcal{L}_0 gehört. D. h. jedes $x \in \mathcal{L}$ hat die Darstellung $x = x_p + x_0$ und somit ist $\mathcal{L} \subset x_p + \mathcal{L}_0$. Umgekehrt gilt mit festem $x_p \in \mathcal{L}$ und beliebigem $x_0 \in \mathcal{L}_0$ die Relation $L(x_p + x_0) = Lx_p + Lx_0 = b$, also $x_p + x_0 \in \mathcal{L}$ und somit $x_p + \mathcal{L}_0 \subset \mathcal{L}$. Diese hier durchgeführten Überlegungen gelten übrigens auch ganz analog für jeden linearen Operator L , der eine Abbildung $X \ni x \mapsto Lx = y \in Y$ zwischen Vektorräumen X und Y vermittelt.

Aus diesen Betrachtungen resultiert folgendes Lösungsverfahren für die Gleichung (2.22): Man bestimme zunächst die Lösungsmenge \mathcal{L}_0 des homogenen Systems $\dot{x} = A(t)x$. Darüber erfolgen im Punkt (3) weitergehende Untersuchungen. Dann bestimme man eine partikuläre Lösung der inhomogenen Gleichung $Lx = b$. Dies kann mittels spezieller Lösungsansätze bzw. in jedem Fall mit der Methode der *Variation der Konstanten* erfolgen (siehe Punkt (6)). Die allgemeine Lösung von (2.25) ergibt sich dann aus $x = x_0 + x_p$.

(3) Dimension des Lösungsraumes \mathcal{L}_0

Wir wollen zeigen, dass $\dim \ker(L) = n$ gilt, d. h. es gibt genau n linear unabhängige Lösungen des linearen homogenen DGL-Systems (2.23). Man benötigt also lediglich die Konkretisierung des Begriffes der *linearen Unabhängigkeit* von Funktionen in einem Funktionenraum (Vektorraum). In unserem Fall ist dies der Raum $\mathcal{C}^1(I, \mathbb{R}^n)$. Für die lineare Unabhängigkeit von Funktionen müssen wir überprüfen: Das Nullelement lässt sich nur als triviale Linearkombination darstellen, also

$$\{\text{für alle } t \in I \text{ gilt } \sum_{i=1}^n c_i x_i(t) = 0\} \quad \Rightarrow \quad \{c_i = 0, \quad i = 1, \dots, n\}.$$

Wir machen zunächst eine weitere Beobachtung: Ist $\varphi(\cdot)$ eine Lösung von $Lx = 0$ und existiert ein $t_0 \in I$ mit $\varphi(t_0) = 0$, so ist $\varphi(t) = 0$ auf I . Dies folgt unmittelbar aus der Unität der Lösung, da $Lx = 0$ stets die triviale Lösung besitzt.

Um $\dim \ker(L) = n \geq 1$ zu zeigen, müssen wir einen geeigneten Vektorraum-Isomorphismus zwischen \mathbb{R}^n und einem Vektorraum V von $\mathcal{C}^1(I, \mathbb{R}^n)$ finden. Für jedes feste $t_0 \in I$ wird dies durch die Abbildung

$$\mathbb{R}^n \ni x_0 \mapsto \varphi(\cdot; t_0, x_0) \in \mathcal{L}_0 \subset \mathcal{C}^1(I, \mathbb{R}^n)$$

von \mathbb{R}^n auf \mathcal{L}_0 geleistet. Dabei ist $\varphi(\cdot; t_0, x_0)$ die (eindeutige) Lösung des AWP $\dot{x} = A(t)x$, $x(t_0) = x_0$. Diese Abbildung ist linear, denn für $\lambda, \mu \in \mathbb{R}$ und $x_0, x_1 \in \mathbb{R}^n$ folgt aus der eindeutigen Lösbarkeit

$$\varphi(\cdot; t_0, \lambda x_0 + \mu x_1) = \lambda \varphi(\cdot; t_0, x_0) + \mu \varphi(\cdot; t_0, x_1).$$

Diese Gleichheit gilt, da auf beiden Seiten der Gleichung eine Lösung von $Lx = 0$ mit dem Anfangswert $\lambda x_0 + \mu x_1$ (zum Zeitpunkt $t = t_0$) steht (Superpositionsprinzip). Aufgrund der Eindeutigkeit folgt aus $\varphi(\cdot; t_0, x_0) = 0$ stets $\varphi(t_0; t_0, x_0) = x_0 = 0$. Somit ist die Abbildung injektiv, die Surjektivität ist offensichtlich. Damit ist gezeigt: Die Abbildung ist ein Vektorraum-Isomorphismus von \mathbb{R}^n auf sein Bild \mathcal{L}_0 . Es gilt also $\dim(\mathcal{L}_0) = \dim \ker(L) = n$.

(4) Fundamentalmatrix

Zunächst sei noch einmal festgestellt: Jede Linearkombination von Lösungen von $Lx = 0$ ist wieder eine Lösung von $Lx = 0$. Es gibt nach (3) genau $n = \dim \mathbb{R}^n$ linear unabhängige Lösungen $\varphi_1, \dots, \varphi_n \in \mathcal{C}^1(I, \mathbb{R}^n)$ von $Lx = 0$. Jedes System $\{\varphi_1, \dots, \varphi_n\}$ von n linear unabhängigen Lösungen heißt *Fundamentalsystem* von $Lx = 0$. Es sei $\{\varphi_1, \dots, \varphi_n\}$ ein Fundamentalsystem von (2.23). Dann heißt die Matrix X mit den Spalten $\varphi_1, \dots, \varphi_n$, also

$$X(t) := [\varphi_1(t) \dots \varphi_n(t)]$$

Fundamentalmatrix. Gilt außerdem $X(t_0) = I$, d. h. $\varphi_{ij}(t_0) = \delta_{ij}$, $1 \leq i, j \leq n$, so heißt $X_{t_0} := X$ *Hauptfundamentalmatrix* (auch *Übergangsmatrix*) zum Zeitpunkt t_0 für (2.23). Ist X eine beliebige Fundamentalmatrix, so wird durch $X_{t_0}(t) := X(t)X(t_0)^{-1}$, $t \in I$, stets eine Übergangsmatrix definiert. Die Hauptfundamentalmatrix X_{t_0} ist die eindeutige globale Lösung des linearen homogenen AWP (auch Matrixdifferenzialgleichung genannt)

$$\dot{X} = A(t)X, \quad X(t_0) = I.$$

Ist X_{t_0} die Hauptfundamentalmatrix zum Zeitpunkt t_0 für (2.23), so wird für jedes $x_0 \in \mathbb{R}^n$ die eindeutige globale Lösung $\varphi(\cdot; t_0, x_0) \in \mathcal{C}^1(I, \mathbb{R}^n)$ des AWP

$$\dot{x} = A(t)x, \quad x(t_0) = x_0 \quad \text{durch} \quad \varphi(t; t_0, x_0) = X_{t_0}(t)x_0$$

gegeben. Insbesondere ist

$$\mathcal{L}_0 = \{X_{t_0}(\cdot)x_0 \mid x_0 \in \mathbb{R}^n\} \subset \mathcal{C}^1(I, \mathbb{R}^n)$$

der Lösungsraum von (2.23). Allgemein heißt jede Lösung von $\dot{X} = A(t)X$ *Lösungsmatrix* der homogenen DGL (2.23).

(5) Wronskideterminante

Ist X eine Lösungsmatrix von (2.23), d. h. jeder Spaltenvektor von X ist Lösung von (2.23), so heißt die Funktion

$$W : I \subset \mathbb{R} \rightarrow \mathbb{R} \quad \text{mit} \quad W(t) := \det X(t) \quad (2.26)$$

die *Wronskideterminante* der Lösungsmatrix $X = [\varphi_1, \dots, \varphi_n]$ oder des Lösungssystems $\{\varphi_1 \dots \varphi_n\}$ von (2.23). Man kann beweisen, dass die Wronskideterminante W eine Lösung der linearen homogenen Differenzialgleichung

$$\dot{y} = \sum_{i=1}^n a_{ii}y, \quad y \in \mathbb{R}.$$

ist. Hieraus folgt sofort die *Formel von Liouville*:

$$W(t) = W(t_0) e^{\int_{t_0}^t \sum_{i=1}^n a_{ii}(s) ds}, \quad \forall t, t_0 \in I. \quad (2.27)$$

Aus der Exponentialdarstellung der Wronskideterminante folgt:

1. Die Wronskideterminante einer Lösungsmatrix von $\dot{x} = A(t)x$ verschwindet entweder identisch oder nirgends.
2. Ein Lösungssystem $\{\varphi_1, \dots, \varphi_n\}$ ist genau dann ein Fundamentalsystem, wenn die zugehörige Wronskideterminante von Null verschieden ist.

(6) Die inhomogene DGL

Wir wenden uns nun dem inhomogenen DGL-System

$$\dot{x} = A(t)x + b(t), \quad t \in I \quad (2.28)$$

zu und erhalten aufgrund der Überlegungen in (2) sofort den Sachverhalt: Die Gesamtheit der Lösungen von (2.28) bildet den affinen Unterraum $\varphi_p + \mathcal{L}_0$ von $\mathcal{C}^1(I, \mathbb{R}^n)$, wobei $\varphi_p \in \mathcal{C}^1(I, \mathbb{R}^n)$ eine beliebige Lösung der inhomogenen Gleichung (2.28) und $\mathcal{L}_0 \subset \mathcal{C}^1(I, \mathbb{R}^n)$ der Lösungsraum der zugehörigen homogenen Gleichung $\dot{x} = A(t)x$ sind.

Es muss also (2.28) „irgendwie“ integriert werden. Üblich ist in der Literatur die Methode der *Variation der Konstanten*. Sie besteht darin, mit dem Ansatz $\varphi_p(t) = X(t)u(t)$ mit unbekanntem $u \in \mathcal{C}^1(I, \mathbb{R}^n)$ und (bekannter) Fundamentalmatrix $X(\cdot)$ ein geeignetes $u(\cdot)$ derart auszurechnen, dass φ_p eine spezielle Lösung von (2.28) darstellt. Wir berechnen u durch Einsetzen von $\varphi_p(t) = X(t)u(t)$ in die DGL (2.28):

$$\begin{aligned} \dot{X}(t)u(t) + X(t)\dot{u}(t) &= A(t)X(t)u(t) + b(t), \\ \underbrace{[\dot{X}(t) - A(t)X(t)]}_{=0}u(t) + X(t)\dot{u}(t) &= b(t). \end{aligned}$$

Es ergibt sich eine DGL für u :

$$\dot{u}(t) = [X(t)]^{-1}b(t), \quad t \in I.$$

Die Lösung dieser Gleichung lautet

$$\varphi_p(t; t_0) = X(t) \int_{t_0}^t [X(s)]^{-1}b(s) ds, \quad t \in I.$$

Wählt man für $X(t)$ die Hauptfundamentalmatrix $X_{t_0}(t)$, dann lautet die Lösung des AWP (mit $x(t_0) = x_0$):

$$\varphi(t; t_0, x_0) = X_{t_0}(t)x_0 + \int_{t_0}^t X_{t_0}(t)X_{t_0}(s)^{-1}b(s) ds, \quad t \in I. \quad (2.29)$$

Für das Auffinden einer partikulären Lösung φ_p muss man im Wesentlichen nur eine Fundamentalmatrix invertieren. Der Rest ist „Rechenarbeit“.

(7) Lösung der adjungierten Gleichung

Wir haben gelernt, dass eine Fundamentalmatrix X durch

$$\dot{X}(t) = A(t)X(t), \quad \det X(t) = W(t) \neq 0, \quad t \in I \quad (2.30)$$

charakterisiert ist. Umgekehrt folgt aus (2.30) auch

$$\dot{X}(t)X^{-1}(t) = A(t), \quad t \in I. \quad (2.31)$$

Ist also eine reguläre differenzierbare $n \times n$ -Matrix $X(\cdot)$ gegeben, dann sind ihre Spalten auf I linear unabhängige Lösungen der DGL $\dot{x} = A(t)x$ mit der durch (2.31) gegebenen Matrix $A(\cdot)$. Die Inverse X^{-1} einer Fundamentalmatrix X genügt ebenfalls einem DGL-System. Aus der Beziehung $X^{-1}X = I$ folgt zunächst durch Differenziation nach t und Beachtung von $\dot{X} = AX$

$$0 = (X^{-1})\dot{X} + X^{-1}\dot{X} = (X^{-1})\dot{X} + X^{-1}AX$$

und daraus die Identität

$$(X^{-1})' = -X^{-1}A(t). \quad (2.32)$$

Geht man in (2.32) zur adjungierten (transponierten) Gleichung über, d. h. zu

$$[(X^{-1})']^T = [(X^{-1})^T]' = -A^T(t)(X^{-1})^T,$$

dann sieht man sofort, dass $(X^{-1})^T$ eine Fundamentalmatrix der adjungierten DGL

$$\dot{y} = -A^T(t)y. \quad (2.33)$$

ist.

2.5 Lineare DGL-Systeme 1. Ordnung mit konstanten Koeffizienten

Für die in der Praxis häufig auftretenden linearen DGL-Systeme

$$\dot{x} = Ax + b(t) \quad (2.34)$$

mit konstanter Koeffizientenmatrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ und $b(\cdot) \in \mathcal{C}^0(I, \mathbb{R}^n)$ lässt sich – zumindest theoretisch – leicht ein Fundamentalsystem des zugehörigen homogenen DGL-Systems $\dot{x} = Ax$ angeben. Mit der allgemeinen Lösung des linearen homogenen Systems kann dann nach der Methode der Variation der Konstanten eine partikuläre Lösung des Systems (2.34) gemäß Abschnitt 2.4, (vi) bestimmt werden. Somit erfolgt die Konzentration unserer Überlegungen auf das Lösen des linearen homogenen DGL-Systems

$$\dot{x} = Ax, \quad x \in \mathbb{R}^n, \quad A \in \mathbb{R}^{n \times n}. \quad (2.35)$$

Lässt man sich von der einfachsten Form eines solchen DGL-Systems, also

$$\dot{x} = ax, \quad x \in \mathbb{R}, \quad a \in \mathbb{R} \quad (2.36)$$

leiten, dann besitzt diese Gleichung die allgemeine Lösung $\varphi(t) = ce^{at}$. Die Idee, um (2.35) zu lösen, ist nun die folgende: Man führt eine *Matrix-Exponentialfunktion* ein, indem man in der Partialsumme $\sum_{k=0}^r a_k x^k$ einer Potenzreihe $\sum_{k=0}^{\infty} a_k x^k$ die Potenzen x^k durch die Potenzen $A^k = A \cdot A \cdots A$ (k -mal, $A^0 = I$) einer Matrix $A \in \mathbb{R}^{n \times n}$ ersetzt. Man erhält auf diese Weise für jedes $r \in \mathbb{N}_0$ eine $n \times n$ -Matrix

$$\sum_{k=0}^r a_k A^k = a_0 I + a_1 A + \dots + a_r A^r. \quad (2.37)$$

Die Matrixreihe $\sum_{k=0}^{\infty} a_k A^k$ heißt konvergent mit Grenzwert $S = (s_{ij}) \in \mathbb{R}^{n \times n}$, wenn jede der (i, j) -Komponenten von (2.37) (das sind jeweils Partialsummen von Zahlenreihen) für $r \rightarrow \infty$ gegen s_{ij} ($1 \leq i, j \leq n$) konvergiert.

Satz 2.14 (Matrix-Exponentialfunktion)

Gegeben seien $A \in \mathbb{R}^{n \times n}$ und $r > 0$. Dann ist die Matrix-Exponentialreihe

$$e^{tA} := \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k = I + tA + \frac{1}{2!} t^2 A^2 + \dots \quad (2.38)$$

absolut und gleichmäßig konvergent für jedes $t \leq r$.

Beweis: Koordinatenweise ist jede (Zahlen-)Potenzreihe eine absolut konvergente Reihe mit dem Konvergenzradius ∞ und in jedem kompakten Teilintervall auch gleichmäßig konvergent. \square

Beispiel 2.15

1. Mit $A := \text{diag}(\lambda_1, \dots, \lambda_n)$ folgt $A^k = \text{diag}(\lambda_1^k, \dots, \lambda_n^k)$ und daraus $e^A = \text{diag}(e^{\lambda_1}, \dots, e^{\lambda_n})$.

2. Ist A eine nilpotente Matrix, d. h. $A^{k+1} = 0$ für ein $k \in \mathbb{N}$, dann bricht die Exponentialreihe für e^A nach endlich vielen Gliedern ab:

$$e^A = I + \frac{1}{1!} A + \dots + \frac{1}{k!} A^k.$$

Speziell: $A := \begin{pmatrix} 0 & 2 & 4 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{pmatrix} \implies e^A = I + \frac{1}{1!} A + \frac{1}{2!} A^2 = \begin{pmatrix} 1 & 2 & 7 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}.$

3. $A := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, $A^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, $A^3 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, $A^4 = I, \dots$

Daraus folgt:

$$e^{tA} = \begin{pmatrix} 1 - \frac{t^2}{2!} + \frac{t^4}{4!} \mp \dots & 0 - t + \frac{t^3}{3!} \mp \dots \\ 0 + t - \frac{t^3}{3!} \pm \dots & 1 - \frac{t^2}{2!} + \frac{t^4}{4!} \mp \dots \end{pmatrix} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}. \quad \blacksquare$$

Aus der Definition (2.38) folgen **Rechenregeln:**

- i. $e^0 = I$ (0 ist die Nullmatrix, I ist die Einheitsmatrix),
- ii. $AB = BA$ (A und B kommutieren) $\implies e^A e^B = e^B e^A = e^{A+B}$,
- iii. **Inverse:** $(e^A)^{-1} = e^{-A}$,
- iv. **Ableitung:** $\frac{d}{dt} e^{tA} = A e^{tA} = e^{tA} A$, $t \in \mathbb{R}$.

Man möge sich an einem einfachen Beispiel (2×2 -Matrix) klarmachen, dass bei der Rechenregel ii. aus $AB \neq BA$ im Allgemeinen auch $e^A e^B \neq e^B e^A \neq e^{A+B}$ folgt. Die Regel iv. folgt aus Satz 2.14, da die Reihe (2.38) gliedweise differenziert werden darf. Mit der Matrix-Exponentialfunktion kann man nun die allgemeine Lösung des linearen homogenen DGL-Systems (2.35) darstellen.

Satz 2.16 (Matrixexponentialdarstellung der Lösung)

Das lineare homogene DGL-System (2.35) mit konstanten Koeffizienten besitzt die allgemeine Lösung

$$\varphi(t; c) = e^{tA} c, \quad c \in \mathbb{R}^n. \quad (2.39)$$

Sie ist für alle $t \in \mathbb{R}$ definiert. Ist $C \in \mathbb{R}^{n \times n}$ eine invertierbare Matrix, dann ist

$$X(t) = e^{tA} C$$

eine Fundamentalmatrix des DGL-Systems (2.35).

Die Lösung des AWP $\dot{x} = Ax$, $x(t_0) = x_0 \in \mathbb{R}^n$ lautet:

$$\varphi(t; t_0, x_0) = e^{(t-t_0)A} x_0. \quad (2.40)$$

Beweis: Der Nachweis für die Formeln (2.39) und (2.40) erfolgt durch Nachrechnen (unter Zuhilfenahme der Rechenregel iv.). Die Unität der Lösung (2.40) des AWP ist durch den Existenz- und Eindeutigkeitsatz von Picard und Lindelöf gesichert. \square

Bemerkung 2.17

Die Abbildung $e^{tA} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ (mit $t_0 = 0$) kann als die Beschreibung für die Bewegung eines Punktes $x_0 \in \mathbb{R}^n$ im Phasenraum \mathbb{R}^n aufgefasst werden. Deshalb spricht man bei dieser Abbildung $\varphi(t; t_0 = 0, \cdot) = e^{tA}(\cdot)$ auch vom *Fluss* des linearen Systems. Eigenschaften des Flusses werden in Satz 2.24 dargelegt. \blacksquare

Um zu einer Lösung von (2.35) zu kommen, müsste man jedes Mal eine unendliche Reihe aufsummieren – eine im Allgemeinen sehr schwierige Aufgabe. Deshalb zeigen wir andere Möglichkeiten, wie man zu einem Fundamentalsystem von (2.35) gelangen kann.

Warum das Bestimmen der Lösungen von (2.35) auf ein Problem der linearen Algebra hinausläuft, soll die folgende Überlegung zeigen. Wir machen den Ansatz

$$x(t) = ce^{\lambda t} \quad \text{mit} \quad c \in \mathbb{C}^n, \lambda \in \mathbb{C} \quad (2.41)$$

und setzen A vorübergehend als konstante komplexe Matrix voraus. Diesen Ansatz in (2.35) eingesetzt, liefert die Gleichung

$$\dot{x} = \lambda c e^{\lambda t} = A c e^{\lambda t}.$$

Somit ist $x(t)$ genau dann eine Lösung von (2.35), wenn

$$A c = \lambda c \tag{2.42}$$

gilt. Ein Vektor $c \neq 0$, der der Gleichung (2.42) genügt, wird ein *Eigenvektor*, die Zahl λ , der zu c gehörige *Eigenwert* der Matrix A genannt (siehe Abschnitt 1.2.2). Somit sind alle Eigenpaare (λ, c) der Gleichung (2.42) oder, was dasselbe bedeutet, alle Lösungen von

$$(A - \lambda I)c = 0 \tag{2.43}$$

zu bestimmen. Bei (2.43) handelt es sich um ein lineares homogenes Gleichungssystem für c . Es hat genau dann eine nichttriviale Lösung, wenn

$$P_n(\lambda) := \det(A - \lambda I) = 0 \tag{2.44}$$

ist. Die Eigenwerte von A sind also die Nullstellen des sogenannten charakteristischen Polynoms $P_n(\lambda)$. Es besitzt n (reelle oder komplexe) Nullstellen, wobei jede Nullstelle gemäß ihrer Vielfachheit gezählt wird. Die Vielfachheit $\chi(\lambda)$ der Nullstelle λ des charakteristischen Polynoms $P_n(\cdot)$ heißt *algebraische Vielfachheit* des Eigenwertes. Bezeichnet man mit $\sigma(A)$ die Menge aller Eigenwerte von A , dann sagt der Fundamentalsatz der Algebra, dass $\sum_{\lambda \in \sigma(A)} \chi(\lambda) = n$ gilt. Die Dimension des Eigenraumes, $\dim \ker(A - \lambda I) =: m(\lambda)$, heißt *geometrische Vielfachheit* des Eigenwertes. Im Fall $\chi(\lambda) = 1$ heißt λ ein *einfacher Eigenwert*. Es gilt stets $1 \leq m(\lambda) \leq \chi(\lambda)$. Ist $m(\lambda) = \chi(\lambda)$ so heißt λ *halbeinfacher Eigenwert*. Für jedes $\lambda \in \sigma(A)$ ist $\dim \ker(A - \lambda I)^{\chi(\lambda)} = \chi(\lambda)$. Die von Null verschiedenen Elemente von $\ker(A - \lambda I)^{\chi(\lambda)}$ heißen verallgemeinerte Eigenvektoren von A und $\ker(A - \lambda I)^{\chi(\lambda)}$ heißt verallgemeinerter Eigenraum zu λ .

Satz 2.18 (Komplexe Lösungen)

Es seien λ, c, A komplex, $c \neq 0$. Die Funktion $\varphi(t) = c e^{\lambda t}$ ist genau dann eine Lösung von (2.35), wenn λ, c ein Eigenpaar von A ist. Die Lösungen $\varphi_i(t) = c_i e^{\lambda_i t}$ ($i = 1, \dots, r$) sind genau dann linear unabhängig, wenn die Vektoren c_i linear unabhängig sind. Dies ist gewiss dann der Fall, wenn alle λ_i , $i = 1, \dots, r$ verschieden sind.

Der **Beweis** des Satzes ist aufgrund der voranstehenden Überlegungen klar. \square

Man erhält z. B. genau dann ein Fundamentalsystem von Lösungen, wenn die Matrix A n verschiedene Eigenwerte besitzt. Der Satz (2.18) gilt natürlich auch für reelle Systeme. Man ist jedoch dann auch an reellen Lösungen interessiert. Die Schwierigkeit ergibt sich dadurch, dass eine reelle Matrix komplexe Eigenwerte besitzen kann, die dann auf komplexe Lösungen gemäß Ansatz (2.41) führen. Man rechnet leicht nach, dass bei reeller Matrix A der Realteil und auch der Imaginärteil einer komplexen Lösung zwei linear unabhängige Lösungen für $\dot{x} = Ax$ darstellen. Mit λ und c sind auch die konjugiert komplexen Größen $\bar{\lambda}$ und \bar{c} wieder Lösungen der Gleichung (2.42). Sie führen auf die zu $\varphi(t) = ce^{\lambda t}$ konjugiert komplexe Lösung $\overline{\varphi(t)} = \bar{c}e^{\bar{\lambda}t}$, bei der die Aufspaltung in Real- und Imaginärteil auf dieselben Lösungen führt. Somit hat man

Satz 2.19 (Reelle Lösungen)

Ist $\lambda := \mu + i\nu$ ($\nu \neq 0$) ein komplexer Eigenwert und $c := a + ib$ ein komplexer Eigenvektor der reellen Matrix A , so ergeben sich aus $\varphi(t) = ce^{\lambda t}$ zwei linear unabhängige reelle Lösungen

$$\begin{aligned} y(t) &= \operatorname{Re} \varphi(t) = e^{\mu t} (a \cos \nu t - b \sin \nu t), \\ y^*(t) &= \operatorname{Im} \varphi(t) = e^{\mu t} (a \sin \nu t + b \cos \nu t). \end{aligned}$$

Beweis: Die lineare Unabhängigkeit der Lösungen rechnet man einfach nach. □

Wir machen jetzt eine Koordinatentransformation. Durch $x = Cy$ mit $\det C \neq 0$ geht $\dot{x} = Ax$ über in

$$\dot{y} = C^{-1}ACy. \tag{2.45}$$

Besitzt nun die Matrix A die n linear unabhängige Eigenvektoren c_1, \dots, c_n und setzt man $C := (c_1, \dots, c_n)$ so ist $AC = (Ac_1, \dots, Ac_n) = (\lambda_1 c_1, \dots, \lambda_n c_n) = CD$, wobei

$$D = \operatorname{diag}(\lambda_1, \dots, \lambda_n) \quad (\text{d. h. } D = (d_{ij}) \text{ mit } d_{ii} = \lambda_i, d_{ij} = 0, i \neq j)$$

eine Diagonalmatrix ist, also $C^{-1}AC = D$ gilt. Das System (2.45) lautet dann

$$\dot{y} = Dy \quad \text{bzw.} \quad \dot{y}_i = \lambda_i y_i \quad (i = 1, \dots, n). \tag{2.46}$$

Ein Fundamentalsystem von Lösungen lautet dazu:

$$Y(t) = \text{diag}(y_1(t) \dots y_n(t)) = \begin{pmatrix} e^{\lambda_1 t} & 0 & \dots & 0 \\ 0 & e^{\lambda_2 t} & \dots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & e^{\lambda_n t} \end{pmatrix},$$

woraus sich aus $CY(t) = X(t)$ das Fundamentalsystem zu $\dot{x} = Ax$ ergibt.

Immer dann, wenn jeder Eigenwert $\lambda \in \sigma(A)$ halbeinfach ist, gibt es n linear unabhängige Eigenvektoren, so dass man damit eine Fundamentalmatrix aufstellen kann. Die Matrix A kann dann in Diagonalgestalt überführt werden (siehe (2.46)). Das angegebene Verfahren versagt, wenn die geometrische Vielfachheit $m(\lambda)$ eines Eigenwertes kleiner als seine algebraische Vielfachheit $\chi(\lambda)$ ist. Da die geometrische Vielfachheit die Maximalzahl linear unabhängiger Eigenvektoren zum Eigenwert λ angibt, ist diese also jetzt kleiner als die algebraische Vielfachheit $\chi(\lambda)$. Somit kann nach der bisherigen Vorgehensweise kein vollständiges Fundamentalsystem von Lösungen gefunden werden. Deshalb geben wir jetzt ein praktikables Verfahren an, wie man trotzdem zu einem Fundamentalsystem gelangt. Der Trick besteht darin, den \mathbb{R}^n mit einer Basis aus Eigenvektoren und Hauptvektoren zu versehen, die die Matrix-Exponentialreihe nach endlich vielen Summanden abbrechen zu lassen.

Praktische Berechnung eines Fundamentalsystems von Lösungen

Zur Bestimmung eines Fundamentalsystems benötigen wir den Begriff des *verallgemeinerten Eigenvektors* bzw. *Hauptvektors*.

Definition 2.20 (Hauptvektor)

Ein Vektor $v \in \mathbb{C}^n$ ($v \neq 0$) heißt *Hauptvektor* der Stufe $l \in \mathbb{N}$ zum Eigenwert λ der Matrix $A \in \mathbb{C}^{n \times n}$, wenn $(A - \lambda I)^l v = 0$ und $(A - \lambda I)^r v \neq 0$ für $r = 1, \dots, l - 1$ gilt.

Hieraus folgt: Hauptvektoren 1. Stufe sind Eigenvektoren. Für Hauptvektoren kann man folgendes zeigen:

- i. Zu jedem k -fachen Eigenwert λ der Matrix A gibt es k linear unabhängige Hauptvektoren, d. h.

$$\dim\{x \in \mathbb{C}^n \mid (A - \lambda I)^k x = 0\} = k.$$

- ii. Hauptvektoren zu verschiedenen Eigenwerten der Matrix A sind linear unabhängig.
- iii. Zu jeder Matrix $A \in \mathbb{C}^{n \times n}$ gibt es in \mathbb{C}^n eine Basis aus Hauptvektoren.
- iv. Ist $A \in \mathbb{R}^{n \times n}$ und v ein Hauptvektor von A zum nichtreellen Eigenwert λ , so sind $\text{Re } v$ und $\text{Im } v$ linear unabhängig.

Wir konstruieren nun eine Fundamentalmatrix. Dabei ist nicht unbedingt e^{tA} zu berechnen, sondern man hat mit $e^{tA}C$ ebenfalls eine Fundamentalmatrix, wenn nur die Matrix C invertierbar ist. Wählt man also

$$C = (v_1, v_2, \dots, v_n)$$

mit einer Basis aus Hauptvektoren von A , so ergeben sich in den Spalten der Fundamentalmatrix Lösungen $\varphi_i(t) = e^{tA}v_i$ in Form endlicher Summen:

- i. Für jeden Eigenvektor u zum Eigenwert λ die Lösung $e^{\lambda t}u$.
- ii. Für jeden Hauptvektor v der Stufe l zum Eigenwert μ die Lösung

$$e^{\mu t} \left[v + \frac{t}{1!} (A - \mu I)v + \frac{t^2}{2!} (A - \mu I)^2 v + \dots + \frac{t^{l-1}}{(l-1)!} (A - \mu I)^{l-1} v \right].$$

Beweis: Zu i. klar.

Zu ii. **Trick:** Es gilt $e^{tA} = e^{\mu t I + t(A - \mu I)} = e^{\mu t} e^{t(A - \mu I)}$. Daraus folgt

$$e^{tA}v = e^{\mu t} \sum_{k=0}^{\infty} \frac{t^k}{k!} (A - \mu I)^k v.$$

Wegen $(A - \mu I)^k v = 0$ für $k \geq l$ bricht die Matrix-Exponentialreihe nach l Summanden ab, und man hat lediglich eine endliche Summe zu berechnen. \square

Beispiel 2.21

Wir betrachten das AWP

$$\dot{x} = Ax, \quad x(0) = x_0, \quad x \in \mathbb{R}^3, \quad A = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \quad (2.47)$$

Gesucht ist der Fluss von (2.47) (siehe Bemerkung (2.17)) und eine partikuläre Lösung zu

$$\dot{x} = Ax + (0, 0, t)^T. \quad (2.48)$$

Wie im vorigen Beispiel werden zunächst die Eigenwerte aus $\det(A - \lambda I) = -\lambda(1 + \lambda)^2 = 0$ ermittelt. Der Eigenwert $\lambda_1 = 0$ tritt einfach auf, während $\lambda_2 = -1$ die algebraische Vielfachheit 2 und die geometrische Vielfachheit 1 besitzt.

Im Fall des Eigenwertes $\lambda_1 = 0$ ergibt sich ein Eigenvektor zu $v_1 = (0, 0, 1)^T$. Für $\lambda_2 = -1$ ergibt sich ein Eigenvektor $v_2 = (0, -1, 1)^T$ und $v_3 = (-1, 0, 1)^T$ als Hauptvektor der Stufe 2 aus $(A + I)v_3 = v_2$. Die drei Basislösungen lauten

$$\begin{aligned}\varphi_1(t) &= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \varphi_2(t) = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} e^{-t}, \\ \varphi_3(t) &= e^{At} v_3 = e^{\lambda_3 t + (A - \lambda_3 I)t} v_3 = e^{-t} e^{(A+I)t} v_3 \\ &= e^{-t} \left[I + \frac{t}{1!} (A+I) \right] v_3 = e^{-t} \left[v_3 + \frac{t}{1!} v_2 \right] = \begin{pmatrix} -1 \\ -t \\ 1+t \end{pmatrix} e^{-t}.\end{aligned}$$

Daraus stellen wir die Lösung zusammen:

$$\begin{aligned}\varphi(t; c) &= c_1 \varphi_1(t) + c_2 \varphi_2(t) + c_3 \varphi_3(t) \\ &= \begin{pmatrix} 0 & 0 & -e^{-t} \\ 0 & -e^{-t} & -te^{-t} \\ 1 & e^{-t} & (1+t)e^{-t} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} =: X(t) c, \quad c \in \mathbb{R}^3.\end{aligned}$$

Die Lösung des AWP lautet

$$\varphi(t; 0, x_0) \equiv \varphi(t; x_0) = X(t) \cdot X^{-1}(0) x_0 = e^{At} x_0.$$

Wir berechnen $X^{-1}(0)$ und erhalten

$$\begin{aligned}e^{tA} &= X(t) \cdot X^{-1}(0) = \begin{pmatrix} 0 & 0 & -e^{-t} \\ 0 & -e^{-t} & -te^{-t} \\ 1 & e^{-t} & (1+t)e^{-t} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ -1 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} e^{-t} & 0 & 0 \\ te^{-t} & e^{-t} & 0 \\ 1 - (1+t)e^{-t} & 1 - e^{-t} & 1 \end{pmatrix}.\end{aligned}$$

Damit ist der Fluss $\varphi(t; \cdot) = e^{tA}(\cdot)$ (siehe Bemerkung 2.17) berechnet und wir kommen zur Bestimmung einer partikulären Lösung von (2.48). Eine partikuläre Lösung φ_p lautet:

$$\begin{aligned}\varphi_p(t) &= \int_0^t e^{A(t-s)} b(s) ds = \int_0^t \begin{pmatrix} e^{-(t-s)} & 0 & 0 \\ (t-s)e^{-(t-s)} & e^{-(t-s)} & 0 \\ 1 - (1+t-s)e^{-(t-s)} & 1 - e^{-(t-s)} & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ s \end{pmatrix} ds \\ &= \int_0^t \begin{pmatrix} 0 \\ 0 \\ s \end{pmatrix} ds = \begin{pmatrix} 0 \\ 0 \\ t^2/2 \end{pmatrix}.\end{aligned}$$

Die allgemeine Lösung ergibt sich nun aus $\varphi(t; 0, x_0) = \varphi_h(t) + \varphi_p(t) = e^{At}x_0 + \varphi_p(t)$. Steht an Stelle des Anfangswertes $x(0) = x_0$ der Anfangswert $x(t_0) = x_0$, dann lautet die Lösung $\varphi(t; t_0, x_0) = \varphi_h(t) + \varphi_p(t) = e^{(t-t_0)A}x_0 + \varphi_p(t)$. ■

Eine lineare DGL n -ter Ordnung hat die Form

$$Lu := u^{(n)} + a_{n-1}(t)u^{(n-1)} + \dots + a_1(t)\dot{u} + a_0(t)u = b(t). \quad (2.49)$$

Sie ist dem DGL-System

$$\dot{x} = \mathbf{A}(t)x + \mathbf{b}(t) \quad \text{mit} \quad (2.50)$$

$$\mathbf{A}(t) := \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \dots & -a_{n-2} & -a_{n-1} \end{pmatrix}, \quad \mathbf{b}(t) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ b(t) \end{pmatrix}$$

äquivalent. Man gewinnt (2.50) aus (2.49) durch die Substitution

$$u = x_1, \quad \dot{u} = x_2, \quad \ddot{u} = x_3, \quad \dots, \quad u^{(n-1)} = x_n.$$

Es gilt wieder der Existenz- und Eindeigkeitsatz: Sind die $a_i(\cdot)$ ($i = 0, \dots, n-1$) und $b(\cdot)$ auf einem Intervall I stetig und ist $t_0 \in I$, dann hat das AWP

$$Lu = b(t), \quad u^{(k)}(t_0) = u_{0k}, \quad k = 0, \dots, n-1 \quad (2.51)$$

genau eine Lösung. Sie existiert auf ganz I und hängt in jedem kompakten Teilintervall von I stetig von den a_i und b ab.

2.6 Autonome Systeme

Die *autonomen DGL-Systeme* sind Sonderfälle der bisher behandelten Differenzialgleichungen. Sie sind dadurch gekennzeichnet, dass f ein zeitunabhängiges Vektorfeld ist. Wir legen ein System der Form

$$\dot{x} = f(x), \quad x \in \Omega \subset \mathbb{R}^n \quad (2.52)$$

zugrunde. Es soll unseren Standardvoraussetzungen genügen, d. h. $f : \Omega \rightarrow \mathbb{R}^n$ ist auf der offenen Menge Ω lokal Lipschitz-stetig. Somit behalten alle bisher erzielten Ergebnisse ihre Gültigkeit. Die rechte Seite von (2.52) ist also jetzt für alle Zeiten $t \in \mathbb{R}$ erklärt. Die Eigenschaften einer DGL (2.52) mit der speziellen rechten Seite $f(x) := Ax$, $A \in \mathbb{R}^{n \times n}$ haben wir bereits im Abschnitt 2.5 kennengelernt.

2.6.1 Allgemeine Aussagen

Ausgehend von einem DGL-System $\dot{x} = f(t, x)$ können wir dieses als mathematisches Modell für ein naturwissenschaftliches oder ökonomisches Problem ansehen. Dabei charakterisiert der Vektor $x = (x_1, \dots, x_n)$ einen Zustand des Systems. Mit der Zeit t beschreibt eine Lösung $t \mapsto x(t)$ des DGL-Systems einen Prozess, wobei $x(t)$ als die *Phase des Prozesses* zur Zeit t betrachtet werden kann. Deshalb nennen wir die x_i auch *Phasenkoordinaten* und das Bild von $t \mapsto x(t)$ wird *Phasenkurve* genannt. Der Vektor $\dot{x}(t)$ ist die *Phasengeschwindigkeit*. Dieser ist zu jedem Zeitpunkt t durch $\dot{x} = f(t, x)$ eindeutig definiert. Im Fall unseres autonomen DGL-Systems (2.52) legt der Zustand x – unabhängig von der Zeit – eindeutig die Phasengeschwindigkeit fest. Unter dieser Besonderheit wollen wir weitere Aussagen für unser DGL-System (2.52) treffen, da im Allgemeinen die t -Abhängigkeit von f eine Beeinflussung des Systems von außen bestimmt. Solche Prozesse verlaufen *heteronom*.

Wird ein physikalisches System durch eine autonome DGL modelliert und der Prozess zu unterschiedlichen Anfangszeiten t_0 und t_1 ($t_0 \neq t_1$) jedoch bei gleicher Anfangsbedingung gestartet, so können wir vermuten und letztlich auch beweisen, dass beide Prozesse identisch (bis auf eine Zeitverschiebung) verlaufen. Wir nennen diese Eigenschaft *Translationsinvarianz* der Lösung. Es wird auf die theoretische Ausführung dieser Behauptung verzichtet. Stattdessen geben wir eine weitere charakteristische Eigenschaft der Integralkurve in nachfolgender Bemerkung an.

Bemerkung 2.22

1. Es sei $\varphi(t; t_0, x_0)$ die Lösung des AWP $\dot{x} = f(x)$, $x(t_0) = x_0$. Dann ist

$$\left. \begin{aligned} \text{graph } \varphi(\cdot; t_0, x_0) &= \{(t, \varphi(t; t_0, x_0)) \mid t \in I_{\max}(t_0, x_0)\} \\ &= \{(t, \varphi(t - t_0; 0, x_0) \mid t - t_0 \in I_{\max}(0, x_0)\} \\ &= \{(s + t_0, \varphi(s; 0, x_0) \mid s \in I_{\max}(0, x_0)\}. \end{aligned} \right\} \quad (2.53)$$

Die letzte Menge ist die um t_0 parallel zur t -Achse verschobene Integralkurve zum Anfangswert $(0, x_0)$ (siehe Abbildung 2.10). Im Fall $n = 1$ besitzt eine autonome DGL keine periodische Lösung. Ausgenommen sind die Punkte mit $f(x) = 0$. Dort ist die existierende Lösung konstant und als solche periodisch mit jeder beliebigen Periode $T > 0$. In allen anderen Punkten ist die Lösung streng monoton.

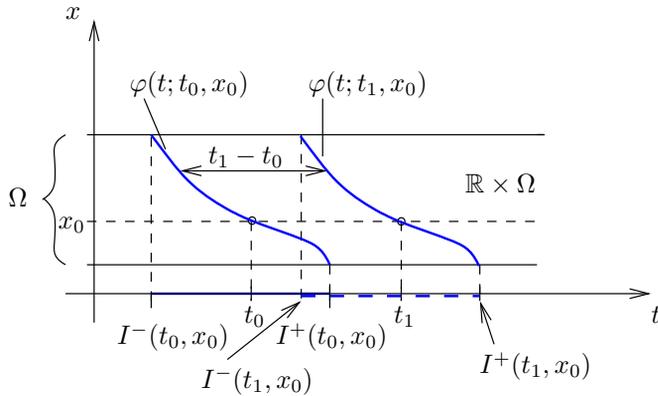


Abb. 2.10 Die Integralkurven zu Anfangswerten (t_0, x_0) und (t_1, x_1) bei festem x_0 sind in bezug auf t translationsinvariant

2. Die Translationseigenschaft ist auch für das Studium periodischer Lösungen und für die Theorie dynamischer Systeme wichtig. ■

Den voranstehenden Überlegungen entnimmt man, dass aufgrund der Unität der Lösung des AWP die Phasenkurven entweder identisch oder disjunkt sind. Die Lösungen $t \mapsto \varphi(t)$ und $t \mapsto \varphi(t - t_0)$, die man erhält, sind zwar im (t, x) -Raum verschieden, aber als Bild im x -Raum identisch. Der Raum, in dem das Verhalten der Variablen x_1, \dots, x_n , parametrisiert durch die Zeit t , beschrieben wird, heißt *Phasenraum*. Ein Punkt im Phasenraum mit den Koordinaten $x_1(t), \dots, x_n(t)$ für ein festes t heißt *Phasenpunkt*. Für wachsendes t wird der Phasenpunkt durch den Phasenraum transportiert. Phasenkurven können sich im Phasenraum nicht schneiden. In unserer Ausgangsgleichung (2.52) ist Ω der Phasenraum der DGL. Wir erläutern diese Begriffsbildungen noch einmal an einem einfachen Beispiel.

Beispiel 2.23

Die Gleichung für den harmonischen Oszillator ist gegeben durch

$$\ddot{x} + x = 0. \quad (2.54)$$

Sie ist autonom, das dazugehörige System lautet:

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = -x_1. \quad (2.55)$$

Die Lösungen der skalaren Gleichung (2.54) sind Linearkombinationen aus $\cos t$ und $\sin t$. Es ist leicht, diese periodischen Lösungen im Lösungsraum $G = \mathbb{R} \times \mathbb{R}^2$ zu zeichnen (siehe Abbildung 2.11 (links)). Die Lösungen können in den $\Omega := \mathbb{R}^2$ -Raum, d. h. in die (x, \dot{x}) -

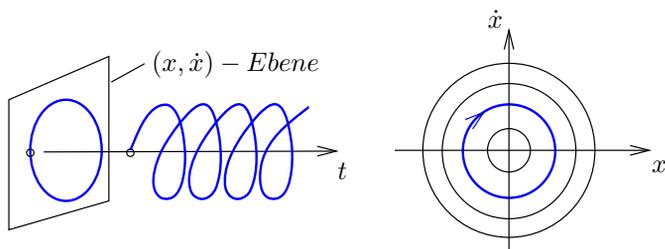


Abb. 2.11 Spiralförmige Lösungen von (2.55) im (x, \dot{x}, t) -Raum (links), Phasenkurven von (2.55) (rechts)

Ebene projiziert werden und erscheinen dort als geschlossene Phasenkurven bzw. Orbits (Abbildung 2.11 (rechts)). Man erkennt also im Phasenraum Ω an den geschlossenen Orbits eine periodisch ablaufende Bewegung. ■

Da die Zeit in der Gleichung (2.54) bzw. (2.55) nicht explizit auftritt, können diese Projektionen auch für Lösungen der allgemeinen Gleichung (2.52) durchgeführt werden. Um Verwechslungen zu vermeiden, werden die Begriffe Lösung, Lösungskurve und Phasenkurve (auch Orbit, Trajektorie oder Bahn) für die DGL (2.52) kurz erläutert. Eine *Lösung* ist eine Funktion $\varphi : I \rightarrow \mathbb{R}^n$ von einem Intervall I in den \mathbb{R}^n , eine *Lösungskurve* ist der Graph $\{(t, \varphi(t)) \in \mathbb{R}^{1+n} \mid t \in I\}$ einer solchen Lösung, also eine Teilmenge des \mathbb{R}^{1+n} . Zwischen beiden besteht kein großer Unterschied, da man Funktionen letztlich durch ihre Graphen definiert. Eine *Phasenkurve* (Orbit etc.) ist die Bildmenge $\{\varphi(t) \in \mathbb{R}^n \mid t \in I\}$ einer (maximalen) Lösung φ , also eine durch die Funktion φ parametrisierte Kurve im \mathbb{R}^n . Wie auch im nichtautonomen Fall nennen wir I_{\max} das maximale Lösungsintervall zum AWP

$$\dot{x} = f(x), \quad x(0) = x_0. \quad (2.56)$$

Wegen der Translationsinvarianz der Zeit t kann in (2.56) stets $t_0 = 0$ gesetzt werden. Mit $\varphi(t; x_0) := \varphi(t; 0, x_0)$ wird die allgemeine Lösung zu (2.56) mit maximalem Definitionsintervall $I_{\max}(x_0) := I_{\max}(0, x_0)$ bezeichnet.

Bisher galt unser Interesse dem Verlauf einzelner Lösungskurven der Abbildung $t \mapsto \varphi(t; x_0)$ des Systems (2.56). Nun beziehen wir einen anderen Standpunkt. Wir halten t fest und wollen wissen, was aus einer bestimmten Menge M von Startpunkten nach der Zeit t wird, d. h. wie sich die Menge $M_t := \{\varphi(t; x_0) \mid x_0 \in M\}$ im Laufe der Zeit verhält. Wir betrachten also die Schar von Abbildungen φ_t und nennen

$$\varphi_t : \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad x_0 \mapsto \varphi_t(x_0) := \varphi(t; x_0)$$

den Fluss zu der DGL (2.56) für jeden beliebigen Anfangswert $x_0 \in \Omega \subset \mathbb{R}^n$. Man spricht auch noch vom Fluss φ_t zum Vektorfeld f . Die Exponentialabbildung $\varphi_t := e^{At}$ erfüllt die folgenden grundlegenden Eigenschaften für alle $x \in \mathbb{R}^n$: $\varphi_0 = I$, $\varphi_s(\varphi_t(x)) = \varphi_{s+t}(x)$ für alle $s, t \in \mathbb{R}$ und $\varphi_{-t}(\varphi_t(x)) = \varphi_t(\varphi_{-t}(x))$ für alle $t \in \mathbb{R}$. Diese Eigenschaften lassen sich sofort aus den Rechengesetzen für die Matrix-Exponentialfunktion herleiten. Der folgende Satz zeigt, dass diese Eigenschaften sogar für den Fluss der DGL (2.56) erhalten bleiben.

Satz 2.24 (Eigenschaften von Flüssen)

Für den Fluss φ_t und das zugehörige maximale Existenzintervall $I_{\max}(x_0) = (I^-(x_0), I^+(x_0))$ des autonomen Systems (2.56) gelten für ein beliebiges aber fest gewähltes $x_0 \in \Omega$ folgende Aussagen:

- i. $-\infty \leq I^-(x_0) < 0 < I^+(x_0) \leq +\infty$,
- ii. $\varphi(0; x_0) = x_0$ (bzw. $\varphi_0 = I$),
- iii. $\varphi(t; \varphi(s; x_0)) = \varphi(t+s; x_0)$ (bzw. $\varphi_t \circ \varphi_s = \varphi_{t+s}$), $\forall s, t$ mit $s, t+s \in I_{\max}(x_0)$,
- iv. $\varphi(-t; \varphi(t; x_0)) = \varphi(0; x_0) = x_0$ (bzw. $\varphi_{-t} \circ \varphi_t = I$), $\forall t \in I_{\max}(x_0)$.

Sämtliche Behauptungen folgen aus entsprechenden Aussagen über die anfangszeitabhängige allgemeine Lösung $\varphi(t; t_0, x_0)$. Für eine detaillierte Beweisführung kann man Aulbach (1997) heranziehen. (Eine Verwechslungsgefahr zwischen I für die Identität einer Abbildung und I als Intervall besteht nicht!)

Die Abbildung 2.12 veranschaulicht das Resultat iii. des Satzes bezüglich eines Anfangspunkts x_0 . Man stelle sich dazu vor, dass nach Anhalten der Zeit zum Zeitpunkt τ die Uhr nicht von τ an los läuft, sondern man die Zeit auf Null zurücksetzt und erst dann die Uhr wieder startet. Dies wird bei der geometrischen Veranschaulichung in Abbildung 2.12 (links) dadurch sichtbar, dass der Kurvenbogen auf der Lösungskurve von $\varphi(t; x_0)$ für t zwischen τ und $\tau + \sigma$ identisch ist mit dem Kurvenbogen zwischen 0 und σ zu derjenigen Lösung, die zum Zeitpunkt 0 in $\varphi(\tau; x_0)$ startet.

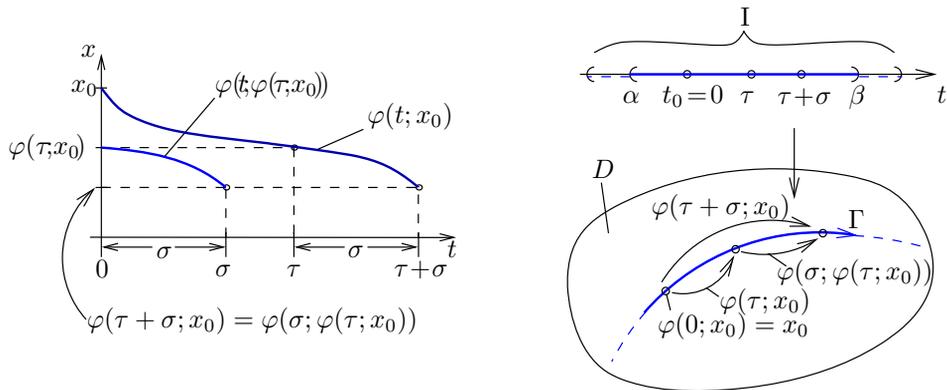


Abb. 2.12 Zur Eigenschaft iii. des Flusses

Bemerkung 2.25

1. Auch bei Flüssen kann man Aussagen über ihr Randverhalten machen: Liegt der Fluss $\varphi(t; x_0)$ der DGL $\dot{x} = f(x)$ für alle $t \in [0, I^+(x_0))$ in einer kompakten Menge K (siehe Abbildung 2.8), so gilt $I^+(x_0) = \infty$. Ist $I^+(x_0)$ endlich, so ist $\varphi(t; x_0)$ entweder auf dem Intervall $[0, I^+(x_0))$ unbeschränkt, oder der Orbit läuft zum Rand von Ω , d. h. es gilt

$$d(\varphi(t; x_0), \partial\Omega) \rightarrow 0 \text{ für } t \rightarrow I^+(x_0) - 0.$$

d bezeichnet den Abstand zwischen $\varphi(t, x_0)$ und dem Rand von Ω .

2. Betrachten wir den (linearen) Fluss $e^{At} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ des linearen Systems $\dot{x} = Ax$, dann erfüllt dieser Fluss – wie bereits erwähnt – alle Eigenschaften aus Satz 2.24.

3. Im allgemeinen Fall einer DGL (2.56) muss es nicht für das gesamte Gebiet Ω eine gleichmäßige Lebensspanne der Menge M geben (siehe nachfolgendes Beispiel). ■

Beispiel 2.26

Wir betrachten das AWP $\dot{x} = x(1 - x)$, $x(0) = x_0$, $x \in \mathbb{R}$ und wollen die Lösungen für alle Anfangswerte $x(0) = x_0$, $x_0 \in \mathbb{R}$ bestimmen. Die Nullstellen der rechten Seite der DGL liefern sofort die stationären Lösungen $\varphi_t(0) = \varphi_t(1) = 0$ für alle Zeiten $t \in \mathbb{R}$. Als nächstes bestimmen wir die Lösungen für alle Anfangswerte $x(0) = x_0$ mit $x_0 > 1$. Trennung der Variablen und Integration beider Seiten der DGL liefert die Lösung

$$\varphi_t(x_0) = \varphi(t; x_0) = \frac{x_0 e^t}{1 + x_0(e^t - 1)}, \quad I_{\max}(x_0) = (\ln(1 - 1/x_0), \infty), \quad x_0 > 1.$$

Analog bestimmt man für die Fälle $0 < x_0 < 1$ und $x_0 < 0$ die Lösungen φ_t . Sie lauten:

$$\varphi(t; x_0) = \begin{cases} \frac{x_0 e^t}{1 + x_0(e^t - 1)}, & I_{\max}(x_0) = \mathbb{R}, \quad 0 < x_0 < 1, \\ \frac{x_0 e^t}{1 + x_0(e^t - 1)}, & I_{\max}(x_0) = (-\infty, \ln(1 - 1/x_0)), \quad x_0 < 0. \end{cases}$$

Diese DGL besitzt keinen globalen Fluss φ_t , da $\bigcap_{x_0 \in \mathbb{R}} I_{\max}(x_0) = \{0\}$. ■

Um das Phasenportrait einer DGL zu entwerfen, benötigt man die Phasenkurven des Systems. Diese können wir mit einem einfachen Trick gewinnen. Wir schreiben (2.52) komponentenweise auf: $\dot{x}_i = f_i(x)$, $i = 1, \dots, n$. Dann benutzen wir eine der Koordinaten von x , z. B. x_1 als neue Koordinate. Dies erfordert die Voraussetzung $f_1(x) \neq 0$. Mittels Kettenregel ergeben sich dann $n - 1$ Gleichungen der Form

$$\frac{dx_2}{dx_1} = \frac{f_2(x)}{f_1(x)}, \quad \dots, \quad \frac{dx_n}{dx_1} = \frac{f_n(x)}{f_1(x)}. \quad (2.57)$$

Die Lösungen dieses Systems im Phasenraum werden Orbits genannt. Die Existenz und Eindeutigkeit der Lösung ist nach Satz 2.8 gesichert, falls die rechte Seite von (2.57) Lipschitz-stetig ist. Die Lösungen von (2.57) beschreiben das Verhalten der Orbits im Phasenraum. Die Orbits im Phasenraum schneiden sich nicht. Die Menge aller Orbits stellen das *Phasenportrait* dar.

Falls $f_1(x)$ Nullstellen besitzt, dafür aber $f_2(x)$ keine Nullstellen hat, dann können wir x_2 als unabhängige Variable verwenden und in (2.57) die Rolle von $f_1(x)$ und $f_2(x)$ vertauschen usw. Ist $f_1(x_0) = f_2(x_0) = \dots = f_n(x_0) = 0$ in irgend welchen Punkten $x_0 \in \Omega$, dann nennt man x_0 *Fixpunkt* oder *Gleichgewichtspunkt* (siehe Definition 2.28). In der Umgebung solcher Punkte ist eine gesonderte Untersuchung erforderlich.

Beispiel 2.27

1. Wir betrachten noch einmal den harmonischen Oszillator $\ddot{x} + x = 0$ (siehe Beispiel 2.23). Der Phasenraum ist zweidimensional (siehe System (2.55)) und die Orbits werden durch eine der Gleichungen

$$\frac{dx_2}{dx_1} = -\frac{x_1}{x_2}, \quad x_2 \neq 0 \quad \text{bzw.} \quad \frac{dx_1}{dx_2} = -\frac{x_2}{x_1}, \quad x_1 \neq 0$$

beschrieben. Die Integration der ersten Gleichung liefert $x_1^2 + x_2^2 = c$ ($c = \text{const.}$). Es handelt sich hierbei um eine Familie von Kreisen in der Phasenebene; $(0, 0)$ ist ein Gleichgewichtspunkt, der *Zentrum* genannt wird (siehe Abbildung 2.13 (links)).

2. Die Gleichung $\ddot{x} - x = 0$ hat ebenfalls einen 2-dimensionalen Phasenraum und die Orbits werden durch die Gleichung

$$\frac{dx_2}{dx_1} = \frac{x_1}{x_2}, \quad x_2 \neq 0$$

beschrieben. Die Integration liefert

$$x_1^2 - x_2^2 = c \quad (c = \text{const.}; \text{ Familie von Hyperbeln}).$$

Der kritische Punkt $(0, 0)$ wird *Sattel* genannt (siehe Abbildung 2.13 (rechts)). In Definition 2.44 gehen wir auf diese Begriffsbildung näher ein. Die Pfeile zeigen die Richtung der Bewegung der Phasenpunkte mit wachsender Zeit an. Die Bewegung einer Menge von Phasenpunkten entlang der dazugehörigen Orbits wird der *Phasenfluss* genannt. ■

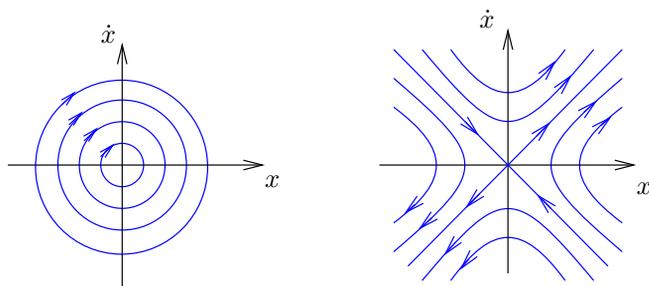


Abb. 2.13 Der Fixpunkt $(0, 0)$ ist ein Zentrum (links), der Fixpunkt $(0, 0)$ ist ein Sattel (rechts)

2.6.2 Kritische Punkte

Im Phasenraum betrachtet ist ein *kritischer Punkt* x_0 ein Punkt, in dem keine Bewegung stattfindet. Dieser Punkt stimmt mit der Gleichgewichtslösung $\varphi(t) = x_0$, $(t \in \mathbb{R})$ der DGL $\dot{x} = f(x)$ überein. Man hat für ihn auch noch andere Begriffe geprägt.

Definition 2.28 (Fixpunkt, Gleichgewichtspunkt)

Der Punkt $x = x_0 \in \Omega$ mit $f(x_0) = 0$ wird *Fixpunkt* oder auch *Gleichgewichtspunkt* der DGL $\dot{x} = f(x)$ genannt.

Aus dem Existenz- und Eindeutigkeitsatz folgt weiter, dass ein Gleichgewichtspunkt niemals in endlicher Zeit erreicht werden kann. Wäre dies der Fall, dann würden sich zwei (verschiedene) Lösungen schneiden, was der Eindeutigkeit widerspricht.

Beispiel 2.29

Wir betrachten die DGL $\dot{x} = -x$, $t \geq 0$. Hier ist $x = 0$ ein Fixpunkt, $x(t) = 0$ für alle $t \geq 0$ ist Gleichgewichtslösung. Die Lösungen lauten $\varphi(t, x_0) = x_0 e^{-t}$, $x_0 \in \mathbb{R}$. Alle Lösungen, die in $x_0 \neq 0$ zur Zeit $t = 0$ starten, laufen in unendlicher Zeit in den Fixpunkt hinein: $\lim_{t \rightarrow \infty} x(t) = 0$. ■

Beispiel 2.30

Bei der DGL $\dot{x} = -x^2$, $t \geq 0$ ist $x = 0$ wieder ein Fixpunkt, $x(t) \equiv 0$ ist Gleichgewichtslösung. Die Lösungen, die in $x_0 \neq 0$ (für $t = 0$) starten, zeigen qualitativ verschiedenes Verhalten für $x_0 > 0$ bzw. $x_0 < 0$. Dies ist klar, wenn man die Lösung $\varphi(t; x_0) = \frac{1}{t+1/x_0}$ ($x_0 \neq 0$) betrachtet. Für $x_0 < 0$ werden die Lösungen in endlicher Zeit unbeschränkt. Für $x_0 > 0$ laufen die Lösungen zum Fixpunkt. Dieses Phänomen nennt man *Attraktion* (Anziehung). ■

Der Leser möge für beide Beispiele die Lösungskurven und den (eindimensionalen) Phasenraum skizzieren.

Definition 2.31 (Attraktor, Repeller)

Ein Fixpunkt $x = x_0$ der DGL $\dot{x} = f(x)$ heißt *Attraktor*, falls eine Umgebung $U(x_0) \subset \mathbb{R}^n$ von $x = x_0$ existiert, so dass aus $x(0) \in U(x_0)$ folgt: $\lim_{t \rightarrow \infty} x(t) = x_0$. Besitzt ein Fixpunkt x_0 diese Eigenschaft für $t \rightarrow -\infty$, dann ist $x = x_0$ ein *Repeller* (auch negativer Attraktor genannt).

Für das Auffinden eines Attraktors (Repellers) ist die grundlegende analytische Herangehensweise dabei die, dass wir stets mit der Linearisierung der DGL in einer Umgebung des kritischen Punktes beginnen. Wir untersuchen die Frage, inwieweit Eigenschaften des linearen Systems in einer noch näher zu erläuternden Art und Weise im nichtlinearen DGL-System zumindest qualitativ erhalten bleiben. Wir setzen voraus, dass f eine Taylor-Reihenentwicklung in der Nähe des kritischen Punktes x_0 – man beachte, dass dort $f(x_0) = 0$ ist – besitzt:

$$f(x) = Df(x_0)(x - x_0) + o(\|x - x_0\|). \quad (2.58)$$

Hier und im Weiteren bezeichnet $Df(x_0) = (\frac{\partial f_i(x_0)}{\partial x_j})$ ($i, j = 1 \dots n$) die Jacobi-Matrix⁸ der ersten partiellen Ableitungen der Funktion $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$. Unter der Linearisierung von $\dot{x} = f(x)$ verstehen wir dann das DGL-System mit konstanter Koeffizientenmatrix

$$\dot{y} = Df(x_0)(y - x_0). \quad (2.59)$$

Substituieren wir $\bar{y} := y - x_0$, dann erhalten wir das System $\dot{\bar{y}} = Df(x_0)\bar{y}$. Wir nennen es wieder

$$\dot{y} = Df(x_0)y \quad (2.60)$$

Mit $A := Df(x_0)$ handelt es sich um das in Abschnitt 2.5 ausführlich studierte System (2.35).

Beispiel 2.32

Die DGL des mathematischen Pendels lautet:

$$\ddot{x} + \sin x = 0, \quad -\pi < x < \pi \pmod{2\pi}$$

oder als DGL-System geschrieben

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = -\sin x_1, \quad (2.61)$$

⁸Wir weichen hier von der bisher verwendeten Schreibweise $f'(x_0)$ ab.

wobei x die Winkelauslenkung von der nach unten orientierten Vertikalen angibt. Die Fixpunkte lauten $(x_1, x_2) = (0, 0), (-\pi, 0), (\pi, 0)$, die Linearisierung am Fixpunkt $(0, 0)$ ist

$$\begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\cos x_1 & 0 \end{pmatrix} \Big|_{(0,0)} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} y_2 \\ -y_1 \end{pmatrix},$$

während an den Fixpunkten $(\pm\pi, 0)$

$$\begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \mp \pi \\ y_2 \end{pmatrix}$$

gilt. Bei der Pendelgleichung treten qualitative Unterschiede im Phasenportrait im Vergleich zur Linearisierung auf. Der Charakter der Fixpunkte bleibt erhalten: $(0, 0)$ ist ein *Zentrum*, $(\pm\pi, 0)$ ist jeweils ein *Sattel*. Das Phasenportrait von (2.61) lässt sich jedoch nicht aus seiner Linearisierung gewinnen (siehe Abbildung 2.14).

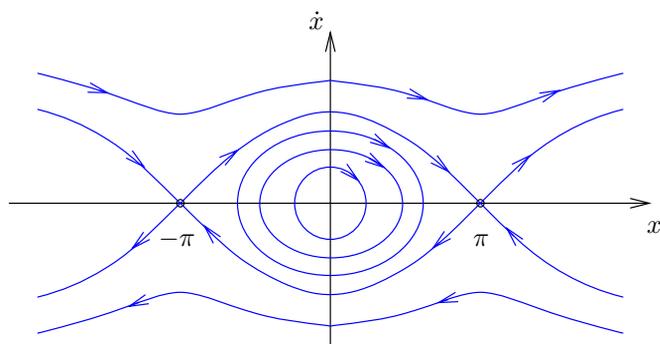


Abb. 2.14 Phasenportrait der Pendelgleichung

Zwei charakteristische Orbits, die die Fixpunkte $(\pm\pi, 0)$ miteinander verbinden, fallen auf. Solche Orbits heißen *heterokline Orbits*. In diesem Zusammenhang kann man auch von *Separatrizen* sprechen. Sie trennen geschlossene von nicht geschlossenen Orbits. ■

Wir wenden uns *periodischen Lösungen* zu (siehe Abbildung 2.14). Wir nehmen an, dass $\varphi(\cdot)$ eine Lösung der Gleichung $\dot{x} = f(x)$, $x \in \Omega \subset \mathbb{R}^n$ ist und weiter ein $T > 0$ existiert, so dass $\varphi(t+T) = \varphi(t)$ für alle $t \in \mathbb{R}$ gilt. Eine solche Lösung heißt periodisch mit Periode T . Falls φ die Periode T hat, dann hat φ auch die Perioden $2T, 3T$ usw. Falls T die kleinste Periode ist, nennen wir $\varphi(\cdot)$ T -periodisch. Eine T -periodische Lösung von $\dot{x} = f(x)$ nimmt also nach der Zeit T wieder den gleichen Wert im \mathbb{R}^n an. Solch eine periodische Lösung erzeugt einen geschlossenen Orbit im Phasenraum. Insgesamt gilt der folgende Zusammenhang:

Satz 2.33

Jede periodische Lösung der DGL $\dot{x} = f(x)$ stimmt mit einem geschlossenen Orbit (im \mathbb{R}^n) überein und umgekehrt ist jeder geschlossene Orbit eine periodische Lösungskurve (im \mathbb{R}^{n+1}).

Der Leser sollte beachten, dass es bei nichtautonomen Systemen auch geschlossene Lösungen in der x - y -Phasenebene gibt, die nicht periodisch sind. Dazu muss man als Beispiel nur das heteronome System

$$\dot{x} = 2ty, \quad \dot{y} = -2tx$$

betrachten, das Lösungen der Form $x(t) = -c_1 \cos(t^2) + c_2 \sin(t^2)$, $y(t) = c_1 \sin(t^2) + c_2 \cos(t^2)$ besitzt. Die dazugehörigen Phasenkurven sind geschlossen, jedoch nicht periodisch. Im autonomen Fall sind jedoch geschlossene Phasenkurven stets periodisch.

2.7 Hilfsmittel zur Konstruktion von Phasenportraits

In diesem Abschnitt wollen wir uns das Ziel stellen, unter Zuhilfenahme der linearen Theorie Aussagen für das nichtlineare DGL-System $\dot{x} = f(x)$ in einer Umgebung eines Fixpunktes zu gewinnen. Eine zentrale Rolle wird dabei der Satz 2.54 von *Hartman* und *Grobman* einnehmen, der eine Aussage darüber macht, dass Phasenportraits in der Umgebung eines hyperbolischen Fixpunktes homöomorph zur linearisierten Gleichung sind. Dieser Satz ermöglicht auch Stabilitätsaussagen am Fixpunkt.

2.7.1 Stabilität linearer Systeme

Zunächst werden die stabilen, instabilen und Zentrumsunterräume E^s , E^u und E^c eines linearen DGL-Systems

$$\dot{x} = Ax, \quad A \in \mathbb{R}^{n \times n} \tag{2.62}$$

erklärt. Es sei $w_j = u_j + iv_j$ ein verallgemeinerter Eigenvektor der reellen Matrix A mit zugehörigem Eigenwert $\lambda_j = a_j + ib_j$. Ist $b_j = 0$, dann ist auch $v_j = 0$. Weiter sei

$$B = \{u_1, \dots, u_k, u_{k+1}, v_{k+1}, \dots, u_m, v_m\}$$

eine Basis im \mathbb{R}^n mit $k + 2(m - k) = n$, also $n = 2m - k$.

Definition 2.34

Mit den Basiselementen u_k, v_k wird definiert:

$$E^s := \text{span} \{u_j, v_j \mid a_j < 0\}, \quad E^u := \text{span} \{u_j, v_j \mid a_j > 0\},$$

$$E^c := \text{span} \{u_j, v_j \mid a_j = 0\}.$$

Die Räume E^s, E^u und E^c sind Unterräume des \mathbb{R}^n , die durch die Real- und Imaginärteile der verallgemeinerten Eigenvektoren w_j in bezug auf die Realteile der Eigenwerte λ_j aufgespannt werden.

Beispiel 2.35

Die Matrix

$$A = \begin{pmatrix} -2 & -1 & 0 \\ 1 & -2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

hat die Eigenwerte $\lambda_{1,2} = -2 \pm i$ und $\lambda_3 = 3$. Der komplexe EW $\lambda_1 = -2 + i$ erzeugt den Eigenvektor $w_1 = u_1 + iv_1 = (0, 1, 0)^T + i(1, 0, 0)^T$. Wegen $a_1 = -2 < 0$ wird durch $(0, 1, 0)^T$ und $(1, 0, 0)^T$ der stabile Unterraum E^s aufgespannt. Analog ergibt sich $E^u = \text{span} \{(0, 0, 1)^T\}$ zum Eigenwert $\lambda_3 = 3$. Wir sehen, dass alle Lösungen in E^s sich

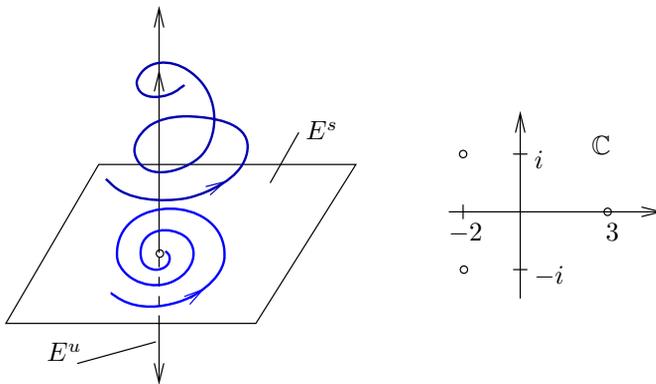


Abb. 2.15 Stabile und instabile Unterräume E^s und E^u (links), Eigenwertdiagramm (rechts)

dem Gleichgewichtspunkt $x = 0$ für $t \rightarrow \infty$ annähern, während sich alle Lösungen aus E^u dem Fixpunkt $x = 0$ für $t \rightarrow -\infty$ nähern (siehe Abbildung 2.15). ■

Beispiel 2.36

Wir betrachten das Eigenwertdiagramm in Abbildung 2.16 (rechts). Wir können es uns z. B. durch die Matrix

$$A = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

erzeugt denken. Zu $\lambda_1 = i$ gehört das Paar reeller Eigenvektoren $u_1 = (0, 1, 0)^T$, $v_1 = (1, 0, 0)^T$ und zu $\lambda_3 = 1$ gehört $u_3 = (0, 0, 1)^T$. Der Zentrumsunterraum E^c ist die (x_1, x_2) -Ebene und E^u ist die x_3 -Achse. Wegen $x_1^2 + x_2^2 = c^2$ liegen alle Phasenkurven auf dem Zylinder (siehe Abbildung 2.16). ■

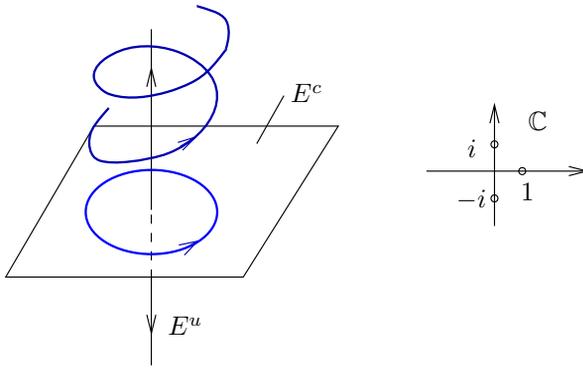


Abb. 2.16 Die Unterräume E^c und E^u (links), Eigenwertdiagramm (rechts)

Die Abbildung 2.16 suggeriert die Vorstellung, dass Lösungen mit einem Anfangspunkt in E^c beschränkt sind. Dass dies nicht so ist, zeigt das einfache

Beispiel 2.37

Gegeben sei das DGL-System

$$\dot{x} = Ax, \quad A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \text{bzw.} \quad \begin{aligned} \dot{x}_1 &= 0 \\ \dot{x}_2 &= x_1. \end{aligned}$$

Die gesamte x_2 -Achse ist Fixpunktgerade und Null ist doppelter Eigenwert. Hierzu gehören $u_1 = (0, 1)^T$ als Eigenvektor und $u_2 = (1, 0)^T$ als verallgemeinerter Eigenvektor. Folglich ist $E^c = \mathbb{R}^2$. Die Lösungen lauten

$$\begin{aligned} x_1(t) &= c_1 \\ x_2(t) &= c_1 t + c_2. \end{aligned}$$

Nur die Lösungen mit $c_1 = 0$ bleiben beschränkt, alle anderen (mit $c_1 \neq 0$) nicht. Der Leser möge sich das Phasenportrait klar machen. ■

Wir hatten das AWP

$$\dot{x} = Ax, \quad A \in \mathbb{R}^{n \times n}, \quad x(0) = x_0 \quad (2.63)$$

betrachtet (siehe Abschnitt 2.5). Wie bekannt, ist die Lösung des AWP (2.63) durch

$$\varphi(t; 0, x_0) = e^{At} x_0$$

gegeben. Die Abbildung $e^{At} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ hatten wir Fluss des linearen Systems genannt. Eine wichtige Rolle spielt im weiteren der *hyperbolische Fluss*.

Definition 2.38 (Hyperbolischer Fluss)

Falls alle Eigenwerte der Matrix $A \in \mathbb{R}^{n \times n}$ einen nichtverschwindenden Realteil haben, dann wird der Fluss $e^{tA} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ *hyperbolischer Fluss* und (2.63) hyperbolisches lineares System genannt.

Definition 2.39 (Invarianter Unterraum)

Ein Unterraum $E \subset \mathbb{R}^n$ heißt *invariant* in bezug auf den Fluss $e^{At} : \mathbb{R}^n \rightarrow \mathbb{R}^n$, falls $e^{At}(E) \subset E$ für alle $t \in \mathbb{R}$ gilt.

Mit diesen Definitionen lassen sich folgende Eigenschaften über die Unterräume E^i , $i = s, u, c$, hier ohne Beweis, formulieren.

Satz 2.40

Sei A eine reelle $n \times n$ -Matrix. Dann gilt

- i. $\mathbb{R}^n = E^s \oplus E^u \oplus E^c$.
- ii. Die E^i , $i = s, u, c$, sind invariant unter dem Fluss e^{At} von (2.63).

Aus Lemma 2.40 ergibt sich noch eine einfache Folgerung. In Abwesenheit rein imaginärer Eigenwerte von A haben wir die Zerlegung

$$\mathbb{R}^n = E^s \oplus E^u.$$

Wegen der Invarianz der Unterräume E^s und E^u in Bezug auf den Fluss e^{tA} , sind nur die Halborbits (das sind Lösungen, die auf $[t_0, \infty)$ betrachtet werden) beschränkt, deren Anfangspunkt $x(t_0)$ in E^s liegt.

Definition 2.41 (Senke, Quelle)

Falls alle Eigenwerte von A einen negativen (positiven) Realteil haben, dann nennt man $x = 0$ eine *Senke* (*Quelle*) für das lineare System (2.63).

Beispiel 2.42

Das lineare System $\dot{x} = Ax$ mit

$$A := \begin{pmatrix} -3 & 1 & 0 \\ -1 & -3 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

hat die Eigenwerte $\lambda_{1,2} = -3 \pm i$ und $\lambda_3 = -1$. Es ist $E^s = \mathbb{R}^3$ und der Ursprung (er ist einziger Fixpunkt des Systems) ist eine Senke. Der Leser möge sich den Verlauf der Orbits klar machen. ■

2.7.2 Linearisierung nichtlinearer DGL-Systeme

In diesem Abschnitt sollen Voraussetzungen aufgestellt werden, unter denen sich das Phasenportrait des nichtlinearen Systems

$$\dot{x} = f(x), \quad f \in \mathcal{C}^1(\Omega, \mathbb{R}^n), \quad \Omega \subset \mathbb{R}^n, \quad \Omega \text{ offen} \quad (2.64)$$

aus dem linearisierten System am Fixpunkt $x = x_0$, also

$$\dot{y} = Df(x_0)y, \quad f(x_0) = 0 \quad (Df := (\partial f_i / \partial x_j)) \quad (2.65)$$

ableiten lässt. Im Allgemeinen ist es nicht möglich, das DGL-System (2.64) zu lösen. Deshalb ist es wichtig, wenigstens qualitative Informationen über das lokale Verhalten der Lösungen in der Nähe eines Fixpunktes zu erhalten. Wenn dies mit Hilfe des Systems (2.65) möglich ist, dann ist das ein enormer Gewinn, da man von (2.65) (prinzipiell) den Fluss und somit die Lösung

$$\varphi_t(\bar{x}_0) \equiv \varphi(t; \bar{x}_0) = e^{tDf(x_0)}\bar{x}_0, \quad \bar{x}_0 \neq x_0$$

kennt. Die Rechtfertigung dieser Vorgehensweise wird durch das *Theorem von Hartman und Grobman* aus dem Jahr 1959 begründet.

Bei der Analyse des Systems (2.64) gehen wir so vor, dass wir zunächst alle Gleichgewichtspunkte von (2.64) bestimmen und dann das Verhalten der Lösung von (2.64) in der Nähe dieser Fixpunkte untersuchen.

Definition 2.43 (Hyperbolischer Gleichgewichtspunkt, Linearisierung)

Wir betrachten das System (2.64).

1. Ein Fixpunkt x_0 heißt *hyperbolischer Gleichgewichtspunkt*, falls alle Eigenwerte der Matrix $Df(x_0)$ nichtverschwindende Realteile besitzen.
2. Das lineare System $\dot{x} = Ax$ mit der Matrix $A = Df(x_0)$ heißt *Linearisierung* von (2.64) am Fixpunkt x_0 .

Nun wird dargelegt, dass das lokale Lösungsverhalten des nichtlinearen Systems (2.64) in der Nähe eines hyperbolischen Gleichgewichtspunktes qualitativ durch seine Linearisierung im Fixpunkt bestimmt ist. Ist $x_0 = 0$ ein Fixpunkt von (2.64), dann ist $f(0) = 0$ und der Satz von Taylor (entsprechende Differenzierbarkeit vorausgesetzt) liefert

$$f(x) = Df(0)x + \frac{1}{2}D^2f(0)(x, x) + \dots$$

Die lineare Funktion $Df(0)x$ ist in der Regel eine gute erste Approximation für f in der Nähe von $x_0 = 0$. Überträgt man diesen Approximationsgedanken auf die Phasenportraits der nichtlinearen und linearisierten DGL-Systeme, dann wird sich herausstellen, dass dies in der Tat nur für hyperbolische Fixpunkte gilt.

Zunächst werden wir die Begriffe Senke (Quelle) aus Definition 2.41 auf unser System (2.64) übertragen.

Definition 2.44 (Senke, Quelle, Sattel für nichtlineare DGL-Systeme)

Ein Fixpunkt x_0 von (2.64) heißt Senke (Quelle), falls alle Eigenwerte der Matrix $Df(x_0)$ negativen (positiven) Realteil haben. Er wird Sattel genannt, falls x_0 hyperbolischer Fixpunkt ist und mindestens einen Eigenwert mit negativem Realteil und mindestens einen Eigenwert mit positivem Realteil besitzt.

Beispiel 2.45

Wir klassifizieren die Fixpunkte des Systems (2.64) mit

$$f(x) := \begin{pmatrix} x_1 \\ 1 + x_1^2 - x_2^2 \end{pmatrix}.$$

Es gilt $f(x_0) = 0$ in $x_0 = (0, \pm 1)$. Die Linearisierung lautet

$$Df(x) = \begin{pmatrix} 1 & 0 \\ 2x_1 & -2x_2 \end{pmatrix}, \quad Df(0, \pm 1) = \begin{pmatrix} 1 & 0 \\ 0 & \mp 2 \end{pmatrix}.$$

Somit ist der Fixpunkt $(0, 1)$ ein Sattel und $(0, -1)$ eine Quelle. ■

2.7.3 Das Hartman-Grobman-Theorem

Für Aussagen wie „das System geht gegen einen Gleichgewichtszustand“ oder „das System nähert sich für große Zeiten einer periodischen Lösung“ benötigt man lediglich ein Phasenportrait, das bis auf homöomorphe oder diffeomorphe Verzerrungen bekannt ist. Es muss also untersucht werden, ob eine Linearisierung des Vektorfeldes am Fixpunkt das Phasenportrait lediglich in diesem Rahmen verändert. Bei mathematischen Modellierungen von physikalischen Systemen hat man es in der Regel mit Parametern zu tun,

die nicht exakt bekannt sind und die somit auch das Vektorfeld verändern. Da ist es natürlich erwünscht, dass diese (kleinen) Störungen des Vektorfeldes f höchstens zu homöomorphen (also stetigen) Veränderungen des Phasenbildes führen.

Der nächste Satz, der ohne Beweis angegeben wird, begründet die Existenz einer glatten invarianten *Mannigfaltigkeit* in einem Fixpunkt des nichtlinearen Systems (2.64). Der hier auftretende Begriff der Mannigfaltigkeit wird nur in dem Rahmen, wie er im Weiteren benötigt wird, erklärt. Mit dem Begriff *Invarianz* ist wieder die Invarianz der Mannigfaltigkeit unter dem Fluss φ_t von (2.64) zu verstehen. Wir können uns auf die Klasse von Mannigfaltigkeiten, die man als Untermannigfaltigkeiten des \mathbb{R}^n auffasst, beschränken. Diese sind reichhaltig genug, um damit beim Studium von Differentialgleichungen auszukommen.

Definition 2.46 (Mannigfaltigkeit)

Eine Menge $M \subset \mathbb{R}^n$ heißt n -dimensionale Mannigfaltigkeit, falls Folgendes gilt:

- i. **Lokale Koordinaten.** Zu jedem Punkt $x \in M$ existieren eine Teilmenge $U \subset M$, die den Punkt x enthält und eine bijektive Abbildung

$$\varphi : U \rightarrow U_\varphi$$

vermittelt, wobei U_φ eine offene Menge des \mathbb{R}^n ist. Die Abbildung φ heißt Kartenabbildung und U_φ nennt man das Kartenbild von U . Das Paar (U, φ) heißt eine Karte von M und $x_\varphi = \varphi(x) = (x_1, \dots, x_n)$ heißt lokale Koordinate des Punktes x in der Karte (U, φ) .

- ii. **Wechsel der lokalen Koordinaten.** Ist (V, ψ) eine zweite Karte für den Punkt x mit der zugehörigen lokalen Koordinate

$$x_\psi = \psi(x),$$

dann ergeben sich für die beiden lokalen Koordinaten des Punktes x die folgenden Transformationsformeln:

$$x_\varphi = \varphi(\psi^{-1}(x_\psi)) \quad \text{bzw.} \quad x_\psi = \psi(\varphi^{-1}(x)).$$

Wir verlangen, dass die beiden zugehörigen Abbildungen

$$\varphi \circ \psi^{-1} : V_\psi \rightarrow U_\varphi \quad \text{bzw.} \quad \psi \circ \varphi^{-1} : U_\varphi \rightarrow V_\psi$$

aus C^k sind. In diesem Fall spricht man von einer C^k -Mannigfaltigkeit.

Die Abbildung φ versteht sich hier als eine Kartenabbildung. Eine Verwechslung mit der Lösung einer DGL besteht nicht.

Anschauliche Interpretation: Wir wählen als Mannigfaltigkeit M die Erdoberfläche (siehe Abbildung 2.17). Ein geographischer Atlas für M besteht dann aus geographischen Karten, die Teilmengen des \mathbb{R}^2 sind. Jede geographische Karte ist das Bild eines Teiles der Erdoberfläche. Dabei kann ein Punkt x der Erdoberfläche in verschiedenen geographischen Karten auftreten (siehe Definition 2.46 ii.).

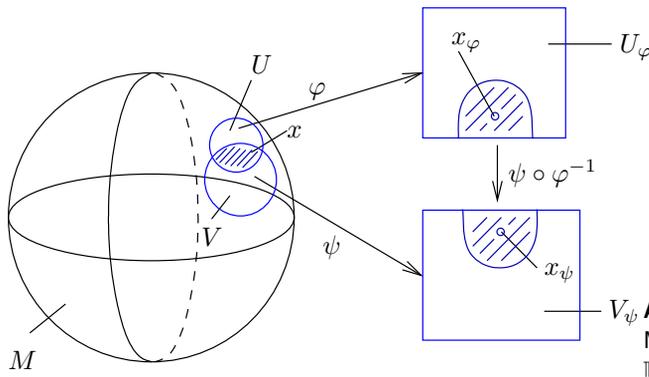


Abb. 2.17 Zweidimensionale Mannigfaltigkeit, eingebettet in \mathbb{R}^3

Den Prototyp für eine eindimensionale Mannigfaltigkeit stellt die Kreislinie dar. Die Kreislinie muss durch mindestens zwei Karten beschrieben werden, da sonst die Bijektivität der Abbildung φ nicht gewährleistet ist. Dabei herrscht eine gewisse Willkür. Um diese zu vermeiden, wird der maximale Atlas eingeführt. Darauf gehen wir hier nicht weiter ein. Der interessierte Leser möge sich in den Büchern von Zeidler (1986) oder Choquet-Bruhat (1983,1988) einen Überblick verschaffen.

Eine Mannigfaltigkeit kann auch durch ein System von Gleichungen

$$f_j(x_1, \dots, x_n) = 0, \quad j = 1, \dots, m, \quad 1 \leq m < n \quad (2.66)$$

erfasst werden. Die Funktionen f_j sind auf der offenen Menge $\Omega \subset \mathbb{R}^n$ glatt, d. h. C^k -Funktionen. Wir verlangen weiter, dass die Funktionalmatrix

$$\frac{\partial(f_1, \dots, f_m)}{\partial(x_1, \dots, x_n)} \quad (2.67)$$

den maximalen Rang m für jeden Punkt $x = (x_1, \dots, x_n) \in \Omega$ besitzt. Dann gilt der

Satz 2.47

Die Menge aller Punkte $x \in \Omega$, die Lösungen der Gleichung (2.66) sind und für die (2.67) maximalen Rang besitzt, bildet eine $(n - m)$ -dimensionale Mannigfaltigkeit.

Man bezeichnet M in diesem Fall auch als eine $(n - m)$ -dimensionale *Untermannigfaltigkeit* des \mathbb{R}^n .

Beweis: Die Aussage ist eine einfache Konsequenz des Satzes über implizite Funktionen. \square

Grob gesprochen heißt eine Teilmenge $N \subset M$ (wobei M eine n -dimensionale Mannigfaltigkeit im Sinne der gegebenen Definition 2.46 ist) eine m -dimensionale Untermannigfaltigkeit von M , falls N lokal wie ein m -dimensionaler linearer Unterraum von \mathbb{R}^n aussieht. Der Tangentialraum $T_x M$ von M im Punkt x besteht in diesem Spezialfall (siehe das Gleichungssystem (2.66)) aus allen reellen n -Tupeln $v = (v_1, \dots, v_n)$, die dem linearisierten System

$$\sum_{k=1}^n \frac{\partial f_j(x)}{\partial x_k} v_k = 0, \quad j = 1, \dots, m, \quad (2.68)$$

genügen. Da der Rang der Matrix (2.67) gleich m ist, stellt $T_x M$ einen $(n - m)$ -dimensionalen linearen Vektorraum dar.

Beispiel 2.48

Für festes $r > 0$ bildet die Menge S^{n-1} ($(n - 1)$ -dimensionale Sphäre im \mathbb{R}^n) aller Lösungen $(x_1, \dots, x_n) \in \mathbb{R}^n$ der Gleichung

$$\sum_{j=1}^n x_j^2 - r^2 = 0 \quad (2.69)$$

eine $(n - 1)$ -dimensionale Mannigfaltigkeit. Die entsprechende Matrix der Linearisierung $(2x_1, \dots, 2x_n)$ hat den Rang 1 in jedem Lösungspunkt von (2.69), weil mindestens eine der Komponenten x_j ungleich Null sein muss. Der Tangentialraum $T_x S^{n-1}$ besteht aus allen n -Tupeln $v = (v_1, \dots, v_n)$ mit $\sum_{j=1}^n x_j v_j = 0$. \blacksquare

Besitzt (2.67) in jedem Punkt x , der Lösung der Gleichung (2.66) ist, den Rang s mit $1 \leq s \leq m < n$, dann bildet die Lösungsmenge von (2.66) eine $(n - s)$ -dimensionale Mannigfaltigkeit M . Der Tangentialraum $T_x M$ wird wieder durch (2.68) gegeben.

Nach diesem kleinen Ausflug in die Begriffswelt der Mannigfaltigkeiten kehren wir zum Studium des Phasenportraits unserer DGL (2.64) in Verbindung mit (2.65) zurück.

Satz 2.49 (Existenz invarianter Mannigfaltigkeiten)

Es sei $\Omega \subset \mathbb{R}^n$ eine offene Menge mit $0 \in \Omega$ und $f \in C^1(\Omega, \mathbb{R}^n)$. Weiter sei φ_t der Fluss des nichtlinearen Systems (2.64), das Vektorfeld f besitzt in $x_0 = 0$ einen Fixpunkt und die Linearisierung $Df(0)$ hat k Eigenwerte mit negativem Realteil und $n - k$ Eigenwerte mit positivem Realteil. Dann existiert eine k -dimensionale differenzierbare Mannigfaltigkeit S , die im Nullpunkt tangential zum stabilen Unterraum E^s des linearisierten Systems (2.65) ist. Für $t \geq 0$ gilt $\varphi_t(S) \subset S$ und außerdem

$$\lim_{t \rightarrow \infty} \varphi_t(\tilde{x}) = 0, \quad \forall \tilde{x} \in S.$$

Des Weiteren existiert eine $n - k$ dimensionale differenzierbare Mannigfaltigkeit U , die im Fixpunkt tangential zum instabilen Unterraum E^u von (2.65) ist. Für $t \leq 0$ gilt $\varphi_t(U) \subset U$ und außerdem

$$\lim_{t \rightarrow -\infty} \varphi_t(\tilde{x}) = 0, \quad \forall \tilde{x} \in U.$$

Beweis: Die Idee besteht darin, eine geeignete Fixpunktgleichung in einem passenden Banach-Raum aufzustellen, so dass der Fixpunkt eine C^1 -Abbildung zwischen \mathbb{R}^n und \mathbb{R}^{n-k} wird und sein Graph die gewünschte invariante Mannigfaltigkeit darstellt (siehe Perko (1996)). \square

Bemerkung 2.50

Falls $f \in C^r(\Omega, \mathbb{R}^n)$, $r \geq 1$, dann sind auch S und U von der Klasse C^r . Ist f analytisch in Ω , dann sind auch S und U invariante analytische Mannigfaltigkeiten. \blacksquare

Beispiel 2.51

Wir betrachten das System

$$\dot{x}_1 = -x_1 - x_2^2, \quad \dot{x}_2 = x_2 + x_1^2. \quad (2.70)$$

Es hat die Fixpunkte $(0, 0)$ und $(-1, -1)$. Wir wollen die invariante stabile Mannigfaltigkeit an E^s im Fixpunkt $(0, 0)$ näherungsweise berechnen. Die Linearisierung des Systems (2.70) lautet

$$\dot{y} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} y, \quad y \in \mathbb{R}^2.$$

Daraus ergeben sich die stabilen und instabilen Unterräume $E^s = \text{span}\{(1, 0)^T\}$ und $E^u = \text{span}\{(0, 1)^T\}$. Nun wird die invariante Mannigfaltigkeit S näherungsweise mittels Potenzreihenansatz gemäß Satz 2.49 und Bemerkung 2.50 berechnet. Wir machen den Ansatz $x_2 = \psi(x_1) = a_2x_1^2 + a_3x_1^3 + \dots$. Aus dem System (2.70) ergibt sich zunächst durch Einsetzen aus der zweiten Gleichung

$$\dot{x}_2 = \psi'(x_1) \cdot \dot{x}_1 = \psi'(x_1)[-x_1 - [\psi(x_1)]^2] = \psi(x_1) + x_1^2$$

eine Bestimmungsgleichung für die unbestimmten Koeffizienten der Potenzreihe. Aus

$$\begin{aligned} (2a_2x_1 + 3a_3x_1^2 + 4a_4x_1^3 + 5a_5x_1^4 + \dots)[-x_1 - a_2^2x_1^4 - 2a_2a_3x_1^5 - \dots] \\ = a_2x_1^2 + a_3x_1^3 + a_4x_1^4 + a_5x_1^5 + \dots + x_1^2 \end{aligned}$$

folgt durch Koeffizientenvergleich

$$a_2 = -\frac{1}{3}, \quad a_3 = a_4 = 0, \quad a_5 = \left(\frac{1}{3}\right)^4.$$

Somit ergibt sich angenähert für S die Darstellung

$$x_2 = -\frac{1}{3}x_1^2 + \left(\frac{1}{3}\right)^4 x_1^5 + o(x_1^5), \quad |x_1| < \delta, \quad \delta > 0.$$

Es ist klar, dass diese Darstellung nur lokal gilt. Abbildung 2.18 zeigt stabile und instabile Mannigfaltigkeiten S und U . Der Leser möge mit einer entsprechenden Rechnung eine genäherte analytische Darstellung für U finden. ■

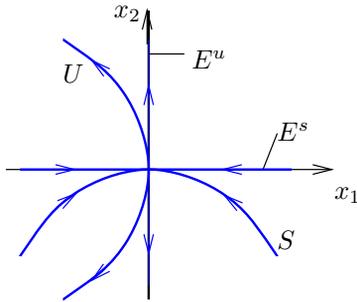


Abb. 2.18 Stabile und instabile invariante Mannigfaltigkeiten E^s und E^u

Da stabile und instabile Mannigfaltigkeit S und U sind nur in einer kleinen Umgebung des Fixpunktes definiert sind, bezeichnen wir sie als lokale stabile und instabile Mannigfaltigkeit. Wir definieren die globale stabile und instabile Mannigfaltigkeit von (2.64) im Ursprung, indem wir Punkte aus S rückwärts in der Zeit fließen lassen und jene in U in der Zeit vorwärts fließen lassen:

$$W^s(0) = \bigcup_{t \leq 0} \varphi_t(S), \quad W^u(0) = \bigcup_{t \geq 0} \varphi_t(U).$$

Man kann zeigen, dass die globalen Mannigfaltigkeiten $W^s(0)$ und $W^u(0)$ eindeutig bestimmt und invariant unter dem Fluss φ_t sind. Außerdem erhält man:

$$\forall x \in W^s(0) : \lim_{t \rightarrow \infty} \varphi_t(x) = 0, \quad \forall x \in W^u(0) : \lim_{t \rightarrow -\infty} \varphi_t(x) = 0.$$

„Nicht ganz analog“ stellt sich ein entsprechendes Resultat über die Zentrumsmannigfaltigkeit $W^c(0)$ dar. Sie ist zwar tangential zum Zentrumsunterraum E^c , aber man kann sie nicht einfach mit Hilfe des asymptotischen Verhaltens von Lösungen charakterisieren wie Abbildung 2.20 zeigt. Es gilt aber der

Satz 2.52 (Zentrumsmannigfaltigkeit)

Sei f wie in Satz 2.49 vorausgesetzt. Weiter wird angenommen, dass $f(0) = 0$ und die Matrix $Df(0)$ k Eigenwerte mit negativem Realteil, j Eigenwerte mit positivem Realteil und $m = n - k - j$ Eigenwerte mit Realteil Null besitzt. Dann existiert eine m -dimensionale Zentrumsmannigfaltigkeit $W^c(0)$ der Klasse C^r , die tangential zum Zentrumsunterraum E^c in $x_0 = 0$ verläuft und invariant unter dem Fluss φ_t ist.

Auf den Beweis verzichten wir und verweisen stattdessen auf die Literatur (Perko (1996) bzw. Guckenheimer und Holmes (1983)). \square

Um den angekündigten Satz von Hartman und Grobman formulieren zu können, benötigen wir noch eine

Definition 2.53 (topologisch äquivalent, topologisch konjugiert)

Zwei DGL-Systeme $\dot{x} = f(x)$ und $\dot{y} = Df(x_0)y$ heißen *topologisch äquivalent* in einer Umgebung eines Punktes x_0 , wenn ein Homöomorphismus $H : U \rightarrow V$ (U und V beides offenen Mengen, die den Punkt x_0 enthalten) existiert, der die Trajektorien von $\dot{x} = f(x)$ in U auf die Trajektorien von $\dot{y} = Df(x_0)y$ in V abbildet und die Orientierung der Phasenkurven erhält.

Die genannten Systeme heißen *topologisch konjugiert*, wenn darüber hinaus die Parametrisierung der Trajektorien in der Zeit erhalten bleibt.

Es werden nun die Flüsse der DGL-Systeme

$$\dot{x} = f(x) \quad \text{und} \quad \dot{y} = Ay \quad \text{mit} \quad A := Df(x_0)$$

in der Nähe eines hyperbolischen Gleichgewichtspunktes x_0 miteinander verglichen. Dabei wird angenommen, dass der Gleichgewichtspunkt x_0 in die Null transformiert wurde.

Satz 2.54 (Hartman und Grobman)

Es seien die offene Menge $\Omega \subset \mathbb{R}^n$ mit $0 \in \Omega$, $f \in C^1(\Omega, \mathbb{R}^n)$, $f(0) = 0$ und φ_t der Fluss des Systems $\dot{x} = f(x)$ gegeben. Weiter möge die Matrix $A := Df(0)$ nur Eigenwerte mit Realteil verschieden von Null haben. Dann gibt es einen Homöomorphismus $H : U \rightarrow V$ ($U, V \subset \mathbb{R}^n$ offene Nullumgebungen), so dass für jedes $\bar{x} \in U$ ein offenes Intervall $I_{\bar{x}} \subset \mathbb{R}$ mit $0 \in I_{\bar{x}}$ existiert und für alle $\bar{x} \in U$ und alle $t \in I_{\bar{x}}$ gilt

$$H \circ \varphi_t(\bar{x}) = e^{tA} H(\bar{x}). \quad (2.71)$$

Mit anderen Worten: Die beiden DGL-Systeme $\dot{x} = f(x)$ und $\dot{y} = Df(0)y$ sind topologisch äquivalent in einer Umgebung des Punktes 0. Darüber hinaus ist die Abbildung H topologisch konjugiert. Sie bildet Orbits von $\dot{x} = f(x)$ auf Orbits von $\dot{y} = Df(0)y$ nahe dem Ursprung ab und erhält die Parametrisierung der Orbits bzgl. der Zeit t . Der interessierte Leser kann den Beweis in Perko (1996) nachlesen.

Manchmal gelingt es, den Homöomorphismus explizit anzugeben. Dies wollen wir an folgendem Beispiel demonstrieren:

Beispiel 2.55

Wir betrachten das System

$$\begin{aligned} \dot{x}_1 &= -x_1 \\ \dot{x}_2 &= x_2 + x_1^2, \end{aligned} \quad x(0) = x_0 = \begin{pmatrix} x_{01} \\ x_{02} \end{pmatrix} \quad (2.72)$$

mit $x_0 \in \mathbb{R}^2$ beliebig aber fest vorgegeben. Aufgrund der einfachen Struktur des nichtlinearen Systems kann der Fluss explizit angegeben werden:

$$\varphi_t(x_0) = \varphi(t; x_0) = \begin{pmatrix} x_{01}e^{-t} \\ x_{02}e^t + \frac{x_{01}^2}{3}(e^t - e^{-2t}) \end{pmatrix}, \quad \forall t \in \mathbb{R}.$$

Das System (2.72) besitzt als einzigen Fixpunkt den Ursprung. Wir suchen Mengen, die invariant unter dem Fluss bleiben und für die $\lim_{t \rightarrow \infty} \varphi_t(x_0) = 0$ bzw. $\lim_{t \rightarrow -\infty} \varphi_t(x_0) = 0$ gilt, sofern man den Punkt x_0 aus diesen Mengen nimmt. Wählen wir die Anfangswerte x_0 aus der Menge $S := \{x_0 \in \mathbb{R}^2 \mid x_{02} = -x_{01}^2/3\}$, dann ist diese Menge invariant unter dem Fluss φ_t :

$$x_0 \in S \quad \Rightarrow \quad \varphi_t(x_0) = \begin{pmatrix} x_{01}e^{-t} \\ -\frac{x_{01}^2}{3}e^{-2t} \end{pmatrix} \in S, \quad (t \geq 0).$$

Eliminiert man t , dann erhält man im Phasenraum die Beziehung $x_2 = -\frac{x_1^2}{3}$, was die Zugehörigkeit zu S bedeutet.

Die Menge $U := \{\bar{x}_0 \in \mathbb{R}^2 \mid x_{01} = 0, x_{02} \in \mathbb{R}\}$ ist ebenfalls invariant unter dem Fluss φ_t :

$$\bar{x}_0 \in U \quad \Rightarrow \quad \varphi_t(\bar{x}_0) = \begin{pmatrix} 0 \\ x_{02}e^t \end{pmatrix} \in U, \quad (t \geq 0).$$

Mit der Kenntnis der Mengen S und U lässt sich das Phasenportrait bereits qualitativ zeichnen (siehe Abbildung 2.19, (rechts)). Da es sich bei dem Fixpunkt Null um einen Sattelpunkt handelt, kann man zum Zeichnen der Orbits auch das linearisierte System

$$\dot{y}_1 = -y_1, \quad \dot{y}_2 = y_2 \quad (2.73)$$

zu Hilfe nehmen (siehe Abbildung 2.19, (links)). Dabei ist auch schön zu sehen, dass S und U tangential im Nullpunkt an $E^s = \text{span}\{(0, 1)^T\}$ und $E^u = \text{span}\{(1, 0)^T\}$ anliegen. Man nennt diese Mengen S und U auch stabile und instabile Mannigfaltigkeiten. Aus der Abbildung 2.19 geht hervor, was damit gemeint ist. Auf eine mathematische Begriffsprägung wird verzichtet.

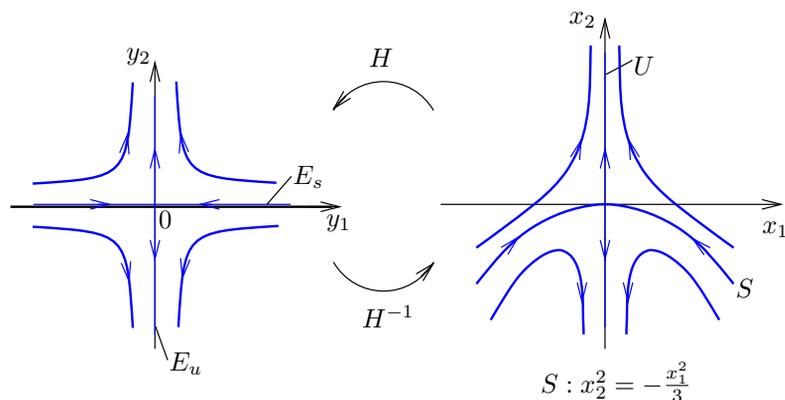


Abb. 2.19 Linearisiertes System (2.73) (links), DGL-System (2.72) (rechts)

Die stabile Mannigfaltigkeit eignet sich als guter Kandidat für einen geeigneten Homöomorphismus H , der nach dem Satz von Hartman und Grobman existiert. Wir betrachten die stetige Abbildung

$$H : \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad H(x_1, x_2) := \begin{pmatrix} x_1 \\ x_2 + \frac{x_1^2}{3} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}. \quad (2.74)$$

Die Inverse von H ist leicht zu berechnen:

$$H^{-1}(y_1, y_2) = \begin{pmatrix} y_1 \\ y_2 - \frac{y_1^2}{3} \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Sie ist wieder eine stetige Abbildung von \mathbb{R}^2 auf \mathbb{R}^2 . Darüber hinaus transformiert die Abbildung unser nichtlineares System (2.72) in das linearisierte System (2.73). Aus (2.74) und (2.72) folgt

$$\begin{aligned} \dot{x}_1 &= \dot{y}_1 = -x_1 = -y_1 \\ \dot{x}_2 &= \dot{y}_2 - \frac{2}{3}y_1 \dot{y}_1 = x_2 + x_1^2 = y_2 - \frac{y_1^2}{3} + y_1^2 \end{aligned}$$

Ersetzt man in der letzten Gleichung noch \dot{y}_1 durch $-y_1$, dann lautet das transformierte „ y -System“:

$$\dot{y}_1 = -y_1, \quad \dot{y}_2 = y_2.$$

Dabei handelt es sich genau um das aus der Linearisierung von (2.72) gewonnene System (2.73). Für die jeweiligen Flüsse rechnet man mit H folgendes nach (Es wird aus Gründen der Vereinfachung $\bar{x} = x$ gesetzt.):

$$\begin{aligned} H(\varphi_t x) &= \begin{pmatrix} x_1 e^{-t} \\ x_2 e^t + \frac{x_1^2}{3} (e^t - e^{-2t}) + \frac{x_2^2}{3} e^{-2t} \end{pmatrix} \\ &= \begin{pmatrix} x_1 e^{-t} \\ x_2 e^t + \frac{x_1^2}{3} e^t \end{pmatrix} = \begin{pmatrix} x_1 e^{-t} \\ (x_2 + \frac{x_1^2}{3}) e^t \end{pmatrix} = \begin{pmatrix} y_1 e^{-t} \\ y_2 e^t \end{pmatrix} \\ &= \begin{pmatrix} e^{-t} & 0 \\ 0 & e^t \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = e^{tA} y = e^{tA} H(x). \end{aligned}$$

Dies bestätigt noch einmal die Beziehung (2.71) in Satz 2.54. ■

Beispiel 2.56

Wir betrachten das System

$$\dot{x}_1 = x_1^2, \quad \dot{x}_2 = -x_2. \quad (2.75)$$

Die Linearisierung am (einzigen) Fixpunkt $x_0 = (0, 0)$ liefert das System

$$\dot{y} = Ay \quad \text{mit} \quad A := Df(0, 0) = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{und} \quad y \in \mathbb{R}^2. \quad (2.76)$$

Die Abbildung 2.20 zeigt das Phasenportrait von (2.75). Jede Lösungskurve die in der

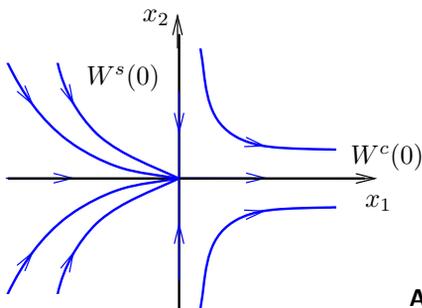


Abb. 2.20 Phasenportrait für das System (2.75)

linken Halbebene startet, kann man sich im Fixpunkt mit der positiven x_1 -Achse „zusammengeflickt“ denken. Sie bilden eine eindimensionale Zentrumsmannigfaltigkeit, die tangential zu E^c im Ursprung ist. Der Fixpunkt $x_0 = (0, 0)$ des Systems (2.75) ist nicht hyperbolisch, da die Linearisierung (2.76) einen Nulleigenwert aufweist. Der Satz von

Hartman und Grobman ist somit nicht anwendbar. Um zu sehen, dass die beiden Phasenportraits tatsächlich nichts miteinander zu tun haben, zeichnen wir die Orbits von (2.76) (siehe Abbildung 2.21). Der stabile Unterraum E^s von (2.76) ist die x_2 -Achse, der

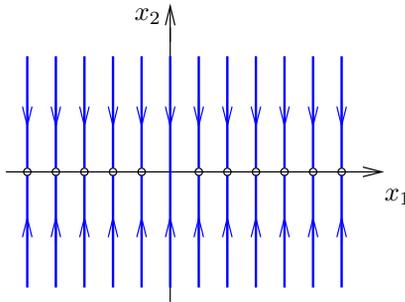


Abb. 2.21 Phasenportrait des Systems (2.76)

Zentrumsunterraum E^c die x_1 -Achse. Der Fluss lautet

$$\varphi_t(x_0) = e^{tA} x_0 = \begin{pmatrix} 0 & 0 \\ 0 & e^{-t} \end{pmatrix} \begin{pmatrix} x_{01} \\ x_{02} \end{pmatrix} = \begin{pmatrix} x_{01} \\ x_{02} e^{-t} \end{pmatrix}, \quad t \geq 0, \quad x_0 \in \mathbb{R}^2.$$

Außerdem enthält das Phasenportrait die x_1 -Achse als Fixpunktgerade. ■

2.8 Aufgaben

Aufgabe 2.1

Skizzieren Sie folgende Kurvenscharen (c , c_1 , c_2 -Scharparameter) und bestimmen Sie ihre dazugehörigen Differenzialgleichungen niedrigster Ordnung:

- a) $y = \frac{c}{1+x^2}$, $c \in \mathbb{R}$, b) $y = \frac{c_1}{4}x^4 + c_2$, $c_1, c_2 \in \mathbb{R}$,
 c) $x^2 + y^2 = c$, $c \in \mathbb{R}$, d) $y = e^{cx}$, $c \in \mathbb{R}$.

Aufgabe 2.2

Skizzieren Sie das Richtungsfeld von a) $\dot{x} = t + x$, b) $\dot{x} = t^2 + x^2$.

Welche Gestalt der Integralkurven entnimmt man dem Richtungsfeld?

Aufgabe 2.3

Lösen Sie das AWP

$$\dot{x} = \frac{1}{t} - \frac{1}{2\sqrt{1-t}}, \quad x(t_0) = x_0, \quad D = (0, 1) \times \mathbb{R}$$

und untersuchen Sie das Verhalten der Lösungen am Rand von D .

Aufgabe 2.4

Lösen Sie das AWP

$$\dot{x} = 2tx^2, \quad D = \mathbb{R}^2, \quad x(t_0) = x_0$$

und skizzieren Sie die Lösungskurven für die Anfangswerte

- a)** $(t_0, x_0) = (1, 1)$, **b)** $(t_0, x_0) = (2, -\frac{1}{3})$, **c)** $(t_0, x_0) = (1, -\frac{1}{2})$, **d)** $(t_0, x_0) = (1, 0)$.

Aufgabe 2.5

Es sei $T : C^0[0, 1] \rightarrow C^0[0, 1]$ definiert durch

$$(Tu)(t) := \frac{1}{2} t^2 + \int_0^t s \cdot u(s) \, ds, \quad t \in [0, 1], \quad u \in C^0([0, 1]).$$

- a)** Zeigen Sie: T ist kontraktiv.
b) Es sei $(u_k)_{k \in \mathbb{N}}$ rekursiv definiert durch $u_1 = 0$ und $u_{k+1} = Tu_k$ ($k \in \mathbb{N}$). Geben Sie eine Darstellung für $u_k(t)$, $k \in \mathbb{N} \setminus \{1\}$, $t \in [0, 1]$ an.
c) Bestimmen Sie $u^* \in C^0([0, 1])$ mit $\|u_k - u^*\| \rightarrow 0$ für $k \rightarrow \infty$.
d) Geben Sie über $Tu^* = u^*$ ein AWP an, für das u^* eine Lösung ist.

Aufgabe 2.6

Berechnen Sie für ein festes $\alpha \in \mathbb{R}$ die Lösung des AWP $\dot{x} = \alpha x$, $x(t_0) = x_0$ mittels der *Picard-Iteration*.

Hinweis: Die Methode der sukzessiven Approximation lässt sich auch auf Differenzialgleichungen übertragen, in dem man das AWP $\dot{x} = f(t, x)$, $x(t_0) = x_0$ durch Integration in eine Integralgleichung $x(t) = x_0 + \int_{t_0}^t f(s, x(s)) \, ds$ umformt und dann die Picard-Iteration definiert:

$$x_{k+1}(t) := x_0 + \int_{t_0}^t f(s, x_k(s)) \, ds, \quad k \in \mathbb{N}_0. \quad (*)$$

Als Startwert nimmt man $x_0(t) \equiv x_0$. Kann man die (gleichmäßige) Konvergenz der Funktionenfolge $(x_k(t))$ beweisen, dann folgt durch Grenzübergang in $(*)$

$$x_\infty = x_0 + \int_{t_0}^t f(s, x_\infty(s)) \, ds,$$

und letztlich die Lösung des AWP.

Aufgabe 2.7

Lösen Sie das AWP

$$(1 - x^2)y'' - xy' = 2, \quad y(0) = y'(0) = 0, \quad |x| < 1,$$

durch einen *Potenzreihenansatz* $y(x) = a_0 + \sum_{n=1}^{\infty} a_n x^n$. Bekanntlich ist

$$\arcsin x = x + \frac{1}{2} \frac{x^3}{3} + \frac{1 \cdot 3}{2 \cdot 4} \frac{x^5}{5} + \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6} \frac{x^7}{7} + \dots, \quad \text{für } |x| < 1,$$

und durch Cauchysche Multiplikation gewinnt man daraus

$$(\arcsin x)^2 = x^2 + \frac{x^4}{3} + \frac{8}{45}x^6 + \dots, \quad \text{für } |x| < 1.$$

Geben Sie unter Berücksichtigung, dass $(\arcsin x)^2$ dem oben gegebenen AWP genügt, ein einfaches Bildungsgesetz für die Koeffizienten der entsprechenden Reihe an.

Aufgabe 2.8

Lösen Sie die Differentialgleichung $\ddot{x} = g(x)$ mit der *Energiemethode*: Überführen Sie $\ddot{x} = g(x)$ in $\dot{x}\ddot{x} = g(x)\dot{x}$. Nutzen Sie die Beziehung $\dot{x}\ddot{x} = \frac{1}{2}(\dot{x}^2)'$, um $\ddot{x} = g(x)$ in eine Differentialgleichung 1. Ordnung zu überführen. Lösen Sie die folgenden Differentialgleichungen

$$\text{a) } x^3\ddot{x} = 1, \quad \text{b) } \ddot{x} = x$$

nach der *Energiemethode*.

Aufgabe 2.9

Es seien zwei Lösungen der zu

$$\dot{x} = 2t \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} x + \begin{pmatrix} \cos t^2 \\ -\sin t^2 \end{pmatrix}$$

gehörenden homogenen Differentialgleichung durch

$$x_1(t) = (\cos t^2, -\sin t^2)^T \quad \text{und} \quad x_2(t) = (\sin t^2, \cos t^2)^T$$

gegeben. Berechnen Sie durch das Verfahren der Variation der Konstanten eine spezielle Lösung der angegebenen Differentialgleichung.

Aufgabe 2.10

Gegeben sei die Differentialgleichung

$$\dot{x} = \frac{t^3 x^3}{1 + x^2} + e^t \cos(x).$$

Geben Sie das maximale Lösungsintervall des zugehörigen AWP mit $x(t_0) = x_0$ an. Begründen Sie ihre Aussage! Zeichnen Sie zusätzlich das Richtungsfeld und die Integralkurven unter Zuhilfenahme eines Computeralgebrasystems.

Aufgabe 2.11

Für die folgenden linearen inhomogenen DGL sind jeweils eine allgemeine Lösung der linearen homogenen und eine partikuläre Lösung der linearen inhomogenen DGL zu berechnen.

$$\text{a) } y'' + y = x \sin x, \quad \text{b) } y'' + 3y' - 4y = e^{-4x} + xe^{-x}.$$

Hinweis: Man kann durch spezielle Ansätze partikuläre Lösungen linearer inhomogener DGL n -ter Ordnung mit konstanten Koeffizienten berechnen. Hat die Inhomogenität die Struktur

$$(b_0 + b_1x + \dots + b_mx) e^{\alpha x} \cdot \begin{cases} \cos \beta x \\ \sin \beta x \end{cases},$$

so wählt man den Ansatz

$$((A_0 + A_1x + \dots + A_mx^m) \cos \beta x + (B_0 + B_1x + \dots + B_mx^m) \sin \beta x) e^{\alpha x},$$

falls $\alpha + i\beta$ keine Nullstelle des charakteristischen Polynoms p ist. Ist $\alpha + i\beta$ eine ν -fache Nullstelle, so setzt man an

$$x^\nu ((A_0 + A_1x + \dots + A_mx^m) \cos \beta x + (B_0 + B_1x + \dots + B_mx^m) \sin \beta x) e^{\alpha x}.$$

Ist die Inhomogenität f der Differenzialgleichung

$$L(y) = f \quad \text{mit} \quad L(y) := a_n \frac{d^n}{dx^n}(y) + \dots + a_1 \frac{d}{dx}(y) + a_0 y, \quad a_n \neq 0$$

eine Summe von Funktionen f_i , $i = 1, \dots, m$, so bestimmt man Partikulärlösungen y_{p_i} zu den Gleichungen $L(y) = f_i$. Die Summe der y_i ist dann eine Lösung von $L(y) = f$ (Superpositionsprinzip).

Aufgabe 2.12

Beweisen Sie für eine lineare homogene DGL n -ter Ordnung mit konstanten Koeffizienten a_i , $i = 0, \dots, n-1$ (siehe (2.49)) die folgende Behauptung: *Die geometrische Vielfachheit eines k -fachen Eigenwertes λ_0 ist $m(\lambda_0) = n - \text{rg}(A - \lambda_0 I) = 1$.*

Hinweis: Verwenden Sie für die DGL die Darstellung (2.50) mit $b(t) \equiv 0$. A ist die Matrix $A(t)$ mit konstanten Einträgen a_i , $i = 0, \dots, n-1$.

Aufgabe 2.13

Lösen Sie folgende Systeme gewöhnlicher Differenzialgleichungen mit konstanten Koeffizienten:

$$\text{a) } \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 2 & 8 \\ 3 & -8 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad \text{b) } \begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} -1 & 1 & -1 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Aufgabe 2.14

Lösen Sie folgende Systeme gewöhnlicher Differenzialgleichungen mit zeitabhängigen Koeffizienten durch Umwandlung in eine Differenzialgleichung 2. Ordnung:

$$\text{a) } \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 0 & -2/t^2 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad \text{b) } \begin{pmatrix} 2t^2 \dot{x} \\ 2t \dot{y} \end{pmatrix} = \begin{pmatrix} -t & 1 \\ t & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Aufgabe 2.15

Lösen Sie die AWP

a) $\dot{x} = x + y, \quad \dot{y} = 4x - 2y, \quad x(0) = 0, y(0) = 5,$

b) $\dot{x} = 3x - 4y, \quad \dot{y} = x - y, \quad x(0) = 3, y(0) = 1,$

c) $\dot{x} = 4x + y - 36t, \quad \dot{y} = -2x + y - 2e^t, \quad x(0) = 0, y(0) = 10.$

Aufgabe 2.16

Bestimmen Sie ein reelles Fundamentalsystem zum System

$$\dot{x} = Ax, \quad x \in \mathbb{R}^4, \quad A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & -2 & 3 & -2 \\ 2 & -1 & 4 & -1 \end{pmatrix}.$$

Aufgabe 2.17Skizzieren Sie das Phasenportrait des Systems $\dot{x} = Ax, x \in \mathbb{R}^2$, für jeden der angegebenen Fälle. Untersuchen Sie den Fixpunkt $x_0 = (0, 0)$ auf Stabilität.

a) $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \text{b) } \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \text{c) } \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$

d) $\begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}, \quad \text{e) } \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \text{f) } \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix}.$

Aufgabe 2.18

Skizzieren Sie die Phasenportraits der folgenden nichtlinearen Systeme in der Nähe ihrer Fixpunkte.

a) $\dot{x}_1 = x_1(1 - x_2) \quad \text{b) } \dot{x}_1 = x_1(1 - x_1)$
 $\dot{x}_2 = x_2(1 - x_1), \quad \dot{x}_2 = -2x_2(1 - \beta x_1) \quad (\beta \in \mathbb{R}).$

Aufgabe 2.19

Fertigen Sie eine Beweisskizze des Satzes von Hartman und Grobman (Satz 2.54) an, indem Sie die Literatur zu Hilfe nehmen (z. B. Perko (1996) S.119).

3 Bifurkation bei gewöhnlichen DGL

Übersicht

3.1	Strukturelle Stabilität	128
3.2	Einige typische Bifurkationen	133
3.3	Aufgaben	148

Änderungen von Kontrollparametern in DGL-Systemen können zu ganz neuen Langzeitmustern der Bewegung führen. Kleine Änderungen in der Erregerfrequenz, der Erregeramplitude oder der Dämpfung können qualitative Änderungen im physikalischen Verhalten hervorrufen.

Die Aufgabe einer Bifurkationsanalyse ist es, diese kritischen Verzweigungsparameter μ_{cr} zu bestimmen und die für $\mu > \mu_{cr}$ abzweigenden neuen Lösungen zu konstruieren. Eine Systematisierung der abzweigenden Lösungen, wie sie für eindimensionale Systeme mit einem Kontrollparameter möglich ist, gestaltet sich für höherdimensionale Systeme mit mehreren Parametern außerordentlich schwierig und ist demzufolge weitestgehend unerforscht.

Wir beschäftigen uns deshalb in diesem Kapitel nur mit den einfachsten *lokalen Bifurkationen* von Gleichgewichtszuständen und periodischen Bewegungen und zwar für Systeme, die nur von einem Parameter abhängen. Nichtlineare Koordinatentransformationen ermöglichen anschließend eine Transformation auf Normalform und damit eine Zuordnung der auftretenden Bifurkationen zu einigen wenigen Grundtypen.

Neben den lokalen gibt es auch *globale Bifurkationen*, bei denen infolge der Variation der Parameter globale qualitative Änderungen der dynamischen Eigenschaften des Systems auftreten, die sich nicht aus lokalen Informationen herleiten lassen. Spontane Entstehung, Vernichtung oder Änderung von Attraktoren sind typisch für globale qualitative Änderungen der Dynamik und damit für globale Bifurkationen. Eine systematische Behandlung dieser Bifurkationen können wir in diesem Buch nicht vornehmen, sondern werden uns im Rahmen der nächsten Kapitel mit theoretischen und numerischer Untersuchungen auf einige wesentliche Fälle beschränken. Für ein Studium theoretischer Grundlagen und

weiterer Beispiele seien die Bücher von Guckenheimer und Holmes (1983), Chow und Hale (1982) und Wiggins (1990) empfohlen.

Das Bifurkationsverhalten kann von zwei Faktoren abhängen:

1. Wie verändert sich das qualitative Verhalten der Lösung der DGL

$$\dot{x} = f(x), \quad x \in \Omega \subset \mathbb{R}^n, \quad f \in C^1(\Omega, \mathbb{R}^n), \quad (3.1)$$

wenn das Vektorfeld f geändert wird? Bleibt das qualitative Verhalten der Lösung bei „kleinen Störungen“ (Begriff wird noch präzisiert) erhalten, dann nennt man das System (3.1) oder das Vektorfeld f strukturell stabil. Die Idee der strukturellen Stabilität geht auf *Andronov*¹ und *Pontrjagin*² aus dem Jahr 1937 zurück. Für zweidimensionale Systeme sind die strukturell stabilen Vektorfelder durch *Peixoto's Theorem* vollständig charakterisiert. Ein entsprechendes Resultat ist für höherdimensionale Vektorfelder ($n \geq 3$) nicht verfügbar.

2. Eine Änderung des qualitativen Verhaltens der Lösung in DGL-Systemen der Form

$$\dot{x} = f(x, \mu), \quad x \in \Omega \subset \mathbb{R}^n, \quad \mu \in \mathbb{R}^k \quad (3.2)$$

kann auch durch Änderung des Parameters μ hervorgerufen werden. Bei uns ist stets $\mu \in \mathbb{R}$. Wir betrachten f in der Umgebung einer Ruhelage $(x_0, \mu_0) \in \Omega \times \mathbb{R}$, d. h. $f(x_0, \mu_0) = 0$ und untersuchen, ob es weitere Ruhelagen in der Nähe des Wertes μ_0 gibt und von welchem Stabilitätstyp diese sind.

Ein wichtiges Prinzip der Bifurkationstheorie wird sich bei den Beispielen herausstellen:

Der Verlust der Stabilität in einem Punkt führt zur Bifurkation der DGL.

3.1 Strukturelle Stabilität

Im Abschnitt 2.7.3 haben wir Phasenbilder des Systems $\dot{x} = f(x)$ mit ihrer Linearisierung an einem Fixpunkt verglichen und erklärt, wann wir solche Phasenportraits noch als gleichwertig ansehen wollen; welche Verzerrungen also zugelassen sein sollen. Wir nennen ein Vektorfeld f *strukturell stabil*, wenn jedes Vektorfeld g , welches nahe bei f bleibt, topologisch äquivalent zu f ist. Ist also z. B. $g(x) := Df(x_0)x$ und sind die Realteile der Eigenwerte von $Df(x_0)$ ungleich Null, dann gibt es einen orientierungserhaltenden Homöomorphismus H , der die Trajektorien des Vektorfeldes f auf die des Vektorfeldes g abbildet. Es finden also nur stetige Deformationen der Phasenportraits zueinander statt.

¹Aleksandr Aleksandrovic Andronov (1901–1952), verfasste Arbeiten über quantenmechanische Statistik, über nichtlineare Schwingungen und zur Störungstheorie, außerdem zur Bifurkation, Stabilität und Topologie der Lösungen dynamischer Systeme.

²Lew Semjonowitsch Pontrjagin (1908–1988), verfasste Arbeiten über quantenmechanische Statistik, 1937 ein Standardwerk zur Störungstheorie nichtlinearer Systeme und lieferte Beiträge zur Bifurkation, Stabilität und Topologie der Lösungen dynamischer Systeme.

Definition 3.1 (Strukturelle Stabilität)

Ein Vektorfeld $f \in \mathcal{C}^1(\Omega, \mathbb{R}^n)$ mit $\Omega \subset \mathbb{R}^n$ und Ω offen heißt *strukturell stabil*, falls es ein $\varepsilon > 0$ gibt, so dass für alle $g \in \mathcal{C}^1(\Omega, \mathbb{R}^n)$ mit

$$\|f - g\|_{\mathcal{C}^1} < \varepsilon \quad (\|f\|_{\mathcal{C}^1} := \sup_{x \in \Omega} |f(x)| + \sup_{x \in \Omega} \|Df(x)\|)$$

folgt, dass f und g topologisch äquivalent auf Ω sind. *Strukturell instabil* heißt f , falls es nicht strukturell stabil ist. (In $\|\cdot\|_{\mathcal{C}^1}$ bezeichnen $|\cdot|$ die Euklidische Norm im \mathbb{R}^n und $\|\cdot\|$ eine der üblichen Matrixnormen.)

An zwei Beispielen wird die strukturelle Stabilität (und Instabilität) erläutert.

Beispiel 3.2

Wir betrachten das lineare System

$$\dot{x} = Ax, \quad x \in \mathbb{R}^n$$

mit konstanter reeller Koeffizientenmatrix $A \in \mathbb{R}^{n \times n}$. Sind die Realteile der Eigenwerte von A verschieden von Null, dann sind diese Systeme strukturell stabil. ■

Beispiel 3.3

Wir betrachten ein einfaches Zweipopulationenmodell, wobei

$x(t)$ die Population der Beutespezies zur Zeit t und

$y(t)$ die Population der Räuberspezies zur Zeit t

bezeichnen (z. B. Karpfen und Hechte). Für jede Population gilt die Wachstumsgleichung

$$\dot{p} = r(t, p)p.$$

Dabei ist $r = \frac{\dot{p}}{p}$ die totale Änderungsrate (die zeitliche Änderung $\dot{p} = \frac{dp}{dt}$, bezogen auf die Gesamtpopulation p) zur Zeit t . Bei zwei Populationen wird natürlich die Wachstumsrate der einen Population durch die andere beeinflusst, d. h. es gilt

$$\dot{x} = r_1(t, x, y)x, \quad \dot{y} = r_2(t, x, y)y.$$

In unserem Modell nehmen wir an, dass die Räuberspezies sich ausschließlich von den Beutespezies ernähren, während die Beutespezies unbegrenzt Nahrung vorfinden. Ein einfacher Ansatz ist von der Form

$$r_1(t, x, y) = \alpha - \beta y, \quad \alpha, \beta > 0.$$

Er lässt die Interpretation zu: Wenn kein Räuber vorhanden ist ($y = 0$) entwickelt sich die Beutepopulation mit der konstanten Wachstumsrate α . Die Anwesenheit der Räuberspezies verringert diese Wachstumsrate und zwar proportional zur Räuberpopulation. Analog wird der Ansatz

$$r_2(t, x, y) = -\gamma + \delta x, \quad \gamma, \delta > 0$$

gemacht. Unter diesen Annahmen erhalten wir die speziellen *Räuber-Beute-Gleichungen* (auch *Volterra-Lotka Gleichungen* genannt):

$$\dot{x} = (\alpha - \beta y)x, \quad \dot{y} = (\delta x - \gamma)y \quad (\alpha, \beta, \gamma, \delta > 0). \quad (3.3)$$

Zum Zeichnen des Phasenportraits gehen wir wieder den üblichen Weg. Es gibt zwei Ruhepunkte $(x_0, y_0) = (0, 0)$ und $(x_1, y_1) = \left(\frac{\gamma}{\delta}, \frac{\alpha}{\beta}\right)$. Die Linearisierung des Vektorfeldes (3.3) liefert

$$Df(x, y) = \begin{pmatrix} \alpha - \beta y & -\beta x \\ \delta y & \delta x - \gamma \end{pmatrix}. \quad (3.4)$$

Der Fixpunkt $(0, 0)$ ist ein Sattelpunkt (zwei im Vorzeichen unterschiedliche reelle Eigenwerte). Der stabile Unterraum ist die y -Achse, der instabile Unterraum die x -Achse. Mit dem Hartman-Grobman-Theorem (vgl. Satz 2.54) ist somit der Phasenverlauf lokal um $(0, 0)$ klar. Der andere kritische Punkt $\left(\frac{\gamma}{\delta}, \frac{\alpha}{\beta}\right)$ ist nicht hyperbolisch ((3.4) hat in $\left(\frac{\gamma}{\delta}, \frac{\alpha}{\beta}\right)$ den doppelten Eigenwert Null) und damit das Theorem von Hartman und Grobman nicht anwendbar. In diesem konkreten Fall kommt man mit der Phasendifferentialgleichung, also mit dem Berechnen eines ersten Integrals für die DGL (3.3) weiter. Die Orbits von (3.3) liegen dann in den Niveaulinien des ersten Integrals. Die Phasendifferentialgleichung lautet:

$$\frac{dy}{dx} = -\frac{y(\gamma - \delta x)}{x(\alpha - \beta y)} \quad (3.5)$$

Diese DGL ist trennbar:

$$\left(\beta - \frac{\alpha}{y}\right) y' = \frac{\gamma}{x} - \delta \quad (x > 0, y > 0).$$

Ihre Lösung lautet

$$\beta y - \alpha \ln y = \gamma \ln x - \delta x + c, \quad c \in \mathbb{R}.$$

Die Orbits von (3.3) liegen also in den Höhenlinien von

$$F(x, y) := \beta y - \alpha \ln y + \delta x - \gamma \ln x = c. \quad (3.6)$$

Damit kann auch das Phasenportrait in der Umgebung des Fixpunktes $\left(\frac{\gamma}{\delta}, \frac{\alpha}{\beta}\right)$ gezeichnet werden. (Den Beweis, dass die Höhenlinien von (3.6) geschlossen sind, überlassen wir dem Leser.) Die Populationsfluktuation der Räuber- und Beutespezies zeigt Abbildung 3.1 (links) während in Abbildung 3.1 (rechts) tatsächlich zu sehen ist, wie die Population von Räubern und Beute oszilliert.

Aus dem Phasenportrait liest man das Langzeitverhalten der beiden Populationen ab, wobei nur die Lösungskurven für uns von Interesse sind, die im 1. Quadranten liegen, da es keine negativen Populationen gibt. Das Verhalten ist überdies stabil gegen kleine Änderungen in den Anfangswerten $\bar{x}_0 > 0, \bar{y}_0 > 0$.

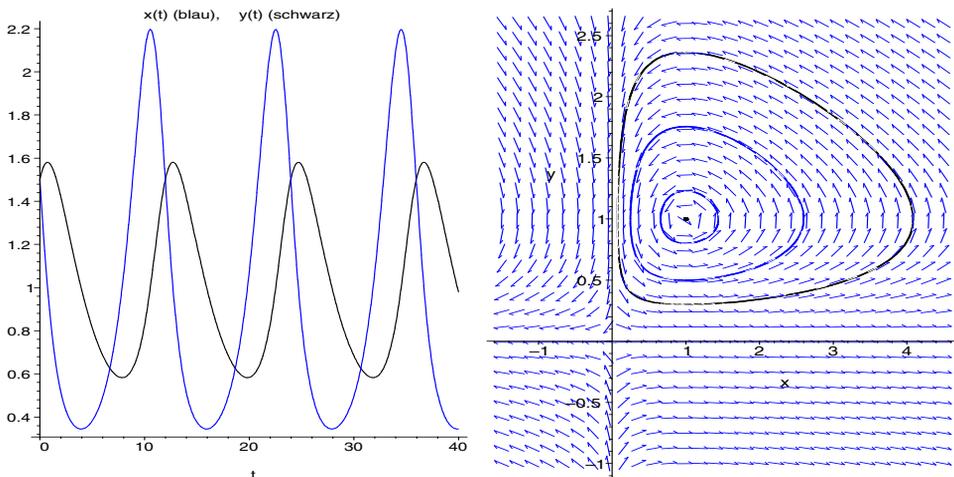


Abb. 3.1 Verlauf von $x(t), y(t)$ (links), Phasenportrait (rechts)

Gilt dagegen $\bar{x}_0 = 0$, ist also am Anfang keine Beutespezies da, so stirbt die Räuberspezies aus ($y(t) \rightarrow 0$ für $t \rightarrow \infty$). Ist $\bar{y}_0 = 0$ (keine Räuberspezies vorhanden), so wächst die Beutepopulation unbegrenzt ($x(t) \rightarrow \infty$ für $t \rightarrow \infty$). Eine kleine Änderung in den Anfangsdaten (z. B. der Übergang von $x_0 = 0$ zu $x_0 > 0$) hat ein total anderes Langzeitverhalten zur Folge.

Stört man das Vektorfeld in Gleichung (3.3), indem man die Nichtlinearität $-(\nu x^2, \mu y^2)^T$ hinzufügt, dann lautet das System

$$\dot{x} = (\alpha - \beta y)x - \nu x^2, \quad \dot{y} = (\delta x - \gamma)y - \mu y^2 \quad (\alpha, \beta, \gamma, \delta > 0). \quad (3.7)$$

Die Linearisierung von (3.7) hat nicht wie die Linearisierung (3.4) einen doppelten Nulleigenwert. Dieser wurde durch die Nichtlinearität $-(\nu x^2, \mu y^2)^T$ „zerstört“. Die Fixpunkte des Systems (3.7) sind wieder hyperbolisch und können mit dem Satz von Hartman und Grobman diskutiert werden. ■

Die hyperbolischen Fixpunkte sind in einem gewissen Sinn stabil, d. h. kleine Änderungen von f können den Charakter des Fixpunktes nicht „zerstören“. Genaue Voraussetzungen macht der folgende

Satz 3.4 (Hirsch (1974))

Es seien $f \in C^1(\Omega, \mathbb{R}^n)$ und $\Omega \subset \mathbb{R}^n$ eine offene Menge, die den hyperbolischen Fixpunkt x_0 der DGL $\dot{x} = f(x)$ enthält. Dann gibt es zu jedem $\varepsilon > 0$ ein $\delta > 0$, so dass für alle $g \in C^1(\Omega; \mathbb{R}^n)$ mit $\|f - g\|_{C^1} < \delta$ ein $y_0 \in K(x_0; \varepsilon)$ existiert, so dass y_0 ein hyperbolischer Fixpunkt von $\dot{x} = g(x)$ ist. Darüber hinaus haben $Df(x_0)$ und $Dg(y_0)$ die gleiche Anzahl von Eigenwerten mit negativem und positivem Realteil.

Eine andere Situation ergibt sich, wenn das DGL-System Orbits enthält, die Sattelpunkte verbinden, sogenannte heterokline Orbits. Dann kann man nicht mehr von strukturell stabilen Vektorfeldern sprechen. Das DGL-System in Beispiel 3.5 besitzt homokline Orbits. Hierbei handelt es sich um Lösungen, die sich in der Vorwärts- und in der Rückwärtszeit gegen ein und denselben Fixpunkt bewegen.

Beispiel 3.5

Wir betrachten das System

$$\dot{x} = y, \quad \dot{y} = \mu y + x - x^3. \quad (3.8)$$

Für $\mu = 0$ liegt ein *Hamiltonsches DGL-System* zugrunde. Darunter versteht man ein DGL-System im \mathbb{R}^2 der Form $\dot{x} = \frac{\partial H}{\partial y}(x, y)$ und $\dot{y} = -\frac{\partial H}{\partial x}(x, y)$, wobei $H : \mathbb{R}^2 \rightarrow \mathbb{R}$ eine C^1 -Funktion ist, die man *Hamiltonfunktion* nennt. Die Hamiltonfunktion für (3.8) lautet: $H(x, y) := \frac{1}{2}(y^2 - x^2) + \frac{x^4}{4}$. Sie hat die Eigenschaft, dass sie entlang jeder Lösungskurve konstant ist. Die Kenntnis einer Hamiltonfunktion hat den Vorteil, dass man das Phasenportrait zeichnen kann, ohne das DGL-System gelöst zu haben. Alle Lösungen liegen in den Niveaulinien $H(x, y) = \text{const}$. Die Niveaulinien für $H(x, y) = c$ sind in Abbildung 3.2 b) gezeichnet. Wir sehen, dass für $\mu = 0$ zwei Fixpunkte $(\pm 1, 0)$ vorliegen, die Zentren sind. Die Linearisierung an den Fixpunkten $(\pm 1, 0)$ besitzt je einen konjugiert komplexen, rein imaginären Eigenwert. Des Weiteren gibt es zwei, zur x -Achse symmetrische homokline Orbits an den Fixpunkt $(0, 0)$, die geschlossene von nicht geschlossenen Orbits trennen. Ihre Vereinigung heißt in diesem Zusammenhang auch Separatrix. Für $\mu = 0$ ist dieses System strukturell instabil auf jeder kompakten Teilmenge $K \subset \mathbb{R}^2$.

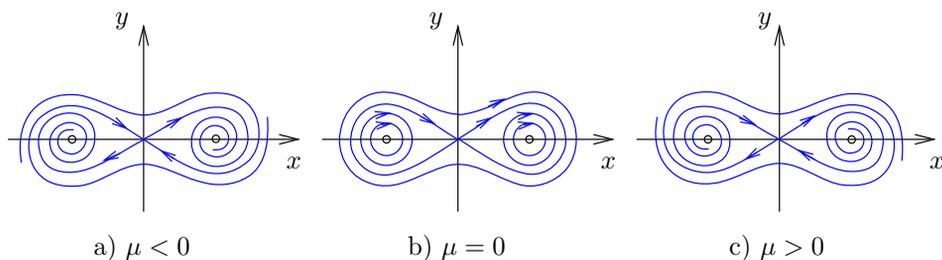


Abb. 3.2 Phasenportrait für die Gleichung (3.8)

Das Phasenportrait wird in Abbildung 3.2 gezeigt und es ist klar, dass das Phasenportrait für $\mu \neq 0$ nicht topologisch äquivalent zu dem mit $\mu = 0$ ist. Es ist sogar so, dass für $\mu \neq 0$ die Separatrix zerstört wird. ■

3.2 Einige typische Bifurkationen

Wir untersuchen drei Haupttypen von Bifurkationen, die auftreten, wenn der Parameter μ variiert wird. Ausgangspunkt unserer Untersuchungen ist die skalare DGL

$$\dot{x} = f(x, \mu), \quad x \in (a, b) \subset \mathbb{R}, \quad \mu \in \mathbb{R}, \quad (3.9)$$

deren rechte Seite $f : \Omega \rightarrow \mathbb{R}$ auf einer offenen Menge $\Omega \subseteq \mathbb{R}^2$ von der Klasse \mathcal{C}^k , $k \geq 1$ ist, und die für einen Parameterwert μ_0 eine stationäre Lösung x_0 (also $f(x_0, \mu_0) = 0$) besitzt. Wir interessieren uns nun für folgende Fragestellungen:

- Unter welchen Bedingungen besitzt die Gleichung (3.9) für Werte des Parameters μ nahe μ_0 weitere stationäre Lösungen in der Nähe von x_0 ?
- Wie viele Ruhelagen gibt es und welchen Stabilitätstyp besitzen sie?

Da die Begriffe Sattel und Knoten bei skalaren DGL keinen Sinn ergeben, wäre die Bezeichnung irreführend. Sie wird jedoch sofort verständlich, wenn man zu der Gleichung (3.9) z. B. die Gleichung $\dot{y} = -y$ hinzufügt und das Ganze als zweidimensionales System betrachtet. Die Bifurkationscharakteristik wird dadurch nicht verändert. Im zweidimensionalen Phasenraum sind dann wieder Sattel- bzw. Knotenpunkte erkennbar. Ein Phasenportrait im (x, μ) -Raum, aus dem insbesondere die Ruhelagen sichtbar sind, nennen wir *Bifurkationsdiagramm* (auch *Verzweigungsdiagramm*).

Zunächst kann man mit Hilfe des Satzes über implizite Funktionen sofort Bedingungen angeben, für die sich die Anzahl der stationären Lösungen in der Nähe von x_0 bei Variation des Parameters μ *nicht* ändern.

Satz 3.6 (Keine Bifurkation von stationären Lösungen)

Für die rechte Seite der DGL (3.9) seien folgende Bedingungen erfüllt: $f : \Omega \rightarrow \mathbb{R}$ sei auf der offenen Menge $\Omega \subseteq \mathbb{R}^2$ von der Klasse \mathcal{C}^k , $k \geq 1$ und es gelte die Beziehung $f_x(x_0, \mu_0) \neq 0$. Dann gibt es eine Umgebung $U \times V$ des Punktes $(x_0, \mu_0) \in \Omega$ und eine \mathcal{C}^k -Funktion $g : V \rightarrow U$ mit $g(\mu_0) = x_0$ und der Eigenschaft

$$f(g(\mu), \mu) = 0 \quad \text{für alle } \mu \in V.$$

Außer den stationären Lösungen $(g(\mu), \mu)$, $\mu \in V$, besitzt die DGL (3.9) keine weiteren stationären Lösungen in $U \times V$.

Die Bedingung $f_x(x_0, \mu_0) \neq 0$ sichert, dass (x_0, μ_0) kein Bifurkationspunkt ist. Umgekehrt ist die Bedingung $f_x(x_0, \mu_0) = 0$ *notwendig*, aber *nicht hinreichend* dafür, dass (x_0, μ_0) ein Bifurkationspunkt ist.

3.2.1 Sattel-Knoten Bifurkation

Wir betrachten das System

$$\dot{x} = \mu - x^2, \quad \dot{y} = -y \quad (\mu \in \mathbb{R}). \quad (3.10)$$

Die Fixpunkte ergeben sich aus $\dot{x} = \dot{y} = 0$. Für (3.10) gibt es (i) keinen, (ii) einen oder (iii) zwei Fixpunkte, die von den Werten μ abhängen. Wir betrachten die drei Fälle separat.

Fall (i): Falls $\mu < 0$ ist, dann gibt es keine reellen Fixpunkte in der Ebene und der Fluss fließt von „rechts nach links“, da $\dot{x} < 0$ ist. Die Sprechweise „der Fluss fließt von rechts nach links“ wird sofort klar, wenn man sich das eindimensionale Phasenportrait (Abbildung 3.3) ansieht. Es entsteht einfach durch Weglassen der zweiten Gleichung in (3.10).

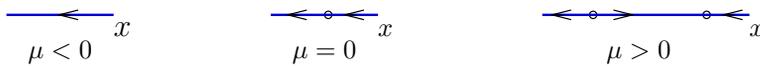


Abb. 3.3 Phasenportrait für $\dot{x} = \mu - x^2$, $x, \mu \in \mathbb{R}$

Fall (ii): Falls $\mu = 0$ ist, dann gibt es genau einen kritischen Punkt im Ursprung. Die Lösungskurven für den Phasenraum ergeben sich aus der Phasendifferentialgleichung

$$\frac{dy}{dx} = \frac{y}{x^2}.$$

Diese separierbare DGL hat die Lösung

$$y(x) = ce^{-1/x}, \quad c \in \mathbb{R}, \quad c = \text{const.}$$

Es ist $\dot{x} < 0$ für alle $x \in \mathbb{R} \setminus \{0\}$. Im eindimensionalen Phasenraum ist der Fluss wieder nach links gerichtet (siehe Abbildung 3.3 für $\mu = 0$).

Fall (iii): Falls $\mu > 0$, dann gibt es zwei reelle Fixpunkte $x_0 = \sqrt{\mu}$ und $x_1 = -\sqrt{\mu}$. Durch die Kurve $\mu - x^2 = 0$ wird die Position der kritischen Punkte der DGL $\dot{x} = \mu - x^2$ bestimmt. Den eindimensionalen Phasenverlauf entnehmen wir wieder Abbildung 3.3 für $\mu > 0$.

Das System (3.10) wird jetzt an den Fixpunkten linearisiert:

$$f_x(\pm\sqrt{\mu}, 0) = \begin{pmatrix} \mp 2\sqrt{\mu} & 0 \\ 0 & -1 \end{pmatrix}.$$

Die Eigenwerte und Eigenvektoren im Fall des Fixpunktes $(\sqrt{\mu}, 0)$ lauten $\lambda_1 = -2\sqrt{\mu}$, $v_1 = (1, 0)^T$ und $\lambda_2 = -1$, $v_2 = (0, 1)^T$. Der Fixpunkt $(\sqrt{\mu}, 0)$ ist ein stabiler Knoten und die stabilen Mannigfaltigkeiten sind orthogonal zueinander. Für den

Fixpunkt $(-\sqrt{\mu}, 0)$ erhalten wir nach entsprechender Betrachtung einen Sattel. Aus dem eindimensionalen Phasenraum (siehe Abbildung 3.3) ergibt sich das zweidimensionale Phasenportrait wie folgt: Man trage in den Fixpunkten der Abbildung 3.3 jeweils eine „stabile y -Richtung“ an (siehe Abbildung 3.4).

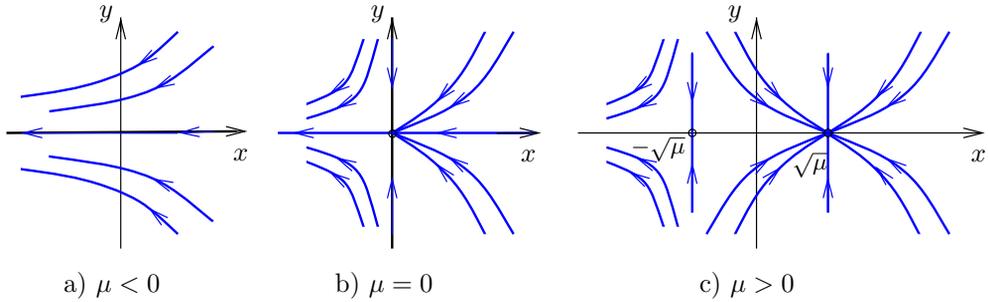


Abb. 3.4 Phasenbilder des Systems $\dot{x} = \mu - x^2, \dot{y} = -y$

Aus den drei Bildern in Abbildung 3.4 kann man auch ein dreidimensionales Phasenportrait zeichnen, indem man die drei Bilder entlang der μ -Achse „übereinanderlegt“ (siehe Abbildung 3.6). Man erkennt deutlich die Bifurkationsparabel $\mu = x^2$ mit dem stabilen und dem instabilen Zweig (verschiedene Punktdarstellungen).

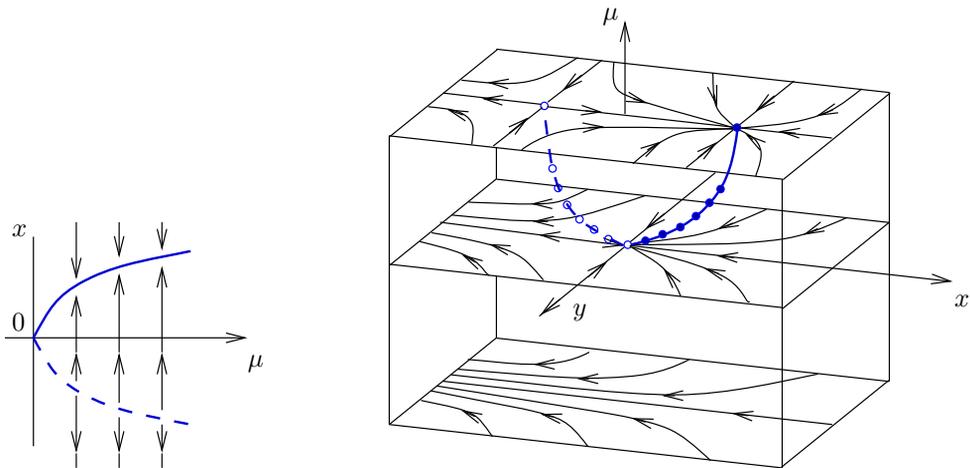


Abb. 3.5 Sattel-Knoten-Bifurkationsdiagramm

Abb. 3.6 Zweidimensionale Sattel-Knoten-Bifurkation: $\dot{x} = \mu - x^2, \dot{y} = -y$

Das Bifurkationsgeschehen kann man im Bifurkationsdiagramm (siehe Abbildung 3.5) kurz und knapp festhalten. Falls $\mu < 0$ ist, dann gibt es keine Fixpunkte und wenn μ durch den Nullpunkt geht, wechselt das qualitative Lösungsverhalten. Es zweigen zwei

Gleichgewichtspunkte vom Ursprung ab, die sich mit wachsendem μ immer weiter voneinander entfernen. Dies ist klar, da die Fixpunkte die Gleichung $\mu = x^2$ erfüllen und das Bifurkationsdiagramm die Form einer Parabel hat. Der durchgezogene Zweig ist der stabile, der gestrichelte Zweig der instabile Fixpunkt. Der Verlust der Stabilität im Ursprung führt also zur Bifurkation der DGL.

Allgemein kann man an einer skalaren parameterabhängigen Differenzialgleichung $\dot{x} = f(x, \mu)$ ihren Sattel-Knoten-Charakter wie folgt erkennen: Zunächst sollte $(0, 0)$ eine nichthyperbolische Ruhelage sein. Dies ist gewährleistet durch die Bedingungen

$$f(0, 0) = 0, \quad f_x(0, 0) = 0.$$

Die Taylorapproximation von $f(x, \mu)$ im Punkt $(0, 0)$ lautet:

$$f(x, \mu) = f_\mu(0, 0)\mu + \frac{1}{2}[f_{xx}(0, 0)x^2 + 2f_{x\mu}(0, 0)x\mu + f_{\mu\mu}(0, 0)\mu^2] + \text{Rest}(x, \mu).$$

Man kann sich nun überlegen, dass das Erfülltsein der Bedingungen

$$f_\mu(0, 0) \neq 0, \quad f_{xx}(0, 0) \neq 0$$

das Auftreten einer Sattel-Knoten-Bifurkation sichert. Kennt man das Vorzeichen von $f_{xx}(0, 0)$, dann kann man Aussagen über die Stabilität der beiden Ruhelagen machen (siehe Aulbach (1997), S. 358 ff.).

Bemerkung 3.7

Bei der skalaren DGL $\dot{x} = f(x)$ kann man aus den Ableitungen der rechten Seite an einem Fixpunkt x_0 auf die Stabilität dieser Ruhelagen schließen:

$$f'(x_0) < 0 \quad \Rightarrow \quad x_0 \text{ asymptotisch stabil}$$

$$f'(x_0) > 0 \quad \Rightarrow \quad x_0 \text{ instabil.}$$

Im Fall $f'(x_0) = 0$ müssen höhere Ableitungen herangezogen werden, wie wir im Fall der Sattel-Knoten-Bifurkation gesehen haben. ■

3.2.2 Transkritische Bifurkation

Wir betrachten das eindimensionale System

$$\dot{x} = \mu x - x^2, \quad \mu \in \mathbb{R} \tag{3.11}$$

und aus den uns bekannten Gründen der besseren Veranschaulichung das zweidimensionale System

$$\dot{x} = \mu x - x^2, \quad \dot{y} = -y \quad (\mu \in \mathbb{R}). \tag{3.12}$$

Die kritischen Punkte von (3.11) sind $x = 0$ und $x = \mu$. Für $\mu = 0$ gibt es nur einen Fixpunkt $x = 0$ und dieser ist nicht hyperbolisch, da $f_x(0, 0) = 0$ gilt. Das Vektorfeld $f(x) = -x^2$ ist strukturell instabil und $\mu = 0$ ist ein Bifurkationspunkt. Für $\mu = 0$ lautet die Zentrumsmannigfaltigkeit $W^c(0) = (-\infty, \infty)$. Das Bifurkationsdiagramm ist in Abbildung 3.7 gezeichnet. Man erkennt, dass am Fixpunkt $x = 0$ ($\mu = 0$), der gleichzeitig Bifurkationspunkt ist, ein Stabilitätswechsel stattfindet. Der Verlust der Stabilität im Ursprung führt wieder zur Bifurkation der DGL (3.11).

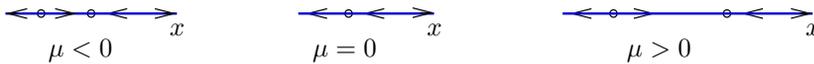


Abb. 3.7 Phasenportrait für die DGL (3.11)

Etwas anschaulicher wird unser Phasenportrait, wenn wir die Gleichung $\dot{y} = -y$ hinzunehmen. Man denke sich also wieder eine stabile Mannigfaltigkeit in den Fixpunkten parallel zur y -Achse angetragen (siehe Abbildung 3.8). Falls $\mu < 0$, dann ist der Ursprung ein stabiler Knoten und $x = \mu$ ist ein Sattel. Falls $\mu = 0$, dann ist $x = 0$ ein nichthyperbolischer Fixpunkt. Die Lösungskurven genügen der Phasendifferenzialgleichung

$$\frac{dy}{dx} = \frac{y}{x^2}, \quad x \neq 0,$$

die die Lösung $y(x) = c e^{-1/x}$ mit $c = \text{const}$ besitzt. Im Fall $\mu > 0$ ist der Ursprung ein Sattel und $x = \mu$ ist ein stabiler Knoten (siehe Abbildung 3.8). Das Stabilitätsverhalten

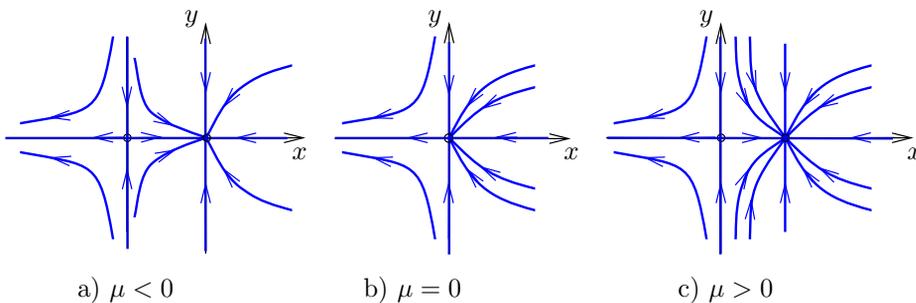


Abb. 3.8 Phasenportrait der DGL (3.12) bei verschiedenen Parameterwerten

hat also im Ursprung wieder gewechselt. Das Bifurkationsdiagramm ist in Abbildung 3.9 gezeichnet. Durchgezogene und gestrichelte Linien geben wieder das Stabilitäts- bzw. Instabilitätsverhalten der Fixpunkte an.

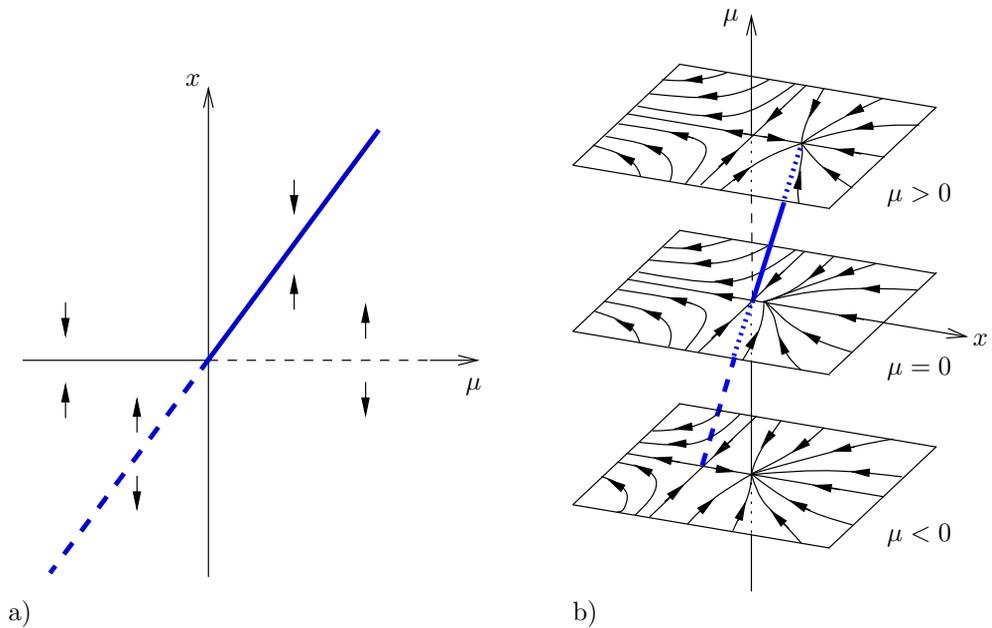


Abb. 3.9 Transkritische Bifurkation: $\dot{x} = \mu x - x^2$, $\dot{y} = -y$.
 a) eindimensional, b) zweidimensional

3.2.3 Die Pitchfork-Bifurkation

Wir betrachten sowohl das eindimensionale System

$$\dot{x} = \mu x - x^3, \quad \mu \in \mathbb{R} \quad (3.13)$$

als auch das zweidimensionale System

$$\dot{x} = \mu x - x^3, \quad \dot{y} = -y \quad (\mu \in \mathbb{R}). \quad (3.14)$$

Eine einfache Rechnung zeigt, dass sowohl (3.13) als auch (3.14) entweder einen Fixpunkt oder drei Fixpunkte besitzt. Für $\mu \leq 0$ ist $x = 0$ einziger Fixpunkt, für $\mu > 0$ gibt es drei Fixpunkte: $x = 0$ und $x = \pm\sqrt{\mu}$. Für $\mu = 0$ ist $x = 0$ nichthyperbolischer Fixpunkt, da $f_x(0, 0) = 0$ ist. Das Vektorfeld $f(x) = -x^3$ ist strukturell instabil. Kleinste Störungen können das Phasenportrait verändern (siehe Abbildung 3.10). Folglich ist $\mu = 0$ ein Bifurkationspunkt.

Das Phasenportrait für die Gleichung (3.13) ist in Abbildung 3.11 aufgezeichnet. Das Bifurkationsdiagramm ist in Abbildung 3.12 a) zu sehen und der Typ der Bifurkation ist treffend Pitchfork-Bifurkation (Heugabel-Bifurkation) genannt.

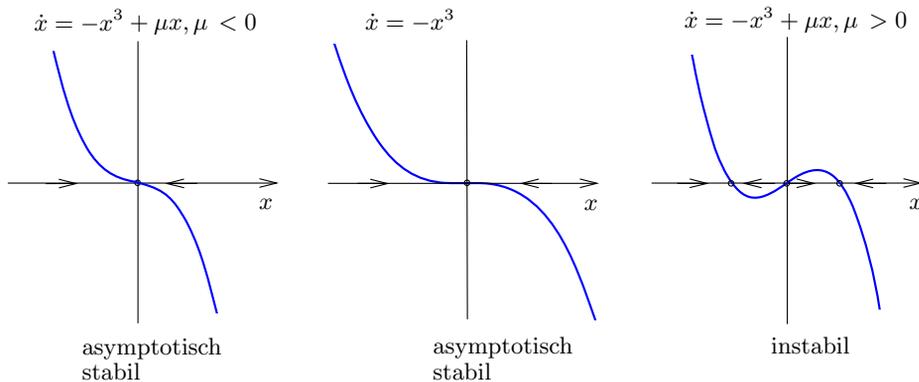


Abb. 3.10 Zur strukturellen Instabilität der trivialen Ruhelage bei skalaren DGL vom Typ $\dot{x} = -x^3$ unter Störungen



Abb. 3.11 Phasenportrait für die DGL (3.13)

Die Abbildung 3.12 b) zeigt das Phasenportrait von Gleichung (3.14). Für $\mu = 0$ erhalten wir aus

$$\frac{dy}{dx} = \frac{y}{x^3}$$

die Lösungen $y(x) = c e^{-1/(2x^2)}$. Der Fixpunkt $x = 0$ ist für $\mu \leq 0$ ein stabiler Knoten. Für $\mu > 0$ haben wir die drei Fixpunkte $(0, 0)$ und $(\pm\sqrt{\mu}, 0)$. Der Ursprung ist jetzt ein Sattelpunkt und die Fixpunkte $(\pm\sqrt{\mu}, 0)$ sind stabile Knoten.

Die Sattel-Knoten-, Transkritische- und Pitchfork-Bifurkation sind sicherlich die häufigsten Typen von Bifurkationen, die in eindimensionalen Systemen auftreten können. Hier ist nur für gewisse Prototypen das Bifurkationsverhalten erklärt worden. Diese Situationen lassen sich auch verallgemeinern und durch Bedingungen an die Ableitungen des Vektorfeldes $f(x, \mu)$ charakterisieren, so wie es im Fall der Sattel-Knoten-Bifurkation beschrieben ist. Beispielsweise tritt die transkritische Bifurkation mit ihrem typischen Bifurkationsdiagramm (siehe Abbildung 3.9) stets dann ein, wenn folgende Bedingungen an f erfüllt sind: $f(0, \mu) = 0$ für alle $\mu \in (-\delta, \delta)$, $\delta > 0$ und $f_x(0, 0) = 0$. Hieran erkennt man sofort, dass es sich um ein anderes Verzweigungsproblem handelt als bei der Sattel-Knoten-Bifurkation. Dort wurde $f_\mu(0, 0) \neq 0$ vorausgesetzt ($f(0, \mu) = 0$ für alle μ verhindert dies). Des Weiteren wird für die transkritische Bifurkation $f_{x\mu}(0, 0) \neq 0$ und $f_{xx}(0, 0) \neq 0$ benötigt. Unter diesen Voraussetzungen kann bis auf Homöomorphie stets das Bifurkationsdiagramm aus Abbildung 3.9 nachgewiesen werden. Die stabilen und instabilen Zweige sind lediglich eine Frage der Vorzeichen der Ableitungen. Auch für die Heugabel-Bifurkation ist eine solche Betrachtungsweise möglich. Es sei hier noch einmal auf Aulbach (1997) verwiesen, wo diese verschiedenen Bifurkationsphänomene theoretisch vollständig ausgearbeitet sind.

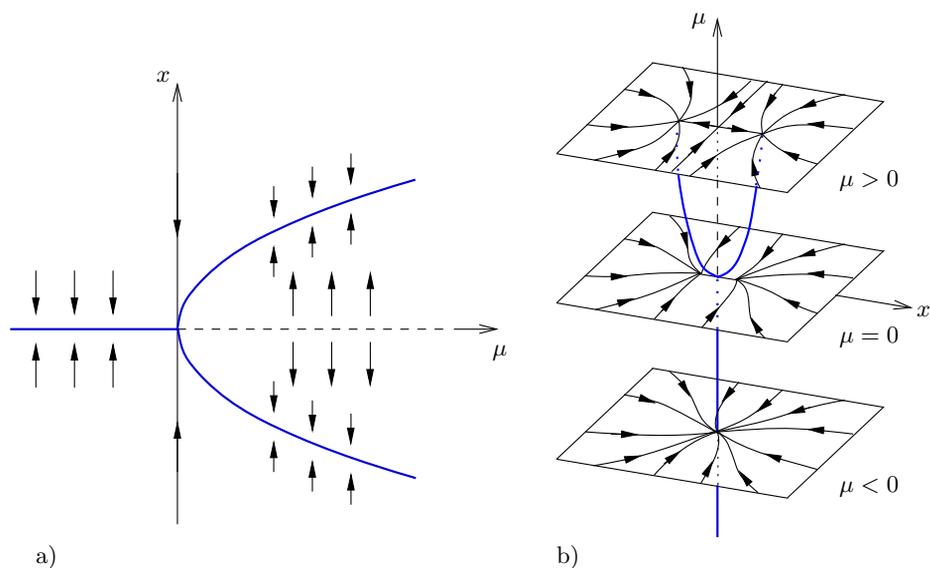


Abb. 3.12 a) Superkritische Heugabel-Bifurkation $\dot{x} = \mu x - x^3$,
 b) Zweidimensionale Heugabel-Bifurkation $\dot{x} = \mu x - x^3$, $\dot{y} = -y$

In eindimensionalen Differentialgleichungen können auch noch andere Bifurkationstypen als die drei beschriebenen auftreten. Falls $f_x(0,0) = \dots = f_x^{(m-1)}(0,0) = 0$ und $f_x^{(m)}(0,0) \neq 0$ gilt, dann spricht man von einem Fixpunkt der Vielfachheit m in $x = 0$. In diesem Fall können höchstens m Fixpunkte zur Bifurkation des Ursprungs Anlass geben. In den Gleichungen (3.10) und (3.11) ist im Bifurkationspunkt $\mu = 0$ der Ursprung $x = 0$ ein Fixpunkt der Vielfachheit zwei, in (3.13) von der Vielfachheit drei.

Wir haben in diesem Abschnitt skalare DGL der Form $\dot{x} = f(x, \mu)$, $\mu \in \mathbb{R}$, stets auch in Verbindung mit einer zweiten Gleichung der Form $\dot{y} = -y$ betrachtet. Dadurch waren u. a. die Charakterisierungen der Fixpunkte als Sattel, Knoten usw. verständlich. Wir haben damit auch gleichzeitig einfachste Bifurkationsphänomene von ebenen Systemen kennengelernt. All diesen Systemen ist gemeinsam, dass am Bifurkationspunkt $\mu = \mu_0$ die Linearisierung $f_x(x_0, \mu_0)$ einen einfachen Nulleigenwert besitzt. Im nächsten Abschnitt betrachten wir einen Typ von Bifurkation der auftritt, wenn die Linearisierung von f ein einfaches Paar rein imaginärer Eigenwerte hat.

3.2.4 Die Hopf-Bifurkation

Allen bisher erwähnten Bifurkationsphänomenen ist gemeinsam, dass die Eigenwerte der Jacobi-Matrix im Gleichgewichtszustand reell sind. Der Fall konjugiert komplexer Eigenwerte ist bisher nicht aufgetreten, obwohl er der Interessanteste ist. Er führt zu einem Klassiker unter den Verzweigungen – der *Hopf-Bifurkation*. Benannt nach dem Mathema-

tiker *Eberhard Hopf*³, der diesen Verzweigungstyp 1942 für Systeme im \mathbb{R}^n studierte. Die Fixpunkte sind alle nicht hyperbolisch. Insofern handelt es sich um strukturell instabile Systeme von DGL, die in Abhängigkeit vom Parameter μ ihr qualitatives Lösungsverhalten ändern. Der Bifurkationstyp wird auch noch einmal im Abschnitt 4.3 behandelt. Wir betrachten ein ebenes System

$$\dot{x} = f(x, \mu), \quad x \in \Omega \subset \mathbb{R}^2, \quad \mu \in \mathbb{R}, \quad (3.15)$$

wobei die Matrix $f_x(x_0, \mu_0)$ ein Paar rein imaginärer Eigenwerte besitzt. Mit dieser Forderung grenzen wir uns gegen die anderen drei Grundtypen von Bifurkationen aus dem vorigen Abschnitt ab. Es ist zu vermuten, dass ein neues Bifurkationsphänomen entsteht. Zunächst untersuchen wir die Fixpunkte von (3.15). Es möge (x_0, μ_0) ein Fixpunkt von (3.15) sein, d. h. es gilt $f(x_0, \mu_0) = 0$. Da die Linearisierung $f_x(x_0, \mu_0)$ nur ein paar rein imaginärer Eigenwerte besitzt, garantiert der Satz über implizite Funktionen, dass es für jedes μ nahe μ_0 einen einzigen Fixpunkt x_μ nahe x_0 gibt. Falls die Eigenwerte von $f_x(x_\mu, \mu)$ die imaginäre Achse für $\mu = \mu_0$ überqueren, dann werden sich die Dimensionen von stabiler und instabiler Mannigfaltigkeit von x_μ ändern; ebenfalls die lokalen Phasenportraits von (3.15).

Um die nachfolgenden Betrachtungen zu erleichtern nehmen wir einige Vereinfachungen vor, die sich in konkreten Fällen leicht bewerkstelligen lassen. Wir setzen voraus, dass die Schar der Ruhelagen x_μ nur aus trivialen Ruhelagen besteht, d. h. dass der Stabilitätswechsel beim Parameterwert Null stattfindet. Die parameterabhängige Koeffizientenmatrix der Linearisierung möge bereits in reeller Jordanscher Normalform vorliegen. Dann kann man das folgende ebene autonome System der Form

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \alpha(\mu) & -\beta(\mu) \\ \beta(\mu) & \alpha(\mu) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} f(x, y, \mu) \\ g(x, y, \mu) \end{pmatrix} \quad (3.16)$$

mit

$$\alpha(0) = 0, \quad \beta(0) \neq 0, \quad f, g \in \mathcal{C}^n, \quad n \geq 4, \quad (3.17)$$

f und g enthalten keine linearen Terme in x und y , betrachten. Letzteres bedeutet:

$$\begin{aligned} f(0, 0, \mu) = f_x(0, 0, \mu) = f_y(0, 0, \mu) = 0, \\ g(0, 0, \mu) = g_x(0, 0, \mu) = g_y(0, 0, \mu) = 0, \end{aligned} \quad (\mu \in (-\varepsilon, \varepsilon), \varepsilon > 0). \quad (3.18)$$

³Eberhard Hopf (1902–1983), österreichisch-amerikanischer Mathematiker, arbeitete auf dem Gebiet der elliptischen DGL, der Hydro- und Aerodynamik und der Ergodentheorie.

Das Überschreiten der imaginären Achse durch das Eigenwertpaar $\alpha(\mu) \pm i\beta(\mu)$ setzen wir als „transversal“ (d. h. die imaginäre Achse wird tatsächlich überquert) voraus. Die entsprechende *Transversalitätsbedingung* lautet

$$\alpha'(0) \neq 0. \quad (3.19)$$

Die Bedingungen, die an die Ableitungen von f gestellt werden müssen, sind im vorliegenden Fall von komplizierterer Natur. Der Ausdruck

$$\frac{A_2}{\beta(0)} + A_3 \neq 0 \quad \text{mit} \quad (3.20)$$

$$A_2 := f_{02}g_{02} - f_{20}g_{20} + f_{11}(f_{20} + f_{02}) - g_{11}(g_{20} + g_{02}),$$

$$A_3 := f_{30} + f_{12} + g_{21} + g_{03},$$

$$f_{ij} := \frac{\partial^{i+j} f(0, 0, 0)}{(\partial x)^i (\partial y)^j}, \quad g_{ij} := \frac{\partial^{i+j} g(0, 0, 0)}{(\partial x)^i (\partial y)^j}$$

sichert letztlich die Hopf-Bifurkation. Wir wollen unsere Ausführungen nun präzisieren.

Satz 3.8 (Hopf-Bifurkation, E. Hopf 1942)

Das System (3.16) habe für alle μ aus einem Intervall $(-\varepsilon, \varepsilon)$ die triviale Lösung und bei $\mu = 0$ gelten die Voraussetzungen (3.17), (3.18) und (3.19). Gilt darüber hinaus die Ungleichung (3.20), dann gibt es eine Nullumgebung $U(0) \subset \mathbb{R}^2$ und ein $\varepsilon_0 > 0$ mit folgenden Eigenschaften:

- i. Das System (3.16) besitzt in $U \times [-\varepsilon_0, \varepsilon_0]$ keine Ruhelagen außer den trivialen $(0, 0, \mu)$, $\mu \in [-\varepsilon_0, \varepsilon_0]$.
- ii. Für $\alpha'(0) > 0$ sind die trivialen Ruhelagen für $\mu \in [-\varepsilon_0, 0)$ asymptotisch stabil und für $\mu \in (0, \varepsilon_0]$ instabil. Im Fall $\alpha'(0) < 0$ gelten umgekehrte Aussagen.
- iii. Für $\mu = 0$ ist die triviale Lösung asymptotisch stabil bzw. instabil, je nach dem ob $\frac{A_2}{\beta(0)} + A_3$ negativ oder positiv ist.
- iv. Für $[\frac{A_2}{\beta(0)} + A_3]\alpha'(0) > 0$ gibt es für jedes $\mu \in [0, \varepsilon_0]$ keine geschlossene Trajektorie in U , während es für jedes $\mu \in [-\varepsilon_0, 0)$ genau eine geschlossene Trajektorie in U gibt. Diese ist anziehend für $\alpha'(0) < 0$ und abstoßend für $\alpha'(0) > 0$. Ein entsprechendes Resultat gilt für den Fall $[\frac{A_2}{\beta(0)} + A_3]\alpha'(0) < 0$.

Beweis: Die Idee besteht darin, dass System (3.16) in Polarkoordinaten zu transformieren, was auf eine Gleichung der Form

$$\dot{r} = \alpha(\mu)r + F(r, \varphi, \mu), \quad \dot{\varphi} = \beta(\mu) + G(r, \varphi, \mu)$$

führt. Die Trajektorien dieses Systems lassen sich dann mit Hilfe der Phasendifferentialgleichung

$$\frac{dr}{d\varphi} = \frac{\alpha(\mu)r + F(r, \varphi, \mu)}{\beta(\mu) + G(r, \varphi, \mu)} \quad (3.21)$$

als Lösungskurven dieser skalaren Differenzialgleichung schreiben. Alle Aussagen des Satzes folgen dann aus der Gleichung (3.21). Man beachte, dass es sich bei diesem Satz um zweigeteilte Voraussetzungen handelt. Der eine Teil betrifft den Linearteil des Systems mit den speziellen Annahmen über das Eigenwertpaar $\alpha(\mu) \pm i\beta(\mu)$, der andere Teil bezieht sich auf die Nichtlinearitäten f und g , genauer auf bestimmte Terme zweiter und dritter Ordnung der Taylor-Entwicklung. Für einen ausführlichen Beweis konsultiere man Aulbach (1997), Satz 7.11.1. \square

An einem Prototyp wollen wir das Hopfsche Resultat näher erläutern.

Beispiel 3.9

Wir betrachten das planare System

$$\begin{aligned} \dot{x} &= -y + x(\mu - x^2 - y^2) \\ \dot{y} &= x + y(\mu - x^2 - y^2). \end{aligned} \quad (3.22)$$

Der Leser überlege sich, dass (3.22) als einzigen Fixpunkt $(x(\mu), y(\mu)) = (0, 0)$ für alle $\mu \in \mathbb{R}$ besitzt. Weiter gilt für die Linearisierung am Fixpunkt

$$D_{xy}f((0, 0), \mu) = \begin{pmatrix} \mu & -1 \\ 1 & \mu \end{pmatrix}. \quad (3.23)$$

Die Eigenwerte von (3.23) lauten: $\lambda_{1,2}(\mu) := \mu \pm i$. Für $\mu < 0$ ist der Gleichgewichtszustand ein *stabiler Fokus* ($\operatorname{Re} \lambda_{1,2}(\mu) < 0$), und für $\mu > 0$ ein *instabiler Fokus* ($\operatorname{Re} \lambda_{1,2}(\mu) > 0$). Man nennt einen *Fokus* auch noch *Spiralpunkt*. Er ist dadurch gekennzeichnet, dass sich die Phasenkurven des DGL-Systems $\dot{x} = Ax$, $A \in \mathbb{R}^{2 \times 2}$ mit wachsender Zeit in den Fixpunkt „spiralen“ (stabiler Fokus). Ein instabiler Fokus „spiralt“ sich mit $t \rightarrow -\infty$ in den Fixpunkt. Für $\mu = 0$ besitzt (3.23) ein Paar rein imaginärer Eigenwerte und man kann zeigen (siehe Perko (1996), S. 142, Theorem 5 und S. 188), dass der Ursprung entweder ein Zentrum oder ein Fokus des Systems (3.22) ist. Für den Entwurf des Phasenportraits ist es geeigneter, das System (3.22) in Polarkoordinaten zu betrachten. Das ist u. a. auch ein Grund, warum in (3.22) die Nichtlinearität genau diese Struktur hat. Führt man die Transformation

$$x(t) = r(t) \cos \varphi(t), \quad y(t) = r(t) \sin \varphi(t)$$

durch, dann ergibt sich das System

$$\dot{r} = r(\mu - r^2), \quad \dot{\varphi} = 1. \quad (3.24)$$

Wir sehen, dass für $\mu = 0$ der Ursprung ein stabiler Fokus ist und für $\mu > 0$ ein stabiler Grenzyklus

$$\gamma_\mu(t) = \sqrt{\mu} (\cos t, \sin t)^T, \quad t \in [0, 2\pi]$$

vorliegt (siehe Abbildung 3.13). Ein *Grenzzzyklus* (engl.: *limit cycle*) ist eine geschlossene Lösungskurve im Phasenraum (bzw. Phasendiagramm) gegen die entweder alle Orbits, die in unterschiedlichen Startpunkten beginnen, konvergieren oder auf der Trajektorie beginnen und dort verbleiben. Entsprechendes gilt für wachsende negative Zeiten.

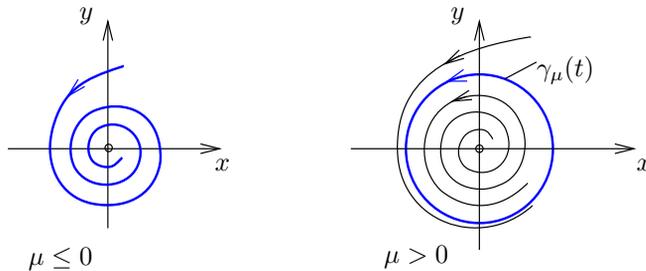


Abb. 3.13 Das Phasenportrait für das System (3.22)

Die Kurve $\gamma_\mu(t)$ repräsentiert eine Familie von Grenzzyklen (Familienparameter μ) dieses Systems. Das Bifurkationsdiagramm zeigt Abbildung 3.14. Die obere Kurve stellt eine einparametrische Familie von Grenzzyklen γ_μ dar, die eine Oberfläche im $\mathbb{R}^2 \times \mathbb{R}$ definieren. Die Bifurkation des Grenzzyklos vom Ursprung tritt im Bifurkationswert $\mu = 0$ auf. Im Bifurkationspunkt $\mu = 0$ tritt wieder ein Stabilitätswechsel ein, welches unser früheres „Bifurkationsprinzip“ bestätigt. ■

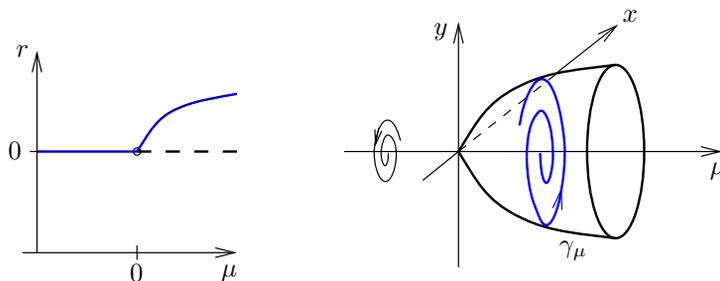


Abb. 3.14 Das Bifurkationsdiagramm von (3.24) und die einparametrische Familie von Grenzzyklen γ_μ , die aus der Hopf-Bifurkation von (3.22) resultiert

Dieses Bifurkationsszenario wird *Hopf-Bifurkation* genannt. Man unterscheidet zwei Fälle von Hopf-Bifurkationen:

- i. Die *superkritische Hopf-Bifurkation* geht im Bifurkationspunkt $\mu = 0$ von einem stabilen Fokus in einen stabilen Grenzzzyklus über, der Ursprung ist ein instabiler Fokus.
- ii. Die *subkritische Hopf-Bifurkation* geht am Bifurkationspunkt $\mu = 0$ von einem instabilen Fokus in einen instabilen Grenzzzyklus über; der Ursprung ist ein stabiler Fokus (siehe Abbildung 3.15).

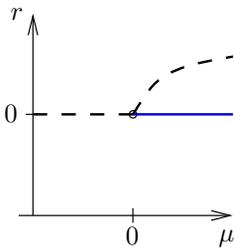


Abb. 3.15 Bifurkationsdiagramm für eine subkritische Hopf-Bifurkation

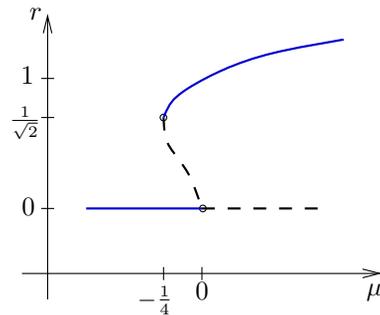


Abb. 3.16 Bifurkationsdiagramm für das System (3.25)

Beispiel 3.10

Abhängig vom Vektorfeld in (3.16) gibt es auch Systeme mit zwei Grenzzyklen. Betrachten wir das System

$$\dot{r} = r(\mu + r^2 - r^4), \quad \dot{\varphi} = 1. \quad (3.25)$$

Der Ursprung ist Fixpunkt für jedes μ . Das System (3.25) durchläuft in $\mu = 0$ eine superkritische Hopf-Bifurkation, d. h. der Fixpunkt $x = y = 0$ ist stabil für $\mu < 0$ und instabil für $\mu > 0$. Dann gibt es noch zwei weitere Fixpunkte:

$$r_1(\mu) = \sqrt{\frac{1}{2} + \sqrt{\frac{1}{4} + \mu}}, \quad \mu \geq -\frac{1}{4} \quad \text{und} \quad r_2(\mu) = \sqrt{\frac{1}{2} - \sqrt{\frac{1}{4} + \mu}}, \quad \mu \in \left[-\frac{1}{4}, 0\right].$$

Das Hopf-Bifurkationsdiagramm ist in Abbildung 3.16 gezeichnet. Daraus lesen wir das folgende Bifurkationsszenario ab: Für $\mu \geq 0$ ist die triviale Lösung instabil, und es gibt einen stabilen Grenzzyklus mit „großer“ Amplitude. (Mit dem Begriff „groß“ wollen wir uns gegenüber der klassischen Hopf-Bifurkation in Abbildung 3.14 abgrenzen, wo die Amplitude von Null auf nur kleine Werte wächst.) Im Parameterwert $\mu = 0$ wird ein weiterer Grenzzyklus, beginnend mit Amplitude Null geboren. Der Stabilitätswechsel erfolgt in $\mu = 0$, und zwar so, dass jetzt die triviale Lösung stabil ist, der Grenzzyklus mit der kleineren Amplitude instabil und der äußere Grenzzyklus wieder stabil ist. Mit fallendem μ , $\mu \in (-\frac{1}{4}, 0)$ wächst die Amplitude des inneren Grenzzyklus an, bis beide Grenzzyklen im Parameterwert $\mu = -\frac{1}{4}$ zu einem Grenzzyklus verschmelzen und der für $\mu > -\frac{1}{4}$ verschwindet. Dabei bleibt für $\mu < 0$ die triviale Lösung stets asymptotisch stabil. In Abbildung 3.17 wird noch einmal im (μ, x, y) -Diagramm (und auch im (x, y) -Diagramm) das Bifurkationsszenario dargestellt. ■

In einem weiteren Beispiel soll u. a. das Bestehen eines Grenzzyklus für alle $\mu \in \mathbb{R}$ gezeigt werden. Außerdem gibt es einen weiteren Grenzzyklus für alle $\mu > -1$.

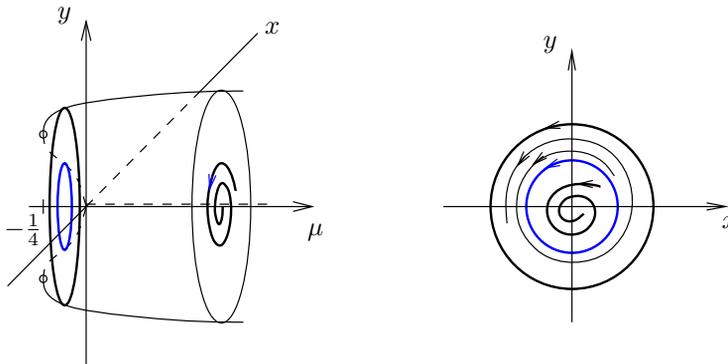


Abb. 3.17 Anschauliche Darstellung der Hopf-Bifurkation für (3.25)

Beispiel 3.11

Wir betrachten das System

$$\begin{aligned}\dot{x} &= -y - x(x^2 + y^2 - 1)(x^2 + y^2 - 1 - \mu), \\ \dot{y} &= x - y(x^2 + y^2 - 1)(x^2 + y^2 - 1 - \mu)\end{aligned}$$

oder in Polarkoordinaten

$$\dot{r} = -r(r^2 - 1)(r^2 - 1 - \mu), \quad \dot{\varphi} = 1. \quad (3.26)$$

Aus (3.26) liest man unmittelbar die beiden Grenzzyklen ab: $r = 1$ für alle $\mu \in \mathbb{R}$ und $r = \sqrt{1 + \mu}$ für alle $\mu > -1$. In den (x, y) -Koordinaten wird der zweite Grenzzyklus durch

$$(x_0(t), y_0(t)) = (\sqrt{1 + \mu} \cos(t), \sqrt{1 + \mu} \sin(t)) \quad (3.27)$$

repräsentiert. Das Bifurkationsdiagramm ist in Abbildung 3.18 dargestellt. Es zeigt in der (μ, r) -Ebene die sogenannte transkritische Bifurkation, die im Bifurkationspunkt $\mu = 0$ ($r = 1$) auftritt. Die gestrichelten (durchgezogenen) Linien geben die Instabilität (Stabilität) der beiden Grenzzyklen an. Interessant ist, dass bei dieser Hopf-Bifurkation als mögliches Bifurkationsszenario die transkritische Bifurkation vorliegt. ■

3.2.5 Zusammenfassung der Bifurkationstypen

In skalaren einparametrischen Differenzialgleichungen der Form $\dot{x} = f(x, \mu)$ können in der Nähe von Gleichgewichtszuständen verschiedene Typen von Bifurkationen auftreten. Den ersten drei in Abschnitt 3.2 genannten Typen ist eine verschwindende erste Ableitung (bzgl. x) der rechten Seite $f(x, \mu)$ am kritischen Punkt gemeinsam. Die transkritische, Sattel-Knoten- und Heugabel-Bifurkation kann jeweils durch eine Normalform charakterisiert werden. Erst durch Voraussetzungen an die Terme höherer Ordnung stellen sich

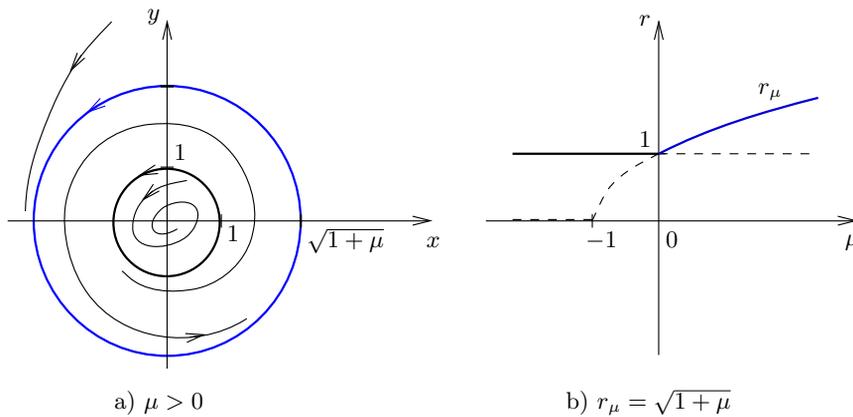


Abb. 3.18 a) Phasenkurven und Grenzzyklen, b) Bifurkationsdiagramm

die entsprechenden Bifurkationszustände ein. Das Verhalten von Lage und Stabilität der Gleichgewichtszustände in der Nähe des Bifurkationspunktes kann kurz wie folgt charakterisiert werden.

Sattel-Knoten Bifurkation

Ein stabiler und ein instabiler Gleichgewichtszustand bewegen sich aufeinander zu, verschmelzen und verschwinden bzw. entstehen im Bifurkationspunkt neu und entfernen sich wieder voneinander.

Transkritische Bifurkation

Ein stabiler und ein instabiler Gleichgewichtszustand bewegen sich aufeinander zu und im Bifurkationspunkt erfolgt ein Stabilitätsaustausch; anschließend entfernen sich die Gleichgewichtszustände voneinander.

Pitchfork-Bifurkation

Ein stabiler Gleichgewichtszustand behält seine Lage, verändert im Bifurkationspunkt jedoch sein Stabilitätsverhalten; zusätzlich entstehen bzw. verschwinden zwei Gleichgewichtszustände derselben Stabilität.

Hopf-Bifurkation:

Bei den vorangegangenen Bifurkationstypen kam der Stabilitätswechsel dadurch zustande, dass der parameterabhängige Koeffizient des Linearteils einer skalaren DGL das Vorzeichen wechselte. Hier betrachten wir ebene autonome Systeme mit der Eigenschaft, dass die Koeffizientenmatrix der Linearisierung ein Paar konjugiert komplexer Eigenwerte besitzt, das bei der Änderung des Parameters die imaginäre Achse überquert. Beim Durchgang der Eigenwerte durch die imaginäre Achse wird ein Stabilitätswechsel im Ursprung vollzogen. Der Fixpunkt ist ein Fokus und beim Übergang vom stabilen zum instabilen Fokus wird ein periodischer Orbit (Grenzzyklus) „geboren“. Dadurch entsteht ein komplett neues Bifurkationsszenario.

Weiterhin existieren abgewandelte Bifurkationen, die aus einer Kombination der beschriebenen Arten bestehen. Die genaue Kenntnis dieser Bifurkationsphänomene ist z. B. für die Synthese und den Entwurfsprozess von Schaltungen bei vorgegebenem Bifurkationsverhalten wichtig. Die Struktursynthese führt über die Simulation des dynamischen Verhaltens zum Strukturbild und dem Schaltungsaufbau für diese Bifurkationsarten. Einen ersten schönen Ansatz zu diesen möglichen Schaltungsaufbauten findet man in der Dissertation von Mohr (2001).

3.3 Aufgaben

Aufgabe 3.1

Beweisen Sie, dass in Beispiel 3.3 die Höhenlinien von $F(x, y) = c$ (siehe Gleichung (3.6)) geschlossen sind.

Aufgabe 3.2

Gegeben sei das konkrete Räuber-Beute-Modell (siehe Beispiel 3.3) der Form

$$\dot{x} = (4 - y)x, \quad \dot{y} = (x - 3)y.$$

Untersuchen Sie die Fixpunkte auf Stabilität, geben Sie eine Lösung an und skizzieren Sie das Phasenportrait.

Aufgabe 3.3

Skizzieren Sie für das System (3.7) das Phasenportrait in der Nähe der Fixpunkte. Kann zur Untersuchung des Systems (3.7) der Satz von Hartman und Grobman (Satz 2.54) heran gezogen werden?

Aufgabe 3.4

Für das System (3.8) in Beispiel 3.5 lautet für $\mu = 0$ die Hamiltonfunktion $H(x, y) := \frac{1}{2}(y^2 - x^2) + \frac{x^4}{4}$. Für welche $c \in \mathbb{R}$ sind die Höhenlinien $\{(x, y) \in \mathbb{R}^2 \mid H(x, y) = c\}$ geschlossene Orbits.

Aufgabe 3.5

Skizzieren Sie die Phasenportraits der zu den skalaren DGL

$$\ddot{u} = \mu \dot{u} + \nu u, \quad \mu, \nu \in \mathbb{R}$$

gehörigen ebenen autonomen Systeme.

Aufgabe 3.6

Skizzieren Sie das Phasenportrait des ebenen autonomen Systems

$$\dot{x} = xy - x^3, \quad \dot{y} = 0.$$

Interpretieren Sie das Ergebnis im Hinblick auf die parameterabhängige Gleichung

$$\dot{x} = \mu x - x^3, \quad \mu \in \mathbb{R}.$$

Aufgabe 3.7

Erstellen Sie das von einem reellen Parameter μ abhängige Phasenportrait des zu der Differenzialgleichung 2. Ordnung

$$\ddot{x} = \mu x - x^3$$

gehörigen Systems 1. Ordnung.

Aufgabe 3.8

Bestimmen Sie für die von dem reellen Parameter μ abhängige skalare DGL

$$\dot{x} = -x + \mu x^3$$

alle nichttrivialen Ruhelagen. In Abhängigkeit von dem Parameter $\mu \in \mathbb{R}$ bestimme man die eindimensionalen Phasenportraits und zeichne diese in ein x, μ -Koordinatensystem. Kennzeichnen Sie die Einzugsgebiete in Bezug auf die jeweilige Ruhelage (Fixpunkt).

Aufgabe 3.9

Gegeben sei eine parameterabhängige skalare DGL

$$\dot{x} = f(x, \mu), \quad \mu \in \mathbb{R}$$

mit einer \mathcal{C}^1 -Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. Die DGL besitze eine differenzierbare Schar $(g(\mu), \mu)$, $\mu \in (-\varepsilon, \varepsilon)$ von Ruhelagen und es gelte $f_x(g(0), 0) \neq 0$. Zeigen Sie, dass damit für jeden (betragsmäßig) hinreichend kleinen Parameterwert μ die Stabilität der Ruhelage $g(\mu)$ festgelegt ist.

Aufgabe 3.10

Gegeben sei eine von zwei Parametern μ und ν abhängige DGL

$$\dot{x} = \mu(1+x) + (\mu+\nu)x^2, \quad \mu, \nu \in \mathbb{R}.$$

Zeigen Sie, dass für jedes feste $\nu \neq 0$ bei $(x, \mu) = (0, 0)$ eine Sattel-Knoten-Bifurkation auftritt und skizzieren Sie das jeweilige Phasenportrait. Welcher Verzweigungstyp liegt für $\nu = 0$ vor?

Aufgabe 3.11

Gegeben sei die von zwei Parametern μ und ν abhängige DGL

$$\dot{x} = \mu x + (\mu + \nu)x^2 + x^3, \quad \mu, \nu \in \mathbb{R}.$$

Zeigen Sie, dass für jeden festen Wert $\nu \neq 0$ bei $(x, \mu) = (0, 0)$ eine transkritische Bifurkation oder eine Pitchfork-Bifurkation vorliegt und skizzieren Sie das jeweilige Phasenportrait.

Aufgabe 3.12

Zeigen Sie, dass in dem System

$$\dot{x} = \mu \sin x + y \cos x - x^2 + xy^2, \quad \dot{y} = \sin(\mu y - x) + x^2 e^{xy}$$

beim Parameterwert $\mu = 0$ im Koordinatenursprung eine Hopf-Bifurkation stattfindet. Welches Stabilitätsverhalten zeigt der Grenzzyklus?

Aufgabe 3.13

Gegeben sei eine skalare DGL 2. Ordnung der Form

$$\ddot{x} = \mu x - x + h(x, \dot{x}, \mu), \quad \mu \in \mathbb{R}.$$

Stellen Sie mit Hilfe des Satzes 3.8 von Hopf eine Bedingung an die Funktion h auf, die für betragsmäßig kleine Werte des Parameters μ nichtkonstante periodische Lösungen dieser DGL liefert.

Aufgabe 3.14

Gegeben sei die skalare DGL 2. Ordnung

$$\ddot{x} = \dot{x}(\mu - x^2) - x, \quad \mu \in \mathbb{R}.$$

Untersuchen Sie, ob die DGL für betragsmäßig kleine Werte des Parameters μ nichtkonstante periodische Lösungen liefert.

4 Analytische Bifurkationstheorie

Übersicht

4.1	Bifurkationsgleichung von Ljapunov-Schmidt	152
4.2	Anwendungsbeispiele	177
4.3	Die Hopf-Bifurkation	211
4.4	Aufgaben	224

Mit dem Begriff *Bifurkation* werden Phänomene beschrieben, die bei nichtlinearen, parameterabhängigen Gleichungssystemen bzw. nichtlinearen DGL-Systemen auftreten. Falls die Systemparameter geändert werden, kann für einen bestimmten Parameterwert ein plötzlicher Wechsel der Lösungsstruktur auftreten. Dann sprechen wir von Bifurkation. Neben der analytischen Bifurkationstheorie gibt es auch die topologische Bifurkationstheorie. Beide Theorien unterscheiden sich jedoch lediglich durch ihre Untersuchungsmethoden – analytische Techniken bzw. Abbildungsgrad-Techniken – nicht durch den Gegenstand an sich. In diesem Kapitel werden (anders als in Kapitel 3) zwei verschiedene Aspekte diskutiert.

Statische Bifurkationstheorie beschäftigt sich mit dem Wechsel der Struktur der Nullstellen in parameterabhängigen Gleichungen bei Änderung des Parameters. Der Parameter kann dem \mathbb{R}^n oder auch einem Funktionenraum entnommen sein und die entwickelten Methoden arbeiten konstruktiv – ein Vorteil der analytischen Bifurkationsmethoden. Eine unmittelbare Anwendung für die statische Theorie sind die Differenzialgleichungen. Die Gleichgewichtslösungen (Fixpunkte) eines DGL-Systems sind genau die Nullstellen des Vektorfeldes.

Dynamische Bifurkationstheorie beschäftigt sich mit der Änderung des Lösungsverhaltens eines DGL-Systems, wenn der Parameter in dem Vektorfeld variiert. Diese Theorie ist eng mit dem Wechsel des qualitativen Verhaltens von Systemen und dem Wechsel ihres Stabilitätsverhaltens verknüpft. Betrachtet man z. B. eine Flüssigkeit, die um eine Achse rotiert, dann bricht bei entsprechenden Winkelgeschwindigkeiten die existierende Gleichgewichtsform in neue Formen auf. Ein anderes Beispiel für ein Bifurkationsproblem ist die Bestimmung der kritischen Kraft bei der Deformation von Stäben, Platten und Schalen. Des Weiteren wird die Theorie auf verschiedene Probleme der nichtlinea-

ren Schwingungen angewendet. Solche Schwingungen spielen eine grundlegende Rolle in elektrischen Schwingkreisen, mechanischen Systemen (d. h. Motoren, Maschinen, Stäben und Platten), in biologischen Systemen usw. Eine Vielzahl von diesen können nicht durch lineare Systeme beschrieben werden. Würde man dies durch geeignete (vereinfachende) Modellannahmen erzwingen, so könnte man solche typischen Phänomene wie Grenzzyklen (siehe Beispiele 3.9 und 4.20) mit diesen linearen Modellen nicht nachweisen. In Abschnitt 4.3 wird ein abstraktes Hopf-Bifurkationstheorem bewiesen (siehe Satz 4.29), welches wesentlich allgemeiner als Satz 3.8 ist. Weitere interessante Modellprobleme zu nichtlinearen Schwingungen werden in den Kapiteln 6 und 7 untersucht.

4.1 Bifurkationsgleichung von Ljapunov-Schmidt

An der einfachen Gleichung

$$F : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad F(x, \mu) := x^2 - \mu = 0 \quad (x, \mu \in \mathbb{R})$$

tritt bereits *Lösungsbifurkation* am Punkt $(x, \mu) = (0, 0)$ ein. Für $\mu < 0$ besitzt die Gleichung keine reelle Lösung, während für $\mu > 0$ die Gleichung zwei reelle Lösungen besitzt.

Für DGL-Systeme kann man ebenfalls Bifurkation in Bezug auf die qualitative Änderung des Phasenportraits an einem bestimmten Parameterwert feststellen. Betrachtet man das DGL-System

$$\dot{x} = A(\mu)x \quad \text{mit} \quad A(\mu) \in \mathbb{R}^{2 \times 2}, \quad x \in \mathbb{R}^2, \quad \mu \in \mathbb{R} \quad (4.1)$$

dann hängen die beiden Eigenwerte stetig von μ ab. Nimmt man einmal an, dass die Eigenwerte $\lambda(\mu) = \alpha(\mu) \pm i\beta(\mu)$ bei $\mu = 0$ die imaginäre Achse überschreiten, dann führt dies zu einem Stabilitätswechsel des Gleichgewichtspunkts $x = 0$: Z. B. von einem stabilen Spiralpunkt ($\operatorname{Re} \alpha(\mu) < 0$) zu einem Zentrum ($\operatorname{Re} \alpha(\mu) = 0$) und dann zu einem instabilen Spiralpunkt ($\operatorname{Re} \alpha(\mu) > 0$). Das System (4.1) wechselt also sein qualitatives Verhalten, wenn μ von Werten $\mu < 0$ über $\mu = 0$ zu $\mu > 0$ variiert und dabei die Eigenwerte $\lambda(\mu)$ die imaginäre Achse überqueren.

Dies führt zu der Beobachtung: Ein Stabilitätswechsel kann dadurch verursacht werden, dass durch äußere Einflüsse die strukturelle Stabilität gestört wird. Die alten Strukturen verlieren ihre Stabilität und werden durch neue ersetzt.

4.1.1 Herleitung einer Bifurkationsgleichung

Betrachtet wird die reelle Gleichung

$$F : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad F(x, \mu) = 0, \quad (x, \mu \in \mathbb{R}) \quad (4.2)$$

mit $F(x_0, \mu_0) = 0$, $F_x(x_0, \mu_0) \neq 0$ und F ist eine \mathcal{C}^1 -Abbildung in einer Umgebung von (x_0, μ_0) . Nach dem impliziten Funktionentheorem (siehe Satz 1.80) kann man (4.2) eindeutig nach x in einer Umgebung von μ_0 auflösen. Es geht also genau eine Lösungskurve durch (x_0, μ_0) (siehe Abbildung 4.1 a)). Eine andere Situation in der Umgebung

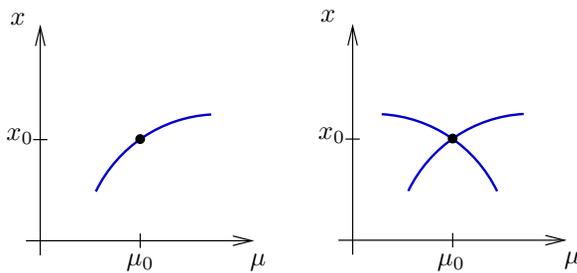


Abb. 4.1 a) Eindeutige Lösung durch (μ_0, x_0) , b) Bifurkationspunkt (μ_0, x_0)

eines Punktes μ_0 ist in Abbildung 4.1 b) dargestellt. Dieses Lösungsphänomen kann bei Gleichungen (4.2) auftreten, für die $F_x(x_0, \mu_0) = 0$ ist. Ein einfaches Beispiel liefert die Gleichung $F(x, \mu) := (\mu - \mu_0)^2 + (x - x_0)^2 = 0$.

Eine entsprechende Situation liegt vor, wenn F auf Banach-Räumen erklärt ist. Hierbei ist μ die Parametervariable und x bezeichnet den Zustand des Systems.

Zunächst wird geklärt, was man unter einem *Bifurkationspunkt* versteht. Betrachtet wird die Gleichung

$$F : X \times \mathbb{K}^m \rightarrow Y, \quad F(x, \mu) = 0 \quad \text{mit } x \in X, \mu \in \mathbb{K}^m, m \in \mathbb{N}, \quad (4.3)$$

wobei X und Y Banach-Räume sind.

Definition 4.1 (Bifurkationspunkt)

Sei X ein Banach-Raum über \mathbb{K} ($= \mathbb{R}, \mathbb{C}$).

1. Der Punkt $(x_0, \mu_0) \in X \times \mathbb{K}^m$ heißt *Bifurkationspunkt* von (4.3), falls

- i. $F(x_0, \mu_0) = 0$,
- ii. Zwei Folgen $(x_n, \mu_n), (y_n, \nu_n) \subset X \times \mathbb{K}^m$ ($x_n \neq y_n$ für alle $n \in \mathbb{N}$) existieren, die Lösungen der Gleichung (4.3) sind und mit $n \rightarrow \infty$ gegen (x_0, μ_0) konvergieren.

2. Der Punkt $\mu_0 \in \mathbb{K}^m$ heißt *Bifurkationspunkt* von (4.3), falls die Gleichung für alle μ stets die triviale Lösung $x = 0$ besitzt, d. h. es gilt $F(0, \mu) = 0$ für alle $\mu \in \mathbb{R}^m$, und vom Punkt $(0, \mu_0)$ mindestens eine (nichttriviale) Lösung $x \neq 0$ abzweigt (siehe Abbildung 4.2).

Letzteres kann auch noch so formuliert werden: μ_0 heißt Bifurkationspunkt von $F = 0$, falls es eine Folge $(x_n, \mu_n) \in X \times \mathbb{R}$ mit $x_n \neq 0$ und $F(x_n, \mu_n) = 0$ gibt, so dass $(x_n, \mu_n) \rightarrow (0, \mu_0)$ gilt.

In den Anwendungen werden die Parameter $\mu \in \mathbb{R}^m$ die Rolle einer Kraft, Druck oder/und Temperatur usw. spielen. Selbstverständlich kann der Parameter auch aus einem abstrakten Parameterraum P (P Banach-Raum) entnommen sein. An der Definition 4.1 ändert sich deshalb nichts.

Häufig liegt für Gleichungen vom Typ (4.3) mit der Linearisierung $F_x(x_0, \mu_0)$ folgende Situation vor: Die Gleichung $F(x, \mu) = 0$ hat für jedes μ die triviale Lösung $x = 0$. Darüber hinaus können an Punkten μ_1, μ_2, \dots , für die

$$\dim \ker(F_x(0, \mu_i)) \geq 1$$

ist, nichttriviale Lösungen von der trivialen Lösung $x = 0$ abzweigen. Im Diagramm (siehe Abbildung 4.2) könnte das folgendes Aussehen hervorrufen. Jene Punkte $(0, \mu_i)$

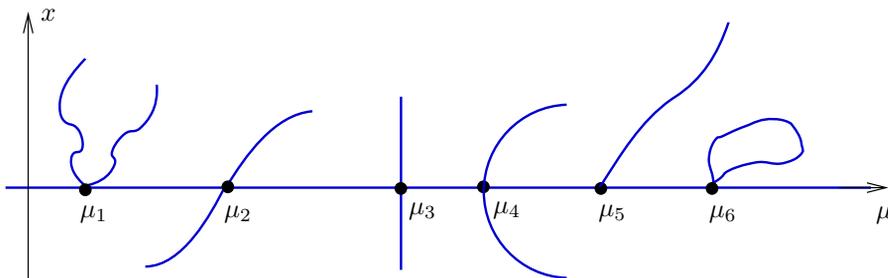


Abb. 4.2 So kann ein Bifurkationsdiagramm aussehen

in Abbildung 4.2, in denen nichttriviale Lösungen von der trivialen Lösung abzweigen, nennt man Bifurkationspunkte.

Wenn man sich mit Bifurkationsproblemen in unendlichdimensionalen Banach-Räumen beschäftigt, wird folgende Strategie verfolgt. Man verwendet Projektionsmethoden – je nach Problemstellung von *E. Schmidt*¹ oder *A. M. Ljapunov*² – um das Problem auf eine

¹Erhard Schmidt (1876–1959), deutscher Mathematiker, promovierte 1905 bei David Hilbert. Zusammen mit David Hilbert wird er als einer der Begründer der modernen abstrakten Funktionalanalysis angesehen. Er lieferte u. a. wesentliche Beiträge zur Theorie der Integralgleichungen.

²Alexander Michailowitsch Ljapunov (1857–1918), russischer Mathematiker. Er leistete grundlegende Beiträge zur Potentialtheorie, der Theorie der Gleichgewichtsfiguren rotierender Flüssigkeiten, der Stabilitätstheorie gewöhnlicher Differenzialgleichungen, sowie zur Wahrscheinlichkeitsrechnung.

endlichdimensionale nichtlineare Gleichung mit endlich vielen reellen oder komplexen Variablen zu reduzieren. Eine solche Gleichung nennt man *Bifurkationsgleichung* (siehe (4.11)). Um zu erkennen, ob es sich um ein Bifurkationsproblem handelt, wird zunächst eine notwendige Bifurkationsbedingung abgeleitet.

Lemma 4.2 (Notwendige Bifurkationsbedingung)

Es seien X und Y Banach-Räume über \mathbb{K} und $F : \mathcal{U}(x_0, \mu_0) \subset X \times \mathbb{K}^m \rightarrow Y$ eine \mathcal{C}^1 -Abbildung in einer Umgebung (x_0, μ_0) . Falls (x_0, μ_0) ein Bifurkationspunkt der Gleichung (4.3) ist, dann ist der lineare stetige Operator $F_x(x_0, \mu_0) \in \mathcal{L}(X, Y)$ nicht invertierbar.

Beweis: Nimmt man an, dass $F_x(x_0, \mu_0)$ invertierbar (auf Y) ist, dann gibt es nach dem Satz 1.80 über implizite Funktionen eine eindeutig bestimmte Lösung $x = x^*(\mu)$ auf $U(\mu_0)$ im Widerspruch dazu, dass es in $U(\mu_0)$ zu $(\mu_n) \subset U(\mu_0)$ zwei Folgen $(x_n), (y_n)$ mit $x_n \neq y_n$ für alle $n \in \mathbb{N}$ und $x_n, y_n \rightarrow x_0$ gibt, sofern $\mu_n \rightarrow \mu_0$ strebt. \square

Die Bedingung „ $F_x(x_0, \mu_0)^{-1}$ existiert nicht auf Y “ ist nicht hinreichend für das Vorliegen eines Bifurkationspunktes, wie das folgende Beispiel zeigt:

Beispiel 4.3

Es seien $X = \mathbb{R}^2$ und $x = (x_1, x_2)$. Dann ist der Punkt $\mu_0 = 1, x_1 = x_2 = 0$ kein Bifurkationspunkt für die Gleichung

$$F(x_1, x_2, \mu) := \begin{pmatrix} x_1 - \mu(x_1 - x_2^3) \\ x_2 - \mu(x_2 + x_1^3) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

obwohl $F_x(0, 0, 1)^{-1}$ nicht existiert. Es ist nämlich $F_x(0, 0, 1) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ und somit $F_x(0, 0, 1)$ nicht invertierbar. Die Multiplikation der ersten Zeile mit x_2 und der zweiten Zeile mit $-x_1$ und anschließender Addition liefert $\mu(x_1^4 + x_2^4) = 0$, d. h. $x_1 = x_2 = 0$ für $\mu \neq 0$. Die Gleichung $F = 0$ besitzt für alle $\mu \in \mathbb{R}$ nur die triviale Lösung. \blacksquare

Ein interessanter Fall tritt ein, wenn $X = Y$ und $F : X \times \mathbb{K} \rightarrow X$ die Darstellung

$$F(x, \mu) := x - \mu(Ax + G(x)) \quad (A \in \mathcal{L}(X, X), G \in \mathcal{C}^1(X, X)) \quad (4.4)$$

hat. In diesem Fall ist $F_x(x_0, \mu_0) = I - \mu_0(A + G'(x_0))$ und man kann folgendes beweisen:

Korollar 4.4

Gegeben sei das nichtlineare Eigenwertproblem

$$x = \mu(Ax + G(x)), \quad x \in X, \mu \in \mathbb{K} \quad (4.5)$$

mit $A : X \rightarrow X$ als linearen kompakten Operator auf dem Banach-Raum X (über \mathbb{K}) und $G : U(0) \subset X \rightarrow X$, definiert auf einer Umgebung von Null in X mit $\|G(x)\|/\|x\| \rightarrow 0$ für $\|x\| \rightarrow 0$. Falls $(0, \mu_0)$ ($\mu_0 \neq 0$) ein Bifurkationspunkt von (4.5) ist, dann ist μ_0 eine charakteristische Zahl für das linearisierte Problem $x = \mu_0 Ax$.

Es ist also $1/\mu_0$ ein Eigenwert von A und gehört somit zum Spektrum $\sigma(A)$. Der einfache Beweis wird in der Aufgabe 4.6 abgefragt. \square

Es kann Punkte μ geben, die zum Spektrum von A gehören, jedoch keine Bifurkationspunkte von $F = 0$ sind, wie folgendes Beispiel zeigt:

Beispiel 4.5

Es seien $X = Y = \mathbb{R}^2$, $\mu \in \mathbb{R}$, $\mu \neq 0$,

$$A := I \text{ und } G : X \rightarrow X \text{ mit } G(x, y) := (y^3, -x^3)^T$$

gegeben. Die Gleichung $F(x, y, \mu) := (x, y)^T - \mu(I(x, y)^T + G(x, y)) = 0$ hat keine Bifurkationspunkte, obwohl $\mu_0 = 1$ ein Eigenwert von A ist. Aus $\lambda x = x + y^3$ und $\lambda y = y - x^3$ ($\lambda := 1/\mu$) folgt, dass $x^4 + y^4 = 0$ und folglich $x = y = 0$ einzige Lösung von $F = 0$ ist. \blacksquare

Wir stellen nun die allgemeine Prozedur von *Ljapunov* und *Schmidt* vor, die zur Gewinnung einer *Bifurkationsgleichung* führt. Beide Vorgehensweisen stellen ein grundlegendes Werkzeug zum Auffinden von Lösungen nichtlinearer Gleichungen dar. Das Ziel besteht darin, die Lösung eines Bifurkationsproblems in unendlichdimensionalen Banach-Räumen auf die Lösung von endlich vielen Gleichungen mit endlich vielen reellen (oder komplexen) Variablen zu reduzieren. Diese Gleichungen enthalten quantitativ alle Lösungen, die auch die Ausgangsgleichung besitzt und werden deshalb auch Bifurkationsgleichungen genannt.

Die von E. Schmidt hergeleitete Bifurkationsgleichung wird mittels *Schmidt-Operator* erzeugt und ist auf den Spezialfall $\text{ind } F_x(0, \mu_0) = 0$ ausgerichtet, während die Variante von A. Ljapunov auch auf den allgemeinen Fall $\text{ind } F_x(0, \mu_0) = n$ ($n \in \mathbb{Z}$) anwendbar ist (siehe Zeidler (1986), Abschnitt 8.6).

Satz 4.6 (Ljapunov-Schmidt Bifurkationsgleichung)

Es seien X und Y Banach-Räume über \mathbb{K} ($= \mathbb{R}$ oder \mathbb{C}) und $F : \mathcal{U}(0, \mu_0) \subset X \times \mathbb{K} \rightarrow Y$ eine \mathcal{C}^k -Abbildung mit $k \geq 1$ und $F(0, \mu_0) = 0$. Weiter sei $T := F_x(0, \mu_0)$ ein Fredholm-Operator vom Index r und $\dim \ker(T) = n \geq 1$, d. h. $Tx = 0$ hat n linear unabhängige Lösungen x_1, \dots, x_n . Dann ist das Bestimmen der Lösungen (x, μ) der Gleichung $F(x, \mu) = 0$ in einer hinreichend kleinen Umgebung um $(0, \mu_0)$ äquivalent zu dem Lösen der Bifurkationsgleichung. Diese besteht aus $n - r$ nichtlinearen \mathcal{C}^k -Gleichungen in $n + 1$ Variablen über \mathbb{K} (siehe (4.11)). Falls F analytisch in einer Umgebung von $(0, \mu_0)$ ist, dann ist der Operator, der die Bifurkationsgleichung (4.11) definiert, ebenfalls analytisch um $(0, \mu_0)$.

Weitere Ausführungen zu diesem Satz findet man in Wainberg und Trenogin (1973). Hinzu kommen in diesem Buch umfangreiche Aussagen über Lösungsmethoden der Bi-

furkationsgleichung. Wir wenden uns nun dem **Beweis** zu.

Die Methode von Ljapunov: Projektionsmethode

Es seien $F \in C^2(X \times \mathbb{R}, Y)$, X, Y Banach-Räume und es gelte $F(0, \mu_0) = 0$. Gemäß Satz 4.2 ist die Stelle $(0, \mu_0)$, für die $F_x(0, \mu_0)$ nicht invertierbar ist, möglicher Bifurkationspunkt von $F = 0$. Wir setzen

$$T := F_x(0, \mu_0) \in \mathcal{L}(X, Y), \quad X_2 := \ker(T) \subset X, \quad Y_1 := R(T) \subset Y. \quad (4.6)$$

Da T ein Fredholm-Operator ist, kann man die Räume X und Y in topologisch direkte Summen zerlegen.

a. X_2 mit $\dim \ker(T) = n$ hat ein topologisch direktes Komplement X_1 in X .

Dies bedeutet, dass ein abgeschlossener Teilraum X_1 von X existiert, so dass $X = X_1 \oplus X_2$ und jedes $x \in X$ in der Form

$$x = u + z, \quad u \in X_1, \quad z \in X_2$$

geschrieben werden kann. Vom Wertevorrat $R(T) = Y_1$ wissen wir Folgendes:

b. $R(T)$ ist abgeschlossen und hat ein topologisches Komplement Y_2 in Y mit endlicher Dimension.

Dies bedeutet, dass wegen $Y = Y_1 \oplus Y_2$ jedes $y \in Y$ in der Form

$$y = v + w, \quad v \in Y_1, \quad w \in Y_2$$

dargestellt werden kann.

Nun kann man Projektionsoperatoren (siehe Seite 34) einführen, die auf die endlichdimensionalen Unterräume X_2 und Y_2 projizieren:

$$P : X \rightarrow X, \quad P(X) := \ker(T), \quad Q : Y \rightarrow Y, \quad (I - Q)(Y) := R(T). \quad (4.7)$$

Um die Projektoren explizit angeben zu können, statten wir die Unterräume X_2 und Y_2 mit Basen aus. Es bilden die $x_1, \dots, x_n \in X$ und die $g_1^*, \dots, g_m^* \in Y^*$ jeweils eine Basis in $\ker(T)$ und $\ker(T^*)$. Dazu gibt es Funktionale $f_i^* \in X^*$ und Elemente $y_j \in Y$, so dass $\{f_i^*, x_i\}$ ($i = 1, \dots, n$) und $\{g_j^*, y_j\}$ ($j = 1, \dots, m$) biorthogonale Systeme bilden (siehe Satz 1.36). Die Elemente $y_j \in Y$ ($j = 1, \dots, m$) sind eine Basis in dem Unterraum Y_2 . Nun kann eine Darstellung für die Projektionsoperatoren explizit angegeben werden:

$$Px := \sum_{i=1}^n \langle f_i^*, x \rangle x_i \in X_2, \quad Qy := \sum_{j=1}^m \langle g_j^*, y \rangle y_j \in Y_2. \quad (4.8)$$

Die Idee von Ljapunov besteht nun darin, die Gleichung

$$F(x, \mu) = 0$$

äquivalent in das System von Gleichungen

$$(I - Q)F(u + z, \mu) = 0, \quad (4.9)$$

$$QF(u + z, \mu) = 0, \quad (4.10)$$

wobei $u := (I - P)x$ und $z := Px$ sind, zu überführen. Wegen der eindeutigen Zerlegung eines Elementes $y \in Y$ in $v \in R(I - Q)$ und $w \in R(Q)$ ist die Zerlegung der Gleichung $F = 0$ in (4.9) und (4.10) ebenfalls eindeutig. Die Gleichung (4.9) kann nach $u = \gamma(z, \mu)$ in der Nähe von $(0, \mu_0)$ mit Hilfe des impliziten Funktionentheorems aufgelöst werden. Wir setzen dann diese Lösung in (4.10) ein und erhalten die Bifurkationsgleichung

$$QF(\gamma(z, \mu) + z, \mu) = 0. \quad (4.11)$$

Falls $z = \varphi(\mu)$ eine Lösung von (4.11) ist, dann ist

$$x = \gamma(\varphi(\mu), \mu) + \varphi(\mu)$$

eine Lösung von $F(x, \mu) = 0$. Man beachte, dass (4.11) aus $\dim(Q(Y))$ -Gleichungen und $(\dim(P(X)) + 1)$ -Variablen besteht.

Wir zeigen nun, dass (4.9) mittels impliziten Funktionentheorems nach u aufgelöst werden kann. Zu diesem Zweck betrachten wir die Gleichung

$$G(u, z, \mu) := (I - Q)F(u + z, \mu) = 0 \quad \text{wobei} \quad G : X_1 \times X_2 \times \mathbb{K} \rightarrow Y_1. \quad (4.12)$$

Mit den eingeführten Projektoren P und Q ist nun

$$X_1 = (I - P)(X), \quad X_2 = P(X) = \ker(T) \quad \text{und} \quad Y_1 = (I - Q)(Y).$$

Wir machen von der Tatsache Gebrauch, dass die Einschränkung

$$F_x(0, \mu_0) : X_1 \rightarrow Y_1 \quad \text{Bezeichnung:} \quad \widehat{T} := F_x(0, \mu_0)|_{X_1}$$

ein linearer, stetiger und bijektiver Operator ist. Man nennt \widehat{T}^{-1} auch *Pseudoresolvente*. Die Abbildung G erfüllt die Bedingung $G(0, 0, \mu_0) = 0$ und hat in Bezug auf die Linearisierung nach u die Gestalt $G_u(0, 0, \mu_0) = \widehat{T}$. Man beachte, dass $(I - Q)\widehat{T} = \widehat{T}$ gilt. Somit ist $G_u(0, 0, \mu_0) \in L(X_1, Y_1)$ bijektiv und das implizite Funktionentheorem liefert die eindeutige Auflösung der Gleichung (4.12) nach u in einer Umgebung von $(0, 0, \mu_0)$. Zur Gewinnung von Lösungen der Bifurkationsgleichung ist es nützlich, folgende äquivalente Umformung zu machen:

$$u = \widehat{T}^{-1}(I - Q)[F_x(0, \mu_0)u - F(u + z, \mu)]. \quad (4.13)$$

Die Umformung wird sofort klar, wenn man beachtet, dass $QF_x(0, \mu_0)u = 0$. Ist F eine C^k -Abbildung mit $k \geq 1$ auf einer Umgebung von $(0, \mu_0) \in X \times \mathbb{R}$, dann erhalten wir C^k -Lösungen $u = \gamma(z, \mu)$ auf einer Umgebung $(0, \mu_0) \in X_2 \times \mathbb{R}$ mittels impliziten Funktionentheorems durch sukzessive Approximation:

$$u_{n+1} = \widehat{T}^{-1}(I - Q)[F_x(0, \mu_0)u_n - F(u_n + z, \mu)], \quad (n = 0, 1, \dots).$$

mit dem Startpunkt $u_0 = 0$. Falls F analytisch in $(0, \mu_0)$ ist, dann erhält man die Lösung γ auch als analytische Funktion in z und μ . Wir können für $z \in X_2$ setzen

$$z := s_1 x_1 + \cdots + s_n x_n.$$

Benutzen wir noch die explizite Darstellung von Q in (4.8), dann erhalten wir die Bifurkationsgleichung (4.11) in der äquivalenten Form

$$\langle g_j^*, F(\gamma(z, \mu) + z, \mu) \rangle = 0, \quad j = 1 \cdots, m. \quad (4.14)$$

Dies sind $(n - r)$ -Gleichungen für die $(n + 1)$ Unbekannten $s_1, \dots, s_n, \mu \in \mathbb{K}$. Da $QF_x(0, \mu_0) = 0$, prüft man leicht nach, dass keine Terme auftreten können, die linear in den Variablen s_1, \dots, s_n sind. Dies verkompliziert das Lösen der Bifurkationsgleichung im allgemeinen Fall enorm. ✓

Die Methode von Schmidt: Schmidt-Operator

In den Anwendungen hat man es häufig mit dem Fall zu tun, dass T ein Fredholm-Operator mit $\text{ind}(T) = 0$ ist. Zur Herleitung einer Bifurkationsgleichung für diesem Fall ist die von E. Schmidt gewählte Variante zu empfehlen. Die Besonderheit besteht in der Wahl des so genannten *Schmidt-Operators* S , dessen stetige Inverse auf ganz Y definiert ist (im Gegensatz zur *Pseudoresolvente* bei Ljapunov). Wir betrachten das Bifurkationsproblem

$$F(x, \mu) = 0, \quad F(0, \mu) = 0, \quad F \in C^k(X \times \mathbb{K}, Y), \quad k \geq 2, \quad \mu \in \mathbb{K} \quad (4.15)$$

zusammen mit dem linearisierten Problem

$$F_x(0, \mu_0)x = 0, \quad x \in X. \quad (4.16)$$

Der Operator F kann auch analytisch in einer Umgebung von $(0, \mu_0)$ sein. Im Hinblick auf die Beispiele ist der Parameter eindimensional (siehe Zeidler (1986) für eine ausführliche Darstellung). Den **Beweis** haben wir wegen der besseren Übersichtlichkeit in mehrere Schritte unterteilt.

1. Schritt: Wir setzen wieder $T := F_x(0, \mu_0)$. Nach Voraussetzung ist $\dim \ker(T) = n$. Es sei $\{x_1, \dots, x_n\}$ eine Basis in $\ker(T)$ und $\{g_1^*, \dots, g_n^*\}$ eine Basis in $\ker(T^*)$. Dass die Dimension der Nullräume von T und T^* gleich sind, folgt unmittelbar aus Satz 1.53, wenn man beachtet, dass $\text{ind}(T) = 0$ vorausgesetzt wurde. Somit können wieder zwei biorthogonale Systeme aufgebaut werden (siehe Satz 1.36):

$$\langle f_i^*, x_j \rangle = \delta_{ij} \quad \text{und} \quad \langle g_i^*, y_j \rangle = \delta_{ij}, \quad i, j = 1, \dots, n. \quad (4.17)$$

Die $y_i, i = 1, \dots, n$, bilden eine Basis in Y_2 . Nun kann der Schmidt-Operator S eingeführt werden:

$$Sx := Tx + \sum_{i=1}^n \langle f_i^*, x \rangle y_i. \quad (4.18)$$

Der Operator $S \in \mathcal{L}(X, Y)$ besitzt eine lineare stetige Inverse $S^{-1} \in \mathcal{L}(Y, X)$. Die Bijektivität von S wird dadurch „hergestellt“, dass die dem Wertevorrat von T fehlenden Dimensionen (in Bezug auf den Raum Y) durch den Term $\sum_{i=1}^n \langle f_i^*, x \rangle y_i$ wieder hinzugefügt werden. Somit ist die Surjektivität klar. Die Injektivität sieht man so. Aus $Sx = 0$ folgt zunächst $Tx = -\sum_{i=1}^n \langle f_i^*, x \rangle y_i$ und daraus

$$\begin{aligned} 0 &= \underbrace{\langle T^* g_j^*, x \rangle}_{=0} = \langle g_j^*, Tx \rangle = -\sum_{i=1}^n \langle f_i^*, x \rangle \langle g_j^*, y_i \rangle \\ &= -\sum_{i=1}^n \langle f_i^*, x \rangle \delta_{ji} = -\langle f_j^*, x \rangle, \quad j = 1, \dots, n. \end{aligned}$$

Weiter folgt aus (4.18) wegen $\langle f_j^*, x \rangle = 0$, $j = 1, \dots, n$ auch $Tx = 0$, d. h. $x \in \ker(T)$ und somit die Darstellung $x = \sum_{i=1}^n \alpha_i x_i$ mit $\alpha_i \in \mathbb{K}$. Aus $\langle f_j^*, x \rangle = 0$ folgt jedoch $\langle f_j^*, x \rangle = \sum_{i=1}^n \alpha_i \langle f_j^*, x_i \rangle = \alpha_j = 0$, $j = 1, \dots, n$, was sofort $x = 0$ nach sich zieht. Betrachtet man den additiven Anteil in (4.18) also $Ux := \sum_{i=1}^n \langle f_i^*, x \rangle y_i$, so handelt es sich hierbei um einen endlichdimensionalen linearen Operator (d. h. einen Operator mit endlichem Wertebereich), der kompakt ist. Nach dem Satz 1.53, ist dann S wieder ein Fredholm-Operator mit Index Null. Nach der Aussage i. des gleichen Satzes ist wegen $\ker(S) = \{0\}$ dann $S^{-1} \in \mathcal{L}(Y, X)$. Aus $Sx_j = y_j$ ergibt sich nun auch noch $x_j = S^{-1}y_j$, $j = 1, \dots, n$.

2. Schritt: Mit Hilfe des Schmidtschen Operators (4.18) kann die Gleichung (4.15) wie folgt umgeformt werden:

$$0 = F(x, \mu) = Tx + F(x, \mu) - Tx$$

Der Operator $T = F_x(0, \mu_0)$ wird gemäß (4.18) durch S ersetzt und man erhält

$$0 = Sx - \sum_{i=1}^n s_i y_i + F(x, \mu) - F_x(0, \mu_0)x \quad (4.19)$$

mit $s_i := \langle f_i^*, x \rangle$, $i = 1, \dots, n$. Löst man (4.19) nach Sx auf und beachtet man, das S invertierbar ist und außerdem $x_j = S^{-1}y_j$, $j = 1, \dots, n$ gilt, dann ergibt sich

$$x = \sum_{i=1}^n s_i x_i + S^{-1}[F_x(0, \mu_0)x - F(x, \mu)] \quad (4.20)$$

$$s_i = \langle f_i^*, x \rangle, \quad i = 1, \dots, n. \quad (4.21)$$

Dieses Gleichungssystem ist äquivalent zu (4.15).

3. Schritt: Die Gleichung (4.20) kann nach x in Abhängigkeit von μ und s_1, \dots, s_n aufgelöst werden (Idee von E. Schmidt). Man schreibt dazu (4.20) als Operatorgleichung in der Form

$$\left. \begin{aligned} G &: X \times \mathbb{K}^n \times \mathbb{K} \rightarrow X, \quad G(x, s_1, \dots, s_n, \mu) = 0 \\ G(x, s_1, \dots, s_n, \mu) &:= x - \sum_{i=1}^n s_i x_i - S^{-1}[F_x(0, \mu_0)x - F(x, \mu)]. \end{aligned} \right\} \quad (4.22)$$

Es ist $G(0, 0, \dots, 0, \mu_0) = 0$ und G ist von derselben Regularitätsklasse wie F . Des Weiteren ist $G_x(0, 0, \dots, 0, \mu_0) = I$ und das implizite Funktionentheorem liefert die Aussage, dass (4.20) bzw. (4.22) in einer Umgebung von $s := (s_1, \dots, s_n) = 0, \mu = \mu_0$ genau eine Lösung besitzt.

Im Fall, dass F analytisch in einer Umgebung von $(x, \mu) = (0, \mu_0)$ ist, lässt sich die Lösung in Form einer Potenzreihe mit banachraumwertigen Koeffizienten präsentieren:

$$x(\mu, s) = \sum_{\sum_{i=0}^n k_i \geq 1} x_{k_0 k_1 \dots k_n} (\mu - \mu_0)^{k_0} s_1^{k_1} \dots s_n^{k_n}, \quad x_{k_0 k_1, \dots, k_n} \in X. \quad (4.23)$$

$$(4.24)$$

Die Lösung der Gleichung (4.20) wird jetzt in (4.21) eingesetzt und man erhält die Bifurkationsgleichung zur Bestimmung der s_1, \dots, s_n in Abhängigkeit von μ :

$$s_i = \sum_{\sum_{i=0}^n k_i \geq 1} b_{k_0 k_1 \dots k_n}^{(i)} (\mu - \mu_0)^{k_0} s_1^{k_1} \dots s_n^{k_n}, \quad i = 1, \dots, n, \quad (4.25)$$

$$b_{k_0 k_1 \dots k_n}^{(i)} := \langle f_i^*, x_{k_0 k_1 \dots k_n} \rangle, \quad i = 1, \dots, n.$$

Die $x_{k_0 k_1 \dots k_n}$ in (4.23) können durch Koeffizientenvergleich gewonnen werden. Insofern sind also die Zahlen $b_{k_0 k_1 \dots k_n}^{(i)}$ berechenbar. Die Reihen (4.23) und (4.25) konvergieren absolut in einer Umgebung um $\mu = \mu_0$ und $s = 0$. Jede nichttriviale Lösung der Bifurkationsgleichung (4.25) erzeugt eine Lösung der Ausgangsgleichung (4.15) und kann in Form der Potenzreihe (4.23) als Funktion $x = x(\mu, s(\mu))$ repräsentiert werden. ✓

Damit ist Satz 4.6 in den beiden angekündigten Varianten bewiesen. □

Wir zeigen nun am Beispiel eines RWP, wie man eine Bifurkationsgleichung aufstellen kann.

Beispiel 4.7

Wir betrachten das RWP

$$\left. \begin{aligned} \dot{x}(t) &= f(t, x(t)), \quad t \in (0, 1) \\ x(0) &= x(1). \end{aligned} \right\} \quad (4.26)$$

Es sei $f \in C^1([0, 1] \times \mathbb{R}^n, \mathbb{R}^n)$. Jede Lösung, die in $t = 0$ startet, erfüllt die Integralgleichung

$$x(t) - x(0) = \int_0^t f(s, x(s)) \, ds$$

für alle t aus ihrem Existenzintervall. Dies bedeutet, dass $x(\cdot)$ die Randbedingungen von (4.26) genau dann erfüllt, wenn

$$G(x_0) := \int_0^1 f(s, x(s; x_0)) \, ds = 0$$

gilt. Hier bezeichnet $x(\cdot; x_0)$ die einzige Lösung von $\dot{x} = f(t, x(t))$ mit $x(0; x_0) = x_0$. Es besteht also das Problem des Lösens der Gleichung

$$G(x_0) = 0 \quad \text{für } G : \mathbb{R}^n \rightarrow \mathbb{R}^n.$$

Man kann nicht direkt das implizite Funktionentheorem anwenden, da (4.26) keine Parameter enthält. Deshalb modifizieren wir das Problem durch Hinzufügen eines multiplikativen Parameters μ zu (4.26), d. h. wir untersuchen das Problem

$$\left. \begin{aligned} \dot{x}(t) &= \mu f(t, x(t)), & t \in (0, 1) \\ x(0) &= x(1). \end{aligned} \right\} \quad (4.26_\mu)$$

Für $\mu = 0$ ist eine n -dimensionale Konstante a eine Lösung. Um den (abstrakten) funktionalanalytischen Kalkül einzusetzen, schreiben wir (4.26 $_\mu$) in Operatorform. Dazu definieren wir Banach-Räume

$$X := \{x \in C^0([0, 1], \mathbb{R}^n) \mid x(0) = x(1)\}, \quad Y := \{y \in C^0([0, 1], \mathbb{R}^n) \mid y(0) = 0\}$$

und Operatoren $L, N : X \rightarrow Y$ mit

$$Lx : t \mapsto x(t) - x(0), \quad N(x) : t \mapsto \int_0^t f(s, x(s)) \, ds, \quad t \in [0, 1].$$

Dann ist das System (4.26 $_\mu$) äquivalent zu der Operatorgleichung

$$G(x, \mu) := Lx - \mu N(x) = 0. \quad (4.27)$$

Der Operator L ist linear und stetig, folglich differenzierbar:

$$L'(x)h = Lh \quad \text{für } h \in X.$$

Der Operator N ist ebenfalls stetig differenzierbar und es gilt

$$N'(x)h : t \mapsto \int_0^t f_x(s, x(s))h(s) \, ds, \quad t \in [0, 1], \quad h \in X.$$

Weiter folgt, dass $G_x(a, 0)h = Lh$ nicht injektiv ist und

$$X_2 := \ker(L) = \text{span}\{\text{const.}\} \subset X$$

aus konstanten Funktionen $c : [0, 1] \rightarrow \mathbb{R}^n$ besteht. Darüber hinaus ist

$$Y_1 := R(L) := \{y \in Y \mid y(1) = y(0) = 0\} \subset Y.$$

Nun gibt es stetige lineare Projektoren P und Q , die auf die abgeschlossenen Unterräume X_2 und Y_1 projizieren und durch

$$Px : t \mapsto x(0), \quad Qy : t \mapsto y(t) - ty(1)$$

gegeben sind. Diese Projektoren bewirken die Zerlegung der Räume X und Y in topologisch direkte Summen: $X = X_1 \oplus X_2$ und $Y = Y_1 \oplus Y_2$. Hierbei sind

$$X_1 := \ker(P) = \{x \in X \mid x(0) = 0\}, \quad X_2 = R(P) = \ker(L) \quad \text{und} \\ Y_1 = R(Q) = R(L), \quad Y_2 := \ker(Q) = \{y \in Y \mid y(t) = ct, c \in \mathbb{R}^n\}.$$

Jetzt lässt sich auch leicht klären, dass L ein F-Operator ist. Es gilt nämlich $\dim \ker(L) = \dim X_2 = n$ und $\text{codim } R(L) = \dim Y_2 = n$. Somit ist L ein F-Operator mit $\text{ind } L = \dim \ker(L) - \text{codim } R(L) = 0$. Damit sind die Vorbereitungen für die *Ljapunov-Schmidt-Reduktion* abgeschlossen. Die Gleichung (4.27) mit $x = x_1 + a$, $x_1 \in X_1$, $a \in X_2$ ist genau dann lösbar, wenn das Paar von Gleichungen

$$G^{(1)}(x_1, a, \mu) := QL(x_1 + a) - \mu QN(x_1 + a) = QLx_1 - \mu QN(x_1 + a) = 0, \\ G^{(2)}(x_1, a, \mu) := (I - Q)L(x_1 + a) - \mu(I - Q)N(x_1 + a) = 0$$

Lösungen besitzt. Nun gilt $R(Q) = Y_1$ und somit ist $QLx_1 = Lx_1$. Hieraus folgt auch sofort:

$$(I - Q)L(x_1 + a) = (I - Q)Lx_1 = Lx_1 - QLx_1 = 0.$$

Mit diesen Vereinfachungen ergibt sich das zur Gleichung (4.27) äquivalente Gleichungssystem

$$G^{(1)}(x_1, a, \mu) = Lx_1 - \mu QN(x_1 + a) = 0, \quad (4.28)$$

$$G^{(2)}(x_1, a, \mu) = (I - Q)N(x_1 + a) = 0. \quad (4.29)$$

Nun ist $G^{(1)}(0, a, 0) = 0$ und $G_{x_1}^{(1)}(0, a, 0)h = Lh$ eine lineare, stetige und bijektive Abbildung von X_1 auf Y_1 . Somit ist die Inverse nach dem Satz 1.34 (Open Mapping Theorem) stetig. Das implizite Funktionentheorem liefert genau eine Lösung

$$x_1 := \varphi(a, \mu)$$

von (4.28) in einer Umgebung von $(a, 0) \in X_2 \times \mathbb{R}$. Außerdem gilt

$$\varphi(a, 0) = 0 \quad \text{und} \quad \varphi_a(a, 0) = 0.$$

Somit genügt es, die Gleichung

$$H(a, \mu) := (I - Q)N(\varphi(a, \mu) + a) = 0$$

in Bezug auf a zu lösen. Da $\dim X_2 = \dim Y_2 = n < \infty$ und $H : X_2 \times \mathbb{R} \rightarrow Y_2$, können wir das implizite Funktionentheorem noch einmal benutzen. Zu diesem Zweck benötigen wir ein $\tilde{a} \in X_2$, für das

$$(I - Q)N(\tilde{a}) = t \int_0^1 f(s, \tilde{a}) \, ds = 0, \quad \text{d. h.} \quad \int_0^1 f(s, \tilde{a}) \, ds = 0$$

und die Gleichung

$$(I - Q)N'(\tilde{a})d = t \left(\int_0^1 f_x(s, \tilde{a}) ds \right) d = tc$$

genau eine Lösung für jedes $c \in \mathbb{R}^n$ besitzt. Die letzte Forderung bedeutet, dass die $(n \times n)$ -Matrix $\int_0^1 f_x(s, \tilde{a}) ds$ regulär sein muss. Mit Blick auf die voran stehenden Betrachtungen hat man folgende Behauptung bewiesen:

- *Es sei $f = (f_1, \dots, f_n) : [0, 1] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ stetig und möge stetige partielle Ableitungen $\frac{\partial f_i}{\partial x_j}$ ($i, j = 1, \dots, n$) besitzen. Weiter erfülle die Funktion f die Bedingungen*

$$\int_0^1 f(s, \tilde{a}) ds = 0, \quad \det \left(\int_0^1 \frac{\partial f_i}{\partial x_j}(s, \tilde{a}) ds \right) \neq 0$$

für ein bestimmtes $\tilde{a} \in \mathbb{R}^n$. Dann existieren ein $\delta > 0$ und eine differenzierbare Abbildung $\mu \mapsto x(\cdot; \mu)$, $|\mu| < \delta$, so dass $x(\cdot; 0) = \tilde{a}$ und die Funktion $x(\cdot; \mu)$ das RWP (4.26 $_{\mu}$) erfüllt.

Um letztlich Lösungen für unser Ausgangsproblem (4.26) zu bekommen, muss bezüglich des Parameters μ eine Kurvenverfolgung bis $\mu = 1$ erfolgen. Numerische Methoden zur Kurvenverfolgung werden dazu in Abschnitt 5.2 vorgestellt. ■

4.1.2 Lösen der Bifurkationsgleichung

Die Bifurkationsgleichung (4.25) zu lösen, ist ein schwieriges Problem. Es gibt eine Vielzahl von mathematischen Methoden, um Lösungen zu erhalten. Diese sind oft abhängig vom vorausgesetzten Regularitätsgrad an F . Im Analytizitätsfall kann der *Vorbereitungssatz von Weierstraß* auf die Verzweigungsgleichungen angewendet werden. Falls nur Differenzierbarkeit von F vorliegt, kann die Gleichung mittels Singularitätentheorie in Normalform gebracht werden. Häufig werden auch Lösungen der Bifurkationsgleichung mit dem Newtondiagramm erzeugt. Detaillierte Betrachtungen findet man z. B. in Wainberg und Trenogin (1973), Zeidler (1986, 1988) und Krasnoselski et al. (1973).

Beispiel 4.8

Für den Fall, dass $T := F_x(0, \mu_0)$ ein F-Operator mit $\text{ind}(T) = 0$ und $\dim \ker(T) = 1$ und $F(0, \mu) = 0$ für $\mu \in U(\mu_0) \subset \mathbb{R}$ ist, wird gezeigt, wie man zu einer nichttrivialen Lösung des Bifurkationsproblems $F(x, \mu) = 0$ kommen kann. Die Voraussetzung $F(0, \mu) = 0$ sichert bereits das Vorhandensein der trivialen Lösung $x = 0$ für alle μ -Werte. Wir verwenden dieselben Bezeichnungen, wie wir sie bei der Herleitung des Gleichungssystems (4.20) und (4.21) eingeführt haben. Die Gleichungen lauten mit $s_1 := s$:

$$\left. \begin{aligned} x &= sx_1 + S^{-1}[F_x(0, \mu_0)x - F(x, \mu)] \\ s &= \langle f_1^*, x \rangle. \end{aligned} \right\} \quad (4.30)$$

Die Auflösung der 1. Gleichung in (4.30) führt je nach Voraussetzung an F , d. h. $F \in \mathcal{C}^k$ oder F analytisch, zu Lösungen in Form von einem Taylor-Polynom mit Restglied bzw. einer absolut konvergenten Potenzreihe in μ und s . Durch implizite Differentiation bzw. Koeffizientenvergleich sind die $x_{k_0 k_1}$ berechenbar. Es ergibt sich für (4.23) mit $n = 1$ die Darstellung:

$$x(\mu, s) = x_1 s - (S^{-1} F_{x\mu}(0, \mu_0) x_1)(\mu - \mu_0) s - \frac{1}{2} (S^{-1} F_{xx}(0, \mu_0)(x_1, x_1)) s^2 - \dots \quad (4.31)$$

Da nach Voraussetzung $F(0, \mu)$ identisch Null ist, verschwinden auch alle partiellen Ableitungen nach μ , d. h. es ist $F_\mu(0, \mu) = F_{\mu\mu}(0, \mu) = \dots = 0$. Da F keine Terme enthält, die frei von x sind, besitzen folglich alle Terme von (4.31) den Faktor s . Setzt man die Lösung (4.31) in die zweite Gleichung von (4.30) ein, dann ergibt sich

$$s = s \langle f_1^*, x_1 \rangle - (\mu - \mu_0) s \langle f_1^*, S^{-1} F_{x\mu}(0, \mu_0) x_1 \rangle - s^2 \langle f_1^*, S^{-1} F_{xx}(0, \mu_0)(x_1, x_1) \rangle - \dots \quad (4.32)$$

Wegen $\langle f_1^*, x_1 \rangle = 1$ und $(S^{-1})^* f_1^* = g_1^*$ (letzteres folgt, indem man in (4.18) zum adjungierten Operator übergeht und die Biorthogonalitätsrelationen (4.17) beachtet) erhält man die Bifurkationsgleichung

$$0 = (\mu - \mu_0) s \langle g_1^*, F_{x\mu}(0, \mu_0) x_1 \rangle + s^2 \langle g_1^*, F_{xx}(0, \mu_0)(x_1, x_1) \rangle + \dots$$

und nach Division durch s

$$0 = (\mu - \mu_0) \langle g_1^*, F_{x\mu}(0, \mu_0) x_1 \rangle + s \langle g_1^*, F_{xx}(0, \mu_0)(x_1, x_1) \rangle + \dots \quad (4.33)$$

Fordert man

$$\langle g_1^*, F_{x\mu}(0, \mu_0) x_1 \rangle \neq 0$$

(was wiederum bedeutet, dass $F_{x\mu}(0, \mu_0) x_1 \notin R(F_x(0, \mu_0))$ ist), dann kann man (4.33) mittels impliziten Funktionentheorems in einer Umgebung von $(\mu, s) = (\mu_0, 0)$ nach μ in Abhängigkeit von s auflösen: $\mu = \mu(s)$ mit $\mu_0 = \mu(0)$. Setzt man diese Lösung in (4.31) ein, so ergibt sich ein nichttrivialer Lösungszweig $x = x(s) := x(\mu(s), s)$. Es gilt nämlich $x(s) \neq 0$ für $0 < |s| < \delta$, $\delta > 0$ und $\lim_{s \rightarrow 0} x(s) = 0$. Folglich hat man mit $(x(s), \mu(s))$ außer der trivialen Lösung $(0, \mu)$ eine weitere Lösung von (4.15) gefunden. ■

Damit hat man auch sofort folgendes Resultat bewiesen (siehe Zeidler (1986), S. 383):

Satz 4.9 (Crandall und Rabinowitz (1971))

Es sei $F : \mathcal{U}(0, \mu_0) \subset X \times \mathbb{K} \rightarrow Y$ und $F \in \mathcal{C}^k$, $k \geq 2$ in einer Umgebung von $(x, \mu) = (0, \mu_0)$ mit $F(0, \mu) = 0$ für alle $\mu \in \mathbb{K}$. Sei weiter $F_x(0, \mu_0)$ ein Fredholm-Operator mit Index Null,

$$\dim \ker(F_x(0, \mu_0)) = 1, \quad x_1 \in \ker(F_x(0, \mu_0))$$

und die Bifurkationsbedingung

$$F_{x\mu}(0, \mu_0)x_1 \notin R(F_x(0, \mu_0)) \quad (4.34)$$

(äquivalent: $\langle g_1^*, F_{x\mu}(0, \mu_0)x_1 \rangle \neq 0$, $g_1^* \in \ker(F_x^*(0, \mu_0))$) erfüllt. Dann existiert in einer Umgebung von $(x, \mu) = (0, \mu_0)$ exakt ein C^{k-1} -Bifurkationszweig $s \mapsto (x(s), \mu(s))$, der durch $(0, \mu_0)$ läuft. Falls F analytisch ist, dann kann man $x(s)$ und $\mu(s)$ in Form von absolut konvergenten Potenzreihen darstellen.

Der *Beweis* wurde bereits in Beispiel 4.8 erbracht. \square

Das Weglassen der Voraussetzung (4.34) kann bereits dazu führen, dass die Gleichung $F = 0$ keinen Bifurkationspunkt besitzt (siehe Beispiel 4.13).

Wir wollen eine **zweite Beweisidee** vorstellen, die grundlegend für den Nachweis von Lösungen von Gleichungen der Form (4.11) ist. Dieser Beweis unterscheidet sich auch dadurch von der ersten Beweisvariante, dass er ohne den Schmidt-Operator S auskommt. Wir weisen eine nichttriviale Lösung nach, indem wir die Gleichung (4.14) in der Form

$$h(\lambda, s) := \langle g^*, F(sx_1 + \gamma(sx_1, \mu_0 + \lambda), \mu_0 + \lambda) \rangle = 0 \quad (4.35)$$

mit $h \in C^k(\mathbb{R} \times \mathbb{R}, \mathbb{R})$, $g^* := g_1^* \in Y^*$, $\mu = \mu_0 + \lambda$ und $z := sx_1$ mit $s \in \mathbb{R}$, $x_1 \in \ker(T)$ schreiben. Der Trick besteht nun darin, das implizite Funktionentheorem noch einmal auf $h = 0$ anzuwenden. Dazu muss h „desingularisiert“ werden, da die triviale Lösung $(\lambda, 0)$ noch vorhanden ist. Zu diesem Zweck führen wir eine Funktion

$$b(\lambda, s) := \begin{cases} h(\lambda, s)/s & \text{für } s \neq 0, \\ h_s(\lambda, 0) & \text{für } s = 0 \end{cases} \quad (4.36)$$

ein. Es ist $b \in C^{k-1}(\mathbb{R} \times \mathbb{R}, \mathbb{R})$ und $b(0, 0) = 0$.

Wir berechnen die ersten partiellen Ableitungen von b . Sie lauten:

$$b_\lambda(0, 0) = \lim_{\lambda \rightarrow 0} \frac{b(\lambda, 0) - b(0, 0)}{\lambda} = \lim_{\lambda \rightarrow 0} \frac{h_s(\lambda, 0) - h_s(0, 0)}{\lambda} = h_{s\lambda}(0, 0).$$

Wegen $h(0, 0) = 0$ folgt zunächst

$$b(0, s) - b(0, 0) = \frac{h(0, s) - h(0, 0)}{s} - h_s(0, 0)$$

und nach Division durch s

$$\frac{b(0, s) - b(0, 0)}{s} = \frac{h(0, s) - h(0, 0) - h_s(0, 0) \cdot s}{s^2} \xrightarrow[s \rightarrow 0]{\text{Taylor}} \frac{1}{2} h_{ss}(0, 0),$$

d. h. $b_s(0, 0) = (1/2)h_{ss}(0, 0)$. \checkmark

Die folgenden Eigenschaften von h werden später benötigt:

$$\text{i. } h(\lambda, 0) = 0, \quad \text{ii. } h_\lambda(0, 0) = h_{\lambda\lambda}(0, 0) = 0, \quad \text{iii. } h_s(0, 0) = 0. \quad (4.37)$$

Zu i. und ii.: Wegen $F(0, \mu) = 0$ ($\mu \in \mathbb{R}$) gilt auch $\gamma(0, \mu) = 0$ und es folgt

$$h(\lambda, 0) = \langle g^*, F(\gamma(0, \mu), \mu) \rangle = 0.$$

Damit ist auch ii. gezeigt.

Zu iii.: Die Differentiation von h nach s liefert:

$$h_s(\lambda, s) = \langle g^*, F_x(sx_1 + \gamma(sx_1, \mu_0 + \lambda), \mu_0 + \lambda)[x_1 + \gamma_z(sx_1, \mu_0 + \lambda)x_1] \rangle \quad (4.38)$$

Wir setzen $s = 0$ und beachten, dass $\gamma(0, \mu_0 + \lambda) = 0$ ist. Dies führt zu

$$h_s(\lambda, 0) = \langle g^*, F_x(0, \mu_0 + \lambda)[x_1 + \gamma_z(0, \mu_0 + \lambda)x_1] \rangle \quad (4.39)$$

und an der Stelle $(\lambda, s) = (0, 0)$

$$h_s(0, 0) = \langle g^*, F_x(0, \mu_0)[x_1 + \gamma_z(0, \mu_0)x_1] \rangle. \quad (4.40)$$

Wir zeigen

$$\gamma_z(0, \mu_0) = 0. \quad (4.41)$$

Wir überlegen uns zunächst, dass $\widehat{T}u = (I - Q)F_x(0, \mu_0)u$, $u \in X_1$ gilt. Man rechnet nach:

$$\begin{aligned} (I - Q)F_x(0, \mu_0)u &= F_x(0, \mu_0)u - QF_x(0, \mu_0)u = F_x(0, \mu_0)u - \langle g^*, F_x(0, \mu_0)u \rangle \\ &= F_x(0, \mu_0)u - \underbrace{\langle F_x^*(0, \mu_0)g^*, u \rangle}_{=0} \\ &= F_x(0, \mu_0)u = \widehat{T}u, \quad \text{da } u \in X_1. \end{aligned}$$

Damit kann (4.12) äquivalent umgeformt werden:

$$\widehat{T}u = (I - Q)F_x(0, \mu_0)u - (I - Q)F(u + z, \mu) \quad (4.42)$$

Bereits gezeigt wurde, dass (4.12) die Lösung $u = \gamma(z, \mu)$ besitzt. Somit gilt auch die Beziehung

$$\gamma(z, \mu) = \widehat{T}^{-1}(I - Q)[F_x(0, \mu_0)\gamma(z, \mu) - F(\gamma(z, \mu) + z, \mu)]$$

für (z, μ) in einer Umgebung von $(0, \mu_0)$. Differentiation nach z an der Stelle $(0, \mu_0)$ liefert für alle $h \in \ker(F_x(0, \mu_0))$

$$\begin{aligned} \gamma_z(0, \mu_0)h &= \widehat{T}^{-1}(I - Q)[F_x(0, \mu_0)\gamma_z(0, \mu_0)h - F_x(\gamma(0, \mu_0), \mu_0)[\gamma_z(0, \mu_0)h + h]] \\ &= \widehat{T}^{-1}(I - Q)[F_x(0, \mu_0)\gamma_z(0, \mu_0)h - F_x(0, \mu_0)\gamma_z(0, \mu_0)h] = 0. \end{aligned}$$

Hieraus folgt $\gamma_z(0, \mu_0) = 0$. Damit geht (4.40) über in

$$h_s(0, 0) = \langle g^*, F_x(0, \mu_0)[x_1] \rangle = \langle F_x^*(0, \mu_0)g^*, x_1 \rangle = 0$$

und Behauptung (4.37) iii. ist bewiesen. ✓

Wir benutzen die Formel (4.39), um die Ableitung $h_{s\lambda}(0, 0)$ zu berechnen:

$$\begin{aligned} h_{s\lambda}(\lambda, 0) &= \langle g^*, F_{x\lambda}(0, \mu_0 + \lambda)[x_1 + \gamma_z(0, \mu_0 + \lambda)x_1] \rangle \\ &\quad + \langle g^*, F_x(0, \mu_0 + \lambda)\gamma_{z\lambda}(0, \mu_0 + \lambda)x_1 \rangle. \end{aligned}$$

Unter Beachtung von (4.41) ergibt sich für $\lambda = 0$ die Beziehung

$$h_{s\lambda}(0, 0) = \langle g^*, F_{x\lambda}(0, \mu_0)x_1 \rangle + \langle g^*, F_x(0, \mu_0)\gamma_{z\lambda}(0, \mu_0)x_1 \rangle.$$

Da $g^*|_{R(T)} = 0$ und $F_x(0, \mu_0)\gamma_{z\lambda}(0, \mu_0)x_1 \in R(T)$ folgt $\langle g^*, F_x(0, \mu_0)\gamma_{z\lambda}(0, \mu_0)x_1 \rangle = 0$ und somit

$$h_{s\lambda}(0, 0) = \langle g^*, F_{x\lambda}(0, \mu_0)x_1 \rangle. \quad (4.43)$$

Aus der Beziehung (4.38) folgt, wenn man $\lambda = 0$ setzt und dann nach s differenziert durch direktes Nachrechnen

$$h_{ss}(0, 0) = \langle g^*, F_{xx}(0, \mu_0)[x_1, x_1] \rangle. \quad (4.44)$$

Nun sind wir in der Lage, den Beweis von Satz 4.9 zu beenden. Aus unseren Berechnungen folgt:

$$\begin{aligned} \alpha &:= b_\lambda(0, 0) = h_{s\lambda}(0, 0) = \langle g^*, F_{x\mu}(0, \mu_0)[x_1] \rangle, \\ \beta &:= b_s(0, 0) = \frac{1}{2} h_{ss}(0, 0) = \frac{1}{2} \langle g^*, F_{xx}(0, \mu_0)[x_1, x_1] \rangle. \end{aligned}$$

Auf Grund von Voraussetzung (4.34) ist $\alpha = \langle g^*, F_{x\mu}(0, \mu_0)[x_1] \rangle \neq 0$. Das implizite Funktionentheorem auf die Gleichung $b = 0$ angewendet, liefert eine einzige Funktion $\lambda \in \mathcal{C}^{-1}(-\varepsilon, \varepsilon)$, so dass $\lambda(0) = 0$ und $b(\lambda(s), s) = 0$ für alle $s \in (-\varepsilon, \varepsilon)$ gilt. Da die Gleichung $h(\lambda, s) = 0$ für $s \neq 0$ äquivalent zu $b(\lambda, s) = 0$ ist, folgt, dass die Bifurkationsgleichung (4.35) durch $\lambda = \lambda(s)$ eindeutig gelöst ist. Gemäß unserer Überlegungen ergibt sich

$$F(sx_1 + \gamma(sx_1, \mu_0 + \lambda(s)), \mu_0 + \lambda(s)) = 0 \quad \text{für alle } s \in (-\varepsilon, \varepsilon).$$

Man beachte, dass

$$(sx_1 + \gamma(sx_1, \mu_0 + \lambda(s)), \mu_0 + \lambda(s)) \neq 0 \quad \text{für } s \neq 0.$$

Folglich ist die Menge S von nichttrivialen Lösungen von $F(x, \mu) = 0$ in einer Umgebung von $(0, \mu_0)$ durch die einzige (kartesische) Kurve

$$\left. \begin{aligned} \mu &= \mu(s) = \mu_0 + \lambda(s), \\ x &= x(s) = sx_1 + \gamma(sx_1, \mu_0 + \lambda(s)) \end{aligned} \right\} \quad (4.45)$$

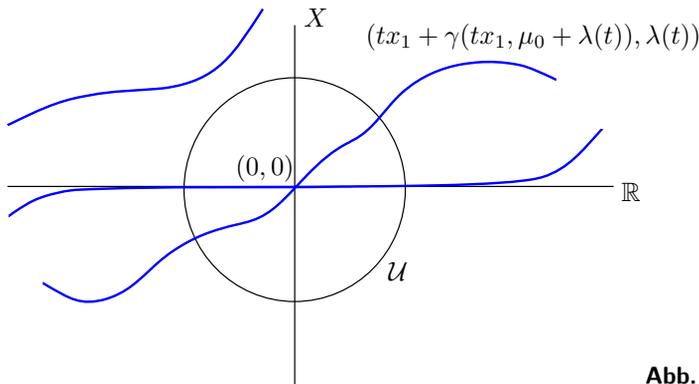


Abb. 4.3 Bifurkationsdiagramm

gegeben, wobei $s \in (-\varepsilon, \varepsilon), s \neq 0$ (siehe Abbildung 4.3 mit $\mu_0 = 0$). Falls F analytisch ist, ergeben sich diese Kurven als absolut konvergente Potenzreihen. Damit ist auch der zweite Beweis von Satz 4.9 beendet. \square

Wir wollen zwei charakteristische Bifurkationsdiagramme betrachten, wie sie bereits bei sehr einfachen Funktionen auftreten und welchen Einfluss sie auf den Stabilitätswechsel bei DGL haben können.

Beispiel 4.10

Die folgenden zwei Funktionen stellen sehr einfache illustrative Beispiele von Bifurkationen dar:

$$f(x, \mu) := \mu x - x^2, \quad g(x, \mu) := \mu x - x^3.$$

Das Bifurkationsdiagramm für die Gleichungen $f(x, \mu) = 0$ und $g(x, \mu) = 0$ zeigt Abbildung 4.4. Wir benutzen diese Funktionen, um die typischen Beispiele des Wechsels der Stabilität bei Differentialgleichungen aufzuzeigen, wenn der nichthyperbolische stationäre Punkt überquert wird. Ein stationärer Punkt x_0 heißt *hyperbolisch* für die Gleichung $\dot{x} = f(x)$, wenn $f(x_0) = 0$ und $\sigma(f'(x_0)) \cap i\mathbb{R} = \emptyset$. In Abbildung 4.4 werden Zweige stationärer Lösungen der Gleichungen $\dot{x} = f(x, \mu), \dot{x} = g(x, \mu)$ mit Angabe ihrer Stabilität gezeigt (s für stabil, u für instabil). \blacksquare

Beispiel 4.11 (Anwendung von Satz 4.9)

Wir wollen die Bifurkationspunkte des RWP

$$\left. \begin{aligned} \dot{x}(t) + \mu x(t) + g(t, x(t), \dot{x}(t), \mu) &= 0, \quad t \in [0, 2\pi] \\ x(0) = x(2\pi), \quad \dot{x}(0) = \dot{x}(2\pi) \end{aligned} \right\} \quad (4.46)$$

bestimmen. In diesem Beispiel werden wir uns auf den Punkt $\mu = 0$ konzentrieren, der ein Eigenwert der Vielfachheit eins des zugehörigen linearisierten Eigenwertproblems

$$\left. \begin{aligned} \dot{x}(t) + \mu x(t) &= 0, \quad t \in [0, 2\pi] \\ x(0) = x(2\pi), \quad \dot{x}(0) = \dot{x}(2\pi) \end{aligned} \right\} \quad (4.47)$$

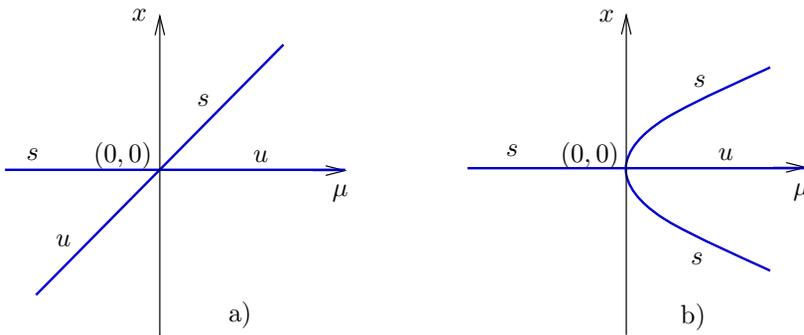


Abb. 4.4 a) Transkritische Bifurkation, b) Pitchfork Bifurkation

ist. Wir betrachten die Funktionenräume

$$X := \{x \in C^2(\mathbb{R}, \mathbb{R}) \mid x \text{ ist } 2\pi\text{-periodisch}\},$$

$$\|x\|_X := \max_{t \in [0, 2\pi]} |x(t)| + \max_{t \in [0, 2\pi]} |\dot{x}(t)| + \max_{t \in [0, 2\pi]} |\ddot{x}(t)|,$$

$$Y := \{y \in C^0(\mathbb{R}, \mathbb{R}) \mid y \text{ ist } 2\pi\text{-periodisch}\}, \quad \|y\|_Y := \max_{t \in [0, 2\pi]} |y(t)|$$

und definieren $F : X \times \mathbb{R} \rightarrow Y$ durch

$$F(x, \mu)(t) := \ddot{x}(t) + \mu x(t) + g(t, x(t), \dot{x}(t), \mu), \quad (4.48)$$

wobei die Funktion g die folgenden Bedingungen erfüllt:

- i. $\mathbb{R}^4 \ni (t, x, y, \mu) \mapsto g(t, x, y, \mu) \in \mathbb{R}$ ist 2π -periodisch in t und stetig in Bezug auf alle vier Variablen (als eine Funktion vom \mathbb{R}^4 in den \mathbb{R});
- ii. die Ableitungen von g bis zur Ordnung p ($p \geq 2$) in Bezug auf t, x, \dot{x}, μ sind stetige Funktionen;
- iii. $g(t, 0, 0, \mu) = 0$ für alle $t, \mu \in \mathbb{R}$;
- iv. $g_x(t, 0, 0, \mu) = g_y(t, 0, 0, \mu) = 0$ für alle $t, \mu \in \mathbb{R}$.

Es folgt aus iii., dass $F(0, \mu) = 0$ für alle $\mu \in \mathbb{R}$ ist. Aus iv. ergibt sich

$$F_x(0, \mu)w = \ddot{w} + \mu w.$$

Daraus können wir schließen, dass

$$\dim \ker(F_x(0, 0)) = 1.$$

ist. Man sieht dies wie folgt. Setzt man $\mu = \lambda^2$, dann ergibt sich aus $\ddot{w} + \lambda^2 w = 0$ die Lösungsschar $w(t) = c_1 \cos \lambda t + c_2 \sin \lambda t$. Ein negatives μ liefert keine periodischen Lösungen. Mit den Randbedingungen in (4.47) wird man auf das Gleichungssystem

$$\begin{pmatrix} 1 - \cos 2\pi\lambda & -\sin 2\pi\lambda \\ \lambda \sin 2\pi\lambda & \lambda(1 - \cos 2\pi\lambda) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

geführt. Dieses besitzt nichttriviale Lösungen c_1, c_2 für alle λ mit $4\lambda \sin^2 \pi\lambda = 0$. Der Eigenwert $\lambda = 0$ liefert $w(t) = \text{const}$, was man natürlich auch sofort erhalten hätte, wenn man die Gleichung $\ddot{w} = 0$ unter Berücksichtigung der Randbedingungen löst. Die anderen Eigenwerte $\mu = \lambda^2 = n^2$, $n \in \mathbb{N}$, ziehen einen zweidimensionalen Kern von $F_x(0, n^2)$ nach sich: $\ker(F_x(0, n^2)) = \text{span}\{\sin nt, \cos nt\}$.

Aus ii. folgt, dass $F \in \mathcal{C}^p(X \times [0, 2\pi], Y)$. Nach Satz 1.42 folgt:

$$Y_1 := R(F_x(0, 0)) = \{w \in Y \mid \int_0^{2\pi} w(t) dt = 0\}.$$

Hieraus resultiert auch, dass Y_1 ein abgeschlossener Teilraum von Y der Kodimension 1 ist. In Beispiel 4.18 wird der Wertevorrat von $F_x(0, 0)$ ohne Zuhilfenahme von Satz 1.42 ausführlich berechnet! Wir setzen

$$x_1 = 1, \quad X_1 := \text{span}\{1\}, \quad X_2 := \{x \in X \mid \int_0^{2\pi} x(t) dt = 0\}.$$

Wegen

$$F_{x,\mu}(0, 0)x_1 = x_1 = 1 \quad \text{und} \quad 1 \notin R(F_x(0, 0)),$$

ist die Bifurkationsbedingung (4.34) aus Satz 4.9 erfüllt. Somit zeigt sich, dass $(0, 0)$ ein Bifurkationspunkt von (4.46) ist. Insbesondere gehört der Punkt $(0, 0) \in X \times \mathbb{R}$ sowohl zum trivialen Lösungsweig $(0, \mu)$, als auch zu dem Zweig

$$\left. \begin{aligned} \mu &= \mu(s), & \mu(0) &= 0, \\ x &= x(s) = s \cdot 1 + \gamma(s \cdot 1, \mu(s)), & \gamma(0, 0) &= 0, \end{aligned} \right\} s \in (-\varepsilon, \varepsilon),$$

der durch $(0, 0)$ läuft. Man sieht, dass $x(s)$ für jedes $s \in (-\varepsilon, \varepsilon)$, $s \neq 0$ eine nichttriviale Lösung ist, die sich aus der Summe einer konstanten Funktion $s \cdot 1$ (in Bezug auf t) und der Funktion $\gamma(s \cdot 1, \mu(s)) \in X_2$ zusammensetzt. ■

Wir wollen nun für den Fall $X = Y$ Lösungen der Gleichung $F = 0$ mit der speziellen Gestalt

$$F(x, \mu) := \mu x - G(x), \quad G \in \mathcal{C}^2(X, X), \quad G(0) = 0, \quad \mu \in \mathbb{R} \quad (4.49)$$

diskutieren. Mögliche Bifurkationspunkte für die Gleichung (4.49) sind Punkte $\mu_0 \neq 0$, für die μ_0 zum Spektrum von $G'(0) \in \mathcal{L}(X, X)$ gehört. Wenn darüber hinaus $G'(0)$ kompakt ist, dann ist μ_0 tatsächlich ein Bifurkationspunkt. Dies wird in nachfolgendem Satz bewiesen.

Satz 4.12

Es sei $G \in \mathcal{C}^2(X, X)$ mit $G(0) = 0$ und $G'(0)$ ein kompakter Operator. Weiter sei $\mu_0 \neq 0$ ein einfacher Eigenwert von $G'(0)$ in dem Sinn, dass

$$\dim \ker(\mu_0 I - G'(0)) = 1, \quad (4.50)$$

$$\ker(\mu_0 I - G'(0)) \cap R(\mu_0 I - G'(0)) = \{0\} \quad (4.51)$$

gilt. Dann ist $\mu = \mu_0$ ein Bifurkationspunkt für $F(x, \mu) := \mu x - G(x) = 0$.

Der Beweis wird als Übungsaufgabe empfohlen (siehe Aufgabe 4.8). \square

Die Bifurkationsresultate aus den Sätzen 4.9 und 4.12 gelten im Allgemeinen nicht, wenn wir nur $\dim \ker(T) = 1$ und $R(T)$ abgeschlossen fordern, ohne (4.34) bzw. (4.51) vorzusetzen. Dies zeigen die nachfolgenden Beispiele.

Beispiel 4.13

Wir setzen $X = Y = \mathbb{R}^2$ und

$$F(x, y, \mu) := \begin{pmatrix} \mu x - y - y^3 \\ \mu y + x^3 \end{pmatrix}.$$

Multipliziert man in $F(x, y, \mu) = 0$ die erste Zeile mit y , die zweite mit x und subtrahiert, dann führt dies auf $y^2 + y^4 + x^4 = 0$. Somit hat $F(x, y, \mu) = 0$ nur die triviale Lösung und folglich *keine* Bifurkationspunkte.

Wir berechnen die Ableitung von F nach (x, y) (Jacobi-Matrix von F in Bezug auf (x, y)) im Punkt $(0, 0, \mu_0)$:

$$F_{(x,y)}(0, 0, \mu_0) \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \mu_0 & -1 \\ 0 & \mu_0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \mu_0 x - y \\ \mu_0 y \end{pmatrix}.$$

Am Punkt $(0, 0, 0)$ ist

$$T := F_{(x,y)}(0, 0, 0) = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}$$

nicht invertierbar. Folglich ist $\ker(F_{(x,y)}(0, 0, 0)) = \text{span} \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$ mit $x_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Außerdem gilt: $R(T) = \text{span} \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$ ist abgeschlossen und $\text{codim } R(T) = 1$. Andererseits ist die gemischte Ableitung nach (x, y) und μ

$$F_{(x,y),\mu}(0, 0, 0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$$

und folglich $F_{(x,y),\mu}(0, 0, 0)x_1 = x_1 \in R(T)$. Weder Satz 4.9 noch 4.12 kann angewendet werden. Die Bedingungen (4.34) und (4.51) sind nicht erfüllt. Die Linearisierung

$$F_{(x,y)}(0, 0, 0) = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}$$

hat am bifurkationsverdächtigen Punkt $\mu_0 = 0$ die algebraische Vielfachheit zwei und die geometrische Vielfachheit eins und ist somit kein einfacher Eigenwert. \blacksquare

Beispiel 4.14

Wir legen Wert auf die Feststellung, dass Satz 4.12 nur auf Gleichungen der speziellen Form $F(x, \mu) := \mu x - G(x) = 0$ angewendet werden kann. Ansonsten kann man Satz 4.9 benutzen, da die Bedingung (4.51) in diesem Satz keine Rolle spielt. Wir betrachten die Abbildung $F: \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^2$,

$$F(x, y, \mu) := \begin{pmatrix} \mu x - y - y^3 \\ \mu x + \mu y + x^3 \end{pmatrix} = \mu \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -y - y^3 \\ \mu x + x^3 \end{pmatrix}.$$

Sie hat nicht die Darstellung (4.49), da G auch von μ abhängt. Somit ist Satz 4.12 nicht anwendbar. Wie in Beispiel 4.13 ist für $\mu_0 = 0$

$$T = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} \quad \text{und} \quad R(T) = \text{span} \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}.$$

Die gemischte Ableitung $F_{(x,y),\mu}(0,0,0)$ ist die Matrix

$$A := F_{(x,y),\mu}(0,0,0) = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

und es gilt

$$A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \notin R(T).$$

Somit ist (4.34) erfüllt und $\mu_0 = 0$ ist ein Bifurkationspunkt von F . Eine nichttriviale Lösung des Gleichungssystems

$$\left. \begin{array}{l} \mu x - y - y^3 = 0 \\ \mu(x + y) + x^3 = 0 \end{array} \right\}$$

erhält man, wenn man Lösungen der Form $y = y^*(x)$ und $\mu = \mu^*(x)$ sucht. Man sieht leicht, dass die Lösungen die Eigenschaft $y^*(x) = -y^*(-x)$ und $\mu^*(x) = \mu^*(-x)$ haben. Der Potenzreihenansatz

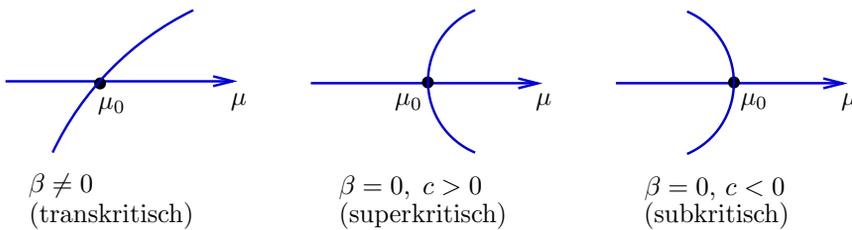
$$\begin{aligned} y^*(x) &:= a_1 x + a_3 x^3 + \dots \\ \mu^*(x) &:= b_0 + b_2 x^2 + b_4 x^4 + \dots \quad (b_0 = 0) \end{aligned}$$

liefert mittels Koeffizientenvergleich die näherungsweise Lösung für einen Bifurkationszweig:

$$y = -x^3 - x^5 - \dots, \quad \mu = -x^2 - x^4 - \dots \quad \blacksquare$$

Bemerkung 4.15

Wenn $\alpha := \langle g_1^*, F_{x\mu}(0, \mu_0)x_1 \rangle = 0$ (d.h die Bedingung (4.34) ist verletzt), dann können weitere verschiedene Bifurkationssituationen auftreten und eine umfangreichere Analyse muss sich anschließen. Für die Lösung der Bifurkationsgleichung sind in der Literatur vielfältige Verfahren entwickelt worden (siehe Chow und Hale (1982), Guckenheimer und Holmes (1983), Krasnoselski et al. (1973), Wainberg und Trenogin (1973), Zeidler (1986) u. a.). Unsere Diskussion beschränkt sich darauf, den „Verlauf der Bifurkationslösung“ in der Nähe des Bifurkationspunktes $(0, \mu_0)$ unter der Voraussetzung $\langle g_1^*, F_{x\mu}(0, \mu_0)x_1 \rangle \neq 0$ zu charakterisieren (siehe Abbildung 4.5). Zur einfacheren Darstellung setzen wir $F \in C^\infty$ voraus.

**Abb. 4.5** Verschiedene Lösungswege

Wie in der zweiten Beweisvariante zu Satz 4.9 bereits gezeigt, löst $\lambda = \mu_0 + \lambda(s)$ die Gleichung $b(\lambda, s) = 0$ (siehe (4.36)). Wir differenzieren $b(\lambda, s) = 0$ implizit nach s und erhalten

$$\lambda'(0) = -\frac{b_s(\mu_0, 0)}{b_\lambda(\mu_0, 0)} = -\frac{\beta}{\alpha}, \quad (\alpha \neq 0) \quad (4.52)$$

wobei

$$\alpha = \langle g_1^*, F_{x\mu}(0, \mu_0)x_1 \rangle, \quad \beta = \frac{1}{2} \langle g_1^*, F_{xx}(0, \mu_0)[x_1, x_1] \rangle. \quad (4.53)$$

Falls $\beta \neq 0$, dann ergibt sich

$$\lambda = \mu_0 - \frac{\beta}{\alpha}s + o(s)$$

und der Bifurkationszweig hat die Parameterdarstellung

$$x = -\frac{\alpha}{\beta}(\lambda - \mu_0)x_1 + o(\lambda - \mu_0).$$

Die Gleichung $F = 0$ besitzt nichttriviale Lösungen, sowohl für $\lambda > 0$ als auch für $\lambda < 0$ (transkritische Bifurkation).

Falls $\beta = 0$, dann findet man

$$2c := \lambda''(0) = -\frac{1}{3\alpha} \langle g_1^*, F_{xxx}(0, \mu_0)[x_1, x_1, x_1] \rangle \quad (4.54)$$

(einfaches Nachrechnen). Falls $\beta = 0$ und $c \neq 0$, dann hat der Bifurkationszweig die Form

$$x = \pm \left(\frac{\lambda - \mu_0}{c} \right)^{1/2} x_1 + O(\lambda - \mu_0).$$

Die vorliegende Formel zeigt, dass falls $c > 0$ (respektive $c < 0$), dann die abzweigende Lösung nach rechts (respektive links) von μ_0 verläuft (superkritische, respektive subkritische Bifurkation; siehe Abbildung 4.5).

Für den Fall, dass F die Darstellung $F(x, \mu) := \mu x - G(x)$ besitzt und G hinreichend glatt ist, ergeben sich die Formeln

$$\beta = -\frac{1}{2\alpha} \langle g^*, G''(0)[x_1, x_1] \rangle \quad \text{und} \quad c = \frac{1}{6\alpha} \langle g^*, G'''(0)[x_1, x_1, x_1] \rangle.$$

Falls G überdies kompakt ist, dann bekommt man mit Hilfe des Abbildungsgrades (siehe Zeidler (1986) und Zeidler (1988)) weitere Resultate geliefert. ■

In vielen Anwendungen tritt der Fall ein, dass $\ker(T)$ nicht eindimensional ist. Deshalb soll jetzt ein Bifurkationsresultat für $\dim \ker(T) > 1$ diskutiert werden. Ein interessantes Theorem zu dieser Situation findet man in Zeidler (1986), S. 392. Es ist für einen mehrdimensionalen Parameterraum E aufgestellt, wobei eine der Voraussetzungen lautet: $\dim(E) = \text{codim}(R(T))$. Auf dieses Theorem wird nicht eingegangen. Hier wird ein Ergebnis von *Ambrosetti* und *Prodi* gebracht.

Satz 4.16 (Ambrosetti und Prodi (1993))

Es seien $F \in C^\infty(X \times \mathbb{R}, Y)$, $F(0, \mu) \equiv 0$, $T := F_x(0, \mu_0)$ ein Fredholm-Operator mit $\text{ind}(T) = 0$ und es mögen die Bezeichnungen wie in (4.6) gelten. Weiter seien $M := F_{x\lambda}(0, \mu_0)$ und $B := F_{xx}(0, \mu_0)$ und es möge ein $z^* \in \ker(T)$, $z^* \neq 0$ existieren, so dass gilt:

- $QMz^* + \frac{1}{2}QB[z^*, z^*] = 0$ (Projektor Q ist in (4.7) definiert),
- die lineare Abbildung $S : \ker(T) \rightarrow Y_2$, $Sv := QMv + QB[z^*, v]$ ist invertierbar.

Dann existiert ein Zweig von nichttrivialen Lösungen von $F(x, \mu) = 0$, der von $(0, \mu_0)$ abzweigt und die Parameterdarstellung

$$\left. \begin{array}{l} \mu = \mu_0 + \lambda \\ x = \chi(\lambda) \end{array} \right\} \quad (4.55)$$

mit $\chi(0) = 0$ und $\chi'(0) = z^*$ besitzt.

Beweis: Bevor wir zum eigentlichen Beweis kommen, wird die Gleichung $F = 0$ entsprechend den Voraussetzungen umgeformt. Mit Hilfe des Taylorschen Satzes erhält man unter Beachtung von $F(0, \mu) \equiv 0$ für F die Darstellung

$$F(x, \mu_0 + \lambda) = Tx + \lambda Mx + \frac{1}{2}B[x, x] + \psi(x, \mu_0 + \lambda) \quad (4.56)$$

mit glattem ψ und

$$\psi(0, \mu) \equiv 0, \quad \psi_x(0, \mu_0) = 0, \quad \psi_{xx}(0, \mu_0) = 0, \quad \psi_{x\mu}(0, \mu_0) = 0. \quad (4.57)$$

Wir suchen Lösungen von $F = 0$ in der Form $x = \lambda(u + z)$ mit $u \in X_1, z \in \ker(T)$. Eine solche Fragestellung werden wir im Beispiel 4.18 noch einmal aufgreifen. Das Einsetzen von $x = \lambda(u + z)$ in (4.56) und Beachten von $Tz = 0$ führt auf die Gleichung

$$\lambda Tu + \lambda^2 M(u + z) + \frac{1}{2} \lambda^2 B[u + z, u + z] + \psi(\lambda(u + z), \mu_0 + \lambda) = 0. \quad (4.58)$$

Wir führen die Ljapunov-Schmidt-Reduktion durch. Das zu (4.58) äquivalente Gleichungssystem lautet:

$$\underbrace{\lambda(I - Q)Tu}_{=\hat{T}u} + \lambda^2(I - Q)M(u + z) + \frac{1}{2} \lambda^2(I - Q)B[u + z, u + z] + (I - Q)\psi(\lambda(u + z), \mu_0 + \lambda) = 0, \quad (4.59)$$

$$\underbrace{\lambda QTu}_{=0} + \lambda^2 QM(u + z) + \frac{1}{2} \lambda^2 QB[u + z, u + z] + Q\psi(\lambda(u + z), \mu_0 + \lambda) = 0, \quad (4.60)$$

wobei Q die Projektion $Q : Y \rightarrow Y_2$ ist. Gemäß der Eigenschaften von ψ (siehe (4.57)) kann man schreiben

$$\psi(\lambda(u + z), \mu_0 + \lambda) = \lambda^3 \tilde{\psi}(u, z, \lambda)$$

mit glattem $\tilde{\psi}$. Die Gleichungen (4.59) und (4.60) sind somit für $\lambda \neq 0$ äquivalent zu

$$\hat{T}u + \lambda(I - Q)M(u + z) + \frac{1}{2} \lambda(I - Q)B[u + z, u + z] + \lambda^2(I - Q)\tilde{\psi}(u, z, \lambda) = 0, \quad (4.61)$$

$$QM(u + z) + \frac{1}{2} QB[u + z, u + z] + \lambda Q\tilde{\psi}(u, z, \lambda) = 0. \quad (4.62)$$

Mit $\tilde{\Phi} = \tilde{\Phi}(u, z, \lambda)$ bezeichnen wir die linke Seite von (4.61). Es gilt sowohl $\tilde{\Phi}(0, z, 0) = 0$, als auch $\tilde{\Phi}_u(0, z, 0) = \hat{T} = T|_{X_1}, \forall z \in X_2$. Folglich kann auf die Gleichung $\tilde{\Phi}(u, z, \lambda) = 0$ der Satz über implizite Funktionen angewendet werden. Für jedes feste $z^* \in X_2$ gibt es eine eindeutig bestimmte glatte Lösung $u = \tilde{\gamma}(z, \lambda)$ in einer Umgebung von $(z, \lambda) = (z^*, 0)$. Damit gilt die Beziehung

$$\hat{T}\tilde{\gamma}(z, \lambda) + \lambda(I - Q)M(\tilde{\gamma}(z, \lambda) + z) + \frac{1}{2} \lambda(I - Q)B[\tilde{\gamma}(z, \lambda) + z, \tilde{\gamma}(z, \lambda) + z] + \lambda^2(I - Q)\tilde{\psi}(\tilde{\gamma}(z, \lambda), z, \lambda) = 0.$$

Diese Gleichung ist insbesondere auch für $\lambda = 0$ erfüllt und da $\tilde{\gamma}(z, 0) = 0$ die einzige Lösung dieser Gleichung ist, gilt $\tilde{\gamma}(z, 0) \equiv 0$ und damit auch

$$\tilde{\gamma}(z, \lambda) = \lambda\gamma(z, \lambda). \quad (4.63)$$

Einsetzen von (4.63) in (4.62) liefert

$$N(z, \lambda) := QM(\lambda\gamma(z, \lambda) + z) + \frac{1}{2}QB[\lambda\gamma(z, \lambda) + z, \lambda\gamma(z, \lambda) + z] + \lambda Q\tilde{\psi}(\lambda\gamma(z, \lambda), z, \lambda) = 0. \quad (4.64)$$

Hier ist N wieder glatt und man beachte auch, dass γ von z^* abhängt. Gemäß der Voraussetzungen des Satzes wird nun z^* so gewählt, dass man eine nichttriviale Lösung für $N = 0$ nachweisen kann. Man gibt sozusagen Richtungen an, in denen Bifurkationslösungen auftreten.

Wir kommen nun zum eigentlichen Beweis. Aus Voraussetzung a. unseres Satzes folgt $N(z^*, 0) = 0$. Die Ableitung $N_z(z^*, 0) = S$ ist nach Voraussetzung b. invertierbar. Das implizite Funktionentheorem ist auf $N(z, \lambda) = 0$ anwendbar und es existiert genau eine Lösung $z = z(\lambda)$, definiert für kleine $|\lambda|$ mit $z(0) = 0$ und $N(z(\lambda), \lambda) = 0$. Folglich finden wir einen Bifurkationszweig der Form

$$x(\lambda) = \lambda(z(\lambda) + \lambda\gamma(z(\lambda), \lambda)).$$

Setzt man $\chi(\lambda) := \lambda(z(\lambda) + \lambda\gamma(z(\lambda), \lambda))$, dann ergibt sich $\chi'(0) = z^*$. Da $x(\lambda) \neq 0$ für kleine $|\lambda|$ mit $\lambda \neq 0$ ist, ergibt sich mit $x = \chi(\lambda)$ eine nichttriviale Lösung von $F = 0$ und damit ein Bifurkationszweig. Dies beweist, dass μ_0 ein Bifurkationspunkt für $F = 0$ ist. \square

Bemerkung 4.17

1. Der Bifurkationszweig, der in Satz 4.16 gefunden wurde, ist im Allgemeinen nicht eindeutig bestimmt. Dies kann daran liegen, dass z^* aus der Gleichung a. in Satz 4.16 nicht eindeutig bestimmt werden kann oder weil es andere Bifurkationslösungen von $F(x, \mu) = 0$, die nicht die Gestalt $\lambda(u + z)$ haben, gibt.
2. Die Lösung der Bifurkationsgleichung $N(z, \lambda) = 0$ ist bezüglich λ parametrisiert. Dies weist darauf hin, dass der Satz 4.16 Anlass zu transkritischen Bifurkationslösungen gibt. \blacksquare

4.2 Anwendungsbeispiele

Im Folgenden wollen wir anhand einiger Beispiele die Arbeitsweise der vorgestellten Methoden und Sätze demonstrieren.

Beispiel 4.18 (Nichtlineare Schwingungsgleichung)

Gesucht sind nichttriviale 2π -periodische Lösungen $x \in \mathcal{C}^2(\mathbb{R}, \mathbb{R})$ der nichtlinearen Pendelgleichung

$$\ddot{x}(t) + \mu \sin x(t) = 0, \quad \mu \in \mathbb{R}. \quad (4.65)$$

Lösung: Wir setzen

$$\begin{aligned} X &:= \{x \in \mathcal{C}^2(\mathbb{R}, \mathbb{R}) \mid x \text{ ist } 2\pi\text{-periodisch}\}, \\ \|x\|_X &:= \max_{t \in [0, 2\pi]} |x(t)| + \max_{t \in [0, 2\pi]} |\dot{x}(t)| + \max_{t \in [0, 2\pi]} |\ddot{x}(t)|, \\ Y &:= \{y \in \mathcal{C}^0(\mathbb{R}, \mathbb{R}) \mid y \text{ ist } 2\pi\text{-periodisch}\}, \quad \|y\|_Y := \max_{t \in [0, 2\pi]} |y(t)|, \\ F &: X \times \mathbb{R} \rightarrow Y, \quad F(x, \mu) : t \mapsto \ddot{x}(t) + \mu \sin x(t). \end{aligned}$$

Man rechnet leicht nach, dass $F(0, \mu) = 0$ für alle $\mu \in \mathbb{R}$ und

$$(F_x(0, \mu)h)(t) = \ddot{h}(t) + \mu h(t)$$

gilt. Folglich ist $\ker(F_x(0, \mu))$ nichttrivial genau dann, wenn $\mu = n^2$, $n \in \mathbb{N} \cup \{0\}$ und

$$\begin{aligned} \ker(F_x(0, 0)) &= \{\text{konstante Funktionen}\} \\ \ker(F_x(0, n^2)) &= \text{span}\{\sin nt, \cos nt\} \quad \text{für } n \in \mathbb{N}. \end{aligned}$$

1. Wir betrachten zuerst den Fall $n = 0$. Es ist

$$R(F_x(0, 0)) = \{y \in Y \mid \int_0^{2\pi} y(s) ds = 0\}.$$

Dies folgt aus $F_x(0, 0)h = \ddot{h} = y$. Integration nach t liefert $\dot{h}(t) - \dot{h}(0) = \int_0^t y(s) ds$ und wegen $0 = \dot{h}(2\pi) - \dot{h}(0) = \int_0^{2\pi} y(s) ds$ hat man den Wertevorrat von $F_x(0, 0)$ ermittelt. Die Kodimension von $R(F_x(0, 0))$ ist eins, da der Komplementärraum zu $R(F_x(0, 0))$ nur aus konstanten Funktionen besteht. Somit ist $F_x(0, 0)$ ein Fredholm-Operator mit $\text{ind } F_x(0, 0) = 0$. (Man kann auch die Vorgehensweise, wie sie in 2. beschrieben wird, anwenden.)

Die Auflösung der Gleichung $\ddot{h} = y$ nach h ergibt sich durch nochmalige Integration der Beziehung $\dot{h}(s) - \dot{h}(0) = \int_0^s y(\tau) d\tau$:

$$\begin{aligned} \int_0^t \dot{h}(s) ds &= \int_0^t \dot{h}(0) ds + \int_0^t \left(\int_0^s y(\tau) d\tau \right) ds \\ h(t) - h(0) &= \dot{h}(0) \cdot t + \int_0^t \left(\int_\tau^t y(\tau) ds \right) d\tau \quad (\text{Vertauschung der Integration}) \\ &= \dot{h}(0) \cdot t + \int_0^t (t - \tau)y(\tau) d\tau. \end{aligned}$$

Für $\dot{h}(0)$ folgt

$$\dot{h}(0) = -\frac{1}{2\pi} \int_0^{2\pi} (2\pi - \tau)y(\tau) d\tau$$

und es ergibt sich

$$h(t) - h(0) = -\frac{t}{2\pi} \int_0^{2\pi} (2\pi - \tau)y(\tau) d\tau + \int_0^t (t - \tau)y(\tau) d\tau. \quad (4.66)$$

Man kann dies weiter umformen:

$$\begin{aligned} h(t) &= h(0) - \frac{t}{2\pi} \int_0^t (2\pi - \tau)y(\tau) \, d\tau - \frac{t}{2\pi} \int_t^{2\pi} (2\pi - \tau)y(\tau) \, d\tau + \int_0^t (t - \tau)y(\tau) \, d\tau \\ &= h(0) + \int_0^t \frac{\tau}{2\pi} (t - 2\pi)y(\tau) \, d\tau + \int_t^{2\pi} \frac{t}{2\pi} (\tau - 2\pi)y(\tau) \, d\tau. \end{aligned}$$

Schließlich erhält man die Form

$$h(t) = h(0) + \int_0^{2\pi} K(t, \tau)y(\tau) \, d\tau, \quad K(t, \tau) := \begin{cases} \frac{\tau}{2\pi} (t - 2\pi), & 0 \leq \tau \leq t \\ \frac{t}{2\pi} (\tau - 2\pi), & t \leq \tau \leq 2\pi. \end{cases}$$

Anmerkung: Aus der Formel (4.66) ergibt sich eine Vermutung für den Wertevorrat von $F_x(0, 0)$. Denkt man sich y in eine Fourier-Reihe $a_0/2 + \sum_{k=1}^{\infty} (a_k \cos kt + b_k \sin kt)$ entwickelt, dann ergibt sich für h (wenn man für y die Funktionen $\sin nt$ und $\cos nt$ einsetzt):

$$\begin{aligned} h(t) - h(0) &= -\frac{t}{2\pi} \int_0^{2\pi} (2\pi - \tau) \sin n\tau \, d\tau + \int_0^t (t - \tau) \sin n\tau \, d\tau \\ &= -\frac{t}{n} + \frac{-\sin nt + nt}{n^2} = -\frac{\sin nt}{n^2} \quad (n \in \mathbb{N}), \\ h(t) - h(0) &= -\frac{t}{2\pi} \int_0^{2\pi} (2\pi - \tau) \cos n\tau \, d\tau + \int_0^t (t - \tau) \cos n\tau \, d\tau \\ &= -\frac{\cos nt - 1}{n^2} \quad (n \in \mathbb{N}). \end{aligned}$$

Die sin- und cos-Anteile in der Fourier-Reihe erzeugen also wieder periodische Funktionen. Für das Absolutglied ergibt sich

$$h(t) - h(0) = -\frac{t}{2\pi} \int_0^{2\pi} (2\pi - \tau) \, d\tau + \int_0^t (t - \tau) \, d\tau = -\pi t + \frac{1}{2} t^2.$$

Hierbei handelt es sich um keine periodische Funktion.

Fazit: Die konstanten Funktionen bilden den Komplementärraum von $R(F_x(0, 0))$. ✓

Des Weiteren gilt $F_{x\mu}(0, 0)(c, 1) = c$ für $c \in \ker(F_x(0, 0))$ und die Voraussetzungen des Satzes von Crandall-Rabinowitz sind erfüllt. Es ist nämlich

$$F_{x\mu}(0, 0)(c, 1) = c \notin R(F_x(0, 0)) \quad (c \in \ker(F_x(0, 0))),$$

da konstante Funktionen nicht zu $R(F_x(0, 0))$ gehören. Vom Punkt $(0, 0)$ zweigt also eine nichttriviale Lösung $s \mapsto (x(s), \mu(s))$ ab!

2. Wir betrachten jetzt den Fall $n \in \mathbb{N}$. Es ist nun $\dim \ker(F_x(0, n^2)) = 2$ und der Satz von Crandall-Rabinowitz (Satz 4.9) nicht anwendbar. Wir greifen deshalb auf die Ljapunov-Schmidt-Reduktion zurück. Wir bezeichnen mit $T := F_x(0, n^2)$. Der Wertevorrat von T ergibt sich aus der einfachen Rechnung $Tx = \ddot{x} + n^2x = y$ mit $x \in X$ und $y \in Y$ (Variation der Konstanten!). Eine partikuläre Lösung x_p lautet:

$$x_p(t) = \frac{\sin nt}{n} \int_0^t y(s) \cdot \cos ns \, ds - \frac{\cos nt}{n} \int_0^t y(s) \cdot \sin ns \, ds. \quad (4.67)$$

Anmerkung: Auch hier kann man sich eine Vorstellung vom Wertevorrat von T machen, indem man sich jede stetige 2π -periodische Funktion aus X bzw. Y wieder als eine Fourier-Reihe $y(t) = a_0/2 + \sum_{k=1}^{\infty} (a_k \cos kt + b_k \sin kt)$ gegeben denkt. Die trigonometrischen Reihen liegen dicht in den Räumen \mathcal{C}^0 bzw. \mathcal{C}^2 (siehe Heuser (1980), Teil 2, S. 66, S. 158). Eingesetzt in die partikuläre Lösung $x_p(t)$ ergeben sich für die Terme

$$\int_0^t y(s) \cdot \begin{Bmatrix} \cos ns \\ \sin ns \end{Bmatrix} ds \quad (n \in \mathbb{N}) \quad (4.68)$$

die Integrale

$$\begin{aligned} \int_0^t \cos ks \cdot \cos ns \, ds &= \frac{\sin(-k+n)t}{2(-k+n)} + \frac{\sin(k+n)t}{2(k+n)}, \quad k \neq n, \\ \int_0^t \cos ks \cdot \sin ns \, ds &= -\frac{\cos(k+n)t}{2(k+n)} + \frac{\cos(k-n)t}{2(k-n)} + \frac{n}{n^2 - k^2}, \quad k \neq n, \\ \int_0^t \sin ks \cdot \sin ns \, ds &= \frac{\sin(-k+n)t}{2(-k+n)} - \frac{\sin(k+n)t}{2(k+n)}, \quad k \neq n, \\ \int_0^t \cos^2 ns \, ds &= \frac{1}{2}t + \frac{1}{4n} \sin(2nt), \quad \int_0^t \sin^2 ns \, ds = \frac{1}{2}t - \frac{1}{4n} \sin(2nt). \end{aligned}$$

Die ersten drei Ausdrücke stellen 2π -periodische Funktionen, die letzten beiden Ausdrücke stellen keine 2π -periodischen Funktionen dar. Somit gibt es zu jeder stetigen und 2π -periodischen Funktion $y(t)$, bei der die „Anteile $\cos nt$ und $\sin nt$ fehlen“, ein Urbild in \mathcal{C}^2 . Dies besagt auch, dass der Komplementärraum zu $R(T)$ durch die Funktionen $\cos nt$ und $\sin nt$ ($n \in \mathbb{N}$) aufgespannt wird und die Kodimension von $R(T)$ zweidimensional ist. \checkmark

Wir wollen nun den Projektionsoperator $Q: Y \rightarrow Y$ aufstellen. Wir fassen Y als Unterraum von $L^2((0, 2\pi), \mathbb{R})$ auf und führen das Skalarprodukt

$$\langle u^*, v \rangle_{L^2} := \frac{1}{\pi} \int_0^{2\pi} u^*(t) \cdot v(t) \, dt \quad (4.69)$$

ein. Wir setzen $x_1 = y_1 := \cos(n\cdot)$, $x_2 = y_2 := \sin(n\cdot)$ und bestimmen zu $\ker(F_x(0, 0)) = \text{span}\{x_1, x_2\}$ und $\text{codim } R(F_x(0, 0)) = \text{span}\{y_1, y_2\}$ biorthogonale Funktionale x_i^* , y_k^* . In unserem Fall sind das wieder die Funktionen $\cos(k\cdot)$ und $\sin(k\cdot)$. Damit ergibt sich gemäß Skalarprodukt:

$$x_i^*, y_k^* \in L^2 \quad \text{mit} \quad \langle x_i^*, x_j \rangle = \delta_{ij} \quad \text{und} \quad \langle y_k^*, y_l \rangle = \delta_{kl}, \quad i, j, k, l = 1, 2$$

und die Projektoren lauten:

$$Px := \sum_{i=1}^2 \langle x_i^*, x \rangle x_i \quad \text{und} \quad Qy := \sum_{k=1}^2 \langle y_k^*, y \rangle y_k, \quad x \in X, y \in Y.$$

Wegen der Auswahl der Funktionen und Funktionale kann man $P \equiv Q$ setzen, genauer: $P = Q|_X$. Der Operator

$$Qy : t \mapsto \frac{\sin nt}{\pi} \int_0^{2\pi} y(s) \cdot \sin ns \, ds + \frac{\cos nt}{\pi} \int_0^{2\pi} y(s) \cdot \cos ns \, ds, \quad y \in Y,$$

ist ein Projektionsoperator, projiziert auf den Unterraum $Z := \text{span} \{\sin(\cdot), \cos(\cdot)\}$ und splittet den Raum Y in die topologisch direkte Summe

$$Y = R(T) \oplus Z \quad \text{mit} \quad Q(Y) = R(Q) =: Z \quad \text{und} \quad (I - Q)(Y) = \ker(Q) = R(T).$$

Achtung: Der Operator

$$\tilde{Q}y : t \mapsto \frac{\sin nt}{\pi} \int_0^{2\pi} y(s) \cdot \cos ns \, ds + \frac{\cos nt}{\pi} \int_0^{2\pi} y(s) \cdot \sin ns \, ds, \quad y \in Y,$$

ist *kein* Projektionsoperator (Warum?).

Entsprechend ist wegen $Py = Qy$ ($y \in Y$):

$$X = \ker(T) \oplus V, \quad \text{wobei} \quad V := \{v \in X \mid Qv = 0\}.$$

Jedes Element x wird dargestellt als $x = u + v$, $u \in \ker(T)$, $v \in V$. Wegen $\dim \ker(T) = \text{codim } R(T) = 2$ ist T wieder ein Fredholm-Operator mit $\text{ind } T = 0$. ✓

Der Operator F kann wie folgt dargestellt werden (man beachte: $F_x(0, n^2)x = Tx = \ddot{x} + n^2x$):

$$\begin{aligned} F(u + v, \mu + n^2) &= \ddot{u} + \ddot{v} + (\mu + n^2) \sin(u + v) \\ &= T(u + v) + \mu(u + v) - (\mu + n^2)(u + v) + (\mu + n^2) \sin(u + v) \\ &= Tv + \mu(u + v) + h(u, v, \mu), \quad u \in \ker(T), v \in V, \end{aligned}$$

wobei

$$\begin{aligned} h(u, v, \mu) &:= -(\mu + n^2)(u + v) + (\mu + n^2) \sin(u + v) \\ &= \mu[\sin(u + v) - (u + v)] + n^2[\sin(u + v) - (u + v)]. \end{aligned}$$

Wegen der speziellen Form von h werden wir versuchen, eine Lösung von (4.65) in der Form $x = \mu(u + v)$ zu finden. Einsetzen in F liefert:

$$F(\mu(u + v), \mu + n^2) = \mu Tv + \mu^2(u + v) + h(\mu u, \mu v, \mu).$$

Nach Ausklammern von μ ist die Gleichung $F(\mu + n^2, \mu(u + v)) = 0$ genau dann erfüllt, wenn

$$Tv + \mu(u + v) + \mu^2 g(u, v, \mu) = 0, \quad (4.70)$$

wobei

$$g(u, v, \mu) := \begin{cases} \frac{h(\mu u, \mu v, \mu)}{\mu^3}, & \mu \neq 0 \\ -\frac{n^2}{6}(u + v)^3, & \mu = 0. \end{cases}$$

Zur Lösung der Gleichung (4.70) wird die *Ljapunov-Schmidt-Methode* benutzt. Wir projizieren dazu die Gleichung (4.70) mit Hilfe des Projektors Q in die Unterräume $R(T)$ und Z und erhalten das zu (4.70) äquivalenten Gleichungssystem:

$$\begin{aligned} 0 &= (I - Q)Tv + \mu(I - Q)(u + v) + \mu^2(I - Q)g(u, v, \mu), \\ 0 &= QTv + \mu Q(u + v) + \mu^2 Qg(u, v, \mu). \end{aligned}$$

Dies lässt sich weiter umformen, wenn man beachtet, dass

$$Qu = u \quad \text{und} \quad (I - Q)(u + v) = v$$

ist. Danach ergibt sich (wenn man aus der zweiten Gleichung noch μ ausklammert):

$$G^{(1)}(u, v, \mu) := Tv + \mu v + \mu^2(I - Q)g(u, v, \mu) = 0, \quad (4.71)$$

$$G^{(2)}(u, v, \mu) := u + \mu Qg(u, v, \mu) = 0. \quad (4.72)$$

Nun gilt für die Gleichung (4.71):

$$G^{(1)}(u^*, 0, 0) = 0 \quad \text{und} \quad G_v^{(1)}(u^*, 0, 0)h = \widehat{T}h \quad \text{mit} \quad \widehat{T} := T|_V, \quad h \in V.$$

\widehat{T} hat eine lineare stetige Inverse auf V und demzufolge ist die Gleichung (4.71) eindeutig nach v in Abhängigkeit von u und μ in einer Umgebung von $(u^*, 0)$ auflösbar:

$$v = \varphi(u, \mu), \quad \varphi(u^*, 0) = 0.$$

Wir setzen φ in die Bifurkationsgleichung (4.72) ein und erhalten

$$G(u, \mu) := G^{(2)}(u, \varphi(u, \mu), \mu) = u + \mu Qg(u, \varphi(u, \mu), \mu) = 0. \quad (4.73)$$

Da $G(u^*, 0) = u^*$, nehmen wir $u^* = 0$ und lösen (4.73) in einer Umgebung von $(0, 0)$ auf. Dies kann wieder unter Zuhilfenahme des impliziten Funktionentheorems erfolgen. Es ist $G(0, 0) = 0$ und $G_u(0, 0) = I|_{\ker(T)}$. Nun ist $G_u(0, 0)$ ein Isomorphismus von $\ker(T)$ nach Z . Somit kann u in Abhängigkeit von μ eindeutig in einer Umgebung um $\mu = 0$ aufgelöst werden. Bezeichnen wir diese Lösung mit $u = \omega(\mu)$, dann haben wir das folgende Resultat bewiesen:

- Jeder Punkt $(0, n^2)$ ist ein Bifurkationspunkt der Gleichung (4.65) und ein nichttrivialer Lösungsweig von 2π -periodischen Lösungen hat die Form

$$x = \mu(\omega(\mu) + \varphi(\omega(\mu, \mu))), \quad \mu \in (-\delta, \delta), \quad \delta > 0.$$

Es sei ausdrücklich bemerkt, dass keine Eindeutigkeit des Bifurkationszweiges bewiesen wurde. Dies liegt an der speziellen Wahl unserer Bifurkationsrichtung: $x = \mu(u + v)$. Es könnte durchaus noch andere Bifurkationsrichtungen geben (siehe Bemerkung 4.17). Um diese herauszubekommen, müsste man die Bifurkationsgleichung ohne Einschränkungen lösen, was für den allgemeinen Fall schwierig ist. ■

Beispiel 4.19 (Sturm-Liouville-Problem)

Es seien $J := [0, \pi]$, $\alpha \in C^1(J, \mathbb{R})$, $\beta \in C^0(J, \mathbb{R})$, $\alpha, \beta > 0$ auf J , $p \in C^2(J \times \mathbb{R} \times \mathbb{R})$ und seien a_0, b_0, a_1, b_1 reelle Zahlen, so dass $(a_0^2 + b_0^2)(a_1^2 + b_1^2) \neq 0$. Wir betrachten das Sturm³-Liouville-Randwertproblem

$$-Lu := -\frac{d}{dt}\left(\alpha \frac{d}{dt}u\right) + \beta u = \lambda u + p(t, u, \frac{du}{dt}), \quad t \in J, \quad (4.74)$$

$$a_0 u(0) + b_0 u'(0) = a_1 u(\pi) + b_1 u'(\pi) = 0, \quad (4.74')$$

wobei λ ein reeller Parameter ist. Über Eigenschaften von L und Lösungen von (4.74), (4.74') konsultiere man auch Ambrosetti und Prodi (1993), S. 7, Theorem 0.7.

Wir müssen (4.74), (4.74') in Form einer Operatorgleichung schreiben. Dazu setzen wir

$$X := \{u \in C^2(J, \mathbb{R}) \mid u \text{ erfüllt (4.74')}\}, \quad Y := C^0(J, \mathbb{R})$$

und definieren

$$F : X \times \mathbb{R} \rightarrow Y, \quad F(u, \lambda) := Lu + \lambda u + p(u). \quad (4.75)$$

Wir verwenden dasselbe Symbol p für den Nemytski-Operator (siehe Abschnitt 1.4), wie auch für die reellwertige Funktion p . Lösungen von (4.74), (4.74') sind dann die Paare (u, λ) mit $F(u, \lambda) = 0$ und $u \in X$. Wir setzen weiter voraus, dass $p = p(t, s, \xi)$ die Bedingungen

$$p(t, 0, 0) \equiv 0, \quad p_s(t, 0, 0) \equiv 0, \quad p_\xi(t, 0, 0) \equiv 0$$

erfüllt. Damit gilt für unser RWP

$$F(0, \lambda) = 0 \text{ für alle } \lambda \text{ und } F_u(0, \lambda) : u \mapsto Lu + \lambda u.$$

Man kann zeigen, dass das lineare Problem

$$\left. \begin{aligned} -Lu(t) &= \lambda u(t) & (t \in J), \\ a_0 u(0) + b_0 u'(0) &= a_1 u(\pi) + b_1 u'(\pi) = 0 \end{aligned} \right\} \quad (4.76)$$

³Jacques Charles Francois Sturm (1803–1855), französischer Mathematiker, nach ihm und Joseph Liouville ist das Sturm-Liouville-Problem benannt. Er ist namentlich auf dem Eiffelturm verewigt.

eine Folge von positiven einfachen Eigenwerten λ_k mit $\lambda_k \rightarrow \infty$ ($k \rightarrow \infty$) besitzt. Die φ_k sind die normierten Eigenfunktionen zu den λ_k , d.h. es ist $\int_0^\pi \varphi_k^2 dt = 1$. Nun wird der Satz 4.9 mit $\mu_0 = \lambda_k$ und $x_1 = \varphi_k$ angewendet. Wir bestimmen den Nullraum und den Wertevorrat von $F_u(0, \lambda_k)$. Es gilt

$$\ker(F_u(0, \lambda_k)) = \text{span} \{ \varphi_k \}.$$

Nach der Fredholmschen Alternative (siehe Satz 1.69) folgt für den Wertevorrat

$$\begin{aligned} R(F_u(0, \lambda_k)) &= [\ker(F_u(0, \lambda_k)^*)]^\perp \\ &= \{ u \in X \mid \langle \psi, u \rangle = 0 \} \quad \text{für alle } \psi \in \ker(F_u(0, \lambda_k)^*). \end{aligned}$$

Somit ergibt sich

$$R(F_u(0, \lambda_k)) = \{ u \in Y \mid \int_0^\pi u \varphi_k dt = 0 \}.$$

Wir definieren noch ψ durch $\langle \psi, u \rangle = \int_0^\pi u \varphi_k dt$. Da $F_{u,\lambda}(0, \lambda_k) : v \mapsto v$ und $a = \langle \psi, \varphi_k \rangle = \int_0^\pi \varphi_k^2 dt = 1$ ist auch die Bedingung (4.34) erfüllt und Satz 4.9 kann angewendet werden und liefert, dass jeder Eigenwert λ_k ein Bifurkationspunkt für $F = 0$ ist. Damit ist bewiesen:

Für jedes $k = 1, 2, \dots$ gibt es eine stetige Familie u_λ von nichttrivialen Lösungen von (4.74), (4.74'), so dass $\|u_\lambda\|_{C^2} \rightarrow 0$ ($\lambda \rightarrow \lambda_k$).

Somit ist der Existenznachweis für Lösungen des Sturm-Liouville-Problems geführt. ■

Wir wollen uns nun noch etwas mit dem geometrischen Charakter der Sätze 4.9-4.12 beschäftigen. Hat man die Ljapunov-Schmidt-Reduktion auf die Gleichung $F(u, \lambda) = 0$ angewendet, verbleibt die Diskussion der Verzweigungsgleichung. Dies ist eine endlichdimensionale Gleichung der Form $\beta(t, \mu) = 0$ mit einer C^2 -Funktion β . Sie hat die Eigenschaften

$$\beta(0, \mu) = 0 \quad \text{für alle } \mu, \quad \beta_t(0, 0) = 0 \quad \text{und} \quad \beta_{t,\mu} \neq 0.$$

Die Rechnung liefert uns zwei Lösungswege: die triviale Lösung und eine nichttriviale Lösung, die Anlass zu einer Bifurkationslösung gibt. Angenommen, dass F durch eine Funktion \tilde{F} mit $\|F - \tilde{F}\|_{C^2} < \varepsilon$ mit kleinem ε gestört wird. Die Störung von F durch \tilde{F} wird sich auch auf β auswirken, in dem Sinn, dass $\beta = 0$ durch die gestörte Bifurkationsgleichung $\tilde{\beta} = 0$ mit $\|\beta - \tilde{\beta}\|_{C^2} < \tilde{\varepsilon}$ ersetzt wird. Im Allgemeinen ist es möglich zu beweisen, dass die Nullstellen von $\tilde{\beta}$ aus zwei Zweigen bestehen, die sich zwar nicht schneiden, jedoch nahe beieinander liegen (siehe Abbildung 4.6).

Wir bringen jetzt noch ein Beispiel für den Fall, dass der Kern von $F_u(0, \lambda^*)$ zweidimensional ist (vgl. Satz 4.16).

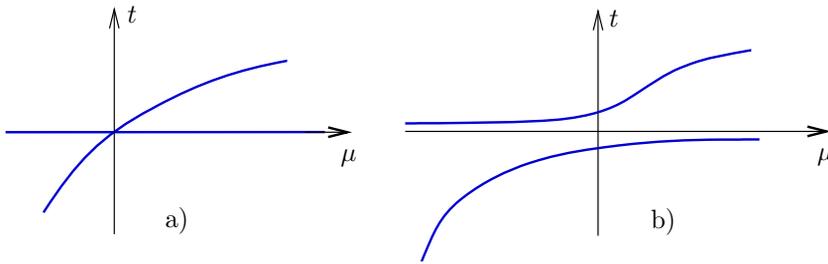


Abb. 4.6 a) Lösungen von $\beta = 0$, b) Lösungen von $\tilde{\beta} = 0$

Beispiel 4.20 (Van-der-Pol-DGL)

Die von dem holländischen Physiker *Balthasar van der Pol* um 1920 aufgestellte Gleichung lautet

$$\ddot{x} + x = \varepsilon(1 - x^2)\dot{x}, \quad (\varepsilon \text{ kleiner reeller Parameter}). \tag{4.77}$$

Sie diente damals der mathematischen Beschreibung von Oszillationen in einem elektrischen Schwingkreis mit Triode. In diesem Fall repräsentiert $x(t)$ die Spannung zur Zeit t (siehe Abbildung 4.7).

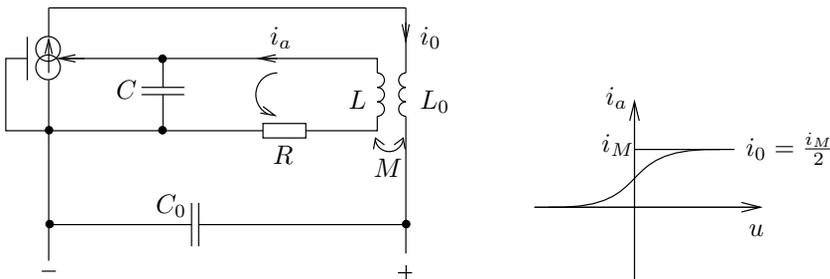


Abb. 4.7 Elektrischer Schwingkreis für die Van-der-Pol-Gleichung (links), Kennlinie der spannungsgesteuerten Stromlinie (rechts)

Auf die Herleitung der Gleichung (4.77) wird nicht eingegangen, siehe dazu Philipow (1971). Die Triodenkennlinie wird im Arbeitspunkt durch $i_a = a_0 + a_1u - a_3u^3$ ($a_0, a_1, a_3 > 0$) approximiert. Darüber hinaus wird sie in vielen anderen physikalischen Systemen und in bestimmten Modellen von chemischen Reaktionen zur Modellierung verwendet. Die Behandlung der Gleichung (4.77) erfolgt hier zunächst mit den bereitgestellten funktionalanalytischen Hilfsmitteln. Wir vergleichen die gewonnene Lösung dann auch mit der *Mittelungsmethode*.

Unser Ziel ist es, für die Gleichung (4.77) eine einzige periodische Lösung für kleine ε nachzuweisen. Eine solche Lösung nennt man wieder Grenzyklus. Dieser wird sich als ein stabiler Grenzyklus für $\varepsilon > 0$ erweisen, d. h. alle Lösungen mit Startwerten $x(t_0) \neq 0$ laufen mit $t \rightarrow \infty$ auf dem periodischen Orbit ein.

Der Leser sollte beachten, dass es auch geschlossene Lösungen in der (x, y) -Phasenebene gibt, die nicht periodisch sind. Dazu muss man als Beispiel nur das heteronome System

$$\dot{x} = 2ty, \quad \dot{y} = -2tx$$

betrachten, welches Lösungen der Form $x(t) = \alpha \cos t^2 + \beta \sin t^2$, $y(t) = -\alpha \sin t^2 + \beta \cos t^2$ besitzt, deren Phasenkurven geschlossen, jedoch nicht periodisch sind. In unserem Fall (autonome DGL) sind geschlossene Phasenkurven jedoch stets periodisch.

Wir wollen nun den Existenznachweis eines Grenzzyklus in der DGL (4.77) führen. Für $\varepsilon = 0$ ist die (x, \dot{x}) -Phasenebene mit konzentrischen Orbits ausgefüllt (siehe Abbildung 4.8 (rechts)). Man fragt nun, was passiert, wenn man dieses Phasenportrait durch die Nichtlinearität $(1 - x^2)\dot{x}$ ($\varepsilon \neq 0$) stört. Das Ergebnis ist der oben beschriebene Grenzzyklus (siehe Abbildung 4.8 (links)), der jetzt nachgewiesen wird. Für $\varepsilon = 0$ liegen Eigenlösungen der Form $x(t) = a \cos t$ mit beliebigem $a \in \mathbb{R}$ vor. Dies führt zu dem Versuch, Lösungen in der Form

$$\begin{aligned} \frac{1}{\omega} &= 1 + \eta, & x &= a_0(1 + \alpha) \cos t + y \\ \eta &= O(\varepsilon), & \alpha &= O(\varepsilon), & y &= O(\varepsilon), & \varepsilon \rightarrow 0 \end{aligned} \quad (4.78)$$

nachzuweisen.

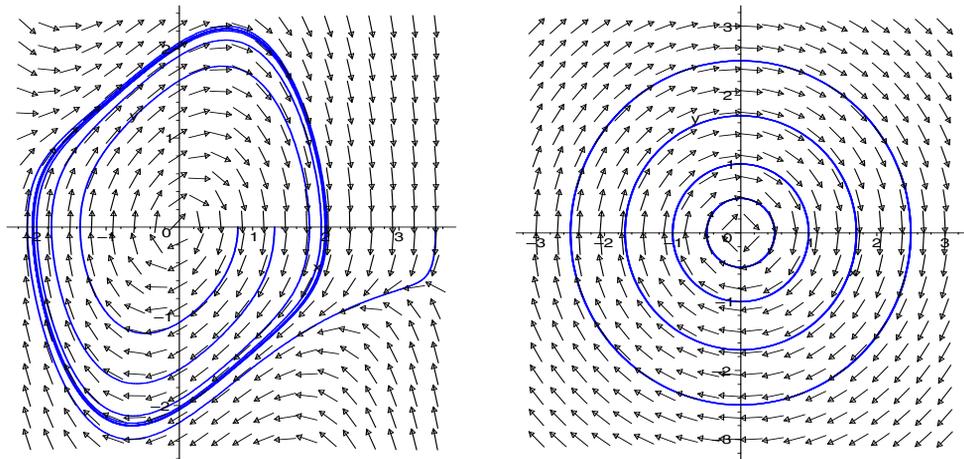


Abb. 4.8 Nur ein (asymptotisch stabiler) periodischer Orbit (links, $\varepsilon > 0$ in (4.77)), ein Kontinuum periodischer Orbits (rechts, $\varepsilon = 0$ in (4.77))

Ein Ziel ist es wieder, die Gleichung (4.77) in ein äquivalentes System zu überführen, welches aus einer eindeutig lösbaren Gleichung und den Bifurkationsgleichungen besteht. Zunächst wird die Gleichung (4.77) in die Form

$$Tx = F(x, \varepsilon), \quad T: \mathcal{C}_{2\pi}^2 \rightarrow \mathcal{C}_{2\pi}^0, \quad Tx := \ddot{x} + x, \quad F(x, \varepsilon) := \varepsilon(1 - x^2)\dot{x} \quad (4.79)$$

gebracht. Die Räume $\mathcal{C}_{2\pi}^k$ bestehen aus allen 2π -periodischen Funktionen, die k -mal stetig differenzierbar auf \mathbb{R} sind. Der Grundgedanke der Projektionsmethode besteht in der Zerlegung des Urbild- und Bildraumes des Operators T in eine direkte Summe „orthogonaler“ Teilräume. Insbesondere werden Null- und Bildraum des Operators T benötigt. Beide wurden in Beispiel 4.18 bestimmt und auch die Projektoren P und Q aufgestellt. Diese Ergebnisse werden übernommen. Zunächst wird $x \in \mathcal{C}_{2\pi}^2$ zerlegt in $x = x_1 + x_2$ mit $x_1 := Px$, $x_2 := (I - P)x$. Hierbei ist $x_1 \in R(P) = N(T)$. Da der (x, \dot{x}) -Raum der Gleichung $Tx = 0$ aus lauter konzentrischen Kreisen besteht und diese auch durch die Wahl von $x_1(t) = a \cos t$, $a \in \mathbb{R}$, erzeugt werden können, genügt es, x in der Form $x = a \cos t + y$ mit $Py = 0$ darzustellen. Die Gleichung

$$G(x, \varepsilon) := Tx - F(x, \varepsilon) = 0 \quad (4.80)$$

geht über in

$$G(x_1 + x_2, \varepsilon) = T(x_1 + x_2) - F(x_1 + x_2, \varepsilon) = Tx_2 - F(x_1 + x_2, \varepsilon) = 0.$$

Nun wird diese Gleichung mittels Q in die entsprechenden Teilräume projiziert:

$$\begin{aligned} (I - Q)[Tx_2 - F(x_1 + x_2, \varepsilon)] &= 0, \\ Q[Tx_2 - F(x_1 + x_2, \varepsilon)] &= 0. \end{aligned} \quad (4.81)$$

Wegen $QTx_2 = 0$ und $Qx_2 = Px_2 = 0$ ergibt sich schließlich (siehe (4.18))

$$Sx_2 = Tx_2 + Px_2 = (I - Q)F(x_1 + x_2, \varepsilon)$$

bzw. in der zu untersuchenden Form

$$x_2 = S^{-1}(F(x_1 + x_2, \varepsilon) - QF(x_1 + x_2, \varepsilon)), \quad (4.82)$$

$$0 = QF(x_1 + x_2, \varepsilon). \quad (4.83)$$

Die Gleichungen (4.82) und (4.83) sind äquivalent zur Ausgangsgleichung (4.79). Wie bereits erwähnt, ist (4.82) eindeutig nach x_2 in Abhängigkeit von x_1 und ε (bei kleiner Norm $\|x_1\|, |\varepsilon|$) auflösbar. Diese Lösung $x_2 = x_2^*(x_1, \varepsilon)$ mit $\|x_1\| \leq r_1$ und $|\varepsilon| \leq r_2$, $r_1, r_2 > 0$, in (4.83) eingesetzt, liefert das System der Bifurkationsgleichungen. In unserem Fall zwei nichtlineare Gleichungen der Form

$$\langle u_1, F(x_1 + x_2^*(x_1, \varepsilon)) \rangle = 0, \quad (4.84)$$

$$\langle v_1, F(x_1 + x_2^*(x_1, \varepsilon)) \rangle = 0, \quad (4.85)$$

mit $u_1(\cdot) = \frac{1}{\sqrt{\pi}} \sin(\cdot)$ und $v_1(\cdot) = \frac{1}{\sqrt{\pi}} \cos(\cdot)$. Das Symbol $\langle \cdot, \cdot \rangle$ bezeichnet das Skalarprodukt, wie in (4.69) eingeführt. Wir lösen dieses nichtlineare Gleichungssystem in drei Schritten.

1. Schritt: Wir beginnen mit einer einfachen aber wichtigen Substitution, die es uns erlaubt, das Aufsuchen von Schwingungen der Gleichung (4.79) mit einer derzeitigen unbekanntem Periode von $T = 2\pi/\omega$ ($\omega \approx 1$) auf die Periode 2π zu transformieren. Dadurch wird es möglich, das Problem in 2π -periodischen Funktionenräumen zu betrachten. Aus diesem Grund setzen wir $\tau = \omega t$. Die Gleichung (4.79) geht dann über in

$$\omega^2 \ddot{x} + x = \omega \varepsilon (1 - x^2) \dot{x}. \quad (4.86)$$

Der „Ableitungspunkt“ wird beibehalten, natürlich bedeutet er jetzt die Ableitung nach τ ; x ist nun eine Funktion von τ . Nun wird $x(\tau) = a_0(1 + \alpha) \cos \tau + y(t)$ in die Gleichung (4.86) eingesetzt und man erhält mit $a := a_0(1 + \alpha)$ und $1 + \eta = \frac{1}{\omega}$

$$\dot{y} + y = F(y, \varepsilon, \eta, \alpha), \quad y \in C_{2\pi}^2. \quad (4.87)$$

Für F ergibt sich nach Rechnung der Ausdruck

$$\begin{aligned} F(y, \varepsilon, \eta, \alpha) &:= -2\eta(a \cos \tau + y) - \eta^2(a \cos \tau + y) + \varepsilon(1 + \eta)f(y, a), \\ f(y, a) &:= \left(\frac{a^2}{4} - 1\right)a \sin \tau + \frac{a^3}{4} \sin 3\tau + a^2 y \sin 2\tau \\ &\quad + ay^2 \sin \tau + \left(1 - \frac{a^2}{2} - \frac{a^2}{2} \cos 2\tau\right)y' - 2ay \cdot y' \cos \tau - y^2 y'. \end{aligned}$$

Der Ansatz $x = a \cos \tau + y$ repräsentiert nicht jedes $x \in C_{2\pi}^2$, da der Anteil „ $\sin \tau$ “ fehlt. Würde man den Ansatz $x = a_0(1 + \alpha) \cos \tau + b_0(1 + \beta) \sin \tau + y$ mit entsprechenden Größen b_0 und β wählen, dann würde die Lösbarkeitsbedingung $PF = 0$, siehe 2. Schritt, anstatt $\frac{1}{4}a_0^2 - 1 = 0$ zweimal die gleiche Beziehung $-1 + \frac{1}{4}a_0^2 + \frac{1}{4}b_0^2 = 0$ liefern. Ohne Einschränkung kann darin $b_0 = 0$ gesetzt werden, ohne Lösungen zu „verlieren“.

2. Schritt: Es folgt die Gewinnung einer notwendigen Bedingung für die Lösung der Gleichung (4.87). Für jede Lösung y der Gleichung (4.87) muss $F \in R(T)$ gelten. Dies ist gleichbedeutend mit

$$PF = 0 \quad \text{bzw.} \quad \langle u_1, F \rangle = \langle v_1, F \rangle = 0. \quad (4.88)$$

Es wird $\langle \sin \tau, F \rangle$ berechnet (man beachte, dass $Py = 0$ gilt):

$$\begin{aligned} \langle \sin \tau, F \rangle &= \varepsilon(1 + \eta) \left[\left(\frac{a^2}{4} - 1\right)a\pi + a^2 \langle \sin \tau, y \sin 2\tau \rangle + a \langle \sin \tau, y^2 \sin \tau \rangle \right. \\ &\quad \left. + \left\langle \sin \tau, \left(1 - \frac{a^2}{2} - \frac{a^2}{2} \cos 2\tau\right)y' \right\rangle - 2a \langle \sin \tau, yy' \cos \tau \rangle - \langle \sin \tau, y^2 y' \rangle \right] = 0. \end{aligned}$$

Nach Division durch $\varepsilon \neq 0$ ergibt sich ein Ausdruck der Form

$$\left(\frac{a_0^2}{4} - 1\right)a_0\pi + 0(y, \eta, \alpha) = 0 \quad \text{für alle } (y, \eta, \alpha) \in U(0).$$

Die Gleichung ist nur erfüllt, falls $\frac{1}{4}a_0^2 - 1 = 0$, also $a_0 = \pm 2$ gewählt wird. Da $-x$ ebenfalls eine Lösung von (4.77) darstellt, liefert die Wahl $a_0 = -2$ keine neue Lösung.

3. Schritt: Die Gleichung (4.87) wird jetzt mittels Projektionsmethode in ein äquivalentes Gleichungssystem überführt. Wegen $P = Q$ wird mit dem Projektor P gearbeitet. Gemäß der Überführung der Gleichung (4.80) in das System (4.82) und (4.83) ergibt sich jetzt mit $Q = P$, $x_1 = a_0(1 + \alpha) \cos \tau$ und $x_2 = y$:

$$y = S^{-1}(F(y, \varepsilon, \eta, \alpha) - PF(y, \varepsilon, \eta, \alpha)), \quad (4.89)$$

$$0 = PF(y, \varepsilon, \eta, \alpha). \quad (4.90)$$

Das System (4.89), (4.90) ist zur Gleichung (4.87) äquivalent. Jede Lösung von (4.87) erzeugt eine Lösung des Systems und umgekehrt. Die Gleichung (4.89) lässt sich mittels impliziten Funktionentheorems (siehe Satz 1.80) eindeutig nach y in Abhängigkeit von $\varepsilon, \eta, \alpha$ auflösen. Wir bezeichnen die Lösung mit $y = y^*(\varepsilon, \eta, \alpha)$. Die Funktion y^* existiert lediglich auf einer (hinreichend kleinen) Umgebung um $\varepsilon = \eta = \alpha = 0$. (Die Voraussetzungen des impliziten Funktionentheorems sind erfüllt, da die rechte Seite in (4.89) keine Linearteile in y enthält.) Nun werden aus der Gleichung (4.89) Eigenschaften der Lösung y^* abgeleitet und einige Koeffizienten der Taylor-Reihe berechnet. Da die rechte Seite von (4.89) analytisch ist, gilt entsprechendes auch für y^* in einer Nullumgebung von ε, η und α . Somit können die Koeffizienten der Lösung y^* z. B. durch Koeffizientenvergleich gewonnen werden. Wir schreiben die Gleichung in der Form $Sy = F - PF$ und bezeichnen die rechte Seite mit $\tilde{F} := F - PF$. Dann ergibt sich

$$\begin{aligned} \tilde{F}(y, \varepsilon, \eta, \alpha) = & -2\eta y - \eta^2 y + \varepsilon(1 + \eta)[2(1 + \alpha)^3 \sin 3\tau \\ & + 4(1 + \alpha)^2 y \sin 2\tau + 2(1 + \alpha)y^2 \sin \tau - (1 + 4\alpha + \alpha^2)y' \\ & - 2(1 + \alpha)^2 y' \cos 2\tau - 4(1 + \alpha)y \cdot y' \cos \tau - y^2 \cdot y' \\ & - 2(1 + \alpha)^2 \langle v_1, y \cos 2\tau \rangle u_1 - 2(1 + \alpha) \langle v_1, y^2 \cos \tau \rangle u_1 \\ & - \frac{1}{3} \langle v_1, y^3 \rangle u_1 + 2(1 + \alpha)^2 \langle u_1, y \cos 2\tau \rangle v_1 \\ & + 2(1 + \alpha) \langle u_1, y^2 \cos \tau \rangle v_1 + \frac{1}{3} \langle u_1, y^3 \rangle v_1]. \end{aligned} \quad (4.91)$$

Für die Lösung $y^*(\varepsilon, \eta, \alpha)$ gilt für hinreichend kleine $\varepsilon, \eta, \alpha$ die Beziehung

$$Sy^* = \tilde{F}(y^*(\varepsilon, \eta, \alpha), \varepsilon, \eta, \alpha), \quad |\varepsilon|, |\eta|, |\alpha| < \delta, \quad \delta > 0. \quad (4.92)$$

Hieraus gewinnt man eine Strukturaussage für die Lösung y^* . Es gilt für $\varepsilon = 0$ die Gleichung

$$Sy^*(0, \eta, \alpha) = \tilde{F}(y^*(0, \eta, \alpha), 0, \eta, \alpha) = -2\eta y^*(0, \eta, \alpha) - \eta^2 y^*(0, \eta, \alpha). \quad (4.93)$$

Diese Gleichung besitzt jedoch nach dem Satz über implizite Funktionen genau eine Lösung y^* . Nun ist $y^*(0, \eta, \alpha) = 0$ einzige Lösung von (4.93), d. h. es gilt die Beziehung

$$y^*(0, \eta, \alpha) = 0 \quad \text{für alle } |\eta|, |\alpha| < \delta. \quad (4.94)$$

Hieraus folgt sofort, dass alle Ableitungen nach η und α sowie auch alle gemischten Ableitungen verschwinden. Dies erleichtert die Berechnung der Anfangsentwicklung von $y^*(\varepsilon, \eta, \alpha)$ erheblich und liefert außerdem die angekündigte Strukturaussage über die Lösungsdarstellung von $y^*(\varepsilon, \eta, \alpha)$:

$$y^*(\varepsilon, \eta, \alpha)(\tau) = -\frac{\varepsilon}{4} \sin 3\tau - \left(\frac{3}{16} \cos 3\tau + \frac{5}{48} \cos 5\tau \right) \varepsilon^2 - \frac{5}{16} (\sin 3\tau) \varepsilon \eta - \frac{3}{4} (\sin 3\tau) \varepsilon \alpha + O_3(\varepsilon, \eta, \alpha), \quad (4.95)$$

bzw. wegen (4.94)

$$y^*(\varepsilon, \eta, \alpha)(\tau) = -\frac{\varepsilon}{4} \sin 3\tau + O(\varepsilon^2) + \varepsilon(O(\eta) + O(\alpha)). \quad (4.96)$$

Der Term $O_3(\varepsilon, \eta, \alpha)$ enthält Glieder der Ordnung ≥ 3 .

Nun wird der Ausdruck $PF(y, \varepsilon, \eta, \alpha)$ berechnet und dann die Lösung (4.96) eingesetzt. Die Gleichung $PF = 0$ zerfällt in zwei nichtlineare Gleichungen in $\varepsilon, \eta, \alpha$. Im Einzelnen ergeben sich aus $\langle u_1, F \rangle = \langle v_1, F \rangle = 0$ die Gleichungen

$$\varepsilon(1 + \eta) \left[2\sqrt{\pi} \alpha (1 + \alpha)(2 + \alpha) + 4(1 + \alpha)^2 \langle v_1, (\cos^2 \tau) y \rangle + 2(1 + \alpha) \langle v_1, (\cos \tau) y^2 \rangle + \frac{1}{3} \langle v_1, y^3 \rangle \right] = 0, \quad (4.97)$$

$$2\sqrt{\pi} (2\eta + \eta^2)(1 + \alpha) + \varepsilon(1 + \eta) \left[4(1 + \alpha)^2 \langle u_1, (\cos^2 \tau) y \rangle + 2(1 + \alpha) \langle u_1, (\cos \tau) y^2 \rangle + \frac{1}{3} \langle u_1, y^3 \rangle \right] = 0. \quad (4.98)$$

Als nächstes wird in (4.97) und (4.98) die Lösung (4.95) eingesetzt und somit die angekündigten Bifurkationsgleichungen gewonnen. Da Lösungen mit Parameter $\varepsilon \neq 0$ gesucht werden, kann in (4.97) zuvor noch durch ε dividiert werden. Die erforderlichen Rechnungen werden dem Leser überlassen. Alle Terme, bis einschließlich zu den Gliedern 2. Ordnung sind exakt berechnet. Es ergibt sich aus (4.97)

$$4\alpha - \frac{1}{8} \varepsilon^2 + O(\varepsilon^4) + O_3(\varepsilon, \eta, \alpha) = 0 \quad (4.99)$$

und aus (4.98)

$$4\eta - \frac{1}{4} \varepsilon^2 + 2\eta^2 + 4\eta\alpha + O(\varepsilon^4) + O_3(\varepsilon, \eta, \alpha) = 0. \quad (4.100)$$

Das System (4.99), (4.100) ist linear in Bezug auf η und α . Der Satz über implizite Funktionen liefert das folgende Resultat: Es existiert ein $\varepsilon_0 > 0$, so dass das System (4.99), (4.100) Lösungen $\eta = \eta^*(\varepsilon)$ und $\alpha = \alpha^*(\varepsilon)$ mit $\eta^*(0) = \alpha^*(0) = 0$ besitzt. Die Lösungen sind für $|\varepsilon| < \varepsilon_0$ analytische Funktionen. Man kann auch hier die Koeffizienten wieder durch Koeffizientenvergleich bestimmen. Die Potenzreihendarstellungen für η^* und α^* (jeweils nach dem quadratischen bzw. kubischen Glied abgebrochen) lauten:

$$\eta^*(\varepsilon) = \frac{1}{16}\varepsilon^2 + O(\varepsilon^3), \quad \alpha^*(\varepsilon) = \frac{1}{32}\varepsilon^2 + O(\varepsilon^3), \quad |\varepsilon| < \varepsilon_0.$$

Setzt man diese Lösungen in (4.95) ein, erhält man für y^* die Darstellung ($|\varepsilon| < \varepsilon_0$)

$$y^*(\tau, \varepsilon) = -\frac{\varepsilon}{4} \sin 3\tau - \varepsilon^2 \left(\frac{3}{16} \cos 3\tau + \frac{5}{48} \cos 5\tau \right) - \frac{11\varepsilon^3}{256} \sin 3\tau + O(\varepsilon^4). \quad (4.101)$$

Nun muss man lediglich noch die Zeit τ gemäß $\tau = \omega t$ nach t und $1 + \eta = \frac{1}{\omega}$ nach ω zurück transformieren. Damit ergibt sich abschließend

$$x^*(t, \varepsilon) = 2(1 + \alpha^*(\varepsilon)) \cos \omega t - \frac{\varepsilon}{4} \sin 3\omega t - O(\varepsilon^2), \quad |\varepsilon| < \varepsilon_0 \quad (4.102)$$

mit

$$\omega = \omega^*(\varepsilon) = \frac{1}{1 + \eta^*(\varepsilon)} = 1 - \frac{1}{16}\varepsilon^2 + O(\varepsilon^4), \quad |\varepsilon| < \varepsilon_0$$

$$\alpha^*(\varepsilon) = \frac{1}{32}\varepsilon^2 + O(\varepsilon^4), \quad |\varepsilon| < \varepsilon_0.$$

Damit ist die Auflösung des Systems (4.84) erfolgt. ■

Im engen Zusammenhang mit der nachgewiesenen periodischen Lösung (4.96) steht die Frage nach der Stabilität des Grenzzyklus. Hierzu gibt es hinreichende Bedingungen für Differenzialgleichungen der Form

$$\ddot{x} + f(x) \cdot \dot{x} + g(x) = 0$$

(siehe z. B. Bogoljubov und Mitropolski (1965), Verhulst (2000)). In unserem Fall erfüllen die Funktionen $f(x) := \varepsilon(x^2 - 1)$ und $g(x) := x$ die dort angegebenen Bedingungen für $\varepsilon > 0$. Diese sichern die Existenz genau eines stabilen Grenzzyklus. Für große x ist $-\varepsilon(1 - x^2) > 0$. Dies verursacht einen Dämpfungseffekt, der die Stabilität des Grenzzyklus vermuten lässt. Die Bifurkation der DGL tritt insofern ein, dass von der Schar konzentrischer Kreise in der Phasenebene ((x, \dot{x}) -Ebene) für $\varepsilon = 0$ nur noch ein Grenzzyklus mit der Amplitude nahe 2 für $\varepsilon > 0$ „übrig bleibt“. Das Phasenporträt der Differenzialgleichung hat sich also bei Durchgang von ε durch den Nullpunkt abrupt geändert. Es ist dabei qualitativ ein völlig anderes Phasenporträt entstanden.

Beispiel 4.21 (Stabilität eines Grenzyklus)

Um die Stabilität eines Grenzyklus zu diskutieren, benötigt man das in Bogoljubov und Mitropolski (1965) bzw. Verhulst (2000) angegebene Kriterium nicht. Man kann stattdessen wie folgt vorgehen. Wir betrachten die Klasse der Differenzialgleichungen der Form

$$\ddot{x} + h(x, \dot{x}) + g(x) = 0 \quad (4.103)$$

mit hinreichend glatten Funktionen h und g . Für den Zweck der Interpretation werden wir weiter annehmen, dass das entsprechende DGL-System

$$\dot{x} = y, \quad \dot{y} = -h(x, y) - g(x) \quad (4.104)$$

einen einzigen Fixpunkt bei $(x, y) = (0, 0)$ besitzt, d. h. die Gleichung $h(x, 0) + g(x) = 0$ besitzt nur die Lösung $x = 0$ und es gelte außerdem $g(0) = 0$. Unter diesen Bedingungen kann (4.103) als Modell für eine Feder (auch für einen elektrischen Schwingkreis) angesehen werden, dessen freie Bewegung durch die Gleichung $\ddot{x} + g(x) = 0$ (konservatives System) beschrieben wird. Die „äußere Kraft $-h(x, \dot{x})$ “ beeinflusst die Bewegung des Systems, indem sie Energie liefert oder absorbiert. Definiert man eine potenzielle Energiefunktion für das Masse-Feder-System durch $V(x, \dot{x}) := \int g(x) dx$ und eine kinetische Energie durch $T(x, \dot{x}) = \frac{1}{2}\dot{x}^2$, dann lautet die Gesamtenergie

$$E(x, \dot{x}) := T(x, \dot{x}) + V(x, \dot{x}) = \frac{1}{2}\dot{x}^2 + \int_0^x g(\xi) d\xi. \quad (4.105)$$

Die zeitliche Ableitung längs Phasenkurven (der Kürze halber mit x statt mit $x(t)$ bezeichnet) liefert

$$\frac{dE(x, \dot{x})}{dt} = \dot{x} \cdot \ddot{x} + g(x) \cdot \dot{x} = \dot{x}(-g(x) - h(x, \dot{x})) + g(x) \cdot \dot{x} = -\dot{x}h(x, \dot{x}). \quad (4.106)$$

Aufgrund dieser Formel lassen sich ohne Kenntnis der Lösung der Gleichung (4.103) Aussagen über das Verhalten von E längs eines Orbits und somit über die Stabilität eines Fixpunktes machen. Darin liegt gerade die Bedeutung dieser Energiefunktion, die man auch *Ljapunov-Funktion* nennt. Hier sei nur so viel gesagt, dass die Ljapunov-Funktion benutzt wird, um die Stabilität eines Fixpunktes in einem dynamischen System zu bestimmen. Leider gibt es keine allgemeine Methode, um eine Ljapunov-Funktion zu konstruieren. Ein guter Kandidat für eine Ljapunov-Funktion sind Erhaltungssätze aus der Physik.

Setzt man in (4.105) $G(x) := \int_0^x g(\xi) d\xi > 0$ voraus ($x \neq 0$), dann ist E eine Ljapunov-Funktion, wenn man noch $\dot{x} \cdot h(x, \dot{x}) > 0$ für $\dot{x} \neq 0$ annimmt. Aus diesen Bedingungen ergibt sich dann die asymptotische Stabilität des Fixpunktes $(x, \dot{x}) = (0, 0) \in \mathbb{R}^2$. Der Leser möge sich mit dem Begriff der Ljapunov-Funktion und der daraus folgenden Stabilität anhand des Buches von Aulbach (1997) beschäftigen.

Für eine nichtlineare Schwingung ohne Reibung mit $h(x, \dot{x}) \equiv 0$ und $x \cdot g(x) > 0$ für $x \neq 0$ ist es naheliegend, die Energiefunktion $E(x, \dot{x}) = \frac{1}{2}\dot{x}^2 + G(x)$ als Ljapunov-Funktion heranzuziehen. Es ist $E(x, \dot{x}) > 0$ für $(x, \dot{x}) \neq (0, 0)$ und $\dot{E}(x(t), \dot{x}(t)) \equiv 0$. Also ist die Nulllösung stabil.

Im Allgemeinen ist es schwierig, isolierte periodische Schwingungen nachzuweisen und sie dann auch noch zu konstruieren. Vor einer weiteren komplizierten Aufgabe steht man dann, wenn auch noch die Stabilität des Grenzzykus untersucht werden soll. Solche Untersuchungen wollen wir nun für die Familie von Gleichungen der Form

$$\ddot{x} + \varepsilon h(x, \dot{x}) + x = 0. \quad (4.107)$$

durchführen. Die Gleichung $\ddot{x} + \varepsilon h(x, \dot{x}) + \omega^2 x = 0$ kann durch einen Variablenwechsel $\tau = \omega t$ in die Form (4.107) (mit einem neuen ε) gebracht werden. Als System lautet (4.107)

$$\dot{x} = y, \quad \dot{y} = -x - \varepsilon h(x, y). \quad (4.108)$$

Wir nehmen an, dass $|\varepsilon| \ll 1$ und $h(0, 0) = 0$ ist, so dass $(0, 0)$ ein Gleichgewichtspunkt ist. Bei der Van-der-Pol-Gleichung ist $h(x, y) := -(1 - x^2)y$ (siehe (4.77)). Die Gleichung (4.107) besitzt für $\varepsilon = 0$ die allgemeine Lösung $x(t) = a \cos(t + \alpha)$, wobei a und α willkürliche Konstanten sind. Soweit das Phasendiagramm der Differentialgleichung betrachtet wird, kann man sich auf den Fall $a > 0$ und $\alpha = 0$ einschränken, da verschiedene Werte von α lediglich zu verschiedenen Zeitanfängen korrespondieren. Die Orbits bleiben unverändert. Die Familie von Orbits der Gleichung

$$\dot{x} = y, \quad \dot{y} = -x \quad (\varepsilon = 0 \text{ in (4.108)}) \quad (4.109)$$

ist in parametrischer Form gegeben durch

$$x(t) = a \cos t, \quad y(t) = -a \sin t, \quad t \in [0, 2\pi]. \quad (4.110)$$

Die Periode all dieser Orbits ist $T = 2\pi$. Für kleines $\varepsilon \neq 0$ haben wir gezeigt, dass sich ein Grenzzyklus einstellt, der in der Nähe einer dieser Lösungen (4.110) liegt und mit $\varepsilon \rightarrow 0$ dagegen konvergiert. Folglich kann man annehmen, dass der Grenzzyklus ungefähr das Aussehen

$$x(t) \approx a \cos t, \quad y(t) \approx -a \sin t \quad \text{und} \quad T \approx 2\pi \quad (4.111)$$

besitzt, wobei T seine Periode ist. Aus (4.106) folgt mit εh anstelle von h durch Integration über eine Periode T für die Energie $E(t)$ ($\dot{x} = y$)

$$E(T) - E(0) = -\varepsilon \int_0^T h(x(t), y(t)) y(t) dt. \quad (4.112)$$

Da der Orbit geschlossen ist, stimmt die Energie im Periodenpunkt T mit $E(0)$ nach einem Umlauf überein. Folglich gilt

$$\int_0^T h(x(t), y(t)) y(t) dt = 0 \quad (4.113)$$

entlang des Grenzyklus. Setzt man die Approximation (4.111) in das Integral ein, ergibt sich

$$\int_0^{2\pi} h(a \cos t, -a \sin t) \sin t dt = 0 \quad (4.114)$$

(nachdem man durch den Faktor $-\varepsilon a$ dividiert hat). Dies ist eine Gleichung, aus der man die unbekannte Amplitude a des Grenzyklus bestimmen kann. Für den Fall der Van-der-Pol-Gleichung (4.77) ergibt sich mit $h(x, y) = -(1 - x^2)y$ die Gleichung

$$\int_0^{2\pi} (1 - a^2 \cos^2 t)(-a \sin t) \sin t dt = 0.$$

Dies führt zu der Bestimmungsgleichung für die Amplitude:

$$1/4a^3\pi - a\pi = 0 \iff 1 - \frac{1}{4}a^2 = 0, \quad (a \neq 0)$$

mit der positiven Lösung $a = 2$. Diese Amplitude kann also für kleine Störungen εh als eine Näherung für die Amplitude des Grenzyklus angesehen werden.

Wir nehmen vorübergehend an, dass die Existenz des Grenzyklus bereits nachgewiesen ist. Dann kann man auch seine Stabilität für kleine ε zeigen. Dazu nehmen wir weiter an, dass die nicht geschlossenen Orbits, die in der Nähe des Grenzyklus liegen, sich auf die periodische Lösung „aufwickeln“. Diese sind dann näherungsweise durch $x \approx a \cos t$, $y \approx -a \sin t$ gegeben, wobei a nahezu konstant über dem Zeitintervall $0 \leq t \leq 2\pi$ (keine Periode) ist. Über diesem Intervall ist die approximierte Energie $E(2\pi) - E(0)$ gegeben durch (4.112) und die Funktion f sei definiert durch

$$f(a) := -\varepsilon a \int_0^{2\pi} h(a \cos t, -a \sin t) \sin t dt, \quad (4.115)$$

mit $a \in K(2; \delta_0)$ ($\delta_0 > 0$). Weiter sei $a_0 \approx a$ die Amplitude für den Grenzyklus, so dass $f(a_0) = 0$ (kein Energieverbrauch!) gilt. Falls der Grenzyklus stabil ist, dann ist in dem inneren Spiralsegment ($a < a_0$) ein Energiegewinn und im äußeren Spiralsegment ($a > a_0$) ein Energieverlust vorhanden. D. h., dass für $\delta > 0$ (hinreichend klein) gilt:

$$\begin{aligned} f(a) &> 0, \quad \text{falls } a_0 - \delta < a < a_0 \\ f(a) &< 0, \quad \text{falls } a_0 < a < a_0 + \delta. \end{aligned} \quad (4.116)$$

Sind die Vorzeichen von $f(a)$ vertauscht, dann ist nach analoger Schlussweise der Grenzyklus instabil. Falls f differenzierbar in einer Umgebung von a_0 ist, dann ist der dazugehörige Grenzyklus stabil, falls $f'(a_0) < 0$ und instabil, falls $f'(a_0) > 0$. Im Fall der Van-der-Pol-Gleichung (4.77) ergibt sich

$$f(a) = \varepsilon a \int_0^{2\pi} (1 - a^2 \cos^2 t) \cdot a \sin t \cdot \sin t dt = \pi \varepsilon a^2 \left(1 - \frac{a^2}{4}\right).$$

Folglich gilt $f'(a) = \pi \varepsilon a(2 - a^2)$ und $f'(2) = -4\pi \varepsilon$, was für $\varepsilon > 0$ die Stabilität und für $\varepsilon < 0$ die Instabilität des Grenzzyklus bewirkt. In Abbildung 4.8 (links) ist ein stabiler Grenzzyklus dargestellt.

Amplituden und Frequenzabschätzungen können auch mittels Polarkoordinatentransformation erfolgen. Wir starten wieder mit der Gleichung (4.107) bzw. dem äquivalenten System (4.108) und wir nehmen an, dass (4.108) mindestens eine periodische Lösung besitzt. Jede Phasenkurve kann durch zeitabhängige Polarkoordinaten $r(t)$ und $\varphi(t)$ parametrisiert werden. Mit $x = r(t) \cos \varphi(t)$ und $y = r(t) \sin \varphi(t)$ geht (4.108) über in

$$\begin{aligned} \dot{r} \cos \varphi - r \dot{\varphi} \sin \varphi &= r \sin \varphi \\ \dot{r} \sin \varphi + r \dot{\varphi} \cos \varphi &= -\varepsilon h(r \cos \varphi, r \sin \varphi) - r \cos \varphi \end{aligned}$$

und liefert ein DGL-System in $\dot{r}, \dot{\varphi}$ der Form

$$\begin{aligned} \begin{pmatrix} \dot{r} \\ \dot{\varphi} \end{pmatrix} &= \begin{pmatrix} \cos \varphi & \sin \varphi \\ -\frac{\sin \varphi}{r} & \frac{\cos \varphi}{r} \end{pmatrix} \begin{pmatrix} r \sin \varphi \\ -\varepsilon h - r \cos \varphi \end{pmatrix} \\ &= \begin{pmatrix} -\varepsilon h \sin \varphi \\ -1 - \frac{\varepsilon}{r} h \cos \varphi \end{pmatrix} \quad (h \text{ steht für } h(r \cos \varphi, r \sin \varphi)). \end{aligned} \tag{4.117}$$

Die Differenzialgleichung für die Phasenkurven lautet folglich

$$\frac{dr}{d\varphi} = \frac{\varepsilon h \sin \varphi}{1 + \frac{\varepsilon}{r} h \cos \varphi}. \tag{4.118}$$

Wir nehmen an, dass ein periodischer Orbit die Periode $T(\varepsilon)$ besitzt. Dann besitzen $r(t)$, $\varphi(t)$ und auch h die Periode $T(\varepsilon)$, d. h. es gilt $r(t_0 + T(\varepsilon)) = r(t_0)$ für alle t_0 und ebenso für alle anderen Größen. In Bezug auf die Koordinate φ sind in (4.118) alle Größen 2π -periodisch. Die Abbildung 4.9 zeigt das Phasendiagramm und die Relation

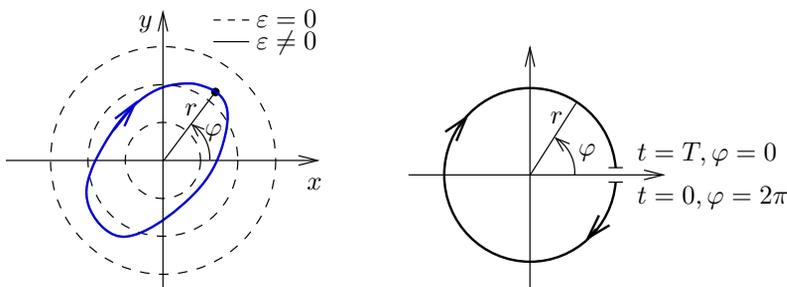


Abb. 4.9 Phasenportrait für $\varepsilon = 0$ und $\varepsilon \neq 0$ (links), Relation zwischen φ und t (rechts)

zwischen φ und t . Ein typischer Grenzzyklus ist folglich durch folgende Randbedingungen beschrieben:

$$\begin{aligned} r &= r_0, \quad \varphi = 2\pi \quad \text{zum Zeitpunkt } t = 0, \\ r &= r_0, \quad \varphi = 0 \quad \text{zum Zeitpunkt } t = T. \end{aligned} \tag{4.119}$$

Wir nehmen jetzt an, dass $|\varepsilon| \ll 1$ gilt und weiter der Grenzzyklus aus einer kleinen Verzerrung, einer Störung einer der Kreise des linearisierten Systems $\ddot{x} + x = 0$ besteht. Wie kann man nun die Periode T und die Kreisfrequenz $\omega = \frac{2\pi}{T}$ angenähert unter Benutzung der Gleichungen (4.117) und (4.118) berechnen? Dazu wird die rechte Seite von (4.118) in Bezug auf ε approximiert. Es ergibt sich

$$\frac{dr}{d\varphi} = \varepsilon h \sin \varphi + O(\varepsilon^2). \quad (4.120)$$

Integriert man (4.120) bezüglich φ , beginnend mit $\varphi = 2\pi$, dann erhält man

$$r(\varphi) - r(2\pi) = O(\varepsilon) \quad \text{oder} \quad r(\varphi) = r_0 + O(\varepsilon), \quad (4.121)$$

da $r(2\pi) = r_0$ gemäß (4.119) gilt. Somit ist die Abweichung vom Kreis vom Radius r_0 von der Ordnung $O(\varepsilon)$. Integriert man (4.120) über das gesamte Zeitintervall des Kreises (4.119) von $\varphi = 2\pi$ bis $\varphi = 0$, dann ergibt sich

$$\begin{aligned} r_0 - r_0 &= \varepsilon \int_{2\pi}^0 h(r \cos \varphi, r \sin \varphi) \sin \varphi \, d\varphi + O(\varepsilon^2) \quad \text{bzw.} \\ 0 &= - \int_0^{2\pi} h(r \cos \varphi, r \sin \varphi) \sin \varphi \, d\varphi + O(\varepsilon) \quad (\varepsilon \neq 0). \end{aligned}$$

Substituiert man r durch (4.121) und entwickelt h an der Stelle r_0 (nullte Näherung), dann ergibt sich

$$\int_0^{2\pi} h(r_0 \cos \varphi, r_0 \sin \varphi) \sin \varphi \, d\varphi = O(\varepsilon).$$

Da diese Beziehung für alle ε (hinreichend klein) gelten muss und die linke Seite der Gleichung nicht von ε abhängt, folgt

$$\int_0^{2\pi} h(r_0 \cos \varphi, r_0 \sin \varphi) \sin \varphi \, d\varphi = 0. \quad (4.122)$$

Diese Beziehung dient zur näherungsweisen Berechnung der Amplitude r_0 des Grenzzyklus. Die entsprechend aussehende Bedingung (4.114) kann aus (4.122) gewonnen werden, indem man die zweite Gleichung in (4.117) nach t integriert und $\varphi = 2\pi - t + O(\varepsilon)$ in (4.122) einsetzt. Der führende Term in (4.122) ist dann genau (4.113).

Um die angenäherte Periodendauer T des Grenzzyklus zu erhalten, geht man wie folgt vor. Für T gilt die Beziehung

$$T = \int_0^T dt = \int_{2\pi}^0 \frac{d\varphi}{\dot{\varphi}} = \int_0^{2\pi} \frac{d\varphi}{1 + \frac{2}{r} h(r \cos \varphi, r \sin \varphi) \cos \varphi},$$

wobei (4.117) benutzt wurde. Ersetzt man wieder r durch $r_0 + O(\varepsilon)$, entwickelt h in Bezug auf r_0 in eine Taylor-Reihe, entwickelt $1/(1 + \varepsilon r_0^{-1} h \cos \varphi)$ mittels geometrischer Reihe und verwendet nur die Glieder bis zur Ordnung $O(\varepsilon)$, dann ergibt sich für T der Ausdruck

$$\begin{aligned} T &= \int_0^{2\pi} (1 - \varepsilon r_0^{-1} h(r_0 \cos \varphi, r_0 \sin \varphi) \cos \varphi + O(\varepsilon^2)) d\varphi \\ &\approx 2\pi - \frac{\varepsilon}{r_0} \int_0^{2\pi} h(r_0 \cos \varphi, r_0 \sin \varphi) \cos \varphi d\varphi + O(\varepsilon^2) \end{aligned} \quad (4.123)$$

mit einem Fehler der Ordnung $O(\varepsilon^2)$. Daraus ergibt sich die Kreisfrequenz

$$\omega = \frac{2\pi}{T} \approx 1 + \frac{\varepsilon}{2\pi r_0} \int_0^{2\pi} h(r_0 \cos \varphi, r_0 \sin \varphi) \cos \varphi d\varphi + O(\varepsilon^2) \quad (4.124)$$

wiederum mit einem Fehler von $O(\varepsilon^2)$. ■

Die in diesem Abschnitt allgemein beschriebene Methode wird nun auf die Van-der-Pol-Gleichung angewendet.

Beispiel 4.22 (Stabilität des Grenzyklus der Van-der-Pol-DGL)

Wir betrachten die Van-der-Pol-Gleichung (4.77) als DGL-System

$$\dot{x} = y, \quad \dot{y} = -x + \varepsilon(1 - x^2)y \quad (4.125)$$

und leiten mit der (zeitabhängigen) Polarkoordinatentransformation

$$x = r(t) \cos \varphi(t), \quad y = r(t) \sin \varphi(t)$$

eine äquivalente Gleichung her. Dies ist bereits allgemein für die Gleichung (4.107) getan. Dort ist $(\dot{x} = y)$

$$h(x, \dot{x}) = h(x, y) = -(1 - x^2)y \quad \text{und} \quad h(r \cos \varphi, r \sin \varphi) = -(1 - r^2 \cos^2 \varphi) r \sin \varphi.$$

Für die Frequenz des Grenzyklus der Van-der-Pol-Gleichung mit $h(x, \dot{x}) := -(1 - x^2)\dot{x}$ und der Amplitude $r_0 = 2$ ergibt sich aus (4.124) der Wert

$$\begin{aligned} \omega &= 1 + \frac{\varepsilon}{4\pi} \int_0^{2\pi} (1 - 4 \cos^2 \varphi)(-2 \sin \varphi) \cos \varphi d\varphi + O(\varepsilon^2) \\ &= 1 + 0 + O(\varepsilon^2). \end{aligned}$$

Der Integrand ist eine ungerade Funktion in Bezug auf $\varphi = \pi$. Somit ist die Frequenz gleich 1 und mit einem Fehler von $O(\varepsilon^2)$ behaftet. ■

Die hier dargestellte Methode erweist sich als durchaus praktikabel und programmierfähig. Falls gewünscht, ist mit dieser Herangehensweise eine noch bessere Approximation von Amplitude und Schwingungsdauer möglich. Es sei jedoch noch einmal bei dieser Methode kritisch angemerkt, dass die Existenz eines Grenzzyklus mit einer anderen Methode (z. B. Ljapunov-Schmidt-Methode oder Poincaré-Methode) gezeigt werden muss. Erst dann macht es Sinn, Amplitude und Frequenz des Grenzzyklus näherungsweise nach den vorgestellten Methoden zu berechnen.

Das Lösungsverhalten einer DGL in der Nähe einer periodischen Lösung kann mit der *Poincaré-Abbildung*⁴ untersucht werden. Sie kann als Nachweis für die Existenz eines Grenzzyklus herangezogen werden. Hierbei ergibt sich eine schöne Anwendung des Satzes 1.80 über implizite Funktionen.

Beispiel 4.23 (Poincaré-Abbildung)

Wir nehmen an, dass wir eine nichtkonstante T -periodische Lösung φ_0 (Bezeichnung: $\varphi_0(t) = \varphi(t; x_0)$) des autonomen DGL-Systems

$$\dot{x} = f(x), \quad x(0) = \xi \quad (4.126)$$

kennen. Wir setzen voraus, dass $f \in \mathcal{C}^1(G)$, $G \subset \mathbb{R}^n$, G offen und bezeichnen mit $\varphi(\cdot; \xi)$ die Lösung des Anfangswertproblems von (4.126), die $\varphi(0; \xi) = \xi$ erfüllt. Wir interessieren uns für das Verhalten weiterer Lösungen, die zur Zeit $t = 0$ nahe $\varphi_0(0) = x_0 = \varphi_0(T)$ starten. Es sei

$$\Sigma := \{x \in \mathbb{R}^n \mid \langle x - x_0, f(x_0) \rangle_{\mathbb{R}^n} = 0\}$$

und $\langle \cdot, \cdot \rangle_{\mathbb{R}^n}$ bezeichnet das Skalarprodukt im \mathbb{R}^n . Um zu zeigen, dass eine Lösung $\varphi(\cdot; \xi)$ auf einem solchen Intervall $[0, t(\xi)]$ existiert, d. h. $\Sigma \cap \mathcal{U}$ (\mathcal{U} ist eine Umgebung von x_0) nach einer positiven Zeit $t(\xi)$ ($t(\xi)$ nahe der Periode T von φ_0) erreicht wird (siehe Abbildung 4.10), benutzt man den Auflösungssatz. Wir lösen dazu die Gleichung

$$\Phi(t, \xi) := \langle \varphi(t; \xi) - x_0, f(x_0) \rangle = 0 \quad (4.127)$$

in der Nähe des Punktes (T, x_0) . Der Punkt (T, x_0) ist ein Lösungselement von (4.127), denn es gilt $\langle \varphi(T; x_0) - x_0, f(x_0) \rangle = \langle \varphi_0(T) - x_0, f(x_0) \rangle = 0$. Weiter folgt für die Ableitung

$$\Phi_t(T, x_0) = \langle f(x_0), f(x_0) \rangle > 0 \quad (f(x_0) \neq 0, \text{ da } \varphi_0 \text{ nicht konstant}).$$

Damit sind alle Voraussetzungen des Auflösungssatzes erfüllt und im Ergebnis erhalten wir eine Funktion $\xi \mapsto t(\xi)$, so dass

$$\Phi(t(\xi), \xi) = 0 \quad \text{für } \xi \in \mathcal{U}(x_0) \text{ und } t(x_0) = T.$$

⁴Jules Henri Poincaré (1854–1912), bedeutender französischer Mathematiker, theoretischer Physiker und Philosoph. Seine Forschungen hatten eine starke Wirkung auf die Astronomie, die Geodäsie, die Potentialtheorie und die Quantenphysik.

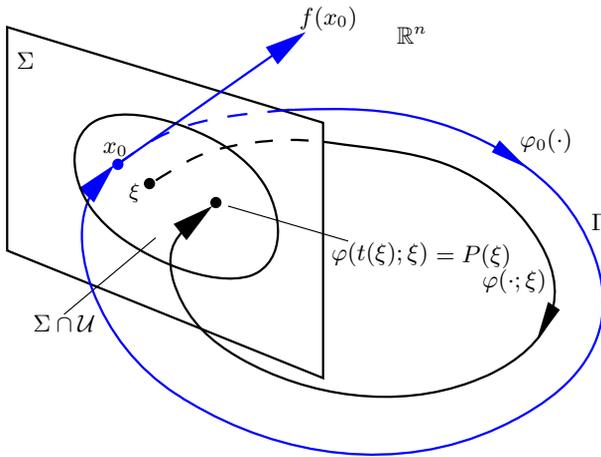


Abb. 4.10 Poincaré-Abbildung

Die Ableitung von Φ nach ξ lautet:

$$\Phi_{\xi}(t, x_0)\eta = \left\langle \frac{\partial \varphi}{\partial \xi}(t; x_0)\eta, f(x_0) \right\rangle = \langle \Xi(t, x_0)\eta, f(x_0) \rangle, \quad \eta \in \mathbb{R}^n.$$

Hierbei ist $\Xi(t, x_0)$ die Fundamentalmatrix des linearen T -periodischen DGL-Systems $\dot{y}(t) = f'(\varphi_0(t))y(t)$. Dies sieht man so: Es gilt $\dot{\varphi}(t; \xi) = f(\varphi(t; \xi))$. Differentiation nach ξ an der Stelle $(t; x_0)$ liefert:

$$\frac{\partial}{\partial \xi} \dot{\varphi}(t; x_0) = f'(\varphi(t; x_0)) \cdot \frac{\partial \varphi}{\partial \xi}(t; x_0) \Leftrightarrow \frac{d}{dt} \frac{\partial \varphi}{\partial \xi}(t; x_0) = f'(\varphi(t; x_0)) \cdot \frac{\partial \varphi}{\partial \xi}(t; x_0).$$

Die Funktion $\Xi(\cdot, x_0) := \frac{\partial \varphi}{\partial \xi}(\cdot; x_0)$ löst die so genannte Variationsgleichung

$$\dot{y}(t) = f'(\varphi(t; x_0))y(t).$$

Daraus folgt dann letztlich zusammen mit (1.59) die Formel

$$\begin{aligned} t'(x_0)\eta &= -\Phi_t(T, x_0)^{-1} \circ \Phi_{\xi}(T, x_0)\eta \\ &= -\frac{1}{\|f(x_0)\|^2} \langle \Xi(T, x_0)\eta, f(x_0) \rangle, \quad \eta \in \mathbb{R}^n. \end{aligned}$$

All dies erlaubt es uns, das Verhalten der so genannten Poincaré-Abbildung

$$P(\xi) := \varphi(t(\xi); \xi), \quad \xi \in \mathcal{U} \cap \Sigma$$

zu untersuchen. ■

Beispiel 4.24 (Nachweis eines Grenzyklus mittels Poincaré-Abbildung)

Am Beispiel der Van-der-Pol-Gleichung soll nun gezeigt werden, wie mittels Poincaré-Abbildung (siehe Beispiel 4.23) eine periodische Lösung nachgewiesen und deren Stabilität untersucht werden kann. Es sei Γ ein periodischer Orbit des Systems

$$\dot{x} = f(x), \quad x(0) = \xi, \tag{4.128}$$

der durch den Punkt x_0 geht und Σ eine Hyperebene, die senkrecht zu Γ im Punkt x_0 verläuft. Die Lösung von (4.128) durch den Punkt ξ zum Zeitpunkt $t = 0$ wird mit $\varphi_t(\xi) = \varphi(t; \xi)$ bezeichnet und Fluss der DGL (4.128) genannt. Man startet nun mit diesem Punkt $\xi \in \Sigma$, der nahe bei x_0 liegt und gelangt nach einer Zeit $\tau(\xi)$ wieder nach Σ zurück. Dieser Punkt wird mit $P(\xi)$ bezeichnet und die Abbildung $P : \Sigma \rightarrow \Sigma$, $\xi \mapsto P(\xi)$ wird Poincaré-Abbildung genannt (siehe Abbildung 4.10). Eigenschaften der Poincaré-Abbildung wurden bereits in Beispiel 4.23 untersucht.

Wir betrachten die nichterregte, schwach gedämpfte Van-der-Pol-Gleichung (4.77). Die ungestörte Gleichung $\ddot{x} + x = 0$ ist explizit lösbar. Die Lösungen mit dem Anfangszustand (x_0, \dot{x}_0) sind durch

$$t \mapsto x_0 \cos t + \dot{x}_0 \sin t$$

gegeben. Die Situation, die für $\varepsilon \neq 0$ bzw. $\varepsilon = 0$ eintritt, zeigt Abbildung 4.8. Wir untersuchen nun, ob die Differenzialgleichung (4.77) eine periodische Lösung besitzt und was über ihre (orbitale) Stabilität ausgesagt werden kann. Die Gewinnung des DGL-Systems wird mit der Einführung der neuen Variablen $x = x_1$, $\dot{x} = -x_2$ vorgenommen. Dies sichert einen Durchlaufsinne der Phasenkurven im mathematisch positiven Sinn und liefert das Van-der-Pol-System (es wird nun wieder $x_1 = x$ und $x_2 = y$ gesetzt)

$$\dot{x} = -y, \quad \dot{y} = x + \varepsilon(1 - x^2)y. \quad (4.129)$$

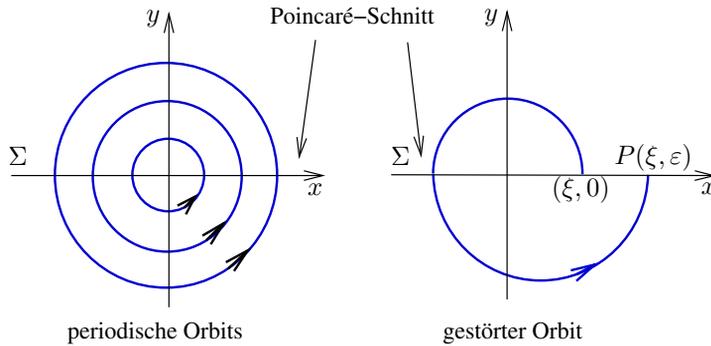
Belässt man ε hinreichend klein, dann bleiben die Orbits des Systems (4.129) mit den Anfangsbedingungen $(x(0), y(0)) = (\xi, 0)$ in der Nähe des Kreises mit Radius ξ mindestens bis zur Rückkehr zur x -Achse. Präziser gilt: Falls

$$t \mapsto \varphi_t(\xi, \varepsilon) := (x(t; \xi, \varepsilon), y(t; \xi, \varepsilon))^T, \quad \xi > 0, \quad \varepsilon \neq 0 \quad (4.130)$$

eine Lösung des Systems (4.129) mit den Anfangsbedingungen $x(0; \xi, \varepsilon) = \xi$ und $y(0; \xi, \varepsilon) = 0$ ist, dann wird diese Lösung zur positiven x -Achse mit dem Schnittpunkt $x(T(\xi, \varepsilon); \xi, \varepsilon)$ zurückkehren. Der Nachweis, dass eine solche Rückkehrzeit $T(\xi, \varepsilon)$ existiert, wurde in Beispiel 4.23 erbracht. Die Funktion $(\xi, \varepsilon) \mapsto P(\xi, \varepsilon)$ mit $P(\xi, \varepsilon) := x(T(\xi, \varepsilon); \xi, \varepsilon)$ heißt auch parametrisierte Rückkehrabbildung. Die Situation wird durch die Abbildung 4.11 illustriert.

Jeder Fixpunkt der Poincaré-Abbildung (bei festem $\varepsilon \neq 0$) liefert eine periodische Lösung des Systems (4.129) mit der Lösung (4.130) und der Periodenzeit $T(\xi, \varepsilon)$. Falls also ξ ein Fixpunkt der Abbildung $\xi \mapsto P(\xi, \varepsilon)$ bzw. eine Nullstelle der dazugehörigen Abstandsfunktion

$$\delta(\xi, \varepsilon) := x(T(\xi, \varepsilon); \xi, \varepsilon) - \xi \quad (4.131)$$

**Abb. 4.11** Poincaré-Schnitt

ist, dann ist $(\xi, 0)$ der Anfangspunkt für eine periodische Lösung des gestörten Systems (4.129). Nun ist klar, dass alles auf das Lösen der Gleichung $\delta(\xi, \varepsilon) = 0$ hinausläuft und man Lösungen $\xi = \xi^*(\varepsilon)$ sucht. Dazu nutzt man Eigenschaften der Funktion $\delta : U \times V \rightarrow \mathbb{R}$, U und V sind offene Teilmengen von \mathbb{R} , aus. Wegen $\delta \in \mathcal{C}^k(U \times V)$ und $\delta(\xi, 0) \equiv 0$ ist das implizite Funktionentheorem nicht unmittelbar auf die Gleichung $\delta(\xi, \varepsilon) = 0$ anwendbar, da $\delta_\xi(\xi, 0) \equiv 0$ und das besagte Theorem gerade $\delta_\xi(\xi, 0) \neq 0$ fordert. Mit einem kleinen Trick wird das implizite Funktionentheorem jedoch wieder anwendbar. Dazu wird eine Taylor-Reihenentwicklung im Punkt $\varepsilon = 0$ durchgeführt. Sie liefert

$$\delta(\xi, \varepsilon) = \varepsilon \delta_\varepsilon(\xi, 0) + O(\varepsilon^2), \quad (4.132)$$

wobei $O(\varepsilon^2)$ das Restglied bezeichnet. Führt man $\Delta(\xi, \varepsilon) := \delta_\varepsilon(\xi, 0) + O(\varepsilon)$ ein, dann schreibt sich (4.132) in der Form $\delta(\xi, \varepsilon) = \varepsilon \Delta(\xi, \varepsilon)$. Nun wird versucht, das implizite Funktionentheorem auf die Gleichung $\Delta(\xi, \varepsilon) = 0$ anzuwenden, da jede Lösung $\xi^*(\varepsilon)$ mit $\Delta(\xi^*(\varepsilon), \varepsilon) = 0$ auch eine Lösung der Gleichung $\delta(\xi, \varepsilon) = 0$ darstellt. Damit ist klar, welche Bedingungen $\Delta(\xi, \varepsilon)$ erfüllen muss: $\Delta(\xi, 0) = 0$ und $\Delta_\xi(\xi, 0) \neq 0$, oder in Bezug auf δ muss $\delta_\varepsilon(\xi, 0) = 0$ und $\delta_{\xi\varepsilon}(\xi, 0) \neq 0$ gelten. Man sagt auch, dass unter den genannten Bedingungen $\xi > 0$ eine einfache Nullstelle der Funktion $\xi \mapsto \Delta(\xi, 0)$ ist. Um das an der Van-der-Pol-Gleichung nachzuweisen, kann man wieder den Übergang zu Polarkoordinaten durchführen und dann die Phasen-DGL (4.118) diskutieren. Hier wird ein Weg eingeschlagen, der die so genannte Variationsgleichung benutzt. Zunächst muss mittels Kettenregel die Ableitung $\delta_\varepsilon(\xi, 0)$ berechnet werden:

$$\delta_\varepsilon(\xi, 0) = \dot{x}(T(\xi, 0); \xi, 0) T_\varepsilon(\xi, 0) + x_\varepsilon(T(\xi, 0); \xi, 0).$$

Wegen $\dot{x}(T(\xi, 0); \xi, 0) = -y(0, \xi, 0) = 0$ verschwinden alle Ableitungen der Funktion $\xi \mapsto \dot{x}(T(\xi, 0); \xi, 0)$ (in Bezug auf ξ). Folglich genügt es, die partielle Ableitung von $x_\varepsilon(T(\xi, 0); \xi, 0)$ zu ermitteln. Dazu wird das System (4.129) in Bezug auf ε (an der

Stelle $\varepsilon = 0$) nebst Anfangsbedingungen differenziert. Es ergibt sich das System (auch Variationsanfangswertproblem genannt):

$$\dot{x}_\varepsilon = -y_\varepsilon, \quad \dot{y}_\varepsilon = x_\varepsilon + (1 - x_\varepsilon^2)y, \quad x_\varepsilon(0; \xi, 0) = 0, \quad y_\varepsilon(0; \xi, 0) = 0. \quad (4.133)$$

Die Lösung dieses Anfangswertproblems lautet

$$t \mapsto (x_\varepsilon(t; \xi, 0), y_\varepsilon(t; \xi, 0))^T.$$

In der üblichen Matrixschreibweise stellt sich (4.133) in der Form

$$\begin{aligned} \dot{w} &= Aw + b(t), \quad w(0) = 0 \quad (w := (x_\varepsilon, y_\varepsilon)^T) \\ A &:= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad b(t) := \begin{pmatrix} 0 \\ (1 - x^2(t; \xi, 0))y(t; \xi, 0) \end{pmatrix} \end{aligned} \quad (4.134)$$

dar. Bekanntermaßen ergibt sich für (4.134) die Lösungsdarstellung mittels der *Formel der Variation der Konstanten*

$$w(t) = \Phi(t)w(0) + \Phi(t) \int_0^t \Phi^{-1}(s)b(s) \, ds \quad (4.135)$$

mit der Fundamentalmatrix

$$\begin{aligned} \Phi(t) &= e^{tA} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}, \quad T(\xi, 0) = 2\pi \quad \text{und} \\ x(t; \xi, 0) &= \xi \cos t, \quad y(t; \xi, 0) = \xi \sin t. \end{aligned}$$

Setzt man dies in (4.135) ein, so erhält man

$$\begin{aligned} \begin{pmatrix} x_\varepsilon(2\pi; \xi, 0) \\ y_\varepsilon(2\pi; \xi, 0) \end{pmatrix} &= \Phi(2\pi) \int_0^{2\pi} \Phi^{-1}(s) b(s) \, ds \\ &= \begin{pmatrix} \int_0^{2\pi} \sin s [(1 - \xi^2 \cos^2 s) \xi \sin s] \, ds \\ \int_0^{2\pi} \cos s [(1 - \xi^2 \cos^2 s) \xi \sin s] \, ds \end{pmatrix}. \end{aligned}$$

Nach der Integration entsteht

$$\delta_\varepsilon(\xi, 0) = \frac{\pi}{4} \xi (4 - \xi^2), \quad \xi > 0 \quad (4.136)$$

und folglich $\xi = 2$ als einfache Nullstelle der Funktion $\xi \mapsto \delta_\varepsilon(\xi, 0)$, da $\delta_{\varepsilon\xi}(2, 0) = -2\pi \neq 0$ gilt. Folglich bleibt der ungestörte periodische Orbit mit Radius 2 bestehen und da $\xi = 2$ einzige Nullstelle ist, werden alle anderen periodischen Orbits des ungestörten Systems

durch die ε -Störung „zerstört“. Insbesondere gibt es genau eine Funktion $\varepsilon \mapsto \xi^*(\varepsilon)$, die in einer Umgebung von $\varepsilon = 0$ definiert ist, für die $\xi^*(0) = 2$ gilt und für jedes ε (ε hinreichend klein) das Van-der-Pol-System (4.129) eine periodische Lösung mit den Anfangsbedingungen $(x(0), y(0)) = (\xi^*(\varepsilon), 0)$ besitzt.

Für die Stabilitätsuntersuchung einer (festen) periodischen Lösung Γ_ε , $\varepsilon \neq 0$ fixiert, gehen wir wie folgt vor. Wir wissen, dass bei gewähltem Startpunkt $(\xi^*(\varepsilon), 0)$ ein periodischer Orbit Γ_ε vorliegt. Für diesen ist

$$\delta(\xi^*(\varepsilon), \varepsilon) = P(\xi^*(\varepsilon), \varepsilon) - \xi^*(\varepsilon) = 0.$$

Um die Stabilität dieses Grenzzyklus nachzuweisen, starten wir mit einem Anfangspunkt $(\xi, 0)$, der nahe bei $(\xi^*(\varepsilon), 0)$ liegt (siehe Abbildung 4.12). Die Stabilität von Γ_ε bestimmt

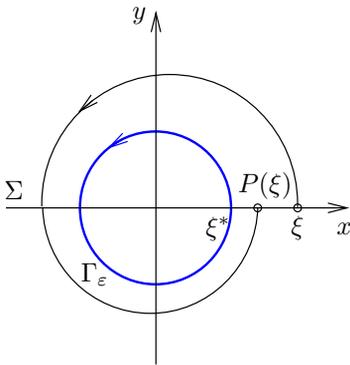


Abb. 4.12 Stabilitätsnachweis

sich allein aus der Kenntnis von $P_\xi(\xi^*, \varepsilon)$. Dies sieht man so. Unsere Abstandsfunktion $\delta(\xi, \varepsilon) = P(\xi, \varepsilon) - \xi$ hat folgende Eigenschaften: Es ist $\delta(\xi^*, \varepsilon) = 0$, δ ist nach ξ differenzierbar und es gilt $\delta_\xi(\xi^*, \varepsilon) = P_\xi(\xi^*, \varepsilon) - 1$. Nach dem Mittelwertsatz der Differentialrechnung gilt für $|\xi - \xi^*| < r_1$, $r_1 > 0$, die Beziehung

$$\delta(\xi, \varepsilon) - \delta(\xi^*, \varepsilon) = \delta_\xi(\xi^* + \vartheta(\xi - \xi^*), \varepsilon)(\xi - \xi^*), \quad \vartheta \in (0, 1). \tag{4.137}$$

Da $\delta_\xi(\xi, \varepsilon)$ stetig ist, besitzt $\delta_\xi(\xi, \varepsilon)$ dasselbe Vorzeichen wie $\delta_\xi(\xi^*, \varepsilon)$, sofern $\delta_\xi(\xi^*, \varepsilon) \neq 0$ gilt. Beachtet man noch, dass $\delta(\xi^*, \varepsilon) = 0$ ist, dann kann man folgende Fallunterscheidung durchführen: Falls $\delta_\xi(\xi^*, \varepsilon) < 0$, dann folgt auch $\delta_\xi(\xi^* + \vartheta(\xi - \xi^*), \varepsilon) < 0$ und aus (4.137) folgt für $\xi - \xi^* < 0$, dass $\delta(\xi, \varepsilon) > 0$ ist. D. h., es ist auch $P(\xi, \varepsilon) - \xi > 0$ bzw. $P(\xi, \varepsilon) > \xi$. Hierbei ist $\delta_\xi(\xi^*, \varepsilon) < 0$ gleichbedeutend mit $P_\xi(\xi^*, \varepsilon) < 1$. Für $\xi - \xi^* > 0$ ist $\delta(\xi, \varepsilon) < 0$. Daraus folgt $P(\xi, \varepsilon) < \xi$ (siehe Abbildung 4.12). Wertet man dieses Ergebnis aus, dann nähert man sich nach jedem Umlauf (unabhängig davon ob man links oder rechts von $(\xi^*, 0)$ startet) der periodischen Lösung Γ_ε an. Da sich die Phasenkurven im Phasenraum nicht schneiden, läuft man auf die periodische Lösung Γ_ε ein. Es ist also Γ_ε unter der Voraussetzung $P_\xi(\xi^*, \varepsilon) < 1$ asymptotisch stabil. In analoger Vorgehensweise ergibt sich für $P_\xi(\xi^*, \varepsilon) > 1$, dass Γ_ε instabil ist.

Ob $P_\xi(\xi^*, \varepsilon)$ größer bzw. kleiner als eins ist, kann man auch über die folgende Beziehung (ohne Heranziehung des Mittelwertsatzes der Differentialrechnung) gewinnen. Man geht von (4.132) aus, differenziert nach ξ und beachtet, dass $\delta_\xi(\xi, \varepsilon) = P_\xi(\xi, \varepsilon) - 1$ gilt. Es ergibt sich dann die Beziehung

$$\delta_\xi(\xi, \varepsilon) = P_\xi(\xi, \varepsilon) - 1 = \varepsilon(\delta_{\varepsilon\xi}(\xi, 0) + O(\varepsilon)). \quad (4.138)$$

Für die Van-der-Pol-Gleichung kann man nun leicht mittels (4.138) die Stabilität des Grenzyklus entscheiden. Aus (4.136) folgt

$$\delta_{\varepsilon\xi}(\xi, 0) |_{\xi=2} = -2\pi < 0,$$

womit das Vorzeichen der runden Klammer in (4.138) festliegt. Somit ist für $\varepsilon < 0$ der Wert $P_\xi > 1$ und für $\varepsilon > 0$ der Wert $P_\xi < 1$, was die Instabilität bzw. die Stabilität des Grenzyklus beweist. ■

Wir bringen im nächsten Beispiel eine Anwendung für den Satz 4.16.

Beispiel 4.25

Gegeben sei eine stetige 2π -periodische Funktion h . Gesucht sind 2π -periodische Lösungen der Gleichung

$$u'' + \lambda u + h(t)u^2 = 0. \quad (4.139)$$

Der Nachweis erfolgt mit dem Satz 4.16. Dazu setzen wir $X := \mathcal{C}_{2\pi}^2$, $Y := \mathcal{C}_{2\pi}$, wobei $\mathcal{C}_{2\pi}^k$ (bzw. $\mathcal{C}_{2\pi}^0$) den Raum der 2π -periodischen C^k -Funktionen (bzw. 2π -periodischen stetigen Funktionen) bezeichnet. Damit lässt sich (4.139) in Form einer Operatorgleichung

$$F : X \times \mathbb{R} \rightarrow Y, \quad F(u, \lambda) = u'' + \lambda u + hu^2 = 0 \quad (4.140)$$

schreiben. Für die Operatoren T, M und B aus Satz 4.16 ergeben sich

$$\begin{aligned} Tz^* &= (z^*)'' + \lambda z^*, & Mz^* &= F_{u\lambda}(0, \mu_0)z^* = z^* \quad \text{und} \\ B[u, v] &= F_{uu}(0, \mu_0)[u, v] = 2huv. \end{aligned}$$

T hat die Eigenwerte $\mu_0 = \lambda_k = k^2$. Der Kern von T wird durch die Eigenfunktionen $\{\cos(kt), \sin(kt)\}$ aufgespannt. Der Wertevorrat $R(T)$ ist L^2 -orthogonal zu $\ker(T)$. Der Projektor

$$Q : Y \rightarrow \ker(T), \quad Qy(t) := \langle y, \cos k(\cdot) \rangle \cos(kt) + \langle y, \sin k(\cdot) \rangle \sin(kt)$$

splittet Y in die direkte Summe

$$Y = (I - Q)(Y) \oplus Q(Y) = R(T) \oplus \ker(T)$$

auf. Nun müssen die Bedingungen (a) und (b) aus Satz 4.16 überprüft werden. Mit $z^* := A \cdot \cos(kt) + B \cdot \sin(kt)$ und

$$QMz^* + \frac{1}{2}QB[z^*, z^*] = Qz^* + \frac{1}{2}Q(2hz^*z^*) = 0$$

ergibt sich die Gleichung

$$\begin{aligned} & \left(\frac{1}{\pi} \int_0^{2\pi} (A \cos(k\tau) + B \sin(k\tau)) \cos(k\tau) \, d\tau \right) \cdot \cos(kt) \\ & + \left(\frac{1}{\pi} \int_0^{2\pi} (A \cos(k\tau) + B \sin(k\tau)) \sin(k\tau) \, d\tau \right) \cdot \sin(kt) \\ & + \left(\frac{1}{\pi} \int_0^{2\pi} h(\tau)[A \cos(k\tau) + B \sin(k\tau)]^2 \cos(k\tau) \, d\tau \right) \cdot \cos(kt) \\ & + \left(\frac{1}{\pi} \int_0^{2\pi} h(\tau)[A \cos(k\tau) + B \sin(k\tau)]^2 \sin(k\tau) \, d\tau \right) \cdot \sin(kt) = 0. \end{aligned}$$

Nutzt man die Beziehungen

$$\int_0^{2\pi} \cos^2(k\tau) \, d\tau = \int_0^{2\pi} \sin^2(k\tau) \, d\tau = \pi, \quad \int_0^{2\pi} \sin(k\tau) \cos(k\tau) \, d\tau = 0,$$

dann ergibt sich ein nichtlineares Gleichungssystem der Form

$$\begin{aligned} A + \frac{1}{\pi} \int_0^{2\pi} h(\tau)[A \cos(k\tau) + B \sin(k\tau)]^2 \cos(k\tau) \, d\tau &= 0 \\ B + \frac{1}{\pi} \int_0^{2\pi} h(\tau)[A \cos(k\tau) + B \sin(k\tau)]^2 \sin(k\tau) \, d\tau &= 0. \end{aligned}$$

Dieses System hat die Form

$$\left. \begin{aligned} A + \mathbf{P}(A, B) &= 0, \\ B + \mathbf{Q}(A, B) &= 0, \end{aligned} \right\} \quad (4.141)$$

wobei \mathbf{P}, \mathbf{Q} homogene Polynome vom Grad zwei sind und deren Koeffizienten von h abhängen. Vom geometrischen Standpunkt kann man sich die Lösungen als transversalen Schnitt zweier Kegelschnitte durch den Ursprung vorstellen. Für ein „generisches“ h besitzt (4.141) eine nichttriviale Lösung $(A^*, B^*) \in \mathbb{R}^2$. Wegen $Mv = v$ folgt aus Bedingung b. (Satz 4.16) zunächst $QM(v) = Qv = v$, da $v \in \ker(T)$. Damit bekommt S die Gestalt

$$\begin{aligned} Sv &= v + QB[z^*, v] = v + Q(2hz^*v) \\ &= v(\cdot) + \left(\frac{2}{\pi} \int_0^{2\pi} h(\tau)[A \cos(k\tau) + B \sin(k\tau)]v(\tau) \, d\tau \right) \cos k(\cdot) \\ & \quad + \left(\frac{2}{\pi} \int_0^{2\pi} h(\tau)[A \cos(k\tau) + B \sin(k\tau)]v(\tau) \, d\tau \right) \sin k(\cdot). \end{aligned} \quad (4.142)$$

Nun muss überprüft werden, unter welchen Bedingungen an A und B die Gleichung $Sv = r$ genau eine Lösung hat. Die (eindeutige) Lösung v hat notwendigerweise das Aussehen

$$v(t) = -\xi \cos(kt) - \eta \sin(kt) + r(t),$$

wobei die ξ und η für die „runden Klammern“ in (4.142) stehen. Wir können schlussfolgern, dass für „gewisse“ $h \in Y$, jedes $\lambda_k = k^2$ für $k = 1, 2, \dots$ eine Bifurkation für $F = 0$ auslöst. Jeder Bifurkationszweig (siehe (4.55)) gibt Anlass zu einer Familie von 2π -periodischen Lösungen der Gleichung (4.139). ■

Beispiel 4.26 (Kleine Schwingungen in DGL 2. Ordnung)

Wir diskutieren ein Problem, welches auf *E. Hopf* (Hopf (1942)) zurückgeht und auf das wir im Unterabschnitt 4.3.2 noch einmal zurückkommen werden. Wir betrachten das DGL-System zweiter Ordnung vom Typ

$$\frac{d^2 u}{dt^2} = f(u), \quad u \in \mathbb{R}^n, \quad f: \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad f(0) = 0. \quad (\text{S})$$

Die Voraussetzungen an f werden später noch präzisiert. Wegen $f(0) = 0$ hat (S) die triviale Lösung $u \equiv 0$. Wir suchen periodische Lösungen von (S) „nahe“ $u \equiv 0$. Das Problem besteht jedoch darin, dass man nicht weiß, wie groß die Periode der gesuchten Lösung ist. Andererseits möchte man das Setup aber in Räumen periodischer Funktionen (mit konstanter und nicht wechselnder Periode) formulieren. Deshalb verwendet man den folgenden Trick. Man skaliert die Zeit: $\tau = \omega t$. Es ist dann $u(t) = u(\tau/\omega) =: \tilde{u}(\tau) = \tilde{u}(\omega t)$ und (S) geht über in

$$\omega^2 \frac{d^2 \tilde{u}}{d\tau^2} = f(\tilde{u}). \quad (\text{S}_\omega)$$

Die Gleichung (S_ω) stellt eine Hilfsgleichung zur Lösung des Problems (S) dar. Falls für ein $\omega > 0$ die Lösung $\tilde{u}(\tau)$ von (S_ω) 2π -periodisch ist, dann ist

$$\tilde{u}(\tau) = \tilde{u}(\tau + 2\pi) = u\left(\frac{\tau}{\omega} + \frac{2\pi}{\omega}\right) = u\left(t + \frac{2\pi}{\omega}\right) = u(t), \quad \forall t \in \mathbb{R} \quad (4.143)$$

und somit u eine T -periodische Lösung von (S) mit $T = 2\pi/\omega$.

Das Problem der Existenz von kleinen Schwingungen nahe $\tilde{u} \equiv 0$ kann nun präzisiert werden. Es sei $\omega^* > 0$ ein Wert mit folgenden Eigenschaften:

- a. Es gibt eine Folge (ω_n) mit $\omega_n \rightarrow \omega^*$.
- b. Es gibt eine Folge (\tilde{u}_n) von 2π -periodischen Lösungen von (S_{ω_n}), so dass $\|\tilde{u}_n\|_\infty \downarrow 0$.

Nach Definition ist dann $u_n(t) := \tilde{u}_n(\omega_n t)$ und (u_n) ist eine Folge von $(2\pi/\omega_n)$ -periodischen Lösungen von (S) mit $\|u_n\|_\infty \downarrow 0$.

Falls $\omega^* > 0$ die Bedingungen a. und b. erfüllt, dann sprechen wir davon, dass (S) kleine Oszillationen in Bezug auf ω^* besitzt. Damit passt das Problem (S_ω) in den Rahmen der Bifurkationstheorie, die Frequenz ω spielt die Rolle eines Parameters.

Wir suchen Lösungen (u, ω) (die Tilde über u lassen wir weg!) der Gleichung

$$F(u, \omega) := \omega^2 \frac{d^2 u}{d\tau^2} - f(u) = 0, \quad f(u)(\tau) := f(u(\tau)). \quad (4.144)$$

Diese Lösungen korrespondieren mit Lösungen von (S_ω) (u aus einem Raum glatter 2π -periodischer Funktionen). Gemäß Voraussetzung gilt $F(0, \omega) \equiv 0$ und falls $\omega^* > 0$ ein Bifurkationspunkt von $F = 0$ ist, dann existiert eine Folge $(u_n, \omega_n) \rightarrow (0, \omega^*)$, so dass $u_n \neq 0$ und $F(u_n, \omega_n) = 0$, für alle $n \in \mathbb{N}$ gilt. Dies bedeutet gerade, dass (a) und (b) erfüllt sind. Da die Umkehrung ebenfalls gilt, können wir schließen, dass (S) kleine Oszillationen in Bezug auf ω^* genau dann besitzt, wenn ω^* ein Bifurkationspunkt von (4.144) ist. Wir setzen nun Folgendes voraus:

- (H1)** $f \in C^2(\mathbb{R}^n, \mathbb{R}^n)$, $f(0) = 0$,
(H2) $A := f'(0)$ ist nicht singulär und hat r , $r \geq 1$, negative reelle Eigenwerte $-\omega_1^2, -\omega_2^2, \dots, -\omega_r^2$ mit $0 < \omega_1 < \omega_2 < \dots < \omega_r$.

Satz 4.27 (Ambrosetti und Prodi (1993), Theorem 3.2)

Die Funktion f möge (H1) und (H2) erfüllen. Weiter sei ω_j so gewählt, dass

- (H3)** $-\omega_j^2$ einfach ist (d. h. ω_j hat die algebraische Vielfachheit eins),
(H4) ω_l/ω_j nicht ganzzahlig für alle $l \neq j$ ist.

Dann besitzt (S) kleine Oszillationen in Bezug auf die Frequenz ω_j .

Beweis: Es bezeichne $C_{2\pi}^k$ den Raum der 2π -periodischen Funktionen der Klasse $C^k(\mathbb{R}, \mathbb{R}^n)$ und $F : C_{2\pi}^2 \times \mathbb{R} \rightarrow C_{2\pi}^0$ ist definiert durch

$$F(u, \omega) := \omega^2 \frac{d^2 u}{d\tau^2} - f(u).$$

Mit Hilfe des Satzes 4.9 wollen wir zeigen, dass $\mu_0 = \omega_j$ ein Bifurkationspunkt der Gleichung $F = 0$ ist. Zu diesem Zweck haben wir die Ableitung

$$L := F_u(0, \omega_j) \in L(C_{2\pi}^2, C_{2\pi}^0), \quad L : u \mapsto \omega_j^2 \frac{d^2 u}{d\tau^2} - Au \quad (A := f'(0)) \quad (4.145)$$

zu untersuchen. Es ist leicht zu sehen, dass für $n = 1$ und $A := -\omega_j^2$ der Kern von L durch die Funktionen $\sin \tau$ und $\cos \tau$ aufgespannt wird, also zweidimensional ist. Dieser Fall wird ausführlich in Abschnitt 4.3.2 diskutiert. Für die Anwendung des Satzes 4.9 benötigen wir jedoch einen eindimensionalen Kern. Deshalb schränken wir F auf die Teilräume

$$X := \{u \in C_{2\pi}^2 \mid u(\tau) = u(-\tau)\}, \quad Y := \{v \in C_{2\pi}^0 \mid v(\tau) = v(-\tau)\}$$

ein. Die Einschränkung $F|_X$ wird ebenfalls wieder mit F bezeichnet. Es ist klar, dass $F(\cdot, \omega)$ den Raum X in Y abbildet, da (S) ein autonomes DGL-System ohne die Ableitung \dot{u} ist. Die Fourier-Reihenentwicklung von $u \in X$ lautet

$$u = u_0 + u_1 \cos \tau + u_2 \cos 2\tau + \dots = \sum_{k \geq 0} u_k \cos k\tau, \quad u_k := \frac{1}{\pi} \int_{-\pi}^{\pi} u(\tau) \cos k\tau \, d\tau,$$

während die Fourier-Reihe von $v \in Y$ die formale Darstellung

$$v = v_0 + v_1 \cos \tau + v_2 \cos 2\tau + \dots = \sum_{k \geq 0} v_k \cos k\tau, \quad v_k := \frac{1}{\pi} \int_{-\pi}^{\pi} v(\tau) \cos k\tau \, d\tau$$

mit $u_k, v_k \in \mathbb{R}^n$ hat. Wir bestimmen nun $\ker(L)$ und $R(L)$.

Lemma 4.28

Es gilt

- i. $\ker(L) = \text{span}\{\xi \cos \tau\}$,
- ii. $R(L) = \{v \in Y \mid \Pi v_1 = 0, \quad v_1 = \frac{1}{\pi} \int_0^{2\pi} v(\tau) \cos \tau \, d\tau\}$ (Π siehe (4.148)).

Beweis: Es wird die Gleichung $Lu = v$ mit $u \in X$ und $v \in Y$ betrachtet. Falls u_k, v_k die Fourier-Koeffizienten von u, v bezeichnen, dann folgen aus $Lu = v$ durch Einsetzen der Fourier-Reihen die Gleichungen (Koeffizientenvergleich)

$$M_k(u_k) := -k^2 \omega_j^2 u_k - Au_k = v_k, \quad (k \in \mathbb{N} \cup \{0\}). \quad (4.146)$$

Der Koeffizientenvergleich kann durchgeführt werden, weil wegen

$$\int_{-\pi}^{\pi} u(\tau) \cos k\tau \, d\tau = -\frac{1}{k^2} \int_{-\pi}^{\pi} u''(\tau) \cos k\tau \, d\tau = o\left(\frac{1}{k^2}\right) \quad (k \rightarrow \infty)$$

die Reihe $\sum_{k \geq 0} u_k \cos k\tau$ gleichmäßig konvergiert und $A \in L(\mathbb{R}^n, \mathbb{R}^n)$ stetig ist (Vertauschung von Operator A und Summation).

Falls $Lu = v$ gilt, dann erfüllen u_k, v_k die Gleichungen (4.146). Umgekehrt folgt: Wenn $u \in X$ und $v \in Y$ und u_k, v_k die Gleichungen (4.146) erfüllen, dann gilt auch $Lu = v$. Damit kann die Bestimmung von $\ker(L)$ und $R(L)$ im \mathbb{R}^n durchgeführt werden.

Wir behaupten, dass die Gleichungen (4.146) für alle $k \neq 1$ genau eine Lösung haben. Falls $k = 0$, dann ist $M_0 = -A$ und A ist invertierbar (A hat keinen Eigenwert Null). Falls $k \neq 0$ und $k \neq 1$, dann folgt aus (H3), dass $-k^2 \omega_j^2$ kein Eigenwert von A ist. Folglich ist M_k ($k \neq 0, 1$) invertierbar und es gilt

$$u_k = M_k^{-1}(v_k) \quad (k \neq 0, 1).$$

Für $k = 1$ ist

$$M_1(u_1) = -\omega_j^2 u_1 - Au_1 = v_1$$

und die Lösbarkeitsbedingung für diese Gleichung wird mit Hilfe der adjungierten Gleichung

$$M_1^T(\eta^*) = \zeta^* \quad (4.147)$$

formuliert. Nach Satz 1.53 ist $M_1(u_1) = v_1$ lösbar genau dann, wenn $\langle \eta^*, v_1 \rangle = 0$ für alle $\eta^* \in \ker(M_1^T)$. Wegen

$$\langle \eta^*, v_1 \rangle = \langle \eta^*, M_1(u_1) \rangle = \langle M_1^T(\eta^*), u_1 \rangle = 0$$

verschwindet $\eta^* \in \mathbb{R}^n$ auf $R(M_1)$.

Nun gibt es ein Element $\xi \in \mathbb{R}^n$, $\xi \in \ker(M_1)$ mit $\langle \eta^*, \xi \rangle = 1$. Wäre $\langle \eta^*, \xi \rangle = 0$, dann würde gemäß Lösbarkeitsbedingung gelten: Es existiert ein \tilde{u}_1 , so dass $M_1 \tilde{u}_1 = \xi$. Daraus würde folgen: $M_1^2 \tilde{u}_1 = M_1 \xi = 0$, was wiederum bedeutet, dass \tilde{u}_1 ein Hauptvektor der Stufe 2 wäre. Es ist aber $-\omega_j^2$ ein einfacher Eigenwert, ein Widerspruch.

Mit $\xi \in \ker(M_1)$ zerlegt der Projektor

$$\Pi : \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad \Pi x := \langle \eta^*, x \rangle \xi \quad (4.148)$$

den \mathbb{R}^n in eine (topologisch) direkte Summe

$$\mathbb{R}^n = \Pi(\mathbb{R}^n) \oplus (I - \Pi)(\mathbb{R}^n) = R(\Pi) \oplus \ker(\Pi). \quad (4.149)$$

Nach diesen Vorüberlegungen können i. und ii. bewiesen werden.

Zu i.: Falls $v = 0$ ist, dann hat (4.146) die Lösung

$$\left. \begin{array}{l} u_k = 0, \text{ falls } k \neq 1 \\ u_1 = a\xi \text{ (} a \in \mathbb{R}, \xi \in \mathbb{R}^n \text{)}. \end{array} \right\}$$

Folglich ist $\ker(L) = \text{span}\{\xi \cos \tau\}$. Damit ist i. bewiesen. \checkmark

Zu ii.: Es gelte zunächst $v = Lu$, d. h. es sei $v \in R(L)$. Aus der voranstehenden Argumentation folgt, dass dann $\Pi v_1 = 0$ und folglich $\{v \in Y \mid \Pi v_1 = 0\} \supset R(L)$ ist.

Umgekehrt sei $v \in Y$ gegeben, so dass $\Pi v_1 = 0$ ($v_1 = \frac{1}{\pi} \int_0^{2\pi} v(\tau) \cos \tau \, d\tau$). Wir müssen zeigen, dass $Lu = v$ unter dieser Voraussetzung eine Lösung u besitzt. Dann ist $R(L) = \{v \in Y \mid \Pi v_1 = 0\}$ gezeigt. Nun hat $Lu = v$ eine formale Lösung der Form

$$u(\tau) = -A^{-1}v_0 + u_1 \cos \tau + \tilde{u}(\tau) \quad (4.150)$$

mit $M_1 u_1 = v_1$ und

$$\tilde{u}(\tau) := \sum_{k \geq 2} u_k \cos k\tau = \sum_{k \geq 2} M_k^{-1} v_k \cos k\tau. \quad (4.151)$$

Wir zeigen, dass \tilde{u} von der Klasse C^2 ist. M_k wird umgeformt:

$$M_k = (-k^2 \omega_j^2 I - A) = -k^2 \omega_j^2 \left(I + \frac{1}{k^2 \omega_j^2} A \right).$$

Da A beschränkt ist, existiert eine gleichmäßige Schranke für $1/(k^2\omega_j^2)A$. Es gilt

$$\left\| \frac{1}{k^2\omega_j^2} A \right\| \leq 1 - \delta < 1 \quad \text{und} \quad \left\| \left(I - \frac{1}{k^2\omega_j^2} A \right)^{-1} \right\| \leq \frac{1}{\delta}$$

für ein $\delta > 0$ und für alle $k > k_0$, wobei k_0 hinreichend groß ist. Für $k > k_0$ folgt, dass M_k invertierbar ist und das gilt

$$\begin{aligned} M_k^{-1} &= (-k^2\omega_j^2 I - A)^{-1} = \frac{-1}{k^2\omega_j^2} \left(I + \frac{1}{k^2\omega_j^2} A \right)^{-1} \\ &= \frac{-1}{k^2\omega_j^2} \left[I - \frac{1}{k^2\omega_j^2} A + \frac{1}{(k^2\omega_j^2)^2} A^2 \mp \dots \right] \quad (\text{Neumannsche Reihe}) \\ &= \frac{-1}{k^2\omega_j^2} I + N_k, \quad \text{wobei } N_k := O\left(\frac{1}{k^4}\right) \quad (k \rightarrow \infty). \end{aligned} \quad (4.152)$$

Wir betrachten von der Funktion \tilde{u} den Reihenrest für $k > k_0$ und bezeichnen ihn wieder mit \tilde{u} . Er lässt sich darstellen als

$$\begin{aligned} \tilde{u}(\tau) &= \sum_{k > k_0} \left(\frac{-1}{k^2\omega_j^2} I + N_k \right) v_k \cos k\tau = w(\tau) + z(\tau) \quad \text{mit} \\ w(\tau) &:= \sum_{k > k_0} -\frac{1}{k^2\omega_j^2} v_k \cos k\tau \quad \text{und} \quad z(\tau) := \sum_{k > k_0} N_k v_k \cos k\tau. \end{aligned} \quad (4.153)$$

Wir zeigen, dass $w \in \mathcal{C}_{2\pi}^2$ ist. Zweimalige partielle Integration liefert

$$\begin{aligned} u_k &\equiv u_k[u] = \langle u(\cdot), \cos k(\cdot) \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} u(\tau) \cos k\tau \, d\tau \quad (u \in \mathcal{C}_{2\pi}^2) \\ &= -\frac{1}{k^2} \frac{1}{\pi} \int_{-\pi}^{\pi} \ddot{u}(\tau) \cos k\tau \, d\tau = -\frac{1}{k^2} \langle \ddot{u}(\cdot), \cos k(\cdot) \rangle = -\frac{1}{k^2} u_k[\ddot{u}]. \end{aligned}$$

Nun folgt aus der Beziehung

$$-\frac{1}{k^2\omega_j^2} v_k = -\frac{1}{k^2\omega_j^2} M_k u_k[u] = \frac{1}{k^4\omega_j^2} M_k u_k[\ddot{u}], \quad (\ddot{u} \in \mathcal{C}_{2\pi}^2)$$

die zweimalige Differenzierbarkeit der Reihe (4.153), was zeigt, dass $w \in \mathcal{C}_{2\pi}^2$.

Bei $z(\tau)$ sieht man die $\mathcal{C}_{2\pi}^2$ -Eigenschaft leicht, da aus (4.153) in Verbindung mit (4.152) sofort

$$N_k v_k = O\left(\frac{1}{k^4}\right) \quad (k \rightarrow \infty)$$

folgt. Damit ist auch das durch (4.151) gegebene \tilde{u} in $\mathcal{C}_{2\pi}^2$ gelegen. Da u durch (4.150) gegeben ist, folgt auch $u \in \mathcal{C}_{2\pi}^2$ und gleichzeitig ist u eine Lösung von $Lu = v$. Dies zeigt die Behauptung ii. des Lemmas. \checkmark

Aus Lemma 4.28 folgt nun, dass $\dim \ker(L) = 1$, $\text{codim } R(L) = 1$ und somit L ein Fredholm-Operator mit Index Null ist. Des Weiteren ist die Bifurkationsbedingung (4.34) aus Satz 4.9 erfüllt. Auf die Situation hier übertragen, lautet sie:

$$F_{u,\omega}(0, \omega_j) : u \mapsto 2\omega_j \frac{d^2 u}{d\tau^2} \quad \text{und} \quad F_{u,\omega}(0, \omega_j)(\xi \cos \tau) = -2\omega_j \xi \cos \tau \notin R(L).$$

Jetzt folgt die Behauptung aus dem Satz 4.9 von Crandall und Rabinowitz. \blacksquare

4.3 Die Hopf-Bifurkation

In diesem Abschnitt wollen wir das Abzweigen einer periodischen Lösung von einer stationären Lösung bei autonomen Systemen erster Ordnung untersuchen. Dieses Problem führt auf Abbildungen, die von einem zweidimensionalen Parameter abhängen. Wir werden zeigen, dass wir durch passende Anwendung der Ljapunov-Schmidt-Reduktion die *Hopf-Bifurkation* erhalten. Wir betrachten das autonome DGL-System der Form

$$\frac{du}{dt} = f(u, \mu), \quad u \in \mathbb{R}^n, \quad f : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n, \quad (S_\mu)$$

dass von einem reellen Parameter μ abhängt. Da die rechte Seite von (S_μ) nicht von der Zeit abhängt (autonome DGL), ist die Periode der Lösung, die wir suchen, a priori unbekannt. Um die Methoden der nichtlinearen Funktionalanalysis einzusetzen, ist es erforderlich, in Funktionenräumen mit fester Periode zu arbeiten. Zu diesem Zweck wird eine Zeitskalierung $\tau = \omega t$ ($\omega > 0$) analog dem Beispiel 4.26 durchgeführt und nun 2π -periodische Lösungen von

$$\omega \frac{d\tilde{u}}{d\tau} = f(\tilde{u}, \mu). \quad (S_{\mu, \omega})$$

gesucht. Falls \tilde{u} eine 2π -periodische Lösung von $(S_{\mu, \omega})$ ist, dann ist $u(t) = u(\tau/\omega) = \tilde{u}(\tau)$ eine $2\pi/\omega$ -periodische Lösung von (S_μ) .

Um die Bifurkation von periodischen Lösungen der Gleichung (S_μ) zu studieren, setzen wir voraus, dass $f(0, \mu) \equiv 0$ gilt. Dann besitzt $(S_{\mu, \omega})$ die triviale Lösung $\tilde{u} \equiv 0$ und wir sind am Auffinden möglicher Bifurkationspunkte (μ_0, ω_0) , an denen *nichttriviale periodische Lösungen* entstehen, interessiert.

4.3.1 Abstraktes Hopf-Bifurkationstheorem

Wir wollen zunächst ein Bifurkationsresultat für Gleichungen

$$F(u, \mu, \omega) = 0 \quad (4.154)$$

mit $F : X \times \mathbb{R}^2 \rightarrow Y$ (X, Y Banach-Räume),

$$F \in \mathcal{C}^2(X \times \mathbb{R}^2, Y) \quad \text{und} \quad F(0, \mu, \omega) = 0 \quad \text{für alle } (\mu, \omega) \in \mathbb{R}^2 \quad (4.155)$$

herleiten, dass sich dann auf autonome DGL-Systeme der Form (S_μ) bzw. $(S_{\mu, \omega})$ anwenden lässt. Wir setzen der Einfachheit halber

$$L := F_u(0, \mu_0, \omega_0), \quad V := \ker(L), \quad R := R(L)$$

und präzisieren die Voraussetzungen: Die Abbildung F möge die Bedingungen (4.155) erfüllen. Weiter fordern wir, dass L die Voraussetzungen (H1) und (H2) erfüllt:

(H1) $\dim(\ker(L)) = 2$,

(H2) $\text{codim}(R(L)) = 2$.

Die Voraussetzung (H2) impliziert, dass $R(L)$ abgeschlossen ist.

Es bezeichnen (in Abänderung der Bezeichnung in (4.6)) W und Z die jeweiligen komplementären Unterräume von V (in X) bzw. von R (in Y). Dann bestehen die Zerlegungen

$$X = V \oplus W, \quad Y = Z \oplus R, \quad (4.156)$$

mit $\dim(Z) = 2$. Wegen der Voraussetzung (H2) kann man auf Z stetig projizieren. Somit existiert ein linearer stetiger Operator $Q : Y \rightarrow Z$ mit $Q(Y) = Z$ und es gilt $Y = Q(Y) \oplus (I - Q)(Y) = R(Q) \oplus \ker(Q)$. Wir setzen noch

$$M := F_{u,\mu}(0, \mu_0, \omega_0) \quad \text{und} \quad N := F_{u,\omega}(0, \mu_0, \omega_0).$$

Bei Abbildungen der Form $F : X \times \mathbb{R} \rightarrow Y$, die von einem reellen Parameter abhängen, ist die gemischte Ableitung $F_{u,\mu}(u_0, \mu_0)$ eine lineare Abbildung von \mathbb{R} auf $\mathcal{L}(X, Y)$: $F_{u,\mu}(u_0, \mu_0) \in \mathcal{L}(\mathbb{R}, \mathcal{L}(X, Y))$. Wir können (und werden) $F_{u,\mu}(u_0, \mu_0)$ mit der linearen Abbildung $h \mapsto F_{u,\mu}(u_0, \mu_0)[h, 1]$ identifizieren. Diese Identifikation ist gerechtfertigt, wenn man den kanonischen Isomorphismus $i : \mathcal{L}(\mathbb{R}, Y) \rightarrow Y$, $i(A) = A(1)$ einführt. Das bedeutet, dass wir statt $F_{u,\mu}(u_0, \mu_0)[h, 1]$ den Ausdruck $F_{u,\mu}(u_0, \mu_0)[h]$ betrachten können.

Satz 4.29 (Abstraktes Hopf-Bifurkationstheorem)

F genüge den Bedingungen (4.155) und es sei (μ_0, ω_0) so gewählt, dass $L := F_u(0, \mu_0, \omega_0)$ die Voraussetzungen (H1) und (H2) erfüllt. Darüber hinaus existiere ein $\tilde{v} \in V$, so dass

$$QM\tilde{v} \quad \text{und} \quad QN\tilde{v} \quad \text{linear unabhängig} \quad (4.157)$$

sind. Dann ist (μ_0, ω_0) ein Bifurkationspunkt für die Gleichung $F(u, \mu, \omega) = 0$.

Die Voraussetzung (4.157) hängt nicht von der Wahl des Projektors Q bzw. von der Wahl des komplementären Teilraumes Z ab.

Beweis: Wir wenden die Ljapunov-Schmidt-Reduktion an und setzen $u = v + w$ mit $v \in V$ und $w \in W$. Die Gleichung $F(u, \mu, \omega) = 0$ splittet sich in das System

$$\left. \begin{aligned} (I - Q)F(v + w, \mu, \omega) &= 0, \\ QF(v + w, \mu, \omega) &= 0 \end{aligned} \right\} \quad (4.158)$$

auf. Wie bereits mehrfach in Abschnitt 4.1 dargelegt, kann die erste Gleichung von (4.158) mit dem impliziten Funktionentheorem gelöst werden. (Es sei daran erinnert, dass $L : W \rightarrow R$ stetig invertierbar ist und $(I - Q)L = L$ gilt.) Dies liefert die Funktion

$$w = \psi(v, \mu, \omega),$$

wobei $\psi \in \mathcal{C}^2$ und auf einer Umgebung von $(0, \mu_0, \omega_0) \in V \times \mathbb{R}^2$, mit Werten in W definiert ist. Darüber hinaus gilt

$$\psi(0, \mu, \omega) \equiv 0, \quad \psi_v(0, \mu_0, \omega_0) = 0.$$

Der Wert der Ableitung $\psi_v(0, \mu_0, \omega_0) = 0$ berechnet sich aus der ersten Gleichung in (4.158), indem man $w = \psi(v, \mu, \omega)$ einsetzt und dann die Identität nach v ableitet. Die Ableitung an der Stelle $(0, \mu_0, \omega_0)$ lautet

$$L \circ [v + \psi_v(0, \mu_0, \omega_0)v] = 0 \quad \text{mit } v \in V = \ker(L).$$

Daraus folgt zunächst $L \circ [\psi_v(0, \mu_0, \omega_0)v] = 0$. Wegen $\psi_v(0, \mu_0, \omega_0)v \in W$ und der Invertierbarkeit von L auf W folgt auch $\psi_v(0, \mu_0, \omega_0)v = 0$ für alle v , was sofort $\psi_v(0, \mu_0, \omega_0) = 0$ nach sich zieht. Substitution der Lösung $\psi(v, \mu, \omega)$ in die zweite Gleichung von (4.158) führt uns zu der Gleichung

$$QF(v + \psi(v, \mu, \omega), \mu, \omega) = 0. \quad (4.159)$$

Unter Berücksichtigung der Voraussetzung (4.157) suchen wir Lösungen von (4.159) in der Form $v = s\tilde{v}$ mit $s \in \mathbb{R}$. Damit ergibt sich die zu lösende Gleichung

$$h : \mathbb{R} \times U(\mu_0, \omega_0) \rightarrow Z, \quad h(s, \mu, \omega) := QF(s\tilde{v} + \psi(s\tilde{v}, \mu, \omega), \mu, \omega) = 0. \quad (4.160)$$

Für $h \in \mathcal{C}^2$ gilt:

$$h(0, \mu, \omega) = QF(\psi(0, \mu, \omega), \mu, \omega) = QF(0, \mu, \omega) \equiv 0.$$

Wir setzen

$$h(s, \mu, \omega) = \chi(s, \mu, \omega) \cdot s$$

und bemerken, dass χ eine Funktion mit Werten in Z und von der Klasse \mathcal{C}^1 ist. Dies folgt aus der Beziehung

$$\frac{h(s, \mu, \omega) - h(0, \mu, \omega)}{s} = \chi(s, \mu, \omega), \quad s \neq 0.$$

Nach dem Grenzübergang ($s \rightarrow 0$) folgt wegen der Differenzierbarkeit von h :

$$\frac{\partial h(0, \mu, \omega)}{\partial s} = \chi(0, \mu, \omega).$$

Damit ergibt sich für χ :

$$\chi(0, \mu, \omega) = \frac{\partial h(0, \mu, \omega)}{\partial s} \stackrel{(4.160)}{=} QF_u(0, \mu, \omega)[\tilde{v} + \psi_v(0, \mu, \omega)\tilde{v}].$$

Da $\psi_v(0, \mu_0, \omega_0) = 0$ folgt, dass

$$\begin{aligned} \chi_\mu(0, \mu_0, \omega_0) &= QF_{u,\mu}(0, \mu_0, \omega_0)[\tilde{v} + \psi_v(0, \mu_0, \omega_0)\tilde{v}] \\ &\quad + QF_u(0, \mu_0, \omega_0)[\psi_{v,\mu}(0, \mu_0, \omega_0)\tilde{v}]. \end{aligned}$$

Unter Benutzung der Tatsache, dass $\psi(0, \mu_0, \omega_0) = 0$ und $QF_u(0, \mu_0, \omega_0)u = QLu = 0$ für alle u ist, finden wir

$$\frac{\partial \chi}{\partial \mu}(0, \mu_0, \omega_0) = QF_{u,\mu}(0, \mu_0, \omega_0)[\tilde{v}] = M\tilde{v}.$$

Entsprechend ergibt sich

$$\frac{\partial \chi}{\partial \omega}(0, \mu_0, \omega_0) = QF_{u,\omega}(0, \mu_0, \omega_0)[\tilde{v}] = N\tilde{v}.$$

Aus Voraussetzung (4.157) schließen wir, dass die Determinante der Jacobi-Matrix $\det(\partial\chi/\partial(\mu, \omega))$, betrachtet an der Stelle $(0, \mu_0, \omega_0)$, verschieden von Null ist. Folglich kann man wieder das implizite Funktionentheorem auf die Gleichung $\chi = 0$ anwenden und man erhält (lokal) Funktionen μ und ω , die von s abhängen. Genauer: Es existieren C^1 -Funktionen $\mu(s)$ und $\omega(s)$ (definiert auf einer Umgebung um $s = 0$), so dass $\mu(0) = \mu_0$, $\omega(0) = \omega_0$ und $\chi(s, \mu(s), \omega(s)) = 0$ gelten. Es folgt, dass

$$QF(s\tilde{v} + \psi(s\tilde{v}, \mu(s), \omega(s)), \mu(s), \omega(s)) = 0$$

und hier der Zweig $u = u_s = s\tilde{v} + \psi(s\tilde{v}, \mu(s), \omega(s))$ zu einer Familie von nichttrivialen Lösungen von $F = 0$ Anlass gibt. Da $u_s \neq 0$ für $s \neq 0$ und $u_s \rightarrow \psi(0, \mu_0, \omega_0) = 0$ für $s \rightarrow 0$ ist (μ_0, ω_0) ein Bifurkationspunkt für $F = 0$. \square

4.3.2 Nichtlineare Schwingungen in autonomen Systemen

Betrachtet wird das autonome System

$$\frac{du}{dt} = f(u, \mu) \tag{S_\mu}$$

unter der Voraussetzung, dass

$$f \in C^2(\mathbb{R}^n \times \mathbb{R}, \mathbb{R}^n) \quad \text{und} \quad f(0, \mu) \equiv 0, \quad \mu \in \mathbb{R} \tag{4.161}$$

gilt. Wegen (4.161) besitzt (S_μ) für jedes μ die triviale Lösung $u(t) \equiv 0$. Wir suchen nach möglichen Werten μ , von denen eine nichtkonstante periodische Lösung von (S_μ) abzweigt. Um das abstrakte Hopf-Theorem 4.29 anzuwenden, führen wir gemäß Abschnitt 4.3.1 wieder einen zusätzlichen reellen Parameter $\omega > 0$ ein. Dies erfolgt durch die Zeitskalierung $\tau = \omega t$. Es ist dann $u(t) =: \tilde{v}(\tau) = \tilde{v}(\omega t)$ und die Gleichung (S_μ) geht über in

$$\omega \frac{d\tilde{v}}{d\tau} = f(\tilde{v}, \mu). \tag{S_{\mu,\omega}}$$

Die Gleichung (S_{μ,ω}) stellt ein Hilfssystem für die Lösung des Problems (S_μ) dar. Falls nämlich für ein $\omega > 0$ die Lösung $\tilde{v}(\tau)$ von (S_{μ,ω}) 2π -periodisch ist, dann ist $u(t) = u(\tau/\omega) = \tilde{v}(\tau)$ eine T -periodische Lösung von (S_μ) mit $T = 2\pi/\omega$ (siehe Periodennachweis in Gleichung (4.143)). Das Theorem 4.29 wird nun unmittelbar auf die Gleichung (S_{μ,ω}) angewendet. Es seien

$$X := \{u \in C^1(\mathbb{R}, \mathbb{R}^n) \mid u(\tau + 2\pi) = u(\tau)\},$$

$$Y := \{y \in C^0(\mathbb{R}, \mathbb{R}^n) \mid y(\tau + 2\pi) = y(\tau)\}$$

(X ist ein Untervektorraum von Y) und wir definieren den Nemytski-Operator

$$F : X \times \mathbb{R}^2 \rightarrow Y, \quad F(\tilde{v}, \mu, \omega) := \omega \frac{d\tilde{v}}{d\tau} - f(\tilde{v}, \mu) = 0. \tag{4.162}$$

Man beachte, dass $F \in C^2(X \times \mathbb{R}^2, Y)$, $F(0, \mu, \omega) \equiv 0$ für alle $(\mu, \omega) \in \mathbb{R}^2$ und $f : Y \times \mathbb{R} \rightarrow Y$ mit $f(0, \mu) \equiv 0$ gilt. In der Gleichung (4.162) ist $f(\tilde{v}, \mu)(\tau) = f(\tilde{v}(\tau), \mu)$. Es wird keine neue Bezeichnung für f eingeführt, da keine Verwechslungsgefahr besteht. Wir wollen zunächst den Zusammenhang zwischen der Bifurkation der Lösungen von (4.162) und der Verzweigung der Lösung von (S_μ) erörtern. Es sei $(\mu_0, \omega_0) \in \mathbb{R}^2$, $\omega_0 > 0$, ein Bifurkationspunkt der Gleichung $F = 0$. Dann existieren Folgen $\mu_n \rightarrow \mu_0$, $\omega_n \rightarrow \omega_0$ und $\tilde{v}_n \in X$, so dass $\tilde{v}_n \rightarrow 0$, $\tilde{v}_n \neq 0$ und $F(\tilde{v}_n, \mu_n, \omega_n) = 0$ gilt. Daraus folgt, dass (\tilde{v}_n) eine Folge von 2π -periodischen Lösungen von (S_{μ_n, ω_n}) ist. Wegen des Zusammenhangs $\tilde{v}_n(\tau) = u_n(t)$ sind die $u_n(t)$ eine Folge von periodischen Lösungen von (S_{μ_n}) , deren Perioden $T_n = 2\pi/\omega_n$ gegen $T_0 = 2\pi/\omega_0$ konvergieren. Natürlich konvergiert die Amplitude der Orbits u_n , wegen $\sup_{t \in \mathbb{R}} \|u_n(t)\| = \sup_{\tau \in \mathbb{R}} \|\tilde{u}_n(\tau)\|$ gegen 0 für $n \rightarrow \infty$. Wenn also das Paar (μ_0, ω_0) ein Bifurkationspunkt der Gleichung $F = 0$ (siehe (4.162)) ist, dann zweigt von $(0, \mu_0)$ eine periodische Lösung von (S_μ) mit der Periode $2\pi/\omega_0$ ab. Aus Gründen der Darstellung setzen wir ab jetzt wieder $\tilde{u} = u$ und „ $'$ “ steht für $d/d\tau$. Um die Bifurkationspunkte von F zu finden, haben wir zunächst das linearisierte System

$$F_u(0, \mu, \omega)u = \omega u' - A_\mu u = 0, \quad A_\mu := f_u(0, \mu)$$

zu betrachten. Wir setzen voraus, dass ein μ_0 existiert, so dass für $A_0 := A_{\mu_0}$ Folgendes gilt:

- (A_0 -1) A_0 ist nicht-singulär (d. h. $\lambda = 0$ ist kein Eigenwert) und hat ein Paar von einfachen, rein imaginären Eigenwerten $\pm i\omega_0$, $\omega_0 > 0$;
- (A_0 -2) A_0 hat keine weiteren Eigenwerte der Form $\pm ik\omega_0$, $k \in \mathbb{N}$, $k \neq 1$.

Die Forderung (A_0 -1) bedeutet für den Eigenwert $i\omega_0$: $\dim \ker(-i\omega_0 I + A_0) = 1$ und falls $c \in \mathbb{C}^n$ ($c \neq 0$) ein zugehöriger Eigenvektor ist, dann ist $c \notin R(-i\omega_0 I + A_0)$. Wir verwenden wieder die Notationen aus Abschnitt 4.3.1:

$$L := F_u(0, \mu_0, \omega_0), \quad V := \ker(L), \quad R := R(L), \quad X = V \oplus W, \quad Y = Z \oplus R$$

und setzen auch hier

$$M := F_{u,\mu}(0, \mu_0, \omega_0), \quad N := F_{u,\omega}(0, \mu_0, \omega_0).$$

Um das abstrakte Hopf-Bifurkationstheorem (Satz 4.29) anwenden zu können, müssen zunächst $\ker(L)$ und $R(L)$ bestimmt werden. Wir betrachten die Gleichung

$$Lu = y, \quad u \in X, \quad y \in Y \tag{4.163}$$

Äquivalent dazu heißt das: Wir suchen 2π -periodische Lösungen von

$$\omega_0 u' - A_0 u = y. \quad (4.164)$$

Zum Auffinden periodischer Lösungen von (4.164) benutzen wir die Fouriersche Methode und setzen

$$u = \sum_{k \in \mathbb{Z}} c_k[u] \cdot e^{ik\tau}, \quad y = \sum_{k \in \mathbb{Z}} c_k[y] \cdot e^{ik\tau},$$

wobei jeweils die $c_k \in \mathbb{C}^n$, $c_{-k} = \overline{c_k}$ ($k \in \mathbb{Z}$) die Fourier-Koeffizienten von u bzw. y bezeichnen (mit \bar{z} bezeichnen wir wie gewohnt die komplex konjugierte Zahl von z).

Wir benötigen einige Resultate aus der Theorie der Fourier-Reihen, die wir kurz zusammenstellen. Für die Fourier-Koeffizienten $c_k[y]$ (respektive $c_k[u]$) gilt

$$c_k[y] := \frac{1}{2\pi} \int_0^{2\pi} y(\tau) e^{-ik\tau} d\tau.$$

Da y eine reellwertige Funktion ist, gilt $c_{-k}[y] = \overline{c_k[y]}$. Für jedes $N \in \mathbb{N}$ ist $P_N(y)$ die endliche Fourier-Reihe von y :

$$(P_N(y))(\tau) := \sum_{|k| \leq N} c_k[y] \cdot e^{ik\tau}.$$

Auf $Y_{\mathbb{C}}$, der Komplexifizierung von Y , betrachten wir die Bilinearform

$$\langle v^*, v \rangle := \frac{1}{2\pi} \int_0^{2\pi} (v^*(\tau), v(\tau))_{\mathbb{C}^n} d\tau \quad \text{für alle } v^*, v \in Y_{\mathbb{C}},$$

wobei $(\cdot, \cdot)_{\mathbb{C}^n}$ das (gewöhnliche) Skalarprodukt auf \mathbb{C}^n bezeichnet:

$$(a, b)_{\mathbb{C}^n} := \sum_{i=1}^n \overline{a_i} b_i \quad \text{für alle } a, b \in \mathbb{C}^n.$$

Der Raum $Y_{\mathbb{C}}$ besteht aus allen (formalen) Ausdrücken $x + iy$ mit $x, y \in Y$ und den bekannten Rechenoperation für komplexe Zahlen, z. B.

$$(\alpha + i\beta)(x + iy) = (\alpha x - \beta y) + i(\beta x + \alpha y) \quad (\alpha, \beta \in \mathbb{R}).$$

Die endliche Reihe $P_N(y)$ konvergiert für $N \rightarrow \infty$ im $L^2((0, 2\pi), \mathbb{R}^n)$ gegen $y \in Y$ genau dann, wenn $\sum_{k \in \mathbb{Z}} \|c_k[y]\|^2 < \infty$ (Satz 1.18). Konvergenz im $L^2((0, 2\pi), \mathbb{R}^n)$ bedeutet:

$$\lim_{N \rightarrow \infty} \langle v^*, P_N(y) \rangle = \langle v^*, y \rangle \quad \text{für alle } v^* \in Y.$$

Falls $u \in X$, dann gilt $c_k[\dot{u}] = ikc_k[u]$ und

$$\sum_{k \in \mathbb{Z}} (1 + k^2) \|c_k[u]\|^2 < \infty. \quad (4.165)$$

Diese Abschätzung folgt aus der Tatsache, dass $\dot{u} \in Y \subset L^2((0, 2\pi), \mathbb{R}^n)$ und somit

$$\sum_{k \in \mathbb{Z}} \|c_k[\dot{u}]\|^2 = \sum_{k \in \mathbb{Z}} k^2 \|c_k[u]\|^2 < \infty.$$

Da auch $\sum_{k \in \mathbb{Z}} \|c_k[u]\|^2 < \infty$, folgt die Konvergenz der Reihe (4.165). Mit der Cauchy-Schwarz-Ungleichung ergibt sich daraus

$$\begin{aligned} \left(\sum_{k \in \mathbb{Z}} \|c_k[u]\| \right)^2 &= \left(\sum_{k \in \mathbb{Z}} \frac{1}{\sqrt{1+k^2}} \sqrt{1+k^2} \|c_k[u]\| \right)^2 \\ &\leq \sum_{k \in \mathbb{Z}} \frac{1}{1+k^2} \sum_{k \in \mathbb{Z}} (1+k^2) \|c_k[u]\|^2 < \infty. \end{aligned}$$

Dies impliziert, dass $u(\tau) = \lim_{N \rightarrow \infty} (P_N(u))(\tau)$ gleichmäßig für alle $t \in \mathbb{R}$ konvergiert. Dies gilt für jedes $u \in X$.

Wir nehmen nun an, dass $u \in X$ und $y \in Y$ die Gleichung $Lu = y$ erfüllen. Da A_0 ein linearer stetiger Operator ist und die Fourier-Reihe von u gleichmäßig konvergiert, liefert der Koeffizientenvergleich das unendliche Gleichungssystem

$$\begin{aligned} \Delta_k \cdot c_k[u] &:= c_k[y] \quad \text{für alle } k \in \mathbb{Z}, \\ \Delta_k \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^n), \quad \Delta_k \cdot c &:= (ik\omega_0 I - A_0)c \quad \text{für alle } c \in \mathbb{C}^n. \end{aligned} \tag{4.166}$$

Man rechnet leicht nach, dass

$$\Delta_{-k} \cdot \bar{c} = \overline{\Delta_k \cdot c} \quad \text{für alle } c \in \mathbb{C}^n, k \in \mathbb{Z}$$

gilt. Mit Δ_k^* wird die Adjungierte von Δ_k bezeichnet und es gilt

$$(\Delta_k^* \cdot a, b) = (a, \Delta_k \cdot b) \quad \text{für alle } a, b \in \mathbb{C}^n.$$

Nun bestimmen wir $\ker(L)$ und $R(L)$.

Lemma 4.30

Es gelten (A_0-1) - (A_0-2) . Dann erfüllt L die Voraussetzungen $(H1)$ - $(H2)$ aus Abschnitt 4.3.1, d. h. L ist ein Fredholm-Operator mit $\dim \ker(L) = \text{codim}(R(L)) = 2$.

Beweis: 1. Zur Bestimmung von $\ker(L)$ untersuchen wir das Gleichungssystem

$$(ik\omega_0 I - A_0)c_k[u] = 0, \quad (k \in \mathbb{Z}). \tag{4.167}$$

Aus den Voraussetzungen (A_0-1) - (A_0-2) folgt, dass die Matrix $(ik\omega_0 I - A_0)$ für $k \neq \pm 1$ invertierbar ist. Somit folgt $c_k[u] = 0$ für $k \neq \pm 1$. Für $k = \pm 1$ gilt die Eigenwertgleichung

$$(\pm i\omega_0 I - A_0)c_{\pm 1}[u] = 0. \tag{4.168}$$

Nach Voraussetzung sind $\pm i\omega_0$ einfache Eigenwerte von A_0 . Falls $\xi \in \mathbb{C}^n$ ein Eigenvektor ist, der $\ker(i\omega_0 I - A_0)$ aufspannt, dann wird $\ker(-i\omega_0 I - A_0)$ durch den konjugiert komplexen Eigenvektor $\bar{\xi}$ aufgespannt. Damit folgt, dass die Funktionen $\xi e^{i\tau}$ und $\bar{\xi} e^{-i\tau}$ eine Basis für $\ker(L)$ bilden. Um eine reellwertige Funktion zu erhalten setzen wir

$$u(\tau) = a\xi e^{i\tau} + \bar{a}\bar{\xi} e^{-i\tau}, \quad a \in \mathbb{C}.$$

Folglich ist

$$\ker(L) = \{a\xi e^{i\tau} + \bar{a}\bar{\xi} e^{-i\tau} \mid \xi \in \ker(\Delta_1), \bar{\xi} \in \ker(\Delta_{-1}), a \in \mathbb{C}\} \quad (4.169)$$

Eine äquivalente Darstellung ist

$$\ker(L) = \{\operatorname{Re}(\xi e^{i\tau}) \mid \xi \in \ker(\Delta_1)\}. \quad (4.170)$$

Danach ist $\dim \ker(L) = 2$ bewiesen. ✓

2. Als Nächstes untersuchen wir den Wertevorrat von L . Dazu betrachten wir die Gleichung (4.164) in Verbindung mit Gleichung (4.166). Wie zuvor sichern die Voraussetzungen (A_0-1) - (A_0-2) , dass (4.166) für $k \neq \pm 1$ eindeutig lösbar ist. Dies führt formal zu einer Fourier-Reihe

$$\sum_{k \neq \pm 1} c_k[u] \cdot e^{ikt}, \quad c_k[u] := \Delta_k^{-1} c_k[y], \quad (4.171)$$

von der wir nun zeigen, dass sie zu X gehört. Δ_k wird umgeformt:

$$\Delta_k := ik\omega_0 I - A_0 = ik\omega_0 \left(I - \frac{1}{ik\omega_0} A_0 \right).$$

Da A_0 beschränkt ist, existiert eine gleichmäßige Schranke für $1/(ik\omega_0)A_0$. Es gilt

$$\left\| \frac{1}{ik\omega_0} A_0 \right\| \leq 1 - \delta < 1 \quad \text{und} \quad \left\| \left(I - \frac{1}{ik\omega_0} A_0 \right)^{-1} \right\| \leq \frac{1}{\delta}$$

für ein $\delta > 0$ und für alle $|k| > k_0$, wobei k_0 hinreichend groß ist. Für $|k| > k_0$ folgt, dass Δ_k invertierbar ist und

$$\|\Delta_k^{-1}\| \leq C_1 \frac{1}{k\omega_0} = C \frac{1}{k}$$

für ein $C > 0$ gilt. Sei c_k , $k \in \mathbb{Z}$, irgend eine Lösung von (4.166) mit $c_{-k} = \bar{c}_k$. Wir zeigen, dass die dazugehörige Fourier-Reihe gegen ein $x \in X$ konvergiert und dass damit $Lx = y$ gilt. Für $|k| > k_0$ folgt $c_k = \Delta_k^{-1} c_k[y]$ und

$$\|c_k\| \leq C \frac{1}{k} \|c_k[y]\|. \quad (4.172)$$

Daraus folgt

$$\left(\sum_{|k| > k_0} \|c_k\| \right)^2 \leq C^2 \left(\sum_{|k| > k_0} \frac{1}{k} \|c_k[y]\| \right)^2 \leq C^2 \sum_{|k| > k_0} \frac{1}{k^2} \sum_{|k| > k_0} \|c_k[y]\|^2 < \infty.$$

Damit ist die Reihe $\sum_{k \in \mathbb{Z}} c_k e^{ik\tau}$ gleichmäßig konvergent gegen eine stetige, 2π -periodische Funktion $x \in X$. Wir behaupten

$$R(L) = \{y \in Y \mid (c_k^*, c_k[y]) = 0, \forall c_k^* \in \ker(\Delta_k^*), k = \pm 1\}. \tag{4.173}$$

Da $\dim \ker(\Delta_k^*) = \dim \ker(\Delta_k)$ und $R(L)$ abgeschlossen (sofern man (4.173) bewiesen hat), folgt aus dem Closed-Range-Theorem (Satz 1.42): $R(L)^\perp = \ker(L^*)$. Damit hat man sofort $\dim R(L)^\perp = \text{codim}(R(L)) = \dim \ker(L^*) = \dim \ker(L)$, d. h. L ist ein Fredholm-Operator mit $\text{ind}(L) = 0$. Es genügt also, die Behauptung (4.173) zu beweisen. i.: Es sei $y \in R(L)$. Dann sind

$$\begin{aligned} \omega_0 \frac{dx}{d\tau}(\tau) - A_0 x(\tau) &= y(\tau), \quad \forall \tau \in \mathbb{R} \quad \text{und} \\ \Delta_k \cdot c_k[x] &= c_k[y], \quad \forall k \in \mathbb{Z}, \end{aligned} \tag{4.174}$$

für ein $x \in X$ erfüllt, d. h. $c_k[y] \in R(\Delta_k), \forall k \in \mathbb{Z}$. Dies ist äquivalent zu

$$(c_k^*, c_k[y]) = 0, \quad \forall c_k^* \in \ker(\Delta_k^*), \forall k \in \mathbb{Z} \tag{4.175}$$

(Lösbarkeitsbedingung für (4.174)). Nun ist $\ker(\Delta_k^*) = \{0\}$ für $k \neq \pm 1$ und da $c_k^* \in \ker(\Delta_k^*)$ genau dann gilt, wenn $\overline{c_k^*} \in \ker(\Delta_{-k}^*)$, folgt, dass y zu der Menge auf der rechten Seite von (4.173) gehört. ✓

ii.: Umgekehrt, falls y ein Element der rechten Seite von (4.173) ist, dann ist $y \in Y$ und (4.175) ist erfüllt. Dies impliziert, dass die Gleichung

$$\Delta_k \cdot c_k = c_k[y] \tag{4.176}$$

mindestens eine Lösung $c_k \in \mathbb{C}^n$ für jedes $k \in \mathbb{Z}$ besitzt. Falls c_k eine Lösung von (4.176), dann ist $\overline{c_k}$ eine Lösung von $\Delta_{-k} \cdot c_{-k} = c_{-k}[y]$.

Nun sei $\{c_k \mid k \in \mathbb{Z}\}$ irgendeine Lösung von (4.176) mit $c_{-k} = \overline{c_k}$. Wir betrachten die Fourier-Reihe $\sum_{k \in \mathbb{Z}} c_k e^{ik\tau}$ und werden zeigen, dass diese Reihe gegen ein $x \in X$ konvergiert und $Lx = y$ gilt. Damit ist $y \in R(L)$ und (4.173) bewiesen. Mit Hilfe der Beziehung (4.172) wurde bereits gezeigt, dass die Reihe $\sum_{k \in \mathbb{Z}} c_k e^{ik\tau}$ gleichmäßig konvergiert und damit die Grenzfunktion eine stetige, reellwertige und 2π -periodische Funktion $x(\tau)$ ist. Wir definieren eine Funktion $z : \mathbb{R} \rightarrow \mathbb{R}^n$ durch

$$z(\tau) := x(0) + \frac{1}{\omega_0} \int_0^\tau (A_0 x(s) + y(s)) ds. \tag{4.177}$$

Der Integrand in (4.177) ist stetig und 2π -periodisch. Folglich ist $z \in X$ und es gilt $z(0) = x(0)$. Falls wir zeigen können, dass $z(\tau) = x(\tau), \forall \tau \in \mathbb{R}$, dann ist $x \in X$ und es gilt $Lx = y$, gemäß (4.177). Aus (4.176) und (4.177) folgt, dass $P_N \dot{z} = \frac{d}{d\tau} P_N x$ ist. Es sei nun $g : \mathbb{R} \rightarrow \mathbb{R}^n$ eine C^1 -Funktion. Dann gilt

$$\begin{aligned} \langle \dot{g}, x \rangle &= \lim_{N \rightarrow \infty} \langle \dot{g}, P_N x \rangle = \lim_{n \rightarrow \infty} (g(\tau), (P_N x)(\tau)) \Big|_{\tau=0}^{\tau=2\pi} - \lim_{N \rightarrow \infty} \langle g, \frac{d}{d\tau} P_N x \rangle \\ &= (g(\tau), x(\tau)) \Big|_{\tau=0}^{\tau=2\pi} - \lim_{N \rightarrow \infty} \langle g, P_N \dot{z} \rangle = (g(\tau), z(\tau)) \Big|_{\tau=0}^{\tau=2\pi} - \langle g, \dot{z} \rangle = \langle \dot{g}, z \rangle. \end{aligned}$$

Somit hat man $\langle \dot{g}, x - z \rangle = 0$ für alle $\dot{g} \in \mathcal{C}^0$. Nun gibt es eine Folge von trigonometrischen Polynomen, die gleichmäßig auf \mathbb{R} gegen eine stetige und periodische Funktion konvergiert (vgl. Heuser (1980), Teil 2, Abschnitt 116). Daraus folgt $x = z$. \checkmark
Damit ist das Lemma bewiesen. \square

Mit diesen Ergebnissen können wir nun die Projektoren $P : X \rightarrow X$ und $Q : Y \rightarrow Y$ aufstellen. Wegen (A₀-1) gibt es ein $b_1 \in \mathbb{C}^n$ und ein $b_1^* \in \mathbb{C}^n$, so dass die Gleichungen

$$A_0 b_1 = i\omega_0 b_1 \quad \text{und} \quad A_0^T b_1^* = -i\omega_0 b_1^* \quad (\text{adjungierte Gleichung}) \quad (4.178)$$

gelten. Man kann die b_1 und b_1^* so wählen, dass sie der Normalisierungsbedingung

$$(b_1^*, b_1)_{\mathbb{C}^n} = -2i \quad (4.179)$$

genügen. Für die Funktionen

$$\zeta(\tau) := b_1 e^{i\tau} \quad \text{und} \quad \zeta^*(\tau) := b_1^* e^{i\tau}$$

gelten die Gleichungen ((4.178) wird mit $e^{i\tau}$ multipliziert)

$$A_0 \zeta(\tau) = i\omega_0 \zeta(\tau) = \omega_0 \dot{\zeta}(\tau), \quad A_0^T \zeta^*(\tau) = -i\omega_0 \zeta^*(\tau) = -\omega_0 \dot{\zeta}^*(\tau). \quad (4.180)$$

Weiter sei $u_1(\tau) := \operatorname{Re} \zeta(\tau)$ und $u_2(\tau) = \operatorname{Im} \zeta(\tau)$. Dann folgt aus (4.169) und (4.173), dass

$$\begin{aligned} \ker(L) &= \operatorname{span}\{u_1, u_2\} = \{\operatorname{Re}(a\zeta) \mid a \in \mathbb{C}\} \quad \text{und} \\ R(L) &= \{z \in Y \mid \langle \zeta^*, z \rangle = 0\}. \end{aligned} \quad (4.181)$$

Nun kann man den Projektor $P \in \mathcal{L}(X, X)$ definieren durch

$$\begin{aligned} P(z) &= \frac{1}{\omega_0} \operatorname{Re}(\langle \zeta^*, A_0 z \rangle \zeta) = \frac{1}{\omega_0} \operatorname{Re}(\langle A_0^T \zeta^*, z \rangle \zeta) \\ &= \frac{1}{\omega_0} \operatorname{Re}(\langle -i\omega_0 \zeta^*, z \rangle \zeta) = \frac{1}{\omega_0} \operatorname{Re}(i\omega_0 \langle \zeta^*, z \rangle \zeta) \\ &= \operatorname{Re}(i \langle \zeta^*, z \rangle \zeta) = Q(z) \quad \text{für alle } z \in Y. \end{aligned} \quad (4.182)$$

Setzt man $\zeta = u_1 + iu_2$ und $\zeta^* = -u_2^* + iu_1^*$, dann erhält (4.182) die Form

$$P(z) = \langle u_1^*, z \rangle u_1 + \langle u_2^*, z \rangle u_2 = Q(z) \quad \text{für alle } z \in Y. \quad (4.183)$$

Da $\langle \zeta^*, \zeta \rangle = -2i$ (siehe (4.179)) und $\langle \zeta^*, \bar{\zeta} \rangle = 0$, folgt durch leichte Rechnung

$$\langle u_i^*, u_j \rangle = \delta_{ij}, \quad i, j = 1, 2.$$

Man rechnet auch nach, dass $P^2 = P$ gilt. Mit diesem Projektor folgt

$$P(z) = 0 \Leftrightarrow \langle u_1^*, z \rangle = \langle u_2^*, z \rangle = 0, \quad R(P) = \ker(L) = \operatorname{span}\{u_1, u_2\}$$

und die Zerlegung des Raumes

$$X = R(P) \oplus \ker(P) = \ker(L) \oplus \ker(P). \quad (4.184)$$

Analog gilt

$$Y = R(Q) \oplus \ker(Q) = R(Q) \oplus R(L), \quad (4.185)$$

wobei $R(L) = \{z \in Y \mid \langle u_1^*, z \rangle = \langle u_2^*, z \rangle = 0\}$.

Nun muss gezeigt werden, dass F die Bedingung (4.157) aus Satz 4.29 erfüllt. Wir benutzen $\tilde{v}(\tau) := \zeta(\tau) = b_1 e^{i\tau} \in \ker(L)$ und berechnen $QF_{u,\mu}(0, \mu_0, \omega_0)\tilde{v}$ und $QF_{u,\omega}(0, \mu_0, \omega_0)\tilde{v}$. Wir starten mit

$$F_u(0, \mu, \omega)\tilde{v} = \omega \frac{d\tilde{v}}{d\tau} - A_\mu \tilde{v} = i\omega b_1 e^{i\tau} - A_\mu (b_1 e^{i\tau}). \quad (4.186)$$

Für die Ableitung in Bezug auf ω ergibt sich zunächst

$$F_{u,\omega}(0, \mu_0, \omega_0)\tilde{v} = i b_1 e^{i\tau} = \frac{d}{d\tau} \tilde{v} = \dot{\tilde{v}}.$$

Mit (4.182) folgt

$$\begin{aligned} QF_{u,\omega}(0, \mu_0, \omega_0)\tilde{v} &= Q\dot{\tilde{v}} = \operatorname{Re}(i\langle \zeta^*, i b_1 e^{i\tau} \rangle \zeta) \\ &= \operatorname{Re}(i \cdot i (b_1^*, b_1) \cdot b_1 e^{i\tau}) \stackrel{(4.179)}{=} 2 \operatorname{Re}(i b_1 e^{i\tau}) = 2 \operatorname{Re}(i\zeta). \end{aligned} \quad (4.187)$$

Um $F_{u,\mu}(0, \mu_0, \omega_0)\tilde{v}$ zu berechnen, sind einige Vorüberlegungen erforderlich. Wir bezeichnen mit $\lambda(\mu) = \alpha(\mu) + i\beta(\mu)$ den Teil von Eigenwerten von A_μ , für die $\lambda(\mu_0) = i\omega_0$ (nämlich $\alpha(\mu_0) = 0, \beta(\mu_0) = \omega_0$) gilt. Da $f \in \mathcal{C}^2$, folgt für die Eigenwerte $\lambda \in \mathcal{C}^1$ bzw. $\alpha, \beta \in \mathcal{C}^1$. Da darüber hinaus $i\omega_0$ einfacher Eigenwert ist, kann man schlussfolgern, dass A_μ eine Familie v_μ von Eigenvektoren mit folgenden Eigenschaften hat:

- i.** Die Abbildung $\mu \mapsto v_\mu$ ist \mathcal{C}^1 , **ii.** $A_\mu(v_\mu) = \lambda(\mu)v_\mu$, **iii.** $v_{\mu_0} = v$.

Mit ii. können wir folgende Umformung machen:

$$A_\mu v = A_\mu(v - v_\mu) + A_\mu v_\mu = A_\mu(v - v_\mu) + \lambda(\mu)v_\mu.$$

Die Differentiation nach μ (an der Stelle $\mu = \mu_0$) unter Verwendung der Abkürzungen $A'_{\mu_0} := dA_\mu/d\mu|_{\mu=\mu_0}$, $v' := dv_\mu/d\mu|_{\mu=\mu_0}$ ergibt:

$$A'_{\mu_0} v = A'_{\mu_0}(v - v_{\mu_0}) - A_0 v' + \lambda'(\mu_0)v_{\mu_0} + \lambda(\mu_0)v'.$$

Mit $v_{\mu_0} = v$ und $\lambda(\mu_0) = i\omega_0$ folgt

$$A'_{\mu_0} v = \lambda'(\mu_0)v + (i\omega_0 I - A_0)v'. \quad (4.188)$$

Wir differenzieren jetzt (4.186) nach μ an der Stelle $(0, \mu_0, \omega_0)$ und erhalten mit (4.188):

$$F_{u,\mu}(0, \mu_0, \omega_0)v = -\lambda'(\mu_0)v - (i\omega_0 I - A_0)v'.$$

Die Anwendung des Projektors Q liefert:

$$\begin{aligned} QF_{u,\mu}(0, \mu_0, \omega_0)v &= \operatorname{Re}(i\langle \zeta^*, F_{u,\mu}(0, \mu_0, \omega_0)v \rangle \zeta) \\ &= \operatorname{Re}(i\langle \zeta^*, -\lambda'(\mu_0)v - (i\omega_0 I - A_0)v' \rangle \zeta) \\ &= \operatorname{Re}(i[\langle \zeta^*, -\lambda'(\mu_0)v \rangle - \langle \zeta^*, (i\omega_0 I - A_0)v' \rangle] \zeta). \end{aligned}$$

Beachtet man nun, dass $(i\omega_0 I - A_0)^T \zeta^* = (-i\omega_0 I - A_0^T) \zeta^* = 0$ (siehe (4.180)), dann ergibt sich mit $v = \tilde{v}$

$$\begin{aligned} QF_{u,\mu}(0, \mu_0, \omega_0)\tilde{v} &= \operatorname{Re}(i\langle \zeta^*, -\lambda'(\mu_0)\tilde{v} \rangle \zeta) = \operatorname{Re}(i(b_1^*, -\lambda'(\mu_0)b_1) \zeta) \\ &= \operatorname{Re}(i(2i \cdot \lambda'(\mu_0)) \zeta) = -2 \operatorname{Re}(\lambda'(\mu_0)\zeta). \end{aligned} \quad (4.189)$$

Um Satz 4.29 anwenden zu können, müssen die Ausdrücke (4.187) und (4.189) linear unabhängig sein, d. h. aus

$$a \cdot \operatorname{Re}(i\zeta) - b \cdot \operatorname{Re}(\lambda'(\mu_0)\zeta) = 0 \quad \text{folgt} \quad a = b = 0.$$

Mit $\zeta = u_1 + iu_2$ folgt $\operatorname{Re}(i\zeta) = -u_2$ und $\operatorname{Re}(\lambda'(\mu_0)\zeta) = \alpha'(\mu_0)u_1 - \beta'(\mu_0)u_2$. Wegen der linearen Unabhängigkeit der Funktionen u_1 und u_2 folgt aus

$$a(-u_2) - b(\alpha'(\mu_0)u_1 - \beta'(\mu_0)u_2) = 0$$

zunächst

$$-a + b\beta'(\mu_0) = 0 \quad \text{und} \quad -b\alpha'(\mu_0) = 0.$$

Die lineare Unabhängigkeit der Vektoren (4.187) und (4.163) tritt also genau für

$$(A_0\text{-3}) \quad \alpha'(\mu_0) \neq 0$$

ein. Damit hat man folgende Aussage bewiesen:

Lemma 4.31

Angenommen die Bedingungen (A₀-1)-(A₀-3) gelten. Dann erfüllt F die Voraussetzung (4.157) aus Satz 4.29.

Bedingung (A₀-3) ist eine sogenannte *Transversalitätsbedingung*. Vom geometrischen Standpunkt aus betrachtet bedeutet dies, dass die Eigenwert-Kurve $\lambda(\mu)$ die imaginäre Achse transversal bei $\mu = \mu_0$ schneidet (siehe Abbildung 4.13).

Die Lemmata 4.30 und 4.31 gestatten nun die Anwendung des Satzes 4.29 auf die Gleichung (4.162) und somit auf das DGL-System (S_μ) .

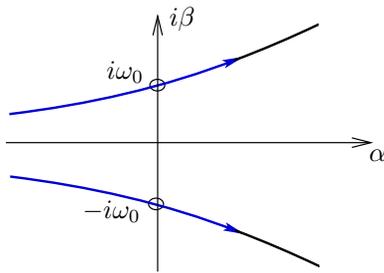


Abb. 4.13 Transversalitätsbedingung $\alpha'(\mu_0) \neq 0$

Satz 4.32 (E. Hopf (1942))

Es sei $f \in \mathcal{C}^2(\mathbb{R}^n \times \mathbb{R}, \mathbb{R}^n)$ mit $f(0, \mu) \equiv 0$ gegeben und für $\mu = \mu_0$ seien die Voraussetzungen (A₀-1)-(A₀-3) erfüllt. Dann zweigt von μ_0 eine periodische Lösung von (S_μ) mit der Periode nahe $2\pi/\omega_0$ ab. Genauer existiert dort eine Umgebung J von $s = 0$, Funktionen $\omega(s), \mu(s) \in \mathcal{C}^1(J)$ und eine Familie u_s von nichtkonstanten, periodischen Lösungen von $(S_{\mu(s)})$, so dass gilt:

- i. $\omega(s) \rightarrow \omega_0, \mu(s) \rightarrow \mu_0$ für $s \rightarrow 0$,
- ii. u_s hat die Periode $T_s = 2\pi/\omega(s)$,
- iii. die Amplitude des Orbits u_s konvergiert gegen Null für $s \rightarrow 0$.

Man überlegt sich, dass das gleiche Ergebnis gilt, wenn man für \tilde{v} die Darstellung $\tilde{v} = \xi e^{i(\tau+\alpha)} + \bar{\xi} e^{-i(\tau+\alpha)}$ für beliebiges α verwendet. Dies liegt daran, dass (S_μ) eine autonome Differentialgleichung ist und die Lösungen unter der Translation $\theta \mapsto (\cdot + \theta)$ invariant bleiben. Dies gilt auch für die Bifurkationslösungen.

Beispiel 4.33 (Van-der-Pol-Gleichung)

Wir betrachten die Van-der-Pol-Gleichung

$$\ddot{x} - (\mu - 3x^2)\dot{x} + x = 0, \tag{4.190}$$

wobei μ ein reeller Parameter ist. Gleichung (4.190) ist äquivalent zu dem System

$$\dot{x} = y + \mu x - x^3, \quad \dot{y} = -x. \tag{4.191}$$

Das linearisierte System lautet

$$\dot{x} = y + \mu x, \quad \dot{y} = -x. \tag{4.192}$$

In diesem Fall ist $A_\mu = \begin{pmatrix} \mu & 1 \\ -1 & 0 \end{pmatrix}$. Die Eigenwerte sind $\lambda(\mu) = \frac{1}{2}(\mu \pm \sqrt{\mu^2 - 4})$. Für $\mu = 0$ ist $\lambda(0) = \pm i$. Es ist leicht zu sehen, dass die Voraussetzungen (A₀-1)-(A₀-3) aus Satz 4.32 mit $\mu_0 = 0$ und $\omega_0 = 1$ erfüllt sind.

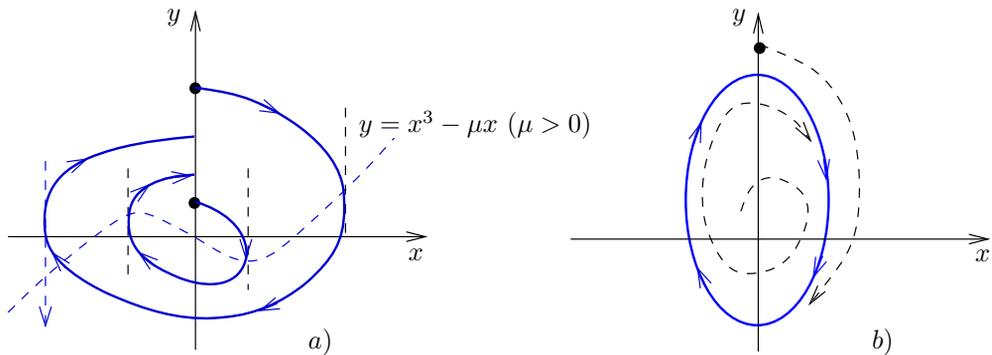


Abb. 4.14 a) Phasenportrait von (4.191), b) Grenzyklus von (4.191)

In diesem speziellen Fall eines Systems im \mathbb{R}^2 ist es möglich, eine direkte Analyse in der (x, y) -Phasenebene zu machen. Man findet leicht, dass für $\mu \leq 0$ das System (4.191) keine periodischen Orbits hat, aber sehr wohl die triviale Lösung $x = y = 0$ vorhanden ist. Außerdem ist die triviale Lösung $x = y = 0$ stabil für alle $\mu < 0$ und instabil für $\mu > 0$. Für $\mu > 0$ hat (4.191) genau eine periodische Lösung, die asymptotisch stabil ist (siehe Abbildung 4.14 b)). ■

4.4 Aufgaben

Aufgabe 4.1

Gegeben sei der lineare Operator $T : X \rightarrow Y$, X, Y lineare Räume über \mathbb{K} . Es sei L ein festes algebraisches Komplement zum Nullraum $\ker(T)$, d. h. L ist linearer Teilraum von X , so dass $X = \ker(T) \oplus L$ gilt. Beweisen Sie die folgenden Behauptungen:

a) Die Einschränkung

$$T : L \rightarrow R(T) \quad (R(T) := T(X)) \quad (4.193)$$

ist linear und bijektiv. Folglich gilt $\operatorname{codim} \ker(T) = \dim R(T)$.

b) Es seien zusätzlich X und Y Banach-Räume, L und $R(T)$ abgeschlossen und der Operator $T : X \rightarrow Y$ stetig. Dann ist der Operator aus (4.193) ein linearer Homöomorphismus.

Aufgabe 4.2

Es seien $\dim X < \infty$, $\dim Y < \infty$ und $S : L \rightarrow R(T)$ bezeichnet die Einschränkung des Operators $T : X \rightarrow R(T) \subset Y$ auf einem linearen Teilraum L von X . L bezeichnet ein festes algebraisches Komplement zum Nullraum $\ker(T)$. Zeigen Sie, dass die Originalgleichung $Tx = b$ für jedes gegebene $b \in R(T)$ die Lösungsmenge

$$S^{-1}b + \ker(T)$$

besitzt und $\dim \ker(T) = \dim X - \dim R(T)$ ist.

Aufgabe 4.3

Es seien $X := C^1[a, b]$, $Y := C^0[a, b]$ und

$$Tx(t) := \dot{x}(t), \quad \forall t \in [a, b].$$

Dann ist der lineare Operator $T : X \rightarrow Y$ ein Fredholm-Operator mit $\text{ind}(T) = 1$.

Aufgabe 4.4

Es sei $T : X \rightarrow Y$ ein linearer Fredholm-Operator, X und Y normierte Räume über \mathbb{K} .

Dann gilt:

- a) T surjektiv $\Leftrightarrow \text{ind}(T) = \dim \ker(T)$.
- b) T injektiv $\Leftrightarrow \text{ind}(T) = \text{codim } R(T)$.
- c) T bijektiv $\Leftrightarrow \text{ind}(T) = \dim \ker(T) = 0$.
- d) Falls X und Y Banach-Räume sind, dann hat die Gleichung

$$Tx = b, \quad x \in X$$

für jedes b genau eine Lösung genau dann, wenn $\text{ind}(T) = \dim \ker(T) = 0$ gilt.

Aufgabe 4.5

Es sei X ein Banach-Raum und $T : X \rightarrow X$ ein linearer kompakter Operator. Man beweise: Ist $R(T)$ abgeschlossen, dann ist $\dim R(T) < \infty$.

Aufgabe 4.6

Geben Sie einen Beweis des Korollars 4.4.

Hinweis: Führen Sie den Beweis indirekt (μ_0 ist also keine charakteristische Zahl von A) und benutzen Sie an Stelle der Gleichung (4.5) den äquivalenten Term

$$x_n = (\mu_n - \mu_0)RAx_n + \mu_n RG(x_n), \quad n \in \mathbb{N},$$

wobei $R := (I - \mu_0 A)^{-1}$ linearer stetiger Operator auf X ist.

Aufgabe 4.7

Bestimmen Sie für das Randwertproblem

$$\left. \begin{aligned} \ddot{u} + \lambda(u - u^3) &= 0, & t \in (0, \pi), \\ u(0) = u(\pi) &= 0. \end{aligned} \right\} \quad (4.194)$$

diejenigen Parameterwerte λ , an denen Bifurkation auftritt.

Hinweis: Führen Sie geeignete Funktionenräume ein und formen Sie (4.194) in eine Operatorgleichung um. (Die Aufgabenstellung kann auch auf elementarem Weg gelöst werden.)

Aufgabe 4.8

Beweisen Sie den Satz 4.12.

Aufgabe 4.9

Wenden Sie den Satz 4.9 auf das RWP (mit den Dirichlet-Bedingungen)

$$\begin{cases} \ddot{x}(t) + \mu x(t) + g(t, x(t), \dot{x}(t), \mu) = 0, & t \in (0, \pi), \\ x(0) = x(2\pi) = 0, \end{cases}$$

an und zeigen Sie, dass jeder Punkt $(0, \mu) = (0, k^2)$, $k \in \mathbb{N}$, ein Bifurkationspunkt ist.

Aufgabe 4.10

Ersetzen Sie die Dirichletschen-Randbedingungen aus Aufgabe 4.9 durch die Neumannschen-Randbedingungen

$$\dot{x}(0) = \dot{x}(\pi) = 0$$

und beweisen Sie, dass jeder Punkt $(0, \mu) = (0, k^2)$, $k \in \mathbb{N} \cup \{0\}$, ein Bifurkationspunkt ist.

Aufgabe 4.11

Wenden Sie den Satz 4.9 auf das RWP

$$\begin{cases} \frac{d^4 x}{dt^4}(t) - \mu x(t) + g(t, x(t), \dot{x}(t), \ddot{x}(t), \dot{\ddot{x}}(t), \mu) = 0, & t \in (0, \pi), \\ x(0) = \ddot{x}(0) = x(\pi) = \ddot{x}(\pi), \end{cases}$$

an und zeigen Sie unter passenden Voraussetzungen an g , dass jeder Punkt $(0, \mu) = (0, k^2)$, $k \in \mathbb{N}$, ein Bifurkationspunkt ist.

Aufgabe 4.12

Berechnen Sie alle Bifurkationslösungen der folgenden Integralgleichung:

$$\lambda x(s) = \frac{2}{\pi} \int_0^\pi [a \sin s \sin t + b \sin(2s) \sin(2t)] [x(t) + x^3(t)] dt, \quad x \in C^0[0, \pi], \quad 0 < b < a.$$

5 Numerik der Gleichgewichtslösungen

Übersicht

5.1	Berechnung von Gleichgewichtslösungen	228
5.2	Parametrisierung von Lösungskurven und Fortsetzungsmethoden	243
5.3	Stabilitäts- und Bifurkationsanalyse	273
5.4	Aufgaben	283

In den vorangegangenen Kapiteln wurden dynamische Systeme als mathematische Objekte eingeführt, die die zeitliche Entwicklung existierender Systeme in Natur-, Ingenieur- und anderen Wissenschaften beschreiben. Zeitkontinuierliche Systeme werden dabei häufig durch gewöhnliche Differenzialgleichungen (DGL) modelliert. Da in den meisten praktischen Anwendungen eine geschlossene Lösung der DGL nicht möglich ist, wollen wir uns in diesem Kapitel mit numerischen Verfahren für folgende Problemklasse befassen: Gegeben ist ein parameterabhängiges C^r -glattes Vektorfeld $f \in C^r(\mathbb{R}^{n+1})$ mit $r \geq 1$, mit dem das autonome System

$$\dot{x} = f(x, \lambda), \quad f : D \times \mathbb{R} \rightarrow \mathbb{R}^n, \quad D \subset \mathbb{R}^n, \quad \text{offen}, \quad \dot{x} \equiv \frac{dx}{dt} \quad (5.1)$$

definiert wird. Dabei ist $\lambda \in \Lambda$ ein reeller Parameter aus einem vorgegebenen endlichen Intervall $\Lambda = [a, b]$. Reale dynamische Systeme besitzen meist eine Vielzahl von Parametern, die wir der Einfachheit wegen bis auf einen einzigen Parameter λ , den sogenannten *Systemparameter* (auch Kontrollparameter genannt), konstant halten wollen.

Beispiel 5.1

1. W. F. Langford (1984) betrachtet ein dynamisches System der Fluidodynamik, das nach mehreren Problemtransformationen auf das DGL-System

$$\begin{aligned} \dot{x}_1 &= (x_3 - 0.7) \cdot x_1 - \omega \cdot x_2 \\ \dot{x}_2 &= \omega x_1 + (x_3 - 0.7) \cdot x_2 \\ \dot{x}_3 &= 0.6 + x_3 - x_3^3/3 - (x_1^2 + x_2^2)(1 + \rho \cdot x_3) + \varepsilon \cdot x_3 \cdot x_1^3 \end{aligned} \quad (5.2)$$

mit den positiven reellen Parametern ω, ρ und ε führt. Mit festen Werten $\omega = 3.5$, $\varepsilon = 0.03$ und dem Systemparameter $\lambda = \rho$ ergibt sich die Darstellung (5.1).

2. Der Meteorologe E. N. Lorenz untersuchte bereits in den 1960er Jahren ein vereinfachtes Modell von Grundgleichungen (vgl. Argyris et al. (1995))

$$\begin{aligned}\dot{x}_1 &= x_2 - x_1 \\ \dot{x}_2 &= \lambda x_1 - x_2 - x_1 x_3 \\ \dot{x}_3 &= x_1 x_2 - x_3\end{aligned}\tag{5.3}$$

mit dem positiven Kontrollparameter λ . Wie (5.2) zeigt auch dieses System trotz scheinbar einfacher Struktur ein erstaunlich reichhaltiges Lösungsverhalten. ■

In diesem Kapitel wollen wir *Gleichgewichtslösungen* parameterabhängiger Systeme (5.1) berechnen, die zugleich Fixpunkte des Flusses darstellen. Die dabei gewonnenen numerischen Techniken werden sich auch für weitere Lösungstypen, wie periodische und quasi-periodische Lösungen, als grundlegend erweisen.

5.1 Berechnung von Gleichgewichtslösungen

Punkte $x^* \in D$ des Phasenraumes, in denen keine Bewegung stattfindet, bezeichnet man als Gleichgewichtspunkte oder Fixpunkte der DGL. Ihre numerische Bestimmung beruht auf den Definitionen 2.28 und 2.31, die wir nun im parameterabhängigen Kontext anführen:

Definition 5.2 (Gleichgewichtspunkt, Attraktor, Repellor)

- i. Ein Punkt $x^* \in D$ heißt *Gleichgewichtspunkt*, *Fixpunkt* oder *kritischer Punkt* der DGL (5.1) zu festem Parameterwert λ^* , wenn $f(x^*, \lambda^*) = 0$ ist.
- ii. Der Gleichgewichtspunkt x^* ist ein (positiver) *Attraktor*, falls eine offene Umgebung $\mathcal{U}(x^*) \subset D$ von x^* existiert, so dass für jede Lösung $x(t)$ von (5.1) mit $x(0) \in \mathcal{U}(x^*)$ stets $\lim_{t \rightarrow \infty} x(t) = x^*$ folgt. Gilt diese Eigenschaft für $t \rightarrow -\infty$, so ist x^* ein *Repellor* (auch negativer Attraktor genannt).

Ist x^* ein Gleichgewichtspunkt zu festem Parameterwert λ^* und $\varphi_t : D \rightarrow \mathbb{R}^n$ die Abbildung des Flusses der DGL (5.1), so ist offenbar $\varphi_t(x^*) = x^*$ für alle $t \in \mathbb{R}$. Damit ist x^* auch ein *Fixpunkt des Flusses*, und die zugehörige konstante Lösung $x(t) = x^*$ wird als *Gleichgewichtslösung* bzw. *Ruhelage* des dynamischen Systems bezeichnet.

Aus dem Existenz- und Eindeutigkeitssatz für die DGL (5.1) ergibt sich, dass ein Gleichgewichtspunkt mit $x(0) \neq x^*$ niemals in endlicher Zeit erreicht werden kann. Deshalb wird man zu seiner Bestimmung eine Integration der DGL unbedingt vermeiden, insbesondere wenn x^* kein Attraktor bzw. Repellor ist. Wir gehen vielmehr direkt von

Definition 5.2 aus und halten vorerst den Kontrollparameter $\lambda^* \in \Lambda = [a, b]$ fest. Damit sind Gleichgewichtspunkte nun als Lösungen eines parameterlosen Gleichungssystems

$$f(x) = 0, \quad f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n \quad (5.4)$$

gesucht. Der Einfachheit halber werde vorausgesetzt, dass f auf D mindestens zweimal stetig differenzierbar ist.

5.1.1 Newton-Verfahren

Der bekannteste Lösungsansatz besteht darin, die Funktion f an einer zuvor ermittelten Näherungslösung x_k zu linearisieren. Entwickeln wir dazu $f(x)$ an der Stelle x_k in die Taylor-Reihe

$$f(x) = f(x_k) + f'(x_k)(x - x_k) + R(x, x_k),$$

so liefert die rechte Seite bei Vernachlässigung des Restgliedes $R(x, x_k)$ die *Linearisierung der Funktion f* am Näherungspunkt x_k . Anstelle der Gleichung $f(x) = 0$ lösen wir nun die linearisierte Gleichung mit der Lösung $x = x_{k+1}$

$$f(x_k) + f'(x_k)(x_{k+1} - x_k) = 0,$$

und erhalten nach Umstellung das *Newton-Verfahren*¹ (engl. *Newton-Raphson method*)

$$x_{k+1} = x_k - [F(x_k)]^{-1}f(x_k), \quad k = 0, 1, 2, 3, \dots \quad (5.5)$$

mit der Jacobi-Matrix

$$F(x) = f'(x) = \left(\frac{\partial f_i}{\partial x_j} \right). \quad (5.6)$$

Das Verfahren nimmt eine iterative Linearisierung der gegebenen Funktion an den berechneten Näherungswerten x_k vor. Dieser einfache Zugang, im eindimensionalen Falle auch als *Tangenten-Näherungsverfahren* bekannt, bildet zugleich eines der grundlegendsten und leistungsfähigsten Näherungsverfahren für differenzierbare Abbildungen.

Bevor wir seine Konverenzeigenschaften näher betrachten, wollen wir eine effiziente algorithmische Darstellung angeben. Die Implementation des Newton-Verfahrens erfordert die Berechnung von $f(x_k)$ und der Jacobi-Matrix $F(x_k)$ sowie die Bestimmung der Näherung x_{k+1} . Nehmen wir an, die Jacobi-Matrix sei voll besetzt, und direkte Verfahren werden zur Lösung der linearen Gleichungssysteme eingesetzt. In jedem Schritt des Newton-Verfahrens ist dann die Invertierung der Jacobi-Matrix $F(x_k)$ erforderlich.

¹Sir Isaac Newton (1643–1727), engl. Physiker, Mathematiker und Astronom; Entdeckung der Gravitation; Newtonsche Axiome; Grundlagen der Differenzial- und Integralrechnung.

Der beträchtliche arithmetische Aufwand lässt sich jedoch auf $\frac{1}{3}$ reduzieren, wenn man stattdessen in jedem Schritt ein lineares Gleichungssystem löst. Das erreichen wir, wenn wir die Verfahrensgleichung (5.5) nach der Newton-Korrektur $d_k = x_{k+1} - x_k$ umstellen und mit der Jacobi-Matrix $F(x_k)$ multiplizieren. So entsteht eine *praktikablere Form des Newton-Verfahrens*:

Löse für $k = 0, 1, 2, \dots$ das lineare Gleichungssystem

$$F(x_k) d_k = -f(x_k), \quad x_{k+1} = x_k + d_k. \quad (5.7)$$

Der *arithmetische Aufwand* pro Newton-Schritt besteht nun in der Lösung eines linearen Gleichungssystems mit n Unbekannten, wofür die LU-Zerlegung oder QR-Zerlegung eingesetzt werden kann (vgl. Hoffmann et al. (2005)). Dafür sind $\mathcal{O}(n^3)$ Gleitpunktoperationen (*floating point operations, flops*) erforderlich. Allerdings sind die n Funktionswerte

ALGORITHMUS 5.3 (Newton-Verfahren)

Function $[x] = \text{newton}(f, F, x_0, \text{tolabs}, \text{tolrel})$

1. Berechne $x = x_0$, $\text{tol} = \text{tolrel} \cdot \|f(x)\| + \text{tolabs}$
2. Do while $\|f(x)\| > \text{tol}$
 1. Berechne Jacobi-Matrix $F(x)$
 2. Zerlege $F(x) = LU$
 3. Löse $LU \cdot d = -f(x)$
 4. $x = x + d$
 5. Berechne $f(x)$
3. Return x

von $f(x_k)$ und die n^2 partiellen Ableitungen von $F(x_k)$ ebenfalls zu berechnen, was zu einem beträchtlichen Funktionsaufwand des Newton-Verfahrens führen kann. Der zugehörige Algorithmus 5.3 erfordert als Input die Funktionen f und F , einen Startwert x_0 und die (absolute und relative) Toleranz $\text{tolabs}, \text{tolrel}$.

Wir wollen nun die Frage behandeln, unter welchen Bedingungen das Newton-Verfahren gegen eine Nullstelle x^* von f konvergiert. Entsprechend den angenommenen Voraussetzungen unterscheidet man drei Formen, die *lokale, semilokale und globale Konvergenz*:

- Wenn vorausgesetzt wird, dass eine Lösung x^* existiert und der Konvergenzsatz dann garantiert, dass eine (Kugel-)Umgebung $K = K[x^*; \delta]$ dieses Punktes existiert, so dass das Iterationsverfahren für alle $x_0 \in K$ gegen x^* konvergiert, so heißt dieses Verfahren *lokal konvergent*. Über die Lage und Größe der Umgebung K ist in praxi

meist nichts bekannt; man weiß nur, dass für hinreichend nahe der Lösung liegende Startpunkte das Verfahren konvergieren wird.

- Muss die Existenz von x^* nicht vorausgesetzt werden und können die Voraussetzungen eines Konvergenzsatzes bereits am Startpunkt x_0 verifiziert werden, so heißt das betreffende Verfahren *semilokal konvergent*. Dazu ist häufig eine Menge D zu beschreiben, so dass für Startpunkte $x_0 \in D$ die Iteration konvergiert. Satz 5.9 liefert ein Beispiel für eine semilokale Konvergenzaussage.
- Ist $D \subset \mathbb{R}^n$ ein im Allgemeinen großer *vorgegebener* Bereich (z. B. eine Kugelumgebung von 0, ein n -dimensionales endliches oder unendliches Intervall), so heißt ein Iterationsverfahren *global konvergent auf D* , falls es für jeden Startpunkt $x_0 \in D$ gegen eine Lösung x^* in D konvergiert. Verfahren dieser Art werden in Abschnitt 5.1.4 behandelt.

Wir wollen zuerst einen lokalen Konvergenzsatz für das Newton-Verfahren formulieren. Dazu betrachten wir Gleichung (5.21) und setzen die Existenz einer regulären Lösung $x^* \in D$ voraus.

Definition 5.4 (Reguläre Lösung)

Eine Lösung $x^* \in D$ heißt *regulär (isoliert)*, wenn

- eine Kugel $K = K[x^*; \delta^*] = \{x \mid \|x - x^*\| \leq \delta^*\}$ um x^* mit $K \in \text{int}(D)$ existiert,
- die Jacobi-Matrix $F(x) = f'(x)$ auf K Lipschitz-stetig ist und
- die Jacobi-Matrix $F(x^*)$ an der Lösung regulär ist.

Die Regularität der Nullstelle stellt in der Tat eine *Standardvoraussetzung* an das zu lösende Problem dar. Bei singularer Jacobi-Matrix $F(x^*)$ ist das Verhalten des Newton-Verfahrens überaus kompliziert: Während in einigen Fällen eine lineare Konvergenz eintritt (z. B. bei skalaren Nullstellenproblemen), versagt das Verfahren in höherdimensionalen Systemen häufig.

Bemerkung 5.5

1. Eine Lösung x^* ist *geometrisch isoliert*, falls eine Umgebung K existiert, in der keine weitere Lösung $x^{**} \neq x^*$ liegt. Die Regularität einer Nullstelle darf deshalb nicht mit deren *geometrischer Isoliertheit* verwechselt werden. Aus der Regularität folgt im übrigen stets die geometrische Isoliertheit einer Lösung.

2. Falls $f \in C^2(D)$ auf der gesamten Menge D ist, so lässt sich leicht zeigen, dass die Bedingung $\text{rang } F(x^*) = n$ bzw. $\det(F(x^*)) \neq 0$ hinreichend für die Regularität einer Lösung $x^* \in \text{int}(D)$. ■

Beispiel 5.6

Für das System mit zweimal stetig differenzierbarem $f \in \mathbb{R}^2$

$$\begin{aligned} f_1(x_1, x_2) &= (x_1 - x_2)^2 = 0 \\ f_2(x_1, x_2) &= x_1 + x_2 - 2 = 0 \end{aligned}$$

erhält man die Jacobi-Matrix

$$F(x) = \begin{pmatrix} 2(x_1 - x_2) & -2(x_1 - x_2) \\ 1 & 1 \end{pmatrix}.$$

An der Lösung $x^* = (1, 1)$ hat deren Determinante den Wert

$$\det(F(x^*)) = \begin{vmatrix} 0 & 0 \\ 1 & 1 \end{vmatrix} = 0,$$

weshalb x^* nicht regulär ist. Modifiziert man die Gleichungen zu

$$f_1(x_1, x_2) = (x_1 - x_2)^2 - 4 = 0$$

$$f_2(x_1, x_2) = x_1 + x_2 - 2 = 0,$$

so ändert man die Jacobi-Matrix $F(x)$ damit nicht. Für die Lösung $x^* = (2, 0)$ besitzt sie nun die Determinante

$$\det(F(x^*)) = \begin{vmatrix} 4 & -4 \\ 1 & 1 \end{vmatrix} = 8 \neq 0,$$

also ist diese Lösung regulär. ■

Vergleichen wir das Konvergenzverhalten des Newton-Verfahrens in diesem Beispiel miteinander, so lässt sich im regulären Fall eine wesentlich höhere Geschwindigkeit erkennen. Wir führen deshalb eine Konvergenzordnung ein und betrachten dazu eine Folge von Näherungen (x_k) , die mit einer allgemeinen Iteration $x_{k+1} = g(x_k)$ erzeugt wird.

Definition 5.7 (Q-Ordnung)

- i. Die Folge (x_k) konvergiert *mindestens mit der Q-Ordnung* $p \geq 1$, falls ein Konvergenzfaktor $Q_+ > 0$ existiert, so dass für alle $k \geq k_0$, $k_0 \in \mathbb{N}$,

$$\|x_{k+1} - x^*\| \leq Q_+ \|x_k - x^*\|^p$$

gilt. Falls $p = 1$ ist, so wird $Q_+ < 1$ gefordert.

- ii. Die Folge (x_k) konvergiert *höchstens mit der Q-Ordnung* $p \geq 1$, falls ein Konvergenzfaktor $Q_- > 0$ existiert, so dass für alle $k \geq k_0$, $k_0 \in \mathbb{N}$, gilt:

$$\|x_{k+1} - x^*\| \geq Q_- \|x_k - x^*\|^p.$$

- iii. Die Folge (x_k) konvergiert *genau mit der Q-Ordnung* $p \geq 1$, falls sie mindestens und höchstens mit Q-Ordnung p konvergiert.
- iv. Das Iterationsverfahren $x_{k+1} = g(x_k)$ konvergiert bezüglich x^* *mindestens mit der Q-Ordnung* p , wenn jede damit erzeugte Folge (x_k) mindestens mit Q-Ordnung p konvergiert.
- v. Existiert eine Folge (x_k) genau mit Q-Ordnung p , so konvergiert auch das Iterationsverfahren $x_{k+1} = g(x_k)$ bezüglich x^* *genau mit Q-Ordnung* p .

Verfahren der Q-Ordnungen 1, 2 bzw. 3 werden als *linear, quadratisch bzw. kubisch konvergent* bezeichnet. Definition 5.7 setzt allerdings nicht voraus, dass die Fehler $\|x_k - x^*\|$ verschieden von Null sind. Tritt dieser Fall für ein $k = k_1 > k_0$ bei einem Verfahren ein, das mit *genau* der Q-Ordnung p konvergiert, so verschwinden alle Fehler der Iterationen für $k \geq k_1$. Andererseits bedeutet $\|x_k - x^*\| > 0$ für ein $k \geq k_0$, dass alle weiteren Fehler ungleich Null sind. Wir setzen desweiteren diesen nichttrivialen Fall und ein Verfahren der genauen Q-Ordnung p voraus. Existiert der Grenzwert

$$q := \lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^p},$$

so bezeichnet man ihn als *asymptotischen Konvergenzfaktor*. Falls jedoch $q = 0$ ist, so erweist sich das Verfahren als *schneller als mit Q-Ordnung p konvergent*. Insbesondere heißen Iterationsverfahren mit

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0$$

überlinear konvergent.

Den beträchtlichen Vorteil eines Verfahrens mit Konvergenzordnung $p > 1$ verdeutlicht man am besten, wenn man die Definition 5.7 rekursiv anwendet, was zur Darstellung

$$\|x_k - x^*\| \leq Q_+^{\frac{p^k - 1}{p - 1}} \|x_0 - x^*\|^p \quad (5.8)$$

führt. Bei Vorliegen linearer Konvergenz hätte sich dagegen

$$\|x_k - x^*\| \leq Q_+^k \cdot \|x_0 - x^*\| \quad (5.9)$$

ergeben. Angenommen, in beiden Fällen liege derselbe Anfangsfehler $\|x_0 - x^*\| = 1.0$ vor und die Konvergenzkonstante sei $Q_+ = 0.1$. Dann schätzt man den Fehler des linear konvergenten Verfahrens durch

$$\|x_k - x^*\| \leq 10^{-k}$$

ab, d. h. pro Iterationsschritt gewinnt man durchschnittlich eine Dezimalziffer an Genauigkeit. Für ein quadratisch konvergentes Verfahren dagegen erhält man

$$\|x_k - x^*\| \leq 10^{-2^k},$$

so dass sich in jedem Iterationsschritt die Zahl der richtigen Ziffern verdoppelt!

Wir können nun einen lokalen Konvergenzsatz für das Newton-Verfahren angeben. Den umfangreichen Beweisgang dazu findet man z. B. bei Hoffmann et al. (2005).

Satz 5.8 (Lokale Konvergenz)

$f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ besitze die reguläre Lösung $x^* \in D$. Dann existieren Konstanten $\delta > 0$ und $Q > 0$ mit $Q\delta < 1$ und eine Kugel $K = K[x^*; \delta]$, so dass folgende Behauptungen gelten:

- i. Das Newton-Verfahren ist für jeden Startwert $x_0 \in K$ durchführbar mit $x_k \in K$ für alle $k \geq 0$.
- ii. $\lim_{k \rightarrow \infty} x_k = x^*$.
- iii. x^* ist die einzige Lösung in der Kugel K .
- iv. Die Konvergenz ist mindestens Q-quadratisch, d. h.

$$\|x_{k+1} - x^*\| \leq Q \|x_k - x^*\|^2 \quad \forall k \geq k_0.$$

Da sich die Regularitätsvoraussetzung auf eine im allgemeinen unbekannte Lösung x^* bezieht, ist sie in der Praxis meist nicht überprüfbar. Geeigneter sind deshalb semilokale Sätze, die entsprechende Voraussetzungen an die vom Anwender des Verfahrens festgelegte *Startlösung* x_0 fordern. Wir geben den bekanntesten Satz an; einen vollständigen Beweis findet man z. B. in Isaacson und Keller (1972).

Satz 5.9 (L. V. Kantorovics)

Für die Startlösung x_0 des Newton-Verfahrens seien folgende *Voraussetzungen* erfüllt:

- i. Es existiert die inverse Matrix $F(x_0)^{-1}$ mit $\|F(x_0)^{-1}\| \leq \alpha$.
- ii. Für die Differenz der ersten beiden Näherungen gelte

$$\|x_1 - x_0\| = \|F(x_0)^{-1}f(x_0)\| \leq \beta.$$

- iii. f sei 2-mal stetig differenzierbar mit

$$\sum_{k=1}^n \left| \frac{\partial^2 f_i(x)}{\partial x_j \partial x_k} \right| \leq \frac{\gamma}{n}, \quad i, j = 1(1)n,$$

für alle $x \in K = K[x_0; 2\beta] = \{x \mid \|x - x_0\| \leq 2\beta\}$.

- iv. $\alpha\beta\gamma \leq 1/2$.

Dann gelten folgende *Behauptungen*:

- i. Das Newton-Verfahren ist für den Startwert $x_0 \in K$ unbeschränkt *durchführbar* mit $x_k \in K$ für alle $k \geq 0$.
- ii. Die Folge (x_k) konvergiert gegen ein x^* mit $f(x^*) = 0$.
- iii. Für den Fehler gilt

$$\|x_k - x^*\| \leq \frac{2\beta}{2^k} \quad k = 0, 1, 2, \dots$$

Falls in Voraussetzung iv. sogar $\alpha\beta\gamma < 1/2$ gilt, so kann zusätzlich nachgewiesen werden, dass die Folge (x_k) mindestens Q-quadratisch konvergiert.

5.1.2 Vereinfachte Newton-Verfahren

Für großdimensionale Systeme ist der Aufwand pro Iterationsschritt auch in der praktikablen Form 5.7 des Newton-Verfahrens zu hoch. Man approximiert die Jacobi-Matrix $F(x_k)$ dann durch eine geeignete konstante Matrix A . Im einfachsten Fall kann man $A = F(x_0)$ wählen und erhält so das *vereinfachte Newton-Verfahren* (*Sehnenverfahren*, engl. *chord method*):

- (1) Berechne die Matrix $A = F(x_0)$ und deren LU-Zerlegung.
- (2) Löse für $k = 0, 1, 2, \dots$ das lineare Gleichungssystem

$$A d^k = -f(x_k), \quad x_{k+1} = x_k + d^k. \quad (5.10)$$

Offenbar ist nun jeder Iterationsschritt wesentlich effizienter, denn der arithmetische Aufwand für eine Vorwärts-Rückwärts-Elimination ist nur von der Ordnung $\mathcal{O}(n^2)$, und es sind lediglich n Funktionswerte in $f(x_k)$ zu berechnen. Dem steht eine langsamere Konvergenz gegenüber. Im Gegensatz zum Newton-Verfahren kann für das Sehnenver-

ALGORITHMUS 5.10 (Sehnenverfahren)

Function $[x] = \text{newton0}(f, F, x_0, \text{tolabs}, \text{tolrel})$

1. Berechne $x = x_0$, $\text{tol} = \text{tolrel} \cdot \|f(x)\| + \text{tolabs}$
2. Berechne Jacobi-Matrix $F(x)$
3. Zerlege $F(x) = LU$
4. Do while $\|f(x)\| > \text{tol}$
 1. Löse $LU \cdot d = -f(x)$
 2. $x = x + d$
 3. Berechne $f(x)$
5. Return x

fahren nur lineare Konvergenz erwartet werden. Um die Konvergenz des Verfahrens zu garantieren, wird wie im Falle des Newton-Verfahrens die Regularität der Nullstelle x^* vorausgesetzt (vgl. Kelley (1995)).

Satz 5.11 (Lokale Konvergenz)

$f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ besitze die reguläre Nullstelle $x^* \in D$. Dann existieren Konstanten $\delta > 0$ und $Q_1 > 0$, so dass folgende Behauptungen gelten:

- i. Das Sehnungsverfahren ist für jeden Startwert $x_0 \in K = K[x^*; \delta]$ durchführbar mit $x_k \in K$ für alle $k \geq 0$.
- ii. $\lim_{k \rightarrow \infty} x_k = x^*$.
- iii. Die Konvergenz ist Q-linear mit der Fehlerschätzung

$$\|x_{k+1} - x^*\| \leq Q_1 \|x_0 - x^*\| \|x_k - x^*\|, \quad k = 0, 1, 2, \dots \quad (5.11)$$

Algorithmus 5.10 erfordert als Input die Funktionen f und F , einen Startwert x_0 und die (absolute und relative) Toleranz $tolabs, tolrel$. Auch diesen Algorithmus kann man leicht als MATLAB-Funktion implementieren.

Falls sich die Jacobi-Matrix $F(x_k)$ von Iteration zu Iteration nur wenig ändert, empfehlen sich mehrere Zwischenschritte mit ein- und derselben Matrix A . Dafür ist jedesmal nur eine LU-Zerlegung erforderlich. Man erhält so eine Kombination aus Newton- und Sehnungsverfahren, die häufig als *Shamanskii-Verfahren* bezeichnet wird. Es lässt sich als eine Variante des Newton-Verfahrens mit m inneren Iterationsschritten interpretieren:

Für $k = 0, 1, 2, \dots$ iteriere

$$\begin{aligned} y_0 &= x_k - [F(x_k)]^{-1} f(x_k), \\ y_{j+1} &= y_j - [F(x_k)]^{-1} f(y_j), \quad j = 0(1)m - 1 \\ x_{k+1} &= y_m. \end{aligned} \quad (5.12)$$

Das Newton-Verfahren liegt offenbar im Sonderfall $m = 0$ vor, während im Grenzfall $m \rightarrow \infty$ dieses Verfahren in das Sehnungsverfahren übergeht.

Nimmt man die Werte y_j der inneren Iteration als Hilfsgrößen an und betrachtet lediglich die Folge (x_k) der äußeren Iterierten, so kann man unter geeigneten Voraussetzungen nachweisen, dass diese Folge mindestens mit Q-Ordnung $m + 2$ gegen die Lösung x^* konvergiert. So kann mit relativ geringem Zusatzaufwand (eine Berechnung von f und eine Vorwärts-Rückwärts-Elimination) eine höhere Konvergenzordnung als beim gewöhnlichen Newton-Verfahren erreicht werden.

Der Algorithmus 5.12 erfordert als Input ebenfalls die Funktionen f und F , einen Startwert x_0 und die (absolute und relative) Toleranz $tolabs, tolrel$. Hinzu kommt die Anzahl m der inneren Iterationen. Sämtliche Iterationswerte können in ein- und derselben Variablen x gespeichert werden. Wird die geforderte Genauigkeit bereits in einer inneren Iteration erreicht, so kann der entsprechende Iterationswert zurückgegeben werden.

Beispiel 5.13

Wir betrachten die Gleichungen

$$f_1(x_1, x_2) = 1 - x_1^2 - x_2^2 = 0, \quad f_2(x_1, x_2) = -2x_1 + x_2 = 0$$

ALGORITHMUS 5.12 (Shamanskii-Verfahren)Function $[x] = \text{shamanskii}(f, F, x_0, \text{tolabs}, \text{tolrel}, m)$

1. Berechne $x = x_0$, $\text{tol} = \text{tolrel} \cdot \|f(x)\| + \text{tolabs}$
2. Do while $\|f(x)\| > \text{tol}$
 1. Berechne Jacobi-Matrix $F(x)$
 2. Zerlege $F(x) = LU$
 3. For $j = 0(1)m$
 1. Löse $LU \cdot d = -f(x)$
 2. $x = x + d$
 3. Berechne $f(x)$
 4. If $\|f(x)\| < \text{tol}$ Return x
3. Return x

und wenden (a) das Newton-Verfahren, (b) das Shamanskii-Verfahren mit den inneren Iterationszahlen $m = 1, 2, 5, 10$ und (c) das Sehnungsverfahren mit denselben Parametersätzen an. In Tabelle 5.3 wird die Anzahl k aller Iterationen des Newton-, des Shamanskii- und des Sehnungsverfahrens gegenübergestellt. In den mit * markierten Fällen wurde die geforderte Genauigkeit tol nicht vollständig erreicht. Beachtet man jedoch, dass das Newton-Verfahren in dritten Aufruf $k = 6$ Berechnungen der Jacobi-Matrix und einen arithmetischen Aufwand von $k \cdot \frac{2}{3}n^3$ erfordert, so liefert das Shamanskii-Verfahren die Lösung bereits für $k = 16$, $m = 10$ mit $\lfloor \frac{k-1}{m+1} \rfloor + 1 = 2$ Jacobi-Matrizen und einem arithmetischen Aufwand von $2 \cdot \frac{2}{3}n^3$. ■

Tab. 5.1 Anzahl der benötigten Iterationen zu Beispiel 5.13.

tolabs tolrel	Startpunkt (x_1, x_2)	Newton- Verfahren	Shamanskii-Verfahren				Sehn- verfahren
			$m = 1$	$m = 2$	$m = 5$	$m = 10$	
10^{-6}	(3,1.7)	5	7	7	10	11*	28*
10^{-12}	(3,1.7)	6	8	9	13	16	50*
10^{-12}	(-3,-2)	6	8	10	13	16	50*

Der Vorteil dieses Verfahrens tritt besonders bei großdimensionalen Gleichungssystemen stärker hervor. Es kann zudem verbessert werden, wenn die Anzahl m ergebnisabhängig gewählt und gegebenenfalls adaptiv geändert wird. Das kann durch Kontrolle der Fehlerentwicklung der inneren Iteration erfolgen (vgl. Kelley (1995)).

5.1.3 Newton-Verfahren mit Differenzenquotienten

Häufig sind die Ableitungen von f nur mit erheblichem Aufwand exakt zu berechnen. Dann kann man das für skalare Gleichungen bekannte Sekantenverfahren verallgemeinern und approximiert die Jacobi-Matrix $F(x_k) = f'(x_k)$ durch eine leichter zu berechnende Matrix $A_k \in \mathbb{R}^{n \times n}$ mittels Differenzenquotienten

$$x_{k+1} = x_k - A_k^{-1} f(x_k), \quad k = 0, 1, 2, \dots \quad (5.13)$$

und einem geeigneten Schrittweitenvektor $h = (h_1, \dots, h_n)^T$, $h_j > 0$. Derartige Verfahren bezeichnet man oft als *Newton-ähnlich* (*Newton-like*), wenn die erzeugten Folgen (x_k) , $k = 0, 1, 2, \dots$, das Verhalten der Newton-Folge aus Verfahren 5.3 approximieren. Besitzt die Gleichung $f(x) = 0$ die reguläre Lösung $x^* \in D \subset \mathbb{R}^n$, so definiert man genauer:

Definition 5.14 (Newton-ähnliches Verfahren)

Eine konvergente Näherungsfolge (x_k) , $k = 0, 1, 2, \dots$, mit $x_k \neq x^*$ in einer Umgebung S von x^* heißt *Newton-ähnlich*, wenn

$$\lim_{k \rightarrow \infty} \frac{\|f(x_k) + f'(x_k)(x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} = 0$$

gilt. Ist jede Folge mit $x_0 \in S$ Newton-ähnlich, so bezeichnet man das Verfahren selbst als *Newton-ähnlich*.

Für das Newton-Verfahren selbst trifft die Definition offenbar zu. Man kann unter der Voraussetzung einer regulären Lösung x^* nachweisen, dass ein Verfahren genau dann überlinear konvergiert, wenn es Newton-ähnlich ist (vgl. Schwetlick (1979)).

Betrachten wir nun Verfahren der Gestalt (5.13) und stellen die Matrix A_k durch Differenzenquotienten dar. Als einfachste Approximation A_k bietet sich die Einpunkt-Approximation mittels der Vorwärts-Differenzenquotienten von $f'(x)$

$$\{\nabla f(x, h)\}_{ij} = \frac{1}{h_j} [f_i(x_1, \dots, x_j + h_j, \dots, x_n) - f_i(x)], \quad i, j = 1 \dots n,$$

an, die mit dem j -ten Einheitsvektor e_j spaltenweise als

$$\nabla f(x, h)e_j = \frac{1}{h_j} [f(x + h_j e_j) - f(x)], \quad j = 1 \dots n \quad (5.14)$$

notiert werden kann. Zuerst ist zu klären, unter welchen Bedingungen das entstehende Verfahren (5.13) durchführbar ist und die erzeugte Folge der Näherungslösungen (x_k) , $k = 0, 1, 2, \dots$ gegen eine Lösung x^* konvergiert. Da anstelle der einfachen Differenzenquotienten (5.14) auch andere Approximationen denkbar sind, soll folgende Verallgemeinerung eingeführt werden:

Definition 5.15 (Streng konsistente Approximation)

$f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ sei auf $D_0 \subset \text{int } D$ differenzierbar. Die Schrittweitenmenge sei

$$H = \{h = (h_1, h_2, \dots, h_n) \in \mathbb{R}^n \mid h_i \neq 0, i = 1, \dots, n, \|h\| \leq r\}.$$

Dann heißt die Abbildung $\nabla f : D_0 \times H \rightarrow \mathbb{R}^{n \times n}$ *streng konsistente Approximation* für $f'(x)$ auf D_0 mit dem Diskretisierungsbereich $H \subset \mathbb{R}^n$, wenn $0 \in \mathbb{R}^n$ Häufungspunkt von H ist und eine Konstante $C > 0$ existiert, so dass

$$\|\nabla f(x, h) - f'(x)\| \leq C \cdot \|h\| \quad (5.15)$$

für alle $(x, h) \in D_0 \times H$ gilt.

Mit der Voraussetzung $f \in C^2(\mathbb{R}^n)$ erfüllt unsere Differenzenapproximation (5.14) offenbar die Bedingung (5.15), d. h. $\lim_{h \rightarrow 0} \nabla f(x, h) = f'(x)$ mit Ordnung 1 in h . Bei Schwetlick (1979) wird dann die Konvergenz des Verfahrens (5.13) mit folgendem Satz nachgewiesen:

Satz 5.16 (Konvergenz)

$f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ besitze die reguläre Lösung $x^* \in D$, und $\nabla f : D_0 \times H \rightarrow \mathbb{R}^{n \times n}$ sei eine streng konsistente Approximation von f' auf einer Umgebung D_0 von x^* . Dann existieren Konstanten $\delta_0 > 0$ und $\varepsilon > 0$, so dass folgende Behauptungen gelten:

- i. Das Verfahren (5.13) ist für jedes $x_0 \in K_0 := K[x^*; \delta_0]$ durchführbar, und $x_k \in K_0$ für $k = 0, 1, 2, \dots$, wenn in jedem Schritt $h_k \in H$ so gewählt wird, dass $\|h_k\| < \varepsilon$ garantiert ist.
- ii. Die Folge (x_k) konvergiert mindestens linear gegen x^* .
- iii. Falls zusätzlich $\lim_{k \rightarrow \infty} h_k = 0$ gilt, so ist die Konvergenz der x_k überlinear.

Damit ist das Verfahren Newton-ähnlich im Sinne der angegebenen Definition, falls die Schrittweitenfolge (h_k) gegen Null konvergiert. In praxi können allerdings wegen der begrenzten Stellenzahl der benutzten Gleitpunktzahlen die Schrittweiten h_j der Differenzenapproximation (5.14) nicht beliebig nahe bei Null gewählt werden, sondern müssen auf jeden Fall größer als die relative Maschinengenauigkeit ε_M sein. Damit erfüllt die reale Schrittweitenmenge H_{real} nicht die Voraussetzung aus Definition 5.15 und der Konvergenzsatz ist nur bedingt aussagekräftig. Beantworten wir deshalb die Frage, wie die Diskretisierungsschrittweiten h_j in $h = (h_1, h_2, \dots, h_n)$ passend zu wählen sind. Neben dem Diskretisierungsfehler $f'(x) - \nabla f(x, h)$ sind dazu nun auch andere Fehlerquellen, insbesondere Rundungsfehler, zu berücksichtigen. Wird nämlich anstelle von $f(x)$ der fehlerbehaftete Funktionswert $\tilde{f}(x) = f(x) + \varepsilon(x)$, $\|\varepsilon(x)\| < \varepsilon_M$ mit der Maschinengenauigkeit ε_M benutzt, so erhält man mit der Voraussetzung $f \in C^2(\mathbb{R}^n)$

ALGORITHMUS 5.17 (Newton-Verfahren mit Differenzen)

Function $[x] = \text{newtondiff}(f, x_0, \text{tolabs}, \text{tolrel})$

1. Berechne $x = x_0$, $\text{tol} = \text{tolrel} \cdot \|f(x)\| + \text{tolabs}$
2. Do while $\|f(x)\| > \text{tol}$
 1. Wähle $h_j = \sqrt{\varepsilon_M}(1 + |x_j|)$, $j = 1(1)n$
 2. Approximiere die Jacobi-Matrix durch A_k mit

$$A_k e_j = \frac{1}{h_j} [f(x + h_j e_j) - f(x)], \quad j = 1(1)n$$
 3. Löse $A_k \cdot d = -f(x)$ nach d
 4. $x = x + d$
 5. Berechne $f(x)$
3. Return x

für den gesamten Approximationsfehler der Differenzenapproximation die Abschätzung

$$\begin{aligned} \|f'(x)e_j - \nabla \tilde{f}(x, h)e_j\| &= \|f'(x)e_j - \frac{1}{h_j}[\tilde{f}(x + h_j e_j) - \tilde{f}(x)]\| \\ &\leq \|f'(x)e_j - \frac{1}{h_j}[f(x + h_j e_j) - f(x)]\| + \frac{2 \max \|\varepsilon(x)\|}{h_j} \\ &\leq Ch_j + \frac{2\varepsilon_M}{h_j} \end{aligned}$$

mit einer funktionsabhängigen Konstanten $C > 0$. Minimierung der rechten Seite über h_j ergibt den Minimalwert $h_j = \sqrt{2\varepsilon_M/C}$, womit $h = \mathcal{O}(\sqrt{\varepsilon_M})$ gilt. Um für betragsgroße Werte x_j auch entsprechend große Diskretisierungsschrittweiten h_j zu erhalten, hat sich eine kombinierte Absolut-Relativ-Wahl

$$h_j = \sqrt{\varepsilon_M}(1 + |x_j|), \quad j = 1(1)n \quad (5.16)$$

gut bewährt. Damit wird zugleich garantiert, dass für Werte von x_j nahe Null keine Stellenauslöschung in den Differenzenquotienten erfolgt. Das entstehende Verfahren ist allerdings nun nicht mehr Newton-ähnlich im engeren Sinne der Definition.

5.1.4 Gedämpfte Newton-Verfahren und globale Konvergenz

Die bisher behandelten Verfahren sind lokal konvergent, da unter geeigneten Voraussetzungen (Glattheit, Regularität) stets eine Konvergenz Umgebung S der Nullstelle x^* der Gleichung

$$f(x) = 0, \quad f: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad D \text{ offen} \quad (5.17)$$

existiert. Über die Lage und Größe der Umgebung S ist in praxi meist nichts bekannt. Um den semilokalen Konvergenzsatz 5.9 von L.V. Kantorovics anwenden zu können, sind schwer nachzuvollziehende Abschätzungen in der Umgebung des Startwertes x_0 durchzuführen, der zudem nahe der Lösung x^* liegen sollte.

Die Menge $S(x^*)$ aller Startwerte x_0 , für die das jeweils betrachtete Verfahren gegen die Nullstelle x^* konvergiert, bildet den *Einzugsbereich von x^** . Um ihn zu vergrößern, kann man zur Bestimmung der Nullstelle x^* von f auch das zugehörige Minimierungsproblem

$$\psi(x) = \frac{1}{2} \|f(x)\|^2 \Rightarrow \text{Min! über } x \in D \subset \mathbb{R}^n, \quad (5.18)$$

mit der Euklidischen Norm $\|x\|$ betrachten. Gesucht ist dann eine Minimalstelle x^* mit $\psi(x^*) = 0$. Ausgehend von einer k -ten Lösungsnaherung x_k lässt sich der neue Wert $x_{k+1} = x_k + d_k$ mit der Newton-Richtung (bei Minimierungsverfahren auch „Such-Richtung“ genannt) $d_k = -f'(x_k)^{-1}f(x_k)$ ermitteln und anschließend kontrollieren, ob die *einfache Abstiegsbedingung*

$$\psi(x_{k+1}) < \psi(x_k)$$

für die Funktion ψ erfüllt ist. Andernfalls kann man den Korrekturvektor d_k mit einem Faktor $\tau_k \in (0, 1]$ dämpfen und mit dem zurückgesetzten Testpunkt

$$x_{k+1} = x_k + \tau_k \cdot d_k \quad \text{mit} \quad d_k = -f'(x_k)^{-1}f(x_k) \quad (5.19)$$

die Abstiegsbedingung erneut überprüfen. Nach eventuell weiteren erforderlichen Reduzierungen $\tau_k := \tau_k \cdot \alpha$ mit dem fest gewählten Skalierungsfaktor $\alpha \in (0, 1)$ liefert der

ALGORITHMUS 5.18 (Gedämpftes Newton-Verfahren)

Function $[x, y] = \text{newtonarmijo}(f, F, x_0, \text{tolabs}, \text{tolrel})$

1. Berechne $x = x_0$, $\text{tol} := \text{tolrel} \cdot \|f(x)\| + \text{tolabs}$
2. Wähle Konstanten $\delta \in (0, \frac{1}{2})$, $\alpha \in (0, 1)$ und $\lambda := 1$
3. Do while $\|f(x)\| > \text{tol}$
 1. Wähle $\lambda := \min(\lambda/\alpha, 1)$
 2. Berechne Jacobi-Matrix $F(x)$
 3. Berechne Newton-Richtung d aus $F(x) \cdot d = -f(x)$
 4. Berechne Testpunkt $y := x + \lambda d$
 5. Do while $\|f(y)\|^2 > (1 - 2\delta\lambda)\|f(x)\|^2$
 1. Reduziere $\lambda := \lambda \cdot \alpha$
 2. Berechne Testpunkt $y := x + \lambda d$
6. Aktualisiere $x := y$
4. Return $x, y = f(x)$

erste akzeptierte τ_k -Wert den *Dämpfungsfaktor* des sogenannten *gedämpften Newton-Verfahrens* (5.19). Häufig wird eine sukzessive Halbierung mit $\alpha = 0.5$ vorgenommen. Bei Minimierungsverfahren hat es sich bewährt, diese einfache Abstiegsbedingung durch eine Bedingung für *hinreichenden Abstieg*

$$\psi(x_{k+1}) < (1 - 2\delta\tau_k) \psi(x_k), \quad \text{mit einer Konstanten } \delta \in (0, \frac{1}{2}) \quad (5.20)$$

zu ersetzen. Offenbar ist $1 - 2\delta\tau_k \in (0, 1)$, wobei Werte $\delta = 10^{-2}, 10^{-4}$ üblich sind. Der erste τ_k -Wert, mit dem diese Bedingung erfüllt ist, liefert ein gedämpftes Newton-Verfahren (5.19) mit der sogenannten *Armijo-Regel* (5.20), die in Algorithmus 5.18 implementiert wurde. Eine detaillierte Begründung dieser und weiterer verbesserter Strategien findet man bei Kosmol (1993).

Das Hauptziel, den Einzugsbereich von x^* zu vergrößern, kann mit dem gedämpften Newton-Verfahren oft erreicht werden. Ist $S \subset \mathbb{R}^n$ ein im Allgemeinen großer vorgegebener Bereich, so ist das Iterationsverfahren *global konvergent auf S*, falls es für jeden Startpunkt $x_0 \in S$ gegen einen Fixpunkt x^* in S konvergiert. Kann die Regularität der Jacobi-Matrix $f'(x)$ auf einer ganzen Niveaumenge vorausgesetzt werden, so lässt sich die globale Konvergenz des gedämpften Newton-Verfahrens für alle Startwerte aus dieser Menge nachweisen (vgl. Schwetlick (1979) zum aufwändigen Beweis des Satzes):

Satz 5.19 (Globale Konvergenz)

Es sei $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ 2-mal stetig differenzierbar auf der offenen Menge D . Zu einem festem $y \in D$ werde die Niveaumenge

$$S = \{x \in D \mid \|f(x)\| \leq \|f(y)\|\}$$

definiert, die als kompakt vorausgesetzt wird. Ist die Jacobi-Matrix $f'(x)$ für alle $x \in S$ invertierbar, so gelten folgende Behauptungen:

- i. Das gedämpfte Newton-Verfahren (5.19) mit Armijo-Regel (5.20) ist für jedes $x_0 \in S$ durchführbar mit $x_k \in S$ und streng monoton fallenden Funktionswerten $\|f(x_k)\|$.
- ii. Die Folge (x_k) konvergiert gegen eine Nullstelle x^* von f in der Menge S .
- iii. Falls das Verfahren nicht nach endlich vielen Schritten mit einer Nullstelle endet, so existiert ein $K \in \mathbb{N}$, bei dem das Verfahren für $k > K$ in das gewöhnliche Newton-Verfahren übergeht und Q-quadratische Konvergenz eintritt.

Die hier eingeführte Dämpfungsstrategie lässt sich nunmehr auch erfolgreich auf das vereinfachte Newton-Verfahren, das Shamanskii-Verfahren und das Newton-Verfahren mit Differenzenquotienten anwenden und dafür globale Konvergenz erreichen. Die komplizierte Konvergenztheorie übersteigt jedoch den Rahmen dieses Buches. Sehr gute und detaillierte Darstellungen des Stoffes findet man bei Deuffhard (2004) und Kelley (1995).

5.2 Parametrisierung von Lösungskurven und Fortsetzungsmethoden

Wir kehren nun zum eingangs genannten Problem (5.1) zurück und bestimmen parameterabhängige Gleichgewichtspunkte stets als Lösungen des Gleichungssystems

$$f(x, \lambda) = 0, \quad f : D \times \Lambda \subset \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n \quad (5.21)$$

mit skalarem Parameter $\lambda \in \Lambda = [a, b]$.

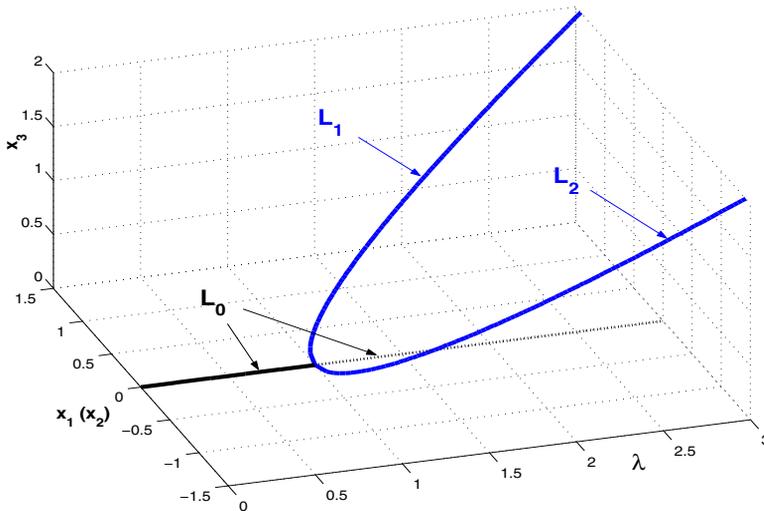


Abb. 5.1

Lösungszweige L_0 (schwarz) und L_1, L_2 (blau) im Lösungsdiagramm zu Beispiel 5.20

Beispiel 5.20

Für $\lambda \in [0, 3]$ berechnen wir alle Gleichgewichtspunkte des Lorenz-Systems (5.3) mittels

$$\begin{aligned} x_2 - x_1 &= 0 \\ \lambda x_1 - x_2 - x_1 x_3 &= 0 \\ x_1 x_2 - x_3 &= 0. \end{aligned}$$

Im Intervall $0 \leq \lambda \leq 1$ existiert lediglich die asymptotisch stabile triviale Lösung $L_0 : x_1 = x_2 = x_3 = 0$, während für $1 < \lambda \leq 3$ außer dieser Lösung L_0 die beiden nichttrivialen Lösungen

$$L_1 : x_1 = x_2 = \sqrt{\lambda - 1}, \quad x_3 = \lambda - 1 \quad \text{und} \quad L_2 : x_1 = x_2 = -\sqrt{\lambda - 1}, \quad x_3 = \lambda - 1$$

gefunden werden. Wegen $x_1 = x_2$ für alle Lösungen wählen wir eine dieser beiden Koordinaten aus und stellen die Lösungszweige im (λ, x_1, x_3) -Lösungsdiagramm der Abbildung 5.1 dar. Beim Durchlaufen des Parameterwertes $\lambda^* = 1$ tritt offenbar eine Lösungsverzweigung ein. Zugleich wird für $\lambda > \lambda^*$ die triviale Lösung L_0 instabil. ■

Abbildung 5.1 legt nahe, die Variablen x und λ zu einer einzigen Variablen $y = (x, \lambda)$ mit $y \in \mathbb{R}^{n+1}$ zusammenzufassen und System (5.21) als unterbestimmtes Gleichungssystem

$$f(y) = 0, \quad f: \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n, \quad f \in C^r(\mathbb{R}^{n+1}) \quad \text{mit } r \geq 1, \quad (5.22)$$

in \mathbb{R}^{n+1} zu behandeln. Um dessen Lösungsmenge \mathcal{M} numerisch bestimmbar zu machen, treffen wir eine Zusatzannahme mit folgender

Definition 5.21 (Regulärer Punkt, regulärer Wert)

- i. Ein Punkt $y \in \mathbb{R}^{n+1}$ heißt *regulärer Punkt* der glatten Abbildung f , wenn die Jacobi-Matrix $Df(y) \in \mathbb{R}^{(n+1) \times n}$ vollen Rang hat, d. h. $\text{rang}(Df(y)) = n$. Andernfalls heißt y *singulärer Punkt*.
- ii. Ein Bildpunkt $w \in \mathbb{R}^n$ heißt *regulärer Wert* von f , wenn die Menge

$$f^{-1}(w) := \{y \in D \times \Lambda \subset \mathbb{R}^{n+1} \mid f(y) = w\}$$

nur aus regulären Punkten von f besteht. Andernfalls heißt w *singulärer Wert*.

Wir nehmen desweiteren an, dass $0 \in \mathbb{R}^n$ ein regulärer Wert von f ist, also die Jacobi-Matrix $Df(y)$ vollen Rang für alle Lösungen von $f(y) = 0$ besitzt. Mit

$$\mathcal{M} = \{y \in \mathbb{R}^{n+1} \mid f(y) = 0, \quad \text{rang}(Df(y)) = n\} \quad (5.23)$$

bezeichnen wir die Menge aller Lösungen von (5.22), die reguläre Punkte von f sind.

Beispiel 5.22

Mit $y = (x_1, x_2, x_3, \lambda) = (y_1, y_2, y_3, y_4)$ lautet die Jacobi-Matrix des Lorenz-Systems

$$Df(y) = \begin{pmatrix} -1 & 1 & 0 & 0 \\ y_4 - y_3 & -1 & -y_1 & y_1 \\ y_2 & y_1 & -1 & 0 \end{pmatrix} \quad \text{wegen } f(y) = \begin{pmatrix} y_2 - y_1 \\ y_4 y_1 - y_2 - y_1 y_3 \\ y_1 y_2 - y_3 \end{pmatrix},$$

womit sich für den Lösungszweig $L_0: y_0 = (0, 0, 0, y_4)$ nur dann $\text{rang} Df(y_0) = 3$ ergibt, wenn $y_4 \equiv \lambda \neq 1$ ist. Der Punkt $y_0 = (0, 0, 0, 1)$ ist hingegen singulär. Also ist $0 \in \mathbb{R}^3$ kein regulärer Wert von f . Schränken wir den Parameterbereich jedoch auf $\Lambda = [1.5, 3]$ ein, so sind alle entsprechenden Lösungen reguläre Punkte von f . ■

5.2.1 Natürliche Parametrisierung

Ist für einen Lösungspunkt $y_0 = (x_0, \lambda_0)$ die Teilmatrix $D_x f(x_0, \lambda_0) \in \mathbb{R}^{n \times n}$ regulär, so kann der natürliche Parameter λ zur lokalen Parametrisierung der Lösungsmenge \mathcal{M} benutzt werden. Nach dem impliziten Funktionentheorem (vgl. Hoffmann et al. (2005)) existiert lokal eine Funktion $x: [\lambda_0 - \varepsilon, \lambda_0 + \varepsilon] \rightarrow \mathbb{R}^n$ mit $\varepsilon > 0$ und $(x(\lambda), \lambda) \in \mathcal{M}$. Ist $D_x f(x, \lambda)$ sogar regulär für alle $(x, \lambda) \in D \times \Lambda$, so ist die Funktion $x(\lambda)$ auf ganz Λ definiert. Wir treffen dazu folgende

Voraussetzung 5.23 (Regularität)

Es sei $D \subset \mathbb{R}^n$ offen und $\Lambda = [a, b]$. $f \in C^r(D \times \Lambda)$, $r \geq 2$, und $D_x f(x, \lambda)$ ist regulär für alle $(x, \lambda) \in D \times \Lambda$.

Mit dieser Regularitätsvoraussetzung gewinnen wir den für das Verfahren grundlegenden

Satz 5.24

Unter Voraussetzung 5.23 existiert die implizit definierte Funktion $x : \Lambda \rightarrow D$ mit den Eigenschaften:

- i. $f(x, \lambda) = 0$ besitzt für jedes $\lambda \in \Lambda$ genau eine Lösung $x = x(\lambda) \in D$.
- ii. Es gilt die Abschätzung $\|x - x(\lambda)\| \leq M_0 \cdot \|f(x, \lambda)\| \quad \forall (x, \lambda) \in D \times \Lambda$, $M_0 > 0$.
- iii. $x(\lambda)$ ist in $\lambda \in \Lambda$ stetig differenzierbar und genügt der Einbettungsgleichung

$$x'(\lambda) = -[D_x f(x(\lambda), \lambda)]^{-1} D_\lambda f(x(\lambda), \lambda). \tag{5.24}$$

Von einer direkten numerischen Integration dieser Einbettungsgleichung – mitunter auch als *Davidenko-DGL* bezeichnet – als Anfangswertproblem sollte wegen der komplizierten rechten Seiten und evtl. fehlender Lösungsstabilität Abstand genommen werden. Numerische Fortsetzungsmethoden nutzen vielmehr die Bestimmungsgleichungen (5.21) zur genauen Berechnung von $x(\lambda)$. Dazu unterteilen wir wie in Abbildung 5.2 das Parameterintervall $\Lambda = [a, b]$ in N Teilintervalle mit den Teilpunkten $a = \lambda_0 < \lambda_1 < \dots < \lambda_N = b$

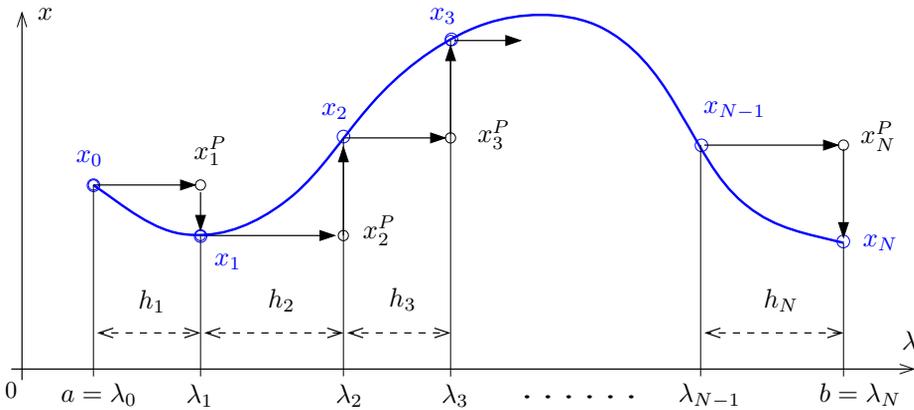


Abb. 5.2 Fortsetzung mit natürlicher Parametrisierung

und bestimmen sukzessive die Punkte $(x(\lambda_j), \lambda_j)$ der Lösungskurve

$$\mathcal{L} = \{(x, \lambda) | (x, \lambda) \in D \times \Lambda \text{ mit } f(x, \lambda) = 0\}, \tag{5.25}$$

beginnend mit $(x(a), a) = (x_0, \lambda_0)$ in der Reihenfolge $(x_0, \lambda_0), (x_1, \lambda_1), \dots, (x_N, \lambda_N)$. Dabei soll abkürzend $x_j = x(\lambda_j)$ notiert werden. Ein *Fortsetzungsschritt* von λ_{j-1} bis λ_j ist dann folgendermaßen aufgebaut:

1. Gegeben sei ein (zuvor ermittelter) Kurvenpunkt (x_{j-1}, λ_{j-1}) .
2. Festlegung einer Fortsetzungsschrittweite $h_j > 0$ und des neuen Parameterwertes $\lambda_j = \lambda_{j-1} + h_j$.
3. Prädiktorschritt: Vorgabe eines Näherungswertes x^P für den neuen Lösungsvektor $x_j = x(\lambda_j)$.
4. Korrektorschritt: Iterative Verbesserung des Prädiktorwertes mittels eines Newton-ähnlichen Verfahrens für Gleichung (5.21).

Prädiktorschritt

Für den *Prädiktorschritt* nutzen wir Informationen über zuvor bestimmte Kurvenpunkte und extrapolieren die Lösungskurve bis zu λ_j :

1. Der *Basis-Prädiktor* nimmt den vorhergehenden Kurvenpunkt als Startwert für den folgenden Korrektorschritt:

$$x^P := x_{j-1} = x(\lambda_{j-1}) \tag{5.26}$$

2. Der *Tangenten-Prädiktor* – auch *Euler-Prädiktor*² genannt – ersetzt die Kurve lokal durch ihre Tangente im vorhergehenden Kurvenpunkt

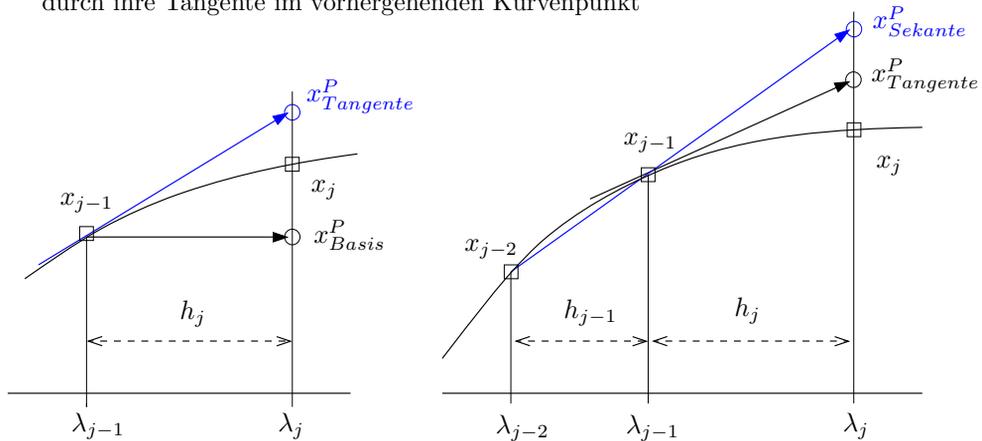


Abb. 5.3 Prädiktoren bei natürlicher Parametrisierung

$$x(\lambda) = x(\lambda_{j-1}) + (\lambda - \lambda_{j-1}) \cdot x'(\lambda_{j-1}), \tag{5.27}$$

wobei $x'(\lambda_{j-1})$ mit der Einbettungsgleichung (5.24) gewonnen werden kann. Wegen $x_{j-1} := x(\lambda_{j-1})$ ergibt sich so die Prädiktorformel

$$x^P := x_{j-1} - h_j \cdot [D_x f(x_{j-1}, \lambda_{j-1})]^{-1} D_\lambda f(x_{j-1}, \lambda_{j-1}).$$

Sie besitzt wie in Abbildung 5.3 links dargestellt eine höhere Genauigkeit als (5.26), erfordert allerdings die Ableitungen $D_x f$ und $D_\lambda f$ sowie die Lösung eines linearen Gleichungssystems.

²Leonhard Euler (1707–1783), Schweizer Mathematiker, arbeitete auf fast allen Gebieten der Mathematik; Leistungen zur Mechanik, Technik, Optik und Astronomie.

3. Legt man durch die zwei Knoten (x_{j-2}, λ_{j-2}) und (x_{j-1}, λ_{j-1}) die Sekante und extrapoliert bis zum Parameterwert λ_j , so liefert dieser *Sekanten-Prädiktor* den in Abbildung 5.3 rechts dargestellten Wert

$$x^P := x_{j-1} + \frac{\lambda_j - \lambda_{j-1}}{\lambda_{j-1} - \lambda_{j-2}} \cdot [x_{j-1} - x_{j-2}]. \quad (5.28)$$

Er benötigt für x^P bereits an der Stelle λ_1 zwei Startwerte, die man sich z. B. mittels des Basis-Prädiktors verschaffen muss. Dieser Ansatz stellt einen praktikablen Zugang dar.

4. Durch Nutzung weiterer zurückliegender Kurvenpunkte $(x_{j-1}, \lambda_{j-1}), (x_{j-2}, \lambda_{j-2}), \dots, (x_{j-k}, \lambda_{j-k})$ kann man mittels *polynomialer Extrapolation* sehr genaue Prädiktorformeln der allgemeinen Form

$$x^P := \alpha_{1j} x_{j-1} + \alpha_{2j} x_{j-2} + \dots + \alpha_{kj} x_{j-k}, \quad \alpha_{ij} \in \mathbb{R}, \quad (5.29)$$

gewinnen und den Sekanten-Prädiktor weiter verbessern. Mit verfügbaren Ableitungen $x'(\lambda_{j-1}), x'(\lambda_{j-2}), \dots, x'(\lambda_{j-k})$ lassen sich mit *Hermite-Interpolation* noch genauere Prädiktorwerte x^P ermitteln (vgl. Deuffhard (2004)).

Um die Genauigkeit dieser und weiterer Prädiktoren zu vergleichen, betrachten wir $x(\lambda)$ an der Stelle $\lambda = \lambda_{j-1}$ und definieren mit $\lambda = \lambda_{j-1} + h$ die lokale Funktion

$$u(h) := x(\lambda) = x(\lambda_{j-1} + h) \quad (5.30)$$

für $h \in I_\varepsilon = [-\varepsilon, \varepsilon], \varepsilon > 0$. Die drei angeführten Prädiktoren lassen sich dann ebenfalls in Abhängigkeit von h angeben:

1. Basis-Prädiktor (5.26): $u_B(h) = x_{j-1}$
2. Tangenten-Prädiktor (5.27): $u_T(h) = x_{j-1} + h \cdot u'(0)$
3. Sekanten-Prädiktor (5.28): $u_S(h) = x_{j-1} + h \cdot (x_{j-1} - x_{j-2}) / (\lambda_{j-1} - \lambda_{j-2})$.

Wir führen nun eine Approximationsordnung p ein: $u(h)$ und $u_P(h)$ seien zwei gegebene Funktionen mit $u, u_P : I_\varepsilon \rightarrow \mathbb{R}^n$. Dann approximiert $u_P(h)$ die Kurve $u(h)$ mit der Ordnung $p \in \mathbb{N}$ an der Stelle $h = 0$, falls es Konstanten $0 < h_0 \leq \varepsilon$ und $K > 0$ gibt, so dass

$$\|u(h) - u_P(h)\| \leq K \cdot |h|^p \quad \forall |h| < h_0 \quad (5.31)$$

gilt, d. h. $u(h) - u_P(h) = O(|h|^p)$ ist. Für die drei Prädiktoren beweist man dann

Satz 5.25

Unter Voraussetzung 5.23 approximieren der Basis-Prädiktor und der Sekanten-Prädiktor mit Ordnung 1, während der Tangenten-Prädiktor mit Ordnung 2 approximiert.

Beweis: Es sei $|h| < h_0$. Für die drei Prädiktoren erhalten wir

$$\begin{aligned} \text{(i)} \quad \|u(h) - u_K(h)\| &= \|u(h) - x_{j-1}\| = \|u(h) - u(0)\| \\ &= \left\| \int_0^1 u'(th) dt h \right\| \leq \int_0^1 \|u'(th)\| dt \cdot |h| \\ &\leq K_B \cdot |h| \quad \text{mit} \quad K_B := \max_{t \in [-\varepsilon, \varepsilon]} \|u'(t)\|. \end{aligned}$$

$$\begin{aligned} \text{(ii)} \quad \|u(h) - u_T(h)\| &= \|u(h) - x_{j-1} - h \cdot u'(0)\| \\ &= \|u(0) + u'(0) \cdot h + \frac{1}{2} u''(\sigma) \cdot h^2 - u(0) - u'(0) \cdot h\| \\ &\leq \frac{1}{2} \max_{t \in [-\varepsilon, \varepsilon]} \|u''(t)\| \cdot |h|^2 = K_T \cdot |h|^2. \end{aligned}$$

$$\begin{aligned} \text{(iii)} \quad \|u(h) - u_S(h)\| &= \|u(h) - u(0) - h \cdot (x_{j-1} - x_{j-2}) / (\lambda_{j-1} - \lambda_{j-2})\| \\ &\leq \|u(h) - u(0)\| + \|x'(\lambda^*) \cdot h\| \quad \text{mit} \quad \lambda^* \in (\lambda_{j-1}, \lambda_{j-2}) \\ &\leq K_1 \cdot |h| + K_2 \cdot |h| = K_S \cdot |h| \end{aligned}$$

mit $K_S := K_1 + K_2$, $K_1 := \max_{t \in [-\varepsilon, \varepsilon]} \|u'(t)\|$, $K_2 := \max_{\lambda \in [\lambda_{j-2}, \lambda_{j-1}]} \|x'(\lambda)\|$. \square

Berücksichtigen wir auch (x_{j-3}, λ_{j-3}) und interpolieren mit den Punkten (x_{j-3}, λ_{j-3}) , (x_{j-2}, λ_{j-2}) , (x_{j-1}, λ_{j-1}) , so gewinnen wir einen *Parabel-Prädiktor*, der unter der Voraussetzung $f \in C^3(D \times \Lambda)$ ebenfalls die Approximationsordnung 2 besitzt.

Korrektorschritt

Wegen seiner schnellen Konvergenz wird für den *Korrektorschritt* häufig das Newton-Verfahren (5.5) genutzt. Mit dem berechneten Prädiktorpunkt (x^P, λ_j) setzen wir $u_0 := x^P$ und gewinnen formal die Näherungslösungen

$$u_{k+1} := u_k - [D_x f(u_k, \lambda_j)]^{-1} f(u_k, \lambda_j), \quad k = 0, 1, 2, \dots \quad (5.32)$$

Numerisch effektiver ist auch hier die Lösung linearer Gleichungssysteme

$$D_x f(u_k, \lambda_j) \cdot \Delta u_k = -f(u_k, \lambda_j), \quad u_{k+1} := u_k + \Delta u_k, \quad k = 0, 1, 2, \dots, K-1. \quad (5.33)$$

Der ermittelte Punkt (x_j, λ_j) mit dem Wert $x_j := u_K$ liegt im Allgemeinen nicht auf der Kurve \mathcal{L} , da der Abbruch des Verfahrens bereits nach endlich vielen Iterationen erfolgt. Ein geeignetes Abbruchkriterium muss deshalb vorgesehen werden. Um die aufwändigen Berechnungen der Jacobi-Matrizen zu vermeiden, werden oft Newton-ähnliche Verfahren mit konstanter Matrix $A_0 \sim D_x f(x^P, \lambda_j)$ bevorzugt, so dass (5.33) nun die Form

$$A_0 \cdot \Delta u_k = -f(u_k, \lambda_j), \quad u_{k+1} := u_k + \Delta u_k \quad (5.34)$$

erhält, z. B. mit einer Differenzenapproximation der Matrixelemente

$$\{A_0\}_{ij} := \frac{1}{\Delta x_j} \{f_i(x_1^P, \dots, x_j^P + \Delta x_j, \dots, x_n^P, \lambda) - f_i(x^P, \lambda)\}, \quad i, j = 1(1)n. \quad (5.35)$$

Die Verifikation des Fortsetzungsverfahrens erfordert den Nachweis der lokalen Konvergenz dieses Korrektorschrittes bei geeigetem Prädiktorwert. Damit ist die Endlichkeit der Anzahl von Fortsetzungsschritten über dem Parameterintervall $\Lambda = [a, b]$ zu begründen. Bei Deuffhard und Hohmann (1993) findet man dazu folgendes

Lemma 5.26 (Fortsetzungsschritt)

Zu $\varepsilon > 0$ sei $\Lambda_\varepsilon = (a - \varepsilon, b + \varepsilon)$ ein offenes Intervall, das $\Lambda = [a, b]$ enthält. Voraussetzung 5.23 sei mit Λ_ε erfüllt. Weiterhin existiere ein $\omega > 0$, mit dem f der Lipschitz-Bedingung

$$\|D_x f(x, \lambda)^{-1} \{D_x f(x + hv, \lambda) - D_x f(x, \lambda)\}v\| \leq h\omega \|v\|^2 \quad (5.36)$$

mit $(x, \lambda) \in D \times \Lambda_\varepsilon$, $x + hv \in D$, genügt. Das Prädiktorverfahren P besitze die Ordnung p , d. h. es existiert $K > 0$ mit

$$\|u(h) - u_P(h)\| \leq K \cdot |h|^p \quad \forall |h| < \varepsilon \quad (5.37)$$

mit $u(h) = x(\lambda_{j-1} + h)$, $\lambda_{j-1} \in \Lambda$ und Prädiktorwert $u_P(h)$. Dann konvergiert das Korrektorverfahren (5.33) mit Startwert $u_0 := x_j^P = u_P(h_j)$ gegen die Lösung $u(h_j) = x(\lambda_j)$ von

$$f(x, \lambda_j) = f(x, \lambda_{j-1} + h_j) = 0, \quad (5.38)$$

falls $0 < h_j < h_{max} := \max \left\{ \varepsilon, \sqrt[2]{2/(K\omega)} \right\}$ ist.

Wechselweise Anwendung des Prädiktor- und Korrektorschrittes zusammen mit der Erhöhung des Parameterwertes λ_j ergibt den Algorithmus 5.28 für das *Prädiktor-Korrektor-Verfahren (PECE)* zur numerischen Lösungsfortsetzung. Um dessen unbeschränkte Ausführbarkeit zu verifizieren, ist eine Auswahlregel für die Schrittweiten h_j anzugeben, die die Durchführbarkeit und Endlichkeit der Fortsetzung sichert.

Satz 5.27 (PECE-Verfahren)

Die Voraussetzungen des Lemmas 5.26 seien erfüllt. Für $\lambda_0 = a$ genüge der Startwert $x_0 \approx x(\lambda_0)$ der Bedingung $\rho := \|x_0 - x(\lambda_0)\| < 2/\omega$ und $K[x(\lambda_0); \rho] \subseteq D$. Wählt man die Schrittweiten h_j mit $h_j := \lambda_j - \lambda_{j-1}$ gemäß

$$h_j := \min\{h_{max}/2, b - \lambda_{j-1}\}, \quad (5.39)$$

so gelten folgende Behauptungen:

- i. Algorithmus 5.28 ist durchführbar, d. h. es existieren die λ_j und $x(\lambda_j)$, $j = 0, 1, 2, \dots$ mit $x(\lambda_j)$ lokal eindeutige Lösung von $f(x(\lambda_j), \lambda_j) = 0$, $j = 0, 1, 2, \dots$
- ii. Algorithmus 5.28 ist endlich, d. h. $\lambda_n = b$ für ein $N \in \mathbb{N}$.

Die Endlichkeit des Korrektorschrittes wird damit allerdings nicht gesichert (vgl. Schwetlick (1979)). Wir verzichten auf den Induktionsbeweis und verweisen auf die sehr guten Darstellungen bei Deuffhard und Hohmann (1993) sowie Deuffhard (2004).

Schrittweitensteuerung

Die richtige Wahl der Fortsetzungsschrittweite $h_j = \lambda_j - \lambda_{j-1}$ ist oft entscheidend für den Erfolg der Fortsetzungsmethode. Bei zu großem Wert kann der Prädiktorpunkt außerhalb des Einzugsbereiches der zu bestimmenden Lösung liegen. Dies führt meist zur Divergenz des Korrektors, mitunter auch zur Konvergenz gegen eine andere unerwünschte Lösung. Die theoretisch korrekte Wahl (5.39) führt oft auf extrem kleine Schrittweiten

ALGORITHMUS 5.28 (Natürliche Parametrisierung)

Function $[N, \lambda, x] = \text{continuation}(f, Df x, a, b, x_0, tol)$

1. Setze $j := 0$, $\lambda_0 := a$, $x^P := x_0$
Wähle Schrittweite h sowie Iterationszahl k_{opt} .
2. (*Startrechnung*) Bestimme $x_0 = x(\lambda_0)$ mit Verfahren (5.33) oder (5.34).
3. Do while $\lambda_j \leq b$
 1. Setze $j := j + 1$.
 2. Setze $\lambda_j := \lambda_{j-1} + h$. Falls $\lambda_j > b$, so $\lambda_j := b$.
 3. (*Prädiktor*) Bestimme x^P mit Verfahren (5.26), (5.27) oder (5.28).
 4. (*Korrektor*) Bestimme $x_j = x(\lambda_j)$ mit Verfahren (5.33) oder (5.34) in k Schritten bis auf eine Genauigkeit tol .
 5. (*Schrittweiten-Bestimmung*)
 - (a) Bestimme $\rho := k_{\text{opt}}/k$.
 - (b) Beschränke $\rho := \max\{\min(\rho, 2), \frac{1}{2}\}$.
 - (c) Setze $h := h * \rho$.
 - (d) Falls $h < \varepsilon_M$ (Maschinengenauigkeit), so STOP.
4. Return $N := j$, λ_j , x_j , $j = 0(1)N$

h_j , womit sich die Zahl N der Fortsetzungsschritte und somit der Gesamtaufwand über alle Maßen erhöht. Eine einfache aber zuverlässige Strategie soll solche λ -Schrittweiten $h = h_j > 0$ wählen, die die Konvergenz des Korrektors gegen die verfolgte Lösung garantiert und zugleich die Anzahl der Rechenoperationen minimiert. Heuristisch lässt sich der von h abhängige Rechenaufwand $A(h)$, bezogen auf einen Einheits-Fortsetzungsschritt, durch die Anzahl

$$A(h) := k(h)/h \quad \text{bei } k(h) \text{ ausgeführten Korrektorschritten}$$

ansetzen. Wird mit einer „optimalen“ Schrittweite $h_{\text{opt}} = \rho \cdot h$ der minimale Aufwand

$$A(h_{\text{opt}}) := k(h_{\text{opt}})/h_{\text{opt}}$$

erreicht, so liefert Division mit den Abkürzungen $k := k(h)$, $k_{\text{opt}} := k(h_{\text{opt}})$

$$\frac{A(h)}{A(h_{\text{opt}})} = \frac{k \cdot h_{\text{opt}}}{h \cdot k_{\text{opt}}} = \frac{k}{k_{\text{opt}}} \cdot \varrho \geq 1.$$

Also ist $\varrho \geq k_{\text{opt}}/k$ sinnvoll. Geben wir eine optimale Anzahl von Korrektoriterationen k_{opt} vor, so setzen wir deshalb heuristisch

$$\rho := k_{\text{opt}}/k. \quad (5.40)$$

Eine zusätzliche Begrenzung $\frac{1}{2} \leq \varrho \leq 2$ (safeguarding) erhöht die Sicherheit der Wahl von $h_{\text{opt}} = \varrho \cdot h$. Wir erhalten so die Schrittweitensteuerung in Algorithmus 5.28. Sie passt die aktuelle Fortsetzungsschrittweite h_j nach jedem Schritt so an, dass eine „optimale“ Zahl k_{opt} von Korrektorschritten ausgeführt wird. Für das Newton-Verfahren hat sich $k_{\text{opt}} = 5$ oder $k_{\text{opt}} = 6$ bewährt. Bei zu geringer Fortsetzungsschrittweite h , insbesondere bei Unterschreiten der relativen Maschinengenauigkeit ε_M , sollte der Algorithmus definiert abgebrochen werden.

Konvergiert der Korrektorschritt nicht, so versagt diese „a-posteriori“-Steuerung allerdings. Deshalb wurden zusätzliche Strategien entwickelt, die nach dem Start des Korrektors dessen Schrittweite „a-priori“ neu bestimmen und den Korrektor wiederholen können. Ein Zugang von Allgower und Georg (1990) wertet das asymptotische Verhalten von Prädiktor und Korrektor aus. Nach Satz 5.49 hat der Tangenten-Prädiktor x^P die Ordnung 2 bezüglich der Fortsetzungsschrittweite h . Benutzen wir als Korrektor das Newton-Verfahren, so liefert (5.32) den ersten Näherungswert

$$x^N := x^P - [D_x f(x^P, \lambda_j)]^{-1} f(x^P, \lambda_j).$$

Nach Ausführung eines vereinfachten Newton-Schrittes mit $D_x f(x^P, \lambda_j)$ ergibt sich auf Grund derselben LU-Zerlegung der Jacobi-Matrix mit geringem Zusatzaufwand

$$x^V := x^N - [D_x f(x^P, \lambda_j)]^{-1} f(x^N, \lambda_j).$$

Aus beiden Werten lässt sich eine Näherung für die lineare Konvergenzrate

$$\kappa := \frac{\|x^V - x^N\|}{\|x^N - x^P\|} = \frac{\|D_x f(x^P, \lambda_j)^{-1} f(x^N, \lambda)\|}{\|D_x f(x^P, \lambda_j)^{-1} f(x^P, \lambda)\|}$$

ermitteln. Gesucht ist die explizite Abhängigkeit des κ von der gewählten Schrittweite $h > 0$, für $h \rightarrow 0$. Allgower und Georg (1990) weisen nach, dass unter den obigen Voraussetzungen eine von h unabhängige Konstante $K(x^P) > 0$ existiert, so dass

$$\kappa = K(x^P) \cdot h^2 + O(h^3), \quad h < h_0 \quad (5.41)$$

gilt. Zu gegebenem Prädiktorwert x^P steht uns die Kontraktionsrate $\kappa_{\text{alt}} = \kappa$ mit dieser Asymptotik zur Verfügung. Gefordert wird eine neue Kontraktionsrate κ_{neu} mit

$$\kappa_{\text{neu}} = K(x^P) \cdot h_{\text{neu}}^2 + O(h_{\text{neu}}^3), \quad h_{\text{neu}} < h_0. \quad (5.42)$$

Der Schrittweiten-Faktor sei wieder $\rho := h_{\text{neu}}/h$. Division der Kontraktionsraten liefert wegen $h_{\text{neu}} = \rho \cdot h$

$$\frac{\kappa_{\text{neu}}}{\kappa_{\text{alt}}} = \frac{K(x^P) \cdot h_{\text{neu}}^2 + O(h_{\text{neu}}^3)}{K(x^P) \cdot h^2 + O(h^3)} = \left(\frac{h_{\text{neu}}}{h}\right)^2 + O(h),$$

woraus sich nach Vernachlässigung der Terme $O(h)$ die einfache Regel für

$$h_{\text{neu}} := h \cdot \sqrt{\frac{\kappa_{\text{neu}}}{\kappa_{\text{alt}}}} \quad \text{bzw.} \quad \rho := \sqrt{\frac{\kappa_{\text{neu}}}{\kappa_{\text{alt}}}}. \quad (5.43)$$

ergibt. In praxi setzt man häufig $\kappa_{\text{neu}} := 1/2$. Bei Benutzung des Sekantenprädiktors lässt sich unter geeigneten Voraussetzungen ebenfalls eine asymptotische Darstellung

$$\kappa = K(x^P) \cdot h + O(h^2) \quad \text{mit Konstante} \quad K(x^P) > 0 \quad (5.44)$$

nachweisen; folglich ist

$$h_{\text{neu}} := h \cdot \frac{\kappa_{\text{neu}}}{\kappa_{\text{alt}}}. \quad (5.45)$$

Eine Implementation dieser verbesserten Schrittweiten-Strategie wird empfohlen, da sie nicht unwesentlich zur numerischen Stabilität von Algorithmus 5.28 beiträgt.

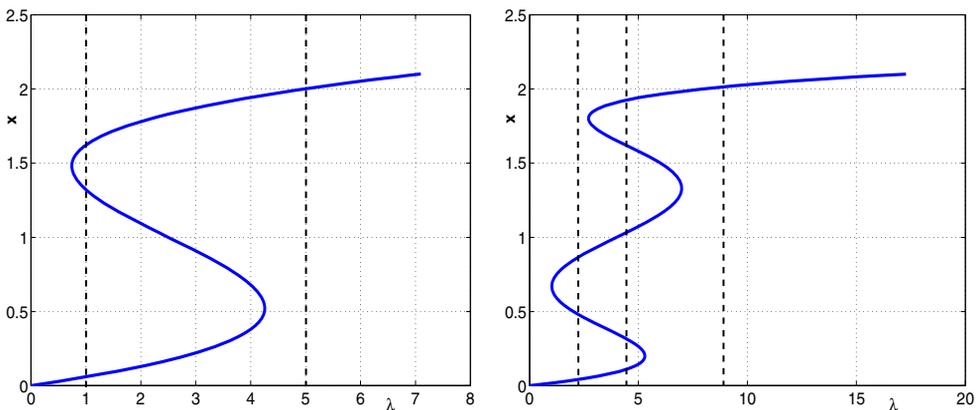


Abb. 5.4 Exakte Lösungskurve \mathcal{L} zu Beispiel 5.29 mit $\mu = 0$ (links) und $\mu = 1$ (rechts)

Beispiel 5.29

Betrachten wir das eindimensionale Beispiel $f(x, \lambda) = 0$ aus Abbildung 5.4 mit

$$f(x, \lambda) := 40\mu x^5 - 200\mu x^4 + (342\mu + 8)x^3 - (226\mu + 24)x^2 + 0.5(91\mu + 37)x - \lambda,$$

wobei $\lambda \geq 0$ der Systemparameter ist. Für festen Parameter $\mu = 0$ treten z. B. bei $\lambda = 1$ drei Lösungen (links), für $\mu = 1$ sogar fünf Lösungen (rechts) auf. Wegen $D_\lambda f(x, \lambda) = -1$ ist jede Lösung $y = (x, \lambda)$ von $f(y) = 0$ ein regulärer Punkt von f im Sinne der Definition 5.21. Die natürliche Parametrisierung nach Algorithmus 5.28 liefert mit Startwerten $(x_0, \lambda_0) = (0, 0)$ die Lösung zuverlässig bis zum ersten Umkehrpunkt (vgl. Abbildung 5.5 links). Dann bricht das Verfahren mit extremer Schrittweitenreduktion zusammen – wie in Abbildung 5.5 rechts dargestellt. ■

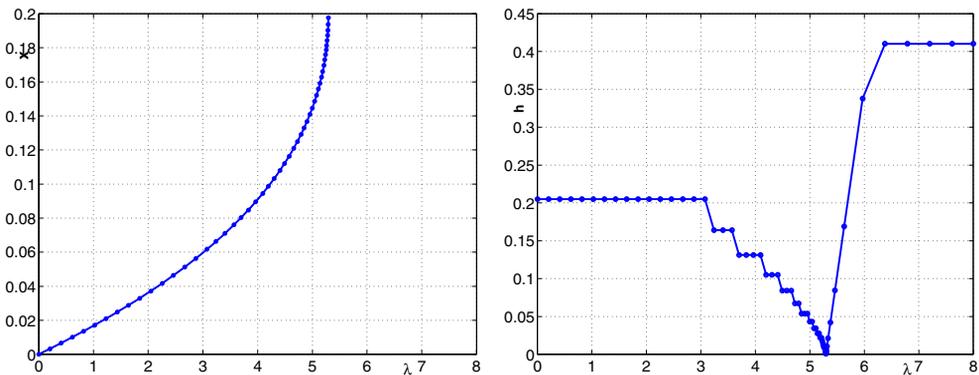


Abb. 5.5 Berechnete Lösung mit $\mu = 1$ (links) und Arbeits-Schrittweite h (rechts)

5.2.2 Bogenlängen-Parametrisierung

Obwohl in jedem Punkt $y = (x, \lambda)$ des Beispiels 5.29 die vollständige Jacobi-Matrix $f'(y) = (D_x f(y), D_\lambda f(y))$ maximalen Rang n hat, versagt der Fortsetzungsalgorithmus an den Stellen $y_0 = (x_0, \lambda_0)$, an denen $D_x f(x_0, \lambda_0)$ singulär wird. Betrachten wir das unterbestimmte System (5.22) und treffen nun folgende

Voraussetzung 5.30

$D \subset \mathbb{R}^{n+1}$ sei offen.

i. $f \in C^r(D)$, $r \geq 2$

ii. Es existiert ein $y_0 \in D \subset \mathbb{R}^{n+1}$ mit $f(y_0) = 0$.

iii. Die erweiterte Jacobi-Matrix $f'(y_0) = (D_x f(y_0), D_\lambda f(y_0))$ hat maximalen Rang n .

Bedingung iii. schwächt die bisherige Regularitätsvoraussetzung ab. Sie garantiert jedoch nach Satz 2.47 aus Abschnitt 2.7, dass die Menge aller Lösungen der Gleichung (5.22) eine eindimensionale Untermannigfaltigkeit \mathcal{L} des \mathbb{R}^n bildet. Aus dem impliziten Funktionentheorem (vgl. Hoffmann et al. (2005)) gewinnen wir dann

Satz 5.31

Unter Voraussetzung 5.30 existiert eine glatte Kurve $\mathcal{L} \in \mathbb{R}^{n+1}$ mit Parametrisierung $y = y(s)$, $s \in J = (s_0 - \delta, s_0 + \delta)$, $\delta > 0$, $y : J \rightarrow \mathbb{R}^{n+1}$, so dass für alle $s \in J$

i. $y(s_0) = y_0$, ii. $f(y(s)) = 0$, iii. $\text{rang}(f'(y(s))) = n$, iv. $y'(s) \neq 0$ gilt.

\mathcal{L} heißt *Lösungspfad* mit der Parametrisierung $y = y(s)$. Einsetzen der Parametrisierung in die Gleichung und Differenziation nach s liefert $f'(y(s)) \cdot y'(s) = 0$. Also liegt der Vektor $y'(s)$ im eindimensionalen Nullraum $N(f'(y(s)))$ und ist orthogonal zu allen n Zeilenvektoren dieser Matrix. Wegen der Rangaussage iii. des Satzes ist folglich die um den Zeilenvektor $y'(s)^T$ erweiterte Jacobi-Matrix

$$H'(y(s)) := \begin{pmatrix} f'(y(s)) \\ y'(s)^T \end{pmatrix} \quad \text{mit} \quad N(f'(y(s))) = \text{span}\{y'(s)\} \quad (5.46)$$

regulär auf dem Pfad \mathcal{L} für $s \in J$. Ein Lösungspfad $\mathcal{L} = \{y(s) \mid f(y(s)) = 0 \quad \forall s \in J\}$ mit der Eigenschaft $\text{rang}(f'(y(s))) = n$ für alle $s \in J$ wird oft als *regulär* bezeichnet. Zur Klassifikation aller Kurvenpunkte $y \in \mathcal{L}$ beweist man leicht

Satz 5.32

Die Bedingung $\text{rang}(f'(y)) = n$ ist genau dann erfüllt, wenn

- i. $D_x f(x, \lambda)$ regulär ist oder
- ii. $\dim N(D_x f(x, \lambda)) = 1$ und $D_\lambda f(x, \lambda) \notin R(D_x f(x, \lambda))$ gilt.

Ein Kurvenpunkt $y(s_0) = (x(s_0), \lambda(s_0)) \in \mathcal{L}$ wird im Falle ii. als *Umkehrpunkt* (*einfacher Grenzpunkt*, *turning point*, *simple limit point*) bezeichnet. Unter Voraussetzung 5.30 besteht jeder reguläre Lösungspfad deshalb nur aus Punkten mit regulärer Jacobi-Matrix $D_x f(x(s_0), \lambda(s_0))$ und einfachen Grenzpunkten. Echte Singularitäten der Ordnung $k \geq 1$ treten hier nicht auf.

Welche praktikablen Parametrisierungen $y = y(s)$ der Kurve \mathcal{L} sind nun einsetzbar? Eine auf Chua und Ushida (1976) zurückgehende Idee führt einen *Parameterwechsel* (*parameter switching*) in der Umgebung von Umkehrpunkten durch. Der natürliche Parameter $\lambda = y_{n+1}$ wird dabei lokal durch eine „geeignere“ andere Komponente y_p , $p \neq n+1$, des Vektors y ersetzt. Offenbar ist y_p besonders dann geeignet, wenn sich diese Komponente beim Durchlaufen des Lösungspfades lokal am stärksten ändert. Benutzt man den Tangenten-Prädiktor $y'(s_j)$ mit $\|y'(s_j)\|_s = 1$, kann man den Index p so bestimmen, dass

$$|y'_p(s_j)| = \max_{i=1(1)n+1} |y'_i(s_j)|$$

ist, während beim Sekanten-Prädiktor zu gegebenen Punkten $y(s_{j-1})$ und $y(s_j)$

$$|y_p(s_j) - y_p(s_{j-1})| = \max_{i=1(1)n+1} |y_i(s_j) - y_i(s_{j-1})|$$

der Auswahl von p dient. Dieser Parameterwechsel kann nach jedem Fortsetzungsschritt erfolgen, womit die Sonderrolle des Parameters λ nach dem ersten Wechsel mit y_p vollkommen entfällt. Wählt man zu p den neuen Parameter $\eta := y_p$, so erhält man das bestimmende System mit n Gleichungen in den n Variablen $y_1, \dots, y_{p-1}, y_{p+1}, \dots, y_{n+1}$

$$f(y_1, \dots, y_{p-1}, \eta, y_{p+1}, \dots, y_{n+1}) = 0 \quad (5.47)$$

zu gegebenem Ausgangspunkt $y^j = (y_1^j, \dots, y_{p-1}^j, \eta^j, y_{p+1}^j, \dots, y_{n+1}^j)$ und $\eta^j := y_p^j$. Nachteilig an dieser Vorgehensweise erweist sich die Gleichbehandlung aller $n+1$ Komponenten von y bei der Wahl von Index p . Wegen unterschiedlicher Größenordnungen der x_i -Komponenten und des Parameters λ sollte deshalb die absolut-relative Abweichung

$$d_p := \max_{i=1(1)n+1} d_i \quad \text{mit} \quad d_i := \frac{|y_i(s_j) - y_i(s_{j-1})|}{|y_i(s_j)| + 1}, \quad i = 1(1)n+1 \quad (5.48)$$

betrachtet werden, die so als gewichtetes Maximum interpretierbar ist. Wir wollen die theoretische Rechtfertigung dieser Strategie jedoch nicht weiter verfolgen, sondern die auf Doedel et al. (2002) zurückgehende Idee der „Pseudo-Bogenlänge“ präferieren, die diesen abrupten Parameterwechsel vermeidet und sich vielfach bewährt hat.

Führen wir – ggf. durch Umparametrisierung der Kurve \mathcal{L} – die Bogenlänge s von $[0, \sigma]$

$$s(\sigma) := \int_0^\sigma \sqrt{\sum_{i=1}^{n+1} [y'_i(\sigma)]^2} d\sigma = \int_0^\sigma \|y'(\sigma)\|_2 d\sigma$$

als Kurvenparameter ein, so genügt der Tangentenvektor in jedem Punkt der Beziehung

$$\|y'(s)\|_2 = 1 \quad \forall s \in J. \quad (5.49)$$

Die Durchlaufgeschwindigkeit einer Kurve ist tatsächlich genau dann in jedem Punkt $s \in J$ gleich Eins, wenn der Parameter s gleich der Bogenlänge ist. Der Tangentenvektor $y'(s)$ wird durch das um (5.49) erweiterte System

$$\left. \begin{array}{l} f'(y(s)) \cdot y'(s) = 0 \\ y'(s)^T y'(s) - 1 = 0 \end{array} \right\} \quad \text{mit} \quad \det \begin{pmatrix} f'(y(s)) \\ y'(s)^T \end{pmatrix} > 0 \quad (5.50)$$

eindeutig bestimmt. Allerdings sollte die numerische Auswertung dieses nichtlinearen impliziten DGL-Systems für $y(s)$ vermieden werden. Nehmen wir dazu an, dass für s_{j-1} die Lösung $y_{j-1} = y(s_{j-1})$ bekannt ist, d. h. $y'(s_{j-1})^T y'(s_{j-1}) - 1 = 0$. Linearisierung von (5.50) bzgl. $y'(s)$ bei y_{j-1} durch lokale Approximation der Bogenlänge liefert dann

$$y'(s_{j-1})^T \frac{y(s) - y(s_{j-1})}{s - s_{j-1}} - 1 = 0,$$

woraus wir nach Multiplikation mit $s - s_{j-1}$ die skalare Zusatzbedingung für y

$$g(y) := y'(s_{j-1})^T (y - y_{j-1}) - (s - s_{j-1}) = 0, \quad g: \mathbb{R}^{n+1} \rightarrow \mathbb{R} \quad (5.51)$$

erhalten. Ergebnis ist das *erweiterte System* für y

$$h(y) := \begin{pmatrix} f(y) \\ g(y) \end{pmatrix} = 0, \quad h: \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{n+1} \quad (5.52)$$

mit der eingeführten skalarwertigen Funktion $g(y)$. Die Jacobi-Matrix dieses Systems

$$h'(y(s)) = \begin{pmatrix} f'(y(s)) \\ g'(y(s)) \end{pmatrix} = \begin{pmatrix} f'(y(s)) \\ y'(s_{j-1})^T \end{pmatrix} \quad (5.53)$$

ist regulär für $s \in J_0 = (s_{j-1} - \delta_0, s_{j-1} + \delta_0)$, $\delta_0 > 0$. Denn nach (5.46) ist $h'(y(s_{j-1}))$ regulär. Wegen der Stetigkeit von $f'(y)$ ist dann auch $h'(y(s))$ regulär.

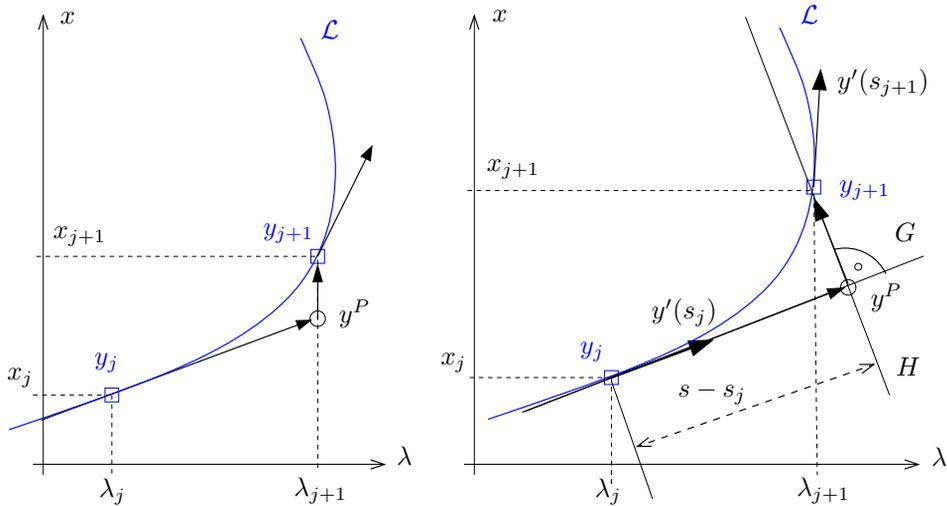


Abb. 5.6 Natürliche Parametrisierung (links) und Pseudo-Bogenlänge (rechts)

Geometrisch lässt sich die Zusatzbedingung mit Abbildung 5.6 (rechts) leicht interpretieren: Wegen der Gleichung $\tilde{y}(s) = y_{j-1} + (s - s_{j-1}) \cdot y'(s_{j-1})$ der Kurventangente G durch y_{j-1} hat der Prädiktorpunkt

$$y^P = y_{j-1} + (s - s_{j-1}) \cdot y'(s_{j-1}) \quad (5.54)$$

den Abstand $s - s_{j-1}$ von y_{j-1} , da $\|y'(s_{j-1})\|_2 = 1$ ist. Die Normal-Hyperebene H zu G durch y^P lautet dann

$$y'(s_{j-1})^T \cdot (y - y^P) = 0.$$

Einsetzen des y^P in H ergibt

$$0 = y'(s_{j-1})^T \{y - y_{j-1} - (s - s_{j-1})y'(s_{j-1})\} = y'(s_{j-1})^T (y - y_{j-1}) - (s - s_{j-1}) = g(y),$$

d. h. $g(y)$ beschreibt die Normal-Hyperebene H zu $y'(s_{j-1})$, die den Abstand $s - s_{j-1}$ von y_{j-1} hat. Wir fassen die Betrachtungen zusammen in

Satz 5.33

Mit den Voraussetzungen 5.30 sei $y_{j-1} = y(s_{j-1}) \in \mathcal{L}$ gegeben. Dann existiert ein $\delta > 0$, so dass für alle s mit $|s - s_{j-1}| < \delta$ das erweiterte System

$$\begin{aligned} f(y) &= 0 \\ g(y) = y'(s_{j-1})^T \cdot (y - y_{j-1}) - (s - s_{j-1}) &= 0 \end{aligned} \quad (5.55)$$

einen eindeutigen glatten Lösungszweig $y = y(s)$ besitzt. Die Jacobi-Matrix $h'(y(s))$ längs des Zweiges ist regulär.

Beweis: $h \in C^r(D \times \Lambda)$, $h(y(s_{j-1})) = 0$ und $h'(y(s_{j-1}))$ ist invertierbar. Mit (5.53) folgt die eindeutige Existenz von $y(s)$ nach dem impliziten Funktionentheorem. \square

ALGORITHMUS 5.34 (Pseudo-Bogenlänge)

Function $[N, y] = \text{pseudoarclength}(f, Df, a, b, x, \text{smax}, \text{tol})$

1. Setze $j := 0$, $\lambda_0 := a$, $x^P := x$, $s_0 := 0$
Wähle Schrittweite h sowie Iterationszahl k_{opt} .
2. (*Startrechnung*) Bestimme $x_0 = x(\lambda_0)$ mit Verfahren (5.33) oder (5.34).
Setze $y_0 = (x_0, \lambda_0)$.
3. Do while ($a \leq \lambda_j \leq b$ and $s_j < \text{smax}$)
 - 3.1. (*Tangentenvektor*) Bestimme $A := f'(y_j)$ und $y'(s_j)$.
 - 3.2. Setze $j := j + 1$ und $s_j := s_{j-1} + h$.
 - 3.3. (*Tangenten-Prädiktor*) Bestimme y^P mittels (5.54).
 - 3.4. (*Newton-Korrektor*) Löse das erweiterte System (5.55) in \mathbb{R}^{n+1}
mit dem Newton-Verfahren oder Newton-ähnlichem Verfahren
 - 3.5. (*Schrittweiten-Anpassung*) wie in Algorithmus 5.28
4. Return $N := j$, $y_j = (x_j, \lambda_j)$, $j = 0(1)N$

Der Bogenlängen-Parameter wird in Abbildung 5.6 rechts anschaulich durch den Tangentenparameter s ersetzt, weshalb man nun von der *Pseudo-Bogenlänge* spricht. Anders als bei der natürlichen Parametrisierung in Abbildung 5.6 links werden nun Umkehrpunkte leicht überwunden. Algorithmus 5.34 nutzt dafür den Tangenten-Prädiktor.

Beispiel 5.35

Deuffhard (2004) behandelt ein fünfdimensionale Stabilitätsproblem aus dem Flugzeugbau mit den Gleichungen

$$\begin{aligned}
 & -3.933x_1 + 0.107x_2 + 0.126x_3 - 9.99x_5 - 45.83\lambda - 0.727x_2x_3 \\
 & \qquad \qquad \qquad + 8.39x_3x_4 - 684.4x_4x_5 + 63.5x_4\lambda = 0 \\
 & \qquad \qquad \qquad -0.987x_2 - 22.95x_4 - 28.37u + 0.949x_1x_3 + 0.173x_1x_5 = 0 \\
 & 0.002x_1 - 0.235x_3 + 5.67x_5 - 0.921\lambda - 0.713x_1x_2 - 1.578x_1x_4 + 1.132x_4\lambda = 0 \\
 & \qquad \qquad \qquad x_2 - x_4 - 0.168u - x_1x_5 = 0 \\
 & \qquad \qquad \qquad -x_3 - 0.196x_5 - 0.0071\lambda + x_1x_4 = 0.
 \end{aligned}$$

Die Steuervariable u wird hier auf Null gesetzt und $-1 \leq \lambda \leq 1$ angenommen. Algorithmus 5.34 liefert zuverlässig die Darstellungen der Lösungskomponenten x_2, x_3, x_4, x_5 in Abbildung 5.7 mit einer vorgegebenen Genauigkeit $\text{tol} = 10^{-8}$. Die Lösungssymmetrie bezüglich $\lambda = 0$ ist gut erkennbar. \blacksquare

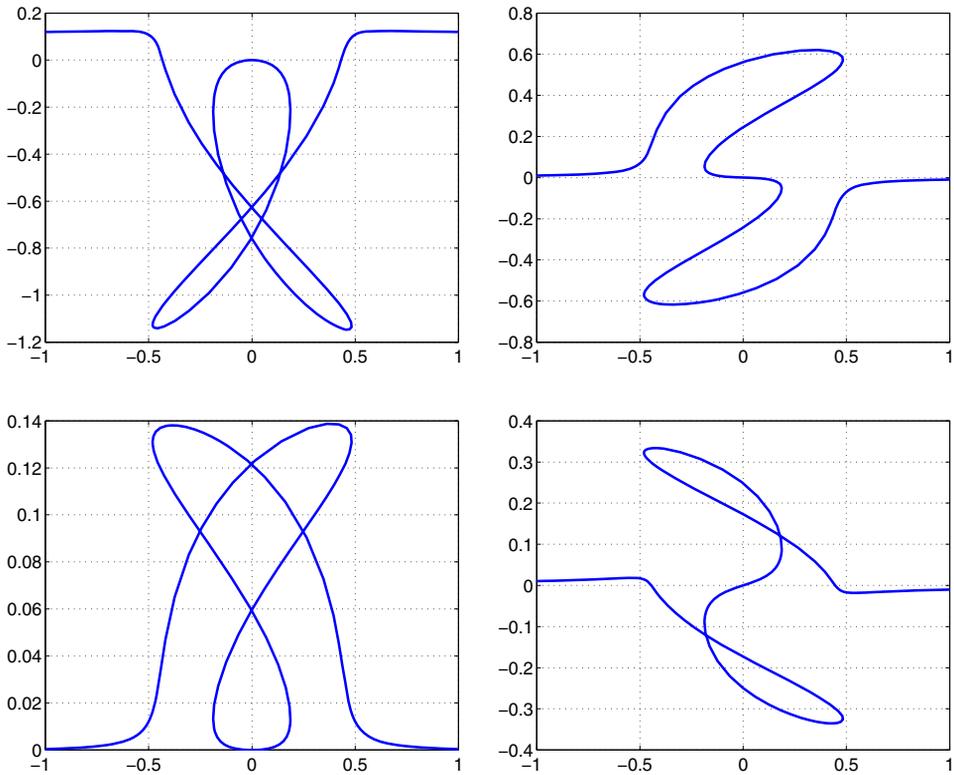


Abb. 5.7 Darstellung der Lösungskomponenten x_2 (links oben), x_3 (rechts oben) und x_4 (links unten), x_5 (rechts unten) über dem Parameter λ

Es bleibt die Frage zu klären, wie wir den in Algorithmus 5.34 – sowohl im Prädiktor- als auch im Korrektorschritt – benötigten normierten Tangentenvektor $y'(s_j)$ effektiv bestimmen. Es sei $A := f'(y(s_j)) \in \mathbb{R}^{n \times (n+1)}$ die im j -ten Schritt ermittelte Jacobi-Matrix. Nach Voraussetzung 5.30 ist ihr Nullraum $N(A)$ eindimensional. Deshalb kann $y'(s_j) \in N(A)$ an Stelle der üblicherweise benutzten Singulärwertzerlegung in diesem speziellen Fall mittels QR-Zerlegung oder LU-Zerlegung gefunden werden. Die QR-Zerlegung der transponierten Jacobi-Matrix A^T führt auf die Darstellung

$$A^T = Q \cdot R = \begin{pmatrix} q_{11} & \cdots & q_{1,n+1} \\ q_{21} & \cdots & q_{2,n+1} \\ \vdots & & \vdots \\ q_{n+1,1} & \cdots & q_{n+1,n+1} \end{pmatrix} \cdot \begin{pmatrix} r_{11} & \cdots & r_{1n} \\ 0 & \ddots & \\ & & r_{nn} \\ 0 & \cdots & 0 \end{pmatrix}$$

mit der orthogonalen Matrix $Q = (q_{ik}) \in \mathbb{R}^{(n+1) \times (n+1)}$ und der hier dargestellten oberen Dreiecksmatrix $R = (r_{ik}) \in \mathbb{R}^{(n+1) \times n}$, deren letzte Zeile nur Nullen enthält. Mit dem

$(n+1)$ -ten Einheitsvektor $e_{n+1} = (0, \dots, 0, 1)^T$ und der Darstellung $z := Qe_{n+1}$ gewinnen wir die letzte Spalte der Matrix Q . Wegen der Orthogonalitätsbedingung $Q^T Q = I$ ist

$$A \cdot z = (QR)^T Q e_{n+1} = R^T Q^T Q e_{n+1} = 0 \quad \text{und}$$

$$\|z\|_2^2 = e_{n+1}^T Q^T Q e_{n+1} = 1,$$

womit der Vektor $y'(s_j) := z$ bis auf das Vorzeichen bereits das gesuchte Nullraumelement ist. Das Vorzeichen des Tangentenvektors muss nun so gewählt werden, dass ein

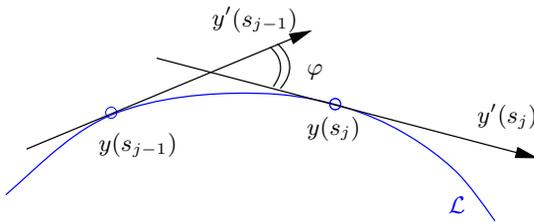


Abb. 5.8

Darstellung zweier aufeinander folgender Tangentenvektoren $y'(s_{j-1})$ und $y'(s_j)$ mit der Lösungskurve \mathcal{L}

Durchlaufwechsel der Lösungskurve \mathcal{L} vermieden wird. Dazu betrachten wir zwei aufeinander folgende Tangentenvektoren $y'(s_{j-1})$ und $y'(s_j)$, wie in Abbildung 5.8 dargestellt. Um einen spitzen Winkel φ zwischen beiden Tangentenrichtungen zu garantieren, fordern wir für das Skalarprodukt der Vektoren

$$\tau := y'(s_{j-1})^T y'(s_j) = \|y'(s_{j-1})\| \cdot \|y'(s_j)\| \cdot \underbrace{\cos \varphi}_{>0} > 0.$$

Falls $\tau < 0$ ist, so kehren wir das Vorzeichen von $y'(s_j)$ um und erhalten insgesamt den fehlenden Schritt 3.1. von Algorithmus 5.34:

■ Berechnung des Tangentenvektors

- Bestimme die Jacobi-Matrix $A := f'(y_j)$.
- QR-Zerlegung von A^T liefert $A^T = Q \cdot R$
- Setze $y'(s_j) := Qe_{n+1}$ mit $(n+1)$ -tem Einheitsvektor e_{n+1}
- Falls $\tau := y'(s_{j-1})^T y'(s_j) < 0$, so setze $y'(s_j) := -y'(s_j)$.

Man überlegt sich leicht, dass in analoger Weise auch die LU-Zerlegung von A^T mit Zeilenvertauschungen zur Ermittlung des Tangentenvektors $y'(s_j)$ genutzt werden kann.

5.2.3 Gauß-Newton-Fortsetzung

Betrachten wir nochmals Abbildung 5.6, so erkennen wir, dass bei zu groß gewählter Fortsetzungsschrittweite h auch die Pseudo-Bogenlänge keinen Schnittpunkt y_j der

Normal-Hyperbene mit der Kurve liefern muss. Daher sind folgende Schritte naheliegend, um vom Bogenlängen-Parameterwert s_{j-1} zu s_j überzugehen:

1. Gegeben sei der Kurvenpunkt $y_{j-1} = y(s_{j-1})$. Festlegung einer geeigneten Fortsetzungsschrittweite $h_j > 0$ und des neuen Parameterwertes $s_j = s_{j-1} + h_j$.
2. Vorgabe eines Prädiktorpunktes $v_0 = v = y^P$ für den neuen Kurvenpunkt y_j , z. B. mit dem Tangenten-Prädiktor

$$y^P := y_{j-1} + h_j \cdot y'(s_{j-1}), \quad \|y'(s_{j-1})\|_2 = 1 \tag{5.56}$$

oder dem Sekanten-Prädiktor (vgl. Abb. 5.9)

$$y^P := y_{j-1} + \frac{h_j}{h_{j-1}} \cdot (y_{j-1} - y_{j-2}). \tag{5.57}$$

3. Bestimmung eines Kurvenpunktes $y_j \in \mathcal{L}$ mit minimalem Abstand von $v = y^P$, d. h. iterative Lösung der Minimierungsaufgabe im Korrektorschritt (vgl. Abb. 5.9)

$$\|y_j - v\|_2 = \min_{f(y)=0} \|y - v\|_2. \tag{5.58}$$

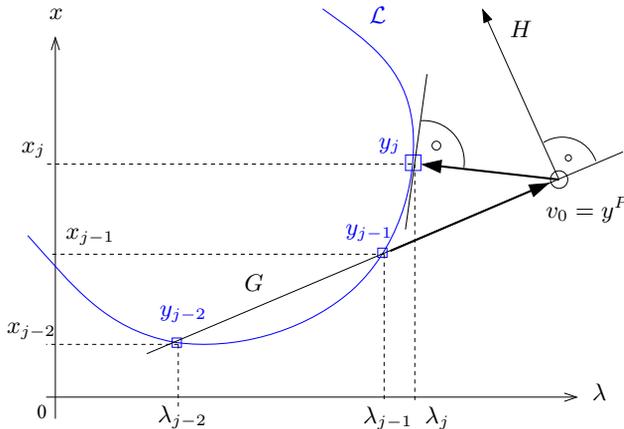


Abb. 5.9 Kurvenverfolgung mittels Projektion

Um dieses Extremwertproblem mit Nebenbedingung zu lösen, definieren wir mit den Multiplikatoren $\mu = (\mu_1, \dots, \mu_n)^\top$ die zugehörige Lagrange-Funktion

$$L(y, \mu) = \frac{1}{2} \|y - v\|_2^2 - \mu^\top f(y) = \frac{1}{2} \sum_{i=1}^{n+1} (y_i - v_i)^2 - \sum_{i=1}^n \mu_i f_i(y). \tag{5.59}$$

Durch Differentiation gewinnen wir die notwendige Bedingung an y_j für ein lokales Minimum in Komponentenform

$$\begin{aligned} (i) \quad \frac{\partial L}{\partial \mu_k} &= f_k(y) = 0, & k &= 1(1)n, \\ (ii) \quad \frac{\partial L}{\partial y_l} &= (y_l - v_l) - \sum_{i=1}^n \mu_i \frac{\partial f_i}{\partial y_l} = 0, & l &= 1(1)n + 1 \end{aligned}$$

bzw. in Vektorschreibweise

$$\begin{aligned} (i) \quad & f(y) = 0 \\ (ii) \quad & y - v = [f'(y)]^\top \cdot \mu, \quad \mu \in \mathbb{R}^n. \end{aligned} \tag{5.60}$$

Multiplikation von Bedingung (ii) an der Lösung y_j mit dem Tangentenvektor $y'(s_j)$ ergibt

$$y'(s_j)^\top \cdot (y_j - v) = [f'(y(s_j)) \cdot y'(s_j)]^\top \mu = 0$$

wegen Voraussetzung 5.30, womit das bestimmende System für $y_j \in \mathbb{R}^{n+1}$ nunmehr

$$\begin{aligned} (i) \quad & f(y) = 0 \\ (ii) \quad & y'(s_j)^T \cdot (y - v) = 0 \end{aligned} \tag{5.61}$$

lautet. Anschaulich wird also ein Minimalpunkt y_j durch die in Abbildung 5.9 dargestellte Orthogonalitätsbedingung $y'(s_j)^T \cdot (y_j - v) = 0$ mit dem Tangentialvektor $y'(s_j)$ definiert. Linearisierung des Systems (5.61) am Näherungspunkt $v \in \mathbb{R}^{n+1}$ mit zugehörigem Tangentenvektor $v'(s_j)$ liefert das lineare Gleichungssystem für den Wert v_{neu} :

$$\begin{aligned} f'(v) \cdot (v_{\text{neu}} - v) &= -f(v) \\ v'(s_j)^T \cdot (v_{\text{neu}} - v) &= 0. \end{aligned} \tag{5.62}$$

Bezeichnen wir $A := f'(v)$, $b := -f(v)$ und $x := v_{\text{neu}} - v$ sowie den Tangentenvektor $t(A) := v'(s_j)$, so lautet das System (5.62) mit diesen Abkürzungen

$$A \cdot x = b, \quad t(A) \cdot x = 0. \tag{5.63}$$

$Ax = b$ beschreibt eine Gerade und die Lösung x^* von (5.63) ist ein Punkt mit $t(A) \perp x$. Geometrisch bedeutet dies, eine Lösung des unterbestimmten linearen Gleichungssystems $Ax = b$ zu finden, die die Minimierungsaufgabe

$$\|x^*\|_2 = \min_{Ax=b} \|x\|_2 \tag{5.64}$$

löst. Deren Lösungen x^* kann man mit der *Pseudoinversen* A^+ von A formal in der Form $x^* = A^+b$ darstellen. Dabei wird A^+ folgendermaßen erklärt:

Definition 5.36 (Pseudoinverse, Moore-Penrose-Inverse)

Es sei $A \in \mathbb{R}^{m \times n}$. Die *Pseudoinverse (Moore-Penrose-Inverse)* $A^+ \in \mathbb{R}^{n \times m}$ von A ist durch die vier Penrose-Bedingungen eindeutig definiert:

- (i) $AA^+A = A$
- (ii) $A^+AA^+ = A^+$
- (iii) $(AA^+)^\top = AA^+$
- (iv) $(A^+A)^\top = A^+A$.

Die geometrische Beobachtung aus Abbildung 5.9 wird durch folgenden Äquivalenzsatz auch theoretisch abgesichert (vgl. Allgower und Georg (1990)):

Satz 5.37

Es sei $A \in \mathbb{R}^{n \times (n+1)}$, $\text{rang}(A) = n$, und $\varphi = t(A)$ der normierte Tangentenvektor, bestimmt durch

$$(i) \quad A \cdot \varphi = 0, \quad (ii) \quad \|\varphi\|_2 = 1, \quad (iii) \quad \det \begin{pmatrix} A \\ \varphi^\top \end{pmatrix} > 0. \quad (5.65)$$

Dann sind für $b \in \mathbb{R}^n$ und $x \in \mathbb{R}^{n+1}$ folgende Aussagen äquivalent:

- (A1) $Ax = b$ und $\varphi^\top x = 0$.
- (A2) $x = A^+ b$
- (A3) x löst die Minimierungsaufgabe (5.64).

Die Berechnung von A^+ im allgemeinen Fall einer (m, n) -Matrix ist allerdings aufwändig und bei Fortsetzungsverfahren nicht praktikabel. Wegen $A \in \mathbb{R}^{n \times (n+1)}$ mit den Voraussetzungen 5.30 des zeilenregulären Falles kann A^+ jedoch als Rechtsinverse

$$A^+ = A^T (AA^T)^{-1} \quad \text{wegen} \quad AA^+ = AA^T (AA^T)^{-1} = I \quad (5.66)$$

direkt ausgerechnet werden. Anwendung der Pseudoinversen von $A := f'(v)$ auf (5.62) ergibt somit den Newtonschritt

$$v_{\text{neu}} = v - A^+ f(v) \quad \text{mit} \quad A^+ = A^T (AA^T)^{-1},$$

dessen iterative Wiederholung das *Gauß-Newton-Verfahren für unterbestimmte nichtlineare Gleichungssysteme* liefert:

$$v_{k+1} = v_k - A^T (AA^T)^{-1} f(v_k) \quad \text{mit} \quad A = f'(v_k), \quad k = 0, 1, 2, \dots \quad (5.67)$$

Liegt der Prädiktorwert des Startpunktes $v_0 = y^P$ hinreichend nahe am Lösungspfad \mathcal{L} , so kann die lokale quadratische Konvergenz dieses Verfahrens nachgewiesen werden (vgl. dazu Allgower und Georg (1990)).

Satz 5.39

Für $f : D \subset \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$ seien die Voraussetzungen 5.30 erfüllt. Dann existiert eine offene Umgebung $U(\mathcal{L}) \subset \mathcal{D}$ des Lösungspfad, mit der folgende Behauptungen gelten:

i. Zu jedem $v \in U(\mathcal{L})$ existiert genau ein $y^* \in \mathcal{L}$ mit

$$\|y^* - v\|_2 = \min_{f(y)=0} \|y - v\|_2.$$

Die dadurch definiert Abbildung $S : U(\mathcal{L}) \rightarrow \mathcal{L}$ ist glatt.

ii. Ist $v_0 \in U(\mathcal{L})$, so auch $v_1 \in U(\mathcal{L})$.

- iii. Zu jedem $v_0 \in U(\mathcal{L})$ konvergiert die Folge (v_k) , $k = 0, 1, 2, \dots$ der Iterierten gemäß (5.67) gegen ein $v_\infty \in \mathcal{L}$.
- iv. Für $v_0 \in U(\mathcal{L})$ gelten folgende Abschätzungen gleichmäßig für $v_k \in U(\mathcal{L})$:

$$\|v_2 - v_1\| \leq C_1 \|v_1 - v_0\|^2$$

$$\|v_\infty - v_1\| \leq C_2 \|v_\infty - v_0\|^2$$

$$\|v_1 - S(v_0)\| \leq C_3 \|v_0 - S(v_0)\|^2$$

$$\|v_\infty - S(v_0)\| \leq C_4 \|v_0 - S(v_0)\|^2$$

Zur praktischen Umsetzung dieses Verfahrens ist die vollständige Jacobi-Matrix $A = f'(y)$ vorzugeben. Zusammen mit dem Sekanten-Prädiktor liefert der Gauß-Newton-Korrektor den Algorithmus 5.38. Im Unterschied zur Parametrisierung mittels Pseudo-

ALGORITHMUS 5.38 (Gauß-Newton-Fortsetzung)

Function $[N, y] = \text{gaussnewton}(f, Df, a, b, x, \text{smax}, \text{tol})$

1. Setze $j := 0$, $\lambda_0 := a$, $x^P := x$, $s_0 := 0$
Wähle Schrittweite h sowie Iterationszahl k_{opt} .
2. (*Startrechnung*) Bestimme $x_0 = x(\lambda_0)$ mit Verfahren (5.33) oder (5.34).
Setze $y_0 = (x_0, \lambda_0)$.
3. Do while ($a \leq \lambda_j \leq b$ and $s_j < \text{smax}$)
 - 3.1. Setze $j := j + 1$ und $s_j := s_{j-1} + h$.
 - 3.2. (*Sekanten-Prädiktor*) Bestimme $v_0 := y^P$ mit Verfahren (5.57),
falls $j > 1$, andernfalls mit Verfahren (5.56).
 - 3.3. (*Gauß-Newton-Korrektor*)
 - (a) Setze $k := 0$.
 - (b) $A := Df(v_k) = f'(v_k)$
 - (c) Löse das reguläre lineare System

$$(AA^T) \cdot b = f(v_k), \quad b \in \mathbb{R}^n$$
 - (d) $v_{k+1} := v_k - A^T b$, $k := k + 1$
 - (e) Falls $\|A^T b\|_2 > \text{tol}$, so gehe zu Schritt (b).
 - (f) Setze $y_j := v_k$.
 - 3.4. (*Schrittweiten-Anpassung*) wie in Algorithmus 5.28
4. Return $N := j$, $y_j = (x_j, \lambda_j)$, $j = 0(1)N$

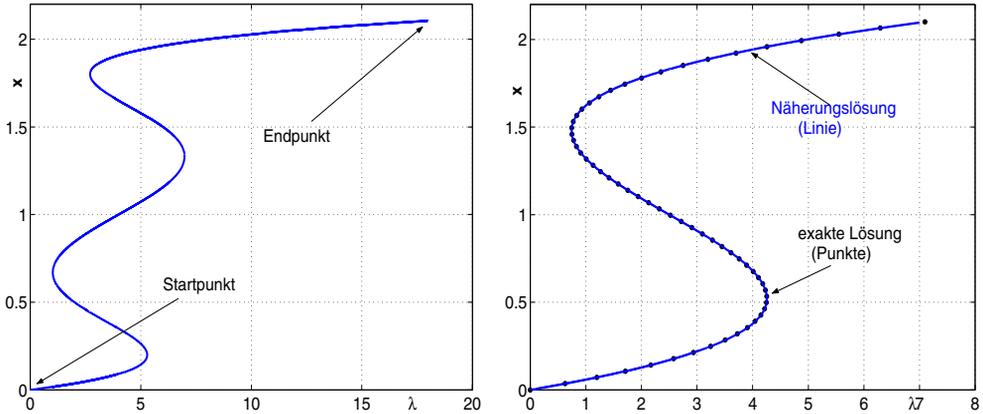


Abb. 5.10 Berechnete Lösung \mathcal{L} zu Beispiel 5.29 mit $\mu = 1$ (links) und $\mu = 0$ (rechts)

Bogenlänge benötigt dieses Verfahren allerdings keine Berechnung des Tangentenvektors $y'(s_j)$ – weder im Prädiktor- noch im Korrektorschritt.

Beispiel 5.40

Wir wenden den Algorithmus auf Beispiel 5.29 mit der Funktion

$$f(x, \lambda) := 40\mu x^5 - 200\mu x^4 + (342\mu + 8)x^3 - (226\mu + 24)x^2 + 0.5(91\mu + 37)x - \lambda$$

an. Für festen Parameter $\mu = 1$ erhalten wir mit Startwerten $(x_0, \lambda_0) = (0, 0)$ und maximaler Fortsetzungsschrittweite $h_{\max} = 0.005$ in Abbildung 5.10 (links) die Lösung mit Genauigkeit $\text{tol} = 10^{-3}$ über alle vier Umkehrpunkte hinweg (vgl. die exakte Darstellung in Abbildung 5.6). Auch bei $\mu = 0$ arbeitet die Fortsetzung zuverlässig und liefert in Abbildung 5.10 (rechts) die Näherungslösung. Zum Vergleich wurden hier exakte Kurvenpunkte eingezeichnet. ■

5.2.4 Fortsetzung mehrdimensionaler Mannigfaltigkeiten

Die üblichen Anwendungsprobleme aus der Praxis hängen meist von mehreren Parametern $\lambda_1, \lambda_2, \dots, \lambda_p$ ab. Um das Lösungsverhalten über einem Bereich des Parameter-raumes \mathbb{R}^p zu studieren, ist dann ein mehrdimensionales Fortsetzungsproblem zu lösen. Anstelle einer Lösungskurve ist nun eine *Lösungsmannigfaltigkeit* des n -dimensionalen Problems

$$f(y) = 0 \quad \text{mit } y = (x, \lambda) \in D \subset \mathbb{R}^{n+p}, \quad p \geq 1, \quad D \text{ offen} \quad (5.68)$$

mit Parametervektor $\lambda \in \mathbb{R}^p$ zu finden. Die Idee der Kurvenverfolgung mittels Pseudo-Bogenlänge, einen Kurvenpunkt y_j durch die in Abbildung 5.6 (rechts) dargestellte Orthogonalitätsbedingung $y'(s_{j-1})^T (y - y^P) = 0$ mit dem Tangentialvektor $y'(s_{j-1})$ zu bestimmen, lässt sich auch auf mehrdimensionale Lösungsmannigfaltigkeiten verallgemeinern. Betrachten wir dazu das Problem (5.68) unter folgender

Voraussetzung 5.41

- i. $f : D \subset \mathbb{R}^{n+p} \rightarrow \mathbb{R}^n$ sei C^r -glatt mit $r \geq 2$ sowie
 ii. Die $n \times (n+p)$ -Matrix $f'(y) = Df(y)$ hat maximalen Rang n für alle $y \in D$.

Dann ist die Menge $\mathcal{M} = \{y \in D \mid f(y) = 0\}$ eine p -dimensionale Mannigfaltigkeit in \mathbb{R}^{n+p} mit C^r -Glattheit (vgl. Rheinboldt (1994)). Zu einem Punkt $y_j \in \mathcal{M}$ bilden

- $T_{y_j}\mathcal{M} = N(Df(y_j))$ den p -dimensionalen Tangentialraum von \mathcal{M} bei y_j und
- $N_{y_j}\mathcal{M} = \{N(Df(y_j))\}^\perp = R(Df(y_j)^T)$ den n -dimensionalen Normalraum,

der das orthogonale Komplement von $T_{y_j}\mathcal{M}$ in \mathbb{R}^{n+p} ist. Nehmen wir nun an, wir hätten p paarweise orthonormierte Vektoren $b_1, b_2, \dots, b_p \in \mathbb{R}^{n+p}$ zur Verfügung, so dass die damit erweiterte $(n+p) \times (n+p)$ -Matrix

$$C := \begin{pmatrix} Df(y_j) \\ B^T \end{pmatrix} \quad \text{mit} \quad B = (b_1, b_2, \dots, b_p) \in \mathbb{R}^{(n+p) \times p} \quad (5.69)$$

regulär ist. Dies ist während der Lösungsfortsetzung realisierbar, wenn beispielsweise $\mathcal{B} = \{b_1, b_2, \dots, b_p\}$ eine orthonormale Basis des Tangentialraumes $T_{y_{j-1}}\mathcal{M}$ eines vorhergehenden Lösungspunktes y_{j-1} darstellt.

Zur Bestimmung von y_j bilden wir in Analogie zu (5.61) am Prädiktorpunkt y^P das erweiterte System

$$g(y) := \begin{pmatrix} f(y) \\ B^T(y - y^P) \end{pmatrix} = 0, \quad g : D \subset \mathbb{R}^{n+p} \rightarrow \mathbb{R}^{n+p}, \quad (5.70)$$

indem wir die n Gleichungen um p Projektionsbedingungen $b_i^T(y - y^P) = 0$, $i = 1(1)p$, ergänzen. Wegen Voraussetzung 5.41(i) ist $g(y)$ C^2 -glatt, und mit (5.69) ist die Jacobi-Matrix $Dg(y_j)$ regulär. Damit sind Newton-Verfahren und vereinfachtes Newton-Verfahren aus Abschnitt 5.1 für $g(y) = 0$ lokal in der Umgebung von y_j ausführbar und konvergieren Q-quadratisch bzw. Q-linear gegen eine gesuchte Lösung y_j mit $g(y_j) = 0$.

Nach Erhalt von y_j ist allerdings noch die Frage zu klären, wie eine Orthonormalbasis $\mathcal{B}_{\text{neu}} = \{b_1, b_2, \dots, b_p\}$ von $T = T_{y_j}\mathcal{M}$ für den nächsten Fortsetzungsschritt zu bestimmen ist. Dazu lösen wir mit der verfügbaren Matrix B die p linearen Gleichungssysteme

$$\begin{pmatrix} Df(y_j) \\ B^T \end{pmatrix} \tau = e_{n+i}, \quad i = 1(1)p, \quad (5.71)$$

deren rechte Seiten die $(n+p)$ -dimensionalen Einheitsvektoren e_{n+i} sind.

Satz 5.42

Die Lösungen $\tau_1, \tau_2, \dots, \tau_p$ von (5.71) sind eindeutig bestimmt und liefern eine Basis des Tangentialraumes $T_{y_j}\mathcal{M}$.

Beweis: Die Koeffizientenmatrix ist offenbar regulär. Aus den ersten n Gleichungen von (5.71) folgt $Df(y_j)\tau_i = 0$ für $i = 1(1)p$, womit $\tau_i \in T_{y_j}\mathcal{M}$ ist. Aus der zusammengefassten Darstellung $B^T(\tau_1, \tau_2, \dots, \tau_p) = I_p$ der übrigen p Gleichungen mit der Einheitsmatrix I_p ergibt sich dann die lineare Unabhängigkeit der Vektoren $\tau_1, \tau_2, \dots, \tau_p$. \square

Schließlich gewinnt man aus dieser Basis mittels *Gram-Schmidt-Orthogonalisierung* oder *QR-Zerlegung* eine Orthonormalbasis $\mathcal{B}_{\text{neu}} = \{b_1, b_2, \dots, b_p\}$ des Tangentialraumes $T = T_{y_j}\mathcal{M}$. Mit diesen Vektoren definieren wir die Matrix $B = (b_1, b_2, \dots, b_p) \in \mathbb{R}^{(n+p) \times p}$ und können damit alle Elemente $y \in \mathbb{R}^{n+p}$ eindeutig in der Form

$$y = y_j + Bs + w \quad \text{mit} \quad s \in \mathbb{R}^p \text{ und } w \in T^\perp$$

darstellen. Betrachten wir an einem derartigen Punkt y das erweiterte System für die Unbekannte $w \in \mathbb{R}^{n+p}$

$$h(w, s) := \begin{pmatrix} f(y_j + Bs + w) \\ B^T w \end{pmatrix} = 0 \quad (5.72)$$

mit dem Parametervektor $s \in \mathbb{R}^p$. Die Abbildung $h : \mathbb{R}^{n+p} \times \mathbb{R}^p \rightarrow \mathbb{R}^{n+p}$ hat offenbar dieselbe C^r -Glattheit wie f . Zudem gilt

$$h(0, 0) = 0 \quad \text{und} \quad D_w h(0, 0) = \begin{pmatrix} Df(y_j) \\ B^T \end{pmatrix} \text{ ist regulär.}$$

Nach dem impliziten Funktionentheorem existieren deshalb Umgebungen $V_s(0) \subset \mathbb{R}^p$ und $V_w(0) \subset \mathbb{R}^{n+p}$ von Null und eine eindeutige Abbildung $w : V_s(0) \rightarrow V_w(0)$, die der Gleichung $h(w(s), s) = 0$ für alle $s \in V_s(0)$ sowie $w(0) = 0$ genügt. Damit kann die Mannigfaltigkeit \mathcal{M} an der Stelle y_j lokal durch die Funktion $y : V_s(0) \rightarrow \mathbb{R}^{n+p}$ mit

$$y(s) := y_j + Bs + w(s), \quad s \in V_s(0), \quad w(0) = 0, \quad w(s) \in V_w(0) \subset T^\perp \quad (5.73)$$

parametrisiert werden. Ausgehend vom Punkt $y_j \in \mathcal{M}$ bestimmen wir durch geeignete Wahl von Parametervektoren $s \in V_s(0)$ Nachbarpunkte $y(s)$ auf der Mannigfaltigkeit, indem wir System (5.72) nach $w(s)$ auflösen. Für den dreidimensionalen Fall sind in Abbildung 5.11 die affinen Unterräume $y_j + T_{y_j}\mathcal{M}$ mit der Orthonormalbasis $\mathcal{B} = \{b_1, b_2\}$ sowie $y_j + N_{y_j}\mathcal{M}$ mit Dimension 1 schwarz dargestellt. Ausgehend von $y_j = y(s_j)$ werden z. B. vier neue Punkte $y_{j+1} = y(s_{j+1})$ auf der blau dargestellten Mannigfaltigkeit \mathcal{M} erzeugt, indem zu vorgegebenen Parameterwerten s_{j+1} und damit berechneten Vektoren $w(s_{j+1})$ (blau) die neuen Vektoren $y_{j+1} := y_j + Bs_{j+1} + w(s_{j+1})$ erzeugt werden. Auf der linken Bildseite erkennt man eine Triangulierung der Mannigfaltigkeit, die im praktischen Algorithmus geeignet vorzunehmen ist. Im nächsten Fortsetzungsschritt kann die Vorgehensweise für jeden Punkt y_{j+1} mit der zugehörigen Basis \mathcal{B} wiederholt werden.

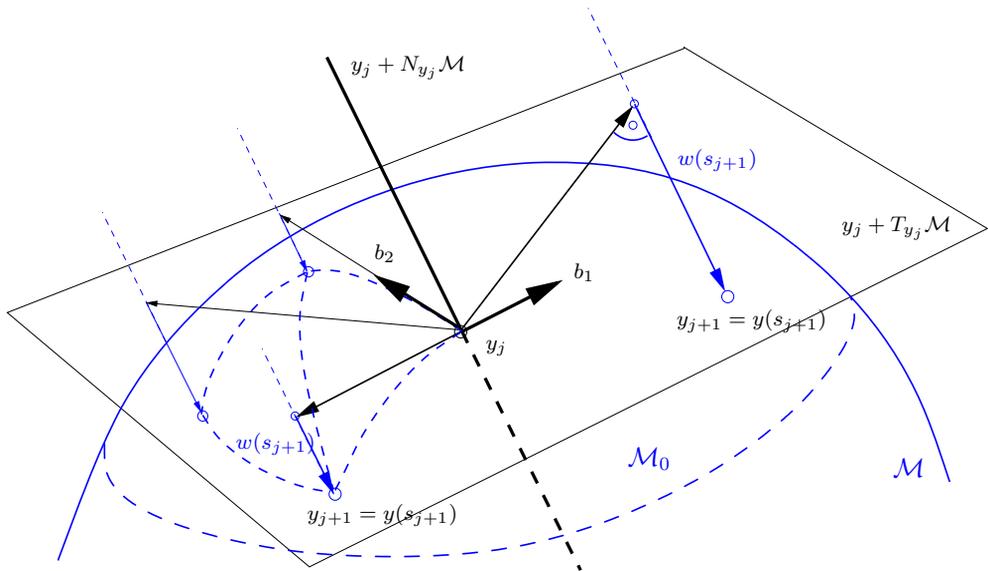


Abb. 5.11 Darstellung von $y_j = y(s_j)$ und vier neuer Punkte $y_{j+1} = y(s_{j+1})$ auf der Mannigfaltigkeit \mathcal{M}

Definition 5.44 (Moving Frame, begleitende Basis)

\mathcal{M}_0 sei eine offene Teilmenge von \mathcal{M} . Ein *Moving Frame* (*begleitende Basis*) der Klasse C^r besteht aus C^r -glatten Funktionen $b_1(y), b_2(y), \dots, b_p(y)$, die eine Basis \mathcal{B} des Tangentialraumes $T_y\mathcal{M}$ für alle $y \in \mathcal{M}_0$ bilden (vgl. Mei (2000)).

Wird eine feste Matrix $B = (b_1, b_2, \dots, b_p) \in \mathbb{R}^{(n+p) \times p}$ so gewählt, dass die erweiterte Matrixfunktion

$$C(y) := \begin{pmatrix} Df(y) \\ B^T \end{pmatrix}$$

für alle $y \in \mathcal{M}_0$ regulär ist, so bilden die p Lösungen $\tau_i(y)$ der Gleichungssysteme

$$C(y)\tau_i = e_{n+i}, \quad i = 1(1)p, \tag{5.74}$$

mit den $(n + p)$ -dimensionalen Einheitsvektoren e_{n+i} einen Moving Frame für \mathcal{M}_0 . Ein Prädiktor-Korrektor-Schritt dieser Vorgehensweise wird in Algorithmus 5.43 vorgestellt. Er berechnet zu gegebenem Punkt y_j und zugehöriger Orthonormalbasis B für die i_{\max} Parametervektoren $s_1, \dots, s_{i_{\max}}$ die zugehörigen Punkte y_{neu}^i , $i = 1(1)i_{\max}$, und aktualisiert die Orthonormalbasis in B_{neu} . Zur weiteren Umsetzung dieses Verfahrens, z. B. mit geeigneten Triangulierungen der Mannigfaltigkeit, empfehlen sich die Darstellungen von Henderson (2007) und Mei (2000).

ALGORITHMUS 5.43 (Fortsetzung von Mannigfaltigkeiten)

Function $[y_{\text{neu}}, B_{\text{neu}}] = \text{moving_frame}(f, Df, y, s, B, \text{tol})$

1. (Punkte-Bestimmung)

For $i = 1(1)i_{\text{max}}$ do: // Durchlaufen von $s = (s_1, \dots, s_{i_{\text{max}}})$

1.1. (Fortsetzungsparameter und Prädiktor)

Setze $s := s_i$ und Startpunkt $w := 0$.

1.2. (Newton-Korrektor für (5.72)):

Do while $\|f(y + Bs + w)\| > \text{tol}$:

$$\text{Löse } \begin{pmatrix} Df(y + Bs + w) \\ B^T \end{pmatrix} \Delta w = - \begin{pmatrix} f(y + Bs + w) \\ B^T w \end{pmatrix}$$

$$w := w + \Delta w$$

1.3. $y_{\text{neu}}^i := y + Bs + w$ **2. (Basis-Bestimmung)**

Mit y löse die linearen Systeme (5.74) nach $\tau_1, \tau_2, \dots, \tau_p$.

3. (Orthonormierung)

Orthonormiere $\tau_1, \tau_2, \dots, \tau_p$, womit die neue Orthonormalbasis

$B_{\text{neu}} = \{b_1, b_2, \dots, b_p\}$ entsteht.

4. Return $y_{\text{neu}}, B_{\text{neu}}$ **5.2.5 Fortsetzung mit Ableitungsprädiktoren**

In den nachfolgenden Kapiteln sind die Fortsetzungsprobleme für periodische und quasi-periodische Orbits durch folgende erschwerende Umstände charakterisiert:

- $f(x, \lambda)$ ist nicht durch einen exakten arithmetischen Ausdruck definiert, sondern Resultat eines numerischen Integrationsverfahrens für DGL. Die Jacobi-Matrix $D_x f(x, \lambda)$ ist so nicht exakt verfügbar und das Newton-Verfahren nicht anwendbar.
- Der Zeitaufwand zur Berechnung eines einzigen Funktionswertes $f(x, \lambda)$ beträgt ein Vielfaches des algebraischen Aufwandes zur Lösung des linearen Gleichungssystems im Korrektorschritt. Rechenzeitverhältnisse des „Funktionswertaufwandes“ zum rein algebraischen Aufwand von über 100 : 1 sind damit keine Seltenheit!
- Die Stabilitäts- und Bifurkationsanalyse erfordert zudem eine gute Approximation der Jacobi-Matrizen $A^* := D_x f(x_j, \lambda_j)$, $j = 0, 1, 2, \dots$, um deren Spektren $\sigma(A^*)$ genau ermitteln zu können.

Als Ausweg bietet sich der vereinfachte Newton-Korrektor (5.34) bei $\lambda = \lambda_j$

$$u_{k+1} := u_k - (A^P)^{-1} f(u_k, \lambda_j), \quad k = 0, 1, 2, \dots \quad (5.75)$$

mit einer Differenzenapproximation A^P gemäß (5.35) an. Neben einem guten Prädiktorwert $x^P = u_0$ mittels Prädiktoren hoher Ordnung ist eine genaue Approximation A^P der zu bestimmenden Jacobi-Matrix $A_j := D_x f(x_j, \lambda_j)$ gefordert. Wenig geeignet sind dafür Matrizen A_{j-1} des vorhergehenden Fortsetzungsschrittes. Konstruiert man jedoch einen Prädiktor höherer Ordnung für A^P mit Hilfe der zurückliegenden A_j -Werte, so entstehen mit geringem Zusatzaufwand effiziente Fortsetzungsverfahren.

Beispiel 5.45

$A_j := D_x f(x_j, \lambda_j)$ bezeichne die Jacobi-Matrizen an den Punkten λ_j mit $x_j = x(\lambda_j)$. Dann konstruieren wir für die Ableitung an der Stelle λ_j den

- Basis-Prädiktor: $A^P := A_{j-1}$,
- Sekanten-Prädiktor: $A^P := A_{j-1} + \frac{\lambda_j - \lambda_{j-1}}{\lambda_{j-1} - \lambda_{j-2}} \cdot [A_{j-1} - A_{j-2}]$ sowie die
- Polynom-Prädiktoren: $A^P := \alpha_1 A_{j-1} + \alpha_2 A_{j-2} + \dots + \alpha_k A_{j-k}$, $\alpha_i \in \mathbb{R}$,
mittels polynomialer Extrapolation. ■

Um den modifizierten Fortsetzungsschritt in $[\lambda_{j-1}, \lambda_j]$ zu verifizieren, benutzen wir einen Konvergenzsatz für approximative Newton-Verfahren von Kelley (1995). Wir betrachten dazu das nichtlineare System $f(x, \lambda_j) = 0$ mit festem Parameterwert λ_j .

Satz 5.46

Falls x_j eine reguläre Nullstelle von $f(x, \lambda_j)$ im Sinne der Definition 5.4 ist, so existieren Konstanten $K_0 > 0$, $\delta_1 > 0$ und $\delta_2 > 0$, so dass für alle u_0 mit $\|u_0 - x_j\| < \delta_1$ und für alle $A \in \mathbb{R}^{n \times n}$ mit $\|A - D_x f(x_j, \lambda_j)\| < \delta_2$ die Iteration $u_{k+1} = u_k - A^{-1} f(u_k, \lambda_j)$, $k = 0, 1, 2, \dots$, Q-linear gegen x_j konvergiert. Mit $e_k := u_k - x_j$ gilt die Abschätzung

$$\|e_{k+1}\| \leq K_0 (\|e_0\| + \|A - D_x f(x_j, \lambda_j)\|) \cdot \|e_k\|, \quad k = 0, 1, 2, \dots \quad (5.76)$$

Die Fehlerschätzung (5.76) zeigt eine gleiche Wichtung der Prädiktoren x^P und A^P , weshalb sich dieselbe Approximationsordnung für beide, d. h. $x^P - x_j = O(h^p)$ und $A^P - D_x f(x_j, \lambda_j) = O(h^p)$ empfiehlt. In Algorithmus 5.47 benutzen wir dazu den bewährten Sekanten-Prädiktor. Um den Fortsetzungsalgorithmus zu verifizieren, führen wir entsprechend Formel (5.31) eine Approximationsordnung für beide Prädiktoren ein. Die Schrittweiten h_j seien darin beschränkt durch $h_{\max} := \max_{j=1(1)N} h_j$.

Voraussetzung 5.48

1. Der Lösungsprädiktor besitze die Ordnung p_1 , d. h. es existiert ein $K_1 > 0$, so dass

$$\|x^P(\lambda) - x(\lambda)\| \leq K_1 h^{p_1}, \quad |h| < h_{\max} \quad (5.77)$$

für $\lambda = \lambda_{j-1} + h$, mit $\lambda_{j-1}, \lambda \in \Lambda$, gilt.

ALGORITHMUS 5.47 (Fortsetzung mit Ableitungsprädiktor)

Function $[N, \lambda, x] = \text{continuation2}(f, Df x, a, b, x, \text{tol})$

1. Setze $j := 0$, $\lambda_0 := a$, $x^P := x$ und $A^P := D_x f(x^P, \lambda_0)$.

Wähle Schrittweite h sowie Iterationszahl k_{opt} .

2. (Startrechnung) Bestimme $x_0 = x(\lambda_0)$ mit Verfahren (5.75).

Setze $A_0 := D_x f(x_0, \lambda_0)$.

3. Do while $\lambda_j \leq b$

3.1. Setze $j := j + 1$.

3.2. Setze $\lambda_j := \lambda_{j-1} + h$. Falls $\lambda_j > b$, so $\lambda_j := b$.

3.3. (Lösungsprädiktor)

Bestimme x^P mit Hilfe von $x_{j-1}, x_{j-2}, \dots, x_0$, z. B. mit

$$x^P := x_{j-1} + \frac{\lambda_j - \lambda_{j-1}}{\lambda_{j-1} - \lambda_{j-2}} \cdot [x_{j-1} - x_{j-2}].$$

3.4. (Ableitungsprädiktor)

Bestimme A^P mit Hilfe von $A_{j-1}, A_{j-2}, \dots, A_0$, z. B. mit

$$A^P := A_{j-1} + \frac{\lambda_j - \lambda_{j-1}}{\lambda_{j-1} - \lambda_{j-2}} \cdot [A_{j-1} - A_{j-2}].$$

3.5. (Korrektor) Bestimme $x_j = x(\lambda_j)$ mit Verfahren (5.75).

3.6. (Ableitungs-Bestimmung) Berechne $A_j = D_x f(x_j, \lambda_j)$.

3.7. (Schrittweiten-Bestimmung) wie in Algorithmus 5.28

4. Return $N := j$ sowie $\lambda_j, x_j, j = 0(1)N$

2. Der Ableitungsprädiktor besitze die Ordnung p_2 , d. h. es existiert ein $K_2 > 0$, so dass

$$\|A^P(\lambda) - D_x f(x(\lambda), \lambda)\| \leq K_2 h^{p_2}, \quad |h| < h_{\text{max}} \quad (5.78)$$

für $\lambda = \lambda_{j-1} + h$, mit $\lambda_{j-1}, \lambda \in \Lambda$, gilt.

Das folgende Lemma garantiert dann, dass das Korrektorverfahren für „hinreichend gute“ Prädiktorwerte x^P und A^P stets konvergiert:

Lemma 5.49

Unter den Voraussetzungen 5.23 und 5.48 existieren Konstanten $K_0 > 0$, $\delta_1 > 0$ und $\delta_2 > 0$, so dass folgende Behauptungen gelten:

i. $f(x, \lambda) = 0$ besitzt für alle $\lambda \in \Lambda$ eine eindeutige Lösung $x(\lambda)$.

ii. Zu festem $\lambda_{j-1} \in \Lambda$ konvergiert das Korrektorverfahren mit $\lambda_j = \lambda_{j-1} + h_j$

$$u_{k+1} = u_k - (A^P)^{-1} f(u_k, \lambda_j), \quad k = 0, 1, 2, \dots$$

gegen $x_j = x(\lambda_j)$, falls $0 < h_j < H := \min \left\{ P\sqrt{\delta_1/K_1}, P\sqrt{\delta_2/K_2} \right\}$ gilt.

iii. Falls $H := \min_{i=1,2} \left\{ P\sqrt{\delta_i/K_i}, P\sqrt{1/(4K_0K_i)} \right\}$ gewählt wird, so gilt

$$\|e_{k+1}\| \leq K_0 (K_1 h^{p_1} + K_2 h^{p_2}) \|e_k\| \leq \frac{1}{2} \|e_k\|, \quad k = 0, 1, 2, \dots \quad (5.79)$$

für $e_k = u_k - x(\lambda_j)$.

Um Algorithmus 5.47 abzusichern, ist wie in Satz 5.27 eine explizite Auswahlregel für die Schrittweiten h_j anzugeben, die die Durchführbarkeit und Endlichkeit der Fortsetzung garantiert. Mittels vollständiger Induktion über j zeigt man leicht

Satz 5.50

Die Voraussetzungen 5.23 und 5.48 seien erfüllt. Ferner genügen die Startwerte x^P und A^P am Startpunkt $\lambda_0 = a$ den Voraussetzungen

$$\|x^P - x(\lambda_0)\| < \delta_1 \quad \text{und} \quad \|A^P - D_x f(x(\lambda_0), \lambda_0)\| < \delta_2$$

mit den nach Lemma 5.49 garantierten Konstanten $\delta_1 > 0$, $\delta_2 > 0$. Wählt man die Schrittweite $h_j := \lambda_j - \lambda_{j-1}$ gemäß

$$h_j := \min \{H/2, b - \lambda_{j-1}\}, \quad \text{mit } H \text{ aus Teil iii. des Lemmas,} \quad (5.80)$$

so ist Algorithmus 5.47 durchführbar, d. h. er liefert zu jedem λ_j genau eine Lösung $x(\lambda_j)$ der Gleichung $f(x, \lambda_j) = 0$. Er ist zudem endlich, d. h. $\lambda_N = b$ für ein $N \in \mathbb{N}$.

Eine effiziente Wahl der Fortsetzungsschrittweiten h_j kann wie in Algorithmus 5.28 erfolgen. Die Berechnung der Jacobi-Matrizen $A_j = D_x f(x_j, \lambda_j)$ in Schritt 3.6 sollte mit der Stabilitätsanalyse verbunden werden, die im nachfolgenden Abschnitt behandelt wird.

Beispiel 5.51

Doogee und Kuznetsov (2002) betrachten das Bratu-Gelfand-System mit Parameter λ

$$\begin{aligned} \dot{x} &= f(x, y, \lambda) = y - 2x + \lambda e^x \\ \dot{y} &= f(x, y, \lambda) = x - 2y + \lambda e^y. \end{aligned} \quad (5.81)$$

Im Parameterintervall $\lambda \in [0.01, 0.367879]$ besitzt das System zwei Gleichgewichtspunkte, die am kritischen Wert $\lambda^* = 0.367879\dots$ zusammenfallen. Abbildung 5.12 zeigt die beiden Lösungszweige L_1 (stabil) und L_2 (instabil) sowie den Fold-Punkt $(\lambda, x, y) = (0.367879\dots, 1, 1)$. Wir notieren nun dieses System zu Testzwecken

$$\left. \begin{aligned} \dot{x}_i &= f_i(x, y, \lambda) = y_i - 2x_i + \lambda e^{x_i} \\ \dot{y}_i &= f_i(x, y, \lambda) = x_i - 2y_i + \lambda e^{y_i} \end{aligned} \right\} \quad \text{für } i = 1, 2, \dots, \frac{n}{2} \quad (5.82)$$

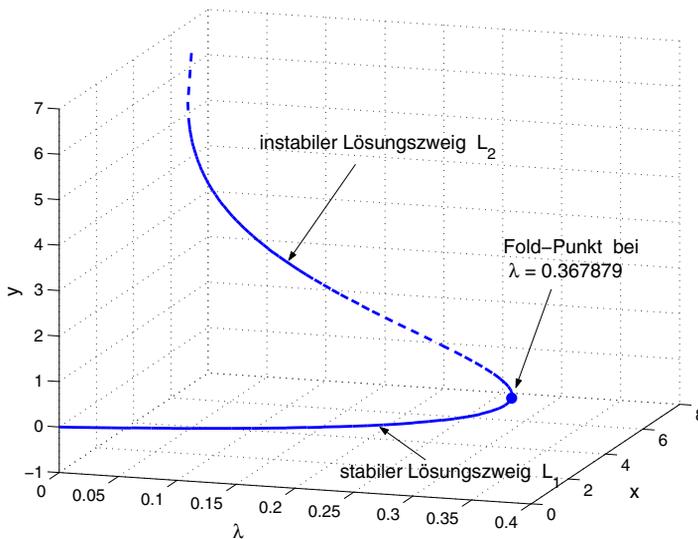


Abb. 5.12 Lösungsweige L_1 und L_2 stabiler und instabiler Gleichgewichtslagen des Bratu-Gelfand-Systems (5.81)

mit Systemgrößen $n = 2, 20, 100, 200, 400$, um damit folgende vier Versionen der Approximation der Jacobi-Matrizen zu vergleichen:

Version 1: Sekanten-Prädiktor x^P und übliche Approximation $A^P := D_x f(x^P, \lambda_j)$

Version 2: Sekanten-Prädiktor x^P und Basis-Prädiktor $A^P := A_{j-1}$

Version 3: Sekanten-Prädiktor x^P und Sekanten-Prädiktor A^P gemäß Algorithmus 5.47

Version 4: Prädiktor 2. Ordnung x^P und Prädiktor 2. Ordnung A^P .

Beginnend mit $\lambda = 0.01$ verfolgen wir den instabilen Zweig L_2 bis zu $\lambda = 0.36787$ und analysieren bei jedem λ_j zudem die Stabilität. Startwerte sind $x_i = y_i = 10$ für $i = 1, 2, \dots, \frac{n}{2}$; die Genauigkeit ist stets $tol = 10^{-10}$. Tabelle 5.2 zeigt die Überlegenheit

Tab. 5.2 Anzahl N der Fortsetzungsschritte und Rechenzeit in sec (Pentium 4 M 1.60 GHz)

n	Version 1		Version 2		Version 3		Version 4	
	N	Zeit	N	Zeit	N	Zeit	N	Zeit
2	69	0.10	280	0.28	50	0.08	37	0.08
20	77	0.36	353	0.85	54	0.20	39	0.16
100	86	2.52	412	7.94	60	1.26	40	0.93
200	88	9.10	433	31.56	62	4.92	41	3.51
400	91	46.13	457	186.3	63	27.42	41	18.74

der Versionen 3 und 4 mit gleicher Approximationsordnung von Lösung und Jacobi-Matrix. Version 4 kann so mit den größten internen Fortsetzungsschrittweiten arbeiten, wie man in Abbildung 5.13 leicht erkennt. ■

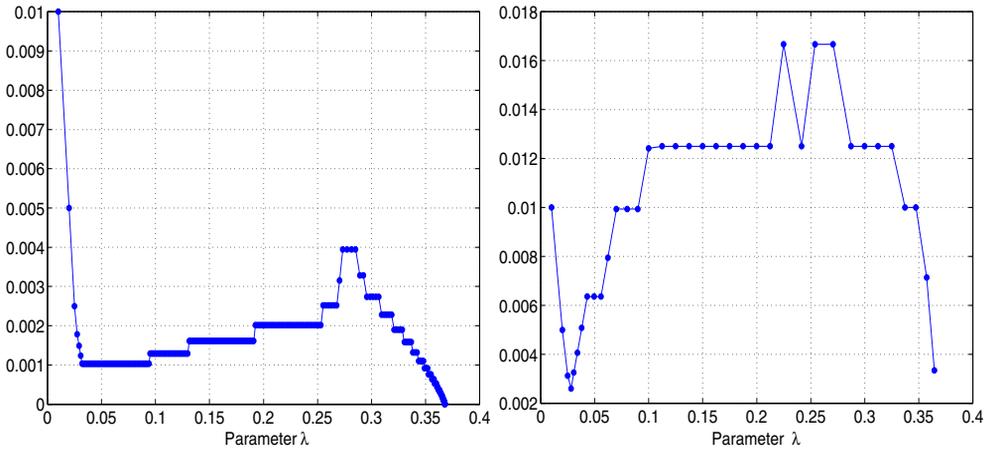


Abb. 5.13 Interne Fortsetzungsschrittweiten h der Versionen 2 (links) und 4 (rechts)

Die Idee, Prädiktoren hoher Ordnung für die benötigten Jacobi-Matrizen zu benutzen, lässt sich mit entsprechendem Aufwand auch bei Bogenlängen-Parametrisierung und Gauß-Newton-Fortsetzung in den Algorithmen 5.34 und 5.38 leicht anwenden.

5.3 Stabilitäts- und Bifurkationsanalyse

Die Klärung des Stabilitätsverhaltens von Gleichgewichtslösungen des Systems

$$\dot{x} = f(x, \lambda), \quad f : D \times \mathbb{R} \rightarrow \mathbb{R}^n, \quad D \subset \mathbb{R}^n, \text{ offen}, \quad \lambda \in \Lambda = [a, b] \quad (5.83)$$

mit C^r -glattem Vektorfeld $f \in C^r(D \times \mathbb{R})$ mit $r \geq 1$, ist von erheblicher praktischer Bedeutung, denn nur asymptotisch stabile Lösungen sind wegen unvermeidlicher Störungen in den Anfangswerten physikalisch relevant. Ein direkter Nachweis der Ljapunov-Stabilität mittels einer Ljapunov-Funktion wie in Abschnitt 4.2 ist im Allgemeinen nicht möglich, weshalb die indirekte Methode aus Abschnitt 2.7.2 für das parameterabhängige System (5.83) eingesetzt wird.

5.3.1 Stabilitätsanalyse

Zu festem Parameterwert $\lambda \in \Lambda$ sei $x^0 = x^0(\lambda)$ ein hyperbolischer Fixpunkt des Flusses. Bestimmen wir die Jacobi-Matrix $A := A(\lambda) = D_x f(x^0(\lambda), \lambda)$ und deren Eigenwerte $\sigma(A(\lambda)) = \{\mu_1(\lambda), \mu_2(\lambda), \dots, \mu_n(\lambda)\}$, die sogenannten *charakteristischen Exponenten*, so liefert uns das Stabilitätskriterium die numerisch überprüfbare Aussage:

- Ist der Realteil $\text{Re } \mu_i(\lambda) < 0$ für alle $i \in \{1, 2, \dots, n\}$, so ist $x^0(\lambda)$ asymptotisch stabil.
- Existiert ein $i \in \{1, 2, \dots, n\}$ mit $\text{Re } \mu_i(\lambda) > 0$, so ist $x^0(\lambda)$ instabil.

Beispiel 5.52

Betrachten wir das eindimensionale Beispiel mit dem Systemparameter $\lambda \in [-1, 1]$

$$\dot{x} = f(x, \lambda) = x(x^3 - x - \lambda),$$

das die beiden Zweige $\mathcal{L}_0 = \{(x, \lambda) \mid x = 0\}$ und $\mathcal{L}_1 = \{(x, \lambda) \mid x^3 - x - \lambda = 0\}$ von Gleichgewichtslösungen besitzt. Auf \mathcal{L}_1 berechnen wir die Ableitung

$$D_x f(x, \lambda) = 4x^3 - 2x - \lambda = 3x(x^2 - 1/3),$$

woraus sich der Eigenwert $\mu = 3x(x - \sqrt{3}/3)(x + \sqrt{3}/3)$ ergibt. Mit $\mu = 0$ erhalten wir die kritischen Punkte $y^0 = (x^0, \lambda_0) = (0, 0)$, $y^1 = (x^1, \lambda_1) = (\sqrt{3}/3, -2\sqrt{3}/9)$ und $y^2 = (x^2, \lambda_2) = (-\sqrt{3}/3, 2\sqrt{3}/9)$ in Abbildung 5.14. Ein Stabilitätswechsel erfolgt offenbar nur an diesen drei Punkten, die die beiden Lösungszweige \mathcal{L}_0 und \mathcal{L}_1 in die dargestellten Abschnitte mit asymptotisch stabilem bzw instabilem Verhalten zerlegen.

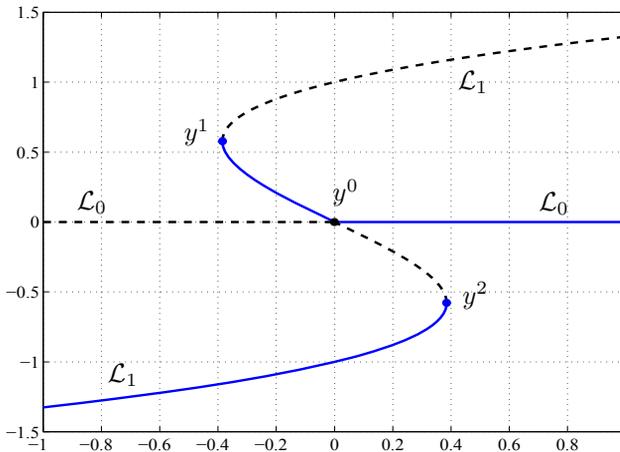


Abb. 5.14 Lösungszweige \mathcal{L}_0 und \mathcal{L}_1 zu Beispiel 5.52: stabile Zweige (blau) und instabile Zweige (schwarz)

Wir modifizieren die DGL nun geringfügig zu der Gleichung

$$\dot{x} = f(x, \lambda) = x(x^3 - x - \lambda^2), \quad \lambda \in [-1, 1],$$

und betrachten deren trivialen Lösungszweig $\mathcal{L}_0 = \{(x, \lambda) \mid x = 0\}$. Wegen der Ableitung $D_x f(x, \lambda) = 4x^3 - 2x - \lambda^2$ lautet der Eigenwert $\mu = -\lambda^2 < 0$ für alle $\lambda \neq 0$. Also ist die Lösung $x^0 = 0$ asymptotisch stabil für alle $\lambda \neq 0$. Bei $\lambda_0 = 0$ verschwindet zwar der Realteil $\operatorname{Re} \mu$, es findet jedoch kein Stabilitätswechsel statt. ■

Über nicht-hyperbolische Fixpunkte kann im nichtlinearen Fall mit dem Eigenwertkriterium nicht entschieden werden. Ist dagegen die *Spektralabszisse* der Matrix A

$$\nu(A) := \max_{\mu_i \in \sigma(A)} \operatorname{Re} \mu_i \quad (5.84)$$

negativ, so ist $x(t) = x^0$ asymptotisch stabile Gleichgewichtslage. Damit können wir leicht den Schritt 3.6 in Algorithmus 5.47 modifizieren:

■ Stabilitätsanalyse

- (a) Berechne die Jacobi-Matrix $A_j := D_x f(x^j, \lambda_j)$.
- (b) Bestimme die Eigenwerte $\mu_1, \mu_2, \dots, \mu_n$ von A_j
und die Spektralabszisse $\nu := \nu(A_j)$.
- (c) Falls $\nu < 0$, so "ASYMPTOTISCH STABIL",
sonst falls $\nu > 0$, so "INSTABIL",
sonst "KEINE AUSSAGE MÖGLICH".

Beispiel 5.53

Wir betrachten nochmals Beispiel 5.35 mit $f(x, \lambda)$, $x = (x_1, x_2, x_3, x_4, x_5)^T$, gegeben durch die fünf dortigen Funktionen. Die Steuervariable ist $u = 0$ und der Parameterbereich $-1 \leq \lambda \leq 1$. Abbildung 5.15 zeigt die numerisch berechnete Lösung x_4 , dargestellt über dem Parameter λ , wobei nun die Punkte mit nachgewiesener Instabilität schwarz gezeichnet wurden. Offenbar sind die Stabilitätswechsel noch genauer einzugrenzen. ■

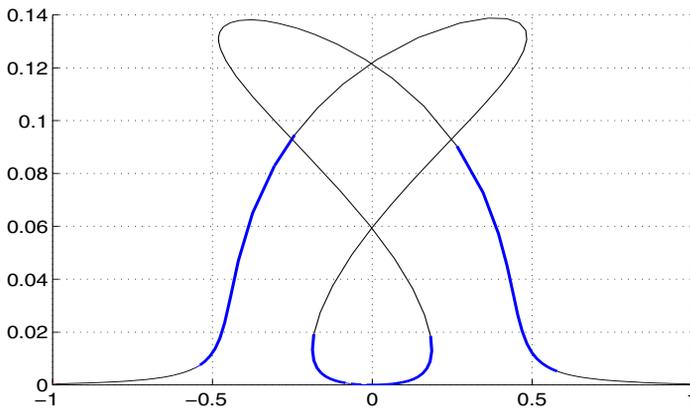


Abb. 5.15
Lösungskomponente x_4 über dem Parameter λ : stabile Zweige (blau) und instabile Zweige (schwarz)

5.3.2 Detektierung lokaler Bifurkationen

Mit der Stabilitätsanalyse kann das Auftreten lokaler Bifurkationen während der Lösungsfortsetzung entdeckt (detektiert) werden. In Abschnitt 2.7.1 werden zur Jacobi-Matrix $A := A(\lambda) = D_x f(x^0(\lambda), \lambda)$ der

- stabile Unterraum E^s der Dimension $n_s = \dim E^s =$ Zahl der μ_i mit $\operatorname{Re} \mu_i < 0$
- instabile Unterraum E^u der Dimension $n_u = \dim E^u =$ Zahl der μ_i mit $\operatorname{Re} \mu_i > 0$
- Zentrumsunterraum E^c der Dimension $n_c = \dim E^c =$ Zahl der μ_i mit $\operatorname{Re} \mu_i = 0$

mit $n_s + n_u + n_c = n$ definiert. Bestimmen wir im j -ten Fortsetzungsschritt diese Dimensionen n_s^j, n_u^j und n_c^j , so signalisiert eine Änderung dieser Werte bei zwei aufeinander

folgenden Schritten die Möglichkeit einer lokalen Bifurkation im (hinreichend kleinen) Intervall $[\lambda_{j-1}, \lambda_j]$. Damit kann Schritt 3.6 in Algorithmus 5.47 leicht modifiziert werden:

■ **Bifurkationsdetektierung**

- (a) Berechne die Jacobi-Matrix $A_j := D_x f(x^j, \lambda_j)$.
- (b) Bestimme die Eigenwerte $\mu_1, \mu_2, \dots, \mu_n$ von A_j und die Dimensionen n_s^j, n_u^j, n_c^j der Unterräume E^s, E^u, E^c .
- (c) Falls $n_s^j \neq n_s^{j-1} \vee n_u^j \neq n_u^{j-1} \vee n_c^j \neq n_c^{j-1}$, so
 "BIFURKATION IN $[\lambda_{j-1}, \lambda_j]$ MÖGLICH".

Die ständige Berechnung aller charakteristischen Exponenten μ_i während der Lösungsfortsetzung ist bei großen Systemen sehr zeitintensiv. So gewinnen einfachere Kriterien an Bedeutung. Allerdings sind diese im Allgemeinen nur notwendig, nicht aber hinreichend für den jeweiligen Bifurkationstyp. Der Nachweis hinreichender Bedingungen gemäß Abschnitt 3.2 ist meistens nicht praktikabel.

Wir nehmen desweiteren an, dass einer der folgenden Bifurkationstypen auftreten kann:

- Sattel-Knoten-Bifurkation (desweiteren auch als Falten-Bifurkation bezeichnet)
- transkritische Bifurkation
- Heugabel-Bifurkation (Pitchfork-Bifurkation) sowie
- Hopf-Bifurkation.

Mit der zusammenfassenden Bezeichnung $y = (x, \lambda)$ und $f(y) = f(x, \lambda)$ sei $y^0 = (x^0, \lambda_0)$ mit dem Gleichgewichtspunkt x^0 gegeben, d. h. $f(y^0) = 0$. Am Punkt y^0 definieren wir die Matrizen der Dimension $n \times n$ und $n \times (n + 1)$

$$A := D_x f(x^0, \lambda_0) \quad \text{und} \quad B := f'(y^0) = (D_x f(x^0, \lambda_0), D_\lambda f(x^0, \lambda_0)).$$

Über die Nullräume von A und B gestattet die Theorie der Kapitel 3 und 4 dann folgende Aussage:

Satz 5.54

Es sei $f \in C^r(D \times \Lambda)$, $r \geq 2$. Dann sind die Nullraumdimensionen

$$\dim N(A) = \begin{cases} 0, & \text{falls } x^0 \text{ regulärer oder Hopf-Punkt} \\ 1, & \text{falls } x^0 \text{ Falten-Punkt, transkritischer oder Heugabel-Punkt ist.} \end{cases}$$

$$\dim N(B) = \begin{cases} 1, & \text{falls } x^0 \text{ regulärer Punkt} \\ 1, & \text{falls } x^0 \text{ Falten- oder Hopf-Punkt} \\ 2, & \text{falls } x^0 \text{ transkritischer oder Heugabel-Punkt ist.} \end{cases}$$

Bei Parametrisierung der Lösungskurve $y = y(s) = (x(s), \lambda(s))$ nach ihrer Bogenlänge oder Pseudo-Bogenlänge s steht zudem der normierte Tangentenvektor $y'(s_0)$ am betrachteten Punkt $y^0 = (x^0, \lambda_0) = (x(s_0), \lambda(s_0))$ zur Verfügung. Damit definieren wir die erweiterte Matrix

$$C := \begin{pmatrix} f'(y(s_0)) \\ y'(s_0)^T \end{pmatrix} = \begin{pmatrix} B \\ y'(s_0)^T \end{pmatrix}, \quad C \in \mathbb{R}^{(n+1) \times (n+1)}.$$

Wegen $f(y(s)) = 0$ folgt durch Differentiation $B \cdot y'(s_0) = 0$, so dass $y'(s_0)$ linear unabhängig von den Zeilen von B ist. Damit ist die Nullraumdimension von C um Eins kleiner als diejenige von B . Mit den Matrizen A, B und C ist dann die in Tabelle 5.3 angegebene teilweise Klassifikation der Bifurkationspunkte möglich.

Tab. 5.3 Nullraumdimension zur Klassifikation der Punkte x^0

Typ der Bifurkationspunkte x^0	Parametrisierung nach λ		Parametrisierung nach s
	A	B	C
regulärer Punkt	0	1	0
Falten-Bifurkation	1	1	0
transkritische Bifurkation	1	2	1
Heugabel-Bifurkation	1	2	1
Hopf-Bifurkation	0	1	0

Um die numerisch wenig geeigneten Dimensions- und Rangbestimmungen zu vermeiden, führen wir mit Hilfe der Determinanten von A und C skalare Testfunktionen ψ ein. Mit $C = (x(s), \lambda(s))$ bezeichnen wir den verfolgten Zweig von Gleichgewichtslösungen mit Parameter $s \in \mathbb{R}$.

Definition 5.55 (Testfunktion)

Es sei \mathcal{U} eine Umgebung von C . Eine stetige Funktion $\psi : \mathcal{U} \rightarrow \mathbb{R}$, die beim Durchlaufen der Kurve $C = (x(s), \lambda(s))$ genau in $s = s_0$ ihr Vorzeichen (strikt) ändert, heißt *Testfunktion* für den Bifurkationspunkt $(x^0, \lambda_0) = (x(s_0), \lambda(s_0))$.

Ist die Testfunktion stetig differenzierbar, so garantieren die Eigenschaften

$$\psi(x(s_0), \lambda(s_0)) = 0, \quad \left. \frac{d}{ds} \psi(x(s), \lambda(s)) \right|_{s=s_0} \neq 0, \quad (5.85)$$

dass s_0 eine reguläre Nullstelle ist und der Bifurkationspunkt (x^0, λ_0) direkt während der Lösungsfortsetzung beobachtet werden kann. Wir wollen nun geeignete Testfunktionen für die genannten Bifurkationstypen einführen.

Detektierung der Sattel-Knoten-Bifurkation (Falten-Bifurkation)

Definieren wir mit den Determinanten die beiden Funktionen auf dem Lösungszweig C

$$\psi_F(y) := \det(D_x f(y)) \quad \text{und} \quad \psi_B(y) := \det \begin{pmatrix} f'(y) \\ y'(s)^T \end{pmatrix}, \quad (5.86)$$

so können $\psi_F(x(s), \lambda(s))$ und $\psi_B(x(s), \lambda(s))$ als Testfunktionen fungieren. Aus Tabelle 5.3 und der Eigenwertdarstellung $\psi_F(y) = \det(D_x f(y)) = \mu_1 \mu_2 \cdots \mu_n$ gewinnen wir zusammen mit den Eigenschaften der betreffenden Bifurkationstypen aus Kapitel 3

Satz 5.56

Es sei $f \in C^r(D \times \Lambda)$, $r \geq 2$. $y^0 = (x^0, \lambda_0)$ gegeben mit $f(y^0) = 0$.

- i. $\psi_F(y)$ und $\psi_B(y)$ sind stetig differenzierbar.
- ii. Ist x^0 regulärer oder Hopf-Punkt, so ist $\psi_F(y^0) \neq 0$.
- iii. Liegt bei x^0 eine Falten-Bifurkation vor, so ist $\psi_F(y^0) = 0$ und $\psi_B(y^0) \neq 0$.
- iv. Ist x^0 ein Falten-Punkt, so wechselt $\psi_F(y)$ bei $y = y^0$ das Vorzeichen.

Beispiel 5.57

In Beispiel 5.52 mit der Funktion $f(x, \lambda) = x(x^3 - x - \lambda)$ verfolgen wir nun den Lösungszweig $\mathcal{L}_1 = \{(x, \lambda) \mid x^3 - x - \lambda = 0\}$ aus Abbildung 5.14. In der Umgebung des Kurvenpunktes $y^1 = (x^1, \lambda_1) = (\sqrt{3}/3, -2\sqrt{3}/9)$ parametrisieren wir die Kurve nach x , d. h. $y(s) = (x(s), \lambda(s)) = (s, s^3 - s)$, und erhalten damit den Wert der Testfunktion

$$\psi_F(y(s)) = 4x(s)^3 - 2x(s) - \lambda(s) = 3s^3 - s, \quad \text{also} \quad \psi_F(y^1) = 0.$$

Wegen der Ableitung an dieser Stelle $y^1 = y(s_1)$ mit $s_1 = \sqrt{3}/3$

$$\left. \frac{d}{ds} \psi(y(s)) \right|_{s=s_1} = 9s_1^2 - 1 = 2 > 0$$

erfolgt bei y^1 ein Nulldurchgang. Für den Tangentenvektor am Punkt y^1 erhalten wir $y'(s_1) = (x'(s_1), \lambda'(s_1))^T = (1, 0)^T$, womit wir die erweiterte Matrix

$$C = \begin{pmatrix} f'(y(s_1)) \\ y'(s_1)^T \end{pmatrix} = \begin{pmatrix} 0 & -\sqrt{3}/3 \\ 1 & 0 \end{pmatrix}$$

berechnen und für die zweite Testfunktion $\psi_B(y_1) = \det C = \sqrt{3}/3 \neq 0$ erhalten. ■

Detektierung der transkritischen und Heugabel-Bifurkation

Falls $\psi_F(y^0) = 0$ und $\psi_B(y^0) = 0$ wird, so ist ein Verzweigungspunkt im engeren Sinne (branching point) – in der Form einer transkritischen bzw. Heugabel-Bifurkation – zu vermuten. Dieses Kriterium ist zwar notwendig, jedoch nicht hinreichend. Die Überprüfung hinreichender Bedingungen ist in praktischen Fällen meist zu aufwändig, weshalb man es bei der Berechnung von $\psi_F(y^0)$ und $\psi_B(y^0)$ belässt.

Beispiel 5.58

Wir setzen Beispiel 5.57 fort und untersuchen den Punkt $y^0 = (x^0, \lambda_0) = (0, 0)$. Offenbar ist $\psi_F(y^0) = 4x(s)^3 - 2x(s) - \lambda(s) = 3s^3 - s$, also $\psi_F(y^0) = 0$. Wir berechnen nun $\psi_B(y^0)$ auf beiden Lösungszweigen: Auf $\mathcal{L}_0 = \{(x, \lambda) \mid x = 0\}$ ist $f'(0, 0) = (0, 0)$ und der Tangentenvektor trivialerweise $y'(0) = (0, 1)^T$, womit $\psi_B(y^0) = \det C = 0$ folgt. Auch auf dem zweiten Lösungsweig \mathcal{L}_1 erhalten wir für $y^0 = 0$ den Wert $\psi_B(y^0) = 0$, womit bei der numerischen Verfolgung beider Zweige stets ein Verzweigungspunkt im engeren Sinne – in diesem Falle mit transkritischer Bifurkation – signalisiert wird. ■

Ein wesentliches Problem beim Auftreten von Verzweigungspunkten $y^* = (x^*, \lambda^*)$ stellt deren Überwindung während der Parameterfortsetzung dar. Denn anders als bei Sattel-Knoten-Bifurkationen (Falten-Punkten) ist nun sowohl die Jacobi-Matrix $A = D_x f(y^*)$ als auch die für die Bogenlängen-Parameterisierung oder die Gauß-Newton-Fortsetzung wesentliche $(n + 1) \times (n + 1)$ -Matrix C singularär und damit die grundlegende Voraussetzung 5.30 iii dieser beiden Methoden verletzt. Damit versagt in der Regel der Newton-Korrektor aller bisherigen Fortsetzungsalgorithmen am Bifurkationspunkt y^* . Der Einzugsbereich des Korrektorverfahrens schrumpft bei Annäherung der Kurvenpunkte $y(s)$ gegen y^* auf Null; das betrifft genauso den anderen schneidenden Lösungsweig. Allerdings beweisen Allgower und Georg (1990) für die Gauß-Newton-Fortsetzung 5.38,

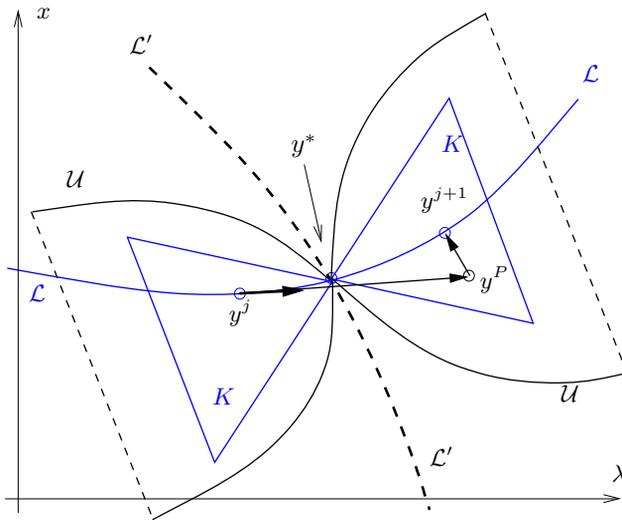


Abb. 5.16 Darstellung der sich schneidenden Lösungswege \mathcal{L} und \mathcal{L}' in der Umgebung des Verzweigungspunktes y^* sowie der Umgebung \mathcal{U} und des Kegels K

dass eine in Abbildung 5.16 dargestellte Kegeleumgebung $K \subset \mathbb{R}^{n+1}$ existiert, so dass das Korrektorverfahren für alle Startwerte $y^0 \in K$ konvergiert:

Satz 5.59

Es sei $y^* \in \mathbb{R}^{n+1}$ transkritischer oder Heugabel-Bifurkationspunkt. Dann existieren eine offene Umgebung \mathcal{U} von $\mathcal{L}_{reg} = \{y \mid f(y) = 0, y \text{ regulärer Punkt von } \mathcal{L}\}$ und Konstanten $\varepsilon > 0, \delta > 0$, so dass gilt:

- i. Der Gauß-Newton-Korrektor in Algorithmus 5.38 konvergiert für jeden Startpunkt $y^0 \in \mathcal{U}$ gegen einen Punkt $y^\infty \in \mathcal{L}$.
- ii. Die Abschätzungen des Satzes 5.39 bleiben für y^0, y^1, y^∞ gültig.
- iii. In der Umgebung \mathcal{U} existiert ein Kegel $K \subset \mathcal{U}$ mit

$$K = \{y \mid y = y^* + [y'(s_0) + z]s\} \quad \text{für } 0 < |s| < \varepsilon \quad \text{und} \quad \|z\|_2 < \delta.$$

In Abbildung 5.16 sind die sich in y^* schneidenden Lösungsbranche \mathcal{L} und \mathcal{L}' mit der Umgebung \mathcal{U} und dem Kegel K dargestellt. Um vom letzten bestimmten Kurvenpunkt y^j während der Fortsetzung stets in der Kegelmenge zu verbleiben, wird ein „Überspringen des Bifurkationspunktes“ y^* empfohlen: Mit hinreichend großer Fortsetzungsschrittweite ist ein Prädiktorwert y^P für den nächsten Kurvenpunkt y^{j+1} so zu bestimmen, dass er wiederum im Kegel K liegt. Dafür empfehlen sich besonders die in Abschnitt 5.2.1 angeführten Prädiktoren hoher Ordnung unter Benutzung zurückliegender Kurvenpunkte. Die algorithmische Umsetzung überlassen wir jedoch dem interessierten Leser.

Detektierung der Hopf-Bifurkation

Da im Fall eines Hopf-Punktes y^* die Jacobi-Matrix $A = D_x f(y^*)$ regulär ist, sind die in (5.86) eingeführten $\psi_F(y)$ und $\psi_B(y)$ nicht als Testfunktionen zu gebrauchen. Das skizzierte Vorgehen zur Bifurkationsdetektierung mittels der Unterräume E^s, E^u, E^c kann durch folgende skalare Testfunktion vereinfacht werden

$$\begin{aligned} \psi_H(y) &:= \prod_{i>k} [\mu_i(y) + \mu_k(y)] \\ &= (\mu_2(y) + \mu_1(y))(\mu_3(y) + \mu_1(y)) \cdots (\mu_n(y) + \mu_1(y)) \cdot \\ &\quad \cdot (\mu_3(y) + \mu_2(y)) \cdots (\mu_n(y) + \mu_2(y)) \cdots (\mu_n(y) + \mu_{n-1}(y)). \end{aligned} \tag{5.87}$$

Führen wir einige Eigenschaften von ψ_H an, um sie als Testfunktion einsetzen zu können:

Satz 5.60

Es sei $f \in C^r(D \times \Lambda)$, $r \geq 2$. $y^* = (x^*, \lambda^*)$ gegeben mit $f(y^*) = 0$.

- i. $\psi_H : D \times \Lambda \rightarrow \mathbb{R}$ ist eine glatte reellwertige Funktion.
- ii. Falls y^* Hopf-Punkt ist, so $\psi_H(y^*) = 0$.
- iii. Bei Parameterfortsetzung nach λ wechselt $\text{Re } \psi_H(y)$ für $\lambda = \lambda^*$ das Vorzeichen.

Eigenschaft (ii) folgt ohne Beschränkung der Allgemeinheit für einen Hopf-Punkt aus $\mu_2(y^*) = \overline{\mu_1(y^*)}$ mit $\text{Re } \mu_1(y^*) = 0$ wegen $\mu_1(y^*) + \mu_2(y^*) = 2\text{Re } \mu_1(y^*) = 0$. Ver-

mittels der Transversalitätsbedingung (3.19) des Satzes 3.8 zeigt man unmittelbar, dass $\frac{d}{d\lambda} \operatorname{Re} \psi_H(x(\lambda), \lambda) \neq 0$ für $\lambda = \lambda^*$ gilt. Ist y^* ein Falten-, Heugabel- oder transkritischer Bifurkationspunkt mit $\mu_1(y^*) = 0$, so ist zusätzlich zu fordern, dass $\mu_i(y^*) \neq -\mu_k(y^*)$ für alle $i, k > 1$ ist, um $\psi_H(y^*) \neq 0$ zu garantieren. Eine Darstellung dieser Testfunktion, die gänzlich ohne Eigenwertbestimmung oder -approximation auskommt, scheint schwer möglich. Wir führen dazu wie Kuznetsov (1995) das sogenannte bialternierende Produkt $A \odot B$ zweier Matrizen ein:

Definition 5.61 (Bialternierendes Produkt)

Das *bialternierende Produkt* $C = A \odot B$ zweier Matrizen $A = (a_{ik}) \in \mathbb{R}^{n \times n}$ und $B = (b_{ik}) \in \mathbb{R}^{n \times n}$ ist eine $m \times m$ -Matrix mit $m = n(n-1)/2$. Die Zeilen von C werden mit dem Multiindex (p, q) , $p = 2(1)n$, $q = 1(1)p - 1$, die Spalten mit dem Multiindex (r, s) , $r = 2(1)n$, $s = 1(1)r - 1$ bezeichnet. Die Matrixelemente berechnet man gemäß

$$C_{(p,q),(r,s)} := \frac{1}{2} \left\{ \begin{vmatrix} a_{pr} & a_{ps} \\ b_{qr} & b_{qs} \end{vmatrix} + \begin{vmatrix} b_{pr} & b_{ps} \\ a_{qr} & a_{qs} \end{vmatrix} \right\}. \tag{5.88}$$

Die Multiindizes (p, q) und (r, s) sind beispielsweise im Falle $n = 4$ folgendermaßen angeordnet:

$$\begin{matrix} (2, 1), (2, 1) & (2, 1)(3, 1) & \dots & (2, 1), (4, 3) \\ (3, 1), (2, 1) & (3, 1)(3, 1) & \dots & (3, 1), (4, 3) \\ (3, 2), (2, 1) & (3, 2)(3, 1) & \dots & (3, 2), (4, 3) \\ (4, 1), (2, 1) & (4, 1)(3, 1) & \dots & (4, 1), (4, 3) \\ (4, 2), (2, 1) & (4, 2)(3, 1) & \dots & (4, 2), (4, 3) \\ (4, 3), (2, 1) & (4, 3)(3, 1) & \dots & (4, 3), (4, 3). \end{matrix} \tag{5.89}$$

Die Bedeutung dieses ansonsten wenig bekannten Produktes ergibt sich aus folgendem

Satz 5.62

Die Matrix A besitze die Eigenwerte $\mu_1, \mu_2, \dots, \mu_n$. Dann hat die Matrix $A \odot A$ die Eigenwerte $\mu_i \mu_k$ und die Matrix $2A \odot I$ die Eigenwerte $\mu_i + \mu_k$, wobei $i = 2(1)n$, $k = 1(1)i - 1$, und I die Einheitsmatrix ist.

Setzen wir zur Bestimmung von $B := 2A \odot I$ die Einheitsmatrix $I = (\delta_{ik})$ in die Definition ein, so liefert (5.88) elementweise

$$B_{(p,q),(r,s)} = \begin{vmatrix} a_{pr} & a_{ps} \\ \delta_{qr} & \delta_{qs} \end{vmatrix} + \begin{vmatrix} \delta_{pr} & \delta_{ps} \\ a_{qr} & a_{qs} \end{vmatrix} = a_{pr} \delta_{qs} - a_{ps} \delta_{qr} + a_{qs} \delta_{pr} - a_{qr} \delta_{ps},$$

mit dem Kronecker-Symbol δ_{ik} , woraus sich die einfache Fallunterscheidung

$$B_{(p,q),(r,s)} = \begin{cases} -a_{ps} & \text{für } r = q \\ a_{pr} & \text{für } r \neq p \text{ und } s = q \\ a_{pp} + a_{qq} & \text{für } r = p \text{ und } s = q \\ a_{qs} & \text{für } r = p \text{ und } s \neq q \\ -a_{qr} & \text{für } s = p \\ 0 & \text{sonst} \end{cases} \quad (5.90)$$

ergibt. Wir können damit die Testfunktion (5.87) für Hopf-Bifurkation in der Darstellung

$$\psi_H(y) = \det(2D_x f(y) \odot I) \quad (5.91)$$

angeben, die das Produkt aller Eigenwerte von $2D_x f(y) \odot I$ liefert.

Beispiel 5.63

$$A = D_x f(y^*) = \begin{pmatrix} 2 & 2 & 7 \\ -5 & -5 & -7 \\ -2 & 5 & 0 \end{pmatrix} \quad \text{mit Eigenwerten } \mu_{1,2} = \pm 7i, \mu_3 = -3$$

sei exemplarisch eine Jacobi-Matrix am Bifurkationspunkt $y^* = (x^*, \lambda^*)$. Wegen $\operatorname{Re} \mu_{1,2} = 0$ und $\operatorname{Re} \mu_3 < 0$ kann hier eine Hopf-Bifurkation vorliegen. Andererseits kann keine der anderen genannten Bifurkationen eintreten, da $\psi_F(y^*) = \det A = -147 \neq 0$ ist. Wir berechnen mittels (5.90) die von A abgeleitete Matrix

$$B = 2A \odot I = \begin{pmatrix} -3 & -7 & -7 \\ 5 & 2 & 2 \\ 2 & -5 & -5 \end{pmatrix} \quad \text{mit Eigenwerten } \beta_{1,2} = -3 \pm 7i, \beta_3 = 0$$

und erhalten daraus den Wert der Testfunktion $\psi_H(y^*) = \det B = 0$. ■

Für große Dimension n der DGL ist der numerische Aufwand $T(n)$ zur Determinantenberechnung von $B = 2A \odot I$ allerdings beträchtlich. Nutzt man die LU-Zerlegung der $m \times m$ -Matrix B , so ergibt sich wegen $m = n(n-1)/2$ asymptotisch

$$T(n) \approx \frac{1}{3}m^3 \approx \frac{1}{3} \left(\frac{1}{2}n^2 \right)^3 \approx \frac{1}{24}n^6,$$

weshalb dann schnelle Approximationen des Spektrums der Jacobi-Matrix A geeigneter erscheinen. Algorithmen zur Berechnung derartiger *Ritz-Werte* und *Petrov-Werte* großdimensionaler Matrizen findet man u. a. bei Hoffmann et al. (2005) in Kapitel 20.

5.4 Aufgaben

Aufgabe 5.1

Betrachten Sie für den Parameter $s \in \mathbb{R}$ folgendes Differenzialgleichungssystem:

$$\begin{aligned}\dot{x}_1 &= s \cdot x_1 - x_2 - x_1 \cdot (x_1^2 + x_2^2) \\ \dot{x}_2 &= x_1 + s \cdot x_2 - x_2 \cdot (x_1^2 + x_2^2).\end{aligned}$$

- Zeichnen Sie ein Phasenportrait dieses Systems für $[-2, 2] \times [-2, 2]$.
- Bestimmen Sie sämtliche Gleichgewichtslösungen (Ruhelagen, stationäre Lösungen) des Differenzialgleichungssystems. Charakterisieren Sie diese Punkte nach Hyperbolizität und Senke-Sattel-Quelle-Eigenschaft.

Aufgabe 5.2

Bestimmen Sie sämtliche Gleichgewichtslösungen des Differenzialgleichungssystems

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = x_1 + x_1^2.$$

Charakterisieren Sie diese Lösungen nach Hyperbolizität und Senke-Sattel-Quelle-Eigenschaft.

Aufgabe 5.3

Untersuchen Sie sämtliche Gleichgewichtslösungen der Differenzialgleichung

$$\dot{x} = x \cdot (x^3 - x - \lambda)$$

mit Parameter $\lambda \in \mathbb{R}$ daraufhin, ob sie hyperbolisch, Sättel, Quellen oder Senken sind! Welcher Zusammenhang besteht zum Regularitätsbegriff?

Aufgabe 5.4

Bestimmen Sie sämtliche Lösungen der Gleichung $f(x, \lambda) = 0$ zu gegebenem Parameter $\lambda \in \mathbb{R}$, und entscheiden Sie, ob diese regulär oder singular sind. Dabei ist jeweils $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$.

- $f(x, \lambda) = \begin{pmatrix} x_1^2 - \lambda^2 \\ x_2^2 + \lambda^2 - 1 \end{pmatrix}$ für $\lambda = 0, \frac{1}{2}, 1$.
- $f(x, \lambda) = \begin{pmatrix} 2x_1^2 - 4 + \lambda \cdot (x_1 + x_2) \\ 2x_2^2 - 4 + \lambda \cdot (x_1 + x_2) \end{pmatrix}$ für beliebiges $\lambda \in \mathbb{R}$.

Aufgabe 5.5

Beweisen Sie, dass unter den Voraussetzungen des Satzes 5.24

- der klassische Prädiktor die Approximationsordnung 1,
- der Tangenten-Prädiktor die Approximationsordnung 2,
- der Sekanten-Prädiktor die Approximationsordnung 1 besitzt.

Aufgabe 5.6

Entwickeln Sie in Analogie zu Formel (5.27) einen Taylor-Prädiktor, der den Prädiktorpunkt x^P bei Kenntnis der 1. und 2. Ableitung $x'(\lambda_{j-1})$ und $x''(\lambda_{j-1})$ bei $\lambda = \lambda_{j-1}$ ermittelt. Geben Sie Berechnungsvorschriften für $x'(\lambda_{j-1})$ und $x''(\lambda_{j-1})$ an, und bestimmen Sie die Approximationsordnung dieser Prädiktorformel. Bewerten Sie den numerischen Aufwand dieser Formel!

Aufgabe 5.7

Entwickeln Sie eine MATLAB-Funktion `continuation` für Algorithmus 5.28 zum Prädiktor-Korrektor-Verfahren mit klassischem Prädiktor bei natürlicher Parametrisierung.

Aufgabe 5.8

Zur Bestimmung der Quadratwurzel $x = \sqrt{\lambda}$, $\lambda > 0$ soll das Prädiktor-Korrektor-Verfahren auf das Nullstellenproblem

$$f(x, \lambda) = x^2 - \lambda = 0, \quad a \leq \lambda \leq b,$$

angewendet werden.

- Notieren Sie die Formeln für den Tangenten-Prädiktor und den Newton-Korrektor.
- Starten Sie mit $\lambda_0 = a = 1$, $x_0 = 1$, und bestimmen Sie numerisch für $1 \leq \lambda \leq 5$ die Werte $x = \sqrt{\lambda}$ mit dem Fortsetzungsverfahren und den Schrittweiten $s = 1$, $s = 2$, $s = 4$.
- Vergleichen Sie die Effizienz der drei Rechnungen, und versuchen Sie, die getroffene Aussage zu verallgemeinern.

Aufgabe 5.9

- Ermitteln Sie alle nichtnegativen Gleichgewichtslagen $x = (x_1, x_2)^T$, d. h. mit $x_1, x_2 \geq 0$, des DGL-Systems

$$\dot{x}_1 = x_1 \cdot (s - x_1^2), \quad \dot{x}_2 = x_2 \cdot (\lambda - (x_2^2 - 1)^2)$$

mit festem reellen $s > 0$ und Parameter $\lambda \in [-2, 2]$.

- Geben Sie die regulären Lösungspfade \mathcal{L} an!
- Bestimmen Sie alle nicht-regulären Lösungen x , und prüfen Sie, ob ein Umkehrpunkt vorliegt!

Aufgabe 5.10

Welche sind die regulären Lösungspfade der Gleichgewichtslösungen von

$$\dot{x} = x \cdot (x^3 - x - \lambda), \quad \lambda \in [-1, 1] ?$$

Verifizieren Sie, welche Kurvenpunkte regulär und welche Umkehrpunkte sind.

Aufgabe 5.11

Untersuchen Sie die Gleichgewichtslage $y^0 = (x_1^0, x_2^0, \lambda_0) = (0, 0, 0)$ des Systems

$$\dot{x}_1 = x_1^p - \lambda, \quad \dot{x}_2 = x_2^q - \lambda \quad \text{mit } p, q \in \mathbb{N}$$

darauffin, ob sie regulär, singular bzw. Umkehrpunkt ist. Führen Sie dabei notwendige Fallunterscheidungen durch!

Aufgabe 5.12

Für die Pseudo-Bogenlängen-Parametrisierung der Kurve $y = y(s)$ nach Algorithmus 5.34 wird der normierte Tangentenvektor $y'(s_j)$ im Punkt s_j benötigt.

- Geben Sie eine Methode zur Berechnung von $y'(s_j)$ mit dem richtigen Vorzeichen an, die die LU-Zerlegung benutzt!
- Entwickeln Sie eine MATLAB-Funktion `pseudoarclength` zur Fortsetzung mit Pseudo-Bogenlängen-Parametrisierung gemäß Algorithmus 5.34 unter Benutzung des Tangenten-Prädiktors und des vereinfachten Newton-Korrektors.
- Erproben Sie diese Funktion anhand der Aufgabe 5.9.

Aufgabe 5.13

Das DGL-System von L. Chua wird im \mathbb{R}^3 durch

$$\begin{aligned} \dot{x} &= \alpha \cdot (y - x - f(x)) \\ \dot{y} &= x - y + z \\ \dot{z} &= -\beta \cdot y \end{aligned} \quad \text{mit } f(x) = \begin{cases} b \cdot x + a - b & \text{für } x \geq 1 \\ a \cdot x & \text{für } |x| < 1 \\ b \cdot x - a + b & \text{für } x \leq -1 \end{cases}$$

und Parametern $\alpha, \beta, a, b \in \mathbb{R} \setminus \{0\}, b \neq -1$, beschrieben.

- Bestimmen Sie alle drei Gleichgewichtslösungen im allgemeinen Fall!
- Im Folgenden seien die Werte $(\alpha, \beta, a, b) = (9, 14\frac{2}{7}, -\frac{8}{7}, -\frac{5}{7})$ gesetzt. Berechnen Sie damit die drei Gleichgewichtslösungen. Sind diese stabil? Bestimmen Sie die Eigenwerte μ_1, μ_2, μ_3 der Jacobi-Matrizen. Wie groß sind die Dimensionen der Unterräume E^s, E^u und E^c ?
- Ermitteln Sie die invarianten Unterräume E^s, E^u, E^c ! Skizzieren Sie im \mathbb{R}^3 die Ruhpunkte, die invarianten Unterräume und die trennenden Ebenen

$$U_1 = \{(x, y, z) \in \mathbb{R}^3 | x = 1\} \quad \text{und} \quad U_{-1} = \{(x, y, z) \in \mathbb{R}^3 | x = -1\}.$$

Wie verhalten sich die Orbits des Systems?

- Stellen Sie die Orbits – etwa mittels MAPLE – graphisch dar!

Aufgabe 5.14

Entwickeln Sie eine MATLAB-Funktion `gaussnewton` zur Gauß-Newton-Fortsetzung gemäß Algorithmus 5.38 unter Benutzung des Sekanten-Prädiktors und des (vereinfachten) Gauß-Newton-Korrektors.

Aufgabe 5.15

Das Differenzialgleichungssystem

$$\begin{aligned}\dot{x}_1 &= \varepsilon \cdot x_1 + x_1^2 \\ \dot{x}_2 &= x_2 \\ \dot{x}_3 &= -x_3\end{aligned}$$

besitzt für alle $\varepsilon \in \mathbb{R}$ die instabile Gleichgewichtslage $x^0 = 0$ (Begründung). Kann $\varepsilon_0 = 0$ dennoch ein Bifurkationspunkt sein? Entscheiden Sie gegebenenfalls, welcher Bifurkationstyp vorliegt.

Aufgabe 5.16

Die Dynamik eines einfachen Jäger-Beute-Ökosystems mit skalierten Populationszahlen $x_1 > 0$ und $x_2 > 0$ wird durch die Differenzialgleichungen

$$\begin{aligned}\dot{x}_1 &= x_1 \cdot (\lambda + x_1) \cdot (1 - x_1) - 3x_1 \cdot x_2 \\ \dot{x}_2 &= -2\lambda \cdot x_2 + x_1 \cdot x_2\end{aligned}$$

beschrieben, wobei $\lambda > 0$ ist.

- Welche Gleichgewichtslösungen besitzt das System? Sind sie stabil?
- Für welche Parameterwerte λ_0 wird das betrachtete Gleichgewicht instabil? Welche Bifurkationen treten auf?
- Für welche Parameter besitzt das System eine periodische Lösung? Ist sie stabil? (Begründung!)
- Skizzieren Sie die Phasenportraits für die verschiedenen Lösungen!

Aufgabe 5.17

Programmieren Sie in MATLAB eine Funktion $B = \text{balt}(A)$ zur Bestimmung des bialternierenden Produkts $B = 2A \odot I$ der Matrix $A \in \mathbb{R}^{n \times n}$.

- Überprüfen Sie anhand der Testmatrizen $A = \text{magic}(n)$, $n = 3, 4, 5, \dots$ aus MATLAB und selbst gewählter Matrizen A die Richtigkeit des Programms.
- Verifizieren Sie mit den Testmatrizen aus MATLAB

$$A = \text{magic}(n) - \text{magic}(n)^T, \quad n = 3, 4, \dots$$

und den Funktionen $\text{eig}(A)$, $\text{eig}(B)$, $\text{det}(B)$ die Aussagen des Satzes 5.62 über die Eigenwerte von B .

- Vergleichen Sie mit selbst gewählten n und Testmatrizen A den Rechenaufwand, gemessen in der Zahl der Gleitpunkt-Operationen (Flops), zur Bestimmung aller Eigenwerte über $\text{eig}(A)$ mit der Zahl der Flops bei $\text{det}(\text{balt}(A))$. Versuchen Sie eine theoretische Aussage zur Zeitkomplexität $T(n)$ für beide Varianten.

6 Numerik periodischer Lösungen

Übersicht

6.1	Periodisch erregte Systeme	289
6.2	Autonome Systeme	306
6.3	Die Poincaré-Abbildung	318
6.4	Lösungsfortsetzung und Bifurkationsanalyse.....	327
6.5	Aufgaben	358

In vielen Anwendungen der Natur- und Ingenieurwissenschaften treten Schwingungsprobleme mit periodischen Lösungen zeitkontinuierlicher dynamischer Systeme auf. Derartige Lösungen befinden sich nicht als Ruhelagen im Gleichgewicht, sondern schwingen mit einer konstanten Schwingungsdauer, der *Periode* T der Bewegung. So besitzt das autonome System von W. F. Langford aus Beispiel 5.1

$$\begin{aligned}\dot{x}_1 &= (x_3 - 0.7)x_1 - \omega x_2 \\ \dot{x}_2 &= \omega x_1 + (x_3 - 0.7)x_2 \\ \dot{x}_3 &= 0.6 + x_3 - x_3^3/3 - (x_1^2 + x_2^2)(1 + \rho x_3) + \varepsilon x_3 x_1^3\end{aligned}\tag{6.1}$$

einen periodischen Orbit, der auf einer toroidalen Fläche verläuft und in Abbildung 6.1 zu sehen ist.

Bezeichnet $\varphi_t : D \rightarrow D$ mit der offenen Menge $D \subset \mathbb{R}^n$ den Fluss eines zeitkontinuierlichen dynamischen Systems mit $t \in \mathbb{R}$, so ist gemäß Kapitel 2 eine Lösung $x(t; x_0) = \varphi_t(x_0)$ durch $x_0 \in D$ *periodisch*, wenn es eine Konstante $T > 0$ mit $\varphi_T(x_0) = x_0$ gibt. Existiert ein kleinstes derartiges $T > 0$, die Periode der Bewegung, so nennt man die Lösung T -periodisch. Ein periodischer Orbit

$$\gamma := \{x \in D \mid x = \varphi_t(x_0), t \in \mathbb{R} \text{ mit } \varphi_T(x_0) = x_0\}$$

sei ein *Grenzykel* gemäß Kapitel 3. Dann ist der Orbit γ isoliert, d. h. es existiert eine Umgebung $U(\gamma)$, die keine weiteren periodischen Orbits enthält.

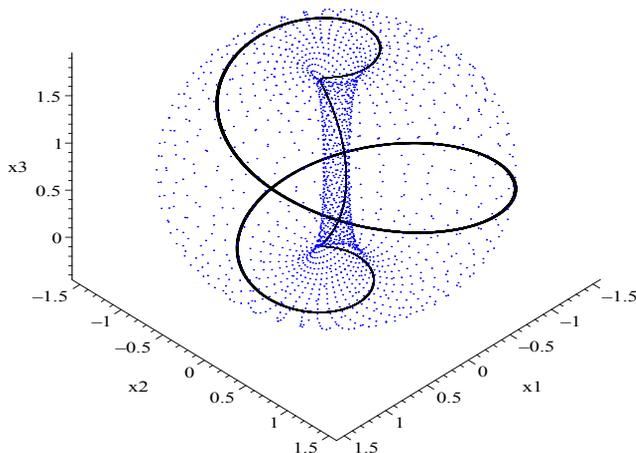


Abb. 6.1 Phasenraum-Darstellung eines periodischen Orbits und des Torus zu Beispiel 5.1. Parameterwerte sind $\omega = 3.5$, $\varepsilon = 0.03$ und $\varrho = 0.25$

Beispiel 6.1

Für feste Parameterwerte $\varepsilon > 0$ und $\omega > 0$ besitzt das nichtlineare autonome System

$$\dot{x}_1 = \varepsilon x_1 - \omega x_2 - x_1(x_1^2 + x_2^2), \quad \dot{x}_2 = \omega x_1 - \varepsilon x_2 - x_2(x_1^2 + x_2^2) \quad (6.2)$$

zum Anfangswert $x_0 = (\sqrt{\varepsilon}, 0)^T$ einen isolierten periodischen Orbit mit der Darstellung

$$\varphi_t(x_0) = \sqrt{\varepsilon}(\cos \omega t, \sin \omega t)^T,$$

wie man leicht nachrechnet. Dagegen hat bei vorgegebenem $\omega > 0$ das lineare System

$$\dot{x}_1 = -\omega x_2, \quad \dot{x}_2 = \omega x_1 \quad (6.3)$$

zu $x_0 = (c, 0)$ mit beliebig gewähltem $c \in \mathbb{R}$ stets einen periodischen Orbit $\varphi_t(x_0) = c(\cos \omega t, \sin \omega t)^T$, so dass keiner dieser Orbits isoliert ist. ■

Falls $\varphi_t(x)$ periodisch mit Periode $T > 0$ ist, so ist aufgrund der Translationsinvarianz zu jedem festen $\tau \in \mathbb{R}$ auch $\varphi_{t+\tau}(x)$ periodisch. Dann gilt

$$\varphi_{t+T}(x) = \varphi_t(x) \quad \forall t \in \mathbb{R}. \quad (6.4)$$

Die Phasenkurve γ eines Zyklus ist zudem geschlossen und einfach zusammenhängend (doppelpunktfrei), wie in Abbildung 6.1 dargestellt.

Bemerkung 6.2

Existenz- und Eindeutigkeitsaussagen über periodische Orbits allgemeiner n -dimensionaler nichtlinearer Systeme (5.1) mit numerisch leicht überprüfbareren Voraussetzungen sind nicht verfügbar. Andererseits schränken spezielle Zusatzannahmen die Problemklasse meist stark ein. Typische Sätze mit hinreichenden Bedingungen für ebene Systeme in \mathbb{R}^2 liefert die Poincaré-Bendixson-Theorie, die man u. a. bei Perko (1996) findet. Eine Verallgemeinerung auf Systeme in \mathbb{R}^n stellt der Satz von R. A. Smith dar (vgl. Reitmann (1996)), dessen Voraussetzungen jedoch im Allgemeinen schwer verifizierbar sind. In der Umgebung eines Hopf-Bifurkationspunktes garantiert der Satz 3.8 von E. Hopf ebenfalls die Existenz eines Grenzyklus. ■

Um die Problemklasse nicht von vornherein einzuschränken, wollen wir deshalb die Existenz eines Grenzzykels stets voraussetzen, sobald dies erforderlich wird.

6.1 Periodisch erregte Systeme

In diesem Abschnitt unterdrücken wir die Parameterabhängigkeit der DGL und suchen für einen festen Parameterwert eine periodische Lösung x^* . Zuerst betrachten wir periodisch angetriebene Schwingungssysteme, deren „Systemantwort“ auf die Erregungsschwingung von Interesse ist. Charakteristisch ist dabei, dass die Periode T des gesuchten Zyklus vorgegeben und damit bekannt ist! Numerische Verfahren sind damit einfacher aufgebaut als im autonomen Fall, den wir anschließend behandeln wollen.

Wir notieren periodisch erregte DGL-Systeme in der Form

$$\dot{x} = f(x, t), \quad f : D \times \mathbb{R} \rightarrow \mathbb{R}^n \quad \text{mit } D \subset \mathbb{R}^n, \quad D \text{ offen,} \quad (6.5)$$

und setzen die Existenz eines Grenzzykels sowie die hinreichende Glattheit des in t periodischen Vektorfeldes f voraus, d. h. alle partiellen Ableitungen bis zur r -ten Ableitung existieren und sind stetig:

Voraussetzung 6.3 (Glattheit, Periodizität und Lösungsexistenz)

- i. $f \in C^r(D \times \mathbb{R})$, $r \geq 2$.
- ii. Es existiert ein $T > 0$, so dass gilt: $f(x, t + T) = f(x, t) \quad \forall (x, t) \in D \times \mathbb{R}$.
- iii. System (6.5) besitzt einen Grenzzykel $x^*(t)$ der Periode T .

Wegen Voraussetzung i. ist offenbar $x^* \in C^{r+1}(\mathbb{R})$. Zudem genügt es nach Voraussetzung ii., die Aufgabe auf einem Intervall $I = [t_0 + T, t_0]$ zu betrachten, wobei stets $t_0 = 0$ gesetzt werden kann. Denn ist $x^*(t)$ für $t \in [0, T]$ bekannt, so kann die Lösung für alle $t \in \mathbb{R}$ periodisch fortgesetzt werden. Betrachten wir drei typische praxisnahe Beispiele:

Beispiel 6.4

1. Die *Duffing-Van-der-Pol-Gleichung* (vgl. Abraham und Shaw (1989), Lynch (2001)) mit $\alpha, \beta, \delta, \lambda, \omega > 0$

$$\ddot{x} + \delta \dot{x} - \beta x + \alpha x^3 = \lambda \cdot \cos(\omega t)$$

beschreibt ein periodisch angetriebenes Pendel mit kubischer Rückstellkraft. Mit der Auslenkung $x_1 = x$ und der Geschwindigkeit der Masse $x_2 = \dot{x}$ ergibt sich das System mit der Erregungsperiode $T_0 = \frac{2\pi}{\omega}$

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = \beta x_1 - \alpha x_1^3 - \delta x_2 + \lambda \cdot \cos(\omega t).$$

Gesucht ist eine sogenannte *harmonische Lösung* derselben Periode $T = T_0$.

2. Die Modellgleichung 2. Ordnung vom Mathieschen Typ (vgl. Philippow und Bütting (1992))

$$\ddot{x} + \alpha \dot{x}^3 - \beta \dot{x} + (1 - \delta \cdot \sin 2t) \cdot x = 0$$

mit $\alpha, \beta, \delta > 0$ und Erregungsperiode $T_0 = \pi$ beschreibt ein *parametrisch erregtes elektrisches Netzwerk*. Außer der stets existierenden trivialen Nulllösung sind sogenannte *2-fach subharmonische Lösungen* mit Periode $T = 2T_0 = 2\pi$ zu ermitteln.

3. E. Philippow und W. Büntig entwickelten in der Monografie Philippow und Büntig (1992) auch Modellgleichungen für subharmonisch reagierende elektrische Netzwerke, die bei harmonischer Erregung eine exakt k -fach subharmonische Schwingung (Strom bzw. Spannung) erzeugen. So sind für

$$\ddot{x} - \varepsilon(1 - x^2 - \dot{x}^2)\dot{x} + x + b(4x^3 - 3x) = B \cdot \cos 3t$$

mit $\varepsilon, b, B > 0$ und Erregungsperiode $T_0 = \frac{2\pi}{3}$ periodische Lösungen mit $T = 3T_0 = 2\pi$ gesucht, die als *3-fach subharmonisch* bezeichnet werden. Im Allgemeinen ist $T = k \cdot T_0$, $k \in \mathbb{N}$, $k > 1$, für jede k -fach subharmonische Lösung. ■

Mit der vorgegebenen Lösungsperiode T ist eine auf dem Periodizitätsintervall $I = [0, T]$ stetig differenzierbare Funktion $x : I \rightarrow \mathbb{R}^n$ gesucht, die dem *Zweipunkt-Randwertproblem*

$$\dot{x} = f(x, t), \quad x(T) = x(0), \quad 0 < t < T, \quad (6.6)$$

genügt. Dafür können die bekannten numerischen Lösungsverfahren wie *Anfangswertverfahren (Schießverfahren)*, *Finite-Differenzen-Verfahren* oder *Kollokationsverfahren* eingesetzt werden. Während die beiden letztgenannten Verfahren unmittelbar eine Diskretisierung der Lösung vornehmen und ursprünglich für statische Randwertprobleme, z. B. Biegeprobleme, entwickelt wurden, trägt das Schießverfahren (auch: ballistische Methode) dem dynamischen Charakter des Problems mit t als Zeitkoordinate Rechnung. Wir wollen deshalb hier nur den ersten Zugang vorstellen und verweisen auf die umfangreiche algorithmische Darstellung der anderen Verfahren in Hoffmann et al. (2006).

6.1.1 Einfaches Schießverfahren

Wir betrachten zu gegebenem Problem (6.6) das assoziierte Anfangswertproblem (AWP)

$$\dot{x} = f(x, t), \quad x(0) = s \quad \text{mit} \quad s \in D \subset \mathbb{R}^n, \quad t \in I = [0, T] \quad (6.7)$$

und nehmen vorerst an, dass (6.7) zu jedem $s \in D$ eine auf I eindeutige Lösung $x(t, s)$ besitzt. Die geforderte Periodizitätsbedingung $x(T, s) = x(0, s)$ wird damit im Allgemeinen nicht zu erfüllen sein. Definieren wir deshalb die Differenzfunktion $g : D \rightarrow D$ mit

$$g(s) := x(T, s) - s, \quad (6.8)$$

so löst $x(t, s^*)$ mit $s^* \in D$ genau dann das Randwertproblem (6.6), wenn der n -dimensionale Vektor $s = s^*$ eine Lösung des Gleichungssystems

$$g(s) = 0, \quad s \in D \subset \mathbb{R}^n \quad (6.9)$$

ist. Damit wird das Schwingungsproblem für eine reelle Vektorfunktion formal auf ein endlich dimensionales Nullstellenproblem in \mathbb{R}^n reduziert, wofür leistungsfähige Verfahren verfügbar sind (vgl. Hoffmann et al. (2005), Kapitel 21).

Um den gesamten Integrationsaufwand für das Randwertproblem niedrig zu halten, setzen wir ein überlinear konvergentes Verfahren zur Gleichungslösung ein. Das Newton-Verfahren für das Gleichungssystem (6.9) lautet in der Standardform

$$s_{k+1} = s_k - [g'(s_k)]^{-1}g(s_k), \quad k = 0, 1, 2, \dots, \quad s_0 \in D \quad (6.10)$$

mit der Jacobi-Matrix (desweiteren als „Newton-Matrix“ bezeichnet)

$$g'(s) = \frac{\partial x}{\partial s}(T, s) - I_n, \quad I_n - n\text{-reihige Einheitsmatrix.}$$

Die aufwändige Invertierung dieser Matrix vermeiden wir durch Umstellung der Verfahrensgleichung (6.10) nach der Newton-Korrektur $d_k = s_{k+1} - s_k$, womit die praktikablere Form des Newton-Verfahrens entsteht:

Iteriere für $k = 0, 1, 2, \dots$, beginnend mit s_0 :

(1) Berechne die Newton-Matrix $g'(s_k) = \frac{\partial x}{\partial s}(T, s_k) - I_n$.

(2) Löse das lineare Gleichungssystem nach d_k

$$g'(s_k) d_k = -g(s_k). \quad (6.11)$$

(3) Korrigiere $s_{k+1} = s_k + d_k$.

Um hieraus einen Algorithmus zu entwickeln, sollten wir zuerst folgende Fragen klären:

1. Wie kann die $(n \times n)$ -Matrixfunktion $\frac{\partial x}{\partial s}(T, s)$ ermittelt werden?
2. Sind alle dabei auftretenden Anfangswertprobleme auf ganz $I = [0, T]$ lösbar?
3. Unter welchen zusätzlichen Voraussetzungen (an das Ausgangsproblem) kann die lokale Konvergenz des Newton-Verfahrens garantiert werden?

Beantworten wir die erste Frage, indem wir eine Lösung $x(t, s)$ des AWP (6.7) in die DGL einsetzen

$$\frac{\partial x}{\partial t}(t, s) = f(x(t, s), t), \quad x(0, s) = s.$$

Differenzieren wir diese Gleichungen nach den Anfangswerten s , so erhalten wir mit der Glattheitsvoraussetzung ein AWP für die Matrixfunktion $X(t, s) = \frac{\partial x}{\partial s}(t, s)$

$$\frac{\partial}{\partial t} \left(\frac{\partial x}{\partial s}(t, s) \right) = \frac{\partial f}{\partial x}(x(t, s), t) \cdot \frac{\partial x}{\partial s}(t, s), \quad \frac{\partial x}{\partial s}(0, s) = I_n. \quad (6.12)$$

Zur Vereinfachung der Notation führen wir nun folgende Begriffe ein:

Definition 6.5 (Variationssystem)

$x(t, s)$ sei die Lösung des AWP (6.6) mit $s \in D$. Die Matrixfunktion $A : I \times D \rightarrow \mathbb{R}^{n \times n}$ mit

$$A(t, s) := \frac{\partial f}{\partial x}(x(t, s), t) = D_x f(x(t, s), t)$$

heißt zugehörige *Jacobi-Matrix*. Die Matrix-DGL für $X : I \times D \rightarrow \mathbb{R}^{n \times n}$ mit

$$\dot{X} = A(t, s) \cdot X, \quad \dot{X} = \frac{\partial}{\partial t} X(t, s) \quad (6.13)$$

heißt zugehöriges *Variationssystem*. Die Matrixfunktion $X : I \times D \rightarrow \mathbb{R}^{n \times n}$ mit

$$\dot{X} = A(t, s) \cdot X, \quad X(0, s) = I_n \quad (I_n \text{ Einheitsmatrix}) \quad (6.14)$$

ist die bei $t = 0$ *normierte Fundamentalmatrix (Hauptfundamentalmatrix)*.

Wegen der Darstellung (6.12) für die gesuchte Matrix gilt offenbar

$$X(t, s) \equiv \frac{\partial x}{\partial s}(t, s), \quad (t, s) \in I \times D,$$

womit sich die gesuchte Ableitung $g'(s_k)$ des Newton-Verfahrens zu

$$g'(s_k) = X(T, s_k) - I_n \quad (6.15)$$

ergibt. Die Hauptfundamentalmatrix $X(t, s)$ kann also durch eine simultane Integration des linearen Variationssystems (6.14) gewonnen werden, womit sich am Intervallende die Newton-Matrix $g'(s_k)$ ergibt. Der Algorithmus 6.6 des *einfachen Schießverfahrens (simple shooting)* liefert zu vorgegebener Startnäherung $s_0 \in \mathbb{R}^n$ und Toleranz tol in Schritt 1 eine Lösung s^* des Gleichungssystems. Mit dem Startvektor $x^*(0) = s^*$ gewinnen wir durch Integration des AWP in Schritt 2 anschließend die gesuchte periodische Lösung $x^*(t) = x(t, s^*)$.

Um die Fragen 2 und 3 zu beantworten, erinnern wir uns daran, dass ein Gleichungssystem $g(s) = 0$ mit dem Newton-Verfahren lokal eindeutig lösbar ist, wenn die Lösung s^* regulär ist (vgl. Hoffmann et al. (2005), Kap.21). Den entsprechenden Begriff für die Lösung $x^*(t)$ des Randwertproblems liefert

Definition 6.7 (Reguläre Lösung)

Voraussetzung 6.3 sei erfüllt. Die Lösung $x^*(t)$ heißt *regulär* (auch: *isoliert*), wenn das assoziierte linear-homogene Randwertproblem

$$\dot{z} = D_x f(x^*(t), t)z, \quad z(T) = z(0) \quad (6.16)$$

zu Randwertproblem (6.6) nur die triviale Lösung $z(t) \equiv 0$ besitzt.

ALGORITHMUS 6.6 (Einfaches Schießverfahren)Function $[s^*] = \text{shooting}(f, D_x f, T, s_0, \text{tol}, \text{kmax})$ 1. Für $k = 0, 1, \dots, \text{kmax}$ iteriere:1.1. (Anfangswertprobleme) Löse die $n + 1$ Systeme für $t \in [0, T]$:

$$\begin{aligned} \dot{x} &= f(x, t) & , & \quad x(0, s_k) = s_k \\ \dot{X} &= D_x f(x(t, s_k), t) \cdot X, & \quad X(0, s_k) = I_n. \end{aligned}$$

Ergebnis: $x(T, s_k) \in \mathbb{R}^n$, $X(T, s_k) \in \mathbb{R}^{n \times n}$

1.2. (Nullstellenaufgabe) Bilde die Funktionen

$$\begin{aligned} g(s_k) &:= x(T, s_k) - s_k \\ g'(s_k) &:= X(T, s_k) - I_n. \end{aligned}$$

1.3. (Newton-Schritt) Löse das lineare System

$$g'(s_k) \cdot d_k = -g(s_k).$$

1.4. Falls $\|d_k\| < \text{tol} \cdot (1 + \|g(s_k)\|)$, so gehe zu Schritt 2.1.5. (Newton-Korrektur) Aktualisiere $s_{k+1} := s_k + d_k$.2. Setze $s^* := s_k$ und löse das AWP

$$\dot{x} = f(x, t), \quad x(0) = s^*, \quad t \in [0, T].$$

3. Return s^*

Die Eindeutigkeit der Null-Lösung z bedeutet, dass die Linearisierung nicht singular ist, womit numerische Verfahren wie das Newton-Verfahren anwendbar werden. Jede reguläre Lösung $x^*(t)$ ist insbesondere lokal eindeutig (auch: geometrisch isoliert), d. h. es existiert eine Konstante $r > 0$, so dass in einer Umgebung $\mathcal{U}[x^*; r] := \{x \in C(I) \mid \sup_{t \in I} \|x(t) - x^*(t)\| \leq r\}$ um die Lösung keine von $x^*(t)$ verschiedene Lösung des Randwertproblems existiert. Für derartige Lösungen findet man bei Keller (1976) den

Satz 6.8

Die reguläre Lösung $x^*(t)$ erfülle die Voraussetzungen 6.3 mit $f \in \mathcal{C}^2(I \times D)$ und der Lipschitz-Konstanten $L := \sup_{I \times D} |D_x f(t, x)|$. Dann existiert eine Umgebung

$$\mathcal{K}[s^*; \varrho] = \{s \in D \mid \|s - s^*\| \leq \varrho\}, \quad \text{mit} \quad \varrho := r e^{-LT}$$

des Anfangswertes $s^* = x^*(0)$ mit folgenden Eigenschaften:

- i. Das AWP (6.7) besitzt zu jedem $s \in \mathcal{K}[s^*; \varrho]$ eine eindeutige Lösung $x(t, s) \in \mathcal{C}^1(I)$.
- ii. Die bei $t = 0$ normierte Fundamentalmatrix $X(t, s)$ existiert als Lösung von (6.14) und genügt der Identität

$$X(t, s) \equiv \frac{\partial x}{\partial s}(t, s), \quad (t, s) \in I \times \mathcal{K}[s^*; \varrho].$$

iii. $x^*(t) = x(t, s^*)$ ist genau dann eine reguläre Lösung des Randwertproblems (6.6), wenn $s^* = x^*(0)$ reguläre Lösung des Gleichungssystems $g(s) = 0$ ist.

Dieser Satz garantiert die Durchführbarkeit des Schießverfahrens 6.6, wenn nur die Startnäherung s_0 hinreichend nahe der Lösung s^* gewählt wird. Aussage iii. sichert die lokale quadratische Konvergenz der Startwerte $s_k \rightarrow s^*$ des Newton-Verfahrens auf Grund des Satzes 5.8.

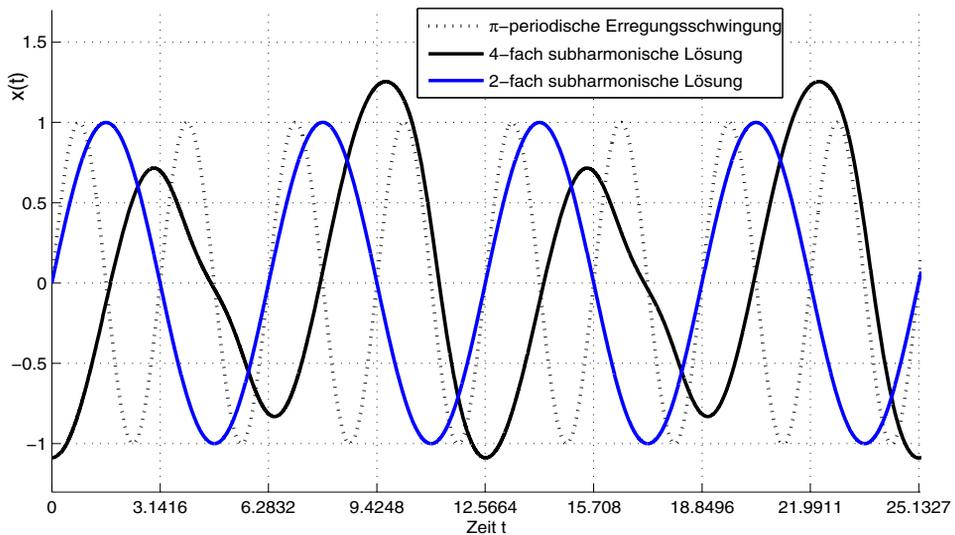


Abb. 6.2 System (6.17) – Erregungsschwingung sowie 2-fach und 4-fach subharmonische Systemantworten

Beispiel 6.9

Das folgende 2-fach subharmonisch reagierende System beschreibt ein elektrisches Netzwerk, das bei harmonischer Erregung als Systemantwort eine subharmonische Schwingung erzeugt. Die Modellgleichung von E. Philippow und W. Büntig (vgl. Philippow und Büntig (1992)) lautet

$$\ddot{x} - \varepsilon(1 - x^2 - \dot{x}^2)\dot{x} + 2bx\dot{x} + x = \hat{B} \sin 2t$$

mit den reellen Parametern $b = \hat{B} = 1$ sowie dem freien Parameter $\varepsilon \in \mathbb{R}$. Transformation in ein System erster Ordnung liefert das nichtautonome Schwingungsproblem

$$\begin{aligned} \dot{x}_1 &= x_2, & x_1(0) - x_1(T) &= 0 \\ \dot{x}_2 &= \varepsilon(1 - x_1^2 - x_2^2)x_2 - 2bx_1x_2 - x_1 + \hat{B} \sin 2t, & x_2(0) - x_2(T) &= 0 \end{aligned} \quad (6.17)$$

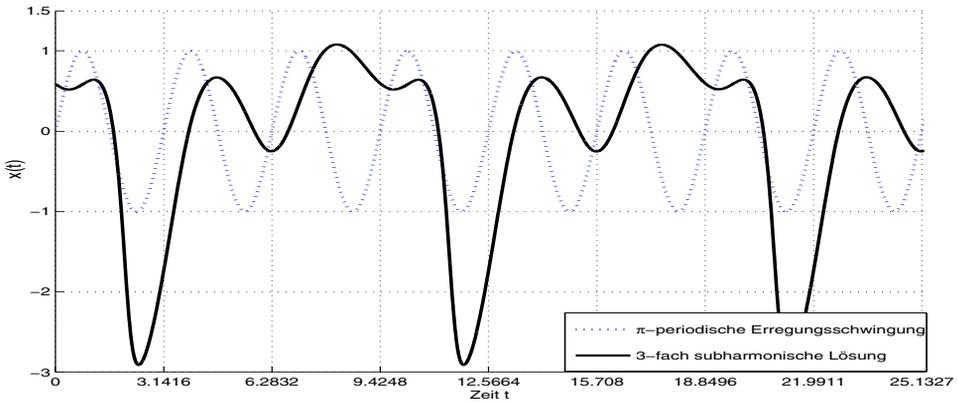


Abb. 6.3 System (6.17) – Erregungsschwingung sowie 3-fach subharmonische Systemantwort

mit Erregungsperiode $T_0 = \pi$ und vorgegebener Lösungsperiode $T = 2\pi$. Untersucht man das Lösungsverhalten für den festen Parameterwert $\varepsilon_0 = 0.4$, so findet man mit Algorithmus 6.6 leicht einen periodischen Orbit der gewünschten Periodendauer 2π . Darüber

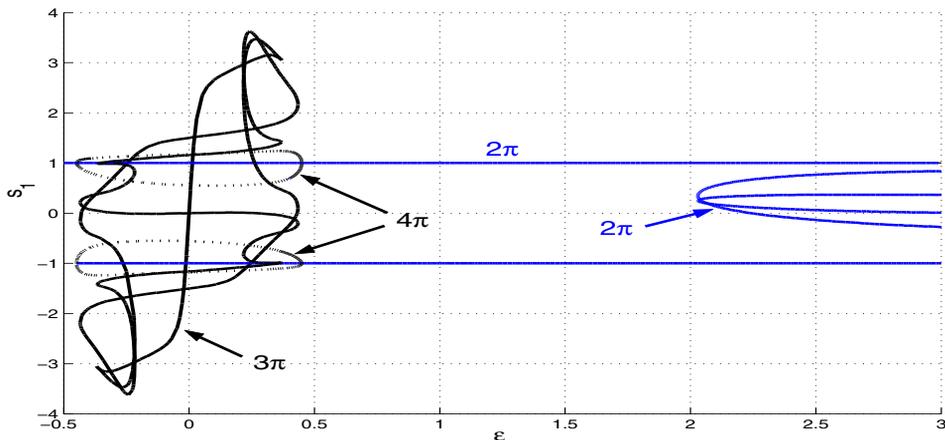


Abb. 6.4 (ε, s_1) -Diagramm der periodischen Lösungen des Systems (6.17)

hinaus existieren aber auch 3-fach bzw. 4-fach subharmonische Lösungen. Die Abbildungen 6.2 und 6.3 zeigen diese Schwingungen. Betrachten wir die 3π -periodische Lösung, deren Anfangswert $s_1 = x_1(0)$ in Abhängigkeit vom Parameter ε in Abbildung 6.4 durch eine geschlossene schwarze Kurve dargestellt ist. Der Versuch einer Lösungsfortsetzung nach dem Kontrollparameter ε , ausgehend von $\varepsilon = 0.36$ und dem Startpunkt $s_0 = x(0) = (0.38697629, 0.09266292)^T$, scheitert mit dem Schießverfahren schon nach wenigen Fortsetzungsschritten. Grund hierfür ist offenbar ein „stark instabiles“ Verhalten der periodischen Lösungen, das zum Versagen jedes numerischen Integrationsverfahrens bereits vor Erreichen des Intervallendes $T = 3\pi$ führt. ■

6.1.2 Stabilitätsanalyse

Betrachten wir Lösungen des soeben behandelten subharmonisch reagierenden Netzwerkes mit Parametern $b = 1$, $\varepsilon = 3$ und $\hat{B} = 2$ in der Umgebung zweier periodischer Lösungen Γ_1 und Γ_2 , so stellen wir unterschiedliches Verhalten fest:

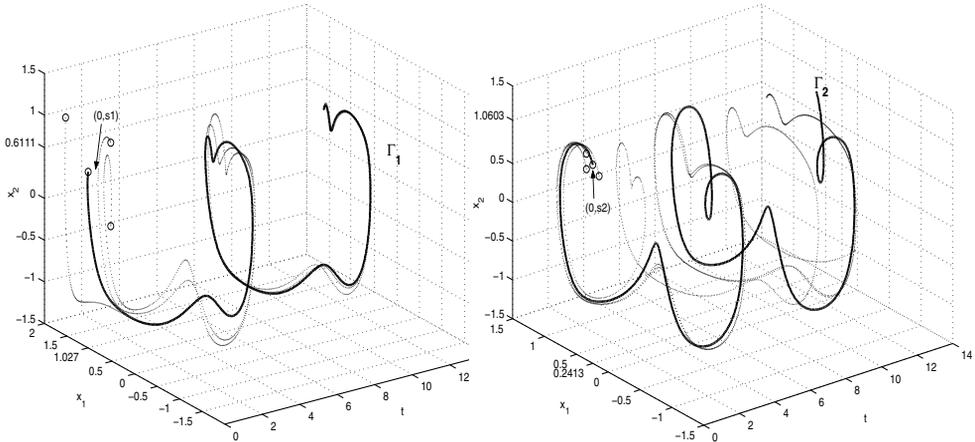


Abb. 6.5 Lösungen in der Umgebung der stabilen Lösung Γ_1 (links) und der instabilen Lösung Γ_2 (rechts)

Während die Lösung Γ_1 im erweiterten Phasenraum \mathbb{R}^3 nahe gelegene Lösungen „anzieht“, wirkt Γ_2 „abstoßend“ (vgl. Abb. 6.5). Wir wenden deshalb den Begriff der Ljapunov-Stabilität (vgl. Hoffmann et al. (2006)) nun auf periodische Orbits an:

Definition 6.10 (Ljapunov-Stabilität periodischer Orbits)

- i. $\{\varphi_t(s^*)\}$, $t \geq 0$, heißt (Ljapunov-)stabil, wenn zu $\varepsilon > 0$ ein $\delta > 0$ existiert, so dass für alle $s \in D$ mit $\|s - s^*\| < \delta$ stets $\|\varphi_t(s) - \varphi_t(s^*)\| < \varepsilon \quad \forall t \geq 0$ gilt.
- ii. Ist $\{\varphi_t(s^*)\}$ (Ljapunov-)stabil und es existiert ein $\delta_1 > 0$, so dass für alle $s \in D$ mit $\|s - s^*\| < \delta_1$ stets $\lim_{t \rightarrow \infty} \|\varphi_t(s) - \varphi_t(s^*)\| = 0$ folgt, so ist $\{\varphi_t(s^*)\}$ asymptotisch (Ljapunov-)stabil.
- iii. Ist $\{\varphi_t(s^*)\}$ nicht (Ljapunov-)stabil, so heißt es (Ljapunov-)instabil.

Auch für periodische Orbits soll nun ein algebraisches Stabilitätskriterium entwickelt werden. Voraussetzung 6.3 sei erfüllt und die Folge der Anfangsvektoren $\{s_k\}$, $k = 0, 1, 2, \dots$, in Algorithmus 6.6 konvergiere gegen eine Nullstelle s^* von $g(s) := x(T, s) - s$. Wegen (6.7) ist dann mit $x(0, s^*) = s^*$ die Funktion $x(t, s^*)$ eine gesuchte T -periodische Lösung. Wir betrachten dazu eine benachbarte Lösung $x(t, s)$ des AWP

$$\dot{x} = f(x, t), \quad x(0, s) = s, \quad s \in \mathcal{K}[s^*; \varrho] \tag{6.18}$$

mit der nach Satz 6.8 garantierten Umgebung $\mathcal{K}[s^*; \varrho]$. Für die Differenzfunktion $z(t) := x(t, s) - x(t, s^*)$ gewinnen wir durch Taylor-Entwicklung

$$\dot{z}(t) = \dot{x}(t, s) - \dot{x}(t, s^*) = f(x(t, s), t) - f(x(t, s^*), t) = \underbrace{D_x f(x(t, s^*), t)}_{A(t, s^*)} \cdot z(t) + \mathcal{O}(\|z(t)\|^2),$$

so dass diese Funktion in 1. Näherung für $\|z(t)\| \ll 1$ dem Variationssystem

$$\dot{z} = A(t, s^*)z, \quad z(0) = z^0 = s - s^* \quad (6.19)$$

genügt. Die zugehörige Jacobi-Matrix $A(t, s^*) = D_x f(x(t, s^*), t)$ ist T -periodisch in t . Denn

$$\begin{aligned} A(t+T, s^*) &= D_x f(x(t+T, s^*), t+T) \\ &= D_x f(x(t, s^*), t+T), \quad \text{da } x(t, s^*) \text{ periodisch ist} \\ &= D_x f(x(t, s^*), t) = A(t, s^*) \quad \forall t \in \mathbb{R}, \end{aligned}$$

da nach Voraussetzung 6.3 auch $D_x f(x, t)$ periodisch in t ist. Die Fundamentalmatrix $X(t, s^*)$ ist mit $X(0, s^*) = I$ nach dem Satz von Liouville regulär für alle $t \in \mathbb{R}$. Deshalb bilden ihre Spaltenvektoren eine Basis, womit sich die Störungsfunktion $z(t)$ als deren Linearkombination

$$\begin{aligned} z(t) &= X(t, s^*)c, \quad c \in \mathbb{R}^n \quad \text{mit } z(0) = X(0, s^*)c = c = z^0 \\ &= X(t, s^*)z^0 = X(t, s^*)(s - s^*) \end{aligned} \quad (6.20)$$

darstellen lässt. Damit ist das Verhalten der mit $s = s^*$ gebildete Matrixfunktion $X(t, s^*)$, die auch *Hauptlösungsmatrix* genannt wird, für $t \rightarrow \infty$ maßgeblich. Die *Floquet-Theorie* liefert eine interessante Darstellung für diese Funktion. In Pontrjagin (1965) findet man den vollständigen Beweis des Satzes von Floquet.¹

Satz 6.11 (Floquet)

Wenn $A(t, s^*)$ eine stetige T -periodische Matrixfunktion ist, so kann die Hauptlösungsmatrix $X(t, s^*)$ in der Floquet-Darstellung

$$X(t, s^*) = G(t) \cdot e^{tR}, \quad t \in \mathbb{R} \quad (6.21)$$

notiert werden. Dabei ist $G(t)$ eine reguläre stetig differenzierbare T -periodische Matrixfunktion und $R \in \mathbb{R}^{n \times n}$ eine konstante Matrix.

Für $t = 0$ ergibt (6.21) die Einheitsmatrix $G(0) = I$, während die T -Periodizität von $G(t)$ die konstante Matrix $M = X(T, s^*) = G(T) \cdot e^{TR} = e^{TR}$ liefert:

¹Gaston Floquet (1847–1920), französischer Mathematiker, Arbeiten zur Theorie gewöhnlicher Differenzialgleichungen.

Definition 6.12 (Monodromiematrix, Floquet-Multiplikatoren)

Die bei $t = T$ gebildete konstante $n \times n$ -Matrix

$$M := X(T, s^*) = \left. \frac{\partial x}{\partial s}(t, s) \right|_{t=T, s=s^*} = e^{TR} \quad (6.22)$$

heißt *Monodromiematrix*. Ihre Eigenwerte m_i , $i = 1(1)n$, sind die *Floquet-Multiplikatoren* (auch: *charakteristische Multiplikatoren*) von $x(t, s^*)$.

Worin liegt nun die Bedeutung dieser Matrix? Wegen der T -Periodizität von $G(t)$ erhält man sofort

$$X(t+T, s^*) = G(t+T)e^{(t+T)R} = G(t)e^{tR}e^{TR} = X(t, s^*) \cdot M \quad \forall t \in \mathbb{R}.$$

Ersetzt man t durch $t+T$, so ergibt sich wiederum

$$X(t+2T, s^*) = X(t+T, s^*) \cdot M = X(t, s^*) \cdot M^2,$$

woraus induktiv

$$X(t+kT, s^*) = X(t, s^*) \cdot M^k, \quad k = 0, 1, 2, \dots \quad \text{und} \quad X(kT, s^*) = M^k \quad (6.23)$$

folgt. Die Monodromiematrix zur k -fachen Periode $T_k = k \cdot T$ ist also gleich der k -ten Potenz der Monodromiematrix M . Die Störungsfunktion (6.20) erhält damit die Form

$$z(t+kT) = X(t+kT, s^*) \cdot (s - s^*) = X(t, s^*) M^k (s - s^*) \quad (6.24)$$

bzw. mit $t = 0$

$$z(kT) = M^k (s - s^*), \quad k = 0, 1, \dots \quad (6.25)$$

$z(kT)$ entsteht also durch eine Vektoriteration mit M , angewandt auf die Anfangsstörung $s - s^*$. Falls $t \neq 0$ in Formel (6.24) ist, so erfolgt eine nachträgliche Korrektur von $z(kT)$ mittels der Matrix $X(t, s^*)$. Über das Verhalten von $z(kT)$ für $k \rightarrow \infty$ entscheiden wie bei der allgemeinen Vektoriteration $z_k = M z_{k-1}$, $k = 1, 2, 3, \dots$, zur Approximation von Eigenwerten einer Matrix M (vgl. dazu Hoffmann et al. (2005), Kapitel 20), die betragsgrößten Eigenwerte von M . Es gilt

Lemma 6.13

Unter der Voraussetzung 6.3 seien m_i die Floquet-Multiplikatoren zu $x(t, s^*)$.

- i. Ist $|m_i| \neq 1$ für alle $i = 1(1)n$, so heißt $x(t, s^*)$ *hyperbolisch*.
- ii. Ist $|m_i| < 1$ für alle $i = 1(1)n$, so ist $x(t, s^*)$ *asymptotisch Ljapunov-stabil*.
- iii. Ist $|m_i| > 1$ für ein i , so ist $x(t, s^*)$ *Ljapunov-instabil*.

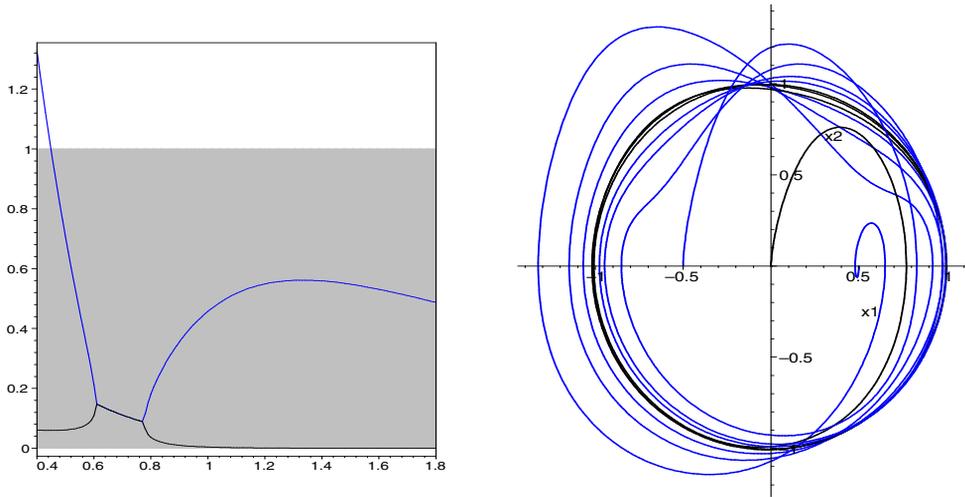


Abb. 6.6 Floquet-Multiplikatoren (links) und Phasenkurven (rechts) zu Beispiel 6.14

Beispiel 6.14

Das subharmonisch reagierende System aus Beispiel 6.4 mit $b = \hat{B} = 1$

$$\begin{aligned} \dot{x}_1 &= x_2, & x_1(0) - x_1(T) &= 0 \\ \dot{x}_2 &= \varepsilon(1 - x_1^2 - x_2^2)x_2 - 2bx_1x_2 - x_1 + \hat{B} \sin 2t, & x_2(0) - x_2(T) &= 0 \end{aligned} \tag{6.26}$$

besitzt für jeden reellen Parameterwert ε die Lösung $x_1(t, s^*) = \sin t, x_2(t, s^*) = \cos t$. Durch partielle Ableitung und Einsetzen der Lösung mit $s^* = (0, 1)^T$ berechnen wir die Jacobi-Matrix

$$A(t, s^*) = D_x f(x(t, s^*), t) = \begin{pmatrix} 0 & 1 \\ -2\varepsilon \sin t \cos t - 2 \cos t - 1 & -2\varepsilon \cos^2 t - 2 \sin t \end{pmatrix}$$

und gewinnen so das Variationssystem

$$\frac{d}{dt} \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} = A(t, s^*) \cdot \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}, \quad \begin{pmatrix} x_{11}(0) & x_{12}(0) \\ x_{21}(0) & x_{22}(0) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

das allerdings nicht analytisch lösbar ist. Numerisch erhalten wir zum Parameterwert $\varepsilon = 1.4$ mittels eines Runge-Kutta-Verfahrens – die Werte wurden mit dem Standardverfahren *dsolve* von MAPLE mit Genauigkeit von 10^{-7} ermittelt – die Hauptlösungsmatrix $X(t, s^*)$, womit sich bei $t = 2\pi$ die Monodromiematrix

$$M = X(2\pi, s^*) = \begin{pmatrix} 0.526710116094 & -0.0733313105237 \\ -0.232313402738 & 0.0326311502789 \end{pmatrix}$$

ergibt. Die Floquet-Multiplikatoren berechnen wir anschließend zu $m_1 = 0.5590706$ und $m_2 = 0.0002706468$, womit Satz 6.13 die asymptotische Stabilität der Lösung garantiert. In Abbildung 6.6 (rechts) ist das Konvergenzverhalten dreier Phasenkurven gegen die Lösung $x(t, s^*) = (\sin t, \cos t)$ dargestellt.

Wiederholung dieser Rechnung für weitere Parameterwerte $\varepsilon \in [0.4, 1.8]$ liefert ein interessantes Stabilitätsverhalten dieser Lösung. In Abbildung 6.6 (links) sind die Absolutbeträge $|m_1|$ (blau) und $|m_2|$ (schwarz) der Floquet-Multiplikatoren dargestellt. Offenbar ist die Lösung für $\varepsilon < 0.45$ instabil und ansonsten asymptotisch stabil. Im Intervall $[0.6, 0.8]$ werden die Multiplikatoren konjugiert komplex, womit aber kein Stabilitätsverlust verbunden ist. Mit der genauen Bestimmung des kritischen Parameterwertes $\varepsilon_0 \approx 0.45$ werden wir uns in Abschnitt 6.4.3 beschäftigen. ■

Mittels der Floquet-Multiplikatoren lässt sich die *Regularität einer periodischen Lösung* $x(t, s^*)$ wesentlich leichter nachweisen als mit der unhandlichen Definition 6.7. Denn mit der Defektfunktion $g(s) := x(T, s) - s$ lautet die Newton-Matrix $g'(s) = X(T, s) - I_n$ des Schießverfahrens (6.6) bzw. an der Lösungsstelle $s = s^*$

$$g'(s^*) = X(T, s^*) - I_n = M - I_n. \quad (6.27)$$

Damit zeigt man leicht das folgende algebraische Regularitäts-Kriterium:

Lemma 6.15

Neben Voraussetzung 6.3 mit $f \in \mathcal{C}^2(I \times D)$ seien alle n Floquet-Multiplikatoren m_i von $x(t, s^)$ verschieden von 1. Dann ist $s^* = x(0, s^*)$ eine reguläre Lösung von $g(s) = 0$, und das Newton-Verfahren konvergiert lokal quadratisch gegen s^* .*

Beweis: Wegen (6.27) besitzt $g'(s^*)$ die Eigenwerte $\mu_1 = m_1 - 1$, $\mu_2 = m_2 - 1$, \dots , $\mu_n = m_n - 1$. Da $m_i \neq 1$ gilt, sind alle Eigenwerte von $g'(s^*)$ ungleich 0. Somit ist $g'(s^*)$ eine reguläre Matrix und s^* eine reguläre Lösung von $g(s) = 0$. Mit dem allgemeinen Konvergenzsatz 5.8 über das Newton-Verfahren folgt die Konvergenzbehauptung. □

Wegen $\lim_{k \rightarrow \infty} s_k = s^*$ gemäß Lemma liefert das Newton-Verfahren nach K Schritten in der Regel eine gute Näherung $g'(s_K) \approx g'(s^*)$, so dass die Monodromiematrix

$$M = g'(s^*) + I_n \approx g'(s_K) + I_n \quad (6.28)$$

mit beliebig hoher Genauigkeit approximiert werden kann.

Häufig steht die Jacobi-Matrix $D_x f(x, t)$ jedoch nicht explizit zur Verfügung. Dann vermeiden wir das Variationssystem vollkommen und ersetzen $g'(s_k)$ durch geeignete Differenzenquotienten, indem der Startwert s_k des k -ten Newtonschrittes mittels eines Vektors $\delta = (\delta_1, \delta_2, \dots, \delta_n)^T$, $\delta_j > 0$, „gestört“ wird. Mit den j -ten Einheitsvektoren e_j definieren wir die gestörten Startvektoren

$$\sigma_0 := s_k, \quad \sigma_j := s_k + \delta_j \cdot e_j, \quad j = 1(1)n. \quad (6.29)$$

Die Schritte 1.1 und 1.2 von Algorithmus 6.6 verändern sich damit und liefern das praktikable Schießverfahren 6.16 ohne Nutzung des Variationssystems. Eine erprobte Strategie wählt zu $s_k = (s_{k1}, s_{k2}, \dots, s_{kn})^T$ die Schrittweiten im k -ten Iterationsschritt zu

$$\delta_j = \sqrt{\varepsilon_M}(1 + |s_{kj}|), \quad j = 1(1)n, \quad \text{mit Maschinengenauigkeit } \varepsilon_M.$$

ALGORITHMUS 6.16 (Schießverfahren ohne Variationssystem)Function $[s^*, m] = \text{approximate_shooting}(f, T, s_0, \text{tol}, kmax)$ 1. Für $k = 0, 1, \dots, kmax$ iteriere:1.1. Wähle $\delta_j = \sqrt{\varepsilon_M}(1 + |s_{kj}|)$, $j = 1(1)n$.1.2. (Anfangswertprobleme) Löse mit $\sigma_0 := s_k$ und $\sigma_j := s_k + \delta_j \cdot e_j$, $j = 1(1)n$, die $n + 1$ Systeme für $t \in [0, T]$:

$$\dot{x} = f(x, t), \quad x(0, \sigma_j) = \sigma_j, \quad j = 0(1)n.$$

Ergebnis: $x(T, \sigma_j) \in \mathbb{R}^n$, $j = 0(1)n$

1.3. (Nullstellenaufgabe) Bilde die Funktionen

$$g(\sigma_j) := x(T, \sigma_j) - \sigma_j, \quad j = 0(1)n \quad \text{und damit}$$

$$g(s_k) := g(\sigma_0) \quad \text{sowie die Differenzenquotienten}$$

$$g_j := \frac{1}{\delta_j} [g(\sigma_j) - g(\sigma_0)], \quad j = 1(1)n,$$

$$g'(s_k) := (g_1, g_2, \dots, g_n).$$

1.4. (Newton-Schritt) Löse das lineare System

$$g'(s_k) \cdot d_k = -g(s_k).$$

1.5. Falls $\|d_k\| < \text{tol} \cdot (1 + \|g(s_k)\|)$, so gehe zu Schritt 2.1.6. (Newton-Korrektur) Aktualisiere $s_{k+1} := s_k + d_k$.2. (Periodische Lösung) Setze $s^* := s_k$ und löse das AWP

$$\dot{x} = f(x, t), \quad x(0) = s^*, \quad t \in [0, T].$$

3. (Monodromiematrix) Setze $M := g'(s_k) + I_n$ und bestimme die Eigenwerte $m = (m_1, m_2, \dots, m_n)$.4. Return s^*, m

Anders als der Lösungsvektor s^* werden jetzt die Monodromiematrix M und somit die Floquet-Multiplikatoren nur mit einer maximalen Genauigkeit von $\sqrt{\varepsilon_M}$ approximiert.

6.1.3 Mehrfach-Schießverfahren

Die Durchführbarkeit der einfachen Schießverfahren 6.6 und 6.16 ist oft nur in unmittelbarer Lösungsnähe gesichert, denn die darin vorkommenden AWP müssen auf dem gesamten Periodizitätsintervall $I = [0, T]$ lösbar sein. Der Konvergenzbereich des Schießverfahrens kann deshalb bei großer Lipschitz-Konstante L und langem Intervall I wegen der Kon-

stanten $\varrho = re^{-LT}$ aus Satz 6.8 extrem klein werden. Im *Mehrfach-Schießverfahren* (*Mehrzielmethode, multiple shooting, parallel shooting*) versuchen wir, diesem Problem mit einer Segmentierung des Lösungsintervalls zu begegnen. In $I = [0, T]$ legen wir so genannte Schießpunkte t_k mit der Eigenschaft

$$0 = t_0 < t_1 < t_2 < \dots < t_m = T$$

fest, die auf $m \geq 1$ Teilintervalle $I_k = [t_k, t_{k+1}]$ führen. $x(t) = x(t; t_k, s_k)$ bezeichne die auf dem k -ten Intervall I_k definierte Lösung des AWP

$$\dot{x} = f(x, t), \quad x(t_k) = s_k, \quad t_k \leq t \leq t_{k+1}, \quad k = 0(1)m - 1. \tag{6.30}$$

Abbildung 6.7 zeigt derartige Lösungen zu vorgegebenen Anfangswerten s_k . An den inne-

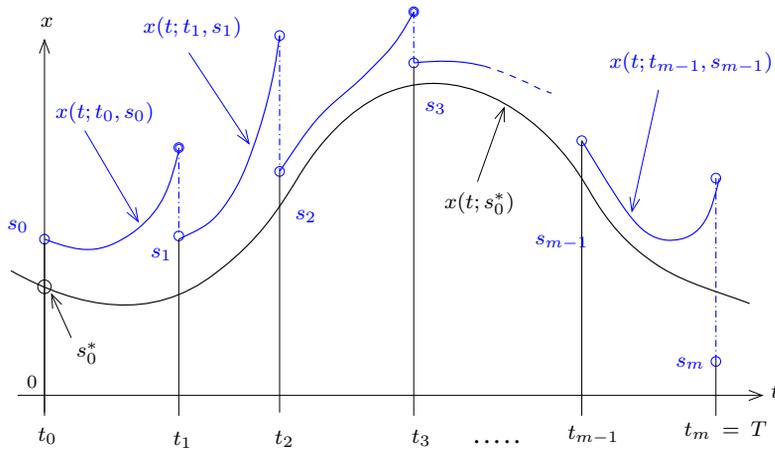


Abb. 6.7 Stückweise Lösung mit der Mehrzielmethode

ren Schießpunkten ergeben sich $m - 1$ Stetigkeitsbedingungen (Matching-Bedingungen) an die gesuchte Lösung mit den unbekanntenen „Schießwerten“ s_0, s_1, \dots, s_{m-1} , ergänzt um die Periodizitätsbedingungen zu

$$\begin{aligned} x(t_k; t_{k-1}, s_{k-1}) &= s_k, & k &= 1(1)m - 1, \\ x(t_m; t_{m-1}, s_{m-1}) &= s_0. \end{aligned} \tag{6.31}$$

Das daraus entstehende große Gleichungssystem für die $n \cdot m$ unbekanntenen Zahlenwerte $s = (s_0, s_1, \dots, s_{m-1})^T \in \mathbb{R}^{n \times m}$ erhält dann folgende Struktur:

$$F(s) := \begin{pmatrix} F_0(s_0, s_1) \\ F_1(s_1, s_2) \\ \vdots \\ F_{m-2}(s_{m-2}, s_{m-1}) \\ F_{m-1}(s_{m-1}, s_0) \end{pmatrix} = \begin{pmatrix} x(t_1; t_0, s_0) - s_1 \\ x(t_2; t_1, s_1) - s_2 \\ \vdots \\ x(t_{m-1}; t_{m-2}, s_{m-2}) - s_{m-1} \\ x(t_m; t_{m-1}, s_{m-1}) - s_0 \end{pmatrix} = 0. \tag{6.32}$$

Das Einfach-Schießverfahren entspricht genau dem Sonderfall $m = 1$. Die theoretischen Grundlagen des Mehrfach-Schießverfahrens findet man dargestellt in den Monografien von Hermann (2004) und Keller (1976). Wir fassen die Aussagen zusammen in

Satz 6.17

Die (hinreichend glatte) Lösung $x^*(t)$ nach Voraussetzung 6.3 sei regulär. Mit $s^* = (s_0^*, s_1^*, \dots, s_{m-1}^*)^T$ gilt:

- i. s^* ist Lösung von $F(s) = 0$ genau dann, wenn $x^*(t) = x(t; t_k, s_k^*)$, $k = 0(1)m - 1$, Lösung des Randwertproblems (6.6) ist.
- ii. Eine Lösung s^* des Gleichungssystems (6.32) ist genau dann regulär, wenn die Lösung $x^*(t)$ des Randwertproblems (6.6) regulär ist.
- iii. Gleichung (6.32) kann mit dem Newton-Verfahren gelöst werden, das Q-quadratisch konvergiert, falls die Startnäherung $s^{(0)} = (s_0^{(0)}, s_1^{(0)}, \dots, s_{m-1}^{(0)})^T$ hinreichend nahe bei s^* liegt.

Um das großdimensionale System $F(s) = 0$ zu lösen, wenden wir das Newton-Verfahren

$$s^{(\nu+1)} = s^{(\nu)} - F'(s^{(\nu)})^{-1}F(s^{(\nu)}), \quad \nu = 0, 1, 2, \dots \quad (6.33)$$

auf F an. Darin bezeichnet $s^{(\nu)}$ die ν -te Newton-Näherung, wogegen s_k die k -te Komponente von s ist. Die Jacobi-Matrix hat wegen (6.32) die zyklisch bidiagonale Blockstruktur

$$F'(s) = \begin{pmatrix} G_0 & -I_n & 0 & \cdots & 0 \\ 0 & G_1 & -I_n & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & G_{m-2} & -I_n \\ -I_n & 0 & \cdots & 0 & G_{m-1} \end{pmatrix} \quad \text{mit} \quad G_k := \frac{\partial x}{\partial s_k}(t_{k+1}; t_k, s_k),$$

wobei $I_n \in \mathbb{R}^{n \times n}$ die Einheitsmatrix bezeichnet. Die $n \times n$ -Matrizen G_k können wir wie beim Einfach-Schießen durch Integration des Variationssystems auf dem Intervall $I_k = [t_k, t_{k+1}]$

$$\dot{X} = A(t; t_k, s_k)X, \quad X(t_k; t_k, s_k) = I_n, \quad k = 0(1)m - 1 \quad (6.34)$$

bestimmen. Die zugehörigen Jacobi-Matrizen A und die Matrizen G_k sind durch

$$A(t; t_k, s_k) := D_x f(x(t; t_k, s_k), t) \quad \text{und} \quad G_k = X(t_{k+1}; t_k, s_k) = \frac{\partial x}{\partial s_k}(t_{k+1}; t_k, s_k)$$

definiert. Mit diesen Erweiterungen lässt sich aus den Grundalgorithmen 6.6 und 6.16 das *Mehrfach-Schießverfahren* 6.18 konstruieren. Was ist bei dessen Anwendung zu beachten?

ALGORITHMUS 6.18 (Mehrfach-Schießverfahren)

Function $[s^*, \mu] = \text{multiple_shooting}(f, D_x f, T, s_0, m, tol, \nu max)$

1. Wähle die Schießpunkte $t_0 = 0 < t_1 < \dots < t_m = T$ und Startnäherungen $s = (s_0, s_1, \dots, s_{m-1})^T$.
2. Für $\nu = 0(1)\nu max$ iteriere:
 - 2.1. (Anfangswertprobleme) Für $k = 0(1)m - 1$ wiederhole:
 - Löse auf $I_k = [t_k, t_{k+1}]$ die $n + 1$ Systeme (6.30) und (6.34).
Ergebnisse: $x(t_{k+1}; t_k, s_k) \in \mathbb{R}^n$, $G_k := X(t_{k+1}; t_k, s_k) \in \mathbb{R}^{n \times n}$
 - 2.2. (Nullstellenaufgabe)
 - Bilde den Vektor $F(s)$ nach (6.32) und die Matrix $F'(s)$.
 - 2.3. (Newton-Schritt) Löse das lineare System

$$F'(s) \cdot d = -F(s).$$
 - 2.4. Falls $\|d\| < tol \cdot (1 + \|F(s)\|)$, so gehe zu Schritt 3.
 - 2.5. (Newton-Korrektur) Aktualisiere $s := s + d$.
3. (Lösung) Setze $s^* := s$ und löse die m AWP

$$\dot{x} = f(t, x), \quad x(t_k) = s_k, \quad t_k \leq t \leq t_{k+1}, \quad k = 0(1)m - 1.$$
4. (Monodromiematrix) Setze $M := G_{m-1} \cdot G_{m-2} \cdots G_1 \cdot G_0$ und bestimme die Eigenwerte $\mu = (\mu_1, \mu_2, \dots, \mu_n)$.
5. Return s^*, μ

1. Zu gegebener Startnäherung $s_0 \in \mathbb{R}^n$ und Toleranz tol müssen die Schießpunkte t_k und zugehörigen Anfangswerte s_k so gewählt werden, dass die AWP des Schrittes 2.1 in den Intervallen $I_k = [t_k, t_{k+1}]$ eindeutig lösbar sind. Satz 6.8 garantiert dies für alle $s_k \in \mathcal{S}[s_k^*; \varrho_k]$ mit den Radien

$$\varrho_k := r e^{-L_k(t_{k+1}-t_k)} \quad \text{und Lipschitz-Konstanten} \quad L_k := \sup_{I_k \times D} |D_x f(t, x)|.$$

Die Durchführbarkeit des Mehrfach-Schießverfahrens ist somit bei hinreichend feiner Zerlegung von I theoretisch gesichert.

2. Problematisch bleibt die Bestimmung einer Startlösung s_0 und der Schießpunkte t_k . Mittels geeigneter Parameterfortsetzung (vgl. dazu Kapitel 5.2) lassen sich in komplizierten Anwendungen die Schießpunkte schrittweise anpassen. Da der arithmetische Aufwand mit zunehmender Gitterfeinheit rasch ansteigt, sollten möglichst wenige Schießpunkte verwendet werden. Algorithmusschritt 2.1 zeigt übrigens, dass der numerische Integrationsaufwand gegenüber dem Einfach-Schießverfahren nicht zunimmt. Der Mehraufwand resultiert lediglich aus den entstehenden größeren linearen Gleichungssystemen. Diese

sollten mittels LU-Zerlegung mit Pivotisierung gelöst werden, wobei die schwache (sparse) Besetzungsstruktur der Jacobi-Matrix bei der Speicherung zu berücksichtigen ist.

3. Mit der ermittelten Lösung $s^* = (s_0^*, s_1^*, \dots, s_m^*)^T$ und zugehörigem Lösungsvektor $x^*(t) := x(t; t_k, s_k^*)$, $t \in I_k$, $k = 0(1)m - 1$, ergeben sich die Hauptlösungsmatrizen $X(t; t_k, s_k^*)$, $k = 0(1)m - 1$, in den m Teilintervallen I_k . Man überlegt sich leicht, dass für beliebiges $t \geq t_k$ die Hauptlösungsmatrix die Produktdarstellung

$$X(t; t_0, s_0^*) = X(t; t_k, s_k^*) \cdot X(t_k; t_{k-1}, s_{k-1}^*) \cdots X(t_2; t_1, s_1^*) \cdot X(t_1; t_0, s_0^*)$$

besitzt. Für die gesuchte Monodromiematrix $M = X(T; 0, s_0^*)$ erhalten wir folglich:

$$M = G_{m-1}^* G_{m-2}^* \cdots G_1^* G_0^* \quad \text{mit} \quad G_k^* := X(t_{k+1}; t_k, s_k^*). \tag{6.35}$$

In Algorithmus 6.18 wird das Variationssystem zur genauen Berechnung der Floquet-Multiplikatoren benutzt. Die Variante mittels Differenzenapproximation aus Algorithmus 6.16 dürfte jedoch insgesamt einfacher umzusetzen sein.

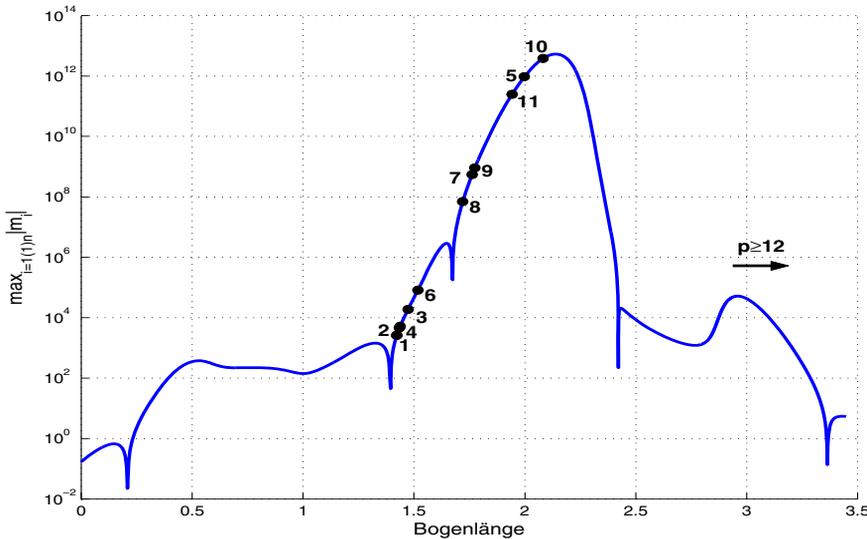


Abb. 6.8 Betragsgößter Multiplikator in Abhängigkeit von der Bogenlänge

Beispiel 6.19

Kommen wir auf das subharmonisch reagierende System aus Beispiel 6.9 mit 3π -periodischer Lösung zurück. Abbildung 6.4 zeigt die 1. Komponente s_1^* dieser Lösung in Abhängigkeit vom Parameter ε , die durch eine geschlossene schwarze Kurve dargestellt ist. Beginnend mit $\varepsilon = 0.36$ und einem Startpunkt $s_0 = x(0) = (0.38697629, 0.09266292)^T$ versagen die Einfach-Schießverfahren 6.6 bzw. 6.16 schon nach wenigen Fortsetzungsschritten für ε mit einem Floquet-Multiplikator der Größenordnung 10^4 . Abbildung 6.8 stellt die maximalen Multiplikatoren der Lösungskurve in Abhängigkeit von deren Bogenlänge logarithmisch skaliert dar. Setzen wir das Mehrfach-Schießverfahren 6.18 mit

m äquidistant gewählten Schießpunkten $t_k = kT/m$ ein, so gelingt die Fortsetzung der Lösungskurve auch bei „starker“ Instabilität. In Abbildung 6.8 sind die Abbruchpunkte der Fortsetzung mit der Anzahl m gekennzeichnet. Erst mit 12 Schießpunkten wird die Fortsetzung in diesem stark instabilen Bereich mit $\max_{i=1(1)n} |m_i| \approx 10^{13}$ gemeistert! Auffällig ist, dass man mit nur fünf Schießpunkten die Instabilität beinahe überwindet, aber erst die Verwendung der doppelten Anzahl besser abzuschneiden vermag. Eine automatische Schießpunktanpassung der Lösungsfortsetzung im instabilen Bereich ist deshalb ratsam. ■

Bemerkung 6.20

Umfangreichere Darstellungen zu diesen und weiteren Verfahrensklassen, wie Finite-Differenzenverfahren und Kollokationsverfahren, findet man in der Spezialliteratur Ascher et al. (1995), Ascher und Petzold (1998), Hermann (2004). Hat man eine gute Lösungsnaheherung $s_k^* \approx x^*(t_k)$, $k = 0(1)m - 1$, auf einem hinreichend feinem Gitter $\{t_k\}$ mit einem dieser Verfahren ermittelt, so lässt sich die Monodromiematrix M stets mittels der Darstellung (6.35)

$$M = G_{m-1}^* G_{m-2}^* \cdots G_1^* G_0^* \quad \text{mit} \quad G_k^* := X(t_{k+1}; t_k, s_k^*)$$

bestimmen. Dazu sind auf den Teilintervallen I_k die Hauptlösungsmatrizen $X(t_{k+1}; t_k, s_k^*)$ durch Integration von (6.34) zu berechnen. ■

6.2 Autonome Systeme

Die numerische Berechnung periodischer Lösungen autonomer DGL-Systeme

$$\dot{x} = f(x), \quad f : D \rightarrow \mathbb{R}^n \quad \text{mit} \quad D \subset \mathbb{R}^n, \quad D \text{ offen}, \quad (6.36)$$

ist wegen der häufig unbekanntenen Lösungsperiode T^* komplizierter als im gerade behandelten periodisch erregten Fall. Wir treffen für diese Problemklasse folgende grundlegende

Voraussetzung 6.21 (Glattheit und Lösungsexistenz)

- i. $f \in C^r(D)$, $r \geq 2$.
- ii. System (6.36) besitzt einen Grenzykel $x^*(t)$ der Periode $T^* > 0$, der keine Gleichgewichtslage ist.

6.2.1 Einfaches Schießverfahren

Anders als im periodisch erregten Fall des vorhergehenden Abschnittes sind nun eine T -periodische Lösung $x \in C^1(D)$ und ihre Periodendauer $T > 0$ gesucht, die auf dem Intervall $I = [0, T]$ dem *Zweipunkt-Randwertproblem*

$$\dot{x} = f(x), \quad x(T) = x(0), \quad 0 < t < T, \quad (6.37)$$

genügen. Diese Aufgabe lässt sich mittels der Zeittransformation $t = \frac{T}{2\pi}\tau$ mit $\tau \in [0, 2\pi]$ wegen $\frac{dx}{d\tau} = \frac{T}{2\pi}\dot{x}(t) = \frac{T}{2\pi}f(x(\tau))$ auf das Einheitsintervall transformieren und in der Form

$$\frac{dx}{d\tau} = \frac{T}{2\pi}f(x), \quad x(2\pi) = x(0), \quad 0 < \tau < 2\pi \quad (6.38)$$

numerisch leichter bearbeiten, da die unbekanntete Konstante T nun lediglich in der DGL auftritt. Wir wollen diese Darstellung erst im nächsten Abschnitt nutzen und vorerst bei der untransformierten Aufgabe (6.37) bleiben. Wegen der Translationsinvarianz der Lösung autonomer DGL (vgl. Abschnitt 2.6) kann jede Lösung $x(t; x_0) = \varphi_t(x_0)$, $t \in \mathbb{R}$, von (6.37) um eine beliebige reelle Phase t_0 verschoben werden.

Beispiel 6.22

Das autonome System in \mathbb{R}^2 mit reellen Parametern $\lambda > 0, \omega > 0$

$$\dot{x}_1 = \lambda x_1 - \omega x_2 - x_1(x_1^2 + x_2^2), \quad \dot{x}_2 = \omega x_1 + \lambda x_2 - x_2(x_1^2 + x_2^2) \quad (6.39)$$

besitzt mit $x_0 = (\sqrt{\lambda}, 0)^T$ zu jedem $\lambda > 0, \omega > 0$ die T -periodische Lösung

$$\varphi_t(x_0) = \sqrt{\lambda} \begin{pmatrix} \cos \omega t \\ \sin \omega t \end{pmatrix},$$

die einen Grenzykel bildet. Offenbar ist mit jedem $t_0 \in \mathbb{R}$ dann auch

$$\psi_t(x_0) = \varphi_{t+t_0}(x_0) = \sqrt{\lambda} \begin{pmatrix} \cos \omega(t+t_0) \\ \sin \omega(t+t_0) \end{pmatrix}$$

T -periodische Lösung der Periode $T = 2\pi/\omega$. ■

Wir gehen vorerst wie im periodisch erregten Fall vor und transformieren das Randwertproblem (6.37) in eine Nullstellenaufgabe in \mathbb{R}^n . Zu vorgegebenen Parametern $s \in \mathbb{R}^n$ und $T > 0$ betrachten wir das AWP

$$\dot{x} = f(x), \quad x(0; s, T) = s \in \mathbb{R}^n, \quad t \in [0, T]. \quad (6.40)$$

Nach Satz 6.8 existiert zu jedem (s, T) in einer Umgebung von (s^*, T^*) mit $s^* = x^*(0)$ eine eindeutige Lösung $x = x(t; s, T)$, $t \in [0, T]$ von (6.40). Definieren wir nun die Abbildung $g: D \times \mathbb{R}^+ \rightarrow D$ mit $D \subset \mathbb{R}^n$ durch

$$g(s, T) := x(T; s, T) - s \quad (6.41)$$

und betrachten die Nullstellenaufgabe

$$g(s, T) = 0. \quad (6.42)$$

Unter Voraussetzung 6.21 ist $x(t; s^*, T^*)$ genau dann eine periodische Lösung mit Periode T^* , wenn (s^*, T^*) dieses Gleichungssystem (6.42) löst. Wie im periodisch erregten Fall definieren wir zu einer Lösung $x(t; s, T)$ des AWP (6.40) das Variationssystem

$$\dot{X} = A(t; s, T)X, \quad X(0; s, T) = I_n \quad \text{mit} \quad A(t; s, T) := Df(x(t; s, T)). \quad (6.43)$$

Mit der Lösung (s^*, T^*) der Gleichung (6.42) und $x(t; s^*, t^*)$ erhalten wir daraus ebenfalls die Hauptlösungsmatrix und die Monodromiematrix

$$X(t; s^*, T^*) = \frac{\partial x}{\partial s}(t; s^*, T^*), \quad M = X(T^*; s^*, T^*) = \frac{\partial x}{\partial s}(T^*; s^*, T^*). \quad (6.44)$$

Lemma 6.23

Unter Voraussetzung 6.21 ist $m_1 = 1$ stets ein Eigenwert der Monodromiematrix M mit zugehörigem Eigenvektor $v(0) = \dot{x}(0; s^*, T^*) = f(s^*)$.

Beweis: Einsetzen der periodischen Lösung $x(t; s^*, T^*)$ und Differentiation nach t liefert

$$\frac{d}{dt} \dot{x}(t; s^*, T^*) = Df(x(t; s^*, T^*)) \cdot \frac{d}{dt} x(t; s^*, T^*) \quad \text{bzw.} \quad \dot{v} = A(t; s^*, T^*) \cdot v$$

mit der Jacobi-Matrix $A(t; s^*, T^*) := Df(x(t; s^*, T^*))$ und der Ableitung $v(t) := \dot{x}(t; s^*, T^*)$, d. h. v genügt dem Variationssystem. Die Hauptlösungsmatrix $X(t; s^*, T^*)$ bildet jedoch ein Fundamentalsystem von Lösungen, so dass $v(t)$ als Linearkombination ihrer (linear unabhängigen) Spaltenvektoren notiert werden kann:

$$v(t) = \dot{x}(t; s^*, T^*) = X(t; s^*, T^*) \cdot c, \quad c \in \mathbb{R}^n.$$

Wegen $v(0) = X(0; s^*, T^*) \cdot c = c$ ergibt sich $v(T^*) = X(T^*; s^*, T^*) \cdot v(0) = M \cdot v(0)$, woraus wegen der T^* -Periodizität von $x(t; s^*, T^*)$ und damit auch von $v(t)$ die Eigenwertbeziehung

$$M \cdot v(0) = m_1 \cdot v(0) \quad \text{mit} \quad m_1 = 1$$

folgt. $v(0) = \dot{x}(0; s^*, T^*) = f(s^*) \neq 0$ ist Eigenvektor von M zu Eigenwert $m_1 = 1$, da $x(0; s^*, T^*) = s^*$ nach Voraussetzung kein Gleichgewicht darstellt. \square

Ist $m_1 = 1$ ein geometrisch einfacher Eigenwert der Monodromiematrix M , so bezeichnet man $x(t, s^*)$ als *reguläre Lösung*. Das trifft insbesondere auf *hyperbolische Lösungen* zu, für die $|m_i| \neq 1$ für $i = 2, 3, \dots, n$ gefordert wird.

Beispiel 6.24

Für das DGL-System (6.39) erhalten wir mit der dort angegebenen periodischen Lösung $x(t; s^*, T^*) = (\sqrt{\lambda} \cos \omega t, \sqrt{\lambda} \sin \omega t)^T$ die periodische Jacobi-Matrix

$$A(t; s^*, T^*) = Df(x(t; s^*, T^*)) = \begin{pmatrix} -2\lambda \cos^2 \omega t & -\omega - \lambda \sin 2\omega t \\ \omega - \lambda \sin 2\omega t & -2\lambda \sin^2 \omega t \end{pmatrix}.$$

Man überzeugt sich leicht durch Einsetzen in das Variationssystem, dass die Hauptlösungsmatrix die Form

$$X(t; s^*, T^*) = \begin{pmatrix} e^{-2\lambda t} \cos \omega t & -\sin \omega t \\ e^{-2\lambda t} \sin \omega t & \cos \omega t \end{pmatrix} = \begin{pmatrix} \cos \omega t & -\sin \omega t \\ \sin \omega t & \cos \omega t \end{pmatrix} \begin{pmatrix} e^{-2\lambda t} & 0 \\ 0 & 1 \end{pmatrix},$$

besitzt, also die Floquet-Gestalt $X(t; s^*, T^*) = G(t) \cdot e^{tR}$ mit den Matrizen

$$G(t) = \begin{pmatrix} \cos \omega t & -\sin \omega t \\ \sin \omega t & \cos \omega t \end{pmatrix} \quad \text{und} \quad R = \begin{pmatrix} -2\lambda & 0 \\ 0 & 0 \end{pmatrix}.$$

Die Monodromiematrix $M = \text{diag}(e^{-4\lambda\pi/\omega}, 1)$ liefert die beiden Floquet-Multiplikatoren $m_1 = 1$, $m_2 = e^{-4\lambda\pi/\omega} \in (0, 1)$, womit die Lösung für alle Werte $\lambda > 0$ hyperbolisch ist. ■

6.2.2 Phasenbedingungen

Wenden wir uns nun der Lösung der Nullstellenaufgabe $g(s, T) = 0$ zu. Mit der partiellen Ableitung $D_s g(s, T)$ ist die Matrix $G \in \mathbb{R}^{n \times n}$ mit

$$G := D_s g(s^*, T^*) = \frac{\partial x}{\partial s}(T^*; s^*, T^*) - I_n = M - I_n \quad (6.45)$$

an der gesuchten Lösung stets singular, denn sie besitzt wegen des Floquet-Multiplikators $m_1 = 1$ stets den Eigenwert 0. Wäre also selbst T^* bekannt, so ist die Nullstelle s^* von $g(s, T^*) = 0$ nicht regulär. Eine Regularisierung kann jedoch durch Einführung einer zusätzlichen *Phasenbedingung* erreicht werden. Wir ergänzen dazu das unterbestimmte System $g(s, T) = 0$ mittels der Funktion $h : D \times \mathbb{R} \rightarrow \mathbb{R}$ durch eine skalare Bedingung $h(s, T) = 0$, die die „Phase“ der Lösungsfunktion festlegt. Das erweiterte System

$$g(s, T) = 0, \quad h(s, T) = 0 \quad (6.46)$$

besitzt nunmehr $n + 1$ Gleichungen für die $n + 1$ Unbekannten $(s, T) \in \mathbb{R}^{n+1}$. Die Wahl von h soll zudem sichern, dass die erweiterte Jacobi-Matrix $\frac{\partial(g, h)}{\partial(s, T)}$ am Lösungspunkt (s^*, T^*) regulär ist. Folgende Phasenbedingungen werden empfohlen:

1. Die *einfache Komponentenbedingung* von Kubíček und Marek (1983) legt die i -te Komponente von s durch

$$h(s, T) := s_i - \alpha \quad \text{für ein } i \in \{1, \dots, n\} \quad (6.47)$$

mit der Konstanten $\alpha \in \mathbb{R}$ fest. Abbildung 6.9 zeigt, dass α im Variationsbereich der Lösungskomponente $x_i(t; s^*, T^*)$ liegen muss, weshalb die Auswahl von α und i problematisch sein kann. Wählen wir andererseits s^* so, dass $\dot{x}_i(0; s^*, T^*) = 0$ wird, so ergibt sich in Abbildung 6.9 die *Komponentenbedingung für die Ableitung* von Seydel (1991)

$$h(s, T) := f_i(s) \quad \text{für ein } i \in \{1, \dots, n\}, \quad (6.48)$$

bei der wiederum die Wahl des i entscheidend für die Verfahrenskonvergenz ist.

- Bei der *Poincaré-Bedingung* aus Keller (1987), Kubíček und Marek (1983), Seydel (1991) wird ein nahe der Lösung liegender Startpunkt s_0 vorausgesetzt. Nutzt man eine Parameterfortsetzung, so eignet sich dafür meist der betreffende Prädiktorwert (vgl. dazu Kapitel 5.1). Mit dem Tangentenvektor $\dot{s}_0 = f(s_0)$ an das Vektorfeld –

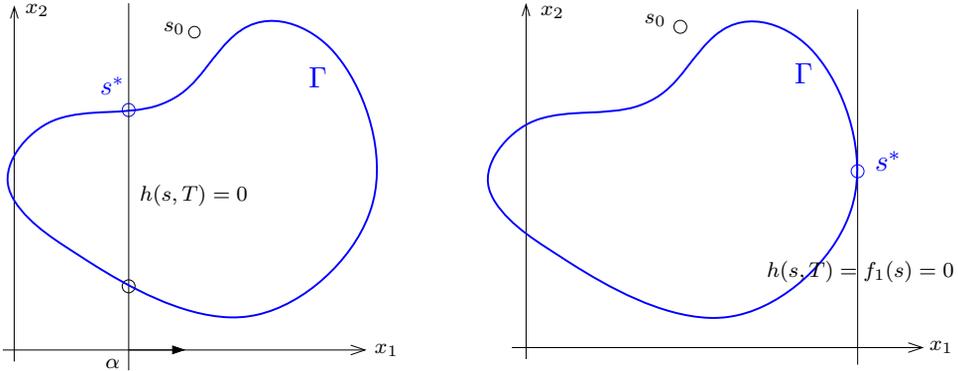


Abb. 6.9 Komponentenbedingung (links) und Ableitungsbedingung (rechts)

oder eine Näherung dafür – beschreibt

$$h(s, T) := f(s_0)^T \cdot (s - s_0) = 0 \tag{6.49}$$

in Abbildung 6.10 (links) anschaulich die Hyperebene durch s_0 mit Normalenvektor $f(s_0)$. Berechnen wir mit dem Newton-Verfahren die sukzessiven Näherungen s_1, s_2, \dots, s_k , so empfiehlt sich der jeweils aktuelle Wert s_{k-1} im k -ten Iterationsschritt als Startpunkt und liefert so die Bedingung von A. I. Mees (vgl. Parker und Chua (1989)) für s_k

$$h(s_k, T) := f(s_{k-1})^T \cdot (s_k - s_{k-1}) = 0, \quad k = 1, 2, 3, \dots \tag{6.50}$$

aus Abbildung 6.10 (rechts). Der Korrekturvektor jedes Iterationsschrittes wird also stets orthogonal zum Tangentenvektor an das Vektorfeld gefordert.

- Berücksichtigen wir die zu $x(0; s_0, T_0) = s_0$ gehörende Lösung $x_0(t) = x(t; s_0, T_0)$ des AWP auf dem gesamten Periodizitätsintervall $[0, T]$, so lässt sich mit dem Skalarprodukt $\langle u, v \rangle$ des Hilbert-Raumes $[\mathcal{L}^2([0, T])]^n$ eine Orthogonalitätsbedingung

$$\langle \dot{x}_0, x - x_0 \rangle := \int_0^T \dot{x}_0(t)^T [x(t; s, T) - x_0(t)] dt = 0$$

für die Abweichung $x(t; s, T) - x_0(t)$ der gesuchten Funktion formulieren. Nach Einsetzen der DGL fordert die *integrale Phasenbedingung* von E. Doedel Beyn (1991), Kuznetsov (1995)

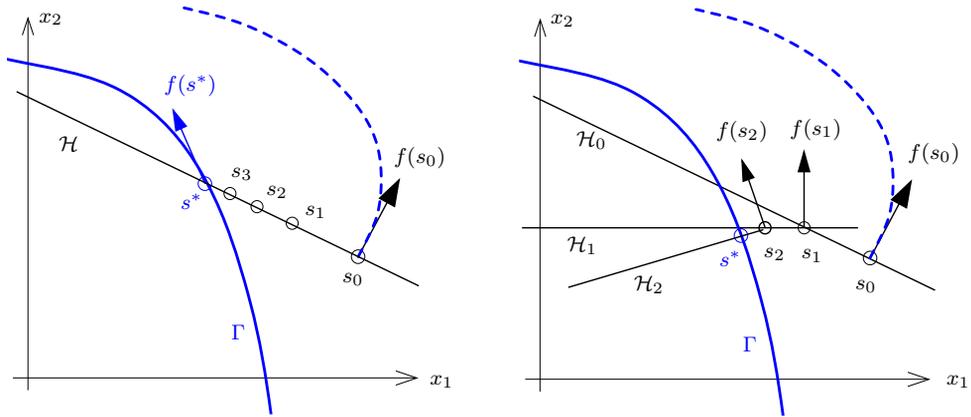


Abb. 6.10 Projektionsbedingung (links) und Bedingung von A. I. Mees (rechts)

$$h(s, T) := \int_0^T f(x_0(t))^T [x(t; s, T) - x_0(t)] dt = 0 \tag{6.51}$$

dass im Funktionenraum $[\mathcal{L}^2([0, T])]^n$ die Differenzfunktion $x - x_0$ orthogonal zur Funktion $\dot{x}_0 = f(x_0)$ sein muss.

Klären wir nun die Frage, unter welchen Voraussetzungen das erweiterte System (6.46) eine reguläre Lösung (s^*, T^*) , also die Ausgangsaufgabe einen Grenzykel der Periode T^* besitzt. Der Einfachheit halber wählen wir vorerst die Poincaré-Phasenbedingung (6.49), womit wir

$$\begin{aligned} g(s, T) &= x(T, s) - s &= 0 \\ h(s, T) &= f(s_0)^T (s - s_0) &= 0 \end{aligned} \tag{6.52}$$

mit der vollständigen Jacobi-Matrix (M ist darin die Monodromiematrix)

$$H := \left. \frac{\partial(g, h)}{\partial(s, T)} \right|_{(s^*, T^*)} = \begin{pmatrix} M - I_n & f(s^*) \\ f(s_0)^T & 0 \end{pmatrix} \tag{6.53}$$

erhalten. Über die erwünschte Regularität der Lösung $(s^*, T^*) \in \mathbb{R}^{n+1}$ zeigen wir folgenden

Satz 6.25

Voraussetzung 6.21 sei erfüllt, und $x(t; s^*, T^*)$ ist reguläre periodische Lösung, d. h. $m_1 = 1$ ist geometrisch einfacher Eigenwert von M . Dann gilt für $s^* = x(0; s^*, T^*)$:

- i. H ist regulär und (s^*, T^*) eine reguläre Lösung des erweiterten Systems (6.52).
- ii. Das Newton-Verfahren konvergiert lokal quadratisch gegen (s^*, T^*) .

Beweisidee: Ersetzt man in (6.49) s_0 durch s^* , so lässt sich für diese „ideale“ Projektionsbedingung $h^*(s, T) = f(s^*)^T(s - s^*) = 0$ die Regularität der zugehörigen Jacobi-Matrix leicht nachweisen. Man zeigt dazu, dass das homogene lineare Gleichungssystem

$$\begin{pmatrix} M - I_n & f(s^*) \\ f(s^*)^T & 0 \end{pmatrix} \cdot \begin{pmatrix} \sigma \\ \tau \end{pmatrix} = 0, \quad \sigma \in \mathbb{R}^n, \tau \in \mathbb{R}$$

nur die triviale Lösung $(\sigma, \tau)^T = 0$ besitzt. Also ist seine Koeffizientendeterminante verschieden von Null. Wegen der stetigen Abhängigkeit dieser Determinante von s_0 ist dann auch $\det H \neq 0$ für hinreichend nahe bei s^* liegende Werte s_0 , womit (s^*, T^*) eine reguläre Nullstelle von (g, h) ist. Zusammen mit der Glattheit von f folgt daraus die lokale Konvergenz des Newton-Verfahrens. \square

Zum entsprechenden Regularitätsnachweis für die integrale Phasenbedingung verweisen wir auf Doedel et al. (2003). Die Monodromiematrix M aus Formel (6.53) gewinnen wir als Nebenprodukt der Lösung: Denn konvergieren die Iterierten (s_k, T_k) des Newton-Verfahrens, so trifft das auch auf die entsprechenden Newton-Matrizen zu. Damit steht uns eine gute Approximation der Matrix H zur Verfügung. Durch Streichen der letzten Zeile und Spalte dieser Matrix entsteht $G = M - I_n$, woraus wir nach Addition der Einheitsmatrix I_n die Monodromiematrix M erhalten.

6.2.3 Technische Realisierung des Schießverfahrens

Wir versuchen nun, den Algorithmus 6.16 des Schießverfahrens ohne Nutzung des Variationsystems mit möglichst geringen Modifikationen auf den autonomen Fall zu übertragen und transformieren dazu das Problem mittels $t = \frac{T}{2\pi}\tau$ auf das Standardintervall $[0, 2\pi]$

$$\dot{x} = \frac{T}{2\pi}f(x), \quad x(2\pi) = x(0), \quad 0 < t < 2\pi. \quad (6.54)$$

Die unabhängige Variable haben wir nachfolgend wiederum mit t bezeichnet. Führen wir die konstante Funktion $T(t) = \text{const}$ ein, so ist zu den Näherungen (s_k, T_k) das transformierte AWP

$$\left. \begin{array}{l} \dot{x} = \frac{T}{2\pi}f(x), \quad x(0; s_k, T_k) = s_k \\ \dot{T} = 0, \quad T(0; s_k, T_k) = T_k \end{array} \right\} \text{ für } t \in [0, 2\pi] \quad (6.55)$$

zu lösen. Die nach Satz 6.8 garantierte Lösung sei $x(t; s_k, T_k), T(t; s_k, T_k)$. Das bestimmende Gleichungssystem mit Phasenbedingung (6.49)

$$\begin{aligned} g(s_k, T_k) &= x(2\pi; s_k, T_k) - s_k = 0 \\ h(s_k, T_k) &= f(s_0)^T(s_k - s_0) = 0 \end{aligned} \quad (6.56)$$

ist dann mit einem Newton-ähnlichen Verfahren zu lösen. Um auf die Notation von Algorithmus 6.16 zurück zu kommen, bezeichnen wir nun die um eine Komponente erweiterten Funktionenvektoren $xx : [0, 2\pi] \rightarrow \mathbb{R}^{n+1}$

$$xx(t) := \begin{pmatrix} x(t) \\ T(t) \end{pmatrix}, \quad ff(xx) := \begin{pmatrix} \frac{T}{2\pi} f(x) \\ 0 \end{pmatrix} \quad \text{mit } ff : DD \subset \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{n+1}$$

sowie die konstanten Vektoren in \mathbb{R}^{n+1} der Einfachheit wegen mit Doppelbuchstaben

$$ss := \begin{pmatrix} s \\ T \end{pmatrix} \quad \text{mit } ss_k := \begin{pmatrix} s_k \\ T_k \end{pmatrix}, \quad ss^* := \begin{pmatrix} s^* \\ T^* \end{pmatrix} \quad \text{und } gg(ss) := \begin{pmatrix} g(s, T) \\ h(s, T) \end{pmatrix}.$$

Dann erhält das AWP (6.55) wiederum die Standardform

$$\frac{d}{dt} xx = ff(xx), \quad xx(0; ss_k) = ss_k, \quad t \in [0, 2\pi], \quad (6.57)$$

und das bestimmende Gleichungssystem (6.56) lautet

$$gg(ss) = 0 \quad \text{mit Lösung } ss^* \in DD \subset \mathbb{R}^{n+1}. \quad (6.58)$$

Mit diesen Vektoren und Funktionen kann Algorithmus 6.16 leicht angepasst werden, indem statt der n Systeme nun stets $n + 1$ Gleichungen, definiert durch (6.56), (6.57) und (6.58), gelöst werden. Zu Startnäherungen s_0 und T_0 liefert der Algorithmus einen Punkt s^* des periodischen Orbits und zugleich dessen Periode T^* bis auf vorgegebene Genauigkeit tol . Schritt 3 in Algorithmus 6.16 ist abzuändern zu:

- 3. (Monodromiematrix)** Setze $GG := gg'(ss_k)$ und reduziere GG um die $(n + 1)$ -te Zeile und Spalte $\Rightarrow G$. Setze $M := G + I_n$ und bestimme davon die Eigenwerte $m = (m_1, m_2, \dots, m_n)$.

Die Monodromiematrix M kann auch hier nur mit einer maximalen Genauigkeit von $\sqrt{\varepsilon_M}$ mit der relativen Maschinengenauigkeit ε_M approximiert werden.

Beispiel 6.26

Das autonome System in \mathbb{R}^3 von W. F. Langford aus Beispiel 5.1

$$\begin{aligned} \dot{x}_1 &= (x_3 - 0.7)x_1 - \omega x_2 \\ \dot{x}_2 &= \omega x_1 + (x_3 - 0.7)x_2 \\ \dot{x}_3 &= 0.6 + x_3 - x_3^3/3 - (x_1^2 + x_2^2)(1 + \varrho x_3) + \varepsilon x_3 x_1^3 \end{aligned} \quad (6.59)$$

besitzt mit den Konstanten $\omega = 3.5$, $\varepsilon = 0$, $\beta = 0.7$ und dem Kontrollparameter $\varrho \in [0.25, 2.00]$ die T -periodische Lösung

$$x(t; s^*, T^*) = \begin{pmatrix} r \cos \omega t \\ r \sin \omega t \\ \beta \end{pmatrix} \quad \text{mit } r := \sqrt{\frac{0.6 + \beta - \frac{1}{3}\beta^3}{1 + \varrho \cdot \beta}} = \frac{1.08888321993989}{\sqrt{1 + 0.7\varrho}}.$$

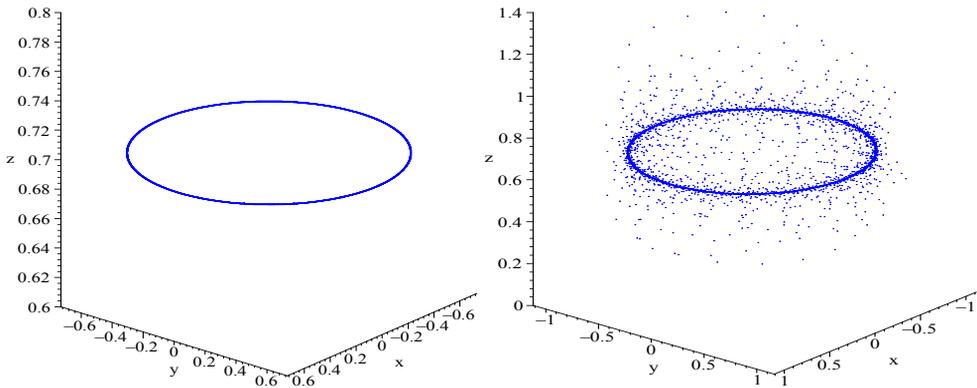


Abb. 6.11 Einschwingen der periodischen Lösungen für $\varrho = 2.0$ (links) und $\varrho = 0.7$ (rechts)

Die Phasenraum-Darstellung der Lösungen in Abbildung 6.11 lässt den langsamen Einschwingprozess für den Parameterwert $\varrho = 0.7$ erkennen. Die exakte Periodendauer $T_{ex}^* = 2\pi/\omega = 1.795\ 195\ 802\ 051\ 310\dots$ ist bekannt und kann damit zur Kontrolle der Resultatgenauigkeit genutzt werden. Algorithmus 6.16 liefert mit Startwerten $s_0 = (1, 0, 1)^T$, $T_0 = 2.0$ und Genauigkeitsschranke $tol = 10^{-9}$ die Werte in Tabelle 6.1. Während die Periodendauer T^* als 4. Lösungskomponente hier mit hoher Genauigkeit approximiert wurde, wird der Standardmultiplikator $m_1 = 1$ (der stets als Kontrollwert dienen kann) nur auf sechs Dezimalstellen geliefert. Die beiden restlichen Floquet-Multiplikatoren $m_{2,3}$ sind in diesem Beispiel stets konjugiert komplex. ■

Die technische Umsetzung der integralen Phasenbedingung (6.51) gestaltet sich schwieriger als die soeben angewandte Poincaré-Bedingung. Da sie in der Regel auf ein zuverlässigeres Lösungsverfahren führt, notieren wir sie nun ebenfalls für die transformierten Gleichungen (6.54):

$$\int_0^{2\pi} [x(t) - x_0(t)]^T \dot{x}_0(t) dt = 0. \quad (6.60)$$

Ist die darin auftretende benachbarte Lösung $x_0(t)$ selbst 2π -periodisch – wenn diese z. B. durch eine Lösungsfortsetzung ermittelt wurde – so ergibt partielle Integration in (6.60) das zu lösende Randwertproblem mit einer Integralnebenbedingung

$$\dot{x} = \frac{T}{2\pi} f(x), \quad x(2\pi) = x(0), \quad \int_0^{2\pi} x(t)^T \dot{x}_0(t) dt = 0, \quad 0 < t < 2\pi. \quad (6.61)$$

Mit einem kleinen Trick gelingt es, diese Aufgabe in die Standardform für Zweipunkt-Randwertprobleme zu überführen. Wir definieren dazu den neuen Funktionenvektor $y : [0, 2\pi] \rightarrow \mathbb{R}^{n+2}$ durch

$$y_i(t) := x_i(t), \quad i = 1(1)n, \quad y_{n+1}(t) := T, \quad y_{n+2}(t) := \int_0^t x(\tau)^T \dot{x}_0(\tau) d\tau$$

Tab. 6.1 Berechnete Lösungen und Floquet-Multiplikatoren zu Beispiel 6.26

ϱ	Lösung (s^*, T^*)	Multiplikatoren m_i	$ m_{2,3} $	$T^* - T_{ex}^*$
2.0	$s_1^* = 6.975\ 518e-1$ $s_2^* = -8.630\ 936e-2$ $s_3^* = 6.999\ 999e-1$ $T^* = 1.795\ 195\ 801\ 964\ 12$	$m_1 = 0.999\ 998\ 877$ $m_2 = -0.597+0.260i$ $m_3 = -0.597-0.260i$	0.651	$-8.72e-11$
0.6	$s_1^* = 9.068\ 526e-1$ $s_2^* = -1.122\ 269e-1$ $s_3^* = 6.999\ 999e-1$ $T^* = 1.795\ 195\ 801\ 974\ 54$	$m_1 = 1.000\ 000\ 348$ $m_2 = -0.937+0.371i$ $m_3 = -0.937-0.371i$	1.008	$-7.68e-11$
0.25	$s_1^* = 9.971\ 034e-1$ $s_2^* = -1.219\ 126e-1$ $s_3^* = 6.999\ 999e-1$ $T^* = 1.795\ 195\ 801\ 983\ 02$	$m_1 = 1.000\ 003\ 622$ $m_2 = -1.167+0.475i$ $m_3 = -1.167-0.475i$	1.260	$-6.83e-11$

und erhalten die äquivalente Standardaufgabe in \mathbb{R}^{n+2} mit $n+2$ Differenzialgleichungen und $n+2$ Randbedingungen

$$\begin{aligned}
 \dot{y}_i &= \frac{y_{n+1}}{2\pi} f_i(y_1, \dots, y_n), & y_i(2\pi) &= y_i(0), \quad i = 1(1)n, \\
 \dot{y}_{n+1} &= 0, & y_{n+2}(2\pi) &= 0, \\
 \dot{y}_{n+2} &= \sum_{i=1}^n y_i \cdot \dot{y}_{i0}, & y_{n+2}(0) &= 0.
 \end{aligned} \tag{6.62}$$

Bei gegebener Startlösung $y_0(t) = (y_{10}(t), y_{20}(t), \dots, y_{n0}(t))^T$ liefert deren Einsetzen in die DGL die hier benötigten Ableitungen $\dot{y}_{i0}(t)$. Auf die Aufgabe (6.62) können dann die bekannten Standardverfahren aus Hoffmann et al. (2006) angewendet werden. Die Umsetzung dieses Zuganges wird dem Leser überlassen.

6.2.4 Stabilitätsanalyse

Anders als im speziellen Beispiel 6.26 hängt die Periodendauer T^* in autonomen Systemen oft vom Kontrollparameter ab. So entsteht das Problem, dass benachbarte periodische Orbits $\{\varphi_t(s^*)\}$ und $\{\varphi_t(s)\}$ verschiedene Periodendauer T^* und T haben können. Für das mathematische Pendel mit der Gleichung

$$\ddot{x} + \sin x = 0 \quad \text{bzw. als System} \quad \dot{x}_1 = x_2, \quad \dot{x}_2 = -\sin x_1 \tag{6.63}$$

sind in Abbildung 6.12 zahlreiche periodische Orbits mit $s^* = (\alpha^*, 0)^T$, $0 < \alpha^* < \pi$ in blau dargestellt.

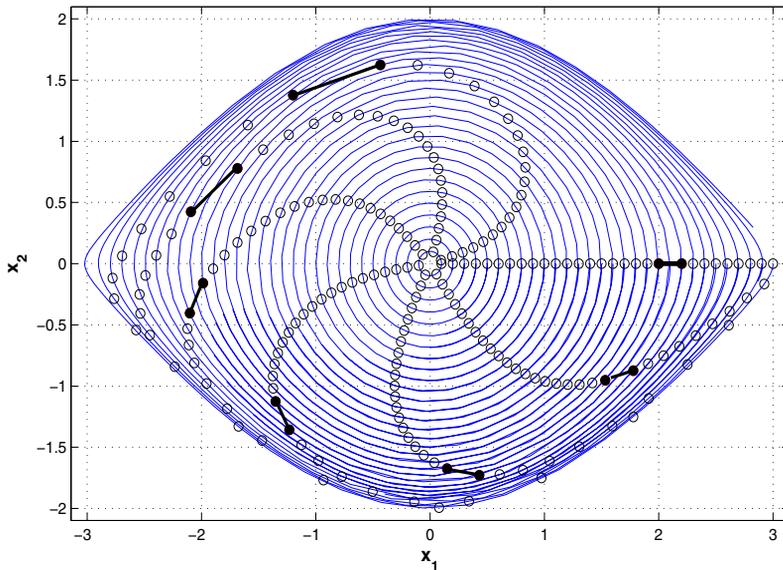


Abb. 6.12 Verlauf der Orbits in (6.63) zu identischen Zeitpunkten

Für kleine Auslenkung α^* wird (6.63) approximativ durch die harmonische Gleichung mit Periodendauer $T^* = 2\pi$ beschrieben, während für $\alpha^* \rightarrow \pi$ die Periode T^* gegen Unendlich tendiert, denn $(\pm\pi, 0)^T$ sind Gleichgewichtspunkte. Durch $s^* = (\alpha^*, 0)^T$ verlaufe der in Abbildung 6.12 mit schwarzen Punkten gekennzeichnete Orbit. Der Abstand $\|\varphi_t(s) - \varphi_t(s^*)\|$ des benachbarten periodischen Orbits durch $s = (\alpha, 0)^T$ (schwarz eingezeichnet) wächst für $t \rightarrow \infty$ sogar über den Wert α^* an. Obwohl beide Orbits im Phasenraum geometrisch nahe beieinander liegen, kann $\{\varphi_t(s^*)\}$ nicht Ljapunov-stabil sein! Wir betrachten deshalb den geometrischen Abstand zweier Phasenkurven, ohne aber dieselbe Zeit t für je zwei Punkte zu fordern. Es sei $\gamma^* = \{x \in D \mid x = \varphi_t(s^*), t \geq 0\}$ der (nicht notwendig periodische) positive Semiorbit durch $s^* \in D$. Zu $s \in D$ bezeichnet dann

$$d(s, \gamma^*) := \inf_{t \geq 0} \|s - \varphi_t(s^*)\| \quad (6.64)$$

den geometrischen Abstand des Punktes von γ^* im Phasenraum. Damit gewinnen wir

Definition 6.27 (Orbitale Stabilität)

- i. $\{\varphi_t(s^*)\}$, $t \geq 0$, heißt *orbital stabil*, wenn zu $\varepsilon > 0$ ein $\delta > 0$ existiert, so dass für alle $s \in D$ mit $\|s - s^*\| < \delta$ stets $d(\varphi_t(s), \gamma^*) < \varepsilon \quad \forall t \geq 0$ gilt.
- ii. Ist $\{\varphi_t(s^*)\}$ orbital stabil und es existiert ein $\delta_1 > 0$, so dass für alle $s \in D$ mit $d(s, \gamma^*) < \delta_1$ stets $\lim_{t \rightarrow \infty} d(\varphi_t(s), \gamma^*) = 0$ folgt, so ist $\{\varphi_t(s^*)\}$ *asymptotisch orbital stabil*.
- iii. Ist $\{\varphi_t(s^*)\}$ nicht orbital stabil, so heißt es *orbital instabil*.

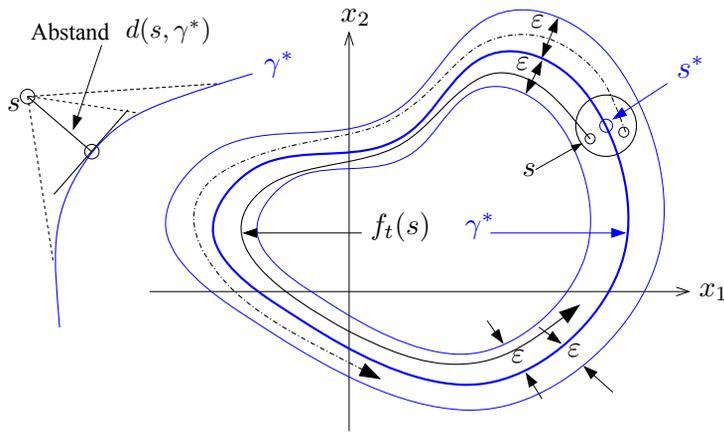


Abb. 6.13 Orbitale Stabilität in der Phasenebene

Abbildung 6.13 demonstriert den Begriff der orbitalen Stabilität in der Phasenebene. Wegen

$$d(\varphi_t(s), \gamma^*) = \inf_{t \geq 0} \|\varphi_t(s) - \gamma^*\| \leq \|\varphi_t(s) - \varphi_t(s^*)\| \quad \forall t \geq 0$$

folgt aus der (asymptotischen) Ljapunov-Stabilität die (asymptotische) orbitale Stabilität. Ist x^* eine Gleichgewichtslösung, so fallen diese beiden Stabilitätsbegriffe zusammen, denn für $x \in D$ gilt nun die Gleichheit $d(\varphi_t(x), \gamma^*) = \|\varphi_t(x) - \varphi_t(x^*)\|$ für alle $t \geq 0$. Eine algebraische Charakterisierung liefert der folgende grundlegende Satz, dessen umfangreichen Beweis man in der Grundlagenliteratur, z. B. bei Pontrjagin (1965), findet:

Satz 6.28 (Andronov & Witt)

$\{\varphi_t(s^*)\}$ sei ein periodischer Orbit des autonomen Systems $\dot{x} = f(x)$ mit den Floquet-Multiplikatoren m_i , $i = 1(1)n$, wobei der Hauptmultiplikator $m_1 = 1$ ist.

- i. Ist $|m_i| < 1$ für $i = 2(1)n$, so ist $\{\varphi_t(s^*)\}$ asymptotisch orbital stabil.
- ii. Ist $|m_i| > 1$ für ein $i \in \{2, \dots, n\}$, so ist $\{\varphi_t(s^*)\}$ orbital instabil.

Beispiel 6.29

Wir betrachten die Ergebnisse aus Beispiel 6.26 für den periodischen Orbit der Periode $T_{ex}^* = 2\pi/\omega = 1.795\ 195\ 802\ 051\ 310\dots$ Während der Satz von Andronov & Witt für den Kontrollparameter $\varrho = 2.0$ die asymptotische Stabilität garantiert, ist der Orbit für die Parameterwerte $\varrho = 0.6$ und $\varrho = 0.25$ instabil (vgl. Tabelle 6.1). Genauere Rechnung liefert für den „kritischen“ Wert $\varrho = 0.615\ 444\ 65\dots$ Floquet-Multiplikatoren mit $m_1 = 1$, $|m_{2,3}| = 1$, so dass mit dem Satz dann keine Stabilitätsaussage möglich ist. ■

6.3 Die Poincaré-Abbildung

6.3.1 Definition und Eigenschaften

Die Poincaré-Abbildung stellt ein wesentliches Hilfsmittel dar, um Existenz, Stabilität und Verzweigungen periodischer Lösungen zu untersuchen und zugleich das Lösungsverhalten geometrisch zu veranschaulichen. So wurde in Abschnitt 1.5 diese Abbildung genutzt, um in Beispiel 4.23 die Anwendung des impliziten Funktionentheorems auf dem Gebiet der Differentialgleichungen zu demonstrieren. Der Nachweis eines Grenzzykels einer autonomen DGL gelang in Abschnitt 4.2 ebenfalls mittels der Poincaré-Abbildung.

Wir wollen aus diesem Grund die Konstruktion und die Eigenschaften dieser wichtigen Abbildung detailliert behandeln, weiter verallgemeinern und schließlich ein überaus effektives numerisches Verfahren zur Berechnung der Abbildungspunkte gewinnen. Wir betrachten dazu das autonome System

$$\dot{x} = f(x), \quad f : D \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n \quad (6.65)$$

mit glattem Vektorfeld $f \in C^r(D)$, $r \geq 2$, und dem damit erzeugten Fluss $\varphi_t : D \rightarrow \mathbb{R}^n$. Nach Voraussetzung 6.21 besitzt das System eine periodische Lösung $\varphi_t(x_0)$ der Periode

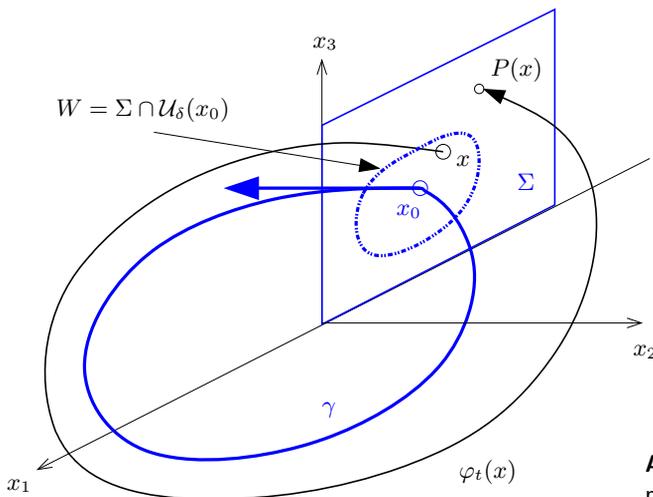


Abb. 6.14 Poincaré-Abbildung mit Ebene Σ in \mathbb{R}^3

T , die eine gegebene Anfangsbedingung $\varphi_0(x_0) = x_0$ erfüllt. Wir legen nun, wie in Abbildung 6.14 dargestellt, eine Hyperebene Σ durch x_0 , die senkrecht auf dem Orbit $\gamma = \{\varphi_t(x_0)\}$ steht und betrachten Lösungen $\varphi_t(x)$, die an Punkten $x \in \Sigma$ aus einer δ -Umgebung $W = \Sigma \cap \mathcal{U}_\delta(x_0)$ starten. Liegt x hinreichend nahe bei x_0 , so ist zu erwarten, dass $\varphi_t(x)$ die Hyperebene ebenfalls nahe x_0 in einem Punkt $P(x)$ schneiden wird. Die

Abbildung $x \mapsto P(x)$ wurde von *Henri Poincaré*² bereits 1881 eingeführt und später nach ihm benannt. Die Existenz und stetige Differenzierbarkeit dieser Abbildung wird z. B. von Perko (1996) nachgewiesen. Die benötigte Rückkehr-Zeit $\tau(x)$ hat folgende Eigenschaft:

Satz 6.30

Es sei $D \subset \mathbb{R}^n$ eine offene Menge und $f \in C^1(D)$. Die periodische Lösung $\varphi_t(x_0)$ habe die Periode T und der durch diese Lösung definierte Orbit γ liege in D .

$$\Sigma = \{x \in \mathbb{R}^n \mid f(x_0)^T(x - x_0) = 0\} \quad (6.66)$$

beschreibt die zu γ orthogonale Hyperebene bei x_0 . Dann existieren ein $\delta > 0$ und eine eindeutig bestimmte Funktion $\tau(x)$, definiert und stetig differenzierbar für alle x aus einer δ -Umgebung $\mathcal{U}_\delta(x_0)$, so daß $\tau(x_0) = T$ ist und $\varphi_{\tau(x)}(x) \in \Sigma \quad \forall x \in \mathcal{U}_\delta(x_0)$ gilt.

Die Rückkehr-Zeit $\tau(x)$ ist also diejenige Zeit, die ein bei $x \in W$ beginnender Orbit $\varphi_t(x)$ benötigt, um wieder in Σ anzukommen. Dieses $\tau(x)$ hängt im Allgemeinen vom Startpunkt x ab und muss nicht identisch mit der Periode T von γ sein. Allerdings konvergiert $\tau(x) \rightarrow T$ für $x \rightarrow x_0$. Die lokal definierte eindeutige Abbildung $P : W \rightarrow \mathbb{R}^n$ mit $W = \mathcal{U}_\delta(x_0) \cap \Sigma$ und

$$P(x) = \varphi_{\tau(x)}(x) \quad (6.67)$$

heißt *lokale Poincaré-Abbildung* (auch: *first return map*) für den Orbit γ an der Stelle x_0 . Folgende Eigenschaften lassen sich für die Abbildung P nachweisen:

1. $P \in C^1(W)$. Falls f sogar analytisch in D ist, so ist auch P analytisch in W .
2. P hat eine stetig differenzierbare inverse Abbildung P^{-1} ; damit ist P ein C^1 -Diffeomorphismus (d. h. eine C^1 -Funktion mit C^1 -Inverser).
3. $x \in \Sigma$ ist ein Fixpunkt von P , d. h. $P(x) = x$ genau dann, wenn $\varphi_t(x)$ ein periodischer Orbit von (6.66) ist.

Der Begriff der Poincaré-Abbildung lässt sich auch auf periodische Orbits verallgemeinern, die auf den in Abschnitt 2.7.3 eingeführten Mannigfaltigkeiten verlaufen: Es sei $\gamma(t)$ ein periodischer Orbit eines Flusses φ_t mit Periode T auf einer Mannigfaltigkeit M . Weiterhin sei x_0 ein Punkt des Orbits mit $x_0 = \gamma(0)$ und Σ ein lokaler transversaler Schnitt, d. h. eine Untermannigfaltigkeit der Kodimension 1, transversal zu γ (also ist $\gamma'(0)$ nicht tangential zu Σ). Es sei $D \subset M \times \mathbb{R}$ ein offenes Gebiet, auf dem der Fluss φ_t definiert ist.

²Mit epochalen Werken zu mathematischen Methoden der Himmelsmechanik gilt Jules Henri Poincaré (1854–1912) als einer der herausragenden Mathematiker und theoretischen Astronomen unserer Zeit und zugleich als Pionier der Theorie dynamischer Systeme.

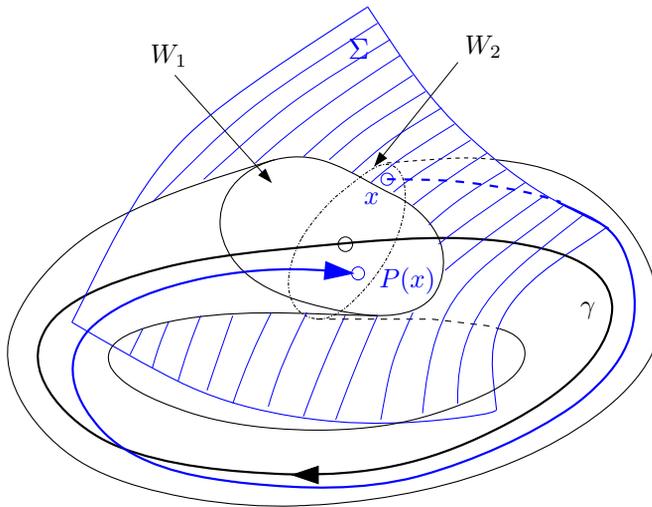


Abb. 6.15 Poincaré-Abbildung auf Mannigfaltigkeiten

Definition 6.31 (Verallgemeinerte Poincaré-Abbildung)

Eine *lokale Poincaré-Abbildung* (*first return map*) von γ ist eine Abbildung $P : W_0 \rightarrow W_1$ mit folgenden Eigenschaften (vgl. Abb. 6.15):

- i. $W_0, W_1 \subset \Sigma$ sind offene Umgebungen von $x_0 \in \Sigma$ und P ist ein C^r -Diffeomorphismus.
- ii. Es existiert eine Funktion $\tau : W_0 \rightarrow \mathbb{R}$, so dass für alle $x \in W_0$ gilt: $(x, \tau(x)) \in D$ und $P(x) = \varphi_{\tau(x)}(x)$.
- iii. Falls $t \in (0, \tau(x))$, so $\varphi_t(x) \notin W_0$.

Die Existenz einer verallgemeinerten Poincaré-Abbildung in der Umgebung eines geschlossenen Orbits wird von Marsden und McCracken (1976) nachgewiesen. In \mathbb{R}^3 kann der Poincaré-Schnitt also eine beliebige glatte Fläche Σ sein, die den Orbit γ allerdings transversal schneiden muss.

Poincaré-Abbildungen lassen sich oft auch ohne Bezug auf einen periodischen Orbit konstruieren. Existiert eine Hyperebene $\Sigma \subset \mathbb{R}^n$, die

- von jeder Lösungskurve $\{\varphi_t(x)\}$ transversal geschnitten wird und
- zu jedem $x \in \Sigma$ einen Rückkehrpunkt $\varphi_t(x) \in \Sigma$, $t > 0$, besitzt,

so definiert Σ einen globalen Poincaré-Schnitt, und die Abbildung $P : \Sigma \rightarrow \Sigma$ mit

$$P(x) = \varphi_{\tau(x)}(x), \quad x \in \Sigma,$$

ist eine *globale Poincaré-Abbildung* mit der Rückkehrzeit $\tau(x)$. Zur Anwendung notieren wir periodisch erregte DGL-Systeme als autonome DGL:

$$\dot{x} = f(x, \theta), \quad \dot{\theta} = 1, \quad (x, \theta) \in \mathbb{R}^n \times \mathbb{S}^1. \quad (6.68)$$

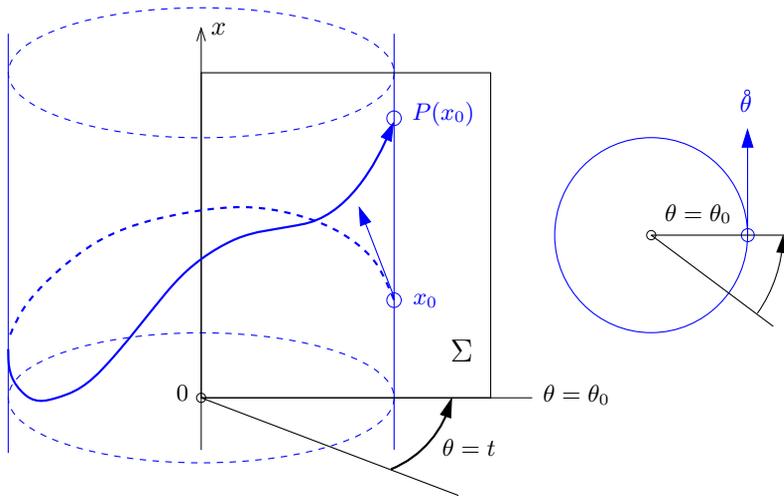


Abb. 6.16 Stroboskop-Abbildung

Der Phasenraum wird dann durch die zylindrische Mannigfaltigkeit $M = \mathbb{R}^n \times \mathbb{S}^1$ beschrieben, wobei $\mathbb{S}^1 = \mathbb{R}(\text{mod } T)$ als zirkulare Komponente die T -Periodizität des Vektorfeldes f in θ widerspiegelt. Für dieses Problem kann man nun den globalen Poincaré-Schnitt

$$\Sigma = \{ (x, \theta) \in \mathbb{R}^n \times \mathbb{S}^1 \mid \theta = \theta_0 \}$$

definieren, da wegen $\dot{\theta} = 1$ alle Lösungen Σ transversal schneiden (vgl. Abb. 6.16). Die Poincaré-Abbildung $P : \Sigma \rightarrow \Sigma$ wird nun definiert durch

$$P(x_0) = \pi\varphi_T(x_0, \theta_0), \tag{6.69}$$

worin φ_t der Fluss von (6.68) ist und π seine Projektion auf die x -Komponente bedeutet. Offenbar ist die Zeit zwischen zwei Abbildungspunkten stets gleich T für alle $x \in \Sigma$. Wählen wir ohne Beschränkung der Allgemeinheit $\theta_0 = 0$ und bezeichnen die Projektion des Flusses auf die x -Komponente mit $\psi_t(x) := \pi\varphi_t(x, 0)$, $x \in D \subset \mathbb{R}^n$, so kommen wir zur *Stroboskop-Abbildung*³ $P : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ mit

$$P(x) = \psi_T(x) = \pi\varphi_T(x, 0).$$

Die k -te Iteration der stroboskopischen Abbildung liefert den Systemzustand zur Zeit $t_0 + kT$, d. h. das System wird periodisch „beleuchtet“. Unter den Voraussetzungen 6.3 ist diese eindeutige Abbildung für alle x in einer Umgebung $\mathcal{U}_\rho(x_0) \subset D$ des Punktes $x_0 = x^*(0)$ definiert.

³Stroboskop (griech.): optisches Gerät zum Beobachten und Messen rasch ablaufender Vorgänge (z. B. Drehzahlmessung) mit Hilfe eines periodisch unterbrochenen Lichtstrahls.

Bemerkung 6.32

Mittels der Poincaré-Abbildung P lassen sich oft Konvergenz- und Stabilitätsaussagen für Fixpunktaufgaben in \mathbb{R}^n , z. B. Fixpunktsätze, anwenden. Denn zu $k \in \mathbb{Z}$ bedeutet die k -fache Anwendung der Poincaré-Abbildung auf ein $x \in D$ den Übergang vom betrachteten zeitkontinuierlichen dynamischen System $DS_1 = (\mathbb{R}, D, \varphi_t)$ zu einem zeitdiskreten System $DS_2 = (\mathbb{Z}, D, \psi_k)$ mit

$$\psi_k(x) := P^k(x) = P(P \cdots P(P(x)) \cdots), \quad x \in D, k \in \mathbb{Z}.$$

Der Übergang von DS_1 zum zeitdiskreten System DS_2 transformiert insbesondere

- T -periodische Lösungen $\varphi_t(x_0)$ in Fixpunkte x_0 von P
- m -fach subharmonische Lösungen, d. h. mT -periodische Lösungen mit $m \in \mathbb{N}, m > 1$, in m -periodische Orbits von DS_2 sowie
- quasi-periodische Lösungen von DS_1 mit 2 Basisfrequenzen in geschlossene Invarianzkurven. von DS_2 ■

6.3.2 Ableitung der Poincaré-Abbildung

Die Ableitung $DP(x_0)$ der Poincaré-Abbildung in einem Punkt x_0 des periodischen Orbits $\gamma = \{\varphi_t(x_0)\}$ erhalten wir, indem wir den Effekt kleiner Variationen der Anfangsbedingungen $x_0 \in \mathbb{R}^n$ analysieren. Zur Vereinfachung der Darstellung verschieben wir den Koordinatenursprung in den Punkt x_0 und benutzen als x_n -Achse den normierten Tangentenvektor $\xi_n = f(x_0)/\|f(x_0)\|$ an den Orbit im Punkt x_0 . Die restlichen $n-1$ orthogonalen Basisvektoren $\xi_1, \xi_2, \dots, \xi_{n-1}$ sollen die Hyperebene Σ aufspannen, wie dies in Abbildung 6.17 für den 3-dimensionalen Fall dargestellt ist. Unter diesen Annahmen

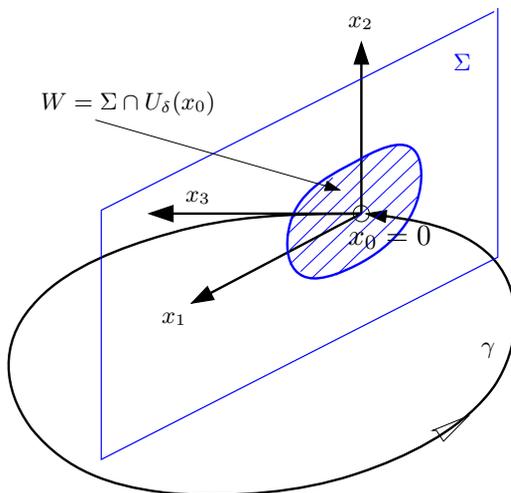


Abb. 6.17 Poincaré-Schnitt zu (6.71)

kann die Ableitung $DP(0)$ der Poincaré-Abbildung im Punkt $x_0 = 0$ angegeben werden. Einen Beweis des Satzes findet man bei Perko (1996).

Satz 6.33

Es sei $f \in C^1(D)$ und $\varphi_t(0)$ bezeichne den periodischen Orbit γ mit Periode T^* in D . Für hinreichend kleines $\delta > 0$ und $x \in \Sigma \cap \mathcal{U}_\delta(0)$ sei $P(x)$ die gerade eingeführte Poincaré-Abbildung. Dann wird die Ableitung $DP(0)$ durch die Matrix

$$DP(0) = \left(\frac{\partial \varphi_i}{\partial x_j}(0, T^*) \right), \quad i, j = 1(1)n - 1, \quad (6.70)$$

definiert, wobei $\varphi(x, t) = \varphi_t(x)$ der durch (6.66) gegebene Fluss ist.

Mit der Definition 6.12 der Monodromiematrix

$$M = \frac{\partial \varphi}{\partial x}(s^*, T^*) = X(s^*, T^*) = e^{T^* R}$$

und $s^* = 0$ erhalten wir $DP(0)$ aus der Monodromiematrix, indem wir deren n -te Zeile und n -te Spalte streichen. Die Eigenwerte von $DP(0)$ sind nun genau die $n - 1$ Floquet-Multiplikatoren m_1, m_2, \dots, m_{n-1} ohne den Hauptmultiplikator $m_n = 1$. Wegen der Invarianz des Spektrums bezüglich linearer Transformationen gilt diese Eigenschaft auch allgemein für beliebige orthogonale Poincaré-Schnitte in einem Kurvenpunkt x_0 .

Beispiel 6.34

Das um eine Gleichung erweiterte Beispiel 6.22 mit Parameter $\lambda > 0$

$$\begin{aligned} \dot{x}_1 &= -x_1 + 2 \\ \dot{x}_2 &= \lambda x_2 - x_3 - x_2(x_2^2 + x_3^2) \\ \dot{x}_3 &= x_2 + \lambda x_3 - x_3(x_2^2 + x_3^2) \end{aligned} \quad (6.71)$$

besitzt zum Punkt $s^* = (2, \sqrt{\lambda}, 0)^T$ und $T^* = 2\pi$ die periodische Lösung mit Hauptlösungsmatrix (vgl. dazu Bsp. 6.24)

$$x(t; s^*, T^*) = \begin{pmatrix} 2 \\ \sqrt{\lambda} \cos t \\ \sqrt{\lambda} \sin t \end{pmatrix}, \quad X(t; s^*, T^*) = \begin{pmatrix} e^{-t} & 0 & 0 \\ 0 & e^{-2\lambda t} \cdot \cos t & -\sin t \\ 0 & e^{-2\lambda t} \cdot \sin t & \cos t \end{pmatrix},$$

was man leicht durch Lösung des Variationssystems bestätigt. Die Monodromiematrix $M = X(T^*; s^*, T^*) = \text{diag}(e^{-2\pi}, e^{-4\lambda\pi}, 1)$ liefert außer dem Hauptmultiplikator $m_3 = 1$ nun die beiden Eigenwerte $m_1 = e^{-2\pi}$, $m_2 = e^{-4\lambda\pi}$ der Ableitung $DP(s^*)$ der in Abbildung 6.17 dargestellten Poincaré-Abbildung. ■

Notieren wir periodisch erregte Systeme wie in (6.68) ebenfalls als autonome DGL in \mathbb{R}^{n+1}

$$\dot{x} = f(x, \theta), \quad \dot{\theta} = 1, \quad (x, \theta) \in \mathbb{R}^n \times \mathbb{S}^1, \quad (6.72)$$

so ergibt sich die leicht nachweisbare

Folgerung 6.35

Der Hauptmultiplikator von (6.72) ist $m_{n+1} = 1$ und die weiteren Eigenwerte von $DP(x_0)$ sind genau die n Floquet-Multiplikatoren m_1, m_2, \dots, m_n der Lösung des periodisch erregten Systems bzw. der Stroboskop-Abbildung $P(x_0) = x(x_0, T)$.

Beweis: Mit dem erweiterten Vektor $y = (y_1, y_2, \dots, y_n, y_{n+1})^T = (x, \theta)^T$ lautet das erweiterte Variationssystem an der periodischen Lösung

$$\dot{Y} = \begin{pmatrix} D_x f(x^*(t), t) & D_t f(x^*(t), t) \\ 0 & 0 \end{pmatrix}, \quad Y(0) = I_{n+1},$$

woraus wir mit der vorgegebenen Lösungsperiode T die erweiterte Monodromiematrix

$$M = Y(T) = \begin{pmatrix} X(T) & \xi(T) \\ 0 & 1 \end{pmatrix}, \quad \xi(T) \in \mathbb{R}^n,$$

erhalten. Darin ist $X(T)$ die Monodromiematrix des periodisch erregten Systems mit den n Floquet-Multiplikatoren m_1, m_2, \dots, m_n , womit die Behauptung folgt. \square

6.3.3 Numerische Approximation der Poincaré-Abbildung

In praktischen Anwendungen kann die Poincaré-Abbildung P selten explizit berechnet werden und muss deshalb durch numerische Lösung der DGL approximiert werden. Für T -periodisch erregte Systeme ist das wegen des speziellen Falles der Stroboskop-Abbildung leicht möglich, wenn wir die Lösung $\varphi(t; t_0, x_0)$ an den vorgegebenen Zeitpunkten $t_k = kT$ mit $k = 1, 2, 3, \dots$ berechnen und als Werte $P^k(x_0)$ abspeichern. Anders verhält es sich jedoch bei autonomen Systemen, wenn ein Durchstoßpunkt der Trajektorie $\varphi(t; t_0, x_0)$ durch eine Hyperebene Σ numerisch zu bestimmen ist. Die Rückkehrzeit $t^* = \tau(x_0)$ gemäß Satz 6.30 ist meist unbekannt und muss deshalb möglichst effizient berechnet werden. Nachfolgend werden zwei Zugänge dafür vorgestellt. Wir lassen als Poincaré-Schnitt beliebige Hyperebenen $\Sigma \subset \mathbb{R}^n$ mit der Gleichung

$$H(x) := \langle v, x - \xi \rangle = v^T(x - \xi) = \sum_{i=1}^n v_i(x_i - \xi_i) = 0, \quad x \in \mathbb{R}^n \quad (6.73)$$

zu, die jedoch die Transversalitätsbedingung $v^T f(\varphi(t^*; t_0, x_0)) > 0$ erfüllen sollen, d. h. der Normalvektor v von Σ bildet mit dem Tangentialvektor an das Vektorfeld in $\varphi(t^*; t_0, x_0)$ einen spitzen Winkel. ξ ist der „Aufpunkt“ der Hyperebene und v der Normalvektor mit $\|v\|_2 = 1$, wie dies für den 2-dimensionalen Fall in Abbildung 6.18 dargestellt ist. Zu gegebenem Startpunkt $x_0 \in \Sigma$ und Startzeit t_0 wollen wir die Lö-

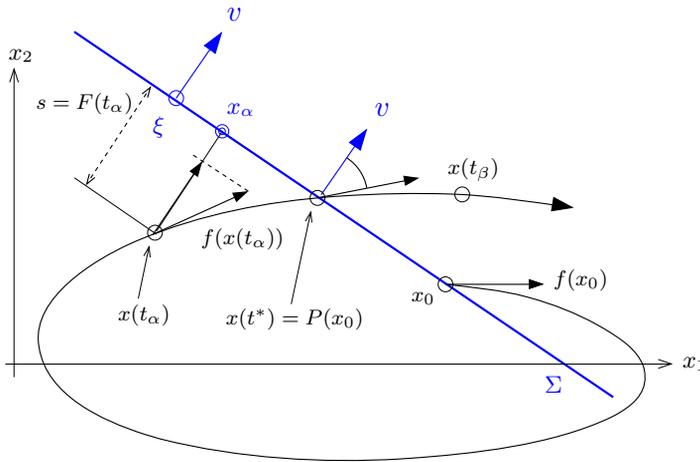


Abb. 6.18 Hyperbene Σ und Poincaré-Schnitt in \mathbb{R}^2

sung der DGL $\dot{x} = f(x)$ vereinfachend mit $x(t) = \varphi(t; t_0, x_0)$ bezeichnen. Eine numerische Integration liefert für $t = t_\alpha$ den letzten „vor der Hyperbene“ liegenden Lösungswert $x(t_\alpha)$, während der folgende Lösungspunkt $x(t_\beta)$ mit Integrationssschrittweite $h = t_\beta - t_\alpha > 0$ bereits jenseits von Σ liegt. Projizieren wir – wie in Abbildung 6.18 dargestellt – den Punkt $x(t_\alpha)$ orthogonal auf die Hyperbene in den Punkt x_α , so erhalten wir mit $x(t_\alpha) = x_\alpha + sv$ aus der Hyperbenen-Gleichung (6.73)

$$H(x(t_\alpha)) := \langle v, x(t_\alpha) - \xi \rangle = v^T(x_\alpha + sv - \xi) = v^T(x_\alpha - \xi) + sv^T v = s$$

mit dem negativen Abstand s .

1. Das *Newton-Verfahren* bestimmt eine Nullstelle t^* der skalarwertigen Abstandsfunktion $F : I^* \rightarrow \mathbb{R}$, definiert durch

$$F(t) := H(x(t)) = v^T(x(t) - \xi) \quad \text{auf} \quad I^* = [t_\alpha, t_\beta]. \quad (6.74)$$

Offenbar ist diese Funktion nach Voraussetzung 6.21 stetig differenzierbar mit $F(t_\alpha) < 0$, $F(t_\beta) > 0$ und $\dot{F}(t) = v^T \dot{x}(t) = v^T f(x(t)) > 0$ für hinreichend kleines $h > 0$ wegen der Transversalitätsbedingung $v^T f(\varphi(t^*; t_0, x_0)) > 0$. Damit erhalten wir das Newton-Verfahren zur Bestimmung des Durchstoß-Zeitpunktes t^* zu

$$t_{k+1} = t_k - \frac{F(t_k)}{\dot{F}(t_k)} = t_k - \frac{v^T(x(t_k) - \xi)}{v^T f(x(t_k))}, \quad k = 1, 2, 3, \dots$$

Als Startwert wählen wir $t_1 = t_\alpha$ mit dem zuvor berechneten Lösungspunkt $x_1 = x(t_\alpha)$. Die Hyperbene Σ legen wir durch einen Punkt ξ möglichst nahe bei $x(t^*)$, z. B. durch den Urbildpunkt $\xi = x_0$, und wählen den Normalvektor $v = f(\xi)/\|f(\xi)\|$, um so die Transversalitätsbedingung zu erfüllen:

Iteriere für $k = 1, 2, 3, \dots$, beginnend mit $t_1 = t_\alpha$, $x_1 = x(t_\alpha)$:

(1) Newton-Schritt:

$$t_{k+1} = t_k - \frac{f(\xi)^T (x_k - \xi)}{f(\xi)^T f(x_k)}$$

(2) Integrationsschritt: Integriere die DGL in $[t_k, t_{k+1}]$ mit einem Schritt

$$\Rightarrow x_{k+1} = \varphi(t_{k+1}; t_k, x_k)$$

$H(x(t_k)) = v^T(x(t_k) - \xi)$ gibt den Abstand des k -ten Näherungspunktes von der Hyperebene Σ an, während $v^T f(x(t_k))$ die Komponente des Vektorfeldes an $x(t_k)$ liefert, die in Richtung von Σ zeigt. Damit liefert die Newton-Korrektur $(v^T(x(t_k) - \xi))/(v^T f(x(t_k)))$ diejenige Zeit, die die Lösung benötigen würde, um von $x(t_k)$ zur Hyperebene zu gelangen, falls das Vektorfeld konstant mit dem Wert $f(x(t_k))$ wäre. Wegen der lokal quadratischen Konvergenz kommt das Newton-Verfahren so mit sehr wenigen Zusatzschritten aus.

2. *M. Hénon* entwickelte 1982 ein noch schnelleres Verfahren (vgl. Anishchenko (1987)), das meist mit einem einzigen Zusatzschritt zur Bestimmung von $x(t^*)$ auskommt. Der Abstand eines beliebigen Punktes $x \in \mathbb{R}^n$ von der eingeführten Hyperebene Σ ist wegen der Darstellung $x = x_\alpha + sv$

$$H(x) = v^T(x - \xi) = v^T(x_\alpha - \xi) + sv^T v = s.$$

Hénons Idee besteht darin, diesen Abstand s als neue unabhängige Variable anstelle von t mittels $s = s(t)$ einzuführen und die DGL zu transformieren. Für deren Lösung $x(t) = \varphi(t; t_0, x_0)$ liefert $s = s(t) = v^T(x(t) - \xi)$ die Ableitung $\dot{s}(t) = v^T f(x(t))$, womit sich die Ableitung der inversen Funktion $t = t(s)$ zu

$$t'(s) = \frac{1}{\dot{s}(t)} = \frac{1}{v^T f(x(t(s)))}$$

ergibt. Führen wir die Funktion $y(s) = x(t(s))$ ein, so transformieren wir sofort die DGL mittels

$$y'(s) = \dot{x}(t(s)) t'(s) = \frac{f(y(s))}{v^T f(y(s))}$$

auf die neue unabhängige Variable s . Zum Zeitpunkt $t = t_\alpha$ liegt der Lösungswert $x(t_\alpha)$ vor, womit wir unmittelbar die Anfangswerte

$$s_\alpha = H(x(t_\alpha)) = v^T(x(t_\alpha) - \xi) < 0 \quad \text{und} \quad y(s_\alpha) = x(t_\alpha)$$

gewinnen. Eine Integration des Systems für $(y, t) \in \mathbb{R}^{n+1}$ im Intervall $[s_\alpha, 0]$ ergibt dann die „Endwerte“ für den Abstand $s = 0$, d. h. den Durchstoß-Zeitpunkt $t(0) = t^*$ und den Bildpunkt $y(0) = x(t(0)) = x(t^*)$ in der Hyperebene Σ :

Gegeben sind t_α und $x_\alpha = x(t_\alpha)$ sowie die Hyperebene Σ mit (ξ, v) .

(1) Anfangswert: Bestimme $s_\alpha = v^T(x_\alpha - \xi) < 0$.

(2) Integrationsschritt: Integriere die $n + 1$ DGL in $[s_\alpha, 0]$

$$y'(s) = \frac{f(y(s))}{v^T f(y(s))}, \quad y(s_\alpha) = x_\alpha,$$

$$t'(s) = \frac{1}{v^T f(y(s))}, \quad t(s_\alpha) = t_\alpha.$$

(3) Endwerte: Durchstoß-Zeit $t^* = t(0)$ und Bildpunkt $x(t^*) = y(0)$.

Beispiel 6.36

Wir vergleichen die beiden numerischen Methoden an Beispiel 6.26 von W. F. Langford mit Parametern $\omega = 3.5$, $\varrho = 0.25$ und $\varepsilon = 0$. Als „Aufpunkt“ der Hyperebene wählen wir stets $\xi = (0, 0.9, 0.7)^T$ und verändern nur den Normalvektor v . Um den Zusatzaufwand zur Errechnung der Durchstoßpunkte zu ermitteln, integrieren wir die DGL mit 5000 Auswertungen der rechten Seiten (desweiteren als f -Bestimmungen bezeichnet).

Tab. 6.2 Zahl der f -Bestimmungen für die Poincaré-Abbildung zu Beispiel 6.36

Normalvektor v	Newton-Verfahren	Hénon-Verfahren
$(-3.15, 0, -0.092)^T$	1307 (26.13 %)	496 (9.92 %)
$(0, 0, -1)^T$	570 (11.41 %)	208 (4.16 %)
$(-1, 0, 0)^T$	1326 (26.51 %)	494 (9.87 %)

Tabelle 6.2 gibt diejenige Zahl von f -Bestimmungen an, die davon für die Berechnung der Poincaré-Abbildung benötigt werden. Obwohl beide Methoden zuverlässig und präzise arbeiten, erweist sich die Hénon-Methode stets als effizienter. ■

Weitere in der Literatur Parker und Chua (1989) empfohlene Ansätze, wie das Bisektionsverfahren zur Nullstellenbestimmung der Abstandsfunktion $F(t)$ aus (6.74) oder die Approximation der Durchstoßzeit t^* durch genaue Interpolation dieser Funktion sind den beiden vorgestellten Verfahren durchweg unterlegen und werden deshalb nicht empfohlen.

6.4 Lösungsfortsetzung und Bifurkationsanalyse

Wie schon in Kapitel 5 behandelt, soll nun das dynamische System von einem reellen Kontrollparameter $\lambda \in \Lambda = [a, b]$ abhängen. Wir wollen die bereits für Gleichgewichtslä-

gen eingeführten Fortsetzungstechniken auch zur Verfolgung periodischer Orbits einsetzen und zugleich die numerische Analyse von Stabilität und auftretenden Bifurkationen periodischer Orbits behandeln.

6.4.1 Numerische Fortsetzung und Stabilitätsanalyse

Wir betrachten nachfolgend parameterabhängige autonome Systeme in $D \subset \mathbb{R}^n$

$$\dot{x} = f(x, \lambda), \quad f : D \times \Lambda \rightarrow \mathbb{R}^n \quad \text{mit } \Lambda = [a, b], \quad D \text{ offen.} \quad (6.75)$$

Alle Betrachtungen lassen sich unmittelbar auf den numerisch einfacheren periodisch erregten Fall mit bekannter Lösungsperiode aus Abschnitt 6.1 übertragen, was nach dem Studium der folgenden Ausführungen gewiss leicht fallen dürfte. Zuerst notieren wir die Voraussetzungen 6.21 für den parameterabhängigen Fall:

Voraussetzung 6.37

Es sei $\Lambda_\varepsilon = (a - \varepsilon, b + \varepsilon) \supset \Lambda$ mit $\varepsilon > 0$.

- i. $f \in C^r(D \times \Lambda_\varepsilon)$, $r \geq 2$.
- ii. System (6.75) besitzt für alle $\lambda \in \Lambda_\varepsilon$ einen Grenzykel $x^*(t, \lambda)$, $t \in \mathbb{R}$, der Periode $T^*(\lambda) > 0$, der keine Gleichgewichtslage ist.

Wir wollen diesen Grenzykel $x^*(t, \lambda)$ samt seiner Periodendauer $T^*(\lambda)$ nun für alle Parameterwerte $\lambda \in \Lambda$ approximieren. Dazu überführen wir die Schwingungsaufgabe in ein Fortsetzungsproblem in \mathbb{R}^{n+1} . Ist $\lambda \in \Lambda$ ein fester Parameterwert, so können wir wie in Abschnitt 6.36 vorgehen. Zu gegebenen Näherungen $s_k \in D$, $T_k > 0$ betrachten wir das folgende transformierte AWP:

$$\left. \begin{aligned} \dot{x} &= Tf(x, \lambda), & x(0; s_k, T_k, \lambda) &= s_k \\ \dot{T} &= 0, & T(0; s_k, T_k, \lambda) &= T_k \end{aligned} \right\} \quad \text{für } t \in [0, 1] \quad (6.76)$$

mit der Lösung $x = x(t, s_k, T_k, \lambda)$, $T = T(t, s_k, T_k, \lambda)$ auf $[0, 1]$. Zusammen mit der Phasenbedingung (6.49) definieren wir die $n + 1$ Funktionen g und h durch

$$\begin{aligned} g(s_k, T_k, \lambda) &:= x(1; s_k, T_k, \lambda) - s_k \\ h(s_k, T_k, \lambda) &:= f(s_0, \lambda)^T (s_k - s_0), \end{aligned} \quad (6.77)$$

deren Nullstellen (s^*, T^*) mit einem Newton-ähnlichen Verfahren zu ermitteln sind. Wie im vorigen Abschnitt führen wir die Notation xx, ss, ff etc. für Vektoren in \mathbb{R}^{n+1} mit

$$xx := \begin{pmatrix} x \\ T \end{pmatrix}, \quad ss := \begin{pmatrix} s \\ T \end{pmatrix}, \quad ff(xx) := \begin{pmatrix} Tf(x, \lambda) \\ 0 \end{pmatrix} \quad \text{mit } ff : DD \subset \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{n+1}$$

ein und notieren das AWP (6.76) in der kompakten Form

$$\dot{xx} = ff(xx, \lambda), \quad xx(0, ss_k, \lambda) = ss_k \quad (6.78)$$

mit vorausgesetzter Lösung $xx = xx(t, ss_k, \lambda)$. Definieren wir nun die Abbildung gg durch

$$gg(ss_k, \lambda) := \begin{pmatrix} g(s_k, T_k, \lambda) \\ h(s_k, T_k, \lambda) \end{pmatrix} = \begin{pmatrix} x(1; s_k, T_k, \lambda) - s_k \\ f(s_0, \lambda)^T (s_k - s_0) \end{pmatrix} \quad (6.79)$$

so ergibt sich das zu lösende System als Fortsetzungsproblem in \mathbb{R}^{n+1}

$$gg(ss, \lambda) = 0, \quad gg : DD \times \Lambda \rightarrow DD, \quad DD \subset \mathbb{R}^{n+1}. \quad (6.80)$$

Gesucht ist eine Funktion $ss = ss(\lambda)$, $ss : \Lambda \rightarrow \mathbb{R}^{n+1}$, mit $gg(ss(\lambda), \lambda) = 0 \quad \forall \lambda \in \Lambda$. Diese Standardaufgabe lösen wir bequem mit den Fortsetzungsmethoden aus Kapitel 5.1, wobei wir uns auf die natürliche Parametrisierung bezüglich λ konzentrieren wollen. Wir passen dazu den Grundalgorithmus 5.28 unserer Problemstellung an:

1. *Prädiktorschritt:* Da die Ableitungen $D_{(ss)}gg(ss, \lambda)$ und $D_\lambda gg(ss, \lambda)$ nur mit großem Aufwand über das Variationssystem zu bestimmen sind, ist der Tangenten-Prädiktor (5.27) nicht zu empfehlen. Besser geeignet sind der
 - Basis-Prädiktor für den ersten Fortsetzungsschritt ($j = 1$):

$$ss^P := ss_{j-1} = ss(\lambda_{j-1}) \quad \text{und der} \quad (6.81)$$

- Sekanten-Prädiktor für die weiteren Fortsetzungsschritte ($j \geq 2$):

Mit den 2 berechneten Knoten $(ss_{j-2}, \lambda_{j-2})$ und $(ss_{j-1}, \lambda_{j-1})$ gewinnen wir

$$ss^P := ss_{j-1} + \frac{\lambda_j - \lambda_{j-1}}{\lambda_{j-1} - \lambda_{j-2}} (ss_{j-1} - ss_{j-2}). \quad (6.82)$$

2. *Korrektorschritt:* Mit dem Startpunkt ss^P lösen wir das erweiterte System (6.80) in \mathbb{R}^{n+1} mit einem Newton-ähnlichem Verfahren mit approximierter Jacobi-Matrix $D_{(ss)}gg(ss, \lambda)$. Dazu greifen wir auf Algorithmus 6.16 (Schießverfahren ohne Variationssystem) zurück und bauen dessen Schritt 1 komplett als Korrektor-Verfahren in den Algorithmus 6.38 zur *numerischen Fortsetzung periodischer Orbits im autonomen Fall* ein. Natürlich sind darin die Vektoren s_k, x, f, g durch die erweiterten Vektoren ss_k, xx, ff, gg zu ersetzen!

Durchführbarkeit und Konvergenz von Algorithmus 6.38 lassen sich – allerdings mit beträchtlichem theoretischen Aufwand – nachweisen. Erschwerend ist, dass die Voraussetzungen (z. B. Glattheit) für das Vektorfeld f , nicht aber für die Funktion gg in (6.80) vorliegen.

Satz 6.39

Voraussetzungen 6.37 seien erfüllt sowie folgende Annahmen:

- i. Die periodische Lösung $x^*(t, \lambda)$ ist nicht-entartet für alle $\lambda \in \Lambda_\varepsilon$.
- ii. Der Startwert $ss_0 = (s_0, T_0)$ von Algorithmus 6.38 liegt hinreichend nahe bei $ss(a) = (s(a), T(a))$, $a = \lambda_0$.

ALGORITHMUS 6.38 (Fortsetzung periodischer Orbits)

Function $[N, \lambda, ss, m] = \text{contperiodauto}(f, a, b, s_0, T_0, tol, kmax)$

1. Setze $j := 0$, $\lambda_0 := a$, $ss^P := ss_0 = (s_0, T_0)$.
Wähle Fortsetzungsschrittweite h sowie Iterationszahl k_{opt} .
2. (Startrechnung)
Bestimme $ss_0 = ss(\lambda_0) = (s(\lambda_0), T(\lambda_0))$ mit Alg. 6.16, Schritt 1.
3. Do while $\lambda_j \leq b$
 - 3.1. Setze $j := j + 1$.
 - 3.2. Setze $\lambda_j := \lambda_{j-1} + h$. Falls $\lambda_j > b$, so $\lambda_j := b$.
 - 3.3. (Prädiktor)
Bestimme ss^P mit den Formeln (6.81) und (6.82).
 - 3.4. (Korrektor)
Bestimme $ss_j = ss(\lambda_j)$ mit Algorithmus 6.16, Schritt 1,
in k Schritten bis auf eine Genauigkeit tol .
 - 3.5. (Schrittweiten-Bestimmung)
 - (a) Bestimme $\rho := k_{\text{opt}}/k$.
 - (b) Beschränke $\rho := \max\{\min(\rho, 2), \frac{1}{2}\}$.
 - (c) Setze $h := h * \rho$.
 - (d) Falls $h < \varepsilon_M$ (Maschinengenauigkeit), so STOP.
 - 3.6. (Monodromiematrix) Setze $GG := D_{(ss)}\mathcal{G}(ss_j, \lambda_j)$ und reduziere um die $(n+1)$ -te Zeile und Spalte $\Rightarrow G$. Setze $M(\lambda_j) := G + I_n$ und bestimme die Eigenwerte $m(\lambda_j) = (m_1(\lambda_j), m_2(\lambda_j), \dots, m_n(\lambda_j))$.
7. Return $N := j$, λ_j , ss_j , $m(\lambda_j)$, $j = 0(1)N$.

Dann existiert ein $h_{\text{max}} > 0$, so dass für alle Schrittweiten $h_j := \lambda_j - \lambda_{j-1} < h_{\text{max}}$ der Algorithmus durchführbar ist, d. h. zu $j = 0, 1, 2, \dots$ existieren λ_j und $ss(\lambda_j)$ mit

$$\mathcal{G}(ss(\lambda_j), \lambda_j) = 0, \quad j = 0, 1, 2, \dots,$$

wobei $ss(\lambda_j)$ reguläre Lösung von (6.80) ist. Der Algorithmus ist endlich, d. h. $\lambda_N = b$ für ein $N \in \mathbb{N}$.

Beweisidee: Es sei $\lambda_{j-1} \in \Lambda_\varepsilon$, $j \geq 1$ fest vorgegeben. Zur Lösung von $\mathcal{G}(ss, \lambda_{j-1}) = 0$ im Korrektorschritt verifiziert man die Voraussetzungen des Konvergenzsatzes des benutzten Newton-ähnlichen Verfahrens, d. h. $\mathcal{G} \in C^r(DD \times \Lambda_\varepsilon)$, $r \geq 2$ und $\mathcal{G}_{ss}(ss, \lambda)$ regulär für alle $(ss, \lambda) \in \mathcal{U}_\delta(ss(\lambda)) \times \Lambda_\varepsilon$. Ist die Fortsetzungsschrittweite h_j hinreichend klein, so kann garantiert werden, dass der Prädiktorwert ss^P in dieser Umgebung

$\mathcal{U}_\delta(ss(\lambda))$ liegt. Anwendung des Satzes 5.26 liefert eine hinreichende Schrittweiteschranke h_{\max} . Wählt man also die Startnäherung $ss_0 = (s_0, T_0)$ des ersten Fortsetzungsschrittes mit $ss_0 \in \mathcal{U}_\delta(ss(\lambda))$ und die konstante Schrittweite $h_j := \frac{1}{2}h_{\max}$, so konvergiert die Korrekturiteration in jedem Fortsetzungsschritt. Zudem erreicht man das rechte Intervallende b in endlich vielen Schritten. \square

Bemerkung 6.40

1. Für periodisch erregte parameterabhängige Systeme

$$\dot{x} = f(x, t, \lambda), \quad t \in [0, T], \quad \lambda \in \Lambda \quad (6.83)$$

lautet das zu (6.80) analoge Gleichungssystem als Fortsetzungsproblem in \mathbb{R}^n

$$g(s, \lambda) = 0, \quad g: D \times \Lambda \rightarrow D, \quad D \subset \mathbb{R}^n. \quad (6.84)$$

Algorithmus 6.38 und Konvergenzsatz 6.39 lassen sich dafür leicht anpassen.

2. Wie bereits für Gleichgewichtslagen in Kapitel 5.1 behandelt, sind auch zur Lösung von (6.80) weitere Parametrisierungen möglich. Empfohlen werden Parametrisierungen mittels

- Pseudo-Bogenlänge gemäß Algorithmus 5.34
- Gauß-Newton-Fortsetzung gemäß Algorithmus 5.38.

Die Angabe von Konvergenzbedingungen anhand der ursprünglich gegebenen Differenzialgleichungen $\dot{x} = f(x, \lambda)$ ist allerdings auch hier nicht trivial. \blacksquare

Zur numerischen Stabilitätsanalyse nutzen wir den Satz 6.28 von Andronov & Witt in jedem Fortsetzungsschritt und ergänzen Schritt 3 in Algorithmus 6.38 durch den Teilschritt:

3.7. (Stabilitätsanalyse) Es sei $m_k(\lambda_j) = 1$ der Hauptmultiplikator.

Setze $\varrho := \max\{|m_i(\lambda_j)| \mid i \neq k, i = 1(1)n\}$.

Falls $\varrho < 1$, so „ASYMPTOTISCH STABIL“,

sonst falls $\varrho > 1$, so „INSTABIL“,

sonst „KEINE AUSSAGE MÖGLICH“.

Bei periodisch erregten Systemen (6.83) ermitteln wir den Spektralradius ϱ über alle n charakteristischen Multiplikatoren $m_i(\lambda_j)$.

Beispiel 6.41

1. Für das 2-fach subharmonisch reagierende System von E. Philippow und W. Büntig aus Beispiel 6.9

$$\ddot{x} - \varepsilon(1 - x^2 - \dot{x}^2)\dot{x} + 2bx\dot{x} + x = \hat{B} \sin 2t, \quad b = \hat{B} = 1 \quad (6.85)$$

mit Kontrollparameter $\varepsilon \in \mathbb{R}$ konnte bereits eine 3-fach subharmonische Systemantwort ermittelt werden (vgl. Beispiel 6.19). Dazu nutzen wir die Parametrisierung mittels Pseudo-Bogenlänge und führen nun zusätzlich eine Stabilitätsanalyse durch. Der An-

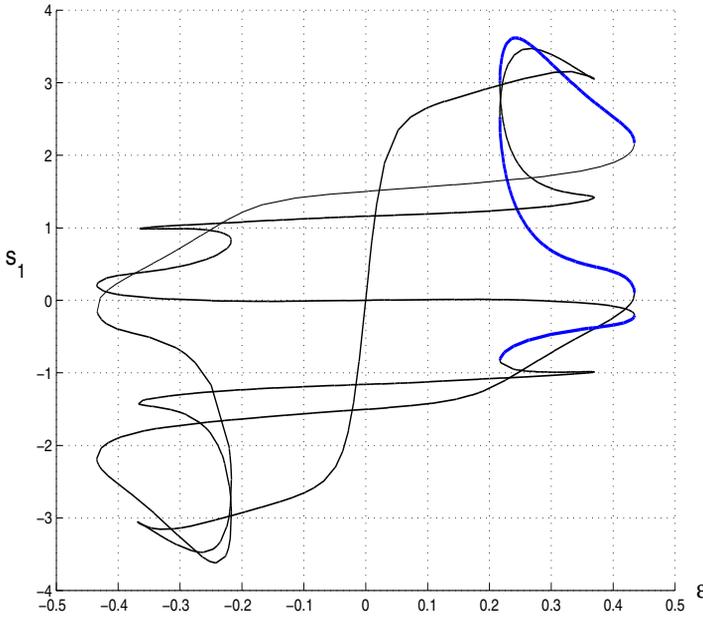


Abb. 6.19
 (ε, s_1) -Diagramm
 mit Stabilitätsverhalten der periodischen Lösungen

fangswert $s_1 = x_1(0)$ dieser 3π -periodischen Lösung ist in Abhängigkeit vom Parameter ε in Abbildung 6.19 dargestellt. Asymptotisch stabile Werte sind blau, während instabile Lösungen schwarz gezeichnet sind.

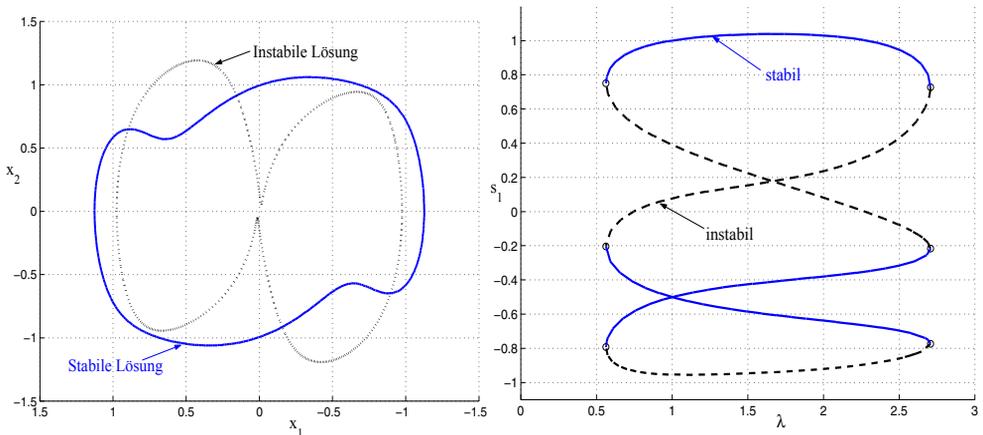


Abb. 6.20 2π -periodische Lösungen für $\lambda = 1.88$ (links) und (λ, s_1) -Lösungsdiagramm (rechts)

2. Philippow und Büntig (1992) entwickelten auch eine Modellgleichung eines dreifach subharmonisch reagierenden Systems der Elektrotechnik

$$\ddot{x} - \varepsilon(1 - x^2 - \dot{x}^2)\dot{x} + x + b(4x^3 - 3x) = \hat{B} \cos 3t \quad (6.86)$$

mit festen Parametern $b = 1.0$ und $\varepsilon = 3.0$, worin nun die Erregungsamplitude $\lambda := \hat{B}$ als Kontrollparameter dient. Neben $\frac{4}{3}\pi$ -periodischen Lösungen besitzt das System (mit $x_1 = x$, $x_2 = \dot{x}$) auch 2π -periodische Lösungen, z. B. für $\lambda = 1.88$ die in Abbildung 6.20 links dargestellte stabile und instabile Lösung. Eine Parameterverfol-

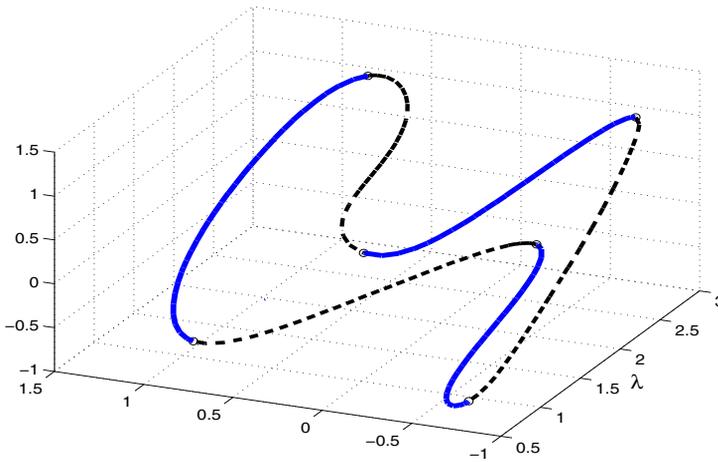


Abb. 6.21
 (λ, s_1, s_2) -Diagramm mit Stabilitätsverhalten der 2π -periodischen Lösungen (blau – stabil, schwarz – instabil)

gung des Anfangswertes $s_1 = x_1(0)$ dieser 2π -periodischen Lösung ist in Abhängigkeit vom Parameter λ in Abbildung 6.20 rechts dargestellt. Die geschlossene Lösungskurve ist übrigens doppeltpunktfrei, was man besser an ihrem 3D-Diagramm der Abbildung 6.21 im (λ, s_1, s_2) -Raum mit den Anfangswerten $s_1 = x_1(0)$, $s_2 = x_2(0)$ erkennen kann. ■

6.4.2 Bifurkationen periodischer Orbits

Wir betrachten nachfolgend parameterabhängige periodisch erregte Systeme

$$\dot{x} = f(x, t, \lambda), \quad (x, t, \lambda) \in \mathbb{R}^n \times \mathbb{S}^1 \times \mathbb{R}, \quad (6.87)$$

mit zugehörigem Fluss $\varphi_t(x; \lambda)$. Autonome Systeme (6.75) können analog zu vorigem Abschnitt untersucht werden. Jeder Parameterwert λ^* , für den der Fluss $\varphi_t(x; \lambda^*)$ nicht strukturell stabil ist, stellt gemäß Abschnitt 3.1 einen Bifurkationswert für λ dar.

Für jeden betrachteten Parameterwert $\lambda \in \Lambda = [a, b]$ wollen wir einen isolierten periodischen Orbit (Grenzykel) voraussetzen, der die Periode $T = 2\pi$ besitzt. Nach Abschnitt 6.3 existiert dazu eine λ -abhängige stroboskopische Abbildung als global definier-

te Poincaré-Abbildung $P : D \times \Lambda \rightarrow \mathbb{R}^n$ mit $D \subset \mathbb{R}^n$. Dem 2π -periodischen Orbit $\varphi_t(x^*(\lambda); \lambda)$ entspricht dann der Fixpunkt $x^*(\lambda)$ von

$$x = P(x, \lambda), \quad (x, \lambda) \in D \times \Lambda \subset \mathbb{R}^n \times \mathbb{R} \quad (6.88)$$

und umgekehrt. Das Stabilitätsverhalten von $x^*(\lambda)$ wird vollständig durch die n Eigenwerte $m_i(\lambda)$, $i = 1, 2, \dots, n$, der Poincaré-Matrix $D_x P(x^*(\lambda), \lambda)$, also durch die charakteristischen Multiplikatoren bestimmt.

Definition 6.42 (Hyperbolizität)

Bei festem λ heißt ein Fixpunkt $x^*(\lambda)$ von P bzw. eine periodische Lösung von (6.87) mit den n Eigenwerten $m_i(\lambda)$ von $D_x P(x^*(\lambda), \lambda)$

- *hyperbolisch*, falls $|m_i(\lambda)| \neq 1$ für $i = 1(1)n$,
- *nicht-entartet*, falls $m_i(\lambda) \neq 1$ für $i = 1(1)n$

gilt. Ein hyperbolischer Fixpunkt $x^*(\lambda)$ heißt

- *Senke*, falls $|m_i(\lambda)| < 1$ für $i = 1(1)n$,
- *Quelle*, falls $|m_i(\lambda)| > 1$ für $i = 1(1)n$,
- *Sattel* in allen anderen Fällen.

Notwendig für eine lokale Bifurkation an einem Parameterwert λ^* ist dann die Nicht-Hyperbolizität des Fixpunktes $x^* = x^*(\lambda^*)$, d. h. für ein $i \in \{1, \dots, n\}$ ist $|m_i(\lambda^*)| = 1$. Verschiebt man den Bifurkationspunkt (x^*, λ^*) in den Nullpunkt – das ist zumindest theoretisch leicht möglich – so lautet die Gleichung am Bifurkationswert nun $P(0, 0) = 0$. Wir erinnern daran, dass nach Guckenheimer und Holmes (1983), Kapitel 1.4., die Jacobi-Matrix $B = D_x P(0, 0)$ die invarianten (stabilen, zentralen und instabilen) Unterräume

- $E_s = \text{span}\{n_s \text{ verallgemeinerte Eigenvektoren zu Eigenwerten mit } |m_i(0)| < 1\}$
- $E_c = \text{span}\{n_c \text{ verallgemeinerte Eigenvektoren zu Eigenwerten mit } |m_i(0)| = 1\}$
- $E_u = \text{span}\{n_u \text{ verallgemeinerte Eigenvektoren zu Eigenwerten mit } |m_i(0)| > 1\}$

der Dimensionen $n_s \geq 0$, $n_c \geq 0$, $n_u \geq 0$ besitzt. Analog zum Fall der Fixpunkte von Flüssen existiert der folgende grundlegende *Reduktionssatz von A. N. Shoshitaishvili für Abbildungen*, den wir hier in der Darstellung von Reitmann (1996) zitieren:

Satz 6.43 (Reduktionssatz für Abbildungen)

P sei ein C^r -Diffeomorphismus mit $r \geq 1$ und $P(0, 0) = 0$. $D_x P(0, 0)$ habe invariante Unterräume E_s, E_c, E_u der Dimensionen $n_s \geq 0$, $n_c \geq 0$, $n_u \geq 0$ mit $n_s + n_c + n_u = n$. Dann gilt:

- i. P ist nahe $(0, 0)$ topologisch äquivalent dem Fixpunktproblem

$$\begin{aligned} u &= g(u, \lambda) = A_c u + h(u, \lambda), \\ v &= A_s v, \\ w &= A_u w \end{aligned} \quad (6.89)$$

mit $u \in \mathbb{R}^{n_c}$, $v \in \mathbb{R}^{n_s}$, $w \in \mathbb{R}^{n_u}$.

- ii. h ist eine C^r -Abbildung mit $h(0,0) = 0$ und $D_u h(0,0) = 0$.
- iii. A_c, A_s, A_u sind Matrizen der Dimensionen $n_c \times n_c$, $n_s \times n_s$ und $n_u \times n_u$ mit den zu E_c, E_s, E_u gehörenden Eigenwerten von $D_x P(0,0)$.

Um unnötige neue Bezeichnungen zu vermeiden, nennen wir den um λ^* verschobenen Parameter wiederum λ , während sich der Phasenraum-Vektor nun aus den drei Komponenten (u, v, w) zusammensetzt. Wegen der Definition 6.42 ist der Fixpunkt $(0,0)$ hyperbolisch, falls die Dimension $n_c = 0$ ist. Ist zusätzlich $n_u = 0$ ($n_s = 0$), so bildet $(0,0)$ eine Senke (Quelle). Lokale Bifurkationen sind deshalb nur möglich, wenn der zentrale Unterraum E_c nichttrivial ist. Man bezeichnet die Menge

$$W_{loc}^c(0) = \{(u, v, w) \mid v = 0, w = 0\}$$

als *lokale Zentrumsmannigfaltigkeit*. Die auf $W_{loc}^c(0)$ reduzierte Poincaré-Abbildung ergibt sich aus (6.89) als Fixpunktgleichung für $g: \mathbb{R}^{n_c} \times \Lambda \rightarrow \mathbb{R}^{n_c}$

$$u = g(u, \lambda) \quad \text{mit} \quad g(u, \lambda) = A_c u + h(u, \lambda). \quad (6.90)$$

Wegen der Bedingungen ii. des Satzes enthält $h(u, \lambda)$ den nichtlinearen Anteil von $g(u, \lambda)$.

Beispiel 6.44

Die parameterabhängige Poincaré-Abbildung $P: \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}^3$ mit

$$P(x, \lambda) = \begin{pmatrix} \lambda x_1 - \omega x_2 - x_1(x_1^2 + x_2^2) \\ \omega x_1 + \lambda x_2 - x_2(x_1^2 + x_2^2) \\ \alpha x_3 \end{pmatrix} \quad \text{und Konstanten} \quad \omega = 0.1, \alpha = 0.995$$

besitzt für alle reellen λ den Fixpunkt $x^*(\lambda) = (0, 0, 0)^T$, so dass keine x -Verschiebung nötig ist. Setzen wir $\lambda^* = \sqrt{1 - \omega^2} = \sqrt{99}/10 = 0.9949874371\dots$, so hat nach Parameterverschiebung die Poincaré-Abbildung

$$P(x, \lambda) = \begin{pmatrix} (\lambda + \lambda^*)x_1 - \omega x_2 - x_1(x_1^2 + x_2^2) \\ \omega x_1 + (\lambda + \lambda^*)x_2 - x_2(x_1^2 + x_2^2) \\ \alpha x_3 \end{pmatrix} \quad \text{mit} \quad \lambda^* = 0.9949874371\dots$$

den Fixpunkt $0 = P(0,0)$. Die zugehörige Jacobi-Matrix und ihre Eigenwerte sind

$$B = D_x P(0,0) = \begin{pmatrix} \lambda^* & -\omega & 0 \\ \omega & \lambda^* & 0 \\ 0 & 0 & \alpha \end{pmatrix} \quad \begin{array}{l} m_1(0) = \lambda^* + i\omega \\ m_2(0) = \lambda^* - i\omega \\ m_3(0) = \alpha. \end{array}$$

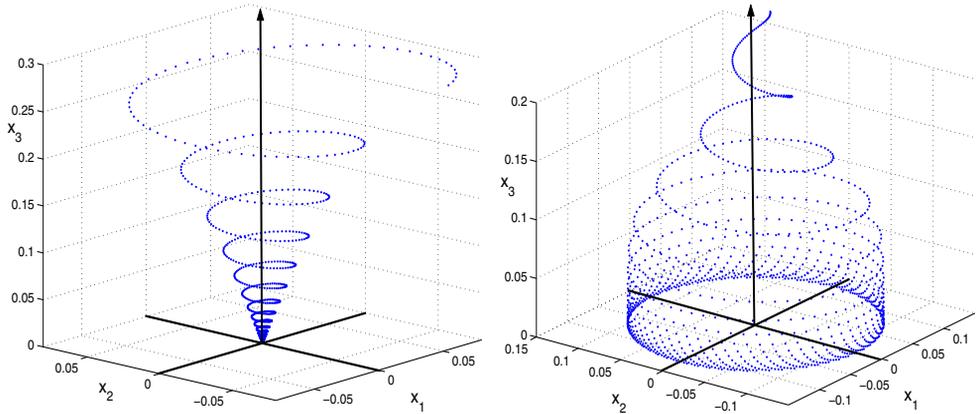


Abb. 6.22 Abbildungspunkte (x^k) des Beispiels 6.44 für $\lambda = -0.004987$ (links) und $\lambda = 0.015013$ (rechts)

Wegen $|m_{1,2}(0)| = 1$ und $|m_3(0)| = 0.995 < 1$ besitzen die invarianten Unterräume die Dimensionen $n_c = \dim E_c = 2$, $n_s = \dim E_s = 1$ und $n_u = \dim E_u = 0$. In Abbildung 6.22 werden Punktfolgen (x^k) , $k = 0, 1, 2, \dots$ mit $x^k = P(x^{k-1}, \lambda)$ für die Parameterwerte $\lambda = -0.004987$ (links) und $\lambda = 0.015013$ (rechts) dargestellt. Die

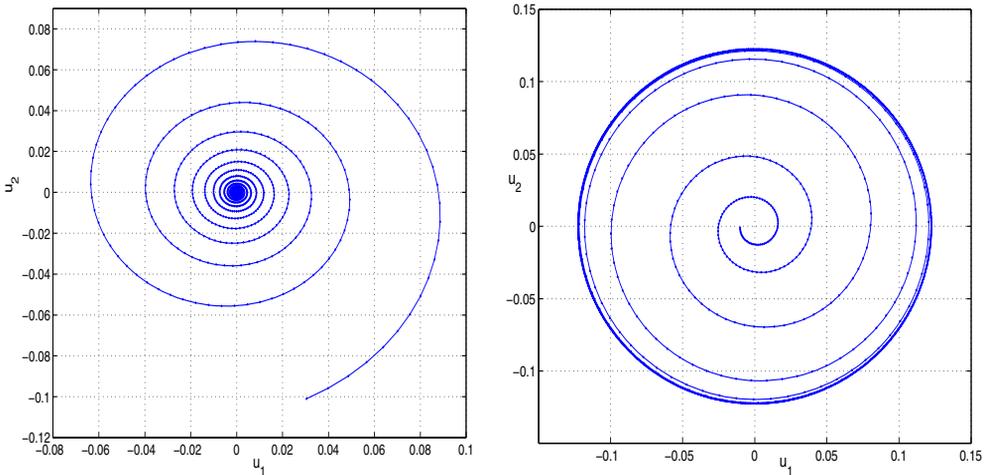


Abb. 6.23 Abbildungspunkte $(u^k) \in E_c$ des Beispiels 6.44 für $\lambda = -0.004987$ (links) und $\lambda = 0.015013$ (rechts)

Zerlegung (6.89) liegt mit den Bezeichnungen $u_1 = x_1$, $u_2 = x_2$, $v_1 = x_3$ bereits vor, weshalb die auf E_c reduzierte Poincaré-Abbildung nun

$$u = g(u, \lambda) \quad \text{mit} \quad g(u, \lambda) = \begin{pmatrix} (\lambda + \lambda^*)u_1 - \omega u_2 - u_1(u_1^2 + u_2^2) \\ \omega u_1 + (\lambda + \lambda^*)u_2 - u_2(u_1^2 + u_2^2) \end{pmatrix}$$

lautet. Wegen $g(u, \lambda) = A_c u + h(u, \lambda)$ mit der Matrix

$$A_c = \begin{pmatrix} \lambda^* & -\omega \\ \omega & \lambda^* \end{pmatrix} \quad \text{und} \quad h(u, \lambda) = \begin{pmatrix} \lambda u_1 - u_1(u_1^2 + u_2^2) \\ \lambda u_2 - u_2(u_1^2 + u_2^2) \end{pmatrix}$$

sind alle Bedingungen des Reduktionssatzes erfüllt. Die Punktfolgen (u^k) , $k = 0, 1, 2, \dots$ der auf E_c reduzierten Poincaré-Abbildung sieht man in Abbildung 6.23. ■

In praktischen Aufgabenstellungen ist eine derartige Reduktion auf die Zentrumsmannigfaltigkeit in der Regel nicht angebar oder wegen der Problemgröße nicht durchführbar. Als einziges praktikables Hilfsmittel dient dann die Beobachtung der charakteristischen Multiplikatoren $m_i(\lambda)$, $i = 1(1)n$, der Poincaré-Abbildung $P(x, \lambda)$ bei Änderung des Kontrollparameters λ . Nicht-Hyperbolizität des Fixpunktes $x^*(\lambda)$ wird für einen Parameterwert λ^* offenbar in folgenden drei *generischen Fällen* eintreten (Zum Begriff der Generizität sei auf Reitmann (1996) verwiesen):

- $m_i(\lambda^*) = +1 \quad \implies \quad$ (reelle) Fold-Bifurkation
- $m_i(\lambda^*) = -1 \quad \implies \quad$ (reelle) Flip-Bifurkation
- $m_i(\lambda^*) \overline{m_i(\lambda^*)} = 1, m_i(\lambda^*) \notin \mathbb{R} \quad \implies \quad$ (komplexe) Hopf-Bifurkation.

Man beachte dabei, dass nicht-reelle Eigenwerte stets als konjugiert komplexe Paare auftreten. Das unterschiedliche Verhalten in diesen drei Fällen wird klar, wenn wir nun die auf $W_{loc}^c(0)$ reduzierte Poincaré-Abbildung (6.100) untersuchen. Wir benutzen hierbei die Notation aus Reitmann (1996) und die Darstellungsweise von Anishchenko (1987).

(1a) Fold-Bifurkation (Sattel-Knoten-Bifurkation)

Taylor-Entwicklung der Funktion $g(x, \lambda)$ aus (6.89) ergibt $g(x, \lambda) = x + \alpha\lambda + \beta x^2 + \dots$ mit

$$g(0, 0) = 0, \quad \frac{\partial g}{\partial x}(0, 0) = m_i(0) = 1, \quad \frac{\partial g}{\partial \lambda}(0, 0) = \alpha \neq 0, \quad \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(0, 0) = \beta \neq 0.$$

In diesem ersten Fall setzen wir voraus, dass $\alpha \neq 0$ und $\beta \neq 0$. Ist $\alpha\beta < 0$, so hat $P(x, \lambda)$ für $\lambda > 0$ zwei Fixpunkte, für $\lambda = 0$ einen Fixpunkt und für $\lambda < 0$ keinen Fixpunkt. Bei $(0, 0)$ tritt eine *superkritische Fold-Bifurkation* ein. Ist $\alpha\beta > 0$, so hat $P(x, \lambda)$ für $\lambda < 0$ zwei Fixpunkte, für $\lambda = 0$ einen Fixpunkt und für $\lambda > 0$ keinen Fixpunkt, weshalb hier eine sogenannte *subkritische Fold-Bifurkation* eintritt. In Analogie zu den Gleichgewichtslagen aus Kapitel 5 nennt man diese Verzweigung oft auch *Sattel-Knoten-Bifurkation*. Sie stellt den generischen Fall dar, wenn ein Multiplikator $m_i(0) = 1$ algebraisch einfacher Eigenwert von $D_x P(0, 0)$ ist.

Beispiel 6.45

Wir betrachten die in den Ursprung $(0, 0)$ transformierte Abbildung

$$g(x, \lambda) = x + \lambda - x^2 \quad \text{mit} \quad (x^*, \lambda^*) = (0, 0).$$

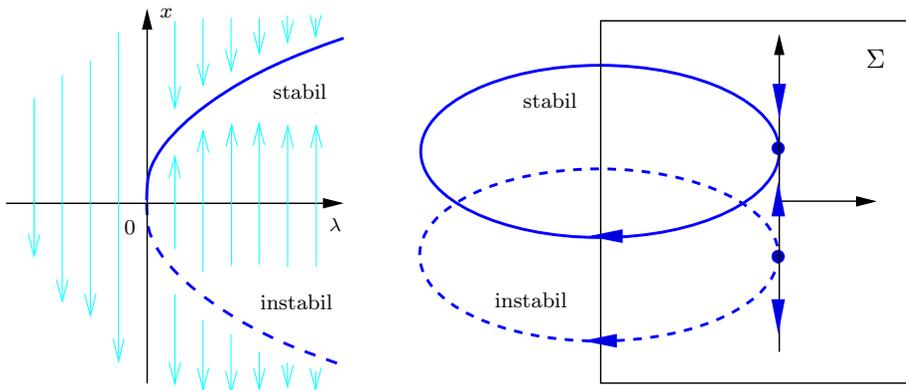


Abb. 6.24 Superkritische Fold-Bifurkation (links) und periodische Lösungen für $\lambda > \lambda^*$ (rechts)

Offenbar ist hier $\alpha = 1$, $\beta = -1$, weshalb bei (x^*, λ^*) eine superkritische Fold-Bifurkation vorliegt (vgl. Abb. 6.24 links). Für Parameterwerte $\lambda > \lambda^*$ liegen zwei periodische Lösungen vor, die in Abbildung 6.24 (rechts) mit ihren Poincaré-Abbildungen veranschaulicht werden. ■

(1b) Transkritische Bifurkation

Taylor-Entwicklung von $g(x, \lambda)$ liefert nun $g(x, \lambda) = x + \alpha\lambda x + \beta x^2 + \dots$ mit

$$g(0,0) = 0, \quad \frac{\partial g}{\partial x}(0,0) = 1, \quad \frac{\partial g}{\partial \lambda}(0,0) = 0, \quad \frac{\partial^2 g}{\partial \lambda \partial x}(0,0) = \alpha \neq 0, \quad \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(0,0) = \beta \neq 0.$$

Hier tritt bei $(0,0)$ eine *transkritische Bifurkation* ein, die häufig mit einem Stabilitätswechsel verbunden ist. Im Gegensatz zu Fall (1a) tritt diese nicht-generische Bifurkation höherer Ordnung seltener in praktischen Anwendungen auf.

Beispiel 6.46

Die in den Ursprung $(0,0)$ transformierte Abbildung laute

$$g(x, \lambda) = x + \lambda x - x^2 - \lambda x^2 \quad \text{mit} \quad (x^*, \lambda^*) = (0,0).$$

Wir stellen fest, dass $\alpha = 1$, $\beta = -1$ ist und so bei (x^*, λ^*) eine transkritische Bifurkation mit Stabilitätswechsel eintritt, die in Abbildung 6.25 (links) dargestellt ist. ■

(1c) Heugabel-Bifurkation

Taylor-Entwicklung von $g(x, \lambda)$ in (6.89) ergibt nun $g(x, \lambda) = x + \alpha\lambda x + \beta x^3 + \dots$ mit

$$g(0,0) = 0, \quad \frac{\partial g}{\partial x}(0,0) = 1, \quad \frac{\partial g}{\partial \lambda}(0,0) = 0 \quad \text{sowie}$$

$$\frac{\partial^2 g}{\partial \lambda \partial x}(0,0) = \alpha \neq 0, \quad \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(0,0) = 0, \quad \frac{1}{6} \frac{\partial^3 g}{\partial x^3}(0,0) = \beta \neq 0.$$

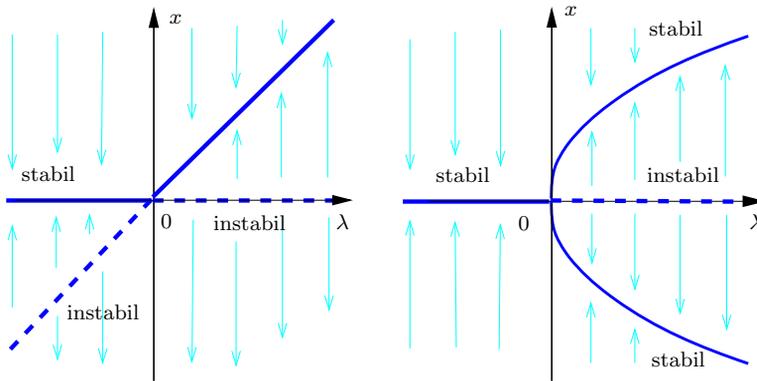


Abb. 6.25 Transkritische Bifurkation (links) und Heugabel-Bifurkation (rechts)

Am Bifurkationspunkt $(x^*, \lambda^*) = (0, 0)$ liegt in diesem Falle eine *Heugabel-Bifurkation* (*pitchfork bifurcation*) vor, bei der ebenfalls der superkritische vom subkritischen Fall unterschieden wird. Diese nicht-generische Bifurkation höherer Ordnung tritt in praktischen Anwendungen mitunter auf, besonders wenn das dynamische System Symmetrien enthält.

Beispiel 6.47

Die transformierte Abbildung sei nun

$$g(x, \lambda) = x + \lambda x - x^3 \quad \text{mit} \quad (x^*, \lambda^*) = (0, 0).$$

Auch hier lauten die Koeffizienten $\alpha = 1$, $\beta = -1$, weshalb bei (x^*, λ^*) die in Abbildung 6.25 rechts dargestellte superkritische Heugabel-Bifurkation vorliegt. Bei wachsendem λ „übergibt“ die triviale Lösung $x^* = 0$ ihre Stabilität bei $\lambda^* = 0$ an die abzweigende Lösung. ■

(2) Flip-Bifurkation (Periodenverdopplung, subharmonische Bifurkation)

Wir nehmen nun mit $m_i(0) = -1$ die Taylor-Entwicklung von $g(x, \lambda)$ in (6.89) vor:

$$g(x, \lambda) = -x + \alpha\lambda x + \beta x^2 + \gamma x^3 + \dots$$

mit

$$g(0, 0) = 0, \quad \frac{\partial g}{\partial x}(0, 0) = m_i(0) = -1 \quad \text{sowie}$$

$$\frac{\partial^2 g}{\partial \lambda \partial x}(0, 0) = \alpha, \quad \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(0, 0) = \beta, \quad \frac{1}{3!} \frac{\partial^3 g}{\partial x^3}(0, 0) = \gamma.$$

In diesem zweiten generischen Fall setzen wir voraus, dass $\alpha \neq 0$ und $\beta^2 + \gamma \neq 0$ ist. Dann gilt: Ist $\alpha(\beta^2 + \gamma) < 0$, so hat $P(x, \lambda)$ für $\lambda < 0$ genau einen Fixpunkt und für $\lambda > 0$ genau einen Fixpunkt sowie zwei Fixpunkte x^{**} der Periode 2, d. h. $x^{**} = P(P(x^{**}, \lambda), \lambda)$. Bei Durchlaufen des Bifurkationspunktes $(0, 0)$ tritt eine *periodenverdoppelnde Bifurkation* (*flip bifurcation*) ein. Ist $\alpha(\beta^2 + \gamma) > 0$, so hat $P(x, \lambda)$ für $\lambda > 0$ genau einen Fixpunkt und für $\lambda < 0$ einen Fixpunkt sowie zwei Fixpunkte x^{**} der Periode 2.

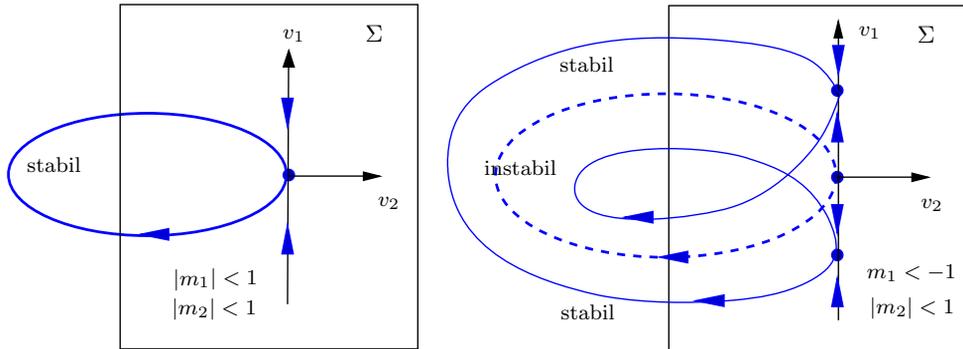


Abb. 6.26 Flip-Bifurkation: Eine T -periodische Lösung für $\lambda < \lambda^*$ (links) und zusätzlich eine $2T$ -periodische Lösung für $\lambda > \lambda^*$ (rechts)

Beispiel 6.48

Die in den Ursprung $(0,0)$ transformierte Abbildung lautet

$$g(x, \lambda) = -x - \lambda x + x^3 \quad \text{mit} \quad (x^*, \lambda^*) = (0, 0).$$

Wegen $\alpha = -1$, $\beta = 0$ und $\gamma = 1$ ist $\alpha(\beta^2 + \gamma) = -1 < 0$, weshalb bei (x^*, λ^*) eine Flip-Bifurkation eintritt. Abbildung 6.26 zeigt links die stabile periodische Lösung für Parameterwerte $\lambda < \lambda^*$, während rechts die nunmehr instabile Lösung sowie die Lösung mit doppelter Periode $2T$ dargestellt ist. Man erkennt leicht, dass die Poincaré-Abbildung nun zwei Fixpunkte der Periode 2 besitzt. ■

Betrachten wir das System von W. F. Langford aus Beispiel 5.1

$$\begin{aligned} \dot{x}_1 &= (x_3 - 0.7)x_1 - \omega x_2 \\ \dot{x}_2 &= \omega x_1 + (x_3 - 0.7)x_2 \\ \dot{x}_3 &= 0.6 + x_3 - x_3^3/3 - (x_1^2 + x_2^2)(1 + \varrho x_3) + \varepsilon x_3 x_1^3. \end{aligned} \quad (6.91)$$

Mit den Konstanten $\omega = 3.5$, $\varrho = 0.25$, $\beta = 0.7$ und dem Kontrollparameter ε erhalten wir die in Abbildung 6.27 dargestellten Lösungen. Zwischen $\varepsilon = 0.03$ und $\varepsilon = 0.06$ tritt eine erste Flip-Bifurkation auf; die nächsten Periodenverdopplungen einer Flip-Bifurkationskaskade liegen vor den Werten $\varepsilon = 0.0675$ und $\varepsilon = 0.07$.

(3) Torus-Bifurkation (Neimark-Sacker-Bifurkation)

Dieser Bifurkationstyp wird mitunter als *Hopf-Bifurkation für Poincaré-Abbildungen* bezeichnet. Das reduzierte System $x = g(x, \lambda)$ in (6.100) ist nun 2-dimensional. Die Jacobi-Matrix $D_x P(x^*, \lambda^*)$ hat ein Paar konjugiert komplexer Eigenwerte $m_i(\lambda^*)$ und $\overline{m_i(\lambda^*)}$ mit $|m_i(\lambda^*)| = 1$ und $n - 2$ charakteristische Multiplikatoren, die nicht auf dem Einheitskreis liegen. Es gilt das folgende *Bifurkations-Theorem von Neimark-Sacker für periodische Orbits* (vgl. Reitmann (1996)):

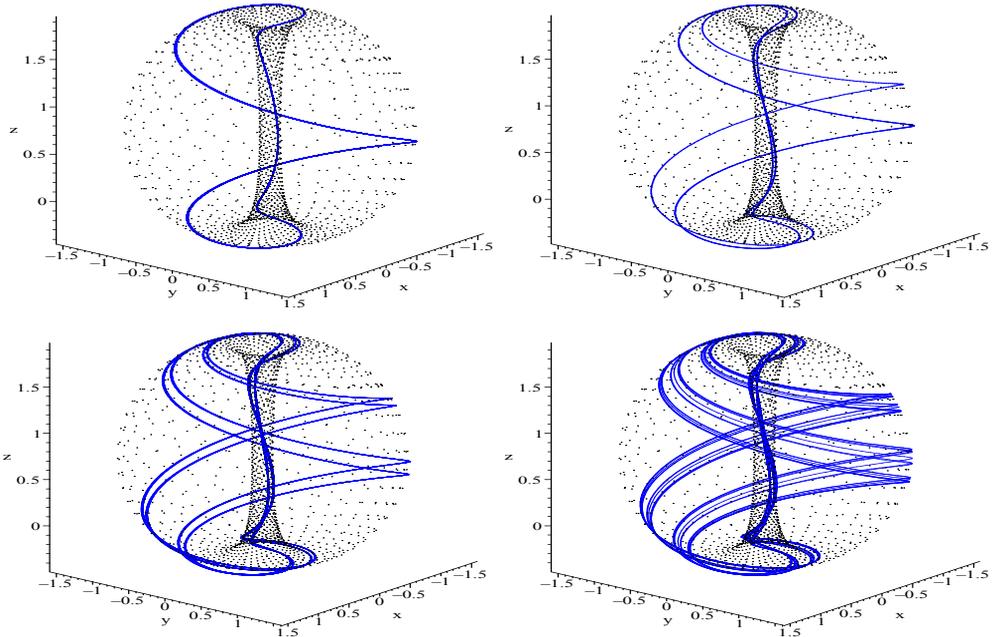


Abb. 6.27 Langford-System des Beispiels 6.26: Parameterwerte sind $\varepsilon = 0.03$ (links oben), $\varepsilon = 0.06$ (rechts oben), $\varepsilon = 0.0675$ (links unten) und $\varepsilon = 0.07$ (rechts unten)

Satz 6.49 (Neimark & Sacker)

Folgende Voraussetzungen seien erfüllt:

- i. $D_x P(0, 0)$ habe zwei Eigenwerte $m_1 = m(0)$ und $m_2 = \overline{m(0)}$ mit $|m(0)| = 1$ und $n - 2$ Eigenwerte $|m_i(0)| \neq 1, i = 3, 4, \dots, n$.
- ii. $g \in C^6$ und $g(0, \lambda) = 0, \lambda \in \mathcal{U}(0)$.
- iii. $D_x g(0, \lambda)$ hat die Eigenwerte $m(\lambda)$ und $\overline{m(\lambda)}$ für alle $\lambda \in \mathcal{U}(0)$ mit

$$d := \frac{d}{d\lambda} |m(\lambda)|_{\lambda=0} \neq 0. \tag{6.92}$$

- iv. $m(0)$ ist keine j -te Wurzel aus 1 mit $j < 5$, d. h. $m(0)^j \neq 1$ für $j = 1, 2, 3, 4$.

Dann gelten folgende Behauptungen:

- i. Es existiert eine glatte Transformation in Polarkoordinaten, die (6.89) nahe $(0, 0)$ in die Form

$$\begin{pmatrix} r \\ \theta \end{pmatrix} \mapsto \begin{pmatrix} r + d\lambda r + \sigma r^3 + \dots \\ \theta + |\arg m(0)| + br^2 + \dots \end{pmatrix} \tag{6.93}$$

mit reellen Parametern b, σ überführt.

- ii. Ist $\sigma \neq 0$, so tritt bei $\lambda^* = 0$ eine *Torus-Bifurkation* für P ein. Für $\lambda < 0$ existiert der triviale Fixpunkt $x^* = 0$, während für $\lambda > 0$ in der Umgebung dieses Fixpunktes eine geschlossene Invarianzkurve $\Gamma \subset \mathbb{R}^n$ für die Abbildung P liegt.

iii. Ist $\sigma < 0$, so heißt die Torus-Bifurkation für $\lambda > 0$ superkritisch, bei $\sigma > 0$ liegt eine subkritische Torus-Bifurkation für $\lambda > 0$ vor.

Das komplizierte Verhalten der Lösungen für $\lambda > 0$ wird durch Abbildung 6.28 veranschaulicht: Die bei $x \in \Gamma$ startende Lösung windet sich um die instabile triviale Lösung

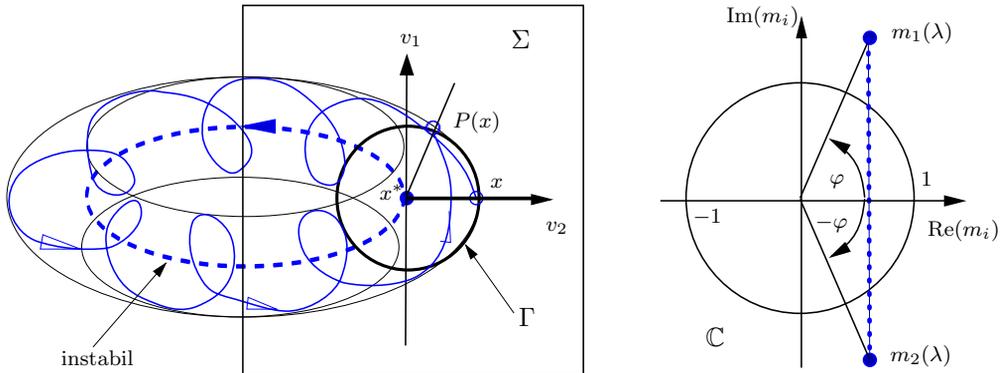


Abb. 6.28 Torus-Bifurkation für $\lambda > \lambda^*$: Invarianter Torus und Invarianzkurve $\Gamma \subset \Sigma$ (links) sowie Multiplikatoren $m_{1,2}(\lambda)$ (rechts)

$x^* = 0$ und liegt auf einem invarianten 2-dimensionalen Torus (daher auch die Bezeichnung „Torus-Bifurkation“). Der Poincaré-Schnitt dieser Torusfläche ist die im Satz postulierte Invarianzkurve $\Gamma \subset \Sigma$. Die Rotationsgeschwindigkeit der Lösung wird für kleines r durch den Winkel $\varphi = |\arg m(0)|$ aus (6.93) bestimmt (vgl. Abb. 6.28 rechts).

Beispiel 6.50

Es sei $x = (x_1, x_2) \in \mathbb{R}^2$ und die Poincaré-Abbildung $P : \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^2$

$$P(x, \lambda) = \begin{pmatrix} sx_1 - \omega x_2 - cx_1(x_1^2 + x_2^2) \\ \omega x_1 + sx_2 - cx_2(x_1^2 + x_2^2) \end{pmatrix}, \quad s, \omega, c \in \mathbb{R}, \quad c > 0. \tag{6.94}$$

Setzen wir $s = \omega = \frac{1}{2}\sqrt{2}(1+\lambda)$, so ist die Gleichgewichtslage $x^* = (0, 0)$ für $\lambda < 0$ asymptotisch stabil und wird für $\lambda > 0$ instabil. Verifizieren wir deshalb die Voraussetzungen des Satzes 6.49 mit $\lambda^* = 0$. Die Eigenwerte der Jacobi-Matrix

$$D_x P(0, \lambda) = \begin{pmatrix} s & -\omega \\ \omega & s \end{pmatrix} \quad \text{sind} \quad \begin{cases} m_1(\lambda) = s + i\omega = \frac{1}{2}\sqrt{2}(1+\lambda)(1+i) \\ m_2(\lambda) = s - i\omega = \frac{1}{2}\sqrt{2}(1+\lambda)(1-i). \end{cases}$$

Mit $|m_{1,2}(\lambda)| = 1 + \lambda$ ist Voraussetzung **i.** erfüllt. Offenbar ist die reduzierte Abbildung $g \equiv P$, womit auch **ii.** gilt. $D_x g(0, \lambda)$ hat damit die Eigenwerte $m(\lambda) = \frac{1}{2}\sqrt{2}(1+\lambda)(1+i)$ und $\overline{m(\lambda)}$ für alle $\lambda \in \mathbb{R}$, welche die *Transversalitätsbedingung* **iii.** mit

$$d = \frac{d}{d\lambda} |m(\lambda)|_{\lambda=0} = 1$$

erfüllen. Wegen $\varphi = |\arg m(0)| = \pi/4$ gilt schließlich auch Voraussetzung **iv**. Gehen wir nun zu Polarkoordinaten (r, θ) mit

$$x_1 = r \cos \theta, \quad x_2 = r \sin \theta \quad \text{bzw. mit} \quad z = x_1 + ix_2 = re^{i\theta}$$

über, so transformiert sich die Poincaré-Abbildung (6.94) in der Umgebung von $(0, 0)$ in

$$\begin{pmatrix} r \\ \theta \end{pmatrix} \mapsto \begin{pmatrix} (1 + \lambda)r - \frac{1}{2}\sqrt{2}cr^3 + \mathcal{O}(r^5) \\ \theta + \frac{\pi}{4} + \frac{c}{\sqrt{2}(1 + \lambda)}r^2 + \mathcal{O}(r^4) \end{pmatrix}. \quad (6.95)$$

Dabei wurden die entstehenden Funktionen in Taylor-Reihen nach r entwickelt und die Restterme zusammengefasst. Der Vergleich mit der Darstellung in Formel (6.93) liefert den Parameterwert $\sigma = -\frac{1}{2}\sqrt{2}c < 0$, weshalb der Satz von Neimark-Sacker eine *superkritische Torus-Bifurkation* für $\lambda^* = 0$ garantiert. ■

Wenn bei der Beschreibung der drei Bifurkationsszenarien vom „Übergang asymptotisch stabiler in instabile Lösungen“ die Rede ist, so trifft dies nach dem Reduktionssatz 6.43 offenbar nur zu, wenn die Dimension n_u des instabilen Unterraumes E_u gleich Null ist. Dies ist der praktisch interessanteste Fall, da nur asymptotisch stabile Lösungen physikalisch sind, also in der Praxis auftreten. Selbstverständlich können auch instabile Lösungen bifurkieren und in instabile Lösungen übergehen, die beim Studium dynamischer Systeme bedeutsam sind.

Beispiel 6.51

Das autonome System von W. F. Langford aus Beispiel 5.1

$$\begin{aligned} \dot{x}_1 &= (x_3 - 0.7)x_1 - \omega x_2 \\ \dot{x}_2 &= \omega x_1 + (x_3 - 0.7)x_2 \\ \dot{x}_3 &= 0.6 + x_3 - x_3^3/3 - (x_1^2 + x_2^2)(1 + \varrho x_3) + \varepsilon x_3 x_1^3 \end{aligned} \quad (6.96)$$

liefert mit den Konstanten $\omega = 3.5$, $\varepsilon = 0$, $\beta = 0.7$ und dem Kontrollparameter ϱ die in Abbildung 6.29 dargestellten Lösungen. Zwischen $\varrho = 0.70$ und $\varrho = 0.60$ tritt offenbar eine Torus-Bifurkation ein. Den genauen Bifurkationswert $\varrho^* = 0.61544465\dots$ ermitteln wir mit den nachfolgend behandelten Verfahren. ■

6.4.3 Detektierung lokaler Bifurkationen

Wir betrachten desweiteren parameterabhängige autonome Systeme mit $n + 1$ DGL

$$\dot{x} = f(x, \lambda), \quad f : D \times \Lambda \rightarrow \mathbb{R}^{n+1} \quad \text{mit} \quad D \subset \mathbb{R}^{n+1}, \quad \Lambda = [a, b] \subset \mathbb{R}, \quad (6.97)$$

die wir mit Algorithmus 6.38 numerisch fortsetzen. Mit der Wahl \mathbb{R}^{n+1} lassen sich die folgenden Betrachtungen stets in \mathbb{R}^n durchführen. Periodisch erregte Systeme (6.87) können in \mathbb{R}^n analog dazu untersucht werden. Während der Lösungsfortsetzung soll das

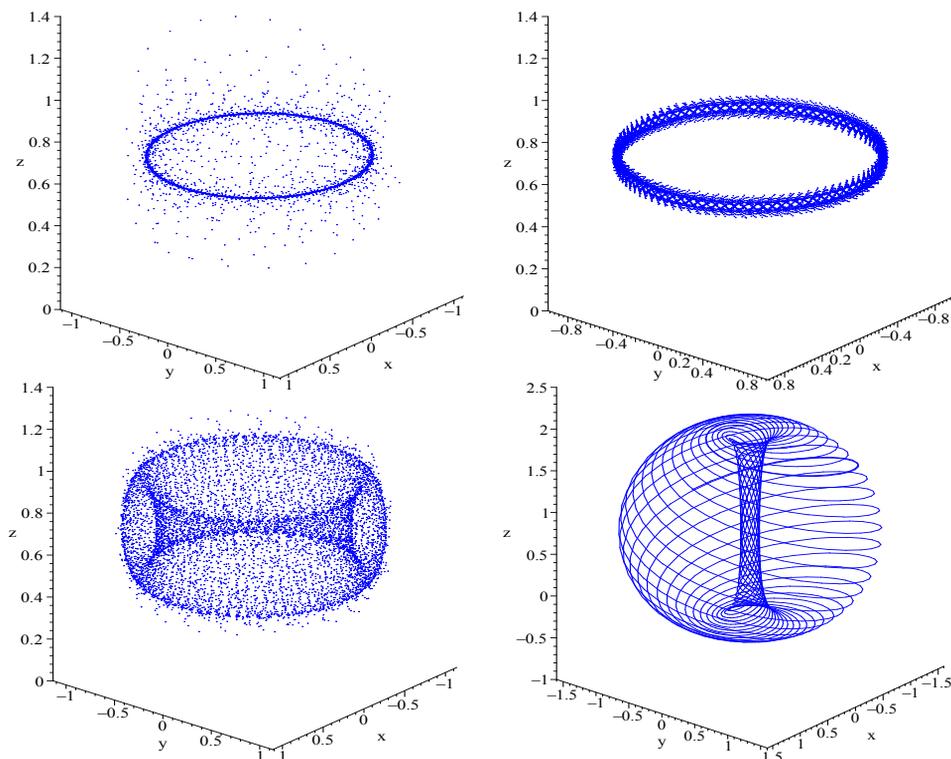


Abb. 6.29 Langford-System des Beispiels 6.26: Parameterwerte sind $\varrho = 0.70$ (links oben), $\varrho = 0.62$ (rechts oben), $\varrho = 0.60$ (links unten) und $\varrho = 0.25$ (rechts unten)

Auftreten der lokalen Bifurkationen des vorigen Abschnitts detektiert werden, um die kritischen Parameterwerte einzugrenzen und numerisch zu bestimmen. Eine direkte algorithmische Umsetzung des Reduktionssatzes 6.43 erscheint höchst unrealistisch, da in jedem Fortsetzungsschritt die auf die lokale Zentrumsmannigfaltigkeit $W_{loc}^c(0)$ reduzierte Poincaré-Abbildung approximiert werden müsste. Gerade bei höherdimensionalen Systemen ist der Aufwand deshalb nicht vertretbar und sollte durch einfachere notwendige Bedingungen ersetzt werden.

Neben der periodischen Lösung $x(t, \lambda)$ mit Anfangswert $s(\lambda) = x(0, \lambda)$ und Periode $T(\lambda)$ approximiert Algorithmus 6.38 auch deren Floquet-Multiplikatoren $m_1(\lambda), m_2(\lambda), \dots, m_n(\lambda), m_{n+1}(\lambda)$. Im vorliegenden autonomen Fall ist der Hauptmultiplikator $m_{n+1}(\lambda) = 1$ zu eliminieren. Anschließend prüfen wir, ob ein Multiplikator $m_i(\lambda)$ mit $i \in \{1, 2, \dots, n\}$ existiert, der eine der folgenden notwendigen Bedingungen erfüllt:

- $m_i(\lambda) = +1$ für eine Fold-, Pitchfork- oder transkritische Bifurkation
- $m_i(\lambda) = -1$ für eine Flip-Bifurkation
- $|m_i(\lambda)| = 1, m_i(\lambda) \notin \mathbb{R}$ für eine Torus-Bifurkation.

Wir führen dazu wie in Kapitel 5 eine skalare Testfunktion ψ ein (vgl. dazu Definition 5.55). Mit $C = (s(\lambda), T(\lambda), \lambda)$ bezeichnen wir den verfolgten Zweig periodischer Lösun-

gen mit Parameter $\lambda \in \Lambda$. \mathcal{U} sei eine Umgebung von C . Eine stetige Funktion $\psi : \mathcal{U} \rightarrow \mathbb{R}$, die genau in $\lambda = \lambda^*$ (strikt) ihr Vorzeichen ändert, werden wir wiederum *Testfunktion für den Bifurkationspunkt* $(s^*, T^*, \lambda^*) = (s(\lambda^*), T(\lambda^*), \lambda^*)$ nennen. Dies trifft zu, wenn die Funktion stetig differenzierbar ist und die Eigenschaften

$$\psi(s(\lambda^*), T(\lambda^*), \lambda^*) = 0, \quad \left. \frac{d}{d\lambda} \psi(s(\lambda), T(\lambda), \lambda) \right|_{\lambda=\lambda^*} \neq 0 \quad (6.98)$$

am Bifurkationspunkt (s^*, T^*, λ^*) besitzt, d. h. λ^* ist eine reguläre Nullstelle.

Betrachten wir die Testfunktion auf dem Lösungszweig C und kürzen sie vereinfachend mit $\tau(\lambda) = \psi(s(\lambda), T(\lambda), \lambda)$ ab, so können zu gegebener Monodromiematrix $M(\lambda)$ leicht folgende Testfunktionen programmiert werden:

$\tau_{\text{Fold}}(\lambda) = \prod_{i=1}^n (m_i(\lambda) - 1)$	für Fold-, Pitchfork- oder transkritische Bifurkationen
$\tau_{\text{Flip}}(\lambda) = \prod_{i=1}^n (m_i(\lambda) + 1)$	für Flip-Bifurkationen
$\tau_{\text{Torus}}(\lambda) = \prod_{\substack{i,j=1 \\ i>j}}^n (m_i(\lambda)m_j(\lambda) - 1)$	für Torus-Bifurkationen

Der Vorzeichenwechsel einer dieser Funktionen ist offenbar notwendig für das Auftreten der entsprechenden Bifurkation. Darstellungen dieser Testfunktionen, die die Eigenwerte der Monodromiematrix M nicht benötigen, lassen sich mit dem *bialternierenden Produkt* $A \odot B$ zweier Matrizen definieren, das wir bereits mit Definition 5.61 in Kapitel 5 eingeführt haben. Das bialternierende Produkt $C = A \odot B$ zweier Matrizen $A = (a_{ik}) \in \mathbb{R}^{n \times n}$ und $B = (b_{ik}) \in \mathbb{R}^{n \times n}$ ist eine $m \times m$ -Matrix mit $m = n(n-1)/2$, deren Zeilen von C mit dem Multiindex (p, q) , $p = 2(1)n$, $q = 1(1)p - 1$, und deren Spalten mit dem Multiindex (r, s) , $r = 2(1)n$, $s = 1(1)r - 1$ bezeichnet werden. Die Matrixelemente berechnen wir gemäß

$$C_{(p,q),(r,s)} := \frac{1}{2} \left\{ \begin{vmatrix} a_{pr} & a_{ps} \\ b_{qr} & b_{qs} \end{vmatrix} + \begin{vmatrix} b_{pr} & b_{ps} \\ a_{qr} & a_{qs} \end{vmatrix} \right\}. \quad (6.99)$$

Wenn die Monodromiematrix $M = (m_{ik}) \in \mathbb{R}^{n \times n}$ die Eigenwerte m_1, m_2, \dots, m_n hat, so folgt aus Satz 5.62, dass die Matrix $C = M \odot M$ die Eigenwerte $m_i m_j$, $i = 1(1)n$, $j = i + 1(1)n$ besitzt. Wegen (6.99) lassen sich die Elemente von $M \odot M$ leicht zu

$$C_{(p,q),(r,s)} := m_{pr} m_{qs} - m_{qr} m_{ps}$$

berechnen und nach dem Schema (5.89) anordnen. Liegt ein konjugiert komplexes Paar von Eigenwerten auf dem Einheitskreis, so verschwindet die Determinante von $M \odot M - I$. Damit lassen sich folgende Testfunktionen angeben, die lediglich Determinantenberechnungen erfordern:

$$\begin{aligned}\tau_{\text{Fold}}(\lambda) &= \det(M - I) && \text{für Fold-, Pitchfork- oder transkritische Bifurkation} \\ \tau_{\text{Flip}}(\lambda) &= \det(M + I) && \text{für Flip-Bifurkation} \\ \tau_{\text{Torus}}(\lambda) &= \det(M \odot M - I) && \text{für Torus-Bifurkation}\end{aligned}$$

Die Rechenzeiten zur Berechnung des bialternierenden Produktes nehmen allerdings mit größerer Dimension n extrem zu, so dass dieser Zugang nur für kleine Systeme empfohlen werden kann. Man überprüfe dies mit selbst gewählten Matrizen M verschiedener Größe. Es sei betont, dass das Verschwinden der Testfunktion lediglich notwendig für eine entsprechende Bifurkation ist. Außer den drei generischen Kodimension-1-Bifurkationen Fold, Flip und Torus sind sogenannte Bifurkationen höherer Ordnung möglich, insbesondere die Pitchfork- und transkritische Bifurkation. Für diese beiden Typen ist $\tau_{\text{Fold}}(\lambda) = \det(M - I) = 0$ notwendig. Kann man die auf die lokale Zentrumsmannigfaltigkeit $W_{loc}^c(0)$ reduzierte Poincaré-Abbildung $g(u, \lambda)$ in der Darstellung

$$u = g(u, \lambda) \quad \text{mit} \quad g(u, \lambda) = A_c u + h(u, \lambda) \quad (6.100)$$

gewinnen, so sind gemäß Abschnitt 6.4.2 notwendige Bedingungen für eine

- transkritische Bifurkation:

$$\frac{\partial g}{\partial \lambda}(0, 0) = 0 \quad \text{und} \quad \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(0, 0) \neq 0$$

- Heugabel-Bifurkation:

$$\frac{\partial g}{\partial \lambda}(0, 0) = 0 \quad \text{und} \quad \frac{1}{2} \frac{\partial^2 g}{\partial x^2}(0, 0) = 0.$$

Wir wollen der Einfachheit wegen die drei generischen Bifurkationstypen Fold, Flip und Torus voraussetzen. Wie approximieren wir dann die auftretenden Bifurkationswerte λ^* ? Nehmen wir an, dass während der Fortsetzung der periodischen Lösung zwei Parameterwerte $\lambda_0 < \lambda_1$ ermittelt werden, für die ein Vorzeichenwechsel der Testfunktion eintritt, d. h.

$$\tau(\lambda_0) \tau(\lambda_1) = \psi(s(\lambda_0), T(\lambda_0), \lambda_0) \psi(s(\lambda_1), T(\lambda_1), \lambda_1) < 0. \quad (6.101)$$

Mit der Stetigkeit von ψ folgt dann, dass ein Wert $\lambda^* \in (\lambda_0, \lambda_1)$ existieren muss, für den $\tau(\lambda^*) = \psi(s(\lambda^*), T(\lambda^*), \lambda^*) = 0$ Null wird. Diese Nullstelle λ^* lässt sich mittels Bisektionsverfahren oder mittels Regula falsi mit Intervallschachtelung sicher eingrenzen. Beide Verfahren haben sich als zuverlässig erwiesen, da eine einmal detektierte und erfasste Nullstelle damit nicht mehr verloren geht. In Abbildung 6.30 wird das Verhalten von Algorithmus 6.52 bei einer Torus-Bifurkation veranschaulicht.

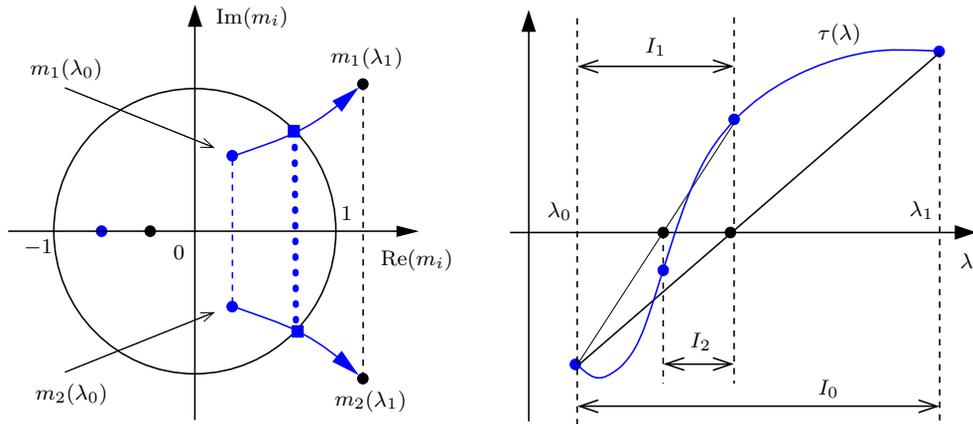


Abb. 6.30 Eingrenzung eines Bifurkationswertes λ^* : Charakteristische Multiplikatoren bei Torus-Bifurkation (links) und Regula falsi mit Intervallschachtelung (rechts)

ALGORITHMUS 6.52 (Einschachtelung von Bifurkationswerten)

Function $[\lambda^*, \Lambda] = \text{bifurkation}(\lambda_0, \lambda_1, \text{tol}, \text{kmax})$

1. (Eignungstest) Berechne $\tau_0 = \tau(\lambda_0)$, $\tau_1 = \tau(\lambda_1)$. Falls $\tau_0\tau_1 > 0$, so STOP.
2. Für $k = 0, 1, \dots, \text{kmax}$ iteriere:
 - 2.1. (Regula falsi) $\lambda^* := \lambda_0 + (\lambda_1 - \lambda_0) \frac{\tau_0}{\tau_0 - \tau_1}$
 - 2.2. (Testfunktion) Bestimme $\tau^* := \tau(\lambda^*)$.
Falls $|\tau^*| < \text{tol}$, so gehe zu Schritt 3.
 - 2.3. (Intervallschachtelung) Falls $\tau^* \cdot \tau_1 < 0$, so
 $\lambda_0 := \lambda^*$, $\tau_0 := \tau^*$, sonst $\lambda_1 := \lambda^*$, $\tau_1 := \tau^*$
 - 2.4. Falls $|\lambda_1 - \lambda_0| < \text{tol}$, so $\lambda^* := (\lambda_0 + \lambda_1)/2$ und gehe zu Schritt 3.
3. Falls $\lambda_0 < \lambda_1$, so setze $\Lambda := [\lambda_0, \lambda_1]$, sonst $\Lambda := [\lambda_1, \lambda_0]$.
4. Return λ^*, Λ

Bemerkungen 6.53

1. Wegen der Regularitätsvoraussetzung (6.98) liefern Newton-Verfahren die Nullstellenapproximationen zwar schneller; sie bergen wegen der genäherten Berechnung der Testfunktion und deren Ableitung allerdings oft die Gefahr einer Divergenz.
2. Im Falle der Fold-Bifurkation stehen keine geeigneten Parameterwerte $\lambda_0 \neq \lambda_1$ des natürlichen Parameters zur Verfügung, die der Bedingung (6.101) genügen. Hier wird eine Neuparametrisierung – z. B. mittels Pseudo-Bogenlänge – erforderlich. Die Ausführungen zu den Testfunktionen lassen sich jedoch leicht auf allgemeine Parametrisierungen der Lösungskurve übertragen und wie Algorithmus 6.52 darstellen. ■

6.4.4 Anwendung bei nichtlinearen energetischen Systemen

Als Anwendung wollen wir ein allgemeines energetisches Dreiphasen-System der Elektrotechnik betrachten, das von Vogt und Büntig (1998), Büntig und Vogt (1998) modelliert und in Vogt und Büntig (2003) ausführlich analysiert wird. Die Gleichungen beschreiben ein grundlegendes mathematisches Modell für Stromversorgungssysteme. Es trägt dem nichtlinearen Verhalten solcher Systeme Rechnung und soll Auslastungsaussagen ermöglichen.

Die Schaltung des Systems ist in Abbildung 6.31 dargestellt. Darin werden drei nichtlineare Ferroresonanzkreise zum Dreiphasen-System gekoppelt. Die einzelnen Phasen werden mit R , S bzw. T bezeichnet. Die Rückführung erfolgt über den rechts gezeichneten Nullleiter. Dabei steht R_0 für den Leitungswiderstand und L_0 für die Induktivität des

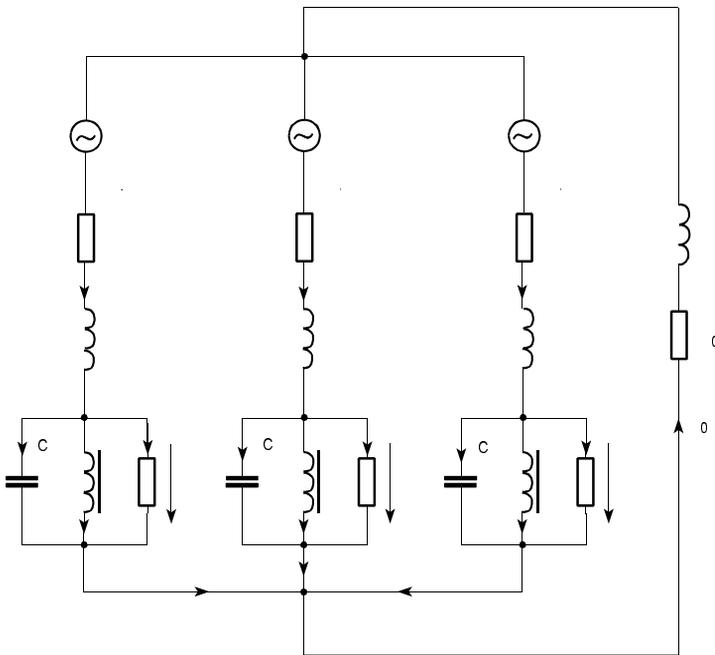


Abb. 6.31 Stern-Stern Schaltung des Dreiphasen-Systems aus Vogt und Büntig (1998)

Nullleiters. Der Generator ist symmetrisch, d. h. mit der Frequenz ω gilt

$$e_R(t) = \hat{E}_0 \sin(\omega t), \quad e_S(t) = \hat{E}_0 \sin(\omega t - 2\pi/3), \quad e_T(t) = \hat{E}_0 \sin(\omega t - 4\pi/3).$$

Ein wesentlicher Bestandteil der nichtlinearen Belastung ist die Induktivität mit dem nichtlinearen Kennlinien-Ansatz

$$i_L = i_L(\psi) = a^* \psi + b^* \psi^n \quad \text{mit} \quad a^*, b^* > 0, \quad n = 3 \quad \text{oder} \quad n = 9. \quad (6.102)$$

Darin bezeichnet ψ den Magnetfluss und die Konstanten a^*, b^*, n sind charakteristisch für das Kernmaterial des Transformators.

Das System wird im allgemeinen Fall eines realen Nullleiters mit $R_0 \neq 0$ und $L_0 \neq 0$ ohne gemeinsamen Eisenkreis durch neun Differenzialgleichungen 1. Ordnung für die Flüsse, Ströme und Spannungen in den drei Phasen beschrieben. Zur Normierung dieser Gleichungen werden dafür die neun dimensionslosen Variablen

- x_R, x_S, x_T – für die normierten Flüsse in den Phasen R, S, T ,
- y_R, y_S, y_T – für die normierten Ströme in den Phasen R, S, T ,
- z_R, z_S, z_T – für die normierten Spannungen in den Phasen R, S, T

eingeführt. Die normierte Zeit $\tau := \omega t$ wollen wir auch weiterhin mit t bezeichnen. Die Nichtlinearität (6.102) des Systems wird durch die Terme $a_\mu x_\mu + b_\mu x_\mu^n$, $\mu \in \{R, S, T\}$ beschrieben. Wir gehen von vollständiger Symmetrie aus und nehmen für alle drei Phasen identische Kenngrößen an. Auch die Amplitude der Erregerspannung wird für alle drei Phasen gleich gesetzt und ist nachfolgend der Hauptverfolgungsparameter λ . Das System (6.103) gliedert sich in drei gekoppelte Gleichungsblöcke – einen für jede Phase:

$$\begin{aligned}
 \dot{x}_R &= z_R \\
 \dot{y}_R &= \frac{q_0}{q_1} (p_2 p_0 \sin t - l_{Sp_0} \sin(t - \frac{2}{3}\pi) - l_{Tp_0} \sin(t - \frac{4}{3}\pi)) \\
 &\quad - (r_1 p_2 p_0) y_R + (s_1 l_{Rp_0}) y_S + (t_1 l_{Rp_0}) y_T \\
 &\quad - \rho l_{Rp_0} (y_R + y_T + y_S) + \frac{1}{r_0} (-p_2 p_0 z_R + l_{Sp_0} z_S + l_{Tp_0} z_T) \\
 \dot{z}_R &= q_1 (y_R - x_R (a + b x_R^8)) - q_2 z_R \\
 \\
 \dot{x}_S &= z_S \\
 \dot{y}_S &= \frac{q_0}{q_3} (-l_{Rp_0} \sin t + p_3 p_0 \sin(t - \frac{2}{3}\pi) - l_{Tp_0} \sin(t - \frac{4}{3}\pi)) \\
 &\quad + (r_1 l_{Sp_0}) y_R - (s_1 p_3 p_0) y_S + (t_1 l_{Sp_0}) y_T \\
 &\quad - \rho l_{Sp_0} (y_R + y_T + y_S) + \frac{1}{s_0} (l_{Rp_0} z_R - p_3 p_0 z_S + l_{Tp_0} z_T) \\
 \dot{z}_S &= q_3 (y_S - x_S (a + b x_S^8)) - q_4 z_S \\
 \\
 \dot{x}_T &= z_T \\
 \dot{y}_T &= \frac{q_0}{q_5} (-l_{Rp_0} \sin t - l_{Sp_0} \sin(t - \frac{2}{3}\pi) + p_4 p_0 \sin(t - \frac{4}{3}\pi)) \\
 &\quad + (r_1 l_{Tp_0}) y_R + (s_1 l_{Tp_0}) y_S - (t_1 p_4 p_0) y_T \\
 &\quad - \rho l_{Tp_0} (y_R + y_T + y_S) + \frac{1}{t_0} (l_{Rp_0} z_R + l_{Sp_0} z_S - p_4 p_0 z_T) \\
 \dot{z}_T &= q_5 (y_T - x_T (a + b x_T^8)) - q_6 z_T
 \end{aligned} \tag{6.103}$$

Die erste Gleichung jeder Phase R, S, T basiert auf dem Induktionsgesetz und verknüpft den normierten Fluss mit der normierten Spannung. Die zweite Gleichung beschreibt die Kopplung zwischen den einzelnen Phasen des Systems. Hier sind auch die Amplituden der Erregerfunktion enthalten. Die dritte Gleichung jeder Phase enthält das Modell der Nichtlinearität und die Belastungsparameter des Systems. Typische Parameterwerte sind

- $a = 0.25, \quad b = 0.75, \quad \rho = 0.10 \quad \text{und} \quad l_R = l_S = l_T = 1.0$
- $q_1 = 1.0, \quad q_2 = 0.50, \quad q_3 = 1.0, \quad q_4 = 0.50, \quad q_5 = 1.0, \quad q_6 = 0.50$
- $r_0 = s_0 = t_0 = 0.05 \quad \text{und} \quad r_1 = s_1 = t_1 = 0.10,$

womit sich die Sekundärparameter

- $p_0 = 1/p_1 \quad \text{mit} \quad p_1 = 1 + l_S + l_R + l_T \quad \text{sowie}$
- $p_2 = 1 + l_S + l_T, \quad p_3 = 1 + l_R + l_T, \quad p_4 = 1 + l_S + l_R$

ergeben. Der Parameter $\lambda = q_0 > 0$ ist frei. Von Bedeutung sind Parameterbereiche für λ , in denen sich stabile 2π -periodische Schwingungen dieses komplizierten Systems einstellen. Folgende Arbeitsschritte zur Systemanalyse werden empfohlen:

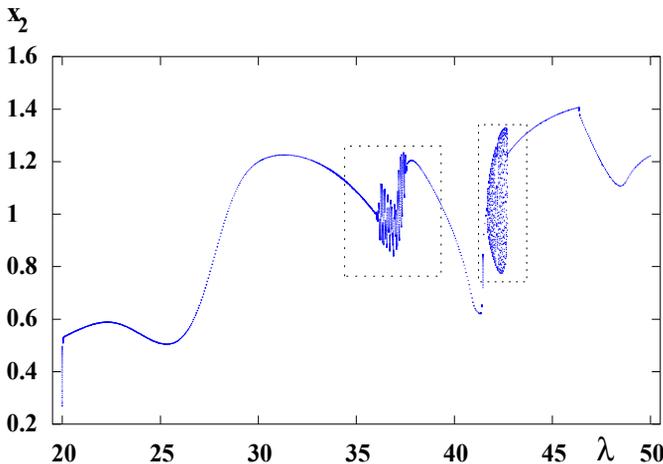


Abb. 6.32

Darstellung der Stroboskop-Abbildung:

In den Parameterintervallen $34 \leq \lambda \leq 38$ und $41 \leq \lambda \leq 43$ erkennt man das Auftreten nicht-periodischer Attraktoren

Einfaches Bifurkationsdiagramm

Mittels Lösungssimulation verschaffen wir uns einen ersten Eindruck vom Auftreten stabiler periodischer Lösungen. Beginnend bei $\lambda = 20$ lassen wir das System über k_0 Perioden einschwingen und stellen dann m Stroboskop-Punkte $x_S(2\pi k)$, $k = k_0(1)k_0 + m - 1$, einer Lösungskomponente grafisch über dem Kontrollparameter λ dar. Diese einfache Fixpunkt-Iteration der Stroboskop-Abbildung – im Ingenieurwesen oft auch “Brute-force“-Methode genannt – veranschaulicht allerdings nur Attraktoren; andere Lösungen können hingegen nicht detektiert werden. Fortsetzung über das Parameterintervall $20 \leq \lambda \leq 50$ ergibt das Diagramm in Abbildung 6.32.

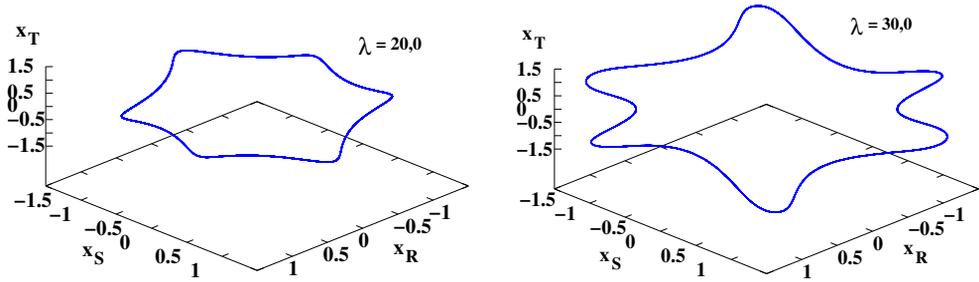


Abb. 6.33 Periodische Lösungen in (x_R, x_S, x_T) für $\lambda = 20,0$ und $\lambda = 30,0$

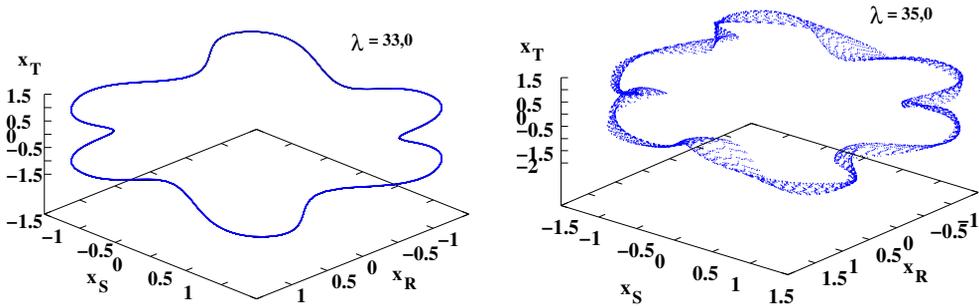


Abb. 6.34 Lösungen in (x_R, x_S, x_T) für $\lambda = 33,0$ und $\lambda = 35,0$

Explizite Lösungsdarstellung in 2D oder 3D

In höherdimensionalen Systemen vermitteln Lösungsgrafiken bei festem Parameterwert λ in der Regel nur geringen Erkenntnisgewinn. Nach Auswahl der drei Koordinaten x_R, x_S, x_T liefert die Projektion der Orbits in den 3D-Unterraum der Magnetflüsse die Abbildungen 6.33 und 6.34. Für $\lambda = 35,0$ vermuten wir hier eine quasi-periodische oder chaotische Lösung.

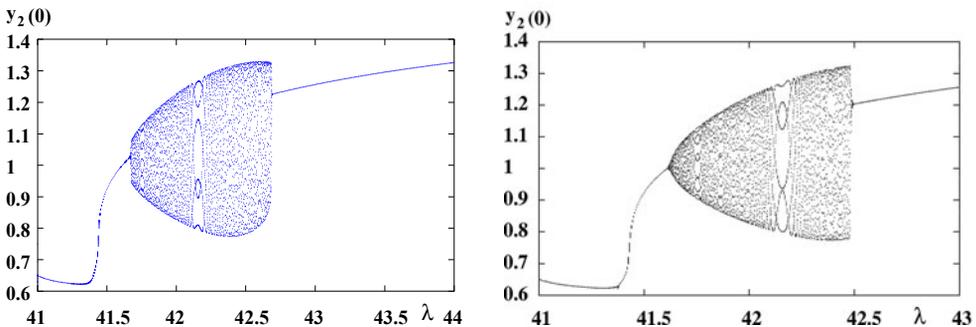


Abb. 6.35 Bifurkationsdiagramm mit λ aufwärts (links) und abwärts (rechts)

Eingrenzung nicht-periodischer Attraktoren

Durch Aufweitung (Zoomen) der Parameterintervalle können nun die Übergänge zu nicht-periodischen Lösungen genauer eingegrenzt werden. Dazu generieren wir die Bifurkati-

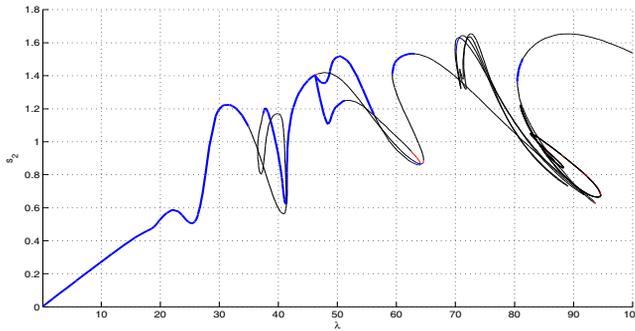


Abb. 6.36 (λ, y_R) -Diagramm (blau – stabil, schwarz – instabil). Neben dem Primärzweig S_1 wurden auch die ermittelten abzweigenden Lösungen eingetragen

onsdiagramme sowohl monoton wachsend in Abbildung 6.35 links als auch fallend in Abbildung 6.35 rechts. Mit etwas Übung lässt der “Torus-Bubble“ bei $\lambda \approx 41.5$ eine superkritische und bei $\lambda \approx 42.5$ eine subkritische Torus-Bifurkation vermuten. Instabile periodische Lösungen sowie abzweigende Lösungen können wir mit derartigen Diagrammen nicht erhalten, womit eine genaue Bifurkationsdetektierung nicht möglich ist.

2D-Bifurkationsdiagramm mittels Lösungsfortsetzung

Nutzen wir Algorithmus 6.16 zusammen mit einer Lösungsfortsetzung, wie in Algorithmus 6.38 dargestellt, so können wir stabile und instabile Lösungen ermitteln. Deren Unterscheidung wird mit der zusätzlichen Stabilitätsanalyse (vgl. Schritt 7 auf S. 331) leicht möglich. Für $\lambda = 0$ hat System (6.103) die Nulllösung, mit der wir die Fortsetzung über das Gesamtintervall $0 \leq \lambda \leq 100$ starten können. Im 2D-Bifurkationsdiagramm

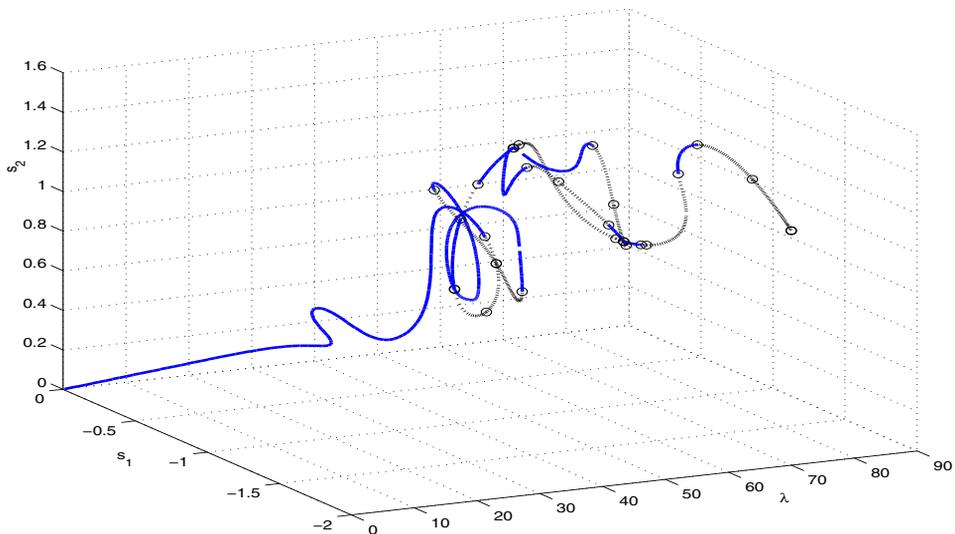


Abb. 6.37 (λ, x_R, y_R) -Diagramm (blau – stabil, schwarz – instabil)

für $(\lambda, y_R(0))$ (vgl. Abb. 6.36) wurden neben dem Primärzweig S_1 auch die ermittelten abzweigenden Lösungen eingetragen.

Bifurkationsdetektierung

Wesentlich informativer sind 3D-Bifurkationsdiagramme, die zwei ausgewählte Lösungskomponenten über dem Kontrollparameter darstellen; in Abbildung 6.37 sind dies $(\lambda, x_R(0), y_R(0))$, $0 \leq \lambda \leq 80$. Offenbar existiert für $0 \leq \lambda \leq 34$ eine stabile 2π -periodische Lösung. Mit der vorgestellten indirekten Methode liefert Algorithmus 6.52 die Bifurkationspunkte in der ersten Spalte von Tabelle 6.3 sowie die detektierten Bifurkationstypen. An den berechneten charakteristischen Multiplikatoren mit $|m_i| \approx 1$ lässt

Tab. 6.3 Detektierte Bifurkationen des energetischen Dreiphasen-Systems

Parameter λ (mit CPO)	Multiplikatoren m_i	Betrag $ m_i $	Typ der Bifurkation	Parameter (AUTO)
34.93632841	$0.9981943 \pm 0.0600340i$	0.9999979	Torus	<u>34.93633</u>
36.48325162	0.9999425	0.9999425	Pitchfork	<u>36.48326</u>
39.50521827	$0.9998027 \pm 0.0195815i$	0.9999944	Torus	<u>39.50487</u>
41.39543821	1.0002848	1.0002848	Pitchfork	<u>41.39548</u>
41.62966402	$0.3740468 \pm 0.9274070i$	0.9999974	Torus	<u>41.62967</u>
42.54245646	$0.6063643 \pm 0.7951860i$	0.9999992	Torus	<u>42.54246</u>
46.25346533	0.9999141	0.9999141	Pitchfork	<u>46.25356</u>
47.19143019	$0.9625519 \pm 0.2709929i$	0.9999716	Torus	<u>47.19146</u>
64.06674375	1.0000024	1.0000024	Pitchfork	<u>64.06676</u>
64.57394419	$0.9744982 \pm 0.2260976i$	1.0003834	Torus	<u>64.57393</u>
64.60599927	1.0021892	1.0021892	Fold	<u>64.60600</u>
59.28615751	0.9998577	0.9998577	Fold	<u>59.28616</u>
63.10205664	$0.9604222 \pm 0.2785375i$	9.9999691	Torus	<u>63.10206</u>
72.27656438	1.0000347	1.0000347	Pitchfork	<u>72.27654</u>
92.53787115	$0.9989053 \pm 0.0468599i$	1.0000039	Torus	<u>92.53791</u>
93.45943866	0.9998071	0.9998071	Pitchfork	<u>93.45937</u>
93.71406409	0.9993964	0.9993964	Fold	<u>93.71406</u>
93.70913803	$0.0969198 \pm 0.7473412i$	0.7535996	Torus	<u>93.69532</u>
80.47671203	0.9998948	0.9998948	Fold	<u>80.47671</u>
81.41454519	1.0000303	1.0000303	Pitchfork	<u>81.41452</u>
84.29459070	$0.8574302 \pm 0.5146328i$	1.0000167	Torus	<u>84.29463</u>

sich die geringere Genauigkeit der Fold-Bifurkationswerte und der Torus-Bifurkation bei $\lambda \approx 93.70913803$ erkennen. Für ausgewählte Grafiken wie in Abbildung 6.38 sind die durchweg richtigen Typangaben jedoch nützlich.

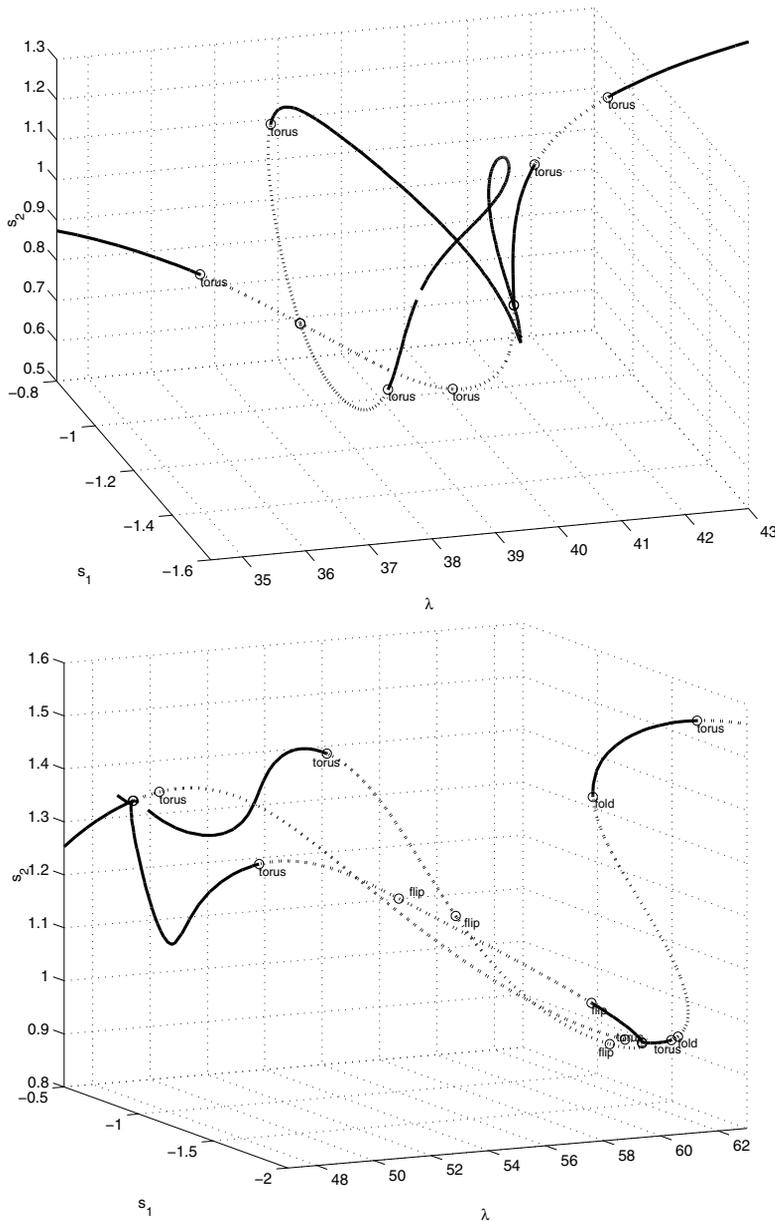


Abb. 6.38 Bifurkationsdiagramm für $35 \leq \lambda \leq 43$ (oben) und $47 \leq \lambda \leq 63$ (unten)

Analyse von Bifurkationsszenarien

Betrachtet man einen ausgewählten Lösungszweig, der zwischen zwei stabilen Zuständen mindestens 2 Bifurkationen durchläuft, so kann man von einem *Bifurkationsszenario* sprechen. Ausgehend von einem stabilen Zustand beginnt das Szenario, wenn ein Multiplikator(paar) den Einheitskreis verlässt und damit einen Stabilitätsverlust signalisiert. Es

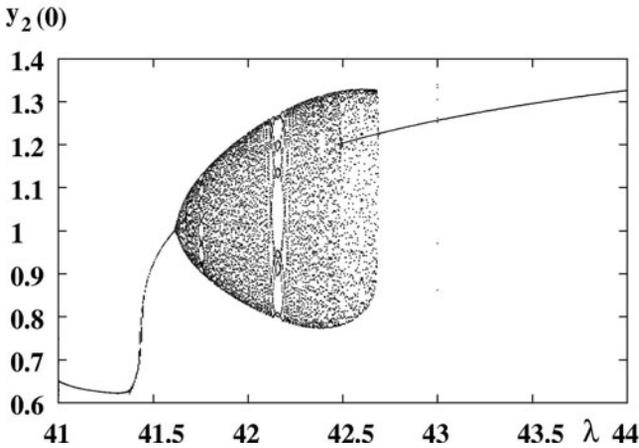


Abb. 6.39 „brute force“-Darstellung von Szenario 1. Der Parameter λ wird sowohl monoton wachsend als auch monoton fallend variiert

endet, wenn alle Multiplikatoren in den Einheitskreis zurückgekehrt sind und die Lösung wieder stabil ist. Jedes Szenario kann durch ein dafür typisches Lösungsdiagramm und gegebenenfalls durch ein „brute-force“-Diagramm beschrieben werden. Untersuchen wir das einfache, aber häufig auftretende Szenario des primären Lösungszweiges S_1 mit zwei Bifurkationen in Diagramm 6.39, das auch in Abbildung 6.38 erkennbar wird. Die Tabelle 6.4 enthält zu jedem Bifurkationspunkt den berechneten Parameterwert λ , den Bifurkationstyp, die drei betragsgrößten Multiplikatoren(paare) und eine schematische Darstellung des Einheitskreises mit den für die Bifurkation wichtigen Floquet-Multiplikatoren.

Tab. 6.4 Bifurkationspunkte von Szenario 1 mit den betragsgrößten Multiplikatoren

Szenario 1 im primären Lösungszweig S_1		
λ / Typ	Multiplikatoren	Skizze
<u>41.62967</u> Torus- Bifurkation	$m_{1,2} = 0.37405 \pm 0.92742i$ $m_3 = 0.53344 + 0.00000i$ $m_{4,5} = 0.46258 \pm 0.17661i$	
<u>42.54246</u> Torus- Bifurkation	$m_{1,2} = 0.60636 \pm 0.79519i$ $m_3 = 0.50508 + 0.00000i$ $m_{4,5} = -0.25929 \pm 0.07309i$	

Das zweite in Abbildung 6.38 erkennbare Szenario liegt im Intervall $34 \leq \lambda \leq 42$. Neben dem primären Lösungszweig S_1 ist der darin dargestellte sekundäre Zweig S_2 mit seinen Stabilitätsverhältnissen dargestellt. Die zugehörigen „brute-force“-Diagramme 6.40 sind nicht aussagekräftig und scheinen nicht mit den Werten in Abbildung 6.38 zu korrespondieren. In Tabelle 6.5 wird das Bifurkationsverhalten beider Lösungszweige S_1

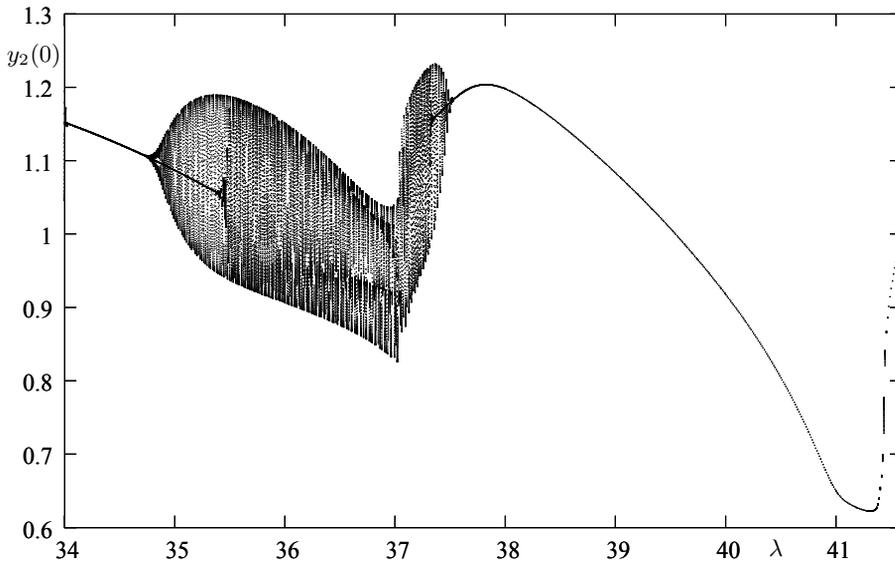


Abb. 6.40 „brute force“-Darstellung von Szenario 2 (aufwärts und abwärts)

und S_2 verdeutlicht. Hier interagieren eine Torus- und eine Pitchfork-Bifurkation des Primärzweiges. Die stabilen Sekundärzweige S_2 für $\lambda > 37.45021$ führen offensichtlich zu den dargestellten Diagrammen. Weitere Bifurkationen bei größeren Parameterwerten $\lambda < 100$ – wie in Abbildung 6.38 dargestellt – wurden von Ernst (2004) mit dem an der TU Ilmenau von Peterseim (2004) entwickelten MATLAB-Paket CPO (Continuation of Periodic Orbits) analysiert und mittels der Ergebnisse des bekannten Paketes AUTO2000 (vgl. Doedel et al. (2002)) verifiziert. Tabelle 6.3 zeigt in der letzten Spalte die mit AUTO berechneten Parameterwerte samt den übereinstimmenden signifikanten Dezimalstellen (unterstrichen) sowie die von beiden Codes stets richtig erkannten Bifurkationstypen. Beide Programmpakete haben eine Torus-Bifurkation bei $\lambda \approx 80.47679$ nicht detektiert, da sie sehr nahe an einer Fold-Bifurkation eintritt. Die gesamte Fortsetzung des primären Lösungszweiges S_1 über das Intervall $[0.1, 100]$ inklusive der Ermittlung aller Verzweigungspunkte aus Tabelle 6.3 nahm auf einem PC (1.5 GHz) mit CPO lediglich 2:30 Minuten in Anspruch.

Generell lässt sich die Vielzahl von Methoden zur Approximation von Bifurkationspunkten in zwei grundlegende Klassen einteilen:

- *Indirekte Methoden* bestimmen zu einer Folge (λ_k) , $k = 0, 1, 2, \dots$ von Parameterwerten die zugehörigen periodischen Lösungen und berechnen damit die Werte $\tau(\lambda_k)$ einer oder mehrerer ausgewählter Testfunktionen. Mittels iterativer Nullstellenbestimmung bzw. durch Interpolation kann so ein lokaler Bifurkationspunkt $(s(\lambda^*), T(\lambda^*), \lambda^*)$ detektiert werden. In vielen Fällen ist die erreichte Genauigkeit indirekter Methoden zur Bestimmung von Bifurkationspunkten *in praxi* ausreichend, wie die CPO-Werte

Tab. 6.5 Bifurkationspunkte von Szenario 2 in den Lösungszweigen S_1 und S_2

Szenario 2 in den Lösungszweigen S_1 und S_2		
λ / Typ	Multiplikatoren	Skizze
<u>34.93633</u> Torus-Bif. in Zweig S_1	$0.99819 \pm 0.06005i$ $0.49487 + 0.00000i$ $-0.19431 \pm 0.79623i$	
<u>36.48326</u> Pitchfork-Bif. in Zweig S_1	$0.99996 + 0.00000i$ $1.01495 \pm 0.03890i$ $-0.13993 \pm 0.62362i$	
<u>37.45021</u> Torus-Bif. in Zweig S_2	$0.99805 \pm 0.06247i$ $0.52999 + 0.00000i$ $-0.11696 \pm 0.13727i$	
<u>39.50487</u> Torus-Bif. in Zweig S_1	$0.99981 \pm 0.01958i$ $4.62004 + 0.00000i$ $-0.08849 \pm 0.12820i$	
<u>41.39548</u> Pitchfork-Bif. in Zweig S_1	$0.99998 + 0.00000i$ $0.68336 \pm 0.01742i$ $-0.05848 \pm 0.14020i$	

in Tabelle 6.3 erkennen lassen.

- *Direkte Methoden* bestimmen einen lokalen Bifurkationspunkt $(s(\lambda^*), T(\lambda^*), \lambda^*)$ als reguläre Lösung eines erweiterten Randwertproblems. Dazu werden die bestimmenden Differenzialgleichungen, Rand- und Phasenbedingungen um weitere Gleichungen zur Berechnung von Nullraumelementen ergänzt. Die Zahl der gesuchten Funktionen
 - verdoppelt sich damit im Fall der Fold- und der Flip-Bifurkation
 - verdreifacht sich bei Torus-Bifurkation.

Eine gute Übersicht über verschiedene *minimal erweiterte Randwertprobleme* findet man in einer Arbeit von Doedel et al. (2003). Diese erweiterten DGL-Systeme sind für den jeweiligen – vorab zu detektierenden – Bifurkationstyp spezifisch und setzen zudem gute Startnäherungen voraus. Im Code AUTO2000 (vgl. Doedel et al. (2002)) werden sie erfolgreich eingesetzt.

6.5 Aufgaben

Aufgabe 6.1

Analysieren Sie das dynamische System aus Moore (1996) mit Parameter $\lambda > 0$

$$\begin{aligned}\dot{x} &= (\lambda - 3)x - 4y + x(z + \frac{1}{5}(1 - z^2)) \\ \dot{y} &= 4x + (\lambda - 3)y + y(z + \frac{1}{5}(1 - z^2)) \\ \dot{z} &= \lambda z - (x^2 + y^2 + z^2).\end{aligned}$$

- Bestimmen Sie sämtliche Gleichgewichtslösungen und untersuchen Sie deren Stabilität.
- Für welche Parameterwerte λ^* tritt Stabilitätsverlust ein? Prüfen Sie, ob eine Hopf-Bifurkation vorliegt.
- Transformieren Sie das System in Zylinderkoordinaten (r, θ, z) mit $(x, y, z) = (r \cdot \cos \theta, r \cdot \sin \theta, z)$ und ermitteln Sie periodische Orbits mit $r(t) = \text{const}, z(t) = \text{const}$. Wie lautet deren Periode T ?
- Für $\lambda^{**} = 2$ existiert ein periodischer Orbit mit $r = 1$ und $z = 1$. (Probieren!) Welche Bifurkation tritt bei $\lambda^{**} = 2$ für diesen Orbit im (r, z) -System auf?

Aufgabe 6.2

Betrachten Sie für den Parameter $\lambda \in \mathbb{R}$ folgendes DGL-System

$$\begin{aligned}\dot{x}_1 &= \lambda x_1 - \omega x_2 - x_1(x_1^2 + x_2^2) \\ \dot{x}_2 &= \omega x_1 + \lambda x_2 - x_2(x_1^2 + x_2^2)\end{aligned}$$

mit dem festen reellen Wert $\omega > 0$.

- Bestimmen Sie periodische Lösungen dieses Systems für Werte $\lambda > 0$. Wie groß ist die Periode?
- Stellen Sie das zugehörige Variationssystem auf, und lösen Sie es. Wie lauten die Monodromiematrix M und die zugehörige Matrix R ?
- Bestimmen Sie die charakteristischen Multiplikatoren und die charakteristischen Exponenten. Welche Stabilitätseigenschaft besitzen die Lösungen?

Aufgabe 6.3

Die Monodromiematrix M eines periodischen Orbits der Periode T hat die Darstellung

$$M = e^{T \cdot R} = \sum_{k=0}^{\infty} \frac{1}{k!} \cdot (T \cdot R)^k$$

mit der Matrix $R \in \mathbb{R}^{n \times n}$ aus der Floquet-Darstellung.

- a) Wie transformieren sich die Eigenwerte μ_j (charakteristische Exponenten) von R in die Eigenwerte m_j (charakteristische Multiplikatoren) von M und umgekehrt?
 b) Welche Beziehung besteht zwischen den Eigenvektoren beider Matrizen?
 c) Überprüfen Sie Ihre Aussagen an den Matrizen

$$R_1 = \begin{pmatrix} -2 & 0 \\ 0 & -1 \end{pmatrix}, \quad R_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Aufgabe 6.4

Entwickeln Sie eine MATLAB- oder eine C-Funktion `shooting` zu Algorithmus 6.16 „Einfaches Schießverfahren ohne Variationssystem“ für periodische Orbits im nicht-autonomen Fall ohne Benutzung des Variationssystems. Ermitteln Sie darin auch die Eigenwerte und zugehörige normierte Eigenvektoren der Monodromiematrix M .

Aufgabe 6.5

Gegeben ist die DGL 2. Ordnung von E. Philippow

$$\ddot{x} - \varepsilon(1 - x^2 - \dot{x}^2)\dot{x} + x + b(4x^3 - 3x) = B \cos 3t$$

mit Parametern b, B, ε , die ein *subharmonisch erregtes elektrisches Netzwerk* beschreibt. Bestimmen Sie 2π -periodische Lösungen für die Parameterwerte

- a) $b = 1.0, \quad B = 1.0, \quad \varepsilon = 3.0,$
 b) $b = 1.0, \quad B = 0.3, \quad \varepsilon = 3.0,$
 c) $b = 1.0, \quad B = 0.3, \quad \varepsilon = 0.75.$

Versuchen Sie, im Fall a) exakte Lösungen zu bestimmen. Welche der erhaltenen Lösungen sind asymptotisch stabil, welche sind instabil?

Aufgabe 6.6

Entwickeln Sie eine MATLAB- oder eine C-Funktion `contperiodauto` zu Algorithmus 6.38 „Fortsetzung periodischer Orbits“ im autonomen Fall mit Stabilitätsanalyse unter Nutzung des Variationssystems. Stellen Sie damit auch die charakteristischen Multiplikatoren zusammen mit dem Einheitskreis in der komplexen Ebene grafisch dar. Beachten Sie dazu Bemerkung 6.40.

Aufgabe 6.7

Gegeben ist das DGL-System 2. Ordnung von Philippow und Büntig (1992)

$$\ddot{x} + \alpha \dot{x}^3 - \beta \dot{x} + (1 + B \sin 2t)x = 0$$

mit Amplitude $B = 0.1$, den Werten $\alpha = \varepsilon - B$, $\beta = \varepsilon/2 - B$ und dem reellen Parameter $\varepsilon > 0$. Es beschreibt ein *parametrisch erregtes elektrisches Netzwerk mit subharmonischer Systemantwort*. Bestimmen Sie 2π -periodische Lösungen, verschieden von der trivialen Lösung $x \equiv 0$, für den Parameterbereich $0 < \varepsilon < 2$ durch Lösungsfortsetzung. Detektieren Sie dabei Bifurkationen und analysieren Sie deren Typen!

Aufgabe 6.8

Zu gegebener (n, n) -Monodromiematrix M sollen die *Testfunktionen* für Bifurkationen periodischer Orbits

$$\begin{aligned}\tau_{\text{Fold}}(\lambda) &= \det(M - I) && \text{für Fold-, Pitchfork- oder transkritische Bifurkation} \\ \tau_{\text{Flip}}(\lambda) &= \det(M + I) && \text{für Flip-Bifurkation} \\ \tau_{\text{Torus}}(\lambda) &= \det(M \odot M - I) && \text{für Torus-Bifurkation}\end{aligned}$$

in MATLAB oder C programmiert werden. Wie lauten entsprechende Testfunktionen, die die Eigenwerte mit `eig(M)` benutzen? Vergleichen Sie an selbst gewählten Matrizen M verschiedener Größe n die Rechenzeiten.

Aufgabe 6.9

Gegeben ist die parameterabhängige Poincaré-Abbildung $P_\mu : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ eines dynamischen Systems durch

$$P_\mu \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} sx - \omega y - cx(x^2 + y^2) \\ \omega x + sy - cy(x^2 + y^2) \end{pmatrix}$$

mit den parameterabhängigen Größen $s = \omega = \frac{1}{2}\sqrt{2}(1 + \mu)$ und der Konstanten $c > 0$.

- Stellen Sie für Punkte $(x, y) \in \mathbb{R}^2$ die Poincaré-Abbildung grafisch dar. Wählen Sie $c = 1$ und verschiedene μ -Werte. Welche Lösungstypen für das dynamische System sind zu vermuten?
- Weisen Sie nach, daß P_μ am Bifurkationswert $\mu = 0$ alle Voraussetzungen des Satzes von Neimark-Sacker für die Torusbifurkation erfüllt.
- Wie sieht der Zweig invarianter Tori für $\mu > 0$ formelmäßig aus?

Aufgabe 6.10

Auf $\bar{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$ sei die Möbius-Transformation $f : \bar{\mathbb{C}} \rightarrow \bar{\mathbb{C}}$ gegeben durch

$$f(z) := \frac{1+z}{1-z};$$

das heißt insbesondere $f(1) = \infty$, $f(\infty) = \lim_{|z| \rightarrow \infty} \frac{1+z}{1-z} = -1$.

- Ist f bijektiv? Auf welche Menge wird $\mathbb{C}^- := \{z \in \mathbb{C} \mid \operatorname{Re}(z) \leq 0\}$ abgebildet? Auf welche Punkte wird die imaginäre Achse abgebildet?
- Betrachten Sie zu $A \in \mathbb{R}^{n \times n}$ die transformierte Matrix

$$B := (I - A)^{-1}(I + A), \quad I - \text{Einheitsmatrix.}$$

Wie werden die Eigenwerte μ und die Eigenvektoren v von A transformiert?

- c) Es sei (x^0, λ_0) eine Gleichgewichtslage und $A = D_x f(x^0, \lambda_0)$ die zugehörige Jacobi-Matrix des Systems $\dot{x} = f(x, \lambda)$. Wie kann man mit Hilfe von **b)** einen Test auf asymptotische Stabilität durchführen, ohne sämtliche Eigenwerte berechnen zu müssen?

Aufgabe 6.11

W. F. Langford betrachtet ein dynamisches System der Hydrodynamik, das nach mehreren analytischen Problemtransformationen auf folgendes DGL-System

$$\begin{aligned} \dot{x} &= (z - 0.7)x - \omega y \\ \dot{y} &= \omega x + (z - 0.7)y \\ \dot{z} &= 0.6 + z - \frac{1}{3}z^3 - (x^2 + y^2)(1 + \varrho z) + \varepsilon z x^3 \end{aligned} \tag{6.104}$$

mit positiven reellen Parametern ω, ϱ und ε führt.

- a) Stellen Sie die Attraktoren im Falle $\omega = 3.5$ und $\varepsilon = 0$ für Parameterwerte $\varrho = 1.0, 0.7, 0.62, 0.60, 0.5, 0.25$ im Phasenraum dar. Bestimmen Sie – durch geeigneten Ansatz und ehrliches Ausrechnen – den periodischen Orbit γ .
- b) Es gelte weiterhin $\omega = 3.5$ und $\varepsilon = 0$. Für welche Parameterwerte ϱ ist dann γ orbital stabil? Bestimmen Sie (formelmäßig und numerisch) den Bifurkationswert ϱ^* , an dem der Stabilitätsverlust von γ eintritt. Welcher Bifurkationstyp liegt vor (Begründung)?
- c) Stellen Sie auch die Attraktoren im Falle $\omega = 3.5$ und $\varrho = 0.25$ für Parameterwerte $\varepsilon = 0, 0.02, 0.04, 0.06, 0.07$ im Phasenraum dar und interpretieren Sie die Typen der auftretenden Bifurkationen.

7 Quasi-periodische Lösungen und invariante Tori

Übersicht

7.1	Quasi-periodische Funktionen	364
7.2	Parametrisierung invarianter Tori	380
7.3	Quasi-periodische invariante Tori	399
7.4	Aufgaben	420

“The world is full of periodic and quasi-periodic motions and oscillations. That is multifrequency oscillations. Mathematically, to a large extent, that means the study of invariant tori of dynamical systems. These constitute the more systematic part of the phase space of a dynamical system as opposed to the chaotic parts and their study is just as (perhaps more) important as the study of chaos.” (zitiert aus Samoilenko (1991))

Periodische Lösungen nichtlinearer dynamischer Systeme wurden in den vergangenen Jahren gründlich analysiert. Ein umfangreiches numerisches Instrumentarium steht dem Anwender zur Verfügung, wovon wir in Kapitel 6 grundlegende Komponenten vorgestellt haben.

Wesentlich weniger untersucht sind hingegen quasi-periodische Bewegungen mit mehreren Basisfrequenzen, treffender auch als „multifrequente Schwingungen“ bezeichnet. Derartige Lösungen bilden ein kompliziertes und sensitives Studienobjekt. Denn sie gehen bei kleinen Änderungen an den Differenzialgleichungen leicht in periodische Lösungen über, weshalb ein „stabilerer Träger“ dieser multifrequenten Schwingungen gesucht ist. Bereits Jules Henry Poincaré (vgl. Kapitel 6.3) wusste, dass eine derartige Menge ein Torus ist. Als Träger quasi-periodischer Lösungen wurden deshalb invariante Tori von N. N. Bogoljubow und Y. A. Mitropolski untersucht. Eine spezielle Theorie existiert für bestimmte Hamiltonsche Systeme, die sogenannte KAM-Theorie (vgl. dazu die Darstellung in Argyris et al. (1995)), und theoretische Resultate liegen für lineare und „schwach nichtlineare“ dissipative Systeme vor. Wichtige Aussagen zu dieser praktisch wesentlichen Problemklasse findet man bei Iooss und Joseph (1990), Samoilenko (1991) und Broer et al. (1996). Numerische Verfahren zur Berechnung, Verfolgung und Stabili-

tätsanalyse derartiger Schwingungen sind jedoch bisher nur eingeschränkt verfügbar und weiterhin Gegenstand der aktuellen Forschung.

Wir wollen uns in diesem Kapitel der recht anschaulichen Terminologie aus Samoilenko (1991) bedienen und nach Einführung wesentlicher Grundbegriffe die bekannten numerischen Zugänge betrachten, die wir auf konkrete praktische Probleme anwenden.

7.1 Quasi-periodische Funktionen

7.1.1 Torusfunktionen

Der p -dimensionale Standardtorus $\mathbb{T}^p = (\mathbb{R}/2\pi\mathbb{Z})^p$ – desweiteren auch kurz als p -Torus bezeichnet – mit den Winkelkoordinaten $\theta_1, \theta_2, \dots, \theta_p$ modulo 2π ist die Menge

$$\mathbb{T}^p = \{\theta \mid \theta = (\theta_1, \dots, \theta_p)^T, \theta_i \in \mathbb{R} \bmod 2\pi, i = 1, \dots, p\}, \quad (7.1)$$

die über $[0, 2\pi)^p$ parametrisiert ist. Speziell ist der 1-Torus $\mathbb{T}^1 = S^1$ und der 0-Torus \mathbb{T}^0 ist ein einziger Punkt. In Abbildung 7.1 wird der 2-Torus \mathbb{T}^2 mit den (lokalen) Winkelkoordinaten θ_1 und θ_2 in \mathbb{R}^3 dargestellt. Die Mannigfaltigkeit wird umkehrbar eindeutig auf das Quadrat $[0, 2\pi)^2$ im Parameterraum abgebildet. Die Koordinatenlinien $\theta_1 = \text{const}$ und $\theta_2 = \text{const}$ sind in beiden Darstellungen eingezeichnet.

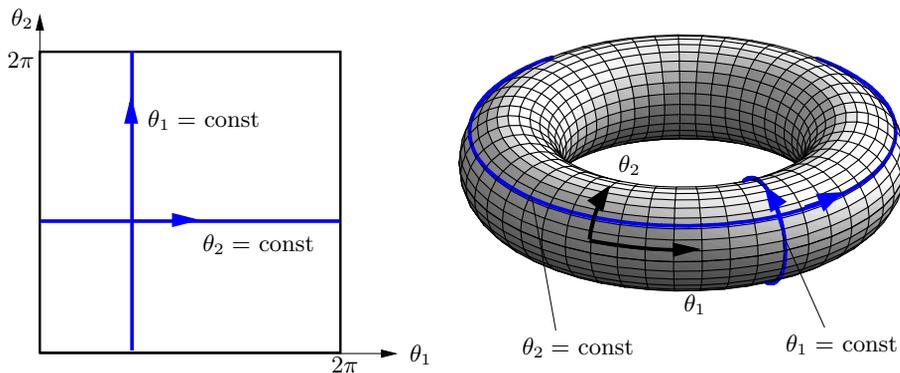


Abb. 7.1 Der 2-dimensionale Standardtorus \mathbb{T}^2 ; links: aufgeschnitten in den Winkelkoordinaten θ_1 und θ_2 ; rechts: dargestellt im Phasenraum \mathbb{R}^3

Jede stetige Funktion $F : \mathbb{T}^p \rightarrow \mathbb{R}^n$ in der Veränderlichen $\theta = (\theta_1, \theta_2, \dots, \theta_p)$, die 2π -periodisch bezüglich $\theta_1, \dots, \theta_p$ ist, heißt *reelle Torusfunktion*. Die Menge aller derartigen Funktionen bildet einen Vektorraum, in dem wir eine Norm

$$\|F\|_0 := \max_{\theta \in \mathbb{T}^p} |F(\theta)| \quad \text{mit der Euklidischen Vektornorm} \quad |F|^2 := \sum_{i=1}^n |F_i|^2$$

eingeführen können. Für einen beliebigen Vektor $F \in \mathbb{R}^n$ soll desweiteren das Betragszeichen $|F|$ stets die Euklidische Norm bezeichnen. Dieser vollständige normierte Raum ist

ein Banach-Raum und wird mit $\mathcal{C}^0(\mathbb{T}^p)$ abgekürzt. Falls F r -mal stetig differenzierbar ist, d. h. F stetige partielle Ableitungen bezüglich aller Variablen θ_i , $i = 1, \dots, p$, bis zur Ordnung r besitzt, so können wir dafür die \mathcal{C}^r -Norm

$$\|F\|_r := \sum_{|\alpha| \leq r} \|D^\alpha F\|_0 \quad \text{mit} \quad D^\alpha F := \frac{\partial^{|\alpha|} F}{\partial \theta_1^{\alpha_1} \partial \theta_2^{\alpha_2} \dots \partial \theta_p^{\alpha_p}} \quad (7.2)$$

eingeführen. Dabei ist $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)$ ein Multi-Index mit $|\alpha| := \sum_{i=1}^p \alpha_i$ und D^α der klassische Differenzialoperator der Ordnung $|\alpha|$ zum Multi-Index α . Mit $\mathcal{C}^r(\mathbb{T}^p)$, $r \geq 0$ bezeichnen wir dann den Raum der r -mal stetig differenzierbaren Torusfunktionen, welcher mit dieser Norm zu einem vollständigen normierten Raum wird.

Beispiel 7.1

Die stetig differenzierbare Funktion $F : \mathbb{T}^3 \rightarrow \mathbb{R}$ mit

$$F(\theta_1, \theta_2, \theta_3) = \pi \sin \theta_1 + \cos \theta_2 \cdot \sin \theta_3 - e^{\cos(\theta_1 - \theta_3)},$$

ist eine \mathcal{C}^r -Torusfunktion für beliebiges $r \in \mathbb{N}$, $r \geq 0$. Dagegen liegt $F : \mathbb{T}^2 \rightarrow \mathbb{R}^4$ mit

$$F(\theta_1, \theta_2) = (\cos(\theta_1 + \theta_2), |\cos \theta_1|, 4\pi, \cos^2 \theta_2)^T.$$

nur in $\mathcal{C}^0(\mathbb{T}^2)$. ■

Betrachten wir nun trigonometrische Polynome N -ten Grades in $\mathcal{C}^0(\mathbb{T}^p)$, die sich mit p -fachen Summen in der Form

$$P(\theta) = \sum_{|k| \leq N} a_k \cos(k, \theta) + b_k \sin(k, \theta) = \sum_{|k| \leq N} P_k e^{i(k, \theta)} \quad (7.3)$$

darstellen lassen. Darin ist $k = (k_1, k_2, \dots, k_p)$ ein Vektor mit ganzzahligen Komponenten, $|k|$ seine Euklidische Norm und $(k, \theta) = k_1 \theta_1 + \dots + k_p \theta_p$. Die Summenzeichen stehen für p -fache Summation. Zwischen den reellen Fourierkoeffizienten a_k und b_k und den komplexen Koeffizienten P_k besteht der Zusammenhang $a_k = \operatorname{Re} P_k$ und $b_k = -\operatorname{Im} P_k$ sowie $P_{-k} = \overline{P_k}$ für reelle Funktionen P . Mit dem Weierstraß'schen Approximationssatz erhalten wir

Satz 7.2

Jede Funktion $F \in \mathcal{C}^0(\mathbb{T}^p)$ kann gleichmäßig durch trigonometrische Polynome approximiert werden, d. h. es existiert eine Folge trigonometrischer Polynome (P_m) , $m = 1, 2, \dots$, mit $\lim_{m \rightarrow \infty} \|F - P_m\|_0 = 0$.

Damit ist der Banach-Raum $\mathcal{C}^0(\mathbb{T}^p)$ der Abschluss des Raumes $\mathcal{P}(\mathbb{T}^p)$ der trigonometrischen Polynome bezüglich der Norm $\|F\|_0$. In diesem Polynomraum $\mathcal{P}(\mathbb{T}^p)$ können wir nun ein Skalarprodukt einführen. Für die zwei trigonometrischen Polynome

$$P(\theta) = \sum_{|k| \leq N} P_k e^{i(k, \theta)} \quad \text{und} \quad Q(\theta) = \sum_{|k| \leq N} Q_k e^{i(k, \theta)}$$

definieren wir folgendes Produkt mit den üblichen Eigenschaften eines Skalarproduktes

$$\langle P, Q \rangle_0 := \frac{1}{(2\pi)^p} \int_0^{2\pi} \cdots \int_0^{2\pi} \langle P, Q \rangle d\theta_1 \dots d\theta_p = \sum_{|k| \leq N} \langle P_k, Q_{-k} \rangle. \quad (7.4)$$

Mit den Koeffizientenvektoren $P_k = (P_k^1, P_k^2, \dots, P_k^n)$ und $Q_k = (Q_k^1, Q_k^2, \dots, Q_k^n)$ bezeichnet darin

$$\langle P_k, Q_{-k} \rangle := \sum_{j=1}^n P_k^j Q_{-k}^j$$

das bekannte Skalarprodukt zweier Vektoren im Euklidischen Raum \mathbb{R}^n . Mit dem Skalarprodukt (7.4) definieren wir die zugehörige Norm $\|P\|$ in $\mathcal{P}(\mathbb{T}^p)$ mittels

$$\|P\|^2 = \langle P, P \rangle_0 = \frac{1}{(2\pi)^p} \int_0^{2\pi} \cdots \int_0^{2\pi} |P|^2 d\theta_1 \dots d\theta_p = \sum_{|k| \leq N} |P_k|^2. \quad (7.5)$$

Der Abschluss von $\mathcal{P}(\mathbb{T}^p)$ bezüglich dieser Norm liefert uns den Hilbert-Raum $\mathcal{H}^0(\mathbb{T}^p)$, dessen Elemente durch die unendlichen trigonometrischen Reihen

$$F(\theta) = \sum_{k \in \mathbb{Z}^p} F_k e^{i(k, \theta)} \quad \text{mit endlicher Summe} \quad \sum_{k \in \mathbb{Z}^p} |F_k|^2 \quad (7.6)$$

gegeben sind. \mathbb{Z}^p bezeichnet darin die Menge aller Vektoren $k = (k_1, \dots, k_p)$ mit ganzzahligen Komponenten. Nach dem Satz von Riesz-Fischer ist jede Reihe $\sum_k F_k e^{i(k, \theta)}$ mit endlicher Summe $\sum_k |F_k|^2$ die Fourier-Reihe einer quadratisch summierbaren Funktion F . Deshalb kann $\mathcal{H}^0(\mathbb{T}^p)$ mit dem Raum $\mathcal{L}^2(\mathbb{T}^p)$ der auf \mathbb{T}^p quadratisch summierbaren Funktionen identifiziert werden.

Mittels der Torusfunktion führen wir den Begriff der quasi-periodischen Funktion ein:

Definition 7.3 (Quasi-periodische Funktion)

Es sei $\omega = (\omega_1, \dots, \omega_p) \in \mathbb{R}^p$ ein Vektor positiver reeller Zahlen. Die ω_i sind *rational unabhängig (inkommensurabel)*, falls für alle $k \in \mathbb{Z}^p$ mit

$$(k, \omega) = \sum_{l=1}^p k_l \omega_l = 0 \quad \text{stets} \quad k = (0, \dots, 0) \quad \text{gilt.}$$

Eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}^n$ der reellen Veränderlichen t mit der Darstellung

$$f(t) = F(\omega t) = F(\omega_1 t, \dots, \omega_p t),$$

wobei $F \in \mathcal{C}^0(\mathbb{T}^p)$ ist, heißt *quasi-periodische Funktion*. Der Vektor $\omega = (\omega_1, \dots, \omega_p)$ heißt *Frequenzbasis* und die Zahlen $\omega_1, \dots, \omega_p$ heißen *Basisfrequenzen* von f .

Periodische Funktionen sind trivialerweise quasi-periodisch mit einer Basisfrequenz ω_1 und Periode $T = 2\pi/\omega_1$. In den praktisch bedeutsamen Fällen mit zwei Basisfrequenzen sprechen wir von *biperiodischen* Funktionen, beim selteneren Auftreten mit drei Basisfrequenzen von *triperiodischen* Funktionen. Einen Orbit γ eines zeitkontinuierlichen dynamischen Systems wird man als quasi-periodisch mit p Basisfrequenzen bezeichnen, falls eine quasi-periodische Funktion F und p Basisfrequenzen ω_i , $i = 1 \dots p$, existieren, so dass sich der Fluss φ_t in der Form

$$\varphi_t(x_0) = F(\omega t) = F(\omega_1 t, \dots, \omega_p t) \quad (7.7)$$

darstellen lässt. In diesem Falle nennt man den invarianten Torus \mathcal{M} selbst quasi-periodisch. Andernfalls wird \mathcal{M} als *resonanter Torus* bezeichnet (vgl. Abb. 7.2).

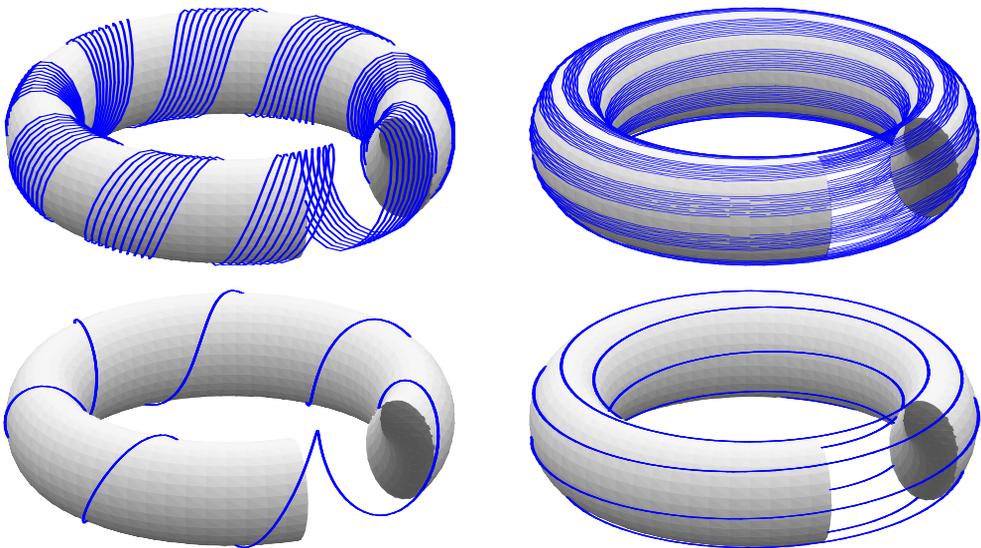


Abb. 7.2 Oben: Quasi-periodische Tori \mathcal{M} mit Basisfrequenzen $\omega_1 = \exp(1), \omega_2 = 3/10$ (links) und $\omega_1 = 3/10, \omega_2 = \exp(1)$ (rechts);

Unten: Resonante Tori \mathcal{M} mit Frequenzen $\omega_1 = 7, \omega_2 = 1$ (links) und $\omega_1 = 1, \omega_2 = 7$ (rechts)

Beispiel 7.4

(a) Mit der Torusfunktion $F(\theta_1, \theta_2, \theta_3)$ aus Beispiel 7.1 und den inkommensurablen Frequenzen $\omega_1 = 1, \omega_2 = \pi, \omega_3 = \sqrt{2}$ ist

$$f(t) = F(\omega_1 t, \omega_2 t, \omega_3 t) = \pi \sin t + \cos(\pi t) \cdot \sin(\sqrt{2} t) - e^{\cos(t - \sqrt{2} t)}$$

eine quasi-periodische (triperiodische) Funktion mit drei Basisfrequenzen $\omega_1, \omega_2, \omega_3$.

(b) Dagegen ist $f(t) = \cos(2.718t) - \sin(3.141592t)$ nicht quasi-periodisch mit zwei Basisfrequenzen (biperiodisch), da die beiden approximierten Frequenzen $\omega_1 = 2.718 \approx e$ und $\omega_2 = 3.141592 \approx \pi$ wegen $(\omega, k) = 392699\omega_1 + (-339750)\omega_2 = 0$ nicht rational unabhängig sind.

(c) Die vektorwertige Funktion $f: \mathbb{R} \rightarrow \mathbb{R}^3$ mit

$$f(t) = (\cos(\sqrt{2} \cdot t) \cdot \sin t, \exp(\cos(\pi \cdot t)) + 1, \sin(\sqrt{8} \cdot t) \cdot \cos t)^T$$

besitzt die drei Basisfrequenzen $\omega_1 = 1$, $\omega_2 = \sqrt{2}$, $\omega_3 = \pi$ und die zugehörige Torusfunktion $F(\theta_1, \theta_2, \theta_3) = (\cos \theta_2 \cdot \sin \theta_1, \exp(\cos \theta_3) + 1, \sin(2\theta_2) \cdot \cos \theta_1)^T$.

(d) Zum Anfangswert $x_0 = (0, 1, 0, 1)^T$ besitzt das DGL-System

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = -x_1, \quad \dot{x}_3 = \pi x_4, \quad \dot{x}_4 = -\pi x_3$$

den Fluss $\varphi_t(x_0) = (\sin t, \cos t, \sin \pi t, \cos \pi t)^T$. Dieser ist biperiodisch wegen der Inkommensurabilität von $\omega_1 = 1$ und $\omega_2 = \pi$ mit der Torusfunktion

$$F(\theta_1, \theta_2) = (\sin \theta_1, \cos \theta_1, \sin \theta_2, \cos \theta_2)^T. \quad \blacksquare$$

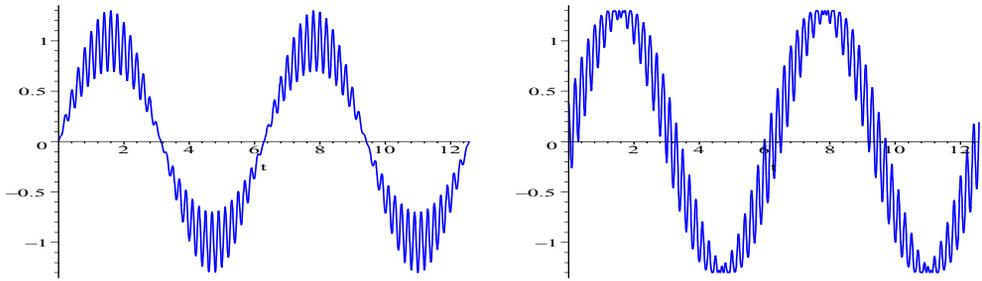


Abb. 7.3 Amplitudenmodulation (links) und Frequenzmodulation (rechts)

Biperiodische Funktionen treten insbesondere bei *Amplitudenmodulation* und *Frequenzmodulation* auf. Mit den Basisfrequenzen $\omega_1 = 1$ und $\omega_2 = 10\pi$ werden in Abbildung 7.3 die Funktionen

■ $f(t) = (1 + 0.3 \cos(\omega_2 t)) \sin(\omega_1 t)$ (Amplitudenmodulation)

■ $g(t) = 1.3 \sin(\omega_1 t + 0.3 \cos(\omega_2 t))$ (Frequenzmodulation)

dargestellt. Gemäß Definition 7.3 sind beide biperiodisch. Wir wollen nun zwei nützliche Eigenschaften quasi-periodischer Funktionen anführen:

Satz 7.5

Die Menge der Werte einer quasi-periodischen Funktion $f(t) = F(\omega t)$ ist *dicht* in der Menge der Werte der Torusfunktion $F(\theta)$.

Insbesondere ist das Supremum von $|f|$ damit gleich der Norm von F , d. h.

$$\sup_{t \in \mathbb{R}} |f(t)| = \max_{\theta \in \mathbb{T}^p} |F(\theta)| = \|F\|_0, \quad (7.8)$$

woraus folgt, dass $f(t) = 0$ für alle $t \in \mathbb{R}$ genau dann gilt, wenn $F(\theta) = 0$ für alle $\theta \in \mathbb{T}^p$ ist. Anschaulich werden quasi-periodische Tori von jedem Orbit γ , der darauf verläuft, dicht gefüllt. Die gesamte Bewegung ist deshalb *ergodisch*, jedoch nicht chaotisch, da sie nicht sensitiv von den Anfangsbedingungen abhängt (vgl. Abb. 7.4). Da jedes $F \in \mathcal{C}^0(\mathbb{T}^p)$ nach Satz 7.2 durch ein trigonometrisches Polynom der Form (7.3) approximiert werden kann, können wir jede quasi-periodische Funktion $f(t) = F(\omega t)$

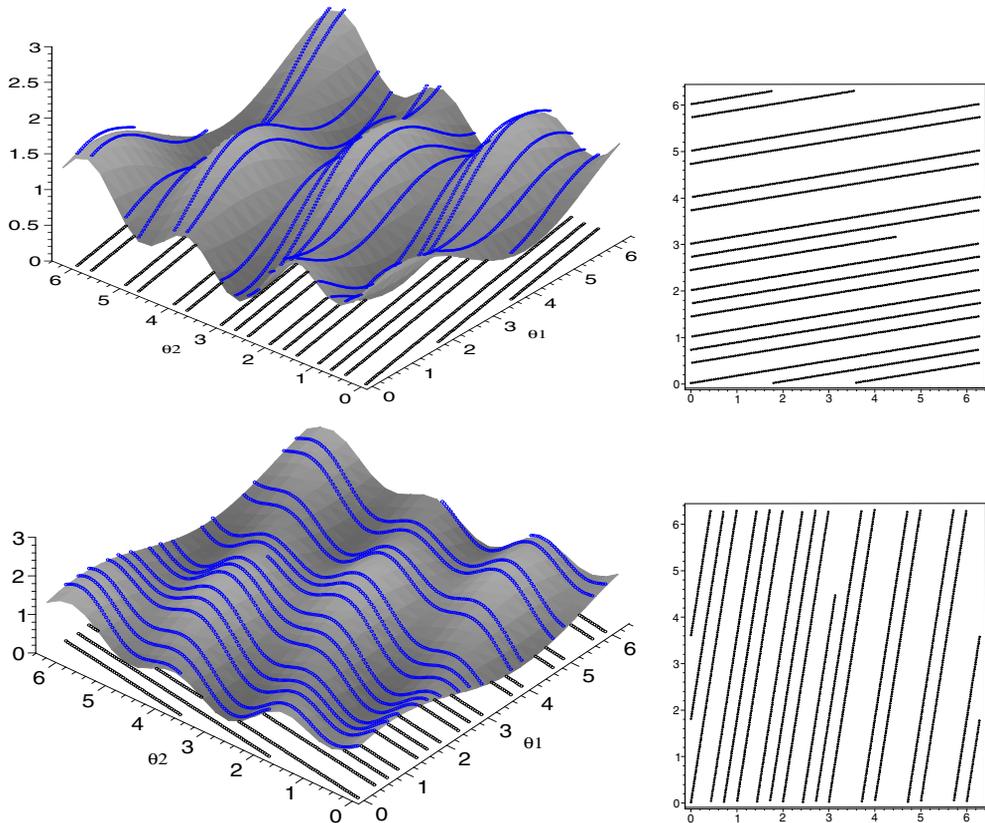


Abb. 7.4 Quasi-periodische Tori \mathcal{M} in (θ_1, θ_2) -Koordinaten mit Orbits γ :
(a) Basisfrequenzen $\omega_1 = 2\pi, \omega_2 = 1$ (oben) und (b) $\omega_1 = 1, \omega_2 = 2\pi$ (unten)

gleichmäßig durch *quasi-trigonometrische Polynome*

$$p_N(t) := P_N(\omega t) = \sum_{|k| \leq N} P_k e^{i(k, \omega)t} \quad (7.9)$$

approximieren, d. h. zu jedem $\varepsilon > 0$ existiert ein Polynom $P_N \in \mathcal{P}(\mathbb{T}^p)$, so dass $\sup_{t \in \mathbb{R}} |f(t) - P_N(\omega t)| < \varepsilon$ ist.

Mit Hilfe quasi-trigonometrischer Polynome lässt sich folgender *Mittelwert-Satz* beweisen:

Satz 7.6

Für jede quasi-periodische Funktion $f(t) = F(\omega t)$ existiert der Grenzwert

$$f_0 := \lim_{T \rightarrow \infty} \frac{1}{T} \int_{\tau}^{T+\tau} f(t) dt = \frac{1}{(2\pi)^p} \int_0^{2\pi} \cdots \int_0^{2\pi} F(\theta) d\theta_1 \dots d\theta_p \quad (7.10)$$

und konvergiert gleichmäßig bezüglich $\tau \in \mathbb{R}$. Der von τ unabhängige Grenzwert f_0 heißt *Mittelwert der Funktion* f .

Die Menge aller quasi-periodischen Funktionen bildet einen Vektorraum, für den wir ebenfalls eine Norm

$$\|f\|_0 := \sup_{t \in \mathbb{R}} |f(t)| \quad \text{mit der Euklidischen Vektornorm } |f|^2 := \sum_{i=1}^n |f_i|^2$$

eingeführen können und den normierten Raum $\mathcal{C}^0(\omega)$ erhalten. Aus der Vollständigkeit des normierten Raumes $\mathcal{C}^0(\mathbb{T}^p)$ folgt zusammen mit der Eigenschaft (7.8), dass $\mathcal{C}^0(\omega)$ ebenfalls ein Banach-Raum ist. Das Skalarprodukt $\langle F, G \rangle_0$ zweier Torusfunktionen $F, G \in \mathcal{C}^0(\mathbb{T}^p)$ induziert mittels (7.10) ein entsprechendes Skalarprodukt $\langle f, g \rangle_0$ der quasi-periodischen Funktionen $f(t) = F(\omega t)$ und $g(t) = G(\omega t)$

$$\langle f, g \rangle_0 = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \langle f(t), g(t) \rangle dt = \langle F, G \rangle_0 \quad (7.11)$$

sowie eine Norm $\|f\| := \sqrt{\langle f, f \rangle_0}$. Die Funktionen $e^{i(k, \omega)t}$, $k \in \mathbb{Z}^p$, bilden dann ein vollständiges Orthonormalsystem in $\mathcal{C}^0(\omega)$. Jede Funktion $f \in \mathcal{C}^0(\omega)$ kann durch eine Fourier-Reihe

$$f(t) = \sum_{k \in \mathbb{Z}^p} f_k e^{i(k, \omega)t} \quad \text{mit Koeffizienten } f_k := \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(t) e^{-i(k, \omega)t} dt$$

dargestellt werden, die die Parseval'sche Gleichung

$$\sum_{k \in \mathbb{Z}^p} |f_k|^2 = \|f\|^2 \quad (7.12)$$

erfüllt. Auch im Banach-Raum $\mathcal{C}^0(\omega)$ können wir den Unterraum $\mathcal{C}^r(\omega)$ aller quasi-periodischen Funktionen f konstruieren, für die eine Torusfunktion $F \in \mathcal{C}^r(\mathbb{T}^p)$, $r \geq 0$, mit $f(t) = F(\omega t)$ existiert. Jede Funktion $f \in \mathcal{C}^r(\omega)$ ist dann r -mal stetig differenzierbar in $t \in \mathbb{R}$ und ihre Ableitungen sind quasi-periodisch. Führen wir auch hier eine Norm

$$\|f\|_r := \|F\|_r \quad \text{für } f(t) = F(\omega t) \quad \text{mit } F \in \mathcal{C}^r(\mathbb{T}^p)$$

ein, so ist der Raum $C^r(\omega)$ ebenfalls vollständig. Betrachten wir schließlich eine Funktion $F(\theta) = \sum_k F_k e^{i(k,\theta)}$ aus dem Hilbert-Raum $\mathcal{H}^0(\mathbb{T}^p)$. Mit einer gegebenen Frequenzbasis $\omega = (\omega_1, \dots, \omega_p)$ bilden wir die quasi-periodische Funktion $f(t) = F(\omega t)$ mit der Reihe

$$f(t) = \sum_{k \in \mathbb{Z}^p} F_k e^{i(k,\omega)t}.$$

Diese Funktionen bilden einen Vektorraum, den wir mit $\mathcal{H}^0(\omega)$ bezeichnen. Zu zwei Funktionen f, g mit Fourier-Reihen $f(t) = \sum_k F_k e^{i(k,\omega)t}$ und $g(t) = \sum_k G_k e^{i(k,\omega)t}$ können wir Skalarprodukt und Norm durch

$$\langle f, g \rangle_0 := \sum_{k \in \mathbb{Z}^p} \langle F_k, G_{-k} \rangle, \quad \|f\|^2 = \langle f, f \rangle_0$$

definieren. Für trigonometrische Polynome $P(\theta) = \sum_{|k| \leq N} P_k e^{i(k,\theta)}$ aus $\mathcal{P}(\mathbb{T}^p)$ bezeichne $\mathcal{P}(\omega)$ den Raum der univariaten Polynome

$$p(t) = \sum_{|k| \leq N} P_k e^{i(k,\omega)t} \quad \text{zur Frequenzbasis } \omega = (\omega_1, \dots, \omega_p).$$

Da $\mathcal{H}^0(\omega)$ bezüglich der Norm $\|f\|$ die Vervollständigung des Polynomraumes $\mathcal{P}(\omega)$ bildet, ist $\mathcal{H}^0(\omega)$ ebenfalls ein Hilbert-Raum.

Damit stehen uns nun die Banach-Räume $C^r(\mathbb{T}^p)$, $r \geq 0$, und der Hilbert-Raum $\mathcal{H}^0(\mathbb{T}^p)$ für Torusfunktionen zur Verfügung. Zu gegebener Frequenzbasis ω konnten wir daraus entsprechende Räume $C^r(\omega)$ und $\mathcal{H}^0(\omega)$ quasi-periodischer Funktionen ableiten.

7.1.2 Quasi-periodische Bewegungen

Nach Einführung der Grundbegriffe wenden wir uns nun wieder der Betrachtung autonomer DGL-Systeme

$$\dot{x} = f(x), \quad f : D \rightarrow \mathbb{R}^n \quad \text{mit } D \subset \mathbb{R}^n, D \text{ offen}, \quad (7.13)$$

zu und setzen voraus, dass eine hinreichend glatte quasi-periodische Lösung

$$x = x(t, x_0) = F(\omega t) \quad \text{mit } F \in C^r(\mathbb{T}^p), r \geq 1, \quad (7.14)$$

durch x_0 existiert, deren Orbit $\gamma = \{x \mid x = x(t, x_0), t \in \mathbb{R}\}$ kompakt in D ist. Nach Satz 7.5 gehört der Abschluss von γ zur Menge D und besteht aus den Punkten, die durch die Darstellung

$$x = F(\theta) \quad \text{mit } \theta \in \mathbb{T}^p \quad (7.15)$$

beschrieben werden. Das dadurch erzeugte Bild des Standardtorus \mathbb{T}^p werden wir mit \mathcal{M} bezeichnen. Welche Eigenschaften besitzt diese Menge?

1. Die Menge \mathcal{M} ist *invariant* bezüglich des Systems (7.13), d. h. sie besteht aus Punkten von Orbits dieses Systems. Setzen wir dazu die Lösung (7.26) in die DGL ein, so liefert die Integration

$$F(\omega t) = F(0) + \int_0^t f(F(\omega\tau)) d\tau, \quad t \in \mathbb{R}.$$

Betrachten wir eine Funktionenfolge $F(\omega t + \omega t_n)$, $n = 1, 2, \dots$ mit dem Grenzwert $\lim_{n \rightarrow \infty} \omega t = \theta \pmod{2\pi}$, so ergibt sich

$$F(\omega t + \theta) = F(\theta) + \int_0^t f(F(\omega\tau + \theta)) d\tau,$$

woraus $x(t, F(\theta)) = F(\omega t + \theta)$ für $t \in \mathbb{R}$, $\theta \in \mathbb{T}^p$ folgt. Also ist die Menge \mathcal{M} invariant bezüglich (7.13).

2. Wegen (7.26) gehört die Lösung $x(t, x_0)$ zu $\mathcal{C}^r(\omega)$. Angenommen, es kann eine zweite Frequenzbasis $\Omega = (\Omega_1, \Omega_2, \dots, \Omega_q)$ mit $q < p$ Basisfrequenzen gefunden werden, so dass $x(t, x_0) \in \mathcal{C}^0(\Omega)$ mit einer entsprechenden Torusfunktion gilt. Dann lässt sich nachweisen, dass die Lösung auch zu $\mathcal{C}^r(\Omega)$ gehört. Für den Fall, dass dies nicht gelingt, führen wir folgenden Begriff ein:

Definition 7.7 (Reale Basisdimension)

Es sei $x(t, x_0) \in \mathcal{C}^0(\omega)$ mit einer Basis $\omega = (\omega_1, \omega_2, \dots, \omega_p)$. Ist jedoch $x(t, x_0) \notin \mathcal{C}^0(\Omega)$ für jede Basis $\Omega = (\Omega_1, \Omega_2, \dots, \Omega_q)$ mit $q < p$, so heißt $p = p^*$ die *reale Dimension der Frequenzbasis* der quasi-periodischen Lösung $x(t, x_0)$.

Die reale Dimension p^* beschreibt zugleich die Dimension der Menge \mathcal{M} in jedem ihrer Punkte. Dies kann folgendermaßen formuliert werden (vgl. Samoilenko (1991), S. 72ff.):

Satz 7.8

Es sei $x(t, x_0) = F(\omega t) \in \mathcal{C}^r(\omega)$ mit $r \geq 1$ und $\omega = (\omega_1, \omega_2, \dots, \omega_p)$ eine Frequenzbasis mit p Frequenzen. Dann ist

$$\text{rang } \frac{\partial F(\theta)}{\partial \theta} = p^*, \quad \theta \in \mathbb{T}^p, \quad (7.16)$$

worin $p^* \leq p$ die reale Dimension der Frequenzbasis von $x(t, x_0)$ ist.

Insbesondere ist die reale Dimension p^* kleiner als die Dimension n des Phasenraumes und gibt die Maximalzahl linear unabhängiger Spalten der Jacobi-Matrix $\frac{\partial F(\theta)}{\partial \theta}$ an.

Beispiel 7.9

Beispiel 7.4(d) lieferte zum Anfangswert $x_0 = (0, 1, 0, 1)^T$ die quasi-periodische Lösung $x(t, x_0) = (\sin t, \cos t, \sin \pi t, \cos \pi t)^T$. Mit der Frequenzbasis $\omega = (\omega_1, \omega_2) = (1, \pi)$ lautet die Torusfunktion

$$F(\theta) = F(\theta_1, \theta_2) = (\sin \theta_1, \cos \theta_1, \sin \theta_2, \cos \theta_2)^T.$$

Die Jacobi-Matrix

$$\frac{\partial F(\theta)}{\partial \theta} = \begin{pmatrix} \cos \theta_1 & 0 \\ -\sin \theta_1 & 0 \\ 0 & \cos \theta_2 \\ 0 & -\sin \theta_2 \end{pmatrix}$$

besitzt für alle $\theta = (\theta_1, \theta_2) \in \mathbb{T}^2$ den Rang 2, womit $p^* = 2$ die reale Dimension der Frequenzbasis ist. ■

3. Die Abbildung $F : \theta \rightarrow F(\theta)$ definiert nicht notwendig einen Homöomorphismus zwischen dem Standardtorus \mathbb{T}^p und der Mannigfaltigkeit \mathcal{M} . Das erkennt man leicht, wenn $F \in \mathcal{C}^0(\mathbb{T}^p)$ so gewählt wird, dass es bezüglich der Variablen θ_1 die Periodendauer $2\pi/m$ mit ganzzahligem $m \geq 2$ hat. Dann liefern zu $\theta^0 = (\theta_1^0, \theta_2^0, \dots, \theta_p^0) = (2\pi/m, 0, 0, \dots, 0)$ die m verschiedenen Punkte $\theta^\mu = \theta + \mu\theta^0$, $\mu = 1, 2, \dots, m$, ein und denselben Bildpunkt $x = F(\theta^\mu) = F(\theta)$ auf \mathcal{M} . Die inverse Abbildung ist somit nicht eindeutig, da die Frequenzbasis von $x(t, x_0)$ auf verschiedene Art gewählt werden kann. Wir führen deshalb den Begriff der maximalen Frequenzbasis ein:

Definition 7.10 (Maximale Frequenzbasis)

Es sei $x(t, x_0) \in \mathcal{C}^0(\omega)$ mit einer Frequenzbasis $\omega = (\omega_1, \omega_2, \dots, \omega_p)$. Dann ist ω *maximal*, wenn für jede andere Frequenzbasis $\Omega = (\Omega_1, \Omega_2, \dots, \Omega_q)$ eine (p, q) -Matrix P mit ganzzahligen Elementen existiert, so dass

$$\omega = P\Omega \quad \text{mit den Spaltenvektoren } \omega, \Omega$$

und $\text{rang } P = p$ gilt.

Die Anzahl p^* der Frequenzen in einer maximalen Basis ist offenbar gleich der Dimension der realen Frequenzbasis der Lösung $x(t, x_0)$. Sind zwei Frequenzbasen ω und Ω maximal, so sind P und P^{-1} ganzzahlige Matrizen. Im periodischen Spezialfall besteht die maximale Basis aus der größten Frequenz $\omega = 2\pi/T$ der Lösung $x(t, x_0)$, die durch die kleinste Periode T bestimmt wird. Die Frage nach der Existenz einer maximalen Frequenzbasis und der Eigenschaften von F beantwortet

Satz 7.11

Für jede quasi-periodische Lösung $x(t, x_0) = F(\omega t) \in C^0(\omega)$ der DGL (7.13) gilt:

- i. Die Funktion $x(t, x_0)$ besitzt eine maximale Frequenzbasis.
- ii. Die Abbildung $F : \theta \rightarrow F(\theta)$ definiert einen Homöomorphismus genau dann, wenn die Frequenzbasis $\omega = (\omega_1, \omega_2, \dots, \omega_p)$ maximal ist.
- iii. Ist darüber hinaus $x(t, x_0) \in C^r(\omega)$ mit $r \geq 0$, so ist die Menge \mathcal{M} (der Abschluss des Orbits durch x_0) C^r -homöomorph zu einem p -Torus, wobei p die reale Dimension der Frequenzbasis der Lösung $x(t, x_0)$ ist.

Die Menge \mathcal{M} , die C^r -homöomorph zu einem p -Torus ist, werden wir als p -dimensionale *toroidale Mannigfaltigkeit mit C^r -Glattheit r* bezeichnen bzw. kürzer als C^r -glatten p -Torus.

7.1.3 Transformation in Toruskoordinaten

Wir wollen nun voraussetzen, dass das autonome DGL-System (7.13) eine quasi-periodische Lösung

$$x(t, x_0) = F(\omega t) \quad \text{mit Torusfunktion } F \in C^{s+1}(\mathbb{T}^p), s \geq 0, \quad (7.17)$$

besitzt, die der Rangbedingung (7.16) mit $p^* = p$ genügt. Dann stellt p die reale Dimension der Frequenzbasis $\omega = (\omega_1, \omega_2, \dots, \omega_p)$ dar. Nach Satz 7.11 kann daraus eine maximale Basis gewonnen werden. Um die p -dimensionale Torusmannigfaltigkeit global in der Form

$$\mathcal{M} = \{x \mid x = F(\theta), \theta \in \mathbb{T}^p\} \quad (7.18)$$

parametrisieren zu können, gehen wir von kartesischen Koordinaten $x = (x_1, \dots, x_n)$ zu den p Toruskoordinaten $\theta = (\theta_1, \dots, \theta_p)$ und den q lokalen (radialen) Koordinaten $u = (u_1, \dots, u_q)$ mit $p + q = n$ über. (θ, t) werden wir desweiteren als *Toruskoordinaten* bezeichnen. Zahlreiche bekannte Koordinatentransformationen lassen sich dann in der allgemeinen Form

$$x = B(\theta)u + b(\theta) \quad (7.19)$$

mit hinreichend glatten und bezüglich jeder θ -Komponente 2π -periodischen Abbildungen $b : \mathbb{T}^p \rightarrow \mathbb{R}^n$ und $B : \mathbb{T}^p \rightarrow \mathbb{R}^{n \times q}$ darstellen. Typisch ist dies in folgenden Szenarien:

- Durchläuft das DGL-System eine Torus-Bifurkation, so garantiert das Bifurkations-Theorem 6.49 von Neimark-Sacker die Existenz eines invarianten Torus \mathcal{M} , wie dies in Abschnitt 6.4 veranschaulicht wurde. Der Orbit der instabilen periodischen Lösung kann durch $b(\theta)$ beschrieben werden. In Abbildung 7.5 wird der dreidimensionale Fall mit dem Poincaré-Schnitt Σ veranschaulicht.

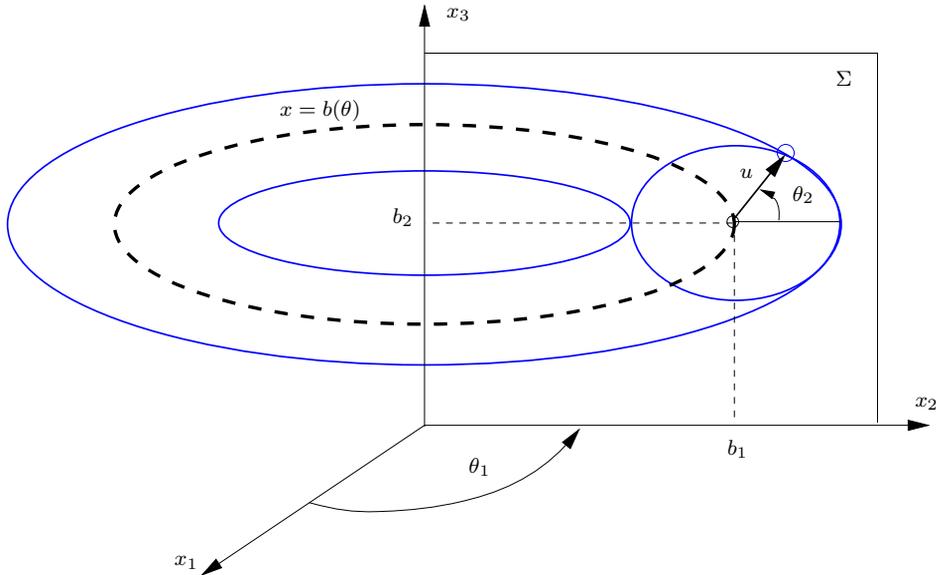


Abb. 7.5 Koordinatentransformation $x = B(\theta)u + b(\theta)$ in \mathbb{R}^3 : kartesische Koordinaten $x = (x_1, x_2, x_3)$ und Toruskordinaten (θ, u) mit $\theta = (\theta_1, \theta_2)$, $u = (u_1)$

- Wird eine Toruslösung bezüglich eines Parameters λ fortgesetzt, so steht für λ_0 ein invarianter „Referenztorus“ \mathcal{M}_0 mit Parametrisierung $b(\theta)$ zur Verfügung, dessen Normalraum $N_{(\theta, u)}\mathcal{M}_0$ durch die q Spaltenvektoren von $B(\theta)$ aufgespannt wird. Der zum Parameter λ gesuchte benachbarte Torus \mathcal{M} kann nun durch (7.19) parametrisiert werden, wie dies der Poincaré-Schnitt Σ in Abbildung 7.6 für $n = 3$ zeigt.

Wir suchen nun eine Darstellung des invarianten p -Torus \mathcal{M} mit der Torusfunktion $u: \mathbb{T}^p \rightarrow \mathbb{R}^q$. Wenden wir die Transformation (7.19) auf eine Lösung der DGL an, d. h. setzen wir $x(t) = B(\theta(t))u(t) + b(\theta(t))$, so geht bei Differenziation in (θ, u) -Koordinaten

$$\mathcal{M} = \{(\theta, u) \mid u = u(\theta), \theta \in \mathbb{T}^p, u \in \mathbb{R}^q\} \subset \mathbb{T}^p \times \mathbb{R}^q \quad (7.20)$$

nach t die DGL (7.13) in die Form

$$(B'(\theta)u + b'(\theta), B(\theta)) \begin{pmatrix} \frac{d\theta}{dt} \\ \frac{du}{dt} \end{pmatrix} = f(B(\theta)u + b(\theta))$$

über, woraus wir unter der zusätzlichen Voraussetzung $\det(B'(\theta)u + b'(\theta), B(\theta)) \neq 0$ das transformierte System

$$\begin{pmatrix} \frac{d\theta}{dt} \\ \frac{du}{dt} \end{pmatrix} = [B'(\theta)u + b'(\theta), B(\theta)]^{-1} f(B(\theta)u + b(\theta)),$$

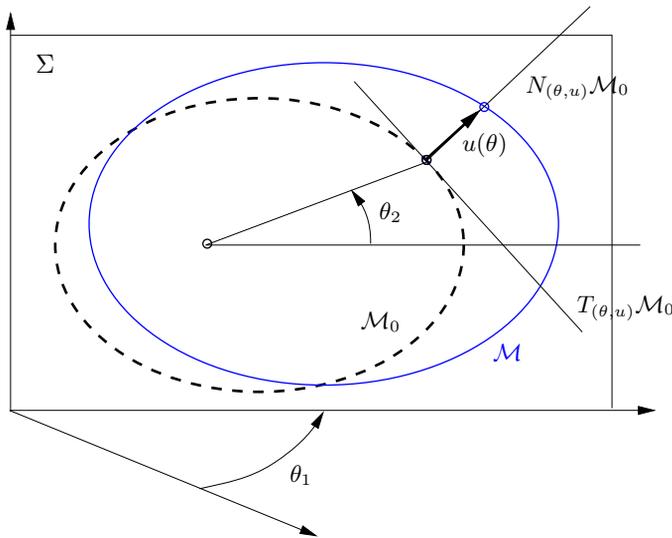


Abb. 7.6 Koordinatentransformation in \mathbb{R}^3 : Fortsetzung des Referenztorus \mathcal{M}_0 (schwarz) zu \mathcal{M} (blau) mit Toruskordinaten (θ, u) mit $\theta = (\theta_1, \theta_2)$, $u = (u_1)$; die Abbildung zeigt den Poincaré-Schnitt Σ

erhalten. In partitionierter Form lautet die DGL in Toruskordinaten $(\theta, u) \in \mathbb{T}^p \times \mathbb{R}^q$

$$\begin{aligned} \frac{d\theta}{dt} &= \Omega(\theta, u) & \text{mit} & \quad \Omega(\theta, u) := L_1(\theta, u)f(B(\theta)u + b(\theta)) \\ \frac{du}{dt} &= R(\theta, u) & \text{mit} & \quad R(\theta, u) := L_2(\theta, u)f(B(\theta)u + b(\theta)). \end{aligned} \tag{7.21}$$

$L_1(\theta, u)$ bezeichnet dabei die aus den ersten p Zeilen der Inversen $C(\theta, u)^{-1}$ von

$$C(\theta, u) := (B'(\theta)u + b'(\theta), B(\theta)) \quad \text{mit} \quad \det(C(\theta, u)) \neq 0 \tag{7.22}$$

bestehende Matrix und $L_2(\theta, u)$ entsprechend die aus den verbleibenden q Zeilen von $C(\theta, u)^{-1}$ bestehende Matrix. Die vorausgesetzte 2π -Periodizität von $B(\theta)$, $b(\theta)$ und $u(\theta)$ bezüglich jeder θ -Komponente überträgt sich unmittelbar auf $\Omega(\theta, u)$ und $R(\theta, u)$. Wegen der Regularität von $C(\theta, u)$ sind diese Funktionen zudem C^s -glatt, wenn $B(\theta)$ und $b(\theta)$ als C^{s+1} -glatt vorausgesetzt werden. Die Regularität der $n \times n$ -Matrix $C(\theta, u)$ auf $\mathbb{T}^p \times D$, $D \subset \mathbb{R}^q$ muss allerdings im konkreten Fall stets verifiziert werden.

Beispiel 7.12

Betrachten wir das System zweier schwach linear gekoppelter Van-der-Pol-Oszillatoren

$$\begin{aligned} \frac{d^2 x_1}{dt^2} - e(1 - x_1^2) \frac{dx_1}{dt} + x_1 &= b(x_2 - x_1) \\ \frac{d^2 x_2}{dt^2} - e(1 - x_2^2) \frac{dx_2}{dt} + (1 + d)x_2 &= b(x_1 - x_2) \end{aligned}$$

mit Dämpfung $e > 0$, Verzerrung $d \geq 0$ und mit $b \geq 0$ als Parameter für die lineare Kopplung der beiden Systeme. In Kapitel 2 wird gezeigt, dass bei fehlender Kopplung

($b = 0$) jeder der beiden Van-der-Pol-Oszillatoren einen asymptotisch stabilen Grenzzyklus S_1 bzw. S_2 mit Basisfrequenzen ω_1 bzw. ω_2 besitzt. Das kartesische Produkt $S_1 \times S_2$ der beiden Grenzzyklen ist ein (asymptotisch stabiler) 2-dimensionaler Torus $\mathcal{M} \subseteq \mathbb{R}^4$. Die Bewegung auf \mathcal{M} ist quasi-periodisch bei inkommensurablen Frequenzen ω_1, ω_2 , andernfalls ist sie periodisch. Das entkoppelte System ($b = 0$) ist struktur stabil in \mathbb{R}^4 (vgl. Broer et al. (1996)), weshalb der invariante Torus bei kleinen Störungen wie z. B. einer schwachen Kopplung oder einer Vergrößerung der Verzerrung d erhalten bleibt. Nach Überführung in ein DGL-System 1. Ordnung

$$\begin{aligned} \dot{x}_1 &= y_1 \\ \dot{y}_1 &= -x_1 + b(x_2 - x_1) + e(1.0 - x_1^2)y_1 \\ \dot{x}_2 &= y_2 \\ \dot{y}_2 &= -x_2 - dx_2 + b(x_1 - x_2) + e(1.0 - x_2^2)y_2 \end{aligned} \quad (7.23)$$

bieten sich deshalb Polarkoordinaten zur Transformation in das (θ, u) -System mit

$$x_i = u_i \cos \theta_i, \quad y_i = -u_i \sin \theta_i, \quad i = 1, 2 \quad (7.24)$$

an, wobei $u_1 > 0$, $u_2 > 0$ gilt. Die Transformation (7.19) lautet dann mit $b(\theta) = 0$

$$x = B(\theta)u + b(\theta) = \begin{pmatrix} \cos \theta_1 & 0 \\ -\sin \theta_1 & 0 \\ 0 & \cos \theta_2 \\ 0 & -\sin \theta_2 \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \end{pmatrix},$$

woraus wir die Matrixfunktion

$$C(\theta, u) = (B'(\theta)u + b'(\theta), B(\theta)) = \begin{pmatrix} -u_1 \sin \theta_1 & 0 & \cos \theta_1 & 0 \\ -u_1 \cos \theta_1 & 0 & -\sin \theta_1 & 0 \\ 0 & -u_2 \sin \theta_2 & 0 & \cos \theta_2 \\ 0 & -u_2 \cos \theta_2 & 0 & -\sin \theta_2 \end{pmatrix}$$

erhalten. Deren Determinante ist $\det(C(\theta, u)) = -u_1 u_2 < 0$ für alle $(\theta_1, \theta_2) \in \mathbb{T}^2$. Das partitionierte System (7.21) mit $p = q = 2$ bekommt dann die recht komplizierte Form

$$\begin{aligned} \dot{\theta}_1 &= 1.0 + \frac{b}{u_1} c_{12} \cos \theta_1 + \frac{e}{2} c_1 \sin 2\theta_1 &=: \Omega_1(\theta_1, \theta_2, u_1, u_2) \\ \dot{\theta}_2 &= 1.0 + d \cos^2 \theta_2 - \frac{b}{u_2} c_{12} \cos \theta_2 + \frac{e}{2} c_2 \sin 2\theta_2 &=: \Omega_2(\theta_1, \theta_2, u_1, u_2) \\ \dot{u}_1 &= bc_{12} \sin \theta_1 + ec_1 u_1 \sin^2 \theta_2 &=: R_1(\theta_1, \theta_2, u_1, u_2) \\ \dot{u}_2 &= \frac{d}{2} u_2 \sin 2\theta_2 - bc_{12} \sin \theta_2 + ec_2 u_2 \sin^2 \theta_2 &=: R_2(\theta_1, \theta_2, u_1, u_2) \end{aligned}$$

mit den Abkürzungen $c_{12} := u_1 \cos \theta_1 - u_2 \cos \theta_2$, $c_1 := 1.0 - u_1^2 \cos^2 \theta_1$ und $c_2 := 1.0 - u_2^2 \cos^2 \theta_2$. ■

Bemerkung 7.13 (Existenz und Eindeutigkeit)

In der Standardliteratur sind relativ wenige Aussagen über die *Existenz und Eindeutigkeit* von Toruslösungen dynamischer Systeme zu finden. Für lineare Systeme bezüglich u in der partitionierten Form

$$\frac{d\theta}{dt} = a(\theta), \quad \frac{du}{dt} = P(\theta)u + f(\theta) \quad (7.25)$$

mit $a, P, f \in C^r(\mathbb{T}^p)$, $r \geq 1$, und $\theta = (\theta_1, \dots, \theta_p)$, $u = (u_1, \dots, u_q)$ werden z. B. bei Samoilenko (1991) notwendige und hinreichende Bedingungen für die Existenz eines invarianten Torus bewiesen. Mit Hilfe der Störungstheorie werden dort auch nichtlineare Systeme der Gestalt

$$\frac{d\theta}{dt} := a(\theta, u, \varepsilon), \quad \frac{du}{dt} = P(\theta, u, \varepsilon)u + f(\theta, \varepsilon) \quad (7.26)$$

betrachtet, wobei a, P und f periodische Funktionen in θ_j , $j = 1(1)p$, mit Periode 2π sind, die stetig bezüglich (θ, u, ε) im Bereich $\|u\|_0 \leq d$, $\theta \in \mathbb{T}^p$, $\varepsilon \in [0, \varepsilon_0]$ sind. Weiterhin gelte $f(\theta, 0) = 0$ für alle $\theta \in \mathbb{T}^p$, d. h. es existiere ein trivialer invarianter Torus $u(\theta) = 0$, $\theta \in \mathbb{T}^p$, von System (7.26) für $\varepsilon = 0$. Dann garantiert Samoilenko (1991) unter zusätzlichen Voraussetzungen an a, P, f und an das für $\varepsilon = 0$ entstehende DGL-System die Existenz eines invarianten Torus für alle $\varepsilon \in [0, \varepsilon_0]$ mit hinreichend kleinem ε_0 . Diese Aussage kann z. B. bei schwach gekoppelten Schwingungssystemen mit hinreichend kleinem Kopplungsparameter ε Anwendung finden, wobei die Überprüfung der entsprechenden Voraussetzungen in der Praxis schwierig sein dürfte. ■

Ein einfaches Beispiel gekoppelter Schwingungssysteme entsteht, wenn zwei nichtlineare DGL mit periodischen Lösungen wie bei Lorenz (1991) betrachtet werden. Das dynamische System in (θ, u) -Koordinaten

$$\frac{d\theta}{dt} = \omega, \quad \frac{du}{dt} = u(1 - u^2) \quad \text{mit Frequenz } \omega > 0 \quad (7.27)$$

besitzt zu beliebigem $u(0) > 0$ eine Lösung mit $\lim_{t \rightarrow \infty} u(t) = 1$. Also ist die Menge

$$\mathcal{M} = \{(\theta, u) \mid u = u(\theta) = 1, \theta \in \mathbb{S}^1\} \subset \mathbb{S}^1 \times \mathbb{R}$$

eine attraktive invariante Mannigfaltigkeit dieser DGL. Wir koppeln nun zwei Systeme des Typs (7.27) und führen dazu eine Kopplungsfunktion $g : \mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{R}^2 \rightarrow \mathbb{R}^4$ ein, die als C^∞ -glatt vorausgesetzt wird. Mit dem Kopplungsparameter ε erhalten wir die Drastellung in partitionierter Form

$$\begin{aligned} \dot{\theta}_1 &= \omega_1 & + & \varepsilon g_1(\theta_1, \theta_2, u_1, u_2) \\ \dot{\theta}_2 &= \omega_2 & + & \varepsilon g_2(\theta_1, \theta_2, u_1, u_2) \\ \dot{u}_1 &= u_1(1 - u_1^2) & + & \varepsilon g_3(\theta_1, \theta_2, u_1, u_2) \\ \dot{u}_2 &= u_2(1 - u_2^2) & + & \varepsilon g_4(\theta_1, \theta_2, u_1, u_2). \end{aligned} \quad (7.28)$$

Mit den *Toruskoordinaten* $\theta = (\theta_1, \theta_2)$ und $u = (u_1, u_2)$ besitzt dieses System bei fehlender Kopplung ($\varepsilon = 0$) die invariante Mannigfaltigkeit

$$\mathcal{M}_0 = \{(\theta, u) \mid u_1(\theta) = u_2(\theta) = 1, \theta \in \mathbb{T}^2\} \subset \mathbb{T}^2 \times \mathbb{R}^2 \quad (7.29)$$

als kartesisches Produkt der zwei invarianten Einheitskreise. \mathcal{M}_0 ist damit diffeomorph zum Standard-2-Torus $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$. Orbits auf der Mannigfaltigkeit sind quasi-periodisch, falls die Basisfrequenzen ω_1 und ω_2 rational unabhängig sind; andernfalls liegt ein periodischer Orbit vor.

Was geschieht nun mit dem invarianten Torus \mathcal{M}_0 bei vorhandener Kopplung? Für kleine Kopplung $|\varepsilon| \leq \varepsilon_s$ liefert die analytische Störungstheorie eine Existenzaussage für den gestörten Torus \mathcal{M}_ε (vgl. Lorenz (1991)):

Satz 7.14

Zu jedem $s \in \{1, 2, 3, \dots\}$ existieren eine Schranke $\varepsilon_s > 0$ und eine \mathcal{C}^s -glatte Funktion $u_s : \mathbb{T}^2 \times (-\varepsilon_s, \varepsilon_s) \rightarrow \mathbb{R}^2$, mit der die Mannigfaltigkeit

$$\mathcal{M}_\varepsilon = \{(\theta, u) \mid u = u_s(\theta, \varepsilon), \theta \in \mathbb{T}^2, |\varepsilon| \leq \varepsilon_s\}$$

invariant unter dem System (7.28) für $|\varepsilon| \leq \varepsilon_s$ ist.

Bei weiterem Anwachsen der Kopplung $|\varepsilon|$ muss allerdings mit dem Zusammenbruch des Torus gerechnet werden, der dann in der Regel in eine periodische Bewegung übergeht. Um die Klasse der numerisch zu behandelnden Probleme nicht unnötig einzuschränken, wollen wir für allgemeine nichtlineare Systeme in partitionierter Form (7.21) die folgende Voraussetzung treffen:

Voraussetzung 7.15 (Existenz)

System (7.21) besitzt einen lokal eindeutigen invarianten p -Torus \mathcal{M} , der in der Form (7.20) mit einer $\mathcal{C}^s(\mathbb{T}^p)$ -Funktion $u(\theta)$ parametrisierbar ist ($s \geq 2$).

Für derartige p -Tori lassen sich Stabilitätsbegriffe einführen, die sich an der Ljapunov-Stabilität von Orbits orientieren, nun aber den gesamten Torus \mathcal{M} betreffen.

Die euklidische Länge des Vektors $\theta = (\theta_1, \dots, \theta_p)^T$ auf dem Torus \mathbb{T}^p sei durch

$$\|\theta\|_{\mathbb{T}^p} = \min \|\theta \bmod 2\pi\| \quad (7.30)$$

definiert, und (θ_t, u_t) sei die Lösung des partitionierten Systems (7.21) mit den Anfangswerten $\theta(0) = \theta_0$ und $u(0) = u_0$. Den Abstand d_t eines Punktes (θ_t, u_t) vom Torus \mathcal{M} in $\mathbb{T}^p \times \mathbb{R}^q$ führt Samoilenko (1991) in der Form

$$d_t := \min_{\theta \in \mathbb{T}^p} \{\|\theta_t - \theta\|_{\mathbb{T}^p}^2 + \|u_t - u(\theta)\|^2\}^{\frac{1}{2}} \quad (7.31)$$

ein. Damit lassen sich folgende Begriffe definieren:

Definition 7.16 (Stabilität von Tori)

- i. Ein invarianter Torus von (7.21) heißt *stabil*, wenn zu jedem $\varepsilon > 0$, $\varepsilon \in \mathbb{R}$, ein $\delta = \delta(\varepsilon) > 0$, $\delta \in \mathbb{R}$, existiert, so dass für alle Lösungen (θ_t, u_t) von (7.21) mit $d_0 < \delta$ gilt: $d_t < \varepsilon$ für alle $t \geq 0$.
- ii. Ein invarianter Torus heißt *asymptotisch stabil*, wenn er stabil ist und ein $\delta_0 > 0$, $\delta_0 \in \mathbb{R}$, existiert, so dass für alle Lösungen (θ_t, u_t) von (7.21) mit $d_0 < \delta_0$ gilt: $d_t < \delta_0$ für alle $t \geq 0$ und

$$\lim_{t \rightarrow +\infty} d_t = 0. \quad (7.32)$$

Ein invarianter Torus der nicht stabil ist, wird als instabil (unstable) bezeichnet. Eine genauere Aussage bezüglich der Konvergenzgeschwindigkeit in (7.32) liefert folgende

Definition 7.17 (Exponentielle Stabilität)

Ein invarianter Torus heißt *exponentiell stabil*, falls er stabil ist und ein $\delta_1 > 0$ existiert, so dass für jede Lösung (θ_t, u_t) von (7.21) mit $d_0 < \delta_1$ die Ungleichung

$$d_t \leq K \cdot e^{-\gamma(t-t_0)} d_{t_0}, \quad t \geq t_0 \quad (7.33)$$

für positive, von t_0 und d_0 unabhängige Konstanten γ und K und $t_0 \in [0, \infty)$ erfüllt ist.

Aus dieser Definition folgt sofort, dass jeder exponentiell stabile Torus auch asymptotisch stabil ist. Stabilitätsnachweise anhand dieser Definitionen sind bei konkreten DGL-Systemen (7.21) meist nicht möglich. Die in den vorhergehenden Kapiteln benutzten *algebraischen Stabilitätskriterien* mittels der Eigenwerte der

- Jacobi-Matrix im Falle von Gleichgewichtslösungen (Kapitel 5) bzw.
- Monodromiematrix (Floquet-Multiplikatoren) für periodische Lösungen (Kapitel 6)

stehen bei invarianten p -Tori allerdings nicht zur Verfügung.

7.2 Parametrisierung invarianter Tori

Nachfolgend wird eine Bestimmungsgleichung für invariante p -Tori hergeleitet, die es gestattet, diese Mannigfaltigkeiten direkt und weitgehend unabhängig von den darauf verlaufenden Orbits zu approximieren. Dieser Zugang erweist sich speziell bei 2-Tori praktikabel und soll dafür algorithmiert werden.

7.2.1 Eine Invarianzgleichung für invariante Tori

Nach der Transformation in (θ, u) -Koordinaten wollen wir eine Bestimmungsgleichung für den invarianten Torus

$$\mathcal{M} = \{(\theta, u) \mid u = u(\theta), \theta \in \mathbb{T}^p, u \in \mathbb{R}^q\} \subset \mathbb{T}^p \times \mathbb{R}^q \quad (7.34)$$

gewinnen. Diese Mannigfaltigkeit ist jedoch genau dann invariant unter dem Fluss des Systems

$$\frac{d\theta}{dt} = \Omega(\theta, u), \quad \frac{du}{dt} = R(\theta, u),$$

wenn zu jedem Punkt $(\theta, u) \in \mathcal{M}$ der Vektor $(\Omega(\theta, u), R(\theta, u))^T \in \mathbb{R}^p \times \mathbb{R}^q$ im Tangentialraum $T_{(\theta, u)}\mathcal{M}$ von \mathcal{M} liegt. Da dieser Unterraum durch die p Vektoren t_1, t_2, \dots, t_p mit

$$t_k = \left(\underbrace{0, 0, \dots, 1, \dots, 0, 0}_{k\text{-ter Einheitsvektor}}, \frac{\partial u_1}{\partial \theta_k}, \dots, \frac{\partial u_q}{\partial \theta_k} \right)^T \in \mathbb{R}^p \times \mathbb{R}^q$$

aufgespannt wird, stellen wir $(\Omega(\theta, u), R(\theta, u))^T$ als Linearkombination dieser Vektoren

$$\sum_{k=1}^p \alpha_k(\theta, u) t_k = \begin{pmatrix} \Omega(\theta, u) \\ R(\theta, u) \end{pmatrix} \quad \text{mit Koeffizientenfunktionen } \alpha_k(\theta, u)$$

dar. Einsetzen der t_k liefert unmittelbar $\alpha_k(\theta, u) = \Omega_k(\theta, u)$ für $k = 1(1)p$. Die zu bestimmende Funktion $u = (u_1, \dots, u_q) : \mathbb{T}^p \rightarrow \mathbb{R}^q$ löst damit das *System partieller Differenzialgleichungen (PDGL) 1. Ordnung* mit q Gleichungen

$$\sum_{k=1}^p \Omega_k(\theta, u) \frac{\partial u}{\partial \theta_k} = R(\theta, u), \quad \theta \in \mathbb{T}^p. \quad (7.35)$$

Da sowohl die Koeffizientenfunktionen $\Omega_k(\theta, u)$, $k = 1(1)p$, als auch die rechten Seiten $R_k(\theta, u)$, $k = 1(1)q$, von den Größen $\theta_1, \dots, \theta_p, u_1, \dots, u_q$ abhängen können, ist das System *quasilinear*. Zudem sind die skalaren Koeffizientenfunktionen $\Omega_k(\theta, u)$ in allen q Gleichungen identisch, so dass (7.35) ein *PDGL-System mit gleichem Hauptteil* darstellt.

Gemäß der Theorie partieller DGL (vgl. z. B. Courant und Hilbert (1968)) ist dieses PDGL-System dem *charakteristischen DGL-System*

$$\begin{aligned} \frac{d\theta_j}{dt} &= \Omega_j(\theta, u) \quad , \quad j = 1(1)p \\ \frac{du_k}{dt} &= R_k(\theta, u) \quad , \quad k = 1(1)q \end{aligned} \quad (7.36)$$

äquivalent. Dies bedeutet insbesondere, dass jede \mathcal{C}^1 -Torusmannigfaltigkeit \mathcal{M} unter dem Vektorfeld $(\Omega(\theta, u), R(\theta, u))^T$ genau dann invariant ist, wenn die \mathcal{C}^1 -Funktion $u(\theta)$ dem quasilinearen System (7.35) für $\theta \in \mathbb{T}^p$ genügt. Beschränken wir θ auf das p -dimensionale Intervall $[0, 2\pi]^p$, so sind periodische Randbedingungen für die Torusfunktion $u(\theta)$

$$u(\theta_1, \dots, \theta_{j-1}, \theta_j, \dots, \theta_p) = u(\theta_1, \dots, \theta_{j-1}, \theta_j + 2\pi, \dots, \theta_p), \quad j = 1(1)p \quad (7.37)$$

hinzuzufügen. Der gesuchte p -Torus \mathcal{M} kann dann als numerische Lösung des Randwertproblems (7.35), (7.37) bestimmt werden.

Beispiel 7.18

1. Das System zweier linear gekoppelter Van-der-Pol-Oszillatoren aus Beispiel 7.12 führt mit den dort definierten Funktionen Ω_1, Ω_2 und u_1, u_2 auf das PDGL-System

$$\begin{aligned} \Omega_1(\theta_1, \theta_2, u_1, u_2) \frac{\partial u_1}{\partial \theta_1} + \Omega_2(\theta_1, \theta_2, u_1, u_2) \frac{\partial u_1}{\partial \theta_2} &= R_1(\theta_1, \theta_2, u_1, u_2) \\ \Omega_1(\theta_1, \theta_2, u_1, u_2) \frac{\partial u_2}{\partial \theta_1} + \Omega_2(\theta_1, \theta_2, u_1, u_2) \frac{\partial u_2}{\partial \theta_2} &= R_2(\theta_1, \theta_2, u_1, u_2) \end{aligned} \quad (7.38)$$

mit den Torusbedingungen

$$\begin{aligned} u_1(\theta_1, \theta_2) &= u_1(\theta_1 + 2\pi, \theta_2), & u_2(\theta_1, \theta_2) &= u_2(\theta_1 + 2\pi, \theta_2), \\ u_1(\theta_1, \theta_2) &= u_1(\theta_1, \theta_2 + 2\pi), & u_2(\theta_1, \theta_2) &= u_2(\theta_1, \theta_2 + 2\pi). \end{aligned}$$

2. Das einfache partitionierte System (7.28) liefert ebenfalls eine derartige quasilineare PDGL, die im entkoppelten Fall ($\varepsilon = 0$) zur Darstellung

$$\begin{aligned} \omega_1 \frac{\partial u_1}{\partial \theta_1} + \omega_2 \frac{\partial u_1}{\partial \theta_2} &= u_1(1 - u_1^2) \\ \omega_1 \frac{\partial u_2}{\partial \theta_1} + \omega_2 \frac{\partial u_2}{\partial \theta_2} &= u_2(1 - u_2^2) \end{aligned} \quad (7.39)$$

führt. Die Lösung $u_1(\theta_1, \theta_2) = u_2(\theta_1, \theta_2) = 1$ auf dem Intervall $[0, 2\pi]^2$ genügt trivialerweise den Torusbedingungen. ■

Eine Approximation von \mathcal{M} entlang der Charakteristiken ist zwar möglich, indem wir das charakteristische System (7.36) als Anfangswertproblem integrieren. Dieser Zugang besitzt jedoch beträchtliche Nachteile in folgenden Fällen:

- Ist der Torus kein Attraktor, so erhalten wir numerisch keine auf \mathcal{M} verlaufenden Orbits. Auch bei „schwach“ attraktiven Tori ist der Zugang ineffizient.
- Sind die Basisfrequenzen rational abhängig, so ist der Torus nicht ergodisch und die berechneten quasi-periodischen bzw. periodischen Orbits füllen den Torus nicht dicht.

Durch die – numerisch zwar ungleich aufwändigere – Integration der PDGL mit den Torusbedingungen können wir uns von diesen Einschränkungen weitgehend befreien.

Beispiel 7.19

Für das System von W. F. Langford aus Beispiel 5.1 mit den Konstanten $\omega = 3.5$, $\varrho = 0.25$, $\beta = 0.7$ und dem Kontrollparameter ε liefert die Integration der gewöhnlichen DGL die in Abbildung 7.7 dargestellten Orbits. Während für $\varepsilon = 0$ der Orbit den ergodischen Torus dicht füllt, gewinnen wir bei $\varepsilon = 0.06$ keine wesentliche Information über die Mannigfaltigkeit. ■

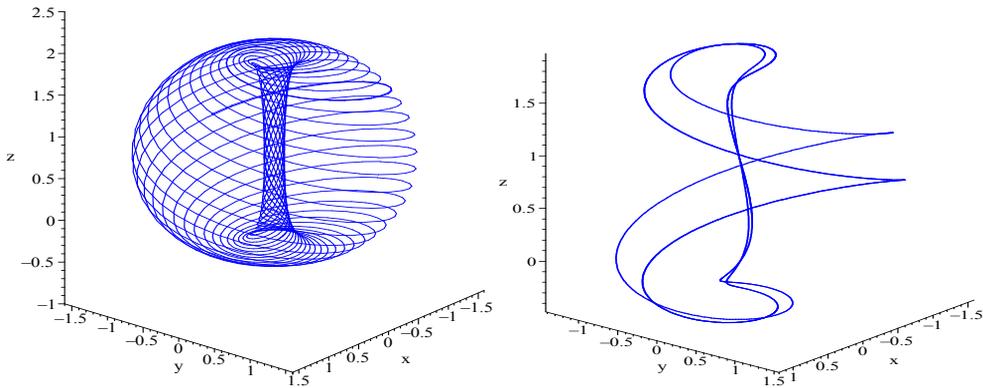


Abb. 7.7 Langford-System des Beispiels 6.26: Orbits für $\varepsilon = 0$ (links) und $\varepsilon = 0.06$ (rechts)

Die numerische Lösung der p -dimensionalen PDGL (7.35) stellt für $p \geq 2$ und große Systemdimension q ein ernstes Problem dar. In den vergangenen Jahren wurden unterschiedliche Ansätze entwickelt und im Falle von 2-Tori erfolgreich erprobt. So erfolgt die Lösung der Torusgleichungen (7.35), (7.37) auf \mathbb{T}^p durch

- *Finite-Differenzen-Methoden* auf einem diskretisierten Torus \mathbb{T}_h^p mit Schrittweiten $h = (h_1, h_2, \dots, h_p)$ im Falle $p = 2$ in Arbeiten von Dieci et al. (1991), Dieci und Bader (1992) sowie Moore (1996),
- eine *Charakteristikenmethode für hyperbolische DGL-Systeme* im Falle $p = 2$ bei Dieci und Lorenz (1995). Die Methode kann als diskretisierte Version der sogenannten Graph-Transformation interpretiert werden.
- Die Lösung der Torusgleichungen auf $\mathbb{T}^{p-1} \times [0, 2\pi]$ wird von Bernet und Vogt (1994) sowie Bernet (1995) mit einem *verallgemeinerten Schießverfahren für hyperbolische PDGL* bewältigt.
- Schließlich wurden auch *Fouriermethoden* auf \mathbb{T}^p bei Mingyou et al. (1997) sowie Chen und Leung (1998) mit Erfolg eingesetzt.

Um eine Lösung $u(\theta_1, \theta_2, \dots, \theta_p)$ der p -dimensionalen PDGL (7.35) numerisch mit vertretbarem Aufwand zu erhalten, ist eine Problemreduktion auf einen $(p - 1)$ -dimensionalen Torus wünschenswert, weil sich dadurch der Rechenaufwand erheblich verringern lässt. Insbesondere geht im praxisrelevanten Fall $p = 2$ die Aufgabe in eine gewöhnliche DGL über, für die umfangreiche Verfahren und Software verfügbar sind.

Wir treffen dazu folgende Voraussetzung, die in der Regel häufig erfüllt ist:

Voraussetzung 7.20

Es existiert ein $j \in \mathbb{N}$ mit $1 \leq j \leq p$, so dass die Funktion $\Omega_j(\theta, u) \neq 0$ für alle $(\theta, u) \in \mathbb{T}^p \times D$ mit $D \subset \mathbb{R}^q$ ist.

Durch einfaches Umnummerieren lässt sich stets $j = p$ erreichen. Wir dividieren sämtliche Gleichungen der PDGL (7.35) durch $\Omega_p(\theta, u)$, womit in dieser „normalisierten“ Gleichung nunmehr $\Omega_p(\theta, u) \equiv 1$ ist. Somit kann $\theta_p := t$ gesetzt und mit der Zeitvariablen t identifiziert werden. Die verbleibenden $p - 1$ Variablen bezeichnen wir wiederum abkürzend mit $\theta := (\theta_1, \dots, \theta_{p-1})$ und erhalten so die *normalisierte Darstellung der PDGL*

$$\frac{\partial u}{\partial t} + \sum_{j=1}^{p-1} \Omega_j(\theta, t, u) \frac{\partial u}{\partial \theta_j} = R(\theta, t, u) \quad (7.40)$$

mit den durch $\Omega_p(\theta, t, u)$ dividierten Funktionen für $j = 1, 2, \dots, p - 1$

$$\Omega_j(\theta, t, u) := \frac{\Omega_j(\theta, t, u)}{\Omega_p(\theta, t, u)} \quad \text{und} \quad R(\theta, t, u) := \frac{R(\theta, t, u)}{\Omega_p(\theta, t, u)}.$$

Die Torusbedingungen (7.37) lauten nunmehr entsprechend diesen Bezeichnungen

$$\begin{aligned} u(\theta_1, \dots, \theta_{j-1}, \theta_j, \dots, \theta_{p-1}, t) &= u(\theta_1, \dots, \theta_{j-1}, \theta_j, \dots, \theta_{p-1}, t + 2\pi), \\ u(\theta_1, \dots, \theta_{j-1}, \theta_j, \dots, \theta_{p-1}, t) &= u(\theta_1, \dots, \theta_{j-1}, \theta_j + 2\pi, \dots, \theta_{p-1}, t), \end{aligned} \quad (7.41)$$

für $j = 1(1)p - 1$.

Periodisch erregte Systeme

Eine wesentliche Problemklasse bilden die bereits in Kapitel 6 ausführlich betrachteten nichtautonomen dynamischen Systeme

$$\frac{dx}{dt} = f(x, t), \quad f: \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n \quad (7.42)$$

mit periodischer Erregung in t . Nach einer eventuellen Zeittransformation kann die Periode stets mit $T = 2\pi$ vorausgesetzt werden, d. h. $f(x, t + 2\pi) = f(x, t)$ für alle (x, t) . Identifizieren wir θ_p mit t , so ergibt sich nach Transformation ein partitioniertes System

$$\begin{aligned} \dot{\theta}_i &= \Omega_i(\theta, t, u), \quad i = 1(1)p - 1 \\ \dot{\theta}_p &= 1 \\ \dot{u} &= R(\theta, t, u) \end{aligned} \quad (7.43)$$

für $(\theta, t) \in \mathbb{T}^p$. Hieraus folgt unmittelbar die normalisierte Darstellung (7.40), die damit bei periodisch erregten DGL stets erreicht wird.

Beispiel 7.21

Das System 6.4 von Philippow und Büntig (1992) eines subharmonisch reagierenden elektrischen Netzwerkes

$$\ddot{x} - \varepsilon(1 - x^2 - \dot{x}^2)\dot{x} + x + b(4x^3 - 3x) = B \cdot \cos 3t \quad (7.44)$$

mit Parametern $\varepsilon, b, B > 0$ und Erregungsperiode $T_0 = \frac{2\pi}{3}$ wurde bereits in Kapitel 6 auf 3-fach subharmonische periodische Lösungen mit $T = 3T_0 = 2\pi$ hin untersucht. Nun wollen wir für verschiedene Werte der Erregungsamplitude B analysieren, ob (i) diese Lösungen auf invarianten 2-Tori liegen und (ii) quasi-periodische Orbits auftreten.

Die DGL 2. Ordnung kann leicht in das System 1. Ordnung

$$\begin{aligned} \dot{x} &= y \\ \dot{y} &= \varepsilon(1 - x^2 - y^2)y - x - b(4x^3 - 3x) + B \cos 3t \quad , \end{aligned}$$

überführt werden. Wir transformieren dieses periodisch erregte System mittels

$$x = u \cos \theta_1, \quad y = u \sin \theta_1$$

in Polarkoordinaten und setzen $t = \theta_2$. Das entstandene partitionierte System

$$\begin{aligned} \frac{d\theta_1}{dt} &= \varepsilon(1 - u^2) \sin \theta_1 \cos \theta_1 - 1 - b \cos^2 \theta_1 (4u^2 \cos^2 \theta_1 - 3) + \frac{1}{u} B \cos \theta_1 \cos 3\theta_2 \\ &= \Omega(\theta_1, \theta_2, u) \\ \frac{d\theta_2}{dt} &= 1 \\ \frac{du}{dt} &= \varepsilon(1 - u^2)u \sin^2 \theta_1 - b \sin \theta_1 (4u^3 \cos^3 \theta_1 - 3u \cos \theta_1) + B \sin \theta_1 \cos 3\theta_2 \\ &= R(\theta_1, \theta_2, u) \end{aligned} \quad (7.45)$$

besitzt offenbar die Gestalt (7.43) mit $p = 2$ und $q = 1$. Der im Parameterbereich $0 \leq \varepsilon \leq 5.0$, $0 < B \leq 1.0$ gesuchte invariante 2-Torus wird somit durch die PDGL

$$\frac{\partial u}{\partial \theta_2} + \Omega(\theta_1, \theta_2, u) \frac{\partial u}{\partial \theta_1} = R(\theta_1, \theta_2, u) \quad (7.46)$$

mit den in (7.45) erhaltenen Funktionen Ω, R und den Torusbedingungen

$$u(\theta_1, \theta_2) = u(\theta_1 + 2\pi, \theta_2), \quad u(\theta_1, \theta_2) = u(\theta_1, \theta_2 + 2\pi) \quad (7.47)$$

beschrieben. Die Gleichung (7.46) ist bereits in normalisierter Darstellung (7.40), so dass keine weiteren Umformungen notwendig sind. ■

7.2.2 Basisdiskretisierungen von 2-Tori

Wenden wir uns nun der numerischen Lösung des Problems (7.40) mit Periodizitätsbedingungen (7.41) zu. Wenn wir – wie bei partiellen DGL üblich – t als „Zeitvariable“ und θ als „Ortsvariablen“ auf dem Grundgebiet $D = [0, T]^p$ mit $T = 2\pi$ interpretieren, so stellt die PDGL ein quasilineares hyperbolisches System 1. Ordnung dar, dessen Lösungen allerdings die Randbedingungen für $j = 1(1)p - 1$

$$\begin{aligned} u(\theta_1, \dots, \theta_{j-1}, 0, \theta_{j+1}, \dots, \theta_{p-1}, t) &= u(\theta_1, \dots, \theta_{j-1}, T, \theta_{j+1}, \dots, \theta_{p-1}, t), \\ \text{sowie} \quad u(\theta_1, \dots, \theta_{p-1}, 0) &= u(\theta_1, \dots, \theta_{p-1}, T) \end{aligned} \quad (7.48)$$

auf dem Rand ∂D erfüllen müssen. Dazu fassen wir bei festem Wert $t \in \mathbb{T}^1$ die Funktion $u(\theta, t)$ als Element des Banach-Raumes C^r -glatter Torusfunktionen in $p - 1$ Variablen

$$\begin{aligned} \mathcal{B}_r &= \{v \in C^r(\mathbb{T}^{p-1}, \mathbb{R}^q) \mid u(\theta_1, \dots, \theta_{j-1}, \theta_j, \theta_{j+1}, \dots, \theta_{p-1}, t) = \\ &= u(\theta_1, \dots, \theta_{j-1}, \theta_j + T, \theta_{j+1}, \dots, \theta_{p-1}, t), \quad j = 1(1)p - 1\} \end{aligned}$$

mit $r \geq 2$ auf und bezeichnen für festes t die Elemente dieses Funktionenraumes mit $v(t) := u(\theta, t)$. Sodann führen wir den Differenzialoperator $\mathcal{F} : \mathcal{B}_r \times \mathbb{T}^1 \rightarrow \mathcal{B}_s$ mit $1 \leq s < r$ ein, der durch

$$\mathcal{F}(v, t) := - \sum_{j=1}^{p-1} \Omega_j(\theta, t, u(\theta, t)) \frac{\partial u(\theta, t)}{\partial \theta_j} + R(\theta, t, u(\theta, t)) \quad (7.49)$$

definiert wird. Beachten wir anschließend die Abhängigkeit der Abbildung $v(t) = u(\theta, t)$ von t , so ergibt sich bei variablem Wert t eine Banach-Raumwertige Funktion $v \in C^r([0, T], \mathcal{B}_r)$. Damit kann (7.40), (7.41) äquivalent durch das Zweipunkt-Randwertproblem (RWP)

$$\frac{dv}{dt} = \mathcal{F}(v, t), \quad v(0) = v(T) \quad \text{mit} \quad T = 2\pi \quad (7.50)$$

für v beschrieben werden. Die Periodizität der Funktionen Ω und R bezüglich t überträgt sich wegen $u \in C^r(\mathbb{T}^p)$ gemäß Voraussetzung auf den Operator \mathcal{F} mit

$$\mathcal{F}(v, t + T) = \mathcal{F}(v, t), \quad t \in \mathbb{T}^1, \quad (7.51)$$

womit (7.50) als periodisch erregtes System betrachtet werden kann.

Bemerkung 7.22

Da (7.50) wegen der Wahl der Banach-Räume \mathcal{B}_r lediglich eine Umformulierung der ursprünglichen Aufgabe darstellt, entspricht jede periodische Lösung v^* von (7.50) genau einer Toruslösung u^* von (7.40) und umgekehrt. Voraussetzung 7.15 sichert somit zusammen mit Voraussetzung 7.20 die Existenz einer periodischen Lösung v^* des Zweipunkt-Randwertproblems. Zudem folgt auch die stetige Differenzierbarkeit von \mathcal{F} bezüglich v und t im Sinne von Fréchet. ■

Schießverfahren

Als Lösungsverfahren für das RWP (7.50) bietet sich die Rückführung dieser Aufgabe mittels des Schießverfahrens 6.6 aus Kapitel 6 an, das nun nicht im endlichdimensionalen Raum \mathbb{R}^n , sondern im unendlichdimensionalen Banach-Raum \mathcal{B} ausgeführt wird. Wir betrachten dazu das entsprechende AWP

$$\frac{dv}{dt} = \mathcal{F}(v(t), t) \quad , \quad v(0) = s \in \mathcal{B}_r \quad (7.52)$$

für $t \in [0, T]$. Gesucht ist ein Element $s^* \in \mathcal{B}_r$, mit dem die Periodizitätsbedingung $v(0) = v(T) = s^*$ erfüllt ist. Wir wollen wie in Kapitel 6 davon ausgehen, dass für Anfangselemente s in einer hinreichend kleinen Umgebung $\mathcal{U}(s^*)$ um s^* dieses AWP eine eindeutige Lösung auf ganz $[0, T]$ besitzt. Bezeichnen wir mit $v(t; s)$ diese Lösung für festes $s \in \mathcal{B}_r$, so lässt sich auf $\mathcal{U}(s^*)$ die Abbildung

$$g(s) := v(T; s) - s, \quad g : \mathcal{U}(s^*) \subset \mathcal{B}_r \rightarrow \mathcal{B}_r \quad (7.53)$$

definieren. Offenbar ist $v^*(0) = s^*$ genau dann eine gesuchte Anfangsbedingung von (7.52), wenn s^* die Gleichung

$$g(s) := v(T; s) - s = 0 \quad (7.54)$$

löst. Zur Bestimmung asymptotisch und speziell exponentiell stabiler Tori kann dafür die Fixpunktiteration (Picard-Iteration) eingesetzt werden. In diesen Fällen ergeben sich die Fixpunktiterierten in der Form

$$s^{(k+1)} = v(T; s^{(k)}), \quad k = 0, 1, 2, \dots \quad (7.55)$$

Nach Diskretisierung des Operators \mathcal{F} entstehen – insbesondere bei 3-Tori – Probleme großer Dimension, wofür Newton-Verfahren sehr aufwändig werden. Deshalb wird von Bernet und Vogt (1994) für die Bestimmung exponentiell stabiler Tori die wesentlich aufwandsärmere Fixpunktiteration empfohlen.

Ist der Torus nicht asymptotisch stabil, so divergiert im Allgemeinen die Fixpunktiteration, d. h. in diesen Fällen muss man Newton- oder Newton-ähnliche Verfahren einsetzen, um die Konvergenz zu erzwingen. Da bei der hier betrachteten Problemstellung die Fréchet-Ableitung $g'(s)$ nicht explizit verfügbar ist, ist man auf deren Approximation angewiesen. Dabei verwendet man eine als Steigung oder dividierte Differenz bezeichnete Approximation $\Delta g : \mathcal{U}(s^*) \times \mathcal{U}(s^*) \subseteq \mathcal{B}_r \times \mathcal{B}_r \rightarrow L(\mathcal{B}_r)$ von g' , wobei $L(\mathcal{B}_r)$ den Raum der linearen Abbildungen von \mathcal{B}_r in \mathcal{B}_r bezeichnet und $\mathcal{U}(s^*)$ eine offene Umgebung von s^* in \mathcal{B}_r ist. Details findet man u. a. bei Bernet (1995). Das *einfache Schießverfahren* im Banach-Raum \mathcal{B}_r erhält damit die abstrakte algorithmische Form 7.23. Dieser Algorithmus bedarf allerdings noch einer Diskretisierung, wofür Finite-Differenzen-Methoden (FDM) und Finite-Elemente-Methoden (FEM) geeignet sind. Wir werden uns nachfolgend auf FDM beschränken; zu FEM verweisen wir auf Hoffmann et al. (2006).

ALGORITHMUS 7.23 (Schießverfahren für Tori)Function $[s^*] = \text{shooting_tori}(\mathcal{F}, T, \text{tol}, \text{kmax})$

1. Wähle $s^{(0)} \in \mathcal{U}(s^*) \subset \mathcal{B}_r$.
2. Für $k = 0, 1, \dots, \text{kmax}$ iteriere:
 - 2.1. Löse AWP (7.52) mit $v(0) = s^{(k)}$ und bestimme $v(t; s^{(k)})$ auf $[0, T]$.
 - 2.2. Definiere $g(s^{(k)}) := v(T; s^{(k)}) - s^{(k)}$ in \mathcal{B}_r .
 - 2.3. Bestimme eine Approximation $A^{(k)} := \Delta g(s^{(k)}, \Delta^{(k)} s)$ der Fréchet-Ableitung $g'(s^{(k)})$.
 - 2.4. Löse in \mathcal{B}_r das lineare Gleichungssystem $A^{(k)} d^{(k)} = -g(s^{(k)})$ und berechne $s^{(k+1)} := s^{(k)} + d^{(k)}$.
 - 2.5. Ist $\|d^{(k)}\|_r < \text{tol}$, so setze $s^* := s^{(k+1)}$ und gehe zu Schritt 3.
3. Return s^*

Basisdiskretisierungen 1. Ordnung

In der nachfolgenden Darstellung beschränken wir uns auf den praxisrelevanten Fall von 2-Tori mit der Variablenbezeichnung $(\theta, t) \in \mathbb{T}^2$. System (7.40) erhält dann die Form

$$\frac{\partial u}{\partial t} + \Omega(\theta, t, u) \frac{\partial u}{\partial \theta} = R(\theta, t, u), \quad (\theta, t) \in \mathbb{T}^2 \quad (7.56)$$

mit $\Omega : \mathbb{T}^2 \times \mathbb{R}^q \rightarrow \mathbb{R}$, $R : \mathbb{T}^2 \times \mathbb{R}^q \rightarrow \mathbb{R}^q$ und den Torusbedingungen

$$u(\theta, t) = u(\theta + 2\pi, t), \quad u(\theta, t) = u(\theta, t + 2\pi).$$

Das zu lösende zugehörige AWP (7.52) lautet entsprechend

$$\frac{\partial u}{\partial t} = -\Omega(\theta, t, u) \frac{\partial u}{\partial \theta} + R(\theta, t, u), \quad u(\theta, 0) = s(\theta) \in \mathcal{C}^r(\mathbb{T}^1). \quad (7.57)$$

Wir diskretisieren nun den Standardtorus \mathbb{T}^2 mit θ -Schrittweite (der „Ortsschrittweite“) $h > 0$ und t -Schrittweite (der „Zeitschrittweite“) $\tau > 0$, wobei wir $\tau/h = \text{const}$ voraussetzen und deshalb h als Diskretisierungsparameter betrachten können. Der diskrete 2-Torus ist die Menge

$$\mathbb{T}_h^2 = \{(\theta_j, t_n) \mid \theta_j = jh, t_n = n\tau, j = 0(1)J - 1, n = 0(1)N - 1\}. \quad (7.58)$$

Ein Einschrittverfahren bezüglich der Zeitvariablen t berechnet Näherungswerte auf der $(n+1)$ -ten Zeitschicht t_{n+1} aus den Lösungsnäherungen der n -ten Schicht t_n . Auf den Gitterpunkten (θ_j, t_n) lassen sich zahlreiche explizite und linear implizite FDM 1. Ordnung zum AWP (7.52) in der allgemeinen Darstellung

$$\frac{1}{\tau} \left[\sum_{\mu=-1}^1 S_\mu^*(\theta_j, t_n, u_j^n) u_{j+\mu}^{n+1} - \sum_{\mu=-1}^1 S_\mu(\theta_j, t_n, u_j^n) u_{j+\mu}^n \right] - R(\theta_j, t_n, u_j^n) = 0 \quad (7.59)$$

mit den Gitterfunktionen $u_j^n \sim u(\theta_j, t_n)$ angeben. Offenbar stellt dieses Verfahren ein 6-Punkt-Differenzenschema dar, bei dem die drei Lösungswerte $u_{j-1}^{n+1}, u_j^{n+1}, u_{j+1}^{n+1}$ auf der $(n + 1)$ -ten Zeitschicht t_{n+1} aus den drei Werten $u_{j-1}^n, u_j^n, u_{j+1}^n$ der n -ten Schicht ermittelt werden (vgl. Abb.7.8). $S_\mu^* \in \mathbb{R}^{q \times q}$ und $S_\mu \in \mathbb{R}^{q \times q}$ mit $\mu = -1, 0, 1$ sind sechs

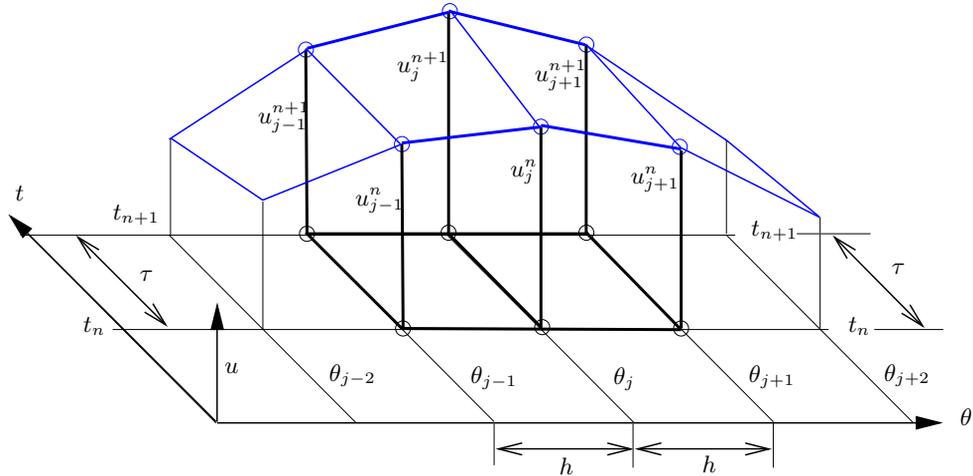


Abb. 7.8 6-Punkt-Differenzenschema: Lösungswerte auf den Zeitschichten t_n und t_{n+1}

Diagonalmatrizen mit Argumenten auf der Zeitschicht t_n . Eine wesentliche Eigenschaft derartiger Verfahren wird durch folgende Definition beschrieben:

Definition 7.24 (FDM von positivem Typ)

Ein durch (7.59) gegebenes Finite-Differenzen-Verfahren (FDM) heißt *von positivem Typ*, wenn die drei Matrizen $S_\mu(\theta, t, u)$, $\mu = -1, 0, 1$, für alle $(\theta, t, u) \in \mathbb{T}^2 \times \mathbb{R}^q$ mit $\|u\| \leq M$ und konstantem $M > 0$ ausschließlich nichtnegative Elemente besitzen.

Ein allgemeiner Konsistenz- und Konvergenzsatz sowohl für explizite als auch linear implizite Verfahren 1. Ordnung der Form (7.59) wurde von Bernet (1995) bewiesen. Mit allen Voraussetzungen lautet er:

Satz 7.25 (Konsistenz und Konvergenz)

Ein Finite-Differenzen-Verfahren (7.59) zum AWP (7.57) genüge folgenden Voraussetzungen:

- i. $\lambda := \tau/h = \text{const}$
- ii. $\Omega(\theta, t, u) \in \mathcal{C}^2(\mathbb{T}^2 \times \mathbb{R}^q, \mathbb{R})$, $R(\theta, t, u) \in \mathcal{C}^2(\mathbb{T}^2 \times \mathbb{R}^q, \mathbb{R}^q)$
- iii. (7.57) besitzt eine lokal eindeutige Lösung $u \in \mathcal{C}^2(\mathbb{T}^2, \mathbb{R}^q)$
- iv. $\sum_{\mu=-1}^1 S_\mu(\theta, t, u) \equiv I$, $\sum_{\mu=-1}^1 S_\mu^*(\theta, t, u) \equiv I$ mit Einheitsmatrix I ,

$$\text{v. } \sum_{\mu=-1}^1 \mu(S_\mu^*(\theta, t, u) - S_\mu(\theta, t, u)) \equiv \lambda C \text{ mit } C = \text{diag}(c_i) \\ \text{und } c_i = \Omega(\theta, t, u), \quad i = 1(1)q.$$

Dann ist das Verfahren *konsistent* in h und τ mit der Ordnung 1. Gilt weiterhin:

- vi.** (7.59) ist von positivem Typ, die Diagonalmatrizen S_μ^* , $\mu = -1, 0, 1$, sind Lipschitz-stetig in u mit der zusätzlichen Bedingung, dass zwei der drei Matrizen S_μ^* nur Diagonalelemente $(s_\mu^*)_k \leq 0$, $k = 1(1)q$, besitzen und
- vii.** bezüglich der Anfangswerte sei $\|u_j^0 - s(\theta_j)\|_\infty \leq K_0 h$ für alle $j \geq 0$, $h \leq h_0$, $K_0 > 0$ erfüllt,

so *konvergiert* (7.59) auf \mathbb{T}_h^2 mit der Ordnung 1, d. h.

$$\|e_j^n\|_\infty = \|z_j^n - z(t_n; \theta_j)\|_\infty \leq Kh, \quad K > 0 \quad \text{für alle } j, n.$$

In dieser allgemeinen Verfahrensdarstellung (7.59) sind alle nachfolgend behandelten expliziten und linear-impliziten 6-Punkt-Schemata enthalten.

Explizite Verfahren vom Upwind-Typ

Bei diesen Verfahren wird die Matrix $C(\theta_j, t_n, u_j^n)$ in die Summe einer Matrix $C^+(\theta_j, t_n, u_j^n)$ mit nur positiven Elementen und einer Matrix $C^-(\theta_j, t_n, u_j^n)$ mit nur negativen Elementen zerlegt:

$$C(\theta_j, t_n, u_j^n) = C^+(\theta_j, t_n, u_j^n) + C^-(\theta_j, t_n, u_j^n). \quad (7.60)$$

(a) Das *explizite Courant-Isaacson-Rees Verfahren (CIR-Verfahren)* ergibt sich mit

$$C^+ = \text{diag}\left(\frac{1}{2}(c_i + |c_i|)\right) \geq 0, \quad C^- = \text{diag}\left(\frac{1}{2}(c_i - |c_i|)\right) \leq 0. \quad (7.61)$$

(b) Das *explizite glatte Upwind-Verfahren* entspricht der allgemeinen Darstellung

$$C^+ = \text{diag}\left(\frac{1}{2}(c_i + \Phi(c_i))\right) \geq 0, \quad C^- = \text{diag}\left(\frac{1}{2}(c_i - \Phi(c_i))\right) \leq 0, \quad (7.62)$$

mit der Funktion $\Phi(d) := \sqrt{\delta^2 + d^2}$ und konstantem $\delta \neq 0$. Offenbar stellt das CIR-Verfahren den Grenzfall für $\delta \rightarrow 0$ mit $\Phi(d) = |d|$ dar. Die Diagonalmatrizen S_μ und S_μ^* , $\mu = -1, 0, 1$, besitzen für alle expliziten Verfahren vom Upwind-Typ die Form

$$\begin{aligned} S_{-1} &= \lambda C^+, & S_{-1}^* &= 0 \\ S_0 &= I - \lambda C^+ + \lambda C^-, & S_0^* &= I \\ S_1 &= -\lambda C^-, & S_1^* &= 0. \end{aligned} \quad (7.63)$$

Nach Einsetzen dieser Matrizen in (7.59) lässt sich das explizite Upwind-Verfahren in der klassischen FDM-Notation mit $\Phi(d) = \sqrt{\delta^2 + d^2}$, $\delta \neq 0$, angeben:

$$\begin{aligned}
 u_j^{n+1} = u_j^n &+ \frac{\lambda}{2} \Omega(\theta_j, t_n, u_j^n) (u_{j-1}^n - u_{j+1}^n) + \\
 &+ \frac{\lambda}{2} \Phi(\Omega(\theta_j, t_n, u_j^n)) (u_{j-1}^n - 2u_j^n + u_{j+1}^n) + \\
 &+ \tau R(\theta_j, t_n, u_j^n), \quad j = 0(1)J - 1.
 \end{aligned}
 \tag{7.64}$$

Wie man leicht erkennt, wird neben den Differenzenapproximationen der 1. Ableitungen nach t und θ ein künstlicher Diffusionsterm 2. Ordnung eingeführt und durch einen zentralen Differenzenquotienten approximiert. Das Differenzenschema zur Berechnung von u_j^{n+1} ist in Abbildung 7.9 dargestellt.

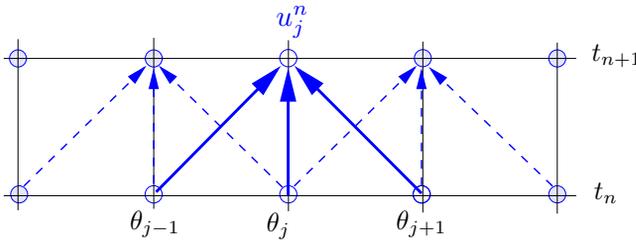


Abb. 7.9 Differenzenschema des expliziten Upwind- und Friedrichs-Verfahrens

(c) Beim *expliziten Friedrichs-Verfahren* wird die Matrix $C(\theta_j, t_n, u_j^n)$ nicht zerlegt, und die Diagonalmatrizen S_μ und S_μ^* , $\mu = -1, 0, 1$, besitzen nun die Form

$$\begin{aligned}
 S_{-1} &= \frac{1}{2}(I + \lambda C), & S_{-1}^* &= 0 \\
 S_0 &= 0, & S_0^* &= I \\
 S_1 &= \frac{1}{2}(I - \lambda C), & S_1^* &= 0.
 \end{aligned}
 \tag{7.65}$$

In der üblichen Notation für FDM erkennt man, dass bei diesem Verfahren an Stelle des Funktionswertes u_j^n der gemittelte Wert $\frac{1}{2} (u_{j+1}^n + u_{j-1}^n)$ benutzt wird:

$$u_j^{n+1} = \frac{1}{2} (u_{j+1}^n + u_{j-1}^n) + \frac{\lambda}{2} \Omega(\theta_j, t_n, u_j^n) (u_{j-1}^n - u_{j+1}^n) + \tau R(\theta_j, t_n, u_j^n). \tag{7.66}$$

Wenn die Lösungswerte u_j^n auf der n -ten Zeitschicht vorliegen, können mit beiden Verfahren die neuen Werte für $j = 0, 1, 2, \dots, J - 1$ leicht bestimmt werden. Eine spezielle Randapproximation in der Umgebung von $\theta_0 = 0$ und $\theta_{J-1} = 2\pi - h$ ist nicht erforderlich, wenn alle Rechnungen bezüglich des Index j modulo J ausgeführt werden.

Für die beiden Verfahren (7.64) und (7.66) sind gemäß ihrer Konstruktion die Voraussetzungen (iv) und (v) des Konvergenzsatzes 7.25 erfüllt, womit ihre Konsistenz gesichert ist. Um auch die Voraussetzung (vi) zu garantieren, bedarf es einer *CFL-Bedingung*.¹

¹Die CFL-Bedingung ist nach den Mathematikern R. Courant, K. Friedrichs und H. Lewy benannt, die sie 1928 einführten.

Satz 7.26

Auf der Menge $G = \{ (\theta, t, u) \mid (\theta, t) \in \mathbb{T}^2, \|u\| \leq M, M \in \mathbb{R}, \text{const} \}$ sei die Schrittweitenbedingung (*Courant-Friedrichs-Lewy-Zahl, CFL-Bedingung*)

$$\lambda \leq \frac{1}{D} \quad \text{mit} \quad D := \max_G \max_{i=1(1)q} |c_i(\theta, t, u)| \quad (7.67)$$

erfüllt. Dann sind das explizite CIR-Verfahren und das explizite Friedrichs-Verfahren von positivem Typ. Ist jedoch die schärfere CFL-Bedingung

$$\lambda \leq \frac{1}{\sqrt{\delta^2 + D^2}}, \quad \delta \neq 0, \quad (7.68)$$

erfüllt, so ist auch das explizite glatte Upwind-Verfahren von positivem Typ.

Beweis: (a) CIR-Verfahren: Auf Grund von (7.61) und (7.63) gilt sofort $S_{-1} \geq 0$ und $S_1 \geq 0$. Die Diagonalelemente s_0^i , $i = 1(1)q$ von S_0 lassen sich in der Form

$$s_0^i = 1 - \frac{\lambda}{2}(c_i + |c_i|) + \frac{\lambda}{2}(c_i - |c_i|) = 1 - \lambda |c_i| \geq 1 - \lambda D$$

abschätzen. Daraus folgt wegen $\lambda \leq 1/D$ unmittelbar $S_0 \geq 0$.

(b) Explizites Friedrichs-Verfahren: In diesem Fall gilt $S_0 = 0$. Für die Diagonalelemente s_1^i , $i = 1(1)q$, von S_1 erhalten wir die Abschätzung

$$s_1^i = \frac{1}{2}(1 - \lambda c_i) \geq \frac{1}{2}(1 - \lambda |c_i|) \geq \frac{1}{2}(1 - \lambda D),$$

woraus wegen $\lambda \leq 1/D$ sofort $S_1 \geq 0$ folgt. Für die Diagonalelemente s_{-1}^i , $i = 1(1)q$, von S_{-1} gilt

$$s_{-1}^i = \frac{1}{2}(1 + \lambda c_i) \geq \frac{1}{2}(1 - \lambda |c_i|),$$

womit analog zu S_1 folgt, dass auch $S_{-1} \geq 0$ erfüllt ist.

(c) Explizites glattes Upwind-Verfahren: Aus $\Phi(d) \geq |d|$ folgt sofort $S_{-1} \geq 0$ und $S_1 \geq 0$. Analog zu (a) zeigen wir für die Diagonalelemente s_0^i , $i = 1(1)q$, von S_0

$$s_0^i = 1 - \lambda \Phi(c_i) = 1 - \lambda \sqrt{\delta^2 + c_i^2} \geq 1 - \lambda \sqrt{\delta^2 + D^2},$$

woraus wegen $\lambda \leq 1/\sqrt{\delta^2 + D^2}$ unmittelbar $S_0 \geq 0$ folgt. \square

Bei allen drei expliziten Verfahren sind die Diagonalmatrizen S_μ^* , $\mu = -1, 0, 1$, unabhängig von u und damit stets Lipschitz-stetig bezüglich u . Ist also zusätzlich die CFL-Bedingung (7.67) bzw. (7.68) erfüllt, so sind die Verfahren für geeignete Startwerte gemäß Voraussetzung (vii) des Satzes 7.25 konvergent.

Linear implizite Verfahren vom Upwind-Typ

Wenn die Funktion $\Omega(\theta, t, u)$ betragsgroß wird, so reduziert die CFL-Bedingung des Satzes 7.26 die t -Schrittweite τ gegenüber der θ -Schrittweite h mitunter um Größenordnungen. Dann empfehlen sich implizite Verfahren, die ohne diese einschränkende Bedingung auskommen. Allerdings entsteht nun auf jeder Zeitschicht ein nichtlineares Gleichungssystem, dessen Lösung mit wachsender Dimension sehr aufwändig wird. Um dies zu vermeiden, wenden wir das *Prinzip der eingefrorenen Koeffizienten* an, indem die Werte aller Koeffizientenfunktionen und rechten Seiten auf der zuletzt berechneten Zeitschicht t_n ermittelt und dort „eingefroren“ werden. Dadurch entstehen linear implizite Verfahren, bei denen pro Zeitschicht lediglich ein lineares Gleichungssystem zu lösen ist.

Wie bei den expliziten Verfahren vom Upwind-Typ zerlegen wir dazu die Matrix

$$C(\theta_j, t_n, u_j^n) = C^+(\theta_j, t_n, u_j^n) + C^-(\theta_j, t_n, u_j^n)$$

mit den beiden Matrizen $C^+(\theta_j, t_n, u_j^n)$ und $C^-(\theta_j, t_n, u_j^n)$. Diese genügen den Beziehungen (7.61) für das CIR-Verfahren bzw. (7.62) für das glatte Upwind-Verfahren mit der Funktion $\Phi(d) := \sqrt{\delta^2 + d^2}$, $\delta \neq 0$, const.

Die Diagonalmatrizen S_μ und S_μ^* , $\mu = -1, 0, 1$, erhalten dann für alle linear impliziten Verfahren vom Upwind-Typ die Form

$$\begin{aligned} S_{-1} &= 0, & S_{-1}^* &= -\lambda C^+ \\ S_0 &= I, & S_0^* &= I - \lambda C^- + \lambda C^+ \\ S_1 &= 0, & S_1^* &= \lambda C^-. \end{aligned} \quad (7.69)$$

Setzen wir diese Matrizen in die allgemeine Darstellung (7.59) ein, so gewinnen wir die übliche FDM-Notation für das *linear implizite Verfahren vom Upwind-Typ*:

$$\begin{aligned} u_j^{n+1} &= u_j^n + \frac{\lambda}{2} \Omega(\theta_j, t_n, u_j^n) (u_{j-1}^{n+1} - u_{j+1}^{n+1}) + \\ &+ \frac{\lambda}{2} \Phi(\Omega(\theta_j, t_n, u_j^n)) (u_{j-1}^{n+1} - 2u_j^{n+1} + u_{j+1}^{n+1}) + \\ &+ \tau R(\theta_j, t_n, u_j^n), \quad j = 0(1)J - 1. \end{aligned} \quad (7.70)$$

Mit $\Phi(d) = |d|$ ergibt sich auch hier das *linear implizite CIR-Verfahren* als Grenzfall. Wie man aus (7.70) und dem Differenzenschema zur Berechnung der Werte u_{j-1}^{n+1} , u_j^{n+1} und u_{j+1}^{n+1} auf der $(n+1)$ -ten Zeitschicht in Abbildung 7.10 erkennt, ist nun ein lineares Gleichungssystem mit Block-tridiagonaler Koeffizientenmatrix zu lösen. Mit je einer Vorwärts- und einer Rückwärtselimination pro t -Schritt verdoppelt sich lediglich der numerische Aufwand gegenüber den expliziten Verfahren.

An den drei Matrizen S_μ , $\mu = -1, 0, 1$, in (7.69) ist sofort erkennbar, dass das Verfahren für alle $\lambda > 0$ von positivem Typ ist. Zudem sind die Diagonalmatrizen S_μ^* , $\mu = -1, 0, 1$, Lipschitz-stetig bezüglich u , und wegen der Darstellung (7.62) von C^+ und C^- gilt

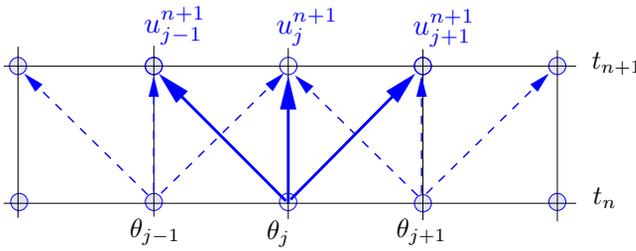


Abb. 7.10 Differenzenschema des impliziten Upwind- und CIR-Verfahrens

weiterhin $S_{-1}^* \leq 0$ und $S_1^* \leq 0$. Somit ist das linear implizite Verfahren vom Upwind-Typ unter den Voraussetzungen (i)–(iii) des Satzes 7.25 konsistent und mit geeigneten Startwerten gemäß Voraussetzung (vii) auch konvergent.

7.2.3 Ein parametrisch erregtes elektrisches Netzwerk

Als Anwendungsbeispiel wollen wir das in Abbildung 7.11 dargestellte parametrisch erregte Netzwerk analysieren, das von Philippow und Büntig (1992) entwickelt wurde. Dieser Schwingkreis beinhaltet eine Kapazität C , einen nichtlinearen Widerstand R und eine lineare zeitveränderliche Induktivität L und wird auch als 2:1-Frequenzteiler genutzt.

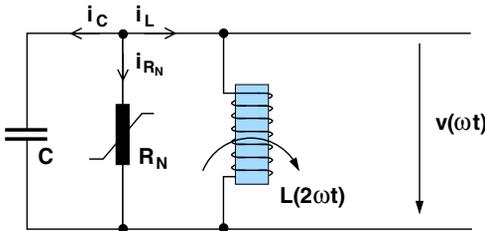


Abb. 7.11 Ein parametrisch erregtes Netzwerk von E. Philippow mit nichtlinearem Widerstand und zeitveränderlicher Induktivität. Die Kennlinie des Widerstandes ist annähernd kubisch und besitzt Bereiche mit negativem Anstieg

Das Modell kann mittels der Knotengleichung $i_C + i_{R_N} + i_L = 0$ hergeleitet werden, worin die drei Ströme durch die Ausdrücke

$$i_C = C \frac{dv}{dt}, \quad i_{R_N} = b_1 v^3 - b_2 v, \quad i_L = \frac{\psi}{L_0(1 + \frac{b}{2} \sin 2\omega t)} \approx \frac{\psi}{L_0} \left(1 - \frac{b}{2} \sin 2\omega t \right)$$

gegeben sind. v bezeichnet den Spannungsabfall über den Komponenten. Mit der Beziehung $\frac{d\psi}{dt} = v$ erhalten wir eine DGL 2. Ordnung für $\psi(t)$, die nach einer Normierung mittels $x := \psi/\psi_0$ und $t := \omega t$ in die Form

$$\ddot{x} + \alpha \dot{x}^3 - \beta \dot{x} + (1 + B \sin 2t)x = 0 \tag{7.71}$$

übergeht. Die Parameter dieser DGL vom Mathieuschen Typ $\alpha = \varepsilon - B$ und $\beta = \frac{\varepsilon}{2} - B$ mit $B, \varepsilon \in \mathbb{R}$ werden so gewählt, dass die Systemantwort $x(t)$ ein 2π -periodisches Signal darstellt, d. h. die Frequenz des Eingangssignals $B \sin 2t$ wird halbiert. Die *periodischen Lösungen* dieses Systems wurden von Hoffmann (1989) und Vogt et al. (2005) für Parameterwerte $B = 0.1$ und $\varepsilon \in [0, 12]$ einer Bifurkationsanalyse unterzogen.

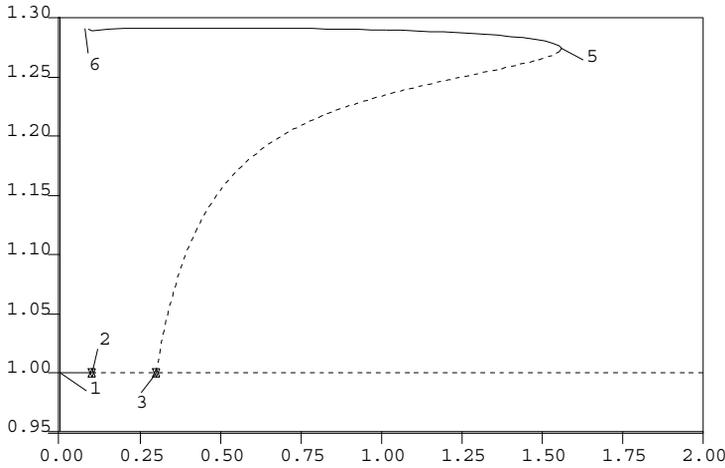


Abb. 7.12 Lösungsdiagramm der periodischen Lösung von DGL (7.71): Für $\varepsilon \approx 0.1$ (Label 2) und $\varepsilon \approx 0.3$ (Label 3) treten Periodenverdopplungen auf und es zweigt jeweils eine 2π -periodische Lösung von der trivialen Gleichgewichtslösung ab. Diese Lösungen verschmelzen bei $\varepsilon \approx 1.56081$ (Label 5)

Für $0.3 < \varepsilon < 1.56081$ besitzt System (7.71) vier periodische Lösungen, zwei stabile und zwei vom Satteltyp, die auf einem 1:2-resonanten invarianten Torus liegen. Für $\varepsilon > 1.56081$ existieren quasi-periodische bzw. resonante Tori, abhängig vom Verhältnis der Basisfrequenzen an ε . Sind die Basisfrequenzen rational abhängig, so sind in diesen Gebieten alle Verfahren, die Quasi-Periodizität voraussetzen, nicht anwendbar.

Das Bifurkationsdiagramm in der $\varepsilon - B$ -Ebene wird in Abbildung 7.13 dargestellt. Die grauen Bereiche heißen *Arnold-Zungen*. Für Parameterwerte innerhalb solcher Arnold-Zungen tritt ein sogenanntes *phase-locking* auf, d. h. auf dem Torus liegt ein stabiler und ein instabiler periodischer Orbit. Die Diagramme in Abbildung 7.12 und 7.13 wurden mit dem Programm AUTO2000 von Doedel et al. (2002) erzeugt.

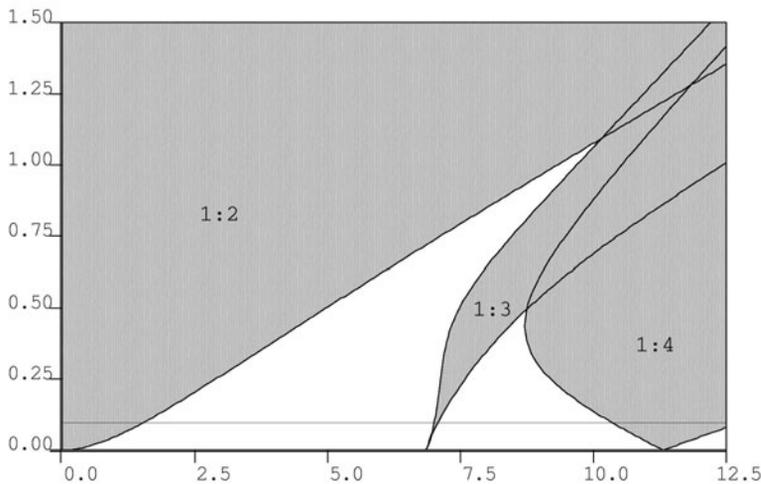


Abb. 7.13 Das Bifurkationsdiagramm in der $\varepsilon - B$ -Ebene. Die grauen Bereiche zeigen die Arnold-Zungen der 1:2-, 1:3- und 1:4-Resonanzen

Um die DGL in die Form des Modellproblems (7.56) zu bringen, transformieren wir sie in das System erster Ordnung

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = -x_1 + \beta x_2 - \alpha x_2^3 - B \sin 2t \cdot x_1. \quad (7.72)$$

Analog zu Beispiel 7.21 führen wir Polarkoordinaten (θ_1, u) mit

$$x_1 = u \cos \theta_1, \quad x_2 = u \sin \theta_1$$

ein und setzen $t \equiv \theta_2$. Mit der Abkürzung $cs = \cos \theta_1 \sin \theta_1$ entsteht das partitionierte System

$$\begin{aligned} \frac{d\theta_1}{dt} &= \beta cs - \sin^2 \theta_1 - \alpha u^2 \sin^3 \theta_1 \cos \theta_1 - (1 + B \sin 2\theta_2) \cos^2 \theta_1 = \Omega(\theta_1, \theta_2, u) \\ \frac{d\theta_2}{dt} &= 1 \\ \frac{du}{dt} &= u(cs + \beta \sin^2 \theta_1) - \alpha u^3 \sin^4 \theta_1 - u(1 + B \sin 2\theta_2)cs = R(\theta_1, \theta_2, u). \end{aligned} \quad (7.73)$$

Der gesuchte invariante Torus \mathcal{M} wird dann durch die partielle DGL

$$\frac{\partial u}{\partial \theta_2} + \Omega(\theta_1, \theta_2, u) \frac{\partial u}{\partial \theta_1} = R(\theta_1, \theta_2, u), \quad (\theta_1, \theta_2) \in \mathbb{T}^2 \quad (7.74)$$

mit den in (7.73) definierten Funktionen Ω, R beschrieben. Wir diskretisieren diese PDGL mit dem linear impliziten Verfahren vom Upwind-Typ (7.70) auf einem (θ, t) -Gitter der Dimension 40×80 und wenden das Schießverfahren für Tori (vgl. Algorithmus 7.23) an. In Abbildung 7.14 werden die Torusquerschnitte, zurücktransformiert in kartesische

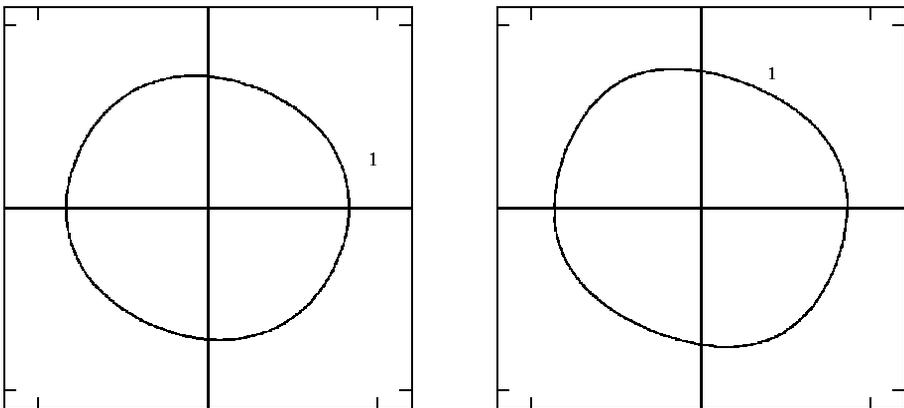


Abb. 7.14 Torusquerschnitte von System (7.74) mit (*) stabilen periodischen Lösungen sowie (□) instabilen periodischen Lösungen

(x_1, x_2) -Koordinaten, für den Parameterwert $\varepsilon = 1.0$ (links) und den kritischen Wert $\varepsilon = 1.56081$ (rechts) dargestellt. Die zuvor berechneten Stroboskop-Abbildungen P_1, P_2 der periodischen Lösungen vom Satteltyp und der stabilen periodischen Lösungen P_3, P_4 wurden ebenfalls eingezeichnet. Am kritischen Parameterwert $\varepsilon \approx 1.56081$ verschmelzen

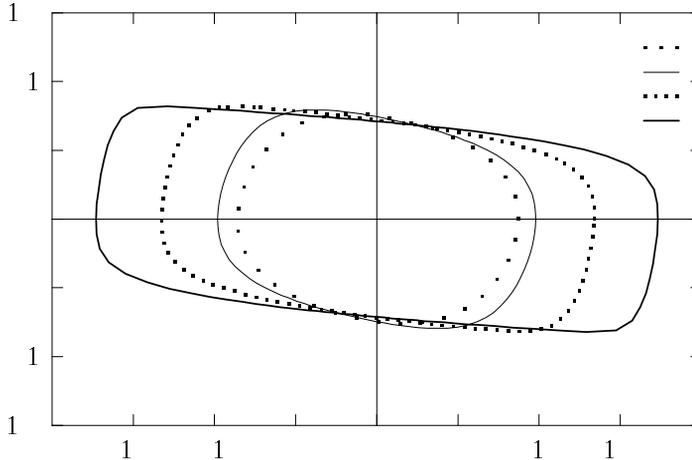


Abb. 7.15 Torusquerschnitte für Parameterwerte $\varepsilon > 1.56081$

je eine stabile und eine Lösung vom Satteltyp durch eine Sattel-Knoten-Bifurkation auf dem Torus. Für $\varepsilon > 1.56081$ existieren dann die in Abbildung 7.15 dargestellten invarianten Tori, deren Attraktivität mit wachsendem ε zunimmt. In Abbildung 7.16 wird das

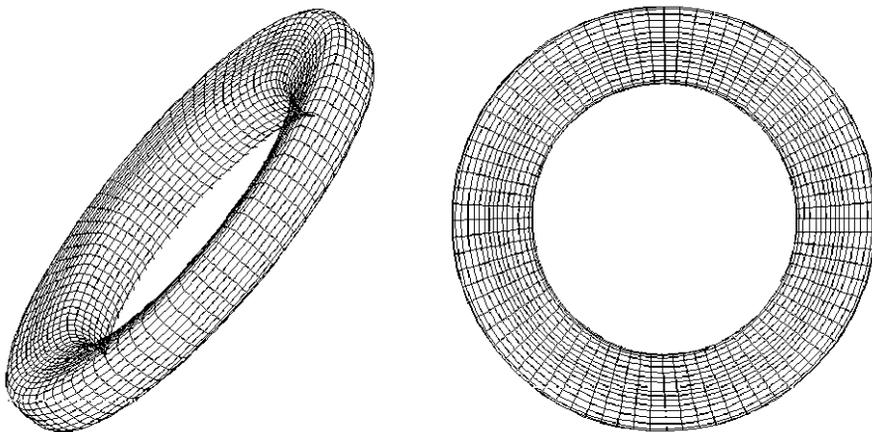


Abb. 7.16 Gittermodell des Torus für $\varepsilon = 5.0$, berechnet mit dem Programm TORUS von Bernet (1995)

Gittermodell des invarianten Torus mit $\varepsilon = 5.0$ in zwei Ansichten dargestellt. Informativer ist allerdings die ermittelte Torusfunktion $u(\theta, t)$ über dem (θ, t) -Gitter in Abbildung

7.17, an der wir die erhebliche Abweichung von \mathcal{M} gegenüber dem Standardtorus \mathbb{T}^2 für den Parameter $\varepsilon = 9.0$ deutlich erkennen können. Aus den Abbildungen wird zugleich ersichtlich, dass sowohl die 1:2-resonanten invariante Tori für $\varepsilon \leq 1.56081$, als auch sämtliche für $\varepsilon > 1.56081$ entstehenden quasi-periodischen und resonante Tori mit dem vorgestellten Verfahren zuverlässig bestimmbar sind.

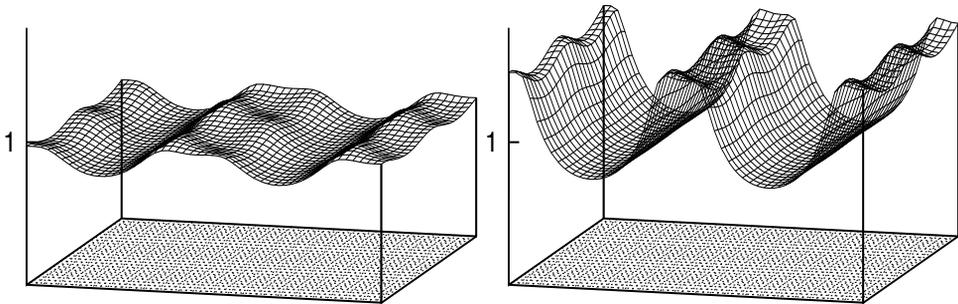


Abb. 7.17 Torusfunktion $u(\theta, t)$ auf dem (θ, t) -Gitter für System (7.74). Links: $B = 0.1$, $\varepsilon = 3$, rechts: $B = 0.1$, $\varepsilon = 9.0$

Fassen wir abschließend die Vorteile dieser Methode zur numerischen Approximation invarianter p -Tori zusammen:

- Während die in den 1980er Jahren von Kaas-Petersen (1985) und Van Veldhuizen (1987) entwickelten ersten Verfahren stark vom Verlauf der Orbits auf dem Torus abhängig waren, besitzt der vorgestellte Zugang über die Lösung des partiellen DGL-Systems (7.40) den Vorteil einer weitgehenden *Unabhängigkeit von den Orbits*. Wie am vorgestellten System (7.71) verdeutlicht, unterscheidet er nicht zwischen quasi-periodischen und resonante p -Tori und ist damit universell einsetzbar.
- Die partielle DGL (7.40) kann numerisch als quasilineares hyperbolisches System zuverlässig mit FDM vom Upwind-Typ gelöst werden – allerdings nur mit Konvergenzordnung 1. Mittels Extrapolation ist eine höhere Konvergenzordnung erreichbar, falls die gegebenen Funktionen hinreichend glatt sind. Dabei lassen *linear implizite Verfahren* (7.70) große t -Schrittweiten zu, womit wegen der θ -Dimension $p - 1$ des entstehenden diskreten Problems der Rechen- und Speicheraufwand erheblich reduziert wird.
- Mit dem *Schießverfahren* 7.23 lassen sich nicht nur asymptotisch stabile p -Tori gemäß Definition 7.16, sondern auch instabile Tori approximieren. Allerdings sollte bei „starker“ Instabilität das Mehrfach-Schießverfahren 6.18 aus Kapitel 6 eingesetzt werden. Den Schwachpunkt des Verfahrens stellt gewiss die erforderliche Ausgangsform (7.21) der DGL in *Toruskoordinaten* $(\theta, u) \in \mathbb{T}^p \times \mathbb{R}^q$ dar. Der Übergang von kartesischen Koordinaten mittels einer Transformation (7.19) gelingt numerisch oft nur bei zuvor bekanntem Lösungsverhalten.

7.3 Quasi-periodische invariante Tori

In diesem Abschnitt wollen wir jede *a-priori*-Transformation in globale Toruskoordinaten $(\theta, u) \in \mathbb{T}^p \times \mathbb{R}^q$ mit $p + q = n$ vermeiden, ebenso eine lokale Transformation in Toruskoordinaten (vgl. Samoilenko (1991)), die meist nur mit großem analytischen Aufwand zu beschaffen sind. Um für das in kartesischen Koordinaten vorgegebene Vektorfeld f eine Invarianzgleichung für einen p -Torus zu gewinnen, müssen wir uns allerdings nun auf *quasi-periodische Tori* beschränken.

Aus der Theorie ist zwar bekannt (vgl. Broer et al. (1996)), dass im dynamischen System externe Parameter vorhanden sein müssen, damit quasi-periodische Tori beobachtet werden können. Aber auch dann treten quasi-periodische Tori nur für solche Mengen von Parameterwerten auf, die die Struktur einer Cantor-Menge haben. Mit anderen Worten: Es gibt eine offene, überall dichte Teilmenge von Parameterwerten, für die kein invarianter quasi-periodischer p -Torus existiert. Beliebige kleine Störungen zerstören zwar die Quasi-Periodizität, nicht aber den Torus!

Da eine numerische Unterscheidung zwischen einem quasi-periodischen und einem schwach resonanten Torus, auf dem periodische Orbits mit sehr großer Periodendauer liegen, kaum möglich ist, dürfen wir davon ausgehen, dass Verfahren für quasi-periodische Tori auch für schwach resonante Tori brauchbare Resultate liefern. Wie schon in Beispiel 7.4 verdeutlicht, sind die im Computer durch Gleitpunktzahlen approximierten Basisfrequenzen schließlich immer rational abhängig.

7.3.1 Eine spezielle Invarianzgleichung für p -Tori

Wir gehen nun von folgendem allgemeinen Modellproblem für $x \in \mathbb{R}^n$, $t \in \mathbb{R}$

$$\frac{dx}{dt} = \dot{x} = f(x, \Omega_1 t, \Omega_2 t, \dots, \Omega_m t) \quad (7.75)$$

mit $m \geq 0$ vorgegebenen *Erregungsfrequenzen* $\Omega := (\Omega_1, \dots, \Omega_m) \in \mathbb{R}^m$ aus, wobei f 2π -periodisch in den Argumenten $2, 3, \dots$ ist. Die Ω_i können als rational unabhängig vorausgesetzt werden; denn wären sie für $m \geq 2$ rational abhängig, so ließe sich ihre Anzahl stets sukzessive reduzieren. Gilt etwa $k_m \Omega_m = \sum_{i=1}^{m-1} k_i \Omega_i$, so substituieren wir Ω_i mittels

$$\begin{aligned} f(x, \Omega_1 t, \Omega_2 t, \dots, \Omega_m t) &= f\left(x, \Omega_1 t, \Omega_2 t, \dots, \Omega_{m-1} t, \frac{1}{k_m} \left(\sum_{i=1}^{m-1} k_i \Omega_i\right) t\right) \\ &=: f_{\text{neu}}(x, \Omega_1 t, \Omega_2 t, \dots, \Omega_{m-1} t) \end{aligned}$$

und untersuchen f_{neu} auf rationale Unabhängigkeit usw. Wie im vorigen Kapitel können wir System (7.75) in eine autonome DGL umwandeln, wenn wir es um m Toruskoordinaten $\theta \in \mathbb{T}^m$ erweitern und das *autonome Modellproblem* erhalten:

$$\dot{x} = f(x, \theta), \quad \dot{\theta} = \Omega \quad \text{mit} \quad (x, \theta) \in \mathbb{R}^n \times \mathbb{T}^m. \quad (7.76)$$

Bemerkung 7.27

Durch die DGL (7.76) wird das C^r -Vektorfeld $f(x, \theta) \frac{\partial}{\partial x} + \Omega \frac{\partial}{\partial \theta}$ auf $\mathbb{R}^n \times \mathbb{T}^m$ gegeben. Unser Modellproblem besitzt einen *invarianten p -Torus \mathcal{M} mit parallelem Fluss (paralleler Dynamik)*, wenn ein Diffeomorphismus $\mathcal{M} \mapsto \mathbb{T}^p$ existiert, der das restringierte C^r -Vektorfeld $(f(x, \theta) \frac{\partial}{\partial x} + \Omega \frac{\partial}{\partial \theta})|_{\mathcal{M}}$ in ein konstantes Vektorfeld

$$\sum_{j=1}^m \Omega_j \frac{\partial}{\partial \theta_j} + \sum_{j=m+1}^p \omega_j \frac{\partial}{\partial \theta_j}$$

auf dem p -dimensionalen Standardtorus \mathbb{T}^p transformiert. ■

Wir wollen nun die Voraussetzungen an unsere Problemstellung formulieren:

Voraussetzung 7.28

i. f ist 2π -periodisch in allen $\Omega_i t$, d. h. für alle $(x, \theta_1, \dots, \theta_m) \in \mathbb{R}^n \times \mathbb{R}^m$ gilt

$$f(x, \theta_1, \dots, \theta_i + 2\pi, \dots, \theta_m) = f(x, \theta_1, \dots, \theta_i, \dots, \theta_m), \quad i = 1(1)m.$$

ii. System (7.75) besitzt eine lokal eindeutige quasi-periodische Lösung $x^* \in C^r(\mathbb{R}, \mathbb{R}^n)$, $r \in \mathbb{N}$, mit p Basisfrequenzen ($m \leq p < m + n$) und der Frequenzbasis

$$(\Omega, \omega) := (\Omega_1, \Omega_2, \dots, \Omega_m, \omega_{m+1}, \dots, \omega_p) \in \mathbb{R}^p.$$

iii. $f \in C^{s+2}(\mathbb{R}^n \times \mathbb{T}^m, \mathbb{R}^n)$ mit $s \in \mathbb{N}$ und $s > r - 1 + \frac{p}{2}$.

Für die freien Frequenzen $\omega \in \mathbb{R}^{p-m}$ der Frequenzbasis (Ω, ω) einer quasi-periodischen Funktion wird im Folgenden stets die verschobene Indizierung $\omega = (\omega_{m+1}, \dots, \omega_p)$ verwendet. Analog zu Kapitel 6 lassen sich nun folgende Modellprobleme unterscheiden:

- autonomer Fall: $p > 0, m = 0$
- periodisch erregter Fall: $p \geq m, m = 1$
- quasi-periodisch erregter Fall: $p \geq m, m > 1$.

Speziell für $p = 1, m = 0$ erhalten wir hieraus den autonomen periodischen Fall, während für $p = m = 1$ ein periodisch erregtes System mit periodischer Systemantwort vorliegt. Zur vorausgesetzten quasi-periodischen Lösung $x^* \in C(\mathbb{R}, \mathbb{R}^n)$ suchen wir eine zugehörige Torusfunktion $u^* \in C(\mathbb{T}^p, \mathbb{R}^n)$ und $p - m$ Basisfrequenzen $\omega_{m+1}, \dots, \omega_p$ mit

$$x^*(t) = u^*(\Omega t, \omega t) = u^*(\Omega_1 t, \dots, \Omega_m t, \omega_{m+1} t, \dots, \omega_p t). \quad (7.77)$$

Wir setzen diesen Ansatz in die DGL (7.75) ein und erhalten nach Differentiation

$$\sum_{i=1}^m \Omega_i \frac{\partial u^*}{\partial \theta_i}(\Omega t, \omega t) + \sum_{j=m+1}^p \omega_j \frac{\partial u^*}{\partial \theta_j}(\Omega t, \omega t) = f(u^*(\Omega t, \omega t), \Omega t).$$

Definieren wir zu beliebigem $u \in C(\mathbb{T}^p, \mathbb{R}^n)$ die Residuumsfunktion $g : \mathbb{R} \rightarrow \mathbb{R}^n$ durch

$$g(t) := \sum_{i=1}^m \Omega_i \frac{\partial u}{\partial \theta_i}(\Omega t, \omega t) + \sum_{i=m+1}^p \omega_i \frac{\partial u}{\partial \theta_i}(\Omega t, \omega t) - f(u(\Omega t, \omega t), \Omega t), \quad (7.78)$$

so verschwindet g auf \mathbb{R} , wenn $u = u^*$ gilt und damit eine Lösung gemäß (7.77) vorliegt. Nach Voraussetzung 7.28 ist $g \in C^{r-1}(\mathbb{R}, \mathbb{R}^n)$ und quasi-periodisch mit Basisfrequenzen $\Omega_1, \dots, \Omega_m, \omega_{m+1}, \dots, \omega_p$. Die zugehörige Torusfunktion $G : \mathbb{T}^p \rightarrow \mathbb{R}^n$ mit der Darstellung

$$g(t) = G(\Omega t, \omega t) = G(\Omega_1 t, \dots, \omega_p t)$$

lautet dann

$$G(\theta) := \sum_{i=1}^m \Omega_i \frac{\partial u}{\partial \theta_i}(\theta) + \sum_{i=m+1}^p \omega_i \frac{\partial u}{\partial \theta_i}(\theta) - f(u(\theta), \theta). \quad (7.79)$$

Gemäß Satz 7.5 gilt damit auch für die Funktionen g und G die Normgleichheit

$$\sup_{t \in \mathbb{R}} |g(t)| = \max_{\theta \in \mathbb{T}^p} |G(\theta)|,$$

womit $g(t) = 0$ für alle $t \in \mathbb{R}$ genau dann gilt, wenn $G(\theta) = 0$ für alle $\theta \in \mathbb{T}^p$ ist. Jede Lösung $u^* \in C(\mathbb{T}^p, \mathbb{R}^n)$ des Torussystems

$$\sum_{i=1}^m \Omega_i \frac{\partial u}{\partial \theta_i}(\theta) + \sum_{i=m+1}^p \omega_i \frac{\partial u}{\partial \theta_i}(\theta) = f(u(\theta), \theta), \quad \theta \in \mathbb{T}^p \quad (7.80)$$

liefert folglich eine quasi-periodische Lösung $x^*(t) = u^*(\Omega t, \omega t)$ der Gleichung $g(t) = 0$, $t \in \mathbb{R}$, mit dem durch (7.78) definierten g und damit auch der gegebenen DGL (7.75).

Die Existenz einer Lösung dieser PDGL ergibt sich mit Voraussetzung 7.28 und Satz 7.5 nach Samoilenko (1991), S.83:

Satz 7.29

$x \in C^r(\mathbb{R}, \mathbb{R}^n)$ sei eine quasi-periodische Funktion mit Frequenzbasis $\omega \in \mathbb{R}^p$ und dem Graphen

$$\gamma := \{(x(t), \omega t) \in \mathbb{R}^n \times \mathbb{T}^p \mid t \in \mathbb{R}\}.$$

Dann existiert eine Torusfunktion $u \in C^r(\mathbb{T}^p, \mathbb{R}^n)$ mit $x(t) = u(\omega t)$, deren Graph

$$\mathcal{M} := \{(u(\theta), \theta) \in \mathbb{R}^n \times \mathbb{T}^p \mid \theta \in \mathbb{T}^p\} = \bar{\gamma},$$

C^r -homöomorph zu \mathbb{T}^p ist.

Folgende Klassifikation dieser *Invarianzgleichung für p -Tori* erscheint nun angebracht:

1. *Response-Fall*: Bei $p = m$ sind alle Basisfrequenzen vorgegeben, keine Phase ist frei und die zweite Summe in (7.80) verschwindet. $x^* = u(\Omega t)$ ist eine „Systemantwort“ (response case).
2. *Torus mit freien Phasen*: Falls $p > m$ ist, so ist die Lösung des Torussystems nicht eindeutig bestimmt. Denn ist $(u(\theta), \omega)$ eine Lösung, so auch

$$(v(\theta), \omega) := (u(\theta + \beta), \omega) \quad \text{mit} \quad \beta = (0, \dots, 0, \beta_{m+1}, \dots, \beta_p)$$

für beliebiges festes $\beta \in \mathbb{R}^p$. Der Torus $u(\theta)$ hat also $(p - m)$ freie Phasen.

3. *Autonomer Fall*: $m = 0 < p$ stellt den „autonomen Grenzfall“ dar, bei dem alle p Phasen frei und damit zu ermitteln sind. Hier verschwindet die erste Summe in (7.80).

Wegen der Vorgabe aller Erregungsfrequenzen Ω bietet der Response-Fall theoretisch und praktisch die geringsten Schwierigkeiten, wogegen die beiden Fälle mit $p > m$ zusätzliche Phasenbedingungen zur gleichzeitigen Bestimmung der unbekanntenen Basisfrequenzen ω erfordern.

Bemerkung 7.30

Die Invarianzgleichung (7.80) verallgemeinert die in den beiden vorhergehenden Kapiteln aufgestellten Bestimmungsgleichungen für Gleichgewichte und periodische Orbits:

- *Gleichgewichtslage (Ruhelage)*: Für $p = 0$ ist $m = 0$, und (7.80) reduziert sich auf

$$0 = f(u), \quad u \in \mathbb{R}^n. \quad (7.81)$$

- *Periodischer Orbit (Grenzyklus)*: Für $p = 1$ und $m = 1$ liegt der periodisch erregte Fall mit bekannter Erregungsfrequenz $\Omega \in \mathbb{R}$ vor:

$$\Omega \frac{du}{dt} = f(u, t). \quad (7.82)$$

$p = 1$ und $m = 0$ liefert den autonomen Fall mit unbekannter Frequenz $\omega \in \mathbb{R}$:

$$\omega \frac{du}{dt} = f(u). \quad (7.83)$$

- *Biperiodischer Orbit mit zwei Basisfrequenzen*: $p = 2$ und $m = 2$ ergibt den quasi-periodisch erregten Fall mit zwei Erregungsfrequenzen $\Omega_1, \Omega_2 \in \mathbb{R}$

$$\Omega_1 \frac{\partial u}{\partial \theta_1} + \Omega_2 \frac{\partial u}{\partial \theta_2} = f(u, \theta_1, \theta_2). \quad (7.84)$$

$p = 2$ und $m = 1$ liefert den periodisch erregten Fall mit einer Erregungsfrequenz Ω und einer unbekanntenen Basisfrequenz $\omega \in \mathbb{R}$

$$\Omega \frac{\partial u}{\partial \theta_1} + \omega \frac{\partial u}{\partial \theta_2} = f(u, \theta_1). \quad (7.85)$$

$p = 2, m = 0$ ist der autonome Fall mit zwei unbekanntenen Basisfrequenzen $\omega_1, \omega_2 \in \mathbb{R}$

$$\omega_1 \frac{\partial u}{\partial \theta_1} + \omega_2 \frac{\partial u}{\partial \theta_2} = f(u). \quad \blacksquare \quad (7.86)$$

Bevor wir die numerische Lösung der PDGL (7.80) in Angriff nehmen, wollen wir sie als Operatorgleichung in geeigneten *Sobolew-Räumen* $\mathcal{H}^s := [\mathcal{H}^s(\mathbb{T}^p)]^n$ notieren. $\mathcal{H}^s(\mathbb{T}^p)$ kann mit dem Raum der bezüglich jeder Variablen θ_j ($j = 1, \dots, p$) 2π -periodischen Funktionen identifiziert werden, die auf \mathbb{T}^p quadratisch summierbar sind und verallgemeinerte Ableitungen bezüglich θ bis einschließlich der Ordnung s besitzen. Im Raum $\mathcal{H}^s(\mathbb{T}^p)$ sind das Skalarprodukt

$$\langle x, y \rangle_{\mathcal{H}^s(\mathbb{T}^p)} := \sum_{|\alpha| \leq s} \langle \partial^\alpha x, \partial^\alpha y \rangle_{\mathcal{H}^0(\mathbb{T}^p)}$$

mit dem verallgemeinerten Differenzialoperator ∂^α der Ordnung $|\alpha|$ zum Multi-Index $\alpha \in [\mathbb{Z}_{\geq 0}]^p$ und die dadurch induzierte Norm $\|x\|_{\mathcal{H}^s(\mathbb{T}^p)}^2 := \langle x, x \rangle_{\mathcal{H}^0(\mathbb{T}^p)}$ definiert. Das Skalarprodukt in $\mathcal{H}^0(\mathbb{T}^p)$ ist dabei durch

$$\langle x, y \rangle_{\mathcal{H}^0(\mathbb{T}^p)} := \frac{1}{(2\pi)^p} \int_{\mathbb{T}^p} x(\theta) \overline{y(\theta)} d\theta$$

gegeben, weshalb speziell $\mathcal{H}^0(\mathbb{T}^p) = \mathcal{L}^2(\mathbb{T}^p)$ ist. Desweiteren genügen die Räume der Einbettungskette

$$\mathcal{L}^2(\mathbb{T}^p) = \mathcal{H}^0(\mathbb{T}^p) \supset \mathcal{H}^1(\mathbb{T}^p) \supset \mathcal{H}^2(\mathbb{T}^p) \supset \dots$$

Nach Voraussetzung 7.28 ist für $s > r - 1 + \frac{p}{2}$ die Einbettung von $\mathcal{H}^{s+1}(\mathbb{T}^p)$ in $\mathcal{C}^r(\mathbb{T}^p)$ kompakt. Gemäß Samoilenko (1991) lassen sich diese Aussagen auch auf die Räume $\mathcal{H}^s := [\mathcal{H}^s(\mathbb{T}^p)]^n$ übertragen. Stattet man die Räume \mathcal{H}^s mit den Skalarprodukten für Räume der Form $[X]^n$ aus, so werden sie Hilbert-Räume. In diesen Hilbert-Räumen definieren wir den verallgemeinerten Differenzialoperator $F : \mathcal{H}^{s+1} \rightarrow \mathcal{H}^s$ mit

$$F(u) := \sum_{i=1}^m \Omega_i \frac{\partial u}{\partial \theta_i}(\theta) + \sum_{i=m+1}^p \omega_i \frac{\partial u}{\partial \theta_i}(\theta) - f(u(\theta), \theta), \quad (7.87)$$

der wohldefiniert und 2-mal Fréchet-differenzierbar ist. Damit wird im *Response-Fall* mit $p = m$ das Newton-Verfahren auf die Operatorgleichung

$$F(u) = 0, \quad u \in \mathcal{H}^{s+1} \quad (7.88)$$

anwendbar und in der Form

$$F'(u^{(k)}) d^{(k)} = -F(u^{(k)}), \quad u^{(k+1)} = u^{(k)} + d^{(k)}, \quad k = 0, 1, 2, \dots \quad (7.89)$$

darstellbar. Unter geeigneten Voraussetzungen an F existiert ein $\delta > 0$, so dass die durch das Verfahren berechnete Funktionenfolge $\{u^{(k)}\}$ für jede Startfunktion $u^{(0)}$ mit $\|u^{(0)} - u^*\| \leq \delta$ wohldefiniert ist und Q-quadratisch gegen die gesuchte Lösung u^* von $F(u) = 0$ konvergiert.

Im *Fall freier Phasen* mit $p > m$ ist die Parametrisierung von u bezüglich $\theta_{m+1}, \dots, \theta_p$ nicht eindeutig durch die Invarianzgleichung festgelegt. Ähnlich wie im autonomen Fall periodischer Orbits in Abschnitt 6.2 können wir nun die unbekanntenen Basisfrequenzen $\omega_{m+1}, \dots, \omega_p$ als künstliche Variablen wählen und die Bestimmungsgleichung um $p - m$ „Phasenbedingungen“ erweitern. Dazu setzen wir voraus, dass während der numerischen Lösungsfortsetzung bereits eine Näherungslösung $\tilde{u} \in \mathcal{H}^{s+1}$ ermittelt wurde. Bestimmen wir den Nullraum

$$\tilde{K} := \text{span}\{\tilde{k}_{m+1}, \dots, \tilde{k}_p\}, \quad \tilde{k}_i := \frac{\partial \tilde{u}}{\partial \theta_i}, \quad i = m + 1, \dots, p,$$

der Linearisierung $L = F'(\tilde{u})$, so können wir die Torusgleichung um $(p - m)$ integrale Phasenbedingungen erweitern

$$\sum_{i=1}^m \Omega_i \frac{\partial u}{\partial \theta_i} + \sum_{i=m+1}^p \omega_i \frac{\partial u}{\partial \theta_i} = f(u(\theta), \theta) \tag{7.90}$$

$$\left\langle \frac{\partial \tilde{u}}{\partial \theta_i}, u \right\rangle_{\mathcal{H}^0} := \frac{1}{(2\pi)^p} \int_{\mathbb{T}^p} \left\langle \frac{\partial \tilde{u}}{\partial \theta_i}, u \right\rangle d\theta = 0, \quad i = m + 1, \dots, p.$$

Mit dem zusammengesetzten Operator $F : \mathcal{H}^{s+1} \times \mathbb{R}^{p-m} \rightarrow \mathcal{H}^s \times \mathbb{R}^{p-m}$ kann dieses erweiterte System nun in der kompakten Darstellung

$$F(u, \omega) := \begin{pmatrix} \sum_{i=1}^m \Omega_i \frac{\partial u}{\partial \theta_i} + \sum_{i=m+1}^p \omega_i \frac{\partial u}{\partial \theta_i} - f(u(\theta), \theta) \\ \int_{\mathbb{T}^p} \left\langle \frac{\partial \tilde{u}}{\partial \theta_{m+1}}, u \right\rangle d\theta \\ \dots\dots\dots \\ \int_{\mathbb{T}^p} \left\langle \frac{\partial \tilde{u}}{\partial \theta_p}, u \right\rangle d\theta \end{pmatrix} = 0 \tag{7.91}$$

notiert werden. Mit beträchtlichem Aufwand kann auch für diesen Fall nachgewiesen werden, dass unter Zusatzvoraussetzungen an F und insbesondere an die Näherungslösung \tilde{u} das Newton-Verfahren

$$(u^{(k+1)}, \omega^{(k+1)}) = (u^{(k)}, \omega^{(k)}) - [F'(u^{(k)}, \omega^{(k)})]^{-1} F(u^{(k)}, \omega^{(k)}) \tag{7.92}$$

mit $k = 0, 1, 2, \dots$ Q-quadratisch gegen die gesuchte Lösung (u^*, ω^*) von $F(u, \omega) = 0$ konvergiert (vgl. Schilder (2004)).

7.3.2 Semidiskretisierung von 2-Tori

Bevor wir die allgemeine Invarianzgleichung (7.80) bzw. die erweiterte Gleichung (7.90) bei Vorliegen freier Phasen weiter bearbeiten, wollen wir die in technischen Anwendungen sehr häufig auftretende periodisch erregte DGL

$$\dot{x} = f(x, t), \quad f : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n \quad (7.93)$$

behandeln, für die eine quasi-periodische Lösung $x(t)$ mit zwei Basisfrequenzen gesucht ist. Typische Anwendungen stellen mechanische und elektrische Schwingkreise dar, z. B. die energetischen Systeme des Abschnittes 6.4.4. Ohne wesentliche Änderungen ist eine Übertragung auf quasi-periodisch erregte Systeme mit zwei Basisfrequenzen möglich, wogegen die Behandlung autonomer Systeme aufwendiger wäre.

Die Voraussetzungen 7.28 seien mit $p = 2$ erfüllt, wobei dann für die gegebene Erregungsfrequenz $\Omega = 1$ angenommen werden kann. Gesucht sind damit die unbekannte Basisfrequenz $\omega \in \mathbb{R}$ und eine Lösung der Invarianzgleichung (7.85). Mit den vereinfachten Bezeichnungen $\theta_1 = t$ und $\theta_2 = \theta$ lautet diese PDGL nun

$$\frac{\partial u}{\partial t} + \omega \frac{\partial u}{\partial \theta} = f(u, t) \quad \text{für } u = u(t, \theta). \quad (7.94)$$

Um die *Galerkin-Methode* bezüglich der Variablen $\theta_1 = t$ anzuwenden, wählen wir periodische Basisfunktionen $\varphi_k : \mathbb{T}^1 \rightarrow \mathbb{C}$ mit

$$\varphi_k(t) := e^{ikt}, \quad k \in \mathbb{Z}, \quad i = \sqrt{-1}. \quad (7.95)$$

Diese Funktionen bilden bekanntermaßen ein Orthonormalsystem im Hilbert-Raum $\mathcal{H}^0(\mathbb{T}^1) = \mathcal{L}^2(\mathbb{T}^1)$ mit dem Skalarprodukt

$$\langle x, y \rangle_{\mathcal{H}^0(\mathbb{T}^1)} := \frac{1}{2\pi} \int_0^{2\pi} x(t) \overline{y(t)} dt.$$

Da wir eine Lösung $u \in \mathcal{H}^s$ der Gleichung (7.94) suchen, definieren wir zu gegebenem *Diskretisierungsparameter* $N \in \mathbb{N}$ die linearen Unterräume $\mathcal{H}_N^0(\mathbb{T}^2) \subset \mathcal{H}^0(\mathbb{T}^2)$ mit

$$\mathcal{H}_N^0(\mathbb{T}^2) := \text{span}\{u \cdot \varphi_k \mid u \in \mathcal{H}^0(\mathbb{T}^1), k = -N(1)N\}$$

und kürzen deren n -faches Produkt analog zu \mathcal{H}^0 mit $\mathcal{H}_N^0 := [\mathcal{H}_N^0(\mathbb{T}^2)]^n$ ab. Für die Semidiskretisierung führen wir zu $u : \mathbb{T}^2 \rightarrow \mathbb{R}^n$, $u \in \mathcal{H}^0$, den Projektionsoperator $P_N : \mathcal{H}^0 \rightarrow \mathcal{H}_N^0$ mit

$$(P_N u)(t, \theta) := \sum_{k=-N}^N u_k(\theta) \varphi_k(t) \quad (7.96)$$

ein. Die vektoriellen Fourier-Koeffizienten $u_k : \mathbb{T}^1 \rightarrow \mathbb{C}^n$ werden dann durch

$$u_k(\theta) := \frac{1}{2\pi} \int_0^{2\pi} u(t, \theta) \overline{\varphi_k(t)} dt.$$

definiert. Die Elemente des Unterraumes \mathcal{H}_N^0 mit der Darstellung

$$u^N = u^N(t, \theta) := \sum_{k=-N}^N u_k(\theta) \varphi_k(t) \quad (7.97)$$

setzen wir nun in PDGL (7.94) ein und wenden darauf den Projektionsoperator P_N an

$$P_N \left(\frac{\partial u^N}{\partial t} + \omega \frac{\partial u^N}{\partial \theta} - f(u^N, t) \right) = 0.$$

Mit der Linearität des Operators P_N und der Ableitung $\dot{\varphi}_k(t) = ik\varphi_k(t)$ der Basisfunktionen gewinnen wir hieraus die Gleichung

$$r^N(t, \theta) := \sum_{k=-N}^N iku_k(\theta) \varphi_k(t) + \sum_{k=-N}^N \omega u'_k(\theta) \varphi_k(t) - \sum_{k=-N}^N h(u^N) \varphi_k(t) = 0,$$

wobei der k -te Fourier-Koeffizient der rechten Gleichungsseite $f(u^N(t, \theta), t)$

$$h_k(u^N) := \frac{1}{2\pi} \int_0^{2\pi} f(u^N(t, \theta), t) \overline{\varphi_k(t)} dt \quad (7.98)$$

lautet und damit nicht von t abhängt. Anwendung der Galerkin-Bedingung, d. h. der Orthogonalität des Residuums $r^N(t, \theta)$ zu allen $2N+1$ Basisfunktionen $\varphi_l(t)$, $l = -N(1)N$,

$$\frac{1}{2\pi} \int_0^{2\pi} r^N(t, \theta) \overline{\varphi_l(t)} dt = 0, \quad l = -N(1)N,$$

liefert schließlich das *Galerkin-System*, auch *Spektralsystem* genannt,

$$\omega u'_k(\theta) = h_k(u^N) - iku_k(\theta), \quad k = -N(1)N \quad (7.99)$$

für die gesuchten komplexwertigen Fourier-Koeffizienten $u_k(\theta)$ der Lösungsapproximation (7.97). Diese Gleichungen bilden ein autonomes System aus $(2N+1)n$ gewöhnlichen Differenzialgleichungen 1. Ordnung für die periodischen Koeffizientenfunktionen $u_k(\theta)$ unbekannter Periodendauer $T = 2\pi/\omega$. Damit führt die Semidiskretisierung bezüglich t durch den Ansatz (7.97) auf ein standardmäßiges Periodizitätsproblem, wofür wir in Kapitel 6 leistungsfähige Verfahren entwickelt haben.

Bevor wir das Spektralsystem und insbesondere die Funktionen $h_k(u^N)$ generieren, führen wir eine kompaktere Notation der $(2N + 1)n$ DGL ein. Dazu fassen wir mit der Matrixfunktion $U : \mathbb{T}^1 \rightarrow \mathbb{C}^{n \times (2N+1)}$ und der Vektorfunktion $\varphi : \mathbb{T}^1 \rightarrow \mathbb{C}^n$ gemäß

$$U(\theta) = (u_{-N}(\theta), \dots, u_0(\theta), \dots, u_N(\theta)) \quad \text{und} \quad \varphi(t) = (\varphi_{-N}(t), \dots, \varphi_0(t), \dots, \varphi_N(t))^T$$

die Funktionen zusammen und können den Ansatz (7.97) damit in der Form

$$u^N(t, \theta) = U(\theta)\varphi(t) \tag{7.100}$$

notieren. Analog fassen wir die Fourier-Koeffizienten $h_k(u^N)$ aus (7.98) zur Matrix

$$H = H(U(\theta)) = (h_{-N}(u^N), \dots, h_0(u^N), \dots, h_N(u^N))$$

zusammen und definieren schließlich die $(2N + 1) \times (2N + 1)$ -Diagonalmatrix $\Phi := \text{diag}(-iN, -i(N - 1), \dots, -2i, -i, 0, i, 2i, \dots, i(N - 1), iN)$. Damit lautet das Spektralsystem nun in Matrixnotation

$$\omega U'(\theta) = H(U(\theta)) - U(\theta)\Phi. \tag{7.101}$$

Aus der Herleitung dieses Systems gewinnen wir im Grenzfall $N \rightarrow \infty$ leicht folgenden

Satz 7.31

Unter den Voraussetzungen 7.28 gilt für $N \rightarrow \infty$: $x(t) = U(\theta)\varphi(t)$ ist eine quasi-periodische Lösung des Originalsystems (7.93) mit zwei Basisfrequenzen $(1, \omega)$ genau dann, wenn $U(\theta)$ eine 2π -periodische Lösung des (unendlich-dimensionalen) Spektralsystems (7.101) mit irrationaler Frequenz ω ist.

Da in der Regel die zweite Basisfrequenz ω unbekannt ist, kann diese formal eliminiert werden. Setzen wir $\theta = \omega\tau$ und damit $U(\theta) = U(\omega\tau) =: Y(\tau)$, so lautet das transformierte Spektralsystem

$$Y'(\tau) = H(Y(\tau)) - Y(\tau)\Phi. \tag{7.102}$$

Gesucht sind nun periodische Lösungen $Y(\tau)$ mit unbekannter Periodendauer T , nach deren Ermittlung sich die zweite Basisfrequenz zu $\omega = 2\pi/T$ ergibt.

Bemerkung 7.32

1. Mit der Lösungsdarstellung (7.97) lässt sich die Invarianzkurve γ_1 der Stroboskop-Abbildung bezüglich θ_1 genauso leicht approximieren wie die zweite Stroboskop-Abbildung γ_2 bezüglich θ_2 :

$$\gamma_1(\theta_2) := \sum_{k=-N}^N u_k(\theta_2)\varphi_k(0) \quad \text{und} \quad \gamma_2(\theta_1) := \sum_{k=-N}^N u_k(0)\varphi_k(\theta_1). \tag{7.103}$$

2. Mittels des Spektralsystems (7.102) können auch Gleichgewichtspunkte bestimmt werden. Jede Matrix $A = (a_{kl})$, die dem System

$$A\Phi = H(A), \quad A \in \mathbb{C}^{n \times (2N+1)} \tag{7.104}$$

genügt, generiert eine 2π -periodische Näherungslösung $x(t) = A\varphi(t)$ der Original-DGL. Im Grenzfall $N \rightarrow \infty$ gilt unter den Voraussetzungen 7.28: $A = (a_{kl})$, $k = 1(1)n$, $l \in \mathbb{Z}$, ist eine Lösung des Gleichungssystems (7.104) und damit Gleichgewichtslage des Spektralsystems, genau dann, wenn $x(t) = A\varphi(t)$ eine 2π -periodische Lösung des Originalsystems (7.93) ist. ■

Diese Aussage stellt ein klassisches Resultat dar (vgl. Hale (1963)), das durch Satz 7.31 auf den Fall quasi-periodischer Lösungen des Originalsystems verallgemeinert wurde. Die Voraussetzung über die Inkommensurabilität der Basisfrequenzen ist dabei unabdingbar, so dass der Fall resonanter Tori gesondert untersucht werden muss.

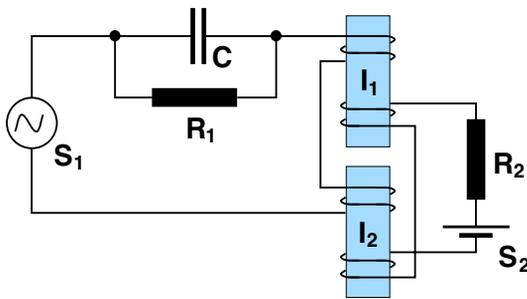


Abb. 7.18 Elektrischer Resonanzkreis mit zwei nichtlinearen Induktivitäten I_1 und I_2 , deren Kennlinie als kubisch angenommen wird. Es besteht aus dem periodisch erregten System mit sinusförmiger Spannungsquelle S_1 und den Sekundärwicklungen mit Gleichspannungsquelle S_2 (Bild aus Vogt et al. (2005))

Anwendung auf einen nichtlinearen elektrischen Resonanzkreis

Wir betrachten ein System von Hayashi (1985)², das den in Abbildung 7.18 dargestellten elektrischen Resonanzkreis mit zwei nichtlinearen Induktivitäten beschreibt. Das System vom Duffing-Typ

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -k_1 x_2 - \frac{1}{8}(x_1^2 + 3x_3^2)x_1 + B \cos t \\ \dot{x}_3 &= -\frac{1}{8}k_2(3x_1^2 + x_3^2)x_3 + B_0 \end{aligned} \tag{7.105}$$

reduziert sich für $x_3 \equiv 0$ und $B_0 = 0$ auf die klassische Duffing-Gleichung. Wir wollen das qualitative Verhalten der Lösungen für die Parameterwerte $B_0 = 0.03$, $B = 0.22$, $k_2 = 0.05$ und den Kontrollparameter $k_1 \in [0.043, 0.09]$ untersuchen. Mit numerischer Integration kann eine Torus-Bifurkation der 2π -periodischen Lösung für $k_1^* \approx 0.1214$ beobachtet werden. Für kleinere Werte $k_1 < k_1^*$ entsteht eine stabile quasi-periodische

²Das System wurde von Yoshinga und Kawakami (1995) behandelt und in Arbeiten von Vogt et al. (2005) und Engert (2009) weiter analysiert.

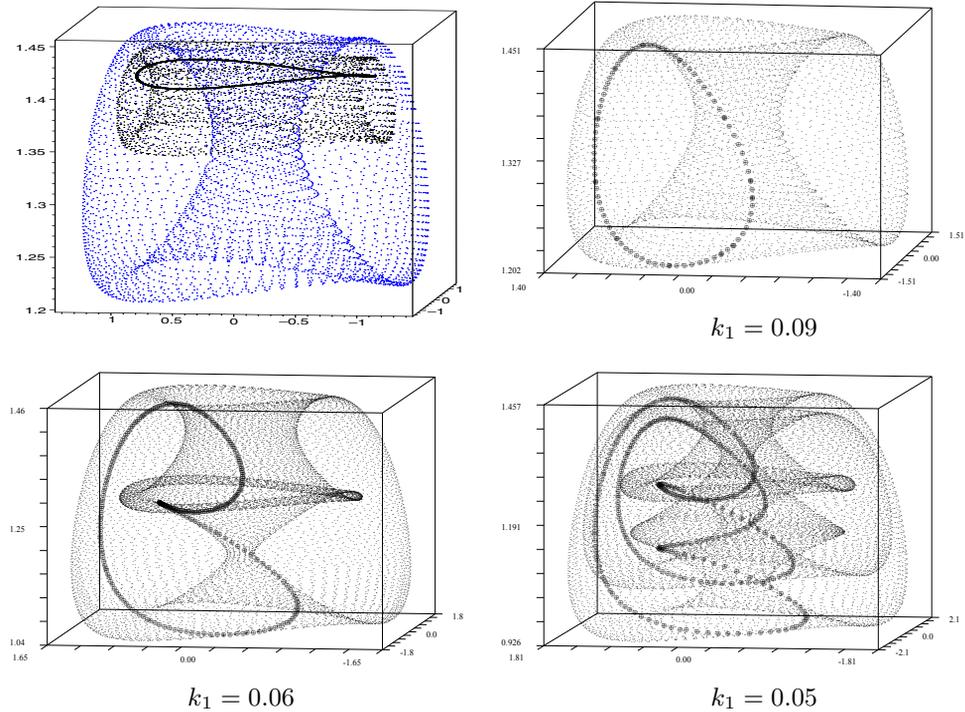


Abb. 7.19 Lösungsszenarien des dynamischen Systems (7.105). Lösungen im Bild oben links: $k_1 = 0.15$ (periodisch), $k_1 = 0.12$ (schwarz) und $k_1 = 0.09$ (blau)

Lösung mit zwei Basisfrequenzen, die in Abbildung 7.19 weiter verfolgt wird. Es wird eine Kaskade von Periodenverdopplungen bezüglich einer der beiden Basisfrequenzen (Torusdopplungen) erkennbar, die schließlich in eine chaotische Lösung übergeht. Die stroboskopische Poincaré-Abbildung mit Periode $T = 2\pi$ wird in Abbildung 7.19 durch die dick eingezeichneten Punkte dargestellt. Zur Bestimmung der 2-Tori mittels Semi-diskretisierung der PDGL (7.94) wählen wir einen reellen Ansatz erster Ordnung für das Spektralsystem (7.96)

$$\begin{pmatrix} u_1(t, \theta) \\ u_2(t, \theta) \\ u_3(t, \theta) \end{pmatrix} = \begin{pmatrix} y_1(\theta) & y_2(\theta) & y_3(\theta) \\ y_4(\theta) & y_5(\theta) & y_6(\theta) \\ y_7(\theta) & y_8(\theta) & y_9(\theta) \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \sin(t) \\ \cos(t) \end{pmatrix} \tag{7.106}$$

bzw. in Matrixform $u(t, \theta) = Y(\theta)\varphi(t)$ mit

$$Y(\theta) = \begin{pmatrix} y_1(\theta) & y_2(\theta) & y_3(\theta) \\ y_4(\theta) & y_5(\theta) & y_6(\theta) \\ y_7(\theta) & y_8(\theta) & y_9(\theta) \end{pmatrix} \quad \text{und} \quad \varphi(t) = \begin{pmatrix} 1 \\ \sin(t) \\ \cos(t) \end{pmatrix},$$

da das nachfolgend genutzte Programmpaket AUTO2000 nur reelle Systeme analysieren kann. Setzen wir Ansatz (7.106) in die PDGL (7.94) ein, so erhalten wir nach einigen Umformungen und Anwendung der Galerkin-Bedingung – hier ist tatsächlich ein Computeralgebra-System wie MAPLE überaus hilfreich – folgendes Spektralsystem, wobei der Strich die Ableitung nach $\theta = \omega t$ kennzeichnet:

$$\begin{aligned}
 y_1' &= y_4 \\
 y_2' &= y_5 + y_3 \\
 y_3' &= y_6 - y_2 \\
 y_4' &= -0.1875 y_1 y_8^2 - k_1 y_4 - 0.1875 y_1 y_9^2 - 0.1875 y_1 y_2^2 - 0.375 y_1 y_7^2 \\
 &\quad - 0.375 y_2 y_7 y_8 - 0.375 y_3 y_7 y_9 - 0.1875 y_1 y_3^2 - 0.125 y_1^3 \\
 y_5' &= -0.375 y_2 y_1^2 - 0.375 y_2 y_7^2 - 0.75 y_1 y_7 y_8 - 0.28125 y_2 y_8^2 + y_6 \\
 &\quad - 0.09375 y_2 y_3^2 - 0.09375 y_2 y_9^2 - k_1 y_5 - 0.09375 y_2^3 \\
 &\quad - 0.1875 y_3 y_9 y_8 \\
 y_6' &= -0.375 y_3 y_7^2 - 0.375 y_3 y_1^2 - 0.28125 y_3 y_9^2 - 0.09375 y_3 y_8^2 \\
 &\quad - 0.1875 y_2 y_9 y_8 + B - y_5 - 0.09375 y_3^3 - k_1 y_6 - 0.75 y_1 y_7 y_9 \\
 &\quad - 0.09375 y_3 y_2^2 \\
 y_7' &= -0.375 k_2 y_9 y_1 y_3 - 0.1875 k_2 y_7 y_2^2 - 0.1875 k_2 y_9^2 y_7 \\
 &\quad - 0.1875 k_2 y_7 y_3^2 - 0.375 k_2 y_7 y_1^2 - 0.125 k_2 y_7^3 + B0 \\
 &\quad - 0.1875 k_2 y_8^2 y_7 - 0.375 k_2 y_8 y_1 y_2 \\
 y_8' &= -0.375 k_2 y_8 y_1^2 - 0.09375 k_2 y_8^3 - 0.75 k_2 y_7 y_1 y_2 + y_9 \\
 &\quad - 0.1875 k_2 y_9 y_3 y_2 - 0.09375 k_2 y_8 y_9^2 - 0.09375 k_2 y_8 y_3^2 \\
 &\quad - 0.375 k_2 y_8 y_7^2 - 0.28125 k_2 y_8 y_2^2 \\
 y_9' &= -0.75 k_2 y_7 y_1 y_3 - 0.09375 k_2 y_9^3 - 0.09375 k_2 y_8^2 y_9 \\
 &\quad - 0.09375 k_2 y_9 y_2^2 - 0.375 k_2 y_9 y_7^2 - 0.1875 k_2 y_8 y_3 y_2 - y_8 \\
 &\quad - 0.375 k_2 y_9 y_1^2 - 0.28125 k_2 y_9 y_3^2.
 \end{aligned}$$

Für dieses System sind nun Ruhelagen und periodische Lösungen gesucht, wobei erstere periodischen und letztere quasi-periodischen Lösungen des Originalsystems (7.105) entsprechen. Die numerische Analyse des Spektralsystems wurde mit dem Programmpaket AUTO2000 durchgeführt (vgl. Doedel et al. (2002)), die Ergebnisse sind in den Abbildungen 7.20 und 7.21 grafisch dargestellt. Die Identifikation periodischer Lösungen des Spektralsystems mit quasi-periodischen Lösungen des Originalsystems gemäß Tabelle 7.1 ist prinzipiell erfolgreich, jedoch stimmen die Stabilitätsverhältnisse der einander zugeordnete

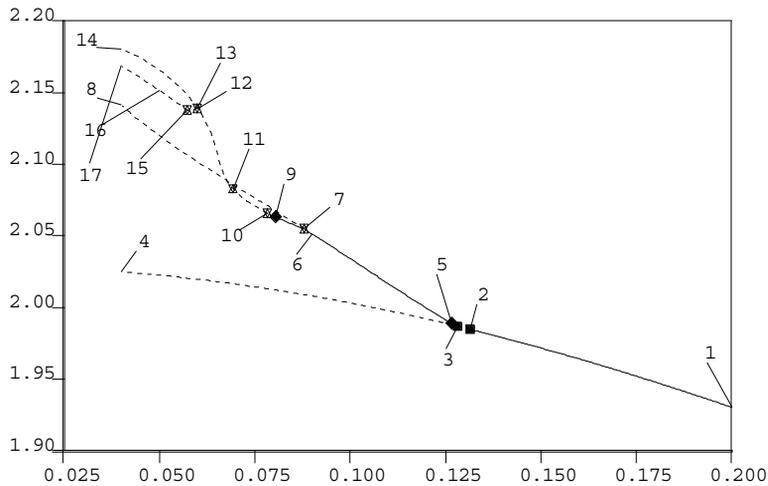


Abb. 7.20 Bifurkationsdiagramm des Spektralsystems für $k_1 \in [0.025, 0.20]$: Hopf-Bifurkationen treten bei $k_1 \approx 0.1315$ und 0.1281 (Label 2 und 3) auf. Der Zweig ab Label 3 für $k_1 \approx 0.1281$ wurde weiter verfolgt. Es tritt eine Kaskade von Periodenverdopplungen (Label 7, 13, 15) auf

ten Objekte nur bedingt überein. Eine zuverlässige Stabilitätsanalyse quasi-periodischer Lösungen ist mittels der hier vorgestellten Spektralmethode noch nicht möglich. Anders als im behandelten Fall lassen sich Spektralsysteme höherer Ordnung, d. h. mit größerem Diskretisierungsparameter N , in der Regel nicht mehr in analytischer Form generieren. Damit entfällt der Vorteil der Semidiskretisierung gegenüber einer kompletten Diskretisierung bezüglich aller unabhängigen Variablen θ_i . Im praxisrelevanten Fall von 2-Tori bei periodisch erregten Systemen bietet sie jedoch eine interessante und überaus effiziente Alternative zur Volldiskretisierung.

Tab. 7.1 Interpretation des Bifurkationsdiagramms in Abbildung 7.20 für das Spektralsystem

Label-Nr.	Spektralsystem	Originalsystem
1	stabile Gleichgewichtslage	stabile periodische Lösung
3	Hopf-Bifurkation	Torus-Bifurkation
6	stabile periodische Lösung	stabiler invarianter Torus
7	Periodenverdopplung	Torus-Doppelung
12	instabile 2-periodische Lösung	stabiler invarianter Doppel-Torus
13	Periodenverdopplung	Torus-Doppelung
16	instabile 4-periodische Lösung	stabiler invarianter Vierfach-Torus
15	Periodenverdopplung	Torus-Doppelung
—	seltener Attraktor	seltener Attraktor

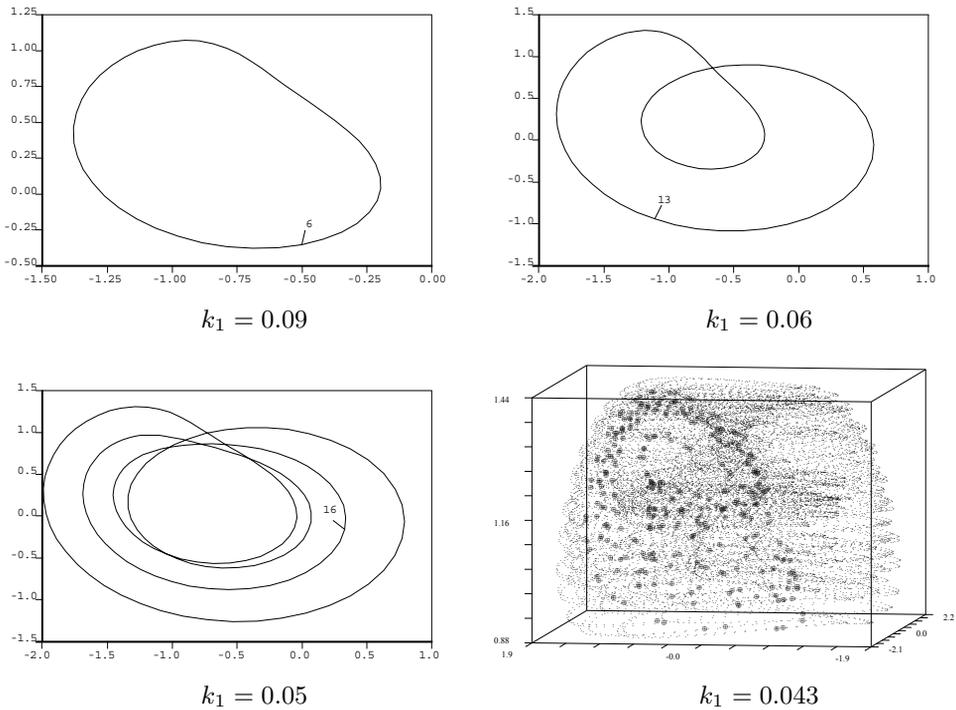


Abb. 7.21 Drei periodische Orbits des Spektralsystems; es tritt offenbar eine Sequenz von Periodenverdopplungen auf. Für $k_1 = 0.043$ wird ein Orbit zusammen mit den Stroboskop-Punkten (fette Punkte) dargestellt

7.3.3 Vollandiskretisierung mit Fourier-Galerkin-Methode

Abgesehen von 2-Tori bei periodisch erregten Systemen ist eine vollständige Diskretisierung zur Lösung der Invarianzgleichung (7.80)

$$\sum_{j=1}^m \Omega_j \frac{\partial u}{\partial \theta_j}(\theta) + \sum_{j=m+1}^p \omega_j \frac{\partial u}{\partial \theta_j}(\theta) = f(u(\theta), \theta), \quad \theta \in \mathbb{T}^p \tag{7.107}$$

sinnvoll. Dazu bieten sich *Finite-Differenzen-Verfahren*, *Galerkin-Verfahren* und *Kollokationsverfahren* gleichermaßen an. Wir nutzen die Methode von Galerkin³ und setzen die gesuchte Lösung $u = u(\theta)$ in Form p -dimensionaler Fourier-Polynome an. Diese *Fourier-Galerkin-Methode*, die mit ihrem Ansatz eine Spektralmethode bildet, erweist sich als besonders erfolgreich, wenn die Funktionen in (7.107) gut durch Fourier-Polynome approximiert werden können. Bei hinreichender Glattheit der Problemdata liefern die Finite-Differenzen-Methode in Schilder (2004) und das Kollokationsverfahren bei Schreiber (2004) ähnliche Resultate.

³ Boris Grigorjevich Galerkin (1871–1945), sowjetischer Ingenieur und Mathematiker, veröffentlichte 1915 seine Methode, die die Tradition der Variationsmethoden (Euler, Lagrange) fortsetzt.

Wir fassen die Invarianzgleichung (7.107) als nichtlineare Operatorgleichung in den im vorigen Abschnitt eingeführten Sobolew-Räumen $\mathcal{H}^s := [\mathcal{H}^s(\mathbb{T}^p)]^n$ auf, und führen den Operator $F : \mathcal{H}^{s+1} \times \mathbb{R}^{p-m} \rightarrow \mathcal{H}^s$ mit

$$F(u, \omega)(\theta) := \sum_{j=1}^m \Omega_j \frac{\partial u}{\partial \theta_j}(\theta) + \sum_{j=m+1}^p \omega_j \frac{\partial u}{\partial \theta_j}(\theta) - f(u(\theta), \theta) \tag{7.108}$$

ein. Mit Hilfe des linearen Differenzialoperators $D : \mathcal{H}^{s+1} \times \mathbb{R}^{p-m} \rightarrow \mathcal{H}^s$ mit

$$D(u, \omega)(\theta) := \sum_{j=1}^m \Omega_j \frac{\partial u}{\partial \theta_j}(\theta) + \sum_{j=m+1}^p \omega_j \frac{\partial u}{\partial \theta_j}(\theta) \tag{7.109}$$

und des nichtlinearen Nemytski-Operators $N : \mathcal{H}^s \rightarrow \mathcal{H}^s$ mit

$$N(u)(\theta) := -f(u(\theta), \theta) \tag{7.110}$$

können wir nun die Operatorgleichung kürzer in der Form

$$F(u, \omega) = D(u, \omega) + N(u) = 0, \quad (u, \omega) \in \mathcal{H}^{s+1} \times \mathbb{R}^{p-m} \tag{7.111}$$

notieren. Diese Gleichung ist im Fall $p > m$ mit $p - m$ freien Basisfrequenzen ω ein schlecht gestelltes Problem, da keine Eindeutigkeit der Lösung vorliegt. Nach der Diskretisierung wollen wir diese Tatsache berücksichtigen und die entstehenden nichtlinearen Gleichungssysteme mit dem Gauß-Newton-Verfahren lösen.

Wir setzen die Lösung u der nichtlinearen Operatorgleichung (7.111) als Fourier-Polynom

$$u^N(\theta) = \sum_{\|k\| \leq N} u_k \varphi_k(\theta) \tag{7.112}$$

mit Multi-Index $k = (k_1, \dots, k_p)$ und den trigonometrischen Basisfunktionen

$$\varphi_k(\theta) = e^{i(k, \theta)}, \quad (k, \theta) = \sum_{j=1}^p k_j \theta_j, \tag{7.113}$$

an. Wir wollen vereinbaren, dass $\|k\| \leq N$ soviel wie $|k_j| \leq N_j$, $j = 1(1)p$ mit dem Diskretisierungsparameter $N := \min\{N_1, N_2, \dots, N_p\}$ bedeutet. Die partiellen Ableitungen von $u^N(\theta)$ berechnen wir nach

$$\frac{\partial u^N}{\partial \theta_j}(\theta) = \sum_{\|k\| \leq N} ik_j u_k \varphi_k(\theta), \quad j = 1(1)p,$$

und erhalten damit für den Differenzialoperator (7.109) die Darstellung

$$D(u^N, \omega)(\theta) = \sum_{\|k\| \leq N} \left(\sum_{j=1}^m \Omega_j k_j + \sum_{j=m+1}^p \omega_j k_j \right) i u_k \varphi_k(\theta). \tag{7.114}$$

Wenden wir die Galerkin-Bedingung (d. h. die Forderung nach Orthogonalität des Residuums bezüglich aller Basisfunktionen) auf den Operator F an, so bedeutet dies

$$\frac{1}{(2\pi)^p} \int_{\mathbb{T}^p} [D(u^N, \omega)(\theta) - f(u^N(\theta), \theta)] \overline{\varphi_k(\theta)} d\theta = 0, \quad \|k\| \leq N,$$

wobei das Integral komponentenweise auszuwerten ist. Mit (7.114) und unter Beachtung der Orthonormalität der Basisfunktionen $\varphi_k(\theta)$ erhalten wir

$$\int_{\mathbb{T}^p} D(u^N, \omega)(\theta) \overline{\varphi_k(\theta)} d\theta = \left(\sum_{j=1}^m \Omega_j k_j + \sum_{j=m+1}^p \omega_j k_j \right) iu_k.$$

Notieren wir einen Fourier-Ansatz M -ter Ordnung für $f(u^N(\theta), \theta)$

$$f(u^N(\theta), \theta) \cong \sum_{\|k\| \leq M} f_k \varphi_k(\theta) \quad \text{mit} \quad f_k = \frac{1}{(2\pi)^p} \int_{\mathbb{T}^p} f(u(\theta), \theta) \overline{\varphi_k(\theta)} d\theta,$$

und $M \geq N$, so führt die Galerkin-Bedingung auf das komplexwertige Gleichungssystem

$$G_k(\tilde{u}, \omega) := \left(\sum_{j=1}^m \Omega_j k_j + \sum_{j=m+1}^p \omega_j k_j \right) iu_k - f_k = 0, \quad \|k\| \leq N \quad (7.115)$$

mit den $m := \prod_{j=1}^p (2N_j + 1)$ vektoriellen Koeffizienten $u_k \in \mathbb{C}^n$ und den reellen Unbekannten $\omega = (\omega_{m+1}, \dots, \omega_p)^T$. Dieses Gleichungssystem ist im Allgemeinen nichtlinear, da die u_k implizit in die Fourier-Koeffizienten f_k eingehen. Im Response-Fall fällt bei obigem Gleichungssystem die zweite Summe weg und das System ist normalbestimmt.

Wir fassen die Unbekannten $u_k \in \mathbb{C}^n$ in dem Supervektor \tilde{u} in lexikographischer Ordnung bezüglich der p Indizes zusammen mittels

$$\tilde{u} = (u_k)_{\|k\| \leq N}^{\text{lex}} := \begin{pmatrix} u_{-N_1|-N_2|\dots|-N_{p-1}|-N_p} \\ u_{-N_1|-N_2|\dots|-N_{p-1}|-N_p+1} \\ \vdots \\ u_{-N_1|-N_2|\dots|-N_{p-1}|N_p} \\ u_{-N_1|-N_2|\dots|-N_{p-1}+1|-N_p} \\ \vdots \\ u_{N_1|N_2|\dots|N_{p-1}|N_p} \end{pmatrix} \in \mathbb{C}^{n \cdot m}. \quad (7.116)$$

In analoger Weise bilden wir mit den Vektoren $G_k(\tilde{u}, \omega)$ in lexikographischer Anordnung einen Supervektor $\tilde{G}(\tilde{u}, \omega) = (G_k(\tilde{u}, \omega))_{\|k\| \leq N}^{\text{lex}}$ und erhalten so die Nullstellenaufgabe

$$\tilde{G}(\tilde{u}, \omega) = 0 \quad \text{mit} \quad \tilde{u} \in \mathbb{C}^{n \cdot m}, \quad \omega \in \mathbb{R}^{p-m}. \quad (7.117)$$

Im Falle $p > m$ ist dieses System unterbestimmt, da es m Unbekannte $u_k \in \mathbb{C}^n$ und $p - m$ reelle Unbekannte ω_j , aber nur nm Gleichungen besitzt. Zur Lösung derartiger algebraischer Gleichungssysteme bietet sich das in Kapitel 5 eingeführte Gauß-Newton-Verfahren (5.67) in der Form

$$\begin{pmatrix} \tilde{u}^{\nu+1} \\ \omega^{\nu+1} \end{pmatrix} = \begin{pmatrix} \tilde{u}^\nu \\ \omega^\nu \end{pmatrix} - [\tilde{G}'(\tilde{u}^\nu, \omega^\nu)]^+ \tilde{G}(\tilde{u}^\nu, \omega^\nu), \quad \nu = 1, 2, \dots \tag{7.118}$$

an. Dabei bezeichnet $[\tilde{G}'(\tilde{u}^\nu, \omega^\nu)]^+$ die Pseudoinverse von $\tilde{G}'(\tilde{u}^\nu, \omega^\nu)$ (vgl. Definition 5.36). Wir notieren dieses Verfahren in der praktikableren Form für $\nu = 0, 1, 2, \dots$

$$\tilde{G}'(\tilde{u}^\nu, \omega^\nu) \begin{pmatrix} z \\ \xi \end{pmatrix} = -\tilde{G}(\tilde{u}^\nu, \omega^\nu), \quad \begin{pmatrix} \tilde{u}^{\nu+1} \\ \omega^{\nu+1} \end{pmatrix} = \begin{pmatrix} \tilde{u}^\nu \\ \omega^\nu \end{pmatrix} + \begin{pmatrix} z \\ \xi \end{pmatrix} \tag{7.119}$$

mit dem lexikographisch geordneten Supervektor $z = (z_k)_{\|k\| \leq N}^{\text{lex}}$ und $\xi = (\xi_{m+1}, \dots, \xi_p)^T$ als Newton-Korrekturen. Die Block-Komponenten der Jacobi-Matrix $\tilde{G}'(\tilde{u}, \omega)$ berechnen wir mittels

$$\begin{aligned} G'_k(\tilde{u}, \omega) &= \left(\frac{\partial}{\partial u_l} G_k(\tilde{u}, \omega), \frac{\partial}{\partial \omega} G_k(\tilde{u}, \omega) \right) \\ &= \left(\left(\frac{\partial}{\partial u_l} G_k(\tilde{u}, \omega) \right)_{\|l\| \leq N}^{\text{lex}}, \frac{\partial}{\partial \omega_{m+1}} G_k(\tilde{u}, \omega), \dots, \frac{\partial}{\partial \omega_p} G_k(\tilde{u}, \omega) \right), \end{aligned}$$

wobei sich für $\|l\| \leq N$ die m komplexwertigen $(n \times n)$ -Blockmatrizen

$$\frac{\partial}{\partial u_l} G_k(\tilde{u}, \omega) = \begin{cases} -\frac{1}{(2\pi)^p} \int_{\mathbb{T}^p} \frac{\partial f}{\partial u} (u^N(\theta), \theta) \varphi_l(\theta) \overline{\varphi_k(\theta)} d\theta & , k \neq l \\ \left(\sum_{j=1}^m \Omega_j k_j + \sum_{j=m+1}^p \omega_j k_j \right) iI_n - \frac{1}{(2\pi)^p} \int_{\mathbb{T}^p} \frac{\partial f}{\partial u} (u^N(\theta), \theta) d\theta & , k = l, \end{cases}$$

ergeben und die $p - m$ Vektoren

$$\frac{\partial}{\partial \omega_j} G_k(\tilde{u}, \omega) = ik_j u_k, \quad j = m + 1, \dots, p$$

lauten. Führen wir nun die Fourier-Transformationen

$$\frac{\partial f}{\partial u} (u(\theta), \theta) =: A(\theta) \cong \sum_{\|l\| \leq M} A_l \varphi_l(\theta) \quad \text{mit} \quad A_l := \frac{1}{(2\pi)^p} \int_{\mathbb{T}^p} A(\theta) \overline{\varphi_l(\theta)} d\theta \tag{7.120}$$

$$-\frac{\partial u}{\partial \theta_j}(\theta) =: h_j(\theta) \cong \sum_{\|l\| \leq M} h_l^j \varphi_l(\theta) \quad \text{mit} \quad h_l^j := \frac{1}{(2\pi)^p} \int_{\mathbb{T}^p} h_j(\theta) \overline{\varphi_l(\theta)} d\theta$$

mit $j = m+1, \dots, p$ ein und beachten, dass $ik_j u_k$ der k -te Fourier-Koeffizient von $\frac{\partial u}{\partial \theta_j}(\theta)$ ist, so erhalten wir

$$\frac{\partial}{\partial u_l} G_k(\tilde{u}, \omega) = \begin{cases} -A_{k-l} & , k \neq l \\ \left(\sum_{j=1}^m \Omega_j k_j + \sum_{j=m+1}^p \omega_j k_j \right) iI_n - A_0 & , k = l \end{cases}$$

und

$$\frac{\partial}{\partial \omega_j} G_k(\tilde{u}, \omega) = -h_k^j.$$

Schließlich setzen wir zur Berechnung der rechten Seite des Systems (7.119) die Funktion

$$b(\theta) := f(u(\theta), \theta) - \sum_{j=1}^m \Omega_j \frac{\partial u}{\partial \theta_j}(\theta) - \sum_{j=m+1}^p \omega_j \frac{\partial u}{\partial \theta_j}(\theta) \quad (7.121)$$

mit der zugehörigen Fourier-Transformation

$$b(\theta) \cong \sum_{\|l\| \leq M} b_l \varphi_l(\theta) \quad \text{mit} \quad b_l := \frac{1}{(2\pi)^p} \int_{\mathbb{T}^p} b(\theta) \overline{\varphi_l(\theta)} d\theta \quad (7.122)$$

ALGORITHMUS 7.33 (Fourier-Galerkin-Verfahren für Tori)

Function $[u, \omega] = \text{Fourier_Galerkin}(f, \text{tol}, \nu_{\max})$

1. (Startnäherung) Wähle $u^{(0)} \in \mathbb{C}^{nm}$ und $\omega^{(0)} \in \mathbb{R}_+^{p-m}$.
2. Für $\nu = 0, 1, \dots, \nu_{\max}$ iteriere:
 - 2.1. Berechne $A(\theta), h_j(\theta)$ nach (7.120) und $b(\theta)$ nach (7.121).
 - 2.2. (FFT) Berechne die zugehörigen Fourier-Koeffizienten A_k, b_k, h_k^j für $\|k\| \leq N$ und $j = m+1, \dots, p$ mittels p -dimensionaler schneller Fourier-Transformation.
 - 2.3. Löse das lineare Gleichungssystem (7.119) nach z und ξ auf.
 - 2.4. Bilde $u^{(\nu+1)} = u^{(\nu)} + z$ und $\omega^{(\nu+1)} = \omega^{(\nu)} + \xi$.
 - 2.5. Ist $\|z\|_{\mathcal{H}^0}^2 + \|\xi\|_{\mathbb{C}^{p-m}}^2 < \text{tol}^2$, so setze $u = u^{(\nu+1)}, \omega = \omega^{(\nu+1)}$ und gehe zu Schritt 3.
3. Return u, ω

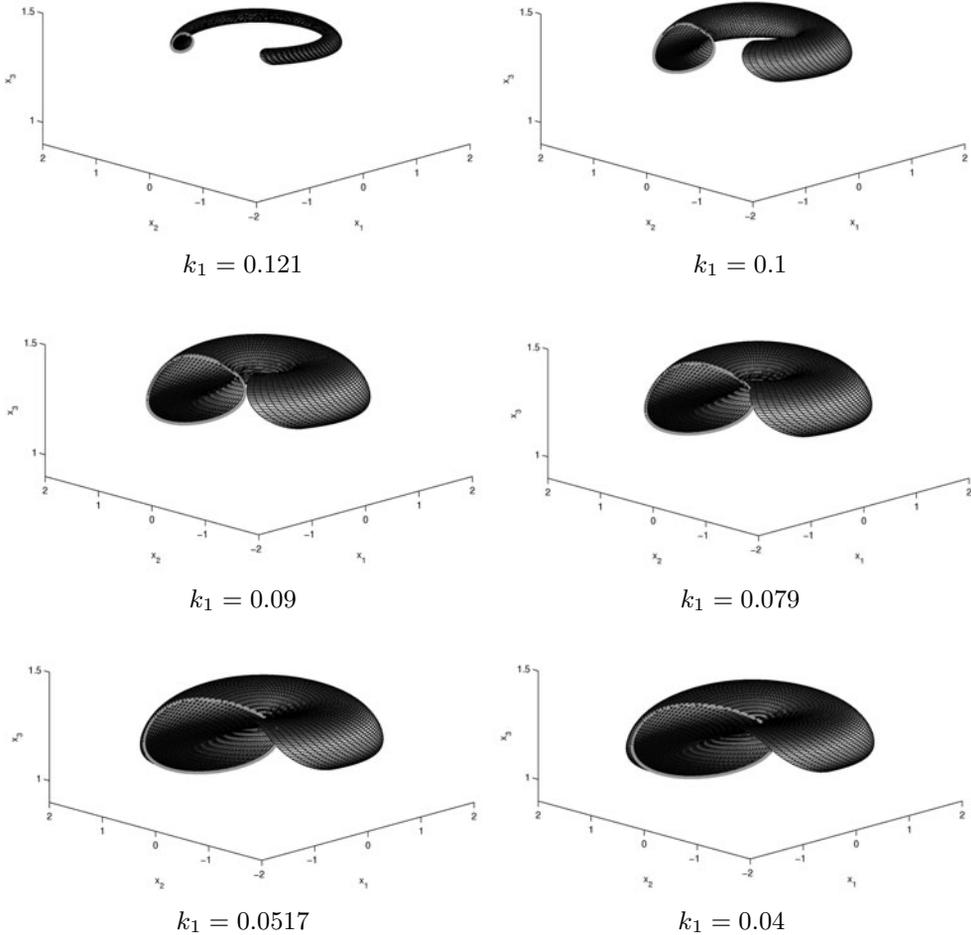


Abb. 7.22 Der einfache invariante Torus des Systems (7.124) für sechs Parameterwerte. Die Darstellung erfolgt in x_1 - x_2 - x_3 . Der Torusschnitt erfolgt stets für $\theta_1 = 0$

an und beachten, dass $b_k = -G_k(\tilde{u}, \omega)$ gilt. Damit gewinnen wir in jedem Newton-Schritt die explizite Darstellung des unterbestimmten linearen Gleichungssystems (7.119) für alle Multiindizes $\|k\| \leq N$ in der Form

$$\left(\sum_{j=1}^m \Omega_j k_j + \sum_{j=m+1}^p \omega_j k_j \right) iz_k - \sum_{\|l\| \leq N} A_{k-l} z_l - \sum_{j=m+1}^p \xi_j h_k^j = b_k \quad (7.123)$$

mit den $m = \prod_{j=1}^p (2N_j + 1)$ unbekanntenen Vektoren $z_l \in \mathbb{C}^n$ und den $p - m$ reellen Unbekannten ξ_j . Eine Lösung minimaler Norm dieses Systems kann nunmehr mit Standard-Software für unterbestimmte lineare Gleichungssysteme, z. B. mit MATLAB ermittelt werden. Eine empfehlenswerte Darstellung zu diesem Thema findet der interessierte Leser bei Gramlich und Werner (2000).

Bemerkung 7.34

1. Algorithmus 7.33 stellt das Newton-Verfahren (7.119) lediglich in seiner Grundform vor. Zu seiner praktischen Realisierung sollten allerdings die in Kapitel 5.1 eingeführten Newton-ähnlichen Verfahren implementiert werden.
2. Bei Vorliegen *parameterabhängiger Systeme* empfiehlt sich der Einbau des Newton-Verfahrens aus Algorithmus 7.33 als Korrektorverfahren in ein leistungsfähiges Prädiktor-Korrektor-Fortsetzungsverfahren (vgl. Kapitel 5.2). Damit stehen die in der Regel schwer zu beschaffenden Startlösungen $u^{(0)}$ und $\omega^{(0)}$ in jedem Fortsetzungsschritt – mit Ausnahme des Anfangsschrittes – stets zur Verfügung. ■

Beispiel 7.35

Wir greifen das System (7.105) des nichtlinearen elektrischen Resonanzkreises von Ch. Hayashi

$$\begin{aligned}
 \dot{x}_1 &= x_2 \\
 \dot{x}_2 &= -k_1 x_2 - \frac{1}{8}(x_1^2 + 3x_3^2)x_1 + B \cos t \\
 \dot{x}_3 &= -\frac{1}{8}k_2(3x_1^2 + x_3^2)x_3 + B_0
 \end{aligned}
 \tag{7.124}$$

mit den Parameterwerten $B_0 = 0.03$, $B = 0.22$, $k_2 = 0.05$ und dem Kontrollparameter $k_1 \in [0.043, 0.09]$ auf und wenden darauf die vorgestellte Spektralmethode an. Mit dem von Schreiber (2004) an der TU Ilmenau entwickelten Analyseprogramm `speccont` wird eine Parameterfortsetzung des einfachen Torus für $k_1 \in [0.04, 0.121]$ und des nach Tabelle 7.1 erhaltenen gedoppelten Torus für $k_1 \in [0.04, 0.077]$ durchgeführt.

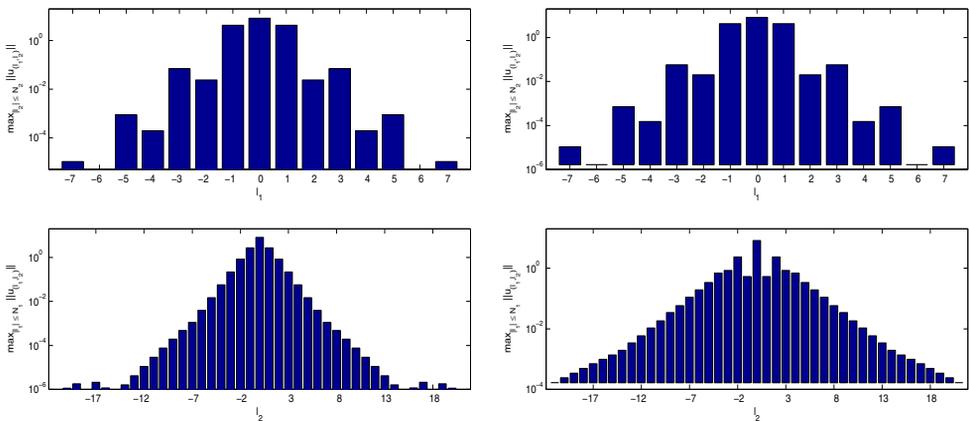


Abb. 7.23 System (7.124): Normmaximale Fourier-Koeffizienten bzgl. θ_1 (oben) und θ_2 (unten) des einfachen Torus (links) und des gedoppelten Torus (rechts) am Parameterwert $k_1 = 0.07$

Mit dem Fourier-Galerkin-Verfahren ist die Fortsetzung der stabilen Tori bezüglich des Parameters k_1 in Bereiche mit sich ändernder Stabilitätseigenschaft problemlos möglich,

da das Fourier-Galerkin-Verfahren keine Stabilitätsvoraussetzungen erfordert. Als Dis-

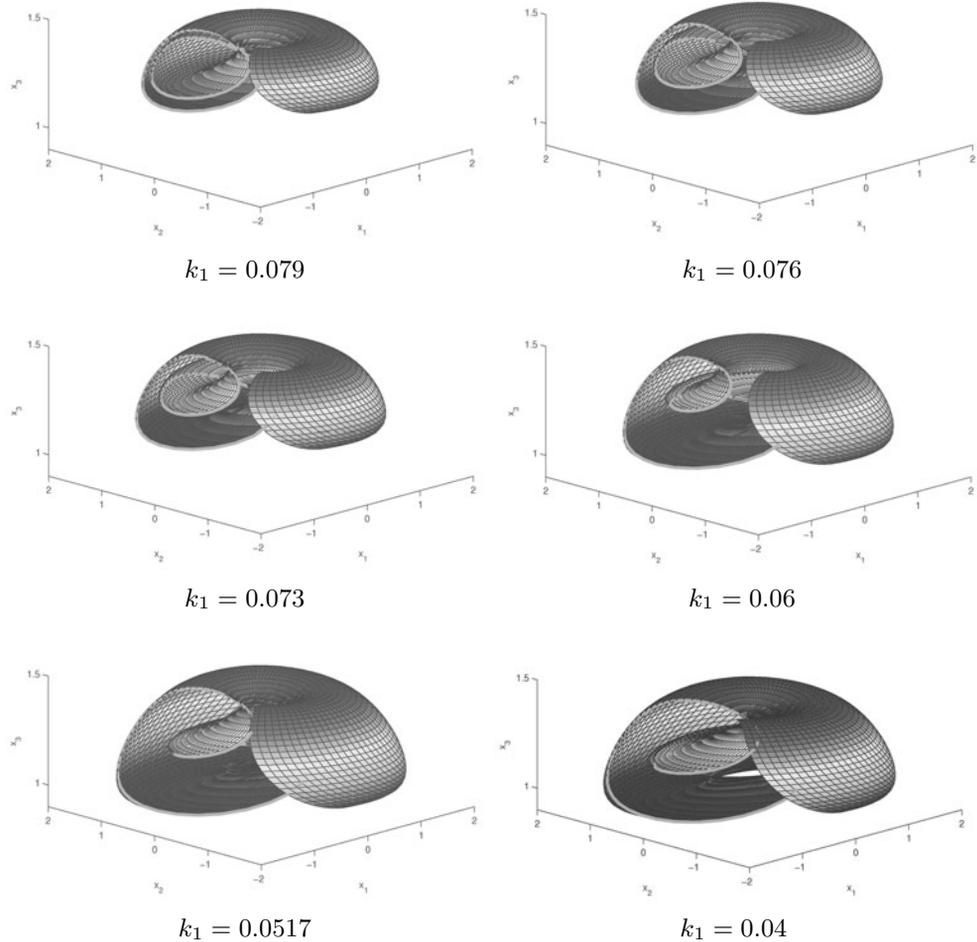


Abb. 7.24 Der gedoppelte invariante Torus des Systems (7.124) zu ausgewählten Parameterwerten sowie der Torusschnitt für $\theta_1 = 0$

cretisierungparameter dienen stets $N_1 = 7$ und $N_2 = 21$. Die Abbildungen 7.22 und 7.24 zeigen einige ausgewählte Tori zu verschiedenen Parameterwerten. Die exponentielle Konvergenz des Spektralansatzes spiegelt sich im exponentiellen Abfall der Normen der Fourier-Koeffizienten wider. In Abbildung 7.23 ist dieser Sachverhalt in logarithmischer Darstellung für den Parameterwert $k_1 = 0.07$ gut zu erkennen. Die Koeffizienten bezüglich θ_1 verhalten sich für beide Tori ähnlich. Das rapide Abfallen der Normen der Koeffizienten lässt sich auch in unmittelbarer Nähe von Verzweigungspunkten beobachten, so dass selbst dort die exponentielle Konvergenz erhalten bleibt. ■

Insgesamt erscheint die *Volldiskretisierung* mit Fourier-Galerkin-Ansatz besonders geeignet, wenn die Lösung hinreichend glatt ist. Dann lässt sich der gesuchte Torus bereits mit wenigen Termen der Fourier-Reihe gut approximieren und die im Verfahren entstehenden Gleichungssysteme sind von akzeptabler Größe. Im Gegensatz zur *Semidiskretisierung* erfordert sie zudem keine analytischen Manipulationen des Ausgangssystems mit Computeralgebrasystemen, deren Grenzen recht schnell erreicht sind. Die numerische schnelle Fourier-Transformation ist hingegen problemlos möglich.

Auf die theoretisch komplizierte Konvergenzanalyse – besonders im Falle freier Phasen mit $p - m$ freien Basisfrequenzen – kann hier nicht eingegangen werden. Weitere Untersuchungen sind ebenfalls bei der Lösungsfortsetzung parameterabhängiger Systeme zum Verhalten des Verfahrens in unmittelbarer Nähe von Resonanzen nötig. Zu diesen und weiteren interessanten Fragen bei der numerischen Behandlung der Bifurkationen invarianter Tori verweisen wir den Leser auf die grundlegenden Beiträge in Doedel und Tuckerman (2000) und Krauskopf et al. (2007) sowie auf den Beitrag Vogt et al. (2005), die allerdings über den Rahmen dieses einführenden Buches weit hinausgehen.

7.4 Aufgaben

Aufgabe 7.1

Untersuchen Sie, welche der folgenden Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}^n$ quasi-periodisch sind:

a) $f(t) = a \cdot \exp(\cos t - \cos(\sqrt{5} \cdot t) + 1) - \cos(\sqrt{10} \cdot t)$ für $a \in \mathbb{R}$,

b) $f(t) = \cos(\exp(1) \cdot t) - \sin(\pi \cdot t)$

c) $f(t) = \cos(2.71828 \cdot t) - \sin(3.14159 \cdot t)$

d) $f(t) = \cos(2.718281828 \cdot t) - \sin(3.141592654 \cdot t)$

e) $f(t) = \begin{pmatrix} \sin(2t) - \cos(\sqrt{2} \cdot t) \\ \sin^2(\sqrt{8} \cdot t) + \cos^2(\sqrt{8} \cdot t) \\ \cos(\sqrt{2} \cdot t) + \sin(2t) \end{pmatrix}$

f) $f(t) = \begin{pmatrix} \cos(\sqrt{8} \cdot t) \cdot \sin t \\ \exp(\cos(\pi \cdot t)) + \cot(\pi\sqrt{2} \cdot t) \\ \tan(\sqrt{2} \cdot t) \cdot \cos t \end{pmatrix}$

g) $f(t) = 1.3 \sin(23.47055t + 0.3 \cos(6\pi t))$.

Geben Sie jeweils die Basisfrequenzen w_i und deren Anzahl an. Ist die erhaltene Frequenzbasis maximal? Bestimmen Sie bei Vorliegen periodischer Funktionen deren (minimale) Periodendauer T .

Aufgabe 7.2

Gegeben sei das System zweier linear gekoppelter Oszillatoren, modelliert durch

$$\begin{aligned}\dot{x}_1 &= \alpha_1 x_1 + \beta_1 y_1 - (x_1^2 + y_1^2)x_1 - \delta(x_1 + y_1 - x_2 - y_2) \\ \dot{y}_1 &= -\beta_1 x_1 + \alpha_1 y_1 - (x_1^2 + y_1^2)y_1 - \delta(x_1 + y_1 - x_2 - y_2) \\ \dot{x}_2 &= \alpha_2 x_2 + \beta_2 y_2 - (x_2^2 + y_2^2)x_2 + \delta(x_1 + y_1 - x_2 - y_2) \\ \dot{y}_2 &= -\beta_2 x_2 + \alpha_2 y_2 - (x_2^2 + y_2^2)y_2 + \delta(x_1 + y_1 - x_2 - y_2)\end{aligned}$$

mit Parameterwerten $\alpha_1 = \alpha_2 = 1.0$, $\beta_1 = 0.55$, $\beta_2 > 0$ und einem reellen Kopplungsparameter $\delta \geq 0$.

- Bestimmen Sie im entkoppelten Fall ($\delta = 0$) periodische und quasi-periodische Orbits des Systems.
- Führen Sie Polarkoordinaten $x_i = u_i \cos \theta_i$, $y_i = -u_i \sin \theta_i$, $i = 1, 2$, ein und transformieren Sie bei beliebigen Parametern die DGL in die partitionierte Form (7.21). Wie lautet die entsprechende partielle DGL (7.35)?
- Entwickeln Sie ein numerisches Verfahren zur Lösung der PDGL (7.35) auf dem Torus \mathbb{T}^2 mittels des expliziten Upwind-Verfahrens (7.64).
- Erproben Sie Ihr Verfahren mit den Parameterwerten (i) $\beta_2 = 0.55$ und (ii) $\beta_2 = 1.578$ mit den Werten $\delta = 0.01, 0.1, 0.2, 0.25, 0.26$ des Kopplungsparameters δ . Vergleichen Sie das Verhalten des Verfahrens in den beiden Fällen (i) und (ii).

Aufgabe 7.3

Gegeben ist die DGL (7.71) von Philippow und Büntig (1992)

$$\ddot{x} + \alpha \dot{x}^3 - \beta \dot{x} + (1 + B \sin 2t)x = 0 \quad (7.125)$$

mit Amplitude $B = 0.1$, den Werten $\alpha = \varepsilon - B$, $\beta = \varepsilon/2 - B$ und dem reellen Parameter $\varepsilon > 0$, die ein parametrisch erregtes elektrisches Netzwerk mit subharmonischer Systemantwort beschreibt.

- Überführen Sie die DGL in ein System 1. Ordnung und danach mittels Polarkoordinaten in die partitionierte Form (7.21). Geben Sie die entsprechende PDGL (7.35) an.
- Lösen Sie die PDGL (7.35) auf dem Torus \mathbb{T}^2 mittels des expliziten Upwind-Verfahrens (7.64) für Parameterwerte $\varepsilon = 1.5, 1.6, 3.0, 6.0, 9.0$. Stellen Sie die Torusquerschnitte wie in Abbildung 7.15 graphisch dar.
- Wie verhält sich das Verfahren in der Nähe von Resonanzen bei $\varepsilon = 1.344$, $\varepsilon = 5.493$ und $\varepsilon = 7.054$?

Aufgabe 7.4

Die DGL (7.125) aus Aufgabe 7.3 mit den Parameterwerten $B = 0.1$, $\alpha = \varepsilon - B$, $\beta = \varepsilon/2 - B$ und $\varepsilon > 0$ soll mit dem Fourier-Galerkin-Verfahren gelöst werden.

- a) Notieren Sie die vereinfachte Invarianzgleichung der Form (7.94) mit $n = 2$ für quasi-periodische 2-Tori. Entwickeln Sie auf der Basis von Algorithmus 7.33 ein Fourier-Galerkin-Verfahren zur Lösung dieser Invarianzgleichung für 2-Tori periodisch erregter Gleichungen, d. h. mit $m = 1$ und $p = 2$.
- b) Erproben Sie das Verfahren am parametrisch erregten Netzwerk für Parameterwerte $\varepsilon = 1.5, 1.6, 3.0, 6.0, 9.0$ und in der Nähe von Resonanzen bei $\varepsilon = 1.344$, $\varepsilon = 5.493$ und $\varepsilon = 7.054$.
- c) Vergleichen Sie die Resultate, die mit unterschiedlichen Diskretisierungsparametern N_1 und N_2 erhalten wurden. Stellen Sie die Beträge der Fourier-Koeffizienten wie in Abbildung 7.23 grafisch dar und bewerten Sie die erhaltenen Approximationen.

Aufgabe 7.5

Entwickeln Sie mit einem Computeralgebra-System, z. B. MAPLE oder MATHEMATICA, ein Verfahren zur Generierung des Spektralsystems (7.102)

$$Y'(\tau) = H(Y(\tau)) - Y(\tau)\Phi$$

bei Semidiskretisierung im Falle $m = 1$ und $p = 2$. Wenden Sie das Verfahren (a) auf das Modell (7.105) des elektrischen Resonanzkreises und (b) auf das parametrisch erregte elektrische Netzwerk (7.125) an.

Literaturverzeichnis

- Abraham R. H. und Shaw C. D. (1989) *Dynamics - The Geometry of Behavior, Part One: Periodic Behavior*. 4th printing, Aerial Press, Santa Cruz, California.
- Allgower E. L. und Georg K. (1990) *Numerical Continuation Methods*. Springer-Verlag, Berlin.
- Alt H. W. (1992) *Lineare Funktionalanalysis*. 2. Auflage, Springer-Verlag, Berlin.
- Amann H. (1983) *Gewöhnliche Differentialgleichungen*. De Gruyter, Berlin, New York.
- Ambrosetti A. und Prodi G. (1993) *A Primer of Nonlinear Analysis*. Cambridge Studies in Advanced Mathematics 34. Cambridge University Press.
- Anishchenko V. S. (1987) *Dynamical Chaos – Basic Concepts*. Teubner-Texte zur Physik, B.G. Teubner Verlag, Leipzig.
- Appell J. und Văth M. (2005) *Elemente der Funktionalanalysis. Vektorräume, Operatoren und Fixpunktsätze*. Vieweg & Sohn, Wiesbaden.
- Argyris J., Faust G. und Haase M. (1995) *Die Erforschung des Chaos*. Vieweg Verlag, Braunschweig und Wiesbaden.
- Ascher U. M., Mattheij R. M. und Russell R. D. (1995) *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*. SIAM Publications, Philadelphia.
- Ascher U. M. und Petzold L. R. (1998) *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations*. SIAM Publications, Philadelphia.
- Aulbach B. (1997) *Gewöhnliche Differentialgleichungen*. Spektrum Verlag, Heidelberg.
- Bernet K. (1995) *Ein Beitrag zur numerischen Approximation und Verfolgung von Toruslösungen parameterabhängiger nichtlinearer dynamischer Systeme*. Dissertation, TU Ilmenau, Shaker Verlag, 105 S.
- Bernet K. und Vogt W. (1994) *Anwendung finiter Differenzenverfahren zur direkten Bestimmung invarianter Tori*. ZAMM 74, No.6 ,T 577 - T 579.
- Beyn W. J. (1991) *Numerical Methods for Dynamical Systems*. In: Light W. (Hrsg.), *Advances in Numerical Analysis*, Volume I. Clarendon Press, Oxford.
- Bogoljubov N. und Mitropolski J. (1965) *Asymptotische Methoden in der Theorie der nichtlinearen Schwingungen*. Akademie-Verlag, Berlin.
- Broer H. W., Huitema G. B. und Sevryuk M. B. (1996) *Quasi-Periodic Motions in Families of Dynamical Systems*. Springer-Verlag, Berlin.
- Büntig W. G. und Vogt W. (1998) *Non-linear Basic Circuits for Modelling Energetic Systems*. In: *Non-Linear Electromagnetic Systems*. ISO Press, Amsterdam, S.531-534.
- Büntig W. G. und Vogt W. (2003) *Numerical Bifurcation Analysis of Nonlinear Power Systems*. Proceedings of 48th Int. Wiss. Kolloquium, 22.-25.09.2003, TU Ilmenau, S. 373-374.
- Chen Y. und Leung A. Y. T. (1998) *Bifurcation and Chaos in Engineering*. Springer-Verlag, London.
- Choquet-Bruhat Y., DeWitt-Morette C. und Dillard-Bleick M. (1983,1988) *Analysis, Manifolds and Physics*. Vol. 1,2. North-Holland, Amsterdam.

- Chow S.-N. und Hale J. K. (1982) *Methods of Bifurcation Theory*. Grundlehren der Math. Wiss. 251, Springer-Verlag, New York.
- Coddington E. A. und Levinson N. A. (1955) *Theory of Ordinary Differential Equations*. McGraw-Hill, New York.
- Courant R. und Hilbert D. (1968) *Methoden der mathematischen Physik II*. 2. Auflage, Springer-Verlag, Berlin.
- Deimling K. (1974) *Nichtlineare Gleichungen und Abbildungsgrade*. Springer-Verlag, Berlin.
- Deuffhard P. und Hohmann A. (1993) *Numerische Mathematik I*. 2. Auflage, W. de Gruyter, Berlin.
- Deuffhard P. (2004) *Newton Methods for Nonlinear Problems*. Springer-Verlag, Berlin.
- Dieci L., Lorenz J. und Russell R. D. (1991) *Numerical Calculation of Invariant Tori*. SIAM Journal on Sci. Stat. Comput., Vol. 12, S.607-647.
- Dieci L. und Lorenz J. (1995) *Computation of Invariant Tori by the Method of Characteristics*. SIAM Journal on Numerical Analysis, Vol. 32, No.5, S.1436-1474.
- Dieci L. und Bader G. (1992) *On Approximating Invariant Tori: Block Iterations and Multigrid Methods for the Associated Systems*. Preprint Nr. 658, Universität Heidelberg.
- Doedel E. J., Paffenroth R. C., Champneys A. R., Fairgrieve T. F., Kuznetsov Y. A., Oldeman B. E., Sandstede B. und Wang X. (2002) *AUTO 2000 - Continuation and bifurcation software for ordinary differential equations (with HomCont), users guide*. Concordia University, Montreal, P.Q., Canada.
- Doedel E. J., Govaerts W. und Kuznetsov Y. A. (2003) *Computation of Periodic Solution Bifurcations in ODEs Using Bordered Systems*. SIAM Journal on Numerical Analysis, Vol. 41, S.401-435.
- Doedel E. J. und Tuckerman L. S. (Hrsg.) (2000) *Numerical Methods for Bifurcation Problems and Large-scale Dynamical Systems*. Springer-Verlag, New York, 471 S.
- Drábek P. und Milota J. (2007) *Methods of Nonlinear Analysis*. Applications to Differential Equations, Birkhäuser, Basel.
- Dunford N. und Schwartz J. T. (1964) *Linear Operators I,II*. Interscience Publisher, New York.
- Engert S. (2009) *Vergleich numerischer Verfahren zur Berechnung des LCE-Spektrums parameterabhängiger zeitkontinuierlicher dynamischer Systeme*. Diplomarbeit, TU Ilmenau, Fakultät für Mathematik und Naturwissenschaften.
- Ernst M. (2004) *Numerische Bifurkationsanalyse nichtlinearer energetischer Systeme*. Diplomarbeit, TU Ilmenau, Fakultät für Mathematik und Naturwissenschaften.
- Fischer W. und Lieb I. (2003) *Funktionentheorie*. 8. Auflage, Vieweg, Braunschweig.
- Göpfert A. und Riedrich T. (1994) *Funktionalanalysis*. 4. Auflage, B.G. Teubner Verlagsgesellschaft, Stuttgart.
- Gramlich G. und Werner W. (2000) *Numerische Mathematik mit MATLAB*. dpunkt Verlag, Heidelberg.
- Guckenheimer J. und Holmes P. (1983) *Nonlinear Oscillations, Dynamical Systems and Bifurcation of Vector Fields*. Applied Mathematical Sciences 42. Springer-Verlag, New York.
- Hale J. K. (1963) *Oscillations in Nonlinear Systems*. Dover Publications, New York.

- Hartmann P. (1960) *A lemma in the theory of structural stability of differential equations*. Proc. Amer. Math. Soc. 11, 610–620.
- Hartman P. (1964) *Ordinary Differential Equations*. John Wiley and Sons, New York.
- Hayashi Ch. (1985) *Nonlinear Oscillations in Physical Systems*. Princeton University Press, Princeton, New Jersey.
- Henderson M. E. (2007) *Higher-Dimensional Continuation*. In: Krauskopf B., Osinga H.M. und Galán-Vioque J. (Hrsg.) *Numerical Continuation Methods for Dynamical Systems*. Springer-Verlag Netherlands.
- Hermann M. (2004) *Numerik gewöhnlicher Differentialgleichungen*. Oldenbourg Verlag, München.
- Heuser H. (1980) *Lehrbuch der Analysis*. Teil 1 und Teil 2, Teubner, Stuttgart.
- Heuser H. (2006) *Funktionalanalysis*. Teubner, Stuttgart.
- Hirsch M. W. und Smale S. (1974) *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press, New York - London, 1974.
- Hoffmann H. (1989) *Berechnung von periodischen Lösungen und ihrer charakteristischen Multiplikatoren im nichtautonomen und autonomen Fall*. Diplomarbeit, TU Ilmenau, Fakultät für Mathematik und Naturwissenschaften.
- Hoffmann A., Marx B. und Vogt W. (2005) *Mathematik für Ingenieure. Band 1: Lineare Algebra, Analysis – Theorie und Numerik*. Pearson Studium, München.
- Hoffmann A., Marx B. und Vogt W. (2006) *Mathematik für Ingenieure. Band 2: Vektoranalysis, Integraltransformationen, Differenzialgleichungen, Stochastik – Theorie und Numerik*. Pearson Studium, München.
- Hopf E. (1942) *Abzweigung einer periodischen Lösung von einer stationären Lösung eines Differentialsystems*. Ber. Math. Phys. Kl. Sächs. Akad. Wiss., Leipzig 94, 3-22.
- Iooss G. und Joseph D. D. (1990) *Elementary Stability and Bifurcation Theory*. 2nd ed., Springer-Verlag, Berlin.
- Isaacson E. und Keller H. B. (1972) *Analyse numerischer Verfahren*. Verlag Harry Deutsch, Frankfurt.
- Kaas-Petersen Ch. (1985) *Computation of Quasi-periodic Solutions of Forced Dissipative Systems*. Journal of Comp. Physics, Vol.58, No.3, S.395-408.
- Kantorowitsch L. W. und Akilow G. P. (1978) *Funktionalanalysis in normierten Räumen*. Akademie Verlag, Berlin.
- Keller H. B. (1976) *Numerical Solution of Two-Point BVP's*. SIAM Publications, Philadelphia.
- Keller H. B. (1987) *Numerical Methods in Bifurcation Problems*. Springer-Verlag, Berlin.
- Kelley C. T. (1995) *Iterative Methods for Linear and Nonlinear Equations*. SIAM Publications, Philadelphia.
- Kosmol P. (1993) *Methoden zur numerischen Behandlung nichtlinearer Gleichungen und Optimierungsaufgaben*. B.G.Teubner, Stuttgart.
- Krasnoselski M. et al. (1973) *Näherungsverfahren zur Lösung von Operatorgleichungen*. Akademie-Verlag, Berlin.

- Krasnoselski M. und Zabreiko P. (1975) *Geometrical Methods in Nonlinear Analysis*. Springer-Verlag, Berlin.
- Krauskopf B., Osinga H. M. und Galán-Vioque J. (Hrsg.) (2007) *Numerical Continuation Methods for Dynamical Systems*. Springer-Verlag Netherlands.
- Kubiček M. und Marek M. (1983) *Computational Methods in Bifurcation Theory and Dissipative Systems*. Springer-Verlag, New York.
- Kuznetsov Y. A. (1995) *Elements of Applied Bifurcation Theory*. Springer-Verlag, New York.
- Langford W. F. (1984) *Numerical Studies of Torus Bifurcations*. In: International Series of Numerical Mathematics, Vol.70, Birkhäuser Verlag, Basel.
- Lorenz J. (1991) *Computation of Invariant Manifolds*. In: Proceedings of the 14th Dundee Biennial Conference on Numerical Analysis, June 1991, ISBN 0-582-08908-5, Dundee.
- Lynch S. (2001) *Dynamical Systems with Applications using MAPLE*. Birkhäuser, Boston - Basel - Berlin.
- Marsden J. E. und McCracken M. (1976) *The Hopf Bifurcation and Its Applications*. Springer-Verlag, New York.
- Mei Zhen (2000) *Numerical Bifurcation Analysis for Reaction-Diffusion Equations*. Springer-Verlag, Berlin.
- Mingyou H., Küpper T. und Masbaum N. (1997) *Computation of Invariant Tori by the Fourier Methods*. SIAM J. Sci. Comput., 18, S. 918-942.
- Mohr T. (2001) *Modellierung von nichtlinearen Differentialgleichungen zum Aufbau elektronischer Schaltungen bei vorgegebenem Bifurkationsverhalten*. Dissertation. Technische Universität Ilmenau, Ilmenau.
- Moore G. (1996) *Computation and parametrisation of invariant curves and tori*. SIAM J. on Numer. Anal., 33, No.6, S. 2333-2358.
- Parker T. S. und Chua L. O. (1989) *Practical Numerical Algorithms for Chaotic Systems*. Springer-Verlag, New York.
- Peixoto M. (1962) *Structural stability on two-dimensional manifolds*. Topology, 1, S.101-120.
- Perko L. (1996) *Differential Equations and Dynamical Systems*. Springer-Verlag, New York.
- Peterseim D. (2004) *Numerische Analyse parameterabhängiger periodischer Orbits nichtlinearer dynamischer Systeme mittels Mehrzielmethode und effizienter Fortsetzungstechniken*. Diplomarbeit, TU Ilmenau, Fakultät für Mathematik und Naturwissenschaften.
- Philippow E. (1971) *Nichtlineare Elektrotechnik*. Geest & Portig, Leipzig.
- Philippow E. S. und Büntig W. G. (1992) *Analyse nichtlinearer dynamischer Systeme der Elektrotechnik*. Carl Hanser Verlag, München.
- Philippow E. S., Büntig W. G. und Vogt W. (1992) *Die Entstehung quasiperiodischer Lösungen in nichtlinearen Systemen der Elektrotechnik*. 37.Int.Wiss.Kolloquium, TU Ilmenau, Band 2, S.151-156.
- Pontrjagin L. S. (1965) *Gewöhnliche Differentialgleichungen*. Deutscher Verlag der Wissenschaften, Berlin.
- Reddy B. Daya (1998) *Introductory Functional Analysis. With Applications to Boundary Value Problems and Finite Elements*, Springer-Verlag, New York.

- Reitmann V. (1996) *Reguläre und chaotische Dynamik*. B.G. Teubner Verlag, Leipzig.
- Rheinboldt W. C. (1994) *Methods for Solving Systems of Nonlinear Equations*. 4th ed., SIAM Publications, Philadelphia.
- Riesz F. und Sz-Nagy B. (1956) *Vorlesungen über Funktionalanalysis*. Verlag der Wissenschaften, Berlin. (English edition: *Functional Analysis*.) Frederick Ungar, New York.
- Robbin J. W. (1968) *On the existence theorem for differential equations*. Proc. Amer. Math. Soc. 19, S.1005-1006.
- Ruzicka M. (2004) *Nichtlineare Funktionalanalysis*. Eine Einführung, Springer-Verlag, Berlin.
- Samoilenko A. M. (1991) *Elements of the Mathematical Theory of Multi-Frequency Oscillations*. Kluwer Academic Publishers, Dordrecht.
- Sanders J. A. und Verhulst F. *Averaging Methods in Nonlinear Dynamical Systems*. Springer-Verlag, New York 1985.
- Sauvigny F. (2004) *Partielle Differentialgleichungen der Geometrie und der Physik*. Grundlagen und Integraldarstellungen, Springer-Verlag, Berlin.
- Schilder F. (2004) *Numerische Approximation quasiperiodischer invarianter Tori unter Anwendung erweiterter Systeme*. Dissertation, TU Ilmenau, Inst. f. Math., 136 S.
- Schilder F. und Vogt W. (2003) *A Generalized Fourier Method for Quasi-periodic Oscillations in Nonlinear Circuits*. Proceedings of 48th Int. Wiss. Kolloquium, 22.-25.09.2003, TU Ilmenau, S.369-370.
- Schilder F. und Vogt W. (2003) *Semidiscretisation Methods for Quasi-periodic Solutions*. PAMM Proceedings in Applied Mathematics and Mechanics, Vol. 2, S. 497 - 498.
- Schreiber S. (2004) *Spektralmethoden und Pseudospektralmethoden zur numerischen Approximation und Verfolgung invarianter 2-Tori*. Diplomarbeit, TU Ilmenau, Fakultät für Mathematik und Naturwissenschaften.
- Schwetlick H. (1979) *Numerische Lösung nichtlinearer Gleichungen*. Deutscher Verlag der Wissenschaften, Berlin.
- Seydal R. (1991) *Practical Bifurcation and Stability Analysis*. Springer-Verlag, New York.
- Van Veldhuizen M. (1987) *A new algorithm for the numerical approximation of an invariant curve*. SIAM J. Sci. Stat. Comput. Vol.8, S.951-962.
- Verhulst F. (2000) *Nonlinear Differential Equations and Dynamical Systems*. Sec. Edition. Springer-Verlag, Berlin.
- Vogt W. (2003) *Numerical Simulation of Nonlinear Circuits - Periodic and Quasi-periodic Responses*. Berichte des IZWR, Jenaer Schriften zur Mathematik und Informatik, Vol. 1, S.73 - 87.
- Vogt W. und Büntig W. G. (1992) *Anwendungsorientierte neuere Verfahren zur Bestimmung und Analyse quasiperiodischer Lösungen*. 37.Int.Wiss.Kolloquium, TU Ilmenau, Band 2, S.145-150.
- Vogt W. und Büntig W. G. (1998) *Nonlinear Circuit Models for the Analysis of Energetic Systems*. Proceedings of the 43rd Internat. Sci. Colloquium, TU Ilmenau, Sept. 21-24,1998, S.138-143.
- Vogt W. und Büntig W. G. (1998) *Non-linear Basic Circuits for Modelling Energetic Systems*. In: Non-Linear Electromagnetic Systems. ISO Press Amsterdam, S.531-534.

- Vogt W. und Büntig W. (2003) *Numerical Bifurcation Analysis of Nonlinear Power Systems*. Proceedings of the 48th Int. Wiss. Kolloquium, TU Ilmenau, 22.-25.09.2003, S. 373-374.
- Vogt W., Schilder F. und Osinga H. M. (2005) *Continuation of Quasi-periodic Invariant Tori*. SIAM Journal on Applied Dynamical Systems, Vol.4, No.3, S.459-488.
- Vogt W., Schilder F., Schreiber S. und Osinga H. M. (2006) *Fourier Methods for Quasi-periodic Oscillations*. International Journal for Numerical Methods in Engineering, Vol. 67, S.629-671.
- Wainberg M. M. und Trenogin W. A. (1973) *Theorie der Lösungsverzweigung bei nichtlinearen Gleichungen*. Akademie-Verlag, Berlin.
- Walter W. (1995) *Analysis 1 und 2*. Springer-Verlag, Berlin.
- Walter W. (1983) *Gewöhnliche Differentialgleichungen*. Springer-Lehrbuch, Berlin.
- Werner D. (1997) *Funktionalanalysis*. Springer-Verlag, Berlin.
- Wiggins S. (1990) *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. Springer-Verlag, New York.
- Yoshinga T. und Kawakami H. (1995) *Bifurcations and Chaotic States in Forced Oscillatory Circuits containing Saturable Inductors*. In: Carroll T. und Pecora L., *Nonlinear Dynamics In Circuits*. World Scientific Publishing.
- Zeidler E. (1995) *Applied Functional Analysis*. Applications to Mathematical Physics, AMS 108, Springer-Verlag, Berlin.
- Zeidler E. (1995) *Applied Functional Analysis*. Main Principles and their Applications, AMS 109, Springer-Verlag, Berlin.
- Zeidler E. (1986) *Nonlinear Functional Analysis and its Applications*. Teil I, Springer-Verlag, Berlin.
- Zeidler E. (1988) *Nonlinear Functional Analysis and its Applications*. Teil IV, Springer-Verlag, Berlin.

Symbolverzeichnis

Die Zahlen geben die Nummern von Seiten an, auf denen das Symbol erscheint.

$R(T)$	Wertebereich der Abbildung T	18
$\text{dom}(T)$	Definitionsbereich der Abbildung T	18
$\ker(T), N(T)$	Kern bzw. Nullraum von T	18
$\text{rang}(T)$	Rang von T ; $\text{rang}(T) := \dim R(T)$	22
$D_x f(x, \lambda)$	Ableitung von f nach x	244
$Df(y)$	Ableitung von f nach y	244
$F_y(x_0, y_0)$	Partielle Ableitung von $F(x, y)$ nach y an der Stelle (x_0, y_0)	60
$f'(x)$	Ableitung der Funktion f im Punkt x	55
$A \odot B$	Bialternierendes Matrixprodukt von A und B	281
δ_{ij}	Kronecker-Symbol	28
$\text{diag}(a_1, \dots, a_n)$	Diagonalmatrix aus den Elementen a_1, \dots, a_n	407
$\dim X$	Dimension des Raumes X	254
$\text{int}(D)$	Menge der inneren Punkte von D	231
$\mathbb{K} = \mathbb{R}, \mathbb{C}$	Körper der reellen bzw. komplexen Zahlen	8
X^n	n -faches kartesisches Produkt der Menge X	403
$A \times B$	Kartesisches Produkt der Mengen A und B	292
$K(a; r)$	Offene Kugel mit Mittelpunkt a und Radius r	4
$K[a; r]$	Abgeschlossene Kugel mit Mittelpunkt a und Radius r	4
$S(a; r)$	Sphäre mit Mittelpunkt a und Radius r	4
e^{tA}	Matrix-Exponentialfunktion	89
$\ x\ $	Norm eines Elementes x	7
$\ x\ _p$	p -Norm; $p = 2$: Euklidische Norm	8
$N_y \mathcal{M}$	Normalraum von \mathcal{M} bei y	265
T^*	Adjungierter Operator zu T	30
X^*	Dualraum von X ; $X^* = \mathcal{L}(X, \mathbb{K})$	29
$\text{ind}(T)$	Index des Operators T	38
$\mathcal{K}(X, Y)$	Menge aller linearen kompakten Operatoren	42
$\ T\ $	Operatornorm	23
$\rho(T)$	Resolventenmenge von T	25
$\sigma(T)$	Spektrum von T	25

$O(f(x))$	Landau'sches Ordnungssymbol O	247
$o(f(x))$	Landau'sches Ordnungssymbol o	54
A^+	Pseudoinverse der Matrix A	261
$\text{rang}(A)$	Rang der Matrix A	253
$(X, \langle \cdot, \cdot \rangle)$	Prähilbert-Raum	9
$(X, \ \cdot \)$	Normierter Raum	7
E^s, E^u, E^c	Unterräume des \mathbb{R}^n	108
M^\perp	Orthogonales Komplement von M	32
$X_1 \oplus X_2$	Direkte Summe von X_1 und X_2	34
\mathbb{C}^n	n -dimensionaler komplexer euklidischer Vektorraum	3
$\mathcal{C}^0[a, b]$	Menge aller stetigen Funktionen auf $[a, b]$	9
$\mathcal{C}^k[a, b]$	Menge aller stetigen und stetig differenzierbaren Funktionen bis zur Ordnung k	45
$\mathcal{L}(X, Y)$	Vektorraum der linearen stetigen Operatoren von X in Y	23
$\mathcal{L}^2(\mathbb{T}^p)$	Hilbert-Raum der auf \mathbb{T}^p quadratisch summierbaren Funktionen	366
$\mathcal{P}(\mathbb{T}^p)$	Raum der trigonometrischen Polynome über \mathbb{T}^p	365
$\text{codim } X$	Kodimension von X	35
l^p	Raum aller Zahlenfolgen, summierbar zur p -ten Potenz	9
\mathbb{R}^n	n -dimensionaler reeller euklidischer Vektorraum	3
$\mathcal{H}^s(\mathbb{T}^p)$	Sobolew-Raum der auf \mathbb{T}^p quadratisch summierbaren Funktionen mit Ableitungen bis zur Ordnung s	403
$\mathcal{L}^2([0, T])$	Hilbert-Raum der auf $[0, T]$ quadratisch summierbaren Funktionen	311
$i = a(h)b$	Sequenz $i = a, a+h, a+2h, \dots, b$ in abgekürzter Notation	248
$\langle x, y \rangle$	Skalarprodukt von x und y	9
$\text{span}\{v_1, \dots, v_m\}$	Linearer Unterraum, der durch die Vektoren v_1, \dots, v_m aufgespannt wird	404
$\nu(A)$	Spektralabszisse der Matrix A	274
\mathbb{S}^1	1-dimensionale Sphäre	320
$T_y \mathcal{M}$	Tangentialraum von \mathcal{M} bei y	265
\mathbb{T}^p	p -dimensionaler Standardtorus	364
$W_{loc}^c(0)$	Lokale Zentrumsmannigfaltigkeit	335
$W(t)$	Wronskideterminante; $W(t) := \det X(t)$	87

Index

- Abbildung, 18
 - bijektive, 17
 - Bild, 18
 - Definitionsbereich, 18
 - Domain, 18
 - homöomorphe, 21
 - isometrisch isomorphe, 20
 - isometrische, 17
 - lineare, 17, 19
 - offene, 27
 - Range, 18
 - stroboskopische, 333
 - topologisch äquivalente, 21
 - Wertebereich, 18
- Ableitung
 - Fréchet-, 54
 - Gâteaux-, 55
 - höhere, 55
- Ableitungsprädiktor, 268, 270
- Abstiegsbedingung, 241
- Ambrosetti A., 175
- Amplitudenmodulation, 368
- Anfangswertproblem, 72
- Approximation
 - Koeffizienten der besten, 15
 - streng konsistente, 239
- Approximationsfehler, 240
- Approximationsordnung, 247, 269
- Approximationssatz
 - von Weierstraß, 365
- Armijo-Regel, 242
- Atlas, 114
- Attraktor, 105, 228, 351, 382
- Auto2000, 356, 357, 395, 410

- Banachraum, 8, 370
- Basis, 267
- Basis-Prädiktor, 246
- Basisdimension, 372
- Basisdiskretisierung, 388
- Basisfrequenzen, 366
- bialternierendes Produkt, 281, 345
- Bifurkation
 - Detektierung, 343, 353
 - Einschachtelung, 347
 - Falten-, 276, 278
 - Flip-, 339
 - Fold-, 337
 - Heugabel-, 276, 278, 338
 - Hopf-, 210, 276, 280
 - Neimark-Sacker-, 340
 - periodischer Orbits, 333
 - Pitchfork-, 170, 276, 339
 - Sattel-Knoten-, 276, 337
 - subharmonische, 339
 - Szenario, 354
 - Torus-, 340
 - transkritische, 170, 276, 278, 338
- Bifurkationsbedingung
 - notwendige, 155
- Bifurkationsdetektierung, 276
- Bifurkationsdiagramm, 350, 352
- Bifurkationsgleichung, 155, 156
 - Ljapunov, 157
 - Ljapunov-Schmidt, 156
 - Schmidt, 159
- Bifurkationslösung
 - subkritisch, 174
 - superkritisch, 174
 - transkritisch, 174, 177
- Bifurkationspunkt, 153, 155
- Bifurkationstheorie
 - dynamische, 151
 - statische, 151
- Bifurkationstyp, 276
- Bifurkationstypen, 133
- biorthogonal, 28
- Bogenlänge, 255
- Bogenlängen-Parameter, 257
- Borel, E., 7
- Bratu-Gelfand-System, 271

- C^1 -Diffeomorphismus, 319
- C^r -Glattheit, 265
- C^r -Torusfunktion, 365
- CFL-Bedingung, 391
- Charakteristik, 382
- charakteristischer Exponent, 273
- charakteristischer Wert, 24
- Chua L., 285
- Courant-Friedrichs-Lewy-Zahl, 392
- Crandall, M., 165

- Dämpfungsfaktor, 242
- Dauidenko-DGL, 245
- De Morgan, 6
- DGL
 - periodisch erregte, 289
 - vom Mathieuschen Typ, 394
- DGL-System
 - charakteristisches, 381
- Diffeomorphismus, 21
- Differential
 - Fréchet-, 54
 - Gâteaux-, 55
- Differenzenapproximation, 239, 240

- Differenzenquotient, 238
- Differenzenschema, 389
- Differenzialgleichung
 - autonomes System, 98
 - explizite, 70
 - hamiltonsche, 132
 - heteronomes System, 98
 - implizite, 70
 - kleine Schwingungen, 205
 - konstante Koeffizienten, Lösung, 91
 - linear beschränkte rechte Seite, 82
 - linear, n-ter Ordnung, 97
 - maximales Lösungsintervall, 100
 - Maximalintervall, 80
 - n-ter Ordnung, 70
 - partielle, 381
 - Phasengeschwindigkeit, 98
 - Phasenkoordinaten, 98
 - Phasenkurve, 98, 100
 - System
 - Dimension des Lösungsraumes, 85
 - erster Ordnung, 83
 - Lösungsraum, 84
 - lineares homogenes, 84
 - lineares inhomogenes, 84
 - mit konstanten Koeffizienten, 89
 - zweiter Ordnung, 205
- Differenzialoperator, 413
- Dimensionsformel, 21
- direkte Summe, 34
- Dirichlet, P., 71
- Diskretisierungsparameter, 388, 413
- Dreiecksungleichung, 7
- Dualraum, 28
- Duffing-Van-der-Pol-Gleichung, 289

- Eigenvektor, 24
- Eigenwert, 24
 - algebraische Vielfachheit, 51, 92
 - einfacher, 92
 - geometrische Vielfachheit, 51, 92, 125
 - halbeinfacher, 92
- Eigenwertdiagramm, 109
- Eigenwerte, 26
- Einbettungsgleichung, 245
- Einschrittverfahren, 388
- Einzugsbereich, 241, 242
- Entweichzeit, 82
- Erregungsfrequenz, 399
- Euler L., 246
- Explosionszeit, 82

- Finite-Differenzen-Methode, 387
- Finities-Differenzen-Verfahren
 - von positivem Typ, 389
- first return map, 319

- Fixpunkt, 104, 228
- Fixpunktiteration, 387
- Floquet
 - Darstellung, 297
 - Multiplikator, 298
 - Theorie, 297
- Fluss, 91, 101
 - paralleler, 400
- Fokus, 143
 - instabiler, 143
 - stabiler, 143
- Folge, 5
 - beschränkte, 5
 - Cauchyfolge, 5
 - konvergente, 5
- Folgenraum, 12
- Fortsetzungsalgorithmus, 269
- Fortsetzungsproblem, 329
- Fortsetzungsschritt, 245, 249
- Fortsetzungsschrittweite, 250, 251
- Fourier
 - Ansatz, 414
 - Galerkin-Methode, 412
 - Galerkin-Verfahren, 417, 418
 - Koeffizient, 13, 14, 406, 407
 - Reihe, 13, 14, 370
 - Transformation, 415, 416
- Fredholmsche Alternative, 37, 50
- Frequenz
 - inkommensurable, 367
- Frequenzbasis, 366, 372, 374, 400
 - maximale, 373
 - reale, 373
- Frequenzmodulation, 368
- Fundamentalmatrix, 86, 292, 297
- Fundamentalsystem, 86
- Funktion
 - biperiodische, 367
 - Mittelwert der, 370
 - quasi-periodische, 366, 420
 - triperiodische, 367
- Funktionentheorem
 - implizites, 253

- Galerkin B.G., 412
- Galerkin-Bedingung, 406, 414
- Galerkin-Methode, 405
- Gauß-Newton-Fortsetzung, 263
- Gauß-Newton-Korrektor, 263, 280
- Gauß-Newton-Verfahren, 413, 415
- Gitterfunktion, 389
- Gleichgewichtslösung, 228
- Gleichgewichtspunkt, 104, 228
- gleichgradig stetig, 44
- Gleichungssystem
 - unterbestimmtes, 261, 262

- Gram-Schmidt-Orthogonalisierung, 266
- Greensche Funktion, 48
- Grenzpunkt, 254
- Grenzyklus, 143, 185, 289
 - Stabilität, 202
 - van-der-Pol, 196
 - Stabilität eines, 191
- Hamiltonfunktion, 132
- harmonischer Oszillator, 100, 103
- Häufungspunkt, 6
- Hauptfundamentalmatrix, 86, 292
- Hauptlösungsmatrix, 297
- Hauptmultiplikator, 324
- Hauptvektor, 94
- Heine, E., 7
- Hénon M., 326
- Heugabel-Bifurkation, 276, 278
- Hilbert-Raum, 366, 403
- Hilbertraum, 11
- Homöomorphismus, 20, 118, 374
- Hopf, E., 140
- Hopf-Bifurkation, 140, 144, 147, 210, 276, 282
 - subkritisch, 144
 - superkritisch, 144
- Hopf-Punkt, 280
- hyperbolischer Fluss, 110
- hyperbolischer Gleichgewichtspunkt, 111
- Hyperbolizität, 334
- inkommensurabel, 366
- Inkommensurabilität, 368
- Integraloperator, 44
- Intervallschachtelung, 346
- invarianter Unterraum, 110
- Invarianzgleichung, 402, 412, 422
- Isometrie, 20
- isometrischer Isomorphismus, 20
- Iterationsverfahren, 232
- Jacobi-Matrix, 229, 253, 273, 275
- Kantorovics L.V., 234
- Karte, 113
- Kartenabbildung, 113
- Kartenbild, 113
- Kernfunktion, 48
- Kompaktheit, 6
 - präkompakt, 6
 - relativ kompakt, 6, 42
- Konsistenz, 389
- Kontraktionsrate, 251
- Kontrollparameter, 229
- Konvergenz, 14, 389
 - globale, 230, 231, 242
 - kubische, 233
 - lineare, 233
 - lokale, 230, 233, 235, 239
 - quadratische, 233, 234
 - semilokale, 230, 231, 234
 - überlineare, 233
- Konvergenzfaktor, 232, 233
- Konvergenzordnung, 233
- Koordinatenwechsel, 113
- Korrektor, 263
- Korrektorschritt, 248
- Korrekturverfahren, 249
- kritischer Punkt, 104, 228
- Kronecker, L., 28
- Kroneckersymbol, 28
- Kugel, 4
- Kurvenschar, 73
- Langford W. F., 313, 343, 361
- limit cycle, 144
- Lindelöf, E., 74
- Linearisierung, 111, 229
- Linielement, 73
- Liouville, J., 71
- Ljapunov, A.M., 154
 - Bifurkationsgleichung, 157
 - Projektionsmethode, 157
- Ljapunov-Funktion, 192
- Ljapunov-Schmidt-Methode, 181
- Ljapunov-Schmidt-Reduktion, 163, 184
- Ljapunov-Stabilität, 273, 296
- lokale Koordinaten, 113
- Lorenz E.N., 228
- Lösung
 - geometrisch isolierte, 231
 - harmonische, 289
 - isolierte, 231
 - periodische, 287
 - quasi-periodische, 371
 - reguläre, 231, 292, 294, 311
 - subharmonische, 290
- Lösungsdiagramm, 243
- Lösungskurve, 100
- Lösungsmannigfaltigkeit, 264
- Lösungspfad, 253, 254, 262
 - regulärer, 254
- Mannigfaltigkeit, 113, 266, 268
 - globale, 117
 - invariante, 378
 - Invarianz der, 113
 - p-dimensionale, 265
 - toroidale, 374
 - Untermannigfaltigkeit, 113
- MAPLE, 299, 410, 422
- MATLAB, 285, 286, 356, 359, 417

- Matrix-Exponentialfunktion, 90
- Matrixdifferenzialgleichung, 86
- Mehrfach-Schießverfahren, 301, 303
- Mehrzielmethode, 302
- Menge
 - abgeschlossene, 4
 - offene, 4
 - präkompakt, 6
 - separable, 17
- Methode
 - direkte, 357
 - indirekte, 356
 - von Galerkin, 412
- metrischer Raum, 3
 - vollständiger, 5
- Minimierungsaufgabe, 261
- Mittelungsmethode, 185
- Modellproblem, 399, 400
- Monodromiematrix, 298
- Moore-Penrose-Inverse, 261
- moving frame, 267
- multiple shooting, 302
- Multiplikator
 - charakteristischer, 298
- Nemytski-Operator, 214, 413
- Netzwerk
 - parametrisch erregtes, 394
- Newton I., 229
- Newton-ähnliches Verfahren, 238
- Newton-Korrektur, 230
- Newton-Matrix, 291
- Newton-Verfahren, 229, 230, 291, 404, 418
 - approximatives, 269
 - gedämpftes, 242
 - mit Differenzen, 240
 - vereinfachtes, 235
- Niveaumenge, 242
- Norm, 7
 - Betragssummen-, 8
 - Maximum-, 8
 - p-Norm, 8
- Norm-Isomorphismus, 20
- Normalraum, 265
- normierter Raum, 7
- Operator
 - adjungierter, 30
 - bijektiver, 19
 - F-Operator, 38
 - Fredholm-, 38
 - identischer, 19
 - injektiver, 19
 - Integral-, 44
 - Kern, 18
 - kompakter, 42
 - linear, beschränkt, 23
 - linear, stetig, 23
 - Nemytski, 57
 - Norm, 23
 - Nullraum, 18
 - Projektions-, 34
 - selbstadjungierter, 32
 - Substitutions-, 57
 - surjektiver, 18
 - symmetrischer, 30, 32
 - Umkehr-, 25
 - Volterra-, 46
- Operatorgleichung, 413
- Orbit, 100, 103
 - biperiodischer, 402
 - heterokliner, 106, 131
 - homokliner, 131
 - isolierter, 288
 - periodischer, 106, 287, 330
 - quasi-periodischer, 367
- orthogonal, 11
- Orthogonalitätsbedingung, 264
- Orthogonalreihe, 14
- Orthonormalbasis, 265
- Orthonormalsystem, 14, 370, 405
 - Vollständigkeit, 14
- p-Torus, 364, 374, 398
 - invarianter, 375
 - Invarianzgleichung für, 402
- Parabel-Prädiktor, 248
- Parameterwechsel, 254
- Parametrisierung
 - natürliche, 245, 250
- Parseval'sche Gleichung, 370
- PECE-Verfahren, 249
- periodische Lösung, 107
- Periodizität, 289
- Phasenbedingung, 309, 404
 - integrale, 310, 404
 - von Poincaré, 310
- Phasendifferenzialgleichung, 137
- Phasenfluss, 103
- Phasenkurve, 100
- Phasenportrait, 103, 132
- Phasenraum, 99, 103
- Philippow E., 394
- Picard, E., 74
- Pitchfork-Bifurkation, 137, 139, 147, 276
- Poincaré
 - Abbildung, 61, 197, 318, 320, 323, 324
 - Bedingung, 310
- Poincaré J. H., 197, 319
- Polarkoordinaten, 377, 396
- Polynom
 - quasi-trigonometrisches, 369

- trigonometrisches, 365
- Prädiktor-Korrektor-Verfahren, 249
- Prädiktorschritt, 246
- Prädiktorverfahren, 249
- Prinzip der eingefrorenen Koeffizienten, 393
- Prodi G., 175
- Projektionsmethode, 157
- Projektionsoperator, 405
- Projektor, 34
- Pseudo-Bogenlänge, 257, 259
- Pseudoinverse, 261
- Pseudoresolvente, 158

- Q-Ordnung, 232, 236
- QR-Zerlegung, 266
- Quelle, 110, 112

- Rabinowitz, P.H., 165
- Randwertproblem, 47, 386
- Randwertproblem (RWP), 161
- rational unabhängig, 366
- Raum
 - Banach-, 8
 - Hilbert-, 11, 16
 - Cauchyfolge im, 16
 - normierter, 7
 - Prähilbert-, 9
 - Skalarprodukt-, 9
- Rechtsinverse, 262
- Reduktionssatz, 334
- reguläre Werte, 25, 26
- regulärer Punkt, 244
- regulärer Wert, 244
- Regularität, 245
- Repellor, 105, 228
- Residuumsfunktion, 401
- Resolvente, 25
- Resolventenmenge, 25
- Resonanzkreis
 - nichtlinearer elektrischer, 408, 418
- response case, 402
- Response-Fall, 403, 414
- Richtungsfeld, 73
- RL-Stromkreis, 69
- Ruhelage, 228

- Sattel, 103, 106, 112
- Sattel-Knoten-Bifurkation, 134, 139, 147, 276, 278
- Satz
 - Closed Range Theorem, 33
 - Hahn-Banach Theorem, 28
 - Hartman-Grobman, 118
 - Hartman-Grobman Theorem, 130
 - Hopf-Bifurkation, 142
 - Hopf-Theorem, 211
 - Implizites Funktionentheorem, 60
 - Open Mapping Theorem, 27
 - Riesz-Schauder Theory, 51
 - über die implizite Funktion, 253
 - von A. N. Shoshitaishvili, 334
 - von Andronov & Witt, 317
 - von Arzelà-Ascoli, 44
 - von Crandall-Rabinowitz, 165
 - von Floquet, 297
 - von Hartman-Grobman, 112
 - von Hopf (1942), 222
 - von Kantorovics, 234
 - von Neimark & Sacker, 341
 - von Peano, 77
 - von Picard-Lindelöf, 77
 - von Riesz, 29
 - von Riesz-Fischer, 366
- Schießpunkt, 304
- Schießverfahren, 290, 292, 300, 301, 306, 387
 - einfaches, 306
 - für Tori, 387
- Schmidt, E., 154
 - Bifurkationsgleichung, 159
 - Schmidt-Operator, 159
- Schrittweitensteuerung, 250, 251
- Schwingung
 - multifrequentiell, 363
- Schwingungsgleichung, 177
- Sehnenverfahren, 235
- Sekanten-Prädiktor, 247, 254
- Semidiskretisierung, 405, 420, 422
- Senke, 110, 112
- Separatrix, 106
- Shamanskii-Verfahren, 236, 237
- Skalarprodukt, 9, 180
- Sobolew-Raum, 403
- Spektralmethode, 412, 418
- Spektralsystem, 406, 407
- Spektralwerte, 25
- Spektrum, 25
 - kontinuierliches, 26
 - Punkt-, 26
 - Residual-, 26
- Sphäre, 4
- Störfunktion, 83
- Stabilität
 - asymptotisch orbitale, 316
 - asymptotische, 296
 - orbitale, 316
 - periodischer Orbits, 296
 - strukturelle, 129
- Stabilitätsanalyse, 273, 275, 296, 315, 331
- Stabilitätstyp, 133
- Standardtorus, 364, 372
- stationärer Punkt, 169
 - hyperbolisch, 169

- Stroboskop-Abbildung, 321, 407
- Sturm, J. C. F., 182
- Sturm-Liouville Problem, 182
- sukzessive Approximation, 78
- Superpositionsprinzip, 84
- Supervektor, 415
- System
 - autonomes, 306
 - Dreiphasen-, 348
 - energetisches, 348
 - parameterabhängiges, 228
 - partitioniertes, 377, 396
 - periodisch erregtes, 384
 - quasilineares hyperbolisches, 386
 - subharmonisch reagierendes, 294
- Systemparameter, 227

- Tangenten-Näherungsverfahren, 229
- Tangenten-Prädiktor, 246, 254
- Tangentenvektor, 255, 259, 261
- Tangentialraum, 265
- Testfunktion, 277, 280, 345
- topologisch äquivalent, 118
- topologisch konjugiert, 118
- Torus
 - asymptotisch stabiler, 380
 - attraktiver, 382
 - ergodischer, 382
 - Existenz und Eindeutigkeit, 378
 - exponentiell stabiler, 380
 - mit freien Phasen, 402
 - Parametrisierung des, 375, 380
 - quasi-periodischer, 367, 369
 - resonanter, 367, 398
 - stabiler, 380
- Torus-Bifurkation, 408
- Torusbedingung, 384
- Torusfunktion, 364, 401
 - reelle, 364
 - stetig differenzierbare, 365
- Toruskoordinaten, 374–376, 399, 400
- Torusmannigfaltigkeit, 382
- Trajektorie, 100
- transkritische Bifurkation, 136, 139, 147, 276
- transversal, 141
- Transversalitätsbedingung, 141, 222

- Umgebung, 4
- Umkehrpunkt, 254
- Ungleichung
 - Cauchy-Schwarzsche, 10
- Untermannigfaltigkeit, 114
- Unterraum
 - instabiler, 107
 - stabiler, 107
 - Zentrums-, 107

- van-der-Pol Balthasar, 184
- Van-der-Pol-
 - Oszillator, 376, 382
- van-der-Pol-
 - Differenzialgleichung, 184, 222
- Variation der Konstanten, 88
- Variationssystem, 292, 303
- Verfahren
 - Friedrichs-, 391, 392
 - implizites CIR-, 393
 - Upwind-, 390, 392–394, 396
 - vom Upwind-Typ, 390
 - von Courant-Isaacson-Rees, 390
- Verzweigungsgleichung, 184
- Volldiskretisierung, 420
- Volterra-Lotka Gleichungen, 130

- Wronskideterminante, 87

- Zahlenfolge
 - komplexe, 5
 - reelle, 5
- Zentrum, 103, 106
- Zentrumsmannigfaltigkeit, 117, 121, 136, 335
- Zentrumsunterraum, 107, 122
- Zweipopulationenmodell, 129