

Mathematische Modellbildung in Zeit, Ebene und Raum

Version Herbst 2020 vom 11. September 2020

Alexander Caspar, Norbert Hungerbühler

ETH Zürich D-MATH

Inhaltsverzeichnis

1 Modellbildung	5
1.1 Einführung und Beispiele	5
1.1.1 Was ist ein (mathematisches) Modell?	5
1.1.2 Erste Annäherung an den Begriff des Modells	8
1.1.3 Wie erstellt man ein Modell?	9
1.1.4 Wie hängen Modell und Realität zusammen?	10
1.1.5 Theorien	11
1.1.6 Wozu braucht man Modelle?	12
1.2 Systeme und Systemanalyse	13
1.2.1 Beispiel: Phosphatkonzentration in einem See	14
1.3 Beispiele von Modellen	16
1.3.1 Das Ohmsche Gesetz	16
1.3.2 Das Sonnensystem	16
1.3.3 Geometrische Optik	17
1.3.4 Algenwachstum und Nitratkonzentration	19
1.3.5 Ein See als linearer Durchflussreaktor	20
1.4 Stoffaustausch an Grenzflächen	23
1.5 Mehrdimensionale Modelle	26
1.6 Die Pocken und das „Dreckige Dutzend“	27
1.6.1 Einige Fakten über die Pocken	27
1.6.2 Impfungen	28
1.7 Daniel Bernoulli und die Pocken	29
1.7.1 Daten aus Halleys Sterbetafel	30
1.7.2 Daniel Bernoullis Modell	31
1.7.3 Bernoullis Folgerung	34

1.8	Das SIR-Modell	35
1.8.1	Ansteckung	35
1.8.2	Heilung	36
1.8.3	Diskussion des SIR Modells	37
1.9	Diskrete Modelle	40
1.9.1	Beispiel Altersstruktur einer Population	41
2	Diskrete Modelle	43
2.1	Eindimensionale Entwicklungen	43
2.1.1	Lineare Entwicklung	43
2.1.2	Exponentielle Entwicklung	45
2.1.3	Begrenzte Entwicklungen	46
2.1.4	Diskrete logistische Entwicklung	48
2.1.5	Fixpunkte eindimensionaler diskreter Entwicklungen	50
2.1.6	Reproduktion	52
2.1.7	Fixpunkte und Konvergenz	55
2.2	Mehrdimensionale Modelle	58
2.2.1	Beispiel Entwicklung Räuber / Beute - Populationen	58
2.2.2	Vereinfachung mit Hilfe der Linearität	62
2.2.3	Explizite Darstellung einer Entwicklung	65
2.2.4	Beispiele Altersstrukturen 3×3	67
2.2.5	Zusammenfassung $n \times n$ - Matrix	70
3	Lineare Modelle	73
3.1	Eindimensionale Lineare Modelle	73
3.1.1	Exponentielle Veränderungen	73
3.1.2	Beschränktes Wachstum	75
3.1.3	Zusammenfassung Lineare DGL mit konstanten Koeffizienten	76
3.2	Höherdimensionale Modelle	78
3.2.1	Beispiele: Serielle Box-Kompartiment-Modelle	79
3.2.2	Ein 3-Box-Modell	82
3.2.3	Weitere Beispiele für (lineare) Kompartiment-Modelle	84
3.2.4	Zusammenfassung für homogen und linear unabhängigen Fall	86
3.3	Vektorräume	87

3.4	Erzeugendensysteme und Basen	91
3.4.1	Linearkombination in einem Vektorraum V	91
3.4.2	Basen	93
3.5	Diagonalisierbarkeit	96
3.5.1	Definition und Anwendung	96
3.6	Exponential einer Matrix	97
3.6.1	Definition, Anwendungen, erste Berechnung von e^A	98
3.6.2	Berechnungsmethode für e^A	99
3.6.3	Berechnung von e^A mit der Jordan-Normalform	103
3.6.4	Anwendung serielles 3-Box-Modell mit Rückfluss.	105
3.6.5	Vergleich der Methoden	107
3.7	Der inhomogene Fall	108
3.7.1	Partikuläre Lösung eines Systems von DGL	109
3.7.2	Lösung bei Diagonalisierbarkeit	110
4	Fourier-Reihen	113
4.1	Einführung	113
4.1.1	Erinnerung periodische Funktionen	113
4.1.2	Beispiele Medikamentenmenge	114
4.1.3	Zusammenfassung Taylor-Reihe	115
4.2	Trigonometrische Polynome / Reihe	117
4.2.1	Fourier-Koeffizienten und Trigonometrische Orthogonalität	117
4.2.2	Was hat das mit Lineare Algebra zu tun?	119
4.3	Euklidische Vektorräume	120
4.3.1	Definition und Beispiele Euklidischer VR	120
4.3.2	Orthogonale Vektoren	121
4.3.3	Normierte Vektorräume	121
4.3.4	Orthonormalbasen	124
4.3.5	Anwendung auf Fourier-Koeffizienten	125
4.4	Orthogonale Projektion	126
4.4.1	Konstruktion	126
4.4.2	Anwendung auf Fourier-Reihen	126
4.4.3	Anwendungen	127
4.5	Ergänzungen und Beispiele	129

4.5.1	Satz von Dirichlet	129
4.5.2	Fourier-Reihe bei allgemeiner Periode T	130
4.5.3	Komplexe Schreibweise	130
5	Nichtlineare Modelle	133
5.1	Eindimensionale Modelle	134
5.1.1	Logistische Gleichung	134
5.1.2	Wachstum von Tumoren	137
5.1.3	Stationäre Lösungen	139
5.1.4	Beschreibung in der Nähe von y_∞ , Linearisierung	139
5.2	Mehrdimensionale Modelle	141
5.2.1	Räuber-Beute-Modell nach Lotka-Volterra	141
5.2.2	Lösungskurven im Lotka-Volterra-Modell	143
5.2.3	Linearisieren in 2 Dimensionen	144
6	Partielle Differentialgleichungen	148
6.1	Einführung	148
6.1.1	ODEs vs. PDEs: Ein Vergleich	148
6.1.2	Begriffe aus der Vektoranalysis	149
6.1.3	Wärmeleitung und Diffusion	154
6.2	Fourier-Methoden	156
6.2.1	Die Wärmeleitungsgleichung	157
6.2.2	Die Laplace-Gleichung	161
6.2.3	Das Fourier-Integral	165
6.2.4	Die Wellengleichung auf \mathbb{R}	168
6.2.5	Die Methode von d'Alembert	169
6.2.6	Filter	174
6.2.7	Das Shannon-Sampling-Theorem	177
6.2.8	Computer-Tomographie	179
6.3	Numerische Methode für die Poisson-Gleichung	182
7	Laplace-Transformation	188
7.1	Definition und Notation	188
7.2	Rechenregeln	190
7.3	Anwendungsbeispiel	194

KAPITEL 1

Modellbildung

1.1 Einführung und Beispiele

Alle quantitativ arbeitenden Natur- und Ingenieurwissenschaften, und inzwischen auch Sozial- und Geisteswissenschaften bedienen sich mathematischer Modelle. Die rasanten Fortschritte in fast allen Aspekten der menschlichen Gesellschaft und insbesondere die Explosion der Technologie des 20. und des beginnenden 21. Jahrhunderts sind eine Folge dieser Entwicklung. Wie ist das zu erklären?

Um uns einer Antwort zu nähern, wenden wir uns in den nächsten Abschnitten den folgenden vier Grundfragen zu:

- Was ist ein Modell?
- Wie erstellt man ein Modell?
- Was ist die Beziehung zwischen Modell und Realität?
- Welchen Nutzen haben Modelle?

1.1.1 Was ist ein (mathematisches) Modell?

Eine Lexikonantwort könnte wie folgt lauten.

Definition 1.1.1. *Ein mathematisches Modell ist die Beschreibung eines realen Phänomens in der Sprache der Mathematik.*

Die Inhalte der Mathematik sind abstrakter Natur. Wie kann es sein, dass die Mathematik sich dennoch so gut eignet, reale Zusammenhänge unserer Welt zu beschreiben? Diese Frage wird unter Philosophen und Wissenschaftstheoretikern kontrovers diskutiert. Schauen wir, was bedeutende Wissenschaftler dazu zu sagen hatten. Lassen wir zuerst Galileo Galilei zu Wort kommen. Er wollte als Erster die qualitative Naturlehre von Aristoteles auf eine quantitative, also mathematische Grundlage stellen. Unter anderem beschrieb er als Erster die physikalischen Fallgesetze,

indem er die Begriffe *Beschleunigung*, *Geschwindigkeit* und *zurückgelegte Wegstrecke* durch eine mathematische Formel miteinander in Verbindung brachte, welche seine Beobachtungen aus zahlreichen Experimenten beschrieb.

Zitat (Galileo Galilei, 1623). *Das Buch der Natur ist in der Sprache der Mathematik geschrieben und ihre Buchstaben sind Dreiecke, Kreise und andere geometrische Figuren, ohne die es ganz unmöglich ist auch nur einen Satz zu verstehen, ohne die man sich in einem dunklen Labyrinth verliert (siehe [galilei1623saggiatore]).*

Abbildung 1.1: Das Bild zeigt Galilei auf einem Porträt von Justus Sustermans aus dem Jahr 1636.



Die oben beschriebene Frage nach der verblüffenden Wirksamkeit der Mathematik brachte der ungarisch-amerikanische Physiker und Nobelpreisträger Eugene Wigner auf den Punkt:

Zitat (Eugene Wigner, 1960). *The miracle of the appropriateness of the language of mathematics for the formulation of the laws of physics is a wonderful gift which we neither understand nor deserve. We should be grateful for it and hope that it will remain valid in future research and that it will extend, for better or for worse, to our pleasure, even though perhaps also to our bafflement, to wide branches of learning (siehe [wigner1960]).*

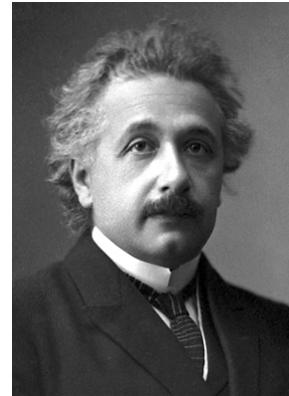
Abbildung 1.2: Eugene Wigner auf dem offiziellen Foto der Nobel-Stiftung anlässlich der Verleihung des Physik Nobelpreises 1963.



Albert Einstein, der die Physik seiner Zeit und unsere Vorstellung von Raum und Zeit durch seine mathematisch formulierten Theorien revolutionierte, sah das Verhältnis von Realität und Mathematik differenziert und durchaus kritisch:

Zitat (Albert Einstein, 1921). *Wie ist es möglich, dass die Mathematik, die doch ein von aller Erfahrung unabhängiges Produkt des menschlichen Denkens ist, auf die Gegenstände der Wirklichkeit so vortrefflich passt? [...] Hierauf ist nach meiner Ansicht kurz zu antworten: Insofern sich die Sätze der Mathematik auf die Wirklichkeit beziehen, sind sie nicht sicher, und insofern sie sicher sind, beziehen sie sich nicht auf die Wirklichkeit (siehe [\[einstein1921\]](#)).*

Abbildung 1.3: Albert Einstein auf einem Foto aus dem Archiv der Nobel-Stiftung. Einstein wurde 1921 mit dem Nobelpreis für Physik ausgezeichnet.



Pflanzt man an vergleichbaren Standorten je einen Apfelkern, so werden die daraus entstehenden Bäume schlussendlich längst nicht genau gleich aussehen. Die Äpfel beider Bäume fallen aber nach genau denselben Fallgesetzen, die schon Galilei mathematisch beschrieben hat. Die Physik gehorcht, ungleich genauer als die belebte Natur, sehr eng und reproduzierbar, strengen und strukturierten Gesetzen. Die Mathematik hält in ihrem riesigen Gebäude entsprechende Strukturen bereit, die sich zur Beschreibung der physikalischen Phänomene eignen. Aber wie sieht es mit Modellen ausserhalb der Physik aus? Hier ist die Antwort komplizierter. Die Mechanismen hinter der belebten Natur sind um ein Vielfaches komplexer als die Gesetze der Physik. Als Kontrast zu Wigners Aussage formulierte der russische Mathematiker Israel Gelfand, der sich intensiv mit Biomathematik und Molekularbiologie befasste, noch in den Sechzigerjahren des zwanzigsten Jahrhunderts auffallend pessimistisch:

Zitat (Israel Gelfand, 1966). *Eugene Wigner wrote a famous essay on the unreasonable effectiveness of mathematics in natural sciences. He meant physics, of course. There is only one thing which is more unreasonable than the unreasonable effectiveness of mathematics in physics, and this is the unreasonable ineffectiveness of mathematics in biology (siehe [\[gelfand1966\]](#)).*

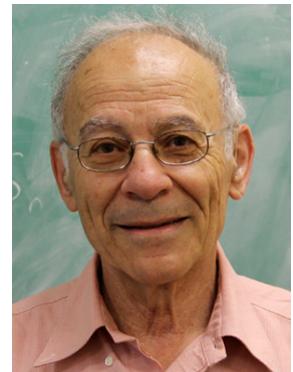
Abbildung 1.4: Israel Gelfand um 1950 als Professor der berühmten Lomonossow-Universität in Moskau.



Avner Friedman ist ein israelisch-US-amerikanischer Mathematiker, der unter anderem das *Mathematical Biosciences Institute* an der Ohio State University gegründet hat. Es ist auch Initiator, Mitautor und Herausgeber Buchreihe *Tutorials in Mathematical Biosciences* [[friedmanI](#)]. Er hält der eher skeptischen Aussage Gelfands nach einer stürmischen Entwicklung im Bereich der Biomathematik im letzten Viertel des 20. Jahrhunderts folgendes entgegen:

Zitat (Avner Friedman, 2010). *The gulf between biology and mathematics is narrowing as each domain tackles the language, concepts, and methods of the other (siehe [/friedman2010/](#)).*

Abbildung 1.5: Avner Friedman als Gründungsdirektor des *Mathematical Biosciences Institute* an der Ohio State University.



Rückblickend können wir folgendes festhalten: Die Anwendungen der Mathematik in den Naturwissenschaften beschränkte sich lange Zeit auf die Astronomie und die Physik. Später kamen Chemie, Ingenieurwissenschaften und Wirtschaftswissenschaften hinzu. Nur vereinzelt blitzte das Potential mathematischer Modelle in den Lifesciences auf: Etwa bei der probabilistischen Beschreibung der Mendelschen Gesetze, bei der Beschreibung der Populationsdynamik bei Räuber-Beute-Modellen (siehe Abschnitt 2.2.1 oder 5.2.1) oder bei der Diskussion der Wirksamkeit von Impfungen (siehe Abschnitt 1.7). Erst im letzten Viertel des 20. Jahrhunderts entwickelten sich Anwendungen der Mathematik in den Lifesciences in geradezu atemberaubendem Tempo und veränderten das Gesicht dieser Wissenschaften grundlegend. So können heute Enzymdynamik, Tumorwachstum und Krebstherapie, Seuchenausbreitung, Bioreaktoren, neuronale Reizleitung, Antibiotikaresistenzen, Zellalterung, Angiogenese, Bildung von Granulomen, die Physiologie des Hörens, usw. mithilfe mathematischer Modelle beschrieben, analysiert und tiefer als je zuvor verstanden und erforscht werden.

1.1.2 Erste Annäherung an den Begriff des Modells

Typischerweise beschränken sich Modelle auf einen eng begrenzten Ausschnitt der Realität. Man spricht von einem *reduktionistischen Ansatz*. Will man beispielsweise die Entwicklung einer Räuber- und einer Beutepopulation beschreiben, wird man im einfachsten Fall ausser Acht lassen, dass die betrachteten Raubtiere auch andere Beutetiere schlagen und dass die betrachteten Beutetiere noch anderen Räubern zum Opfer fallen. Und erst recht wird man das ganze umgebende Ökosystem ausblenden. Eine Kritik an dieser Einschränkung liegt auf der Hand: Was nicht ins Modell passt, wird weggelassen, auf die Gefahr hin, dass man die grösseren Zusammenhänge aus dem Auge verliert. Deshalb ist zur Beschreibung komplexer Systeme eine Synthese von Modellen angezeigt. Dies entspricht dann einem *holistischen*, also ganzheitlichen Ansatz. Bei alledem muss man sich immer vor Augen halten, dass Modelle in jedem Fall eine *vereinfachte* Beschreibung der Realität sind. Man beachte auch, dass man ein bestimmtes Phänomen oft auf unterschiedliche Weise modellieren kann: Je nach Fragestellung wird man etwa beim freien Fall den Luftwiderstand vernachlässigen oder im Modell integrieren. Auch die mathematischen Werkzeuge können sich von Modell zu Modell unterscheiden: Zum Beispiel in der Optik wird man je nachdem Lichtstrahlen geometrisch als einfache Geraden darstellen, oder man wird auf eine Beschreibung des Lichts als elektromagnetische Welle zurückgreifen.

1.1.3 Wie erstellt man ein Modell?

Die Modellbildung ist keine mathematische Disziplin: Es sind Kenntnisse aus der jeweiligen Disziplin über den zu modellierenden Vorgang erforderlich. Es ist von entscheidender Bedeutung, dass die Mechanismen des Phänomens genau bekannt und verstanden sind. Erst dann ist es möglich, diese mathematisch mithilfe von Formeln zu beschreiben und zu analysieren. Es liegt auf der Hand, dass in diesem Prozess Kenntnisse aus beiden Disziplinen erforderlich sind, also aus der Fachrichtung des betrachteten Phänomens und aus der Mathematik.

Bevor man mit dem Erstellen eines Modells beginnt, mache man sich klar:

- Welchen Ausschnitt der Realität will man modellieren? Das heisst, die *Systemgrenzen* müssen festgelegt werden.
- Welche Effekte sind vernachlässigbar, welche relevant?
- Welche mathematischen Strukturen sind geeignet um das in Frage stehende Phänomen zu beschreiben? Infrage kommen gewöhnliche Differentialgleichungen (engl. *Ordinary Differential Equations ODEs*), partielle Differentialgleichungen (engl. *Partial Differential Equations PDEs*), stochastische Methoden (aus der Statistik und der Wahrscheinlichkeitstheorie), lineare Algebra, Geometrie, ...

Bei der Beantwortung dieser Fragen helfen typischerweise Überlegungen der folgenden Art.

- Bilanzierung von Masse, Energie, Teilchenzahl, Anzahl Individuen einer Population, ...
- Aufstellen einer stöchiometrischen Reaktionsgleichung bei chemischen Vorgängen
- Betrachten von Wachstums-, Zerfalls- und Interaktionsprozessen: beispielsweise beim Algenwachstum, bei radioaktivem Zerfall, bei Ansteckungen durch Kontakt von kranken mit gesunden Individuen
- Betrachtung von Transportprozessen: beispielsweise Austausch von Stoffen über eine Grenzschicht hinweg, Diffusion von Wärme oder von Stoffen in einem ruhenden Medium, Verfrachten von Stoffen durch ein bewegtes Medium (*Advektion* oder *Konvektion*, beispielsweise der Transport eines Medikaments mit dem Blutstrom oder die Verteilung von Giftstoffen mit dem Wind)
- Anwenden von statistischen Methoden bei Systemen mit vielen Freiheitsgraden (bei subjektiver oder objektiver Unsicherheit)

Als nächstes muss man sich darüber klar werden, welche Fragen man mit dem Modell beantworten will. Also zum Beispiel: Will man Prognosen über das Langzeitverhalten machen? Will man den Output eines Prozesses optimieren? Wie genau oder detailliert sollen die Aussagen sein? Die Fragestellung diktiert dann unter anderem, welche Daten gesammelt, welche Prozesse modelliert, und welche Parameter als relevant zu betrachten sind.

Die typische Etappen der Modellbildung sind dann:

- Beobachten
- Daten sammeln: Messungen

- Daten ordnen: Suche nach mathematischen Mustern
- Verstehen: Erkennen der Grundprinzipien, Erstellen des mathematischen Modells
- Überprüfen (Validieren) des Modells: Vergleich mit der Realität
- Anpassen, Verfeinern oder Verallgemeinern

Wir werden weiter unten Beispiele zu diesen Schritten sehen.

1.1.4 Wie hängen Modell und Realität zusammen?

Immer wieder sind in der Vergangenheit Unglücksfälle (Einsturz von Gebäuden, Börsencrash, Dammbrüche) geschehen, die darauf zurückzuführen waren, dass inadäquate Modelle verwendet wurden. Das heisst, im Modell waren diese Unglücksfälle gar nicht möglich oder vorhesehbar gewesen, weil entscheidende Effekte bei der Modellbildung nicht berücksichtigt oder nicht bekannt waren. Als wichtigen Grundsatz halten wir daher fest:

$$\text{Modell} \neq \text{Realität}$$

Das Zusammenspiel zwischen Modell und Realität wird in der Abbildung 1.6 deutlich.

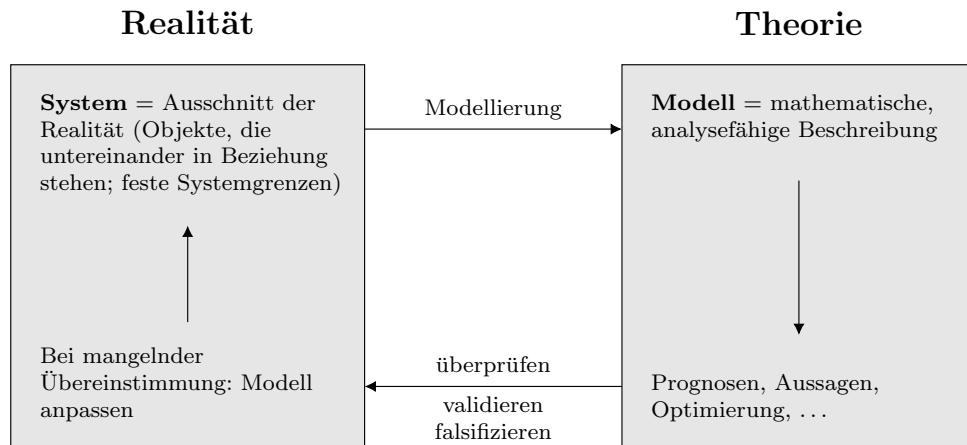


Abbildung 1.6: Schematische Darstellung des Verhältnisses von Modell und Realität.

Dabei ist zu beachten:

- Die (mathematische) Analyse liefert nur *innerhalb* der Theorie korrekte Resultate.
- Ist die mathematische Analyse zu schwierig, helfen numerische Simulationen und Berechnungen. Deren Resultate sind im Allgemeinen nur Approximationen: Nur falls Fehlerschranken mathematisch bewiesen wurden, sind die numerischen Aussagen (innerhalb dieser Schranken) verlässlich.

Der zuletzt erwähnte Punkt legt nahe, den oben formulierten Grundsatz um ein Element zu erweitern:

Realität \neq Mathematisches Modell \neq Numerisches Modell

1.1.5 Theorien

Modelle bauen auf Theorien auf. Letztere entwickeln sich häufig über einen längeren Zeitraum, und sie stellten in der Vergangenheit oftmals historisch bedeutsame Fortschritte dar.

Beispiele

- Newtonsche Mechanik
- Einsteins Relativitätstheorie
- Quantenmechanik (Werner Heisenberg, Erwin Schrödinger, ...)
- Teilchenphysik (Vorhersage des Higgs Bosons)
- Statik (Festigkeit von Bauwerken)
- Mendelsche Gesetze (Gregor Mendel, Vererbungslehre)
- Evolutionstheorie (Charles Darwin, Alfred Russel Wallace)
- Doppelhelixmodell der DNS (James Watson, Francis Crick)
- Zellbiologie (Antoni van Leeuwenhoek, Robert Hooke)

Mit der Entwicklung, Bedeutung und mit den Grenzen von Theorien befassen sich die Erkenntnis- und die Wissenschaftstheorie. Ein bedeutender Exponent dieser philosophischen Richtung war Karl Popper, der Begründer des *Kritischen Rationalismus*. Popper stellte unmissverständlich die Grenzen der Gültigkeit von Theorien fest:

Zitat (Karl Popper, 1934). *Über Folgerungen [von Theorien] wird im Zusammenhang mit der praktischen Anwendung, den Experimenten usw. entschieden. Fällt die Entscheidung positiv aus, werden die Folgerungen anerkannt, verifiziert, so hat das System die Prüfung vorläufig bestanden; wir haben keinen Anlass, es zu verwerfen. Fällt eine Entscheidung negativ aus, werden Folgerungen falsifiziert, so trifft ihre Falsifikation auch das System, aus dem sie deduziert wurden. Die positive Entscheidung kann das System immer nur vorläufig stützen; es kann durch spätere negative Entscheidungen immer wieder umgestossen werden (siehe [popper1935]).*

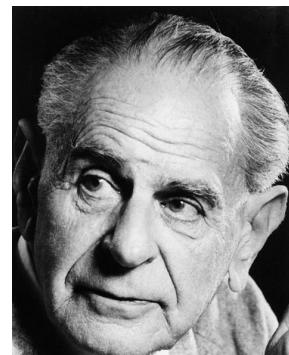
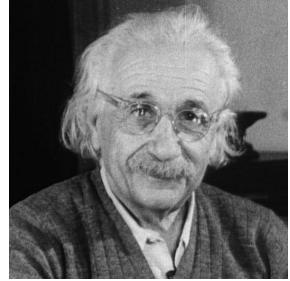


Abbildung 1.7: Karl Popper um 1980. Er war Träger zahlreicher Auszeichnungen und Ehrendoktorwürden.

Eine Theorie kann demnach nie abschliessend als Ganzes verifiziert, sondern nur falsifiziert werden: Egal wieviele Aussagen einer Theorie man überprüft und in der Realität als zutreffend beobachtet hat, man kann sich dennoch nie sicher sein, dass alle Sätze der Theorie mit der Wirklichkeit übereinstimmen. Hingegen genügt ein Beispiel einer Aussage der Theorie, die nicht mit der Realität übereinstimmt, um sie als unzureichend zu entlarven. Ist eine Theorie also falsifiziert, so muss sie modifiziert werden, um den Bereich ihrer verifizierbaren Aussagen zu erweitern. Dies ist jedoch nicht der einzige Grund für die Evolution von Theorien:

Zitat (Albert Einstein). *Neue Theorien entstehen zum einen, wenn neue Fakten nicht mit bestehenden Theorien erklärt werden können, zum anderen durch das Streben nach Vereinheitlichung und Vereinfachung zu einer Theorie als Ganzem (siehe [einstein1950]).*

Abbildung 1.8: Albert Einstein etwa 1950. Nach der nationalsozialistischen Machtergreifung in Deutschland emigrierte Einstein 1933 in die USA und arbeitete dort bis zu seinem Tode 1955 am *Institute for Advanced Study* in Princeton.



Erklären sowohl eine komplizierte, undurchschaubare Theorie (etwa die Epizykeltheorie des geozentrischen Weltbildes) als auch eine auf wenigen einfachen Grundsätzen beruhende Theorie (zum Beispiel das heliozentrische Weltbild) dieselben Erscheinungen der Wirklichkeit, wird man der einfacheren Theorie den Vorzug geben.

1.1.6 Wozu braucht man Modelle, was kann man mit ihnen machen?

Modelle dienen selten nur einem einzigen Zweck. Sie sind wahre Multitalente:

- *Modelle führen zu einem besseren Verständnis der Realität.* Welche Ursache hat welche Wirkung? Zum Beispiel, welche Parameter beeinflussen die Ausbreitung einer Seuche? Bei einer Pandemie wie etwa bei Covid-19 lassen sich mithilfe von Modellen Aussagen darüber treffen, wie sich zum Beispiel Social-Distancing-Massnahmen oder Maskenpflicht auf die Reproduktionsrate und damit die Ausbreitung der Seuche auswirken.
- *Modelle dienen der Optimierung.* Beispiele sind die Verringerung des Luftwiderstands und somit des Treibstoffverbrauchs bei Fahrzeugen durch aerodynamisch optimierte Formgebung oder die optimale Dosierung von Medikamenten.
- *Mit Modellen lassen sich numerische Simulationen durchführen.* Dies ist insbesondere dann der Fall, wenn die theoretische oder formelmässige Analyse zu komplex ist (so etwa bei Klimamodellen).
- *Mit Modellen lassen sich virtuelle Experimente durchführen.* Dann nämlich, wenn reale Experimente zu teuer, zu gefährlich oder schlichtweg undurchführbar sind. Zum Beispiel: Welche Erhöhung des Treibhausgases CO₂ in der Atmosphäre führt zu welchem globalen Temperaturanstieg?
- *Mit Modellen lassen sich Prognosen erstellen.* Modelle erlauben Vorhersagen über die Zukunft, zum Beispiel das Wetter von morgen.

1.2 Systeme und Systemanalyse

Die *Systemanalyse* ist der erste Schritt der Modellbildung: Dabei versucht man sich ein schematisches Bild der Situation zu machen, die interessierenden Objekte und deren Beziehungen zu identifizieren und die Systemgrenzen festzulegen. Über die Systemgrenzen hinweg bestehen Wechselwirkungen mit der Umwelt. Auch diese sind für die Systemanalyse von Interesse. In der Realität bestehen dabei in der Regel Rückkopplungseffekte. Solche Effekte werden bei der Systemanalyse vernachlässigt (siehe Abbildung 1.9). Dieses Vorgehen ist beispielsweise dann gerechtfertigt, wenn die Umwelt im Vergleich zum betrachteten System ein grosses Reservoir bildet in dem die Wirkung des Systems praktisch verschwindet. Stellen sich die Rückkopplungseffekte jedoch als zu gross heraus, müssen die Systemgrenzen erweitert werden, und zwar so, dass die Rückkopplungseffekte innerhalb des Systems verlaufen.

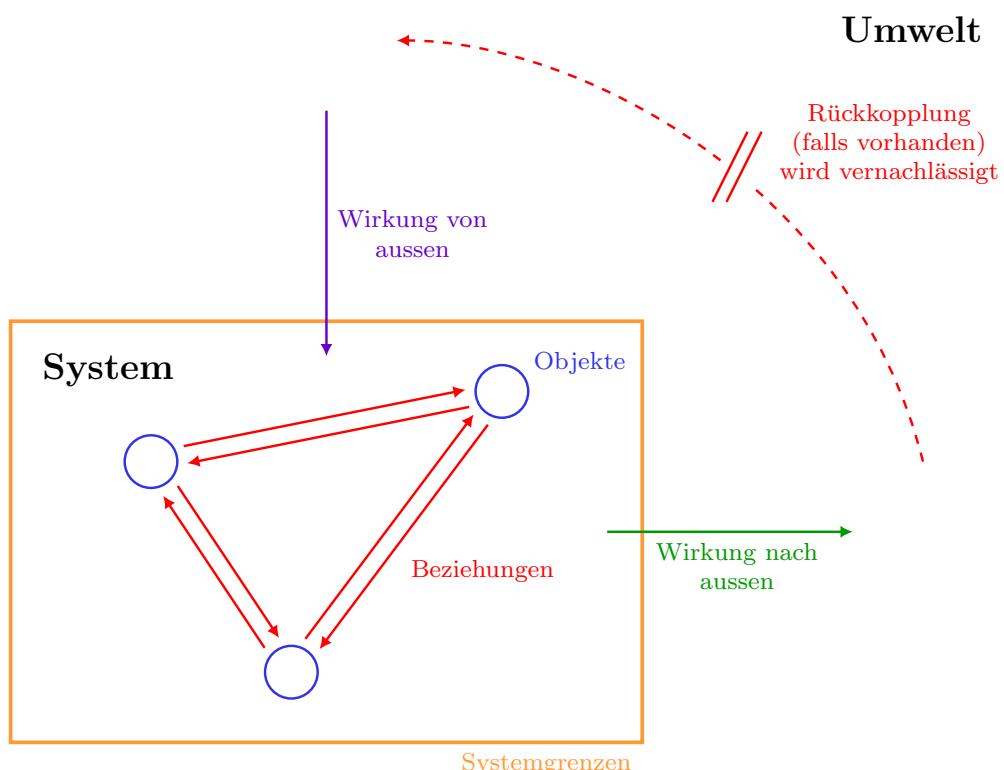


Abbildung 1.9: Schematische Darstellung eines Systems und seiner Systemgrenzen.

Die folgende Definition eines Systems hat sich als sinnvoll erwiesen:

Definition 1.2.1. Ein **System** ist eine gedankliche Konstruktion, die einen durch festgelegte **Grenzen** definierten Teil der Umwelt abbildet und dabei die **Objekte** und deren **Beziehungen** identifiziert.

Die Beziehungen zwischen den Objekten sowie die Wirkungen von aussen und nach aussen sind

dann auf der Grundlage einer passender Theorie in die Sprache der Mathematik zu übersetzen. Auf diese Weise führt die Systemanalyse zu einem mathematischen Modell.

Definition 1.2.2. Ein **Mathematisches Modell** ist eine vereinfachte Beschreibung eines komplexen Systems mit Hilfe mathematischer Formeln und von **Systemvariablen**. Die Anzahl der Systemvariablen heisst **Dimension des Modells**.

Beispiele für Systemvariablen sind Konzentration, Temperatur oder Mengenangaben.

1.2.1 Beispiel: Phosphatkonzentration in einem See

Phosphate werden in landwirtschaftlichen Düngemitteln sowie in Waschmitteln eingesetzt. Auch der Mensch scheidet Phosphate aus. Gelangen Phosphate über die Fliessgewässer in einen See, so beschleunigen sie dort das Wachstum der Algen. Bei dieser Eutrophierung wird dem Wasser schlussendlich Sauerstoff entzogen (siehe Abschnitt 1.3.4). Im schlimmsten Fall wird organisches Leben im See unmöglich. Aus diesem Grund ist es sehr wichtig, über den Phosphathaushalt eines Sees Bescheid zu wissen. Wir stellen dazu zunächst den uns interessierenden Ausschnitt des Systems See in Abbildung 1.10 vereinfacht dar.

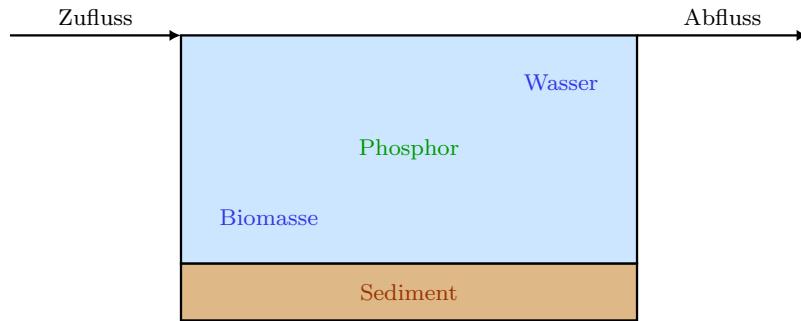


Abbildung 1.10: Der See mit seinen Systemgrenzen.

Dabei sind folgende Überlegungen gemacht worden, respektive weiter zu machen:

- Festlegen der Systemgrenzen
- Definieren der Systemvariablen
- Definieren der inneren und äusseren Relationen: Dabei fliessen empirische Überlegungen ein.

Der zuletzt genannte Punkt verknüpft die Systemvariablen miteinander: Welche Systemvariablen stehen miteinander in Beziehung? Dazu sind nun genauere Überlegungen nötig. Wir betrachten die Komponenten des Systems vereinfacht als räumlich vollständig durchmischt (homogen). Man nennt dies dann ein *Boxmodell*. Dieser Ansatz hat den Vorteil, dass die Konzentration eines Stoffes im See durch eine einzige, zeitabhängige Variable beschrieben werden kann, die nicht vom Ort abhängt. Der Nachteil ist natürlich, dass die Annahme dauernder und vollständiger Durchmischung im realen See nur näherungsweise erfüllt ist. In unserem Beispiel erhält man

ein Modell mit drei Systemvariablen: Masse des Phosphors im Seewasser, in der Biomasse und im Sediment. Das Modell hat also die Dimension 3. Die Situation ist in der Abbildung 1.11 schematisch dargestellt (die Abkürzungen sind unten beschrieben).

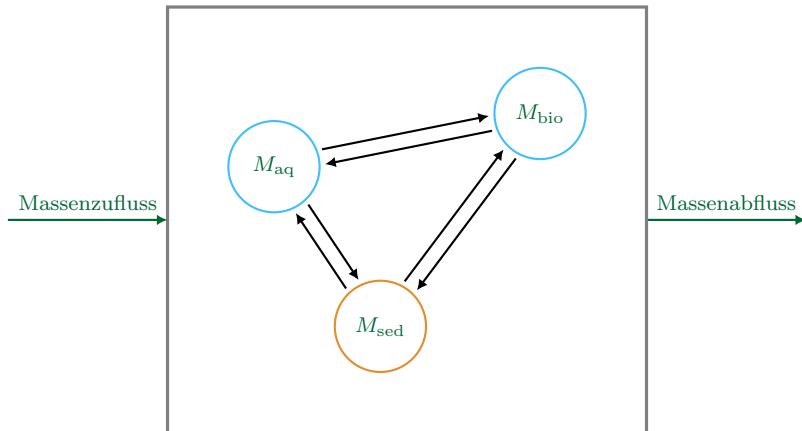


Abbildung 1.11: Schematische Darstellung des Systems See als Boxmodell.

Es ist nützlich, sich die Masseinheiten zu notieren. Sie helfen bei der Analyse und den weiteren Überlegungen. Nur mathematische Beziehungen, bei denen auch die Masseinheiten konsistent sind, können richtig sein. In unserem Beispiel haben wir:

- Masseninput: Masse an Phosphor im Zufluss [kg sec^{-1}]
- Massenabfluss: Masse an Phosphor im Abfluss [kg sec^{-1}]
- M_{aq} : Masse an Phosphor im Wasser [kg]
- M_{bio} : Masse an Phosphor in der Biomasse [kg]
- M_{sed} : Masse an Phosphor im Sediment [kg]

Diese Größen hängen im Allgemeinen von der Zeit t und den Anfangsbedingungen ab. Statt mit den Massen rechnet man gern mit den Konzentrationen, da diese einfacher zu messen sind. Über die Relationen (Pfeile) sind Modellannahmen zu treffen. Das werden wir weiter unten tun.

Verschiedene Boxmodelle können miteinander kombiniert werden, um auf diese Weise ein komplexeres System abzubilden. Dabei ist der Output eines Systems der Input eines anderen Systems. In unserem Beispiel könnte man dann eine Seenkette modellieren, wobei der Ausfluss des einen Sees in den Zufluss eines anderen Sees mündet. Um Ende, nach der Analyse, lassen sich dann beispielsweise Grenzwerte für die Phosphatkonzentration im Zufluss berechnen, damit der See gesund bleibt.

1.3 Beispiele von Modellen

1.3.1 Das Ohmsche Gesetz

Das Ohmsche Gesetz (Georg Simon Ohm, 1826) postuliert einen proportionalen Zusammenhang zwischen der angelegten Spannung über einem Widerstand und der resultierenden Stromstärke. Es gilt für eine grosse Zahl von Widerständen respektive Materialien und innerhalb gewisser Grenzen (es versagt aber beispielsweise bei sehr hohen Spannungen):

$$U = R \cdot I$$

Dabei bezeichnet U die Spannung über einem Widerstand [Volt], R den Widerstand [Ohm], und I die Stromstärke [Ampère]. Das Gesetz wurde gefunden, indem verschiedene Messwerte für Stromstärke und Spannung in ein Koordinatensystem eingetragen wurden. Die Messpunkte liegen auf einer Geraden durch den Ursprung, deren Steigung gerade die Proportionalitätskonstante, eben den Widerstand, darstellt.

Auch der menschliche Körper ist näherungsweise ein Ohmscher Widerstand. Berechnungen mit Hilfe des Ohmschen Gesetzes sind daher etwa bei Stromunfällen und Reiztherapien relevant. Das Ohmsche Gesetz bildet auch eine Grundlage für das Elektrokardiogramm und für die Elektroenzephalografie.

1.3.2 Das Sonnensystem

Am Sonnensystem kann man besonders gut die Entwicklung von Theorien verfolgen. Auf einfachen Beobachtungen (also auf Phänomenologie) basierend, entstanden im alten Griechenland die ersten schriftlich überlieferten Theorien über das Sonnensystem:

- Aristarchos von Samos (3. Jahrhundert vor Christus) postulierte ein heliozentrisches Weltbild.
- Ptolemäus (2. Jahrhundert vor Christus) setzte die Erde in den Mittelpunkt des Universums, vertrat also ein geozentrisches Weltbild.

Das geozentrische Weltbild wurde in der Folge ausgebaut und in der sogenannten Epizykeltheorie immer mehr verfeinert, um die genauer werdenden Messungen zu erklären. Die Theorie war jedoch ausserordentlich kompliziert und beantwortete die Frage nach dem *Warum* in keiner Weise. Schliesslich kam es mit Nikolaus Kopernikus (1473–1543) zum *Paradigmenwechsel*. Kopernikus stellte die Sonne in den Mittelpunkt seines Weltbildes und konnte so mit einem Schlag den Lauf der Gestirne viel einfacher erklären. Dieser Theoriesprung erfolgte also aufgrund des Strebens nach Vereinfachung.

- Nikolaus Kopernikus (Anfang des 16. Jahrhunderts) etablierte das heliozentrische Weltbild.

Die Theorie des Kopernikus fand anfangs nur wenige Anhänger, zumal sie Aussagen der Bibel zu widersprechen schien, was sie in den Augen der Kirche als ketzerisch erscheinen liess. In der Folge wurden jedoch immer genauere Beobachtungen gemacht, die mit der Epizykeltheorie nur noch durch Hinzunahme immer weiterer Epizyklen erklärbar gewesen wären.

- Tycho Brahe (Ende des 16. Jahrhunderts): Immer präzise Messungen wecken wachsende Zweifel am geozentrischen Weltbild.

Es war dann das Verdienst von Johannes Kepler, dem heliozentrischen Weltbild seine endgültige Form zu geben. Sein Modell basierte auf den Daten von Tycho Brahe. Sein Ansatz war zwar mathematisch, jedoch nur beschreibender Natur: Er erklärte nicht den Grund für die beobachteten Phänomene.

- Johannes Kepler (Anfang des 17. Jahrhunderts) formulierte die drei Keplerschen Gesetze.
 1. Keplersches Gesetz: Die Planeten bewegen sich auf elliptischen Bahnen, in deren einem gemeinsamen Brennpunkt die Sonne steht.
 2. Keplersches Gesetz: Eine von der Sonne zum Planeten gezogene Strecke überstreicht in gleichen Zeiten gleich grosse Flächen.
 3. Keplersches Gesetz: Die Quadrate der Umlaufzeiten zweier Planeten verhalten sich wie die dritten Potenzen der grossen Bahnhalbachsen.

Ein mathematisch-physikalischen Modell, welches auch erklären konnte, warum es ausgerechnet zu dieser Form der Keplerschen Gesetze kommt, entstand, als Isaac Newton das Gesetz der Gravitation lüftete:

- Isaac Newton (Ende des 17. Jahrhunderts): Keplers Gesetze folgen mathematisch aus dem Gravitationsgesetz. Zwei Massen m_1 und m_2 im Abstand r ziehen sich mit einer Kraft F an, die proportional zu beiden Massen und umgekehrt proportional zum Quadrat des Abstands ist. Also

$$F = G \frac{m_1 m_2}{r^2}.$$

Der Proportionalitätsfaktor G heisst Gravitationskonstante und kann durch Beobachtung und Messung bestimmt werden. Freilich erklärt auch die Theorie von Newton nicht, warum das Gravitationsgesetz so und nicht anders lautet.

Newton's Mechanik stellte sich später als eine Näherung der relativistischen Mechanik heraus. Tatsächlich konnte die Newtonsche Theorie auch nicht alle astronomischen Beobachtungen erklären. Unter dem Einfluss der übrigen Planeten verläuft die Bahn des Planeten Merkur nicht exakt auf einer Ellipse, sondern das Perihel (der sonnennächste Punkt) unterliegt einer langsamen Drehbewegung. Diese Periheldrehung erfolgt jedoch langsamer, als die Newtonsche Mechanik es voraussagt.

- Albert Einstein (Anfang des 20. Jahrhunderts): Die Periheldrehung des Merkur wird durch die allgemeine Relativitätstheorie erklärt.

1.3.3 Geometrische Optik

Am Beispiel der geometrischen Optik wird deutlich, wie das Streben nach Vereinheitlichung zur Triebfedern der Modellbildung werden kann. Dieser Ansatz führte im Fall der Optik schliesslich sogar zu einer Erweiterung der Theorie.

Bereits Euklid formulierte in seinen *Elementen*, in denen er das Wissen seiner Zeit zusammenfasste, im Buch über die Optik seine drei Gesetze zur Lichtausbreitung:

- Licht breitet sich gradlinig aus.
- Bei der Reflexion an einem Spiegel sind Einfall- und Ausfallwinkel des Lichtstrahls gleich gross.
- Bei der Reflexion an einem Spiegel liegen der einfallende Lichtstrahl, der ausfallende Lichtstrahl und die Normale zur Spiegelebene in einer Ebene¹.

Tatsächlich folgen aber *alle drei* Gesetze schon aus einem *einzigsten Prinzip*!

Prinzip 1.3.1 (Prinzip von Heron). *Licht bewegt sich immer entlang von kürzesten Bahnen.*

Insbesondere folgt daraus das zweite Gesetz Euklids, wie die Abbildung 1.12 zeigt. Ein Lichtstrahl bewegt sich vom Punkt A zum Punkt B , wobei er an einem Spiegel (dargestellt durch die waagerechte Gerade) reflektiert wird. Nach Heron erfolgt die Reflexion in demjenigen Punkt X , für den der Lichtweg kürzestmöglich ist. Da der Streckenzug von A über X nach B gleich lang ist, wie der Streckenzug von A über X zum gespiegelten Punkt B' , wird das Minimum durch die Gerade AB' realisiert (der kürzeste Weg zwischen zwei Punkten ist die Gerade). Für den so gefundenen Punkt X stimmen in der Tat Einfall- und Ausfallwinkel überein.

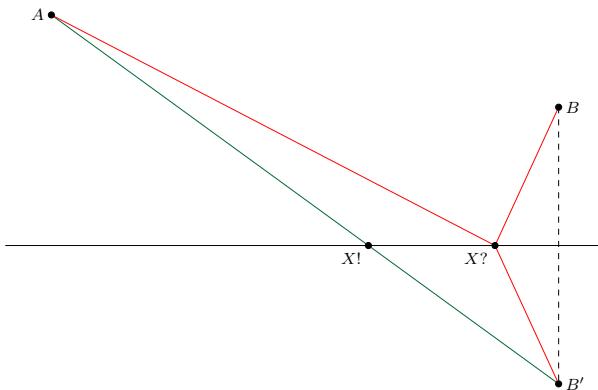


Abbildung 1.12: Herleitung des Reflexionsgesetzes aus dem Prinzip von Heron.

Das Phänomen der Lichtbrechung wird durch das Prinzip von Heron jedoch nicht erklärt. Pierre de Fermat modifizierte deshalb das Heronsche Prinzip wie folgt:

Prinzip 1.3.2 (Prinzip von Fermat). *Licht wählt Wege mit möglichst kurzer Durchlaufzeit².*

Das heisst, Licht wählt nicht den geometrisch kürzesten Weg, sondern den Weg kürzester Reisezeit. Wäre die Lichtgeschwindigkeit in jedem Medium gleich, wäre diese Unterscheidung nicht

¹ Anders ausgedrückt: Bei der Reflexion stimmen Einfalls- und Ausfallsebene überein: Die Einfallsebene wird gebildet aus dem einfallenden Lichtstrahl und der Normalen der Spiegelebene im Reflexionspunkt. Die Ausfallsebene ist entsprechend aufgespannt vom ausfallenden Lichtstrahl und der Normalen der Spiegelebene.

² Sowohl beim Prinzip von Heron, als auch beim Prinzip von Fermat handelt es sich genau genommen nicht immer um ein Minimum. Etwa bei der Reflexion in einem Hohlspiegel nimmt das Licht unter Umständen einen Weg maximaler Länge. Allgemein ist die Weglänge respektive die Reisezeit des Lichts ein Extremum (Minimum, Maximum oder Sattelpunkt).

relevant. Licht bewegt sich jedoch zum Beispiel in Glas oder in Wasser langsamer als im Vakuum oder in Luft. Um den Effekt zu illustrieren stellen wir uns vor, im Punkt A in der obigen Figur befindet sich ein Rettungsschwimmer, der einem in Not geratenen Schwimmer im Punkt B' zu Hilfe kommen will. Die horizontale Gerade stellt die Küstenlinie dar. Obwohl der direkte, grüne Weg der kürzeste ist, wird der Rettungsschwimmer den roten Weg wählen, da er auf dem Strand schneller laufen als im Wasser schwimmen kann!

Das Prinzip von Fermat liefert dann mit einem Schlag sowohl die drei Gesetze Euklids, als auch das Gesetz der Lichtbrechung von Snellius (siehe Übungen). Es ist benannt nach Willebrord van Roijen Snell.

Die mathematische Formulierung des Prinzips von Fermat,

$$|\nabla u| = \frac{1}{c},$$

heisst Eikonalgleichung (u bezeichnet die Laufzeit, c die ortsabhängige Lichtgeschwindigkeit) und regiert die geometrische Optik. Der Lichtweg im Auge wird durch die Eikonalgleichung bestimmt. Auf dieser Gleichung beruhen die Methoden der refraktiven Chirurgie am Auge (Laserkorrekturen, Implantate).

1.3.4 Algenwachstum und Nitratkonzentration

In der Limnologie, der Wissenschaft der Binnengewässer (insbesondere der Seen), sind Nitrat- und Phosphatkonzentration eines Sees wichtige Indikatoren für dessen Gesundheit. Bei intensiver Landwirtschaft werden die ausgebrachten Düngemittel in die Gewässer gespült, was zu hohen Nährstoffkonzentrationen in den Seen führen kann. Bei hohen Nitrat- und Phosphatkonzentrationen gedeihen Algen und Cyanobakterien besonders gut. Im Extremfall führt eine Überdüngung zu einer so hohen Konzentration von Algen, dass bei deren natürlichem Absterben das Ökosystem aus dem Gleichgewicht gerät: Der mikrobielle Abbau der abgestorbenen Algen geschieht unter Sauerstoffverbrauch, so dass schliesslich dem Wasser der gesamte gelöste Sauerstoff entzogen wird. Damit wird allen aeroben Lebewesen im Wasser die Lebensgrundlage entzogen, der See „kippt um“. Wir wollen nun den Zusammenhang zwischen Nährstoffkonzentration und Algenwachstum modellieren. Dabei verwenden wir das Boxmodell aus dem Abschnitt 1.2.1 wobei wir uns auf Nitrat anstelle von Phosphat als Nährstoff beschränken.

Sei $X(t)$ die Algenkonzentration in einem See. Bei *konstanter* Nitratkonzentration wächst die Algenkonzentration laut einer Messreihe gemäss

$$X'(t) = \lambda X(t). \quad (1.1)$$

X' bezeichnet hierbei die Ableitung von X nach der Zeit t . Die Formel (1.1) ist plausibel: Je höher die Algenkonzentration, desto schneller ist die Zunahme dieser Konzentration. Der Wachstumskoeffizient $\lambda = \lambda(C)$ hängt in zunächst unbekannter Weise von der Nitratkonzentration C ab. Das heisst es gilt

$$\lambda(C) = \frac{X'(t)}{X(t)} = (\ln X(t))' =: W.$$

Um $\lambda(C)$ zu bestimmen, wurden über einen gewissen Zeitraum folgende Werte gemessen:

- Die Nitratkonzentration C im See,

- das spezifische Wachstum W der Algenpopulation im See.

Bezogen auf eine willkürlich gewählte Referenzkonzentration C_0 ergab sich die in Abbildung 1.13 dargestellte Verteilung der Messwerte.

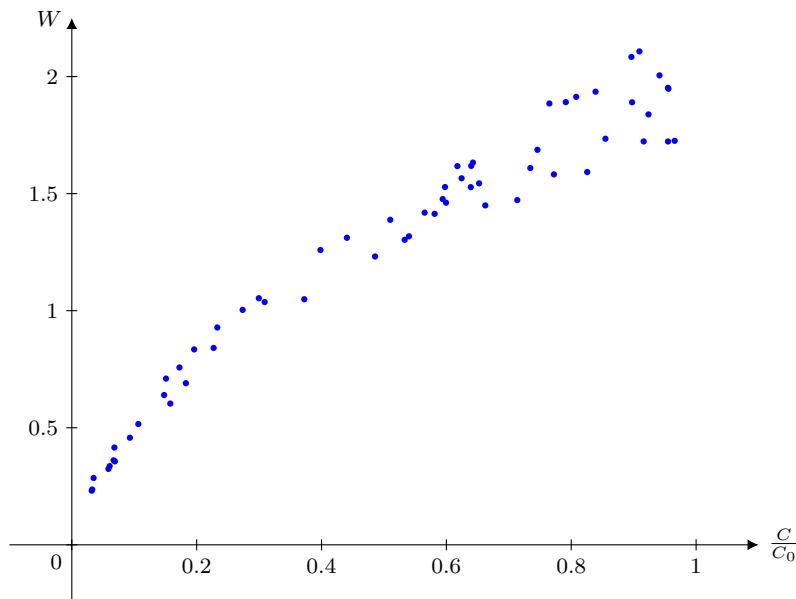


Abbildung 1.13: Darstellung der Messpunkte im Koordinatenystem: normierte Nitratkonzentration $\frac{C}{C_0}$ auf der horizontalen Achse, spezifisches Wachstum W auf der vertikalen Achse.

Für den Zusammenhang zwischen der normierten Nitratkonzentration und dem spezifischen Wachstum vermutet man aufgrund der Form der Punktwolke ein Potenzgesetz der Art

$$W = W_0 \left(\frac{C}{C_0} \right)^\alpha \quad (1.2)$$

mit zunächst unbekannten Werten W_0 und α . Durch Logarithmieren gelangt man zu

$$\log W = \log W_0 + \alpha \log \frac{C}{C_0}.$$

Die logarithmierten Daten sind in Abbildung 1.14 dargestellt.

Für die grüne Regressionsgerade findet man den Achsenabschnitt $\log W_0 = 0.7$ und die Steigung $\alpha = 0.61$.

Wir haben also auf diese Weise die Modellgleichung (1.2) für das spezifische Algenwachstum in Abhängigkeit der Nitratkonzentration erhalten!

1.3.5 Ein See als linearer Durchflussreaktor

Das in diesem Abschnitt beschriebene Beispiel ist typisch für lineare Modelle mit einer einzigen Systemvariablen, nämlich der Konzentration eines Stoffes, der vom Zufluss in den See

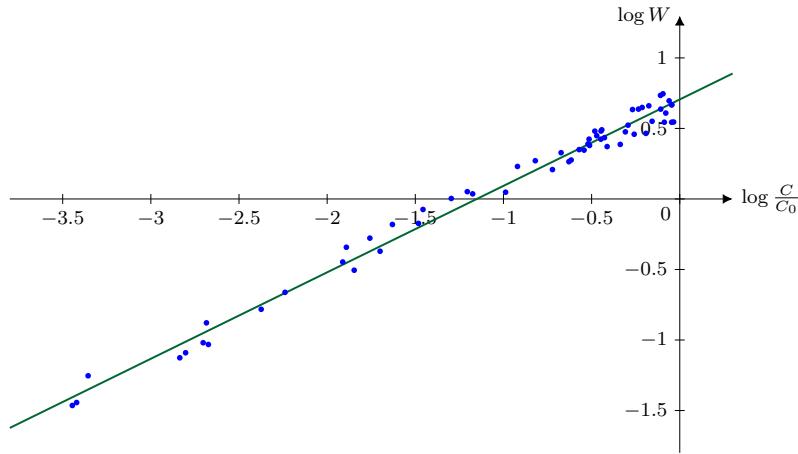


Abbildung 1.14: Darstellung der logarithmierten Daten.

eingetragen, im See mit einer bestimmten Rate abgebaut und schliesslich durch den Ausfluss ausgeschwemmt wird. Wir betrachten also einen See mit einem Zu- und einem Abfluss, wobei das Seewasser als jederzeit vollständig durchmischt angenommen wird. Durch den Zufluss wird nun also ein Stoff (zum Beispiel Nitrat oder Phosphat) in den See transportiert. Wir wählen folgende Bezeichnungen:

V = Volumen des Sees [m^3] (konstant)

$Q = Q(t)$ = Zufluss = Abfluss [m^3s^{-1}] (Wassermenge pro Zeiteinheit)

$C_{\text{in}} = C_{\text{in}}(t)$ = Konzentration eines Stoffs im Zufluss [kg/m^3]

$C = C(t)$ = Konzentration des Stoffs im Seewasser [kg/m^3]

k = Abbaurate des Stoffs im Seewasser [sec^{-1}] (konstant)

Die Masse $M(t)$ des Stoffes im See zur Zeit t ist somit gegeben durch $M(t) = VC(t)$. Um zu einem Modell zu kommen stellen wir die *Massenbilanz* des Stoffs im See auf. Dazu betrachten wir ein kurzes Zeitintervall $[t, t + \Delta t]$. Zur Zeit t befindet sich eine bestimmte Masse $M(t)$ des Stoffes im See, zur Zeit $t + \Delta t$ ist es die Masse $M(t + \Delta t)$. Die Änderung der Masse $M(t + \Delta t) - M(t)$ kommt auf drei Arten zustande: Durch Zufluss, durch Abbau im See und durch Abfluss des Stoffes im betrachteten Zeitintervall. Also gilt für dieses Zeitintervall:

Änderung der Stoffmenge im See =

$$= \text{Zufuhr des Stoffs} - \text{Abfuhr des Stoffs} - \text{Abbau des Stoffs} \quad (1.3)$$

Die einzelnen Terme in (1.3) lassen sich nun leicht angeben, wenn wir annehmen, dass im betrachteten kurzen Zeitintervall die Konzentrationen näherungsweise kontant bleiben:

- Änderung der Stoffmenge im See = $M(t + \Delta t) - M(t) = V(C(t + \Delta t) - C(t))$.
- Zufuhr des Stoffs = $C_{\text{in}}Q(t)\Delta t$.

- Abfuhr des Stoffs = $C(t)Q(t)\Delta t$.
- Abbau des Stoffs = $kM(t)\Delta t = kVC(t)\Delta t$.

Die Gleichung (1.3) lautet demnach

$$V(C(t + \Delta t) - C(t)) = QC_{\text{in}}(t)\Delta t - C(t)Q(t)\Delta t - kVC(t)\Delta t.$$

Wir dividieren nun diese Gleichung durch Δt und lassen dann Δt gegen 0 streben. Dabei verwandelt sich auf der linken Seite der Differenzenquotient in die Ableitung:

$$\lim_{\Delta t \rightarrow 0} \frac{C(t + \Delta t) - C(t)}{\Delta t} = C'(t).$$

Im Limes $\Delta t \rightarrow 0$ erhalten wir also die Systemgleichung

$$VC'(t) = QC_{\text{in}} - QC - kVC. \quad (1.4)$$

Diese Gleichung können wir auch wie folgt interpretieren: $VC'(t)$ ist die momentane Änderungsrate der Phosphatmasse im Seewasser. QC_{in} ist die momentane Zuflussrate, QC die momentane Abflussrate und kVC die momentane Abbaurate des Phosphats. Wir merken uns für die Zukunft, dass wir statt die Massenbilanz auch gleich die *Bilanz der momentanen Änderungsraten* aufstellen können:

$$\begin{aligned} & \text{momentane Änderungsrate der Masse} = \\ & = \text{momentane Zuflussrate} - \text{momentane Abflussrate} - \text{momentane Abbaurate} \end{aligned}$$

Dividiert man schliesslich die Gleichung (1.4) durch das konstante Volumen V erhält man

$$C' = \frac{Q}{V}C_{\text{in}} - C \left(\frac{Q}{V} + k \right). \quad (1.5)$$

Wir nehmen nun an, alle auftretenden Koeffizienten (also auch Q und C_{in}) seien konstant. Dann ist die obige Systemgleichung eine lineare inhomogene Differentialgleichung für $C(t)$ mit konstanten Koeffizienten. Die allgemeine Lösung ist demnach die Summe einer *partikulären Lösung* und der *allgemeinen Lösung der zugehörigen homogenen Gleichung*. Eine partikuläre Lösung ist zum Beispiel die stationäre (d. h. die zeitlich konstante) Lösung, die wir mit C_∞ bezeichnen. Setzen wir den Ansatz $C(t) = C_\infty = \text{konstant}$ (also insbesondere $C'_\infty = 0$) in (1.5) ein, so können wir nach C_∞ auflösen und finden:

$$C_\infty = \frac{QC_{\text{in}}}{Q + kV}.$$

Die zu (1.5) gehörende homogene Gleichung lautet

$$C' = -C \left(\frac{Q}{V} + k \right) \quad (1.6)$$

und hat die allgemeine Lösung

$$C_{\text{hom}}(t) := \lambda e^{-t(k+Q/V)},$$

für beliebiges $\lambda \in \mathbb{R}$. Somit lautet die allgemeine Lösung von (1.5)

$$C(t) = C_\infty + C_{\text{hom}}(t) = C_\infty + \lambda e^{-t(k+Q/V)}.$$

Verlangen wir zur Zeit $t = 0$ den Anfangswert $C(0) = C_0$, so ist λ bestimmt und wir finden

$$C(t) = \textcolor{red}{C}_\infty + (\textcolor{blue}{C}_0 - \textcolor{red}{C}_\infty) e^{-t(k+Q/V)}.$$

$C_{\text{hom}}(t) = C(t) - \textcolor{red}{C}_\infty$ löst die homogene Gleichung (1.6) und klingt in der *Halbwertszeit*

$$T_{1/2} = \frac{\ln 2}{k + Q/V}$$

jeweils um die Hälfte ab. Das heisst, nach Ablauf dieser Zeit $T_{1/2}$ verringert sich die Differenz von $C(t)$ zum stationären Zustand C_∞ jeweils um die Hälfte.

Statt der Halbwertszeit wird oft die *Anpassungszeit* als Referenz gebraucht, also die Zeit, um auf 5% = 1/20 abzuklingen. Die übliche kurze Rechung liefert

$$T_{1/20} = \frac{\ln 20}{k + Q/V}.$$

Für die numerische Daten wählen wir nun ein Gewässer von der Grösse des Hallwiler Sees, der in den 1980er Jahren umzukippen drohte und seit 1985 künstlich belüftet werden muss:

- Volumen des Hallwiler Sees: $V = 2.15 \times 10^8 \text{ m}^3$
- Phosphatbelastung des Hallwiler Sees im Jahr 1985: $C_0 = 2.5 \times 10^{-4} \text{ kg/m}^3$
- Der Hallwiler See wird vom Aabach gespeist und entwässert. Er führt im Mittel etwa 1.4 m^3 Wasser pro Sekunde, also $Q = 1.4 \text{ m}^3/\text{sec}$.
- Heutige Phosphatbelastung des Aabachs im Mittel: $C_{\text{in}} = 4 \times 10^{-5} \text{ kg/m}^3$. In der Vergangenheit war die Phosphatbelastung höher.
- Geschätzte Abbaurate des Phosphats im Hallwiler See: $k = 3 \times 10^{-9} \text{ sec}^{-1}$

Mit diesen Werten ergibt sich eine Konzentration $C_\infty \approx 2.7 \times 10^{-5} \text{ kg/m}^3$, die gut übereinstimmt mit dem langfristigen Ziel von $3 \times 10^{-5} \text{ kg/m}^3$ Phosphat im Hallwiler See. Als Halbwertszeit ergibt sich ein Wert von 2.3 Jahren, in guter Übereinstimmung mit den realen Messungen im See.

Die Abbildung 1.15 zeigt den Verlauf der Lösung für den oben erwähnten Anfangswert $C_0 = 2.5 \times 10^{-4} \text{ kg/m}^3$. Da $C_0 > C_\infty$ fällt die Lösung $\textcolor{green}{C}(t)$ exponentiell gegen den Wert C_∞ ab. Die gestrichelt eingetragene Lösung entspringt einem Anfangswert $C_0 < C_\infty$ und wächst exponentiell gegen den Wert C_∞ .

1.4 Stoffaustausch an Grenzflächen

Es wird geschätzt, dass rund 600 Millionen Menschen weltweit an Eisenmangelanämie leiden. Vor allem in Entwicklungsländern fehlt es an Geld und Medikamenten, um Patienten zu behandeln, und eisenhaltige Nahrungsmittel wie rotes Fleisch, Hülsenfrüchte und Nahrungsmittelzusätze sind zu teuer. Der Biomedizin-Doktorand der kanadischen Universität von Guelph, Christopher Charles, kam daher bei seiner Arbeit in Kambodscha auf die folgende Idee: Da dort eine bestimmte Fischart, der „try kantrop“, als Glücksbringer gilt, liess er etwa 200 Gramm schwere Fische aus Eisen herstellen und verteilte sie in der Bevölkerung. Der Fisch wird dann einfach

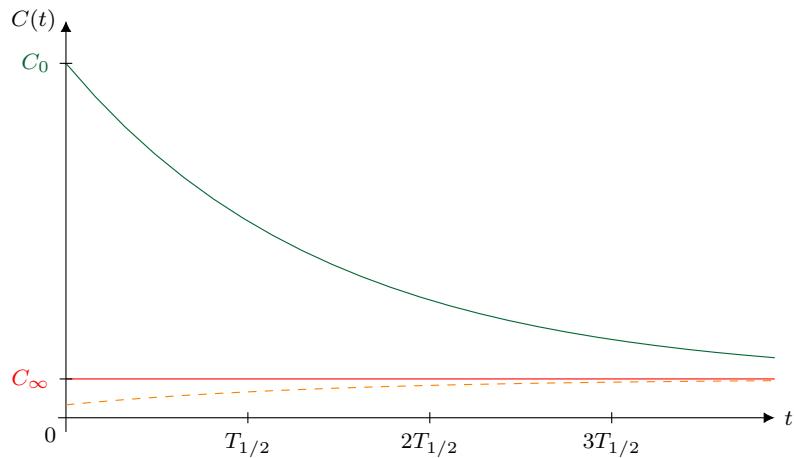
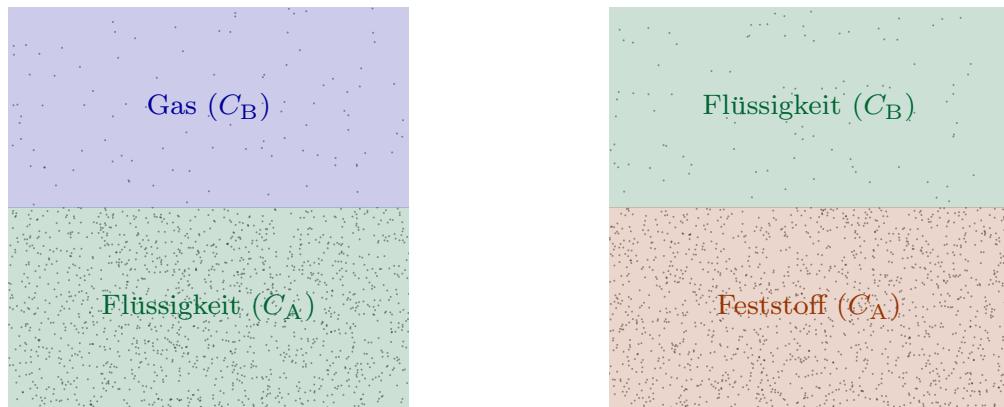


Abbildung 1.15: Die Entwicklung der Phosphatkonzentration $C(t)$ im Hallwiler See.

während der Zubereitung von Suppe und Reis in den Kochtopf gelegt und gibt dort Spuren von Eisen an die Speisen ab. Auf diese Weise wird dem Eisenmangel begegnet. Inzwischen wird der Fisch auch kommerziell vom Unternehmen *Lucky Iron Fish* in grossen Stückzahlen produziert. Wie gross muss die Oberfläche des Fisches sein, damit nicht zuviel und nicht zuwenig Eisen ins Kochwasser abgegeben wird? In diesem Abschnitt gehen wir ganz allgemein der Frage nach, wie sich der Stoffaustausch an einer Grenzfläche quantitativ erfassen lässt.

Von einer Grenzfläche spricht man allgemein, wenn zwei unterschiedliche Medien A und B aneinander stoßen. Das Bild unten zeigt schematisch eine solche Situation, wenn eine Gasphase auf eine flüssige Phase trifft (links) oder wenn eine flüssige Phase auf eine feste Phase trifft (rechts). Interessante Effekte ergeben sich dabei, wenn in beiden Medien ein Stoff S mit (unterschiedlicher) Konzentration C_A und C_B gelöst ist. Im *Gleichgewicht* (welches sich mit der Zeit einstellt, wenn das System in Ruhe gelassen wird) findet man durch Messung, dass sich die beiden Konzentrationen proportional zueinander verhalten.



Gesetz 1.4.1 (Gesetz von Henry im Gleichgewichtszustand). *Ein Stoff S sei in einem Zweipha-*

sensystem A, B gelöst. Im Gleichgewicht, bei niedriger Konzentration gilt:

$$\frac{C_A}{C_B} = K_{A/B}.$$

Die Proportionalitätskonstante $K_{A/B}$ heisst Henry-Koeffizient und hängt von den beiden Stoffen A und B ab.

Dieses Gesetz geht auf den englischen Chemiker William Henry zurück, der es 1803 publizierte.

Im Gleichgewicht ändern sich die Konzentrationen C_A und C_B nicht. Das bedeutet nicht, dass kein Stoffaustausch zwischen den beiden Meiden A und B stattfindet, sondern, dass im Durchschnitt pro Zeiteinheit die gleiche Menge des Stoffs S von A nach B wandert, wie in der umgekehrten Richtung. Der Massenfluss ist netto also gleich Null.

Befindet sich das System *ausserhalb des Gleichgewichts*, versucht dieses, durch Austausch des Stoffs S über die Grenzfläche, den Gleichgewichtszustand zu erreichen. Dabei gilt folgendes.

Gesetz 1.4.2 (Gesetz von Henry ausserhalb des Gleichgewichtszustands). *Ist das Zweiphasensystem A, B mit dem darin gelösten Stoff S nicht im Gleichgewicht, ist der Stoffaustausch an der Grenzfläche netto gegeben durch*

$$F_{A/B} = v_{A/B}(C_A - C_A^{eq}).$$

Dabei ist

$F_{A/B}$ = Massenfluss netto von A nach B pro Flächen- und Zeiteinheit

$v_{A/B}$ = Austauschgeschwindigkeit

$C_A^{eq} := C_B K_{A/B}$ = Konzentration des Stoffes S in A, die mit Konzentration C_B in Gleichgewicht wäre.

Um das Gesetz besser zu verstehen, sind die folgenden Bemerkungen nützlich:

- Die Austauschgeschwindigkeit $v_{A/B}$ ist eine Konstante, die von den Medien A und B sowie den Eigenschaften der Grenzfläche abhängt.
- Ist $C_A = C_A^{eq}$, so ist das System im Gleichgewicht, also der Nettoaustausch des Stoffs S gleich Null.
- Ist $C_A > C_A^{eq}$, d. h. befindet sich im Vergleich zum Gleichgewicht zuviel Stoff S in der Phase A, so ist $F_{A/B}$ positiv und es fliesst netto mehr des Stoffs S von Phase A nach B als umgekehrt.
- Ist $C_A < C_A^{eq}$, d. h. befindet sich im Vergleich zum Gleichgewicht zuwenig Stoff S in der Phase A, so ist $F_{A/B}$ negativ und es fliesst netto mehr des Stoffs S von Phase B nach A als umgekehrt.

Mit Hilfe der beiden Gesetze von Henry lassen sich insbesondere physiologisch relevante Vorgänge modellieren, etwa den Gas austausch in der Lunge zwischen der Atemluft und dem Blut. Dieses Beispiel wird in den Übungen ausführlich behandelt. Die Grenzfläche wird in diesem Fall von den Alveolen (Lungenbläschen) gebildet. Im nächsten Abschnitt wird das Gesetz von Henry auf die Situation in einem See angewandt: Dort findet nämlich ein Stoffaustausch zwischen dem Seewasser und der Sedimentschicht am Seegrund statt.

1.5 Mehrdimensionale Modelle: Ein lineares Zweiboxmodell

Dieses Beispiel ist typisch für ein mehrdimensionales lineares System. Wir betrachten dabei einen See mit Zu- und Abfluss, sowie das Seesediment. Uns interessiert die Konzentration eines Stoffs im Seewasser respektive im Sediment. Wir vereinbaren folgende Bezeichnungen

V_1 = Volumen des Sees (konstant)

V_2 = Volumen des Sediments (konstant)

A = Fläche der Grenzschicht zwischen Seewasser und Sediment

$Q = Q(t)$ = Zufluss = Abfluss (Wassermenge pro Zeiteinheit)

$C_{\text{in}} = C_{\text{in}}(t)$ = Konzentration eines Stoffs im Zufluss

$C_1 = C_1(t)$ = Konzentration des Stoffs im Seewasser

$C_2 = C_2(t)$ = Konzentration des Stoffs im Sediment

Auch hier gelangen wir zur Systemgleichung, indem wir die *Bilanz der Änderungsraten* betrachten (siehe Abschnitt 1.3.5). Für den See sieht diese so aus:

$$\begin{aligned} \text{momentane Änderungsrate der Stoffmenge im Seewasser} &= \\ &= \text{momentane Zuflussrate} - \text{momentane Abflussrate} - \text{Abbaurate des Stoffs durch Sedimentation} \end{aligned}$$

Und für das Sediment stellt sich die Bilanz der Änderungsraten so dar:

$$\begin{aligned} \text{momentane Änderungsrate der Stoffmenge im Sediment} &= \\ &= \text{momentane Zufuhrrate des Stoffs durch Sedimentation} \end{aligned}$$

Wie im Abschnitt 1.3.5 gilt nun:

$V_1 C'_1(t)$ = momentane Änderungsrate der Stoffmenge im Seewasser

QC_{in} = momentane Zuflussrate des Stoffs im Seewasser

QC_1 = momentane Abflussrate des Stoffs aus dem Seewasser

$v_{1/2}(C_1 - C_1^{\text{eq}})A$ = momentane Abbaurate des Stoffs durch Sedimentation, wobei $C_1^{\text{eq}} = C_2 K_{1/2}$ (nach dem Gesetz von Henry 1.4.2). Diese Menge geht dem Seewasser verloren und gleichzeitig kommt sie dem Sediment zugute.

$V_2 C'_2(t)$ = momentane Änderungsrate der Stoffmenge im Sediment

Setzen wir alles in die Bilanzen der Änderungsraten ein, erhalten wir die Systemgleichungen. Die erste Gleichung gilt für den See und die zweite für das Sediment:

$$V_1 C'_1(t) = QC_{\text{in}} - QC_1 - v_{1/2}(C_1 - C_2 K_{1/2})A$$

$$V_2 C'_2(t) = v_{1/2}(C_1 - C_2 K_{1/2})A.$$

Mit

$$C := \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}$$

lässt sich dieses System von zwei Differentialgleichungen kompakt schreiben als

$$C' = MC + B, \quad (1.7)$$

wobei man die Matrix M und den Vektor B direkt aus den Systemgleichungen abliest:

$$M = \begin{pmatrix} -\frac{Q + Av_{1/2}}{V_1} & \frac{v_{1/2}K_{1/2}A}{V_1} \\ \frac{v_{1/2}A}{V_2} & -\frac{v_{1/2}K_{1/2}A}{V_2} \end{pmatrix}$$

und

$$B = \begin{pmatrix} \frac{QC_{\text{in}}}{V_1} \\ 0 \end{pmatrix}.$$

Sind alle Koeffizienten konstant und die Matrix M invertierbar, so erhält man durch den Ansatz $C(t) = C_\infty = \text{konstant}$ aus (1.7) sofort eine stationäre Lösung:

$$C_\infty = -M^{-1}B.$$

Die zu (1.7) gehörige homogene Gleichung lautet

$$C' = MC. \quad (1.8)$$

Die allgemeine Lösung von (1.8) kann mit Hilfe der Exponentialreihe (siehe Abschnitt 3.6) so geschrieben werden:

$$C_{\text{hom}}(t) = e^{\textcolor{red}{Mt}} \lambda = \sum_{k=0}^{\infty} \frac{M^k t^k}{k!} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix},$$

wobei λ ein beliebiger Vektor ist. Der (formale) Beweis erfolgt einfach durch Einsetzen. Somit lautet die allgemeine Lösung von (1.7)

$$C(t) = C_\infty + e^{\textcolor{red}{Mt}} \lambda.$$

Verlangt man zur Zeit $t = 0$ die Anfangsdaten $C(0) = C_0$, so wird λ dadurch festgelegt. Man findet als Lösung sofort

$$C(t) = C_\infty + e^{\textcolor{red}{Mt}} (C_0 - C_\infty).$$

Wie sich diese Lösung verhält, werden wir im Kapitel 3 untersuchen.

1.6 Die Pocken und das „Dreckige Dutzend“

1.6.1 Einige Fakten über die Pocken

Die Pocken oder Blattern waren während Jahrhunderten eine Geissel der Menschheit. Zahlreiche Pockenepidemien sind historisch belegt, sogar in der Bibel wird eine Krankheit mit allen Anzeichen einer Pockeninfektion beschrieben (Exodus, 2. Mose 9,9 bis 9,11). Bei den Pocken handelt

sich um eine Viruserkrankung mit der lateinischen Bezeichnung *Variola*. Nach einer Inkubationszeit von meist 12 bis 14 Tagen zeigen die Infizierten folgende Symptome: Fieber, Schüttelfrost und schliesslich starke Blasenbildung auf der Haut. Je nach Pockenstamm liegt die Letalität bei unbehandelten Pocken bei 10% bis 90%, im Durchschnitt wird von 30% berichtet. Patienten, welche die Krankheit überleben, sind gezeichnet durch teilweise entstellende Narben, die bei der Abheilung der Blasen entstanden sind. In schweren Fällen führt die Krankheit auch zu Erblindung, Gehörlosigkeit, Lähmungen, Lungenentzündung oder Hirnschäden. Menschen, die eine Pockenerkrankung überstanden haben, sind gegen die Krankheit lebenslang immun.

1.6.2 Impfungen

In China sind seit rund 3000 Jahren einfache Immunisierungen gegen Pocken bekannt. Auch in Indien sind ähnliche Techniken überliefert. Diese einfache Form der Impfung wird *Variolation* genannt und besteht darin, einen Gesunden mit einer abgeschwächten Form der Erreger zu infizieren. So wurde zerriebener Schorf von den Pusteln erkrankter Personen geschnupft oder in die Haut geritzt.

In Europa ist die Methode der Variolation durch Lady Mary Wortley Montagu bekannt geworden. Als Frau eines britischen Diplomaten, der in Istanbul stationiert war, hatte sie von der in der Türkei praktizierten Methode der Variolation gehört. Lady Mary hatte 1713 ihren Bruder durch die Pocken verloren und war 1715 selber daran erkrankt. Sie beschloss daher 1721, gegen den Rat ihrer englischen Ärzte, ihre Kinder durch Variolation impfen zu lassen. Zurück in England machte sie die Methode bekannt, stiess jedoch bei ihren Zeitgenossen so lange auf Unverständnis und Skepsis, bis sie den englischen König George I. überreden konnte, seine Enkel impfen zu lassen.

Auf einer Beobachtung, die durch ein Dokument aus dem Jahr 1765 belegt ist, beruht eine weitere Impfmethode. Man hatte nämlich festgestellt, dass Bauern nach einer Kuhpockeninfektion vor echten Pocken geschützt waren. Tatsächlich sind die Erreger der Kuhpocken und der Pocken verwandt, sodass die Antikörper, welche sich bei einer Kuhpockeninfektion bilden, auch gegen den Erreger der Pocken wirken.

Etwa ab 1771 wurde diese Methode in England und Niedersachsen erprobt. Ernsthaft untersucht und vorangetrieben wurde die Entwicklung der Pockenimpfung vom englischen Landarzt Edward Jenner, der 1796 seine Methode der *Vakzination*³ durch Vaccinia-Viren (Kuhpocken) breit einsetzte und den Impferfolg und die Impfrisiken dabei systematisch untersuchte. Seine Experimente würden heute üblichen ethischen Grundsätzen widersprechen. So infizierte er ein von ihm geimpftes Kind absichtlich mit Pocken - es überlebte. Jenners Methode wurde damals von der etablierten Wissenschaft nicht akzeptiert und lächerlich gemacht. Trotz persönlicher Anfeindungen setzte sich seine Idee schliesslich durch. 1979 erklärte die Weltgesundheitsorganisation WHO nach systematischen Impfkampagnen die Welt für pockenfrei. Somit hat Jenner im Grunde allen Pockenopfern, die es ohne seine Impfung noch gegeben hätte, das Leben gerettet und ist damit einer der bedeutendsten Wohltäter der Menschheitsgeschichte.

Die Ausrottung der Pocken gilt als Meilenstein der Medizingeschichte. Bei anderen Krankheiten



Abbildung 1.16:
Edward Jenner

³Auf Englisch und Französisch heisst heute Impfung ganz allgemein *vaccination*.

ist dies noch nicht gelungen. Immerhin könnte die Kinderlähmung bald der Vergangenheit angehören: Derzeit werden nur noch Fälle in Nigeria, Pakistan und Somalia registriert. Die WHO hat sich zum Ziel gesetzt, auch die Masern komplett zurückzudrängen. In Amerika und Skandinavien ist dies bereits gelungen. In andern Ländern ist hingegen die Impfquote noch nicht hoch genug. In den Industrieländern kommen auch andere Infektionskrankheiten wie Typhus, Pest, Milzbrand oder Diphtherie nur noch selten vor. Diese Krankheiten komplett auszurotten, gilt aber als fast unmöglich, da die Erreger – anders als bei Pocken oder Masern – nicht nur den Menschen als Wirt nutzen, sondern auch Tiere. Das ist auch das Problem bei den *Coronaviridae*, den Coronaviren, zu denen auch SARS-CoV-2 zählt.

Es wird vermutet, dass in den Biowaffenarsenalen mancher Staaten noch letzte Stämme des Pockenvirus gelagert werden. Dort zählen die Pocken zusammen mit Milzbrand, Anthrax, hämmorrhagischem Fieber und anderen schlimmen Seuchen zum *Dreckigen Dutzend* der international geächteten Biowaffen.

1.7 Daniel Bernoulli und die Pocken

Impfungen bringen gewisse Risiken mit sich. Mit einer gewissen Wahrscheinlichkeit führen Impfungen zu Komplikationen, die im Extremfall sogar tödlich enden können. Vom Standpunkt des Individuums aus betrachtet, wird daher eine Impfung als ein konkretes Risiko empfunden, welches bei der Entscheidung für eine Impfung bewusst eingegangen werden muss. Dieses Impfrisiko ist subjektiv nur schwer gegen das eher abstrakte Risiko einer möglichen künftigen Ansteckung mit der Krankheit (und deren Folgen) abzuschätzen⁴. Dieses Dilemma stellt sich nicht, wenn man statt das Individuum die Gesamtbevölkerung in den Fokus nimmt: Über die Gesamtbevölkerung sind statistische Aussagen darüber möglich, ob eine Impfung die mittlere Lebenserwartung erhöht, und wenn ja um wie viel.

Genau dieser Frage ging der berühmte Basler Mathematiker und Physiker Daniel Bernoulli (1700–1782) nach:

Bernoullis Grundfrage (1760): Kann trotz Impfrisiko die Lebenserwartung einer Population durch die Pockenimpfung erhöht werden?

Am 16. April 1760 stellte die *Académie Royale des Sciences de Paris* eine öffentliche Lesung eines Werkes von Daniel Bernoulli vor: seinen *Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, & des avantages de l'inoculation pour la prévenir* (siehe [bernoulli]). Darin wirft Bernoulli just die oben formulierte Frage auf.



Abbildung 1.17: Daniel Bernoulli

⁴Stellen Sie sich vor, jemand bietet Ihnen eine Spritze an, die Ihr Leben um 20 Jahre verlängert, an der Sie jedoch mit einer Wahrscheinlichkeit von 1:1000 sofort sterben. Nehmen Sie die Spritze an?

1.7.1 Daten aus Halley's Sterbetafel

Aber wie sollte Bernoulli diese Grundfrage beantworten können? Als erstes brauchte er gesicherte Daten darüber, wie lang die Menschen seiner Zeit lebten. Er fand eine sogenannte Sterbetafel⁵, die der Astronom Edmond Halley⁶ 1693 aufgestellt hatte. In dieser Sterbetafel verzeichnete Halley anhand von Geburts- und Sterberegistern für eine Population von 1000 Einjährigen in der Stadt Breslau (Schlesien, heute Polen), wie viele von ihnen nach n Jahren noch am Leben waren.

Age.	Per- sons												
	Curr.												
1	1000	8	580	15	623	22	585	29	539	35	481	7	5517
2	855	9	570	16	621	23	579	30	531	37	472	14	4564
3	798	10	551	17	615	24	573	31	523	38	453	21	4270
4	750	11	533	18	610	25	567	32	515	39	453	28	3964
5	732	12	546	19	604	26	560	33	507	40	445	35	3504
6	710	13	640	20	593	27	553	34	495	41	436	42	3178
7	652	14	634	21	562	28	546	35	490	42	427	49	2709
Age Cur-	Per- sons												
43	417	50	340	57	272	64	222	71	331	78	53	63	1574
44	407	51	335	58	262	65	192	72	120	79	49	70	1264
45	357	52	324	59	252	65	182	73	105	80	41	84	692
46	387	53	313	60	242	67	172	74	58	81	34	100	253
47	377	54	302	61	232	68	162	75	86	82	28	82	34000
48	367	55	292	62	222	69	152	76	78	83	23	Sum Total	
49	357	56	282	63	212	70	142	77	68	84	20		

Abbildung 1.18: Halley's Sterbetafel aus [halley].

Halley hatte Breslau ausgewählt, weil diese Stadt kaum Immigration und Emigration kannte und daher unverfälschte Daten erwartet werden konnten. An der Tafel kann man beispielsweise ableSEN, dass nur 855 der 1000 Kinder den zweiten Geburtstag erreichten. Immerhin 20 überlebten das 84. Altersjahr. In der Tabelle rechts summiert Halley die Anzahl der gelebten Jahre. Insgesamt haben die 20 Menschen, die das 84. Altersjahr überlebt hatten, zusammen noch 107 Jahre gelebt.

Daniel Bernoulli schätzte, dass ursprünglich 1'300 Neugeborene in Halley's Statistik vorkamen. (Halley hatte diese Zahl leider nicht vermerkt). Da total 34'000 Jahre gelebt wurden, ergibt das eine durchschnittliche Lebenserwartung von

$$\frac{34'000}{1300} \approx 26.15 \text{ Jahre} \approx 26 \text{ Jahre und knapp 2 Monate.}$$

Zum Vergleich betrachten wir die heutige mittlere Lebenserwartung.

- In der Schweiz: 84.4 Jahre bei Frauen, respektive 79.7 Jahre bei Männern.
 - In afrikanischen Staaten: zum Teil unter 35 Jahre bei Männern.

Bernoulli entwickelte zur Beantwortung seiner Frage das erste epidemiologische Modell der Geschichte.

⁵Solche Sterbetafeln sind heute in der Lebensversicherungsbranche eines der wichtigsten Hilfsmittel, um Prämien zu berechnen.

⁶Halley hat auch den nach ihm benannten Kometen entdeckt.

1.7.2 Daniel Bernoullis Modell

Daniel Bernoulli versuchte nun ein zu Halley's Daten passendes Modell einer Bevölkerungspopulation zu erstellen. Unter Population verstehen wir hier eine ausgewählte Gruppe von gleichaltrigen Individuen, die ab der Geburt über die Jahre hinweg beobachtet wird. Dazu traf Bernoulli folgende Modellannahmen. Die Zeiteinheit ist dabei das Jahr.

$$\xi(t) = \text{Anzahl Überlebende in der Population zur Zeit } t$$

$$s(t) = \text{Anzahl ansteckbare Individuen in der Population zur Zeit } t$$

$$\frac{1}{n} = \text{Wahrscheinlichkeit, in einer Zeiteinheit (Jahr) mit Pocken infiziert zu werden}$$

$$\frac{1}{m} = \text{Wahrscheinlichkeit, dass ein Infizierter an Pocken stirbt}$$

$$\frac{1}{N} = \text{Wahrscheinlichkeit, an der Impfung zu sterben}$$

Es wird zudem angenommen, dass die überstandene Krankheit zu lebenslanger Immunität führt. Konkret rechnete Bernoulli mit den Werten $n = m = 8$ und $N = 200$. Das heisst, im Schnitt würde etwa jeder 8. innerhalb eines Jahres an Pocken erkranken, und ebenfalls jeder 8. an Pocken Erkrankte würde der Krankheit erliegen. Auf der anderen Seite nimmt Bernoulli an, dass jeder 200. an den Folgen der Impfung stirbt.

Bernoullis Rechnung

Bernoulli berechnet nun im Rahmen seiner Modellannahmen nacheinander die Abnahme der Population aufgrund von Todesfällen durch Pocken respektive aus anderen Gründen: Offenbar gilt

$$\xi(t + \Delta t) - \xi(t) =: \Delta \xi = -\text{Anzahl Todesfälle im Zeitintervall } [t, t + \Delta t].$$

Man beachte dass $\Delta \xi < 0$ ist (daher kommt das **Minuszeichen**).

Von den zur Zeit t ansteckbaren $s(t)$ Personen werden nach Ablauf einer Zeiteinheit $\frac{s(t)}{n}$ neu Infizierte erwartet. Entsprechend gilt:

$$\frac{s(t) \cdot \Delta t}{n} = \text{Anzahl im Zeitintervall } [t, t + \Delta t] \text{ neu Angesteckte.} \quad (1.9)$$

Wir nehmen der Einfachheit halber an, dass eine Pockeninfektion entweder überlebt wird, oder instantan zum Tode führt. Von den oben berechneten im Zeitintervall $[t, t + \Delta t]$ Neuangesteckten werden $1/m$ sterben:

$$\frac{s(t) \cdot \Delta t}{n} \frac{1}{m} = \text{Anzahl Pockentote, die im Zeitintervall } [t, t + \Delta t] \text{ angesteckt wurden.}$$

An dieser Stelle können wir eine erste einfache Folgerung ziehen, nämlich:

$$-\Delta \xi - \frac{s(t) \Delta t}{mn} = \text{Anzahl Todesfälle aus anderen Gründen (als Pocken) im Zeitintervall } [t, t + \Delta t]. \quad (1.10)$$

Der Anteil der Ansteckbaren $s(t)$ in der Gesamtpopulation $\xi(t)$ ist $s(t)/\xi(t)$. Somit ist wegen (1.10) die Anzahl der Ansteckbaren, die in $[t, t + \Delta t]$ aus anderen Gründen sterben

$$\frac{s(t)}{\xi(t)} \left(-\Delta \xi - \frac{s(t) \Delta t}{mn} \right). \quad (1.11)$$

Weiter gilt

$$s(t + \Delta t) - s(t) =: \Delta s = - \text{Abnahme der Anzahl ansteckbarer Personen im Zeitintervall } [t, t + \Delta t].$$

(Man beachte wieder das Minuszeichen.) Diese Abnahme hat zwei Gründe:

- Infektion mit Pocken (mit oder ohne Todesfolge)
- Todesfälle aus anderen Gründen

Beides hatten wir weiter oben, in (1.9) und (1.11), berechnet. Also erhalten wir

$$-\Delta s = \frac{s(t)\Delta t}{n} - \frac{s(t)}{\xi(t)} \left(\Delta \xi + \frac{s(t)\Delta t}{mn} \right).$$

Nach Division durch Δt und dem Limes $\Delta t \rightarrow 0$ ergibt sich:

$$-s'(t) = \frac{s(t)}{n} - \frac{s(t)\xi'(t)}{\xi(t)} - \frac{s^2(t)}{\xi(t) \cdot mn}.$$

Multiplizieren wir dies mit $\xi(t)/s^2(t)$, so erhalten wir

$$-\frac{s'(t)\xi(t)}{s^2(t)} = \frac{\xi(t)}{ns(t)} - \frac{\xi'(t)}{s(t)} - \frac{1}{mn}.$$

Die beiden orangen Terme können wir zusammenfassen und als Ableitung eines Quotienten schreiben:

$$\left(\frac{\xi(t)}{s(t)} \right)' = \frac{s(t)\xi'(t) - s'(t)\xi(t)}{s^2(t)} = \frac{\xi(t)}{ns(t)} - \frac{1}{mn}.$$

Dies ist wunderbarerweise eine lineare Differentialgleichung für

$$q(t) := \frac{\xi(t)}{s(t)},$$

nämlich

$$q'(t) = \frac{q(t)}{n} - \frac{1}{mn}. \quad (1.12)$$

Eine stationäre Lösung von (1.12) ist offenbar die konstante Funktion

$$q_\infty \equiv \frac{1}{m}.$$

Die allgemeine Lösung der zu (1.12) gehörenden homogenen Gleichung lautet

$$q_{\text{hom}}(t) = C e^{t/n}.$$

Somit ist die allgemeine Lösung von (1.12)

$$q(t) = q_\infty + q_{\text{hom}}(t) = \frac{1}{m} + C e^{t/n}.$$

Nun können wir noch C aus der Anfangsbedingung bestimmen: Es gilt

$$q(0) = \frac{\xi(0)}{s(0)} = 1,$$

denn am Anfang sind alle Individuen der Population infizierbar. Auf diese Weise finden wir schliesslich

$$\frac{\xi(t)}{s(t)} = q(t) = \frac{1}{m} \left((m-1) e^{t/n} + 1 \right). \quad (1.13)$$

Somit ist die Anzahl infizierbarer Personen

$$s(t) = \xi(t) \cdot \frac{m}{(m-1)e^{t/n} + 1}.$$

Da $\xi(t)$ für die von Halley betrachtete Population aus dessen Tabelle bekannt ist, lassen sich daraus $s(t)$ und alle davon ableitbaren Grössen berechnen! Man beachte, dass der hellblaue Faktor für $t = 0$ den Wert 1 hat, und mit wachsendem t immer kleiner wird. Das muss auch so sein, denn der Anteil an immun gewordenen Personen wächst im Laufe der Zeit.

Wie wirkt sich Impfen auf die Entwicklung der Population aus?

Bisher haben wir in der betrachteten Population nur Individuen, die nach überstandener Krankheit immun wurden. Jetzt wollen wir diese Population vergleichen mit einer gleichartigen Population, die zu Anfang geimpft wurde.

Erinnern wir uns: Wir hatten weiter oben bei (1.10) berechnet dass

$$-\Delta\xi - \frac{s(t)\Delta t}{mn} = \text{Anzahl Todesfälle aus anderen Gründen (als Pocken) im Zeitintervall } [t, t + \Delta t].$$

Nun dividieren wir dies durch Δt , nehmen den Limes $\Delta t \rightarrow 0$ und dividieren gerade noch durch die momentane Populationsgrösse $\xi(t)$. Dies ergibt

$$\mu(t) := -\frac{1}{\xi(t)} \left(\xi'(t) + \frac{s(t)}{mn} \right) = \text{Todesrate aus anderen Gründen pro Person.}$$

Die Bedeutung von $\mu(t)$ ist also folgende. Wenn wir wissen wollen, wieviele Personen in einer Gruppe von p Personen in der Zeitspanne $[t, t + \Delta t]$ aus anderen Gründen als Pocken sterben, so ist diese Zahl gleich $p \Delta t \mu(t)$.

Aber wie kommen wir weiter? An dieser Stelle hatte Bernoulli eine geniale Idee. Er führte sozusagen virtuell eine Vergleichspopulation ein, in der keine Pocken vorkommen:

$$z(t) = \text{Anzahl einer Population, die keine Pocken kennt, zur Zeit } t.$$

Die Todesrate in dieser Population pro Person ist also gerade unser $\mu(t)$! Das bedeutet, dass

$$z'(t) = -\mu(t)z(t) = \frac{1}{\xi(t)} \left(\xi'(t) + \frac{s(t)}{mn} \right) \cdot z(t).$$

Wir dividieren nun beidseits flugs durch $z(t)$ und erhalten

$$(\ln z(t))' = \frac{z'(t)}{z(t)} = \frac{\xi'(t)}{\xi(t)} + \frac{s(t)}{\xi(t)mn} = (\ln \xi(t))' + \frac{s(t)}{\xi(t)mn}. \quad (1.14)$$

Den braunen Quotienten $q(t) = \xi(t)/s(t)$ kennen wir noch von (1.13): Setzen wir also $q(t)$ in (1.14) ein, so erhalten wir

$$\left(\ln \frac{z(t)}{\xi(t)} \right)' = \frac{s(t)}{\xi(t)mn} = \frac{\frac{1}{n}}{m q(t)},$$

wobei wir linkerhand noch das Logarithmengesetz für Quotienten verwendet haben. Schliesslich setzen wir für $q(t)$ noch die in (1.13) gefundene Lösung ein:

$$\left(\ln \frac{z(t)}{\xi(t)} \right)' = \frac{\frac{1}{n}}{(m-1)e^{t/n} + 1} = \left(\ln \frac{e^{t/n}}{(m-1)e^{t/n} + 1} \right)'.$$

Die Ableitungen beider Klammerausdrücke stimmen also überein, das heisst sie sind bis auf eine additive Konstante gleich:

$$\ln \frac{z(t)}{\xi(t)} = \ln \left(C \cdot \frac{e^{t/n}}{(m-1)e^{t/n} + 1} \right). \quad (1.15)$$

Um die Konstante C zu bestimmen, verwenden wir, dass für $t = 0$ nach Annahme die virtuelle pockenfreie Population gleich gross ist, wie die reale Population, das heisst $\xi(0) = z(0)$. In (1.15) gilt somit für $t = 0$, dass

$$\ln 1 = \ln(C \cdot \frac{1}{m}),$$

also $C = m$. Jetzt können wir (1.15) nach der gesuchten Populationsgrösse $z(t)$ auflösen und erhalten als Resultat

$$z(t) = \xi(t) \cdot \frac{m e^{t/n}}{(m-1)e^{t/n} + 1}. \quad (1.16)$$

Und wieder: Da $\xi(t)$ aus Halley's Tabelle bekannt ist, kann daraus nun auch $z(t)$ berechnet werden!

1.7.3 Bernoullis Folgerung

Bernoulli konnte aus den Resultaten seiner Rechnung nun die mittlere Lebenserwartung der realen Population mit der mittleren Lebenserwartung in der pockenfreien (geimpften) Population vergleichen.

Erinnern wir uns zunächst, dass wir angenommen haben, dass eine von N gleich zu Anfang geimpften Personen in der Folge an Komplikationen stirbt. Das heisst, wir erhalten mit (1.16):

$$\begin{aligned} Z(t) &:= \left(1 - \frac{1}{N}\right) z(t) \\ &= \left(1 - \frac{1}{N}\right) \xi(t) \frac{m e^{t/n}}{(m-1)e^{t/n} + 1} \end{aligned} \quad (1.17)$$

= Grösse einer Population zur Zeit t , die systematisch bei der Geburt geimpft wurde.

Indem Bernoulli mit Hilfe der Werte $\xi(t)$ aus Halley's Tabelle die gelebten Jahre aufsummierte und durch die Anzahl Individuen (also durch $\xi(0) = 1'300$) teilte, erhielt er die *Mittlere Lebenserwartung ohne Impfung*:

$$\frac{1}{\xi(0)} \cdot (\xi(1) + \xi(2) + \xi(3) + \dots) \approx 26 \text{ Jahre und 2 Monate.} \quad (1.18)$$

Mit Hilfe der Formel aus (1.17) und den Werten $m = n = 8$ und $N = 200$, sowie erneut mit den Werten von Halley, bestimmte er ebenso die *Mittlere Lebenserwartung mit Impfung*:

$$\frac{1}{Z(0)} \cdot (Z(1) + Z(2) + Z(3) + \dots) \approx 29 \text{ Jahre und 7 Monate.}$$

Die Erhöhung der Lebenserwartung von über 10% bewog Bernoulli, die Pockenimpfung zu empfehlen. Dieser Schlussfolgerung widersprach unter anderem der berühmte Mathematiker, Philosoph und Aufklärer Jean-Baptiste d'Alembert in einem Vortrag am 12. November 1760 vor der *Académie des Sciences de Paris* heftig (siehe [dalembert]). Selbst 1971 plädierten US-amerikanische Epidemiologen gegen die Pockenimpfung (siehe [bailey]), wenige Jahre vor der Ausrottung der Krankheit (durch Impfung). Die Impfdebatte hält bis heute an.

1.8 Das SIR-Modell

In diesem Abschnitt modellieren wir die *Ausbreitung* einer Krankheit durch Ansteckung. Dazu beobachten wir eine feste Population, in der sich die Krankheit ausbreitet und treffen zunächst folgende *Modellannahmen*:

- Die Krankheit ist von begrenzter Dauer.
- Jeder Kranke wird wieder gesund.
- Die überstandene Krankheit führt zur lebenslangen Immunität.
- Die Population ist isoliert (keine Zu- oder Abwanderung, keine Todesfälle).

Wir unterscheiden nun drei Gruppen innerhalb unserer Population:

$$S(t) = \text{Anzahl Ansteckbare zur Zeit } t \text{ (susceptible)}$$

$$I(t) = \text{Anzahl Infizierte zur Zeit } t \text{ (infected)}$$

$$R(t) = \text{Anzahl Geheilte zur Zeit } t \text{ (recovered)}$$

Jedes Individuum der Population gehört zu genau einer dieser Gruppen.

1.8.1 Ansteckung

Den Mechanismus der Ansteckung modellieren wir durch folgende Annahmen:

- Trifft eine kranke Person eine ansteckbare, wird letztere mit Wahrscheinlichkeit $\textcolor{red}{p}$ infiziert.
- Die Wahrscheinlichkeit eines solchen Treffens im Zeitintervall $[t, t + \Delta t]$ ist proportional (mit Proportionalitätsfaktor $\textcolor{red}{q}$) zu $S(t)$, $I(t)$ und Δt . Diese Annahme ist vernünftig: In einer doppelt so langen Zeitspanne können doppelt so viele Treffen erwartet werden. Ist die Zahl der Ansteckbaren (oder Infizierten) doppelt so hoch, verdoppelt sich auch die Wahrscheinlichkeit, dass ein Ansteckbarer einen Infizierten trifft.

Somit gilt für $\Delta S := S(t + \Delta t) - S(t)$,

$$\Delta S = -\textcolor{red}{p} \textcolor{red}{q} \cdot S(t) I(t) \Delta t.$$

Division durch Δt und Limes $\Delta t \rightarrow 0$ liefert eine erste Systemgleichung:

$$S'(t) = -\textcolor{red}{p} \textcolor{red}{q} \cdot S(t) I(t) = -\textcolor{red}{c} \cdot S(t) I(t), \quad (1.19)$$

wobei $\textcolor{red}{c} := \textcolor{red}{p} \textcolor{red}{q}$.

1.8.2 Heilung

Über den Krankheitsverlauf respektive die Heilung machen wir folgende Annahme:

- Die Krankheit dauert im Schnitt n Tage. Greift man zufällig einen Kranken heraus, ist dieser mit Wahrscheinlichkeit $\frac{1}{n}$ im i -ten Tag seiner Krankheit ($i = 1, 2, \dots, n$).

Insbesondere werden also jeden Tag im Schnitt $\frac{1}{n}$ der Kranken wieder gesund. Also gilt: Im Zeitintervall $[t, t + \Delta t]$ werden im Schnitt

$$\Delta R = R(t + \Delta t) - R(t) = \frac{1}{n} I(t) \Delta t$$

Kranke wieder gesund. Wir dividieren erneut durch Δt und betrachten den Limes $\Delta t \rightarrow 0$. Zudem wählen wir noch die Abkürzung $w := 1/n$. Der Parameter w heisst Heilungskoeffizient. Dies ergibt eine zweite Systemgleichung

$$R'(t) = w I(t). \quad (1.20)$$

Da die Summe $S(t) + I(t) + R(t)$ konstant ist, und damit die Ableitung gleich 0 ist, erhalten wir mit (1.19) und (1.20) für die Ableitung der Infizierbaren die dritte Systemgleichung,

$$I'(t) = -S'(t) - R'(t) = \underbrace{cS(t)I(t)}_{\text{neu Infizierte}} - \underbrace{wI(t)}_{\text{Geheilte}}. \quad (1.21)$$

Zusammen bilden die Gleichungen (1.19), (1.20) und (1.21) das *SIR-Modell*. Wir schreiben dieses nochmals kurz und kompakt hin:

Das *SIR-Modell* ist durch das folgende nichtlineare System von Differentialgleichungen beschrieben:

$$S' = -cSI \quad (1.22)$$

$$I' = cSI - wI \quad (1.23)$$

$$R' = wI \quad (1.24)$$

Beispiel 1.8.1. Bei $N = 1000$ Studierenden einer Uni schleppt eine Person eine Krankheit ein. Wir nehmen als Zeiteinheit einen Tag an und $c = 0.002$, $n = 5$, also $w = 0.2$. Die Anfangsbedingungen lauten dann

$$S(0) = 999,$$

$$I(0) = 1,$$

$$R(0) = 0.$$

Die numerische Lösung des Systems ist in der Abbildung 1.19 dargestellt.

Die numerische Lösung erweckt den Anschein, als strebe die Anzahl Infizierbarer $S(t)$ für $t \rightarrow \infty$ gegen Null. Dies ist sogar recht plausibel, denn in einer Population, in der niemand stirbt und niemand geboren wird, sollte man annehmen, dass jedes Individuum früher oder später angesteckt wird. Wir werden bei der anschliessenden Analyse aber sehen, dass dies nicht so ist.

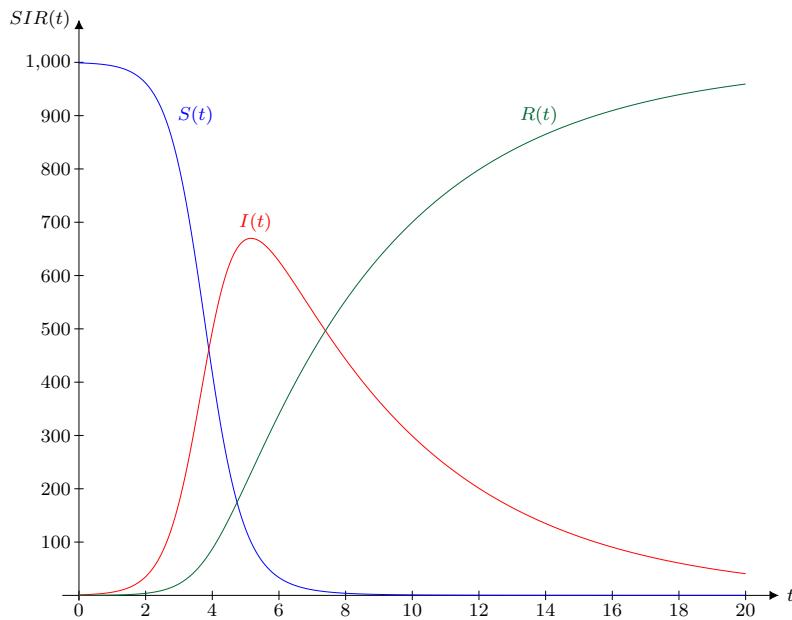


Abbildung 1.19: Der typische Verlauf einer Epidemie am Beispiel des SIR-Modells.

1.8.3 Diskussion des SIR Modells

Das Differentialgleichungssystem des SIR-Modells ist zwar nicht explizit lösbar, dennoch sind eine ganze Reihe interessanter Aussagen möglich.

Monotonie

Solange I und $S > 0$ sind, folgt aus (1.22), dass $S'(t) < 0$. Insbesondere ist $S(t)$ also strikt monoton fallend.

Höhepunkt der Epidemie

Im Höhepunkt der Krankheit, das heisst zu dem Zeitpunkt t_{\max} für den die Anzahl infizierter Personen $I(t_{\max})$ maximal ist, gilt $I'(t_{\max}) = 0$. Daraus lässt sich hier zwar t_{\max} nicht explizit berechnen, aber aus (1.23) ergibt sich immerhin die Beziehung

$$S(t_{\max}) = \frac{w}{c}.$$

Das heisst am Wert von S kann man ablesen, wann die Anzahl Infizierter I maximal ist (siehe auch die Abbildung 1.20 am Ende des Abschnitts).

Invariante

Die Gleichungen (1.22) und (1.23) enthalten nur die Funktionen S und I . Damit ist die Gleichung (1.24) entkoppelt. Insbesondere hatten wir es ja so eingerichtet, dass die Summe $S(t) + I(t) + R(t)$ konstant (also eine *Invariante*) ist. Somit gilt

$$R(t) = S(0) + I(0) + R(0) - (S(t) + I(t)).$$

Der Zustandsraum

Um die Ausbreitung der Krankheit besser zu verstehen, ist es günstig, die Trajektorien im *Zustandsraum* zu betrachten. Das bedeutet, wir betrachten die Lösungen in der (S, I) -Ebene, also die parametrisierten Kurven

$$t \mapsto (S(t), I(t)) \quad (1.25)$$

(siehe die Abbildung 1.20 am Ende des Abschnitts). Dazu berechnen wir mit Hilfe von (1.22) und (1.23)

$$\frac{dI}{dS} = \frac{\frac{dI}{dt}}{\frac{dS}{dt}} = \frac{cSI - wI}{-cSI} = \frac{w}{cS} - 1. \quad (1.26)$$

Das heisst, die Ableitung der Funktion $I(S)$ ist durch den expliziten Ausdruck rechts gegeben. Eine Stammfunktion hiervon ist schnell gefunden:

$$I(S) = \frac{w}{c} \ln S - S + k.$$

Dabei bezeichnet k die Integrationskonstante. Diese Funktionen $I(S)$ sind konkav, und die parametrisierten Kurven (1.25) werden wegen $\frac{dS}{dt} < 0$ von rechts nach links durchlaufen. Die Abbildung 1.20 am Ende des Abschnitts zeigt diese Kurven für verschiedene Werte von k . Der Anfangspunkt $(S(0), I(0))$ bestimmt den Wert von k , das heisst die konkrete Lösungskurve. Anhand dieser Kurven können wir nun den qualitativen Verlauf der Epidemie ablesen und diskutieren. Insbesondere können wir die Frage beantworten, ob zwangsläufig die gesamte Population von der Seuche befallen wird.

Das qualitative Verhalten

Für $t = 0$ erhalten wir aus (1.26)

$$\frac{dI}{dS}(0) = \frac{w}{cS(0)} - 1.$$

Daraus lassen sich nun interessante Folgerungen ziehen:

- Ist $S(0) > \frac{w}{c}$, so nimmt I zunächst zu, bis $S(t)$ den Wert $\frac{w}{c}$ erreicht. Dann ist, wie wir zuvor gesehen haben, die Anzahl I infizierter Personen maximal. Nachher sinkt $I(t)$ monoton gegen 0, und $S(t)$ nimmt weiter gegen einen Grenzwert $S_\infty > 0$ ab. Die Krankheit stirbt also aus, bevor sie alle Individuen erwischt hat! Die Abbildung 1.20 illustriert diese Beobachtung.
- Ist $S(0) < \frac{w}{c}$, so folgt wegen der eingangs festgestellten Monotonie für alle Zeiten $S(t) < \frac{w}{c}$. Somit klingt I strikt monoton gegen 0 und S gegen einen positiven Grenzwert S_∞ ab.

Für die Gesundheitsbehörden ist für den Fall einer Epidemie der kritische Wert $\frac{w}{c}$ von zentraler Bedeutung. Wenn es gelingt, diesen Wert gross zu halten, ist es möglich, die unkontrollierte Ausbreitung der Seuche zu verhindern. Der Heilungskoeffizient $w = 1/n$ hängt von der Krankheitsdauer n ab. Letztere lässt sich zwar typischerweise nur schwer beeinflussen, aber indem man infizierte Personen bereits nach $n' < n$ Tagen isoliert, lässt sich der Wert von w dennoch erhöhen. Andererseits kann man im Vorfeld durch Impfungen dafür sorgen, dass die Anzahl infizierbarer Personen kleiner ist als der kritische Wert, das heisst $S < \frac{w}{c}$. Die Epidemie klingt dann von alleine ab. Oder man sorgt durch Massnahmen wie Quarantäne oder Änderung des Verhaltens in der Bevölkerung (Mundschutz, Händewaschen, Benutzen von Kondomen) dafür, dass der Wert der Variablen p und q kleiner wird. Dann wird nach (1.19) auch $c = pq$ kleiner, und somit verschiebt sich der kritische Wert $\frac{w}{c}$ nach oben, womit selbst bei einer geringeren Impfquote die Epidemie abklingt.

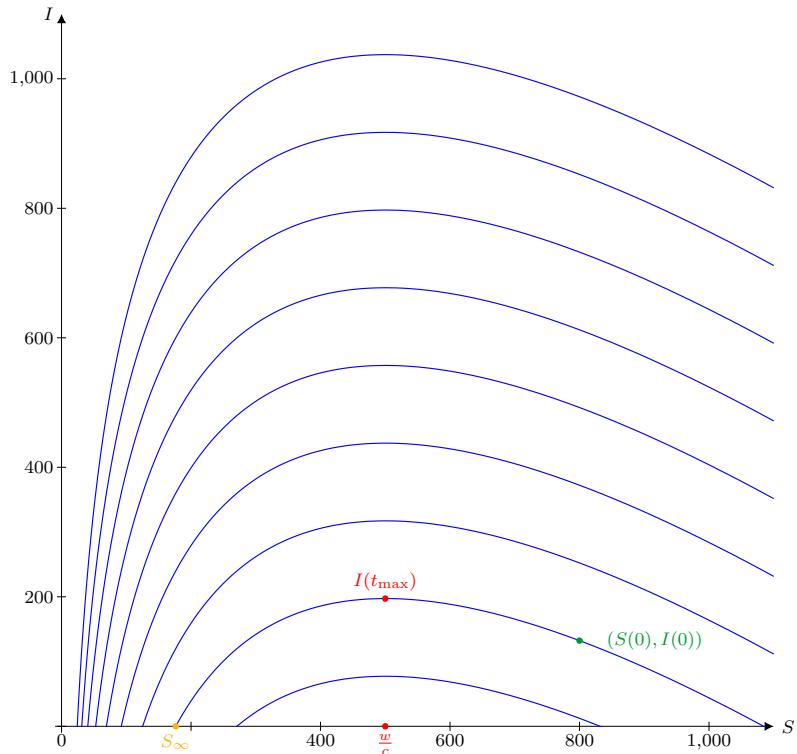


Abbildung 1.20: Der Verlauf der Epidemie in der (S, I) -Ebene.

Die Abbildung 1.21 illustriert den Erfolg solcher Massnahmen, nämlich den Krankheitsverlauf aus dem Beispiel 1.8.1 mit den geänderten Parametern $c' = 0.001$, $n' = 2$, also $w' = 0.5$. Im Beispiel 1.21 war der Wert $w/c = 100 < 500 = w'/c'$. Der Verlauf der Epidemie ist daher nun einerseits deutlich verlangsamt und das Gesundheitswesen wird nicht überlastet, andererseits wird ein viel geringerer Teil der Population überhaupt von der Krankheit befallen.

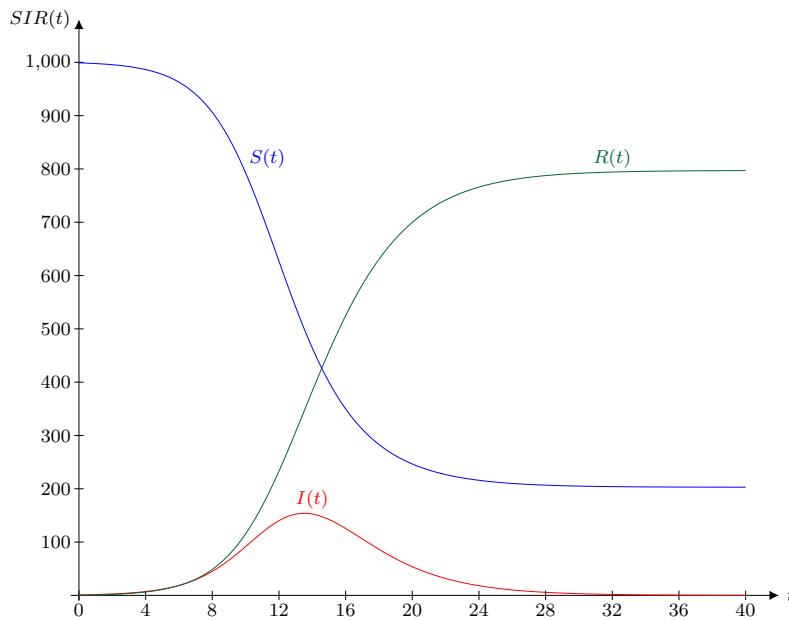


Abbildung 1.21: Ein milderer Verlauf der Epidemie wird erreicht durch geeignete Massnahmen zur Bekämpfung der Ausbreitung der Seuche.

1.9 Diskrete Modelle

Schauen wir nochmals einfache Modelle an und fragen zum Beispiel: Wie entwickelt sich im Laufe der Zeit

1. die Zahl von Bakterien eines Stammes?
2. das Wachstum von Krebszellen?
3. die Verschmutzung eines Sees?
4. die Konzentration eines Medikaments?
5. die Ausbreitung eines Gerüchts, einer Krankheit, Erfindung?
6.

Mathematische Modelle helfen uns, diese Fragen teilweise zu beantworten. Zum Beispiel: Stirbt eine Population aus oder wird sie zu gross? Bisher hatten wir dafür jeweils eine kontinuierliche Entwicklung betrachtet, die durch eine Funktion $t \mapsto f(t)$ beschrieben wird.

In der Praxis messen wir in der Regel zu bestimmten Zeitpunkten eine Grösse und versuchen, aus der Gesamtheit dieser Werte eine allgemeine Gesetzmässigkeit zu finden.

Wähle eine Zeiteinheit (ZE): 1 Jahr, 1 Stunde oder noch kleiner. Seien dann

a_0 = Anfangswert (zum Beispiel die Enzymkonzentration im metabolischen Netzwerk eine Zelle)

a_1 = Wert nach 1 Zeitschritt a_2 = Wert nach 2 Zeitschritten

\vdots

a_n = Wert nach n Zeitschritten

\vdots

Das bedeutet: Jedem Zeitschritt n mit $n = 0, 1, 2, 3, \dots$ ordnen wir eine Zahl (einen Messwert) zu, den wir a_n nennen. Dies ist dann eine diskrete Beschreibung, da wir nur für die Zeitschritte n den Wert a_n kennen.

Formal schreiben wir: $n \mapsto a_n$

Oft gibt es auch die Bezeichnungen oder Notationen: $n \mapsto N_n$, $i \mapsto N_i$ oder $i \mapsto a_i$ oder ...

Bemerkung:

Mathematisch ist eine diskrete Entwicklung $n \mapsto a_n$ eine Folge $(a_n)_n$ reeller Zahlen. In der Anwendung sind die Zeitschritte beschränkt, bei einer Folge bricht der Index n nicht ab.

1.9.1 Beispiel Altersstruktur einer Population

Oft messen wir im n -ten Zeitschritt nicht einen Wert a_n sondern mehrere, zum Beispiel drei Werte x_n, y_n, z_n , welche wir zu einem Vektor

$$v_n := \begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix}$$

zusammenfassen. Wir haben damit eine Zuordnung (Abbildung)

$$n \mapsto v_n = \begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix}.$$

Der Vektor v_n beschreibe zum Beispiel die Altersverteilung in einer Population zur Zeit n :

x_n = Anzahl der Zellen, welche genau 1 ZE überlebt haben: 1. Altersstufe.

y_n = Anzahl der Zellen, welche genau 2 ZE überlebt haben: 2. Altersstufe.

z_n = Anzahl der Zellen, welche genau 3 ZE überlebt haben: 3. Altersstufe.

Ein Entwicklungsmodell sei wie folgt gegeben: Für $i = 1, 2, 3$ ist

r_i = durchschnittliche Anzahl der Zellen, welche aus einer Zelle der i -ten Altersstufe entstehen. Gezählt werden nur die neuen Zellen, welche 1 ZE überleben.

q_i = Anteil der Zellen, welche eine ZE überleben und in die $(i+1)$ -te Altersstufe übergehen. Zellen überleben höchstens 3 ZE.

Wie können wir die Entwicklung $v_0, v_1, v_2, v_3, \dots, v_n, \dots$ beschreiben? Aus den Angaben erhalten wir drei Gleichungen für den $(n+1)$ -ten Zeitschritt.

1. Altersstufe: Anzahl der Zellen, welche genau 1 ZE überlebt haben oder anders: Wie viele Zellen x_{n+1} werden beim Übergang $n \rightarrow n+1$ erzeugt? Es entstehen

- aus x_n neu $r_1 \cdot x_n$ Stück
- aus y_n neu $r_2 \cdot y_n$ Stück
- aus z_n neu $r_3 \cdot z_n$ Stück
- insgesamt also $x_{n+1} = r_1 x_n + r_2 y_n + r_3 z_n$

2. Altersstufe: Anzahl Zellen, welche aus einer Zelle der 1. Stufe entstehen: $y_{n+1} = q_1 x_n$.

3. Altersstufe: Anzahl Zellen, welche aus einer Zelle der 2. Stufe entstehen: $z_{n+1} = q_2 y_n$.

Insgesamt erhalten wir drei Gleichungen

$$\begin{aligned} x_{n+1} &= r_1 x_n + r_2 y_n + r_3 z_n \\ y_{n+1} &= q_1 x_n \\ z_{n+1} &= q_2 y_n, \end{aligned}$$

welche uns den Vektor

$$v_{n+1} = \begin{pmatrix} x_{n+1} \\ y_{n+1} \\ z_{n+1} \end{pmatrix}$$

liefern. Mit Hilfe der Matrizenrechnung unten können wir die Entwicklung $v_0, v_1, \dots, v_n, v_{n+1}, \dots$ elegant beschreiben.

Wir werden mit diesen diskreten Modell beginnen und dann die Theorie für die kontinuierlichen Modelle entwickeln.

KAPITEL 2

Diskrete Modelle

2.1 Eindimensionale Entwicklungen

Wir steigen mit diskreten eindimensionalen Entwicklungen

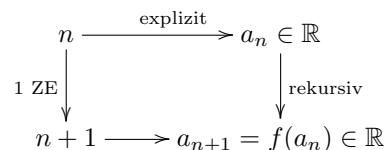
$$n \mapsto a_n = \text{Meswert nach } n \text{ Zeitschritten}$$

ein. Dabei unterscheiden wir zwischen einer *rekursiven* oder einer *expliziten* Darstellung der Entwicklung.

Rekursive heisst: Der Wert a_n ist abhängig von einem oder mehreren vorherigen Werten a_{n-1}, a_{n-2}, \dots

Explizit bedeutet, dass der a_n durch eine Formel berechnet, die nur von n abhängt. Die vorherigen Werte brauchen wir nicht zu kennen.

Formal können wir uns die Situation in einem Diagramm vorstellen:



Dabei ist f die *Reproduktionsfunktion*, die wir später genauer kennenlernen.

2.1.1 Lineare Entwicklung

Erste elementare eindimensionale Entwicklungen $n \mapsto a_n$ sind lineare Entwicklungen:

Definition 2.1.1. Eine (eindimensionale) Entwicklung heisst linear, wenn der Wert a_n in jedem Zeitschritt um eine konstante additive Grösse $l \neq 0$ zu- oder abnimmt:

$$a_n = a_{n-1} + l.$$

Wir berechnen die explizite Darstellung einer linearen Entwicklung:

$$\begin{aligned} a_n &= a_{n-1} + l \\ &= (a_{n-2} + l) + l = a_{n-2} + 2l \\ &= (a_{n-3} + l) + 2l \\ &= \dots = a_0 + n \cdot l. \end{aligned}$$

Also zusammen $a_n = a_0 + l \cdot n.$

Bemerkung:

Eine lineare Entwicklung ist durch l und den Startwert a_0 eindeutig charakterisiert.

Beispiele

- Beim Menschen gilt für Kopfhaar- und Nagelwuchs, je nach Individuum und Körperteil:

Haare: 0,3 – 0,5 mm pro ZE = Tag (in Zyklen)

Nägel: 0,5 – 1,2 mm pro ZE = Woche

- Eine Population wachse von $a_0 = 100$ an linear. Nach drei Jahren sei $a_3 = 160$. Wie lautet das Entwicklungsmodell?

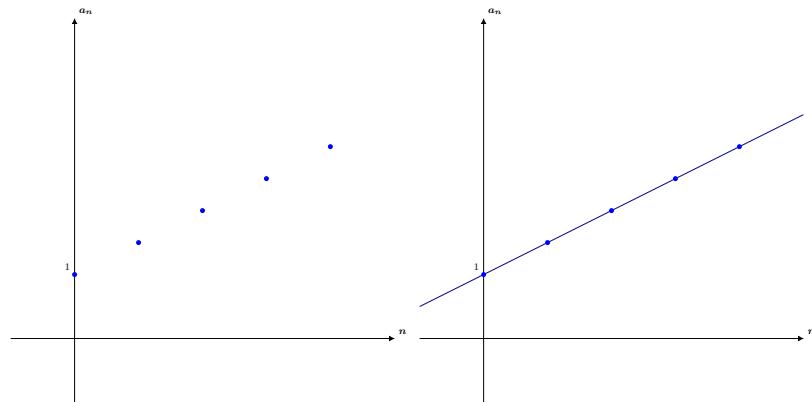
Eine lineare Entwicklung ist nach Definition von der Form $a_n = a_0 + l \cdot n$. Mit $a_0 = 100$ und $a_3 = 160$ folgt $a_3 = 160 = 100 + 3 \cdot l$, also $l = \frac{160 - 100}{3} = 20$. Und somit $a_n = 100 + 20 \cdot n$.

Veranschaulichung

Sei $n \mapsto a_n$ eine lineare Entwicklung mit $a_0 = 1, l = \frac{1}{2}$. Dann erhalten wir eine Tabelle mit Werten.

n	0	1	2	3	4	\dots
a_n	1	$\frac{3}{2}$	2	$\frac{5}{2}$	3	\dots

Wir zeichnen diese Werte als Punkte in ein (n, a_n) -Koordinatensystem ein. Verbinden wir die Punkte, entsteht eine Gerade mit Steigung $\frac{1}{2} (= l)$ und Achsenabschnitt 1 ($= a_0$).



2.1.2 Exponentielle Entwicklung

Ein Einzeller teile sich pro Stunde zweimal. Wie viele Zellen gibt es nach 1, 2, 3, 4, ... Stunden?

Stunden	0	1	2	3	4	...
Anzahl	1	4	16	64	256	...

Wir sehen, dass $a_n = 4 \cdot a_{n-1}$ mit $a_0 = 1$ und erhalten damit eine rekursive Formel. Durch rekursives Einsetzen erhalten wir wieder die explizite Darstellung:

$$\begin{aligned} a_n &= 4a_{n-1} \\ &= 4(4a_{n-2}) \\ &= 4^2 a_{n-2} \\ &= 4^2 (4a_{n-3}) = 4^3 a_{n-3} = \dots \\ &= 4^n a_0 = 4^n \cdot 1 = 4^n. \end{aligned}$$

Definition 2.1.2. Eine (eindimensionale) Entwicklung heisst exponentiell, wenn der Wert a_n in jedem Zeitschritt um einen konstanten multiplikativen Faktor $g > 0$ zu- oder abnimmt:

$$a_n = g \cdot a_{n-1}.$$

Bemerkung:

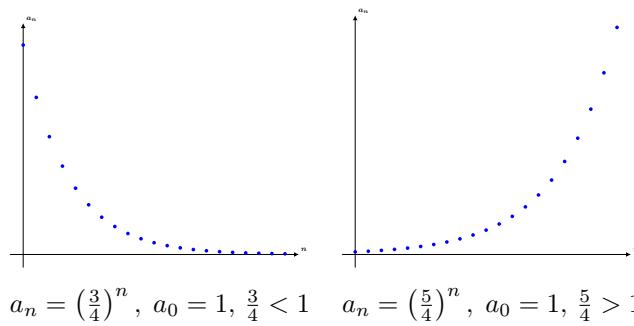
- Die Population nimmt zu, falls $g > 1$; sie nimmt ab, falls $g < 1$.
- Die Berechnung der expliziten Formel erfolgt wie im Beispiel oben:

$$a_n = ga_{n-1} = g(ga_{n-2}) = g^2 a_{n-2} = \dots = g^n a_0,$$

also $a_n = g^n \cdot a_0.$

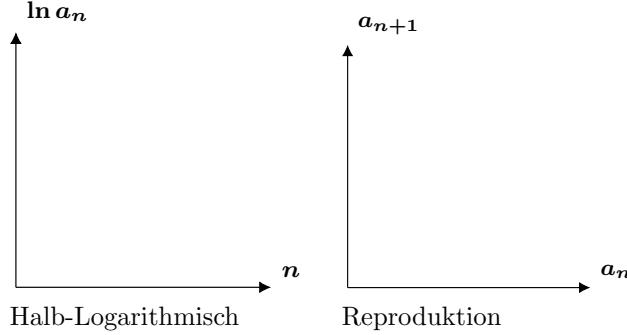
Darstellungsformen

Wie sehen Schaubilder in einem (n, a_n) -Koordinatensystem aus?



Wir erhalten wieder einen bekannten Verlauf mit exponentiell fallenden oder wachsenden Kurven.

Beispiel 2.1.3. Weitere Darstellungsformen erhalten wir in einem $(n, \ln(a_n))$ -Koordinatensystem oder (a_n, a_{n+1}) -Koordinatensystem:



Diese diskutieren wir in den Übungsaufgaben.

Beispiel 2.1.4. Eine Population habe zu Beginn den Bestand a_0 und nehme jährlich um 10% ab. Wann hat sie sich mindestens halbiert? Oder anders gefragt: Wie gross ist ihre Halbwertzeit?

Wenn die Population nach jeder ZE um einen Prozentsatz $-p$ abnimmt, gilt

$$a_n = a_{n-1} - p \cdot a_{n-1} = (1 - p) \cdot a_{n-1}.$$

Die Population entwickelt sich damit exponentiell mit $g = 1 - p < 1$ und nimmt jede ZE um den Faktor $1 - p$ ab, also explizit:

$$a_n = (1 - p)^n \cdot a_0.$$

Wir suchen n mit $a_n = 1/2 \cdot a_0$:

$$a_n = (1 - p)^n \cdot a_0 = \frac{1}{2} a_0.$$

Kürzen wir $a_0 \neq 0$, suchen wir damit n mit $(1 - p)^n = \frac{1}{2}$:

$$(1 - p)^n = \frac{1}{2} \implies \ln(1 - p)^n = \ln \frac{1}{2} \implies n \cdot \ln(1 - p) = \ln \frac{1}{2} \implies n = \frac{\ln \frac{1}{2}}{\ln(1 - p)}.$$

Mit $-p = 0,1$ und $1 - p = 0,9$ folgt $n \approx 6,58$. Also hat sich die Population nach 7 Jahren halbiert. Die Halbwertzeit ist somit 7 Jahre.

2.1.3 Begrenzte Entwicklungen

Bisher haben wir in einer Idealisierung angenommen, dass eine Entwicklung keinen Rahmenbedingungen unterliegt. Nach jeder ZE kommt ein konstanter Anteil hinzu (Geburten = γ) und ein konstanter Anteil fällt weg (Sterbefälle = σ). Die Population ändert sich von der aktuellen ZE n zur nächsten $n + 1$:

$$a_{n+1} = a_n + \underbrace{\gamma \cdot a_n - \sigma \cdot a_n}_{\text{Beschreibt Veränderung}} = a_n + p \cdot a_n = g \cdot a_n$$

mit $g = 1 + p = 1 + \gamma - \sigma$.

Im Beispiel oben ist $\gamma - \sigma = -0,1$.

In der Natur spielen aber Begrenzung des Lebensraumes, der Nahrung etc. eine wichtige Rolle. Daher führen wir eine bf Korrektur ein: $a_{n+1} = g \cdot a_n - k_n = a_n + p \cdot a_n - k_n$. In jedem Zeitschritt n erhalten wir eine Korrektur k_n .

Beispiel 2.1.5 (Konstante Ernte). Angenommen, wir entnehmen nach jeder ZE einen konstanten (nicht zu grossen) Teil H aus der Population ($H = \text{Harvest} = \text{Ernte}$). Dann ist $k_n = H$ unabhängig von n . Wie sieht die explizite Darstellung zu

$$a_{n+1} = a_n + p \cdot a_n - H$$

aus?

Wir setzen $g = 1 + p$ und erhalten $a_{n+1} = g a_n - H$. Starten wir mit $a_1 = g a_0 - H$, erhalten wir einen Schritt weiter

$$a_2 = g a_1 - H = g(g a_0 - H) - H = g^2 a_0 - H(g + 1)$$

und

$$a_3 = g a_2 - H = g(g^2 a_0 - H(g + 1)) - H = g^3 a_0 - H(g^2 + g + 1)$$

usw. Für a_n ergibt sich damit

$$a_n = g^n a_0 - H \sum_{k=0}^{n-1} g^k \stackrel{*}{=} g^n a_0 - H \frac{1 - g^n}{1 - g} = g^n \left(a_0 + \frac{H}{1 - g} \right) - \frac{H}{1 - g},$$

wobei wir (mit $g \neq 1$) die Formel für die geometrische Reihe $\sum_{k=0}^{n-1} g^k \stackrel{*}{=} \frac{1 - g^n}{1 - g}$ verwendet haben.

$$\text{Also lautet die explizite Darstellung des Modells } a_n = g^n \left(a_0 + \frac{H}{1 - g} \right) - \frac{H}{1 - g}.$$

CAVE

Spätestens hier fragen wir uns, wie gross die Ernte H für ein sinnvolles Wachstum sein darf. Wir müssen $0 < H < p \cdot a_0$ annehmen. Das können wir rekursiv oder explizit begründen:

Rekursiv Wir möchten, dass die Population trotz Ernte wächst. Die rekursive Darstellung sagt

$$a_{n+1} = a_n + p a_n - H$$

und gleich zu Beginn soll $a_1 > a_0$ gelten, also:

$$a_1 > a_0 \implies a_1 - a_0 > 0 \implies (a_0 + p a_0 - H) - a_0 > 0 \implies p a_0 - H > 0 \implies p a_0 > H > 0.$$

Explizit Damit die Population wächst und $a_n > 0$ gilt, muss der Ausdruck in der Klammer

$$a_0 + \frac{H}{1 - g} > 0$$

sein. Wir rechnen, mit $1 - g = -p$:

$$a_0 + \frac{H}{1 - g} = a_0 - \frac{H}{p} > 0 \implies a_0 > \frac{H}{p} \implies p \cdot a_0 > H > 0.$$

Beispiel 2.1.6 (Kapazitätsgrenze). Gegeben sei eine Entwicklung durch die rekursive Darstellung

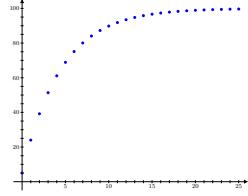
$$a_{n+1} = a_n + \underbrace{p \cdot (K - a_n)}_{\text{Beschreibt Veränderung}},$$

mit Konstanten $0 \leq p \leq 1$ und K , $K \geq a_0$. Wie sieht das qualitative Verhalten der Entwicklung aus?

Nehmen wir zum Beispiel an, dass zu Beginn K wesentlich grösser als a_0 ist:

- Dann ist $K - a_n \approx K$ für kleine n , und die Entwicklung wächst zu Beginn linear, denn
$$a_{n+1} = a_n + p \cdot (K - a_n) \approx a_n + pK = a_n + l \text{ mit } l = pK.$$
- Mit der Zeit wird der Bestand a_n immer grösser und damit $K - a_n$ immer kleiner, es kommt immer weniger hinzu, und die Entwicklung stagniert $a_{n+1} \approx a_n$, denn $K - a_n \approx 0$.

Beispiel mit $a_0 = 5, p = 0.2, K = 100$:



- Wenn wir K als die maximale Kapazität interpretieren, ist $K - a_n$ die in der n -ten ZE freie Kapazität. Diese bestimmt mit einer Proportionalität p die Veränderung, da

$$a_{n+1} = a_n + p \cdot (K - a_n) \implies a_{n+1} - a_n = p \cdot (K - a_n) \sim K - a_n$$

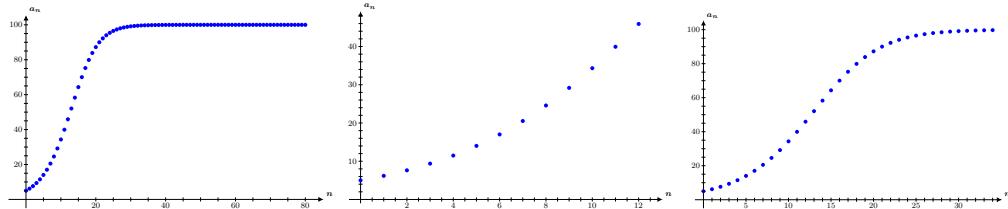
- Eine Variante des Modells betrachtet die Proportionalität $(K - a_n)/K$ als Vergleich zur maximalen Kapazität K . Um die Rechnungen einfacher zu halten, nehmen wir den Nenner mit in die Konstante p .

Es ist $a_{n+1} = a_n + p \cdot (K - a_n) = a_n + p \cdot a_n - (p \cdot (2a_n - K))$, und damit die Korrektur $k_n = p \cdot (2a_n - K)$ abhängig vom Bestand a_n zum Zeitpunkt n .

2.1.4 Diskrete logistische Entwicklung

Wichtige begrenzte Entwicklungen sind diskrete logistische Entwicklungen. Eine logistische Entwicklung wächst zunächst exponentiell (und nicht linear wie oben), sie bremst dann und ist wieder durch eine Kapazität K begrenzt.

Der Verlauf ist typischerweise



mit einer rekursiven Darstellung,

$$a_{n+1} = a_n + \underbrace{p \cdot a_n \cdot (K - a_n)}_{\text{Beschreibt Veränderung}},$$

mit $0 < p < 1$. Im Beispiel oben sind $a_0 = 5$, $p = \frac{1}{400}$ und $K = 100$.

Qualitative Analyse

Schauen wir die Entwicklung qualitativ an und nehmen an, dass zu Beginn K wesentlich grösser als a_0 ist:

- Dann ist $K - a_n \approx K$ für kleine n , und die Entwicklung wächst zu Beginn exponentiell, denn

$$a_{n+1} = a_n + p \cdot a_n(K - a_n) \approx a_n + p \cdot a_n \cdot K = (1 + pK) \cdot a_n$$
 mit $g = 1 + pK$.
- Mit der Zeit wird der Bestand a_n immer grösser und damit $K - a_n$ immer kleiner, es kommt immer weniger hinzu, und die Entwicklung stagniert $a_{n+1} \approx a_n$, denn $K - a_n \approx 0$.

Bemerkung:

Schreiben wir die logistische Gleichung mit geeigneten Konstanten F und G ist

$$a_{n+1} = F \cdot a_n - G \cdot a_n^2,$$

und die Korrektur $k_n = Ga_n^2$ hängt damit sogar quadratisch vom aktuellen Bestand ab.

Beispiel 2.1.7. Angenommen, die Korrektur k_n in $a_{n+1} = g \cdot a_n - k_n$ ist proportional zur Anzahl der möglichen Begegnungen im beschränkten Lebensraum ($= B_n$), das heisst, es gibt eine Konstante c mit $k_n = c \cdot B_n$. Berechnen wir zunächst B_n , die Anzahl der Begegnungen. Dies ergibt sich aus folgender Überlegung:

Auf einer Party befinden sich H Personen. Für die erste Person gibt es $H - 1$ Möglichkeiten einer anderen Person die Hand zu geben (Begegnung), für die zweite Person $H - 2$ usw. Also haben wir insgesamt für die Handschläge (Begegnungen):

$$(H - 1) + (H - 2) + \dots + 1 = 1 + 2 + 3 + \dots + H - 1 = \sum_{i=1}^{H-1} i.$$

Mit der bekannten Summenformel von Gauss $\sum_{i=1}^N i = \frac{N(N+1)}{2}$ erhalten wir

$$\sum_{i=1}^{H-1} i = \frac{(H-1)((H-1)+1)}{2} = \frac{(H-1)H}{2}.$$

Im n -ten Zeitschritt ist die Populationsgrösse a_n , also $B_n = \frac{(a_n - 1)a_n}{2}$ und

$$k_n = c \cdot B_n = c \cdot \frac{(a_n - 1)a_n}{2}.$$

Setze diese k_n in $a_{n+1} = g \cdot a_n - k_n$ ein:

$$\begin{aligned} a_{n+1} &= g \cdot a_n - k_n \\ &= g \cdot a_n - c \cdot \frac{(a_n - 1)a_n}{2} = \left(g + \frac{c}{2}\right) \cdot a_n - \frac{c}{2} \cdot a_n^2, \end{aligned}$$

und damit $a_{n+1} = F \cdot a_n - G \cdot a_n^2$ mit $F = g + \frac{c}{2}$ und $G = \frac{c}{2}$.

Bemerkung:

Alternativ folgt die Formel $B_n = \frac{(a_n - 1)a_n}{2}$ auch mit Hilfe der Kombinatorik, denn die Anzahl der ungeordneten Stichproben von 2 Elementen ohne Zurücklegen bei a_n gegebenen Elementen ist $\binom{a_n}{2} = \frac{(a_n - 1)a_n}{2}$.

2.1.5 Fixpunkte eindimensionaler diskreter Entwicklungen

Hat eine Entwicklung a_n einen Wert, bei dem ändert sie sich nicht mehr ändert, so nennen wir diesen Wert $\tilde{a}_n = \tilde{a}$ einen Fixpunkt.

Definition 2.1.8. Gilt bei einem Bestand $\tilde{a}_n = \tilde{a}$, dass $\tilde{a}_n = \tilde{a}_{n+1} = \tilde{a}$ so heisst dieser Fixpunkt.

Beispiele

- Linear: Die Folge $a_{n+1} = a_n + l$ mit $l \neq 0$ hat keinen Fixpunkt $< \infty$.
- Exponentiell: Die Folge $a_{n+1} = ga_n$ (mit $g \neq 1$) hat allein den Fixpunkt $\tilde{a} = 0$, denn $a_{n+1} = a_n = ga_n$ woraus folgt dass $g = 1$.
- Begrenzt: Die Folge $a_{n+1} = a_n + p(K - a_n)$ hat den Fixpunkt $\tilde{a} = K$.

Logistische Beispiele

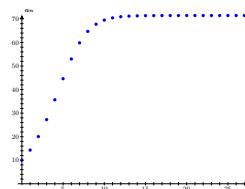
Die allgemeine logistische Entwicklung $a_{n+1} = a_n + pa_n(K - a_n)$ hat die Fixpunkte $\tilde{a} = 0$ und $\tilde{a} = K$. Schauen wir zwei Fälle genauer an:

- Logistisch unter Bejagung: Die Entwicklung beschreiben wir durch

$$a_{n+1} = a_n + \underbrace{pa_n(K - a_n)}_{\text{Wachstum}} - \underbrace{qa_n}_{\text{Jagd}}.$$

Dann sind die Fixpunkte $\tilde{a} = 0$ und $\tilde{a} = K - \frac{q}{p}$.

Beispiel mit $a_0 = 10, p = 0.007, q = 0.2$ und $K = 100$:



- Logistisch mit Nahrungsminimum: Mit $a_{n+1} = a_n + p(a_n - T)(K - a_n)$ und $T < K$. Der Term $p(a_n - T)$ entscheidet über das Wachstum.

Wie sieht die Entwicklung qualitativ aus? Das untersuchen wir in den Übungen, indem wir mit $T < K$ und $a_n > 0$ die Fälle

$$a_n < T < K \quad T < a_n < K, \quad T < K < a_n$$

unterscheiden.

Veränderungsrate

Sei $n \mapsto a_n$ eine diskrete Entwicklung. Die (absolute) Zu-/Abnahme im n -ten Zeitschritt ist die Differenz $a_{n+1} - a_n$, und die relative Zu-/Abnahme zum aktuellen Bestand $a_n \neq 0$ ist dann

$$\frac{a_{n+1} - a_n}{a_n}.$$

Beispiel 2.1.9. Sei $n \mapsto a_n$ eine diskrete exponentielle Entwicklung: $a_{n+1} = g a_n$. Einsetzen von $a_{n+1} = g a_n$ in $a_{n+1} - a_n$ liefert $a_{n+1} - a_n = g a_n - a_n = (g - 1)a_n$.

Die relative Zu-/Abnahme ist also $\frac{a_{n+1} - a_n}{a_n} = g - 1$.

Man kann dies allgemein wie folgt definieren.

Definition 2.1.10. Für eine diskrete Entwicklung $n \mapsto a_n$ heisst die Zahl $r_n = \frac{a_{n+1} - a_n}{a_n}$ die **Veränderungsrate** im n -ten Zeitschritt.

Für $r_n > 0$ **Wachstumsrate**, für $r_n < 0$ **Zerfallsrate**.

Beispiele

- Exponentiell: Es ist $r_n = \frac{a_{n+1} - a_n}{a_n} = g - 1$ eine Konstante:

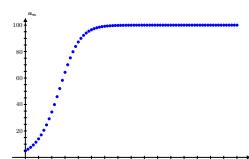
Wenn $g > 1$, folgt $r_n > 0$, also Wachstum.

Wenn $g < 1$ folgt $r_n < 0$, also Zerfall.

- Linear: $a_{n+1} = a_n + l$, also $r_n = \frac{a_{n+1} - a_n}{a_n} = \frac{l}{a_n}$. Da $a_n = a_0 + ln$, folgt $r_n = \frac{l}{a_0 + ln}$, und es gilt $r_n = \frac{l}{a_0 + ln} \rightarrow 0$ für $n \rightarrow \infty$.

Das heisst, "Gemessen am aktuellen Bestand a_n , kommt immer weniger hinzu."

- Logistisch: Sei $a_{n+1} = \frac{5}{4}a_n - \frac{1}{400}a_n^2 = a_n + \frac{1}{4}a_n - \frac{1}{400}a_n^2 = a_n + \frac{1}{400}a_n(100 - a_n)$ mit $a_0 = 5$. Dann ist der zeitliche Verlauf



Die Rate r_n berechnet sich hier mit $r_n = \frac{a_{n+1} - a_n}{a_n} = \frac{\left(\frac{5}{4}a_n - \frac{1}{400}a_n^2\right) - a_n}{a_n} = \frac{1}{4} - \frac{1}{400}a_n$.

Ist die Rate gleich Null, so verändert sich die Entwicklung nicht mehr. Für welchen Bestand ist das der Fall? Wir rechnen

$$\tilde{r} = 0 \implies \tilde{r} = \frac{1}{4} - \frac{1}{400}\tilde{a} = 0 \implies \tilde{a} = 100.$$

Im logistischen Beispiel sehen wir (mit dem Computer!), dass sich die Population genau dem Fixpunkt $\tilde{a} = 100$ annähert, in dem $r = 0$ ist. Allgemein gilt Folgendes.

1. Ein Bestand einer Entwicklung ist genau dann ein Fixpunkt $\tilde{a} > 0$, wenn dort die Veränderungsrate $\tilde{r} = 0$ ist, $r_n = \frac{a_{n+1} - a_n}{a_n} = 0 \iff a_{n+1} - a_n = 0 \iff a_{n+1} = a_n$.
2. Wenn wir wissen, dass sich die Entwicklung $a_n, n = 1, 2, \dots$ einem Bestand \tilde{a} beliebig annähert, muss dieser ein Fixpunkt \tilde{a} sein, und damit $\tilde{r} = 0$.

CAVE

Die umgekehrte Aussage gilt im Allgemeinen nicht. Für einen Fixpunkt \tilde{a} (also $\tilde{r} = 0$) folgt nicht automatisch, dass $a_n, n = 1, 2, \dots$ sich einem Bestand \tilde{a} beliebig annähert.

Beispiel 2.1.11. Sei (a_n) eine Folge mit $a_{n+1} = k \cdot 25 \cdot a_n (1 - \frac{1}{500}a_n)$ für $k = 1, 2, 3, 4$. Für jedes k gibt es einen Fixpunkt $\tilde{a}^{(k)} > 0$, aber nur für $k = 1$ und $k = 2$ gibt es eine konvergente Folge, das heisst, $\tilde{a}^{(k)} = \lim_{n \rightarrow \infty} a_n$. Mit Hilfe der Differentialrechnung werden wir sehen, wann und warum das so ist.

2.1.6 Reproduktion

Sei eine diskrete Entwicklung durch eine Folge $(a_n)_n$ gegeben. Die Folge sei rekursiv definiert, das heisst, wir berechnen den Wert a_{n+1} durch den Wert a_n . Mathematisch heisst das, dass wir eine Abbildung oder eine Funktion

$$\begin{aligned} f : \mathbb{R} &\rightarrow \mathbb{R}, \\ x &\mapsto f(x), \end{aligned}$$

haben, welche die Folge: $a_{n+1} = f(a_n)$ beschreibt. Dabei betrachten wir f als Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ und nehmen an, dass die Variable x kontinuierlich variiert, denn wir wissen a-priori nicht, welche Werte a_n wir jeweils in f einsetzen, um den Folgewert a_{n+1} zu bestimmen. Wir bezeichnen im weiteren Verlauf die Folgenglieder mit x_n statt mit a_n .

In den Anwendungen ist es sinnvoll, $x_n \geq 0$ und $f(x_n) \geq 0$ anzunehmen.

Definition 2.1.12. Eine Folge $(x_n)_n$ sei rekursiv definiert und durch eine Funktion f beschrieben:

$$x_{n+1} = f(x_n).$$

Dann heisst f **Reproduktionsfunktion**.

Bemerkung:

Schreiben wir $f(x)$ als $f(x) = x + \delta(x)$, dann gibt die Funktion

$$\delta : \mathbb{R} \rightarrow \mathbb{R}, x \mapsto \delta(x) = f(x) - x,$$

die (absolute) Veränderung in einem Zeitschritt an.

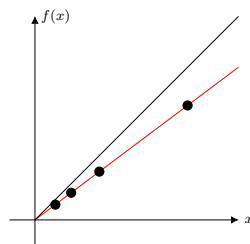
Beispiele 2.1.13.

1. Linear:

Für $x_{n+1} = x_n + l$ ist die Reproduktionsfunktion f mit $f(x) = x + l$, mit der konstanten Funktion $\delta = l$.

2. Exponentiell:

Für ist $x_{n+1} = gx_n$ ist $f(x) = gx$. Es ist $x_{n+1} = gx_n = (1+p)x_n$, dann sind $f(x) = x + px$ und δ eine lineare Funktion mit $\delta(x) = px$. Der Graph ist eine Gerade durch den Nullpunkt mit Steigung g :



Falls $g = 1$, erhalten wir die Winkelhalbierende $y = x$ mit Steigung 1, und die Entwicklung ist konstant, $x_{n+1} = f(x_n)$.

Falls $g < 1$ ist die Steigung kleiner als 1, und die Entwicklung schrumpft, die Werte $f(x_n)$ werden entlang dem Graphen von f immer kleiner.

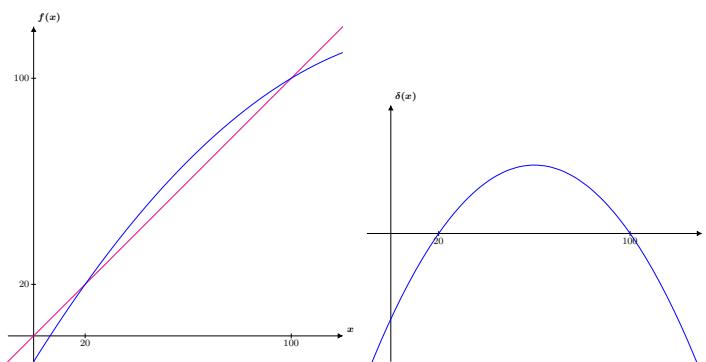
Falls $g > 1$ ist die Steigung grösser als 1, und die Entwicklung wächst. Es ergibt sich eine Gerade oberhalb der Winkelhalbierenden.

3. Logistische Entwicklungen:

Bei einer Entwicklung mit einem **Nahrungsminimum** gilt: $x_{n+1} = x_n + p(x_n - T)(K - x_n)$. Damit ist die Reproduktionsfunktion

$$f(x) = x + \underbrace{p(x - T)(K - x)}_{\text{Veränderung } \delta(x)}, \quad T < K.$$

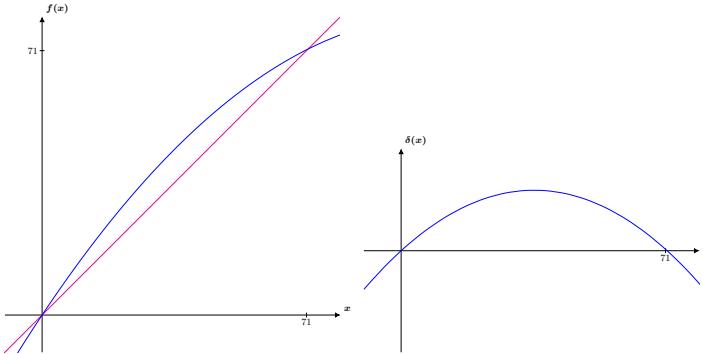
Für $p = 0.005, T = 20, K = 100$ sind die Graphen



Unter Bejagung erhalten wir die Funktion f mit

$$f(x) = x + \underbrace{px(K-x)}_{\text{Veränderung } \delta(x)} - qx$$

und für $p = 0.007, q = 0.2, K = 100$ die Graphen



4. Tumorzellen

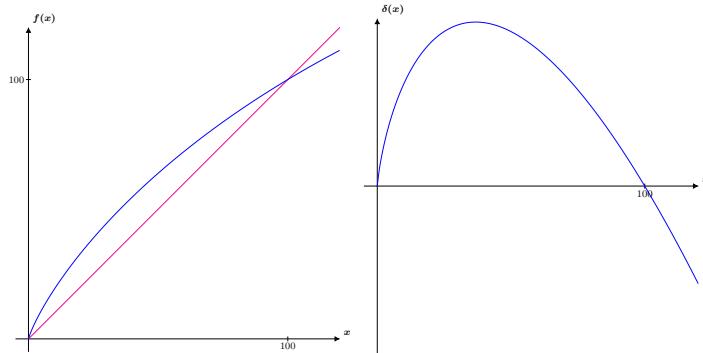
Das Wachstum von Tumorzellen lässt sich annähernd durch die Entwicklung

$$a_{n+1} = a_n + r a_n \ln \left(\frac{K}{a_n} \right)$$

beschreiben, also ist die Reproduktionsfunktion f gegeben durch

$$f(x) = x + rx \cdot \ln \left(\frac{K}{x} \right).$$

Als Beispielgraphen erhalten wir mit $r = 0.4, K = 100$



Explizite Darstellung durch Iteration

Mit der n -fachen Komposition der Reproduktionsfunktion f mit sich (Iteration) können wir aus x_0 den Wert x_n berechnen:

$$\begin{aligned} x_n &= f(x_{n-1}) = f(f(x_{n-2})) = (f \circ f)(x_{n-2}) \\ &= f^2(x_{n-2}) = f^2(f(x_{n-3})) = (f \circ f \circ f)(x_{n-3}) \\ &\quad \vdots \\ &= f^{n-1}(x_1) = f^{n-1}(f(x_0)) = (\underbrace{f \circ f \circ \dots \circ f}_{n \text{ St\xfck}})(x_0) \\ &= f^n(x_0) \quad \text{also } x_n = f^n(x_0) \text{ eine "explizite Darstellung"} \end{aligned}$$

R\xfcckw\rts mit der Umkehrabbildung

Sei $x_{n+1} = f(x_n)$ und sei f umkehrbar. Mit der Umkehrabbildung f^{-1} können wir die Entwicklung rückwärts durchlaufen:

$$x_{n+1} = f(x_n) \implies f^{-1}(x_{n+1}) = f^{-1}(f(x_n)) \implies f^{-1}(x_{n+1}) = \underbrace{(f^{-1} \circ f)(x_n)}_{=x_n}$$

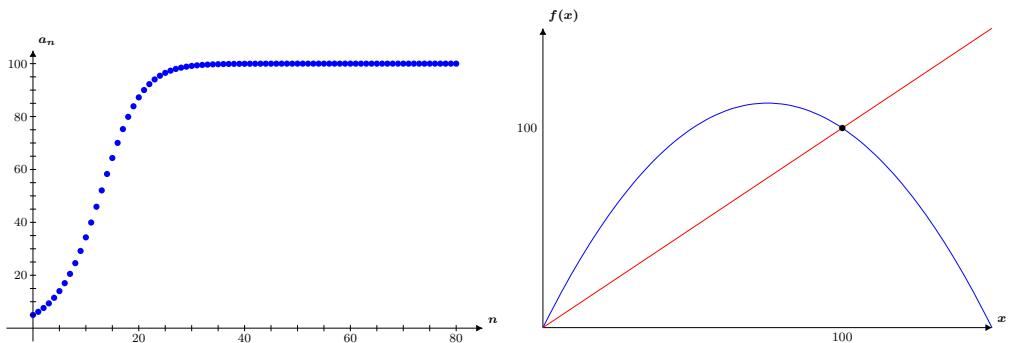
und damit $x_n = f^{-1}(x_{n+1})$ und $x_0 = (f^{-1})^n(x_n)$. Kennen wir also x_n und f^{-1} , können wir unseren Startwert x_0 bestimmen.

2.1.7 Fixpunkte und Konvergenz

Wir wollen nun entscheiden, wann ein Fixpunkt einer Reproduktionsfunktion f tatsächlich dann auch ein Grenzwert einer Folge mit $a_{n+1} = f(a_n)$ ist.

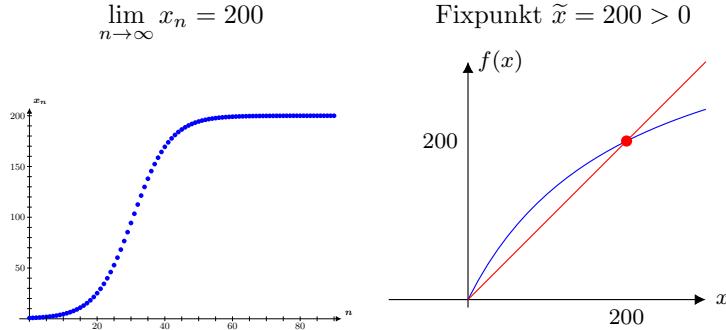
Wie unterscheiden sich die folgenden Beispiele?

- Sei $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = \frac{5}{4}x - \frac{1}{400}x^2$. Dann ist $x_{n+1} = \frac{5}{4}x_n - \frac{1}{400}x_n^2 = f(x_n)$ mit Fixpunkt $\tilde{x} = 100 > 0$.

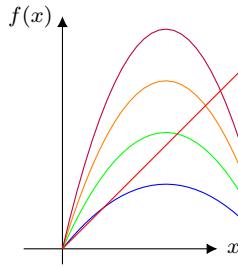


2. Sei $x_{n+1} = f(x_n) = g \cdot x_n \cdot \frac{C}{C + g \cdot x_n}$ mit $C, g > 0$ Konstanten. Es gibt genau dann einen Fixpunkt $\tilde{x} > 0$, wenn $g > 1$ (Übung).

Beispiel mit $g = 1.2, C = 1200$:



3. Seien $f_k : \mathbb{R} \rightarrow \mathbb{R}$ mit $f_k(x) = k \cdot 25 \cdot x \cdot \left(1 - \frac{1}{500}x\right)$ mit $k = 1, 2, 3, 4$.



Jedes f_k hat einen Fixpunkt $\tilde{x}^{(k)} > 0$, aber nur bei f_1 und f_2 gibt es eine Konvergenz von $x_{n+1} = f_k(x_n)$, d.h. $\tilde{x}^{(k)} = \lim_{n \rightarrow \infty} x_n$.

Konvergenzkriterien

Sei $x_{n+1} = f(x_n)$ eine Entwicklung mit stetig differenzierbarer Reproduktionsfunktion $f : \mathbb{R} \rightarrow \mathbb{R}$ und Fixpunkt \tilde{x} : $f(\tilde{x}) = \tilde{x}$. Wann ist $\tilde{x} = \lim_{n \rightarrow \infty} x_n$?

Es gilt:

Falls $|f'(\tilde{x})| < 1$, konvergiert die Folge (x_n) , $x_{n+1} = f(x_n)$ gegen \tilde{x} für jeden Startwert x_0 in der Nähe von \tilde{x} .

Falls $|f'(\tilde{x})| > 1$, ist \tilde{x} kein Grenzwert der Folge (x_n) , $x_{n+1} = f(x_n)$ für jeden Startwert $x_0 \neq \tilde{x}$.

Falls $|f'(\tilde{x})| = 1$, ist zunächst keine Entscheidung mit f' möglich.

Bemerkungen: Anschaulich verstehen wir dies so:

- Falls $|f'(\tilde{x})| > 1$, ist die Steigung der Tangente in $x_n = \tilde{x}$ gross. Der Punkt $x_{n+1} = f(x_n)$ hat dann einen sehr grossen Abstand von $x_n = \tilde{x}$.
- Oder noch anschaulicher: Falls $|f'(\tilde{x})| > 1$, ist die Veränderungsgeschwindigkeit zu gross, und die Entwicklung kann im Fixpunkt nicht stoppen.

Etwas formaler können wir diese Aussagen mit dem Mittelwertsatz auf der Differentialrechnung begründen:

- Sei eine Folge $x_{n+1} = f(x_n)$ mit Fixpunkt \tilde{x} und zusätzlich $|f'(\tilde{x})| < 1$ gegeben.
- Wir überlegen uns, dass damit für jedes n gelten muss $|x_{n+1} - \tilde{x}| < |x_n - \tilde{x}|$. Das bedeutet: Im Laufe der Zeit nähert sich die Folge schliesslich immer näher dem Fixpunkt \tilde{x} an.
- Es ist

$$\begin{aligned} x_{n+1} - \tilde{x} &= f(x_n) - \tilde{x} && \text{Definition der Folge} \\ &= f(x_n) - f(\tilde{x}) && \tilde{x} \text{ ist ein Fixpunkt} \\ &= (x_n - \tilde{x})f'(x^*) && \text{für eine Stelle } x^*, \text{ Mittelwertsatz} \end{aligned}$$

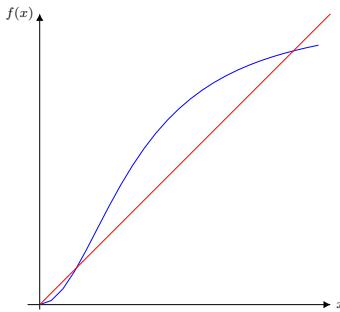
- Das bedeutet für die Beträge mit den Rechenregeln der Betragsfunktion

$$|x_{n+1} - \tilde{x}| = |(x_n - \tilde{x})f'(x^*)| = |x_n - \tilde{x}| \cdot |f'(x^*)|$$

- Nach Voraussetzung $|f'(\tilde{x})| < 1$. Nehmen wir an, dass f' stetig ist und dass x_n in der Nähe von \tilde{x} ist, damit auch $|f'(x^*)| < 1$. Und zusammen $|x_{n+1} - \tilde{x}| = |x_n - \tilde{x}| \cdot \underbrace{|f'(x^*)|}_{<1} < |x_n - \tilde{x}|$

Beispiele

- Sei f mit $f(x) = \frac{\frac{3}{2}x^2}{1 + \frac{x^2}{4}}$ und $f'(x) = \frac{3x}{\left(1 + \frac{x^2}{4}\right)^2}$. Am Graphen



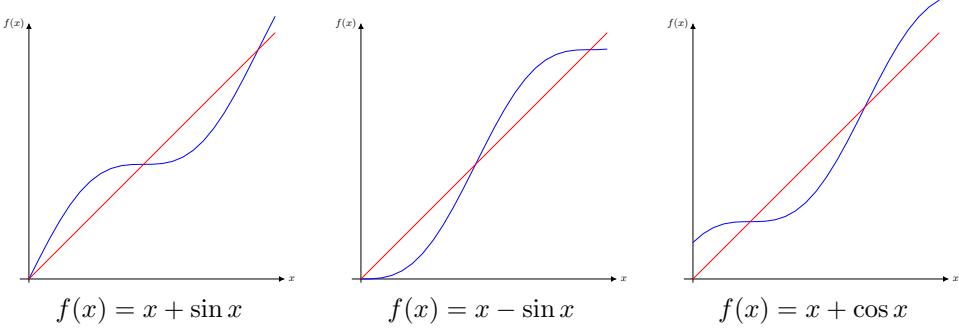
erkennen wir die unterschiedlichen Steigungen der Tangente in den drei Fixpunkten.

Numerisch sind dies

Fixpunkt	$x_0 = 0$	$x_1 = 3 - \sqrt{5} \approx 0.76$	$x_2 = 3 + \sqrt{5}$
	$ f'(0) = 0 < 1$	$ f'(3 - \sqrt{5}) \approx 1.75 > 1$	$ f'(3 + \sqrt{5}) \approx 0.25 < 1$

- Wenden Sie das Kriterium auf die drei Beispiele oben an.
- Wir betrachten (als Übung) drei Funktionsvorschriften für eine Funktion und nennen sie jeweils f

$$f(x) = x \pm \sin x \quad \text{und} \quad f(x) = x + \cos x$$



2.2 Mehrdimensionale Modelle

Wir betrachten nun mehrdimensionale Modelle, in denen einer Zeiteinheit mehrere Werte zugeordnet werden. Wir haben also eine Abbildung $n \mapsto v_n = \begin{pmatrix} x_1^{(n)} \\ \vdots \\ x_m^{(n)} \end{pmatrix} \in \mathbb{R}^m$.

2.2.1 Beispiel Entwicklung Räuber / Beute - Populationen

Gegeben seien zwei Populationen mit Räubern und Beute, beschrieben durch Folgen $(x_n)_n$ (Räuber) und $(y_n)_n$ (Beute). Das heisst, die Population x_n wächst auf Kosten der Population y_n .

$$\begin{aligned} x_{n+1} &= a \cdot x_n + b \cdot y_n, && (\text{Räuber}), \\ y_{n+1} &= c \cdot y_n - d \cdot x_n, && (\text{Beute}), \end{aligned}$$

mit $0 < a, b, d < 1$ und $c > 1$.

Bemerkung:

Angenommen, $b = d = 0$, dann vereinfachen sich die Gleichungen $x_{n+1} = a \cdot x_n$ und $y_{n+1} = c \cdot y_n$ und die Räuber x sind eine exponentiell schrumpfende ($a < 1$) und die Beute y eine exponentiell wachsende Population ($c > 1$).

Zu jedem n haben wir nun $m = 2$ Werte, die Grösse der Räuber und die Grösse der Beute. Die beiden Gleichungen geben an, nach welcher Gesetzmässigkeit sich die Populationen in der nächsten Generation verändert.

Wir arbeiten nun mit den folgendem Zahlenbeispiel weiter:

$$\begin{aligned} x_{n+1} &= \frac{1}{5}x_n + \frac{2}{5}y_n && (\text{Räuber}), \\ y_{n+1} &= \frac{13}{10}y_n - \frac{3}{5}x_n && (\text{Beute}). \end{aligned}$$

Grundfragen

Wir fragen uns:

1. Wie entwickeln sich $\{x_n\}_n$, $\{y_n\}_n$ im Laufe der Zeit?
2. Gibt es ein Gleichgewicht, eine Monotonie- oder ein asymptotisches Verhalten?
3. Gibt es Startwerte (x_0, y_0) , bei denen nichts passiert? Das heisst,

$$x_n = x_0, \quad y_n = y_0, \quad n = 1, 2, 3, \dots$$

4. Kann es sein, dass sich die Population ändert, aber das Verhältnis nicht? Das heisst,

$$\frac{x_n}{y_n} = \frac{x_{n+1}}{y_{n+1}}, \quad y_n \neq 0 \neq y_{n+1}?$$

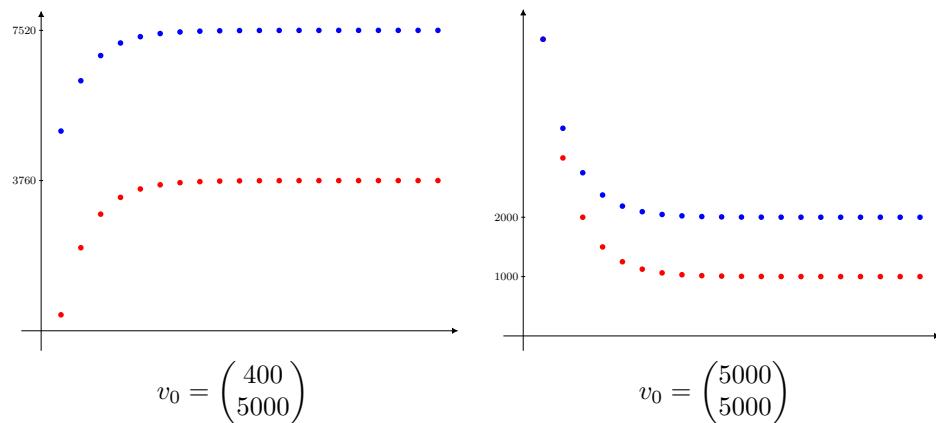
Darstellungen mit Computerhilfe

Wieder fassen wir die beiden Werte x_n und y_n zu einem Vektor

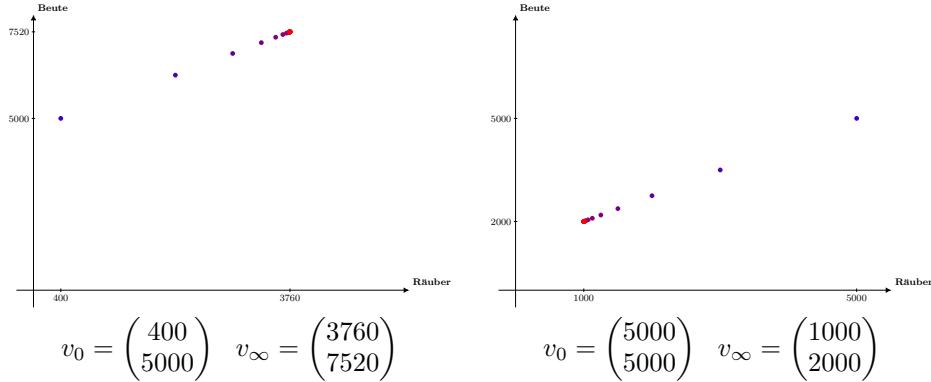
$$v_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix}$$

zusammen.

Der Computer zeigt uns die Entwicklung im Laufe der Zeit bei unterschiedlichen Startvektoren v_0 mit ● Räuber und ● Beute:



In einer anderen Darstellung wählen wir ein (Räuber,Beute)-Koordinatensystem: Dabei sind $\bullet = v_0, \dots, \bullet = v_\infty$.



Wir sehen, dass für beide Startvektoren v_0 die Punkte v_n auf einer Geraden liegen und diese Geraden parallel sind. Können wir dies erklären?

Auftritt Lineare Algebra

Oben hatten wir die beiden Gleichungen für eine Räuber / Beute-Entwicklung

$$\begin{aligned} x_{n+1} &= \frac{1}{5}x_n + \frac{2}{5}y_n && \text{(Räuber),} \\ y_{n+1} &= \frac{13}{10}y_n - \frac{3}{5}x_n = -\frac{3}{5}x_n + \frac{13}{10}y_n && \text{(Beute).} \end{aligned}$$

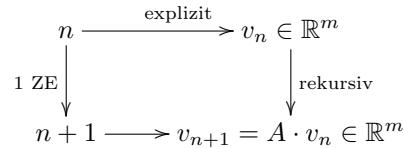
Wir schreiben diese kompakter als ein Matrix-Vektor-Produkt: $v_{n+1} = A \cdot v_n$ mit den Vektoren v_n und v_{n+1} der Länge 2 sowie einer (2×2) -Matrix A :

$$v_{n+1} = \begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix}, \quad v_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix} \quad \text{und} \quad A = \begin{pmatrix} \frac{1}{5} & \frac{2}{5} \\ -\frac{3}{5} & \frac{13}{10} \end{pmatrix}.$$

Die beiden Gleichungen fassen wir dann mit Matrix-Vektor-Produkt $v_{n+1} = A \cdot v_n$ zusammen, denn es gilt:

$$\begin{aligned} A \cdot v_n &= \begin{pmatrix} \frac{1}{5} & \frac{2}{5} \\ -\frac{3}{5} & \frac{13}{10} \end{pmatrix} \cdot \begin{pmatrix} x_n \\ y_n \end{pmatrix} && \text{Definition } A \text{ und } v_n \\ &= \begin{pmatrix} \frac{1}{5}x_n + \frac{2}{5}y_n \\ -\frac{3}{5}x_n + \frac{13}{10}y_n \end{pmatrix} && \text{Definition Matrix-Vektor-Produkt} \\ &= \begin{pmatrix} \frac{1}{5}x_n + \frac{2}{5}y_n \\ \frac{13}{10}y_n - \frac{3}{5}x_n \end{pmatrix} && \text{Vertauschen Summanden} \\ &= \begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} && \text{Gleichungen Räuber/Beute-Modell} \\ &= v_{n+1}. \end{aligned}$$

Wir haben also folgendes Diagramm als Verallgemeinerung des eindimensionalen Falles:



Schauen wir uns die Entwicklungen für unterschiedliche Startvektoren an:

1. Für den Startvektor $v_0 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ gilt $v_1 = Av_0 = v_0$. Wenn wir zu Beginn für jeden Räuber zwei Beutetiere haben, bleiben die Populationen konstant.

Mathematisch sagen wir: Der Startvektor v_0 ist ein Eigenvektor zum Eigenwert 1.

2. Für den Startvektor $\tilde{v}_0 = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$ gilt $v_1 = A\tilde{v}_0 = \frac{1}{2}\tilde{v}_0$ und weiter

$$\begin{aligned} v_2 &= A \cdot v_1 = A \cdot \left(\frac{1}{2}\tilde{v}_0 \right) = \frac{1}{2}A \cdot \tilde{v}_0 = \frac{1}{2} \cdot \frac{1}{2}\tilde{v}_0 \\ &= \left(\frac{1}{2} \right)^2 \cdot \tilde{v}_0 \\ v_3 &= A \cdot v_2 = A \cdot \left(\left(\frac{1}{2} \right)^2 \cdot \tilde{v}_0 \right) = \left(\frac{1}{2} \right)^2 A \cdot \tilde{v}_0 \\ &= \left(\frac{1}{2} \right)^3 \cdot \tilde{v}_0 \\ &\vdots \\ v_n &= \left(\frac{1}{2} \right)^n \tilde{v}_0 \end{aligned}$$

Da $\left(\frac{1}{2}\right)^n \rightarrow 0$ für $n \rightarrow \infty$, sterben die Populationen aus, wenn wir zu Beginn im Verhältnis 4 : 3 mehr Räuber als Beute haben.

Mathematisch bedeutet die Gleichung $A\tilde{v}_0 = \frac{1}{2}\tilde{v}_0$, dass der Vektor \tilde{v}_0 ein Eigenvektor zum Eigenwert $\frac{1}{2}$ ist.

Folgende Fragen aus dem 1. Kapitel können wir nun beantworten:

1. Gibt es Startwerte (x_0, y_0) , bei denen nichts passiert, das heisst,

$$x_n = x_0, \quad y_n = y_0, \quad n = 1, 2, 3, \dots ?$$

Antwort:

Für $v_0 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ gilt $v_1 = Av_0 = v_0$, und damit $x_1 = x_0, y_1 = y_0$, also auch

$$x_n = x_0, \quad y_n = y_0, \quad n = 1, 2, 3, \dots$$

2. Kann es sein, dass sich die Population ändert, aber das Verhältnis nicht, das heisst,

$$\frac{x_n}{y_n} = \frac{x_{n+1}}{y_{n+1}}, \quad y_n \neq 0 \neq y_{n+1}?$$

Antwort:

Für $v_0 = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$ gilt $v_1 = A \cdot v_0 = \frac{1}{2} \cdot v_0$ also

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \frac{1}{2} \cdot \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \implies \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \cdot x_0 \\ \frac{1}{2} \cdot y_0 \end{pmatrix} \implies \frac{x_1}{y_1} = \frac{\frac{1}{2}x_0}{\frac{1}{2}y_0} \implies \frac{x_1}{y_1} = \frac{x_0}{y_0}$$

und damit für alle n

$$\frac{x_n}{y_n} = \frac{x_{n+1}}{y_{n+1}}, \quad y_n \neq 0 \neq y_{n+1}.$$

2.2.2 Vereinfachung mit Hilfe der Linearität

Erinnerung: Seien $w \in \mathbb{R}^m$ und $v_1, v_2, \dots, v_n \in \mathbb{R}^m$.

- Der Vektor w ist eine Linearkombination der Vektoren v_1, v_2, \dots, v_n , falls es Zahlen $\lambda_1, \lambda_2, \dots, \lambda_n$ gibt mit
- $$w = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n.$$
- Lineare Gleichungssysteme liefern eine systematische Antwort auf die Fragen:
 - Können wir w als Linearkombination schreiben?
 - Falls ja: Wie berechnen wir die Zahlen $\lambda_1, \lambda_2, \dots, \lambda_n$?

Beispiel 2.2.1. Seien v_1 und v_2 in \mathbb{R}^2 die Vektoren von oben:

$$v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ und } v_2 = \begin{pmatrix} 4 \\ 3 \end{pmatrix}.$$

Für $w = \begin{pmatrix} 400 \\ 5000 \end{pmatrix}$ erhalten wir eine Linearkombination

$$\begin{pmatrix} 400 \\ 5000 \end{pmatrix} = 3760 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + (-840) \begin{pmatrix} 4 \\ 3 \end{pmatrix} \implies w = 3760v_1 + (-840)v_2,$$

also $\lambda_1 = 3760$ und $\lambda_2 = -840$.

Für $w = \begin{pmatrix} 5000 \\ 5000 \end{pmatrix}$ erhalten wir eine Linearekombination

$$\begin{pmatrix} 5000 \\ 5000 \end{pmatrix} = 1000 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + 1000 \begin{pmatrix} 4 \\ 3 \end{pmatrix} \implies w = 1000v_1 + 1000v_2,$$

also $\lambda_1 = \lambda_2 = 1000$.

Erinnerung: Eine wichtige Rechenregel beim Matrix-Vektor-Produkt ist die Linearität.

Für eine Matrix A , Vektoren v, u und Zahlen α, β gilt

$$A \cdot (\alpha v + \beta u) = \alpha A \cdot v + \beta A \cdot u.$$

Für längere Linearkombinationen gilt dann

$$A \cdot (\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n) = \lambda_1 A \cdot v_1 + \lambda_2 A \cdot v_2 + \dots + \lambda_n A \cdot v_n.$$

Etwas abstrakter sieht die Situation so aus, dass uns eine $(m \times n)$ -Matrix A eine lineare Abbildung

$$\phi_A : \mathbb{R}^n \rightarrow \mathbb{R}^m, \quad v \mapsto \phi_A(v) = A \cdot v$$

liefert. Es gilt also:

$$\phi_A(\alpha v + \beta u) = \alpha \phi_A(v) + \beta \phi_A(u)$$

Anwendung auf unser Räuber/Beute-Beispiel

Linearität hilft uns wie folgt: Wir betrachten unser Räuber / Beute-Modell mit den Startvektoren von oben

- $v_0 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ mit $v_1 = Av_0 = v_0$
- $\tilde{v}_0 = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$ mit $v_1 = A\tilde{v}_0 = \frac{1}{2}\tilde{v}_0$.

Sei w_0 ein Startvektor gegeben als eine Linearkombination von v_0 und \tilde{v}_0 mit $w_0 = \alpha v_0 + \beta \tilde{v}_0$.

1. Nach einer ZE ist wegen der Linearität:

$$w_1 = A \cdot w_0 = A \cdot (\alpha v_0 + \beta \tilde{v}_0) = \alpha A \cdot v_0 + \beta A \cdot \tilde{v}_0 = \alpha \cdot v_0 + \beta \frac{1}{2} \tilde{v}_0$$

2. Nach zwei ZE:

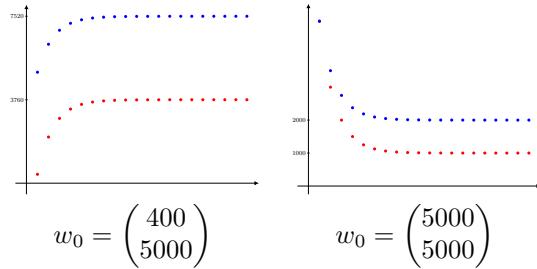
$$\begin{aligned} w_2 &= A \cdot w_1 = A \cdot \left(\alpha \cdot v_0 + \beta \frac{1}{2} \tilde{v}_0 \right) = \alpha \cdot A \cdot v_0 + \beta A \cdot \left(\frac{1}{2} \tilde{v}_0 \right) \\ &= \alpha \cdot v_0 + \beta \left(\frac{1}{2} \right)^2 \tilde{v}_0 \end{aligned}$$

3. Und nach n ZE:

$$w_n = \alpha \cdot v_0 + \beta \left(\frac{1}{2} \right)^n \tilde{v}_0.$$

Da $\left(\frac{1}{2}\right)^n \rightarrow 0$, nähert sich die Population immer weiter dem Vektor $w_\infty = \alpha v_0$ an.

Mit der Linearität verstehen wir auch die vom Computer gezeigten Entwicklungen bei unterschiedlichen Startvektoren w_0 mit • Räuber, • Beute.



Links stabilisiert sich die Entwicklung beim Vektor $w_\infty = \begin{pmatrix} 3760 \\ 7520 \end{pmatrix} = 3760 \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ und rechts bei $w_\infty = \begin{pmatrix} 1000 \\ 2000 \end{pmatrix} = 1000 \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Das hätten wir auch a priori gesehen, wenn wir die Koeffizienten α und β berechnet hätten. Es gilt

Links

$$\begin{pmatrix} 400 \\ 5000 \end{pmatrix} = 3760 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + (-840) \begin{pmatrix} 4 \\ 3 \end{pmatrix} \implies w_0 = 3760v_0 + (-840)\tilde{v}_0$$

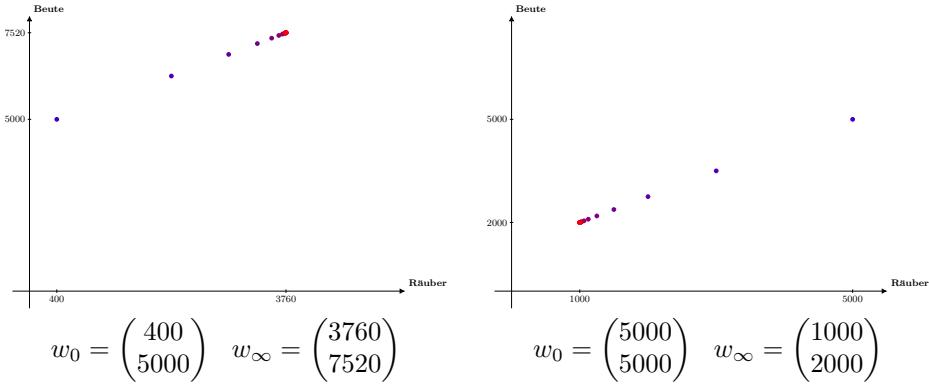
$$\implies w_\infty = 3760v_0$$

Rechts

$$\begin{pmatrix} 5000 \\ 5000 \end{pmatrix} = 1000 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + 1000 \begin{pmatrix} 4 \\ 3 \end{pmatrix} \implies w_0 = 1000v_0 + 1000\tilde{v}_0$$

$$\implies w_\infty = 1000v_0$$

In der Darstellung im (Räuber,Beute)-Koordinatensystem hatten wir mit $\bullet = w_0$, $\circ = w_1 \dots$



Wir sehen, dass für beide Startvektoren w_0 die Punkte w_n jeweils auf einer Geraden liegen und diese Geraden parallel sind. Dies können wir nun verstehen:

1. Bei Start mit $w_0 = \begin{pmatrix} 400 \\ 5000 \end{pmatrix}$ wandern die Punkte w_1, w_2, \dots zum Punkt $w_\infty = \begin{pmatrix} 3760 \\ 7520 \end{pmatrix}$.
2. Bei Start mit $w_0 = \begin{pmatrix} 5000 \\ 5000 \end{pmatrix}$ wandern die Punkte w_1, w_2, \dots zum Punkt $w_\infty = \begin{pmatrix} 1000 \\ 2000 \end{pmatrix}$.
3. Oben hatten wir gesehen: $w_n = \alpha \cdot v_0 + \beta \left(\frac{1}{2}\right)^n \tilde{v}_0$. Die Punkte w_n liegen auf der Geraden mit Parameterform

$$g : r = r_0 + \lambda u = \alpha \cdot v_0 + \lambda \tilde{v}_0.$$

Dabei ist $\alpha = 3760$ für $w_0 = \begin{pmatrix} 400 \\ 5000 \end{pmatrix}$ und $\alpha = 1000$ für $w_0 = \begin{pmatrix} 5000 \\ 5000 \end{pmatrix}$.

4. In beiden Fällen gilt $\left(\frac{1}{2}\right)^n \rightarrow 0$.

Der "Anteil" in Richtung $\tilde{v}_0 = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$ wird im Laufe der Zeit immer kleiner. Die Population nähert sich so immer weiter dem Vektor $\alpha \cdot v_0$ an.

Bemerkung:

Wir halten fest, dass Eigenvektoren helfen, eine Entwicklung $w_{k+1} = A \cdot w_k$ zu untersuchen. Wenn es uns gelingt, einen Startvektor w_0 als Linearkombination von Eigenvektoren zu schreiben.

2.2.3 Explizite Darstellung einer Entwicklung

Sei ein Entwicklungsmodell gegeben: $v_{n+1} = A \cdot v_n$.

Wie bei den eindimensionalen Modellen setzen wir diese Iteration rückwärts fort:

$$\begin{aligned} v_{n+1} &= A \cdot v_n = A \cdot (A \cdot v_{n-1}) = A \cdot (A \cdot (A \cdot v_{n-2})) \\ &= \dots \\ &= A \cdot (A \cdot (A \cdot (\dots (A \cdot v_0)))) \dots \end{aligned}$$

mit

$$\underbrace{A \cdot (A \cdot (A \cdot (\dots A \cdot v_0))) \dots}_{n+1 \text{ Stück}}.$$

Die Rechenregeln der Matrixmultiplikation erlaubt uns, die Klammern anders zu setzen:

$$\underbrace{A \cdot (A \cdot (\dots A)) \dots}_{n+1 \text{ Stück}} = \underbrace{(A \cdot A \cdot A \cdot \dots A)}_{n+1 \text{ Stück}} = A^{n+1}.$$

Vor allem können wir v_{n+1} so berechnen: $v_{n+1} = A \cdot v_n = A^{n+1} \cdot v_0$ und erhalten damit eine explizite Darstellung der Entwicklung.

Zyklische oder periodische Entwicklung

Gegeben sei eine Entwicklung mit $v_n = A \cdot v_{n-1} = A^n \cdot v_0$. Angenommen, es gibt eine Potenz k mit $A^k = E_n$. Dann gilt

$$v_{n+k} = A^{n+k} \cdot v_0 = A^n A^k \cdot v_0 = A^n E_n \cdot v_0 = A^n \cdot v_0 = v_n.$$

Nach k Zeitschritten erhalten wir also wieder den Vektor v_n . Die Entwicklung zeigt eine periodisches Verhalten mit einer Periode der Länge k .

Beispiel 2.2.2. Sei $A = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{3} \\ 6 & 0 & -6 \\ 0 & \frac{1}{2} & 0 \end{pmatrix}$. Dann ist $A^3 = E_3$. (Nachrechnen!)

Für eine Entwicklung $v_n = A \cdot v_{n-1} = A^n \cdot v_0$ gilt dann das periodische Verhalten:

$$v_3 = A^3 \cdot v_0 = v_0 \text{ und damit } v_{n+3} = v_n.$$

Rückwärts mit der inverse Matrix

Ein Entwicklungsmodell sei durch eine invertierbare Matrix A gegeben:

$$v_{n+1} = A \cdot v_n.$$

Mit der inversen Matrix können wir die Entwicklung rückwärts durchlaufen und für v_{n+1} den Vorgänger v_n berechnen:

$$\begin{aligned} v_{n+1} = A \cdot v_n &\implies A^{-1} \cdot v_{n+1} = A^{-1}(A \cdot v_n) \\ &\implies A^{-1} \cdot v_{n+1} = \underbrace{(A^{-1} \cdot A)}_{=E_n} \cdot v_n \end{aligned}$$

und damit $v_n = A^{-1} \cdot v_{n+1}$.

Iterationen mit Eigenwerten/-vektoren

Die Beobachtungen in dem Beispiel oben werden wir nun verallgemeinern und dabei auf die Theorie der Eigenwerte und -vektoren zurückgreifen. Unser Ziel ist weiterhin, damit eine Entwicklung $w_{k+1} = A \cdot w_k$ mit Hilfe der Eigenschaften der Matrix A zu verstehen.

Sei $A = \begin{pmatrix} \frac{1}{5} & \frac{2}{5} \\ -\frac{3}{5} & \frac{13}{10} \end{pmatrix}$. Wie oben gesehen gilt $Av_0 = v_0$ für $v_0 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ und für $\tilde{v}_0 = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$ gilt $A\tilde{v}_0 = \frac{1}{2}\tilde{v}_0$.

Wir erinnern an die allgemeine

Definition 2.2.3. Sei A eine quadratische Matrix.

1. Eine Zahl λ heisst Eigenwert von A (EW), wenn es einen Vektor v mit $v \neq 0$ und

$$A \cdot v = \lambda v$$

gibt. Dabei ist λ eine komplexe Zahl, die allenfalls reell ist.

2. Jeder Vektor v mit $v \neq 0$ heisst Eigenvektor von A (EV) zum Eigenwert λ , wenn er die Gleichung $A \cdot v = \lambda v$ erfüllt.

re

Beispiel 2.2.4. Die Matrix $A = \begin{pmatrix} \frac{1}{5} & \frac{2}{5} \\ -\frac{3}{5} & \frac{13}{10} \end{pmatrix}$ hat die Eigenwerte 1 und $\frac{1}{2}$.

Der Vektor $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ ist ein Eigenvektor zum EW 1, der Vektor $\begin{pmatrix} 4 \\ 3 \end{pmatrix}$ ein Eigenvektor zum EW $\frac{1}{2}$.

Sei $v_{n+1} = A \cdot v_n$ ein 2×2 -Entwicklungsmodell mit einem Startvektor v_0 , von dem wir wissen, dass er ein EV ist, das heisst, es gibt eine Zahl λ mit $A \cdot v_0 = \lambda v_0$. Es ist $v_1 = A \cdot v_0 = \lambda v_0$ und weiter

$$\begin{aligned} v_2 &= A \cdot v_1 = A \cdot (\lambda v_0) = \lambda A \cdot v_0 = \lambda \cdot \lambda \cdot v_0 = \lambda^2 \cdot v_0 \\ v_3 &= A \cdot v_2 = A \cdot (\lambda^2 v_0) = \lambda^2 A \cdot v_0 = \lambda^3 \cdot v_0 \\ &\vdots \\ v_n &= \lambda^n v_0. \end{aligned}$$

Ausserdem halten wir noch fest:

Bemerkungen:

1. Sei v ein EV von A zum EW λ , dann ist auch αv ein EV zum EW λ , denn

$$A(\alpha v) = \alpha A(v) = \alpha \lambda v = \lambda(\alpha v).$$

Seien v_1 und v_2 EV von A zum gleichen EW λ , dann ist auch $w = \alpha v_1 + \beta v_2$ ein EV zum EW λ , denn

$$\begin{aligned} Aw &= A(\alpha v_1 + \beta v_2) = \alpha A(v_1) + \beta A(v_2) = \alpha \lambda v_1 + \beta \lambda v_2 = \lambda(\alpha v_1 + \beta v_2) \\ &= \lambda w. \end{aligned}$$

2. Sei λ ein k -facher EW. Dann gibt es höchstens k linear unabhängige EV zum EW λ .

3. Die EV zu unterschiedlichen EW sind linear unabhängig.

Zusammenfassung 2×2 -Matrix

Betrachten wir zunächst den Spezialfall $n = 2$. Im allgemeinen Fall müssen wir später Restriktionen an die Matrix A beachten.

Sei A eine 2×2 -Matrix mit v_1 einem EV zum EW λ_1 und v_2 einem EV zum EW $\lambda_2 \neq \lambda_1$.

Sei w_0 ein Startvektor. Da $\lambda_1 \neq \lambda_2$, sind v_1 und v_2 linear unabhängig. Damit gibt es Zahlen α_1 und α_2 mit $w_0 = \alpha_1 v_1 + \alpha_2 v_2$. Nach n ZE ist wegen der Linearität:

$$w_n = A \cdot w_{n-1} = \alpha_1 \cdot \lambda_1^n \cdot v_1 + \alpha_2 \cdot \lambda_2^n \cdot v_2.$$

- Für $|\lambda_1| < 1, |\lambda_2| < 1$ sind $\lambda_1^n, \lambda_2^n \rightarrow 0$ und damit $w_n \rightarrow 0$.

Populationen sterben aus.

- Für $\lambda_1 = 1$ und $|\lambda_2| < 1$ sind $\lambda_1^n = 1$ und $\lambda_2^n \rightarrow 0$ und damit $w_n \rightarrow \alpha_1 v_1$:

Gleichgewicht der Populationen.

- Für $|\lambda_1| > 1$ und $|\lambda_2| < 1$ sind $\lambda_1^n \rightarrow \infty$ und $\lambda_2^n \rightarrow 0 \implies w_n \rightarrow \infty$.

Populationen "explodieren" in Richtung v_1 .

- Allgemeiner gilt noch:

Sei $|\lambda_1| > |\lambda_2|$. Für jeden Startvektor $w_0 = \alpha_1 v_1 + \alpha_2 v_2$ mit $\alpha_1 \neq 0$ nähern sich die Vektoren $w_1, w_2, \dots, w_n, \dots$ immer mehr der Richtung von v_1 an.

Um dies noch zu präzisieren, betrachten wir anstatt der Folge w_n die Folge der normierten Vektoren: $\frac{w_0}{|w_0|}, \frac{w_1}{|w_1|}, \frac{w_2}{|w_2|}, \dots$. Die Folge $\left(\frac{w_k}{|w_k|} \right)$ konvergiert gegen den normierten EV $\frac{v_1}{|v_1|}$ zum EW λ_1 .

2.2.4 Beispiele Altersstrukturen 3×3

Kommen wir zurück zum Beispiel der Verteilung einer Altersstruktur in der Einleitung: Der Vektor v_n beschreibe zum Beispiel die Altersverteilung in einer Population zur Zeit n :

x_n = Anzahl der Zellen, welche genau 1 ZE überlebt haben: 1. Altersstufe.

y_n = Anzahl der Zellen, welche genau 2 ZE überlebt haben: 2. Altersstufe.

z_n = Anzahl der Zellen, welche genau 3 ZE überlebt haben: 3. Altersstufe.

Ein Entwicklungsmodell sei wie folgt gegeben: Für $i = 1, 2, 3$ ist

r_i = Durchschnittliche Anzahl der Zellen, welche aus einer Zelle der i -ten Altersstufe entstehen.
Gezählt werden nur die neuen Zellen, welche 1 ZE überleben.

q_i = Anteil der Zellen, welche eine ZE überleben und in die $(i+1)$ -te Altersstufe übergehen.
Zellen überleben höchstens 3 ZE.

Insgesamt hatten wir die drei Gleichungen

$$x_{n+1} = r_1 x_n + r_2 y_n + r_3 z_n$$

$$y_{n+1} = q_1 x_n$$

$$z_{n+1} = q_2 y_n,$$

welche uns den Vektor $v_{n+1} = \begin{pmatrix} x_{n+1} \\ y_{n+1} \\ z_{n+1} \end{pmatrix}$ liefern. Mit Hilfe der Matrizenrechnung können wir dieses Verfahren wie oben angekündigt elegant beschreiben.

In Matrix-Vektor-Schreibweise folgt $v_{n+1} = \begin{pmatrix} r_1 & r_2 & r_3 \\ q_1 & 0 & 0 \\ 0 & q_2 & 0 \end{pmatrix} v_n$.

Der Computer zeigt bei einigen Zahlenbeispiele folgende Phänomene, welche wir mit EW und EV verstehen und untersuchen werden.

	Matrix	Startvektor	Entwicklung
1.	$\begin{pmatrix} 0 & 0 & 6 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{pmatrix}$	$\begin{pmatrix} 12 \\ 12 \\ 12 \end{pmatrix}$	Zyklus mit 3 ZE
2.	$\begin{pmatrix} 0 & 0 & 6 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{pmatrix}$	$\begin{pmatrix} 24 \\ 12 \\ 4 \end{pmatrix}$	Gleichgewichtszustand $\begin{pmatrix} 24 \\ 12 \\ 4 \end{pmatrix}$
3.	$\begin{pmatrix} 0 & 1 & 3 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{pmatrix}$	$\begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}$	Konvergenz zum Gleichgewicht $\begin{pmatrix} 24 \\ 12 \\ 4 \end{pmatrix}$
4.	$\begin{pmatrix} 0 & 2 & 3 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$	Die Folgen $\left(\frac{x_{n+1}}{x_n} \right), \left(\frac{y_{n+1}}{y_n} \right), \left(\frac{z_{n+1}}{z_n} \right)$ konvergieren jeweils gegen eine feste Zahl λ Die Folge $\left(\frac{v_n}{ v_n } \right)$ der normierten v_n konvergiert.

Die ersten beiden Entwicklungen 1. und 2. untersuchen wir in den Aufgaben.

Im 3. Beispiel erhalten wir komplexe Eigenwerte. Mit diesen können wir am Ende dieses Kapitels die Entwicklung dann weiter beschreiben und verstehen.

Es ist $p_A(\lambda) = -\lambda^3 + \frac{1}{2}\lambda + \frac{1}{2} = (\lambda - 1)(-\lambda^2 - \lambda - \frac{1}{2})$. Dann ist ein EW $\lambda_1 = 1$ und die weiteren die Nullstellen von $-\lambda^2 - \lambda - \frac{1}{2}$.

Nun ist aber

$$\begin{aligned} -\lambda^2 - \lambda - \frac{1}{2} = 0 &\implies \lambda^2 + \lambda + \frac{1}{2} = 0 \\ &\implies \lambda_{2,3} = -\frac{1}{2} \pm \sqrt{\frac{1}{4} - \frac{1}{2}} = -\frac{1}{2} \pm \sqrt{-\frac{1}{4}}. \end{aligned}$$

Um die Wurzel $\sqrt{-\frac{1}{4}}$ aus einer negativen Zahl zu ziehen, benötigen wir Komplexe Zahlen.

Beispiel mit dominantem Eigenwert

Diskutieren wir das 4. Beispiel mit $A = \begin{pmatrix} 0 & 2 & 3 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \end{pmatrix}$ und $v_0 = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. Diese definieren eine Iteration: $v_{n+1} = \begin{pmatrix} x_{n+1} \\ y_{n+1} \\ z_{n+1} \end{pmatrix} = A \cdot v_n = A^{n+1} \cdot v_0$.

Wann und wieso konvergieren die Folgen $\left(\frac{x_{n+1}}{x_n}\right)$, $\left(\frac{y_{n+1}}{y_n}\right)$, $\left(\frac{z_{n+1}}{z_n}\right)$ jeweils gegen eine feste Zahl λ ?

Für grosse n ist $\frac{x_{n+1}}{x_n} \approx \lambda$, $\frac{y_{n+1}}{y_n} \approx \lambda$, $\frac{z_{n+1}}{z_n} \approx \lambda$ oder

$$x_{n+1} \approx \lambda \cdot x_n, \quad y_{n+1} \approx \lambda \cdot y_n, \quad z_{n+1} \approx \lambda \cdot z_n$$

und als Vektoren

$$A \cdot v_n = v_{n+1} = \begin{pmatrix} x_{n+1} \\ y_{n+1} \\ z_{n+1} \end{pmatrix} \approx \begin{pmatrix} \lambda x_n \\ \lambda y_n \\ \lambda z_n \end{pmatrix} = \lambda \cdot \begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix} = \lambda v_n.$$

Das heisst, dass für grosse n der Vektor v_n immer mehr ein EV von A zum EW λ wird, und dass die Vektoren v_0, v_1, v_2, \dots im Laufe immer mehr in Richtung eines EV zeigen.

Um das Experiment und die Überlegung zu bestätigen, geben wir Folgendes an: (Nachrechnen!)

Charakteristisches Polynom $p_A(\lambda) = -\lambda^3 + \lambda + \frac{3}{8}$,

Eigenwerte $\lambda_1 = -\frac{1}{2}$, $\lambda_2 = \frac{1}{4}(1 - \sqrt{13})$, $\lambda_3 = \frac{1}{4}(1 + \sqrt{13})$.

Als Eigenvektoren wählen wir

$$v_1 = \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 7 - \sqrt{13} \\ 1 - \sqrt{13} \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 7 + \sqrt{13} \\ 1 + \sqrt{13} \\ 1 \end{pmatrix}.$$

Übung: Es gilt $A \cdot v_i = \lambda_i \cdot v_i$ für $i = 1, 2, 3$.

Entwicklung

1. Die drei EW sind paarweise voneinander verschieden \implies 3 EV $v_1, v_2, v_3 \in \mathbb{R}^3$ sind linear unabhängig \implies für jeden Startvektor w_0 gibt es $\alpha_1, \alpha_2, \alpha_3$ mit

$$w_0 = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3$$

2. Es gilt wieder mit der Linearität: $w_n = \alpha_1 \lambda_1^n \cdot v_1 + \alpha_2 \lambda_2^n \cdot v_2 + \alpha_3 \lambda_3^n \cdot v_3$.
3. Es gilt: $|\lambda_1|, |\lambda_2| < 1$ und $|\lambda_3| > 1$. Damit $\lambda_1^n, \lambda_2^n \rightarrow 0$ und $\lambda_3^n \rightarrow \infty$

4. Falls $\alpha_3 = 0 \implies w_n \rightarrow 0$.

Falls $\alpha_3 \neq 0 \implies w_n \rightarrow \alpha_3 \lambda_3^n \cdot v_3$, das heisst, w_n konvergiert in Richtung des EV v_3 .

Um dies noch zu präzisieren, betrachten wir anstatt der Folge w_n die Folge der normierten Vektoren: $\frac{w_0}{\|w_0\|}, \frac{w_1}{\|w_1\|}, \frac{w_2}{\|w_2\|}, \dots$

Diese Folge $\left(\frac{w_k}{\|w_k\|}\right)$ konvergiert gegen den normierten EV $\frac{v_3}{\|v_3\|}$ zum EW λ_3 .

Beispiel 2.2.5. Seien $\alpha_1 = \alpha_2 = 1$ und $\alpha_3 = 0$, dann ist

$$w = v_1 + v_2 = \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix} + \begin{pmatrix} 7 - \sqrt{13} \\ 1 - \sqrt{13} \\ 1 \end{pmatrix} = \begin{pmatrix} 9 - \sqrt{13} \\ -1 - \sqrt{13} \\ 2 \end{pmatrix}.$$

2.2.5 Zusammenfassung $n \times n$ -Matrix

Gegeben sei eine $n \times n$ -Matrix A .

Wir können die Entwicklung $w_{k+1} = A \cdot w_k$ mit Hilfe der Eigenwerte und EV bestimmen, wenn wir die Voraussetzungen an die Matrix A gut unterscheiden.

Nehmen wir an, dass es n linear unabhängige EV von A gibt: v_1, v_2, \dots, v_n , mit EW $\lambda_1, \lambda_2, \dots, \lambda_n$. Die λ_i sind dabei nicht unbedingt voneinander verschieden.

Sei w_0 ein Startvektor. Da die EV $v_i, i = 1, \dots, n$ linear unabhängig sind, gibt es Zahlen $\alpha_1, \alpha_2, \dots, \alpha_n$ mit $w_0 = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$.

Nach k Zeiteinheiten gilt wieder wegen der Linearität:

$$w_k = \alpha_1 \cdot \lambda_1^k v_1 + \alpha_2 \lambda_2^k v_2 + \dots + \alpha_n \lambda_n^k v_n.$$

1. Falls für jeden EW $|\lambda_i| < 1$ gilt, folgt $\lambda_i^k \rightarrow 0$ und damit $w_k \rightarrow 0$. Dies gilt auch, wenn λ_i komplexe Zahlen sind.

Die Populationen sterben aus.

2. Sei zum Beispiel $\lambda_1 = 1$ ein einfacher EW, und für alle anderen EW λ_i gelte $|\lambda_i| < 1$.

Dann gilt $\lambda_1^k = 1$ und $\lambda_i^k \rightarrow 0$ und damit $w_k \rightarrow \alpha_1 v_1$. Dies gilt auch, wenn $\lambda_i, i = 2, 3, \dots, n$ komplexe Zahlen sind.

Es gibt ein Gleichgewicht der Populationen.

3. Seien zum Beispiel λ_1 ein einfacher reeller EW mit $\lambda_1 > 1$, und für alle anderen EW λ_i gelte $|\lambda_i| < 1$.

Dann folgt $\lambda_1^k \rightarrow \infty$ und für die anderen EW $\lambda_i^k \rightarrow 0$ und damit $w_k \rightarrow " \infty \cdot v_1 "$. Dies gilt auch, wenn $\lambda_i, i = 2, 3, \dots, n$ komplexe Zahlen sind.

Die Populationen explodieren in Richtung v_1 .

4. Allgemeiner gilt wieder (vergleiche den 2×2 -Fall):

Sei zum Beispiel λ_1 ein einfacher EW mit $|\lambda_1| > |\lambda_i|, i = 2, 3, 4, \dots, n$.

Für jeden Startvektor $w_0 = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$ mit $\alpha_1 \neq 0$ nähern sich die Vektoren $w_1, w_2, \dots, w_k, \dots$ immer mehr der Richtung von v_1 an. Das heisst wieder:

Die Folge $\left(\frac{w_k}{\|w_k\|} \right)$ konvergiert gegen den normierten EV $\frac{v_1}{\|v_1\|}$. Der EW λ_1 *dominiert*.

Dies gilt auch, wenn λ_i komplexe Zahlen sind.

Anwendungen Komplexe EW / EV

Im 3. Beispiel

Matrix	Startvektor	Entwicklung
$\begin{pmatrix} 0 & 1 & 3 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{pmatrix}$	$\begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}$	Konvergenz zum Gleichgewicht $\begin{pmatrix} 24 \\ 12 \\ 4 \end{pmatrix}$

hatten wir

$$p_A(\lambda) = -\lambda^3 + \frac{1}{2}\lambda + \frac{1}{2} = (\lambda - 1)(-\lambda^2 - \lambda - \frac{1}{2}),$$

mit EW $\lambda_1 = 1$ und $\lambda_{2,3} = -\frac{1}{2} \pm \sqrt{-\frac{1}{4}}$.

Mit komplexen Zahlen erhalten wir also

$$\lambda_1 = 1 \in \mathbb{R}, \quad \lambda_2 = -\frac{1}{2} + \frac{1}{2}i \in \mathbb{C}, \quad \lambda_3 = -\frac{1}{2} - \frac{1}{2}i \in \mathbb{C}$$

und der Betrag der EW ist

$$|\lambda_2| = \left| -\frac{1}{2} + \frac{1}{2}i \right| = \sqrt{\frac{1}{2}} = |\lambda_3| = \left| -\frac{1}{2} - \frac{1}{2}i \right| < 1.$$

Was bedeutet dies für die Entwicklung?

Es ist $\lambda_2 = re^{i\varphi}$ mit $r < 1$, da $r = |\lambda_2| < 1$. Für $n \rightarrow \infty$, ist $r^n \rightarrow 0$. Der Radius wird immer kleiner und die Folge der komplexen Zahlen $\lambda_2, \lambda_2^2, \lambda_2^3, \dots$ dreht sich um den Nullpunkt spiralförmig. Für λ_3 gelten analoge Überlegungen.

Geben wir noch Eigenvektoren an:

$$v_1 = \begin{pmatrix} 6 \\ 3 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} -3i \\ -\frac{3}{2} + \frac{3}{2}i \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 3i \\ -\frac{3}{2} - \frac{3}{2}i \\ 1 \end{pmatrix}$$

Entwicklung

1. Die drei EW sind paarweise voneinander verschieden \implies 3 EV $v_1, v_2, v_3 \in \mathbb{C}^3$ sind linear unabhängig \implies für jeden Startvektor w_0 gibt es $\alpha_1, \alpha_2, \alpha_3$ mit

$$w_0 = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3.$$

2. Es gilt wieder mit Linearität

$$w_n = \alpha_1 \lambda_1^n \cdot v_1 + \alpha_2 \lambda_2^n \cdot v_2 + \alpha_3 \lambda_3^n \cdot v_3.$$

3. Es gilt: $\lambda_1 = 1$ und $|\lambda_2| < 1, |\lambda_3| < 1$.

Damit $\lambda_1^n = 1$ konstant für alle n und $\lambda_2^n, \lambda_3^n \rightarrow 0$.

4. Falls $\alpha_1 = 0 \implies w_n \rightarrow 0$.

Falls $\alpha_1 \neq 0 \implies w_n \rightarrow \alpha_1 \lambda_1^n \cdot v_1 = \alpha_1 \cdot v_1$, das heisst, w_n konvergiert zum Vektor αv_1 .

Der Computer zeigt zum Beispiel:

- Starten wir mit $w_0 = \begin{pmatrix} 0 \\ -3 \\ 2 \end{pmatrix} = v_2 + v_3$. Dann konvergiert die Folge $w_1, w_2, w_3, \dots, w_n, \dots$ gegen den Nullvektor.
- Mit Startvektor $w_0 = \begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}$ konvergiert die Folge $w_1, w_2, w_3, \dots, w_n, \dots$ gegen den Vektor $\begin{pmatrix} 24 \\ 12 \\ 4 \end{pmatrix} = 4 \cdot \begin{pmatrix} 6 \\ 3 \\ 1 \end{pmatrix}$.

Um dies a priori zu bestimmen, müssen wir in der Gleichung

$$w_0 = \begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix} = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3,$$

den Koeffizienten α_1 bestimmen, das heisst, das LGS $C \cdot x = b$ lösen, mit

$$C = \begin{pmatrix} 6 & -3i & 3i \\ 3 & -\frac{3}{2} + \frac{3}{2}i & -\frac{3}{2} - \frac{3}{2}i \\ 1 & 1 & 1 \end{pmatrix}, \quad \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}, \quad b = \begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}.$$

$$\text{Es ist } \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 - \frac{7}{3}i \\ 3 + \frac{7}{3}i \end{pmatrix}.$$

KAPITEL 3

Lineare Modelle

3.1 Eindimensionale Lineare Modelle

Wir beginnen mit Beispielen für (Gewöhnlichen) Differentialgleichungen, welche helfen, eindimensionale (kontinuierliche) Populationsentwicklungen zu beschreiben.

Dafür betrachten wir eine differenzierbare Funktion $f : \mathbb{R} \rightarrow \mathbb{R}, t \mapsto f(t)$. Diese bezeichnen wir oft mit $f = N$ und damit $f(t) = N(t)$. Den Wert $N(t)$ interpretieren wir als die Grösse der Population zur Zeit t und die Ableitung $N'(t)$ als die momentane Veränderungsgeschwindigkeit.

3.1.1 Exponentielle Veränderungen

Sei $t \mapsto N(t)$ eine kontinuierliche Entwicklung mit der Ausgangspopulation $N(0) = N_0$ und konstanter Veränderungsrate. Das heisst, für alle t gilt $\frac{N'(t)}{N(t)} = r$.

Dies ist ein erstes Beispiel einer Gewöhnlichen Differentialgleichung (DGL). Schreiben wir die Gleichung um:

$$\frac{N'(t)}{N(t)} = r \implies N'(t) = r \cdot N(t) \text{ oder kurz } N' = rN.$$

Das bedeutet anschaulich: Zur jeder Zeit t ist die Veränderungsgeschwindigkeit proportional zur aktuellen Grösse. Die Gleichung charakterisiert sich durch:

- Eine Lösung dieser DGL ist eine Funktion f mit der Eigenschaft:
Die Ableitungsfunktion f' ist gleich der Ausgangsfunktion multipliziert mit einem konstanten Faktor: $r \cdot f$.
- Als Lösung einer DGL suchen wir also eine Funktion und nicht eine Zahl, welche diese DGL erfüllt.
- Im Gegensatz zu einer Gleichung der Form $F'(t) = \frac{1000}{1 + \frac{1}{2}t}$ taucht hier in der Gleichung $N'(t) = r \cdot N(t)$ auf der rechten Seite auch die gesuchte Funktion $N : t \mapsto N(t)$ auf. Wir

können also nicht einfach eine Stammfunktion für die rechte Seite suchen:

$$F(t) = \int F'(t)dt = \int \frac{1000}{1 + \frac{1}{2}t} dt = 2000 \ln \left(1 + \frac{1}{2}t \right) + C, C = \text{Konstante.}$$

Allgemeine Lösung

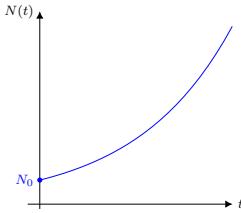
Sei zunächst $r = 1$, damit suchen wir eine Funktion N mit $N' = N$.

Wir erinnern uns, dass dies für die Exponentialfunktion gilt: $(e^t)' = e^t$. Setze $N(t) = e^t$. Dann ist $N'(t) = (e^t)' = e^t = N(t)$. Für eine Konstante c betrachten wir $N(t) = c \cdot e^t$. Die Funktion erfüllt auch die DGL, denn $N'(t) = (c \cdot e^t)' = c \cdot (e^t)' = c \cdot e^t = N(t)$.

Um eine eindeutige Lösung zu finden, verwenden wir die Anfangsbedingung: $N(0) = N_0$. Damit stellen wir ein Anfangswertproblem (AWP), bei welchem wir eine DGL $N' = N$ und einen Startwert $N(t_0) = N_0$ haben.

Dann ist hier $N(t_0) = N(0) = c \cdot e^0 = c \implies N_0 = c$ und $N(t) = N_0 \cdot e^t$.

Genauso erhalten wir für $N'(t) = r \cdot N(t)$ als Lösung $N(t) = N_0 \cdot e^{rt}$ mit Graphen der Form:



Eindeutigkeit

Aussagen über Existenz und Eindeutigkeit einer Lösung einer DGL oder eines Anfangswertproblems (AWP) erfordern mehr Theorie. Bei einer Modellierung ist es sinnvoll oder vielleicht zwingend, dass wir mit einer eindeutigen Lösung arbeiten, oder zumindest entscheiden können, ob es allenfalls weitere Möglichkeiten geben könnte.

Als ein elementares Beispiel schauen wir uns die exponentielle Entwicklung an und fragen uns, wieso die Exponentialfunktion die einzige Lösung der Gleichung $N' = N$ liefert?

Sei f eine weitere Funktion, welche die Gleichung $f' = f$ erfüllt. Sei g eine Hilfsfunktion mit $g(t) = f(t) \cdot e^{-t}$. Dann ist

$$\begin{aligned} g'(t) &= (f(t) \cdot e^{-t})' \\ &= f'(t) \cdot e^{-t} + f(t) \cdot (-e^{-t}) && \text{Produkt- und Kettenregel} \\ &= f'(t) \cdot e^{-t} - f'(t) \cdot e^{-t} && f' = f \\ &= 0 \\ \implies g(t) &= C \text{ konstant} = f(t) \cdot e^{-t} && \text{Definition von } g(t) \\ \implies f(t) &= C \cdot e^t \end{aligned}$$

Mit der Anfangsbedingung $f(0) = N_0$ ist dann $f(t) = N_0 \cdot e^t$.

Gleiches gilt für $N' = rN$ mit Startwert $N(t_0) = N_0$: Auch hier liefert die Exponentialfunktion die eindeutige Lösung: $f(t) = N_0 \cdot e^{rt}$.

3.1.2 Beschränktes Wachstum

Exponentielles Wachstum ist eine Idealisierung. In der Natur spielen Begrenzung des Lebensraumes, der Nahrung etc. eine wichtige Rolle. Daher führen wir eine Korrektur ein. Wir erhalten eine DGL der Form

$$N'(t) = r(K - N(t)), \quad 0 < N(0) = N_0 < K \text{ und } r > 0$$

Dies ist nach $N' = rN$ ein weiteres Beispiel einer Linearen DGL 1. Ordnung.

Eine Lösung dieser DGL ist eine Funktion f mit der Eigenschaft:

- Auf der linken Seite steht die Ableitungsfunktion f' .
- Setzen wir f auf der rechten Seite ein, erhalten wir $r(K - f)$.
- Oder anders gesagt: Für jedes t aus den Definitionsbereichen D_f und $D_{f'}$ gilt:

$$f'(t) = r(K - f(t)).$$

- Für $t = 0$ muss $f(0)$ einen vorgegeben Wert N_0 annehmen.
- Als Lösung einer DGL suchen wir also wieder eine Funktion und nicht eine Zahl, welche diese DGL erfüllt.

Ohne die¹ Lösung explizit zu kennen, können wir qualitativ das Verhalten beschreiben:

1. Es ist nach Voraussetzung $0 < N(0) < K$ und somit $K - N(0) > 0$. Mit $r > 0$ ist

$$N'(0) = r(K - N(0)) > 0 \quad \text{und} \quad N'(t) = r(K - N(t)) > 0,$$

solange $N(t) < K$, und die Population N ist streng monoton wachsend, solange $N(t) < K$.

2. In der Regel ist zu Beginn $N(t)$ viel kleiner als K , also $K - N(t) \approx K$:

$$\begin{aligned} N'(t) &= r(K - N(t)) \approx rK \implies \text{konstantes Wachstum} \\ &\implies N(t) \approx (r \cdot K) \cdot t + C \text{ lineare Entwicklung} \end{aligned}$$

3. Mit $N(t) \rightarrow K$ folgt $K - N(t) \rightarrow 0$, und wir erhalten immer langsameres Wachstum, da die rechte Seite der DGL $N'(t) = r(K - N(t))$ immer kleiner wird. Also wird die Geschwindigkeit N' immer kleiner.
4. Falls $N(t_\infty) = K$, folgt $N'(t_\infty) = r(K - K) = 0$ und es gibt kein Wachstum mehr. Damit ist K eine obere Schranke (maximale Kapazität). Die Population, beschrieben durch N , wächst im Laufe der Zeit und nähert sich immer weiter der Asymptote $N(t_\infty) = K$ an.

¹Wir nehmen hier und im Verlauf an, dass eine Lösung existiert und eindeutig ist.

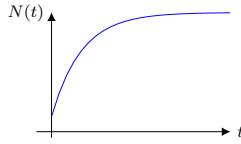
5. Wie verändert sich allenfalls die Wachstumsgeschwindigkeit N' ? Wo liegen allfällige Wendepunkte? Dafür rechnen wir $N''(t)$ aus

$$\begin{aligned} N''(t) &= (N')'(t) = (r(K - N(t)))' = -rN'(t) \stackrel{\text{DGL}}{=} -r(r(K - N(t))) \\ &= -r^2(K - N(t)) \text{ mit der einzigen Nullstelle } K = N(t_\infty). \end{aligned}$$

Es gibt somit keine Wendepunkte für $t < t_\infty$.

Die Lösung berechnet sich mit allgemeinen Methoden, s.u.

$$N(t) = (N_0 - K) \cdot e^{-rt} + K$$



Probe: Es sind

$$N(0) = (N_0 - K) \cdot e^{-r \cdot 0} + K = (N_0 - K) \cdot 1 + K = N_0 - K + K = N_0.$$

und

$$\begin{aligned} N'(t) &= ((N_0 - K) \cdot e^{-rt} + K)' \\ &= (N_0 - K) \cdot (e^{-rt})' \\ &= (N_0 - K) \cdot (-r)e^{-rt} \\ &= (N_0 - K) \cdot (-r)e^{-rt} - rK + rK \\ &= r(-((N_0 - K) \cdot e^{-rt} + K)) + rK \\ &= r(-N(t)) + rK \\ &= r(K - N(t)). \end{aligned}$$

3.1.3 Zusammenfassung Lineare DGL mit konstanten Koeffizienten

Definition 3.1.1. Eine inhomogene Lineare DGL mit **konstanten Koeffizienten** ist eine Gleichung der Form

$$y' = ay + b, \text{ mit Konstanten } a, b.$$

Falls $b = 0$, heisst die DGL homogen.

Wir übersetzen die beiden Beispiele oben in diese Notation:

Beispiele

Exponentiell: Für $N' = rN$ sind $N = y, a = r$ und $b = 0$.

Beschränktes Wachstum: Für $N' = r(K - N)$ rechnen wir und erhalten nach Auflösen der Klammer $N' = rK - rN = -rN + rk$, und es sind $a = -r$ und $b = rK$.

Lösungen

Wir geben nun unterschiedliche Fälle an, in denen Funktionen eine lineare DGL mit konstanten Koeffizienten der Form $y' = ay + b$ lösen.

Stammfunktion Falls $a = 0$ haben wir die Gleichung $y' = b$, die wir direkt mit der Integralrechnung lösen $y = \int b dt = b \cdot t + C$.

Stationäre Lösung

Eine stationäre Lösung von $y' = ay + b$ ist eine konstante Lösung y_∞ mit $y'_\infty = 0$. Damit ist (für $a \neq 0$)

$$0 = ay_\infty + b \implies y_\infty = -\frac{b}{a}.$$

Homogene Lösung

Falls $b = 0$, d.h. die DGL ist homogen, haben wir die allgemeine Lösung $y_H(x) = C \cdot e^{ax}$.

Allgemeine Lösung

Eine stationäre Lösung zusammen mit der Lösung der *homogenen* Lösung liefern die allgemeine Lösung

$$y(x) = y_H + y_\infty = C \cdot e^{ax} - \frac{b}{a}.$$

Die Probe wie oben ergibt:

$$\begin{aligned} y'(x) &= \left(C \cdot e^{ax} - \frac{b}{a} \right)' = Ca \cdot e^{ax} && \text{Kettenregel und Konstante verschwindet} \\ &= a(C \cdot e^{ax}) = a \left(C \cdot e^{ax} - \frac{b}{a} + \frac{b}{a} \right) && \text{Erweiterung um Null} \\ &= a \left(y(x) + \frac{b}{a} \right) && \text{Einsetzen der Lösung } y(x) = C \cdot e^{ax} - \frac{b}{a} \\ &= ay(x) + b && \text{Ausklammern} \end{aligned}$$

Das heisst jede Lösung der DGL ist von dieser Form.

Anfangswertproblem

Geben wir noch einen Start $y(0) = y_0$ vor, so erhalten wir für die Konstante C oben

$$y(0) = y_0 = C \cdot e^0 - \frac{b}{a} = C - \frac{b}{a} \implies C = y_0 + \frac{b}{a}.$$

Setzen wir dieses C ein, erhalten wir

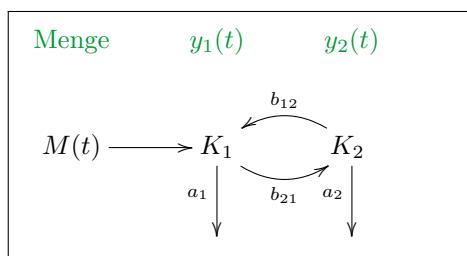
$$y(x) = \left(y_0 + \frac{b}{a} \right) \cdot e^{ax} - \frac{b}{a} = y_0 \cdot e^{ax} + \frac{b}{a} (e^{ax} - 1).$$

3.2 Höherdimensionale Modelle

In diesem Kapitel vertiefen wir die bisherigen Beobachtungen und Diskussionen. Einige Dinge werden dabei repetiert, andere systematischer dargestellt, und einiges Neues kommt natürlich auch hinzu.

Wir hatten bereits gesehen, dass die Entwicklung einer Substanz M (eines Medikament, eines Markers, Schadstoffs, ...) in einem System aus Kompartimenten (Organen) durch ein System von linearen DGLn beschrieben werden kann. Betrachten wir dazu ein einfaches Beispiel:

Beispiel 3.2.1. Wir verfolgen eine Substanz in zwei Organen (Kompartimente K_1 und K_2). Die Situation stellt sich schematisch wie folgt dar:



Mit dieser Darstellung meinen wir Folgendes: Zur Zeit t

- befindet sich die Menge $y_i(t)$ der Substanz im Kompartiment K_i ,
- wird dem Kompartiment K_1 die Menge $M(t)$ zugeführt,
- baut das Kompartiment K_i die aktuell gegebene Menge mit der Rate a_i ab,
- gibt das Kompartiment K_1 die dort vorhandene Menge mit der Rate b_{21} an das Kompartiment K_2 ab - und umgekehrt.
- Allgemeint heisst das: Kompartiment K_j gibt die dort vorhandene Menge mit der Rate b_{ij} an K_i ab. Als Faustregel haben wir damit “Zielindex zuerst”.

Wir stellen uns, unter anderem, folgende Fragen:

- Wie entwickeln sich im Laufe der Zeit t die Mengen y_1 und y_2 ?
- Gibt es zum Beispiel stationäre Zustände, bei denen die jeweilige Menge in den beiden Kompartimenten konstant ist?

Die schematische Darstellung übersetzt sich in ein mathematisches Modell

Mathematisches Modell: System von linearen DGL

$$\begin{aligned} y'_1(t) &= M(t) - a_1 y_1(t) - b_{21} y_1(t) + b_{12} y_2(t) \\ y'_2(t) &= b_{21} y_1(t) - b_{12} y_2(t) - a_2 y_2(t) \end{aligned}$$

Beachte:

- Es sind $y_i : I \rightarrow \mathbb{R}$ auf einem Intervall differenzierbare Funktionen. Dabei ist I jeweils so gewählt, dass es für die Anwendung und die Theorie günstig und sinnvoll ist.
- Pfeile aus dem Kompartiment K_i hinaus übersetzen sich in ein Minuszeichen.
- Pfeile in ein Kompartiment K_i hinein übersetzen sich in ein Pluszeichen. Mit dieser Konvention können nur Terme der Form $\pm a_{ij}y_i$ und $\pm b_{ij}y_j$ vorkommen.

Diese Gleichungssystem übersetzen wir in die Sprache der Linearen Algebra. Wir erhalten ein (2×2) -System linearer DGL 1. Ordnung

$$\begin{cases} y'_1(t) = M(t) - (a_1 + b_{21}) y_1(t) & +b_{12} y_2(t) \\ y'_2(t) = b_{21} y_1(t) & -(b_{12} + a_2) y_2(t) \end{cases}$$

Die Matrixschreibweise ist dann

$$\begin{pmatrix} y'_1(t) \\ y'_2(t) \end{pmatrix} = \begin{pmatrix} -(a_1 + b_{21}) & b_{12} \\ b_{21} & -(b_{12} + a_2) \end{pmatrix} \cdot \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} + \begin{pmatrix} M(t) \\ 0 \end{pmatrix}$$

und kürzer $y'(t) = A \cdot y(t) + g(t)$ mit $y'(t) = \begin{pmatrix} y'_1(t) \\ y'_2(t) \end{pmatrix}$, $y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix}$ und

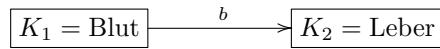
$$g(t) = \begin{pmatrix} M(t) \\ 0 \end{pmatrix}, \quad A = \begin{pmatrix} -(a_1 + b_{21}) & b_{12} \\ b_{21} & -(b_{12} + a_2) \end{pmatrix}.$$

Zur Vereinfachung nehmen wir nun **homogene** DGL im System $y'(t) = A \cdot y(t)$. Anschaulich haben wir keinen Zu- oder Abfluss in dem System.

3.2.1 Beispiele: Serielle Box-Kompartiment-Modelle

Ein 2-Box-Modell

Gegeben sei mit $0 < b$ die folgende einfache Situation:



Das heisst, die Substanz wandert mit Rate b aus dem Blut in die Leber. Damit erhalten wir zwei (lineare) DGLn:

$$y'_1 = -by_1 \text{ und } y'_2 = by_1.$$

Für die exponentielle Entwicklung $y'_1 = -by_1$ können wir direkt die allgemeine Lösung angeben:

$$y'_1 = -by_1 \implies y_1(t) = C_1 e^{-bt}.$$

Diese Lösung setzen wir dann in die 2. DGL $y'_2 = by_1$ oben ein:

$$\begin{aligned} y'_2(t) = by_1(t) &= b(C_1 e^{-bt}) \implies y_2(t) = \int bC_1 e^{-bt} dt \\ &= -C_1 e^{-bt} + C_2 \end{aligned}$$

Die Konstanten C_1 und C_2 können wir aus Anfangsbedingungen bestimmen.

$$y'_1 = -by_1, \quad y_1(0) = y_{10} \quad (3.1)$$

$$y'_2 = by_1, \quad y_2(0) = y_{20} \quad (3.2)$$

Wieder hat (3.1) die Lösung $y_1(t) = y_{10}e^{-bt}$ nun mit $C_1 = y_{10} = y_1(0)$. Einsetzen in (3.2) liefert mit dem Hauptsatz

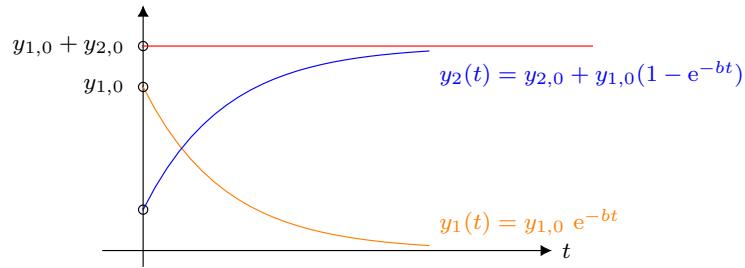
$$\begin{aligned} y_2(t) &= y_{20} + \int_0^t b y_1(\tau) d\tau = y_{20} + b y_{10} \int_0^t e^{-b\tau} d\tau \\ &= y_{20} + b y_{10} \left. \frac{e^{-b\tau}}{-b} \right|_0^t \\ &= y_{20} - y_{10} (e^{-bt} - 1). \end{aligned}$$

Bemerkungen:

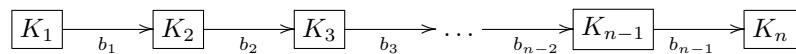
- Im Laufe der Zeit f für $t \rightarrow \infty$ ist mit dem negativen Exponenten $y_1(t) \rightarrow 0$ und damit

$$y_2(t) = y_{20} - y_{10} (e^{-bt} - 1) \rightarrow y_{20} - y_{10}(-1) = y_{10} + y_{20} = C_2$$

wie auch bei den Graphen der Lösungen sehen.



- Analoges gilt auch für n -Kompartiment-Systeme (siehe Übungen):



Alternative Lösung mit Linearer Algebra

Wir schreiben nun die Lösungen y_1 und y_2 in einen Vektor $y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix}$ und erhalten:

$$\begin{aligned} y(t) &= \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = \begin{pmatrix} C_1 e^{-bt} \\ -C_1 e^{-bt} + C_2 \end{pmatrix} \\ &= C_1 e^{-bt} \begin{pmatrix} 1 \\ -1 \end{pmatrix} + C_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{aligned}$$

Die beiden Vektoren $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ und $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ sind Eigenvektoren der Matrix $A = \begin{pmatrix} -b & 0 \\ b & 0 \end{pmatrix}$ zu den Eigenwerten $-b$ und 0 . Aber warum, und woher kommt diese Matrix A ?

Wir schreiben das Gleichungssystem (3.1) und (3.2) für die beiden DGLn $y'_1 = -by_1$ und $y'_2 = by_1$ als ein Matrix-Vektor-Produkt $y'(t) = Ay(t)$ mit

$$y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix}, \quad y'(t) = \begin{pmatrix} y'_1(t) \\ y'_2(t) \end{pmatrix} \quad \text{und} \quad A = \begin{pmatrix} -b & 0 \\ b & 0 \end{pmatrix}.$$

Rechnen wir das Matrix-Vektor-Produkt $y'(t) = Ay(t)$ auf der rechten Seite aus, folgt:

$$y'(t) = Ay(t) = \begin{pmatrix} -b & 0 \\ b & 0 \end{pmatrix} \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = \begin{pmatrix} (-b)y_1(t) + 0 \cdot y_2(t) \\ by_2(t) + 0 \cdot y_1(t) \end{pmatrix} = \begin{pmatrix} (3.1) \\ (3.2) \end{pmatrix}$$

Mit $\mathbf{y}_0 = \begin{pmatrix} y_{10} \\ y_{20} \end{pmatrix}$ lautet die Lösung, wie im 1. Kapitel angedeutet, $y(t) = e^{\mathbf{A}t} \mathbf{y}_0$, und wir werden später sehen, dass in unserm Fall $e^{\mathbf{A}t} = \begin{pmatrix} e^{-bt} & 0 \\ 1 - e^{-bt} & 1 \end{pmatrix}$.

Übung: Vergleiche diese Darstellung mit der Lösung oben:

1. Berechnen Sie mit diesem $e^{\mathbf{A}t} = \begin{pmatrix} e^{-bt} & 0 \\ 1 - e^{-bt} & 1 \end{pmatrix}$ das Matrix-Vektor-Produkt $e^{\mathbf{A}t} \mathbf{y}_0$.
2. Vergleichen Sie die Koordinatenfunktionen des Produktvektors mit den Lösungsfunktionen der DGLn (3.1) und (3.2).

Bemerkung:

Um die allgemeine Lösung $t \mapsto y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix}$ eines n -Kompartiments-Modells zu berechnen,

können wir versuchen, diese direkt zu bestimmen, oder die Methoden der Linearen Algebra für das DGL-System $y' = Ay$ mit einer $n \times n$ Matrix A anzuwenden. Dabei sind

$$y'(t) = \begin{pmatrix} y'_1(t) \\ y'_2(t) \\ \vdots \\ y'_n(t) \end{pmatrix}, \quad y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix}, \quad y_k : \mathbb{R} \rightarrow \mathbb{R}, t \mapsto y_k(t).$$

Hier wird uns vor allem die Theorie der Eigenwerte und -vektoren einer Matrix helfen

Folgende Aussage ermöglicht, sich in vielen Fällen die Lösung noch einfacher zu beschaffen:

Sei λ ein EW zum EV v von A , das heisst, es gilt $\mathbf{A}v = \lambda v$. Dann ist $y(t) = e^{\lambda t} v$ eine Lösung des DGL-Systems $y' = Ay$.

Das rechnen wir nach:

$$\begin{aligned} y'(t) &= \underbrace{(\mathbf{e}^{\lambda t} \mathbf{v})'}_{\text{Koordinate ableiten}} = \lambda e^{\lambda t} \mathbf{v} = \underbrace{e^{\lambda t} \lambda \mathbf{v}}_{\text{EV und EW}} = \underbrace{e^{\lambda t} A \mathbf{v}}_{\text{EV und EW}} = A e^{\lambda t} \mathbf{v} \\ &= A y(t). \end{aligned}$$

Anwendung auf das System (3.1)-(3.2): Die EW der Dreiecksmatrix $A = \begin{pmatrix} -b & 0 \\ b & 0 \end{pmatrix}$ stehen auf der Diagonalen: $\lambda_1 = -b$, $\lambda_2 = 0$ und mit zugehörigen EV

$$v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Somit haben wir zwei **Basislösungen** $e^{\lambda_i t} v_i$ und durch Linearkombination finden wir die allgemeine Lösung

$$y(t) = C_1 e^{-bt} \begin{pmatrix} 1 \\ -1 \end{pmatrix} + C_2 \underbrace{e^{0t}}_{=1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Übung:

- Bestimmen Sie die Koeffizienten C_i mit der Anfangsbedingung

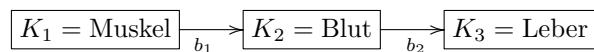
$$y(0) = C_1 \begin{pmatrix} 1 \\ -1 \end{pmatrix} + C_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} y_{10} \\ y_{20} \end{pmatrix}$$

als Lösungen eines 2×2 -LGS.

- Vergleichen Sie Ihr Ergebnis mit $e^{At} y_0$.

3.2.2 Ein 3-Box-Modell

Gegeben sei nun folgende Situation:



mit $0 < b_1 < b_2$. Diesmal geben wir konkrete Anfangswerte vor.

Dann erhalten wir ein System von drei DGLn als AWP (Anfangswertproblem):

$$\begin{array}{ll} \text{Für } K_1 : y'_1 = -b_1 y_1 & y_1(0) = y_0 \\ \text{Für } K_2 : y'_2 = -b_1 y_1 - b_2 y_2 & y_2(0) = 0 \\ \text{Für } K_3 : y'_3 = -b_2 y_2 & y_3(0) = 0 \end{array}$$

Auch hier sind die Lösungen elementar zu finden:

Für K_1 : Wie oben ist $y_1 = y_0 e^{-b_1 t}$, wobei wir die Konstante grad schon passend zum Anfangswert gewählt haben.

Für K_2 : Setze die Lösung $y_1 = y_0 e^{-b_1 t}$ in die 2. DGL ein:

$$\begin{aligned} y'_2 &= b_1 y_1 - b_2 y_2 \\ &= b_1 (y_0 e^{-b_1 t}) - b_2 y_2 \quad \xrightarrow{\text{Var. d. Konst}} \quad y_2(t) = \frac{y_0 b_1}{b_2 - b_1} (e^{-b_1 t} - e^{-b_2 t}). \end{aligned}$$

Dabei wurde Konstante in der allgemeinen Lösung so gewählt, dass die Anfangsbedingung $y_2(0) = 0$ erfüllt ist.

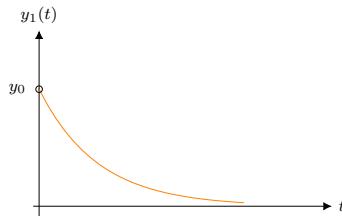
Vergleiche y_2 mit der „Bateman-Funktion“ in den Übungen. Dort diskutieren wir auch die Annahme $b_2 > b_1$.

Für K_3 : Setze die gefundene Lösung für K_2 ein:

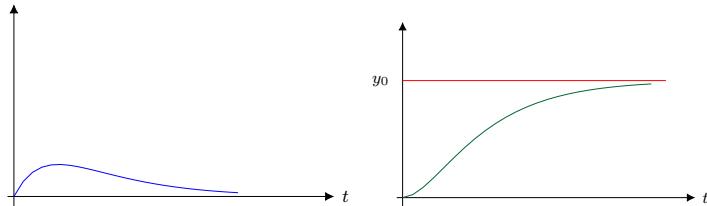
$$\begin{aligned} y'_3 &= b_2 y_2, \quad y_3(0) = 0 \implies y_3(t) = \int_0^t b_2 y_2(\tau) d\tau \\ &= \frac{y_0}{b_2 - b_1} (b_1 (e^{-b_2 t} - 1) - b_2 (e^{-b_1 t} - 1)). \end{aligned}$$

Übung

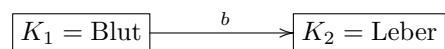
Die Lösungskurve für y_1 ist



Welcher Graph beschreibt jeweils die Lösungen y_2 und y_3 ?



Vergleichen Sie die Kurven auch mit dem Verlauf für das 2-Box-Modell



Mit Lineare Algebra

Wir schreiben das Systems oben

$$\begin{aligned} y'_1 &= -b_1 y_1 \\ y'_2 &= b_1 y_1 - b_2 y_2 \\ y'_3 &= b_2 y_2 \end{aligned}$$

in der Form $y' = Ay$ mit $A = \begin{pmatrix} -b_1 & 0 & 0 \\ b_1 & -b_2 & 0 \\ 0 & b_2 & 0 \end{pmatrix}$. Die EW lesen wir wieder direkt ab $\lambda_1 = -b_1$, $\lambda_2 = -b_2$ und $\lambda_3 = 0$. Die EW sind paarweise voneinander verschieden und damit zugehörige EV linear unabhängig:

$$v_1 = \begin{pmatrix} -b_1 + b_2 \\ b_1 \\ -b_2 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Die allgemeine Lösung des Systems lautet also

$$y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{pmatrix} = C_1 e^{-b_1 t} v_1 + C_2 e^{-b_2 t} v_2 + C_3 v_3 \quad (\text{wieder: } e^0 = 1).$$

Die Koeffizienten C_i bestimmt man aus der Anfangsbedingung

$$y(0) = C_1 \begin{pmatrix} -b_1 + b_2 \\ b_1 \\ -b_2 \end{pmatrix} + C_2 \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} + C_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} y_0 \\ 0 \\ 0 \end{pmatrix}.$$

Dieses lineare Gleichungssystem hat Dreiecksform und kann daher direkt gelöst werden:

$$C_1 = \frac{y_0}{-b_1 + b_2}, \quad C_2 = \frac{b_1 y_0}{-b_1 + b_2}, \quad C_3 = y_0.$$

Also

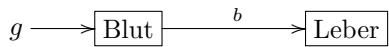
$$y(t) = \frac{y_0 e^{-b_1 t}}{-b_1 + b_2} \begin{pmatrix} -b_1 + b_2 \\ b_1 \\ -b_2 \end{pmatrix} + \frac{b_1 y_0 e^{-b_2 t}}{-b_1 + b_2} \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} + y_0 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Übung. Vergleichen Sie die unterschiedlichen Darstellungen der Lösungen des Modells

$$K_1 = \text{Muskel} \xrightarrow{b_1} K_2 = \text{Blut} \xrightarrow{b_2} K_3 = \text{Leber}$$

3.2.3 Weitere Beispiele für (lineare) Kompartiment-Modelle

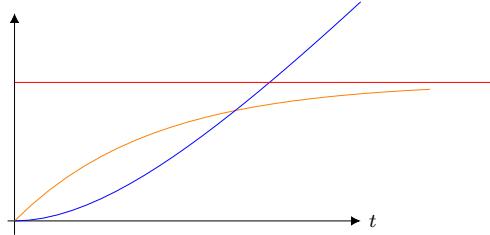
1. Entwicklung eines Medikaments bei konstanter Infusion g :



Zur Zeit t befindet sich die Medikamentenmenge

- $y_1(t)$ im Blut,
- $y_2(t)$ in der Leber,
- mit den Startwerten $y_1(0) = 0$ und $y_2(0) = 0$.

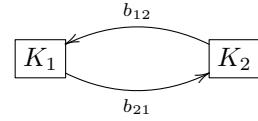
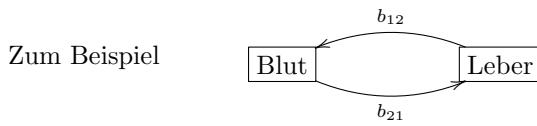
Der Verlauf der Lösungskurven ist



Siehe auch Übungsaufgaben.

2. Beispiele für Modelle mit Rückfluss

- (a) Ein 2-Box-Modell mit Rückfluss sieht so aus:



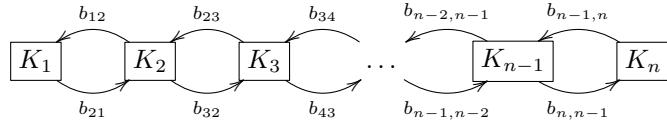
Die Entwicklung wird beschrieben durch das System

$$y'_1 = -b_{21}y_1 + b_{12}y_2, \quad y'_2 = b_{21}y_1 - b_{12}y_2,$$

oder in Matrixschreibweise $y'(t) = Ay(t)$ mit

$$A = \begin{pmatrix} -b_{21} & b_{12} \\ b_{21} & -b_{12} \end{pmatrix}.$$

- (b) Ein n -Box-Modell mit Rückfluss



wird durch folgendes System von DGLn beschrieben:

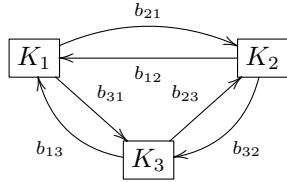
$$y'_1 = -b_{21}y_1 + b_{12}y_2, \quad y'_2 = b_{21}y_1 - (b_{32} + b_{12})y_2 + b_{23}y_3, \quad \dots,$$

$$y'_n = b_{n,n-1}y_{n-1} - b_{n-1,n}y_n$$

Die dazu gehörige Matrix lautet demnach

$$A = \begin{pmatrix} -b_{21} & b_{12} & 0 & \dots & 0 \\ b_{21} & -(b_{32} + b_{12}) & b_{23} & 0 & \vdots \\ 0 & b_{32} & -(b_{43} + b_{23}) & b_{34} & 0 \\ \vdots & \dots & \ddots & \ddots & b_{n-1,n} \\ 0 & \dots & \dots & 0 & b_{n,n-1} \\ & & & & -b_{n-1,n} \end{pmatrix}$$

(c) Ein 3-Box-Modell mit zyklischem Rückfluss hat folgende Form:



Mit dem zugehörigen System von DGLn:

$$\begin{aligned} y'_1 &= -(b_{31} + b_{21})y_1 + b_{12}y_2 + b_{13}y_3 \\ y'_2 &= b_{21}y_1 - (b_{12} + b_{32})y_2 + b_{23}y_3 \\ y'_3 &= b_{31}y_1 + b_{32}y_2 - (b_{13} + b_{23})y_3 \end{aligned}$$

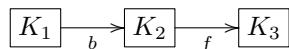
und als Matrix des Systems $y'(t) = Ay(t)$ sehen wir:

$$A = \begin{pmatrix} -(b_{31} + b_{21}) & b_{12} & b_{13} \\ b_{21} & -(b_{12} + b_{32}) & b_{23} \\ b_{31} & b_{32} & -(b_{13} + b_{23}) \end{pmatrix}$$

Übung:

Berechnen Sie jeweils die Determinante und entscheiden Sie, ob das System $y'(t) = Ay(t)$ eine stationäre Lösung y_∞ hat. Was bedeutet das jeweils für das Modell?

3. Ein **nicht** lineares Beispiel:



Dabei ist b eine Konstante und f eine Funktion von y_2 . Die Entwicklung wird nun beschrieben durch ein System von DGL

$$y'_1 = -by_1, \quad y'_2 = by_1 - \frac{c \cdot y_2}{1 + \frac{y_2}{a}}, \quad y'_3 = \frac{c \cdot y_2}{1 + \frac{y_2}{a}} \text{ mit Konstanten } a, c.$$

Der Ausdruck $\frac{cy_2}{1 + \frac{y_2}{a}}$ ist die Michaelis-Menten-Wachstumsfunktion. Bei der Modellierung biologischer Prozesse spielt diese eine prominente Rolle. In der Tat werden die meisten Prozesse in der Natur nicht linear sein. Wir können aber versuchen, solche Modelle zu linearisieren.

3.2.4 Zusammenfassung für homogen und linear unabhängigen Fall

Definition 3.2.2. Sei $A \in M_{n \times n}$.

1. Dann ist $y' = Ay$ oder $y'(t) = Ay(t)$ ein homogenes lineares $n \times n$ -DGL-System mit

$$y'(t) = \begin{pmatrix} y'_1(t) \\ y'_2(t) \\ \vdots \\ y'_n(t) \end{pmatrix}, \quad y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix}, \quad y_k : \mathbb{R} \rightarrow \mathbb{R}, t \mapsto y_k(t).$$

2. Eine Lösung ist eine Funktion $y : \mathbb{R} \rightarrow \mathbb{R}^n$, $t \mapsto y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix}$, sodass für alle t gilt
 $y'(t) = Ay(t)$.

Theorem 3.2.3. Sei $y' = Ay$ wie oben mit λ ein EW und EV $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ von A .

Dann gilt:

1. Die Funktion $y : \mathbb{R} \rightarrow \mathbb{R}^n$, $t \mapsto y(t) = e^{\lambda t} v = \begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix} = \begin{pmatrix} x_1 e^{\lambda t} \\ x_2 e^{\lambda t} \\ \vdots \\ x_n e^{\lambda t} \end{pmatrix}$ ist eine Lösung des DGL-Systems $y' = Ay$.

2. **FALLS** v_1, v_2, \dots, v_n lineare unabhängige EV sind, ist die allgemeine Lösung mit n Konstanten C_1, \dots, C_n die Funktion

$$t \mapsto y(t) = C_1 e^{\lambda_1 t} v_1 + C_2 e^{\lambda_2 t} v_2 + \dots + C_n e^{\lambda_n t} v_n.$$

CAVE

Wie bei den LGS kann ein EW auch mehrfach auftauchen. Wichtig ist, dass die n EV linear unabhängig sind. Wir werden nun die Theorie so weiter entwickeln, dass wir auch den allgemeinen Fall lösen können.

3.3 Vektorräume

Um die Lineare Algebra auf allgemeine Lineare DGL-Systeme anwenden zu können, führen wir abstrakte Vektorräume ein, bzw. wir erinnern daran. Wir kennen schon den Vektorraum \mathbb{R}^n (und analog \mathbb{C}^n). Diese reichen aber nicht. Wir interessieren uns für die Situation, dass eine 2×2 -Matrix ein DGL-System $y' = Ay$ definiert.

Betrachten wir nun die Menge von Funktionen

$$C^1(\mathbb{R}, \mathbb{R}^2) = \left\{ f : \mathbb{R} \rightarrow \mathbb{R}^2 \mid f(t) = \begin{pmatrix} f_1(t) \\ f_2(t) \end{pmatrix}, f_1, f_2 \in C^1(\mathbb{R}) \right\}$$

so ist der Lösungsraum des Systems die (Teil-)Menge

$$\mathcal{L}_A = \left\{ f : \mathbb{R} \rightarrow \mathbb{R}^2 \mid f \in C^1(\mathbb{R}, \mathbb{R}^2) \text{ und } f'(t) = Af(t) \text{ für alle } t \right\}.$$

Um deren Struktur zu verstehen, führen wir zwei Operationen ein (Addition und Skalare Multiplikation), die $\mathcal{L}_A \subset C^1(\mathbb{R}, \mathbb{R}^2)$ zu einem Untervektorraum machen.

Erinnerung an die Vektorräume $\mathbb{R}^n, \mathbb{C}^n$

Wir kennen schon den Vektorraum \mathbb{R}^n mit den folgenden Operationen:

- die Addition zweier Vektoren ($v, w \in \mathbb{R}^n \implies v + w \in \mathbb{R}^n$)
- die Skalare Multiplikation einer Zahl mit einem Vektor ($\lambda \in \mathbb{R}, v \in \mathbb{R}^n \implies \lambda v \in \mathbb{R}^n$)

Analog können wir so den Vektorraum \mathbb{C}^n einführen.

Es gelten dann die üblichen Rechenregeln: Seien $\lambda, \mu \in \mathbb{R}$ reelle Zahlen und $u, v, w \in \mathbb{R}^n$ Vektoren. Dann gelten:

$$\begin{array}{lll} v + w &= w + v & \text{Kommutativgesetz} \\ u + (v + w) &= (u + v) + w & \text{Assoziativgesetz} \\ \lambda(v + w) &= \lambda v + \lambda w & \text{Distributivgesetz} \\ (\lambda + \mu)v &= \lambda v + \mu v & \text{Distributivgesetz} \\ (\lambda\mu)v &= \lambda(\mu v) & \text{Assoziativgesetz} \end{array}$$

Schreiben wir $\mathbb{R}^n = V$, so gilt zudem (und wieder analog für \mathbb{C}^n):

Abgeschlossenheit.

- Für $v, w \in V$ ist auch $v + w \in V$,
- für $\lambda \in \mathbb{R}$ und $v \in V$ ist $\lambda v \in V$, für $1 \in \mathbb{R}$ ist $1v = v$.

Nullvektor. In V gibt es ein Element 0 mit $0 + v = v$, für alle $v \in V$.

Inverses bezüglich Addition. Zu jedem $v \in V$ gibt es ein Element $-v \in V$ mit $v + (-v) = 0$.

Definition 3.3.1. Sobald nun auf einer Menge V zwei Operationen

$$\begin{aligned} V \times V &\rightarrow V, \quad (v, w) \mapsto v + w \\ \mathbb{R} \times V &\rightarrow V, \quad (\lambda, v) \mapsto \lambda v \end{aligned}$$

definiert sind, für die alle oben genannten Rechenregeln gelten, heisst V ein reeller Vektorraum.

Bemerkungen:

- Mit anderen Worten: Ein \mathbb{R} - (oder \mathbb{C} -) Vektorraum (VR) ist eine Menge von Objekten (zum Beispiel n -Tupel oder Funktionen) zusammen mit 2 Verknüpfungen
 - + : $v, w \in V \implies v + w \in V$ Vektoraddition
 - : $\lambda \in \mathbb{R}, v \in V \implies \lambda \cdot v \in V$ Skalare Multiplikation,
 welche die Regeln oben erfüllen.
- Ein komplexer Vektorraum ist ganz analog mit \mathbb{C} anstelle von \mathbb{R} definiert.
- Aus den Rechenregeln folgt unter anderem, dass es in einem Vektorraum nur einen Nullvektor gibt, und dass $-v = (-1) \cdot v$ eindeutig bestimmt ist. Ebenso kann man beweisen, dass ein beliebiges Vielfaches des Nullvektors $\lambda \cdot 0$ wieder der Nullvektor ist, und dass das Produkt eines beliebigen Vektors mit dem Skalar 0 den Nullvektor liefert.

Beispiele von Funktionenräumen

Uns interessieren Vektorräume, deren Elemente Funktionen sind. Diese Elemente heissen Vektoren.

1. Sei $\mathcal{F}(\mathbb{R}, \mathbb{R})$ die Menge aller Funktionen $\mathbb{R} \rightarrow \mathbb{R}$.

Für $f, g \in \mathcal{F}(\mathbb{R}, \mathbb{R})$ führen wir ein neues + ein

$$\underbrace{f + g}_{\text{neu}} : \mathbb{R} \rightarrow \mathbb{R} \quad x \mapsto (f + g)(x) = f(x) \underbrace{+}_{\text{in } \mathbb{R}} g(x)$$

und mit $\lambda \in \mathbb{R}$

$$\underbrace{\lambda \cdot f}_{\text{neu}} : \mathbb{R} \rightarrow \mathbb{R} \quad x \mapsto (\lambda \cdot f)(x) = \lambda \underbrace{\cdot}_{\text{in } \mathbb{R}} f(x).$$

Dann erfüllt $\mathcal{F}(\mathbb{R}, \mathbb{R})$ die Regeln oben, wobei der Nullvektor $0 \in \mathcal{F}(\mathbb{R}, \mathbb{R})$ die konstante Nullfunktion ist.

2. In $\mathcal{F}(\mathbb{R}, \mathbb{R})$ betrachten wir nur die Funktionen, für welche die n -te Ableitung $f^{(n)}$ existiert und stetig ist. Diese fassen wir zusammen mit

$$C^n(\mathbb{R}) = \{f : \mathbb{R} \rightarrow \mathbb{R} \mid f^{(n)} \text{ existiert und ist stetig}\}.$$

Ausserdem ist $C^\infty(\mathbb{R}) = \{f : \mathbb{R} \rightarrow \mathbb{R} \mid f \text{ ist beliebig oft differenzierbar}\}$. Zusammen mit den unter 1. definierten Operationen sind auch $C^n(\mathbb{R})$ und $C^\infty(\mathbb{R})$ Vektorräume.

3. In $\mathcal{F}(\mathbb{R}, \mathbb{R})$ ist $\mathcal{P}_{\leq n}$ die Menge aller Polynom-Funktionen

$$p : \mathbb{R} \rightarrow \mathbb{R} \quad x \mapsto \sum_{k=0}^d a_k x^k, \quad d \leq n$$

vom Grad $\leq n$ ein Vektorraum.

CAVE

Wir brauchen $\leq n$. Die Polynome vom Grad $= n$ bilden keinen Vektorraum.

Seien zum Beispiel $p(x) = 1 + x^2$ und $q(x) = 4 - x^2$ vom Grad 2, dann ist $(p + q)(x) = 5$ vom Grad 0.

Untervektorräume

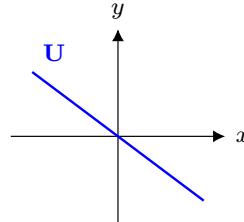
Definition 3.3.2. Sei V ein Vektorraum. Eine nicht leere Teilmenge $U \subset V$ heisst Unterraum oder Untervektorraum von V (U-VR), wenn U bezüglich + und \cdot abgeschlossen ist, das heisst,

$$+ : v, w \in U \implies v + w \in U$$

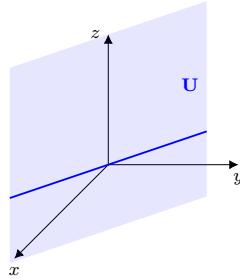
$$\cdot : \lambda \in \mathbb{R}, v \in U \implies \lambda \cdot v \in U.$$

Beispiele 3.3.3.

1. Sei $U \subset \mathbb{R}^2$ die Menge $U = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \mid 3x + 4y = 0 \right\}$. Dies ist eine Gerade in der Ebene durch den Nullpunkt.



2. Sei $U \subset \mathbb{R}^3$ die Menge $U = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid 3x + 4y = 0 \right\}$. Dies ist eine Ebene im Raum durch den Nullpunkt.



3. Die Menge der stetigen reellen Funktionen auf einem Intervall, $C([a, b], \mathbb{R})$, ist ein Unterraum des Vektorraums aller reeller Funktionen auf dem Intervall, $\mathcal{F}([a, b], \mathbb{R})$.
4. Die Menge $\mathcal{P}_{<\infty}$ aller Polynome ist ein Vektorraum.
5. Das Folgende ist eine Kette ineinander geschachtelter Unterräume:

$$\begin{aligned} \mathcal{F}(\mathbb{R}, \mathbb{R}) &\supset C^0(\mathbb{R}, \mathbb{R}) \supset C^1(\mathbb{R}, \mathbb{R}) \supset C^2(\mathbb{R}, \mathbb{R}) \supset \dots \\ &\dots \supset C^\infty(\mathbb{R}, \mathbb{R}) \supset \mathcal{P}_{<\infty} \supset \dots \supset \mathcal{P}_{\leq 2} \supset \mathcal{P}_{\leq 1} \supset \mathcal{P}_{\leq 0} \supset \{0\} \end{aligned}$$

6. Sei $A \in M_{m \times n}$. Dann ist der **Kern** von A $\ker A = \{x \in \mathbb{R}^n \mid Ax = 0\}$ ein UR von \mathbb{R}^n . Tatsächlich gilt für $u, v \in \ker A$ und $\lambda \in \mathbb{R}$:
- $A(u + v) = Au + Av = 0 + 0 = 0$
 - $A(\lambda u) = \lambda Au = \lambda 0 = 0$

Beispiel Lösungsraum eines DGL-Systems

Sei $C^1(\mathbb{R}, \mathbb{R}^2) = \left\{ f : \mathbb{R} \rightarrow \mathbb{R}^2 \mid f(t) = \begin{pmatrix} f_1(t) \\ f_2(t) \end{pmatrix}, f_1, f_2 \in C^1(\mathbb{R}) \right\}$.

Sei A eine 2×2 -Matrix, welche ein DGL-System $y' = Ay$ definiert. Der Lösungsraum

$$\mathcal{L}_A = \{f : \mathbb{R} \rightarrow \mathbb{R}^2 \mid f \in C^1(\mathbb{R}, \mathbb{R}^2) \text{ und } f'(t) = Af(t) \text{ für alle } t\}$$

ist ein Untervektorraum $\mathcal{L}_A \subset C^1(\mathbb{R}, \mathbb{R}^2)$, denn

1. Die Nullfunktion 0 erfüllt das System $y' = Ay \implies 0 \in \mathcal{L}_A$.
2. Seien $f, g \in \mathcal{L}_A$. Dann gilt mit Ableitungsregeln und Linearität

$$\begin{aligned} f' = Af, \quad g' = Ag \implies f' + g' &= (f + g)' = Af + Ag = A(f + g) \\ &\implies f + g \in \mathcal{L}_A. \end{aligned}$$

3. **Übung:** Zeigen Sie die Abgeschlossenheit für die Skalare Multiplikation.

Erinnerung: Allfällige **stationäre** Lösungen $f \in \mathcal{L}_A$ sind Lösungen mit $f' = Af = 0$. Dies sind die Eigenvektoren zum Eigenwert $\lambda = 0$.

Diese Überlegungen gelten auch für eine beliebige $n \times n$ -Matrix, und wir fragen:

Wie können wir den Lösungsraum \mathcal{L}_A des DGL-Systems $y' = Ay$ beschreiben?

Betrachten wir zunächst ein Beispiel:

Beispiel 3.3.4. Sei $A = \begin{pmatrix} \frac{1}{5} & \frac{2}{5} \\ -\frac{3}{5} & \frac{13}{10} \end{pmatrix}$ mit EW $\lambda_1 = 1, \lambda_2 = \frac{1}{2}$ und linear unabhängigen EV $v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, v_2 = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$. Dann gilt

1. Wie oben in 3.2.1 festgestellt, hat das System $y' = Ay$ sicher die Lösungen

$$t \mapsto e^t \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ und } t \mapsto e^{\frac{1}{2}t} \begin{pmatrix} 4 \\ 3 \end{pmatrix}.$$

Auch jede Linearkombination mit

$$y(t) = C_1 e^t \begin{pmatrix} 1 \\ 2 \end{pmatrix} + C_2 e^{\frac{1}{2}t} \begin{pmatrix} 4 \\ 3 \end{pmatrix} = C_1 \begin{pmatrix} e^t \\ 2e^t \end{pmatrix} + C_2 \begin{pmatrix} 4e^{\frac{1}{2}t} \\ 3e^{\frac{1}{2}t} \end{pmatrix}$$

ist eine Lösung.

2. Wir werden uns später davon überzeugen, dass tatsächlich alle Lösungen in \mathcal{L}_A von dieser Form sind, und damit die allgemeine Lösung $t \mapsto y(t) = C_1 e^t \begin{pmatrix} 1 \\ 2 \end{pmatrix} + C_2 e^{\frac{1}{2}t} \begin{pmatrix} 4 \\ 3 \end{pmatrix}$ ist.

CAVE

Wichtig ist hier, dass die EV linear unabhängig sind.

3.4 Erzeugendensysteme und Basen

3.4.1 Linearkombination in einem Vektorraum V

Erinnerung: Seien $w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{pmatrix} \in \mathbb{R}^m$ und $v_1, v_2, \dots, v_n \in \mathbb{R}^m$. Der Vektor w ist eine Linearkombination der

Vektoren v_1, v_2, \dots, v_n , falls es Zahlen $\alpha_1, \alpha_2, \dots, \alpha_n$ gibt mit $w = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$. Die Summe von Linearkombinationen von Vektoren und ebenso ein Vielfaches einer solchen Linearkombination ist wieder eine Linearkombination derselben Vektoren.

Seien nun V ein VR und $v_1, \dots, v_m \in V$. Der Vektor $w = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_m v_m = \sum_{i=1}^m \alpha_i v_i$ ist eine Linearkombination von v_1, v_2, \dots, v_m .

Definition 3.4.1. Seien $v_1, \dots, v_m \in V$. Der von v_1, \dots, v_m aufgespannte U -Vektorraum $U \subset V$ ist

$$\begin{aligned} U &= \{w \in V \mid w \text{ ist eine Linearkombination von } v_1, \dots, v_m\} \\ &= \left\{ w \in V \mid \text{Es gibt } \alpha_1, \dots, \alpha_m \in \mathbb{R} \text{ mit } w = \sum_{i=1}^m \alpha_i v_i \right\}. \end{aligned}$$

Notation: $U = \langle \{v_1, \dots, v_m\} \rangle$.

Beispiele

- Die Spaltenvektoren e_i der Einheitsmatrix in $M_{n \times n}$ spannen \mathbb{R}^n auf.
- Die Polynome $1, x, x^2, \dots, x^n$ spannen $\mathcal{P}_{\leq n}$ auf.
- Das **Bild** von $A \in M_{m \times n}$

$$\text{Im } A = \{Ax \mid x \in \mathbb{R}^n\}$$

wird aufgespannt von den Spaltenvektoren von A und ein Unterraum des \mathbb{R}^m .

- $\langle \{1, x, x^2, x^3, \dots, x^n\} \rangle = \mathcal{P}_{\leq n}$

- Blut → Leber

$$A = \begin{pmatrix} -b & 0 \\ b & 0 \end{pmatrix}$$

mit EW $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Mit den **Basislösungen** $t \mapsto e^{\lambda_i t} v_i$ hatten wir die allgemeine Lösung

$$y(t) = C_1 e^{-bt} \begin{pmatrix} -1 \\ 1 \end{pmatrix} + C_2 \underbrace{e^{0t}}_{=1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Jede Lösung des Systems ist Linearkombination dieser Basislösungen und damit

$$\mathcal{L}_A = \left\langle \left\{ t \mapsto e^{-bt} \begin{pmatrix} -1 \\ 1 \end{pmatrix}, t \mapsto \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} \right\rangle$$

- Sei $A = \begin{pmatrix} \frac{1}{5} & \frac{2}{5} \\ -\frac{3}{5} & \frac{13}{10} \end{pmatrix}$. Der Lösungsraum $\mathcal{L}_A \subset C^1(\mathbb{R}, \mathbb{R}^2)$ des Systems $y' = Ay$ wird aufgespannt von $t \mapsto \begin{pmatrix} e^t \\ 2e^t \end{pmatrix}$ und $t \mapsto \begin{pmatrix} 4e^{\frac{1}{2}t} \\ 3e^{\frac{1}{2}t} \end{pmatrix}$.

Umgekehrt fragen wir:

Sei V ein VR.

Finden wir ein „Erzeugendensystem“ $\{v_1, \dots, v_m\}$ von V ? Das heisst: Gibt es $\{v_1, \dots, v_m\}$ mit $\langle \{v_1, \dots, v_m\} \rangle = V$?

Mit dem zentralen Beispiel:

Sei $A \in M_{n \times n}$ mit $y' = Ay$. Welche Funktionen in $C^1(\mathbb{R}, \mathbb{R}^n)$ erzeugen $\mathcal{L}_A \subset C^1(\mathbb{R}, \mathbb{R}^2)$?

3.4.2 Basen

Definition 3.4.2. Sei V ein VR. Ein unverkürzbares (minimales) Erzeugendensystem \mathcal{B} von V heisst Basis von V . Das heisst, dass keine echte Teilmenge \mathcal{B} erzeugt den VR V .

Theorem 3.4.3. Seien V ein VR und $\mathcal{B} = \{b_1, \dots, b_n\}$ eine Teilmenge von Vektoren.

1. \mathcal{B} ist genau dann eine Basis von V , wenn b_1, \dots, b_n linear unabhängig sind und den VR V erzeugen.
2. \mathcal{B} ist genau dann eine Basis von V , wenn es für jeden Vektor $v \in V$ eindeutig bestimmte Koeffizienten $\alpha_i \in \mathbb{R}$ gibt mit $v = \sum_{i=1}^n \alpha_i b_i$.

Beispiele

- Die Spaltenvektoren e_i der Einheitsmatrix in $M_{n \times n}$ sind eine Basis von \mathbb{R}^n .
- Die Polynome $1, x, x^2, \dots, x^n$ bilden eine Basis von $\mathcal{P}_{\leq n}$.
- Sei $A = \begin{pmatrix} \frac{1}{5} & \frac{2}{5} \\ -\frac{3}{5} & \frac{13}{10} \end{pmatrix}$. Die Funktionen $t \mapsto \begin{pmatrix} e^t \\ 2e^t \end{pmatrix}$ und $t \mapsto \begin{pmatrix} 4e^{\frac{1}{2}t} \\ 3e^{\frac{1}{2}t} \end{pmatrix}$ bilden eine Basis des Lösungsraums $\mathcal{L}_A \subset C^1(\mathbb{R}, \mathbb{R}^2)$ des Systems $y' = Ay$.
- Je n linear unabhängige Vektoren im \mathbb{R}^n bilden eine Basis.

Anwendungsbeispiel 2 × 2-DGL-System Sei $A \in M_{2 \times 2}$ mit EW $\lambda_1 \neq \lambda_2$ mit EV v_1, v_2 . Dann sind die EV v_1, v_2 linear unabhängig, und wir wissen:

- Jede LK der Lösungen $t \mapsto e^{\lambda_i t} v_i$ eine Lösung von $y' = Ay$.
- Zu jeder Anfangsbedingung $y(0) = y_0 \in \mathbb{R}^2$ gibt es reelle Koeffizienten C_1 und C_2 mit $y(0) = C_1 v_1 + C_2 v_2 = y_0$.
- Also ist $\mathcal{L}_A = \langle \{t \mapsto e^{\lambda_1 t} v_1, t \mapsto e^{\lambda_2 t} v_2\} \rangle = \langle \{e^{\lambda_1 t} v_1, e^{\lambda_2 t} v_2\} \rangle$.
- Die Funktionen $t \mapsto e^{\lambda_i t} v_i$ sind linear unabhängig und damit ist $\{e^{\lambda_1 t} v_1, e^{\lambda_2 t} v_2\}$ sogar eine Basis von \mathcal{L}_A , denn: Aus $\alpha_1 e^{\lambda_1 t} v_1 + \alpha_2 e^{\lambda_2 t} v_2 = 0$ folgt nämlich für $t = 0$: $\alpha_1 v_1 + \alpha_2 v_2 = 0$, also tatsächlich $\alpha_1 = \alpha_2 = 0$.
- Zusammengefasst: Die allgemeine Lösung ist $t \mapsto y(t) = C_1 e^{\lambda_1 t} v_1 + C_2 e^{\lambda_2 t} v_2$.

Eigenbasen

Definition 3.4.4. Eine Basis von \mathbb{R}^n aus EV einer Matrix $A \in M_{n \times n}$ heisst Eigenbasis.

Verallgemeinerung des Beispiels Hat $A \in M_{n \times n}$ die EW $\lambda_1, \dots, \lambda_n$ (nicht unbedingt paarweise voneinander verschieden) und eine Eigenbasis v_1, \dots, v_n .

Dann ist $\{t \mapsto e^{\lambda_1 t} v_1, \dots, t \mapsto e^{\lambda_n t} v_n\}$ eine Basis von $\mathcal{L}_A \subset C^1(\mathbb{R}, \mathbb{R}^n)$.

Bleibt das Problem: Nicht alle Matrizen $A \in M_{n \times n}$ besitzen eine Eigenbasis!

Wir werden sehen, dass dann die Spalten von $e^{At} = \sum_{k=0}^{\infty} \frac{t^k A^k}{k!}$ eine Basis des Lösungsraums \mathcal{L}_A bilden, und wir werden lernen, wie sich e^{At} ohne Reihe (oder Computer) einfacher und strukturell berechnen lässt.

Zunächst führen wir noch weitere Begriffe und Konzepte aus der Linearen Algebra ein.

Dimension

1. Jeder VR besitzt mindestens eine Basis. Diese kann dann auch jeweils unendlich viele Vektoren enthalten.
2. Sei V ein von endlich vielen Vektoren aufgespannter VR. Dann haben alle Basen die gleiche Anzahl n Vektoren, und n heißt **Dimension von V** . Notation: $n = \dim V = \dim_{\mathbb{R}} V$.

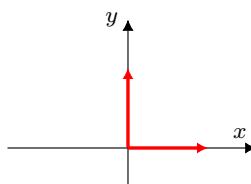
Beispiele

- (a) $\dim_{\mathbb{R}} \mathbb{R}^n = n$, $\dim_{\mathbb{C}} \mathbb{C}^n = n$ und $\dim_{\mathbb{R}} \mathbb{C}^n = 2n$
 - (b) $\dim \mathcal{P}_{\leq n} = n + 1$
 - (c) $\dim C^n(\mathbb{R}) = \infty$. Das heißt, dass eine endliche Anzahl an Vektoren (Funktionen) nicht ausreicht, dem Raum $C^n(\mathbb{R})$ zu erzeugen.
3. Je n linear unabhängige Vektoren in einem VR V der Dimension n bilden eine Basis.

Wahl einer Basis und Koordinaten

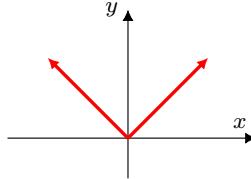
Sei $\dim V = n$. Die Wahl einer Basis $B = \{b_1, \dots, b_n\}$ von V bedeutet, dass wir die Vektorräume V und \mathbb{R}^n identifizieren können: $\varphi_B : V \rightarrow \mathbb{R}^n$, $v = \sum_{i=1}^n \alpha_i b_i \mapsto \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$. Dabei sind die $\alpha_i \in \mathbb{R}$ eindeutig, hängen aber von der Wahl einer Basis ab. Es sind die **Koordinaten** des Vektors v bezüglich dieser gewählten Basis B .

Beispiel 3.4.5. Seien $n = 2$ und $v = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$. Wählen wir zunächst die Standardbasis $B = \{e_1, e_2\}$



Dann ist $v = \begin{pmatrix} 4 \\ 2 \end{pmatrix} = 4 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = 4e_1 + 2e_2$. Bezuglich dieser Basis hat der Vektor v die Koordinaten 4 und 2, die wir im Koordinatenvektor $\varphi_B(v) = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$ zusammenfassen.

Sei nun als Basis $B = \{b_1, b_2\}$ mit $b_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ und $b_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ gewählt.



(Warum ist das eine Basis von \mathbb{R}^2 ?)

Dann suchen wir Zahlen α_1 und α_2 mit

$$v = \begin{pmatrix} 4 \\ 2 \end{pmatrix} = \alpha_1 b_1 + \alpha_2 b_2 = \alpha_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Der Koordinatenvektor bezüglich dieser Basis ist $\varphi_B(v) = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$.

Übung: Lösen Sie das LGS oben und bestimmen Sie damit $\varphi_B(v)$.

Die Abbildung φ_B zwischen den beiden Vektorräumen V mit \mathbb{R}^n ist eine umkehrbare Abbildung (bijektiv), die zudem linear ist. Sie erfüllt folgende Eigenschaften:

$$\varphi_B(v + w) = \varphi_B(v) + \varphi(w) \quad \text{und} \quad \varphi_B(\lambda v) = \lambda \varphi_B(v).$$

Ein solches φ_B heisst „Isomorphismus“, $V \xrightarrow{\sim} \mathbb{R}^n$.

Anwendung: Basen für den Lösungsraum \mathcal{L}_A

Sei A eine $n \times n$ -Matrix, welche ein DGL-System $y' = Ay$ definiert. Der Lösungsraum

$$\mathcal{L}_A = \{f : \mathbb{R} \rightarrow \mathbb{R}^n \mid f \in C^1(\mathbb{R}, \mathbb{R}^n) \text{ und } f'(t) = Af(t) \text{ für alle } t\}$$

ist ein Untervektorraum $\mathcal{L}_A \subset C^1(\mathbb{R}, \mathbb{R}^n)$. Es gilt:

1. Für jeden Anfangsvektor $f_0 \in \mathbb{R}^n$ gibt es genau eine Lösung $f \in \mathcal{L}_A$ mit Anfangswert

$$f(0) = f_0 = \begin{pmatrix} f_1(0) \\ f_2(0) \\ \vdots \\ f_n(0) \end{pmatrix}.$$
2. Es gibt einen Isomorphismus $\Phi : \mathbb{R}^n \rightarrow \mathcal{L}_A$, $f_0 \mapsto$ Lösung f mit $f_0 = f(0)$.
3. Es gilt $\dim \mathcal{L}_A = n$.

Unsere Aufgabe:

Suche eine geeignete Basis von \mathcal{L}_A . Das heisst, n linear unabhängige Lösungsfunktionen!

Wir wissen schon, dass für eine $n \times n$ -Matrix A gilt:

1. Sei v ein EV von A zum EW λ . Dann ist $t \mapsto y(t) = e^{\lambda t}v \in \mathbb{R}^n$ eine Lösung von $y' = Ay$ mit $y(0) = v$ (Siehe 3.2.1).
2. Angenommen A hat n linear unabhängige EV v_1, \dots, v_n zu EW $\lambda_1, \dots, \lambda_n$. (Die λ_i sind nicht unbedingt alle voneinander verschieden.)
Dann bilden die Funktionen $t \mapsto e^{\lambda_i t} v_i$, $i = 1, \dots, n$ eine Basis von \mathcal{L}_A .

Aber was ist, wenn es keine n linear unabhängigen EV von A gibt? Was können wir dann über den Lösungsraum \mathcal{L}_A sagen?

3.5 Diagonalisierbarkeit

Um diese Aufgabe anzugehen, wenden wir die Strukturtheorie von quadratischen Matrizen an.

3.5.1 Definition und Anwendung

Definition 3.5.1. Eine Matrix $A \in M_{n \times n}(\mathbb{R})$ heisst diagonalisierbar, wenn es eine invertierbare Matrix T gibt, mit $T^{-1}AT = J = D(\lambda_1, \lambda_2, \dots, \lambda_n) = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$ mit EW λ_i von A .

In diesem Fall sind die Spalten von T die EV von A , das heisst, $T = (v_1 \ v_2 \ \dots \ v_n)$ mit $Av_i = \lambda_i v_i$.

Die Theorie der Linearen Algebra verrät uns die Äquivalenz

Eine Matrix $A \in M_{n \times n}(\mathbb{R})$ ist diagonalisierbar

- \iff Die Matrix A hat n linear unabhängige EV v_1, \dots, v_n zu den EW $\lambda_1, \dots, \lambda_n$ – allenfalls sind mehrfache EW dabei.
- \iff Es gibt eine Basis des \mathbb{R}^n aus EV von A (Eigenbasis).

Die Hauptanwendung dieser Aussage erfolgt auf ein System $y' = Ay$:

Die Funktionen $t \mapsto e^{\lambda_i t} v_i, i = 1, \dots, n$ bilden genau dann eine Basis des Lösungsraumes \mathcal{L}_A , wenn die Matrix A diagonalisierbar ist, mit Eigenbasis v_1, v_2, \dots, v_n zu den EW λ_i .

Beispiele

- Die Matrix $A = \begin{pmatrix} 0 & -1 & 1 \\ -3 & -2 & 3 \\ -2 & -2 & 3 \end{pmatrix}$ hat die EW $-1, 1, 1$ mit zugehörigen EV $\begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$.

Mit $T = \begin{pmatrix} 1 & 1 & -1 \\ 3 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix}$ ist die Diagonalform $T^{-1}AT = \text{diag}(-1, 1, 1) = J = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

- Falls alle EW einer Matrix B 1-fach sind, d.h. λ_i paarweise voneinander verschieden, so ist B diagonalisierbar

$$J = \begin{pmatrix} \lambda_1 & 0 & & 0 \\ \lambda_2 & 0 & & \\ & \ddots & \ddots & \\ 0 & & \lambda_{n-1} & 0 \\ & & & \lambda_n \end{pmatrix}.$$

Eigenwerte/-vektoren einer hermiteschen Matrix

Wir schauen noch einen Fall an, in dem sich allfällige Symmetrien eines Modells in Diagonalisierbarkeit übersetzen.

Eine Matrix $A = (a_{ij}) \in M_{n \times n}(\mathbb{C})$ mit $A^T = \bar{A}$ heisst hermitesch. Falls alle $a_{ij} \in \mathbb{R}$ und $A = A^T$, heisst A symmetrisch. Für ein solches A gilt:

- Alle EW von A sind reell.
- Zu einem k -fachen EW gehören k linear unabhängige EV.
- Es gibt insgesamt n linear unabhängige EV, d.h. es gibt eine Eigenbasis des \mathbb{C}^n .

Also ist eine symmetrische/hermitesche Matrix A immer diagonalisierbar, und wir können in diesem Fall das Lösungraum \mathcal{L}_A gut beschreiben.

3.6 Exponential einer Matrix

Für $n = 1$ ist die Matrix $A = a \in \mathbb{R}$ eine Zahl, und wir sind im eindimensionalen Fall der DGL $y' = ay$ (exponentielle Entwicklung). Dort hatten wir die allgemeinen Lösung $y(t) = Ce^{at}$ und gesehen, dass dies die einzigen Lösungen sind. In unserer neuen Sprache heisst das:

Der Lösungsraum $\mathcal{L}_{A=a} \subset C^1(\mathbb{R})$ ist ein 1-dimensionaler U-VR erzeugt von der Funktion $t \mapsto e^{at}$:

$$f \in \mathcal{L}_a \iff f(t) = Ke^{at}, \text{ für alle } t$$

Im Fall $n \geq 2$ suchen wir ein Analogon zu e^{at} für die Lösung eines $n \times n$ -Systems $y' = Ay$.

3.6.1 Definition, Anwendungen, erste Berechnung von e^A

Erinnerung: Was ist e^x ? Gegeben durch eine Potenzreihe: $x \mapsto e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \dots + \frac{1}{n!}x^n + \dots$

Definition 3.6.1. Sei $A \in M_{n \times n}(\mathbb{R})$ (oder $A \in M_{n \times n}(\mathbb{C})$). Dann schreiben wir

$$e^A = E_n + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \dots + \frac{1}{n!}A^n + \dots$$

Bemerkungen:

1. Für jedes $A \in M_{n \times n}(\mathbb{R})$ (oder $A \in M_{m \times n}(\mathbb{C})$) konvergiert die Reihe e^A . Wir wollen hier zunächst nicht präzisieren, was wir unter Konvergenz verstehen.
2. Beachte, dass $e^A = (\eta_{ij})$ wieder eine $n \times n$ -Matrix ist.
3. Sei A eine Funktion

$$A : \mathbb{R} \rightarrow M_{m \times n}, t \mapsto A(t) = (a_{ij}(t)) \text{ mit } a_{ij} : \mathbb{R} \rightarrow \mathbb{R}, t \mapsto a_{ij}(t).$$

Falls $a_{ij} \in C^1(\mathbb{R})$, dann ist die Abbildung A differenzierbar mit $A'(t) = (a'_{ij}(t))$.

Für $n = 1$ ist

$$A : \mathbb{R} \rightarrow M_{m \times 1}, \text{ das heisst, } A \in C^1(\mathbb{R}, \mathbb{R}^m).$$

Theorem 3.6.2. Sei $A \in M_{n \times n}(\mathbb{R})$. Dann gilt:

1. $t \mapsto e^{tA} = E_n + tA + \frac{1}{2}t^2A^2 + \frac{1}{3!}t^3A^3 + \dots + \frac{1}{n!}t^nA^n + \dots$ ist eine differenzierbare Funktion $\mathbb{R} \rightarrow M_{n \times n}$ mit Ableitung $(e^{tA})' = (\eta'_{ij}(t)) = A \cdot e^{tA}$.
2. Die Spaltenvektoren der Matrix $e^{tA} = (\eta_{i1}(t) \ \eta_{i2}(t) \ \dots \ \eta_{in}(t))$ (mit $i = 1, \dots, n$) bilden eine Basis des Lösungsraumes \mathcal{L}_A von $y' = Ay$.
3. Wie im Fall $n = 1$ gilt

- (a) Die allgemeine Lösung y von $y' = Ay$ schreibt sich als

$$y(t) = \underbrace{e^{tA}C}_{\text{Matrix mal Vektor}} \in \mathbb{R}^n \text{ mit } C = \begin{pmatrix} C_1 \\ \vdots \\ C_n \end{pmatrix} \quad C_i \text{ konstant.}$$

- (b) Sei $y_0 \in \mathbb{R}^n$. Dann ist $y(t) = e^{tA}y_0$ die eindeutige Lösung der DGL-Systems $y' = Ay$ mit $y(0) = y_0$.

Zusammenfassung: Bestimmung einer Basis von \mathcal{L}_A

Seien $A \in M_{n \times n}(\mathbb{R})$ und $y' = Ay$ ein DGL-System.

Wie finden wir eine Basis des Lösungsraums $\mathcal{L}_A \subset C^1(\mathbb{R}, \mathbb{R}^n)$?

Falls A nicht diagonalisierbar ist (oder wir darüber nicht entscheiden können), verwende die Exponentialabbildung

$$t \mapsto e^{tA} = E_n + tA + \frac{1}{2}t^2 A^2 + \dots + \frac{1}{k!}t^k A^k + \dots$$

Das Theorem oben sagt uns:

Die Spalten der Matrix e^{tA} bilden eine Basis von \mathcal{L}_A .

Problem: Wie berechnen wir e^{tA} - ohne den Computer?

3.6.2 Berechnungsmethode für e^A

Schauen wir uns zunächst Fälle an, in denen A eine möglichst einfache Gestalt hat. Für eine beliebige Matrix werden wir diese dann anwenden.

Diagonalmatrizen

Das Exponential einer Diagonalmatrix $A = \begin{pmatrix} a_{11} & 0 & \dots & \dots & 0 \\ 0 & a_{22} & 0 & \dots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & a_{nn} \end{pmatrix}$ ist ebenfalls eine Diagonalmatrix: $e^A = \begin{pmatrix} e^{a_{11}} & 0 & \dots & \dots & 0 \\ 0 & e^{a_{22}} & 0 & \dots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & e^{a_{nn}} \end{pmatrix}$.

Blockdiagonalmatrizen

Eine Matrix der Form $A = \begin{pmatrix} A_1 & & & & \\ & A_2 & & & \\ & & \ddots & & \\ & & & A_k & \end{pmatrix}$ heisst **Blockdiagonalmatrix**. Dabei sind alle A_i quadratische Matrizen entlang der Diagonalen und sonst alle Einträge 0.

Für die obige Blockdiagonalmatrix A gilt

$$A^j = \begin{pmatrix} A_1^j & & & & \\ & A_2^j & & & \\ & & \ddots & & \\ & & & A_k^j & \end{pmatrix} \implies e^A = \begin{pmatrix} e^{A_1} & & & & \\ & e^{A_2} & & & \\ & & \ddots & & \\ & & & & e^{A_k} \end{pmatrix}.$$

Vertauschbare Matrizen

Für zwei $n \times n$ -Matrizen A und B gilt

$$\boxed{AB = BA} \iff e^{A+B} = e^A e^B.$$

Beispiel 3.6.3. Sei

$$A = \begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix} = \underbrace{\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}}_{=B} + \underbrace{\begin{pmatrix} 0 & 3 \\ 0 & 0 \end{pmatrix}}_{=C}$$

Es gilt $BC = CB$ und damit $e^A = e^{B+C} = e^B e^C$. Für die Diagonalmatrix B ist $e^B = \begin{pmatrix} e^2 & 0 \\ 0 & e^2 \end{pmatrix}$ und für e^C schauen wir die Definition als Reihe an:

$$\begin{aligned} e^C &= E_2 + C + \frac{1}{2}C^2 + \dots = E_2 + C + \frac{1}{2}\begin{pmatrix} 0 & 3 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 3 \\ 0 & 0 \end{pmatrix} + \dots \\ &= E_2 + C + \underbrace{\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} + \dots}_{\text{Nur noch Nullmatrizen}} = \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

Damit ist

$$e^A = e^{B+C} = e^B e^C = \begin{pmatrix} e^2 & 0 \\ 0 & e^2 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e^2 & 3e^2 \\ 0 & e^2 \end{pmatrix}.$$

Bemerkungen:

1. Für den allgemeinen Fall schreiben wir e^{A+B} gemäß Definition als Potenzreihe. Dann verwenden wir $(A+B)^n = \sum_{k=0}^n \binom{n}{k} A^k B^{n-k}$ und vertauschen Sie die Summationsreihenfolge.
2. Auch wenn sich die Faktoren A, B nicht vertauschen lassen, gibt es Möglichkeiten, $e^A e^B$ zu berechnen: Zum Beispiel besagt die Baker-Campbell-Hausdorff-Formel, dass

$$e^A e^B = e^{A+B+\frac{1}{2}[A,B]+\frac{1}{3!}[A,[A,B]]+\dots},$$

wobei $[A, B] = AB - BA$.

Inverse Matrix

Wenden wir das Kriterium der Vertauschbarkeit auf die Inverse Matrix an:

Die Matrix e^A ist invertierbar mit $(e^A)^{-1} = e^{-A}$, denn:

$$(-A)A = A(-A) \implies e^A e^{-A} = e^{A-A} = e^0 = E_n \implies (e^A)^{-1} = e^{-A}.$$

Da die Inverse $(e^A)^{-1} = e^{-A}$ jeweils existiert, muss also auch immer $\det(e^A) \neq 0$ sein. Diese Zahl können wir noch genauer bestimmen.

Eigenwerte und Determinante von e^A

Sei λ ein EW mit EV v der Matrix A . Dann rechnen wir

$$\begin{aligned}
 (e^A) v &= \left(\sum_{k=1}^{\infty} \frac{1}{k!} t^k A^k \right) v && \text{Definition} \\
 &= \left(\sum_{k=1}^{\infty} \frac{1}{k!} t^k A^k \cdot v \right) && \text{Erlaubt die Theorie} \\
 &= \left(\sum_{k=1}^{\infty} \frac{1}{k!} t^k \lambda^k \cdot v \right) && \text{Da EW, EV und } A^k \cdot v = \lambda^k \cdot v \\
 &= \left(\sum_{k=1}^{\infty} \frac{1}{k!} t^k \lambda^k \right) \cdot v && \text{Erlaubt die Theorie} \\
 &= e^\lambda \cdot v && \text{Definition}
 \end{aligned}$$

Damit ist also v ein EV der Matrix e^A zum EW e^λ .

Daraus ergibt sich direkt die Determinante $\det(e^A) = \text{Produkt der Eigenwerte der Matrix}$.

Seien $\lambda_1, \lambda_2, \dots, \lambda_n$ die EW der Matrix A . Dann sind $e^{\lambda_1}, e^{\lambda_2}, \dots, e^{\lambda_n}$ die EW der Matrix e^A , und es gilt

$$\det(e^A) = e^{\lambda_1} \cdot e^{\lambda_2} \cdot \dots \cdot e^{\lambda_n} = e^{\lambda_1 + \lambda_2 + \dots + \lambda_n}.$$

Exponential eines Jordan-Blocks

Definition 3.6.4. Eine $n \times n$ -Matrix der Form $J = \begin{pmatrix} \lambda & 1 & 0 & \dots & \\ 0 & \lambda & 1 & 0 & \dots \\ & & \ddots & \ddots & \\ 0 & \dots & & \lambda & 1 \\ 0 & \dots & & & \lambda \end{pmatrix}$ heisst **Jordan-Block**

Nach Marie Ennemond Camille Jordan (1838 – 1922)



Damit haben wir folgende Aussage über das Matrix-Exponential eines Jordan-Blocks

Lemma 3.6.5. Sei $t \in \mathbb{R}$, dann gilt:

$$e^{Jt} = e^{\lambda t} \begin{pmatrix} 1 & t & \frac{t^2}{2!} & \frac{t^3}{3!} & \cdots & \frac{t^{n-1}}{(n-1)!} \\ 0 & 1 & t & \frac{t^2}{2!} & \frac{t^3}{3!} & \cdots \\ & \ddots & \ddots & \ddots & & \\ 0 & \cdots & 1 & t & \frac{t^2}{2!} & \\ 0 & \cdots & & 1 & t & \\ 0 & \cdots & & & 1 & \end{pmatrix}.$$

Auf den Nebendiagonalen stehen also just die Summanden der Taylor-Reihe e^t .

Dies rechnen wir mit der Definition nach:

$$J = \begin{pmatrix} \lambda & 1 & 0 & \cdots & \\ 0 & \lambda & 1 & 0 & \cdots \\ & \ddots & \ddots & \ddots & \\ 0 & \cdots & & \lambda & 1 \\ 0 & \cdots & & & \lambda \end{pmatrix} = \underbrace{\lambda E_n}_A + \underbrace{\begin{pmatrix} 0 & 1 & 0 & \cdots & \\ 0 & 0 & 1 & 0 & \cdots \\ & \ddots & \ddots & \ddots & \\ 0 & \cdots & 0 & 1 & \\ 0 & \cdots & & 0 & \end{pmatrix}}_B.$$

Es gilt $AB = BA$. Also folgt für $t \in \mathbb{R}$,

$$e^{Jt} = e^{At+Bt} = e^{\lambda t} e^{Bt} = e^{\lambda t} E_n \left(E_n + Bt + \frac{t^2}{2!} B^2 + \frac{t^3}{3!} B^3 + \dots \right).$$

Es ist

$$B^2 = \begin{pmatrix} 0 & 0 & 1 & 0 & \cdots & \\ 0 & 0 & 0 & 1 & 0 & \cdots \\ & \ddots & \ddots & \ddots & & \\ 0 & \cdots & 0 & 0 & 1 & \\ 0 & \cdots & & 0 & 0 & \\ 0 & \cdots & & & 0 & \end{pmatrix}$$

usw. und damit B^k eine Matrix, die auf der k -ten Nebendiagonale 1 und sonst lauter 0 hat.
Also ist

$$B^{n-1} = \begin{pmatrix} 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}, B^n = 0, B^{n+1} = 0, \text{ usw.}$$

Beispiel 3.6.6. $J = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix} \Rightarrow e^{tJ} = e^{3t} \begin{pmatrix} 1 & t & \frac{1}{2}t^2 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix}.$

Konjugation

Sei T eine invertierbare Matrix und $T^{-1}AT = J$, so gilt

$$e^A = Te^JT^{-1}, \quad \text{Reihenfolge!}$$

denn: Aus $T^{-1}AT = J \iff A = TJT^{-1}$ folgt

$$\begin{aligned} A^k &= (TJT^{-1})^k = \underbrace{(TJ\textcolor{red}{T}^{-1})(\textcolor{red}{T}J\textcolor{green}{T}^{-1})(\textcolor{green}{T}J\textcolor{blue}{T}^{-1}) \dots (\textcolor{blue}{T}J\textcolor{blue}{T}^{-1})}_{k \text{ Faktoren}} \\ &= TJ^kT^{-1} \end{aligned}$$

weil sich die farbigen Faktoren aufheben. Also

$$e^A = \sum_{k=0}^{\infty} \frac{\textcolor{orange}{A}^k}{k!} = \sum_{k=0}^{\infty} \frac{\textcolor{orange}{T}J^k\textcolor{orange}{T}^{-1}}{k!} = T \left(\sum_{k=0}^{\infty} \frac{J^k}{k!} \right) T^{-1} = Te^J T^{-1}.$$

Beispiel 3.6.7. Sei $A = \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix}$. Wir möchten die Konjugation anwenden: Wählen wir als $T = \begin{pmatrix} 1 & -2 \\ 1 & 1 \end{pmatrix}$ die Matrix mit EV von A als Spalten, und somit $T^{-1} = \frac{1}{3} \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix}$, so wird $T^{-1}AT$ diagonal mit den EW von A auf der Diagonalen:

$$T^{-1}AT = \begin{pmatrix} 5 & 0 \\ 0 & 2 \end{pmatrix} = D \text{ und } e^{tD} = \begin{pmatrix} e^{5t} & 0 \\ 0 & e^{2t} \end{pmatrix},$$

Somit

$$e^{tA} = Te^{tD}T^{-1} = \frac{1}{3} \begin{pmatrix} e^{5t} + 2e^{2t} & 2e^{5t} - 2e^{2t} \\ e^{5t} - e^{2t} & 2e^{5t} + e^{2t} \end{pmatrix}.$$

Bemerkung:

Gelingt es, A durch „Konjugation“ $T^{-1}AT = J$ auf eine einfache Form zu bringen, so können wir das Exponential von A leichter berechnen. Tatsächlich sagt der Satz über die Jordan-Normalform (siehe unten), dass dies immer erreicht werden kann: Die Matrix J besteht aus lauter Jordanblöcken entlang der Diagonalen.

3.6.3 Berechnung von e^A mit der Jordan-Normalform

Falls A nicht diagonalisierbar ist, hilft uns der Satz von Jordan bei der Berechnung von e^A .

Theorem 3.6.8 (Jordan-Normalform (JNF)). *Sei $A \in M_{n \times n}(\mathbb{C})$.*

1. Dann existiert eine invertierbare Matrix T , sodass $T^{-1}AT = J$ für eine Jordan-Blockdiagonalmatrix

$$J = \begin{pmatrix} \boxed{J_1} & & & \\ & \boxed{J_2} & & \\ & & \ddots & \\ & & & \boxed{J_k} \end{pmatrix}.$$

2. Jeder Jordan-Block hat die Form

$$J_i = \begin{pmatrix} \lambda_i & 1 & & 0 & & \\ \lambda_i & 1 & & & & \\ & \ddots & \ddots & & & \\ 0 & & & \lambda_i & 1 & \\ & & & & \lambda_i & \end{pmatrix} \in M_{n_i \times n_i}.$$

3. Die λ_i sind die EW von A , nicht unbedingt alle voneinander verschieden.

Bemerkungen:

1. Die Anzahl der Jordan-Blöcke zum EW λ_i und die Länge eines einzelnen Jordan-Blocks lassen sich (theoretisch, explizit und per Hand) berechnen. Die Entwicklung des zugehörigen Algorithmus sprengt den Rahmen unserer Ausführungen.
2. Falls A diagonalisierbar ist, ist die Jordan-Normalform eine Diagonalmatrix $J = D(\lambda_1, \dots, \lambda_n)$.

Anwendung / Zusammenfassung Exponential der Jordan-Normalform.

Seien $A \in M_{n \times n}(\mathbb{C})$ und $T^{-1}AT = J$ die Jordan-Normalenform. Dann ist

$$e^{tJ} = \begin{pmatrix} e^{tJ_1} & & & 0 \\ & e^{tJ_2} & & \\ & & \ddots & \\ 0 & & & e^{tJ_k} \end{pmatrix} \text{ mit } J = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k \end{pmatrix},$$

wobei die J_i die Jordan-Blöcke von J sind.

$$\text{Für } J = J_i = \begin{pmatrix} \lambda & 1 & & 0 \\ & \lambda & 1 & \\ & & \ddots & \ddots \\ 0 & & & \lambda & 1 \end{pmatrix} \text{ gilt } e^{Jt} = e^{\lambda t} \begin{pmatrix} 1 & t & \frac{t^2}{2!} & \frac{t^3}{3!} & \cdots & \frac{t^{n-1}}{(n-1)!} \\ 0 & 1 & t & \frac{t^2}{2!} & \frac{t^3}{3!} & \cdots \\ & & \ddots & \ddots & \ddots & \\ 0 & \dots & 1 & t & \frac{t^2}{2!} & \\ 0 & \dots & & 1 & t & \\ 0 & \dots & & & & 1 \end{pmatrix}.$$

Beispiele Jordan-Blöcke zu einem EW.

Wir betrachten den Teil der JNF mit dem EW λ_i auf der Diagonale und geben - bis auf Permutationen- die Möglichkeiten in Abhängigkeit von der Vielfachheit des EW λ_i an:

EW 1-fach. Es gibt eine Möglichkeit:

$$(\lambda_i).$$

Falls alle EW 1-fach, d.h. es gibt n Stück mit $\lambda_i \neq \lambda_j$, dann ist

$$J = \begin{pmatrix} \lambda_1 & 0 & & 0 \\ & \lambda_2 & 0 & \\ & & \ddots & \ddots \\ 0 & & & \lambda_{n-1} & 0 \\ & & & & \lambda_n \end{pmatrix}.$$

EW 2-fach. Hier gibt es zwei Möglichkeiten wir den Jordan-Block der Länge 2:

$$\begin{pmatrix} \lambda_i & 0 \\ & \lambda_i \end{pmatrix} \text{ oder } \begin{pmatrix} \lambda_i & 1 \\ & \lambda_i \end{pmatrix}.$$

EW 3-fach. Hier haben die Möglichkeiten

$$\begin{pmatrix} \lambda_i & 1 \\ \lambda_i & 1 \\ \lambda_i & 1 \end{pmatrix} \text{ oder } \begin{pmatrix} \lambda_i & 0 \\ \lambda_i & 1 \\ \lambda_i & 1 \end{pmatrix} \text{ oder } \begin{pmatrix} \lambda_i & 0 \\ \lambda_i & 0 \\ \lambda_i & 0 \end{pmatrix}.$$

Zahlenbeispiele

Sei $A = \begin{pmatrix} 1 & -3 & -2 \\ -1 & 1 & -1 \\ 2 & 4 & 5 \end{pmatrix}$. Dann liefert *der Computer*

$$T = \begin{pmatrix} -1 & 1 & -1 \\ -1 & 0 & 0 \\ 2 & 0 & 1 \end{pmatrix}, \quad J = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Die Matrix ist **nicht** diagonalisierbar, es gibt keine Eigenbasis.

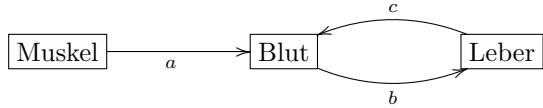
Sei $A = \begin{pmatrix} 0 & -1 & 1 \\ -3 & -2 & 3 \\ -2 & -2 & 3 \end{pmatrix}$. Dann liefert *der Computer*

$$T = \begin{pmatrix} 1 & 1 & -1 \\ 3 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix}, \quad J = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Die Matrix ist **diagonalisierbar**, es gibt eine **Eigenbasis**.

3.6.4 Anwendung serielles 3-Box-Modell mit Rückfluss.

Gegeben sei folgendes 3-Box-Kompartiment-Modell für $0 < a, b, c < 1$:



Das zugehörige DGL-System $y' = Ay$ ist $y' = \begin{pmatrix} -a & 0 & 0 \\ a & -b & c \\ 0 & b & -c \end{pmatrix} y$.

Ein EW von A ist immer $\lambda = 0$ (Übung: Warum?) mit EV $v \neq 0$.

Dann ist $\{t \mapsto y_\infty(t) = e^{0t}v = v\}$ eine stationäre Lösung.

Wir wählen die Zahlen a, b, c einmal so, dass A **diagonalisierbar** ist, und einmal so, dass dies **nicht der Fall** ist, und untersuchen das qualitative Verhalten der Lösungskurven.

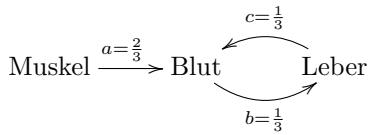
Erster Fall: A diagonalisierbar.

Wir suchen eine Basis des Lösungsraumes \mathcal{L}_A für $a = \frac{2}{3}, b = \frac{1}{3}, c = \frac{2}{3}$.

Die Matrix $A = \frac{1}{3} \begin{pmatrix} -2 & 0 & 0 \\ 2 & -1 & 2 \\ 0 & 1 & -2 \end{pmatrix}$ hat die Eigenwerte $\lambda_{1,2,3} = -1, -\frac{2}{3}, 0$ mit den Eigenvektoren $(v_1, v_2, v_3) = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 2 \\ 1 & 1 & 1 \end{pmatrix}$. Die drei EW sind einfach (da alle paarweise verschieden), und damit bilden die drei linear unabhängigen EV eine Eigenbasis von A . Somit erhalten wir als Basis von \mathcal{L}_A :

$$\left\{ t \mapsto e^{-t} \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, t \mapsto e^{-\frac{2}{3}t} \underbrace{\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}}_{\text{stationär}}, t \mapsto \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} \right\}.$$

Zweiter Fall: A nicht diagonalisierbar.



Es wird beschrieben durch $y' = Ay$ mit $A = \begin{pmatrix} -a & 0 & 0 \\ a & -b & c \\ 0 & b & -c \end{pmatrix} = \begin{pmatrix} -\frac{2}{3} & 0 & 0 \\ \frac{2}{3} & -\frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & -\frac{1}{3} \end{pmatrix}$, und den EW $\lambda_{1,2} = -\frac{2}{3}, \lambda_3 = 0$ mit EV $v_{1,2} = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, v_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$. Also besitzt \mathcal{L}_A keine Eigenbasis und ist daher nicht diagonalisierbar. Aber für

$$T = \begin{pmatrix} 0 & -3 & 0 \\ -1 & 3 & 1 \\ 1 & 0 & 1 \end{pmatrix} \text{ folgt } T^{-1}AT = \begin{pmatrix} -\frac{2}{3} & 1 & 0 \\ 0 & -\frac{2}{3} & 0 \\ 0 & 0 & 0 \end{pmatrix} = J.$$

Die Lösung von $y' = Ay$ mit Anfangswert $y(0) = y_0$ lautet

$$y(t) = e^{At}y_0 = Te^{Jt}T^{-1}y_0 = T \begin{pmatrix} e^{-\frac{2}{3}t} & te^{-\frac{2}{3}t} & 0 \\ 0 & e^{-\frac{2}{3}t} & 0 \\ 0 & 0 & 1 \end{pmatrix} T^{-1}y_0.$$

Anders gesagt (Übung), erhalten wir eine Basis \mathcal{L}_A :

$$\left\{ t \mapsto e^{-\frac{2}{3}t} \begin{pmatrix} 3 \\ t-3 \\ -t \end{pmatrix}, t \mapsto e^{-\frac{2}{3}t} \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, t \mapsto \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}.$$

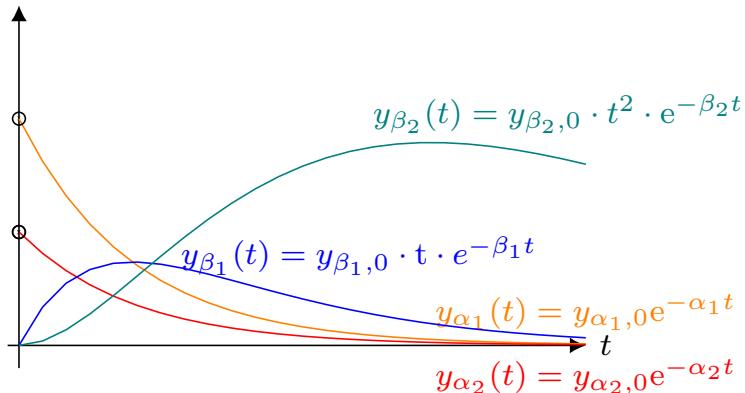
Diagonalisierbar vs. nicht diagonalisierbar.

Der Unterschied zwischen einer (nicht-)diagonalisierbaren Matrix A , spiegelt sich qualitativ im Konvergenzverhalten der Lösungsfunktionen des DGL-Systems $y' = Ay$ wieder. Die Terme der Lösungen sind beschrieben durch:

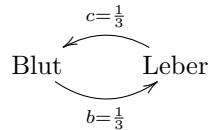
Diagonalsierbar: In diesem Fall haben Basisvektoren die Form $t \mapsto e^{-\alpha t} \cdot v$ mit einem EV v .

Nicht diagonalsierbar: Hier finden wir eine Basis mit Funktionen $t \mapsto te^{-\alpha t} \cdot v$ oder allgemeiner durch $t \mapsto q(t)e^{-\alpha t} \cdot v$ für ein Polynom q . Dabei ist v nicht notwendig ein Eigenvektor.

Das folgende Bild zeigt das Verhalten solcher Funktionen:



Anschaulich heisst dies, dass es im Fall der Nicht-Diagonalsierbarkeit zu einem Stau kommen kann. Bei Diagonalsierbarkeit haben wir jeweils einen exponentiellen Abbau. In dem Beispiel oben ist es anschaulich naheliegend, da der Austausch



mit der gleichen Rate erfolgt.

3.6.5 Vergleich der Methoden

Wir betrachten das Beispiel $y' = A \cdot y$ mit $A = \begin{pmatrix} 5 & 1 \\ -4 & 1 \end{pmatrix}$, dem doppelten EW $\alpha = 3$ und EV $\begin{pmatrix} 1 \\ -2 \end{pmatrix} \cdot t$ mit $t \in \mathbb{R}, t \neq 0$. Wir finden also nicht direkt eine Basis des Lösungsraumes \mathcal{L}_A .

Ohne das Matrix-Exponential können wir die allgemeine Lösung dann mit Hilfe einer linearen DGL 2. Ordnung bestimmen:

- Wir erhalten für die erste Komponente y_1 die DGL

$$y_1'' - 6y_1' + 9y_1 = 0$$

- Die zugehörige charakteristische Gleichung mit doppelter Nullstelle liefert

$$y_1(x) = (C_1 + C_2x)e^{3x}$$

- Zusammen erhalten wir: $y(x) = C_1 e^{3x} \begin{pmatrix} 1 \\ -2 \end{pmatrix} + C_2 e^{3x} \begin{pmatrix} x \\ 1-2x \end{pmatrix}$

Nun zeigt Computer $e^{xA} = \begin{pmatrix} e^{3x}(1+2x) & e^{3x}x \\ -4e^{3x}x & e^{3x}(1-2x) \end{pmatrix} = (b_1(x) \ b_2(x))$.

Frage: Sei zum Beispiel $y(0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Liefern beide Darstellung dieselbe Lösung y ?

Um dies zu beantworten, rechnen wir

$$e^{xA}y_0 = \begin{pmatrix} e^{3x}(1+2x) & e^{3x}x \\ -4e^{3x}x & e^{3x}(1-2x) \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} e^{3x} + 3e^{3x}x \\ e^{3x} - 3e^{3x}x \end{pmatrix}$$

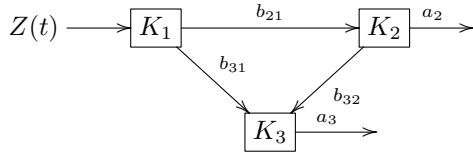
aus.

Im anderen Fall setzen wir $x = 0$ ein $y(x) = C_1 e^{3x} \begin{pmatrix} 1 \\ -2 \end{pmatrix} + C_2 e^{3x} \begin{pmatrix} x \\ 1-2x \end{pmatrix}$ und lösen dann das LGS $y(0) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ für die beiden gesuchten Konstanten C_1 und C_2 . Wir erhalten $C_1 = 1$ und $C_2 = 3$. Mit diesen ergibt sich dann wieder $y(x) = \begin{pmatrix} e^{3x} + 3e^{3x}x \\ e^{3x} - 3e^{3x}x \end{pmatrix}$.

3.7 Der inhomogene Fall

Bisher können wir allein ein *homogenes* System $y' = Ay$ lösen.

In der Praxis tauchen *inhomogene* Systeme $y' = Ay + g$ auf, wenn wir zum Beispiel in einem Kompartiment-System eine Zufuhr gegeben haben, durch eine Funktion $t \mapsto Z(t)$:



Im Kompartiment K_i haben wir die Menge $y_i(t)$ und deren Veränderungsgeschwindigkeit $y'_i(t)$. Die Entwicklung (mit allen Raten positiv) wird dann beschrieben durch das inhomogene System $y'_1(t) = Z(t) - (b_{31} + b_{21})y_1(t)$, $y'_2(t) = b_{21}y_1(t) - (a_2 + b_{32})y_2(t)$, $y'_3(t) = b_{31}y_1(t) + b_{32}y_2(t) - a_3y_3(t)$, oder in der Matrixschreibweise $y'(t) = Ay(t) + g(t)$ mit

$$A = \begin{pmatrix} -(b_{31} + b_{21}) & 0 & 0 \\ b_{21} & -(a_2 + b_{32}) & 0 \\ b_{31} & b_{32} & -a_3 \end{pmatrix}, \quad g(t) = \begin{pmatrix} Z(t) \\ 0 \\ 0 \end{pmatrix}.$$

Erinnerung: Allgemeine Lösung einer inhomogenen DGL

Wir kennen zwei Verfahren zum Lösen einer inhomogenen DGL $y'(x) = p(x)y(x) + q(x)$.

1. Variation der Konstanten: Die allgemeine Lösung ist

$$y(x) = (K_0(x) + C) e^{P(x)},$$

mit

- $P'(x) = p(x)$ und C eine Konstante
- $K_0(x) \in \int q(x) e^{-P(x)} dx$ eine beliebige fest gewählte Stammfunktion von $x \mapsto q(x)e^{-P(x)}$.

2. Partikulärlösungen einer DGL

- Sei y_H die allgemeine Lösung der homogenen DGL $y'(x) = p(x)y(x)$.
- Sei y_{sp} eine spezielle Lösung von $y'(x) = p(x)y(x) + q(x)$.
- Dann ist $y = y_H + y_{sp}$ die allgemeine Lösung der inhomogenen DGL $y'(x) = p(x)y(x) + q(x)$.
- Um eine spezielle Lösung y_{sp} zu finden, gibt es neben genauem Hinschauen abhängig von q unterschiedliche Ansätze. Meist ist Variation der Konstanten die zielführende Methode.

Im höher dimensionalen Fall haben wir drei (analoge) Hauptmethoden zur Lösung des inhomogenen Systems:

- Finden einer partikulären Lösung
- Entkoppeln des Systems bei Diagonalsierbarkeit
- Variation der Konstanten

3.7.1 Partikuläre Lösung eines Systems von DGL

Seien nun $A \in M_{n \times n}$ und $g \in C^1(\mathbb{R}, \mathbb{R}^n)$. Gesucht werden die Lösungen $y \in C^1(\mathbb{R}, \mathbb{R}^n)$ des **inhomogenen** Systems

$$y' = Ay + g. \quad (\text{I})$$

Für $g = 0$ konstant, ist das zugehörige **homogene** System

$$y' = Ay. \quad (\text{H})$$

Es gilt analog zum 1-dimensionalen Fall:

Die allgemeine Lösung von (I) ist eine partikuläre Lösung von (I) plus die allgemeine Lösung von (H).

Die allgemeine Lösung für (H) ist $y_H(t) = e^{tA}C$ für beliebiges $C \in \mathbb{R}^n$. Somit ist die allgemeine Lösung von (I):

$$\{t \mapsto y(t) = y_{sp}(t) + e^{tA}C\} \in C^1(\mathbb{R}, \mathbb{R}^n).$$

Die Lösung des AWP für (I) mit $y(0) = y_0 \in \mathbb{R}^n$ lautet somit

$$\{t \mapsto y(t) = y_{sp}(t) + e^{tA}(y_0 - y_{sp}(0))\} \in C^1(\mathbb{R}, \mathbb{R}^n).$$

Beispiel 3.7.1. Für $y_{sp} = y_\infty$ ein Stationärzustand:

Sei $Z(t) = Z$ konstant. Dann ist das System gegeben durch

$$A = \begin{pmatrix} -(b_{31} + b_{21}) & 0 & 0 \\ b_{21} & -(a_2 + b_{32}) & 0 \\ b_{31} & b_{32} & -a_3 \end{pmatrix}, \quad g(t) = g = \begin{pmatrix} Z \\ 0 \\ 0 \end{pmatrix}.$$

Falls A invertierbar ist, können wir einen stationären Zustand y_∞ angeben: Für y_∞ steht auf der linken Seite der Gleichung $y' = Ay + g$ und damit

$$0 = Ay_\infty + g \implies Ay_\infty = -g \implies y_\infty = -A^{-1}g.$$

In unserem Beispiel ist $\det A \neq 0$, da alle Raten > 0 angenommen wurden.

Damit ist die AWP-Lösung von

$$\begin{aligned} y'(t) &= Ay(t) + g \\ y(0) &= y_0 \end{aligned}$$

gleich $y(t) = y_\infty + e^{tA}(y_0 - y_\infty)$.

3.7.2 Lösung bei Diagonalisierbarkeit

Seien $A \in M_{n \times n}$ und $g \in C^1(\mathbb{R}, \mathbb{R}^n)$.

Wir suchen wieder die Lösungen $y \in C^1(\mathbb{R}, \mathbb{R}^n)$ des **inhomogenen** Systems $y' = Ay + g$. Wir nehmen an, dass A diagonalisierbar ist. Es gibt also eine invertierbare Matrix T mit

$$T^{-1}AT = J = D(\lambda_1, \lambda_2, \dots, \lambda_n) = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}.$$

Sei $x(t) = T^{-1}y(t)$. Dann ist

$$y(t) = Tx(t) \implies y'(t) = (Tx(t))' = Tx'(t)$$

und

$$\begin{aligned} y'(t) = Ay(t) + g(t) &\implies Tx'(t) = A(Tx(t)) + g(t) \\ &\implies T^{-1}(Tx'(t)) = T^{-1}((AT)x(t) + g(t)) \\ &\implies \underbrace{(T^{-1}T)}_{=E_n} x'(t) = \underbrace{(T^{-1}AT)}_{=D} x(t) + \underbrace{T^{-1}g(t)}_{=h(t)}. \end{aligned}$$

Wir erhalten ein neues inhomogenes DGL-System: $x'(t) = Dx(t) + h(t)$ oder in Koordinaten:

$$\begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix} = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{pmatrix}.$$

Wir sehen dann n inhomogene DGLn 1. Ordnung

$$\begin{aligned} x'_1 &= \lambda_1 x_1 + h_1 \\ x'_2 &= \lambda_2 x_2 + h_2 \\ &\vdots \\ x'_n &= \lambda_n x_n + h_n. \end{aligned}$$

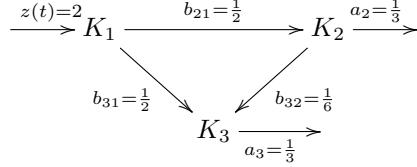
Diese sind entkoppelt und können separat gelöst werden, zum Beispiel mit der Variation der Konstanten. Dann ist

$$x_i(t) = e^{\lambda_i t} \int e^{-\lambda_i s} h_i(s) ds + C_i e^{\lambda_i t}$$

und die gesuchte Lösung

$$y = Tx.$$

Beispiel Entkopplung bei konstanter Zufuhr



wird beschrieben durch $y'(t) = Ay(t) + g(t)$ mit $A = \begin{pmatrix} -1 & 0 & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{6} & -\frac{1}{3} \end{pmatrix}$, dem konstanten Vektor

$$g(t) = \begin{pmatrix} z(t) \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} \text{ und den EW } \color{red}{-1}, \color{green}{-\frac{1}{2}}, \color{blue}{-\frac{1}{3}}. \text{ Für eine Eigenbasis haben wir die Vektoren} \\ \begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Damit gilt für $T = \begin{pmatrix} -2 & 0 & 0 \\ 2 & -1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$ dass $T^{-1}AT = \text{diag}(\color{red}{-1}, \color{green}{-\frac{1}{2}}, \color{blue}{-\frac{1}{3}})$.

Substitution $y = Tz$, also $y' = Tz'$, liefert

$$y' = Ay + g \implies Tz' = ATz + g.$$

Multiplikation von links mit T^{-1} ergibt

$$z' = T^{-1}ATz + T^{-1}g = \text{diag}(\color{red}{-1}, \color{green}{-\frac{1}{2}}, \color{blue}{-\frac{1}{3}})z + T^{-1}g.$$

Da $T^{-1} = \begin{pmatrix} -\frac{1}{2} & 0 & 0 \\ -1 & -1 & 0 \\ \frac{3}{2} & 1 & 1 \end{pmatrix}$ lautet $z' = \text{diag}(\color{red}{-1}, \color{green}{-\frac{1}{2}}, \color{blue}{-\frac{1}{3}})z + T^{-1}g$ ausgeschrieben:

$$\begin{aligned} z'_1 &= \color{red}{-1}z_1 - 1 \implies z_1(t) = c_1 e^{\color{red}{-t}} - 1 \\ z'_2 &= \color{green}{-\frac{1}{2}}z_2 - 2 \implies z_2(t) = c_2 e^{\color{green}{-\frac{t}{2}}} - 4 \\ z'_3 &= \color{blue}{-\frac{1}{3}}z_3 + 3 \implies z_3(t) = c_3 e^{\color{blue}{-\frac{t}{3}}} + 9. \end{aligned}$$

Wir sehen also, dass hier die **Lösung für z** einfach(er) zu finden ist, da die Substitution das ursprüngliche System für y **entkoppelt** hatte! Daraus ergibt sich Lösung mit Substitution: $y(t) = Tz(t)$.

Mehrdimensionale Variation der Konstanten

Wir betrachten nun ein inhomogenes Anfangswertproblem

$$y'(t) = Ay(t) + g(t), \quad A \in M_{n \times n}, \quad g \in C^1(\mathbb{R}, \mathbb{R}^n), \quad y(0) = y_0 \in \mathbb{R}^n.$$

Dann ist die Lösung des Anfangswertproblems gegeben durch

$$y(t) = e^{tA}y_0 + \int_0^t e^{(t-\tau)A}g(\tau)d\tau \quad \text{Mit Integration in jeder Koordinate.}$$

Wir machen wie im eindimensionalen den Ansatz für das homogene DGL-System der Form $y(t) = e^{tA}C(t)$ mit einer Funktion $C \in C^1(\mathbb{R}, \mathbb{R}^n)$.

Wegen der Linearität der Matrix-Vektor-Multiplikation $e^{tA} \cdot C(t)$ gilt die Produktregel (*) auch für matrix- oder vektorwertige Funktionen, und wir rechnen mit $(e^{tA})' \stackrel{**}{=} Ae^{tA}$:

$$y'(t) = (e^{tA}C(t))' \stackrel{*}{=} (e^{tA})'C(t) + e^{tA}C'(t) \stackrel{**}{=} Ae^{tA}C(t) + e^{tA}C'(t).$$

Der Ansatz und die Ableitung werden nun in $y'(t) = Ay(t) + g(t)$ eingesetzt:

$$Ae^{tA}C(t) + e^{tA}C'(t) = Ae^{tA}C(t) + g(t).$$

Nach Kürzen und Umstellen (mit $(e^{tA})^{-1} = e^{-tA}$) ist

$$C'(t) = e^{-tA}g(t) \implies C(t) = \int_0^t e^{-uA}g(u)du + \begin{pmatrix} K_1 \\ \vdots \\ K_n \end{pmatrix},$$

mit $K_1, \dots, K_n \in \mathbb{R}$.

Dieses $C(t)$ setzen wir in $y(t) = e^{tA}C(t)$ ein, es folgt die allgemeine Lösung der inhomogenen Differentialgleichung

$$y(t) = e^{tA}C(t) = e^{tA} \begin{pmatrix} K_1 \\ \vdots \\ K_n \end{pmatrix} + e^{tA} \int_0^t e^{-\tau A}g(\tau)d\tau.$$

Mit dem Anfangswert $y(0) = y_0$ ergibt sich mit $\int_0^0 e^{-\tau A}g(\tau)d\tau = 0$ für den Vektor $\begin{pmatrix} K_1 \\ \vdots \\ K_n \end{pmatrix} = y_0$.

Beispiel 3.7.2. Mit dieser Methode finden wir (in den Übungen) die Lösung des Systems

$$y'(t) = Ay(t) + g(t), \quad A = \begin{pmatrix} -2 & 1 \\ -1 & 0 \end{pmatrix}, \quad g(t) = \begin{pmatrix} 1 \\ e^{-t} \end{pmatrix}, \quad y(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

als

$$y(t) = \begin{pmatrix} \frac{1}{2}t^2e^{-t} + 2te^{-t} \\ \frac{1}{2}t^2e^{-t} + 3te^{-t} + 2e^{-t} - 1 \end{pmatrix}.$$

KAPITEL 4

Fourier-Reihen

4.1 Einführung

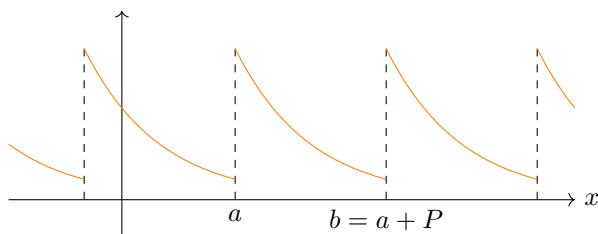
In diesem Kapitel diskutieren wir Fourier-Reihen. Damit können wir periodische Funktionen approximieren und besser verstehen. Periodische Funktionen modellieren periodische oder zyklische Phänomene, die in den Anwendungen natürlich oder auch erzwungen auftauchen.

4.1.1 Erinnerung periodische Funktionen

Definition 4.1.1. Seien $P > 0$ und $f : \mathbb{R} \rightarrow \mathbb{R}$ eine Funktion. Falls für alle $x \in \mathbb{R}$ gilt $f(x + P) = f(x)$, so heisst f **periodisch** mit Periode P .

Bemerkungen

1. Für eine Periode P ist auch $k \cdot P$ ($k \in \mathbb{Z}$) eine Periode: $f(x + kP) = f(x)$.
2. Die kleinste Zahl $P > 0$ mit $f(x + P) = f(x)$ für alle $x \in \mathbb{R}$ heisst **Hauptperiode** oder **kleinste** oder **primitive** Periode.
3. Es genügt also, eine periodische Funktion mit Hauptperiode P auf einem Intervall der Länge P zu kennen und zu untersuchen.
4. Umgekehrt untersuchen wir Funktionen f , welche auf einem Intervall $[a, b[$ definiert sind, und setzen f periodisch fort mit Periode $P = b - a$.



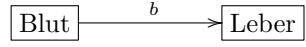
5. Seien f_1, f_2, \dots, f_n mit Periode P . Dann hat jede Linearkombination $g = \sum_{i=1}^n \alpha_i f_i$ auch Periode P , ebenso (endliche) Produkte der f_i .
6. Zunächst beschränken wir uns auf den Fall $P = 2\pi$. Falls f die Periode T hat, führen wir dies auf den Fall $P = 2\pi$ zurück, denn:

Sei f mit $f(x+T) = f(x)$. Dann hat die Funktion g mit $g(x) = f\left(\frac{Tx}{2\pi}\right)$ die Periode 2π :

$$\begin{aligned} g(x+2\pi) &= f\left(\frac{T(x+2\pi)}{2\pi}\right) = f\left(\frac{Tx}{2\pi} + \frac{T2\pi}{2\pi}\right) = f\left(\frac{Tx}{2\pi} + T\right) = f\left(\frac{Tx}{2\pi}\right) \\ &= g(x) \end{aligned}$$

4.1.2 Beispiele Medikamentenmenge

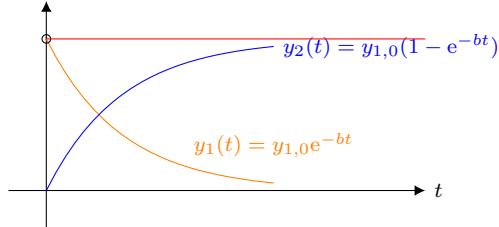
Bei den linearen Modellen hatten wir die Entwicklung eines Medikaments in Blut und Leber als Beispiel eines 2-Box-Modells



betrachtet: Zur Zeit t befindet sich die Medikamentenmenge $y_1(t)$ im Blut, $y_2(t)$ in der Leber, mit den Startwerten $y_1(0) = y_{1,0}$ und $y_2(0) = 0$.

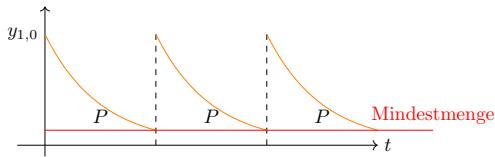
Die Entwicklungen $t \mapsto y_1(t), t \mapsto y_2(t)$ sind gegeben durch:

$$\begin{aligned} y_1(t) &= y_{1,0} \cdot e^{-bt}, \\ y_2(t) &= y_{1,0} \cdot (1 - e^{-bt}). \end{aligned}$$



Wollen wir sicherstellen, dass eine Mindestmenge \mathcal{M} im Blut mit $0 < \mathcal{M} < y_{1,0}$ nicht unterschritten wird, müssen wir die Infusion periodisch neu starten, mit Anfangswert $y_{1,0}$.

Die Periode P berechnet sich als $P = \frac{1}{b}(\ln y_{1,0} - \ln \mathcal{M})$ (siehe Übungen). Der Verlauf der Funktion $t \mapsto y_{1,\mathcal{M}}(t)$ ist dann

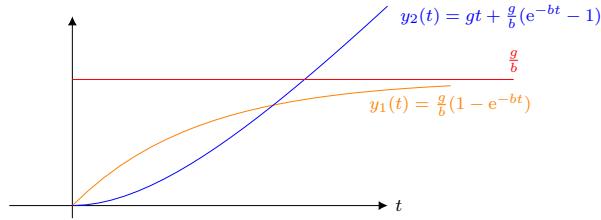


Konstante Infusion. Betrachten wir die Entwicklung eines Medikaments in Blut und Leber bei konstanter Infusion g :



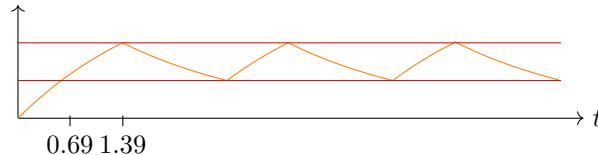
Dann sind die Entwicklungen $t \mapsto y_1(t), t \mapsto y_2(t)$ gegeben durch:

$$y_1(t) = \frac{g}{b}(1 - e^{-bt}), \quad y_2(t) = gt + \frac{g}{b}(e^{-bt} - 1)$$



Wir möchten die Menge des Medikaments im Blut innerhalb eines Intervalls $[Y_2, Y_1]$ halten. Dies gelingt durch periodischen Unterbruch und Neustart der Infusion:

1. Angenommen, wir unterbrechen die Infusion bei $t = T_1$ mit einer Menge im Blut Y_1 .
2. Von diesem Zeitpunkt baut das Blut das Medikament exponentiell ab, wie im 1. Fall.
3. Um eine Mindestmenge von Y_2 zu garantieren, starten wir die Infusion wieder zum Zeitpunkt T_2 . Dieser berechnet sich als $T_2 = T_1 + \frac{1}{b} \left(\ln \frac{Y_1}{Y_2} \right)$. (Siehe Übungen.)



4.1.3 Zusammenfassung Taylor-Reihe

Wir erinnern zunächst an die Taylor-Reihe einer (differenzierbaren) Funktion f . Damit können die Funktionswerte $f(x)$ allenfalls beschrieben werden.

Sei $f \in C^\infty$. In der Theorie der Potenzreihen untersuchen wir die Frage: Können wir f um x_0 in eine Potenzreihe entwickeln? Das heisst: Gibt es Koeffizienten a_k mit $f(x) = \sum_{k=0}^{\infty} a_k(x - x_0)^k$?

Dabei gilt es zu beachten:

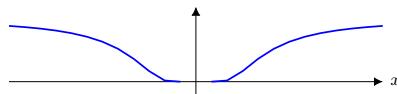
1. Falls sich f in dieser Weise darstellen lässt, gilt $a_k = \frac{f^{(k)}(x_0)}{k!}$ (Taylorkoeffizienten).

2. Wenn wir f in eine Potenzreihe entwickeln, brauchen Definitionsbereich D_f von f und Konvergenzbereich D_T der Taylorreihe nicht übereinzustimmen.

Beispiel: Es ist f mit $f(x) = \frac{1}{1+x^2} = \sum_{k=0}^{\infty} (-1)^k x^{2k}$ auf ganz \mathbb{R} definiert, konvergiert aber nur für $|x| < 1$.

3. Es gibt Funktionen f deren Taylor-Reihe in einer Umgebung von x_0 konvergiert, und trotzdem stimmt f nur im Punkt x_0 mit ihrer Taylor-Reihe überein.

Beispiel: Die Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ mit $f(x) = \begin{cases} e^{-\frac{1}{x^2}} & \text{für } x \neq 0 \\ 0 & \text{für } x = 0 \end{cases}$ hat die konstante Taylor-Reihe 0.



4. Um sicher zu sein, dass f durch ihre Taylor-Reihe dargestellt wird, muss man also den Konvergenzbereich D_T bestimmen und entscheiden, ob die Reihe auf $D_f \cap D_T$ gegen f konvergiert.

Erst dann ist $f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k$ für $x \in D_f \cap D_T$.

Fourier-Reihen helfen uns nun, **periodische** Funktionen darzustellen.

Stellen wir gegenüber:

Taylor-Reihe:

Nach Brook Taylor, 1685 – 1731



Falls $f(x) = \sum_{i=0}^{\infty} \textcolor{blue}{a}_k (\textcolor{red}{x} - x_0)^k$ so gilt

$$\textcolor{blue}{a}_k = \frac{f^{(k)}(x_0)}{k!}.$$

Dabei f muss also beliebig oft differenzierbar sein.

Fourier-Reihe:

Nach J.-B. Joseph Fourier, 1768 – 1830



Versuche $f(x)$ zu schreiben als

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

Dann muss also 2π -periodisch sein. Wir werden auch Formel für a_k, b_k finden und sehen, dass wir weiter nur wenig für f voraussetzen müssen.

4.2 Trigonometrische Polynome / Reihe

Definition 4.2.1. *Die Funktion*

$$\mathbb{R} \rightarrow \mathbb{R}$$

$$x \mapsto \frac{a_0}{2} + \sum_{k=1}^{\textcolor{red}{N}} (a_k \cos(kx) + b_k \sin(kx))$$

heisst **trigonometrisches Polynom vom Grad $\textcolor{red}{N}$** - mit Periode 2π .

Für $\textcolor{red}{N} \rightarrow \infty$ erhalten wir einen Ausdruck der Form

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(kx) + b_k \sin(kx)) \quad \text{trigonometrische Reihe.}$$

Falls die Reihe (z.B. für alle x) konvergiert, ist die Summen-Funktion periodisch mit Periode 2π .

Wir untersuchen die Fragen

1. Wie können wir ein periodisches f approximieren?
2. Welche periodischen Funktionen lassen sich durch eine trigonometrische Reihe darstellen? Für welche f und welche x ist $f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(kx) + b_k \sin(kx))$
3. Wie berechnen wir die Koeffizienten a_k und b_k ?

4.2.1 Fourier-Koeffizienten und Trigonometrische Orthogonalität

Wir widmen uns zunächst den Koeffizienten a_k und b_k .

Angenommen, wir haben eine Funktion f , für die es Zahlen a_k und b_k gibt, mit

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

Wir wenden nun folgende Formeln an, um die Koeffizienten in diesen Fall zu berechnen:

Trigonometrische Orthogonalitäten

Woher dieser Name kommt, wird sich weiter unten klären.

Mit Hilfe der Partiellen Integration gelten

$$1. \int_{-\pi}^{\pi} \cos(kx) \cos(nx) dx = \begin{cases} 2\pi & n = k = 0 \\ \pi & n = k \neq 0 \\ 0 & n \neq k \end{cases}$$

$$2. \int_{-\pi}^{\pi} \sin(kx) \sin(nx) dx = \begin{cases} 0 & n = k = 0 \\ \pi & n = k \neq 0 \\ 0 & n \neq k \end{cases}$$

$$3. \int_{-\pi}^{\pi} \sin(kx) \cos(nx) dx = 0$$

Jetzt multiplizieren wir $f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(kx) + b_k \sin(kx))$ mit $\cos(nx)$ und berechnen

$$\begin{aligned} \int_{-\pi}^{\pi} f(x) \cdot \cos(nx) dx &= \int_{-\pi}^{\pi} \frac{a_0}{2} \cdot \cos(nx) dx \\ &\quad + \sum_{k=1}^{\infty} a_k \int_{-\pi}^{\pi} \cos(kx) \cos(nx) dx + \sum_{k=1}^{\infty} b_k \int_{-\pi}^{\pi} \sin(kx) \cos(nx) dx \end{aligned}$$

Vertauschen von Summe und Integral ist hier erlaubt

$$\begin{aligned} &= \int_{-\pi}^{\pi} \frac{a_0}{2} \cdot \cos(nx) dx + \sum_{k=1}^{\infty} a_k \int_{-\pi}^{\pi} \cos(kx) \cos(nx) dx + 0, \text{ wegen 3. oben} \\ &= \frac{a_0}{2}(2\pi) + a_n \cdot \pi, \text{ wegen 1. oben} \end{aligned}$$

Also insgesamt ist $a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx) dx$ für jedes n .

Wenn wir $f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(kx) + b_k \sin(kx))$ mit $\sin(nx)$ multiplizieren, folgen Formeln für die b_k .

Wir fassen zusammen und definieren:

Definition 4.2.2. Sei f eine 2π -periodische Funktion. Die Zahlen

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

heissen **Fourier-Koeffizienten** von f . Sie definieren die **Fourier-Reihe** von f

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(kx) + b_k \sin(kx)).$$

Bleibt die zentrale Frage: Für welche f und x ist

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

Beispiel 4.2.3. Mit Trigonometrischen Identitäten/Additionstheoremen erhalten wir direkt:

$$\begin{aligned}
 \cos^2(x) &= \frac{1}{2} + \frac{1}{2} \cos(2x) \\
 \sin^3(x) &= \frac{3}{4} \sin(x) - \frac{1}{4} \sin(3x) \\
 (\cos^2 \sin^3)(x) &= \frac{1}{8} \sin(x) + \frac{1}{16} \sin(3x) - \frac{1}{16} \sin(5x) \\
 (\cos^3 \sin^2)(x) &= \frac{1}{8} \cos(x) - \frac{1}{16} \cos(3x) - \frac{1}{16} \cos(5x) \\
 &\vdots
 \end{aligned}$$

4.2.2 Was hat das mit Lineare Algebra zu tun?

Wir definieren zunächst:

1. Sei $C^0([-\pi, \pi])$ der Vektorraum der stetigen Funktionen auf dem Intervall $[-\pi, \pi]$.
2. In $C^0([-\pi, \pi])$ betrachten wir die Funktionen (als Vektoren)

$$c_0(x) = \frac{1}{\sqrt{2\pi}}, c_n(x) = \frac{1}{\sqrt{\pi}} \cos(nx), s_m(x) = \frac{1}{\sqrt{\pi}} \sin(mx), n, m \geq 1$$
3. Sei T_N der von c_n und s_m aufgespannte Untervektorraum, mit $n, m \leq N$.

Bemerkungen:

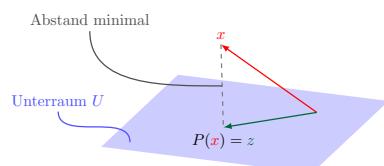
1. Statt $[-\pi, \pi]$ könnten wir auch ein anderes Intervall der Länge 2π wählen.
2. Die Koeffizienten $\frac{1}{\sqrt{2\pi}}$ und $\frac{1}{\sqrt{\pi}}$ sind eine Normierung für später.
3. Es gilt $\dim T_N = 2N + 1$.

Sei $f \in C^0([-\pi, \pi])$. Falls $f \in T_N$ ist f eine Linearkombination der Vektoren c_n und s_n . Es gibt also Zahlen α_n und β_n mit

$$f(x) = \sum_{n=0}^N \alpha_n c_n + \sum_{m=1}^N \beta_m s_m = \underbrace{\frac{\alpha_0}{\sqrt{2\pi}} + \sum_{k=1}^N \left(\frac{\alpha_k}{\sqrt{\pi}} \cos(kx) + \frac{\beta_k}{\sqrt{\pi}} \sin(kx) \right)}_{\text{ein trig. Polynom}}$$

Sei $\mathbf{f} \in C^0([-\pi, \pi])$, aber $\mathbf{f} \notin T_N$.

Welche Funktion in T_N ist die beste Approximation der Funktion \mathbf{f} ? Diese Approximation erfolgt durch die orthogonale Projektion $P(\mathbf{f})$ von \mathbf{f} auf den Untervektorraum T_N



Dabei ist die Länge von $\mathbf{f} - P(\mathbf{f})$ (der Abstand) minimal, wenn $P(\mathbf{f})$ die orthogonale Projektion von \mathbf{f} auf T_N ist. Für wachsendes N wird die Approximation immer besser.

Bemerkung:

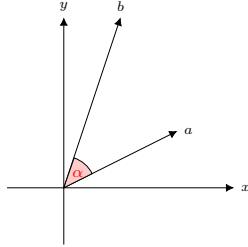
Um dies zu berechnen, müssen wir zuerst im VR $C^0([-\pi, \pi])$ Längen und (rechte) Winkel messen können. Aus den VR-Axiomen lassen sich diese Strukturen nicht ableiten, es handelt sich also um einen Zusatz.

4.3 Euklidische Vektorräume

4.3.1 Definition und Beispiele Euklidischer VR

Mit dem Standardskalarprodukt können wir im \mathbb{R}^n Längen und Winkel angeben. Wir erinnern uns:

- Das bekannte Standardskalarprodukt für $a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \in \mathbb{R}^2$ ist gegeben durch $a \cdot b = \langle a, b \rangle = a_1 b_1 + a_2 b_2 = |a||b| \cdot \cos(\alpha)$ mit $|a| = \sqrt{\langle a, a \rangle}$, und es sind a und b senkrecht $\iff \langle a, b \rangle = 0$.



- Allgemein für \mathbb{R}^n ist ein Skalarprodukt gegeben durch $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$.

Jetzt verallgemeinern wir die Situation und analysieren, welche Eigenschaft dieses bekannte Skalarprodukt hat. Das führt zur Definition:

Definition 4.3.1. Sei V ein reeller VR. Ein **Skalarprodukt** ordnet zwei Vektoren $x, y \in V$ eine Zahl zu $(x, y) \mapsto \langle x, y \rangle \in \mathbb{R}$ und erfüllt dabei folgende Regeln:

Symmetrisch Für alle $x, y \in V$ gilt $\langle x, y \rangle = \langle y, x \rangle$.

Bilinear im zweiten Argument

- Für alle $x, y, z \in V$ gilt $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$.
- Für alle $x, y \in V, \alpha \in \mathbb{R}$ gilt $\langle x, \alpha y \rangle = \alpha \langle x, y \rangle$.

Wegen der Symmetrie gilt diese Eigenschaft auch im ersten Argument.

Positiv definit

- Für alle $x \in V$: $\langle x, x \rangle \geq 0$.
- $\langle x, x \rangle = 0 \iff x = 0$.

Ein reeller Vektorraum V zusammen mit einem Skalarprodukt $\langle \cdot, \cdot \rangle$ heißt **euklidischer Vektorraum** $(V, \langle \cdot, \cdot \rangle)$.

Beispiele

- Der VR \mathbb{R}^n wird mit dem Standardskalarprodukt $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ für $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ und $y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ zu einem euklidischen VR.
- Auf $C^0([a, b])$ ist ein Skalarprodukt gegeben durch $\langle f, g \rangle = \int_a^b f(x)g(x)dx$, für zwei stetige Funktionen $f, g : [a, b] \rightarrow \mathbb{R}$.

4.3.2 Orthogonale Vektoren

Definition 4.3.2. Zwei Vektoren x, y in einem euklidischen Vektorraum heißen **orthogonal**, falls $\langle x, y \rangle = 0$. Die Notation ist in diesem Fall $x \perp y$.

Beispiele

- Bezüglich des Standardskalarprodukts in \mathbb{R}^2 gilt $\begin{pmatrix} a \\ b \end{pmatrix} \perp \begin{pmatrix} b \\ -a \end{pmatrix}$.
- In $C^0([- \pi, \pi])$ mit Skalarprodukt $\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)g(x)dx$ betrachten wir $f_n(x) = \cos(nx)$ mit $n = 0, 1, 2, 3, \dots$. Dann gilt $f_n \perp f_m$, falls $m \neq n$ (Übung).

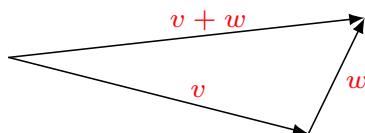
4.3.3 Normierte Vektorräume

Definition 4.3.3. Sei V ein VR. Eine **Norm** auf V ist eine Abbildung

$$\| \cdot \| : V \rightarrow \mathbb{R}, \quad v \mapsto \|v\|$$

mit den Eigenschaften

- Für alle $v \in V$: $\|v\| \geq 0$.
- $\|v\| = 0 \iff v = 0$.
- Für alle $v \in V$ und alle $\alpha \in \mathbb{R}$: $\|\alpha v\| = |\alpha| \|v\|$.
- Für alle $v, w \in V$ gilt die **Dreiecksungleichung**: $\|v + w\| \leq \|v\| + \|w\|$



Ein VR V mit einer Norm $\|\cdot\|$ heisst normierter VR $(V, \|\cdot\|)$. Ein Vektor v mit $\|v\| = 1$ heisst **Einheitsvektor**.

Beispiele

- Auf \mathbb{R}^n haben wir die euklidische Norm $\|x\| = \sqrt{\langle x, x \rangle} = \sqrt{\sum_{i=1}^n x_i^2}$.

- Es gibt auf \mathbb{R}^n auch weitere Normen, zum Beispiel

die **Maximumsnorm** $\|x\|_\infty = \max\{|x_1|, \dots, |x_n|\}$

die **p -Norm** $\|x\|_p = (|x_1|^p + \dots + |x_n|^p)^{\frac{1}{p}}$ mit $1 \leq p < \infty$.

- Beispiele für Normen auf $M_{m \times n}$. Sei A eine $m \times n$ -Matrix:

$$\text{Hilbert-Schmidt-Norm } \|A\|_2 = \sqrt{\sum_{i,j} |a_{ij}|^2}$$

$$\text{Spaltenmaximumsnorm } \|A\|_{\text{SM}} = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|$$

$$\text{Zeilenmaximumsnorm } \|A\|_{\text{ZM}} = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$$

Definition 4.3.4. Sei V mit dem Skalarprodukt $\langle \cdot, \cdot \rangle$ ein euklidischer VR.

Dann ist durch $\|x\| := \sqrt{\langle x, x \rangle}$ für $x \in V$ eine Norm auf V definiert. Sie heisst die vom Skalarprodukt **induzierte Norm**.

- Auf \mathbb{R}^n induziert das Standardskalarprodukt $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ die euklidische Norm.
- Das Skalarprodukt auf $C^0([a, b])$ gegeben durch $\langle f, g \rangle = \int_a^b f(x)g(x)dx$ induziert die Norm

$$\|f\|_{L^2} = \sqrt{\int_a^b f^2(x)dx}.$$

Bemerkungen

1. Eine Norm $\|\cdot\|$ wird genau dann von einem Skalarprodukt induziert, wenn die **Parallelogrammregel** gilt:

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2)$$

Das Skalarprodukt erhalten wir in diesem Fall aus der Norm zurück durch

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

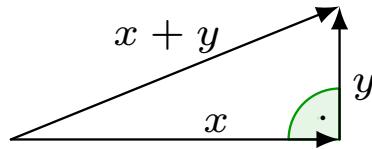
2. In einem euklidischen VR mit induzierter Norm $x \mapsto \|x\| = \sqrt{\langle x, x \rangle}$ gelten

- Wir rechnen

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

Also gilt der **Satz von Pythagoras**:

$$\langle x, y \rangle = \iff x \perp y \implies \|x + y\|^2 = \|x\|^2 + \|y\|^2,$$



- Für $\lambda \in \mathbb{R}$ folgt in der blauen Formel mit λy anstelle von y :

$$0 \leq \|x + \lambda y, x + \lambda y\|^2 = \|x^2\| + 2\lambda\langle x, y \rangle + \lambda^2\|y\|^2$$

Wählen wir für $\lambda = -\frac{\langle x, y \rangle}{\|y\|^2}$ ergibt sich daraus die

Ungleichung von Cauchy-Schwarz: $|\langle x, y \rangle| \leq \|x\|\|y\|$

- Aus der Ungleichung von Cauchy-Schwarz folgt dann

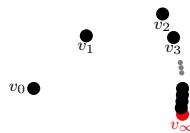
$$\begin{aligned}\|x + y\|^2 &= \|x\|^2 + 2\langle x, y \rangle + \|y\|^2 \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \\ &= (\|x\| + \|y\|)^2,\end{aligned}$$

also die Dreiecksungleichung.

Abstand und Konvergenz

Sei $(V, \|\cdot\|)$ ein normierter VR. Dann ist $\|x - y\|$ ein Mass für den Abstand der Vektoren x und y . Jede Norm liefert daher einen Konvergenzbegriff:

Definition 4.3.5. Sei (v_n) eine Folge im normierten VR V . Dann konvergiert die Folge gegen $v \in V$, falls $\lim_{n \rightarrow \infty} \|v_n - v\| = 0$, mit Schreibweise $\lim_{n \rightarrow \infty} v_n = v$.



Für einen endlichdimensionalen VR V mit zwei Normen $\|\cdot\|_a$ und $\|\cdot\|_b$ gilt:

$$\lim_{n \rightarrow \infty} \|v_n - v\|_a = 0 \iff \lim_{n \rightarrow \infty} \|v_n - v\|_b = 0.$$

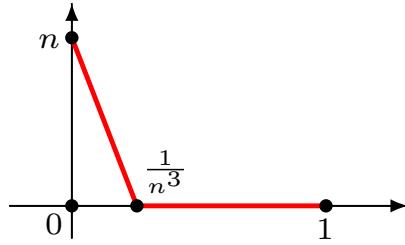
CAVE

In endlichdimensionalen Räumen spielt also für die Konvergenz die Wahl der Norm keine Rolle. Dies ist nicht mehr so in unendlichdimensionalen Räumen!

Dies illustriert folgendes Beispiel

Beispiel / Übung. Die Folge $(f_n)_{n \in \mathbb{N}} \in C^0([0, 1])$ mit

$$f_n(x) = \begin{cases} n - n^4 x & \text{falls } x \in [0, \frac{1}{n^3}], \\ 0 & \text{sonst} \end{cases}$$



konvergiert bezüglich der Norm $\|f\|_{L^2} = \sqrt{\int_0^1 f^2(x) dx}$ gegen die Funktion $f(x) \equiv 0$, obwohl $f_n(0) \rightarrow \infty$ für $n \rightarrow \infty$, also keine punktweise Konvergenz!

4.3.4 Orthonormalbasen

Seien e_1, e_2, \dots, e_k Einheitsvektoren, welche paarweise orthogonal sind. Das bedeutet:

1. $\langle e_i, e_j \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$
2. e_1, e_2, \dots, e_k sind linear unabhängig. Insbesondere: Falls $k = \dim V$, so bilden e_1, e_2, \dots, e_k eine Basis.

Definition 4.3.6. Eine Basis aus paarweise orthogonalen Einheitsvektoren (in einem endlichdimensionalen VR) heisst **Orthonormalbasis (ONB)**.

Bemerkungen

1. Jeder endlichdimensionale euklidische VR hat eine ONB.
2. Sei $\{e_1, e_2, \dots, e_n\}$ eine ONB des VR V . Für jedes $v \in V$ gilt dann:

$$v = \sum_{i=1}^n \langle v, e_i \rangle e_i$$

Der Koeffizient $\langle v, e_i \rangle$ ist die i -te Koordinate von v bezüglich der Basis e_1, e_2, \dots, e_n , denn mit $v = \sum_{i=1}^n a_i e_i$ ist $\langle v, e_k \rangle = \left\langle \sum_{i=1}^n a_i e_i, e_k \right\rangle = \sum_{i=1}^n a_i \langle e_i, e_k \rangle = a_k$.

Beispiel 4.3.7. Für die Funktionen

$$c_0(x) = \frac{1}{\sqrt{2\pi}}, c_n(x) = \frac{1}{\sqrt{\pi}} \cos(nx), \quad s_m(x) = \frac{1}{\sqrt{\pi}} \sin(mx), \quad n, m \geq 1$$

in $C^0([-\pi, \pi])$ gilt mit den Trigonometrischen Orthogonalitäten oben

1. $\langle c_n, c_k \rangle = 1$ und $\langle c_n, c_k \rangle = 0$, für $n \neq k$;
2. $\langle s_n, s_n \rangle = 1$ und $\langle s_n, s_k \rangle = 0$, für $n \neq k$;
3. $\langle c_k, s_n \rangle = 0$

Somit bilden diese Funktionen eine ONB im von ihnen aufgespannten VR.

4.3.5 Anwendung auf Fourier-Koeffizienten

Wir möchten eine 2π -periodische Funktion f durch ein trigonometrisches Polynom

$$x \mapsto f(x) = \frac{a_0}{2} + \sum_{k=1}^N (a_k \cos(kx) + b_k \sin(kx))$$

approximieren, und suchen dafür a_k und b_k .

Sei T_N der von c_n und s_m aufgespannte Untervektorraum, mit $n, m \leq N$. Da $c_0, c_1, \dots, c_N, s_1, \dots, s_N$ eine ONB bilden, gilt für jede Funktion $f \in T_N$

$$\begin{aligned} f(x) &= \sum_{n=0}^N \langle f, c_n \rangle c_n + \sum_{m=1}^N \langle f, s_m \rangle s_m \\ &= \underbrace{\langle f, c_0 \rangle}_{\text{gesuchtes } \frac{a_0}{2}} \frac{1}{\sqrt{2\pi}} + \underbrace{\sum_{n=1}^N \langle f, c_n \rangle}_{\text{gesuchtes } a_n} \frac{1}{\sqrt{\pi}} \cos(nx) + \underbrace{\sum_{m=1}^N \langle f, s_m \rangle}_{\text{gesuchtes } b_m} \frac{1}{\sqrt{\pi}} \sin(mx) \end{aligned}$$

Es ist für

$n = 0$:

$$\frac{a_0}{2} = \langle f, c_0 \rangle \frac{1}{\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(x) \frac{1}{\sqrt{2\pi}} dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx.$$

Also

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) dx.$$

$n \geq 1$:

$$a_n = \langle f, c_n \rangle \frac{1}{\sqrt{\pi}} = \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} f(x) \cos(nx) dx = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx) dx.$$

$m \geq 1$:

$$b_m = \langle f, s_m \rangle \frac{1}{\sqrt{\pi}} = \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} f(x) \sin(mx) dx = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(mx) dx.$$

Wir erhalten die bekannten Formeln aus 4.2.1 für die Fourier-Koeffizienten a_n und b_n .

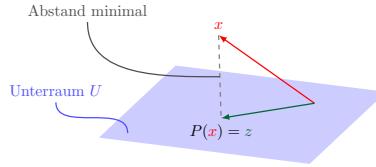
4.4 Orthogonale Projektion

Bleibt die Fragen, was mit $f \notin T_N$ ist? Welche Funktion in T_N ist die beste Approximation der Funktion $\textcolor{red}{f}$? Wie oben angekündigt erfolgt diese Approximation durch die orthogonale Projektion $P(\textcolor{teal}{f})$ von f auf den Untervektorraum T_N

4.4.1 Konstruktion

Sei V ein euklidischer VR mit Skalarprodukt $\langle \cdot, \cdot \rangle$. Dann gilt

1. Die Orthogonalprojektion eines Vektors x auf einen Vektor $y \neq 0$ ist $P(x) = z = \frac{\langle x, y \rangle}{\langle y, y \rangle} y$.
2. Sei U ein UR von V . Seien e_1, \dots, e_n eine ONB von U . Dann ist $P(\textcolor{teal}{x}) = \textcolor{teal}{z} = \sum_{i=1}^n \langle x, e_i \rangle e_i$ die Orthogonalprojektion von $\textcolor{red}{x} \in V$ auf U .



3. Für alle $v \in U$ gilt $\textcolor{cyan}{x} - \textcolor{teal}{z} \perp v$. Wir zeigen, dass $\textcolor{cyan}{x} - \textcolor{teal}{z}$ tatsächlich senkrecht auf jedem Basisvektor e_j von U steht:

$$\begin{aligned}\langle \textcolor{cyan}{x} - \textcolor{teal}{z}, e_j \rangle &= \langle x, e_j \rangle - \langle \textcolor{teal}{z}, e_j \rangle = \langle x, e_j \rangle - \left\langle \sum_{k=1}^n \langle x, e_k \rangle e_k, e_j \right\rangle \\ &= \langle x, e_j \rangle - \sum_{k=1}^n \langle x, e_k \rangle \langle e_k, e_j \rangle = \langle x, e_j \rangle - \langle x, e_j \rangle = 0\end{aligned}$$

4. Die Projektion $P(\textcolor{teal}{x}) = \textcolor{teal}{z}$ hat den kleinsten Abstand eines Punktes in U von $\textcolor{red}{x}$:

Pythagoras: $\|\textcolor{red}{x} - P(\textcolor{teal}{x})\| \leq \|\textcolor{red}{x} - u\|$ für alle $u \in U$.

4.4.2 Anwendung auf Fourier-Reihen

Sei nun $V = C^0([-\pi, \pi])$ mit dem Skalarprodukt $\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)g(x)dx$. Wir betrachten den UR $U = T_N$ mit der ONB $c_0, c_1, \dots, c_N, s_1, \dots, s_N$.

1. Die Projektion $P_N(f) \in T_N$ von $\textcolor{red}{f} \in V$ auf T_N ist das Fourier-Polynom vom Grad N

$$\begin{aligned}P_N(f) &= \langle f, c_0 \rangle c_0 + \sum_{k=1}^N \langle f, c_k \rangle c_k + \langle f, s_k \rangle s_k \\ &= \frac{a_0}{2} + \sum_{k=1}^N a_k \cos(kx) + b_k \sin(kx)\end{aligned}$$

mit den zuvor berechneten **Fourier-Koeffizienten**

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

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

2. Für $N \rightarrow \infty$ wird die Approximation vom $P_N(f)$ an f immer besser, und wir erhalten die Fourier-Reihe f . Genauer:

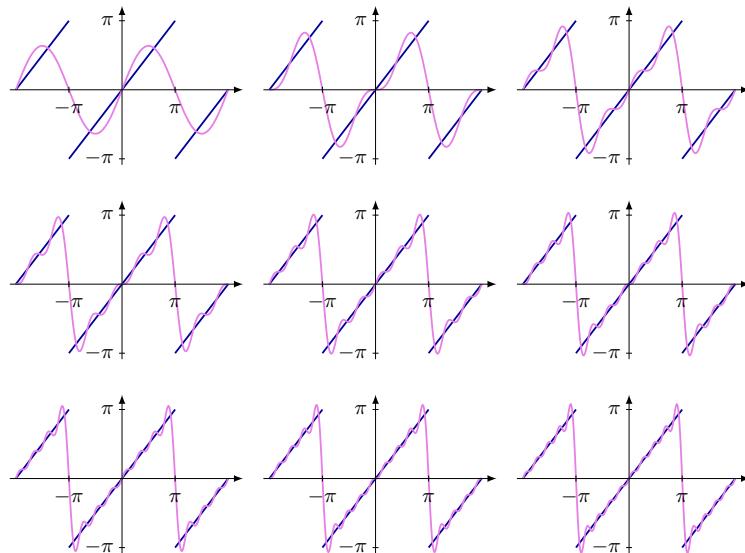
$$\lim_{N \rightarrow \infty} \|f - P_N(f)\|_{L^2} = \lim_{N \rightarrow \infty} \int_{-\pi}^{\pi} (f(x) - P_N(f)(x))^2 dx = 0.$$

Die Beziehung besagt, dass $P_N(f)$ den Limes f im quadratischen Mittel approximiert.

4.4.3 Anwendungen

Gibbs-Tower

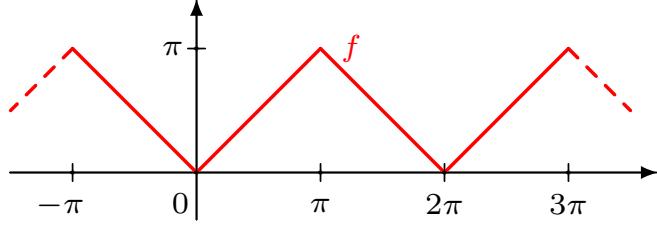
An den Sprungstellen zeigt die Konvergenz das sogenannte Gibbs-Phänomen. Schauen wir uns dies an der 2π -periodischen Fortsetzung der Funktion f mit $f :]-\pi, \pi[\rightarrow \mathbb{R}, x \mapsto x$ an:



Der kleine „Turm“, der sich links und rechts der Sprungstelle bildet heisst **Gibbs-Tower**. Er wird mit wachsendem N nicht kleiner, nur schmäler. Hat $f \in C^n(\mathbb{R})$ die Periode T , so konvergieren die Koeffizienten c_k für $k \rightarrow \infty$ der T -periodischen Fourier-Reihe umso schneller nach Null, je grösser n ist.

Lösen einer DGL mit Fourier

Fourier Reihen helfen uns beim Lösen einer inhomogenen linearen DGL: Suchen wir zum Beispiel die 2π -periodische Lösung der DGL $y'(x) + y(x) = f(x)$ für folgende Funktion f :



Wählen wir als **Ansatz**: $y(x) = \frac{\alpha_0}{2} + \sum_{k=1}^{\infty} \alpha_k \cos(kx) + \beta_k \sin(kx)$. Durch Ableiten auf beiden Seiten folgt $y'(x) = \sum_{k=1}^{\infty} -k\alpha_k \sin(kx) + k\beta_k \cos(kx)$. Damit rechnen wir:

$$\begin{aligned} y'(x) + y(x) &= \frac{\alpha_0}{2} + \sum_{k=1}^{\infty} (\alpha_k + k\beta_k) \cos(kx) + (\beta_k - k\alpha_k) \sin(kx) \\ &\stackrel{!}{=} f(x) \\ &= \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos(kx) + b_k \sin(kx) \end{aligned}$$

Sobald die Fourier-Koeffizienten a_k, b_k von f berechnet sind, können wir die gesuchten Koeffizienten α_k, β_k von y durch Koeffizientenvergleich bestimmen. Da f eine gerade Funktion ist, sind alle $b_k = 0$. Für a_k gilt

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(kx) dx = \frac{2}{\pi} \int_0^{\pi} x \cos(kx) dx.$$

Für $k = 0$ ist also $a_0 = \pi$. Für $k > 0$ verwenden wir partielle Integration:

$$\begin{aligned} a_k &= \frac{2}{\pi} \int_0^{\pi} x \cos(kx) dx = \frac{2}{\pi} \left(x \underbrace{\frac{\sin(kx)}{k}}_0 \Big|_0^\pi - \int_0^{\pi} \frac{\sin(kx)}{k} dx \right) \\ &= \frac{2}{\pi} \frac{\cos(kx)}{k^2} \Big|_0^\pi = \frac{2}{k^2 \pi} (\cos(k\pi) - 1) = \frac{2}{k^2 \pi} ((-1)^k - 1) = \\ &= \begin{cases} 0 & \text{falls } k \text{ gerade} \\ \frac{-4}{k^2 \pi} & \text{falls } k \text{ ungerade} \end{cases}. \end{aligned}$$

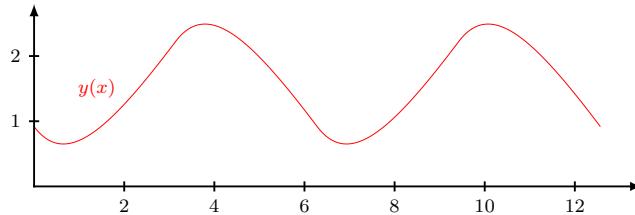
Also gilt $\alpha_0 = a_0 = \pi$, $\beta_k - k\alpha_k = b_k = 0$, also $\beta_k = k\alpha_k$, und somit $a_k = \alpha_k + k\beta_k = \alpha_k + k^2\alpha_k = \alpha_k(1 + k^2)$. Also

$$\alpha_k = \begin{cases} 0 & k \text{ gerade} \\ \frac{-4}{k^2(1+k^2)\pi} & k \text{ ungerade} \end{cases}, \quad \beta_k = \begin{cases} 0 & k \text{ gerade} \\ \frac{-4}{k(1+k^2)\pi} & k \text{ ungerade} \end{cases}.$$

Wenn wir die ungeraden Zahlen mit $k = 2n + 1$ durchlaufen, erhalten wir schliesslich

$$y(x) = \frac{\pi}{2} - \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{\cos((2n+1)x)}{(2n+1)^2(1+(2n+1)^2)} + \frac{\sin((2n+1)x)}{(2n+1)(1+(2n+1)^2)}.$$

Die Reihe konvergiert wegen der dritten Potenz von n im Nenner sehr rasch. Für einen Plot der Lösung genügen daher schon wenige Terme der Reihe:



4.5 Ergänzungen und Beispiele

Es bleibt die Frage, für welche x gilt: $f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(kx) + b_k \sin(kx))$?

4.5.1 Satz von Dirichlet

Definition 4.5.1. Eine Funktion $f : [-\pi, \pi] \rightarrow \mathbb{R}$ heisst **stückweise stetig differenzierbar**, wenn f mit Ausnahme endlich vieler Punkte stetig differenzierbar ist, und falls in jedem Ausnahmepunkt z und in den Randpunkten die einseitigen Grenzwerte

$$\underbrace{\lim_{\substack{x \rightarrow z \\ x < z}} f(x)}_{=f(z^-)} \quad \underbrace{\lim_{\substack{x \rightarrow z \\ x > z}} f(x)}_{=f(z^+)} \quad \lim_{\substack{x \rightarrow z \\ x < z}} f'(x) \quad \lim_{\substack{x \rightarrow z \\ x < z}} f'(x) \quad \text{in } \mathbb{R} \text{ existieren.}$$

Theorem 4.5.2 (Dirichlet). Sei $f : [-\pi, \pi] \rightarrow \mathbb{R}$ stückweise stetig differenzierbar. Dann gilt

1. Die Fourier-Reihe von f konvergiert in allen inneren Stetigkeitspunkten x von f gegen $f(x)$.
2. In einer Sprungstelle z ist der Grenzwert der Reihe $\frac{f(z^+) + f(z^-)}{2}$.

Bemerkung:

Da die Fourier-Reihe 2π -periodisch ist, gilt der Satz von Dirichlet auf ganz \mathbb{R} , wenn f die Periode 2π hat.

4.5.2 Fourier-Reihe bei allgemeiner Periode T

Sei $f : \left[-\frac{T}{2}, \frac{T}{2}\right] \rightarrow \mathbb{R}$ stückweise stetig differenzierbar. Dann ist $g(x) = f\left(\frac{Tx}{2\pi}\right)$ auf $[-\pi, \pi]$ definiert und hat die 2π periodische Fourier-Reihe $g(x) = \frac{a_0}{2} + \sum_{k=0}^{\infty} a_k \cos(kx) + b_k \sin(kx)$ für alle Stetigkeitspunkte x .

Aus $g(x) = f\left(\frac{Tx}{2\pi}\right) \implies f(x) = g\left(\frac{2\pi}{T}x\right)$ erhalten wir so die T -periodische Fourier-Reihe von f . In den Stetigkeitspunkten gilt

$$f(x) = \frac{a_0}{2} + \sum_{k=0}^{\infty} a_k \cos\left(k\frac{2\pi x}{T}\right) + b_k \sin\left(k\frac{2\pi x}{T}\right)$$

und mit der Substitution $z = \frac{2\pi x}{T}$ bekommen wir

$$\begin{aligned} a_k &= \frac{1}{\pi} \int_{-\pi}^{\pi} g(z) \cos(kz) dz = \frac{1}{\pi} \int_{-T/2}^{T/2} g\left(\frac{2\pi z}{T}\right) \cos\left(k\frac{2\pi z}{T}\right) \frac{2\pi}{T} dx \\ &= \frac{2}{T} \int_{-T/2}^{T/2} f(x) \cos\left(k\frac{2\pi x}{T}\right) dx \quad \text{für } k = 0, 1, \dots \text{ und analog} \\ b_k &= \frac{2}{T} \int_{-T/2}^{T/2} f(x) \sin\left(k\frac{2\pi x}{T}\right) dx \quad \text{für } k = 1, 2, \dots \end{aligned}$$

4.5.3 Komplexe Schreibweise

Führen wir noch die komplexe Schreibweise ein, mit der es sich oft einfacher rechnen lässt. Sei

$$f : \mathbb{R} \rightarrow \mathbb{C}, \quad t \mapsto \underbrace{f(t) = g(t) + ih(t)}_{\text{Annahme: immer möglich}} \in \mathbb{C}, \quad i^2 = -1$$

Dann ist

$$\int_a^b f(t) dt = \int_a^b g(t) dt + i \int_a^b h(t) dt.$$

Die Formel von Euler

Wegen der Taylor-Reihe gilt:

$$\begin{aligned} \cos(x) &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} \dots \\ i \sin(x) &= ix - \frac{ix^3}{1!} + \frac{ix^5}{3!} - \frac{ix^7}{5!} + \dots \\ \hline \cos(x) + i \sin(x) &= 1 + ix - \frac{x^2}{2!} + \frac{ix^3}{3!} + \frac{x^4}{4!} - \frac{ix^5}{5!} - \frac{x^6}{6!} \dots = e^{ix} \end{aligned}$$

Aus (I)

$$e^{ix} = \cos(x) + i \sin(x)$$

und (II)

$$e^{-ix} = \cos(x) - i \sin(x)$$

folgt durch Addition/Subtraktion (I) ± (II)

$$\frac{e^{ix} + e^{-ix}}{2} = \cos(x) \quad \frac{e^{ix} - e^{-ix}}{2i} = \sin(x)$$

Komplexe vs. Trigonometrische Koeffizienten

1. Wir rechnen mit der Eulerschen Formel um:

$$\begin{aligned} f(x) &= \frac{a_0}{2} + \sum_{k=0}^{\infty} a_k \cos(kx) + b_k \sin(kx) \\ &= \frac{a_0}{2} + \sum_{k=0}^{\infty} \frac{a_k}{2} (e^{ikx} + e^{-ikx}) + \frac{b_k}{2i} (e^{ikx} - e^{-ikx}) \\ &= \frac{1}{2} \left(a_0 + \sum_{k=1}^{\infty} (a_k - ib_k) e^{ikx} + \sum_{k=1}^{\infty} (a_k + ib_k) e^{-ikx} \right) \\ &= \sum_{k=-\infty}^{\infty} c_k e^{ikx} \end{aligned}$$

2. Wir können die komplexen Koeffizienten c_k in die trigonometrischen a_k, b_k umrechnen und umgekehrt. Mit

$$c_k = \begin{cases} \frac{1}{2}(a_k - ib_k) & \text{für } k > 0 \\ \frac{a_0}{2} & \text{für } k = 0 \\ \frac{1}{2}(a_{-k} + ib_{-k}) & \text{für } k < 0 \end{cases} \quad \text{resp.} \quad \begin{cases} a_k = c_k + c_{-k} \\ a_0 = 2c_0 \\ b_k = i(c_k - c_{-k}) \end{cases}$$

3. Tatsächlich (siehe Beispiel unten) lassen sich die komplexen Fourier-Koeffizienten c_k einfacher direkt ausrechnen, nämlich:

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} dx$$

In der Tat gilt z.B. für $k > 0$ mit Hilfe der Eulerschen Formel:

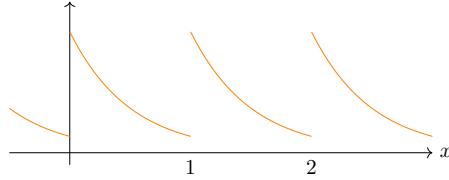
$$\begin{aligned} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} dx &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) (\cos(kx) - i \sin(kx)) dx \\ &= \frac{1}{2} \left(\frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(kx) dx - i \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(kx) dx \right) \\ &= \frac{1}{2} (a_k - ib_k) \end{aligned}$$

Zusammengefasst haben wir damit

Theorem 4.5.3. Sei $f :]-\frac{T}{2}, \frac{T}{2}[\rightarrow \mathbb{C}$. Dann ist die komplexe Fourier-Reihe von f

$$\sum_{k=-\infty}^{\infty} c_k e^{k \frac{2\pi i}{T} t} \quad \text{mit} \quad c_k = \frac{1}{T} \int_0^T f(t) e^{-k \frac{2\pi i}{T} t} dt.$$

Beispiel 4.5.4. Sei $f(x) = e^{-x}$ für $0 \leq x < 1$ periodisch fortgesetzt, mit Periode 1. Dann hat f eine 1-periodische Fourier-Reihe, die f in allen Stetigkeitspunkten darstellt.



Dabei ist der komplexe Koeffizient c_k

$$\begin{aligned} c_k &= \frac{1}{T} \int_0^T \exp(-x) \exp\left(-k \frac{2\pi i}{T} x\right) dx = \int_0^1 \exp(-(1+2k\pi i)x) dx \\ &= -\frac{1}{1+2k\pi i} \exp(-(1+2k\pi i)x) \Big|_0^1 = -\frac{1}{1+2k\pi i} (\exp(-1-2k\pi i) - 1) \\ &= \frac{1 - \frac{1}{e}}{1 + 2k\pi i}. \end{aligned}$$

Die trigonometrischen a_k, b_k erhalten wir zum Beispiel durch

$$a_k = c_k + c_{-k} = \frac{1 - \frac{1}{e}}{1 + 2k\pi i} + \frac{1 - \frac{1}{e}}{1 - 2k\pi i} = \frac{2 - 2\frac{1}{e}}{1 + 4k^2\pi^2}.$$

Übung.

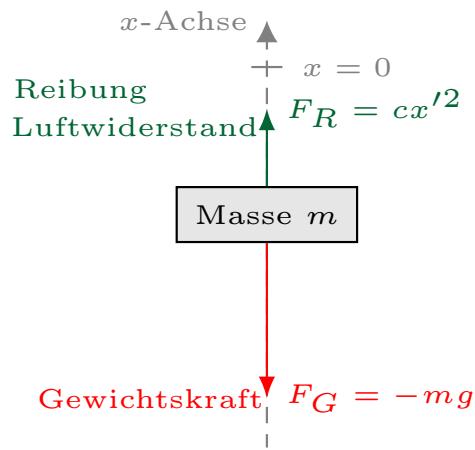
1. Berechne a_k auch mit der Integralformel.
2. Berechne b_k auf diese beiden Weisen.

KAPITEL 5

Nichtlineare Modelle

Viele, die meisten, Phänomene sind nicht linear. Beispiele ergeben sich direkt, wenn wir die vereinfachten Annahmen bei den linearen Modellen ignorieren:

1. Der „freie Fall“ wird im einfachsten Modell durch die lineare DGL $mx'' = -mg$ beschrieben. Dabei ist $x(t)$ die Position der Masse m zur Zeit t . Das lineare Modell lässt die (nicht lineare) **Reibung** ausser Acht, und ebenso die Tatsache, dass in Wahrheit g nicht linear von x abhängt.



2. Exponentielles Wachstum (etwa von Algen) wird linear beschrieben durch $\dot{y}(t) = \alpha y(t)$. Die Lösung $y(t) = y_0 e^{\alpha t}$ ist für grosse t unrealistisch. Nicht lineare Modelle (mit begrenztem Wachstum) sind besser, zum Beispiel gegeben durch die Logistische DGL $y' = y(1 - y)$.
3. Die Eulersche Knicklast ist ein rein nicht lineares Phänomen.
4. Der Henry-Koeffizient ist nicht wirklich konstant, etc.

Lineare Modelle sind daher meist jeweils eine Approximation und nicht-lineare Gleichungen sind meist schwerer zu lösen als lineare: Explizite Lösungen gibt es nicht oder lassen sich nicht angeben.

Frage: Wie analysieren wir also nicht lineare Probleme?

Numerische Antworten - Cave!

Numerische Resultate sind fehlerbehaftet. Daher sind kaum gesicherte Aussagen möglich.

Qualitative Antworten

Qualitative Aussagen durch mathematische Analyse und entsprechende Methoden liefern Aussagen etwa über das Langzeit- oder Globalverhalten:

asymptotisches Verhalten für $t \rightarrow \infty$, Periodizität von Lösungen etc.

5.1 Eindimensionale Modelle

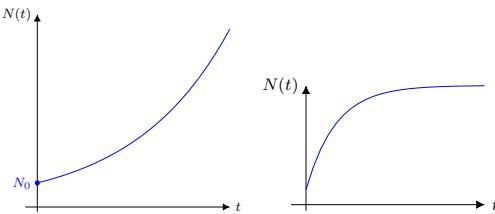
5.1.1 Logistische Gleichung

Wir hatten die beiden linearen Modelle

Exponentiell $N'(t) = r_1 \cdot N(t)$. Die Entwicklung ist exponentiell wachsend.

Beschränkt $N'(t) = r_2 \cdot (K - N(t))$ (oder $N'(t) = r_2 \cdot \frac{K - N(t)}{K}$)

Diese Entwicklung ist **zunächst linear** und wird dann gebremst.



Die Experimente zeigen, dass wir **zunächst ein exponentielles Wachstum** erwarten (genug Platz, Nahrung etc.), die Entwicklung aber dann gebremst wird.

Wir kombinieren die rechte Seite der beiden ersten Entwicklungen: $r_1 \cdot N(t)$ und $r_2 \cdot (K - N(t))$,

$$N'(t) = r_3 \cdot N(t) \cdot (K - N(t)).$$

Wir haben (mit der Konstanten $r_3 = r_1 \cdot r_2$)

$$N'(t) = r_3 \cdot N(t) \cdot (K - N(t)) = r_3 K \cdot N(t) \cdot \left(1 - \frac{N(t)}{K}\right)$$

und damit (mit der Konstanten $r = r_3 \cdot K$)

$$N'(t) = r \cdot N(t) \cdot \left(1 - \frac{N(t)}{K}\right)$$

mit $0 < N_0 = N(0) < K$ und $r > 0$. Dies ist ein Beispiel einer nicht-linearen DGL.

Ohne die Lösung explizit zu kennen, können wir qualitativ das Verhalten beschreiben:

1. Es ist $0 < N(0) < K$ und somit $0 < \frac{N(0)}{K} < 1$ und $1 - \frac{N(0)}{K} > 0$. Mit $r > 0$ sind $N'(0) = rN(0) \left(1 - \frac{N(0)}{K}\right) > 0$ und $N'(t) = rN(t) \left(1 - \frac{N(t)}{K}\right) > 0$ solange $N(t) < K$.

Die Population N ist streng monoton wachsend, solange $N(t) < K$.

2. In der Regel ist zu Beginn $N(t)$ viel kleiner als K und damit $\frac{N(t)}{K} \approx 0$ und $1 - \frac{N(t)}{K} \approx 1$, also

$$N'(t) \approx rN(t) \implies \text{exponentielles Wachstum}.$$

3. Mit $N(t) \rightarrow K$ folgt $\frac{N(t)}{K} \rightarrow 1$ und $1 - \frac{N(t)}{K} \rightarrow 0$. Wir erhalten immer langsameres Wachstum, da die rechte Seite der DGL $N'(t) = rN(t) \left(1 - \frac{N(t)}{K}\right)$ immer kleiner wird. Die Wachstumsgeschwindigkeit N' wird immer kleiner.

4. Falls $N(t_\infty) = K$, folgt $\frac{N(t_\infty)}{K} = 1$ sowie $N'(t_\infty) = rK \left(1 - \frac{K}{K}\right) = 0$ und es gibt kein Wachstum mehr. Damit ist K eine obere Schranke (maximale Kapazität).

Die Entwicklung wächst im Laufe der Zeit und nähert sich immer weiter der Asymptote $N(t_\infty) = K$ an.

5. Wie verändert sich allenfalls die Wachstumsgeschwindigkeit N' ? Wo liegen allfällige Wendepunkte? Wir schreiben kurz $N = N(t)$ und berechnen N'' :

$$N'' = (N')' = \left(rN \left(1 - \frac{N}{K}\right)\right)' = \left(rN - \frac{r}{K}N^2\right)' = rN' - \frac{r}{K}(N^2)'.$$

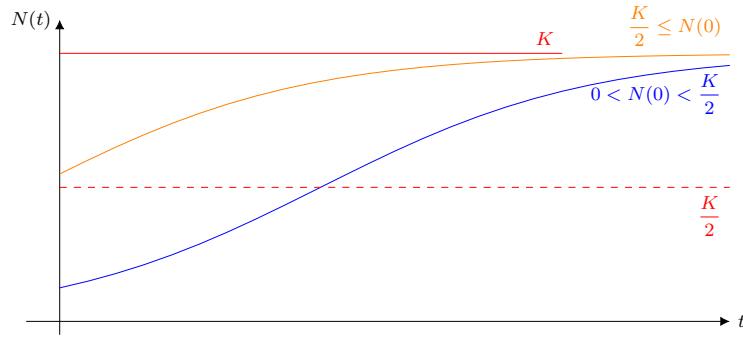
Mit Produktregel ist $(N^2)' = 2N \cdot N'$. Damit folgt

$$N'' = rN' - \frac{r}{K}(N^2)' = rN' - \frac{2r}{K}N \cdot N' = rN' \left(1 - \frac{2}{K}N\right).$$

Da $rN' > 0$ suchen wir Nullstellen von $1 - \frac{2}{K}N$. Dies ist $N = \frac{K}{2}$.

Sei t_w der Zeitpunkt, an dem $N(t_w) = \frac{K}{2}$ gilt. Dann hat die 2. Ableitung der Lösungskurve eine Nullstelle. Sei zu Beginn $N(0) = N_0 < \frac{K}{2}$. Dann ist $\frac{2N(0)}{K} < 1$ und $1 - \frac{2N(0)}{K} > 0$. Da $rN' > 0$, ist zunächst $N'' > 0$, und das Vorzeichen der 2. Ableitung wechselt zur Zeit t_w von + nach -. Also hat N dort einen Wendepunkt. Der Graph von N geht von einer Linkskurve ($N'' > 0$) in eine Rechtskurve ($N'' < 0$) über. Für $t < t_w$ ist N' streng monoton wachsend (Linkskurve) und für $t > t_w$ ist N' streng monoton fallend (Rechtskurve).

Starten wir mit $N_0 = N(0) \geq \frac{K}{2}$, gibt es keinen Wendepunkt, und die Wachstumsgeschwindigkeit ist monoton fallend.



Die Lösung berechnet sich zum Beispiel durch Trennung der Variablen und ist:

$$N(t) = \frac{N_0 K e^{rt}}{K + N_0(e^{rt} - 1)} = \frac{N_0 K}{N_0 + (K - N_0)e^{-rt}}.$$

Entwicklung mit $N(0) > K$

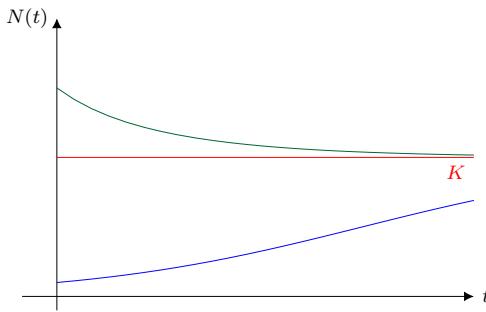
Schauen wir noch den Fall an, wenn unsere Anfangspopulation grösser als K ist. Das kann zum Beispiel durch eine Verunreinigung oder einen Unfall geschehen. Dann haben wir

$$N'(t) = r \cdot N(t) \left(1 - \frac{N(t)}{K}\right), \quad 0 < K < N(0), r > 0,$$

und es folgt

$$\begin{aligned} 0 < K < N(0) &\implies \frac{N(0)}{K} > 1 \\ &\implies 1 - \frac{N(0)}{K} < 0 \\ &\implies N'(0) = r \cdot N(0) \left(1 - \frac{N(0)}{K}\right) < 0 \\ &\implies r \cdot N(t) \left(1 - \frac{N(t)}{K}\right) < 0, \text{ für alle } t \end{aligned}$$

Also ist N streng monoton fallend, solange $K < N(t)$.



5.1.2 Wachstum von Tumoren

In einem weiteren Modell beim Wachstum von Tumoren (nach Gompertz) wird die Anzahl der Zellen mit der DGL

$$N'(t) = r N(t) \ln \left(\frac{K}{N(t)} \right), \quad 0 < N_0 = N(0) < K \text{ und } r > 0, \quad \boxed{\ln(e) = 1}$$

beschrieben. Dies ist ein weiteres Beispiel einer nicht-linearen DGL.

Eine Lösungsfunktion f muss wieder diese DGL und $f(0) = N_0$ erfüllen. Ohne die Lösung explizit zu kennen, können wir qualitativ das Verhalten beschreiben:

1. Es ist $0 < N(0) < K$ und somit $\frac{K}{N(0)} > 1$ und $\ln \left(\frac{K}{N(t)} \right) > 0$. Mit $r > 0$,

$$N'(t) = \underbrace{rN(t)}_{>0} \underbrace{\ln \left(\frac{K}{N(t)} \right)}_{>0} > 0, \text{ solange } N(t) < K,$$

und die Population N ist streng monoton wachsend, solange $N(t) < K$.

2. Für $N(t) \rightarrow K$ ist $\frac{K}{N(t)} \rightarrow 1$ und $\ln \left(\frac{K}{N(t)} \right) \rightarrow 0$. Wir erhalten immer langsameres Wachstum, da die rechte Seite der DGL, $N'(t) = rN(t) \ln \left(\frac{K}{N(t)} \right)$, immer kleiner wird. Die Wachstumsgeschwindigkeit wird immer kleiner.

3. Falls $N(t_\infty) = K$, folgt $\frac{K}{N(t_\infty)} = 1$ und $N'(t_\infty) = rK \ln(1) = 0$ und es gibt kein Wachstum mehr. Damit ist K eine obere Schranke (Kapazität).

Die Entwicklung wächst im Laufe der Zeit und nähert sich immer weiter der Asymptote $N(t_\infty) = K$ an.

4. Wie verändert sich allenfalls die Wachstumsgeschwindigkeit N' ? Wo liegen allfällige Wendepunkte?

Wir schreiben kurz $N = N(t)$ und berechnen N'' (mit Produktregel *):

$$N'' = (N')' = \left(rN \ln \left(\frac{K}{N} \right) \right)' \stackrel{*}{=} rN' \cdot \ln \left(\frac{K}{N} \right) + rN \cdot \ln \left(\frac{K}{N} \right)'.$$

Mit der Kettenregel ist $\ln \left(\frac{K}{N} \right)' = -\frac{N'}{N}$. Damit folgt

$$\begin{aligned} N'' &= rN' \cdot \ln \left(\frac{K}{N} \right) - rN \cdot \frac{N'}{N} = rN' \cdot \ln \left(\frac{K}{N} \right) - rN' \\ &= rN' \cdot \left(\ln \left(\frac{K}{N} \right) - 1 \right). \end{aligned}$$

Für allfällige Wendepunkte brauchen wir die Nullstellen von N'' .

Da $rN' > 0$, suchen wir Nullstellen von $\left(\ln\left(\frac{K}{N}\right) - 1 \right)$:

$$\ln\left(\frac{K}{N}\right) - 1 = 0 \implies \ln\left(\frac{K}{N}\right) = 1 \implies \frac{K}{N} = e \implies N = \frac{K}{e}.$$

Sei t_w der Zeitpunkt, an dem $N(t_w) = \frac{K}{e}$ gilt. Dann hat die 2. Ableitung der Lösungskurve eine Nullstelle.

Sei zu Beginn $N(0) = N_0 < \frac{K}{e}$. Dann ist $\frac{K}{N(0)} > e$ und $\ln\left(\frac{K}{N(0)}\right) > \ln(e) = 1$, also

$$\ln\left(\frac{K}{N(0)}\right) - 1 > 0.$$

Da auch $rN' > 0$, wechselt zur Zeit t_w das Vorzeichen der 2. Ableitung

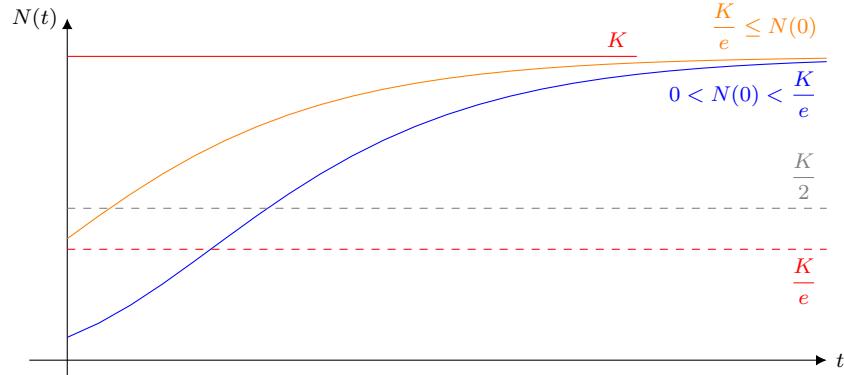
$$N'' = rN' \cdot \left(\ln\left(\frac{K}{N}\right) - 1 \right)$$

von $+$ nach $-$. Also hat N dort einen Wendepunkt. Der Graph von N geht von einer Linkskurve ($N'' > 0$) in eine Rechtskurve ($N'' < 0$) über.

Für $t < t_w$ ist N' streng monoton wachsend (Linkskurve) und für $t > t_w$ ist N' streng monoton fallend (Rechtskurve).

Starten wir mit $N_0 = N(0) \geq \frac{K}{e}$, gibt es keinen Wendepunkt, und die Wachstums geschwindigkeit ist monoton fallend.

Die Lösung ist $N(t) = Ke^{\left(\ln\left(\frac{N_0}{K}\right)e^{-rt}\right)}$ und der Verlauf abhängig von der Startpopulation $N(0)$.



Bemerkung:

Da $e > 2$, ist $\frac{K}{2} > \frac{K}{e}$. Im logistischen Fall beginnt die Wachstumsgeschwindigkeit, beim grösseren Bestand $N = \frac{K}{2}$ zu fallen.

5.1.3 Stationäre Lösungen

Seien $y = y(t)$ und $y' = F(y)$. Hier ist wichtig, dass die rechte Seite $F(y)$ nur von y abhängt.

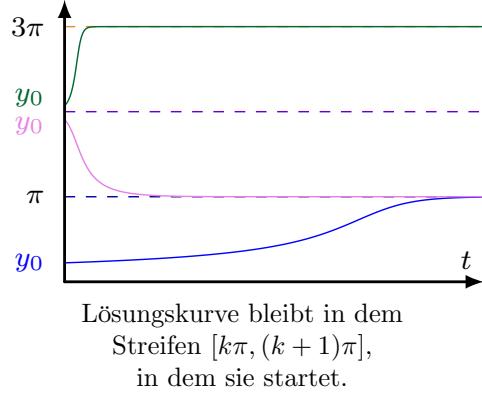
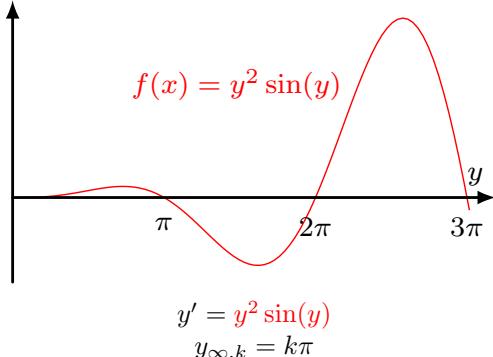
- Die von t unabhängigen Lösungen heißen stationäre Lösung, Gleichgewichtslösung oder auch Fixpunkt.
- Es gilt: y_∞ ist eine stationäre Lösung $\iff y_\infty$ ist konstant $\iff y'_\infty \equiv 0 \iff F(y_\infty) = 0$. Gleichgewichtslösungen sind also gerade die Nullstellen von F .
- Anschaulich ist die Population bei einer stationären Lösung im Gleichgewicht, d.h. $y'_\infty = 0$, es gibt also keine Veränderung.
- Wegen der Eindeutigkeit der Lösung des AWPs schneiden sich zwei Lösungen nicht. Daher sind stationäre Lösungen Kandidaten für Asymptoten.

5.1.4 Beschreibung in der Nähe von y_∞ , Linearisierung

Pendent ist eine Antwort auf folgende Frage.

Wie verhalten sich die Lösungen in der Nähe einer Gleichgewichtslösung y_∞ ? Konvergieren sie gegen y_∞ , oder werden sie von y_∞ abgestossen? Können wir ein Kriterium für die Konvergenz in der Nähe y_∞ angeben?

Beispiel 5.1.1. Schauen wir die Frage für das Beispiel $y' = y^2 \sin(y)$ und mit einer Lösung f , welche in einem Streifen $[k\pi, (k+1)\pi]$ startet.



Welche stationäre Lösung $y_{\infty,k} = n\pi$ zieht f an? Welches $y_{\infty,k}$ ist ein Attraktor, welches ein Repeller?

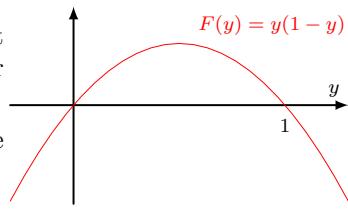
$$f(t) \longrightarrow \begin{cases} k\pi = y_{\infty,k} & ? \\ (k+1)\pi = y_{\infty,k+1} & ? \end{cases} \quad \text{für } t \rightarrow \infty$$

Untersuchen wir die Frage oben mit der logistischen DGL $y' = y(1-y) = F(y)$ genauer:

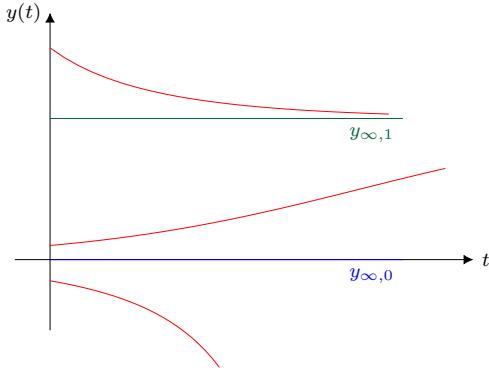
Diese DGL hat die stationären Lösungen $y_{\infty,0} = 0$ und $y_{\infty,1} = 1$. Beachten Sie, dass die Steigung einer Lösung (also y') am Wert der Funktion $F(y)$ abgelesen werden kann:

Die Steigung einer Lösung bei $y = 0$ ist gleich Null, nimmt dann zu, ist bei $y = \frac{1}{2}$ am grössten und nimmt dann wieder ab bis sie bei $y = 1$ wieder den Wert Null erreicht.

Ist $y < 0$ oder $y > 1$ ist der Wert von $F(y) < 0$, die Steigung der Lösung also negativ.



Dies ergibt folgendes Richtungsfeld mit roten Lösungen:



Beobachtungen: Die grüne stationäre Lösung $y_{\infty,1}$ zieht andere Lösungen in ihrer Nähe an und heisst daher **Attraktor**. Die blaue stationäre Lösung $y_{\infty,0}$ stösst andere Lösungen in ihrer Nähe ab und heisst daher **Repeller**.

Wie lässt sich *allgemein* das Verhalten von Lösungen in der Nähe von stationären Lösungen genauer bestimmen? Wir wissen bereits:

Die stationären Lösungen von $y' = F(y)$ sind die Nullstellen von F .

Idee ist nun: Linearisieren von $y' = F(y)$ bei der stationären Lösung y_∞ .

Das bedeutet, wir machen den **Ansatz** $y(t) = y_\infty + h(t)$ für eine kleine Störung $h(t)$. Einsetzen in die DGL liefert

$$y' = h' = F(y_\infty + h) \stackrel{\text{Taylor}}{=} \underbrace{F(y_\infty)}_{=0} + F'(y_\infty)h + \text{ThO}.$$

Dabei enthalten die **Terme höherer Ordnung (ThO)** Potenzen h^2, h^3, \dots die vernachlässigbar klein sind. Linearisieren bedeutet nun, dass wir die ThO weglassen, d.h. h erfüllt näherungsweise

$$h' = F'(y_\infty)h, \text{ also } h(t) = h_0 e^{F'(y_\infty)t}.$$

Dies in den Ansatz eingesetzt ergibt die Näherungslösung zur Anfangsbedingung $y(0) = y_0$:

$$y(t) = y_\infty + (y_0 - y_\infty)e^{F'(y_\infty)t} \quad (5.1)$$

Wir halten also fest:

- Die stationären Lösungen der nicht linearen Gleichung $y' = F(y)$ sind die Nullstellen y_∞ von F .

- Die Lösung (5.1) der linearisierten Gleichung verhält sich in der Nähe der Gleichgewichtslösung y_∞ wie die exakte Lösung von $y' = F(y)$:
 - Ist $F'(y_\infty) < 0$ gilt $e^{F'(y_\infty)t} \rightarrow 0$ für $t \rightarrow \infty$. Für Lösungen y , die nahe genug bei y_∞ starten folgt: $y(t) \rightarrow y_\infty$ für $t \rightarrow \infty$.
Die stationäre Lösung y_∞ ist in diesem Fall ein **Attraktor** oder synonym ein **stabiles Gleichgewicht**.
 - Ist $F'(y_\infty) > 0$ so gilt: Egal wie nahe eine Lösung y bei y_∞ startet, $y(t)$ entfernt sich rasch von y_∞ .
Die stationäre Lösung y_∞ ist in diesem Fall ein **Repeller** oder synonym ein **instabiles Gleichgewicht**.
- Gedankenstütze durch eine pendelnde Stange (mit Reibung): Die untere Gleichgewichtslage ist stabil, die obere instabil.

5.2 Mehrdimensionale Modelle

Gegeben seien zwei Populationen mit Räubern x und Beute y .

Erinnern wir an die diskrete Beschreibung durch zwei Folgen.

Räuber (x_n) entwickele sich gemäss $x_{n+1} = ax_n + by_n$

Beute (y_n) entwickele sich gemäss $y_{n+1} = cy_n - dx_n$

Dabei sind typischerweise $0 < a, b, d < 1$ und $c > 1$. Das heisst, die Population x_n wächst auf Kosten der Population y_n . Die beiden Gleichungen schreiben wir kompakt als ein Matrix-Vektor-Produkt: $v_{n+1} = Av_n$ mit den Vektoren $v_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix} \in \mathbb{R}^2$ und der 2×2 -Matrix $A = \begin{pmatrix} a & b \\ -d & c \end{pmatrix}$.

Für kontinuierliche Entwicklungen erhalten wir mit den linearen Modellen eine analoge Situation für eine Räuberpopulation $x(t)$ und eine Beutepopulation $y(t)$:

Räuber $x(t)$ entwickelt sich gemäss $x'(t) = \alpha x(t) + \beta y(t)$

Beute $y(t)$ entwickelt sich gemäss $y'(t) = \gamma y(t) - \delta x(t)$

Oder in Matrixschreibweise $\begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ -\delta & \gamma \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$. Oder in anderer Notation:

$$y' = Ay, \quad \begin{pmatrix} y'_1(t) \\ y'_2(t) \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ -\delta & \gamma \end{pmatrix} \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix}.$$

5.2.1 Räuber-Beute-Modell nach Lotka-Volterra

Nun untersuchen wir Modelle, die sich nicht als ein System $y' = Ay$ beschreiben lassen.

Beispiel 5.2.1. Betrachten wir zunächst die Entwicklung einer Population x mit folgenden Annahmen:

- Die Wachstumsrate der Population $\frac{x'}{x}$ ist nicht konstant.
- Sie hängt von der zur Zeit t vorhandenen Nahrung $n(t)$ ab.
- Es soll Wachstum nur dann geben, wenn genügend Nahrung vorhanden ist, d.h. wenn $n(t) > M$.

Also erhalten wir insgesamt ein Modell $x'(t) = p(t)x(t)$ mit $p(t) = n(t) - M$.

Sei die Nahrung n nun gegeben durch eine zweite Population y von Beutetieren. Dann erhalten wir ein nicht lineares kontinuierliches Räuber-Beute-Modell. Die Modellannahmen sind:

- $x(t)$: Anzahl der Räuber zur Zeit t .
- $y(t)$: Anzahl der Beutetiere zur Zeit t .
- $s > 0$: Sterberate der Räuber, wenn keine Beute vorhanden ist, das heisst $\Delta x = x(t + \Delta t) - x(t) = -sx(t)\Delta t$.
- $r > 0$: Reproduktionsrate der Beute in Abwesenheit von Räubern und bei grossem Nahrungsangebot, d.h. $\Delta y = y(t + \Delta t) - y(t) = ry(t)\Delta t$.
- Die Anzahl der Begegnungen im Zeitraum Δt zwischen Räuber und Beutetier sei $\alpha xy\Delta t$ mit $\alpha > 0$. Diese Annahme ist plausibel: Ist die Dichte der Räuber doppelt so hoch, wird auch die Wahrscheinlichkeit einer Begegnung mit einem Beutetier verdoppelt. Ebenso, wenn die Dichte der Beutetiere oder die Zeitdauer Δt sich verdoppelt.
- Eine Begegnung eines Räubers mit einem Beutetier führt mit einer gewissen Wahrscheinlichkeit dazu, dass letzteres gefressen wird. Die Jagderfolgswahrscheinlichkeit sei β . Die Abnahme der Beutetiere durch gefressen werden ist demnach gegeben durch $\Delta y = -\alpha\beta xy\Delta t$.
- Der Jagderfolg erhöht die Überlebenswahrscheinlichkeit und die Fertilität der Räuber. Diesen Effekt beschreiben wir durch den Term $\Delta x = \alpha\gamma xy\Delta t$, indem wir eine instantane (sofortige) Wirkung postulieren.

Die Änderung der Räuberpopulation setzt sich also zusammen aus der **Abnahme durch natürlichen Tod** und **Zunahme durch Reproduktion**: $\Delta x = -sx(t)\Delta t + \alpha\gamma xy\Delta t$. Die Änderung der Beutepopulation setzt sich entsprechend zusammen aus der **Zunahme durch Reproduktion** und **Dezimierung durch Frass**: $\Delta y = ry(t)\Delta t - \alpha\beta xy\Delta t$. Nach Division durch Δt und Grenzübergang $\Delta t \rightarrow 0$ erhalten wir die Lotka-Volterra Differentialgleichungen:

$$\begin{aligned} x'(t) &= -sx(t) + \alpha\gamma xy \\ y'(t) &= ry(t) - \alpha\beta xy. \end{aligned}$$

Division durch Δt und Grenzübergang $\Delta t \rightarrow 0$ liefert das nicht lineare System

$$\left. \begin{aligned} x' &= -sx + \alpha\gamma xy \\ y' &= ry - \alpha\beta xy \end{aligned} \right\}$$

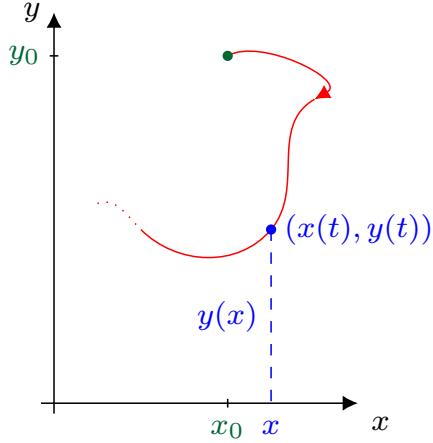
Der Einfachheit halber betrachten wir nur den Fall $r = s = \alpha\beta = \alpha\gamma = 1$:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} -x + xy \\ y - xy \end{pmatrix} \quad (5.2)$$

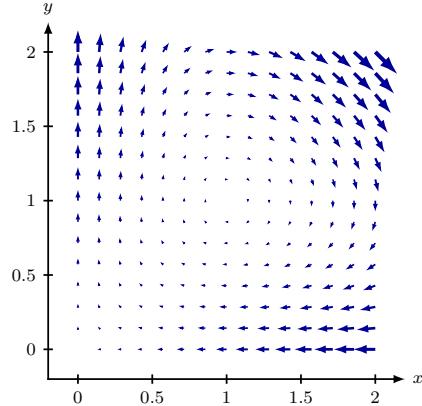
Leider kann keine explizite Lösung gefunden werden. Einen Ausweg suchen wir im nächsten Abschnitt.

5.2.2 Lösungskurven im Lotka-Volterra-Modell

Wir beschreiben die **Lösungskurven** lokal in der (x, y) -Ebene als Graph einer Funktion $x \mapsto y(x)$: Wie bestimmt die Zahl der Räuber $\textcolor{red}{x}$ die Zahl der Beute $\textcolor{green}{y}$?



Betrachten wir zunächst das Vektorfeld, welches durch die beiden DGLs definiert wird, sehen wir ein periodisches Verhalten einer Lösungskurve: $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} -x + xy \\ y - xy \end{pmatrix}$



Das heisst, dass der Punkt $(x(t), y(t))$ der Lösung auf den geschlossenen Höhenlinien im Uhrzeigersinn herum wandert und also periodisch ist.

Die zweite Zeile des DGL Systems (5.2) und die Verallgemeinerte Kettenregel (*) liefern:

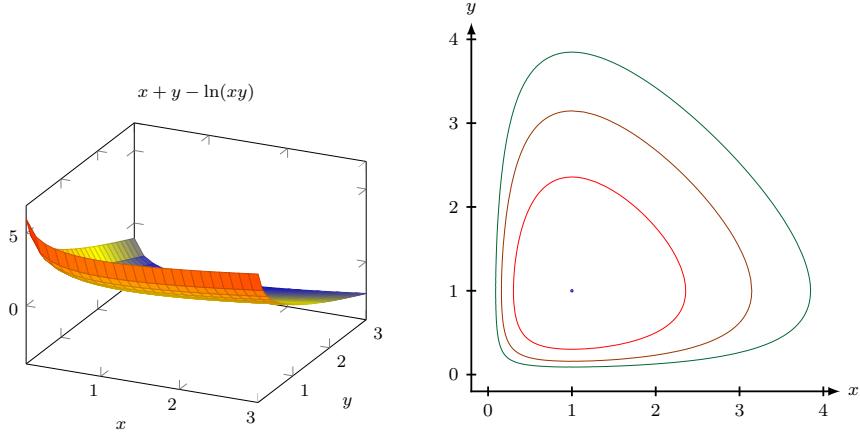
$$\textcolor{green}{y} - \textcolor{orange}{xy} = y'(t) = y'(x(t)) \stackrel{(*)}{=} y'(x)x'(t) = y'(x) \underbrace{(-x + xy)}_{\text{1. Zeile}}.$$

Nach Umstellen entsteht die separierbare DGL $y'(x) = -\frac{y(1-x)}{x(1-y)} = \frac{dy}{dx}$, welche wir mit Tren-

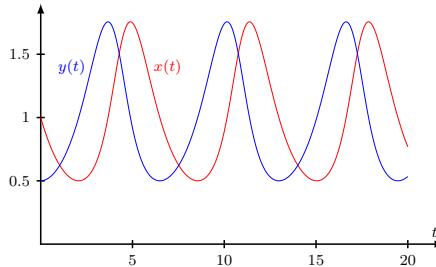
nung der Variablen lösen:

$$\left(\frac{1}{x} - 1 \right) dx = \left(-\frac{1}{y} + 1 \right) dy \underset{\text{Integration}}{\implies} \ln(x) - x + C = -\ln(y) + y.$$

Die Konstante C und Gleichung $C = \textcolor{teal}{x} + \textcolor{violet}{y} - \ln(xy)$ legen also eine Beziehung zwischen Räuber und Beute fest. Kennen wir also zum Beispiel die Räuber ist mit C auch die Beute bestimmt. Die Funktion $f(x, y) := \textcolor{teal}{x} + \textcolor{violet}{y} - \ln(xy)$ ist konvex, d.h. der Graph von f ist eine Talsenke. Die Höhenlinien $\{(x, y) \mid f(x, y) = C\}$ sind daher geschlossene Kurven:



Tatsächlich bestätigt die numerische Lösung unseren qualitativen Befunde:



Wir halten fest:

- Die Lösungen der Lotka-Volterra DGL $\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} -x + xy \\ y - xy \end{pmatrix}$ sind periodisch.
- Die Räuber können die Beute nicht ausrotten. Ebenso wenig sterben die Räuber aus.
- Einziger Gleichgewichtspunkt des Systems ist $(x, y) = (1, 1)$.

5.2.3 Linearisieren in 2 Dimensionen

Das Räuber-Beute Modell von Lotka-Volterra ist ein nicht lineares System von DGLn.

Ein solches System setzt sich zusammen aus einer zweimal stetig differenzierbare Funktion $F : \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \mapsto \begin{pmatrix} F_1(y_1, y_2) \\ F_2(y_1, y_2) \end{pmatrix}$ und einer unbekannte Funktion $t \mapsto y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix}$, welche durch die Gleichung

$$y' = F(y) \quad (5.3)$$

verknüpft sind.

Wie im eindimensionalen Fall (5.1.3) ist $y_\infty = \begin{pmatrix} y_{\infty,1} \\ y_{\infty,2} \end{pmatrix} = \text{konstant}$ eine stationäre Lösung und dazu äquivalent sind

$$y'_\infty = 0 \iff F(y_\infty) = \begin{pmatrix} F_1(y_{\infty,1}, y_{\infty,2}) \\ F_2(y_{\infty,1}, y_{\infty,2}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Also gilt wieder:

Die stationären Lösungen von $y' = F(y)$ sind die Nullstellen von F .

Durch Linearisieren von (5.3) bei der stationären Lösung y_∞ versuchen wir die tatsächliche Lösung des Systems zu beschreiben.

Zweidimensionale Taylor-Entwicklung

Dazu brauchen wir zuerst die 2-dimensionale Taylor-Entwicklung:

Lemma 5.2.2. *Sei $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ zweimal stetig differenzierbar, so gilt*

$$f(x_1 + h_1, x_2 + h_2) = f(x_1, x_2) + \frac{\partial f(x_1, x_2)}{\partial x_1} h_1 + \frac{\partial f(x_1, x_2)}{\partial x_2} h_2 + \text{ThO}.$$

Für eine geeignete Konstante C lassen sich die Terme höherer Ordnung (ThO) abschätzen durch $| \text{ThO} | < C(h_1^2 + h_2^2)$.

Beweis. Es gilt aufgrund der eindimensionalen Taylor-Entwicklung nacheinander

$$\begin{aligned} f(x_1 + h_1, x_2 + h_2) &= f(x_1, x_2 + h_2) + \frac{\partial f(x_1, x_2 + h_2)}{\partial x_1} h_1 + \text{ThO} \\ f(x_1, x_2 + h_2) &= f(x_1, x_2) + \frac{\partial f(x_1, x_2)}{\partial x_2} h_2 + \text{ThO} \\ \frac{\partial f(x_1, x_2 + h_2)}{\partial x_1} &= \frac{\partial f(x_1, x_2)}{\partial x_1} + \frac{\partial^2 f(x_1, x_2)}{\partial x_1 \partial x_2} h_2 + \text{ThO}. \end{aligned}$$

Einsetzen der farbigen Terme in der ersten Zeile liefert das Gewünschte. \square

Um das DGL-System $y' = F(y)$ (5.3) im stationären Punkt y_∞ zu linearisieren, brauchen wir den **Ansatz** $y(t) = y_\infty + h(t)$. Einsetzen in (5.3) liefert mit Taylor

$$\begin{aligned} y' &= \underbrace{y'_\infty}_{=0} + \color{red}h'\color{black} = F(y_\infty + h) = \begin{pmatrix} F_1(y_{\infty,1} + h_1, y_{\infty,2} + h_2) \\ F_2(y_{\infty,1} + h_1, y_{\infty,2} + h_2) \end{pmatrix} \\ &= \underbrace{\begin{pmatrix} F_1(y_{\infty,1}, y_{\infty,2}) \\ F_2(y_{\infty,1}, y_{\infty,2}) \end{pmatrix}}_{=0} + \underbrace{\begin{pmatrix} \frac{\partial F_1(y_{\infty,1}, y_{\infty,2})}{\partial y_1} & \frac{\partial F_1(y_{\infty,1}, y_{\infty,2})}{\partial y_2} \\ \frac{\partial F_2(y_{\infty,1}, y_{\infty,2})}{\partial y_1} & \frac{\partial F_2(y_{\infty,1}, y_{\infty,2})}{\partial y_2} \end{pmatrix}}_{=:DF(y_\infty)} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \text{ThO}. \end{aligned}$$

Also lautet die DGL für h :

$$\textcolor{red}{h}' = \textcolor{blue}{DF}(\textcolor{blue}{y}_\infty) \textcolor{green}{h}. \quad (5.4)$$

Mit $\textcolor{orange}{y} = \textcolor{blue}{y}_\infty + h \implies h = \textcolor{orange}{y} - \textcolor{blue}{y}_\infty$ erhalten wir daraus die **linearisierte Gleichung**:

$$\textcolor{blue}{y}' = \textcolor{blue}{DF}(\textcolor{blue}{y}_\infty)(\textcolor{blue}{y} - \textcolor{blue}{y}_\infty). \quad (5.5)$$

Die Lösung von (5.4) ist $h(t) = e^{DF(y_\infty)t}h_0$. Dies eingesetzt in den Ansatz $\textcolor{orange}{y} = \textcolor{blue}{y}_\infty + h$ ergibt für die Anfangsdaten $y(0) = y_0$ die Lösung von (5.5)

$$y(t) = \textcolor{blue}{y}_\infty + e^{DF(y_\infty)t}(y_0 - \textcolor{blue}{y}_\infty). \quad (5.6)$$

Es gilt dann der Satz

Theorem 5.2.3 (Linearisierung von Systemen: Satz von Hartman-Grobman). *Falls alle Eigenwerte von $DF(y_\infty)$ Realteil verschieden von 0 haben, verhält sich das linearisierte System in der Nähe eines Gleichgewichtspunkts y_∞ gleich wie die echte Lösung.*

Beispiele.

1. Für die Lösungen der Lotka-Volterra DGL $\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} -x + xy \\ y - xy \end{pmatrix}$ hilft uns diese Aussage leider nicht - warum?

2. Wir betrachten das System $y' = F(y)$ für $y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix}$ und $F : \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \mapsto \begin{pmatrix} y_1 - y_1^2 y_2 \\ -y_2 + y_1 y_2 \end{pmatrix}$.

Offenbar ist $y_\infty = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ eine stationäre Lösung. Es gilt

$$DF(y_\infty) = \begin{pmatrix} \frac{\partial F_1(1,1)}{\partial y_1} & \frac{\partial F_1(1,1)}{\partial y_2} \\ \frac{\partial F_2(1,1)}{\partial y_1} & \frac{\partial F_2(1,1)}{\partial y_2} \end{pmatrix} = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix}.$$

Die EW der Matrix sind $-\frac{1}{2} \pm \frac{i}{2}\sqrt{3}$. Somit ist y_∞ ein Attraktor, und zwar ein Wirbel.

3. Wir betrachten das System von DGLn $y' = F(y)$ mit

$$F : \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \mapsto \begin{pmatrix} 1 - 2y_1 + y_1 y_2 \\ 3 - 2(y_1 + y_2) + y_1 y_2 \end{pmatrix}.$$

Dann ist $\textcolor{blue}{y}_\infty = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ eine stationäre Lösung denn $F(\textcolor{blue}{1}, \textcolor{blue}{1}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Man rechnet

$$DF(\textcolor{blue}{1}, \textcolor{blue}{1}) = \begin{pmatrix} -1 & 1 \\ -1 & -1 \end{pmatrix}.$$

Diese Matrix hat die EW $\lambda_1 = -1 + i$, $\lambda_2 = -1 - i$ mit EV $v_1 = \begin{pmatrix} -i \\ 1 \end{pmatrix}$, $v_2 = \begin{pmatrix} i \\ 1 \end{pmatrix}$.

Also lässt sich A diagonalisieren, und es gilt für $T = \begin{pmatrix} -i & i \\ 1 & 1 \end{pmatrix}$ und $T^{-1} = \frac{1}{2} \begin{pmatrix} i & 1 \\ -i & 1 \end{pmatrix}$,

dass $T^{-1}AT = \text{diag}(-1+i, -1-i) =: d$. Somit

$$\begin{aligned} e^{DF(y_\infty)t} &= Te^{\textcolor{brown}{d}t}T^{-1} = T \begin{pmatrix} e^{(-1+i)t} & 0 \\ 0 & e^{(-1-i)t} \end{pmatrix} T^{-1} \\ &= T \begin{pmatrix} e^{-t}(\cos t + i \sin t) & 0 \\ 0 & e^{-t}(\cos t - i \sin t) \end{pmatrix} T^{-1} \\ &= e^{-t} \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}. \end{aligned}$$

Also ist die Lösung der linearisierten Gleichung mit Anfangsbedingung $y(0) = y_0 = \begin{pmatrix} y_{0,1} \\ y_{0,2} \end{pmatrix}$,

dass $y(t) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + e^{-t} \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \begin{pmatrix} y_{0,1} - 1 \\ y_{0,2} - 1 \end{pmatrix}$. D.h. diese Lösung nähert sich auf einer logarithmischen Spirale dem stationären Punkt y_∞ .

Alternativ lässt sich die Gleichung für h , also $h' = DF(1,1)h$, natürlich auch so lösen: Da die DGL reelle Koeffizienten hat, sind Real- und Imaginärteil der Lösung $e^{\lambda_1 t} v_1$ auch wieder Lösungen, ebenso Linearkombinationen davon:

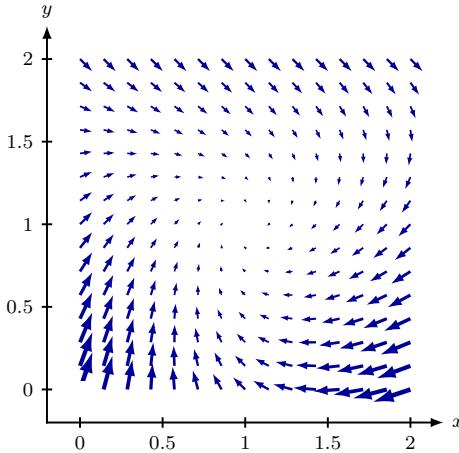
$$\begin{aligned} \text{Re}(e^{\lambda_1 t} v_1) &= \text{Re}\left(e^{(-1+i)t} \begin{pmatrix} -i \\ 1 \end{pmatrix}\right) = \text{Re}\left(e^{-t}(\cos t + i \sin t) \begin{pmatrix} -i \\ 1 \end{pmatrix}\right) = e^{-t} \begin{pmatrix} \sin t \\ \cos t \end{pmatrix} \\ \text{Im}(e^{\lambda_1 t} v_1) &= \text{color{orange}}{e^{-t}} \begin{pmatrix} -\cos t \\ \sin t \end{pmatrix}. \end{aligned}$$

Also

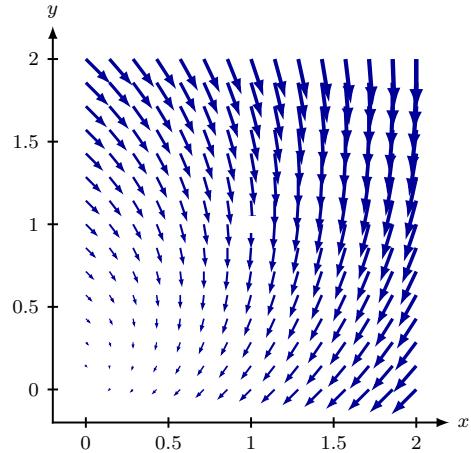
$$h(t) = -c_1 \text{color{orange}}{e^{-t}} \begin{pmatrix} -\cos t \\ \sin t \end{pmatrix} + c_2 e^{-t} \begin{pmatrix} \sin t \\ \cos t \end{pmatrix} = e^{-t} \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

und dann $y(t) = y_\infty + h(t)$ in Übereinstimmung mit dem Resultat von vorher.

Vergleichen wir das Strömungsbild des echten nicht linearen Systems (links) mit der linearisierten Situation (rechts) bestätigt sich die gute Übereinstimmung in der Nähe des Fixpunktes $(1, 1)$:



Nicht lineares System



Lineare Approximation

KAPITEL 6

Partielle Differentialgleichungen

6.1 Einführung

Bisher haben wir Systeme behandelt, die lediglich von der Zeit abhängen. Insbesondere bei den Boxmodellen war dies nur möglich durch die vereinfachende Annahme der vollständigen Durchmischung. Zum Beispiel bei der Beschreibung der Konzentration von Phosphat in einem See nimmt man an, dass diese im ganzen See gleich gross ist. Von dieser Annahme, die in der Realität nur sehr näherungsweise zutrifft, wollen wir uns nun lösen. Der Grund ist, dass zahlreiche Phänomene sowohl von der Zeit als auch vom Ort abhängen.

6.1.1 ODEs vs. PDEs: Ein Vergleich

Mit Hilfe von **Gewöhnlichen Differentialgleichungen** (ODEs für *ordinary differential equations*) modelliert man Vorgänge, die von *einer einzigen* unabhängigen Variablen (oft die Zeit t) abhängen. Von dieser unabhängigen Variablen hängen dann eine oder mehrere veränderliche Grössen ab, zum Beispiel die Konzentration $C(t)$ eines Medikaments in einem Organ, oder die Grösse einer Räuber- und einer Beutepopulation. Je nach Ordnung der ODEs und Anzahl gesuchter Funktionen sind eine entsprechende Anzahl von Anfangsbedingungen nötig, um die Lösung zu bestimmen.

Beispiel 6.1.1. Hängt die Änderungsrate der Konzentration $C(t)$ einer Substanz (zum Beispiel nichtlinear) in einem durchmischten Reaktor von der Konzentration selber und der Zeit t ab, wird der entsprechende Vorgang wie folgt beschrieben:

$$\begin{aligned} \text{ODE: } & C'(t) = f(C(t), t) \\ \text{AB: } & C(0) = C_0 \quad (\text{Anfangsbedingung}). \end{aligned}$$

Partielle Differentialgleichungen (PDEs für *partial differential equations*) begegnen einem, wenn der betrachtete Vorgang von zwei oder mehr unabhängigen Variablen (etwa Zeit t und Ort $x \in \mathbb{R}^3$) anhängt. Je nachdem können eine oder mehr abhängige Variablen (zum Beispiel die Konzentration $C(x, t)$ eines Medikaments und/oder die Temperatur $T(x, t)$) im Spiel sein. Neben Anfangsbedingungen sind hier gegebenenfalls auch Randbedingungen zu berücksichtigen.

Beispiel 6.1.2. Die Temperaturrentwicklung $T(x, t)$ in einem Gebiet $\Omega \subset \mathbb{R}^3$ wird, wie wir im Abschnitt 6.1.3 sehen werden, wie folgt modelliert:

$$\begin{aligned}\text{PDE: } \quad & \frac{\partial T(x, t)}{\partial t} = \Delta T(x, t) \\ \text{AB: } \quad & T(x, 0) = T_0(x) \quad \text{für } x \in \Omega \text{ (Anfangsbedingung)} \\ \text{RB: } \quad & T(x, t) = f(x, t) \quad \text{für } x \in \partial\Omega, t > 0 \text{ (Randbedingung).}\end{aligned}$$

Dabei bezeichnet

$$\Delta = \sum_{i=1}^n \frac{\partial^2}{\partial x_i^2}$$

den Laplace-Operator in \mathbb{R}^n , und $\partial\Omega$ bezeichnet den Rand von Ω .

Bevor wir uns ans Lösen solcher PDEs machen, halten wir kurz inne und stellen uns zwei Fragen:

- Was für Phänomene führen auf Modelle, welche mit Hilfe von PDEs beschrieben werden?
- Wie erfolgt die Modellbildung mit PDEs?

Die erste Frage hat mit dem Vergleich ODEs vs. PDEs bereits eine erste Antwort gefunden: Sobald Phänomene betrachtet werden, bei denen sowohl Zeit als auch Ort eine Rolle spielen, kommen PDEs natürlicherweise ins Spiel. Dies gilt insbesondere, wenn sich die Grundannahme bei Boxmodellen (vollständige Durchmischung) als zu grob herausstellt. So findet eben zum Beispiel die Algenblüte in den Weltmeeren nicht überall gleichzeitig und gleich stark statt, sondern, sie hängt von der Zeit *und* vom Ort ab. Erwärmt man einen Metallstab an einem Ende, breitet sich die Temperaturerhöhung allmählich im Laufe der Zeit zum andern Ende aus.

6.1.2 Begriffe aus der Vektoranalysis

Wir beginnen mit einer Zusammenstellung wichtiger Begriffe aus der Vektoranalysis.

Geometrische und physikalische Interpretation eines Differentialgleichungssystems

Eine $n \times n$ -Matrix definiere ein homogenes System von Differentialgleichungen $x' = Ax$. In jedem Punkt $\textcolor{red}{x} \in \mathbb{R}^n$ ist $\textcolor{blue}{Ax} \in \mathbb{R}^n$ ein Vektor. Wir erhalten damit ein **Vektorfeld**

$$F : \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad \textcolor{red}{x} \mapsto \textcolor{blue}{Ax}$$

(siehe Abbildung 6.1): In jedem Punkt $\textcolor{red}{x}$ denken wir uns den Vektor $\textcolor{blue}{Ax}$ befestigt. Eine Lösung von $x' = Ax$ ist dann eine stetig differenzierbare Funktion

$$\mathbb{R} \rightarrow \mathbb{R}^n, \quad t \mapsto x(t), \tag{6.1}$$

welche für alle t die Gleichung $x'(t) = Ax(t)$ erfüllt. Eine solche Lösung (6.1) stellt somit geometrisch eine Kurve im \mathbb{R}^n dar: Für $n = 2$ ist

$$\textcolor{red}{t} \mapsto x(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}$$

eine ebene Kurve.

Für $n = 3$ ist

$$t \mapsto x(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix}$$

eine Raumkurve.

Interpretieren wir t als Zeit, so ist die Geschwindigkeit $x'(t)$ entlang der Kurve in jedem Punkt $x(t) \in \mathbb{R}^n$ der dort vorgegebene Vektor $Ax(t) \in \mathbb{R}^n$. Die Lösungskurve ist also in jedem Punkt tangential an das Vektorfeld. In der Abbildung 6.1 ist eine rote Lösungstrajektorie eingezeichnet und in zwei ihrer Punkte die grünen Tangentialvektoren.

Sei das Vektorfeld F zum Beispiel das Geschwindigkeitsfeld einer Flüssigkeit. Ein Teilchen wird zur Zeit $t = 0$ am Ort x_0 in die Strömung geworfen. Das Teilchen wird dann von der Strömung mitgerissen: In jedem Punkt x seiner Reise hat es also die durch die Strömung dort gegebenen Geschwindigkeit Ax . Das heisst, die Reise des Teilchens $t \mapsto x(t)$ realisiert genau die Lösung des Anfangswertproblems $x'(t) = Ax, x(0) = x_0$. Es befindet sich zur Zeit t am durch die Lösung gegebenen Ort $x(t)$ (siehe Abbildung 6.1).

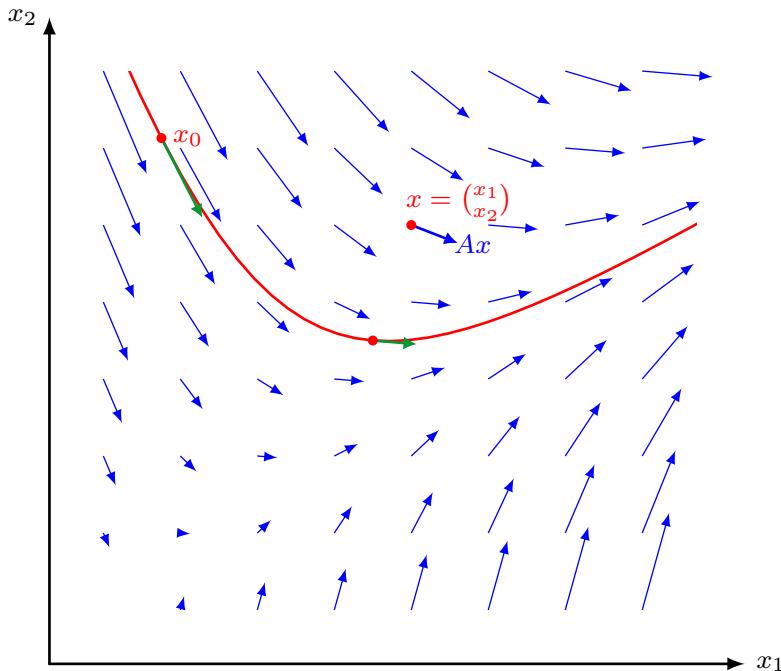


Abbildung 6.1: Interpretation der Lösung des Systems $x'(t) = Ax(t)$ mit dem Anfangswert $x(0) = x_0$.

Nun repetieren wir noch weitere Begriffe der Vektoranalysis:

Definition 6.1.3. Sei $u : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ differenzierbar. Dann ist

$$\text{grad } u = \nabla u = \left(\frac{\partial u}{\partial x_1}, \dots, \frac{\partial u}{\partial x_n} \right)$$

der **Gradient von u** (in kartesischen Koordinaten).

Geometrische Interpretation. Stellt man sich $u(x)$ als „Höhe über Meer“ des Punktes x vor, so beschreibt der Graph von u eine Landschaft. Es gilt:

- Der Gradient $\nabla u(x)$ steht senkrecht auf der Niveaumenge

$$M_c = \{z \in \Omega : u(z) = c\}$$

welche x enthält (man denke an die Höhenlinien auf einer Landkarte): Siehe Abbildung 6.2.

- Der Gradient zeigt in Richtung des stärksten Anstiegs von u .

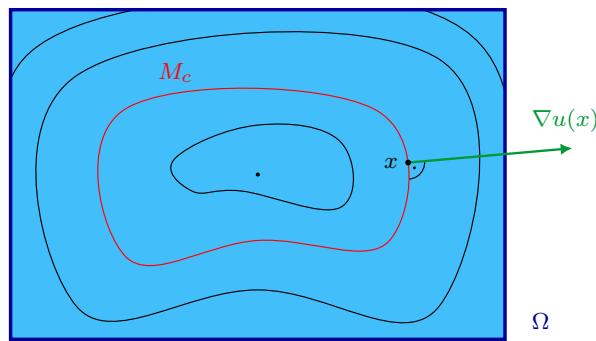


Abbildung 6.2: Geometrische Interpretation des Gradienten.

Physikalische Interpretation. Ein Vektorfeld $v : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}^3$ kann auch aufgefasst werden als ein Kraftfeld: In jedem Punkt $x \in \Omega$ wirkt eine Kraft $v(x)$. Beispiele sind das Gravitationsfeld der Erde oder das elektrische Feld einer Ladungsverteilung. Jedenfalls heisst das Vektorfeld **konservativ**, wenn es ein Potential $\phi : \Omega \rightarrow \mathbb{R}$ besitzt, das heisst, falls $\nabla \phi = v$. Die Arbeit des Kraftfeldes v entlang einer Kurve

$$\gamma : [a, b] \rightarrow \Omega$$

ist dann (gemäss „Arbeit = Kraft \cdot Weg“¹) gegeben durch das Wegintegral

$$\begin{aligned} \int_{\gamma} v \cdot ds &= \int_a^b v(\gamma(t)) \cdot \dot{\gamma}(t) dt \\ &= \int_a^b \nabla \phi(\gamma(t)) \cdot \dot{\gamma}(t) dt \\ &= \int_a^b \frac{d}{dt} \phi(\gamma(t)) dt \\ &= \phi(\gamma(b)) - \phi(\gamma(a)). \end{aligned}$$

Das heisst, die Arbeit hängt nur vom Anfangs- und vom Endpunkt des Weges ab, jedoch nicht von dessen Verlauf.

¹Der Punkt \cdot bedeutet hier und im Folgenden das Skalarprodukt.

Beispiel 6.1.4. Lokal wird das Kraftfeld der Erde, welches auf eine Masse m wirkt, dargestellt durch

$$v : \mathbb{R}^3 \rightarrow \mathbb{R}^3, \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} 0 \\ 0 \\ -mg \end{pmatrix}.$$

Die Gewichtskraft mg wirkt senkrecht nach unten. Die Funktion

$$\phi : \mathbb{R}^3 \rightarrow \mathbb{R}, \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto -mgz$$

ist dann ein Potential (die potentielle Energie der Masse m), da $\nabla\phi(x) = v(x)$. Die Arbeit, die man aufwenden muss, um die Masse m vom Niveau z_1 auf das Niveau z_2 zu heben, ist die Differenz der potentiellen Energie, also $\phi(z_2) - \phi(z_1) = mgz_2 - mgz_1 = mg(z_2 - z_1)$, unabhängig vom Weg.

Definition 6.1.5. Sei $v : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ ein differenzierbares Vektorfeld. Dann ist

$$\operatorname{div} v = \sum_{i=1}^n \frac{\partial v_i}{\partial x_i}$$

die Divergenz von v (in kartesischen Koordinaten).

Interpretation in 2 Dimensionen. Das Vektorfeld $v : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ beschreibt die Geschwindigkeit einer Flüssigkeit (siehe Abbildung 6.3). Dann ist die Flüssigkeitsmenge, die pro Zeiteinheit durch den Rand $\partial\Omega$ eines Gebietes Ω nach aussen fliesst

$$\int_{\partial\Omega} v \cdot n \, ds \stackrel{\text{Gauss}}{=} \int_{\Omega} \operatorname{div} v \, dV, \quad (6.2)$$

wobei wir den Divergenzsatz von Gauss verwendet haben.

Wird nirgends Flüssigkeit erzeugt oder vernichtet, so ist diese Nettomenge für beliebige Wahl von Ω gleich 0, und daher $\operatorname{div} v \equiv 0$. Gilt umgekehrt $\operatorname{div} v \equiv 0$, so ist die Nettomenge für beliebiges Ω gleich 0. Wir haben also Massenerhaltung: Es fliesst gleichviel Flüssigkeit in Ω hinein wie hinaus.

Stellt man sich aber vor, die Flüssigkeit fliesse zwischen einer soliden Grund- und einer porösen Deckplatte, so kann lokal durch Infiltration oder Diffusion Flüssigkeit zu- oder abgeführt werden: Siehe Abbildung 6.4. Da (6.2) für beliebiges Ω gilt, gibt $\operatorname{div} v$ also gerade an, wieviel Flüssigkeit lokal pro Flächeneinheit hinzukommt oder weggeht. Man nennt daher $\operatorname{div} v$ auch die **Quellenstärke** des Feldes v .

Interpretation in 3 Dimensionen. Wir betrachten ein kompressibles Gas mit Dichte $\rho(x, t)$, welches ein Gebiet $\Omega \subset \mathbb{R}^3$ mit einer vom Ort $x \in \mathbb{R}^3$ abhängigen Geschwindigkeit $v(x)$ durchströmt: Siehe Abbildung 6.5.

Die zur Zeit t in Ω enthaltene Masse an Gas ist

$$M(t) = \int_{\Omega} \rho \, dV.$$

Deren Änderungsrate ist also

$$\frac{dM}{dt} = \frac{d}{dt} \int_{\Omega} \rho \, dV = \int_{\Omega} \frac{d\rho(x, t)}{dt} \, dV. \quad (6.3)$$

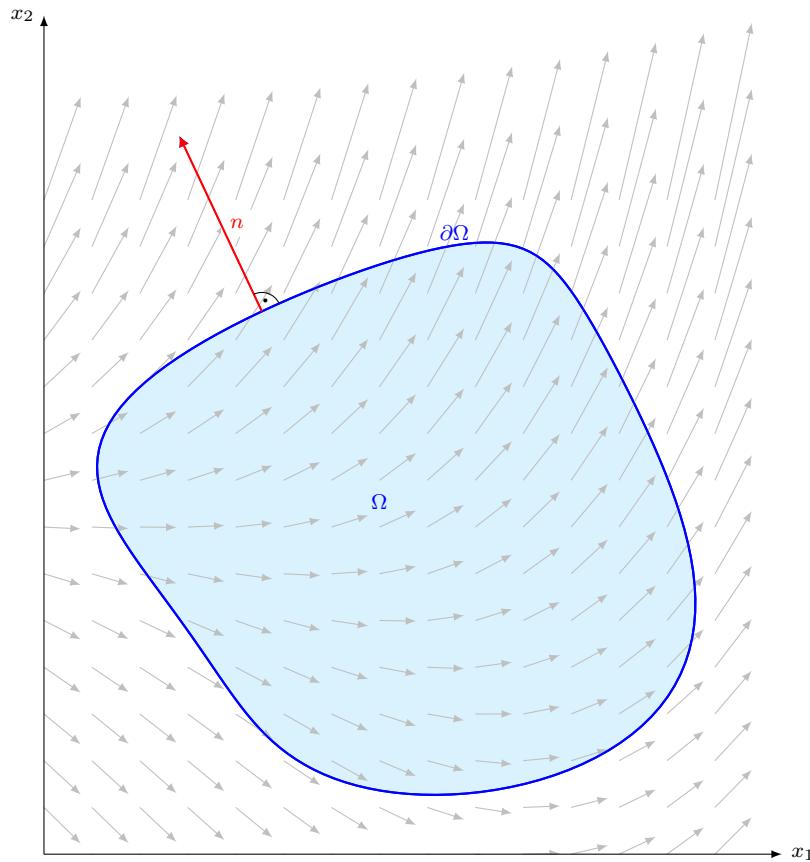


Abbildung 6.3: Blick von oben auf eine fliessende Flüssigkeit. Die Vektoren des Vektorfeldes beschreiben die Fliessgeschwindigkeit im jeweiligen Punkt.

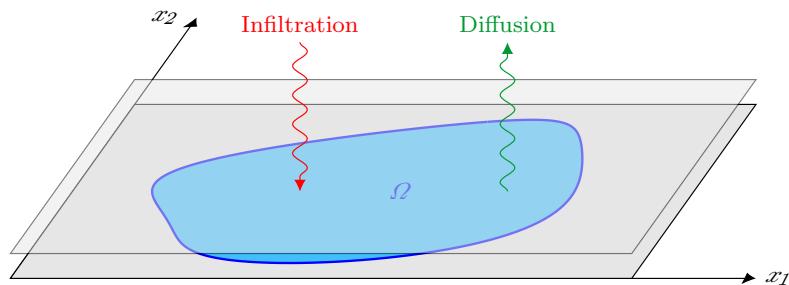


Abbildung 6.4: Fliessende Flüssigkeit zwischen zwei Platten.

Die in Ω enthaltene Masse an Gas ändert sich, wenn netto mehr oder weniger Gas in Ω hinein- als hinausfliesst. Das heisst, die Änderungsrate (6.3) der Masse auch (minus!) die netto aus Ω

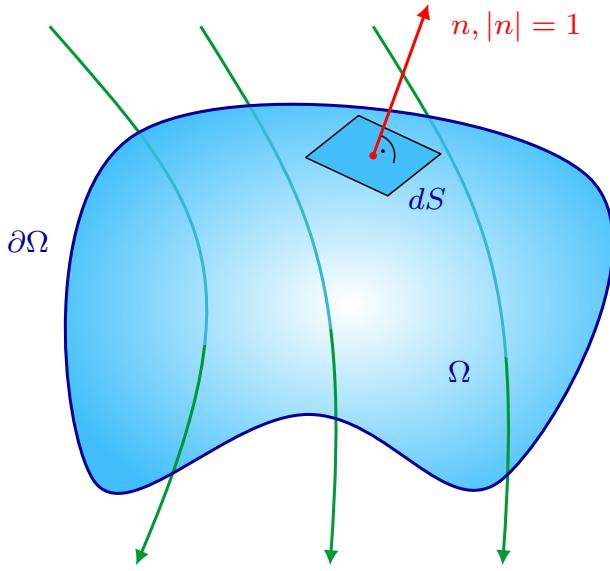


Abbildung 6.5: Ein Gas durchströmt ein Gebiet $\Omega \subset \mathbb{R}^3$. Drei grüne Stromlinien sind eingezeichnet.

ausströmende Gasmasse, nämlich

$$\frac{dM}{dt} = - \int_{\partial\Omega} \rho v \cdot n \, dS \stackrel{\text{Gauss}}{=} - \int_{\Omega} \operatorname{div}(\rho v) \, dV, \quad (6.4)$$

wobei wir wieder den Divergenzsatz von Gauss verwendet haben. Gleichsetzen von (6.3) und (6.4) liefert

$$\int_{\Omega} \left(\frac{d\rho}{dt} + \operatorname{div}(\rho v) \right) \, dV = 0.$$

Da diese Gleichung für alle Ω gilt, muss der Integrand Null sein, d.h.

$$\frac{d\rho}{dt} + \operatorname{div}(\rho v) = 0.$$

Diese Gleichung heisst **Massenerhaltungs- oder Kontinuitätsgleichung**.

6.1.3 Wärmeleitung und Diffusion

Stellt man eine Pfanne auf eine heiße Herdplatte, so breitet sich die Wärme der Platte nach und nach im Metall der Pfanne aus. Gibt man einen Tropfen Tinte in ein Glas Wasser, so verteilt sich die Tinte mit der Zeit und färbt das Wasser schliesslich gleichmässig, ohne dass man umröhrt. Beide genannten Prozesse, die Wärmeleitung und die Diffusion werden mathematisch durch dieselbe Gleichung beschrieben. Diese wollen wir nun herleiten.

Sei $u(x, t)$ die Temperatur (oder die Konzentration einer Stoffes) zur Zeit $t \geq 0$ und am Ort x in einem Festkörper $\Omega \subset \mathbb{R}^3$. Wie breitet sich die Wärme (oder der Stoff) aus?

Modellsituation: Wir betrachten einen kleinen, quaderförmigen Ausschnitt von Ω . Da der Quader klein ist, dürfen wir ∇u als annähernd konstant betrachten, und wir richten die Längsachse des Quaders in Richtung von ∇u aus. Schliesslich wählen wir auch noch das Koordinatensystem entlang der Quaderachsen, wobei die x_1 -Achse in Richtung ∇u zeigen soll (siehe Abbildung 6.6).

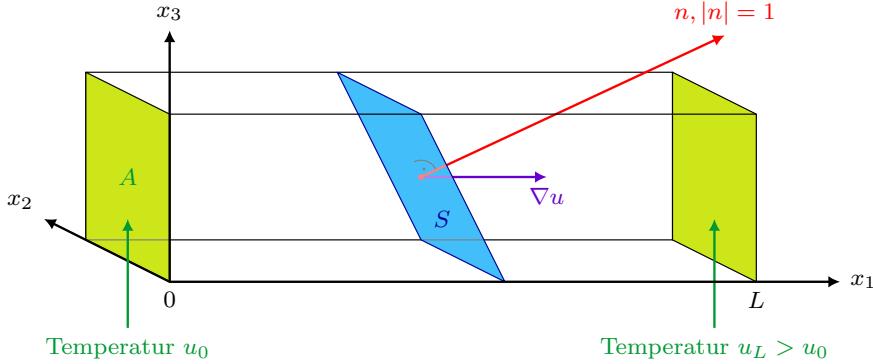


Abbildung 6.6: Wärmefluss in einer einfachen Modellsituation.

Die Wärmemenge dW , die pro Zeit dt **von rechts nach links** durch den Quader der Länge L fliessst ist proportional zur Querschnittsfläche A , zur Temperaturdifferenz $u_L - u_0$ und umgekehrt proportional zu L . Diese Annahmen sind unmittelbar plausibel und lassen sich im Experiment überprüfen. Demnach gilt

$$\frac{dW}{dt} = k \frac{u_L - u_0}{L} A. \quad (6.5)$$

Die Materialkonstante k heisst **Wärmeleitfähigkeit**. Die Richtung der Wärmeleitung ist dabei durch den zweiten Hauptsatz der Thermodynamik gegeben („von warm nach kalt“).

Wir gehen nun von der parallel zur x_3 -Achse geneigten Fläche S mit Einheitsnormalenvektor n aus. Dann gilt $A = S \cos(\alpha)$, wenn α der Winkel zwischen n und der x_1 -Achse bezeichnet. Für kleines L approximiert der Differenzenquotient die partielle Ableitung nach x_1 . Da ∇u parallel zur x_1 -Achse ist, erhalten wir daher aus (6.5)

$$\begin{aligned} \frac{dW}{dt} &= k \frac{u_L - u_0}{L} A \\ &= k \frac{u_L - u_0}{L} S \cos(\alpha) \\ &= k \frac{\partial u}{\partial x_1} S \cos(\alpha) \\ &= k \nabla u \cdot n S, \end{aligned}$$

wobei wie zuvor mit dem Punkt \cdot das Euklidische Skalarprodukt gemeint ist. Dies ist **minus** die Wärmemenge, die pro Zeiteinheit durch S in Richtung n fliessst.

Allgemein ist damit der Wärmefluss durch $\partial\Omega$ nach aussen

$$\phi(\partial\Omega) = - \int_{\partial\Omega} k \nabla u \cdot n dS$$

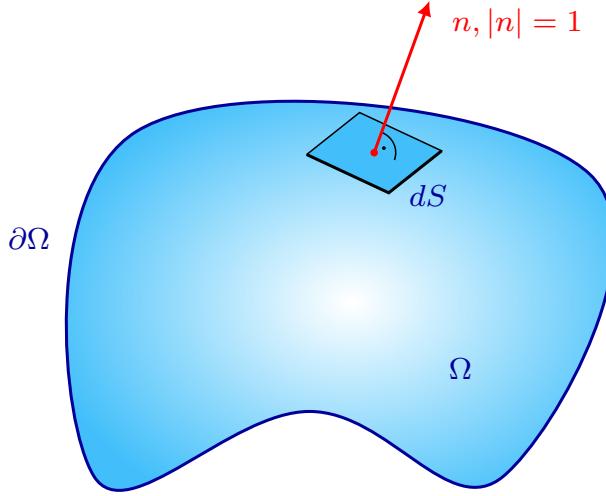


Abbildung 6.7: Wärmefluss aus einem Gebiet $\Omega \subset \mathbb{R}^3$ heraus.

(siehe Abbildung 6.7).

Die pro Masse gespeicherte Wärmemenge ist proportional zur Temperatur: Die in Ω enthaltene Wärmemenge ist daher

$$W(\Omega) = \int_{\Omega} c \rho u dV,$$

wobei c die **spezifische Wärmekapazität** pro Masseneinheit des Stoffes, und ρ seine Dichte ist. Somit gilt

$$\begin{aligned} \frac{dW(\Omega)}{dt} &= \int_{\Omega} c \frac{\partial u}{\partial t} \rho dV = -\phi(\partial\Omega) = \\ &= \int_{\partial\Omega} k \nabla u \cdot n dS \stackrel{\text{Gauss}}{=} \int_{\Omega} \operatorname{div}(k \nabla u) dV \end{aligned}$$

wobei im letzten Schritt wieder der Divergenzsatz von Gauss zum Einsatz kam. Diese Gleichung gilt für beliebiges Ω , daher müssen die roten Integranden gleich sein:

$$c \rho \frac{\partial u}{\partial t} = \operatorname{div}(k \nabla u).$$

Ist k konstant, so erhält man daraus mit $D := \frac{k}{c\rho} > 0$ die **Wärmeleitungs- oder Diffusionsgleichung**

$$\frac{\partial u}{\partial t} = D \Delta u.$$

6.2 Fourier-Methoden zur Lösung von PDEs

Die Methoden zur expliziten Lösung von partiellen Differentialgleichungen sind nicht sehr zahlreich. Zu den wichtigsten zählen Fourier-Methoden, die wir in diesem Abschnitt vorstellen. Da-

neben sind numerische Methoden im Einsatz um konkrete PDEs zu lösen, sowie analytische Methoden um allgemeine qualitative Aussagen über Lösungen zu machen. Davon wird weiter unten die Rede sein.

6.2.1 Die Wärmeleitungsgleichung

Wir beginnen mit einer einfachen Modellsituation um die Technik der Fourier-Methoden zu illustrieren.

Beispiel 6.2.1. Wärmeleitung im geschlossenen Draht. Auf der geschlossenen Drahtschleife der Länge L messen wir die Position x als Entfernung ab einem willkürlichen Referenzpunkt 0 . Zu Beginn liegt eine Hälfte des Drahts in einem Wärmebad der Temperatur 1°C , der Rest hat Temperatur 0°C . Zur Zeit $t = 0$ wird der Draht aus dem Wärmebad gehoben, und die resultierende Wärmeverteilung ist zu berechnen.

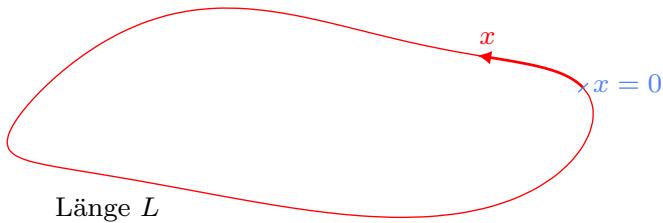


Abbildung 6.8: Wärmeverteilung in der geschlossenen Drahtschleife.

Zu lösen ist also eine partielle Differentialgleichung (PDE), nämlich die Wärmeleitungsgleichung, mit einer Anfangsbedingung (AB), nämlich die Temperaturverteilung zur Zeit $t = 0$, und einer Randbedingung (RB) welche angibt, dass der Draht sich nach der Länge L schliesst:

$$\text{PDE: } \frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2} \quad \text{für } x \in \mathbb{R}, t > 0 \quad (6.6)$$

$$\text{AB: } u(x, 0) = u_0(x) := \begin{cases} 1 & \text{für } x \in [0, \frac{L}{2}[\\ 0 & \text{für } x \in [\frac{L}{2}, L[\end{cases} \quad (6.7)$$

$$\text{RB: } u(x + L, t) = u(x, t) \quad \text{für } x \in \mathbb{R}, t \geq 0. \quad (6.8)$$

1. Lösungsschritt: Separation der Variablen. Wir suchen zunächst sehr spezielle Lösungen, welche Produkt einer Funktion vom Ort x und einer Funktion der Zeit t sind. Das heisst, wir suchen Lösungen der PDE (6.6) von der Form $u(x, t) = X(x)T(t)$. Setzen wir diesen Ansatz in (6.6) ein, erhalten wir $\dot{T}X = DX''T$, und Division durch DXT ergibt

$$\frac{1}{D} \frac{\dot{T}(t)}{T(t)} = \frac{X''(x)}{X(x)} \quad \text{für alle } x \in \mathbb{R}, t > 0. \quad (6.9)$$

Halten wir in dieser Gleichung $x =: x_0$ fest, folgt, dass $\frac{1}{D} \frac{\dot{T}(t)}{T(t)} = \frac{X''(x_0)}{X(x_0)} =: \kappa$ für alle $t > 0$ gilt. Insbesondere gilt dann für $t =: t_0$, dass $\kappa = \frac{1}{D} \frac{\dot{T}(t_0)}{T(t_0)} = \frac{X''(x)}{X(x)}$ für alle $x \in \mathbb{R}$. Das heisst, beiden

Seiten von (6.9) sind gleich einer Konstanten κ . Es gelten also die beiden ODEs

$$\dot{T} = \kappa DT \quad (6.10)$$

$$X'' = \kappa X. \quad (6.11)$$

Für $\kappa \geq 0$ besitzt (6.11) nur die periodische Lösung $X \equiv 0$. Wir setzen also $\kappa := -\omega^2 < 0$. Dann besitzt (6.11) die periodischen Lösungen

$$X(x) = A \cos(\omega x) + B \sin(\omega x).$$

Die Periode L liegt genau dann vor, wenn $\omega = \omega_n := \frac{2\pi n}{L}$ für $n \in \{0, 1, 2, \dots\}$. Das heisst, die L -periodischen Lösungen von (6.11) sind

$$X_n(x) = A_n \cos(\omega_n x) + B_n \sin(\omega_n x).$$

Entsprechend erhalten wir nun für jedes $n \in \{0, 1, 2, \dots\}$ aus (6.10) $\dot{T}_n = -D\omega_n^2 T_n$ die Lösungen

$$T_n(t) = \exp(-D\omega_n^2 t)$$

und als Lösungen von (PDE) und (RB) die Funktionen

$$\begin{aligned} u_n(x, t) &= X_n(x)T_n(t) \\ &= (A_n \cos(\omega_n x) + B_n \sin(\omega_n x)) \exp(-D\omega_n^2 t). \end{aligned}$$

Diese Lösungen nennen wir **Basislösungen**.

2. Lösungsschritt: Superposition. Die PDE (6.6) und die Randbedingung (6.8) sind linear und homogen. Daher sind Linearkombinationen und aus den Basislösungen gebildete Reihen wieder Lösungen von (6.6) und (6.8): Die Funktion

$$u(x, t) = \sum_{n=0}^{\infty} \left(A_n \cos\left(\frac{2\pi n}{L}x\right) + B_n \sin\left(\frac{2\pi n}{L}x\right) \right) \exp\left(-D\frac{4\pi^2 n^2}{L^2}t\right)$$

ist (Konvergenz der Reihe vorausgesetzt) für beliebige Wahl der A_n, B_n eine Lösung der PDE (6.6) und der Randbedingung (6.8).

Idee: Man wähle die Koeffizienten A_n, B_n so, dass auch die Anfangsbedingung (6.7) erfüllt ist! Das heisst

$$u(x, 0) = \sum_{n=0}^{\infty} \left(A_n \cos\left(\frac{2\pi n}{L}x\right) + B_n \sin\left(\frac{2\pi n}{L}x\right) \right) \stackrel{!}{=} u_0(x).$$

Dies ist offenbar gerade dann der Fall, wenn die A_n, B_n als Fourier-Koeffizienten der Funktion u_0 gewählt werden. Diese sind einfach mit den bekannten Formeln aus dem Kapitel ?? zu berechnen, und man findet

$$\begin{aligned} A_0 &= \frac{1}{2}, \quad A_n = 0 \text{ für } n \geq 1, \\ B_n &= \frac{2}{n\pi} \text{ für } n \text{ ungerade}, \quad B_n = 0 \text{ für } n \text{ gerade}. \end{aligned}$$

Setzen wir diese Koeffizienten ein, erhalten wir die Lösung

$$u(x, t) = \frac{1}{2} + \frac{2}{\pi} \sum_{\substack{n=1 \\ n \text{ ungerade}}}^{\infty} \frac{1}{n} \sin\left(\frac{2\pi n}{L}x\right) \exp\left(-D\frac{4\pi^2 n^2}{L^2}t\right).$$

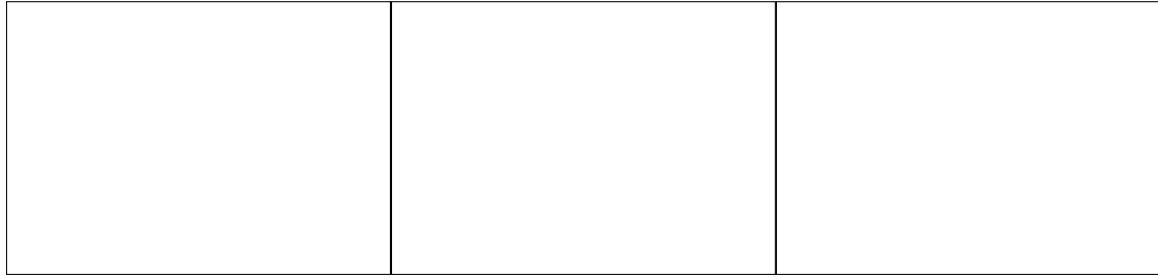


Abbildung 6.9: Die Wärmeverteilung im geschlossenen Draht (click to play).

In Abbildung 6.9 ist die Lösung für drei unterschiedlich lange Zeitintervalle als Animation zu sehen. Dabei wurden die numerischen Werte $D = L = 1$ gewählt.

Interpretation: Auf dem kurzen Zeitintervall $[0, 0.002]$ dringt die Erwärmung respektive Abkühlung kaum vor: Das Innere der Intervalle $[0, 1/2]$ und $[1/2, 1]$ behält die Temperatur 1°C respektive 0°C . Bei Verbrennungsverletzungen hängt also die Schwere (Tiefe) wesentlich von der Kontaktzeit des Gewebes mit der Hitzequelle ab. Ist der Kontakt nur kurz, ist die Eindringtiefe der Hitze gering und damit die Gewebebeschäden nur oberflächlich. Es kommt zu Verbrennungen vom Grad I oder IIa. Ist der Kontakt mit der Hitzequelle länger, gehen die Gewebebeschäden tiefer und es resultieren Verbrennungen vom Grad IIb oder III, welche operative Eingriffe erforderlich machen.

Bemerkung: An der Lösungsformel liest man ab, dass $u(x, t)$ für $t \rightarrow \infty$ gegen die stationäre (das heißt zeitunabhängige) Lösung $A_0 = \frac{1}{L} \int_0^L u_0(x) dx$ der PDE (6.6) und der Randbedingung (6.8) konvergiert. Auch ohne Kenntnis einer expliziten Lösungsformel lassen sich qualitative Informationen über die Lösung gewinnen, oft sogar einfacher, als mit einer expliziten Lösungsformel. Dazu betrachten wir einige Beispiele:

A priori Aussagen über die Lösung

- Integriert man die PDE (6.6) über x von 0 bis L , ergibt sich

$$\frac{d}{dt} \int_0^L \mathbf{u} dx = \int_0^L \mathbf{u}_t dx = \int_0^L \mathbf{u}_{xx} dx = u_x \Big|_0^L = 0,$$

wobei wir die Notation

$$\begin{aligned} u_t &:= \frac{\partial u}{\partial t}, \\ u_{xx} &:= \frac{\partial^2 u}{\partial x^2}, \end{aligned}$$

etc. benutzen. Das heißt, die **Wärmemenge im Draht** ist zeitlich konstant,

$$\int_0^L u(x, t) dx = \text{konst.}, \quad (6.12)$$

was physikalisch plausibel ist, da weder Wärme zu- noch abgeführt wird.

- Multipliziert man die PDE (6.6) mit u und integriert dann über x von 0 bis L , ergibt sich

$$\frac{1}{2} \frac{d}{dt} \int_0^L u^2 dx = \int_0^L uu_t dx = \int_0^L u u_{xx} dx \stackrel{\substack{\text{partielle} \\ \text{Integration}}}{=} - \int_0^L u_x^2 dx \leq 0.$$

Das heisst, das Integral $\int_0^L u^2(x, t) dx$ ist monoton fallend in t .

- Auf der Kelvin-Skala sind Temperaturen u wegen des dritten Hauptsatzes der Thermodynamik immer strikt grösser als 0. Also ist der Logarithmus $\ln u$ wohldefiniert. Multipliziert man nun die PDE (6.6) mit $\ln u$ und integriert dann über x von 0 bis L , ergibt sich

$$\begin{aligned} \frac{d}{dt} \int_0^L u \ln u dx &= \frac{d}{dt} \int_0^L (u \ln u - u) dx = \\ &= \int_0^L u_t \ln u dx = \int_0^L u_{xx} \ln u dx \stackrel{\substack{\text{partielle} \\ \text{Integration}}}{=} - \int_0^L \frac{u_x^2}{u} dx \leq 0, \end{aligned}$$

wobei wir im ersten Schritt (6.12) verwendet haben. Das heisst, die sogenannte **Entropie**

$$-\int_0^L u(x, t) \ln u(x, t) dx$$

ist in der Zeit t monoton wachsend. Das ist der zweite Hauptsatz der Thermodynamik.

- Betrachten wir $v(x, t) := u(x, t) - \varepsilon t$ (für $\varepsilon > 0$), und nehmen an, v nehme zur Zeit t_0 in x_0 ein Maximum an, d. h.

$$v(x_0, t_0) = \max_{x \in \mathbb{R}} v(x, t_0).$$

Dann gilt

$$0 \geq v_{xx}(x_0, t_0) = u_{xx}(x_0, t_0) = u_t(x_0, t_0) = v_t(x_0, t_0) + \varepsilon,$$

und somit $-\varepsilon \geq v_t(x_0, t_0)$. Aufgrund von Stetigkeit gilt dann

$$-\frac{\varepsilon}{2} \geq v_t(x, t)$$

in einer ganzen Umgebung von (x_0, t_0) . Insbesondere gilt für $t < t_0$ nahe genug bei t_0 ,

$$\max_{x \in \mathbb{R}} v(x, t) \geq v(x_0, t) > v(x_0, t_0) = \max_{x \in \mathbb{R}} v(x, t_0).$$

Da diese Überlegung unabhängig von t_0 gemacht wurde, ist $\max_{x \in \mathbb{R}} v(x, t)$ für alle $t \geq 0$ strikt monoton fallend. Im Limes $\varepsilon \searrow 0$ folgt

$$\max_{x \in \mathbb{R}} u(x, t) \text{ ist monoton fallend für } t \geq 0.$$

Dies ist ein Spezialfall des sogenannten **Maximumprinzips** für die Wärmeleitungsgleichung. Wir werden diesem Prinzip später bei der Diskussion der Potentialgleichung im Theorem 6.2.4 nochmal begegnen.

Betrachtet man $-u$ statt u , so folgt das entsprechende **Minimumprinzip**:

$$\min_{x \in \mathbb{R}} u(x, t) \text{ ist monoton wachsend für } t \geq 0.$$

Das Maximumprinzip besagt also, dass sich bei der Wärmeleitung (und entsprechend bei der Diffusion) Wärme nicht lokal konzentrieren kann, ohne dass dort geheizt wird. Ebenso ist wegen des Minimumprinzips spontane lokale Abkühlung nicht möglich. Dies steht im Gegensatz zu Wellenphänomenen, wo sich Wellen lokal aufschaukeln können, was etwa bei der Nierensteinzertrümmerung durch Schall explizit ausgenutzt wird.

6.2.2 Die Laplace-Gleichung

Eine stationäre, also nicht von der Zeit abhängige Lösung der Wärmeleitungsgleichung erfüllt die **Laplace-** oder **Potentialgleichung**

$$\Delta u = 0.$$

Wegen $\Delta u = \operatorname{div} \nabla u$ ist dann eine Lösung u das Potential des divergenzfreien Vektorfeldes ∇u . Elektrostatische Felder, oder Gravitationsfelder sind von dieser Art. Lösungen der Laplace-Gleichung heissen **harmonische Funktionen**.

Als Beispiel suchen wir die stationäre Lösung der Wärmeleitungsgleichung auf einer Kreisscheibe B vom Radius r_0 , also die Wärmeverteilung auf B , die sich nach langer Zeit einstellt, wenn auf dem Rand eine bestimmte Temperatur vorgegeben ist. Dazu statten wir B mit Polarkoordinaten aus: Der Punkt P mit den kartesischen Koordinaten (x, y) wird dabei durch die Polarkoordinaten (r, φ) beschrieben (siehe Abbildung 6.10).

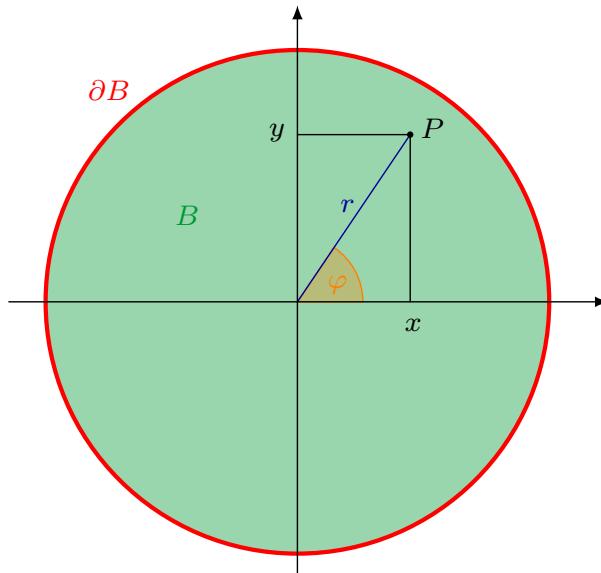


Abbildung 6.10: Die Kreisscheibe B und Polarkoordinaten.

Es ist dann allerdings nötig, den Laplace-Operator ebenfalls in Polarkoordinaten auszudrücken:

Theorem 6.2.2. *In Polarkoordinaten (r, φ) lautet der Laplace-Operator*

$$\Delta u = u_{rr} + \frac{1}{r}u_r + \frac{1}{r^2}u_{\varphi\varphi}.$$

Beweis. In Polarkoordinaten gilt $x = r \cos \varphi$, $y = r \sin \varphi$ (siehe Abbildung 6.10). Dann liefert die Kettenregel:

$$\begin{aligned} u_r &= u_x \cos \varphi + u_y \sin \varphi \\ u_\varphi &= -u_x r \sin \varphi + u_y r \cos \varphi. \end{aligned}$$

Dividiert man die zweite Gleichung durch r , ist die resultierende Gleichungsmatrix orthogonal, also lesen wir ab:

$$\begin{aligned} u_x &= u_r \cos \varphi - \frac{1}{r} u_\varphi \sin \varphi \\ u_y &= u_r \sin \varphi + \frac{1}{r} u_\varphi \cos \varphi. \end{aligned}$$

Zweimaliges Anwenden dieser Regel für die Ableitung nach x ergibt

$$u_{xx} = \cos \varphi \frac{\partial}{\partial r} (u_r \cos \varphi - \frac{1}{r} u_\varphi \sin \varphi) - \frac{1}{r} \sin \varphi \frac{\partial}{\partial \varphi} (u_r \cos \varphi - \frac{1}{r} u_\varphi \sin \varphi).$$

Analog berechnet man u_{yy} . Daraus ergibt sich die Behauptung durch Addition: $\Delta u = u_{xx} + u_{yy}$. \square

Wir suchen also eine Funktion $u(r, \varphi)$, $0 \leq r \leq r_0$, die 2π -periodisch in φ ist, so dass

$$\text{PDE: } \Delta u = u_{rr} + \frac{1}{r} u_r + \frac{1}{r^2} u_{\varphi\varphi} = 0 \quad \text{in } B \quad (6.13)$$

$$\text{RB: } u(r_0, \varphi) = \xi(\varphi) \quad \text{auf } \partial B. \quad (6.14)$$

Dabei beschreibt ξ die gegebene Temperatur am Rand des Kreises.

1. Lösungsschritt: Separation der Variablen. Wir machen den **Ansatz**: $u(r, \varphi) = f(r)g(\varphi)$. Einsetzen in die PDE liefert

$$f''g + \frac{1}{r} f'g + \frac{1}{r^2} fg'' = 0,$$

wobei ' die Ableitung nach r , und ' die Ableitung nach φ bezeichnen. Nach Multiplikation mit $\frac{r^2}{fg}$ ergibt sich

$$\frac{1}{f} (r^2 f'' + rf') = -\frac{g''}{g} =: \omega^2.$$

Da links eine Funktion von r und rechts eine von φ steht, müssen beide Seiten eine Konstante ω^2 sein (wobei wir das Vorzeichen bereits geeignet gewählt haben). Es sind also zwei ODEs zu lösen:

$$g'' = -\omega^2 g, \quad r^2 f'' + rf' - \omega^2 f = 0.$$

Die Lösungen der Gleichung links sind

$$g(\varphi) = A \cos(\omega \varphi) + B \sin(\omega \varphi).$$

Da nur 2π -periodische Lösungen infrage kommen, ist $\omega = \omega_n := n \in \mathbb{N}_0$, d.h.

$$g_n(\varphi) = A_n \cos(n\varphi) + B_n \sin(n\varphi).$$

Die rechte Gleichung

$$r^2 f''_n + rf'_n - n^2 f_n = 0$$

(eine Eulersche Differentialgleichung) löst man mit dem Ansatz $f_n(r) = r^\alpha$. Einsetzen liefert

$$\alpha(\alpha - 1) + \alpha - n^2 = 0,$$

also $\alpha = \pm n$. Da wir keine Singularitäten wollen, kommt nur $\alpha = n$ infrage, also $f_n(r) = r^n$.

Damit erhalten wir für $n \in \mathbb{N}_0$ die Basislösungen

$$u_n(r, \varphi) = f_n(r)g_n(\varphi) = r^n (A_n \cos(n\varphi) + B_n \sin(n\varphi)).$$

2. Lösungsschritt: Superposition der Basislösungen. Da die PDE (6.13) linear und homogen ist, ist auch

$$u(r, \varphi) = A_0 + \sum_{n=1}^{\infty} r^n (A_n \cos(n\varphi) + B_n \sin(n\varphi))$$

eine Lösung (Konvergenz der Reihe vorausgesetzt).

Idee: Wähle die Koeffizienten A_n, B_n so, dass die Randbedingung erfüllt ist, das heisst

$$\begin{aligned} u(r_0, \varphi) &= A_0 + \sum_{n=1}^{\infty} r_0^n (A_n \cos(n\varphi) + B_n \sin(n\varphi)) \stackrel{!}{=} \xi(\varphi) \\ &= \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos(n\varphi) + b_n \sin(n\varphi)), \end{aligned}$$

wobei auf der unteren Zeile die Fourier-Reihe von ξ steht. Durch Koeffizientenvergleich finden wir

$$A_0 = \frac{a_0}{2}, \quad A_n = \frac{a_n}{r_0^n}, \quad B_n = \frac{b_n}{r_0^n}.$$

Somit lautet die endgültige Lösung

$$u(r, \varphi) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \frac{r^n}{r_0^n} (a_n \cos(n\varphi) + b_n \sin(n\varphi)). \quad (6.15)$$

So erhalten wir zum Beispiel für

$$\xi(\varphi) := \begin{cases} 1 & \text{für } 0 \leq \varphi < \pi \\ 0 & \text{für } \pi \leq \varphi < 2\pi \end{cases}$$

die Lösung

$$u(r, \varphi) = \frac{1}{2} + \frac{2}{\pi} \sum_{\substack{n=1 \\ n \text{ ungerade}}}^{\infty} \frac{r^n}{n} \sin(n\varphi).$$

In der Abbildung 6.11 ist die Lösung dargestellt: Die Temperatur $u(r, \varphi)$ ist als Graph über der Kreisscheibe B dargestellt. Der Graph ist so eingefärbt, dass **rot** heiss ($u = 1$), und **blau** kalt ($u = 0$) andeutet.

Wir formen die allgemeine Lösung (6.15) noch um, indem wir die Fourier-Koeffizienten einsetzen:

$$\begin{aligned} u(r, \varphi) &= \frac{a_0}{2} + \sum_{n=1}^{\infty} \frac{r^n}{r_0^n} (a_n \cos(n\varphi) + b_n \sin(n\varphi)) \\ &= \frac{1}{2} \int_0^{2\pi} \xi(t) dt + \sum_{n=1}^{\infty} \frac{1}{\pi} \left(\int_0^{2\pi} \xi(t) \cos(nt) dt \cos(n\varphi) + \int_0^{2\pi} \xi(t) \sin(nt) dt \sin(n\varphi) \right) \left(\frac{r}{r_0} \right)^n \\ &= \frac{1}{2\pi} \int_0^{2\pi} \xi(t) \left[1 + 2 \sum_{n=1}^{\infty} (\cos(nt) \cos(n\varphi) + \sin(nt) \sin(n\varphi)) \left(\frac{r}{r_0} \right)^n \right] dt. \end{aligned}$$

Das Additionstheorem für den Kosinus liefert für den roten Ausdruck

$$\cos(nt) \cos(n\varphi) + \sin(nt) \sin(n\varphi) = \cos(n(t - \varphi)).$$

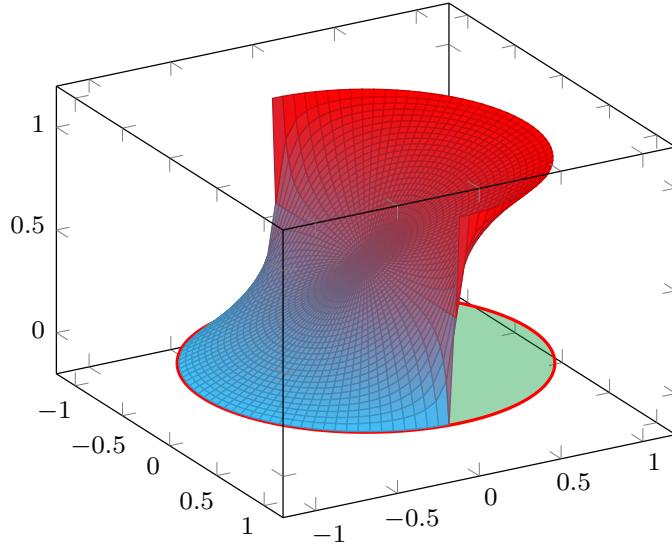


Abbildung 6.11: Die Lösung der Laplace-Gleichung auf der Kreisscheibe B : Auf der vertikalen Achse ist die Temperatur u abgetragen.

Die eckige Klammer lässt sich dann mit Hilfe der Formel für die geometrische Reihe umformen:

$$\begin{aligned} [\dots] &= \operatorname{Re} \left(1 + 2 \sum_{n=1}^{\infty} \left(\frac{r}{r_0} e^{i(t-\varphi)} \right)^n \right) \\ &= \operatorname{Re} \left(1 + 2 \frac{\frac{r}{r_0} e^{i(t-\varphi)}}{1 - \frac{r}{r_0} e^{i(t-\varphi)}} \right) \\ &= \operatorname{Re} \frac{r_0 + r e^{i(t-\varphi)}}{r_0 - r e^{i(t-\varphi)}} = \frac{r_0^2 - r^2}{r_0^2 - 2r r_0 \cos(t - \varphi) + r_0^2}. \end{aligned}$$

Setzen wir dies oben für die eckige Klammer ein, ergibt sich die Lösung ohne Umweg über die Fourier-Reihe von ξ :

$$u(r, \varphi) = \frac{1}{2\pi} \int_0^{2\pi} \xi(t) \frac{r_0^2 - r^2}{r_0^2 - 2r r_0 \cos(t - \varphi) + r_0^2} dt.$$

Dies ist die **Formel von Poisson**.

Für $r = 0$ ergibt sich insbesondere

$$u(0, \varphi) = \frac{1}{2\pi} \int_0^{2\pi} \xi(t) dt.$$

Das lässt sich als Satz formulieren.

Theorem 6.2.3 (Mittelwerteigenschaft harmonischer Funktionen). *Sei $u : \Omega \rightarrow \mathbb{R}$ in einem Gebiet Ω harmonisch (das heisst $\Delta u = 0$). Sei B eine Kreisscheibe mit Zentrum z , die samt Rand ∂B in Ω liegt. Dann ist $u(z)$ der Mittelwert der Werte von u auf ∂B .*

Bemerkung: Die Mittelwerteigenschaft gilt in jeder Dimension. Man überlege sich, was die Mittelwerteigenschaft in $n = 1$ Dimension bedeutet.

Aus der Mittelwerteigenschaft folgt das folgende Theorem.

Theorem 6.2.4 (Maximumprinzip für harmonische Funktionen). *Nimmt eine harmonische Funktion u in einem inneren Punkt x_0 ihres zusammenhängenden Definitionsbereichs ein Maximum an, so ist u eine Konstante.*

Beweis. Wir nehmen an, dass u nicht konstant ist. Also existiert ein $x \in \Omega$ mit $u(x) < u(x_0) := M$. Da Ω zusammenhängend ist, können wir x_0 mit x durch einen Weg γ verbinden, der in Ω verläuft. Sei P der letzte Punkt auf γ , in dem u den Wert M annimmt und K ein Kreis um P in Ω (siehe Abbildung 6.12).

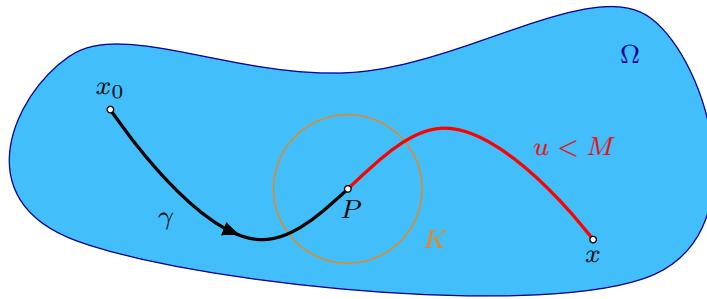


Abbildung 6.12: Zum Beweis des Maximumprinzips.

Dann ist $u \leq M$ auf K , und $u < M$ in der Nähe des Punktes, wo γ den Kreis K verlässt. Aus der Mittelwerteigenschaft folgt dann $M = u(P) > M$. Widerspruch. \square

Bemerkungen: Auch das Maximumprinzip gilt in beliebigen Dimensionen. Durch Betrachten von $-u$ folgt ein entsprechendes Minimumprinzip.

Interpretation: Eine stationäre Temperaturverteilung oder eine stationäre Konzentration eines Diffusionsprozesses nimmt immer in Randpunkten den maximalen und den minimalen Wert an. Wenn ein Medikament sich in einem Organ durch Diffusion verteilt, so gilt die entsprechende Aussage auch für diese Situation.

6.2.3 Das Fourier-Integral

Mithilfe von Fourier-Reihen konnten wir periodische Funktionen als Überlagerung von Sinus- und Kosinusfunktionen darstellen. Dies möchten wir nun ausdehnen auf nicht-periodische Funktionen, die auf ganz \mathbb{R} definiert sind. Dazu stellen wir $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ zunächst auf dem Intervall $[-\frac{T}{2}, \frac{T}{2}]$ als

Fourier-Reihe dar, mit dem Hintergedanken, T anschliessend gegen Unendlich streben zu lassen:

$$\begin{aligned}\varphi(x) &= \sum_{k \in \mathbb{Z}} \textcolor{blue}{c_k} e^{ik\frac{2\pi}{T}x} \\ &= \sum_{k \in \mathbb{Z}} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \varphi(y) e^{-ik\frac{2\pi}{T}y} dy e^{ik\frac{2\pi}{T}x} \\ &= \frac{1}{2\pi} \int_{-\frac{T}{2}}^{\frac{T}{2}} \varphi(y) \frac{2\pi}{T} \sum_{k \in \mathbb{Z}} e^{ik\frac{2\pi}{T}(x-y)} dy.\end{aligned}$$

Der rote Ausdruck ist die Riemann-Summe für das Integral

$$\int_{-\infty}^{\infty} e^{i\lambda(x-y)} d\lambda.$$

Für $T \rightarrow \infty$ ergibt sich somit

$$\varphi(x) = \int_{-\infty}^{\infty} \underbrace{\left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(y) e^{-i\lambda y} dy \right)}_{=: \hat{\varphi}(\lambda)} e^{i\lambda x} d\lambda.$$

Interpretation: Die Funktion φ wird als Superposition von (komplexen) Schwingungen $e^{i\lambda x}$ der Kreisfrequenz λ dargestellt, wobei $\hat{\varphi}(\lambda)$ deren Amplitude angibt. Dies können wir als Theorem formulieren:

Theorem 6.2.5. *Sei φ eine auf \mathbb{R} stückweise glatte Funktion, für die*

$$\int_{-\infty}^{\infty} |\varphi(x)| dx < \infty$$

gilt. Dann ist die Fourier-Transformierte

$$\hat{\varphi}(\lambda) := \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(y) e^{-i\lambda y} dy$$

für alle $\lambda \in \mathbb{R}$ definiert, und falls φ in $x \in \mathbb{R}$ stetig ist, gilt

$$\varphi(x) = \int_{-\infty}^{\infty} \hat{\varphi}(\lambda) e^{i\lambda x} d\lambda.$$

Statt für jede Funktion φ bei der Berechnung ihrer Fourier-Transformierten auf die Definition zurückzugreifen, existieren einige praktische Rechenregeln.

Theorem 6.2.6 (Rechenregeln).

- Für Konstanten $a, b \in \mathbb{C}$, gilt $\widehat{a\varphi + b\psi} = a\hat{\varphi} + b\hat{\psi}$.
- $\widehat{\varphi'}(\lambda) = i\lambda\hat{\varphi}(\lambda)$.

Beweis. Die erste Regel folgt sofort aus der Linearität des Integrals. Die zweite folgt durch partielle Integration:

$$\begin{aligned}\widehat{\varphi}'(\lambda) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \underset{\uparrow}{\varphi'(y)} e^{-i\lambda y} dy \\ &= \underbrace{\frac{1}{2\pi} \varphi(y) e^{-i\lambda y}}_{=0} \Big|_{-\infty}^{\infty} + i\lambda \underbrace{\frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(y) e^{-i\lambda y} dy}_{=\widehat{\varphi}(\lambda)}.\end{aligned}$$

Der erste Term verschwindet, denn aus

$$\int_{-\infty}^{\infty} |\varphi(y)| dy < \infty$$

folgt, dass für geeignete Folgen $x_i \rightarrow \infty, y_i \rightarrow -\infty$, gilt

$$\lim_{i \rightarrow \infty} \varphi(x_i) = \lim_{i \rightarrow \infty} \varphi(y_i) = 0.$$

Damit ist die Formel bewiesen. \square

Merke: Die letztgenannte Regel erlaubt es also, mit Hilfe der Fourier-Transformation eine Ableitung in eine Multiplikation zu verwandeln! Diese Beobachtung ist spektakulär, denn sie erlaubt es, eine Differentialgleichung in eine algebraische Gleichung zu verwandeln.

Die Fourier-Transformation einer Summe von zwei Funktionen ist also die Summe ihrer Fourier-Transformierten. Wie steht es mit der Fourier-Transformation eines Produkts von zwei Funktionen? Bevor wir diese Frage beantworten können benötigen wir noch eine Definition.

Definition 6.2.7. Seien $f, g : \mathbb{R} \rightarrow \mathbb{R}$ zwei Funktionen mit

$$\int_{-\infty}^{\infty} |f(x)| dx < \infty, \quad \int_{-\infty}^{\infty} |g(x)| dx < \infty.$$

Dann ist die **Faltung** von f und g definiert als

$$(f * g)(x) := \int_{-\infty}^{\infty} f(y) g(x-y) dy.$$

Theorem 6.2.8 (Weitere Rechenregeln, ohne Beweis).

- $F(x) := f(a(x-d))$ für $a, d \in \mathbb{R} \implies \hat{F}(\lambda) = \frac{1}{a} e^{-i\lambda d} \hat{f}\left(\frac{\lambda}{a}\right)$.
- $F(x) = e^{idx} f(x)$ für $d \in \mathbb{R} \implies \hat{F}(\lambda) = \hat{f}(\lambda - d)$.
- $F(x) := (f * g)(x) \implies \hat{F}(\lambda) = 2\pi \hat{f}(\lambda) \hat{g}(\lambda)$.
- $F(x) := f(x)g(x) \implies \hat{F}(\lambda) = (\hat{f} * \hat{g})(\lambda)$.
- $\hat{f}(x) = \frac{1}{2\pi} f(-x)$.
- $F(x) := \int_a^x f(y) dy \implies \hat{F}(\lambda) = \frac{1}{i\lambda} \hat{f}(\lambda)$, falls F die Voraussetzungen im Theorem 6.2.5 erfüllt.

- Bei der Fourier-Transformation nach x gilt für eine Funktion $f : (x, t) \mapsto f(x, t)$, die noch von einem Parameter t abhängt: $\widehat{f}_t = (\hat{f})_t$. D.h. die Ableitung nach t vertauscht mit der Fourier-Transformation bezüglich x .

Oben haben wir partielle Differentialgleichungen mithilfe von Fourier-Reihen gelöst, wenn periodische Rand- oder Anfangsbedingungen im Spiel waren. Mit den Fourier-Integralen lassen sich nun PDEs lösen, wo keine periodischen Bedingungen bestehen.

6.2.4 Die Wellengleichung auf \mathbb{R}

Wir betrachten die Wellengleichung für die Funktion $u(x, t)$, $x, t \in \mathbb{R}$. Die Funktion $u(x, t)$ beschreibt die Amplitude der Welle am Ort x zur Zeit t . Schall oder elektromagnetische Wellen mit Ausbreitungsgeschwindigkeit $c > 0$ werden auf diese Weise modelliert:

$$\text{PDE: } u_{tt} = c^2 u_{xx} \quad \text{für } x, t \in \mathbb{R} \quad (6.16)$$

$$\text{AB}_1: \quad u(x, 0) = \varphi(x) \quad \text{für } x \in \mathbb{R} \quad (6.17)$$

$$\text{AB}_2: \quad u_t(x, 0) = \psi(x) \quad \text{für } x \in \mathbb{R}. \quad (6.18)$$

Die PDE (6.6) heisst **Wellengleichung**. Dabei beschreibt die Anfangsbedingung (6.17) die Anfangsamplitude der Welle (die Form), und (6.18) die Anfangsgeschwindigkeit.

Idee: Wir Fourier-transformieren das Problem bezüglich der Variable x (t wird dabei als Parameter aufgefasst):

$$\widehat{\text{PDE}} : \quad \hat{u}_{tt}(\lambda, t) = (i\lambda)^2 c^2 \hat{u}(\lambda, t) \quad \text{für } \lambda, t \in \mathbb{R} \quad (6.19)$$

$$\widehat{\text{AB}}_1 : \quad \hat{u}(\lambda, 0) = \hat{\varphi}(\lambda) \quad \text{für } \lambda \in \mathbb{R} \quad (6.20)$$

$$\widehat{\text{AB}}_2 : \quad \hat{u}_t(\lambda, 0) = \hat{\psi}(\lambda) \quad \text{für } \lambda \in \mathbb{R}. \quad (6.21)$$

Für festes λ ist (6.19) eine ODE in t mit der Lösung

$$\hat{u}(\lambda, t) = A(\lambda) \cos(c\lambda t) + B(\lambda) \sin(c\lambda t).$$

Aus (6.20) und (6.21) folgt

$$\begin{aligned} \hat{u}(\lambda, 0) &= A(\lambda) \stackrel{!}{=} \hat{\varphi}(\lambda) \\ \hat{u}_t(\lambda, 0) &= c\lambda B(\lambda) \stackrel{!}{=} \hat{\psi}(\lambda). \end{aligned}$$

Also $A(\lambda) = \hat{\varphi}(\lambda)$ und $B(\lambda) = \frac{\hat{\psi}(\lambda)}{c\lambda}$, und somit

$$\hat{u}(\lambda, t) = \hat{\varphi}(\lambda) \cos(c\lambda t) + \frac{\hat{\psi}(\lambda)}{c\lambda} \sin(c\lambda t).$$

Damit kennt man die Fourier-Transformierte der Lösung. Also finden wir mit dem Theorem 6.2.5 die Lösung selber:

$$u(x, t) = \int_{-\infty}^{\infty} \left(\hat{\varphi}(\lambda) \cos(c\lambda t) + \frac{\hat{\psi}(\lambda)}{c\lambda} \sin(c\lambda t) \right) e^{i\lambda x} d\lambda.$$

In Abbildung 6.13 sind zwei animierte Beispiele zu sehen (click to play).

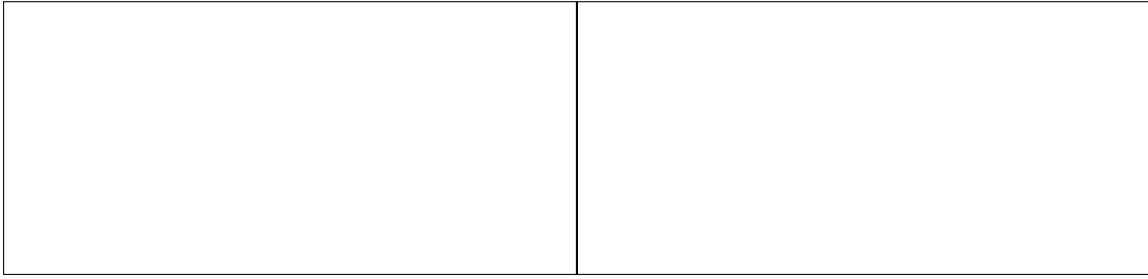


Abbildung 6.13: Links sind zwei gegeneinander laufende Wellen zu sehen, die sich beim Aufeinandertreffen addieren. Diese Eigenschaft von Wellen wird beim Nierensteinzertrümmerer ausgenutzt: Wo sich der Stein befindet werden Starke Schallwellen geeigneter Frequenz aus verschiedenen Richtungen konzentriert. Rechts ist eine zerfließende Welle zu sehen. Die Welle breitet sich nach beiden Seiten aus, während im Zentrum, so sie startete, allmählich Ruhe einkehrt. Dies ist eine Eigenschaft, welche Wellen in allen ungeraden Dimensionen haben (siehe Abschnitt 6.2.5).

6.2.5 Die Methode von d'Alembert

Wir betrachten nochmals die Wellengleichung. In beliebig vielen Raumdimensionen lautet sie

$$u_{tt} = c^2 \Delta u. \quad (6.22)$$

Wir wollen insbesondere verstehen, wie die Konstante $c > 0$ als Ausbreitungsgeschwindigkeit interpretiert werden kann. Dabei ist $u(x, t)$ eine Funktion der Zeit t und des Ortes $x \in \mathbb{R}^n$, $n = 1, 2, \dots$. Der Δ -Operator wirkt, wie bei der Wärmeleitungsgleichung, nur auf die Ortskoordinaten x . Die Gleichung (6.22) beschreibt ganz allgemein die Ausbreitung von Wellen. Ein Beispiel ist der Schall. Im medizinischen Bereich kommt er als Ultraschall bei kardiologischen Untersuchungen und in der Pränataldiagnostik zum Einsatz. In Gasen und Flüssigkeiten breiten sich Druckunterschiede in Form von Longitudinalwellen im Medium aus: u ist dann der Druck im Medium. In Festkörpern kann sich Schall auch als Transversalwelle ausbreiten, wobei die elastische Deformierung des Materials senkrecht zur Ausbreitungsrichtung schwingt. In diesem Fall beschreibt u die Auslenkung des Materials aus der Ruhelage. Im Vakuum kann sich Schall nicht ausbreiten. Elektromagnetische Wellen kommen hingegen auch ohne Trägermedium aus. In diesem Fall ist u ein Vektor: die elektrische oder magnetische Feldstärke.

Im Gegensatz zur Wärmeleitungsgleichung, die einen Ausgleich der Anregung herbeiführt, bewirkt die Wellengleichung eine Ausbreitung der Anregung, und zwar ist die Ausbreitungsgeschwindigkeit, wie wir sehen werden, gerade c .

Wir behandeln den Lösungsansatz von d'Alembert. Betrachten wir dazu ein beidseitig unendlich langes Rohr, das mit einem Gas der Temperatur T gefüllt ist. Zur Zeit $t = 0$ erfolgt in einem auf das Intervall $[a, b]$ begrenzten Teil des Rohres eine Anregung, zum Beispiel eine Explosion. Ist $u(x, t)$ die Abweichung vom Normaldruck, so genügt u der eindimensionalen Wellengleichung

$$u_{tt} = c^2 u_{xx}, \quad x \in \mathbb{R}, t > 0, \quad (6.23)$$

$c^2 = \gamma RT$, wobei R die Gaskonstante und γ den adiabatischen Exponenten bezeichnet. Gesucht ist nun eine Lösung von (6.23), die überdies den Anfangsbedingungen

$$u(x, 0) = \varphi(x), \quad u_t(x, 0) = \psi(x) \quad (6.24)$$

genügt. Randbedingungen gibt es nicht. Nach d'Alembert führen wir nun neue Koordinaten ξ, η ein,

$$\begin{cases} \xi &:= x + ct, \\ \eta &:= x - ct, \end{cases} \quad (6.25)$$

beziehungsweise umgekehrt

$$\begin{cases} x &= \frac{1}{2}(\xi + \eta), \\ t &= \frac{1}{2c}(\xi - \eta). \end{cases}$$

Auf diese Weise erhalten wir die neue unbekannte Funktion

$$\tilde{u}(\xi, \eta) := u\left(\frac{1}{2}(\xi + \eta), \frac{1}{2c}(\xi - \eta)\right).$$

Welche PDE erfüllt nun \tilde{u} ? Es gilt

$$\begin{aligned} u_t &= c(\tilde{u}_\xi - \tilde{u}_\eta), & u_{tt} &= c^2(\tilde{u}_{\xi\xi} - 2\tilde{u}_{\xi\eta} + \tilde{u}_{\eta\eta}), \\ u_x &= \tilde{u}_\xi + \tilde{u}_\eta, & u_{xx} &= \tilde{u}_{\xi\xi} + 2\tilde{u}_{\xi\eta} + \tilde{u}_{\eta\eta}. \end{aligned}$$

Somit haben wir folgenden Zusammenhang zwischen u und \tilde{u} :

$$u_{tt} = c^2 u_{xx} \iff \tilde{u}_{\xi\eta} = 0. \quad (6.26)$$

Aus $\frac{\partial}{\partial \xi} \left(\frac{\partial \tilde{u}}{\partial \eta} \right) = 0$ folgt, dass $\frac{\partial \tilde{u}}{\partial \eta}$ nicht von ξ abhängen darf, also eine Funktion allein von η sein muss, das heisst

$$\frac{\partial \tilde{u}}{\partial \eta} = G'(\eta) \quad (6.27)$$

für eine zu bestimmende Funktion G . Nun integrieren wir noch (6.27) und erhalten

$$\tilde{u}(\xi, \eta) = G(\eta) + F(\xi). \quad (6.28)$$

Man beachte dabei, dass bei der Integration nach η die Integrationskonstante von der Variablen ξ abhängen darf. Wir haben daher an ihre Stelle eine Funktion $F(\xi)$ gesetzt.

Damit haben wir die allgemeine Lösung von (6.23) gefunden:

$$u(x, t) = F(x + ct) + G(x - ct).$$

Wir müssen noch F und G so bestimmen, dass die Anfangsbedingungen (6.24) erfüllt sind:

$$F(x) + G(x) = \varphi(x), \quad (6.29)$$

$$c F'(x) - c G'(x) = \psi(x). \quad (6.30)$$

Integrieren wir (6.30) von 0 bis x , so ergibt sich

$$F(x) - G(x) - (F(0) - G(0)) = \frac{1}{c} \int_0^x \psi(y) dy.$$

Wir dürfen noch annehmen, dass $F(0) = G(0)$ ist und erhalten

$$F(x) - G(x) = \frac{1}{c} \int_0^x \psi(y) dy. \quad (6.31)$$

Aus (6.29) und (6.31) folgt nun

$$F(x) = \frac{1}{2}(\varphi(x) + \frac{1}{c} \int_0^x \psi(y) dy)$$

$$G(x) = \frac{1}{2}(\varphi(x) - \frac{1}{c} \int_0^x \psi(y) dy)$$

und somit

$$u(x, t) = \frac{1}{2} [\varphi(x + ct) + \varphi(x - ct) + \frac{1}{c} \left(\int_0^{x+ct} \psi(y) dy - \int_0^{x-ct} \psi(y) dy \right)]$$

$$= \frac{1}{2} (\varphi(x + ct) + \varphi(x - ct)) + \frac{1}{2c} \int_{x-ct}^{x+ct} \psi(y) dy. \quad (6.32)$$

Um dieses Resultat zu diskutieren, betrachten wir zunächst den Fall $\psi \equiv 0$ (siehe Abbildung 6.14).

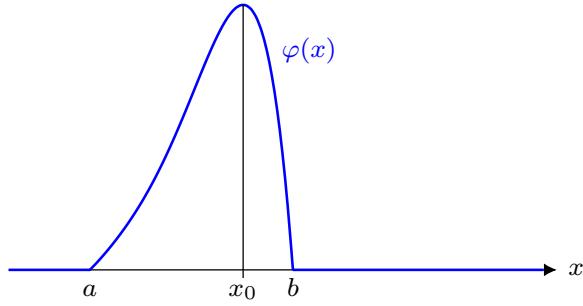


Abbildung 6.14: Anfangsbedingung $u(x, 0) = \varphi(x)$, $u_t(x, 0) = 0$.

Für festes $t > 0$ stellt $x \mapsto \varphi(x - ct)$ den um ct nach links verschobenen Graphen von φ dar, und $x \mapsto \varphi(x + ct)$ den um ct nach rechts verschobenen Graphen: Siehe Abbildung 6.15. Beide Funktionen $\varphi(x \pm ct)$ sind selber Lösungen der Wellengleichung. Lässt man t laufen, so bewegt sich also der eine Graph mit Geschwindigkeit c nach links, der andere nach rechts. Gemäß (6.32) ist die Lösung u also die Überlagerung dieser beiden Wellen.

Die Ausbreitung der Anregung lässt sich noch besser in der (x, t) -Ebene verfolgen: In der Figur 6.16 ist die Menge der Punkte (x, t) schraffiert, für die $u(x, t) \neq 0$ ist.

Die Formel (6.32) lässt sich (immer noch im Fall $\psi \equiv 0$) auch folgendermassen schreiben:

$$u(x, t) = \frac{1}{2} (u(x + ct, 0) + u(x - ct, 0)).$$

In Worten: Die Störung an der Stelle x zur Zeit t ist gleich dem arithmetischen Mittel der Anregungen an den Stellen $x - ct$ und $x + ct$ zur Zeit 0.

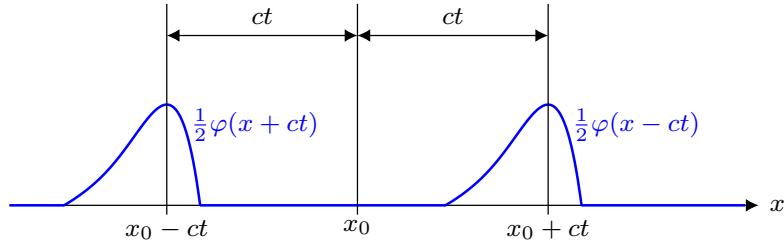


Abbildung 6.15: Die Lösung der Wellengleichung als Superposition einer nach links und einer nach rechts laufenden Welle.

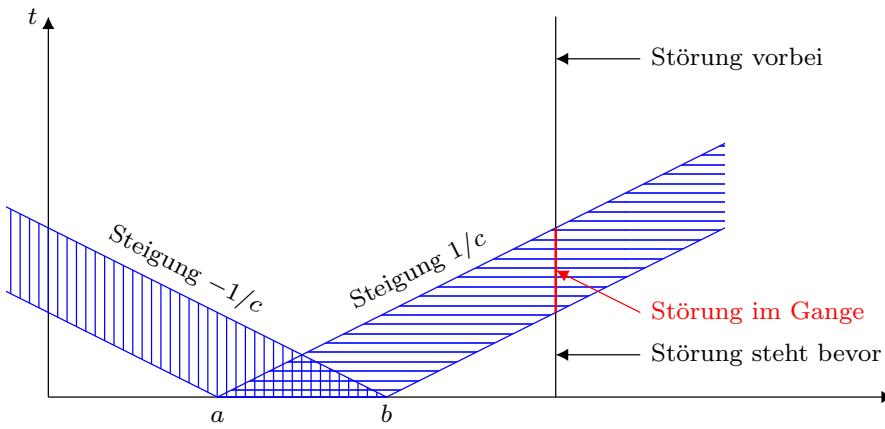


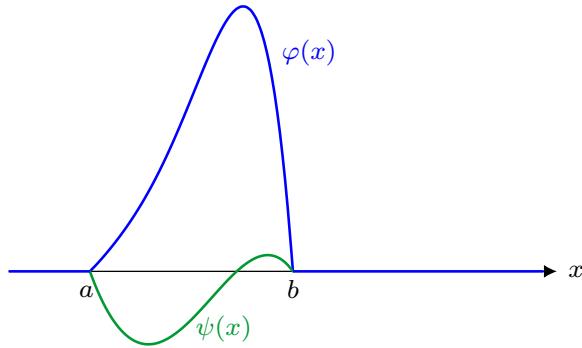
Abbildung 6.16: Einflussgebiet der Anregung im Fall $\psi \equiv 0$

Wir betrachten zweitens den Fall, wo die Anfangsgeschwindigkeit $u_t(x, 0) = \psi(x)$ nicht identisch verschwindet. Wir setzen aber nach wie vor voraus, dass die Anregung zur Zeit $t = 0$ auf das Intervall $[a, b]$ beschränkt ist (siehe Figur 6.17):

$$\varphi(x) = \psi(x) = 0 \quad \text{für alle } x \notin [a, b]. \quad (6.33)$$

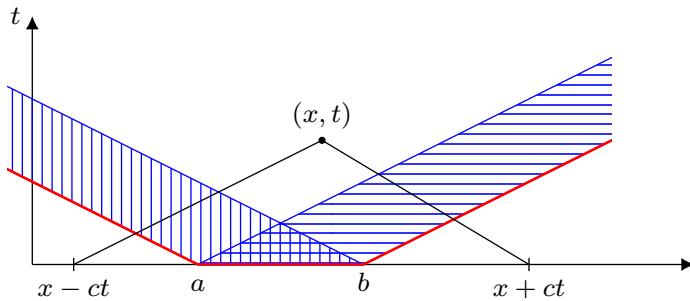
Wir wollen auch hier das ‘‘Ausbreitungsdiagramm’’ zeichnen (vergleiche die Figur 6.16). Dazu schraffieren wir das Gebiet in der (x, t) -Ebene, für das (möglicherweise) $u(x, t) \neq 0$ ist. Dabei ist $u(x, t)$ durch (6.32) gegeben: Ist $u(x, t) \neq 0$, so muss wegen (6.33) wenigstens eine der folgenden Bedingungen erfüllt sein:

- (i) $x + ct \in [a, b]$.
- (ii) $x - ct \in [a, b]$.
- (iii) Das Intervall $[x - ct, x + ct]$ (= Integrationsintervall in (6.32)) hat mit dem Intervall $[a, b]$ Punkte gemeinsam.

Abbildung 6.17: Allgemeine Anfangsbedingungen auf dem Intervall $[a, b]$.

Die Punkte, die einer der ersten beiden Bedingungen genügen, bilden zusammen die schon in der Figur 6.16 dargestellte Menge.

Wir haben diese Menge in der Figur 6.18 noch einmal eingetragen; ferner für einen fest gewählten Punkt (x, t) das Intervall $[x - ct, x + ct]$. Man erkennt, dass man den Punkt (x, t) beliebig oberhalb der roten Linie verschieben darf, ohne die Bedingung (iii) zu verletzen.

Abbildung 6.18: Durch φ verursachte Störung

Im Ganzen resultiert also der in Figur 6.19 dargestellte „Störungsbereich“ in der (x, t) -Ebene.

Allerdings hat die Störung für alle Punkte (x, t) für die $t \geq \frac{x-a}{c}$ und $t \geq -\frac{x-b}{c}$ den konstanten Wert

$$u(x, t) = \int_a^b \psi(y) dy \quad (6.34)$$

und dieser Wert ist wegen Impulserhaltung gleich Null, wenn die Anregung z. B. durch eine Explosion zustande gekommen ist. Das heisst, in Wahrheit ist eine zeitabhängige Störung auch im Fall $\psi \neq 0$ auf das in Figur 6.18 schraffierte Gebiet beschränkt. Nachdem diese Störung vorbei ist, hat die Lösung den in (6.34) angegebenen konstanten Wert. Diese Beobachtung gilt nicht nur in einer Raumdimensionen, sondern analog auch in allen ungeraden Raumdimensionen, insbesondere also im \mathbb{R}^3 . Wenn uns nämlich jemand etwas zuruft, so hören wir dessen Stimme, doch

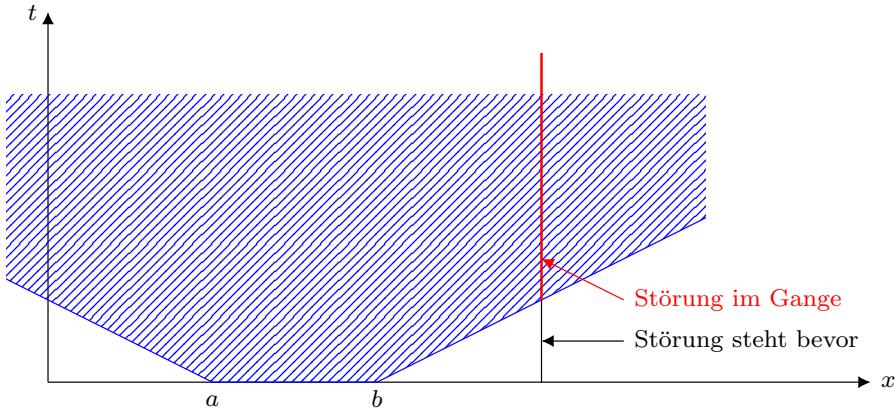


Abbildung 6.19: Störungsbereich bei allgemeinen Anfangsbedingungen.

sobald uns die Schallwellen passiert haben, herrscht wieder Ruhe. In geraden Raumdimensionen ist dies anders. Dort hält die zeitabhängige Störung für alle Zeiten an. In der Tat verursacht ein ins Wasser geworfener Stein auf der (zweidimensionalen) Wasseroberfläche anhaltende Wellen, die nur aufgrund der in Wirklichkeit vorhandenen Dämpfung schliesslich verschwinden. Dieses Phänomen wird **Huygenssches Prinzip** genannt. Es ist also ein grosses Glück, dass wir in einem dreidimensionalen Raum leben: In einer geraden Dimension wäre die akustische Verständigung äusserst schwierig.

6.2.6 Filter

Bei akustischen Signalen sind gelegentlich störende Nebengeräusche zu hören, die mit geeigneten Filtern entfernt werden können. Auch bei der Bildverarbeitung kommen Filter zum Einsatz. Auf diese Weise können Bilder aus einem Magnetresonanztomographen oder einem Computertomographen nachbearbeitet und für die Diagnose optimiert werden. Die Idee hinter diesen Filtern ist folgende: Wir haben gesehen, dass wir zum Beispiel periodische Signale mithilfe von Fourier-Reihen als Überlagerung von Sinus- und Kosinusfunktionen schreiben können. Will man beispielsweise hohe Frequenzen aus dem ursprünglichen Signal herausfiltern, nimmt man von der Fourier-Reihe einfach nur diejenigen Terme, welche kleineren Frequenzen entsprechen. Mithilfe des Fourier-Integrals lässt sich diese Idee auch für nicht-periodische Signale realisieren. Wir beginnen mit der folgenden Definition, welche Funktionen charakterisiert, die in ihrer Fourier-Reihe nur Frequenzen bis zu einer bestimmten Grenzfrequenz enthalten.

Definition 6.2.9 (b-Band-beschränkte Funktionen). $f : \mathbb{R} \rightarrow \mathbb{R}$ heisst *b-Band-beschränkt*, falls $\hat{f}(\lambda) = 0$ für alle λ mit $|\lambda| \geq b$.

Interpretation: Eine *b*-Band beschränkte Funktion enthält (in der Fourier-Darstellung) keine Frequenzen, die höher als *b* sind. Daher sind solche Funktionen besonders glatt.

Definition 6.2.10 (Die Sinus Cardinalis Funktion sinc). Die Funktion

$$\text{sinc}(x) := \frac{1}{2} \int_{-1}^1 e^{i\lambda x} d\lambda = \begin{cases} \frac{\sin(x)}{x} & \text{falls } x \neq 0 \\ 1 & \text{falls } x = 0 \end{cases}$$

ist 1-Band-beschränkt, denn

$$\text{sinc}(x) = \int_{-\infty}^{\infty} \widehat{\text{sinc}}(\lambda) e^{i\lambda x} d\lambda$$

und somit

$$\widehat{\text{sinc}}(\lambda) = \frac{1}{2} \chi_{]-1,1[}(\lambda).$$

Dabei bezeichnet χ die **charakteristische Funktion**: Für eine beliebige Menge $A \subset \mathbb{R}$ ist

$$\chi_A(x) = \begin{cases} 1 & \text{falls } x \in A \\ 0 & \text{falls } x \notin A \end{cases}$$

(siehe Abbildung 6.20).

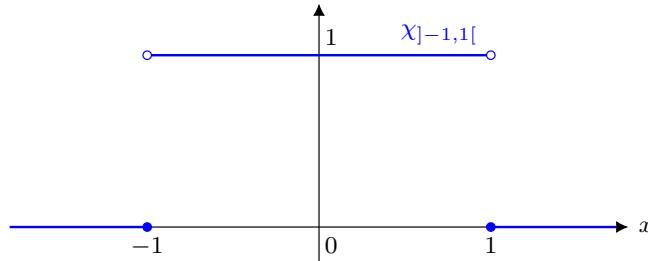


Abbildung 6.20: Graph der Funktion $\chi_{]-1,1[}$

Mit Hilfe der Rechenregeln für die Fourier-Transformation folgt somit für $F_b(x) := \frac{b}{\pi} \text{sinc}(bx)$:

$$\widehat{F}_b(\lambda) = \frac{1}{2\pi} \chi_{]-b,b[}(\lambda).$$

Also lässt sich mit F_b ein **Tiefpassfilter** mit Cutoff-Frequenz b bilden, denn

$$\widehat{f * F}_b(\lambda) = \widehat{f}(\lambda) \chi_{]-b,b[}(\lambda)$$

entfernt alle Frequenzen $|\lambda| \geq b$ aus der Fourier-Darstellung der Funktion f . Wir illustrieren die Wirkung des Tiefpassfilters in der Abbildung 6.21. Oben links ist ein „verrauschtes“ Signal f zu sehen. Es kann sich zum Beispiel um ein aufgezeichnetes akustisches Signal handeln oder um ein EKG-Signal. Rechts daneben ist die Fourier-Transformierte \widehat{f} zu sehen. Man beachte, dass neben tiefen auch hohe Frequenzen vertreten sind. Letztere sind für das ausgefranste Aussehen des Graphen von f verantwortlich. Multipliziert man \widehat{f} mit $\chi_{]-b,b[}$, so werden alle Frequenzen in \widehat{f} entfernt, die ausserhalb von $]-b, b[$ liegen. Der Graph von $\widehat{f} \chi_{]-b,b[}$ ist unten rechts zu sehen: Es sind nur die Frequenzen aus dem Intervall $]-b, b[$ verblieben. Die inverse Fourier-Transformierte von $\widehat{f} \chi_{]-b,b[}$ entsteht aus f durch Faltung mit F_b . Der Graph $f * F_b$ ist unten links zu sehen: Das Rauschen ist verschwunden!

Anwendung Akustik: Indem man ein verrausches Signal einem Tiefpassfilter unterwirft, lassen sich die störenden hohen Frequenzanteile entfernen (siehe Abbildung 6.22). Analog zu Tiefpassfiltern lassen sich Hochpassfilter realisieren: $f - f * F_b$ enthält keine Frequenzen mehr zwischen $-b$ und b . Ebenso lassen sich ganz gezielt einzelne Frequenzbänder löschen.

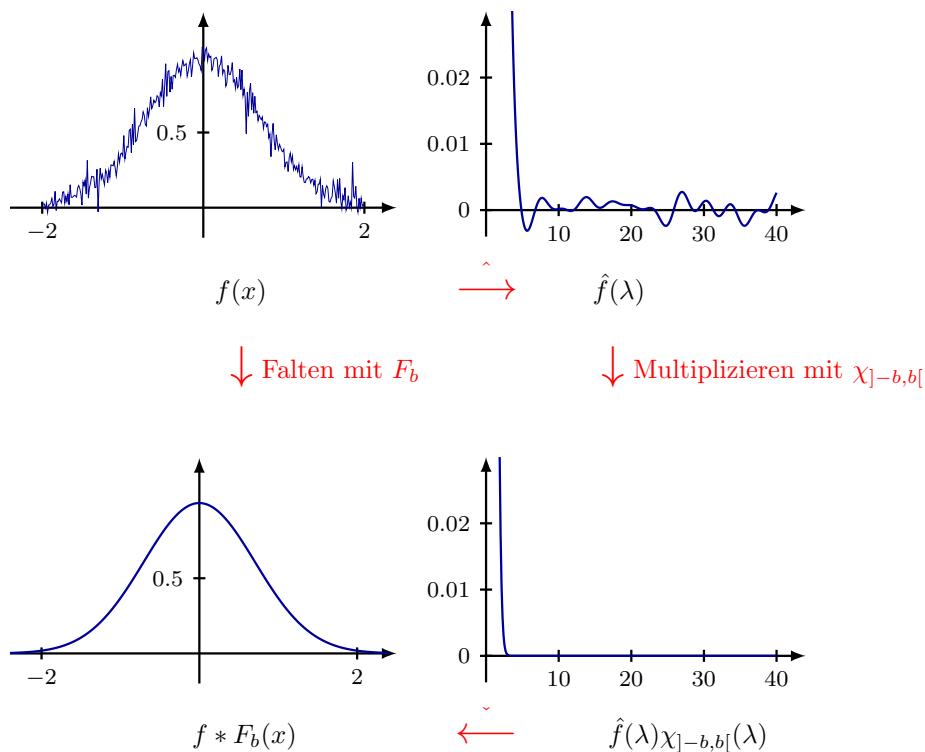


Abbildung 6.21: Anwendung eines Tiefpassfilters auf ein verrauschtes Signal.

(Rauschen)

(Rein)

Verrauschtes Signal

Gefiltertes Signal

Abbildung 6.22: Ein verrauscht akustisches Signal (links) und das gefilterte Signal (rechts). Zum Anhören [clicken](#).

Anwendung Bildverarbeitung: Genauso lassen sich auch unerwünschte Strukturen aus Bildern entfernen. Wird etwa ein gerastertes Bild eingescannt so ergeben sich oft sogenannte Moiré-Muster, ein Effekt, der auftritt, wenn sich zwei Gitter unterschiedlicher Periode überlagern. Der Effekt der Filterung ist in der Abbildung 6.23 gut zu sehen.

Bemerkungen:

- In der Praxis benutzt man die diskrete Fourier-Transformation DFT.



Bild mit Moiré-Muster



Gefiltertes Bild

Abbildung 6.23: Links ist ein eingescanntes Bild aus einer Zeitung mit einem typischen Moiré-Muster zu sehen, rechts dasselbe Bild nachdem es einem Tiefpassfilter unterworfen wurde. Die Bilder zeigen Michel Plancherel, einen Schweizer Pionier der Theorie der Fourier-Transformation.

- Die Fourier-Transformation ist auch in Dimensionen $n \geq 2$ möglich: In \mathbb{R}^n definiert man

$$\begin{aligned}\hat{f}(\lambda) &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(x) e^{ix \cdot \lambda} dx \quad (\text{Fourier-Transformation}) \\ f(x) &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \hat{f}(\lambda) e^{ix \cdot \lambda} d\lambda \quad (\text{Rücktransformation}),\end{aligned}$$

wobei $x \cdot \lambda$ das Skalarprodukt bezeichnet.

Achtung: Die Konstanten vor dem Integral werden in der Literatur unterschiedlich auf die Hin- und die Rücktransformation verteilt.

6.2.7 Das Shannon-Sampling-Theorem

Daten werden auf dem Computer digital gespeichert. Handelt es sich bei den Daten zum Beispiel um ein akustisches Signal oder ein EKG-Signal ist es gegeben durch eine Funktion $f(t)$, welche die Signallstärke als Funktion der Zeit t angibt. Aber $f(t)$ kann nur an diskreten Zeitpunkten t_i abgetastet und gespeichert werden. Bei dieser Diskretisierung geht alle Information zwischen den Messzeitpunkten t_i verloren. Bei der Verarbeitung und der Interpretation der Daten ist es entscheidend zu wissen, wieviel Information verloren gegangen ist, respektive wie gut man die ursprüngliche Funktion f durch die diskreten, aufgezeichneten Daten rekonstruieren kann. Diese Frage wollen wir in diesem Abschnitt untersuchen.

Eine gegebene Funktion, zum Beispiel $f : [0, 1] \rightarrow \mathbb{R}$, kann durch **Abtasten** (sampeln) in Punkten $x_i := \frac{i}{n}$, $i = 0, 1, \dots, n$, diskretisiert werden. Das heißt, man merkt sich nur die Werte $(f(x_0), f(x_1), \dots, f(x_n))$. Der umgekehrte Vorgang, nämlich aus den Werten $(f(x_0), f(x_1), \dots, f(x_n))$ die ganze Funktion $f : [0, 1] \rightarrow \mathbb{R}$ (wenigstens approximativ) zu rekonstruieren nennt man **Interpolieren**. Im Allgemeinen ist f durch die Werte in endlich vielen Punkten natürlich nicht eindeutig festgelegt. Es gilt jedoch das folgende erstaunliche Theorem. Es besagt, dass

b -bandbeschränkte Funktionen *exakt* aus den gesampelten Daten rekonstruiert werden können, wenn die Messpunkte x_i nahe genug beieinander liegen.

Theorem 6.2.11 (Shannon²). *Sei $f : \mathbb{R} \rightarrow \mathbb{R}$ b -Band-beschränkt und $h \leq \frac{\pi}{b}$. Dann ist f an allen Stetigkeitsstellen eindeutig bestimmt durch die Werte $f(kh)$, $k \in \mathbb{Z}$, und es gilt*

$$f(x) = \sum_{k \in \mathbb{Z}} f(kh) \operatorname{sinc}\left(\frac{\pi}{h}(x - kh)\right) \quad (6.35)$$

und für $|\lambda| \leq \frac{\pi}{h}$,

$$\hat{f}(\lambda) = \frac{h}{2\pi} \sum_{k \in \mathbb{Z}} f(kh) e^{-ikh\lambda}. \quad (6.36)$$

Beweis. Nach Voraussetzung ist $\hat{f}(\lambda) = 0$, für $\lambda \notin [-\frac{\pi}{h}, \frac{\pi}{h}]$. Daher ist (6.36) gerade die Fourier-Reihendarstellung von \hat{f} auf $[-\frac{\pi}{h}, \frac{\pi}{h}]$:

$$\hat{f}(\lambda) = \sum_{k \in \mathbb{Z}} c_k e^{\frac{2k\pi i \lambda}{2\pi/h}} = \sum_{k \in \mathbb{Z}} c_k e^{kh i \lambda}$$

mit

$$c_k = \frac{1}{2\pi/h} \int_{-\pi/h}^{\pi/h} \hat{f}(\lambda) e^{-\frac{2k\pi i \lambda}{2\pi/h}} d\lambda = \frac{h}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\lambda) e^{-kh i \lambda} d\lambda = \frac{h}{2\pi} f(-kh).$$

Die Formel (6.35) erhält man danach aus (6.36) wie folgt: Zunächst ist

$$\chi_{]-1,1[}\left(\frac{\lambda}{\pi/h}\right) = \chi_{]-\frac{\pi}{h}, \frac{\pi}{h}[}(\lambda).$$

Somit gilt für alle $\lambda \in \mathbb{R}$,

$$\hat{f}(\lambda) = \frac{h}{2\pi} \sum_{k \in \mathbb{Z}} f(kh) e^{-ikh\lambda} \chi_{]-1,1[}\left(\frac{\lambda}{\pi/h}\right).$$

Nach den Rechenregeln ist aber

$$\frac{h}{\pi} e^{-ikh\lambda} \frac{1}{2} \chi_{]-1,1[}\left(\frac{\lambda}{\pi/h}\right)$$

gerade die Fourier-Transformierte der Funktion $\operatorname{sinc}\left(\frac{\pi}{h}(x - kh)\right)$. □

Anwendung: CD, MP3 und andere Audioformate speichern das analoge akustische Signal digital. Es wird in einer gewissen hohen Taktfrequenz abgetastet und die zu diesen Zeitpunkten gemessenen Amplitude abgespeichert. Diese Daten werden dann dem Wiedergabegerät eingespeist.

Die menschliche Hörschwelle liegt bei rund $20 \cdot 10^3$ Hz, das heisst Töne oberhalb der Kreisfrequenz $b = 2\pi \cdot 20 \cdot 10^3$ spielen für die Wahrnehmung (fast) keine Rolle. Nach Shannon ist daher eine Abtastrate

$$\frac{1}{h} \geq \frac{b}{\pi} = 40 \cdot 10^3 \text{Hz}$$

nötig für eine Rekonstruktion des (hörbaren Teils des) Tonsignals aus seiner Digitalisierung. Tatsächlich wird für CDs oder MP3 eine Abtastrate von 44.1kHz verwendet.

²Claude Elwood Shannon (1916–2001) war ein US-amerikanischer Mathematiker und Elektroingenieur. Er gilt als Begründer der Informationstheorie.

6.2.8 Computer-Tomographie

Bildgebende Diagnoseverfahren wie die Computertomographie oder die Magnetresonanztomographie sind aus der modernen Medizin nicht mehr wegzudenken. Sie erlauben einen präzisen, nichtinvasiven Blick ins Innere des Körpers. Aber wie ist das möglich? Um dies Frage zu beantworten müssen wir kurz ausholen.

Das Absorptionsgesetz: Durchdringt ein Röntgenstrahl eine dünne Schicht, wird er abgeschwächt (siehe Abbildung 6.24).

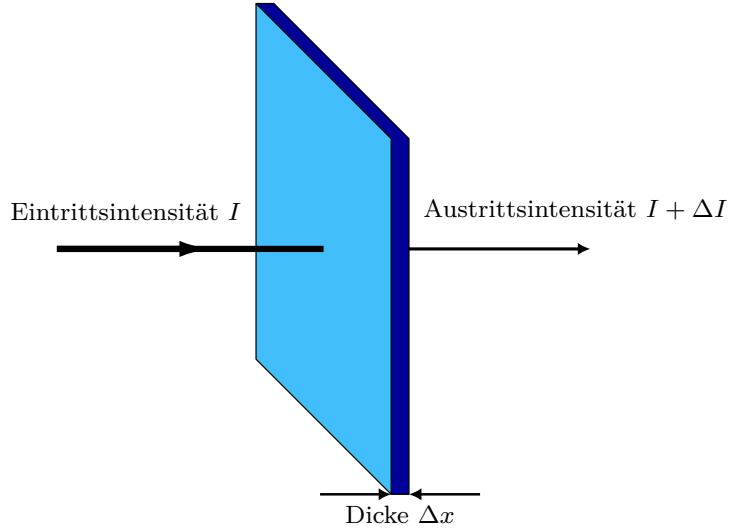


Abbildung 6.24: Absorption an einer dünnen Schicht.

Es gilt

$$\Delta I = -f I \Delta x,$$

wobei $f \geq 0$ die Absorptionsdichte des Materials bezeichnet. Betrachten wir nun ein inhomogenes Medium, in dem also die Absorptionsdichte f von der Eindringtiefe x abhängt (siehe Abbildung 6.25).

Dann ergibt sich aus

$$\frac{1}{I(x)} \frac{\Delta I}{\Delta x} = -f(x)$$

für $\Delta x \rightarrow 0$ die Differentialgleichung

$$\frac{d}{dx} \ln I(x) = \frac{I'(x)}{I(x)} = -f(x)$$

mit der Lösung

$$I(x) = I_0 \exp\left(-\int_0^x f(t) dt\right).$$

Beim Austritt aus dem Material hat sich also der Röntgenstrahl abgeschwächt auf den Wert

$$I_1 = I_0 \exp\left(-\int_{\mathbb{R}} f(t) dt\right).$$

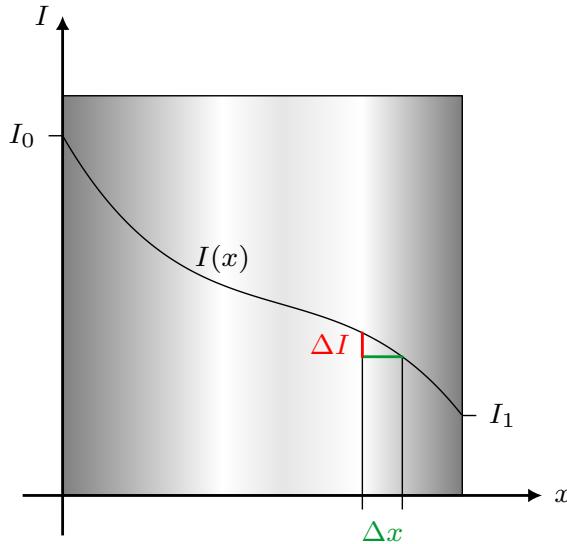


Abbildung 6.25: Absorption in einem inhomogenen Medium.

Durch Messung lässt sich also für jede Gerade L in der Ebene E der Wert

$$-\ln\left(\frac{I_1}{I_0}\right) = \int_L f(t)dt \quad (6.37)$$

ermitteln.

Stellen wir uns nun vor, wir möchten eine exakte Diagnose bei einem Verdacht auf einen Hirntumor stellen. Es ist klar, dass ein simples Röntgenbild nicht hilfreich ist, denn die inneren Strukturen im Gehirn werden vom Schädelknochen verdeckt. Dennoch soll ein Bild eines Querschnitts durch den Kopf entlang einer Ebene E erzeugt werden (siehe Abbildung 6.26).

In der Ebene E messen wir entlang von beliebig vielen Geraden L die Werte (6.37). Die Frage ist nun, ob aus diesen Messdaten die Funktion $f(x)$ für $x \in E$ rekonstruiert werden kann. Interpretiert man dann den Wert $f(x)$ als Farbwert des Pixels an der Position x , erhält man das gewünschte Bild des Querschnitts.

Um die Frage zu beantworten parametrisieren wir jede Gerade L in der Ebene E durch einen Winkel θ und ihren Abstand s vom Ursprung (siehe Abbildung 6.27).

Wir betrachten:

- $G =$ die Menge der Geraden L in der Ebene E .
- $A =$ die Menge der Absortionsfunktionen $f : E \rightarrow \mathbb{R}$.
- $B =$ die Menge der Funktionen $g : G \rightarrow \mathbb{R}$.

Bei gegebener Absortionsfunktionen $f : E \rightarrow \mathbb{R}$ erhält man also via (6.37) eine Abbildung $g : G \rightarrow \mathbb{R}$ zuordnen. Dies gibt Anlass zur folgenden Definition:

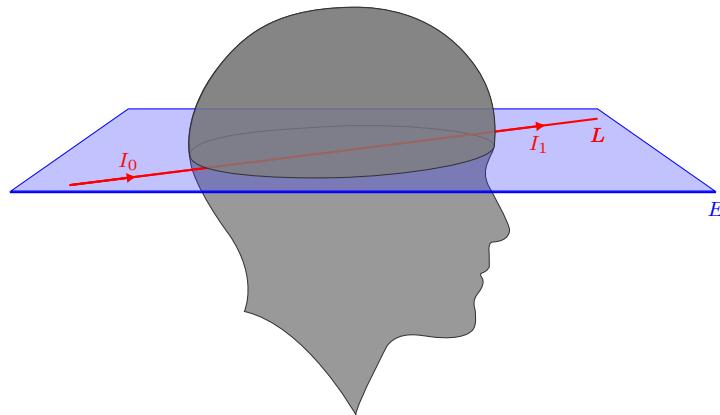
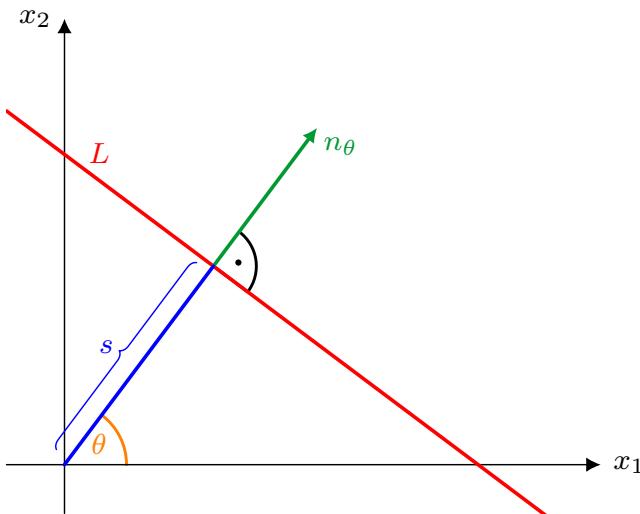


Abbildung 6.26: Querschnitt durch einen Kopf.

Abbildung 6.27: Parametrisierung der Geraden durch θ und s .

Definition 6.2.12 (Die Radon-Transformation). *Die Radon-Transformation $R : A \rightarrow B$ ordnet jeder Funktion $f \in A$ eine Funktion $Rf \in B$ wie folgt zu.*

$$\begin{aligned} R : A &\rightarrow B \\ f &\mapsto Rf, \text{ wobei } Rf(L) := \int_L f(x) dx. \end{aligned}$$

Notation: Ist die Gerade L durch (θ, s) gegeben, schreibt man

$$Rf(L) = Rf(\theta, s) = R_\theta f(s).$$

Die Frage ist also, ob f (die Absorptionsfunktion, also das Bild des Querschnitts) aus Rf (den

Messungen) rekonstruiert werden kann.

Die erste Rekonstruktionsformel stammt von Johann Radon (1917). 1963 fertigte Alan MacLeod Cormack auf Radons Erkenntnissen aufbauend einen ersten Computer-Tomographen. Das erste marktfähige Gerät baute Sir Godfrey Newbold Hounsfield 1972. Beide wurden für ihre Leistung 1979 mit dem Nobelpreis in Medizin ausgezeichnet. Radon erlebte diese Entwicklung leider nicht mehr.

Mit Hilfe der Fourier-Transformation ist eine sehr einfache Rekonstruktionsformel zu finden:

Theorem 6.2.13 (Fourier-Rekonstruktion). *Es gilt*

$$\widehat{R_\theta f}(\sigma) = \hat{f}(\sigma n_\theta).$$

Beweis. Sei n_θ ein auf L senkrechter Einheitsvektor, und n_θ^\perp ein dazu senkrechter Einheitsvektor in Richtung L . Dann gilt:

$$\begin{aligned}\widehat{R_\theta f}(\sigma) &= \frac{1}{2\pi} \int_{\mathbb{R}} R_\theta f(s) e^{-is\sigma} ds \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} f(sn_\theta + yn_\theta^\perp) dy e^{-is\sigma} ds \\ &= \frac{1}{2\pi} \int_{\mathbb{R}^2} f(x) e^{-i\sigma n_\theta \cdot x} dx = \hat{f}(\sigma n_\theta),\end{aligned}$$

wobei in der letzten Zeile die Substitution $x = sn_\theta + yn_\theta^\perp$ benutzt wurde, das heisst $s = n_\theta x$, $dx = dy$. \square

In der Praxis kennt man Rf nur auf einer endlichen Menge von Geraden L . Die Kunst besteht darin, genügend viele geeignete Geraden für die Messungen zu wählen um trotzdem eine adäquate Rekonstruktion von f zu berechnen. Wählt man zu wenige Geraden, wird das Bild ungenau, es enthält Artefakte und kleine Details werden nicht erfasst. In der Abbildung 6.28 sind Computer-tomogramme eines künstlichen Schädels zu sehen. Die Auflösung ist mit Absicht grob gewählt um den Effekt des Undersampling zu demonstrieren.

Neben der Fourier-Methode existieren weitere Rekonstruktionsalgorithmen, etwa die **Filtered Backprojection Methode** oder die **Methode von Kacmarz**.

6.3 Numerische Methode für die Poisson-Gleichung

Die Poisson-Gleichung taucht bei einer Vielzahl biologischer und medizinischer Situationen auf, etwa bei der Untersuchung des stationären Verhaltens bei der Diffusion durch Membranen, bei der Musterbildung in der Embryologie, oder beim Wachstums von avaskulären soliden Tumoren (siehe zum Beispiel [Murray] oder [Roose]). Die Poisson-Gleichung lautet

$$\Delta u(x) = f(x), \quad x \in \Omega \subset \mathbb{R}^n. \tag{6.38}$$

Dabei ist $u : \mathbb{R}^n \rightarrow \mathbb{R}$ die gesuchte Funktion und $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ist gegeben. $\Omega \subset \mathbb{R}^n$ ist ein festes Gebiet, in dem die Gleichung zu lösen ist, wobei Randbedingungen auf dem Rand $\partial\Omega$ von Ω gegeben sind, im einfachsten Fall

$$u(x) = 0, \quad \text{für } x \in \partial\Omega. \tag{6.39}$$

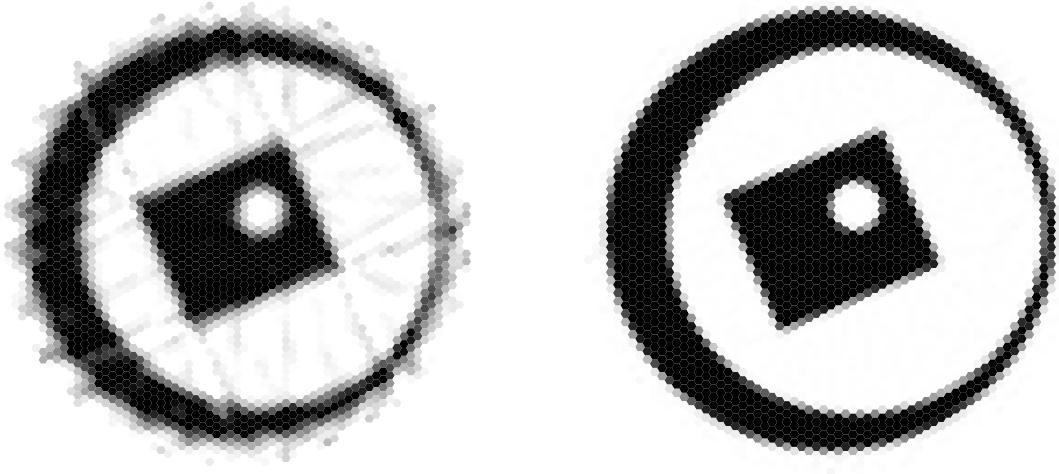


Abbildung 6.28: Das Bild links zeigt ein Tomogramm einer Abtastung bei 19 Winkeln und je 30 Geraden. Das rechts Bild zeigt das Tomogramm bei korrektes Sampling: Um Details dieser Grösse aufzulösen genügen Abtastungen bei 99 Winkeln und je 158 Geraden.

Als Anwendung stellen wir uns ein transdermales Pflaster vor, mit dem ein im Pflaster enthalter Wirkstoff über die Haut in den Körper diffundiert. Arzneistoffe, die sich auf diese Weise verabreichen lassen sind beispielsweise Nitroglycerin (bei Angina pectoris), Clonidin (bei Bluthochdruck), Rotigotin (zur Behandlung der Parkinson-Krankheit) oder Nikotin (zur Raucherentwöhnung). $u(x)$ beschreibt dann die Wirkstoffkonzentration in der Haut am Punkt $x \in \Omega$, wobei $\Omega \subset \mathbb{R}^2$ die Region bezeichnet, welche das Pflaster auf der Haut bedeckt. $f(x)$ ist dabei die Wirkstoffkonzentration im Pflaster. Die von der Zeit t abhängige Diffusion des Wirkstoffs in die Haut kann dann beschrieben werden durch

$$u_t(x, t) = \Delta u(x, t) + f(x) \quad \text{für } t \geq 0, x \in \Omega \quad (6.40)$$

$$u(x, t) = 0 \quad \text{für } x \in \partial\Omega \text{ auf dem Rand und } t \geq 0 \quad (6.41)$$

$$u(x, 0) = 0 \quad \text{für } x \in \Omega \text{ zur Zeit } t = 0 \quad (6.42)$$

Wenn das Pflaster schon eine gewisse Zeit auf der Haut liegt, stellt sich eine stationäre (von der Zeit unabhängige) Situation ein, also $u_t = 0$. Das heisst (6.38), (6.39) beschreiben die stationäre Situation. Der Einfachheit halber nehmen wir an, dass $f(x) = 1$ konstant und $\Omega = (0, 1) \times (0, 1)$ ein quadratisches Pflaster ist. Das Problem

$$\Delta u(x) = -1 \quad x \in \Omega = (0, 1) \times (0, 1) \quad (6.43)$$

$$u(x) = 0 \quad \text{für } x \in \partial\Omega \quad (6.44)$$

kann in diesem einfachen Fall zwar mithilfe von Fourier-Reihen analytisch gelöst werden, aber sobald das Gebiet Ω eine kompliziertere Form aufweist, ist man auf numerische Lösungsmethoden angewiesen. Eine solche, nämlich das **Gauss-Seidel Relaxationsverfahren**, wollen wir nun vorstellen.

Die zentrale Idee des Verfahrens ist die Approximation des Differentialquotienten durch den Differenzenquotienten: Sei $v : \mathbb{R} \rightarrow \mathbb{R}$ eine glatte Funktion. Wir entwickeln v im Punkt x in eine

Taylor-Reihe bis zur Ordnung 3: Für ein festes $h \geq 0$ gilt

$$v(x+h) = v(x) + hv'(x) + \frac{h^2}{2}v''(x) + \frac{h^3}{6}v'''(\xi)$$

$$v(x-h) = v(x) - hv'(x) + \frac{h^2}{2}v''(x) - \frac{h^3}{6}v'''(\eta)$$

für ein $\xi \in]x, x+h[$ und ein $\eta \in]x-h, x[$. Wir dividieren die Differenz der obigen Gleichungen durch $2h$ und erhalten

$$\frac{v(x+h) - v(x-h)}{2h} = v'(x) + h^2 C \approx v'(x). \quad (6.45)$$

Die linke Seite von (6.45) heisst **zentraler erster Differenzenquotient** und approximiert $v'(x)$. Der Fehlerterm ist $h^2 C$ mit $C = \frac{1}{12}(v'''(\xi) + v'''(\eta))$, er geht also von der Größenordnung h^2 gegen Null. Ganz analog erhält man den **zentralen zweiten Differenzenquotienten**:

$$\frac{v(x+h) - 2v(x) + v(x-h)}{h^2} = v''(x) + Ch^2 \approx v''(x). \quad (6.46)$$

Zur Diskretisierung überziehen wir Ω mit einem Gitter der Maschenweite $h = 1/n, n \in \mathbb{N}$ (siehe Abbildung 6.29).

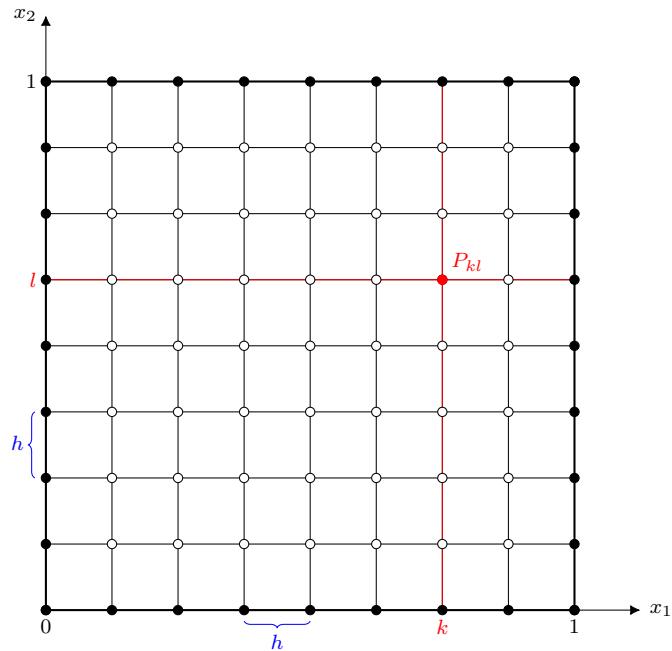


Abbildung 6.29: Gitter in Gauss-Seidel-Verfahren

Wir führen eine Nummerierung der Gitterpunkte ein: Wir bezeichnen die Gitterpunkte mit P_{kl} (der erste Index nummeriert in horizontaler, der zweite in vertikaler Richtung). Die Randbedingung $u = 0$ auf $\partial\Omega$ können wir leicht diskretisieren: Wir setzen

$$u(P_{kl}) = 0$$

für alle Randgitterpunkte P_{kl} (die schwarzen Punkte \bullet in der Abbildung 6.29).

Zur Diskretisierung von $\Delta u(x, y) = -1$ benutzen wir (6.46)

$$\begin{aligned}\Delta u = u_{xx} + u_{yy} &\approx \frac{1}{h^2} (u(x+h, y) - 2u(x, y) + u(x-h, y)) + \\ &+ \frac{1}{h^2} (u(x, y+h) - 2u(x, y) + u(x, y-h)).\end{aligned}$$

Also gilt

$$\Delta u \approx \frac{1}{h^2} (u(x+h, y) + u(x, y+h) + u(x-h, y) + u(x, y-h) - 4u(x, y)). \quad (6.47)$$

Setzen wir $u_{kl} := u(P_{kl})$, so wird (6.47):

$$\Delta u(P_{kl}) \approx \frac{1}{h^2} (u_{k+1,l} + u_{k,l+1} + u_{k-1,l} + u_{k,l-1} - 4u_{kl}).$$

Somit lautet die diskretisierte Form von $\Delta u = -1$

$$u_{k+1,l} + u_{k,l+1} + u_{k-1,l} + u_{k,l-1} - 4u_{kl} = -h^2. \quad (6.48)$$

(6.48) ist ein lineares Gleichungssystem mit ebensovielen Unbekannten u_{kl} , wie innere Gitterpunkte vorhanden sind (sagen wir N). Da jeder innere Gitterpunkt (\circ in der Abbildung 6.29) eine Gleichung liefert, haben wir auch N Gleichungen. Man beachte dabei folgendes: Liegt ein P_{kl} in der Nachbarschaft eines (oder zweier) Randgitterpunktes (wie in Figur 6.30), so ist $u_{k+1,l}$,

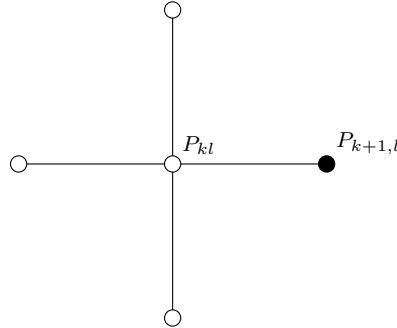


Abbildung 6.30: Randgitterpunkt

nicht als unbekannt, sondern als gegeben zu betrachten, denn es gilt ja $u_{k+1,l} = 0$.

Ein bekannter Satz der linearen Algebra sagt über solche $N \times N$ Gleichungssysteme:

Entweder besitzt das lineare Gleichungssystem $Ax = b$ (A eine $N \times N$ Matrix, x, b Vektoren) genau eine Lösung (bei beliebiger Wahl von b),

oder das zugehörige homogene System $Ax = 0$ besitzt eine nichttriviale Lösung.

Wir wollen nun nachweisen, dass für unser Gleichungssystem das „Oder“ nicht zutrifft, dass also der „Entweder“-Fall gilt!

Lemma 6.3.1. *Verschwindet u_{kl} in allen Randgitterpunkten P_{kl} und ersetzen wir in (6.48) die rechte Seite durch 0, so besitzt das entsprechend modifizierte, homogene Gleichungssystem*

$$u_{k+1,l} + u_{k,l+1} + u_{k-1,l} + u_{k,l-1} - 4u_{kl} = 0 \quad (6.49)$$

nur die Null-Lösung, das heisst für alle k und l ist

$$u_{kl} = 0.$$

Beweis. Zunächst schreiben wir (6.49) in suggestiver Form

$$u_{kl} = \frac{1}{4}(u_{k+1,l} + u_{k,l+1} + u_{k-1,l} + u_{k,l-1}). \quad (6.50)$$

Der Wert von u_{kl} ist das arithmetische Mittel der Werte seiner vier Nachbarpunkte. Dies entspricht der Mittelwerteigenschaft der harmonischen Funktionen.

Wir wollen nun annehmen, u sei 0 in allen Randpunkten, ohne identisch zu verschwinden. Nimmt zum Beispiel u positive Werte in inneren Gitterpunkten an, so ist das Maximum M von u positiv und wird in einem inneren Gitterpunkt P_{kl} angenommen. Dort gilt

$$M = u_{kl} = \frac{1}{4}(u_{k+1,l} + u_{k,l+1} + u_{k-1,l} + u_{k,l-1}).$$

Da $u_{k+1,l}, u_{k,l+1}, u_{k-1,l}$ und $u_{k,l-1} \leq M$, schliessen wir

$$u_{k+1,l} = u_{k,l+1} = u_{k-1,l} = u_{k,l-1} = M.$$

Ist unter diesen Punkten ein Randgitterpunkt, so ist dies bereits ein Widerspruch. Sind alle Nachbarn von P_{kl} innere Gitterpunkte, so kann man für sie die obige Schlussweise wiederholen: Auch in ihren Nachbarpunkten muss $u = M$ sein. Nach endlich vielen Schritten trifft man auf einen Randpunkt und erhält den Widerspruch.

Nimmt u keine positiven (aber dafür negative) Werte an, so hat man statt des Maximums das Minimum zu betrachten. \square

Somit besitzt das lineare Gleichungssystem (6.48) genau eine Lösung. Aber wie berechnen wir diese?

Der sicherlich falsche Weg wäre, das System hinzuschreiben und mit einem Eliminationsverfahren anzugehen (bei 100 inneren Punkten hätte die Matrix 10'000 Elemente, die meisten davon Null). Besser ist eine iterative Lösung:

- (i) Man wähle eine Startnäherung: z. B. $u_{kl}^{(0)} := 0$ für alle Gitterpunkte P_{kl} .
- (ii) Man berechne die $(n+1)$ -te Näherung $u_{kl}^{(n+1)}$ aus der n -ten Näherung $u_{kl}^{(n)}$ wie folgt: Man setze für alle inneren Punkte P_{kl}

$$u_{kl}^{(n+1)} := \frac{1}{4}(u_{k+1,l}^{(n)} + u_{k,l+1}^{(n)} + u_{k-1,l}^{(n)} + u_{k,l-1}^{(n)} + h^2)$$

Die Folge $u_{kl}^{(n)}$ strebt dann für $n \rightarrow \infty$ gegen die Lösung u_{kl} von (6.48).

Dies ist das **Gauss-Seidel-Verfahren**. Auf den Konvergenzbeweis wollen wir hier verzichten. Die Abbildung 6.31 zeigt die numerische Gauss-Seidel-Lösung bei $n = 50$.

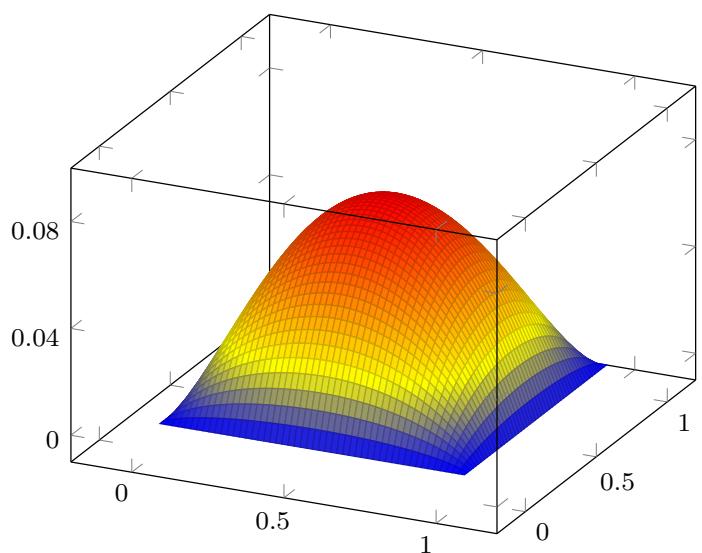


Abbildung 6.31: Numerische Lösung von (6.43), (6.44).

KAPITEL 7

Laplace-Transformation

7.1 Definition und Notation

Neben der Fourier-Transformation ist eine weitere Integraltransformation von grosser praktischer Bedeutung. Zunächst führen wir einige grundlegende Begriffe ein.

Definition 7.1.1. Eine Funktion

$$F : t \mapsto F(t)$$

einer reellen Variablen t mit reellen oder komplexen Werten heisst **Originalfunktion**, wenn F folgenden vier Bedingungen genügt:

- (i) F ist auf der ganzen reellen Achse definiert.
- (ii) F ist stückweise stetig differenzierbar.
- (iii) Für $t < 0$ gilt $F(t) = 0$.
- (iv) F wächst für $t \rightarrow \infty$ höchstens exponentiell, d.h. es gibt reelle Konstanten σ und M derart, dass

$$|\textcolor{red}{F}(t)| \leq \textcolor{blue}{M} e^{\sigma t} \quad (7.1)$$

für alle $t \geq 0$.

Die Menge aller Originalfunktionen wird **Originalraum** der Laplace-Transformation genannt.

Abbildung 7.1: Eine typische Originalfunktion $\textcolor{red}{F}$

Zu (iii): Falls für eine Funktion F die Bedingung (iii) nicht „von Natur aus“ erfüllt ist, so kann man (iii) immer erzwingen: Die Funktion

$$G(t) := \begin{cases} F(t) & \text{für } t \geq 0 \\ 0 & \text{für } t < 0 \end{cases}$$

erfüllt nämlich (iii). Wir werden im folgenden nicht jedes Mal ausdrücklich sagen, dass die betrachteten Funktionen für $t < 0$ Null sind. So reden wir z. B. von der „Originalfunktion“ $F(t) = \cos(\omega t)$ und meinen damit die Funktion

$$F(t) = \begin{cases} \cos(\omega t) & \text{für } t \geq 0 \\ 0 & \text{für } t < 0 \end{cases}.$$

Zu (iv): Gilt (7.1) für σ , so auch für jedes $\sigma' > \sigma$. Die kleinste Zahl σ_0 mit der Eigenschaft, dass (7.1) für jedes $\sigma > \sigma_0$ gilt, heisst **Wachstumskoeffizient** von F . (Es kann sein, dass (7.1) auch noch für σ_0 selber gilt, muss aber nicht).

Beispiel 7.1.2. (i) Die Funktion $t \mapsto e^{t^2}$ ist keine Originalfunktion, denn (iv) ist für kein σ erfüllt! Dies kann man etwa durch Logarithmieren von (7.1) einsehen.

(ii) Die **Heaviside Sprungfunktion**

$$H : t \mapsto \begin{cases} 1 & \text{für } t \geq 0 \\ 0 & \text{für } t < 0 \end{cases}$$

ist eine Originalfunktion. Ihr Wachstumskoeffizient ist $\sigma_0 = 0$.

(iii) Sei $n = 1, 2, \dots$. Die Funktion

$$F(t) := \begin{cases} t^n & \text{für } t \geq 0 \\ 0 & \text{für } t < 0 \end{cases}$$

ist eine Originalfunktion. Wir zeigen, dass (7.1) für jedes $\sigma > 0$ gilt. Bekanntlich gilt für jedes $\sigma > 0$

$$\lim_{t \rightarrow \infty} t^n e^{-\sigma t} = 0,$$

$t^n e^{-\sigma t}$ besitzt daher im Intervall $0 \leq t < \infty$ ein Maximum M . Daraus folgt

$$t^n \leq M e^{\sigma t} \quad \text{für alle } t \geq 0.$$

F hat also den Wachstumskoeffizienten $\sigma_0 = 0$.

(iv) Sei $a = \alpha + i\beta$ (α, β reell). Die Funktion

$$F : t \mapsto \begin{cases} e^{at} = e^{\alpha t} \cdot e^{i\beta t} & \text{für } t \geq 0 \\ 0 & \text{für } t < 0 \end{cases}$$

ist eine Originalfunktion mit Wachstumskoeffizient $\sigma_0 = \alpha$.

Sei nun F eine Originalfunktion mit Wachstumskoeffizient σ_0 . Dann gilt das folgende Lemma.

Lemma 7.1.3. *Für jede komplexe Zahl $s = \sigma + i\omega$ mit $\sigma > \sigma_0$ existiert das Integral*

$$\int_0^\infty e^{-st} F(t) dt.$$

Beweis. Wir wählen eine Zahl σ_1 mit $\sigma_0 < \sigma_1 < \sigma$. Dann gilt:

$$\begin{aligned} \left| \int_0^\infty e^{-st} F(t) dt \right| &\leq \int_0^\infty |e^{-st}| |F(t)| dt = \int_0^\infty e^{-\sigma t} |F(t)| dt \\ &\leq \int_0^\infty e^{-\sigma t} M e^{\sigma_1 t} dt = M \int_0^\infty e^{-(\sigma-\sigma_1)t} dt \\ &= -\frac{M}{\sigma-\sigma_1} e^{-(\sigma-\sigma_1)t} \Big|_0^\infty = \frac{M}{\sigma-\sigma_1} < \infty. \end{aligned}$$

□

Wegen Lemma 7.1.3 ist nun folgende Definition möglich:

Definition 7.1.4. Sei F eine Originalfunktion mit dem Wachstumskoeffizienten σ_0 . Die komplexe Funktion f einer komplexen Variablen,

$$f : s \mapsto \int_0^\infty e^{-st} F(t) dt \in \mathbb{C} \quad (\operatorname{Re} s > \sigma_0),$$

heisst **Laplace-Transformierte** oder **Bildfunktion** von F . Die Menge aller Laplace-Transformierten heisst **Bildraum**. Die Abbildung, die jeder Originalfunktion ihre Bildfunktion zuordnet, heisst **Laplace-Transformation**. Ist f Laplace-Transformierte von F , so schreiben wir dafür

$$f = \mathcal{L}[F]$$

oder das sogenannte **Doetsch-Symbol**

$$F \circ \bullet f \quad \text{oder} \quad f \bullet \circ F$$

(\circ markiert die Originalfunktion, \bullet die Bildfunktion).

Konvention: Wir werden für die Originalfunktionen Grossbuchstaben und für die zugehörigen Bildfunktionen die entsprechenden Kleinbuchstaben verwenden. Die Laplace-Transformation ordnet jeder Funktion des Originalraumes eine Funktion des Bildraumes zu:

7.2 Rechenregeln und Beispiele für die Laplace-Transformation

Um das Rechnen mit der Laplace-Transformation zu erleichtern, gibt es, ähnlich wie bei der Fourier-Transformation, einige einfache Regeln.

Theorem 7.2.1 (Rechenregeln für die Laplace-Transformation).

Regel 1: Linearität Die Laplace-Transformation ist linear: Es gilt

$$\mathcal{L}[aF + bG] = a\mathcal{L}[F] + b\mathcal{L}[G]$$

für beliebige komplexe Zahlen a und b .

Regel 2: Ähnlichkeitssatz Für jedes reelle $\alpha > 0$ und $f = \mathcal{L}[F]$ gilt

$$F(\alpha t) \circlearrowleft \bullet \frac{1}{\alpha} f\left(\frac{s}{\alpha}\right).$$

Regel 3: Differentiationssatz Sei F für $t > 0$ stetig, und mit F sei auch die Ableitung F' eine Originalfunktion. Dann gilt

$$F'(t) \circlearrowleft \bullet sf(s) - F(0),$$

wobei mit $F(0)$ der rechtsseitige Grenzwert $\lim_{t \rightarrow 0^+} F(t)$ gemeint ist. Die Laplace-Transformation verwandelt also eine Ableitung (bis auf eine Konstante $F(0)$) in eine Multiplikation mit s .

Regel 3ⁿ Wiederholte Anwendung von Regel 3 liefert

$$F^{(n)}(t) \circlearrowleft \bullet s^n f(s) - s^{n-1}F(0) - s^{n-2}F'(0) - \dots - F^{(n-1)}(0).$$

Diese Regel gilt für $n = 0, 1, 2, \dots$, vorausgesetzt, $F', F'', \dots, F^{(n)}$ sind Originalfunktionen und $F, F', F'', \dots, F^{(n-1)}$ sind für $t > 0$ stetig. Wiederum sind die Werte $F(0), F'(0), \dots, F^{(n-1)}(0)$ als rechtsseitige Grenzwerte zu verstehen.

Regel 4: Multiplikationssatz Der Ableitung im Bildraum entspricht die Multiplikation mit $-t$ im Originalraum:

$$-tF(t) \circlearrowleft \bullet f'(s).$$

Regel 4ⁿ Durch Iteration erhält man die allgemeine Formel:

$$(-t)^n F(t) \circlearrowleft \bullet f^{(n)}(s), \quad n = 0, 1, 2, \dots$$

Regel 5: Integrationssatz Der Integration im Originalraum entspricht die Division durch s im Bildraum:

$$\int_0^t F(\tau) d\tau \circlearrowleft \bullet \frac{1}{s} f(s).$$

Regel 6: Divisionssatz Es besitze F den Wachstumskoeffizienten σ_0 , und neben F sei auch die Funktion $G(t) := \frac{1}{t} F(t)$ Originalfunktion. Dann gilt für $\text{Re } s > \sigma_0$,

$$G(t) = \frac{F(t)}{t} \circlearrowleft \bullet \int_s^\infty f(u) du.$$

Der Integration im Bildraum entspricht also die Division durch t im Originalraum.

Regel 7: Verschiebungssatz Für jedes $T_0 > 0$ gilt

$$F(t - T_0) \circlearrowleft \bullet e^{-sT_0} f(s).$$

Der Verschiebung im Originalraum um T_0 entspricht die Multiplikation mit dem Dämpfungsfaktor e^{-sT_0} im Bildraum.

Regel 8: Dämpfungssatz Für beliebiges komplexes a gilt

$$e^{at} F(t) \xrightarrow{\quad} f(s - a).$$

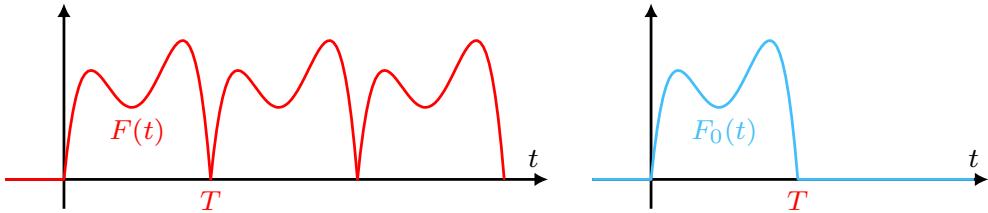
Der Verschiebung um a im Bildraum entspricht also die Multiplikation mit dem Dämpfungsfaktor e^{at} im Originalraum.

Regel 9: Satz über periodische Funktionen Es sei F eine periodische Funktion mit Periode $T > 0$, d.h.

$$F(t + T) = F(t) \quad \text{für } t \geq 0.$$

Dann gilt

$$F(t) \xrightarrow{\quad} \frac{1}{1 - e^{-sT}} \int_0^T e^{-st} F(t) dt = \frac{1}{1 - e^{-sT}} \mathcal{L}[F_0](s).$$



Regel 10: Faltungssatz Seien F und G Originalfunktionen und f und g ihre Laplace-Transformierten.

Dann ist die Faltung

$$(F * G)(t) = \int_0^t F(\tau)G(t - \tau) d\tau$$

ebenfalls eine Originalfunktion, und es gilt

$$F * G \xrightarrow{\quad} f g.$$

Die Laplace-Transformierte verwandelt also eine Faltung im Originalraum in ein Produkt im Bildraum.

Die Regeln sind einfach zu beweisen (siehe zum Beispiel das Buch über PDEs, das zur Vorlesung empfohlen wurde).

Beispiel 7.2.2. (i) Sei H die Heaviside-Funktion von Beispiel 7.1.2,

$$\mathcal{L}[H(t)] = \int_0^\infty e^{-st} dt = -\frac{1}{s} e^{-st} \Big|_0^\infty = \frac{1}{s} \quad \text{für } \operatorname{Re} s > 0.$$

(ii) Die Bildfunktion f zur Originalfunktion $F : t \mapsto e^{at}$ mit $a = \alpha + i\beta$ ist

$$f(s) = \int_0^\infty e^{-st} e^{at} dt = \int_0^\infty e^{-(s-a)t} dt = \frac{1}{s - a} \quad \text{für } \operatorname{Re} s > \alpha.$$

(iii) Sei ω_0 reell. Es gilt $\sin(\omega_0 t) = \frac{1}{2i}(e^{i\omega_0 t} - e^{-i\omega_0 t})$. Im letzten Beispiel hatten wir

$$e^{i\omega_0 t} \xrightarrow{\quad} \frac{1}{s - i\omega_0} \quad \text{und} \quad e^{-i\omega_0 t} \xrightarrow{\quad} \frac{1}{s + i\omega_0}$$

für $\operatorname{Re} s > 0$. Aus Regel 1 folgt somit

$$\sin(\omega_0 t) \xrightarrow{\quad} \frac{1}{2i} \left(\frac{1}{s - i\omega_0} - \frac{1}{s + i\omega_0} \right) = \frac{\omega_0}{s^2 + \omega_0^2}.$$

(iv) Anwenden der Regel 3 liefert nun

$$\frac{d}{dt} \sin(\omega_0 t) = \omega_0 \cos(\omega_0 t) \xrightarrow{\quad} \frac{s\omega_0}{s^2 + \omega_0^2}.$$

Also gilt (laut Regel 1)

$$\cos(\omega_0 t) \xrightarrow{\quad} \frac{s}{s^2 + \omega_0^2}. \quad \circlearrowright$$

Kann es sein, dass zwei verschiedene Originalfunktionen dieselbe Laplace-Transformierte besitzen? Oder umgekehrt gefragt: Ist bei gegebener Bildfunktion die zugehörige Originalfunktion eindeutig bestimmt? Die Antwort gibt der folgende Satz, den wir hier nicht beweisen.

Theorem 7.2.3. *Es seien F_1, F_2 zwei Originalfunktionen, und es sei*

$$\mathcal{L}[F_1] = \mathcal{L}[F_2].$$

Dann ist an allen Stellen t , an denen F_1 und F_2 stetig sind,

$$F_1(t) = F_2(t).$$

Nach Satz 7.2.3 unterscheiden sich also zwei Originalfunktionen einer Bildfunktion höchstens an ihren Umstetigkeitsstellen voneinander. Es ist sinnvoll, zwei solche Funktionen als gleich anzusehen. Satz 7.2.3 sagt dann, dass bei gegebener Bildfunktion die zugehörige Originalfunktion *eindeutig* bestimmt ist. Die Zuordnung

$$\text{Bildfunktion } f \mapsto \text{Originalfunktion } F$$

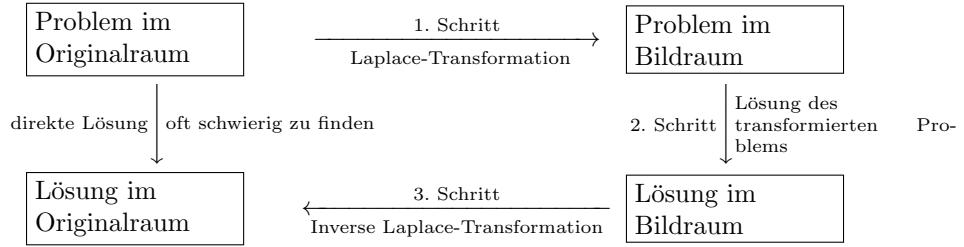
wird **inverse Laplace-Transformation** oder **Rücktransformation** genannt. Man schreibt dafür

$$F = \mathcal{L}^{-1}[f] \text{ oder } f \xrightarrow{\quad} F.$$

Nun können wir die Grundidee der Methode der Laplace-Transformation darlegen:

Es sei eine gewöhnliche oder partielle Differentialgleichung zu lösen. Zur Lösung des Problems gehen wir wie folgt vor:

1. Schritt: Wir wenden auf die Gleichung die Laplace-Transformation an und übersetzen so das Problem in den Bildraum.
2. Schritt: Wir lösen das Problem im Bildraum.
3. Schritt: Wir transformieren die im Bildraum gefundene Lösung in den Originalraum zurück.



7.3 Anwendungsbeispiel

Nun wollen wir die Laplace-Transformation zur Lösung einer gewöhnlichen Differentialgleichung einsetzen.

Beispiel 7.3.1. Man löse das Anfangswertproblem

$$\begin{cases} F'' + 2F' + F = te^{-t} + e^t \\ F(0) = 0, \quad F'(0) = 0. \end{cases} \quad (7.2)$$

Lösung: Wir transformieren die Differentialgleichung (7.2). Zuerst die linke Seite:

$$\begin{aligned} \mathcal{L}[F'' + 2F' + F] &\stackrel{\text{R.1}}{=} \mathcal{L}[F''] + 2\mathcal{L}[F'] + \mathcal{L}[F] \\ &\stackrel{\text{R.3}}{=} s^2 f(s) - sF(0) - F'(0) + 2sf(s) - 2F(0) + f(s) \\ &= (s^2 + 2s + 1)f(s) = (\mathbf{s} + 1)^2 f(s). \end{aligned}$$

Nun die rechte Seite:

$$\begin{aligned} \mathcal{L}[te^{-t} + e^t] &\stackrel{\text{R.1}}{=} \mathcal{L}[te^{-t}] + \mathcal{L}[e^t] \\ &\stackrel{\text{R.4}}{=} -\frac{d}{ds}\mathcal{L}[e^{-t}] + \mathcal{L}[e^t] \\ &\stackrel{\text{R.8}}{=} -\frac{d}{ds}\mathcal{L}[1](s+1) + \mathcal{L}[1](s-1) = -\frac{d}{ds}\frac{1}{s+1} + \frac{1}{s-1} \\ &= \frac{1}{(s+1)^2} + \frac{1}{s-1}. \end{aligned}$$

Gleichsetzen **beider Seiten** liefert

$$f(s) = \frac{1}{(s+1)^4} + \frac{1}{(s-1)(s+1)^2}.$$

Man beachte, dass im Bildraum die Lösung des Problems ganz einfach durch algebraische Umformung zu finden war! Nachdem das Problem im Bildraum gelöst ist, müssen wir die Lösung noch zurücktransformieren.

Der erste Term ist noch sehr einfach:

$$\begin{aligned} \frac{1}{(s+1)^4} &\stackrel{\text{R.8}}{\longrightarrow} \bullet \mathcal{L}^{-1}\left[\frac{1}{s^4}\right]e^{-t} = \mathcal{L}^{-1}\left[-\frac{1}{6}\left(\frac{1}{s}\right)''''\right]e^{-t} \\ &\stackrel{\text{R.4}}{=} \frac{1}{6}t^3\mathcal{L}^{-1}\left[\frac{1}{s}\right]e^{-t} \stackrel{\text{Bsp. 7.2.2(i)}}{=} \frac{1}{6}t^3e^{-t}. \end{aligned}$$

Der zweite Term wird zuerst in Partialbrüche zerlegt:

$$\frac{1}{(s-1)(s+1)^2} = \frac{A}{s-1} + \frac{B}{s+1} + \frac{C}{(s+1)^2}.$$

Durch eine der zahlreichen Methoden findet man

$$A = -B = \frac{1}{4}, \quad C = -\frac{1}{2}.$$

Somit

$$\begin{aligned} \frac{1}{(s-1)(s+1)^2} &\circlearrowleft \frac{1}{4}\mathcal{L}^{-1}\left[\frac{1}{s-1}\right] - \frac{1}{4}\mathcal{L}^{-1}\left[\frac{1}{s+1}\right] - \frac{1}{2}\mathcal{L}^{-1}\left[\frac{1}{(s+1)^2}\right] \\ &\stackrel{\text{R.8}}{=} \frac{1}{4}\mathcal{L}^{-1}\left[\frac{1}{s}\right]e^t - \frac{1}{4}\mathcal{L}^{-1}\left[\frac{1}{s}\right]e^{-t} - \frac{1}{2}\mathcal{L}^{-1}\left[\frac{1}{s^2}\right]e^{-t} \\ &\stackrel{\text{R.4}}{=} \frac{1}{4}e^t - \frac{1}{4}e^{-t} - \frac{1}{2}te^{-t}. \end{aligned}$$

Damit lautet die Lösung

$$F(t) = \left(\frac{t^3}{6} - \frac{t}{2} - \frac{1}{4}\right)e^{-t} + \frac{1}{4}e^t. \quad \circlearrowright$$

In ähnlicher Weise kann man die Laplace-Transformation verwenden, um partielle Differentialgleichungen, Integralgleichungen oder Differenzengleichungen zu lösen.