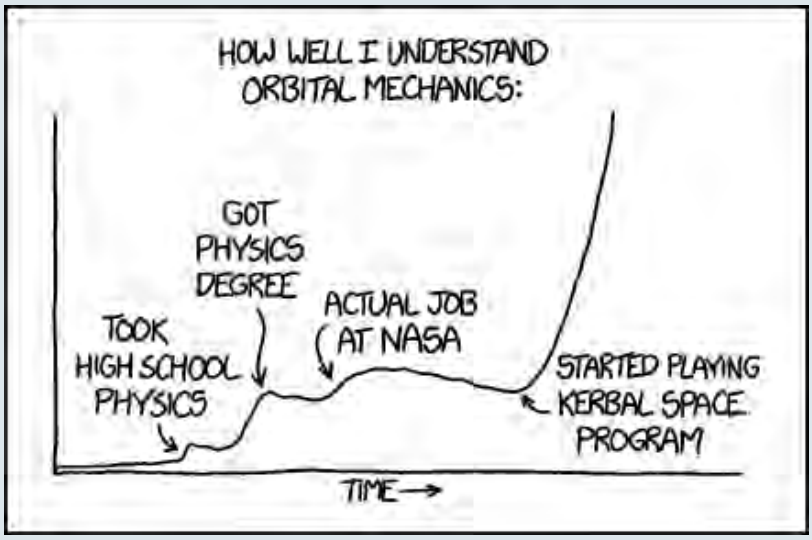
JLU
NEUE WEGE. SEIT 1607.

JUSTUS-LIEBIG-
UNIVERSITÄT
GIESSEN



Skript zum Seminar

Tutorium zu Physik und E-Technik in der Raumfahrt

Oktober 2024, Version 3.0.0

IPI

Version 3.0.0

Satz: LaTeX

Kontakt:

I. Physikalisches Institut
AG Ionentriebwerke
Heinrich-Buff-Ring 16
35392 Gießen
E-Mail: Kristof.Holste@physik.uni-giessen.de

Skript zum Seminar

Tutorium zu Physik und E-Technik in der Raumfahrt BRF-G-01

Dr. Kristof Holste

18. Oktober 2024

INHALTSVERZEICHNIS

VORWORT

Das ≫Tutorium zur Raumfahrt≪ wird seit der Einführung des Studiengangs *Physik und Technologie für Raumfahrtanwendungen* im Jahr 2017 an der Justus-Liebig-Universität Gießen angeboten. Ursprünglich im 2. Semester angesiedelt, wurde es auf das erste Semester erweitert und ist seit dem Wintersemester 2023/2024 in den ersten drei Semestern fester Bestandteil des Curriculums. Es ist die erste studiengangsspezifische Veranstaltung, die nicht aus dem klassischen Kanon der Physik- und Elektrotechnik-Studiengänge stammt.

Von seiner Form her ist das Tutorium ein ganz gewöhnliches Seminar, wie es an Universitäten unzählige davon gibt. Seminare unterscheiden sich von Vorlesungen darin, dass eine stärkere Interaktion zwischen Studierenden untereinander und Studierenden und Lehrenden vorliegen sollte, eine aktive Beteiligung von Studierenden wird in Seminaren daher gefordert. Vorlesungen mögen einen Überblick über ein Fachgebiet geben, in Seminaren werden einzelne Themen einer Vorlesung diskutiert, vertieft, eingeübt und vieles mehr. Es ist kein Zufall, dass die deutsche Übersetzung von Seminar nichts anderes als Baumschule bedeutet, hier werden in ihre Köpfe die - hoffentlich - richtigen Vorstellungen und Gedanken ausgesät, die dann in der Zukunft Blüten tragen können.

Die Zielsetzung des Tutoriums besteht darin, die grundlegenden Inhalte der Vorlesungen in Physik und Elektrotechnik gezielt mit raumfahrtspezifischen Themen zu verknüpfen. Vertiefte und spezialisierte Aspekte werden in späteren Vorlesungen behandelt. Seit dem Sommersemester 2024 erfolgt die Modulabschlussprüfung in Form eines schriftlichen Tests, der am Termin der letzten Lehrveranstaltung stattfindet. Der Test umfasst sowohl Multiple-Choice-Fragen als auch Rechenaufgaben, wobei alle Rechenaufgaben im Verlauf des Tutoriums ausführlich behandelt wurden.

Einige grundsätzliche Anmerkungen zu diesem Skript und dessen Rolle: Dieses Dokument ist über einen längeren Zeitraum hinweg entstanden und wird kontinuierlich ergänzt und korrigiert. Es ist daher möglich, dass es inhaltliche Fehler oder Redundanzen enthält. Wie alle Materialien, die Sie im Rahmen Ihres Studiums lesen, sollte es kritisch hinterfragt werden. Dieses Skript ist kein professionelles Lehrbuch und sollte nicht als alleinige Quelle zur Prüfungsvorbereitung dienen (das gilt nicht nur für diesen Kurs). Vielmehr soll es als Anregung dienen, sich mithilfe weiterführender Literatur intensiver mit bestimmten Themen

auseinanderzusetzen – insbesondere mit denen, die wir in den Übungen vertiefen.

Nutzen Sie die Möglichkeit, Lehrbücher als PDF kostenlos über die Universitätsbibliothek zu beziehen. Ob in digitaler Form oder als gedrucktes Buch – Lehrbücher sind ein unverzichtbarer Bestandteil der Prüfungsvorbereitung. Studierende, die sich ausschließlich auf Skripte und Internetquellen verlassen, haben – wie wir jedes Semester bei etwa 30 bis 50 % der Studierenden feststellen – häufig Schwierigkeiten, den Stoff in der notwendigen Tiefe zu verstehen. Das Internet kann selbstverständlich eine wertvolle Informationsquelle sein, sollte jedoch nicht als Ersatz für fundierte Lehrbücher dienen. Nutzen Sie die Ressourcen der Bibliothek der JLU.

Auch wenn die Wahl eines Lehrbuchs oft Geschmackssache ist, möchten wir hier einige Empfehlungen aussprechen:

- Die vierbändige Lehrbuchreihe Experimentalphysik von Wolfgang Demtröder gibt es bereits seit über 20 Jahren (2021 erschien die 9. Auflage) und ist dementsprechend beliebt. Sie bieten eine gute Mischung aus Experimenten und deren theoretischer Behandlung. Band 1 dieser Reihe *Mechanik und Wärme* ist als Einstieg in das Studium und als Begleitbuch zum Tutorium uneingeschränkt zu empfehlen.

- Das Buch *Grundlagen der Orbitmechanik* von Volker Maiwald deckt sich inhaltlich sehr gut mit den Inhalten des Tutorials. Es ist didaktisch sinnvoll aufgebaut und diesem Skript nicht unähnlich.

- Feynmans Vorlesungen über Physik sind nicht unbedingt das richtige Buch für Anfänger. Hier wird wenig bis gar nicht gerechnet, dafür aber viel erklärt. Wer sich mit der Mathematik schwer tut, wird sicher wenig Motivation haben, ein dickes Lehrbuch über das *große Ganze* der Physik zu lesen. Wer aber diese Hürde genommen hat, findet hier eine Fülle von Informationen, die über das typische Lehrbuchwissen hinausgehen.

- Wer die Bahnmechanik bis ins Detail verstehen will, dem sei das Buch *Orbital Mechanics* von Prussing und Conway empfohlen, das zu den Standardwerken auf diesem Gebiet zählt. Es ist anspruchsvoll, aber mit etwas Mühe zu bewältigen. Außerdem ist es angenehm kompakt.

Für das Wintersemester 2022/23 wurde dieses Skript um ein Kapitel zu mathematischen Grundlagen, Rechentechniken und Kegelschnitten erweitert, die für ein naturwissenschaftlich-technisches Studium und die

erfolgreiche Teilnahme an dieser Lehrveranstaltung unerlässlich sind. Die Erfahrung der letzten Jahre hat gezeigt, dass viele der mathematischen Methoden, die im Tutorium benötigt werden, erst spät in den Mathematikvorlesungen behandelt werden. Dies führte dazu, dass mathematische Konzepte vorzeitig und beiläufig im Tutorium eingeführt werden mussten – ein Vorgehen, das uns didaktisch nicht überzeugte.

Daher haben wir beschlossen, das Tutorium mit einer Einführung in die wesentlichen mathematischen Rechenverfahren zu beginnen, jedoch in stark komprimierter Form. Die Grundbegriffe der Vektor- und Infinitesimalrechnung werden in den ersten Sitzungen des Tutoriums in Vorlesungsform vorgestellt. Dabei wird bewusst auf eine vollständige Darstellung der Zusammenhänge verzichtet – dafür gibt es separate Vorlesungen und umfassende Mathematikbücher, die diese Themen weitaus detaillierter behandeln können. Ziel dieser Einführung ist es, die Studierenden mit den wichtigsten mathematischen Methoden vertraut zu machen, sodass sie in der Lage sind, Fragestellungen der Newtonschen Mechanik mit Raumfahrtbezug zu bearbeiten, ohne an der erforderlichen Mathematik zu scheitern.

An dieser Stelle sei darauf hingewiesen, dass der Inhalt des mathematischen Kapitels nicht als Prüfungsstoff dient, der gezielt abgefragt wird. Vielmehr verfolgen wir das Motto *non discere debemus ista, sed didicisse*[1]. Das bedeutet, es wird erwartet, dass Sie sich diese Inhalte im Laufe des ersten Semesters selbstständig aneignen, da ein Verständnis der zentralen Themen des Tutoriums ohne diese Grundlagen nicht möglich ist.

Wer nach einem Semester Schwierigkeiten hat, eine einfache Funktion wie $f(x) = \frac{1}{x}$ abzuleiten oder zu integrieren, sollte möglicherweise in Betracht ziehen, ein Studienfach ohne Mathematik zu wählen. Dies ist nicht als Überheblichkeit gemeint, sondern als realistische Einschätzung: Weder Ihnen noch uns als Lehrenden ist geholfen, wenn Sie viel Zeit in ein Studium investieren, das letztlich wenig Erfolgsaussichten hat. Dennoch gilt auch: Der Stoff des Tutoriums ist für jeden erlernbar, der bereit ist, die notwendige Zeit und Mühe zu investieren – der eine braucht vielleicht mehr, der andere weniger.

K. Holste, Gießen, 18. Oktober 2024

1 Aus einem Brief Senecas an Lucilius, in dem er nach dem Sinn des Grundstudiums fragt.

1

PHYSIKALISCHE RECHENMETHODEN

1.1 Funktionen

Funktionen sind das tägliche Handwerkszeug eines Wissenschaftlers im naturwissenschaftlich-technischen Bereich. Man kann sich eine Funktion als eine eindeutige Abbildung einer Menge auf eine andere vorstellen. Diese Mengen können Mengen von Zahlen sein, beispielsweise die Menge der natürlichen Zahlen $\mathbb{N}$. Ein Beispiel für eine Funktion aus der Physik ist die ideale Gasgleichung

$$p = \frac{N \cdot k_\mathrm{B} \cdot T}{V}, \tag{1.1}$$

die einen Zusammenhang zwischen dem Druck p und den drei Variablen Temperatur T, Teilchenzahl N und Volumen V sowie einer Naturkonstanten k_B, der Boltzmann-Konstanten, herstellt. Wo diese Funktion herkommt, ist erstmal unerheblich – sie wird später in der Experimentalphysik behandelt. Man muss lediglich wissen, was man für die jeweiligen Variablen einsetzen kann. Wir müssen also einen Bereich festlegen, den diese Variablen einnehmen können, den Definitionsbereich D. Die Funktionswerte liegen dann – je nach Funktionstyp – in einem bestimmten Wertebereich W. Sowohl D als auch W stellen Mengen von Zahlen dar.

Wenn man Temperatur und Teilchenzahl in der Funktion p als Konstanten betrachtet, also einen festen Wert annimmt, dann hat man eine Funktion von nur noch einer Variablen vorliegen, also ein $p = p(V)$. Wenn man Funktionen allgemein untersucht, dann wird meist die Form $y = f(x)$ verwendet. Hier steht y für den Funktionswert und x für das Argument der Funktion, also für die unabhängige Variable. Eine Funktion hat eine sehr wichtige Eigenschaft, die Eindeutigkeit des Funktionswertes. Jedes Element des Definitionsbereichs wird auf genau ein Element des Wertebereichs abgebildet. Es spielt dabei keine Rolle, ob verschiedene Elemente des Definitionsbereichs auf das gleiche Element des Wertebereichs abgebildet werden (im Extremfall beispielsweise durch eine Funktion $f(x) = \text{const.}$, die nur ein Element im Wertebereich hat). Ausgeschlossen ist der Fall, in dem ein Element des Definitionsbereichs auf zwei Elemente des Wertebereichs abgebildet wird (vgl. Abb. 1.1).

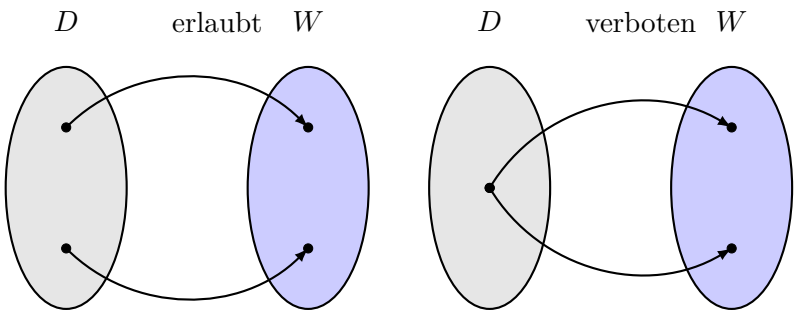
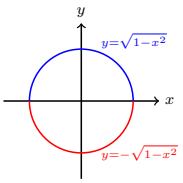


Abbildung 1.1: Zu den wesentlichen Eigenschaften einer Funktion gehört die Eindeutigkeit des Funktionswerts. Die linke Darstellung stellt diese Eindeutigkeit her. Beide Elemente der Definitionsmenge D werden auf verschiedene Elemente des Wertebereichs W abgebildet. Es ist auch nicht verboten, dass beide Pfeile auf das gleiche Element des Wertebereichs zeigen. Die rechte Darstellung stellt keine Funktion dar. Hier wird ein Element des Definitionsbereichs auf zwei verschiedene Elemente des Wertebereichs abgebildet, die Eindeutigkeit geht verloren.

So ist z. B. die Schreibweise $f(x) = \pm\sqrt{x}$ keine Funktion, $f(x) = x^2$ hingegen schon.

1.1.1 Analytische Darstellung

Es gibt verschiedene Möglichkeiten, eine Funktion analytisch darzustellen. Die in Tabelle 1.2 dargestellten Möglichkeiten spielen eine zentrale Rolle in der Physik. Sehr oft hat man es mit der expliziten Darstellung zu tun, bei der man den Funktionswert y eindeutig durch eine Variable ausdrückt. Demgegenüber steht die implizite Darstellung, bei der x und y auf der gleichen Seite stehen, man beide also in die Form $F(x, y) = 0$ bringen kann. Das mag für einfache lineare Zusammenhänge wie $x + y = 0$ unnötig erscheinen, weil man durch einfaches Umstellen sofort die explizite Darstellung $y = -x$ bekommt, ergibt aber bei komplexeren Zusammenhängen wie $x^2 + y^2 - 1 = 0$ schon einen Sinn. Aus diesem Zusammenhang ergeben sich bereits zwei explizite Funktionen der Form $y_1 = \sqrt{1 - x^2}$ und $y_2 = -\sqrt{1 - x^2}$. Man kann eine implizite Darstellung daher auch als zusammenfassende Darstellung mehrerer expliziter Darstellungen auffassen.



Die implizite Darstellung $x^2 + y^2 - 1 = 0$ definiert übrigens den Einheitskreis, also einen Kreis mit Radius 1, der bei der Diskussion der Winkelfunktionen noch eine wichtige Rolle spielen wird.

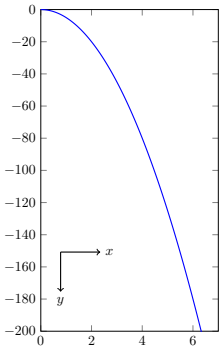
Die Parameterform tritt immer dann auf, wenn verschiedene Variablen explizit von genau einer anderen Variablen abhängen. In der Mechanik

Tabelle 1.1: Darstellungsarten von Funktionen

Darstellungsart	**Beispiel(e)**
explizite Darstellung	$f(x) = x^2$, $y = x^2$
implizite Darstellung	$x^2 + y^2 = 1$
Parameterform	$x(t) = vt, y(t) = \frac{1}{2}gt^2$

ist dies häufig der Fall, wenn beispielsweise die Ortskoordinaten (x, y, z) eines Objekts von der Zeit t abhängen. Der reibungsfreie freie Fall eines Körper der Masse m im Schwerefeld der Erde (mit Erdbeschleunigung g) kann in Parameterdarstellung als

$$\begin{aligned} x(t) &= v_0 t \\ y(t) &= -\frac{1}{2}gt^2 \end{aligned} \tag{1.2}$$



dargestellt werden. Hierbei steht v_0 für eine anfängliche Geschwindigkeit in x-Richtung. Diese Geschwindigkeit sorgt für eine Zeitabhängigkeit der x-Koordinate, wäre $v_0 =$ würde sich die x-Komponente über die Zeit t nicht ändern. Wenn dies der Fall wäre, dann hätte man nur noch eine Ortsvariable, die sich mit t ändert, was die Parameterdarstellung überflüssig machen würde. Wenn nun beide Ortskoordinaten von t abhängen und man eine Darstellung der Form $y(x)$ will, dann kann man eine Variablentransformation durchführen, indem man $t = x/v_0$ in $y(t)$ einsetzt. Man erhält dadurch

$$y(x) = -\frac{1}{2}g\frac{1}{v_0^2}x^2, \tag{1.3}$$

also einen parabelförmigen Verlauf, dem der Körper beim freien Fall folgt.

1.1.2 Graphische Darstellung

Funktionen können in Form einer graphischen Darstellung visualisiert werden. Bei physikalischen Funktionen kann man dadurch oft Zusammenhänge besser erkennen, beispielsweise wie sich eine Funktion bei sehr großen oder sehr kleinen Werten der variablen Größen, also der Argumente, verhält. Eine Funktion kann in einem kartesischen Koordinatensystem wie in Abbildung 1.2 als Menge von geordneten Wertepaaren $(x, f(x))$ dargestellt werden. Ein kartesisches Koordinatensystem wird

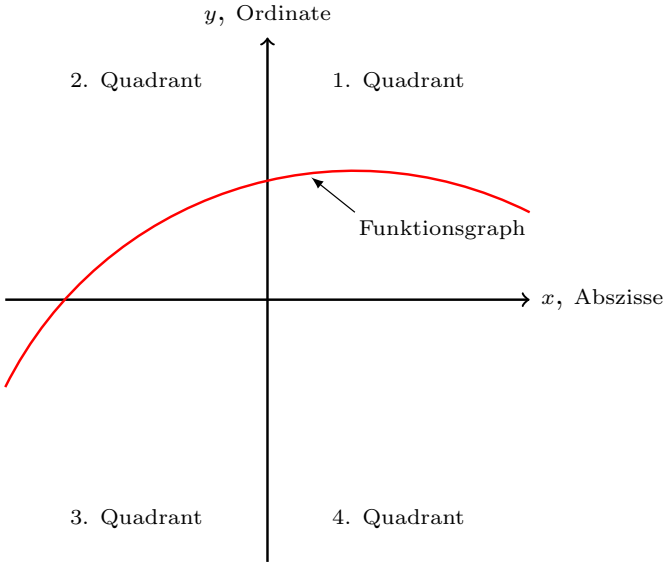
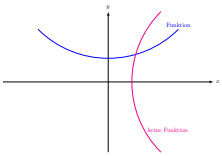


Abbildung 1.2: Funktion



in vier Quadranten eingeteilt, die x-Achse bezeichnet man als Abszisse, die y-Achse als Ordinate. Die Eindeutigkeit von Funktionen kann man am Graphen auch gut erkennen, da zu jedem x-Wert ein und nur ein y-Wert existiert. Kann man anhand des Graphen einem x-Wert zwei oder mehr y-Werte zuordnen, dann liegt keine Funktion vor.

1.1.3 Definitionsbereich

Unter dem Definitionsbereich D versteht man die Menge an Werten, die man als Argument in die Funktion einsetzen kann. Es gibt hier eine Vielzahl von möglichen Schreibweisen, wie man diese Mengen bezeichnen kann. In der Physik gibt man den Definitionsbereich dann auch gar nicht mehr an, weil sowieso meist klar ist, welche Werte als Argument in Frage kommen. Häufig hat man es mit der Menge der reellen Zahlen zu tun, die in einigen Fällen auf bestimmte Bereiche beschränkt sind. So ist z. B. die absolute Temperatur, die man im SI-System in Kelvin angibt, auf die positiven reellen Zahlen beschränkt. Gehört die Null als untere Grenze zum Definitionsbereich, so ist die Menge an der unteren Grenze abgeschlossen. Wird die Null nicht zur Menge dazu gezählt, dann ist die Menge an dieser Stelle offen. Liegen x-Werte in einem offenen Intervall, das nach unten durch a und nach oben durch b begrenzt wird, so kann man dieses offene Intervall als

$$(a, b) = \{x \, \epsilon \, \mathbb{R} \; / \; a < x < b\} \tag{1.4}$$

schreiben. Für ein abgeschlossenes Intervall gilt entsprechend

$$[a, b] = \{x \in \mathbb{R} \ / \ a \leq x \leq b\}. \tag{1.5}$$

Schauen wir uns als Beispiel die Funktion $y = \sqrt{1 - x^2}$ an. Im Bereich der reellen Zahlen darf keine negative Zahl unter der Wurzel stehen, somit gilt also für den Definitionsbereich

$$D = [-1, 1] = \{x \in \mathbb{R} \ / \ |x| \leq 1\}. \tag{1.6}$$

1.1.4 Gerade und ungerade Funktionen

Es gibt Funktionen, die eine sehr interessante Symmetrieeigenschaft haben können und die man entweder als gerade oder ungerade Funktionen bezeichnet. Eine gerade Funktion ist dabei über die Eigenschaft

$$f(x) = f(-x) \tag{1.7}$$

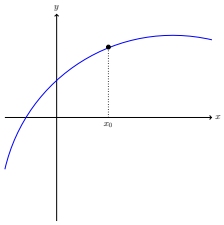
definiert. Dies bedeutet, dass eine Spiegelung des Teils der Funktion links der y-Achse an ebendieser Achse der Funktion auf der rechten Seite der Ordinate entspricht. Wir haben hier also Achsenspiegelung vorliegen. Eine ungerade Funktion ist über die Bedingung

$$f(x) = -f(-x) \tag{1.8}$$

definiert. Hier liegt eine doppelte Spiegelung sowohl an der Ordinate als auch an der Abszisse vor, die dadurch in eine Punktspiegelung am Ursprung übergeht. Es mag am Anfang des Studiums gar nicht so klar sein, worin der Nutzen dieser Symmetrien besteht, aber das wird hoffentlich an späterer Stelle deutlich. Ein Nutzen besteht darin, dass man ungerade und gerade Funktionen bei entsprechender Normierung als orthonormales Basissystem für beliebige Funktionen nehmen und man eine beliebige Funktion durch diese Basisfunktionen ausdrücken kann. Dieses Verfahren wird im zweiten Semester in der Elektrodynamik unter dem Begriff Fourierreihe eingeführt. Im ersten Semester kommt ein ähnliches Verfahren zur Anwendung, welches man als Taylorreihe bezeichnet und das an späterer Stelle noch genauer beschrieben wird. Es gibt noch weitere nützliche Eigenschaften, die im Studium hilfreich sind. So ist - als Vorwegnahme, da noch nicht über Integrale gesprochen wurde - beispielsweise das Integral einer ungeraden Funktion $f(x)$ über den gesamten Definitionsbereich der reellen Zahlen immer Null, d. h. es gilt

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 0. \tag{1.9}$$

1.1.5 Grenzwert und Stetigkeit



Beim Grenzwert oder Limes einer Funktion handelt es sich um einen Wert $f(x_0)$, dem sich die Funktionswerte $f(x)$ annähern, wenn man immer näher an die betrachtete Stelle x_0 herangeht. Die Schreibweise

$$\lim_{x \to x_0} f(x) = f(x_0) \tag{1.10}$$

drückt diesen Zusammenhang in kompakter Form aus. Betrachten wir dies am Beispiel der Funktion

$$f(x) = \frac{x^2 + 2x + 1}{5x^2 + 3} \tag{1.11}$$

und der Grenzwertbetrachtung für $x \to \infty$. Wir können die Funktion durch Ausklammern von x^2 in die Form

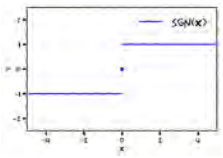
$$f(x) = \frac{1 + \frac{2}{x} + \frac{1}{x^2}}{5 + \frac{3}{x^2}} \tag{1.12}$$

bringen und dann die Grenzwertbetrachtung für die einzelnen Terme in Zähler und Nenner durchführen:

$$\lim_{x \to \infty} \frac{1 + \frac{2}{x} + \frac{1}{x^2}}{5 + \frac{3}{x^2}} = \frac{1}{5}. \tag{1.13}$$

Es gibt eine Vielzahl von Funktionen, bei denen die Grenzwertbildung nicht ganz so einfach durchzuführen ist. Betrachten wir als Beispiel die Signum-Funktion[1] $\operatorname{sgn}(x)$:

$$\operatorname{sgn}(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0 \end{cases} \tag{1.14}$$



Wir wollen nun den Grenzwert dieser Funktion an der Stelle $x = 0$ betrachten. Wir müssen uns also auf der Funktion auf diesen Wert zubewegen und stellen zunächst fest, dass es einen deutlichen Unterschied macht, ob man sich von der positiven oder der negativen Seite her annähert. Bei der Grenzwertbetrachtung kann es also eine Rolle spielen, von welcher Seite man kommt. Man spricht daher auch von

1 Auch Vorzeichenfunktion genannt (*signum*, lat. »Zeichen«)

links- und rechtsseitigen Grenzwerten. Der linksseitige Grenzwert der Signum-Fuktion beträgt

$$\lim_{x \to x_0^-} \operatorname{sgn}(x) = -1, \tag{1.15}$$

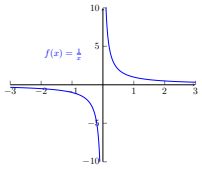
der rechtsseitige entsprechend

$$\lim_{x \to x_0^+} \operatorname{sgn}(x) = 1. \tag{1.16}$$

Man sieht an der Signum-Funktion aber auch, dass beide Grenzwerte für $x \to 0$ gar nicht mit dem Funktionswert $\operatorname{sgn}(0)$ an der Stelle $x = 0$ übereinstimmen. Die Signum-Funktion weist an dieser Stelle einen Sprung auf und spricht von einer Unstetigkeitsstelle.

Stetigkeit einer Funktion

Eine Funktion ist stetig an einer Stelle x_0, wenn ein Grenzwert an dieser Stelle existiert und dieser gleichzeitig mit dem Funktionswert identisch ist.

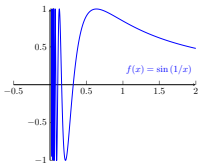


Unstetigkeitsstellen sind bei physikalischen Funktion oft mit Singularitäten verbunden, also Stellen, an denen die Funktionswerte z. B. über alle Grenzen wachsen kann. So hat beispielsweise die Funktion $f(x) = 1/x$ an der Stelle $x = 0$ eine Singularität, hier gilt

$$\lim_{x \to x_0^-} \frac{1}{x} = -\infty, \tag{1.17}$$

und

$$\lim_{x \to x_0^+} \frac{1}{x} = \infty. \tag{1.18}$$



Verkettet man die Funktion $f(x) = 1/x$ mit der Sinus-Funktion und betrachtet den rechts- oder linksseitigen Grenzwert für $x \to 0$, so hat man einen interessante Funktion vorliegen, die bei kleiner werdenden Argumenten immer schneller oszilliert. Da der Wert $x = 0$ nicht Teil des Definitionsbereichs ist, liegt dort in jedem Fall eine Unstetigkeitsstelle vor. Was in der näheren Umgebung von $x = 0$ passiert, kann man aber wegen des immer dichter werdenden Kurvenverlaufs gar nicht genau sagen.

1.1.6 Verkettung

Funktionen können über verschiedene Methoden so miteinander kombiniert werden, dass eine neue Funktion entsteht. Seien $f(x)$ und $g(x)$ zwei beliebige Funktionen, dann kann man z. B. durch Addition eine heute Funktion $h(x)$ erzeugen:

$$h(x) = f(x) + g(x). \tag{1.19}$$

Man kann natürlich die beiden Funktionen voneinander subtrahieren oder auch multiplizieren oder durcheinander teilen. Es dürfte klar sein, dass man bei diesen Operationen darauf achten muss, dass die Definitionsbereiche der beiden einzelnen Funktionen nicht mit dem Definitionsbereich der zusammengesetzten Funktion übereinstimmen muss. Neben dieser algebraischen Verknüpfung von Funktionen gibt es eine weitere Möglichkeit, eine neue Funktion zu erzeugen, nämlich über eine Verkettung. Wir betrachten dies am Beispiel der Funktionen $f(x) = x^2$ und $g(x) = \cos x$. Wir bezeichnen den Funktionswert von $f(x)$ als z, d. h. $z = f(x)$. Diesen Funktionswert nehmen wir dann als Argument der Funktion $g(x)$ und bezeichnen diesen Funktionswert als y, also $y = g(z)$. Es gilt also

$x \xrightarrow{f} z \xrightarrow{g} y.$

$$y = g(f(x)) = (g \circ f)(x), \tag{1.20}$$

und somit für die Funktionen aus dem Beispiel

$$(g \circ f)(x) = g(x^2) = \sin x^2. \tag{1.21}$$

Man kann sich mit Hilfe dieser beiden Funktion klar machen, dass die Reihenfolge der Verkettung eine Rolle spielt. Es gilt nämlich

$$(f \circ g)(x) = f(\sin(x)) = (\sin x)^2. \tag{1.22}$$

Die Verkettung von Funktionen ist also nicht kommutativ:

$$(f \circ g)(x) \neq (g \circ f)(x). \tag{1.23}$$

Beachten Sie, dass die Nicht-Kommutativität nicht bedeutet, dass es trotzdem zwei Funktionen geben kann, die bei wechselnder Reihenfolge der Verkettung zum gleichen Resultat kommen. Sie können dies am Beispiel $f(x) = 1/x$ und $g(x) = x^2$ überprüfen. Verkettete Funktionen treten in der Physik häufig auf, man sollte daher die Rechenregeln, die verkettete Funktionen betreffen, schnell sicher beherrschen lernen, v. a. was das Ableiten verketteter Funktionen betrifft.

1.1.7 Umkehrfunktion

Ein weiterer Weg, aus einer gegebenen Funktion $f(x)$ eine neue Funktion zu erschaffen, ist die Bildung der Umkehrfunktion $f^{-1}(x)$, die man durch Vertauschen von Argument und Funktionswert erhält. Für die Funktion $f(x) = x^2$ lautet dementsprechend die Umkehrfunktion $f^{-1}(x) = \sqrt{x}$. Eine wichtige Eigenschaft der Umkehrfunktion, die wir später bei Ableiten bestimmter Funktionen ausnutzen werden, betrifft die Verkettung mit ihrer Ausgangsfunktion. Es gilt: Die Verkettung von Funktionen ist also nicht kommutativ:

$$(f^{-1} \circ f)(x) = x. \tag{1.24}$$

Prüfen wir dies am gegebenen Beispiel:

$$(f^{-1} \circ f)(x) = \sqrt{f(x)} = \sqrt{x^2} = x. \tag{1.25}$$

Umkehrfunktion

Gibt es zu einer Funktion $f : D \to W$ eine Funktion $f^{-1} : W \to D$ mit den Eigenschaften

- $(f^{-1} \circ f)(x) = x$ für alle $x \in D$
- $(f \circ f^{-1})(y) = y$ für alle $y \in W$

so nennt man f^{-1} die Umkehrfunktion von f.

Bei der Umkehrfunktion muss man wieder auf den Definitionsbereich aufpassen. Es ist zudem so, dass nicht jede Funktion vollumfänglich umkehrbar ist. An unserem Beispiel $y = x^2$ kann man das gut erkennen. Die Funktion beschreibt eine Parabel und – abgesehen für $x = 0$ – gibt es zu jedem y-Wert immer genau zwei x-Werte. Dies spiegelt sich nach Vertauschen von x und y in der Beziehung $x = y^2$ bzw. $y = \pm\sqrt{x}$ wider. Man kann sich in solchen Fällen damit behelfen, dass man sich auf eines der beiden möglichen Vorzeichen festlegt (vgl. Abb. 1.3). Bei physikalisch-technischen Problemen kommt man immer wieder zu einer solchen Problematik, bei der man aus zwei möglichen Vorzeichen ein Vorzeichen auswählen muss. Hierzu muss man dann physikalisch vernünftig argumentieren, was aber gar nicht immer auf den ersten Blick so einfach sein muss. Vorzeichen gehören in der Physik erfahrungsgemäß zu den schwierigeren Aspekten.

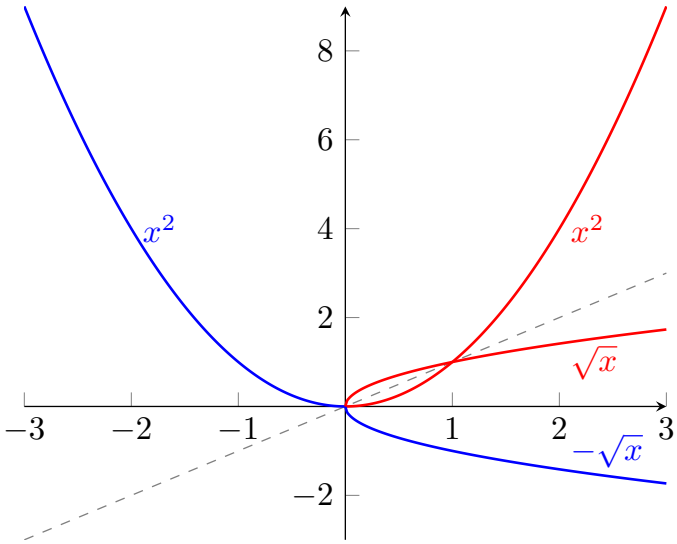


Abbildung 1.3: Die Funktion $f(x) = x^2$ und ihre Umkehrfunktionen $f(x) = \pm\sqrt[2]{x}$.

Es gibt Funktionen, die man ohne diese Problematik vollumfänglich umkehren kann, die streng monotonen Funktionen. Darunter versteht man Funktionen, die entweder nur wachsen oder fallen. Für den Fall einer streng monoton fallenden Funktion gilt für zwei Argumente x_1 und x_2 mit $x_1 < x_2$

$$f(x_1) > f(x_2), \tag{1.26}$$

für eine streng monoton steigende Funktion entsprechend

$$f(x_1) < f(x_2). \tag{1.27}$$

Wichtig ist hier die strenge Monotonie. Bei Funktionen, die nur monoton steigend oder fallen sind, bei denen also am Beispiel einer fallenden Funktion ein $\leq$ statt eines $<$ steht, ist diese vollumfängliche Umkehrbarkeit schon nicht mehr gegeben.

1.2 Wichtige Funktionen

Nachdem es bisher um allgemeine Eigenschaften von Funktionen gegangen ist, soll es nun um die wichtigsten Funktionen gehen, die einem immer wieder über den Weg laufen. Bei vielen dieser Funktionen wird auch gleich die Umkehrfunktion mitbehandelt und die elementaren Rechenregeln dieser Funktionen vorgestellt.

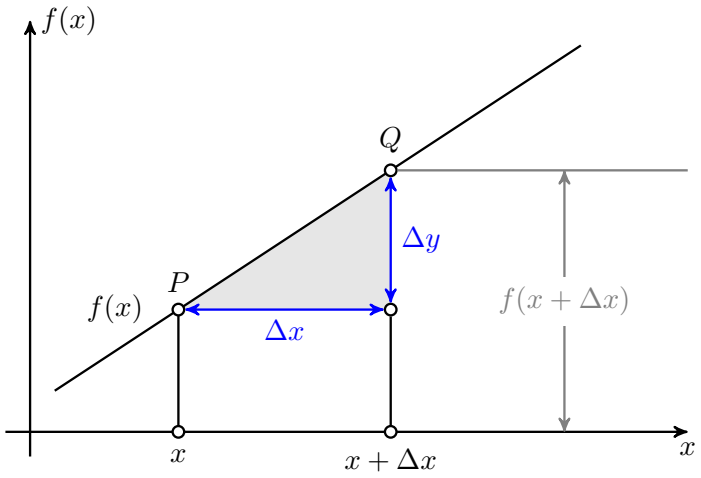


Abbildung 1.4: Die affine Funktion $f(x) = mx + b$.

1.2.1 Affine Funktionen

Affine Funktionen können in der Form

$$f(x) = mx + b \tag{1.28}$$

geschrieben werden. Hierbei sind m und b zwei Konstanten. Aus der Schule sollte Ihnen diese Schreibweise unter dem Begriff »Geradengleichung« bekannt sein. Die Größe m wird als Steigung der Geraden bezeichnet, also als Änderung der Geraden in y-Richtung pro zurückgelegter Änderung in x-Richtung, d. h. $m = \Delta y/\Delta x$. Die y-Koordinate des Schnittpunktes der Geraden entspricht der Größe b, was man sofort für $x = 0$ sieht (vgl. Abb. 1.4). Der Definitionsbereich der affinen Funktion ist $D = W = (-\infty, \infty)$.

Eine Gerade weist einen linearen Verlauf auf und man könnte geneigt sein, sich die Frage zu stellen, warum man nicht einfach »lineare Funktionen« zu dieser Art von Funktion sagt. Der Grund liegt darin, dass eine lineare Funktion in der Auslegung der Mathematik die Zahl Null auf die Null abbilden muss. Die lineare Funktion ist also ein Spezialfall der affinen Funktionen für $b = 0$. Für $b \neq 0$ spricht man daher streng genommen von linear affinen Funktionen, wobei im Alltagsgebrauch eigentlich mehr von »Geraden« gesprochen wird – zumindest in der Experimentalphysik.

1.2.2 Potenzen und Wurzeln

Rationale Funktionen

Bei den rationalen Funktionen hat man es mit Potenzen x^n zu tun, wobei dies abkürzend für

$$x^n = \overbrace{x \cdot x \cdot x \cdots x}^{n\text{-mal}} \tag{1.29}$$

steht. Für Potenzen sind folgende Rechenregeln relevant.

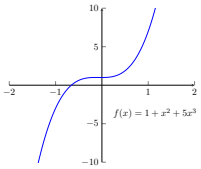
Rechenregeln: Potenzen

$$x^{-n} = \frac{1}{x^n} \qquad x^0 = 1$$

$$x^n \cdot x^m = x^{n+m} \qquad (x^n)^m = x^{n \cdot m}$$

POLYNOMFUNKTIONEN Eine Summe der Form

$$P_n(x) = \sum_{i=0}^{n} a_i x^i = a_0 x^0 + a_1 x^1 + a_2 x^2 + \ldots + a_n x^n \tag{1.30}$$

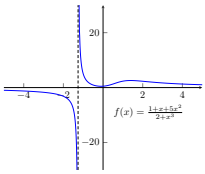


für $n \geq 0$ wird als Polynom n-ten Grades bezeichnet. Bezeichnet x eine Variable, dann spricht man von einer Polynumfunktion bzw. einer ganzrationalen Funktion. Eine Funktion 3. Grades wäre beispielsweise der Ausdruck $f(x) = 1 + x^2 + 5x^3$. Polynomfunktionen spielen in der Physik eine wichtige Rolle. Eine Anwendung ist z. B. das »Fitten« von Daten. Darunter versteht man das Bestimmen einer Funktion, die einen Verlauf von experimentellen Daten sinnvoll beschreiben, was oft durch Polynomfunktionen der Fall ist. Der Grund hierfür ist, dass man viele Funktionen in eine Form wie in Gl. 1.30 entwickeln kann, man spricht dann von einer Taylor-Entwicklung. Zudem gibt eine Reihe von Polynomen, die wichtige Differentialgleichungen der Physik lösen. So sind beispielsweise die Laguerre-Polynome Lösung der Schrödingergleichung für das Wasserstoffatom.

GEBROCHENRATIONALE FUNKTIONEN Eine Funktion, die man als Quotient zwei Polynomfunktionen in der Form

$$f(x) = \frac{P_n}{Q_m} = \frac{a_0 x^0 + a_1 x^1 + a_2 x^2 + \ldots + a_n x^n}{b_0 x^0 + b_1 x^1 + b_2 x^2 + \ldots + a_m x^m} \tag{1.31}$$

schreiben kann, wird allgemein als rationale Funktion bezeichnet. Hierbei bezeichnen n und m den Grad des jeweiligen Polynoms. Ist das Polynom im Nenner vom Grad $m = 0$, so hat man eine »normale« Polynomfunktion vorliegen. Für $m > 0$ hat man eine gebrochenrationale Funktion vorliegen, wobei man für $n < m$ von einer echt gebrochenrationalen Funktion spricht, bei $n > m$ von einer unecht gebrochenrationalen Funktion. Letztere kann durch Polynomdivision als Summe einer Polynomfunktion mit einer echt gebrochenrationale Funktion dargestellt werden.



MEHRERE VARIABLEN Rationale Funktionen sind nicht auf eine Variable beschränkt. Es ist in der Physik sogar die Regel, dass man es mit rationalen Funktionen mehrerer Variablen zu tun hat. So berechnet sich beispielsweise der Gesamtwiderstand R einer Parallelschaltung aus zwei Widerständen R_1 und R_2 gemäß

$$R = R(R_1, R_2) = \frac{R_1 R_2}{R_1 + R_2}. \tag{1.32}$$

1.2.3 Wurzelfunktionen

Abschließend soll es noch kurz um die Wurzelfunktionen gehen, also um die Umkehrfunktionen der Potenzfunktionen. Die Wurzelfunktion

$$f(x) = x^{\frac{1}{n}} = \sqrt[n]{x} \qquad (n \text{ ungerade}) \tag{1.33}$$

hat also die Potenzfunktion $f(x) = x^n$ als Umkehrfunktion. Bei geraden n muss man berücksichtigen, dass es bei der Potenzfunktion zwei x-Werte für einen y-Wert gibt, nämlich x und $-x$. Die Umkehrfunktion, die man durch einfaches Vertauschen von x und y gewinnt, würde also einem Wert ihres Definitionsbereichs zwei Elemente des Wertebereichs zuordnen, wäre also keine Funktion. Man muss daher den Definitionsbereich einschränken, um eine eindeutige Umkehrfunktion zu definieren. Üblicherweise gibt man die möglichen Umkehrfunktionen für die verschiedenen Definitionsbereiche als Fallunterscheidung an.

Wir verdeutliche dies am Beispiel der Funktion $f(x) = x^2$. Betrachten wir nur den positiven Ast dieser Funktion, so entspricht der Definitionsbereich der Menge der positiven reellen Zahlen $\{x \in \mathbb{R} \;/\; x \geq 0\}$. Da bei Bildung der Umkehrfunktion Definitionsbereich und Wertebereich vertauschen,[2] entspricht der Wertebereich für die zugehörige

2 Der Definitionsbereich der Ausgangsfunktion ist der Wertebereich der Umkehrfunktion. Der Wertebereich der Ausgangsfunktion ist der Definitionsbereich der Umkehrfunktion.

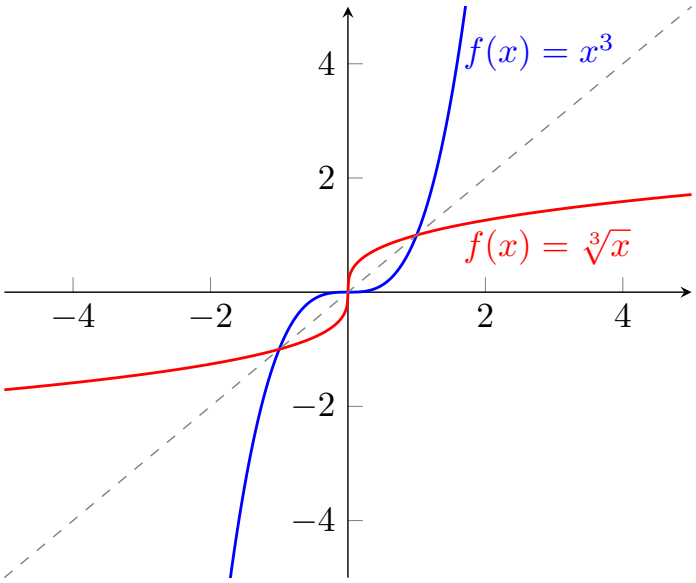


Abbildung 1.5: Die Funktion $f(x) = x^3$ und ihre Umkehrfunktion $f(x) = \sqrt[3]{x}$.

Umkehrfunktion dem Definitionsbereich der Ausgangsfunktion, also den positiven reellen Zahlen. Da der Wertebereich der Ausgangsfunktion ebenfalls den positiven reellen Zahlen entspricht, hat die Umkehrfunktion nur die positiven reellen Zahlen als Definitionsbereich. Man kann also die Umkehrfunktion für diesen Teil der Parabel problemlos bilden.

Der linke Parabelast hat als Definitionsbereich die negativen reellen Zahlen, die zugehörige Umkehrfunktion entsprechend diesen Wertebereich. Alle Werte der Umkehrfunktionen müssen also kleiner als Null sein. Der Wertebereich der Ausgangsfunktion entspricht den positiven reellen Zahlen, ist also der Definitionsbereich der Umkehrfunktion. Wir haben für beide Parabeläste also den gleichen Definitionsbereich, nämlich $\{x \, \epsilon \, \mathbb{R} \; / \; x \geq 0\}$. Das entspricht der Tatsache, dass es für reelle Zahlen keine Lösung für $\sqrt{x}$ gibt, wenn $x < 0$ ist. Wir fassen zusammen:

$$f^{-1}(x) = \begin{cases} \sqrt{x} & (\text{rechterParabelast}) \\ -\sqrt{x} & (\text{linkerParabelast}) \end{cases} \tag{1.34}$$

Schauen wir uns nun an, was bei ungeraden Exponenten passiert: als Beispielfunktion nehmen wir $f(x) = x^3$. Betrachtet man den Graphen der Funktion, so sieht man, dass es sich um eine streng monoton steigende Funktion handelt. Die Funktion ist also umkehrbar, ohne dass man sich um eine Fallunterscheidung Gedanken machen muss. Die

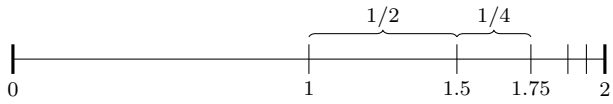


Abbildung 1.6: Eine Summe unendlich vieler positiver Glieder, die alle von Null verschiedenen sind, kann dennoch auf einen endlichen Wert konvergieren. Die Reihe $\sum_{i=0}^{\infty} 1/2^i$ konvergiert beispielsweise gegen den Wert 2. In der Summe wird immer die Hälfte des vorigen Beitrages addiert. Je näher man der Zwei kommt, desto kleiner werden die dazukommenden Beiträge, bis sie im Grenzfall $i \to \infty$ verschwinden.

zugehörige Umkehrfunktion lautet $f^{-1}(x) = \sqrt[3]{x}$ für alle $x \in \mathbb{R}$. Aber Achtung: Wir haben bei der Umkehrfunktion zugelassen, dass negative Zahlen unter der (dritten) Wurzel stehen. Das ist prinzipiell möglich, wenn man vorsichtig vorgeht, es produziert aber auch Widersprüche an anderer Stelle. Betrachten wir die Gleichung $-2 = \sqrt[3]{-8}$ und quadrieren unter der Wurzel und ziehen gleichzeitig die Quadratwurzel, so erhält man

$$-2 = \sqrt[3\cdot 2]{(-8)^2} = \sqrt[6]{64} = \sqrt[6]{(8)^2} = \sqrt[3]{(8)} = 2. \tag{1.35}$$

Man kann diesen Widerspruch verhindern, wenn man für negative x-Werte der Umkehrfunktion Beträge verwendet. Die Umkehrfunktion wird daher in einigen Lehrbüchern als

$$f^{-1}(x) = \begin{cases} \sqrt[3]{x} & \text{für } x \geq 0 \\ -\sqrt[3]{|x|} & \text{für } x < 0 \end{cases} \tag{1.36}$$

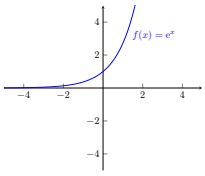
geschrieben.

1.2.4 Exponentialfunktionen

Eine Funktion der Form

$$f(x) = a^x \qquad (a > 0) \tag{1.37}$$

wird als Exponentialfunktion bezeichnet. Man bezeichnet a als Basis und x als Exponenten. Man kann als Basis jede beliebige Zahl einsetzen, die größer als Null ist. Übliche Basen sind 10 sowie die irrationale Eulersche Zahle $\mathrm{e} = 2.7182\ldots$ Die Eulersche Zahl lässt sich als Reihe der Form



$$\mathrm{e} = 1 + \frac{1}{1} + \frac{1}{2} + \frac{1}{3\cdot 2} + \frac{1}{4\cdot 3\cdot 2} + \ldots \tag{1.38}$$

schreiben. Die Funktion e^x ergibt sich daraus zu

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots . \tag{1.39}$$

Das Ausrufezeichen steht für eine spezielle mathematische Operation, die Fakultät und ist $n!$ eine abkürzende Schreibweise für

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 1. \tag{1.40}$$

Die Reihendarstellung in Gl. 1.39 hat unendlich viele Terme. Alle diese Terme sind von Null verschieden (und größer als Null). Häufig wird in diesem Zusammenhang die Frage gestellt, wie eine Summe unendlich vieler Beiträge einen endlichen Wert haben kann. Man muss hier beachten, dass die Beiträge der Fakultäten ebenfalls anwachsen werden und irgendwann größer als die Potenzwerte im Zähler werden. Je nach x-Wert gibt es irgendwann nur noch Beiträge, die kleiner als Eins sind und immer kleiner werden, bis sie letztlich gegen Null konvergieren. Die Tatsache, dass eine unendliche Summe immer kleiner werdender Beiträge einen endlichen Wert ergibt, kann man sich an der Reihe

$$\sum_{i=0}^{\infty} \frac{1}{2^i} = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots = 2 \tag{1.41}$$

mit Hilfe eines Zahlenstrahls (vgl. Abb. 1.6) verdeutlichen. Ein mathematischer Beweis könnte folgendermaßen aussehen: Wir nehmen an, dass diese Summe den endlichen Wert S ergeben soll. Nun multiplizieren wir diese Summe mit Zwei:

$$2S = \sum_{i=0}^{\infty} 2 \cdot \frac{1}{2^i} = 2 + 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots \tag{1.42}$$

Man sieht, dass diese Reihe bis auf den ersten Term mit der ursprünglichen Reihe identisch ist. Bildet man die Differenz $2S - S$, so erhält man das gesuchte Ergebnis $S = 2$.

Die Reihe aus Gl. 1.41 ist auch als das Paradoxon von Zeno bekannt. Es geht in etwa so: ein Pfeil wird auf ein Ziel geschossen. Zunächst muss der Pfeil die Hälfte der Strecke zurücklegen. Dann die nächste Hälfte der Hälfte, also ein Viertes der Gesamtstrecke. Dann wieder die Hälfte der übrig gebliebenen Strecke usw. Der Pfeil wird sein Ziel nach dieser Argumentation niemals erreichen können, da es immer noch ein kleines Teilstück geben wird, das zurückgelegt werden muss.[3] Natürlich ist das Unsinn, aber es verdeutlicht doch das Bestreben der Mathematik, widerspruchsfreie Argumentationen zu erschaffen. Es hätte dem Schöpfer des Paradoxon auffallen können, dass man bereits die erste zurückgelegt

3 Das Paradoxon von Zeno entspricht inhaltlich der Argumentation beim Wettrennen von Achilles mit der Schildkröte, die einen kleinen Vorsprung hat.

Hälfte in kleinere Intervalle hätte aufteilen können, wodurch diese erste Hälfte auch niemals erreicht worden wäre, was beim Paradoxon aber stillschweigend angenommen wird.

Ein typisches Verhalten von Exponentialfunktionen ist das schnelle Wachstum, das sich irgendwann immer gegenüber anderen Funktionen, die ebenfalls wachsen, durchsetzt. Man kann zeigen, dass

$$\lim_{x\to\infty} \frac{a^x}{x^n} = \infty \tag{1.43}$$

gilt. Die Exponentialfunktion »gewinnt« am Ende immer gegen eine Potenzfunktion, egal wie groß das n ist. Folgende Regeln sind bei Exponentialfunktionen relevant:

Rechenregeln: Exponentialfunktionen

$$a^0 = 1 \qquad a^x \cdot a^y = a^{x+y}$$

$$\frac{a^x}{a^y} = a^{x-y} \qquad (a^x)^y = a^{xy}$$

$$a^{-x} = \frac{1}{a^x}$$

Warum kommen Exponentialfunktionen eigentlich so häufig in den naturwissenschaftlichen Fächern vor? Vielleicht haben Sie sich schon einmal gefragt, warum z. B. die Auf- und Enladekurve eines Kondensators oder der radiaktive Zerfall durch eine e-Funktion beschrieben werden. Gibt es vielleicht ein übergeordnetes Prinzip, dass diese beiden völlig verschiedenen Phänomene zu Grunde liegt? Die Ursache liegt darin, dass Exponentialfunktionen Veränderungen einer physikalischen Größe beispielsweise mit der Zeit t beschreiben, die vom Wert der physikalischen Größe zum betrachteten Zeitpunkt t abhängen. Schauen wir uns das am radioaktiven Zerfall an: Wir haben eine bestimmte Menge an Plutonium zur Zeit t_0 vorliegen. Die Menge an Plutonium können wir durch die Teilchenzahl N beschreiben, die Menge zur Zeit t_0 sei N_0. In einem kleinen Zeitintervall Δt – so sagt es uns das Experiment – wird die Menge an Plutoniumteilchen um ΔN abnehmen und diese Abnahme ist proportional zum Produkt $N_0 \cdot \Delta t$. Zu jeder anderen Zeit gilt genau der gleiche Zusammenhang. Die Abnahme der Teilchenzahl ist immer proportional zur gerade vorhandenen Teilchenzahl N und zum Zeitintervall Δt, d. h.

$$\Delta N \propto N \cdot \Delta t. \tag{1.44}$$

Aus physikalischer Sicht muss man jetzt nur noch die Proportionalitätskonstante finden und beachten, dass man noch ein Minuszeichen braucht, da die Teilchenzahl abnimmt. Als Vorgriff auf die Kapitel über Differential- und Integralrechnung sei hier schon gesagt, dass man bei kleinen Änderungen statt des ΔN ein $\mathrm{d}N$ schreibt, analog $\mathrm{d}t$ für Δt. Falls Ihnen die folgenden Argumentationen nicht einleuchten, arbeiten Sie erst die Grundlagen für das Ableiten und Integrieren von Funktionen durch und gehen Sie dann nochmal durch den Text, es könnte sich lohnen. Dieses infinitesimale Intervall $\mathrm{d}N$ entspricht dem $\mathrm{d}x$, das vermutlich beim Integrieren in der Schule vorgekommen sein dürfte und dessen näherer Sinn erfahrungsgemäß vielen nicht ganz klar ist, außer dass es sagt, nach welcher Variablen man integriert. Gleiches gilt natürlich auch für das $\mathrm{d}t$, denn es gibt keinen Grund, warum die eine Größe relevanter sein sollte als die andere. Bleiben wir hier auf Schulniveau und deuten $\mathrm{d}N$ und $\mathrm{d}t$ genau als die Größen, nach denen man integrieren muss, in unserem Beispiel also N und t. Bezeichnen wir die Proportionalitätskonstante noch mit k, so erhält man

$$\mathrm{d}N = -k \cdot N \cdot \mathrm{d}t. \tag{1.45}$$

Bringt man das N auf die linke Seite und integriert beide Seiten über die jeweilige Größe, so steht da

$$\int_{N_0}^{N} \frac{\mathrm{d}N'}{N'} = -k \int_0^t \mathrm{d}t'. \tag{1.46}$$

Das Integral auf beiden Seiten ausgeführt ergibt:

$$\ln \frac{N}{N_0} = -kt. \tag{1.47}$$

Wenn man jetzt nach N auflöst (dazu braucht man die Umkehrfunktion des *Logarithmus naturalis*, also die e-Funktion), so erhält man das Gesetz des radioaktiven Zerfalls:

$$N(t) = N_0 \exp\{-kt\}. \tag{1.48}$$

Wir haben hier die Darstellung $\exp\{-kt\}$ statt e^{-kt} verwendet. Beide Schreibweisen sind gleichwertig, übersichtlicher ist erstere.

Wie sieht das nun beim Kondensator aus? Wir betrachten einen zunächst ungeladenen Kondensator, der zur Zeit t_0 an eine Spannungsquelle angeschlossen wird. Wir wissen aus Experimenten, dass der Kondensator sich auflädt. Wir können beispielsweise den elektrischen Strom beim Aufladen messen, oder das Magnetfeld, das in der Umgebung der elektrischen Leitungen entsteht usw. Wir wissen auch, dass sich

gleichnamige Ladungen abstoßen. Die Spannungsquelle »schiebt« nun elektrische Ladungen auf die Kondensatorplatten. Wir betrachten hier nur die Platte, die sich positiv auflädt, für die negative Platte gilt genau das gleiche. Die erste positive Ladung kann ohne große Anstrengung auf die Kondensatorplatte fließen. Bei der zweiten Ladung sieht es schon anders aus, da ja bereits eine positive Ladung auf der Platte sitzt und diese eine Kraft auf die zweite Ladung ausübt, die abstoßend wirkt. Die dritte Ladung hat es wiederum etwas schwerer, auf die Platte zu kommen, da dann bereits zwei Ladungen dort sitzen und sich die abstoßende Kraft verdoppelt hat. Diese Argumentation kann man fortsetzen, bis keine Ladung mehr auf die Platte gelangt, da die Spannungsquelle nicht die dafür benötigte Energie bereitstellen kann. Die Zunahme der Ladungsmenge $\mathrm{d}Q$ auf der Kondensatorplatte wird also mit der Zeit t immer kleiner und ist zu jeder Zeit proportional zur Ladungsmenge Q auf der Platte. Es gilt also wieder $\mathrm{d}Q = -k \cdot Q \cdot \mathrm{d}t$ und somit der gleiche funktionale Zusammenhang:

$$Q(t) = Q_0 \exp\{-kt\}. \tag{1.49}$$

Aus Gl. 1.49 erhält man durch Ableiten den bekannten Ausdruck für den Aufladestrom:

$$I(t) = I_0 \exp\{-kt\}. \tag{1.50}$$

Es gibt nun eine Vielzahl von weiteren physikalischen Größen, die einer solchen Abhängigkeit folgen. Wenn Sie mit einer Vakuumpumpe eine Kammer evakuieren, dann hängt die Menge Gasteilchen, die zur Zeit t abgepumpt werden können, von der vorhanden Anzahl der Teilchen N zu dieser Zeit in der Kammer ab. Also folgt die Abnahme des Drucks, der proportional zur Teilchenzahl ist, dem bekannten Ausdruck

$$p(t) = p_0 \exp\{-kt\}. \tag{1.51}$$

Es kann sein, dass diese Gesetzmäßigkeit noch verfeinert werden muss – bei der Vakuumpumpe sind wir davon ausgegangen, dass die Pumpe immer mit der gleichen Saugleistung arbeitet, was in der Realität nicht der Fall ist –, die grundlegende Abhängigkeit bleibt aber bestehen. Die zugehörigen Konstanten k ergeben sich aus dem physikalischen Problem. Man kann hier ausnutzen, dass in Logarithmen und e-Funktionen immer eine Größe der Dimension Zahl stehen muss.[4]

4 Früher hat man von einer dimensionslosen Größe gesprochen, aber dies ist nach EN ISO 80000 veraltet. Das Internationale Einheitensystem (SI) ordnet jeder physikalischen Größe eine Dimension zu. Eine »dimensionslose« Größe ist mit dem SI daher nicht vereinbar. Im Laborjargon und auch in den Vorlesungen werden Sie diesen Ausdruck jedoch häufig zu hören bekommen. Das hängt damit

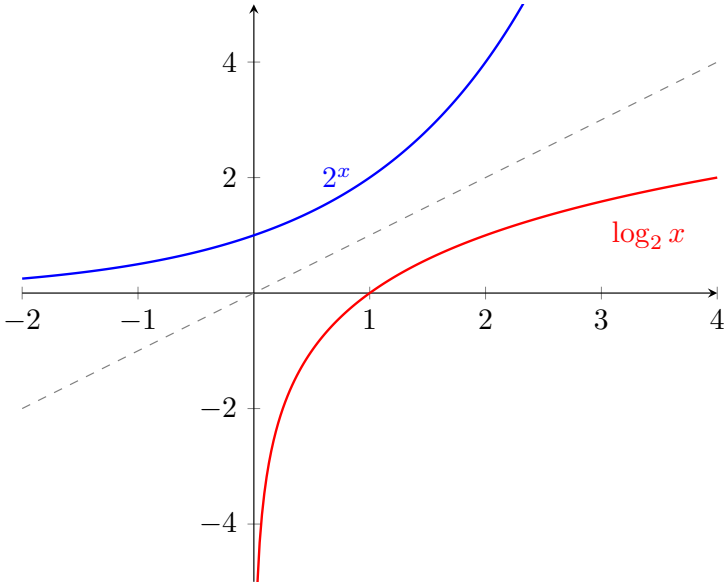


Abbildung 1.7: Graphische Darstellung der Funktionen $f(x) = 2^x$ und $f(x) = \log_2 x$.

1.2.5 Logarithmusfunktionen

Die Umkehrfunktion von $y = a^x$ wird als Logarithmus von y zur Basis a bezeichnet (vgl. Abb. 1.7):

$$x = \log_a y. \tag{1.52}$$

Wir haben hier eine eindeutige Umkehrfunktion vorliegen, d. h. brauchen uns nicht weiter um Vorzeichen zu kümmern. Aus $a^0 = 1$ folgt sofort die wichtige Beziehung $\log_a 1 = 0$. Während jede Exponentialfunktion irgendwann einmal eine beliebige Potenzfunktion übersteigt, ist es beim Logarithmus genau umgekehrt:

$$\lim_{x\to\infty} \frac{\log_a x}{x^n} = 0 \tag{1.53}$$

Der Logarithmus »verliert« am Ende immer. Zwei spezielle Logarithmen sind allgegenwärtig in der Physik, der dekadische Logarithmus $\log_{10}$ zur Basis 10 sowie der natürliche Logarithmus $\log_e = \ln$ zur Basis e. Folgende Rechenregeln sind anzuwenden:

zusammen, dass Lehrende an Universitäten häufig die Begrifflichkeiten verwenden, die sie selbst einmal gelernt haben und Normen meist nicht so ernst genommen werden, wie sie sollten.

Rechenregeln: Logarithmen

$$\log(x \cdot y) = \log x + \log y \qquad \log x^y = y \cdot \log x$$
$$\log \frac{x}{y} = \log x - \log y \qquad \log 1 = 0$$

Man kann sich diese Rechenregeln aus den Potenzgesetzen herleiten, wenn man ausnutzt, dass eine Funktion verknüpft mit ihrer Umkehrfunktion gerade dem Argument der Funktion entspricht:

$$x = a^{\log_a(x)} = \log_a\left(a^x\right) \tag{1.54}$$

Daraus folgt:

$$\begin{aligned} \log_a(x \cdot y) &= \log_a\left(a^{\log_a(x)} \cdot a^{\log_a(y)}\right) \\ &= \log_a\left(a^{\log_a(x)+\log_a(y)}\right) \\ &= \log_a(x) + \log_a(y). \end{aligned} \tag{1.55}$$

Das Beweisen der anderen Rechenregeln erfolgt analog. Die gleiche Methode ermöglicht es auch, einen Logarithmus einer bestimmten Basis a durch eine andere Basis b auszudrücken:

$$\begin{aligned} \log_a(x) &= \log_a\left(b^{\log_b(x)}\right) \\ &= \log_b(x) \cdot \log_a(b). \end{aligned} \tag{1.56}$$

Im letzten Schritt haben wir die Beziehung $\log x^y = y \cdot \log x$ ausgenutzt. Diese kann man aus dem Potenzgesetz $(a^m)^n$ herleiten. Zunächst folgt daraus

$$\log_a\left(a^m\right)^n = \log_a\left(a^{mn}\right) = m \cdot n. \tag{1.57}$$

Sei nun $a^m = x$, also $m = \log_a x$, und $n = y$, so folgt der zu beweisende Zusammenhang.

1.2.6 Winkelfunktionen

Die Winkelfunktionen Sinus und Kosinus spielen eine zentrale Rolle in der Physik bei der Beschreibung von periodischen Vorgängen wie Schwingungen und Wellen, da sie eine Periodizität der Form $\sin(x + 2\pi) = \sin x$ bzw. $\cos(x + 2\pi) = \cos x$ aufweisen. Grenzwerte für $x \to \infty$ sind für Sinus und Cosinus nicht definiert. Tangens und Kotangens sind aus Sinus und Kosinus abgeleitete Winkelfunktionen der Form

$$\tan x = \frac{\sin x}{\cos x} \qquad \cot x = \frac{\cos x}{\sin x}. \tag{1.58}$$

Tabelle 1.2: Winkel in Grad und im Bogenmaß

Grad	360°	180°	90°	45°	60°	30°
Bogenmaß	2π	π	$\pi/2$	$\pi/4$	$\pi/3$	$\pi/6$

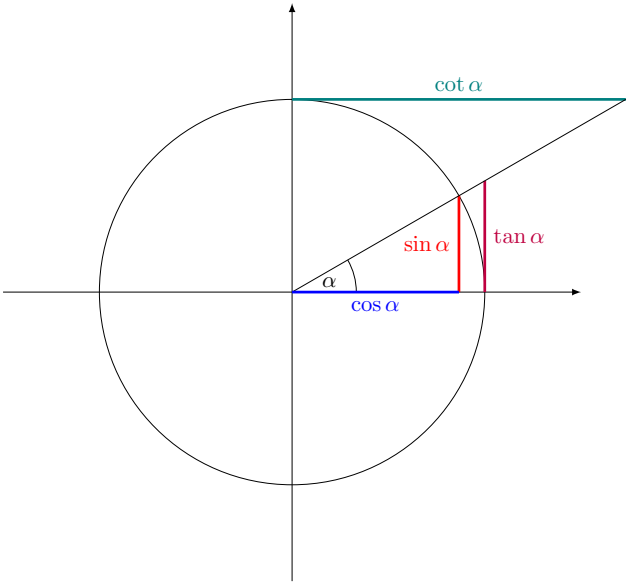


Abbildung 1.8: Definition der Winkelfunktionen Sinus, Kosinus, Tangens und Kotangens mit Hilfe des Einheitskreises.

Tangens und Kotangens sind ebenfalls periodische Funktionen, haben allerdings eine kürzere Periode von π statt 2π. Es gibt mehrere Möglichkeiten, einen Winkel anzugeben. Aus der Schule dürfte die Angabe in Grad bekannt sein. In der Physik wird sehr häufig auf das Bogenmaß zurückgegriffen. Die Idee hierbei ist, dass die Länge eines Bogens auf dem Einheitskreis ein Maß für den Winkel α des Bogens ist.

Die Definition der Winkelfunktionen über den Einheitskreis ist anschaulich (vgl. Abb. 1.8), allerdings wird auf universitärem Niveau häufig auf andere Definitionen zurückgegriffen, die es erleichtern, bestimmte Eigenschaften dieser Funktionen zu beweisen. Eine mögliche Definition läuft über die Darstellung mit Hilfe komplexer Exponentialfunktionen, was wir aber erst nach Einführung der komplexen Zahlen besprechen

wollen. Eine weitere wichtige Definition ist die Darstellung als Reihe. Es gilt:

$$\begin{aligned} \sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots \\ \cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots \end{aligned} \tag{1.59}$$

Man sieht an Gl. 1.59, dass der Sinus sich nur aus Beiträgen ungerader Potenzen zusammensetzt, der Kosinus analog nur aus geraden Potenzen.Das bedeutet, dass es sich – wie wir bereits in Abschnitt 1.1.4 festgestellt haben – beim Sinus um eine ungerade Funktion handelt, beim Kosinus um eine gerade Funktion. An der Reihendarstellung sieht man auch, dass man für kleine x-Werte eine Näherung machen kann, die Ihnen öfters über den Weg laufen wird. Für kleine Argumente gilt in sehr guter Näherung die Beziehung $\sin x \approx x$ und $\cos x \approx 1$. Man spricht hier auch von der Kleinwinkelnäherung, die es Ihnen beispielsweise überhaupt erst erlaubt, bestimmte schwingende Systeme wie das Fadenpendel analytisch zu lösen, dann aber eben nur für kleine Auslenkungen.

Es sollen nun noch ein paar wichtige Anmerkungen zu den Winkelfunktionen gemacht werden. Der Einheitskreis ist definiert als ein Kreis mit Radius Eins. Dies bedeutet, dass für jeden Punkt (x, y) auf dem Einheitskreis die Beziehung $x^2 + y^2 = 1$ erfüllt sein muss. Nun stehen Sinus und Kosinus für nichts anderes als diese beiden Koordinaten für einen bestimmten Punkt ausgedrückt durch den Winkel α. Dies bedeutet, dass

$$\sin^2 \alpha + \cos^2 \alpha = 1 \tag{1.60}$$

für jeden Winkel α erfüllt sein muss. Beachten Sie, dass $\sin^2 x$ eine abkürzende Schreibweise für $(\sin x)^2$ ist. Gleichung 1.60 verknüpft als Sinus und Kosinus und erlaubt es, eine Reihe weiterer nützlicher Verknüpfungen der Winkelfunktionen wie beispielsweise

$$\sin x = \frac{\tan x}{\sqrt{1 + \tan^2 x}}. \tag{1.61}$$

Häufig hat man es auch mit Winkelfunktionen zu tun, die zwei Terme im Argument haben. Hier finden die sogenannten Additionstheoreme ihre Anwendung, die man in Tabellenwerken nachschlagen kann. Hier seien stellvertretend für die Additionstheoreme

$$\begin{aligned} \sin x + y &= \sin x \cdot \cos y + \sin y \cdot \cos x \\ \cos x + y &= \cos x \cdot \cos y - \sin x \cdot \cos y \end{aligned} \tag{1.62}$$

als Beispiel genannt.

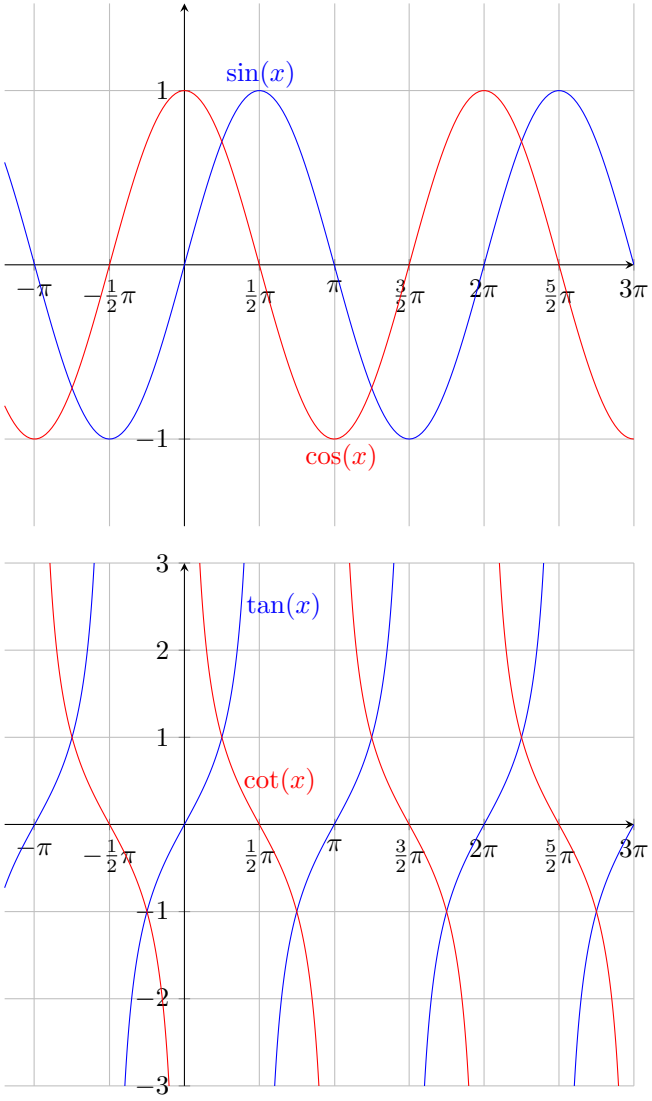


Abbildung 1.9: Graphische Darstellung der Winkelfunktionen.

1.2.7 Arcusfunktionen

Die Arcusfunktionen[5] arcsin, arccos und arctan sind die Umkehrfunktionen der Winkelfunktionen sin, cos und tan. Die Winkelfunktionen Sinus und Kosinus sind für alle $x \in \mathbb{R}$ definiert, bilden aber als periodische Funktionen nur auf das Intervall $[-1, 1]$ ab, wodurch es ein Problem mit der Eindeutigkeit der Umkehrfunktion gibt. Es ist Konvention, für die Umkehrfunktion des Sinus nur das Intervall $[-\pi/2, \pi/2]$ als Defi-

5 *arcus* lat. Bogen

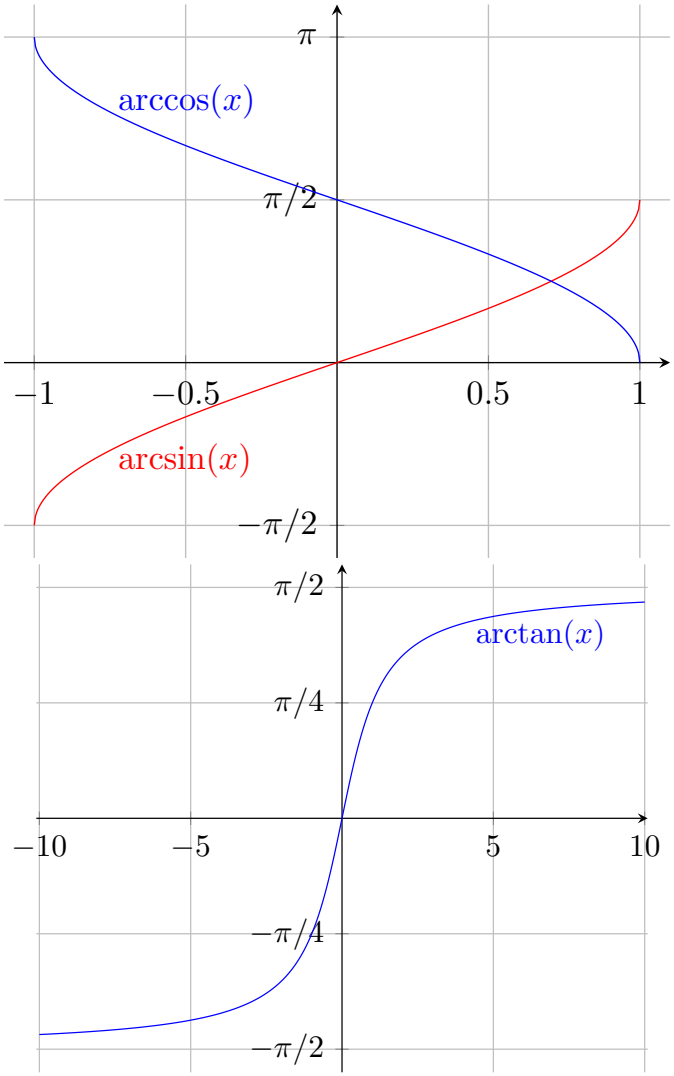


Abbildung 1.10: Graphische Darstellung der Arcusfunktionen.

nitionsbereich zu betrachten. Der Sinus ist in diesem Intervall streng monoton. Dies bedeutet, dass der Wertebereich der Umkehrfunktion, also des Arcussinus diesem Intervall entspricht. Für den Kosinus und den Tangens verwendet man aus dem gleichen Grund das Intervall $[0, \pi]$. Als Definitionsbereich wird für beide Umkehrfunktionen das Intervall $[-1, 1]$ verwendet.

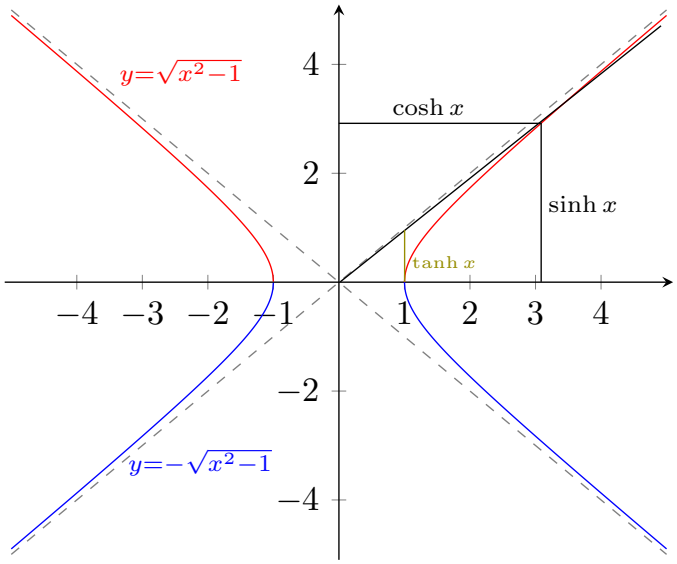


Abbildung 1.11: Graphische Darstellung der hyperbolischen Funktionen als Schnitte von Geraden an der Einheitshyperbel. Zeichnet man ausgehend vom Ursprung eine Gerade auf einen beliebigen Punkt der Einheitshyperbel entsprechen $\sinh(x)$ und $\cosh x$ der senkrechten und horizontalen Projektion auf x- und y-Achse. Der $\tanh x$ ergibt sich als Länge der vertikalen Linie ausgehend vom Scheitel der Hyperbel, die die vom Ursprung ausgehende Gerade schneidet.

1.2.8 Hyperbolische Funktionen

Die trigonometrischen Funktionen haben wir als Schnitte von Geraden mit dem Einheitskreis $x^2 + y^2 = 1$ darstellen werden (vgl. Abb. 1.8). Die hyperbolischen Funktionen $\sinh(x)$, $\cosh x$, $\tanh x$ und $\coth x$ können auch als Schnitte dargestellt werden, allerdings mit der Einheitshyperbel $x^2 - y^2 = 1$. Die beiden Zweige der Einheitshyperbel sind folglich durch die Beziehungen $y = \pm\sqrt{x^2 - 1}$ gegeben. Die Konstruktion ist in Abb. 1.11 gezeigt.

Eine interessante Eigenschaft der hyperbolischen Funktionen ist ihre Darstellung als Kombination von e-Funktionen:

Hyperbolische Funktionen

$$\sinh x = \frac{e^x - e^{-x}}{2} \qquad \cosh x = \frac{e^x + e^{-x}}{2}$$

$$\tanh x = \frac{\sinh x}{\cosh x} \qquad \coth x = \frac{\cosh x}{\sinh x}$$

Abbildung 1.12: Ein an zwei Enden fixiertes durchhängendes Seil im Schwerefeld der Erde folgt dem Kosinus hyperbolicus (vgl. Abb. 1.13). Man spricht hier auch von einer Kettenlinie oder Katenoide. Bildquelle: Wikipedia

Für die trigonometrischen Funktionen gibt es eine recht ähnliche Darstellung über e-Funktionen, allerdings braucht man dafür die komplexen Zahlen, die wir noch nicht eingeführt haben. Wir werden dies an entsprechender Stelle nachholen.

Die hyperbolischen Funktionen mögen zunächst vielleicht etwas exotisch erscheinen, sie haben aber eine Reihe von wichtigen Anwendungen in der Physik. Zunächst sei gesagt, dass viele Phänomene in der Physik durch e-Funktionen beschrieben werden können, somit sind hyperbolische Funktionen aus dieser Perspektive schon einmal interessant. Sie tauchen beispielsweise in der Relativitätstheorie bei der Beschreibung des Lorentz-Boosts auf. Es gibt zudem eine Übereinstimmung des Verlaufs des Kosinus hyperbolicus mit einem durchhängenden Seil, das an seinen beiden Enden im Schwerefeld der Erde fest eingespannt wird.

Alle vier hyperbolischen Funktionen sind in Abbildung 1.13 graphisch dargestellt. Der Kosinus hyperbolicus ähnelt auf dem ersten Blick einer Parabel, ist aber keine, da er deutlich schneller wächst. Betrachtet man seine Definition als Summe zweier Exponentialfunktionen, so sieht man, dass er sich für sehr große x-Werte der Funktion $0.5 \cdot \exp\{x\}$ annähert. Wie wir bereits erwähnt haben, wächst eine Exponentialfunktion immer schneller als eine beliebige Polynomfunktion.

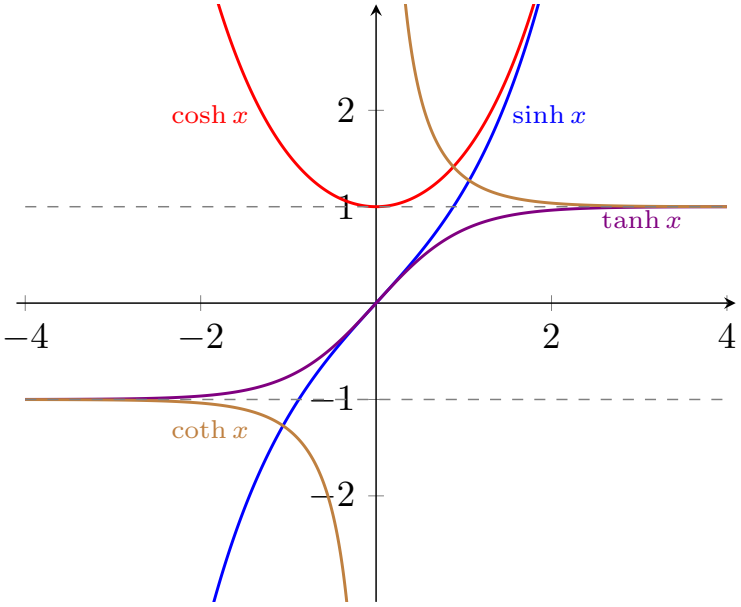


Abbildung 1.13: Graphische Darstellung der hyperbolischen Funktionen. Die Funktionen $\sinh x$ und $\cosh x$ nähern sich für größer werdende x-Werte von der jeweils anderen Seite einer gemeinsamen Asymptote.

Aus der Definitionsgleichung $x^2 - y^2 = 1$ kann man über Pythagoras die wichtige Beziehung

$$\sinh^2 \alpha = 1 - \cosh^2 \alpha \tag{1.63}$$

herleiten, die das Analogon zur Beziehung $\sin^2 \alpha = 1 - \cos^2 \alpha$ darstellt.

1.2.9 Areafunktionen

Die Umkehrfunktionen der hyperbolischen Funktionen bezeichnet man als Areafunktionen[6]. Wir wollen diese hier nicht weiter im Detail besprechen, nur darauf hinweisen, dass aus der Definition der hyperbolischen Funktion als Summe von e-Funktionen die Areafunktionen durch den natürlichen Logarithmus dargestellt werden können:

6 *area*, lat. Fläche

Areafunktionen

$$\operatorname{arsinh} x = \ln x + \sqrt{x^2 + 1} \qquad \operatorname{arcosh} x = \ln x + \sqrt{x^2 - 1}$$

$$\operatorname{artanh} x = \frac{1}{2} \ln \frac{1 + x}{1 - x}$$

Wir werden auf die trigonometrischen und hyperbolischen Funktionen sowie deren Umkehrfunktionen nochmal im Zusammenhang mit den komplexen Zahlen und der Eulerschen Formel zurückkommen.

1.3 Vektoren

In der Physik gibt es skalare Größen, die nur durch einen Zahlenwert gekennzeichnet sind. Das kann z. B. die Temperatur T eines Gases sein oder die Masse eines Himmelskörpers. Es gibt aber auch Größen, denen man intuitiv eine Richtung zuweisen würde, beispielsweise einer Geschwindigkeit. Hier kommt es offensichtlich nicht nur darauf an, wie schnell man fährt, sondern auch wohin. Wir wollen diese Alltagserfahrung für die Definition von Vektoren nutzen:

Definition: Vektoren

Unter einem Vektor versteht man eine gerichtete Größe, die durch eine Richtung und durch eine Länge (Betrag) gekennzeichnet ist.

Die Zeit t ist eine Größe, bei der man ein wenig aufpassen muss, denn sie hat eine Richtung. In der statistischen Physik, die in den höheren Semestern gelehrt wird, bestimmt die Zunahme der Entropie die Richtung der Zeit. Sie unterscheidet sich daher gar nicht so sehr von einer beliebigen Raumrichtung, die man z. B. durch die x-Achse beschreiben kann, mit dem Unterschied, dass die Zeit nur in eine Richtung läuft. Es ist daher vielleicht gar nicht verwunderlich, dass sie in der speziellen Relativitätstheorie als vierte Raumkoordinate eingeführt wird. Abseits dieser hoffentlich nicht zu verwirrenden Anmerkung macht man aber in der klassischen Mechanik keinen Fehler, wenn man die Zeit als skalare Größe betrachtet.

Zur Unterscheidung von Skalaren verwendet man bei Vektoren häufig einen Pfeil über dem Symbol. Für die Kraft F schreibt man dann $\vec{F}$, wenn man den Richtungscharakter ausdrücken will. Den Betrag deutet man dadurch an, dass man den Pfeil weglässt, da es sich dann ja wieder um eine skalare Größe handelt. In manchen Lehrbüchern v. a. aus dem anglo-amerikanischen Raum werden statt der Pfeile fettgedruckte Buchstaben wie $\mathbf{F}$ verwendet, in sehr alten Büchern und Fachaufsätzen auch »exotische« Darstellungen von Buchstaben wie $\mathfrak{F}$. Ebenfalls zu finden ist ein Strich $\underline{F}$ unter der Größe, das dann ebenfalls einen Vektor bezeichnet. In Lehrbüchern wird in der Regel ganz am Anfang auf die verwendete Notation hingewiesen. Die unterschiedlichen Schreibweisen haben mit den verschiedenen Vorstellungen zu tun, wie sich ein Verlag ein gelungenes Layout vorstellt - hier gibt es im Laufe der Zeit entsprechende Moden. Aus typographischer Sicht sind Fettdruck (außer bei Überschriften) und Unterstriche (immer) in einem Dokument nicht optimal.

In der Physik spielen Vektoren eine wichtige Rolle bei der Beschreibung von räumlichen Bewegungen z. B. von Massenpunkten. Man verwendet hier verschiedene Koordinatensysteme, in denen man die Position eines Körpers durch Angabe von Koordinaten angibt. Das kartesische Koordinatensystem[7] zählt hierbei zu den bekanntesten. Der Ortsvektor $\vec{r}$ hat in kartesischen Koordinaten die Form

$$\vec{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{1.64}$$

Man zerlegt also einen Vektor auf seine Koordinaten auf den drei senkrecht zueinander stehenden Achsen. Man kann einen Vektor als Spaltenvektor wie in Gl. 1.64 schreiben, genausogut kann man aber die Schreibweise als Zeilenvektor $\vec{r} = (x, y, z)$ verwenden. Der Betrag (die Länge) eines Vektors folgt aus Pythagoras und berechnet sich über den Ausdruck

$$|\vec{r}| = \sqrt{x^2 + y^2 + z^2}. \tag{1.65}$$

Der Nullvektor $\vec{0} = (0, 0, 0)$ hat im Unterschied zu allen anderen Vektoren keine Richtung. Aus mathematischer Sicht ist unsere Vektordefinition damit nicht präzise, da der Nullvektor nicht enthalten ist, sondern vielmehr eine vereinfachte Gebrauchsdefinition. In den mathematischen Vorlesungen werden Sie lernen, dass Vektoren eigentlich als Elemente eines Vektorraums definiert sind, die in definierter Art und Weise miteinander verknüpft werden können. Verknüpfungen können Addition

7 Das kartesische Koordinatensystem ist benannt nach dem französischen Philosophen und Mathematiker René Descartes (1596-1650).

und Multiplikation sein. In einem solchen Vektorraum muss es dann ein neutrales Element geben, das alle Elemente über eine Verknüpfung auf sich selbst abbildet. Bei der Addition wäre dieses neutrale Element der Nullvektor.

Vektoren mit einer Länge von Eins bezeichnet man als Einheitsvektoren. Häufig wird der Buchstabe $\vec{e}$ als Symbol für Einheitsvektoren verwendet. Man kann aus jedem Vektor $\vec{r}$ einen zugehörigen Einheitsvektor $\vec{e}_r$ über die Beziehung

$$\vec{e}_r = \frac{\vec{r}}{r} \tag{1.66}$$

konstruieren. Dieser Vektor zeigt dann in die gleiche Richtung wie $\vec{r}$, hat aber eben nur die Länge Eins. Zwei Vektoren $\vec{a}$ und $\vec{b}$ werden als gleich bezeichnet, wenn sie den gleichen Betrag und die gleiche Richtung haben. Bei kartesischen Koordinaten bedeutet dies, dass zwei gleiche Vektoren in allen Komponenten übereinstimmen, wobei man hier hinzufügen muss, dass hier immer die Komponenten vom Anfangspunkt des jeweiligen Vektors gemeint sind, schließlich können sich die beiden Vektoren an völlig unterschiedlichen Stellen befinden, also völlig andere absolute Komponenten haben, obwohl sie den gleichen Vektor darstellen. Den Vektor $-\vec{a}$ bezeichnet man als den inversen Vektor zu $\vec{a}$. Vektor und inverser Vektor haben die gleiche Länge und stehen antiparallel zueinander.

1.3.1 Kartesische Koordinaten

Kartesische Koordinaten sind aus der Schule bekannt und auch im Physikstudium das anfangs bevorzugte Koordinatensystem. Man macht sich hier den Isomorphismus des euklidischen Raums mit dem Koordinatenraum R^3 zu Nutze und beschreibt den Ort eines Objekts durch einen Ortsvektor $\vec{r}$.

Definition: Ortsvektor $\vec{r}$

Unter dem Ortsvektor $\vec{r}$ versteht man einen Vektor, der von einem festen Bezugspunkt zum Ort eines Objekts zeigt.

Hier bedeutet »fest« nicht, dass man ein raumfestes Bezugssystem auswählen muss, sondern nur, dass man einen einmal gewählten Bezugspunkt beibehält. Ob sich dieser Bezugspunkt selbst auch bewegt, ist dabei nicht relevant.

Neben der bereits vorgestellten Darstellung in Gl.1.64 ist die Darstellung als Linearkombination von Einheitsvektoren $\vec{e}_x$, $\vec{e}_y$, $\vec{e}_z$ üblich:

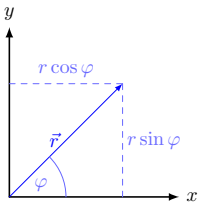
$$\vec{r} = x\vec{e}_x + y\vec{e}_y + \vec{e}_z = x\begin{pmatrix}1\\0\\0\end{pmatrix} + y\begin{pmatrix}0\\1\\0\end{pmatrix} + z\begin{pmatrix}0\\0\\1\end{pmatrix}. \tag{1.67}$$

Koordinaten werden durch Vektoren beschrieben. Es ist daher unerlässlich, dass man sich mit den fundamentalen Rechenoperationen von Vektoren auskennt.

Es gibt in allen Koordinatensystemen zwei wichtige Ausdrücke, die man im späteren Alltagsgeschäft beim Integrieren braucht, das infinitesimale Flächenelement $\mathrm{d}A$ und das infinitesimale Volumenelement $\mathrm{d}V$. In kartesischen Koordinaten lassen sich diese beiden Ausdrücke sehr einfach schreiben als $\mathrm{d}A = \mathrm{d}x \cdot \mathrm{d}y$ (für ein Flächenelement in der xy-Ebene) sowie als $\mathrm{d}V = \mathrm{d}x \cdot \mathrm{d}y \cdot \mathrm{d}z$.

1.3.2 Polarkoordinaten

Wir sind bei der Einführung von Vektoren zunächst von kartesischen Koordinaten ausgegangen, da diese unserer Vorstellung des Raumes recht nahe kommen. In der Physik gibt es noch andere Koordinatensysteme, die sich für spezielle Probleme eher anbieten, als kartesische Koordinaten. Es ist beispielsweise nicht immer sinnvoll, eine zweidimensionale Drehbewegung durch die Koordinaten x und y auszudrücken, sondern durch einen Abstand r von der Drehachse und den Drehwinkel φ. Der Drehwinkel wird für gewöhnlich relativ zur x-Achse angegeben. Koordinaten dieser Art nennt man Polarkoordinaten. Der Übergang von Polarkoordinaten zu kartesischen Koordinaten wird durch



$$\begin{aligned} x &= r\cos\varphi \\ y &= r\sin\varphi \end{aligned} \tag{1.68}$$

beschrieben, der Übergang von kartesischen Koordinaten zu Polarkoordinaten über

$$\begin{aligned} r &= \sqrt{x^2 + y^2} \\ \varphi &= \arctan\left(\frac{x}{y}\right). \end{aligned} \tag{1.69}$$

Ein Beispiel für eine Funktion in Polarkoordinaten ist die archimedische Spirale, die man als

$$r(\varphi) = a \cdot \varphi \tag{1.70}$$

schreiben kann. Setzt man dieses r in Gl. 1.68 ein, erhält man die Darstellung der Spirale in kartesischen Koordinaten:

$$\begin{aligned} x &= a\varphi \cos\varphi \\ y &= a\varphi \sin\varphi. \end{aligned} \tag{1.71}$$

Ein weiteres wichtiges Beispiel ist die Kreisbewegung mit konstanter Umlaufgeschwindigkeit v bzw. Winkelgeschwindigkeit ω. In Polarkoordinaten kann sie beschrieben werden durch

$$\begin{aligned} \varphi &= \omega \cdot t \\ r &= \text{const}, \end{aligned} \tag{1.72}$$

in kartesischen Koordinaten durch

$$\begin{aligned} x &= r \cdot \cos\varphi \\ y &= r \cdot \sin\varphi. \end{aligned} \tag{1.73}$$

Für das Flächenelement in ebenen Polarkoordinaten erhält man:

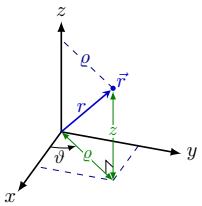
$$A = \int \mathrm{d}A = \int_0^R r\,\mathrm{d}r \int_0^{2\pi} \mathrm{d}\phi = \pi R^2. \tag{1.74}$$

An dieser Stelle lässt sich nochmal der Vorteil gegenüber kartesischen Koordinaten zeigen. Würde man die Integration kartesisch durchführen, so hätte man folgende Integration durchzuführen:

$$A = \int \mathrm{d}A = \int_{-R}^{R} \mathrm{d}x \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \mathrm{d}y = \int_{-R}^{R} \mathrm{d}x 2\sqrt{R^2-x^2} = \pi R^2. \tag{1.75}$$

Die Integration ist hier zwar noch relativ einfach durchzuführen, aber man kann vermutlich erahnen, dass es bei Kugelkoordinaten schon deutlich aufwändiger sein wird, da hier drei Koordinaten über $R^2 = x^2 + y^2 + z^2$ verknüpft sind.

1.3.3 Zylinderkoordinaten



Zylinderkoordinaten ϱ, φ, z erweitern die ebenen Polarkoordinaten um eine weitere Dimension, die als kartesische Koordinate dargestellt wird. Die entsprechenden Transformationsgleichungen lauten

$$\begin{aligned} x &= \varrho \cdot \cos\varphi \\ y &= \varrho \cdot \sin\varphi \\ z &= z, \end{aligned} \tag{1.76}$$

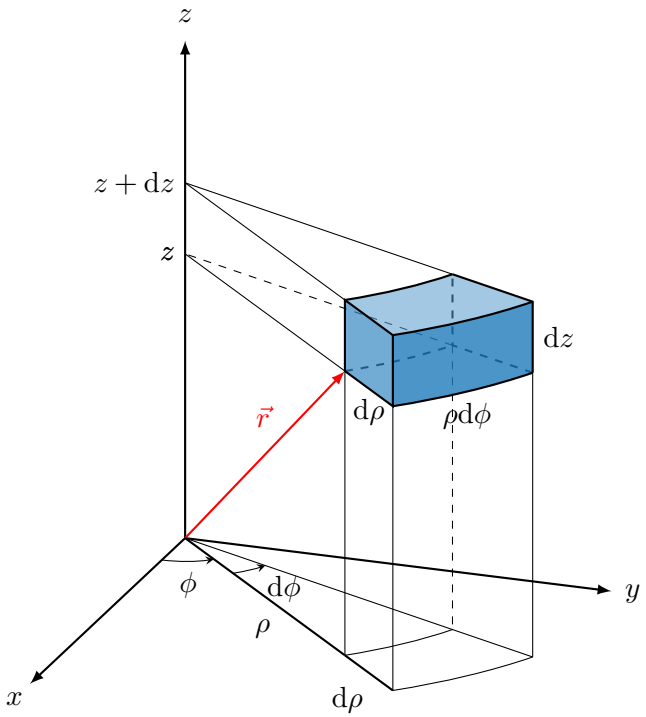


Abbildung 1.14: Zylinderkoordinaten

sowie

$$\begin{aligned} \varrho &= \sqrt{x^2 + y^2} \\ \varphi &= \arctan\left(\frac{y}{x}\right) \\ z &= z. \end{aligned} \tag{1.77}$$

Das Flächenelement $\mathrm{d}A$ in Zylinderkoordinaten kann man definieren als

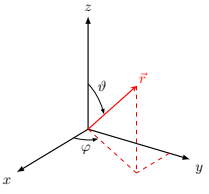
$$\mathrm{d}A = \rho \, \mathrm{d}\phi \, \mathrm{d}z. \tag{1.78}$$

Das Volumenelement ergibt sich durch Multiplikation mit einem kleinen Intervall $\mathrm{d}\rho$, also

$$\mathrm{d}V = \rho \, \mathrm{d}\rho \, \mathrm{d}\phi \, \mathrm{d}z. \tag{1.79}$$

Eine sinnvolle Übung zum Selbststudium ist die Berechnung der Mantelfläche sowie des Volumens eines Zylinders.

1.3.4 Kugelkoordinaten



Kugelkoordinaten r, φ, ϑ sind ebenfalls eine Erweiterung der ebenen Polarkoordinaten um eine dritte Dimension. Diese dritte Koordinate ist

hier der Winkel ϑ zwischen z-Achse und Vektor $\vec{r}$. Die z-Achse wird oft als Polarachse bezeichnet, ϑ als Polarwinkel und φ als Azimutwinkel. Die Transformationsgleichungen lauten

$$\begin{aligned} x &= r \cdot \sin\vartheta \cos\varphi \\ y &= r \cdot \sin\vartheta \sin\varphi \\ z &= r \cdot \cos\vartheta, \end{aligned} \tag{1.80}$$

sowie

$$\begin{aligned} r &= \sqrt{x^2 + y^2 + z^2} \\ \varphi &= \arctan\left(\frac{y}{x}\right) \\ \vartheta &= \arccos\frac{z}{\sqrt{x^2 + y^2 + z^2}}. \end{aligned} \tag{1.81}$$

Das Flächenelement $\mathrm{d}A$ kann man sich aus kleinen Drehungen des Vektors $\vec{r}$ um $\mathrm{d}\theta$ bzw. seiner Projektion auf die Äquatorialebene $r\sin\theta$ um $\mathrm{d}\phi$ zusammensetzen. Es gilt:

$$\mathrm{d}A = r^2 \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\phi. \tag{1.82}$$

Analog für das Volumenelement, welches man durch Ausdehnung des Flächenelements um $\mathrm{d}r$ erhält:

$$\mathrm{d}V = \mathrm{d}A \cdot \mathrm{d}r = r^2 \sin\theta \,\mathrm{d}r \,\mathrm{d}\theta \,\mathrm{d}\phi. \tag{1.83}$$

Was ist jetzt daran der Vorteil? Berechnen wir daraus doch einmal das Volumen einer Kugel. Dazu müssen wir eine dreifache Integration durchführen. Da die drei Koordinaten voneinander unabhängig sind, kann man diese Integration schrittweise durchführen, also beispielsweise erst über die Integration über r und danach über die beiden Winkel θ und ϕ:

$$V = \int \mathrm{d}V = \underbrace{\int_0^R r^2 \,\mathrm{d}r}_{=\frac{1}{3}R^3} \underbrace{\int_0^\pi \sin\theta \,\mathrm{d}\theta}_{=2} \underbrace{\int_0^{2\pi} \mathrm{d}\phi}_{=2\pi} = \frac{4}{3}\pi R^3. \tag{1.84}$$

Man erhält also das erwartete Ergebnis. Für die Oberfläche einer Kugel folgt analog:

$$A = \int \mathrm{d}A = R^2 \underbrace{\int_0^\pi \sin\theta \,\mathrm{d}\theta}_{=2} \underbrace{\int_0^{2\pi} \mathrm{d}\phi}_{=2\pi} = 4\pi R^2 \tag{1.85}$$

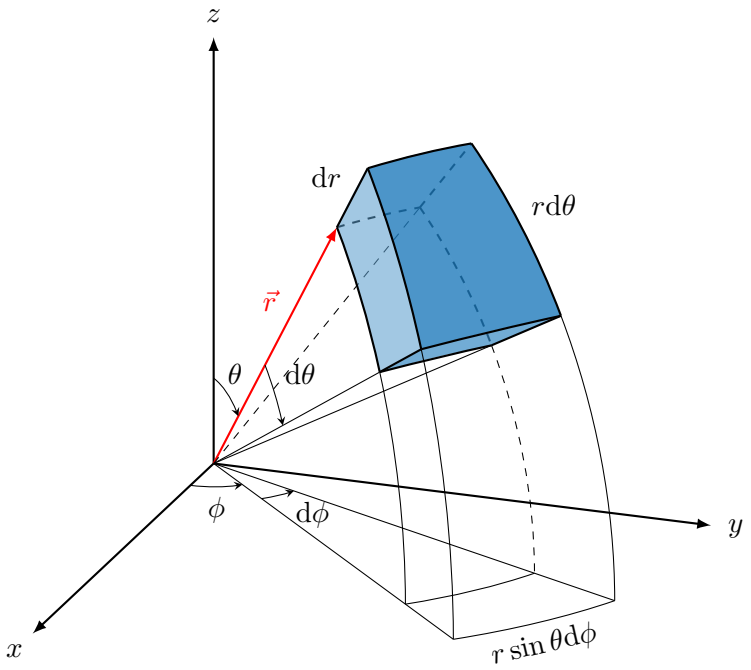


Abbildung 1.15: Kugelkoordinaten

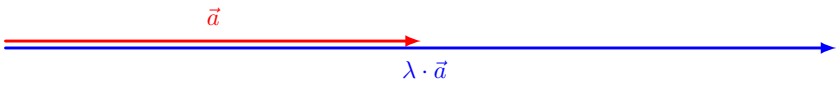


Abbildung 1.16: Multiplikation eines Vektors mit einem Skalar.

1.3.5 Rechenregeln für Vektoren

Multiplikation mit einem Skalar

Die Multiplikation eines Vektors $\vec{a}$ mit einem Skalar λ führt zu einer Verkürzung ($0 < \lambda < 1$) oder Verlängerung ($\lambda > 1$) des Vektors, bei $\lambda < 0$ wird zusätzlich die Richtung des Vektors umgekehrt. Es gilt

$$\vec{b} = \lambda\vec{a} = \begin{pmatrix} \lambda a_x \\ \lambda a_y \\ \lambda a_z \end{pmatrix}. \tag{1.86}$$

Der Betrag eines mit einem Skalar multiplizierten Vektors ergibt sich zu

$$|\,\lambda\vec{a}\,| = |\,\lambda\,| \cdot |\,\vec{a}\,| = \sqrt{\lambda^2 a_x^2 + \lambda^2 a_y^2 + \lambda^2 a_z^2}. \tag{1.87}$$

Es gelten Assoziativität

$$\beta\,(\alpha\vec{a}) = (\beta\alpha)\,\vec{a} \tag{1.88}$$

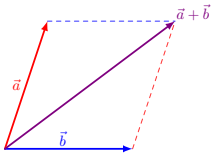
und Distributivität

$$\begin{aligned} \alpha\left(\vec{a}+\vec{b}\right) &= \alpha\vec{a}+\beta\vec{b} \\ (\alpha+\beta)\,\vec{a} &= \alpha\vec{a}+\beta\vec{a} \end{aligned} \tag{1.89}$$

Addition von Vektoren

Vektoren werden komponentenweise addiert. Dies bedeutet, dass die Summe der Vektoren $\vec{a}$ und $\vec{b}$ berechnet wird gemäß

$$\vec{c} = \vec{a}+\vec{b} = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} + \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix} = \begin{pmatrix} a_x+b_x \\ a_y+b_y \\ a_z+b_z \end{pmatrix}. \tag{1.90}$$

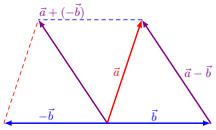


Man kann sich die Addition geometrisch so vorstellen, dass man den Vektor $\vec{b}$ an die Spitze des Vektors $\vec{a}$ anhängt. Die Addition ist kommutativ

$$\vec{a}+\vec{b} = \vec{b}+\vec{a} \tag{1.91}$$

und assoziativ

$$\left(\vec{a}+\vec{b}\right)+\vec{c} = \vec{a}+\left(\vec{b}+\vec{c}\right). \tag{1.92}$$



Die Subtraktion $\vec{a}-\vec{b}$ zweier Vektoren $\vec{a}$ und $\vec{b}$ kann formal als Addition des Vektors $\vec{a}$ mit dem Vektor $-\vec{b}$ interpretiert werden. Der resultierende Vektor kann so verschoben werden, dass er von der Pfeilspitze von $\vec{b}$ zur Pfeilspitze von $\vec{a}$ zeigt.

Differenz von Vektoren

Der resultierende Vektor $\vec{a}-\vec{b}$ zeigt von $\vec{b}$ nach $\vec{a}$.

1.3.6 Skalarprodukt

Wir kommen nun zu einer Verknüpfung zweier Vektoren, die in der Physik eine herausragende Rolle spielt. Das Skalarprodukt (manchmal auch inneres Produkt genannt) verknüpft dabei zwei Vektoren über eine Multiplikation derart, dass als Ergebnis eine skalare Größe herauskommt.

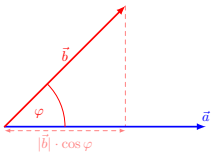
Definition: Skalarprodukt

Das Skalarprodukt der Vektoren $\vec{a}$ und $\vec{b}$ ist definiert als

$$\vec{a} \cdot \vec{b} = |\vec{a}| \cdot \left|\vec{b}\right| \cdot \cos\varphi. \tag{1.93}$$

Dabei entspricht φ dem Winkel zwischen den beiden Vektoren.

Den Grundgedanken beim Skalarprodukt kann man sich am Beispiel der Arbeit verdeutlichen. Arbeit wird aus physikalischer Sicht geleistet, wenn ein Körper, auf den eine Kraft wirkt, bewegt wird. Es spielt dabei eine wesentliche Rolle, ob man diese Bewegung parallel zur Kraftrichtung oder unter einem beliebigen Winkel ausführt. Heben wir einen Körper so an, dass er vertikal nach oben bewegt wird, so stehen Weg und Kraft parallel zueinander. Bewegt man den Körper vertikal auf einem reibungsfreien Untergrund (z. B. eine Eisfläche), so wird physikalisch überhaupt keine Arbeit geleistet. Bewegt man den Körper nun unter einem Winkel zwischen 0 und 90° nach oben, so kann man diese Bewegung in zwei Anteile zerlegen: einen, der in Richtung der Kraft zeigt und dementsprechend Arbeit leistet und einen Anteil senkrecht dazu, der keine Arbeit leistet. Das Skalarprodukt erledigt die Projektion des Weges auf die Kraft oder allgemein: das Skalarprodukt projiziert einen Vektor $\vec{b}$ auf einen anderen Vektor $\vec{a}$. Beachten Sie, dass diese Argumentation in beide Richtungen läuft. Die nebenstehende Abbildung zeigt die Projektion von $\vec{b}$ auf $\vec{a}$. Genauso gut könnte man $\vec{a}$ auf $\vec{b}$ projizieren. Natürlich wird dadurch in der Regel eine andere (Projektions-)Länge herauskommen, aber da beim Skalarprodukt entweder die Projektion von $\vec{b}$ auf $\vec{a}$ mit a multipliziert wird oder die Projektion von $\vec{a}$ auf $\vec{b}$ mit b kommt am Ende das gleiche Ergebnis heraus.



Man kann die Projektion von $\vec{b}$ auf $\vec{a}$ auch als Vektor $\vec{b}_a$ ausdrücken. Die Projektion zeigt in Richtung von $\vec{a}$. Die Richtung von $\vec{a}$ kann über den Einheitsvektor $\vec{e}_a = \vec{a}/a$ ausgedrückt werden. Es gilt

$$\vec{b}_a = |\vec{b}| \cos\varphi \frac{\vec{a}}{a} = |\vec{a}| \cdot |\vec{b}| \cos\varphi \frac{\vec{a}}{a^2}. \tag{1.94}$$

Das Skalarprodukt ist kommutativ

$$\vec{a} \cdot \vec{b} = \vec{b} \cdot \vec{a}, \tag{1.95}$$

assoziativ bei Multiplikation mit einem Skalar

$$(\alpha\vec{a})\,\vec{b} = \alpha\left(\vec{a}\vec{b}\right) = \alpha\vec{a}\vec{b} \tag{1.96}$$

und distributiv

$$\vec{a}\left(\vec{b}+\vec{c}\right)=\vec{a}\vec{b}+\vec{a}\vec{c}. \tag{1.97}$$

Achtung: Es gibt keine Assoziativität bei Multiplikation von drei Vektoren:

$$\left(\vec{a}\vec{b}\right)\vec{c}\neq\vec{a}\left(\vec{b}\vec{c}\right). \tag{1.98}$$

Zwei Vektoren mit Komponenten $\vec{a}=(a_x,a_y,a_z)$ und $\vec{b}=(b_x,b_y,b_z)$ werden skalar gemäß der Vorschrift

$$\vec{a}\cdot\vec{b}=a_xb_x+a_yb_y+a_zb_z \tag{1.99}$$

multipliziert. Man kann gar nicht stark genug betonen, wie nützlich dieser Ausdruck ist. Sind zwei Vektoren in kartesischen Komponenten gegeben, kann man ohne großen Aufwand die Beträge und das Skalarprodukt ausrechnen und damit sofort den Winkel zwischen den beiden Vektoren bestimmen, ohne, dass man sich groß um die Geometrie Gedanken machen zu müssen. Der Winkel ergibt sich dann aus der Beziehung

$$\varphi=\arccos\frac{\vec{a}\cdot\vec{b}}{|\vec{a}|\,|\vec{b}|}. \tag{1.100}$$

Zwei Vektoren stehen also senkrecht zueinander, wenn $\vec{a}\cdot\vec{b}=0$ ist. Man kann in diesem Fall den einen Vektor nicht auf den anderen projizieren.

Die Rechenvorschrift nach Gl. 1.99 ist eine Folge einer anderen wichtigen Darstellung von Vektoren über die Einheitsvektoren in die drei Raumrichtungen x,y und z. Man kann $\vec{a}$ (und analog auch $\vec{b}$) auch in folgender Form schreiben:

$$\vec{a}=a_x\vec{e}_x+a_y\vec{e}_y+a_z\vec{e}_z. \tag{1.101}$$

Die Einheitsvektoren $\vec{e}_{x,y,z}$ zeigen in die jeweilige Achsenrichtung. Beim kartesischen Koordinatensystem stehen die Achsen senkrecht zueinander, so dass man eine recht einfache Verknüpfung der Einheitsvektoren über das Skalarprodukt findet:

$$\begin{aligned}\vec{e}_x\cdot\vec{e}_x=\vec{e}_y\cdot\vec{e}_y=\vec{e}_z\cdot\vec{e}=1\\ \vec{e}_x\cdot\vec{e}_y=\vec{e}_x\cdot\vec{e}_z=\vec{e}_y\cdot\vec{e}_z=0.\end{aligned} \tag{1.102}$$

Die Multiplikation von $\vec{a}$ und $\vec{b}$ sieht damit folgendermaßen aus:

$$\vec{a}\cdot\vec{b}=\left(a_x\vec{e}_x+a_y\vec{e}_y+a_z\vec{e}_z\right)\cdot\left(b_x\vec{e}_x+b_y\vec{e}_y+b_z\vec{e}_z\right). \tag{1.103}$$

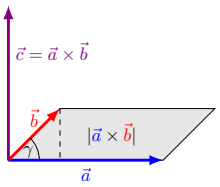
Multipliziert man die Klammer aus und wendet die Beziehungen aus Gl. 1.102 an, folgt die bekannte Beziehung aus Gl. 1.99. Eine abkürzende Schreibweise des Skalarprodukts nutzt das Summenzeichen:

$$\vec{a}\cdot\vec{b}=\sum_{i=1}^{3}a_ib_i. \tag{1.104}$$

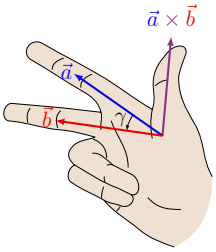
Bei dieser Schreibweise nummeriert man die Komponenten, anstatt die jeweilige Raumrichtung als Index anzuhängen, was aber rechnerisch keinen Unterschied macht.

1.3.7 Vektorprodukt

Das Vektorprodukt (auch äußeres Produkt oder Kreuzprodukt genannt) ist eine Besonderheit des dreidimensionalen Raumes. Im Unterschied zum Skalarprodukt, bei dem eine Zahl als Ergebnis der Multiplikation herauskommt, liefert das Vektorprodukt einen Vektor als Ergebnis. An diesen Vektor sind bestimmte Eigenschaften geknüpft, die wir besprechen wollen. Für zwei linear unabhängige Vektoren $\vec{a}$ und $\vec{b}$ (das sind zwei Vektoren, die nicht parallel zueinander sind und somit eine Ebene aufspannen) liefert das Vektorprodukt $\vec{c} = \vec{a} \times \vec{b}$ einen Vektor $\vec{c}$ mit folgenden Eigenschaften:



- Der Vektor $\vec{c}$ steht senkrecht zu $\vec{a}$ und zu $\vec{b}$, also senkrecht zu der von $\vec{a}$ und $\vec{b}$ aufgespannten Ebene.
- Dem Vektor $\vec{c}$ wird eine Länge zugeordnet, die der Fläche des von $\vec{a}$ und $\vec{b}$ aufgespannten Parallelogramms entspricht. Bezeichnet γ den Winkel zwischen $\vec{a}$ und $\vec{b}$, so folgt hieraus sofort die Beziehung $|\vec{c}| = |\vec{a}|\,|\vec{b}| \sin\gamma$.
- Die Vektoren $\vec{a}$, $\vec{b}$ und $\vec{c}$ bilden ein Rechtssystem. Soll bedeuten, dass die Vektoren $\vec{a}$ und $\vec{b}$ durch Zeige- und Mittelfinger der rechten Hand angedeutet werden können und der Daumen dann die Orientierung von $\vec{c}$ angibt. Dies bedeutet insbesondere auch, dass das Vektorprodukt antikommutativ ist.



Zwei Vektoren, die parallel zueinander stehen, spannen keine Ebene (und somit auch kein Parallelogramm) auf. Die Parallelogrammfläche ist folglich Null, somit auch das Vektorprodukt. Wir können für beliebige Vektoren $\vec{a}$ und $\vec{b}$ eine allgemeine Formel für die Berechnung des Vektorprodukts herleiten, wenn wir ausnutzen, wie sich die kartesischen Einheitsvektoren zueinander verhalten, wenn man das Kreuzprodukt auf sie anwendet. Zwei zueinander senkrechte Einheitsvektoren spannen eine Fläche von Eins (in beliebiger Einheit z. B. m^2) auf. Die Länge des resultierenden Vektors ist also auch ein Einheitsvektor. Es gilt insbesondere

$$\vec{e}_x \times \vec{e}_y = \vec{e}_z \qquad \vec{e}_y \times \vec{e}_z = \vec{e}_x \qquad \vec{e}_z \times \vec{e}_x = \vec{e}_y \tag{1.105}$$

und

$$\vec{e}_x \times \vec{e}_z = -\vec{e}_y \qquad \vec{e}_y \times \vec{e}_x = -\vec{e}_z \qquad \vec{e}_z \times \vec{e}_y = -\vec{e}_x. \tag{1.106}$$

Wenn man die beiden Vektoren $\vec{a}$ und $\vec{b}$ als Linearkombination von Einheitsvektoren $\vec{e}_{x,y,z}$ darstellt, das Vektorprodukt bildet, die Klammern ausmultipliziert und die eben genannten Regeln beachtet, erhält man einen einfachen Ausdruck, um das Vektorprodukt zu bilden:

$$\vec{c} = \vec{a} \times \vec{b} = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} \times \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix} = \begin{pmatrix} a_y b_z - a_z b_y \\ a_z b_x - a_x b_z \\ a_x b_y - a_y b_x \end{pmatrix}. \tag{1.107}$$

Das Vektorprodukt ist wie bereits gesagt antikommutativ, d. h.

$$\vec{a} \times \vec{b} = -\vec{b} \times \vec{a}, \tag{1.108}$$

assoziativ bei Multiplikation mit einem Skalar

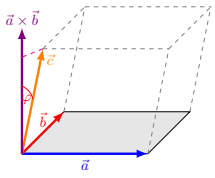
$$(\alpha\vec{a}) \times \vec{b} = \alpha \left(\vec{a} \times \vec{b} \right) = \alpha\vec{a} \times \vec{b} \tag{1.109}$$

und distributiv

$$\vec{a} \times \left(\vec{b} + \vec{c} \right) = \vec{a} \times \vec{b} + \vec{a} \times \vec{c}. \tag{1.110}$$

1.3.8 Anwendungen von Skalar- und Vektorprodukt

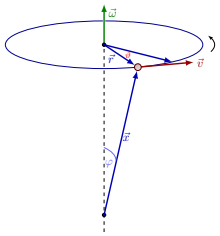
Das Vektorprodukt kann dazu genutzt werden, das Volumen V eines Spats (auch Parallelflach oder Parallelepiped genannt) zu berechnen, welches von drei linear unabhängigen Vektoren $\vec{a}$, $\vec{b}$ und $\vec{c}$ aufgespannt wird. Ein Spat ist also ein geometrischer Körper, der aus sechs Parallelogrammen besteht, wobei jeweils zwei gegenüberliegende Parallelogramme deckungsgleich (kongruent) und zueinander parallel sind. Der Betrag des Vektorprodukts entspricht gerade einer dieser Parallelogrammflächen (je nachdem, welche Vektoren man wählt). Die Projektion des dritten Vektors auf den resultierenden Vektor des Kreuzprodukts entspricht der Höhe h des Spats, entsprechend liefert $(\vec{a} \times \vec{b})\vec{c}$ nichts anderes als das Produkt aus Grundfläche A und Höhe h des Spats, also eines Volumens V. Dieses kann je nach Reihenfolge der Vektoren auch ein negatives Vorzeichen haben, weshalb noch der Betrag des Spatprodukts gebildet wird. Das Spatprodukt $V = \left[\vec{a}\vec{b}\vec{c} \right]$ berechnet sich nach



$$V = \overbrace{\left| \vec{a} \times \vec{b} \right|}^{A} \cdot \overbrace{|\vec{c}| \cos\varphi}^{h} \tag{1.111}$$

Eine weitere Anwendung des Kreuzprodukts ist die Beschreibung von Drehbewegungen. Betrachten wir einen Körper, der sich auf einer Kreisbahn mit Radius r befindet. Die Geschwindigkeit des Körpers soll sich betragsmäßig nicht ändern, die Winkelgeschwindigkeit ω ist also

eine Konstante. Wir können uns nun einen beliebigen Punkt auf der Drehachse als Ursprung unseres Koordinatensystems auswählen. Der Vektor $\vec{x}$ soll von diesem Ursprung auf den Körper zeigen. Der Vektor $\vec{r}$ zeigt vom Mittelpunkt der Kreisbewegung zum Körper. Beide Vektoren ändern ihre Lage mit der Bewegung des Körpers um die Drehachse. Die Geschwindigkeit v kann aus der zurückgelegten Strecke s des Körpers pro Zeitintervall t bestimmt werden. Die Strecke entspricht gerade dem Bogenstück $r\vartheta$ der Kreisbahn, also



$$s = \frac{r\vartheta}{t}. \tag{1.112}$$

Den Radius r können wir durch den Betrag des Vektors $\vec{x}$ und dem Winkel φ zwischen Drehachse und $\vec{x}$ ausdrücken:

$$r = |\vec{x}| \cdot \sin\varphi. \tag{1.113}$$

Daraus folgt:

$$|\vec{v}| = r\omega = |\vec{x}| \cdot \sin\varphi \cdot |\vec{\omega}| = |\vec{\omega} \times \vec{x}|. \tag{1.114}$$

Da wir den Ursprung beliebig auf die Drehachse gelegt haben, können wir ihn auch o. B. d. A. auch in den Mittelpunkt der Kreisbahn legen, müssen es aber nicht. Es hilft aber manchmal, die Orientierung leichter zu prüfen, also ob ein Rechtssystem vorliegt, was der Fall ist. Für die Geschwindigkeit gilt also vektoriell der Ausdruck

$$\vec{v} = \vec{\omega} \times \vec{x}. \tag{1.115}$$

1.4 Komplexe Zahlen

Den komplexen Zahlen ist ein eigener Abschnitt gewidmet, um deren Bedeutung in der Physik besonders zu betonen. Schauen wir uns dazu ein wenig die Geschichte ihrer Entstehung an. Erstmals in der mathematischen Literatur erwähnt wurden die komplexen Zahlen in der Zeit der italienischen Renaissance. Nachdem man bereits um etwa 1500 die negativen Zahlen als Lösung der Gleichung $x + 4 = 2$ identifiziert hatte, hat man sich an Gleichungen der Art $x^2 = -9$ probiert. Bekannt zu dieser Zeit war die Tatsache, dass das Quadrat einer beliebigen Zahl entweder größer oder gleich Null war, so dass diese Gleichung mit den reellen Zahlen $\mathbb{R}$ nicht lösbar ist. Geholfen hat man sich, indem man eine eingebildete (imaginäre) Größe eingeführt hat, die folgende Eigenschaft haben muss:

$$i^2 = -1. \tag{1.116}$$

Damit konnte man dann als Lösung von Gleichung $x^2 = -9$ den Wert $x = i\sqrt{9} = 3i$ angeben, hatte aber noch keine konkrete Vorstellung davon, was das bedeuten soll. Der Mathematiker Gerolamo Cardano hat sich in seinem Buch *Artis magnae sive de regulis algebraicis liber unus* aus dem Jahr 1545 etwas grundlegendere Gedanken gemacht. Er hat dort eine Aufgabe diskutiert, bei der es um zwei Zahlen geht, deren Summe 10 und deren Produkt 40 ergeben soll. Es ging also um eine Lösung der Gleichung $x(10 - x) = 40$ oder ausgeschrieben um

$$x^2 - 10x + 40 = 0. \tag{1.117}$$

Die aus der Schule hoffentlich bekannte *pq*-Formel war auch zur damaligen Zeit schon bekannt, und so konnte Cardano als Lösung der quadratischen Gleichung die Zahlen $x_1 = 5 + \sqrt{-15}$ und $x_2 = 5 - \sqrt{-15}$ angeben. Der Unterschied zum ersten Beispiel ist der, dass hier eine Lösung in der Art vorkommt, dass man hier eine Verknüpfung aus einer reellen Zahl und einer weiteren reellen Zahl, die mit der imaginären Größe i verbunden ist, vorliegen hat. Man definierte damit die komplexen Zahlen formal als Zahlen, die in der Form

$$z = a + ib \tag{1.118}$$

vorliegen, wobei a und b reelle Zahlen sind. Man bezeichnet a als Realteil und b als Imaginärteil der komplexen Zahl z. In manchen Lehrbücher wird dafür auch $a = \mathfrak{R}(a + ib)$ und $b = \mathfrak{I}(a + ib)$ bzw. $a = \mathrm{Re}(a + ib)$ und $b = \mathrm{Im}(a + ib)$ geschrieben. Wichtig: Real- und Imaginärteil sind jeweils reelle Zahlen. Ihr volles Existenzrecht haben die komplexen Zahlen erhalten, als Carl Friedrich Gauß mit ihnen den Fundamentalsatz der Algebra beweisen konnte: Im Bereich der komplexen Zahlen hat jedes nicht konstante Polynom mit reellen Koeffizienten mindestens eine Nullstelle. Etwas vornehmer ausgedrückt sagt man auch, dass die komplexen Zahlen algebraisch abgeschlossen sind. Unser Beispiel $x^2 - 10x + 40$ ist genau ein solches Polynom, Cardano hat dafür die beiden genannten Lösung aus dem Bereich der komplexen Zahlen angegeben.

Komplexe Zahlen spielen bei der Behandlung von Schwingungen eine wichtige Rolle. Sie sind Lösung der Differentialgleichungen, die Schwingungen beschreiben und mathematisch deutlich einfacher zu handhaben als Sinus- und Kosinus-Terme, die ebenfalls als Lösungen in Frage kommen. Die Behandlung von Wechselstromgrößen vereinfacht sich kolossal, wenn man mit komplexen Größen rechnet. Ganze Lehrbücher der Elektrotechnik können auf wenige Seiten gekürzt werden, wenn man mit komplexen Zahlen rechnet. Im späteren Verlauf des Studiums, wenn die Quantenmechanik behandelt wird, sind die komplexen Zahlen dann der Normalfall. Die gesamte Quantenmechanik funktioniert nur vernünftig, wenn man mit komplexen Zahlen rechnet.

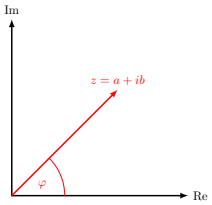
1.4.1 Rechengesetze

Komplexe Zahlen werden nach folgenden Regeln verknüpft:

$$\begin{aligned}(a+ib)+(x+iy) &= (a+x)+i(b+y)\\ (a+ib)\cdot(x+iy) &= ax-by+i(ay+bx)\end{aligned} \tag{1.119}$$

Man kann die komplexen Zahlen als Punkt (a, b) einer Ebene definieren, die von der reellen (als x-Achse) und der imaginären Achse (als y-Achse) aufgespannt wird. Die Rechenregeln sehen dann folgendermaßen aus:

$$\begin{aligned}\begin{pmatrix}a\\b\end{pmatrix}+\begin{pmatrix}x\\y\end{pmatrix} &= \begin{pmatrix}a+x\\b+y\end{pmatrix}\\ \begin{pmatrix}a\\b\end{pmatrix}\cdot\begin{pmatrix}x\\y\end{pmatrix} &= \begin{pmatrix}ax-by\\ay+bx\end{pmatrix}\end{aligned} \tag{1.120}$$



Man kann eine komplexe Zahl $a + ib$ also auch in der Form

$$\begin{pmatrix}a\\b\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix}a+\begin{pmatrix}0\\1\end{pmatrix}b \tag{1.121}$$

schreiben. Es gibt hier also eine gewisse Ähnlichkeit mit der Darstellung von Vektoren. Die imaginäre Einheit i ist nun auch keine abstrakte (imaginäre) Größe mehr, sondern einfach der Punkt $(0, 1)$ auf der imaginären Achse. Die reellen Zahlen werden dadurch zu einem Spezialfall der komplexen Zahlen, nämlich als diejenigen mit Imaginärteil von Null. Man bezeichnet die komplexen Zahlen mit dem Symbol $\mathbb{C}$, wobei damit nicht nur die Zahlenmenge gemeint ist, sondern auch damit verbundenen Verknüpfungen der Elemente (also Addition und Multiplikation). Es gilt also

$$\mathbb{C} = \{z = a + ib \;/\; a, b \,\epsilon\, \mathbb{R}\}. \tag{1.122}$$

Die Darstellung der komplexen Zahlen in der Ebene findet später Anwendung in der Elektrotechnik bei der Behandlung von Strom und Spannung im Wechselstromkreis. So lässt sich z. B. der Phasenwinkel zwischen diesen beiden Größen elegant geometrisch darstellen.

Der Betrag einer komplexen Zahl ist definiert als

$$|z| = |a + ib| = \sqrt{x^2 + y^2}. \tag{1.123}$$

Diese Gleichung folgt sofort aus der Darstellung in der Ebene und Anwendung des Lehrsatzes von Pythagoras.

Als Konjugation in $\mathbb{C}$ bezeichnet man die Spiegelung einer komplexen Zahl $z = x + iy$ an der reellen Achse. Als Kennzeichnung dieser

gespiegelten Zahl $\overline{z}$ wird häufig ein Strich über der Zahl gezogen. Es gilt $\overline{z} = x - iy$. Man bezeichnet $\overline{z}$ auch als komplex konjugierte Zahl zu z. Der Betrag einer komplexen Zahl lässt sich damit auch als $|z| = z\overline{z}$ schreiben. Bei der Division zweier komplexer Zahlen z_1 und z_2 erweitert man mit der komplex konjugierten Zahl, die im Nenner steht, um das Ergebnis der Division in gewohnter Darstellung $a + ib$ zu erhalten:

$$\frac{z_1}{z_2} = \frac{z_1\overline{z_2}}{z_2\overline{z_2}} = a + ib. \tag{1.124}$$

1.4.2 Die Eulersche Formel

Wir können komplexe Zahlen in einer zweidimensionalen Ebene darstellen, die von der imaginären und der reellen Achse aufgespannt wird. Formal ähnelt dies der Darstellung eines Vektors in kartesischen Koordinaten und es gibt nun überhaupt keinen Grund, nicht auch auf Polarkoordinaten zu wechseln. Dazu führen wir den Einheitskreis $\mathbb{S}$ in der komplexen Ebene ein:

$$\mathbb{S} = \{z \,\epsilon\, \mathbb{C} \;/\; |z| = 1\}. \tag{1.125}$$

Eine komplexe Zahl, die auf dem Einheitskreis liegt, kann in Polarkoordinaten durch die Beziehung

$$z = \cos\varphi + i\sin\varphi \tag{1.126}$$

ausgedrückt werden. Interessant wird es nun, wenn man zwei komplexe Zahlen in dieser Polarkoordinatenform miteinander multipliziert. Man sieht dann, dass diese Multiplikation einer Addition der Winkel (im Bogenmaß) und einer Multiplikation der Beträge der beiden Zahlen entspricht, allerdings sei für diese Thematik auf die mathematischen Spezialvorlesungen verwiesen. Wir wollen uns hier einen ganz speziellen Zusammenhang verdeutlichen, der beim Lösen von Schwingungsgleichungen sehr nützlich ist und den man als Eulersche Formel bezeichnet. Wir hatten bereits eine Reihendarstellung für die Funktion kennengelernt (siehe Gl. 1.39):

$$\mathrm{e}^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots. \tag{1.127}$$

Man kann diese Darstellung auch für einen imaginären Exponenten ix durchführen:

$$\mathrm{e}^{ix} = 1 + ix + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \frac{(ix)^4}{4!} + \ldots. \tag{1.128}$$

Bei allen geraden Exponenten fällt die imaginäre Einheit i weg und das Vorzeichen ist entweder negativ (wenn für i^n das Ergebnis aus $n/2$ ungerade) oder positiv (wenn für i^n das Ergebnis aus $n/2$ gerade). Bei allen ungeraden Exponenten bleibt die imaginäre Einheit bestehen, es findet aber auch ein Vorzeichenwechsel statt (ungerade Exponenten negativ, gerade Exponenten positiv). Man kann die Reihe also auch in die Form

$$\mathrm{e}^{ix} = \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots\right) + i\left(x + -\frac{x^3}{3!} + \frac{x^5}{5!} - \ldots\right). \tag{1.129}$$

bringen. Nun entspricht die erste Reihe gerade der Reihendarstellung des Kosinus, die zweite Reihe der des Sinus. Damit gilt dann

$$\mathrm{e}^{ix} = \cos x + i \sin x, \tag{1.130}$$

was sich beim Rechnen als vorteilhaft erweist, wie wir beim Lösen der Differentialgleichung eines harmonischen Oszillators sehen werden.

1.4.3 Bezug zu den Winkelfunktionen

Die Eulersche Formel verknüpft die (komplexe) Exponentialfunktion mit der Sinus- und Kosinus-Funktion. Der Zusammenhang ist eigentlich recht einfach, wenn man sich die Situation graphisch veranschaulicht. Abbildung 1.17 zeigt den komplexen Einheitskreis, auf dem ein willkürlich gewählter Punkt e^{ix} hervorgehoben ist. Zusätzlich sind zwei weitere Punkte eingezeichnet: der konjugiert komplexe Wert e^{-ix} (der durch Spiegelung an der reellen Achse entsteht) sowie sein am Ursprung gespiegelter Wert $-\mathrm{e}^{-ix}$. Wir wissen, dass sich zwei komplexe Zahlen wie Vektoren addieren. Bildet man die Summe $\mathrm{e}^{ix} + \mathrm{e}^{-ix}$, dann liegt die resultierende komplexe Zahl genau bei A, ist also rein reell. Der Kosinus, also die Projektion von e^{ix} auf die reelle Achse, entspricht gerade dem halben Wert dieser Summe:

$$\cos x = \frac{\mathrm{e}^{ix} + \mathrm{e}^{-ix}}{2}. \tag{1.131}$$

Bildet man die Summe $\mathrm{e}^{ix} + (-\mathrm{e}^{-ix}) = \mathrm{e}^{ix} - \mathrm{e}^{-ix}$, so liegt die resultierende komplexe Zahl auf der imaginären Achse bei Punkt B. Die Projektion von e^{ix} auf die imaginäre Achse entspricht gerade $i \sin x$, was wiederum dem halben Wert der Summe $\mathrm{e}^{ix} - \mathrm{e}^{-ix}$ entspricht. Daraus folgt

$$\sin x = \frac{\mathrm{e}^{ix} - \mathrm{e}^{-ix}}{2i}. \tag{1.132}$$

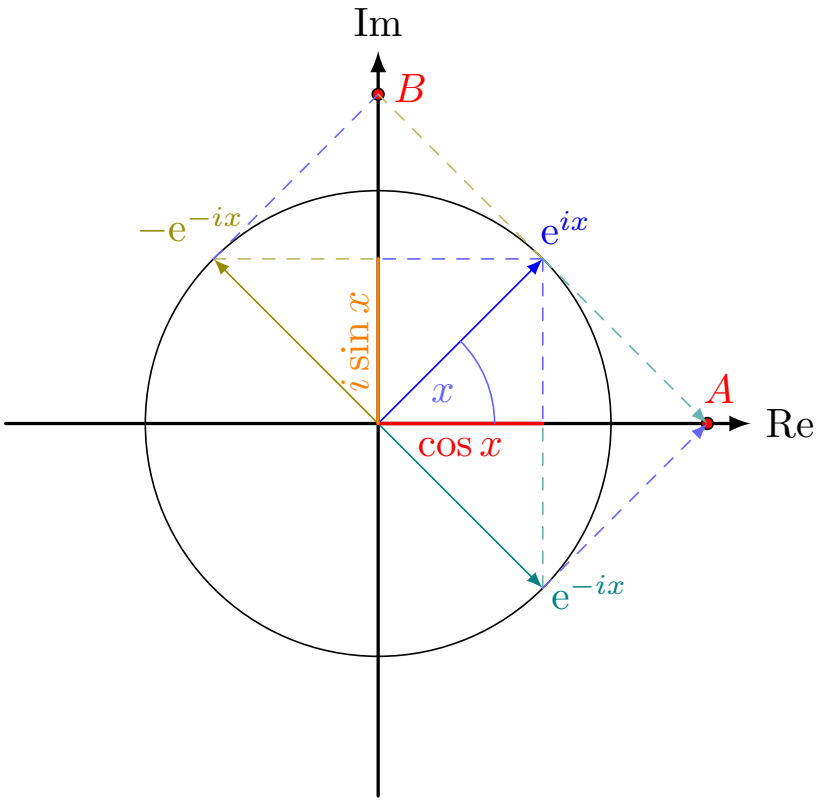


Abbildung 1.17: Darstellung von Sinus und Kosinus in der komplexen Ebene.

Die Verknüpfung zwischen trigonometrischen und hyperbolischen Funktionen erhält man, indem man in den Gleichungen 1.131 und 1.132 das x im Exponenten imaginär werden lässt, also die Transformation $x \to ix$ durchführt. Dadurch ergibt sich

$$\cos ix = \frac{\mathrm{e}^{x} + \mathrm{e}^{-x}}{2} = \cosh x \tag{1.133}$$

sowie

$$\sin ix = \frac{\mathrm{e}^{x} + \mathrm{e}^{-x}}{2} = i \sinh x \tag{1.134}$$

1.5 Differentialrechnung

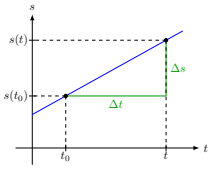
1.5.1 Differentialquotient

Wir kommen nun zu einem ganz wesentlichen Punkt in den modernen Naturwissenschaften, den man aus der Schule meistens unter dem Begriff

»Ableiten« kennengelernt hat. Die Idee beim Ableiten ist die Näherung eines beliebigen funktionalen Zusammenhangs durch eine affine Funktion. Wenn man also eine Funktion $f(x)$ an einem Funktionswert x betrachtet, so will man »abschätzen«, welcher Funktionswert $f(x + \Delta x)$ sich an einem Punkt $x + \Delta x$ ergibt. Wir wollen dabei unser Δx so wählen, dass sich mindestens eine quadratische Abhängigkeit der Ungenauigkeit dieser Schätzung ergibt. In anderen Worten: Verkleinert man Δx um den Faktor 10, so soll $f(x + \Delta x$ um den Faktor 100 am realen Wert der Funktion dran sein. Wir suchen also eine Funktion, die wir $f'(x)$ nennen und für die der Zusammenhang

$$f(x + \Delta x) = f(x) + f'(x) \cdot \Delta x + \mathcal{O}((\Delta x)^2) \tag{1.135}$$

gelten soll. Die Notation $\mathcal{O}((\Delta x)^2)$ steht für Terme höherer Ordnung (ab dem quadratischen Term wird in dem Beispiel gezählt) und bedeutet, dass man alle Terme mit Ordnung $n \geq 2$ vernachlässigen kann. Der führende Beitrag in unserer Näherung soll also die lineare Näherung an die Funktion $f(x)$ sein.

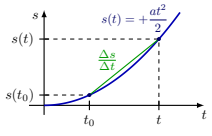


Betrachten wir die Situation anhand eines Beispiels: Ein Auto, das wir idealisiert als Massenpunkt[8] beschreiben, fahre mit konstanter Geschwindigkeit v eine Strecke Δs. Zur Anfangszeit t_0 sei das Auto an der Stelle $s(t_0)$, zur Zeit t an der Stelle s. Wir können für diese Bewegung die Geschwindigkeit definieren als

$$v = \frac{s(t) - s(t_0)}{t - t_0}. \tag{1.136}$$

Eine lineare Beziehung der Form wie in Gl. 1.135 erhalten wir durch Umstellen:

$$s(t) = s(t_0) + v(t - t_0). \tag{1.137}$$



Nun sind funktionale Zusammenhänge in der Realität selten linear. Verändert man im Zeitintervall $t - t_0$ die Geschwindigkeit, indem man beispielsweise mit konstanter Beschleunigung a die Geschwindigkeit erhöht, so ändert sich diese quadratisch mit der Zeit t. Es gilt in diesem Fall die Beziehung

$$s(t) = \frac{1}{2}at^2, \tag{1.138}$$

wenn das Auto sich zur Zeit $t = 0$ am Ort $s = 0$ befunden hat. Man kann nun zwei beliebige Punkte s_1 und s_2 festlegen, sich die Zeiten t_1 und t_2 anschauen, zu denen das Auto an diesen Orten war und eine Geschwindigkeit nach Gl. 1.136 definieren. Man bekommt allerdings bei großen Zeitabständen dann aber auch nur eine mittlere Geschwindigkeit,

8 Die Definition des Massenpunktes folgt an späterer Stelle.

die unter Umständen sogar Null ergeben kann. Interessanter wäre aber eine Definition, die uns die Geschwindigkeit zu einem gegebenen Zeitpunkt t angibt. Hier kann man sich behelfen, wenn man das Zeitintervall Δt sehr klein macht. Im Grenzfall $\Delta t \to 0$ bekommt man dann die so genannte Momentangeschwindigkeit, die man also als

$$v = \lim_{\Delta x \to 0} \frac{s(t + \Delta t) - s(t)}{\Delta t} \tag{1.139}$$

schreiben kann.

Diese Überlegungen übertragen wir nun auf eine Funktion f und halten fest: eine Funktion $f(x)$ ist an der Stelle x differenzierbar, wenn der Grenzwert

$$\lim_{\substack{\Delta x \to 0 \\ \Delta x \neq 0}} \frac{f(x + \Delta x) - f(x)}{\Delta x} \tag{1.140}$$

existiert. Diesen Grenzwert nennt man die Ableitung von $f(x)$ und bezeichnet sie mit $f'(x)$. Ohne die Grenzwertbildung wird der Bruch auch als Differenzenquotient bezeichnet, mit Grenzwertbildung spricht man vom Differentialquotienten. Eine Funktion ist also genau dann ableitbar, wenn der Grenzwert nach Gl. 1.140 existiert.

Ein Hinweis zur Bezeichnung von Ableitungen: Aus der Schule kennt man vermutlich das bekannte $f'(x)$. In der Physik werden Ihnen sehr häufig auch Darstellungen der Form $\dot{x}(t)$ begegnen. Hier ist x die Funktion und die Zeit t die Variable. Zeitableitungen werden in der Physik mit einem Punkt über dem Formelzeichen gekennzeichnet, der Strich ist den anderen Ableitungstypen (vornehmlich nach Ortskoordinaten) vorbehalten. Zweifache Ableitungen schreibt man entsprechend $f''(x)$ bzw. $\ddot{x}(t)$, bei höheren Ableitungen (z. B. n-mal) schreibt man einfach $f^{(n)}(x)$. Bei zeitlichen Ableitungen höherer (n-ter) Ordnung gibt es die Schreibweise

$$\overset{n}{\dot{x}}, \tag{1.141}$$

in der Physik ist das $\ddot{x}$ aber auch völlig ausreichend, höhere Zeitableitungen sind eher selten anzutreffen. Eine andere Schreibweise für die Ableitung ist die Darstellung als Bruch $\mathrm{d}f / \mathrm{d}x$. Das $\mathrm{d}f$ steht dabei für eine sehr (infinitesimal) kleine Differenz, die sonst üblicherweise den Buchstaben Δ zur Kenntlichmachung verpasst bekommt. Höhere Ableitungen werden häufig auch in der Form

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}f}{\mathrm{d}x}\right) = \frac{\mathrm{d}^2 f}{\mathrm{d}x^2} = f''(x) \tag{1.142}$$

dargestellt. Man sollte sich mit den verschiedenen Schreibweisen für Ableitungen und deren Bedeutung früh vertraut machen.

Fassen wir die Darstellungen für die erste Ableitung zusammen:

$$f'(x) = \frac{\mathrm{d}f}{\mathrm{d}x} = \lim_{\substack{\Delta x \to 0 \\ \Delta x \neq 0}} \frac{\Delta f}{\Delta x} = \lim_{\substack{\Delta x \to 0 \\ \Delta x \neq 0}} \frac{f(x + \Delta x) - f(x)}{\Delta x}. \tag{1.143}$$

Schauen wir uns die Bildung einer Ableitung mit Hilfe des Differentialquotienten nach Gl. 1.140 am Beispiel der Funktion $f(x) = x^2 + 1$ genauer an. Für den ersten Term im Zähler des Bruchs benötigen wir das $f(x + \Delta x)$, also den Funktionswert für das Argument $x + \Delta x$. Das ergibt: $f(x + \Delta x) = x^2 + 2x\Delta x + (\Delta x)^2 + 1$. Einsetzen in Gl. 1.143 ergibt

$$\begin{aligned}
\frac{\mathrm{d}f}{\mathrm{d}x} &= \lim_{\substack{\Delta x \to 0 \\ \Delta x \neq 0}} \frac{x^2 + 2x\Delta x + (\Delta x)^2 + 1 - (x^2 + 1)}{\Delta x} \\
&= \lim_{\substack{\Delta x \to 0 \\ \Delta x \neq 0}} \frac{2x\Delta x + (\Delta x)^2}{\Delta x} \\
&= \lim_{\substack{\Delta x \to 0 \\ \Delta x \neq 0}} 2x + \Delta x \\
&= 2x,
\end{aligned} \tag{1.144}$$

wie man es nach der allgemeinen Ableitungsvorschrift $f'(x) = nx^{n-1}$ für eine Funktion $f(x) = x^n$ erwarten würde. Für einen Beweis dieser allgemeinen Vorschrift sei auf die entsprechenden mathematischen Vorlesungen verwiesen.

1.5.2 Das Differential

Vielleicht wundern Sie sich über die Schreibweise $f'(x) = \mathrm{d}f/\,\mathrm{d}x$, die den Eindruck erweckt, als wäre der Differentialquotient nichts weiter als das Verhältnis zweier infinitesimal kleiner Größen $\mathrm{d}f$ und $\mathrm{d}x$. Im strengen mathematischen Sinn muss man an dieser Stelle vorsichtig sein, allerdings kann man in vielen Fällen mit den kleinen Größen $\mathrm{d}x$ und $\mathrm{d}f$ aber so rechnen, als wären das eigenständig Größen. Das Differential $\mathrm{d}f$ kann man dadurch (tautologisch) definieren als

$$\mathrm{d}f = \mathrm{d}f \cdot 1 = \mathrm{d}f \cdot \underbrace{\frac{\mathrm{d}x}{\mathrm{d}x}}_{=1} = \left(\frac{\mathrm{d}f}{\mathrm{d}x}\right) \mathrm{d}x = f'(x)\,\mathrm{d}x. \tag{1.145}$$

Vergleichen wir das mit Gl. 1.135. Das Differential $\mathrm{d}f$ nach Gl. 1.145 gibt die Änderung der Funktion f bei kleiner Variation der variablen Größe x um ein $\mathrm{d}x$ an. Hierbei wird nur der lineare Anteil dieser Änderung berücksichtigt. Terme höherer Ordnung werden hier also vernachlässigt.

Es spielt insbesondere bei Funktionen mit mehreren Variablen eine wichtige Rolle und wird später noch etwas ausführlicher diskutiert.

1.5.3 Differenzierbarkeit von Funktionen

Nicht jede Funktion lässt sich differenzieren, also in einem bestimmten Bereich eindeutig linear annähern. Ein guter Indikator für Differenzierbarkeit ist die Stetigkeit der Funktion, da jede Funktion, die differenzierbar ist, stetig sein muss, während jede nicht-stetige Funktion an der Stelle der Unstetigkeit nicht differenzierbar ist. Allerdings ist die Stetigkeit einer Funktion nicht ausreichend. Bekanntes Beispiel ist die Betragsfunktion $f(x) = |x|$, die an der Stelle $x = 0$ zwar stetig ist, aber keine eindeutige lineare Approximation zulässt. Ein anderes Beispiel ist die Funktion $f(x) = \sqrt{x}$, die an der Stelle $x = 0$ ebenfalls nicht differenzierbar ist. Betrachten wir dazu den Differenzenquotient, wenn man sich von positiven x-Werten dem Wert $x = 0$ annähert. Es gilt hierbei

$$\frac{f(x) - f(0)}{x - 0} = \frac{\sqrt{x} - 0}{x - 0} = \frac{1}{\sqrt{x}}. \tag{1.146}$$

Im Grenzfall eines immer kleiner werdenden Intervalls, das durch die Werte $x_0 = 0$ und $\Delta x = x - x_0 = x$ bestimmt ist, also im Grenzfall

$$\lim_{x \to 0} \frac{1}{\sqrt{x}}, \tag{1.147}$$

ergibt sich ein unendlich großer Grenzwert. Die Ableitung ist an dieser Stelle also nicht eindeutig definierbar. Jede differenzierbare Funktion muss durch eine lineare Approximation der Form $f(x) \approx f(x_0) + f'(x) \cdot (x - x_0)$ darstellbar sein, was für $f(x) = \sqrt{x}$ an der Stelle $x_0 = 0$ nicht möglich ist. Geometrisch lässt sich das so interpretieren, dass die Tangente an die Funktion $\sqrt{x}$ im Nullpunkt parallel zur y-Achse steht (mit dieser hier sogar identisch ist), somit kein Steigungsdreieck angelegt werden kann.

Es gibt eine Reihe von Ableitungen wichtiger Funktionen, die man immer parat haben sollte und die in Tabelle 1.3 zusammengefasst sind. Hinzu kommen eine Reihe von wichtigen Rechenregeln, die ebenfalls häufig gebraucht werden (f und g sind Funktionen der Variablen x):

- Faktorregel:
$$(af)' = a \cdot f' \tag{1.148}$$
- Summenregel:
$$(f + g)' = f' + g' \tag{1.149}$$

Tabelle 1.3: Ableitungen wichtiger Funktionen.

$f(x)$	$f'(x)$	$f(x)$	$f'(x)$
x^n	$n \cdot x^{n-1}$	$\log_a x$	$\dfrac{1}{x \cdot \ln a}$
a	0	$\ln x$	$\dfrac{1}{x}$
$\sin x$	$\cos x$	$\sinh x$	$\cosh x$
$\cos x$	$-\sin x$	$\cosh x$	$\sinh x$
e^x	e^x	$\tanh x$	$\dfrac{1}{\cosh^2 x}$
a^x	$(\ln a) \cdot a^x$	$\tan x$	$\dfrac{1}{\cos^2 x}$

- Produktregel:
$$(fg)' = f'g + fg' \tag{1.150}$$
- Quotientenregel:
$$\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2} \tag{1.151}$$
- Kettenregel:
$$(f \circ g)' = g'(f(x)) \cdot f'(x) \tag{1.152}$$

Diese Rechenregeln lassen sich recht einfach aus der Definition des Differentialquotienten herleiten. Meist muss man dabei nur ein paar ≫Tricks≪ anwenden, wie das geschickte Erweitern der Formeln, die Ihnen nach einiger Zeit aber auch nicht mehr wie Tricks vorkommen

werden. Wir führen dies exemplarisch am Beispiel der Produktregel durch. Es gilt

$$\begin{aligned}(fg)' &= \lim_{\Delta x\to 0}\frac{f(x+\Delta x)g(x+\Delta x)-f(x)g(x)}{\Delta x}\\ &= \lim_{\Delta x\to 0}\frac{f(x+\Delta x)g(x+\Delta x)-\overbrace{f(x)g(x+\Delta x)+f(x)g(x+\Delta x)}^{=0}-f(x)}{\Delta x}\\ &= \lim_{\Delta x\to 0}\frac{(f(x+\Delta x)-f(x))\overbrace{g(x+\Delta x)}^{\approx g(x)}}{\Delta x}+\lim_{\Delta x\to 0}\frac{f(x)(g(x+\Delta x)-g(x))}{\Delta x}\\ &= f'\cdot g+f\cdot g'\end{aligned} \tag{1.153}$$

Die Kettenregel schauen wir uns an einem Beispiel an: Sei $y=\sin x^2$. Wir können die Funktion $y(x)$ in zwei Teilfunktionen f und g zerlegen:

$$\begin{aligned}f(x) &= x^2\\ g(y) &= \sin y.\end{aligned} \tag{1.154}$$

Daraus folgt mit den allgemeinen Ableitungsregeln der Teilfunktionen:

$$\begin{aligned}f'(x) &= 2x\\ g'(y) &= \cos y.\end{aligned} \tag{1.155}$$

Mit Gl. 1.152 erhält man

$$y'(x) = \cos x^2\cdot 2x. \tag{1.156}$$

Die Verkettung dieser zwei Funktionen lässt sich auch als $g(f(y))$ darstellen. Wir können uns die Kettenregel auch über die Beziehung

$$\frac{\mathrm{d}g}{\mathrm{d}x}=\frac{\mathrm{d}g}{\mathrm{d}x}\cdot\overbrace{\frac{\mathrm{d}f}{\mathrm{d}f}}^{=1}=\frac{\mathrm{d}g}{\mathrm{d}f}\cdot\frac{\mathrm{d}f}{\mathrm{d}x}. \tag{1.157}$$

herleiten, an der man gut die aus der Schule bekannte Formulierung »innere Ableitung · äußere Ableitung« ablesen kann.

1.5.4 Ableitung der Umkehrfunktion

Die Kettenregel ist ein sehr mächtiges Werkzeug, da sie es auch ermöglicht, die Ableitung komplizierterer Funktionen zu bestimmen. Es sollte bekannt sein, dass die Funktion $f(x)=\sin x$ die Ableitung $f'(x)=\cos x$ hat. Wie sieht es aber mit der Umkehrfunktion des Sinus, also mit

$f^{-1}(x) = \arcsin x$ aus? Wenn wir die Ableitung der Ausgangsfunktion $f(x)$ kennen, können wir die Ableitung der Umkehrfunktion bestimmen. Wir nutzen hierzu aus, dass die Verkettung von Funktion und zugehöriger Umkehrfunktion dem Argument x der Funktion entspricht, also dass

$$(f^{-1} \circ f)(x) = x \tag{1.158}$$

gilt. Am Beispiel $f(x) = x^2$ und $f^{-1}(x) = \sqrt{x}$ kann man sich das gut verdeutlichen. Es gilt $\sqrt{x^2} = x$. Leitet man Gl. 1.158 nach x ab, so erhält man auf der rechten Seite Eins, auf der linken Seite kann man die Kettenregel anwenden:

$$(f^{-1} \circ f)' = (f^{-1})' \cdot f' = (f^{-1}(f(x)))' \cdot f' = 1. \tag{1.159}$$

Daraus folgt

$$(f^{-1}(f(x)))' = \frac{1}{f'(x)}. \tag{1.160}$$

Jetzt muss man nur noch dafür sorgen, dass beide Seiten der Gleichung durch die gleiche Variable ausgedrückt werden. Mit $y = f(x)$ und $x = f^{-1}(y)$ (letzeres ist die Definition der Umkehrfunktion) folgt schließlich

$$(f^{-1}(y)' = \frac{1}{f'(f^{-1}(y))}. \tag{1.161}$$

Schauen wir uns das am Beispiel $f(x) = \sin x$ bzw. $f^{-1}(y) = \arcsin y$ genauer an. Aus Gl. 1.160 folgt

$$(f^{-1}(y))' = \frac{1}{\cos x} = \frac{1}{\cos\underbrace{(\arcsin y)}_{=x}} = \frac{1}{\sqrt{1 - \sin^2(\arcsin y)}}. \tag{1.162}$$

Mit $y = \sin(\arcsin y)$ und Umbenennung der Variable y in x bekommen wir einen Ausdruck für die Ableitung von $f(x) = \arcsin x$:

$$(\arcsin x)' = \frac{1}{\sqrt{1 - x^2}}. \tag{1.163}$$

1.5.5 Parameterform

Gegeben seien zwei Funktionen $x(t)$ und $y(t)$, die beide von der Variable t abhängen. Eine solche Funktion könnte beispielsweise einen Kreis darstellen, den man in der x, y-Ebene zeichnen kann. Für die Ableitung $\mathrm{d}y/\,\mathrm{d}x$ kann man nun schreiben:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\frac{\mathrm{d}y}{\mathrm{d}t}}{\frac{\mathrm{d}x}{\mathrm{d}t}} = \frac{\dot{y}}{\dot{x}}. \tag{1.164}$$

Bleiben wir beim Kreis und drücken die beiden Variablen aus über

$$\begin{aligned} x(t) &= r\cos\omega t \\ y(t) &= r\sin\omega t. \end{aligned} \tag{1.165}$$

Daraus folgt

$$\begin{aligned} \dot{x}(t) &= -r\omega\sin\omega t = \omega y \\ \dot{y}(t) &= r\omega\cos\omega t = \omega x \end{aligned} \tag{1.166}$$

und somit

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\dot{y}}{\dot{x}} = -\frac{y}{x}. \tag{1.167}$$

Wir prüfen Gl. 1.167 auf Plausibilität. Für $x > 0$ und $y \approx 0$ wird die Steigung $\mathrm{d}y/\,\mathrm{d}x$ sehr groß und negativ, für $x = 0$ erhält man eine horizontale Tangente, wie man es erwartet.

1.5.6 Ableitung von Vektoren

Die Bewegung eines Massenpunktes im Raum sei durch einen zeitlichen veränderlichen Vektor $\vec{r}(t) = (x(t), y(t), z(t))$ beschrieben. Der Vektor $\vec{r}$ wird dabei als Ortsvektor bezeichnet. Die Bahnkurve bezeichnet man als Trajektorie.[9] Die Ableitung des Ortsvektors $\vec{r}(t)$ nach der Zeit t ist die Geschwindigkeit $\vec{v}(t)$ des Massenpunktes. Die Ableitung ist dabei auf jede Komponente des Vektors anzuwenden, d. h.

$$\vec{v} = \lim_{\Delta t\to 0} \begin{pmatrix} \dfrac{x(t+\Delta t) - x(t)}{\Delta t} \\ \dfrac{y(t+\Delta t) - y(t)}{\Delta t} \\ \dfrac{z(t+\Delta t) - z(t)}{\Delta t} \end{pmatrix} = \begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}. \tag{1.168}$$

Der Geschwindigkeitsvektor $\vec{v}$ steht tangential zur Trajektorie, sein Betrag ist gleich der Geschwindigkeit in Richtung der Trajektorie. Für die Ableitung von Verknüpfungen von Vektoren gelten die gleichen Regeln wie für gewöhnliche Ableitungen, so z. B. für das Kreuzprodukt die Produktregel:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\vec{a}\times\vec{b}\right) = \frac{\mathrm{d}\vec{a}}{\mathrm{d}t}\times\vec{b} + \vec{a}\times\frac{\mathrm{d}\vec{b}}{\mathrm{d}t}. \tag{1.169}$$

9 Das Wort »Trajektorie« wird Tra-jek-to-ri-e ausgesprochen, also am Ende so wie Studie, Aktie, Serie, Hostie. Es wird nicht Tra-jek-to-rie wie in Energie oder Amnesie ausgesprochen.

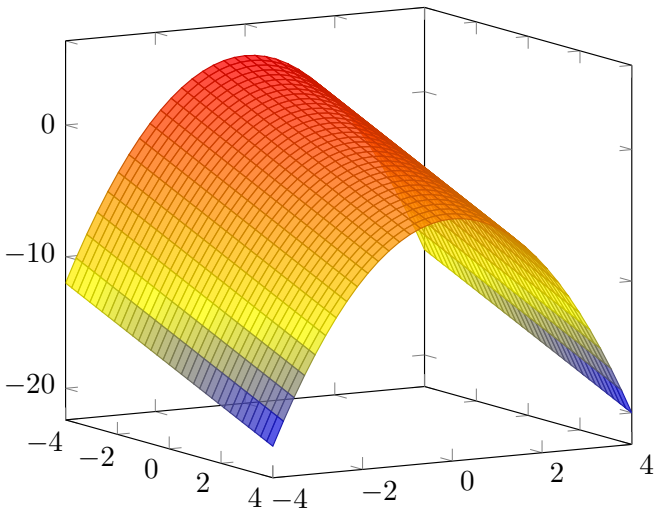


Abbildung 1.18: Die Funktion $f(x, y) = -y^2 - x$ als 3D-Plot.

1.5.7 Partielle Ableitung

Bisher haben wir nur Funktionen mit einer Variablen bezüglich ihrer Ableitung betrachtet. Funktionen können aber von mehreren Variablen abhängen. So kann beispielsweise ein Höhenprofil z eines Gebirges als Funktion zweier räumlicher Größen x und y ausgedrückt werden, d. h. als

$$z = f(x, y). \tag{1.170}$$

Für jede Kombination von Wertepaaren (x, y) gibt es eine eindeutige Größe z, in unserem Beispiel eine Höhe. Die Funktion $z = f(x, y)$ definiert eine Fläche, die – wenn man sich auf ihr bewegen würde – in verschiedenen Richtungen verschiedene Steigungen hat. Man kann sich das an Abb. 1.18 verdeutlichen, in der die Funktion $z = -y^1 - x$ dargestellt ist. Hier gibt es Bereiche geringer und hoher Steigung, ähnlich einem Berggrat, der seitlich steil abfällt. Die Steigung einer solchen Funktion ist also richtungsabhängig (vgl. Abb. 1.19). Partielle Ableitungen betrachten die Veränderung einer mehrdimensionalen Funktion in Richtung einer Koordinate, alle anderen Richtungen werden dabei als fest angenommen. Für eine Funktion $f(x, y)$ sind die partiellen Ableitung nach x und y folgendermaßen definiert:

$$\begin{aligned} \frac{\partial f}{\partial x} &= \lim_{\Delta x \to 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x} \\ \frac{\partial f}{\partial y} &= \lim_{\Delta y \to 0} \frac{f(x, y + \Delta x) - f(x, y)}{\Delta y}. \end{aligned} \tag{1.171}$$

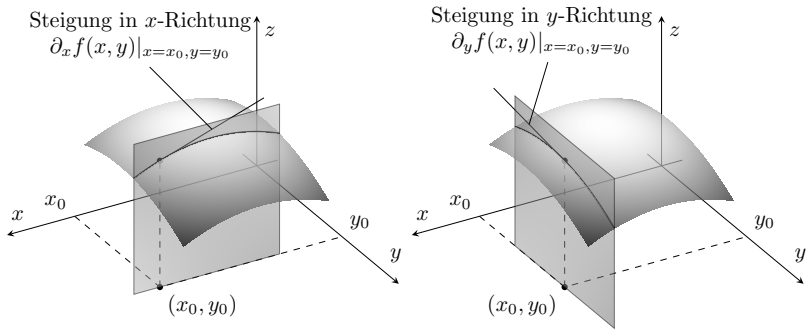


Abbildung 1.19: Die Steigung einer Funktion der Form $z = f(x, y)$ bezogen auf einen Punkt (x_0, y_0) ist abhängig von der Richtung, in die man sich bewegt.

Betrachten wir das Beispiel $f(x, y) = 3x^2y + \sin xy$. Die partiellen Ableitungen ergeben sich zu

$$\begin{aligned} \left.\frac{\partial f}{\partial x}\right|_{y=\text{const.}} &= 6xy + y \cdot \cos xy \\ \left.\frac{\partial f}{\partial y}\right|_{x=\text{const.}} &= 3x^2 + x \cdot \cos xy. \end{aligned} \tag{1.172}$$

Analog zu den »normalen« Ableitungen gibt es bei den partiellen ebenfalls höhere Ableitungen der Form $\partial^2 / \partial x^2$. Es sind zudem gemischte Ableitungen der Form

$$\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial y \partial x}. \tag{1.173}$$

Man kann zeigen, dass die Reihenfolge der partiellen Ableitung nach verschiedenen Variablen keine Rolle spielt (Satz von SCHWARZ). Es gilt also

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y}. \tag{1.174}$$

1.5.8 Differential von $f(x, y)$

Das Differential $\mathrm{d}f$ wurde bereits in Abs. 1.5.2 eingeführt. Für zwei Koordinaten x, y können wir es als

$$\mathrm{d}f = \left(\frac{\partial f}{\partial x}\right) \mathrm{d}x + \left(\frac{\partial f}{\partial y}\right) \mathrm{d}y \tag{1.175}$$

schreiben. Eine kleine Änderung des Funktionswertes $\mathrm{d}f$ kann geschrieben werden als Summe der Änderungen in x- und y-Richtung. Wir schauen uns eine Eigenschaft des Differentials für implizite Darstellungen von Funktionen an, also für Darstellungen der Form $F(x, y) = 0$. Betrachten wir den Einheitskreis $x^2 + y^2 - 1 = 0$ und beschränken uns auf positive y-Werte. In diesem Fall gilt:

$$\mathrm{d}F = \left(\frac{\partial F}{\partial x}\right) \mathrm{d}x + \left(\frac{\partial F}{\partial y}\right) \mathrm{d}y = 0. \tag{1.176}$$

Wir können dies in die Form

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{\left(\frac{\partial F}{\partial x}\right)}{\left(\frac{\partial F}{\partial y}\right)} \tag{1.177}$$

bringen. Wir können also die Ableitung $\mathrm{d}y/\,\mathrm{d}x$ mit den partiellen Ableitungen von F verknüpfen. Am Beispiel des Einheitskreises bedeutet dies, dass

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{2x}{2y} = -\frac{x}{y} \tag{1.178}$$

gilt, was wir bereits als Ergebnis parametrischer Ableitungen in Abs. 1.5.5 erhalten haben.

1.5.9 Zeitableitungen

In der Physik interessiert oft die zeitliche Veränderung von physikalischen Größen. Solche Veränderungen werden durch Differentialgleichungen beschrieben. Zum Lösen dieser Differentialgleichungen benötigt man eine Reihe von mathematischen Methoden, die im Verlauf des Studiums gelehrt werden. Als Basiswissen sollte man aber die Techniken der Differentialrechnung beherrschen, also das, was man gemeinhin als »ableiten« bezeichnet. Darunter versteht man die Berechnung lokaler Veränderungen von Funktionen, die man üblicherweise als Tangentensteigung der Funktion angibt. Die Tangentensteigung wird als Grenzwert der Sekantensteigung für sehr kleine Δx berechnet. Die Größe x kann für eine beliebige Größe stehen (Ort, Zeit, Konzentration, Druck usw.), muss als nicht unbedingt eine Länge bezeichnen. Den Grenzwert bezeichnet man als Differentialquotienten $\mathrm{d}f/\,\mathrm{d}x$:

$$f'(x) = \frac{\mathrm{d}f}{\mathrm{d}x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}. \tag{1.179}$$

Für den Differentialquotienten gibt es bei räumlichen Ableitungen die Schreibweise $f'(x)$. Bei zeitlichen Ableitungen hat sich die Konvention durchgesetzt, einen Punkt über die Funktion statt eines Strichs daneben zu setzen, also:

$$\dot{f}(t) = \frac{\mathrm{d}f}{\mathrm{d}t} = \lim_{\Delta t \to 0} \frac{f(t+\Delta t) - f(t)}{\Delta t}. \tag{1.180}$$

Zeitableitungen werden also abgekürzt in der Punktschreibweise geschrieben. Höhere Ableitungen werden durch mehr Punkte gekennzeichnet, wobei man eigentlich fast nur Ableitungen bis zweiter Ordnung braucht. Das erspart Schreibarbeit und macht Formeln übersichtlicher. So lässt sich beispielsweise die Bewegungsgleichung eines gedämpften harmonischen Oszillators mit Kreisfrequenz ω sehr einfach in der Form

$$\ddot{x} + 2\beta \cdot \dot{x} + \omega^2 x = 0 \tag{1.181}$$

statt

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + 2\beta\frac{\mathrm{d}x}{\mathrm{d}t} + \omega^2 x = 0 \tag{1.182}$$

schreiben. Wir werden den harmonischen Oszillator an späterer Stelle im Detail besprechen, also keine Angst, wenn die Formel Ihnen an dieser Stelle nichts sagt.

1.6 Integralrechnung

1.6.1 Das unbestimmte Integral

Die Integration einer Funktion lässt sich als Umkehrung der Differentiation auffassen. Bei der Differentiation geht es darum, die Ableitung $f'(x)$ einer Funktion $f(x)$ zu bestimmen. Bei der Integration ist es nun so, dass uns eine Funktion $f(x)$ vorliegt, die wir als Ableitung einer Funktion $F(x)$ interpretieren können und wir diese Funktion $F(x)$ bestimmen wollen. Es gilt also:

$$F'(x) = f(x). \tag{1.183}$$

Die Funktion $F(x)$ wird als Stammfunktion bezeichnet. Unter Integrieren versteht man in erster Linie das Auffinden dieser Stammfunktion. Die Schreibweise hierfür lautet:

$$F(x) = \int f(x)\,\mathrm{d}x. \tag{1.184}$$

Eine Reihe wichtiger Stammfunktionen sind in Tabelle 1.4 zusammengefasst.

Tabelle 1.4: Wichtige Stammfunktionen (ohne additive Konstante).

$f(x)$	$F(x)$	$f(x)$	$F(x)$
x^n	$\frac{1}{n+1}x^{n+1}$	a	$a \cdot x$
$\frac{1}{x}$	$\ln x$	e^x	e^x
e^{ax}	$\frac{1}{a}\mathrm{e}^{ax}$	a^x	$\frac{1}{\ln a}a^x$
$\sin x$	$-\cos x$	$\cos x$	$\cos x$

Es gibt zu einer Funktion $f(x)$ eine unendliche Anzahl von Stammfunktionen, da $F(x)$ nur bis auf eine beliebige additive Konstante C bestimmbar ist. So ergibt z. B. die Ableitung der Funktion $F(x) = 2x^2 + 5$ die Funktion $f(x) = 4x$. Die Funktion $F(x) = 2x^2 + 7$ hat aber die selbe Ableitung, ist also auch Stammfunktionen von $f(x)$, genauso wie jede andere Konstante, die man additiv hinzufügt.

In der Physik hat man es häufig mit unbestimmten Integralen zu tun. Der Wert der Integrationskonstanten C ergibt sich dabei oft aus den Randbedingungen der physikalischen Problemstellung. Betrachten wir den freien Fall eines (punktförmigen) Körpers im Schwerefeld der Erde, also unter Vernachlässigung der Luftreibung. Die Bewegung kann in einer Dimension beschrieben werden, wir definieren hierzu eine z-Achse, deren positive Richtung nach oben zeigt. Die Erdbeschleunnigung $\vec{g}$ ist dieser Achse entgegen gerichtet, es gilt also $\vec{g} = -g\vec{e}_z$. Die zeitliche Ableitung der Geschwindigkeit ist die Beschleunigung, in unserem Fall $|g| = \text{const.}$. Wenn wir die Beschleunigung nach der Zeit integrieren, so erhält man folglich die Geschwindigkeit v:

$$v(t) = \int (-g)\,\mathrm{d}t = -gt + C_1. \tag{1.185}$$

Das negative Vorzeichen ist eine Folge der Festlegung des Koordinatensystems. Die Konstante C_1 kann aus den Anfangsbedingungen bestimmt werden. Wird der Körper z. B. zur Zeit $t = 0$ mit einer Anfangsgeschwindigkeit v_0 entgegen Richtung der wirkenden Beschleunigung gestartet, so erhält man

$$v(t) = -gt + v_0. \tag{1.186}$$

Integriert man nochmals nach der Zeit, so erhält man den Ort $z(t)$:

$$z(t) = \int v(r)\,\mathrm{d}t = -\int gt\,\mathrm{d}t + \int v_0\,\mathrm{d}t = -\frac{1}{2}gt^2 + v_0 t + C_2. \quad (1.187)$$

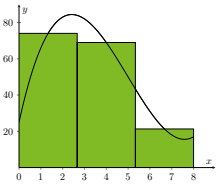
Die Konstante C_2 wird wieder über eine Anfangsbedingung festgelegt. Wir legen fest, dass bei $t = 0$ der Körper aus einer Höhe $z(t = 0) = z_0$ startet. Daraus erhält man die Bewegungsgleichung für den freien Fall mit Anfangsgeschwindigkeit v_0 und Starthöhe z_0:

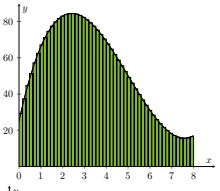
$$z(t) = -\frac{1}{2}gt^2 + v_0 t + z_0 \quad (1.188)$$

Prüfen Sie durch zweimaliges Ableiten nach der Zeit, ob diese Bewegungsgleichung auf die konstante Erdbeschleunigung $a = -g$ führt.

1.6.2 Das bestimmte Integral

Beim bestimmten Integral geht es darum, einen konkreten Zahlenwert auszurechnen, also nicht wie beim unbestimmten Integral eine neue Funktion zu bestimmen. Dieser Zahlenwert entspricht der Fläche unterhalb der Funktion innerhalb eines Intervalls (a, b). Sie wird auf der anderen Seite durch die x-Achse begrenzt. Die Grundidee, diese Fläche zu berechnen, ist relativ einfach. Man führt die Gesamtfläche auf eine Summe von kleineren Rechteckflächen (vertikalen Streifen) zurück, die alle die gleiche Breite Δx haben und die einfach addiert werden. Die Höhe dieser Rechteckflächen ist dabei durch die Funktionswerte $f(x)$ gegeben. Zerlegt man das Intervall (a, b) in n äquidistante Teilstücke der Breite $(b - a)/n$, so kann man auf der x-Achse eine Folge von x-Werten definieren, die durch

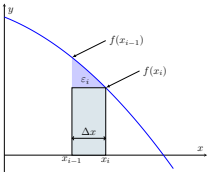




$$x_i = a + i \cdot \Delta x \qquad i = 0, \ldots, n \quad (1.189)$$

beschrieben werden. Wir prüfen auf Plausibilität: für $i = 0$ ist $x_0 = a$, für $i = n$ ist $x_n = b$. Man kann an den drei Beispielen am Rand erkennen, dass die Fläche umso genauer wird, je mehr Teilstücke man verwendet, je kleiner also das Δx wird. Die Gesamtfläche erhält man durch Addition der Teilflächen s_i:

$$s_{(a,b)} = s_1 + s_2 + \cdots + s_n = \sum_{i=1}^{n} s_i. \quad (1.190)$$



Beachten muss man hier, dass die Rechteckfläche mit der Bezeichnung s_i zum Intervall (x_{i-1}, x_i) gehört, da zwar der Index i bei Null beginnt, man die Nummerierung der Flächen aber mit 1 anfängt. Die Fläche s_i erhält man nun durch Bildung der Summe

$$s_i = f(x_i)\Delta x + \varepsilon_i. \tag{1.191}$$

Dabei steht das ε_i für eine Restfläche, die nicht vom Produkt $f(x_i)\Delta x$ erfasst wird. Die Gesamtfläche kann damit geschrieben werden als

$$s_{(a,b)} = \sum_{i=1}^{n} f(x_i)\Delta x + \sum_{i=1}^{n} \varepsilon_i. \tag{1.192}$$

Im Fall immer kleiner werdender Intervalle Δx werden die einzelnen Restflächen immer geringer, so dass

$$s_{(a,b)} \approx \sum_{i=1}^{n} f(x_i)\Delta x \tag{1.193}$$

gilt. Im Grenzfall unendlich kleiner Intervalle geht dies in die exakte Fläche unter der Kurve über und führt zur Definition des bestimmten Integrals:

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{\Delta x \to 0} \sum_{i=1}^{n} f(x_i)\Delta x. \tag{1.194}$$

Die Variable x ist die Integrationsvariable, $\mathrm{d}x$ das Differential. Hier wird auch nochmal der Bezug zur Darstellung der Ableitung als $(\mathrm{d}f\,/\,\mathrm{d}x)$ deutlich. Nehmen wir als Beispiel, dass die Ableitung einer Funktion f nach x eine Konstante a ergibt:

$$\frac{\mathrm{d}f}{\mathrm{d}x} = a. \tag{1.195}$$

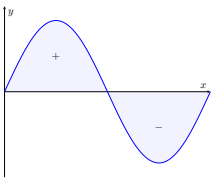
Man kann nun das $\mathrm{d}x$ auf die rechte Seite bringen und erhält

$$\mathrm{d}f = a\,\mathrm{d}x. \tag{1.196}$$

Wir können nun auf beiden Seiten eine Integration durchführen, auf der linken Seite nach f auf der rechten nach x, wobei es keine Rolle spielt, ob wir das bestimmte oder unbestimmte Integral bilden:

$$\begin{aligned} \int_{f_a}^{f_b} \mathrm{d}f &= a \int_{x_a}^{x_b} \mathrm{d}x \\ f_b - f_a = \Delta f &= a(x_b - x_a) = a\Delta x \end{aligned} \tag{1.197}$$

Die Integration führt also von den infinitesimal kleinen Intervallen d zu endlich großen Intervallen Δ, in denen sich die reale Welt abspielt. Die Idee sowohl der Integration als auch der Differentiation ist die lineare Approximation von funktionalen Zusammenhängen in sehr kleinen Intervallen, die man mathematisch einfacher behandeln kann sowie die Zurückführung auf endliche (nicht-infinitesimale) Größen.



Die Fläche zwischen positiven Funktionswerten und x-Achse wird positiv gezählt, zwischen negativen Funktionswerten und x-Achse entsprechend negativ. Daraus folgt z. B. sofort, dass

$$\int_0^{2\pi} \sin x \, \mathrm{d}x = 0 \tag{1.198}$$

sein muss.

Wir haben den Differentialquotienten beim Ableiten auf Plausibilität geprüft, indem wir gezeigt haben, dass die Ableitung von $f(x) = x^2$ zu $f'(x) = 2x$ führt. Wir wollen dieses Plausibilitätsprüfung nun auch für die Integration einer einfachen Funktion, genauer für $f(x) = x$, durchführen. Wir wissen natürlich, dass

$$\int_a^b x \, \mathrm{d}x = \left[\frac{1}{2}x^2\right]_a^b = \frac{b^2 - a^2}{2} \tag{1.199}$$

ergibt. Betrachten wir nun Gl. 1.193 für die Funktion $f(x)$. Die Funktion sagt aus, dass ihr Argument dem Funktionswert entspricht. Als Argumente der Funktion nehmen wir die Werte $x_i 0 = a + i \cdot \Delta x$, die wir durch Intervallbildung erhalten, was bedeutet, dass wir das Δx als

$$\Delta x = \frac{b - a}{n} \tag{1.200}$$

definieren. Es gilt also

$$f(x_i) = f(a + i\Delta x) = a + i\Delta x. \tag{1.201}$$

Somit gilt für die Fläche $s_{(a,b)}$:

$$\begin{aligned} s_{(a,b)} &= \sum_{i=1}^{n} (a + i\Delta x)\Delta x \\ &= \sum_{i=1}^{n} a\Delta x + \sum_{i=1}^{n} i(\Delta x)^2 \\ &= a\Delta x \sum_{i=1}^{n} 1 + (\Delta x)^2 \sum_{i=1}^{n} i \\ &= a\Delta x \cdot n + (\Delta x)^2 \frac{n(n+1)}{2} \end{aligned} \tag{1.202}$$

Mit der Intervalldefinition von Δx und ein paar einfachen Umformungen erhält man:

$$s_{(a,b)} = a(b - a) + \frac{1}{2}(b - a)^2 \left(1 + \frac{1}{n}\right). \tag{1.203}$$

Im Grenzfall $\Delta x \to 0$, der gleichbedeutend mit $n \to \infty$ ist, kommt man auf

$$s_{(a,b)} = \frac{b^2 - a^2}{2}, \tag{1.204}$$

was dem bekannten Ergebnis entspricht.

1.6.3 Rechenregeln für bestimmte Integrale

Folgende Rechenregeln für bestimmte Integrale werden häufig gebraucht:

- Fallen die Integrationsgrenzen zusammen, ist die resultierende Fläche Null
$$\int_a^a f(x)\,\mathrm{d}x = 0 \tag{1.205}$$
- Integrale können in Teilintegrale zerlegt werden
$$\int_a^c f(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \int_b^c f(x)\,\mathrm{d}x \tag{1.206}$$
Hierbei ist zu beachten, dass b in Intervall (a, c) liegen muss.
- Vertauscht man die Integrationsgrenzen, ändert sich das Vorzeichen
$$\int_a^b f(x)\,\mathrm{d}x = -\int_b^a f(x)\,\mathrm{d}x \tag{1.207}$$

1.6.4 Partielle Integration

Das Integrieren von Funktionen ist im Gegensatz zum Differenzieren deutlich komplizierter. Es gibt daher eine Vielzahl von Techniken, die man sich aneignen kann wie beispielsweise das Substituieren. Ziel hierbei ist es, dass man das Integral in eine Form bringt, deren Lösung man kennt bzw. deren Lösung in einer Formelsammlung zu finden ist. Eine Technik, die auf der Produktregel der Differentiation basiert, soll hier exemplarisch vorgestellt werden: die partielle Integration (engl. *integration by parts*). Erinnern wir uns an die Produktregel:

$$(fg)' = f'g + fg'. \tag{1.208}$$

Wir können diese Umstellen und eine Integration durchführen:

$$\begin{aligned}\int f'g\,\mathrm{d}x &= \int (fg)'\,\mathrm{d}x - \int fg'\,\mathrm{d}x \\ &= fg - \int fg'\,\mathrm{d}x.\end{aligned} \tag{1.209}$$

Führen wir als Beispiel die Integration von $f(x) = x\mathrm{e}^x\,\mathrm{d}x$ durch. Man kann nun frei auswählen, welche der beiden Teilfunktionen das g und welche das f' sein soll. Wählt man $g(x) = x$, so ergibt die Ableitung $g'(x) = 1$. Wir erhalten damit

$$\int x\mathrm{e}^x\,\mathrm{d}x = x\mathrm{e}^x - \int 1 \cdot \mathrm{e}^x\,\mathrm{d}x = \mathrm{e}^x(x-1). \tag{1.210}$$

Tabelle 1.5: Ableitungen der Form $f'(x) = nx^{n-1}$ für verschiedene n.

n	-2	-1	0	1	2	3
$f'(x)$	$-2x^{-3}$	$-x^{-2}$	0	x^0	$2x^1$	$3x^2$

Wir können die Methode des partiellen Integrierens auch auf $f(x) = \ln x$ anwenden:

$$\int \ln x \, \mathrm{d}x = \int 1 \cdot \ln x = x \cdot \ln x - \int x \frac{1}{x} \, \mathrm{d}x = x \cdot \ln x - x = x(\ln x - 1). \tag{1.211}$$

1.6.5 Logarithmen einmal anders betrachtet

Abschließend zum Abschnitt über Integration soll nochmal über Logarithmen gesprochen werden. Der Zugang zu Logarithmen aus der Sicht eines Mathematikers unterscheidet sich von unserer bisherigen anwendungsorientierten Betrachtungsweise, bei der es mehr um die benötigten Rechenregeln gegangen ist. Betrachten wir zunächst noch einmal die Ableitungsregel für Funktionen der Form $f(x) = x^n$. Hier gilt

$$f'(x) = nx^{n-1}. \tag{1.212}$$

Tabelle 1.5 zeigt eine Reihe von Ableitungen für aufsteigende Werte von n, beginnend mit $n = -2$. Was auffällt ist, dass es in dieser Reihe keine Ableitung der Form $f'(x) = x^{-1}$ gibt. Nun ist eine Funktion dieser Form keineswegs ungeeignet, eine Ableitung darzustellen. Bis auf die Stelle $x = 0$ ist diese Funktion stetig differenzierbar, somit als Ableitung für eine noch zu findende Funktion geeignet. In der Mathematik geht man nun folgenden Weg: man definiert die gesuchte Funktion für $x > 0$ über ein Integral, welches folgende Form haben soll:

$$\ln x = \int_1^x u^{-1} \, \mathrm{d}u. \tag{1.213}$$

Die Einschränkung auf $x > 0$ ergibt sich aus der Tatsache, dass man nicht über die Singularität integrieren kann. Die obere Integrationsgrenze x ist entspricht dem Argument der Funktion $\ln x$. Die untere Integrationsgrenze beträgt 1. Dieser Wert ist als untere Integrationsgrenze zwingend notwendig, da

$$\int_1^x u^{-1} \, \mathrm{d}u = [\ln u]_1^x = \ln x - \ln 1 \tag{1.214}$$

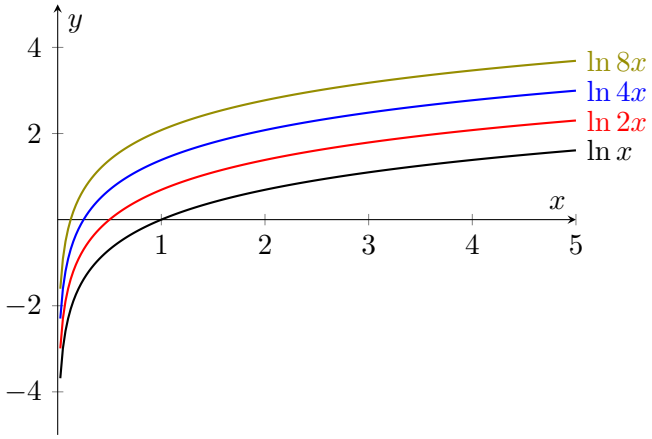


Abbildung 1.20: Graphische Darstellung der Funktion $\ln nx$ für $n = 1, 2, 4, 8$.

ergibt und Gl. 1.213 als Definitionsgleichung gelten soll. Somit muss $\ln 1 = 0$ sein. Wir kennen also damit bereits einen Funktionswert der gesuchten Funktion. Man sieht, dass diese Bedingung auch über die Integraldefinition erfüllt ist, wenn man als obere Integrationsgrenze $x = 1$ verwendet:

$$\ln 1 = \int_1^1 u^{-1}\,\mathrm{d}u = 0 \tag{1.215}$$

Wenn die beiden Integrationsgrenzen gleich sind, beträgt die Fläche null. Ebenfalls kann aus Gl. 1.213 geschlossen werden, dass $\ln x > 0$ für $x > 1$ sein muss, ebenso $\ln x < 0$ für $0 < x < 1$.

Der Integrand in Gl. 1.213 ist positiv, somit muss die Funktion $\ln x$ streng monoton steigend sein. Man kann durch Grenzwertbetrachtungen zeigen, dass für $x \to \infty$ die Steigung von $\ln x$ immer geringer wird, da ja die Steigung durch x^{-1} beschrieben wird, diese bei steigenden x-Werten asymptotisch gegen null läuft. Für $x \to 0$ wird analog die Steigung von $\ln x$ unendlich groß. Zusammen mit der Bedingung $\ln x < 0$ für $x < 1$ folgt daraus, dass $\ln x \to -\infty$ für $x \to 0$ gelten muss.

Man kann sich nun weitere Eigenschaften unserer gesuchten Funktion überlegen. Wenn die Ableitung von $\ln x$ die Funktion x^{-1} ergibt, dann kann man mit Hilfe der Kettenregel sich überlegen, dass für einen konstanten Faktor a die Funktion $\ln(ax)$ die Ableitung

$$(\ln(ax))' = \frac{1}{x} \tag{1.216}$$

hat. Die Ableitungen von $\ln x$ und $\ln ax$ haben also dieselbe Steigung (siehe Abb. 1.20), somit unterscheiden sich die beiden Funktionen nur um eine additive Konstante c (die ja beim Ableiten verschwindet). Es gilt also

$$\ln(ax) = \ln x + c. \tag{1.217}$$

Setzt man in dieser Gleichung $x = 1$ ein, erhält man $c = \ln a$. Daraus folgt unmittelbar die Beziehung

$$\ln(ab) = \ln a + \ln b. \tag{1.218}$$

Setzt man $a = b$, so folgt daraus dann

$$\ln a^2 = \ln a + \ln a = 2 \ln a. \tag{1.219}$$

Verallgemeinert muss dann gelten, dass

$$\ln x^n = n \cdot \ln x \tag{1.220}$$

ist. Da unsere Funktion stetig ist, kann n für beliebige reelle Exponenten stehen.

Bisher kennen wir für die Funktion $\ln x$ nur einen einzigen Funktionswert, nämlich $\ln 1 = 0$. Erweitert man dies mit

$$\ln 1 = \ln\left(\frac{1}{x} \cdot x\right) = \ln\frac{1}{x} + \ln x = 0 \tag{1.221}$$

folgt damit unmittelbar

$$\ln\frac{x}{y} = \ln x - \ln y. \tag{1.222}$$

Wie sieht es nun mit anderen Funktionswerten aus? Betrachten wir hierzu das Integral

$$\ln(1 + x) = \int_1^{1+x} \frac{1}{u}\,\mathrm{d}u. \tag{1.223}$$

Wir führen eine Variablentransformation mit $u = 1 + v$ durch, wodurch wir $\mathrm{d}u = \mathrm{d}v$ erhalten. Beachten Sie, dass die Variable u der Variable x entspricht, die Umbenennung ergibt sich aus der Tatsache, dass Integrationsvariable und Variablen der Integrationsgrenzen andere Benennungen erfordern. Wir müssen die Integrationsgrenzen entsprechend mit transformieren und erhalten

$$\ln(1 + x) = \int_0^x \frac{1}{1+v}\,\mathrm{d}v. \tag{1.224}$$

Der Integrand entspricht einer binomischen Reihe und kann als

$$(1+v)^{-1} = \sum_{k=0}^{\infty}(-v)^k = 1 - v + v^2 - v^3 + v^4 - v^5 + \ldots \tag{1.225}$$

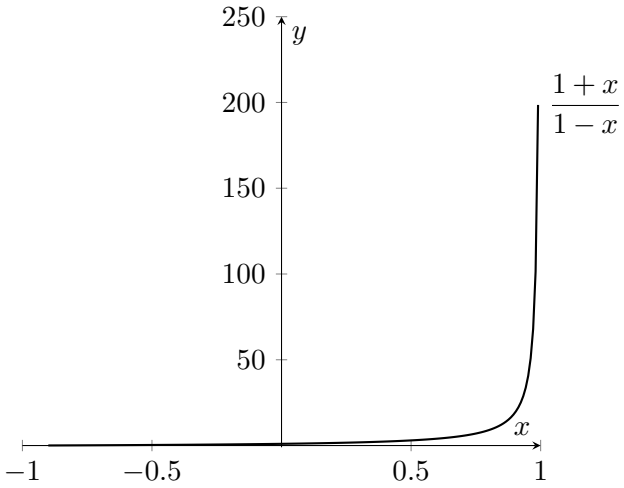


Abbildung 1.21: Graphische Darstellung der Funktion $\frac{1+x}{1-x}$.

geschrieben werden. Für Werte $|v| < 1$ konvergiert diese Reihe, so dass wir die Integration über die einzelnen Summanden ausführen können, um ein endliches Ergebnis zu bekommen:

$$\begin{aligned} \ln(1+x) &= \int_0^x \left(1 - v + v^2 - v^3 + v^4 - \dots\right) \mathrm{d}v \\ &= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots \end{aligned} \tag{1.226}$$

Wir können damit für das Intervall $(0, 2)$ alle Werte der Funktion $\ln x$ darstellen, was natürlich etwas unbefriedigend ist, da alle Werte $x > 2$ noch nicht abgedeckt sind. Um diesem Dilemma zu entkommen, bedient man sich einer geschickten Wahl des Funktionsarguments von $\ln x$. Man setzt hier den Ausdruck $X = (1+x)/(1-x)$ ein und erhält damit

$$\ln X = \ln\left(\frac{1+x}{1-x}\right) = \ln(1+x) - \ln(1-x), \tag{1.227}$$

wodurch man alle Werte von 0 bis unendlich abbilden kann. Die Funktion $X = (1+x)/(1-x)$ ist in Abb. 1.21 dargestellt. Alle Zahlen $0 < X < \infty$ werden für $|x| < 1$ abgebildet. Damit ist es nun möglich, alle Funktionswerte von $\ln x$ darzustellen, nämlich als Differenz zweier Logarithmen aus dem Bereich 0 bis 2.

Die Funktion $\ln x$ ist eine streng monoton wachsende Funktion und folglich auch umkehrbar. Die Umkehrfunktion wird als e^x bezeichnet, somit gilt also

$$\ln \mathrm{e}^x = x. \tag{1.228}$$

Wir leiten nun beide Seiten ab und erhalten unter Berücksichtigung der Kettenregel:

$$\frac{1}{\mathrm{e}^x} \cdot (\mathrm{e}^x)' = 1. \tag{1.229}$$

Daraus folgt, dass

$$(\mathrm{e}^x)' = \mathrm{e}^x \tag{1.230}$$

sein muss. Die Ableitung der e-Funktion e^x ergibt e^x.

Wir haben ausgehend von einer Lücke in der bekannten Ableitungsvorschrift ($f'(x) = nx^{n-1}$), also dem Fehlen einer Ableitung der Form x^{-1} eine Funktion $\ln x$ über ein Integral definiert und konnten wichtige Eigenschaften dieser Funktion daraus bestimmen. Natürlich handelt es sich bei der Funktion um den natürlichen Logarithmus, was die Bezeichnung bereits angedeutet hat. Wir konnten zudem aus dieser Definition zeigen, dass die Ableitung der Umkehrfunktion, also von e^x wieder die Funktion e^x ergibt, gänzlich ohne die Ableitungsdefinition über den Differentialquotienten zu verwenden. Man kommt also mit ein wenig Nachdenken häufig auf verschiedenen Wegen zu neuen Erkenntnissen, was dieser abschließende Abschnitt zeigen will.

2 KEGELSCHNITTE

2.1 Historischer Überblick

Unter Kegelschnitten versteht man eine Gruppe von Kurven, die man geometrisch als Schnitte einer Ebene mit der Oberfläche eines unendlich ausgedehnten Doppelkegels interpretieren kann (vgl. Abb. 2.1). Sofern die Ebene die Kegelspitze enthält, erhält man entweder einen Punkt oder zwei sich schneidende Geraden. Man spricht hier von entarteten Kegelschnitten. In allen anderen Fällen ergeben sich die sog. nicht-zerfallenden Kegelschnitte (Ellipse, Parabel, Hyperbel), die man gemeinhin unter Kegelschnitten versteht und um die es in diesem Kapitel gehen soll.

Kegelschnitte werden in der mathematischen und physikalischen Ausbildung meist etwas stiefmütterlich behandelt, spielen aber dennoch eine wichtige Rolle in der Raumfahrt und auch in der Optik. Kegelschnitte sind Lösungen des Zweikörperproblems der Mechanik und beschreiben die möglichen Trajektorien von Planeten und Kometen aufgrund der gravitativen Wechselwirkung.

Erstmal erwähnt wurden Kegelschnitte in den Werken von Menaichmos (ca. 380 bis 320 v. Chr.), einem griechischen Mathematiker der Platonischen Akademie[1]. Menaichmos entdeckte die ≫Parabel≪ und die ≫Hyperbel≪. Die Gesamtheit der Kegelschnitte wurden dann in dem großen Werk ≫Konika≪ (Über Kegelschnitte) von Apollonius von Perge beschrieben. Dieses Buch bildete die Grundlage für die spätere Beschreibung der Bewegung der Planeten durch Ptolemäus (ca. 100-160 n. Chr.).

Wir werden an anderer Stelle die geometrischen Parameter der Kegelschnitte mit physikalischen Größen in Verbindung bringen. Um überhaupt zu erkennen, dass diese Verbindung besteht, ist es unabdingbar, sich mit den grundlegenden Begriffen aus mathematischer Sicht vertraut zu machen. Dies soll in einem überschaubaren Umfang getan

1 Der Begriff ≫Akademie≪, der heutzutage für eine Gelehrtengesellschaft steht, bezieht sich auf Akademos, einer Gestalt aus der griechischen Mythologie. Akademos rettete die Stadt Athen vor der Zerstörung durch die Dioskuren (Söhne des Zeus) Kastor und Polydeukes. Diese wollten ihre Schwester Helena, die von Theseus entführt und versteckt worden war, befreien und drohten mit der Zerstörung von Athen. Akademos verriet den beiden den Aufenthaltsort von Helena. Helena wurde später nochmal entführt, was den Trojanischen Krieg auslöste.

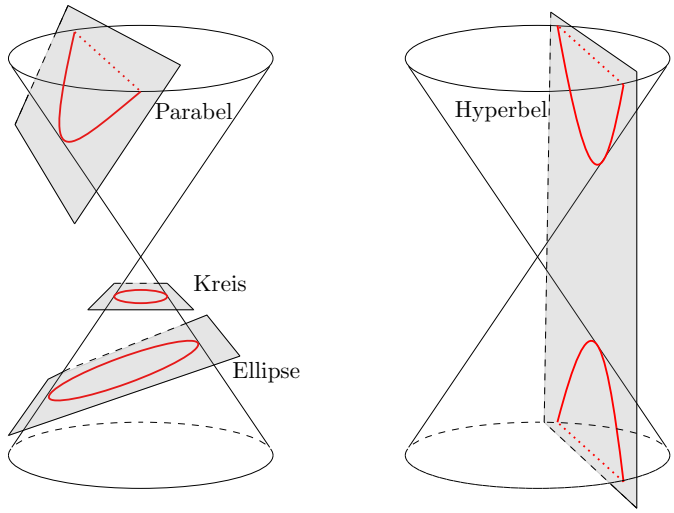


Abbildung 2.1: Kegelschnitte

werden, da es zu Kegelschnitten eine Fülle an Informationen gibt, man sich hier schnell in weniger wichtigen Details verlaufen kann.

2.2 Kreis und Ellipse

Beginnen wir mit dem Kreis K_r, den man als eine Menge von Punkten (x, y) mit einem festen Abstand r vom Kreismittelpunkt definieren kann. Das kann man schreiben als

$$K_r = \{(x, y) / x^2 + y^2 = r^2\}. \tag{2.1}$$

Teilt man nun die definierende Bedingung $x^2 + y^2 = r^2$ durch r^2, so ergibt sich die Darstellung

$$\frac{x^2}{r^2} + \frac{y^2}{r^2} = 1, \tag{2.2}$$

von der man auf die Ellipsengleichung 2.3 verallgemeinern kann.

Ellipsengleichung

Eine Ellipse in Mittelpunktsform ist definiert als Menge aller Punkte $P(x, y)$ mit

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \tag{2.3}$$

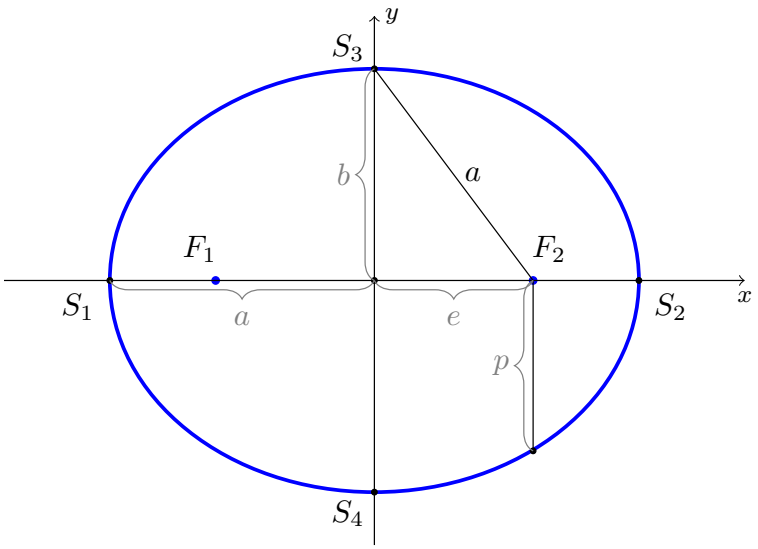


Abbildung 2.2: Wichtige Kenngrößen einer Ellipse.

Man spricht hier auch von der Normalenform bzw. Mittelpunktsgleichung der Ellipse. Eine Ellipse ist also eine Menge von Punkten (x, y), die Gl. 2.3 erfüllt. Die Größen a und b bezeichnet man als große und kleine Halbachse der Ellipse, der Kreis ist damit ein Spezialfall der Ellipse für $a = b = r$. Die Halbachse a entspricht dem Umkreis der Ellipse, die Halbachse b dem Innenkreis. Eine Ellipse, deren Mittelpunkt mit dem Ursprung eines kartesischen Koordinatensystems zusammenfällt und deren große Halbachse o. B. d. A. in x-Richtung zeigt, ist in Abb. 2.2 dargestellt. Der Ellipse können zwei Brennpunkte zugeordnet werden, die mit F_1 und F_2 bezeichnet werden. Würde man in den Brennpunkt eines elliptischen Körpers eine kleine Lampe aufstellen, die in alle Richtungen Licht abstrahlt, so wird sämtliches Licht im anderen Brennpunkt gebündelt werden (und von dort dann zurück in den Brennpunkt geworfen werden, in dem die Lampe steht). Steht die Lampe an beliebiger Stelle innerhalb der Ellipse, so wird sämtliches Licht in den beiden Brennpunkten fokussiert. In dem lesenswerten Buch *Feynmans verschollene Vorlesung. Die Bewegung der Planeten um die Sonne* von David L. Goodstein und Judith R. Goodstein wird auf Grundlage dieser Reflexionseigenschaft und ein paar geometrischen Überlegungen gezeigt, dass aus den Ellipsenbahnen ein Kraftgesetz mit einer $1/r^2$-Abhängigkeit wirken muss. Die Beweisführung entspricht in etwa der Darstellung von Isaac Newton in seinem 1886 veröffentlichten Hauptwerk »Philosphiae Naturalis Principia Mathematica«, der auf Basis der Keplerschen Beobachtungen sein Gravitationsgesetz beweisen konnte.

Die fokussierende Eigenschaft von Ellipsen wird technisch vielfältig ausgenutzt. Eine Anwendung ist die gezielte Aufheizung einer Probe. Hierzu wird eine Heizquelle in einen der Brennpunkte, die Probe in den anderen Brennpunkt gelegt. Wärmestrahlung der Quelle wird gerichtet auf den anderen Brennpunkt geleitet, die Probe wird dabei von allen Seiten angestrahlt. Eine andere (medizinische) Anwendung ist die Zertrümmerung von Nierensteinen durch Stoßwellen. Der Patient liegt dabei in einer wassergefüllten elliptischen Wanne und wird so positioniert, dass die im zweiten Brennpunkt ausgesendeten Stoßwellen den Nierenstein treffen, der sich im ersten Brennpunkt befindet, und auch erst dort durch ihre hohe Intensität das (störende) Gewebe schädigen, während das gesunde Gewebe außerhalb des Fokus nicht geschädigt wird.

In der Optik wird dem Brennpunkt eine Brennweite f zugeordnet. Dieser entspricht bei allen Kegelschnitten dem Abstand des Brennpunkts zum nächstgelegenen Scheitelpunkt. Bei der Parabel ist dies gerade $f = p/2$, bei Ellipse und Hyperbel $f = |a - e|$.

Der Abstand zwischen Ursprung und Brennpunkt wird als lineare Exzentrizität e bezeichnet und hat die Dimension einer Länge. Zieht man von einem der beiden Brennpunkte eine vertikale Linie bis zur Ellipse, erhält man den Halbparameter p der Ellipse (engl. *semi-latus rectum*). Die Strecke der Verbindungslinie von Brennpunkt zu den Punkten $(0, \pm b)$ entspricht der großen Halbachse a. Die Ellipse hat insgesamt vier Scheitelpunkte S_1, S_2, S_3, S_4. Die Scheitelpunkte S_1 und S_2 bezeichnet man als Hauptscheitel, die beiden anderen als Nebenscheitel. Die beiden Hauptscheitel sind dadurch gekennzeichnet, dass sie die größte Kurvenkrümmung aufweisen, die beiden Nebenscheitel weisen die kleinste Krümmung der Ellipsenbahn auf. Aus Symmetriegründen haben die beiden Hauptscheitel den Abstand $2a$, die beiden Nebenscheitel den Abstand $2b$.

Kepler hat herausgefunden, dass sich die Planeten auf Ellipsenbahnen bewegen, in deren einem Brennpunkt die Sonne steht. Die beiden Scheitelpunkte, die der Sonne am nächsten bzw. am weitesten entfernt sind, also die beiden Hauptscheitel, werden als Apsiden (Singular: Apsis) bezeichnet. Der nächstgelegene Scheitelpunkt ist die Periapsis, der weitest entfernte die Apoapsis. Für viele Himmelskörper gibt es spezielle Bezeichnungen für die Apsiden, die meist an griechische oder lateinische Bezeichnungen angelehnt sind (vgl. Tabelle 2.1). Es schadet nicht, zumindest für Sonne und Erde die entsprechenden Begriffe zu kennen.

Aus Abb. 2.2 kann man ein paar nützliche Beziehungen ableiten. Nach Pythagoras gilt für die lineare Exzentrizität der Ellipse:

$$e^2 = a^2 - b^2. \tag{2.4}$$

Tabelle 2.1: Bezeichnungen von Periapsis und Apoapsis für verschiedene Himmelskörper

Himmelskörper	Periapsis	Apoapsis	Ursprung
Sonne	Perihel	Aphel	griech. helios[1]
Erde	Perigäum	Apogäum	griech. gaia[2]
Mond	Periselen	Aposelen	griech. selene[3]
Mars	Periares	Apoares	griech. ares[4]
Jupiter	Perijovum	Apojovum	lat. iupiter[5]

1. Helios ist in der griechischen Mythologie der Sonnengott.
2. Gaia steht für die Urmutter in der griechischen und römischen Mythologie.
3. Selene ist die Mondgöttin in der griechischen Mythologie, eine Titanin, und Namensgeberin der chemischen Elements Selen. Sie ist nicht zu verwechseln mit der Mondgöttin Artemis, einer Göttin des Olymps, die als Namensgeberin in der Raumfahrt beliebt ist (siehe Artemis-Programm der NASA).
4. Griechischer Gott des Krieges, des Blutbades und Massakers; einer der zwölf Götter des Olymps.
5. Oberster Gott der römischen Religion.

Halbparameter

Der Halbparameter p einer Ellipse ist definiert als

$$p = \frac{b^2}{a}. \tag{2.5}$$

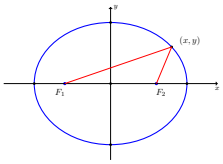
Dieser Ausdruck gilt auch für die anderen Kegelschnitte.

Der Halbparameter p wird u. a. für eine parametrisierte Formulierung der Ellipse (als auch der anderen Kegelschnitte) verwendet, die man als Scheitelgleichung bezeichnet (siehe Abs. 2.4.3). Der Halbparameter entspricht zudem dem Krümmungsradius der Ellipse an den Hauptscheiteln. Man kann recht einfach zeigen, dass der Punkt (e, p) auf der Ellipse liegt, indem man diese Werte in Gl. 2.3 einsetzt.

Eine weitere wichtige Definition der Ellipse beruht auf einer geometrischen Konstruktion über zwei festgelegt Brennpunkte und der großen Halbachse a.

Ellipse

Die Ellipse wird definiert als die Menge aller Punkte (x, y), deren Abstand zu den beiden Brennpunkten $F_{1,2}$ genau $2a$ beträgt.



Die beiden roten Linien in nebenstehender Abbildung haben also in Summe die Länge $2a$. Diese Art der Konstruktion wird als Gärtnerdefinition der Ellipse bezeichnet, da man z. B. mit zwei Pflöcken, einer Schnur der Länge $2a$ und einem Stock eine Ellipse in einem Blumenbeet zeichnen kann. Sei nun der Ortsvektor des linken Brennpunkts $\vec{g}$, der des rechten $\vec{f}$. Ein beliebiger Punkt auf der Ellipse sei mit $\vec{r} = (x, y)$ bezeichnet.

Gärtnerdefinition der Ellipse

Die Gärtnerdefinition der Ellipse besagt, dass

$$|\vec{r} - \vec{f}| + |\vec{r} - \vec{g}| = 2a. \tag{2.6}$$

Dies kann man auch schreiben als

$$\sqrt{(e - x)^2 + y^2} + \sqrt{(-e - x)^2 + y^2} = 2a, \tag{2.7}$$

was durch ein paar kleinere Rechenschritte in die Ellipsengleichung umgeformt werden kann.

2.3 Parabel und Hyperbel

Parabel und Hyperbel kann man aus bahnmechanischer Sicht zur gleichen Klasse von Trajektorien zuordnen. Im Gegensatz zu den geschlossenen Kreis- und Ellipsenbahnen handelt es sich bei Parabel- und Hyperbelbahnen um offene Bahnkurven. Ein Komet, der sich auf einer solchen Bahn befindet und beispielsweise an der Erde vorbeifliegt, wird niemals zurückkommen. Für die Hyperbel gibt es eine über die Halbachsen a und b definierte Gleichung (Normalform), die ähnlich der Ellipsengleichung ist.

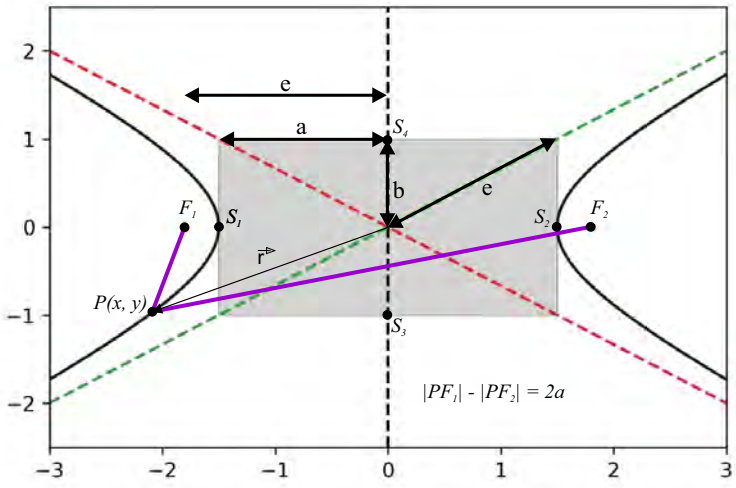


Abbildung 2.3: Geometrische Parameter der Hyperbel deren Definition über eine Gärtnerdefinition. Es gilt $|\overline{PF_1} - \overline{PF_2}| = 2a$

Hyperbelgleichung

Eine Hyperbel in Mittelpunktsform ist definiert als Menge aller Punkte $P(x, y)$ mit

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \tag{2.8}$$

Dabei gilt

$$e^2 = a^2 + b^2, \tag{2.9}$$

mit den Brennpunkten $\vec{f} = (e, 0)$ und $\vec{g} = -\vec{f} = (-e, 0)$. Die geometrischen Parameter der Hyperbel sind in Abb. 2.3 näher erläutert. Man kann auch eine »gärtnerartige« Definitionsgleichung für die Hyperbel angeben:

$$|\vec{f} - \vec{r}| - |\vec{g} - \vec{r}| = 2a. \tag{2.10}$$

Für $x = e$ erhält man aus Gl. 2.8 die Beziehung $y^2 = b^4/a^2$ bzw. $y = \pm b^2/a = p$. Somit entspricht der Halbparameter – wie bei der Ellipse – dem Wert der Ordinate am Brennpunkt $y(e)$. Aus der Gärtner-Definition der Hyperbel kann auf die Normalform (Mittelpunktsgleichung) geschlossen werden (und umgekehrt).

Bei der Parabel verwendet man als Definition Gl. 2.11.

Parabelgleichung

Eine Parabel in Mittelpunktsform ist definiert als Menge aller Punkte $P(x, y)$ mit

$$y^2 = 2px, \tag{2.11}$$

Der (alleinige) Brennpunkt der Parabel liegt bei $\vec{f} = (p/2, 0)$, somit gilt auch hier $y(e) = \pm p$. Zudem kann die Parabel – wie auch die anderen Kegelschnitte – über eine Leitlinie bei $x = -p/2$ definiert werden, was in Abschnitt 2.6.3 genauer besprochen wird. Da die Parabel keinen Mittelpunkt hat, kann man keine entsprechende Mittelpunktsgleichung definieren. Zunächst soll es aber um die Darstellung von Kegelschnitten in verschiedenen Bezugssystemen geben.

2.4 Darstellung von Kegelschnitten

In der Physik spielen eine Reihe von verschiedenen Darstellungen von Kegelschnitten eine wichtige Rolle, die hier vorgestellt werden sollen. Wir gehen hier zunächst von einer Ellipse aus und zeigen dann, wie man auf die anderen Kegelschnitte kommt.

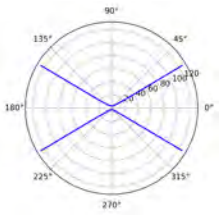
2.4.1 Polarkoordinaten vom Mittelpunkt

Die einfachste Darstellung von Polarkoordinaten geht vom Mittelpunkt der Ellipse aus. Man setzt in Gl. 2.3 $x = r\cos\varphi$ und $y = r\sin\varphi$ ein und erhält die Beziehung

$$r_{\text{Ellipse}} = \frac{ab}{\sqrt{a^2\sin^2\varphi + b^2\cos^2\varphi}}. \tag{2.12}$$

Für den Kreis erhält man mit $r = a = b$ die Identität $r = r$. Die Darstellung der Hyperbel ist durch Vorzeichenwechsel in der Wurzel gegeben:

$$r_{\text{Hyperbel}} = \frac{ab}{\sqrt{a^2\sin^2\varphi - b^2\cos^2\varphi}}. \tag{2.13}$$



Wir führen noch eine neue Größe ein, um Kegelschnitte zu beschreiben, die numerische Exzentrizität ε, die definiert ist als

$$\varepsilon = \frac{e}{a} = \frac{\sqrt{a^2 - b^2}}{a} \tag{2.14}$$

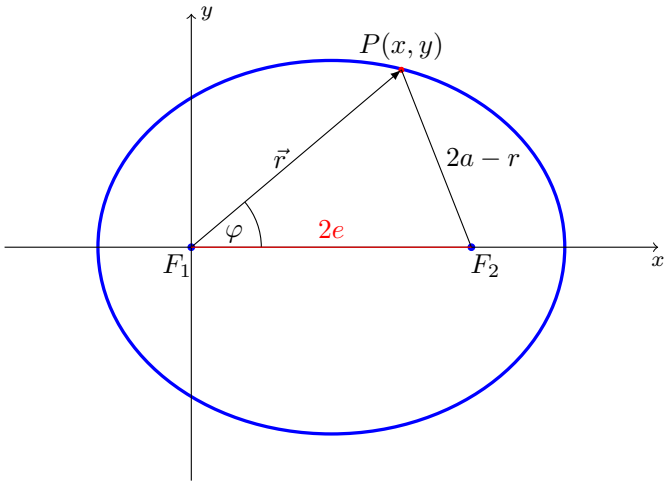


Abbildung 2.4: Zur Herleitung der Ellipsengleichung in Polarkoordinaten auf den Brennpunkt bezogen.

Mit der Identität $\sin^2 + \cos^2 = 1$ kann man dann am Beispiel der Ellipse auch

$$r_{\text{Ellipse}} = \frac{b}{\sqrt{1 - \varepsilon^2 \cos^2 \varphi}}. \tag{2.15}$$

schreiben.

2.4.2 Polarkoordinaten vom Brennpunkt

Zur Beschreibung von Planetenbewegungen und Satellitenbahnen ist es zweckmäßiger, wenn man den Ursprung des Koordinatensystems in den Brennpunkt der Ellipse legt, an dem sich auch der Zentralkörper befindet. Gerade wenn die Masse des Zentralkörpers deutlich größer als die des umlaufenden Körpers ist, ist diese Festlegung besonders sinnvoll. Bei ähnlich schweren Körpern, die sich gegenseitig umlaufen, wäre die Wahl eines Bezugspunktes, der sich in Ruhe befindet, sinnvoller. Hier würde man dann den Schwerpunkt des Systems auswählen. In den meisten Fällen, die im Tutorium besprochen werden, macht man keinen großen Fehler, wenn man davon ausgeht, dass Schwerpunkt und Massezentrum des Zentralkörpers zusammenfallen. Die Herleitung der Ellipsengleichung in brennpunktbezogenen Polarkoordinaten ist eigentlich recht einfach, wird aber in vielen Lehrbüchern unnötig verkompliziert (teilweise werden für die Herleitung sogar quadratische Gleichungen gelöst, was dem Verständnis nicht gerade förderlich ist). Die benötigten geometrischen Zusammenhänge für diese Herleitung sind in Abb. 2.4 dargestellt. Wir gehen von der bekannten Gärtner-Definition

der Ellipse aus, die besagt, dass die Summe der beiden Streckenteile von den Brennpunkten zum jeweiligen Punkt P der Ellipse gerade $2a$ beträgt. Die Verbindungslinie vom Brennpunkt F_1 zum Punkt P wird durch den Vektor $\vec{r}$ beschrieben, der entsprechend den Betrag r haben soll. Somit ergibt sich für die zweite Verbindungslinie eine Strecke von $2a - r$. Der Abstand zwischen den Brennpunkten beträgt $2e$ (rote Linie in der Abbildung). Halten wir das nochmal gesondert fest:

$$\begin{aligned} \overline{F_1P} &= r \\ \overline{F_1P} + \overline{F_2P} &= 2a \\ \overline{F_2P} &= 2a - r. \end{aligned} \tag{2.16}$$

Mit Hilfe des Kosinussatzes (man betrachte hierzu den Vektor $2\vec{e}$ von F_1 nach F_2, bilde die Differenz $2\vec{e} - r$ und dann das Betragsquadrat unter Ausnutzung des Skalarprodukts) kann man nun schreiben:

$$(2a - r)^2 = (2e)^2 + r^2 - 4er\cos\varphi. \tag{2.17}$$

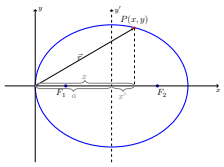
Mit den Definitionen von Halbparameter, linearer und numerischer Exzentrizität erhält man nach wenigen Rechenschritten Gl. 2.18.

Kegelschnitte in Polarkoordinaten

$$r = \frac{p}{1 - \varepsilon\cos\varphi}. \tag{2.18}$$

Diese Formel ist auch gültig für eine Parabel (mit $\varepsilon = 1$) sowie für eine Hyperbel (mit $\varepsilon > 1$). Für $0 < \varepsilon < 1$ erhält man eine Ellipse, für $\varepsilon = 0$ einen Kreis.

2.4.3 Scheitelgleichung



Die letzte wichtige Darstellung der Kegelschnitte, die v. a. in der geometrischen Optik vorkommt, ist die Scheiteldarstellung. Hierzu wird der Nullpunkt in einen der beiden Hauptscheitel gelegt (in unserem Beispiel in den linken), so dass man eine Koordinatentransformation in x-Richtung um die große Halbachse a durchführt, die y-Komponente bleibt von dieser Transformation unberührt. Mit $x = x' + a$ und $y = y'$ eingesetzt in die Ellipsengleichung erhält man die Scheitelgleichung

$$y^2 = 2px + \left(\varepsilon^2 - 1\right) \cdot x^2 \tag{2.19}$$

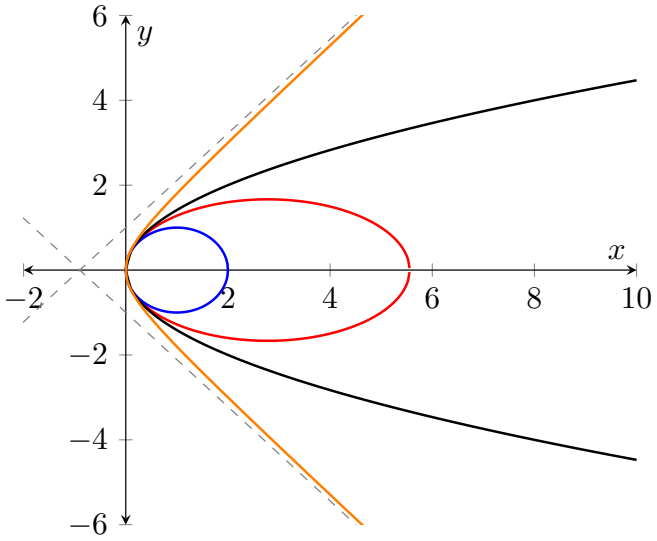


Abbildung 2.5: Scheiteldarstellung der Kegelschnitte Kreis (blau), Ellipse (rot), Parabel (schwarz) und Hyperbel (orange).

der Ellipse für $0 < \varepsilon < 1$. Mit den entsprechenden numerischen Exzentrizitäten für Kreis, Parabel und Ellipse erhält man die korrespondierenden Gleichungen. Als Beispiel sind in Abb. 2.5 für den Halbparameter $p = 4.05$ und verschiedene Exzentrizitäten die entsprechenden Kegelschnitte in Scheiteldarstellung gezeigt.

Die Asymptotengleichung der Hyperbel kann man über eine Grenzwertbetrachtung ermitteln. Leitet man die Scheitelgleichung $y = y(x)$ nach x ab, so erhält man

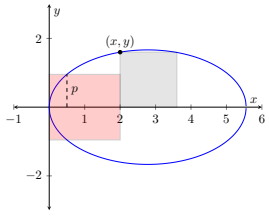
$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{p + \frac{p}{a}x}{\sqrt{px + \frac{p}{a}x^2}} = \frac{p}{\sqrt{px + \frac{p}{a}x^2}} + \frac{\frac{p}{a}x}{\sqrt{px + \frac{p}{a}x^2}}. \tag{2.20}$$

Im Grenzfall $x \to \infty$ verschwindet der erste Term, den zweiten kann man umschreiben als

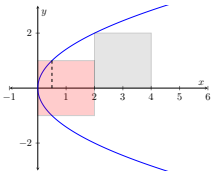
$$\frac{\frac{p}{a}}{\sqrt{\frac{px}{x^2} + \frac{p}{ax^2}x^2}} = \frac{\frac{p}{a}}{\sqrt{\frac{p}{x} + \frac{p}{a}}}. \tag{2.21}$$

Der erste Term unter der Wurzel wird ebenfalls im Grenzfall $x \to \infty$ verschwinden, übrig bleibt

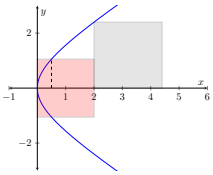
$$\frac{\mathrm{d}y}{\mathrm{d}x}_{x\to\infty} = \frac{b}{a}. \tag{2.22}$$



(a) Ellipse, $p = 1$, $\varepsilon = 0.8$



(b) Parabel, $p = 1$, $\varepsilon = 1.0$



(c) Hyperbel, $p = 1$, $\varepsilon = 1.2$

Abbildung 2.6: Konstruktion von Ordinatenquadrat (grau) und Sperrungsrechteck (rot) für Kegelschnitte in Scheiteldarstellung $y^2 = 2px + (\varepsilon^2 - 1)x^2$.

Die Hyperbel nähert sich also asymptotisch den Geraden

$$y(x) = \pm\frac{b}{a}x \tag{2.23}$$

an.

2.5 Herkunft der Bezeichnungen

Da den Kegelschnitten ein eigenes Kapitel gewidmet ist, soll hier auch noch die Namensgebung enträtselt werden, zumal man gerade im Internet hierzu auch einigen Unsinn findet. Die Benennung wird durch den Vergleich zweier Flächen motiviert, dem Ordinatenquadrat und dem Sperrungsrechteck. Die Konstruktion ist recht einfach, wenngleich vielleicht nicht sofort ersichtlich sein wird, warum man die Flächen überhaupt so wählt. Versuchen wir es zunächst ohne den tieferen Hintergrund nachzuvollziehen. Die Situation ist in Abb. 2.6 für drei Kegelschnitte mit gleichem Halbparameter $p = 1$ dargestellt. Wir wählen einen willkürlichen Punkt (x, y), im Beispiel beträgt die x-Koordinate immer $x = 2$, die y-Koordinate folgt aus der Formel $y^2 = 2px + (\varepsilon^2 - 1)x^2$ für den jeweiligen Kegelschnitt. Die y-Koordinate ist die relevante Größe für das Ordinatenquadrat (graue Fläche), die Strecke $\overline{xy}$ ist eine Seitenlinie dieses Quadrats. Das Sperrungsrechteck erhält man, indem man den Halbparameter als halbe Rechteckhöhe annimmt, dieses Rechteck linksseitig bis zur y-Achse und rechtsseitig bis zum Ordinatenquadrat laufen lässt (rote Fläche).

Die Fläche des Sperrungsrechtecks beträgt $2px$, was man geometrisch sofort in Abb. 2.6 sieht. Die Fläche des Ordinatenquadrats beträgt y^2, denn y ist ja gerade der Ordinatenwert unseren Punktes (x, y). Bilden wir die Differenz aus Ordinatenquadrat und Sperrungsfläche, so erhält man

$$y^2 - 2px = (\varepsilon^2 - 1)x^2. \tag{2.24}$$

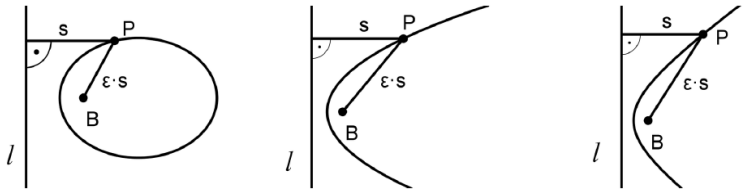


Abbildung 2.7: Leitliniendefinition von Ellipse, Parabel und Hyperbel (aus O. Grund, Geometrie-Kegelschnitte). Die Leitlinie stellt eine Beziehung zwischen der Länge der Verbindungslinie eines beliebigen Punktes $P(x, y)$ des Kegelschnitts zu dessen Brennpunkt und der Länge s einer vertikalen Linie l her, die man als Leitlinie bezeichnet. Dabei gilt, dass die Länge der Strecke zur Leitlinie für jeden Punkt P der Länge der Verbindungslinie zum Brennpunkt multipliziert mit der Exzentrizität des Kegelschnitts entspricht.

Für Ellipsen gilt $0 < \varepsilon < 1$, somit wird die rechte Seite dieser Gleichung kleiner null. Die Fläche des Ordinatenquadrats ist immer geringer als die Fläche des Sperrungsrechtecks. Der Ellipse »mangelt« es also an genug Fläche in diesem Vergleich, das griechische Wort für *ermangeln* heißt *elleipein*. Es mangelt der Ellipse also nicht an einer Kreisform, wie es manchmal behauptet wird. Bei der Parabel ($\varepsilon = 1$) sind Sperr- und Ordinatenfläche immer gleich groß, *paraballein* heißt *gleichkommen*. Im Falle einer Hyperbel wird die Ordinatenfläche immer größer als die Sperrfläche sein, *hyperballein* bedeutet *überschießen*.[2]

2.6 Leitliniendefinition

2.6.1 Allgemeines

Kegelschnitte können auf verschiedene Arten definiert werden. Neben der bereits besprochenen Gärtnerdefinition gibt es eine weitere Abstandsdefinition zu einer Linie, die man als Leitlinie (engl. *directrix*) bezeichnet. Die Situation ist in Abb. 2.7 gezeigt: Die Leitlinie ist mit l bezeichnet, der Brennpunkt des jeweiligen Kegelschnitts mit B. Für einen beliebigen Punkt P auf dem Kegelschnitt kann nun die Verbindungslinie zum Brennpunkt $s = \varepsilon \cdot \overline{PB}$ gezogen werden, wobei ε für die numerische Exzentrizität steht. Ausgehend vom Punkt P wird nun eine gerade Strecke der Länge s so wie in Abb. 2.7 gezogen (parallel zur

2 Der Vollständigkeit halber sei auch noch der Hinweis auf die rhetorischen Figuren und sprachlichen Mittel gegeben. Eine Ellipse ist z. B. in der Linguistik das Auslassen von Bestandteilen eines Satzes (»Wer da?« statt »Wer ist da?«). In der mathematischen Schreibweise werden ebenfalls häufig Ellipsen (meist drei Punkte) verwendet, z. B. in $x_1 + x_2 + \cdots + x_n$.

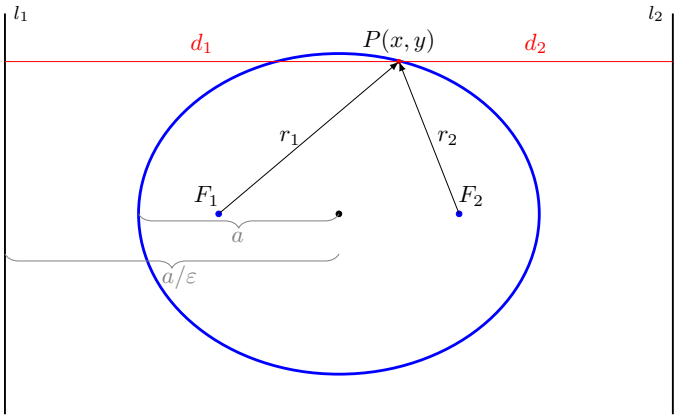
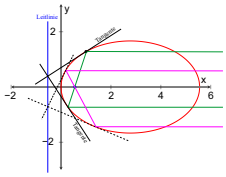


Abbildung 2.8: Definition der Ellipse über Leitlinien. Jedem Brennpunkt F wird dabei eine Leitlinie zugeordnet. Der Abstand beider Leitlinien zueinander beträgt $2a/\varepsilon$.

Verbindungslinie beider Brennpunkte). Es wird nun behauptet, dass für jeden Kegelschnitt die Verbindungsstrecke $\overline{PB}$ für alle Punkte P des Kegelschnitts das $\varepsilon \cdot s$-fache der Strecke s ist.



Im Fall der Parabel ist $\varepsilon = 1$, so dass diese beiden Wegstrecken gleich groß sind (siehe Abs. 2.6.3). Die Leitlinie der Parabel befindet sich zudem immer an der gleichen Position, während bei Ellipse und Hyperbel eine Abhängigkeit von der Exzentrizität gegeben ist. Man kann sich die Leitlinie folgendermaßen über eine Tangenbetrachtung konstruieren, hier am Beispiel einer Ellipse gezeigt: ein einfallender (Licht)Strahl (grün) falle parallel zur x-Achse von rechts kommend auf die Ellipse und werden zum Brennpunkt reflektiert, passiert diesen und wird dann beim nächsten Zusammentreffen mit der Ellipse wirder in Richtung der x-Achse reflektiert. Es ist dabei egal, ob man vom oberen oder unteren Strahl ausgeht, die Situation ist symmetrisch. An beiden Treffpunkten des Strahls mit der Ellipse wird die Tangente eingezeichnet, beide Tangenten treffen sich an einem Punkt, der auf der Leitlinie liegt. Würde man diese Konstruktion für alle Punkte der Ellipse durchführen, ergibt sich genau die Leitlinie. Für einen zweiten Punkt der Ellipse sind die beiden Tangenten (gestrichelte Linien) eingezeichnet.

2.6.2 Leitlinie einer Ellipse

Abbildung 2.8 zeigt eine Ellipse mit großer Hslbachse a, den beiden Brennpunkten F_1 und F_2 sowie den beiden zugehörigen Leitlinien l_1 und l_2. Wir wissen aus der Gärnerdefinition, dass für einen beliebigen Punkt $P(x, y$ die Beziehung $r_1 + r_2 = 2a$ gelten muss. Seien nun die

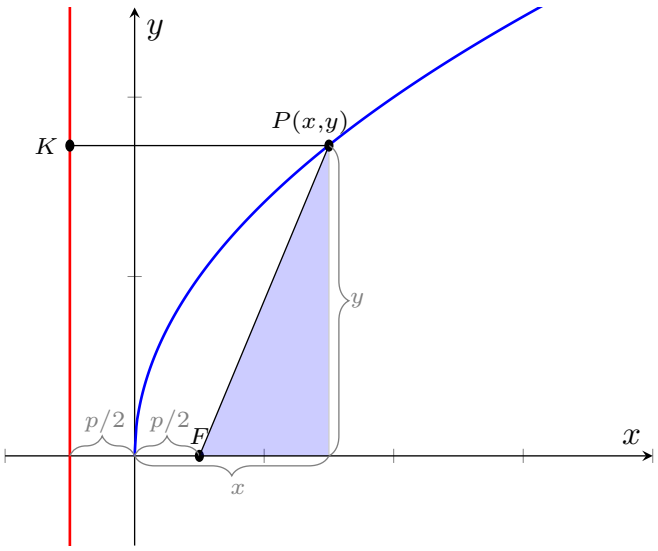


Abbildung 2.9: Definition der Parabel über die Leitlinie bei $x = -0.5p$. Es gilt $\overline{PF} = \overline{PK}$.

Leitlinien in der Entfernung a/ε vom Mittelpunkt der Ellipse vertikal laufend. Der Abstand zwischen den beiden Leitlinien beträgt dann $d_1 + d_2 = 2a/\varepsilon$. Für die Ellipse soll - so die Behauptung - gelten, dass die Strecke $r_1 = \varepsilon d_1$ beträgt (analog gilt natürlich auch $r_2 = \varepsilon d_2$). Es gilt somit

$$\begin{aligned} r_1 + r_2 &= \varepsilon d_1 + \varepsilon d_2 \\ &= \varepsilon(d_1 + d_2) \\ &= \varepsilon\left(\frac{2a}{\varepsilon}\right) \\ &= 2a. \end{aligned} \tag{2.25}$$

Unsere Ellipsendefinition über die Leitlinie ist also identisch mit der Gärtnerdefinition.

2.6.3 Leitlinie einer Parabel

Wir haben für die Ellipse die Gärtnerdefinition eingeführt und gezeigt, dass diese mit ihrer Normalform identisch ist. Eine sehr ähnliche Konstruktionsvorschrift gibt es auch für die Hyperbel (siehe Abb. 2.3). Für die Parabel gibt es keine Gärtner-Vorschrift, die man zeichnerisch umsetzen kann. Formal gibt es hier einen zweiten Brennpunkt im Unendlichen, so dass die Summe der beiden Abstände zwischen einem beliebigen Punkt P der Parabel und den beiden Brennpunkten unend-

lich groß ist, was sich zeichnerisch nicht umsetzen lässt. Umsetzen lässt sich aber die geometrischen Konstruktionsvorschrift über eine Leitlinie.

Eine Parabel ist definiert nach unseren bisherigen Betrachtungen ($\varepsilon = 1$ in Abb. 2.7) als die Menge alle Punkte $P(x, y)$, die von einem festen Punkt, dem Brennpunkt, und einer festen Geraden, der Leitlinie, eine festen Abstand haben (siehe Abb. 2.9).

Dies bedeutet, dass

$$\overline{PF} = \overline{PK} \tag{2.26}$$

die Definitionsgleichung der Parabel $y^2 = 2px$ erfüllen muss. Der Abstand $\overline{PK}$ entspricht gerade $x + p/2$. Die Strecke $\overline{PF}$ kann aus dem bläulichen Dreieck in Abb. 2.9 über den Satz des Pythagoras als

$$\overline{PF} = \sqrt{\left(x - \frac{1}{2}p\right)^2 + y^2} \tag{2.27}$$

ausgedrückt werden. Somit folgt

$$\overline{PF} = \sqrt{\left(x - \frac{1}{2}p\right)^2 + y^2} = x + \frac{1}{2}p. \tag{2.28}$$

Quadriet man beide Seiten und löst nach y aus, so erhält man die Parabelgleichung $y^2 = 2px$. Die Definition der Parabel als Menge von Punkten, die die genannte Abstandsdefinition erfüllen, entspricht der Definition der Parabel über die Scheitelgleichung.

Die Parabel weist bei der Reflexion von Licht und anderen elektromagnetischen Wellen eine einzigartige Eigenschaft auf. Alle einfallenden Strahlen werden bei konkavem Einfall in den Brennpunkt fokussiert, egal bei welcher Höhe relativ zur Grundlinie diese Strahlen einfallen. Das ist nützlich, wenn man beispielsweise geringe Strahlintensitäten messen will und wird bei Parabolantennen ausgenutzt. Allerdings ist die Herstellung parabolischer Optiken technisch herausfordernd und teuer, weswegen alternativ kugelförmige Optiken verwendet werden, die aber nur für sehr achsennahe Strahlen eine ähnlich gutes Fokussierverhalten wie der Parabelspiegel aufweisen. Man spricht dann von der sphärischen Aberration[3], also einem Abbildungsfehler von Linsen.

2.6.4 Leitlinie einer Hyperbel

Abschließend soll es noch um die Leitliniendarstellung der Hyperbel gehen. Da die Hyperbel über zwei Brennpunkte verfügt, gibt es entsprechend zwei Leitlinien, die zwischen den Scheiteln im Abstand $x = \pm a/\varepsilon$

3 erratisch: verwirrt, verstreut

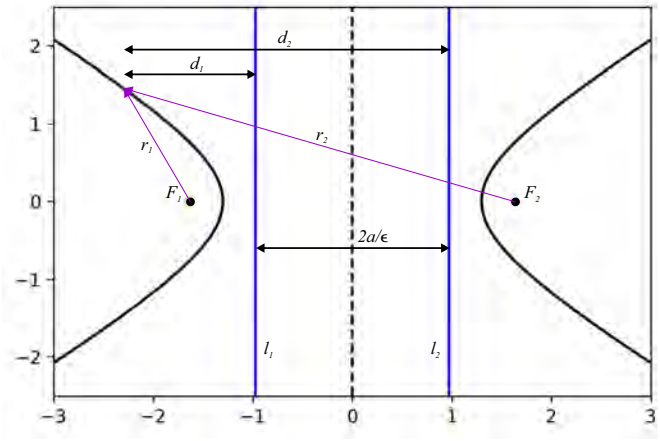


Abbildung 2.10: Definition der Hyperbel über die Leitlinien bei $x = \pm a/\varepsilon$. Es gilt $r_2 - r_1 = 2a$. Diese Beziehung kann zur Positionsbestimmung genutzt werden. Wenn zwei synchronisierte Sender in den Brennpunkten positioniert werden, so ergibt sich für jede konstante Laufzeit Δt eine charakteristische Hyperbelbahn. Man kann also mit Laufzeitmessungen sich den Weg auf einer Hyperbelbahn suchen. Ein weiteres Senderpaar an anderen Brennpunkten erweitert dieses Messprinzip um den Schnittpunkt von Hyperbeln, was eine genaue Positionsbestimmung ermöglicht.

vom Mittelpunkt vertikal stehen (siehe Abb. 2.10). Es gilt für alle Punkte $P(x, y)$ der Hyperbel die Beziehung

$$\frac{r_1}{d_1} = \frac{r_2}{d_2} = \varepsilon. \tag{2.29}$$

Daraus folgt

$$\begin{aligned} r_2 - r_1 &= \varepsilon d_2 - \varepsilon d_1 \\ &= \varepsilon(d_2 - d_1) \\ &= \varepsilon \left(\frac{2a}{\varepsilon} \right) \\ &= 2a, \end{aligned} \tag{2.30}$$

was der Gärtner-Definition der Hyperbel entspricht.

Die besprochenen Eigenschaften der Hyperbel haben wichtige Anwendungen im Alltag. Die Tatsache, dass die Differenz der beiden Wegstrecken r_1 und r_2 auf einer Hyperbel eine Konstante ist, kann zur Positionsbestimmung z. B. von Schiffen verwendet werden. Nehmen wir einmal an, dass es zwei Sender gibt, die miteinander synchronisiert sind und zur gleichen Zeit ein definiertes Signal aussenden, welches eindeutig auf den jeweiligen Sender zurückzuführen ist. Die Signale sollen eine bekannte Phasenbeziehung zueinander haben. Beide Sender sollen nun in den Brennpunkten einer Hyperbel liegen (das kann

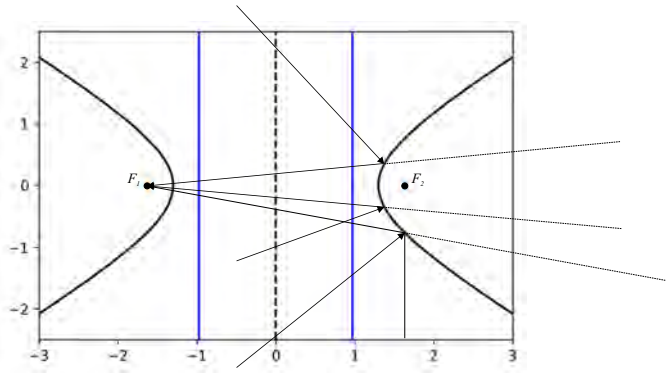


Abbildung 2.11: Reflexion von elektromagnetischen Strahlen an einer Hyperbelfläche.

man bei zwei Punkten natürlich immer annehmen). Da beide Signale sich mit Lichtgeschwindigkeit ausbreiten, wird ein Schiff, wenn es sich auf einer Hyperbelbahn fortbewegt, immer die gleiche Zeitdifferenz Δt messen und kann somit seinen Kurs über eine Differenzmessung der Laufzeit der empfangenen Signale bestimmen. Erweitert man dieses Szenario um weitere Paare von Sendestationen, so kann man Schnitte von Hyperbeln über die verschiedenen Laufzeitmessungen bestimmen, die letztlich eine sehr genaue Positionsbestimmung ermöglichen. Die Genauigkeit der Positionsbestimmung hängt dabei maßgeblich von der zeitlichen Qualität der ausgesendeten Signale ab.

Hyperbolische Spiegel werden in Cassegrain-Teleskopen zur Brennweitenverlängerung (größeres ε führt zu größerer Brennweite) verwendet, wodurch man kompaktere Bauformen erreichen kann. Eine Weiterentwicklung stellt das Ritchey-Chrétien-Cassegrain-Teleskop dar (zwei hyperbolische Spiegel anstatt der Kombination parabolisch-hyperbolisch), das beim *Hubbly Space Telescope* zum Einsatz kommt. Auch das *Very Large Telescope* in Chile basiert auf zwei hyperbolischen Spiegeln. Die Reflexion von Lichtstrahlen an einer hyberbolischen Fläche ist exemplarisch in Abb. 2.11 gezeigt.

3 NEWTONSCHE MECHANIK

3.1 Der Massenpunkt

Die Mechanik ist ein Teilgebiet der Physik, das sich mit der Bewegung und den Verformungen von Objekten sowie den dabei wirkenden Kräften beschäftigt. Die Größe dieser Objekte kann stark variieren – von winzigen, punktförmigen Massen bis hin zu großen Strukturen wie Galaxien. Die Newtonsche Mechanik, benannt nach Isaac Newton, stellt die erste umfassende mathematische Beschreibung mechanischer Vorgänge dar. Sie wurde in Newtons Werk *Philosophiae Naturalis Principia Mathematica* von 1687 formuliert. Newton konzentrierte sich in seiner Theorie auf idealisierte Körper, sogenannte Punktmassen oder Massenpunkte. Dieser Teil der Mechanik wird daher auch als Punktmechanik bezeichnet. Verformungen von Körpern spielen in der klassischen Newtonschen Mechanik keine Rolle und werden in diesem Skript nicht behandelt.

Massenpunkt

Bei einem Massenpunkt handelt es sich um ein idealisiertes, punktförmiges Objekt, dessen einzige Eigenschaft seine Masse ist.

Das »Punkt« in Massenpunkt deutet an, dass dieser Massenpunkt sich irgendwo im Raum befindet. Die Lage im Raum wird durch seine Punktkoordinaten ausgedrückt.

Das Konzept des Massenpunkts wird immer dann verwendet, wenn die räumliche Ausdehnung eines Objekts für den betrachteten physikalischen Sachverhalt keine Rolle spielt. Ein Beispiel dafür ist die Beschreibung der Planetenbewegung um die Sonne, bei der die Planeten als Massenpunkte betrachtet werden. Die Größe der Planeten ist im Vergleich zu den Abständen zwischen ihnen vernachlässigbar klein. Massenpunkte besitzen keine inneren Freiheitsgrade, das bedeutet, sie können weder um sich selbst rotieren noch in irgendeiner Weise angeregt werden. Volumenabhängige Kräfte, wie etwa der Auftrieb, wirken ebenfalls nicht auf Massenpunkte. Sie können daher als die grundlegenden

Bausteine der Mechanik betrachtet werden. In der Elektrodynamik wird das Konzept des Massenpunkts erweitert: Hier definiert man ein punktförmiges Objekt, das eine elektrische Ladung trägt. Da Ladung stets an eine Masse gekoppelt ist, besitzt dieses Elementarteilchen auch eine Masse. Ein solches Teilchen, das sowohl Ladung als auch Masse trägt, wird als Punktladung bezeichnet.

Wir beschäftigen uns in diesem Skript - solange es nicht explizit gesagt wird - ausschließlich mit Massenpunkten[1] bzw. im späteren Verlauf dann mit Punktladungen.

3.2 Bezugs- und Koordinatensysteme

Zur Beschreibung eines mechanischen Systems benötigt man ein Modell für Raum und Zeit, in dem sich die Massenpunkte bewegen. Newton postulierte einen absoluten Raum, relativ zu dem alle Bewegungen physikalischer Objekte stattfinden. In der klassischen (newtonschen) Mechanik erfolgt die Beschreibung in einem dreidimensionalen euklidischen Punktraum ε^3, der auch als Ortsraum bezeichnet wird. Die Position eines Massenpunkts im Ortsraum wird durch drei Ortskoordinaten bestimmt. Häufig werden hierfür die kartesischen Koordinaten $\vec{r} = (x, y, z)$ verwendet, es können jedoch auch andere Koordinatensysteme gewählt werden, wie beispielsweise Polar-, Kugel- oder Zylinderkoordinaten, die wir später kennenlernen werden. Die Koordinaten eines Massenpunkts können sich mit der Zeit ändern, sodass er zu einem früheren oder späteren Zeitpunkt eine andere Position einnehmen kann.

Man könnte sich nun fragen, ob es eine maximale Anzahl von Koordinaten gibt, mit denen sich ein Massenpunkt vollständig beschreiben lässt. Aus dem Alltag wissen wir, dass ein Körper eine Geschwindigkeit haben muss, um sich im Raum zu bewegen. Geschwindigkeit ist formal nichts anderes als die Veränderung der Position eines Körpers innerhalb eines bestimmten Zeitintervalls. Wenn man also den Ort und die Änderung dieses Orts pro Zeiteinheit (die Geschwindigkeit) eines Körpers zu einem bestimmten Zeitpunkt t kennt, kann man seinen zukünftigen (oder vergangenen) Ort berechnen. Ein Massenpunkt ist daher vollständig beschrieben, wenn seine Position (x, y, z) und seine Geschwindigkeitskomponenten (v_x, v_y, v_z) zu einem beliebigen Zeitpunkt t bekannt sind. Mit diesen sechs Angaben lässt sich seine Bewegung für alle Zeiten bestimmen.

1 Eine kurze Anmerkung zur Schreibweise: auch wenn es sich nur um eine Masse handelt, hängt man an die Masse noch ein „n“ an, schreibt also Massenpunkt statt Massepunkt, so wie eben bei allen zweisilbigen Wörtern dieser Bauart, beispielsweise bei Sonnensystem statt Sonnesystem.

Es gibt jedoch eine entscheidende Erweiterung: Was passiert, wenn sich die Geschwindigkeit $\vec{v} = (v_x, v_y, v_z)$ mit der Zeit t ändert? Für eine vollständige Beschreibung eines Massenpunkts reicht es dann nicht mehr aus, nur den Ort und die Geschwindigkeit anzugeben. Man muss zusätzlich die Größe kennen, die diese Geschwindigkeitsänderung verursacht. In der Newtonschen Mechanik ist diese Größe die Kraft. Kräfte bewirken eine Änderung der Geschwindigkeit, allerdings nur unter der Annahme, dass die Masse des Körpers konstant bleibt, daher musste eine allgemeinere Beschreibung gefunden werden: In der klassischen Mechanik wurde erkannt, dass Kräfte den Impuls $\vec{p}$ eines Körpers verändern. Der Impuls ist das Produkt aus Masse und Geschwindigkeit: $\vec{p} = m\vec{v}$. Um die Bewegung eines Massenpunkts vollständig zu beschreiben, verwendet man daher den Impuls anstelle der Geschwindigkeit. Der Zustand eines Massenpunkts zu einem bestimmten Zeitpunkt wird somit durch seine Ortskoordinaten (x, y, z) und seine Impulskomponenten (p_x, p_y, p_z) beschrieben. Diese sechs Größen spannen den sogenannten Phasenraum auf, einen sechsdimensionalen Raum, in dem der Zustand des Massenpunkts vollständig erfasst wird.[2]

Jedem Objekt wird zu jedem Zeitpunkt t ein Punkt $P \in \varepsilon^3$ im Ortsraum zugeordnet. Diese Zuordnung erfolgt mithilfe sogenannter Bezugssysteme. Ein Bezugssystem legt fest, von welchem Standpunkt aus ein System beobachtet wird und wie die Zeitmessung innerhalb dieses Systems definiert ist. Die Formulierung *von wo aus* bedeutet dabei nicht zwingend, dass ein fixer Bezugspunkt gewählt werden muss. Beispielsweise kann man die Bewegung der benachbarten Planeten im Bezugssystem der um die Sonne rotierenden Erde beschreiben. Ebenso könnte man einen festen Bezugspunkt wählen, der immer an derselben Stelle im Raum bleibt. Allerdings ist es problematisch, festzustellen, ob dieser Punkt tatsächlich ortsfest ist – denn dies erfordert wiederum ein Referenzsystem. Dieses Problem erinnert an das klassische Henne-Ei-Dilemma und führt häufig zu Verwirrung, besonders zu Beginn des Studiums. Newtons Konzept des absoluten Raums postuliert zwar, dass es feste Punkte im Raum geben muss, die unverändert bleiben, liefert jedoch keine praktische Methode zur Bestimmung eines solchen absoluten Punktes.

2 man spricht in der theoretischen Physik von verallgemeinerten Koordinaten und verallgemeinerten Geschwindigkeiten, was bedeuten soll, dass zur Beschreibung eines Objekts im Phasenraum auch andere physikalische Größen zur Beschreibung der Bewegung herangezogen werden können, beispielsweise die Auslenkung eines Fadenpendels um den Winkel φ statt seiner Auslenkung in kartesischen Koordinaten. Wichtig ist hierbei, dass die gewählten Koordinaten voneinander unabhängig sind, also eine Koordinate nicht als Linearkombination von anderen beschreibenden Koordinaten dargestellt werden kann.

Bezugssysteme sind somit keine absolut greifbaren Größen, sondern Annahmen, die man zur Beschreibung eines physikalischen Problems trifft. Sie ermöglichen es, einem physikalischen Objekt im Raum drei Ortskoordinaten und eine Zeitkoordinate zuzuordnen und dadurch dessen Bewegung zu beschreiben. Ein Bezugssystem ist eine physikalische Konstruktion, die festlegt, wie und von wo aus Bewegungen und physikalische Prozesse beobachtet und gemessen werden. Es ist an einen Beobachter gebunden (im verallgemeinerten Sinn) und definiert nicht nur die Position eines Objekts im Raum, sondern auch, wie die Zeit gemessen wird. Ein Bezugssystem umfasst also sowohl räumliche als auch zeitliche Komponenten.

Ein Koordinatensystem hingegen ist ein mathematisches Werkzeug, das verwendet wird, um Positionen in einem Raum zu beschreiben, indem es einem Punkt eindeutige Koordinaten zuweist (z.B. x, y, z in einem kartesischen System). Es ist Teil eines Bezugssystems, aber ein Bezugssystem geht über die bloße Angabe von Koordinaten hinaus.

Bezugssystem

Ein Bezugssystem legt fest, in welchem Rahmen ein physikalisches Problem beschrieben wird. Es ist eine Annäherung an die Realität. Verwechseln Sie nicht Bezugssystem und Koordinatensystem. Letztere kommen ins Spiel, wenn man ein Bezugssystem festgelegt hat.

Der angesprochene Zirkelschluss löst sich auf, wenn man zulässt, dass ein frei gewähltes Bezugssystem die Referenz für alle anderen darauf bezogenen Bezugssysteme darstellt. Beispiel: ein ruhender Beobachter an einem Bahnhof sieht einen anfahrenden (also für ihn beschleunigten) Zug, in dem ein (leerer) Kinderwagen steht - idealisiert mit reibungsfreien Rädern. Für den Beobachter auf dem Bahnhof sieht es so aus, als ob der Kinderwagen an der gleichen Stelle bleibt – bis er natürlich irgendwann am Ende des Zuges an die Wand knallt, aber darum soll es nicht gehen. Für ihn wirkt (wenn man mal einen unendlich langen Zug annimmt) die ganze Zeit keine Kraft. Machen wir die Annahme (das ist hier der springende Punkt), dass das Bezugssystem des Beobachters ein Inertialsystem sein soll, dann handelt es sich beim Zug um ein beschleunigtes Bezugssystem. Für einen Fahrgast im Zug sieht es wiederum so aus, als ob der Kinderwagen durch eine Kraft beschleunigt wird, da er sich für ihn beschleunigt nach hinten (entgegen der Fahrtrichtung) bewegt. Diese Kraft ist allerdings nur eine Folge der Beschleunigung seines Bezugssystems, folglich keine physikalische, sondern eine Schein-

kraft. Scheinkräfte zeichnen sich dadurch aus, dass es zu ihnen keine Gegenkraft gibt, sie sind also keine Kräfte im Sinne des dritten Newtonschen Axioms (genaueres hierzu findet man in Abschnitt 3.4). Wenn man ein Bezugssystem mit dem Kinderwagen verbindet, so sieht man den Kinderwagen wiederum die ganze Zeit in Ruhe - in diesem Fall befindet man sich wieder in einem Inertialsystem. Für den Kinderwagen sieht es so aus, als ob der Zug sich beschleunigt bewegt.

Was passiert nun, wenn der Kinderwagen durch irgendetwas am Rollen gehindert wird, beispielsweise durch eine Reibungskraft? Antwort: er wird mit dem Zug beschleunigt. Man könnte auch sagen, dass er nun Teil des Zuges ist, genauso wie alle anderen fest verbunden Teile wie die Sitze usw. Das mit dem Kinderwagen beschleunigte Bezugssystem ist nun mit dem Bezugssystem des Fahrgasts verbunden - und bis auf die absolute Festlegung, wo der Ursprung eines Koordinatensystems liegen soll, auch gleich. In diesem beschleunigten Bezugssystem sieht es nun so aus, als ob die Gesamtkraft auf den Kinderwagen Null ergibt. Solange der Kinderwagen im beschleunigten Zug in Ruhe bleibt, wird die Trägheitskraft F_{T} des Kinderwagens durch seine (Haft)Reibungskraft F_{HR} kompensiert. Reicht diese Reibungskraft nicht mehr aus, dann setzt sich der Kinderwagen (mit Masse m) mit der Beschleunigung $\vec{a} = (\vec{F_{\mathrm{T}}} - \vec{F_{\mathrm{HR}}})/m$ in Bewegung.

Wie kann nun aber eine physikalische Kraft eine Scheinkraft kompensieren? Die Antwort ist einfach: sie tut es überhaupt nicht. Die Reibungskraft ist eine physikalische Kraft und demzufolge gibt es eine gleichgroße Gegenkraft. Der Kinderwagen drückt auf den Boden und im Gegenzug drückt der Boden des Zuges auf den Kinderwagen. Dies passiert auf mikroskopischer Ebene so, dass der Boden, der eine Rauigkeit aufweist, eine Gegenkraft auf den Kinderwagen ausübt, die man als Haftreibungskraft bezeichnet. Sie ist eine Reaktion der Kraft des Kinderwagens auf den Boden. Diese Andruckkraft des Kinderwagens auf den Boden ist eine Folge der Trägheit des Kinderwagens, sich der Beschleunigung des Zuges zu widersetzen. Falls die Trägheitskraft größer als die Reibungskraft werden sollte, dann bewegt sich der Kinderwagen mit einer Beschleunigung nach hinten, die proportional zur Kraftdifferenz von Trägheitskraft und Reibungskraft ist.

Wir sind jetzt die ganze Zeit davon ausgegangen, dass unser Beobachter am Bahnhof in Ruhe und daher sich in einem Inertialsystem befindet. Wir wissen aber auch, dass die Erde sich um eine Achse dreht. Als Folge kann das Inertialsystem des Beobachters nur in guter Näherung als Inertialsystem angesehen werden. Wir haben aufbauend auf dieser Näherung weitere Bezugssysteme definiert. Ein willkürlich gewähltes Bezugssystem wurde als Inertialsystem angenommen (was es

ja in guter Näherung auch ist) und darauf aufbauend wurden andere Bezugssysteme definiert. Dies soll nochmal ein Hinweis auf den relativen Charakter der Festlegung von Bezugssystemen sein.

Es gibt erstmal keine allgemeine Regel, ob ein ruhendes Bezugssystem besser ist als ein bewegtes, das hängt letztlich vom mechanischen Problem ab, das man beschreiben will. Es ist aber oft so, dass eine geschickte Wahl des Bezugssystems den Rechenaufwand erleichtert. Wir werden später im Detail sehen, wie sich die Wahl des Bezugssystems auf die Beschreibung des mechanischen Problems auswirkt. Dabei wird eine spezielle Art von Bezugssystem, das so genannte Inertialsystem, eine wichtige Rolle spielen.

Wie sieht es nun mit der Zeit aus: hier geht man in der Newtonschen Mechanik von einer absoluten Zeit aus. Unter »absolut« versteht man die Annahme, dass die Zeit in allen Bezugssystemen gleich verläuft. Eine Uhr in einem bewegten Bezugssystem wird genauso schnell laufen wie in einem ruhenden Bezugssystem. In der speziellen Relativitätstheorie hängt die verstreichende Zeit vom Bezugssystem ab, d. h. einmal synchronisierte Uhren müssen danach nicht für alle Zeiten synchron laufen.

Nachdem ein Bezugssystem festgelegt wurde, wird im nächsten Schritt ein Koordinatensystem eingeführt. Koordinatensysteme und Bezugssysteme sind nicht dasselbe und keine Synonyme. So kann beispielsweise ein kartesisches Koordinatensystem sowohl in einem ruhenden als auch in einem bewegten Bezugssystem verwendet werden. Ohne die Angabe eines Bezugssystems ist die Einführung eines Koordinatensystems jedoch sinnlos. Der Ursprung des Koordinatensystems wird in der Regel im Bezugspunkt platziert, allerdings ist dies keine zwingende Voraussetzung.

Der erste Schritt zur Beschreibung eines physikalischen Systems besteht in der Festlegung eines Bezugssystems, der zweite in der Wahl eines geeigneten Koordinatensystems. Ein Koordinatensystem ist dann geeignet, wenn es die mechanische Bewegung präzise beschreibt und idealerweise vereinfacht. Häufig nutzt man dabei die Symmetrien des Problems. Bei Rotationssymmetrie bieten sich zum Beispiel Kugelkoordinaten an. In der Mechanik (und auch in anderen Disziplinen) wird man im Wesentlichen mit vier Arten von Koordinatensystemen arbeiten: kartesischen Koordinaten, ebenen Polarkoordinaten, Kugelkoordinaten und Zylinderkoordinaten. Betrachtet man ebene Polarkoordinaten als Spezialfall der Kugelkoordinaten, reduziert sich diese Anzahl auf drei Systeme. Bei der Wahl von Bezugssystem und Koordinatensystem gibt es im Grunde drei Ansätze: einen durchdachten, einen weniger optimalen und einen absolut ungeeigneten. Letzteren sollte man selbstverständlich vermeiden.

An dieser Stelle der wichtige Hinweis: Alle vier Koordinatensysteme sollten im Laufe des Studiums – und zwar möglichst früh – sicher beherrscht werden. Sie werden regelmäßig verwendet, und ohne ein solides Verständnis kommt man irgendwann nicht mehr weiter. Der positive Aspekt ist, dass es bei diesen vier Systemen bleibt. In Spezialvorlesungen oder Abschlussarbeiten können gelegentlich andere Koordinatensysteme, wie elliptische, vorkommen, aber das bleibt die Ausnahme.

3.3 Der Impuls

Wie misst man den Bewegungszustand eines Körpers? Diese Frage mag auf den ersten Blick trivial erscheinen, fand aber erst durch Newton eine schlüssige Antwort. Stellen Sie sich einen Zug vor, der langsam auf Sie zurollt. Trotz der geringen Geschwindigkeit wäre es schwer, ihn nur durch körperliche Anstrengung zum Stillstand zu bringen. Es scheint also eine Eigenschaft zu geben, die den Zug daran hindert, seinen Bewegungszustand zu ändern. Welche Eigenschaft könnte das sein?

Der Zug besitzt sowohl Masse als auch Geschwindigkeit. Die Erfahrung zeigt, dass es schwieriger ist, einen Körper mit großer Masse zu bremsen als einen leichteren Körper bei gleicher Geschwindigkeit. Mit der Masse ist also eine Eigenschaft des Körpers verbunden, die man als Trägheit bezeichnet. Dennoch reicht die Geschwindigkeit allein nicht aus, um den Bewegungszustand vollständig zu beschreiben. Ebenso lehrt die Erfahrung, dass leichte Körper mit hoher Geschwindigkeit schwerer abzubremsen sind als Körper derselben Masse mit niedriger Geschwindigkeit. Weder Masse noch Geschwindigkeit allein genügen also, um den Bewegungszustand eines Körpers vollständig zu erfassen.

Newton führt auf Basis solcher Überlegungen den Impuls als charakteristische Größe der Bewegung ein. In seinem Hauptwerk heißt es: ≫quantitas motus est mensura ejusdem orta ex velocitate et quantitate materiae conjunctim≪.[3]

Definition: Impuls $\vec{p}$

Der Impuls $\vec{p}$ eines Objekts der Masse m und der Geschwindigkeit $\vec{v}$ ist definiert als das Produkt

$$\vec{p} = m \cdot \vec{v} \tag{3.1}$$

3 Die Bewegungsmenge ist ein Maß, das sich aus der Geschwindigkeit und der Menge der Materie zusammen ergibt.

Die Definition des Impulses nach Gleichung 3.1 gehört zu den fundamentalsten Definitionen in der Physik. Sie bildet die Basis für Newtons Formulierung der klassischen Mechanik und muss im Kopf eines Studierenden der Physik immer abrufbar sein.

Wichtig ist, sich klarzumachen, dass der Impuls eine vom Bezugssystem abhängige Größe ist.[4] Ein Auto der Masse m, das mit konstanter Geschwindigkeit $\vec{v}$ an mir vorbeifährt, hat in meinem Bezugssystem (in dem ich als ruhend angenommen werde) den Impuls $m\vec{v}$. Im Bezugssystem, das mit dem Auto verbunden ist, beträgt der Impuls hingegen Null. Es geht beim Impuls – wie bei vielen anderen physikalischen Größen – daher oft weniger um den absoluten Wert, sondern vielmehr um die relativen Unterschiede im jeweils festgelegten Bezugssystem.

Auch wenn dies an anderer Stelle ausführlicher behandelt wird, soll hier folgendes angemerkt werden: Aus der Alltagserfahrung wissen wir, dass eine Waage ein hohes Gewicht anzeigt, wenn ein Zug darauf gestellt wird. Daraus lässt sich ableiten, dass dieses Gewicht im Schwerefeld der Erde und die damit verbundene Masse etwas mit der Trägheit zu tun haben könnten. Aber Vorsicht: Die Waage zeigt das Gewicht an, also die Schwerkraft, die durch die Erdanziehung auf den Zug wirkt. Diese Masse kann man als *schwere Masse* bezeichnen, die im Gravitationsgesetz vorkommt und als Parameter für die gegenseitige Anziehung dient.

Die Annahme, dass die Trägheit eines Körpers durch diese schwere Masse beschrieben werden kann, ist jedoch nicht selbstverständlich. Es könnte sein, dass nur ein Teil der trägen Masse zur Gravitationskraft beiträgt. Obwohl Newton und vor ihm Galilei durch ihre Fallexperimente erkannten, dass schwere und träge Masse identisch zu sein schienen, betrachteten sie dies als Zufall. Erst Einstein erkannte in dieser Übereinstimmung mehr als einen Zufall und formulierte daraus sein Äquivalenzprinzip, das letztlich zur Allgemeinen Relativitätstheorie führte (die in diesem Skript jedoch keine weitere Rolle spielen wird).

3.4 Die Newtonschen Axiome

Die Newtonschen Axiome bilden das Fundament der Mechanik und sind – wie der Name schon sagt – Axiome. Ich betone hier bewusst »leider«, da Axiome eigentlich Aussagen sein sollten, die unmittelbar einleuchten und keiner weiteren Erklärung bedürfen. Die Erfahrung mit

4 Das gilt eigentlich für alle physikalischen Größen, weshalb der Festlegung eines Bezugssystems so viel Bedeutung beigemessen wird. Leider wird dieser Aspekt in der physikalischen Grundausbildung nicht immer ausreichend vermittelt.

Studierenden in den Anfangssemestern zeigt jedoch, dass dies oft nicht der Fall ist. Deshalb werde ich diese Axiome hier ausführlich erläutern. Es gibt insgesamt drei eigenständige Axiome sowie ein zugehöriges Corollarium, das in vielen Lehrbüchern oft übergangen wird. In diesem Skript werde ich es als viertes Axiom behandeln. Beginnen wir mit dem ersten Axiom.

Lex Prima

Jeder Körper verharrt in seinem Zustand der Ruhe oder gleichförmigen Bewegung, wenn er nicht durch einwirkende Kräfte gezwungen wird, seinen Bewegungszustand zu ändern.

Im *Lex Prima* fällt zunächst auf, dass Newton keinen Unterschied zwischen dem Zustand der Ruhe und der gleichförmigen Bewegung macht. Dies ist konsequent und sinnvoll, da der Bewegungszustand eines Körpers vom gewählten Bezugssystem abhängt: In einem Bezugssystem ist der Körper in Bewegung, im anderen in Ruhe. Newton postuliert zudem das Beharrungsvermögen eines Körpers in seinem aktuellen Bewegungszustand, was als Trägheit bezeichnet wird. Trägheit ist ein Konzept, das mit der Masse des Körpers verknüpft ist und aus der Alltagserfahrung stammt: Den Bewegungszustand eines schweren Körpers zu ändern, erfordert mehr Aufwand als bei einem leichten.

Da der Bewegungszustand jedoch nicht allein von der Masse abhängt – da Masse keine Bewegungsgröße ist – führt Newton den Begriff des Impulses ein. Er behauptet, dass eine Änderung des Bewegungszustands, also des Impulses, nur dann erfolgt, wenn eine Kraft auf den Körper einwirkt. Die Kraft ist für Newton also die Größe, die den Impuls eines Körpers verändert. Daraus lässt sich folgern: In Abwesenheit von Kräften bleibt der Impuls eines Körpers erhalten. Für einen einzelnen Körper führt dies direkt zum Impulserhaltungssatz.

Impulserhaltungssatz für einen Körper

Bei einem System aus genau einem Massenpunkt, auf den keine äußere Kraft wirkt, ist der Impuls des Massenpunkts eine Erhaltungsgröße.

Später werden wir diesen Ansatz auf Systeme mit mehreren Massenpunkten erweitern und dabei genauer auf den hier eingeführten Begriff

der »äußeren« Kraft eingehen. An dieser Stelle kann man sich jedoch bereits merken, dass in einem System, das nur aus einem einzelnen Massenpunkt besteht, jede auf ihn wirkende Kraft eine äußere Kraft sein muss. Da kein weiterer Körper im System vorhanden ist, gibt es keine Möglichkeit für eine innere Wechselwirkung.

Lex Secunda

Die Änderung der Bewegung ist der Einwirkung der bewegenden Kraft proportional und geschieht nach der Richtung derjenigen geraden Linie, nach welcher jene Kraft wirkt.

Mit der Änderung der Bewegung ist natürlich die zeitliche Änderung des Impulses gemeint. Diese Impulsänderung ist nicht nur proportional zur einwirkenden Kraft, sie ist die Kraft. Newton definiert Kraft als die zeitliche Änderung des Impulses eines Körpers. Die »Richtung derjenigen geraden Linie« bezieht sich darauf, dass Kräfte als Vektoren beschrieben werden und von einem Ursprungspunkt aus auf den Körper einwirken. Dies lässt sich in der bekannten Definitionsgleichung der Kraft zusammenfassen.

Definition: Kraft $\vec{F}$

$$\vec{F} = \frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = \dot{\vec{p}} \tag{3.2}$$

Die Kraft entspricht der zeitlichen Änderung des Impulses. Konkreter: Die Kraft entspricht der Tangente der Impuls-Zeit-Funktion, also der zeitlichen Ableitung des Impulses nach der Zeit. Gleichung 3.2 sollte Ihnen in Fleisch und Blut übergehen. Verwenden Sie als Definitionsgleichung der Kraft stets diesen Ausdruck und nicht die vereinfachte Schulbuchformel $\vec{F} = m\vec{a}$, da diese insbesondere bei Raketenantrieben, die Schub erzeugen, nicht anwendbar ist, um den Schub zu beschreiben.

Leiten wir die Schubgleichung her, um zu verstehen, was damit gemeint ist. Schub ist nichts anderes als eine Kraft, die eine Rakete beschleunigt. Raketenantriebe erzeugen Schub, indem sie (meist) kontinuierlich eine Masse an Treibstoff mit konstanter Ausströmgeschwindigkeit $\vec{c}$ ausstoßen. Wenn der Schub $\vec{T}$ als Kraft verstanden werden soll, muss er sich in der Definitionsgleichung $\vec{F} = \dot{\vec{p}}$ wiederfinden. Leiten wir also

den Impuls des ausgestoßenen Treibstoffs $m\vec{c}$ nach der Zeit ab und definieren diesen Ausdruck als Schub $\vec{T}$:

$$\vec{T} = \dot{\vec{p}} = \frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(m \cdot \vec{c}) = m \underbrace{\frac{\mathrm{d}\vec{c}}{\mathrm{d}t}}_{=0} + \frac{\mathrm{d}m}{\mathrm{d}t}\vec{v} = \dot{m}\vec{c} \tag{3.3}$$

Die Rakete erfährt durch diese Schubkraft eine Beschleunigung $\vec{a}$ in die entgegengesetzte Richtung. Zu einer gegebenen Zeit t, bei der die Rakete die Masse m haben soll, können wir das Kraftgesetz, also die Wirkung des Schubs auf die Rakete folgendermaßen schreiben:

$$\vec{F} = m\vec{a} = m\frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = -\frac{\mathrm{d}m}{\mathrm{d}t}\vec{c} \tag{3.4}$$

Man will als Raketenbauer in der Regel wissen, welche Geschwindigkeit v die Rakete zu einer bestimmten Zeit t hat und wie diese Geschwindigkeit mit der ausgestoßenen Masse zusammenhängt. Bei einer chemischen Rakete kann man nicht davon ausgehen, dass die Massenänderung der Rakete vernachlässigbar ist, so dass man hier die Massenänderung pro Masse $\mathrm{d}m/m$ berücksichtigen muss.[5] Wir schreiben Gleichung 3.3 daher in der Form

$$m\,\mathrm{d}\vec{v} = -\,\mathrm{d}m \cdot \vec{c} \tag{3.5}$$

und teilen noch durch die Masse m:

$$\mathrm{d}\vec{v} = -\frac{\mathrm{d}m}{m} \cdot \vec{c}. \tag{3.6}$$

Diesen Ausdruck kann bei nun integrieren:

$$\int_{v_0}^{v_1} \mathrm{d}\vec{v} = -\vec{c}\int_{m_0}^{m_1} \frac{\mathrm{d}m}{m}. \tag{3.7}$$

Als Ergebnis erhält man die Geschwindigkeitsänderung $\Delta\vec{v}$ als Funktion der ausgestoßenen Masse, die so genannte Raketengrundgleichung bzw. Ziolkowski-Gleichung.

Ziolkowski-Gleichung

$$\Delta\vec{v} = \vec{v_1} - \vec{v_0} = \vec{c} \cdot \ln\left(\frac{m_0}{m_1}\right) \tag{3.8}$$

5 Bei einer Ionenrakete mag diese Näherung vielleicht gerechtfertigt sein, da hier deutlich weniger Treibstoff ausgestoßen wird.

Die Raketengleichung wird im weiteren Verlauf noch näher analysiert werden. Wir haben bei der Herleitung bereits das dritte Newtonsche Axiom angewendet, dass in seiner Lehrbuchform folgendermaßen aussieht:

Lex Tertia

Die Wirkung ist stets der Gegenwirkung gleich, oder die Wirkungen zweier Körper aufeinander sind stets gleich und von entgegengesetzter Richtung.

Dieses Axiom wird häufig in der Kurzform »actio = reactio« zusammengefasst und besagt, dass Kräfte, die physikalischer Natur sind, immer paarweise auftreten.

Wir wollen dieses Axiom etwas mehr präzisieren, um die Sache etwas klarer zu machen. Newton bezieht sich in seiner Formulierung nur auf eine Wechselwirkung, die Gravitation. Sein drittes Axiom kann aber auf alle vier Wechselwirkungen der Physik erweitert werden. Es bietet sich an, folgende Formulierung zu verwenden:

Lex Tertia Version 2

Übt Körper A per Wechselwirkung B die Kraft $\vec{F}$ auf Körper C aus, so übt Körper C per gleicher Wechselwirkung die entgegengesetzt gerichtete Kraft $-\vec{F}$ auf Körper A aus. Kraft und Gegenkraft greifen also niemals am gleichen Körper an.

Die Betonung liegt hier vor allem auf der gleichen Wechselwirkung. In Abbildung 3.1 soll verdeutlicht werden, was dies bedeutet. Man sieht hier einen Klotz, der auf einer Unterlage im Schwerefeld der Erde liegt. Im Schwerefeld der Erde wirkt auf den Klotz natürlich die Gewichtskraft $\vec{F}_\mathrm{G}$, die man o. B. d. A. im Schwerpunkt angreifen lässt. Zusätzlich muss auf den Klotz eine entgegengesetzt gleich große Nettokraft wirken, damit er in Ruhelage bleibt. Diese Nettokraft wird unter idealisierten Annahmen vollständig durch die Abstoßungskraft des Tisches aufgebracht, die man als Normalkraft $\vec{F}_\mathrm{N}$ bezeichnet.

Es kann aber auch noch andere Kräfte geben, die entgegen der Gewichtskraft wirken, beispielsweise die Auftriebskraft, wenn der Klotz einer Atmosphäre ausgesetzt ist und auf einem Tisch liegt. Hier liegen

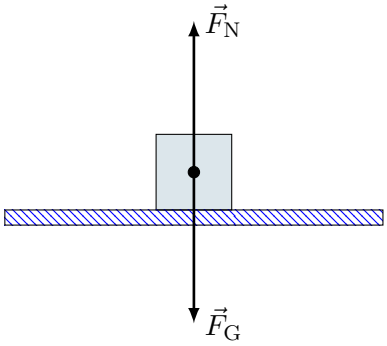


Abbildung 3.1: Der Zusammenhang zwischen Normalkraft und Gewichtskraft ist anders als man meist denkt. In Lehrbüchern wird häufig eine solche Darstellung verwendet, wenn die Kräfte dargestellt werden sollen, die auf einen Körper wirken, der auf einer Unterlage liegt. Studierende interpretieren dies häufig so, als wäre die Normalkraft, also die abstoßende Kraft der Unterlage auf den Körper die Gegenkraft zur Gravitationskraft, die nach unten wirkt. Dem ist aber nicht so. Die Gegenkraft ist die Anziehungskraft, die der Klotz auf die Erde ausübt. Diese greift im Erdmittelpunkt an.

also bereits zwei Kräfte vor, die entgegen der Gewichtskraft wirken und den Körper in der Waage halten. Nach dem dritten Newtonschen Axiom kann also keine der beiden Kräfte (Normalkraft oder Auftriebskraft) die Gegenkraft zur Gravitationskraft auf den Klotz sein, denn es handelt sich um völlig andere Wechselwirkungen.

Die Normalkraft wird z. B. von der elektromagnetischen Wechselwirkung aufgebracht. Da die Gravitationskraft (Wechselwirkung B) auf den Klotz (Körper A) von der Erde (Körper C) ausgeübt wird, muss die Gegenkraft entsprechend von Körper A auf Körper C ausgeübt werden. Der Klotz zieht also mit der betragsgleichen Kraft F_G die Erde an. Unter Vernachlässigung des Luftdrucks und der Rotation der Erde drückt der Klotz dann natürlich mit einer Kraft (actio) auf den Tisch, die betragsmäßig der Gewichtskraft entspricht. Der Tisch stößt entsprechend den Klotz mit einer gleichgroßen und entgegengesetzt gerichteten Kraft (reactio) ab, eben mit der Normalkraft $\vec{F}_\mathrm{N}$. Normalkräfte sind also Coulombkräfte und haben ihre alleinige Ursache in der elektromagnetischen Wechselwirkung. Gleiches gilt übrigens auch für Seilkräfte.

Auf Grundlage dieser drei Axiome lässt sich ein Bezugssystem definieren, das die Grundlage der Newtonschen Mechanik darstellt, das Inertialsystem.

Definition: Inertialsystem

Unter einem Inertialsystem versteht man ein Bezugssystem, in dem die drei Newtonschen Axiome gelten.

Diese Definition impliziert, dass es Bezugssysteme geben muss, in denen die Newtonschen Axiome nicht gelten. Wie man sich in diesem Fall helfen kann, wird im Abschnitt über beschleunigte Bezugssysteme diskutiert. Für das Alltagsgeschäft kann man sich ein Inertialsystem als ruhendes oder mit konstanter Geschwindigkeit bewegtes Bezugssystem vorstellen. Nehmen wir an, dass Σ ein ruhendes System und damit ein Inertialsystem ist. Das System Σ' sei ein sich in beliebige Richtung gleichförmig bewegtes Bezugssystem, das zur Zeit $t = 0$ identisch mit Σ sei. Die Geschwindigkeit von Σ' relativ zu Σ sei $\vec{v}$. Ein Massenpunkt m sei in Σ zur Zeit t am Ort $\vec{r}$. Im System Σ' wird sich dieser Massenpunkt am Ort

$$\vec{r}\,' = \vec{r} - \vec{v} \cdot t \tag{3.9}$$

befinden. Wirkt nun ausgehend vom System Σ auf diesen Massenpunkt eine Kraft $\vec{F}$, dann ist das System Σ' genau dann auch ein Inertialsystem, wenn in diesem System die gleiche Kraft beobachtet wird. Es gilt also in Σ:

$$\vec{F} = m \cdot \ddot{r} = m\frac{\mathrm{d}^2\vec{r}}{\mathrm{d}t^2}. \tag{3.10}$$

Im System Σ' gilt wiederum:

$$\vec{F}\,' = m \cdot \ddot{r}\,' = m\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}\vec{r}\,'}{\mathrm{d}t} + \vec{v}\right) = m\frac{\mathrm{d}^2\vec{r}}{\mathrm{d}t^2} = \vec{F}. \tag{3.11}$$

Bezugssysteme, die sich relativ zueinander mit konstanter Geschwindigkeit bewegen, sind also Inertialsysteme. Bewegt sich das System Σ' mit Beschleunigung $\vec{a}$ relativ zu Σ, so sieht ein Beobachter in Σ' eine Kraft $\vec{F}\,' = \vec{F} + m\vec{a}$. Wir werden diese Situation am Beispiel der Atwoodschen Fallmaschine im Detail diskutieren. Am Beispiel der mit einem beschleunigten Körper verbundenen Bezugssysteme werden wir zudem diskutieren, wie man die Newtonschen Axiome »retten« kann, wie man also trotz des zusätzlichen Kraftterms mit den Axiomen rechnen kann. Der Trick besteht darin, dass man so genannte Scheinkräfte einführt, die diesen zusätzlichen Kraftterm kompensieren. Scheinkräfte sind Kräfte, die keine Gegenkraft haben, also ihre alleinige Ursache in der Beschleunigung des Koordinatensystems haben.

Man sollte immer im Hinterkopf behalten, dass viele Bezugssysteme in unserem Alltag keine Inertialsysteme sind, aber innerhalb gewisser Näherungen durch ein Inertialsystem beschrieben werden können. Ein mit der Erdoberfläche verbundenes Bezugssystem ist beispielsweise kein Inertialsystem, da die Erde rotiert. Dennoch kann man mechanische Bewegungen in der Nähe der Erdoberfläche näherungsweise so beschreiben, als wäre dies ein Inertialsystem. Ein Inertialsystem ist mit einer Annahme verknüpft, die man an das System stellt, das man beschreiben will.

Beenden wir diesen Abschnitt mit dem vierten Newtonschen Axiom.

Lex Quarta

Zwei Kräfte, die am gleichen Massenpunkt angreifen, setzen sich zur Diagonalen des von ihnen gebildeten Parallelogramms zusammen.

Dieses Corollarium besagt, dass sich Kräfte wie Vektoren addieren. Durch das so entstehende Kräfteparallelogramm wird axiomatisch auch das Superpositionsprinzip eingeführt, d. h. jede Kraftwirkung kann unabhängig von anderen Kräften betrachtet werden.

3.5 Trägheitskräfte

An dieser Stelle soll nochmal auf den Begriff der Trägheit näher eingegangen werden. Nehmen wir an, dass von einem beliebigen Inertialsystem aus betrachtet auf einen Körper der Masse m und Geschwindigkeit $\vec{v}$ eine Kraft $\vec{F}_{\text{wirk}}$ wirkt. Diese Kraft bewirkt eine Änderung des Impulses $\vec{p}$ dieses Körpers gemäß des zweiten Newtonschen Axioms:

$$\frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = \vec{F}_{\text{wirk}}. \tag{3.12}$$

Unter der weiteren Annahme, dass die Masse des Körpers konstant sei, folgt daraus der Ausdruck

$$m \cdot \vec{a} = \vec{F}_{\text{wirk}}. \tag{3.13}$$

Wir erweitern unsere Überlegung und lassen nun mehrere Kräfte auf den Körper wirken. Diese Kräfte kann man nach der Lex Quarta vektoriell zu einer Gesamtkraft addieren und erhält somit

$$m \cdot \vec{a} = \sum_i \vec{F}_{\text{wirk},i}. \tag{3.14}$$

In Prüfungen bekommen wir häufig auf die Frage »Was versteht man unter einer Kraft?« die Antwort »Masse mal Beschleunigung« zu hören. In Gleichung 3.14 steht aber, dass dieses Produkt aus Masse und Beschleunigung sich aus der Summe aller auf den Körper einwirkenden Kräfte ergibt. Die wirkenden Kräfte können unterschiedlicher Natur sein, beispielsweise die Gravitationskraft zwischen Massen oder die Coulomb-Kraft zwischen Ladungen. Eine Kraft setzt erstmal eine Wechselwirkung zwischen Körpern voraus und einer Folge aus dieser Wechselwirkung. Die Folge ist, dass sich der Bewegungszustand des Körpers, auf den eine Kraft wirkt, ändert. Wir haben gelernt, dass Körper, die eine Masse haben, träge sind. Sie behalten ihren Bewegungszustand bei, bis eine Kraft auf sie wirkt. Kräfte bewirken eine Änderung des durch diese Trägheit charakterisierten Bewegungszustand. Man bezeichnet daher die resultierende Kraft, die auf einen Körper wirkt, als Trägheitskraft. Es gibt also Kräfte, die auf einen Körper wirken und die ihre Ursache in einer der vier bekannten fundamentalen Wechselwirkungen haben[6] und einen Einfluss dieser Kräfte auf den Bewegungszustand eines Körpers. Letzteres ist das, was man aus der Schule als $\vec{F} = m\vec{a}$ kennt und was allgemeiner als $\vec{F} = \mathrm{d}\vec{p}/\,\mathrm{d}t$ formuliert wird. Die Trägheit eines Körpers hängt von einer Eigenschaft ab, die man als Masse bezeichnet, genauer als träge Masse m_t. In der Impulsdefinition steht also genauer gesagt nicht einfach nur eine Masse, sondern die träge Masse eines Körpers.

Trägheitskräfte sind nun Kräfte, die proportional zur trägen Masse eines Körpers sind (wie es z. B. in $\vec{F} = m\vec{a}$ unschwer zu erkennen ist). Es gibt unter den vier fundamentalen physikalischen Wechselwirkungen nur eine Kraft, die ebenfalls proportional zur Masse eines Körpers ist: die Gravitationskraft. Die Masse, die hier vorkommt, wird als schwere Masse m_s bezeichnet. Es ist alles andere als selbstverständlich, dass diese beiden Massen identisch sein müssen. Die träge Masse ist ein Maß dafür, wie sehr sich ein Körper einer Änderung seines Bewegungszustands widersetzt. Die schwere Masse gibt lediglich an, wie schwer ein Körper ist, also welche Gewichtskraft auf ihn wirkt. Es hat sich jedoch in allen bisherigen Messungen herausgestellt, dass $m_\mathrm{t} = m_\mathrm{s}$. Die Gleichheit von träger und schwerer Masse wird als Äquivalenzprinzip der Physik bezeichnet. In diesem Fall gilt für den freien Fall in Nähe der Erdoberfläche

$$m_\mathrm{t} \cdot a = m_\mathrm{s} \cdot g \tag{3.15}$$

und somit $a = g$.

Es gibt aufgrund des Äquivalenzprinzips die Hypothese, dass es sich bei der Gravitationskraft um eine Scheinkraft handeln könnte.

6 gravitative, elektromagnetische, schwache und starke Wechselwirkung

Um das zu illustrieren, wird gerne der Beobachter im beschleunigten geschlossenen Aufzug herangezogen. Dieser Beobachter kann formal nicht unterscheiden, ob er mit seinem Aufzug auf der Erdoberfläche steht und dort einer nach unten wirkenden Beschleunigung $\vec{g}$ ausgesetzt ist, oder ob er sich im leeren Raum befindet (also in einem Gebiet, in dem keine große Masse in der Nähe ist) und der Aufzug die ganze Zeit mit der Beschleunigung $-\vec{g}$ nach oben beschleunigt wird.

3.6 Gravitation

Nachdem nun der Kraftbegriff im 2. Newtonschen Axiom definiert wurde, wollen wir uns mit der Kraft beschäftigen, die zentral für dieses Skript ist, der Gravitationskraft $\vec{F}_\mathrm{G}$ zwischen zwei Massen m_1 und m_2.

Definition: Gravitationskraft $\vec{F}_\mathrm{G}$

$$\vec{F}_\mathrm{G} = -\frac{\gamma m_1 m_2}{|\vec{r}_2 - \vec{r}_1|^2} \cdot \frac{\vec{r}_2 - \vec{r}_1}{|\vec{r}_2 - \vec{r}_1|} \tag{3.16}$$

Die Proportionalitätskonstante γ wird als Gravitationskonstante bezeichnet. In der so geschriebenen Form stellt $\vec{F}_\mathrm{G}$ die Kraft dar, die von Masse m_1 auf Masse m_2 ausgeübt wird, da $\vec{r}_2 - \vec{r}_1$ der Verbindungsvektor von Masse 1 zu Masse 2 ist. Das negative Vorzeichen soll andeuten, dass die Kraft anziehend ist. Diese Darstellung ist die allgemeinste Schreibweise der Gravitationskraft. Wenn man das Bezugssystem in eine der beiden Massen gelegt hätte, dann hätte man sich in ein beschleunigtes Bezugssystem gesetzt und kann die Newtonschen Axiome nicht benutzen. In vielen Fällen und insbesondere bei der Bewegung von Planeten um die Sonne bzw. von Satelliten um einen Planeten kann man in guter Näherung davon ausgehen, dass die Masse des einen Körpers deutlich größer als die des anderen ist. Der großen Masse gibt man sinnvollerweise einfach den Buchstaben M, der kleinen Masse den Buchstaben m. Es gilt also $m \ll M$.

In dieser Näherung ist das mit der großen Masse verbundene Bezugssystem also ein Inertialsystem. Bei genauerem Hinsehen ist es so, dass auch der größere Körper eine Bewegung durchführt, also nicht fest im Raum steht. Bei Orbitalbewegungen ist es so, dass beide Massen um den gemeinsamen Schwerpunkt kreisen. Wenn man also ein richtiges Inertialsystem des Zwei-Körper-Problems verwenden will, dann muss man sich in diesen Schwerpunkt setzen. Oder man wählt einen beliebigen

fest Punkt im Raum als Bezugssystem, dann werden die Gleichungen nur etwas unhandlicher. Im Fall der Näherung $M \gg m$ gilt

$$\vec{F}_{\mathrm{G}} = -\frac{\gamma m M}{r^2}\vec{e}_r, \tag{3.17}$$

was etwas einfacher von der Form her ist. Wir werden hauptsächlich mit dieser Formulierung arbeiten, die uns schließlich zum Kepler-Problem führt und die uns im weiteren Verlauf beschäftigen wird.

Wir haben das mit dem Schwerpunkt verbundene Bezugssystem des Zweikörper-Problems als »richtiges Inertialsystem« bezeichnet. Wie sieht das nun in der realen Welt aus, die ja aus vielen Massen besteht? Betrachtet man das Sonnensystem unter der Annahme, dass keine äußeren Kräfte auf die Massen wirken, so ist der Schwerpunkt des Gesamtsystems ein Inertialsystem. Nun ist es aber so, dass das Sonnensystem Teil eines größeren Gesamtsystems, der Milchstraße, ist. In der Milchstraße sollen sich über 100 Milliarden weitere Sterne befinden, die alle über die Gravitation miteinander wechselwirken - wenn auch nur schwach. Zudem kreisen alle diese Sterne - also auch unser Sonnensystem - um ein galaktisches Zentrum. Der Schwerpunkt des Sonnensystems befindet sich also auf einer rotierenden Bahn und ist somit bei näherer Betrachtung also doch kein Inertialsystem. Man könnte geneigt sein, dass das Baryzentrum unserer Galaxie ein Inertialsystem sein könnte, aber es gibt eine Vielzahl weitere Galaxien, die alle relativ zueinander in Bewegung sind und natürlich über die Gravitation miteinander verbunden sind. Ob es also überhaupt ein Inertialsystem gibt, ist demnach fraglich.[7] Es lässt sich also festhalten, dass ein Inertialsystem ein idealisiertes Bezugssystem ist. Wir machen die Annahme, dass ein Bezugssystem ein Inertialsystem ist und rechnen mit den Newtonschen Axiomen. Ob es sich dabei wirklich um ein Inertialsystem handelt, ist gar nicht der springende Punkt. Es reicht, wenn unsere Annahme in guter Näherung richtig ist, wenn wir unser physikalisches Problem durch die Annahme eines Inertialsystems hinreichend gut beschrieben wird. Aber zurück zum Kepler-Problem:

7 Wenn es aber mindestens ein Inertialsystem geben sollte, dann gibt es entsprechend des 1. Newtonschen Axioms auch unendlich viele weitere.

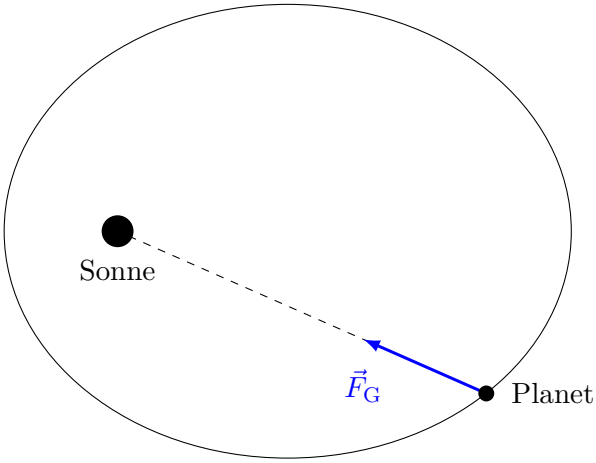


Abbildung 3.2: Das Kepler-Problem am Beispiel des elliptischen Umlaufs eines Planeten um die in einem der Brennpunkte der Ellipse ruhende Sonne. Eingezeichnet ist die auf den Planeten wirkende Gravitationskraft der Sonne.

Das Keplerproblem

Als Keplerproblem bezeichnet man ein spezielles Zwei-Körper-Problem, bei dem zwei Körper nur über die Gravitation miteinander wechselwirken und der eine Körper deutlich schwerer als der andere ist.

Das Kepler-Problem ist in Abbildung 3.2 am Beispiel eines elliptischen Orbits eines Planeten um die Sonne dargestellt. Kepler hat als erster die Bewegung der Planeten auf Ellipsenbahnen um die im Brennpunkt der Sonne ruhende Sonne proklamiert. Ellipsen sind also Bahnformen, die sich als Lösung des Kepler-Problems ergeben. Sie sind allerdings nicht die einzig möglichen Bahnformen, sondern Teil eines größeren Ensembles von Bahnformen, die man allgemein als Kegelschnitte bezeichnet. Zu den Kegelschnitten gehören Kreise, Ellipsen, Parabeln und Hyperbeln. Wir werden an späterer Stelle sehen, dass die Art der Bahnkurve mit der mechanischen Gesamtenergie des sich bewegenden Objekts zusammenhängt. Zunächst sollen aber die drei Keplerschen Gesetze vorgestellt werden, die allesamt aus empirischen Beobachten von Kepler Anfang des 17. Jahrhunderts gewonnen wurden.

1. Keplersches Gesetz

Die Planeten bewegen sich auf Ellipsen, in deren einem Brennpunkt die Sonne steht.

Im anderen Brennpunkt der Ellipse befindet sich demzufolge nichts. Das erste Keplersche Gesetz war lange Zeit einzuordnen als eine empirische Beobachtung, die vielleicht einen kleinen Kreis von Astronomen interessierte, aber allgemein eher als unwichtig eingestuft wurde - bis Newton etwa 100 Jahre später auf Grundlage von Keplers Beobachtungen sein Gravitationsgesetz (also Gleichung 3.16) beweisen konnte. Sie sind also ein schönes Beispiel dafür, dass Wissenschaft nicht sinnvolles Wissen produzieren muss, wie das heute gerne im Wissenschaftsbetrieb gesehen wird, sondern Wissen ganz allgemein und völlig zweckfrei schaffen sollte. Ob gewonne Erkenntnisse einen Nutzen haben, kann sich erst viele Jahre später entscheiden.

Newton konnte anhand der Keplerschen Beobachtungen zeigen, dass die Ellipsenbahnen exakte Lösungen des nach ihm benannten Zwei-Körper-Problems sind, wenn die wirkende Kraft eine $1/r^2$-Abhängigkeit aufweist. Bevor wir diesen Zusammenhang im Detail diskutieren, führen wir das nächste Kepler-Gesetz ein.

2. Keplersches Gesetz

Der Radiusvektor von der Sonne zum Planeten überstreicht in gleichen Zeiten gleiche Flächen.

Dieses Gesetz, dass auch als Flächensatz bekannt ist, ist eine empirische Beobachtung, die mit der Erhaltung einer noch nicht eingeführten Größe, dem Drehimpuls $\vec{L}$, zusammenhängt. Wir holen dies hier nach:

Definition: Drehimpuls $\vec{L}$

$$\vec{L} = \vec{r} \times \vec{p} \tag{3.18}$$

Der Drehimpuls $\vec{L}$ ist das Kreuzprodukt aus Ortsvektor $\vec{r}$ und Impuls $\vec{p}$ eines Körpers. Für die Angabe eines Ortsvektors benötigen wir ein

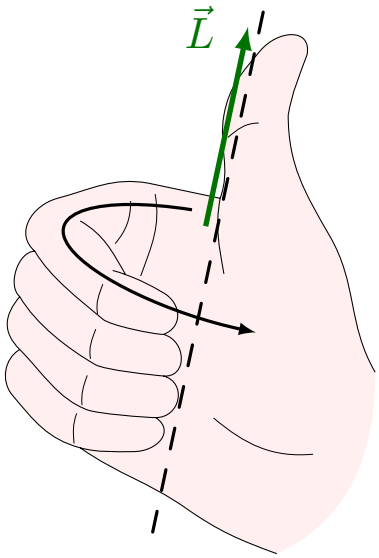


Abbildung 3.3: Die Lage des Drehimpulsvektors bestimmt man mit der rechten Hand. Hierbei folgen alle Finger (außer dem Daumen) dem Drehsinn der Bewegung, der abgespreizte Daumen dann die Lage von $\vec{L}$.

Bezugssystem. Es ist nahe liegend, dass man beim Kepler-Problem das mit dem schweren Körper verbundene Bezugssystem wählt. Die Drehachse geht dann entsprechend der Rotationsrichtung senkrecht aus der Ebene, in der die Bewegung stattfindet (in unserem Fall die Papierebene) hinein (bei Bewegung im Uhrzeigersinn) oder hinaus (bei Bewegung entgegen des Uhrzeigersinns). Dieses Vorgehen ist in Abbildung 3.3 veranschaulicht. Die Physik würde sich allerdings nicht ändern, wenn man einen anderen Bezugspunkt für den Ortsvektor wählt, nur der mathematische Aufwand erhöht sich.

Mit der Definition des Drehimpulses kann man nun eine wichtige Aussage bewiesen werden, nämlich, dass der Drehimpuls beim Keplerproblem eine Erhaltungsgröße ist. Dazu müssen wir zuerst eine Definition dieses Begriffs einführen. Wir werden dann zeigen, dass die Erhaltung des Drehimpulses identisch mit Keplers zweitem Gesetz ist.

Definition: Erhaltungsgröße

Eine Größe $\vec{A}$ wird als Erhaltungsgröße bezeichnet, wenn sie sich über die Zeit nicht ändert, wenn also gilt:

$$\frac{\mathrm{d}\vec{A}}{\mathrm{d}t} = \dot{\vec{A}} = 0 \tag{3.19}$$

Wenn der Drehimpuls sich über die Zeit nicht ändert, muss seine Zeitableitung verschwinden. Man könnte geneigt sein zu argumentieren, dass sich ja bei einer orbitalen Bewegung, z. B. auf einer Ellipse, die ganze Zeit über sowohl der Ortsvektor als auch der Impuls ändert, somit auch der Drehimpuls nicht konstant sein kann. Dies ist allerdings eine Fehlannahme, da ja das Kreuzprodukt zu betrachten ist und das durchaus eine Konstante sein könnte. Leiten wir also $\vec{L}$ nach der Zeit ab. Da es sich um ein Produkt handelt, muss die Produktregel der Differentialrechnung verwendet werden. Zudem muss man beachten, dass die Reihenfolge beim Kreuzprodukt nicht getauscht wird, da $\vec{a} \times \vec{b} = -\vec{b} \times \vec{a}$:

$$\frac{\mathrm{d}\vec{L}}{\mathrm{d}t} = \dot{\vec{L}} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\vec{r} \times \vec{p}\right) = \dot{\vec{r}} \times \vec{p} + \vec{r} \times \dot{\vec{p}} \tag{3.20}$$

Wir haben also einmal den Ortsvektor nach der Zeit abgeleitet und den Impuls konstant gehalten, danach den Ortsvektor konstant gehalten und den Impuls nach der Zeit abgeleitet. Da wir die zeitliche Änderung dieser beiden Größen noch gar nicht kennen, schreiben wir also einfach nur einen Punkt über die Größe, um anzudeuten, dass dies die Zeitableitung der Größe sein soll. Betrachten wir den ersten Term, so können wir ausnutzen, dass die zeitliche Ableitung des Ortsvektors nicht anderes als die Geschwindigkeit des Körpers ist, d. h.

$$\dot{\vec{r}} = \vec{v} = \frac{\mathrm{d}\vec{r}}{\mathrm{d}t}. \tag{3.21}$$

Der Impuls $\vec{p}$ ist im Prinzip nichts anderes als der gleiche Geschwindigkeitsvektor, der mit der Masse m des Körpers skaliert wird. Somit kann man für den ersten Term schreiben: $\dot{\vec{r}} \times m\vec{v} = m(\vec{v} \times \vec{v}) = 0$. Das Kreuzprodukt eines Vektors mit sich selbst ist immer Null. Für den zweiten Term verwenden wir die Definitionsgleichung der Kraft entsprechend des zweiten Newtonschen Axioms: $\dot{\vec{p}} = \vec{F}$. Daraus folgt eine weitere wichtige physikalische Größe:

Definition: Drehmoment $\vec{M}$

$$\vec{M} = \dot{\vec{L}} = \vec{r} \times \vec{F} \tag{3.22}$$

Das Drehmoment ist also die physikalische Größe, die Drehimpulse ändert, so wie es Kräfte bei Impulsen tun. Machen wir uns mit Hilfe von Abbildung 3.2 klar, wie Ortsvektor $\vec{r}$ und Kraft $\vec{F}$ zueinander stehen. Die Gewichtskraft ist dort bereits eingezeichnet. Der Ortsvektor würde von der Sonne zum Ort des Planeten zeigen, somit auf der gestrichelten Linie liegen. Die beiden Vektoren sind also kollinear zueinander, somit gilt

$$\frac{\mathrm{d}\vec{L}}{\mathrm{d}t} = 0 \tag{3.23}$$

bei Zentralkräften. Die Erhaltung des Drehimpulses gehört zu den wichtigsten Eigenschaften von Orbitalbewegungen. Ohne diese Erhaltungsgröße wäre Raumfahrt, wie wir sie heute kennen und betreiben, nicht möglich.

Was hat nun der Drehimpuls $\vec{L}$ mit dem zweiten Keplergesetz zu tun? Hierzu muss man sich die mathematische Bedeutung des Kreuzprodukts in Erinnerung rufen. Das Kreuzprodukt $\vec{a} \times \vec{b}$ zweier Vektoren $\vec{a}$ und $\vec{b}$ ergibt einen neuen Vektor $\vec{c}$, der senkrecht zu den beiden anderen Vektor steht und mit diesen ein Rechtssystem bildet (im Sinne der Rechte-Hand-Regel). Die Länge dieses neuen Vektors entspricht der Fläche des Parallelogramms, das von den beiden Vektoren $\vec{a}$ und $\vec{b}$ aufgespannt wird. Hier besteht also der Zusammenhang zum Keplerschen Flächengesetz. In Abbildung 3.4 ist dieser Zusammenhang visualisiert. Das Vektorprodukt $\frac{1}{2}\vec{r} \times \mathrm{d}\vec{r}$ entspricht der grün hinterlegten Fläche, also der halben Fläche des von den beiden Vektoren aufgespannten Parallelogramms. Die Gesamtfläche des Parallelogramms entspricht der Summe aus der grüner und hellblauer Teilfläche. Wir bezeichnen die halbe Parallelogrammfläche als $\mathrm{d}\vec{A}$ und teilen den zugehörigen Ausdruck durch das Zeitintervall $\mathrm{d}t$:

$$\frac{\mathrm{d}\vec{A}}{\mathrm{d}t} = \frac{1}{2}\,\vec{r} \times \frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \frac{1}{2}\vec{r} \times \vec{v} = \frac{1}{2m}\vec{L}. \tag{3.24}$$

Da es sich bei $\mathrm{d}t$ formal um eine skalare Größe handelt, können wir sie an beliebiger Stelle einsetzen, sinnvollerweise aber beim $\mathrm{d}\vec{r}$, da wir dann wieder die Geschwindigkeit $\vec{v}$ erhalten. Im letzten Schritt haben wir die

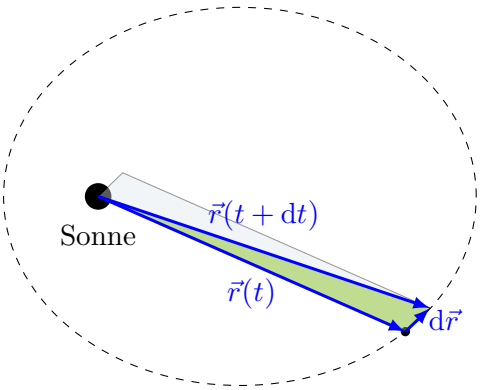


Abbildung 3.4: Zum geometrischen Zusammenhang zwischen Kreuzprodukt und vom Ortsvektor $\vec{r}$ überstrichener Fläche. In einem Zeitintervall bewegt sich ein Körper auf der Ellipse um $\mathrm{d}\vec{r}$ weiter. Die von $\vec{r}$ dabei überstrichene grüne Fläche (wir machen die Annahme, dass sich $\vec{r}$ bei kleinen Zeitintervallen nicht ändert) entspricht dem Kreuzprodukt $0.5 \cdot \vec{r} \times \mathrm{d}\vec{r}$.

Definitionsgleichung des Drehimpulses verwendet. Da der Drehimpuls bei Zentralkräften eine Erhaltungsgröße ist, folgt:

$$\frac{\mathrm{d}\vec{A}}{\mathrm{d}t} = \text{const.} \tag{3.25}$$

Das zweite Kepler-Gesetz ist also nicht anderes als die empirische Formulierung des Drehimpulserhaltungssatzes. Dass wir die Fläche $\vec{A}$ als Vektor definiert haben, bedeutet formal nur, dass die Fläche eine Orientierung hat, was aber für die eigentliche Aussage keine Relevanz hat.

Kommen wir nun zum dritten der Keplerschen Gesetze.

3. Keplersches Gesetz

Die Quadrate der Umlaufzeiten T_i der Planeten (i) verhalten sich zueinander wie die Kuben der großen Halbachsen a_i.

Mathematisch lässt sich das so ausdrücken, dass für einen beliebigen Planeten mit Umlaufzeit T und großer Halbachse a gilt:

$$\frac{T^2}{a^3} = \text{const.} \tag{3.26}$$

Wir zeigen dies am Beispiel einer kreisförmigen Bewegung eines Planeten um die Sonne. Die Beschränkung auf Kreisbahnen ist der mathematischen Einfachheit geschuldet, natürlich kann man dies auch für

elliptische Orbits zeigen. Um dies zu zeigen, müssen wir zuerst noch einen Exkurs über Kreisbahnen und der dort - wenn man aus einem Inertialsystem heraus argumentiert - vorliegenden Zentripetalkraft.

3.6.1 Zentripetalkraft

Die Zentripetalkraft ist ein Konzept, welches Studierenden erfahrungsgemäß große Verständnisschwierigkeiten bereitet. Dies scheint an der Fehlvorstellung zu liegen, dass diese Kraft häufig als eigenständige physikalische Kraft angesehen wird und nicht als (resultierende) Trägheitskraft einer auf einen Körper wirkenden physikalischen (Gesamt-)Kraft. Wir wollen daher an dieser Stelle versuchen, dieser Größe etwas mehr Aufmerksamkeit zu widmen. Führen wir zunächst eine Gebrauchsdefinition ein:

Definition: Zentripetalkraft

Unter der Zentripetalkraft versteht man in Inertialsystemen eine bei kreisförmigen Bewegungen auftretende Kraft, die zum Mittelpunkt der Kreisbewegung gerichtet ist. Sie ist die Folge der Wirkung einer resultierenden physikalischen Kraft, beispielsweise der Gravitationskraft.

Zentripetalkräfte treten also nur in Inertialsystemen auf, sind also nicht zu verwechseln mit Zentrifugalkräften. Die treten nur in Nicht-Inertialsystemen auf. Was ist nun mit »Wirkung« gemeint? Der Begriff der Wirkung ist in der Physik eigentlich für etwas ganz anderes belegt (für eine Größe, die die Einheit Js hat), dies ist in unserem Kontext nicht gemeint. Wirkung ist hier das, was als Folge einer (aus möglicherweise vielen Einzelkräften resultierenden) Kraftwirkung passiert, nämlich eine Änderung des Impulses. Machen wir uns das an einer beliebig ausgewählten Formel fest:

$$\underbrace{\frac{\mathrm{d}p}{\mathrm{d}t}}_{\text{Wirkung der Kraft}} = \underbrace{G\frac{mM}{r^2}}_{\text{wirkende Kraft}} \tag{3.27}$$

In diesem Beispiel ist die physikalisch wirkende Kraft die Gravitationskraft, die Wirkung ist die Impulsänderung des Objekts mit Masse

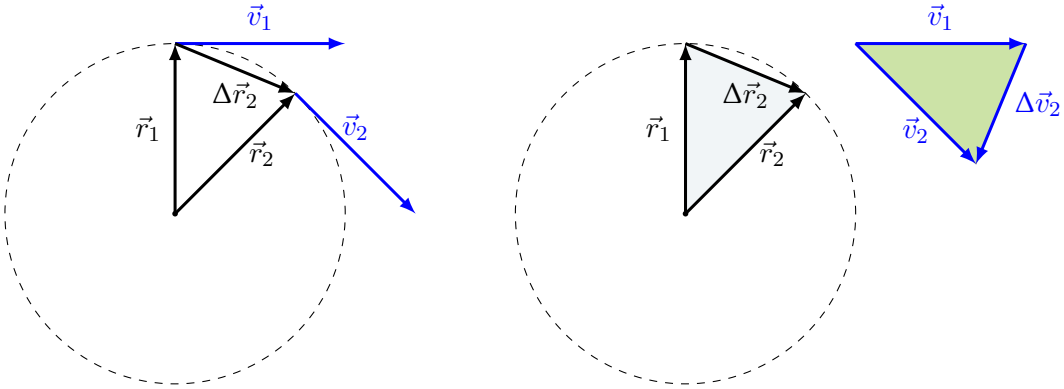


Abbildung 3.5: Zur Herleitung der Zentripetalbeschleunigung. Hierzu nutzt man die Ähnlickeit des grau und des grün hinterlegten Dreiecks aus.

m. Wenn diese Masse nun konstant bleibt, so kann man den Ausdruck etwas mehr konkretisieren:

$$\underbrace{m\frac{\mathrm{d}v}{\mathrm{d}t} = ma}_{\text{Wirkung der Kraft}} \quad = \quad \underbrace{G\frac{mM}{r^2}}_{\text{wirkende Kraft}} \tag{3.28}$$

Bei konstanter Masse m besteht die Wirkung der Kraft also in einer Beschleunigung dieser Masse. Bei Kreisbewegungen, also Bahnen mit konstantem Radius r bei denen der Körper einen konstanten Geschwindigkeitsbetrag v aufweist, kann man diese Beschleunigung weiter präzisieren, sie nimmt eine spezielle mathematische Form an.

Definition: Zentripetalbeschleunigung a_{zp}

$$a_{\mathrm{zp}} = \frac{v^2}{r} \tag{3.29}$$

Wie kommt man nun auf diesen Ausdruck? Eine mögliche Herleitung ist in Abbildung 3.5 gezeigt. Ein Körper sei zur Zeit t_1 am Ort $\vec{r}_1$, zur Zeit t_2 dann am Ort $\vec{r}_2$. Da er sich auf einer Kreisbahn befinden soll, sind die Beträge der beiden Ortsvektoren natürlich gleich groß. Dadurch, dass die einige physikalisch wirkende Kraft in Richtung des Kreismittelpunkt zeigen muss (wir betrachten ja hier eine Zentralkraft), kann sich auch am Betrag der Geschwindigkeit nichts ändern, da diese Kraftwirkung immer senkrecht zur Tangentialgeschwindigkeit $\vec{v}$ steht und das Superpositionsprinzip uns sagt, dass die Kraft keinen Einfluss auf dieses $\vec{v}$ hat. Vektoriell ändert sich natürlich die Richtung der

Geschwindigkeit, was durch das grün hinterlegte Dreieck angedeutet wird, welches man durch Parallelverschiebung von $\vec{v}_2$ an $\vec{v_1}$ erhält. Die Richtungsänderung wird mit $\Delta\vec{v}$ bezeichnet, die Ortsänderung mit $\Delta\vec{r}$.

Zwischen den beiden Dreiecken (grün und hellblau) besteht eine Ähnlichkeit, die man nun in folgender Form ausnutzt (Anwendung des Strahlensatzes):

$$\frac{\Delta r}{r} = \frac{\Delta v}{v} \tag{3.30}$$

Wir stellen diesen Ausdruck nun nach Δv um und teilen durch Δt:

$$\frac{\Delta v}{\Delta t} = \frac{v}{r}\frac{\Delta r}{\Delta t}. \tag{3.31}$$

Nun können wir eine Grenzwertbetrachtung durchführen und die in Gleichung 3.32 formulierte mittlere Zentripetalbeschleunigung in die momentane Zentripetalbeschleunigung überführen:

$$a_{\mathrm{zp}} = \lim_{\Delta t\to 0}\frac{\Delta v}{\Delta t} = \frac{v}{r}\left(\lim_{\Delta t\to 0}\frac{\Delta r}{\Delta t}\right). \tag{3.32}$$

Im letzten Term konnte man also die Momentangeschwindigkeit

$$v = \lim_{\Delta t\to 0}\frac{\Delta r}{\Delta t} \tag{3.33}$$

einsetzen und bekommt so den bekannten Ausdruck für die Zentripetalbeschleunigung. Der aufmerksame Leser wird vielleicht bemerkt haben, dass die Geschwindigkeitsänderung in Abbildung 3.5 gar nicht zum Kreismittelpunkt zeigt, wenn wir dieses Dreieck an die Spitze von $\vec{r}_1$ legen. Das liegt daran, dass wir zeichnerisch ein recht großes $\Delta\vec{r}$ angesetzt haben. Machen wir dieses Intervall immer kleiner, so wird das $\Delta\vec{v}$ immer mehr in Richtung des Kreismittelpunkts zeigen und im Grenzfall $\Delta t \to 0$ dann schließlich genau dorthin.

Die Zentripetalkraft ist also die Wirkung aller auf einen Körper wirkenden Kräfte, der eine Kreisbahn mit Radius r vollführt, wenn man in einem Intertialsystem argumentiert. Diese Feststellung so so fundamental, dass wir sie als Merksatz festhalten:

Merksatz zur Zentripetalkraft

Bewegt sich ein Körper in einem Inertialsystem auf einer Kreisbahn mit konstantem Radius r, dann muss die Summe aller auf diesen Körper wirkenden Kräfte eine Zentripetalkraft sein, d. h.

$$\sum_i \vec{F}_i = \frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = -m\frac{v^2}{r}\vec{e}_r = -m\omega^2 r\vec{e}_r \tag{3.34}$$

Im letzten Schritt haben wir ausgenutzt, dass zwischen Winkelgeschwindigkeit ω und Geschwindigkeit v auf einer Kreisbahn mit Radius r der Zusammenhang $v = \omega r$ besteht. Bei Bewegungen auf beliebigen Bahnen ist der Ausdruck etwas allgemeiner formuliert:

$$\vec{v} = \vec{\omega} \times \vec{r} \tag{3.35}$$

Bei der Kreisbewegung steht aber $\vec{r}$ immer senkrecht auf $\vec{v}$, so dass sich für den Betrag der Geschwindigkeit die Beziehung $v = \omega r$ ergibt.

3.6.2 Zentrifugalkraft

Bei der Zentrifugalkraft handelt es sich um eine so genannte Scheinkraft, also um eine Kraft, die in beschleunigten Bezugssystemen eingeführt wird, um weiterhin mit den Newtonschen Bewegungsgleichungen rechnen zu können. Im Fall der Zentrifugalkraft geht es wieder um eine Kreisbewegung mit konstantem Radius r. Diesmal betrachten wir die Bewegung allerdings nicht aus einem Bezugssystem heraus, sondern in einem mit der kleinen Masse m verbundenen Bezugssystem. Dieses Bezugssystem folgt der kleinen Masse so, dass Sie als Beobachter diese Masse die ganze Zeit über an der selben Stelle stehen sehen. Sie erscheint in Ruhe und wir interpretieren dies im Rahmen der Newtonschen Axiome so, dass auf sie keine Kraft wirkt. Halten wir das in einem Merksatz fest:

Merksatz zur Zentrifugalkraft

Bewegt sich ein Körper auf einer Kreisbahn mit konstantem Radius r und wird von einem mit der bewegten Masse aus verbundenen Bezugssystem aus beobachtet, dann muss die Summe aller auf diesen Körper wirkenden Kräfte Null ergeben, d. h.

$$\sum_i \vec{F}_i = 0 \tag{3.36}$$

Nun haben wir natürlich das Problem, dass es auf unserer Kreisbahn aus physikalischer Sicht eine wirkende Kraft gibt, nämlich die Gravitationskraft. Diese müssen wir mit einer gleichgroßen, aber entgegengesetzt gerichteten Kraft addieren, damit in Summe Null herauskommt. Hierzu nehmen wir einfach die sich einstellende Wirkung aus der Inertialsystem-Betrachtung, tauschen das Vorzeichen und fassen zusammen:

$$\frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = \sum_i \vec{F}_i + m\frac{v^2}{r}\vec{e}_r = \sum_i \vec{F}_i + m\omega^2 r\vec{e}_r = 0 \tag{3.37}$$

Vergleicht man Gleichungen 3.34 und 3.37 miteinander, stellt man fest, dass diese in ihrer physikalischen Aussage identisch sind. Beide Gleichungen liefern die gleichen Lösungen, wenn man beispielsweise die Geschwindigkeit v der Kreisbahn sucht. Man kann also ohne Verständnis der Auswirkung der Wahl des verwendeten Bezugssystems ein richtiges Ergebnis erhalten, aber wirklich befriedigend ist das natürlich nicht.

3.6.3 Alternative Betrachtung

Es gibt eine weitere Möglichkeit, die Zentripetalbeschleunigung (und damit natürlich auch die Zentripetalkraft) herzuleiten. Wir beschreiben die Kreisbewegung dazu in ebenen Polarkoordinaten. Zu einer gegebenen Zeit t befindet sich der Körper an einem Ort $\vec{r}(t)$. Mit zugehörigem Einheitsvektor $\vec{e}_r$ kann man also schreiben:

$$\vec{r} = r\vec{e}_r. \tag{3.38}$$

Unter der Beschleunigung versteht man die zweite Ableitung des Ortes nach der Zeit. Bei einer Bewegung auf einer Kreisbahn mit Radius R sollte sich daher der gleiche Ausdruck ergeben wie bei unserer geometrischen Herleitung. Nun können wir Gleichung 3.38 zweimal nach der Zeit ableiten und schauen, was dabei herauskommt. Für die erste Ableitung gilt (da es sich um ein Produkt zweier Größen handelt, nämlich dem skalaren r und dem Einheitsvektor $\vec{e}_r$, können wir die Produktregel verwenden):

$$\dot{\vec{r}} = \vec{v} = \dot{r}\vec{e}_r + r\dot{\vec{e}}_r. \tag{3.39}$$

Bei einer Kreisbahn ist der Radius immer gleich groß, so dass $\dot{r} = 0$ gelten muss. Es stellt sich nun die Frage, wie man den Einheitsvektor nach der Zeit ableiten muss. Dazu muss man eigentlich nur wissen, wie dieser Vektor aussieht und dann jede Zeile nach t ableiten. Der Einheitsvektor $\vec{e}_r$ setzt sich aus den x- und y-Koordinaten bestimmen. Die x-Komponente beträgt beispielsweise $x = r\cos\varphi$, wobei φ der Winkel zwischen x-Achse und Ortsvektor $\vec{r}$ ist. Für die y-Achse gilt entsprechend $y = r\sin\varphi$, also

$$\vec{r} = \begin{pmatrix} x \\ y \end{pmatrix} = r \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix} = r\vec{e}_r. \tag{3.40}$$

Der radiale Einheitsvektor setzt sich also aus Sinus und Kosinus des Winkels φ zusammen. Der Winkel selbst ist natürlich zeitabhängig,

da sich der Körper zu verschiedenen Zeiten an verschiedenen Orten befindet. Mit der Kettenregel findet man:

$$\dot{\vec{e}}_r = \dot{\varphi} \begin{pmatrix} -\sin\varphi \\ \cos\varphi \end{pmatrix} = \dot{\varphi}\vec{e}_\varphi. \tag{3.41}$$

Man kann sich geometrisch recht einfach klarmachen, dass der Vektor $\vec{e}_\varphi$ senkrecht zu $\vec{e}_r$ steht, die Länge Eins hat und in Richtung der Änderung des Winkels φ zeigt. Die zeitliche Änderung des Winkels wird als Winkelgeschwindigkeit $\omega = \dot{\varphi}$ bezeichnet. Bei einer Kreisbahn ist die Winkelgeschwindigkeit konstant, also $\dot{\omega} = 0$. Wir halten fest:

$$\vec{v} = \dot{r}\vec{e}_r + r\dot{\varphi}\vec{e}_\varphi. \tag{3.42}$$

Die Beschleunigung erhält man durch nochmaliges Ableiten nach der Zeit, wobei wir bereits $\ddot{\vec{r}} = 0$ und $\ddot{\varphi} = 0$ berücksichtigen:

$$\vec{a} = \dot{\vec{v}} = \ddot{\vec{r}} = -r\dot{\varphi}^2\vec{e}_r = -r\omega^2\vec{e}_r = -\frac{v^2}{r}\vec{e}_r. \tag{3.43}$$

Wir bekommen also duch zweimaliges Ableiten des Ortsvektors nach der Zeit unter Berücksichtigung der Besonderheiten einer Kreisbahn (konstanter Radius, konstanter Betrag der Umlaufgeschwindigkeit) den bekannten Ausdruck für die Zentripetalbeschleunigung.

3.7 Energie, Arbeit und Kraft

Wir verlassen nun den Bereich der Kräfte und wenden uns einer skalaren Größe zu, die in der Physik eine besonders wichtige Rolle spielt: der Energie. Zu Beginn soll ein Auszug aus den Feynman-Vorlesungen zum Thema Energieerhaltung verdeutlichen, wie fundamental diese Größe ist:

> There is a fact, or if you wish, a law, governing all natural phenomena that are known to date. There is no known exception to this law—it is exact so far as we know. The law is called the conservation of energy. It states that there is a certain quantity, which we call energy, that does not change in the manifold changes which nature undergoes. That is a most abstract idea, because it is a mathematical principle; it says that there is a numerical quantity which does not change when something happens. It is not a description of a mechanism, or anything concrete; it is just a strange fact that

> we can calculate some number and when we finish watching nature go through her tricks and calculate the number again, it is the same. (Something like the bishop on a red square, and after a number of moves—details unknown—it is still on some red square. It is a law of this nature.)

Der Begriff »Energie« ist jedem geläufig, doch lassen Sie uns kurz vergessen, was wir darüber zu wissen glauben, und mit einer einfachen Definition auf Basis des obigen Zitats beginnen:

Energie

Energie ist ein Zahlenwert (mit einer Einheit), die wir auf irgendeine Art und Weise berechnen können.

Man könnte einwenden, dass dies für alle physikalischen Größen gilt, und damit läge man völlig richtig. Auch der Impuls ist ein berechenbarer Zahlenwert mit Einheit. Wie wir gesehen haben, hängt der Impuls vom Bezugssystem ab und ist daher keine universelle Größe, die in jeder Betrachtung gleich bleibt – im Gegensatz zu Naturkonstanten oder (in der klassischen Mechanik) zur Masse eines Körpers. Auch bei der Energie spielt das Bezugssystem eine Rolle: Ein Körper mit der Geschwindigkeit v in einem Bezugssystem besitzt dort kinetische Energie, während er in einem Bezugssystem, in dem er ruht, keine kinetische Energie hat.

Es gibt zudem bestimmte Energieformen, insbesondere die potentiellen Energien, die eine zentrale Rolle in der Physik spielen. Bei diesen Formen bestimmt nicht nur das Bezugssystem den Energiewert, sondern vor allem die Lage eines Körpers relativ zu anderen Körpern. Man kann einem Punkt im Bezugssystem willkürlich einen beliebigen Energiewert zuordnen, wobei andere Punkte relativ dazu unterschiedliche Energiewerte haben. Wichtig ist, dass letztlich nur der Energieunterschied zwischen zwei Punkten von Bedeutung ist.

Die Wahl des Bezugspunkts für die Energie ist willkürlich, und es ist häufig sinnvoll (aber nicht zwingend), dem Bezugspunkt eine Energie von Null zuzuordnen, da dies die Berechnungen in der Regel vereinfacht. Man könnte jedoch genauso gut einen anderen Wert wie 5,345 oder $\pi/2$ wählen. Die Null ist jedoch meist die praktischere Wahl. Halten wir zunächst allgemein fest, wie man den Nullpunkt der potenziellen Energie definiert.

Wo liegt der Nullpunkt der potentiellen Energie

Der Nullpunkt der potentiellen Energie liegt dort, wo man ihn hinlegt.

Bei potentiellen Energien ist der absolute Zahlenwert irrelevant – entscheidend ist der Wert relativ zu einem festgelegten Bezugswert.

Der Energiebegriff wird noch komplexer, wenn man berücksichtigt, dass er auch von der Wahl der Systemgrenzen abhängt. Ein Körper kann innerhalb bestimmter Systemgrenzen potenzielle Energie besitzen, während er innerhalb anderer Grenzen keine hat, da sonst Widersprüche auftreten würden. Dieses Thema wird im Abschnitt 3.7.4 genauer behandelt, wenn es um innere und äußere Kräfte und deren Einfluss auf die Gesamtenergie eines Systems geht.

3.7.1 Potentielle Energie

Den Begriff der potenziellen Energie haben wir bereits eingeführt, aber es fehlt noch eine Definition. Potentielle Energie ist eine skalare Größe (im SI-System gemessen in Joule), die mit einer zugehörigen Kraft durch eine mathematische Operation verknüpft ist. Anders ausgedrückt: Bestimmte Kräfte lassen sich durch ein Potential bzw. eine potentielle Energie beschreiben. In der Mechanik verwenden wir die Begriffe Potential und potentielle Energie synonym und nutzen den Buchstaben V. Erst in der Elektrodynamik wird zwischen diesen Begriffen unterschieden. Der Zusammenhang zwischen potentieller Energie V und der Kraft $\vec{F}$ lautet:

$$\vec{F} = -\mathrm{grad}\, V \tag{3.44}$$

Was bedeutet nun der Begriff »grad«? Er bezeichnet eine mathematische Operation, die auf eine skalare Potentialfunktion angewendet wird – die Bildung des Gradienten dieser Funktion. In kartesischen Koordinaten und unter Berücksichtigung aller drei Raumrichtungen ist der Gradient einer Funktion $f(x, y, z)$ wie folgt definiert:

$$\begin{aligned} \mathrm{grad} f(x,y,z) &= \nabla f(x,y,z) \\ &= \vec{e}_x \frac{\partial f(x,y,z)}{\partial x} + \vec{e}_y \frac{\partial f(x,y,z)}{\partial y} + \vec{e}_z \frac{\partial f(x,y,z)}{\partial z}. \end{aligned} \tag{3.45}$$

Hier haben wir einen neuen »Buchstaben« für den Gradienten eingeführt, den sogenannten Nabla-Operator ∇. Dieser dient als abkürzende

Schreibweise, da Physiker – so sagt man – tendenziell schreibfaul sind. Der Nabla-Operator spielt eine zentrale Rolle in der Elektrodynamik, da die gesamten Maxwell-Gleichungen mit ihm formuliert werden (können). Nehmen Sie diesen Operator daher ernst; er ist keine akademische Spielerei, die nur in seltenen Fällen Anwendung findet.

Machen wir ein Beispiel: Aus der Schule oder aus der Experimentalphysik-Vorlesung sollte man das Federkraftgesetz kennen. Eine Feder mit Federkonstante D wird aus ihrer Gleichgewichtslage (diese definieren wir als $x = 0$) um eine Strecke x gedehnt (die Richtung der Dehnung definieren wir als positive x-Richtung), die Feder reagiert mit einer rücktreibenden Federkraft $F = -Dx$ in Richtung der Gleichgewichtslage. Wenn es sich bei der Federkraft um eine konservative Kraft handeln sollte, dann muss es ein Potential zu dieser Kraft geben, für das gilt (wir haben hier ein eindimensionales Problem und können die partielle Ableitung durch den Differentialquotienten ersetzen):

$$-Dx = -\frac{\mathrm{d}V(x)}{\mathrm{d}x} \tag{3.46}$$

Der mathematisch geübte Leser wird gleich erkennen, dass die gesuchte Funktion die Form $V(x) = \frac{1}{2}Dx^2$ hat. Man kann natürlich auch ganz formal vorgehen, erstmal die Minuszeichen kürzen und dann das Ganze in eine integrierbare Form bringen, d. h.

$$\mathrm{d}V = D\,x\,\mathrm{d}x. \tag{3.47}$$

Bilden wir nun das unbestimmte Integral:

$$\int \mathrm{d}V = D \int x\,\mathrm{d}x, \tag{3.48}$$

also

$$V = \frac{1}{2}Dx^2 + C. \tag{3.49}$$

Die Integrationskonstante C kann beliebig festgelegt werden. Üblicherweise wählt man für die Ruheposition bzw. Gleichgewichtslage der Feder die Koordinate $x = 0$ und legt dort willkürlich fest, dass $V(x = 0) = 0$ ist. Wie bereits erwähnt, könnte man auch jeden anderen Wert wählen, da es letztlich nur auf den Energieunterschied zum Bezugspunkt ankommt. Energien sind – wie wir anfangs festgehalten haben – nichts anderes als Zahlenwerte, die relativ zu einem gewählten Bezugspunkt angegeben werden.

Man sollte sich Energie nicht wie einen Eimer vorstellen, den man „füllen“ kann, etwa wie Wasser. Energie ist keine greifbare Größe, sondern bleibt ein abstrakter Begriff – eine Zahl, die wir berechnen können.

Wir beschäftigen uns in dieser Veranstaltung hauptsächlich mit der Gravitationskraft $\vec{F}_\mathrm{G}$ und es stellt sich natürlich die Frage, ob diese Kraft konservativ ist bzw. ob es zu dieser Kraft ein zugehöriges Potential gibt. Die Gravitationskraft ist eine sogenannte Zentralkraft, was durch eine geschickte Wahl des Bezugssystems deutlich wird. »Geschickt« bedeutet zum Beispiel, dass man den Massenmittelpunkt eines deutlich schwereren Zentralkörpers als Ursprung des Bezugssystems wählt, wodurch in guter Näherung ein Inertialsystem entsteht. Bei ähnlich schweren Massen muss der Bezugspunkt im Schwerpunkt der beiden Massen liegen, wodurch man ein echtes Inertialsystem erhält, nicht nur eine Näherung. Da wir uns hauptsächlich mit dem Keplerproblem befassen, können wir in guter Näherung von einem Inertialsystem ausgehen, wenn das Bezugssystem im Massenmittelpunkt der schwereren Masse platziert wird.

Eine Zentralkraft weist nur eine radiale Abhängigkeit der Kraftstärke auf, Winkelabhängigkeiten gibt es bei Zentralkräften also nicht. Generell ist bei Zentralfeldern die Verwendung von Kugelkoordinaten sinnvoll. Der Nabla-Operator ∇ hat in Kugelkoordinaten (r, ϑ, φ) die Form

$$\nabla = \vec{e}_r \frac{\partial}{\partial r} + \vec{e}_\vartheta \frac{1}{\vartheta} \frac{\partial}{\partial \vartheta} + \vec{e}_\varphi \frac{1}{r \sin \vartheta} \frac{\partial}{\partial \varphi}. \tag{3.50}$$

Wenn es nur eine radiale Abhängigkeit gibt, dann bleibt auch nur die $\vec{e}_r$-Komponente übrig, aus der partiellen Ableitung wird dann der Differentialquotient. Die Fragestellung ist also, ob es eine Funktion V gibt, für die gilt:

$$\vec{F}_\mathrm{G} = -\frac{\gamma m M}{r^2} \vec{e}_r = -\vec{e}_r \frac{\mathrm{d}V}{\mathrm{d}r} \tag{3.51}$$

Wir gehen genauso wie beim Federpotential vor und bilden das Integral

$$\int \mathrm{d}V = \int \frac{\gamma m M}{r^2} \, \mathrm{d}r. \tag{3.52}$$

Man erhält somit die Beziehung

$$V(r) = -\frac{\gamma m M}{r} + C. \tag{3.53}$$

Auch hier kann man die Integrationsvariable durch Wahl einer Randbedingung des Potentials festlegen. Eine übliche Konvention ist die Annahme, dass das Gravitationspotential im Unendlichen Null werden soll, also $V(r \to \infty) = 0$. Dadurch ergibt sich $C = 0$ und für das Gravitationspotential entsprechend

$$V(r) = -\frac{\gamma m M}{r}. \tag{3.54}$$

In Abbildung 3.6 ist der Verlauf des Gravitationspotentials am Beispiel der Erde als Zentralkörper dargestellt. Man könnte sich nun fragen, wie

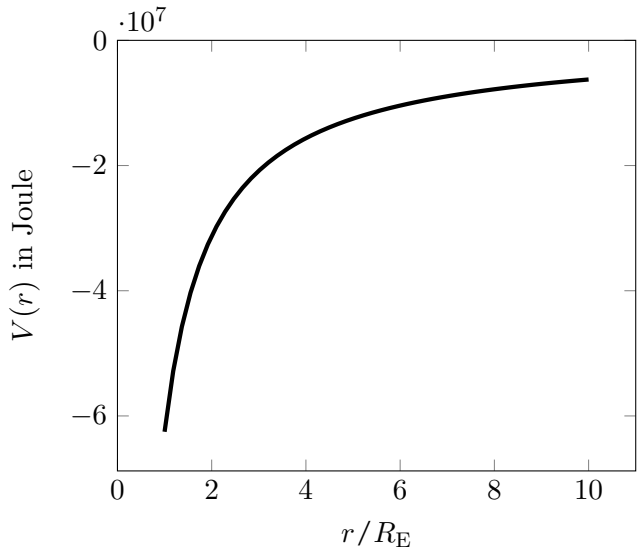


Abbildung 3.6: Verlauf des Gravitationspotentials am Beispiel der Erde als Zentralkörper für einen Körper der Masse $m = 1\,\mathrm{kg}$ als Funktion des radialen Abstands r in Vielfachen des Erdradius R_E. Für $r > R_\mathrm{E}$ kann das Potential durch den $-1/r$-Verlauf beschrieben werden, daher beginnt der Plot bei $r/R_\mathrm{E} = 1$. Im Inneren der Erde verläuft das Potential nicht hyperbolisch, aber das wird an anderer Stelle besprochen.

dieser hyperbolische Kurvenverlauf mit der aus der Schule bekannten potentiellen Gravitationsenergie der Form mgr zusammenhängt, wobei g die Erdbeschleunigung, m die Masse eines Körpers und r die Höhe relativ zu einem Bezugspunkt ist. Schauen wir uns das Gravitationspotential in der Nähe der Erdoberfläche an. Für einen Punkt genau auf der Erdoberfläche ($r = R_\mathrm{E}$) gilt

$$V(R_\mathrm{E}) = -\frac{\gamma m M}{R_\mathrm{E}}. \tag{3.55}$$

Für einen Punkt $R_E + r$ entsprechend

$$V(r + R_\mathrm{E}) = -\frac{\gamma m M}{r + R_\mathrm{E}}. \tag{3.56}$$

Wir klammern den Erdradius aus und bekommen

$$V(r + R_\mathrm{E}) = -\frac{\gamma m M}{R_\mathrm{E}\left(1 + \dfrac{r}{R_\mathrm{E}}\right)} \tag{3.57}$$

Mit der Substitution $x = r/R_\mathrm{E}$ kann man den Bruch als $(1 + x)^{-1}$ schreiben. Da $x \ll 1$ für geringe Höhen r gilt, kann man eine Taylorent-

wicklung um $x = 0$ durchführen. Es gilt (bei Abbruch der Reihenentwicklung nach dem linearen Term)

$$(1+x)^{-1} \approx 1 - x. \tag{3.58}$$

Somit kann man das Potential annähern durch

$$V(r + R_\mathrm{E}) = -\frac{\gamma m M}{R_\mathrm{E}} \frac{1}{\left(1 + \frac{r}{R_\mathrm{E}}\right)} \approx -\frac{\gamma m M}{R_\mathrm{E}} \left(1 - \frac{r}{R_\mathrm{E}}\right). \tag{3.59}$$

Das kann man in die Form

$$V(r + R_\mathrm{E}) = -\frac{\gamma m M}{R_\mathrm{E}} + \frac{\gamma m M}{R_\mathrm{E}^2} r = -\frac{\gamma m M}{R_\mathrm{E}} + mgr \tag{3.60}$$

bringen. Das Potential an der Stelle $r + R_\mathrm{E}$ setzt sich also aus dem Potential an der Erdoberfläche plus einem additiven Term mgr zusammen. Wenn man also nur die Änderung der potentiellen Energie relativ zum Bezugspunkt Erdoberfläche haben will, dann kann man in guter Näherung diese Formel benutzen. Natürlich kann man auch jeden anderen Bezugspunkt auswählen, die Wahl der Erdoberfläche war ja willkürlich gewählt. An jeder beliebigen Stelle der allgemeinen Potentialkurve kann eine Taylorentwicklung durchgeführt werden, somit ist der Bezugspunkt also völlig beliebig. Er liegt dort, wo man ihn hinlegt.

3.7.2 Arbeit

Das Konzept der Arbeit spielt eine zentrale Rolle in der Physik. Wir haben es bereits indirekt verwendet, als wir zur Berechnung des Potentials das Wegintegral über eine konservative Kraft gebildet haben. Nun wollen wir dies verallgemeinern und betrachten, welche Arbeit notwendig ist, um einen Körper auf einem beliebigen Weg durch ein Kraftfeld zu bewegen. Die wirkende Kraft $\vec{F}$ unterliegt dabei keiner speziellen Bedingung und muss nicht konservativ sein. Auch der Weg ist nicht näher spezifiziert; es kann ein beliebiger Pfad zwischen zwei Punkten A und B sein.

An jedem Punkt des Weges wirkt die Kraft in eine bestimmte Richtung. Für eine infinitesimal kleine Wegstrecke $\mathrm{d}\vec{s}$ ist nur die Komponente der Kraft relevant, die in Richtung dieser Strecke zeigt. Die Arbeit, die dabei am Körper verrichtet wird, beträgt $\mathrm{d}W = \vec{F} \cdot \mathrm{d}\vec{s}$. Die gesamte Arbeit W entlang des Weges erhält man durch das Aufsummieren dieser einzelnen Beiträge, was in einer Integration resultiert:

Definition: Arbeit

Unter Arbeit versteht man das Wegintegral der Kraft:

$$W = \int_A^B \mathrm{d}W = \int_A^B \vec{F} \cdot \mathrm{d}\vec{s}. \tag{3.61}$$

3.7.3 Kinetische Energie

Bei der kinetischen Energie handelt es sich um die Bewegungsenergie eines Körpers der Masse m. Der Energiegehalt ändert sich, wenn der Körper seine Geschwindigkeit ändert, was voraussetzt, dass der Körper beschleunigt wurde. Für eine Beschleunigung ist es erforderlich, dass an dem Körper eine Kraft angreift, somit eine Arbeit geleistet wird, die zur Erhöhung der kinetischen Energie des Körpers führt. Die an dem Körper geleistete Arbeit ist allgemein (und ohne nähere Festlegung von Anfangs- und Endposition des Wegintegrals) definiert als

$$W = \int \vec{F} \cdot \mathrm{d}\vec{s}. \tag{3.62}$$

Wir erweitern diesen Ausdruck mit dem Zeitintervall $\mathrm{d}t$:

$$W = \int \vec{F} \cdot \frac{\mathrm{d}\vec{s}}{\mathrm{d}t}\,\mathrm{d}t = \int \vec{F} \cdot \vec{v}\,\mathrm{d}t. \tag{3.63}$$

Die Kraft $\vec{F}$ kann man für einen Körper konstanter Masse ausdrücken als $\vec{F} = m \cdot \vec{a}$. Wir setzen ein:

$$W = \int \vec{F} \cdot \vec{v}\,\mathrm{d}t = \int m \cdot \vec{a} \cdot \vec{v}\,\mathrm{d}t. \tag{3.64}$$

Die Beschleunigung $\vec{a}$ ist definiert als $\vec{a} = \mathrm{d}\vec{v}/\,\mathrm{d}t$. Die Masse m ist eine Konstante und kann vor das Integral gezogen werden. Wir setzen dies ebenfalls ein und erhalten:

$$W = \int m \cdot \vec{a} \cdot \vec{v}\,\mathrm{d}t = m \cdot \int \frac{\mathrm{d}\vec{v}}{\mathrm{d}t}\vec{v}\,\mathrm{d}t = m \cdot \int \vec{v}\,\mathrm{d}\vec{v} = \frac{1}{2}mv^2 \tag{3.65}$$

Die kinetische Energie wird häufig mit W_{kin} abgekürzt, oft auch mit T, was etwas weniger Schreibarbeit bedeutet und was daher in diesem Skript bevorzugt wird. Sie ergibt sich nach unserer Herleitung zu dem bekannten Ausdruck

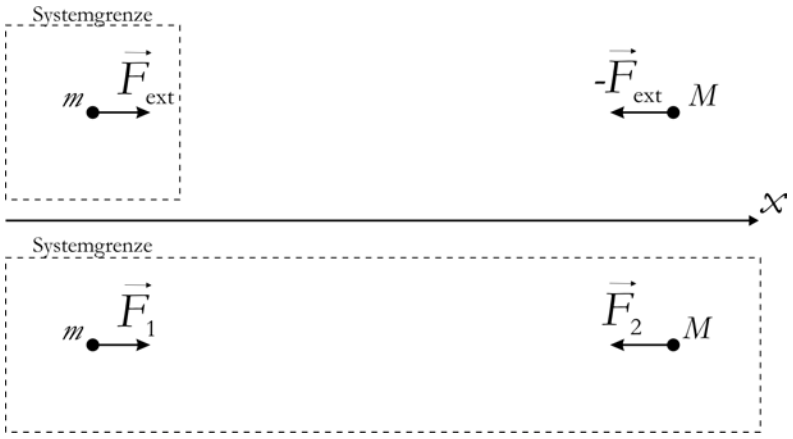


Abbildung 3.7: Die Wahl meiner Systemgrenze definiert, welche Kräfte als innere und äußere Kräfte gezählt werden. Im oberen Fall übt die Masse M eine Kraft auf die Masse m aus. Innerhalb des definierten Systems soll nur die Masse m liegen, weswegen die auf sie wirkende Kraft eine externe Kraft darstellt. Die nach dem dritten Newtonschen Axiom auf die Masse M wirkende Reaktionskraft muss es natürlich auch geben, soll aber zunächst keine Rolle bei unserer Betrachtung spielen. Durch die (externe)Kraftwirkung wird die Masse m beschleunigt, die Kraft $\vec{F}_{\text{ext}}$ leistet also Arbeit an m und erhöht die Gesamtenergie des Systems, die sich nur aus kinetischer Energie von m zusammensetzt. Im unteren Fall sind beide Massen Teil des definierten Systems, beide Kräfte zählen daher also zu den inneren Kräften. Die zentrale Aussage dieses Abschnitts lautet, dass innere Kräfte, sofern sie konservativer Natur sind, keine Arbeit am System leisten, als Folge daher die Gesamtenergie eines Systems sicher nicht ändert, wenn nur innere konservative Kräfte im Spiel sind.

Definition: Kinetische Energie

$$T = \frac{1}{2}mv^2 \tag{3.66}$$

3.7.4 Innere und äußere Kräfte

Für die Beschreibung eines physikalischen Sachverhalts ist es häufig sinnvoll, nur einen Ausschnitt des physikalischen Problems zu betrachten, der von besonderem Interesse ist. Man definierte hierzu eine Systemgrenze, die man sich genauer anschaut und betrachtet die Wechselwirkung mit der Umgebung als äußere Einflüsse.

In den Feynman-Vorlesungen wird dieses Vorgehen für den Begriff ≫Energieerhaltung≪ am Beispiel eines Kinderzimmers und der Anzahl an Bauklötzen in diesem Zimmer näher ausgeführt. Die Bauklötze stehen dabei für die Energie, die in einem System (dem Zimmer) vorhanden ist. Eine Mutter stellt fest, dass egal was tagsüber mit den Klötzen passiert, abends immer die gleiche Anzahl davon vorhanden ist. Die Energie scheint also erhalten zu sein. Manchmal stellt diese Mutter aber auch fest, dass ein paar Klötze mehr im Zimmer sind. Offenbar sind sie von einem Spielkameraden ins Zimmer gebracht worden. Sie stammen also aus der Umgebung des Systems. Es kann auch vorkommen, dass das Kind ein paar Klötze aus dem Fenster des Zimmers in den Garten wirft, was die Anzahl im Zimmer reduziert. Die Botschaft hier ist, dass ein System mit seiner Umgebung in einer Beziehung steht. Der Fall eines vollständig abgeschlossenen Systems ist dabei der einfachste und auch der langweiligste Fall. Wenn das Zimmer keine Tür und keine Fenster hat, können auch keine Klötze rein oder raus. Übersetzt heißt dies, dass die Energie in einem abgeschlossenen System eine Erhaltungsgröße sein muss. Das ist richtig, aber auch banal. Interessanter ist die Tatsache, dass die Energie, die in einem System steckt, verändert werden kann. Dies passiert genau dann, wenn an einem System Arbeit durch eine äußere Kraft verrichtet wird.

Die Festlegung einer Systemgrenze ist willkürlich, daher kann man durch Änderung der Systemgrenzen jede äußere Kraft in eine innere Kraft umwandeln und umgekehrt. Es gibt also keinen Unterschied in der physikalischen Natur von inneren und äußeren Kräften, es ist nur eine Frage der Betrachtungsweise. Wichtig ist aber, dass die physikalische Beschreibung der Bewegung eines Körpers nicht von der Wahl der Systemgrenzen abhängt, man also für verschiedene Systeme die gleiche Physik erhält. Es darf also nicht zu Widersprüchen kommen, wenn ein physikalischer Sachverhalt von verschiedenen Systemgrenzen aus beschrieben wird.[8] Wir beginnen zunächst mit einer Definition des Zusammenhangs zwischen Arbeit und Energie:

8 Die Darstellung dieses Abschnitts folgt inhaltlich den Lehrbüchern *Physik für Wissenschaftler und Ingenieure* von Tipler sowie *Klassische Mechanik* von Rainer Müller. Insbesondere das letztgenannte ist eines der wenigen Lehrbücher überhaupt, das sich ausführlich mit dem Begriff der Systemgrenze auseinandersetzt.

Zusammenhang zwischen Energie und Arbeit

Unter der Arbeit W versteht man den Transfer von Energie in ein System hinein oder aus einem System heraus durch Verschiebung eines Objekts des Systems unter externer Kraftwirkung.

Wichtig ist hierbei, dass es sich um eine externe Kraft handeln muss. Es gibt hier nun zwei mögliche Vorzeichen der Arbeit, deren physikalische Bedeutung festgelegt werden muss. Wir werden dies hier folgendermaßen tun:

- $W > 0$: das System bekommt Energie aus der Umgebung. Die Energie des Systems erhöht sich.
- $W < 0$: das System verliert Energie, die Energie des Systems nimmt also ab.

In Abbildung 3.7 ist der gleiche physikalische Sachverhalt mit zwei unterschiedlichen Systemgrenzen dargestellt. Im obigen Beispiel wirkt auf eine Masse m eine äußere Kraft $\vec{F}_{\text{ext}}$. Auf die besondere Natur dieser Kraft soll hier noch gar nicht eingegangen werden, dennoch können wir bereits einiges über unser System aussagen. Zunächst dürfte klar sein, dass der Gesamtimpuls unseres Systems, der hier nur aus dem Impuls der Masse m besteht, nicht erhalten sein wird, da die Masse durch die äußere Kraft ja beschleunigt wird. Ferner leistet die äußere Kraft bei Bewegung in Kraftrichtung eine positive Arbeit (Bewegung in positiver x-Richtung) an dem System, wodurch sich die Energie des Systems erhöhen wird. Auch das ist nicht weiter verwunderlich, da die Geschwindigkeit der Masse m sich erhöht. Die durch $\vec{F}_{\text{ext}}$ geleistete Arbeit wird also in kinetischer Energie der Masse m gespeichert.

Man kann übrigens genauso argumentieren, wenn man die Systemgrenze nur um die Masse M legen würde. In diesem Fall wäre $-\vec{F}_{\text{ext}}$ die externe Kraft, die Masse M bewegt sich in Richtung der Masse m (also in negativer x-Richtung), weswegen die geleistete Arbeit an diesem System ebenfalls positiv ist und in kinetischer Energie der Masse M steckt.

Jetzt legen wir unsere Systemgrenze so fest, dass beide Massen sich innerhalb des Systems befinden sollen. Die zwischen den Massen wirkenden Kräfte sind nun ausschließlich innere Kräfte, durch das dritte Newtonsche Axiom ist zudem festgelegt, dass $\vec{F}_1 = -\vec{F}_2$ gelten muss.

Für die Änderung des Gesamtimpulses $\Delta\vec{p}$ des Systems in einem Zeitintervall $\mathrm{d}t$ gilt daher

$$\Delta\vec{p} = \Delta\vec{p}_1 + \Delta\vec{p}_2 = \int \overbrace{\left(\vec{F}_1 + \vec{F}_2\right)}^{=0} \mathrm{d}t = 0. \tag{3.67}$$

Der Gesamtimpuls unseres Systems ändert sich also nicht, wenn ausschließlich innere Kräfte wirken. Wie verhält es sich nun mit der Energie des Systems? Bei der Einzelbetrachtung der beiden Massen, bei der die Systemgrenze jeweils nur eine Masse umfasst, hat sich die Gesamtenergie des Systems erhöht, da eine äußere Kraft Arbeit am System geleistet hat. Wenn jedoch beide Massen Teil des Systems sind, gibt es per Definition keine äußere Kraft. Aber ist die Energie des Systems dadurch automatisch eine Erhaltungsgröße?

Beide Massen werden durch die gegenseitige Kraftwirkung beschleunigt, wodurch ihre kinetischen Energien zunehmen. Da kinetische Energien bei endlichen Geschwindigkeiten immer positiv sind, wächst die gesamte kinetische Energie der beiden Massen stetig – zumindest bis sie aufeinander prallen, was wir hier vernachlässigen.

Hier scheint ein Widerspruch vorzuliegen: Die Energie des Systems nimmt zu, obwohl keine äußere Kraft wirkt. Dieser Widerspruch lässt sich auflösen, wenn neben der kinetischen Energie T auch eine weitere Energieform berücksichtigt wird: die potentielle Energie V. Für die weitere Betrachtung führen wir den Begriff der mechanischen Gesamtenergie ein, der nützlich sein wird:

Definition: Mechanische Gesamtenergie E

Die mechanische Gesamtenergie E eines Systems ist die Summe aus der kinetischen Energie T und der potentiellen Energie V der im System enthaltenen Teilchen.

Im gleichen Maße, wie die kinetische Energie unseres Systems zunimmt, muss die potentielle Energie abnehmen, sodass die mechanische Gesamtenergie des Systems konstant bleibt. Allerdings ist die Existenz einer potentiellen Energie an bestimmte Eigenschaften der wirkenden Kräfte gekoppelt. Nur konservative Kräfte können über ein skalares Potentialfeld V beschrieben werden, wobei der Zusammenhang durch $\vec{F} = -\nabla V$ gegeben ist. Wenn nur konservative Kräfte vorliegen, können wir folgenden wichtigen Satz aufstellen.

Erhaltung der mechanischen Gesamtenergie

In einem System von Massenpunkten, in dem nur konservative Kräfte wirken, ist die mechanische Gesamtenergie des Systems eine Erhaltungsgröße.

Wenn im System beispielsweise Reibungskräfte vorliegen, können wir nicht davon ausgehen, dass die mechanische Gesamtenergie erhalten ist, da es zu Reibungskräften kein zugehöriges Potential gibt. Energieerhaltung im mechanischen Sinne bezeichnet also immer nur die beiden Energieformen *kinetisch* und *potentiell*. Die Existenz einer potentiellen Energie im System erfordert, dass die wirkenden inneren Kräfte konservativ sind. Unter dieser Annahme kann man festhalten:

Wirkung innerer Kräfte

Innere Kräfte leisten keine Arbeit am System, sofern es sich um konservative Kräfte handelt.

Äußere Kräfte, ob konservativ oder nicht-konservativ, haben dagegen einen direkten Einfluss auf die mechanische Gesamtenergie eines Systems:

Wirkung äußerer Kräfte

Äußere Kräfte ändern die mechanische Gesamtenergie eines Systems. Sie können die mechanische Gesamtenergie sowohl erhöhen als auch absenken.

Der folgende Abschnitt verdeutlicht den Sachverhalt anhand eines einfachen Beispiels: Eine Kiste befindet sich im Schwerefeld der Erde und wird von einer Person auf eine Höhe $\Delta z = h$ angehoben (siehe Abb. 3.8). Als Bezugssystem wählen wir ein mit der Erdoberfläche verbundenes, das wir als ruhendes Inertialsystem betrachten. Dabei ist es wichtig, die Begriffe »Bezugssystem« und »System« nicht zu verwechseln. Das System besteht aus der Kiste, der Erde und dem von der Erde erzeugten Gravitationsfeld.

Innerhalb dieses Systems gibt es innere Kräfte – konkret die Gravitationskräfte zwischen der Kiste und der Erde. Alle anderen Kräfte,

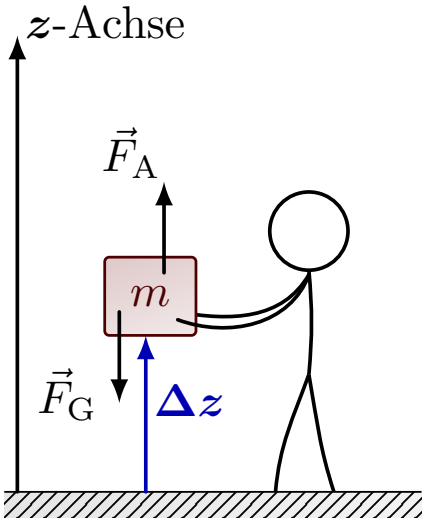


Abbildung 3.8: Am Beispiel des Anhebens einer Kiste der Masse m soll der Begriff der Arbeit, die an einem System geleistet wird, verdeutlicht werden. Die Kiste wird natürlich als Massenpunkt behandelt.

die nicht durch die Gravitationswechselwirkung dieser beiden Objekte verursacht werden, sind als äußere Kräfte zu betrachten. Wir nehmen weiterhin an, dass die Erde sich durch die Gravitationskraft der Kiste kaum beeinflussen lässt, sodass wir eine vereinfachte Form des Keplerproblems betrachten. Die Erde bleibt in Ruhe, während sich die Kiste bewegt.

Die Kiste war vor dem Hochheben auf dem Boden in Ruhe und soll sich nach dem Hochheben ebenfalls in Ruhe befinden. Es ist dabei unerheblich, ob die Person die Kiste in dieser Lage in Ruhe hält oder ob die Kiste auf eine Ablagefläche, z. B. auf einen Tisch gestellt wird. Im physikalischen Sinn verrichtet die Person keine Arbeit, wenn sie die Kiste in der Höhe h in Ruhe hält, auch wenn das die Person anders empfinden mag.

Die Kiste ist zu Beginn in Ruhe, hat also die kinetische Energie $T = 0$. Sie befindet sich im Schwerefeld der Erde, hat also eine potentielle Energie V, die wir o. B. d. A. Null setzen, wenn die Kiste auf dem Boden steht. Die Gesamtenergie der Kiste ist also zu Beginn $E = 0$. Der Person übt durch das Hochheben eine äußere Kraft $\vec{F}_\mathrm{A}$ auf unser System aus. Die an dem System geleistete Arbeit W berechnet sich nach

$$W_{0\to h} = \int_0^h mg\,\mathrm{d}z = mgh. \tag{3.68}$$

Die an dem System wirkende Kraft ist $\vec{F}_\mathrm{A} = mg\vec{e}_z$. Sie entspricht vom Betrag der Gewichtskraft $\vec{F}_\mathrm{G} = -mg\vec{e}_z$. Nach dem Hochheben verfügt die Kiste über die potentielle Energie $V = mgh$. Die mechanische Gesamtenergie hat sich also ebenfalls um diesen Betrag erhöht.

Beim Absenken der Kiste hat die Person immer noch die Kraft $\vec{F}_\mathrm{A}$ in positive z-Richtung aufzubringen, da ja immer noch die Gewichtskraft kompensiert werden muss. Berechnet man die an dem System geleistete Arbeit beim Absenken, so erhält man

$$W_{h\to 0} = \int_h^0 mg\,\mathrm{d}z = -mgh. \tag{3.69}$$

Hier muss man ein wenig auf die Vorzeichen achten. Die Bewegung $\mathrm{d}z$ läuft zwar in negative z-Richtung ab, also $\mathrm{d}\vec{z} = -\,\mathrm{d}z\vec{e}_z$, allerdings wird $\mathrm{d}z$ ebenfalls negativ, da der z-Wert am Ende kleiner als am Anfang ist, also $\mathrm{d}z = z_\mathrm{Ende} - z_\mathrm{Anfang} < 0$. Die geleistete Arbeit ist also negativ, was bedeutet, dass Energie aus dem System entnommen wurde. Das ist sinnvoll, da die Kiste am Boden ja wieder die mechanische Gesamtenergie $E = 0$ hat.

Selbst an diesem einfachen Beispiel gibt es bereits viel Raum für ausgedehnte Diskussionen. Manche Physiker könnten argumentieren, dass zum Anheben der Kiste eine geringfügig größere Kraft als mg benötigt wird, da die Kiste sonst nicht angehoben werden könnte. Das ist natürlich korrekt und entspricht unserer Alltagserfahrung. Allerdings soll die Kiste am Ende der Bewegung wieder in Ruhe sein, weshalb man beim Abbremsen eine etwas geringere Anhebekraft aufwenden muss. Der anfängliche ≫Ruck≪ beim Anheben muss also durch eine negative Beschleunigung am Ende ausgeglichen werden.

Um die Sache zu vereinfachen, betrachten wir die Bewegung idealisiert, da dies das Endergebnis nicht beeinflusst. In dieser idealisierten Vorstellung wird die Kiste in infinitesimal kleinen Schritten angehoben, sodass keine Geschwindigkeitszunahme stattfindet.

Wir erweitern unsere Betrachtung nun auf den Fall, dass die Person die Kiste auf die Höhe h gebracht hat und sie dann fallen lässt. Hier wird es interessant, denn in unserem System, bestehend aus der Kiste und der Erde (Gravitationspotential), gibt es nun keine äußeren Kräfte mehr. Wir haben bereits festgestellt, dass äußere Kräfte die mechanische Gesamtenergie eines Systems erhöhen oder senken können. Doch was passiert bei inneren (konservativen) Kräften?

Wenn unsere Annahme korrekt ist, muss die Gesamtenergie bei $z = 0$ genauso groß sein wie bei $z = h$. Wir gehen davon aus, dass die

Gravitationskraft Arbeit an der Kiste verrichtet, und unser nächster Schritt ist, das Vorzeichen dieser Arbeit zu bestimmen.

$$W_{h\to 0} = \int_h^0 mg\,\mathrm{d}z = mgh. \tag{3.70}$$

Die Arbeit, die durch die Gravitationskraft an der Kiste verrichtet wird, trägt ein positives Vorzeichen. Nach unserer Vorzeichenkonvention würde dies bedeuten, dass dem System Energie zugeführt wird. Doch das ergibt bei einer inneren Kraft keinen Sinn, da nichts von außen in das System eingebracht wird. Tatsächlich wird die potentielle Energie innerhalb des Systems in eine andere Energieform umgewandelt, nämlich in die kinetische Energie der Kiste.

Kurz bevor die Kiste den Boden erreicht, besitzt sie eine kinetische Energie T, die dem Betrag der anfänglichen potentiellen Energie entspricht. Die Gesamtenergie des Systems hat sich durch die innere Kraft weder erhöht noch verringert, sondern lediglich von einer Form in eine andere umgewandelt.

Gehen wir die Fragestellung zur Sicherheit noch von einer anderen Seite an. Wenn die Behauptung stimmt, dass innere (konservative) Kräfte keine Arbeit leisten, die Gesamtenergie des Systems also eine Konstante ist, dann kann man den Ausdruck für die Gesamtenergie nach der Zeit ableiten. Wenn diese Ableitung Null ergibt, dann sollten die richtige Bewegungsgleichung der Kiste herauskommen. Da die Gleichung der Gesamtenergie für alle z-Werte gelten muss, wird also die Höhe h allgemein durch die Variable z ersetzt. Die Geschwindigkeit ist demnach also auch eine Funktion von z, also $v = v(z) = \dot{z}$ Gehen wir es an:

$$\dot{E} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}m\dot{z}^2 + mgz\right) = \frac{1}{2}2m\ddot{z}\dot{z} + mg\dot{z} = m\dot{z}\left(\ddot{z} + g\right) = 0 \tag{3.71}$$

Den trivialen Fall $\dot{z} = 0$ können wir ignorieren, da das System dann statisch wäre (ebenso den Fall $m = 0$). Der andere Fall bedeutet, dass $\ddot{z} = -g$ sein muss. Zweimalige Integration ergibt die Bewegungsgleichung:

$$z(t) = z_0 + v_0 \cdot t - \frac{1}{2}gt^2 = h - \frac{1}{2}gt^2. \tag{3.72}$$

Wir erhalten also die bekannte Gleichung für den freien Fall unter der Annahme einer konstanten Gesamtenergie, was gleichbedeutend mit unserer Aussage ist, dass innere Kräfte keine Arbeit leisten.

In unserer bisherigen Betrachtung haben wir das System so definiert, dass sowohl die Kiste als auch die Erde bzw. ihr Gravitationsfeld Teil des Systems sind. Nun könnte man sich die Frage stellen, was passiert, wenn man die Systemgrenze so zieht, dass nur die Kiste im System enthalten

ist. Die Erde ist weiterhin in der Umgebung vorhanden und übt eine Gravitationskraft auf die Kiste aus. Der Unterschied besteht darin, dass die Gravitationskraft in diesem Fall als äußere Kraft betrachtet werden muss. Da äußere Kräfte die Gesamtenergie eines Systems verändern, muss das auch hier gelten, andernfalls wäre die Physik widersprüchlich.

Wenn wir die Kiste aus der Höhe h fallen lassen, ändert sich die Betrachtung: Bei inneren, konservativen Kräften setzt sich die Gesamtenergie aus kinetischer und potentieller Energie zusammen. In diesem Fall liegt jedoch keine innere Kraft vor, weshalb es nicht sinnvoll ist, von einer potentiellen Energie zu sprechen. Die Höhe h verliert an Bedeutung, da es in diesem Szenario keine potentielle Energie gibt. Die mechanische Energie des Systems wird ausschließlich durch die kinetische Energie bestimmt.

Was passiert, wenn die Kiste fällt, ist klar: Sie wird durch die Gravitationskraft $\vec{F}_\mathrm{G}$ beschleunigt und gewinnt kinetische Energie. Die Gesamtenergie des Systems nimmt zu (zumindest bis die Kiste auf den Boden trifft). Diese Argumentation ist widerspruchsfrei.

Viele Physiker haben hier Schwierigkeiten. Oft wird behauptet, dass immer potentielle Energie im System vorhanden ist, auch wenn die Systemgrenze nur die Kiste umschließt. Diese Argumentation stützt sich darauf, dass die potentielle Energie eine Eigenschaft des Raumes ist, der den Zentralkörper (in diesem Fall die Erde) umgibt, und dass die Gravitationskraft als konservative Kraft eine potentielle Energie impliziert. Da sich das System im Einflussbereich des Gravitationsfeldes bewegt, müsse es demnach potentielle Energie im System geben.

Diese Argumentation widerspricht jedoch dem, was wir bisher über innere und äußere Kräfte gesagt haben, und ist Unsinn. Wenn die Gravitationskraft als äußere Kraft betrachtet wird, ergibt der Begriff des Potentials im System keinen Sinn, da in einem solchen Fall die gesamte mechanische Energie erhalten bliebe. Wie wir gesehen haben, ändert sich die mechanische Energie jedoch, was bedeutet, dass in dieser Betrachtung keine potentielle Energie existiert.

Energie, wie wir eingangs festgestellt haben, ist immer eine Frage der Betrachtung – eine Zahl, die wir berechnen können oder manchmal auch nicht. Aus rechnerischer Sicht ist die Verwendung von Potentialen selbstverständlich sinnvoll, und die Argumentation in diesem Abschnitt soll keinesfalls ein Plädoyer dagegen sein, mit potentiellen Energien zu arbeiten. Im Gegenteil: Es soll lediglich verdeutlichen, dass die Physik keine Widersprüche erzeugt, wenn man die Systemgrenzen ändert.

Gehen wir unser Beispiel der fallenden Kiste nochmal an, ohne dass wir wissen, dass sich diese Kiste in einem Potentialfeld befindet. Wir stellen aus Sicht eines mit der Erdoberfläche verbundenen Inertialsys-

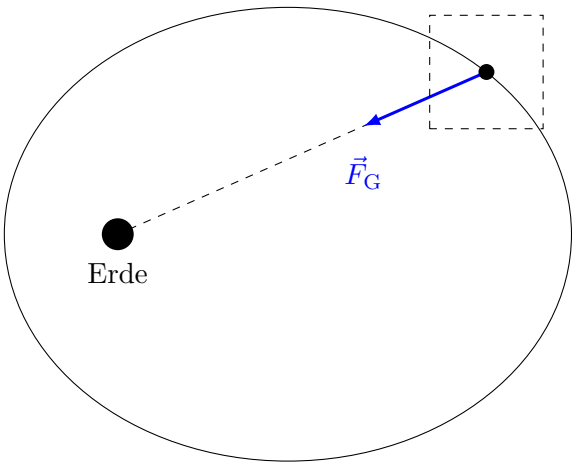


Abbildung 3.9: Die Frage der Systemgrenze (gestricheltes Quadrat) am Beispiel eines Satelliten im Erdorbit

tems lediglich fest, dass auf die Kiste eine Kraft wirkt. Wenn wir die Systemgrenze um die Kiste ziehen, dann ist diese Kraft eine externe Kraft und folglich wird sich die Gesamtenergie des Systems ändern:

$$W = \int_h^0 \vec{F} \cdot \mathrm{d}\vec{z} = \int_h^0 (-mg)\,\mathrm{d}z = mhg. \tag{3.73}$$

Die äußere Kraft $\vec{F}_\mathrm{A} = \vec{F}_\mathrm{G} = -m\vec{g} = -mg\vec{e}_z$ leistet Arbeit und erhöht die Gesamtenergie im System der Kiste (da $W > 0$). War die Kiste zu Beginn in Ruhe ($v = 0$), so hat sie nach Durchlaufen der Höhendifferenz h die Energie $E = \frac{1}{2}mv^2 = mgh$. Man erhält die gleiche Endgeschwindigkeit nach Durchlaufen der Höhendifferenz h wie im Fall der Betrachtung des Systems aus Kiste und Erde. Im einen Fall (dem mit potentieller Energie) leistet die innere (Gravitations-)Kraft keine Arbeit, die mechanische Gesamtenergie bleibt konstant, im anderen Fall (keine potentielle Energie) leistet die äußere (Gewichts-)Kraft Arbeit, die mechanische Gesamtenergie erhöht sich.

Wie sieht das nun am Beispiel eines Satelliten im Orbit aus, wie in Abbildung 3.9 dargestellt? Formal dürfte es keinen Unterschied geben, da in beiden Fällen die Gravitationskraft auf einen Massenpunkt betrachtet wird. Legen wir die Systemgrenze um den Satelliten, dann ist die Gravitationskraft der Erde auf diesen Satelliten eine äußere Kraft. Als Konsequenz dieser äußeren Kraft ist der Impuls des Satelliten keine Erhaltungsgröße. Wenn wir die Systemgrenze also nur um den Satelliten ziehen, dann ist auch die mechanische Gesamtenergie keine Erhaltungsgröße, denn der Begriff der potentiellen Energie ist unter dieser Annahme nicht sinnvoll definierbar, er würde Widersprüche produzieren. In der

Tat ist es ja auch so, dass die mechanische Gesamtenergie, die in diesem Fall nur aus der kinetischen Energie besteht, bei einem elliptischen Umlauf nicht konstant. Im Perigäum ist die Geschwindigkeit des Satelliten am größten, die Gesamtenergie entsprechend auch. Im Apogäum ist die Geschwindigkeit am geringsten. Bei einem vollständigen Umlauf um die Erde wird Netto keine Arbeit von der externen Kraft geleistet, was eine Folge davon ist, dass $\vec{F}_\mathrm{G}$ eine konservative Kraft ist. Hier könnte man nun geneigt sein, dass doch ein Potential im Raum sein muss, schließlich haben wir einige Sätze vorher gesagt, dass für die Energieerhaltung die konservativen Kräfte auch innere Kräfte sein müssen. Das ist aber wieder ein Fehlschluss, denn dieser Sachverhalt spiegelt lediglich die Tatsache wider, dass beide Argumentationen (die mit und die ohne potentielle Energie) zum gleichen Ergebnis führen, wenn der Satellit wieder an der gleichen Stelle steht - auch hier ist die physikalische Argumentation widerspruchsfrei.

Übrigens lässt sich die Behauptung, dass bei der Verwendung einer potentiellen Energie immer zwei Körper zum System gehören müssen, recht einfach einsehen, da bei Wirkung einer weiteren externen Kraft immer eine Wechselwirkung mit beiden Körpern stattfindet. Hebt man wie in unserem Beispiel die Kiste hoch, so »zieht« man immer auch an der Erde, da diese ja in gravitativer Wechselwirkung mit der Kiste steht. Das fällt natürlich durch den großen Massenunterschied zwischen Erde und Kiste nicht weiter ins Gewicht, ist aber Tatsache. Zu behaupten, dass potentielle Energie ohne potentialerzeugenden Körper nicht Teil des Systems sein soll, erscheint daher nicht sinnvoll.

Wir sind bisher von einer elliptischen Bahn ausgegangen, aber das ganze lässt sich auf beliebige Kegelschnittbahnen übertragen, so beispielsweise auch auf die Kreisbahn, die zusätzlich eine Besonderheit aufweist. Legt man die Systemgrenze wieder um den Satelliten, so wirkt zwar eine äußere Kraft, allerdings steht diese immer senkrecht zur momentanen Bewegungsrichtung, so dass hier $|\vec{F} \cdot \mathrm{d}\vec{s}| = |F| \cdot |\,\mathrm{d}s| \cdot \cos(\vec{F}, \mathrm{d}\vec{s}) = 0$ gilt. Die äußere Kraft leistet also keine Arbeit. Gleiches gilt für die Erhaltung des Bahndrehimpulses, hier allerdings nicht nur für den Kreis, sondern für alle Kegelschnitte. Es gilt der folgende Satz:

Satz von der Erhaltung des Drehimpulses

In einem System von Massenpunkten, in dem keine äußeren Drehmomente wirken, ist der Bahndrehimpuls eine Erhaltungsgröße.

Wir wir bereits an anderer Stelle gesehen haben, steht der Ortsvektor bei Zentralkräften immer kollinear zur wirkenden Kraft, und in Folge ergibt sich ein äußeres Drehmoment von Null.

4

WEITERFÜHRENDE BEISPIELE

Wir sind nun mit den wichtigsten Grundlagen fertig, die im Tutorium behandelt werden. In diesem Kapitel wird versucht, die Grundlagen durch Beispiele zu vertiefen sowie weitere Bezüge zur Physik in der Raumfahrt herzustellen.

4.1 Das dritte Keplergesetz und die Kreisbahn

Newton hat auf Grundlage des dritten Keplergesetzes seine Formel für die Gravitationskraft beweisen können. Wir wollen diesen Weg in umgekehrter Richtung gehen und uns aus Gründen mathematischer Einfachheit auf einen kreisförmigen Orbit eines Körpers der Masse m beschränken. Wir argumentieren in einem mit dem Zentralkörper der Masse M verbundenen Bezugssystem (also einem Inertialsystem in guter Näherung) und legen den Ursprung des Koordinatensystems in den Bezugspunkt. Die Masse des Zentralkörpers soll dabei deutlich größer als die Masse des kleinen Körpers sein, also $M \gg m$. Bei einer Kreisbewegung mit Radius R resultiert die Gravitationskraft in ihrer Wirkung als Zentripetalkraft:

$$m\omega^2 R = \gamma \frac{mM}{R^2}. \tag{4.1}$$

Wir kürzen die Masse m und bringen die Gleichung in die Form

$$\omega^2 R^3 = \gamma M = \text{const.} \tag{4.2}$$

Das Produkt aus Gravitationskonstante und Masse des Zentralkörpers wird als Standardgravitationsparameter $\mu = \gamma M$ bezeichnet. Der Standardgravitationsparameter ist für viele Körper im Sonnensystem genauer bekannt als seine beiden Faktoren. Das liegt daran, dass man μ durch astronomische Beobachtungen mit hoher Genauigkeit bestimmen kann, während man beispielsweise zur Bestimmung der Gravitationskonstanten sehr aufwändige Experimente (z. B. durch das Cavendish-Experiment) durchführen muss. Wir können nun ausnutzen, dass die Winkelgeschwindigkeit ω mit der Umlaufzeit über $\omega = 2\pi/T$ verknüpft ist. Wir erhalten

$$\frac{R^3}{T^2} = \text{const}, \tag{4.3}$$

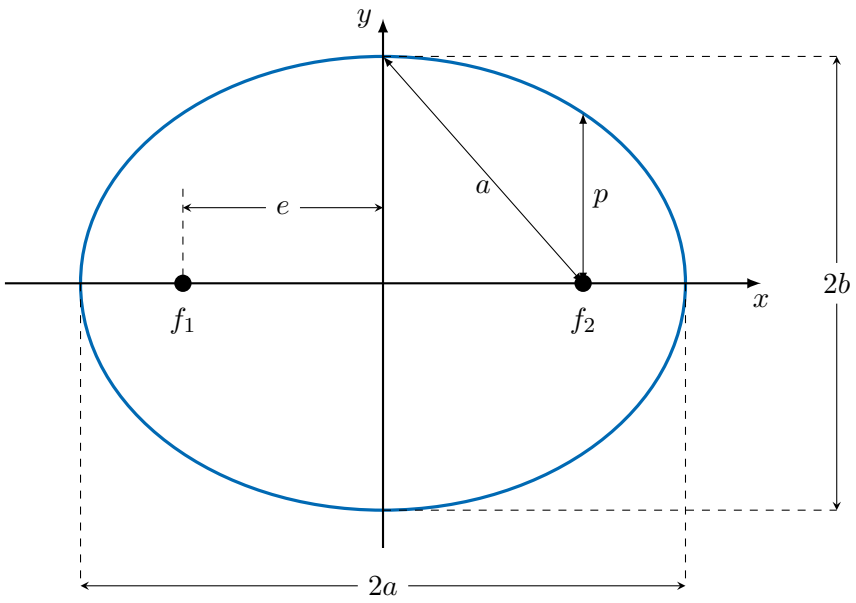


Abbildung 4.1: Ellipse mit Halbachsen a und b, Brennpunkten f_1, f_2, linearer Exzentrizität e und Halbparameter p.

also das 3. Keplersche Gesetz. Die Annahme einer Kreisbahn ist natürlich eine Näherung, die zu hinterfragen ist, denn es ist keineswegs selbstverständlich, dass dieses Ergebnis auch für elliptische Bahnen gelten muss. Kepler hat ja explizit in seinem ersten Gesetz von elliptischen Bahnformen gesprochen und Newton wird die Gültigkeit seines Gravitationsgesetzes sicherlich nicht nur für eine Kreisbahn gezeigt haben. Für den Beweis des Zusammenhangs zwischen den Kepler-Gesetzen und dem Newtonschen Gravitationsgesetz müssen wir uns etwas mehr mit der Theorie elliptischer Bahnkurven beschäftigen.

4.2 Das dritte Keplergesetz und die elliptische Bahn

Eine Ellipse kann dargestellt werden als Menge aller Punkte (x, y), die folgende Bedingung erfüllen:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \tag{4.4}$$

Hierbei stehen a und b für die große und die kleine Halbachse der Ellipse (vgl. Abbildung 4.1. Der Ursprung des Koordinatensystems liegt dabei im Mittelpunkt der Ellipse. Wie wir bereits festgestellt haben, bieten sich ebene Polarkoordinaten besser an zur Beschreibung der elliptischen Orbitalbewegung. Beim Keplerproblem ist es zudem sinnvoll,

den Ursprung von Bezugs- und Koordinatensystem in den Mittelpunkt des Zentralkörpers zu legen, also in einen der beiden Brennpunkte der Ellipse. Es gibt daneben noch alternative Darstellungen, beispielsweise durch Festlegung des Bezugspunktes in die Apo- oder Periapsis, die aber in diesem Skript nicht weiter diskutiert werden. Legen wir den Ursprung in den linken Brennpunkt (also $x' = x - e$), so erhält man:

$$\frac{(x-e)^2}{a^2} + \frac{y^2}{b^2} = 1. \tag{4.5}$$

Nun setzt man für x und y die bekannten Zusammenhänge ein:

$$\frac{(r \cdot \cos\varphi - e)^2}{a^2} + \frac{(r \cdot \sin\varphi)^2}{b^2} = 1. \tag{4.6}$$

Nach einigen Rechenschritten erhält man folgende Darstellung für $r(\varphi)$:

$$r = \frac{b^2}{a - e \cdot \cos\varphi} = \frac{p}{1 - \varepsilon\cos\varphi}. \tag{4.7}$$

Im letzten Schritt wurden die Beziehungen $p = b^2/a$ sowie $\varepsilon = e/a$ ausgenutzt. Die Größe ε wird als numerische Exzentrizität bezeichnet. Hätte man den anderen Brennpunkt als Bezugspunkt gewählt, hätte sich lediglich das Vorzeichen im Nenner vor dem Cosinus-Term geändert. Gleichung 4.7 ist nicht nur für Ellipsen gültig, sondern für alle Kegelschnitte. Die Art des Kegelschnitts ist dabei über die numerische Exzentrizität festgelegt:

- $\varepsilon = 0$: Kreis
- $0 < \varepsilon < 1$: Ellipse
- $\varepsilon = 1$: Parabel
- $\varepsilon > 1$: Hyperbel

Wir können also für Kegelschnitte folgende Beziehung angeben:

Kegelschnitte in Polarkoordinaten

$$r = \frac{p}{1 \pm \varepsilon\cos\varphi}. \tag{4.8}$$

Wir machen eine Plausibilitätsbetrachtung. Für $\varepsilon = 0$ erhält man $r = p$. Da beim Kreis $a = b = r$ gilt, folt der triviale Zusammenhang $r = r$. Nicht besonders aufregend, aber formal richtig. Für die Bedingung einer

Ellipse, also für $0 < \varepsilon < 1$, erhält man eine gebundene Bahn, da $r(\varphi)$ keine Polstellen hat, der Nenner für einen vollständigen Umlauf nicht Null werden kann. Im Fall der Parabel, also für $\varepsilon = 1$ ergibt sich eine Singularität, der Nenner wir für einen speziellen Winkel Null. Die Bahn kann also nicht geschlossen sein. Gleiches gilt für die Bedingung der Hyperbelbahn, also für $\varepsilon > 1$.

Halten wir nochmal fest, was wir bisher wissen: Die Gravitationskraft ist eine Zentralkraft, d. h. die Richtung der Kraft auf einen Körper, der sich auf einer Ellipse bewegen soll, zeigt zum Zentralkörper. Die Kraftrichtung und somit auch die Beschleunigung hat nur eine radiale Komponente. Es gilt für die radiale Komponente die Beziehung

$$a = \ddot{r} - r^2\dot{\varphi}^2. \tag{4.9}$$

Wir haben bereits gezeigt, dass die Flächengeschwindigkeit $\dot{A}$ als Folge der Erhaltung des Bahndrehimpulses eine Erhaltungsgröße ist, was bedeutet, dass $\ddot{A} = 0$ ist. Für die Flächengeschwindigkeit gilt der Zusammenhang

$$\dot{A} = \frac{1}{2}r^2\dot{\varphi} = \text{const.} \tag{4.10}$$

Wir stellen nach $\dot{\varphi}$ um und setzen in Gleichung 4.10 ein:

$$\ddot{a} = \ddot{r} - \frac{4\dot{A}^2}{r^3}. \tag{4.11}$$

Den Ausdruck für $\ddot{r}$ können wir mit Hilfe der Bahngleichung berechnen. Wir leiten zunächst einmal nach der Zeit ab:

$$\dot{r} = -\frac{\varepsilon}{p}r^2\dot{\varphi}\sin\varphi = -\frac{\varepsilon}{p}2\dot{A}\sin\varphi. \tag{4.12}$$

Diese Beziehung können wir problemlos weiter nach der Zeit ableiten:

$$\ddot{r} = -\frac{\varphi}{p}2\dot{A}\dot{\varphi}\cos\varphi. \tag{4.13}$$

Wir setzen ebenfalls in Gleichung 4.10 ein und stellen etwas um:

$$\ddot{a} = -\frac{4\dot{A}}{r^2}\left(\frac{\varepsilon}{p}\cos\varphi + \frac{1}{r}\right). \tag{4.14}$$

Wir erkennen in dem Klammerausdruck eine Variante der Ellipsengleichung in Polarkoordinaten und halten fest:

$$\ddot{a} = -\frac{\dot{A}^2}{pr^2}. \tag{4.15}$$

Die Kraft, die den Körper auf der Umlaufbahn hält, hat also die Form:

$$F = -m\frac{\dot{A}^2}{pr^2}. \tag{4.16}$$

Im Prinzip kann man aus diesem Ausdruck bereits die wesentliche Information entnehmen: das Kraftgesetz folgt einer r^{-2}-Abhängigkeit. Alle anderen Größen dieser Gleichung sind Konstanten, so dass man die Gravitationskraft in der Form $F_{\mathrm{G}} = -\mathrm{const} \cdot r^{-2}$ angeben kann. Die Konstante kann beispielsweise man durch experimentelle Beobachtungen bestimmen. Man kann aber noch ein paar weiterführende Betrachtungen durchführen, um direkt auf das bekannte Kraftgesetz zu schließen.

Die Umlaufzeit T für eine elliptische Bahn können wir berechnen, indem wir die Fläche der Ellipse, also $A_{\mathrm{E}} = \pi ab$ durch die Flächengeschwindigkeit dividieren:

$$T = \frac{\pi ab}{\dot{A}}. \tag{4.17}$$

Wir quadrieren diesen Ausdruck und teilen durch a^3:

$$\frac{T^2}{a^3} = \frac{\pi^2 b^2}{a\dot{A}^2} = \frac{\pi^2 p}{\dot{A}^2} = \xi. \tag{4.18}$$

Hierbei steht ξ für die Keplerkonstante. Es gilt also

$$\frac{\dot{A}^2}{p} = \frac{1}{\xi}, \tag{4.19}$$

und somit

$$F = -m\frac{4\pi^2}{\xi}\frac{1}{r^2}. \tag{4.20}$$

Die Keplerkonstante lässt sich bem Keplerproblem ($M \gg m$) ausdrücken als

$$\xi = \frac{4\pi^2}{\gamma M}. \tag{4.21}$$

Damit erhalten wir das bekannte Gravitationsgesetz von Newton in der Form

$$F = -\frac{\gamma mM}{r^2}. \tag{4.22}$$

Das Newtonsche Gravitationsgesetz kann also auch unter Annahme einer beliebigen Kegelschnittbahn hergeleitet werden und ist somit für alle antriebslosen Bewegungen in der Raumfahrt die maßgebliche Kraft.

4.3 Oberth-Effekt

Beim Oberth-Effekt dreht sich alles um die Fragestellung, ob es auf einem elliptischen Orbit Orte gibt, an denen es energetisch effizienter ist, ein Triebwerk anzumachen, um eine Geschwindigkeitsänderung Δv zu erzielen. Die Antwort ist ja und aus mathematischer Sicht ist das auch recht einfach zu zeigen. Sei v die Geschwindigkeit vor dem Schubmanöver

und $E_{\text{kin},2}$ die dazugehörige kinetische Energie, dann berechnet sich die kinetischen Energie $E_{\text{kin},2}$ nach dem Schubstoß zu

$$E_{\text{kin},2} = \frac{1}{2}m(v+\Delta v)^2 = \overbrace{\frac{1}{2}mv^2}^{E_{\text{kin},1}} + mv\Delta v + \frac{1}{2}m\Delta v^2. \tag{4.23}$$

Die Änderung der kinetischen Energie beträgt also

$$\Delta E_{\text{kin}} = E_{\text{kin},2} - E_{\text{kin},1} = m\Delta v\left(v + \frac{1}{2}\Delta v\right). \tag{4.24}$$

Man erkennt an dieser Gleichung, dass es zu einer größeren Energiezunahme kommt, wenn man das Schubmanöver bei höherer Geschwindigkeit ausführt. Bei einer Ellipsenbahn ist die Bahngeschwindigkeit an der Periapsis am größten, daher ist es dort am effizientesten, wenn man in diesem Bereich das Triebwerk benutzt. Bei kleiner Geschwindigkeitsänderung Δv kann man $\Delta v^2 \approx 0$ annehmen, so dass aus Gleichung 4.24

$$\Delta E_{\text{kin}} = mv\Delta v \tag{4.25}$$

wird. Teilt man Gleichung 4.25 durch das Zeitintervall Δt und bildet den Grenzwert $\Delta t \to 0$, so ergibt sich die Leistung P als

$$P = \frac{\mathrm{d}E_{\text{kin}}}{\mathrm{d}t} = mv\frac{\mathrm{d}v}{\mathrm{d}t} = mva = mv\frac{F}{m} = Fv. \tag{4.26}$$

Hier entspricht F dem Schub, der vom Triebwerk erzeugt wurde.

4.4 Lenz-Runge-Vektor

Wir haben bisher zeigen können, dass die mechanische Gesamtenergie E und der Bahndrehimpuls $\vec{L}$ beim Kepler-Problem Erhaltungsgrößen sind. Die Gesamtenergie legt dabei die Art der Bahnkurve (Kreis, Ellipse, Parabel, Hyperbel) fest, während die Erhaltung des Bahndrehimpulses ausdrückt, dass die Bewegung in einer festen Ebene abläuft. Das Keplerproblem reduziert sich also auf ein zweidimensionales Problem und kann folglich mit zwei unabhängigen Koordinaten beschrieben werden, beispielsweise durch ebene Polarkoordinaten. Der Bahndrehimpuls sagt uns nun zwar, dass die Bewegung in einer festen Ebene abläuft, aber es ist vorstellbar, dass beispielsweise eine Ellipse sich in dieser Ebene drehen könnte; der Drehimpuls wäre dabei dennoch immer gleich groß. Mit Hilfe eines weiteren Vektor, den wir als $\vec{A}$ bezeichnen, kann man zeigen, dass unter bestimmten Bedingungen die Ellipse (aber auch die anderen Kegelschnitte) in der Ebene fixiert sind.

Lenz-Runge-Vektor

Der Vektor $\vec{A} = \vec{v} \times \vec{L} + V(r)\vec{r}$ wird als verallgemeinerter, zum Zentralpotential $V(r)$ gehöriger Lenz-Runge-Vektor bezeichnet und stellt für bestimmte Potentialtypen eine Erhaltungsgröße dar.

Man kann zeigen, dass für alle Potentiale der Form $V = \lambda/r$ der Vektor $\vec{A}$ erhalten ist, d. h., dass $\mathrm{d}\vec{A}/\,\mathrm{d}t = 0$ gelten muss.

Die Bedeutung des Lenz-Runge-Vektors kann man sich mit Hilfe einiger einfacher Rechnungen klarmachen. Zunächst bilden wir das Skalarprodukt des Vektors mit sich selbst:

$$\vec{A} \cdot \vec{A} = v^2 L^2 + \frac{2\lambda}{r} \vec{L} \underbrace{(\vec{r} \times \vec{v})}_{\vec{L}/m} + \lambda^2 \tag{4.27}$$

$$= v^2 L^2 + \frac{2\lambda}{mr} L^2 + \lambda^2 \tag{4.28}$$

$$= \frac{2L^2}{m} \underbrace{\left[\frac{1}{2} m v^2 + V\right]}_{=E} + \lambda^2 \tag{4.29}$$

$$= \frac{2L^2}{m} E + \lambda^2. \tag{4.30}$$

Wir sehen, dass der Vektor $\vec{A}$ einen konstanten von Null verschiedenen Betrag hat. Es handelt sich also nicht um den Nullvektor. Die Lage des Vektors bestimmen wir über folgendes Skalarprodukt:

$$\vec{L}\vec{A} = \underbrace{\vec{L} \cdot (\vec{v} \times \vec{L})}_{=0} + \vec{L} V(r) \vec{r} \tag{4.31}$$

$$= V(r) \cdot m \cdot \vec{r} \underbrace{\left(\vec{r} \times \dot{\vec{r}}\right)}_{=0} = 0 \tag{4.32}$$

Der Lenz-Runge-Vektor liegt also in der Bahnebene. Die genau Lage in dieser Ebene kann man mit Hilfe des Skalarprodukts $\vec{A}\vec{r}$ ermitteln:

$$\vec{A}\vec{r} = (\vec{v} \times \vec{L} + V(r)\vec{r})\vec{r} \tag{4.33}$$

$$= (\vec{v} \times \vec{L})\vec{r} + V(r) r^2 \tag{4.34}$$

$$= \vec{L} \cdot (\vec{r} \times \vec{v}) + V(r) r^2 \tag{4.35}$$

$$= \frac{L^2}{m} + V(r) r^2 = \tag{4.36}$$

$$= \frac{L^2}{m} + \lambda r = A \cdot r \cdot \cos\varphi \tag{4.37}$$

Aufgelöst nach r unter Annahme eines anziehenden Kepler-Potentials ($\lambda = -|\lambda|$) erhält man

$$r = \frac{\frac{L^2}{\lambda m}}{1 + \frac{A}{\lambda}\cos\varphi}, \tag{4.38}$$

was formal der Darstellung eines Kegelschnitts in Polarkoordinaten entspricht. Durch einen Vergleich mit $r = p/(1 + \varepsilon\cos\phi)$. Daraus folgt $p = L^2/\lambda m$ und $\varepsilon = A/\lambda$. Den letzten Term kann man umschreiben:

$$\varepsilon = \frac{1}{\lambda}\sqrt{\frac{2L^2E}{m} + \lambda^2} \tag{4.39}$$

$$= \sqrt{1 + \frac{2EL^2}{m\lambda^2}} \tag{4.40}$$

Geschlossene Bahnen (Kreis, Ellipse) sind definiert über die Exzentrizität $\varepsilon < 1$. Dies ist nur dann erfüllbar, wenn $E < 0$. Für die Parabel gilt $\varepsilon = 1$. Dies ist nur erfüllbar mit $E = 0$. Analog sind Hyperbelbahnen ($\varepsilon > 1$) nur dann möglich, wenn $E > 0$.

Wir wissen also, dass $\vec{A}$ in der Kegelschnittebene liegt und dass die Bahnkurve die Form $r = p/(1 + \varepsilon \cdot \cos\varphi)$ hat. Diese Formel beschreibt die Bewegung eines Punktes auf einem Kegelschnitt vom Brennpunkt aus. Der Winkel φ ist definiert als Winkel zwischen Radius- und Lenz-Runge-Vektor. Für $\varphi = 0$ wird der Radiusvektor minimal, der Lenz-Runge-Vektor muss demnach in Richtung des Perihel zeigen.

4.5 Der harmonische Oszillator

Viele Systeme in der Physik können in einen Zustand gebracht werden, den man als Schwingung bezeichnet. Hierzu muss ein Körper, der sich in einem stabilen lokalen Potentialminimum befindet, durch eine kleine Störung ausgelenkt werden. Gemäß der Beziehung $\vec{F} = -\nabla V$ wird immer eine rücktreibende Kraft auf den Körper wirken, die ihn in Richtung des lokalen Minimums zurücktreibt. Die Situation ist in Abbildung 4.2 am Beispiel eines eindimensionalen Potentialverlaufs in x-Richtung dargestellt. Das lokale Minimum muss hinreichend tief sein, so dass der Körper bei einer kleinen Auslenkung nicht aus der Gleichgewichtslage gebracht werden kann. Es darf sich also um kein metastabiles Gleichgewicht handeln. Der Vollständigkeit halber ist auch die Position für ein instabiles Gleichgewicht dargestellt. Befindet sich ein Körper auf einem Potentialplateau, so wirkt bei Auslenkung überhaupt keine Kraft. Diese Situation wird als neutrale Gleichgewichtslage bezeichnet.

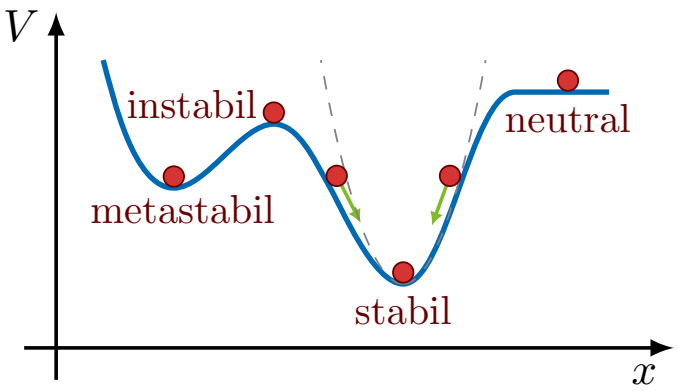


Abbildung 4.2: Gleichgewichtspositionen für verschiedene Positionen eines Körpers in einem beliebigen Potentialverlauf.

Im Falle eines harmonischen Oszillators wird eine zusätzliche Bedingung an das Potential gestellt: es muss einem parabelförmigen Verlauf haben (gestrichelte Linie in Abbildung 4.2). In diesem Fall stellt sich ein lineares Kraftgesetz ein, so dass für die rücktreibende Kraft am Beispiel unseres gezeigten Potentialverlaufs $\vec{F} = -kx$ geschrieben werden kann, wobei wir den Ursprung unseres Koordinatensystems gedanklich in die Position des Minimums legen. Die Bewegungsgleichung des Körpers ergibt sich aus dem zweiten Newtonschen Axiom:

$$m\ddot{x} = -kx \tag{4.41}$$

bzw. als:

$$\ddot{x} = -\frac{k}{m}x = -\omega^2 x. \tag{4.42}$$

Gesucht ist also eine Funktion $x(t)$, deren zweite Zeitableitung $\ddot{x}(t)$ bis auf einen Faktor ω^2 mit dieser Funktion identisch ist. Eine Funktion, die nach zweimaligem Ableiten wieder in die gleiche Funktion übergeht, ist beispielsweise die e-Funktion. Auch Sinus- und Cosinus-Funktionen erfüllen diese Bedingung. Da es sich bei der Oszillatorgleichung um eine Differentialgleichung zweiter Ordnung handelt, müssen zwei Randbedingungen im Lösungsansatz gegeben sein. Eine Darstellung des allgemeinsten Lösungsansatzes ist beispielsweise

$$x(t) = A \cdot \sin(\omega t) + B \cdot \cos(\omega t). \tag{4.43}$$

Es lässt sich leicht prüfen, dass dieser Lösungsansatz die Gleichung eines harmonischen Oszillators erfüllt.

Wie sieht es nun aus, wenn man kein parabellförmiges Potential vorliegen hat, wenn also kein lineares Kraftgesetz vorliegt. In diesem Fall ist es

hilfreich, dass Potential durch eine Taylorreihe zu approximieren. Für ein lokales Minimum verschwindet dabei der lineare Term, so dass man für kleine Auslenkungen immer ein parabolisches Potential annehmen kann. Klassisches Beispiel für diesen Fall ist das Fadenpendel: hier liegt für beliebige Auslenkungen eine rücktreibende Kraft der Form $F = -mg \sin \varphi$ vor, wobei φ die Winkelauslenkung des Fadens (mit Länge l) relativ zur Gleichgewichtslage beschreibt. Die Bewegungsgleichung hat in diesem Fall die Form

$$\ddot{\varphi} = -\frac{g}{l} \sin \varphi, \tag{4.44}$$

was man eigentlich nur numerisch lösen kann. Im Fall kleiner Winkelauslenkungen kommt allerdings die Kleinwinkelnäherung zum Tragen. Man kann in diesem Fall also die Näherung $\sin \varphi \approx \varphi$ durchführen. Dadurch wird

$$\ddot{\varphi} = -\frac{g}{l} \varphi, \tag{4.45}$$

was der Bewegungsgleichung eines harmonischen Oszillators entspricht. Wir halten fest:

Kleinwinkelnäherung

Die Kleinwinkelnäherung entspricht der Approximation eines parabelförmigen Potentialverlaufs an ein beliebiges lokales Potentialminimum.

4.6 Atwoodsche Fallmaschine

Bei einer idealen Atwoodschen Fallmaschine sind zwei beliebige Massen ($M > m$) über ein masseloses, reibungsfreies Seil miteinander verbunden. Dieses Seil läuft dabei über eine ebenfalls masselose Umlenkrolle. Die Anordnung befindet sich im Schwerefeld der Erde. Wir zeichnen zunächst alle im System vorliegenden Kräfte in die Zeichnung ein. Es wirken die Gravitationskraft $\vec{F}_1$ auf die Masse m, die Gravitationskraft $\vec{F}_2$ auf die Masse M sowie die betragsgleichen (aber entgegengerichteten) Seilkräfte $\vec{T}$ an den Befestigungen des Seils an den beiden Massen.

Für das Aufstellen der Bewegungsgleichungen der einzelnen Massen werden die Seilkräfte benötigt. Da $M > m$ gilt, wird sich die Masse m entgegen der hier nach unten zeigenden z-Achse bewegen. Es gilt:

$$ma\vec{e}_z = -mg\vec{e}_z + T\vec{e}_z \tag{4.46}$$

$$-Ma\vec{e}_z = -Mg\vec{e}_z + T\vec{e}_z \tag{4.47}$$

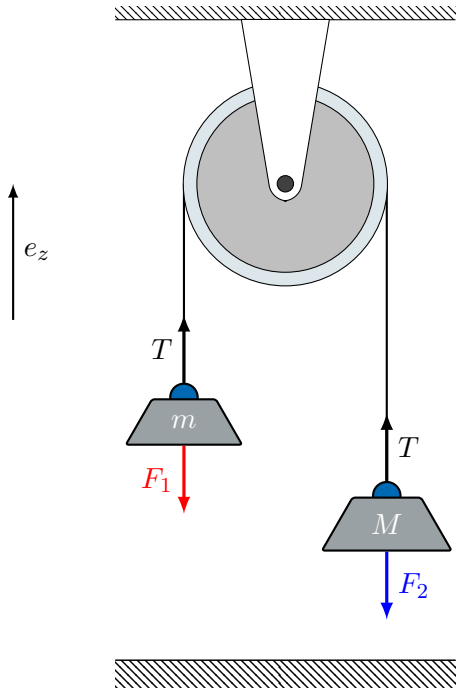


Abbildung 4.3: Die Atwoodsche Fallmaschine eignet sich gut, um mechanische Bewegungsvorgänge zu studieren.

Die beiden Massen sind über das Seil und somit über die Seilkräfte miteinander verbunden. Stellt man z. B. Gl. 4.46 nach T um, also

$$T\vec{e}_z = mg\vec{e}_z + ma\vec{e}_z, \tag{4.48}$$

und setzt dies in Gl. 4.47 ein, so erhält man mit

$$(M+m)a\vec{e}_z = (M-m)g\vec{e}_z \tag{4.49}$$

die Bewegungsgleichung des Gesamtsystems (also für beide Massen).

Die Betragsgleichheit der Seilkräfte folgt übrigens aus dem Drehimpulssatz für die Rolle mit Radius r und Massenträgheitsmoment J: $J\ddot{\varphi} = T_1 r - T_2 r$. Für die masselose Rolle gilt $J = 0$ und somit $T_1 = T_2 = T$. Man sieht an Gleichung 4.49, dass $(M+m)$ die träge Masse und $(M-m)$ die schwere Masse darstellt.

Während der Bewegung muss die Rolle das Gewicht der beiden Massen m und M kompensieren, d. h. die Belastung auf die Umlenkrolle beträgt $(M+m)\cdot g$. Wenn die große Masse M nun auf dem Boden ruht, so zieht die kleine Masse an der Umlenkrolle mit ihrem Gewicht mg. Da sich die kleine Masse ebenfalls in Ruhe befindet und das Seil gespannt ist, muss eine betragsgleiche Kraft auf der anderen Seite ziehen. Auf die Rolle wirkt demnach eine Belastung von $2mg$. Die „überschüssige“

Gewichtskraft des großen Gewichts M, d. h. $(M-m)g$, drückt auf die Unterlage und wird entsprechend von dieser kompensiert.

Aus Gleichung 4.46 erhält man durch Auflösen nach der Beschleunigung a:

$$a\vec{e}_z = -g\vec{e}_z + \frac{T}{m}\vec{e}_z. \tag{4.50}$$

Eingesetzt in Gleichung 4.47 und aufgelöst nach T ergibt:

$$T = 2\left(\frac{Mm}{M+m}\right)g \tag{4.51}$$

Für den Fall, dass die beiden Massen gleich groß sind, also $m = M$ gilt, folgt das erwartete Ergebnis, dass $T = mg$ beträgt.

Wie untersuchen nun, wie sich die Bewegungsgleichungen ändern, wenn die Fallmaschine mit der Beschleunigung a nach oben bewegt wird? Danach schlussfolgern wir aus dem Ergebnis, was passieren würde, wenn sich die Aufhängung von der Decke löst. Sei die z-Achse hier entgegen der Richtung der Erdbeschleunigung $\vec{g}$ orientiert. Würde sich die Umlenkrolle nicht drehen, so wäre nach einer Zeit t die Masse m an der Position z_1. Durch die Rollbewegung (verursacht durch die größere Masse M) kommt es zusätzlich noch zu einer relativen Verschiebung z_{rel}. Die Masse m befindet sich also an der Position $z_m = z_1 + z_{\text{rel}}$. Für die Masse M gilt entsprechend $z_M = z_2 - z_{\text{rel}}$ (siehe auch die Abbildung am Schluss dieser Aufgabe). Da die Beschleunigung a dafür sorgt, dass beide Massen die gleiche Wegstrecke zurücklegen, gilt $z_1 = z_2 = z$. Für die Beschleunigung der beiden Massen kann man folglich schreiben:

$$a_m = a + a_{\text{rel}}, \tag{4.52}$$

$$a_M = a - a_{\text{rel}}. \tag{4.53}$$

Für die Bewegungsgleichungen kann man somit schreiben:

$$ma_m = m(a + a_{\text{rel}}) = T - mg, \tag{4.54}$$

$$Ma_M = M(a - a_{\text{rel}}) = T - Mg. \tag{4.55}$$

Multiplikation von Gl. 4.54 mit M und von Gl. 4.55 mit m und Addition der beiden Gleichungen führt zu

$$T = 2\frac{mM}{m+M}(a+g). \tag{4.56}$$

Die Seilkraft erhöht sich also, wenn die Fallmaschine nach oben beschleunigt wird. Wenn sich die Aufhängung von der Decke löst, dann befindet sich das ganze System im freien Fall. Da alle Körper im freien Fall die gleiche Beschleunigung erfahren (also auch das Seil und die

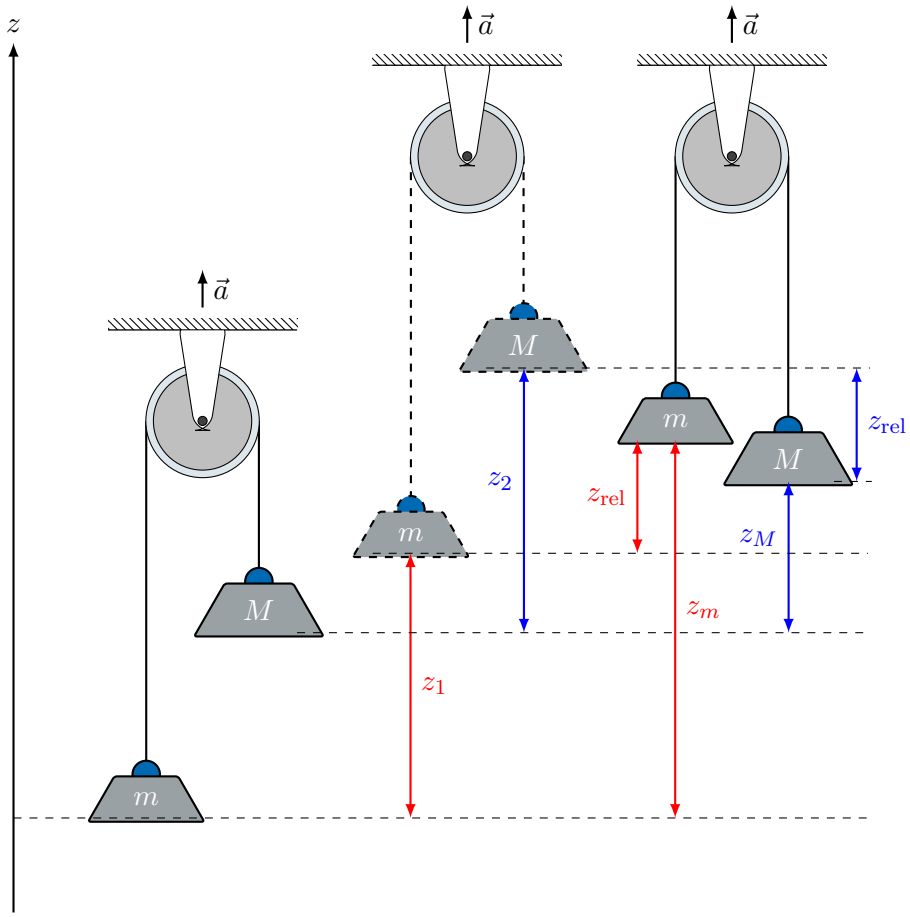


Abbildung 4.4: Die mittlere gestrichelte Darstellung entspricht der Situation, dass sich die Fallmaschine nach oben bewegt hat, die Massen sich aber relativ zur Umlenkrolle noch auf der selben Position befinden. Die rechte Darstellung berücksichtigt die Bewegung der unterschiedlichen Massen über die Rolle.

Aufhängung), gibt es keine Abrollbewegung mehr. Die relativen Positionen der beiden Massen zueinander sind fest. Als Folge dessen gilt auch für die Seilkraft: $T = 0$. Man sieht dies auch an Gl. 4.56, da im freien Fall die Beziehung $a = -g$ gilt.

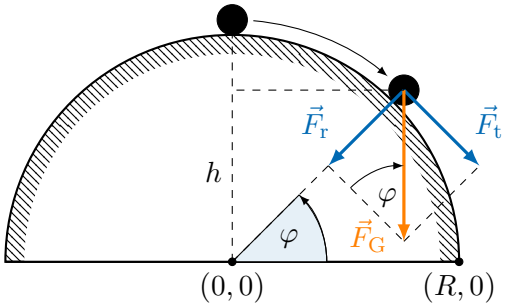
4.7 Gleitende Punktmasse auf Halbkugel

Folgende Situation sei gegeben: vom höchsten Punkt einer Halbkugel beginnt eine Punktmasse aus der Ruhe reibungsfrei im Schwerefeld der Erde ($g = \text{const.}$) hinabzugleiten. Es soll untersucht werden, bei welchem Winkel φ sich die Punktmasse von der Oberfläche der Kugel löst?

4.7.1 Verwendung des Energieerhaltungssatzes

Es ist sinnvoll, die Bewegung der Punktmasse auf der Kugelfläche in zwei zueinander senkrecht Komponenten zu zerlegen, in eine Komponente tangential und in eine Komponente vertikal zur Bewegungsrichtung. Die Bewegung in tangentialer Richtung sorgt dafür, dass die Punktmasse beschleunigt wird, sich also die Geschwindigkeit bei hinab gleitender Bewegung erhöht. Die radiale Komponente liefert uns die Information, wann die Punktmasse sich von der Kugeloberfläche löst. Solange sich die Punktmasse auf der Kugeloberfläche und damit auf einer Kreisbahn bewegt, muss – von einem Inertialsystem aus betrachtet – die Summe aller radialen Kräfte, die auf die Punktmasse wirken, der Zentripetalkraft entsprechen, die benötigt wird, damit die Bewegung auf einer Kreisbahn abläuft.

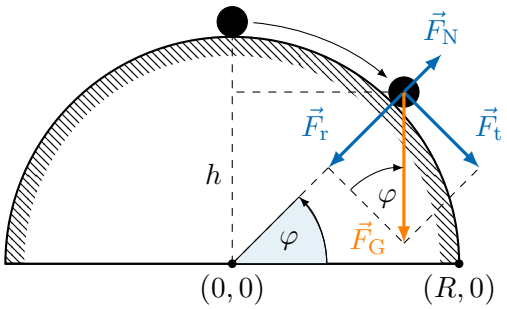
Was für Kräfte wirken nun auf die Punktmasse aus Sicht eines Inertialsystems? Zunächst natürlich die Gewichtskraft $\vec{F}_\mathrm{G}$ mit einer tangentialen Komponente $\vec{F}_\mathrm{t}$ und einer vertikalen Komponente $\vec{F}_\mathrm{r}$.



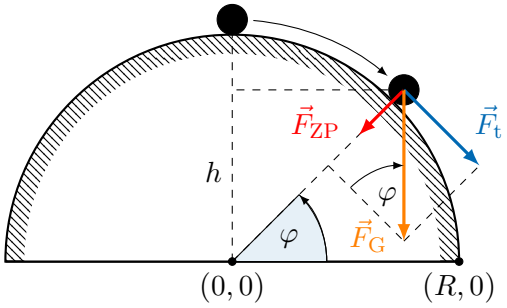
Wir müssen uns nun überlegen, was die Wirkung dieser radialen Gewichtskraftkomponente ist. Geht man zunächst einmal von der Situation aus, dass die Punktmasse an der höchsten Stelle der Kugel sich in Ruhe befindet, dann wird die Gewichtskraft, die auf die Punktmasse wirkt, von der Auflagekraft der Kugelfläche kompensiert. Wenn die Punktmasse nun herunter gleitet, dann wirkt nun die radiale Komponente der Gravitationskraft auf die Kugelfläche und es gibt eine Reaktion der Kugelfläche auf die Punktmasse, also eine Auflagekraft $\vec{F}_\mathrm{N}$. Diese Auflagekraft entspricht aber betragsmäßig nicht der vollständigen radialen Komponente der Gewichtskraft, denn für eine Bewegung auf einer Kreisbahn ist es zwingend erforderlich, dass die Summe der radialen Kräfte (in einem Inertialsystem) der Zentripetalkraft $\vec{F}_\mathrm{ZP}$ entspricht, d.h. es muss gelten:

$$\vec{F}_\mathrm{r} + \vec{F}_\mathrm{N} = \vec{F}_\mathrm{ZP} = \frac{mv^2}{R} \tag{4.57}$$

Zeichnen wir nun für die gezeigte Situation die Normalkraft ein:



Bildet man nun die resultierende Kraft in radialer Richtung, sieht die Situation folgendermaßen aus:



Die Auflagekraft wird also für größer werdende Geschwindigkeit der Punktmasse immer geringer, da eine immer größere Zentripetalkraft für die Kreisbewegung aufgebracht werden muss. Es gibt nun einen Punkt, an dem die gesamte radiale Komponente der Gewichtskraft als Zentripetal wirkt und entsprechend die Auflagekraft verschwindet. An diesem Grenzpunkt beginnt die Punktmasse die Kugeloberfläche zu verlassen. Das liegt anschaulich daran, dass die Punktmasse in diesem Grenzpunkt weiter tangential beschleunigt wird, sich also die Geschwindigkeit weiter erhöht. Für eine weitere Bewegung auf der Kugelfläche müsste also eine größere Zentripetalkraft aufgebracht werden, was aber nicht möglich ist, da man maximal nur die radiale Komponente von $\vec{F}_\mathrm{G}$ zur Verfügung hat.

Die radiale Komponente der Gewichtskraft $\vec{F}_\mathrm{r}$ ergibt sich zu

$$\vec{F}_\mathrm{r} = \vec{F}_\mathrm{G} \cdot \sin\varphi = mg \sin\varphi \, \vec{e}_\mathrm{r}. \tag{4.58}$$

Die für die Kreisbahn mit Radius R benötigte Zentripetalkraft $\vec{F}_\mathrm{ZP}$ berechnet sich zu

$$\vec{F}_\mathrm{ZP} = \frac{mv^2}{R} \, \vec{e}_\mathrm{r}. \tag{4.59}$$

Der Grenzfall ist also gegeben durch $\vec{F}_\mathrm{r} = \vec{F}_\mathrm{ZP}$, wodurch sich betragsmäßig die Beziehung

$$\cancel{m}g \sin\varphi = \frac{\cancel{m}v^2}{R} \tag{4.60}$$

ergibt. Die Geschwindigkeit lässt sich aus der Erhaltung der mechanischen Gesamtenergie berechnen. Bei $\varphi = 90°$ hat die Punktmasse, bezogen auf die durch $\varphi = 0$ definierte horizontale Ebene, die mechanische Gesamtenergie E von

$$E = E_{\text{pot}} = mgR. \tag{4.61}$$

In der Höhe h hat die Punktmasse die Gesamtenergie

$$E = E_{\text{pot}} + E_{\text{kin}} = mgh + \frac{1}{2}mv^2 = mgR. \tag{4.62}$$

Die Höhe h ist gegeben durch die Beziehung $h = R\sin\varphi$. Somit folgt:

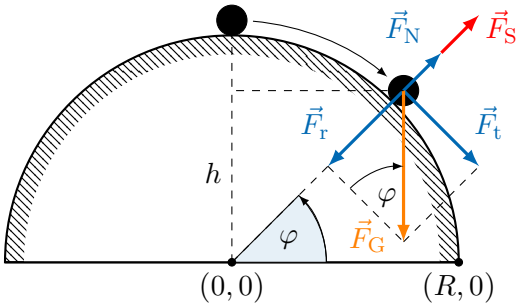
$$v^2 = 2g(h - R\sin\varphi). \tag{4.63}$$

Eingesetzt in Gl. 4.60 führt zum Ergebnis

$$\varphi = \arcsin\left(\frac{2}{3}\right) = 41.81° \tag{4.64}$$

Wie würde die ganze Situation in einem mit der Punktmasse verbundenen Bezugssystem aussehen? Hier muss die Bedingung erfüllt sein, dass die Summe aller radialen Kräfte gleich Null ergibt.[1] Das lässt sich natürlich nur erfüllen, wenn wir eine weitere Kraft einführen, die Scheinkraft $\vec{F}_{\text{S}}$, die man bei Kreisbewegungen als Zentrifugalkraft bezeichnet. Dies bedeutet also:

$$\vec{F}_{\text{r}} + \vec{F}_{\text{N}} + \vec{F}_{\text{S}} = 0 \tag{4.65}$$



1 Dies muss natürlich auch für die tangentiale Komponente gelten, aber da wir beide Komponente als linear voneinander unabhängig betrachten dürfen (Superpositionsprinzip), ist es ausreichend, wenn wir uns nur die radiale Komponente anschauen.

4.7.2 Lösung der Bewegungsgleichung

Wir haben im vorigen Abschnitt den Energieerhaltungssatz ausgenutzt. Es stellt sich daher die Frage, ob man auch ohne Energieerhaltungssatz eine Lösung des Problems finden kann. Die Antwort ist: ja, aber es ist mathematisch etwas aufwendiger und bedarf einiger kleiner »Tricks«.

Für die Bewegung auf der Kugeloberfläche mit konstantem Radius R kann man für ein kleines Kreisbogenelement $\mathrm{d}s$ schreiben:

$$\mathrm{d}s = R\,\mathrm{d}\vartheta. \tag{4.66}$$

Hierbei ist ϑ der Winkel zwischen vertikaler Achse und Ortsvektor der Punktmasse (Polarwinkel). Dieser Winkel ist hier zu bevorzugen, da man unter Verwendung von φ im weiteren Verlauf der Rechnung Probleme mit den Randbedingungen bekommt. Leitet man das Kreisbogenelement zweimal nach der Zeit t ab, so erhält man für die Beschleunigung a:

$$a = \ddot{s} = R\ddot{\vartheta}. \tag{4.67}$$

Für die Geschwindigkeitszunahme der Punktmasse ist die tangentiale Komponente $\vec{F}_\mathrm{r}$ der Gewichtskraft $\vec{F}_\mathrm{G}$ zur Kreisbewegung relevant. Für diese gilt $\vec{F}_\mathrm{r} = mg \cdot \sin\vartheta$, womit man für die Winkelbeschleunigung folgenden Ausdruck erhält:

$$\ddot{\vartheta} = \frac{g}{R}\sin\vartheta. \tag{4.68}$$

Diese Gleichung wird mit $\dot{\vartheta}$ multipliziert. Berücksichtigt man, dass

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\dot{\vartheta}^2\right) = \dot{\vartheta}\ddot{\vartheta} \tag{4.69}$$

und

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\cos\vartheta\right) = -\dot{\vartheta}\sin\vartheta \tag{4.70}$$

gilt, so folgt:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\dot{\vartheta}^2\right) = -\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{g}{R}\cos\vartheta\right) \tag{4.71}$$

bzw.

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\dot{\vartheta}^2 + \frac{g}{R}\cos\vartheta\right) = 0. \tag{4.72}$$

Der Ausdruck in der Klammer soll der Konstanten C_1 entsprechen. Mit den Randbedingungen $\vartheta(t=0) = 0$ und $\dot{\vartheta} = 0$ folgt für diese Konstante

$$C_1 = \frac{g}{R}. \tag{4.73}$$

Somit erhält man einen einfachen Zusammenhang für die Winkelgeschwindigkeit in Abhängigkeit des Polarwinkels:

$$\dot{\vartheta}^2 = \frac{2g}{R}\left(1 - \cos\vartheta\right). \tag{4.74}$$

Die Normalkraft $\vec{F}_\mathrm{N}$ im Falle der Bewegung auf der Kreisbahn ergibt sich aus der Differenz von radialer Komponente der Gewichtskraft und der für die Kreisbahn benötigten Zentripetalkraft:

$$\vec{F}_\mathrm{N} = mg\cos\vartheta - mR\dot{\vartheta}^2. \tag{4.75}$$

Ersetzt man nun $\dot{\vartheta}^2$ mit Gleichung 4.74, so folgt die Beziehung $\vartheta = \arccos\left(2/3\right) = 48.19\,^\circ$. Da $\vartheta + \varphi = 90\,^\circ$ gilt, folgt für den Winkel $\varphi = 41.81\,^\circ$, was dem Ergebnis bei Verwendung des Energieerhaltungssatzes entspricht.

4.8 Feynman und der schrumpfende Asteroid

In dem Begleitbuch zu den Feynman-Vorlesungen über Physik findet sich eine interessante Fragestellung, die hier diskutiert werden soll: ein Satellit mit der Masse m bewegt sich auf einer kreisförmigen Umlaufbahn um einen Asteroiden mit der Masse M ($M \gg m$). Was würde mit dem Satelliten passieren, wenn die Masse des Asteroiden plötzlich um die Hälfte schrumpfen würde? Wie das passieren könnte: Der Satellit befindet sich auf einer Umlaufbahn in großem Abstand zu dem Asteroiden, um einen Atomtest auf dem Asteroiden zu beobachten. Die Explosion stößt die Hälfte des Asteroiden ab, ohne dass dies direkte Auswirkungen auf den entfernten Satelliten hat. Gefragt wird, ob und wie sich die Umlaufbahn des Satelliten ändert. Nicht gegeben ist der Hinweis, dass eine energetische Betrachtung der Situation durchzuführen ist. Gebundene Bahnen sind durch eine mechanische Gesamtenergie von $E < 0$ gekennzeichnet, ungebundene Bahnen durch $E \geq 0$.

Legt man den Nullpunkt des Gravitationspotentials des Asteroiden ins Unendliche, so kann man für die mechanische Gesamtenergie (Summe aus kinetischer Energie T und potentieller Energie V) des Satelliten schreiben:

$$E = T + V = \frac{1}{2}mv^2 - \frac{\gamma mM}{r}. \tag{4.76}$$

An dieser Formel sieht man, dass für $E < 0$ der Satellit im Schwerefeld des Asteroiden gefangen ist (wir zeigen das explizit auch noch weiter unten im Text). Es liegt in diesem Fall also eine gebundene Bahn vor. Im Fall $E = 0$ hätte der Satellit so viel kinetische Energie, dass er

mit $v = 0$ im Unendlichen ankommen würde (Definition der Fluchtgeschwindigkeit), was einer Parabelbahn entspricht. Bei $E > 0$ liegt eine Hyperbelbahn vor. Parabel- und Hyperbelbahn sind also ungebundene Bahnen.

In welchem Zusammenhang stehen nun E, T und V? Da eine kreisförmige Umlaufbahn angenommen wird, wirkt eine Zentripetalkraft mv_B^2/R, wobei R hier der Radius der Kreisbahn ist. Die Geschwindigkeit v_B der Kreisbahn ergibt sich aus der Gleichheit von Zentripetalkraft und Gravitationskraft zu

$$v_\mathrm{B} = \sqrt{\frac{\gamma M}{R}}. \tag{4.77}$$

Setzt man diese Geschwindigkeit in Gleichung 4.76 ein, so folgt:

$$E = \frac{1}{2}\frac{\gamma m M}{R} - \frac{\gamma m M}{R} = -\frac{1}{2}\frac{\gamma m M}{R}. \tag{4.78}$$

Das bedeutet, dass $E = -\frac{1}{2}|V|$ und $|V| = 2T$ ist. Daraus folgt für die Gesamtenergie:

$$E = T - 2T = -T. \tag{4.79}$$

Da die kinetische Energie immer positiv ist, folgt eindeutig, dass für eine Kreisbahn eine negative Gesamtenergie vorliegen muss, die vom Betrag der kinetischen Energie entspricht, nur eben mit negativem Vorzeichen. Halbiert man nun in Gleichung 4.78 die Masse M des Asteroiden im Ausdruck für die potentielle Energie V, so folgt:

$$E = \frac{1}{2}\frac{\gamma m M}{R} - \frac{\gamma m \left(\frac{1}{2}M\right)}{R} = 0. \tag{4.80}$$

Im Ausdruck für die kinetische Energie T wird die Masse M nicht halbiert, da dieser Term aus der Gleichheit von Zentripetal- und Gravitationskraft stammt und sich die kinetische Energie des Satelliten beim Schrumpfen der Asteroidenmasse ja nicht ändert. Im Falle einer mechanischen Gesamtenergie von $E = 0$ folgt, dass sich der Satellit nun auf einer Parabelbahn befindet und nicht mehr an den Asteroiden gebunden ist.

5 BAHNMECHANIK

5.1 Einführung

In diesem Kapitel sollen wichtige Manöver der Orbitmechanik wie der Hohmann-Transfer, das elektrische Aufspiralen, das Swing-By-Manöver und andere diskutiert werden. Die benötigten physikalischen Grundlagen werden – sofern dies noch nicht in anderen Kapiteln geschehen ist – ausführlich vorgestellt.

5.2 Koordinatensysteme und Zeitangaben

Die Berechnung der Position und der Geschwindigkeit eines Himmelskörper erfordert die Angabe eines Bezugssystems (inklusive Koordinatensystem) sowie eine Festlegung der Zeitmessung bzw. -angabe. Es gibt sowohl bei der Festlegung des Bezugssystems als auch bei der Zeitmessung eine Vielzahl von Möglichkeiten. Historisch waren erdgebundene Bezugssysteme die zuerst genutzten (z. B. das Ptolemäische System), später wurden dann verstärkt sonnenzentrierte Systeme verwendet. Bei der Positionsangabe von Satelliten, die sich in einem Orbit um die Erde befinden, ist wiederum ein erdgebundenes System die naheliegende Wahl. Zeitsysteme können beispielsweise auf der Bewegung der Sonne oder auf der Rotation der Erde basieren. Aus dieser Vielzahl an Möglichkeiten sollen hier beispielhaft ein paar grundlegende Bezugs- und Zeitsysteme vorgestellt werden. Für tiefergehende Informationen sei auf die einschlägige Fachliteratur verweisen.

5.2.1 Positionsbestimmung auf der Erdoberfläche

Bevor es um die Positionsbestimmung von Himmelskörper gehen soll, ist es hilfreich, sich nochmal mit der Angabe einer Position auf der Erdoberfläche zu beschäftigen, da hier viele Analogien bestehen und der Zugang dank eines gewissen Vorwissens einfacher ist. Für die Angabe eines beliebigen Ortes auf der Erdkugel gibt man typischerweise die geographische Breite, also die Entfernung in Grad nördlich oder südlich des Äquators, und die geographische Länge, d. h. die Entfernung in

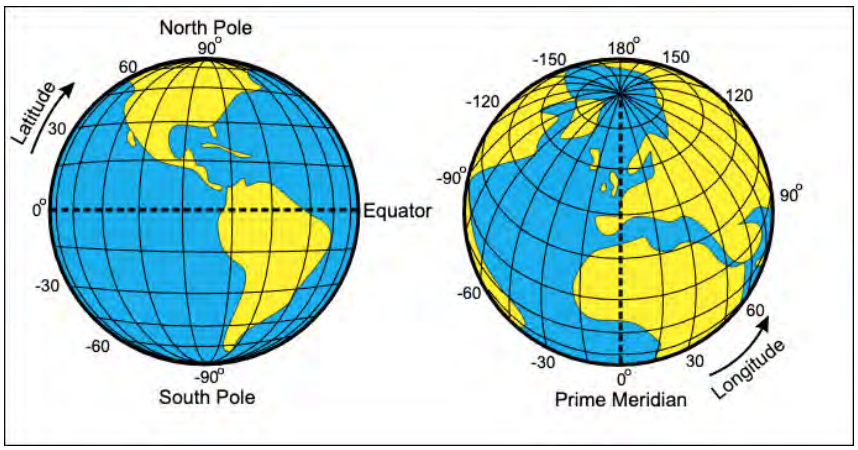


Abbildung 5.1: Definition von geographischer Breite (*Latitude*) und geographischer Länge (*Longitude*). Die Breitengrade laufen von $-90°$ bis $90°$, wobei der Äquator die Null-Linie festlegt. Die Längengrade laufen ausgehend von einem Referenzmedian von 0 bis $\pm 180°$, wobei in östlicher Richtung positiv und in westlicher Richtung negativ gezählt wird. Dieser Referenzmedian, also die Null-Linie der Längengradskala, ist seit Oktober 1884 durch den sog. Nullmeridian festgelegt, der durch Greenwich bei London verläuft.

Grad östlich oder westlich des Nullmeridians an. Der Nullmeridian ist dabei eine imaginäre Linie, die durch das Königliche Observatorium im Vereinigten Königreich verläuft und sich vom Nord- bis zum Südpol erstreckt. Er definiert die Längengradlinie von 0°, so wie der Äquator die 0°-Breitengrad-Linie festlegt. Die Festlegung ist dabei willkürlich und bis zur Internationalen Meridian-Konferenz im Oktober 1884 in Washington waren verschiedene Nullmeridiane in Benutzung. Letztlich wurde aber auf dieser Konferenz der Greenwich-Meridian als Bezugsgröße für die internationale Angabe von Längengraden beschlossen. Breitenkreise haben ihren Namen daher, dass sie alle verschiedene Durchmesser (also Breiten) haben, während Längenkreise alle den gleichen Durchmesser bzw. die gleiche Länge aufweisen.

Halten wir also fest, dass die Position auf der Erdoberfläche durch zwei sphärische Koordinaten (geographische Breite und Länge) vollständig bestimmt ist. Durch Angabe eine Höhe (typischerweise auf den Meeresspiegel referenziert) würde man auch noch einen beliebigen Punkt über (oder unter) der Erdoberfläche angeben können. Referenzpunkt für Breitenkreise ist die senkrecht zur Rotationsachse der Erde laufende Äquatorebene, für Längenkreise verwendet man den Nullmeridian. Gießen liegt beispielsweise (in Dezimalgradangabe) bei 50.58398° N und 8.67793° E. Die Angabe lässt sich einfach in andere Darstellungen wie Grad, Minuten, Sekunden usw. umrechnen.

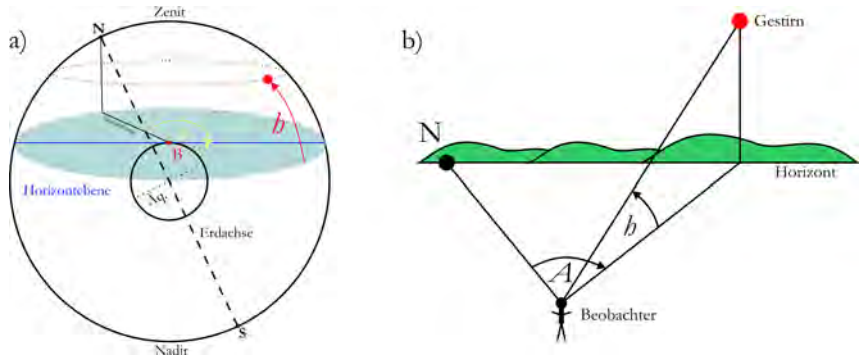


Abbildung 5.2: a) Zusammenhang zwischen Himmels- und Erdkugel. b) Angabe der Position eines Sterns über Azimutwinkel A und Höhenwinkel h.

5.2.2 Positionsbestimmung von Gestirnen

Zur Beschreibung der Position von Sternen gibt es verschiedene Bezugssysteme. Das älteste System verwendet dabei eine die Erde umgebende Himmelskugel (siehe Abb. 5.2 a), also eine gedachte Hohlkugel mit sehr großem Radius, auf dem die Sterne in irgendeiner Form angebracht sind.[1] Auch wenn die Sterne in der Realität sehr unterschiedliche Abstände zum Beobachter B haben, spielt dies für die Winkelangabe keine Rolle. Alle Sterne haben hier den gleichen fiktiven Abstand zum Beobachter.

Die Position eines Sterns auf der Himmelskugel kann nun - in Analogie zur Positionsbestimmung auf der Erdoberfläche - über die Angabe von zwei sphärischen Winkel geschehen. Sehr einfach ist die Angabe des Höhenwinkels h (auch *Elevation* genannt), der auf den Horizont des Beobachters bezogen wird, sowie die Angabe des Azimutwinkels A, der sich gegenüber dem geographischen Norden (manchmal wird auch der geographische Süden als Referenz verwendet) ergibt (siehe Abb. 5.2 b). Die Angabe der Position über diese beiden Winkel wird als Horizontsystem bezeichnet. Der Vorteil des Horizontsystems (manchmal in der Literatur auch als Horzintalsystem bezeichnet) liegt in der einfachen Positionsbestimmung über zwei intuitiv gut vorstellbare Winkel, der Nachteil liegt darin, dass verschiedene Beobachter verschiedene Winkel messen und sich die Winkelangaben durch die Rotation der Erde auch noch zeitlich ändern. Die Sterne (und auch die Sonne) bewegen sich im zeitlichen Verlauf von Osten nach Westen, also entgegengesetzt zur Rotationsrichtung der Erde in Richtung Osten.

---

1 Das Wort ≫Firmament≪ für Himmelszelt leitet sich aus dem lateinischen Wort ≫Befestigungsmittel, Stütze≪ ab, deutet also an, dass die Sterne auf der Himmelskugel fest angebracht sind.

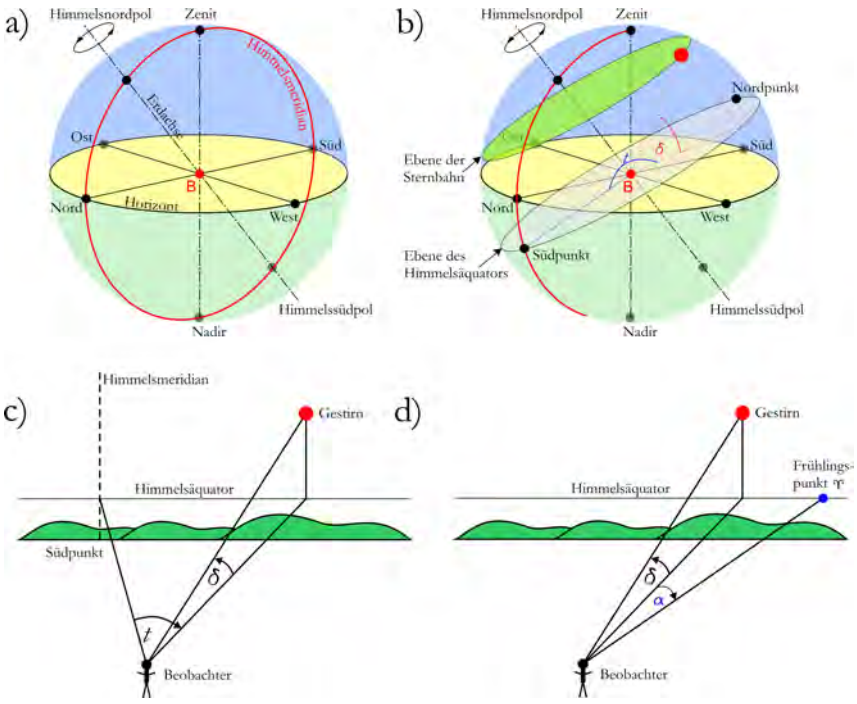


Abbildung 5.3: a) Definition des Himmelsmeridians als Meridianlinie, die durch Zenit, Himmelsnord- und Südpol, sowie den Nadir verläuft. b) Ortsfestes äquatoriales Koordinatensystem mit Deklination δ und Stundenwinkel t. c) gleiches Koordinatensystem wie in (b) in anderer Darstellung d) Rotierendes äquatoriales Koordinatensystem unter Verwendung des Frühlingspunktes ♈

Um nun zu einer Darstellung zu gelangen, die ganz oder teilweise unabhängig von der Erdrotation ist, verwendet man die sog. äquatorialen Koordinatensysteme, wobei hier noch zusätzlich unterschieden wird, ob die Erde rotiert oder ortsfest ist. Beim ortsfesten Äquatorialsystem sind die Bezugspunkte die Höhe δ über dem Himmelsäquator, die man als Deklination bzw. himmliche Breite bezeichnet sowie der Stundenwinkel t, der vom Himmelsmeridian aus gemessen wird. Den Himmelsäquator erhält man als Extrapolation des Erdäquators an die Himmelskugel, den Himmelsmeridian als Verbindungslinie von Zenit, Nadir, Nord- und Südpol (bezogen auf die Himmelskugel) sowie Nord- und Südpunkt (siehe Abb. 5.3 a–c).

Den Himmelsnordpol bestimmt man in der Regel über den Nordstern (auch als Polarstern bekannt), da dieser ungefähr am nördlichen Schnittpunkt der Erdachse mit der Himmelskugel liegt. Er liegt im Sternbild *Kleiner Bär* (manchmal auch als *Kleiner Wagen* bezeichnet). Der Begriff Polarstern ist ein übergeordneter Begriff für einen Stern, der gerade am oder in unmittelbarer Nähe vom Himmelsnordpol liegt. Durch

die Präzession der Rotationsachse der Erde mit einer Periodendauer von etwa 25 800 Jahren (das Platonische Jahr) zeigt die Erdachse irgendwann auf einen völlig anderen Stern in einem anderen Sternbild, der dann als Polarstern dient. Zudem gibt es Polarsterne für die nördliche und für die südliche Hemisphäre.

Der Stundenwinkel t hat seinen Namen daher, dass er für Zeitmessungen verwendet werden kann, weswegen er manchmal auch im Zeitmaß (statt im Gradmaß) angegeben wird. Eine Stunde entspricht dabei einer Winkeländerung von 15°, was einer vollständigen Drehung um 360° in 24 Stunden entspricht. Der Nachteil der Angabe einer Sternenposition über den Stundenwinkel ist seine Ortsabhängigkeit aufgrund der Erdrotation. Um hier zu einer rotationsunabhängigen Beschreibung zu gelangen, muss man einen mitrotierenden Bezugspunkt auf dem Himmelsäquator festlegen. Hier hat sich der sog. Frühlingspunkt (auch bekannt als Widderpunkt bzw. vernal equinox) etabliert, der dem frühjährlichen Schnittpunkt des Himmelsäquators mit der Ekliptik (das ist die Bewegung der Sonne auf der Himmelskugel) entspricht. Den über den Frühlingspunkt definierten Polarwinkel bezeichnet man als Rektaszension α (siehe Abb. 5.3 d). Die Rektaszension kann als Analogon zur Angabe der geographischen Breite auf der Himmelskugel interpretiert werden.

5.2.3 Siderischer und synodischer Umlauf

Als synodische Periode (*synodos*; griechisch für Versammlung) bezeichnet man die Zeitspanne, die benötigt wird, bis ein Himmelskörper von der Erde aus gesehen wieder genau an der gleichen Stelle steht. Im Unterschied bezeichnet die siderische Periode die Zeitspanne, die ein Himmelskörper für einen vollständigen Umlauf benötigt, also einen Winkel von 2π zurücklegt. Als Referenz für die siderische Periode wird üblicherweise der Fixsternhintergrund verwendet (*sidus*; lateinisch für Stern). Die Situation ist in Abbildung 5.4 am Beispiel der Rotation des Mondes um die Erde abgebildet. Zu Beginn stehen Sonne, Erde und Mond auf einer Verbindungslinie bzw. der Mond steht in Opposition zur Sonne von der Erde aus gesehen. Nach der siderischen Zeitspanne T_{sid} hat der Mond sich einmal um die Erde gedreht, die Erde hat sich in dieser Zeit um den Winkel α um die Sonne gedreht. Von der Erde aus gesehen steht der Mond nach T_{sid} nicht an der gleichen Position, es fehlt ihm genau der Winkel α, damit er auf der Verbindungslinie Sonne–Erde steht. Man kann hier zunächst festhalten, dass für die

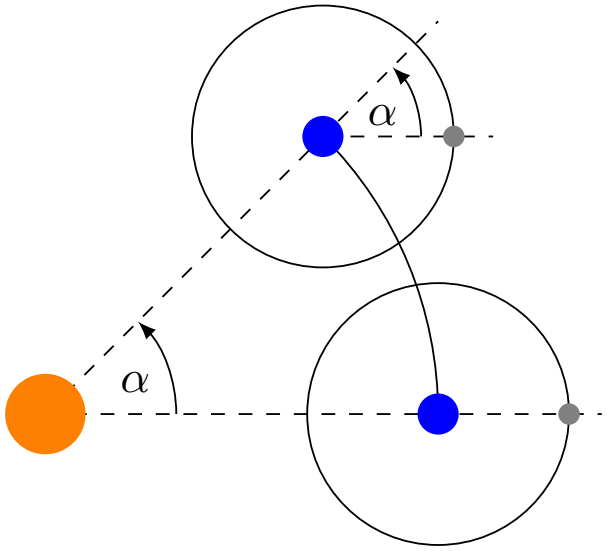


Abbildung 5.4: Synodischer und siderischer Umlauf am Beispiel der Rotation des Mondes um die Erde.

synodische Zeitspanne des Mondes $T_{\text{syn}} > T_{\text{sid}}$ gelten muss. Wir können sogar genau festhalten, dass

$$\frac{2\pi}{T_{\text{sid}}} = \frac{2\pi + \alpha}{T_{\text{syn}}} = \frac{2\pi}{T_{\text{syn}}} + \frac{\alpha}{T_{\text{syn}}} \tag{5.1}$$

gelten muss, da der Mond sich ja die ganze Zeit mit der gleichen Winkelgeschwindigkeit dreht. Der Winkel α kann über die Winkelgeschwindigkeit der Drehung der Erde um die Sonne ($1\,\text{a} = 1$ siderisches Jahr) ausgedrückt werden:

$$\frac{\alpha}{T_{\text{syn}}} = \frac{2\pi}{1\text{a}}, \tag{5.2}$$

wodurch

$$\frac{1}{T_{\text{sid}}} = \frac{1}{T_{\text{syn}}} + \frac{1}{1\text{a}} \tag{5.3}$$

folgt. Der siderische Monat beträgt 27,322 Tage, das siderische Jahr 365,256 Tage. Mit Gleichung 5.3 errechnet man damit einen synodischen Monat von 29,53 Tagen bzw. von 29 Tagen, 12 Stunden, 44 Minuten und 2,9 Sekunden.

5.2.4 Zeit- und Datumsformate

Wie bereits erwähnt verschiebt sich wegen der Präzession der Erde die Position des Frühlingspunktes bezogen auf den Sternenhintergrund

mit einer Periodendauer von 25 800 Jahren. Diese Verschiebung macht pro Jahr etwa 50 Bogensekunden aus, also 0.7° in 50 Jahren. Bei der erstmaligen Festlegung des Frühlingspunktes vor etwa 2000 Jahren stand die Sonne im Sternbild Widder, weswegen man auch vom Widderpunkt spricht und dieser Punkt das Symbol ♈ eines Widderkopfes hat. Mittlerweile steht die Sonne bei Frühlingsbeginn im Sternbild Fische. Um eine Vergleichbarkeit von Positionsangaben zu ermöglichen, werden Angaben von Himmelskoordinaten auf sogenannte Epochen bezogen. Eine Epoche entspricht einer Momentaufnahme des Fixsternhimmels zur Festlegung von Deklination und Rektaszension. Epochen werden in Übereinkunft der International Association of Geodesy (IAG) und der Internationalen Union für Geodäsie und Geophysik (IUGG) seit 1984 in Zeitabständen von 50 Jahren und über Angabe des Julianischen Datums (JD) festgelegt. Zur Zeit befinden wir uns in der Epoche J2000.0 (J steht für Julianisch), der »Schnappschuss« ist auf den 1. Januar 2000 11:59 UTC datiert. Das entsprechende Julianische Datum ist 2 451 450.0. Das Julianische Datum ist eine Zeitangabe, die die Tage seit dem 1. Januar 4713 v. Chr. zählt, wobei die Nachkommastelle den Tagesbruchteil angibt (also z. B. 0.5 für einen halben Tag). Da sich hier recht große Zahlen ergeben, verwendet man alternativ das Modifizierte Julianische Datum (MJD), welches über

$$\mathrm{MJD} = \mathrm{JD} - 2400000.5 \tag{5.4}$$

definiert ist. Der Nullpunkt liegt damit am 17. November 1858 um 00:00 UTC. Der folgende Code-Schnipsel berechnet aus der Angabe von Jahr (Y), Monat (M), Tag (D), Stunde (H), Minute (Min) und Sekunde (S) das modifizierte julianische Datum. Umgesetzt ist die Berechnung innerhalb einer Funktion, die Programmiersprache ist Python. Damit die Funktion ausführbar ist, muss die Bibliothek NUMPY eingebunden werden.

```
# calculates modified Julian day from "normal" date
def calc_MJD(Y, M, D, H, Min, S):
    if M > 2:
        Y = Y
    else:
        Y = Y - 1
        M = M + 12
    D = D + H/24 + Min/1440 + S/86400
    JD = np.floor(365.25*(Y + 4716)) + np.floor(30.6001*(M
    +1)) + D + 2 - np.floor(Y/100) + np.floor(Y/400) -
    1524.5
    MJDate = JD - 2400000.5
    return MJDate
```

Unter dem Begriff UTC verbirgt sich die koordinierte Weltzeit (coordinated universal time). Bezugspunkt der UTC ist der Nullmeridian, die Zeitmessung basiert auf der Sekunde des Internationalen Einheitensystems (SI), man spricht hier auch von der ≫Atomsekunde≪. UTC sollte nicht mit der Greenwich Mean Time (GMT) verwechselt werden, kann aber im Alltag damit gleichgesetzt werden. GMT bezieht sich auf den mittleren Sonnentag und weicht daher von der UTC ab, diese Abweichungen werden durch Schaltsekunden aber ausgeglichen. Laut der Physikalischen Technischen Bundesanstalt wird darauf geachtet, dass der Unterschied zwischen UTC und GMT immer kleiner als 0.9 Sekunden bleibt. Aus der UTC abgeleitet sind die Begriffe Central European Time (UTC+01:00) und Central European Summer Time (UTC+02:00). UTC ist eine gebräuchliche Zeitangabe in der Luft- und Raumfahrt. Ausgehend von der UTC definiert man 24 Zeitzonen, die jeweils 15° breit sind und eine Stunde Zeitunterschied zueinander aufweisen, wobei man hier auch noch diverse Sonderfälle (z. B. nicht ganzstündige Zeitzonen oder große Staaten, die nur wenige Zeitzonen intern haben wollen) berücksichtigen muss (siehe Abb. **??**). Neben der UTC gibt es auch noch die Universal Time (UT), die im Unterschied zur UTC auch Schwankungen der Erdrotation (beispielsweise durch Gezeitenkräfte) einbezieht. Da Raumfahrt neben der zivilen Nutzung auch eine sicherheitspolitische Komponente und daher eine Verbindung zur rüstungsnahen Industrie hat, kann es auch vorkommen, dass Datums- und Zeitangaben im Datum/Zeit-Format (*date-time group*, DTG) der NATO verwendet werden. Hier wird mit einer Folge aus sechs Ziffern, vier Buchstaben und zwei weiteren Ziffern gearbeitet. So steht

$$251600\text{Bapr}24 \tag{5.5}$$

für den ersten Veranstaltungstermin im Sommersemester 2024 (25. Tag des Monats, 16:00 Uhr, Zeitzone B (deutsche Sommerzeit, ansonsten A), April im Jahr 2024). Der inernationale Standard für numerische Datums- und Zeitangaben ist in der ISO 8601 niedergeschrieben. Für den eben genannten Zeitpunkt in DTG folgt in ISO-Schreibweise

$$2024\text{-}04\text{-}25\text{T}16{:}00{+}02{:}00. \tag{5.6}$$

5.2.5 Die klassischen Orbitalelemente

Für die Angabe der Position und der Geschwindigkeit eines Himmelskörpers im Orbit gibt es verschiedene Möglichkeiten. Naheliegend ist zunächst die Angabe kartesischer Vektoren $\vec{r}$ und $\vec{v}$, allerdings kann man sich darunter schwer die konkrete Form und Lage des Orbits

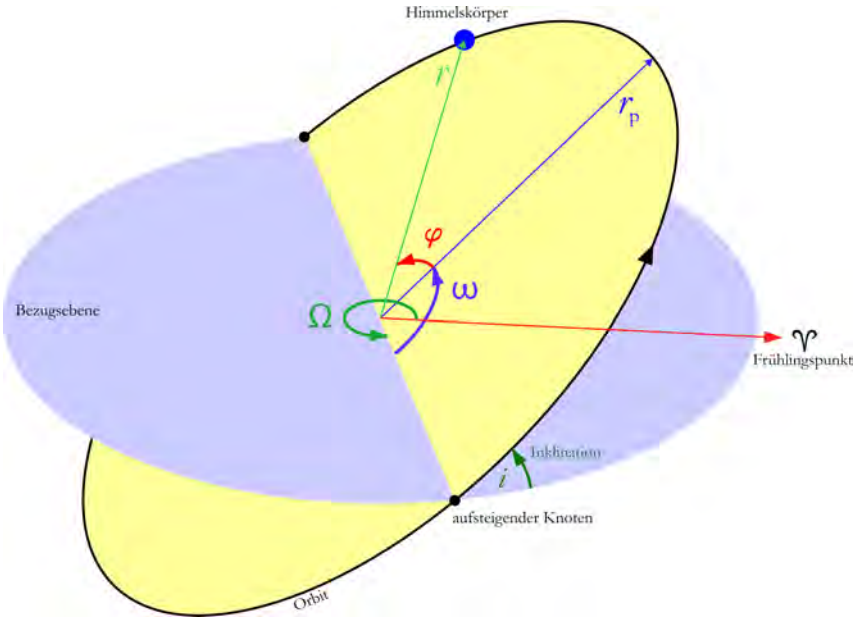


Abbildung 5.5: Zur Definition der sechs klassischen Orbitelemente: große Halbachse a, Exzentrizität ε, Inklination i, Rektaszension des aufsteigenden Knotens Ω, das Argument des Perizentrums ω sowie die wahre Anomalie φ.

im Raum vorstellen. Für einen Orbit beliebiger Form als Lösung des Keplerproblems haben wir bereits die Parameterdarstellung

$$r = \frac{a(1-\varepsilon^2)}{1+\varepsilon\cos\varphi}. \tag{5.7}$$

kennengelernt. Je nach Wert der Exzentrizität ε und der großen Halbachse a können wir eine konkrete Form des Orbits als Funktion des Winkels φ angeben, wissen also, wie die Form des Orbits in einer bestimmten Ebene aussieht. Der Winkel φ bei einer elliptischen Bahn wird in der Orbitmechanik als wahre Anomalie bezeichnet. Generell beschreibt der Begriff ≫Anomalie≪ den momentanen Winkel eines Himmelskörpers bezogen auf die Periapsis des Kegelschnitts. Die Größen ε und a bestimmen also die Form des elliptischen Orbits. Für die Lage des Orbits im Raum werden drei weitere Größen verwendet, die Inklination i, die Rektaszension des aufsteigenden Knotens Ω sowie das Argument des Perizentrums (also der Periapsis) ω. Über die wahre Anomalie $\varphi(t)$, also über den Winkel zwischen Periapsis und dem Ortsvektor der aktuellen Position des Satelliten auf der elliptischen Bahn ist die Position eindeutig festgelegt. Die Zeit ist dabei implizit in der wahren Anomalie enthalten. Die sechs klassischen Orbitelemente sind in Abb. 5.5 dargestellt und werden im einzelnen nun besprochen. Das

Bezugssystem ist im folgenden nun ein äquatoriales, rotierendes und kartesisches Bezugssystem mit Koordinaten (X, Y, Z) und entsprechenden Einheitsvektoren $\vec{e}_{x,y,z}$.

INKLINATION Die Inklination i, die der Neigung der Ebene der Satellitenbahn zur Bezugsebene entspricht (also dem Winkel zwischen dem Einheitsvektor $\vec{e}_z$ und dem Drehimpulsvektor), lässt sich am einfachsten über den massenspezifischen Drehimpuls $\vec{h} = \vec{L}/m$ berechnen. Sind Ortsvektor $\vec{r}$ und Geschwindigkeit $\vec{v}$ im Referenzkoordinatensystem (X, Y, Z) gegeben (siehe Abb. 5.5), so folgt mit $\vec{h} = \vec{r} \times \vec{v}$ für die Inklination der Ausdruck

$$i = \arccos\left(\frac{h_z}{h}\right). \tag{5.8}$$

Über die Inklination kann die Form des Orbits bezogen auf die Äquatorebene der Erde weiter spezifiziert werden. Bei einer Inklination von 0 bzw. 180° spricht man von einem äquatorialen, bei 90° von einem polaren Orbit. Bei Inklinationen von $0° \leq i < 90°$ liegt ein prograder Orbit (der Satellit bewegt sich in Richtung der Erdrotation) vor, bei $90° < i \leq 180°$ ein retrograder Orbit (Bewegung entgegen der Erdrotation).

REKTASZENSION DES AUFSTEIGENDEN KNOTENS Ω Bezugspunkt dieses Winkels (im englischen auch *Right Ascension of Ascending Node*, RAAN, genannt) ist die zum Frühlingspunkt zeigende X-Achse. Gemessen wird zur Verbindungslinie zwischen Erdmittelpunkt und dem Punkt, an dem der Satellit die Bezugsebene (X-Y-Ebene) passiert. Diese Verbindungslinie können wir durch den Vektor $\vec{n} = \vec{h} \times \vec{e}_z$ beschreiben. Daraus folgt sofort

$$\Omega = \arccos\left(\frac{n_x}{n}\right). \tag{5.9}$$

Die Rektaszension des aufsteigenden Knotens kann Werte zwischen 0 und 360° annehmen. Da der Kosinus eine gerade Funktion ist, hat man nun das Problem, dass Gl. 5.9 nicht unterscheiden kann, wenn Winkel $\geq 180°$ vorliegen. Man sieht das bereits sehr gut bei 180°, da hier $\cos(\Omega) = (-n_x/n) = (n_x/n)$ gilt. Daher muss eine Fallunterscheidung durchgeführt werden: Winkel, die größer als 180° sind, liegen vor, wenn die Y-Komponente von $\vec{n}$ negativ wird. In diesem Fall berechnet man die Rektaszension nach

$$\Omega = 360° - \arccos\left(\frac{n_x}{n}\right), \tag{5.10}$$

in allen anderen Fällen nach Gl. 5.9. Die Rektaszension des aufsteigenden Knotens ist unbestimmt für äquatoriale Orbits, also für $i = 0$ oder $i = 180°$.

EXZENTRIZITÄT e Hier bedient man sich des sog. Exzentrizitätsvektors $\vec{e}$, der aus dem Lenz-Runge-Vektor abgeleitet wird.[2] Mit dem Standardgravitationsparameter μ und dem Einheitsvektor des Ortsvektors $\vec{e}_r$ berechnet er sich nach

$$\vec{e} = \frac{\vec{v} \times \vec{h}}{\mu} - \vec{e}_r. \tag{5.11}$$

Der Betrag dieses Vektors entspricht der gesuchten Exzentrizität.

GROSSE HALBACHSE a Hat man die Exzentrizität e berechnet, ergibt sich die große Halbachse aus

$$a = \frac{h^2/\mu}{1 - e^2}. \tag{5.12}$$

ARGUMENT DES PERIZENTRUMS ω Diese Größe gibt die Orientierung der Periapsis relativ zur Rektaszension des aufsteigenden Knotens an und berechnet sich aus dem Skalarprodukt des Exzentrizitätsvektors $\vec{e}$ und des Rektaszensionsvektors $\vec{n}$ zu

$$\omega = \arccos\left(\frac{\vec{n} \cdot \vec{e}}{n \cdot e}\right). \tag{5.13}$$

Der Winkel ω kann zwischen 0 und 360° liegen, daher muss man auch hier wegen des Kosinus eine Fallunterscheidung durchführen: da $\omega > 180°$ für negative z-Komponenten von $\vec{e}$ vorliegt, gilt in diesem Fall

$$\omega = 360° - \arccos\left(\frac{\vec{n} \cdot \vec{e}}{n \cdot e}\right). \tag{5.14}$$

Zu beachten ist, dass wie bei der Inklination für $i = 0$ bzw. $i = 180°$ und zusätzlich auch für $e = 0$, also für kreisförmige Bahnen, kein Argument des Perigäums angegeben werden kann.

WAHRE ANOMALIE φ Die bisher definierten Orbitelemente haben die Form und Lage der Ellipse relativ zu unserem geozentrischen äquatorialen Koordinatensystem festgelegt. Die wahre Anomalie φ gibt nun die genaue Position des Satelliten auf dem Orbit an. Es handelt sich dabei um den Winkel zwischen dem Ortsvektor $\vec{r}$ und dem Exzentrizitätsvektor $\vec{e}$, womit sich der gesuchte Winkel durch

$$\varphi = \arccos\left(\frac{\vec{e} \cdot \vec{r}}{e \cdot r}\right). \tag{5.15}$$

2 Verwechseln Sie diesen Vektor nicht mit den Einheitsvektoren. e_x ist die x-Komponente des Exzentrizitätsvektors, $\vec{e}_x$ der Einheitsvektor in x-Richtung.

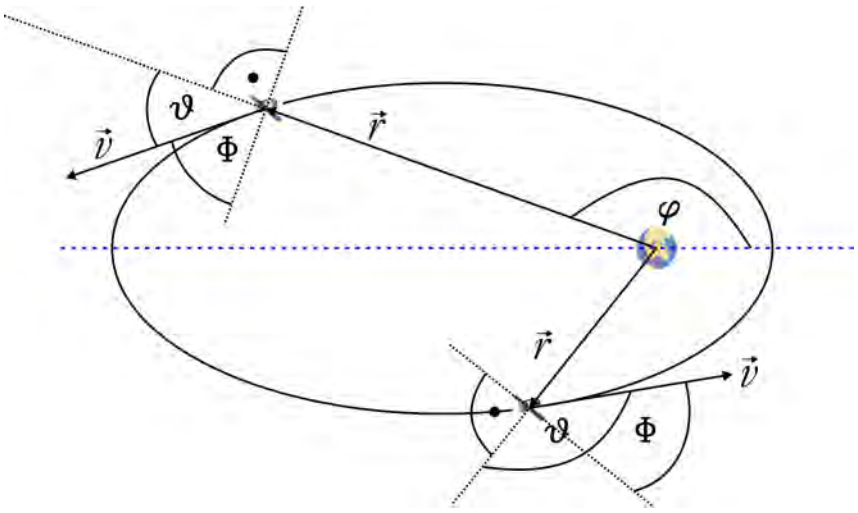


Abbildung 5.6: Fallunterscheidung zur Bestimmung der wahren Anomalie φ.

berechnen lässt. Da die wahre Anomalie zwischen 0 und 360° liegt, muss wieder eine Fallunterscheidung durchgeführt werden, die sich etwas von der bisherigen Methodik unterscheidet. Die Situation ist in Abb. 5.6 dargestellt. Es geht letztlich darum, wie man festlegen kann, ob sich der Satellit ober- oder unterhalb der Ebene befindet, die im Schnittbild durch die blaue gestrichelte Linie definiert ist. Hierzu bedient man sich des Winkels ϑ, der über das Skalarprodukt von Orts- und Geschwindigkeitsvektor definiert ist. Es gilt

$$\vec{r} \cdot \vec{v} = rv \cos(\vartheta). \tag{5.16}$$

Mit der Winkelbeziehung $\phi = 90° - \vartheta$, folgt daraus

$$\vec{r} \cdot \vec{v} = rv \sin(\phi). \tag{5.17}$$

Für eine Bewegung von der Peri- zur Apoapsis ist ϕ positiv, somit auch das Skalarprodukt $\vec{r}\vec{v}$, für eine Bewegung von der Apo- zur Periapsis nimmt ϕ negative Werte an (da $\vartheta > 90°$ wird), entsprechend ist auch das Skalarprodukt der beiden Vektoren negativ. Somit kann man also über das Vorzeichen des Skalarproduktes von Orts- und Geschwindigkeitsvektor eine Fallunterscheidung durchführen. Dies bedeutet, dass

$$\vec{r} \cdot \vec{v} < 0 \Leftrightarrow 180° < \varphi < 360°. \tag{5.18}$$

Mit den definierten Bahnelementen ist eine eindeutige Beschreibung der Bahnform und Position des Satelliten für ein beliebiges zu einer bestimmten Zeit t gegebenes Vektorpaar $\vec{r}$ und $\vec{v}$ möglich. Es sei an dieser Stelle erwähnt, dass es auch noch andere Orbitelemente gibt, die man zur Beschreibung nutzen könnte, die aber im Rahmen des Tutoriums keine Rolle spielen werden.

5.3 Das Swing-By-Manöver

Das Swing-By-Manöver (auch als *Gravity Assist* bekannt) ist eine Technik, bei der ein Raumfahrzeug seine Bahnenergie und seinen Drehimpuls durch eine Annäherung an einen Himmelskörper verändert. Es ermöglicht eine Verringerung des Treibstoffverbrauchs und der Flugzeit. Mehrere interplanetare Missionen wie die berühmten Voyager-, Mariner- oder Galileo-Missionen haben dieses Verfahren angewandt und wurden dadurch auch überhaupt erst ermöglicht. Aus der Beobachtung der Änderung von Kometenbahnen nach der Annäherung an Jupiter war in der Astronomie das Phänomen der Energie- und Drehimpulsänderung schon seit dem späten 19. Jahrhundert bekannt. In der Raumfahrt hat sich der Swing-By in den 1960er Jahren etabliert, interessanterweise durch die Berechnungen eines jungen Studenten namens Michael Minovitch, der durch einen Semesterferien-Job Zugriff auf eine damalige Großrechenanlage an der *University of California, UCLA* (Los Angeles, USA) hatte und diesen Zugriff für seine Kalkulationen nutzte, er hat im Prinzip numerisch das Dreikörperproblem gelöst. Erste schriftliche Erwähnungen dieses Manövers stammen aus den 1920er Jahren vom russischen Wissenschaftler Friedrich Zander. Minovitch konnte zeigen, dass in der Mitter der 1970er Jahre eine seltene Planetenkonstellation von Jupiter, Saturn, Uranus und Neptun, die nur etwa alle 176 Jahre auftritt, die man gezielt für eine Reihe von Swing-By-Manövern nutzen könnte, um in nur 12 Jahre alle diese Riesenplaneten anfliegen zu können. Ohne Swing-By hätte man hierfür etwa 30 Jahre gebraucht. Er nutzte hierzu den damals schnellsten Computer, der verfügbar war, einen IBM 7090.[3] Das Swing-By-Manöver wurde erstmals im Dezember 1973 eingesetzt, um die heliozentrische Geschwindigkeit der Raumsonde Pioneer 10 beim Vorbeiflug am Jupiter zu erhöhen, damit sie aus dem Sonnensystem entkommen konnte. James van Allen[4] stellte 2003 in seinem Artikel ≫Gravitational assist in celestial mechanics-a tutorial≪ im American Journal of Physics fest, dass während der Begegnung die heliozentrische Geschwindigkeit von Pioneer 10 von 9.8 km/s auf 22.4 km/s erhöht wurde, wodurch sich die kinetische Energie um den Faktor 5.2 erhöhte und die ursprüngliche elliptische Bahn in eine hyperbolische Bahn umgewandelt wurde. Während der Wechselwirkung verringerte sich die

3 Der IBM 7090 wurde damals auch verwendet, um die Mondflüge des Apollo-Programms zu simulieren. Seine Rechenleistung lag bei etwa 230000 Rechenoperationen pro Sekunde, also bei einer Taktfrequenz von 0.23 MHz, was aus heutiger Sicht nicht wirklich viel ist.

4 Nach van Allen sind auch die Strahlungsgürtel der Erde benannt.

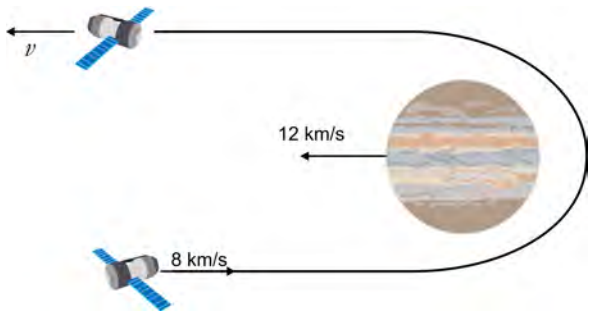


Abbildung 5.7: Das Swing-By-Manöver als elastischer Stoß.

Geschwindigkeit des Jupiters um $2.1 \cdot 10^{-24}$ km/s, was man getrost als vernachlässigbar bezeichnen kann.

5.3.1 Swing-By als elastischer Stoß

Bevor der Swing-by im Detail betrachtet wird, ist es hilfreich, sich dieses Manöver in stark vereinfachter Form als elastischen Stoß vorzustellen. Dabei wird angenommen, dass die Masse des Himmelskörpers deutlich größer ist als die Masse des vorbeifliegenden Satelliten. Die Situation ist in Abb. 5.7 dargestellt: Ein Satellit fliegt mit einer Geschwindigkeit von 8 km/s auf einen Planeten zu, der sich mit einer Geschwindigkeit von 12 km/s in die entgegengesetzte Richtung bewegt. Beide Geschwindigkeiten sind von einem ruhenden Bezugspunkt aus angegeben, beispielsweise vom Zentrum des Sonnensystems. Die Fragestellung ist nun, welche Geschwindigkeit v der Satellit in diesem Bezugssystem nach dem Swing-By-Manöver hat.

Zur Beantwortung dieser Frage ist es hilfreich, zunächst das Bezugssystem zu wechseln. Wir setzen uns hierzu in das bewegte Bezugssystem des Planeten. In diesem Bezugssystem kommt der Satellit mit einer Relativgeschwindigkeit von 20 km/s auf uns zu, »stößt« mit dem Planeten und fliegt dann – da es sich um einen elastischen Stoß handelt – mit einer Relativgeschwindigkeit von 20 km/s von uns weg. Jetzt ändern wir wieder unser Bezugssystem und betrachten die Situation vom ruhenden Bezugspunkt. Da der Satellit sich mit 20 km/s vom Planeten wegbewegt, der Planet aber selbst eine Geschwindigkeit von 12 km/s hat, entspricht die Endgeschwindigkeit des Satelliten nach dem Manöver 32 km/s. Der Satellit hat also 24 km/s an Geschwindigkeit gewonnen, während der am Planeten vorbei geflogen ist. Die hierfür benötigte Energie stammt aus der kinetischen Energie des Planeten, der zwar etwas langsamer geworden ist, was aber angesichts des großen Massenunterschieds nicht wirklich messbar ist.

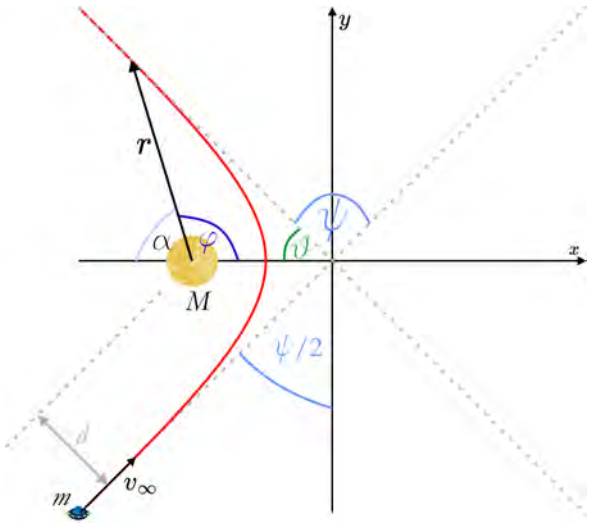


Abbildung 5.8: Hyperbelbahn des Swing-By-Manövers und Bezeicnung der relevanten Winkel.

5.3.2 Swing-By im Detail

Wir betrachten nun eine Raumsonde mit Masse m, die einen Planten passiere. Der Vorgang ist mit allen relevanten Größen in Abb. 5.8 dargestellt. Um das Swing-By-Manöver sinnvoll beschreiben zu können, ist es hilfreich, eine neue Größe, den Stoßparameter d einzuführen, der dem kleinsten Abstand der geradlinigen Verlängerung der Flugbahn aus dem Unendlichen vom Mittelpunkt des Planeten entspricht. Diese geradlinige Verlängerung verlaufe zudem unter einem Winkel φ gegen die Symmetrieachse, was dem halben Öffnungswinkel der Hyperbel entspricht. Es ist davon auszugehen, dass bei kleineren Stoßparametern ein größerer Ablenkwinkel zu erwarten ist, da die Raumsonde tiefer in das Potentialfeld des Planeten eintaucht.[5] Es werden nun auch bei der detaillierteren Betrachtung zwei Bezugssystemen verwendet, einmal das Bezugssystem des Planeten sowie das Bezugssystem der Sonne, was wir als heliozentrisches System bezeichnen wollen. Die Geschwindigkeit der Raumsonde in großer Entfernung vom Planeten sei endlich (also $v > 0$) und mit v_∞ bezeichnet. Aus dem Bezugssystem des Planeten, die wir in guter Näherung als Inertialsystem betrachten, bewegt sich die Raumsonde auf einer Hyperbelbahn, da die Gesamtenergie (bei Vernachlässigung des Einflusses anderer Himmelskörper) der Raumsonde größer Null ist. Der Zeitraum des Swing-By-Manövers sei im Vergleich

5 Natürlich muss es auch einen minimalen Stoßparameter geben, ab dem die Raumsonde mit dem Planeten kollidiert, was hier aber nicht weiter diskutiert wird.

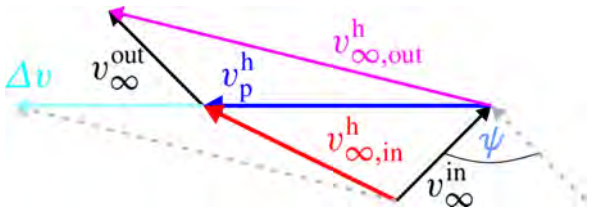


Abbildung 5.9: Vektorielle Darstellung des Swing-By-Manövers. Der Index h steht für heliozentrische Vektoren.

zur Umlaufperiode des Planeten gering, der Geschwindigkeitsvektor des Planeten soll sich also während des Manövers nicht ändern.

Wir wollen nun den Ablenkwinkel ψ der Raumsonde als Funktion des Gravitationsparameters GM des Planeten, des Stoßparameters d und der Geschwindigkeit v_∞ berechnen. Die Hyperbelbahn beschreiben wir durch die bekannte Darstellung in Polarkoordinaten gemäß

$$r = \frac{p}{1+\varepsilon\cos\varphi} = \frac{a(1-\varepsilon^2)}{1+\varepsilon\cos\varphi}. \tag{5.19}$$

Der Periapsenabstand r_p ist durch die Beziehung

$$r_p = a(1-\varepsilon) \tag{5.20}$$

gegeben, der Betrag des Bahndrehimpuls L_∞ im Unendlichen durch

$$L_\infty = mv_\infty d. \tag{5.21}$$

Da wir ein klassisches Keplerproblem mit einer wirkenden Zentralkraft vorliegen haben, handelt es sich beim Bahndrehimpuls um eine Konstante der Bewegung. Die Geschwindigkeit in der Periapsis kann aus der Vis-Viva-Gleichung zu

$$v_\mathrm{p} = \sqrt{\frac{GM}{a}\frac{1+\varepsilon}{1-\varepsilon}} \tag{5.22}$$

bestimmt werden. Der Bahndrehimpuls in der Periapsis ergibt sich damit zu

$$L_\mathrm{p} = mr_\mathrm{p}v_\mathrm{p} = ma(1-\varepsilon)\sqrt{\frac{GM}{a}\frac{1+\varepsilon}{1-\varepsilon}} = m\sqrt{GMa(1-\varepsilon^2)}, \tag{5.23}$$

was wegen der Drehimpulserhaltung $L_\infty = mdv_\infty$ entsprechen muss. Es gilt als

$$v_\infty d = \sqrt{GMa(1-\varepsilon^2)}. \tag{5.24}$$

Aus der Vis-Viva-Gleichung können wir zudem eine Beziehung zwischen

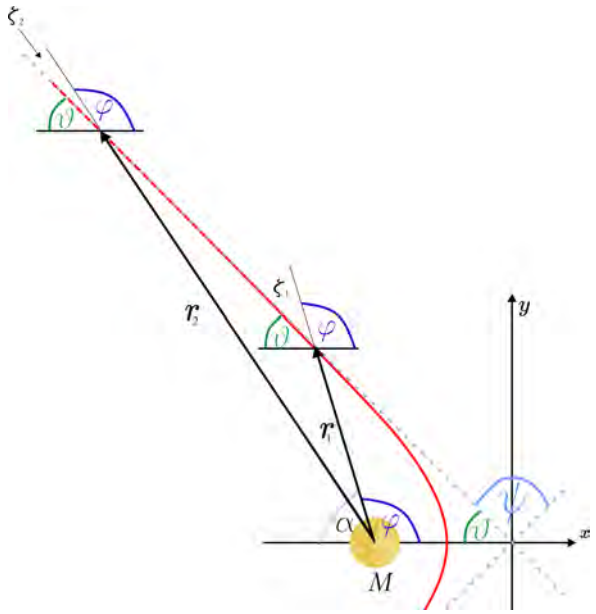


Abbildung 5.10: Winkelbetrachtung beim Swingby für große r. Läuft r gegen Unendlich, wird der Winkel ζ zwischen ϑ und φ genau Null, so dass die Summe von ϑ und φ genau 180° ergibt.

der großen Halbachse a der Hyperbelbahn und der Geschwindigkeit v_∞ herstellen, die im weiteren Verlauf hilfreich sein wird. Für $r \to \infty$ folgt aus der Vis-Viva-Gleichung

$$a = -\frac{GM}{v_\infty^2}, \tag{5.25}$$

was eingesetzt in Gl. 5.24 zu

$$\sqrt{\varepsilon^2 - 1} = \frac{dv_\infty^2}{GM} \tag{5.26}$$

führt. Für $r \to \infty$ fallen zudem die Winkel α und ϑ zusammen (vgl. Abb. 5.10), so dass

$$\varphi + \vartheta = \pi \tag{5.27}$$

gilt. Aus Gl. 5.19 folgt im Grenzfall $r \to \infty$, dass der Nenner gegen Null laufen muss, also $1 + \varepsilon \cos\varphi \to 0$ gelten muss. Ersetzt man hier φ mit $(\pi - \vartheta)$, so erhalten wir

$$\cos\vartheta = \frac{1}{\varepsilon}, \tag{5.28}$$

also eine Beziehung zwischen dem Öffnungswinkel der Hyperbel und der numerischen Exzentrizität ε. Aus Abb. 5.8 kann der Zusammenhang

$$\frac{1}{2}\psi = \frac{1}{2}\pi - \varphi \tag{5.29}$$

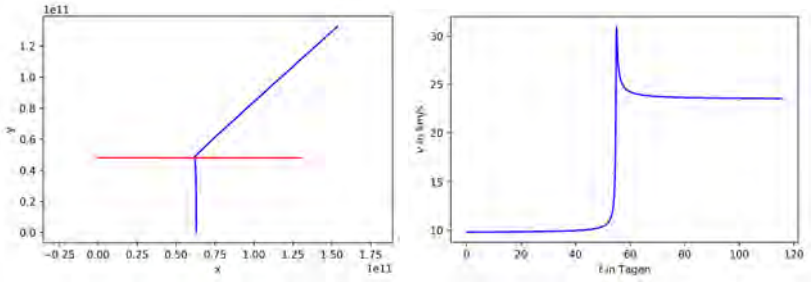


Abbildung 5.11: Links: Numerische Berechnung des Swing-By-Manövers am Jupiter. Die Bahn des Jupiters ist in rot dargestellt, die des Satelliten in blau. **Rechts**: Zunahme der kinetischen Energie des Satelliten im heliozentrischen Bezugssystem. Die Daten wurden mit dem im Abs. 5.3.3 gezeigten Programmcode berechnet.

zwischen dem Streuwinkel ψ und dem Öffnungswinkel φ abgeleitet werden. Bildet man von dieser Gleichung den Tangens, so erhält man

$$\tan\left(\frac{\psi}{2}\right) = \frac{\cos(\varphi)}{\sin(\varphi)}. \tag{5.30}$$

Hierbei wurde ausgenutzt, dass $\tan(\varphi) = -\tan(-\varphi)$ und $\cos(\varphi - \pi/2) = -\sin(\varphi)$ sowie $\sin(\varphi - \pi/2) = \cos(\varphi)$ gilt. Durch Einsetzen von Gl. 5.28 und ein paar Umformungsschritten erhält man schließlich das gesuchte Ergebnis in der Form

$$\tan\left(\frac{\psi}{2}\right) = \frac{GM}{dv_\infty^2}. \tag{5.31}$$

In Abb. 5.9 sind die relevanten Geschwindigkeitsvektoren für das Manöver dargestellt. Im planetaren Bezugssystem ändert sich der Betrag der Geschwindigkeit v_∞ nicht, es findet lediglich eine Drehung des Vektors um den Winkel ψ statt. Das entspricht der bekannten Tatsache, dass im Schwerpunktsystem – wir gehen hier ja davon aus, dass $M >> m$ ist, somit der Schwerpunkt mit unserem Bezugspunkt in guter Näherung zusammenfällt – jeder Stoßpartner beim elastischen Stoß seine kinetische Energie behält.

5.3.3 Swing-by am Jupiter numerisch gerechnet

Da sich das Swing-By-Manöver letztlich am besten numerisch lösen lässt, soll dies hier in vereinfachter Form für einen Vorbeiflug eines Satelliten am Jupiter durchgeführt werden. Der Jupiter soll dabei eine Geschwindigkeit von 13.1 km/s haben, der Satellit von 9.8 km/s. Der hier zur Verfügung gestellte Code vernachlässigt den Einfluss des Satelliten

auf den Planeten und ist nicht darauf ausgelegt, ein optimiertes Manöver durchzuführen, sondern soll mehr das didaktische Verständnis, wie man die Situation in einem Programm umsetzen könnte, trainieren. Im Code wird Bewegungsgleichung nach dem einfachsten Algorithmus, dem Euler-Verfahren, gelöst. Bei jedem Zeitschritt berechnen wir Zeit, x- und y-Koordinate für Jupiter und Satelliten. Dann berechnen wir die Beschleunigung des Satelliten an der augenblicklichen Position relativ zum Planeten. Wir rechnen die neue Position des Satelliten (und auch des Planeten) aus, indem wir die jeweilige Geschwindigkeit mit dem Zeitschritt multiplizieren und zur Position addieren. Analog rechnen wir die neue Geschwindigkeit aus, indem wir die jeweilige Beschleunigung mit dem Zeitschritt multiplizieren und zur Geschwindigkeit addieren.

Der Abstand zwischen Satellit und Jupiter ist so gewählt, dass der Satellite außerhalb der Einflusssphäre (*sphere of influence*, SOI) des Jupiters startet. Hierunter versteht man einen Bereich, ab dem der Einfluss der Gravitation des Planeten im Vergleich zum Einfluss der Sonne keinen nennenswerte Einfluss mehr auf die Bewegung des Satelliten haben soll. Beim Jupiter beträgt der SOI-Radius etwa $4.8 \cdot 10^7$ km. Am gezeigten Beispielcode wird eine Erhöhung der Geschwindigkeit auf 23.5 km/s erzielt, also eine Erhöhung um das 2.4fache der Anfangsgeschwindigkeit. Das ist vergleichbar mit den Bedingungen von Pionieer 10 aus dem anfangs erwähnten Paper von van Allen.

```
import numpy as np
import matplotlib.pyplot as plt
MGamma=1.2669e17     # Standardgravitationsparameter des
    Jupiters
msat = 1000          # Masse des Satelliten
Rjup = 69911000      # Radius des Jupiters

### Startbedingungen fuer Jupiter
xjup = 0*Rjup
yjup = 4.8e10
vxjup = 13100
vyjup = 0

### Startbedingungen fuer Satelliten
xsat = 6.3e10
ysat = 0
vxsat = 0
vysat = 9800

tmax = 10000000
deltat = 15.0
steps = int(tmax/deltat)

### Startwerte setzen
```

```
t = 0.0
vsat = np.sqrt(vxsat**2 + vysat**2)
vjup = np.sqrt(vxjup**2 + vyjup**2)
# Abstand zwischen Jupiter und Satellit
xdist = xjup - xsat
ydist = yjup - xsat
r=np.sqrt(xdist**2+ydist**2)
xwerte_jup = [xjup]
ywerte_jup = [yjup]
vxwerte_jup = [vxjup]
vywerte_jup = [vyjup]
vwerte_jup = [vjup]
xwerte_sat = [xsat]
ywerte_sat = [ysat]
vxwerte_sat = [vxsat]
vywerte_sat = [vysat]
vwerte_sat = [vsat]
rwerte = [r]
twerte = [t]
i = 0
while i < steps:
    t = t + deltat
    twerte.append(t)
    xjup = xjup + deltat*vxjup
    yjup = yjup + deltat*vyjup
    xwerte_jup.append(xjup)
    ywerte_jup.append(yjup)
    xsat = xsat + deltat*vxsat
    ysat = ysat + deltat*vysat
    xwerte_sat.append(xsat)
    ywerte_sat.append(ysat)
    xdist = xjup - xsat
    ydist = yjup - ysat
    r = np.sqrt(xdist**2+ydist**2)
    rwerte.append(r)
    ax = MGamma*xdist/r**3
    ay = MGamma*ydist/r**3
    vxsat = vxsat + ax*deltat
    vysat = vysat + ay*deltat
    v = np.sqrt(vxsat**2 + vysat**2)
    vwerte_sat.append(v)
    i = i + 1
    if r <= Rjup:
        print("CRASH")
        break

xwerte_sat = np.array(xwerte_sat)
ywerte_sat = np.array(ywerte_sat)
twerte = np.array(twerte)
vwerte_sat = np.array(vwerte_sat)
rwerte = np.array(rwerte)
vwerte_sat = vwerte_sat/1000
```

```
76 twerte = twerte / (24*60*60)
77 plt.figure(dpi=300)
78 plt.plot(xwerte_sat,ywerte_sat,'b')
79 plt.plot(xwerte_jup,ywerte_jup,'r')
80 plt.plot(0,0,'b,')
81 plt.xlabel('x')
82 plt.ylabel('y')
83 plt.axis("equal")
84 circle1 = plt.Circle((0, 4.8e10), Rjup, color='b')
85 plt.gca().add_patch(circle1)
86 plt.show()
87 plt.figure(dpi=300)
88 plt.plot(twerte,vwerte_sat,'b')
89 #plt.plot(0,0,'b,')
90 plt.xlabel(r'$t$ in Tagen')
91 plt.ylabel(r'$v$ in km/s')
92 plt.show()
93 rmin = np.min(rwerte)/Rjup
94 vend = vwerte_sat[len(vwerte_sat)-1]
95 print("Minimale Annaeherung (in Vielfachen von R_Jupiter): "
      , rmin)
96 print("Endgeschwindigkeit des Satelliten (in km/s): ", vend)
```

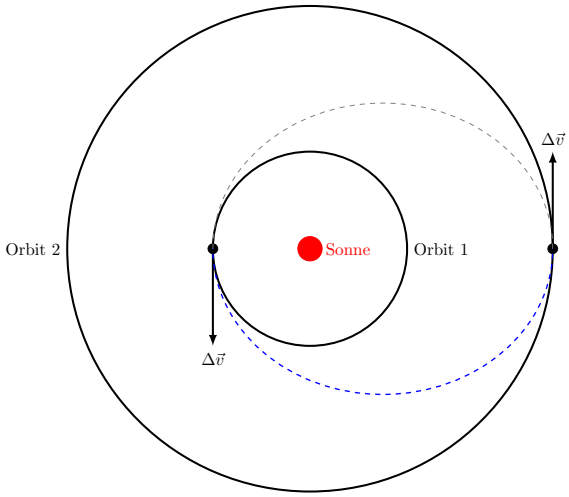


Abbildung 5.12: Darstellung des Hohmann-Transfers, einem elliptischen Transfer zwischen zwei zirkularen Orbits auf Basis zweier Schubmanöver des Raumfahrzeugs.

5.4 Hohmann-Transfer

Im Jahr 1925 stellte Walter Hohmann die Vermutung an, dass es einen optimalen Transfer eines Raumfahrzeugs zwischen zwei koplanaren, konzentrischen, kreisförmigen Orbits gibt, der mit genau zwei impulsiven Geschwindigkeitsänderungen (Impulsen) auskommt und von allen möglichen Zwei-Impuls-Transferbahnen energetisch optimal ist, d. h. eine minimale Menge an Treibstoff benötigt. Dieser Transfer wird als Hohmann-Transfer bezeichnet. Das erste Schubmanöver bringt das Fahrzeug auf die Transferbahn, während das zweite Schubmanöver die Umlaufbahn auf den endgültigen Radius zirkularisiert. Jede Geschwindigkeitsänderung ist parallel zum momentanen Geschwindigkeitsvektor, was offensichtlich das Manöver mit dem geringsten Treibstoffverbrauch darstellt (vgl. Abb. 5.12). Die große Halbachse a_H dieser Hohmann-Ellipse entspricht der halben Summe der Radien r_1 und r_2 der beiden Kreisbahnen ($r_1 < r_2$), d. h.

$$a_\mathrm{H} = \frac{r_1 + r_2}{2}. \tag{5.32}$$

Auf der inneren Kreisbahn mit Radius r_1 habe das Raumfahrzeug die Geschwindigkeit v_1. Wenn hier nun das Schubmanöver durchgeführt wird, dann wird aus dem Mittelpunkt der Kreisbahn automatisch der Brennpunkt der sich neu einstellenden Ellipsenbahn und das Raumfahrzeug befindet sich in der Periapsis der Ellipse. Seine neue Geschwin-

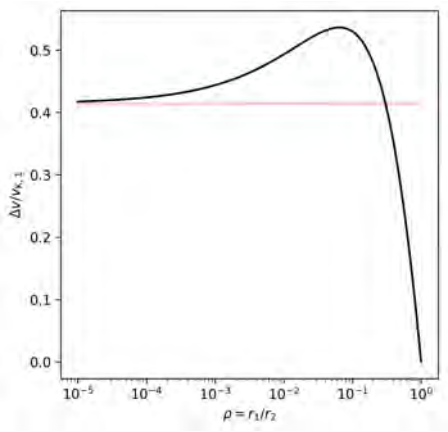


Abbildung 5.13: Antriebsbedarf für einen Hohmann-Übergang von einer niederen auf eine höhere Bahn. Kleinere Werte von ρ entsprechen bei gegebener Startbahn mit Radius r_1 einem zunehmenden Radius r_2.

digkeit an der Periapsis v_p ist also die größte Geschwindigkeit dieser Ellipsenbahn. Es gilt

$$v_p = v_1 + \Delta v_1. \tag{5.33}$$

Beim Erreichen der Kreisbahn r_2 befindet sich das Raumfahrzeug in der Apoapsis der Ellipsenbahn, hat also die geringste Geschwindigkeit v_p dieser Bahnkurve. Diese Geschwindigkeit ist nicht ausreichend groß, um auf der Kreisbahn zu verweilten, das Raumfahrzeug würde ohne erneutes Schubmanöver auf der Ellipsenbahn weiterfliegen und die Kreisbahn verlassen. Es muss also am Ort der Apoapsis erneut die Geschwindigkeit erhöht werden. Die benötigte Geschwindigkeit beträgt v_2 und lässt sich wie jeder Geschwindigkeit einer Kreisbahn mit Radius r aus der Formel $v_r = \sqrt{\mu/r}$ berechnen. Die Geschwindigkeit der elliptischen Bahn kann aus der Vis-Viva-Gleichung berechnet werden. Es gilt an der Apoapsis:

$$v_2 = v_a + \Delta v_2. \tag{5.34}$$

Das benötigte Δv_H für den Hohmann-Transfer beträgt $\Delta v_\mathrm{H} = \Delta v_1 + \Delta v_2$. Man kann zeigen, dass diese Größe mit $\rho = r_1/r_2$ durch die Beziehung

$$\Delta v(\rho) = \Delta v_1 + \Delta v_2 = v_{1,\mathrm{K}} \cdot \left((1-\rho) \cdot \sqrt{\frac{2}{1+\rho}} + \sqrt{\rho} - 1 \right). \tag{5.35}$$

gegeben ist. Hierbei bezeichnen Δv_1 und Δv_2 die Geschwindigkeitsdifferenzen in Peri- und Apoapsis und $v_{1,\mathrm{K}}$ die Geschwindigkeit der Start-Kreisbahn.

In der Realität sind die Annahmen, die beim Hohmann-Transfer gemacht werden, nicht gegeben. Planetenbahnen sind elliptisch (wenn

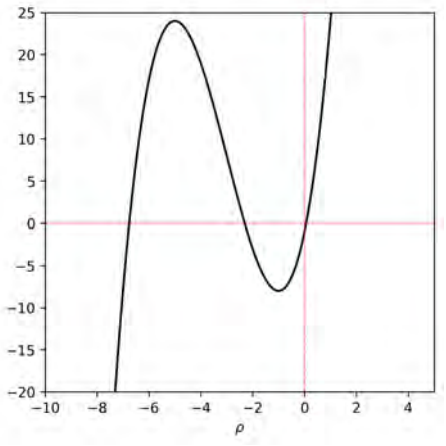


Abbildung 5.14: Graphische Darstellung von $\rho^3 + 9\rho^2 + 15\rho - 1 = 0$.

auch oft nur mit geringer Exzentrizität) und auch nicht unbedingt perfekt koplanar. Dennoch eignet sich diese Berechnung für eine grobe Abschätzung, ob sich eine Mission mit den zur Verfügung stehenden Triebwerken realisieren lässt.

Prüfen wir Gl. 5.35 der Vollständigkeit halber noch auf Plausibilität. Für $r_1 = r_2$, d. h. für einen nicht-stattfindenden Hohmann-Transfer, beträgt $\Delta v = 0$, was sicherlich sinnvoll ist. Für $r_2 \to \infty$ folgt $\rho \to 0$, somit

$$\Delta v = \Delta v_{1,\text{Flucht}} = v_{1,\text{K}} \left(\sqrt{2} - 1\right) = v_{1,\text{Flucht}} - v_{1,\text{K}}. \tag{5.36}$$

Für einen Orbittransfer ins Unendliche wird also die Fluchtgeschwindigkeit des Startorbits abzüglich der bereits vorhandenen Kreisgeschwindigkeit benötigt, was auch ein vernünftiges Ergebnis darstellt.

Die Funktion $\Delta v(\rho)/v_{1,\text{K}}$ ist graphisch in Abb. 5.13 für einen Orbittransfer mit $r_2 > r_1$ dargestellt. Die rote gestrichelte Linie in dieser Abbildung entspricht $\Delta v_{1,\text{Flucht}}/v_{1,\text{K}}$, also dem Geschwindigkeitsinkrement, das man benötigt, um von der Kreisbahn mit Radius r_1 auf die Fluchtgeschwindigkeit ($r_2 \to \infty$) zu kommen normiert auf die Geschwindigkeit $v_{1,\text{K}}$. Das Maximum dieser Funktion erhält man aus der Extremwertbedingung $\mathrm{d}(\Delta v(\rho)/\,\mathrm{d}\rho) = 0$. Nach einigen Umformungsschritten gelangt man zur kubischen Gleichung

$$\rho^3 + 9\rho^2 + 15\rho - 1 = 0. \tag{5.37}$$

Die Funktion hat drei Nullstellen, wobei zwei davon im negativen Bereich von ρ liegen und physikalisch nicht sinnvoll sind, da $\rho > 0$ gelten muss. Die gesuchte Nullstelle liegt bei $\rho = 0.064176$. Der Antriebsbedarf des Hohmann-Transfers wird also maximal bei $r_2 = 15.852 \cdot r_1$. Für

$r_2 > 15.852 \cdot r_1$ sinkt der Antriebsbedarf wieder. Das liegt daran, dass der Bedarf für Δv_2 schneller abfällt, als der Bedarf für Δv_1 steigt.

5.5 Energetische Betrachtung des Hohmann-Transfers

Nachdem der Hohmann-Transfer im Detail in Abschnitt 5.4 besprochen wurde, soll hier nochmal eine energetische Betrachtung durchgeführt werden. Hohmanns Vermutung, das sein beschriebener Orbittransfer der energetisch günstigste ist, soll hier bewiesen werden. Die Beweisführung orientiert sich an der Darstellung von John E. Prussing (*Simple Proof of the Global Optimality of the Hohmann Transfer*, J. Guidance, Vol. 15, No. 4: Engineering Notes, 1992), die man als Studierender im Grundstudium verstehen kann. Dennoch hier gleich der Hinweis: der Beweis ist nicht ganz trivial. Man sollte sich klar machen, dass Hohmann selbst seine Vermutung nicht beweisen konnte. Es hat sogar noch fast 40 Jahre gedauert, bis überhaupt jemand seine Hypothese bestätigen konnte (das war D. F. Lawden im Jahr 1963). Es ist also keine Schande, wenn man sich mit diesem Abschnitt schwer tut. Da das Paper von Prussing für einen Anfänger vielleicht etwas schwierig zu lesen ist, wird hier versucht, die einzelnen Schritte sehr ausführlich darzustellen.

Energetisch günstig im Sinne eines Orbittransfers bedeutet, dass dieser den geringsten Treibstoffverbrauch aller möglichen Transferbahnen aufweist. Dies ist gleichbedeutend mit der Tatsache, dass das benötigte gesamte (totale) $\Delta\vec{v}_\mathrm{T}$ für den Orbitwechsel minimal sein muss. Schauen wir uns den Sachverhalt mit Hilfe von Abbildung 5.15 genauer an. Die Ausgangssituation ist einfach: wir wollen ein Raumfahrzeug von einer Kreisbahn mit Radius r_1 auf eine Kreisbahn mit Radius r_2 befördern. Dazu machen wir hier die Annahme, dass $r_1 < r_2$ sein soll, zudem sollen beide Kreisbahnen in der gleichen Ebene liegen und es sind insgesamt nur zwei impulsive Schubmanöver zugelassen. An irgendeiner Stelle der Kreisbahn r_1 wird kurz das Triebwerk des Raumfahrzeugs gezündet und eine Geschwindigkeitsänderung $\Delta\vec{v}_1$ herbeigeführt. Wir machen dabei die Annahme, dass die Geschwindigkeitsänderung in beliebige Richtung zeigen kann, denn für unsere Beweisführung wollen wir keine konkrete Festlegung *a priori* treffen. Eine Geschwindigkeitsänderung, die zu einer Ellipse führen würde, die innerhalb des inneren Kreises liegt, ist aber wenig sinnvoll, da dann ja nicht die äußere Kreisbahn erreicht werden kann. Wir werden also ein paar Kriterien definieren müssen, damit überhaupt ein sinnvoller Orbittransfer zustande kommen kann. Das Raumfahrzeug befindet sich nach dem Schubmanöver nicht mehr auf einer Kreisbahn mit Geschwindigkeit $\vec{v}_{1,\,\mathrm{K}}$, sondern auf einer Ellipse.

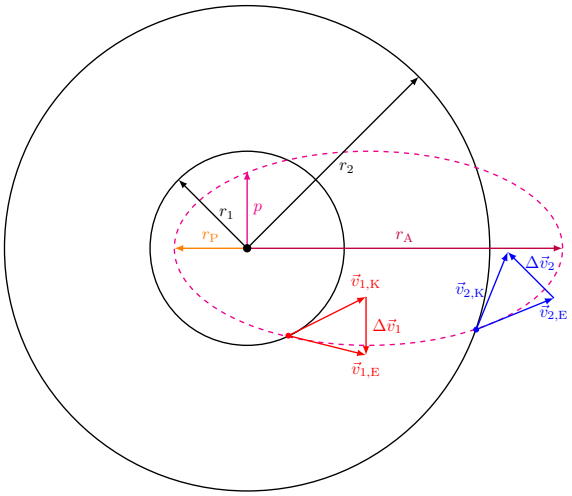


Abbildung 5.15: Allgemeine Darstellung eines elliptischen Orbittransfers zwischen zwei koplanaren, zirkularen Orbits mittels zweier impulsiver Schubmanöver. Das erste Manöver bewirkt eine Änderung der Geschwindigkeit derart, dass das Raumfahrzeug den zirkularen Orbit r_1 verlässt und sich auf danach auf einer elliptischen Bahn bewegt. Beim Erreichen der zweiten Kreisbahn wird mit einem zweiten Schubmanöver der elliptische Orbit so zirkularisiert, dass sich das Raumfahrzeug danach auf einer Kreisbahn mit Radius r_2 befindet. Damit dieser elliptische Transfer überhaupt möglich ist, müssen gewisse Randbedingungen oder geometrische Anforderungen an die elliptische Bahn gestellt werden (siehe Fließtext). Diese Randbedingungen schränken die möglichen Ellipsenbahnen, die wir durch die Parameter p und ε charakterisieren ein (vgl. Abbildung 5.16). Um zu zeigen, dass der Hohmann-Transfer der energetisch optimalste Orbittransfer zwischen den beiden Orbits ist, muss man nur noch innerhalb des eingeschränkten Parameterbereichs argumentieren, was die Beweisführung erheblich vereinfacht.

Der Geschwindigkeitsvektor der Ellipse an diesem Punkt sei $\vec{v}_{1,\,\mathrm{E}}$. Wichtig hier ist die Tatsache, dass der Mittelpunkt der Kreisbahnen und der Brennpunkt der Ellipse zusammenfallen. Am äußeren Kreisorbit mit Radius r_2 wird wieder ein Schubmanöver durchgeführt, damit die Ellipsenbahn in die äußere Kreisbahn übergeht. In Abbildung 5.15 ist die Ellipse so gezeichnet, dass die Hauptachse horizontal liegt. Prinzipiell kann die Ellipse aber beliebig im Raum liegen. Da Kreismittelpunkte und Brennpunkt der Ellipse aber immer zusammenfallen und man das ganze demnach beliebig rotieren kann, ist die Darstellung in der Abbildung zur besseren Übersicht so gewählt worden.

Wie sehen nun unsere Rahmenbedingungen für einen sinnvollen Orbittransfer aus? Bevor wir das präzisieren, wollen wir nochmal die Ellipsengleichung in Polarkoordinaten anschauen:

$$r = \frac{p}{1 + \varepsilon \cos \varphi} \tag{5.38}$$

Die numerische Exzentrizität ε einer Ellipse ist definiert über die Bedingung $0 < \varepsilon < 1$. Diese Gleichung soll nun die Hohmann-Ellipse der Transferbahn beschreiben. Für einen Orbittransfer muss diese Ellipse die beiden Kreisbahnen schneiden oder zumindest berühren. Schauen wir uns das mit Hilfe von Abbildung 5.15 an. Die Ellipse hat – wie alle Ellipsen – ein paar charakteristische Größen. Hier von Bedeutung ist der Halbparameter p sowie die Abstände vom Brennpunkt zu den beiden Apsiden, also den Periapsis- und den Apoapsisabstand (r_P, r_A). Betrachten wir zunächst die Periapsis. Den Abstand r_P erhält man aus Gleichung 5.38 für den Winkel $\varphi = 0$, also

$$r_\mathrm{P} = \frac{p}{1 + \varepsilon}. \tag{5.39}$$

Die Behauptung ist nun folgende: Ist $r_\mathrm{P} > r_1$, dann schneidet die Ellipse niemals die innere Kreisbahn. Das mag auf den ersten Blick gar nicht klar sein, da man sich eine Ellipse gerne als beliebige eiförmige Struktur vorstellt und man sich sicherlich beim Blick auf Abbildung 5.15 eine solche Form vorstellen kann, die die innere Kreisbahn schneidet, aber einen Periapsisabstand mit $r_\mathrm{P} > r_1$ aufweist. Die Vorstellung trügt aber, da unsere Hohmann-Ellipse eben nicht jede beliebige Form annehmen kann, da wir den Brennpunkt festgenagelt haben. Der Periapsisabstand ist nämlich mit dem Halbparameter p der Ellipse verknüpft. Ersetzt man in Gleichung 5.39 den Radius r_P durch r_1, so folgt $p = r_1(1 + \varepsilon)$. Da $1 + \varepsilon > 1$ für eine Ellipsenbahn gilt, folgt, dass der Halbparameter p größer als der Radius r_1 wird. Dann kann die Transferellipse aber die Kreisbahn nicht mehr schneiden. Wir halten also als erste Randbedingung fest, dass

$$r_\mathrm{P} = \frac{p}{1 + \varepsilon} \leq r_1 \tag{5.40}$$

gelten muss.

Etwas eingängiger ist die zweite Randbedingung. Damit die Transferellipse die Kreisbahn mit Radius r_2 erreichen kann, muss die Bedingung

$$r_\mathrm{A} = \frac{p}{1 - \varepsilon} \geq r_2 \tag{5.41}$$

erfüllt sein. Man kann diese beiden Randbedingungen etwas übersichtlicher in der Form

$$\begin{aligned} p &\leq r_1(1 + \varepsilon) \\ p &\geq r_2(1 - \varepsilon) \end{aligned} \tag{5.42}$$

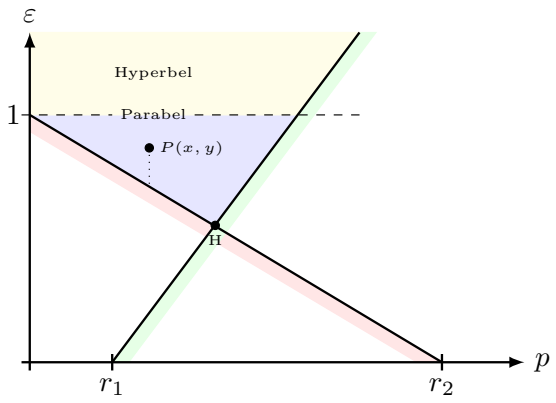


Abbildung 5.16: Graphische Darstellung möglicher (p, ε)-Werte von Transferellipsen zwischen zwei koplanaren, konzentrischen Orbits mit Radien r_1 und r_2. Die Ellipsenbahnen sind einerseits durch die numerische Exzentrizität eingeschränkt ($0 < \varepsilon < 1$), andererseits durch die Rahmenbedingungen, die sich aus Gl. 5.42 ergeben. Der Punkt H entspricht dem Transferorbit mit kleinster numerischer Exzentrizität, was wiederum dem Hohmann-Transfer entspricht.

darstellen. Wir haben damit bereits einen Bereich eingegrenzt, in dem sich mögliche Werte von p bewegen. Stellen wir diese beiden Ungleichungen nach ε um, so erhalten wir analog den Wertebereich für die numerischen Exzentrizitäten. Prussing hat für seine weitere Beweisführung eine sinnvolle graphische Darstellung möglicher Parameter von Transferellipsen in Form eines p-ε-Diagramms gefunden, die in Abbildung 5.16 nachgezeichnet wurde.

Geht man beispielsweise von $p \leq r_1(1 + \varepsilon)$ aus, so kann man dies in die Form $\varepsilon = \varepsilon(p)$ bringen:

$$\varepsilon \geq \frac{p}{r_1} - 1. \tag{5.43}$$

Mit dem elliptischen Transfer vereinbare Exzentrizitäten müssen also die Bedingung aus Gleichung 5.43 erfüllen. Der kleinste Wert für ein gegebenes p ist also eine Gerad mit positiver Steigung (diese beträgt r_1^{-1}; die Gerade schneidet die ε-Achse bei -1). Aus der zweiten Randbedingung folgt:

$$\varepsilon \geq -\frac{p}{r_2} + 1. \tag{5.44}$$

Dies entspricht eine Geraden mit negativer Steigung r_2^{-1} und Schnittpunkt mit der ε-Achse bei $+1$. Diese beiden Geraden sind in Abbildung 5.16 eingezeichnet, der nicht erlaubte Bereich kleinerer Werte (unterhalb der Geraden ist farbig hervorgehoben). Da wir nur Ellipsen betrachten, gibt es eine weitere Randbedingung, nämlich $\varepsilon < 1$.

Wir erhalten dadurch einen Bereich erlaubter p, ε-Werte für unseren Transferorbit, welcher der bläulich hervorgehobenen Dreiecksfläche in Abbildung 5.16 entspricht.

Was beschreiben eigentlich die Punkte auf den Geraden für Transfer-Ellipsenbahnen? Für die Gerade mit positiver Steigung gilt $r_\mathrm{P} = r_1$, analog gilt für die Gerade mit negativer Steigung die Beziehung $r_\mathrm{A} = r_2$. Für die Gerade positiver Steigung entspricht dies einer Ellipse, deren Periapsis mit r_1 zusammenfällt. In diesem Fall würde ein Schubmanöver in Richtung des Geschwindigkeitsvektors der Kreisbahn zeigen, da ja die Tangente der Kreisbahn mit der Tangente der Ellipsenbahn zusammenfällt. Die gleiche Überlegung kann für die zweite Gerade mit fallender Steigung durchgeführt werden.

Schauen wir uns nun die Geschwindigkeitsänderung $\Delta\vec{v}$ etwas näher an. Wir haben bei beiden Manövern eine Geschwindigkeit v_K, die für den zirkularen Orbit steht, eine zweite v_E für den elliptischen Orbit an entsprechender Stelle, an dem der Orbitwechsel stattfindet (also beim Schnittpunkt von Ellipsenbahn und Radius r_2). Der Winkel zwischen den beiden Geschwindigkeitsvektoren sein mit θ bezeichnet. Es gilt also entweder $\Delta\vec{v} = \vec{v}_\mathrm{E} - \vec{v}_\mathrm{K}$ oder $\Delta\vec{v} = \vec{v}_\mathrm{K} - \vec{v}_\mathrm{E}$, je nachdem, welche Geschwindigkeitsänderung man sich anschaut. Die Geschwindigkeitsänderungen sind natürlich nicht gleich groß. Mit Hilfe des Kosinussatzes kann man das für beide Geschwindigkeitsänderungen in die gleiche Form bringen:

$$\Delta v^2 = \underbrace{v_\mathrm{E}^2}_{\text{Ellipse}} + \underbrace{v_\mathrm{K}^2}_{\text{Kreis}} - 2\,\underbrace{v_\mathrm{E} v_\mathrm{K} \cos\theta}_{\text{Skalarprodukt}}\,. \tag{5.45}$$

Der gemischte Term enthält das Skalarprodukt der beiden Geschwindigkeitsvektoren, also die jeweilige Projektion des einen auf den anderen Vektor. Wir definieren die Geschwindigkeit $v_\theta = v_\mathrm{E} \cos\theta$ als Projektion von v_E auf v_K und fassen das ganze zusammen als

$$\Delta v^2 = v_\mathrm{E}^2 + v_\mathrm{K}^2 - 2 v_\mathrm{K} v_\theta. \tag{5.46}$$

Sei nun M die Masse des Zentralkörpers, der im Mittelpunkt der beiden konzentrischen Kreise und im Brennpunkt der Ellipse sitzt. Die Geschwindigkeit für die Bewegung auf der jeweiligen Kreisbahn i ist gegeben durch

$$v_{i,\mathrm{K}}^2 = \frac{\gamma M}{r_i}. \tag{5.47}$$

Die Geschwindigkeit auf einer Ellipsenbahn für einen bestimmten Radiusvektor r kann aus der Vis-Viva-Gleichung berechnet werden:

$$v_{i,\mathrm{E}}^2 = \gamma M \left(\frac{2}{r_i} - \frac{1}{a}\right) = \gamma M \left(\frac{2}{r_i} + \frac{\varepsilon^2 - 1}{p}\right). \tag{5.48}$$

Im letzten Schritt wurde die Beziehung $p = a(1 - \varepsilon^2)$ berücksichtigt, wodurch man die Ellipsenbahngeschwindigkeit in Abhängigkeit der beiden Parameter p und ε ausdrücken kann. Dies ist wichtig, da wir ja den einen Parametersatz dieser beiden Größen suchen, der zu einem minimalen Treibstoffverbrauch führt.

Jetzt muss noch das v_θ bestimmt werden. Hier nutzt man aus, dass auf einer Kreisbahn der Radiusvektor $\vec{r}$ immer senkrecht auf den Geschwindigkeitsvektor $\dot{\vec{r}}$ steht. Da im Moment des impulsiven Schubmanövers immer Kreis- und Ellipsenbahn zusammenfallen, muss auch der Drehimpuls beim Übergang zwischen den beiden Bahnformen konstant bleiben. Im Falle der Kreisbahn ergibt sich daher die Beziehung $r\dot{r} = h$, wobei h für den spezifischen Bahndrehimpuls stehen soll, was nichts anderes ist als der Bahndrehimpuls $\vec{L}$ geteilt durch die Masse des sich bewegenden Körpers (in unserem Fall also die Masse des Raumfahrzeugs). Im Falle der Ellipse haben wir die Beziehung $h = r v_\mathrm{E} \sin(\theta + \pi)$. Das Argument des Sinus setzt sich zusammen aus den 90° zwischen $\vec{r}$ und $\dot{\vec{r}}$ der Kreisbahn sowie des Winkels θ zwischen $\dot{\vec{r}}$ der Kreisbahn und $\vec{v}_\mathrm{E}$ der Ellipse. Aus der Gleichheit der spezifischen Drehimpulse von Kreisbahn und Ellipse sowie der Beziehung $\sin(\theta + \pi) = \cos\theta$ folgt:

$$v_\theta = \frac{h}{r} = \frac{\sqrt{\mu p}}{r} = \frac{\sqrt{\gamma M p}}{r}. \tag{5.49}$$

Im letzten Schritt haben wir den Zusammenhang $h^2 = \mu p$ ausgenutzt, der aber hier nicht hergeleitet wird. Für die Geschwindigkeitsänderung i können wir nun (mit ein paar kleineren Umformungen) folgende Beziehung aufstellen:

$$\Delta v_i^2 = \gamma M \left(\frac{3}{r_i} + \frac{\varepsilon^2 - 1}{p} - 2\sqrt{\frac{p}{r_i^3}} \right). \tag{5.50}$$

Werden wir nun konkreter und schreiben die totale Geschwindigkeitsänderung auf:

$$\Delta v_\mathrm{T}^2 = \Delta v_1(p, \varepsilon) + \Delta v_2(p, \varepsilon). \tag{5.51}$$

Hier steht $\Delta v_1(p, \varepsilon)$ für den Übergang von der Kreisbahn r_1 auf die Ellipse und $\Delta v_2(p, \varepsilon)$ für den Übergang von der Ellipse auf die Kreisbahn r_2. Schauen wir uns nun an, wie sich das Δv_i verhält, wenn man die Exzentrizität ε variiert:

$$\frac{\partial \Delta v_i}{\partial \varepsilon} = \frac{1}{\Delta v_i} \gamma M \frac{\varepsilon}{p}. \tag{5.52}$$

Für die gesamte Geschwindigkeitsänderung bei Variation von ε gilt damit:

$$\frac{\partial \Delta v_\mathrm{T}}{\partial \varepsilon} = \frac{\gamma M \varepsilon}{p} \left(\frac{1}{\Delta v_1} + \frac{1}{\Delta v_2} \right) > 0. \tag{5.53}$$

Man sieht, dass Δv_T monoton wächst, wenn ε größer wird. Für einen gegebenen Punkt $P(p, \varepsilon)$ der erlaubten Transferellipsen kann daher das Δv_T verringert werden, wenn man bei festem p die numerische Exzentrizität absenkt. Das minimale ε liegt für ein festes p daher auf einer der beiden Geraden $p = r_1(1+\varepsilon)$ bzw. $p = r_2(1-\varepsilon)$. Es reicht also aus, sich nur die Punkte auf den beiden Geraden anzuschauen.

Die Abhängigkeit von Δv_i von p kann durch Einsetzen der jeweiligen Geradengleichungen überwunden werden. Wir kennzeichnen zur besseren Übersicht diese Geschwindigkeitsänderungen mit einem Dach über dem v (also $\Delta\hat{v}$) und führen diese Rechnung exemplarisch für die steigende Gerade durch. Wir setzen also $p = r_1(1+\varepsilon)$ ein:

$$\begin{aligned}\Delta\hat{v}_i^2 &= \gamma M \left(\frac{3}{r_i} - 2\sqrt{\frac{r_1(1+\varepsilon)}{r_i^3}} + \frac{\varepsilon^2 - 1}{r_1(1+\varepsilon)} \right) \\ &= \gamma M \left(\frac{3}{r_i} - 2\sqrt{\frac{r_1(1+\varepsilon)}{r_i^3}} - \frac{1-\varepsilon}{r_1} \right).\end{aligned} \tag{5.54}$$

Wenn nun gezeigt werden kann, dass $\partial\Delta v_i/\partial\varepsilon > 0$ auf beiden Geraden gilt, dann ist der Punkt mit dem kleinsten ε der energetisch günstigste Punkt. Am Beispiel der ersten Gerade müssen wir dies für zwei Terme durchführen, da $\Delta\hat{v}_\mathrm{T} = \Delta\hat{v}_1 + \Delta\hat{v}_2$ zu berücksichtigen ist. Im ersten Term wird $r_i = r_1$ im zweiten Term $r_i = r_2$ eingesetzt. Man erhält somit:

$$\begin{aligned}2\Delta\hat{v}_1 \frac{\partial\Delta\hat{v}_1}{\partial\varepsilon} &= \frac{\gamma M}{r_1} \left(1 - \frac{1}{\sqrt{1+\varepsilon}} \right) > 0 \\ 2\Delta\hat{v}_2 \frac{\partial\Delta\hat{v}_2}{\partial\varepsilon} &= \frac{\gamma M}{r_1} \left(1 - \frac{1}{(r_2/r_1)^{3/2}\sqrt{1+\varepsilon}} \right) > 0.\end{aligned} \tag{5.55}$$

Man sieht sofort, dass damit $\partial\Delta\hat{v}_\mathrm{T}/\partial\varepsilon > 0$ gelten muss. Der energetisch günstigste Übergang liegt also bei der kleinsten numerischen Exzentrizität der Geraden mit positiver Steigung. Eigentlich ist man damit schon fertig, aber formal kann man die ganze Prozedur auch für die andere Gerade durchführen:

$$\begin{aligned}2\Delta\hat{v}_1 \frac{\partial\Delta\hat{v}_1}{\partial\varepsilon} &= \frac{\gamma M}{r_2} \left((r_2/r_1)^{3/2} \frac{1}{\sqrt{1-\varepsilon}} - 1 \right) > 0 \\ 2\Delta\hat{v}_2 \frac{\partial\Delta\hat{v}_2}{\partial\varepsilon} &= \frac{\gamma M}{r_2} \left(\frac{1}{\sqrt{1-\varepsilon}} - 1 \right) > 0.\end{aligned} \tag{5.56}$$

Auch in diesem Fall folgt $\partial\Delta\hat{v}_\mathrm{T}/\partial\varepsilon > 0$. Der energetisch günstigste Übergang ist also derjenige mit geringster numerischer Exzentrizität,

was dem Punkt H in Abbildung 5.16 – also dem Hohmann-Transfer – entspricht.

Der Hohmann-Transfer ist wohl mit der einfachste Bahnübergang (nur zwei impulsive Manöver), dennoch ist die Beweisführung, dass er energetisch bei zwei impulsiven Manövern optimal ist, alles andere als trivial. Ähnliche Berechnungen bei etwas komplexeren Transferbahnen durchzuführen, ist nochmal deutlich komplizierter und wahrlich kein Theme einer Einführungsveranstaltung, sondern eher einer Spezialvorlesung über Bahnmechaniken. Bereits die Rechnung des Transfers zwischen zwei elliptischen Bahnen statt zwischen zwei Kreisbahnen erhöht den Rechenaufwand signifikant, sofern nicht starke Vereinfachungen wie koapsidale Ellipsen angenommen werden.

Man kann zeigen, dass für Kreisbahnen, die weit auseinander liegen (genau gesagt für $r_2 > 11.94\, r_1$), immer genau ein bi-eliptischer Transfer gefunden werden kann, der energetisch günstiger als der Hohmann-Transfer ist, somit gilt unsere hier durchgeführte Berechnung auch nur eingeschränkt. Für $r_2 > 15.58\, r_1$ ist sogar jeder bi-eliptischer Transfer energetisch günstiger als ein Hohmann-Transfer. Nachteil dieser bi-eliptischen Transferbahnen ist allerdings die deutlich längere Zeit, die es braucht, bis man am Ziel angekommen ist.

5.6 Sternfeld-Trajektorien

Wir haben bereits den Hohmann-Transfer als optimalen Zwei-Impuls-Transfer kennen gelernt. Nun gibt es in der Raumfahrt keine Beschränkung der Anzahl von Schubmanövern, und vielleicht findet sich unter dieser Voraussetzung ein anderer Transfer, der mit weniger Treibstoff auskommt als der Hohmann-Transfer. Ein solches Manöver wurde 1934 von Ari Sternfeld[6] vorgestellt und ist heute als bielliptischer Transfer bekannt.

6 Ari Abramowitsch Sternfeld (1905–1980) gehörte zu den Pionieren der Raumfahrt und teilte mit vielen von ihnen das Schicksal, in ihrer wissenschaftlichen Blütezeit von der Wissenschaftsgemeinde als Spinner belächelt zu werden. Erst nach einer entbehrungsreichen Zeit ohne regelmäßiges Einkommen oder mit Beschäftigungen in völlig fachfremden Bereichen wurde ihnen die gebührende Aufmerksamkeit zuteil. Es lohnt sich, sich mit den Hauptfiguren der Pionierzeit (Ziolkowski, Hohmann, von Braun, Sternfeld usw.) zu beschäftigen, da man hier den Perspektivenwechsel von der Raumfahrt als Nische für Phantasten zu einer der wichtigsten angewandten Wissenschaften der Gegenwart studieren kann.

Wir gehen von Gl. 5.35 aus, um den bielliptischen Transfer zu beschreiben.:

$$\Delta v(\rho) = v_{1,\mathrm{K}} \cdot \left((1-\rho) \cdot \sqrt{\frac{2}{1+\rho}} + \sqrt{\rho} - 1 \right). \tag{5.57}$$

Das bielliptische Manöver kann man sich so vorstellen, dass man mit dem ersten Schubmanöver weit über den Zielorbit hinausschießt, um dann mit dem zweiten Manöver den Zielorbit zu erreichen. Man kann sich diesem zweistufigen Übergang annähern, indem man im ersten Schritt einen Orbittransfer ins Unendliche durchführt und im zweiten Schritt von dort wieder zurückfliegt. Für den Transfer von der Kreisbahn 1 ins „Unendliche“ ($\rho \to \infty$) ergibt sich der auf $v_{1,\mathrm{K}}$ normierte Antriebsbedarf zu

$$\frac{\Delta v_1}{v_{1,\mathrm{K}}} = \sqrt{2} - 1. \tag{5.58}$$

Um vom „Unendlichen“ zur Kreisbahn 2 zu gelangen, wird der auf $v_{2,\mathrm{K}}$ normierte Antriebsbedarf

$$\frac{\Delta v_2}{v_{2,\mathrm{K}}} = \sqrt{2} - 1. \tag{5.59}$$

benötigt. Hier wurde ausgenutzt, dass aus Symmetriegründen der Antriebsbedarf vom Unendlichen nach r_2 genauso groß ist, wie der Antriebsbedarf von r_2 ins Unendliche. Der gesamte Antriebsbedarf (jetzt nur auf $v_{1,\mathrm{K}}$ normiert) ist dann durch die Summe der beiden Terme gegeben, also durch

$$\frac{\Delta v}{v_{1,\mathrm{K}}} = \frac{\Delta v_1}{v_{1,\mathrm{K}}} + \frac{\Delta v_2}{v_{1,\mathrm{K}}}. \tag{5.60}$$

Mit

$$\frac{\Delta v_2}{v_{1,\mathrm{K}}} = \left(\sqrt{(2)} - 1 \right) \frac{v_{2,\mathrm{K}}}{v_{1,\mathrm{K}}} = \left(\sqrt{2} - 1 \right) \sqrt{\rho} \tag{5.61}$$

folgt

$$\frac{\Delta v(\rho)}{v_{1,\mathrm{K}}} = \left(\sqrt{2} - 1 \right) (1 + \sqrt{\rho}). \tag{5.62}$$

Der Schnittpunkt der Funktionen nach Gl. 5.35 und Gl. 5.62 (siehe Abb. 5.17) markiert den Orbitwechsel, bei dem der Hohmann-Transfer weniger effizient wird als der bielliptische Orbitwechsel. Die Berechnung des Schnittpunktes der beiden Funktionen kann durch die Berechnung der Nullstelle der Differenzfunktion erfolgen. Das folgende Skript diese Nullstelle mit Hilfe der Bisektionsmethode:

```
1 import numpy as np
2 import sympy as sy
3 from scipy import optimize
```

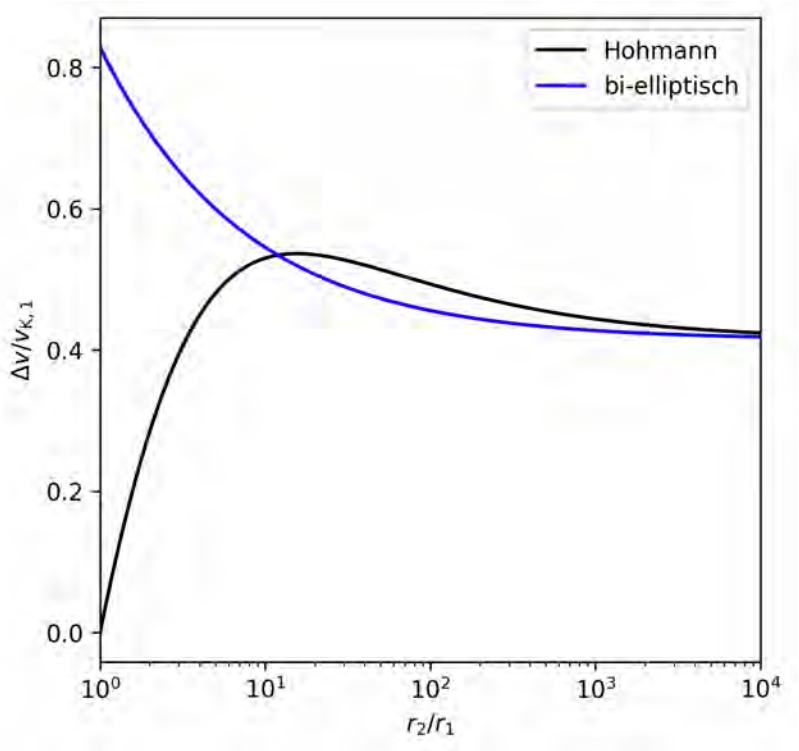


Abbildung 5.17: Auf die Geschwindigkeit $v_{1,\,\mathrm{K}}$ normierter Antriebsbedarf Δv in Abhängigkeit vom Verhältnis der Radien der Start- und Zielbahn. Es wird angenommen, dass beide Bahnen kreisförmig sind.

```
import matplotlib.pyplot as plt
def Bisektionsuchfunktion(f, x1, x2, tol=1e-8):
    n = 0
    if f(x1)*f(x2) < 0 and x1<x2:
        f_Links = f(x1)
        while True:

            x_3 = (x1 + x2)/2
            f_Rechts = f(x_3)
            n += 1
            if f_Rechts == 0 or (x2 - x1)/np.abs(x2 + x1) <
    tol:
                Nullstelle = x_3
                break
            elif f_Links*f_Rechts < 0:
                x2 = x_3
            else:
                x1 = x_3
                f_Links = f_Rechts
        Iterationsnummer = n
        return Nullstelle, Iterationsnummer
```

```
24      else:
25          print("Fehler (Die Bisektion kann keine Nullstelle
        im ggb. Intervall finden)")
26  f = lambda x: (1-x)*np.sqrt(2/(1+x)) + np.sqrt(x) - 1 - ((np
        .sqrt(2) - 1)*(1 + np.sqrt(x)))
27  nullstelle, itnum = Bisektionsuchfunktion( f , x1=1e-5, x2
        =1)
28  print("Nullstelle bei: ", nullstelle)
29  print("r_2 / r_1 = ", 1/nullstelle)
```

Das Skript berechnet die Nullstelle des Radienverhältnisses zu

$$r_2/r_1 = 11.939, \tag{5.63}$$

was dem in der Literatur üblichen Wert entspricht.

6

COMPUTERGESTÜTZTE NUMERISCHE MODELLIERUNGEN

Der Einsatz von Computern ist in der heutigen Wissenschaft nicht mehr wegzudenken. Experimente werden fast ausschließlich am Computer mittels *Computer Aided Design* (CAD) geplant, Messdaten in Mengen aufgenommen, die noch vor wenigen Jahren in kürzester Zeit große Festplatten gefüllt hätten, die heutzutage jeder in seinem Heimcomputer verbaut hat. Heutige Experimente am CERN produzieren beispielsweise Datenmengen in der Größenordnung von einem Petabyte (also 1024 Terabyte), pro Sekunde wohlgemerkt. Oftmals sind wissenschaftliche Experimente so einzigartig, dass man sich für Datenaufnahme und -auswertung seine eigenen Routinen mit Hilfe einer Programmiersprache schreiben muss. Es gibt natürlich hilfreiche Bibliotheken in jeder dieser Sprachen, die einem das Leben erleichtern und mit deren Hilfe man sich viele Routineaufgaben »zusammenklicken« kann, aber ohne grundlegende Kenntnis von Programmiersprachen kommt man heutzutage nicht mehr sehr weit in der Wissenschaft.

Ein sinnvoller Einsatz von Computern und oft auch der erste Kontakt mit Programmiersprachen im Studium ist die numerische Modellierung von physikalischen oder technischen Fragestellungen, die man entweder nur mühsam oder überhaupt nicht analytisch lösen kann. In diesem Kapitel sollen solche Beispiele exemplarisch durchgegangen werden, damit man frühzeitig im Studium einige dieser Methoden erlernt.

Aus didaktischen Gründen werden alle Beispiele in der Programmiersprache Python vorgestellt. Python gehört zu den Sprachen, die relativ einfach zu erlernen sind und viele sinnvolle Bibliotheken bereitstellen, z. B. für die graphische Darstellung von Ergebnissen. Es gibt unglaublich viele Tutorials zu Python im Internet, dazu viele Skripte diverser Universitäten, die frei verfügbar sind und mit deren Hilfe man sich in Python hinreichend intensiv einarbeiten kann, um die in diesem Skript gezeigten Beispiele nachvollziehen zu können.

Eine Programmiersprache lernt man eigentlich nur dadurch, indem man diese Sprache anwendet. Das mag anfangs mühsam sein, aber es lohnt sich. Bei Python kommt man dazu auch recht schnell zu brauchbaren Ergebnissen, die Lernkurve steigt in der Regel schnell an.

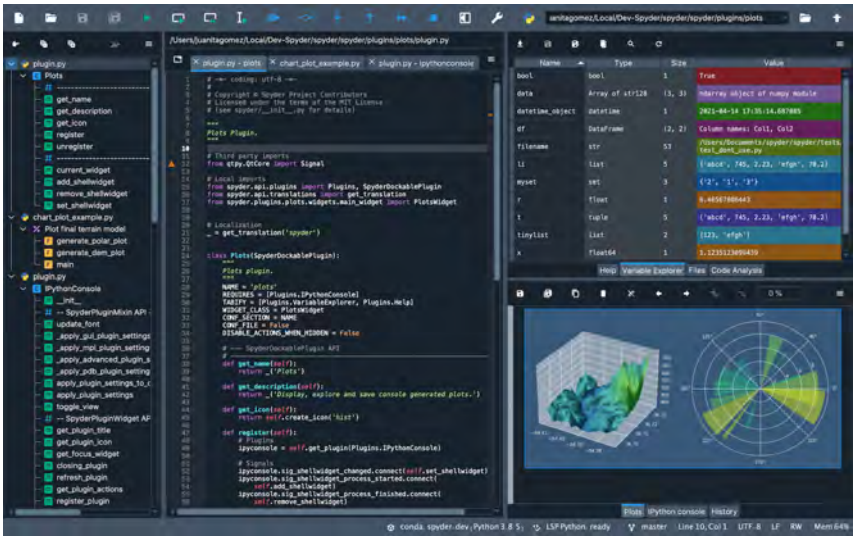


Abbildung 6.1: Screenshot der Entwicklungsumgebung SPYDER. Diese ermöglicht die Eingabe und Ausführung von Python-Code, die Überwachung von Variablen, der Ausgabe von Berechnungen in der integrierten Konsole und der Ausgabe von Grafiken.

6.1 Was man braucht, um loszulegen

Python lässt sich auf PCs mit Windows, Linux und macOS, bei letzterem muss man wohl bei der Installation darauf achten, welcher Chip verbaut ist (Intel oder M1). Es ist empfehlenswert, dass man gleich zu einer Python-Distribution greift, bei der nicht nur das »nackte« Python installiert wird, sondern auch gleich eine Reihe von wichtigen zusätzlichen Bibliotheken. Der Platzhirsch unter den Distributionen ist wohl Anaconda, welches neben Python auch noch die Programmiersprache *R* mitliefert, zusätzlich auch noch die Entwicklungsumgebung SPYDER, die den Einstieg erleichtert. Hier sind zudem die wichtigen Bibliotheken NUMPY, MATPLOTLIB, SCIPY und SYMPY sofort verfügbar.

6.2 Python

Python wurde von Guido van Rossum in der ersten Hälfte der 1990er Jahre entwickelt, die erste Vollversion erschien 1994. Seinen Namen verdankt es der englischen Komikertruppe Monty Python, hat also nichts mit der namensgleichen Schlange zu tun. Im Laufe der Jahre haben sich zwei getrennte Versionszweige (Version 2 und 3) entwickelt, die eine zeitlang parallel weiterentwickelt wurden. Mittlerweile wird Python 2 von der Python Software Foundation (das ist die Organisation,

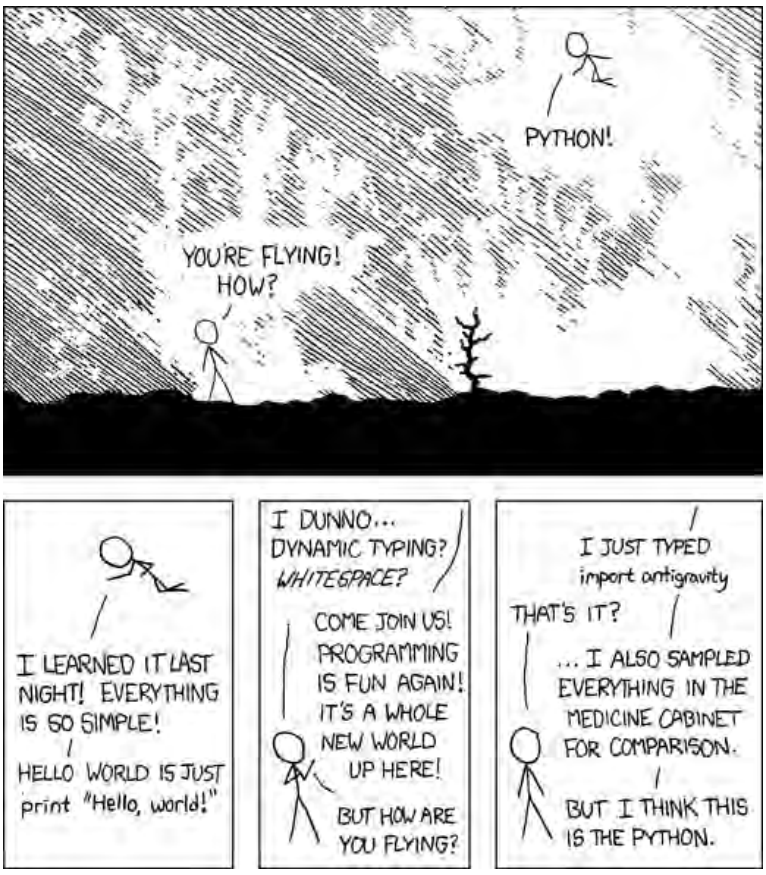


Abbildung 6.2: Ein xkcd-Comic über Python. Der dargestellte print-Befehl ist übrigens in Python 2 geschrieben und nicht mit dem print-Befehl der Version 3 kompatibel.

die hinter Python steht und auch die Rechte an Python hat) nicht mehr weiterentwickelt (die letzte Version ist Python 2.7.18 aus dem Jahr 2020). Wir verwenden in diesem Skript ausschließlich Python 3. Man sollte aber im Hinterkopf behalten, dass es im Internet viele Beispielcodes gibt, die im älteren Python geschrieben sind und in der Regel nicht kompatibel mit Python 3 sind.

6.3 Ein paar Grundlagen von Python

6.3.1 Das obligatorische Hallo-Welt-Programm

In guter alter Tradition soll auch hier mit der Ausgabe des Strings »Hallo Welt!« begonnen werden. Dieses Beispielprogramm dient hauptsächlich dazu zu prüfen, ob man Python richtig installiert hat. Der Befehl zur Ausgabe in der Konsole lautet print(), die Ausgabe wird in die Klammer geschrieben. Strings werden in Python in Gänsefüßchen gesetzt (wobei es egal ist, ob man " oder ' verwendet). Das Hallo-Welt-Programm sieht also folgendermaßen aus:

```
1 print("Hallo Welt!")
```

Man muss also keine Bibliotheken für die Ausgabe vorher importieren, sondern kann sofort loslegen. Der print-Befehl kann natürlich nicht nur Strings ausgeben, sondern im Prinzip alles, was man will, was folgende Zeilen verdeutlichen sollen:

```
1 a = 15   # definiert Variable a und weist Wert 15 zu
2 print(a) # Ausgabe des Werts der Variable
```

Wenn man mehrere Werte ausgeben muss, dann wird es oft schnell unübersichtlich. Man kann daher die Ausgabe mit einem String kombinieren, der angibt, welche Variable gerade gezeigt wird:

```
1 a = 15
2 print("Wert der Variable a: ", a)
```

Man kann auch mehrere Variablen auf einmal ausgeben:

```
1 a = 15
2 b = 20
3 print("Wert a: ", a, " Wert b: ", b)
```

Das Leerzeichen vor Wert b dient dazu, dass zwischen dem Zahlenwert von a und dem nachfolgenden Wort ebenfalls ein Leerzeichen einzufügen, damit die Lesbarkeit erhalten bleibt. Alternativ kann man auch die üblichen Metazeichen für Zeilenvorschub und Tabulator verwenden:

```
1 a = 15
2 b = 20
3 c = 0.1
4 print("Wert a: ", a, "\tWert b: ", b, "\nWert c: ", c)
```

Halten wir ein paar Unterschiede zu anderen Programmiersprachen fest:

- bei Python muss man die Art der Variable nicht festlegen. Schreibt man beispielsweise `a = 10`, dann erkennt Python, dass man eine Integer-Zahl eingegeben hat. Schreibt man `a = 10.0`, dann

wird automatisch eine Gleitkommazahl erkannt. Dieses Prozedere wird als dynamische Typisierung bezeichnet. Teilt man z. B. zwei Integer-Werte so, dass eine Gleitkommazahl als Ergebnis herauskommt, dann wandelt Python das Ergebnis entsprechend um (man kann das verhindern, wenn man zum Teilen statt `/` den Operator `//` verwendet).

- es gibt kein Befehlsabschlusszeichen wie in C++, also kein `;`
- jede Zeile wird nacheinander abgearbeitet, so dass man zur Festlegung von Variablen einfach untereinander schreibt, wie in den gezeigten Beispielen. Will man mehrere Variablen in einer Zeile festlegen, dann geht das natürlich auch:

```
a,b,c = 15,20,0.1
print("Wert a: ", a, "\tWert b: ", b, "\nWert c: ", c)
```

6.3.2 Funktionen am Beispiel des modifizierten Julianischen Kalenders

Funktionen sind ein nützliches Hilfsmittel, um bestimmte Anweisungen und Routinen im Code wiederholt auszuführen. Das kann beispielsweise die Durchführung einer mathematischen Operation sein, aber auch die graphische Ausgabe in einer gewünschten Darstellung. Eine Funktion wird mit dem Befehl `def` gefolgt von einem Funktionsnamen und runden Klammern definiert:

```
def funktionsname():
    tue etwas...
```

Die runden Klammer sind charakteristisch für Funktionen. So ist z. B. auch der `print()`-Befehl nichts anderes als eine Funktion, an die eine Variable, ein String oder ein Boolescher Wert übergeben und auf der Konsole ausgegeben wird. Wenn eine Funktion zwei Variablen a und b addieren soll, dann werden die übergebenen Variablen in die Klammer geschrieben:

```
def addiere(a,b):
    c = a + b
    return c
```

Es ist dabei völlig unerheblich, welche Namen die Variablen außerhalb der Funktion hatten. Es kommt nur auf die Reihenfolge an, in der sie an die Funktion übergeben werden. Die Variablenbezeichnungen innerhalb der Funktion sind unabhängig von der Bezeichnung außerhalb der Funktion. Mit `return` kann die Funktion angewiesen werden, einen Wert zurückzugeben (hier das Ergebnis der Addition). Was auffällt ist,

dass die Befehlszeilen in der Funktion eingerückt sind. SPYDER wird dies automatisch durchführen und Python verlangt diese Einrückung. Die Idee ist es, dass der Code dadurch lesbarer werden soll. Das Einrücken mag am Anfang ungewohnt sein, ist aber sinnvoll und man gewöhnt sich schnell daran. Eine Funktion kann man einfach durch ihren Namen (und den benötigten Variablen) aufrufen, wie folgendes Beispiel zeigen soll. Hier wird das Ergebnis der Addition auch gleich in einer Variable gespeichert, die dann über `print()` ausgegeben wird.

```
def addiere(a,b):
    c = a + b
    return c
addition = addiere(1,2)
print(addition)
```

Natürlich ist eine Funktion, die zwei Zahlen addiert, nur ein Minimalbeispiel, um die Verarbeitungweise zu verstehen. Python kann das Addieren natürlich auch ganz ohne Funktion durchführen. Ein etwas komplexeres Beispiel sei im folgenden Code gegeben. Hier geht es um die Umrechnung des »normalen« Datums in das modifizierte julianische Datum, abgekürzt als MJD. Das MJD ist eine in der Astronomie übliche Tagezählung, die Zeit wird dabei in Tagen und Tagesbruchteilen angegeben. Das erleichtert beispielsweise die Berechnung von Zeitdifferenzen enorm. Der Code dafür ist allerdings etwas komplizierter und soll hier gar nicht in jedem Detail durchgegangen werden (für uns ist das hier einfach eine Funktion, die mehrere Eingaben wie Jahr (Y), Montat (M), Tag (D), Stunde (H), Minute (Min) und Sekunde (S) verlangt und eine Ausgabe (das MJD) hat), es kommt uns auf die Umsetzung als Code an:

```
import numpy as np
def calc_MJD(Y, M, D, H, Min, S):
    if M > 2:
        Y = Y
    else:
        Y = Y - 1
        M = M + 12
    D = D + H/24 + Min/1440 + S/86400
    JD = np.floor(365.25*(Y + 4716)) + np.floor(30.6001*(M
    +1)) + D + 2 - np.floor(Y/100) + np.floor(Y/400) -
    1524.5
    MJDate = JD - 2400000.5
    return MJDate
summertime = calc_MJD(2021, 3, 28, 2, 0, 0)
print(summertime)
```

Für die Berechnung des MJD wird eine spezielle Funktion namens `floor` gebraucht. Was genau diese Funktion mit einer Zahl macht, können Sie sich in Python selbst anschauen. Wichtig ist hier, dass diese Funktion

nicht Teil der Standardroutinen von Python ist. Wir binden diese Funktion daher über eine Bibliothek namens NUMPY ein, was für *Numerical Python* steht. Damit Python weiß, dass die Funktion aus NUMPY verwendet werden soll, muss man vor den Funktionsnamen entweder `numpy` schreiben. Alternativ kann man auch ein Kürzel verwenden, üblich ist die Abkürzung `np`.

6.3.3 Graphische Ausgabe mit Matplotlib

Häufig will man seine Ergebnisse graphisch ausgeben. In Python gibt es dazu eine Reihe von Bibliotheken, die dies ermöglichen. Die Bibliothek MATPLOTLIB gehört hierbei zu den bekanntesten Bibliotheken und ist bei einer Anaconda-Installation automatisch dabei. Die Einbindung erfolgt dabei meist über:

```
1 import matplotlib.pyplot as plt
```

Es soll nun ein Plot mit einer Sinus- und einer Cosinus-Funktion erstellt werden. Diese beiden Funktionen müssen über eine Zusatzbibliothek integriert werden, beispielsweise über NUMPY. Danach muss eine passende x-Achse definiert werden. Hier gibt es verschiedene Möglichkeiten. Über NUMPY kann mit der Funktion `linspace` ein Array von Zahlen erstellt werden:

```
1 import matplotlib.pyplot as plt
2 import numpy as np
3 x = np.linspace(0,10,100)
```

Das Array enthält Zahlen von 0 bis 10, die Anzahl der Einträge beträgt 100. Das erzeugte Array hat ein spezielles Format, welches von NUMPY erzeugt und verwendet wird. Übergibt man dieses Array nun an eine NUMPY-Funktion, so wird automatisch ein neues Array gleicher Länge erzeugt, in dem zu jedem x-Wert der errechnete Funktionswert eingetragen wird. Wir nennen dieses neue Array $y1$ für die Sinus- und $y2$ für die Cosinus-Funktion und können danach mit der grafischen Ausgabe beginnen.

```
1 import matplotlib.pyplot as plt
2 import numpy as np
3 x = np.linspace(0,10,100)
4 y1 = np.sin(x)
5 y2 = np.cos(x)
6 plt.figure(figsize = (7,5), dpi=300)
7 plt.plot(x, y1, label = "Sinus", color = "blue")
8 plt.plot(x, y2, label = "Cosinus", color = "red")
9 plt.xlabel("$x$-Achse")
10 plt.ylabel("$y$-Achse")
11 plt.legend()
```

```
12 plt.show()
```

Hier sind bereits eine Reihe von zusätzlichen Eingabeparametern untergebracht, die man nicht alle defaultmäßig eingeben muss. Die Eingabe `plt.figure()` ist hier die minimale Variante. Sie erzeugt ein *figure*-Objekt, in welches die Arrays geplottet werden können. In die Klammer kann nun eine Reihe von Ergänzungen untergebracht werden. Meist ist die Standardeinstellung recht grobpixelig, weswegen `dpi = 300` die *dots-per-inch* auf 300 erhöht. `figsize(7,5)` ändert die Größe der Abbildung in vertikaler und horizontaler Richtung. Bei Abbildungen ist meistens die x-Richtung etwas größer, empfehlenswert ist ein Verhältnis ähnlich des Goldenen Schnitts. Manchmal ist es aber auch sinnvoll, dass beide Achsen gleich skalieren, beispielsweise, wenn man ein kreisförmiges Objekt darstellen will. Der folgende Beispielcode erzeugt die bereits bekannten beiden Funktionen sowie weitere Plots. Spielen Sie mit den Parametern und schauen sich an, was sich für Auswirkungen ergeben.

```
1  import matplotlib.pyplot as plt
2  import numpy as np
3
4  x = np.linspace(0,10,100)
5  y1 = np.sin(x)
6  y2 = np.cos(x)
7
8  plt.figure(figsize = (7,5), dpi=300)
9  plt.plot(x, y1, label = "Sinus", color = "blue")
10 plt.plot(x, y2, label = "Cosinus", color = "red")
11 plt.xlabel("$x$-Achse")
12 plt.ylabel("$y$-Achse")
13 plt.legend()
14 plt.show()
15
16 plt.figure(figsize = (7,5), dpi=300)
17 plt.plot(y1, y2, label = "Sinus-Cosinus", color = "blue")
18 plt.xlabel("$x$-Achse")
19 plt.ylabel("$y$-Achse")
20 plt.legend()
21 plt.show()
22
23 plt.figure(figsize = (5,5), dpi=300)
24 plt.plot(y1, y2, label = "Sinus-Cosinus", color = "blue")
25 plt.xlabel("$x$-Achse")
26 plt.ylabel("$y$-Achse")
27 plt.legend()
28 plt.show()
29
30 plt.figure(figsize = (5,5), dpi=300)
31 plt.plot(x, 1- np.exp(-x), label = "saturation", color = "
       blue")
32 plt.xlabel("$x$-Achse")
```

```
33 plt.ylabel("$y$-Achse")
34 plt.legend()
35 plt.show()
```

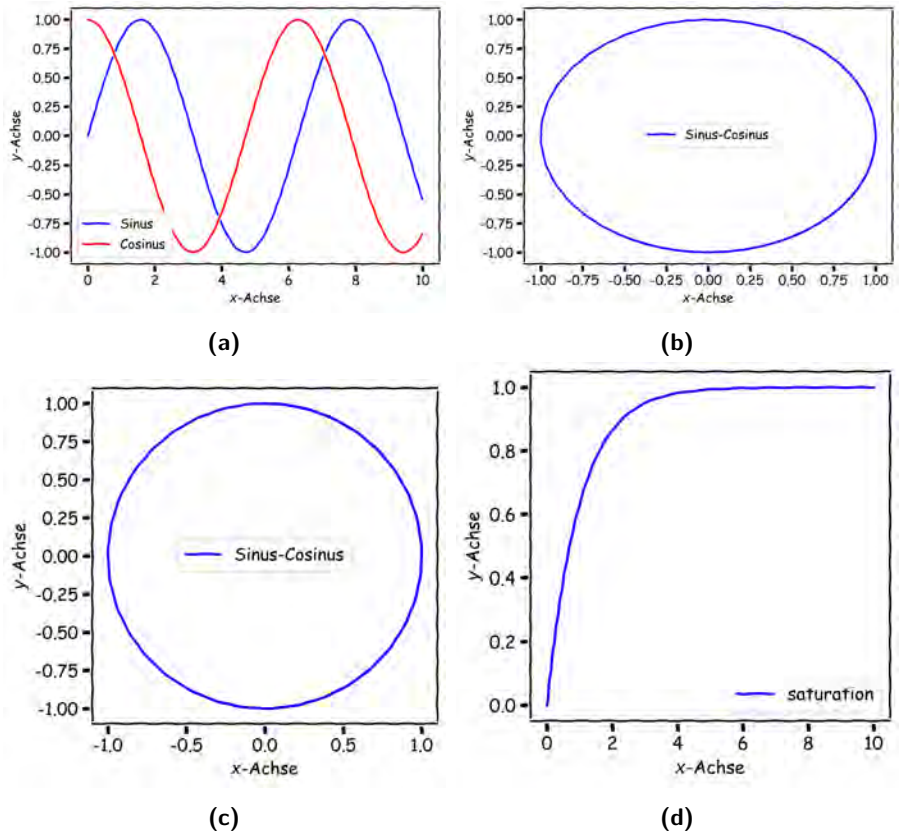


Abbildung 6.3: Mit MATPLOTLIB erzeugte Plots. Die Plots (a) und (b) sind bis auf ihre Achsenskalierung identisch. Hier sieht man gut, dass man das Seitenverhältnis von Plots immer im Auge behalten sollte.

Wie man sieht, übernimmt MATPLOTLIB die Achsenskalierung automatisch, sofern keine individuellen Einstellungen vorgenommen werden. Natürlich kann man so einen Plot nach belieben anpassen. Die Dollarzeichen bei den Achsenbezeichnungen, also `$x$-Achse` führen dazu, dass x und y als kursive Variablen gesetzt werden, was bei Variablen üblich ist. Die Dollarnotation ist der LaTeX-Schreibweise entnommen, MATPLOTLIB kann LaTeX-Syntax in Teilen umsetzen.

6.3.4 Listen

Wir haben bereits die Verwendung von NUMPY-Arrays kennengelernt. Python verwendet von Haus aus eine andere Art von Sammelobjekt,

die sogenannte Liste. Eine (leere) Liste wird relativ einfach mit eckigen Klammern erzeugt:

```
meine_Liste = []
```

In diese Liste können nun beliebige Dinge geschrieben werden. Dabei muss es sich auch nicht unbedingt um Zahlen handeln. Will man eine Zahl anhängen, verwendet man die `.append()`-Methode:

```
meine_Liste = []
meine_Liste.append(5.0)
meine_Liste.append("Hallo Welt!")
meine_Liste.append(True)
print(meine_Liste)
```

Die Ausgabe ist `[5.0, 'Hallo Welt!', True]`. Neben diesen Listen gibt es in Python noch ein Sammelobjekt namens *Dictionary*. Der Unterschied zur Liste besteht in der Art der Indizierung. Eine Element einer Liste kann man mit `meine_Liste[Index]` auswählen. Der Index ist dabei eine Zahl, wobei man darauf achten muss, dass die Indizierung bei Python wie bei vielen anderen Programmiersprachen auch bei null anfängt. Die Indizierung eines Dictionaries läuft über einen String. Das mag in gewissen Situationen Vorteile haben, beispielsweise wenn man ein Wörterbuch erstellt, woher wohl auch der Name zu stammen scheint. Erzeugt wird es mit geschweiften Klammern:

```
mein_Dictionary = {}
mein_Dictionary['null'] = 'zero'
```

Wenn man nun wissen will, was unter dem Index `null` zu finden ist, wird Python `zero` ausgeben. Auf diese Weisen lassen sich eben Wörterbücher erzeugen. Sicherlich gibt es auch noch andere sinnvolle Verwendungen eines solchen Dictionaries. Persönlich habe ich diesen Datentypen aber noch nie verwendet, weswegen ich hier auch nicht weiter darauf eingehe.

6.3.5 Schleifen

Python kennt `for`- und `while`-Schleifen. Wir haben mit Hilfe von NUMPY in einem der vorigen Beispiele ein Array von Zahlen erstellt. Mit dem Kommando `x = np.linspace(0,10,11)` wird beispielsweise ein Array von 0 bis 10 in Abständen von 1 erzeugt. Dies wollen wir nun mit einer Liste durchführen, wofür wir Schleifen verwenden können. Natürlich kann man bei kurzen Listen einfach per Hand die Zahlen eintragen, also

```
meine_Liste = [0,1,2,3,4,5,6,7,8,9,10]
```

verwenden, aber das wird bei großen Listen schnell ineffizient. Mit einer `for`-Schleife wird das deutlich effizienter:

```
meine_Liste = []
for i in range(0,11):
    meine_Liste.append(i)
```

Wir definieren hier einen Iterator `i`, der von 0 bis 10 läuft und hängen diesen Iterator an die Liste in jedem Schleifendurchgang an. Die Ausgabe von Python für diese Liste ist `[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]`. Will man die gleiche Ausgabe mit einer `while`-Schleife erzeugen, muss man eine Abbruchbedingung der Schleife definieren. Eine Möglichkeit wäre wie folgt:

```
meine_Liste = []
i = 0
while i <= 10:
    meine_Liste.append(i)
    i += 1
```

Hier muss man den Iterator (`i = 0`) außerhalb der Schleife definieren und händisch bei jedem Schleifendurchgang um 1 erhöhen (`i += 1`). Die Ausgabe ist identisch mit der der `for`-Schleife.

6.3.6 Liste oder Numpy-Array

Wir haben gesehen, dass wir die Sinuswerte eines NUMPY-Arrays recht einfach ausrechnen und in ein neues Array schreiben können. Mit einer Liste geht das nicht so einfach. Hier muss man in jedem Schleifendurchgang den spezifischen Funktionswert berechnen und in eine eigene Liste anhängen. NUMPY macht bei seiner Berechnung letztlich auch nichts anderes, allerdings ist der Syntax etwas einfacher und – da NUMPY eine in C geschriebene Python-Bibliothek ist – auch deutlich schneller. Letztlich führt beides zum gleichen Ergebnis und eine Möglichkeit, die beiden Funktionen Sinus und Cosinus mit Listen zu erzeugen und graphisch darzustellen, könnte folgendermaßen aussehen:

```
import numpy as np
import matplotlib.pyplot as plt

x = []
y1 = []
y2 = []

for i in range(0,101):
    x_wert = i/10
    x.append(x_wert)
    y1_wert = np.sin(x_wert)
    y1.append(y1_wert)
    y2_wert = np.cos(x_wert)
    y2.append(y2_wert)
```

```
15
16 plt.figure(figsize = (7,5), dpi=300)
17 plt.plot(x, y1, label = "Sinus", color = "blue")
18 plt.plot(x, y2, label = "Cosinus", color = "red")
19 plt.xlabel("$x$-Achse")
20 plt.ylabel("$y$-Achse")
21 plt.legend()
22 plt.show()
```

Natürlich gibt es auch eine Vielzahl anderer Möglichkeiten, die gleiche Ausgabe zu bekommen. Man kann Listen auch in NUMPY-Arrays umwandeln und umgekehrt, so dass man je nach Bedarf beide Möglichkeiten kombinieren kann.

Schauen wir aber zur Sicherheit uns die beiden Datentypen nochmal genauer an. Wir definieren uns zunächst eine Python-Liste und ein NUMPY-Array und geben uns beides nacheinander auf der Konsole aus:

```
1 import numpy as np
2 meine_liste = [0, 1, 2, 3, 4]  # Python-Liste
3 mein_array = np.array([0, 1, 2, 3, 4])  # NumPy-Array
4 # Hier die Ausgabe
5 print(meine_liste)
6 print(mein_array)
```

Das Ergebnis sieht so aus:

```
1 [0, 1, 2, 3, 4]
2 [0 1 2 3 4]
```

Der visuell einige Unterschied besteht in den Kommata bei den Listen. Wir schauen uns nun an, was passiert, wenn man die Liste mit der Zahl 2 »multipliziert«:

```
1 meine_liste = [0, 1, 2, 3, 4]  # Python-Liste
2 print(meine_liste*2)
```

Die Ausgabe aus der Konsole:

```
1 [0, 1, 2, 3, 4, 0, 1, 2, 3, 4]
```

Intuitiv hätte man vielleicht erwartet, dass jeder Eintrag der Liste mit 2 multipliziert wird. Dem ist aber nicht so. Es wird stattdessen einfach die gleiche Liste nochmal angehängt. Das gleiche Ergebnis hätte man auch bekommen, wenn man die Liste mit sich selbst addiert hätte. Um jeden Eintrag der Liste mit zwei zu multiplizieren, muss man über eine Schleife auf jeden einzelnen Eintrag zugreifen, z. B. so:

```
1 meine_liste = [0, 1, 2, 3, 4]  # Python-Liste
2 for i in range(len(meine_liste)):
3     meine_liste[i] = meine_liste[i]*2
4 print(meine_liste)
```

Prüfen Sie das Ergebnis der Ausgabe. Das Kommando `len()` gibt die Anzahl der Einträge der Liste zurück. Das ist praktisch, da man nicht per Hand nachzählen muss.

In NUMPY führt die Multiplikation

```
import numpy as np
mein_array = np.array([0, 1, 2, 3, 4])
print(mein_array*2)
```

zu

```
[0 2 4 6 8]
```

wie man es von der Vektormultiplikation kennt. Arrays in NUMPY sind also mächtige Werkzeuge und den gewöhnlichen Listen in Python weit überlegen. Arrays lassen sich über `np.linspace()` und `np.arange()` anlegen. Die Details dazu findet man in unzähligen Tutorials und auch in der offiziellen NUMPY-Dokumentation (https://numpy.org/doc/).

6.4 Beispielcodes

Mit den gezeigten Grundlagen lässt sich schon eine ganze Menge anstellen. Die folgenden Beispiele sind geeignet, sich in Python zu vertiefen und nebenbei auch etwas über numerische Methoden zu lernen.

6.4.1 Monte-Carlo Integration

Es kommt vor, dass man eine Funktion integrieren muss, aber die Stammfunktion nicht einfach zu finden ist. Es kommt auch vor, dass es manchmal gar keine (analytische) Stammfunktion gibt. Nehmen wir die Funktion

$$f(x) = \frac{2}{\sqrt{x}} \int_0^\alpha e^{-x^2} \, \mathrm{d}x \tag{6.1}$$

als Beispiel. In Lehrbüchern findet man die Information, dass die Fehlerfunktion erf(x) eine Stammfunktion dieses Integrals sein soll. Das ist richtig, aber auch etwas verwirrend, denn eigentlich ist erf(x) nur eine abkürzende Schreibweise für das Integral selbst. Diese Funktion ist nämlich nicht als geschlossener Ausdruck darstellbar, sondern kann nur numerisch bestimmt werden. Wir wollen daher das Integral mit Hilfe der Monte-Carlos-Integration numerisch berechnen. Die Vorgehensweise ist denkbar einfach: Wir zeichnen die Funktion in einem festgelegten x-Intervall. Der maximale Funktionswert legt die y-Achse fest. Nun erzeugen wir Zufallszahlen, die für Koordinaten (x, y) stehen und schauen, ob unser Koordinatenwert y größer oder kleiner gleich dem Funktionswert ist. Ist er größer, dann zählen wir dies als »Fehlwurf«,

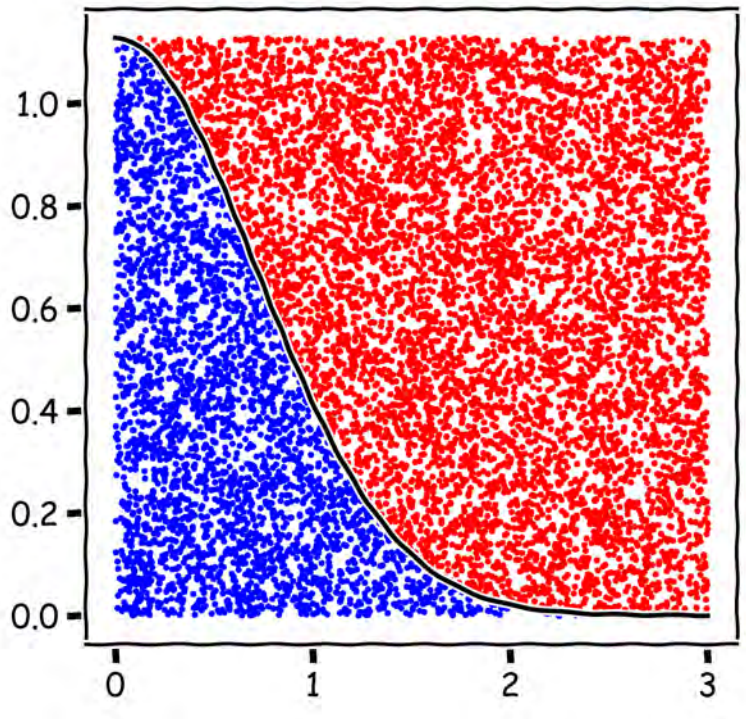


Abbildung 6.4: Berechnung des Integrals von $f(x) = \frac{2}{\sqrt{x}} \int_0^\alpha e^{-x^2}\,\mathrm{d}x$ für $\alpha = 3$ mit Hilfe der Monte-Carlo Methode. Liegt ein Koordinatenwert unterhalb des Funktionswerts, wird die Koordinate blau gezeichnet, oberhalb des Funktionswerts entsprechend rot.

liegt der Wert innerhalb der Integralfläche, dann bezeichnen wir das als ≫Treffer≪. Wir führen diese Zufallsbetrachtung für eine Vielzahl von Zufallskoordinaten durch und berechnen das Verhältnis der Treffer zur Gesamtzahl der Zufallskoordinaten (Trefferquote). Da wir die Fläche des Rechecks kennen, das durch unserer Achsen aufgespannt wird (maximaler x-Wert multipliziert mit dem maximalen Funktionswert y), entspricht die Fläche unter unserer Funktion dieser Gesamtfläche multipliziert mit der Trefferquote. Der folgende Code führt diese Integration durch und gibt den errechneten Wert aus. Zudem wird auch eine anschauliche Abbildung in einem etwas anderen Layout erzeugt (vgl. Abbildung 6.4).

```
1 import random
2 import matplotlib.pyplot as plt
3 import numpy as np
4
5 werte = []
6 nr = 0
7 hit = 0
8
```

```
X_hit = []
Y_hit = []

X_no_hit = []
Y_no_hit = []

def myfunction(variable):
    funktionswert = 2/np.sqrt(np.pi)*np.exp(-x*x)
    return funktionswert

x_max = 3
for i in range (1,10000):
    nr = nr + 1
    x = random.random()*x_max
    y = random.random()*2/np.sqrt(np.pi)
    if y <= myfunction(x):
        hit = hit + 1
        X_hit.append(x)
        Y_hit.append(y)
    else:
        X_no_hit.append(x)
        Y_no_hit.append(y)
full_area = 2/np.sqrt(np.pi)*x_max
area = full_area*hit/nr
my_x = np.linspace(0,x_max,100)
my_y = 2/np.sqrt(np.pi)*np.exp(-my_x*my_x)
print("Integral = ", area)
plt.xkcd()
plt.figure(figsize = (5,5), dpi = 300)
plt.scatter(X_no_hit, Y_no_hit, color = "red", s = 3)
plt.scatter(X_hit, Y_hit, color = "blue", s = 3)
plt.plot(my_x, my_y, color = "black")
plt.show()
```

Der tabellierte Wert für erf(3) beträgt 0.99997791. Das entspricht in etwa dem Ergebnis unseres Codes, wobei man darauf achten muss, dass hinreichend viele Versuche durchgeführt werden müssen. Bei großen Mengen an Zufallszahlen ist es ratsam, die graphische Ausgabe zu deaktivieren, da die Listen sehr groß werden und die Ausgabe als Abbildung langsam wird.

6.4.2 Gedämpfter harmonischer Oszillator

Der gedämpfte harmonische Oszillator für eine eindimensionale Bewegung in x-Richtung lässt sich in der Form

$$\ddot{x} + 2D\dot{x} + \omega^2 x = 0 \tag{6.2}$$

schreiben. Der mittlere Term beschreibt dabei die auftretende Reibungsbeschleunigung, D wird auch als Abklingkonstante bezeichnet.

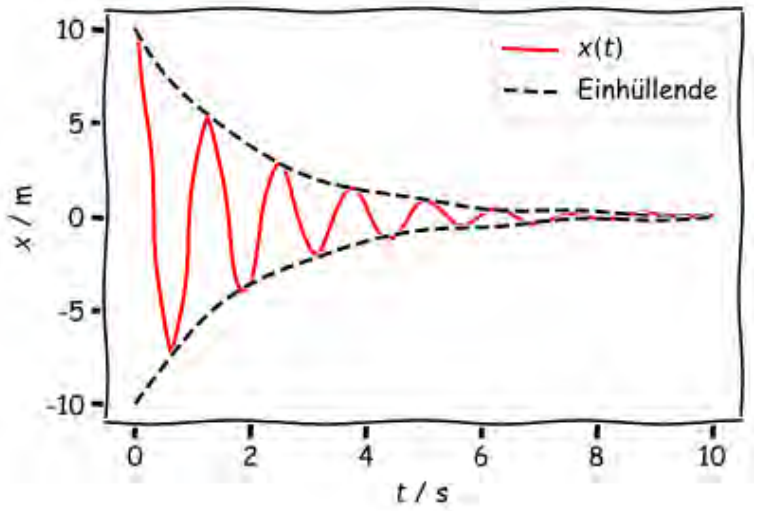


Abbildung 6.5: Der gedämpfte harmonische Oszillator.

Man kann eine solche Differentialgleichung mit relativ wenig Aufwand numerisch lösen. Verwenden wollen wir dazu das sogenannte Euler-Verfahren, welches zu den einfachsten Lösungsverfahren gehört. Man zählt es zu den Runge-Kutta-Verfahren, die eine ganze Klasse von ähnlichen Lösungsmethoden zusammenfassen. Beim Euler-Verfahren nutzt man aus, dass wir innerhalb eines Intervalls $\mathrm{d}x$ die Veränderung der Größen bestimmen können. Wir wissen ja, dass sich die Geschwindigkeit über die Beziehung $v = a\,\mathrm{d}t = \ddot{x}\,\mathrm{d}t$ ändert. Für den Weg x gilt analog $x = v\,\mathrm{d}t$. Wir machen also folgendes: Zunächst müssen wir unsere Randbedingungen anschauen, diese sind zwingend erforderlich, wenn man eine Differentialgleichung numerisch lösen will. Wir machen die Annahme, dass bei $t = 0$ das System um $x = 10$ Einheiten ausgelenkt ist und keine Anfangsgeschwindigkeit habe. Zu dieser Zeit wirkt eine Beschleunigung $a = -2D\dot{x}_0 - \omega^2 x$, wobei ja $\dot{x}_0 = 0$ aus den Randbedingungen folgt. Damit rechnen wir also die Beschleunigung a zu dieser Zeit aus und berechnen die Geschwindigkeit v_{t_1}. Mit dieser Geschwindigkeit wird das System nun in einem kleinen Zeitintervall sich bewegen und dabei die Strecke $x_1 = v_1\,\mathrm{d}t$ zurücklegen. Die beiden Werte x_1 und v_1 schreiben wir jeweils in eine Liste (in der wir auch schon unsere Anfangswerte geschrieben haben). Dieses Spiel wiederholen wir jetzt ausgehend von unserem neuen Koordinatenpunkt. Da das System jetzt eine Geschwindigkeit ungleich Null hat, wird nun auch der Reibungsterm wichtig. Wenn wir unser Zeitintervall $\mathrm{d}t$ hinreichend klein wählen, dann approximieren wir unsere Lösungsfunktion (also die gesuchte Bewegungsgleichung) durch lineare Teilstücke, erhalten somit

also eine numerische Lösung des gedämpften harmonischen Oszillators. Der folgende Code führt diese Berechnung durch. Da man eine Vielzahl von kleinen Teilstücken berechnet und addiert, ist eine Berechnung über eine Schleife unabdingbar.

```
import matplotlib.pyplot as plt
import numpy as np

t_max = 10
x_max = 10
t = 0
dt = 0.01
v = 0
x = 10
omega = 5

D = 0.5 # Abklingkonstante

num_points = int(t_max/dt)

X = []    # x-Werte
V = []    # v-Werte
T = []    # t-Werte
i = 0

while i <= num_points:
    print("Zeit: ", t, "\t Ort: ", x, " \t Geschwindigkeit: ", v)
    X.append(x)
    V.append(v)
    T.append(t)
    v += -omega**2 * x * dt - 2*D*v*dt
    x += v*dt
    t += dt
    #print(i)
    i += 1
    #if i == num_points:
        #break
t1 = np.linspace(0, t_max, 1000)
damping = x_max*np.exp(-D*t1)
plt.figure()
plt.plot(T,X, color = "red", label = "$x(t)$")
plt.plot(t1, damping, color = "black", linestyle = "dashed", label = "Einhuellende")
plt.plot(t1, -damping, color = "black", linestyle = "dashed")
plt.xlabel("$t$ / s")
plt.ylabel("$x$ / m")
plt.legend()
plt.show()
```

Mit diesem Code wird auch zusätzlich die aus der analytischen Lösung sich ergebende Einhüllende gezeichnet, damit man die numerische Berechnung gleich überprüfen kann. Dieses Lösungsverfahren lässt sich nicht nur für den harmonischen Oszillator anwenden, sondern für beliebige Bewegungen von Körpern unter dem Einfluss von Kräften. Auch komplexere Berechnungen wie beispielsweise der Bewegung von Teilchen in einem Plasma werden auf Teilchenebene mit ähnlicher Herangehensweise numerisch modelliert. Man verwendet dann zwar Runge-Kutta-Verfahren höherer Genauigkeit und noch ein paar weitere Methoden wie das Leapfrog-Verfahren, um numerische Problematiken elegant zu lösen, aber der Grundgedanke ist unserer hier durchgeführten Rechnung sehr ähnlich.