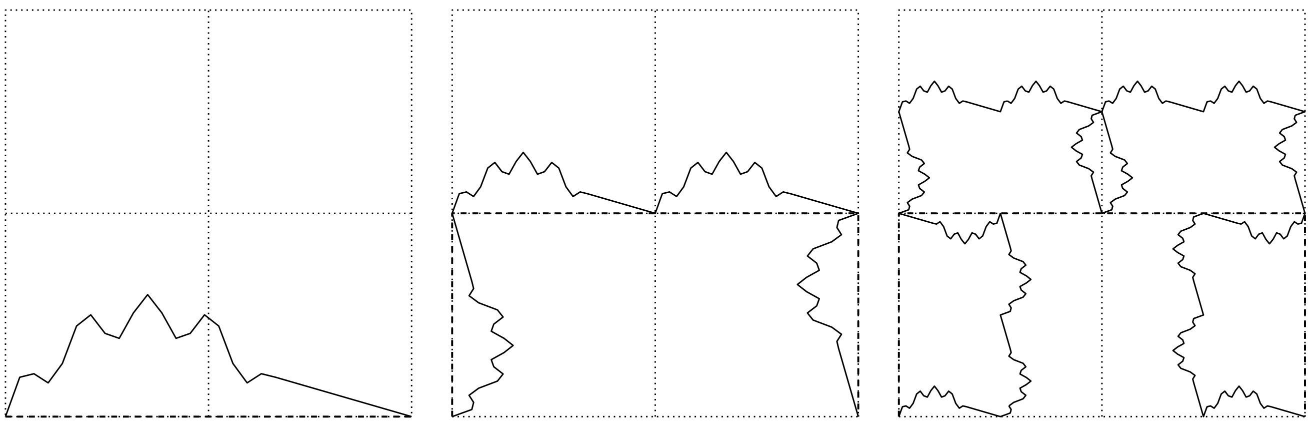
Analysis mit gleichmäßigen Fehlerschranken

H. Karcher, Version Okt. 2002

Dieser Text ist eine Ausarbeitung des Analysisteils meiner Vorlesung Mathematik I im WS1999/2000 in Bonn. Die Vorlesung Analysis I im WS2002/2003 wurde nach diesem Skript gehalten.

Kapitel



Vorwort

Details aus diesem Manuskript habe ich in all meinen Anfängervorlesungen ausprobiert. Das hier vorgestellte Gesamtkonzept weicht ein Semester lang stark von dem Standardaufbau ab, mündet aber mit Beginn des zweiten Semesters wieder in den Standardaufbau ein. Nur im WS99/00 habe ich die Erstsemestervorlesung konsequent nach diesem Plan gehalten. Der Anhang enthält eine 1:10 komprimierte (englische) Kurzfassung. Da ich die Stetigkeit erst am Ende des ersten Semesters behandele, wird sehr schnell angenommen, ich wolle die Rolle der Stetigkeit reduzieren. Ich werde durch Prüfungsergebnisse in der Ansicht bestärkt, daß ich sogar ein besseres Verständnis der Stetigkeit erreiche, weil sie erst zu einem Zeitpunkt behandelt wird, in dem schon eine ganze Reihe von Erfahrungen mit Analysisargumenten und Analysiszielen gesammelt sind.
Ich versuche, als zentrale Strategie in den Vordergrund zu rücken, daß neue Objekte der Analysis gewonnen werden, indem ("konvergente") Approximationen aus bereits bekannten Objekten konstruiert werden - irrationale Zahlen werden durch rationale Zahlen approximiert, neue Funktionen wie exp, sin, cos werden durch Polynome oder rationale Funktionen approximiert. Dieses "Konstruieren durch Approximation" ist auf der einen Seite erstaunlich leistungsfähig, auf der anderen Seite erfordert es sorgfältiges Argumentieren, um z.B. für neue, nur approximierbare Funktionen deren Ableitungen zu bestimmen. Die Bedeutung dieser neuen Funktionen innerhalb und außerhalb der Mathematik liegt so weitgehend an ihren Ableitungseigenschaften, daß ein Vorverständnis des Begriffs Ableitung notwendig ist, ehe wir uns ihnen zuwenden können. Diese Vorverständis wird an den Polynomen erworben.
Dies Manuskript beginnt mit zwei vorbereitenden Kapiteln. Zuerst werden an neun Beispielen absolute und relative Fehler besprochen. Das letzte Beispiel diskutiert die Kreistangente. Die beobachteten Eigenschaften erlauben, auch für quadratische Parabeln Tangenten zu definieren und wichtige Eigenschaften (wie Hohlspiegel) zu erklären.
Die Argumente sind verallgemeinerungsfähig und erlauben, in Kapitel 3 Polynome zu differenzieren. Die Abweichung der Polynome von ihren Tangenten ist ein wichtiges Detail. Die beobachtete quadratische Abweichung ist die Grundlage der Differenzierbarkeitsdefinition, mit expliziten Fehlertermen. Damit werden in Kapitel 4 die Differentiationsregeln bewiesen. In Kapitel 5 tritt der auch weiterhin zentrale Monotoniesatz zum ersten Mal auf. Er kann ohne die Vollständigkeit bewiesen werden, da alle bisher behandelten Funktionen es rechtfertigen, gleichmäßige Fehlerschranken vorauszusetzen. Die Anwendungen des Monotoniesatzes gehen schon bei Polynomen schnell über das hinaus, was man durch Ausrechnen verifizieren kann. Zum Beispiel stellt er folgende im weiteren viel benutzte Tatsache zur Verfügung:
Kennt man für ein Polynom P in einem Intervall $[r, R]$ eine Schranke $|P''| \leq K$, so folgt

$$x, a \in [r, R] \;\Rightarrow\; |P(x) - P(a) - P'(a) \cdot (x - a)| \leq \frac{K}{2}|x - a|^2.$$

Nach diesen Vorbereitungen sollen jetzt neue Funktionen mit interessanten Ableitungseigenschaften "durch Approximation" konstruiert werden. Dazu benötigen wir konvergente Folgen und die Vollständigkeit der reellen Zahlen (Kapitel 5). Dabei profitieren wir von zwei mathematischen Sachverhalten: Erstens, da die approximierenden Funktionenfolgen rational sind, haben wir den Monotoniesatz zur Verfügung und alle benötigten Konvergenzabschätzungen ergeben sich aus diesem einen Handwerkszeug. Zweitens, ebenfalls mit dem Monotoniesatz, folgen Abschätzungen

$$|f_n(x) - f_n(y)| \le L \cdot |x - y|, \; L \text{ unabhängig von } n,$$

und schon Archimedes hat formalisiert, wie daraus dieselbe Ungleichung für die Grenzfunktion folgt. Mit anderen Worten: Der Monotoniesatz für rationale Funktionen liefert eine Alternative zur Behandlung von Grenzfunktionen mit Hilfe gleichmäßiger Konvergenz.

Die komplexen Zahlen sind innerhalb und außerhalb der Mathematik fundamental. In Kapitel 6 wird der Inhalt der Kapitel 3-5 für komplexe Zahlen und komplexe Funktionen wiederholt, diesmal mit Einsatz der Vollständigkeit von $\mathbb{R}$.

Neben Anwendungen des Monotoniesatzes ist "das" andere Werkzeug für Konvergenzbeweise die Majorisierung, sehr oft durch die geometrische Reihe. Dies wird in Kapitel 8 an den Potenzreihen geübt.

Gleichzeitig mit der Differentialrechung wurde im 17. Jahrhundert eine sehr weitgehende Verallgemeinerung des Summierens, nämlich die Integration, entwickelt. Alle bisher behandelten Funktionen sind "gleichmäßig durch stückweise lineare Funktionen approximierbar". Mit diesem Begriff wird in Kapitel 9 der Hauptsatz der Differential- und Integralrechnung bewiesen.

In Kapitel 10 werden die stetigen Funktionen eingeführt als die Funktionen, die konvergente Folgen auf konvergente Folgen abbilden. Die Hauptsätze für stetige Funktionen können nach dem vorhergehenden Training mit weniger Mühe bewiesen werden als zu Beginn der Analysisausbildung.

Verständigungsschwierigkeiten

Die folgenden Punkte verursachen in der Mathematikausbildung immer wieder Schwierigkeiten; vielleicht hilft ein früher Hinweis, sie zu mildern.

ERSTENS, selbstverständlich muß man die grundlegenden Definitionen auswendig wissen. Es ist grotesk, wie viele Studierende sich dazu nicht bringen können.

ZWEITENS, ein richtig auswendig gelernter und wiedergegebener mathematischer Satz verrät noch keinerlei Verständis, allen juristischen Einwendungen zum Trotz. Es kommt darauf an, die neuen mathematischen Begriffe in *selbst* formulierten (deutschen) Sätzen zu richtigen Argumenten zusammmenzusetzen.

DRITTENS, es gibt mathematische Definitionen, die in der Umgangssprache nicht akzeptiert werden würden: Wir nennen eine Menge von Objekten einen Vektorraum, wenn gewisse Eigenschaften, nämlich Relationen zwischen den Elementen des Vektorrraums, erfüllt sind - wir sagen aber nicht, was die einzelnen Elemente "sind". Oder, wir nennen eine Funktion stetig, wenn eine unendliche Menge von Implikationen richtig ist - aber

wir sagen nicht, wie diese Implikationen zu verifizieren sind.
VIERTENS, die grammatische Struktur mathematischer Aussagen ist viel empfindlicher gegen kleine Veränderungen, als das in anderen sprachlichen Kontexten der Fall ist. Insbesondere ist es sehr ungewohnt, eine (von Mathematikern definierte) Vokabel nur genau so zu benutzen, wie sie definiert ist - statt daß man einfach "weiß", was sie bedeutet, und daher drauf los reden kann.
FÜNFTENS, gibt es einen Konflikt zwischen dem Verständnis von Details und dem Wahrnehmen der großen Linie. Lernende fühlen sich verunsichert durch Einzelheiten, die sie nicht verstanden haben. Daher wenden sie den Einzelheiten ihre Aufmerksamkeit zu. Das führt aber dazu, daß die größeren Ziele nicht wahrgenommen werden. Versuche, diese Ziele herauszuarbeiten, müssen notwendiger Weise Einzelheiten ignorieren. Und dann beginnt wieder die Verunsicherung...

Absolute Fehler und Relative Fehler

In diesem Abschnitt werden Vorkenntnisse über Ungleichungen gesammelt. Einfache Analysissituationen können schon mit den Begriffen der Prozentrechnung angemessen beschrieben werden. Der Umgang mit Fehlern, schon ehe man die Begriffe der Analysis zur Verfügung hat, bereitet auf ein gutes intuives Verständnis der Ziele der Analysis vor.

Beispiel 1. Absolute und relative Unterschiede und ihr Verhalten bei Addition und Multiplikation von Konstanten.
Die gleichbedeutenden Aussagen $x \in [0.9, 1.1]$ oder $0.9 \le x \le 1.1$ besitzen auch folgende Formulierungen, die nicht den Inhalt der Aussage, wohl aber die Akzente verändern:

x ist von 1 höchstens um 0.1 verschieden ("absoluter Unterschied"),
x ist von 1 höchstens um 10% verschieden ("relativer Unterschied").

Sprachregelung: Die Formulierung "x hat zu a den relativen oder auch prozentualen Unterschied $:= |x - a|/|a|$" wird verständlicher Weise nur benutzt, falls $a \neq 0$. Oft wird auch $0 < x, a$ (stillschweigend) vorausgesetzt.
Bei Addition von oder Multiplikation mit Konstanten verhalten sich die absoluten und die relativen Unterschiede verschieden. Aus der ersten Aussage ($0.9 \le x \le 1.1$) folgt:

$x + 1$ ist von 2 höchstens um 0.1 verschieden,
$x + 1$ ist von 2 höchstens um 5% verschieden,
$2x$ ist von 2 höchstens um 0.2 verschieden,
$2x$ ist von 2 höchstens um 10% verschieden
$x - 0.7$ ist von 0.3 höchstens um 0.1 verschieden,
$x - 0.7$ ist von 0.3 höchstens um 33%(!) verschieden,
$0.5x$ ist von 0.5 höchstens um 0.05 verschieden,
$0.5x$ ist von 0.5 höchstens um 10% verschieden.

Besondere Aufmerksamkeit verdient das enorme Anwachsen *relativer* Unterschiede beim *Subtrahieren* fast gleich großer Zahlen.
Beim Quadrieren verhalten sich Abweichungen nach oben bzw. unten verschieden:

x^2 liegt im Intervall $[0.81, 1.21]$,

daher ist x^2 von 1 höchstens um 21% verschieden; es ist nicht üblich, hier zu sagen: höchstens um 19% kleiner als 1, höchstens um 21% größer als 1. Man sagt jedoch auch: Wenn x "ungefähr" 10% von 1 verschieden ist, dann ist x^2 "ungefähr" 20% von 1 verschieden.
Übertrage diese Beispiele auf die Ausgangssituation $x \in [99, 101]$.

Beispiel 2. Größenordnungen von Potenzen.
Während im ersten Beispiel der Unterschied zwischen nicht genau festgelegten Zahlen und präzisen Konstanten in Worte gefaßt wird, betrachten wir jetzt die *unterschiedliche*

Größenordnung von Potenzen. Beachte: In Intervallen bei 0 und bei ∞ herrscht entgegengesetztes Verhalten!

Für alle $x \in (0,\ 0.1]$ gilt:
x^2 ist höchstens 10% von x,
x^3 ist höchstens 1% von x.

Für alle $x \in (0,\ 0.02]$ gilt:
x^2 ist höchstens 2% von x,
x^3 ist höchstens 0.04% von x.

Im Intervall $[20, \infty)$ gilt:
x ist höchstens 5% von x^2,
x ist höchstens 0.25% von x^3.

Im Intervall $[100, \infty)$ gilt:
x ist höchstens 1% von x^2,
x ist höchstens 0.01% von x^3.

Überlege, wie hier die Prozentzahlen von den Intervallenden abhängen. Z.B. im ersten Fall: $x^2 = x \cdot x \leq 0.1 \cdot x$, und das bedeutet eben, x^2 ist höchstens 10% von x, nämlich $x^2/x \leq 0.1$.

Beispiel 3. Beim Bilden der Reziproken, $x \to 1/x$, ändern sich (kleine) relative Unterschiede (fast) nicht.
Vorausgesetzt werde ein relativer Unterschied von höchstens 5%. Also, mit a als 100%,

$$|x - a|/|a| = |x/a - 1| \leq 0.05.$$

Rechnungen sind oft einfacher ohne Betragstriche. Aus $1 - 0.05 \leq x/a \leq 1 + 0.05$ folgt

$$1/(1 + 0.05) \leq a/x \leq 1/(1 - 0.05) \text{ oder } -0.05/(1 + 0.05) \leq a/x - 1 \leq 0.05/(1 - 0.05).$$

Schließlich
$$\frac{|1/x - 1/a|}{|1/a|} = \frac{|x - a|}{|x|} = |1 - \frac{a}{x}| \leq \frac{0.05}{1 - 0.05}.$$

Man sieht, weil $|x/a|$ nach Voraussetzung "wenig"($\leq 5\%$) von 1 verschieden ist, gilt für den Kehrwert a/x fast dasselbe und der relative Unterschied zwischen $1/x$ und $1/a$ ist höchstens "wenig" größer als 0.05. Was ändert sich, wenn die Voraussetzung von 5% auf 1% verschärft wird?

Beispiel 4. Fehler
Bisher haben wir von "Unterschieden" gesprochen. Es ist üblich von "Fehlern" zu sprechen, wenn theoretisch präzise Zahlen durch numerische Werte ungefähr mitgeteilt werden. Ich erläutere das an zwei berühmten Beispielen:

(i) Archimedes' Abschätzung für die Kreiszahl π, nämlich $3\frac{10}{71} < \pi < 3\frac{1}{7}$, drückt man dann so aus:
Der numerische Wert $3\frac{1}{7}$ für π hat einen absoluten Fehler von höchstens $3\frac{10}{70} - 3\frac{10}{71} \sim 0.002$ und einen relativen Fehler von höchstens $(\frac{10}{70} - \frac{10}{71})/3.14 \sim 0.064 \cdot 10^{-2}$, also von höchstens 0.064%.

(ii) Oder, aus der Ungleichung $1.4^2 = 1.96 < 2 < 2.25 = 1.5^2$ folgt für die mit dem Symbol $\sqrt{2}$ bezeichnete Zahl die Ungleichung: $1.4 < \sqrt{2} < 1.5$, also:
Der numerische Näherungswert 1.5 für $\sqrt{2}$ hat einen absoluten Fehler von höchstens 0.1 und einen relativen Fehler von höchstens $0.1/1.4 \sim 0.072$ oder höchstens 7.2%.

Beispiel 5. Verbesserung von Näherungswerten durch Argumentieren mit relativen und absoluten Fehlern.
Die Beispiele 1,3,4 werden kombiniert, um (dem Leser) bekannte Näherungen für $\sqrt{2}$ zu verbessern. Wir erinnern an die binomische Formel $(a+b)(a-b) = a^2 - b^2$, spezialisieren sie zu $(\sqrt{2}+1)(\sqrt{2}-1) = 2 - 1 = 1$ und benutzen die Umformung

$$\sqrt{2} = \frac{1}{(1+\sqrt{2})} + 1.$$

Wir lesen die rechte Seite als "Wert der Funktion $x \mapsto 1/(1+x)+1$ an der Stelle $x = \sqrt{2}$", und ich argumentiere jetzt, daß das Einsetzen eines Näherungswert x für $\sqrt{2}$ eine *bessere* Näherung liefert: Nach Beispiel 1 ist der relative Fehler von $1 + x$ (für $1 + \sqrt{2}$) kleiner als der von x; nach Beipiel 3 ist der relative Fehler von $1 + x$ fast derselbe wie der von $1/(1 + x)$; nach Beipiel 1 ist der relative Fehler von $1/(1 + x) + 1$ noch einmal kleiner als der von $1/(1 + x)$; also ist $1/(1 + x) + 1$ eine bessere Näherung für $\sqrt{2}$ als x.
Dies verfolgen wir noch einmal mit Zahlenwerten:
Weil 1.5 Näherungswert für $\sqrt{2}$ mit absolutem Fehler < 0.1 ist, ist 2.5 ein Näherungswert für $1 + \sqrt{2}$ mit demselben absoluten Fehler und daher mit einem relativen Fehler von höchstens $0.1/2.4 \sim 0.042$ oder 4.2%. Dann ist $1/2.5 = 0.4$ Näherungswert für $1/(1 + \sqrt{2})$ mit ungefähr demselben relativen Fehler von 4.2%, also mit einem absoluten Fehler von $0.4 \cdot 4.2/100 < 0.017$. (Berechne sicherheitshalber diese absoluten Fehler direkt aus den Näherungen 1.4 und 1.5 für $\sqrt{2}$ durch Einsetzen in $1/(1 + \sqrt{2})$.) Damit ist $1 + 0.4$ ein Näherungswert für $1 + 1/(1 + \sqrt{2}) = \sqrt{2}$ mit demselben absoluten Fehler (< 0.017) und einem relativen Fehler von $0.017/1.4 < 0.0122$, d.h. weniger als 1.22%.
Die geschickte Anwendung der binomischen Formel, nämlich die Umformung von $(\sqrt{2} + 1)(\sqrt{2} - 1) = 1$ *in* $\sqrt{2} = 1 + 1/(1 + \sqrt{2})$*, erlaubt also, aus einer Näherung mit höchstens* 7.2% *relativem Fehler eine Näherung mit weniger als* 1.3% *Fehler zu machen. Wir haben also in einem Schritt den relativen Fehler fast auf den sechsten Teil reduziert.* Natürlich können wir diese Verbesserung wiederholen:

> $1 + 1/(1 + 1.4) = 1 + 1/(1 + 7/5) = \frac{17}{12} \sim 1.4167$ und $1.4167 - \sqrt{2} \sim 0.0025$ ergibt einen relativen Fehler von $0.0025/1.4 < 0.0018$ oder weniger als 0.18% (grob gerechnet ein Sechstel von 1.3%).

Beispiel 6. Irrationalität von $\sqrt{2}$, indirekter Beweis.
Ich hoffe, daß es etwas Erstaunen verursacht, daß es solche "verbessernden" Formeln gibt. Wir sehen daher diese binomische Formel noch einmal an. Das Symbol $\sqrt{2}$ kommt darin zweimal vor. Sollte man sich nicht fragen, was passiert, wenn man umgekehrt wie in Beispiel 5 auflöst? Das liefert $\sqrt{2} = 1/(\sqrt{2} - 1) - 1$. Wenn man die vorherige Rechnung gründlich nachvollzogen hat, sieht man jetzt: Einsetzen einer Näherung (für $\sqrt{2}$) in die rechte Seite liefert als Ergebnis eine *schlechtere* Näherung! Wer kann so etwas wollen?

Wenn man sich die einzusetzende Näherung (für $\sqrt{2}$) als Dezimalzahl vorstellt, passiert in der Tat nichts Nützliches. Stellt man sich die Näherung jedoch als *gekürzten* Bruch p/q vor, mit $1 < p/q < 2$, so folgt durch Erweitern des ersten Doppelbruches mit dem Nenner von p/q :

$$1/(\frac{p}{q} - 1) - 1 = \frac{q}{p-q} - 1 = \frac{2q-p}{p-q},$$

und wegen $\quad q < p < 2q \quad$ folgt $\quad 0 < p - q < q.$

Man erhält also einen Bruch mit *kleinerem Nenner* ! Diese Beobachtung erlaubt, aus einer für Approximationen unbrauchbaren Formel einen *indirekten Beweis* für die Irrationalität von $\sqrt{2}$ zu gewinnen :
Angenommen $\sqrt{2}$ wäre rational, also gleich einem Bruch, $\sqrt{2} = P/Q$. Da es zwischen 1 und 2 nur *endlich viele* Brüche mit Nenner $\leq Q$ gibt, so kann man unter diesen den Bruch mit *kleinstem Nenner* suchen, der gleich $\sqrt{2}$ ist, $\sqrt{2} = p/q$. Aber dann liefert die vorgehende Rechnung, also

$$\frac{p}{q} = \sqrt{2} = \frac{1}{\sqrt{2}-1} - 1 = \frac{2q-p}{p-q},$$

einen Bruch mit *kleinerem Nenner* $(p - q)$ als der ausgesuchte Bruch p/q mit *kleinstem* Nenner. Dieser Widerspruch widerlegt die Annahme, $\sqrt{2}$ wäre rational. — Natürlich kann man sich in diesem Beweis p/q als den vollständig gekürzten Bruch vorstellen. Aber der Beweis ist so formuliert, daß diese Kürzungsregel nicht benutzt wird sondern nur das "Prinzip vom kleinsten Verbrecher": Falls $\sqrt{2}$ rational ist, so gibt es auch ein kleinstes Beispiel dafür. Aber, ein "kleinstes" Beispiel eingesetzt in die den Nenner verkleinernde Formel zeigt, daß es ein kleinstes Beispiel gar nicht geben kann.

Beispiel 7. Irrationalität von $\sqrt{2}$, "direkter" Beweis mit Hilfe absoluter Fehler. Wir stellen die Frage: Wie weit ist der Wert des Polynoms $P(x) := x^2 - 2$ an der rationalen Stelle $x = p/q$ von 0 verschieden? Wir wollen die Antwort $|P(p/q)| \geq 1/q^2$ herleiten, in Worten: *Das Polynoms P hat an der Stelle $x = p/q$ einen Wert vom Betrag mindestens* $1/q^2$. Haben wir diese Behauptung für gekürzte Brüche bewiesen, so gilt sie für alle Erweiterungen erst recht, weil die Behauptung dann schwächer ist. Daher dürfen wir annehmen, daß in dem Bruch alle Faktoren 2 gekürzt sind, d.h.

entweder p ist ungerade, oder, falls nicht, so ist q ungerade.

Im ersten Fall ist der Zähler von $P(p/q) = (p^2 - 2q^2)/q^2$ **ungerade**, also mindestens vom Betrag 1 und daher $|P(p/q)| \geq 1/q^2$. Im zweiten Fall ist p^2 mindestens durch $2 \cdot 2$ teilbar, $2q^2$ aber nur durch einen Faktor 2; in diesem Fall ist der Zähler von $P(p/q)$ das Doppelte einer ungeraden Zahl, also mindestens vom Betrage 2. In jedem Fall erhalten wir also: Das Polynom $P(x) := x^2 - 2$, dessen Nullstellen die Wurzeln aus 2 sind, hat an den rationalen Stellen $x = p/q$ einen Wert mindestens vom Betrage $1/q^2$. Das Polynom hat also keine

rationalen Nullstellen.
Es ist nützlich, ein Gespür dafür zu entwickeln, wie gut die hergeleiteten Ungleichungen sind. Obwohl im Beweis von $|P(p/q)| \geq 1/q^2$ nur ein sehr einfaches Argument benutzt wurde, gilt tatsächlich für unendlich viele Brüche das Gleichheitszeichen. Erstens, für $p/q := 3/2$ gilt in der Ungleichung $|P(p/q)| \geq 1/q^2$ Gleichheit, und zweitens, für die unendlich vielen Brüche $7/5, 17/12, 41/29, \ldots$, die in Beispiel 5 aus $p/q := 3/2$ erzeugt werden, ebenfalls, weil sich $|\text{Zähler}^2 - 2{\cdot}\text{Nenner}^2|$ bei der Verbesserung nicht ändert:

$$|(2q+p)^2 - 2(p+q)^2| = |2q^2 - p^2| = \ldots = |2 \cdot 2^2 - 3^2|.$$

Beispiel 8. Mittelwerte und Fehler.
Mittelwerte können oft zur Verkleinerung von Fehlern eingesetzt werden. Ich erkläre hier die drei wichtigsten Mittelwerte und die Ungleichungen zwischen ihnen. Dann benutze ich sie, um Näherungen für Quadratwurzeln *schneller* zu verbessern als in Beispiel 5.
Gegeben seien positive Zahlen $0 < a, b$. Dann bezeichnet man als

Arithmetisches Mittel die Zahl $A(a,b) := (a+b)/2$,
Geometrisches Mittel die Zahl $G(a,b) := \sqrt{a \cdot b}$,
Harmonisches Mittel die Zahl $H(a,b) := 1/((1/a + 1/b)/2) = 2ab/(a+b)$.

$A(a,b)$ ist der Mittelpunkt der Strecke mit den Endpunkten a, b. $G(a,b)$ ist die Seitenlänge des Quadrates, das denselben Flächeninhalt hat wie das Rechteck mit den Seiten a, b. Daher ist $G(a,b)$ auch die Höhe des rechtwinkligen Dreiecks, dessen Hypothenusenabschnitte die Längen a und b haben (Höhensatz). Der Umkreisradius dieses Dreiecks ist $A(a,b)$ und daher gilt

$G(a,b) \leq A(a,b)$.
Dies folgt auch aus der binomischen Formel: $0 \leq (\sqrt{a} - \sqrt{b})^2 = a + b - 2\sqrt{a \cdot b}$.

Zwischen den drei Mitteln bestehen einfache Beziehungen:

$1/H(a,b) = A(1/a, 1/b)$ und $G(a,b)^2 = H(a,b) \cdot A(a,b)$. Wegen $G \leq A$ folgt daraus $H(a,b) \leq G(a,b)$.

Der Abstand zwischen harmonischem und arithmetischem Mittel aus a, b ist unter Umständen viel kleiner als der Abstand zwischen a und b, nämlich:

$$A(a,b) - H(a,b) = \frac{a+b}{2} - \frac{2ab}{a+b} = \frac{(a+b)^2 - 4ab}{2(a+b)} = \frac{a-b}{2} \cdot \frac{a-b}{a+b}.$$

Man sieht sofort, daß der Abstand der Mittel immer kleiner als der halbe Abstand von a und b ist. Aber, nachdem wir in Beispiel 2 gesehen haben, daß kleine Zahlen beim Quadrieren viel kleiner werden, verstehen wir, daß $A(a,b) - H(a,b)$ viel kleiner als $|a-b|$ ist, falls $a-b$ nur Bruchteile von $a+b$ beträgt.

Anwendung auf $\sqrt{3}$. Es sei a ein Näherungswert für $\sqrt{3}$ und $a > \sqrt{3}$. Dann ist $b := 3/a$ ein zu kleiner Näherungswert. Arithmetisches Mittel $a_1 := A(a, 3/a)$ und Harmonisches

Mittel $b_1 := H(a,3/a) = 3/a_1$ sind dann *näher* bei einander liegende Zahlen, die das unveränderte geometrische Mittel $G(a,3/a) = \sqrt{3}$ zwischen sich haben. Es lohnt sich, mit dem Taschenrechner auszuprobieren, eine wie rasante Verbesserung durch einige Schritte dieser *quadratischen Fehlerverkleinerung* erzielt wird:

$$a_1 - b_1 = A(a,3/a) - H(a,3/a) = \frac{a-3/a}{2} \cdot \frac{a-3/a}{a+3/a} < \left(\frac{a-3/a}{2}\right)^2 = \left(\frac{a-b}{2}\right)^2.$$

Beispiel 9. Abweichung der Kreistangenten vom Kreis.
In der Differentialrechnung werden Tangenten eine große Rolle spielen. Vor der Differentialrechnung kommen "Tangenten" nur als Kreistangenten vor. Wir wollen genauer den

Einheitskreis $K := \{(x,y);\ x^2+y^2=1\}$ und seine oberste Tangente $T := \{(x,y);\ y=1\}$

betrachten. Wenn man x von 0 an vergrößert, entfernt man sich auf der Tangente vom Berührpunkt B, gleichzeitig wird der Abstand vom Kreis größer, aber anscheinend zuerst nur sehr wenig. Kann man das in deutlicheren Worten ausdrücken? In so deutlichen Worten, daß man eine Tangentendefinition für kompliziertere Kurven als Kreise erfinden kann?
Der Punkt $P = (x,1) \in T$ hat den Abstand $|x|$ vom Berührpunkt $B = (0,1) \in T \cap K$, vergleiche die folgende Figur. Der Punkt P hat nach dem Satz des Pythagoras den Abstand $\sqrt{1+x^2}$ vom Nullpunkt, dem Mittelpunkt von K. Daher hat P den Abstand $\sqrt{1+x^2}-1$ von diesem Einheitskreis. Wir wollen mit Beipiel 2 ausrechnen, daß dieser Abstand *viel* kleiner als $|x|$ ist, zumindest, falls P in der Nähe von B ist (in der folgenden Rechnung wird zuerst mit $\sqrt{1+x^2}+1$ erweitert, dann $(a-1)(a+1) = a^2-1$ benutzt, schließlich der Nenner durch Weglassen von x^2 unter der Wurzel verkleinert, also der Bruch vergrößert):

$$0 \le \sqrt{1+x^2}-1 = (\sqrt{1+x^2}-1)\cdot\frac{\sqrt{1+x^2}+1}{\sqrt{1+x^2}+1} = \frac{x^2}{\sqrt{1+x^2}+1} \le \frac{x^2}{2}$$

und daher z.B. $\quad |x| \le 0.1 \Rightarrow \sqrt{1+x^2}-1 \le \frac{|x|}{2}\cdot|x| \le 0.05\cdot|x|.$

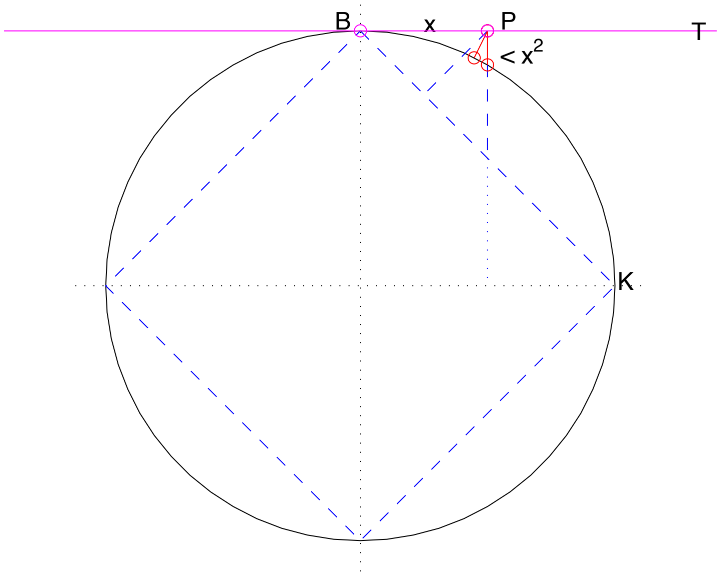
D.h., nahe bei B, genauer für $|x| \le 0.1$, ist der Abstand zwischen P und K höchstens 5% des Abstandes $|x|$ von P nach B, und für $|x| \le 0.01$ ist der Abstand von P nach K höchstens 0.5% von $|x|$, usw. wie Beispiel 2. Hiermit wird also nachgerechnet, daß die Tangente sich so langsam von dem Kreis entfernt (wenn $|x|$ wächst), daß die Formulierung "die Tangente berührt den Kreis" gerechtfertigt wird.

Falls man den Abstand zwischen Kreis und Tangente lieber senkrecht zur Tangente (statt wie eben senkrecht zum Kreis) mißt, so hat man den Abstand zwischen $(x,1) \in T$ und $(x,\sqrt{1-x^2}) \in K$ zu nehmen. Abgesehen davon, daß man dann offenbar die Beschränkung

$|x| \leq 1$ braucht, passiert (bis auf den Wegfall des Faktors 1/2) dasselbe:

$$0 \leq 1 - \sqrt{1-x^2} = (1 - \sqrt{1-x^2}) \cdot \frac{1+\sqrt{1-x^2}}{1+\sqrt{1-x^2}} = \frac{x^2}{1+\sqrt{1-x^2}} \leq x^2.$$

Der Abstand zwischen P und K ist höchstens gleich dem **Quadrat** des Abstandes zwischen P und Berührpunkt B, also nach Beispiel 2 eben "viel" kleiner als dieser.



Wie viel langsamer entfernt sich die Tangente vom Kreis als vom Quadrat?

Stellt man sich die Gerade T und das dem Kreis einbeschriebene Quadrat Q aus Metall vor, so würde man auch hier sagen: Die Gerade *berührt* das Quadrat. Aber diese physikalische Berührung ist *nicht* gemeint, wenn in der Mathematik formuliert wird: Die Tangente *berührt* den Kreis. Den Unterschied können wir mit Ungleichungen gut ausdrücken: Der Abstand zwischen dem Punkt $P \in T$ und dem Quadrat (parallel zur y-Achse gemessen) ist ebenso groß wie der Abstand zwischen P und B, nämlich gleich x, *unabhängig* davon wie klein x ist. Ich wiederhole, daß der Abstand von P und K kleiner als x^2 ist und daher ein *um so kleinerer* Prozentanteil von x je kleiner x ist. Dieser große Unterschied zwischen der Berührung von T und K einerseits und T und Q andererseits ändert sich auch nicht, wenn wir die Abstände von P senkrecht zu Q und K messen (der senkrechte Abstand vom Quadrat is $x/\sqrt{2}$, der senkrechte Abstand vom Kreis ist $< x^2/2$).

Dies ist ein wichtiges Beispiel dafür, daß die Bedeutung von Vokabeln in der Mathematik eventuell abweichend von der Umgangssprache festgelegt wird. Die Vokabel *berühren* hat eben für jemanden, der Stromleitungen verlegt, völlig berechtigter Weise eine andere Bedeutung als für jemanden, der sagt: Eine Kurve wird von ihrer Tangente berührt.

Tangenten der quadratischen Parabel

Ableitungen und Tangenten sind zentrale Begriffe der Analysis. Am Kreis war das noch nicht zu merken, denn wir bezeichnen diejenige Gerade durch einen Kreispunkt als Tangente, die senkrecht zum Radius ist. Offenbar werden dabei keine Begriffe der Analysis benutzt. Wir wollen nun an einer etwas weniger symmetrischen Kurve, dem Graphen der quadratischen Funktion $x \mapsto x^2$, sehen, daß wir auch hier nicht im Zweifel sein können, welche Gerade wir Tangente nennen wollen. Die am Kreis und an der Parabel beobachteten Eigenschaften dieser "naiven" Tangenten führen im nächsten Abschnitt zu einer *Tangentendefinition*. Diese Definition ist zwar noch nicht die, die seit dem *Ende des 19. Jahrhunderts* verwendet wird, aber es wird lange dauern, bis wir Funktionen angeben können, die auch "differenzierbar" sein sollen, obwohl unsere vorläufige Definition mehr verlangt, als diese komplizierten Funktionen erfüllen. Bis weit über die Behandlung der rationalen Funktionen hinaus wird die vorläufige Definition uns gestatten, mit einfacheren Argumenten und expliziteren Ungleichungen auszukommen als die endgültige Definition.

Die Tangente im tiefsten Punkt eines Kreises hat die Eigenschaft, daß alle rechtsseitigen Sehnen größere (nämlich positive) Steigung haben als die (horizontale) Tangente, während alle linksseitigen Sehnen kleinere Steigung als die Tangente haben. Durch jeden Punkt (a, a^2) des Graphen der Funktion $x \mapsto P(x) := x^2$ gibt es nun genau eine Gerade mit derselben Eigenschaft. Dazu berechnen wir für jedes $x \neq a$ die

Sehnensteigung: $$\frac{P(x) - P(a)}{x - a} = \frac{x^2 - a^2}{x - a} = x + a.$$

Linksseitig: $x < a \Rightarrow x + a < 2a,$ rechtsseitig: $a < x \Rightarrow 2a < x + a.$

Daher hat der Graph der linearen Funktion ℓ mit der Steigung $2a$ und dem Wert $\ell(a) = a^2$, also $\ell(x) := 2a \cdot (x - a) + a^2 = 2a \cdot x - a^2$ die am Kreis beobachtete Eigenschaft, nämlich, die Steigung dieser Geraden ist kleiner als die Steigung der rechtsseitigen Sehnen der Parabel und größer als die Steigung der linksseitigen Sehnen.

Weiter zeigt der Satz des Pythagoras, das jede Kreistangente das Innere des Kreises nicht trifft. Man kann auch sagen, die Tangente läßt den Kreis auf einer Seite. Die eben gefundene Gerade hat dieselbe Eigenschaft für die Parabel, d.h., sie liegt unterhalb der Parabel, denn

$$P(x) - \ell(x) = x^2 - (2a \cdot x - a^2) = (x - a)^2 \geq 0.$$

Wir können uns schon durch diese beiden Eigenschaften veranlaßt sehen, den Graph der linearen Funktion ℓ als *Parabeltangente* zu bezeichen. Aber es gibt weitere Eigenschaften, die das unterstützen. Ähnlich wie beim Kreis untersuchen wir jetzt, wie langsam sich die Parabeltangente von der Parabel entfernt. Wir betrachten einen (vom Berührpunkt verschiedenen) Punkt $(x, \ell(x))$, $x \neq a$ auf dieser Tangente. Sein Abstand vom Berührpunkt $(a, \ell(a))$ ist nach dem Satz des Pythagoras *mindestens* gleich dem Abstand $|x-a|$ der x-Koordinaten. Der Abstand von der Parabel ist *höchstens* gleich der parallel zur y-Achse gemessenen Entfernung, also höchstens gleich dem Abstand $P(x)-\ell(x) = (x-a)^2$ zwischen den Punkten $(x, \ell(x))$ und $(x, P(x))$. (Der Eindruck, dieser Abstand parallel zur y-Achse sei unnötig ungünstig, ist richtig; aber kein anderer Abstand ist so einfach zu berechnen, und selbst dieser "ungünstige" Abstand ist ja quadratisch klein und daher "sehr" klein.) Genau wie beim Kreis ist also das *Verhältnis* der Abstände des Tangentenpunktes einerseits von der Parabel, andererseits vom Berührpunkt $\leq (x-a)^2/|x-a|$. Der Tangentenpunkt $(x, \ell(x))$ ist also viel näher an der Parabel als am Berührpunkt der Tangente, die Tangente berührt die Parabel so gut, daß das *Verhältnis* dieser Abstände sogar proportional zum Abstand $|x-a|$ vom Berührpunkt abgeschätzt ist.

Wir treiben den Vergleich zwischen Kreis und Parabel noch weiter, weil der quantitative Vergleich zwischen Neuem und gut bekanntem Alten die wichtigste Arbeitsmethode der Analysis ist. Wir zeigen: Zu jedem Parabelpunkt (a, a^2) gibt es einen Kreis, so daß die Parabel *zwischen dem Kreis und seiner Tangente verläuft*, siehe die folgende Abbildung. Die Parabel wird also von ihrer Tangente mindestens ebenso gut berührt wie dieser Kreis.

Beweis. Welchen Kreis sollen wir nehmen? Die Definition der Kreistangente sagt, daß wir den Mittelpunkt des Kreises auf derjenigen Geraden durch (a, a^2) wählen sollen, die *senkrecht* zur Parabeltangente ist. Diese Gerade heißt Normale der Parabel und ist Graph der Funktion

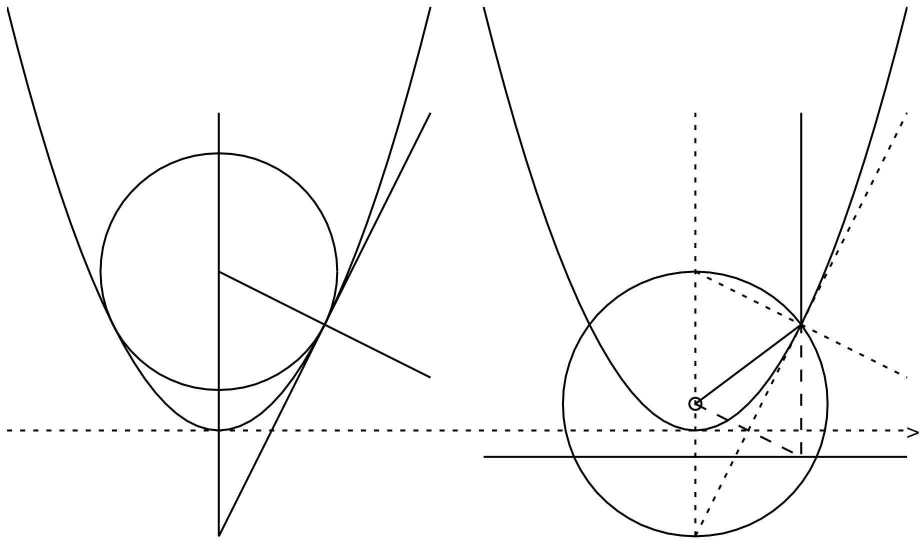
$$n(x) := -\frac{1}{2a} \cdot (x-a) + a^2 = -\frac{1}{2a} \cdot x + \frac{1}{2} + a^2.$$

(Erinnerung: Zu einander senkrechte Geraden haben die Steigungen m und $-1/m$.)

Stellt man sich Kreise vor, die von oben in die Parabel fallen, so bleiben sie mit dem Mittelpunkt auf der y-Achse liegen. Wir wählen also den Schnittpunkt $(0, n(0))$ der Normalen mit der y-Achse als Mittelpunkt und den Abstand zu (a, a^2) als Radius r. Nach Pythagoras finden wir $r^2 = 1/4 + a^2$ und haben damit eine aussichtsreiche

Kreisgleichung $$(x-0)^2 + (y - \frac{1}{2} - a^2)^2 = r^2 = \frac{1}{4} + a^2.$$

Wie kann man feststellen, daß die Punkte der Parabel das Innere dieses Kreise *nicht* treffen? Wenn wir $(x, y) = (x, x^2)$ in die linke Seite einsetzen, muß ein Wert $\geq r^2$ herauskommen! Es ist besser, noch r^2 zu subtrahieren, weil man eine Ungleichung der Form $\ldots \geq 0$ häufig leichter als richtig erkennen kann. Außerdem benutze die binomische



Links: Parabel zwischen Kreis und Tangente.
Rechts: Hohlspiegeleigenschaft, Brennpunkt und Leitlinie, Thaleskreis.

Formel $(u-v)^2 = u^2 - 2uv + v^2$ mit $u = x^2$ und $v = \frac{1}{2} + a^2$. Weil Quadrate immer ≥ 0 sind, hat man die gewünschte Ungleichung:

$$\begin{aligned} x^2 + (x^2 - \frac{1}{2} - a^2)^2 - r^2 &= x^2 + x^4 - 2x^2(\frac{1}{2} + a^2) + (\frac{1}{2} + a^2)^2 - (\frac{1}{4} + a^2) \\ &= x^4 - 2x^2a^2 + a^4 = (x^2 - a^2)^2 \\ &\geq 0. \end{aligned}$$

Damit haben wir so viele Vergleiche zwischen Kreis und Parabel angestellt, daß wir **definieren** können:

Die Gerade $\ell(x) := a^2 + 2a(x-a)$ heißt Tangente der Parabel $x \to x^2$ an der Stelle $x = a$. Die Steigung $2a$ dieser Geraden nennen wir Steigung der Parabel an der Stelle a.

Anwendungsbeispiel:

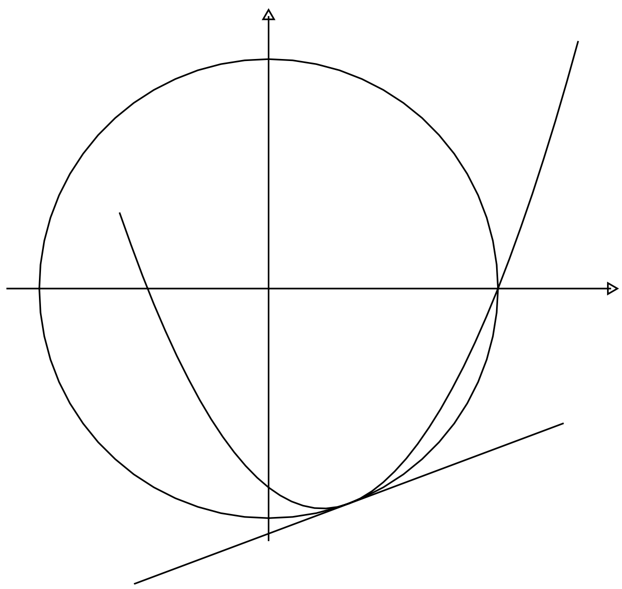
Ich finde wichtig, daß die neue Definition sofort neuartige Erklärungen ermöglicht, z.B. die perfekte Hohlspiegeleigenschaft der Parabel. Nachdem wir mit der Tangente auch die Normale der Parabel definiert haben, verallgemeinern wir das Reflektionsgesetz: Lichtstrahlen werden an Kurven so reflektiert, daß dort, wo sie auftreffen, gilt: Einfallswinkel (gegen die Normale) gleich Austrittswinkel. Reflektiert man nun parallel zur y-Achse einfallende Strahlen an der Parabel (x, x^2) (rechter Teil der vorhergehenden Abbildung), so gehen die reflektierten Strahlen durch den Punkt $(0, \frac{1}{4})$, weil dieser Punkt Mittelpunkt des

Thaleskreises für das aus y-Achse, Normale und Tangente gebildete (punktierte) rechtwinklige Dreieck ist. Daher heißt dieser Punkt *Brennpunkt* der Parabel. Nachdem man diesen Punkt entdeckt hat, sieht man leicht eine weitere geometrische Eigenschaft der Parabel: Jeder Punkt (x, x^2) der Parabel ist gleich weit von diesem Brennpunkt und von der "Leit"geraden $y = -\frac{1}{4}$ entfernt, denn

$$\begin{aligned}\text{Abstand vom Brennpunkt} &= \sqrt{(x-0)^2 + (x^2 - \frac{1}{4})^2} \\ &= |x^2 + \frac{1}{4}| = \text{Abstand von der Leitgeraden.}\end{aligned}$$

Wir haben jetzt *zwei verschiedene Definitionen* der Parabel, und ich möchte den Unterschied noch einmal betonen. Erstens handelte es sich um den Graphen der Funktion $x \mapsto x^2$, zweitens handelte es sich um die Menge der Punkte in der euklidischen Ebene, die gleich weit von einem Punkt ("Brennpunkt") und einer Geraden ("Leitgerade") entfernt sind. Die zweite Eigenschaft ist der Definition des Kreises offenbar ähnlicher als die erste. Auch die Beschreibung der Parabeltangenten wird an die beiden Definitionen angepaßt. Verwenden wir die geometrische Definition der Parabel, so ist die Tangente in einem Parabelpunkt P definiert als *Winkelhalbierende* der Strecken von P zum Brennpunkt und von P zur Leitgerade. Betrachten wir die Parabel als Graph der Funktion $x \mapsto x^2$, so ist die Tangente *Graph einer linearen Funktion.* Die Tangente an der Stelle $x = a$ ist der Graph der linearen Funktion $\ell(x) := a^2 + 2a(x-a)$. Im folgenden Abschnitt sollen zunächst Tangenten an Graphen von Funktionen definiert werden. Das werden wieder Graphen linearer Funktionen sein. Danach kommen wir auf Tangenten von allgemeineren Kurven zurück.

Aufgabe. Geben Sie die Gleichung der Parabel an, die den Einheitskreis in dem Punkt (a, b), mit $a^2 + b^2 = 1,\ b \neq 0$, von innen berührt und durch $(1, 0)$ geht, wie in dem nebenstehenden Bild.
Mit anderen Worten, zu jedem Kreispunkt (a, b), in dem die Tangente *nicht parallel zur y-Achse* ist, gibt es eine Parabel, so daß der Kreis in dem Interval $[a - l, a + l]$, mit $l := 1 - |a|$, zwischen Parabel und Parabeltangente liegt.



Damit können wir sagen: Kreise und Parabeln werden von ihren Tangenten gleich gut berührt, denn zuerst haben wir gesehen, daß die Parabel an jeder Stelle zwischen einem Kreis und dessen Tangente verläuft, und die Aufgabe zeigt das Umgekehrte, daß ein Kreis an jeder Stelle (an der seine Tangente nicht vertikal ist) zwischen einer (vertikalen) Parabel und deren Tangente verläuft. – Die geometrische Definition der Parabel erlaubt natürlich auch, gedrehte Parabeln zu betrachten. Dann spielen die Kreispunkte mit vertikaler Tangente keine Sonderrolle. Ich habe die vertikalen Parabeln bevorzugt, weil wir uns zunächst auf Graphen von Funktionen konzentrieren werden.

Aufgabe zur Dreiecksungleichung.
a) Zeigen Sie für $a, b \in \mathbb{R}$ die sogenannte Dreiecksungleichung $|a + b| \leq |a| + |b|$.
b) Unter was für Voraussetzungen an a, b gilt $|a + b| = |a| + |b|$?
c) Folgern Sie aus a) für $a, b, c \in \mathbb{R}$: $|a + b + c| \leq |a| + |b| + |c|$.
d) Folgern Sie ähnlich: $a_k \in \mathbb{R}, k = 1 \dots n \;\Rightarrow\; |\sum_{k=1}^{n} a_k| \leq \sum_{k=1}^{n} |a_k|$.
Der Name "Dreiecksungleichung" ist bei der zweidimensionalen Verallgemeinerung dieser Ungleichung einleuchtender.

Ableitungen und Tangenten der Polynome

Es ist in Ordnung, bei *einzelnen* einfachen Kurven, wie Kreisen und Parabeln, dafür geeignet erscheinende Geraden als *Tangenten* zu bezeichnen, obwohl man noch keine allgemeine Definition hat. Jetzt soll für eine Klasse von Kurven, nämlich die *Graphen recht spezieller Funktionen*, die Graphen von Polynomen, der Begriff Tangente erklärt und später Differentiationsregeln bewiesen werden. Dazu sind *Definitionen* notwendig. Die zur Definition verwendeten Eigenschaften werden motiviert durch das, was bei den Tangenten des Graphen der quadratischen Funktion $x \mapsto x^2$ beobachtet wurde.

Es ist leicht, eine lineare Funktion ℓ mit der Steigung m anzugeben, die an der Stelle a den Wert b hat, nämlich

$$\ell(x) := m \cdot (x-a) + b.$$

Weiter verabreden wir: Falls ℓ an der Stelle a Tangente des Polynoms P ist, so definieren wir die Steigung m der Tangente auch als die *Steigung des Polynoms* P an der Stelle a. Und umgekehrt, falls es gelingt, die Steigung m eines Polynoms P an der Stelle a zu definieren, so soll die Gerade $\ell(x) := m \cdot (x-a) + P(a)$ Tangente sein. Auch in Formeln läßt sich die Differenz zwischen dem Polynom P und der linearen Funktion ℓ leicht mit Hilfe der Differenz von Steigungen ausdrücken:

$$\begin{aligned} P(x) - \ell(x) = P(x) - P(a) - m \cdot (x-a) &= \left(\frac{P(x)-P(a)}{x-a} - m \right) \cdot (x-a) \\ &= (\text{Sehnensteigung} - \text{Geradensteigung}) \cdot (x-a) \end{aligned}$$

Mit anderen Worten, wir müssen nur entweder den Begriff *Steigung* (einer Funktion an einer Stelle a) oder den Begriff *Tangente* (der Funktion an dieser Stelle) definieren, der jeweils andere Begriff ist dann in einfacher Weise dazu gegeben. Zur Veranschaulichung zeichnen wir immer den Graphen $(x, f(x))$, x im Definitionsbereich der Funktion f. Die Tangente (an der Stelle a) ist Graph einer linearen Funktion $x \mapsto f(a) + m \cdot (x-a)$.

Wir nehmen uns zunächst die Potenzfunktionen $P(x) := x^n$ vor und zeigen, daß wir die Steigung m an einer Stelle a so definieren können, daß die Werte des Polynoms P ebenso "wenig" von der linearen Funktion $x \mapsto P(a) + m \cdot (x-a)$ abweichen, wie wir es bei der quadratischen Funktion $x \mapsto x^2$ beobachtet haben. Dann machen wir diese Approximationseigenschaft zu unserer Tangentendefinition und zeigen, daß alle Polynome berechenbare Tangenten haben.

Wir beginnen mit dem Ausrechnen der

Sehnensteigung: $\dfrac{P(x)-P(a)}{x-a} := \dfrac{x^n - a^n}{x-a} = x^{n-1} + x^{n-2}a + \ldots + xa^{n-2} + a^{n-1}.$

Beweis. Multipliziere die letzte Gleichung mit $(x-a)$. Links kürzt sich der Nenner $(x-a)$. Rechts kommen nach dem Ausmultiplizieren x^n und $-a^n$ alleine vor, alle übrigen Terme treten doppelt auf, einmal mit positivem, einmal mit negativem Vorzeichen, so daß sie alle wegfallen.

Wir setzen zunächst $0 < a, x$ voraus, um die Situation möglichst übersichtlich zu haben, so lange wir noch nicht wissen, was wir erwarten sollen. Dann sind offenbar die rechtsseitigen Sehnensteigungen (also $a < x$) *größer* als na^{n-1} und die linkseitigen Sehnensteigungen $(x < a)$ sind *kleiner* als na^{n-1}. Wir müssen also na^{n-1} als Steigung von P an der Stelle a erwarten. Zu dieser Steigung gehört als *erwartete Tangente*

$$T_a(x) := na^{n-1} \cdot (x-a) + a^n.$$

Wir werden diese Erwartung als berechtigt ansehen, wenn wir zeigen können, daß diese lineare Funktion von dem Graphen von P nicht stärker abweicht, als wir es bei $x \mapsto x^2$ beobachtet haben. Vorher bemerken wir noch, daß es uns an der Kreistangente gar nicht stört, daß *nicht alle* Punkte der Tangente nahe an dem Kreis sind. Wir wollen daher von einer Tangente nur verlangen, daß sie sehr nahe bei ihrer Kurve bleibt, wenn wir nicht zu weit vom Berührpunkt weg sind. Aus diesem Grund machen wir folgende Voraussetzung: x ist nicht weiter als $r > 0$ von a entfernt, also $|x-a| \le r$. (Bei Polynomen könnten wir immer $r = 1$ verabreden, aber wir haben schon bei Kreisen gesehen, daß das unpraktisch ist, etwa weil der Kreisradius ja viel kleiner als 1 sein kann. Außerdem werden wir weiteren Funktionen wie $x \mapsto 1/x$ begegnen, bei denen nur ausreichend kleine Intervalle $[a-r, a+r]$ um die Berührstelle a sinnvoll für Approximationen sind.)

Nach diesen Verabredungen folgern wir als nächstes aus der etwas länglichen Formel für die Sehnensteigung eine leichter zu handhabende Ungleichung:

Voraussetzung: $x \in [a-r, a+r]$ und $R := |a| + r$

Steigungsungleichung: $\left|\dfrac{x^n - a^n}{x-a}\right| \le n \cdot R^{n-1}.$

Um die Differenz zwischen x^n und der erwarteten Tangente $T_a(x)$ zu bearbeiten, muß diese Steigungsungleichung für alle Exponenten zwischen 1 und $n-1$ benutzt werden:

$$|\text{Funktionswert} - \text{Tangentenwert}| = |x-a| \cdot |\text{Sehnensteigung} - \text{Tangentensteigung}| =$$

$$\begin{aligned}
|x^n - T_a(x)| &= |x-a| \cdot \left|x^{n-1} + x^{n-2}a + \ldots + xa^{n-2} + a^{n-1} - n \cdot a^{n-1}\right| \\
&= |x-a| \cdot \left|(x^{n-1} - a^{n-1}) + (x^{n-2}a - a^{n-1}) + \ldots + (xa^{n-2} - a^{n-1})\right| \\
&\le |x-a| \cdot \big(|x^{n-1} - a^{n-1}| + |a| \cdot |x^{n-2} - a^{n-2}| + \ldots + |a^{n-2}| \cdot |x-a|\big) \\
&\le (x-a)^2 \cdot R^{n-2} \cdot \big((n-1) + (n-2) + \ldots + 1\big) \quad \text{(Steigungsungleichung)} \\
&= \frac{n(n-1)}{2} R^{n-2} \cdot (x-a)^2.
\end{aligned}$$

Mit dieser Ungleichung sind wir am Ziel, denn sie besagt erstens, daß der Graph von $P : x \mapsto x^n$ in dem betrachteten Intervall zwischen zwei Parabeln mit gleicher Tangente T_a liegt, nämlich:

$$x \in [a-r, a+r] \quad \Rightarrow \quad P_-(x) \ \le x^n \le \ P_+(x) \qquad \text{mit}$$
$$P_-(x) := T_a(x) - \frac{n(n-1)}{2} R^{n-2} \cdot (x-a)^2, \ P_+(x) := T_a(x) + \frac{n(n-1)}{2} R^{n-2} \cdot (x-a)^2,$$

und die hergeleitete Ungleichung besagt zweitens, daß die Sehnensteigungen von P in $[a-r, a+r]$ von der Zahl na^{n-1}, also der Steigung von $x \mapsto T_a(x)$, höchstens um $\frac{n(n-1)}{2} R^{n-2} \cdot |x-a|$ verschieden sind – also um so weniger verschieden, je näher x bei a ist. Die Idee, kompliziertere Funktionen, nämlich zunächst Potenzfunktionen, mit quadratischen Funktionen zu vergleichen, hat geklappt und rechtfertigt folgende Definition für Funktionen f, die nicht unbedingt auf ganz $\mathbb{R}$ definiert sind sondern nur auf einem Intervall $[\alpha, \omega] \subset \mathbb{R}$:

Definition der elementaren Differenzierbarkeit:
Eine Funktion $f : [\alpha, \omega] \to \mathbb{R}$ heißt an der Stelle $a \in [\alpha, \omega]$ *ableitbar* (oder *differenzierbar*) mit der Ableitung (oder Steigung) $f'(a) = m$, falls gilt:
Es gibt ein Interval $[a-r, a+r]$ und eine Konstante K so daß

$$x \in [a-r, a+r] \cap [\alpha, \omega] \Rightarrow \big| f(x) - f(a) - m \cdot (x-a) \big| \le K \cdot |x-a|^2,$$

oder gleichbedeutend mit Sehnensteigungen (d.h. $x \neq a$) formuliert

$$x \in [a-r, a+r] \cap [\alpha, \omega] \Rightarrow \left| \frac{f(x) - f(a)}{x-a} - m \right| \le K \cdot |x-a|.$$

Bemerkungen. (i) *Rechtfertigungen.* Zu einer solchen axiomatischen Definition muß man einige Fragen stellen: Sind die Steigungen $f'(a)$ durch die Definition *eindeutig* bestimmt? Das Eindeutigkeitslemma beantwortet die Frage mit ja. Sind genügend viele Funktionen differenzierbar? Der folgende Satz zeigt dies für Polynome, das nächste Kapitel für rationale Funktionen, und mit der Behandlung der Vollständigkeit werden unsere Konstruktionsmöglichkeiten sehr umfangreich. Natürlich muß man auch fragen: Ist die Bezeichnung "Steigung" angemessen? Die Antwort "ja" hat zwei Teile: Aus der Definition folgt unmittelbar, daß wachsende Funktionen f Ableitungen $f' \ge 0$ haben (im Anschluß an die Differenzierbarkeit der Polynome), die Umkehrung, der Monotoniesatz, ist ein Hauptresultat der Analysis: Funktionen mit $f' \ge 0$ sind in der Tat (schwach) wachsend.
(ii) *Andere Definitionen.* Mit dieser Definition sind nur solche Funktionen differenzierbar, die von ihren Tangenten ebenso gut approximiert werden, wie es vorher bei Kreis und

Parabel beobachtet wurde (oder deren Sehnensteigungen nicht stärker von der Tangentensteigung verschieden sind, als bei $x \mapsto x^2$ beobachtet). Wir werden später lernen, wesentlich kompliziertere Funktionen zu konstruieren. Diese Funktionen können mit einer etwas schwächeren, weniger expliziten Eigenschaft als "differenzierbar" definiert werden. Das ist sinnvoll, weil die Sätze, die wir für differenzierbare Funktionen beweisen, richtig bleiben, wenn die schwächere Definition benutzt wird. Zunächst begegnen wir nur Funktionen, die von ihren Tangenten nicht stärker abweichen als die quadratische Funktion $x \mapsto x^2$.

Eindeutigkeitslemma. Falls eine Funktion f an der Stelle a differenzierbar ist, so ist die Steigung m eindeutig bestimmt.

Indirekter Beweis. Hätten wir zwei verschiedene Steigungen $m_1 \neq m_2$, so hätten wir auf Grund der Definition Konstanten $r_i, K_i,\ i = 1, 2$, mit denen gilt:

$$|x - a| \le r_i \Rightarrow |f(x) - f(a) - m_i(x - a)| \le K_i|x - a|^2, \quad \text{i=1,2}$$

Daraus folgt mit der Dreiecksungleichung

$$|x - a| \le \min(r_1, r_2) \Rightarrow |(m_1 - m_2)(x - a)| \le (K_1 + K_2) \cdot |x - a|^2,$$

also auch

$$0 < |x - a| < \min(r_1, r_2) \Rightarrow |m_1 - m_2| \le (K_1 + K_2) \cdot |x - a|.$$

Aber die letzte Ungleichung ist falsch, wenn wir x so nahe bei a wählen, daß gilt

$$0 < |x - a| < \min(r_1, r_2, \frac{|m_1 - m_2|}{2(K_1 + K_2)}).$$

Satz. Polynome $P : x \mapsto \sum_{k=0}^{n} a_k x^k$ sind *differenzierbar* mit $P'(a) = \sum_{k=1}^{n} a_k k a^{k-1}$. Dabei kann die Länge des Intervalls $[a - r, a + r]$ in der Definition unabhängig von dem Polynom gewählt werden, die Konstante K wird aus den Koeffizienten des Polynoms und aus den Intervallgrenzen explizit berechnet.

Aufgabe. Spezialisiere jeden Schritt des folgenden Beweises auf Polynome vom Grade 3.

Beweis. Wir erläutern die Beweismethode zuerst an einer einfacheren Aussage, indem wir die Steigungsungleichung von Potenzen auf Polynome ausdehnen.

Das Intervall $[a-r, a+r]$ sei (beliebig) gewählt, definiere dazu wie oben $R := |a|+r$. Dann liefert die Steigungsungleichung für Potenzen

$$x \in [a - r, a + r] \Rightarrow |a_k x^k - a_k a^k| \le |a_k| k R^{k-1} \cdot |x - a|,\ k = 1 \dots n.$$

Diese Ungleichungen können addiert werden. Erinnerung: die Dreiecksungleichung lautet $|a + b| \le |a| + |b|$, oder für mehrere Summanden $|\sum_k b_k| \le \sum_k |b_k|$. Damit ergibt sich:

Voraussetzung: $\quad x \in [a - r,\ a + r],\ R := |a| + r,$

Steigungsungleichung: $\quad \left| \sum_{k=0}^{n} a_k x^k - \sum_{k=0}^{n} a_k a^k \right| \le \left(\sum_{k=1}^{n} |a_k| k R^{k-1} \right) \cdot |x - a|.$

Die eigentliche Behauptung entsteht ebenso, indem die Ungleichungen addiert werden, die die Abweichung der einzelnen Potenzen von ihren Tangenten ausdrücken:

$$|a_k x^k - a_k \cdot (a^k + k a^{k-1}(x-a))| \le |a_k| \frac{k(k-1)}{2} R^{k-2} |x-a|^2.$$

Definiere eine Konstante $K := \sum_{k=1}^n |a_k| \frac{k(k-1)}{2} R^{k-2}$, die alle Abweichungsbeiträge der einzelnen Potenzen zusammenfaßt, und addiere (für $x \in [a-r, a+r]$) zur

Tangentenabweichung $$\left| \sum_{k=0}^n a_k x^k - \sum_{k=0}^n a_k a^k - (\sum_{k=1}^n a_k k a^{k-1}) \cdot (x-a) \right| \le K \cdot |x-a|^2.$$

Bezeichnung: Das die Steigungen berechnende Polynom $P'(x) := \sum_{k=1}^n a_k k x^{k-1}$ heißt *Ableitung* des Polynoms $P(x) := \sum_{k=0}^n a_k x^k$. Der Wert $P'(a)$ heißt Steigung von P an der Stelle a. Damit schreibt sich die Tangentenabweichung für Polynome übersichtlicher:

Voraussetzung: $$x \in [a-r,\, a+r\,],\ P(x) := \sum_{k=0}^n a_k x^k,\ P'(x) := \sum_{k=1}^n a_k k x^{k-1},$$

$$R := |a| + r,\ K := \sum_{k=1}^n |a_k| \frac{k(k-1)}{2} R^{k-2}.$$

Tangentenabweichung: $$|P(x) - P(a) - P'(a)(x-a)| \le K|x-a|^2,$$

oder für Steigungen ($x \ne a$): $$\left| \frac{P(x) - P(a)}{x-a} - P'(a) \right| \quad \le \quad K \cdot |x-a|.$$

Der folgende Satz mit Beweis gehört noch zum unmittelbaren Verständnis der Differenzierbarkeitsdefinition. U.a. zeigt der Satz: an inneren Extremstellen c gilt $f'(c) = 0$.

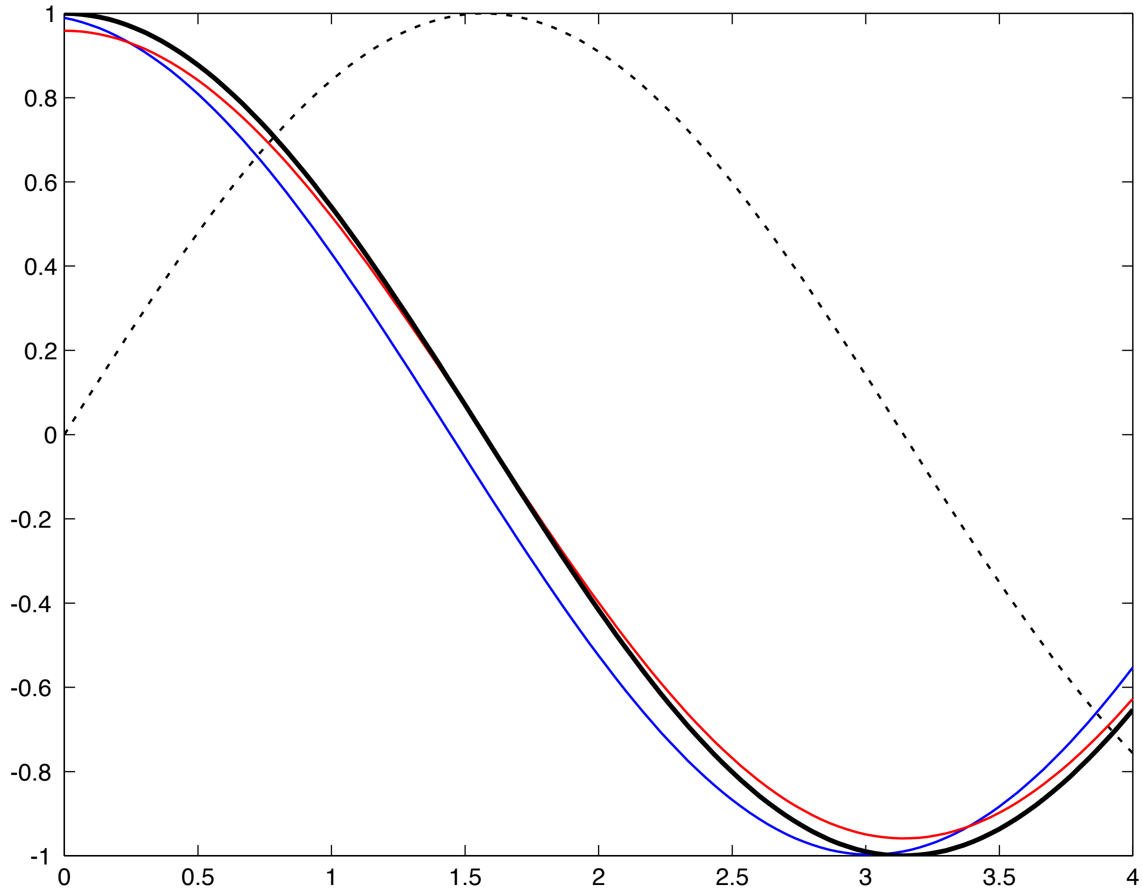
Satz. Wachstum und nichtnegative Steigung.
Wachsenden Funktionen f (also $x \le y \Rightarrow f(x) \le f(y)$) haben an jeder Stelle nichtnegative Steigung $f'(a) \ge 0$. – Die Ungleichungen $\le$ können nicht durch $<$ ersetzt werden, wie die streng wachsende Funktion $x \mapsto f(x) := x^3$ mit $f'(0) = 0$ zeigt.
Indirekter Beweis. Angenommen, an einer Stelle a sei die Ableitung negativ, $f'(a) < 0$. Die Definition liefert uns Konstanten r, K mit

$$|x-a| \le r \;\Rightarrow\; |f(x) - f(a) - f'(a)(x-a)| \le K \cdot |x-a|^2.$$

Wir verkleinern r zu $\min(r, |f'(a)|/2K) > 0$ und folgern weiter:

$$a-r \le x < a \;\Rightarrow\; f(x) \ge f(a) + f'(a)(x-a) - K \cdot |x-a|^2 \ge f(a) + f'(a)(x-a)/2 > f(a),$$
$$a < x \le a+r \;\Rightarrow\; f(x) \le f(a) + f'(a)(x-a) + K \cdot |x-a|^2 \le f(a) + f'(a)(x-a)/2 < f(a).$$

Mit anderen Worten, für x nahe bei a und $x < a$ ist $f(x) > f(a)$, während rechts von a die Funktionswerte $< f(a)$ sind – wie man es von einer negativen Ableitung erwartet, aber im *Widerspruch* zur vorausgesetzten Monotonie von f.



Vergleich von Ableitung und Differenzenquotienten.

Die Ableitung f' der punktierten Funktion f ist fett gezeichnet. Sie wird von zwei Differenzenquotienten (mit der recht großen Schrittweite $h = 0.5$) approximiert,

von einseitigen: $(f(x+h) - f(x))/h$ und von symmetrischen: $(f(x+h) - f(x-h))/2h$.

Die symmetrischen Differenzenquotienten sind in der Nähe der Wendestellen von f zu klein, aber sonst erheblich genauer als die einseitigen. Je kleiner die Schrittweite ist, um so mehr sind die symmetrischen Quotienten überlegen; bei quadratischen Funktionen geben sie sogar genau die Ableitung, unabhängig von der Schrittweite: $((a+h)^2 - (a-h)^2)/2h = 2a$.

Rationale Funktionen, Differentiationsregeln.

Die Differenzierbarkeitsdefinition kann auch auf die Funktion $x \to f(x) := 1/x$ angewandt werden. Sie ist im Definitionsbereich $\mathbb{R} \setminus \{0\}$ differenzierbar mit der Ableitung $f'(a) = -1/a^2$. Aus dieser Funktion und den Polynomen kann man alle rationalen Funktionen (nämlich die Quotienten von Polynomen) gewinnen. Dabei werden die Differentiationsregeln wichtig. Sie besagen: Falls zwei Funktionen f, g differenzierbar sind, so sind auch $f + g, f \cdot g, f \circ g$ differenzierbar und die Ableitungen dieser zusammengesetzten Funktionen können *explizit* aus den Werten und Ableitungen von f, g berechnet werden. Obwohl der Funktionsbegriff sich von Newton bis Hilbert stark verändert hat und dementsprechend auch die Differenzierbarkeitsdefinition, sind die Differentiationsregeln *dieselben* geblieben. Wir erläutern den Grund dafür.

Das folgende Beispiel, ist aus zwei Gründen wichtig. Erstens dehnt es, mit der Kettenregel, die Menge der Funktionen, die wir differenzieren können, sehr aus. Zweitens kann man nicht mehr wie bei Polynomen das Intervall $[a - r, a + r]$, auf dem Funktion und Tangente verglichen werden, völlig unabhängig von der Stelle a wählen, weil natürlich 0 *nicht* in diesem Intervall liegen soll.

Beispiel $f(x) = 1/x$. Wir setzen voraus $0 < a, x$. Dann gilt für die Sehnensteigungen:

$$\frac{1/x - 1/a}{x - a} = \frac{-1}{x \cdot a}$$

Offenbar liegt die Zahl $-1/a^2$ (wie in den anderen Potenzbeispielen) zwischen den rechtsseitigen ($a < x$) und den linksseitigen ($x < a$) Sehnensteigungen, so daß nur der Wert $-1/a^2$ als Ableitung erwartet werden kann. Damit ist $T_a(x) := 1/a - (x - a)/a^2$ die erwartete Tangente. Wir berechnen die Differenz zwischen $f(x)$ und erwarteter Tangente:

$$f(x) - T_a(x) = x^{-1} - (a^{-1} - a^{-2} \cdot (x - a)) = (\frac{1}{a^2} - \frac{1}{x \cdot a}) \cdot (x - a) = \frac{(x - a)^2}{x \cdot a^2}.$$

Wir vermeiden durch die Wahl $r := |a|/2$, daß 0 im Intervall $[a - r, a + r]$ liegt. Dies ergibt mit $f'(a) = -1/a^2$ und mit der Konstanten $K := 2/a^3$ daß die *Ungleichung aus der Differenzierbarkeitsdefinition erfüllt* ist:

$$0 < \frac{a}{2} \le x \quad \Rightarrow \quad 0 \le f(x) - f(a) - f'(a) \cdot (x - a) \le +\frac{2}{a^3}(x - a)^2.$$

Ich habe hier zusätzlich $0 < a$ vorausgesetzt, weil dann die Rechnung zeigt, daß der Graph dieser Funktion *oberhalb* jeder Tangente und *unterhalb* quadratischer Funktionen mit denselben Tangenten liegt. (Für $a < 0$ ist es umgekehrt.)

Aufgabe. Folgere ähnlich für $f(x) = x^{-n}$ und $n > 1$:

$$0 < \frac{a}{2} \leq x \quad \Rightarrow \quad 0 \leq f(x) - f(a) - f'(a) \cdot (x-a) \leq \frac{n(n+1)}{(a/2)^{n+2}} \cdot (x-a)^2.$$

Ableitungsregeln. Offenbar wird man nur für Einführungsbeispiele, zur Erläuterung der Definition, Ableitungen direkt aus der Definition bestimmen wollen. Danach sucht man weniger aufwendige Verfahren, also z.B. Regeln, wie man aus den Ableitungen von zwei Funktionen f, g die Ableitungen von $f \cdot g$ und $f \circ g$ berechnet. Allgemeiner sucht man zu jeder neuen Konstruktion von Funktionen aus einfacheren zu dieser Konstruktion gehörige Ableitungsregeln. Wir beantworten zunächst die Fragen: Wie müssen die Regeln aussehen, die zu Summe, Produkt und Komposition von Funktionen gehören? Danach beweisen wir diese Regeln. Als nächstes bemerken wir, daß wir Kurven in der Ebene durch *Paare von Funktionen* beschreiben können; wie differenziert man diese? Später konstruieren wir neue Funktionen als *Umkehrfunktionen*, welche Differentiationsregel gehört dazu? In späteren Kapiteln wird die Konstruktion neuer Funktionen als Grenzfunktion von Approximationen durch einfachere Funktionen besonders wichtig sein; gibt es auch dazu Differentiationsregeln?

Linearkombinationen.

Wir haben schon Polynome als Linearkombinationen von Potenzen behandelt und können daraus ablesen: Falls die Polynome P, Q die Ableitungen P', Q' haben und $\alpha, \beta \in \mathbb{R}$ sind, so gilt:

$$(\alpha \cdot P + \beta \cdot Q)' = \alpha \cdot P' + \beta \cdot Q'.$$

Diese Regel kann für komplizierte Funktionen nicht anders lauten, also

$$(\alpha \cdot f + \beta \cdot g)' = \alpha \cdot f' + \beta \cdot g'.$$

Aufgabe. Der Beweis ist eine direkte Anwendung der Dreiecksungleichung.

Produktregel

Wenn es überhaupt Differentiationsregeln gibt, müssen sie auch für die einfachsten Funktionen gelten. Um zu sehen, wie eine Produktregel lauten könnte, multiplizieren wir zuerst lineare Funktionen.

$$\ell_1(x) = w_1 + m_1(x-a), \quad \ell_2(x) = w_2 + m_2 \cdot (x-a)$$
$$(\ell_1 \ell_2)(x) = w_1 w_2 + (w_1 m_2 + w_2 \cdot m_1) \cdot (x-a) + m_1 \cdot m_2 \cdot (x-a)^2.$$

Offenbar hat die quadratische Funktion $\ell_1\ell_2$ bei a die Ableitung $(\ell_1 \cdot \ell_2' + \ell_2 \cdot \ell_1')(a)$. Wenn es also eine Produktregel gibt, muß sie wohl so lauten:

$$(f_1 \cdot f_2)'(a) = f_1'(a) \cdot f_2(a) + f_1(a) \cdot f_2'(a).$$

Kettenregel, Kompositionsregel

Wieder probieren wir, wie die Regel lauten könnte. Gegeben sei:

$\ell(x) = A + m \cdot (x - a),\ m \neq 0$

und, es gibt Zahlen R, K so daß gilt

$Y \in [A - R, A + R] \ \Rightarrow\ |f(Y) - f(A) - f'(A) \cdot (Y - A)| \leq K \cdot |Y - A|^2.$

Nach Einsetzen von $Y = \ell(x)$ usw in die Ungleichung für f finden wir:

$$|x - a| \leq R/m \ \Rightarrow\ |f(\ell(x)) - f(\ell(a)) - f'(\ell(a)) \cdot m \cdot (x - a)| \leq K \cdot m^2 \cdot |x - a|^2.$$

Die Komposition $f \circ \ell$ hat also die Ableitung: $(f \circ \ell)'(a) = f'(\ell(a)) \cdot \ell'(a)$.
Dies erwarten wir auch für nicht lineares ℓ als allgemeine Differentiationsregel.

Kommentar. Im 19. Jahrhundert hat sich durchgesetzt, Funktionen auch dann schon differenzierbar zu nennen, wenn sie gut, aber nicht ganz so gut, wie wir es bei den Polynomen beobachtet haben, durch lineare Funktionen (weiterhin Tangenten genannt) approximierbar sind. Wesentlich ist, daß der Unterschied zwischen $f(x)$ und Tangente $T_a(x)$ bei a schneller klein wird als proportional zum Abstand $|x - a|$. Beispiele für solche gröberen aber genügend guten Fehler sind: $\mathrm{const} \cdot |x - a|^{1+p}$ mit $0 < p \leq 1$. (Nur die linearen Funktionen erlauben $p > 1$.) Dem Vorteil solcher Fehler, nämlich explizit zu sein, steht als Mangel gegenüber, daß man solche expliziten Abschätzungen nicht in allen interessanten Fällen beweisen kann. Die endgültige Definition (die 200 Jahre jünger als Newton ist!) arbeitet mit noch gröberen Fehlern, die ich erst etwas später benutzen werde. Man drückt das so aus: Die Fehler sind $\leq \epsilon \cdot |x - a|$; dabei ist jedoch ϵ nicht eine feste Konstante, das wäre nicht gut genug; man darf ein beliebig kleines $\epsilon > 0$ als Faktor wählen, aber dann muß man dafür dadurch bezahlen, daß die Fehlerabschätzung nur in immer kleineren Intervallen gilt. Wir werden später definieren (aber jetzt schon kommentieren):

Endgültige Differenzierbarkeitsdefinition. Zu jedem Fehlerfaktor $\epsilon > 0$ gibt es ein (eventuell winzig kleines) $\delta(\epsilon) = \delta > 0$, so daß wenigstens in dem Intervall $[a - \delta, a + \delta]$ gilt:

$$|x - a| \leq \delta \Rightarrow |f(x) - f(a) - f'(a) \cdot (x - a)| \leq \epsilon \cdot |x - a|.$$

Selbst diese Formulierung gibt es noch in Variationen: Zu jedem ϵ kann man dies δ auf einem großen Intervall $[R, S]$ *unabhängig* von $a \in [R, S]$ finden, oder noch schwächer: δ muß sogar in Abhängigkeit von a gefunden werden.

Trotz all dieser Variationsmöglichkeiten bei der Formulierung von Definitionen und Voraussetzungen von Sätzen sind die Differentiations*regeln* immer dieselben. Es gilt nämlich: Wenn zwei Funktionen f und g "so-und-so-gut" durch ihre Tangenten approximiert werden, dann haben $f \cdot g$ und $f \circ g$ die aus den Vorüberlegungen erwarteten Ableitungen und sie werden von ihren Tangenten im selben Sinne wie f und g "so-und-so-gut" approximiert. Man kann dieses "so-und-so-gut" in Formeln so audrücken:

$$f(x) - f(a) - f'(a) \cdot (x - a) = \phi(x, a) \cdot |x - a|,$$

und *je nach den gerade benötigten Voraussetzungen* hat man:

$$x \in [a - r, a + r] \;\Rightarrow\; |\phi(x, a)| \leq K \cdot |x - a|^p, \;\; (0 < p \leq 1)$$

oder, zu jedem $\epsilon > 0$ gibt es ein $\delta > 0$ so daß gilt

$$x \in [a - \delta, a + \delta] \;\Rightarrow\; |\phi(x, a)| \leq \epsilon,$$

und zwar entweder mit Konstanten r, K oder δ, die noch von a abhängen, oder besser, mit einer Wahl dieser Konstanten, die gleichmäßig für alle a in einem gerade interessierenden Teilbereich des Definitionsgebietes von f gilt. All diese Variationsmöglichkeiten kommen vor, aber die Beweise der Differentiationsregeln müssen nur in der letzten Zeile an die genauen Voraussetzungen angepaßt werden. Ich beweise zunächst die Produktregel, dabei gehe ich erst am Ende auf die genauen Voraussetzungen ein. Wie die meisten Autoren nehmen wir an, dass r bzw. δ bereits so klein gewählt ist, dass $|\phi(x, a)| \leq 1$ gilt.

Beweis der Produktregel:
Voraussetzung: $f_i(x) = f_i(a) + f_i'(a) \cdot (x - a) + \phi_i(x, a) \cdot |x - a|, \; i = 1, 2$ soll gelten mit Funktionen ϕ_i, die in einer der eben besprochenen Weisen "kleine" Fehler beschreiben. Dann liefert

$$f_1(x) \cdot f_2(x) =$$
$$\Big(f_1(a) + f_1'(a) \cdot (x - a) + \phi_1(x, a) \cdot |x - a|\Big) \cdot \Big(f_2(a) + f_2'(a) \cdot (x - a) + \phi_2(x, a) \cdot |x - a|\Big)$$

nach Ausmultiplizieren die Differenz zwischen Produkt und erwarteter Tangente:

$$(f_1 \cdot f_2)(x) \; - \; \big((f_1 \cdot f_2)(a) + (f_1' \cdot f_2 + f_1 \cdot f_2')(a) \cdot (x - a)\big) =$$
$$\Big[(f_1(a) + f_1'(a) \cdot (x - a)) \cdot \phi_2(x, a) + (f_2(a) + f_2'(a) \cdot (x - a)) \cdot \phi_1(x, a)$$
$$+ (f_1'(a) \cdot f_2'(a) + \phi_1(x, a) \cdot \phi_2(x, a)) \cdot |x - a|\Big] \cdot |x - a|.$$

Zunächst vereinfachen wir nur unter der Annahme, daß a, x aus einem beschränkten (aber evtl. sehr großen) Intervall sind, etwa $a, x \in [-R, R]$. Damit ist der Ausdruck in den eckigen Klammern (beachte $|\phi(x,a)| \leq 1$)

$$[\ldots] \leq const_1 \cdot \phi_2 + const_2 \cdot \phi_1 + (1 + |(f_1' \cdot f_2')(a)|) \cdot |x - a|,$$

wobei offenbar die Konstanten $const_1, const_2$ aus $f_i(a), f_i'(a)$ und der Intervallänge $2R$ explizit berechnet werden können. Eine solche Kombination von ϕ_1 und ϕ_2 ist aber im selben Sinne "klein", wie es für ϕ_1 und ϕ_2 vorausgesetzt wurde, weil der dritte Term $(1 + |(f_1' \cdot f_2')(a)|) \cdot |x - a|$ die schärfste Forderung, "klein" zu sein, erfüllt. Zum Beispiel: Aus

$$|x - a| \leq r_1 \Rightarrow |\phi_1(x,a)| \leq M_1 \cdot |x - a|^p, \text{ und } |x - a| \leq r_2 \Rightarrow |\phi_2(x,a)| \leq M_2 \cdot |x - a|^p$$

folgt (beachte $|x - a| \leq |x - a|^p$)

$$\begin{aligned}
&|x - a| \leq \min(r_1, r_2, 1) \Rightarrow \\
&\big|(const_1 \cdot \phi_2 + const_2 \cdot \phi_1)(x,a) + (1 + |(f_1' \cdot f_2')(a)|) \cdot |x - a|\big| \leq \\
&\qquad\qquad \leq (const_1 \cdot M_2 + const_2 \cdot M_1 + const_3) \cdot |x - a|^p \\
&\qquad\qquad = M_3 \cdot |x - a|^p.
\end{aligned}$$

Der Unterschied zwischen dem Produkt $f_1 \cdot f_2$ und der erwarteten Tangente ist also in der Tat ein im selben Sinne kleiner Fehler wie für die f_i vorausgesetzt, im Beispiel $\leq M_3|x - a|^{p+1}$ im Intervall $|x - a| \leq \min(r_1, r_2, 1)$. Alle späteren Beweise von Produktregeln folgen genau diesem Muster, insbesondere kann das hier mit $\cdot$ bezeichnete Produkt später irgendein bilineares Produkt sein (Skalarprodukt, Kreuzprodukt, Matrixprodukt, …).

Beweis der Kettenregel.

Im Beweis der Kettenregel ist es wesentlich, mit Hilfe der Voraussetzungen die innere Funktion so genau zu kontrollieren, daß ihre Werte in ein so kleines Intervall fallen, daß die Voraussetzungen über die äußere Funktion anwendbar sind. In der vorausgeschickten Rechnung für $f \circ \ell$ war das nur die Folgerung $|x - a| \leq R/m \Rightarrow |\ell(x) - \ell(a)| = |m(x - a)| \leq R$, für eine nichtlineare innere Funktion ist hier mehr Mühe erforderlich. – Anders als bei der Produktregel formuliere ich den Beweis nicht für alle Sorten "kleiner Fehler" auf einmal, ich beschränke mich auf die bei den Polynomen beobachteten quadratischen Fehler. Ich verschiebe die Diskussion der ϵ-δ-Fehler, da wir ohnehin in höherdimensionalen Situationen auf den Beweis der Kettenregel zurückkommen müssen.

Voraussetzung:

$$|x - a| \leq r_1 \Rightarrow |f(x) - f(a) - f'(a) \cdot (x - a)| \leq M_1(x - a)^2, \quad f(a) = A \text{ und}$$

$$|Y-A| \le R_2 \Rightarrow |F(Y)-F(A)-F'(A)\cdot(Y-A)| \le M_2 \cdot |Y-A|^2.$$

Dann folgt, falls $|f(x)-f(a)| \le R_2$ ist (!!), als erster wichtiger Schritt:

$$(*) \qquad |F(f(x)-F(f(a))-F'(f(a))\cdot(f(x)-f(a))| \le M_2 \cdot |f(x)-A|^2.$$

Ferner (Dreiecksungleichung in der Voraussetzung über f):

$$(**) \qquad |x-a| \le r_1 \Rightarrow |f(x)-f(a)| \le (|f'(a)|+M_1\cdot r_1)\cdot|x-a|$$

und (Multiplikation der Voraussetzung über f mit $|F'(f(a))|$) :

$$|x-a| \le r_1 \Rightarrow |F'(f(a))\cdot(f(x)-f(a))-F'(f(a))\cdot f'(a)\cdot(x-a)| \le |F'(f(a))|\cdot M_1\cdot|x-a|^2.$$

Jetzt muß man nur noch mit dieser letzten Ungleichung in (*) $F'(f(a))\cdot(f(x)-f(a))$ durch $F'(f(a))\cdot f'(a)\cdot(x-a)$ ersetzen, die Dreiecksungleichung benutzen und die Konstanten sortieren zu: Const $:= M_2(|f'(a)|+M_1\cdot r_1)^2+|F'(f(a))|\cdot M_1$,
um

$$|F\circ f(x)-F\circ f(a)-F'\circ f(a)\cdot f'(a)\cdot(x-a)| \le \text{Const}\cdot(x-a)^2$$

zu erhalten. Aber nun darf unter keinen Umständen übersehen werden, daß innerhalb des Beweises $|f(x)-A| \le R_2$ garantiert werden muß, sonst kann die Voraussetzung über F gar nicht auf $Y=f(x)$ angewandt werden. Beachten Sie unbedingt, daß dafür das ursprüngliche r_1 (eventuell) noch verkleinert werden muß: Nur wenn wir

$$|x-a| \le r := \min\left(r_1,\ R_2\cdot(|f'(a)|+M_1\cdot r_1)^{-1}\right)$$

voraussetzen, folgt aus der Voraussetzung über f, nach der Umarbeitung in (**), wirklich, daß $|x-a| \le r$ impliziert $|f(x)-A| \le R_2$, so daß die behaupteten Ungleichungen über $F\circ f$ allein aus den Voraussetzungen über f und F folgen. – Der Beweis mit ϵ-δ-Fehlern wird nach demselben Muster geführt werden, allerdings stehen weder in den Voraussetzungen *explizite* Intervalle, noch kann der Beweis daraus explizite Intervalle machen, in denen die behaupteten Ungleichungen richtig sind.

Anwendungen

Inversenregel: $$(\frac{1}{f})' = -\frac{f'}{f^2}.$$

Beweis. Benutze $F(x) = 1/x, F'(x) = -1/x^2$ (s.o.) und die gerade bewiesene Kettenregel. Damit haben wir auch eine

Quotientenregel: $$(\frac{f}{g})' = (f\frac{1}{g})' = \frac{f'}{g} - \frac{f \cdot g'}{g^2} = \frac{f'g - fg'}{g^2} = \frac{f}{g} \cdot (\frac{f'}{f} - \frac{g'}{g}).$$

Die letzte Form der Quotientenregel ist z.B. bei der Diskussion prozentualer (oder: relativer) Fehler nützlich. In naturwissenschaftlichen Anwendungen sind die quadratischen Fehler oft vernachlässigbar klein, dann gibt die Ableitung an, wie sich

Fehler der Argumente auf Fehler der Werte einer Funktion auswirken:

Absolute Fehler: $$\triangle f := f(x) - f(a) \approx f'(a) \cdot (x - a)$$

Oft interessiert man sich mehr für die sogenannten *relativen Fehler*, für die man durch $f(a) > 0$ dividieren muß:

Relative Fehler: $$\frac{\triangle f}{f} := \frac{f(x) - f(a)}{f(a)} \approx \frac{f'}{f}(a) \cdot (x - a).$$

Bezeichnung: Für positive Funktionen $f > 0$ heißt f'/f deren *Wachstumsrate*; z.B. ist die Geburtenrate solch ein Quotient. Diese Wachstumsraten f'/f kommen in naturwissenschaftlichen Anwendungen sicher so häufig vor wie die Steigungen f'.

Kurven in der Ebene. Mit Paaren von Funktionen $f, g : [a, b] \to \mathbb{R}$ können wir Kurven in der Ebene beschreiben. Z.B. liefert $f(t) := 2t/(1 + t^2)$, $g(t) = (1 - t^2)/(1 + t^2)$ eine Abbildung von $\mathbb{R}$ auf den Einheitskreis (außer $(0, -1)$):

$$(f, g) : \mathbb{R} \to \mathbb{R}^2, \quad t \to \big(\frac{2t}{(1 + t^2)}, \frac{(1 - t^2)}{(1 + t^2)}\big), \qquad f(t)^2 + g(t)^2 = 1.$$

An dieser Abbildung ist z.B. bemerkenswert, daß rationale Punkte in $\mathbb{R}$ auf Kreispunkte mit rationalen Koordinaten abgebildet werden und auch umgekehrt ($p^2 + q^2 = 1 \Rightarrow t := (1 + q)/p$). Andererseits werden wir sehen, daß der Kreis nicht mit konstanter Geschwindigkeit durchlaufen wird. (Funktionen, die das leisten, können wir unter den rationalen Funktionen nicht finden, sin und cos werden am Ende des Abschnitts über die Vollständigkeit der reellen Zahlen konstruiert.)

Um zur Definition von Ableitungen und Tangenten zu kommen, betrachten wir den einfachsten Fall zuerst. f und g seien lineare Funktionen:

$$f(t) := p_1 + m_1 \cdot t, \qquad g(t) := p_2 + m_2 \cdot t.$$

Dann können wir schreiben:

$$p(t) := (f,g)(t) = (p_1,p_2) + (m_1,m_2)\cdot t, \qquad \frac{p(t_2)-p(t_1)}{t_2-t_1} = (m1,m2).$$

Alle Durchschnittsgeschwindigkeiten dieser Bewegung sind also konstant, $v = (m_1,m_2)$. Natürlich werden wir verabreden, daß wir in diesem übersichtlichen Fall die *Momentangeschwindigkeit* als v definieren.

Kompliziertere Kurven sollen nun mit diesen einfachsten, linearen Kurven verglichen werden.

Strategie: Eine Kurve $(f_1,f_2) : [a,b] \to \mathbb{R}^2$ hat an der Stelle t_0 die "Ableitung" oder "Geschwindigkeit" $v = (f_1'(t_0), f_2'(t_0))$, wenn die lineare Kurve

$$t \to (f_1(t_0), f_2(t_0)) + (f_1'(t_0), f_2'(t_0))\cdot(t-t_0)$$

in der Nähe von $t = t_0$ genügend wenig von $(f_1(t), f_2(t))$ abweicht. Es muß noch präzisiert werden, was mit "genügend wenig abweicht" gemeint ist. Für diese Präzisierung verallgemeinern wir mit Hilfe des Satzes von Pythagoras die "kleinen Fehler", die wir über f_1, f_2 voraussetzen. Gegeben sei

$$f_i(t) - f_i(t_0) - f_i'(0)\cdot(t-t_0) = \phi_i(t,t_0)\cdot|t-t_0|, \quad (i=1,2),$$

wobei $\phi_i(t,t_0)$ die gerade benutzten "kleinen Fehler" beschreibt, also je nach Situation explizite Fehler wie $|\phi_i(t,t_0)| \le M_i \cdot |t-t_0|$ in Intervallen $[t_0-r_i, t_0+r_i]$ oder ϵ-δ-Fehler $|\phi_i(t,t_0)| \le \epsilon$ in Intervallen $[t_0-\delta, t_0+\delta]$. (Hier sind M_i und $r := \min(r_1,r_2)$ Konstanten, die von t_0 abhängen dürfen oder in günstigeren Fällen unabhängig von t_0 sind (man sagt dann, die Konstanten können *gleichmäßig* gewählt werden), während bei den ϵ-δ-Fehlern zu *jedem* $\epsilon > 0$ ein $\delta > 0$ vorhanden sein muß , ebenfalls i.a. $\delta = \delta(\epsilon, t_0)$, oder besser ein $\delta = \delta(\epsilon)$ für alle t_0 (gleichmäßig) .)

Unabhängig von einer derartigen Präzisierung der "kleinen Fehler" gilt jedenfalls wegen des Satzes von Pythagoras $\qquad t \in [t_0 - r, t_0 + r] \Rightarrow$

$$\big|(f_1,f_2)(t) - (f_1,f_2)(t_0) - (f_1',f_2')(t_0)\cdot(t-t_0)\big| \le \sqrt{\phi_1(t,t_0)^2 + \phi_2(t,t_0)^2}\cdot|t-t_0|.$$

Und wieder kann man für jede Sorte kleiner Fehler ϕ_1, ϕ_2 leicht einsehen, daß auch $\sqrt{\phi_1^2+\phi_2^2}$ ein kleiner Fehler *derselben Art* ist, z.B. für ein p mit $0 < p \le 1$

$$|t-t_0| \le r \Rightarrow |\phi_i(t,t_0)| \le M_i \cdot |t-t_0|^p \quad (i=1,2)$$

hat zur Folge

$$|t - t_0| \le r \Rightarrow \sqrt{\phi_1(t, t_0)^2 + \phi_2(t, t_0)^2} \le \sqrt{M_1^2 + M_2^2} \cdot |t - t_0|^p =: M \cdot |t - t_0|^p.$$

Mit anderen Worten: *Wir können das Differenzieren von Funktionen $f : [a, b] \to \mathbb{R}$ leicht auf das Differenzieren von Kurven $(f, g) : [a, b] \to \mathbb{R}^2$ verallgemeinern.*
Die einfache Differentiationsregel hierzu lautet "komponentenweise Differenzieren":

Ableitung von Kurven: $$(f, g)'(t) = \big(f'(t),\ g'(t)\big).$$

Nachdem wir jetzt Funktionen und Kurven differenzieren können, haben wir die anfänglichen Überlegungen ohne Definitionen, mit denen wir Tangenten für andere Kurven als Kreise gesucht haben, ganz in den Aufbau der Theorie eingebunden.

Beispiel. Für die eben angegebene Kreiskurve

$$k(t) := \Big(\frac{2t}{(1+t^2)}, \frac{(1-t^2)}{(1+t^2)}\Big) \;=\; \Big(\frac{2t}{(1+t^2)}, \frac{2}{(1+t^2)} - 1\Big)$$

finden wir

$$k'(t) = \frac{2}{(1+t^2)^2} \cdot (1 - t^2,\ -2t), \qquad |k'(t)| = \frac{2}{(1+t^2)}.$$

Offenbar ist die Geschwindigkeit (und damit die Tangente) immer senkrecht zum Radius, aber der Betrag der Geschwindigkeit dieser Kreisbewegung ist nicht konstant.

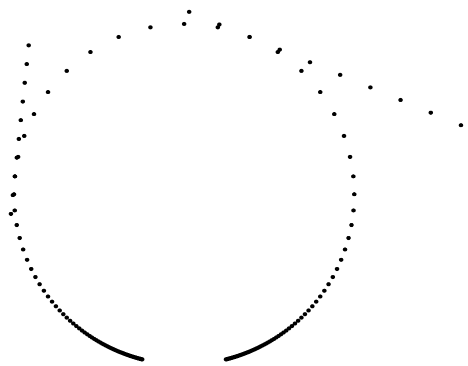


Bild (nicht Graph) dieser Kreisparametrisierung, mit zwei parametrisierten Tangenten.

Kommentar. Wir können jetzt eine große Klasse von Funktionen und Kurven differenzieren (und mit Hilfe der Differentiationsregeln auch schnell differenzieren), aber wir haben noch keinen Satz kennen gelernt, der aus Ableitungsvoraussetzungen Folgerungen über die Funktion zieht. Diesem wichtigen Analysisthema wenden wir uns im nächsten Abschnitt zu.

Der Monotoniesatz

> Der Monotoniesatz und seine Verwandten erlauben, aus Ableitungsvoraussetzungen Eigenschaften der Funktion zu beweisen. Man muß erstens einsehen, warum solche Folgerungen zentrale Argumente der Analysis sind. Zweitens, mit um so schwächeren Differenzierbarkeitsvoraussetzungen man arbeitet, um so subtilere Argumente benötigt man zum Beweis des Monotoniesatzes. Üblicher Weise wird die Vollständigkeit der reellen Zahlen in einem indirekten Beweis benutzt. Die bei allen bisher besprochenen Funktionen vorhandenen Gleichmäßigkeitseigenschaften erlauben jedoch, direkte Beweise ohne die Vollständigkeit der reellen Zahlen zu führen. Diese Beweise funktionieren daher schon, wenn man nur die rationalen Zahlen kennt. Sie erlauben, ein wesentliches Stück quantitativer Analysis der rationalen Funktionen zu entwickeln, ehe man sich der Vollständigkeit zuwenden muß.

Die eben besprochenen Regeln zum Berechnen von Ableitungen erklären natürlich nicht, welchen *Nutzen* das Berechnen von Ableitungen haben könnte. Mit dem Monotoniesatz wenden wir uns dieser Frage zu. Während wir die bisherigen Ableitungseigenschaften von Polynomen einfach ausrechnen konnten, wird das jetzt nicht mehr der Fall sein, wir müssen argumentieren, und wir beweisen Aussagen, die nicht durch Nachrechnen verifiziert werden können.

Zunächst wiederhole ich die Definition mit dem Zusatz "gleichmäßig":

Definition. Eine Funktion $f\colon (\alpha, \omega) \to \mathbb{R}$ heißt *gleichmäßig differenzierbar* mit sogar quadratischer Approximation der Tangenten, falls gilt:

Es gibt zu jedem $a \in (\alpha, \omega)$ ein Intervall $[a-r, a+r]$ und Konstanten m und K, (wobei (r, K) ausdrücklich nicht von a abhängen sollen), so daß gilt:

$$x \in (\alpha, \omega),\ |x-a| \le r \Rightarrow |f(x) - f(a) - m\cdot(x-a)| \le K\cdot(x-a)^2.$$

Bezeichnungen: $m = f'(a)$ heißt Ableitung oder Steigung von f bei a.

$T(x) = f(a) + f'(a)\cdot(x-a)$ heißt Tangente von f bei a.

Von jetzt an wird ein Argument wichtig, das schon auf Archimedes zurückgeht:

Archimedes-Argument: Um eine Ungleichung $a \le b$ zu beweisen, genügt es, für jede positive Zahl $p > 0$ zu beweisen: $a \le b + p$.

Beweis. Wäre $a \le b$ falsch, also $a > b$, so wähle $p := \frac{1}{2}(a-b) > 0$. Die nach Voraussetzung beweisbare Ungleichung $a \le b+p$ liefert $a \le b + \frac{1}{2}(a-b)$, also $\frac{1}{2}a \le \frac{1}{2}b$, im Widerspruch zu der Annahme $a > b$.

Ich kenne keinen anderen Beweis, der den Titel *Prototyp eines indirekten Beweises* mehr

verdient als dieser.
Das Archimedes-Argument wird in der Analysis außerordentlich oft benutzt; Archimedes selber hat mit spezielleren Voraussetzungen gearbeitet, nämlich nur mit Stammbrüchen $p = 1/n$; dies wird im nächsten Kapitel nach der Formulierung des Archimedes Axioms für die reellen Zahlen ausführlicher erläutert.

Nach diesen Vorbemerkungen kommen wir zum Kern, dem Prototyp eines Satzes, der aus Voraussetzungen über f' auf Eigenschaften von f schließt, nämlich dem

Monotoniesatz. f sei differenzierbar mit einer Ableitung $f' \geq 0$, dann ist f schwach wachsend:

$$x \leq y \Rightarrow f(x) \leq f(y).$$

Folgerungen 1 - 6. Die Behauptung dieses Satzes wird natürlich durch die sprachliche Formulierung ("Steigung") nahe gelegt. Trotzdem liegt der Beweis nicht auf der Hand. Der Satz ist auch schon zur Behandlung von Polynomen sehr nützlich.
Als Werbung für den Monotoniesatz zähle ich einige unmittelbare Folgerungen auf:

1.) Verallgemeinerter Monotoniesatz.
Funktionen mit der größeren Ableitung wachsen schneller:
$$x \leq y \text{ und } f' \leq g' \text{ impliziert } f(y) - f(x) \leq g(y) - g(x),$$
denn $(g - f)$ erfüllt die Voraussetzungen des Monotoniesatzes.

2.) Schrankensatz.
Eine Schranke für f' ist auch eine Schranke für alle Sehnensteigungen von f:
$$|f'| \leq L \Rightarrow |f(y) - f(x)| \leq L \cdot |y - x|,$$
denn mit der linearen Funktion $g(x) := L \cdot x$ gilt $-g' \leq f' \leq g'$. Mit der ersten Folgerung haben wir für $x \leq y$: $\;-L \cdot (y - x) \leq f(y) - f(x) \leq +L \cdot (y - x)$.
Funktionen mit dieser Eigenschaft spielen eine so große Rolle, daß sie einen Namen haben:

Definition. Funktionen, die $|f(y) - f(x)| \leq L \cdot |y - x|$ erfüllen, heißen *lipschitzstetig* (oder auch *dehnungsbeschränkt*) mit Lipschitzschranke (oder Dehnungsschranke) L.

3.) Zweite Ableitung und Tangente, einseitiges Berühren.
Aus $f'' \geq 0$ folgt: f liegt oberhalb jeder Tangente.
Beweis. Sei $T_a(X) := f(a) + f'(a) \cdot (x - a)$ die Tangente bei a, dann gilt für die Hilfsfunktion $h := f - T_a$ zunächst $h'' \geq 0$. Ferner ist $h'(a) = 0$, also folgt: $x \leq a \Rightarrow h'(x) \leq h'(a) = 0,\; a \leq x \Rightarrow h'(x) \geq 0$. Daher ist nach dem Monotoniesatz h rechts von a wachsend, links von a fallend, wegen $h(a) = 0$ folgt daraus $h \geq 0$ oder $f \geq T_a$.

Anwendung. Die Bernoulli'sche Ungleichung: $-1 \le x, n \in \mathbb{N} \Rightarrow (1+x)^n \ge (1+n\cdot x)$ ist eine unmittelbare Folge, denn $((1+x/n)^n)'' \ge 0$, falls $-1 \le x$.
Auch bei vielen anderen Polynombeispielen ist die Anwendung von 3.) ein besonders bequemes Argument.

4.) Zweite Ableitung und Abweichung von der Tangente.
Aus $|f''| \le K$ folgt: f weicht von seiner Tangente bei a höchstens um $\frac{1}{2}K\cdot(x-a)^2$ ab.
Denn, für die Hilfsfunktion $h(x) := f(x) - T_a(x)$ gilt $h(a) = 0 = h'(a), -K \le h'' \le K$.
Wir wenden Folgerung 1 zweimal an, zuerst:

$$x \ge a \to -K\cdot(x-a) \le h'(x) - h'(a) = h'(x) \le +K\cdot(x-a),$$
$$x \le a \to -K\cdot(a-x) \le h'(x) - h'(a) = h'(x) \le +K\cdot(a-x)$$

und dann noch einmal:

$$-\frac{1}{2}K\cdot(x-a)^2 \le h(x) - h(a) = f(x) - T_a(x) \le \frac{1}{2}K\cdot(x-a)^2.$$

Diese Folgerung erklärt also die Bedeutung der Konstanten in der quadratischen Abweichung von der Tangente, jede Schranke für die zweite Ableitung ist eine geeignete Konstante. Dies Ergebnis ist oft wesentlich besser als die für Polynome aus den Beträgen der Koeffizienten ausgerechnete "Tangentenabweichung" in Kapitel 3.

5.) Zweite Ableitung und Abweichung von der Sehne.
Aus $0 \le f''$ folgt: In jedem Intervall $[a,b]$ liegt f unterhalb der Sehne S_{ab}.
Die Sehne ist die lineare Funktion $S_{ab}(x) := (f(a)\cdot(b-x) + f(b)\cdot(x-a))/(b-a)$.
Für die Hilfsfunktion $h(x) := f(x) - S_{ab}(x)$ gilt $h'' \ge 0,\ h(a) = 0 = h(b)$.
Erstens ist h' wachsend wegen $h'' \ge 0$.
Nun sei $c \in (a,b)$ beliebig. Entweder ist $h'(c) \ge 0$, dann ist $h' \ge 0$ im Intervall $[c,b]$ (h' ist wachsend!) und daher $h(c) \le h(b) = 0$; oder $h'(c) < 0$, dann ist $h' \le h'(c) < 0$ im Intervall $[a,c]$ (h' ist wachsend!) und daher $0 = h(a) > h(c)$. In beiden Fällen ist also $h(c) \le 0$ und damit f unterhalb der Sehne.
Gilt $f'' \le B$, so erfüllt $h,\ h(x) := -0.5B(x-a)(b-x) + S_{ab}(x) - f(x)$ wie eben $0 \le h''$.
Aus $h \le 0$ folgt $S_{ab}(x) - 0.5B(x-a)(b-x) \le f(x)$.

6.) Multiplikativer Monotoniesatz (od. Schrankensatz) für positive Funktionen.
Wir hatten im Zusammenhang mit relativen Fehlern die Wachstumsraten von Funktionen kennen gelernt. Wegen der Quotientenregel gilt:

$$\frac{f'}{f} \le \frac{g'}{g} \quad\Rightarrow\quad (\frac{g}{f})' = \frac{g}{f}\cdot(\frac{g'}{g} - \frac{f'}{f}) \ge 0.$$

Nach dem Monotoniesatz ist dann g/f wachsend, also gilt:

$$a \le x \ \text{ und } \ \frac{f'}{f} \le \frac{g'}{g} \Rightarrow \frac{g(a)}{f(a)} \le \frac{g(x)}{f(x)} \Rightarrow \frac{f(x)}{f(a)} \le \frac{g(x)}{g(a)}.$$

Wir illustrieren diese multiplikative Version durch eine ausführlichere Anwendung, bei der ich vor allem den geringen Rechenaufwand hervorheben möchte:
Betrachte für $0 \le x$ und $m \in \mathbb{N}$ die Polynome $f_m(x) := (1 + x/m)^m$. Ihre Wachstumsraten sind für kleine x wenig von 1 verschieden, deshalb werden sich diese Polynome als gute Approximationen der Exponentialfunktion erweisen. Falls $m < n$ ist, haben wir

$$\frac{f_m'}{f_m}(x) = \frac{1}{1 + x/m} \le \frac{1}{1 + x/n} = \frac{f_n'}{f_n}(x).$$

Zusammen mit $f_m(0) = 1 = f_n(0)$ folgt aus dem multiplikativen Monotoniesatz:

$$0 \le x \text{ und } m < n \Rightarrow (1 + \frac{x}{m})^m \le (1 + \frac{x}{n})^n.$$

Es ist sehr viel mühsamer diese Monotonie (bei festem x von $n \to (1 + x/n)^n$) ohne den multiplikativen Monotoniesatz zu beweisen.
Nachdem man diese wachsende Folge hat, entsteht die Frage: Wie stark wächst sie? Bleibt sie bei festem x beschränkt? Trotz des einfachen Aussehens dieser Polynome kann man eine Schranke nicht einfach ausrechnen, aber mit dem multiplikativen Monotoniesatz geht es schnell. Unser Ansatz verdient jedoch einen Kommentar: Interpretiert man $1 + x$ als Gutschrift von x Zinsen auf das Anfangskapital am Jahresende, so bedeutet $(1 + x/2)^2$ zweimaliges Gutschreiben der halben Zinsen, usw. Betrachtet man die Sache rückwärts und möchte, daß das Anfangskapital um x Zinsen vom Endkapital kleiner ist als dieses, so hat man die Anfangssumme durch $(1 - x)$ zu dividieren, um die Endsumme zu bekommen. Entsprechend hat man $1/(1 - x/k)^k$ für k Gutschriften von x/k Rückwärtszinsen.
Betrachte daher

$$h_j(x) := \frac{1}{(1 - x/j)^j} \ \text{ für } \ x \in [0, j);$$

es gilt

$$h_j(0) = 1, \ \text{und } \ \frac{h_j'}{h_j}(x) = \frac{1}{1 - x/j} \ge 1 \ge \frac{1}{1 + x/n} = \frac{f_n'}{f_n}(x).$$

Somit liefert der multiplikative Monotoniesatz ohne weitere Rechnungen

$$0 \le x < j \Rightarrow (1 + x/n)^n \le 1/(1 - x/j)^j.$$

Die Zinseszins-Interpretation der ersten Formel hat also zu einer wachsenden und einer fallenden Folge rationaler Funktionen geführt, die bequeme Schranken von einander sind.

So weit Folgerungen aus dem Monotoniesatz, nun zu seinem Beweis. Am üblichsten ist ein indirekter Beweis, der die Vollständigkeit der reellen Zahlen (vgl. nächstes Kapitel) wesentlich benutzt. Da ich bisher ohne diese Vollständigkeit ausgekommen bin – es brauchten ja nicht mehr Zahlen als die rationalen bekannt zu sein – möchte ich auch diesen wichtigen Satz zunächst ohne die Vollständigkeit herleiten. Das hat noch den zusätzlichen Nutzen, daß der Beweis ein Schritt in Richtung Integralrechnung ist. Andererseits hat der Verzicht auf die Vollständigkeit seinen Preis: Wenn die benutzten Differenzierbarkeitsvoraussetzungen tatsächlich so schwach sind, daß für die Konstante K und die Länge r derjenigen Intervalle, auf denen die quadratische Ungleichung nach Voraussetzug gilt, *nicht* von a unabhängige Werte genommen werden können, sondern daß diese Konstanten K, r ("leider") in Abhängigkeit von a gewählt werden müssen, dann braucht man die Vollständigkeit, um im indirekten Beweis die Stelle zu finden, an der der Widerspruch auftritt.
Mir scheint, daß die Raffinesse dieser verfeinerten Argumente deutlicher hervortritt, wenn wir zunächst in einfacheren Situationen mit einfacheren Argumenten auskommen und dann sehen, an welchen Schwierigkeiten diese einfacheren Argumente scheitern.
Im Augenblick haben wir nur die rationalen Funktionen zur Verfügung, die, wie wir gesehen haben, in besonders guter Weise von ihren Tangenten approximiert werden. Auch die nächsten Funktionen, die wir konstruieren werden, *sin, cos, exp,* allgemeiner Potenzreihen, haben diese guten Eigenschaften. Es wird erheblicher Anstrengungen bedürfen, weniger gut differenzierbare Funktionen zu konstruieren. Daher machen wir zunächst die stärkeren "Gleichmäßigkeits"-Voraussetzungen.

Beweis des Monotoniesatzes bei gleichmäßiger Differenzierbarkeit.
Hauptvoraussetzung: f sei auf $[\alpha, \omega]$ differenzierbar und $f' \geq 0$.
Zusatzvoraussetzung: Es gibt positive Konstanten r, K so daß für die Tangentenapproximation "gleichmäßig" gilt:

$$a, x \in [\alpha, \omega],\ |a - x| \leq r \ \Rightarrow\ |f(x) - f(a) - f'(a) \cdot (x - a)| \leq K \cdot |x - a|^2.$$

Diese Zusatzvoraussetzung erlaubt, auf die Vollständigkeit zu verzichten, also z.B. die rationalen Funktionen nur über $\mathbb{Q}$ zu betrachten.

Behauptung: f ist schwach monoton wachsend, d.h.

$$x \leq y \in [\alpha, \omega] \quad \Rightarrow \quad f(x) \leq f(y).$$

Bemerkung: Bei der Zusatzvoraussetzung kommt es darauf an, daß die Konstanten r, K von a unabhängig sind; man sagt: "gleichmäßig" gelten. Der folgende Beweis funktioniert für alle Sorten "kleiner Fehler", sofern die daran beteiligten Konstanten gleichmäßig gewählt werden können. Ein Beweis ohne diese Gleichmäßigkeit folgt nach der Diskussion der Vollständigkeit, vgl. S.71, S.122.

Beweis: Teile das Teilintervall $[x, y]$ in n gleiche Teile, wobei n mindestens so groß ist, daß $(y - x)/n \leq r$ gilt, also $n \geq (\omega - \alpha)/r$. Bezeichne $t_k := x + \frac{k}{n} \cdot (y - x), k = 0, \ldots, n$.
Dann hat man wegen der Zusatzvoraussetzung eine Konstante K unabhängig von n, k so daß die folgende rechte Ungleichung gilt, und wegen $f' \geq 0$ dann auch die linke:

$$-K \cdot (t_k - t_{k-1})^2 \leq -K \cdot (t_k - t_{k-1})^2 + f'(t_{k-1}) \cdot (t_k - t_{k-1}) \leq f(t_k) - f(t_{k-1}).$$

In diese Ungleichungen wird $t_k - t_{k-1} = (y - x)/n$ eingesetzt und über $k = 1, \ldots, n$ summiert:

$$x \leq y \quad \Rightarrow \quad -K \cdot (y - x)^2 \cdot \frac{1}{n} \leq f(y) - f(x).$$

Schließlich wird dies mit dem Archimedes Argument zu der Behauptung $0 \leq f(y) - f(x)$ verbessert. – Ich fasse den Beweis noch einmal zusammen: Wendet man die Differenzierbarkeitsdefinition auf "kleine" Intervalle an, so erhält man die Teilintervall-Behauptung bis auf "sehr kleine" Fehler; summiert man die Teilergebnisse, so erhält man die Behauptung bis auf einen irrelevanten $1/n$-Archimedes-Fehler.

Aufgabe. Es sei die Kurve $(f, g) : [\alpha, \omega] \to \mathbb{R}^2$ gleichmäßig differenzierbar mit beschränkter Ableitung: $\big|(f, g)'\big| := \sqrt{f'^2 + g'^2} \leq L$. Modifiziere den vorhergehenden Beweis etwas und zeige, daß L eine Dehnungsschranke der Kurve ist:
$x, y \in [\alpha, \omega] \;\Rightarrow\; \big|(f, g)(x) - (f, g)(y)\big| := \sqrt{(f(x) - f(y))^2 + (g(x) - g(y))^2} \leq L \cdot |x - y|$.
Ähnliche Argumente liefern uns weitere Eigenschaften von gleichmäßig differenzierbaren Funktionen (insbesondere also von Polynomen).

Satz. Dehnungsschranke für f'.
Voraussetzung. f sei gleichmäßig differenzierbar mit Fehlerkonstante K auf allen Teilintervallen der Länge $2r$.
Behauptung. Die Ableitung f' ist dehnungsbeschränkt mit Dehnungsschranke $2K$, also

Dehnungsschranke für f': $\quad a, b \in [\alpha, \omega] \Rightarrow |f'(a) - f'(b)| \leq 2K \cdot |b - a|$.

Beweis: Zunächst gilt für jede Funktion h wegen der Dreiecksungleichung (Wiederholung: $|C - A| \leq |C - B| + |B - A|$), daß Dehnungsschranken auf $[a, b]$ und auf$[b, c]$ Dehnungsschranken auf $[a,c]$ zur Folge haben, etwa

$$\begin{aligned} a<b<c, \quad |h(b)-h(a)| &\le L\cdot|b-a|, \quad |h(c)-h(b)| \le L\cdot|c-b| \\ \Rightarrow \qquad |h(c)-h(a)| &\le L\cdot(|b-a|+|c-b|) = L\cdot|c-a|. \end{aligned}$$

Es genügt daher, derartige Lipschitzschranken auf kleinen Teilintervallen zu beweisen. Deshalb können wir in der Behauptung zusätzlich $|b-a|\le r$ voraussetzen.
Dann liefert die Voraussetzung zunächst:

$$|f(b)-f(a)-f'(a)\cdot(b-a)| \le K\cdot(b-a)^2,$$
$$|f(a)-f(b)-f'(b)\cdot(a-b)| \le K\cdot(a-b)^2,$$

danach folgt mit der Dreiecksungleichung die behauptete Dehnungsschranke für f':

$$|(f'(a)-f'(b))\cdot(b-a)| \le 2K\cdot(b-a)^2, \quad \text{oder mit } a\ne b: \; |f'(a)-f'(b)| \le 2K\cdot|a-b|.$$

Höhere Ableitungsinformationen, Taylorapproximation. In der vierten Folgerung aus dem Monotoniesatz hatten wir aus Schranken für die zweite Ableitung die quadratische Abweichung von der Tangente hergeleitet. Diese Idee kann iteriert werden. Wir werden immer höhere Potenzen als Fehlerterme bekommen. Für kleine Argumentdifferenzen sind höhere Potenzen $|x-a|^k$ viel kleiner als niedrigere Potenzen (Kap. 1, Beispiel 2). In Fällen, in denen die höheren Ableitungen nicht zu schnell sehr groß werden, können wir daher wesentlich bessere Approximationen als durch die Tangente erwarten. Als Beispiel behandeln wir die

Voraussetzung an die 4. Ableitung: $\qquad f:[-R-R]\to\mathbb{R},\ |f^{(4)}|\le M.$

Behauptung: $\big|f(x)-(f(0+f'(0)\cdot x+f''(0)\cdot x^2/2+f'''(0)\cdot x^3/6\ )\big| \le M\cdot x^4/24.$

Beweis. Die schon bewiesene 4. Folgerung besagt für f'' wegen $|(f'')''|\le M$:
$-M\cdot x^2/2 \ \le\ f''(x)-(f''(0)+f'''(0)\cdot x) \ \le\ M\cdot x^2/2 = \big(M\cdot x^3/6\big)'.$
Daraus folgt mit dem verallgemeinerten Monotoniesatz (1. Folgerung) für $x\ge 0$
$-M\cdot x^3/6 \ \le\ f'(x)-f'(0)-(f''(0)\cdot x+f'''(0)\cdot x^2/2) \ \le\ M\cdot x^3/6 = \big(M\cdot x^4/24\big)',$
und für $x\le 0$ sind die Ungleichungen andersherum. Eine weitere Anwendung des verallgemeinerten Monotoniesatzes gibt in beiden Fällen, $x\ge 0$ und $x\le 0$, die Behauptung.
Bemerkung. Es ist nicht schwer, diese Aussage auf höhere Ableitungen zu verallgemeinern. Auch wenn M viel größer als $B:=\max|f''|$ ist, so gibt es nach Kapitel 1 immer ein Intervall um 0, in dem der Fehler $M\cdot x^4/24$ viel kleiner als die Tangentenabweichung $B\cdot x^2/2$ ist. Diese Beobachtung führt in günstigen Fällen zu einer Folge immer besserer

Approximationen, genannt Taylor Approximationen. Wir kommen darauf zurück, wenn die Funktionen sin, cos definiert werden sollen.

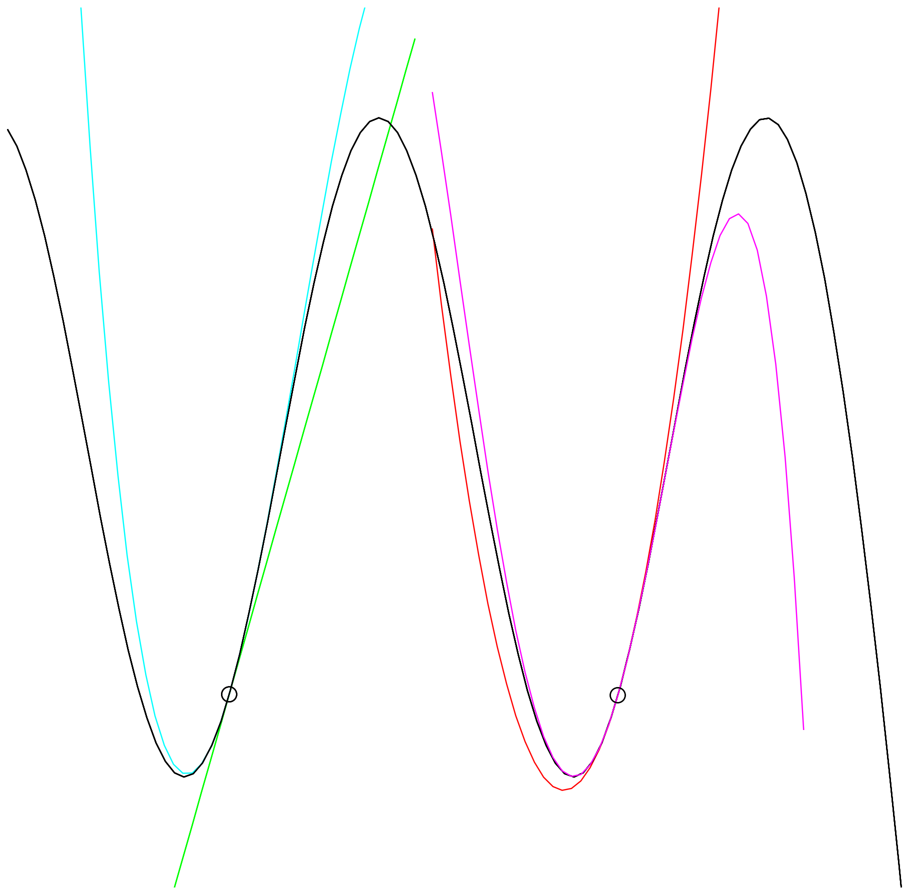


Illustration von Taylorapproximationen

Taylorpolynome ungerader Ordnung bleiben im allgemeinen (d.h. wenn die nächsthöhere Ableitung $\neq 0$ ist) oberhalb oder unterhalb der approximierten Funktion. Taylorpolynome gerader Ordnung wechseln im allgemeinen von einer Seite auf die andere. An der linken Stelle ist die Tangente und das 3. Taylorpolynom gezeigt, an der rechten Stelle das 2. und 4. Taylorpolynom. – Da an Wendestellen w gilt $f''(w) = 0$, stimmt das 2. Taylorpolynom mit der Tangente überein; daher werden die Graphen kubischer Funktionen von ihren Wendetangenten geschnitten.

Anfänge der Integralrechnung. Zu Newton's Zeit entwickelten sich Differentialrechnung ("Tangentenproblem") und Integralrechnung ("verallgemeinerte Summen") zunächst unabhängig von einander, ehe sie als Umkehrungen von einander ("Hauptsatz der Differential- und Integralrechnung") erkannt wurden. Die Vollständigkeit der reellen Zahlen wurde erst 200 Jahre später formuliert. Die nächste Anwendung des Monotoniesatzes reicht nahe an die Integralrechnung heran, so nahe, wie es ohne die Vollständigkeit geht.

Dazu erkläre ich zunächst sogenannte Riemann Summen. Zu deren Definition benötigen wir folgende Bezeichnungen. $[\alpha, \omega]$ liege im Definitionsbereich einer Funktion f und $[a, b] \subset [\alpha, \omega]$ sei ein Teilintervall mit einer sogenannten Einteilung $a = t_0 < t_1 < \ldots < t_n = b$. Als "Größe" einer Einteilung wird die Länge des längsten Teilintervalls angesehen, etwa $\delta := \max |t_j - t_{j-1}|$. Außerdem seien Zahlen $\tau_j \in [t_{j-1}, t_j],\ j = 1 \ldots n$ gewählt, sogenannte Zwischenstellen. Damit machen wir die

Definition: Jede der Summen $\sum_{j=1}^{n} f(\tau_j) \cdot (t_j - t_{j-1})$ heißt **Riemann Summe** von f zum Intervall $[a, b]$.

Als Vorbereitung unseres Satzes betrachte als Beispiel, für festes k, die Summe $S_N := \sum_{n=1}^{N} n^k$, also eine Riemann Summe der Funktion $x \to x^k$. Das hilft uns herauszufinden, wie schnell S_N als Funktion von N ungefähr wächst. Wir vergleichen Summanden mit Werten der Funktion x^k:

$x \in [n, n+1] \Rightarrow (x-1)^k \le n^k \le x^k$ (umgekehrte Ungleichungen falls $k < 0$).

Diese drei Funktionen sind Ableitungen, nämlich:

$x \in [n, n+1] \Rightarrow (\frac{1}{k+1}(x-1)^{k+1})' \le (n^k \cdot x)' \le (\frac{1}{k+1} \cdot x^{k+1})'$.

Daher folgt aus dem Monotoniesatz (1. Folgerung):

$\frac{1}{k+1}((n)^{k+1} - (n-1)^{k+1}) \le n^k \cdot (n+1-n) \le \frac{1}{k+1} \cdot ((n+1)^{k+1} - n^{k+1})$

(Bei Potenzen war der Monotoniesatz ja trivial, deshalb können diese Ungleichungen auch direkt verifiziert werden.) Schließlich summieren wir über n von 1 bis N:

$\frac{1}{k+1} N^{k+1} \le \sum_{n=1}^{N} n^k \le \frac{1}{k+1}((N+1)^{k+1} - 1)$.

Die linke und die rechte Schranke stimmen in der höchsten Potenz von N überein. Das ist also die Größenordnung, mit der die Summe wächst, $\sum_{n=1}^{N} n^k \sim \frac{1}{k+1} N^{k+1}$. Die übrigen Terme sind um so kleinere Prozentteile dieses Hauptterms, je größer N ist.

Zur Diskussion dieses Beispiels war wichtig, daß die Summanden Werte einer Funktion waren, die wir als Ableitung einer anderen Funktion schreiben konnten. Das ist z.B. bei allen Polynomen der Fall; auch alle Potenzfunktionen $x \mapsto x^k$ mit negativem Exponenten $k < -1$ sind uns als Ableitungen bekannt. Für diese Beziehung gibt es folgende

Definition. Eine Funktion F heißt *Stammfunktion* einer Funktion f, falls gilt $F' = f$.

Satz. Riemann Summen und gleichmäßig differenzierbare Stammfunktionen.
Riemann-Summen von Ableitungen f' können mit Hilfe der Stammfunktion f bis auf "kleine" Fehler berechnet werden. Genauer, es gilt die
Behauptung: $|f(b) - f(a) - \sum_{j=1}^{n} f'(\tau_j) \cdot (t_j - t_{j-1})| \le K \cdot |b-a| \cdot \delta$,

wobei K aus der gleichmäßigen Differenzierbarkeit von f stammt.
Zusatz: Gilt für $t \in [t_{j-1}, t_j]$ sogar: $f'(t) \leq f'(\tau_j)$, so liefert derselbe Beweis

$$f(b) - f(a) \leq \sum_{k=1}^{n} f'(\tau_j) \cdot (t_j - t_{j-1}),$$

entsprechend gilt die umgekehrte Ungleichung, falls $f'(t) \geq f'(\tau_j)$.

Beweis. In jedem Teilintervall gilt

$$\begin{aligned} &|(f(t_j) - f(t_{j-1})) - f'(\tau_j) \cdot (t_j - t_{j-1})| \\ = &|f(t_j) - f(\tau_j) + f(\tau_j) - f(t_{j-1}) - f'(\tau_j) \cdot (t_j - \tau_j + \tau_j - t_{j-1})| \leq \\ \leq &K \cdot (|t_j - \tau_j|^2 + |\tau_j - t_{j-1}|^2) \leq K \cdot \delta \cdot |t_j - t_{j-1}|. \end{aligned}$$

Summiert man diese Ungleichungen über $j = 1, \ldots, n$ und beachtet $\sum |t_j - t_{j-1}| = b - a$, so folgt die Behauptung.

Bemerkungen: Erstens wird man glauben, daß dieser Satz, wie im vorhergeschickten Beispiel, eine recht genaue Kontrolle über gewisse Summen ermöglicht. Zweitens ist wichtig, daß die Fehler $K \cdot |b - a| \cdot \delta$ um so kleiner sind, je kleiner die Länge δ des längsten Teilintervalls ist. Veranschaulicht man zum Beispiel jeden Summanden $f'(\tau_j) \cdot (t_j - t_{j-1})$ durch ein Rechteck der Höhe $f'(\tau_j)$ über dem Teilintervall $[t_{j-1}, t_j]$, so erkennt man, daß die Riemann-Summen den "Flächeninhalt" unter dem Graphen von f' über dem Intervall $[a, b]$ approximieren; außerdem sind die größte und die kleinste Riemann-Summe zu einer festen Einteilung höchstens um $K \cdot |b - a| \cdot \delta$ verschieden von der Differenz der Stammfunktionswerte $f(b) - f(a)$. Wir haben zwar noch keine Definition des Flächeninhalts unter dem Graphen von f' gegeben, aber unsere Fehlerabschätzung ist so gut, daß wir jetzt von jeder Definition des Flächeninhalts erwarten dürfen, daß aus dem Archimedes-Axiom folgt,

daß $f(b) - f(a)$ der Flächeninhalt unter dem Graphen von f' ist.

Eine andere Zahl als $f(b) - f(a)$ kommt jedenfalls als Flächeninhalt nicht mehr in Frage. Schließlich ist unsere Abschätzung des Unterschieds zwischen $f(b) - f(a)$ und Riemannsummen bereits eine Hälfte des Hauptsatzes der Differential- und Integralrechnung; die verbleibenden Schwierigkeiten kommen daher, daß es viele Funktionen wie $g(x) := 1/x$ gibt, von denen man nicht, wie bei Polynomen, von vonherein weiß, daß sie Ableitungen anderer Funktionen sind.
Damit sind wir an der Grenze dessen, was sich vor der Vollständigkeit der reellen Zahlen machen läßt.

Reelle Zahlen, Vollständigkeit

Die rationalen Zahlen reichen nicht aus, um Umkehrfunktionen, Existenz von Grenzfunktionen wie exp oder auch nur alle Punkte einer Strecke zu behandeln. Für eine axiomatische Beschreibung der reellen Zahlen braucht man die Begriffe *Nullfolge, konvergente Folge, Intervallschachtelung, Cauchyfolge.* In deren Definitionen greift das Archimedes Axiom ein. Das Vollständigkeitsaxiom erlaubt, den Satz von Bolzano-Weierstraß zu beweisen, Umkehrfunktionen zu definieren und neue Funktionen wie exp, sin, cos als Grenzwerte von Approximationen zu konstruieren.

Probleme, zu deren Behandlung die rationalen Zahlen nicht ausreichen.

Für alle bisher besprochenen Definitionen und Sätze genügte es, wenn man unter dem Wort "Zahlen" die rationalen Zahlen verstand. (Natürlich durfte man auch schon mehr wissen, die Aussagen und Sätze der ersten Abschnitte bleiben richtig, wenn das Wort "Zahlen" als reelle Zahlen gelesen wird). Ich muß nun erklären, in welcher Weise die rationalen Zahlen uns nicht genügen. Daß einzelne Zahlen wie $\sqrt{2}$ irrational sind, ist sicher von der Schule bekannt, aber wenn man seinen Blick zu sehr auf einzelne Zahlen lenkt, verkennt man die Größe des Problems. Ich erkläre zuerst, warum wir mehr Funktionen brauchen als die bisher betrachteten Quotienten von Polynomen.

Die Quadratwurzel ist ein spezielles Beispiel einer *Umkehrfunktion.* Es ist anschaulich sehr leicht, sich zu dem Graphen irgendeiner streng monoton wachsenden (z.B. rationalen) Funktion f den Graphen der Umkehrfunktion vorzustellen: Spiegelt man die Punkte $(x, f(x))$ des Graphen von f an der Winkelhalbierenden des 1. Quadranten, so erhält man den Graphen der Umkehrfunktion. Auf diese Weise erhält man aus dem Graph von $f(x) := x^2$ als Graph der Wurzelfunktion eine horizontale Parabel. Bei dieser Spiegelung gehen Geraden der Steigung m in Geraden der Steigung $1/m$ über, daher hat man sicher keine Zweifel, was die Ableitung der Umkehrfunktion g sein sollte:

$$\text{Ableitung der Umkehrfunktion:}\qquad f'(a) = m,\ g(f(x)) = x \ \Rightarrow\ g'(f(a)) = 1/m,$$

übrigens auch in Übereinstimmung mit der Kettenregel: $1 = (g \circ f)'(x) = g'(f(x)) \cdot f'(x)$.

Vor diesem einfachen anschaulichen Hintergrund ist es ärgerlich, daß wir das nicht mit Zahlen nachmachen können. Es würde ja nicht helfen, dieser ersten Umkehrfunktion zuliebe alle Quadratwurzeln rationaler Zahlen zu den rationalen Zahlen hinzu zu nehmen. Erstens würden die Wurzeln aus diesen neuen Zahlen immer noch fehlen, und zweitens, was ist mit all den anderen "einfachen" Umkehrfunktionen? Offenbar sind die rationalen Zahlen im Umgang mit Umkehrfunktionen sehr unzureichend.

Andere Wünsche kommen aus den Naturwissenschaften. Eine Funktion mit Wachstumsrate $f'/f = 1$ erweist sich an vielen Stellen als ebenso fundamental wie die geläufigeren Funktionen mit Steigung $f' = 1$. Funktionen mit $f'/f = 1$ wachsen in Intervallen der Länge a um den *konstanten Faktor* $f(a)/f(1)$, weil die Funktion $h(x) := f(x+a)/f(x)$ nach der Quotientenregel die Ableitung 0 also den Wert $h(x) = h(0) = f(a)/f(0)$ hat. (Funktionen mit $f' = 1$ wachsen in Intervallen der Länge a *additiv* um die Konstante $f(a) - f(0)$.) Derartige "Exponentialfunktionen" findet man aber unter den rationalen Funktionen nicht: Angenommen, wir hätten $f = P/Q = f'$, dann folgte mit der Quotientenregel für Polynome

$$\frac{P}{Q} = \left(\frac{P}{Q}\right)' = \frac{QP' - PQ'}{Q^2} \quad \Rightarrow \quad PQ = QP' - PQ',$$

aber das ist unmöglich, denn der Grad von $QP' - PQ'$ ist um 1 kleiner als der Grad von PQ. Aber nicht nur das, Exponentialfunktionen oder auch die zur Beschreibung von Schwingungen und Kreisbewegungen notwendigen trigonometrischen Funktionen $(\sin, \cos \ldots)$ kann man mit den rationalen Zahlen allein nicht einmal definieren. Die Exponentialfunktion z.B. hat außer bei 0 an keiner anderen rationalen Stelle einen rationalen Wert, so daß man auf dem Graphen nur diesen einen rationalen Punkt hat.

Wie ungeheuer viele weitere Zahlen wir wünschen, wird durch folgende Überlegung deutlich. Natürlich hätten wir gern, daß der Abstand jedes Punktes einer Strecke von deren Anfangspunkt durch eine *Zahl* angegeben werden kann. Davon sind wir mit den rationalen Zahlen allein *sehr weit* entfernt. Um einzusehen wie weit, ordnen wir die rationalen Zahlen zuerst so an, daß wir sie zählen können: einem (ungekürzten) Bruch m/n ordnen wir den Punkt (m, n) im ersten Quadranten zu. Diese Punkte zählen wir, indem wir die kurzen Diagonalen der Reihe nach zählen:

$$(1,1);\ (2,1),(1,2);\ (3,1),(2,2),(1,3);\ (4,1),(3,2),(2,3),(1,4);\ \ldots.$$

Nun müssen wir entweder die nicht vollständig gekürzten Brüche beim Zählen weglassen oder wir stören uns nicht dran, jede rationale Zahl immer wieder mal zu zählen. Nachdem wir so die rationalen Zahlen im Einheitsintervall *der Reihe nach aufzählen* können, kommt jetzt eine zweite Konstruktion. Wir wollen jede rationale Zahl mit etwas Platz umgeben, sie mit einem Intervall überdecken. Erstaunlicher Weise können wir die benötigte Gesamtlänge ℓ dieser Intervalle beliebig klein halten: Wir wählen eine Zahl $\ell < 1$, dann verwenden wir für die erste rationale Zahl bereits den halben Vorrat und überdecken sie mit einem Intervall der Länge $\ell/2$. Für die zweite rationale Zahl verwenden wir wieder die Hälfte des noch vorhandenen Vorrats und überdecken sie mit einem Intervall der Länge $\ell/2^2$. So fahren wir fort, und bei jedem Schritt ist das zuletzt verwendete Intervall so lang wie der noch verbleibende Vorrat. Über die n. rationale Zahl legen wir ein Intervall der Länge $\ell/2^n$ und so groß ist auch der verbleibende Vorrat. Jede rationale Zahl wird also irgendwann von einem Intervall zugedeckt, aber die Gesamtlänge der verwendeten Intervalle ist auf Grund unserer Konstruktion $\leq \ell$. Da wir ℓ so klein, wie wir wollen, wählen können, machen die rationalen Zahlen nur einen beliebig kleinen Anteil des Einheitsintervalles aus, alle übrigen Punkte müssen mit "neuen" Zahlen beschrieben werden.

Tatsächlich lösen die reellen Zahlen all diese Probleme: Rationale Funktionen f, die auf einem reellen Intervall $[\alpha, \omega]$ monoton wachsend sind, besitzen auf dem reellen Bildintervall $[f(\alpha), f(\omega)]$ definierte Umkehrfunktionen; exp, sin, cos können als reellwertige Funktionen auf $\mathbb{R}$ definiert werden; und jeder *Punkt* des Einheitsintervalls wird durch genau eine reelle *Zahl* beschrieben, den Abstand vom Anfangspunkt. Diese Erfolge haben allerdings ihren Preis. Wir können die einzelnen reellen Zahlen nicht ähnlich explizit angeben wie die rationalen Zahlen. Wir werden eine reelle Zahl als gegeben ansehen, wenn wir eine Folge als "konvergent" nachgewiesen haben; deren "Grenzwert" ist dann die reelle Zahl. Diesen Grenzwert kennen wir in der Regel nicht explizit, wir kennen nur die Elemente der Folge als Approximationen. Argumentieren mit reellen Zahlen wird daher Argumentieren mit Approximationen bedeuten.
Bezeichnung: Das Wort "Folge" wird als Name für eine Abbildung der natürlichen Zahlen $\mathbb{N}$ in eine Menge (aus Zahlen, aus Punkten, aus Funktionen ...) verwendet; wir schreiben Folgen als $\{a_n\}, \{P_n\}, \{f_n\}, \ldots$.

Die benötigten Begriffe.

Wir streben eine axiomatische Beschreibung der reellen Zahlen an. Dazu benötigen wir die vier eng zusammenhängenden Begriffe *Nullfolge, konvergente Folge, Intervallschachtelung* und *Cauchyfolge*. Alle vier sind von komplizierterer logischer Struktur, als in den bisherigen Abschnitten nötig war; sie können nicht mehr mit *expliziten* Ungleichungen definiert werden. Diese Schwierigkeit brauchen wir allerdings nur einmal zu überwinden, die Definitionen der drei letzten Begriffe können leicht mit Hilfe von Nullfolgen formuliert werden, den Begriff Nullfolge werde ich ausführlich diskutieren.

Zurückführende Definitionen.

Konvergente Folge: Eine Zahlenfolge $\{a_n\}$ heißt *konvergent* gegen a, wenn die Folge $\{a_n - a\}$ eine Nullfolge ist. Die Zahl a heißt *Grenzwert* der Folge $\{a_n\}$, $a = \lim_{n\to\infty} a_n$.
Es ist beabsichtigt, nicht-rationale Zahlen als Grenzwerte konvergenter Folgen rationaler Zahlen zu beschreiben.

Intervallschachtelung: Eine Folge von Intervallen, $\{[a_n, b_n]\}$ heißt *Intervallschachtelung*, falls erstens jedes folgende Intervall in dem vorhergehenden enthalten ist, $[a_{n+1}, b_{n+1}] \subset [a_n, b_n]$, und falls zweitens die Folge $\{b_n - a_n\}$ eine Nullfolge ist. Liegt eine Zahl c in

allen Intervallen einer Intervallschachtelung, $c \in [a_n, b_n]$, so heißt c *Grenzwert der Intervallschachtelung*.
Es ist beabsichtigt, die Intervalle $[a_n, b_n]$ als Approximationen des Grenzwertes anzusehen.
Cauchyfolge: Eine Zahlenfolge $\{a_n\}$ heißt *Cauchyfolge*, falls es eine Nullfolge $\{r_n\}$ gibt, so daß für alle $m > n$ gilt: $|a_m - a_n| \le r_n$.
Cauchyfolgen werden definiert, damit man über konvergente Folgen reden kann, *ohne* deren Grenzwert erwähnen zu müssen.
Diskussion des Begriffs Nullfolge. Da mit diesem Begriff etwas grundsätzlich Neues, etwas von allen vorhergehenden Begriffen sehr Verschiedenes, erfaßt werden soll, kann ich nur mit Beispielen erläutern, was die neue Definition leisten soll. Daher hilft es sicher, wenn man die folgenden Überlegungen bis zur Definition der Nullfolge mehr als einmal liest.
Beispiel. Die Folge $r_n := 1/n$ soll eine Nullfolge sein.
Vergleiche. Falls $\{r_n\}$ eine Nullfolge ist und $|a_n| \le r_n$, so soll auch $\{a_n\}$ eine Nullfolge sein. Falls c eine Konstante ist, so soll auch $\{c \cdot r_n\}$ eine Nullfolge sein.
Folgerungen. Wir zeigen, daß man mit den drei abgeleiteten Begriffen schon argumentieren kann, wenn man nur diese drei aufgezählten Eigenschaften von Nullfolgen verwendet:
(i) Falls c Grenzwert einer Intervallschachtelung $\{[a_n, b_n]\}$ ist, so ist c auch Grenzwert der Folgen $\{a_n\}$ und $\{b_n\}$. Beweis: Nach Voraussetzung ist $\{r_n := b_n - a_n\}$ eine Nullfolge. Es folgt $|c - a_n| \le r_n,\ |b_n - c| \le r_n$, d.h. $\lim a_n = c = \lim b_n$.
(ii) Es sei $\{[a_n, b_n]\}$ eine Intervallschachtelung und $\{c_n\}$ eine Folge mit $c_n \in [a_n, b_n]$, so ist $\{c_n\}$ eine Cauchyfolge. Insbesondere sind $\{a_n\}$ und $\{b_n\}$ Cauchyfolgen.
Beweis: Zunächst folgt aus $m \ge n$ nach Voraussetzung $c_m \in [a_m, b_m] \subset [a_n, b_n]$, also wegen $|c_m - c_n| \le |b_n - a_n| = r_n$ die Behauptung.
(iii) Falls $0 \le q < 1$ so ist $\{a_n := q^n\}$ eine Nullfolge. Beweis: Wegen $0 \le q < 1$ und der Summenformel der geometrischen Reihe haben wir einen Vergleich ("Majorisierung") mit einer schon bekannten Nullfolge:

$$n \cdot q^n \le \sum_{j=1}^{n} q^j = \frac{1 - q^{n+1}}{1 - q} \qquad \text{also} \qquad q^n \le \frac{1}{1-q} \cdot \frac{1}{n} = \frac{const}{n}.$$

(iv) Falls $0 \le q < 1$ so konvergiert die geometrische Reihe $a_n := \sum_{j=1}^{n} q^j$ gegen $1/(1-q)$.
Beweis: Die Differenz $a_n - 1/(1-q) = q^{n+1}/(1-q)$ ist nach (iii) eine Nullfolge.
(v) Falls $\{a_n\}$ gegen a konvergiert und f eine dehnungsbeschränkte Funktion ist (d.h. $|f(x) - f(y)| \le L \cdot |x - y|$), so konvergiert die Bildfolge $\{f(a_n)\}$ gegen $f(a)$. – Dies Ergebnis erlaubt uns, Tangentensteigungen als Grenzwerte von Sehnensteigungen anzusehen.
Beweis: Setze beide Definitionen ein, $|f(a_n) - f(a)| \le L \cdot |a_n - a| \le L \cdot r_n$, und verwende die Vergleichseigenschaften der Nullfolgen.

Propaganda. Diese Beispiele sind nicht etwa "geschickt" ausgesucht, es kommt wirklich nur darauf an, den Begriff Nullfolge zu verstehen.
Als nächstes ist für die Definition der Nullfolge eine Eigenschaft wichtig, die schon von Eudoxos formuliert und von Archimedes sehr wirkungsvoll benutzt worden ist:
Gegeben zwei von 0 verschiedene Längen, dann kann man die kürzere so oft abtragen, bis man die größere übertrifft.
Da wir die Längen von Strecken durch Zahlen beschreiben wollen, müssen diese Zahlen ebenfalls die Eigenschaft des Eudoxos haben. In unserer Literatur hat dies Axiom zwei Namen:

Axom des Archimedes, Axiom des Eudoxos. Zu jeder (auch noch so großen) Zahl R gibt es eine natürliche Zahl $n \in \mathbb{N}$, die größer ist als R, d.h. $R < n \cdot 1 = n$.
Dieses Axiom besagt in anderen Worten, daß zwischen der 0 und der Menge aller Stammbrüche, $\{1/n;\ n \in \mathbb{N}\}$ keine weitere Zahl liegt: Zu jeder (auch noch so kleinen) Zahl $r > 0$ gibt es $n \in \mathbb{N}$ mit $1/r < n$ oder $0 < 1/n < r$. Wir formulieren diese Eigenschaft, die in vielen Beweisen vorkommt, als Merkregel:

Archimedes Strategie. Um die Ungleichnung $a \leq b$ zu beweisen, gnügt es, die (unendlich vielen) schwächeren Ungleichungen $a \leq b + const/n$ für alle $n \in \mathbb{N}$ zu beweisen.
Die Archimedes Strategie ist deshalb so wichtig, weil es erstaunlich häufig vorkommt, daß die schwächeren Ungleichungen $a \leq (b + const/n)$ *viel* leichter zu beweisen sind, als die eigentlich gewünschte Ungleichung $a \leq b$.
Indirekter Beweis. Wäre $b < a$, so setze $r := const/(a - b) > 0$. Mit dem Archimedes Axiom finde eine natürliche Zahl n mit $r < n$. Daraus folgt $const/n < a - b$ und $b + const/n < a$, im Widerspruch zur Voraussetzung $a \leq b + const/n$ für alle $n \in \mathbb{N}$.
Ich werde weiter unten einen "angeordneten Körper" beschreiben, in dem das Archimedes Axiom nicht gilt, in dem es "unendlich kleine" Elemente zwischen 0 und allen Stammbrüchen $\{1/n;\ n \in \mathbb{N}\}$ gibt.

Jetzt könnten wir den ersten *Definitionsversuch* machen: Eine Folge $\{r_n\}$ heißt Nullfolge, falls es eine Konstante *const* gibt, so daß für alle $n \in \mathbb{N}$ gilt $|r_n| \leq const/n$. Die oben genannten Eigenschaften von Nullfolgen (und damit auch die Folgerungen (i)-(v)) wären damit richtige Aussagen. Tatsächlich formuliert dieser Definitionsversuch eine zu starke Bedingung für den Begriff Nullfolge, obwohl alle bisher aufgezählten Wünsche an diesen Begriff erfüllt sind. Was kann man noch wünschen? Zum Beispiel ist die monoton fallende Folge $\{1/\sqrt{n};\ n \in \mathbb{N}\}$ keine Nullfolge im Sinne des ersten Definitionsversuchs, obwohl aus dem Archimedes Axiom folgt, daß zwischen 0 und den Elementen dieser Folge keine weitere Zahl liegen kann. Unser letzter Wunsch wird durch dieses Beispiel motiviert: Eine monoton fallende Folge $\{r_n\}$ soll dann eine Nullfolge sein, wenn zwischen 0 und der Menge aller r_n keine weitere Zahl liegt. Und das drückt sich mit dem Archimedes Axiom endgültig so aus:

Definition des Begriffs Nullfolge. Eine Folge $\{r_n\}$ heißt *Nullfolge*, falls es zu jeder natürlichen Zahl $k \in \mathbb{N}$ einen Index n_k gibt mit der Eigenschaft:

$$n \geq n_k \quad \Rightarrow \quad |r_n| \leq \frac{1}{k}.$$

Variationen. In der Literatur findet man auch die anscheinend stärkere Definition: Eine Folge $\{r_n\}$ heißt *Nullfolge*, falls es zu jeder positiven Zahl $\epsilon > 0$ einen Index n_ϵ gibt mit der Eigenschaft:

$$n \geq n_\epsilon \quad \Rightarrow \quad |r_n| \leq \epsilon.$$

Äquivalenzbeweis. Eine Folge $\{r_n\}$ erfülle unsere schwächere Nullfolgeneigenschaft und ein $\epsilon > 0$ sei gegeben. Wie findet man n_ϵ? Mit dem Archimedes Axiom finde zunächst eine natürliche Zahl $k > 1/\epsilon$ und dazu ein n_k so daß $n \geq n_k \;\Rightarrow\; |r_n| \leq 1/k$. Wegen $1/k < \epsilon$ können wir $n_\epsilon := n_k$ wählen.

Die endgültige Nullfolgendefinition kann natürlich in die Definitionen der drei abgeleiteten Begriffe eingesetzt werden, um alternative Formulierungen zu bekommen, in denen das Wort Nullfolge (wie vielfach in der Literatur) nicht vorkommt:

Expandierte Definition der Konvergenz. Eine Folge $\{a_n\}$ heißt konvergent gegen a, falls es zu jedem $\epsilon > 0$ einen Index n_ϵ gibt, so daß gilt:

$$n \geq n_\epsilon \quad \Rightarrow \quad |a_n - a| \leq \epsilon.$$

Expandierte Definition des Begriffs Cauchyfolge. Eine Folge $\{a_n\}$ heißt Cauchyfolge, falls es zu jedem $\epsilon > 0$ einen Index n_ϵ gibt, so daß gilt:

$$m, n \geq n_\epsilon \quad \Rightarrow \quad |a_m - a_n| \leq \epsilon.$$

Diese Umformulierungen dienen dem besseren Vergleich mit der Literatur. Inhaltlich bedeuten sie dasselbe wie die explizit auf den Begriff Nullfolge zurückgeführten Definitionen. Der direkte Rückgriff auf die Nullfolgen führt vielleicht zu etwas zielstrebigeren Konvergenzbeweisen, weil ein Konvergenzbeweis natürlich beendet ist, sobald eine Majorisierung $|a_n - a| \leq r_n$ mit *irgendeiner* schon bekannten Nullfolge $\{r_n\}$ geglückt ist.

Ergänzung: Eine nicht Archimedesche Anordnung.

Definition. Ein Polynom heißt *koeffizienten-positiv*, in Formeln $P \overset{k}{>} 0$, falls der niedrigste von 0 verschiedene Koeffizient von P positiv ist. Entsprechend werden Polynome verglichen,

P heißt *koeffizientenweise größer als* Q, $P \overset{k}{>} Q$, genau dann wenn $P - Q \overset{k}{>} 0$.
Aufgabe. Wie bei reellen Zahlen gilt: Summe und Produkt koeffizienten-positiver Polynome sind koeffizienten-positiv; für $P \neq 0$ ist $P^2 \overset{k}{>} 0$. Anders als bei reellen Zahlen gilt zwar für $P(x) := x$, $P \overset{k}{>} 0$, aber es gibt kein $n \in \mathbb{N}$ mit $P \overset{k}{>} 1/n$.
Aufgabe. Diese Anordnung kann auf rationale Funktionen P/Q ausgedehnt werden. Nimm an, daß Faktoren x in P/Q gekürzt sind und definiere $P/Q \overset{k}{>} 0$ genau dann wenn entweder $P, Q \overset{k}{>} 0$ oder $-P, -Q \overset{k}{>} 0$. Definiere ähnlich wie für Polynome $P/Q \overset{k}{>} P1/Q1$. Auf diese Weise wird der Körper der rationalen Funktionen zu einem angeordneten, aber nicht Archimedesch angeordneten Körper.

Reelle Zahlen. In der Einführungsliteratur zur Analysis hat sich durchgesetzt, die reellen Zahlen axiomatisch zu beschreiben als einen angeordneten Körper, in dem das Archimedes Axiom und das Vollständigkeitsaxiom gilt. Daß es solch einen Körper wirklich gibt, wird nur noch in Spezialliteratur bewiesen. Nachdem ich die Rolle des Archimedes Axioms für die Definition des Konvergenzbegriffs erläutert habe, fehlt uns daher nur noch die Vollständigkeit, für die ich drei übliche Formulierungen besprechen werde. Hier die

Erste Formulierung des Vollständigkeitsaxioms:

> Zu jeder Intervallschachtelung $[a_k, b_k]_{k\in\mathbb{N}}$ gibt es eine (und offenbar dann genau eine) reelle Zahl r, die in allen Intervallen der Schachtelung liegt, $r \in [a_k, b_k]$ für alle $k \in \mathbb{N}$. r wird auch Grenzwert der Intervallschachtelung genannt.

Eindeutigkeitslemma. Falls $r_1 \leq r_2$ Grenzwerte einer Intervallschachtelung $[a_k, b_k]_{k\in\mathbb{N}}$ sind, so gilt $r_1 = r_2$.
Beweis. Nach Voraussetzung liegen r_1, r_2 in allen Intervallen $[a_k, b_k]$, also gilt $0 \leq r_2 - r_1 \leq b_k - a_k$ für alle k. Weil also $0 \leq r_2 - r_1$ unterhalb einer Nullfolge liegt, folgt aus dem Archimedes Axiom $r_2 - r_1 = 0$.

Wir müssen nun einsehen, daß wir durch das Vollständigkeitsaxiom wirklich genügend viele Zahlen bekommen. Dazu zeige ich erstens den Satz von Bolzano-Weierstraß, der besagt, daß wir wirklich so viele Zahlen wie Punkte auf einer Strecke haben. In der folgenden einfachen Formulierung wäre der Satz nicht beweisbar, wenn wir uns mit weniger als der endgültigen Nullfolgendefinition zufrieden gegeben hätten. Danach folgt die Existenz von Umkehrfunktionen. Schließlich konstruieren wir die wichtigen Funktionen exp, sin, cos.

Satz von Bolzano-Weierstraß
Jede monoton wachsende und beschränkte Folge $\{a_k\}$ ist konvergent.

Beweis. Wir konstruieren eine Intervallschachtelung $\{[a_k, b_k]\}$, deren linke Intervallenden die Punkte der gegebenen monotonen Folge $\{a_k\}$ sind, und deren rechte Intervallenden immer bessere obere Schranken sind. Die Schwierigkeit besteht darin, die b_k so zu wählen, daß $b_k - a_k$ eine Nullfolge ist. Da die Folge $\{a_k\}$ beschränkt ist, bezeichnen wir eine gegebene obere Schranke mit b_1; außerdem sei der erste "Erfolgsindex" $e_1 := 1$.
Nun wird das folgende Verfahren (für jedes m) wiederholt:
Angenommen für alle $j,\ j \le m$ seien die Erfolgsindices $e_j \ge j$ schon definiert und für $k \le e_m$ auch die oberen Schranken b_k. (Induktionsanfang: Dies ist für $m = 1$ der Fall.)
Betrachte den nächsten Index $k := e_m + 1$ und dazu den Mittelpunkt $c_k := \frac{1}{2}(a_k + b_{k-1})$.
Entweder c_k ist wieder obere Schranke der ganzen Folge $\{a_k\}_{k\in\mathbb{N}}$, dann setze: $b_k := c_k$, $e_{m+1} := k > e_m$.
Oder c_k ist *nicht* obere Schranke, dann gibt es einen Index $e_{m+1} > k$ mit $a_{e_{m+1}} > c_k$. Da wir mit c_k keine verbesserte obere Schranke gefunden haben, verwenden wir die letzte obere Schranke weiter und setzen für alle k mit $e_m < k \le e_{m+1}$: $b_k = b_{e_m}$.
Damit ist mindestens ein weiteres b_k definiert und $|a_{e_{m+1}} - b_{e_{m+1}}| \le \frac{1}{2}|a_{e_m} - b_{e_m}|$. Diese Ungleichung besagt, daß die Intervallängen $b_k - a_k$ der konstruierten Schachtelung wirklich eine Nullfolge bilden. Mit dem Vollständigkeitsaxiom finden wir einen Grenzwert $r \in [a_k, b_k]$ dieser konstruierten Intervallschachtelung und damit auch der gegebenen monoton wachsenden Folge $\{a_k\}, r = \lim_{k\to\infty} a_k$.
Offenbar gilt $a_k \le r$ für alle $k \in \mathbb{N}$, und es gibt keine kleinere obere Schranke der a_k. Auf diese Beobachtung kommen wir noch zurück.

Ähnlich beweisen wir jetzt den oben durch das Spiegeln des Graphen von f nahegelegten

Satz über die Surjektivität monotoner Funktionen. $f\colon [a, b] \to \mathbb{R}$ sei monoton und dehnungsbeschränkt. Dann ist f surjektiv auf das Bild $[A_1, B_1] := [f(a), f(b)]$. Falls f streng monoton ist, existiert eine auf dem reellen Intervall $[A_1, B_1]$ definierte Umkehrfunktion.

Beweis. Sei $C \in [A_1, B_1]$ beliebig und L Dehnungsschranke von f. Wir konstruieren eine Intervallschachtelung in $[a, b]$, deren Bildschachtelung gegen C konvergiert:
Setze $a_1 = a, b_1 = b$, also $f(a_1) = A_1, f(b_1) = B_1$.
Betrachte $D := f((a_1 + b_1)/2)$.
Ist $D \ge C$ so setze $a_2 = a_1, b_2 = (a_1 + b_1)/2$, andernfalls $a_2 = (a_1 + b_1)/2,\ b_2 = b_1$.
Dies Verfahren wird wiederholt:
Sei $[a_m, b_m]$ das zuletzt definierte Intervall, von dem wir nach Konstruktion wissen $C \in f([a_m, b_m])$. Betrachte $D := f((a_m + b_m)/2)$.

Ist $D \geq C$, so setze $a_{m+1} = a_m,\ b_{m+1} = (a_m + b_m)/2$,
andernfalls $a_{m+1} = (a_m + b_m)/2,\ b_{m+1} = b_m$.
Offenbar ist $\{[a_k, b_k]\}$ eine Intervallschachtelung mit $C \in f([a_k, b_k])$ für alle $k \in \mathbb{N}$. Weil f eine Dehnungsschranke hat, ist diese Bildschachtelung konvergent gegen C, Folgerung (vi) aus der Diskussion von Nullfogen. Wir zitieren nun das Vollständigkeitsaxiom für die *Intervallschachtelung im Definitionsbereich* und finden $c \in [a_k, b_k]$ für alle $k \in \mathbb{N}$. Dann ist $|f(c) - C| \leq L \cdot |b_k - a_k|$, also $f(c) = C$ mit dem Archimedes Axiom. – Wird zusätzlich strenge Monotonie vorausgesetzt, so ist c das einzige Urbild von C, also eine Umkehrfunktion definiert.
Beispiel: Für $f : [0, \infty] \to [0, \infty],\ f(x) := x^2,\ \ f'(x) = 2x$ ist also jetzt eine Umkehrfunktion, die Wurzelfunktion, definiert: $f^{-1} : [0, \infty] \to [0, \infty],\ f^{-1}(y) = \sqrt{y},\ f^{-1} \circ f(x) = x$.
Wegen der Kettenregel erwarten wir:

$$(f^{-1})'(f(x)) = \frac{1}{2x} \ \text{ oder } \ (f^{-1})'(y) = \frac{1}{2\sqrt{y}}.$$

Dies können wir ohne Differentiationsregel direkt aus der Differenzierbarkeitsdefinition folgern, indem wir eine quadratische Abschätzung beweisen:

$$\begin{aligned}
\sqrt{x} - \sqrt{a} - \frac{1}{2\sqrt{a}} \cdot (x - a) &= \Big(\frac{1}{\sqrt{x} + \sqrt{a}} - \frac{2}{\sqrt{a}}\Big) \cdot (x - a) \\
&= \frac{\sqrt{a} - \sqrt{x}}{2\sqrt{a}(\sqrt{x} + \sqrt{a})} \cdot (x - a) = \frac{-(x-a)^2}{2\sqrt{a} \cdot (\sqrt{x} + \sqrt{a})^2}, \qquad \text{also}
\end{aligned}$$

$$\frac{a}{2} \leq x \leq 2a \ \Rightarrow \ \frac{-1}{4\sqrt{a}^3} \cdot (x-a)^2 \leq \ \sqrt{x} - \Big(\sqrt{a} + \frac{1}{2\sqrt{a}} \cdot (x - a)\Big) \ \leq 0.$$

Bemerkung. Um zu betonen, daß zu jeder neuen Konstruktion von Funktionen eine Differentiationsregel gehört, beweise ich die früher schon plausibel gemachte Differentiationsregel für Umkehrfunktionen. Dieser Beweis trägt nichts zur Diskussion der Vollständigkeit bei und kann daher auch später gelesen werden.
Beweis der Differentiationsregel für Umkehrfunktionen.
Voraussetzung: $f : [a, b] \to [A, B]$ sei streng monoton; die nach dem vorgehenden Satz existierende Umkehrfunktion heiße $f^{-1} : [A, B] \to [a, b]$. Ferner sei f bei $c \in [a, b]$ differenzierbar, $f(c) = C,\ f(x) = X$ und $f'(c) > 0$.
Wir benutzen Differenzierbarkeit mit quadratischem Fehler, also die Existenz von Konstanten r, K so daß

$$|x - c| \leq r \Rightarrow |f(x) - f(c) - f'(c) \cdot (x - c)| \leq K \cdot (x - c)^2$$

Behauptung: $(f^{-1})'(C) = 1/f'(c).$

Beweis. Setze die Bezeichnungen in die Voraussetzung ein und dividiere durch $f'(c)$:

$$(*)\qquad |x-c|\le r \Rightarrow \left|\frac{1}{f'(c)}\cdot(X-C)-f^{-1}(X)+f^{-1}(C)\right|\le\frac{K}{f'(c)}\cdot(x-c)^2$$

Diese Ungleichung ist schon fast die, die wir brauchen; rechts muß nur noch $(x-c)^2$ durch $(X-C)^2$ abgeschätzt werden, außerdem müssen die Voraussetzungen an x in Voraussetzungen an X verwandelt werden. Wie im Beweis der Kettenregel (bitte vergleichen) folgt aus der Differenzierbarkeitsdefinition mit der Dreiecksungleichung ein beidseitiger Vergleich zwischen $|x-c|$ und $|f(x)-f(c)|$:

$$|x-c|\le r \;\Rightarrow\; (f'(c)-K\cdot r)\cdot|x-c|\le\; |f(x)-f(c)| \;\le(f'(c)+K\cdot r)\cdot|x-c|.$$

Damit die linke Ungleichung etwas nützt, verkleinern wir r zu $r_\_ := \min(r, f'(c)/2K)$ und vergröbern die letzte Ungleichung zu einer guten zweiseitigen Abschätzung:

$$|x-c|\le r_\_ \;\Rightarrow\; \frac{1}{2}f'(c)\cdot|x-c|\le|X-C|\le\frac{3}{2}f'(c)\cdot|x-c|.$$

Bisher gilt die Ungleichung $(*)$ unter der Voraussetzung $|x-c|\le r$ und die untere Abschätzung für $|X-C|$ unter der Voraussetzung $|x-c|\le r_\_$. Damit können wir jetzt eine Voraussetzung an $X-C$ formulieren, aus der $|x-c|\le r_\_$ folgt, nämlich $|X-C|\le R := \frac{1}{2}f'(c)\cdot r_\_$. Damit haben wir $(*)$ unter Voraussetzungen an die Argumente X von f^{-1}. Schließlich vergröbern wir $(*)$ durch Einsetzen von $|x-c|\le\frac{2}{f'(c)}\cdot|X-C|$ zu

$$|X-C|\le R \;\Rightarrow\; \left|f^{-1}(X)-f^{-1}(C)-\frac{1}{f'(c)}\cdot(X-C)\right|\le\frac{4K}{f'(c)^3}\cdot|X-C|^2.$$

Dies ist die gewünschte quadratische Tangentenapproximation, die $(f^{-1})'(C) = 1/f'(c)$ zeigt. – Der Beweis überträgt sich auf ϵ-δ-Fehler: Es ist leicht, $K\cdot(x-c)^2$ durch $\epsilon\cdot|x-c|$ zu ersetzten; mehr argumentative Schwierigkeiten macht, durch *Voraussetzungen* an $|X-C|$ zu garantieren, daß die Voraussetzungen für die Gültigkeit der benutzten Abschätzungen für f erfüllt sind. Wie beim Beweis der Kettenregel muß man nur auf dieses Detail zurückkommen, wenn man diese Sätze unter der schwächeren endgültigen Differenzierbarkeitsvoraussetzung beweisen will.

Konstruktion differenzierbarer Funktionen als Grenzfunktionen. Dies ist das dritte der Rechtfertigungsprobleme, an dem sich das Vollständigkeitsaxiom beweisen muß. Hier benötigen wir etwas längere Argumentationsketten, denn es kann ja nicht ganz einfach sein, die *Ableitung* von Funktionen zu bestimmen, deren *Werte* wir nicht berechnen, sondern nur approximieren können.

Erste Konstruktion der Exponentialfunktion. Wir verwenden die Zinseszins-Funktionen, für die wir als Anwendung des multiplikativen Monotoniesatzes folgende Ungleichungen bewiesen hatten:

$$0 \leq x < j \Rightarrow f_n := (1 + x/n)^n \leq 1/(1 - x/j)^j =: h_j(x).$$

Die Wachstumsraten dieser Funktionen waren:

$$\frac{f_n'}{f_n}(x) = \frac{1}{1 + x/n} \leq 1 \leq \frac{1}{1 - x/j} = \frac{h_j'}{h_j}(x).$$

Wir sehen, daß die Wachstumsraten der f_n monoton wachsend und die der h_j monoton fallend gegen 1 konvergieren. Wir können also hoffen, daß diese Zinseszins-Funktionen gegen die Exponentialfunktion konvergieren. Wegen der Nullstellen im Nenner betrachten wir jetzt ein festes Intervall $[0, 4]$ und nehmen nur die Funktionen mit $n, j \geq 8$. Dann liefern die Funktionswerte geschachtelte Intervalle:

$$x \in [0, 4],\ n \geq 8 \quad \Rightarrow \quad [1, h_8(4)] \supset [f_8(x), h_8(x)] \supset [f_n(x), h_n(x)] \supset [f_{n+1}(x), h_{n+1}(x)].$$

Tatsächlich haben wir sogar eine Intervallschachtelung, denn die Differenzen $h_n(x) - f_n(x)$ werden durch eine uns schon bekannte Nullfolge majorisiert. Setze $q := 1 - (x/n)^2 < 1$ und benutze $(1 - q^n) = (1 - q)(1 + q + \ldots + q^{n-1}) \leq (1 - q) \cdot n = x^2/n$:

$$0 \leq h_n(x) - f_n(x) = h_n(x) \cdot \Big(1 - \big(1 - (x/n)^2\big)^n\Big) \leq h_8(4) \cdot x^2/n \leq 256x^2/n\,.$$

Für jedes $x \in [0, 4]$ konvergiert daher die Intervallschachtelung $[f_n(x), h_n(x)]$ *wegen des Vollständigkeitsaxioms*. Wir nennen den Grenzwert $E(x)$. Wir erwarten, daß diese Grenzfunktion differenzierbar ist und daß $E'(x) = E(x)$ gilt. Der Beweis ist überraschend einfach, weil die Funktionen f_n die Voraussetzungen des Monotoniesatzes erfüllen und weil offenbar gilt:

$$x \in [0, 4],\ n \geq 8 \ \Rightarrow\ 0 \leq f_n''(x) \leq f_n'(x) \leq f_n(x) \leq h_8(4) = 256.$$

Der Monotoniesatz liefert zuerst die Lipschitzschranke 256 für alle f_n in $[0, 4]$, also

$$|f_n(x) - f_n(y)| \leq 256|x - y|.$$

Die Differenz zwischen Grenzfunktion und Approximation ist in unserer Konstruktion schon durch eine Nullfolge majorisiert:

$$|E(x) - f_n(x)| \leq 256x^2/n\,,$$

daher folgt mit der Archimedes Strategie *dieselbe* Lipschitzschranke für die Grenzfunktion:

$$x, y \in [0, 4] \Rightarrow |E(x) - E(y)| \leq 256|x - y|.$$

Diese Argumentation mit Abschätzungen, die für alle Approximationen in gleicher Weise gelten und die sich daher mit dem Archimedes Axiom sofort auf die Grenzfunktion ausdehnen lassen, liefern uns auch die Ableitungseigenschaften. Die Schranke für die zweite Ableitung und die Folgerung 4 (Tangentenabweichung) aus dem Monotoniesatz liefert:

$$x, y \in [0,4],\ n \geq 8\ \Rightarrow |f_n(y) - f_n(x) - f_n'(x)(y-x)| \leq 128(y-x)^2.$$

Wieder sind die Differenzen $E(y) - f_n(y),\ E(x) - f_n(x),\ E(x) - f_n'(x)$ schon durch Nullfolgen abgeschätzt, also liefert die Archimedes Strategie

$$x, y \in [0,4] \Rightarrow |E(y) - E(x) - E(x)(y-x)| \leq 128(y-x)^2.$$

Diese Abschätzung bedeutet: Die Funktion $y \mapsto E(y)$ wird von der linearen Funktion $y \mapsto E(x) - E(x)(y-x)$ so gut approximiert, daß diese die Tangente der Funktion E an der Stelle x ist, vergleiche die Differenzierbarkeitsdefinition. Daher gilt insbesondere $E'(x) = E(x)$.
Damit haben wir die erste Grenzfunktion konstruiert und ihre Ableitung bestimmt.

Natürlich müssen wir uns noch von der Einschränkung auf das Intervall $[0,4]$ befreien. Dazu definieren wir Funktionen $E_k(x) := E(x/k)^k : [0,4k] \to \mathbb{R}$. Es gilt $E_k(0) = E(0) = 1$ und aus der Kettenregel folgt $E_k' = E_k$. Damit ergibt die Quotientenregel $(E_k/E_j)' = 0$, d.h. je zwei dieser Funktionen stimmen auf dem gemeinsamen Teil ihrer Definitionsbereiche überein. Dieser Trick dehnt also den Definitionsbereich auf $[0,\infty)$ aus. Die Formeln gelten auch für $k = -1$, also ist mit $E(-x) := 1/E(x)$ der Definitionsbereich auf ganz $\mathbb{R}$ ausgedehnt. Ebenso folgt das Additionstheorem aus $h(x) := E(x)E(a)/E(a+x),\ h' = 0$.
Bezeichnung. Die so erhaltene Grenzfunktion heißt *Exponentialfunktion*,

$$\exp : \mathbb{R} \to \mathbb{R},\ \exp(0) = 1,\ \exp' = \exp, \quad \exp(a+x) = \exp(a) \cdot \exp(x).$$

Taylorapproximation von exp. Ich habe zur Konstruktion von exp die Zinseszins-Funktionen benutzt, weil sie obere und untere Approximationen liefern. Noch bekannter sind die Taylor Approximationen von exp. Das n-te Taylorpolynom von exp ist das Polynom T_n vom Grade n, dessen Wert und dessen Ableitungen bis zur Ordnung n bei 0 mit denen der Exponentialfunktion übereinstimmen. Wegen

$$1 = \exp(0) = \exp'(0) = \exp''(0) = \ldots = \exp^{(n)}(0)$$

kann man diese Taylorpolynome leicht hinschreiben:

Taylorpolynome von exp: $\quad T_n(x) := \sum_{k=0}^{n} \frac{x^k}{k!}, \quad T_n' = T_{n-1}.$

Für $x \in [0,\infty)$ ist die Folge $\{T_n(x)\}$ monoton wachsend. Wegen $T_n'(x) \le T_n(x)$ und dem multiplikativen Schrankensatz ist diese Folge durch $\exp(x)$ nach oben beschränkt, also nach dem Satz von Bolzano-Weierstraß konvergent. Andererseits gilt für die unteren Zinseszins-Approximationen $f_n^{(k)}(0) \le 1,\ k = 0 \dots n$. Daher sind alle Koeffizienten der Polynome f_n höchstens so groß wie die der Taylor Polynome T_n, d.h.

$$x \in [0,\infty) \ \Rightarrow\ f_n(x) \le T_n(x) \le \exp(x).$$

Daher konvergieren auch die $T_n(x)$ gegen $\exp(x)$. – Will man die Konstruktion mit den Taylorpolynomen beginnen, so muß man sich vor allem eine obere Schranke für die $T_n(x)$ verschaffen, vgl. Leibniz-Reihen.

Aufgabe. Für jedes $n \in \mathbb{N}$ ist $\{a_k := k^n/\exp(k)\}$ eine Nullfolge. (Wie groß muß x sein, damit $x^{n+1} \le \exp(x)$ ist? Es ist nicht nach einer guten Abschätzung gefragt. Vgl. S.56.)

Irrationalität von e. Um die Qualität dieser Taylorapproximation zu demonstrieren, zeige ich:

$$e\!:= \exp(1) \text{ ist irrational.}$$

Beweis. Für $1/e = \exp(-1) = \sum_{k=0}^{\infty}(-1)^k/k!$ können wir wegen der alternierenden Vorzeichen sofort eine Intervallschachtelung angeben:

$$a_n := \sum_{k=0}^{2n-1} \frac{(-1)^k}{k!},\ \ b_n := \sum_{k=0}^{2n} \frac{(-1)^k}{k!},\ \ n = 1, 2, \dots$$

$$a_{n+1} - a_n = \frac{1}{(2n)!} - \frac{1}{(2n+1)!} > 0,\ \ b_n - b_{n+1} = \frac{1}{(2n+1)!} - \frac{1}{(2n+2)!} > 0$$

$$0 < b_n - a_n = \frac{1}{(2n)!}.$$

Daraus folgt für jedes N:

$$0 \ne |1/e - \sum_{k=0}^{N} (-1)^k/k!\ | < 1/(N+1)!$$

Wäre nun $1/e = P/Q$ mit $P, Q \in \mathbb{N}$, so wähle in der letzten Abschätzung $N\!:= Q$ und bringe alles auf den Hauptnenner $Q!$:

$$0 \ne \frac{|P \cdot (Q-1)! - \text{ GanzeZahl}\ |}{Q!} < \frac{1}{(Q+1)!}$$

Da ein Bruch $\ne 0$ mit Nenner $Q!$ mindestens die Größe $1/Q!$ hat, ist die letzte Ungleichung ein Widerspruch, der die Annahme $1/e = P/Q$ widerlegt. (Die "Ganze Zahl" ist $\sum_{k=0}^{Q}(-1)^k \cdot Q!/k!$.)

Leibniz-Reihen. Die eben benutzte bequeme Intervallschachtelung für $1/e$ ist Spezialfall einer Situation, die häufig genug ist, um einen Namen zu haben, wir beschreiben die Intervallschachtelung der *Leibniz-Reihen:*
Voraussetzung: $\{r_k\}$ sei eine manonton fallende Nullfolge.
Behauptung: $a_n := \sum_{k=0}^{2n-1}(-1)^k r_k, \quad b_n := \sum_{k=0}^{2n}(-1)^k r_k$ definiert eine Intervallschachtelung.
Beweis:

$$a_{n+1} - a_n = r_{2n} - r_{2n+1} \geq 0,\ b_n - b_{n+1} = r_{2n+1} - r_{2n+2} \geq 0,\ b_n - a_n = r_{2n}.$$

Beispiele. Erstens kann man dies auf einem negativen Intervall $[-R, 0]$ auf die Taylorpolynome T_n von exp anwenden, wenn man sich auf Indices $n \geq R$ beschränkt. Zweitens,

Konstruktion der trigonometrischen Funktionen sin und cos. Wir wollen versuchen, Funktionen $f, g\colon \mathbb{R} \to \mathbb{R}$ mit $f^2(t) + g^2(t) = 1, f'^2(t) + g'^2(t) = \text{const.}$ zu konstruieren, weil dann die Kurve $t \to (f(t), g(t))$ den Einheitskreis mit konstanter Geschwindigkeit parametrisiert. Wegen der Kettenregel können wir const $= 1$ annehmen (ersetze $f(t)$ durch $f(t/\sqrt{\text{const.}}))$.
Angenommen, wir könnten mehrfach differenzierbare Funktionen f, g mit diesen Eigenschaften finden, dann folgte durch Differenzieren von

$f^2 + g^2 = 1,\ f'^2 + g'^2 = 1$:

$2 \cdot f \cdot f' + 2 \cdot g \cdot g' = 0$ und $2f' \cdot f'' + 2g' \cdot g'' = 0$.

Die beiden Gleichungen lesen wir als Skalarprodukte: Der Geschwindigkeitsvektor (f', g') steht wie erwartet senkrecht auf dem Ortsvektor (f, g), und der Beschleunigungsvektor (f'', g'') ist senkrecht zu (f', g') also proportional zu (f, g). Der Proportionalitätsfaktor ergibt sich durch Weiterdifferenzieren von $f \cdot f' + g \cdot g' = 0$ aus $f \cdot f'' + g \cdot g'' + f'^2 + g'^2 = 0$ als (-1). Also gilt notwendig: $(f'', g'') + (f, g) = 0$.
Wir verabreden noch, daß unsere Kreisparameterisierung beginnen soll in $(f(0), g(0)) = (1, 0)$, und zwar mit der Geschwindigkeit $(f'(0), g'(0)) = (0, 1)$.
Nach diesen Vorüberlegungen können wir alle Ableitungen bei $t = 0$ angeben, die die Funktionen f und g haben müßten (wenn es sie gäbe, und wenn sie die benutzten Differenzierbarkeitseigenschaften hätten).

$$f(0) = 1, f'(0) = 0, f''(0) = -1,\ f'''(0) = 0, f^{(2k)}(0) = (-1)^k, f^{(2k+1)}(0) = 0;$$
$$g(0) = 0, g'(0) = 1, g''(0) = 0,\ g'''(0) = -1,\ g^{(2k)}(0) = (0), g^{(2k+1)}(0) = (-1)^k.$$

Natürlich ist es jetzt leicht, Polynomfolgen anzugeben, deren Ableitungen in 0 zu immer höherer Ordnung mit diesen hypothetischen Werten übereinstimmen:

$$P_n(t) = 1 - \frac{t^2}{2} + \frac{t^4}{4!} - + \ldots = \sum_{k=0}^{n} \frac{(-1)^k}{(2k)!} \cdot t^{2k},$$
$$Q_n(t) = t - \frac{t^3}{6} + \frac{t^5}{5!} - + \ldots = \sum_{k=0}^{n} \frac{(-1)^k}{(2k+1)!} \cdot t^{2k+1}.$$

Sie erfüllen:

$$Q_n' = P_n, \quad P_n' = -Q_{n-1}, \quad P_n'' = -P_{n-1}, \quad Q_n'' = -Q_{n-1}.$$

Ich betrachte diese Polynome genauer:

$$P_{n+1}(t) = P_n(t) - \frac{(-1)^n}{(2n+2)!} \cdot t^{2n+2}, Q_{n+1}(t) = Q_n(t) - \frac{(-1)^n}{(2n+3)!} \cdot t^{2n+3}.$$

Sie bieten sich an für eine Anwendung des Leibnizkriteriums: Wähle $R > 0$ fest und betrachte nur die Polynome mit $n \geq R$.
Dann gilt: Für $n \geq R$ und $|t| \leq 2R$ sind $\{|t|^{2n}/(2n)!\}_{n\in\mathbb{N}}$ und $\{|t|^{2n+1}/(2n+1)!\}_{n\in\mathbb{N}}$ monotone Nullfolgen, weil beim Erhöhen von n im Zähler ein kleinerer Faktor als im Nenner hinzukommt.
Für $2m \geq R$ und $|t| \leq 2R$ haben wir daher die Leibniz'schen Intervallschachtelungen für die gesuchten Grenzfunktion f und g

$$\{[P_{2m+1}(t), P_{2m}(t)]\}, \quad \{[Q_{2m+1}(t), Q_{2m}(t)]\}.$$

Wegen des Vollständigkeitsaxioms *existieren* damit f und g als Grenzfunktionen dieser Intervallschachtelungen.
Jetzt benötigen wir noch die Differenzierbarkeitseigenschaften dieser Grenzfunktionen. Das ist nun besonders einfach, weil die Leibniz'sche Intervallschachtelung uns bereits Schranken unabhängig von n liefert. Die Archimedes Strategie macht daraus dieselben Schranken für die Grenzfunktionen:

$$2m \geq 2, \quad 0 \leq t \leq 4 \quad \Rightarrow$$
$$-7 \leq P_1(t) \leq P_{2m+1}(t) \leq \ldots f(t) \ldots \leq P_{2m}(t) \leq P_2(t) \leq 4$$
$$-7 \leq Q_1(t) \leq Q_{2m+1}(t) \leq \ldots g(t) \ldots \leq Q_{2m}(t) \leq Q_2(t) \leq 4$$

denn $0 \leq t \leq 4 \quad \Rightarrow$

$$P_2(t) = 1 - \frac{t^2}{2}(1 - \frac{t^2}{12}) \leq 4 \qquad Q_2(t) = t \cdot (1 - \frac{t^2}{6} \cdot (1 - \frac{t^2}{20})) \leq 4$$
$$-7 \leq 1 - \frac{t^2}{2} = P_1(t) \qquad -7 \leq t - \frac{t^3}{3!} = Q_1(t).$$

Aus diesen Schranken für die Polynome folgt wegen $Q_n' = P_n, P_n' = -Q_{n-1}$ und aus dem Schrankensatz

$$n \geq 4, \quad 0 \leq s,t \leq 4 \;\Rightarrow\; |P_n(s) - P_n(t)| \leq 7 \cdot |s-t|, \quad |Q_n(s) - Q_n(t)| \leq 7 \cdot |s-t|.$$

Dann liefert die Dreiecksungleichung

$$|f(s) - f(t)| \leq 7 \cdot |s-t| + \text{ Nullfolge }, \quad |g(s) - g(t)| \leq 7 \cdot |s-t| + \text{ Nullfolge.}$$

und das Archimedes-Argument beseitigt die Nullfolgen. Damit haben wir die (noch verbesserungsfähige) Dehnungsschranke 7 für die Grenzfunktionen im Intervall $[0,4]$.
Auf dieselbe Weise haben wir von n unabhängige Schranken für die zweiten Ableitungen $P_n'' = -P_{n-1}, Q_n'' = -Q_{n-1}$ und damit auch von n unabhängige Schranken für die Abweichung von der Tangente. Wieder mit dem Archimedes Argument bekommen wir daraus dieselben Ungleichungen für die beiden Grenzfunktionen:

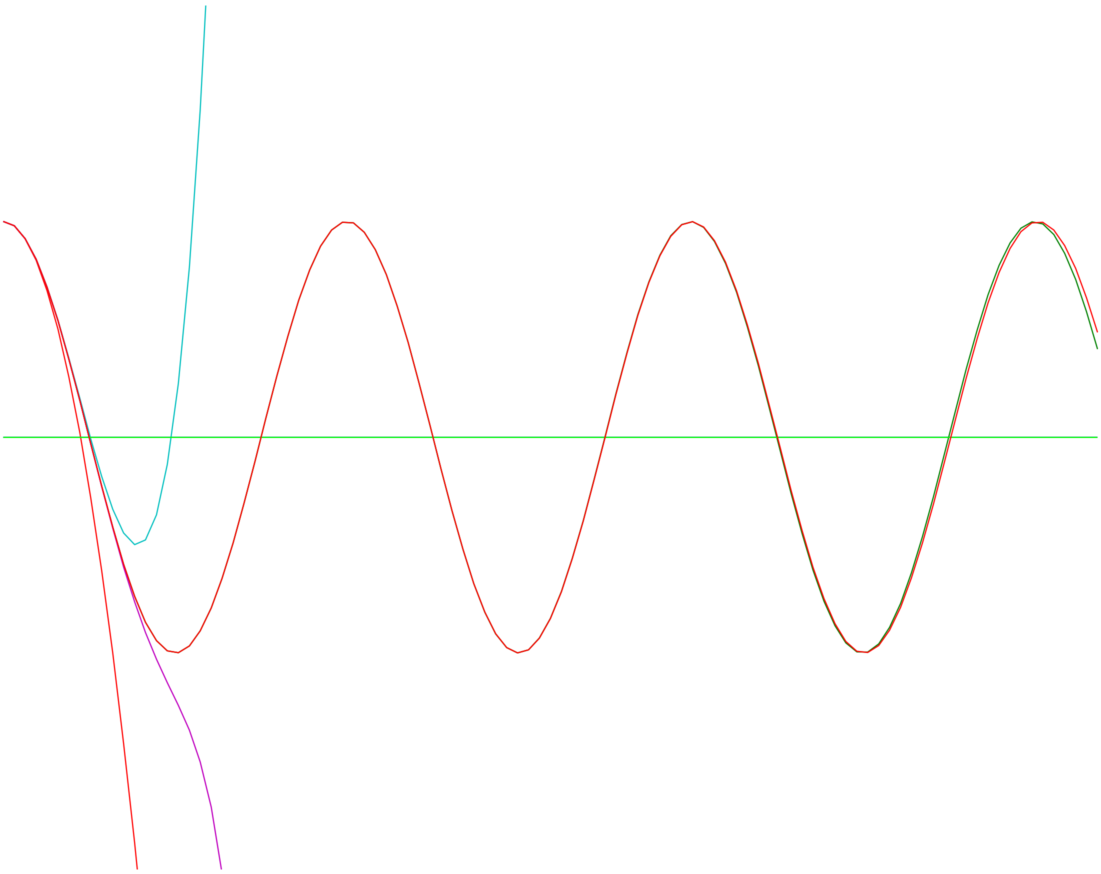
$$\begin{aligned}
&n \geq 4, \quad 0 \leq s,t \leq 4 \Rightarrow \\
&|P_n(s) - P_n(t) + Q_{n-1}(t)(s-t)| \leq 4 \cdot |s-t|^2 \Rightarrow |f(s) - f(t) + g(t)(s-t)| \leq 4 \cdot |s-t|^2, \\
&|Q_n(s) - Q_n(t) - P_n(t) \cdot (s-t)| \leq 4 \cdot |s-t|^2 \Rightarrow |g(s) - g(t) - f(t)(s-t)| \leq 4 \cdot |s-t|^2.
\end{aligned}$$

Die letzten Ungleichungen besagen, daß die Grenzfunktionen differenzierbar sind (genauer sogar gleichmäßig differenzierbar mit quadratischer Tangentenapproximation).

Es gilt: $\quad f' = -g, \quad g' = f, \quad f(0) = 1, \ g(0) = 0.$

Diese Gleichungen zeigen, daß f' und g' wieder differenzierbar sind, $f'' = -f, g'' = -g$, usw.: Die Grenzfunktionen sind unendlich oft differenzierbar. (Wenn man will, kann man jetzt bessere Fehlerkonstanten nachschieben: Aus $f'^2 + g'^2 = 1$ und dem Schrankensatz folgt, daß 1, statt 7, Dehnungsschranke für f und g ist. Ebenso hat man $|f''| \leq 1, |g''| \leq 1$, daher ist die Tangentenabweichung $\leq 0.5|s-t|^2$.)
Bezeichnungen: $f = \cos, \ g = \sin, \ \sin' = \cos, \ \cos' = -\sin, \ \ \tan := \frac{\sin}{\cos}, \ \ \tan' = 1 + \tan^2$.
Die erste positive Nullstelle von cos heißt $\pi/2$; wegen $P_2(t) \leq P_6(t) \leq \cos(t) \leq P_4(t)$ und $P_4(\sqrt{3}) = -1/8, P_6(1.5) \geq 0.07$ gilt: $\quad \sqrt{2} \leq 1.5 + 0.07 \leq \pi/2 \leq \sqrt{3} - 1/8 \sim 1.607$.

Additionstheoreme. Für Funktionen h mit $h'' = -h$ hat die Funktion $K(t) := h(t)^2 + h'(t)^2$ die Ableitung $K'(t) = 2h'(t)(h(t) + h''(t)) = 0$, sie ist also konstant. Daher haben zwei Funktionen $h1, h2$ mit gleichen Anfangswerten $h1(a) = h2(a), \ h1'(a) = h2'(a)$ eine Differenz h mit Anfangswerten $h(a) = 0 = h'(a)$, also $0 = K(a) = K(t)$, also $h = 0$. Daher stimmen die Funktionen $h1(t) := \sin(a+t)$ und $h2(t) := \sin(a)\cos(t) + \cos(a)\sin(t)$ überein, und ihre Ableitungen $h1'(t) := \cos(a+t) = h2'(t) := -\sin(a)\sin(t) + \cos(a)\cos(t)$ ebenfalls. – Damit können f, g von dem Intervall $[0,4]$ auf ganz $\mathbb{R}$ fortgesetzt werden.



Approximationen von $\cos$**.**

Die Figur zeigt den Graph von cos und die Graphen der 2., 4. und 6. Taylorpolynome T_n. Außerdem wird eine sehr viel bessere Approximation gezeigt:
Aus $\cos(2t) = \cos^2(t) - \sin^2(t) = 2\cos^2(t) - 1$ folgt $\cos(4t) = 8\cos^4(t) - 8\cos^2(t) + 1$. Daher kann man $\cos(t)$ aus $\cos(t/4)$ mit Hilfe des Polynoms $P4(x) := 8x^2(x^2-1)+1$ berechnen: $\cos(4t) = P4(\cos(t/4))$. Da nun die Taylorpolynome für *kleine Argumente viel bessere* Approximationen sind, kann man die Taylorapproximation mit dieser Konsequenz der Additionstheoreme kombinieren. Gezeigt ist $\cos(t) \sim P4 \circ P4 \circ T_4(t/16)$, und die über mehrere Perioden erstreckte Rechnung soll vermitteln, wie gut die Approximation in dem eigentlich nur benötigten Intervall $[0, \pi/2]$ ist.

Umkehrfunktionen von $\exp, \tan, \sin$**.** Weil $\exp : \mathbb{R} \to (0, \infty)$ streng monoton ist und auf jedem endlichen Intervall eine beschränkte Ableitung hat, haben wir eine differenzierbare Umkehrfunktion, genannt *Logarithmusfunktion*, $\log\colon (0,\infty) \to \mathbb{R}$ mit $\log(\exp(x)) = x$ und mit der Kettenregel:

$$\log'(\exp(x)) = \frac{1}{\exp'(x)} \quad \text{oder mit} \quad y := \exp(x)\colon\ \log'(y) = \frac{1}{y}.$$

Wegen $\exp(x)^n = \exp(nx)$ haben wir $n \log(\sqrt[n]{y}) = \log(y)$. Dies wird kombiniert mit einer einfachen Anwendung des Monotoniesatzes:

$$y \geq 1 \;\Rightarrow\; 0 < \log'(y) \leq 1 \;\Rightarrow\; \log(y) \leq 1 \cdot (y-1) \;\Rightarrow\; \log(y) \leq n \cdot (\sqrt[n]{y} - 1).$$

Diese Ungleichung zeigt, daß der Logarithmus eine außerordentlich langsam wachsende Funktion ist. Die früher gestellte Frage "Wie groß muß x sein damit $x^m \leq \exp(x)$ ist?" besitzt jetzt eine sehr einfache Antwort: Das ist genau dann der Fall, wenn $m \cdot \log(x) \leq x$ ist, also sicher dann, wenn $m \cdot \log(x) \leq m \cdot 2(\sqrt{x} - 1) \leq 2m\sqrt{x} \leq x$ oder $4m^2 \leq x$ ist.
Die Exponentialfunktionen zu beliebiger Basis $a > 0$ werden mit Hilfe des Logarithmus definiert:

$$\exp_a(x) := a^x := \exp(x \log a), \quad \text{also} \quad \exp_a' = \log(a) \exp_a = \exp_a'(0) \exp_a. \qquad \text{Daher}$$

$$a^{x+h} = a^x \cdot a^h \quad \text{und} \quad \frac{\exp_a(x+h) - \exp_a(x)}{h} = \frac{\exp_a(h) - 1}{h} \cdot \exp_a(x) \geq \exp_a'(0) \exp_a(x).$$

Aus dem Additionstheorem folgt also, daß man die Grenzwerte der Differenzenquotienten der Exponentialfunktionen überall kennt, sobald man den Grenzwert bei $x = 0$ kennt.

Auch $\tan : (-\pi/2, +\pi/2) \to \mathbb{R}$ ist eine streng monotone differenzierbare Funktion mit beschränkter Dehnungsschranke auf jedem abgeschlossenen Teilintervall $[-a, b]$. Daher haben wir eine differenzierbare Umkehrfunktion, genannt *Arcustangensfunktion*, $\arctan : \mathbb{R} \to (-\pi/2, +\pi/2)$ mit $\arctan(\tan(x)) = x$ und mit der Kettenregel

$$\arctan'(\tan(x)) = \frac{1}{\tan'(x)} \quad \text{oder mit} \quad y := \tan(x): \; \arctan'(y) = \frac{1}{1+y^2}.$$

Wir heben hervor, daß die Differentialgleichungen $\exp' = \exp$, $\tan' = 1 + \tan^2$ zur Folge haben, daß die Ableitungen der Umkehrfunktionen *rationale* Funktionen sind. Insbesondere sind daher die weiteren Ableitungen der Umkehrfunktionen leicht zu berechnen.

Schließlich ist auch $\sin : (-\pi/2, +\pi/2) \to (-1, +1)$ eine streng monotone differenzierbare Funktion mit beschränkter Dehnungsschranke auf jedem abgeschlossenen Teilintervall $[-a, b]$. Daher haben wir eine differenzierbare Umkehrfunktion, genannt *Arcussinusfunktion*, $\arcsin\colon (-1, +1) \to (-\pi/2, +\pi/2)$ mit $\arcsin(\sin(x)) = x$ und mit der Kettenregel

$$\arcsin'(\sin(x)) = \frac{1}{\sin'(x)} \quad \text{oder mit} \quad y := \sin(x): \; \arcsin'(y) = \frac{1}{\sqrt{1-y^2}}.$$

Wie sin ist auch arcsin eine *ungerade Funktion*, so daß alle geraden Ableitungen bei 0 verschwinden. $\arcsin''(y) = y(1-y^2)^{-3/2}$, $\arcsin'''(0) = 1$. Das dritte Taylorpolynom von

arcsin bei 0 ist daher $T_3(y) = y + y^3/6$. Wir illustrieren an einem Iterationsverfahren zur Berechnung von $\pi/2$, wie gut diese Taylorpolynome auch diese neuen Funktionen approximieren. Angenommen, a_1 sei ein Näherungswert für $\pi/2$. Wir haben oben gesehen, daß wir $b_1 := \cos(a_1)$ sehr genau ausrechnen können. Wegen $\cos(\pi/2 - x) = \sin(x)$ ist $\pi/2 = a_1 + \arcsin(b_1)$. Wir setzen daher $a_2 := a_1 + (b_1 + b_1^3/6)$, $b_2 := \cos(a_2)$, usw.
Beispiel:
$a_1 = 1, b_1 = 0.5403, \quad a_2 = 1.5666, b_2 = 0.0042059, \quad a3 = 1.5708, b_3 = 9.881 \cdot 10^{-14}$.

Andere Vollständigkeitsformulierungen. Es kommt oft vor, daß man eine einzelne Folge hat, deren Konvergenz man vermutet, die man aber nicht ohne weiteres mit einer Intervallschachtelung in Verbindung bringen kann (z.B. die Folge $T_n(x)$ der Taylorpolynome der Exponentialfunktion). Man müßte also einer Folge ansehen können, ob sie konvergiert, ohne ihren Grenzwert zu kennen. eine notwendige Bedingung liefert die Dreiecksungleichung, denn für eine gegen a konvergente Folge $\{a_k\}$ gilt:
Zu jedem $n \in \mathbb{N}$ gibt es einen Index k_n so daß gilt:

$$k, l \geq k_n \Rightarrow |a_k - a| \leq \frac{1}{n}, \quad |a_\ell - a| \leq \frac{1}{n}, \quad \text{also:} \quad |a_k - a_\ell| \leq \frac{2}{n}.$$

In der letzten Ungleichung kommt der Grenzwert a nicht mehr vor! Das veranlaßt die schon früher formulierte Definition der Cauchyfolge, die ich wiederhole:
Eine Folge $\{a_k\}$ heißt *Cauchyfolge* wenn gilt: Zu jedem $n \in \mathbb{N}$ gibt es ein k_n so daß

$$k, \ell \geq k_n \Rightarrow |a_k - a_\ell| \leq \frac{1}{n}.$$

Mit diesem Begriff haben wir die

Zweite Formulierung des Vollständigkeitsaxioms

Jede reelle Cauchyfolge besitzt eine reelle Zahl r als Grenzwert.

Diese zweite Formulierung der Vollständigkeit impliziert die erste: Gegeben eine Intervallschachtelung $\{[a_k, b_k]\}$. Dann gilt für $k \leq \ell :\ a_k \leq a_\ell \leq b_k$, also ist $|a_\ell - a_k| \leq |b_k - a_k|$ und $\{|b_k - a_k|\}$ ist nach Voraussetzung eine Nullfolge. Daher ist $\{a_k\}$, und ebenso $\{b_k\}$, Cauchyfolge. Nach der zweiten Formulierung der Vollständigkeit existiert ein (gemeinsamer) Grenzwert r. Dieser liegt in allen Intervallen der Schachtelung. Denn gäbe es auch nur ein Intervall $[a_k, b_k]$ das r nicht enthält, so folgte für alle $\ell \geq k$ erstens $r \notin [a_\ell, b_\ell]$ und dann sogar, daß r von a_ℓ (und b_ℓ) einen Abstand $\geq d := \min(|r - a_k|, |r - b_k|) > 0$ hat, im Widerspruch zur Konvergenz von $\{a_\ell\}$ gegen r. Aus diesem Widerspruch folgt, daß r tatsächlich in allen Intervallen der gegebenen Intervallschachtelung liegt, so daß diese im

Sinne der ersten Formulierung der Vollständigkeit konvergiert.
Für die umgekehrte Richtung definieren wir zu einer gegebenen Cauchyfolge $\{c_k\}$ eine Intervallschachtelung $\{[a_m, b_m]\}$, so daß jedes Intervall unendlich viele der c_k enthält. Das geschieht in mehreren Schritten:

1.) Die ganze Cauchyfolge $\{c_k\}$ liegt in einem Intervall $[a_1, b_1]$.
Beweis. Die Definition liefert: Zu $n = 2$ gibt es k_2 so daß $k, \ell \geq k_2 \Rightarrow |c_k - c_\ell| \leq \frac{1}{2}$.
Daraus folgt für alle ℓ: $a_1 := \min\{c_k; k \leq k_2\} - \frac{1}{2} \leq c_\ell \leq \max\{c_k; k \leq k_2\} + \frac{1}{2} =: b_1$.
2.) Betrachte $d := \frac{1}{2}(a_1 + b_1)$. Entweder liegen in $[a_1, d]$ unendlich viele Elemente der Cauchyfolge, dann setze $a_2 = a_1, b_2 = d$. Oder das ist nicht der Fall, dann setze $a_2 = d, b_2 = b_1$. Dies Verfahren wird wiederholt.
3.) Es sei $[a_m, b_m]$ das zuletzt gewählte Intervall, das unendlich viele Elemente der Cauchyfolge enthält. Betrachte $d = \frac{1}{2}(a_m + b_m)$. Falls $[a_m, d]$ unendlich viele Elemente der Cauchyfolge enthält, setze $a_{m+1} = a_m, b_{m+1} = d$. Falls das nicht der Fall ist, setze $a_{m+1} = d, b_{m+1} = b_m$, usw.
4.) Da jedes Intervall eine Hälfte des vorhergehenden ist, ist $\{[a_m, b_m]\}$ eine Intervallschachtelung, hat also nach der vorausgesetzten ersten Formulierung der Vollständigkeit einen Grenzwert r.
5.) Um zu zeigen, daß $\{c_k\}$ gegen r konvergiert, müssen wir zeigen, daß $\{c_k - r\}$ Nullfolge ist. Es sei also $\epsilon > 0$ gegeben. Wähle erstens m so daß $|b_m - a_m| \leq \epsilon/2$. Wähle zweitens nach Definition der Cauchyfolge k_ϵ so daß $k, \ell \geq k_\epsilon \Rightarrow |c_k - c_\ell| \leq \epsilon/2$. Daraus folgern wir mit der Dreiecksungleichung: $k \geq k_\epsilon \Rightarrow |c_k - r| \leq \epsilon$: Dazu sei $c_{\ell *}$, $\ell * \geq k_\epsilon$ eines der unendlich vielen Elemente der Cauchyfolge in $[a_m, b_m]$. Dann gilt: $|c_{\ell *} - r| \leq |b_m - a_m| \leq \epsilon/2$, also schließlich $k \geq k_\epsilon \Rightarrow |c_k - r| \leq |c_k - c_{\ell *}| + |c_{\ell *} - r| \leq \epsilon$.

Beispiel. Sei $\{a_k\}$ eine "geometrisch majorisierte" Folge, d.h. es gelte für alle k

$$(q1) \qquad |a_{k+2} - a_{k+1}| \leq q \cdot |a_{k+1} - a_k|, \;\; q < 1.$$

Behauptung. Dann ist $\{a_k\}$ Cauchyfolge.
Beweis. Aus der Voraussetzung folgern wir zunächst durch $k-$malige Anwendung eine Ungleichung, die ebenfalls den Namen "geometrische Majorisierung" verdient:

$$(q2) \qquad |a_{k+2} - a_{k+1}| \leq q^k \cdot |a_1 - a_0|.$$

Aus $(q2)$ folgt mit der Dreiecksungleichung (wir setzen $\ell > k$ voraus):

$$|a_\ell - a_k| \leq \sum_{m=k}^{\ell-1} |a_{m+1} - a_m| = |a_1 - a_0| \cdot \sum_{m=k}^{\ell-1} q^{m-1} \;\; \leq \frac{|a_1 - a_0|}{1 - q} \cdot q^{k-1}.$$

Diese Majorisierung durch eine bekannte Nullfolge zeigt, daß $\{a_k\}$ Cauchyfolge ist, also in $\mathbb{R}$ einen Grenzwert hat.

Dieses einfache Argument hat eine viel zitierte Konsequenz, das

Kontraktionslemma. f sei eine Abbildung mit Dehnungsschranke $q < 1$, es gilt also $|f(y) - f(x)| \le q \cdot |y - x|$. Es sei a_0 beliebig. *Dann definiert* $\underline{a_{k+1} := f(a_k), k \in \mathbb{N},}$ *eine geometrisch majorisierte Cauchyfolge mit* $a_\infty = f(a_\infty)$, *einem Fixpunkt von* f.

Das Kontraktionslemma hat eine Prestigeformulierung: **Banachscher Fixpunktsatz.**

Aufgabe. Warum ist der Grenzwert $a := \lim_{k\to\infty} a_k$ Fixpunkt von f, also $a = f(a)$?

Beispiel. $f(x) := 1/(2 + x)$, also $0 \le x \Rightarrow |f'(x)| \le 1/4$, erfüllt mit $a_0 = 0$ die Voraussetzungen des Kontraktionslemmas. Die Folge $a_{k+1} := f(a_k), k \in \mathbb{N}$ liefert die Teilbrüche eines konvergenten Kettenbruchs

$$\frac{1}{2},\ \frac{2}{5} = \frac{1}{2 + \frac{1}{2}},\ \frac{5}{12} = \frac{1}{2 + \frac{1}{2+1/2}},\ f(\frac{5}{12}),\ \ldots, \frac{1}{2 + \frac{1}{2+1/2\ldots}}\ ,$$

der gegen den Fixpunkt $a = \sqrt{2} - 1 = f(a)$ konvergiert.

Die dritte Formulierung der Vollständigkeit geht auf Dedekind zurück. Sie beruht auf der folgenden

Definition. Betrachte eine nichtleere Teilmenge $\emptyset \ne A \subset \mathbb{R}$. Eine Zahl s heißt *kleinste obere Schranke* oder *obere Grenze* oder *Supremum* von A, Bezeichung: $s = \sup A$, falls gilt:

1.) s ist obere Schranke von A (d.h. $a \in A \Rightarrow a \le s$).

2.) Es gibt keine kleinere obere Schranke (d.h. $r < s \Rightarrow$ es gibt ein $a \in A$ mit $r < a$).

Dritte Formulierung des Vollständigkeitsaxioms

Jede nicht leere nach oben beschränkte Teilmenge A von $\mathbb{R}$ besitzt eine kleinste obere Schranke, auch obere Grenze genannt und mit $\sup A$ bezeichnet.

Bemerkung. Der Satz von Bolzano-Weierstraß folgt hieraus unmittelbar. Umgekehrt wird diese dritte Formulierung so ähnlich wie jener Satz bewiesen. Die Supremum-Formulierung der Vollständigkeit gilt als die eleganteste. Allerdings verlangt sie eine sicherere Beherrschung der Argumente der Analysis als die beiden anderen.

Argumentationsübung. Vorausgesetzt werde die dritte Vollständigkeitsformulierung und gegeben sei eine Intervallschachtelung $\{[a_k, b_k]\}$. Wir zeigen, daß diese Schachtelung konvergiert. Nach Voraussetzung hat die durch b_1 beschränkte Folge $\{a_k\}$ eine kleinste obere Grenze c. Da jedes b_k eine obere Schranke der Folge $\{a_k\}$ ist, folgt für jedes k: $c \le b_k$, also $c \in [a_k, b_k]$. Damit ist c Grenzwert der Intervallschachtelung.

Anwendung. Definition der Bogenlänge dehnungsbeschränkter Kurven $p : [a, b] \to \mathbb{R}^2$.
Voraussetzung: $|p'| \leq L$, also $|p(y) - p(x)| \leq L \cdot |y - x|$.
Betrachte eine beliebige Einteilung $a = x_0 < x_1 < \ldots < x_n = b$ des Definitionsintervalls $[a, b]$. Dann gilt für die Summe der Sehnenlängen

$$\sum_{k=1}^{N} |p(x_k) - p(x_{k-1})| \leq \sum_{k=1}^{N} L \cdot |x_k - x_{k-1}| = L \cdot |b - a|.$$

Damit haben wir eine von der Einteilung unabhängige obere Schranke für die Länge aller Sehnenzüge, können also mit der dritten Vollständigkeitsformulierung **definieren:**

$$\text{Länge}\,(p\big|_{[a,b]}) := \sup\{\sum_{k=1}^{N} |p(x_k) - p(x_{k-1})|; a = x_0 < \ldots < x_N = b\}.$$

Offenbar führt diese anschauliche Definition nicht zu Berechnungsformeln. Diese finden wir im Abschnitt Integralrechnung.

Erfolgsbilanz. Die reellen Zahlen haben uns also erlaubt, durch Zitieren des Vollständigkeitsaxioms die am Anfang geschilderten Existenzprobleme zu lösen. Mit Hilfe unserer gleichmäßigen Fehlerschranken konnten wir durch Zitieren des Archimedes Axioms unsere Kenntnisse über die Approximationen auf die Grenzfunktionen ausdehnen. Außer daß die reellen Zahlen in ihrer Gesamtheit nicht vorstellbar sind, sind beim Umgang mit den reellen Zahlen keine erheblichen technischen Schwierigkeiten aufgetaucht, die Anwendung beider Axiome besteht in einfachem Zitieren. Ich wiederhole mit Nachdruck, daß das an der Verwendung *gleichmäßiger Fehlerschranken* liegt. Das Zusammenspiel der reellen Zahlen mit der noch zu besprechenden Stetigkeit wird neue Argumentationen erfordern.

Komplexe Zahlen

Nach der Einführung der reellen Zahlen ist ein Rückblick auf die Definitionen und Beweise der Differentialrechnung zweckmäßig. Für diese Wiederholung möchte ich die komplexen Zahlen zur Verfügung haben. Etwa von 1500-1800 haben Mathematiker mit schlechtem Gewissen mit der "imaginären Einheit", $\mathbf{i}^2 = -1$, gerechnet. Ab 1800 wurden die komplexen Zahlen u. a. von Gauß als zweidimensionale Zahlen verständlich gemacht. Sie sind aus Elektrotechnik, Quantentheorie und weiten Teilen der Mathematik nicht wegzudenken.
Die komplexen Zahlen werden so definiert, daß die euklidische Ebene zu einem vollständigen Zahlkörper wird. Die Multiplikation ist mit der euklidischen Geometrie verträglich: $|z \cdot w| = |z| \cdot |w|$. Die komplexen rationalen Funktionen sind mit denselben Argumenten differenzierbar wie die reellen rationalen Funktionen. Die ("zweidimensionalen") komplexen Funktionen müssen anders veranschaulicht werden als die ("eindimensionalen") reellen Funktionen. Es gibt eine "gemischte Kettenregel", in der reelle und komplexe Differenzierbarkeit gemischt werden: $c : I \to \mathbb{C},\quad f : \mathbb{C} \to \mathbb{C} \;\Rightarrow\; (f \circ c)' = (f' \circ c) \cdot c'$.
Wir beweisen für (nicht unbedingt gleichmäßig) differenzierbare Funktionen den Schrankensatz: $|f'| \le L \Rightarrow |f(w) - f(z)| \le L \cdot |w - z|$.

Erstes Ziel ist, die euklidische Ebene zu einem Zahlkörper zu machen. Mit $\mathbf{e}, \mathbf{i}$ bezeichnen wir eine Orthonormalbasis. Die reelle Gerade stellen wir uns als die Vielfachen des ersten Basisvektors vor und schreiben daher statt $r \cdot \mathbf{e}$ kürzer r. Die additive Gruppe des $\mathbb{R}$-Vektorraums $\mathbb{C} := \operatorname{span}_{\mathbb{R}}\{\mathbf{e}, \mathbf{i}\}$ ist die additive Gruppe des gesuchten Zahlkörpers:
$(a \cdot \mathbf{e} + b \cdot \mathbf{i}) + (x \cdot \mathbf{e} + y \cdot \mathbf{i}) = (a + x) \cdot \mathbf{e} + (b + y) \cdot \mathbf{i}$.
Außerdem kennen wir die Multiplikation von $\mathbb{R} \times \mathbb{C}$ nach $\mathbb{C}$ (bzw. von $\mathbb{C} \times \mathbb{R}$ nach $\mathbb{C}$) schon (als die Multiplikation mit Skalaren in dem $\mathbb{R}$-Vektorraums $\mathbb{C}$:
$(r \cdot \mathbf{e}) \cdot (x \cdot \mathbf{e} + y \cdot \mathbf{i}) = r \cdot (x \cdot \mathbf{e} + y \cdot \mathbf{i}) = rx \cdot \mathbf{e} + ry \cdot \mathbf{i} = (x \cdot \mathbf{e} + y \cdot \mathbf{i}) \cdot (r \cdot \mathbf{e})$.
Insbesondere: $\mathbf{e} \cdot \mathbf{i} = \mathbf{i} = \mathbf{i} \cdot \mathbf{e}$.
Um $\mathbb{C}$ zu einem kommutativen Ring zu machen, fehlt uns nur noch das, was die historische Vorgeschichte nahe legt:

$$\mathbf{i} \cdot \mathbf{i} := \; -\mathbf{e}.$$

Auf beliebige Linearkombinationen ausgedehnt haben wir daher folgende Definition der Multiplikation

$$(a \cdot \mathbf{e} + b \cdot \mathbf{i}) \cdot (x \cdot \mathbf{e} + y \cdot \mathbf{i}) := (ax - by) \cdot \mathbf{e} + (ay + bx) \cdot \mathbf{i}\,.$$

Diese Multiplikation hat wegen $(a \cdot \mathbf{e} + b \cdot \mathbf{i}) \cdot (a \cdot \mathbf{e} - b \cdot \mathbf{i}) = (a^2 + b^2) \cdot \mathbf{e}$

ein Inverses: $(a \cdot \mathbf{e} + b \cdot \mathbf{i})^{-1} = (a \cdot \mathbf{e} - b \cdot \mathbf{i})/(a^2 + b^2)$.

Damit ist $\mathbb{C}$ ein kommutativer Körper.

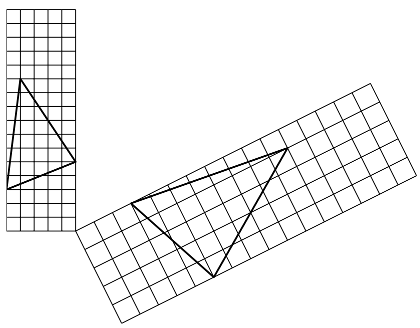
Verträglichkeit der komplexen Zahlen mit der euklidischen Geometrie

Um sich an die komplexen Zahlen als zweidimensionale Zahlen zu gewöhnen, ist deren Verträglichkeit mit der euklidischen Geometrie wichtig.
(i) Die Addition liefert die Translationen, $\mathrm{Trans}_a : z \to a + z,\ a, z \in \mathbb{C}$.

Definition der zu $z = x + \mathbf{i}y$ konjugierten komplexen Zahl: $\overline{z} := x - y \cdot \mathbf{i}$.
Definition der euklidischen Länge von $z = x + y \cdot \mathbf{i}$: $|z| := \sqrt{x^2 + y^2} = \sqrt{z\overline{z}}$.

(ii) **Satz** (Verträglichkeit der euklidischen Geometrie mit der Multiplikation)
Für $c := a + b \cdot \mathbf{i}$ und $z := x + y \cdot \mathbf{i}$ gilt:
1.) $\overline{c \cdot z} = \overline{z} \cdot \overline{c}$, $|c \cdot z| = |c| \cdot |z|$.
2.) Für $c \neq 0$ ist die Abbildung $z \mapsto c \cdot z$ eine Ähnlichkeitsabbildung: Jedes Dreieck $\{z_1, z_2, z_3\}$ wird auf das Dreieck $\{c \cdot z_1, c \cdot z_2, c \cdot z_3\}$ abgebildet, dessen Seitenlängen $|c|$ mal so lang sind. Für $c \neq 1$ ist nur 0 ein Fixpunkt. Daher ist die Abbildung $z \mapsto c \cdot z$ eine Drehung um 0 gefolgt von einer Streckung mit dem Faktor $|c|$.
Beweis.
1.) $\overline{c \cdot z} = (ax - by) - (ay + bx) \cdot \mathbf{i} = (a - b \cdot \mathbf{i}) \cdot (x - y \cdot \mathbf{i}) = \overline{c} \cdot \overline{z}$.
$|c \cdot z|^2 = (c \cdot z) \cdot \overline{(c \cdot z)} = c \cdot (z \cdot \overline{z}) \cdot \overline{c} = |c|^2 \cdot |z|^2$.

2.) $|(c \cdot z_i) - (c \cdot z_j)| = |c(z_i - z_j)| =$
$|c| \cdot |z_i - z_j|$,
$i, j \in \{1, 2, 3\}$.



Multiplikation mit einer komplexen Zahl $c \neq 0$ ist eine winkeltreue Abbildung, $z \mapsto cz$.

(iii) **Satz** (Verträglichkeit der Inversion Inv: $z \mapsto 1/z$ mit Geraden und Kreisen)
Durch die Inversionsabbildung $z \mapsto 1/z$ werden abgebildet:

1. Geraden durch 0 auf Geraden durch 0, (nämlich $r \to r \cdot (c/|c|)$ auf $r \to 1/r \cdot (\bar{c}/|c|)$,)
2. Geraden nicht durch 0 auf Kreise durch 0, Kreise durch 0 auf Geraden nicht durch 0,
3. Kreise nicht durch 0 auf Kreise nicht durch 0.

Beweis. Kreise mit Mittelpunkt m und Radius r beschreiben wir als $\{z\ ;\ |z-m| = r\}$ oder besser $K_r(m) := \{z\ ;\ (z-m)(\bar{z}-\bar{m}) = r^2\}$. Falls $m\bar{m} = r^2$ ist, gilt $0 \in K_r(m)$.

Das Bild dieser Kreise unter der Inversionsabbildung Inv: $z \mapsto w = 1/z$ ist

$$\mathrm{Inv}(K_r(m)) = \{w = 1/z\ ;\ (z-m)(\bar{z}-\bar{m}) - r^2 = 0 = (1-mw)(1-\bar{m}\bar{w}) - r^2 w\bar{w}\ \},$$

also, falls $m\bar{m} = r^2$, folgende Gerade, die nicht durch 0 geht:

$$\mathrm{Inv}(K_r(m)) = \{w\ ;\ 1 = mw + \bar{m}\bar{w}\},$$

und falls $m\bar{m} \neq r^2$, $M := \bar{m}/(m\bar{m} - r^2)$, $R^2 := M\bar{M} - 1/(m\bar{m} - r^2)$, dann der Kreis:

$$\mathrm{Inv}(K_r(m)) = \{w\ ;\ (w-M)(\bar{w}-\bar{M}) = R^2\ \}.$$

Die Bilder von Geraden sind leicht zu bestimmen. Beachte $\mathrm{Inv} \circ \mathrm{Inv} = \mathrm{id}$.

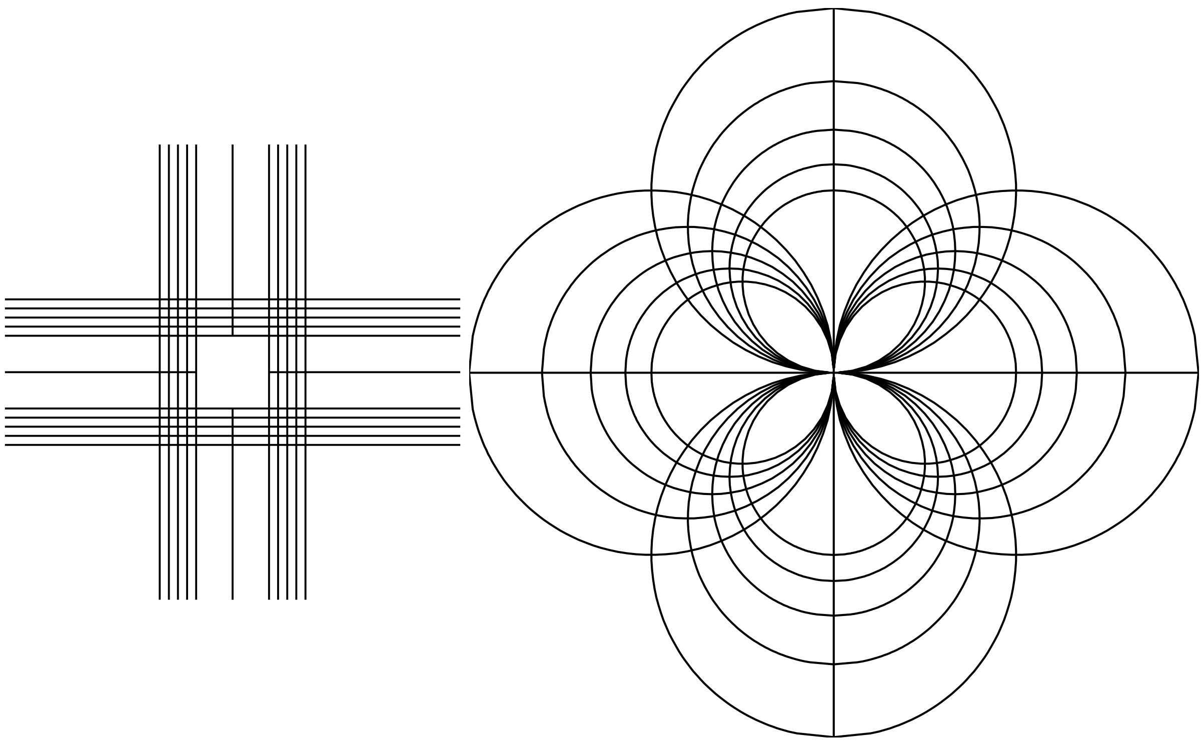


Bild unter $z \mapsto 1/z$ von vier achsenparallelen (zu kurz gezeichneten) Streifen.
Die parallelen Geraden werden in Kreise abgebildet, die die Achsen in 0 berühren.

Zusatz. Diese Kreistreue dehnt sich aus auf die Abbildungen (genannt Möbius Transformationen) $z \mapsto (az+b)/(cz+d) \neq const$, die sich im Falle $c \neq 0$, $ad - bc \neq 0$ schreiben lassen als

$$z \mapsto \frac{(az+b)}{(cz+d)} = \frac{a}{c} - \frac{(ad-bc)/c^2}{(z+d/c)},$$

also als Komposition von zwei Translationen (um a/c und d/c), einer Drehstreckung (mit dem Faktor $-(ad-bc)/c^2$) und einer Inversion, d.h. von lauter kreistreuen Abbildungen.

Satz: **Vollständigkeit von $\mathbb{C}$**

Die komplexen Zahlen sind vollständig, d. h. jede Cauchyfolge $\{z_k\}$, $z_k \in \mathbb{C}$, konvergiert.

Beweis. $\{z_k = x_k + y_k \cdot \mathbf{i}\}$ ist genau dann komplexe Cauchyfolge, wenn sowohl $\{x_k\}$ wie $\{y_k\}$ reelle Cauchyfolgen sind, nämlich wegen

$$|z_\ell - z_k|^2 = |x_k - x_\ell|^2 + |y_k - y_\ell|^2.$$

Erfreulicher Weise sind keine neuen Anstrengungen nötig, die im reellen geleistete Arbeit war ausreichend. Wir werden auch im Komplexen durch Approximation und Zitieren der Vollständigkeit neue interessante Funktionen gewinnen können.

Rationale Funktionen.

Die formalen Rechnungen sind von denen im Reellen kaum zu unterscheiden, mehr Mühe werden wir mit dem Entwickeln anschaulicher Vorstellungen haben. Zuerst machen wir für Polynome *dieselben Rechnungen* wie im Reellen, z. B.:

$$|a|, |z| \le R \Rightarrow |z^n - a^n - n \cdot a^{n-1} \cdot (z-a)| \le \frac{n \cdot (n-1)}{2} \cdot R^{n-2} \cdot |z-a|^2.$$

In Worten: Die Potenzfunktionen $P : z \mapsto P(z) := z^n$ sind gleichmäßig durch die linearen Funktionen $z \to a^n + n \cdot a^{n-1} \cdot (z-a)$ mit quadratischem Fehler approximierbar. Deswegen nennen wir auch im komplexen Fall die Funktion $P' : z \mapsto P'(z) := nz^{n-1}$ "Ableitung" der Funktion P.

Wie im Reellen dehnt sich diese *gleichmäßige lineare Approximierbarkeit* mit quadratischem Fehler auf Polynome aus:

$$|a|, |z| \le R,\ P(z) := \sum_{k=0}^{n} a_k z^k,\ P'(z) = \sum_{k=1}^{n} a_k k z^{k-1},\ K := \sum_{k=2}^{n} |a_k| \frac{k(k-1)}{2} R^{k-2}$$
$$\Rightarrow \qquad |P(z) - P(a) - P'(a)(z-a)| \le K \cdot |z-a|^2.$$

Die Funktion $f : z \mapsto f(z) := 1/z$ ist bei $a \ne 0$ differenzierbar mit $f'a) = -1/a^2$ und auch gleichmäßig differenzierbar, wenn man einen Abstand $|a| \ge r > 0$ von 0 einhält:

$$0 < r \le |a|,\ |z-a| \le r/2 \Rightarrow |\frac{1}{z} - \frac{1}{a} + \frac{z-a}{a^2}| \le \frac{2}{a^2}|z-a|^2.$$

Ehe wir versuchen, diese *Ableitungen der Polynome* anschaulicher zu machen, müssen wir fragen: Wie kann man sich überhaupt komplexe *Funktionen* anschaulich vorstellen? Wir können nicht mehr den aus dem Reellen gewohnten Graph von f, $\{(z, f(z)) \in \mathbb{C} \times \mathbb{C}\}$ benutzen, weil $\mathbb{C} \times \mathbb{C}$ vierdimensional ist. Aber was sonst?
In großen Bereichen der Mathematik werden die Vokabeln "Funktion" und "Abbildung" synonym gebraucht. Im reell eindimensionalen hat sich das nicht durchgesetzt, weil eindimensionale "Bilder" zu dürftig sind. Für komplexe Funktionen dagegen erhält man brauchbare Veranschaulichungen, indem man sie als Abbildungen ansieht: Man male in den Definitionsbereich der Funktion f ein (nicht zu einfaches) Bild und betrachte, was unter der Abbildung f aus diesem Bild wird! Landkarten der Erde sind Beispiele für dies Verfahren. Auch die Texturen der Computergraphik sind aus diesem Grund eingeführt. Für viele mathematische Zwecke ist ein quadratisches Gitternetz im Definitionsbereich von f schon ein "genügend kompliziertes Muster", wie das folgende Beispiel für $f(z) = z^2$ zeigt. Man beachte, daß das Bild aus dem Definitionsbereich in der Nähe der Nullstelle der Ableitung am stärksten verzerrt wird. Die entfernteren quadratischen Netzmaschen werden alle auf nur geringfügig verformte Quadrate abgebildet. Die Netzkurven im Bild sind Parabeln, wie man aus $z = x+i\cdot b \to u(x)+i\cdot v(x) := (x^2-b^2)+i\cdot 2b\cdot x,\ \ u = v^2/(4b^2)-b^2$ sieht.

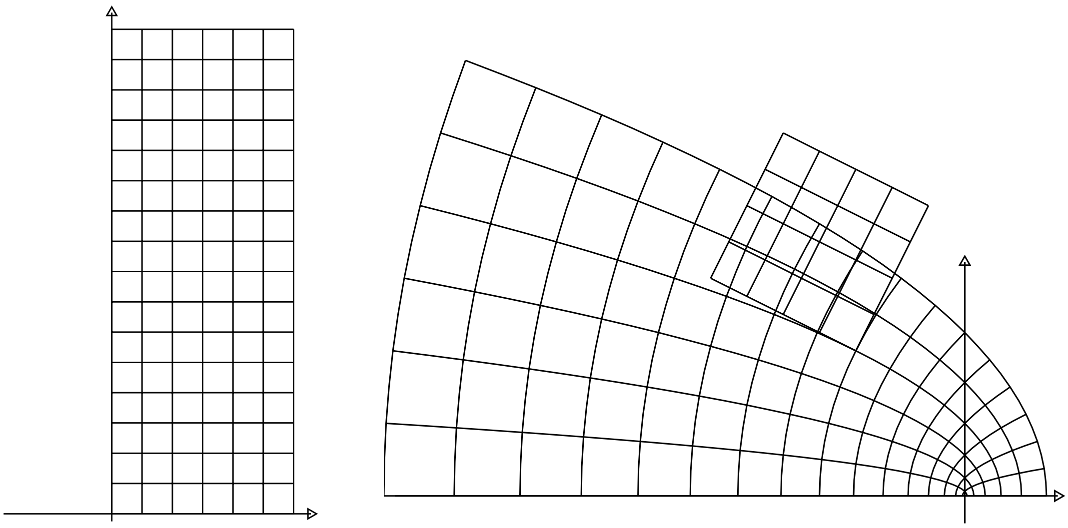


Bild unter $z \to z^2$, mit einer linearer Approximation
Die Gittermaschen bei $f(a)$ im Bild sind um so kleiner, je kleiner $|f'(a)|$ ist, $f'(0) = 0$.

Wir zeigen die starke Verformung in der Nähe von Nullstellen der Ableitung und die weitgehende Respektierung der weiter entfernten quadratischen Netzmaschen an einem zweiten Beispiel, an $z \to z - z^3/3$. Die Nullstellen der Ableitung sind bei ± 1.

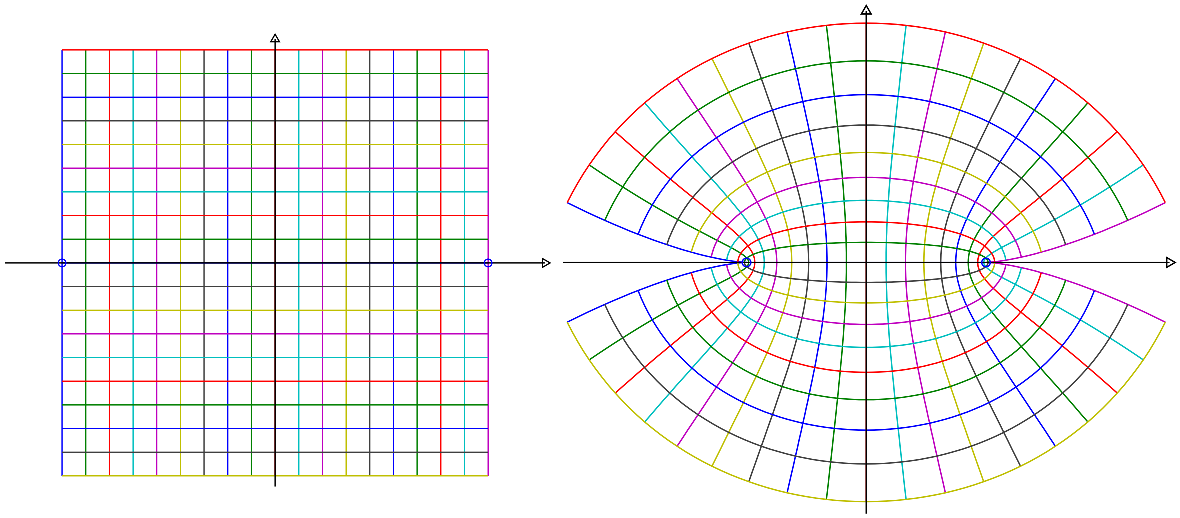


Bild unter $z \to z - z^3/3$. Die Nullstellen von f' bei ± 1 sind markiert.

Es ist instruktiv, auch andere Muster im Definitionsbereich abzubilden, z.B. Polarkoordinatennetze. Das verschiebe ich etwas, weil es einfacher ist zu sagen, worauf man sich bei diesen Bildern konzentrieren soll, wenn die Theorie etwas weiter entwickelt ist. Wie im Reellen schließt sich an die bisher differenzierten einfachen Beispiele die Behandlung der Differentiationsregeln an. Wir haben erstens die zum Reellen analogen Regeln, die im nächsten Satz formuliert sind. Deren Beweise kann man wörtlich übernehmen. Außerdem haben wir ein neues konzeptionelles Problem: Die Netzlinien in den gezeigten Bildern sind Kurven in der Euklidischen Ebene, die wir ja schon differenzieren können. Aber der Vergleich zwischen den Kurven im Definitionsbereich und im Bildbereich soll jetzt benutzt werden, um die komplexe Funktion (= Abbildung) besser zu verstehen. Daher brauchen wir eine Regel, die die Ableitung der *komplexen* Funktion benutzt, um aus den *Kurventangenten* im Definitionsbereich die *Kurventangenten* im Bildbereich auszurechnen. Ich nenne diese Regel *gemischte Kettenregel*. Auch hier ist der Beweis sehr ähnlich zu früheren. Ich führe ihn vor, weil er die Konzepte der reellen und der komplexen Differenzierbarkeit mischt. Zunächst also die Formulierung der schon bekannten Regeln:

Satz. **Die Ableitungsregeln.**

f, g seien komplexe Funktionen, für die gilt: Es gibt andere komplexe Funktionen f', g', die die in den Voraussetzungen genannten guten linearen Approximationen liefern (es spielt keine Rolle, ob die Konstanten $K_f, K_g \ldots$ von a abhängen oder gleichmäßig gelten, solange man nur in Voraussetzung und Behauptung *dasselbe* verlangt):

Voraussetzungen: $|z-a| \le r \Rightarrow |f(z) - f(a) - f'(a)\cdot(z-a)| \le K_f \cdot |z-a|^2,$

$$|z-b| \le r \Rightarrow |g(z) - g(b) - g'(b)\cdot(z-b)| \le K_g \cdot |z-b|^2.$$

Dann gelten die **Differentiationsregeln:**

$$(\alpha f + \beta g)' = \alpha \cdot f' + \beta \cdot g' \qquad (a = b)$$
$$(f{\cdot}g)' = f' \cdot g + f \cdot g' \qquad (a = b)$$
$$(f{\circ}g)' = (f' \circ g) \cdot g' \qquad (a = g(b))$$

Auf Grund dieser Regeln sind die komplexen rationalen Funktionen $z \to P(z)/Q(z)$ außerhalb der Nullstellen des Nenners (in dieser bisher noch unanschaulichen, nur durch Ungleichungen beschriebenen Weise) komplex differenzierbar. Die Beweise ergeben, daß die zusammengesetzten Funktionen $\alpha \cdot f + \beta \cdot g$, $f \cdot g$, $f \circ g$ durch die aus den Ableitungen berechneten linearen Funktionen mit quadratischem Fehler approximiert werden, wie es über f, g vorausgesetzt wurde. (Spätere Resultate zeigen, daß es, *anders als im Reellen*, mit gröberem Fehler komplex differenzierbare Funktionen gar nicht gibt.)

Wir kommen zur gemischten Kettenregel und wiederholen unsere Frage: Gegeben eine differenzierbare Kurve $c : I \mapsto \mathbb{C}$ und eine komplex differenzierbare Abbildung f, wie können wir die Ableitung der Bildkurve $\hat{c} = f \circ c : I \to \mathbb{C}$ mit den Mitteln der Differentialrechnung erhalten? Bei differenzierbaren ebenen Kurven $c : I \to \mathbb{C}$ wissen wir bereits, daß die Ableitung eine weitere Kurve $c' : I \to \mathbb{C}$ ist, mit deren Hilfe man gut approximierende Tangenten bekommt, z. B. es gibt r_c, K_c (i.a. abhängig von t_0) so daß

$$|t - t_0| \leq r_c \Rightarrow |c(t) - c(t_0) - c'(t_0)(t - t_0)| \leq K_c \cdot |t - t_0|^2.$$

Satz. **Gemischte Kettenregel**

Sei c wie eben und $\hat{c} := f \circ c$, ferner $a := c(t_0)$ und f komplex differenzierbar bei a, also:

Voraussetzung: $|z - a| \leq r_f \Rightarrow |f(z) - f(a) - f'(a) \cdot (z - a)| \leq K_f \cdot |z - a|^2.$

Behauptung: $(f \circ c)'(t_0) = f'(c(t_0)) \cdot c'(t_0),$

d. h. die lineare Approximation $z \to f(a) + f'(a) \cdot (z - a)$ von f bildet die Tangente $t \to c(t_0) + c'(t_0) \cdot (t - t_0)$ von c auf die Tangente $t \to \hat{c}(t_0) + \hat{c}'(t_0) \cdot (t - t_0)$ von $\hat{c}$ ab. Als Ungleichung:

$$|t - t_0| \leq r_{\hat{c}} \Rightarrow |f \circ c(t) - f \circ c(t_0) - f'(c(t_0)) \cdot c'(t_0) \cdot (t - t_0)| \leq K_{\hat{c}} \cdot |t - t_0|^2.$$

Beweis. Der Beweis folgt dem alten Muster. Ich gebe ihn an, weil hier reelle und komplexe Differenzierbarkeit gemischt werden. Aus den Differenzierbarkeitsvoraussetzungen erhält man mit der Dreiecksungleichung Dehnungsschranken von c: $\ell := |c'(t_0)| + K_c \cdot r_c$ im Intervall $[t_0 - r_c, t_0 + r_c]$ und von f: $L := |f'(a)| + K_f \cdot r_f$ in der Scheibe $\{z;\ |z - a| \leq r_f\}$. Zunächst wird r_c verkleinert, $r_1 := \min(r_c, r_f/\ell)$, so daß das erlaubte r_1-Intervall um t_0 durch c in die für f erlaubte r_f-Scheibe um $a = c(t_0)$ abgebildet wird:

$$|t - t_0| \leq r_1 \Rightarrow |c(t) - c(t_0)| \leq \ell \cdot r_1 \leq r_f.$$

(Dies ist ein wichtiger Schritt, der im Beweis der Kettenregel nicht übersehen werden darf.)

Die hieraus folgende Dehnungsschranke $L \cdot l$ für $\hat{c}$, die nur wenig größer als das Produkt der Ableitungen $|f'(a)| \cdot |c'(t_0)|$ ist, ist in Übereinstimmung mit der zu beweisenden Regel. Für deren Beweis ist der Hauptschritt das Einsetzen der Werte von c (beachte $c(t_0) = a$) in die Differenzierbarkeitsungleichung für f:
$|t-t_0| \le r_1 \Rightarrow |c(t)-c(t_0)| \le r_f \Rightarrow |f(c(t))-f(a)-f'(a)\cdot(c(t)-c(t_0))| \le K_f \cdot |c(t)-c(t_0)|^2$.
Hier wird der vorläufige Term $f'(a) \cdot (c(t) - c(t_0))$ ersetzt durch den erwünschten Term $f'(a) \cdot c'(t_0) \cdot (t-t_0)$ und der Fehler $\le |f'(a)| \cdot K_c \cdot |t-t_0|^2$ wird mit der Dreiecksungleichung zu dem anderen Fehler $K_f \cdot |c(t) - c(t_0)|^2 \le K_f \cdot l^2 \cdot |t - t_0|^2$ geschlagen.
(Erinnerung: $|A - B| \le F_1, |B - C| \le F_2 \Rightarrow |A - C| = |(A - B) - (C - B)| \le F_1 + F_2$.)
Um die Fehler zusammenzufassen, definieren wir $K_{\hat{c}} := K_f \cdot \ell^2 + |f'(a)| \cdot K_c$, und $r_{\hat{c}} := r_1$. Damit haben wir das Ziel erreicht, es gibt $r_{\hat{c}}, K_{\hat{c}}$ so daß gilt:

$$
\begin{aligned}
|t - t_0| \le r_{\hat{c}} \Rightarrow & |\hat{c}(t) - \hat{c}(t_0) - \hat{c}'(t_0) \cdot (t - t_0)| = \\
& = |f(c(t)) - f(c(t_0)) - f'(c(t_0)) \cdot c'(t_0) \cdot (t - t_0)| \le K_{\hat{c}} \cdot |t - t_0|^2.
\end{aligned}
$$

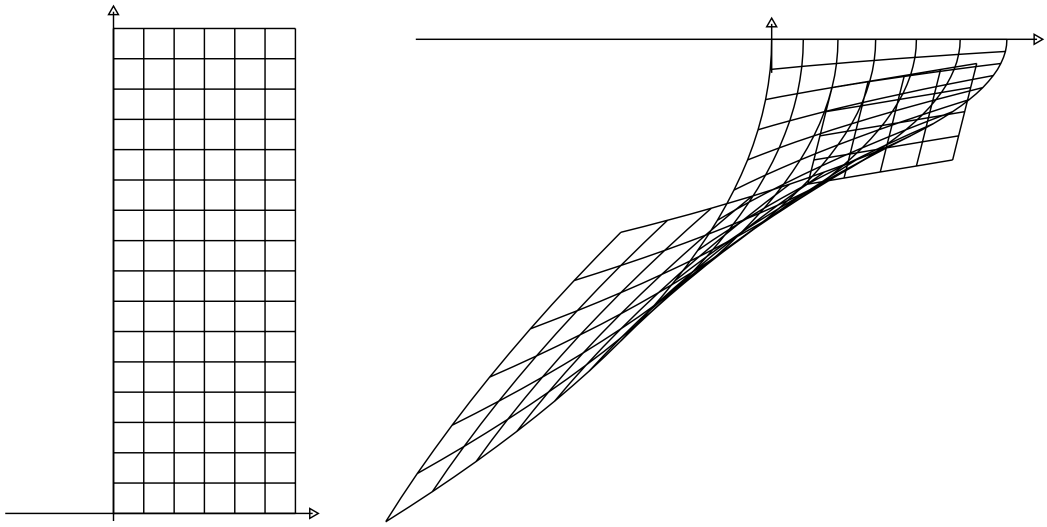
Aus dieser gemischten Kettenregel folgt eine wesentliche Eigenschaft komplex differenzierbarer Abbildungen, die großen Einfluß auf das Aussehen der Figuren hat.

Satz über die Winkeltreue komplex differenzierbarer Abbildungen.

Sei f komplex differenzierbar. Die Abbildung f ist an allen Stellen a mit $f'(a) \neq 0$ winkeltreu.

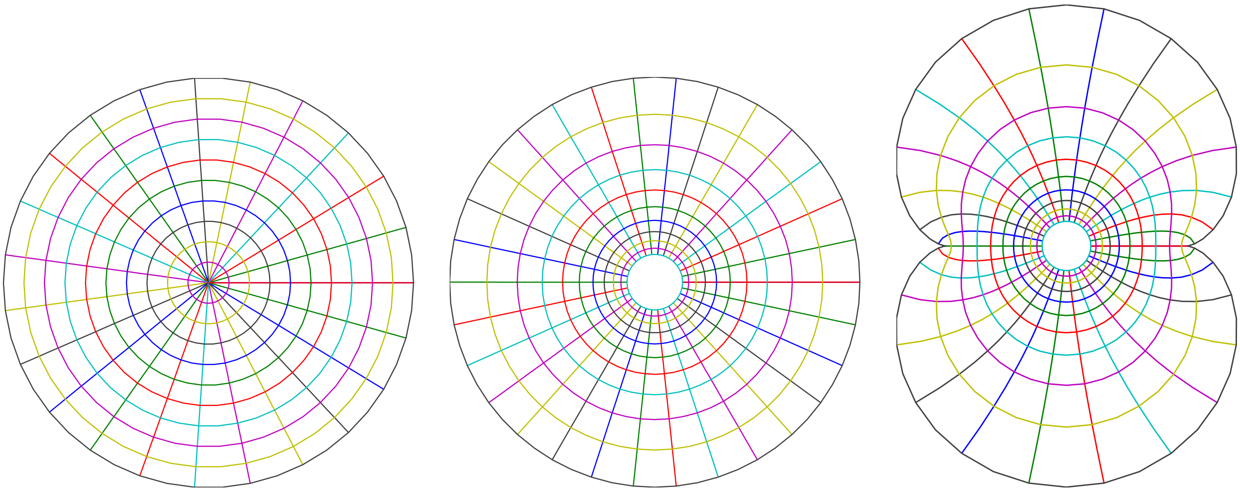
Beweis. Die linearen Approximationen $z \mapsto f(a) + f'(a) \cdot (z - a)$ sind nach dem ersten Satz dieses Kapitels *Drehstreckungen* und damit winkeltreu. Aus der gemischten Kettenregel folgt: Schneiden sich zwei differenzierbare Kurven c_1, c_2 in a, so ist der Winkel ihrer Tangenten *derselbe* wie der Winkel der Tangenten der Bildkurven $f \circ c_1$, $f \circ c_2$ in $f(a)$; diese Eigenschaft heißt *Winkeltreue* der Abbildung f.

Veranschaulichung von Funktionen. Wegen dieser Winkeltreue ist es zweckmäßig, bei der Veranschaulichung komplexer Abbildungen f (wie in den vorhergehenden Illustrationen) orthogonale Kurvenscharen abzubilden, weil leicht im Bild zu erkennen ist, ob sich auch die Bildkurven orthogonal schneiden. Noch besser ist es, von quadratischen Gitternetzen auszugehen; da auch deren Diagonalen sich senkrecht schneiden, gilt dies auch für das Kurvennetz der Bildkurven. Solche Netze wirken auf das Auge wie krummlinige Quadrate, und schon kleine Abweichungen fallen sofort auf. (Man betrachte Bilder von $z \to z + a \cdot \bar{z}^2$ mit kleinem a.) Daher ist die vorgeschlagene Veranschaulichung komplexer Funktionen so gut, daß man leicht erkennt, wenn sie nicht komplex differenzierbar sind. Nachdem verabredet ist, wie Funktionen veranschaulicht werden sollen, kann man weiter fragen, wie man eine Veranschaulichung der Ableitung der Figur hinzufügen kann. Da die Ableitungen als lineare Approximationen $z \to f(a) + f'(a) \cdot (z - a)$ ebenfalls Abbildungen sind, werden sie ebenso dargestellt: Wir bilden ein kleineres Stück der Figur im Definitionsbereich *durch die lineare Approximation* ab und fügen deren Bild zum Bild unter f hinzu. Dann gilt: Im Punkte $f(a)$ haben die *beiden Netzkurven* des Bildes von f durch $f(a)$ als Tangenten die *beiden Gitterlinien* des Bildes der linearen Approximation von f (an der Stelle a).

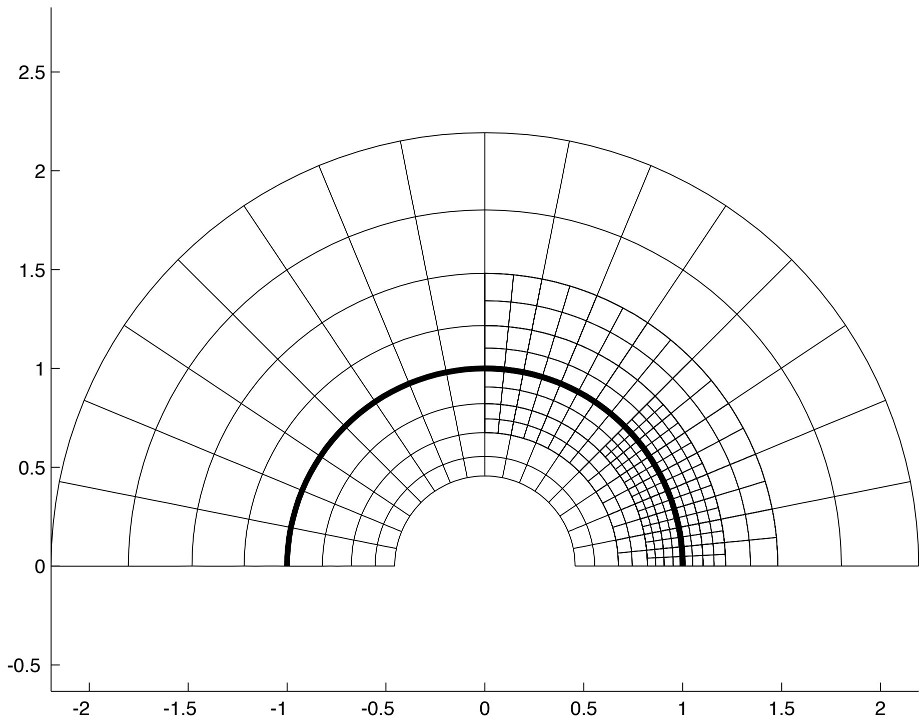


Offensichtlich nicht winkeltreues Bild unter $z \to \bar{z} + z^2/4$.
Das Bildnetz ist nicht "infinitesimal quadratisch" und das Bild hat eine "Faltkante".

Nachdem wir jetzt wissen, worauf in den Bildern geachtet werden soll, komme ich zu Darstellungen in Polarkoordinaten. Wir sehen jetzt, warum die üblichen Polarkoordinaten, also mit radialen Strahlen und mit äquidistanten Kreisen, zur Darstellung von winkeltreuen Abbildungen nicht so gut sind: Die Parametervierecke sind zwar orthogonal, aber die Seitenverhältnisse dieser "krummen" Rechtecke sind in beiden Richtungen unbeschränkt, so daß die Winkeltreue einer Abbildung nicht beurteilt werde kann. Kann man bessere Polarkoordinaten machen? Wir wollen die radialen Strahlen behalten und beginnen mit einem einzigen Kreisring mit den Radien 1 und R. Eine winkeltreue(!) zentrische Streckung mit dem Faktor R vergrößert den Kreisring zu einem außen anschließenden mit den Radien R, R^2. Dies kann man wiederholen. Das Ergebnis zeigt das folgende Bild. Die neuen "winkeltreuen", "im Kleinen quadratischen" Polarkoordinaten vertragen sich besonders gut mit der Inversion: durch die Abbildung $z \to 1/z$ wird das Netz im Innern des Einheitskreises mit dem Netz im äußeren vertauscht. Auch das Verhalten von $z \to z^2$ in einer kleinen Scheibe um die Nullstelle $z = 0$ der Ableitung ist in Polarkoordinaten übersichtlich: Aus Kreisen um 0 (Radius $= r$) werden Kreise um 0 (Radius $= r^2$), aus radialen Strahlen werden radiale Strahlen, aber der Winkel in 0 wird *verdoppelt*, weil die Punkte $c, |c| = 1$ auf dem Einheitskreis durch die Multiplikation mit c ihren Bogenabstand von 1 verdoppeln.



Gewöhnliche und winkeltreue ("lokal quadratische") Polarkoordinaten, sowie Bild des lokal quadratischen Netzes unter $z \mapsto z(1 - z^2/3)$. Auffällig sind die Nullstellen von f'.



Winkeltreue Polarkoordinaten veranschaulichen $z \mapsto z^2$und $z \mapsto 1/z$.

Die feiner karierten Teile der Figur werden durch $z \mapsto z^2$ auf die gröber karierten Teile abgebildet, also werden die Winkel der radialen Strahlen verdoppelt. – Die Inversion $z \mapsto 1/z$ vertauscht Inneres und Äußeres des Einheitskreises, und auch die obere und untere Halbebene. Aber ein konformes Polarkoordinatennetz wird auf sich abgebildet. Vgl. S.72

Ehe wir uns der Konstruktion neuer Funktionen zuwenden, benötigen wir noch einen Ersatz für den im Komplexen nicht einmal formulierbaren Monotoniesatz. Wir finden ihn im sogenannten Schrankensatz (Folgerung 4 zum Monotoniesatz). Da wir jetzt die Vollständigkeit bereits haben, beweise ich den Satz unter schwächeren Voraussetzungen als bisher.

Schrankensatz für komplex differenzierbare Funktionen.

Voraussetzung. Auf der Strecke zwischen z und w sei f komplex differenzierbar mit $|f'| \leq L$. Wir verlangen nur eine *nicht-gleichmäßige Tangentenapproximation*, d.h.
Zu jedem a zwischen z und w gibt es Konstanten $r_a > 0$, $K_a > 0$ (die ausdrücklich von a abhängen dürfen) so daß gilt:

$$|c - a| \leq r_a \Rightarrow |f(c) - f(a) - f'(a) \cdot (c - a)| \leq K_a \cdot |c - a|^2.$$

Behauptung. $$|f(z) - f(w)| \leq L \cdot |z - w|.$$

Beweis. Wegen der nicht gleichmäßigen Voraussetzungen muß der Beweis indirekt geführt werden. Er verläuft für andere "kleine" Fehler als die hier verwendeten quadratischen Fehler ebenso. Angenommen, die Behauptung sei falsch, d.h.

$$\begin{array}{rcll} |f(z) - f(w)| & > & L \cdot |z - w|, & \text{und damit für geeignetes } n \\ |f(z) - f(w)| & \geq & (L + 1/n) \cdot |z - w|. & \end{array}$$

Setze $z_0 = z$, $w_0 = w$ und betrachte den Mittelpunkt $m = \frac{1}{2}(z_0 + w_0)$. Dann gilt wenigstens eine der beiden Ungleichungen:

$$|f(z_0) - f(m)| \geq (L + 1/n) \cdot |z_0 - m|, \qquad |f(m) - f(w_0)| \geq (L + 1/n) \cdot |m - w_0|,$$

denn andernfalls folgte aus der Dreiecksungleichung

$$|f(z_0) - f(w_0)| \leq |f(z_0) - f(m)| + |f(m) - f(w_0)| < (L + 1/n) \cdot (|z_0 - w_0|).$$

ein Widerspruch zur Anfangsannahme. Wenn die erste Ungleichung gilt, setze $z_1 = z_0$, $w_1 = m$, andernfalls setze $z_1 = m$, $w_1 = w_0$.
Dies Verfahren wird wiederholt und $z_0, \ldots, z_k$, $w_0, \ldots, w_k$ seien bereits definiert mit
$|f(z_k) - f(w_k)| \geq (L + 1/n) \cdot |z_k - w_k|$, $|z_k - w_k| = 2^{-k} \cdot |z_0 - w_0|$.
Sei wieder $m = \frac{1}{2}(z_k + w_k)$. In mindestens einer der beiden Hälften haben wir die Ungleichung mit demselben Faktor, also können z_{k+1}, w_{k+1} definiert werden mit
$|f(z_{k+1} - f(w_{k+1})| \geq (L + 1/n) \cdot |z_{k+1} - w_{k+1}|$, $|z_{k+1} - w_{k+1}| = |z_k - w_k|/2$.
Damit sind $\{z_k\}$ und $\{w_k\}$ geometrisch majorisierte Cauchyfolgen, die wegen der Vollständigkeit konvergieren und zwar gegen denselben Grenzwert c. Diese komplexe Zahl c liegt nach Konstruktion für jedes k auf der Strecke zwischen z_k und w_k (evtl. $c = z_k$ oder $= w_k$). Deswegen haben wir **mindestens** eine der beiden Ungleichungen

$$|f(z_k) - f(c)| \geq (L + 1/n) \cdot |z_k - c| > 0 \quad \text{oder} \quad |f(w_k) - f(c)| \geq (L + 1/n) \cdot |w_k - c| > 0.$$

Diese Ungleichungen lassen einen Widerspruch entstehen, wenn wir die Differenzierbarkeitsvoraussetzungen für f an der Stelle c verwenden (ein Schlüsselpunkt der Beweisstrategie):

Wähle k so groß, daß $|z_k - w_k| < \min(r_c, 1/(2nK_c))$ gilt. Dann folgt für alle z auf der Strecke zwischen z_k und w_k offenbar $|z - c| \le r_c$ und daraus weiter:

$$|f(z) - f(c) - f'(c) \cdot (z - c)| \le K_c \cdot |z - c|^2,$$

also wegen $|f'(c)| \le L$ und $|z - c| \le 1/(2nK_c)$:

$$|f(z) - f(c)| \le (L + 1/2n) \cdot |z - c|,$$

im Widerspruch zu den aus der Konstruktion hergeleiteten Ungleichungen für $z = z_k$ oder für $z = w_k$.

Bemerkung. Die Struktur dieses Beweises: "Mit Hilfe der Vollständigkeit wird in einem indirekten Beweis die Stelle lokalisiert, an der ein Widerspruch zu den Voraussetzungen entsteht", findet sich in vielen Beweisen der Analysis. Andere berühmte Beispiele sind der Beschränktheitssatzsatz für stetige Funktionen, der Äquivalenzsatz für verschiedene Stetigkeitsdefinitionen und, später, der Cauchysche Integralsatz.

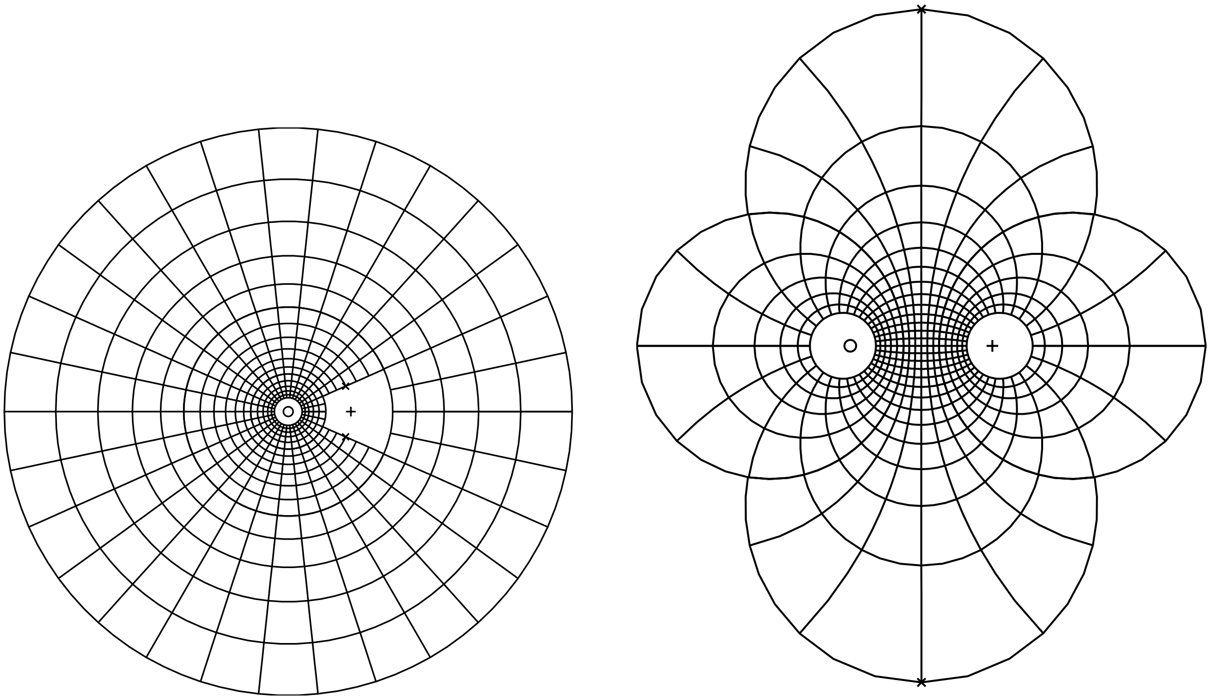


Abbildung winkeltreuer Polarkoordinaten unter $f(z) := (z + 1)/(z - 1)$. Die Abbildung bildet Geraden und Kreise auf Geraden und Kreise ab, so daß man extrapolieren kann, was außerhalb des gezeichneten Netzes passiert. Wegen $f(f(z)) = z$ kann man jede der beiden Figuren als Definitionsbereich, die andere als Bildbereich ansehen. f vertauscht 0 und -1 (markiert o), bildet das Innere des Einheitskreises auf die linke Halbebene ab, das äußere auf die rechte, und umgekehrt. Das Loch um $+1$ (markiert $+$) in der einen Figur wird auf das Äußere der anderen Figur abgebildet. Das suggeriert, von einem "Punkt ∞" zu reden, der mit $+1$ vertauscht wird. Rechts sieht man also Umgebungen von 0 und ∞ abgebildet.

Komplexe Potenzreihen

Die Vollständigkeit der komplexen Zahlen und der Schrankensatz erlauben uns, Grenzfunktionen komplexer Funktionen zu betrachten. Nach der Diskussion der Vollständigkeit hatten wir nur drei besonders wichtige Grenzfunktionen konstruiert. Mit den Potenzreihen konstruieren wir eine große Klasse von Grenzfunktionen, die viele Eigenschaften mit den Polynomen teilen. Mit ihrer Behandlung wird die einfachste und wichtigste Majorisierungstechnik, der Vergleich mit der geometrischen Reihe, geübt. Die Potenzreihen sind eine wichtige Klasse von Funktionen, z.B. können fast alle Funktionen, die individuelle Namen haben, als Potenzreihen beschrieben werden.

Definition. Gegeben sei eine Folge $a_k \in \mathbb{C}$. Bilde dazu die

Polynomfolge: $$P_n(z) := \sum_{k=0}^{n} a_k \cdot z^k.$$

Sowohl die Polynomfolge $\{z \mapsto P_n(z)\}$ wie auch (im Falle der Konvergenz) deren Grenzfunktion $z \mapsto P_\infty(z)$ werden als *Potenzreihe* bezeichnet und meist als $\sum_{k=0}^{\infty} a_k z^k$ geschrieben.

Satz **Majorisierung durch einen Konvergenzpunkt.**
Voraussetzung. Für ein $z_0 \neq 0$ konvergiert $\{P_n(z_0)\}$. Setze $r := |z_0|$.
Variation. Da dann $\{b_k := a_k z_0^k\}$ jedenfalls beschränkt ist (nämlich: zu $\epsilon := 1 > 0$ wähle n_1 mit $m, n \geq n_1 \Rightarrow |b_m - b_n| \leq 1$; dann ist $|b_k| \leq \max_{j=1\ldots n_1}(|b_j| + 1)$, wird oft diese **schwächere Voraussetzung** gemacht, also: $|a_k| \cdot r^k \leq M$ für alle $k \in \mathbb{N}$.
Behauptung. Sei $0 < q < 1$ und $|z| \leq q \cdot r$. Dann ist die Folge $\{P_n(z)\}$ q-geometrisch majorisiert und damit für alle z mit $|z| < |z_0|$ konvergent.

Beweis. $|P_{n+1}(z) - P_n(z)| = |a_{n+1} \cdot z^{n+1}| \leq M \cdot q^{n+1}$, also ist $\{P_n(z)\}$ Cauchyfolge:

$$|P_{n+m}(z) - P_n(z)| \leq \frac{M}{1-q} \cdot q^{n+1}.$$

Kontraposition. Falls $\{P_n(z_1)\}$ nicht konvergiert, so konvergiert $\{P_n(z)\}$ für kein z mit $|z| > |z_1|$.

Dieser Satz gibt eine erste Übersicht über das
Konvergenzverhalten von Potenzreihen:
1.) Es gibt Potenzreihen, die für kein $z \neq 0$ konvergieren, $a_k := k^k$.
2.) Es gibt Potenzreihen, die für jedes $z \in \mathbb{C}$ konvergieren. Beispiele: Die uns schon bekannten Potenzreihen für exp, sin, cos konvergieren für alle $z \in \mathbb{R}$, daher konvergieren sie

nach dem Satz auch für alle $z \in \mathbb{C}$.
3.) Für alle übrigen Potenzreihen gibt es ein $R > 0$, so daß für $|z| < R$ die Folge $\{P_n(z)\}$ konvergiert und für $|z| > R$ divergiert. (Für $|z| = R$ wird keine allgemeine Aussage gemacht.) Diese Zahl R heißt *Konvergenzradius* der Potenzreihe. Offenbar gilt:

$$R = \sup\{|z|; P_n(z)\text{konvergiert}\}.$$

Beachten Sie von Anfang an: In der offenen Scheibe $|z| < R$ hat man zwar Konvergenz, aber *keine quantitative Kontrolle*; jedoch kann man in abgeschlossenen kleineren Scheiben $|z| \le q \cdot R$, $(0 < q < 1)$ mit geometrischen Reihen majorisieren.
Ich schicke zwei Tricks als Handwerkszeug zum Umgang mit Potenzreihen voraus. Der erste bezieht sich auf den Umgang mit den differenzierten Potenzreihen, dazu müssen wir nur den Umgang mit der geometrischen Reihe selbst beherrschen. Der zweite Trick zeigt, wie man damit fertig wird, daß auf dem Rand eines Konvergenzkreises kein einziger Konvergenzpunkt zu liegen braucht (wie bei der Ableitung der geometrischen Reihe), so daß man nicht wie im vorhergehenden Beweis argumentieren kann. Für den ersten Trick differenzieren wir die Summenformel der *endlichen* geometrischen Reihe und lassen aus dem Resultat einfach die negativen Terme weg:

Voraussetzung: $0 \le x < 1$,

Behauptung, bequeme Majorisierungen**:**

$$\sum_{k=0}^{n} x^k = \frac{1-x^{n+1}}{1-x} \le \frac{1}{1-x},$$

$$\left(\sum_{k=0}^{n} x^k\right)' = \sum_{k=1}^{n} kx^{k-1} = \frac{-(n+1)x^n}{1-x} + \frac{1-x^{n+1}}{(1-x)^2} \le \frac{1}{(1-x)^2},$$

$$\left(\sum_{k=0}^{n} x^k\right)'' = \sum_{k=2}^{n} k(k-1)x^{k-2} = \frac{-(n+1)nx^{n-1}}{1-x} - \frac{2(n+1)x^n}{(1-x)^2} + 2\frac{1-x^{n+1}}{(1-x)^3} \le \frac{2}{(1-x)^3}.$$

Voraussetzung: Der Konvergenzradius der Potenzreihe $\{\sum_{k=0}^{n} a_k \cdot z^k\}$ sei $0 < R < \infty$.

Behauptung: Es gibt $M > 0$, so daß $|z| \le q \cdot R \Rightarrow |a_k \cdot z^k| \le M \cdot (\sqrt{q})^k$, also hat man

bequeme Schranken:

$$\sum_{k=0}^{n} |a_k \cdot z^k| \le \frac{M}{1-\sqrt{q}},$$

$$\sum_{k=1}^{n} |a_k k \cdot z^{k-1}| \le \frac{M}{(1-\sqrt{q})^2},$$

$$\sum_{k=2}^{n} |a_k k(k-1) \cdot z^{k-2}| \le \frac{2M}{(1-\sqrt{q})^3}.$$

Beweis. Wähle einen Punkt $z_0 := \sqrt{q} \cdot R$. Nach Definition des Konvergenzradius konvergiert die Potenzreihe bei z_0. Daher ist (wie im ersten Beweis) die Folge $\{a_k z_0^k\}$ beschränkt, d.h. es gibt $M > 0$ so daß $|a_k z_0^k| \leq M$. Daraus folgt die behauptete geometrische Majorisierung $|z| \leq q \cdot R \Rightarrow |a_k z^k| \leq M(\sqrt{q})^k$ mit $\sqrt{q}$ statt q. Um die Schranken zu beweisen, muß nur noch der erste Trick mit $x = \sqrt{q}$ angewandt werden.

Man kann den Konvergenzradius aus den Koeffizienten a_k der Potenzreihe bestimmen. Ich werde die folgenden Kriterien nicht viel benutzen, aber deren Beweise zeigen einen wichtigen Trick: Der Beweis unseres ersten Satzes legt vielleicht die Vermutung nahe, daß man auf der Kreisscheibe $|z| \leq q \cdot R$ (R der Konvergenzradius) tatsächlich mit der geometrischen Reihe mit demselben Faktor q majorisieren kann. Das ist nicht richtig, aber man kann ja, wie im zweiten Trick, mit einem etwas größeren Faktor $\sqrt{q}$, $q < \sqrt{q} < 1$ auch noch zufrieden sein.
Das erste, einfachere, der beiden folgenden Kriterien gibt nur eine hinreichende Bedingung:

Quotientenkriterium. Falls $R := \lim_{k\to\infty} |a_k/a_{k+1}|$ existiert, so ist R der Konvergenzradius.
Beispiel: Hieraus folgt ein einfacher Beweis der Konvergenz der Taylorreihe von exp.

Beweis. Ich behandele nur den Fall $0 < R < \infty$. Wähle $q \in (0,1)$ mit der Absicht, q beliebig nahe an 1 zu wählen und zu zeigen, daß der Konvergenzradius zwischen $q \cdot R$ und R/q liegt.
Wegen der vorausgesetzten Konvergenz von $\{|a_k/a_{k+1}|\}$ kann man zu jedem $q \in (0,1)$ ein k_q finden, so daß gilt:

$$k \geq k_q \Rightarrow \sqrt{q} \cdot R \leq \left| \frac{a_k}{a_{k+1}} \right| \leq \frac{1}{\sqrt{q}} \cdot R.$$

Dann hat man für alle $k \geq k_q$ eine einfache, aber entscheidende Abschätzung:

$$\begin{aligned} |z| \leq q \cdot R \ \Rightarrow \ |P_{k+1}(z) - P_k(z)| &= |a_{k+1} \cdot z^{k+1}| \leq q \cdot R \cdot |a_{k+1}| \cdot |z|^k \leq \\ &\leq \sqrt{q} \cdot |a_k| \cdot |z|^k = \sqrt{q} \cdot |P_k(z) - P_{k-1}(z)|. \end{aligned}$$

In der Scheibe $|z| \leq q \cdot R$ ist die Potenzreihe also $\sqrt{q}$*-geometrisch majorisiert* und damit konvergent, d. h. $q \cdot R$ ist kleiner oder gleich dem Konvergenzradius. (Daß das Anfangsstück der Potenzreihe, nämlich für $k < k_q$, nicht betrachtet wird, stört die Konvergenz nicht.)
Umgekehrt folgt aus $|a_{k+1}| \cdot R \geq \sqrt{q} \cdot |a_k|$:
$|z| \geq R/q \Rightarrow |P_{k+1}(z) - P_k(z)| \geq (1/\sqrt{q}) \cdot |P_k(z) - P_{k-1}(z)|$,
also Divergenz, d. h. R/q ist größer oder gleich dem Konvergenzradius.

Das zweite Kriterium ist notwendig und hinreichend:

Wurzelkriterium. Setze $s := \limsup \sqrt[k]{|a_k|} := \lim_{j\to\infty}(\sup_{k\geq j} \sqrt[k]{|a_k|})$. Dann gilt:
1.) Genau dann, wenn $s = \infty$ ist, konvergiert die Potenzreihe nur für $z = 0$.
2.) Genau dann, wenn $s = 0$ ist, konvergiert die Potenzreihe für alle $z \in \mathbb{C}$.
3.) Ist $0 < s < \infty$, so gilt für den Konvergenzradius R:

$$R = \frac{1}{s} = 1/\limsup \sqrt[k]{|a_k|}.$$

Beweis. Ich behandle nur 3.).
Erstens wird $R \cdot s \leq 1$ gezeigt, mit dem einzigen Beweis dieses Kapitels, der nicht die geometrische Reihe benutzt:
Für jedes z, für das $\{P_k(z)\}$ konvergiert, ist $\{a_k z^k\}$ beschränkt, etwa $|a_k \cdot z^k| \leq M$ für alle $k \in \mathbb{N}$. Nun folgt erstens wegen der Monotonie der k-ten Wurzel und zweitens weil der Graph dieser Wurzelfunktion unterhalb der Tangente bei 1 liegt

$$\sqrt[k]{|a_k|} \cdot |z| \leq \sqrt[k]{M} \leq 1 + \frac{M-1}{k}, \quad \text{also weiter (zuletzt mit dem Archimedes Axiom)}$$
$$s \cdot |z| = \limsup \sqrt[k]{|a_k|} \cdot |z| \leq \sup_{k\geq j}(1 + \frac{M}{k}) = 1 + \frac{M}{j}, \quad \Rightarrow \quad s \cdot |z| \leq 1.$$

Dies gilt für jedes $|z| < R$ (= Konvergenzradius) und damit ist $s \cdot R \leq 1$ bewiesen.
Zweitens wird für jedes $q \in (0,1)$ gezeigt $s \cdot R \geq q$. Beide Ungleichungen und das Archimedes Argument liefern dann $s \cdot R = 1$. Wegen $0 < s = \limsup \sqrt[k]{|a_k|} < \infty$ können wir zu jedem $q \in (0,1)$ ein k_q finden, so daß gilt:

$$k \geq k_q \Rightarrow \sup \sqrt[k]{|a_k|} \leq s/\sqrt{q}\ , \ |a_k| \cdot (\sqrt{q}/s)^k \leq 1.$$

Also gilt für alle $k \geq k_q$ und für $|z| \leq q/s$

$$|P_{k+1}(z) - P_k(z)| = |a_{k+1} \cdot z^{k+1}| \leq |a_{k+1} \cdot (q/s)^{k+1}| \leq \sqrt{q}^{k+1}.$$

In der Scheibe $|z| \leq q/s$ ist daher die Potenzreihe $\sqrt{q}$-geometrisch majorisiert (jedenfalls für $k \geq k_q$) und daher konvergent; d.h. q/s ist kleiner oder gleich dem Konvergenzradius R. Damit ist auch die linke der beiden Ungleichungen $q \leq s \cdot R \leq 1$ bewiesen.

Aufgabe. Behandele die Fälle 1.) und 2.) des Wurzelkriteriums.

Das Konvergenzverhalten der Potenzreihen kennen wir jetzt genügend gut, aber wir wissen noch nichts über die *Eigenschaften der Grenzfunktionen.* Insbesondere:

"Sind sie differenzierbar? Und wenn ja, wie findet man ihre Ableitung?"

Wir werden das mit der geometrischen Majorisierung der differenzierten Potenzreihen $\{P'_n(z)\}$, $\{P''_n(z)\}$ beantworten, die wir oben schon behandelt haben ("zweiter Trick").

Aufgabe. Man kann die differenzierten Potenzreihen auch anders behandeln (immer mit geometrischer Majorisierung). Wiederhole: $(n+1)\cdot q^n \le 1+q+\ldots+q^n \le 1/(1-q)$ und folgere daraus (indem q durch $q^{1/2}$, $q^{1/3}$ usw. ersetzt und die Ungleichung quadriert usw. wird):

$$\begin{array}{rcl}(n+1)\cdot q^n = (n+1)\cdot(q^{1/2})^n\cdot(q^{1/2})^n & \le & (q^{1/2})^n/(1-q^{1/2}),\\ (n+1)^2\cdot q^n = (n+1)^2\cdot(q^{2/3})^n\cdot(q^{1/3})^n & \le & (q^{1/3})^n/(1-q^{1/3})^2,\\ (n+1)^k\cdot(q^{k/(k+1)})^n\cdot(q^{1/(k+1)})^n & \le & (q^{1/(k+1)})^n/(1-q^{1/(k+1)})^k.\end{array}$$

Daher ist für jedes feste k die Folge $\{a_n := n^k\cdot q^n\}$ Nullfolge.

Satz. **Geometrische Majorisierung der differenzierten Reihen**

Voraussetzung. Die Potenzreihe mit den Koeffizienten a_k habe einen Konvergenzradius $R \ge r > 0$. Wähle $q \in (0,1)$.

Behauptung. Die differenzierten Potenzreihen $\{P_n'(z)\}$, $\{P_n''(z)\}$ sind in der Scheibe $|z| \le q\cdot r$ geometrisch majorisiert, damit konvergent. Sie sind, abhängig von der Wahl von q, aber unabhängig von n in der Scheibe $|z| \le q\cdot r$ *beschränkt* (Schranken im Beweis).

Daraus folgt, daß die differenzierten Potenzreihen denselben Konnvergenzradius haben wie die gegebene Potenzreihe.

Beweis. Da nach Voraussetzung die Potenzreihe $\{P_n(z)\}$ für $|z| \le \sqrt{q}\,\cdot\, r$ konvergiert, ist insbesondere $a_k\cdot(\sqrt{q}r)^k$ beschränkt: $|a_k|\cdot\sqrt{q}^k\cdot r^k \le M$ (Trick 2 oben). Daraus folgt in der kleineren Scheibe $|z|\le q\cdot r$ die geometrische Majorisierung der Potenzreihe selbst und ihrer Ableitungen:

Zunächst:

$|P_n(z) - P_{n-1}(z)| = |a_n\cdot z^n| \le |a_n\cdot q^n\cdot r^n| \le M\cdot\sqrt{q}^n.$

Wegen (Trick 1 oben)

$\sum\sqrt{q}^k \le 1/(1-\sqrt{q}),\ \sum k\sqrt{q}^{k-1} \le 1/(1-\sqrt{q})^2,\ \sum k(k-1)\sqrt{q}^{k-2} \le 1/(1-\sqrt{q})^3$

haben wir für alle $n\in\mathbb{N}$ und alle $|z|\le q\cdot r$ die Schranken

$$\begin{aligned}|P_n(z)| &\le M/(1-\sqrt{q}) =: C_q,\\ |P_n'(z)| &\le M/r\cdot 1/(1-\sqrt{q})^2 =: L_q,\\ |P_n''(z)| &\le M/r^2\cdot 1/(1-\sqrt{q})^3 =: B_q.\end{aligned}$$

Die Schranke L_q liefert mit dem Archimedes Argument eine Dehnungsschranke für die Grenzfunktion. Die Schranke B_q liefert ebenso die quadratische Abweichung von der erwarteten Ableitung (wer erwartet nicht $\lim P_k'$?) und damit die (gleichmäßige) Differenzierbarkeit der Grenzfunktion:

Satz. **Ableitung der Grenzfunktion**

Die differenzierte Potenzreihe $\{P_n'(z)\}$ konvergiert in der Scheibe $|z| < r$ gegen die Ableitung der Grenzfunktion P_∞ der Potenzreihe $\{P_n(z)\}$. In jeder kleineren Scheibe $D_{qr} = \{z\ ;\ |z| \leq q \cdot r\}$ approximieren die Tangenten gleichmäßig quadratisch:

$$z, a \in D_{qr} \Rightarrow |P_\infty(z) - P_\infty(a) - \lim P_n'(a)(z-a)| \leq \frac{B_q}{2}|z-a|^2.$$

Daher ist P_∞ differenzierbar und die Differentiationsregel lautet $(P_\infty)' = \lim P_n'$.

Beweis. Wir wiederholen auch noch den Beweis, daß L_q Dehnungsschranke der Grenzfunktion ist. Aus $|P_n'(z)| \leq L_q$ folgt für das Polynom P_n aus dem für Polynome bewiesenen Schrankensatz:
$z, w \in D_{q\cdot r} \Rightarrow |P_n(w) - P_n(z)| \leq L_q \cdot |w - z|$.
Diese Ungleichung gilt für jedes n. Also liefert das Archimedes Argument für die Grenzfunktion P_∞ dieselbe Dehnungsschranke:
$z, w \in D_{q\cdot r} \Rightarrow |P_\infty(w) - P_\infty(z)| \leq L_q \cdot |w - z|$.
Dieselbe Argumentation wiederholen wir für die Tangentenapproximation. Die Grenzfunktion der Potenzreihe $\{P_n'(z)\}$ heiße $Q_\infty(z)$. Aus $|P_n''(z)| \leq B_q$ folgt durch Anwendung des Schrankensatzes
$z, w \in D_{q\cdot r} \Rightarrow |P_n(w) - P_n(z) - P_n'(z)\cdot(w-z)| \leq B_q \cdot |w-z|^2$.
Daher gilt für die Grenzfunktionen wegen der Dreiecksungleichung:
$z, w \in D_{q\cdot r} \Rightarrow |P_\infty(w) - P_\infty(z) - Q_\infty(z)\cdot(w-z)| \leq B_q \cdot |w-z|^2 +$ Summe dreier Nullfolgen.
Das Archimedes Argument beseitigt die Nullfolgen:
$z, w \in D_{q\cdot r} \Rightarrow |P_\infty(w) - P_\infty(z) - Q_\infty(z)\cdot(w-z)| \leq B_q \cdot |w-z|^2$.
Diese Ungleichung besagt nun, daß P_∞ die (gleichmäßige) Differenzierbarkeitsdefinition erfüllt mit der Ableitung $(P_\infty)' = Q_\infty$. Mit anderen Worten: "Man kann Potenzreihen (in $D_{q\cdot r}$) gliedweise, wie Polynome, differenzieren". Die Potenzreihen sind Taylorreihen:

$$\left(\sum_{k=0}^{\infty} a_k\ z^k\right)' = \sum_{k=1}^{\infty} a_k \cdot k \cdot z^{k-1}, \qquad a_k = \frac{P_\infty^{(k)}(0)}{k!},\ P_\infty(z) = \sum_{k=0}^{\infty} \frac{P_\infty^{(k)}(0)}{k!}\ z^k.$$

Bemerkung. Insbesondere ist jede Potenzreihe Ableitung einer anderen Potenzreihe:

$$\left(\sum_{k=0}^{\infty} a_k \cdot \frac{z^{k+1}}{k+1}\right)' = \sum_{k=0}^{\infty} a_k\ z^k.$$

Beispiel. Die Funktion $f(z) = 1/z$ konnten wir bisher nicht als Ableitung einer anderen Funktion schreiben, wohl aber in einem Teil ihres Definitionsbereichs als geometrische Reihe. Dort haben wir jetzt auch eine Stammfunktion:

$$|z-1| < 1 \Rightarrow \frac{1}{z} = \frac{1}{1-(1-z)} = \sum_{k=0}^{\infty}(1-z)^k = \left(-\sum_{k=0}^{\infty} \frac{(1-z)^{k+1}}{k+1}\right)'.$$

In größeren Kreisscheiben geht das (mit etwas längeren Formeln) ebenso:

$$
\begin{aligned}
|z-R|<R \Rightarrow \frac{1}{z} &= \frac{1}{R-(R-z)} = \frac{1}{R}\cdot\sum_{k=0}^{\infty}(1-\frac{z}{R})^k = \\
&= \left(\sum_{k=0}^{\infty}\left[\frac{-1}{k+1}(1-\frac{z}{R})^{k+1} + (\frac{1}{k+1}(1-\frac{1}{R})^{k+1}\right]\right)'.
\end{aligned}
$$

Die Grenzfunktion der letzten Reihe bezeichnen wir mit $L_R(z)$, es gilt $L_R(1) = 0$ und $L_R'(z) = 1/z$ in der Scheibe $|z-R|<R$. Ist $L_R(z)$ vielleicht eine Logarithmusfunktion?

Zur Exponentialfunktion. Zu reellen konvergenten Potenzreihen definiert man komplex fortgesetzte Funktionen:

$$
\exp(z) := \sum_{k=0}^{\infty} \frac{z^k}{k!},\ \sin(z) := \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k+1}}{(2k+1)!},\ \cos(z) := \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k}}{(2k)!}.
$$

exp hat dieselbe Differentialgleichnung und dieselbe Funktionalgleichnug wie im Reellen:

$$
\exp' = \exp \quad \text{und} \quad \exp(a)\cdot\exp(z) = \exp(a+z).
$$

Beweis. Die Ableitung findet man durch gliedweises Differenzieren. Der Quotient

$$
q(z) := \exp(a)\cdot\exp(z)/\exp(a+z)
$$

erfüllt $q(0) = 1$ und wegen der Quotientenregel $q'(z) = 0$. Aus dem Schrankensatz folgt $q(z) = 1$ für alle z. (Ohne den Schrankensatz ist das wesentlich mühsamer.)

Durch Vergleich der Potenzreihen findet man Eulers Formeln:

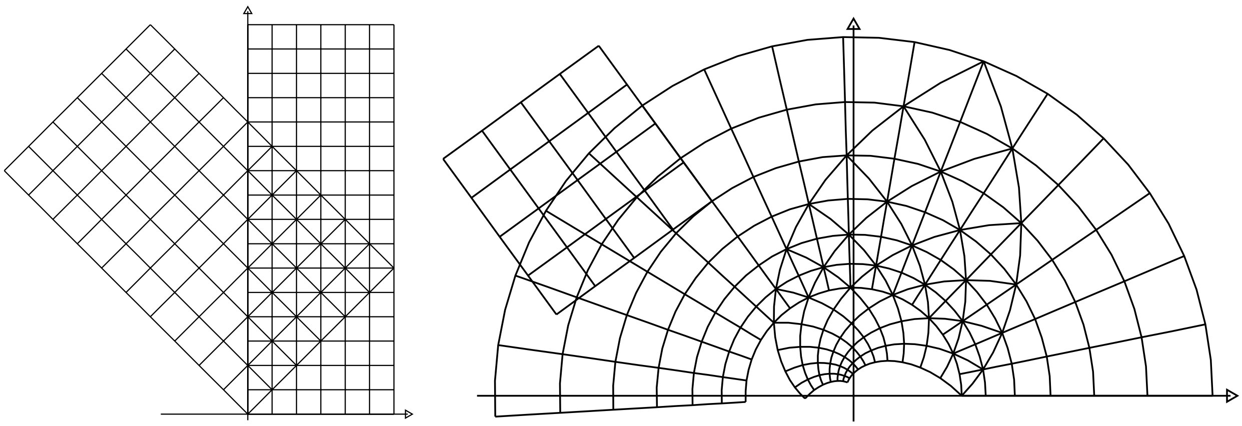
$$
\begin{aligned}
&\exp(iz) = \cos(z) + i\cdot\sin(z)\,, \quad \exp(2\pi i) = \exp(0) = 1, \\
&\cos z = (\exp(iz) + \exp(-iz))/2, \\
&\sin z = (\exp(iz) - \exp(-iz))/2i.
\end{aligned}
$$

Aus dem Additionstheorem für exp folgen durch Einsetzen die Additionstheoreme von sin und cos. Das im Reellen verwendete Argument funktioniert nicht, weil im Komplexen nicht gilt $a^2+b^2 = 0 \Rightarrow a = 0 = b$. Ein Beispiel:

$$
\begin{aligned}
\cos(z+w) &= \frac{1}{2}(\exp(iz+iw) + \exp(-iz-iw)) = \\
&= \frac{1}{2}(\exp(iz)\cdot\exp(iw) + \exp(-iz)\cdot\exp(-iw)) \\
&= +\frac{1}{2}(\exp(iz)+\exp(-iz))\cdot\frac{1}{2}(\exp(iw)+\exp(-iw)) \\
&\quad -\frac{1}{2i}(\exp(iz)-\exp(-iz))\cdot\frac{1}{2i}(\exp(iw)-\exp(-iw)) \\
&= \cos z\cdot\cos w - \sin z\cdot\sin w.
\end{aligned}
$$

Aus $\exp(2\pi i) = \exp(0) = 1$ folgt daher, daß cos und sin auch im Komplexen 2π-periodisch sind:

$$\cos(z + 2\pi) = \cos(z), \quad \sin(z + 2\pi) = \sin(z).$$



Winkeltreue Polarkoordinaten erhält man im **Bild der Exponentialfunktion.** Die orthogonalen 45°-Diagonalen links werden auf orthogonale Diagonalkurven (Spiralen) abgebildet. Man sieht auch eine lineare Approximation.

Umkehrfunktion von exp **im Komplexen.** Da exp $2\pi i$-periodisch ist, gibt es *keine* eindeutige Umkehrfunktion. Sei $\ell(z)$ *eine* (innere) Umkehrfunktion, also $\exp(\ell(z)) = z$. Falls wir schon wüßten, daß wir nach der Kettenregel differenzieren könnten, so folgte:

$$1 = \exp'(\ell(z)) \cdot \ell'(z) = \ z \cdot \ell'(z) = 1, \quad \ell'(z) = 1/z.$$

Nun haben wir gerade (zumindest in Kreisen $|z - R| < R$) Potenzreihen $L_R(z)$ mit Ableitung $L'_R(z) = \ 1/z$ und mit $L_R(1) = 0$ konstruiert. Sind sie Umkehrfunktionen der Exponentialfunktion? Die Funktion $h(z) := \exp(L(z))/z$ erfüllt

$$h(1) = 1 \text{ und } h'(z) = \exp'(L(z)) \cdot L'(z)/z - \exp(L(z))/z^2 = 0,$$

sie ist also nach dem Schrankensatz konstant 1, und $\exp(L_R(z)) = z$ ist in $|z - R| < R$ bewiesen!

Wir stellen nun noch fest, daß die expandierenden Kreise $\{z; |z - R| < R\}$ die rechte Halbebene $\{z : \text{Re } z > 0\}$ ausschöpfen und daß je zwei der Funktionen $L_R(z)$ auf dem Durchschnitt ihrer Konvergenzkreise (also auf dem kleineren Kreis) übereinstimmen:

$$r \le R, \ |z - r| < r \ \Rightarrow \ h(z) := L_R(z) - L_r(z) \ \Rightarrow \ h(1) = 0, \ h'(z) = 0.$$

Diese (bisher) auf der rechten Halbebene definierte (innere) Umkehrfunktion von exp heißt

Hauptwert der komplexen Logarithmusfunktion, Log (z),

$$\text{Log }(1) = 0 \ , \ \exp(\text{Log } z) = z \ , \ \text{Log }'(z) = 1/z.$$

Als Anwendung definieren wir *komplexe Potenzen*

$$z^\alpha := \exp(\alpha \cdot \mathrm{Log}\ z)\ ,\ (\alpha \in \mathbb{C}, \mathrm{Re}\ z > 0).$$

Mit der Kettenregel folgt eine Verallgemeinerung der Differentiationsregel von ganzen auf komplexe Potenzen:

$$(z^\alpha)' = \alpha/z \cdot \exp(\alpha \cdot \mathrm{Log}\ z) = \alpha \cdot z^{\alpha-1}.$$

Aufgabe. Die Funktion $z \to (1+z)^\alpha$ und ihre Taylorreihe um $z = 0$ (Konvergenzradius 1) haben bei $z = 0$ den Wert 1 und sie haben beide die Wachstumsrate $\alpha/(1+z)$. Ihr Quotient ist daher 1. Bestimmen Sie die Koeffizienten dieser Potenzreihe.

Nicht approximierende Taylorreihe. Wir hatten schon erwähnt, daß die Taylorpolynome einer Funktion f nur dann als konvergente Approximationen erwartet werden sollten, wenn die Ableitungen $f^{(k)}(x)$ nicht zu schnell wachsen. Dazu gehört noch ein Beispiel.

Definition: $\quad f(0) := 0, \quad x \in \mathbb{R} \setminus \{0\}\ :\ f(x) := \exp(-1/x^2).$

Behauptung: $\quad f^{(k)}(0) := 0, \quad x \in \mathbb{R} \setminus \{0\}\ :\ f'(x) := \dfrac{2}{x^3} \exp(-1/x^2).$

Beweis. Aus unserer (reellen) Approximation von exp durch die Zinseszins-Funktionen, also aus $0 \le x \Rightarrow \exp(-x) \le 1/(1+x/n)^n$, folgt

$$0 \le f(x) \le \frac{1}{(1+1/(x^2n))^n} \le n^n \cdot x^{2n}, \qquad |f'(x)| \le 2n^n \cdot |x|^{2n-3}.$$

Schon für $n = 1$ folgt aus der ersten Ungleichung $f'(0) = 0$. Die zweite ergibt für $n = 3$:

$$|f'(x) - f'(0) - 0 \cdot x| \le 27|x|^3, \text{ also } f''(0) = 0.$$

Nun folgt ein Induktionsbeweis. Bei jedem Differenzieren von $\exp(-1/x^2)$ erhält man einen Faktor $2/x^3$, das Differenzieren der rationalen Faktoren vergrößert die Nenner nicht so stark. Daher gibt es Konstanten C_k so daß gilt:

$$|x| \le 1 \Rightarrow |f^{(k)}(x)| \le C_k n^n \cdot |x|^{2n-3k}.$$

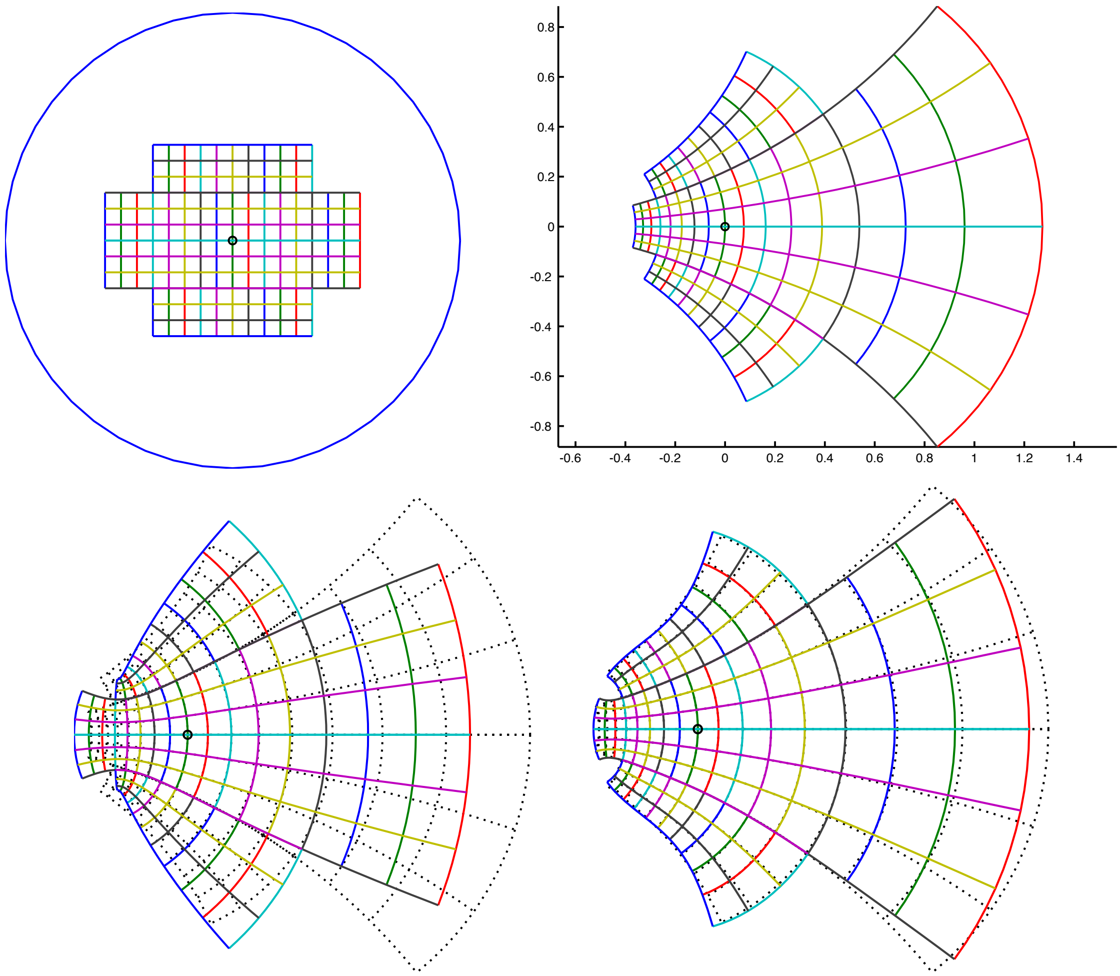
Wählt man $2n \ge 3k + 2$ so folgt

$$|x| \le 1 \Rightarrow |f^{(k)}(x) - 0 - 0 \cdot x| \le C_k n^n |x|^2, \text{ also } f^{(k)}(0) = 0.$$

Alle Taylorpolynome dieser Funktion f sind also 0. Die Taylorreihe konvergiert fabelhaft, aber deren Grenzfunktion 0 hat nichts mit der Funktion f zu tun.

So viel einstweilen zu Potenzreihen. Sie haben den Zweck erfüllt, daß sie uns erlaubt haben, alle Funktionen, von denen Sie vielleicht auf der Schule gehört haben, als komplexe Funktionen zu definieren. Die komplexe Funktionentheorie hält noch allerhand Überraschungen bereit. – Die folgende Aufgabe führt zu einem noch nicht behandelten Problem.

Aufgabe. Gegeben sei eine Funktion f mit $f'' = -f$, $f^2 + f'^2 = 1$. Definiere eine Funktion g durch $g(x) := f(x/3) \cdot (3 - 4f(x/3)^2)$.
Zeige: $g'' = -g$. Falls $f(0) = 0$ ist, so auch $f'(0) = g'(0)$ und unser reeller Beweis der Additionstheoreme würde zeigen $f = g$. Im Komplexen kennen wir noch keinen Satz, aus dem dieses Ergebnis routinemäßig folgt, oder?



Taylorapproximation der geometrischen Reihe $z/(1-z) = z + z^2 + z^3 + \dots$
Links oben ist die abgebildete Figur im Innern des Konvergenzkreises der Reihe gezeigt. Rechts oben ist das Bild unter $z \to z/(1-z)$ dargestellt; da es sich um eine Möbius-Transformation handelt, sind alle Netzkurven Kreise, vgl. S. 63. In den unteren beiden Bildern wird die Figur durch das 3. bzw. 5. Taylorpolynom abgebildet; das richtige Bild ist punktiert dazu gezeichnet. Der Entwicklungspunkt ist markiert (o). – Die Bilder zeigen, daß man auch in dieser zweidimensionalen Situation die Approximation veranschaulichen kann, obwohl man nicht auf die gewohnte Funktionsdarstellung durch Graphen zurückgreifen kann.

Riemannsummen, Integrale, Stammfunktionen

Tangenten mancher Kurven sind schon im Altertum betrachtet worden und ge wisse anschauliche Vorstellungen von der “Steigung” eines Bergpfades hat sicher jeder Wanderer. Im Vergleich dazu sind Integrale ein völlig neuartiges Konzept. Sie erlauben z.B. aus der Geschwindigkeit einer Bewegung die Bahnkurve durch eine Art “Summation” zu rekonstruieren. Wir definieren die Integrale mit Hilfe von Riemannsummen und interpretieren die Definition als “kontinuierliche Summation”. Der Hauptsatz der Differential- und Integralrechnung besagt, daß die Integration Umkehrung der Differentiation ist, daß man Integrale mit Hilfe von Stammfunktionen berechnen kann.

Das Konzept der Stammfunktion ist einfach: F heiß Stammfunktion von f, falls $F' = f$ gilt. Wir benutzen diesen Begriff für reelle oder komplexe Funktionen, aber auch für Kurven $f : I \to \mathbb{R}^3$ (meist wird dann f als Geschwindigkeit von F interpretiert). Wir wissen schon, daß Potenzreihen Stammfunktionen haben, und wir wissen vor allem, daß sich $F(b) - F(a)$ wenig von Riemann-Summen von f im Intervall $[a, b]$ unterscheidet. Dieser Sachverhalt wird jetzt zur Integralrechnung ausgebaut. Vorher beweisen wir den bisher aufwendigsten Existenzsatz, der auf andere Weise als die Potenzreihen neue Funktionen liefert:

Existenzsatz für Stammfunktionen. L-dehnungsbeschränkte Abbildungen $f : I \to \mathbb{R}$ (oder $f : I \to \mathbb{R}^2, \mathbb{R}^d$) haben Stammfunktionen $F : I \to \mathbb{R}$ (oder $\mathbb{R}^2, \mathbb{R}^d$), d.h. $F' = f$.

Bemerkung. Wir werden sehen, daß dieser Satz den noch fehlenden Teil des Hauptsatzes der Differential- und Integralrechnung enthält.

Beweis. Wegen der Voraussetzung: $x, y \in I \Rightarrow |f(y) - f(x)| \leq L \cdot |y - x|$ können wir f stückweise linear mit gleichmäßiger Fehlerkontrolle approximieren:
Teile das Intervall $I = [x_0, x_n]$ in n gleiche Teile $x_0 < x_1 < \ldots < x_n$, mit Abständen $x_j - x_{j-1} = (x_n - x_0)/n = |I|/n$ und definiere die folgende “**S**ehnen**a**pproximationen” sa_n von f:

$$x \in [x_{j-1}, x_j] \Rightarrow sa_n(x) := \frac{(x_j - x) \cdot f(x_{j-1}) + (x - x_{j-1}) \cdot f(x_j)}{x_j - x_{j-1}}.$$

Alle Steigungen von sa_n sind Sehnensteigungen von f , also durch L beschränkt:

$$x, y \in I \to |sa_n(y) - sa_n(x)| \leq L \cdot |y - x|.$$

Nun hat die Differenz $f - sa_n$ die Dehnungsschranke $2L$, ferner stimmen f und sa_n an allen Teilpunkten überein, $f(x_j) = sa_n(x_j)$ $(j = 0, \ldots, n)$, also $\min_j |x - x_j| \leq |x_n - x_0|/2n$.

Daher folgt für die maximale Abweichung zwischen f und sa_n:

$$x \in I \Rightarrow |f(x) - sa_n(x)| \leq 2L \cdot \min_j |x - x_j| \leq \frac{L \cdot |I|}{n}.$$

Wir nennen diese Eigenschaft:

"f ist gleichmäßig stückweise linear approximierbar".

Für die Funktionen sa_n können wir explizit stückweise **q**uadratische **S**tammfunktionen $SQ_n,\ SQ_n' = sa_n$ angeben, zunächst in jedem Teilintervall:

$$x \in [x_{j-1}, x_j] \Rightarrow SQ_n(x) := C_{j-1} + \frac{1}{2} \frac{(x - x_{j-1})^2 \cdot f(x_j) - (x_j - x)^2 \cdot f(x_{j-1})}{x_j - x_{j-1}}.$$

Schließlich setzen wir z. B. $C_0 = 0$ und wählen C_1 so, daß die beiden Definitionen von $SQ_n(x_1)$ auf $[x_0, x_1]$ und auf $[x_1, x_2]$ denselben Wert ergeben. Das wird wiederholt: Sind die Funktionswerte $C_0, C_1, \ldots, C_{j-1}$ an den Teilpunkten $x_0, ..., x_{j-1}$ schon gewählt, so wird C_j so gewählt, daß die Definitionen von $SQ_n(x_j)$ auf $[x_{j-1}, x_j]$ und auf $[x_j, x_{j+1}]$ übereinstimmen. Natürlich gilt jetzt

$$x \in I \Rightarrow SQ_n(x_0) = 0\ ,\ SQ_n'(x) = sa_n(x).$$

Aus dem Monotoniesatz und aus $x \in I \Rightarrow |sa_n(x) - sa_m(x)| \leq \left(\frac{1}{n} + \frac{1}{m}\right) \cdot L \cdot |I|$ folgt weiter, daß die $\{SQ_n(x)\}$ für $x \in I$ Cauchyfolgen bilden:

$$\begin{aligned} x \in I \Rightarrow |SQ_n(x) - SQ_m(x)| &\leq \max |\big(SQ_n - SQ_m\big)'(x)| \cdot |x - x_0| \\ &\leq \left(\frac{1}{n} + \frac{1}{m}\right) \cdot L \cdot |I|^2. \end{aligned}$$

Wegen der Vollständigkeit konvergiert $\{SQ_n(x)\}$ gegen die Werte einer *Grenzfunktion* $x \mapsto SQ_\infty(x)$. Wieder wegen des Monotoniesatzes (für stückweise quadratische Funktionen mit $|SQ_n'(x) - sa_n(a)| \leq L|x - a|$) gilt:

$$a, x \in I \Rightarrow |SQ_n(x) - SQ_n(a) - sa_n(a) \cdot (x - a)| \leq L \cdot |x - a|^2.$$

Daraus folgt mit dem Archimedes-Argument (andere Worte dafür sind: mit $\lim_{n \to \infty}$)

$$x, y \in I \Rightarrow |SQ_\infty(y) - SQ_\infty(x) - f(x) \cdot (y - x)| \leq L \cdot |y - x|^2.$$

Daher erfüllt die Grenzfunktion SQ_∞ in I die Differenzierbarkeitsdefinition mit der Ableitung $SQ_\infty'(x) = f(x)$, wir haben also eine Stammfunktion für f konstruiert.

Die nächste Aufgabe ist, das Integral einer Funktion mit Hilfe ihrer Riemannsummen zu definieren. Dazu muß präzisiert werden, daß alle Riemannsummen zu *genügend feinen Unterteilungen* so nahe bei einander liegen, daß nur eine einzige Zahl als deren "verallgemeinerter Grenzwert" angesehen werden kann.

Definition des Integrals. Sei $f : I \to \mathbb{R}^1$ eine Funktion. Sei $x_0 < x_1 < \ldots < x_n$ eine Unterteilung U des Intervalls $I = [a, b] = [x_0, x_n]$. Wir definieren die *Feinheit* dieser Unterteilung durch

$$\delta(U) := \max\{|x_j - x_{j-1}|;\ j = 1, \ldots, n\}.$$

Schließlich sei W_U die Menge der Werte von Riemann-Summen von f für die Unterteilung U, also

$$w \in W_U \Leftrightarrow w = \sum_{j=1}^{n} f(\tau_j) \cdot (x_j - x_{j-1}) \text{ mit } \tau_j \in [x_{j-1}, x_j].$$

Die Funktion f heißt *Riemann-integrierbar* auf $I = [a, b]$ falls gilt:

$$\begin{aligned}\lim_{n\to\infty} \sup\{W_U\ ;\ \delta(U) \le 1/n\} &=: \limsup_{\delta(U)\to 0} W_U \\ &= \liminf_{\delta(U)\to 0} W_U := \lim_{n\to\infty} \inf\{W_U\ ;\ \delta(U) \le 1/n\}.\end{aligned}$$

Diese Bedingung besagt, daß es nur eine reelle Zahl gibt, nämlich $\limsup_{\delta(U)\to 0} W_U = \liminf_{\delta(U)\to 0} W_U$, die von Riemannsummen zu immer feineren Unterteilungen approximiert wird. Wir nennen sie "Integral von f über $[a, b]$" und schreiben $\int_I f$ oder $\int_a^b f(x)dx$:

Integral von f über $[a, b]$: $$\int_I f := \int_a^b f(x)dx := \limsup_{\delta(U)\to 0} W_U = \liminf_{\delta(U)\to 0} W_U.$$

Eine Funktion $f = (f^1, f^2, \ldots, f^d) : I \to \mathbb{R}^d$ heißt integrierbar, falls alle Komponentenfunktionen f^k integrierbar sind. Wir integrieren komponentenweise:

$$\textstyle\int_I f := (\int_I f^1, \int_I f^2, \ldots, \int_I f^d).$$

Auch $\int_I f$ kann durch Riemannsummen approximiert werden, denn mit dem Satz von Pythagoras kann man die Approximationsfehler der Komponenten zusammenfassen:

> Zu jeder Fehlerschranke $1/n$ gibt es eine Feinheitsschranke δ_n so daß gilt:
>
> Sei U eine Unterteilung mit Feinheit $\delta(U) \le \delta_n$
>
> und sei $\mathfrak{RS}(f) = \sum_k f(\tau_k)(t_k - t_{k-1})$ eine Riemannsumme zur Unterteilung U,
>
> dann folgt: $$\left|\int_I f - \mathfrak{RS}(f)\right| \le \frac{1}{n}.$$

Unser Existenzsatz für Stammfunktionen und unsere Abschätzung von Riemannsummen von f durch eine Stammfunktion F von f liefert nun den

Hauptsatz der Differential- und Integralrechnung.
L-dehnungsbeschränkte $f : I = [a,b] \to \mathbb{R}^d$ sind Riemann integrierbar; f besitzt eine Stammfunktion $F : I \to \mathbb{R}^d,\ F' = f$, und es gilt

$$\int_I f := \int_a^b f(x)dx = F(b) - F(a).$$

Beweis. Mit F bezeichnen wir die nach dem vorhergehenden Existenzsatz konstruierte Stammfunktion F von f. Aus der Dehnungsschranke L für f wurde in diesem Existenzsatz bewiesen:

$$x, y \in I \Rightarrow |F(y) - F(x) - f(x) \cdot (y - x)| \leq L \cdot |y - x|^2.$$

Wähle ein $n \in \mathbb{N}$ beliebig. Dann gilt für jede Einteilung U mit Feinheit $\delta(U) \leq 1/n$ und *jede* Riemannsumme $\mathfrak{RS}(f) = \sum_{j=1}^n f(\tau_j) \cdot (x_j - x_{j-1})$ zu dieser Einteilung nach dem Satz am Ende des Kapitels zum Monotoniesatz

$$|F(b) - F(a) - \mathfrak{RS}(f)| \leq \frac{L \cdot |I|}{n}.$$

Damit ist die Differenz zwischen $F(b) - F(a)$ einerseits und beliebigen Riemannsummen zu Einteilungen mit Feinheit $\leq 1/n$ durch eine *Nullfolge* abgeschätzt; das Integrierbarkeitskriterium ist also erfüllt und $F(b) - F(a) = \limsup_{\delta(U)\to 0} W_U = \liminf_{\delta(U)\to 0} W_U$ ist festgestellt.

Kommentar. Die Differentialrechnung als Kalkül zur Bestimmung von Tangenten und die Integralrechnung als verallgemeinerte Summation sind unabhängig von einander entwickelt worden. Es war eine große Entdeckung, daß die Integration als Umkehrung der Differentiation aufgefaßt werden kann. Da der Vorrat an Funktionen mit bekannten Stammfunktionen von Beginn an groß war, gestattete diese Entdeckung die explizite Berechnung vieler Integrale. Natürlich war das ein großer Erfolg, weil die Definition des Integrals ja nicht direkt zu einem Berechnungsverfahren führt (sondern nur zur Approximation).

Eine etwas **andere Formulierung des Hauptsatzes** ist ebenfalls wichtig:
Hat eine Funktion $F : I \to \mathbb{R}^d$ eine dehnungsbeschränkte Ableitung $f = F' : I \to \mathbb{R}^d$, so kann man F aus f und dem Anfangswert $F(a)$ durch Integration rekonstruieren:

$$F(x) - F(a) = \int_a^x f(t)dt.$$

Die Voraussetzung "dehnungsbeschränkt" wird im nächsten Abschnitt zu "stetig" abgeschwächt; der Hauptsatz der Differential- und Integralrechnung ist allerdings knapp 200 Jahre älter als die Stetigkeit.

Zur Interpretation des Integrals stellen wir als erstes Eigenschaften gegenüber, die die große formale Ähnlichkeit zwischen Summen und Integralen hervorheben:

1.) $$\sum_k(\alpha \cdot a_k + \beta \cdot b_k) \;=\; \alpha \cdot \sum_k a_k + \beta \cdot \sum_k b_k, \qquad \text{linear}$$

$$\int(\alpha \cdot f(x) + \beta \cdot g(x))dx = \alpha \cdot \int f(x)dx + \beta \cdot \int g(x)dx. \qquad \text{linear}$$

2.) $$\sum_{k=1}^{m} a_k + \sum_{k=m+1}^{n} a_k = \sum_{1}^{n} a_k, \qquad \text{assoziativ}$$

$$\int_a^b f + \int_b^c f \;=\; \int_a^c f, \qquad \text{intervalladditiv}$$

3.) $$a_k = b_{k+1} - b_k \Rightarrow \sum_{k=1}^{n} a_k = b_{n+1} - b_1, \qquad \text{Teleskopsumme}$$

$$f = F' \Rightarrow \int_a^b f(x)dx = F(b) - F(a), \qquad \text{Hauptsatz}$$

4.) $$a_k \le b_k \Rightarrow \sum a_k \le \sum b_k, \qquad \text{Monotonie}$$

$$f \le g \;\Rightarrow \int_a^b f \;\le \int_a^b g, \qquad \text{Monotonie}$$

5.) $$|\sum a_k| \;\le\; \sum |a_k|, \qquad \text{verallgemeinerte Dreiecksungleichung}$$

$$|\int_a^b f(x)dx| \le \int_a^b |f(x)|dx, \qquad \text{kontinuierliche Dreiecksungleichung.}$$

Beweis der kontinuierlichen Dreiecksungleichung für Funktionen $f : I \to \mathbb{R}^d$ unter Verwendung irgendeiner Norm $|.|$ auf $\mathbb{R}^d$. Wähle — zu jedem $n \in \mathbb{N}$ — die Feinheit von Unterteilungen U des Intervalls $I = [a, b]$ so klein, daß für alle so feinen Unterteilungen die Riemann-Summen zu den beiden Integralen um höchstens $1/n$ von den jeweiligen Integralen verschieden sind. Für die Riemann-Summen gilt die Dreiecksungleichung:

$$\Big|\sum_{j=1}^{n} f(\tau_j) \cdot (x_j - x_{j-1})\Big| \le \sum_{j=1}^{n} |f(\tau_j)| \cdot (x_j - x_{j-1}),$$

daher folgt für die Integrale (und das Archimedes-Axiom beseitigt den Fehler $2/n$):

$$\Big|\int_a^b f(x)dx\Big| \le \int_a^b |f(x)|dx + \frac{2}{n} \quad \text{für jedes} \quad n \in \mathbb{N}.$$

Numerische Integralberechnungen. Die Analogie zwischen Summen und Integralen wird durch numerische Berechnungsformeln nachdrücklich unterstrichen, so daß auch dadurch Integrale anschaulicher werden. Der Integraldefinition entsprechend wird das Integrationsintervall I in kleine Teilintervalle aufgeteilt. Die Hauptidee besteht darin, daß die Integrale *differenzierbarer* Funktionen f sich auf kleinen Intervallen schon durch sehr einfache Formeln so gut approximieren lassen, daß die Summe über alle Teilintervalle immer noch eine gute Approximation für $\int_I f$ liefert. Die beiden einfachsten Approximationen bestehen darin, daß man (i) das Integral der Sehne zwischen den Endpunkten berechnet und (ii) das Integral der Tangente in der Intervallmitte berechnet:

Sehnentrapez: $$\int_a^b f(x)dx \sim \frac{f(a)+f(b)}{2}(b-a)$$

Tangententrapez: $$\int_a^b f(x)dx \sim f(\frac{a+b}{2})(b-a)$$

Wie gut sind diese Approximationen? Im Kapitel Monotoniesatz hatten wir aus Schranken für die zweite Ableitung die Unterschiede zwischen f und der Tangente, bzw. zwischen f und der Sehne abgeschätzt:

Tangente in c: $\quad T_c(x) := f(c) + f'(c)(x-c),$

Sehne zwischen a, b: $\quad S_{ab}(x) := \big(f(a)(b-x) + f(b)(x-a)\big)/(b-a).$

Voraussetzung: $\quad a \le c, x \le b, \quad 0 \le f'' \le B.$

Monotoniesatz $\Rightarrow$ $$0 \le f(x) - T_c(x) \le \frac{1}{2}B(x-c)^2,$$
$$0 \le S_{ab}(x) - f(x) \le \frac{1}{2}B(x-a)(b-x).$$

Die *Fehlerabschätzung* besteht nun einfach darin, die Monotonie des Integrals auszunutzen und diese Ungleichungen zu integrieren:

Voraussetzung: $\quad 0 \le f'' \le B, \quad c := \frac{a+b}{2}.$

Tangententrapez: $$0 \le \int_a^b f(x)dx - f(c)(b-a) \le \frac{B}{24}(b-a)^3,$$

Sehnentrapez: $$0 \le \frac{f(a)+f(b)}{2}(b-a) - \int_a^b f(x)dx \le \frac{B}{12}(b-a)^3.$$

Aus der Voraussetzung $|f''| \le B$ erhält man die entsprechenden Betragsungleichungen. Aber es ist nützlich zu wissen: Solange die zweite Ableitung das Vorzeichen nicht wechselt, ist die eine Näherung zu klein, die andere zu groß. Daher stellt sich die Frage: Kann man die beiden Approxmationen geschickt mitteln, um eine wesentlich bessere zu erhalten? Wenn die Fehlerabschätzungen gut sind, ist die Sehnentrapez-Näherung nur halb so gut wie die andere. Das stimmt mit Archimedes Ergebnis überein, daß der Flächeninhalt zwischen Parabel und Tangente nur halb so groß ist wie der zwischen Parabel und Sehne. Ein erster Versuch sollte also in einer 1:2-Mittelung bestehen. Wir finden die

Simpsonregel: $$\int_a^b f(x)dx \sim \big(\frac{f(a)+f(b)}{2} + 2\cdot f(\frac{a+b}{2})\big)\frac{(b-a)}{3}.$$

Probiert man sie an der quadratischen Parabel aus, so findet man den richtigen Wert:

$$\int_a^b x^2 dx = \frac{1}{3}(b^3-a^3) = \big(\frac{a^2+b^2}{2} + 2(\frac{a+b}{2})^2\ \big)\frac{(b-a)}{3}.$$

Tatsächlich liefert die Simpsonregel sogar für jedes *kubische* Polynom P den richtigen Wert, denn wenn c der Intervallmittelpunkt ist, können wir schreiben $x^3 = (x-c+c)^3 = (x-c)^3 +$ quadratischer Rest. Der quadratische Rest wird durch die Simpsonregel richtig integriert und das Integral von $(x-c)^3$ ist 0, genau wie seine Simpson-Approximation. Wir können daher sagen: Die Simpsonregel berechnet das Integral desjenigen kubischen Polynoms P, das mit f in den drei Werten $f(a), f(\frac{a+b}{2}), f(b)$ und der Ableitung $f'(\frac{a+b}{2})$ übereinstimmt. Eine Fehlerabschätzung kommt daher, daß eine Schranke für die vierte Ableitung, etwa $|f^{(4)}| \le C$, den Unterschied zwischen f und dem kubischen Polynom P kontrolliert, und mit dem Integral dieses Unterschiedes ist der Fehler der Simpsonregel abgeschätzt. Der Beweis dieser sehr häufig benutzten Fehlerabschätzung ist ein etwas komplizierteres Beispiel für Argumentationen mit dem Monotoniesatz. Ich vereinfache die Bezeichnungen durch die Annahme $a = -b$. Über die Funktion $h := f - P$ haben wir dann folgende

Voraussetzung: $\quad h(\pm a) = 0 = h(0),\ h'(0) = 0,\ |h^{(4)}| \le C,$

Behauptung: $\quad -a \le x \le a \ \Rightarrow\ |h(x)| \le \dfrac{C}{24}x^2(a^2-x^2),$

Simpsonfehler: $$\left|\int_a^b f(x)dx - \frac{1}{6}\left(f(a) + 4f(\frac{a+b}{2}) + f(b)\right)(b-a)\right| \le \frac{C}{2880}(b-a)^5.$$

Beweis. Wie bei den ersten Monotoniesatzanwendungen ist ein Spezialfall, nämlich

$$h(\pm a) = 0 = h(0),\ h'(0) = 0,\ 0 \le h^{(4)} \quad\Rightarrow\quad h \le 0,$$

die Hauptaussage. Ist das bewiesen und $|h^{(4)}| \le C$ gegeben, so erfüllen $g_+(x) := +h(x) - \frac{C}{24}x^2(a^2-x^2)$, $g_-(x) := -h(x) - \frac{C}{24}x^2(a^2-x^2)$ die Voraussetzung $0 \le g_\pm^{(4)}$ und aus $g_\pm \le 0$ folgt die Behauptung für $|h|$. Danach folgt die Simpson-Abschätzung mit Hilfe der Stammfunktion $(a^2x^3/3 - x^5/5)' = x^2(a^2-x^2)$.

Nun zur Hauptaussage. Aus $0 \le h^{(4)}$ haben wir als erste Anwendungen des Monotoniesatzes schon gefolgert, daß h'' oberhalb jeder Tangente und unterhalb jeder Sehne liegt. Daraus schließen wir indirekt $h''(0) \le 0$, denn wäre $h''(0) > 0$ so wäre die Tangente in 0 an h'' mindestens in einem der Intervalle $[-a, 0], [0, a]$ strikt positiv, also h'' auch (nämlich oberhalb der Tangente); daher wäre h' in wenigstens einem dieser Intervalle streng wachsend, und wegen $h'(0) = 0$ folgt daraus, daß auch h in wenigstens einem der Intervalle streng monoton wäre (links fallend,rechts wachsend), im Widerspruch zu $h(-a) = h(0) = h(a)$.
Für das Ende des Beweises zitiere ich aus dem nächsten Kapitel, weil die noch zu beweisenden Sätze die hier benötigte Argumentation sehr viel übersichtlicher machen. Falls die Funktion h positive Werte hätte, so hätte sie auch ein positives Maximum, etwa $h(c)$. Notwendige Bedingung für ein Maximum ist (neben $h'(c) = 0$) auch $h''(c) \le 0$. Da h'' unterhalb jeder Sehne bleibt, ist $h'' \le 0$ auf dem ganzen Intervall zwischen 0 und c. Wie am Anfang des Beweises folgt, daß h zwischen 0 und c unterhalb der (horizontalen) Tangente in 0 bleibt, aber das steht im Widerspruch zu $h(c) > 0$.
Zum Vergleich führe ich den Beweis auch allein mit den bewiesenen Hilfsmitteln zu Ende. Angenommen, h hätte positive Werte, etwa $h(x) > 0$ und z.B. $0 < x < a$, dann müßten in $[0, x]$ positive Werte von h' und in $[x, a]$ negative Werte von h' vorkommen. Damit h' von $h'(0) = 0$ auf positive Werte wachsen kann, muß h'' irgendwo in $[0, x]$ positive Werte haben, z.B. $h''(c) > 0, c \in (0, x)$; und damit h' von seinen positiven Werten wieder auf negative Werte (spätestens in $[x, a]$) fallen kann, muß h'' *rechts* von seinen positiven Werten, rechts von c, ebenfalls negative Werte haben, z.B. $h''(d) < 0, d \in (c, a)$. Das ist der gewünschte Widerspruch, denn da h'' ja unterhalb seiner Sehnen liegt, kann h'' in $(0, d\,]$ nur Werte < 0 haben und nicht $h''(c) > 0$.

Weitere Interpretationen von Integralen. Stellt man sich, für Funktionen $f : [a, b] \to \mathbb{R}$, jeden Summanden $f(\tau_j) \cdot (x_j - x_{j-1})$ einer Riemann-Summe für f als Rechteck vor, mit dem Intervall $[x_{j-1}, x_j]$ auf der x-Achse als horizontaler Seite und der Höhe $f(\tau_j)$ darüber, so erwartet man, daß das Integral den Flächeninhalt zwischen x-Achse und Graph von f ausrechnet. Da wir im Augenblick noch keine *Definition* des Flächeninhalts krummlinig begrenzter Figuren geben können, können wir dieses anschauliche Bild nicht zu einem Satz ausbauen. Das gelingt aber für 1-dimensionale “Inhalte”:

Berechnung der Bogenlänge mit Integralen: Länge $(p([a, x]) = \int_a^x |p'(t)|dt$.

Wir hatten die Bogenlänge dehnungsbeschränkter Kurven $p : [a, b] \to \mathbb{R}^d$ (damals $d = 2$) *definiert* als

Länge $(p([a, b])) := \sup\{\sum_{k=1}^{N} |p(x_k) - p(x_{k-1})|; \ldots$

$\ldots\ a = x_0 < x_1 < \ldots < x_N = b$ Einteilung von $[a, b]$ $\}$.

Wir setzen nun zusätzlich voraus, daß $p(\ )$ eine differenzierbare Kurve mit (z. B.) dehnungsbeschränkter Ableitung $p' : [a,b] \to \mathbb{R}^d$ ist. Dann gilt nach dem Hauptsatz der Differential- und Integralrechnung und der kontinuierlichen Dreiecksungleichung:

$$|p(x_k) - p(x_{k-1})| = \Big| \int_{x_{k-1}}^{x_k} p'(x)dx \Big| \le \int_{x_{k-1}}^{x_k} |p'(x)|dx,$$

also

$$\sum_{k=1}^{N} |p(x_k) - p(x_{k-1})| \le \int_a^b |p'(x)|dx.$$

Das Integral ist damit eine obere Schranke für die Länge beliebiger Sehnenzüge, und wir erhalten nach Definition der Bogenlänge als Supremum:

$$\text{Länge } (p[a,b]) \le \int_a^b |p'(x)|dx.$$

Sowohl die Länge wie das Integral betrachten wir als Funktionen des Intervallendes:

$$L(t) := \text{Länge } (p([a,t])), \quad F(t) := \int_a^t |p'(x)|dx,$$

und wir wollen die Gleichheit dieser Funktionen zeigen. Wegen $L(a) = F(a) = 0$ würde dazu $L'(t) = F'(t)$ genügen. Die schon bewiesene Ungleichung schreiben wir um als Ungleichung zwischen Differenzenquotienten:

$$\frac{|p(t_2) - p(t_1)|}{t_2 - t_1} \le \frac{L(t_2) - L(t_1)}{t_2 - t_1} = \frac{\text{Länge } (p([t_1,t_2]))}{t_2 - t_1} \le \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} |p'(t)|dt.$$

Wegen der Differenzierbarkeit von $p(\ )$ ist der Differenzenquotient ganz links auf kleinen Intervallen um t_1 wenig von $p'(t_1)$ verschieden; wegen der Dehnungsschranke von $p'(\ )$ ist der Term ganz rechts wenig von $|p'(t_1)|$ verschieden. Genauer, wir finden zu jedem $n \in \mathbb{N}$ ein $\delta = \delta(n, t_1)$, so daß gilt

$$t_2 \in [t_1 - \delta, t_1 + \delta] \Rightarrow \Big| \frac{p(t_2) - p(t_1)}{t_2 - t_1} - p'(t_1) \Big| \le \frac{1}{n}, \quad \Big| \frac{1}{t_2 - t_1} \cdot \int_{t_1}^{t_2} |p'(t)|dt - |p'(t_1)| \Big| \le \frac{1}{n},$$

also wegen der Dreiecksungleichung

$$t_2 \in [t_1 - \delta, t_1 + \delta] \Rightarrow |p'(t_1)| - \frac{1}{n} \le \frac{L(t_2) - L(t_1)}{t_2 - t_1} \le |p'(t_1)| + \frac{1}{n}.$$

Damit ist $L(\ )$ differenzierbar mit der Ableitung $L'(t_1) = |p'(t_1)|$ für jedes $t_1 \in [a,b]$. Es folgt $F = L$ und wir haben bewiesen:

$$\text{Länge } (p([a,x]) = \int_a^x |p'(t)|dt.$$

Dies Ergebnis formulieren wir noch einmal mit anderen Worten. Wir interpretieren den Parameter t der Kurve $t \to p(t)$ als "Zeit", also zum Zeitpunkt t befindet sich ein sich bewegender Punkt p an der Stelle $p(t)$. Dann besitzt $|p'(t)|$ die Interpretation (Betrag der) "Momentangeschwindigkeit", und die Länge $L(t)$ ist der bis zum Zeitpunkt t zurückgelegte Weg. In Formeln:

$$\text{Zurückgelegter Weg } = L(t) = \int_0^t |p'(\tau)| d\tau = \text{ Integral der Geschwindigkeit.}$$

Setzen wir die Definition des Integrals mit Hilfe der Riemann-Summen noch einmal ein, so besagt diese Gleichung, daß der zurückgelegte Weg dadurch erhalten wird, daß zunächst die Riemann-Summen zu den Zeitabschnitten $[\tau_{k-1}, \tau_k]$ und den zugehörigen Durchschnittsgeschwindigkeiten $|p(\tau_k) - p(\tau_{k-1})|/(\tau_k - \tau_{k-1})$ betrachtet werden; die zurückgelegten Weglängen, genauer die Durchschnittsstrecken $|p(\tau_k) - p(\tau_{k-1})|$, werden zur Riemann-Summe aufsummiert, und das Integral ist der Grenzwert dieser Weglängen. — Die Umgangssprache besitzt eigentlich keine Formulierungen für das, was hier von dem Integral geleistet wird. Ich finde nun, daß unser Vergleich von Riemann-Summen und Integral ein genügend anschauliches Bild vermittelt, um damit den folgenden (umgangssprachlichen) Worten einen Sinn zu geben:

> Das Integral $\int_a^t |p'(\tau)| d\tau$ "summiert kontinuierlich" die Wegbeiträge, die im Laufe der Zeit durch die Bewegung $t \to p(t)$ mit der Momentangeschwindigkeit $|p'(t)|$ zurückgelegt werden.

Mittelwerteigenschaft von Integralen

Zunächst erinnere ich daran, daß der Schwerpunkt S von Punkten P_k mit der Position $p_k \in \mathbb{R}^3$ und der Masse m_k als Mittelwert berechnet wird:

$$\text{Schwerpunkt } S = \sum_{k=1}^{N} \frac{m_k}{M} \cdot p_k, \quad \text{Gesamtmasse } M = \sum m_k \ .$$

Wir wollen für eine Funktion $f : [a, b] \to \mathbb{R}^3$ das Integral $\frac{1}{b-a} \int_a^b f(t) dt$ als Mittelwert interpretieren. Wir betrachten Riemann-Summen $\sum_{k=1}^N f(\tau_k) \cdot (t_k - t_{k-1})/(b-a)$. Diese können wir offenbar als Mittelwert der $f(\tau_k)$ mit den Massen $(t_k - t_{k-1})/(b - a)$ interpretieren. Jede Verfeinerung der Einteilung für die Riemann-Summen bezieht immer mehr Funktionswerte $f(\tau_k)$ in diese Mittelung ein. Daher gibt die umgangssprachliche Formulierung:

$$\int_a^b f(t) \frac{dt}{b-a} \text{ ist ein kontinuierlicher Mittelwert der Werte } f(t)$$

eine brauchbare Umschreibung des Sachverhaltes.

Stetigkeit und gleichmäßige Konvergenz

Gegen Ende des 19. Jahrhunderts hat sich die Stetigkeit als ein Grundbegriff der Analysis durchgesetzt, etwa 200 Jahre nach Beginn der Differential- und Integralrechnung. Ich kann den Gründen für diese Entwicklung hier nicht nachgehen, ich muß mich mit einfacheren als den historischen Motivationen begnügen. Wegen der großen Bedeutung, die konvergente Folgen schon in unserem bisherigen Aufbau gehabt haben, ist eine naheliegende Frage, die zu den stetigen Funktionen führt, die folgende: *Welche Funktionen bilden jede in ihrem Definitionsbereich konvergente Folge auf eine konvergente Bildfolge ab*? Alle Beweise zu Hauptsätzen über stetige Funktionen brechen ohne die Vollständigkeit zusammen. Zuerst bespreche ich die Sätze, die sich leicht mit der Folgen-Stetigkeit zeigen lassen. Dann komme ich zur ϵ-δ-Stetigkeit und den übrigen Hauptsätzen. Schließlich erkläre ich die Konvergenzeigenschaften stetiger Funktionenfolgen. Den Abschluß bilden berühmte Beispiele stetiger Funktionen von Cantor, Weierstraß und Hilbert.

Definition der Stetigkeit. Eine Funktion f heißt *folgen-stetig bei* a, falls jede in ihrem Definitionsbereich gegen a konvergente Folge $\{a_k\}$, $\lim a_k = a$, auf eine gegen $f(a)$ konvergente Folge $\{f(a_k)\}$, $\lim f(a_k) = f(a)$, abgebildet wird. Die Funktion f heißt stetig, wenn dies für jeden Punkt a des Definitionsbereichs gilt.

Wir benutzen diese Definition für reelle oder komplexe Funktionen, für Kurven $c : \mathbb{R} \to \mathbb{R}^d$, für Abbildungen $F : \mathbb{R}^d \to \mathbb{R}^e$ und im weiteren Verlauf in immer allgemeineren Situationen. Natürlich sind dehnungsbeschränkte Abbildungen stetig: $|F(x) - F(y)| \le L \cdot |x - y|$ und $\lim a_k = a$ liefert $|F(a_k) - F(a)| \le L \cdot |a_k - a|$, also $\lim F(a_k) = F(a)$. Die Definition hat unmittelbar zur Folge:

Satz. Summen, Produkte und Kompositionen stetiger Funktionen sind stetig.

Was man sich darüber hinaus unter stetigen Funktionen vorzustellen hat, wird erst klarer, wenn wir leistungsfähige Konstruktionsmittel haben, mit denen am Ende dieses Abschnitts interessante Beispiele konstruiert werden. Die ersten drei Sätze allerdings handeln von so anschaulichen Eigenschaften stetiger Funktionen, daß sie oft benutzt werden, ohne daß ihre Beweisnotwendigkeit wahrgenommen wird. Zunächst der

Zwischenwertsatz. $f : [a, b] \to \mathbb{R}$ sei stetig und $w \in \mathbb{R}$ mit $f(a) < w < f(b)$. Dann gibt es $c \in (a, b)$ mit $f(c) = w$. (Es genügt, diesen Satz für $w = 0$ zu beweisen: betrachte $f - w$.)

Beweis. Wir konstruieren eine Intervallschachtelung. Der Anfang ist $a_1 := a$, $b_1 := b$. Nun seien $[a_1, b_1] \supset [a_2, b_2] \supset \cdots \supset [a_n, b_n]$ schon konstruiert mit $f(a_n) \le 0$, $f(b_n) \ge 0$.

Betrachte $m := (a_n + b_n)/2$. Ist $f(m) \geq 0$, so setze $a_{n+1} = a_n$, $b_{n+1} = m$, andernfalls setze $a_{n+1} = m$, $b_{n+1} = b_n$. Aus der Stetigkeit folgt, daß f im Grenzpunkt c dieser Intervallschachtelung $\{[a_n, b_n]\}$ eine Nullstelle hat:

$$0 \geq \lim f(a_n) = f(c) = \lim f(b_n) \geq 0.$$

Aufgabe. Benutze eine andere Formulierung der Vollständigkeit, um diesen Satz zu beweisen.

Funktionen, die gleichmäßige Dehnungsschranken haben (z.B. Polynome), sind auf beschränkten Intervallen (a, b) beschränkt durch $f((a + b)/2) \pm L \cdot (b - a)/2$. Die rationale Funktion $f(x) = 1/x$ ist auf dem Intervall $(0, 1)$ offenbar stetig, aber nicht beschränkt. Derartige Unglücke können nur am Rande passieren, denn es gilt der

Beschränktheitssatz. $f : [a, b] \to \mathbb{R}$ sei stetig. Dann ist f beschränkt.

Beweis. Angenommen f habe keine Schranke auf $[a, b]$, dann konstruieren wir einen Widerspruch mit Hilfe einer Intervallschachtelung $\{[a_n, b_n]\}$, auf deren Intervallen f nicht beschränkt ist. Anfang: $a_1 = a$, $b_1 = b$. Nun sei $[a_1, b_1] \supset \cdots \supset [a_n, b_n]$ schon so konstruiert, daß f auf diesen Intervallen nicht beschränkt ist und zusätzlich $\max(f(a_n), f(b_n)) > n$ gilt. Betrachte $m = (a_n + b_n)/2$. Falls f auf $[a_n, m]$ nicht beschränkt ist, setze $a_{n+1} = a_n$, $b_{n+1} = m$; andernfalls ist f auf $[m, b_n]$ nicht beschränkt, dann setze $a_{n+1} = m$, $b_{n+1} = b_n$. Weil f auf diesem Intervall nicht beschränkt ist, gibt es $x \in (a_{n+1}, b_{n+1})$ mit $f(x) > n + 1$ und wir ersetzen noch a_{n+1} oder b_{n+1} durch x, so daß weiterhin f auf $[a_{n+1}, b_{n+1}]$ nicht beschränkt ist. Diese Intervallschachtelung hat einen Grenzpunkt $c \in [a, b]$. Wegen der Stetigkeit gilt $\lim f(a_n) = f(c) = \lim f(b_n)$, im Widerspruch zu $\max(f(a_n), f(b_n)) > n$.

Fast ebenso beweist man den folgenden sehr häufig zitierten Satz, den

Satz vom Maximum. $f : [a, b] \to \mathbb{R}$ sei stetig. Dann nimmt f sein Maximum an, d. h. es gibt ein $c \in [a, b]$ mit $f(c) \geq f(x)$ für alle $x \in [a, b]$.

Beweis. Da f auf $[a, b]$ schon als beschränkt nachgewiesen ist, können wir definieren

$$S := \sup\{f(x); x \in [a, b]\}, \quad S < \infty.$$

Wir konstruieren eine Intervallschachtelung $\{[a_n, b_n]\}$ mit $S = \sup f([a_n, b_n])$. Anfang: $a_1 = a$, $b_1 = b$. Sei $[a_n, b_n]$ bereits konstruiert, setze $m = (a_n + b_n)/2$. Falls $S = \sup f([a_n, m])$ ist, setze $a_{n+1} = a_n$, $b_{n+1} = m$; andernfalls muß $S = \sup f([m, b_n])$ gelten, wir setzen $a_{n+1} = m$, $b_{n+1} = b_n$. Wegen dieser Wahl gibt es $x \in [a_{n+1}, b_{n+1}]$

mit $f(x) > S - 1/(n+1)$ und wir ersetzen noch a_{n+1} oder b_{n+1} durch x, so daß weiterhin $S = \sup f([a_{n+1}, b_{n+1}])$ ist. Der Konvergenzpunkt dieser Intervallschachtelung sei $c \in [a,b]$. Wegen der Definition von S und wegen der Stetigkeit von f gilt $\lim f(a_n) = f(c) = \lim f(b_n) \leq S$. Aus $\max(f(a_n), f(b_n)) > S - 1/n$ folgt dann (wie immer mit dem Archimedes Axiom) $f(c) = S$.

Die Definition von *folgen-stetig* sagt etwas über den Wert $f(a)$ an der Stelle a, wenn man die Werte konvergenter Nachbarstellen kennt: $f(a) = \lim f(a_n)$. Wir fragen in der umgekehrten Richtung: Wenn man $f(a)$ kennt, was weiß man über die Werte an Nachbarstellen? Der folgende Satz formuliert eine einfache Antwort, die aber nur indirekt bewiesen werden kann.

Satz über positive Nachbarn. $f : (a,b) \to \mathbb{R}$ sei folgen-stetig bei $c \in (a,b)$ und positiv, $f(c) > 0$. Dann ist f positiv in einem Intervall $(c - \delta, c + \delta)$.

Indirekter Beweis. Zunächst wähle n_1 so groß, daß gilt $(c - 1/n_1, c + 1/n_1) \subset (a,b)$. Entweder ist f positiv auf $(c - 1/n_1, c + 1/n_1)$, oder es gibt ein $a_1 \in (c - 1/n_1, c + 1/n_1)$ mit $f(a_1) \leq 0,\ a_1 \neq c$. Im zweiten Fall wähle $n_2 > n_1$ so, daß $a_2 \notin (c - 1/n_2, c + 1/n_2)$; entweder ist f positiv auf $(c - 1/n_2, c + 1/n_2)$ oder es gibt $a_3 \in (c - 1/n_2, c + 1/n_2)$ mit $f(a_3) \leq 0$. Wiederholung dieses Verfahrens liefert entweder ein Intervall um c, auf dem f wie gewünscht positiv ist oder eine gegen c konvergente Folge $\{a_k\}$ mit $f(a_k) \leq 0$, also den Widerspruch $f(c) = \lim f(a_k) \leq 0$.

Ein derart mühsamer Beweis für eine so einfach erscheinende Aussage macht andere Charakterisierungen der Stetigkeit wünschenswert. Die folgende Definition geht den umgekehrten Weg wie die Folgen-Stetigkeit: Wenn man $f(a)$ kennt, dann kann man etwas über die Nachbarwerte sagen: Es gibt zu jeder *erlaubten* Abweichung von $f(a)$ ein Intervall $[a - \delta, a + \delta]$, in dem *keine größeren* Abweichungen von $f(a)$ vorkommen, genauer:

Definition der ϵ-δ-Stetigkeit. A, B seien Teilmengen von $\mathbb{R}, \mathbb{C}$ oder $\mathbb{R}^d$. Eine Funktion $f : A \to B$ heißt *ϵ-δ-stetig bei $a \in A$* wenn folgende Eigenschaft richtig ist:
Zu jeder Fehlerschranke $\epsilon > 0$ gibt es einen Garantie-Radius $\delta > 0$ so daß

$$x \in A,\ |x - a| \leq \delta \ \Rightarrow\ |f(x) - f(a)| \leq \epsilon.$$

Falls dies für jedes $a \in A$ gilt, heißt f ϵ-δ-stetig auf A.
(Die Worte “Fehlerschranke, Garantie-Radius” sollen die mit der Definition verfolgte Absicht unterstreichen, sie können weggelassen werden.)

Satz über äquivalente Stetigkeitsdefinitionen.
Eine Funktion $f : (a,b) \to \mathbb{R}$ ist genau dann folgen-stetig, wenn sie ϵ-δ-stetig ist.

Beweis. a) f sei ϵ-δ-stetig und $\{c_k\}$ sei eine gegen c konvergente Folge. Wir wollen zeigen: Dann konvergiert auch $f(c_k)$ gegen $f(c)$. Es sei also $\epsilon > 0$ gegeben, und wir müssen k_ϵ finden, so daß gilt

$$k \geq k_\epsilon \Rightarrow |f(c_k) - f(c)| \leq \epsilon.$$

Die ϵ-δ-Stetigkeit von f liefert zu jedem $\epsilon > 0$ ein Garantieintervall $(c - \delta, c + \delta)$ mit $x \in (c - \delta, c + \delta) \Rightarrow |f(x) - f(c)| \leq \epsilon$. Dann liefert die Konvergenz von $\{c_k\}$ ein k_δ mit $k \geq k_\delta \Rightarrow |c_k - c| \leq \delta$. Setzt man daher $k_\epsilon := k_\delta$, so kann man diese Folgerungen zusammensetzen, um $\lim f(c_k) = f(c)$ zu erhalten:

$$k \geq k_\epsilon = k_\delta \Rightarrow |c_k - c| \leq \delta \Rightarrow |f(c_k) - f(c)| \leq \epsilon.$$

b) Umgekehrt müssen wir aus der Folgen-Stetigkeit von f folgern, daß zu jedem $\epsilon > 0$ ein $\delta > 0$ existiert, so daß gilt

$$|x - c| < \delta \Rightarrow |f(x) - f(c)| \leq \epsilon.$$

Dieser Beweis kann nur indirekt geführt werden und ist dem des letzten Satzes sehr ähnlich. Wäre für eine Fehlerschranke ϵ^* das Finden eines solchen Garantieintervalls $(c - \delta, c + \delta)$ unmöglich, so könnten wir eine Folge definieren die zu einem Widerspruch führt:
Zuerst wähle n_1 so daß $(c - 1/n_1, c + 1/n_1) \subset (a, b)$ und $a_1 \in (c - 1/n_1, c + 1/n_1)$ mit $|f(a_1) - f(c)| > \epsilon^*$. Nun seien $n_1 < n_2 < \ldots < n_k$ und $a_1, a_2, \ldots, a_k$ schon gewählt mit $a_j \in (c - 1/n_j, c + 1/n_j)$ und $|f(a_j) - f(c)| > \epsilon^*$. Dann wähle n_{k+1} so groß, daß $a_k \notin (c - 1/n_{k+1}, c + 1/n_{k+1})$; wegen der (indirekten) Annahme ist kein δ gut genug, also finden wir $a_{k+1} \in (c - 1/n_{k+1}, c + 1/n_{k+1})$ mit $|f(a_{k+1}) - f(c)| > \epsilon^*$. Die Folge $\{a_k\}$ ist gegen c konvergierend konstruiert, aber im Widerspruch zur vorausgesetzten Folgen-Stetigkeit ist $|f(a_{k+1}) - f(c)| > \epsilon^*$. Daher kann diese Folge in Wahrheit nicht konstruiert werden, stattdessen erhalten wir nach endlich vielen Schritten ein $\delta = 1/n_k > 0$ mit $|x - c| \leq \delta \Rightarrow |f(x) - f(c)| \leq \epsilon^*$. Damit ist die ϵ-δ-Stetigkeit von f bei c bewiesen.

Aufgabe. Zeige mit diesem ϵ-δ-Kriterium, daß Summe, Produkt und Komposition stetiger Funktionen stetig sind. Das ist mühsamer als mit der ersten Definition. (Dagegen ist der Satz über positive Nachbarn jetzt unmittelbar klar: Wähle $\epsilon = f(c)/2 > 0$, dann gilt in dem hierzu gehörenden Garantieintervall: $f(c) - f(x) \leq f(c)/2$ oder $0 < f(c)/2 \leq f(x)$; damit haben wir eine positive untere Schranke.)

Der Zwischenwertsatz und der Satz vom Maximum werden mit der ϵ-δ-Stetigkeit nicht leichter beweisbar als vorher. Der Beschränktheitssatz besitzt jedoch einen völlig anderen Beweis, der an einer wichtigen Eigenschaft des Intervalls $[a, b]$ hängt, die sich zu sehr kurzen Argumenten mit der ϵ-δ-Stetigkeit verbinden läßt. Wir beweisen zunächst diesen

Überdeckungssatz von Heine-Borel. Zu jedem Punkt $x \in [a,b]$ sei ein $\delta_x > 0$ gegeben. Dann gibt es *endlich* viele $x_k \in [a,b]$ und die zugehörigen $\delta_k := \delta_{x_k}$, so daß $[a,b]$ von den Intervallen $(x_k - \delta_k, x_k + \delta_k)$ überdeckt wird, d. h. $[a,b] \subset \bigcup_{k=1}^{K}(x_k - \delta_k, x_k - \delta_k)$.

Indirekter Beweis. Die Argumentation ähnelt dem Beweis des Beschränktheitssatzes. Wir konstruieren eine Intervallschachtelung $\{[a_n, b_n]\}$, so daß die $[a_n, b_n]$ nicht von endlich vielen Intervallen $(x-\delta_x, x+\delta_x)$ überdeckt werden: Wir nehmen an, das gegebene Intervall $[a,b] =: [a_1, b_1]$ sei *nicht* durch endlich viele der Intervalle $(x-\delta_x, x+\delta_x)$ überdeckbar. Dann konstruieren wir eine Intervallschachtelung $\{[a_n, b_n]\}$, so daß die $[a_n, b_n]$ ebenfalls nicht von endlich vielen der Intervalle $(x-\delta_x, x+\delta_x)$ überdeckt werden. $[a_n, b_n]$ sei schon konstruiert. Definiere $m := (a_n + b_n)/2$. Dann ist $[a_n, m]$ oder $[m, b_n]$ nicht endlich überdeckbar und eines dieser beiden Intervalle liefert $[a_{n+1}, b_{n+1}]$. Schließlich sei c der Grenzpunkt dieser Intervallschachtelung. Offenbar überdeckt das eine Intervall $(c - \delta_c, c + \delta_c)$ die Intervalle $[a_n, b_n]$, die genügend klein sind: $|b_n - a_n| = |b_1 - a_1| \cdot 2^{-n} < \delta_c$. Der Widerspruch zeigt, daß es diese Intervallschachtelung nicht geben kann, d. h. $[a,b]$ ist endlich überdeckbar.

Mit dem Satz von Heine Borel variieren wir zunächst den Beschränktheitssatz.

Definition von lokal beschränkt. Eine Funktion f heißt *lokal beschränkt*, wenn es zu jedem Punkt x ihres Definitionsbereichs ein $\delta_x > 0$ gibt, so daß f auf $(x - \delta_x, x + \delta_x)$ beschränkt ist.
Stetige Funktionen sind lokal beschränkt, denn zu jedem c im Definitionsbereich gibt es $\delta_1 > 0$ mit $\quad |x - c| \le \delta_1 \Rightarrow |f(x) - f(c)| \le 1 \Rightarrow |f(x)| \le |f(c)| + 1$

Satz. Jede auf $[a,b]$ nur lokal beschränkte Funktion f ist sogar global beschränkt.

Beweis. Überdecke $[a,b]$ durch endlich viele der Intervalle $(x - \delta_x, x + \delta_x)$. Das Maximum der *endlich vielen* vorausgesetzten lokalen Schranken auf den Intervallen $(x_k - \delta_k, x_k + \delta_k)$ ist dann eine Schranke für f auf $[a,b]$.

Mit dieser Argumentation hat man eine leicht anwendbare Beweismethode entwickelt. Die Formulierungen dieser Beweise sind stark standardisiert. Man hat sie jedoch erst verstanden, wenn man weiß, was für Variationen der Wortwahl zulässig sind. Meiner Meinung nach sind dafür mündliche Diskussionen unerläßlich. Wir zeigen den

Satz über gleichmäßige Stetigkeit. Die Funktion $f : [a,b] \to \mathbb{R}$ sei stetig. Dann ist f sogar gleichmäßig stetig, d.h. zu jedem $\epsilon > 0$ gibt es ein $\delta > 0$, das ausdrücklich von x *unabhängig* gewählt werden kann, so daß gilt:

$$|x - y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon.$$

Beweis. Zu jedem $\epsilon > 0$ und jedem $x \in [a,b]$ gibt es wegen der Stetigkeit von f bei x ein $\delta_x > 0$, so daß gilt:

$$x, y \in [a,b] \text{ und } |x-y| < 2 \cdot \delta_x \Rightarrow |f(x) - f(y)| < \epsilon/2.$$

Beachte, daß das Einfügen der Faktoren 2 und 1/2 mit der Definition der Stetigkeit verträglich ist. Endlich viele der $(x-\delta_x, x+\delta_x)$ überdecken $[a,b]$, also $[a,b] \subset \bigcup_{k=1}^{K}(x_k-\delta_k, x_k+\delta_k)$. Setze $\delta := \min_{k=1\dots k} \delta_k$. Wir zeigen, daß dies für unsere Behauptung genügt: Zu jedem $x \in [a,b]$ wähle zuerst x_k so daß $x \in (x_k - \delta_k, x_k + \delta_k)$; dann gilt $|f(x) - f(x_k)| < \epsilon/2$. Ist nun $|x-y| < \delta$ so folgt $y \in (x_k - 2\delta_k, x_k + 2\delta_k)$, also auch $|f(y) - f(x_k)| < \epsilon/2$. Mit der Dreiecksungleichung hat man $|f(x) - f(y)| < \epsilon$ erreicht.

Satz über gleichmäßige Approximierbarkeit. Stetige Funktionen $f : [a,b] \to \mathbb{R}$ sind gleichmäßig stückweise linear approximierbar, d. h. zu jedem $\epsilon > 0$ gibt es eine stückweise lineare Funktion ℓ mit $|f(x) - \ell(x)| < \epsilon$ für alle $x \in [a,b]$.

Beweis. Wegen der gleichmässigen Stetigkeit von f kann man zu jedem $\epsilon > 0$ ein $\delta > 0$ so wählen, daß gilt

$$x, y \in [a,b], |x-y| \le \delta \Rightarrow |f(x) - f(y)| \le \frac{\epsilon}{2}.$$

Dazu wähle eine Einteilung $a = t_0 < t_1 < \ldots < t_N = b$ mit $|t_j - t_{j-1}| < \delta$. Nun definiere eine stückweise lineare Funktion ℓ als lineare Interpolation zu den Stützstellen t_j,

$$t \in [t_{j-1}, t_j] \Rightarrow \ell(t) := f(t_{j-1}) + \frac{f(t_j) - f(t_{j-1})}{t_j - t_{j-1}} \cdot (t - t_{j-1}).$$

Dann gilt in der Tat, daß f von l gleichmäßig approximiert wird:

$$t \in [a,b] \;\Rightarrow\; |f(t) - \ell(t)| \le |f(t) - f(t_j)| + |\ell(t) - \ell(t_j)| \le \frac{\epsilon}{2} + \frac{\epsilon}{2}.$$

Aufgabe. Formuliere für den Satz vom Maximum einen Beweis mit Hilfe des Satzes von Heine-Borel, beginnend mit: Angenommen für jedes $x \in [a,b]$ wäre $f(x) < S := \sup f([a,b])$, dann folgte aus der Stetigkeit von f bei $x \ldots$.

Als nächstes wollen wir den Hauptsatz der Differential-und Integralrechnung von dehnungsbeschränkten auf stetige Funktionen verbessern. Dazu brauchen wir die

Endgültige Differenzierbarkeitsdefinition. Eine Funktion $f : (\alpha, \omega) \to \mathbb{R}$ heißt *differenzierbar* bei $a \in (\alpha, \omega)$ mit Ableitung $m = f'(a)$ und Tangente $T_a(x) := f(a) + m \cdot (x-a)$, falls gilt:

$$|f(x) - T_a(x)| = \Phi_a(x)|x-a| \text{ und } \Phi_a(x) \text{ ist stetig bei } x = a \text{ mit } \Phi_a(a) = 0$$

oder:

$$|\frac{f(x)-f(a)}{x-a} - m| = \Phi_a(x) \text{ und } \Phi_a(x) \text{ ist stetig bei } x = a \text{ mit } \Phi_a(a) = 0$$

oder nach Einsetzen der Definition der Stetigkeit von Φ_a :

Zu jedem $\epsilon > 0$ gibt es $\delta > 0$ so daß: $|x-a| \le \delta \Rightarrow |f(x) - T_a(x)| \le \epsilon \cdot |x-a|$.

Aufgabe. Im Kapitel Komplexe Zahlen wurde der Schrankensatz bewiesen, ohne daß Gleichmäßigkeitsvoraussetzungen gemacht wurden. Benutze die dort verwendete Beweisstrategie (erst ganz am Ende ist die Anpassung an die endgültige Definition nötig), um jetzt den *Monotoniesatz* für differenzierbare $f : [\alpha, \omega] \to \mathbb{R}$ zu beweisen:

$$f' \ge 0 \;\Rightarrow\; f \text{ ist monoton wachsend, d.h. } x \le y \Rightarrow f(x) \le f(y).$$

Satz über die Existenz von Stammfunktionen und Integralen:

Stetige Funktionen $f : [\alpha, \omega] \to \mathbb{R}$ besitzen Stammfunktionen F, d.h. $F' = f$.

Deshalb sind sie integrierbar mit

$$a, b \in [\alpha, \omega] \Rightarrow F(b) - F(a) = \int_a^b f(x)dx.$$

Beweis. Der Beweis verläuft ebenso wie der für dehnungsbeschränkte f im Kapitel Integralrechnung, wir müssen ihn nur an die ϵ-δ-Fehler anpassen.

Für jedes $1/n > 0$ sei $\ell_n : [\alpha, \omega] \to \mathbb{R}$ eine gleichmäßige stückweise lineare Approximation mit $x \in [\alpha, \omega] \Rightarrow |f(x) - \ell_n(x)| \le 1/n,\ |\ell_n(x) - \ell_m(x)| \le (1/n + 1/m)$, wie im vorhergehenden Satz bewiesen. Dabei wurde schon benutzt, daß f gleichmäßig stetig ist, d.h.: Zu dem gewählten $\epsilon > 0$ gibt es $\delta > 0$ so daß

$$x, y \in [\alpha, \omega],\ |x-y| \le \delta \Rightarrow |f(x) - f(y)| \le \epsilon \Rightarrow |\ell_n(x) - \ell_n(y)| \le 2/n + \epsilon.$$

Wieder haben die ℓ_n stückweise quadratische Stammfunktionen $Q_n : [\alpha, \omega] \to \mathbb{R}$ mit $Q_n' = \ell_n,\ Q_n(0) = 0$, und aus der vorhergehenden Abschätzung folgt für die Tangentenapproximation von Q_n aus dem Monotoniesatz

$$x, y \in [\alpha, \omega],\ |x-y| \le \delta \Rightarrow |Q_n(x) - Q_n(y) - \ell_n(y)(x-y)| \le (2/n + \epsilon) \cdot |x-y|.$$

Ebenfalls aus dem Monotoniesatz folgt, daß die $\{Q_n\}$ eine Cauchyfolge bilden, nämlich:

$$x \in [\alpha, \omega] \Rightarrow |Q_n(x) - Q_m(x)| \le \max(|\ell_n - \ell_m|) \cdot |x - \alpha| \le |\omega - \alpha| \cdot (1/n + 1/m).$$

Mit dem Vollständigkeitsaxiom haben wir die Grenzfunktion Q_∞ dieser Cauchyfolge. Mit dem Archimedes Axiom dehnt sich die Tangentenapproximation der Q_n auf die Grenzfunktion aus:

$$x, y \in [\alpha, \omega],\ |x-y| \le \delta \Rightarrow |Q_\infty(x) - Q_\infty(y) - f(y)(x-y)| \le \epsilon \cdot |x-y|.$$

Da wir zu jedem $\epsilon > 0$ ein $\delta > 0$ gefunden haben, so daß diese Aussage gilt, ist die Differenzierbarkeitsdefinition für Q_∞ erfüllt mit $Q_\infty' = f$. Die Grenzfunktion ist also Stammfunktion von f.

Wie im Kapitel Integralrechnung können Riemannsummen von f bis auf kleine Fehler durch

Differenzen der Stammfunktion berechnet werden. Daher ist f integrierbar mit

$$a, b \in [\alpha, \omega] \Rightarrow Q_\infty(b) - Q_\infty(a) = \int_a^b f(x)dx, \;\; Q'_\infty = f.$$

Folgen stetiger Funktionen.

Hiermit sind die wesentlichen Eigenschaften, die stetige Funktionen als einzelne Funktionen haben, bereits abgehandelt. Als nächstes untersuchen wir Folgen stetiger Funktionen f_n. Insbesondere lernen wir eine neue Methode zur Konstruktion von Grenzfunktionen kennen im Satz über gleichmäßig konvergente Funktionenfolgen. Für alles folgende machen wir die

Mindestvoraussetzungen für Funktionenfolgen $\{f_n(x)\}$.
f_n sei stetig und $\{f_n(x)\}$ sei konvergent für jedes $x \in D \subset$ Definitionsbereich (f_n).
Diese Voraussetzungen erlauben, eine Grenzfunktion zu definieren:

$$x \in D, \; f(x) := \lim_{n\to\infty} f_n(x).$$

Aber f braucht *nicht* stetig zu sein. Beispiel:
$D = [0,1], \;\; f_n(x) := x^n, \;\; f(x) := \lim_{n\to\infty} f_n(x)$ gibt $f(1) = 1, \;\; f(x) = 0$ für $x \in [0,1)$.

Die folgenden Sätze handeln also davon, unter welchen zusätzlichen Voraussetzungen man *stetige Grenzfunktionen* bekommt. Zunächst verwenden wir die Ideen aus den früheren Abschnitten: Falls eine Fehlerkontrolle unabhängig von n (auch: "gleichmäßig in n" oder: "gleichgradig in n") möglich ist, so überträgt sich dies auf die Grenzfunktion:

Satz über gleichgradige Fehlerkontrolle.
Voraussetzung. Zu jedem $\epsilon > 0$ und jedem $x \in D$ gibt es ein $\delta_x > 0$, ausdrücklich unabhängig von n, so daß gilt:

$$x, y \in D \;, \; |y - x| \le \delta_x \Rightarrow |f_n(y) - f_n(x)| \le \epsilon.$$

Behauptung: Die Grenzfunktion ist stetig, nämlich:

$$x, y \in D \;, \; |y - x| \le \delta_x \Rightarrow |f(y) - f(x)| \le \epsilon.$$

Beweis. $r_n := |f(x) - f_n(x)|$ und $s_n := |f(y) - f_n(y)|$ sind nach Voraussetzung Nullfolgen. Aus der gleichmäßigen Fehlerkontrolle für die f_n und der Dreiecksungleichung folgt

$$x, y \in D, \; |y - x| \le \delta_x \Rightarrow |f(y) - f(x)| \le \epsilon + r_n + s_n,$$

und das Archimedes-Argument beseitigt die Nullfolgen.

Dieser Satz genügt, um am Ende dieses Kapitels zwei der Beispiele, die Cantortreppe und Hilberts flächenfüllende Kurve, zu behandeln. Aber im Gegensatz zu den vorhergehenden Abschnitten, in denen solche von n unabhängigen Fehlerkontrollen sich schon beinahe aufdrängten, ist bei interessanten Konstruktionen mit stetigen Funktionen das Gegenteil der Fall: die Fehlerkontrolle wird oft für größere Folgenindices immer schlechter (wie bei dem dritten Beispiel am Ende, dem Weierstraßoszillator). Ein erfolgreicher Ausweg ist die Definition eines neuen Begriffs. Statt die Fehler unabhängig von n zu kontrollieren, machen wir die Konvergenz unabhängig von der Stelle des Definitionsbereichs:

Definition gleichmäßig konvergenter Funktionenfolgen. Die Funktionenfolge $\{f_n\}$ heißt gleichmäßig konvergent (genauer: gleichmäßig Cauchy) auf D, falls zu jedem $\epsilon > 0$ ein n_ϵ existiert, so daß für alle Punkte $x \in D$ gilt:

$$m, n \geq n_\epsilon \Rightarrow |f_m(x) - f_n(x)| \leq \epsilon.$$

Man sagt auch: Die Konvergenz (für $n \to \infty$) von $\{f_n(x)\}$ ist gleichmäßig in x.
Die Zahl $||f - g|| := \sup\{|f(x) - g(x)|; x \in D\}$ heißt Abstand der Funktionen f, g.

Hauptsatz über gleichmäßig konvergente Funktionenfolgen. $f_n : (a, b) \to \mathbb{R}$ sei eine gleichmäßig konvergente Folge stetiger Funktionen mit $f(x) := \lim_{n\to\infty} f_n(x)$.
Behauptung: Die Grenzfunktion f ist stetig.

Beweis. Da die Behauptung von anderer Natur ist als alle vorhergehenden, lernen wir eine neue Beweismethode kennen, die ich als $\epsilon/3$-Argument zitiere. $\epsilon > 0$ sei gegeben, wähle zunächst wegen der gleichmäßigen Konvergenz n_ϵ, so daß gilt:

$$x \in (a, b),\ n \geq n_\epsilon \Rightarrow |f(x) - f_n(x)| \leq \frac{\epsilon}{3}.$$

Dann wird für ein $n^* \geq n_\epsilon$ die Stetigkeit der einen Funktion f_{n^*} benutzt und ein δ_x gewählt, so daß gilt:

$$x, y \in (a, b),\ |x - y| < \delta_x \Rightarrow |f_{n^*}(y) - f_{n^*}(x)| \leq \frac{\epsilon}{3}.$$

Aus der Dreiecksungleichung folgt die Behauptung:

$$\begin{aligned} x, y \in (a, b),\ |x - y| < \delta_x \Rightarrow |f(x) - f(y)| = & \\ &= |(f(x) - f_{n^*}(x)) + (f_{n^*}(x) - f_{n^*}(y)) + (f_{n^*}(y) - f(y))| \leq \\ &\leq |f(x) - f_{n^*}(x)| + |f_{n^*}(x) - f_{n^*}(y)| + |f_{n^*}(y) - f(y)| \leq \\ &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3}. \end{aligned}$$

Bemerkung. Anders als früher hat man keine Kontrolle darüber, wie stark δ verkleinert werden muß, um ein verkleinertes ϵ zu garantieren, denn zu einem kleineren ϵ muß man eine Approximation f_{n^*} der Grenzfunktion mit (i.a.) größerem Index n^* benutzen, so daß man damit rechnen muß, ein sehr viel kleineres δ_x wählen zu müssen. Bei Anwendung

dieses Satzes hat man daher keine quantitative Kontrolle der Eigenschaften der Grenzfunktion mehr. Dies paßt zur Definition der Stetigkeit, durch die ja auch keinerlei quantitative Kontrolle verlangt wird.
Die beiden bewiesenen Sätze haben unmittelbare Anwendungen auf Integration und Differentiation der Grenzfunktion einer konvergenten Funktionenfolge.

Satz über Integration konvergenter Funktionenfolgen. Die Funktionen $f_n : [a, b] \to \mathbb{R}$ seien stetig und die Folgen $\{f_n(x)\}$ seien konvergent gegen $f(x)$. Wir wissen schon, daß die stetigen Funktionen f_n Stammfunktionen haben, also integrierbar sind.
a) Falls die $\{f_n\}$ gleichmäßig konvergieren, so gilt für die Grenzfunktion

$$\lim \int_a^b f_n(x)dx = \int_a^b \lim f_n(x)dx = \int_a^b f(x)dx.$$

b) Falls die Fehlerkontrolle der f_n gleichmäßig in n ist, so gilt für die Grenzfunktion

$$\lim \int_a^b f_n(x)dx = \int_a^b \lim f_n(x)dx = \int_a^b f(x)dx.$$

Beweis. Die Voraussetzungen a) oder b) erlauben, die beiden Stetigkeitssätze für Grenzfunktionen anzuwenden, d.h. die Grenzfunktion f ist stetig, also integrierbar. Die gleich mäßige Konvergenz a) verträgt sich in besonders einfacher Weise mit der Integration:

$$||f - g|| \le \epsilon \Rightarrow |\int_a^b f(x)dx - \int_a^b g(x)dx| \le \int_a^b ||f - g||dx = ||f - g|| \cdot |b - a|.$$

Durch Anwenden auf $g = f_n,\ n \ge n_\epsilon$ folgt die Vertauschungsregel $\lim \int = \int \lim$.

Im Fall b) der gleichmäßigen Fehlerkontrolle können wir den Unterschied zwischen Riemann-Summen und Integralen unabhängig von n kontrollieren:
Zu $\epsilon > 0$ wähle $\delta > 0$ so daß unabhängig von n gilt:

$$x, y \in [a, b], |x - y| \le \delta \Rightarrow |f_n(y) - f_n(x)| \le \epsilon.$$

Danach sei $a = t_0 < t_1 < \ldots < t_N = b$ eine Einteilung mit $|t_j - t_{j-1}| < \delta$. Sei $\Sigma_1^{(n)}$ irgendeine Riemannsumme von f_n zu dieser Einteilung,

$$\Sigma_1^{(n)} := \sum_{j=1}^{N} f_n(\tau_j) \cdot (t_j - t_{j-1}), \quad \tau_j \in [t_{j-1}, t_j]$$

und $\Sigma_2^{(n)}$ eine Riemannsumme für f_n zu irgendeiner feineren Einteilung. Dann folgt aus der gleichmäßigen Fehlerkontrolle

$$|\Sigma_1^{(n)} - \Sigma_2^{(n)}| \leq \epsilon \cdot \sum_{j=1}^{N} |t_j - t_{j-1}| = \epsilon \cdot |b-a|,$$

und daraus nach Definition des Integrals

$$|\Sigma_1^{(n)} - \int_a^b f_n(x)| \leq \epsilon \cdot |b-a|.$$

Diese Abschätzung gilt wegen der gleichmäßigen Fehlerkontrolle (und mit Zitieren des Archimedes Axioms) auch für jede Riemannsumme Σ_1 der Grenzfunktion f zur gleichen Intervalleinteilung:

$$|\Sigma_1 - \int_a^b f(x)dx| \leq \epsilon \cdot |b-a|.$$

Dies sind nun die Voraussetzungen für ein $\epsilon/3$-Argument: Da $\Sigma_1^{(n)} - \Sigma_1$ (bei gleich gewählten Zwischenstellen τ_j) Summe von N festen Nullfolgen ist, können wir $n \geq n_\epsilon$ so groß wählen, daß $|\Sigma_1^{(n)} - \Sigma_1| \leq \epsilon \cdot |b-a|$ ist, also:

$$n \geq n_\epsilon \Rightarrow |\int_a^b f_n(x)dx - \int_a^b f(x)dx| \leq 3\epsilon \cdot |b-a|.$$

Diese Abschätzung ist die Behauptung:

$$\lim_{n\to\infty} \int_a^b f_n(x)dx = \int_a^b f(x)dx = \int_a^b \lim f_n(x)dx,$$

mit anderen Worten, wieder gilt die Vertauschungsregel: $\lim \int = \int \lim$.

Satz über Differentiation der Grenzfunktionen konvergenter Folgen. $f_n : [a,b] \to \mathbb{R}$ sei stetig differenzierbar (d. h. f_n' sei stetig). Die Folgen $\{f_n(a)\}$ und $\{f_n'(x)\}$ seien für alle $x \in [a,b]$ konvergent mit $\lim f_n(a) =: f(a)$, $\lim f_n'(x) =: g(x)$. Diese Voraussetzungen genügen noch nicht, um die Differenzierbarkeit einer Grenzfunktion f der f_n mit $f' = g$ zu beweisen. Daher erfülle die Folge der *Ableitungen* $\{f_n'\}$ zusätzlich eine der Voraussetzungen a) oder b) im letzten Satz über die Integration konvergenter Folgen. Dann konvergiert f_n gegen eine Grenzfunktion f und man kann Differentiation und Grenzwert vertauschen:

$$f'(x) = \left(\lim_{n\to\infty} f_n(x)\right)' = \lim_{n\to\infty} (f_n'(x)) = g(x).$$

Beweis. Da f_n' die Stammfunktion f_n hat, gilt $\int_a^x f_n'(t)dt = f_n(x) - f_n(a)$.
Da die linke Seite und $\{f_n(a)\}$ konvergieren, folgt zunächst, daß $\{f_n(x)\}$ konvergiert. Außerdem konvergiert die linke Seite gegen eine Stammfunktion G von g:

$$G(x) := \int_a^x g(t)dt = \lim_{n\to\infty} \int_a^x f_n'(t)dt = f(x) - f(a).$$

Aber $f(x) = f(a) + G(x)$ zeigt $f'(x) = G'(x) = g(x)$, unsere Behauptung.

Bemerkung: Die vorgeführten Argumente zum Umgang mit stetigen Funktionen überstehen viele Verallgemeinerungen. Sie haben sich in den letzten 100 Jahren nicht mehr geändert und wohl ihre endgültige Form erreicht. Ich nehme an, jeder Leser kann die gezeigten Beweise für stetige Kurven $c : I \to \mathbb{R}^d$ wiederholen. Auch über Verallgemeinerungen auf Abbildungen $F : \mathbb{R}^m \to \mathbb{R}^d$, braucht man nicht lange nachzudenken. Im Gegensatz dazu liegt es nicht auf der Hand, wie man die Differenzierbarkeit solcher Abbildungen F definieren sollte. Auch bei dem Versuch, die Differentiation umzukehren, Stammfumktionen zu definieren, treten neuartige Phänomene auf, die es für eindimensionale Funktionen nicht gibt. Für die höherdimensionale Differentialrechnung sollte man daher die Stetigkeitsargumente schon kennen gelernt haben. Aus diesem Grund ist es üblich, auch für eindimensionale Funktionen die Stetigkeit vor der Differentialrechnung zu behandeln. Ich habe diese Reihenfolge geändert, weil mir so ein nahtloserer Anschluß an die Vorkenntnisse möglich war. Außerdem können wir jetzt überraschende stetige Funktionen konstruieren; diese sind bei der üblicheren Reihenfolge weit von der Stetigkeitsdefinition getrennt, weil auch noch die Differentialrechnung vor den Funktionenfolgen behandelt wird.

Beispiele stetiger Funktionen

Die intuitive Seite des Funktionsbegriffs, also das, was man sich als normales Verhalten von Funktionen vorstellt, wird natürlich geprägt durch die Funktionen, die man am besten kennt. Ohne Zweifel sind das die explizit berechenbaren rationalen Funktionen. Sie erzeugen ein zu harmloses Bild vom typischen Verhalten stetiger Funktionen. Ich will deshalb drei Beispiele vorstellen, die, wenn man sich an den rationalen Funktionen orientiert, ein sehr ungewöhnliches Verhalten haben, die aber unter den stetigen Funktionen keine Sonderlinge sind. Wenn man zu einem dieser Beispiele irgendeine der bisher bekannten "harmlosen" Funktionen addiert, dann bleibt das "ungewöhnliche" Verhalten der folgenden Beispiele erhalten. Im Kreise der stetigen Funktionen fallen sie also keineswegs als ungewöhnlich auf.

Außerdem ist es üblich, sich während der Ausbildung an anschauliche Kurzformen mathematischer Argumente zu gewöhnen. Auch daher ist es zweckmäßig, ein angemessenes Bild von den zu einem Begriff gehörenden Beispielen zu haben. Das erste Beispiel ist

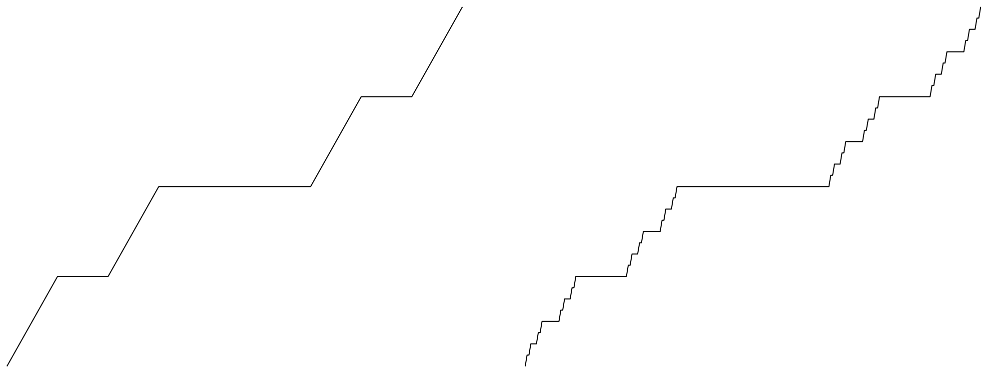
Die Cantortreppe. Ich erinnere daran, daß eine differenzierbare Funktion mit Ableitung $f' = 0$ notwendig konstant ist. Ein Versuch, die Voraussetzungen dieses wichtigen Satzes abzuschwächen, wird durch die Cantortreppe widerlegt:

Die Cantortreppe ist eine stetige und monoton von 0 auf 1 wachsende Funktion, deren Ableitung "fast überall" (s.u.) verschwindet. Sie wird mit Hilfe von stückweise linearen stetigen monotonen Approximationen $f_k : [0, 1] \to [0, 1]$ beschrieben, die mit wachsendem k auf immer mehr Teilintervallen von $[0, 1]$ konstant sind und dort mit der Grenzfunktion übereinstimmen. Die erste stückweise lineare Funktion ist:

$$f_1(x) := \begin{cases} \frac{3}{2}\cdot x & \text{für } 0 \le x \le \frac{1}{3} \\ \frac{1}{2} & \text{für } \frac{1}{3} \le x \le \frac{2}{3} \\ \frac{3}{2}\cdot x - \frac{1}{2} & \text{für } \frac{2}{3} \le x \le 1 \end{cases}$$

Die folgenden Funktionen f_k werden rekursiv definiert.Um f_{k+1} zu definieren, wird der Graph von f_k horizontal auf ein Drittel, vertikal auf die Hälfte gestaucht; zwei solche verkleinerten Kopien definieren f_{k+1} auf dem linken und rechten Drittel von $[0,1]$, in der Mitte ist f_{k+1} gleich $1/2$. In Formeln:

$$f_{k+1}(x) := \begin{cases} \frac{1}{2}\cdot f_k(3x) & \text{für } 0 \le x \le \frac{1}{3} \\ \frac{1}{2} & \text{für } \frac{1}{3} \le x \le \frac{2}{3} \\ \frac{1}{2}\cdot (1 + f_k(3x-2)) & \text{für } \frac{2}{3} \le x \le 1 \end{cases}$$



Zwei Approximationen der monotonen stetigen Cantortreppe, $f' = 0$ fast überall.
Die Stetigkeit dieser Approximationen ist gleichmäßig (od. gleichgradig) in n:

$$x, y \in [0,1], |x-y| < 3^{-n} \quad \Rightarrow \quad |f_k(x) - f_k(y)| < 2^{-n}$$

Die Approximationen konvergieren gleichmäßig in x (geometrisch majorisiert):

$$|f_k(x) - f_{k+1}(x)| \le \tfrac{1}{3}\cdot 2^{-(k+1)}.$$

Die Funktion f_k ist daher auf $(2^k - 1)$ disjunkten Teilintervallen konstant mit den (wachsenden) Werten $j \cdot 2^{-k}$, $(j = 1, \ldots, 2^k - 1)$; ich nenne diese Intervalle "Plateauintervalle". Am Rande und zwischen den Plateauintervallen bleiben 2^k Intervalle der Länge 3^{-k}, auf denen f_k mit der Steigung $(3/2)^k$ von einem Plateau zum nächsten wächst; ich nenne diese Intervalle "Leiterintervalle". Die rekursive Definition erreicht, daß f_{k+1} mit f_k auf den Plateauintervallen von f_k übereinstimmt; jedes Leiterintervall von f_k wird in drei Teile geteilt, jedes mittlere Drittel wird ein neues Plateau auf der mittleren Höhe, die übrigen Teile sind die 2^{k+1} Leiterintervalle von f_{k+1}, dort ist die Steigung von f_{k+1} 3/2-mal so groß wie die von f_k. Insbesondere gilt:

$$x \in [0,1] \;\Rightarrow\; |f_{k+1}(x) - f_k(x)| \le \tfrac{1}{3} \cdot 2^{-(k+1)}.$$
$$x, y \in [0,1], |x-y| < 3^{-n} \quad\Rightarrow\quad |f_k(x) - f_k(y)| < 2^{-n}$$

Die Folge $\{f_k\}$ stetiger Funktionen ist also gleichmäßig (in x) geometrisch majorisiert, konvergiert also gegen eine stetige Grenzfunktion. (Wir haben auch eine gleichgradige (in k) Fehlerkontrolle und auch aus diesem Grund eine stetige Grenzfunktion.) Diese Grenzfunktion ist auf allen (offenen) Plateauintervallen von f_k (für jedes k) *differenzierbar* mit der Ableitung 0, die übrigen Punkte liegen in 2^k Intervallen, deren Gesamtlänge $(2/3)^k$ eine Nullfolge ist. Das wird zusammengefaßt durch die Formulierung: Die Cantortreppe ist "fast überall" differenzierbar mit Ableitung 0.

Beispiel 2: Der Weierstraßoszillator ist eine stetige jedoch nirgends differenzierbare Funktion, die ebenfalls aus gleichmäßig geometrisch majorisierten Approximationen konstruiert wird. Baustein ist die Sägezahn-Funktion (mit floor(x) bzw. ceil(x) werden die nächst kleinere bzw. größere ganze Zahl bezeichnet):
$\text{säg}\,(x) := \min(x - \text{floor}\,(x), \text{ceil}\,(x) - x) = \text{Abstand}(x, \mathbb{Z})$.

$$f_n(x) := \sum_{k=1}^{n} 2^{-k} \cdot \text{säg}\,(4^k \cdot x).$$

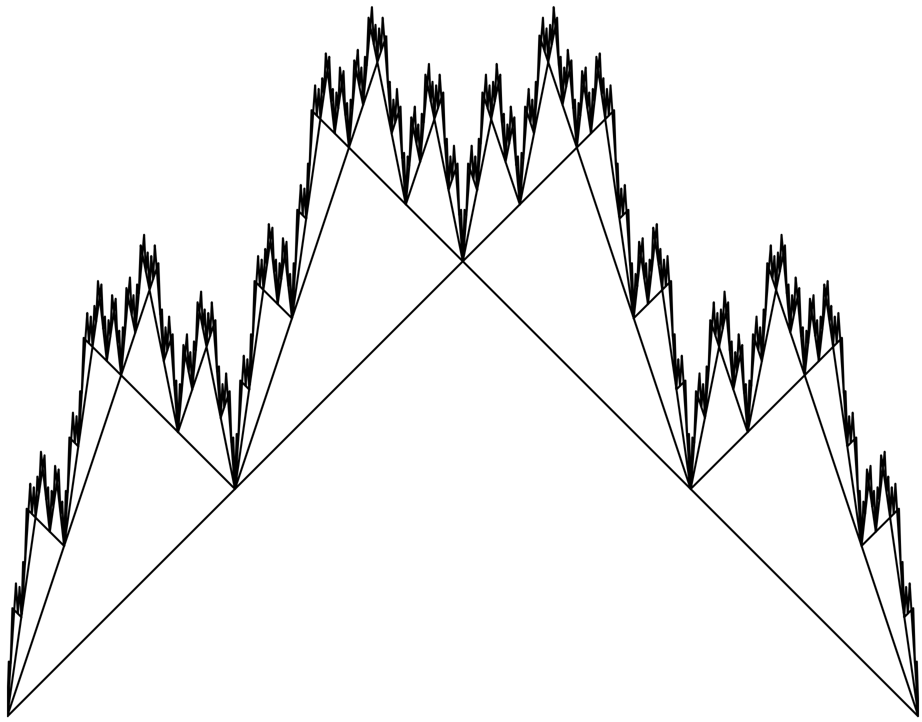
Offenbar gilt für alle x: $\qquad |f_n(x) - f_{n+1}(x)| \le 2^{-(n+1)}$,
d.h. die Folge $\{f_n\}$ ist eine gleichmäßig konvergente Folge stetiger Funktionen, die *stetige* Grenzfunktion $W(x) := \lim f_n(x)$ ist der Weierstraßoszillator (diesmal haben wir *keine* gleichgradige Fehlerkontrolle). Wir müssen begründen, warum W nirgends differenzierbar ist. Dazu bemerken wir zunächst, daß an rationalen Stellen mit Nenner 4^{-n} die Grenzfunktion aus der n-ten Approximation berechnet werden kann, also:

$$\begin{aligned} x := j \cdot 4^{-n},\ \Delta x = 0.5 \cdot 4^{-n} \ \Rightarrow\ & \text{säg}(4^n x) = 0,\ f_{n-1}(x) = f_n(x) = \ldots = W(x) \\ & f_n(x + \Delta x) = f_{n+1}(x + \Delta x) = \ldots = W(x + \Delta x) \\ & W(x + \Delta x) \ge W(x) + 2^{-(n+1)} \\ & \frac{W(x + \Delta x) - W(x)}{\Delta x} \ge 2^n \end{aligned}$$

Jede Zahl mit Nenner 4^{-n} ist auch Zahl mit Nenner 4^{-N} für $N \ge n$, daher gehen von allen Punkten $j \cdot 4^{-n}$ Sehnen mit Steigung $\ge 2^N$ über Intervallen der Länge $0.5 \cdot 4^{-N}$ aus, so daß an diesen Stellen kein Grenzwert $\lim_{N\to\infty}$ dieser Sehnensteigungen existiert.

Schließlich sei $\tilde{x} \in (j\cdot 4^{-n},\ j\cdot 4^{-n}+0.5\cdot 4^{-n}) = (x_0, x_1)$. Dann ist $W(x_1)-W(x_0) \geq 2^{-(n+1)}$, also $\max(|W(\tilde{x}) - W(x_0)|,\ |W(\tilde{x}) - W(x_1)|) \geq 0.5 \cdot 2^{-(n+1)}$ und $\max(\bar{x} - x_1,\ x_2 - \bar{x}) \leq 2^{-2n-1}$. Deshalb gibt es auch von $\tilde{x}$ aus Sehnen, deren Steigungsbetrag $\geq 2^n$ ist, so daß wieder kein endlicher Grenzwert der Sehnensteigungen mit einem Endpunkt $(\tilde{x},\ W(\tilde{x}))$ existiert.

Man kann zusätzlich zeigen, daß der Graph von W über jedem noch so kleinen Intervall *unendliche* Länge hat, so daß man den Graphen von W nicht mehr sehr gut eine "ohne abzusetzen zeichenbare Kurve" nennen kann (was manchmal als angebliche Veranschaulichung stetiger Funktionen benutzt wird). – Die Abschätzungen der Länge sind einfacher, wenn man die Variante $\sum 2^{-n} \cdot \text{säg}(8^n x)$ betrachtet.



Weierstraß' stetige, nirgends differenzierbare Funktion ist Grenzwert einer gleichmäßig konvergenten Folge: $W(x) = \sum_{n=0}^{\infty} 2^{-n} \cdot \text{säg}(4^n x)$.

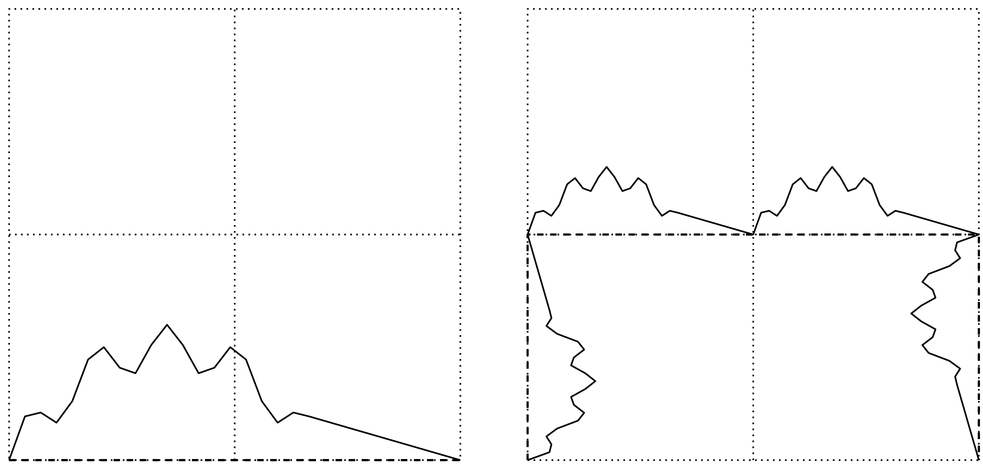
Beispiel 3: Hilberts flächenfüllende Kurve. Die Graphen stetiger Funktionen sind "Kurven" mit eineindeutiger Projektion auf die x-Achse; dies ist ein untypisch einfaches Verhalten für stetige Kurven: Hilberts flächenfüllende Kurve trifft alle Punkte des Einheitsquadrates. Auch diese Kurve wird als Grenzwert gleichmäßig konvergenter Approximationen konstruiert, verstehen muß man nur, wie man aus einer Approximation die nächste gewinnt.

Sei $c_1; [0,1] \to [0,1]$ eine stetige Kurve mit $c_1(0) = (0,0),\ c_1(1) = (1,0)$; sie verbindet also die beiden unteren Eckpunkte des Einheitsquadrates im ersten Quadranten.

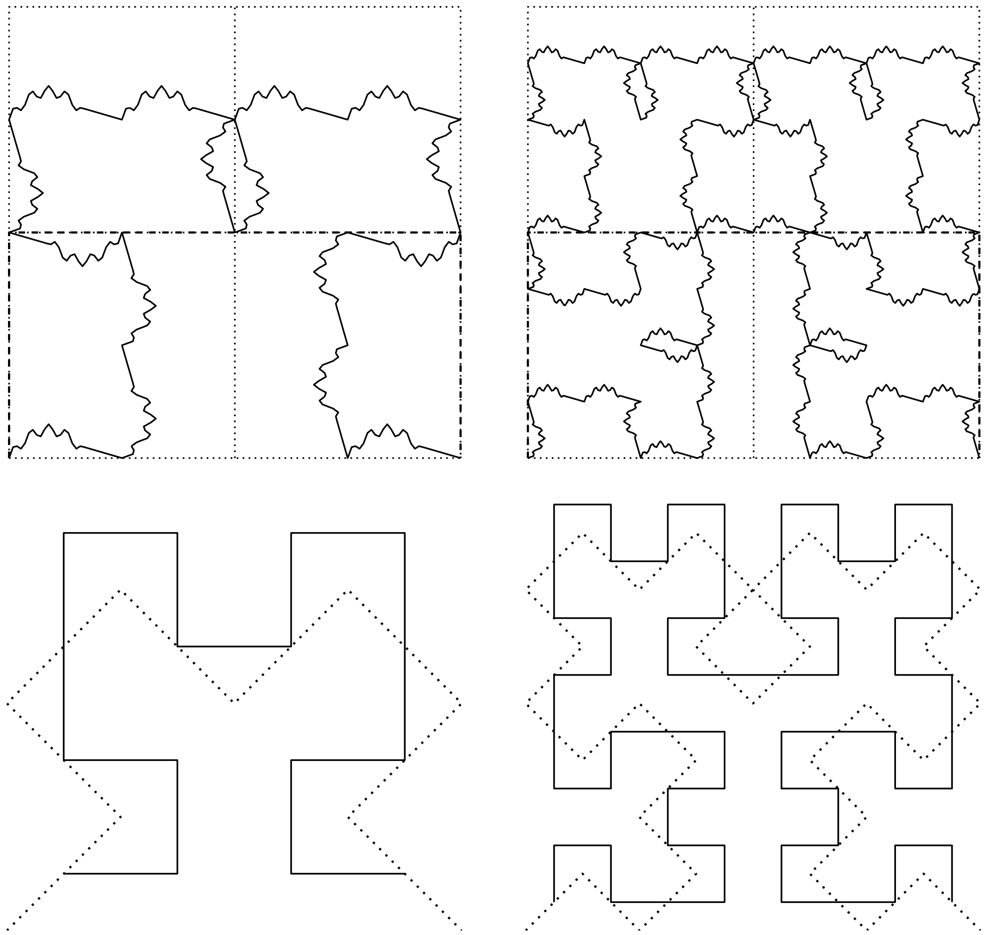
$\tilde{c}(x) := \frac{1}{2}c_1(4\cdot x),\ 0 \le x \le 0.25$, ist eine halb so große Kopie, die aber nur mit einem Intervall der Länge 1/4 parametrisiert ist. Daher können *vier* dieser *halb* so großen Kopien aneinandergelegt werden zu einer neuen Kurve $c_2 : [0,1] \to [0,1]\times[0,1]$, mit den Ecken $c_2(0) = (0,0),\ c_2(0.25) = (0,0.5),\ c_2(0.5) = (0.5,0.5),\ c_2(0.75) = (1,0.5),\ c_2(1) = (1,0)$, wobei die vier Teilstücke nacheinander in den vier Teilquadraten $[0,0.5]\times[0,0.5],\ [0,0.5]\times[0.5,1],\ [0.5,1]\times[0.5,1],\ [0.5,1]\times[0,0.5]$ liegen. Die Kurve c_2 verbindet weiterhin die unteren Eckpunkte des Einheitsquadrates. (s. Abbildungen)
Offenbar gilt: Falls jeder Punkt aus dem Einheitsquadrat einen Abstand $\le a$ vom Bild von c_1 hat, so hat jeder der Punkte aus einem der vier Teilquadrate einen Abstand $\le \frac{1}{2}\cdot a$ von dem entsprechenden Teilstück der Kurve c_2. Ebenso sieht man:

$$|c_3(s) - c_2(s)| \le \frac{1}{2}\cdot|c_2(s) - c_1(s)|.$$

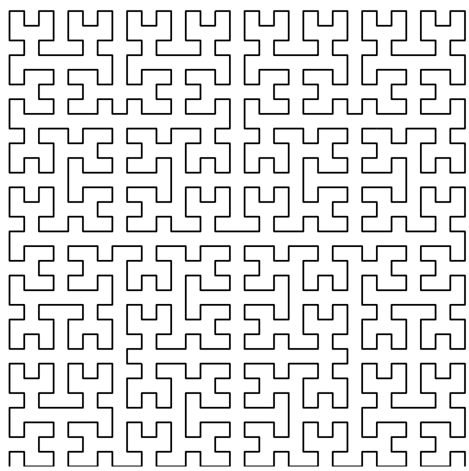
Deshalb ist die iterativ konstruierte Kurvenfolge $\{c_k\}$ gleichmäßig geometrisch majorisiert, konvergiert also gegen eine *stetige* Grenzkurve. Ferner ist *jeder* Punkt P des Einheitsquadrates Grenzwert der Folge der zu P nächsten Punkte P_k auf den Approximationen c_k. Deshalb liegt P auf der Grenzkurve, und diese füllt das ganze Einheitsquadrat aus. – Die Bilder der approximierenden Kurven werden etwas unübersichtlich, weil bei jeder Iteration mehr Doppelpunkte auftreten. Von Hilbert stammt eine Modifikation, bei der alle Approximationen injektiv sind, vgl. die folgenden Kurven, die nur aus achsenparallelen Stücken bestehen. Andererseits spricht für die Doppelpunkte, daß sie bereits Bilder der Grenzkurve sind; tatsächlich sind sogar alle Punkte der Grenzkurve Grenzwerte von diesen Doppelpunkten.



Hilbert's flächenfüllende Kurve. Hilberts Iteration einer beliebigen Kurve, die $(0,0)$ und $(1,0)$ verbindet, besteht aus vier halb so großen Kopien, die längs des gestrichelten Iterationsgerüstes aneinander gehängt sind. Das Verfahren wird mit der iterierten Kurve wiederholt.



Hilbert's injektive Approximationen verbinden die Flankenmitten der gestrichelten Hilbert-Iterationen der Startkurve $x \to \min(x, 1-x)$.



Rückblickend halten wir fest, daß alle drei Beispiele mit Hilfe gleichmäßig geometrisch majorisierter Approximationen konstruiert wurden. Dies ist ein fundamentales Verfahren zur Konstruktion neuer Funktionen, dem wir das nächste Mal bei den gewöhnlichen Differentialgleichungen begegnen werden.

Analysis via Uniform Error Bounds

Hermann Karcher, Bonn. Version May 2001

The standard analysis route proceeds as follows:

1.) Convergent sequences and series, completeness of the reals,
2.) Continuity, main theorems, trivial examples,
3.) Differentiability via limits,
4. or 5.) Integrals,
5. or 4.) Properties of limit functions via uniform convergence.

I propose the following reorganization of the material:

1.) Differentiation of polynomials, emphasis on tangent approximation,
2.) Differentiation rules, differentiation of rational functions,
3.) The monotonicity theorem via uniform estimates, before completeness,
4.) Completeness and limit functions, proofs via uniform estimates,
5.) Integrals, defined as generalized sums, computed with antiderivatives,
6.) Continuity, with less trivial examples.

Arguments **why** one would **want** such a reorganization are: The start is closer to the background of the student, one spends more time on differentiation where technical skills have to be trained, and discusses continuity when logical skills are more developed.

The main reasons **why** such a reorganization **is possible** are: The properties of limit functions follow more easily from uniform error bounds than from uniform convergence (e.g. if one has a pointwise convergent sequence of functions $\{f_n\}$ with a uniform Lipschitz bound L, then the limit function also has L as Lipschitz bound). The first such uniform error bounds follow for polynomials in a pre-analysis fashion. These bounds suffice to prove the Monotonicity Theorem before completeness. Finally, the Monotonicity Theorem shows that many interesting converging sequences indeed have uniform error estimates.

1. Derivatives of Polynomials

The elementary formula

$$x^k - a^k = (x-a) \cdot (x^{k-1} + x^{k-2}a + \ldots + a^{k-1})$$

is good enough to replace all calls on continuity. The first consequence is

$$x, a \in [-R, R] \Rightarrow |x^k - a^k| \leq kR^{k-1} \cdot |x-a|.$$

One only needs the triangle inequality to get for polynomials $P(X) := \sum_{k=0}^{n} a_k X^k$:

$$x, a \in [-R, R] \Rightarrow |P(x) - P(a)| \leq \left(\sum |a_k| k R^{k-1}\right) \cdot |x - a| =: L \cdot |x - a|.$$

In other words: *From the data which give the polynomial and from the interval on which we want to study it we can explicitly compute a Lipschitz bound.* This conforms with the common expectation (which is disappointed by continuous functions): Differences between function values, $|P(x) - P(a)|$, increase no worse than proportionally to the difference of the arguments, $|x - a|$.

The derivative controls this rough proportionality more precisely. The term $(x^{k-1}+x^{k-2}a+\ldots+a^{k-1})$, which multiplies $(x-a)$ in the above elementary formula, differs "little" from ka^{k-1} on "small" intervals around a. The first inequality above makes this precise:

$$x \in [a-r, a+r], R := |a| + r \Rightarrow$$
$$\left|(x^{k-1} + x^{k-2}a + \ldots + a^{k-1}) - ka^{k-1}\right| \leq \frac{k(k-1)}{2} R^{k-2} \cdot |x - a|$$

hence

$$\left|x^k - a^k - ka^{k-1}(x-a)\right| \leq \frac{k(k-1)}{2} R^{k-2} \cdot |x-a|^2.$$

Observe that $|x - a|^2$ is less than 1% of the argument difference $|x - a|$ if $r < 0.01$.
Again, the triangle inequality extends this to polynomials $P(X) := \sum_{k=0}^{n} a_k X^k$ and again *the error bound is explicitly computable from the data of the function and the interval in question*:

$$x \in [a-r, a+r], R := |a| + r,\ P'(a) := \sum_{k=1}^{n} a_k k a^{k-1} \Rightarrow$$
$$|P(x) - P(a) - P'(a)(x-a)| \leq \left(\sum |a_k| k(k-1)/2 R^{k-2}\right) \cdot |x-a|^2 =: K \cdot |x-a|^2.$$

Application: At interior extremal points the derivative has to vanish. (If $P'(a) > 0$ and $0 < x - a < P'(a)/K$ then $P(x) > P(a)$, etc.)

Note on Completeness. Consider the step function, which jumps from 0 to 1 at $\sqrt{2}$, but consider as its domain only the rational numbers. This function is differentiable, but **not** uniformly. On the other hand, if a function is uniformly continuous (or uniformly differentiable) on a **dense** subset of an interval, then, using completeness, one can extend the function to the whole interval without loosing continuity (or differentiability) and without changing the visible behaviour of the function. This is clearly the case for polynomials.

Moreover, the above estimates make sense in any field between $\mathbb{Q}$ and $\mathbb{R}$ which the student happens to know. Even in $\mathbb{C}$ they are useful. Therefore one can indeed discuss differentiability before completeness. – Of course completeness remains essential if one wants to define inverse functions or limit functions. I chose to discuss completeness immediately before constructing limit functions, because I view this as the more spectacular application, but I am not advertising that choice here, I am merely saying: *If one decides to work with uniform error bounds then this imposes fewer restrictions on where in the course one chooses to discuss completeness*, while uniform convergence does need prior knowledge of completeness.

2. Differentiation Rules, Derivatives of Rational Functions

Differentiation rules are ment to compute derivatives of "complicated" functions, built out of "simpler" ones, from the derivatives of the "simpler" functions. Since linear combinations, products and compositions of polynomials give again (easy to differentiate) polynomials the promise looks limited. However, there are some other functions which can be differentiated directly from their definitions, and this will broaden our possibilities considerably. We start with $x \to 1/x$ and observe $(1/x - 1/a) = (a-x)/(x \cdot a)$. While in the case of polynomials we could compute suitable error constants for *any* interval $[a-r, a+r]$ we now have to avoid division by zero. With this extra care we have:

$$0 < a/2 \le x \Rightarrow |1/x - 1/a| = \frac{|x-a|}{x \cdot a} \le \frac{2}{a^2} \cdot |x-a|$$

$$0 < a/2 \le x \Rightarrow (1/x - 1/a + \frac{x-a}{a^2}) = \frac{(x-a)^2}{x \cdot a^2} \begin{cases} \ge 0 \\ \le 2a^{-3} \cdot (x-a)^2. \end{cases}$$

Composition of polynomials with this one extra function gets us to all rational functions. Of course, the proof of the chain rule must pay attention to the distances from zeros which occur in the denominators.

In a similar way we handle the square root function. Here extra attention is needed for the domain of the function: As long as we only know rational numbers, the domain is rather thin, it contains only the squares of rational numbers (which, however, are still a dense subset). The following computation remains valid as more numbers become known.

$$0 < a/2 \le x \Rightarrow |\sqrt{x} - \sqrt{a}| = \frac{|x-a|}{\sqrt{x}+\sqrt{a}} \le \frac{1}{1.7} a^{-1/2} \cdot |x-a|$$

$$0 < a/2 \le x \Rightarrow (\sqrt{x} - \sqrt{a} - \frac{x-a}{2\sqrt{a}}) = \frac{-(x-a)^2}{2\sqrt{a}(\sqrt{x}+\sqrt{a})^2} \begin{cases} \le 0 \\ \ge 0.2a^{-3/2} \cdot (x-a)^2. \end{cases}$$

If we compose this function with polynomials then the discussion of the domain quickly becomes unmanagable; the computation shows that we know, even with error bounds, what the derivative of the square root function has to be, well before we know enough about numbers to be able to use this function freely.

All known proofs of the differentiation rules have the following property:
Given two differentiable functions f, g we abbreviate their tangent functions as
$l_f(x) := f(a) + f'(a)(x-a),\ l_g(x) := g(a) + g'(a)(x-a)$.
Then, if we assume that the differences $f - l_f,\ g - l_g$ are "small", then also the differences
$(\alpha \cdot f + \beta \cdot g) - (\alpha \cdot l_f + \beta \cdot l_g), \quad f \cdot g - l_f \cdot l_g, \quad f \circ g - l_f \circ l_g$
are in the same sense "small".
To prove something we have to specify "small". Because of the functions known so far we may say: $f - l_f$ is "**small**" means, there exists an interval $[a-r,\ a+r]$ and a constant K such that

$$x \in [a-r,\ a+r] \Rightarrow |f(x) - l_f(x)| \le K \cdot |x-a|^2.$$

Later $f - l_f$ is "**small**" may have a more subtle interpretation:
For every $\epsilon > 0$ there exists $\delta > 0$ such that

$$x \in [a-\delta,\ a+\delta] \Rightarrow |f(x) - l_f(x)| \le \epsilon \cdot |x-a|.$$

I find it important to emphasize that the form of the tangent approximation and the form of the differentiation rules **do not depend on the specification of "small"**. Moreover, even the strategy of the proofs does not depend on what exactly we mean by "small": if we change the definition of "small" in the assumptions, we can follow these changes through the proofs and end up with the changed conclusion. This possibility, to repeat the proofs under slightly changed assumptions, allows one to emphasize what is essential in these proofs.

3. Monotonicity Theorem and Related Results

A fundamental fact of analysis is that rough information about the derivative f' of a function f allows to deduce sharper information about f. As an example, take Lipschitz bounds:

$$\text{If} \qquad x \in [\alpha, \omega] \Rightarrow |f'(x)| \le L \qquad \text{then}$$
$$a, b \in [\alpha, \omega] \Rightarrow |f(b) - f(a)| \le L \cdot |b-a|.$$

For functions with uniform tangent approximations, i.e. for all the functions discussed so far, this fundamental theorem can be proved without invoking (even before discussing) the completeness of the reals. The only tool needed is Archimedes Principle, which is obvious for the rationals and an axiom for the reals:

Archimedes Principle

$$\text{If } 0 \leq r \text{ and } r \leq \frac{1}{n} \text{ for all } n \in \mathbb{N} \text{ then } r = 0.$$

The most intuitive result from the family of theorems that exploit derivative information probably is the

Monotonicity Theorem
Main assumption:

$$x \in [\alpha, \omega] \Rightarrow f'(x) \geq 0.$$

Technical assumption replacing completeness: The function f can be **uniformly** approximated by its derivatives, i.e. there exist positive constants r, K such that

$$x, a \in [\alpha, \omega], |x - a| \leq r \Rightarrow |f(x) - f(a) - f'(a)(x - a)| \leq K \cdot |x - a|^2.$$

Then f is nondecreasing:

$$a, b \in [\alpha, \omega], a < b \Rightarrow f(a) \leq f(b).$$

Note: The uniformity in the technical assumption is the crucial part, not the quadratic error bound; if one assumes
For each $\epsilon > 0$ exists a $\delta > 0$, which can be chosen **independent** of a, such that

$$x, a \in [\alpha, \omega], |x - a| \leq \delta \Rightarrow |f(x) - f(a) - f'(a)(x - a)| \leq \epsilon \cdot |x - a|,$$

then the strategy of the following proof also works.

Proof. For any $x, y \in [\alpha, \omega],\ x < y$ with $|x-y| \leq r$ we have (because of the two assumptions of the theorem):

$$-K \cdot |x - y|^2 \leq f(y) - f(x) - f'(x)(y - x) \leq f(y) - f(x).$$

Apply this to sufficiently short subintervals $[t_{j-1}, t_j]$ of the interval $[a, b]$, i.e. put $t_j := a + j/n \cdot (b - a), 0 \leq j \leq n$, with $(b - a)/n \leq r$ (Archimedes!) and have:

$$-K(b - a)^2/n^2 \leq f(t_j) - f(t_{j-1}),\ j = 1 \dots n,$$

then sum for $j = 1 \dots n$:

$$-K(b - a)^2/n \leq f(b) - f(a).$$

Archimedes Principle improves this to the desired claim $0 \leq f(b) - f(a)$. (The previous inequality changes under the ϵ-δ-assumption to $-K(b - a) \cdot \epsilon \leq f(b) - f(a)$, which also implies the theorem.)

Immediate consequences are:

Generalized monotonicity:

$$f' \leq g',\ a < b \Rightarrow f(b) - f(a) \leq g(b) - g(a)$$

Explicit bounds:

$$m \leq f' \leq M,\ a < b \Rightarrow m \cdot (b-a) \leq f(b) - f(a) \leq M \cdot (b-a)$$

Multiplicative version:

$$0 < f, g,\quad \frac{f'}{f} \leq \frac{g'}{g},\ a < b \Rightarrow (\frac{g}{f})' = \frac{g}{f}(\frac{g'}{g} - \frac{f'}{f}) \geq 0 \Rightarrow \frac{f(b)}{f(a)} \leq \frac{g(b)}{g(a)}$$

Iterated application to second derivatives:

$$\begin{aligned} |f''| \leq B,\ a < x\ &\Rightarrow -B(x-a) \leq f'(x) - f'(a) \leq B(x-a) \\ &\Rightarrow \frac{-B}{2}(x-a)^2 \leq f(x) - f(a) - f'(a)(x-a) \leq \frac{B}{2}(x-a)^2. \end{aligned}$$

These are strong improvements over the error bounds which were initially computed from the coefficients of polynomials. As illustration consider the Taylor polynomials $T_n(X)$ for a (not yet constructed) function with $f' = -f,\ f(0) = 1$:

$$T_n(x) := \sum_{k=0}^{n} (-1)^k \frac{x^k}{k!},\ 0 \leq x \leq 1,\ T_n'(x) = -T_{n-1}(x).$$

Since this is a Leibniz series we have the nested intervals:

$$[1-x, 1] \supset [T_1(x), T_2(x)] \supset [T_3(x), T_4(x)] \supset \ldots \supset [T_{2n-1}(x), T_{2n}(x)]$$

hence in particular: $|T_n''(x)| \leq 1$. The "second derivative consequence" of the monotonicity theorem now implies uniform bounds for polynomials of **arbitrarily large** degree:

$$x, a \in [0,1] \Rightarrow |T_n(x) - T_n(a) + T_{n-1}(a)(x-a)| \leq \frac{1}{2}(x-a)^2.$$

Clearly, if we only had the *existence* of a limit function T_∞ then Archimedes Principle would imply without further words differentiability and derivative of the limit function:

$$x, a \in [0,1] \Rightarrow |T_\infty(x) - T_\infty(a) + T_\infty(a)(x-a)| \leq \frac{1}{2}(x-a)^2.$$

4. Completeness and Limits of Sequences of Functions

At most occasions I have prefered to begin the discussion of completeness with nested intervals; immediate applications are Leibniz series like the just mentioned Taylor polynomials. Thus we have interesting limit functions together with their derivatives right from the start.

Cauchy sequences are the second step; the most important applications are sequences or series which are dominated by the geometric series (i.e. power series and the contraction lemma). Existence of *sup* and *inf* for bounded nonempty sets gives the most compact formulation for dealing with completeness. I have used the standard text book arguments, only the applications change since the results of the earlier sections are of significant technical help.

Note. Since so much emphasis was placed on computable error bounds I mention that the theorem "*Monotone increasing, bounded sequences converge*" is a very significant exception. In my opinion this intuitively desirable theorem is a major justification of the standard limit definition: Because the convergence speed of monotone increasing, bounded sequences can be slowed down arbitrarily (namely by repeating the elements of the sequence more and more often) one **cannot** prove the monotone sequence theorem with any version of a limit definition which requires more explicit error control than the standard definition. By contrast, in numerical analysis one tries to use sequences which converge at least as **fast** as some geometric sequence, i.e. one has more explicit control like $|a_{n+p} - a_n| \leq C \cdot q^n$.

I illustrate how the monotonicity theorem often allows short replacements of standard induction proofs.

Bernoulli's inequality is a version of the monotonicity theorem:
$f(x) := (1+x)^n \geq 1 + n \cdot x = f(0) + f'(0) \cdot x$, since $f''(x) \geq 0$ for $x \geq -1$.

The monotonicity of $n \to f_n(x) := (1 + x/n)^n,\ 0 \leq x$ and the decreasing of $n \to g_n(x) := (1 - x/n)^{-n},\ 0 \leq x < n$ requires no technical skill since $(f_n'/f_n)(x) = 1/(1+x/n) \leq (f_{n+1}'/f_{n+1}(x) \leq 1 \leq 1/(1-x/n) = (g_n'/g_n)(x) \leq (g_{n-1}'/g_{n-1})(x)$.

The following are convenient estimates of the geometric series and its derivatives which give the desired uniform constants when dealing with power series:

$$0 \leq x < 1, \qquad \sum_{k=0}^{n} x^k = \frac{1-x^{n+1}}{1-x} \leq \frac{1}{1-x}$$

$$\left(\sum_{k=0}^{n} x^k\right)' = \sum_{k=1}^{n} k x^k = \frac{-(n+1)x^n}{1-x} + \frac{1-x^{n+1}}{(1-x)^2} \leq \frac{1}{(1-x)^2}$$

$$\left(\sum_{k=0}^{n} x^k\right)'' = \frac{-(n+1)nx^{n-1}}{1-x} - \frac{2(n+1)x^n}{(1-x)^2} + 2\frac{1-x^{n+1}}{(1-x)^3} \leq \frac{2}{(1-x)^3}.$$

As mentioned, the standard proof of the Contraction Lemma uses explicit error bounds and our limit arguments are in the same spirit: Functions $f : M \to M$ with $|f'| \leq q < 1$

are contracting, $|f(x) - f(y)| \leq q \cdot |x - y|$. And for contracting maps sequences generated by iteration, $a_{n+1} := f(a_n)$, are geometrically dominated and hence Cauchy:

$$|a_{n+p} - a_n| \leq \frac{|a_1 - a_0|}{1 - q} \cdot q^n.$$

An example of a contracting rational map with the irrational golden ratio as fixed point is

$$f(x) := 1/(1 + x),\ f : [1/2, 1] \to [1/2, 1],\ |f'(x)| \leq | - (1 + x)^{-2}| \leq 4/9.$$

The approximating sequence $a_0 = 1, a_1 = 1/(1+1), \ldots, a_n = 1/(1+1/(1+1/\ldots))$ consists of the optimal approximations from continued fractions. This is to show that sequences do appear in my approach, but they are handled with the monotonicity theorem and not presented as the road to differentiation.

The main question about limit functions is, of course, "What is their derivative?" Usually one employs uniform convergence of the differentiated sequence and the main theorem connecting differentiation and integration. I illustrate the use of uniform error bounds in the case of power series $P_n(X) := \sum_{k=0}^{n} a_k X^k$. The basic tool in both approaches is the comparison with a geometric series and it will better bring out the differences if I add an assumption which simplifies either approach: a bound for the coefficients, $|a_k| \leq C$.
Then we have for all x with $|x| \leq q < 1$

$$|P_{n+m}(x) - P_n(x)| \leq \sum_{n<k} |a_k| q^k \leq \frac{C}{1 - q} q^{n+1},$$

which shows that $\{P_n(x)\}$ is a Cauchy sequence (in fact uniformly for $|x| \leq q$). For the first derivatives we obtain uniform bounds by comparing with the geometric series

$$|P_n'(x)| \leq \sum_{k\geq 1} |a_k| k q^{k-1} \leq \frac{C}{(1 - q)^2} =: L.$$

The Monotonicity Theorem implies **for all** n the Lipschitz bound

$$|x|, |y| \leq q \Rightarrow |P_n(y) - P_n(x)| \leq L \cdot |y - x|$$

and Archimedes Principle extends this uniform estimate to the limit function

$$|x|, |y| \leq q \Rightarrow |P_\infty(y) - P_\infty(x)| \leq L \cdot |y - x|.$$

Similarly we employ our estimate of the geometric series to get uniform bounds for the second derivatives

$$|P_n''(x)| \leq \sum_{k\geq 2} |a_k| k(k - 1) q^{k-2} \leq \frac{C}{(1 - q)^3} =: K,$$

we use the Monotonicity Theorem to get uniform tangent approximations

$$|x|, |y| \le q \Rightarrow |P_n(y) - P_n(x) - P_n'(x)(y-x)| \le K \cdot |y-x|^2,$$

and Archimedes Principle again extends these uniform bounds to the limit

$$|x|, |y| \le q \Rightarrow |P_\infty(y) - P_\infty(x) - \lim_{n\to\infty} P_n'(x)(y-x)| \le K \cdot |y-x|^2.$$

This proof of the differentiability and the determination of the derivative of the limit function clearly extends to complex power series, a first step to higher dimensional analysis. Another immediate extension is to the differentiation of curves $c := (c_1, c_2, c_3) : [a,b] \to \mathbb{R}^3$, which is important in itsself but also a prerequisite for analysis in $\mathbb{R}^n$.

5. Integrals, Riemann Sums, Antiderivatives

Some notion of tangent (and hence some version of derivative) was known centuries before Newton. Similarly, summation of infinitesimals had already troubled the Greeks, and Archimedes determination of the area bounded by a parabola and a secant was definite progress. Against the background of this early knowledge I find the conceptual progress achieved with the definition of the integral even more stunning than that achieved with differentiation. I try to teach integrals as a fantastic generalization of sums, they allow for example to "continuously sum" the velocity of an object in order to obtain the distance which it travelled. With this goal in mind I think it is fundamental that Riemann sums of a function f can be computed up to controlled errors if one knows an antiderivative F of the given f, i.e. $F' = f$. I use the standard definition of the integral in terms of Riemann sums. I add a construction of an antiderivative F for a continuous f which is in the spirit of uniform error control, but from now on the differences to the standard approach are not very pronounced.

Let $f : [a,b] \to \mathbb{R}^3$ be given. Consider a subdivision of $[a,b]$, i.e. $a = t_0 < t_1 < \ldots < t_n = b$ and choose intermediate points $\tau_j \in [t_{j-1}, t_j],\ j = 1 \ldots n$. For these data we **define** the

Riemann Sum

$$\mathfrak{RS}(f) := \sum_{j=1}^{n} f(\tau_j)(t_j - t_{j-1}).$$

Now, if f is at least continuous and $F' = f$ then the difference $|F(b) - F(a) - \mathfrak{RS}(f)|$ is "small"; how "small" depends on the precise assumption for f, for example a Lipschitz bound, $x, y \in [a,b] \Rightarrow |f(y) - f(x)| \le L \cdot |y-x|$, implies the

Error Bound

$$|F(b) - F(a) - \mathfrak{RS}(f)| \le L \cdot (b-a) \cdot \max_{1\le j\le n} |t_j - t_{j-1}|.$$

This estimate (or other versions) follows from the Monotonicity Theorem and the triangle inequality; first note the derivative bound:

$$x \in [t_{j-1}, t_j] \Rightarrow |(F(x) - f(\tau_j) \cdot x)'| = |f(x) - f(\tau_j)| \leq L \cdot |x - \tau_j| \leq L \cdot |t_j - t_{j-1}|,$$

hence

$$|F(t_j) - F(t_{j-1}) - f(\tau_j)(t_j - t_{j-1})| \leq L \cdot |t_j - t_{j-1}|^2,$$

so that summation over $j = 1 \dots n$ gives the claimed error bound.

Continuity (see below) of f implies with the same arguments an ϵ-δ-error bound:

$$\max_{1 \leq j \leq n} |t_j - t_{j-1}| \leq \delta \Rightarrow |F(b) - F(a) - \mathfrak{RS}(f)| \leq \epsilon \cdot (b - a).$$

I prefer to define the integral for vector valued functions (for direct application to integrals of velocities). The notion of limit has to be generalized to cover "convergence" of Riemann sums: For every $\epsilon > 0$ there is a subdivison of $[a, b]$ such that **for all finer subdivisions** and all choices of intermediate points τ_j the corresponding Riemann sums differ by less than ϵ. Almost the same argument as used to prove the error bound gives: For continuous (or better) f the Riemann sums converge; the limit is called the integral of f over $[a, b]$, notation $\int_a^b f(x)dx$. Moreover, if $F' = f$ then the error bound proves

$$F(b) - F(a) = \int_a^b f(x)dx.$$

The triangle inequality for Riemann sums gives the **Triangle Inequality for Integrals**

$$a < b \Rightarrow \left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|dx.$$

Also other analogies between sums and integrals need to be discussed.
I describe a construction of an antiderivative which is independent of the definition of the integral (its proof, based on uniformity, works in Banach spaces).

Theorem. Let f be Lipschitz continuous on $[a, b]$ (soon: uniformly continuous). Then one can approximate f uniformly by piecewise linear "secant" functions s_n. These have piecewise quadratic antiderivatives and a limit of these is an antiderivative of f. In more detail:

$$s_n(t_j) := f(t_j),\ j = 0, \dots, n,\ (a = t_0 < t_1 < \dots < t_n = b)$$

$$s_n(x) := \frac{f(t_{j-1}) \cdot (t_j - x) + f(t_j) \cdot (x - t_{j-1})}{t_j - t_{j-1}} \text{ for } x \in [t_{j-1}, t_j]$$

$$|f(x) - s_n(x)| \leq L \cdot \max |t_j - t_{j-1}| =: r_n$$

Obviously, s_n has a piecewise quadratic antiderivative $S_n, S_n^{'}(x) = s_n(x), S_n(0) = 0$. From the Monotonicity Theorem and

$$|S_{n+m}^{'}(x) - S_n^{'}(x)| \leq 2r_n$$

we have the (uniform) Cauchy property:

$$|S_{n+m}(x) - S_n(x)| \leq 2(b-a)r_n,$$

which, by completeness, gives a limit function S_∞. But we again have uniform error bounds:

$$|(S_n(x) - S_n(c) - s_n(c)(x-c))'| = |s_n(x) - s_n(c)| \leq L \cdot |x-c|,$$

hence

$$|S_n(x) - S_n(c) - s_n(c)(x-c)| \leq L \cdot |x-c|^2.$$

Archimedes Principle gives the final result:

$$|S_\infty(x) - S_\infty(c) - f(c)(x-c)| \leq L \cdot |x-c|^2,$$

which says that S_∞ is differentiable and $S'_\infty = f$, as claimed.

Note in this proof: The approximation $S_n(b) - S_n(a)$ for $S_\infty(b) - S_\infty(a) = \int_a^b f(x)dx$ is a frequently used numerical approximation for the integral of f.

6. Continuity, Theorems and Examples

The arguments in this last section are the standard arguments. My point is that with the experience of the previous sections one can achieve a better understanding of continuity and, moreover, this takes less time than a treatment of continuity in the early parts of a course. Since convergent sequences were such an essential tool (for getting limit functions) we ask: "What kind of functions are compatible with convergence?" We define:

$f : A \subset \mathbb{R}^d \to \mathbb{R}^e$ is called **sequence continuous** at $a \in A$
if **every** sequence $a_n \in A$ which converges to the limit $a \in A$ has its image sequence $f(a_n)$ converging to $f(a)$.

We see that linear combinations, products, compositions of sequence continuous functions are (directly from the definition) sequence continuous. But $1/f$ causes a problem:

If $f(a) \neq 0$ then we would like to find an interval $[a-\delta, a+\delta]$ on which f **is not zero**.

If such an interval could be found then on it $1/f$ would clearly be sequence continuous. It is well known that such a $\delta > 0$ can only be found with an indirect proof. But this proof is essential for understanding continuity. It is also very similar to the equivalence of sequence continuous and ϵ-δ-continuous.

Definition: f is ϵ-δ-continuous at a if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$x \in [a - \delta, a + \delta\,] \Rightarrow |f(x) - f(a)| \leq \epsilon.$$

It remains to prove the main theorems and to show examples. First I give short summaries of the proofs to recall what kind of arguments are involved. 1.) The intermediate value theorem: By interval halfing construct a Cauchy sequence which converges to a preimage of the given intermediate value. 2.) Boundedness on complete, bounded sets: In an indirect proof construct by interval halfing a Cauchy sequence on which the function f is unbounded, a contradiction to the continuity at the limit point of the Cauchy sequence. 3.) Extremal values are assumed on complete, bounded sets: Since the function is bounded by the previous result we have $\sup f$ and $\inf f$; with interval halfing we find Cauchy sequences $\{a_n\}$ such that the sequences of values, $\{f(a_n)\}$, converge to $\sup f$ resp. $\inf f$. 4.) Uniform continuity on complete, bounded sets: Necessarily indirect, if for some $\epsilon^* > 0$ no $\delta > 0$ is good enough then we find a pair of Cauchy sequences $\{a_n\}, \{b_n\}$ which have the **same** limit but $|f(a_n) - f(b_n)| \geq \epsilon^*$, a contradiction to the continuity at the common limit point. 5.) Uniformly convergent sequences of continuous functions have a continuous limit function: To given $\epsilon > 0$ choose an $\epsilon/3$-approximation from the sequence and for this (continuous!) approximation find $\delta > 0$ for $\epsilon/3$-deviations; this δ guarantees with the triangle inequality at most ϵ-deviations for the limit function.

From the construction of examples I recall that comparison with a geometric sequence is the main tool. I find it misleading to call the following examples "weird", as if continuity "really" were much more harmless. 1.) The polygonal approximations of Hilbert's cube filling curve are explicitly continuous: To guarantee value differences $\leq 2^{-n}$ the arguments have to be closer than 8^{-n}; Archimedes Principle concludes the same for the limit function. 2. Similarly for Cantor's staircase, a monotone increasing continuous function which is differentiable with derivative 0 except on a set of measure zero: The piecewise linear approximations satisfy: To guarantee value differences $\leq 2^{-n}$ the arguments have to be closer than 3^{-n} and Archimedes Principle does the rest. 3.) And continuous but nowhere differentiable functions can be obtained as obviously uniform limits of sums of faster and faster oscillating continuous functions like $\sum_k 2^{-k} \sin(8^k x)$.

In view of the much better-than-continuos properties of all the functions we have so far been concerned with, I find it not so obvious how to demonstrate the usefulness of the notion of continuity. The proof of the existence of solutions of ordinary differential equations based on the Contraction Lemma in the complete metric space of **continuous** functions (sup-norm) is my lowest level convincing example.

Since the Monotonicity Theorem and its consequences are usually derived from the (essentially one-dimensional) Mean Value Theorem of Differentiation, I finally show that the theorems which conclude from assuming derivative bounds can be obtained with shorter proofs (again valid in Banach spaces).

Theorem: Derivative bounds are Lipschitz bounds.

More precisely, let $A \subset \mathbb{R}^d$ be a convex subset and let $F : A \to \mathbb{R}^e$ be ϵ-δ-differentiable, and for emphasis: without any uniformity assumption. Assume further a bound on the derivative: $x \in A \Rightarrow |TF(x)| \le L$. Then

$$a, b \in A \Rightarrow |F(a) - F(b)| \le L \cdot |a - b|.$$

Indirect proof via halfing. If the inequality were not true then we had some $a, b \in A$ with $|F(a) - F(b)| > L \cdot |a - b|$, i.e. with some fixed $\eta > 0$ we had $|F(a) - F(b)| \ge (L + \eta) \cdot |a - b|$. Let $m := (a + b)/2$ be the midpoint. Then for either a, m or m, b the same inequality has to hold (since otherwise, by the triangle inequality, $|F(a) - F(b)| < (L + \eta) \cdot |a - b|$). In other words, we have a_1, b_1 with half the distance $|a_1 - b_1| = |a - b|/2$, but still

$$|F(a_1) - F(b_1)| \ge (L + \eta) \cdot |a_1 - b_1|.$$

This procedure can be repeated, we get a pair of Cauchy sequences $\{a_n\}, \{b_n\}$ with the same limit c between a_n and b_n on the closed segment from a to b, but with the inequalities:

$$|F(a_n) - F(b_n)| \ge (L + \eta) \cdot |a_n - b_n|.$$

By differentiability of F at $c \in A$ we have for $\epsilon = \eta/2$ a $\delta > 0$ such that

$$\begin{aligned} x \in A, |x - c| \le \delta \Rightarrow |F(x) - F(c) - TF|_c \cdot (x - c)| &\le \epsilon \cdot |x - c| \\ \Rightarrow |F(x) - F(c)| &\le (L + \epsilon) \cdot |x - c|. \end{aligned}$$

Choose n so large that $|a_n - c|, |b_n - c| \le \delta$ so that the last inequality holds for $x = a_n$ and $x = b_n$. Add the inequalities and observe $|a_n - c| + |b_n - c| = |a_n - b_n|$ to obtain the contradiction

$$(L + \eta) \cdot |a_n - b_n| \le |F(a_n) - F(b_n)| \le (L + \eta/2) \cdot |a_n - b_n| < (L + \eta) \cdot |a_n - b_n|.$$