Argumente der Analysis

H. Karcher, Version 30.9.99

Dieser Text ist eine Kurzfassung des Analysisteils meiner Vorlesung Mathematik I im WS1999/2000. Es ist beabsichtigt, daß die Darstellung für Übungsleiter ausführlich genug ist.
Für Erstsemester ist der Text als roter Faden gedacht. Eventuell als fehlend empfundene Zwischenrechnungen werden in der Vorlesung ausgeführt.

Kapitel

0. Vorwort

Dies Manuskript ist aus Erfahrungen mit meinen Anfängervorlesungen entstanden. Ziele sind: Die Begriffe der Analysis möglichst gut an die Vorkenntnisse der Lernenden anzubinden und dann die erstaunliche Leistungsfähigkeit des typischen Argumentierens mit diesen Begriffen darzustellen; 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 solche nur approximierbaren 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 kann an den Polynomen erworben werden und zwar aus zwei Gründen: Erstens kann man deren Ableitungen definieren und explizit berechnen, ohne daß man weitere Hilfsbegriffe der Analysis benötigt (Kapitel 1), und zweitens kann man eine wichtige Umkehrung des Ableitens beweisen noch ehe man mehr reelle Zahlen als die rationalen Zahlen kennt (Kapitel 2). Diese beiden Kapitel werden auch zur Aufarbeitung der Vorkenntnisse der Studienanfänger benutzt. Am Ende haben wir 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 3). 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

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

und schon Archimedes hat formalisiert, wie daraus dieselbe Ungleichung für die Grenzfunktion folgt. Mit anderen Worten: Der Monotoniesatz für rationale Funktionen erlaubt, die Grenzfunktionen zu behandeln, ehe gleichmäßige Konvergenz zur Verfügung steht.

Die komplexen Zahlen sind innerhalb und außerhalb der Mathematik fundamental. In Kapitel 4 wird der Inhalt der Kapitel 1-3 für komplexe Zahlen wiederholt, diesmal mit Einsatz der Vollständigkeit von $\mathbb{R}$.

Neben Anwendungen des Monotoniesatzes ist "das" andere Werkzeug für Konvergenzbeweise die Majorisierung durch die geometrische Reihe. Dies wird in Kapitel 5 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 6 der Hauptsatz der Differential- und Integralrechnung bewiesen.
In Kapitel 7 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 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 sonst 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...

1. Differenzieren von Polynomen

Ziel des Abschnitts: Bei Potenzfunktionen kann man mit Monotonieargumenten ausrechnen, was man als Steigung bezeichnen muß, ehe man mathematische Definitionen gemacht hat und ehe man die Vollständigkeit der reellen Zahlen zur Verfügung hat. Zum

Begriff Steigung gehört der Begriff Tangente. Diese neuen Tangenten können mit Kreistangenten verglichen werden: Der Graph einer Potenzfunktion verläuft an jeder Stelle a zwischen der Tangente bei a und einem von der anderen Seite berührenden Kreis. Dieses anschauliche Ergebnis erlaubt, Tangenten durch Ungleichungen zu beschreiben; diese Beschreibung dehnt sich auf Polynome aus.

Polynome $P(x) = a_0 + a_1 x + \ldots + a_n \cdot x^n$ sind Funktionen, deren Werte so leicht berechnet werden können, daß die für die Analysis typischen Argumente mit Grenzwerten und der Vollständigkeit der reellen Zahlen eine untergeordnete Rolle spielen. Insbesondere können auch Ableitungen und Integrale explizit ausgerechnet werden. Ehe wir jedoch neue Funktionen mit speziellen Ableitungseigenschaften konstruieren können, müssen wir wissen, was Ableitungen bei den Polynomen bedeuten und was für Aussagen über Polynome man insbesondere mit dem Monotoniesatz beweisen kann. Außderdem muß ich gewisse Rechenfertigkeiten üben, damit das Verständnis der Analysisargumente nicht durch Probleme mit den Vorkenntnissen erschwert wird. Nur bei den Polynomen werden Diskussionen, die zur Definition der Ableitung führen, eine Rolle spielen. Später ist die Definition der Ableitung zuerst da und leitet dann die Konstruktion neuer Funktionen.

Wir veranschaulichen Polynomfunktionen durch ihre Graphen, Polynome vom Grad ≤ 1 haben *Geraden* als Graphen; die Steigung dieser Geraden heißt Ableitung des Polynoms. Für Polynome höheren Grades kennen wir die Durchschnittssteigung oder Sekantensteigung m über einem Intervall $[a, b]$:

$$m = \frac{P(b) - P(a)}{b - a}.$$

Unser Problem besteht nun darin, die Steigung in einem Punkt zu definieren - die Formel macht ja für $b = a$ keinen Sinn. Das Problem ist in physikalischer Einkleidung noch plausibler: Ein Punkt P befinde sich zum Zeitpunkt t an der Stelle $p(t)$, wobei man sich die Bahn $t \to p(t)$ nach Belieben in $\mathbb{R}$, $\mathbb{R}^2$ oder $\mathbb{R}^3$ vorstellen kann. Dann sind die Quotienten $(p(t_2) - p(t_1))/(t_2 - t_1)$ *Durchschnitts*geschwindigkeiten, und das Problem besteht darin, *Momentangeschwindigkeiten* zu jedem Zeitpunkt zu definieren.
Wir betrachten den Kreis unter diesem Gesichtspunkt. Die Formulierung: "Die Tangente steht senkrecht auf dem Radius" kann nicht für andere Kurven übernommen werden. Stattdessen vergleichen wir Sehnensteigungen mit der Tangentensteigung im tiefsten Punkt: Alle Sehnensteigungen im rechten unteren Viertelkreis sind positiv und im linken unteren Viertelkreis negativ, die Steigung 0 der Tangente im tiefsten Punkt liegt dazwischen. Diese Eigenschaft kann für reine Potenzen, $P(x) = x^n$, mit Hilfe der Summe der endlichen geometrischen Reihe $(1 + q + \ldots + q^{n-1}) \cdot (1 - q) = 1 - q^n$ gezeigt werden. Die Durchschnittssteigung im Intervall $[a, b]$ ist:

$$\frac{b^n - a^n}{b - a} = b^{n-1} + b^{b-2} \cdot a + \ldots + b \cdot a^{n-2} + a^{n-1}, \quad q = \frac{a}{b}.$$

Die folgenden Ungleichungen sind übersichtlicher, wenn wir zunächst annehmen $0 \le a < b$, damit haben wir:

$$0 \le a < b \quad \Rightarrow \quad n \cdot a^{n-1} < \frac{b^n - a^n}{b - a} < n \cdot b^{n-1}$$

Diese Ungleichung besitzt folgende Interpretationen: Bei festem a sind *alle rechtsseitigen* (also $b > a$) Durchschnittsgeschwindigkeiten $> n \cdot a^{n-1}$, bei festem b sind *alle linksseitigen* Durchschnittssteigungen $< n \cdot b^{n-1}$; wenn das Intervall $[a, b]$ "klein" ist, liegen diese unteren und oberen Schranken für die Sehnensteigungen $(b^n - a^n)/(b - a)$ "nahe" beieinander. Wir erklären "nahe" noch genauer. Dazu benutzen wir die Ungleichung noch einmal, für $n - 1$ statt n:

$$0 \le a < b \quad \Rightarrow \quad 0 < b^{n-1} - a^{n-1} < (n-1) \cdot b^{n-2} \cdot (b - a).$$

Beide Ungleichungen zusammen ergeben:

$$0 \le a < b \quad \Rightarrow \quad 0 < \frac{b^n - a^n}{b - a} - n \cdot a^{n-1} < n \cdot (n-1) \cdot b^{n-2} \cdot (b - a).$$

Die sprachliche Zusammenfassung wird übersichtlicher, wenn wir noch eine obere Schranke R für die Zahlen a, b benutzen und damit die Ungleichung vergröbern:
Satz: Es sei R eine feste Zahl und $0 \le a < b \le R$. Dann gilt

$$0 < \frac{b^n - a^n}{b - a} - n \cdot a^{n-1} < n \cdot (n-1) \cdot R^{n-2} \cdot (b - a).$$

Interpretation: Der Unterschied zwischen den Durchschnittssteigungen $(b^n - a^n)/(b - a)$ und der Zahl $n \cdot a^{n-1}$ ist abgeschätzt durch $Konstante \times Intervallänge = (n \cdot (n-1) \cdot R^{n-2}) \times |b - a|$. Das veranlaßt die

Definition: Die Steigung der Graphen der Potenzfunktion $x \to P(x) = x^n$ im Punkte a ist $n \cdot a^{n-1}$. Diese Zahl heißt auch Ableitung von P bei a. Bezeichnung: $P'(a) = n \cdot a^{n-1}$.
Zusatz: Die lineare Funktion $T(x)$, die bei a denselben Wert und dieselbe Steigung wie P hat, also $T(x) = P(a) + P'(a) \cdot (x - a)$, heißt Tangente oder beste lineare Approximation von P.

Rechtfertigung: Die Bezeichnung "beste lineare Approximation" wird dadurch gerechtfertigt, daß, wegen des eben bewiesenen Satzes, der Unterschied zwischen Werten der Funktion und der Tangente schneller als proportional zu $|x - a|$ klein wird, nämlich:

$$\begin{aligned}&x, a \in [-R, R] \Rightarrow \\ &|P(x) - T(x)| = |x^n - a^n - n \cdot a^{n-1} \cdot (x - a)| \le (n \cdot (n-1) \cdot R^{n-2}) \cdot |x - a|^2.\end{aligned}$$

(Wir heben ab jetzt die Beschränkung auf positive Intervalle auf.)

Die Benutzung der vom Kreis her bekannten Vokabel "Tangente" wird dadurch gerechtfertigt, daß der Graph der Potenzfunktion an jeder Stelle $(a, P(a))$ zwischen einem Kreis und

dessen Tangente hindurchläuft. Ich zeige das nur für $a, x \geq 0$. Erstens, der Graph liegt oberhalb der Tangente wegen
$0 \leq a, x \Rightarrow P(x) - T(x) = (x^{n-1} + x^{n-2} + ... + xa^{n-2} + a^{n-1} - na^{n-1}) \cdot (x - a) \geq 0$,
denn der erste Faktor hat dasselbe Vorzeichen wie der zweite, $(x - a)$.
Zweitens, einen Kreis mit derselben Tangente aber auf der anderen Seite des Graphen bestimmen wir zunächst für die Normalparabel (vgl. die folgende Abbildung):
Die zu der Tangente im Berührpunkt senkrechte Gerade nennen wir die Normale; diese beiden Geraden sind:

$$T(x) = P(a) + P'(a) \cdot (x - a), \quad N(x) = P(a) - \frac{1}{P'(a)} \cdot (x - a).$$

Speziell für die quadratische Parabel $P(x) = x^2$ vereinfacht sich das zu

$$T(x) = -a^2 + 2a \cdot x, \quad N(x) = +(a^2 + \frac{1}{2}) - \frac{1}{2a} \cdot x.$$

Der Kreis um den Schnittpunkt $(0, a^2 + \frac{1}{2})$ der Normale mit der y-Achse und durch $(a, P(a))$ hat die Gleichung

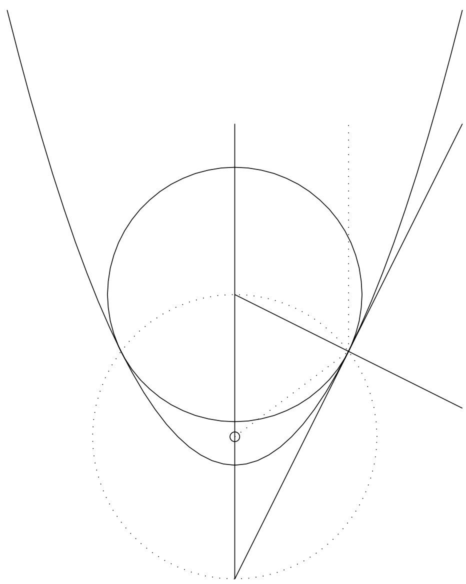
$$(x - 0)^2 + (y - m)^2 = r^2, \quad m = a^2 + \frac{1}{2}, \; r^2 = a^2 + \frac{1}{4}.$$

Er hat in $(a, P(a))$ dieselbe Tangente T wie der Graph von P. Der Graph von P liegt außerhalb dieses Kreises, denn: Ein Punkt (x, y) liegt außerhalb des Kreises mit Mittelpunkt (m_1, m_2) und Radius r, wenn gilt

$$(x - m_1)^2 + (y - m_2)^2 - r^2 \geq 0.$$

Also mit $y = x^2$ und m, r wie eben

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



Parabel zwischen Kreis und Tangente. Brennpunktkonstruktion punktiert.

Für höhere Potenzen $n > 2$ werden die Kreise um den Schnittpunkt der Normalen mit der y-Achse zu groß, wenn a nahe Null ist. Als Rechtfertigung für die Bezeichnung Tangente genügt allerdings der Fall $n = 2$, denn der bewiesene Satz besagt, daß der Graph von x^n bei a zwischen zwei quadratischen Parabeln mit derselben Tangente $T(x)$ verläuft:

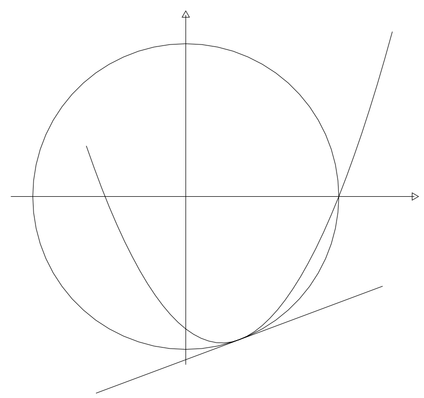
$$x, a \in [-R, R] \Rightarrow$$
$$-n(n-1)R^{n-2} \cdot (x-a)^2 \leq x^n - a^n - n \cdot a^{n-1} \cdot (x-a) \leq +n(n-1)R^{n-2} \cdot (x-a)^2.$$

Anwendungsbeispiel:

Ich finde wichtig, daß die neue Definition sofort neuartige Erklärungen ermöglicht, z.B. die 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), 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 rechtwinklige Dreieck ist. Daher heißt dieser Punkt *Brennpunkt* der Parabel. Nachdem man diesen Punkt entdeckt hat, sieht man leicht, daß jeder Punkt (x, x^2) der Parabel gleich weit vom Brennpunkt und von der Geraden $y = -\frac{1}{4}$ entfernt ist:

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

Aufgabe. Geben Sie die Gleichung der Parabel an, die den Einheitskreis in (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 senkrecht* 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.



Ausdehnung der Definition

Die zuletzt erhalten Eigenschaft, nämlich daß der Graph einer Potenzfunktion bei a zwi-

schen zwei quadratischen Parabeln liegt, dehnt sich wie folgt auf Summen und Linearkombinationen aus:
DEFINITION. Zu einem Polynom

$$P(x) = \sum_{k=0}^{n} a_k \cdot x^k$$

nennen wir das Polynom

$$P'(x) := \sum_{k=1}^{n} a_k \cdot k \cdot x^{k-1}$$

die *Ableitung* von P, und die Gerade $T(x) = P(a) + P'(a) \cdot (x - a)$ heißt *Tangente* (des Graphen) von P bei a.
RECHTFERTIGUNG. Zu jedem Polynom P und jedem Intervall $[-R, R]$ definieren wir

$$M(P, R) = M := \sum_{k=2}^{n} |a_k| \cdot k \cdot (k-1) R^{k-2}.$$

Dann ergibt die Addition unserer quadratischen Ungleichungen:

$$x, a \in [-R, R] \Rightarrow -M \cdot (x-a)^2 \leq P(x) - P(a) - P'(a) \cdot (x-a) \leq M \cdot (x-a)^2.$$

Diese Ungleichung besagt in der Tat, daß der Graph von P bei a zwischen zwei quadratischen Parablen mit gleicher Tangente liegt. Damit sind die Bezeichnungen beste lineare Approximation und Tangente (vergleichbar mit Kreistangente oder Parabeltangente) für Polynome berechtigt. – Wir werden später sowohl solche als auch schwächere Ungleichungen zur *Definition* von Ableitungen benutzen. Hier treten sie dagegen als *Eigenschaften* der Polynome auf, die ohne den Begriff des Grenzwertes bewiesen worden sind.

Zur Übung dehne ich die Definition noch auf negative Potenzen aus. Wir wären auch ohne das schon so weit, daß wir uns *Sätzen* über Ableitungen zuwenden könnten.
Wir betrachten $f(x) = x^{-n}$ und setzen voraus $0 < a, x$. Dann gilt für die Durchschnittssteigungen:

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

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

Dies ist wieder an den Intervallenden x bzw. a am kleinsten oder größten. Daher können wir wieder die Steigung als die Zahl definieren, die zwischen den linksseitigen und den rechtsseitigen Durchschnittsteigungen liegt.
DEFINITION von *Ableitung* und *Tangente* für $f(x) = x^{-n}$ an der Stelle $x = a$:

$$f'(a) = -n \cdot a^{-n-1}, \quad T(x) = f(a) + f'(a) \cdot (x - a).$$

Ich verifiziere nur für $n = 1$, daß $f(x) - T(x)$ wieder eine quadratische Fehlerabschätzung erlaubt, weil wir solche Aussagen mit vor uns liegenden Sätzen leichter bekommen werden als mit (immerhin durchführbaren) expliziten Rechnungen:

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

Also gilt für $f(x) = x^{-1}$:

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

Damit verläuft auch der Graph dieser Funktion an jeder Stelle zwischen einer Parabel und deren Tangente - die Benutzung der alten Vokabeln ist gerechtfertigt.
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

Ziel dieses Abschnitts: Zu jeder neuen Konstruktion von Funktionen sollte man zugehörige Ableitungsregeln haben. Wir beantworten folgende Fragen: Wie müssen die Regeln aussehen, die zu Summe, Produkt und Komposition von Funktionen gehören? Wie beweist man sie? Wir kennen den Begriff Umkehrfunktion, welche Differentiationsregel gehört dazu? Wir können mit Paaren von Funktionen Kurven in der Ebene beschreiben, wie differenziert man diese?

Linearität.

Falls f, g die Ableitungen f', g' haben und $\alpha, \beta \in \mathbb{R}$ sind, so gilt:

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

Dies folgt unmittelbar aus der Addition quadratischer Ungleichungen.

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 $(\ell_1 \cdot \ell_2' + \ell_2 \cdot \ell_1')(a)$. Wir erwarten also die Regel:

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

Kettenregel, Kompositionsregel

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

$$\ell(x) = A + m \cdot (x - a), \quad |f(Y) - f(A) - f'(A) \cdot (Y - A)| \leq \text{const} \cdot |Y - A|^2.$$

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

$$|f(\ell(x)) - f(\ell(a)) - f'(\ell(a)) \cdot m \cdot (x - a)| \leq \text{const} \cdot m^2 \cdot |x - a|^2.$$

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 $T(x)$ bei a schneller klein wird als proportional zum Abstand $|x - a|$. Beispiele für solche gröberen aber genügend guten Fehler sind: $\text{const} \cdot |x - a|^{1+p}$ 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 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:

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 noch 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 Voraussetzungen gilt:

$$|\phi(x,a)| \leq \text{const} \cdot |x-a|^p \quad \text{oder} \quad |\phi(x,a)| \leq \epsilon,$$

entweder in einem festen Intervall um a oder für jedes ϵ in einem geeigneten δ-Intervall um a. Bei den folgenden Beweisen sollte man nach Möglichkeit derartige Variationsmöglichkeiten vor Augen haben.

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 Weise kleine Fehler beschreiben.

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

liefert:

$$\begin{aligned} (f_1 \cdot f_2)(x) - (f_1 \cdot f_2)(a) &- (f_1' \cdot f_2 + f_1 \cdot f_2'(a)) \cdot (x-a) = \\ \Big[(f_1(a) + f_1'(a) \cdot (x-a)) \cdot \phi_2(x,a) &+ (f_2(a) + f_2'(a) \cdot (x-a)) \cdot \phi_1(x,a) \\ &+ (f_1'(a) \cdot f_2'(a) + \phi_1(x,a) \cdot \phi_2(x,a)) \cdot |x-a|\Big] \cdot |x-a| \end{aligned}$$

Für $a, x \in [-R, R]$ ist der Ausdruck in den eckigen Klammern $\leq \mathit{const}_1 \cdot \phi_2 + \mathit{const}_2 \cdot \phi_1 + 2R \cdot (f_1' \cdot f_2')(a) \cdot |x-a|$, wobei die Konstanten aus $f_i(a), f_i'(a)$ und der Intervallänge $2R$ bestimmt 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. Zum Beispiel:
Aus

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

folgt

$$\begin{aligned} &|x-a| \leq \min(r_1, r_2) \Rightarrow |(\mathit{const}_1 \cdot \phi_2 + \mathit{const}_2 \cdot \phi_1)(x,a) + 2R \cdot (f_1' \cdot f_2')(a) \cdot |x-a|| \\ &\leq (\mathit{const}_1 \cdot M_2 + \mathit{const}_2 \cdot M_1 + \mathit{const}_3 \cdot 2R) \cdot |x-a| = M_3 \cdot |x-a|. \end{aligned}$$

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. Anders als bei der Produktregel formuliere ich den Beweis daher nicht für alle Sorten "kleiner Fehler" auf einmal, ich beschränke mich auf die bei den Polynomen beobachteten quadratischen Fehler und verschiebe die Diskussion der ϵ-δ-Fehler.

Voraussetzung:

$|x-a| \le r_1 \Rightarrow |f(x) - f(a) - f'(a) \cdot (x-a)| \le M_1 (x-a)^2, \quad f(a) = A$ 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:

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

Ferner (Dreiecksungleichung in der Voraussetzung):

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

und (Multiplikation 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 die dritte Ungleichung in die erste einsetzen und die Konstanten sortieren:

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

Dabei ist erstens

$\text{Const} = M_2 (|f'(a)| + M_1 \cdot r_1)^2 + |F'(f(a))| \cdot M_1,$

und zweitens 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

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

voraussetzen, folgt aus der Voraussetzung über f wirklich $|f(x) - A| \le R_2$, so daß die vorgeführten Abschätzungen allein aus den Voraussetzungen über f und F folgen. - Der Beweis mit ϵ-δ-Fehlern wird nach demselben Muster geführt werden.

Anwendungen

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

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:

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

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

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

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'.

Umkehrfunktionen

Für monoton wachsende, surjektive Funktionen $f\colon [a,b] \to [A,B]$, meist veranschaulicht durch ein Bild von Graph $f := \{(x, f(x)); x \in [a,b]\}$ entsteht der Graph der Umkehrfunktion $f^{-1}\colon [A,B] \to [a,b]$ aus dem Graphen von f durch Spiegelung an der Winkelhalbierenden, $(x,y) \to (y,x)$.

Daher erwarten wir als Ableitung

$$(f^{-1})'(f(x)) = \frac{1}{f'(x)},$$

in Übereinstimmung mit der Kettenregel:

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

Beispiel: Für $f(x) = x^2, \quad f'(x) = 2x, \quad f^{-1}(y) = \sqrt{y}$
erwarten wir daher:

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

In der Tat:

$$\sqrt{x} - \sqrt{a} - \frac{1}{2\sqrt{a}} \cdot (x-a) = \left(\frac{1}{\sqrt{x}+\sqrt{a}} - \frac{2}{\sqrt{a}}\right) \cdot (x-a)$$
$$= \frac{\sqrt{a}-\sqrt{x}}{2\sqrt{a}(\sqrt{x}+\sqrt{a})} \cdot (x-a) = \frac{-(x-a)^2}{2\sqrt{a} \cdot (\sqrt{x}+\sqrt{a})^2}, \text{also}$$
$$\frac{a}{2} \le x \le 2a \ \Rightarrow \ \frac{-1}{4\sqrt{a}^3} \cdot (x-a)^2 \le \ \sqrt{x} - \left(\sqrt{a} + \frac{1}{2\sqrt{a}} \cdot (x-a)\right) \ \le 0.$$

Das Beispiel macht deutlich, daß, anders als bei allem vorhergehenden, die Diskussion der Umkehrfunktionen abhängt von dem, was man über Zahlen weiß. Z.B. liegt 2 nicht im Bild von $f(x) = x^2$, wenn man nur die rationalen Zahlen, also nicht $x = \sqrt{2}$, kennt. Diese Fragen werden im Abschnitt Vollständigkeit der reellen Zahlen geklärt. Dagegen hängt die Form der Ableitungsregeln von solchen Kenntnissen nicht ab: Die folgenden Ungleichungen sind eben nur gemeint für Zahlen, von denen man weiß (oder später erfährt), daß sie im Definitionsbereich der Umkehrfunktion f^{-1} liegen.

Beweis der quadratischen Tangentenapproximation für Umkehrfunktionen.

Voraussetzung: $0 < f'(a), f$ monoton und
$|x-a| \le r \Rightarrow |f(x) - f(a) - f'(a) \cdot (x-a)| \le M \cdot (x-a)^2$
Setze $X := f(x), A := f(a)$. Dies sind Punkte im Definitionsbereich von f^{-1}: $x = f^{-1}(X)$, $a = f^{-1}(A)$. Wir setzen diese Bezeichnungen in die Voraussetzung ein und dividieren durch $f'(a)$:

$$(*) \quad |\frac{1}{f'(a)} \cdot (X-A) - f^{-1}(X) + f^{-1}(A)| \le \frac{M}{f'(a)} \cdot (x-a)^2$$

Wie im Beweis der Kettenregel (bitte vergleichen) gilt:
$|x-a| \le r \Rightarrow (f'(a) - M \cdot r) \cdot |x-a| \le |f(x) - f(a)| \le (f'(a) + M \cdot r) \cdot |x-a|$.
Damit die linke Ungleichung etwas nützt, verkleinern wir r zu $r_- := \min(r, f'(a)/2M)$ und vergröbern die letzte Ungleichung:

$$\frac{1}{2} f'(a) \cdot |x-a| \le |X-A| \le \frac{3}{2} f'(a) \cdot |x-a|.$$

Bisher gilt die Ungleichung $(*)$ unter der Voraussetzung $|x-a| \le r$. Mit der unteren Abschätzung für $|X-A|$ haben wir jetzt: $|X-A| \le R := \frac{1}{2} f'(a) \cdot r_-$ und damit $(*)$. Schließlich vergröbern wir $(*)$ durch $|x-a| \le \frac{2}{f'(a)} \cdot |X-A|$ zu

$$|f^{-1}(X) - f^{-1}(A) - \frac{1}{f'(a)} \cdot (X-A)| \le \frac{3M}{f'(a)^3} \cdot |X-A|^2.$$

Dies ist die gewünschte quadratische Tangentenapproximation. Der Beweis überträgt sich auf ϵ-δ-Fehler: Es ist leicht, $M \cdot (x-a)^2$ durch $\epsilon \cdot |x-a|$ zu ersetzten; mehr argumentative Schwierigkeiten macht, durch Voraussetzungen an $|X-A|$ zu garantieren, daß die Voraussetzungen für die Gültigkeit der benutzten Abschätzungen für f erfüllt sind.

Kurven

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:
$f(t) = p_1 + m_1 \cdot t$, $g(t) = p_2 + m_2 \cdot t$. 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 $|\phi_i(t, t_0)| \leq M_i \cdot |t - t_0|$ in Intervallen $[t_0 - r, t_0 + r]$ oder ϵ-δ-Fehler $|\phi_i(t, t_0)| \leq \epsilon$ in Intervallen $[t_0 - \delta, t_0 + \delta]$. (Hier sind M_i und r Konstanten, die von t_0 abhängen dürfen oder in günstigeren Fällen gleichmäßig gelten, 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 (gleichmäßig) ein $\delta = \delta(\epsilon)$ für alle t_0.

Unabhängig von einer derartigen Präzisierung der "kleinen Fehler" gilt jedenfalls wegen des Satzes von Pythagoras:

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

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 dieser Art ist, z.B.

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

hat zur Folge

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

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.*

Beispiel. Für die anfangs angegebene Kreiskurve

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

finden wir

$$k'(t) = \frac{2}{(1 + t^2)^2} \cdot (1 - t^2, \; -2t), \qquad |k'(t)| = \frac{4}{(1 + t^2)^2}.$$

Offenbar ist der Betrag der Geschwindigkeit dieser Kreisbewegung nicht konstant.

2. Der Monotoniesatz

Ziel dieses Abschnitts: 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. Da wir die Vollständigkeit der reellen Zahlen noch nicht besprochen haben, führe ich noch einmal vor, wie Gleichmäßigkeitsvoraussetzungen erlauben, Beweise ohne die Vollständigkeit der reellen Zahlen zu führen.

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.

Daher halten wir unsere Voraussetzungen fest:
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.

Für solche Funktionen haben wir die Differentiationsregeln bewiesen. Die benutzte Sprache erfordert zwei kleine Rechtfertigungen. Erstens, eine Funktion f kann an einer Stelle a nicht zwei verschiedene Tangenten haben: Hätten wir zwei verschiedene Steigungen $m_1 \ne m_2$ so folgte

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

also auch

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

Aber diese Ungleichung ist falsch, wenn wir x so wählen, daß gilt

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

Zweitens: Ist eine Funktion f schwach wachsend (d.h. $a \le x \Rightarrow f(a) \le f(x)$), so verlangt die umgangssprachliche Bedeutung des Wortes Steigung, daß wir "Steigung" ≥ 0 beweisen können; in der Tat, aus

$$a \le x \le a+r \Rightarrow -m\cdot(x-a) \le f(x) - f(a) + m\cdot(x-a) \le K\cdot(x-a)^2,$$

folgt

$$a < x \le a+r \quad \Rightarrow \quad -m \le K\cdot|x-a|.$$

Wäre nun $m < 0$, so würde diese Ungleichung falsch, wenn wir x so wählen, daß gilt

$$a < x \le a + \min(r, |m|/2K).$$

Die hier benutzte Argumentation geht schon auf Archimedes zurück. Ich halte sie fest als:
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. Wir wollen nämlich meistens nicht aus Voraussetzungen über f Eigenschaften von f' herleiten sondern umgekehrt aus Voraussetzungen über f' Eigenschaften von f folgern. Prototyp ist der

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

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

Bemerkung. 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.) Funktionen mit der größeren Ableitung wachsen schneller:
$x \leq y$ und $f' \leq g'$ 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:
$$|f'| \leq L \text{ impliziert } |f(y) - f(x)| \leq L \cdot |y - x|,$$
denn mit der linearen Funktion $g(x) := L \cdot x$ gilt $-g' \leq f' \leq g'$. Mit der ersten Folgerung haben wir für $x \leq y: \ \ -L \cdot (y - x) \leq f(y) - f(x) \leq +L \cdot (y - x)$.

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

3.) Aus $f'' \geq 0$ folgt: f liegt oberhalb jeder Tangente. Sei $T(X) := f(a) + f'(a) \cdot (x - a)$ die Tangente bei a, dann gilt für die Hilfsfunktion $h := f - T$ zunächst $h'' \geq 0$. Ferner ist $h'(a) = 0$, also folgt: $x \leq a \Rightarrow h'(x) \leq 0, a \leq x \Rightarrow h'(x) \geq 0$. Nach dem Monotoniesatz ist h rechts von a wachsend, links von a fallend, wegen $h(a) = 0$ folgt daraus $h \geq 0$ oder $f \geq T$.
Die Bernoulli'sche Ungleichung: $-1 \leq x, n \in \mathbb{N} \ \Rightarrow \ (1 + x)^n \geq (1 + n \cdot x)$ ist ein uns schon bekannter Spezialfall.

Bemerkung: Schon bei recht einfachen Polynombeispielen ist die Anwendung von 3.) wesentlich kürzer als ein direktes Nachrechnen.

4.) Aus $|f''| \leq K$ folgt: f weicht von seiner Tangente bei a höchstens um $\frac{1}{2}K \cdot (x - a)^2$ ab. Denn, für die Hilfsfunktion $h(x) := f(x) - T(x)$ gilt $h(a) = 0 = h'(a), -K \leq h'' \leq K$.
Wir wenden Folgerung 1 zweimal an, zuerst:

$$\begin{array}{l} x \geq a \rightarrow -K \cdot (x - a) \leq h'(x) - h'(a) \leq +K \cdot (x - a), \\ x \leq a \rightarrow -K \cdot (a - x) \leq h'(x) - 0 \leq +K \cdot (a - x) \end{array},$$

und dann noch einmal:

$$-\tfrac{1}{2}K \cdot (x - a)^2 \leq h(x) - 0 \leq \tfrac{1}{2}K \cdot (x - a)^2.$$

Bemerkung: Diese Folgerung erklärt also die Bedeutung der Konstanten in der quadratischen Abweichung von der Tangente.

5.) Multiplikative Version 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:

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

Anwendung dieser multiplikativen Version: Betrachte für $0 \le x$ und $m < n \in \mathbf{N}$ die Polynome $f(x) = (1+x/m)^m$ und $g(x) = (1+x/n)^n$, die wir bei der Exponentialfunktion wieder treffen werden. Ihre Wachstumsraten sind

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

Zusammen mit $f(0) = 1 = g(0)$ 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 z.B. beschränkt? Trotz des einfachen Aussehens dieser Polynome kann man eine Schranke nicht einfach ausrechnen. Mit dem multiplikativen Monotoniesatz geht es schnell, der Ansatz verdient jedoch einen Kommentar: Interpretiert man $1+x$ als Gutschrift von x Zinsen 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 kleiner ist als das Endkapital, so hat man die Anfangssumme durch $(1-x)$ zu dividieren, um die Endsumme zu bekommen. Entsprechend $1/(1-x/k)^k$ für k Gutschriften von x/k Rückwärtszinsen.
Betrachte daher

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

es gilt

$$h(0) = 1, \ \text{ und } \ \frac{h'}{h}(x) = \frac{1}{1-x/k} \ge 1 \ge \frac{1}{1+x/n} = \frac{g'}{g}(x).$$

Somit liefert der multiplikative Monotoniesatz

$$0 \leq x < k \Rightarrow (1 + x/n)^n \leq 1/(1 - x/k)^k.$$

Das Interpretieren der Formel hat also zu einem brauchbaren Ansatz geführt.

Nach diesen Folgerungen aus dem Monotoniesatz komme ich 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 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 Voraussetzungen.

Erster Beweis des Monotoniesatzes

VORAUSSETZUNG: f sei auf $[\alpha, \omega]$ differenzierbar und $f' \geq 0$.

ZUSATZVORAUSSETZUNG: Es gibt positive Konstanten r, K so daß für die Tangentenapproximation "gleichmäßig" gilt:

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

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 besprochenen "kleinen Fehler", sofern die daran beteiligten Konstanten gleichmäßig gewählt werden können. Ein Beweis ohne diese Gleichmäßigkeit folgt nach der Diskussion der Vollständigkeit, vgl. S.39.

Beweis: Teile das Intervall $[x, y]$ in n gleiche Teile, wobei n mindestens so groß ist, daß $(y-x)/n \le r$ gilt. Bezeichne $t_k = x + \frac{k}{n} \cdot (y-x), k = 0, \ldots, n$. Dann gilt wegen $f' \ge 0$ und der Zusatzvoraussetzung:

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

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

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

Schließlich wird dies mit dem Archimedes Argument zu der Behauptung $0 \le f(y) - f(x)$ verbessert.

Dieselbe Argumentation liefert uns weitere Eigenschaften von differenzierbaren Funktionen, die gleichmäßig mit quadratischen Fehlern von ihren Tangenten approximiert werden.

Gleichmässigkeitsvoraussetzung: f sei auf $[\alpha, \omega]$ definiert und es gibt positive Konstanten K, r so daß gilt:

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

Folgerung 1: Die Ableitung f' ist dehnungsbeschränkt:

$$a, b \in [\alpha, \omega] \Rightarrow |f'(a) - f'(b)| \le 2K \cdot |b-a|.$$

Beweis: Zunächst gilt für jede Funktion h wegen der Dreiecksungleichung (Wiederholung: $|C-A| \le |C-B| + |B-A|$)

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

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

In der nächsten Folgerung betrachten wir sogenannte Riemann Summen. Zur Definition sei $a < b \in [\alpha, \omega]$ und sei $a = t_0 < t_1 < \ldots < t_n = b$ eine Einteilung des Intervalls $[a, b]$. Das längste Teilintervall hat die Länge $\delta := \max |t_k - t_{k-1}|$, außerdem seien Zahlen $\tau_k \in [t_{k-1}, t_k],\ k = 1 \ldots n$ gewählt.

Definition: Jede der Summen $\sum_{k=1}^{n} f(\tau_k) \cdot (t_k - t_{k-1})$ heißt **Riemann Summe** von f.

Betrachte als Beispiel für festes k die Summe $S_N = \sum_{n=1}^{N} n^k$, also eine Riemann Summe der Funktion $x \to x^k$. Wir möchten wissen, wie schnell S_N als Funktion von N ungefähr wächst. Wir vergleichen mit 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 und:

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

Daher folgt aus dem Monotoniesatz:

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

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

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

Die linke und die rechte Schranke stimmen in der höchsten Potenz von N überein.

Folgerung 2: Riemann-Summen von Ableitungen f' können bis auf kleine Fehler berechnet werden. Genauer, es gilt die

Behauptung: $|f(b) - f(a) - \sum_{k=1}^{n} f'(\tau_k) \cdot (t_k - t_{k-1})| \le K \cdot |b-a| \cdot \delta$.

Zusatz: Gilt für $t \in [t_{k-1}, t_k]$: $f'(t) \le f'(\tau_k)$, so liefert derselbe Beweis

$$f(b) - f(a) \le \sum_{k=1}^{n} f'(\tau_k) \cdot (t_k - t_{k-1});$$

entsprechend gilt die umgekehrte Ungleichung $\ge$, falls $f'(t) \ge f'(\tau_k)$.

Beweis. In jedem Teilintervall gilt

$$\begin{aligned} &|f(t_k) - f(\tau_k) + f(\tau_k) - f(t_{k-1}) - f'(\tau_k) \cdot (t_k - \tau_k + \tau_k - t_{k-1})| \le \\ &\le K \cdot (|t_k - \tau_k|^2 + |\tau_k - t_{k-1}|^2) \le K \cdot \delta \cdot |t_k - t_{k-1}|. \end{aligned}$$

Summiert man diese Ungleichungen über $k = 1, \ldots, n$ und beachtet $\sum |t_k - t_{k-1}| = b - a$, 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 $f'(\tau_k) \cdot (t_k - t_{k-1})$ durch ein Rechteck der Höhe $f'(\tau_k)$ über dem Teilintervall $[t_{k-1}, t_k]$, 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 $f(b) - f(a)$. 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 erwarten dürfen, daß aus dem Archimedes-Axiom folgt, daß $f(b) - f(a)$ der Flächeninhalt unter dem Graphen von f' ist. Schließlich ist die 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.

3. Die Vollständigkeit der reellen Zahlen

Ziele dieses Abschnitts: Die Konstruktion von nicht rationalen Funktionen wie *exp, sin, cos* erfordert die Vollständigkeit der reellen Zahlen und den Begriff der Konvergenz. Vorausgeschickt wird das Archimedes Axiom. Die Vollständigkeit wird mit Intervallschachtelungen, Cauchyfolgen und mit kleinsten oberen Schranken diskutiert. Die Verbindung von Geometrie und Analysis stellt der Satz von Bolzano-Weierstraß her, nämlich daß jede monoton wachsende und beschränkte Folge in $\mathbb{R}$ konvergiert.
Die Körperaxiome werden in der Literatur so einheitlich behandelt, daß ich darauf nicht weiter eingehe. Dasselbe gilt für die meisten Anordungunsaxiome, hervorheben möchte ich nur das:
Archimedes Axiom:
Zu jeder reellen Zahl $r > 0$ gibt es ein $n \in \mathbb{N}$ mit $r > 1/n$.
Die häufigste Anwendung ist: Kann man für eine Zahl r zeigen: $r \leq 1/n$ für jedes $n \in \mathbb{N}$, so gilt $r \leq 0$. Umgangssprachlich könnte man sagen: Es gibt keine unendlich kleinen reellen Zahlen. Da wir die reellen Zahlen nicht konstruieren, erläutern wir das Archimedes Axiom mit folgendem Beispiel.
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 $P \overset{k}{>} 1/n$.

Die beweistechnische Bedeutung dieses Axioms ist sehr groß und verdient, als strategische Merkregel festgehalten zu werden:

Archimedes Strategie: Es genügt, reelle Ungleichungen $a \leq b$ bis auf Archimedes Fehler zu beweisen: $a \leq b + 1/n, n \in \mathbb{N}$ hat $a \leq b$ zur Folge.

Wir kommen zur Vollständigkeit. Als Zahlen standen im Altertum nur die rationalen Zahlen zur Verfügung. Es wurde zwar geometrische Streckenverhältnisse, wie Quadratdiagonale zu Seitenlänge, betrachtet, solche "Verhältnisse von Größen" konnten aber nicht durch Zahlen ausgedrückt werden. Die Ungleichungen, die ich bei den Polynomen diskutiert habe, hätten auch schon vor 2000 Jahren Sinn gemacht, sie hängen nicht davon ab, daß man mehr Zahlen als nur die rationalen Zahlen kennt. Im Zusammenhang mit den Umkehrfunktionen wird der Mangel an ausreichend vielen Zahlen allerdings ein Problem: Betrachte die Umkehrfunktion P^{-1} des monotonen Polynoms $P(X) = X^3$. Was ist $P^{-1}(2)$, wenn man die Irrationalzahl $\sqrt[3]{2}$ nicht kennt? Um den "anschaulichen" Satz zu beweisen: "Ein monotones Polynom P bildet jedes Intervall $[a, b]$ surjektiv auf das Bildintervall $[P(a), P(b)]$ ab, d.h. jede Zahl $R \in [P(a), P(b)]$ ist ein Wert $R = P(r)$ für eine Zahl $r \in [a, b]$ " braucht man jedenfalls genügend viele Zahlen.
Ein anderes Beispiel. Das Betrachten relativer Fehler

$$\frac{f(x) - f(a)}{f(a)} \sim \frac{f'}{f}(a) \cdot (x - a)$$

hatte uns auf die Wachstumsrate f'/f positiver Funktionen geführt. Um diese Wachstumsrate besser zu verstehen, suchen wir Funktionen mit konstanter Wachstumsrate, etwa $f'/f = 1$ oder $f' = f$. Wir ziehen Folgerungen aus dieser Ableitungsinformation:
Behauptung: Solche Funktionen haben folgende Funktionalgleichung:

$$\frac{f(x)}{f(0)} = \frac{f(x + a)}{f(a)},$$

d.h. sie wachsen in gleichen Intervallen um den gleichen Faktor (statt des *additiven* Wachstums bei konstanter Steigung).
Beweis. Die Quotientenfunktion $f(x + a)/f(x)$ hat die Ableitung 0, hat also den konstanten Wert $f(a)/f(0)$.

Offensichtlich gibt es kein Polynom mit konstanter Wachstumsrate und auch unter den rationalen Funktionen findet man keine solchen. Gibt es sie überhaupt? Wenigstens geht es beinahe, wir kennen ja schon wachsende und fallende Funktionenfolgen, deren Wachstumsraten sich mit größerem Index immer weniger von 1 unterscheiden:

$$\begin{aligned}
0 \leq x,\ f_n(x) = (1 + x/n)^n,\ \frac{f_n'}{f_n}(x) = \frac{1}{1 + x/n},\ f_n(x) \leq f_{n+1}(x);\\
0 \leq x < k,\ g_k(x) = \frac{1}{(1 - x/k)^k},\ \frac{g_k'}{g_k}(x) = \frac{1}{1 - x/k},\ g_{k+1}(x) \leq g_k(x);\\
f_n(x)\ \leq g_k(x).
\end{aligned}$$

Tatsächlich wird auch der Abstand beliebig klein:

$$0 \leq g_n(x) - f_n(x) = g_n(x) \cdot (1 - f_n(x)/g_n(x))^n) = g_n(x) \cdot (1 - (1 - \frac{x^2}{n^2})^n),$$

was nur noch übersichtlicher gemacht werden muß. Z.B. im Intervall [0, 4] vergröbern wir den ersten Faktor für $n \geq 8$ zu

$$g_n(x) \leq g_8(x) \leq g_8(4) = 256.$$

Für den zweiten Faktor $h(x) := (1 - (1 - x^2/n^2)^n)$ gilt wegen $h(0) = 0, 0 \leq h'(x) = (1 - x^2/n^2)^{n-1} \cdot 2x/n \leq 2x/n$ also mit dem Monotoniesatz

$$0 \leq x \leq n \Rightarrow 0 \leq h(x) \leq \frac{x^2}{n}.$$

Dasselbe kann mit der geometrischen Reihe erreicht werden (setze $q = (1 - x^2/n^2)$):

$$h(x) = (1 - q^n) = (1 - q) \cdot (1 + q + \ldots + q^{n-1}) \leq n \cdot (1 - q) = \frac{x^2}{n}.$$

Die wachsende Folge $f_n(x)$ und die fallende Folge $g_n(x)$ kommen sich also so nahe, wie man mit Blick auf das Archimedes Axiom nur wünschen kann. Genauer:

$$0 \leq x \leq 4 \Rightarrow 0 \leq g_n(x) - f_n(x) \leq 256 \cdot \frac{x^2}{n}.$$

Wenn wir in dieser Situation sagen könnten: "Es gibt eine Grenzfunktion", dann könnten wir auch leicht zeigen, daß sie konstante Wachstumsrate hat. Das Vollständigkeitsaxiom wird genau das leisten. Wir brauchen dazu einige Begriffe.
Das Archimedes Axiom läßt uns verabreden: Eine Folge $\{r_k\}_{k\in\mathbb{N}}$ heißt Nullfolge, wenn sie schließlich kleiner als jede Zahl $1/n$ wird, genauer:
Definition der Nullfolge: Eine Folge $\{r_k\}_{k\in\mathbb{N}}$ heißt Nullfolge, wenn es zu jeder Zahl $1/n$ einen Folgenindex $k(n)$ gibt, so daß gilt: $k \geq k(n) \Rightarrow |r_k| \leq 1/n$.
Definition der konvergenten Folge: Eine Folge$\{a_k\}_{k\in\mathbb{N}}$ heißt konvergent gegen a, genau dann wenn $\{a_k - a\}_{k\in\mathbb{N}}$ eine Nullfolge ist.
Beispiele. Viele Übungsaufgaben werden beantwortet durch:
Falls f dehnungsbeschränkt ist, $|f(x) - f(y)| \leq L \cdot |x - y|$, und a_k konvergent gegen a, so ist $f(a_k)$ konvergent gegen $f(a)$, denn $a_k - a$ ist nach Voraussetzung Nullfolge, daher auch $f(a_k) - f(a)$ wegen

$$|f(a_k) - f(a)| \leq L \cdot |a_k - a|.$$

(Die Dehnungsschranke verschafft man sich durch eine Schranke für f' mit dem Monotoniesatz; wir haben so ein überraschend leistungsfähiges Konvergenzkriterium.)
Der Nachteil der Konvergenzdefinition ist, daß man zu ihrer Formulierung den Grenzwert

kennen muß. Diese Definition genügt also noch nicht, um in unserem Beispiel auch nur ausdrücken zu können: "Die Funktionen $f_n(x), g_n(x)$ konvergieren gegen die gewünschte Grenzfunktion". Mit den beiden ersten Definitionen kann man also nur üben, man kann noch nicht etwas Neues erfassen.

Definition der Intervallschachtelung: Eine monoton wachsende Folge $\{a_k\}$ und eine monoton fallende Folge $\{b_k\}$ bilden eine Intervallschachtelung $\{[a_k, b_k]\}$, falls gilt:

1.) $a_k \leq b_k$ für alle $k \in \mathbb{N}$ (und damit auch $a_m \leq b_n$ für alle $m, n \in \mathbb{N}$)

2.) $\{b_k - a_k\}$ ist eine Nullfolge.

Beispiel. Für jedes $x \in [0, 4]$ haben wir gezeigt, daß $\{[f_n(x), g_n(x)]\}$ eine Intervallschachtelung ist.

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 $k \in \mathbb{N}$.

Folgerung Offenbar sind $a_k - r$ und $b_k - r$ Nullfolgen, die Folgen a_k und b_k konvergieren also gegen r.

Es gibt noch andere Formulierungen der Vollständigkeit. Ich komme darauf zurück. Vorher diskutiere ich Anwendungen. Da Nullfolgen offenbar eine zentrale Rolle spielen, sollte man über einen Vorrat an "bekannten" Nullfolgen verfügen, denn eine unbekannte Folge $\{u_k\}$ ist sicher dann eine Nullfolge, wenn man sie mit einer schon bekannten Nullfolge $\{r_k\}$ für alle $k \in \mathbb{N}$ (oder auch nur: für alle genügend großen $k \in \mathbb{N}, k \geq K$) majorisieren kann: $|u_k| \leq |r_k|$.

Wegen $n \cdot q^n \leq q + \ldots + q^n \leq q/(1-q)$ für $0 < q < 1$ haben wir die besonders beliebten Nullfolgen:

$$q^n \leq \frac{1}{n} \cdot \frac{q}{1-q}, \quad n \cdot q^n \leq \frac{1}{n} \cdot \frac{q}{(1-\sqrt{q})^2}, \ldots$$

(quadriere die erste Ungleichung, ersetze q durch $\sqrt{q}$ und multipliziere mit n).

Daß wir in der Tat durch das Vollständigkeitsaxiom genügend viele Zahlen für die geometrische Anschauung bekommen, zeigt der zentrale

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 $k_1 = 1$.

Nun wird das folgende Verfahren (für jedes m) wiederholt:

Angenommen für alle $j,\ j \leq m$ seien die "Erfolgsindices" $e_j \geq j$ schon definiert und für

$k \leq e_m$ auch die oberen Schranken b_k. (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: $e_{m+1} := k,\ b_k := c_k$.
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 $e_m < k \leq e_{m+1}$: $b_k = b_{e_m}$.
Damit ist mindestens ein weiteres b_k definiert und $|a_{k_{m+1}} - b_{k_{m+1}}| \leq \frac{1}{2}|a_{k_m} - b_{k_m}|$, d.h. wir erhalten die gewünschte Intervallschachtelung und mit dem Vollständigkeitsaxiom einen Grenzwert $r \in [a_k, b_k]$ der Folge $\{a_k\}$.

Ähnlich beweisen wir jetzt den oben formulierten "anschaulichen"
Satz über die Surjektivität monotoner Funktionen. $f: [a, b] \to \mathbf{R}$ sei monoton und dehnungsbeschränkt. Dann ist f surjektiv auf das Bild $[A_1, B_1] := [f(a), f(b)]$.

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 $D \geq C$ 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. 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.
Bemerkungen. 1.) Setzt man zusätzlich voraus, daß f streng monoton ist, dann ist die Umkehrabbildung auf dem Bildintervall definiert. Jetzt sollte man die Ableitungsregel für die Umkehrfunktion noch einmal ansehen.
2.) Der Satz kann verallgemeinert werden: dehnungsbeschränkte Funktionen sind surjektiv auf ihr Bild. Dazu müssen noch $\min f$ und $\max f$ diskutiert werden. Diese Erweiterungen benötige ich nicht zum Erläutern der Vollständigkeit, sie werden bei den stetigen Funktionen besprochen.
Die letzte der früher angeschnittenen und noch offenen Fragen ist die Existenz einer Funktion mit Wachstumsrate 1. Nach Zitieren des Vollständigkeitsaxioms für die Intervallschachtelung $[f_n(x), g_n(x)]$ haben wir jedenfalls eine Grenzfunktion $E(x)$, zunächst für

$x \in [0,4]$.
Wir stellen zunächst eine grobe Dehnungsschranke für diese Grenzfunktion her:

$$x \in [0,4] \;\Rightarrow\; 0 \le f_n'(x) = \frac{1}{1+\frac{x}{n}} \cdot f_n(x) \le 1 \cdot g_8(x) \le 256.$$

Mit dem Monotoniesatz folgt:

$$x, y \in [0,4] \;\Rightarrow\; |f_n(x) - f_n(y)| \le 256 \cdot |x-y|$$

Ferner sind $u_n = E(x) - f_n(x)$ und $v_n = E(y) - f_n(y)$ Nullfolgen (nämlich $0 \le u_n, v_n \le 256 \cdot 16/n$). Aus dem Archimedes Axiom folgt

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

Wichtig ist mir an diesem Ergebnis, daß zu den uns schon bekannten Argumenten nur das Zitieren des Vollständigkeitsaxioms hinzu kommt, um Grenzfunktionen und deren Eigenschaften zu erhalten. Es hat keine nachteiligen Folgen, daß wir mit einer sehr groben Dehnungsschranke zufrieden sind; diese genügt bereits, um die *richtige* Ableitung zu finden, und die Bestimmung der Ableitung der Grenzfunktion verläuft dann ähnlich wie die Herleitung der groben Dehnungsschranke. (Wie vorher ist $x, y \in [0,4]$ und $n \ge 8$):
Aus

$$0 \le f_n'' = ((1+x/n)^{n-1})' = \frac{n-1}{n} \cdot (\frac{1}{1+x/n})^2 \cdot f_n(x) \le g_8(x) \le 256$$

folgt mit dem Monotoniesatz (Folgerung 4):

$$|f_n(y) - f_n(x) - f_n'(x) \cdot (y-x)| \le 128 \cdot (y-x)^2.$$

Wegen

$$f_n(x) - f_n'(x) = \frac{x/n}{1+x/n} \cdot f_n'(x) \le \frac{1}{n} \cdot 4 \cdot g_8(4)$$

ist $f_n(x) - f_n'(x)$ Nullfolge, also auch $f_n'(x) - E(x)$. Daher haben wir mit der Dreiecksungleichung:

$$|E(y) - E(x) - E(x) \cdot (y-x)| \le 128 \cdot (y-x)^2 + \text{ Nullfolgen},$$

schließlich mit dem Archimedes Argument:

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

Diese Ungleichung besagt, daß E differenzierbar ist, $E'(x) = E(x)$ und daß die Tangenten mit quadratischen Fehlern approximieren.
Schließlich wollen wir uns noch von der Einschränkung auf das Intervall $[0,4]$ befreien. Der

Multiplikative Monotoniesatz (Folgerung 5) besagt, daß der Quotient von zwei Funktionen mit $f'(x) = f(x),\ g'(x) = g(x)$ konstant ist. Für jede der Funktionen $f_k(x) := E(x/k)^k$, $(k \in \mathbb{N})$ gilt nun mit der Kettenregel:

$$f_k(0) = E(0) = 1, \quad f_k'(x) = k \cdot E(\frac{x}{k})^{k-1} \cdot E'(\frac{x}{k}) \cdot \frac{1}{k} = f_k(x),$$

also $f_k(x) = E(x)$ für $0 \le x \le 4$, während der Definitionsbereich von $f_k(x)$ viel größer ist, nämlich: $[0, 4 \cdot k]$. (Natürlich stimmen auch die $f_k(x)$ auf den Durchschnitten ihrer Definitionsbereiche überein). Damit haben wir für alle $x \in [0, \infty)$ eine Funktion $E(x)$ mit $E(0) = 1, E'(x) = E(x)$ definiert. Derselbe Trick noch einmal: $E(-x) := 1/E(x)$ $(k = -1$ in der Rechnung) setzt die Definition auf $(-\infty, 0]$ fort.
Bezeichnung:
Die so erhaltene Funktion heißt "Exponentialfunktion", $x \to \exp(x), \exp' = \exp$.

Damit hat das Vollständigkeitsaxiom seine Leistungsfähigkeit demonstriert. Allerdings ist noch ein Zusatz zur Exponentialfunktion nötig, da eine andere als die eben verwendete Approximation wesentlich bekannter ist. Wir bemerken zunächst, daß alle Ableitungen der Exponentialfunktion wegen $\exp'' = \exp' = \exp$ usw. bei 0 den Wert 1 haben. Außerdem ist es leicht, ein Polynom T_n vom Grade n hinzuschreiben, dessen Ableitungen vom Grade $\le n$ alle bei 0 den gewünschten Wert 1 haben:

$$T_n(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} = \sum_{k=0}^{n} \frac{x^k}{k!}.$$

Bezeichnung: $T_n(x)$ heißt n-tes Taylorpolynom von exp bei 0.
Die Ableitungen der eben benutzten Polynomapproximation n-ten Grades:
$f_n(x) = (1 + x/n)^n$ sind bei 0 aber ≤ 1, daher haben wir (für $0 \le x$) erstens:

$$(1 + \frac{x}{n})^n \le T_n(x).$$

Andererseits ist $T_n'(x) = T_{n-1}(x) \le T_n(x)$, also mit dem multiplikativen Monotoniesatz zweitens:

$$T_n(x) \le \exp(x).$$

Damit konvergiert auch die Folge $\{T_n(x)\}$ der Taylorpolynome gegen $\exp(x)$.

Um die Qualität dieser Taylorapproximation zu demonstrieren, zeige ich:
$e := \exp(1)$ 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:

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

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 $1/Q!$:

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

Da ein Bruch $\neq 0$ 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 mononton 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:

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

Beispiel. Die 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 folgt durch Differenzieren von

$f^2 + g^2 = 1,\ f'^2 + g'^2 = 1$:
$2 \cdot f \cdot f' + 2 \cdot g \cdot g' = 0,\ 2f' \cdot f'' + 2g' \cdot g'' = 0.$

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: $(f'', g'') + (f, g) = 0$.
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:

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

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 $N \in \mathbb{N}$ fest und betrachte nur die Polynome mit $n \geq N$.
Dann gilt: Für $n \geq N$ und $|t| \leq 2N$ sind $\{|t|^{2n}/(2n)!\}_{n\in\mathbb{N}}$ und $\{|t|^{2n+1}/(2n+1)!\}_{n\in\mathbb{N}}$ monotone Nullfolgen.
Für $2m \geq N$ und $|t| \leq 2N$ haben wir daher die Leibniz'schen Intervallschachtelungen für die gesuchten Grenzfunktion f und g

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

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 (und damit dieselben Schranken für die Grenzfunktionen):

$$\begin{gathered} 2m \geq 2, \quad 0 \leq t \leq 4 \quad \Rightarrow \\ -7 \leq P_3(t) \leq P_{2m+1}(t) \leq \ldots f(t) \ldots \leq P_{2m}(t) \leq P_2(t) \leq 4 \\ -7 \leq Q_3(t) \leq Q_{2m+1}(t) \leq \ldots g(t) \ldots \leq Q_{2m}(t) \leq Q_2(t) \leq 4 \end{gathered}$$

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

$$\begin{aligned} P_2(t) =& 1 - \frac{t^2}{2}(1 - \frac{t^2}{12}) \leq 4 \\ Q_2(t) =& t \cdot (1 - \frac{t^2}{6} \cdot (1 - \frac{t^2}{20})) \leq 4 \\ -7 \leq& 1 - \frac{t^2}{2} + \frac{t^4}{4!} - \frac{t^6}{6!} = P_3(t) \\ -7 \leq& t - \frac{t^3}{3!} + \frac{t^5}{5!} - \frac{t^7}{7!} = Q_3(t). \end{aligned}$$

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 die Folgerung aus dem Monotoniesatz über den Fehler bei der Approximation durch Tangenten. Wieder mit Archimedes Argument bekommen wir dieselbe Ungleichung für die Grenzfunktion:

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

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

$$\text{Es gilt:}\quad f' = -g, \quad g' = f.$$

Diese Gleichungen zeigen, daß f' und g' wieder differenzierbar sind, $f'' = -f, g'' = -g$, usw.: Die Grenzfunktionen sind unendlich oft differenzierbar.
Bezeichnung: $f = \cos,\ g = \sin$.

Zum Abschluß komme ich noch zu anderen 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:

$$k, l \geq k(n) \Rightarrow |a_k - a| \leq \frac{1}{n}, \quad |a_\ell - a| \leq \frac{1}{n}, \quad |a_k - a_\ell| \leq \frac{2}{n}.$$

In der letzten Ungleichung kommt der Grenzwert a nicht mehr vor! Das veranlaßt die
Definition. Eine Folge $\{a_k\}$ heißt *Cauchyfolge* wenn gilt: Zu jedem $n \in \mathbb{N}$ gibt es $k(n)$ so daß

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

Derartige Folgen sehen offenbar konvergenten Folgen sehr ähnlich. Tatsächlich haben wir:

Zweite Formulierung des Vollständigkeitsaxioms

Jede reelle Cauchyfolge besitzt einen Grenzwert.

Diese zweite Formulierung der Vollständigkeit impliziert die erste: Gegeben eine Intervallschachtelung $\{[a_k, b_k]\}$. Dann gilt für $\ell \geq k:\ a_k \leq a_\ell \leq b_k$, also ist $|a_\ell - a_k| \leq |b_k - a_k|$ Nullfolge. Daher ist $\{a_k\}$, und ebenso $\{b_k\}$, Cauchyfolge. Nach der zweiten Formulierung existiert ein (gemeinsamer) Grenzwert r, der natürlich in allen Intervallen der Schachtelung liegt.
Für die umgekehrte Richtung müssen wir zu einer Cauchyfolge eine Intervallschachtelung definieren. Das geschieht in mehreren Schritten:

1.) Eine Cauchyfolge $\{c_k\}$ ist beschränkt.
Beweis: Zu $n = 2$ finde $k(2)$ sodaß $k, \ell \geq k(2) \Rightarrow |c_k - c_\ell| \leq \frac{1}{2}$. Daraus folgt: $\min\{c_k; k \leq k(2)\} - \frac{1}{2} \leq c_\ell \leq \max\{c_k; k \leq k(2)\} + \frac{1}{2}$. Nenne die gefundene untere Schranke a_1, die obere 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.) $\{[a_m, b_m]\}$ ist eine Intervallschachtelung, hat also nach Voraussetzung einen Grenzwert r. Nach Konstruktion ist $\{c_k - r\}$ Nullfolge, also $\{c_k\}$ konvergent gegen r.

BEISPIEL. Sei $\{a_k\}$ eine "geometrisch majorisierte" Folge, d.h. es gelte

$|a_{k+2} - a_{k+1}| \leq q \cdot |a_{k+1} - a_k|,\ q < 1$. Dann ist $\{a_k\}$ Cauchyfolge.

Beweis. Es sei ohne Einschränkung $\ell > k$. Aus der Voraussetzung folgern wir zunächst durch $k-$malige Anwendung: $|a_{k+2} - a_{k+1}| \leq q^k \cdot |a_1 - a_0|$.
Danach ergibt die Dreiecksungleichung:

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

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

ANWENDUNG. f sei eine Abbildung mit Dehnungsschranke $q < 1$, $|f(y) - f(x)| \leq q \cdot |y - x|$ und a_0 beliebig. Dann ist $a_{k+1} := f(a_k), k \in \mathbb{N}$, eine geometrisch majorisierte Cauchyfolge.

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 $0 \leq x \Rightarrow |f'(x)| \leq 1/4$, erfüllt mit $a_0 = 0$ die Voraussetzungen und liefert den konvergenten Kettenbruch

$$\frac{1}{2 + \frac{1}{2+1/2\ldots}}$$

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

Eine weitere Formulierung der Vollständigkeit geht im wesentlichen auf Dedekind zurück.

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

1.) s ist obere Schranke von A (d.h. $a \in A \Rightarrow a \leq s$).
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, genannt $\sup A$.

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.

Anwendung: Die Bogenlänge dehnungsbeschränkter Kurven $p\colon [a,b] \to \mathbb{R}^2$.
Voraussetzung: $|p'| \le L$, also $|p(y) - p(x)| \le L \cdot |y - x|$.
Betrachte eine beliebige Einteilung $x = 0 < x_1 < \ldots < x_n = b$. Dann gilt für die Summe der Sehnenlängen

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

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

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

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

4. Komplexe Zahlen

Ziele dieses Abschnitts:
Definition der komplexen Zahlen.
Verträglichkeit der Multiplikation mit der euklidischen Geometrie: $|z \cdot w| = |z| \cdot |w|$.
Vollständigkeit.
Komplexe Differenzierbarkeit der rationalen Funktionen.
Gemischte Kettenregel für $c : I \to \mathbb{C}, f : \mathbb{C} \to \mathbb{C} : (f \circ c)' = (f' \circ c) \cdot c'$.
Veranschaulichung komplexer Funktionen.
Schrankensatz: $|f'| \le L \Rightarrow |f(w) - f(z)| \le L \cdot |w - z|$.

Etwa von 1500-1800 haben Mathematiker mit schlechtem Gewissen mit der "imaginären Einheit", $i^2 = -1$, 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.

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 $\mathbb{C} := \mathrm{span}_{\mathbb{R}}\{\mathbf{e}, \mathbf{i}\}$ 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 $\mathbb{C}$ schon:
$(r \cdot \mathbf{e}) \cdot (x \cdot \mathbf{e} + y \cdot \mathbf{i}) = rx \cdot \mathbf{e} + ry \cdot \mathbf{i} + (x \cdot \mathbf{e} + y \cdot i) \cdot (r \cdot \mathbf{e})$.
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, was die historische Vorge-

schichte nahe legt:

$$\mathbf{i} \cdot \mathbf{i} := -\mathbf{e},$$

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

Um sich an die komplexen Zahlen als zweidimensionale Zahlen zu gewöhnen, ist deren Verträglichkeit mit der euklidischen Geometrie wichtig. Die Addition $z \to a + z$, $a, z \in \mathbb{C}$, liefert die Translationen.
Definition der euklidischen Länge von $z = x + y \cdot \mathbf{i}$:

$$|z| := \sqrt{x^2 + y^2}.$$

Definition der zu $z = x + \mathbf{i}y$ konjugierten komplexen Zahl:

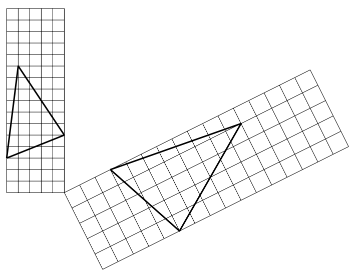
$$\overline{z} := x - y \cdot \mathbf{i}.$$

Satz (Verträglichkeit der euklidischen Geometrie mit der Multiplikation)
Für $c = a + b \cdot \mathbf{i}$ und $z = x + y \cdot \mathbf{i}$ gilt:
1.) $\overline{c \cdot z} = \overline{z} \cdot \overline{c}, \quad |z|^2 = z \cdot \overline{z}, \quad |c \cdot z| = |c| \cdot |z|.$
2.) Die Abbildung $z \to c \cdot z$ ist für $c \neq 0$ 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 $z \to c \cdot z$ 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.) $|c \cdot (z_i - z_j)| = |c| \cdot |z_i - z_j|,$
$i, j \in \{1, 2, 3\}.$



Multiplikation mit einer komplexen Zahl $c \neq 0$ ist eine winkeltreue Abbildung

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

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

Wir werden also auch im Komplexen durch Approximation neue interessante Funktionen gewinnen können.

Für Polynome können wir dieselben Rechnungen wie im Reellen machen, z. B.:
$|a|, |z| \le R \Rightarrow |z^n - a^n - n \cdot a^{n-1} \cdot (z-a)| \le n \cdot (n-1) \cdot R^{n-2} \cdot |z-a|^2$.
In Worten:
Die Potenzfunktionen $z \to z^n$ sind durch die linearen Funktionen $z \to a^n + n \cdot a^{n-1} \cdot (z-a)$ mit quadratischem Fehler approximierbar.

Wie im Reellen dehnt sich diese lineare Approximierbarkeit mit quadratischem Fehler auf Polynome oder auch $z \to 1/z$ (bei $a \neq 0$) aus. Schließlich können wir auch die Beweise der Differentiationsregeln einfach wiederholen:

Satz: f, g seien komplexe Funktionen, für die gilt: Es gibt andere komplexe Funktionen f', g', die gute lineare Approximationen liefern:
$|z-a| \le r \Rightarrow |f(z) - f(a) - f'(a) \cdot (z-a)| \le K_f \cdot |z-a|^2$,
$|z-a| \le r \Rightarrow |g(z) - g(a) - g'(a) \cdot (z-a)| \le K_g \cdot |z-a|^2$.
Dann gelten die Differentiationsregeln:
$(\alpha f + \beta g)' = \alpha \cdot f' + \beta \cdot g'$
$(f \cdot g)' = f' \cdot g + f \cdot g'$
$(f \circ g)' = (f' \circ g) \cdot g'$.
Auch für die zusammengesetzten Funktionen $\alpha \cdot f + \beta \cdot g$, $f \cdot g$, $f \circ g$ gilt, daß die aus den Ableitungen berechneten linearen Funktionen mit quadratischem Fehler approximieren.

Insbesondere sind damit die rationalen Funktionen $z \to P(z)/Q(z)$ außerhalb der Nullstellen des Nenners in dieser komplexen (und bisher noch unanschaulichen Weise) differenzierbar.

Ehe wir versuchen, diese Ableitungen anschaulicher zu machen, müssen wir fragen: Wie kann man sich 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 1-dimensionalen hat sich das nicht durchgesetzt, weil 1-dimensionale "Bilder" zu dürftig sind. Für komplexe Funktionen dagegen erhält man brauchbare Veranschaulichungen so: 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:

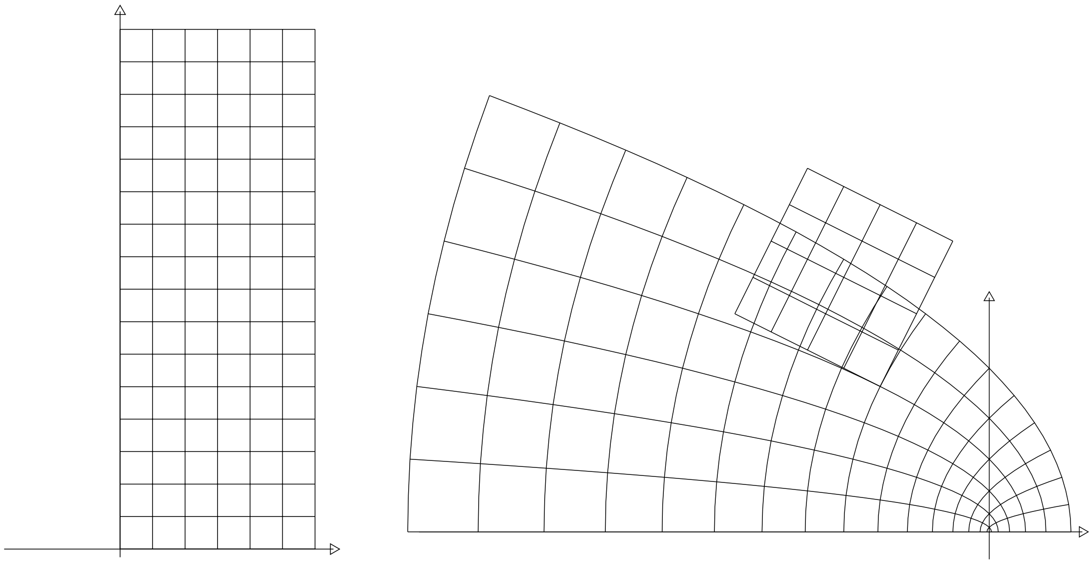


Bild unter $z \to z^2$, mit einer linearer Approximation

Offenbar stellt sich hierbei ein neues Problem: Gegeben eine z. B. differenzierbare Kurve $c : I \to \mathbb{C}$, wie können wir die Ableitung der Bildkurve $\tilde{c} = f \circ c : I \to \mathbb{C}$ 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.
$|t - t_0| \le r_c \Rightarrow |c(t) - c(t_0) - c'(t_0)(t - t_0)| \le K_c \cdot |t - t_0|^2$.

Satz (Gemischte Kettenregel)
Sei c wie eben und $\tilde{c} = f \circ c$, ferner $a = c(t_0)$ und f komplex differenzierbar:
$|z - a| \le r_f \Rightarrow |f(z) - f(a) - f'(a) \cdot (z - a)| \le K_f \cdot |z - a|^2$.
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 $t \to \tilde{c}(t_0) + \tilde{c}'(t_0) \cdot (t - t_0)$ von $\tilde{c}$ ab. Als Ungleichung:
$|t - t_0| \le r_{\tilde{c}} \Rightarrow |f \circ c(t) - f \circ c(t_0) - f'(c(t_0)) \cdot c'(t_0) \cdot (t - t_0)| \le K_{\tilde{c}} \cdot |t - t_0|^2$.

Beweis. Der Beweis folgt dem alten Muster. Ich gebe ihn an, weil hier reelle und komplexe Differenzierbarkeit gemischt werden.
Abkürzung: $\ell := |c'(t_0)| + K_c \cdot r_c$, $L := |f'(a)| + K_f \cdot r_f$. Zunächst wird r_c verkleinert, $r_1 := \min(r_c, r_f/\ell)$, so daß gilt:

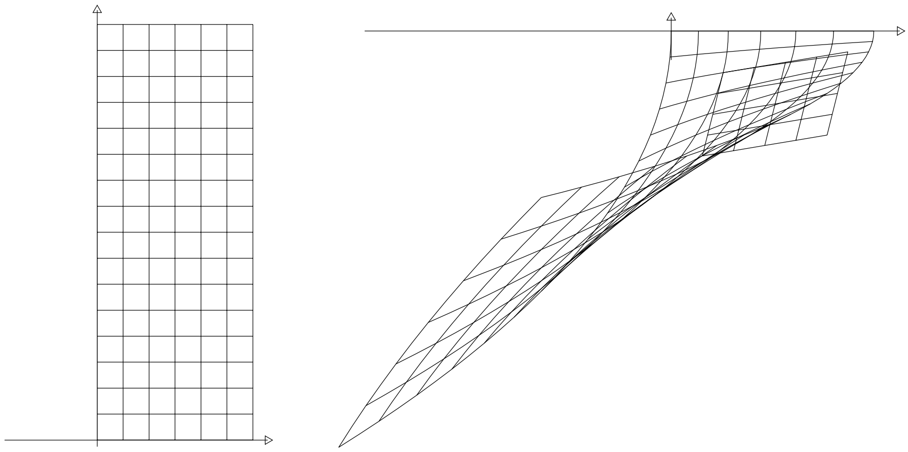
$|t-t_0| \le r_1 \Rightarrow |c(t)-c(t_0)| \le \ell \cdot r_1 \le r_f$,
also:
$|t-t_0| \le r_1 \Rightarrow |\tilde{c}(T)-\tilde{c}(t_0)| \le L \cdot |c(t)-c(t_0)| \le L \cdot \ell \cdot |t-t_0|$, und, mit $c(t_0)=a$,
$|t-t_0| \le r_1 \Rightarrow |f(c(t))-f(a)-f'(a)\cdot(c(t)-c(t_0))| \le K_f \cdot |c(t)-c(t_0)|^2$.
Hier wird $f'(a)\cdot(c(t)-c(t_0))$ ersetzt durch $f'(a)\cdot c'(t_0)\cdot(t-t_0)$, der Fehler $\le |f'(a)|\cdot K_c \cdot |t-t_0|^2$ wird mit der Dreiecksungleichung zu dem anderen Fehler geschlagen.
Erinnerung: $(|A-B| \le F_1, |B_C| \le F_2 \Rightarrow |A-C| = |(A-B)-(C-B)| \le F_1+F_2)$.
Mit der Abkürzung: $K_{\tilde{c}} := K_f \cdot L^2 \cdot \ell^2 + |f'(a)| \cdot K_c$, $r_{\tilde{c}} := r_1$ haben wir das Ziel erreicht:

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

Folgerung. (Winkeltreue komplex differenzierbarer Abbildungen.)
Sei f komplex differenzierbar. An allen Stellen a mit $f'(a) \ne 0$ ist die Abbildung f winkeltreu.

Beweis. Die linearen Approximationen $z \mapsto f(a)+f'(a)\cdot(z-a)$ sind nach dem ersten Satz 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.

Wegen dieser Folgerung ist es zweckmäßig, bei der Veranschaulichung komplexer Abbildungen f 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. Die hieran angepaßte Veranschaulichung der Ableitung besteht darin, in das Bild von f auch das Bild der linearen Approximation $z \to f(a) + f'(a)\cdot(z-a)$ einzuzeichnen. Im Punkte $f(a)$ sind dann die beiden Gitterlinien des Bildes der linearen Approximation Tangenten an die beiden Netzkurven des Bildes von f durch $f(a)$.



Nicht winkeltreues Bild unter $z \to \bar{z} + z^2/4$

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.
Voraussetzung. Auf der Strecke zwischen z und w sei f komplex differenzierbar mit $|f'| \le L$. Wir verlangen nur eine *nicht-gleichmäßige Tangentenapproximation:*
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:
$|c-a| \le r_a \Rightarrow |f(c) - f(a) - f'(a)\cdot(c-a)| \le K_a \cdot |c-a|^2$.
Behauptung. $|f(z) - f(w)| \le L \cdot |z-w|$.

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, also

$$\begin{array}{lcll} |f(z)-f(w)| & > & L\cdot|z-w|, & \text{z.B.} \\ |f(z)-f(w)| & \ge & (L+1/n)\cdot|z-w|. & \end{array}$$

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:
$|f(z_0) - f(m)| \ge (L+1/n)\cdot|z_0 - m|$
$|f(m) - f(w_0) \ge (L+1/n)\cdot|m - w_0|$.
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
$|f(z_k) - f(w_k)| \geq (L + 1/n) \cdot |z_k - w_k|, \quad |z_k - w_k| = 2^{-k} \cdot |z_0 - w_0|$.
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
$|f(z_{k+1} - f(w_{k+1})| \geq (L + 1/n) \cdot |z_{k+1} - w_{k+1}|, \quad |z_{k+1} - w_{k+1}| = 2^{-k-1} \cdot |z_0 - w_0|$.
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 eine der beiden Ungleichungen
$|f(z_k) - f(c)| \geq (L + 1/n) \cdot |z_k - c| \quad$ oder
$|f(w_k) - f(c)| \geq (L + 1/n) \cdot |w_k - c|$.
Der Widerspruch entsteht nun, wenn wir die Differenzierbarkeitsvoraussetzungen für f an der Stelle c verwenden:
Wähle k so groß, daß $|z_k - w_k| < \min(r_c, 1/(2nK_c))$ gibt, dann folgt für alle z auf der Strecke zwischen z_k und w_k:
$|f(z) - f(c) - f'(c) \cdot (z - c)| \leq K_c \cdot |z - c|^2$,
also wegen $|f'(c)| \leq L$ und $|z - c| \leq 1/(2nK_c)$:
$|f(z) - f(c) \leq (L + 1/2n) \cdot |z - c|$,
im Widerspruch zu den aus der Konstruktion hergeleiteten Ungleichungen für z_k oder 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 Zwischenwertsatz und der Cauchysche Integralsatz.

5. Potenzreihen

Ziel des Abschnitts: Ich behandele Potenzreihen, weil fast alle Funktionen, die individuelle Namen haben, als Potenzreihen beschrieben werden können. Außerdem wird die einfachste Majorisierungstechnik, der Vergleich mit der geometrischen Reihe, geübt. Potenzreihen verallgemeinern Polynome.

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 $\{P_n(z)\}$ wie auch (im Falle der Konvergenz) deren Grenzfunktion 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 $\{a_k z_0^k\}$ jedenfalls beschränkt ist, 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

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

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. Da wir über die uns schon bekannten Potenzreihen für exp, sin, cos wissen, daß sie für alle $z \in \mathbb{R}$ konvergieren, 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 $|z| \leq q \cdot R$, $(0 < q < 1)$ mit geometrischen Reihen majorisieren.

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 auch mit einem etwas größeren Faktor zufrieden sein. Ich benutze $0 < q < 1 \Rightarrow q < \sqrt{q} < 1$. — Das 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.

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 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 gilt für alle $k \geq k_q$

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

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 $s := \limsup \sqrt[k]{|a_k|}$. 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 Abschnitts, 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 $k \in \mathbb{N}$. Wegen der Monotonie der k-ten Wurzel folgt
$$\sqrt[k]{|a_k|} \cdot |z| \leq \sqrt[k]{M} \leq 1 + \tfrac{M}{k}, \text{ also } \quad s \cdot |z| = \limsup \sqrt[k]{|a_k|} \cdot |z| \leq 1.$$
Dies gilt für jedes $|z| < R$ (= Konvergenzradius), damit ist $s \cdot R \leq 1$ bewiesen.
Zweitens wird für jedes $q \in (0,1)$ gezeigt $q \leq s \cdot R$; 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:
$k \geq k_q \Rightarrow \sup \sqrt[k]{|a_k|} \leq s/\sqrt{q}\ ,\ |a_k| \cdot \left(\sqrt{q}/s\right)^k \leq 1.$

Also gilt für alle $k \geq k_q$ und $|z| \leq q/s$
$|P_{k+1}(z) - P_k(z)| = |a_{k+1} \cdot z^{k+1}| \leq |a_{k+1} \cdot (q/s)^{k+1}| \leq \sqrt{q}^{k+1}$.
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.

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 durch geometrische Majorisierung der differenzierten Potenzreihen $\{P_n'(z)\}$, $\{P_n''(z)\}$ beantworten. Das einzige Problem dabei sind die beim Differenzieren größer werdenden Koeffizienten: $a_k \to k \cdot a_k \to k \cdot (k-1) \cdot a_k$. Wir werden damit genau wie in den letzten Beweisen fertig: Wenn wir im Definitionsbereich den Konvergenzradius um den Faktor q verkleinern, so sind wir zufrieden, die Werte der Potenzreihen mit schlechter konvergenten geometrischen Reihen vergleichen zu können. Dazu genügt uns die schon erwähnte Ungleichung
$(n+1) \cdot q^n \leq 1 + q + \ldots + q^n \leq 1/(1-q)$ mit ihren Folgerungen (in dem q durch $q^{1/2}$, $q^{1/3}$ usw. ersetzt und die Ungleichung quadriert usw. wird):

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

Bemerkung. Es ist nicht notwendig, die differenzierten Reihen $\{P'(z)\}$, $\{P''(z)\}$ mit genau diesen Ungleichungen zu behandeln, man muß nur irgendeine Majorisierung schaffen. Daß für jedes feste k und $|q| < 1$ auch $n \to n^k \cdot q^n$ eine Nullfolge ist, ist jedenfalls ein weit verbreiteter Trick.
Satz (Geometrische Majorisierung der differenzierten Reihen).
Voraussetzung. Die Potenzreihe mit den Koeffizienten a_k habe einen Konvergenzradius $\geq r > 0$. Wähle $q \in (0,1)$.
Behauptung. Die differenzierten Potenzreihen $\{P_n'(z)\}$, $\{P_n''(z)\}$ sind in der Scheibe $|z| \leq q \cdot r$ geometrisch majorisiert, damit konvergent, und in Abhängigkeit von der Wahl von q *beschränkt* (Schranken im Beweis).

Beweis. Da nach Voraussetzung $\{P_n(z)\}$ für $|z| \leq \sqrt{q} \cdot r$ konvergiert, ist insbesondere $a_k \cdot (\sqrt{q}r)^k$ beschränkt: $|a_k| \cdot \sqrt{q}^k \cdot r^k \leq M$.
Daraus folgt in der kleineren Scheibe $|z| \leq q \cdot r$ die geometrische Majorisierung der Potenzreihe selbst:
$|P_k(z) - P_{k-1}(z)| = |a_k \cdot z^k| \leq |a_k \cdot q^k \cdot r^k| \leq M \cdot \sqrt{q}^k$
und wegen $\Sigma \sqrt{q}^k \leq 1/(1-\sqrt{q})$ die Schranke

$$|P_k(z)| = |a_0| + M/(1-\sqrt{q}) =: C_q \quad \text{für alle} \;\; k \in \mathbb{N}.$$

Setzen wir nun $q_0 = \sqrt{q}$, $q_1 = q_0^{1/2}$, $q_2 = q_0^{1/3}$, $q_j = q_0^{1/(j+1)}$, so finden wir für die differenzierten Reihen in derselben Scheibe $|z| \leq \sqrt{q} \cdot r$ die geometrische Majorisierung:

$|P_k' - P_{k-1}'(z)| = |k \cdot a_k \cdot z^{k-1}| \leq M/r \cdot k \cdot q_0^k \leq M/r \cdot q_1^k/(1-q_1)$,
also wieder mit $\sum q_1^k \leq 1/(1-q_1)$ die Schranke

$$|P_k'(z)| \leq |a_1| + M/r \cdot 1/(1-q_1)^2 =: L_q \quad \text{für alle} \quad k \in \mathbb{N};$$

$|P_k''(z) - P_{k-1}''(z)| = |k \cdot (k-1) \cdot a_k z^{k-2}| \leq M/r^2 \cdot k^2 q_0^k \leq M/r^2 \cdot q_2^k/(1-q_2)^2$,
also die Schranke

$$|P_k''(z) - P_{k-1}''(z)| \leq |a_2| + M/r^2 \cdot 1/(1-q_2)^3 =: B_q \quad \text{für alle} \quad k \in \mathbb{N}.$$

Da der eben geführte Majorisierungsbeweis nicht nur die Konvergenz der differenzierten Reihen liefert, sondern in der Scheibe $|z| \leq q \cdot r$ auch Schranken $|P_n'(z)| \leq L_q$, $|P_n''(z)| \leq B_q$ herleitet, ergibt sich sofort die Differenzierbarkeit der Grenzfunktion zusammen mit sogar quadratischer Approximation der Tangenten:

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

Beweis. 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 Archimedesargument 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:
$z, w \in D_{q \cdot r} \Rightarrow |P_\infty(w) - P_\infty(z) - Q_\infty(z) \cdot (w - z)| \leq B_q \cdot |w - z|^2 +$ 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_∞ differenzierbar ist mit der Ableitung $P_\infty' = Q_\infty$.
Mit anderen Worten: "Man kann Potenzreihen (in $D_{q \cdot r}$) gliedweise differenzieren",

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

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 als geometrische Reihe:

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

oder auch:

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

Der Vorteil des komplizierteren Ausdrucks ist, daß der Konvergenzkreis größer ist. Die Grenzfunktion der letzten Reihe bezeichnen wir mit $L_R(z)$, es gilt also $L_R'(z) = 1/z$ in der Scheibe $|z-R| < R$ und $L_R(1) = 0$.

Zur Exponentialfunktion. Für die durch die konvergente Potenzreihe definierte komplexe Exponentialfunktion

$$\exp(z) := \sum_{k=0}^{\infty} \frac{z^k}{k!}$$

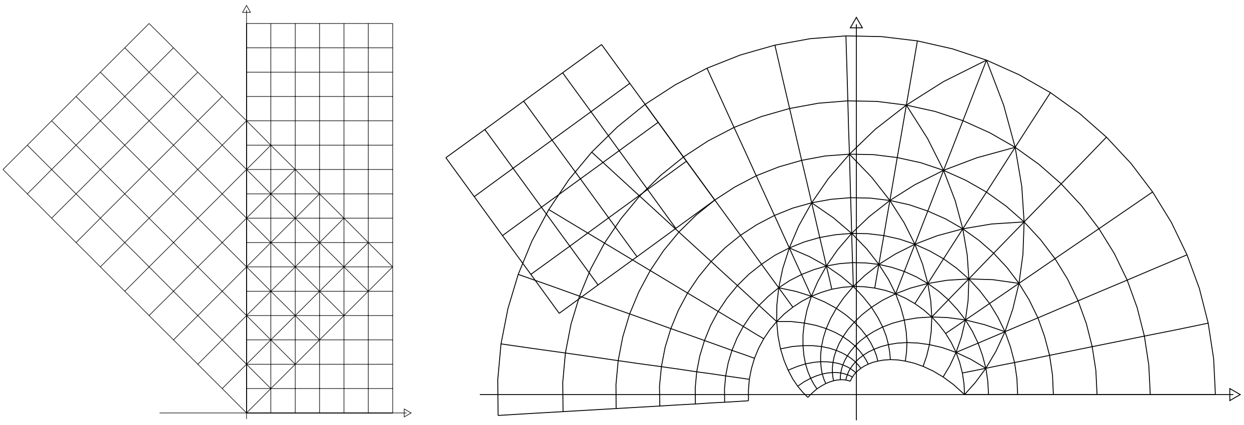
gilt

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

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 $q(z) = 1$.



Winkeltreue Polarkoordinaten als Bild der Exponentialfunktion,
mit Diagonalkurven und linearer Approximation.

Durch Vergleich der Potenzreihen findet man Eulers Formeln:
$\exp(iz) = \cos(z) + i \cdot \sin(z)$, $\exp(2\pi i) = \exp(0) = 1$,
$\cos z = (\exp(iz) + \exp(-iz))/2$,
$\sin z = (\exp(iz) - \exp(-iz))/2i$.
Daraus folgen die Additionstheoreme von sin und cos, z. B.:

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

Aus $\exp(2\pi i) = \exp(0) = 1$ folgt daher, daß cos und sin auch im Komplexen 2π-periodisch sind:
$\cos(z + 2\pi) = \cos(z)$, $\sin(z + 2\pi) = \sin(z)$.
Da exp $2\pi i$-periodisch ist, gibt es keine eindeutigen Umkehrfunktionen. Sei $\ell(z)$ eine (innere) Umkehrfunktion, also $\exp(\ell(z)) = z$.
Falls wir schon wüßten, daß wir nach der Kettenregel differenzieren könnten, so folgte:
$\exp'(\ell(z)) \cdot \ell'(z) = 1$, $z \cdot \ell'(z) = 1$, $\ell'(z) = 1/z$.
Nun haben wir gerade (zumindest in Kreisen $|z - R| < R$) Potenzreihen $L_R(z)$ mit $L'_R = 1/z$ und $L_R(1) = 0$ konstruiert. Sind sie Umkehrfunktionen der Exponentialfunktion? Die Funktion $h(z) := \exp(L(z))/z$ erfüllt $h(1) = 1$ 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 ($h(z) := L_R(z) - L_r(z) \Rightarrow h(1) = 0,\ h'(z) = 0$). 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
$z^\alpha := \exp(\alpha \cdot \text{Log } z)\ ,\ (\alpha \in \mathbb{C}, \text{Re } z > 0).$
Mit der Kettenregel folgt eine Verallgemeinerung der Differentiationsregel von ganzen auf komplexe Potenzen:
$(z^\alpha)' = \alpha/z \cdot \exp(\alpha \cdot \text{Log } z) = \alpha \cdot z^{\alpha-1}.$
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.

Aufgabe. Die Funktion $z \to (1+z)^\alpha$ und ihre Potenzreihe um $z = 0$ 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.
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: $g'' = -g$, und, falls $f(0) = 0$, so auch: $f = g$.

6. Stammfunktionen und Integrale

Ziele des Abschnitts: Konstruktion von Stammfunktionen für dehnungsbeschränkte Funktionen. Definition des Integrals, Hauptsatz der Differential- und Integralrechnung und Interpretation des Integrals als kontinuierliche Summe.

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:

Satz. 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 $\mathbb{R}^2, \mathbb{R}^d$).

Bemerkung. Wir werden sehen, daß dieser Satz den Hauptteil 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 Fehlerkontrolle approximieren:
Teile das Intervall $I = [x_0, x_n]$ in n gleiche Teile $x_0 < x_1 < \ldots < x_n$, $x_j - x_{j-1} = (x_n - x_0)/n = |I|/n$ und definieren die folgende "Sehnenapproximationen" 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:

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

Da die Differenz $f - sa_n$ die Dehnungsschranke $2L$ hat und da f und sa_n an den Teilpunkten übereinstimmen, $f(x_j) = sa_n(x_j)$ $(j = 0, \ldots, n)$, folgt für die maximale Abweichung zwischen f und sa_n:

$$x \in I \Rightarrow |f(x) - sa_n(x)| \leq L/n \cdot |x_n - x_0|.$$

Wir nennen diese Eigenschaft:
"f ist gleichmäßig stückweise linear approximierbar".
Für die Funktionen sa_n können wir explizit stückweise quadratische Stammfunktionen 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

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

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

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

Wegen des Vollständigkeitsaxioms konvergiert $\{SQ_n(x)\}$ gegen die Werte einer Grenzfunktion $SQ_\infty(x)$. Wieder wegen des Monotoniesatzes (für stückweise quadratische Funktionen) gilt:

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

Daraus folgt mit dem Archimedes-Argument

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

Also ist die Grenzfunktion SQ_∞ differenzierbar mit $SQ'_\infty(x) = f(x)$, wir haben eine Stammfunktion für f gefunden.

Definition des Integrals. Sei $f : I \to \mathbb{R}^d$ eine Funktion, über die wir noch nichts weiter voraussetzen. 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:

$$\limsup_{\delta(U) \to 0} W_U = \liminf_{\delta(U) \to 0} W_U,$$

und diese Zahl wird mit $\int_I f$ oder $\int_a^b f(x)dx$ bezeichnet und "Integral von f über $[a, b]$" genannt. 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.
Dehnungsbeschränkte $f : I \to \mathbb{R}^d$ sind Riemann integrierbar; f besitzt eine Stammfunktion $F : I \to \mathbb{R}^d,\ F' = f$, und es gilt

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

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 Differtential- und Integralrechnung ist allerdings knapp 200 Jahre älter als die Stetigkeit.

Zur Interpretation des Integrals.
Zunächst hat man eine formale Ähnlichkeit mit Summen:

1.) $\sum_k(\alpha \cdot a_k + \beta \cdot b_k) = \alpha \cdot \sum_k a_k + \beta \cdot \sum_k b_k$,	linear
$\int(\alpha \cdot f(x) + \beta \cdot g(x))dx = \alpha \cdot \int f(x)dx + \beta \cdot \int g(x)dx.$	linear
2.) $\sum_{k=1}^m a_k + \sum_{k=m+1}^n a_k = \sum_1^n a_k$,	assoziativ
$\int_a^b f + \int_b^c f = \int_a^c f$,	intervalladditiv
3.) $a_k = b_{k+1} - b_k \Rightarrow \sum_{k+1}^n a_k = b_{n+1} - b_1$,	Teleskopsumme
$f = F' \Rightarrow \int_a^b f(x)dx = F(b) - F(a)$,	Hauptsatz
4.) $a_k \leq b_k \Rightarrow \sum a_k \leq \sum b_k$,	Monotonie
$f \leq g \;\Rightarrow \int_a^b f \;\leq \int_a^b g$,	Monotonie
5.) $\lvert \sum a_k \rvert \leq \sum \lvert a_k \rvert$,	verallgemeinerte Dreiecksungleichung
$\lvert \int_a^b f(x)dx \rvert \leq \int_a^b \lvert f(x) \rvert dx$,	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:

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

daher folgt für die Integrale:

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

Das Archimedes-Axiom beseitigt den Fehler $2/n$.

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

Wir hatten die Bogenlänge dehnungsbeschränkter Kurven $p : [a,b] \to \mathbb{R}^d$ (damals $d = 2$) definiert als

Länge $(p([a,b])) := \sup\{\sum_{k=1}^N |p(x_k) - p(x_{k-1})|;$
$a = x_0 < x_1 < \ldots < x_N = b$ eine 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:

$$|p(x_k) - p(x_{k-1})| = \Big|\int_{x_{k-1}}^{x_k} p'(x)sx\Big| \leq \int_{x_{k-1}}^{x_k} |p'(x)|dx,$$

also

$$\sum_{k=1}^N |p(x_k) - p(x_{k-1})| \leq \int_a^b |p'(x)|dx.$$

Das Integral ist damit eine obere Schranke für die Länge beliebiger Sehnenzüge, und wir erhalten nach Definition der Bogenlänge:

$$\text{Länge } (p[a,b]) \leq \int_a^b |p'(x)|dx.$$

Sowohl die Länge wie das Integral betrachten wir als Funktionen der oberen Grenze:

$$L(t) := \text{Länge } (p([a,t])), \quad F(t) := \int_a^t |p'(x)|dx,$$

und wir wollen die Gleichheit dieser Funktionen zeigen. Wegen $L(a) = F(a) = 0$ würde dazu $L'(t) = F'(t)$ genügen. Dazu betrachten wir die Differenzenquotienten:

$$\frac{|p(t_2) - p(t_1)|}{t_2 - t_1} \leq \frac{L(t_2) - L(t_1)}{t_2 - t_1} = \frac{\text{Länge } (p([t_1, t_2]))}{t_2 - t_1} \leq \frac{1}{t_2 - t_1}\int_{t_1}^{t_2} |p'(t)|dt.$$

Links wegen der Differenzierbarkeit von $p(\ )$, rechts wegen der Dehnungsbeschränktheit von $p'(\ )$ finden wir zu jedem $n \in \mathbb{N}$ ein $\delta(n, t_1)$, so daß gilt

$$t_2 \in [t_1 - \delta, t_1 + \delta] \Rightarrow \left| \frac{p(t_2) - p(t_1)}{t_2 - t_1} - p'(t_1) \right| \leq \frac{1}{n}, \quad \left| \frac{1}{t_2 - t_1} \cdot \int_{t_1}^{t_2} |p'(t)| dt - |p'(t_1)| \right| \leq \frac{1}{n},$$

also

$$t_2 \in [t_1 - \delta, t_1 + \delta] \Rightarrow |p'(t_1)| - \frac{1}{n} \leq \frac{L(t_2) - L(t_1)}{t_2 - t_1} \leq |p'(t_1)| + \frac{1}{n}.$$

Damit ist $L(\ )$ differenzierbar mit der Ableitung $L'(t_1) = |p'(t_1)|$ für alle $t_1 \in [a, b]$. Damit ist $F = L$ bewiesen und

$$\text{Länge } (p([a, x]) = \int_a^x |p'(t)| dt.$$

gezeigt.

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:

$$\text{Zurückgelegter Weg } \ L(t) = \int_0^t |p'(t)| dt = \text{ Integral der Geschwindigkeit.}$$

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 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:

$$\text{Schwerpunkt } S = \sum_{k=1}^{N} \frac{m_k}{M} \cdot p_k, \quad \text{Gesamtmasse } M = \sum m_k \ .$$

Nun sei $t \to m(t) > 0$ eine positive Funktion mit $\int_a^b m(t)dt = 1$. Wir wollen für eine Funktion $f : [a, b] \to \mathbb{R}^3$ das Integral $\int_a^b f(t) \cdot m(t)dt$ als Mittelwert interpretieren. Wir betrachten Riemann-Summen $\sum_{k=1}^N f(\tau_k) \cdot m(\tau_k) \cdot (t_k - t_{k-1})$. Diese können wir offenbar als Mittelwert der $f(\tau_k)$ mit den Massen $m(\tau_k) \cdot (t_k - t_{k-1})$ interpretieren, *wenn* wir die Zwischenwerte τ_k für die Riemann-Summen so wählen, daß gilt:

$$m_k := m(\tau_k) \cdot (t_k - t_{k-1}) = \int_{\tau_{k-1}}^{\tau_k} m(\tau)d\tau, \quad \sum m_k = \int_a^b m(t)dt = 1.$$

(Die Möglichkeit dieser Wahl haben wir noch nicht bewiesen, ich verweise auf den Zwischenwertsatz im nächsten Abschnitt.)
Jede Verfeinerung der Einteilung für die Riemann-Summen bezieht immer mehr Funktionswerte $f(\tau_k)$ in diese Mittelung ein. Daher gibt die umgangssprachliche Formulierung:

$\int_a^b f(t) \cdot m(t)dt$ ist ein kontinuierlicher Mittelwert der Werte $f(t)$,
gemittelt mit der Massenverteilung $m(t)$

eine brauchbare Umschreibung des Sachverhaltes.

7. Stetigkeit und gleichmäßige Konvergenz

Ziele dieses Abschnitts: Stetige Funktionen als Konvergenz erhaltende Funktionen. ϵ-δ-Kriterium. Hauptsätze über stetige Funktionen. Überdeckungssatz von Heine-Borel. Konvergenz von Folgen stetiger Funktionen. Integration und Differentiation der Grenzfunktionen. Neuartige Beispiele stetiger Funktionen (Cantor, Weierstraß, Hilbert).

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 die Frage nach den konvergenzerhaltenden Funktionen wohl naheliegend.

Definition. Eine Funktion f heißt *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 $F : \mathbb{R} \to \mathbb{R}^d$ und im weiteren Verlauf in immer allgemeineren Situationen. Natürlich sind dehnungsbeschränkte Abbildungen stetig: $|F(x) - F(y)| \leq L \cdot |x - y|$ und $\lim a_k = a$ liefert $|F(a_k) - F(a)| \leq L \cdot |a_k - a|$, 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, am Ende dieses Abschnitts. Vorher beweisen wir wichtige Eigenschaften.

Satz. $f : (a, b) \to \mathbb{R}$ sei stetig bei $c \in (a, b)$ und positiv, $f(c) > 0$. Dann ist f positiv in einem Intervall $(c - \delta, c + \delta)$.

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 $f(a_1) \leq 0$. Im zweiten Fall wähle $n_2 > n_1$ so, daß $a_2 \notin (c - 1/n_2, x + 1/n_2)$; 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.

Satz. Eine Funktion $f : (a, b) \to \mathbb{R}$ ist genau dann stetig bei $c \in (a, b)$, falls es zu jeder Fehlerschranke $\epsilon > 0$ ein Garantieintervall $(c - \delta, c + \delta)$ gibt $(\delta > 0)$, so dass gilt:

$$x \in (c - \delta, c + \delta) \Longrightarrow |f(x) - f(c)| \leq \epsilon.$$

Beweis. a) Diese Eigenschaft sei erfüllt und $\{a_k\}$ 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

$$k \geq k_\epsilon \Longrightarrow |f(c_k) - f(c)| \leq \epsilon.$$

Zunächst bestimme zu ϵ ein Garantieintervall $(c - \delta, c + \delta)$ und dann mit der Konvergenz von $\{c_k\}$ ein k_δ mit
$k \geq k_\delta \Rightarrow |c_k - c| \leq \delta \Rightarrow |f(c_k) - f(c)| \leq \epsilon$.
D. h. man kann $k_\epsilon := k_\delta$ wählen.
b) Umgekehrt müssen wir aus der Stetigkeit von f folgern, daß zu jedem $\epsilon > 0$ ein $\delta > 0$ existiert, so daß gilt
$|x - c| < \delta \Rightarrow |f(x) - f(c)| < \epsilon$.
Wäre das Finden eines solchen Garantieintervalls $(c - \delta, c + \delta)$ für eine Fehlerschranke ϵ^* unmöglich, so liefert derselbe Beweis wie im vorhergehenden Satz einen Widerspruch: Wir können dann nämlich eine gegen c konvergente Folge $\{c_k\}$ finden mit $|f(c_k) - f(c)| \geq \epsilon^*$, im Widerspruch zu $\lim f(c_k) = f(c)$.

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 erfordert unser erster Satz nun gar keinen Beweis mehr: 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.)

Eine anschauliche Eigenschaft stetiger Funktionen etabliert der folgende Satz, der

Zwischenwertsatz. $f : [a, b] \to \mathbb{R}$ sei stetig, $f(a) < 0$, $f(b) > 0$. Dann gibt es $c \in (a, b)$ mit $f(c) = 0$. (Das gilt nicht nur für 0 sondern für jedes $w \in [f(a), f(b)]$.)

Beweis. Wir konstruieren eine Intervallschachtelung $a_1 = a$, $b_1 = b$. Ferner seien $[a_1, b_1] \supset [a_2, b_2] \supset \cdots \supset [a_n, b_n]$ schon konstruiert mit $f(a_n) \leq 0$, $f(b_n) \geq 0$. 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.

Offenbar ist die stetige Funktion $f(x) = 1/x$ auf dem Intervall $(0, 1)$ 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 hat 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: $a_1 = a$, $b_1 = b$. Nun sei $[a_1, b_1] \supset \cdots \supset [a_n, b_n]$ schon so konstruiert, daß f auf diesen Intervallen nicht beschränkt ist. Betrachte $m = (a_n + b_n)/2$. Falls f auf $[a_n, m]$ nicht beschränkt ist, setze $a_{n+1} = a_n$, $b_{b+1} = m$; andernfalls kann f auf $[m, b_n]$ nicht beschränkt sein, dann setze $a_{n+1} = m$, $b_{n+1} = b_n$. Diese Intervallschachtelung hat einen Grenzpunkt $c \in [a, b]$, in dem f stetig ist. Wir finden also zu $\epsilon = 1$ ein Garantieintervall $(c - \delta_1, c + \delta_1) \cap [a, b]$, auf dem gilt:
$x \in (c - \delta_1, c + \delta_1) \Rightarrow |f(x) - f(c)| \leq 1 \Rightarrow |f(x)| \leq |f(c)| + 1.$
Auf dem Garantieintervall ist f also beschränkt. Außerdem gilt $[a_n, b_n] \subset (c - \delta_1, c + \delta_1)$ falls nur n so groß ist, daß $|b_n - a_n| = |b_1 - a_1| \cdot 2^{-n} < \delta_1$. Der Widerspruch besteht darin, daß nach Konstruktion f auf $[a_n, b_n]$ nicht beschränkt ist.

Im Laufe der Zeit hat sich gezeigt, daß dieser Beweis die Stetigkeit von f nur wenig benötigt aber stattdessen wesentlich an einer wichtigen Eigenschaft des Intervalls $[a, b]$ hängt. Es gilt nämlich der

ÜBERDECKUNGSSATZ von Heine-Borel. Zu jedem Punkt $x \in [a, b]$ sei ein $\delta_x > 0$ gewählt. 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. $[a, b] \subset \bigcup_k (x_k - \delta_k, x_k - \delta_k)$.

Beweis. Die Argumentation ist fast dieselbe wie im letzten Beweis. 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: $[a_n, b_n]$ sei schon konstruiert und $m = (a_n + b_n)/2$; dann ist $[a_n, m]$ oder $[m, b_n]$ nicht endlich überdeckbar und liefert $[a_{n+1}, b_{n+1}]$. Schließlich sei c der Grenzpunkt der Intervallschachtelung. Offenbar überdeckt das eine Intervall $(c-\delta_c, c+\delta_c)$ die Intervalle $[a_n, b_n]$ mit $|b_1-a_n| = |b_1-a_1| \cdot 2^{-n} < \delta_c$. 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 den Beschränktheitssatz. 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. Offenbar gilt dann: Jede auf $[a, b]$ lokal beschränkte Funktion f ist sogar beschränkt, denn das Maximum der endlich vielen Schranken auf den Intervallen $(x_k - \delta_k, x_k + \delta_k)$ einer endlichen Überdeckung genügt. Mit dieser Argumentation hat man eine leicht anwendbare Beweismethode entwickelt, 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: $x, y \in [a, b]$ und $|x - y| < 2 \cdot \delta_x \Rightarrow |f(x) - f(y)| < \epsilon/2$. Endlich viele der Intervalle $(x - \delta_x, x + \delta_x)$ 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 mit $x \in (x_k - \delta_k, x_k + \delta_k)$; 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.

Bemerkung. Diese Beweise sind in ihren Formulierungen stark standardisiert. Man hat sie jedoch erst verstanden, wenn man weiß, was für Variationen zulässig sind. Meiner Meinung nach sind dafür mündliche Diskussionen unerläßlich.

Approximationssatz. 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

$$x, y \in [a,b], |x-y| < \delta \Rightarrow |f(x) - f(y)| < \frac{\epsilon}{2}.$$

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, $t \in [t_{j-1}, t_j] \Rightarrow \ell(t) = f(t_{j-1}) + (f(t_j) - f(t_{j-1}))/(t_j - t_{j-1}) \cdot (t - t_{j-1})$. Dann gilt

$$|f(t) - \ell(t)| \leq |f(t) - f(t_j)| + |\ell(t) - \ell(t_j)| < \frac{\epsilon}{2} + \frac{\epsilon}{2}.$$

Korollar. Stetige Funktionen f besitzen Stammfunktionen F und sind integrierbar mit

$$F(b) - F(a) = \int_a^b f(x)dx, \quad F'(x) = f(x).$$

Beweis. Unsere frühere Konstruktion hängt nur von der stückweise linearen Approximierbarkeit ab.

Wir kommen zu dem letzten der berühmten Sätze über stetige Funktionen, dem

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 $S = \sup f([a_n, b_n])$: $a_1 = a$, $b_1 = b$; 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 $a_{n+1} = m$, $b_{n+1} = b_n$. Der Konvergenzpunkt dieser Intervallschachtelung sei $c \in [a,b]$, wir behaupten $S = f(c)$. Nach Definition von S gilt $f(c) \leq S$; wäre $f(c) < S$ so setze $\epsilon = (S - f(c))/2$ und wähle δ, so daß

$$|x - c| < \delta \Rightarrow |f(x) - f(c)| < \epsilon \Rightarrow f(x) < f(c) + \epsilon = S - \epsilon.$$

Damit ist $(S - \epsilon)$ obere Schranke für f auf $(c - \delta, c + \delta)$ und auf allen $[a_n, b_n] \subset (c - \delta, c + \delta)$, im Widerspruch zu $S = \sup f([a_n, b_n])$.

Aufgabe. Formuliere 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 $x \ldots$.

Erstaunlicherweise sind die wesentlichen Eigenschaften, die stetige Funktionen als einzelne Funktionen haben, hiermit 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.

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:

$D = [0, 1]$, $f_n(x) := x^n$, $f(1) = 1$, $f(x) = 0$ für $x \in [0, 1)$.

Wie in den früheren Abschnitten gilt: 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 gleichmäßige Fehlerkontrolle.

Voraussetzung. Zu jedem $\epsilon > 0$ und jedem $x \in D$ gibt es ein $\delta_x > 0$, ausdrücklich unabhängig von n, so daß gilt:

$$x, y \in D\ ,\ |y - x| < \delta_x \Rightarrow |f_n(y) - f_n(x)| < \epsilon.$$

Behauptung: Für die Grenzfunktion gilt:

$$x, y \in D\ ,\ |y - x| < \delta_x \Rightarrow |f(y) - f(x)| < \epsilon.$$

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

$$x, y \in D,\ |y - x| < \delta_x \Rightarrow |f(y) - f(x)| < \epsilon + r_n + s_n,$$

und das Archimedes-Argument beseitigt die Nullfolgen.

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 für größere Folgenindices immer schlechter. Ein erfolgreicher Ausweg ist, die Konvergenz "gleichmäßig" vorauszusetzen:

Definition. 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:

$$m, n \geq n_\epsilon \Longrightarrow |f_m(x) - f_n(x)| \leq \epsilon.$$

Ich sage 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:

$$x \in (a, b), n \geq n_\epsilon \Longrightarrow |f(x) - f_n(x)| < \frac{\epsilon}{3}.$$

Dann wird für ein $n^* \geq n_\epsilon$ die Stetigkeit der einen Funktionen f_{n^*} benutzt und ein δ_x gewählt, so daß gilt:

$$x, y \in (a, b), |x - y| < \delta_x \Longrightarrow |f_{n^*}(y) - f_{n^*}(x)| < \frac{\epsilon}{3}.$$

Aus der Dreiecksungleichung folgt die Behauptung:

$$\begin{aligned} x, y \in (a, b), |x - y| < \delta_x \Rightarrow |f(x) - f(y)| = \\ &= |(f(x) - f_{n^*}(x)) + (f_{n^*}(x) - f_{n^*}(y)) + (f_{n^*}(y) - f(y))| < \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3}. \end{aligned}$$

Bemerkung. Anders als früher hat man keine Kontrolle darüber, wie stark verkleinert δ zu kleinerem ϵ gewählt werden muß, denn zu einem kleineren ϵ muß man eine Approximation f_{n^*} der Grenzfunktion mit größerem Index n^* benutzen, so daß man damit rechnen muß, ein sehr viel kleineres δ_x wählen zu müssen.
Der bewiesene Satz hat unmittelbare Anwendungen auf Integration und Differentiation.

Satz über Integration konvergenter Folgen. 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 Funktionen f_n Stammfunktionen haben, also integrierbar sind.
a) Falls die $\{f_n\}$ gleichmäßig konvergieren, so gilt

$$\lim \int_a^b f_n(x)dx = \int_a^b \lim f_n(x)dx = \int_a^b f(x)dx.$$

b) Falls die Fehlerkontrolle der f_n gleichmäßig in n ist, so gilt

$$\lim \int_a^b f_n(x)dx = \int_a^b \lim f_n(x)dx = \int_a^b f(x)dx.$$

Beweis. Die Voraussetzungen a) oder b) sichern, daß die Grenzfunktion f stetig, also integrierbar ist. Die gleichmäßige Konvergenz a) verträgt sich in besonders einfacher Weise mit der Integration:

$$||f-g|| < \epsilon \Rightarrow |\int_a^b f(x)dx - \int_a^b g(x)dx| \leq \int_a^b ||f-g||dx = ||f-g|| \cdot |b-a|.$$

Daher folgt die Vertauschungsregel $\lim \int = \int \lim$.
Im Fall 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:

$$x, y \in [a,b], |x-y| < \delta \Rightarrow |f_n(y) - f_n(x)| < \epsilon.$$

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

$$|\Sigma_1^{(n)} - \Sigma_2^{(n)}| < \epsilon \cdot \textstyle\sum_{j=1}^N |t_j - t_{j-1}| = \epsilon \cdot |b-a|,$$

und nach Definition des Integrals also auch

$$|\Sigma_1^{(n)} - \textstyle\int_a^b f_n(x)| < \epsilon \cdot |b-a|.$$

Diese Abschätzung gilt wegen der gleichmäßigen Fehlerkontrolle auch für jede Riemann-summe Σ_1 der Grenzfunktion f zur gleichen Intervalleinteilung:

$$|\Sigma_1 - \textstyle\int_a^b f(x)dx| < \epsilon \cdot |b-a|.$$

Dies sind nun die Voraussetzungen für ein $\epsilon/3$-Argument: Da $\Sigma_1^{(n)} - \Sigma_1$ Summe von N festen Nullfolgen ist, können wir $n \geq n_\epsilon$ so groß wählen, daß $|\Sigma_1^{(n)} - \Sigma_1| < \epsilon \cdot |b-a|$ ist, also:

$$n \geq n_\epsilon \Rightarrow |\textstyle\int_a^b f_n(x)dx - \int_a^b f(x)dx| < 3\epsilon \cdot |b-a|.$$

Diese Abschätzung besagt gerade:

$$\lim_{n\to\infty} \int_a^b f_n(x)dx = \int_a^b f(x)dx = \int_a^b \lim f_n(x)dx,$$

wie behauptet.

Satz über Differentiation konvergenter Folgen. $f_n : [a,b] \to \mathbb{R}$ sei stetig differenzierbar (d. h. f_n' sei stetig); $\{f_n(a)\}$ und $\{f_n'(x)\}$ für alle $x \in [a,b]$ seien konvergent mit

$\lim f_n'(x) =: g(x)$. Außerdem: $\{f_n'\}$ erfülle eine der beiden Voraussetzungen a) oder b) über die Integration der Folge $\{f_n'\}$. 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. Außerdem treten ziemlich bald 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 leichterer und genetischerer 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 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. Das erste Beispiel ist
Die Cantortreppe. Ich erinnere daran, daß eine differenzierbare Funktion mit Ableitung 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 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 mit der Grenzfunktion übereinstimmen. Die erste Funktion ist:

$$f_1(x) := \left\{ \begin{array}{ll} \frac{3}{2} \cdot x & \text{für } 0 \le x \le \frac{1}{3} \\ \frac{1}{2} & \text{für } \frac{1}{3} \le x \le \frac{2}{3} \\ \frac{3}{2} \cdot x - \frac{1}{2} & \text{für } \frac{2}{3} \le x \le 1 \end{array} \right\}$$

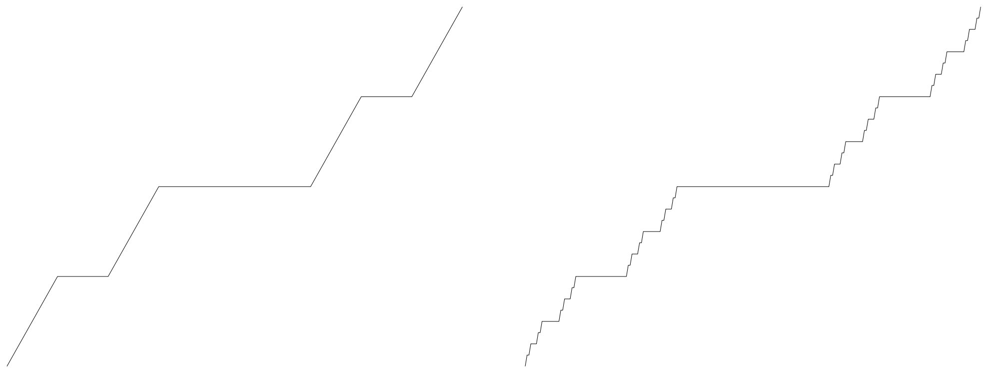
Die folgenden Funktionen f_k werden rekursiv definiert.Um f_{k+1} zu definieren, wird der Graph von f_k horizontal auf ein Drittel, vertikal auf die Hälfte gestaucht; zwei solche verkleinerten Kopien definieren f_{k+1} auf dem linken und rechten Drittel von $[0,1]$, in der Mitte ist f_{k+1} gleich $1/2$. In Formeln:

$$f_{k+1}(x) := \left\{ \begin{array}{ll} \frac{1}{2} \cdot f_k(3x) & \text{für } 0 \le x \le \frac{1}{3} \\ \frac{1}{2} & \text{für } \frac{1}{3} \le x \le \frac{2}{3} \\ \frac{1}{2} \cdot (1 + f_k(3x-2)) & \text{für } \frac{2}{3} \le x \le 1. \end{array} \right.$$

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:

$$x \in [0,1] \;\Rightarrow\; |f_{k+1}(x) - f_k(x)| \le \tfrac{1}{3} \cdot 2^{-(k+1)}.$$

Die Folge $\{f_k\}$ stetiger Funktionen ist also gleichmäßig geometrisch majorisiert, konvergiert also gegen 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 der Gesamtlänge $(2/3)^k$. Das wird zusammengefaßt durch die Formulierung: Die Cantortreppe ist “fast überall” differenzierbar mit Ableitung 0.



Zwei Approximationen der monotonen stetigen Cantortreppe, $f' = 0$ fast überall.
Die Stetigkeit dieser Approximationen ist gleichmäßig 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:
$|f_k(x) - f_{k+1}(x)| \leq \frac{1}{3} \cdot 2^{-(k+1)}$.

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):
$\text{säg}(x) := \min(x - \text{floor}(x), \text{ceil}(x) - x) = \text{Abstand}(x, \mathbb{Z})$.

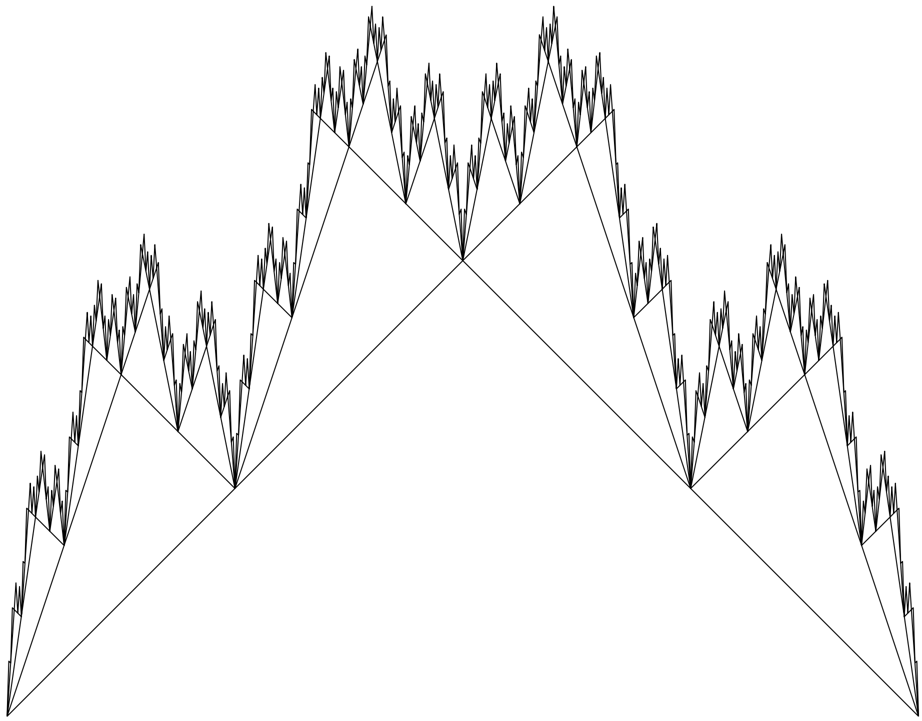
$$f_n(x) := \sum_{k=1}^{n} 2^{-k} \cdot \text{säg}(4^k \cdot x).$$

Offenbar gilt für alle x: $\qquad |f_n(x) - f_{n+1}(x)| \leq 2^{-(n+1)}$,
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. 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:

$$\begin{aligned} x := j \cdot 4^{-n},\ \Delta x = 0.5 \cdot 4^{-n} \ \Rightarrow\ & f_{n-1}(x) = f_n(x) = \ldots = W(x) \\ & f_n(x + \Delta x) = f_{n+1}(x + \Delta x) = \ldots = W(x + \Delta x) \\ & W(x + \Delta x) \geq W(x) + 2^{-(n+1)} \\ & \frac{W(x + \Delta x) - W(x)}{\Delta x} \geq 2^{n+2} \end{aligned}$$

Jede Zahl mit Nenner 4^{-n} ist auch Zahl mit Nenner 4^{-N} für $N \geq n$, daher gehen von allen Punkten $j \cdot 4^{-n}$ Sehnen mit Steigung $\geq 2^{N+2}$ über Intervallen der Länge $0.5 \cdot 4^{-N}$ aus, so daß 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 $\max(|W(\tilde{x}) - W(x_0)|,\ |W(\tilde{x}) - W(x_1)|) \geq 2^{-(n+2)}$. Deshalb gibt es auch von $\tilde{x}$ aus Sehnen, deren Steigungsbetrag $\geq 2^{n+1}$ 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).



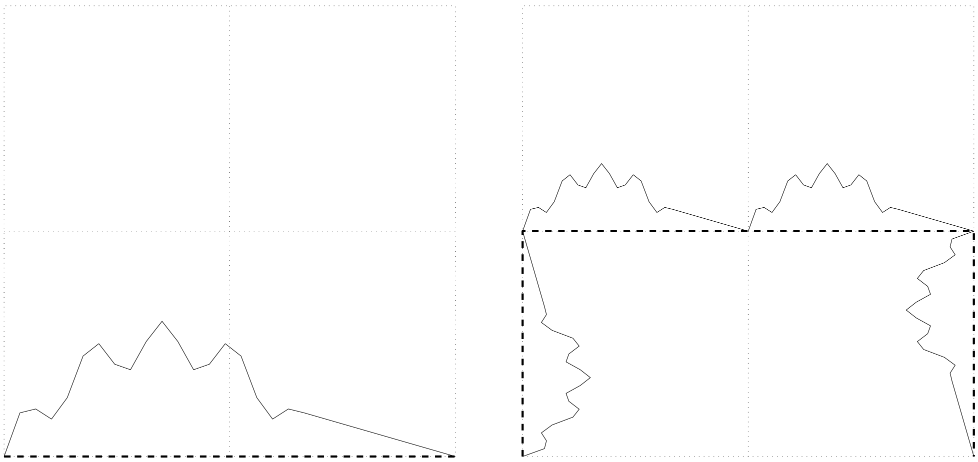
Weierstraß stetige, nirgends differenzierbare Funktion ist Grenzwert einer gleichmäßig konvergenten Folge; $W(x) = \sum_{n=0}^{\infty} 2^{-n} \cdot \text{säg}(4^n x)$.

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 $c_1(0) = (0,0),\ c_1(1) = (1,0)$.
$\tilde{c}(x) := \frac{1}{2} c_1(4 \cdot x),\ 0 \leq x \leq 0.25$, 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
$c_2 : [0,1] \to [0,1] \times [0,1]$,

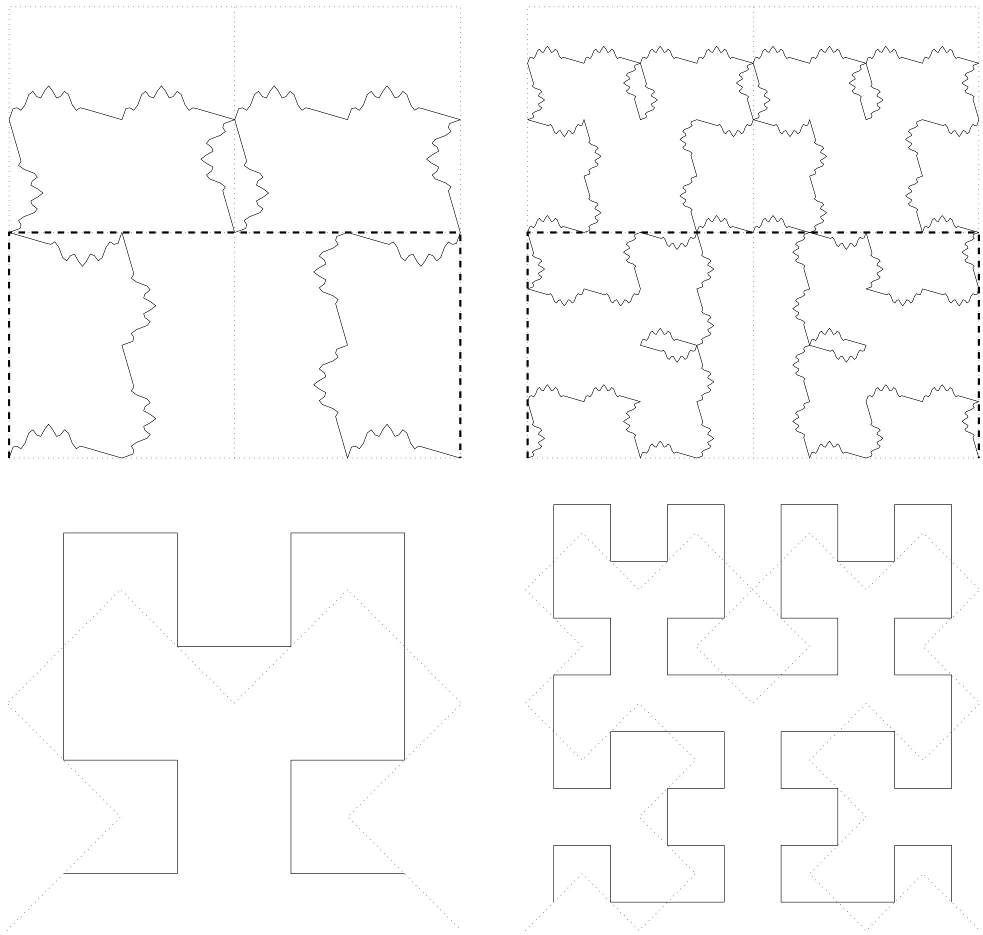
$c_2(0) = (0,0)$, $c_2(0.25) = (0, 0.5)$, $c_2(0.5) = (0.5, 0.5)$, $c_2(0.75) = (1, 0.5)$, $c_2(1) = (1, 0)$, wobei die vier Teilstücke nacheinander in den vier Teilquadraten
$[0, 0.5] \times [0, 0.5]$, $[0, 0.5] \times [0.5, 1]$, $[0.5, 1] \times [0.5, 1]$, $[0.5, 1] \times [0, 0.5]$ liegen, vergleiche die Abbildungen.
Offenbar gilt: Falls jeder Punkt aus dem Einheitsquadrat einen Abstand $\leq a$ vom Bild von c_1 hat, so hat jeder der Punkte aus einem der vier Teilquadrate einen Abstand $\leq \frac{1}{2} \cdot a$ von dem entsprechenden Teilstück der Kurve c_2. Ebenso sieht man:

$$|c_3(s) - c_2(s)| \leq \frac{1}{2} \cdot |c_2(s) - c_1(s)|.$$

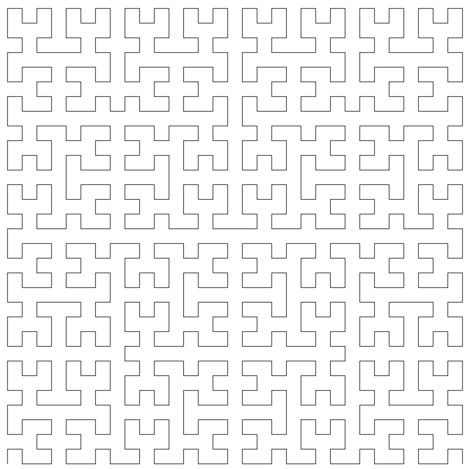
Deshalb ist die iterativ konstruierte Kurvenfolge $\{c_k\}$ gleichmäßig geometrisch majorisiert, konvergiert also gegen eine *stetige* Grenzkurve. Ferner ist jeder Punkt P des Einheitsquadrates Grenzwert der Folge der zu P nächsten Punkte P_k auf den Approximationen c_k, deshalb liegt P auf der Grenzkurve, 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.