

Anfang der Differentialrechnung.

In der ersten Hälfte des 17. Jahrhunderts war von Fermat und später von Descartes die analytische Geometrie entwickelt worden. Damit konnten Kurven durch Gleichungen definiert werden. Beide Autoren haben für einige gleichungsdefinierte Kurven schon Tangenten bestimmen können. Aber erst die Differentialrechnung, sowohl in der von Leibniz wie in der von Newton entwickelten Form, verallgemeinerte das alte Konzept "Steigung über einem Intervall" (oder "Sehnensteigung") in das neue Konzept "an die Kurve angepasste Steigung über einem Punkt" (oder "Tangentensteigung"). Das war in der Geometrie ein sensationeller Erfolg noch ehe Newton die Differentialrechnung auf die Physik anwandte. Aber es war nur der Anfang der Differentialrechnung. Besonders wichtig wurde, dass man von Voraussetzungen über diese punktwisen Steigungen zurückschließen konnte auf Eigenschaften der differenzierten Kurven. Der Monotoniesatz besagt: *Eine Funktion mit positiver Ableitung ist streng wachsend.* Daher heißen die Ableitungen auch "Steigungen".

Dasselbe noch einmal in einem physikalischen Kontext: Zurückgelegte Wege werden mit einer von der Zeit abhängigen Funktion $s(t)$ modelliert. Die experimentell zugänglichen Differenzenquotienten, also $(s(t_2) - s(t_1))/(t_2 - t_1)$, heißen in diesem Zusammenhang *Durchschnittsgeschwindigkeiten*. Mit den Grenzwerten dieser Durchschnittsgeschwindigkeiten werden wieder Ableitungen $s'(t)$ definiert. Wegen des Monotoniesatzes gilt wieder: Wenn diese Ableitungen positiv sind, wachsen die Wegfunktionen $s(t)$. **Deshalb** heißen die Ableitungen $s'(t)$ auch Momentangeschwindigkeiten $v(t)$. Und wieder wegen dieses Satzes heißen deren Ableitungen (momentane) Beschleunigungen.

Warum ich finde, dass die heutigen Schulbücher der Analysis Unrecht tun:

Es wird wohl niemand bezweifeln, dass die Menschen schon vor 1600 Lineale an Kurven legen und ungefähre Tangenten zeichnen konnten. Das modern gewordene sogenannte "graphische Differenzieren" macht daher keinen Versuch, den Schülerinnen und Schülern zu erklären, welch enormen begrifflichen Fortschritt Leibniz und Newton erreicht haben.

Mein Taschenrechner berechnet den Differenzenquotienten $(\sin(x+h) - \sin(x))/h$ mit $h = 10^{-12}$ als 0. Da alle Taschenrechner notwendig mit endlicher Stellenzahl rechnen, bekommt man dies Ergebnis immer, wenn man ausreichend kleine Differenzen benutzt. Solche Rechnungen liefern also nur dann plausible Zahlen, wenn man sehr deutlich bei endlichen Differenzen bleibt. Sie erklären also auch nicht, inwiefern Leibniz und Newton etwas radikal Neues geschaffen haben.

Die größte Wirkung der neuen Differentialrechnung war, dass sie eine Formulierung der Gesetze der Mechanik ermöglichte. Newton's "Principia" waren der Beginn der Theoretischen Physik und im Gefolge davon, der auf die Gesetze der Mechanik gegründeten Industrialisierung. Die Physik steht heute zur Motivierung der Analysis auf der Schule nicht mehr zur Verfügung. Damit ist es schwierig, den Schülerinnen und Schülern verständlich zu machen, dass sie mit der Analysis eine geistige Revolution kennen lernen.

Vor Erfindung der analytische Geometrie, waren die meisten Kurven durch geometrische Konstruktionen definiert, die die Kurventangenten mitlieferten (Ellipsen, Parabeln, Hyperbeln, Rollkurven wie die Cardioide, durch Gelenkmechanismen definierte Kurven wie die Lemniskate,...). In der analytischen Geometrie konnten Kurven durch Formeln definiert werden, deren Tangenten die Differentialrechnung berechnete. Im Vergleich mit der theoretischen Mechanik und ihren technischen Konsequenzen ist die Berechenbarkeit von Tangenten - ohnehin nur der Anfang der Analysis - eine deutlich schwächere Motivation. Auf der Schule

kommen parametrisierte Kurven $c(t) = (x(t), y(t))$ nicht vor, Kurven sind nur Graphen von Funktionen. Das schränkt die Möglichkeiten weiter ein. Also, kann man die Analysis überhaupt als Fortschritt vorführen?

Das Einfachste, was ich dazu kenne, ist die Brennspegeleigenschaft der Parabel und ihre geometrische Konstruktion aus Brennpunkt und Leitlinie. Beides wird möglich, wenn man die Ableitung der quadratischen Parabel zur Verfügung hat. Das zeigt dann: Ableitungen ermöglichen Argumente und Einsichten, die man ohne diesen Begriff nicht haben kann. Und das muss im Analysisunterricht immer wieder vorkommen: *Ableitungen ermöglichen Argumente und Einsichten, die man ohne diesen Begriff nicht haben könnte.*

Die quadratische Parabel und der Nutzen ihrer Tangenten.

Dieser Abschnitt ist lang, weil es um die **erste** Tangente nach der Kreistangente geht.

1. Definition der Tangente durch Vergleich mit Kreisen.

Der Kreis ist die erste Kurve, bei der “Tangenten” (aus dem Lateinischen: Berührende) vorkommen. Eine Tangente ist definiert als die Gerade durch einen Kreispunkt, die dort senkrecht auf dem Radius ist. Wegen des Satzes des Pythagoras sind alle anderen Punkte der Tangente weiter vom Kreismittelpunkt entfernt als der Berührungspunkt – der Kreis liegt auf einer Seite jeder seiner Tangenten. Ich formuliere eine weitere offensichtliche Eigenschaft, die sich sofort auf die Parabel übertragen läßt: Legt man einen Kreis so, dass er die x-Achse von oben berührt, so sind die rechtsseitigen Sehnensteigungen positiv, also größer als die Steigung 0 der Tangente und die linksseitigen Sehnensteigungen sind negativ.

Bei der **Normalparabel** $P(x) = x^2$ sind an der Stelle $x = a$

die rechtsseitigen Sehnensteigungen $(x^2 - a^2)/(x - a) = x + a > 2a$,

die linksseitigen Sehnensteigungen $(a^2 - x^2)/(a - x) = x + a < 2a$.

Wir erwarten wegen der Erfahrung am Kreis, dass die Tangentensteigung zwischen den rechtsseitigen und linksseitigen Sehnensteigungen liegen sollte, also $= 2a$ ist. Die Tangente ist dann die Gerade durch den Parabelpunkt (a, a^2) mit der Steigung $2a$, also

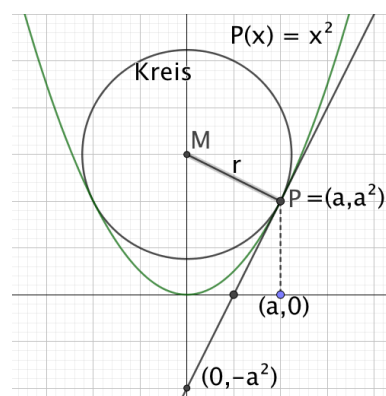
$$T_a(x) = 2a \cdot (x - a) + a^2 = 2a \cdot x - a^2.$$

Was kann man anführen, um zu rechtfertigen, diese Gerade “Tangente” zu nennen?

Die einfachste Eigenschaft ist: Die Parabel liegt oberhalb dieser Geraden, denn

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

Da das Wort Tangente “Berührende” bedeutet, wollen wir sehen, dass der Graph der Parabel an jeder Stelle zwischen einem Kreis und der gemeinsamen Tangente hindurchläuft, also mindestens so gut berührt wird wie der Kreis. Wir stellen uns vor, dass wir einen nicht zu kleinen Kreis von oben in die Parabel rutschen lassen, bis er stecken bleibt, weil er auf beiden Seiten die Parabel berührt. Sein Mittelpunkt liegt dann auf der y-Achse und der Radius zu einem Berührungspunkt steht senkrecht auf der **Tangente**. Das erlaubt uns, diese Kreise zu finden: Die Gerade durch $P = (a, a^2)$ senkrecht zur Tangente heißt



“Normale”. Sie hat die Gleichung $N_a(x) = \frac{-1}{2a} \cdot (x - a) + a^2 = \frac{-1}{2a} \cdot x + \frac{1}{2} + a^2$.

Dieser Kreis hat also den Mittelpunkt $M = (0, N_a(0)) = (0, \frac{1}{2} + a^2)$ und der Abstand r zum Berührungspunkt $P = (a, a^2)$ ist $r = \sqrt{1/4 + a^2}$. Die Kreisgleichung ist:

$$x^2 + (y - (1/2 + a^2))^2 = 1/4 + a^2.$$

Um sicher zu sein, dass die Parabel wirklich zwischen Kreis und Kreistangente hindurch läuft, müssen wir noch einsehen, dass alle Punkte (x, x^2) außerhalb dieses Kreises liegen. Wir müssen also zeigen

$$\begin{aligned} x^2 + (x^2 - (1/2 + a^2))^2 &\geq 1/4 + a^2 \\ \text{oder: } x^2 + x^4 - x^2 \cdot (1 + 2a^2) + 1/4 + a^2 + a^4 &\geq 1/4 + a^2 \\ \text{oder offensichtlich: } x^4 - 2a^2 \cdot x^2 + a^4 &= (x^2 - a^2)^2 \geq 0. \end{aligned}$$

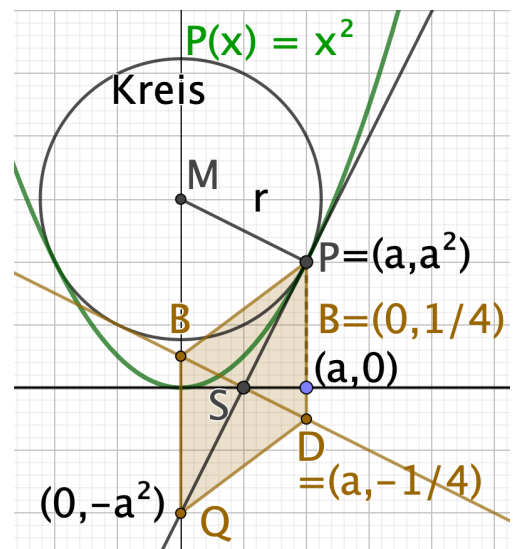
Solche einklemmenden Kreise sind *nicht eindeutig*: jeder kleinere Kreis mit derselben Tangente hat die Parabel zwischen sich und seiner Tangente. – *Verglichen mit dem, was wir vom Kreis her kennen, ist jetzt der Name **Tangente der Parabel** gerechtfertigt.*

2. Folgerungen aus der Kenntnis der Parabeltangente.

2.1. Geometrische Definition der Parabel

Die Tangente in $P = (a, a^2)$ schneidet die x-Achse in $S = (a/2, 0)$ und die y-Achse in $Q = (0, -a^2)$. Die vorhergehende Zeichnung wird durch die *Senkrechte zur Tangente* in S ergänzt. Da sie die Steigung $-1/2a$ hat, schneidet sie die y-Achse in $B = (0, 1/4)$ und die Vertikale $x = a$ in $D = (a, -1/4)$. Das Viereck $PBQD$ ist ein **Rhombus**, weil sich die Diagonalen halbieren und senkrecht auf einander stehen. Dieser aus der Tangente konstruierte Rhombus macht die geometrische Definition der Parabel kristallklar: es gilt $|PB| = |PD|$ oder *Die Punkte der Parabel sind gleich weit vom "Brennpunkt" B und der "Leitlinie" $y = -1/4$ entfernt.*

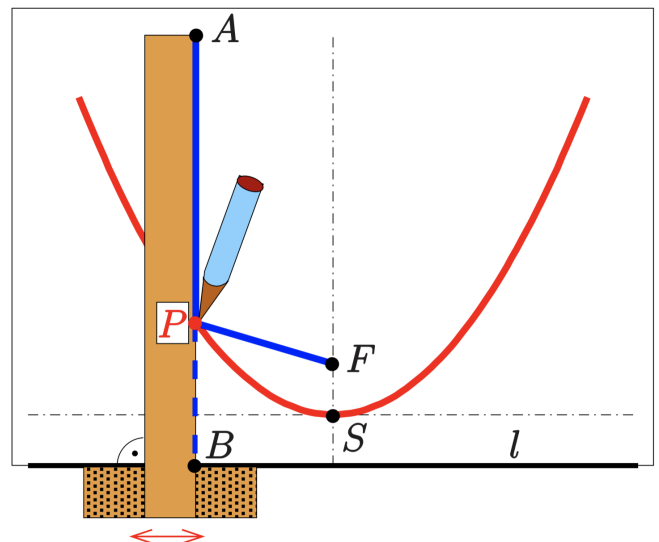
Die Kenntnis der Tangente der Parabel (= Graph von $P(x) = x^2$) liefert also deren geometrische Definition.



2.2. Fadenkonstruktion der Parabel

Fadenkonstruktionen von Ellipse und Hyperbel sind sehr alt. Nach *Wikipedia* ist die abgebildete sehr schöne Fadenkonstruktion der Parabel von Jakob Steiner (1796-1863) gefunden worden.

Wie oben folgt aus der Dreiecksungleichung, dass die Winkelhalbierende des Winkels $\angle(FPB)$ die so definierte Parabel auf einer Seite läßt, also diese Eigenschaft der Kreistangente hat. Es handelt sich also um eine Konstruktion, die nicht nur die Kurve sondern auch ihre Tangente liefert.



2.3. Reflektion an der Parabel

Die "Satellitenschüsseln" genannten Empfangsantennen haben die Form von Flächen, die durch Rotation von Parabeln um ihre Achse entstehen. An metallischen Flächen werden Lichtstrahlen und Radiowellen so reflektiert, dass an jeder Stelle der Einfallswinkel zur Normale gleich dem Ausfallswinkel ist. Setzt man diese physikalische Eigenschaft voraus, so folgt aus der Kenntnis der Parabeltangente von Neuem, was eben für die Parabel auf andere Weise bewiesen wurde:

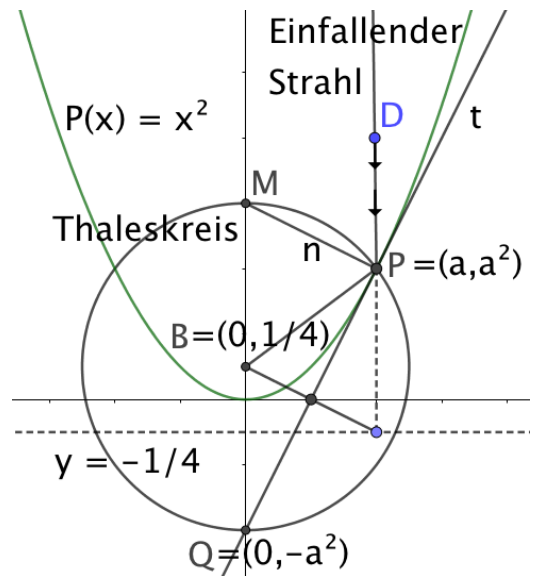
Parallel zur Parabelachse ankommende Strahlen werden so reflektiert:

$\angle DPM = \angle MPB$ (Reflektionsgesetz)

$\angle DPM = \angle PMB$ (Wechselwinkel an Parallelen) \Rightarrow
das Dreieck BPM ist gleichschenkelig, $|BM| = |BP|$.

Weil die Normale den Winkel $\angle DPB$ halbiert, halbiert die Tangente den Komplementwinkel. Daher ist auch das Dreieck BQP gleichschenkelig, $|BP| = |BQ|$ und B ist Mittelpunkt des Thaleskreises um MPQ , also $B = (0, 1/4)$ für alle ankommenden Strahlen.

Der in der geometrischen Definition schon *Brennpunkt* genannte Punkte B hat also seinen Namen daher, dass parallel einfallende Lichtstrahlen zu diesem Punkt hin reflektiert werden. Die Kenntnis der Parabeltangente erklärt also die Form von Empfangsantennen mit dem Empfänger im Brennpunkt und ebenso die Form der Spiegel astronomischer Spiegelteleskope.



2.4 Die quadratische Parabel und der Anfang der theoretischen Physik.

Newton hat sich auch für die Geometrie von Kurven interessiert, aber sein Hauptinteresse galt der Physik. Der berühmte Apfel fiel ebenfalls mit einem quadratischen Gesetz $s(t) = -\frac{g}{2} \cdot t^2$. Das Minuszeichen kommt daher, dass wir uns vertikale Koordinaten nach oben positiv vorstellen, der Apfel aber nach unten fällt. Eine physikalische Konstante ist notwendig, weil die Argumente Zeiten, die Werte aber Längen sind. Auf der Erdoberfläche ist diese Konstante $g = 9,81 \text{ m/s}^2$. Dieser Zahlenwert wird im Folgenden ignoriert. Die rechtsseitigen ($t > t_0$) und linksseitigen ($t < t_0$) Differenzenquotienten $(s(t) - s(t_0))/(t - t_0)$ heißen jetzt *Durchschnittsgeschwindigkeiten*. Die rechtsseitigen sind $-g \cdot \frac{t+t_0}{2} < -g \cdot t_0$, die linksseitigen sind $-g \cdot \frac{t+t_0}{2} > -g \cdot t_0$. Die einzige Zahl, die dazwischen passt, heißt seit Newton *Momentangeschwindigkeit* $v(t_0) = -g \cdot t_0$. (Da das für jedes t_0 gilt, schreiben wir $v(t) = -g \cdot t$.)

Das sieht zunächst harmlos aus, aber die wirkenden Kräfte verändern explizit die Momentangeschwindigkeiten und erst mittelbar die Durchschnittsgeschwindigkeiten. Das erlaubte Newton die Formulierung seiner heute *drittes Newtonsches Gesetz* genannten grundlegenden Gleichung der Mechanik:

$$(\text{wirkende}) \text{ Kraft} = \text{Masse} \cdot (\text{bewirkte}) \text{ Beschleunigung.}$$

Mit der in diesem Gesetz auftretenden “Beschleunigung” ist die *momentane* Änderung der *Momentangeschwindigkeit* gemeint. Ich kann mir nicht vorstellen, wie man dieses Gesetz mit den Durchschnittsgrößen formulieren könnte. Ich betrachte daher die Idee, den experimentell messbaren Durchschnittsgeschwindigkeiten $(s(t) - s(t_0))/(t - t_0), t \neq t_0$ die **nur** im beschreibenden Modell definierbaren Momentangeschwindigkeiten hinzuzufügen, als **phantastischen Fortschritt**.

Im Beispiel des fallenden Apfels sind alle Durchschnittsänderungen der Momentangeschwindigkeit $(v(t) - v(t_0))/(t - t_0)$ konstant und gleich $-g$. Deshalb hat man in diesem Beispiel keine Probleme, auch die Beschleunigung als konstant und gleich $-g$ anzusehen.

Soweit haben wir nur im physikalischen Kontext wiederholt, was im geometrischen Kontext erfolgreich war. Für die Physik ist das jedoch nicht genug, denn man möchte, im Beispiel, aus der konstanten Anziehungskraft auf das quadratische Bewegungsgesetz schließen. Dazu genügt zur Definition von Tangenten **nicht**, was ich bisher gesagt habe: Tangenten von Kurven sollen sich “so ähnlich” verhalten wie Kreistangenten. Und es genügt auch nicht, reelle Zahlen gegen einen Grenzwert “gehen” zu lassen. Man benötigt eine Definition, die erlaubt, aus der Existenz eines Grenzwertes auf das zurückzuschließen, was in seiner Definition benutzt wurde. Mit einer solchen Definition läßt sich dann der Monotoniesatz beweisen, also etwa diese Aussage: Wenn alle Tangenten einer Funktion nicht negative Steigung haben, dann ist die Funktion schwach wachsend. Daraus folgt: Wenn alle Tangenten die Steigung 0 haben, dann ist die Funktion weder wachsend noch fallend, also konstant. Daraus folgt dann, dass die Geschwindigkeit von Newton’s Apfel nicht durch eine andere Funktion beschrieben werden kann als durch $v(t) = 0 - g \cdot t$ und der Fall des Apfels aus der Höhe H beschrieben werden muss durch $s(t) = H - \frac{g}{2}t^2$. Das Fallgesetz wird also als Folge des Monotoniesatzes erhalten.

Probleme, die beim Aufbau der Differentialrechnung auftreten.

Eine Definition, mit der man aus Ableitungsvoraussetzungen Funktionseigenschaften folgern kann, muss mit Ungleichungen formuliert werden. Leider sind Ungleichungen sehr weitgehend aus dem Mathematikunterricht verschwunden, obwohl wir im täglichen Leben wirklich täglich die verschiedenartigsten konkreten oder abstrakten Dinge *der Größe nach* vergleichen. Um 1800, also zur Zeit von Gauß, war man noch weit von den heutigen Definitionen der Begriffe *konvergente Folge*, *Grenzwert*, *Tangente*, *Stetigkeit* entfernt. Während des 19. Jahrhunderts wurde intensiv daran gearbeitet. Bis 1900 hatten sich Definitionen dieser Begriffe durchgesetzt, die extrem verschieden von Formulierungen sind, die in der Umgangssprache vorkommen. Das hat zu einem sehr festgelegten Aufbau der Analysis geführt. Ich glaube, fast alle in Deutschland in Mathematik ausgebildeten Personen haben die Analysis so kennen gelernt: *Zuerst werden konvergente Folgen und ihre Grenzwerte erklärt, dann folgt der Begriff stetig (oft zuerst als folgenstetig) und am Schluss wird die Definition der Differenzierbarkeit mit Hilfe des Grenzwertbegriffs oder der Stetigkeit formuliert.* Das ist so fest etabliert, dass ich noch niemanden getroffen habe, der es für ernsthaft erwähnenswert hält, dass die Differentialrechnung bereits 200 glorreiche Jahre hinter sich hatte, als schließlich Grenzwerte und Stetigkeit ihre heutige Formulierung erhielten.

Deshalb möchte ich hervorheben, dass in der oben besprochenen Ausdehnung des Begriffs Tangente von Kreistangenten auf Parabeltangenten Grenzwerte oder Stetigkeit nicht vorkommen. Dafür gibt es zwei Gründe. Erstens, die Potenzfunktionen sind so einfach, dass sich bei dem Vergleich der rechtsseitigen und linksseitigen Sehnensteigungen der Faktor $(x - a)$ im Nenner aus dem Zähler $(f(x) - f(a))$ kürzen läßt, so dass man sehen kann, dass nur die Zahl $2a$ als Steigung der Tangente in Frage kommt (weil $2a$ die einzige Zahl ist, die zwischen den linksseitigen und rechtsseitigen Sehnensteigungen liegt). Zweitens, bei dem Nachweis, dass die vermutete Gerade mit guten Gründen Tangente genannt werden kann, kommt keine Division durch $(x - a)$ vor, weil das Argument sich darauf konzentriert, wie wenig der Graph der Parabel von der vermuteten Tangente abweicht - nämlich weniger als ein passender Kreis von seiner Tangente abweicht.

Diese Argumentation läßt sich auf alle Potenzfunktionen ausdehnen (und danach mit Differentiationsregeln auf alle rationalen Funktionen), aber ich kenne nur für die Parabel so schöne geometrische Eigenschaften, die unmittelbar zeigen, dass die definierten Tangenten nützlich sind. Ich finde, dass es zu wenig ist, das bloße Ausrechnen von Ableitungen als Hauptsache zu betrachten. Die Differentialrechnung hätte nicht ihr enormes Prestige erreicht, wenn sie nur Tangenten bestimmen und mit deren Hilfe Extremstellen finden und Wendetangenten berechnen könnte - ein Eindruck, den heutige Schulbücher erwecken.

Ehe ich zu Vorschlägen komme, muß ich auch die Schwierigkeiten ansprechen, die die endgültigen Definitionen vom Ende des 19. Jahrhunderts für die Schule verursachen. Für mich steht an erster Stelle die Tatsache, dass alle auf der Schule vorkommenden Funktionen schon Newton bekannt waren. Daher kann man keine Beispiele vorführen, für die die Definitionen so gemacht werden mußten, wie es im 19. Jahrhundert erarbeitet wurde. Ohne solche Beispiele schießt man aber mit Kanonen auf Spatzen. Das zweite Problem sind die vielen Misserfolge, die sich im Laufe der Zeit angesammelt haben. Von den Leuten, die Abitur gemacht haben, aber später nichts mehr mit Mathematik zu tun hatten, habe ich nur einen einzigen kennen gelernt, der bereit war, ein an den Analysisunterricht anknüpfendes Gespräch zu führen. Und niemand wird leugnen, dass auch noch viele Studienanfänger im Fach Mathematik erhebliche Schwierigkeiten mit der Stetigkeit haben.

Wie kann es nach der quadratischen Parabel weitergehen?

1. Übersicht über meinen Vorschlag.

Mein wichtigstes Anliegen ist, dass Schülerinnen und Schüler im Mathematikunterricht logische Argumentation kennen und verwenden lernen, Argumente also, die **nichts** benutzen dürfen, was nicht in Definitionen und Voraussetzungen zur Verfügung gestellt wird, und deren Folgerungen nur logische Schlussweisen benutzen.

Meine wichtigste Vereinfachung ist, statt der Grenzwerte von Sehnensteigungen die Approximation durch Tangenten ins Zentrum der Diskussion zu rücken - wie bei der quadratischen Parabel. Dadurch, dass man in der Definition auf die Division durch $(x - a)$ verzichtet, wird die Differentialrechnung der rationalen Funktionen unabhängig von Grenzwertdiskussionen. (Damit keine Missverständnisse entstehen: Grenzwerte sind unvermeidlich und von *unmittelbarem* Nutzen, überall wo irrationale Zahlen auftreten.) Das wird möglich durch die anschauliche Verabredung (Formulierung mit Ungleichungen weiter unten):

Eine Funktion f hat an der Stelle $x = a$ die Tangente $T_a(x) = m \cdot (x - a) + f(a)$, wenn es zwei Kreise mit dieser gemeinsamen Tangente gibt, so dass sich der Graph von f in einem kleinen Intervall um a zwischen den beiden Kreisen hindurchzwängt.

Da ich den Nutzen von Tangenten heutzutage nicht mit Anwendungen aus der Physik vorführen kann (bzw. darf), muss ich mich beschränken auf die Fehlerkontrolle numerischer Verfahren zur Differentiation, Integration und Berechnung nicht-rationaler Funktionen wie

$$\sqrt{x}, \exp, \sin.$$

Ein Bonus der Konzentration auf die Tangenten statt auf die Grenzwerte der Differenzenquotienten besteht darin, dass man folgende übersichtliche Formulierung für alle Differentiationsregeln hat:

Setzt man eine kompliziertere Funktion h aus einfacheren Funktionen f, g zusammen (z.B. $h = \alpha \cdot f + \beta \cdot g$, $h = f \cdot g$, $h = f \circ g$), so erhält man die Tangenten von h , indem man die Tangenten von f und g ebenso zusammensetzt wie diese Funktionen. (Details weiter unten.)

Aus meiner Sicht sind die Differentiationsregeln nur Bequemlichkeitsregeln, weil alles Schwierige schon in den Voraussetzungen steht. Ich betrachte den Monotoniesatz und seine Folgerungen als die einzige wesentliche Satzgruppe aus den Anfängen der Differentialrechnung. Deshalb führe ich einen Beweis vor, der auf der Schule behandelt werden könnte.

2. Tangentendefinition wie bei der Parabel.

Es ist einfacher, Ungleichungen mit Parabeln zu handhaben als Ungleichungen mit Kreisen. Da wir für die Parabel schon gezeigt haben, dass sich zu jedem Punkt ein Kreis finden lässt, so dass der Graph der Parabel zwischen Kreis und Kreistangente verläuft, können wir in der folgenden Formulierung Parabeln statt Kreisen verwenden.

Tangentendefinition:

Die Funktion f hat an der Stelle a die **Tangente** $T_a(x) = m \cdot (x - a) + f(a)$, wenn es ein Intervall $[a - \delta, a + \delta]$ und eine Konstante B gibt, so dass folgende *Approximationsungleichung* gilt:

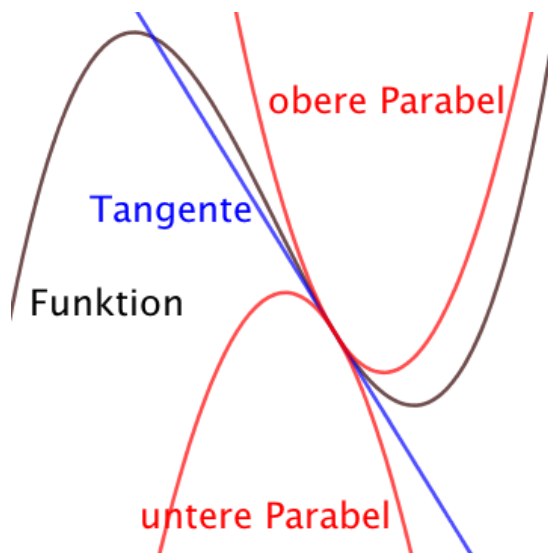
$$x \in [a - \delta, a + \delta] \Rightarrow |f(x) - T_a(x)| \leq B \cdot |x - a|^2.$$

Der Graph von f verläuft also im Intervall $[a - \delta, a + \delta]$ zwischen einer oberen und einer unteren Parabel mit der gemeinsamen Tangente T_a . Wir zeigen im nächsten Abschnitt, dass die Tangente von f bei a durch diese Definition eindeutig bestimmt ist. Das wird durch eine entsprechende Bezeichnung betont:

$$m = f'(a).$$

Bei Funktionen, die diese Eigenschaft haben, sind die Tangenten also besonders leicht zu finden und sie sind **bessere** Approximationen von f als mit der akzeptierten Tangentendefinition des 19. Jahrhunderts verlangt wird.

Funktionen, die nach der endgültigen Definition *zweimal* differenzierbar sind und für die im Intervall $[a - \delta, a + \delta]$ gilt $|f''| \leq 2B$, werden - als Folgerung aus dem Monotoniesatz - von ihrer Tangente bei a so gut approximiert, wie das in der vorgeschlagenen (schärferen) Definition verlangt wird.



3. Erste Konsequenz der Definition: Eindeutigkeit der Tangente.

In der Tangentendefinition kommt nicht explizit vor, dass die Tangente bei a *eindeutig* bestimmt ist. Wegen der Bilder, die die Schülerinnen und Schüler gesehen haben, werden sie wohl kaum Zweifel an der Eindeutigkeit haben. Obwohl man daher über die Wichtigkeit eines Beweises geteilter Meinung sein kann, füge ich einen *Beweis der Eindeutigkeit* an, denn Vertrautheit mit der Definition kann nur mit Beweisen erreicht werden, da die Anwendungen sich auf das Einüben der Differentiationsregeln konzentrieren.

Weil der Umgang mit Ungleichungen auf der Schule kaum vorkommt, schicke ich einige Bemerkungen voraus. (Ausführlicher in: Hilfestellung zum Umgang mit Ungleichungen.)

Es seien $a, b \in \mathbb{R}$ und $p, q \in \mathbb{R}_+$. Dann bedeuten folgende Ungleichungen dasselbe:

$$\begin{aligned} |a| \leq p &\Leftrightarrow -p \leq a \leq p \Leftrightarrow -p \leq -a \leq p \\ |b| \leq q &\Leftrightarrow -q \leq b \leq q \Leftrightarrow -q \leq -b \leq q \end{aligned}$$

Ungeübte können die Ungleichungsketten leichter addieren als mit den Beträgen argumentieren (zum Lesen sind die Beträge übersichtlicher), also

$$|a| \leq p \wedge |b| \leq q \Rightarrow -p - q \leq a \pm b \leq +p + q \Leftrightarrow |a \pm b| \leq p + q.$$

Beweis: Daraus wird nun gefolgert, dass *zwei* Tangenten an den Graphen einer Funktion f an der Stelle $(a, f(a))$ *übereinstimmen*. Wir setzen also für alle $x \in [a - \delta, a + \delta]$ die zwei Approximationsungleichungen für (möglicher Weise) zwei Tangenten voraus:

$$\begin{aligned} |(f(x) - f(a)) - m_1 \cdot (x - a)| &\leq B_1 \cdot (x - a)^2 \\ |(f(x) - f(a)) - m_2 \cdot (x - a)| &\leq B_2 \cdot (x - a)^2 \end{aligned}$$

und wir schließen daraus für die Differenz der Terme in den Beträgen

$$|m_1 \cdot (x - a) - m_2 \cdot (x - a)| \leq (B_1 + B_2) \cdot (x - a)^2.$$

Diese Ungleichung kann für alle $x \neq a$, $x \in [a - \delta, a + \delta]$ durch $|x - a|$ dividiert werden:

$$0 < |x - a| \leq \delta \Rightarrow |m_1 - m_2| \leq (B_1 + B_2) \cdot |x - a|.$$

Die Formulierung des folgenden Schlußschrittes ist Geschmacksache. Ich betone gern, wie oft man Archimedes' Argument einsetzen kann und wähle für alle $n \in \mathbb{N}$, $n > 1/\delta$ die Argumente

$x_n := a + 1/n$, $0 < |x_n - a| < \delta < 1/n$. Das gibt:

$$0 \leq |m_1 - m_2| \leq (B_1 + B_2)/n, \text{ also mit Archimedes: } m_1 = m_2.$$

Man kann das indirekte Argument aus dem Beweis für Archimedes' Schlussweise auch explizit wiederholen:

Angenommen $m_1 \neq m_2$, dann wähle $x := a + \min(\delta, \frac{|m_1 - m_2|}{2(B_1 + B_2)}) \neq a$, um folgenden Widerspruch zu erhalten: $0 < |m_1 - m_2| \leq |m_1 - m_2|/2$. Daher ist $m_1 \neq m_2$ unmöglich.

4. Bausteine für Differentiationsregeln:

Direkt aus der Definition berechnete Beispiele.

Zweifellos werden viel mehr Ableitungen mit den Differentiationsregeln berechnet als direkt aus der Definition bestimmt. Damit dieser Prozess beginnen kann, müssen einige erste Ableitungen direkt aus der Definition hergeleitet werden.

Die höheren Potenzen werden mit der Produktregel differenziert, aber es hilft dem Umgang mit der Definition, nach der quadratischen Parabel zwei weitere Beispiele direkt zu differenzieren. Dazu verallgemeinern wir die dritte binomische Formel ein wenig:

$$\begin{aligned} (x - a) \cdot (x + a) &= x^2 - a^2, \\ (x - a) \cdot (x^2 + x \cdot a + a^2) &= x^3 - a^3, \\ (x - a) \cdot (x^3 + x^2 \cdot a + x \cdot a^2 + a^3) &= x^4 - a^4. \end{aligned}$$

Im ersten Schritt geht es nur darum, durch Betrachten der rechtsseitigen und linksseitigen Sehnensteigungen eine *gute Vermutung* zu bekommen, wie groß die Tangentensteigung sein könnte. Daher kann ich, zur Vermeidung von Fallunterscheidungen, die Voraussetzung

$$0 < x, a \text{ machen.}$$

$$\begin{aligned} f(x) := x^3, \quad \frac{f(x) - f(a)}{x - a} &= x^2 + x \cdot a + a^2 \begin{cases} \geq 3a^2 & \text{für } x > a \\ \leq 3a^2 & \text{für } x < a \end{cases} & f'(a) = 3a^2 \\ f(x) := x^4, \quad \frac{f(x) - f(a)}{x - a} &= x^3 + x^2 \cdot a + x \cdot a^2 + a^3 \begin{cases} \geq 4a^3 & \text{für } x > a \\ \leq 4a^3 & \text{für } x < a \end{cases} & f'(a) = 4a^3 \end{aligned}$$

Im zweiten Schritt muss die *Approximationsungleichung* bewiesen werden, ohne Einschränkung an die Vorzeichen von x und a . Die Potenzfunktionen sind so einfach, dass man $\delta = 1$ nehmen könnte. Ich benutze weiter δ , damit man den Einfluß der Länge des Intervalls sieht. Mit der vermuteten Steigung wird $T_a(x) := f'(a) \cdot (x - a) + f(a)$ definiert und die Differenz $f(x) - T_a(x)$ abgeschätzt. So wird die Vermutung bestätigt.

$$\begin{aligned} |x^3 - (3a^2 \cdot (x - a) + a^3)| &= |(x^2 + x \cdot a + a^2) - 3a^2| \cdot |x - a| = |x + 2a| \cdot |x - a|^2 \\ &\leq (3|a| + \delta) \cdot (x - a)^2 & \text{also } B_{x^3} := 3|a| + \delta \\ |x^4 - (4a^3 \cdot (x - a) + a^4)| &= |(x^3 + x^2 \cdot a + x \cdot a^2 + a^3) - 4a^3| \cdot |x - a| \\ &= |(x^2 + x \cdot a + a^2) + (x + a) \cdot a + a^2| \cdot |x - a|^2 \\ &\leq (6a^2 + 4|a| \cdot \delta + \delta^2) \cdot (x - a)^2 & \text{also } B_{x^4} := 6a^2 + 4|a| \cdot \delta + \delta^2 \end{aligned}$$

Wie gesagt, wegen der Produktregel braucht man diese Rechnungen nicht zu machen. Der direkte Zugriff auf die Tangentendefinition ist nur notwendig, wenn man etwas Neues macht, sonst genügen die Differentiationsregeln. Mit der Definition sollte aber geübt werden.

Inverse Funktion, Wurzelfunktion

Zwei weitere Funktionen, die ich direkt mit der Definition, also ohne Anwendung der Regeln, differenzieren würde, sind $f(x) = 1/x$ und $f(x) = \sqrt{x}$. Damit das Intervall $[a - \delta, a + \delta]$ nicht aus dem Definitionsbereich herausragt, wird $\delta \leq |a|/2$ gewählt. Aufgeschrieben ist nur der Fall $a > 0$ und mit der Voraussetzung $x \in [a - \delta, a + \delta]$. Die *vermuteten* Tangentensteigungen folgen wieder aus der Betrachtung der rechts- und linksseitigen Sehnensteigungen.

$$f(x) := 1/x, \quad \frac{f(x) - f(a)}{x - a} = \frac{-1}{a \cdot x} \quad \begin{cases} \geq -1/a^2 & \text{für } x > a \\ \leq -1/a^2 & \text{für } x < a \end{cases} \quad f'(a) = -1/a^2$$

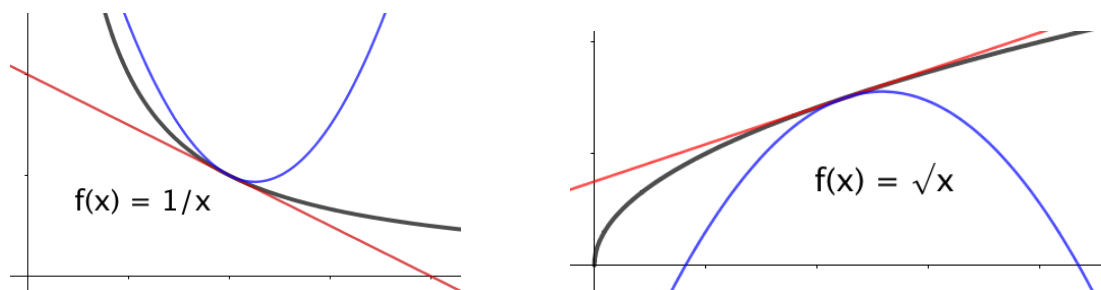
$$f(x) := \sqrt{x}, \quad \frac{f(x) - f(a)}{x - a} = \frac{1}{\sqrt{x} + \sqrt{a}} \quad \begin{cases} \leq \frac{1}{2\sqrt{a}} & \text{für } x > a \\ \geq \frac{1}{2\sqrt{a}} & \text{für } x < a \end{cases} \quad f'(a) = \frac{1}{2\sqrt{a}}$$

Zu *beweisen* ist wieder die *Approximationsungleichung* für die Differenz $f(x) - T_a(x)$ unter der Voraussetzung $x \in [a - \delta, a + \delta]$. Das Vorzeichen dieser Differenz zeigt der Term nach dem Gleichheitszeichen, daher bleibt der Graph in $[a - \delta, a + \delta]$ auf einer Seite der Tangente.

$$0 \leq \frac{1}{x} - \left(\frac{-1}{a^2} \cdot (x - a) + \frac{1}{a} \right) = \frac{(x - a)^2}{a^2 x} \leq \frac{1}{a^2(a - \delta)} \cdot (x - a)^2$$

$$0 \geq \sqrt{x} - \left(\frac{x - a}{2\sqrt{a}} + \sqrt{a} \right) = \frac{(\sqrt{a} - \sqrt{x}) \cdot (x - a)}{2\sqrt{a} \cdot (\sqrt{x} + \sqrt{a})} \geq \frac{-1}{8\sqrt{a}(a - \delta)} \cdot (x - a)^2$$

Wir lesen ab $B_{1/x} := 1/(a^2(a - \delta))$ und $B_{\sqrt{x}} := 1/(8\sqrt{a}(a - \delta))$. Meiner Meinung nach genügt das, um zu verstehen, wie die Definition auf Beispiele angewendet wird. Natürlich unterstützen Bilder das Verständnis:



Bemerkung. Die berechneten Tangenten von $f(x) = x^3$ schneiden den Graphen ein weiteres Mal bei $x = -2a$. Die vom Kreisbeispiel suggerierte Vorstellung, Tangenten ließen ihre Kurven auf einer Seite, ist also falsch. Auch eine Beschränkung auf die in der Definition vorkommenden kleinen Intervalle $[a - \delta, a + \delta]$ hilft nicht, denn die x-Achse sieht nicht nur wie die Tangente $T_0(x) = 0 \cdot (x - a) + 0^2$ des Graphen von $f(x) = x^3$ aus, sie ist sogar nach unserer schärferen Definition Tangente, denn $|x^3 - T_0(x)| = |(x - 0)^3| \leq \delta \cdot (x - 0)^2$. Diese Tangente schneidet ihre Kurve im *Berührungspunkt*. Man sieht daran, dass die allgemeine Charakterisierung von Tangenten durch *„lassen die Kurve auf einer Seite“* ebenso falsch ist wie die allgemeine Behauptung, eine Funktion wechsele an einer Nullstelle ihr Vorzeichen, auch wenn die Gegenbeispiele dafür (wie $f(x) = x^2$) klar in der Minderheit sind.

Transzendente Funktionen

Die transzendenten Funktionen werden meistens an einer späteren Stelle in einem Analysislehrgang behandelt. Für Leser, die die Analysis schon kennen, möchte ich schon hier hinzufügen, dass auch die Ableitungen dieser Funktionen so erhalten werden können, dass man mit Formeln, die ohnehin vorkommen, zeigt, dass die Graphen dieser Funktionen *zwischen* dem Graphen einer rationalen Funktion und dessen Tangente verlaufen.

Selbstverständlich benötigt man Grenzwerte, um die transzendenten Funktionen zu definieren. Ich finde auch nicht, dass Grenzwerte so schwierig zu erklären sind, dass deren Eliminierung aus der Schulmathematik gerechtfertigt ist. Aber, da die überall benutzten Taschenrechner meistens nur Approximationen ausgeben und auch viele Eingaben nur Approximationen sind ($1/3, \sqrt{2}, \pi, \exp(1), \dots$), ist es vordringlich, **Approximationen** zu erklären. Deshalb dehne ich die begonnene Erklärung, Ableitungen mit Approximationen statt mit Grenzwerten zu bestimmen, hier auf die transzendenten Funktionen aus.

Für die **Sinusfunktion** scheinen alle Autoren von der Ungleichung

$$0 \leq x < \pi/2 \Rightarrow \sin x \leq x \leq \tan x = \sin x / \cos x$$

auszugehen, auch mein Text A6, S. 5. Zusammen mit $0 < a < 1 \Rightarrow a < \sqrt{a} < 1$ verwende ich das hier so:

$$\begin{aligned} 0 \leq x < \pi/2 \Rightarrow x \geq \sin x \geq x \cdot \cos x &= x \cdot \sqrt{1 - \sin^2(x)} \geq x \cdot (1 - \sin^2(x)) \\ &\geq x - x^3 \end{aligned}$$

Hier ist $tg(x) := x$ die *schon bekannte* Tangente bei $x = 0$ der Unterfunktion $f_u(x) := x - x^3$ und der Graph von Sinus verläuft *zwischen* dieser Unterfunktion und ihrer Tangente (wegen $\sin(-x) = -\sin x$ im Intervall $-\pi/2 < x < \pi/2$).

Deshalb ist $tg(x)$ auch die Tangente von $\sin(x)$ bei $x = 0$.

Die Ableitung an anderen Stellen wird auf unterschiedliche Weise mit einem Additionstheorem bestimmt.

Die Definition der **natürlichen Exponentialfunktion** wird meistens im Anschluss an die Ableitung der Exponentialfunktionen $f_a(x) := a^x$ definiert. Sie kann auch als Grenzwert von unteren und oberen Approximationen definiert werden:

$$-n < x < n \Rightarrow \left(1 + \frac{x}{n}\right)^n \leq \exp(x) \leq \left(1 - \frac{x}{n}\right)^{-n},$$

insbesondere: $-1 < x < 1 \Rightarrow (1 + x) \leq \exp(x) \leq (1 - x)^{-1}$.

Auch hier ist $tg(x) := 1 + x$ Tangente der Oberfunktion $f_o(x) = (1 - x)^{-1}$. Und der Graph von \exp verläuft *zwischen* dieser Oberfunktion und ihrer Tangente bei $x = 0$. Man braucht zwar Grenzwerte um $\exp(x)$ zu definieren, aber man braucht keine Grenzwerte für die Ableitung der Grenzfunktion $\exp(x)$, weil, wie in allen anderen Beispielen, der Graph dieser Grenzfunktion zwischen einer Oberfunktion und deren Tangente verläuft.

Deshalb ist $tg(x)$ auch die Tangente von $\exp(x)$ bei $x = 0$.

Für diese transzendenten Funktionen folgt die Differenzierbarkeit an Stellen $x \neq 0$ aus den Additionstheoremen, aus $e^{x+y} = e^x \cdot e^y$, $\sin(x+y) = \sin(x)\cos(y) + \cos(x)\sin(y), \dots$

5. Differentiationsregeln

Differentiationsregeln haben den Sinn, solche Funktionen ohne Rückgriff auf die Tangenten-
definition zu differenzieren, die aus "einfacheren" Funktionen – deren Ableitungen bekannt
sind – zusammengesetzt sind. In Differentiationsregeln kommen also mehrere Funktionen
vor und wir müssen unterscheiden können, welche Tangentengleichung zu welcher Funktion
gehört. Deshalb verabreden wir eine etwas kompliziertere Bezeichnung:

Die Tangente des Graphen von f an der Stelle a ist T_a^f ,
oder mit Argumenten: $T_a^f(x) := f'(a) \cdot (x - a) + f(a)$
und in $[a - \delta^f, a + \delta^f]$ gilt $|f(x) - T_a^f(x)| \leq B^f \cdot (x - a)^2$.

Schließlich wird die Formulierung der Beweise übersichtlicher, wenn wir dem Unterschied
zwischen Funktion und Tangente ebenfalls einen (etwas umfangreichen) Namen geben:

$U_a^f(x - a) := f(x) - T_a^f(x)$, $|U_a^f(x - a)| \leq B^f \cdot (x - a)^2$ bzw.
mit $h := (x - a)$ auch $U_a^f(h) = f(a + h) - (f'(a) \cdot h + f(a))$, $|U_a^f(h)| \leq B^f \cdot h^2$.

In den Voraussetzungen der Differentiationsregeln steht, dass die Approximationsungleichun-
gen für die zum Zusammensetzen benutzten Funktionen erfüllt sind. Die Beweise bestehen
nur aus der Kontrolle der Approximationsfehler mit der Dreiecksungleichung. Das gilt ebenso,
wenn man später lernt, mit der endgültigen Differenzierbarkeitsdefinition zu arbeiten. Die
folgenden Beweise zitieren also **nicht** frühere Ergebnisse sondern enthalten **alle** wesentli-
chen Details. Für alle Regeln gilt die übersichtliche Zusammenfassung: *Um die Tangenten
der zusammengesetzten Funktion h zu erhalten, muss man mit den Tangenten der Baustein-
funktionen f, g dasselbe machen wie mit f und g .*

Voraussetzungen: Die Funktionen f, g seien an der Stelle a differenzierbar. Das heißt, es
gibt $\delta^f, B^f, \delta^g, B^g$ so dass

$$\begin{aligned} x \in [a - \delta^f, a + \delta^f] &\Rightarrow |f(x) - T_a^f(x)| = |U_a^f(x - a)| \leq B^f \cdot (x - a)^2, \\ x \in [a - \delta^g, a + \delta^g] &\Rightarrow |g(x) - T_a^g(x)| = |U_a^g(x - a)| \leq B^g \cdot (x - a)^2. \end{aligned}$$

Beide Ungleichungen gelten mit $\delta := \min(\delta^f, \delta^g) > 0$ im Intervall $[a - \delta, a + \delta]$.

Behauptung 1. Differenzieren von Linearkombinationen

Mit reellen Konstanten $\alpha, \beta \in \mathbb{R}$ definieren wir die **Linearkombination** $h := \alpha \cdot f + \beta \cdot g$.
Dann ist h differenzierbar bei $x = a$ mit

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

Beweis 1. Die beiden Gleichungen,

$$f(x) = T_a^f(x) + U_a^f(x - a) \quad \text{und} \quad g(x) = T_a^g(x) + U_a^g(x - a),$$

die in $[a - \delta, a + \delta]$ gelten, werden mit α bzw. β multipliziert und addiert. Das liefert h
als die *Linearkombination der Tangenten* plus die *Linearkombination der Fehler*:

$$h(x) = \alpha \cdot T_a^f(x) + \beta \cdot T_a^g(x) + \alpha \cdot U_a^f(x - a) + \beta \cdot U_a^g(x - a).$$

Für die Linearkombination der Fehler liefert die Dreiecksungleichung ($|u + v| \leq |u| + |v|$):

$$|\alpha \cdot U_a^f(x - a) + \beta \cdot U_a^g(x - a)| \leq (|\alpha| \cdot B^f + |\beta| \cdot B^g) \cdot (x - a)^2.$$

Mit der Definition $B^h := (|\alpha| \cdot B^f + |\beta| \cdot B^g)$ gilt dann in $[a - \delta, a + \delta]$

$$|h(x) - ((\alpha \cdot f'(a) + \beta \cdot g'(a)) \cdot (x - a) + h(a))| \leq B^h \cdot (x - a)^2.$$

Das beweist die Differenzierbarkeit von h und $h'(a) = \alpha \cdot f'(a) + \beta \cdot g'(a)$.

Die Differentiationsregel für Linearkombinationen gilt also wegen der Dreiecksungleichung
(angewandt auf die Fehler).

Behauptung 2.**Produktregel**

Das Produkt $h := f \cdot g$ ist bei $x = a$ differenzierbar und

$$h'(a) = f'(a) \cdot g(a) + f(a) \cdot g'(a).$$

Beweis 2. Die beiden Tangenten-Approximationen

$$f(x) = T_a^f(x) + U_a^f(x - a) \quad \text{und} \quad g(x) = T_a^g(x) + U_a^g(x - a)$$

die in $[a - \delta, a + \delta]$ gelten, werden multipliziert. Das liefert $h(x)$ als *Produkt der Tangenten* (als Hauptterm) plus drei Fehlerterme.

Der Beweis besteht – ohne Zitate anderer Ergebnisse – allein im Zusammenfügen aller schon in den Voraussetzungen stehenden Fehlerterme mit Hilfe der Dreiecksungleichung!

Multipliziert man zwei Zerlegungen in Hauptteil plus “unwesentlichen” Fehleranteil, also

$$F_1 = H_1 + U_1, \quad F_2 = H_2 + U_2,$$

so ergibt sich mit der Dreiecksungleichung $|u + v + w| \leq |u| + |v| + |w|$:

$$\begin{aligned} |F_1 \cdot F_2 - H_1 \cdot H_2| &= |H_1 \cdot U_2 + H_2 \cdot U_1 + U_1 \cdot U_2| \\ &\leq |H_1 \cdot U_2| + |H_2 \cdot U_1| + |U_1 \cdot U_2|. \end{aligned}$$

Die Fehlerterme sind *schon in der Voraussetzung* so abgeschätzt: $\leq B \cdot (x - a)^2$.

Wenn man daher Schranken für H_1 , H_2 finden kann, dann lassen sich die unwesentlichen Terme zusammenfassen mit einer Abschätzung $\leq B_U \cdot (x - a)^2$.

H_1 , H_2 sind die linearen Tangentenfunktionen, für sie finden wir leicht Schranken:

$$|H_1| = |T_a^f(x)| \leq |f'(a)| \cdot \delta + |f(a)| =: M^f, \quad |H_2| = |T_a^g(x)| \leq |g'(a)| \cdot \delta + |g(a)| =: M^g,$$

mit denen wir die Fehlerterme so zusammenfassen können:

$$\begin{aligned} |H_1 \cdot U_2| + |H_2 \cdot U_1| + |U_1 \cdot U_2| &\leq (B^f \cdot M^g + B^g \cdot M^f + B^f \cdot B^g \cdot \delta^2) \cdot (x - a)^2 \\ &=: B_U \cdot (x - a)^2 \end{aligned}$$

$H_1 \cdot H_2$ ist das Produkt der beiden linearen Tangentenfunktionen $T_a^f(x) \cdot T_a^g(x)$, damit

$$\begin{aligned} H_1 \cdot H_2 &= (f'(a) \cdot (x - a) + f(a)) \cdot (g'(a) \cdot (x - a) + g(a)) \\ &= \underline{(f'(a)g(a) + g'(a)f(a)) \cdot (x - a) + h(a)} + f'(a)g'(a)(x - a)^2. \end{aligned}$$

Der nicht unterstrichene quadratische Term gehört noch zu den Fehlern, er wird diesen hinzugefügt:

$$|h(x) - ((f'(a)g(a) + g'(a)f(a)) \cdot (x - a) + h(a))| \leq (B_U + |f'(a)g'(a)|) \cdot (x - a)^2.$$

Diese quadratische Kontrolle der Differenz zwischen h und einer linearen Funktion zeigt nach Definition, dass diese lineare Funktion die Tangente von $h = f \cdot g$ bei a ist:

$$\begin{aligned} T_a^h(x) &= h'(a) \cdot (x - a) + h(a), \\ \text{mit } h'(a) &= f'(a)g(a) + g'(a)f(a). \end{aligned}$$

Das waren die beiden einfacheren Differentiationsregeln. Mit der **Kettenregel** ist es komplizierter, da ich in Schulbüchern der letzten 50 Jahre nur Beweise gefunden habe, die entweder den Zusatz enthalten, der Beweis gelte nur für streng monotone Funktionen, oder in denen ohne Zögern mit $0/0$ erweitert wird, nämlich in der Form

$$\frac{u(v(x)) - u(v(a))}{v(x) - v(a)} \cdot \frac{v(x) - v(a)}{x - a}.$$

Deshalb möchte ich zuerst erklären, wie leicht man sich zweimal differenzierbare Funktionen verschaffen kann, deren Graphen viele horizontale Stücke enthalten, die also als innere Funktionen bei der zitierten Erweiterung *nicht vorkommen dürfen*.

Die Funktion $f(x) := \sin(x) + x$ ist streng monoton wachsend, hat aber an allen ungeraden Vielfachen von π *horizontale* Wendetangenten. Man kann an jeder dieser Wendestellen den Graphen durchtrennen, etwas auseinander schieben und mit einer horizontalen Strecke wieder verbinden. Diese geänderte Kurve ist Graph einer zweimal differenzierbaren Funktion mit vielen Konstanzintervallen.

Im Beweis der Kettenregel wird eine Eigenschaft der inneren Funktion benutzt, die ich als Hilfssatz vorab formulieren möchte.

Hilfssatz:

Die Ableitung ist beinahe eine Dehnungsschranke.

Genauer, aus der Approximationsungleichung der Tangente

$$x \in [a - \delta, a + \delta] \Rightarrow |f(x) - T_a^f(x)| \leq B \cdot |x - a|^2$$

folgt mit der Dreiecksungleichung, dass f Abstände von der Stelle a höchstens um den Faktor $L := |f'(a)| + B \cdot \delta$ vergrößert, also

Behauptung: $x \in [a - \delta, a + \delta] \Rightarrow |f(x) - f(a)| \leq L \cdot |x - a|.$

Beweis. Wir benutzen folgende Variante der Dreiecksungleichung:

$$|u| = |(u - v) + v| \leq |u - v| + |v|, \text{ also } |u - v| \geq |u| - |v|.$$

Dann ergibt sich aus der Approximationsungleichung der Tangente für $|x - a| \leq \delta$

$$\begin{aligned} B \cdot |x - a|^2 &\geq |f(x) - T_a^f(x)| = |(f(x) - f(a)) - f'(a) \cdot (x - a)| \geq \\ &\geq |f(x) - f(a)| - |f'(a)| \cdot |x - a|, \end{aligned}$$

und damit

$$|f(x) - f(a)| \leq (|f'(a)| + B \cdot \delta) \cdot |x - a| = L \cdot |x - a|.$$

Bei der Kettenregel haben innere und äußere Funktion unterschiedliche Definitionsbereiche. Um Verwechslungen möglichst auszuschließen, bezeichne ich die äußere Funktion und ihre Argumente mit großen Buchstaben.

Es geht jetzt um die Funktion $h := F \circ f$ oder $h(x) = F(f(x))$.

Voraussetzungen der Kettenregel: f ist differenzierbar bei a und F ist differenzierbar bei $A = f(a)$, also

$$\begin{aligned} x \in [a - \delta^f, a + \delta^f] &\Rightarrow |f(x) - T_a^f(x)| \leq B^f \cdot (x - a)^2 \\ X \in [A - \delta^F, A + \delta^F] &\Rightarrow |F(X) - T_A^F(X)| \leq B^F \cdot (X - A)^2 \end{aligned}$$

Damit $f(x)$ in das Intervall $[A - \delta^F, A + \delta^F]$ fällt, muss mit Hilfe des Hilfssatzes weiter vorausgesetzt werden, dass δ^f klein genug ist, also notfalls verkleinert werden muss:

$$\delta^f \leq \delta^F / L \text{ und } x \in [a - \delta^f, a + \delta^f] \Rightarrow f(x) \in [A - \delta^F, A + \delta^F].$$

Diese Kontrolle der inneren Funktion ist ein wichtiger Teil des Beweises.

Behauptung: Die Funktion $h = F \circ f$ ist bei a differenzierbar mit

$$h'(a) = F'(f(a)) \cdot f'(a).$$

Für die Tangenten gilt $T_a^h(x) = T_A^F \circ T_a^f(x)$. Mit anderen Worten:

Die Tangente der Komposition $F \circ f$ ist die Komposition der Tangenten.

Beweis: Wie immer geht es darum, die Differenz zwischen $F \circ f$ und der erwarteten Tangente $T_a^h(x) = T_A^F \circ T_a^f(x)$ zu beherrschen. Das legt folgenden Anfang nahe:

$$\begin{aligned} |F \circ f(x) - T_A^F \circ T_a^f(x)| &= |F \circ f(x) - T_A^F \circ f(x) + T_A^F \circ f(x) - T_A^F \circ T_a^f(x)| \leq \\ &\leq |F \circ f(x) - T_A^F \circ f(x)| + |T_A^F \circ f(x) - T_A^F \circ T_a^f(x)| \end{aligned}$$

und diese beiden Summanden müssen weiter behandelt werden.

Nach Voraussetzung über F und Einsatz des Hilfssatzes haben wir für den ersten Summanden (beachte $A = f(a)$):

$$|F \circ f(x) - T_A^F \circ f(x)| \leq B^F \cdot (f(x) - f(a))^2 \leq B^F \cdot L \cdot (x - a)^2.$$

Nach Voraussetzung über f haben wir für den zweiten Summanden (beachte, dass T_A^F die Steigung $F'(A)$ hat, also $T_A^F(Y) - T_A^F(Z) = F'(A) \cdot (Y - Z)$):

$$\begin{aligned} |T_A^F(f(x)) - T_A^F(T_a^f(x))| &= \left| F'(A) \cdot (f(x) - (f'(a)(x - a) + f(a))) \right| \\ &\leq |F'(A)| \cdot B^f \cdot (x - a)^2. \end{aligned}$$

Beides zusammen beweist die Behauptung:

$$|F \circ f(x) - T_A^F \circ T_a^f(x)| \leq (B^F \cdot L + |F'(A)| \cdot B^f) \cdot (x - a)^2.$$

Ich wiederhole: Die Tangente der Komposition $F \circ f$ bei a ist die Komposition der Tangente von F bei $A = f(a)$ und der Tangente von f bei a . Das ist ein sehr schönes Ergebnis.

Außerdem: *Der Beweis beruht wirklich nur auf der Dreiecksungleichung.*

Damit ist gesagt, was zunächst zu Definition und Berechnung von Ableitungen gesagt werden muss. Später, wenn man (mit der Vollständigkeit!) Umkehrfunktionen, \exp und \sin behandelt, kommen weitere Differentiationsregeln hinzu. Aber wir müssen noch auf das für Physik und numerische Mathematik wichtige Problem zurückkommen:

*Wie schließt man zurück von Informationen über
die Ableitung auf Eigenschaften der Funktion?*

In der Hochschulmathematik gehen solche Resultate von dem *Mittelwertsatz der Differentialrechnung* aus, der aber einen so aufwendigen Beweis hat, dass er nicht zu meiner Vorstellung von “Einführung in die Differentialrechnung” passt. (Ich erinnere daran, dass der Begriff “stetig” 200 Jahre jünger als die Differentialrechnung ist. Euler und Gauß kannten den Begriff “stetig” nicht.) Eine unmittelbare Folge des Mittelwertsatzes ist der Monotoniesatz:

Eine Funktion mit positiver Ableitung ist wachsend.

Ich habe in dem Text A5: *Loblied auf den Monotoniesatz* auf Seite 7 einen Beweis aufgeschrieben, der direkt von der ϵ - δ -Definition der Differenzierbarkeit ausgeht, natürlich die Vollständigkeit benötigt und wie in allen ähnlichen Fällen indirekt ist, aber *keine weiteren* Begriffe benutzt. Damit ist die Einführung in die Differentialrechnung beendet, die weitere Entwicklung wird von Anwendungen des Monotoniesatzes dominiert, die Arbeit mit Analysis beginnt.

Für Interessenten füge ich noch einen erheblich elementareren Beweis des Monotoniesatzes an, der aber das Reizwort “gleichmäßig” benutzt.

Da wir bis hier nur mit Funktionen zu tun haben, die schon mit der vorgeschlagenen schärferen Differenzierbarkeitsdefinition als differenzierbar erkannt werden können, kann man den Beweis aus Text A5 weiter vereinfachen. Ich gehe noch einen Schritt weiter. In vielen weit über die Schule hinausreichenden Situationen kann man die in der Definition auftretenden Konstanten δ^f, B^f in dem gerade betrachteten Intervall *unabhängig* von der Stelle a wählen (man sagt auch: *gleichmäßig wählen*). Dann kann man im Beweis sogar auf die Vollständigkeit verzichten und ihn als direkten Beweis organisieren.

Dass man einen Monotoniesatz ohne Vollständigkeit, also z.B. nur mit Kenntnis der rationalen Zahlen, beweisen kann, liegt nicht daran, dass wir nur eine Verschärfung der Differenzierbarkeitsdefinition betrachten, es geht mit der ϵ - δ -Definition ebenso, wenn man die entsprechenden Gleichmäßigkeitsvoraussetzungen macht. Deshalb denke ich, dass diese Version des Monotoniesatzes eine interessante Variante ist. Bei den Differenzierbarkeits**definitionen** spielt es keine entscheidende Rolle, ob man an die reellen oder nur an die rationalen Zahlen denkt, denn diese Definitionen sagen nur, wie man erkennt, dass eine *gegebene* Zahl m Ableitung $f'(a)$ von f bei a ist. Nur wenn zur Definition von f oder f' irrationale Zahlen nötig sind, kommt man nicht mit den rationalen Zahlen aus.

Das Wort “gleichmäßig” mag eine gewisse Abneigung auslösen, weil es zuerst in der Kombination “gleichmäßig stetig” auftritt und dort mit dem eher schwierigen Beweis verbunden ist, dass eine auf $[a, b]$ stetige Funktion dort sogar(!) gleichmäßig stetig ist. Meiner Meinung nach macht das Adjektiv “gleichmäßig” Begriffe der Analysis einfacher, nicht schwieriger. Und das in so großem Maße, dass die (oft geringe) Mehrarbeit zum Nachweis der Gleichmäßigkeit einer Voraussetzung mehr als gerechtfertigt ist. Zum Beispiel erlaubt eine so verbesserte Voraussetzung oft, einen indirekten Beweis in einen direkten zu verwandeln, wie im folgenden Beispiel.

Der rationale Monotoniesatz

Voraussetzung: Die Funktion f sei in dem Intervall $[\alpha, \omega]$ **gleichmäßig** differenzierbar. Das soll heißen: Es gibt Konstanten δ^f, B^f so dass für alle $a \in [\alpha, \omega]$ ein $m \in \mathbb{R}$ so existiert:

$$x \in [a - \delta^f, a + \delta^f] \cap [\alpha, \omega] \Rightarrow \\ -B^f \cdot (x - a)^2 \leq f(x) - (m \cdot (x - a) + f(a)) \leq B^f \cdot (x - a)^2$$

Hauptvoraussetzung: Für die Ableitung gilt: $f' \geq 0$.

Behauptung:

$$\boxed{x < y \in [\alpha, \omega] \Rightarrow f(x) \leq f(y).}$$

Beweis: Wir werden mehrfach benutzen, dass aus $f' \geq 0$ folgt

$$\xi < \eta \in [x, y] \text{ und } \eta - \xi < \delta^f \Rightarrow \\ \Rightarrow -B \cdot (\eta - \xi)^2 \leq f(\eta) - f(\xi) - f'(\xi) \cdot (\eta - \xi) \leq f(\eta) - f(\xi).$$

Wähle $n > (y - x)/\delta^f$ und teile das Intervall $[x, y]$ in n gleiche Teile, deren Länge $< \delta^f$ ist. Wende auf jedes Teilintervall die eben hergeleitete Ungleichung $-B \cdot (\eta - \xi)^2 \leq f(\eta) - f(\xi)$ an. Diese n Ungleichungen werden summiert, wobei sich an jeder inneren Intervallgrenze die Funktionswerte wegheben, $f(\text{linkes Intervallende}) - f(\text{rechter Intervallanfang}) = 0$. Es bleibt:

$$n > (y - x)/\delta^f \Rightarrow n \cdot (-B \cdot (y - x)^2/n^2) = \frac{-B(y-x)^2}{n} \leq f(y) - f(x).$$

Nun argumentieren wir wie Archimedes, dass diese Ungleichungen zur Folge haben

$$0 \leq f(y) - f(x),$$

also die Behauptung des Monotoniesatzes.

Ich wiederhole: Die Voraussetzung der *gleichmäßigen* Differenzierbarkeit macht diesen Beweis möglich. Wenn man mit der ϵ - δ -Definition der Differenzierbarkeit arbeitet, muss man verlangen, dass zu jedem $\epsilon > 0$ das $\delta > 0$ *gleichmäßig* gewählt werden kann. Dann kann man diesen Beweis anpassen. Natürlich muss man sich mehr anstrengen, um die notwendigen Voraussetzungen als *gleichmäßig* nachzuweisen. Ob man das will oder nicht, ist Geschmackssache. Ich finde auf jeden Fall erwähnenswert, dass man bei gleichmäßigen Voraussetzungen bis hier ohne die Vollständigkeit auskommen kann. Dagegen ist die Vollständigkeit gänzlich unverzichtbar, um die transzendenten Funktionen \exp , \log , \sin , \cos , \tan zu definieren. Selbst bei den Quadratwurzeln ist die Vollständigkeit das bequemere Hilfsmittel, obwohl man sich wegen $\sqrt{a} \cdot \sqrt{a} = a$ und $x^2 < a \Leftrightarrow |x| < \sqrt{a}$ ein gutes Stück weit mit symbolischem Rechnen bewegen kann.

Zusatz: Was ist mit den strikten Ungleichungen, gilt $f' > 0 \Rightarrow f(x) < f(y)$?

Im Archimedes Argument hilft die strenge Ungleichung in der Voraussetzung nicht. Aus $\forall_{n \in \mathbb{N}} -1/n < a$ folgt eben nur $0 \leq a$. Daher kann man den obigen Beweis nicht auf strikte Ungleichungen ausdehnen. Man kann jedoch eine einfache Bemerkung über monotone Funktionen f benutzen:

$$\text{Hat man } x < y \text{ mit } f(x) = f(y) \text{ so folgt aus der Monotonie} \\ \xi \in (x, y) \Rightarrow f(x) \leq f(\xi) \leq f(y) = f(x).$$

Daher ist f konstant in $[x, y]$ und man hat für $\xi \in (x, y)$: $f'(\xi) = 0$.

Wenn also im Monotoniesatz statt $f' \geq 0$ vorausgesetzt wird $f' > 0$, so schließt man zuerst, dass f schwach wachsend ist und zeigt mit der letzten Bemerkung (indirekt!), dass $f(x) = f(y)$ unmöglich ist, also dass f streng wachsend sein muss.

Kommentar zur Vollständigkeit

Selbstverständlich kommen auf der Schule Quadratwurzeln früher vor als die Differentialrechnung. Deshalb sind Schülerinnen und Schüler der Vollständigkeit der reellen Zahlen schon begegnet, wenn sie mit Differentialrechnung beginnen. Zu diesem ersten Kontakt mit der Vollständigkeit habe ich in dem Text A3 *“Die Entwicklung der Grenzwertdefinition beginnt bei Archimedes”* geschrieben. Für die Vollständigkeit gilt: *Wenn man mit irrationalen Zahlen zu tun hat, braucht man sie – wenn man mit den rationalen Zahlen auskommt, braucht man sie nicht.* Da auch das Vollständigkeitsaxiom der reellen Zahlen aus der zweiten Hälfte des 19. Jahrhunderts stammt, habe ich auf den vorhergehenden Seiten beim Aufbau der Anfänge der Differentialrechnung betont, dass man diesen Anfang mit den rationalen Zahlen bewältigen kann. Klar, dass diese Betonung Geschmacksache ist, denn die Differentialrechnung wurde nicht für diese Anfänge geschaffen. Schon bei den nächsten Schritten – Umkehrfunktionen, Exponentialfunktionen, trigonometrische Funktionen – hängen die *Definitionen* unvermeidbar von der Vollständigkeit ab. Das steht nicht im Widerspruch dazu, dass frühe Tafeln trigonometrischer Funktionen viel älter und die Neperschen Logarithmentafeln immer noch 250 Jahre älter als die Vollständigkeit sind, denn diese Tafelwerke wurden auch zu ihrer Entstehungszeit als Approximationen aufgefasst.

Was ich zu Quadratwurzeln im Text A3 geschrieben habe, wiederhole ich hier nicht. Ich beginne mit einer Besprechung der Exponentialfunktionen, wobei die Betonung etwas anders sein wird als in dem Text A2, von dem ich mir ja vorstelle, dass er parallel zu Schulbüchern, also vergleichend, gelesen werden soll.

Exponentialfunktionen

Ich kenne zwei plausible Vorgehensweisen. Eine betont den Umgang mit konvergenten Folgen, die Hauptarbeit fließt in die Potenzregel $\exp(x + y) = \exp(x)\exp(y)$, Differentialrechnung tritt erst im abschließenden Differenzierbarkeitsbeweis auf. Die andere argumentiert von Anfang an mit dem Monotoniesatz, die Hauptarbeit braucht man für die Differenzierbarkeit der als Grenzfunktion definierten Exponentialfunktion.

Skizze des *ersten* Falls: Beginn mit 2^m . Mit Hilfe n -ter Wurzeln wird dann $2^{m/n}$ definiert. Für irrationale Exponenten habe ich in Schulbüchern keine Definition gesehen. Dazu wird die Vollständigkeit vorausgesetzt. Man kann also zu jedem reellen $x \in \mathbb{R}$ monoton wachsende *rationale* Folgen r_n und monoton fallende *rationale* Folgen s_n finden mit

$$r_n \leq x \leq s_n \text{ und } 0 \leq s_n - r_n \leq 1/n, \text{ also } x = \lim_{n \rightarrow \infty} r_n = \lim_{n \rightarrow \infty} s_n.$$

Nun *soll definiert werden* $2^x := \lim_{n \rightarrow \infty} 2^{r_n} = \lim_{n \rightarrow \infty} 2^{s_n}$. Für die benötigte Abschätzung muss zuerst für **rationale** $r, s \in \mathbb{Q}$ das Potenzgesetz $2^r \cdot 2^s = 2^{r+s}$ gefolgert werden. Dann hat man:

$$0 \leq 2^{s_n} - 2^{r_n} = 2^{r_n}(2^{s_n - r_n} - 1) \leq 2^{s_1}(2^{1/n} - 1) \leq 2^{s_1}/n,$$

wobei die letzte Ungleichung aus der Bernoullischen Ungleichung $(1 + x)^n \geq 1 + nx$ mit $x := 1/n$ folgt: $(1 + 1/n)^n \geq 2 \Rightarrow 1 + 1/n \geq 2^{1/n}$.

Damit ist 2^x für alle $x \in \mathbb{R}$ definiert und das Potenzgesetz kann endgültig gezeigt werden:

$$x, y \in \mathbb{R} \Rightarrow 2^x \cdot 2^y = 2^{x+y}.$$

Dann folgt aus $(2^{x+h} - 2^x)/h = 2^x \cdot (2^h - 1)/h$, dass 2^x differenzierbar ist, wenn es bei $x = 0$ differenzierbar ist. Mit $h := \pm 1/n$ bekommt man eine fallende Folge $s_n = n(2^{1/n} - 1)$ oberhalb von einer wachsenden Folge $r_n = n(1 - 2^{-1/n})$. Dieselbe Bernoulli-Abschätzung $0 \leq s_n - r_n = n \cdot 2^{-1/n}(2^{1/n} - 1)^2 \leq n \cdot 1/n^2$ liefert die Differenzierbarkeit bei $x = 0$. Mit

$c := \lim_{h \rightarrow 0} (2^h - 1)/h$ setze: $f(x) := 2^{x/c}$. Die Kettenregel zeigt $f' = f$. Man ist also bei der natürlichen Exponentialfunktion angekommen.

Den *zweiten Zugang* erkläre ich ausführlicher, weil dies ein Text zur Differentialrechnung ist. Man beginnt mit Prozentrechnung: Zu einem Anfangswert sollen $p\%$ addiert werden oder von einem Endwert sollen $p\%$ subtrahiert werden. Dazu wird $x = p/100$ gesetzt und man hat entweder

$$\text{Endwert} = \text{Anfangswert} \cdot (1 + x) \text{ oder } \text{Anfangswert} = \text{Endwert} \cdot (1 - x).$$

Was ändert sich, wenn man n -mal $(p/n)\%$ addiert bzw. subtrahiert? Man muss n -mal mit $(1 + x/n)$ bzw. mit $(1 - x/n)$ multiplizieren! Beide Male vom Anfangswert ausgehend hat man also:

$$\text{Endwert} = \text{Anfangswert} \cdot (1 + x/n)^n \text{ oder } \text{Endwert} = \text{Anfangswert} \cdot (1 - x/n)^{-n}.$$

Um dies auch für große n besser untersuchen zu können, definieren wir die Funktionen

$$f_n(x) := (1 + x/n)^n \text{ und } g_n(x) := (1 - x/n)^{-n}.$$

Da diese Funktionen bei multiplikativen Vergleichen entstanden sind, benötigen wir den **multiplikativen Monotoniesatz**:

Voraussetzung: Die Wachstumsraten der positiven Funktionen f, g erfüllen: $f'/f \leq g'/g$.

Behauptung: Die Quotientenfunktion $h := f/g$ ist fallend: $x < y \Rightarrow \frac{f(y)}{g(y)} \leq \frac{f(x)}{g(x)}$.

Beweis: $h' = f'/g - fg'/g^2 = (f/g) \cdot (f'/f - g'/g) \leq 0$.

Um diesen Satz zu benutzen, werden die Wachstumsraten berechnet:

$$-m < x \Rightarrow \frac{f'_m}{f_m}(x) = \left(1 + \frac{x}{m}\right)^{-1} \text{ und } x < n \Rightarrow \frac{g'_n}{g_n}(x) = \left(1 - \frac{x}{n}\right)^{-1},$$

also mit dem multiplikativen Monotoniesatz, $f_m(0) = 1 = g_n(0)$, $0 \leq x < n$ wegen

$$\frac{1}{1 + x/m} \leq \frac{1}{1 + x/(m+1)} \leq 1 \leq \frac{1}{1 - x/(n+1)} \leq \frac{1}{1 - x/n}$$

sofort: $1 \leq \dots \leq f_m(x) \leq f_{m+1}(x) \leq \dots \leq g_{n+1}(x) \leq g_n(x)$.

Für $-m < x \leq 0$ hat man in den Wachstumsraten \geq statt \leq und damit

wie eben: $0 < f_m(x) \leq f_{m+1}(x) \leq \dots \leq g_{n+1}(x) \leq g_n(x) \leq \dots \leq 1$.

Da auch die Differenz $g_n - f_n$ z.B. für $|x| \leq 2$ gut abgeschätzt werden kann

$$\begin{aligned} 0 \leq g_n(x) - f_n(x) &= f_n(x) \cdot \left(\left(\frac{1}{1 - x^2/n^2} \right)^n - 1 \right) \leq \\ &[\text{benutze im Nenner die Bernoullische Ungleichung für } n \geq 8] \\ &\leq g_8(x) \cdot \left(\frac{1}{1 - x^2/n} - 1 \right) \leq 0.75^{-8} \cdot \frac{x^2/n}{0.5} \leq 20 \cdot 2^2/n \end{aligned}$$

folgt aus der Vollständigkeit, dass die Funktionen f_n und g_n für $|x| \leq 2$ gegen eine gemeinsame Grenzfunktion f_∞ konvergieren. Weil die Wachstumsraten der Funktionen f_n, g_n gegen 1 konvergieren, können wir hoffen, dass f_∞ differenzierbar ist mit Wachstumsrate 1. Wer die klassischen Differenzierbarkeitsbeweise mit gleichmäßig konvergenten Funktionenfolgen kennen gelernt hat, wird überrascht sein, wie einfach das aus dem Monotoniesatz

folgt. Schließlich, nachdem man für die Grenzfunktion in $|x| \leq 2$ nun $f' = f$ weiß, liefern weitere Anwendungen des Monotoniesatzes die Ausdehnung der Definition auf alle $x \in \mathbb{R}$ und die Gültigkeit der Potenzgesetze.

Differenzierbarkeit der Grenzfunktion

Anders als in den allgemeineren klassischen Sätzen können wir zusätzliche Voraussetzungen benutzen. Zunächst folgt für $m, n \geq 8$ und $|x| \leq 2$

$$0 < f'_m(x), g'_n(x) \leq g'_8(2) \leq 0.75^{-9} < 14$$

und damit aus dem Monotoniesatz:

$$-2 \leq x < y \leq 2 \Rightarrow 0 \leq f_m(y) - f_m(x) \leq 14(y - x),$$

$$0 \leq g_n(y) - g_n(x) \leq 14(y - x)$$

und daraus mit dem Eudoxos Axiom wegen $|f_m(x) - f_\infty(x)| \leq 80/m$:

$$-2 \leq x < y \leq 2 \Rightarrow 0 \leq f_\infty(y) - f_\infty(x) \leq 14(y - x).$$

Die Grenzfunktion f_∞ ist also in harmloser Weise stetig, man sagt: “Lipschitz stetig”.

Für die Differenzierbarkeit werden ebenfalls bekannte Schranken für f''_m, g''_n benutzt:

$$0 < f''_m(x), g''_n(x) \leq g''_8(2) \leq \frac{9}{8} 0.75^{-9} < 16,$$

also mit dem Monotoniesatz “Abweichung von der Tangente” bei $x = a$:

$$0 \leq f_m(x) - (f_m(a) + f'_m(a)(x - a)) \leq 8(x - a)^2,$$

und wieder mit Eudoxos und mit

$$\lim_{m \rightarrow \infty} f_m(x) = f_\infty(x), \quad \lim_{m \rightarrow \infty} f_m(a) = f_\infty(a), \quad \lim_{m \rightarrow \infty} f'_m(a) = f_\infty(a) \text{ schließlich:}$$

$$0 \leq f_\infty(x) - (f_\infty(a) + f_\infty(a)(x - a)) \leq 8(x - a)^2.$$

Die lineare Funktion $f_\infty(a) + f_\infty(a)(x - a)$ weicht bei $x = a$ von f_∞ nur um den quadratischen Fehler $8(x - a)^2$ ab, ist also dort die Tangente. Also ist f_∞ differenzierbar mit $f'_\infty = f_\infty$. Der übliche Name dieser Funktion ist $\exp = f_\infty$.

Ausdehnung des Definitionsbereichs

Da $f_\infty(x)$ für $|x| \leq 2$ definiert ist, ist $h_n(x) := f_\infty(x/n)^n$ für $|x| \leq 2n$ definiert, wegen der Kettenregel folgt $h'_n = h_n$ und der Monotoniesatz liefert (beachte $h_n(0) = 1$): Auf dem Durchschnitt ihrer Definitionsbereiche stimmen h_n und h_m überein. Auf diese Weise haben wir für jedes $x \in \mathbb{R}$ eine Definition für $\exp(x)$ und es gilt auf \mathbb{R} : $\exp' = \exp$.

Potenzregeln

Für die Hilfsfunktion $h(x) := \exp(x) \cdot \exp(-x)$ folgt $h' = 0$, also mit dem Monotoniesatz $h(x) = h(0) = 1$ und daher $\exp(-x) = 1/\exp(x)$.

Für die Hilfsfunktion $h(x) := \exp(x + a) \cdot \exp(-x)$ folgt $h' = 0$, also mit dem Monotoniesatz $h(x) = h(0) = \exp(a)$ und daher $\exp(x + a) = \exp(x) \cdot \exp(a)$.

Die letzte Potenzregel, $\exp(x \cdot a) = \exp(x)^a$, muss warten, bis die Potenzfunktion $x^a, x > 0$ für beliebige reelle a definiert und ihre Ableitung, $a \cdot x^{a-1}$, bestimmt wurde.

Umkehrfunktionen

Umkehrfunktionen werden nur für streng monotone Funktionen definiert, wenn nötig nur auf Teilintervallen ihres Definitionsbereichs. Zu einer solchen Funktion $x \rightarrow f(x) =: y$ ist die Umkehrfunktion $u, y \rightarrow u(y)$, definiert durch

$$u(f(x)) = x \text{ oder gleichwertig } f(u(y)) = y.$$

Zeichnet man den Graphen von f auf eine durchsichtige Folie und sieht sie so von der Rückseite an, dass die y-Achse horizontal nach rechts zeigt, so sieht man den Graphen von u . Zeichnet man die Graphen von f und u in dasselbe Koordinatensystem, so liegen die beiden spiegelsymmetrisch zur Winkelhalbierenden des ersten Quadranten. Man erwartet daher, dass die Umkehrfunktion differenzierbar ist.

Die Umkehrfunktion der streng monotonen Funktion $f(x) := x^3$ ist nicht differenzierbar bei $y = 0$. Da wir an der Ableitung der Umkehrfunktion interessiert sind, verschärfen wir die Voraussetzungen noch etwas: Wir betrachten nur Funktionen f , deren *Ableitung nicht verschwindet* - nötigenfalls wieder durch Einschränkung auf ein Teilintervall des Definitionsbereichs. Außerdem kann $f' > 0$ angenommen werden, andernfalls betrachte $f(-x)$.

Differentiationsregel für Umkehrfunktionen

Bezeichnungen: Tangente $T_a^f(x) = f(a) + f'(a)(x - a)$, $y = f(x)$, $b = f(a)$, $u(y) = x$.

Voraussetzung: $f' > 0$ und $|x - a| \leq \delta \Rightarrow |f(x) - T_a^f(x)| \leq B(x - a)^2$.

Behauptung: $T_b^u(y) = u(b) + (1/f'(a))(y - b)$, also $u'(b) = 1/f'(a) = 1/f'(u(b))$.

Beweis:

Wir schreiben die Voraussetzung ohne Beträge und passen die Bezeichnungen an die Umkehrfunktion an ($|x - a| \leq \delta$ ist immer vorausgesetzt):

$$\begin{aligned} (1) \quad & -B(x - a)^2 \leq f(x) - f(a) - f'(a)(x - a) \leq B(x - a)^2 \quad \text{oder} \\ (2) \quad & -B(x - a)^2 \leq y - b - f'(a)(u(y) - u(b)) \leq B(x - a)^2 \quad \text{oder} \\ (3) \quad & -\frac{B}{f'(a)}(x - a)^2 \leq \frac{1}{f'(a)}(y - b) - u(y) + u(b) \leq \frac{B}{f'(a)}(x - a)^2. \end{aligned}$$

Der mittlere Term ist schon die Differenz zwischen $u(y)$ und der behaupteten Tangente, $T_b^u(y) = u(b) + (1/f'(a))(y - b)$. Nun müssen die quadratischen Fehlerterme $(x - a)^2$ noch durch $(y - b)^2$ ersetzt werden. Dazu machen wir nur die Hälfte der Umbenennungen von Zeile 1 zu Zeile 2 und ordnen die Terme etwas anders:

$$(f'(a) - B(x - a)) \cdot (x - a) \leq y - b \leq (f'(a) + B(x - a)) \cdot (x - a).$$

Da wir $|x - a|$ durch $|y - b|$ abschätzen wollen, verkleinern wir nötigenfalls δ , so dass $B \cdot \delta < f'(a)/2$ ist. Dann folgt $f'(a)/2 \leq f'(a) - B|x - a|$, also $|x - a| \leq \frac{2}{f'(a)}|y - b|$. Damit verbessert sich Gleichung (3) zu

$$(4) \quad |x - a| \leq \delta \Rightarrow |u(y) - (u(b) + \frac{1}{f'(a)}(y - b))| \leq \frac{2B}{f'(a)^2}(y - b)^2.$$

Das ist fast unser Ziel - nur soll es aus der Kleinheit von $|y - b|$ gefolgert werden, nicht aus der von $|x - a|$. Dazu definiere $b_1 := f(a - \delta) < b$ und $b_2 := f(a + \delta) > b$. Das ergibt:

$$(5) \quad \underline{y \in [b_1, b_2]} \Rightarrow |x - a| \leq \delta \Rightarrow |u(y) - (u(b) + \frac{1}{f'(a)}(y - b))| \leq \frac{2B}{f'(a)^2}(y - b)^2,$$

wie es die Differenzierbarkeitsdefinition für $y \rightarrow u(y)$ bei $y = b$ verlangt.

Sehnensteigungen von a^x , $a > 1$, bei $x = 0$.

Für die Differenzierbarkeit einer Funktion f bei $x = 0$ genügt es nicht, die Sehnensteigungen $(f(h) - f(0))/h$ nur für $h = 1/n, n \in \mathbb{N}$ zu betrachten, sondern für alle $0 < h \leq \delta, h \in \mathbb{R}$.

Angenommen, man könnte $f_r(x) = (x^r - 1^r)/r$ differenzieren: $f'_r(x) = x^{r-1}$, dann folgte:

1. Für jedes $a > 1$ und $r > 0$ gilt für $h(x) := a^x$:

Die Sehnensteigungen sind $(h(r) - h(0))/r = (a^r - 1)/r = f'_r(a)$.

2. Für $0 < r < s$ und $a > 1$ gilt $1 \leq x \leq a \Rightarrow f'_r(x) \leq f'_s(x)$, also $f_r(a) \leq f_s(a)$.

Also: Aus der Kenntnis der (gebrochenen) Potenzfunktion $f_r(x)$ folgt die Monotonie der Sehnensteigungen bei $x = 0$ der Exponentialfunktion $h(x) = a^x$ und damit die Differenzierbarkeit von h bei $x = 0$.

Leider ist der Zugang zu den gebrochenen Potenzfunktionen am einfachsten über die Umkehrfunktion $\ln(x)$ von $\exp(x)$. – Der direkte Weg über $r \in \mathbb{N}$, $r \in \mathbb{Q}$, $r \in \mathbb{R}$ ist möglich aber mühsam.

Aus diesem Grund habe ich mich für folgende Strategie entschieden: Jede neue Konstruktion von Funktionen – z.B.: Umkehrfunktionen, implizit definierte Funktionen, Grenzfunktionen – erfordert den Beweis ihrer eigenen Differentiationsregel.

Für den Beweis der Differenzierbarkeit von Grenzfunktionen gibt es

Erstens den häufigeren Weg mit **gleichmäßig konvergenten** Funktionenfolgen und *zweitens* den an Arzela-Ascoli angelehnten Weg mit **gleichgradig differenzierbaren** Funktionenfolgen.

Im Kontext der Schulmathematik hat man fast immer Schranken für zweite Ableitungen zur Verfügung und kann daher leicht den zweiten Weg gehen. Z.B. ist das für

$$(1 + x/n)^n \leq f_\infty(x) \leq (1 - x/n)^{-n}$$

der Fall, so dass dies ein Weg zur Exponentialfunktion ist, bei dem man sich nicht den ganzen Weg lang mit immer neuen kleinen Detailschwierigkeiten auseinandersetzen muss.