

7.2. Ableitungen und lineare Approximation

Eindimensionale Ableitungen und Differentialquotienten

einer Funktion bekommt man bekanntlich als Limes von **Differenzenquotienten**

$$f'(a) = \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h} = \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a}.$$

Die Steigung der Tangente in einem Punkt mit der Ordinate a ist nichts anderes als die Ableitung, vorausgesetzt natürlich, daß diese Tangente existiert, d.h. daß die Funktion zumindest in diesem Punkt **differenzierbar** ist (insbesondere keinen Knick und keinen Sprung hat). Es gilt dann

$$f(a+h) = f(a) + f'(a)h + o(h)$$

wobei $o(h)$ eine Funktion ist, die schneller gegen 0 konvergiert als h , d.h.

$$\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0.$$

Beispiel 1: Parabel und Tangente

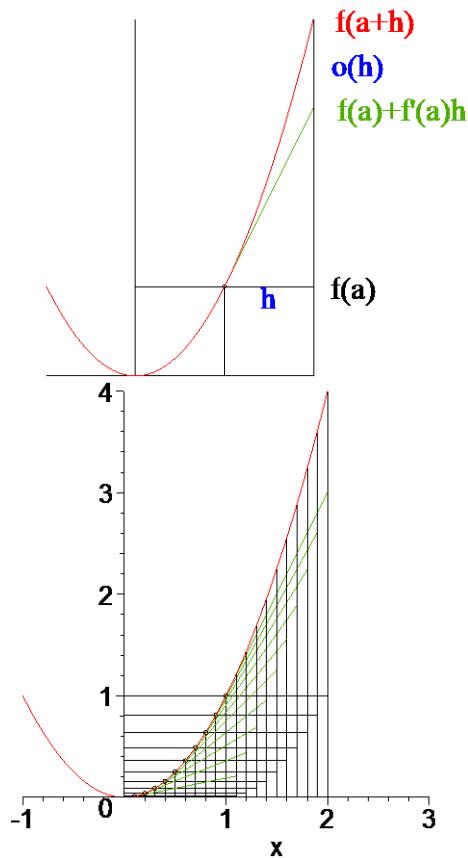
$$f(x) = x^2$$

$$f(a+h) = f(a) + f'(a)h + o(h)$$

$$(a+h)^2 = a^2 + 2ah + h^2$$

$$f'(a) = 2a$$

$$o(h) = h^2$$



Einige spezielle Ableitungen

Die Ableitung der Potenzfunktion x^n ist $n x^{(n-1)}$:

$$\lim_{h \rightarrow 0} \frac{(a+h)^n - a^n}{h} = n a^{(n-1)} + (\lim_{h \rightarrow 0} q(h))$$

mit einem Polynom $q(h)$ ohne konstantes Glied.

Die Ableitung der Exponentialfunktion $e(x) = e^x$ ist e^x :

$$\lim_{h \rightarrow 0} \frac{e(a+h) - e(a)}{h} = e(a) \left(\lim_{h \rightarrow 0} \frac{e(h) - 1}{h} \right) = \left(1 + \left(\lim_{h \rightarrow 0} \sum_{k=1}^{\infty} \frac{h^k}{(k+1)!} \right) \right) e(a) = e(a).$$

Die Ableitung der Logarithmusfunktion $\ln(x)$ ist $\frac{1}{x}$:

$$\lim_{x \rightarrow a} \frac{\ln(x) - \ln(a)}{x - a} = \lim_{y \rightarrow b} \frac{y - b}{e(y) - e(b)} = \frac{1}{e(b)} = \frac{1}{a} \quad (\text{mit } y = \ln(x) \text{ und } b = \ln(a)).$$

Die Ableitung der Sinusfunktion $\sin(x)$ ist $\cos(x)$:

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\sin(a+h) - \sin(a)}{h} &= \lim_{h \rightarrow 0} \frac{\sin(a) \cos(h) + \cos(a) \sin(h) - \sin(a)}{h} \\ &= \sin(a) \left(\lim_{h \rightarrow 0} \frac{\cos(h) - 1}{h} \right) + \cos(a) \left(\lim_{h \rightarrow 0} \frac{\sin(h)}{h} \right) = \cos(a). \end{aligned}$$

Dazu berechnen wir

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\cos(h) - 1}{h} &= \lim_{h \rightarrow 0} \frac{\cos(h)^2 - 1}{h(\cos(h) + 1)} = \\ \lim_{h \rightarrow 0} \frac{-\sin(h)^2}{h(\cos(h) + 1)} &= \left(\lim_{h \rightarrow 0} \frac{-\sin(h)}{h} \right) \left(\lim_{h \rightarrow 0} \frac{\sin(h)}{\cos(h) + 1} \right) = 0. \end{aligned}$$

Die Ableitung der Cosinusfunktion $\cos(x)$ ist $-\sin(x)$
(mit einer ähnlichen Rechnung).

Lineare Approximation

Entsprechende Approximationsmöglichkeiten hat man auch für Funktionen zwischen Teilmengen mehrdimensionaler Räume, wobei man Tangenten im Falle von Kurven durch

Tangentialvektoren, im Falle von Flächen bzw. Funktionsgebirgen hingegen durch

Tangentialebenen zu ersetzen hat. Aus der linearen Algebra wissen wir andererseits, daß man Geraden, Ebenen und allgemeinere Unterräume mit Hilfe linearer Abbildungen und diese wiederum mit Matrizen beschreiben kann.

Gegeben sei eine Funktion f zwischen einer Teilmenge A des R_n und einer Teilmenge B des R_m . Wir wollen Funktionswerte $f(a+h)$ in der Nähe eines Punktes (bzw. Vektors) a im Inneren des Definitionsbereichs A annähern (approximieren), indem wir zum Funktionswert $f(a)$ den Wert einer geeigneten linearen Funktion, angewandt auf einen Verschiebungsvektor h aus R_n , addieren.

Die Ableitung

der Funktion f an der Stelle a ist eine mit $f'(a)$ oder $Df(a)$ bezeichnete Matrix (!) aus $R^{(m \times n)}$. Sie existiert und ist festgelegt durch die Gleichung

$$f(a + h) = f(a) + f'(a)h + o(h)$$

falls die **Restfunktion** $o(h)$ schneller gegen 0 geht als der Vektor (!) h . Präzise bedeutet das:

$$\lim_{h \rightarrow 0} \frac{o(h)}{|h|} = 0,$$

oder explizit:

Zu jedem $\varepsilon > 0$ gibt es ein $\delta > 0$, so daß aus $|h| < \delta$ stets $|o(h)| < \varepsilon |h|$ folgt.

Im Allgemeinen hängt die Funktion $o(h)$ nicht nur von h , sondern auch von der Funktion f und meist auch von der gewählten Stelle a ab. Existiert die Ableitung in jedem Punkt einer Teilmenge T von A , so wird dadurch eine neue Abbildung

$$Df = f' \text{ von } T \text{ nach } R^{(m \times n)}$$

definiert, die jedem Punkt aus T die Ableitung in diesem Punkt zuordnet.

Das (totale) Differential

von f an der Stelle a ist die zugehörige lineare Abbildung

$$df(a) \text{ von } R_n \text{ nach } R_m \text{ mit } df(a)(h) = f'(a)h.$$

Falls es existiert, nennt man f (**total**) **differenzierbar** im Punkt a .

Die Tangentialfunktion

$$Tf(a)(x) = f(a) + f'(a)(x - a)$$

beschreibt anschaulich die lineare Approximation an f in der Nähe des Punktes a , wobei wir hier den variablen Punkt $a + h$ mit x bezeichnen haben.

Wie man mehrdimensionale Ableitungen, Differentiale und Tangentialfunktionen generell berechnet, werden wir in Kürze mit Hilfe sogenannter **partieller Ableitungen** sehen. Zunächst betrachten wir einige einfache Beispiele, wo solche Rechnungen überflüssig sind.

Spezialfall 1: Affine Abbildungen

Die einfachsten und "glattesten" differenzierbaren Funktionen sind natürlich diejenigen, die sich selbst linear approximieren, d.h. selbst **linear** oder wenigstens **affin** sind, also von der Form

$$f(x) = Mx + c$$

mit einer Matrix M aus $R^{(m \times n)}$ und einem konstanten Vektor c aus R_m .

Für solche Abbildungen f ist $f'(a) = M$ konstant.

Spezialfall 2: Eindimensionale Ableitungen

Für $m = 1$ und $n = 1$ ist $f'(a)$ eine Zahl, und zwar die übliche **Ableitung (Steigung)** im Punkt a .

Die Tangentialfunktion

$$Tf(a)(x) = f(a) + f'(a)(x - a)$$

ist dann durch eine "Punkt-Steigungs-Gleichung" gegeben.

Beispiel 2: Kubische Parabel

$$f(x) = x^3$$

$$f(a+h) = f(a) + f'(a)h + o(h)$$

$$(a+h)^3 = a^3 + 3a^2h + 3ah^2 + h^3$$

$$f'(a) = 3a^2$$

$$o(h) = 3ah^2 + h^3$$

$$Tf(a)(x) = 3a^2x - 2a^3$$

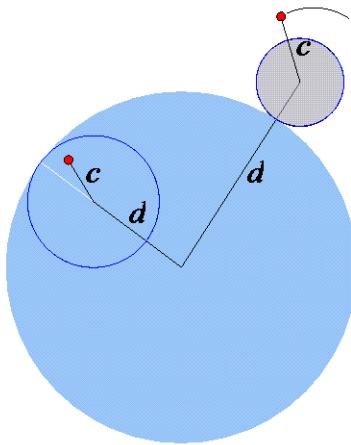
Spezialfall 3: Ableitungen von Kurven

Eine auf einem Intervall definierte Funktion f nach R_m beschreibt eine **Kurve** im m -dimensionalen Raum, insbesondere eine **ebene Kurve** für $m = 2$ und eine **Raumkurve** für $m = 3$. Der **Spaltenvektor** $f'(a)$ ist ein **Tangentialvektor** an die Kurve im Punkt a , physikalisch interpretierbar als **Geschwindigkeitsvektor**, falls man die Eingangsvariable als Zeitparameter auffaßt.

Sein Betrag ist die (**skalare**) **Geschwindigkeit** (die man z.B. auf der Skala des Tachometers abliest).

Beispiel 3: Ebene Rollkurven

enstehen etwa durch Abrollen eines kleinen um einen großen Kreis.



Bei fest gewählten Zahlen c (Abstand vom Mittelpunkt des abrollenden Kreises mit Radius 1) und d (Abstand der Mittelpunkte) sind sie gegeben durch

$$k(t) = \begin{bmatrix} d \cos(t) - c \cos(dt) \\ d \sin(t) - c \sin(dt) \end{bmatrix}.$$

Die Ableitung wird für die beiden Koordinatenfunktionen separat bestimmt.

$$\begin{aligned}k'_1(t) &= -d \sin(t) + c d \sin(d t) \\k'_2(t) &= d \cos(t) - c d \cos(d t).\end{aligned}$$

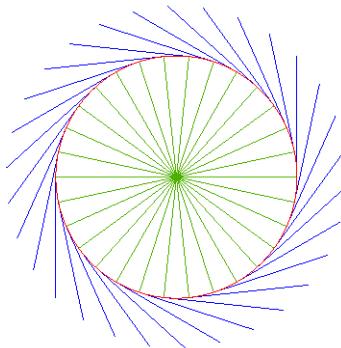
Die Tangenten zum Zeitpunkt t haben dann die komponentenweise Parameterdarstellung

$$\begin{aligned}k_1(t) + s k'_1(t) &= d \cos(t) - c \cos(d t) + s d (-\sin(t) + c \sin(d t)) \\k_2(t) + s k'_2(t) &= d \sin(t) - c \sin(d t) + s d (\cos(t) - c \cos(d t)).\end{aligned}$$

Für den Sonderfall $c = 0$ haben wir einfach eine Rotation um den Ursprung im Abstand d . Der Tangentialvektor

$$[d \cos(t), -d \sin(t)]^T$$

hat die gleiche Länge wie der rotierende Radiusvektor und steht senkrecht auf diesem.



Dies eröffnet eine besonders anschauliche Möglichkeit, die Ableitung von Sinus und Cosinus zu bestimmen:

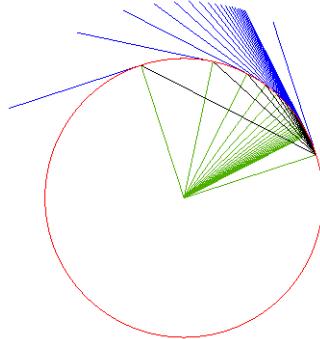
Der Radiusvektor des Einheitskreises mit Spitze beim Winkel t (im Bogenmaß!) ist

$$\mathbf{r}(t) = [\cos(t), \sin(t)]^T.$$

Seine Ableitung ist ein Tangentialvektor, also senkrecht dazu. Der Quotient aus der Länge des Differenzvektors zwischen zwei Radien und dem entsprechenden Kreisbogenstück

$$\frac{|\mathbf{r}(t+h) - \mathbf{r}(t)|}{h}$$

nähert sich dem Wert 1, wenn h gegen 0 strebt. Also muß der Ableitungsvektor die Länge 1 haben.

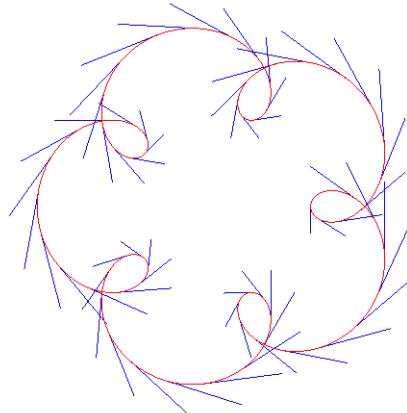


Es bleibt nur die Möglichkeit

$$\mathbf{r}'(t) = [-\sin(t), \cos(t)]^T,$$

und wir haben die Ableitungen von Cosinus und Sinus ohne Rechnerei mit einem Schlag erledigt.

Für den Fall $c = 2$ und $d = 6$ zeichnen wir noch einmal die Rollkurve und dazu einige Tangentialvektoren.



Die **skalare Geschwindigkeit** ist die Länge des jeweiligen Tangentialvektors, in diesem Beispiel also

$$\sqrt{(-d \sin(t) + c d \sin(d t))^2 + (d \cos(t) - c d \cos(d t))^2} = d \sqrt{1 + c^2 - 2 c \cos((d-1)t)}$$

Wie man an der Länge der Tangentialvektoren sieht, wird in den stärker gekrümmten Kurven abgebremst. Die obige Rollkurve hat nicht eine Sechser-, sondern eine Fünfer-Symmetrie, denn das Verhältnis des großen zum kleinen Kreisradius ist gleich

$$\frac{d-1}{1} = 5.$$

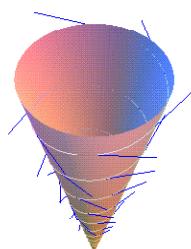
Das Kurvenbild wiederholt sich also bei einer Umdrehung fünf mal.

Beispiel 4: Eine räumliche Schneckenlinie

wächst mit einem exponentiellen Faktor c^t . Bei Rotation um die Wachstumsachse entsteht daher folgende Kurve:

$$\mathbf{f}(t) = \begin{bmatrix} c^t \cos(t) \\ c^t \sin(t) \\ d c^t \end{bmatrix}, \mathbf{f}'(t) = \begin{bmatrix} c^t \ln(c) \cos(t) - c^t \sin(t) \\ c^t \ln(c) \sin(t) + c^t \cos(t) \\ d c^t \ln(c) \end{bmatrix}$$

$c = 1.05, d = 3$



Spezialfall 4: Ableitungen von Flächen

Eine Funktion f von einer Teilmenge des n -dimensionalen Raumes R_n nach R beschreibt eine **Hyperfläche**, im Falle $n = 2$ eine gewöhnliche **Fläche**. Die **Tangentialebene** im Punkt a ist dann gegeben durch die obige Tangentialfunktion $Tf(a, x)$. Man nennt den **Zeilenvektor** $f'(a)$ in diesem Fall den **Gradienten** im Punkt a . Der lineare Zuwachs

$$f'(a)(x - a)$$

ist als **Skalarprodukt** der Zeile $f'(a)$ mit der Spalte $x - a$ zu verstehen.

Im Folgenden interessiert uns vor allem der Fall $n = 2$. Hier ist alternativ folgende Notation üblich: Man bezeichnet die beiden Komponenten

- eines **festen Punktes** bzw. Ortsvektors mit x_0 und y_0 ,
- eines **variablen Punktes** bzw. Vektors mit x und y .

Der **Differenzvektor** hat dann die Komponenten $h_1 = x - x_0$ und $h_2 = y - y_0$, und die **Tangentenfunktion** einer im Punkt (x_0, y_0) differenzierbaren Funktion f lautet

$$Tf(x_0, y_0)(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0),$$

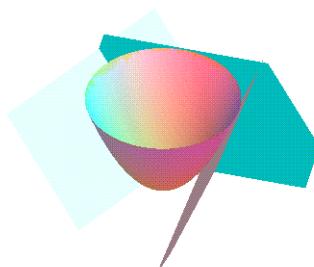
wobei $f_x(x_0, y_0)$ und $f_y(x_0, y_0)$ die Komponenten des Gradienten $f'(x_0, y_0)$ sind, die so genannten **partiellen Ableitungen** nach x und y . Man erhält sie, indem man nach einer der Variablen ableitet und dabei die andere als Konstante betrachtet. Mehr davon in Abschnitt 7.3.) Wenn man es ganz genau nimmt, muss man hier überall eigentlich

$$f\left(\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}\right) \text{ statt } f(x_0, y_0) \text{ usw.}$$

schreiben. Das ist aber ebenso unbequem wie platzraubend und wird fast nie gemacht.

Beispiel 5: Paraboloid mit Tangentialebenen

$$\begin{aligned} f(x, y) &= x^2 + y^2 \\ f(x_0 + h_1, y_0 + h_2) &= x_0^2 + y_0^2 + 2x_0 h_1 + 2y_0 h_2 + h_1^2 + h_2^2 \\ f'(x_0, y_0) &= [2x_0, 2y_0] \\ o(h) &= f(h) \\ Tf(x_0, y_0)(x, y) &= x_0^2 + y_0^2 + 2x_0(x - x_0) + 2y_0(y - y_0) \end{aligned}$$



Wagen wir uns jetzt noch an beliebige Dimensionen m und n .

Koordinatenfunktionen und Jacobi-Matrix

Im allgemeinsten Fall einer beliebigen Funktion f von R_n nach R_m betrachtet man die **Koordinatenfunktionen**

$$f_i \text{ von } R_n \text{ nach } R \text{ mit } f(a)^T = [f_1(a), \dots, f_m(a)].$$

Falls f in a differenzierbar ist, hat die **Jacobi-Matrix** genannte Ableitung $Df(a) = f'(a)$ als Zeilen die Gradienten

$$Df_i(a).$$

Es ist also speziell für $m = 2$ und $n = 2$ (nach Umbenennung von x_1 in x und von x_2 in y):

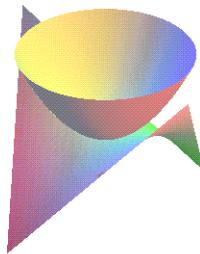
$$Df(x, y) = \begin{bmatrix} f_{1,x}(x, y) & f_{1,y}(x, y) \\ f_{2,x}(x, y) & f_{2,y}(x, y) \end{bmatrix}$$

Dabei ist $f_{1,x}(x, y)$ die Ableitung von $f_1(x, y)$ nach x bei konstant gehaltenem y , usw.

Beispiel 6: Eine Jacobi-Matrix

$$\begin{aligned} f(x, y) &= \begin{bmatrix} x^2 + y^2 \\ xy \end{bmatrix} \\ Df(x, y) &= \begin{bmatrix} 2x & 2y \\ y & x \end{bmatrix} \end{aligned}$$

Die beiden Koordinatenfunktionen in einem Bild:



Differentiationsregeln

Die aus der eindimensionalen Differentialrechnung bekannten Regeln für das Ableiten zusammengesetzter Funktionen treffen bei richtiger Interpretation auch für höhere Dimensionen zu. Was wir über stetige Funktionen gesagt haben, gilt entsprechend abgewandelt auch für zwei differenzierbare Funktionen f und g .

Summenregel

Sind f und g auf der gleichen Menge differenzierbar, so auch $f+g$ und $f-g$, und es gilt

$$\begin{aligned} (f+g)' &= f' + g', \\ (f-g)' &= f' - g'. \end{aligned}$$

Produktregel

Sind f und g differenzierbare Funktionen von einer Menge A reeller Zahlen nach R_m und bedeutet $f * g$ das elementweise Skalarprodukt (oder das Vektorprodukt, falls $m = 3$), so ist auch die Funktion $f * g$ von A nach R wieder differenzierbar, und für a aus A gilt:

$$(f * g)'(a) = f'(a) * g(a) + f(a) * g'(a).$$

Speziell gilt für den eindimensionalen Fall (in verkürzter Schreibweise):

$$(f \cdot g)' = f'g + fg'.$$

Kettenregel

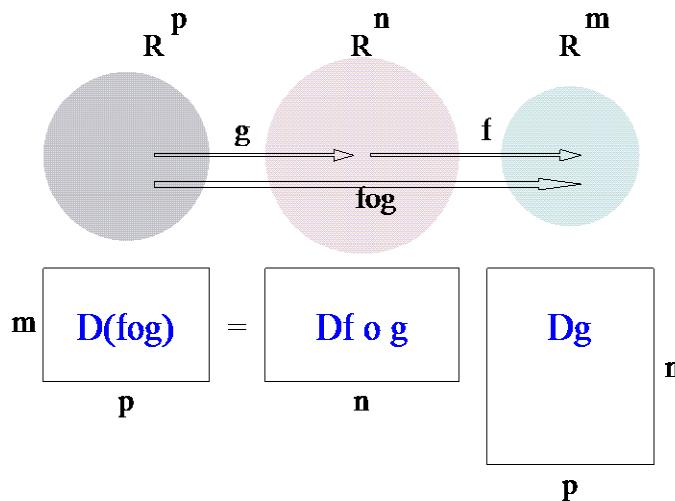
Ist der Wertebereich von g eine Teilmenge des Definitionsbereichs von f , so ist die durch

$$(f \circ g)(x) = f(g(x))$$

definierte Verknüpfung $f \circ g$ wieder differenzierbar mit den Ableitungen

$$(f \circ g)'(a) = f'(g(a)) g'(a).$$

Dabei steht auf der rechten Seite das Matrizenprodukt!



Rezept:

äußere Ableitung bilden - einsetzen - innere Ableitung bilden - multiplizieren

Das Bilden der inneren Ableitung (nach g) wird **Nachdifferenzieren** genannt. Der Beweis der Kettenregel mit Hilfe der linearen Approximation ist relativ einfach, wir lassen ihn hier aber weg.

Die Kettenregel ist eines der wichtigsten Werkzeuge für die Berechnung verschiedenster Ableitungen. Komplizierte Funktionen können mit ihrer Hilfe in einfachere "zerlegt" werden.

Beispiel 7: Die Produktabbildung

p von R^2 nach R ist gegeben durch

$$p(x, y) = xy.$$

Vergleichen wir die explizite Gleichung

$$p(x_0 + h_1, y_0 + h_2) = x_0 y_0 + y_0 h_1 + x_0 h_2 + h_1 h_2$$

mit der theoretischen Formel

$$p(x_0 + h_1, y_0 + h_2) = p(x_0, y_0) + p'(x_0, y_0) \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + o(h_1, h_2)$$

so sehen wir:

$$p'(x, y) = [y, x].$$

Mit der Kettenregel folgt nun sofort die Produktregel:

$$(f g)' = (p \circ (f, g))' = (p' \circ (f, g)) (f', g') = g f' + f g' .$$

Analog geht es in höheren Dimensionen mit dem Skalarprodukt (und im R_3 sogar mit dem Vektorprodukt, aber das wird etwas komplizierter).

Beliebige Exponenten und Basen

Für Potenzfunktionen x^c mit **konstanten nichtnegativen Exponenten** c liefert die Kettenregel die Ableitung

$$D(x^c) = D(e(c \ln(x))) = e(c \ln(x)) \frac{c}{x} = \frac{c}{x} x^c = c x^{(c-1)} ,$$

während c^x bei **konstanter Basis** c die folgende Ableitung hat:

$$D(c^x) = D(e(x \ln(c))) = e(x \ln(c)) \ln(c) = c^x \ln(c).$$

Inversionsregel

Ist f invertierbar und differenzierbar, so auch die Umkehrfunktion g mit $f(x) = y \Leftrightarrow x = g(y)$, und es gilt

$$Dg(x) = Df(g(x))^{(-1)}.$$

Auch diese Gleichung, bei der die rechte Seite die **Inverse der Matrix** $Df(g(x))$ bedeutet, folgt unmittelbar aus der Kettenregel:

Wegen $f \circ g(x) = x$ ist $D(f \circ g)(x)$ die Einheitsmatrix.

(Leider kann man nicht so einfach die Differenzierbarkeit der Umkehrfunktion einer differenzierbaren Funktion beweisen; aber das ist für die Praxis auch weniger wichtig.)

Ableitungen von Umkehrfunktionen

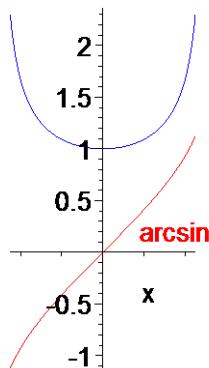
Mit der Inversionsregel erhält man viele wichtige Formeln für das Differenzieren (und später auch das Integrieren) gängiger Funktionen, z.B.

Funktion	Ableitung	Umkehrfunktion	Ableitung
x^c	$c x^{(c-1)}$	$x^{\left(\frac{1}{c}\right)}$	$\frac{1}{c} x^{\left(\frac{1}{c}-1\right)}$
e^x	e^x	$\ln(x)$	$\frac{1}{x}$
$\sin(x)$	$\cos(x) = \sqrt{1 - \sin(x)^2}$	$\arcsin(x)$	$\frac{1}{\sqrt{1-x^2}}$
$\cos(x)$	$-\sin(x) = -\sqrt{1 - \cos(x)^2}$	$\arccos(x)$	$\frac{-1}{\sqrt{1-x^2}}$
$\tan(x)$	$1 + \tan(x)^2$	$\arctan(x)$	$\frac{1}{1+x^2}$
$\cot(x)$	$-1 - \cot(x)^2$	$\text{arccot}(x)$	$-\frac{1}{1+x^2}$

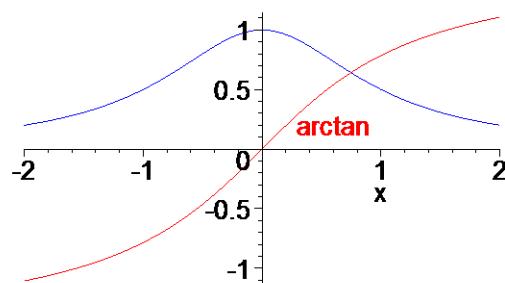
sowie analoge Formeln für die "hyperbolischen" Funktionen $\sinh(x)$, $\cosh(x)$ usw.

Beachten Sie, daß $-\arcsin(x)$ und $\arccos(x)$ zwar die gleiche Ableitung haben, aber nicht die selbe Funktion darstellen, sondern sich um die additive Konstante $\frac{\pi}{2}$ unterscheiden:

$$\arccos(x) = \frac{\pi}{2} - \arcsin(x), \text{ und analog } \text{arccot}(x) = \frac{\pi}{2} - \arctan(x).$$



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



$$\arctan(x), \arctan'(x) = \frac{1}{1+x^2}$$