Kettenregel:

Seien $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ offene Mengen, $x_0 \in U$, $y_0 \in V$, es seien $f: U \to \mathbb{R}^m$, $g: V \to \mathbb{R}^k$ Abbildungen, es sei $y_0 = f(x_0)$, f sei in x_0 und g in y_0 differenzierbar.

Dann ist auch $(g \circ f): U \to \mathbb{R}^k$ in x_0 differenzierbar, und es gilt:

$$D(g \circ f)(x_0) := Dg(y_0) \circ Df(x_0)$$

Um die Abschätzungen im Beweis dieses zentralen Satzes übersichtlich durchführen zu können, benötigen wir den Begriff der ***Matrix-Norm.***

Einschub zu Matrix-Normen

Ist $A \in M_{m \times n}(\mathbb{R})$ eine reelle Matrix mit m Zeilen und n Spalten, so setzen wir

$$\|A\| := \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^{\ 2}}\ ,$$

in direkter Verallgemeinerung der euklidischen Norm eines Vektors. Dann gilt für $A \in M_{m \times n}(\mathbb{R})$, $B \in M_{n \times k}(\mathbb{R})$, $x \in \mathbb{R}^n$

$$\|AB\| \le \sqrt{mk}\, \|A\| \|B\|,\ \|Ax\| \le \sqrt{m}\, \|A\| \|x\|$$

Beweis:

Das Matrizenprodukt $C=AB$ entsteht durch Multiplikation der Zeilen von A mit den Spalten von B. Wir haben also $c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}$. Dies lässt sich interpretieren als Skalarprodukt $\left\langle \begin{pmatrix} a_{i1} \\ \vdots \\ a_{in} \end{pmatrix} \begin{pmatrix} b_{1k} \\ \vdots \\ b_{nk} \end{pmatrix} \right\rangle$. Damit lässt sich die Cauchy-Schwarzsche Ungleichung anwenden:

$$c_{ik}^{\ 2} = \left\langle \begin{pmatrix} a_{i1} \\ \vdots \\ a_{in} \end{pmatrix} \begin{pmatrix} b_{1k} \\ \vdots \\ b_{nk} \end{pmatrix} \right\rangle^2 \le \left(\sum_{j=1}^{n} a_{ij}^{\ 2} \right) \left(\sum_{j=1}^{n} b_{jk}^{\ 2} \right) \le \|A\|^2 \|B\|^2 .$$

Durch Aufsummieren aller $c_{ik}^{\ 2}$ erhalten wir das gewünschte Ergebnis. Analog erhalten wir die zweite Ungleichung.

Es gibt viele verschiedene Möglichkeiten, die Norm einer Matrix zu definieren; wir hatten ja die verschiedenen Varianten bei der Definition der Norm eines Vektors.

Nun zum Beweis der Kettenregel:

Sei $\varepsilon > 0$ vorgegeben. Wir müssen ein $\delta > 0$ finden, so dass

$$\left\| (g \circ f)(x) - (g \circ f)(x_0) - \left(Dg\left(f(x_0)\right) \circ Df(x_0) \right)(x - x_0) \right\| \le \varepsilon \|x - x_0\|$$

Der Ausdruck links vom Ungleichheitszeichen ist gleich

$$\left\| \begin{array}{l} g(f(x)-g(f(x_0))-Dg(f(x_0))(f(x)-f(x_0)) \\ +Dg(f(x_0))((f(x)-f(x_0)-Df(x_0)(x-x_0)) \end{array} \right\| \leq$$

(*) $\|g(f(x)-g(f(x_0))-Dg(f(x_0))(f(x)-f(x_0))\|+\|Dg(f(x_0))((f(x)-f(x_0)-Df(x_0)(x-x_0))\| \leq$

Aufgrund der Differenzierbarkeit von g in $y_0 = f(x_0)$ folgt nun, dass der erste Summand in obiger Summe weiter abgeschätzt werden kann: Zu $\varepsilon_1 > 0$ gibt es $\delta_1 > 0$, so dass für $y \in V$ und $\|y-y_0\| < \delta_1$ gilt: $\|g(y)-g(y_0)-Dg(y_0)(y-y_0)\| \leq \varepsilon_1 \|y-y_0\|$. Wir benötigen aber eine Abschätzung mit $\|x-x_0\|$ und rechnen daher weiter

$$\|y-y_0\| = \|f(x)-f(x_0)\| = \|f(x)-f(x_0)-Df(x_0)(x-x_0)+Df(x_0)(x-x_0)\| \leq$$

$$\|f(x)-f(x_0)-Df(x_0)(x-x_0)+Df(x_0)(x-x_0)\| \leq \|f(x)-f(x_0)-Df(x_0)(x-x_0)\|+\|Df(x_0)(x-x_0)\| \leq$$

Den ersten dieser beiden Summanden können wir über die Differenzierbarkeit von f in x_0 weiter abschätzen: zu $\varepsilon_2 > 0$ gibt es $\delta_2 > 0$, so dass für $x \in U$ und $\|x-x_0\| < \delta_2$:

$$\|f(x)-f(x_0)-Df(x_0)(x-x_0)\| \leq \varepsilon_2 \|x-x_0\|$$

Andererseits ist $\|Df(x_0)(x-x_0)\| \leq \sqrt{m}\,\|Df(x_0)\|\|x-x_0\|$, so dass wir auch den 2. Summanden abgeschätzt haben. (Hier wurde das erste Mal eine Matrixnorm benutzt.) Auf Grund der Stetigkeit von f in x_0 gibt es zu $\varepsilon_3 := \delta_1 > 0$ ein $\delta_3 > 0$, so dass für $x \in U$ und $\|x-x_0\| < \delta_3$: $\|y-y_0\| = \|f(x)-f(x_0)\| < \varepsilon_3 = \delta_1$. Setzen wir nun $\delta := \min\{\delta_2, \delta_3\}$, so können wir unsere Teilergebnisse zusammenfassen zu:

Ist $x \in U$ und $\|x-x_0\| < \delta$, so gilt

$$\|g(f(x)-g(f(x_0))-Dg(f(x_0))(f(x)-f(x_0)\| \leq \varepsilon_1 \|f(x)-f(x_0)\| \leq \varepsilon_1 \left(\varepsilon_2 + \sqrt{m}\,\|Df(x_0)\|\right)\|x-x_0\|$$

Damit sind wir mit dem ersten Summanden in (*) durch. Für den zweiten gilt:

$\|Dg(f(x_0))((f(x)-f(x_0)-Df(x_0)(x-x_0))\| \leq \sqrt{k}\,\|Dg(f(x_0))\|\|(f(x)-f(x_0)-Df(x_0)(x-x_0)\| \leq$
$\varepsilon_2 \sqrt{k}\,\|Dg(f(x_0))\|\|x-x_0\|$ falls $x \in U$ und $\|x-x_0\| < \delta$, insgesamt also

Ist $x \in U$ und $\|x-x_0\| < \delta_4$, so gilt

$$\|(g \circ f)(x)-(g \circ f)(x_0)-\left(Dg(f(x_0)) \circ Df(x_0)\right)(x-x_0)\| \leq$$

$$\left(\varepsilon_1\left(\varepsilon_2 + \sqrt{m}\,\|Df(x_0)\|\right)+\varepsilon_2 \sqrt{k}\,\|Dg(f(x_0))\|\right)\|x-x_0\|$$

Wir sind nun fertig, wenn wir den Ausdruck $\left(\varepsilon_1\left(\varepsilon_2 + \|Df(x_0)\|\right)+\varepsilon_2 \|Dg(f(x_0))\|\right)$ kleiner als ε machen können. Dies ist offenbar der Fall, wenn wir $\varepsilon_2 = \dfrac{\min\left\{1, \dfrac{\varepsilon}{\sqrt{k}\,\|Dg(f(x_0))\|}\right\}}{2}$ und

$\varepsilon_1 = \dfrac{\varepsilon}{2\left(1+\sqrt{m}\,\|Df(x_0)\|\right)}$ wählen, denn dann gilt

$$\left(\varepsilon_1\left(\varepsilon_2+\sqrt{m}\,\|Df(x_0)\|\right)+\varepsilon_2\sqrt{k}\,\|Dg(f(x_0))\|\right) \le \frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon\,.$$

Bemerkung: Einfacher geht’s nicht.