Proof of the chain rule

Theorem (Chain Rule)**.** *If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{x}_0 \in \mathbb{R}^n$ and $g : \mathbb{R}^m \to \mathbb{R}^l$ is differentiable at $\mathbf{y}_0 = f(\mathbf{x}_0) \in \mathbb{R}^m$, then $g \circ f : \mathbb{R}^n \to \mathbb{R}^l$ is differentiable at $\mathbf{x}_0$, with derivative*

$$\mathbf{D}(g \circ f)(\mathbf{x}_0) = \mathbf{D}g(f(\mathbf{x}_0))\mathbf{D}f(\mathbf{x}_0).$$

Proof. We must show that

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{\|g(f(\mathbf{x})) - g(f(\mathbf{x}_0)) - \mathbf{D}g(f(\mathbf{x}_0))\mathbf{D}f(\mathbf{x}_0)\,(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0$$

which means we must show that, for all $\epsilon > 0$ there exists a $\delta > 0$ such that $\|\mathbf{x} - \mathbf{x}_0\| < \delta$ implies that

$$\frac{\|g(f(\mathbf{x})) - g(f(\mathbf{x}_0)) - \mathbf{D}g(f(\mathbf{x}_0))\mathbf{D}f(\mathbf{x}_0)\,(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} < \epsilon. \tag{1}$$

Since $\mathbf{D}f(\mathbf{x}_0)$ and $\mathbf{D}g(\mathbf{y}_0)$ are linear maps, there exist constants $C_{Df} \in \mathbb{R}$ and $C_{Dg} \in \mathbb{R}$ such that

$$\|\mathbf{D}f(\mathbf{x}_0)\mathbf{v}\| \le C_{Df}\,\|\mathbf{v}\|, \quad \text{for all } \mathbf{v} \in \mathbb{R}^n, \tag{2}$$

and

$$\|\mathbf{D}g(\mathbf{y}_0)\mathbf{w}\| \le C_{Dg}\,\|\mathbf{w}\|, \quad \text{for all } \mathbf{w} \in \mathbb{R}^m. \tag{3}$$

We will also use the fact that, since f is differentiable at $\mathbf{x}_0$, then f is continuous there.

Let $\epsilon > 0$ be given. By differentiability of g, we can choose a $\delta_{g'} > 0$ so that

$$\|\mathbf{y} - \mathbf{y}_0\| < \delta_{g'} \implies \frac{\|g(\mathbf{y}) - g(\mathbf{y}_0) - \mathbf{D}g(\mathbf{y}_0)(\mathbf{y} - \mathbf{y}_0)\|}{\|\mathbf{y} - \mathbf{y}_0\|} < \frac{\epsilon}{3C_{Df}}. \tag{4}$$

Then, by *continuity* of f, we can choose a $\delta_f > 0$ so that

$$\|\mathbf{x} - \mathbf{x}_0\| < \delta_f \implies \|f(\mathbf{x}) - f(\mathbf{x}_0)\| = \|\mathbf{y} - \mathbf{y}_0\| < \delta_{g'}, \tag{5}$$

using $\delta_{g'}$ as the ϵ_f in the definition of continuity for f at $\mathbf{x}_0$. Finally, by *differentiability* of f, we choose $\delta_{f'} > 0$ so that

$$\|\mathbf{x} - \mathbf{x}_0\| < \delta_{f'} \implies \frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{D}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} < \min\left(\frac{\epsilon}{3C_{Dg}}, C_{Df}\right). \tag{6}$$

Now let $\delta = \min(\delta_{g'}, \delta_f, \delta_{f'})$. We will show that when $\|\mathbf{x} - \mathbf{x}_0\| < \delta$, (1) is satisfied. Indeed,

$$\begin{aligned}
&\frac{\|g(f(\mathbf{x})) - g(f(\mathbf{x}_0)) - \mathbf{D}g(f(\mathbf{x}_0))\mathbf{D}f(\mathbf{x}_0)\,(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} \\
&= \frac{\|g(\mathbf{y}) - g(\mathbf{y}_0) - \mathbf{D}g(\mathbf{y}_0)(\mathbf{y} - \mathbf{y}_0) + \mathbf{D}g(\mathbf{y}_0)(\mathbf{y} - \mathbf{y}_0) - \mathbf{D}g(\mathbf{y}_0)\mathbf{D}f(\mathbf{x}_0)\,(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} \\
&\le \frac{\|g(\mathbf{y}) - g(\mathbf{y}_0) - \mathbf{D}g(\mathbf{y}_0)(\mathbf{y} - \mathbf{y}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} + \frac{\|\mathbf{D}g(\mathbf{y}_0)(\mathbf{y} - \mathbf{y}_0) - \mathbf{D}g(\mathbf{y}_0)\mathbf{D}f(\mathbf{x}_0)\,(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} \\
&\le \frac{\epsilon}{4C_{Df}} \frac{\|\mathbf{y} - \mathbf{y}_0\|}{\|\mathbf{x} - \mathbf{x}_0\|} + C_{Dg} \frac{\|(\mathbf{y} - \mathbf{y}_0) - \mathbf{D}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} \\
&= \frac{\epsilon}{3C_{Df}} \frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{D}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \mathbf{D}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} + C_{Dg} \frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{D}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} \\
&\le \frac{\epsilon}{3C_{Df}} \frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{D}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} + \frac{\epsilon}{3C_{Df}} \frac{\|\mathbf{D}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} + C_{Dg} \frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - \mathbf{D}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} \\
&\le \frac{\epsilon C_{Df}}{3C_{Df}} + \frac{\epsilon C_{Df}}{3C_{Df}} \frac{\|\mathbf{x} - \mathbf{x}_0\|}{\|\mathbf{x} - \mathbf{x}_0\|} + \frac{\epsilon C_{Dg}}{3C_{Dg}} \\
&= \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.
\end{aligned}$$

In the first line we add and subtract the same term from the numerator, then use the triangle inequality, then use (3), (4) and (5), then add and subtract again, then use the triangle inequality again, and finally use (2) and (6). $\square$