12. Chain rule

Theorem 12.1 (Chain Rule)**.** *Let $U \subset \mathbb{R}^n$ and let $V \subset \mathbb{R}^m$ be two open subsets. Let $f\colon U \longrightarrow V$ and $g\colon V \longrightarrow \mathbb{R}^p$ be two functions. If f is differentiable at P and g is differentiable at $Q = f(P)$, then $g \circ f\colon U \longrightarrow \mathbb{R}^p$ is differentiable at P, with derivative:*

$$D(g \circ f)(P) = (D(g)(Q))(D(f)(P)).$$

It is interesting to untwist this result in specific cases. Suppose we are given

$$f\colon \mathbb{R} \longrightarrow \mathbb{R}^2 \qquad \text{and} \qquad g\colon \mathbb{R}^2 \longrightarrow \mathbb{R}.$$

So $f(x) = (f_1(x), f_2(x))$ and $w = g(y, z)$. Then

$$Df(P) = \begin{pmatrix} \frac{df_1}{dx}(P) \\ \frac{df_2}{dx}(P) \end{pmatrix} \qquad \text{and} \qquad Dg(Q) = (\frac{\partial g}{\partial y}(Q), \frac{\partial g}{\partial z}(Q)).$$

So

$$\frac{d(g \circ f)}{dx} = D(g \circ f)(P) = Dg(Q)Df(P) = \frac{\partial g}{\partial y}(Q)\frac{df_1}{dx}(P) + \frac{\partial g}{\partial z}(Q)\frac{df_2}{dx}(P).$$

Example 12.2. *Suppose that $f(x) = (x^2, x^3)$ and $g(y, z) = yz$. If we apply the chain rule, we get*

$$D(g \circ f)(x) = z(2x) + y(3x^2) = 5x^4.$$

On the other hand $(g \circ f)(x) = x^5$, and of course

$$\frac{dx^5}{dx} = 5x^4.$$

Now suppose that

$$f\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2 \qquad \text{and} \qquad g\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$$

So $f(x, y) = (f_1(x, y), f_2(x, y))$ and $w = g(u, v)$. Then

$$Df(P) = \begin{pmatrix} \frac{\partial f_1}{\partial x}(P) & \frac{\partial f_2}{\partial x}(P) \\ \frac{\partial f_2}{\partial x}(P) & \frac{\partial f_2}{\partial x}(P) \end{pmatrix} \qquad \text{and} \qquad Dg(Q) = (\frac{\partial g}{\partial u}(Q), \frac{\partial g}{\partial v}(Q)).$$

In this case

$$\begin{aligned} D(g \circ f) &= (\frac{\partial(g \circ f)}{\partial x}, \frac{\partial(g \circ f)}{\partial y}) \\ &= (\frac{\partial g}{\partial u}(Q)\frac{\partial f_1}{\partial x}(P) + \frac{\partial g}{\partial v}(Q)\frac{\partial f_2}{\partial x}(P), \frac{\partial g}{\partial u}(Q)\frac{\partial f_1}{\partial y}(P) + \frac{\partial g}{\partial v}(Q)\frac{\partial f_2}{\partial y}(P)). \\ &= (\frac{\partial g}{\partial u}(Q)\frac{\partial u}{\partial x}(P) + \frac{\partial g}{\partial v}(Q)\frac{\partial v}{\partial x}(P), \frac{\partial g}{\partial u}(Q)\frac{\partial u}{\partial y}(P) + \frac{\partial g}{\partial v}(Q)\frac{\partial v}{\partial y}(P)) \\ &= (\frac{\partial g}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial g}{\partial v}\frac{\partial v}{\partial x}, \frac{\partial g}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial g}{\partial v}\frac{\partial v}{\partial y}), \end{aligned}$$

since $u = f_1(x, y)$ and $v = f_2(x, y)$. Notice that in the last line we were a bit sloppy and dropped P and Q.

If we split this vector equation into its components we get

$$\begin{aligned}\frac{\partial(g \circ f)}{\partial x} &= \frac{\partial g}{\partial u}(Q)\frac{\partial f_1}{\partial x}(P) + \frac{\partial g}{\partial v}(Q)\frac{\partial f_2}{\partial x}(P)\\ \frac{\partial(g \circ f)}{\partial y} &= \frac{\partial g}{\partial u}(Q)\frac{\partial f_1}{\partial y}(P) + \frac{\partial g}{\partial v}(Q)\frac{\partial f_2}{\partial y}(P).\end{aligned}$$

Again, we could replace f_1 by u and f_2 by v in these equations, and maybe even drop P and Q.

Example 12.3. *Suppose that* $f(x, y) = (\cos(xy), e^{x-y})$ *and* $g(u, v) = u^2 \sin v$. *If we apply the chain rule, we get*

$$\begin{aligned}D(g \circ f)(x) &= (2u \sin v(-y \sin xy) + u^2 \cos v(e^{x-y}), -2u \sin vx \sin xy - u^2 \cos ve^{x-y}\\ &= (2 \cos(xy) \sin(e^{x-y})(-y \sin xy) + \cos^2(xy) \cos(e^{x-y})e^{x-y}, \dots).\end{aligned}$$

In general, the (i, k) entry of $D(g \circ f)(P)$, that is

$$\frac{\partial(g \circ f)_i}{\partial x_k}$$

is given by the dot product of the ith row of $Dg(Q)$ and the kth column of $Df(P)$,

$$\frac{\partial(g \circ f)_i}{\partial x_k} = \sum_{j=1}^{m} \frac{\partial g_i}{\partial y_j}(Q)\frac{\partial f_j}{\partial x_i}(P).$$

If $z = (g \circ f)(P)$, then we get

$$\frac{\partial z_i}{\partial x_k} = \sum_{j=1}^{m} \frac{\partial z_i}{\partial y_j}(Q)\frac{\partial y_j}{\partial x_i}(P).$$

We can use the chain rule to prove some of the simple rules for derivatives. Suppose that we have

$$f\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m \qquad \text{and} \qquad g\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m.$$

Suppose that f and g are differentiable at P. What about $f + g$? Well there is a function

$$a\colon \mathbb{R}^{2m} \longrightarrow \mathbb{R}^m,$$

which sends $(\vec{u}, \vec{v}) \in \mathbb{R}^m \times \mathbb{R}^m$ to the sum $\vec{u} + \vec{v}$. In coordinates $(u_1, u_2, \dots, u_m, v_1, v_2, \dots, v_m)$,

$$a(u_1, u_2, \dots, u_m, v_1, v_2, \dots, v_m) = (u_1 + v_1, u_2 + v_2, \dots, u_m + v_m).$$

Now a is differentiable (it is a polynomial, linear even). There is function

$$h\colon \mathbb{R}^n \longrightarrow \mathbb{R}^{2m},$$

which sends Q to $(f(Q), g(Q))$. The composition $a \circ h\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is the function we want to differentiate, it sends P to $f(P) + g(P)$. The chain rule says that that the function is differentiable at P and

$$D(f+g)(P) = Df(P) + Dg(P).$$

Now suppose that $m = 1$. Instead of a, consider the function

$$m\colon \mathbb{R}^2 \longrightarrow \mathbb{R},$$

given by $m(x, y) = xy$. Then m is differentiable, with derivative

$$Dm(x, y) = (y, x).$$

So the chain rule says the composition of h and m, namely the function which sends P to the product $f(P)g(P)$ is differentiable and the derivative satisfies the usual rule

$$D(fg)(P) = g(P)D(f)(P) + f(P)D(g)(P).$$

Here is another example of the chain rule, suppose

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta. \end{aligned}$$

Then

$$\begin{aligned} \frac{\partial f}{\partial r} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} \\ &= \frac{\partial f}{\partial x}\cos\theta + \frac{\partial f}{\partial y}\sin\theta. \end{aligned}$$

Similarly,

$$\begin{aligned} \frac{\partial f}{\partial \theta} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta} \\ &= -\frac{\partial f}{\partial x} r\sin\theta + \frac{\partial f}{\partial y} r\cos\theta. \end{aligned}$$

We can rewrite this as

$$\begin{pmatrix} \frac{\partial}{\partial r} \\ \frac{\partial}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{pmatrix}$$

Now the determinant of

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{pmatrix}$$

is

$$r(\cos^2\theta + \sin^2\theta) = r.$$

So if $r \neq 0$, then we can invert the matrix above and we get

$$\begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{pmatrix} = \frac{1}{r} \begin{pmatrix} r\cos\theta & -\sin\theta \\ r\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial r} \\ \frac{\partial}{\partial \theta} \end{pmatrix}$$

We now turn to a proof of the chain rule. We will need:

Lemma 12.4. *Let $A \subset \mathbb{R}^n$ be an open subset and let $f\colon A \longrightarrow \mathbb{R}^m$ be a function.*

If f is differentiable at P, then there is a constant $M \geq 0$ and $\delta > 0$ such that if $\|\overrightarrow{PQ}\| < \delta$, then

$$\|f(Q) - f(P)\| < M\|\overrightarrow{PQ}\|.$$

Proof. As f is differentiable at P, there is a constant $\delta > 0$ such that if $\|\overrightarrow{PQ}\| < \delta$, then

$$\frac{\|f(Q) - f(P) - Df(P)\overrightarrow{PQ}\|}{\|\overrightarrow{PQ}\|} < 1.$$

Hence

$$\|f(Q) - f(P) - Df(P)\overrightarrow{PQ}\| < \|\overrightarrow{PQ}\|.$$

But then

$$\begin{aligned} \|f(Q) - f(P)\| &= \|f(Q) - f(P) - Df(P)\overrightarrow{PQ} + Df(P)\overrightarrow{PQ}\| \\ &\leq \|f(Q) - f(P) - Df(P)\overrightarrow{PQ}\| + \|Df(P)\overrightarrow{PQ}\| \\ &\leq \|\overrightarrow{PQ}\| + K\|\overrightarrow{PQ}\| \\ &= M\|\overrightarrow{PQ}\|, \end{aligned}$$

where $M = 1 + K$. □

Proof of (12.1). Let's fix some notation. We want the derivative at P. Let $Q = f(P)$. Let P' be a point in U (which we imagine is close to P). Finally, let $Q' = f(P')$ (so if P' is close to P, then we expect Q' to be close to Q).

The trick is to carefully define an auxiliary function $G\colon V \longrightarrow \mathbb{R}^p$,

$$G(Q') = \begin{cases} \frac{g(Q') - g(Q) - Dg(Q)(\overrightarrow{QQ'})}{\|\overrightarrow{QQ'}\|} & \text{if } Q' \neq Q \\ \vec{0} & \text{if } Q' = Q. \end{cases}$$

Then G is continuous at $Q = f(P)$, as g is differentiable at Q. Now,

$$\frac{(g \circ f)(P') - (g \circ f)(P) - Dg(Q)Df(P)(\overline{P}P')}{\|\overline{P}P'\|}$$

$$= Dg(Q)\frac{f(P') - f(P) - Df(P)(\overrightarrow{PP'})}{\|\overline{P}P'\|} + G(f(P'))\frac{\|f(P') - f(P)\|}{\|\overrightarrow{PP'}\|}.$$

As P' approaches P, note that

$$\frac{f(P') - f(P) - Df(P)(\overrightarrow{PP'})}{\|\overline{P}P'\|},$$

and $G(P')$ both approach zero and

$$\frac{\|f(P') - f(P)\|}{\|\overline{P}P'\|} \leq M.$$

So then

$$\frac{(g \circ f)(P') - (g \circ f)(P) - Dg(Q)Df(P)(\overrightarrow{PP'})}{\|\overline{P}P'\|},$$

approaches zero as well, which is what we want. □