THE CHAIN RULE
BASED ON LECTURE NOTES BY JAMES MCKERNAN

Theorem 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 $g = g(y, z)$. 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 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

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

In general, if $f = f(x_1, \dots, x_n)$ and $g = g(y_1, \dots, y_n)$ then 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 $Df(p)$,

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

If $y = f(x)$ and $z = g(y)$ then we get

$$\frac{\partial z_i}{\partial x_k} = \sum_{j=1}^{m} \frac{\partial z_i}{\partial y_j}\frac{\partial y_j}{\partial x_k}.$$

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 3. *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 $\delta > 0$ such that if $\|q - p\| < \delta$, then

$$\|f(q) - f(p)\| < (K + 1)\|q - p\|,$$

where K is the Frobenius norm of $Df(p)$.

Proof. As f is differentiable at p, there is a constant $\delta > 0$ such that if $\|q - p\| < \delta$, then

$$\frac{\|f(q) - f(p) - Df(p)(q - p)\|}{\|q - p\|} < 1.$$

Hence

$$\|f(q) - f(p) - Df(p)(q - p)\| < \|q - p\|.$$

But then

$$\begin{aligned} \|f(q) - f(p)\| &= \|f(q) - f(p) - Df(p)(q - p) + Df(p)(q - p)\| \\ &\leq \|f(q) - f(p) - Df(p)(q - p)\| + \|Df(p)(q - p)\| \\ &\leq \|(q - p)\| + K\|(q - p)\| \\ &= (K + 1)\|(q - p)\|, \end{aligned}$$

□

Proof of (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 $G\colon V \longrightarrow \mathbb{R}^p$,

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

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

$$\begin{aligned} &\frac{(g \circ f)(p') - (g \circ f)(p) - Dg(q)Df(p)(p' - p)}{\|p' - p\|} \\ &= \frac{Dg(q)(f(p') - f(p)) - Dg(q)(q' - q) + g(q') - g(q) - Dg(q)Df(p)(p' - p)}{\|p' - p\|} \\ &= Dg(q)\frac{f(p') - f(p) - Df(p)(p' - p)}{\|p' - p\|} + \frac{g(q') - g(q) - Dg(q)(q' - q)}{\|p' - p\|} \\ &= Dg(q)\frac{f(p') - f(p) - Df(p)(p' - p)}{\|p' - p\|} + G(f(p'))\frac{\|f(p') - f(p)\|}{\|p' - p\|}. \end{aligned}$$

As p' approaches p, note that

$$\frac{f(p') - f(p) - Df(p)(p' - p)}{\|p' - p\|},$$

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

$$\frac{\|f(p') - f(p)\|}{\|p' - p\|} \leq K + 1.$$

So then

$$\frac{(g \circ f)(p') - (g \circ f)(p) - Dg(q)Df(p)(p' - p)}{\|p' - p\|},$$

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