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. $\square$

13. Implicit functions

Consider the curve $y^2 = x$ in the plane $\mathbb{R}^2$,

$$C = \{ (x, y) \in \mathbb{R}^2 \,|\, y^2 = x \}.$$

This is not the graph of a function, and yet it is quite close to the graph of a function.

Given any point on the graph, let's say the point $(2, 4)$, we can always find open intervals U containing 2 and V containing 4 and a smooth function $f\colon U \longrightarrow V$ such that $C \cap (U \times V)$ is the graph of f.

Indeed, take $U = (0, \infty)$, $V = (0, \infty)$ and $f(x) = \sqrt{x}$. In fact, we can do this for any point on the graph, apart from the origin. If it is above the x-axis, the function above works. If the point we are interested in is below the x-axis, replace V by $(0, -\infty)$ and $f(x) = \sqrt{x}$, by $g(x) = -\sqrt{x}$.

How can we tell that the origin is a point where we cannot define an implicit function? Well away from the origin, the tangent line is not vertical but at the origin the tangent line is vertical. In other words, if we consider

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

given by $F(x, y) = y^2 - x$, so that C is the set of points where F is zero, then

$$DF(x, y) = (-1, 2y).$$

The locus where we run into trouble, is where $2y = 0$. Somewhat amazingly this works in general:

Theorem 13.1 (Implicit Function Theorem)**.** *Let $A \subset \mathbb{R}^{n+m}$ be an open subset and let $F\colon A \longrightarrow \mathbb{R}^m$ be a $\mathcal{C}^1$-function. Suppose that*

$$(\vec{a}, \vec{b}) \in S = \{ (\vec{x}, \vec{y}) \in A \,|\, F(\vec{x}, \vec{y}) = \vec{0} \}.$$

Assume that

$$\det \left(\frac{\partial F_i}{\partial y_j} \right) \neq 0.$$

Then we may find open subsets $\vec{a} \in U \subset \mathbb{R}^n$ and $\vec{b} \in V \subset \mathbb{R}^m$, where $U \times V \subset A$ and a function $f\colon U \longrightarrow V$ such that $S \cap (U \times V)$ is the graph of f, that is,

$$F(\vec{x}, \vec{y}) = \vec{0} \qquad \text{if and only if} \qquad \vec{y} = f(\vec{x}).$$

where $\vec{x} \in U$ and $\vec{y} \in V$.

Let's look at an example. Let

$$F\colon \mathbb{R}^3 \longrightarrow \mathbb{R},$$

be the function

$$F(x_1, x_2, y) = x_1^3 x_2 - x_2 y^2 + y^5 + 1.$$

Let

$$S = \{\, (x_1, x_2, y) \in \mathbb{R}^3 \mid F(x_1, x_2, y) = 0 \,\}.$$

Then $(1, 3, -1) \in S$. Let's compute the partial derivatives of F,

$$\begin{aligned}
\frac{\partial F}{\partial x_1}(1, 3, -1) &= 3x_1^2 x_2\Big|_{(1,3,-1)} = 9 \\
\frac{\partial F}{\partial x_2}(1, 3, -1) &= (x_1^3 - y^2)\Big|_{(1,3,-1)} = 0 \\
\frac{\partial F}{\partial y}(1, 3, -1) &= (-2x_2 y + 5y^4)\Big|_{(1,3,-1)} = 11.
\end{aligned}$$

So

$$DF(1, 3, -1) = (9, 0, 11).$$

Now what is important is that the last entry is non-zero (so that the 1×1 matrix (1) is invertible). It follows that we may find open subsets $(1, 3) \in U \subset \mathbb{R}^2$ and $-1 \in V \subset \mathbb{R}$ and a $\mathcal{C}^1$ function $f \colon U \longrightarrow V$ such that

$$F(x_1, x_2, f(x_1, x_2)) = 0.$$

It is not possible to write down an explicit formula for f, but we can calculate the partial derivatives of f.

Define a function

$$G \colon U \longrightarrow \mathbb{R},$$

by the rule

$$G(x_1, x_2) = F(x_1, x_2, f(x_1, x_2)) = 0.$$

On the one hand,

$$\frac{\partial G}{\partial x_1} = 0 \qquad \text{and} \qquad \frac{\partial G}{\partial x_2} = 0.$$

On the other hand, by the chain rule,

$$\frac{\partial G}{\partial x_1} = \frac{\partial F}{\partial x_1}\frac{\partial x_1}{\partial x_1} + \frac{\partial F}{\partial x_2}\frac{\partial x_2}{\partial x_1} + \frac{\partial F}{\partial x_3}\frac{\partial f}{\partial x_1}$$

Now

$$\frac{\partial x_1}{\partial x_1} = 1 \qquad \text{and} \qquad \frac{\partial x_2}{\partial x_1} = 0.$$

So

$$\frac{\partial f}{\partial x_1} = -\frac{\frac{\partial F}{\partial x_1}}{\frac{\partial F}{\partial x_3}}.$$

Similarly

$$\frac{\partial f}{\partial x_2} = -\frac{\frac{\partial F}{\partial x_2}}{\frac{\partial F}{\partial x_3}}.$$

So

$$\frac{\partial f}{\partial x_1}(1,3) = -\frac{\frac{\partial F}{\partial x_1}(1,3,-1)}{\frac{\partial F}{\partial x_3}(1,3,-1)} = -\frac{9}{11},$$

and

$$\frac{\partial f}{\partial x_2}(1,3) = -\frac{\frac{\partial F}{\partial x_2}(1,3,-1)}{\frac{\partial F}{\partial x_3}(1,3,-1)} = -\frac{0}{11} = 0.$$

Definition 13.2. *Let $A \subset \mathbb{R}^n$ be an open subset and let $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ be a function.*

The ***directional derivative*** *of f in the direction of the unit vector $\hat{u}$ is*

$$D_{\hat{u}} f(P) = \lim_{h \to 0} \frac{f(P + h\hat{u}) - f(P)}{h}.$$

If $\hat{u} = \hat{e}_i$ then,

$$D_{\hat{e}_i} f(P) = \frac{\partial f}{\partial x_i}(P),$$

the usual partial derivative.

Proposition 13.3. *If f is differentiable at P then*

$$D_{\hat{u}} f(P) = Df(P) \cdot \hat{u}.$$

Proof. Since A is open, we may find $\delta > 0$ such that the parametrised line

$$r\colon (-\delta, \delta) \longrightarrow A,$$

given by $r(h) = f(P) + h\hat{u}$ is entirely contained in A. Consider the composition of r and f,

$$f \circ r\colon \mathbb{R} \longrightarrow \mathbb{R}.$$

Then

$$\begin{aligned} D_{\hat{u}} f(P) &= \left(\frac{d(f \circ r)}{dh}\right)(0) \\ &= D(r(0)) \cdot Dr(0) \\ &= Df(P) \cdot \hat{u}. \end{aligned}$$

□

Note that we can also write

$$D_{\hat{u}} f(P) = \nabla f(P) \cdot \hat{u}.$$

Note that the directional derivative is largest when

$$\hat{u} = \frac{\nabla f(P)}{\|\nabla f(P)\|},$$

so that the gradient always points in the direction of maximal change (and in fact the magnitude of the gradient, gives the maximum change). Note also that the directional derivative is zero if $\hat{u}$ is orthogonal to the gradient and that the directional derivative is smallest when

$$\hat{u} = -\frac{\nabla f(P)}{\|\nabla f(P)\|}.$$

Proposition 13.4. *If $\nabla f(P) \neq 0$ then the tangent hyperplane Π to the hypersurface*

$$S = \{\, Q \in \mathbb{R}^n \mid f(Q) - f(P) = 0 \,\},$$

is the set of all points Q which satisfy the equation

$$\nabla f(P) \cdot \overrightarrow{PQ} = 0.$$

Remark 13.5. *If f is $\mathcal{C}^1$, then f is the graph of some function, locally about P.*

Proof. By definition, the point Q belongs to the tangent hyperplane if and only if there is a curve

$$r \colon (-\delta, \delta) \longrightarrow S,$$

such that

$$r(0) = P \qquad \text{and} \qquad r'(0) = \overrightarrow{PQ}.$$

Now, since $r(h) \in S$ for all $h \in (-\delta, \delta)$, we have $F(r(h)) = 0$. So

$$\begin{aligned} 0 &= \frac{dF(r(h))}{dh}(0) \\ &= \nabla F(r(0)) \cdot r'(0) \\ &= \nabla F(P) \cdot \overrightarrow{PQ}. \end{aligned}$$

□