# Chain Rule

In this activity, we introduce the multi-variable chain rule, and we use it to compute derivatives of compositions of functions.

## The Chain Rule

Let's begin by recalling the chain rule from single variable calculus. If we have differentiable functions $f$ and $g$, then we can compute the derivative of the composition as $(f \circ g)'(x) = f'(g(x))g'(x)$.

**Example 1.** *Let $f(x) = \sin(x)$ and $g(x) = x^2$. Then $f'(x) = \cos(x)$, $g'(x) = 2x$, and we can differentiate the composition $(f \circ g)(x) = \sin(x^2)$ using the chain rule:*

$$(f \circ g)'(x) = f'(g(x))g'(x)$$
$$= \cos(g(x)) \cdot 2x$$
$$= \cos(x^2) \cdot 2x.$$

*We can also use the chain rule to differentiate the composition $(g \circ f)(x) = \sin^2(x)$.*

$$(g \circ f)'(x) = g'(f(x))f'(x)$$
$$= 2(f(x))\cos(x)$$
$$= 2\sin(x)\cos(x)$$

The multi-variable chain rule is similar, with the derivative matrix taking the place of the single variable derivative, so that the chain rule will involve matrix multiplication. We also need to pay extra attention to whether the composition of functions is even defined.

**Theorem 1.** *Suppose $\mathbf{f} : Y \subset \mathbb{R}^p \to \mathbb{R}^n$ and $\mathbf{g} : X \subset \mathbb{R}^m \to \mathbb{R}^p$ are defined on open sets $Y \subset \mathbb{R}^p$ and $X \subset \mathbb{R}^m$, respectively. Suppose that $\mathbf{g}(X) \subset Y$, so the image of $\mathbf{g}$ is contained in the domain of $\mathbf{f}$. Suppose further that $\mathbf{g}$ is differentiable at some point $\mathbf{x}_0 \in X$, and that $\mathbf{f}$ is differentiable at $\mathbf{y}_0 = \mathbf{g}(\mathbf{x}_0) \in Y$.*

*Then the composition $\mathbf{f} \circ \mathbf{g}$ is differentiable at $\mathbf{x}_0$, and*

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

Learning outcomes:
Author(s):

Although the conditions sound complicated, essentially they're just requiring that all of the derivatives mentioned actually exist. Note the similarities to the single variable chain rule.

**Proof**     super great proof of chain rule                                                   ∎

# A Special Case

We'll now consider a special case of the chain rule, when we have a composition $f \circ g$ of functions $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{g} : \mathbb{R} \to \mathbb{R}^n$. Note that $f$ is a scalar function, and we can think of $\mathbf{g}$ as a curve in $\mathbb{R}^n$.

Let's look at what the chain rule tells us in this case. For any $x_0 \in \mathbb{R}$, we have

$$D(f \circ \mathbf{g})(x_0) = Df(\mathbf{g}(x_0))D\mathbf{g}(x_0).$$

Writing $\mathbf{g}(x) = (g_1(x), ..., g_n(x))$ in terms of its components, we have

$$D\mathbf{g} = \begin{pmatrix} \partial g_1/\partial x \\ \vdots \\ \partial g_n/\partial x \end{pmatrix}.$$

Since $\mathbf{g}$ only has one input variable, we can rewrite this as

$$D\mathbf{g}(x) = \begin{pmatrix} g_1'(x) \\ \vdots \\ g_n'(x) \end{pmatrix}.$$

Now that we've sorted out $D\mathbf{g}$, let's consider $Df$. Since $f$ is a scalar-valued function, $Df$ will consist of only one row,

$$Df = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} & \cdots & \dfrac{\partial f}{\partial x_n} \end{pmatrix}.$$

For $Df(\mathbf{g}(x_0))$, we would evaluate these partial derivatives at $\mathbf{g}(x_0)$:

$$Df(\mathbf{g}(x_0)) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(\mathbf{g}(x_0)) & \cdots & \dfrac{\partial f}{\partial x_n}(\mathbf{g}(x_0)) \end{pmatrix}.$$

Now let's turn our attention back to the composition $f \circ \mathbf{g}$. Putting together our results from above, we have

$$D(f \circ \mathbf{g})(x_0) = Df(\mathbf{g}(x_0))D\mathbf{g}(x_0)$$

$$= \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(\mathbf{g}(x_0)) & \cdots & \dfrac{\partial f}{\partial x_n}(\mathbf{g}(x_0)) \end{pmatrix} \begin{pmatrix} g_1'(x_0) \\ \vdots \\ g_n'(x_0) \end{pmatrix}$$

$$= \left( \left( \dfrac{\partial f}{\partial x_1}(\mathbf{g}(x_0)) \right) g_1'(x_0) + \cdots + \left( \dfrac{\partial f}{\partial x_n}(\mathbf{g}(x_0)) \right) g_n'(x_0) \right)$$

Since $f \circ \mathbf{g} : \mathbb{R} \to \mathbb{R}$ is a single variable function, its derivative matrix at $x_0$ only has one entry, which is $\dfrac{df \circ \mathbf{g}}{dx}(x_0)$. So, we can rewrite the above as

$$\frac{df \circ \mathbf{g}}{dx}(x_0) = \left( \frac{\partial f}{\partial x_1}(\mathbf{g}(x_0)) \right) g_1'(x_0) + \cdots + \left( \frac{\partial f}{\partial x_n}(\mathbf{g}(x_0)) \right) g_n'(x_0).$$

This gives us a special case of the Chain Rule, that can be useful when we have a composition of functions $\mathbb{R} \to \mathbb{R}^n \to \mathbb{R}$.

# Examples

**Example 2.** *Let's consider the functions $\mathbf{f}(x, y) = (x^2 + y^2, xy)$ and $\mathbf{g}(x, y) = (3xy, x - y, 7y^2)$. First, let's decide how we can compose these functions.*

*Which composition(s) exist?*

*Multiple Choice:*

(a) *Neither $\mathbf{f} \circ \mathbf{g}$ nor $\mathbf{g} \circ \mathbf{f}$ exists.*

(b) *$\mathbf{f} \circ \mathbf{g}$ exists, but $\mathbf{g} \circ \mathbf{f}$ does not.*

(c) *$\mathbf{g} \circ \mathbf{f}$ exists, but $\mathbf{f} \circ \mathbf{g}$ does not.* ✓

(d) *$\mathbf{f} \circ \mathbf{g}$ and $\mathbf{g} \circ \mathbf{f}$ both exist.*

*We'll compute the derivative matrix $D(\mathbf{g} \circ \mathbf{f}$ in two ways: using the chain rule, and directly.*

*Let's begin by using the chain rule. We'll have*

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

*so we'll start by computing the derivative matrices $D\mathbf{f}$ and $D\mathbf{g}$.*

$$D\mathbf{f}(x, y) = \begin{pmatrix} \boxed{2x} & \boxed{2y} \\ \boxed{y} & \boxed{x} \end{pmatrix}$$

$$D\mathbf{g}(x, y) = \begin{pmatrix} \boxed{3y} & \boxed{3x} \\ \boxed{1} & \boxed{-1} \\ \boxed{0} & \boxed{14y} \end{pmatrix}$$

*Now, for $D\mathbf{g}(\mathbf{f}(\mathbf{x}))$, we need to input $\mathbf{f}(x, y)$ into $D\mathbf{g}$.*

$$D\mathbf{g}(\mathbf{f}(x, y)) = D\mathbf{g}(x^2 + y^2, xy)$$

$$= \begin{pmatrix} \boxed{3(xy)} & \boxed{3(x^2 + y^2)} \\ \boxed{1} & \boxed{-1} \\ \boxed{0} & \boxed{14(xy)} \end{pmatrix}$$

*To compute $D(\mathbf{g} \circ \mathbf{f})(x, y)$, we multiply matrices, and obtain*

$$D(\mathbf{g} \circ \mathbf{f})(x, y) = \begin{pmatrix} \boxed{3(xy)} & \boxed{3(x^2 + y^2)} \\ \boxed{1} & \boxed{-1} \\ \boxed{0} & \boxed{14(xy)} \end{pmatrix} \begin{pmatrix} \boxed{2x} & \boxed{2y} \\ \boxed{y} & \boxed{x} \end{pmatrix}$$

$$= \begin{pmatrix} \boxed{9x^2 y + 3y^3} & \boxed{3x^3 + 9xy^2} \\ \boxed{2x - y} & \boxed{2y - x} \\ \boxed{14xy^2} & \boxed{14x^2 y} \end{pmatrix}.$$

*Let's verify our answer, by computing the $D(\mathbf{g} \circ \mathbf{f})(x, y)$ directly, without using the chain rule. We'll begin by finding $\mathbf{g} \circ \mathbf{f}$.*

$$\begin{aligned} (\mathbf{g} \circ \mathbf{f})(x, y) &= \mathbf{g}(f(x, y)) \\ &= \mathbf{g}(x^2 + y^2, xy) \\ &= \boxed{(3(x^2 + y^2)(xy), (x^2 + y^2) - (xy), 7(xy)^2)}. \end{aligned}$$

*This simplifies to $(\mathbf{g} \circ \mathbf{f})(x, y) = (3x^3 y + 3xy^3, x^2 + y^2 - xy, 7x^2 y^2)$. We can then compute the derivative matrix.*

$$(\mathbf{g} \circ \mathbf{f})(x, y) = \begin{pmatrix} \boxed{9x^2 y + 3y^3} & \boxed{3x^3 + 9xy^2} \\ \boxed{2x - y} & \boxed{2y - x} \\ \boxed{14xy^2} & \boxed{14x^2 y} \end{pmatrix}$$

*We see that this gives the same result as using the chain rule.*

**Example 3.** *special case*

**Example 4.** *Consider the function $g : \mathbb{R}^2 \to \mathbb{R}^2$, given by $\mathbf{g}(r, \theta) = (r \cos \theta, r \sin \theta)$. We can think of this function as converting from polar coordinates to Cartesian coordinates. Suppose we have a function $f : \mathbb{R}^2 \to \mathbb{R}$, and we want to know how $f$ changes with respect to polar coordinates. We could find this by computing*

$$D(f \circ \mathbf{g}) = \begin{pmatrix} \dfrac{\partial (f \circ \mathbf{g})}{\partial r} & \dfrac{\partial (f \circ \mathbf{g})}{\partial \theta} \end{pmatrix}.$$

*For specific functions $f$, we could probably compute this derivative matrix directly. However, if we use the chain rule, we can find a general formula for the derivative matrix. The chain rule tells us*

$$D(f \circ \mathbf{g})(r, \theta) = Df(\mathbf{g}(r, \theta))D\mathbf{g}(r, \theta).$$

*We don't know precisely what $Df$ will be, but we can write it in terms of its partial derivatives:*

$$Df = \begin{pmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{pmatrix}.$$

*Given* $\mathbf{g}(r, \theta) = (r\cos\theta, r\sin\theta)$, *we can compute the derivative matrix of* $\mathbf{g}$.

$$D\mathbf{g}(r, \theta) = \begin{pmatrix} \partial g_1/\partial r & \partial g_1/\partial\theta \\ \partial g_2\partial r & \partial g_2/\partial\theta \end{pmatrix}$$

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

*Then, from the chain rule, we have*

$$D(f\circ\mathbf{g})(r, \theta) = Df(\mathbf{g}(r, \theta))D\mathbf{g}(r, \theta)$$

$$= \begin{pmatrix} \dfrac{\partial f}{\partial x}(\mathbf{g}(r, \theta)) & \dfrac{\partial f}{\partial y}(\mathbf{g}(r, \theta)) \end{pmatrix} \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

$$= \begin{pmatrix} \cos\theta\dfrac{\partial f}{\partial x}(\mathbf{g}(r, \theta)) + \sin\theta\dfrac{\partial f}{\partial y}(\mathbf{g}(r, \theta)) & -r\sin\theta\dfrac{\partial f}{\partial x}(\mathbf{g}(r, \theta)) + r\cos\theta\dfrac{\partial f}{\partial y}(\mathbf{g}(r, \theta)) \end{pmatrix}$$

*Since*

$$D(f\circ\mathbf{g}) = \begin{pmatrix} \dfrac{\partial(f\circ\mathbf{g})}{\partial r} & \dfrac{\partial(f\circ\mathbf{g})}{\partial\theta} \end{pmatrix},$$

*we then have that*

$$\frac{\partial(f\circ\mathbf{g})}{\partial r} = \cos\theta\frac{\partial f}{\partial x}(\mathbf{g}(r, \theta)) + \sin\theta\frac{\partial f}{\partial y}(\mathbf{g}(r, \theta))$$

$$= \cos\theta\frac{\partial f}{\partial x}(r\cos\theta, r\sin\theta) + \sin\theta\frac{\partial f}{\partial y}(r\cos\theta, r\sin\theta),$$

*and*

$$\frac{\partial(f\circ\mathbf{g})}{\partial\theta} = -r\sin\theta\frac{\partial f}{\partial x}(\mathbf{g}(r, \theta)) + r\cos\theta\frac{\partial f}{\partial y}(\mathbf{g}(r, \theta))$$

$$= -r\sin\theta\frac{\partial f}{\partial x}(r\cos\theta, r\sin\theta) + r\cos\theta\frac{\partial f}{\partial y}(r\cos\theta, r\sin\theta).$$

# Conclusion