Derivation of the chain rule

Assume $f(x)$ and $g(x)$ are differentiable functions. We want to show that the derivative of $f(g(x))$ equals $f'(g(x))g'(x)$, which is the chain rule for derivatives:

$$[f(g(x))]' = f'(g(x))g'(x).$$

Before we begin, I'd like to remark on the notation used to define derivatives. I happen to like the Greek letter δ (lowercase *delta*), so I defined the derivative of $f(x)$ as

$$f'(x) = \lim_{\delta\to 0} \frac{f(x+\delta) - f(x)}{\delta}.$$

Instead, we could use the variable Δ (uppercase *delta*) and write

$$f'(x) \equiv \lim_{\Delta\to 0} \frac{f(x+\Delta) - f(x)}{\Delta}.$$

In fact, we can use *any* variable for the limit expression. All that matters is that we divide by the *same* non-zero quantity as the quantity added to x inside the function, and that this quantity goes to zero. If we're not careful with our choice of limit variable we could run into trouble. Specifically, the definition of a limit depends on a "small nonzero number Δ," which is then used in the limit $\Delta \to 0$. The condition $\Delta \neq 0$ is essential because the expression $\frac{f(x+\Delta)-f(x)}{\Delta}$ is not well defined when $\Delta = 0$, since it leads to a divide-by-zero error.

In order to avoid any possibility of such errors, we define the following piecewise function

$$R(y,b) \equiv \begin{cases} \frac{f(y)-f(b)}{y-b} & \text{if } y \neq b, \\ f'(b) & \text{if } y = b. \end{cases}$$

Observe the function $R(y,b)$ is continuous in y, when we treat b as a constant. This follows from the definition of the derivative formula and the assumption that $f(x)$ is differentiable. Using the function $R(y,b)$, we can write the formula for the derivative of $f(x)$ as $f'(x) = \lim_{\Delta\to 0} R(x+\Delta, x)$. Note this formula is valid even in the case $\Delta = 0$.

To prove the chain rule, we'll need the function $R\big(g(x+\delta), g(x)\big)$ which is defined as follows:

$$R\big(g(x+\delta), g(x)\big) \equiv \begin{cases} \frac{f(g(x+\delta))-f(g(x))}{g(x+\delta)-g(x)} & \text{if } g(x+\delta) \neq g(x), \\ f'(g(x)) & \text{if } g(x+\delta) = g(x). \end{cases}$$

Okay, we're done with the preliminaries, so we can get back to proving the chain rule $[f(g(x))]' = f'(g(x))g'(x)$. We start with the limit expression for the left hand side of the equation:

$$[f(g(x))]' = \lim_{\delta \to 0} \frac{f(g(x+\delta)) - f(g(x))}{\delta}.$$

Observe that the fraction inside the limit can be written as

$$\frac{f(g(x+\delta)) - f(g(x))}{\delta} = R\big(g(x+\delta), g(x)\big) \frac{g(x+\delta) - g(x)}{\delta}.$$

This is the most tricky part of the proof so let's analyze carefully why this equation holds. We must check the equation holds in the two special cases in the definition of $R\big(g(x+\delta), g(x)\big)$.

Case A Whenever $g(x+\delta) \neq g(x)$ we have

$$\begin{aligned} \frac{f(g(x+\delta)) - f(g(x))}{\delta} &= \frac{f(g(x+\delta)) - f(g(x))}{\delta} \frac{g(x+\delta) - g(x)}{g(x+\delta) - g(x)} \\ &= \frac{f(g(x+\delta)) - f(g(x))}{g(x+\delta) - g(x)} \frac{g(x+\delta) - g(x)}{\delta} \\ &= R\big(g(x+\delta), g(x)\big) \frac{g(x+\delta) - g(x)}{\delta} \end{aligned}$$

Case B For points where $g(x+\delta) = g(x)$ we have

$$\frac{f(g(x+\delta)) - f(g(x))}{\delta} = \frac{0}{\delta} = 0,$$

and

$$R\big(g(x+\delta), g(x)\big) \frac{g(x+\delta) - g(x)}{\delta} = f'(g(x)) \frac{0}{\delta} = 0.$$

Thus, the equation $\frac{f(g(x+\delta)) - f(g(x))}{\delta} = R\big(g(x+\delta), g(x)\big) \frac{g(x+\delta) - g(x)}{\delta}$ holds in both cases.

We can now rewrite the limit expression for $[f(g(x))]'$ using the equation established above:

$$\begin{aligned} [f(g(x))]' &= \lim_{\delta \to 0} \frac{f(g(x+\delta)) - f(g(x))}{\delta} \\ &= \lim_{\delta \to 0} \left(\underbrace{R\big(g(x+\delta), g(x)\big)}_{F_1} \underbrace{\frac{g(x+\delta) - g(x)}{\delta}}_{F_2} \right) \end{aligned}$$

We're trying to evaluate a limit expression that is the product of two factors $\lim_{\delta \to 0} F_1 F_2$. The limit of a product exists if the limits of both factors $\lim_{\delta \to 0} F_1$ and $\lim_{\delta \to 0} F_2$ exist. Before we proceed, we must evaluate the limit $\delta \to 0$ for both factors to ensure they exist.

To obtain the limit of the first factor, we'll rely on the continuity of the functions $g(x)$ and $R(y, b)$:

$$\lim_{\delta\to 0} g(x+\delta) = g(x) \quad \text{and} \quad \lim_{\Delta\to 0} R\big(b+\Delta, b\big) = R\big(b, b\big) = f'(b).$$

We define the quantity $\Delta \equiv g(x+\delta) - g(x)$ and using the continuity of $g(x)$, we can establish $\Delta \to 0$ as $\delta \to 0$. We are therefore allowed to change the limit variable from δ to Δ, and evaluate the limit of the first factor as follows:

$$\begin{aligned}\lim_{\delta\to 0} F_1 &= \lim_{\delta\to 0} R\big(g(x+\delta), g(x)\big)\\ &= \lim_{\Delta\to 0} R\big(g(x)+\Delta, g(x)\big)\\ &= R\big(g(x), g(x)\big) = f'(g(x)).\end{aligned}$$

We also know the limit of the second factor exists because it corresponds to the derivative of $g(x)$:

$$\lim_{\delta\to 0} F_2 = \lim_{\delta\to 0} \frac{g(x+\delta) - g(x)}{\delta} = g'(x),$$

and we assumed $g(x)$ is differentiable so its derivative must exists.

Since the limits of both factors $\lim_{\delta\to 0} F_1$ and $\lim_{\delta\to 0} F_2$ exist and are well defined, we can now complete the proof:

$$\begin{aligned}[f(g(x))]' &= \lim_{\delta\to 0}\left(R\big(g(x+\delta), g(x)\big)\frac{g(x+\delta)-g(x)}{\delta}\right)\\ &= \left(\lim_{\delta\to 0} R\big(g(x+\delta), g(x)\big)\right)\left(\lim_{\delta\to 0}\frac{g(x+\delta)-g(x)}{\delta}\right)\\ &= f'(g(x))g'(x).\end{aligned}$$

This establishes the validity of the chain rule $[f(g(x))]' = f'(g(x))g'(x)$.

Alternate notation

The presence of so many primes and brackets can make derivative formulas difficult to read. As an alternative, we sometimes use the *Leibniz notation* for derivatives. The three rules of derivatives in Leibniz notation are written as follows:

- Linearity: $\frac{d}{dx}(\alpha f(x) + \beta g(x)) = \alpha\frac{df}{dx} + \beta\frac{dg}{dx}$
- Product rule: $\frac{d}{dx}(f(x)g(x)) = \frac{df}{dx}g(x) + f(x)\frac{dg}{dx}$
- Chain rule: $\frac{d}{dx}\left(f(g(x))\right) = \frac{df}{dg}\frac{dg}{dx}$

Some authors prefer the notation $\frac{df}{dx}$ for the derivative of $f(x)$ because it is more evocative of a rise-over-run calculation.