Theorem: Suppose $I, J \subseteq \mathbb{R}$ are open intervals, $a \in I$, and both $g: I \mapsto J$ and $f: J \mapsto \mathbb{R}$ are differentiable at $a$ and $g(a)$, respectively. Then $f \circ g: I \mapsto \mathbb{R}$ is differentiable at $a$ and $(f \circ g)'(a) = (f'(g(a)) \cdot g'(a))$.

Comment about notation and history: In applications it is common to write $u = g(x)$ and $y = f(u)$ and rephrase the chain-rule as the very suggestive form \( \frac{dy}{dx} = \frac{du}{dx} \cdot \frac{du}{dx} \). While inappropriate for most of our class, in all fairness, this suggestive notation allows for very effective computations in many settings. Moreover, this effectiveness (and suggestiveness) are certainly major reasons why the whole world now uses the notation of differentials introduced by Leibniz, and hardly anybody uses Newton’s formalism!

Leibniz’ notation is often understood as an infinitesimal version (“limits”) of difference quotients

\[
\frac{\Delta y}{\Delta x} = \frac{\Delta y}{\Delta u} \cdot \frac{\Delta u}{\Delta x}, \quad \text{meaning} \quad \frac{(f \circ g)(x) - (f \circ g)(a)}{x - a} = \frac{(f(g(x)) - f(g(a))}{g(x) - g(a)} \cdot \frac{g(x) - g(a)}{x - a}
\]

But just because the left-hand side of either equation makes sense does not mean that the right-hand side of the equations must make sense: When considering the limit of the quotient as $x$ approaches $a$, we only consider $x \neq a$, assuring that there is no division by zero on the left. However, the right-hand side is undefined when $g(x) - g(a) = 0$. The case when $g(x) = g(a)$ for all $x$ in some open interval containing $a$ is boring, and handled easily. The case when $g(x) = g(a)$ only for some $x$ a positive distance away from $a$ is also easy: Just restrict $g$ to an interval containing $a$ where $g(x) = g(a)$ only when $x = a$.

A typical trouble maker to have in mind is the function $g: \mathbb{R} \mapsto \mathbb{R}$ that is defined by $g(0) = 0$ and $g(x) = \max\{0, x^2 \sin \frac{1}{x}\}$ which in every interval $(-\delta, \delta)$ about $a = 0$ takes the value $g(x) = g(0) = 0$ on infinitely many intervals near zero. This situation can still be handled directly, but it does not look terribly appealing . . . to get the formula for the chain-rule one may work with sequences of distinct terms $(x_n)_{n=1}^{\infty}$ that converge to $a$ and that are such that $\forall n, \ g(x_n) \neq g(a)$. But to prove differentiability this way requires more work. An elegant alternative approach is given below. It is worthwhile remembering the general structure of this construction.

Proof. Suppose $I, J \subseteq \mathbb{R}$ are open intervals, $a \in I$, and both $g: I \mapsto J$ and $f: J \mapsto \mathbb{R}$ are differentiable at $a$ and $g(a)$, respectively. Since $f$ is differentiable at $g(a)$, the function $\varphi: J \mapsto \mathbb{R}$ defined by

\[
\varphi(u) = \begin{cases} 
  f'(g(a)) \cdot g'(a) & \text{if } u = g(a) \\
  \frac{f'(u)}{g'(u)} & \text{else}
\end{cases}
\]

is continuous at $g(a)$. Moreover, since $g$ is continuous at $a$, $(\varphi \circ g)$ is continuous at $a$. Note that the identity $\varphi(g(x)) \cdot (g(x) - g(a)) = (f \circ g)(x) - (f \circ g)(a)$ holds for all $x \in I$, irrespective of whether $g(x) \neq g(a)$ or $g(x) = g(a)$. Thus $(f \circ g)$ is differentiable at $a$ and

\[
(f' \circ g)(x) \cdot g'(a) = \lim_{x \to a} \varphi(g(x)) \cdot \frac{(g(x) - g(a))}{x - a} = \lim_{x \to a} \frac{(f \circ g)(x) - (f \circ g)(a)}{x - a} = (f \circ g)'(a)
\]

Exercises:

- Use the chain rule to prove the product rule: Start with $F: u \mapsto u^2$ and $G = (f + g)$ and use that $ab = \frac{1}{2}((a + b)^2 - (a^2 + b^2))$.
- Simplify as much as possible by repeatedly applying the chain (and product rules): $(f \circ g)'$, $(f \circ g)''$, . . . , and, as a challenge, $(f \circ g)^{(n)}$.
- Suppose that $f: \mathbb{R} \mapsto \mathbb{R}$ is differentiable at $a \in \mathbb{R}$ and $f'(a) \neq 0$. Assuming that $f^{-1}$ exists and is differentiable at $f(c)$, use the chain-rule to derive a formula for $(f^{-1})'(f(c))$.
- Suppose that $f: \mathbb{R} \mapsto \mathbb{R}$ is differentiable at $a \in \mathbb{R}$ and $f'(a) \neq 0$. Prove that there exists some interval $(c - \delta, c + \delta)$ on which $f$ is invertible, and that $f^{-1}$ is differentiable at $f(c)$.
- Note that the theorem does not require that either $f$ or $g$ have continuous derivatives, nor require that either function is differentiable on an open interval containing that point. Verify that neither proof accidentally uses such hypotheses. Inspect how the chain rule works when composing a usual suspect $x \mapsto x^2 \sin \frac{1}{x}$ when composed with itself or another function (either order).