Chain Rule!!

Solvers: Ali, Ethan, Ryan
Writeup: Ethan

Let X, Y, Z be normed vector spaces, and $g : X \to Y$, $f : Y \to Z$ be functions. For some $p \in X$, suppose that g is totally differentiable at p and f is totally differentiable at $g(p)$. Then the function $f \circ g$ is totally differentiable at p, and $(f \circ g)' = f'(g(p)) \circ g'(p)$.

Proof. Let $L_f = f'(g(p))$ and $L_g = g'(p)$. First, let's prove that $L_f \circ L_g$ is a linear bounded operator. For $a, b \in X$, $c \in \mathbb{R}$,

$$L_f(L_g(ca + b)) = L_f(cL_g(a) + L_g(b)) = cL_f(L_g(a)) + L_f(L_g(b))$$

Thus $L_f \circ L_g$ is linear. Now, since L_f, L_g are bounded linear operators, there exist constants $M_f, M_g > 0$ such that for all $x \in X$, $y \in Y$,

$$\|L_f(y)\|_Z \le M_f \|y\|_Y \text{ and } \|L_g(x)\|_Y \le M_g \|x\|_X$$

Then we have that for all $x \in X$,

$$\|L_f(L_g(x))\|_Z \le M_f \|L_g(x)\|_Y \le M_f \cdot M_g \|x\|_X$$

so it is bounded as well. Now we can move on to the long part of the proof.

Consider the case if $g(p + t) - g(p) = 0$ for some $t \in X$ in every open neighbourhood around 0. Then it is probably true that $g(x)$ is constant when x is sufficiently close to p. We can quickly verify that $L_g = 0$ because

$$\lim_{h \to 0} \frac{\|g(p + h) - g(p)\|_Y}{\|h\|_X} = \lim_{h \to 0} 0 = 0$$

From this, we can see that

$$\lim_{h \to 0} \frac{\|f(g(p + h)) - f(g(p))\|_Z}{\|h\|_X} = 0,$$

so $(f \circ g)' = 0 = L_f(L_g)$.

Otherwise, there is some open ball $B(0, r)$ such that $g(p + h) - g(p) \ne 0$ for all $h \in B(0, r)$.

Let $L_f = f'(g(p))$ and $L_g = g'(p)$. Since these are bounded linear operators, there exist constants $M_f, M_g > 0$ such that for all $x \in X$, $y \in Y$,

$$\|L_f(y)\|_Z \le M_f \|y\|_Y \text{ and } \|L_g(x)\|_Y \le M_g \|x\|_X$$

Let $\varepsilon > 0$. By our assumption, there exists positive numbers δ_f, δ_g such that for h_f, h_g with $\|h_f\|_X < \delta_f$, $\|h_g\|_X < \delta_g$,

$$\frac{\|f(g(p) + h_f) - f(g(p)) - L_f(h_f)\|_Z}{\|h_f\|_X} < \min\left\{\sqrt{\varepsilon}, \frac{\varepsilon}{4M_g}\right\} \tag{1}$$

and

$$\frac{\|g(p + h_g) - g(p) - L_g(h_g)\|_Y}{\|h_g\|_X} < \min\left\{\frac{\sqrt{\varepsilon}}{4}, \frac{\varepsilon}{2M_f}\right\} \tag{2}$$

As well, since g being totally differentiable at p implies continuity at p, then for all $\|h_c\|_X < \delta_c$,

$$\|g(p + h_c) - g(p)\| < \delta_f \tag{3}$$

Define $\varepsilon_f = \min\left\{\sqrt{\varepsilon}, \frac{\varepsilon}{4M_g}\right\}$ and $\varepsilon_g = \min\left\{\frac{\sqrt{\varepsilon}}{4}, \frac{\varepsilon}{2M_f}\right\}$, and let $\delta = \min\{\delta_g, \delta_c, r\}$. For $h \in X$ with $\|h\|_X < \delta$,

$$\frac{\|f(g(p + h)) - f(g(p)) - L_f(L_g(h))\|_Z}{\|h\|_X}$$

$$\le \frac{\|f(g(p + h)) - f(g(p)) - L_f(g(p + h) - g(p))\|_Z + \|L_f(g(p + h) - g(p) - L_g(h))\|_Z}{\|h\|_X}$$

Consider the first term. Because of (3), we can apply (1) with $h_f = g(p+h) - g(p)$ and get that

$$\frac{\|f(g(p+h)) - f(g(p)) - L_f(g(p+h) - g(p))\|_Z}{\|h\|_X} < \frac{\|g(p+h) - g(p)\|}{\|h\|_X}\varepsilon_f$$

Then applying (2) with $h_g = h$,

$$\leq \frac{\|g(p+h) - g(p) - L_g(h)\|_Y + \|L_g(h)\|_Y}{\|h\|_X}\varepsilon_f \leq \frac{\sqrt{\varepsilon}}{4}\varepsilon_f + M_g\varepsilon_f \leq \frac{\sqrt{\varepsilon}}{4}\sqrt{\varepsilon} + M_g\frac{\varepsilon}{4M_g}$$

$$= \frac{\varepsilon}{2}$$

Now, we examine the second term. Using the fact that L_f is a bounded linear operator and (2),

$$\frac{\|L_f(g(p+h) - g(p) - L_g(h))\|_Z}{\|h\|_X} \leq \frac{M_f\|g(p+h) - g(p) - L_g(h)\|_Y}{\|h\|_X} < \frac{M_f\varepsilon}{2M_f} = \frac{\varepsilon}{2}$$

Adding the two together we conclude that

$$\frac{\|f(g(p+h)) - f(g(p)) - L_f(g(p+h) - g(p))\|_Z + \|L_f(g(p+h) - g(p) - L_g(h))\|_Z}{\|h\|_X}$$

$$< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

and we are done.

$\square$