Proof of chain rule via linearization

The chain rule states that if g is differentiable at c, and f is differentiable at $g(c)$, then $f \circ g$ is differentiable at c, and

$$(f \circ g)'(c) = f'(g(c))\, g'(c).$$

Sometimes the following "proof" of the chain rule is provided, to convince one that this formula is reasonable:

$$\begin{aligned}
(f\circ g)'(c) &= \lim_{x \to c} \frac{f(g(x)) - f(g(c))}{x - c} \\
&\overset{(*)}{=} \lim_{x \to c} \frac{f(g(x)) - f(g(c))}{g(x) - g(c)} \cdot \frac{g(x) - g(c)}{x - c} \\
&= \lim_{g(x) \to g(c)} \frac{f(g(x)) - f(g(c))}{g(x) - g(c)} \cdot \lim_{x \to c} \frac{g(x) - g(c)}{x - c} \\
&= f'(g(c))\, g'(c)
\end{aligned}$$

Convincing as it is, this is technically not a correct proof, since eg. in step $(*)$, one needs to divide by $g(x) - g(c)$, and $g(x) - g(c)$ may well be zero for many x that is in a deleted neighborhood of c.

Below we aim to give a "better" proof of chain rule.

The key insight is that of <u>linearization</u>.

Recall that a linear function on $\mathbb{R}$ is a function of the form $L(x) = ax+b$, where $a, b \in \mathbb{R}$ are constants.

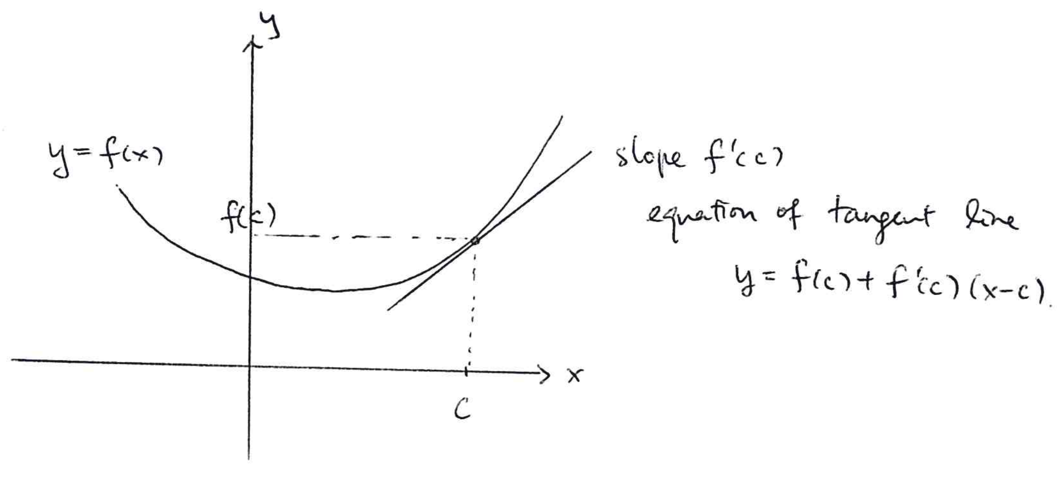
The graphs of such L's are straight lines (with finite slopes)

If f is a function that is differentiable at a point $c \in \mathbb{R}$, then one can find the tangent line of its graph at $x=c$. The tangent line is a straight line (with finite slope), and is hence the graph of a linear function.

In fact, the tangent line at $x=c$ has ~~equation~~ equation.

$$y = f(c) + f'(c)\cdot(x-c).$$

(Verify!). If f is differentiable at c, then the graph of f should be well-approximated by its tangent line near $x=c$:



Hence the function $f(x)$ should be well-approximated by the linear function defining the tangent line, i.e.

$$f(x) \approx f(c) + f'(c)(x-c) \qquad \text{when } x \approx c.$$

To make this precise, suppose f is defined in a neighborhood of some $c \in \mathbb{R}$, and that f is differentiable at c. For h in a neighborhood of 0, define

$$e(h) = f(c+h) - f(c) - f'(c)\, h.$$

Then setting $h = x-c$, we see that

$$f(x) = f(c) + f'(c)(x-c) + e(x-c) \qquad \forall\, x \text{ in a neighborhood of } c;$$

also, for $h \neq 0$, but close to 0,

$$\frac{e(h)}{h} = \frac{f(c+h) - f(c)}{h} - f'(c)$$

so by differentiability of f at c,

$$\lim_{h \to 0} \frac{e(h)}{h} \text{ exists, and } \lim_{h \to 0} \frac{e(h)}{h} = 0.$$

This proves the following proposition:

<u>Proposition 1</u> Suppose f is defined in a neighborhood of some $c \in \mathbb{R}$, and that f is differentiable at c.

Then $\exists$ a function e, defined in a neighborhood of 0, s.t.

$$\begin{cases} f(x) = f(c) + f'(c)(x-c) + e(x-c) & \forall x \text{ in a neighborhood of } c \\ \lim_{h \to 0} \frac{e(h)}{h} \text{ exists and equals } 0. \end{cases}$$

This is a precise way of saying that "$f(x) \approx f(c) + f'(c)(x-c)$ $\forall x \approx c$, if f is differentiable at c".

What is often useful (conceptually) is the following converse:

<u>Proposition 2</u> Suppose f is defined in a neighborhood of some $c \in \mathbb{R}$.

If $\exists$ $\alpha \in \mathbb{R}$, and if $\exists$ a function e defined in a neighborhood of 0, s.t.

$$\begin{cases} f(x) = f(c) + \alpha(x-c) + e(x-c) & \forall x \text{ in a neighborhood of } c \\ \lim_{h \to 0} \frac{e(h)}{h} \text{ exists and equals } 0, \end{cases}$$

then f is differentiable at c, and $f'(c) = \alpha$.

Proof Suppose f is as in the Proposition. Then $\forall x$ in a deleted neighborhood of c, we have

$$\frac{f(x)-f(c)}{x-c} = \frac{\alpha(x-c)+e(x-c)}{x-c}$$

$$= \alpha + \frac{e(x-c)}{x-c}$$

Since $\lim_{h\to 0} \frac{e(h)}{h} = 0$, we have $\lim_{x\to c} \frac{e(x-c)}{x-c}$ exists & equals 0.

As a result,

$\lim_{x\to c} \frac{f(x)-f(c)}{x-c}$ exists, and equals $\alpha + 0 = \alpha$

$\therefore$ f is differentiable at c, and $f'(c) = \alpha$.

∎

Now we are ready for the proof of chain rule.

Proof of chain rule

Suppose g is differentiable at c, and $a = g(c)$.

Suppose also that f is differentiable at a.

Then by Proposition 1, $\exists$ functions e_1, e_2 defined in a neighborhood of 0, s.t.

$$\begin{cases} g(x) = g(c) + g'(c)(x-c) + e_1(x-c) & \forall x \text{ in a neighborhood of } c \quad (1) \\ \lim_{h\to 0} \dfrac{e_1(h)}{h} = 0 & (2) \end{cases}$$

$$\begin{cases} f(y) = f(a) + f'(a)(y-a) + e_2(y-a) & \forall y \text{ in a neighborhood of } a \quad (3) \\ \lim_{h\to 0} \dfrac{e_2(h)}{h} = 0 & (4) \end{cases}$$

Hence

$$f(g(x)) = f(a) + f'(a)(g(x)-a) + e_2(g(x)-a) \qquad (\text{by } (3))$$

$$= f(g(c)) + f'(g(c))(g(x)-g(c)) + e_2(g(x)-g(c))$$

$$= f(g(c)) + f'(g(c))\left[g'(c)(x-c) + e_1(x-c)\right] + e_2(g'(c)(x-c) + e_1(x-c)) \qquad (\text{by } (1))$$

$$= f(g(c)) + f'(g(c))\,g'(c)\,(x-c) + e(x-c)$$

where we define

$$e(h) = f'(g(c))\,e_1(h) + e_2(g'(c)h + e_1(h))$$

for h in a neighborhood of 0.

If we could show

(5) ... $\lim_{h\to 0} \dfrac{e(h)}{h}$ exists and equals 0,

then by Proposition 2, we have $f \circ g$ differentiable at c,

with $(f\circ g)'(c) = f'(g(c))\,g'(c)$

Hence it remains to prove ⑤

Now

$$\frac{e(h)}{h} = f'(g(c)) \frac{e_1(h)}{h} + \frac{e_2(g'(c)h + e_1(h))}{h}.$$

and $\lim_{h \to 0} \frac{e_1(h)}{h}$ exists and equals 0.

Hence to prove ⑤, it suffices to prove

⑥ ... $\lim_{h \to 0} \frac{e_2(g'(c)h + e_1(h))}{h}$ exists and equals 0.

To do so, recall

$$\lim_{h \to 0} \frac{e_2(h)}{h} = 0.$$

Hence if we define a new function $\tilde{e}_2$ in a neighborhood of 0,

by

$$\tilde{e}_2(h) = \begin{cases} \frac{e_2(h)}{h} & \text{if } h \neq 0 \\ 0 & \text{if } h = 0 \end{cases}$$

Then

$$e_2(h) = \tilde{e}_2(h) \cdot h \qquad \forall\, h \text{ in a neighborhood of } 0,$$

and $\tilde{e}_2$ is continuous at $h = 0$.

Now

(7) ... $\lim_{h\to 0} (g'(c)h + e_1(h))$ exists and equals 0,

Since $g'(c)h + e_1(h) = g'(c)h + \frac{e_1(h)}{h}\cdot h \to g'(c)\cdot 0 + 0\cdot 0 = 0$ as $h\to 0$

Hence for h sufficiently close to 0,

$$\frac{e_2(g'(c)h + e_1(h))}{h} = \tilde{e}_2(g'(c)h + e_1(h))\left[\frac{g'(c)h + e_1(h)}{h}\right]$$

$$= \tilde{e}_2(g'(c)h + e_1(h))\left[g'(c) + \frac{e_1(h)}{h}\right]$$

But by (7) & the continuity of $\tilde{e}_2$ at $h=0$, we get

$$\lim_{h\to 0} \tilde{e}_2(g'(c)h + e_1(h)) \text{ exists and equals } \tilde{e}_2(0) = 0.$$

Also, $\lim_{h\to 0} \left(g'(c) + \frac{e_1(h)}{h}\right)$ exists and equals $g'(c)$.

Hence altogether,

$$\lim_{h\to 0} \frac{e_2(g'(c)h + e_1(h))}{h} \text{ exists and equals } 0\cdot g'(c) = 0.$$

This proves (5), and concludes the proof of chain rule. ∎

We remark that the characterization of differentiability via linearization gives the correct definition of differentiability of functions defined in higher dimensions in $\mathbb{R}^n$. and the above proof of chain rule is valid there as well. This is another reason why we want to look at this proof.