Math 1010A Term 1 2017
Food for thought about the chain rule

The following exercise provides a guided attempt to prove the one-variable chain rule. The last question is intended to give you a first glimpse of the multivariable chain rule.

1. Let $f\colon I \to \mathbb{R}$ be a function defined on an open interval I, and $c \in I$.

   (a) Show that if there exists a function $\phi\colon I \to \mathbb{R}$ such that

   $$f(x) = f(c) + \phi(x)(x-c) \quad \text{for all } x \in I,$$

   and such that ϕ is continuous at c, then f is differentiable at c, and $f'(c) = \phi(c)$. (Hint: Compute the quotient $\frac{f(x)-f(c)}{x-c}$, and let x tend to c.)

   (b) Show that the converse of (a) also holds, in the sense that if f is differentiable at c, then there exists a function $\phi\colon I \to \mathbb{R}$ such that

   $$f(x) = f(c) + \phi(x)(x-c) \quad \text{for all } x \in I, \tag{1}$$

   and such that ϕ is continuous at c. How does the value of $\phi(c)$ depend on f? (Hint: The identity (1) *defines* $\phi(x)$ for you already, for all $x \in I$ that is not equal to c. Just make $\phi(x)$ the subject of (1)! Now figure out what $\phi(x)$ should be when $x = c$, if ϕ were to be continuous at c.)

   Note that the above shows that if f is differentiable at c, then for $x \simeq c$, we have

   $$f(x) = f(c) + \phi(x)(x-c) \simeq f(c) + \phi(c)(x-c) = f(c) + f'(c)(x-c)$$

   The (linear) function $x \mapsto f(c) + f'(c)(x-c)$ is just the equation of the tangent line to the graph of f through $(c, f(c))$. Hence if f is differentiable at c, then when x is very close to c, $f(x)$ is very close to the tangent line through $(c, f(c))$. Every differentiable function is almost linear (locally)!

2. In order to prove the chain rule, sometimes the following heuristic argument is given: let f be differentiable at $g(c)$, and g be differentiable at c. Then

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

   Why is this not quite a completely rigorous proof?

3. We are going to give a correct proof of the chain rule in this exercise. Let f be differentiable at $g(c)$, and g be differentiable at c.

   (a) Using Question 1b, show that there exists functions ϕ and ψ, such that

   $$g(x) = g(c) + \phi(x)(x - c) \quad \text{for all } x \text{ near } c,$$

   $$f(y) = f(g(c)) + \psi(y)(y - g(c)) \quad \text{for all } y \text{ near } g(c),$$

   and such that ϕ and ψ are continuous at c and $g(c)$ respectively. Note that $\phi(c) = g'(c)$, $\psi(g(c)) = f'(g(c))$.

   (b) Using part (a), show that

   $$f(g(x)) = f(g(c)) + \psi(g(x))\phi(x)(x - c) \quad \text{for all } x \text{ near } c.$$

   (c) Show that $\psi(g(x))\phi(x)$ is continuous at $x = c$, and is equal to $f'(g(c))g'(c)$ at $x = c$.

   (d) Using parts (b) and (c), together with Question 1a, conclude that $f \circ g$ is differentiable at $x = c$, with $(f \circ g)'(c) = f'(g(c))g'(c)$.

4. Having settled the one-variable chain rule, let's try to guess what the multivariable chain rule may be like.

   (a) To make sense of the multivariable chain rule, one needs to know the notion of partial derivatives. If $f(x, y)$ is a function of two real variables, then $\frac{\partial f}{\partial x}$ is the derivative of $f(x, y)$ by holding y fixed and differentiating only with respect to x; similarly $\frac{\partial f}{\partial y}$ is the derivative of $f(x, y)$ by holding x fixed and differentiating only with respect to y. Checkpoint: if $f(x, y) = x^2 \sin y$, let's verify that

   $$\frac{\partial f}{\partial x} = 2x \sin y, \quad \frac{\partial f}{\partial y} = x^2 \cos y.$$

   (b) Let's begin with a simple situation. Suppose we have a smooth function $f(x, y)$ of two real variables, and both x and y are differentiable functions of one real variable t. We want to differentiate with respect to t, i.e. to compute

   $$\frac{d}{dt} f(x(t), y(t)).$$

   To guess what the answer may be like, let's assume that f is the simplest kind of functions, namely linear functions. In other words, let's assume that $f(x, y) = ax + by$ for some constants a and b. In that case,

      (i) What is $\frac{d}{dt} f(x(t), y(t))$?

(ii) What is $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$?

(iii) Verify that

$$\frac{d}{dt}f(x(t),y(t)) = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

in this special case.

It turns out that this is true in general: if $f(x,y)$ is a smooth function of two real variables, and $x(t)$, $y(t)$ are differentiable functions of one real variable t, then

$$\frac{d}{dt}f(x(t),y(t)) = \frac{\partial f}{\partial x}(x(t),y(t))\frac{dx}{dt} + \frac{\partial f}{\partial y}(x(t),y(t))\frac{dy}{dt}.$$

(How does this compare with the one-variable chain rule? It should specialize to the one-variable chain rule if f is independent of y!)

(c) More generally, write $x = (x_1,\dots,x_n) \in \mathbb{R}^n$, $t = (t_1,\dots,t_m) \in \mathbb{R}^m$. If f is a differentiable function of x and x is a differentiable function of t, then the chain rule states that

$$\frac{\partial}{\partial t_j}f(x(t)) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial t_j} \quad \text{for any } 1 \leq j \leq m;$$

here the partial derivatives of f on the right hand side are all evaluated at the point $x(t)$. In matrix notations, this says

$$\begin{pmatrix} \frac{\partial f}{\partial t_1} & \frac{\partial f}{\partial t_2} & \cdots & \frac{\partial f}{\partial t_m} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{pmatrix} \begin{pmatrix} \frac{\partial x_1}{\partial t_1} & \frac{\partial x_1}{\partial t_2} & \cdots & \frac{\partial x_1}{\partial t_m} \\ \frac{\partial x_2}{\partial t_1} & \frac{\partial x_2}{\partial t_2} & \cdots & \frac{\partial x_2}{\partial t_m} \\ \vdots & & & \vdots \\ \frac{\partial x_n}{\partial t_1} & \frac{\partial x_n}{\partial t_2} & \cdots & \frac{\partial x_n}{\partial t_m} \end{pmatrix}.$$

(We abused notation and wrote $\frac{\partial f}{\partial t_j}$ in place of $\frac{\partial}{\partial t_j}f(x(t))$ on the left hand side, as is often customary.) It is not a surprise that the notion of matrix multiplication is compatible with the chain rule, since matrix multiplications are defined so as to capture the composition of linear maps, and every differentiable map on an Euclidean space is almost linear!