Differentiability and the Chain Rule

September 24, 2010

This note will give an alternate definition of differentiability and derivation of the chain rule. I learned this idea from some notes of Michael Range. He attributes the idea to Caratheodory. The usual definition is as follows.

Definition 1. *Let f be a function defined in a neighborhood of a point a in $\mathbb{R}^n$. f is differentiable at a if there is a vector $c = (c_1, c_2, \ldots, c_n) \in \mathbb{R}^n$ so that*

$$\lim_{x \to a} \frac{f(x) - f(a) - \sum_{i=1}^n c_i(x_i - a_i)}{|x - a|} = 0. \tag{1}$$

In this definition, $|x - a|$ can be any one of the p-norms, $|x-a|_p = \left(\sum_1^n |x_i - a_i|^p \right)^{1/p}$. If we use the 1-norm we can write

$$|x|_1 = \sum_1^n \sigma_i x_i, \tag{2}$$

where σ_i is the sign of x_i. The numbers c_i are defined to be the partial derivatives of f at a, $\frac{\partial f}{\partial x_i}(a) = c_i$ and $\nabla f(a) = (c_1, \ldots, c_n)$.

Theorem 1. *f is differentiable at a if and only if there are functions $q_i(x), i = 1, \ldots, n$ which are continuous at a, such that*

$$f(x) = f(a) + \sum_1^n q_i(x)(x_i - a_i). \tag{3}$$

Proof. Assume f is differentiable. Let

$$r(x) = \frac{f(x) - f(a) - \sum_{i=1}^n c_i(x_i - a_i)}{|x - a|_1}. \tag{4}$$

By (1), $r(x) \to 0$ as $x \to a$. Using (2) rewrite (4) as

$$f(x) = f(a) + \sum_1^n (c_i + \sigma_i r(x))(x_i - a_i), \tag{5}$$

where σ_i is the sign of $x_i - a_i$. Now let $q_i(x) = c_i + \sigma_i r(x)$. Since $r(x) \to 0$, q_i is continuous at a and we have proved (3).

Assume (3), where q_i is continuous at a. Then let $c_i = q_i(a)$. We can write $q_i(x) = c_i + r_i(x)$, where $r_i \to 0$ as $x \to a$. Then

$$\frac{f(x) - f(a) - \sum_{i=1}^n c_i(x_i - a_i)}{|x - a|} = \frac{\sum_1^n r_i(x)(x_i - a_i)}{|x - a|}, \tag{6}$$

and

$$\left|\frac{\sum_1^n r_i(x)(x_i - a_i)}{|x-a|}\right| \le \sum_1^n |r_i(x)|,$$

since $|x_i - a_i|/|x-a| \le 1$. This goes to 0 as $x \to a$. □

Remark: $q_i(a) = \dfrac{\partial f}{\partial x_i}(a)$.

Theorem 2. *(Chain rule) Let f be differentiable at $a \in \mathbb{R}^n$. Let $\gamma(t) = (\gamma_1(t), \dots, \gamma_n(t))$ be differentiable at 0 and $\gamma(0) = a$. Then $g(t) = f(\gamma(t))$ is differentiable at 0 and*

$$g'(0) = \nabla f(a) \cdot \gamma'(0) = \sum_1^n \frac{\partial f}{\partial x_i}(a)\gamma_i'(0).$$

Proof. The proof is a string of equations. Differentiability of γ_i implies $\gamma_i(t) = \gamma_i(0) + s_i(t)t$ where s_i is continuous at 0 and $s_i(0) = \gamma_i'(0)$.

$$\begin{aligned} g(t) &= f(\gamma(0)) + \sum_1^n q_i(\gamma(t))(\gamma_i(t) - \gamma_i(0)) \\ &= g(0) + \sum_1^n q_i(\gamma(t))(s_i(t))t \\ &= g(0) + \left(\sum_1^n q_i(\gamma(t))(s_i(t))\right) t. \end{aligned}$$

The expression $\sum_1^n q_i(\gamma(t))(s_i(t)$ is continuous at 0 and its value at 0 is

$$\sum_1^n c_i\gamma_i'(0) = \sum_1^n \frac{\partial f}{\partial x_i}(a)\gamma_i'(0).$$

□