Asymptotic Notation and The Chain Rule

Nikhil Srivastava

September 3, 2015

In class I pointed out that the definition of the derivative:

$$f'(z) = \lim_{\Delta z \to 0} \frac{f(z+\Delta z) - f(z)}{\Delta z}$$

is equivalent, after rearrangement, to the existence of a number $f'(z) \in \mathbb{C}$ and an "error function" $\Phi(\Delta z)$ such that

$$f(z+\Delta z) = f(z) + f'(z)\Delta z + \Phi(\Delta z), \tag{1}$$

and

$$\lim_{\Delta z \to 0} \frac{\Phi(z)}{\Delta z} = 0.$$

Thus, the existence of a derivative at z is equivalent to the existence of a *local linear approximation* (1) in a neighborhood of z where the error $\Phi(\Delta z)$ decays more rapidly than the perturbation Δz. The derivative is simply the coefficient that appears in this linear approximation. This is not circular, but a perfectly self-contained way to *define* the derivative.

1 Asymptotic Notation

Often, we don't really care much about what the error function Φ is, but just that it approaches zero sufficiently rapidly. For situations like these, there is a handy and fully rigorous notation called "little-oh" notation. Given two functions $f(z)$ and $g(z)$ and a number a, we say that $f(z) = o(g(z))$ as $z \to a$ if

$$\lim_{z \to a} \left| \frac{f(z)}{g(z)} \right| = 0.$$

For instance, we have $z^2 = o(z)$ as $z \to 0$ since $\lim_{z \to 0} |z^2/z| = 0$. and similarly $\Delta z = o(1)$ as $\Delta z \to 0$. Often it will be clear from the context what a is and we will not write it explcitly.

There is a related notation called "big oh" notation: we say that $f(z) = O(g(z))$ as $z \to a$ if there is a constant $M > 0$ such that

$$|f(z)| \le M|g(z)|$$

in a punctured neighborhood of a. Note that in some cases of interest we may take $a = \infty$, and then a punctured neighborhood of ∞ is a set of the form $|z| > 1/\epsilon$. You may have encountered $O()$ notation when analyzing the running times of algorithms in computer science courses, and it is the same concept.

It is important to understand that despite the "=", both these notions are actually not equalities but *inequalities*; $f = o(g)$ means that $|f| < |g|$ in the limit, and $f = O(g)$ means that $|f| \leq M|g|$ in the limit. When we use an expression like $o(\Delta z)$ in an arithmetic expression, we mean that there exists some function $\Psi(\Delta z)$ such that $\Psi(\Delta z) = o(\Delta z)$, and we are using asymptotic notation as an unnamed placeholder for this function.

Here are some useful and easy to verify properties of little oh notation, which will be useful when we use it to write a proof of the chain rule.

(A) If $c \in \mathbb{C}$ and $f(z) = o(g(z))$ then $cf(z) = o(g(z))$.

(B) If $f(z) = o(g(z))$ and $g(z) = O(h(z))$ as $z \to a$ then $f(z) = o(h(z))$.

(C) If $f_1(z) = o(g(z))$ and $f_2(z) = o(g(z))$ then $(f_1 + f_2)(z) = o(g(z))$.

Proof. For (A), observe that

$$\lim_{z \to a} \frac{cf(z)}{g(z)} = c \lim_{z \to a} \frac{f(z)}{g(z)} = 0.$$

For (B), observe that since $f(z) = o(g(z))$ the limit

$$\lim_{z \to a} \left| \frac{f(z)}{g(z)} \right|$$

exists and $g(z) \neq 0$ in some punctured neighborhood $D^o(a, \epsilon_1) = \{z : 0 < |z - a| < \epsilon_1\}$ of a. Since $g(z) = O(h(z))$, there is a constant $M > 0$ and another punctured neighborhood $D^o(a, \epsilon_2) \subset D^o(a, \epsilon_1)$ such that $|h(z)| \geq g(z)/M$ (and in particular $h(z) \neq 0$) in $D^o(a, \epsilon_2)$. Thus, we can write

$$\lim_{z \to a} \left| \frac{f(z)}{h(z)} \right| \leq \lim_{z \to a} \frac{M|f(z)|}{|g(z)|} = 0,$$

as desired.

(C) is left as an exercise. □

The above rules are also useful when keeping track of error terms while manipulating truncated infinite series.

2 The Chain Rule

The chain rule is a consequence of the fact that differentiation produces the linear approximation to a function at a point, and that the derivative is the coefficient appearing in this linear approximation.

Suppose we are interested in computing the derivative of $(f \circ g)(z) = f(g(z))$ at z, where g is differentiable at z and f is differentiable at $g(z)$. Then the definition (1) implies that:

$$g(z + \Delta z) = g(z) + g'(z)\Delta z + o(\Delta z) \tag{2}$$

for a number $g'(z)$ which we call the derivative of g at z. Similarly, for $w = g(z)$ we have for every perturbation Δw:

$$f(w + \Delta w) = f(w) + f'(w)\Delta w + o(\Delta w), \tag{3}$$

as $\Delta w \to 0$. Take[1]

$$\Delta w = g(z + \Delta z) - g(z) = g'(z)\Delta z + o(\Delta z),$$

and observe that $\Delta w = O(\Delta z)$ as $\Delta z \to 0$, since

$$\lim_{\Delta z \to 0} \left| \frac{g'(z)\Delta z + o(\Delta z)}{\Delta z} \right| = |g'(z)|,$$

and we can take M to be any number larger than $|g'(z)|$. Then:

$$\begin{aligned}
(f \circ g)(z + \Delta z) &= f(g(z + \Delta z)) \\
&= f\left(g(z) + \Delta w\right) \\
&= f(g(z)) + f'(g(z))\Delta w + o(\Delta w) \\
&= f(g(z)) + f'(g(z))\left(g'(z)\Delta z + o(\Delta z)\right) + o(\Delta w) \\
&= f(g(z)) + f'(g(z)) \cdot g'(z)\Delta z + f'(g(z)) \cdot o(\Delta z) + o(\Delta w) \\
&= f(g(z)) + f'(g(z)) \cdot g'(z)\Delta z + o(\Delta z) \quad \text{by properties (A),(B), and (C).}
\end{aligned}$$

Thus, we must have $(f \circ g)'(z) = f'(g(z))g'(z)$, as desired.

[1] strictly speaking, we are defining Δw to be a function of Δz — it is the pertubation in $w = f(z)$ induced by Δz.