MA 8, WEEK 4: BIG O AND LITTLE O NOTATION, DIFFERENTIATION

1. Big O and Little o Notation

We'll introduce Big O and Little o notation in greater generality than was given in lecture.

Definition 1. We say $f(x) = O(g(x))$ as $x \to a$ if there exists a constant C such that

$$|f(x)| \leq C|g(x)|$$

in some punctured neighborhood of a, that is, for $x \in (a-\delta, a+\delta)\backslash\{x\}$ for some $\delta > 0$.

We say $f(x) = o(g(x))$ as $x \to a$ if

$$\lim_{x \to a} \frac{f(x)}{g(x)} = 0.$$

Remark 2. This notation is standard, but it is not good. Namely, the statement "$f(x) = O(g(x))$" has a specific meaning, but "$O(g(x)) = f(x)$" is meaningless. That is why "$f(x) = O(g(x))$" is often called an *abuse of notation*, because equality doesn't behave like it should. Some newer textbooks in computer science use "$f(x) \in O(g(x))$" to get around this, but the standard notation has unfortunately proliferated through various disciplines for decades, so we're stuck with it.

Prof. Katz tries to avoid this by saying "$f(x)$ is $O(g(x))$" which is how you would pronounce "f(x) = O(g(x))." The bottom line is, think of "being $O(g(x))$" or "being $o(g(x))$" as a *property* of a function.

Remark 3. In lecture, Prof. Katz gives this definition for $a = 0$. This is also how you prove it in your homework assignment.

In computer science, or in applied mathematics/engineering, we usually take $a = \infty$. Things are usually more intuitive from this point of view, in which you can think of the functions $g(x)$ as "dominating" $f(x)$ if $f(x) \in O(g(x))$ or $f(x) \in o(g(x))$. This also works to some extent for $a = 0$, but you'll see in the later examples it's slightly trickier.

Recall that the $a = \infty$ and $a = 0$ aren't unrelated. Recall from last time that we can define limits as $x \to \infty$ by inverting x and taking $x \to 0$.

Learning big O and little o notation is useful for many reasons. First, the language (especially Big O) is ubiquitous in any practical application of mathematics, especially in complexity theory in computer science and in engineering. Also, introducing the notation allows us to make many of the important proofs in this class more intuitive. Furthermore, it allows us to treat the familiar notion of "error" rigorously, which you'll see in any field that uses mathematics and statistics (namely, every discipline of science and engineering). However, the downside is that it looks

Date: October 22, 2013.

scary at first and you'll probably be confused the first time you see it. But trust that it's not too hard, and that it's worth becoming familiar with these ideas.

Since these ideas are most intuitive when sending $x \to \infty$, for our first examples, let's begin by **assuming that** $a = \infty$.

Example 4. We'll begin with big O:

- $x = O(x)$
- $60x^2 = O(x^2)$
- $x^2 = O(x^3)$
- $x^3 \neq O(x^2)$
- $x^2 \neq O(x \log x)$

For little O:

- $10x = o(x^2)$
- $2x^2 \notin o(x^2)$
- $\frac{1}{x} = o(1)$

Note that the limit a is important. For instance, $\frac{1}{x} = o(1)$ as $a \to \infty$, but $\frac{1}{x} \neq o(1)$ as $a \to 0$.

Remark 5. We usually prove results assuming $a = 0$ since changing $x - a$ to x is just a change of coordinates. You may have noticed this with power series as well. When proving results, we normally assume that the power series are expanded around 0.

The main way we use big O and little o notation is with power series (expanded about a) (especially Taylor expansions), as in the following familiar examples (with $a = 0$):

- $e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + O(x^4)$ as $x \to 0$
- $\frac{1}{1-x} = 1 + x + x^2 + O(x^3) = 1 + x + x^2 + o(x^2)$ as $x \to 0$
- $|x|^3 = O(x^3) = o(x^2)$ as $x \to 0$. As an exercise, check the following results.
- $\cosh(x) = O(e^x) = o(e^{5x/4})$ as $x \to \infty$.
- $1/\sin(x) = O(1/x) = o(1/x^{3/2})$ as $x \to 0$.

Note that saying the error is $o(x^n)$ is a stronger statement than saying the error is $O(x^n)$ (as $x \to 0$). However, in the most common cases, we are usually choosing between saying that the error is $o(x^2)$ or that the error os $O(x^3)$ (as $x \to 0$), in which case the latter statement is strongest.

Here are some general properties. (Please remember that the "=" sign here is not symmetric due to the abuse of notation.)

(1) $f(x) = O(f(x))$
(2) If $f(x) = o(g(x))$, then $f(x) = O(g(x))$.
(3) If $f(x) = O(g(x))$, then $O(f(x)) + O(g(x)) = O(g(x))$.
(4) If $f(x) = O(g(x))$, then $o(f(x)) + o(g(x)) = o(g(x))$.
(5) Let $c \neq 0$, then $c \cdot O(g(x)) = O(g(x))$ and $c \cdot o(g(x)) = o(g(x))$.
(6) $O(f(x))O(g(x)) = O(f(x)g(x))$
(7) $o(f(x))O(g(x)) = o(f(x)g(x))$
(8) If $g(x) = o(1)$, then $\frac{1}{1+o(g(x))} = 1 + o(g(x))$ and $\frac{1}{1+O(g(x))} = 1 + O(g(x))$.

The notation itself makes the results slightly confusing, but they make sense if you just think about the definitions and rule out of the absurdly untrue statements.

As a general rule, don't "divide" by some $o(g(x))$ of $O(g(x))$. The intuitive reason for this is that these usually represent small terms, and so can be zero (or may be close enough to zero cause the sum to blow up).

The proofs of these propositions are easy, and more or less involves unwinding the definitions. Let's work through (3).

Proof. Let $h(x) = O(f(x))$ and $j(x) = O(g(x))$, so there exist constants A and B such that

$$\begin{aligned} |h(x)| &\le A|f(x)| \\ |j(x)| &\le B|g(x)|. \end{aligned}$$

on a neighborhood around a. From our assumption, $f(x) = O(g(x))$, so there exists a constant C such that

$$|f(x)| \le C|g(x)|$$

on a neighborhood around a. Take the smaller of the two neighborhoods around a, so that all the inequalities hold. Then

$$\begin{aligned} |h(x) + j(x)| &\le |h(x)| + |j(x)| \\ &\le A|f(x)| + B|g(x)| \\ &\le AC|g(x)| + B|g(x)| \\ &\le (AC + B)|g(x)|. \end{aligned}$$

Therefore, $h(x) + j(x) = O(g(x))$. □

Here's an example of how we can use o notation to simplify a computation.

Example 6. Let's try and find the power series expansion of $\sin(x)\cos(x)$ up to an $o(x^4)$ error. Recall that

$$\begin{aligned} \sin(x) &= x - \frac{x^3}{6} + o(x^4) \\ \cos(x) &= 1 - \frac{x^2}{2} + \frac{x^4}{24} + o(x^4). \end{aligned}$$

We compute the product, and in doing so absorb all terms of $o(x^4)$ and smaller into $o(x^4)$:

$$\begin{aligned} \sin(x)\cos(x) &= \left(x - \frac{x^3}{6} + o(x^4)\right)\left(1 - \frac{x^2}{2} + \frac{x^4}{24} + o(x^4)\right) \\ &= x - \frac{x^3}{2} + \frac{x^5}{24} + x \cdot o(x^4) - \frac{x^3}{6} + \frac{x^5}{12} - \frac{x^7}{144} - \frac{x^3}{6} \cdot o(x^4) \\ &\quad + o(x^4) - \frac{x^2}{2} \cdot o(x^4) + \frac{x^4}{24} \cdot o(x^4) + o(x^4)o(x^4) \\ &= x - \frac{x^3}{2} - \frac{x^3}{6} + o(x^4) \\ &= x - \frac{2}{3}x^3 + o(x^4). \end{aligned}$$

What's nice about this is that the 4th order Taylor polynomial is $x - \frac{2}{3}x^3$ (c.f. Apostol 7.9), so this provides an alternative way to compute this.

Remark 7. This kind of procedure is useful, for instance, if it is very costly to compute. (Think transmitting a message from a remote location with the worst roaming charges you can imagine, like being annihilated by the hostile rebel forces around your hidden location.) Because instead of computing the many arithmetic operations needed to compute sine and cosine and then multiplying the values together, you can just plug in the value of x for $x - \frac{2}{3}x^3$ and use only ~ 5 operations and get something within $o(x^4)$ accuracy.

For a more pedestrian example, if you need to compute really really big numbers with any hope of accuracy, you need to use some kind of procedure like the above. (Indeed, almost all robust industrial computer programs make heavy use of such estimates.)

2. Differentiation

Definition 1. For a function f defined on some neighborhood $(a-\delta, a+\delta)$, we say that f is **differentiable** at a iff the limit

$$\lim_{h\to 0} \frac{f(a+h) - f(a)}{(a+h) - a}$$

exists. If it does, denote this limit as $f'(a)$; we will often call this value the **derivative** of f at a.

Proposition 2. *The derivative of $f(x) = 1/x$ at any point $a \neq 0$ is $-1/a^2$.*

Proof. Pick any point $a \neq 0$ in $\mathbb{R}$: then, a direct calculation of the limit tells us that

$$\begin{aligned}
\lim_{h\to 0} \frac{f(a+h) - f(a)}{(a+h) - a} &= \lim_{h\to 0} \frac{\frac{1}{a+h} - \frac{1}{a}}{h} \\
&= \lim_{h\to 0} \frac{\frac{a}{(a)\cdot(a+h)} - \frac{a+h}{(a)\cdot(a+h)}}{h} \\
&= \lim_{h\to 0} \frac{\frac{a-a-h}{(a)\cdot(a+h)}}{h} \\
&= \lim_{h\to 0} \frac{-h}{(h)\cdot(a)\cdot(a+h)} \\
&= \lim_{h\to 0} \frac{-1}{(a)\cdot(a+h)}
\end{aligned}$$

But this is just the quotient of a pair of polynomials, the denominator of which is nonzero as $h \to 0$. Because polynomials are continuous everywhere, we then know that we can pass the limit through the quotient operation, and just evaluate the numerator and denominator's limits separately. Hence, this limit is $-1/a^2$, as desired. □

This is probably very familiar to the limit calculations that you did in your previous calculus course. Indeed, such calculations provide the outline for the proofs of limits in this course, you just need to apply the rigorous methods at the right steps.

The above example was a rather quick and direct calculation; our second example, however, is a bit trickier:

Proposition 3. *The derivative of $f(x) = e^x$ at any point a is e^a.*

Before we begin this proof, however, we probably should define our terms: what do we even mean by e^x here, anyways? Assume that we don't know any of the familiar properties of e and only know that we have a power series expansion

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Proof. We start by simply examining the derivative as a limit:

$$\begin{aligned}
\lim_{h\to 0} \frac{f(a+h)-f(a)}{(a+h)-a} &= \lim_{h\to 0} \frac{e^{a+h}-e^a}{h} \\
&= \lim_{h\to 0} \frac{e^a e^h - e^a}{h} \\
&= \lim_{h\to 0} e^a \cdot \frac{e^h - 1}{h} \\
&= \lim_{h\to 0} e^a \cdot \frac{\left(\sum_{n=0}^{\infty} \frac{h^n}{n!}\right) - 1}{h}
\end{aligned}$$

(Notice here that we've used the fact $e^{a+h} = e^a e^h$, despite the fact that this is not obvious given our definition of e^x as a power series, and not as some kind of exponentiation! If you want, you can directly prove this fact: just multiply the polynomials $\left(1 + \frac{a}{1} + \frac{a^2}{2!} + \ldots + \frac{a^n}{n!}\right)$ and $\left(1 + \frac{h}{1} + \frac{h^2}{2!} + \ldots + \frac{h^n}{n!}\right)$ by each other, and by clever grouping/use of the binomial expansion formula, see that you have $\left(1 + \frac{a+h}{1} + \frac{(a+h)^2}{2!} + \ldots + \frac{(a+h)^n}{n!}\right)$.)

Once you're persuaded that this was legitimate, return to our expression above. In order to bring the -1 and the division by h into our sum $\sum_{n=0}^{\infty} \frac{h^n}{n}$, we express this sum as a limit, and then use the fact that limits play nicely with arithmetic:

$$\begin{aligned}
\lim_{h\to 0} \frac{f(a+h)-f(a)}{(a+h)-a} &= \lim_{h\to 0} e^a \cdot \frac{\left(\lim_{n\to\infty} 1 + \frac{h}{1} + \frac{h^2}{2!} + \ldots + \frac{h^n}{n!}\right) - 1}{h} \\
&= \lim_{h\to 0} e^a \cdot \frac{\left(\lim_{n\to\infty} \frac{h}{1} + \frac{h^2}{2!} + \ldots + \frac{h^n}{n!}\right)}{h} \\
&= \lim_{h\to 0} e^a \cdot \left(\lim_{n\to\infty} \frac{1}{1} + \frac{h}{2!} + \ldots + \frac{h^{n-1}}{n!}\right) \\
&= \lim_{h\to 0} e^a \cdot \left(\sum_{n=1}^{\infty} \frac{h^{n-1}}{n!}\right)
\end{aligned}$$

The series on the right has an infinite radius of convergence by the ratio test. In particular, it is strictly less than the series $\sum_{n=1}^{\infty} \frac{h^n}{n!}$, which we know to be e^h. Therefore, if we use the squeeze theorem and squish our limit between e^a and $e^a \cdot e^h$, because both $\lim_{h\to 0} e^a = \lim_{h\to 0} e^a \cdot e^h = e^a$, the squeeze theorem tells us that the derivative of e^x at a is e^a, as well. □

Remark 4. One thing to notice in the above proof: we've assumed that $\lim_{h\to 0} e^h$ is 1. This is a familiar property of e^x, but since we only assumed that we knew the power series expansion, it's worth noting.

It's not too hard to prove this—try comparing its series to the series $\sum \frac{x^n}{2^{n-1}}$. First, notice that term-wise, we have $\frac{x^n}{n!} < \frac{x^n}{2^{n-1}}$, for any n: this is because on the one side you're taking the product $1 \cdot 2 \cdot \ldots n$, which grows much faster than just $2 \cdot 2 \cdot 2 \ldots \cdot 2$. Now, notice that we can actually evaluate the second series: $\sum \frac{x^n}{2^{n-1}}$ is just a geometric series that converges to $2 \cdot \frac{1}{1-\frac{x}{2}}$. (In general, the geometric series $\sum_{n=}^{\infty} y^n$ converges to $\frac{1}{1-y}$ whenever $0 < y < 1$: prove this!) When $x \to 0$ in the above expression, the expression goes to 1; so our larger series goes to 1 as x goes to 0. Therefore, by the squeeze theorem, because we can also bound it below by just its first term 1, the e^x series must also go to 1 as $x \to 0$.

In general, in your work in this class, you can assume familiarity with the exponential function and its basic properties (e.g. the ones you find in the book). However, it's nice to know that, if you wanted to, you could derive all of the familiar properties rigorously using our new tools and definitions.

3. Differentiation: Tools

Here are a few fundamental results that we will use constantly in this class to study differentiable functions.

Proposition 1. *For f, g a pair of functions differentiable at a and α, β a pair of constants,*

$$(\alpha f(x) + \beta g(x))'(a) = \alpha f'(a) + \beta g'(a).$$

Proposition 2. *For f, g a pair of functions differentiable at a,*

$$(f(x) \cdot g(x))'(a) = f'(a) \cdot g(a) + g'(a) \cdot f(a).$$

Proposition 3. *For f a function differentiable at $g(a)$ and g a function differentiable at a,*

$$(f(g(x)))'(a) = f'(g(a)) \cdot g'(a).$$

Much like for continuous functions, most of your derivative calculations should come through the results we've established in class (how to take derivatives of trig functions, polynomials, and e^x) and just algebra with the above three rules. (You could, in theory, work out everything single such thing from the beginning straight from the definitions, but of course, we want to save ourselves some work.)

Here's a quick illustration of how you might go about this.

Example 4. For any positive a,

$$(\ln(x))'(a) = \frac{1}{a}.$$

Proof. To do this, we should probably define what $\ln(x)$ is! Define $\ln(x)$ as the function from $(0, \infty)$ to $\mathbb{R}$, that takes $x \in \mathbb{R}$ to the unique value $y \in \mathbb{R}$ such that $e^y = x$. That is, $\ln(x)$ is the function that undoes e^x: it is the unique function such that $e^{ln(x)} = x$, for any positive x. (Note that mathematicians usually use $\log(x)$ to mean $\ln(x)$. We almost never use base 10 logarithms.)

Using this definition, examine the quantity $e^{\ln(x)}$. On one hand, we know that this function is just x by definition; so

$$\left(e^{\ln(x)}\right)' = (x)' = 1.$$

On the other hand,

$$\begin{aligned}\left(e^{\ln(x)}\right)' &= e^{\ln(x)} \cdot (\ln(x))', \text{ (by the chain rule)} \\ &= x \cdot (\ln(x))'.\end{aligned}$$

Equating both sides, we have that

$$\begin{aligned} 1 &= x \cdot (\ln(x))' \\ \Rightarrow \frac{1}{x} &= (\ln(x))'.\end{aligned}$$

So we're done! No epsilons or deltas or convergence tests needed! (Of course, they're all working very hard in the background.) □