LITTLE o'S AND BIG $\mathcal{O}$'S.

Suppose you have a function $f(x)$ with $f(a) = 0$ and you want to consider how quickly the function goes to zero around a. Then ideally you would want to find a simple function g (for example $g(x) = (x-a)^n$) which also vanishes at a such that g and f are almost equal around a. The little o and big O notation, want to express something like that, but only state that f goes to zero *faster* than g. For error terms this is of course sufficient (you just want to know that the error term is small), so they are used mostly in that context.

Definition 1. *We say $f(x) = \mathcal{O}(g(x))$ as $x \to a$ if there exists a constant M such that $|f(x)| \leq M|g(x)|$ in some punctured neighborhood of a, that is for $x \in (a-\delta, a+\delta) \setminus \{a\}$ for some value of δ.*

We say $f(x) = o(g(x))$ as $x \to a$ if $\lim_{x\to a} \frac{f(x)}{g(x)} = 0$. This implies that there exists a punctured neighborhood of a on which g does not vanish.

As notation this means that whenever we write $\mathcal{O}(x^4)$ we mean an unspecified function $h(x)$ which satisfies the requirement that $h(x) = \mathcal{O}(x^4)$. Likewise, whenever we write $o(x^4)$ you should read that as denoting a further unspecified function $h(x)$ which satisfies $h(x) = o(x^4)$. This means that we don't quite know what this error term is, but we have singled out its most important property (how small it is around the point we are looking at).

This allows us to write the error term for the Taylor polynomial approximation in a simple way

Theorem 1. *Let $f : U \to \mathbb{R}$ be an $(n+1)$-times continuously differentiable map. Let $P_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k$ be the n-th order Taylor polynomial of f at a, then*

$$f(x) = P_n(x) + \mathcal{O}((x-a)^{n+1}).$$

If $f : U \to \mathbb{R}$ is only n-times continuously differentiable, then we find

$$f(x) = P_n(x) + o((x-a)^n).$$

Inversely, suppose $Q_n(x)$ is a polynomial of degree at most n, such that

$$f(x) = Q_n(x) + o((x-a)^n)$$

then Q_n is the n-th order Taylor polynomial of f around a.

In writing this down you get to choose what the value of n is: the larger you choose n, the better your approximation of the function is (the error term becomes smaller, at least around the point a), but the more complicated your approximation is (it is a polynomial of higher degree).

Let us now write down some examples; the first two are derived from Taylor polynomials, the rest can be checked directly.

- We have $e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \mathcal{O}(x^4)$ as $x \to 0$;
- We have $\frac{1}{1-x} = 1 + x + x^2 + \mathcal{O}(x^3) = 1 + x + x^2 + o(x^2)$ as $x \to 0$;
- We have $|x|^3 = \mathcal{O}(x^3) = o(x^2)$ as $x \to 0$;

- We have $\cosh(x) = \mathcal{O}(e^x) = o(e^{\frac{5}{4}x})$ as $x \to \infty$;
- We have $1/\sin(x) = \mathcal{O}(1/x) = o(1/x^{3/2})$ as $x \to 0$;

The difference between big $\mathcal{O}$ and little o is exemplified here. $o(x^2)$ is smaller than $\mathcal{O}(x^2)$, so using little o tells you more than using big $\mathcal{O}$; but in all these examples we had to put bigger functions after the o than after the $\mathcal{O}$, in which case it says less after all. That is saying the error is $o(x^2)$ is stronger than saying the error is $\mathcal{O}(x^2)$ (as $x \to 0$), but unfortunately, often you have to choose between saying the error is $o(x^2)$ or $\mathcal{O}(x^3)$ (as $x \to 0$), in which case the latter is the strongest statement.

We want to use these error behaviors to simplify our calculations. Therefore it is vital that we can simply calculate with them. The following rules hold for calculating with these errors

Theorem 2. *(1) $f(x) = \mathcal{O}(f(x))$;*

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

For example the fourth rule should be read as "the sum of any two functions h and k, where $h = o(f(x))$ and $k = o(g(x))$ is itself $h + k = o(g(x))$ (as long as $f(x) = \mathcal{O}(g(x))$)". Do not ever divide simply by some $o(g(x))$ or $\mathcal{O}(g(x))$, as we only know these terms are small, they could be 0. The proofs are included for completeness, they are not that interesting, and you may just skim or skip them.

Proof. The proofs are all really basic, consisting mostly of painstakingly checking the definitions.

(1) Indeed $|f(x)| \leq 1 \cdot |f(x)|$, i.e. you can just use the bound $M = 1$;
(2) Indeed if $\lim_{x\to a} f(x)/g(x) = 0$ then there for $\epsilon = 1$ there exists a δ, such that if $0 < |x - a| < \delta$, then $|f(x)/g(x)| < 1$, so $|f(x)| < |g(x)|$.
(3) If $h(x) = \mathcal{O}(f(x))$ and $j(x) = \mathcal{O}(g(x))$, then there exist constants M_1, M_2, and M_3 such that $|h(x)| \leq M_1|f(x)|$, $|j(x)| \leq M_2|g(x)|$, and $|f(x)| \leq M_3|g(x)|$ in a neighborhood around a. We then conclude that $|h(x) + j(x)| \leq |h(x)| + |j(x)| \leq M_1|f(x)| + M_2|g(x)| \leq M_1M_3|g(x)| + M_2|g(x)| = (M_1M_3 + M_2)|g(x)|$. Thus $h(x) + j(x) = \mathcal{O}(g(x))$;
(4) Similarly if $h = o(f(x))$ and $j = o(g(x))$ then (with again M_3 such that $|f(x)| \leq M_3|g(x)|$ in a neighborhood of a) we have $|h(x)/g(x)| = |h(x)/f(x)||f(x)/g(x)| \leq M_3|h(x)/f(x)|$, so if $\lim_{x\to a} h(x)/f(x) = 0$ we have $\lim_{x\to a} h(x)/g(x) = 0$ by the squeezing lemma. We conclude $\lim_{x\to a}(h(x) + j(x))/g(x) = \lim_{x\to a} h(x)/g(x) + \lim_{x\to a} j(x)/g(x) = 0 + 0 = 0$;
(5) If $h(x) = \mathcal{O}(g(x))$ then there exists M such that $|h(x)| \leq M|g(x)|$, so $|ch(x)| = c|h(x)| \leq cM|g(x)|$, so $ch(x) = \mathcal{O}(g(x))$ as well. Likewise if $h(x) = o(g(x))$, then we have $\lim_{x\to a} ch(x)/g(x) = c \lim_{x\to a} h(x)/g(x) = c \cdot 0 = 0$, so $ch(x) = o(g(x))$.
(6) If $h(x) = \mathcal{O}(f(x))$ and $j(x) = \mathcal{O}(g(x))$ there exist M_1 and M_2 such that in a neighborhood of a we have $|h(x)| \leq M_1|f(x)|$ and $|j(x)| \leq M_2|g(x)|$.

We conclude that in this neighborhood we have $|h(x)j(x)| = |h(x)||j(x)| \leq M_1M_2|f(x)||g(x)| = M_1M_2|f(x)g(x)|$. Thus $h(x)j(x) = \mathcal{O}(f(x)g(x))$;

(7) If $h(x) = o(f(x))$ and $j(x) = \mathcal{O}(g(x))$ then we know that $\lim_{x\to a} h(x)/f(x) = 0$ and that $j(x)/g(x)$ is bounded. By a previous homework problem we have seen that this implies that $\lim_{x\to a} \frac{h(x)}{f(x)} \cdot \frac{j(x)}{g(x)} = 0$ as well.

(8) To prove $1/(1 + o(g(x))) = 1 + o(g(x))$, suppose $h(x) = o(g(x))$, then

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

In the last step we used that we know the limits of $h(x)/g(x)$ and of $g(x)/1$ as $x \to a$ and that these limits equal 0. In particular we see that $\frac{1}{1+h(x)} - 1 = o(g(x))$. Now if $h(x) = \mathcal{O}(g(x))$ we choose M such that $|h(x)| \leq M|g(x)|$ in a neighborhood of a. Then we get

$$\frac{1}{1+h(x)} - 1 = \frac{h(x)}{1+h(x)} = h(x)\frac{1}{1+\frac{h(x)}{g(x)}g(x)}$$

Since $g(x) = o(1)$ we see that $|g(x)| < \frac{1}{2M}$ in a close enough neighborhood of a. Then $\left|\frac{h(x)}{g(x)}g(x)\right| < M\frac{1}{2M} = \frac{1}{2}$ in a neighborhood around a. Thus $\left|\frac{1}{1+\frac{h(x)}{g(x)}g(x)}\right| < 2$ in this neighborhood. We conclude that in this neighborhood $\left|\frac{1}{1+h(x)} - 1\right| < 2|h(x)| < 2M|g(x)|$, so it is $\mathcal{O}(g(x))$.

□

In the case the functions f and g are polynomials these rules simplify to the following

Proposition 1. *Around 0 we have*

(1) $x^a = \mathcal{O}(x^b)$ *for all* $b \leq a$ *and* $x^a = o(x^b)$ *for all* $b < a$;

(2) $\mathcal{O}(x^a) + \mathcal{O}(x^b) = \mathcal{O}(x^{\min(a,b)})$, $o(x^a) + o(x^b) = o(x^{\min(a,b)})$, *and*

$$\mathcal{O}(x^a) + o(x^b) = \begin{cases} o(x^b) & b < a \\ \mathcal{O}(x^a) & b \geq a \end{cases}$$

(3) $c\mathcal{O}(x^a) = \mathcal{O}(x^a)$ *and* $c\, o(x^a) = o(x^a)$;

(4) $x^b\mathcal{O}(x^a) = \mathcal{O}(x^{a+b})$, *and* $x^b o(x^a) = o(x^{a+b})$;

(5) $\mathcal{O}(x^a)\mathcal{O}(x^b) = \mathcal{O}(x^{a+b})$, *and* $\mathcal{O}(x^a)o(x^b) = o(x^{a+b})$, *and* $o(x^a)o(x^b) = o(x^{a+b})$

This allows us to find some simple rules for calculating Taylor polynomials.

Theorem 3. *If f and g are n times continuously differentiable and $P_n(x)$, respectively $Q_n(x)$, is the n-th order Taylor polynomial of f, respectively g, around a, then the n-th order Taylor polynomial of $f(x)g(x)$ is given by $P_n(x)Q_n(x)$ with the terms $(x-a)^s$ with $s > n$ omitted.*

Proof. Write $R_n(x)$ for the n-th degree polynomial obtained by removing the terms $(x-a)^s$ with $s > n$ from $P_n(x)Q_n(x)$. We see that $R_n(x) = P_n(x)Q_n(x) + \mathcal{O}((x-a)^{n+1}) = P_n(x)Q_n(x) + o((x-a)^n)$. Now we see that $f(x) = P_n(x) + o((x-a)^n)$ and $g(x) = Q_n(x) + o((x-a)^n)$. Hence

$$f(x)g(x) = P_n(x)Q_n(x) + P_n(x)o((x-a)^n) + Q_n(x)o((x-a)^n) + o((x-a)^n)o((x-a)^n)$$
$$= R_n(x) - o((x-a)^n) + o((x-a)^n) + o((x-a)^n) + o((x-a)^{2n}) = R_n(x) + o((x-a)^n).$$

Thus we can use the inverse characterization of the Taylor polynomials to conclude that R_n is the n-th order Taylor polynomial of $f(x)g(x)$ around a. $\square$

In a similar way we get the Taylor polynomial for compositions of functions.

Theorem 4. *Suppose f and g are n times continuously differentiable. Let $P_n(x)$ be the n-th order Taylor polynomial of f around $g(a)$ and $Q_n(x)$ the n-th order Taylor polynomial of g around a. Then the n-th degree polynomial obtained by removing terms $(x-a)^s$ for $s > n$ from $P_n(Q_n(x))$ is the n-th order Taylor polynomial of $f(g(x))$.*

Proof. Write $R_n(x)$ for the polynomial obtained by removing terms $(x-a)^s$ for $s > n$ from $P_n(Q_n(x))$. Then $R_n(x) = P_n(Q_n(x)) + o((x-a)^n)$ as before. Now we see that

$$f(g(x)) = P_n(Q_n(x) + o((x-a)^n)) + o((Q_n(x) - g(a))^n)$$

First we notice that $Q_n(a) = g(a)$, as Q_n is the Taylor polynomial of g around a, so $Q_n(x) - g(a) = \mathcal{O}(x-a)$. Therefore we get that $o((Q_n(x) - g(a))^n) = o((x-a)^n)$. Then we observe that $P_n(Q_n(x) + o((x-a)^n))$ equals a sum of terms of the form $(x-a)^k o((x-a)^n)^l$ for some $k, l \geq 0$. Thus either $l = 0$, there are no $o((x-a)^n)$ terms, and we get some term from $P_n(Q_n(x))$, or $l > 0$ and this term is $o((x-a)^n)$ (or even smaller). We conclude that $P_n(Q_n(x) + o((x-a)^n)) = P_n(Q_n(x)) + o((x-a)^n)$. Together we get that

$$f(g(x)) = P_n(Q_n(x)) + o((x-a)^n) = R_n(x) + o((x-a)^n).$$

By the inverse characterization of the Taylor polynomials we see that R_n is the n-th order Taylor polynomial of $f(g(x))$. $\square$

Let us now give some examples.

- $\sin(x)\cos(x)$: We know $\sin(x) = x - \frac{x^3}{6} + o(x^4)$ and $\cos(x) = 1 - \frac{1}{2}x^2 + \frac{1}{24}x^4 + o(x^4)$. Thus their product is (where we throw all terms $o(x^4)$ and smaller in one pile in the second step)

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

Thus the fourth order Taylor polynomial is $x - \frac{2}{3}x^3$. In further calculations I will immediately remove the terms which combine into the error term (in this case $o(x^4)$). Another, lot quicker way of calculating this is by using the doubling formula for the sine.

$$\sin(x)\cos(x) = \frac{1}{2}\sin(2x) = \frac{1}{2}((2x) - \frac{(2x)^3}{6} + o((2x)^4)) = x - \frac{8}{2\cdot 6}x^3 + o(x^4).$$

Fortunately the result remains the same.

- $\cos(\sqrt{1+\log(1+x)}-1)$ around $x=0$. You don't really want to differentiate this function a few times to calculate its Taylor polynomial. Fortunately we can first write (differentiating $\log(1+x)$)

$$\log(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + o(x^4)$$

Then taking the square root we see that (differentiating $\sqrt{1+u}$)

$$\sqrt{1+u} = 1 + \frac{1}{2}u - \frac{1}{8}u^2 + \frac{1}{16}u^3 - \frac{5}{128}u^4 + o(u^4)$$

Plugging in $u = \log(1+x)$ gives us

$$\begin{aligned}
\sqrt{1+\log(1+x)} = 1 &+ \frac{1}{2}\left(x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + o(x^4)\right) - \frac{1}{8}\left(x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + o(x^4)\right)^2 \\
&+ \frac{1}{16}\left(x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + o(x^4)\right)^3 - \frac{5}{128}\left(x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + o(x^4)\right)^4 \\
&+ o((x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + o(x^4))^4) \\
= 1 &+ \frac{1}{2}\left(x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + o(x^4)\right) - \frac{1}{8}\left(x^2 - x^3 + \frac{11}{12}x^4 + o(x^4)\right) \\
&+ \frac{1}{16}\left(x^3 - \frac{3}{2}x^4 + o(x^4)\right) - \frac{5}{128}\left(x^4 + o(x^4)\right) + o(x^4) \\
= 1 &+ \frac{1}{2}x + (-\frac{1}{2}\cdot\frac{1}{2} - \frac{1}{8}\cdot 1)x^2 + (\frac{1}{2}\cdot\frac{1}{3} + \frac{1}{8}\cdot 1 + \frac{1}{16}\cdot 1)x^3 \\
&+ (-\frac{1}{2}\cdot\frac{1}{4} - \frac{1}{8}\cdot\frac{11}{12} - \frac{1}{16}\frac{3}{2} - \frac{5}{128}\cdot 1)x^4 + o(x^4) \\
= 1 &+ \frac{1}{2}x - \frac{3}{8}x^2 + \frac{17}{48}x^3 - \frac{143}{384}x^4 + o(x^4).
\end{aligned}$$

Thus we subtract 1 and plug this in in

$$\cos(v) = 1 - \frac{1}{2}v^2 + \frac{1}{24}v^4 + o(v^4)$$

to get (note $o(v^4) = o(x^4)$, so we can simplify that term immediately)

$$\begin{aligned}
\cos(\sqrt{1+\log(1+x)}-1) &= 1 - \frac{1}{2}\left(\frac{1}{2}x - \frac{3}{8}x^2 + \frac{17}{48}x^3 - \frac{143}{384}x^4 + o(x^4)\right)^2 \\
&\quad + \frac{1}{24}\left(\frac{1}{2}x - \frac{3}{8}x^2 + \frac{17}{48}x^3 - \frac{143}{384}x^4 + o(x^4)\right)^4 + o(x^4) \\
&= 1 - \frac{1}{2}\left(\frac{1}{4}x^2 - \frac{3}{8}x^3 + \frac{95}{192}x^4 + o(x^4)\right) + \frac{1}{24}\left(\frac{1}{16}x^4 + o(x^4)\right) + o(x^4) \\
&= 1 - \frac{1}{8}x^2 + \frac{3}{16}x^3 - \frac{47}{192}x^4 + o(x^4)
\end{aligned}$$

Admittedly the calculation here was somewhat tedious (the complexity quickly increases for higher orders of approximation, if you want just the third order term we already have to do less than half the work), but we did not have to calculate any ugly derivatives. I challenge you to try and differentiate $\cos(\sqrt{1+\log(1+x)}-1)$, which might well be more work. Moreover the power of the method lies in the fact that, as long as you know low order approximations of the basic functions appearing, you know it of all functions you can make with these functions.

It should be noted that not all relevent kinds of behavior as $x \to 0$ can be expressed by powers of x; Other terms you will often see are $o(\log(x)^k)$. Around infinity the relevant behaviors are, apart from polynomial growth/decay ($\mathcal{O}(x^n)$) (growth for $n > 0$, decay for $n < 0$), exponential growth/decay $\mathcal{O}(e^{ax})$ (for $a > 0$, respectively $a < 0$), and logarithmic growth/decay ($\mathcal{O}(\log(x)^n)$). The exponential behaviour is growing or decaying fastest, while logarithmic is slowest, with polynomial growth somewhere in between. If $a, n, k > 0$ are positive numbers we get the sequence (with the left hand side being the smallest and the right hand side the largest).

$$\mathcal{O}(e^{-ax}) = \mathcal{O}(x^{-n}) = \mathcal{O}(\log(x)^{-k}) = \mathcal{O}(1) = \mathcal{O}(\log(x)^k) = \mathcal{O}(x^n) = \mathcal{O}(e^{ax}).$$

That is: any function that is $\mathcal{O}(e^{-ax})$ is also $\mathcal{O}(x^{-n})$, and any function that is $\mathcal{O}(x^{-n})$ is also $\mathcal{O}(\log(x)^{-k})$, etc. (But not the other way around!)