

Taylor's theorem and big-oh notation

Course outcomes

Course outcome	What it is about
<i>Scientific Literacy</i>	I can explain and give reference to important results about numerical methods and theory in a precise manner with references.
<i>Core Knowledge</i> (Facts and Intuition)	I can explain important facts about numerical methods and the intuition behind them.
<i>Analysis</i>	I can give rigorous analysis of numerical methods.
<i>Computation</i>	I can write programs that implement numerical methods in an efficient and collaborative way, including analyzing and evaluating working programs.

Take-aways

After studying these notes, we will be able to

- use Taylor's theorem in an effective way, (*Analysis*)
 - choose a 'right' order of Taylor polynomial for use (trial and error is okay)
 - check the smoothness (a.k.a. 'regularity') condition,
 - draw a meaningful conclusion by using the theorem correctly,
- translate big-oh notations to their precise definition, (*Scientific Literacy*)
 - state the definition with help of reference,
 - check whether specific examples satisfy the definition,
- explain what a big-oh notation tells us, (*Core Knowledge*)
 - describe the main tendency of a quantity in an intuitive language,
 - tell which part is negligible and which part matters,
 - use correct big-oh notation after operations.

Taylor's theorem

Taylor's theorem

Theorem (Taylor's theorem with Lagrange remainder)

Suppose $\delta > 0$ and a real valued function f defined on $I = (x_0 - \delta, x_0 + \delta)$ is $k + 1$ times differentiable. Then, for any $x \in I$, we have

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \dots + \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + \frac{f^{(k+1)}(\xi)}{(k+1)!}(x - x_0)^{k+1},$$

where $\xi \in (x, x_0)$ if $x < x_0$ or $\xi \in (x_0, x)$ if $x_0 < x$.

[proof of Taylor's theorem with Lagrange remainder 1](#)

[proof of Taylor's theorem with Lagrange remainder 2](#)

Big-oh notation

Motivating examples

Example: Big-oh notation with exponential in easy language

Example (Taylor series of exponential - easy language)

Recall from calculus, for any $x \in \mathbb{R}$, we have

$$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \dots + \frac{1}{n!}x^n + \dots$$

Note that this "equality" is "more true" near $x_0 = 0$ since the series is expanded around $x_0 = 0$. For example, if $x = 0.1$, then the sum becomes

$$e^{0.1} = 1 + 0.1 + \frac{0.01}{2} + \frac{0.001}{6} + \frac{0.0001}{24} + \dots$$

After the first few terms, the magnitude of the remaining terms are so small that it does not change the whole sum much.

	partial sum	error
0	1.000000	1.051709e-01
1	1.100000	5.170918e-03
2	1.105000	1.709181e-04
3	1.105167	4.251409e-06
4	1.105171	8.474231e-08
5	1.105171	1.408981e-09
6	1.105171	2.009237e-11
7	1.105171	2.511324e-13
8	1.105171	3.108624e-15
9	1.105171	4.440892e-16
10	1.105171	4.440892e-16

► code

For example, after eye-ball the table above, one may want to focus on the first four terms (the rows number 0 through 3) and thinks "just remember there are some other terms, but they are minor". We denote this as

$$\begin{aligned}
 e^x &= 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \frac{1}{24}x^4 + \dots \\
 &= 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \mathcal{O}(x^4).
 \end{aligned}$$

The quotation marked part corresponds to the big-oh part.

Definition

There are slightly different variants of the definitions. We follow essentially Strichartz's definitions as defined in his excellent book (p. 147) [\[1\]](#).

Definition: Big-oh for finite quantities

Definition (Big oh for finite quantities)

Let $x_0 \in \mathbb{R}$ and f and g are real valued functions defined near x_0 . If there exist (fixed) constants $C > 0$ and $\delta > 0$ such that

$$|f(x)| \leq C|g(x)| \quad \text{for all } x \in (x_0 - \delta, x_0 + \delta),$$

or,

$$\frac{|f(x)|}{|g(x)|} \leq C \quad \text{for all } x \in (x_0 - \delta, x_0 + \delta) \setminus \{x_0\},$$

then we write

$$f(x) = \mathcal{O}(g(x)) \quad \text{as } x \rightarrow x_0,$$

and say $f(x)$ is a big oh of $g(x)$ near x_0 or $f(x)$ is of order of $g(x)$ near x_0 .

Definition: Big-oh for growing quantities

Definition (Big oh for growing quantities)

Let f and g are real valued functions defined on near ∞ (i.e., $[N, \infty)$ for some $N \in \mathbb{R} \cup \{-\infty\}$). If there exist (fixed) constants $C > 0$ and $M > 0$ such that

$$|f(x)| \leq C|g(x)| \quad \text{for all } x \in [M, \infty),$$

then we write

$$f(x) = \mathcal{O}(g(x)) \quad \text{as } x \rightarrow \infty,$$

and say $f(x)$ is a big oh of $g(x)$ as x grows or $f(x)$ is of order of $g(x)$ as x grows.

Notation (error form of big-oh)

Let f, g, h be real valued functions defined near x_0 . If $f(x) - g(x) = \mathcal{O}(h(x))$ as $x \rightarrow x_0$, then we write

$$f(x) = g(x) + \mathcal{O}(h(x)) \quad \text{as } x \rightarrow x_0.$$

Interpretation

Interpretation (Controlled/bounded by a simpler function)

Suppose $f(x) = \mathcal{O}(h(x))$ as $x \rightarrow x_0$. To use big-oh notation in a useful way, we usually set:

f	h
function we want to study	a simpler function than f

Then, the big-oh condition says that *the magnitude of f can be controlled or bounded by g* .

Interpretation (error form of big-oh)

The error form is the most common form of big-oh (e.g., $f(x) = g(x) + \mathcal{O}(h(x))$ as $x \rightarrow x_0$). Intuitively, we can interpret it as follows:

- *the error caused by $g(x)$ in place of $f(x)$ is no worse than $h(x)$ (up to constant multiple), or*
- *$f(x)$ is made up of $g(x)$ and something that behaves like $h(x)$ (up to constant multiple), which is assumed to be small.*

Example: Big-oh notation with cosine

Example (Taylor series of cosine)

Recall from calculus, for any $x \in \mathbb{R}$, we have

$$\cos x = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots,$$

where the equality is in the sense of infinite sum: the sum on the right hand side converges to the left hand side. Note that this "equality" is "more true" near $x_0 = 0$ since the series is expanded around $x_0 = 0$. For example, if $x = 0.1$, then the sum becomes

$$\cos(0.1) = 1 - \frac{0.01}{2} + \frac{0.0001}{24} - \frac{0.000001}{720} + \dots,$$

After the first few terms, the magnitude of the remaining terms are so small that it does not change the whole sum much. For example, one may want to focus on the first three terms and write

$$\begin{aligned} \cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots \\ &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \mathcal{O}(x^6). \end{aligned}$$

The last equality is not trivial because it involves infinitely many terms, but it is true (See [Justification of lumping infinite sum](#)).

Question

Give the interpretation of the big-oh notation in the previous example. (Type in a short answer)

(Reminder) This is **about atmosphere**, not getting it right.

1. Think for a short time.
2. Share your guess with your pair.
3. Type your answer in clicker.
4. Feel free to say out loud.

Example: Big-oh notation for complexity of matrix multiplication

Example (complexity of matrix multiplication)

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

Suppose we want to multiply 3×3 matrix and 3×1 column vector. Let us count how many operations are needed to do this. To get a , we need 3 (real number) multiplications and 2 (real number) additions. We need the same amount of computation for b and c . If we increase the sizes to 4×4 (matrix) and 4×1 (vector), or more generally, to $n \times n$ (matrix) and $n \times 1$ (vector), we need the following number of operations.

size	(\times)	($+$)	total
3	$3 \times 3 = 9$	$2 \times 3 = 6$	15
4	$4 \times 4 = 16$	$3 \times 4 = 12$	28
n	?	?	?

Question

Fill the table and find the best big-oh notation.

Answer choice	big-oh
(A)	$\mathcal{O}(n)$
(B)	$\mathcal{O}(n^2)$
(C)	$\mathcal{O}(n^3)$
(D)	I don't see a good answer.

(Reminder) This is **about atmosphere**, not getting it right.

1. Think for a short time.

2. Share your guess with your pair.
3. Type your answer in clicker.
4. Feel free to say out loud.

Properties

We state and prove some frequently used properties of big-oh relations for the finite quantities (see the boardwork for the proofs). Similar results hold also for the growing quantities.

Proposition (sum)

Suppose $f_1(x) = \mathcal{O}(g_1(x))$, $f_2(x) = \mathcal{O}(g_2(x))$, and $g_1(x) = \mathcal{O}(g_2(x))$ as $x \rightarrow x_0$. Then,

$$f_1(x) + f_2(x) = \mathcal{O}(g_2(x)) \quad \text{as } x \rightarrow x_0.$$

Proof of sum of big-oh's

Example of sum of big-oh's

Corollary (sum of the same big-oh terms)

Suppose $f_1(x) = \mathcal{O}(g(x))$ and $f_2(x) = \mathcal{O}(g(x))$ as $x \rightarrow x_0$. Then,

$$f_1(x) + f_2(x) = \mathcal{O}(g(x)) \quad \text{as } x \rightarrow x_0.$$

Proof of sum of big-oh's: same order

Example of sum of big-oh's: same order

Proposition (product)

Suppose $f_1(x) = \mathcal{O}(g_1(x))$ and $f_2(x) = \mathcal{O}(g_2(x))$ as $x \rightarrow x_0$. Then,

$$f_1(x)f_2(x) = \mathcal{O}(g_1(x)g_2(x)) \quad \text{as } x \rightarrow x_0.$$

Proof of product of big-oh's

Corollary (product directly added to order)

Suppose $f(x) = \mathcal{O}(g(x))$ as $x \rightarrow x_0$. Then,

$$f(x)h(x) = \mathcal{O}(g(x)h(x)) \quad \text{as } x \rightarrow x_0.$$

Proof of product of big-oh's: directly added to order

Examples

Example: Application of Taylor's theorem

Example (Accuracy of difference quotient)

Suppose that $f : \mathbb{R} \rightarrow \mathbb{R}$ is continuously differentiable near $x_0 \in \mathbb{R}$. Then, we have

$$\frac{f(x_0 + h) - f(x_0)}{h} = f'(x_0) + \mathcal{O}(h)$$

Interpretation

Accuracy of difference quotient

Question

$$\begin{aligned}\cos x &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \mathcal{O}(x^6) \quad \text{as } x \rightarrow 0. \\ c(n) &= 2n^2 - n = \mathcal{O}(n^2) \quad \text{as } n \rightarrow \infty,\end{aligned}$$

Give a second look at the examples of $\cos(x)$ and the complexity of matrix multiplication. Which part of each of those quantities is negligible and which part matters?

(Reminder) This is **about atmosphere**, not getting it right.

1. Think for a short time.
2. Share your guess with your pair.
3. Type your answer in clicker.
4. Feel free to say out loud.

Remarks

Remark: Advantage of big-oh

Remark (Advantage of big-oh)

Big-oh notation allows us to forget about unimportant details (i.e., exact behaviors of a function) and to focus on what really matters (i.e., what is the magnitude of the function like) when it is helpful.

Remark: Big-oh term may or may not matter

Remark (Big-oh term may or may not matter)

Sometimes big-oh term is the one that matters. This is typically true for growing quantities.

Sometimes, big-oh term is the one that is negligible. This is usually the case for finite quantities.

Remark: Big-oh is about inequality, not equality

Warning (Big-oh is about inequality, not equality)

Consider $f(x) = x^3$ and $g(x) = x^2$ around $x_0 = 0$. If x is close enough to $x_0 = 0$, say $\delta = 0.1$ (hence $x \in (-0.1, 0.1)$), we have

$$|f(x)| = |x|^3 \leq 0.1|x|^2.$$

Using big-oh notation, it follows

$$f(x) = \mathcal{O}(g(x)),$$

with the constant $C = 0.1 > 0$ appearing in the definition.

However, we do not have " $g(x) = \mathcal{O}(f(x))$ as $x \rightarrow 0$."

To see this, observe that it is impossible to choose a fixed constant $C > 0$ that satisfies "for all $x \in (-0.1, 0.1)$, $|x|^2 \leq C|x|^3$ ": taking smaller and smaller x 's, $C|x|^3$ eventually becomes smaller than $|x|^2$ no matter how big C you choose.

Even if we try different $\delta \neq 0.1$, the situation is not changed. This is because the violation of big-oh condition (i.e., inequality flip) happens due to x 's that are close to 0. Thus we conclude

$$g(x) \neq \mathcal{O}(f(x)) \quad \text{as } x \rightarrow 0.$$

Observe, however, as you can see from the above reasoning, big-oh encapsulates inequalities.

Remark: The order of error term depends on the context

Remark (The order of error term depends on the context)

Consider the following example. If we want to focus on the terms up to 2nd order, we write

$$\begin{aligned} e^x &= 1 + x + \frac{1}{2}x^2 + \underbrace{\frac{1}{6}x^3 + \cdots + \frac{1}{n!}x^n + \cdots}_{= \mathcal{O}(|x|^3)} \\ &= 1 + x + \frac{1}{2}x^2 + \mathcal{O}(|x|^3) \quad \text{as } x \rightarrow 0. \end{aligned}$$

However, while continuing working on, we may need more terms. If it turns out we need to include one more term, we write

$$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \mathcal{O}(|x|^4) \quad \text{as } x \rightarrow 0.$$

Both are true. But one thing may not be useful while the other is. It depends the situation.

Appendix

Justification of lumping infinite sum

In the [example of Taylor series of exponential](#), we claimed

$$\frac{1}{6}x^3 + \frac{1}{24}x^4 + \dots = \mathcal{O}(|x|^3) \quad \text{as } x \rightarrow 0.$$

Let us justify this.

(step 1)

We prove:

$$\frac{g(x)}{x^3} \rightarrow C \quad \text{as } x \rightarrow 0,$$

where $C > 0$ is some constant and

$$g(x) := \frac{1}{6}x^3 + \frac{1}{24}x^4 + \dots.$$

for some $C > 0$. By the [theorem on power series](#),

$$\frac{g(x)}{x^3} = \frac{1}{6} + \frac{1}{24}x + \dots$$

is continuous on the interval of convergence, which is $(-\infty, \infty)$ in this case. Therefore,

$$\lim_{x \rightarrow 0} \frac{g(x)}{x^3} = \lim_{x \rightarrow 0} \left(\frac{1}{6} + \frac{1}{24}x + \dots \right) = \frac{1}{6} + 0 + 0 \dots = \frac{1}{6},$$

where we use the continuity of functions defined by power series in the second equality (see the [Theorem on power series](#)). Thus, this proves the (step 1) with $C = 1/6$.

(step 2)

Since $g(x)/x^3 \rightarrow C$ as $x \rightarrow 0$, by the definition of the limit, for $\epsilon = 1$, there exists $\delta > 0$ such that, for any $0 < |x| < \delta$, we have

$$\left| \frac{g(x)}{x^3} - C \right| < 1.$$

Multiplying through by $|x|^3$ and using the inverse triangle inequality, i.e. $|a| - |b| \leq |a - b|$ for all $a, b \in \mathbb{R}$, we have

$$|g(x)| - C|x|^3 \leq |g(x) - Cx^3| < |x|^3.$$

Rearranging, we obtain

$$g(x) \leq |g(x)| \leq (C + 1)|x|^3.$$

Q.E.D.

Error term via Taylor theorem

The most common situation where big-oh notation is used is when analyzing errors using Taylor theorem. Suppose that the context of the analysis somehow hints that knowing only the terms up to 2nd order is enough. Suppose you are studying the difference $f(x) - f(a)$ for some x close to a , say, $x \in (a - \delta, a + \delta)$. Then, by [Taylor theorem](#), we have

$$f(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \underbrace{\frac{f'''(\xi)}{3!}(x - a)^3}_{\mathcal{O}(|x-a|^3)},$$

for some $\xi \in (x, a)$ or $\xi \in (a, x)$ depending on $x < a$ or $a < x$. If $f'''(x)$ is continuous, then $|f'''(x)|$ attains the maximum M_0 on a closed subinterval of $[a - \delta/2, a + \delta/2] \subset (a - \delta, a + \delta)$ ([Extreme value theorem](#)). Thus, for all $x \in (a - \delta/3, a + \delta/3)$, we have

$$\left| \frac{f'''(\xi)}{3!}(x - a)^3 \right| \leq C|x - a|^3,$$

where $C = M_0/3!$. Therefore, we conclude

$$\frac{f'''(\xi)}{3!}(x - a)^3 = \mathcal{O}(|x - a|^3).$$

Theorem on power series

Theorem (Power series)^[2]

Suppose that f has a Taylor expansion around $x = a$ with a radius of convergence $R > 0$ or $R = \infty$, that is,

$$f(x) = \sum_{n=0}^{\infty} c_n (x - a)^n \quad \forall x \in (a - R, a + R).$$

Then, it is differentiable, hence continuous. Furthermore, its derivative and integral have the same radius of convergence as f and they are given by

$$f'(x) = \sum_{n=0}^{\infty} (n+1)c_n (x - a)^n \quad \forall x \in (a - R, a + R),$$

$$\int f(x) dx = C + \sum_{n=0}^{\infty} \frac{c_n}{n+1} (x - a)^{n+1} \quad \forall x \in (a - R, a + R).$$

Extreme value theorem

Theorem (Extreme value theorem)

Suppose f is continuous on a closed interval $[a, b]$. Then, it attains the maximum and minimum. That is, there are $x_1, x_2 \in [a, b]$ such that $f(x_1) = M$, $f(x_2) = m$, and $m \leq f(x) \leq M$ for all $x \in [a, b]$.

Question

What part would you make boldfaced in the Extreme value theorem if you were a professor? Why? If you feel like it, answer the same question to other theorems.

(Reminder) This is **about atmosphere**, not getting it right.

1. Think for a short time.
2. Share your guess with your pair.
3. Type your answer in clicker.
4. Feel free to say out loud.

Summary

- Big 'oh' notation allows us to focus only on the most important term out of (possibly) infinitely many terms. (See [Advantage of big-oh](#))
 - In some cases, all terms except the big-oh term matter. (See [Example: Big-oh notation with cosine](#))
 - In other cases, only the big-oh term matters. (See [Example: Big-oh notation for complexity of matrix multiplication](#))
- Big 'oh' is essentially *inequality*. It is not an equality, though its notation uses '=' sign. (See [Remark: Big-oh is about inequality, not equality](#))
- Eventually, people use big-oh notation at an intuitive level. But it is good to see it works rigorously as well to convince yourself. (See [Appendix](#) for some rigorous approaches)

1. Strichartz (2000) *The Way of Analysis*, Jones & Bartlett Learning. ↩
2. James Stewart (2012) *Calculus Early Transcendentals*, 7th Edition, Brooks Cole ↩