Big-oh stuff

Definition. You should know this definition *by heart* and be able to give it, if asked.

> Let f and g both be functions from R^+ to R^+. Then f is $O(g)$ (pronounced "big-oh") if and only if there exists positive constants C_1 and C_2 such that $f(n) \leq C_1 g(n)$ whenever $n > C_2$.

This is the *definition* of what it means to say f is $O(g)$. So, a real proof that f is $O(g)$ requires providing the constants C_1 and C_2 and proving the result above. Furthermore, this definition also applies for functions defined only on natural numbers, only on positive integers, etc. A small modification of the definition is made when the functions can have negative values (e.g., $f(n) = 3n^2 - 500$).

Most of the time you won't be asked to provide the constants, but rather to be able to guess intelligently (and back up your guess if asked) whether a function is big-oh of another function.

Comment 1: In some very mathematically oriented presentations, $O(g)$ is defined to be the set *of functions f such that $\exists C_1, C_2 \geq 0$ so that $f(n) \leq C_1 g(n)$ for all $n > C_2$. Therefore, you may find statements like "$f \in O(g)$" instead of "f is $O(g)$". Do not be surprised - these mean the same thing.*

Comment 2: Often people write "$f(n)$ is $O(g(n))$" instead of "f is $O(g)$". Technically, "$f(n)$" refers to the value of the function f on input n, rather than the function itself. However, by this point in time, both are acceptable ways of communicating the same point. Just don't be surprised when you see either one, as they mean the same thing.

Finding the positive constants C_1 and C_2. But now that you know the definition of what big-oh means, you can see the following statements will be true.

- **Technique 1:** Suppose $\exists C \geq 0$ so that

$$lim_{n\to\infty} \frac{f(n)}{g(n)} \leq C$$

  Then f is $O(g)$. Furthermore, for any $C' > C$, $\exists C_2 > 0$ such that $f(n) \leq C' g(n)$ for all $n > C_2$.

- **Technique 2:** Suppose $\exists C \geq 0$ such that

$$\lim_{n\to\infty} (\log_2 f(n) - \log_2 g(n)) \leq C$$

  Then f is $O(g)$. Furthermore, for any $C' > C, \exists C_2 > 0$ such that $f(n) \leq 2^{C'} g(n)$ for all $n > C_2$.

Comments:

1. Note that we require that $C' > C$. Why do we do this? The reason is that the limit may approach C from above, making it necessary to pick some constant bigger than whatever the limit is. For example if $f(n) = n^2 + 1$ and $g(n) = n^2$, then $f(n) > g(n)$ for all n, and yet $lim_{n \to \infty} \frac{f(n)}{g(n)} = 1$. If we set $C_1 = 1$, we definitely would not be able to find C_2 and prove that $f(n) \leq C_1 g(n)$ for $n > C_2$. On the other hand, for *any* $C_1 > 1$, we could find a corresponding value for C_2. For example, if we pick $C_1 = 2$, then we would need C_2 to satisfy $n^2 + 1 \leq 2n^2$ whenever $n \geq C_2$. This is equivalent to $n^2 \geq 1$, and so $C_2 = 1$ works.

2. For Technique 1, if $lim_{n \to \infty} \frac{f(n)}{g(n)} = C > 0$, then f and g are *both* big-oh of each other! (Note that saying that the limit is $C > 0$ means that C is a real number, and so by definition is not ∞.) This situation is described by saying that f is $\Theta(g)$ (from which it follows that g is $\Theta(f)$). The precise definition of $\Theta(g)$ is the set of functions f such that $\exists C_1, C_2, N$ positive real numbers such that whenever $n > N$ then $C_1 g(n) \leq f(n) \leq C_2 g(n)$. Thus we may also say $f \in \Theta(g)$ instead of saying f is $\Theta(g)$.

3. For Technique 1, if $lim_{n \to \infty} \frac{f(n)}{g(n)} = 0$, then f is $O(g)$ but the converse is not true.

4. For Technique 1, if $lim_{n \to \infty} \frac{f(n)}{g(n)} = \infty$, then f is not $O(g)$, but g is $O(f)$.

5. Note that for Technique 2, you can use any base you want for the log; this means that you can use ln (the natural log, which is $\log_e$), or log base 2 (written $\log_2$), or any base you like.

These two results may help you find the constant C_1, or determine that f is not $O(g)$! However, you will still need to find the constant C_2, and this will probably require that you know how to use logarithms, among other things.

To use these techniques you should be able to compute limits, something you may not right now be comfortable with. Furthermore, just knowing that the limit exists (which it may not) doesn't make it straightforward to pick the constant.

Example #1: Consider for example the following pair of functions

- $f(n) = n^2$
- $g(n) = n^3$

Let's use the techniques we've described to find the constants C_1 and C_2 to establish that f is $O(g)$.

If you compute $\lim_{n \to \infty} \frac{f(n)}{g(n)}$, you will get 0. But setting $C_1 = 0$ won't work. The reason is that that ratio approaches its limit *from above*. So you need to pick some constant greater than the limit, not equal to the limit. (Conversely, if the ratio approaches its limit from below, you can pick the constant C_1 to be that limit, but setting it to be bigger is always safe.)

So pick $C_1 = 1$, and then solve for C_2. This one is easy: $C_2 = 1$ works just fine.

Example #2: Here's another example, where it's a bit harder to figure out the answer.

- $f(n) = 3^{\sqrt{n}}$
- $g(n) = 2^n$

For this pair of functions, we wish to determine whether f is $O(g)$ and whether g is $O(f)$.

Part 1: Determining if f is $O(g)$.

Trying to figure out whether f is $O(g)$ using $\lim_{n\to\infty} \frac{f(n)}{g(n)}$ gives you something harder to compute. You'll need to use L'Hôpital's rule, but that may be something you are not comfortable with in this context. Let's try the second approach, which is based on logarithms. (Even this you may not be comfortable with!) Let's take logs using base 2. Then we get

- $\log_2 f(n) = \log_2(3^{\sqrt{n}}) = \sqrt{n}\log_2(3)$. Note that $1 < \log_2(3) < 2$, and so $\log f(n) < 2\sqrt{n}$.
- $\log_2 g(n) = log_2(2^n) = n$

We continue:

$$\log_2 f(n) - \log_2 g(n) = \sqrt{n}\log_2(3) - n < 2\sqrt{n} - n.$$

Note that when $n > 4$, $\sqrt{n} > 2$. Furthermore, when $n > 4$, it is easy to see that $2\sqrt{n} < n$. Hence,

$$\log_2 f(n) - \log_2 g(n) \leq 2\sqrt{n} - n < 0$$

How do we use this?

The analysis given above shows that if $\log f(n) - \log g(n) < \log C_1$ for large enough n, then f is $O(g)$. Therefore, if we can find a non-negative constant C_1 such that $\log C_1 > 0$, we will have established that f is $O(g)$.

What values of C_1 satisfy $\log C_1 > 0$? Answer: all $C_1 \geq 1$.

Setting $C_1 = 1$ then makes sense. What value would you give for C_2? The analysis here shows that $C_2 = 4$ works. Thus we have shown that f is $O(g)$.

Part 1: Determining if g is $O(f)$.

We will see if we can find the constants C_1 and C_2 such that $\log_2 g(n) - \log_2 f(n) < C_1'$ for all $n > C_2'$.

Recall that $2 > \log_2(3) > 1$, and that if $n > 9$ then $n > 3\sqrt{n}$. Therefore, for $n > 9$,

$$\log g(n) - \log f(n) = n - \sqrt{n}\log_2(3) > n - 2\sqrt{n} > \sqrt{n}$$

Since

$$\lim_{n\to\infty} \sqrt{n} = \infty,$$

it follows that

$$\lim_{n\to\infty} [\log_2 g(n) - \log_2 f(n)] \geq \lim_{n\to\infty} \sqrt{n} = \infty$$

Thus, g is not $O(f)$.

Example #3: Consider the following pair of functions:

- $f(n) = 100n^2$
- $g(n) = n^3 + 3$

From your training, you know that f is $O(g)$ but not vice-versa. We will use the first approach to find the constants C_1 and C_2 to prove that f is $O(g)$. Using L'Hôpital's Rule (applied twice!!), we find that

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = \lim_{n\to\infty} \frac{f'(n)}{g'(n)} = \lim_{n\to\infty} \frac{200n}{3n^2} = \lim_{n\to\infty} \frac{200}{3n} = 0$$

Hence, we can let C_1 be any positive real number. Let's pick $C_1 = 100$ just to make life easy. Now we want to solve for C_2. That is, we want to find $C_2 > 0$ such that

$$f(n) \leq C_1 g(n) \text{ for all } n > C_2$$

Substituting, we see that we want to find C_2 such that $100n^2 \leq 100(n^3 + 3) = 100n^3 + 300$ for all $n > C_2$. It is easy to see that $C_2 = 1$ works.

Note that we could have picked $C_1 = 1$, and then solved for C_2, but it would have been a more intricate piece of algebra if we did that, to find C_2.

Suppose we had tried to prove that g is $O(f)$. By the same analysis as done above,

$$\lim_{n\to\infty} \tfrac{g(n)}{f(n)} = \lim_{n\to\infty} \tfrac{g'(n)}{f'(n)} = \lim_{n\to\infty} \tfrac{3n^2}{200n} = \lim_{n\to\infty} \tfrac{3n}{200} = \infty$$

(Note that we applied L'Hôpital's Rule twice in this derivation.) Hence, the limit as $n \to \infty$ of $\frac{g(n)}{f(n)}$ is not bounded from above, and so g is not $O(f)$.

Example #4: Now let's look at a more complicated example. Let $g(n) = n$, and let $f(n)$ be defined as follows:

$$f(n) = \begin{cases} 5 & \text{if n is even} \\ n & \text{if n is odd} \end{cases}$$

Thus, $f(1) = 1, f(2) = 5, f(3) = 3, f(4) = 5, f(5) = 5, f(6) = 5, f(7) = 7, f(8) = 5$, etc. We want to know if f is $O(g)$. It is challenging to apply either technique to this pair of functions, since the value of $f(n)$ depends on the parity of n. For example, if we try to apply Technique 1, we find

$$\frac{f(n)}{g(n)} = \begin{cases} \frac{5}{n} & \text{if n is even} \\ 1 & \text{if n is odd} \end{cases}$$

In other words, $\lim_{n\to\infty} \frac{f(n)}{g(n)}$ *does not exist.*

When we look at Technique 2, we get

$$\log f(n) = \begin{cases} \log 5 & \text{if n is even} \\ 1 & \text{if n is odd} \end{cases}$$

Thus,

$$\lim_{n\to\infty} [\log f(n) - \log g(n)] = \begin{cases} \log 5 - 1 & \text{if n is even} \\ 0 & \text{if n is odd} \end{cases}$$

Hence, once again, the limit does not exist.

In other words, these two techniques only work when the limit exists! This does not mean that f is not $O(g)$, but only that a different technique must be used to determine whether f is $O(g)$. However, note that for all $n \geq 5$, $\frac{f(n)}{g(n)} \leq 1$. Hence, even though the limit does not exist, setting $C_1 = 1$ and $C_2 = 5$ allows use to prove that f is $O(g)$.

What about the converse? Is g big-oh of f?

$$\frac{g(n)}{f(n)} = \begin{cases} \frac{n}{5} & \text{n is even} \\ 1 & \text{n is odd} \end{cases}$$

Since $\lim_{n\to\infty} \frac{n}{5} = \infty$, it follows that g is not $O(f)$.

Proving that a function f is $O(g)$ Let f and g be functions from R^+ to R^+. To prove that f is $O(g)$, you need to do one of the following:

- Find constants C_1 and C_2 such that $f(n) \leq C_1 g(n)$ whenever $n > C_2$, and prove that the inequality holds. Proving that the inequality holds might in turn require calculus or induction, or at a minimum some algebra, so it's not enough to just write down the constants.
- Use one of the techniques given above (which provide implicit proofs that these constants exist).

Obviously it's easier to use the techniques than to find the constants and prove the inequality holds. Therefore, it's a good idea to learn how to use the techniques.

The skills you need to use these techniques are mostly from pre-calculus and calculus, and you are probably rusty. Please practice!

- You will need to be able to compute logarithms using any base.

- You will need to use L'Hôpital's Rule.
- You need to be able to compute limits.
- You will need to compute derivatives of potentially complicated functions (such as n^{3n} or $(1+n)^n$).

Practice questions. Try to answer each of the following questions, any of which could appear on the examlet. You could expect questions like these, even if they are not identical.

1. Provide the definition of the set of functions $f(n)$ that are $O(n^2)$.
2. Provide the definition of the set of functions that are $\Theta(n^2)$.
3. Provide the constants C_1 and C_2 proving that 3^n is $O(3^n - 2^n)$.
4. Solve for $H(n)$: $3^{n^2-1} = 4^{H(n)}$
5. Compute $\lim_{n\to\infty} \frac{\ln n}{n}$
6. Compute $\log_2(3f(n)^n)$
7. Compute $\log_3(5n^2 4^n)$
8. Determine (no proof requested), for each pair of functions below, whether (a) f is $O(g)$ but not vice-versa, (b) g is $O(f)$ but not vice-versa, (c) both are big-oh of each other, or (d) neither is big-oh of each other You should only concern yourself with values $n \geq 1$.
    - $f(n) = n^2$ and $g(n) = \log(n^n)$
    - $f(n) = (\log n)^n$ and $g(n) = \sqrt{n}$
    - $f(n) = \log(n^{500})$ and $g(n) = 100$
    - $f(n) = (\log n)^{500}$ and $g(n) = 100$
    - $f(n) = 100 + \frac{3}{n}$ and $g(n) = 5$
9. Sally says f is $O(g)$, but Bob notes that $f(n) > g(n)$ for all n, and so says f cannot be $O(g)$. What do you think of this argument? Assuming that $f(n) > g(n)$ is true, can Sally possibly be right? Or is Bob always right?