Math Facts

Victor Shoup

1 Introduction

In this class, we will often consider functions $f : \mathbb{Z}_{>0} \to \mathbb{R}$; that is, functions f that map each positive integer n to some real number $f(n)$.[1] For example, $f(n)$ might represent some upper bound on the number of steps that a particular algorithm executes on inputs of size n. We want to be able to answer questions about how $f(n)$ behaves as n gets really large. For example, does $f(n)$ itself tend towards infinity as n gets very large? And if so, how *fast* does $f(n)$ tend towards infinity? For example, does $f(n)$ grow *quadratically* (i.e., at a rate proportional to n^2), or much faster, such as *exponentially* (i.e., at a rate proportional to c^n for some constant $c > 1$)?

Mathematicians and computer scientists have developed some concepts and terminology that allow us to pose such questions in a fairly precise way. An example of that is so-called "big-O" notation, which you may have seen in a data structures class. We will discuss that notation in detail later on. However, we will begin with a review of a more basic concept, namely, that of the *limit* of $f(n)$ as n gets very large. This is a concept that should be familiar to you from calculus, but we will review it here.

2 Limits of functions

We start with some examples.

Example 2.1. Consider the function $f : \mathbb{Z}_{>0} \to \mathbb{R}$ that maps each positive integer n to the number $1/n$. That is, $f(n) = 1/n$ for each positive integer n. As n gets very large, the number $f(n)$ gets very close to 0 (even though it is never actually *equal* to 0). Nevertheless, we say that the *limit of $f(n)$ as n tends to infinity* is equal to 0. $\square$

Example 2.2. Consider the function $f : \mathbb{Z}_{>0} \to \mathbb{R}$ defined by $f(n) = 1 - 1/n$. As n gets very large, the number $f(n)$ gets very close to 1 (even though it is never *equal* to 1). In this case, we say that the *limit of $f(n)$ as n tends to infinity* is equal to 1. $\square$

Example 2.3. Consider the function $f : \mathbb{Z}_{>0} \to \mathbb{R}$ defined by $f(n) = n$. Unlike the previous two examples, the value $f(n)$ does not get close to any one number as n very large. Indeed, as n gets very large, the number $f(n)$ grows without bound. In this case, we say that the *limit of $f(n)$ as n tends to infinity* is equal to *infinity*. $\square$

Example 2.4. Consider the function $f : \mathbb{Z}_{>0} \to \mathbb{R}$ defined by $f(n) = (-1)^n$. That is, $f(n) = 1$ for even integers n, and $f(n) = -1$ for odd integers n. As n gets large, the value $f(n)$ just "bounces back and forth" between 1 and -1, but it does not remain very close any one number, nor does it grow without bound. In this case, we say that the *limit of $f(n)$ as n tends to infinity* does not

[1] $\mathbb{Z}$ is the set of integers, and $\mathbb{Z}_{>0}$ the set of positive integers. $\mathbb{R}$ is the set of real numbers, and $\mathbb{R}_{>0}$ is the set of positive real numbers.

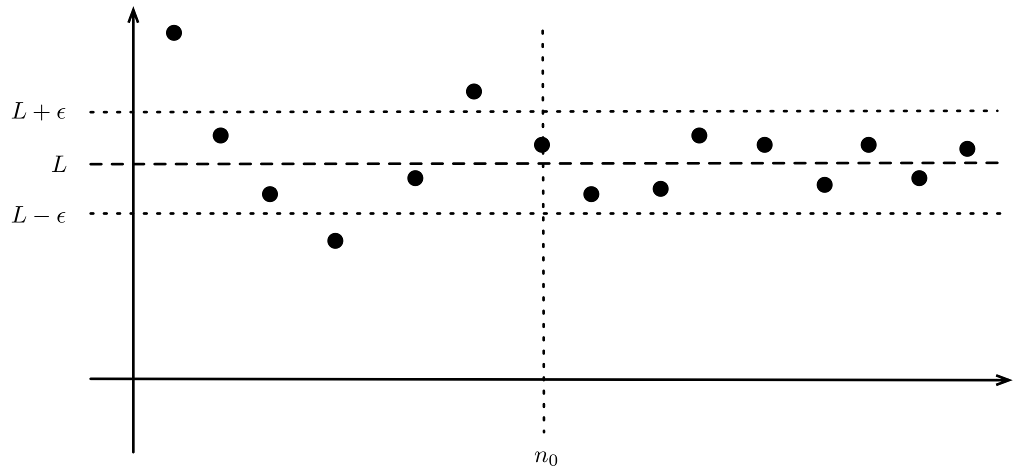


Figure 1: Visualizing a limit

exist. While it is easy to cook up functions, such as this one, for which this limit does not exist, they will not arise very much (if at all) in practice. □

More generally, consider an arbitrary function $f : \mathbb{Z}_{>0} \to \mathbb{R}$. That is, f maps each positive integer n to some real number $f(n)$. Let L be some real number. Intuitively, we say that the *limit of $f(n)$ as n tends to infinity* is equal to L if the following holds:

> as n gets very large, the value of $f(n)$ gets very close to L.

The formal definition is as follows:

Definition 2.1. *Let $f : \mathbb{Z}_{>0} \to \mathbb{R}$, and let $L \in \mathbb{R}$. We say that the* limit of $f(n)$ as n tends to infinity *is equal to L if the following holds:*

> *for every $\epsilon > 0$, there exists an integer n_0, such that $|f(n) - L| < \epsilon$ for all integers $n \geq n_0$.*

This definition says that $f(n)$ is very close to L, i.e., $L + \epsilon < f(n) < L + \epsilon$, as long as n is very large, i.e., $n \geq n_0$. This holds for *every* choice of ϵ, no matter how small. However, as ϵ gets smaller, the value of n_0 (which will, in general, depend on ϵ) may get larger.

See Fig. 1 to see how to visualize a limit. Here, we are graphing the points $(n, f(n))$ in the x-y plane, each represented as a dot in the figure. If the limit is L, then if we draw a very narrow band around the horizontal line $y = L$ (between the lines $y = L - \epsilon$ and $y = L + \epsilon$) then far enough out (to the right of the vertical line $x = n_0$) all of the points will lie within the band. If we draw an even narrower band, then we may have to go farther out.

Notation: we will write

$$\lim_{n\to\infty} f(n) = L$$

to mean that the *limit of $f(n)$ as n tends to infinity* is equal to L.

Hearkening back Examples 2.1 and 2.2, we see that

$$\lim_{n\to\infty} 1/n = 0$$

and

$$\lim_{n\to\infty}(1-1/n)=1.$$

Notice, however, that the above definition does not really capture the situation in Example 2.3. Intuitively, we say that the *limit of $f(n)$ as n tends to infinity* is equal to *infinity* if the following holds:

> as n gets very large, the absolute value of $f(n)$ grows without bound.

The formal definition is as follows:[2]

Definition 2.2. *Let $f:\mathbb{Z}_{>0}\to\mathbb{R}$. We say that the* limit of $f(n)$ as n tends to infinity *is equal to* infinity *if the following holds:*

> *for every $B>0$, there exists an integer n_0, such that $|f(n)|>B$ for all integers $n\geq n_0$.*

Notation: we will write

$$\lim_{n\to\infty}f(n)=\infty$$

to mean that the *limit of $f(n)$ as n tends to infinity* is equal to ∞.

So now, in Example 2.3, we can say

$$\lim_{n\to\infty}n=\infty.$$

Example 2.5. You can often just "eyeball" a simple function to determine if it tends to infinity or zero as n tends to infinity. Functions like n, n^2, $\sqrt{n}$, or even $\ln(n)$, which grow without bound, tend to infinity, while their reciprocals, $1/n$, $1/n^2$, $1/\sqrt{n}$, $1/\ln(n)$, tend to zero. $\Box$

2.1 Relation to other notions of limits

In calculus, when you first studied limits, you probably considered functions $g:\mathbb{R}\to\mathbb{R}$, that is, functions mapping real numbers to real numbers, and considered limits like

$$\lim_{x\to c}g(x),$$

where c is a real number. Indeed, this is the type of limit used to define the notion of the derivative of a function.

You probably also studied limits of the form

$$\lim_{x\to\infty}g(x).$$

The only difference between this type of limit and the one we are talking about here is that we are considering functions $f:\mathbb{Z}_{>0}\to\mathbb{R}$ whose domains are the positive integers, rather than the real numbers. These two notions of a limit are pretty much the same. In fact, in many situations, we may be working with a function $f:\mathbb{Z}_{>0}\to\mathbb{R}$ whose domain can be naturally extended to all positive real numbers. Indeed, this is the case for the functions f in Examples 2.1–2.3. In such cases, if the limit of $f(x)$ as x tends to infinity over the real numbers exists, then the limit of $f(n)$ as n tends to infinity over the integers also exists, and these two limits are equal.

[2]Note the absolute value in the definition. Because of that, our notion of infinity is an "unsigned infinity". This simplifies a number of things and will be sufficient for our purposes. Note that some texts will define notions of "positive" and "negative" infinity. In any event, we will usually be working with functions that output only nonnegative values (at least for very large inputs), and so this distinction will typically not matter.

Again, suppose that $f : \mathbb{Z}_{>0} \to \mathbb{R}$ is a function that maps each positive integer n to a value $f(n)$. Suppose we define $a_n := f(n)$. Then we can think of the function f as an *infinite sequence* of numbers

$$a_1, a_2, a_3, \ldots.$$

This is why you will often see calculus textbooks define the notion of the *limit of a sequence*, which is *exactly* the same as the notion of the *limit of $f(n)$ as n tends to infinity* that we have defined here.

Finally, note that there is nothing special about defining the domain of our functions f to be the integers starting from 1. We can just as well start from 0, or -1, or 17. Since we are interested in what happens to $f(n)$ as n gets very large, it really does not matter where we start.

2.2 Properties of limits

We state some standard properties of limits. See any calculus textbooks for proofs.[3]

The first property says that a function cannot have more than one limit.

Theorem 2.3. *Let $f : \mathbb{Z}_{>0} \to \mathbb{R}$. Let $L_1, L_2 \in \mathbb{R} \cup \{\infty\}$. If $\lim_{n\to\infty} f(n) = L_1$ and $\lim_{n\to\infty} f(n) = L_2$, then $L_1 = L_2$.*

The second property says that the limit of a constant function is the constant itself.

Theorem 2.4. *Let $f : \mathbb{Z}_{>0} \to \mathbb{R}$ and let $c \in \mathbb{R}$. If $f(n) = c$ for all $n \in \mathbb{Z}_{>0}$, then $\lim_{n\to\infty} f(n) = c$.*

The third property essentially says that the limit of the sum, difference, product, or quotient of two functions is just the sum, difference, product, or quotient of their limits (although there are some corner cases, like dividing by zero, to watch out for).

Theorem 2.5. *Let $f_1, f_2 : \mathbb{Z}_{>0} \to \mathbb{R}$. Let $L_1, L_2 \in \mathbb{R} \cup \{\infty\}$, and suppose*

$$\lim_{n\to\infty} f_1(n) = L_1 \quad \text{and} \quad \lim_{n\to\infty} f_2(n) = L_2.$$

Then we have:

(i) $\lim\limits_{n\to\infty} (f_1(n) + f_2(n)) = L_1 + L_2$, *unless L_1 and L_2 are both ∞.*

(ii) $\lim\limits_{n\to\infty} (f_1(n) - f_2(n)) = L_1 - L_2$, *unless L_1 and L_2 are both ∞.*

(iii) $\lim\limits_{n\to\infty} (f_1(n) \cdot f_2(n)) = L_1 \cdot L_2$, *unless one of L_1 and L_2 is 0 and the other is ∞.*

(iv) $\lim\limits_{n\to\infty} (f_1(n)/f_2(n)) = L_1/L_2$, *unless L_1 and L_2 are both 0 or both ∞.*

In using this theorem, arithmetic expressions on limits involving ∞ and a number $c \in \mathbb{R}$ is to be interpreted as follows:[4]

- $\infty \pm c = c \pm \infty = \infty$;
- $c \cdot \infty = \infty \cdot c$ for $c \neq 0$;
- $\infty \cdot \infty = \infty$;

[3] Depending on your calculus class, you may have only seen these properties stated (and proved) for functions with real domains, rather than integer domains, but the properties (and their proofs) are essentially the same.

[4] Note that the expression "$c/0 = \infty$" makes sense because we are working with an "unsigned" infinity.

- $c/\infty = 0$;
- $\infty/c = \infty$;
- $c/0 = \infty$ for $c \neq 0$.

Note that the following expression on limits have no good interpretation:

$$\infty \pm \infty, \quad 0 \cdot \infty, \quad \infty \cdot 0, \quad \infty/\infty, \quad \text{and} \quad 0/0.$$

If you read the above theorem carefully, you will see that none of these expressions arise.

This theorem is very powerful, as the following examples illustrate.

Example 2.6. Let

$$f(n) = \frac{3n^2 - 2n + 6}{n^2 + 4n - 7}.$$

The goal is to compute $\lim_{n\to\infty} f(n)$. Divide both the numerator and denominator by n^2, obtaining

$$f(n) = \frac{3 - 2/n + 6/n^2}{1 + 4/n - 7/n^2}.$$

We see that $\lim_{n\to\infty} 2/n = 0$ and $\lim_{n\to\infty} 6/n^2 = 0$. and so

$$\lim_{n\to\infty} (3 - 2/n + 6/n^2) = 3 - 0 + 0 = 3.$$

Similarly,

$$\lim_{n\to\infty} (1 + 4/n - 7/n^2) = 1 + 0 - 0 = 1.$$

It follows that

$$\lim_{n\to\infty} f(n) = \frac{\lim_{n\to\infty}(3 - 2/n + 6/n^2)}{\lim_{n\to\infty}(1 + 4/n - 7/n^2)} = \frac{3}{1} = 3.$$

Sometimes, one can express the above computation more succinctly as

$$f(n) = \frac{3 - 2/n + 6/n^2}{1 + 4/n - 7/n^2} \to \frac{3 - 0 + 0}{1 + 0 - 0} \to \frac{3}{1} \to 3. \quad \square$$

Example 2.7. Let

$$f(n) = \frac{2n^2 + 100}{n^3 + 5n + 6}.$$

The goal is to compute $\lim_{n\to\infty} f(n)$. Dividing numerator and denominator by n^3, we obtain

$$f(n) = \frac{2/n + 100/n^3}{1 + 5/n^2 + 6/n^3} \to \frac{0 + 0}{1 + 0 + 0} \to \frac{0}{1} \to 0. \quad \square$$

Example 2.8. Let

$$f(n) = \frac{n^2 + 100}{5n + 6}.$$

The goal is to compute $\lim_{n\to\infty} f(n)$. Dividing numerator and denominator by n^2, we obtain

$$f(n) = \frac{1 + 100/n^2}{5/n + 6/n^2} \to \frac{1 + 0}{0 + 0} \to \frac{1}{0} \to \infty. \quad \square$$

Example 2.9. Let $f(n) = n^2 - 100n$. We want to show that $\lim_{n\to\infty} f(n) = \infty$. Note that we cannot simply apply Theorem 2.5 with $f_1(n) := n^2$ and $f_2(n) := 100n$: both $f_1(n)$ and $f_2(n)$ tend to infinity, and $\infty - \infty$ is one of the cases not covered by this theorem.

Nevertheless, it is intuitively clear that $\lim_{n\to\infty} f(n) = \infty$. Indeed, we know that n^2 tends to infinity. Moreover, while we are subtracting off the term $100n$, which also tends to infinity, it does so "much more slowly" than n^2, so it should not make a big difference as n gets very large. For example, when $n = 10^9$, we have $n^2 = 10^{18}$, and we are subtracting off $100n = 10^{11}$, which is a tiny compared to n^2. Indeed, for $n = 10^9$, we have

$$f(n) = 10^{18} - 10^{11} = 10^{18}(1 - 10^{-7}) = 10^{18} \cdot 0.9999999,$$

and so $f(n)$ is very well approximated by n^2 for such large values of n.

To make this argument more rigorous, let us rewrite $f(n)$ as $n^2 \cdot (1 - 100/n)$. Then we have

$$f(n) = n^2 \cdot (1 - 100/n) \to \infty \cdot (1 - 0) \to \infty \cdot 1 \to \infty. \quad \square$$

We state a couple more property of limits.

Theorem 2.6. *Let $f : \mathbb{Z}_{>0} \to \mathbb{R}$ and suppose that $\lim_{n\to\infty} f(n) = L$ where $L \in \mathbb{R}$. Furthermore, suppose g is function mapping real numbers to real numbers such that $\lim_{x\to L} g(x) = M \in \mathbb{R}\cup\{\infty\}$. Then we have*

$$\lim_{n\to\infty} g(f(n)) = M.$$

In particular, if g is continuous at L, then

$$\lim_{n\to\infty} g(f(n)) = g(L).$$

Recall from calculus that g is continuous at L means that $\lim_{x\to L} g(x) = g(L)$. Intuitively, continuous functions are those that can be drawn on paper without lifting your pencil.

Example 2.10. Consider again the function $f(n)$ defined in Example 2.6. Define $h(n) = \sqrt{f(n)}$. That is,

$$h(n) = \sqrt{\frac{3n^2 - 2n + 6}{n^2 + 4n - 7}}.$$

We saw in Example 2.6 that $\lim_{n\to\infty} f(n) = 3$. Moreover, the function $g(x) = \sqrt{x}$ is certainly continuous at $x = 3$. Therefore,

$$\lim_{n\to\infty} h(n) = \lim_{n\to\infty} g(f(n)) = g(L) = \sqrt{3}. \quad \square$$

The next theorem is analogous to Theorem 2.6, but covers the case where $\lim_{n\to\infty} f(n) = \infty$.

Theorem 2.7. *Let $f : \mathbb{Z}_{>0} \to \mathbb{R}_{>0}$, where $\lim_{n\to\infty} f(n) = \infty$.*[5] *Furthermore, suppose $g : \mathbb{R}_{>0} \to \mathbb{R}$, where $\lim_{x\to\infty} g(x) = M \in \mathbb{R} \cup \{\infty\}$. Then we have*

$$\lim_{n\to\infty} g(f(n)) = M.$$

[5] Note that here, we have to insist that $f(n)$ itself, rather than $|f(n)|$ (as in Definition 2.2), grows without bound as n increases. This is why we assume that $f(n) > 0$ for all n.

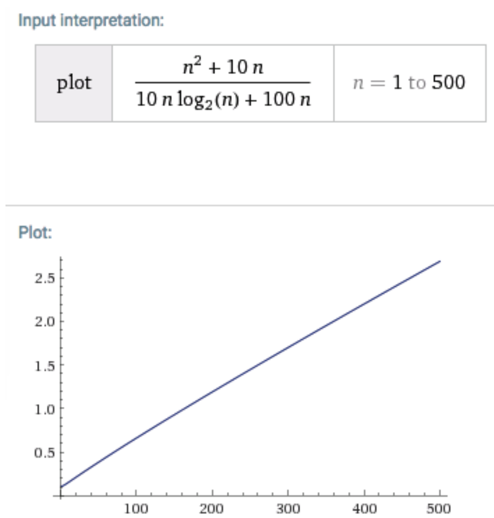


Figure 2: Graph of $f(n)/g(n)$

3 Asymptotic notation: big-O, big-Theta, and all that

Suppose that on inputs of size n, Algorithm 1 takes time

$$f(n) = n^2 + 10n$$

while Algorithm 2 takes time

$$g(n) = 10n \log_2(n) + 100n$$

Since

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = \infty,$$

eventually, as n gets larger, Algorithm 1 will get slower and slower than Algorithm 2. You can see this graphically in Fig. 2. Although Algorithm 1 is faster for smaller inputs (up to about $n = 200$), for larger inputs, Algorithm 1 is slower, and its performance relative to Algorithm 2 just gets worse as the input gets larger. We say that Algorithm 1 is *asymptotically slower than* Algorithm 2, and that Algorithm 2 is *asymptotically faster than* Algorithm 1.

In this class, we are mainly concerned with the *rate of growth* of the running time of an algorithm as a function of the *input size*. For this purpose, we really only need to worry about the "high order term" of this function. Moreover, we will typically ignore the leading constant as well. For example, for $f(n) = n^2 + 10n$, we just say $f(n) = O(n^2)$, and for $g(n) = 10n \log_2(n) + 100n$, we just say $g(n) = O(n \log_2(n))$. Recall that $\log_b(n) = \log_c(n)/\log_c(b)$, so if we are ignoring constant factors, then the base of the logarithm does not really matter (as long as it is a *constant*). So with $g(n)$ as above, we can also just say $g(n) = O(n \log(n))$, without even specifying the base of the logarithm.

There are a number of reasons for ignoring leading constants and lower order terms when analyzing the running times of algorithms:

- The lower order terms and constants depend on myriad implementation details, the quality of the compiler, the instruction set architecture of the machine, and the micro-architecture of the machine. By ignoring constants and lower order terms, we can prove results about algorithms that are meaningful across a wide range of implementations.

- When comparing algorithms, even though we are ignoring constants and lower order terms, if Algorithm 2 is *asymptotically faster* than Algorithm 1 (as in the example above), then for *any* implementation, Algorithm 2 will be faster *in practice* than Algorithm 1 for large enough inputs.

  Moreover, as computers get faster and faster (as they tend to do), such asymptotic results only gain in importance: as machines get faster, the input sizes that can be processed in a practical amount of time get ever larger.

3.1 Big-O: formal definition

We now formally define big-O notation.

Definition 3.1 (big-O). *Let f and g be functions from $\mathbb{Z}_{>0}$ to $\mathbb{R}$. We say $f = O(g)$ if*

$$|f(n)| \leq c|g(n)|$$

for some constant c and all sufficiently large n. More precisely: there exists a some positive real number c and a positive integer n_0 such that

$$|f(n)| \leq c|g(n)|$$

for all $n \geq n_0$.

Intuitively, $f = O(g)$ means f grows *no faster* than g. The relation "$f = O(g)$" is analogous to "$f \leq g$". In particular, it is reflexive (i.e., $f = O(f)$) and transitive (i.e., $f = O(g)$ and $g = O(h)$ implies $f = O(h)$).

Example 3.1. Let $g(n) = 10n\log_2(n) + 100n$. We will show that $g(n) = O(n\log_2(n))$. Instead of working directly from the definition of big-O, it is easier to consider the limit

$$\lim_{n\to\infty} \frac{g(n)}{n\log_2(n)}.$$

We have:

$$\begin{aligned}
\lim_{n\to\infty} \frac{g(n)}{n\log_2(n)} &= \lim_{n\to\infty} \frac{10n\log_2(n) + 100n}{n\log_2(n)} \\
&= \lim_{n\to\infty} \frac{10 + 100/\log_2(n)}{1} \qquad \text{(dividing numerator and denominator by } n\log_2(n)\text{)} \\
&= \frac{10+0}{1} = 10.
\end{aligned}$$

By the definition of a limit: for all $\epsilon > 0$, we have

$$10 - \epsilon < \frac{g(n)}{n\log_2(n)} < 10 + \epsilon$$

for all sufficiently large n. In particular, setting $\epsilon := 0.1$:

$$g(n) < 10.1 \cdot n\log_2(n)$$

for all sufficiently large n. $\square$

Why only for “sufficiently large n”?

Note that in the definition of $f = O(g)$, we only require that $|f(n)| \leq c|g(n)|$ for all sufficiently large n, that is, for all $n \geq n_0$ for some n_0. The reader may wonder why we cannot simply insist that this inequality holds for *all* positive integers n? Indeed, to satisfy the definition, we need to show that there exist a suitable c and n_0. Perhaps we can just set $n_0 = 1$ and then choose c to be large enough so that $|f(n)| \leq c|g(n)|$ for all $n \geq n_0 = 1$.

One reason why this does not always work is that the functions $f(n)$ and $g(n)$ may not even be well defined for all positive integers n. For example, if $f(n) = n/\log_2(n)$, then $f(n)$ is not well defined for $n = 1$. Even though Definition 3.1 is technically for functions whose domains consist of all positive integers, we will sometimes adapt it to functions whose domains contain all integers $n \geq n_1$ for some n_1.

Even if $f(n)$ and $g(n)$ are well defined for all positive integers n, if for some values of n, we have $f(n) \neq 0$ but $g(n) = 0$, then no matter how large we choose the value c, we will not have $|f(n)| \leq c|g(n)|$ for all n. For example, suppose $f(n) = 10n$ and $g(n) = n\log_2(n)$. Then for $n = 1$, we cannot have $|f(n)| \leq c|g(n)|$ for any positive number c.

In general, what we can say is the following. Suppose that $f = O(g)$. Moreover, suppose that $f(n)$ and $g(n)$ are defined for all $n \geq n_1$ and $g(n) \neq 0$ for all $n \geq n_1$. Then there exists a positive number d such that $|f(n)| \leq d|g(n)|$ for all $n \geq n_1$.

To see why this is so, observe that since $f = O(g)$, we know that there exists c and n_0 such that $|f(n)| \leq c|g(n)|$ for all $n \geq n_0$. Now, if $n_1 \geq n_0$, then we can just set $d := c$, and this will ensure that $|f(n)| \leq d|g(n)|$ for all $n \geq n_1$. Otherwise, if $n_1 < n_0$, we can set

$$d := \max\Big(\ |f(n_1)/g(n_1)|, |f(n_1+1)/g(n_1+1)|, \ldots, |f(n_0-1)/g(n_0-1)|,\ c\ \Big).$$

and this will ensure that $|f(n)| \leq d|g(n)|$ for all $n \geq n_1$.

Implicit big-O notation

If $f(n) = 2n^2 + 10n + 1$, we may write

$$f(n) = 2n^2 + O(n)$$

This means that

$$f(n) = 2n^2 + h(n)$$

for some “anonymous” function $h = O(n)$. In this example, $h(n) = 10n + 1$. This is useful in situations where we do not want to completely ignore the constant in the high-order term.

As another example, suppose $f(n) = 3n^2 - 20n + 5$. We may write

$$f(n) = 3n^2 + O(n).$$

This means that

$$f(n) = 3n^2 + h(n)$$

for some “anonymous” function $h = O(n)$. In this example, $h(n) = -20n + 5$. Note that in the definition of big-O, we take absolute values, so it makes sense to say that a function that may be negative is $O(n)$.

As another example, suppose we know that

$$\ln(n) - 1 \leq f(n) \leq \ln(n) + 2$$

for all n. Then we can write

$$f(n) = \ln(n) + O(1).$$

In general, "$O(1)$" denotes some "anonymous" function that is bounded by a constant in absolute value.

3.2 Big-Omega and big-Theta notation

We write $f = \Omega(g)$ to mean $g = O(f)$.

Intuitively, $f = \Omega(g)$ means f grows *at least as fast* as g. The relation "$f = \Omega(g)$" is analogous to "$f \geq g$". In particular, it is reflexive (i.e., $f = \Omega(f)$) and transitive (i.e., $f = \Omega(g)$ and $g = \Omega(h)$ implies $f = \Omega(h)$).

We write $f = \Theta(g)$ to mean $f = O(g)$ and $g = O(f)$. This is equivalent to saying that

$$d|g(n)| \leq |f(n)| \leq c|g(n)|$$

for some constants c, d and all sufficiently large n.

Intuitively, $f = \Theta(g)$ means f and g grow *at the same rate*. The relation "$f = \Theta(g)$" is analogous to "$f = g$". In particular, it is reflexive (i.e., $f = \Theta(f)$), symmetric (i.e., $f = \Theta(g)$ if and only if $g = \Theta(f)$), and transitive (i.e., $f = \Theta(g)$ and $g = \Theta(h)$ implies $f = \Theta(h)$). That is, the relation "$f = \Theta(g)$" is an *equivalence relation.*

Example 3.2. Let $f(n) = n^2 + 10n$ and $g(n) = 2n^2 + n$. We will show that $f = \Theta(g)$. As in Example 3.1, we do this by taking the limit of the ratio $f(n)/g(n)$ as n tends to infinity.

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = \lim_{n\to\infty} \frac{n^2 + 10n}{2n^2 + n} = \lim_{n\to\infty} \frac{1 + 10/n}{2 + 1/n} = \frac{1+0}{2+0} = \frac{1}{2}$$

It follows from the definition of a limit that

$$0.25 \cdot g(n) \leq f(n) \leq 0.75 \cdot g(n)$$

for all sufficiently large n. Indeed, just plug in $L = 0.5$ and $\epsilon = 0.25$ into Definition 2.1, and we see that for all sufficiently large n,

$$-0.25 \leq f(n)/g(n) - 0.5 \leq 0.25.$$

which implies

$$0.25 \leq f(n)/g(n) \leq 0.75. \quad \square$$

3.3 Little-o notation

Definition 3.2 (little-o). *Let f and g be functions from $\mathbb{Z}_{>0}$ to $\mathbb{R}$. We say $f = o(g)$ if*

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0.$$

Equivalently, $f = o(g)$ if for every positive number ϵ, we have

$$|f(n)| < \epsilon|g(n)| \tag{1}$$

for all sufficiently large n. This just follows from the definition of a limit.

Intuitively, $f = o(g)$ means f grows *strictly more slowly* than g. The relation "$f = o(g)$" is analogous to "$f < g$". In particular, it is transitive (i.e., $f = o(g)$ and $g = o(h)$ implies $f = o(h)$).

Example 3.3. Let $f(n) = 2n^2 + 10n$ and $g(n) = n^3 + n$. We have

$$\begin{aligned} \lim_{n\to\infty} \frac{f(n)}{g(n)} &= \lim_{n\to\infty} \frac{2n^2+10n}{n^3+n} = \lim_{n\to\infty} \frac{2/n+10/n^2}{1+1/n^2} \\ &= \frac{0+0}{1+0} = 0 \end{aligned}$$

Therefore, $f = o(g)$. $\square$

There is a simple connection between little-o and big-O that is analogous to the connection between "less than" and "less than or equal to":

Theorem 3.3. *Suppose $f = o(g)$. Then we have:*

(i) $f = O(g)$, *and*

(ii) $g \neq O(f)$.

Another connection between little-o and big-O that is analogous to the connection between "less than" and "less than or equal to" is the following transitivity property:

Theorem 3.4. *Let f, g, and h be functions from $\mathbb{Z}_{>0}$ to $\mathbb{R}$. Then we have:*

(i) if $f = O(g)$ and $g = o(h)$, then $f = o(h)$;

(ii) if $f = o(g)$ and $g = O(h)$, then $f = o(h)$.

We leave the proofs of these two theorems to the reader: they follows directly from the definitions.

Implicit little-o notation

If $f(n) = 2n^2 + 10n$, we may write $f(n) = 2n^2 + o(n^2)$. This means that $f(n) = 2n^2 + h(n)$ for some "anonymous" function $h = o(n^2)$. In this example, $h(n) = 10n$.

Again, if $f(n) = 2n^2 + 10n$, we may also write $f(n) = 2n^2(1 + o(1))$. Again, this means $f(n) = 2n^2(1 + h(n))$ for some anonymous function $h(n) = o(1)$. In this example, $h(n) = (10n)/(2n^2) = 5/n$. Saying $h(n) = o(1)$ just means that $\lim_{n\to\infty} h(n) = 0$. In general, "$o(1)$" denotes some "anonymous" function that tends to zero as n tends to infinity.

As another example, if $f(n) = n^2(\ln(n))^3$, we may write $f(n) = n^{2+o(1)}$. Again, this means $f(n) = n^{2+h(n)}$ for some anonymous function $h(n) = o(1)$. In this example, $h(n)$ is implicitly defined by $n^{h(n)} = (\ln(n))^3$, and taking logarithms, this implies $h(n)\ln(n) = 3\ln(\ln(n))$. In other words,

$$h(n) = \frac{3\ln(\ln(n))}{\ln(n)}.$$

We shall see below (in Example 3.7) that $\lim_{n\to\infty} h(n) = 0$.

We have discussed implicit big-O and little-o notation, as these are often used in practice; however, one can also use implicit big-Omega and big-Theta notation.

More little-o examples

Here, we present a few examples of the little-o relation that arise in many applications.

Example 3.4. Let $f(n) = n^\alpha$ and $g(n) = n^\beta$, where $\alpha < \beta$ are constants. The we have

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = \lim_{n\to\infty} \frac{n^\alpha}{n^\beta} = \lim_{n\to\infty} n^{\alpha-\beta} = 0.$$

Therefore, $f = o(g)$. $\square$

Example 3.5. Let $f(n) = \ln(n)$ and $g(n) = n^\beta$, where $\beta > 0$ is a constant. We want to compute $\lim_{n\to\infty} f(n)/g(n)$. In order to compute this limit, we shall employ L'hospital's rule. However, L'hospital's rule only applies to functions defined over the reals. But as discussed in §2.1, the limit of $f(n)/g(n)$ as n tends to infinity over the integers is equal to the limit of $f(x)/g(x)$ as x tends to infinity over the real numbers (provided this latter limit exists). So we have

$$\begin{aligned}\lim_{n\to\infty} \frac{f(n)}{g(n)} &= \lim_{x\to\infty} \frac{f(x)}{g(x)} = \lim_{x\to\infty} \frac{\ln(x)}{x^\beta} = \lim_{x\to\infty} \frac{x^{-1}}{\beta x^{\beta-1}} \quad \text{(L'hospital's rule)} \\ &= \lim_{x\to\infty} \frac{1}{\beta x^\beta} = 0\end{aligned}$$

Therefore, $f = o(g)$. $\square$

The above example says that the logarithm of n grows more slowly than any positive power (no matter how small) of n. We can generalize this, as follows:

Example 3.6. Let $f(n) = (\ln(n))^\alpha$ and $g(n) = n^\beta$, where α, β are positive constants. Then we have

$$\begin{aligned}\lim_{n\to\infty} \frac{f(n)}{g(n)} &= \lim_{n\to\infty} \frac{(\ln(n))^\alpha}{n^\beta} \\ &= \lim_{n\to\infty} \left(\frac{\ln(n)}{n^{\beta/\alpha}}\right)^\alpha = \left(\lim_{n\to\infty} \frac{\ln(n)}{n^{\beta/\alpha}}\right)^\alpha = 0^\alpha = 0.\end{aligned}$$

Here, we have combined Theorem 2.6 (applied with $f(n) := \ln(n)/n^{\beta/\alpha}$ and $g(x) := x^\alpha$) with the result of the previous example (which implies that $\lim_{n\to\infty} \ln(n)/n^{\beta/\alpha} = 0$).

Therefore, $f = o(g)$. $\square$

This last example says that any positive power (no matter how large) of the logarithm of n grows more slowly than any positive power (no matter how small) of n. This holds for $\log_2(n)$ in place of n, and more generally, with $\log_c(n)$ for any constant $c > 1$.

So, for example, we have $(\log_{10}(n))^2 = o(n^{0.5})$. Indeed, for $n > 50000$, the ratio $(\log_{10}(n))^2/n^{0.5}$ is less than $1/10$. As another example, we have $(\log_{10}(n))^{100} = o(n^{0.01})$. However, in this example, n has to be *very* large for the ratio $(\log_{10}(n))^{100}/n^{0.01}$ to become small. Indeed, this ratio is larger than 1 for all n below 10^{46000}.

Example 3.7. Let $f(n) = \ln(\ln(n))$ and $g(n) = (\ln(n))^\beta$, where $\beta > 0$ is a constant. Then we have

$$\begin{aligned}\lim_{n\to\infty} \frac{f(n)}{g(n)} &= \lim_{n\to\infty} \frac{\ln(\ln(n))}{(\ln(n))^\beta} \\ &= \lim_{n\to\infty} \frac{\ln(n)}{n^\beta}. \\ &= 0.\end{aligned}$$

Here, we have combined Theorem 2.7 (applied with $f(n) := \ln(n)$ and $g(x) := \ln(x)/x^{\beta}$) with Example 3.5 (where we showed that $\lim_{x\to\infty} \ln(x)/x^{\beta} = 0$).

Therefore, $f = o(g)$. $\square$

Example 3.8. We can apply the same argument that we made in Example 3.6 to the result from the previous example, to obtain

$$\lim_{n\to\infty} \frac{(\ln(\ln(n)))^{\alpha}}{(\ln(n))^{\beta}} = 0$$

for any positive constants α, β. Therefore, $(\ln(\ln(n)))^{\alpha} = o\big((\ln(n))^{\beta}\big)$.

More generally, we have:

$$\lim_{n\to\infty} \frac{(\ln(f(n)))^{\alpha}}{(f(n))^{\beta}} = 0$$

for any function $f : \mathbb{Z}_{>0} \to \mathbb{R}_{>0}$ such that $\lim_{n\to\infty} f(n) = \infty$. So we have $(\ln(f(n)))^{\alpha} = o\big((f(n))^{\beta}\big)$. $\square$

3.4 General tools for comparing functions

Suppose we have two functions $f, g : \mathbb{Z}_{>0} \to \mathbb{R}$. Typically (but not always), we have one of three possibilities:

(i) f grows strictly more slowly than g, i.e., $f = o(g)$,

(ii) g grows strictly more slowly than f, i.e., $g = o(f)$, or

(iii) f and g grow at the same rate, i.e., $f = \Theta(g)$.

Typically (but not always), we can determine which of these three possibilities hold simply by looking the limit of the ratio $|f/g|$:

Theorem 3.5 (Limit Comparison Theorem for Rates of Growth). *Let $f, g : \mathbb{Z}_{>0} \to \mathbb{R}$. Suppose that*

$$\lim_{n\to\infty} \frac{|f(n)|}{|g(n)|} = L \in \mathbb{R} \cup \{\infty\}.$$

Then we have:

(i) if $L = 0$, then $f = o(g)$;

(ii) if $L = \infty$, then $g = o(f)$;

(iii) if $0 < L < \infty$, then $f = \Theta(g)$.

A more concrete tool

In this class, we will most often run into functions that are a sum of terms, where each term of the form $cn^{\alpha}(\ln(n))^{\beta}$, for constants c, α, β, and $c \neq 0$. For such a function, we can sort the terms: $cn^{\alpha}(\ln(n))^{\beta}$ is a *higher order term* than $c'n^{\alpha'}(\ln(n))^{\beta'}$ if

- $\alpha > \alpha'$, or
- $\alpha = \alpha'$ and $\beta > \beta'$.

For two such functions f and g, we can compare their highest terms:

(i) If g's highest term is higher than f's: $f = o(g)$

(ii) If f's highest term is higher than g's: $g = o(f)$

(iii) Otherwise, $f = \Theta(g)$.

Example 3.9. Let $f(n) = n^2 \log_2(n) + 10n^2$ and $g(n) = n^3 - n$.

We can use the Limit Comparison Theorem:

$$\frac{f(n)}{g(n)} = \frac{n^2 \log_2(n) + 10n^2}{n^3 - n} = \frac{\log_2(n)/n + 10/n}{1 - 1/n^2} \to \frac{0+0}{1-0} \to 0.$$

Therefore, case (i) of the theorem applies, and $f = o(g)$.

We can also simply use the above "concrete tool". The highest order term of $f(n)$ is $n^2 \log_2(n)$, i.e., $cn^{\alpha}(\ln(n))^{\beta}$, where $\alpha = 2$ and $\beta = 1$. The highest order term of $g(n)$ is n^3, i.e., $n^{\alpha'}(\ln(n))^{\beta'}$, where $\alpha' = 3$ and $\beta' = 0$. Since g's highest term is higher than f's highest term, case (i) of the concrete tool applies, and $f = o(g)$. □

3.5 Asymptotic equality

Definition 3.6 (asymptotic equality). *Let f and g be functions from $\mathbb{Z}_{>0}$ to $\mathbb{R}$. We say $f \sim g$ if*

$$\lim_{n \to \infty} \frac{f(n)}{g(n)} = 1.$$

When $f \sim g$, we also say that f and g are *asymptotically equal*. Note that $f \sim g$ is equivalent to saying that $f = g \cdot (1 + o(1))$.

Clearly, if $f \sim g$, then $f = \Theta(g)$, but $f \sim g$ says much more: not only do f and g grow at the same rate, but they are very good approximations of each other (as n gets large). Indeed, suppose we think of $f(n)$ as an approximation to $g(n)$. Recall that the *relative error* $\epsilon(n)$ is defined as

$$\epsilon(n) := \left| \frac{f(n) - g(n)}{g(n)} \right| = \left| \frac{f(n)}{g(n)} - 1 \right|.$$

So $f(n)/g(n) \to 1$ implies that the relative error $\epsilon(n)$ tends to 0 as n gets large.

The relation $f \sim g$ is reflexive (i.e., $f \sim f$), symmetric (i.e., $f \sim g$ if and only if $g \sim f$), and transitive (i.e., $f \sim g$ and $g \sim h$ implies $f \sim h$). That is, the relation $f \sim g$ is an *equivalence relation*.

Example 3.10. Let $f(n) = 2n^2 + 3n$ and $g(n) = 2n^2 - 4n + 1$. We leave it to the reader to verify that

$$\lim_{n \to \infty} \frac{f(n)}{g(n)} = 1.$$

Therefore, $f \sim g$. □

Example 3.11. Let $f(n) = 2n^2 + 3n$ and $g(n) = n^2 - 4n + 1$. We leave it to the reader to verify that

$$\lim_{n \to \infty} \frac{f(n)}{g(n)} = 2.$$

Therefore, $f = \Theta(g)$ (i.e., f and g grow at the same rate), but $f \nsim g$ (i.e., f and g are *not* asymptotically equal). □

We close this section on asymptotic equality with a useful theorem which also serves as a nice example:

Theorem 3.7. *Suppose* $f, g : \mathbb{Z}_{>0} \to \mathbb{R}_{>0}$. *Further suppose that* $f = \Theta(g)$ *and that* $\lim_{n\to\infty} g(n) = \infty$. *Then we have:*

$$\ln(f(n)) \sim \ln(g(n)).$$

Proof. Observe that $f = \Theta(g)$ implies

$$c < \frac{f(n)}{g(n)} < d$$

for some positive constants c and d and all sufficiently large n. Taking logarithms, we have

$$\ln(c) < \ln(f(n)) - \ln(g(n)) < \ln(d).$$

In other words,

$$\ln(f(n)) = \ln(g(n)) + O(1).$$

Dividing by $\ln(g(n))$, we have

$$\frac{\ln(f(n))}{\ln(g(n))} = 1 + O(1/\ln(g(n))).$$

By the assumption that $\lim_{n\to\infty} g(n) = \infty$, we have $\lim_{n\to\infty} 1/\ln(g(n)) = 0$. It follows that

$$\lim_{n\to\infty} \frac{\ln(f(n))}{\ln(g(n))} = 1.$$

Therefore, $\ln(f(n)) \sim \ln(g(n))$. $\square$

4 Summation

In analyzing algorithms, we will often need to add things up.

For example, suppose we execute the following algorithm:

```
for i in [1..n] do
    for j in [1..i] do
        HERE
```

The notation means that in the outer loop, i ranges over $1, \ldots, n$, while the inner loop, j ranges over $1, \ldots, i$. The question is, how many times does the line "HERE" get executed?

Clearly, for each iteration of the outer for loop, it gets executed i times. So if S is the total number of times is gets executed, then

$$S = 1 + 2 + \cdots + n = \sum_{i=1}^{n} i.$$

The reader will hopefully recall the formula

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}.$$

So we have

$$S = \sum_{i=1}^{n} i = \frac{n(n+1)}{2} = \frac{n^2}{2} + O(n) = \Theta(n^2)$$

What about more complicated sums, like $\sum_{i=1}^{n} i^2$ or $\sum_{i=1}^{i} 1/i$? While there are well known formulas and approximations for these, we next present a general tool that you can use to easily derive them from scratch.

4.1 Approximating sums by integrals

Suppose we want to estimate the sum

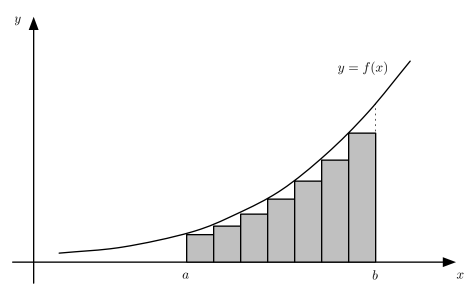
$$f(a) + f(a+1) + \cdots + f(b-1) + f(b) = \sum_{i=a}^{b} f(i),$$

where a and b are integers, with $a \leq b$.

Suppose that the function $f(x)$ is continuous and *non-decreasing* on $[a, b]$. Observe that the sum

$$\sum_{i=a}^{b-1} f(i).$$

is equal to the area of the shaded area in this figure:



Here, each rectangle has width 1. Also, recall that the integral

$$\int_a^b f(x)dx$$

is the area under the curve between a and b. Therefore, we have

$$\sum_{i=a}^{b-1} f(i) \leq \int_a^b f(x)dx,$$

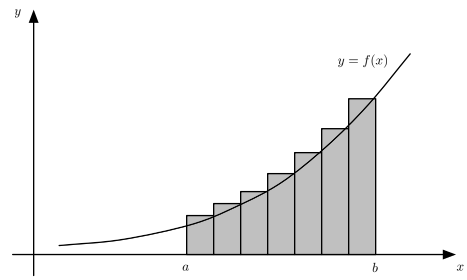
or

$$\sum_{i=a}^{b} f(i) \leq \int_a^b f(x)dx + f(b).$$

Now suppose that the function $f(x)$ is continuous and *non-increasing* on $[a, b]$. The sum

$$\sum_{i=a}^{b-1} f(i).$$

is equal to the area of the shaded area in this figure:



Therefore, we have

$$\int_a^b f(x)dx \leq \sum_{i=a+1}^{b} f(i),$$

or

$$\int_a^b f(x)dx + f(a) \leq \sum_{i=a}^{b} f(i),$$

We can put these observations together as a general theorem:

Theorem 4.1. *Suppose f is continuous and* monotone *(non-decreasing or non-increasing) on $[a, b]$. Then*

$$\int_a^b f(x)dx + m \;\leq\; \sum_{i=a}^{b} f(i) \;\leq\; \int_a^b f(x)dx + M,$$

where $m := \min(f(a), f(b))$, and $M := \max(f(a), f(b))$

Example 4.1 (Sum of squares). Estimate

$$Q_n := 1^2 + 2^2 + \cdots + n^2 = \sum_{i=1}^{n} i^2.$$

We have

$$\int_1^n x^2 dx = \left[\frac{x^3}{3}\right]_1^n = \frac{n^3}{3} - \frac{1}{3}$$

Therefore

$$\frac{n^3}{3} + \frac{2}{3} \leq Q_n \leq \frac{n^3}{3} + n^2 - \frac{1}{3}.$$

In other words,

$$\frac{2}{3} \leq Q_n - \frac{n^3}{3} \leq n^2 - \frac{1}{3}$$

In particular,

$$Q_n = n^3/3 + O(n^2). \quad \square$$

Example 4.2 (Harmonic numbers). Estimate

$$H_n := \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n} = \sum_{i=1}^{n} \frac{1}{i}.$$

We have

$$\int_1^n \frac{1}{x}dx = \Big[\ln(x)\Big]_1^n = \ln(n) - \ln(1) = \ln(n).$$

Therefore,

$$\ln(n) + \frac{1}{n} \le H_n \le \ln(n) + 1.$$

In other words,

$$\frac{1}{n} \le H_n - \ln(n) \le 1.$$

In particular,

$$H_n = \ln(n) + O(1).$$

The value H_n is called the nth *harmonic number*. $\square$

4.2 Geometric series

An important type of summation that arises in many situations is called a *geometric series*. Let r be a real number and let n be a positive integer. We define the geometric series

$$S_n(r) := \sum_{i=0}^{n} r^i = 1 + r + r^2 + \cdots + r^n.$$

The following fact is well known: for $r \neq 1$, we have

$$S_n(r) = \frac{1 - r^{n+1}}{1 - r} = \frac{r^{n+1} - 1}{r - 1}.$$

One can easily derive this as follows:

$$(1 - r)S_n(r) = (1 + r + \cdots + r^n) - (r + r^2 + \cdots r^{n+1}) = 1 - r^{n+1}$$

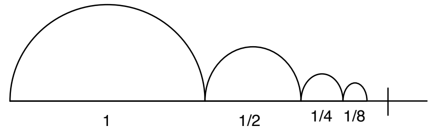
Example 4.3. We have

$$S_n(2) = 1 + 2 + 4 + \cdots + 2^n = 2^{n+1} - 1. \quad \square$$

Example 4.4. We have

$$S_n(1/2) = 1 + 1/2 + 1/4 + \cdots + 1/2^n = 2 - 1/2^n.$$

This is illustrated for $n = 3$ in the following diagram:



One sees that as n gets larger, the sum $S_n(1/2)$ gets closer and closer to 2, but it never quite gets there. $\square$

4.3 Infinite series

Consider again Example 4.4, where we saw that

$$S_n(1/2) = 1 + 1/2 + 1/4 + \cdots + 1/2^n = 2 - 1/2^n.$$

We can consider the corresponding *infinite series*

$$1 + 1/2 + 1/4 + \cdots = \sum_{i=0}^{\infty} 1/2^i.$$

This series has an infinite number of terms in it. However, we can easily make sense of it in terms of the limit of a sequence of finite sums. That is, the value of this infinite series is simply *defined* to be

$$\lim_{n\to\infty} S_n(1/2) = \lim_{n\to\infty} (2 - 1/2^n) = 2.$$

More generally, for $|r| < 1$:

$$\sum_{i=0}^{\infty} r^i = \lim_{n\to\infty} \sum_{i=0}^{n} r^i = \lim_{n\to\infty} \frac{1 - r^{n+1}}{1 - r} = \frac{1}{1 - r}.$$

Even more generally, consider an infinite sequence of numbers

$$a_1, a_2, a_3, \ldots.$$

Recall that such an infinite sequence is really just another way of talking about the function $f : \mathbb{Z}_{>0} \to \mathbb{R}$, where $f(i) = a_i$ for $i = 1, 2, 3, \ldots$.[6] We define the value of the infinite series

$$\sum_{i=1}^{\infty} a_i$$

to be the limit of the sequence of partial sums

$$S_1, S_2, S_3 \ldots$$

where

$$S_n := \sum_{i=1}^{n} a_i = a_1 + \cdots + a_n.$$

That is,

$$\sum_{i=1}^{\infty} a_i = \lim_{n\to\infty} \sum_{i=1}^{n} a_i.$$

This limit may be finite, infinite, or it may not exist at all.

For the remainder of this section, we shall assume that all of the a_i's are nonnegative after some point, i.e., for all sufficiently large i. This simplifies a number of things. For example, it is a general fact that in this case, the limit of the sequence of partial sums is equal to some $L \in \mathbb{R}$, or is equal to ∞ (in particular, the partial sums cannot just "bounce around", as in Example 2.4). If it is equal to $L \in \mathbb{R}$, we say the series $\sum_{i=1}^{\infty} a_i$ *converges to* L, and we write $L = \sum_{i=1}^{\infty} a_i < \infty$; otherwise, we say it *diverges (to infinity)*, and we write $\sum_{i=1}^{\infty} a_i = \infty$.

If $\sum_{i=1}^{\infty} a_i < \infty$, it seems intuitive that the a_i's must get small as i tends to infinity, i.e., that $\lim_{i\to\infty} a_i = 0$. We state this as a theorem:

[6] Also, there is nothing special about having the indices start at 1 — we could just have easily started at $i = 0$ or any other number for that matter.

Theorem 4.2. *Consider the series $\sum_{i=1}^{\infty} a_i$, where $a_i \geq 0$ for all sufficiently large i. If $\sum_{i=1}^{\infty} a_i < \infty$, then $\lim_{i\to\infty} a_i = 0$.*

Interestingly, the converse of this theorem is not true. That is, even if $\lim_{i\to\infty} a_i = 0$, we may still have $\sum_{i=1}^{\infty} a_i = \infty$.

Example 4.5. Using the result from Example 4.2, we have

$$\sum_{i=1}^{\infty} \frac{1}{i} = \lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{i} \geq \lim_{n\to\infty} \ln(n) = \infty.$$

Thus, we have $\sum_{i=1}^{\infty} 1/i = \infty$, even though $\lim_{i\to\infty} 1/i = 0$. $\square$

So to get a convergent series, not only do the terms have to go to zero, they have to go to zero *fast enough.*

There are various tests that can be used to determine when an infinite series converges or diverges. We state a couple of them.

Theorem 4.3 (Ratio Test). *Consider the series $\sum_{i=1}^{\infty} a_i$, where $a_i \geq 0$ for all sufficiently large i. Let $L := \lim_{i\to\infty} a_{i+1}/a_i$. Then we have:*

(i) if $L < 1$, then $\sum_{i=1}^{\infty} a_i < \infty$;

(ii) if $L > 1$, then $\sum_{i=1}^{\infty} a_i = \infty$;

(iii) If $L = 1$ (or the limit does not exist), the test is inconclusive *(i.e., the series may either converge or diverge).*

The intuition behind the ratio test is that to a large degree, the series $\sum_{i=1}^{\infty} a_i$ behaves much like a geometric series with $r = L$.

Example 4.6. Consider the series

$$\sum_{i=1}^{\infty} \frac{i}{2^i}.$$

So $a_i = i/2^i$. Applying the ratio test, we have

$$\frac{a_{i+1}}{a_i} = \frac{i+1}{2^{i+1}} \cdot \frac{2^i}{i} = \frac{i+1}{2i} = \frac{1+1/i}{2} \to \frac{1+0}{2} \to 1/2.$$

We conclude that the series converges. $\square$

Example 4.7. Consider the series

$$\sum_{i=1}^{\infty} \frac{1}{i}.$$

So $a_i = 1/i$. Applying the ratio test, we have

$$\frac{a_{i+1}}{a_i} = \frac{1}{i+1} \cdot \frac{i}{1} = \frac{i}{i+1} = \frac{1}{1+1/i} \to \frac{1}{1+0} \to 1.$$

The ratio test is inconclusive in this case. We already know, however, by Example 4.5, that this series diverges. $\square$

Theorem 4.4 (Comparison test). *Consider the two series $\sum_{i=1}^{\infty} a_i$ and $\sum_{i=1}^{\infty} b_i$, where the a_i's and b_i's are nonnegative for all sufficiently large i. Moreover, assume that $a_i = O(b_i)$. Then we have:*

(i) $\sum_{i=1}^{\infty} b_i < \infty$ *implies* $\sum_{i=1}^{\infty} a_i < \infty$;

(ii) $\sum_{i=1}^{\infty} a_i = \infty$ *implies* $\sum_{i=1}^{\infty} b_i = \infty$.

Example 4.8. Consider the series

$$\sum_{i=1}^{\infty} \frac{3i}{2^i - i}.$$

Applying the comparison test with the convergent series in Example 4.6, we see that since

$$\frac{3i}{2^i - i} = O\left(\frac{i}{2^i}\right),$$

we see that this series must converge as well. □

Example 4.9. Consider the series

$$\sum_{i=2}^{\infty} \frac{1}{\sqrt{i}(\ln(i))^2}.$$

Note that we start the series at $i = 2$, to prevent division by zero. Applying the comparison test with the divergent series in Example 4.5, we see that since

$$\frac{1}{i} = O\left(\frac{1}{\sqrt{i}(\ln(i))^2}\right),$$

we see that this series must diverge as well. □