1 Introduction

In number theory, combinatorics, calculus, computer science and other fields, it is useful to compare the rough order of magnitude of functions. The big-$O$, little-$o$ and related notation were invented for this purpose. They were created by Bachmann in the 1890s, popularized by Edmund Landau in the early 20th century, and expanded by Hardy, Vinogradov and Knuth.

1.1 Big-$O$ notation

Definition 1. Given functions $f : D \to \mathbb{C}$ and $g : D \to [0, \infty)$ defined on the same domain, the notation “$f = O(g)$” means that for some constant $C > 0$, $|f(x)| \leq Cg(x)$ for all $x \in D$. The notation “$f(x) = h(x) + O(g(x))$” means that $f(x) - h(x) = O(g(x))$.

The domain $D$ may be any set. Commonly $D$ is a subset of $\mathbb{R}$, $\mathbb{N}$, $\mathbb{R}^2$. Sometimes the domain is unspecified, if it is understood from the context.

Examples.

- $x^2 + 23x + 149 = O(x^2)$, on $D = [1, \infty)$.
- $x^2 = O(x^3)$ on $[1, \infty)$. FALSE for $D = [0, \infty)$ (hint: $x$ near zero)
- $(1 + x)^8 = 1 + 8x + O(|x|^2)$ for $|x| \leq 1$ (binomial theorem);
- $\sin x = O(1)$. Understood that the domain is $\mathbb{R}$.
- $\sin x = x + O(|x|^3)$ for $|x| \leq 1$; an example of a Taylor polynomial approximation.
- $n^{1/t} = O(1)$ for $(n, t) \in \mathbb{N} \times \mathbb{R}$. 
\begin{itemize}
  \item $xe^y = O(|x|)$, for $(x, y) \in \mathbb{R} \times (-\infty, 1]$.
  \item $[x] = x + O(1)$ for $x \in \mathbb{R}$;
  \item $\frac{xy}{x^2+y^2} = O(1)$ for $(x, y) \in \mathbb{R}^2 \setminus \{(0, 0)\}$.
  \item $x^{100} = O(e^x)$ on $[0, \infty)$. Is it true on $\mathbb{R}$?
  \item $\frac{1}{1+x} = 1 - x + x^2 + O(|x|^3)$. What is a valid domain?
\end{itemize}

For functions $f$ or $g$ of several variables, sometimes one variable is more important, and the other is treated as a parameter; e.g., $f(x) = x^a$, here $x$ is the main variable and $a$ is a parameter. One can indicate that there is a constant, depending on a parameter, by putting a subscript on the big-O-symbol.

\textbf{Examples.}

\begin{itemize}
  \item $(1 + x)^a = 1 + O_a(|x|)$, for $x \in \mathbb{C}$ with $|x| \leq \frac{1}{2}$, and $a \in \mathbb{R}$. Think of the first few terms in the binomial theorem.
  \item $x^a = O_a(e^x)$, $x \in [1, \infty)$, $a \geq 0$. This says that any power is eventually majorized by the exponential function.
  \item $\log x = O_\varepsilon(x^\varepsilon)$, $x \geq 1$, $\varepsilon > 0$. The point here is to take $\varepsilon$ as small as one likes, the implied constant depending on $\varepsilon$. This is a very useful inequality which is used frequently in number theory.
  \item $\pi(x) = \text{li}(x) + O_\varepsilon(x^{1/2+\varepsilon})$ for any $\varepsilon > 0$. This is the Riemann Hypothesis.
  \item $\log(1 - p^{-a}) \ll p^{-a}$ for $a \geq 1$, $p$ a prime. The implied constant is \textbf{absolute}, meaning that it does not depend on $p$ or on $a$.
\end{itemize}
**Definition 2** (Vinogradov, 1920s). The notation \( f \ll g \) means the same as \( f = O(g) \). Similarly, \( f \gg g \) means the same as \( g = O(f) \). Finally, \( f \asymp g \) means \( g \ll f \ll g \), that is, \( f \) and \( g \) have the same order of magnitude.

The notation \( f \asymp g \) for \( x \in D \) implies that there are constants \( c_1 \geq c_2 > 0 \) so that

\[
c_2 g \leq f \leq c_1 g \quad (x \in D).
\]

**Examples.**

- \( 12x^2 + 234x \log x \ll x^2 \) for \( x \geq 1 \).
- \( 12x^2 + 234x \log x = x^2 \) for \( x \geq 1 \).
- \[
\int_2^x \left( \frac{t}{\log t} \right)^2 + t \sin \sqrt{t} \, dt \ll \int_2^x \left( \frac{t^2}{\log^2 t} \right) \, dt \ll \frac{x^3}{\log^2 x}; \text{ this looks awful using } O\text{-notation.}
\]
- Let \( n \in \mathbb{N} \). \( O(n) = O(n^2) \) is true, \( O(n^2) = O(n) \) is false. The first means that if \( f(n) \ll n \) then \( f(n) \ll n^2 \) (with the second implied constant depending on the first one), the converse being false.
- \( \sin x = x \) for \( 0 < x \leq \pi/2 \).
- \( f(x) = g(x) + O(1) \) for \( x \in D \) is equivalent to \( e^{f(x)} = e^{g(x)} \) for \( x \in D \).
- If \( 0 < f(x) = O(1) \) then \( \log(1 + f(x)) = f(x) \). Again, the constants implied by the second relation depend on the implied constant in the \( O(1) \) bound.
- \( (1 + x)^a - 1 \asymp_{a,b} x \) for \( 0 \leq x \leq b, \ a > 0 \).
- In the expression \( e^{O(\sqrt{x})} \) it is understood that \( x \geq 0 \).

Alternative notation was created by Donald Knuth in 1976, and is most useful when used inside of an expression. It is popular in computer science and combinatorics, but not so much in number theory.

**Definition 3** (Knuth, 1976). \( f = \Omega(g) \) means the same as \( f \gg g \); \( f = \Theta(g) \) means \( f \asymp g \).

**Examples.**

- \( f(x) = e^{-\Omega(x)} \) (\( x \geq 1 \)) means that there is a constant \( C > 0 \) so that \( f(x) \geq e^{-Cx} \) for all \( x \geq 1 \).
- \( (1 + x)^8 = 1 + x + \Theta(x^2) \) for \( 0 < x < 1 \).
- \( (1 + x)^8 = x^8 + \Theta(x^7) \) for \( x \geq 1 \).
1.2 Little-o notation and relatives

Definition 4. If \( f : D \to \mathbb{R}, \ g : D \to \mathbb{R} \) are functions, then

\[ f(x) = o(g(x)) \ (x \to a) \quad \text{means} \quad \lim_{x \to a} \frac{f(x)}{g(x)} = 0. \]

Here \( a \) can be a finite or infinite quantity. The notation

\[ f(x) \sim g(x) \ (x \to a) \quad \text{means} \quad \lim_{x \to a} \frac{f(x)}{g(x)} = 1. \]

In this case, we say that \( f(x) \) is asymptotic to \( g(x) \) as \( x \to a \).

Informally, \( f(x) = o(g(x)) \) means that \( f \) has smaller order of growth than \( g \) as \( x \to a \).

The value of \( a \) is often unspecified when it is understood by context (usually \( a = \infty \) if unspecified).

Examples.

- \( x^2 = o(x^3) \ (x \to \infty) \).
- \( 12x^2 + 234x \log x \sim 12x^2 \ (x \to \infty) \).
- \( \log x = o_{\varepsilon}(x^{\varepsilon}) \) as \( x \to \infty \), for any \( \varepsilon > 0 \). the subscript means that the rate of convergence of the limit depends on \( \varepsilon \).
- \( x^a = o_{a,b}(e^{bx}) \) as \( x \to \infty \), for any positive \( a, b \).
- \( \sin x \sim x \) as \( x \to 0 \).
- \( \pi(x) \sim \frac{x}{\log x} \); the Prime Number Theorem.
- \( f(x) = g(x) + o(1) \) as \( x \to a \) is equivalent to \( e^{f(x)} \sim e^{g(x)} \) as \( x \to a \).
- True or False: \( f(x) = o(g(x)) \) as \( x \to \infty \) implies that \( e^{f(x)} = o(e^{g(x)}) \) as \( x \to \infty \) ?

What about the converse?
- \( f(x) = o(1) \) \( (x \to a) \) is another way of writing \( \lim_{x \to a} f(x) = 0 \).
- \( \frac{1}{x-y} = o_y(1) \) as \( x \to \infty \), for any fixed \( y \);
- \( \arctan(y/x) \sim_x \pi/2 \) as \( y \to \infty \) for any fixed \( x > 0 \).
- \( \arctan(y/x) \sim_x -\pi/2 \) as \( y \to \infty \) for any fixed \( x < 0 \).
- \( \sum_{n=1}^{\infty} \frac{1}{n^s} \sim \frac{1}{s-1} \) as \( s \to 1^+ \). A basic property of the Riemann zeta function.
Definition 5. If $S$ is an infinite subset of the positive integers, we say that property $P$ holds for almost all $n \in S$ if

$$\#\{n \leq x : n \in S, n \text{ has property } P\} \sim \#\{n \leq x : n \in S\} \quad (x \to \infty).$$

In other words, the set of $n \in S$ below $x$ that do not satisfy property $P$ is relatively small compared to the set of all $n \in S, n \leq x$.

If $S$ is not specified, it is understood that $S = \mathbb{N}$.

Examples.

- Almost all positive integers are composite.
- Almost all primes are odd.
- Almost all even $n \geq 2$ are the sum of two primes (a theorem of Vinogradov).
- For almost all primes $p$, $p + 2$ is composite (a consequence of sieve theory).
- For every $\varepsilon > 0$, $n$ has between $(1 - \varepsilon) \log \log n$ and $(1 + \varepsilon) \log \log n$ distinct prime factors for almost every $n$. (Note the variation “almost every”). This is a famous theorem of Hardy and Ramanujan.