Chapter 2

Asymptotic notations

2.1 The "oh" notations

Terminology	**Notation**	**Definition**
Big oh notation	$f(s) = O(g(s)) \quad (s \in S)$	There exists a constant c such that $\|f(s)\| \leq c\|g(s)\|$ for all $s \in S$
Vinogradov notation	$f(s) \ll g(s) \quad (s \in S)$	Equivalent to "$f(s) = O(g(s))$ $(s \in S)$"
Order of magnitude estimate	$f(s) \asymp g(s) \quad (s \in S)$	Equivalent to "$f(s) \ll g(s)$ and $g(s) \ll f(s)$ $(s \in S)$".
Small oh notation	$f(s) = o(g(s)) \quad (s \to s_0)$	$\lim_{s \to s_0} f(s)/g(s) = 0$
Asymptotic equivalence	$f(s) \sim g(s) \quad (s \to s_0)$	$\lim_{s \to s_0} f(s)/g(s) = 1$
Omega estimate	$f(s) = \Omega(g(s)) \quad (s \to s_0)$	$\limsup_{s \to s_0} \|f(s)/g(s)\| > 0$.

Table 2.1: Overview of asymptotic terminology and notation. In these definitions S denotes a set of real or complex numbers contained in the domain of the functions f and g, and s_0 denotes a (finite) real or complex number or $\pm\infty$.

A very convenient set of notations in asymptotic analysis are the so-called "big oh" (O) and "small-oh" (o) notations, and their variants. These notations are in widespread use and are often used without further explanation. However, in order to properly apply these notations and avoid mistakes resulting from careless use, it is important to be aware of their precise definitions.

In this section we give formal definitions of the "oh" notations and their variants, show how to work with these notations, and illustrate their use with a number of examples. Tables 2.1 and 2.2 give an overview of these notations.

Short-hand form	**Full form**
$f(s) = O(g(s)) \quad (s \to s_0)$	There exists a constant $\delta > 0$ such that $f(s) = O((g(s))$ $(\lvert s - s_0\rvert \le \delta)$.
$f(x) = O(g(x))$	There exists a constant x_0 such that $f(x) = O((g(x))$ $(x \ge x_0)$.
$f(x) = o(g(x))$	$f(x) = o(g(x))$ $(x \to \infty)$.

Table 2.2: Notational conventions and shortcuts for commonly occurring asymptotic expressions.

2.1.1 Definition of "big oh", special case

We consider first the simplest and most common case encountered in asymptotics, namely the behavior of functions of a real variable x as $x \to \infty$. Given two such functions $f(x)$ and $g(x)$, defined for all sufficiently large real numbers x, we write

$$f(x) = O(g(x))$$

as short-hand for the following statement: *There exist constants x_0 and c such that*

$$|f(x)| \le c|g(x)| \quad (x \ge x_0).$$

If this holds, we say that $f(x)$ **is of order** $O(g(x))$, and we call the above estimate a O**-estimate** ("big oh estimate") for $f(x)$. The constant c called the O**-constant**, and the range $x \ge x_0$ the **range of validity** of the O-estimate.

In exactly the same way we define the relation "$f(n) = O(g(n))$" if f and g are functions of an integer variable n.

Note that the O-constant c is not unique; if the above inequality holds with a particular value c, then obviously it also holds with any constant c' satisfying $c' > c$. Similarly, the constant x_0 implicit in the range of an O-estimate may be replaced by any constant x_0' satisfying $x_0' \geq x_0$.

The value of the O-constant c is usually not important; all that matters is that such a constant exists. In fact, in many situations it would be quite tedious (though, in principle, possible) to work out an explicit value for c, even if we are not interested in getting the best-possible value. **The beauty of the O-notation is that it allows us to express, in a succinct and suggestive manner, the existence of such a constant without having to write down the constant.**

Example 2.1. We have

$$x = O(e^x).$$

Proof. By the definition of an O-estimate, we need to show that there exist constants c and x_0 such that $x \leq ce^x$ for all $x \geq x_0$. This is equivalent to showing that the quotient function $q(x) = x/e^x$ is bounded on the interval $[x_0, \infty)$, with a suitable x_0. To see this, observe that the function $q(x)$ is nonnegative and continuous on the interval $[0, \infty)$, equal to 0 at $x = 0$ and tends to 0 as $x \to \infty$ (as follows, e.g., from l'Hopital's Rule). Thus this function is bounded on $[0, \infty)$, and so the O-estimate holds with $x_0 = 0$ and any value of c that is an upper bound for $q(x)$ on $[0, \infty)$.

In this simple example it is easy to determine an explicit, and best-possible, value for the O-constant c. Indeed, the above argument shows that the best-possible constant is $c = \max_{0 \leq x < \infty} q(x)$, with $q(x) = xe^{-x}$, and the maximum of $q(x)$ is easily determined by methods of calculus: The derivative of $q(x)$ equals $q'(x) = e^{-x}(1-x)$, which has a unique zero at $x = 1$. Thus that $q(x)$ attains its maximal value at $x = 1$, and so $c = q(1) = e^{-1}$ is the best-possible constant for the range $x \geq 0$. □

Example 2.2. The argument used in the previous example works in much more general situations. For example, consider the functions $f(x) = (x+1)^A$ and $g(x) = \exp((\log x)^{1+\epsilon})$, where A and ϵ are arbitrary positive constants. Since the quotient $q(x) = f(x)/g(x)$ of these two functions is nonnegative and continuous on the interval $[1, \infty)$ and tends to 0 as $x \to \infty$, it is bounded on this interval. Thus, we have the O-estimate

$$(x + 1)^A = O\left(\exp\left((\log x)^{1+\epsilon}\right)\right)$$

in the range $x \geq 1$. However, in contrast to the previous example, working out an explicit value for the O-constant c would be rather tedious. The

O-notation allows us to ignore these complications: all we need to know is the *existence* of a constant, and this, as we have seen, is easy to establish with general continuity or compactness arguments.

Example 2.3. If $P(x) = \sum_{k=0}^{n} a_k x^k$ is a polynomial of degree n, then

$$P(x) = O(x^n).$$

Proof. For $x \geq 1$ we have

$$|P(x)| \leq \sum_{i=0}^{n} |a_i| x^i \leq \left(\sum_{i=0}^{n} |a_i| \right) x^n,$$

so the required inequality holds with $x_0 = 1$ and $c = \sum_{i=0}^{n} |a_i|$. □

Example 2.4. The relation

$$f(x) = O(1)$$

simply means that $f(x)$ is bounded as $x \to \infty$.

2.1.2 Dependence on parameters

In many cases, the functions involved in an O-estimate depend on one or more parameters. It may then be important to know whether the O-constant depends on these parameters or can be chosen independently of the parameters. If the constant (possibly) depends on one or more parameters, it is customary to indicate this dependence by placing the parameters as subscripts to the O-symbol and writing, for example, O_λ, O_k, or $O_{k,\epsilon}$. The same convention applies, if the constant depends on a parameter arising in the range of an estimate (rather than the functions to be estimated).

To avoid mistakes, it is a good practice to explicitly indicate the dependence of O-estimates on any parameters by using the subscript notation, and we will generally adhere to this practice.

If it is possible to choose the constant in an O-estimate independent of some parameter occurring in the definition of the function or the range of the estimate, we say that the estimate is **uniform** (or **holds uniformly**) with respect to the given parameter. Uniform estimates are more informative and more useful than nonuniform estimates, and obtaining uniform estimates or making non-uniform estimates uniform (e.g., by making the dependence on parameters explicit) is a desirable goal.

Example 2.5. In Example 2.2 we showed, by a simple continuity argument, that, for any positive constants A and ϵ, we have $(x+1)^A = O(\exp((\log x)^{1+\epsilon}))$ in the range $x \geq 1$. While, in this case, the range $x \geq 1$ could be chosen independently of the constants A and ϵ, this is not true for the O-constant c. Thus, to indicate the (possible) dependence of the O-constant on A and ϵ, we should write this O-estimate more precisely as

$$(x+1)^A = O_{A,\epsilon}\left(\exp\left((\log x)^{1+\epsilon}\right)\right) \quad (x \geq 1).$$

In general, the subscript notation simply says that the constant *may* depend on the indicated parameters, not that it is not possible (for example, through a more clever argument) to find a constant independent of the parameters. However, in this particular example, it is easy to see that the constant *necessarily* has to depend on both parameters A and ϵ.

2.1.3 Definition of "big oh", general case

If $f(s)$ and $g(s)$ are functions of a real or complex variable s and S is an arbitrary set of (real or complex) numbers s (belonging to the domains of f and g), we write

$$f(s) = O(g(s)) \quad (s \in S),$$

if there exists a constant c such that

$$|f(s)| \leq c|g(s)| \quad (s \in S).$$

To be consistent with our earlier definition of "big oh" we make the following convention: **If a range is not explicitly given, then the estimate is assumed to hold for all sufficiently large values of the variable involved, i.e., in a range of the form $x \geq x_0$, for a suitable constant x_0.**

Example 2.6. Given any positive constant $r < 1$, we have

$$\log(1+z) = O_r(|z|) \quad (|z| < r).$$

Proof. Note that the function $\log(1+z)$ is analytic in the open unit disk $|z| < 1$ and has power series expansion

$$\log(1+z) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} z^n \quad (|z| < 1).$$

Hence, in $|z| < 1$ we have

$$|\log(1+z)| \le \sum_{n=1}^{\infty} \frac{1}{n}|z^n| \le \sum_{n=1}^{\infty} |z|^n = \frac{1}{1-|z|}|z|.$$

If now z is restricted to the disk $|z| < r$ (with $r < 1$), then the above bound becomes $\le (1-r)^{-1}|z|$, so the required inequality holds with constant $c = c(r) = (1-r)^{-1}$. (This is an example where the O-constant depends on a parameter occurring the definition of the range.) $\square$

Further generalizations to functions of more than one variable can be made in an obvious manner.

Example 2.7. For any positive real number p we have

$$(x+y)^p = O_p\left(x^p + y^p\right) \quad (x, y \ge 0).$$

More generally, we have, for any positive integer n and any positive real number p,

$$(a_1 + \cdots + a_n)^p = O_{n,p}\left(a_1^p + \cdots + a_n^p\right) \quad (a_1, \ldots, a_n \ge 0),$$

where now the O-constant depends on both n and p.

Proof. The estimate (in the second, more general, form) can be proved via Hölder's inequality; alternatively, it follows immediately from the simple observation

$$\left(\sum_{i=1}^{n} a_i\right)^p \le \left(n \max_i a_i\right)^p = n^p (\max_i a_i)^p \le n^p \sum_{i=1}^{n} a_i^p. \qquad \square$$

2.1.4 "Oh" terms in arithmetic expressions

By a term $O(g(s))$ in an arbitrary arithmetic expression we mean a function $f(s)$ that satisfies the inequality in the definition of the O-estimate. In other words, an O-term can be thought of as a "black box" hiding some unknown function, and the only information we have about this function is that it satisfies the appropriate inequality.

This is a natural and useful convention that greatly simplifies the notation when working with O-expressions. For example, this convention allows us to write the relation

$$\log(1+x) - x = O(x^2) \quad (|x| \le 1/2)$$

more naturally as

$$\log(1+x) = x + O(x^2) \quad (|x| \le 1/2).$$

The latter can be thought of as a succinct form of the following rather unwieldy statement. "$\log(1+x)$ *is equal to* x *plus a function that, in absolute value, is bounded by a constant times* x^2 *in the range* $|x| \le 1/2$*.*"

Example 2.8. Power series expansions naturally lead to O-estimates in the above more generalized sense. In particular, if $f(z)$ is a function analytic in some disk $|z| < R$, then for any $r < R$ and any fixed positive integer n, we have, by Taylor's theorem,

$$f(z) = \sum_{k=0}^{n} a_k z^k + O_{r,n}(|z|^{n+1}) \quad (|z| < r),$$

where the a_k are the Taylor coefficients of $f(z)$.

Example 2.9. A term $O(1)$ simply stands for a bounded function. For example, the "floor function" $[x]$ satisfies

$$[x] = x + O(1),$$

since $|[x] - x| \le 1$.

2.1.5 The Vinogradov "$\ll$" notation

This notation was introduced by the Russian number theorist I.M. Vinogradov as an alternative to the O-notation. Along with the closely related notations "$\gg$" and "$\asymp$", it has all become standard in number theory, though it is less common in other areas of mathematics. In the case of functions of a real variable x and (implicit) ranges of the form $x \ge x_0$, these three notations are defined as follows:

- "$f(x) \ll g(x)$" is equivalent to "$f(x) = O(g(x))$".
- "$f(x) \gg g(x)$" is equivalent to "$g(x) \ll f(x)$".
- "$f(x) \asymp g(x)$" means that both "$f(x) \ll g(x)$" and "$g(x) \gg f(x)$" hold.

These definitions generalize in an obvious manner to more general functions and ranges.

If $f(x) \asymp g(x)$, we say that $f(x)$ **and** $g(x)$ **have the same order of magnitude**. From the definition it is easy to see that "$f(x) \asymp g(x)$" holds if and only if *there exist positive constants* c_1 *and* c_2 *and a constant* x_0 *such that*

$$c_1|g(x)| \le |f(x)| \le c_2|g(x)| \quad (x \ge x_0). \tag{2.1}$$

As with the O-notation, dependence on parameters may be indicated by putting the parameters as subscripts to the "$\ll$" or "$\gg$" symbols. For example, the estimate $(x+1)^A = O_{A_\epsilon}(\exp((\log x)^{1+\epsilon}))$, which we considered in Example 2.2, could have been written in the equivalent form

$$(x+1)^A \ll_{A,\epsilon} \exp\left((\log x)^{1+\epsilon}\right).$$

The primary advantage of the Vinogradov notation over the O-notation is a typographical one: If the function $g(x)$ is a complicated expression (for example, a sum of several integrals), then $f(x) \ll g(x)$ looks much cleaner than $f(x) = O(g(x))$ (which would require an oversized set of parentheses). In addition, the Vinogradov notation provides an easy way to express lower bounds by using the symbol "$\gg$" instead of "$\ll$", and the "$\asymp$" symbol allows one to express two O-estimates in a single statement.

The Vinogradov notation has the drawback that, unlike the O-notation, it does not extend to terms in arithmetic expressions. Thus, for example, while one can rewrite the estimate

$$\pi(x) - \frac{x}{\log x} = O\left(\frac{x}{\log x)^2}\right)$$

in an equivalent manner as

$$\pi(x) = \frac{x}{\log x}\left(1 + O\left(\frac{1}{\log x}\right)\right),$$

only the first version can be stated using the Vinogradov "$\ll$" notation:

$$\pi(x) - \frac{x}{\log x} \ll \frac{x}{(\log x)^2}.$$

Thus, depending on the situation, one or the other of these two notations may be more convenient to use, and we will use both notations interchangeably throughout this course, rather than settle on one particular type of notation.

Example 2.10. For any positive integer n and any positive real number p we have

$$(a_1 + \cdots + a_n)^p \asymp_{p,n} a_1^p + \cdots + a_n^p \quad (a_1, \ldots, a_n \geq 0).$$

Proof. The upper bound of this estimate (i.e., the "$\ll$" portion of "$\asymp$") was established (quite easily) in Example 2.7, but the proof of the lower bound is just as simple: Since

$$a_1^p + \cdots + a_n^p \leq n(\max(a_1, \ldots, a_n))^p \leq n\,(a_1 + \cdots + a_n)^p\,,$$

we obtain the "$\gg$" portion of the estimate with constant $1/n$. □

This example is a good illustration of the benefits of the "$\asymp$" notation. With this notation, the asserted two-sided estimate we claimed takes a concise, and suggestive, one-line form, whereas the same estimate in the O-notation would have required two somewhat clumsy looking O-relations.

Example 2.11. We have

$$\sqrt{\log y} \asymp \sqrt{\log x} \quad (x^{1/2} \leq y \leq x^2, x \geq 1).$$

Proof. This follows immediately on noting that the function $f(y) = \sqrt{\log y}$ is increasing and satisfies

$$f(x^{1/2}) = \sqrt{\log x^{1/2}} = 2^{-1/2}\sqrt{\log x} = 2^{-1/2} f(x) \quad (x \geq 1).$$

and, similarly, $f(x^2) = 2^{1/2} f(x)$. □

Example 2.12. If $f(x)$ and $g(x)$ are positive functions, then

$$f(x) \asymp g(x)$$

holds if and only if

$$\log f(x) = \log g(x) + O(1).$$

This follows immediately from the explicit version (2.1) of the relation "$\asymp$".

2.1.6 Other variants of the O-notation

Some other notations that are equivalent to or related to the O-notation and which are occasionally used are the following. All of these notations are non-standard and do not have a generally accepted meaning, so they should be avoided, or at least precisely defined before use.

- In some areas of analysis (especially harmonic analysis), the symbol "$\lesssim$" is used with the same meaning as "$\ll$".
- The symbol "$\lll$" is sometimes used to indicate that one function is "of smaller order of magnitude" than another function, usually in the sense that the ratio between the two functions tends to 0 (i.e., the equivalent of the o-notation defined below). In their book "Concrete Mathematics", Graham, Knuth, and Patashnik use the symbol "$\prec$" in the same sense. However, neither of these notation is very widespread.
- In numerical applications the value of an O-constant is important. One notation that refines the O-notation by keeping track of constants is the θ-notation, which means the same as the O-notation with constant $c = 1$. For example, since $|\log(1+z)| \leq \sum_{n=1}^{\infty} |z|^n/n \leq |z|/(1-|z|) \leq 2|z|$ for $|z| \leq 1/2$, we have, using the θ-notation, $\log(1+z) = \theta(2|z|)$ for $|z| \leq 1/2$.
- The symbol "$\approx$" is sometimes used with the same meaning as $\asymp$. However, more commonly, this symbol is used in an informal manner (e.g., in heuristic arguments) to indicate that one quantity is "approximately" equal to another quantity.

2.1.7 The "small oh" notation and asymptotic equivalence

The notation

$$f(x) = o(g(x)) \quad (x \to \infty)$$

means that $g(x) \neq 0$ for sufficiently large x and $\lim_{x\to\infty} f(x)/g(x) = 0$. If this holds, we say that $f(x)$ **is of smaller order than** $g(x)$. This is equivalent to having an O-estimate $f(x) = O(g(x))$ with a constant c that can be chosen arbitrarily small (but positive) and a range $x \geq x_0(c)$ depending on c. Thus, an o-estimate is stronger than the corresponding O-estimate.

A closely related notation is that of asymptotic equivalence:

$$f(x) \sim g(x) \quad (x \to \infty)$$

means that $g(x) \neq 0$ for sufficiently large x and $\lim_{x\to\infty} f(x)/g(x) = 1$. If this holds, we say that $f(x)$ **is asymptotic (or "asymptotically equivalent") to $g(x)$ as $x \to \infty$.** Just as an o-estimate refines the O-estimate, the asymptotic equivalence relation $f(x) \sim g(x)$ refines the order of magnitude estimate $f(x) \asymp g(x)$.

By an **asymptotic formula** for a function $f(x)$ we mean a relation of the form $f(x) \sim g(x)$, where $g(x)$ is a "simple" function.

In much the same way as the O-notation, the o-notation can be generalized to functions for complex variables, and to more general limits: If $f(s)$ and $g(s)$ are functions of a real or complex variable s and s_0 is a real or complex number or infinity, we write

$$f(s) = o(g(s)) \quad (s \to s_0),$$

if the limit $\lim_{s\to s_0} f(s)/g(s)$ exists and is equal to 0, Asymptotic formulas with respect to the limit $s \to s_0$ are defined analogously.

It is important to keep in mind that the o-notation is always with respect to a given limiting process. If a limiting process is not explicitly given (in a form like "$x \to x_0$"), the limit is usually understood to be taken as the variable tends to infinity.

In the same way as we have done with the O-notation, we allow o-terms to appear inside arithmetic expressions: a term $o(g(x))$ stands for a function $f(x)$ that satisfies $\lim_{x\to\infty} f(x)/g(x) = 0$ (but on which we have no further information). With this convention the asymptotic formula $f(x) \sim g(x)$ is easily seen to be equivalent to either of the relations

$$f(x) = g(x) + o(g(x))$$

or

$$f(x) = g(x)(1 + o(1)).$$

Another related notation that is used, for example, in number theory, is the Ω-notation. This notation simply means the opposite of "small oh": Namely, we write

$$f(x) = \Omega(g(x)) \quad (x \to \infty),$$

if the relation $f(x) = o(g(x))$ is *false*, i.e., if $\limsup_{x\to\infty} |f(x)/g(x)| > 0$. Analogous definitions apply for the case of more general functions or limits. For example, we have $\sin x = \Omega(1)$ as $x \to \infty$, and $\sin x = \Omega(x)$ as $x \to 0$.

Note that the relation $f(x) = \Omega(g(x))$ is *not* equivalent to $f(x) \gg f(x)$. Indeed, the latter means that $|f(x)| > c|g(x)|$ holds, with some positive constant c, for *all* sufficiently large x, whereas $f(x) = \Omega(g(x))$ only requires this inequality to hold for *arbitrarily large* values of x.

2.1.8 O-estimates versus o-estimates

An o-estimate is a qualitative, rather than quantitative, statement: $f(x) = o(g(x))$ simply means that the quotient $f(x)/g(x)$ tends to 0 as $x \to \infty$, but it says nothing about the rate of convergence. In almost all cases where o-estimates (or, equivalently, asymptotic formulas) are known, these estimates arise as corollaries to more precise O-estimates: An O-estimate of the form $f(x) = O(g(x)/\psi(x))$ with some explicit function $\psi(x)$ (such as $\psi(x) = \log x$) that tends to infinity as $x \to \infty$ implies the o-estimate $f(x) = o(g(x))$ and provides more information. The chief advantage of o-estimates and asymptotic formulas is that they are easy to state and make for clean and easy-to-remember theorems. However, in the course of proving such estimates, it is almost always advisable to carry the argument through with O-estimates, and only at the very end, if necessary, make the transition to an o-estimate. The main reason for this is that working with o-terms is fraught with pitfalls, whereas O-terms can be manipulated fairly easily and safely, as we will show below.

2.1.9 An illustration: Estimates for the prime counting function

To illustrate the various notations introduced here, we present a list of estimates for the prime counting function $\pi(x)$, the number of primes $\leq x$, which have been proved over the past century or so, or put forth as conjectures. Each of these estimates represented a major milestone in our understanding of the behavior of $\pi(x)$.

Chebysheff's estimate: This estimate establishes the correct order of magnitude of $\pi(x)$:

$$\pi(x) \asymp \frac{x}{\log x} \quad (x \geq 2).$$

The Prime Number Theorem (PNT): In its simplest and most basic form, the PNT gives an asymptotic formula for $\pi(x)$:

$$\pi(x) \sim \frac{x}{\log x} \quad (x \to \infty).$$

This result, arguably the most famous result in number theory, had been conjectured by Gauss, who, however, was unable to prove it. It was eventually proved in the late 19th century, independently and at about the same time, by Jacques Hadamard and Charles de la Vallée Poussin.

PNT with modest error term: A more precise version of the above form of the PNT shows that the *relative* error in the above asymptotic formula is of order $O(1/\log x)$:

$$\pi(x) = \frac{x}{\log x}\left(1 + O\left(\frac{1}{\log x}\right)\right) \quad (x \geq 2).$$

This version, while far from the best-known version of the PNT, is sharp enough for many applications.

PNT with "classical" error term: To be able to state more precise versions of the PNT, the function $x/\log x$ as approximation to $\pi(x)$ is too crude; a better approximation is provided by the "logarithmic integral",

$$\mathrm{Li}(x) = \int_2^x \frac{dt}{\log t} \quad (x \geq 2).$$

With $\mathrm{Li}(x)$ as main term in the approximation to $\pi(x)$, the relative error in the approximation can be shown to be much smaller than any negative power of $\log x$. Indeed, the analytic method introduced by Hadamard and de la Vallée Poussin in their proof of the PNT yields the estimate

$$\pi(x) = \mathrm{Li}(x)\left(1 + O\left(\exp(-c\sqrt{\log x})\right)\right) \quad (x \geq 3),$$

where c is a positive constant. This result, which is now more than 100 years old, can be considered the "classical" version of the PNT with error term.

PNT with Vinogradov-Korobov error term: The only significant improvement in the error term for the PNT obtained during the past 100 years is due to I.M. Vinogradov and A. Korobov, who improved the above classical estimate to

$$\pi(x) = \mathrm{Li}(x)\left(1 + O_\epsilon\left(\exp(-(\log x)^{3/5-\epsilon}\right)\right) \quad (x \geq 3),$$

for any given $\epsilon > 0$. The Vinogradov-Korobov result is some 50 years old, but it still represents essentially the sharpest known form of the PNT.

PNT with conjectured error term: A widely believed conjecture is that the "correct" relative error in the PNT should be about $1/\sqrt{x}$. More precisely, the conjecture states that

$$\pi(x) = \mathrm{Li}(x)\left(1 + O_\epsilon\left(x^{-1/2+\epsilon}\right)\right) \quad (x \geq 3)$$

holds for any given $\epsilon > 0$. This conjecture is known to be equivalent to the Riemann Hypothesis. It is interesting to compare the size of the (relative) error term in this *conjectured* form of the PNT with that in the sharpest *known* form of the PNT, i.e., the Vinogradov-Korobov estimate cited above: To this end, note that

$$\exp\left(-(\log x)^{3/5-\epsilon}\right) \geq \exp\left(-(\log x)^{3/5}\right) \gg_\epsilon x^{-\epsilon} \quad (x \geq 3)$$

for any $\epsilon > 0$. Thus, while the conjectured form of the PNT involves a relative error of size $O_\alpha(x^{-\alpha})$ for *any* fixed exponent $\alpha < 1/2$, our present knowledge does not even give such an estimate for *some* positive value of α.

Omega estimate: It is known that the relative error in the PNT cannot be of order $O(x^{-\alpha})$ with an exponent $\alpha > 1/2$. Using the "Omega" notation introduced above, this can be expressed as follows: For any $\alpha > 1/2$, we have

$$\pi(x) - \mathrm{Li}(x) = \Omega\left(\mathrm{Li}(x)x^{-\alpha}\right) \quad (x \to \infty).$$

2.2 Working with the "oh" notations

Recall that an O-term in an arithmetic expression or an equation represents a function that satisfies the inequality implicit in the definition of an O-estimate. With this convention, expressions involving several O-terms have a well-defined meaning. However, we have to be careful when working with such terms as these are not ordinary arithmetic expressions and cannot be manipulated in the same way. Fortunately, most arithmetic operations are permissible with O-terms.

2.2.1 Rules for "big oh" and "small oh" estimates

We now list some basic rules for manipulating O-terms. For simplicity, we state these only for functions of a real variable x and do not explicitly indicate the range (which thus, by our convention, is of the form $x \geq x_0$). However, the same rules hold in the more general context of functions of a complex variable s and O-estimates valid in a general range $s \in S$.

- **Constants in O-terms:** If C is a positive constant, then the estimate $f(x) = O(Cg(x))$ is equivalent to $f(x) = O(g(x))$. In particular, the estimate $f(x) = O(C)$ is equivalent to $f(x) = O(1)$.

- **Transitivity:** O-estimates are transitive, in the sense that if $f(x) = O(g(x))$ and $g(x) = O(h(x))$, then $f(x) = O(h(x))$.

- **Multiplication of O-terms:** If $f_i(x) = O(g_i(x))$ for $i = 1, 2$, then $f_1(x)f_2(x) = O(g_1(x)g_2(x))$.

- **Pulling out factors:** If $f(x) = O(g(x)h(x))$, then $f(x) = g(x)O(h(x))$. This property allows us to factor out main terms from O-expressions. For example, we can write the relation $f(x) = x + O(x/\log x)$ as $f(x) = x(1 + O(1/\log x))$. The latter relation is more natural as it clearly shows the *relative* error in the approximation of $f(x)$ by x.

- **Summation of O-terms:** If $f_i(x) = O(g_i(x))$ for $i = 1, 2, \ldots, n$, where the O-constants are independent of i, then

$$\sum_{i=1}^{n} f_i(x) = O\left(\sum_{i=1}^{n} |g_i(x)|\right).$$

In other words, O's can be pulled out of sums, provided the summands are replaced by their absolute values and the O-constants do not depend on the summation index. The same holds for infinite series $\sum_{i=1}^{\infty} f_i(x)$ in which each term satisfies an O-estimate of the above type (again with an O-constant that is independent of the summation index i).

- **Integration of O-terms:** If $f(x)$ and $g(x)$ are integrable on finite intervals and satisfy $f(x) = O(g(x))$ for $x \ge x_0$, then

$$\int_{x_0}^{x} f(y)dy = O\left(\int_{x_0}^{x} |g(y)|dy\right) \quad (x \ge x_0).$$

In other words, O's can be pulled out of or integrals provided the integrand is replaced by its absolute value.

Proofs. These rules are straightforward consequences of the definition of an O-estimate. As an example, we give a proof for the last rule. Suppose $f(x)$ and $g(x)$ are integrable on finite intervals and satisfy $f(x) = O(g(x))$ for $x \ge x_0$. Thus there exists a constant c such that $|f(x)| \le c|g(x)|$ holds for all $x \ge x_0$. But then we have, for $x \ge x_0$,

$$\left|\int_{x_0}^{x} f(y)dy\right| \le \left|\int_{x_0}^{x} |f(y)|dy\right| \le c\left|\int_{x_0}^{x} |g(y)|dy\right|.$$

Hence

$$\int_{x_0}^{x} f(y)dy = O\left(\int_{x_0}^{x} |g(y)|dy\right) \quad (x \geq x_0),$$

as desired. □

Rules for o-estimates. Some, but not all, of the above rules for O-estimates carry over to o-estimates. For example, the first four rules also hold for o-estimates. On the other hand, this is not the case for the last two rules. For instance, if $f(x) = e^{-x}$ and $g(x) = 1/x^2$, then $f(x) = o(g(x))$ as $x \to \infty$. On the other hand, the integrals $F(x) = \int_1^x f(x)dy$ and $G(x) = \int_1^x g(y)$ are equal to $e^{-1} - e^{-x}$ and $1 - 1/x$, respectively, and satisfy $\lim_{x\to\infty} F(x)/G(x) = e^{-1}$, so the relation $F(x) = o(G(x))$ does not hold. This example illustrates the difficulties and pitfalls that one may encounter when trying to manipulate o-terms. To avoid these problems, it is advisable to work with O-estimates rather than o-estimates, whenever possible.

2.2.2 Equations involving O-terms

In all examples we considered so far, all O-terms occurred on the right-hand side of the equation. It is useful to further extend the usage of the O-notation by allowing equations in which O-terms arise on both sides, provided one takes care in properly interpreting such an equation. In particular, **equations in which there are O-terms on both sides are not symmetric and should be read left to right.** For example, the relation

$$O(\sqrt{x}) = O(x) \quad (x \geq 1),$$

is to be understood in the sense that *any function $f(x)$ satisfying $f(x) = O(\sqrt{x})$ for $x \geq 1$ also satisfies $f(x) = O(x)$ for $x \geq 1$*, a statement that is obviously true. On the other hand, if we interchange the left- and right-hand sides of the above equation, we get

$$O(x) = O(\sqrt{x}) \quad (x \geq 1),$$

which, when interpreted in the same way (i.e., read left to right), is patently false.

For similarly obvious reasons, O-terms in equations cannot be cancelled; after all, each O-term stands for a function satisfying the appropriate O-estimate, and multiple instances of the same O-term (say, multiple terms $O(x)$) in general it will represent different functions. For example, from $f(x) = \log x + O(1/x)$ and $g(x) = \log x + O(1/x)$ we can only conclude that

$f(x) = g(x) + O(1/x)$, i.e., that $f(x)$ and $g(x)$ differ (at most) by a term of order $O(1/x)$, but not that $f(x)$ and $g(x)$ are equal.

2.2.3 Simplifying O-expressions

The following are some transformation rules which often allow one to dramatically simplify messy expressions involving O-terms.

$$\begin{aligned} \frac{1}{1+O(\phi(x))} &= 1 + O(\phi(x)), \\ (1+O(\phi(x)))^p &= 1 + O_p(\phi(x)), \\ \log(1+O(\phi(x))) &= O(\phi(x)), \\ \exp(O(\phi(x))) &= 1 + O(\phi(x)). \end{aligned}$$

Table 2.3: Some common transformations of O-expressions, valid when $\phi(x) \to 0$. Here p is any real or complex parameter.

These relations are to be interpreted from left to right as described in the preceding subsection. For example, the first estimate means that any function $f(x)$ satisfying $f(x) = 1/(1 + O(\phi(x)))$ also satisfies $f(x) = 1 + O(\phi(x))$.

The above relations follow immediately from the following basic O-estimates, which are easily proved (e.g., via the first-order Taylor formula):

$$\begin{aligned} \frac{1}{1+z} &= 1 + O(|z|), \\ (1+z)^p &= 1 + O_p(|z|), \\ \log(1+z) &= O(|z|), \\ e^z &= 1 + O(|z|), \end{aligned}$$

Table 2.4: Some basic O-estimates, valid for $z \to 0$, i.e., with a range $|z| \leq \delta$, for a suitable constant $\delta > 0$. Here p is any real or complex parameter.

2.2.4 Some asymptotic tricks

Factoring out dominant terms. A simple, but very effective technique in asymptotic analysis is to identify a dominant term in an estimate and then factor out this term. This often facilitates subsequent estimates, and it leads to a relation that clearly displays the *relative* error, which is usually more informative than the *absolute* error.

Example 2.13. As a simple example illustrating this technique, we try to determine the behavior of the function

$$f(x) = \sqrt{x^2 + 1}.$$

as $x \to \infty$. We begin by noting that the term x^2 is the dominant term under the square root sign, so we expect $f(x)$ to be close to $\sqrt{x^2} = x$. To make this precise, we factor out the term x^2, to get $f(x) = x\sqrt{1 + 1/x^2}$. Since for $x \geq 2$ we have $1/x^2 \leq 1/4$, we can estimate $\sqrt{1 + 1/x^2}$ using the binomial series expansion of $(1+y)^\alpha$, which is valid, for example, in $|y| \leq 1/2$. Taking only the first term gives $\sqrt{1 + 1/x^2} = 1 + O(x^{-2})$, and hence

$$f(x) = x\left(1 + O\left(\frac{1}{x^2}\right)\right) = x + O\left(\frac{1}{x}\right).$$

Taking more terms in the series would lead to correspondingly more precise estimates for $f(x)$.

Example 2.14. The technique of factoring out dominant can also be useful when applied only to parts of an arithmetic expression, such as the argument of a logarithm or the denominator of a fraction. For example, let

$$f(x) = \log(\log x + \log\log x).$$

In the argument of the logarithm the term $\log x$ is dominant. We factor out this term, use the functional equation of the logarithm along with the expansion $\log(1 + y) = y + O(y^2)$, which is valid in $|y| \leq 1/2$. Setting

$$L = \log x, \quad L_2 = \log\log x = \log L,$$

we then get (for sufficiently large x)

$$\begin{aligned} f(x) &= \log(L + L_2) = \log(L(1 + L_2/L)) \\ &= L_2 + \log(1 + L_2/L) \\ &= L_2 + \frac{L_2}{L} + O\left(\frac{L_2^2}{L^2}\right). \end{aligned}$$

Taking logarithms. Another sometimes very useful technique in asymptotic analysis is to take logarithms in order to transform products to sums and exponentials to products.

Example 2.15. Consider the function

$$f(x) = (\log x + \log\log x)^{1/\sqrt{\log\log x}}.$$

This is a rather fierce looking function, and its behavior as $x \to \infty$ (for example, the question whether it is bounded) is anything but obvious.

Taking logarithms, we can answer such questions. We have

$$\log f(x) = \frac{\log(\log x + \log\log x)}{\sqrt{\log\log x}},$$

and we recognize the numerator as the expression estimated in the above example. Using the notation and result of this example, we get

$$\begin{aligned}\log f(x) &= \frac{1}{\sqrt{L_2}}\left(L_2 + \frac{L_2}{L} + O\left(\frac{L_2^2}{L^2}\right)\right)\\ &= \sqrt{L_2} + \frac{\sqrt{L_2}}{L} + O\left(\frac{L_2^{3/2}}{L^2}\right).\end{aligned}$$

To get back to $f(x)$, we exponentiate, using the estimate $e^z = 1 + O(|z|)$, valid for $|z| \leq 1$, say. Thus,

$$f(x) = \exp\left\{\sqrt{\log\log x} + \frac{\sqrt{\log\log x}}{\log x}\right\}\left(1 + O\left(\frac{(\log\log x)^{3/2}}{(\log x)^2}\right)\right).$$

In particular, we now see that $f(x)$ tends to infinity as $x \to \infty$.

Swapping main and error terms in convergent series and integrals. A common problem in asymptotic analysis is that of estimating partial sums $S(x) = \sum_{n\leq x} a_n$ of an infinite series $\sum_{n=1}^{\infty} a_n$. While the sums $S(x)$ can rarely be evaluated in closed form, it is usually easy to get estimates for the summands of the form $a_n = O(\phi(n))$. Applying such an estimate directly to the summands in $S(x)$ would lead to an error term of size $O(\sum_{n\leq x} |\phi(n)|)$, which is at best $O(1)$ (unless $\phi(n) = 0$ for all n). However, if the series $\sum_{n=1}^{\infty} |\phi(n)|$ (and hence also $\sum_{n=1}^{\infty}$) converges, we can use the following trick to obtain an estimate for $S(x)$ with error term tending to zero as $x \to \infty$. Namely, we extend the range of summation in $S(x) = \sum_{n\leq x} a_n$ to infinity and write $S(x) = S - R(x)$, where $S = \sum_{n=1}^{\infty} a_n$ and $R(x) = \sum_{n>x} a_n$. Applying now the estimate $a_n = O(\phi(n))$ to the tails $R(x)$ of the series then leads to an estimate with error term $O(\sum_{n>x} |\phi(n)|)$. The convergence of the series $\sum_{n=1}^{\infty} |\phi(n)|$ implies that this error term tends to zero, and usually it is easy to obtain more precise estimates for this error term.

Example 2.16. Consider the sum

$$S(x) = \sum_{\leq x} \left(\frac{1}{n} - \log \left(1 + \frac{1}{n} \right) \right) = \sum_{n \leq x} a_n.$$

The terms in this series satisfy $a_n = O(1/n^2)$ for all n, since $x - x^2/2 \leq \log(1 + x) \leq x$ for $0 \leq x \leq 1$ (which can be seen, for example, from the fact that $\log(1 + x) = x - x^2/2 + x^3/3 - \dots$ is an alternating series with decreasing terms). Substituting this estimate directly into the terms in $S(x)$ would only give the estimate

$$S(x) = O\left(\sum_{n \leq x} \frac{1}{n^2} \right) = O(1).$$

However, the trick of extending the summation to infinity leads to an estimate with error term $O(1/x)$,

$$S(x) = S + O\left(\sum_{n>x} \frac{1}{n^2} \right) = S + O\left(\frac{1}{x} \right),$$

where $S = \sum_{n=1}^{\infty} (1/n - \log(1 + 1/n))$ is some (finite) constant.

Note that the method does not give a value for this constant. This is an intrinsic limitation of the method, but in most cases the series simply do not have an evaluation in "closed form" and trying to find such a evaluation would be futile. One can, of course, estimate this constant numerically by computing the partial sums of the series.

Extending the range of an O-estimate. According to our convention, an asymptotic estimate for a function of x without an explicitly given range is understood to hold for $x \geq x_0$ for a suitable x_0. This is convenient as many estimates (e.g., $\log \log x = O(\sqrt{\log x})$), do not hold, or do not make sense, for small values of x, and the convention allows one to just ignore those issues. However, there are applications in which it is desirable to have an estimate involving a simple explicit range for x, such as $x \geq 1$, instead of an unspecified range like $x \geq x_0$ with a “sufficiently large” x_0. This can often be accomplished in two steps as follows: First establish the desired estimate for $x \geq x_0$, with a certain x_0. Then use direct (and usually trivial) arguments to show that the estimate also holds for $1 \leq x \leq x_0$.

Example 2.17. One form of the Prime Number Theorem states that

$$\pi(x) = \mathrm{Li}(x) + O\left(\frac{x}{(\log x)^2}\right).$$

Suppose we have established this estimate for $x \geq x_0$, with a suitable (and possibly quite large) constant x_0. To show that the same estimate in fact holds for $x \geq 2$, we argue as follows: Assume $x_0 \geq 2$ (otherwise there is nothing to prove) and consider the range $2 \leq x \leq x_0$. In this range the functions $\pi(x)$ and $\mathrm{Li}(x)$ are bounded from above, so we have

$$|\pi(x) - \mathrm{Li}(x)| \leq c_1 \quad (2 \leq x \leq x_0)$$

with some constant c_1 depending on x_0. (For example, since both $\pi(x)$ and $\mathrm{Li}(x)$ are nondecreasing functions, we could take $c_1 = \pi(x_0) + \mathrm{Li}(x_0)$.) On the other hand, in the same range the function in the error term is bounded from below by a positive constant, i.e., we have

$$\frac{x}{(\log x)^2} \geq c_2 \quad (2 \leq x \leq x_0)$$

with some positive constant c_2 (e.g., $c_2 = 2(\log 2)^{-2}$). Hence we have

$$|\pi(x) - \mathrm{Li}(x)| \leq c\frac{x}{\log x)^2} \quad (2 \leq x \leq x_0)$$

with $c = c_1 c_2^{-1}$, which proves the desired estimate for the range $2 \leq x \leq x_0$.

2.3 Asymptotic series

In the next chapter, we will show that the logarithmic integral $\mathrm{Li}(x) = \int_2^x (\log t)^{-1} dt$ satisfies

$$\mathrm{Li}(x) = \frac{x}{\log x}\left(\sum_{k=0}^{n-1} \frac{k!}{(\log x)^k} + O_n\left(\frac{1}{(\log x)^n}\right)\right)$$

for any *fixed* positive integer n (and a range of the form $x \geq x_0(n)$). This estimate is reminiscent of the approximation of an analytic function by the partial sums of its power series. Indeed, setting $z = (\log x)^{-1}$ and $a_k = k!$, the expression in parentheses in the above estimate for $\mathrm{Li}(x)$ takes the form

$$\sum_{k=0}^{n-1} a_k z^k + O_n(|z|^n).$$

The latter expression is of the form of the usual n-term Taylor approximation to an analytic function with power series $\sum_{k=0}^{\infty} a_k z^k$. However, there is one significant difference: With the above choice of coefficients a_k, the series $\sum_{k=0}^{\infty} a_k z^k$ diverges at all $z \neq 0$, and thus does not represent an analytic function.

This is an example of a very common phenomenon in asymptotic analysis that gives rise to the concept of an "asymptotic series". Roughly speaking, an asymptotic series for a given function is an infinite series that has the same approximation properties as the Taylor series expansion of an analytic function, but which does not (necessarily) converge. More formally, we define an asymptotic series as follows:

Definition. Let $f(x)$ be a function defined for all sufficiently large x and let $\phi_0(x), \phi_1(x), \ldots$ be a sequence of functions satisfying

$$\phi_{n+1}(x) = o(\phi_n(x)) \quad (x \to \infty)$$

for each n. A (formal) series of the form

$$\sum_{k=0}^{\infty} a_k \phi_k(x)$$

is called an **asymptotic series** for a function $f(x)$, as $x \to \infty$, if, for each n, the truncation of this series at n approximates $f(x)$ to within $o(\phi_n(x))$, i.e., if

$$f(x) = \sum_{k=0}^{n} a_k \phi_k(x) + o(\phi_n(x)) \quad (x \to \infty).$$

If this holds, we write[1]

$$f(x) \sim \sum_{k=0}^{\infty} a_k \phi_k(x) \quad (x \to \infty).$$

Asymptotic series with respect to other limiting processes, such as $x \to 0$, are defined analogously. Moreover, these definitions can be generalized to functions of complex variables in an obvious manner.

The above definition is sufficiently general to apply to nearly all situations one encounters in practice. For example, the above-mentioned series occurring in the expansion of $\mathrm{Li}(x)$ is an asymptotic series in this sense with basis functions $\phi_k(x) = (\log x)^{-k}$.

Any power series $\sum_{k=0}^{\infty} a_k z^k$ that has positive radius of convergence is an asymptotic series, as $z \to 0$, for the function it represents within the disk of convergence, with $\phi_k(z) = z^k$ as the basis functions.

While asymptotic series share many properties with ordinary power series, there are also some notable differences. The most glaring difference is, of course, the fact that, in general, an asymptotic series does not converge; it "represents" the function only in an asymptotic sense. However, there are other differences as well. In particular, a function is not uniquely determined by its asymptotic series expansion.

Example 2.18. If $f(x)$ has the asymptotic series expansion

$$f(x) \sim \sum_{k=0}^{\infty} a_k x^{-k} \quad (x \to \infty),$$

then any function $g(x)$ satisfying $g(x) = f(x) + O_n(x^{-n})$ for every fixed positive integer n (e.g., $g(x) = f(x) + e^{-x}$) has the same asymptotic series expansion. This follows immediately from the definition of an asymptotic series.

[1]The notation "$\sim$" here is the same as that used for asymptotic equivalence (as in "$f(x) \sim g(x)$"), though it has a very different meaning. The usage of the symbol "$\sim$" in two different ways is somewhat unfortunate, but is now rather standard, and alternative notations (such as using the symbol "$\approx$" instead of "$\sim$" in the context of asymptotic series) have their own drawbacks. In practice, the intended meaning is usually clear from the context. Since most of the time we will be dealing with the symbol "$\sim$" in the asymptotic equivalence sense, we make the convention that, *unless otherwise specified, the symbol "$\sim$" is to be interpreted in the sense of an asymptotic equivalence.*