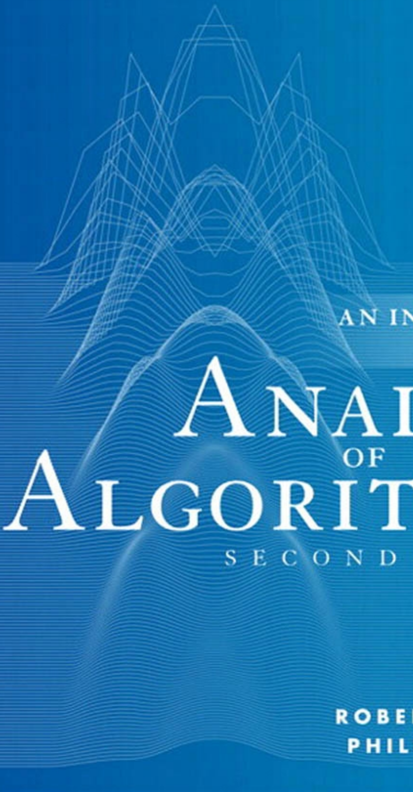

AN INTRODUCTION TO THE

# ANALYSIS OF ALGORITHMS

SECOND EDITION

ROBERT SEDGEWICK
PHILIPPE FLAJOLET

# AN INTRODUCTION TO THE ANALYSIS OF ALGORITHMS

*Second Edition*

This process leads immediately to concise, accurate, and precise approximations to solutions for linear recurrences. For example, consider the recurrence

$$a_n = 2a_{n-1} + a_{n-2} - 2a_{n-3}, \qquad n > 2; \quad a_0 = 0, a_1 = a_2 = 1.$$

We found in §3.3 that the generating function for the solution is

$$a(z) = \frac{f(z)}{g(z)} = \frac{z}{(1+z)(1-2z)}.$$

Here $\beta = 2$, $\nu = 1$, $g'(1/2) = -3$, and $f(1/2) = 1/2$, so Theorem 4.1 tells us that $a_n \sim 2^n/3$, as before.

**Exercise 4.14** Use Theorem 4.1 to find an asymptotic solution to the recurrence

$$a_n = 5a_{n-1} - 8a_{n-2} + 4a_{n-3} \qquad \text{for } n > 2 \text{ with } a_0 = 1,\ a_1 = 2, \text{ and } a_2 = 4.$$

Solve the same recurrence with the initial conditions on $a_0$ and $a_1$ changed to $a_0 = 1$ and $a_1 = 2$.

**Exercise 4.15** Use Theorem 4.1 to find an asymptotic solution to the recurrence

$$a_n = 2a_{n-2} - a_{n-4} \qquad \text{for } n > 4 \text{ with } a_0 = a_1 = 0 \text{ and } a_2 = a_3 = 1.$$

**Exercise 4.16** Use Theorem 4.1 to find an asymptotic solution to the recurrence

$$a_n = 3a_{n-1} - 3a_{n-2} + a_{n-3} \qquad \text{for } n > 2 \text{ with } a_0 = a_1 = 0 \text{ and } a_2 = 1.$$

**Exercise 4.17** [Miles, cf. Knuth] Show that the polynomial $z^t - z^{t-1} - \ldots - z - 1$ has $t$ distinct roots and that exactly one of the roots has modulus greater than 1, for all $t > 1$.

**Exercise 4.18** Give an approximate solution for the "$t$th-order Fibonacci" recurrence

$$F_N^{[t]} = F_{N-1}^{[t]} + F_{N-2}^{[t]} + \ldots + F_{N-t}^{[t]} \qquad \text{for } N \ge t$$

with $F_0^{[t]} = F_1^{[t]} = \ldots = F_{t-2}^{[t]} = 0$ and $F_{t-1}^{[t]} = 1$.

**Exercise 4.19** [Schur] Show that the number of ways to change an $N$-denomination bill using coin denominations $d_1, d_2, \ldots, d_t$ with $d_1 = 1$ is asymptotic to

$$\frac{N^{t-1}}{d_1 d_2 \ldots d_t (t-1)!}.$$

(See Exercise 3.55.)

**4.2 Asymptotic Expansions.** As mentioned earlier, we prefer the equation $f(N) = c_0 g_0(N) + O(g_1(N))$ with $g_1(N) = o(g_0(N))$ to the equation $f(N) = O(g_0(N))$ because it provides the constant $c_0$, and therefore allows us to provide specific estimates for $f(N)$ that improve in accuracy as $N$ gets large. If $g_0(N)$ and $g_1(N)$ are relatively close, we might wish to find a constant associated with $g_1$ and thus derive a better approximation: if $g_2(N) = o(g_1(N))$, we write $f(N) = c_0 g_0(N) + c_1 g_1(N) + O(g_2(N))$.

The concept of an *asymptotic expansion*, developed by Poincaré (cf. [6]), generalizes this notion.

**Definition** Given a sequence of functions $\{g_k(N)\}_{k\geq 0}$ having the property that $g_{k+1}(N) = o(g_k(N))$ for $k \geq 0$, the formula

$$f(N) \sim c_0 g_0(N) + c_1 g_1(N) + c_2 g_2(N) + \ldots$$

is called an *asymptotic series* for $f$, or an *asymptotic expansion* of $f$. The asymptotic series represents the collection of equations

$$\begin{aligned}
f(N) &= O(g_0(N)) \\
f(N) &= c_0 g_0(N) + O(g_1(N)) \\
f(N) &= c_0 g_0(N) + c_1 g_1(N) + O(g_2(N)) \\
f(N) &= c_0 g_0(N) + c_1 g_1(N) + c_2 g_2(N) + O(g_3(N)) \\
&\;\;\vdots
\end{aligned}$$

and the $g_k(N)$ are referred to as an *asymptotic scale*.

Each additional term that we take from the asymptotic series gives a more accurate asymptotic estimate. Full asymptotic series are available for many functions commonly encountered in the analysis of algorithms, and we primarily consider methods that could be extended, in principle, to provide asymptotic expansions describing quantities of interest. We can use the $\sim$-notation to simply drop information on error terms or we can use the $O$-notation or the $o$-notation to provide more specific information.

For example, the expression $2N\ln N + (2\gamma - 2)N + O(\log N)$ allows us to make far more accurate estimates of the average number of comparison required for quicksort than the expression $2N\ln N + O(N)$ for practical values of $N$, and adding the $O(\log N)$ and $O(1)$ terms provides even more accurate estimates, as shown in Table 4.1.

Asymptotic expansions extend the definition of the $\sim -$ notation that we considered at the beginning of §4.1. The earlier use normally would involve just one term on the right-hand side, whereas the current definition calls for a series of (decreasing) terms.

Indeed, we primarily deal with *finite* expansions, not (infinite) asymptotic series, and use, for example, the notation

$$f(N) \sim c_0 g_0(N) + c_1 g_1(N) + c_2 g_2(N)$$

to refer to a finite expansion with the implicit error term $o(g_2(N))$. Most often, we use finite asymptotic expansions of the form

$$f(N) = c_0 g_0(N) + c_1 g_1(N) + c_2 g_2(N) + O(g_3(N)),$$

obtained by simply truncating the asymptotic series. In practice, we generally use only a few terms (perhaps three or four) for an approximation, since the usual situation is to have an asymptotic scale that makes later terms extremely small in comparison to early terms for large $N$. For the quicksort example shown in Table 4.1, the "more accurate" formula $2N\ln N + (2\gamma - 2)N + 2\ln N + 2\gamma + 1$ gives an absolute error less than .1 already for $N = 10$.

**Exercise 4.20** Extend Table 4.1 to cover the cases $N = 10^5$ and $10^6$.

The full generality of the Poincaré approach allows asymptotic expansions to be expressed in terms of *any* infinite series of functions that decrease (in an $o$-notation sense). However, we are most often interested in a very

| $N$ | $2(N+1)(H_{N+1}-1)$ | $2N\ln N$ | $+(2\gamma-2)N$ | $+2(\ln N+\gamma)+1$ |
|---|---|---|---|---|
| 10 | 44.43 | 46.05 | 37.59 | 44.35 |
| 100 | 847.85 | 921.03 | 836.47 | 847.84 |
| 1000 | 12,985.91 | 13,815.51 | 12,969.94 | 12,985.91 |
| 10,000 | 175,771.70 | 184,206.81 | 175,751.12 | 175,771.70 |

**Table 4.1** Asymptotic estimates for quicksort comparison counts

restricted set of functions: indeed, we are very often able to express approximations in terms of decreasing powers of $N$ when approximating functions as $N$ increases. Other functions occasionally are needed, but we normally will be content with an asymptotic scale consisting of terms of decreasing series of products of powers of $N$, $\log N$, iterated logarithms such as $\log\log N$, and exponentials.

When developing an asymptotic estimate, it is not necessarily clear *a prioiri* how many terms should be carried in the expansion to get the desired accuracy in the result. For example, frequently we need to subtract or divide quantities for which we only have asymptotic estimates, so cancellations might occur that necessitate carrying more terms. Typically, we carry three or four terms in an expansion, perhaps redoing the derivation to streamline it or to add more terms once the nature of the result is known.

***Taylor expansions.*** Taylor series are the source of many asymptotic expansions: each (infinite) Taylor expansion gives rise to an asymptotic series as $x \to 0$. Table 4.2 gives asymptotic expansions for some of the basic functions, derived from truncating Taylor series. These expansions are classical,

| | |
|---|---|
| exponential | $e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + O(x^4)$ |
| logarithmic | $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} + O(x^4)$ |
| binomial | $(1+x)^k = 1 + kx + \binom{k}{2}x^2 + \binom{k}{3}x^3 + O(x^4)$ |
| geometric | $\frac{1}{1-x} = 1 + x + x^2 + x^3 + O(x^4)$ |
| trigonometric | $\sin(x) = x - \frac{x^3}{6} + \frac{x^5}{120} + O(x^7)$ |
| | $\cos(x) = 1 - \frac{x^2}{2} + \frac{x^4}{24} + O(x^6)$ |

**Table 4.2** Asymptotic expansions derived from Taylor series ($x \to 0$)

and follow immediately from the Taylor theorem. In the sections that follow, we describe methods of manipulating asymptotic series using these expansions. Other similar expansions follow immediately from the generating functions given in the previous chapter. The first four expansions serve as the basis for many of the asymptotic calculations that we do (actually, the first three suffice, since the geometric expansion is a special case of the binomial expansion).

For a typical example of the use of Table 4.2, consider the problem of finding an asymptotic expansion for $\ln(N-2)$ as $N \to \infty$. We do so by pulling out the leading term, writing

$$\ln(N-2) = \ln N + \ln\Bigl(1 - \frac{2}{N}\Bigr) = \ln N - \frac{2}{N} + O\Bigl(\frac{1}{N^2}\Bigr).$$

That is, in order to use Table 4.2, we find a substitution ($x = -2/N$) with $x \to 0$.

Or, we can use more terms of the Taylor expansion to get a more general asymptotic result. For example, the expansion

$$\ln(N + \sqrt{N}) = \ln N + \frac{1}{\sqrt{N}} - \frac{1}{2N} + O\Bigl(\frac{1}{N^{3/2}}\Bigr)$$

follows from factoring out $\ln N$, then taking $x = 1/\sqrt{N}$ in the Taylor expansion for $\ln(1+x)$. This kind of manipulation is typical, and we will see many examples of it later.

**Exercise 4.21** Expand $\ln(1 - x + x^2)$ as $x \to 0$, to within $O(x^4)$.

**Exercise 4.22** Give an asymptotic expansion for $\ln(N^\alpha + N^\beta)$, where $\alpha$ and $\beta$ are positive constants with $\alpha > \beta$.

**Exercise 4.23** Give an asymptotic expansion for $\dfrac{N}{N-1}\ln\dfrac{N}{N-1}$.

**Exercise 4.24** Estimate the value of $e^{0.1} + \cos(.1) - \ln(.9)$ to within $10^{-4}$, without using a calculator.

**Exercise 4.25** Show that

$$\frac{1}{9801} = 0.000102030405060708091011\cdots47484950\cdots$$

to within $10^{-100}$. How many more digits can you predict? Generalize.

**Nonconvergent asymptotic series.** Any convergent series leads to a full asymptotic approximation, but it is very important to note that the converse is *not* true—an asymptotic series may well be divergent. For example, we might have a function

$$f(N) \sim \sum_{k\geq 0} \frac{k!}{N^k}$$

implying (for example) that

$$f(N) = 1 + \frac{1}{N} + \frac{2}{N^2} + \frac{6}{N^3} + O\Big(\frac{1}{N^4}\Big)$$

even though the infinite sum does not converge. Why is this allowed? If we take any fixed number of terms from the expansion, then the equality implied from the definition is meaningful, as $N \to \infty$. That is, we have an infinite collection of better and better approximations, but the point at which they start giving useful information gets larger and larger.

**Stirling's formula.** The most celebrated example of a divergent asymptotic series is *Stirling's formula*, which begins as follows:

$$N! = \sqrt{2\pi N}\Big(\frac{N}{e}\Big)^N\Big(1 + \frac{1}{12N} + \frac{1}{288N^2} + O\Big(\frac{1}{N^3}\Big)\Big).$$

| $N$ | $N!$ | $\sqrt{2\pi N}\Big(\frac{N}{e}\Big)^N\Big(1+\frac{1}{12N}+\frac{1}{288N^2}\Big)$ | absolute error | relative error |
|---|---|---|---|---|
| 1 | 1 | 1.002183625 | .0022 | $10^{-2}$ |
| 2 | 2 | 2.000628669 | .0006 | $10^{-3}$ |
| 3 | 6 | 6.000578155 | .0006 | $10^{-4}$ |
| 4 | 24 | 24.00098829 | .001 | $10^{-4}$ |
| 5 | 120 | 120.0025457 | .002 | $10^{-4}$ |
| 6 | 720 | 720.0088701 | .009 | $10^{-4}$ |
| 7 | 5040 | 5040.039185 | .039 | $10^{-5}$ |
| 8 | 40,320 | 40320.21031 | .210 | $10^{-5}$ |
| 9 | 362,880 | 362881.3307 | 1.33 | $10^{-5}$ |
| 10 | 3,628,800 | 3628809.711 | 9.71 | $10^{-5}$ |

**Table 4.3** Accuracy of Stirling's formula for $N!$