1.1 Скорость роста функций

Сравнивая два алгоритма сортировки, мы установили, что время работы одного (сортировка вставками) примерно пропорционально n^2, а другого (сортировка слиянием) — $n \log n$. Каковы бы ни были коэффициенты пропорциональности, для достаточно больших n первый алгоритм работает быстрее. Анализируя алгоритм, можно стараться найти точное количество выполняемых им действий. Но в большинстве случаев игра не стоит свеч, и достаточно оценить асимптотику роста времени работы алгоритма при стремлении размера входа к бесконечности (asymptotic efficiency). Если у одного алгоритма асимптотика роста меньше, чем у другого, то в большинстве случаев он будет эффективнее для всех входов, кроме совсем коротких. (Хотя бывают и исключения.)

1.1.1 Асимптотические обозначения

Хотя во многих случаях эти обозначения используются неформально, полезно начать с точных определений.

Θ-обозначение

Ранее мы говорили, что время $T(n)$ работы алгоритма сортировки вставками на входах длины n есть $\Theta(n^2)$. Точный смысл этого утверждения такой: найдутся такие константы $c_1, c_2 > 0$ и такое число n_0, что $c_1 n^2 \leqslant T(n) \leqslant c_2 n^2$ при всех $n \geqslant n_0$. Вообще, если $g(n)$ — некоторая функция, то запись $f(n) = \Theta(g(n))$ означает, что найдутся такие $c_1, c_2 > 0$ и такое n_0, что $0 \leqslant c_1 g(n) \leqslant f(n) \leqslant c_2 g(n)$ для всех $n \geqslant n_0$ (см. рис. (а)). (Запись $f(n) = \Theta(g(n))$ читается так: «эф от эн есть тэта от же от эн».)

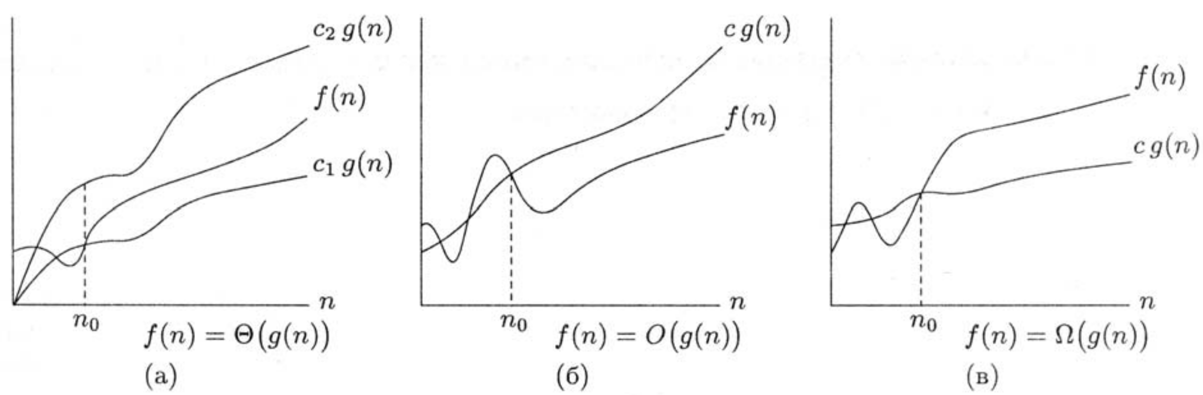

Рис. 1.1: Иллюстрация к определениям $f(n) = \Theta(g(n))$, $f(n) = O(g(n))$ и $f(n) = \Omega(g(n))$.

Разумеется, это обозначение следует употреблять с осторожностью: установив, что $f_1(n) = \Theta(g(n))$ и $f_2(n) = \Theta(g(n))$, не следует заключать, что $f_1(n) = f_2(n)$!

Определение $\Theta(g(n))$ предполагает, что функции $f(n)$ и $g(n)$ **асимптотически неотрицательны** (asymptotically nonnegative), т. е. неотрицательны для достаточно больших значений n. Заметим, что если функции f и g строго положительны, то можно исключить n_0 из определения (изменив константы c_1 и c_2 так, чтобы для малых n неравенство также выполнялось).

Если $f(n) = \Theta(g(n))$, то говорят, что $g(n)$ является **асимптотически точной оценкой** (asymptotically tight bound) для $f(n)$. На самом деле это отношение симметрично: если $f(n) = \Theta(g(n))$, то $g(n) = \Theta(f(n))$.

Например, проверим, что $(1/2)n^2 - 3n = \Theta(n^2)$. Согласно определению надо указать положительные константы c_1, c_2 и число n_0 так, чтобы неравенства

$$c_1 n^2 \leqslant \frac{1}{2}n^2 - 3n \leqslant c_2 n^2$$

выполнялись для всех $n \geqslant n_0$. Разделим выражение на n^2:

$$c_1 \leqslant \frac{1}{2} - \frac{3}{n} \leqslant c_2$$

Видно, что для выполнения второго неравенства достаточно положить $c_2 = 1/2$. Первое будет выполнено, если (например) $n_0 = 7$ и $c_1 = 1/14$.

Другой пример использования формального определения: покажем, что $6n^3 \neq \Theta(n^2)$. В самом деле, пусть найдутся такие c_2 и n_0, что $6n^3 \leqslant c_2 n^2$ для всех $n \geqslant n_0$. Но тогда $n \leqslant c_2/6$ для всех $n \geqslant n_0$ — что явно не так.

Отыскивая асимптотически точную оценку для суммы, мы можем отбрасывать члены меньшего порядка, которые при больших n становятся малыми по сравнению с основным слагаемым. Заметим также, что коэффициент при старшем члене роли не играет (он может повлиять только на выбор констант c_1 и c_2). Например, рассмотрим квадратичную функцию $f(n) = an^2 + bn + c$, где a, b, c —некоторые константы и $a > 0$. Отбрасывая члены младших порядков и коэффициент при старшем члене, находим, что $f(n) = \Theta(n^2)$. Чтобы убедиться в этом формально, можно положить $c_1 = a/4$, $c_2 = 7a/4$ и $n_0 = 2 \cdot \max(|b|/a, \sqrt{|c|/a})$ (проверьте, что требования действительно выполнены). Вообще, для любого полинома $p(n)$ степени d с положительным старшим коэффициентом имеем $p(n) = \Theta(n^d)$.

Упомянем важный частный случай использования Θ-обозначений: $\Theta(1)$ обозначает ограниченную функцию, отделённую от нуля некоторый положительной константой при достаточно больших значениях аргумента. (Из контекста обычно ясно, что именно считается аргументом функции.)

O- и Ω-обозначения

Запись $f(n) = \Theta(g(n))$ включает в себя две оценки: верхнюю и нижнюю. Их можно разделить. Говорят, что $f(n) = O(g(n))$, если найдётся такая константа $c > 0$ и такое число n_0, что $0 \leqslant f(n) \leqslant cg(n)$ для всех $n \geqslant n_0$ (см. рис. (б)). Говорят, что $f(n) = \Omega(g(n))$, если найдется такая константа $c > 0$ и такое число n_0, что $0 \leqslant cg(n) \leqslant f(n)$ для всех $n \geqslant n_0$ (см. рис. (в)). Эти записи читаются так: «эф от эн есть о большое от же от эн», «эф от эн есть омега большая от же от эн».

По-прежнему мы предполагаем, что функции f и g неотрицательны для достаточно больших значений аргумента. Легко видеть, что выполнены следующие свойства:

Теорема 2.1. *Для любых двух функций $f(n)$ и $g(n)$ свойство $f(n) = \Theta(g(n))$ выполнено тогда и только тогда, когда $f(n) = O(g(n))$ и $f(n) = \Omega(g(n))$.*

Для любых двух функций $f(n)$ и $g(n)$ свойства $f(n) = O(g(n))$ и $g(n) = \Omega(f(n))$ равносильны.

Как мы видели, $an^2+bn+c = \Theta(n^2)$ (при положительных a). Поэтому $an^2+bn+c = O(n^2)$. Другой пример: при $a > 0$ можно написать $an+b = O(n^2)$ (положим $c = a+|b|$ и $n_0 = 1$). Заметим, что в этом случае $an + b \neq \Omega(n^2)$ и $an + b \neq \Theta(n^2)$.

Асимптотические обозначения (Θ, O и Ω) часто употребляются внутри формул. Например, ранее мы получили рекуррентное соотношение

$$T(n) = 2T(n/2) + \Theta(n)$$

для времени работы сортировки слиянием. Здесь $\Theta(n)$ обозначает некоторую функцию, про которую нам важно знать лишь, что она не меньше $c_1 n$ и не больше $c_2 n$ для некоторых положительных c_1 и c_2 и для всех достаточно больших n.

Часто асимптотические обозначения употребляются не вполне формально, хотя их подразумеваемый смысл обычно ясен из контекста. Например, мы можем написать выражение

$$\sum_{i=1}^{n} O(i),$$

имея в виду сумму $h(1)+h(2)+\cdots+h(n)$, где $h(i)$ — некоторая функция, для которой $h(i) = O(i)$. Легко видеть, что сама эта сумма как функция от n есть $O(n^2)$.

Типичный пример неформального использования асимптотических обозначений — цепочка равенств наподобие $2n^2 + 3n + 1 = 2n^2 + \Theta(n) = \Theta(n^2)$. Второе из этих равенств ($2n^2+\Theta(n) = \Theta(n^2)$) понимается при этом так: какова бы ни была функция $h(n) = \Theta(n)$ в левой части, сумма $2n^2 + h(n)$ есть $\Theta(n^2)$.

o- и ω-обозначения

Запись $f(n) = O(g(n))$ означает, что с ростом n отношение $f(n)/g(n)$ остаётся ограниченным. Если к тому же

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0, \tag{1.1}$$

то мы пишем $f(n) = o(g(n))$ (читается «эф от эн есть о малое от же от эн»). Формально говоря, $f(n) = o(g(n))$, если для всякого положительного $\epsilon > 0$ найдётся такое n_0, что $0 \leqslant f(n) \leqslant \epsilon g(n)$ при всех $n \geqslant n_0$. (Тем самым запись $f(n) = o(g(n))$ предполагает, что $f(n)$ и $g(n)$ неотрицательны для достаточно больших n.)

Пример: $2n = o(n^2)$, но $2n^2 \neq o(n^2)$.

Аналогичным образом вводится ω-обозначение: говорят, что $f(n)$ есть $\omega(g(n))$ («эф от эн есть омега малая от же от эн»), если для всякого положительного c найдется такое n_0, что $0 \leqslant cg(n) \leqslant f(n)$ при всех $n \geqslant n_0$. Очевидно, $f(n) = \omega(g(n))$ равносильно $g(n) = o(f(n))$. Пример: $n^2/2 = \omega(n)$, но $n^2/2 \neq \omega(n^2)$.

Сравнение функций

Введённые нами определения обладают некоторыми свойствами транзитивности, рефлексивности и симметричности.

Транзитивность:
$f(n) = \Theta(g(n))$ и $g(n) = \Theta(h(n))$ влечет $f(n) = \Theta(h(n))$,
$f(n) = O(g(n))$ и $g(n) = O(h(n))$ влечет $f(n) = O(h(n))$,
$f(n) = \Omega(g(n))$ и $g(n) = \Omega(h(n))$ влечет $f(n) = \Omega(h(n))$,
$f(n) = o(g(n))$ и $g(n) = o(n))$ влечет $f(n) = o(h(n))$,
$f(n) = \omega(g(n))$ и $g(n) = \omega(h(n))$ влечет $f(n) = \omega(h(n))$.

Рефлексивность:
$f(n) = \Theta(f(n)), \qquad f(n) = O(f(n)), \qquad f(n) = \Omega(f(n))$.

Симметричность:
$f(n) = \Theta(g(n))$ если и только если $g(n) = \Theta(f(n))$.

Обращение:
$f(n) = O(g(n))$ если и только если $g(n) = \Omega(f(n))$,
$f(n) = o(g(n))$ если и только если $g(n) = \omega(f(n))$.

Можно провести такую параллель: отношения между функциями f и g подобны отношениям между числами a и b:

$$\begin{aligned}
f(n) &= O(g(n)) && \approx \quad a \leqslant b,\\
f(n) &= \Omega(g(n)) && \approx \quad a \geqslant b,\\
f(n) &= \Theta(g(n)) && \approx \quad a = b,\\
f(n) &= O(g(n)) && \approx \quad a < b,\\
f(n) &= \omega(g(n)) && \approx \quad a > b.
\end{aligned}$$

Параллель эта, впрочем, весьма условна: свойства числовых неравенств не переносятся на функции. Например, для любых двух чисел a и b всегда или $a \leqslant b$, или $a \geqslant b$, однако нельзя утверждать, что для любых двух (положительных) функций $f(n)$ и $g(n)$ или $f(n) = O(g(n))$, или $f(n) = \Omega(g(n))$. В самом деле, можно проверить, что ни одно из этих двух соотношений не выполнено для $f(n) = n$ и $g(n) = n^{1+\sin n}$ (показатель степени в выражении для $g(n)$ меняется в интервале от 0 до 2). Заметим ещё, что для чисел $a \leqslant b$ влечёт $a < b$ или $a = b$, в то время как для функций $f(n) = O(g(n))$ не влечёт $f(n) = o(g(n))$ или $f(n) = \Theta(g(n))$.

1.1.2 Стандартные функции и обозначения

Монотонность

Говорят, что функция $f(n)$ **монотонно возрастает** (is monotonically increasing), если $f(m) \leqslant f(n)$ при $m \leqslant n$. Говорят, что функция $f(n)$ **монотонно убывает** (is monotonically decreasing), если $f(m) \geqslant f(n)$ при $m \leqslant n$. Говорят, что функция $f(n)$ **строго возрастает** (is strictly increasing), если $f(m) < f(n)$ при $m < n$. Говорят, что функция $f(n)$ **строго убывает** (is strictly decreasing), если $f(m) > f(n)$ при $m < n$.

Целые приближения снизу и сверху

Для любого вещественного числа x через $\lfloor x \rfloor$ (the floor of x) мы обозначаем его целую часть, т. е. наибольшее целое число, не превосходящее x. Симметричным образом, $\lceil x \rceil$ (the ceiling of x) обозначает наименьшее целое число, не меньшее x. Очевидно,

$$x - 1 < x - 1 < \lfloor x \rfloor \leqslant x \leqslant \lceil x \rceil < x + 1$$

для любого x. Кроме того,

$$\lceil n/2 \rceil + \lfloor n/2 \rfloor = n$$

для любого целого n. Наконец, для любого x и для любых целых положительных a и b имеем

$$\lceil \lceil x/a \rceil /b \rceil = \lceil x/ab \rceil$$

и

$$\lfloor \lfloor x/a \rfloor /b \rfloor = \lfloor x/ab \rfloor$$

(чтобы убедиться в этом, полезно заметить, что для любого z и для целого n свойства $n \leqslant z$ и $n \leqslant \lfloor z \rfloor$ равносильны).

Функции $x \mapsto \lfloor x \rfloor$ и $x \mapsto \lceil x \rceil$ монотонно возрастают.

Многочлены

Многочленом (полиномом) степени d **от переменной** n (polynomial in n of degree d) называют функцию

$$p(n) = \sum_{i=0}^{d} a_i n^i$$

(d — неотрицательное целое число). Числа $a_0, a_1, \ldots, a_d$ называют **коэффициентами** (coefficients) многочлена. Мы считаем, что **старший коэффициент** a_d не равен нулю (если это не так, уменьшим d — это можно сделать, если только многочлен не равен нулю тождественно).

Для больших значений n знак многочлена $p(n)$ определяется старшим коэффициентом (остальные члены малы по сравнению с ним), так что при $a_d > 0$ многочлен $p(n)$ асимптотически положителен (положителен при больших n) и можно написать $p(n) = \Theta(n^d)$.

При $a \geqslant 0$ функция $n \mapsto n^a$ монотонно возрастает, при $a \leqslant 0$ — монотонно убывает. Говорят, что функция $f(n)$ **полиномиально ограничена** (is polynomially bounded), если $f(n) = n^{O(1)}$, или, другими словами, если $f(n) = O(n^k)$ для некоторой константы k.

Экспоненты

Для любых вещественных m, n и $a \neq 0$ имеем

$$\begin{aligned} a^0 = 1, (a^m)^n &= a^{mn}, \\ a^1 = a, (a^m)^n &= (a^n)^m, \\ a^{-1} = 1/a, a^m a^n &= a^{m+n}. \end{aligned}$$

При $a \geqslant 1$ функция $n \mapsto a^n$ монотонно возрастает.

Мы будем иногда условно полагать $0^\circ = 1$.

Функция $n \mapsto a^n$ называется **показательной функцией**, или **экспонентой** (exponential). При $a > 1$ показательная функция растёт быстрее любого полинома: каково бы ни было b,

$$\lim_{n \to \infty} \frac{n^b}{a^n} = 0$$

или, другими словами, $n^b = o(a^n)$.. Если в качестве основания степени взять число $e = 2,71828\ldots$, то экспоненту можно записать в виде ряда

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{x^k}{k!},$$

где $k! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot k$.

Для всех вещественных x выполнено неравенство

$$e^x \geqslant 1 + x,$$

которое обращается в равенство лишь при $x = 0$. При $|x| \leqslant 1$ можно оценить e^x сверху и снизу так: I

$$1 + x \leqslant e^x \leqslant 1 + x + x^2.$$

Можно сказать, что $e^x = 1 + x + \Theta(x^2)$ при $x \to 0$, имея в виду соответствующее истолкование обозначения Θ (в котором $n \to \infty$ заменено на $x \to 0$).

При всех x выполнено равенство $\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x$.

Логарифмы

Мы будем использовать такие обозначения:

$$\begin{aligned}
\log n = \log_2 n \qquad &(\text{двоичный логарифм}),\\
\ln n = \log_e n, \qquad &(\text{натуральный логарифм}),\\
&\log^k n = (\log n)^k,\\
\log\log n = \log(\log n) \qquad &(\text{повторный логарифм}).
\end{aligned}$$

Мы будем считать, что в формулах знак логарифма относится лишь к непосредственно следующему за ним выражению, так что $\log n + k$ есть $\log(n) + k$ (а не $\log(n+k)$. При $b > 1$ функция $n \mapsto \log_b n$ (определённая при положительных n) строго возрастает. Следующие тождества верны при всех $a > 0$, $b > 0$, $c > 0$ и при всех n (если только основания логарифмов не равны 1):

$$\begin{aligned}
a &= b^{\log_b a},\\
\log_b a^n &= n \log_b a,\\
\log_b a &= \frac{1}{\log_a b},\\
\log_c ab &= \log_c a + \log_c b,\\
\log_b a &= \frac{log_c a}{\log_c b},\\
a^{\log_b c} &= c^{\log_b a},\\
\log_b\left(\frac{1}{a}\right) &= -\log_b a.
\end{aligned}$$

При изменении основания логарифма умножается на константу, поэтому в записи типа $O(\log n)$ можно не уточнять, каково основание логарифма. Мы будем чаще всего иметь дело с двоичными логарифмами (они появляются, когда задача делится на две части) и потому оставляем за ними обозначение log.

Для натурального логарифма есть ряд (который сходится при $|x| < 1$):

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - \dots$$

При $x > -1$ справедливы неравенства

$$\frac{x}{1+x} \leqslant \ln(1+x) \leqslant x,$$

которые обращаются в равенства лишь при $x = 0$.

Факториалы

Запись $n!$ (читается «эн факториал», "n factorial") обозначает произведение всех целых чисел от 1 до n. Полагают $0! = 1$, так что $n! = n \cdot (n-1)!$ при всех $n = 1, 2, 3, \dots$.

Сразу же видно, что $n! \leqslant n^n$ (каждый из сомножителей не больше n). Более точная оценка даётся **формулой Стерлинга** (Stirling's approximation), которая гласит, что

$$n! = \sqrt{2\pi n}(\frac{n}{e})^n(1 + \Theta(1/n)).$$

Из формулы Стирлинга следует, что

$$n! = o(n^n),$$
$$n! = \omega(2^n),$$
$$\log(n!) = \Theta(n \log n).$$

Справедлива также следующая оценка:

$$\sqrt{2\pi n}(\frac{n}{e})^n \leqslant n! \leqslant \sqrt{2\pi n}(\frac{n}{e})^n e^{1/12n}.$$

Числа Фибоначчи

Последовательность **чисел Фибоначчи** (Fibonacci numbers) определяется рекуррентным соотношением:

$$F_0 = 0, \quad F_1 = 1, \quad F_i = F_{i-1} + F_{i-1} \text{при} i \geqslant 2.$$

Другими словами, в последовательности Фибоначчи

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55 \dots$$

каждое число равно сумме двух предыдущих. Числа Фибоначчи связаны с так называемым отношением **золотого сечения** (golden ratio) φ и с сопряжённым с ним числом $\hat{\varphi}$:

$$\varphi = \frac{1+\sqrt{5}}{2} = 1,61803\dots,$$
$$\hat{\varphi} = \frac{1-\sqrt{5}}{2} = -0,61803\dots.$$

Именно, имеет место формула

$$F_i = \frac{\varphi^i - \hat{\varphi}^i}{\sqrt{5}},$$

которую можно доказать по индукции. Поскольку $|\hat{\varphi}| < 1$, слагаемое $|\hat{\varphi}^i/\sqrt{5}|$ меньше $1/\sqrt{5} < 1/2$, так что F_i равно числу $\varphi^i/\sqrt{5}$, округлённому до ближайшего целого.

Числа F_i быстро (экспоненциально) растут с ростом i.