Entwurf, Analyse und Umsetzung von Algorithmen

O-Notation, L'Hopital

Prof. Dr. Rolf Backofen

Bioinformatics Group / Department of Computer Science

Entwurf, Analyse und Umsetzung von Algorithmen



Structure

O-Notation

Motivation / Definition Examples

Ω -Notation

Θ-Notation

Runtime

Summary Limit / Convergence L'Hôpital / l'Hospital Practical use

$\mathcal{O} ext{-Notation}$

Motivation

We are interested in:

- Example: sorting
 - Runtime of Minsort "is growing as"
 n²
 - Runtime of Heapsort "is growing as" n log n
- ▶ Growth of a function in runtime T(n)
 - ▶ the role of constants (e.g. 1ns) is minor
 - ▶ it is enough if relation holds for some $n \ge ...$
- Describe the growth of the function more formally
 - by the means of Landau-Symbols [Wik]):
 - \triangleright $\mathcal{O}(n)$ (Big O of n),
 - $ightharpoonup \Omega(n)$ (Omega of n),
 - $\triangleright \ \Theta(n)$ (Theta of n)

Definition

Big \mathcal{O} -Notation:

- ▶ Consider the function: $f: \mathbb{N} \to \mathbb{R}, n \mapsto f(n)$
 - N: Natural numbers → input size
 - $ightharpoonup \mathbb{R}$: Real numbers \rightarrow runtime

Example:

- f(n) = 3n
- $ightharpoonup f(n) = 2 n \log n$
- $f(n) = \frac{1}{10}n^2$ $f(n) = n^2 + 3 n \log n 4 n$

O-Notation Definition

Big \mathcal{O} -Notation:

- Given two functions f and g:
 - $f,g:\mathbb{N}\to\mathbb{R}$
- ▶ **Intuitive**: f is Big-O of g (f is $\mathcal{O}(g)$)
 - \triangleright ... if f relative to g does not grow faster than g
 - ▶ the growth rate matters, not the absolute values

Definition

Big \mathcal{O} -Notation:

- ▶ Informal: $f = \mathcal{O}(g)$
 - "=" corresponds to "is" not "is equal to"
 - ▶ ... if for some value n_0 for all $n \ge n_0$
 - ▶ $f(n) \le C \cdot g(n)$ for a constant C
 - $(f = \mathcal{O}(g)$: From a value n_0 for all $n \ge n_0 \to f(n) \le C \cdot g(n)$
- ▶ Formal: $f \in \mathcal{O}(g)$

Formal: $f \in \mathcal{O}(g)$ $\mathcal{O}(g) = \{ f : \mathbb{N} \to \mathbb{R} \longrightarrow \exists n_0 \in \mathbb{N}, \exists C > 0, \forall n \geq n_0 : f(n) \leq C \cdot g(n) \}$ "set of "for which" "it exists" "for all" "such that" all functions"

$\mathcal{O}\text{-Notation}$

Examples

Illustration of the Big O-Notation:

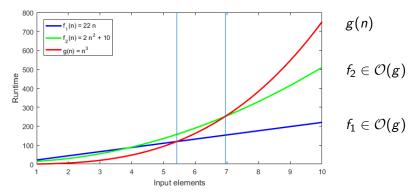


Figure: Runtime of two algorithms f_1, f_2

Example:

- ► f(n) = 5 n + 7, g(n) = n⇒ $5 n + 7 \in \mathcal{O}(g)$ ⇒ $f \in \mathcal{O}(g)$
- Intuitive:

$$f(n) = 5 n + 7 \rightarrow \text{linear growth}$$

Attention

 $f(n) \le g(n)$ is not guaranteed, better is $f(n) \le C \cdot g(n) \ \forall n \ge n_0$.

\mathcal{O} -Notation

We have to proof: $\exists n_0, \exists C, \forall n \geq n_0$: $5n + 7 \leq C \cdot n$.

$$5 n + 7 \le 5 n + n \text{ (for } n \ge 7)$$

= $6 n$

$$\Rightarrow n_0 = 7, C = 6$$

Alternate proof:

$$5 n + 7 \le 5 n + 7 n \text{ (for } n \ge 1\text{)}$$

$$= 12 n$$

$$\Rightarrow n_0 = 1, C = 12$$

Examples

Big O-Notation:

- ► We are only interested in the term with the highest-order, the fastest growing summand, the others will be ignored
- ▶ f(n) is limited from above by $C \cdot g(n)$

Examples:

$$2 n^{2} + 7 n - 20 \in \mathcal{O}(n^{2})$$

$$2 n^{2} + 7 n \log n - 20 \in$$

$$7 n \log n - 20 \in$$

$$5 \in$$

$$2 n^{2} + 7 n \log n + n^{3} \in$$

O-Notation Examples

Harder Example:

- Polynomes are simple
- More problematic: combination of complex functions

$$2\sqrt{x} + 3\ln x \in \mathcal{O}(??)$$

Ω -Notation

Definition

Omega-Notation:

- ► Intuitive:
 - ▶ $f \in \Omega(g)$, f is growing at least as fast as g
 - ► So the same as Big-O but with *at-least* and not *at-most*

Formal:
$$f \in \Omega(g)$$

$$\Omega(g) = \{ f : \mathbb{N} \to \mathbb{R} \mid \exists n_0 \in \mathbb{N}, \ \exists C > 0, \ \forall n \geq n_0 : \ f(n) \geq C \cdot g(n) \}$$
"in $O(n)$
we had \leq "

Proof

Example:

Proof of
$$f(n) = 5n + 7 \in \Omega(n)$$
:

$$\underbrace{5n+7}_{f(n)} \geq \underbrace{1 \cdot n}_{g(n)} \quad (\text{for } n \geq 1)$$

$$\Rightarrow n_0 = 1, C = 1$$

Ω -Notation

Examples

Illustration of the Omega-Notation:

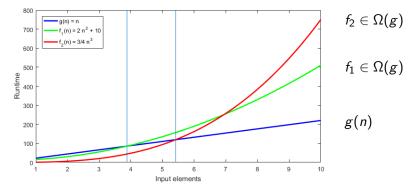


Figure: Runtime of two algorithms f_1, f_2

Ω -Notation

Examples

Big Omega-Notation:

- We are only interested in the term with the highest-order, the fastest growing summand, the others will be ignored
- ▶ f(n) is limited from underneath by $C \cdot g(n)$

Examples:

$$2 n^{2} + 7 n - 20 \in \Omega(n^{2})$$

$$2 n^{2} + 7 n \log n - 20 \in$$

$$7 n \log n - 20 \in$$

$$5 \in$$

$$2 n^{2} + 7 n \log n + n^{3} \in$$

⊖-Notation

Theta-Notation:

- ▶ **Intuitive**: *f* is Theta of *g* . . .
 - ightharpoonup ... if f is growing as much as g
 - $f \in \Theta(g)$, f is growing at the same speed as g

Formal:
$$f \in \Theta(g)$$

$$\Theta(g) = \underbrace{\mathcal{O}(g) \cap \Omega(g)}_{Intersection}$$

Example:

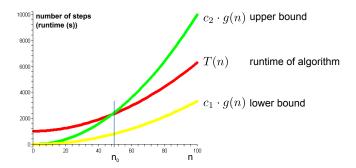
$$f(n) = 5 n + 7, \ f(n) \in \mathcal{O}(n), \ f(n) \in \Omega(n)$$

$$\Rightarrow f(n) \in \Theta(n)$$

Proof for $\mathcal{O}(g)$ and $\Omega(g)$ look at slides 11 and 17

Θ -Notation

Graphs



▶ f and g have the same "growth"

Runtime

Landau-Symbol Summary

Big O-Notation $\mathcal{O}(n)$:

- f is growing at most as fast as g
- $ightharpoonup C \cdot g(n)$ is the upper bound

Big Omega-Notation $\Omega(n)$:

- f is growing at least as fast as g
- $ightharpoonup C \cdot g(n)$ is the lower bound

Big Theta-Notation $\Theta(n)$:

- f is growing at the same speed as g
 - $ightharpoonup C_1 \cdot g(n)$ is the lower bound
 - $ightharpoonup C_2 \cdot g(n)$ is the upper bound

Runtime

Common Runtimes

Table: Common runtime types

| Runtime | Growth |
|---|------------------------------|
| $f \in \Theta(1)$ | constant time |
| $f \in \Theta(\log n) = \Theta(\log_k n)$ | logarithmic time |
| $f\in\Theta(n)$ | linear time |
| $f \in \Theta(n \log n)$ | n-log-n time (nearly linear) |
| $f \in \Theta(n^2)$ | squared time |
| $f \in \Theta(n^3)$ | cubic time |
| $f\in\Theta(n^k)$ | polynomial time |
| $f \in \Theta(k^n), f \in \Theta(2^n)$ | exponential time |

- So far discussed:
 - ► Membership in O(...) proofed by hand: Explicit calculation of n_0 and C
 - ► **However:** Both hint at limits in calculus

Limit / Convergence

Definition of "Limit"

- ► The limit L exists for an infinite sequence f_1, f_2, f_3, \ldots if for all $\epsilon > 0$ one $n_0 \in \mathbb{N}$ exists, such that for all $n \geq n_0$ the following holds true: $|f_n L| \leq \epsilon$
- ▶ A function $f: \mathbb{N} \to \mathbb{R}$ can be written as a sequence $\Rightarrow \lim_{n \to \infty} f_n = L$

The limit is converging:

$$\forall \epsilon > 0 \ \exists n_0 \in \mathbb{N} \ \forall n \geq n_0 \colon |f_n - L| \leq \epsilon$$

Limit / Convergence

- Example for the proof of a limit
- Function $f(n) = 2 + \frac{1}{n}$ with limes $\lim_{n \to \infty} f(n) = 2$
- "Engineering" solution: use $n = \infty$

$$\frac{1}{\infty} = 0 \Rightarrow \lim_{n \to \infty} f(n) = \lim_{n \to \infty} 2 + \frac{1}{n} = 2$$

Limit / Convergence

- Now a more formal proof for $\lim_{n\to\infty} 2 + \frac{1}{n} = 2$
- ▶ We need to show: for all given ϵ there is an n_0 such that for all $n \ge n_0$

$$\left|2 + \frac{1}{n} - 2\right| = \left|\frac{1}{n}\right| \le \epsilon$$

- ▶ E.g.: for $\epsilon = 0.01$ we get $\frac{1}{n} \le \epsilon$ for $n \ge 100$
- In general

$$n_0 = \left\lceil \frac{1}{\epsilon} \right\rceil$$

Then we get:

$$\left|\frac{1}{n}\right| = \frac{1}{n} \le \frac{1}{n_0} = \frac{1}{\left\lceil \frac{1}{\epsilon} \right\rceil} \le \frac{1}{\frac{1}{\epsilon}} = \epsilon \quad \Box$$

Limit / Convergence

Let $f, g: \mathbb{N} \to \mathbb{R}$ with an existing limit

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

Hence the following holds:

$$f \in \mathcal{O}(g)$$
 \Leftrightarrow $\lim_{n \to \infty} \frac{f(n)}{g(n)} < \infty$ (1)

$$f \in \Omega(g)$$
 \Leftrightarrow $\lim_{n \to \infty} \frac{f(n)}{g(n)} > 0$ (2)

$$f \in \Theta(g)$$
 \Leftrightarrow $0 < \lim_{n \to \infty} \frac{f(n)}{g(n)} < \infty$ (3)

Limit / Convergence

$$f \in \mathcal{O}(g) \Leftrightarrow \lim_{n \to \infty} \frac{f(n)}{g(n)} < \infty$$

Forward proof (\Rightarrow) :

$$f \in \mathcal{O}(g) \overset{\text{def. of } \mathcal{O}(n)}{\Rightarrow} \exists n_0, \ C \ \forall n \ge n_0 : \ f(n) \le C \cdot g(n)$$
$$\Rightarrow \exists n_0, \ C \ \forall n \ge n_0 : \frac{f(n)}{g(n)} \le C$$
$$\Rightarrow \qquad \lim_{n \to \infty} \frac{f(n)}{g(n)} \le C \quad \Box$$

Limit / Convergence

Backward proof (\Leftarrow) :

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

$$\Rightarrow \lim_{n \to \infty} \frac{f(n)}{g(n)} = C \qquad \text{For some } C \in \mathbb{R} \text{ (Limit)}$$

$$\stackrel{\text{def. limes}}{\Rightarrow} \exists n_0, \ \forall n \ge n_0: \qquad \frac{f(n)}{g(n)} \le C + \varepsilon \quad (e.g. \ \varepsilon = 1)$$

$$\Rightarrow \exists n_0, \ \forall n \ge n_0: \qquad f(n) \le \underbrace{(C+1)}_{O-notation \ constant} \cdot g(n)$$

$$\Rightarrow f \in \mathcal{O}(g) \quad \square$$

Intuitive:

$$\lim_{n\to\infty} 2 + \frac{1}{n} = 2 + \frac{1}{\infty} = 2$$

- With L'Hôpital:
 - ▶ Let $f, g : \mathbb{N} \to \mathbb{R}$
 - If $\lim_{n \to \infty} f(n) = \lim_{n \to \infty} g(n) = \infty/0$ $\Rightarrow \lim_{n \to \infty} \frac{f(n)}{g(n)} = \lim_{n \to \infty} \frac{f'(n)}{g'(n)}$
- Holy inspiration

you need a doctoral degree for that

The limit can not be determined in the way of an Engineer:

$$\lim_{n\to\infty}\frac{\ln(n)}{n}=\frac{\lim_{n\to\infty}\ \ln(n)}{\lim\limits_{n\to\infty}\ n} \quad \stackrel{\text{plugging in}}{\longrightarrow} \quad \frac{\infty}{\infty}$$

Determine the limit using L'Hôpital:

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

Using L'Hôpital:

Numerator: $\mathbf{f(n)} : n \mapsto \ln(n)$ Denominator: $\mathbf{g(n)} : n \mapsto n$ $\Rightarrow f'(n) = \frac{1}{n}$ (derivation from Numerator) $\Rightarrow g'(n) = 1$ (derivation from Denominator)

$$\lim_{n\to\infty}\frac{f'(n)}{g'(n)}=\lim_{n\to\infty}\frac{1}{n}=0\ \Rightarrow\ \lim_{n\to\infty}\frac{f(n)}{g(n)}=\lim_{n\to\infty}\frac{\ln(n)}{n}=0$$

What can we take for granted without proofing?

- Only things that are trivial
- ▶ It is always better to proof it

Examples:

$$\lim_{n\to\infty}\frac{1}{n}=0 \qquad \text{is trivial}$$

$$\lim_{n\to\infty}\frac{1}{n^2}=0 \qquad \text{is trivial}$$

$$\lim_{n\to\infty}\frac{\log(n)}{n}=0 \qquad \text{use L'Hopital}$$

O-Notation Practical use

Practical use:

- It is much easier to determine the runtime of an algorithm by using the $\mathcal{O}\text{-Notation}$
 - 1. Computing rules
 - 2. Practical use

Characteristics

► Transitivity:

$$f \in \Theta(g) \land g \in \Theta(h) \rightarrow f \in \Theta(h)$$

 $f \in \mathcal{O}(g) \land g \in \mathcal{O}(h) \rightarrow f \in \mathcal{O}(h)$
 $f \in \Omega(g) \land g \in \Omega(h) \rightarrow f \in \Omega(h)$

Symmetry:

$$f \in \Theta(g) \leftrightarrow g \in \Theta(f)$$

 $f \in \mathcal{O}(g) \leftrightarrow g \in \Omega(f)$

Reflexivity:

$$f \in \Theta(f)$$
 $f \in \Omega(f)$ $f \in \mathcal{O}(f)$

Calculation Rules

Trivial:

$$f \in \mathcal{O}(f)$$

 $C \cdot \mathcal{O}(f) = \mathcal{O}(f)$
 $\mathcal{O}(f+C) = \mathcal{O}(f)$

Addition:

$$\mathcal{O}(f) + \mathcal{O}(g) = \mathcal{O}(\max\{f, g\})$$

Multiplication:

$$\mathcal{O}(f) \cdot \mathcal{O}(g) = \mathcal{O}(f \cdot g)$$

Runtime Complexity

- ▶ The input size for all examples is *n*
- Basic operations

$$i1 = 0$$
 $\mathcal{O}(1)$

Sequences of basic operations

| i1 = 0 | $\mathcal{O}(1)$ | |
|----------|---|--|
| i2 = 0 | $egin{array}{c} \mathcal{O}(1) \ \mathcal{O}(1) \end{array} ight\} \hspace{1cm} 327 \cdot \mathcal{O}(1) = \mathcal{O}(1) \end{array}$ | |
| ••• | \cdots | |
| i327 = 0 | $\mathcal{O}(1)$ | |

Runtime Complexity

Loops

Runtime Complexity

Loops

$$\begin{array}{c|c} \text{for i in range}(0, \ n): \\ \hline \text{for j in range}(0, \ n-1): \\ \text{a1[i][j]} = 0 \\ \dots \\ \text{a137[i][j]} = 0 \\ \end{array} \begin{array}{c|c} \mathcal{O}(n) \\ \hline \mathcal{O}(n-1) \\ \hline \mathcal{O}(n-1) \\ \hline \mathcal{O}(1) \\ \dots \\ \hline \mathcal{O}(1) \\ \end{array} \begin{array}{c|c} \mathcal{O}(n) \cdot \mathcal{O}(n) \\ \hline \mathcal{O}(n) \cdot \mathcal{O}(n) \\ \hline \mathcal{O}(1) \\ \dots \\ \hline \mathcal{O}(1) \\ \hline \mathcal{O}(1) \\ \end{array} \begin{array}{c|c} \mathcal{O}(n) \cdot \mathcal{O}(n) \\ \hline \mathcal{O}(n) \cdot \mathcal{O}(n) \\ \hline \mathcal{O}(n) \cdot \mathcal{O}(n) \\ \hline \mathcal{O}(n) \cdot \mathcal{O}(n^2) \\ \hline \mathcal{O}(n) \cdot \mathcal{O}(n) \\ \hline \mathcal{O}(n) \cdot \mathcal{O}(n^2) \\ \hline \mathcal{O}($$

Runtime Complexity

Conditions

Arithmetic mean

- ► Input: List *x* with *n* numbers
- ▶ Output: a[i] is the arithmetic mean of x[0] to x[i]

```
def arithMean(x):
    a = [0] * len(x)
    for i in range(0, len(x)):
        s = 0
        for j in range(0, i+1):
            s = s + x[j]
        a[i] = s / (i+1)
    return a
```

O-Notation Runtime complexity

▶ How often will the instructions in the loop be executed, when the problem has size *n*?

$$1+2+\ldots+n=\frac{n\cdot(n+1)}{2}\in\mathcal{O}(n^2)$$

Way of speaking:

- With the \mathcal{O} -Notation we look at the behavior of a function when $n \to \infty$
- ▶ We only analyze the runtime when $n \ge n_0$
- We talk about asymptotic analysis, when we discuss cost, runtime, etc. as $\mathcal{O}(\ldots)$, $\Omega(\ldots)$ or $\Theta(\ldots)$

Attention:

- If you are using **asymptotic analysis**, you can not make any predictions about the runtime of smaller input sizes $(n < n_0)$
- For small input sizes (mostly n < 10), the runtime is predictably small
- $ightharpoonup n_0$ does not necessarily have to be small

Discussion

Examples:

- Let A and B be algorithms
 - A has the runtime f(n) = 80 n
 - ▶ B has the runtime $g(n) = 2 n \log_2 n$
- ▶ So $f = \mathcal{O}(g)$ but **not** $\Theta(g)$
 - ightharpoonup
 ightharpoonup A is asymptotic faster than B
 - ▶ ⇒ There is an n_0 for that $n \ge n_0$: $f(n) \le g(n)$

\mathcal{O} -Notation

Discussion

When is A faster then B?

We search the minimal n_0 :

$$f(n_0) = g(n_0)$$

$$80 n_0 = 2 n_0 \log_2 n_0$$

$$40 = \log_2 n_0$$

$$n_0 = 2^{40}$$

$$= (2^{10})^4 = (1024)^4$$

$$\approx (10^3)^4 = 10^{12}$$

$$\approx 1 \text{ trillion}$$

A ist faster than B if n_0 has more than 1 trillion elements

Runtime Examples

Continued

Logarithm of different bases differ only by a constant

$$\log_a n = \frac{\log_b n}{\log_b a} = \frac{1}{\log_b a} \cdot \log_b n$$

- ▶ Hence: $\log_a n \in \Theta(\log_b n)$
- For exponent this does not hold

$$3^n \notin \Theta(2^n)$$

▶ Proof: Use equation (1) from Slide 31

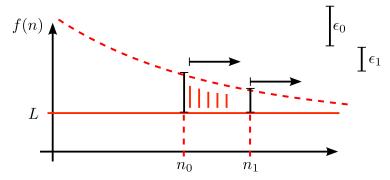
$$3^n \in \mathcal{O}(2^n) \Leftrightarrow \lim_{n \to \infty} \frac{3^n}{2^n} < \infty$$

► However:

$$\lim_{n \to \infty} \frac{3^n}{2^n} = \lim_{n \to \infty} \left(\frac{3}{2}\right)^n = \infty$$

Additional Figure

▶ Figure for slide 28



Further Literature

General

[MS08] Kurt Mehlhorn and Peter Sanders.
 Algorithms and data structures, 2008.
 https://people.mpi-inf.mpg.de/~mehlhorn/
 ftp/Mehlhorn-Sanders-Toolbox.pdf.

Further Literature

▶ Big O notation

[Wik] Big O notation https://en.wikipedia.org/wiki/Big_O_notation