Datenstrukturen und Algorithmen

Vorlesung 2: Asymptotische Effizienz (K3)

Joost-Pieter Katoen

Lehrstuhl für Informatik 2
Software Modeling and Verification Group

http://www-i2.informatik.rwth-aachen.de/i2/dsal12/

10. April 2012

RWTH AACHEN UNIVERSITY

Übersicht

Übersicht

Laufzeit von Algorithmen

Betrachte

Die Laufzeit eines Algorithmus ist keine Zahl, sondern eine Funktion.

Sie gibt die Laufzeit des Algorithmus für jede Eingabelänge an.

Worst-Case Laufzeit

Die Worst-Case Laufzeit $W(n)$ für Eingabelänge n ist die längste Laufzeit aus allen Eingaben mit Länge n.

Best-Case Laufzeit

Die Best-Case Laufzeit $B(n)$ für Eingabelänge n ist die kürzeste Laufzeit aus allen Eingaben mit Länge n.

Asymptotische Betrachtung

Die exakte Bestimmung der Funktionen $A(n)$, $B(n)$ und $W(n)$ ist üblicherweise sehr schwierig. Außerdem:

- ist sie von zweifelhaftem Nutzen für Vergleiche:
  Ist etwa $W(n) = 1021n$ besser als $W(n) = \frac{1}{2}n^2$?
- wollen wir maschinenabhängige Konstanten (z. B. Rechnergeschwindigkeit), Initialisierungsaufwand, usw. ausklammern.

Daher: Normalerweise keine exakte sondern asymptotische Betrachtung.

- Betrachte Wachstum der Laufzeit für $n \longrightarrow \infty$.
- Kurze Eingaben und konstante Faktoren werden vernachlässigt.
- Anschaulich: Wir lassen Glieder niedriger Ordnung weg, z. B.:

$$W(n) = 3n^4 + 5n^3 + 10 \in O(n^4)$$

(d. h. n^4 ist dominierender Faktor für $n \longrightarrow \infty$)

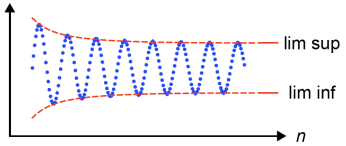
- So erhalten wir untere/obere Schranken für $A(n)$, $B(n)$ und $W(n)$!
- Mathematische Zutat: Asymptotische Ordnung von Funktionen.

Grenzwerte

Limes inferior und Limes superior

Sei $(x_i)_{i\in\mathbb{N}}$ die Folge $x_1, x_2, \ldots$. Dann:

1. $\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \left(\inf_{m\geqslant n} x_m \right)$
2. $\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \left(\sup_{m\geqslant n} x_m \right)$



Einige Fakten

1. Existieren $\liminf_{n\to\infty} x_n$ und $\limsup_{n\to\infty} x_n$: $\liminf_{n\to\infty} x_n \leqslant \limsup_{n\to\infty} x_n$.
2. Existiert $\lim_{n\to\infty} x_n$ dann: $\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n = \lim_{n\to\infty} x_n$.

Sind f, g differenzierbar, dann gilt $\lim_{n\longrightarrow\infty} \frac{g(n)}{f(n)} = \lim_{n\longrightarrow\infty} \frac{g'(n)}{f'(n)}$. L'Hôpital

Übersicht

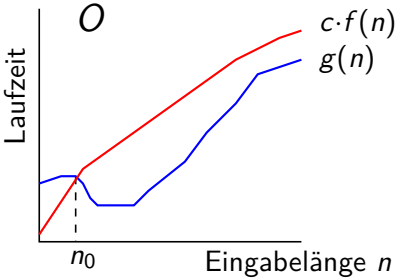
Die Klasse Groß-O (I)

Seien f und g Funktionen von $\mathbb{N}$ (Eingabelänge) nach $\mathbb{R}_{\geqslant 0}$ (Laufzeit) und $c > 0$.

$O(f)$ ist die *Menge* von Funktionen, die nicht schneller als f wachsen.

- $g \in O(f)$ heißt: $c \cdot f(n)$ ist obere Schranke für $g(n)$.

Diese Eigenschaft gilt ab einer Konstanten n_0; Werte unter n_0 werden vernachlässigt.



Die Klasse Groß-O liefert eine obere Schranke für die Komplexität einer Funktion.

- Kleinste obere Schranken sind von größtem Nutzen!
  $g \in O(n^2)$ sagt mehr als $g \in O(n^3)$.

Die Klasse Groß-O (II)

Definition

$g \in O(f)$ gdw. $\exists c > 0$, n_0 mit $\forall n \geqslant n_0 : 0 \leqslant g(n) \leqslant c \cdot f(n)$.

Definition (alternativ)

$g \in O(f)$ gdw. $\limsup_{n \longrightarrow \infty} \frac{g(n)}{f(n)} = c \geqslant 0$ mit $c \neq \infty$.

Theorem

Es seien $f, g : \mathbb{N} \longrightarrow \mathbb{R}$ zwei Funktionen. Es sei nur endlich oft $f(n) = 0$. Dann:

$\limsup_{n \longrightarrow \infty} \frac{g(n)}{f(n)}$ *existiert gdw.* $\exists c > 0$, n_0 *mit* $\forall n \geqslant n_0 : g(n) \leqslant c \cdot f(n)$.

Die Klasse Groß-O (III)

Theorem

Es seien $f, g : \mathbb{N} \longrightarrow \mathbb{R}$ zwei Funktionen. Es sei nur endlich oft $f(n) = 0$. Dann existiert $\limsup_{n \longrightarrow \infty} \frac{g(n)}{f(n)}$ gdw. $\exists c > 0, n_0. \forall n \geqslant n_0 : g(n) \leqslant c \cdot f(n)$.

Beweis.

„$\Longrightarrow$“: Sei $\limsup_{n \longrightarrow \infty} \frac{g(n)}{f(n)} = c < \infty$. Für $\varepsilon \geqslant 0$ es folgt $c + \varepsilon \geqslant \frac{g(n)}{f(n)}$ und $f(n) \neq 0$ bis auf endlich viele Ausnahmen. Ab einem $n_0 \in \mathbb{N}$ gilt für alle $n \geqslant n_0$ also: $c + \varepsilon \geqslant \frac{g(n)}{f(n)}$; und damit: $g(n) \leqslant (c + \varepsilon) \cdot f(n)$.

„$\Longleftarrow$“: Gegeben seien nun n_0', $c > 0$ so dass $\forall n \geqslant n_0' : g(n) \leqslant c \cdot f(n)$. Ab einem $n_0 \geqslant n_0'$ gilt (wie oben) außerdem $f(n) \neq 0$ für alle $n \geqslant n_0$. Damit ist $\forall n \geqslant n_0 : 0 \leqslant \frac{g(n)}{f(n)} \leqslant c$.

Die Folge $a_n = \frac{g(n)}{f(n)}$ ist in $[0, c]$, also beschränkt und abgeschlossen. Dann existiert $\limsup_{n \longrightarrow \infty} a_n < \infty$. □

Die Klasse Groß-O (IV)

Definition

$g \in O(f)$ gdw. $\exists c > 0$, n_0 mit $\forall n \geqslant n_0 : 0 \leqslant g(n) \leqslant c \cdot f(n)$.

Definition (alternativ)

$g \in O(f)$ gdw. $\limsup_{n \longrightarrow \infty} \frac{g(n)}{f(n)} = c \geqslant 0$ mit $c \neq \infty$.

Beispiel

Betrachte $g(n) = 3n^2 + 10n + 6$. Dann ist:

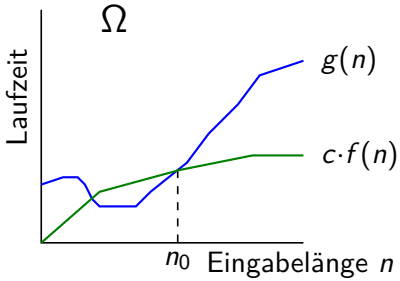
- $g \notin O(n)$, da $\limsup_{n \longrightarrow \infty} g(n)/n = \infty$.
- $g \in O(n^2)$, da $g(n) \leqslant 20n^2$ für $n \geqslant 1$.
- $g \in O(n^3)$, da $g(n) \leqslant \frac{1}{10} n^3$ für n hinreichend groß.

Die Klasse Groß-Omega (I)

$\Omega(f)$ ist die *Menge* von Funktionen, die nicht langsamer als f wachsen.

- $g \in \Omega(f)$ heißt: $c \cdot f(n)$ ist untere Schranke für $g(n)$.

Diese Eigenschaft gilt ab einer Konstanten n_0; Werte unter n_0 werden vernachlässigt.



Die Klasse Groß-Omega liefert eine untere Schranke für die Komplexität einer Funktion.

- Größte untere Schranken sind von größtem Nutzen! $g \in \Omega(n^2)$ sagt mehr als $g \in \Omega(n)$.

Die Klasse Groß-Omega (II)

Definition

$g \in \Omega(f)$ gdw. $\exists c > 0$, n_0 mit $\forall n \geqslant n_0 : c \cdot f(n) \leqslant g(n)$.

Definition (alternativ)

$g \in \Omega(f)$ gdw. $\liminf_{n \longrightarrow \infty} \frac{g(n)}{f(n)} > 0$.

Beispiel

Betrachte $g(n) = 3n^2 + 10n + 6$. Dann ist:

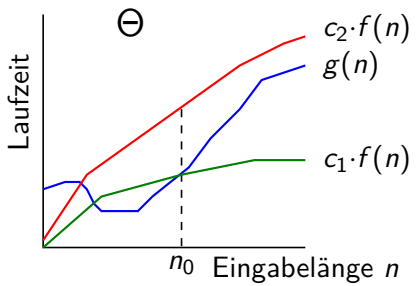
- $g \in \Omega(n)$, da $\liminf_{n \longrightarrow \infty} g(n)/n = \infty > 0$.
- $g \in \Omega(n^2)$, da $\liminf_{n \longrightarrow \infty} g(n)/n^2 = 3 > 0$.
- $g \notin \Omega(n^3)$, da $g(n) \leqslant 5n^3$ für $n \geqslant 2$.

Die Klasse Groß-Theta (I)

$\Theta(f)$ ist die *Menge* von Funktionen, die genauso schnell wie f wachsen.

- $g \in \Theta(f)$ heißt: $c_2 \cdot f(n)$ ist obere Schranke **und** $c_1 \cdot f(n)$ ist untere Schranke für $g(n)$.

Diese Eigenschaft gilt ab einer Konstanten n_0; Werte unter n_0 werden vernachlässigt.



Die Klasse Groß-Theta liefert eine obere und untere Schranke für die Komplexität einer Funktion.

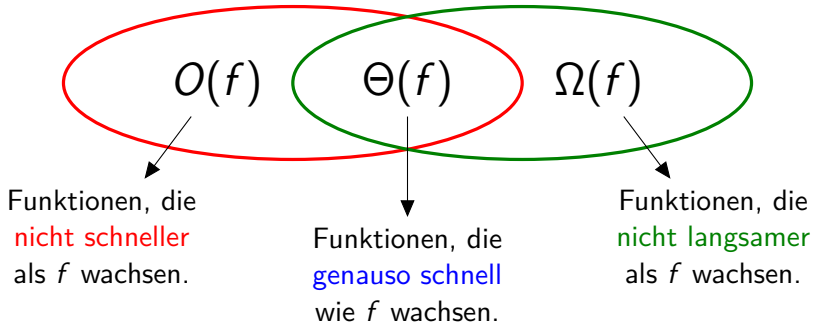
Definition

$g \in \Theta(f)$ gdw. $\exists c_1, c_2 > 0$, n_0 mit $\forall n \geqslant n_0 : c_1 \cdot f(n) \leqslant g(n) \leqslant c_2 \cdot f(n)$.

Die Klassen O, Ω und Θ

Beziehung zwischen O, Ω und Θ

$g \in \Theta(f)$ gdw. $g \in O(f)$ und $g \in \Omega(f)$.



Lemma

$g \in \Theta(f)$ *wenn* $\lim_{n \longrightarrow \infty} \frac{g(n)}{f(n)} = c$ *für ein* $0 < c < \infty$.

Die Klasse Groß-Theta (II)

Definition

$g \in \Theta(f)$ gdw. $\exists c_1, c_2 > 0$, n_0 mit $\forall n \geqslant n_0 : c_1 \cdot f(n) \leqslant g(n) \leqslant c_2 \cdot f(n)$.

Lemma

$g \in \Theta(f)$ *wenn* $\lim_{n \longrightarrow \infty} \frac{g(n)}{f(n)} = c$ *für ein* $0 < c < \infty$.

Beispiel

Betrachte $g(n) = 3n^2 + 10n + 6$. Dann ist:

- $g \notin \Theta(n)$, da zwar $g \in \Omega(n)$, aber $g \notin O(n)$.
- $g \in \Theta(n^2)$, da $\lim_{n \longrightarrow \infty} g(n)/n^2 = 3$.
- $g \notin \Theta(n^3)$, da zwar $g \in O(n^3)$, aber $g \notin \Omega(n^3)$.

Beispiel: Fibonacci-Zahlen

Wachstum einer Kaninchenpopulation

- Zu Beginn gibt es ein Paar geschlechtsreifer Kaninchen.
- Jedes neugeborene Paar wird im zweiten Lebensmonat geschlechtsreif.
- Jedes geschlechtsreife Paar wirft pro Monat ein weiteres Paar.
- Sie sterben nie und hören niemals auf.

$$Fib(0) = 0 \quad \text{und} \quad Fib(1) = 1$$
$$Fib(n+2) = Fib(n+1) + Fib(n) \quad \text{für } n \geqslant 0.$$

n	0	1	2	3	4	5	6	7	8	9	...
$Fib(n)$	0	1	1	2	3	5	8	13	21	34	...

$Fib(n) \in O(2^n)$ und $Fib(n) \in \Omega\left(2^{\frac{n}{2}}\right)$

Einige elementare Eigenschaften

Reflexivität

- $f \in O(f)$, $f \in \Omega(f)$, $f \in \Theta(f)$.

Transitivität

- Aus $f \in O(g)$ und $g \in O(h)$ folgt $f \in O(h)$.
- Aus $f \in \Omega(g)$ und $g \in \Omega(h)$ folgt $f \in \Omega(h)$.
- Aus $f \in \Theta(g)$ und $g \in \Theta(h)$ folgt $f \in \Theta(h)$.

Symmetrie von Θ

- $f \in \Theta(g)$ gdw. $g \in \Theta(f)$.

Beziehung zwischen O und Ω

- $f \in O(g)$ gdw. $g \in \Omega(f)$.

Beispiel

Die folgenden 3 Aussagen sind alle gültig:

(a) $\log_a n \in \Theta(\log_b n)$, (b) $\log_b n \in \Theta(\log_a n)$, (c) $O(\log_a n) = O(\log_b n)$.

Beweis.

Wir beweisen (c). Zu zeigen: $\exists c_1, c_2 > 0$, so dass

$$\underbrace{\log_a n \leqslant c_1 \cdot \log_b n}_{\log_a n \in O(\log_b n)} \quad \text{und} \quad \underbrace{\log_b n \leqslant c_2 \cdot \log_a n}_{\log_b n \in O(\log_a n)}.$$

$$\text{Dann: } log_a n \leqslant c_1 \cdot \log_b n \;\Leftrightarrow\; \log_a n \leqslant c_1 \cdot \frac{\log_a n}{\log_a b} \;\Leftrightarrow\; \log_a b \leqslant c_1$$

Wähle $c_1 \geqslant \lceil \log_a b \rceil$.

Analog erhalten wir $\log_b a \leqslant c_2$; dann wähle $c_2 \geqslant \lceil \log_b a \rceil$.

Die Aussagen (a) und (b) folgen auf ähnliche Weise. □

Die Klassen Klein-O, Klein-Omega

$o(f)$ ist die *Menge* von Funktionen, die echt langsamer als f wachsen.

Definition

$g \in o(f)$ gdw. $\forall c > 0$, $\exists n_0$ mit $\forall n \geqslant n_0 : 0 \leqslant g(n) < c \cdot f(n)$.

$\omega(f)$ ist die *Menge* von Funktionen, die echt schneller als f wachsen.

Definition

$g \in \omega(f)$ gdw. $\forall c > 0$, $\exists n_0$ mit $\forall n \geqslant n_0 : c \cdot f(n) < g(n)$.

Beziehung zwischen o und ω

- $f \in o(g)$ gdw. $g \in \omega(f)$.

Übersicht

Platzkomplexität

Platzkomplexität

Unter der Platzkomplexität eines Problems versteht man den (minimalen) Bedarf an Speicherplatz eines Algorithmus zur Lösung dieses Problems, in Abhängigkeit von der Länge der Eingabe.

Platzkomplexität

- Nicht nur die Zeitkomplexität, sonder auch der Speicherbedarf ist wichtig!
- Dilemma: Eine Reduktion der Zeitkomplexität führt oft zur Erhöhung der Platzkomplexität, und vice versa.
- Dies werden wir in später in der DSAL Vorlesung öfters feststellen.

Beispiel: Platzkomplexität von Liedern [Knuth, 1984]

Beispiel

Betrachte eine Lied mit n Wörter, d. h. die Eingabelänge ist n.

Was ist die benötigte Platzkomplexität $S(n)$ um ein Lied der Länge n zu singen?

Obere und untere Schranken

1. $S(n) \in O(n)$, da höchstens n verschiedene Wörter gespeichert werden müssen.
2. $S(n) \in \Omega(1)$, da wir mindestens eine Sache über das Lied wissen müssen, um es singen zu können.

Kann man die Platzkomplexität durch Refrains (= Kehrverse) reduzieren?

Refrains

Refrain

Die Wiederkehr von textlich/musikalisch (wenigstens überwiegend) identischen Zeilen am Schluss einer Strophe oder zwischen den Strophen.

Beispiel

Soo.. Bye, bye miss American Pie
Drove me Chevy to the levee but the levee was dry
Them good old boys were drinking whiskey and rye?
Singing this will be the day that I die
this will be the day that I die *[Don McLean]*

Platzkomplexität

Speichere den Refrain einmal und singe ihn $O(n)$ Mal.

Dann: $S(n) \in O(n)$, da die Anzahl der Wörter immer noch $O(n)$ ist; z. B. bei Strophelänge = Refrainlänge halbiert sich der Speicherbedarf.

Die k Weihnachtstage

Reduktion der Platzkomplexität

Reduziere $S(n)$ durch eine bestimmte, sich ändernde Liedstruktur, etwa:

On the kth day of Xmas, my true love gave to me $gift_k$, $gift_{k-1}, \ldots, gift_1$
On the $(k-1)$st day of Xmas, my true love gave to me $gift_{k-1}, \ldots, gift_1$
. . . . . .
On the first day of Xmas, my true love gave to me a bottle of wine

Bekanntere Variante: „Old MacDonald had a farm“.

Platzkomplexität

Die benötigte Zeit, um das Lied zu singen ist (betrachte keine Refrains):

$$n = \sum_{i=1}^{k} i = k \cdot \left(\frac{k+1}{2}\right) \in \Theta(k^2)$$

Da $n \in \Theta(k^2)$ folgt $k \in O(\sqrt{n})$, also $S(n) \in O(\sqrt{n})$.

100 Bierflaschen

Eine weitere Vereinfachung

Ein (sehr) langweiliges Lied für lange Autofahrten:

n bottles of beer on the wall, n bottles of beer
You take one down and pass it around
$n-1$ bottles of beer on the ball
. . . . . *[Andy Kaufman]*

Platzkomplexität

$S(n) \in O(\log n)$, da nur der Wert von n von Bedeutung ist. Dafür reichen $\log n$ Bits aus.

Es geht jedoch noch etwas einfacher, nämlich indem man auf das Zählen verzichtet.

Untere Schranke?

Ein Lied mit $S(n) \in \Theta(1)$

That's the way, uh-huh, uh-huh
I like it, uh-huh, huh
. . . . . *[KC & the Sunshine Band, 1977]*