Informatik II: Algorithmen und Datenstrukturen SS 2017

Vorlesung 6b, Mittwoch, 31. Mai 2017
(Dynamische Felder, Teil 2: amortisierte Analyse)

Prof. Dr. Hannah Bast
Lehrstuhl für Algorithmen und Datenstrukturen
Institut für Informatik
Universität Freiburg

Blick über die Vorlesung heute

- Drumherum
  - Das offensichtlich Richtige — Warum so schwer?
  - Nächste Woche ist Pfingsten — Keine Vorlesungen
- Dynamische Felder, Teil 2
  - Vergrößerungsstrategien — Fortsetzung von VL6a
  - Laufzeitanalyse — amortisierte Analyse
  - Potenzialfunktion — Konzept dahinter
  - ÜB6: kurze Besprechung dazu und Gelegenheit für Fragen

Das offensichtlich Richtige tun 1/5

- Warum so schwer? … Ihre Kommentare dazu
  – Frage sehr allgemein ... dafür viele individuelle Antworten!
  – Kenne das nur, wenn das offensichtliche ein Übungsblatt ist
  – Das ist nichts für Bad Boys... und Ladys lieben Bad Boys
  – Mensch: 99% Körperintelligenz, 0.9% Kulturintelligenz, 0.1% Individualität … Gefühl fehlender Kontrolle kein Wunder
  – Die Wahrheit ist, wir haben nicht die Kontrolle
  – Satan ist überall! Ihr habt den Teufel zum Vater! (vgl. Johannes 8, 44) Gelobt seist du Christi!
  – No man chooses evil because it is evil; he only mistakes it for happiness, the good he seeks.

Das offensichtlich Richtige tun 2/5

- Präfrontaler Kortex
  - Linke Seite: leitet **Aktionen** ein
  - Rechte Seite: **widersteht** Versuchungen
  - Mitte: **wägt ab** zwischen Alternativen
  - Folgen bei Verletzung oder Zerstörung dieser Areale:

    Keine Langzeitplanung mehr möglich

    Entscheidungsunfähigkeit

    Krankhaftes Beharren und Inflexibilität

    Emotionale Verflachung, Enthemmung, Euphorie ohne Grund

Das offensichtlich Richtige tun 3/5

- Präfrontaler Kortex, Evolution
  - Die Kortikalisierung des Gehirns ist wohl der wesentlichste Faktor für die besondere Intelligenz des Homo Sapiens

    Viel wichtiger als die Zunahme des Gehirnvolumens
  - Kam in der Evolution als letztes dazu
  - Entsprechend wird in existenziellen **Stress-Situationen** und bei **Energiemangel** der präfrontale Kortex auch als erstes "abgeschaltet" bzw. auf "Sparflamme" gestellt

    Dann übernehmen entwicklungsgeschichtlich ältere Teile des Gehirns die Kontrolle … das will man (meistens) nicht

Das offensichtlich Richtige tun 4/5

- Präfrontaler Kortex, Unterstützung
  - Genügend **Schlaf** und genügend Nahrung

    Das ist quasi die Grundversorgung, wenn die fehlt ist das so wie wenn der Strom abgeschaltet wurde

    Gehirn verbraucht sehr viel Energie (in der Form von Glukose)
  - Vermeidung von negativem (existenziellen) Stress

    Sonst geht die ganze Energie in die Stress-Reaktion
  - Willenskraft verhält sich in vielerlei Hinsicht wie Muskelkraft

    Es ist eine begrenzte Ressource, die irgendwann aufgebraucht ist und dann Zeit zur Regeneration braucht

    Ganz wichtig: man kann (und sollte) sie trainieren

Das offensichtlich Richtige tun 5/5

- Präfrontaler Kortex, Training
  - Bewegung … das erfordert Planung und Willenskraft

    Trainiert Körper **und** Geist → zwei Fliegen mit einer Klappe
  - Sich etwas vornehmen und es dann auch umsetzen

    Kann etwas ganz einfaches sein, wie z.B. für X Wochen immer fünf Minuten vorher zu einem Termin kommen

    Sollte nicht-trivial sein, aber auch keine Überforderung

    Wenn das Ziel erreicht ist, oder es langweilig wird, eine andere Aufgabe angehen (aber eine nach der anderen)
  - Und wie gesagt: genügend **schlafen** und vernünftig essen

    Sonst geht schon rein energietechnisch gar nichts

Laufzeitanalyse 1/7

- Vorbetrachtungen
  - Unsere bisherigen Laufzeitanalysen haben angenommen, dass wir nur Elemente hinzufügen (mit pushBack)
  - Dann können wir leicht ausrechnen, wann es zu einer Reallokation kommt

    Mit der Verdoppelungstrategie bei: 1, 2, 4, 8, …
  - Im Folgenden wollen wir unsere Analyse verallgemeinern auf beliebige Abfolgen von pushBack und popBack

    Dann können wir nicht mehr so leicht vorhersagen, wann realloziert werden muss

Laufzeitanalyse 2/7

size

capacity

- Notation
  - Gegeben n Operationen $O_1, \ldots, O_n$

    Eine beliebige Abfolge von **pushBack** und **popBack**
  - Sei s_i die Anzahl Elemente **nach** Operation O_i $(s_0 := 0)$
  - Sei c_i die Größe des Feldes **nach** Operation O_i $(c_0 := 0)$

    Es gilt immer $c_i \geq s_i$ (Feld muss immer "groß genug" sein)
  - Sei wie in VL6a gestern T_i die Zeit für Operation O_i
    - Falls Reallokation nicht nötig: $T_i \leq A$
    - Falls Reallokation nötig: $T_i \leq A + B \cdot s_i$

    für irgendwelche Konstanten A und B unabhängig von n

Laufzeitanalyse 3/7

- Wir analysieren folgende Variante
  - Falls Operation O_i ein pushBack ist:
    - Reallokation genau dann, wenn $s_{i-1} = c_{i-1}$
    - Vergrößerung so, dass danach $c_i = 2 \cdot s_i$
  - Falls Operation O_i ein popBack ist:
    - Reallokation genau dann, wenn $4 \cdot s_{i-1} \leq c_{i-1}$
    - Verkleinerung so, dass danach $c_i = 2 \cdot s_i$
  - In beiden Fällen ist also direkt nach der Reallokation

    $c_i = 2 \cdot s_i$ … also das interne Feld exakt **doppelt** so groß
  - Das Feld ist immer zu mindestens **einem Viertel** voll

    ÜB6: Verallgemeinern auf $s_i \geq f \cdot c_i$ … für $f \in (0,1)$ beliebig

zum Beispiel:

$s_{i-1} = 17 = c_{i-1}$
(vor der Operation)

$s_i = 18$, $c_i = 36$
(nach der Operation)

$s_{i-1} = 12$, $c_{i-1} = 50$
(vor der Operation)

$s_i = 11$; $c_i = 22$
(nach der Operation)

Laufzeitanalyse 4/7

nicht genau:
$X+1$ Operationen à 2000 €
$\Rightarrow \frac{2000€}{1001} < 2€$ / OP.

- Beweisidee
  - Nach einer teuren Op. kommen viele billige Operationen

    Teuer sind nur Operationen, wo realloziert werden muss
  - **Genauer:** wenn nach einer Operation die $\overset{1000€}{X}$ gekostet hat $\overset{1000}{X}$ Operationen kommen die alle nur $\overset{1€}{1}$ kosten, sind die Gesamtkosten bei n Operationen höchstens $\overset{2000€}{2 \cdot n}$ 2€ / OP.
  - **Allgemeiner:** wenn nach einer Operation mit Kosten $\overset{5000€}{c_1 \cdot X}$ $\overset{1000}{X}$ Operationen kommen mit Kosten $\overset{2€}{c_2}$, dann sind die Gesamtkosten bei n Operationen höchstens $\overset{7000€}{(c_1 + c_2) \cdot n}$ 7€ / OP.

    Man kann die Kosten der teuren Operationen quasi auf die billigen Operationen "umlegen" (amortisieren)

    Analogie: neue Heizung (teuer) spart dann monatlich Öl

Laufzeitanalyse 5/7

Realloz. Realloz.

$O_i \underbrace{\quad \cdots \quad}_{\text{dazwischen keine Realloz.}} O_j$

- Formaler Beweis
  - **Lemma:** Wenn bei O_i eine Reallokation stattfindet, und die nächste Reallokation danach bei O_j, dann $j - i \geq s_i / 2$

    Nächste Reallokation frühestens nach $s_i/2$ Operationen
  - **Korollar:** Seien die Kosten einer Operation O_i ohne Reallokation $T_i \leq A$ und mit Reallokation $T_i \leq A + B \cdot s_i$

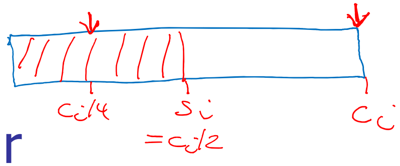
    Dann ist $T_1 + T_2 + \ldots + T_n \leq (A + 3B) \cdot n$

    Eine Operation kostet also im Durchschnitt $\leq A + 3B = O(1)$
  - Wir beweisen auf den nächsten beiden Folien erst das Lemma und dann, weil es so schön ist, auch das Korollar

Laufzeitanalyse 6/7

- Beweis des Lemmas
  - Zu zeigen: Wenn bei O_i eine Reallokation stattfindet, und die nächste Reallokation danach bei O_j, dann $j - i \geq s_i / 2$
  - Nach O_i ist auf jeden Fall $c_i = 2 \cdot s_i$

    Egal ob es ein pushBack oder ein popBack war



  - Nächste Vergrößerung wenn $s_j = c_i = 2 \cdot s_i$

    Also nach frühestens $s_i \geq s_i / 2$ Operationen
  - Nächste Verkleinerung wenn $s_j = \lfloor c_i / 4 \rfloor = \lfloor s_i / 2 \rfloor$

    Also nach frühestens $s_i - \lfloor s_i / 2 \rfloor \geq s_i / 2$ Operationen
  - Das Lemma gilt also in jedem Fall

Laufzeitanalyse 7/7

- Beweis des Korollars
  - Seien $O_{i_1}, O_{i_2}, \ldots, O_{i_k}$ die Op.n bei denen realloziert in chronologischer Reihenfolge, also: $i_1 < i_2 < \ldots < i_k$
  - Damit $T_1 + T_2 + \ldots + T_n \leq A \cdot n + B \cdot \underline{(s_{i_1} + s_{i_2} + \ldots + s_{i_k})}$
  - Zu zeigen: dann $T_1 + T_2 + \ldots + T_n \leq (A + 3B) \cdot n$

LEMMA $\Rightarrow$ $\underbrace{i_2 - i_1 \geq s_{i_1}/2}_{\Leftrightarrow\ s_{i_1} \leq 2\cdot(i_2 - i_1)}$, $\underbrace{i_3 - i_2 \geq s_{i_2}/2}_{s_{i_2} \leq 2\cdot(i_3 - i_2)}$, ...

Dann: $s_{i_1} + \cdots + s_{i_k} \leq 2\cdot(i_2 - i_1) + 2\cdot(i_3 - i_2) + \cdots + 2\cdot(i_k - i_{k-1}) + s_{i_k}$

$= 2\cdot(\overbrace{i_2 - i_1 + i_3 - i_2 + \cdots + i_k - i_{k-1}}^{\text{sog. Teleskopsumme}}) + s_{i_k}$

$= i_k - i_1$

Also: $\sum_{i=1}^{n} T_i \leq A\cdot n + B\cdot(2\underbrace{(i_k - i_1)}_{\leq n} + \underbrace{s_{i_k}}_{\leq n}) \leq (A + 3\cdot B)\cdot n$ ☒

weil: alle $i_1, \ldots, i_k \leq n$ weil: n OP.

weil: worst case, n mal push Back

Beweis mit Potenzialfunktion 1/5

- Variante des Beweises
  - Der Beweis auf den vorherigen Folien hat die Kosten für eine Folge von Operationen quasi "zu Fuß" analysiert
  - Man kann solche Beweise auch mit Hilfe einer sogenannten **Potenzialfunktion** führen
  - Intuitiv misst die Potenzialfunktion, wie lange es bis zur nächsten teuren Operation dauert

    Teure Operationen (wie unser reallocate) sollen die Potenzialfunktion erhöhen, und zwar um mindestens X, wenn die Kosten der Operation $\Theta(X)$ waren

    Billige Operationen sollen die Potenzialfunktion nur geringfügig (ideal: um eine Konstante) erniedrigen

Beweis mit Potenzialfunktion 2/5

wenn $\Phi_i - \Phi_{i-1}$ groß,
dann darf O_i teuer sein

- Potenzialfunktionen, Mastertheorem
  - Gegeben eine Folge von n Operationen $O_1, \ldots, O_n$ auf einer beliebigen Datenstruktur
  - Sei Φ eine Potenzialfunktion, wobei Φ_i = der Wert der Potenzialfunktion **nach** O_i und Φ_0 = Wert am Anfang ≥ 0
  - Sei T_i die Laufzeit für O_i mit $T_i \leq A + B \cdot (\Phi_i - \Phi_{i-1})$
  - Dann ist die Gesamtlaufzeit $\Sigma\, T_i = O(n + \Phi_n)$
  - **Beweis:**

$$\sum_{i=1}^{n} T_i \leq A \cdot n + B \cdot \underbrace{(\Phi_1 - \Phi_0 + \Phi_2 - \Phi_1 + \Phi_3 - \Phi_2 + \ldots + \Phi_n - \Phi_{n-1})}$$

wie vorher eine Teleskopsumme
zammenfallen: $\Phi_n - \Phi_0$

$$= A \cdot n + B \cdot (\Phi_n - \underbrace{\Phi_0}_{\geq 0}) \leq A \cdot n + B \cdot \Phi_n$$

$$= O(n + \Phi_n)$$

Beweis mit Potenzialfunktion 3/5

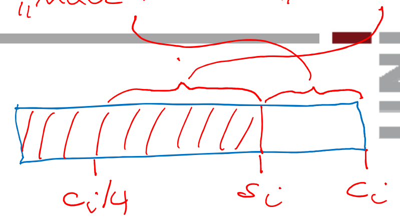
- Anwendung des Satzes für dynamische Felder
  - Wie vorher s_i = Anz. Elemente und c_i = Kapazität **nach** O_i
  - Definiere $\Phi_i := \min \{ c_i - s_i , s_i - c_i / 4 \}$ … $\Phi_0 := 0$

    Minimum Anzahl freier Plätze nach links und nach rechts
  - Dann gilt $T_i \leq A' + B' \cdot (\Phi_i - \Phi_{i-1})$ … Beweis nächste Folie
  - Und damit gemäß Satz $\Sigma\, T_i = O(n + \underbrace{\Phi_n}_{\leq n}) = O(n)$

Beweis mit Potenzialfunktion 4/5

$\Phi_i = \min\{ c_i - s_i , s_i - c_i/4 \}$

Platz „nach rechts“ Platz „nach links“



- Beweis, dass $T_i \leq A' + B' \cdot (\Phi_i - \Phi_{i-1})$
  - Wir unterscheiden zwei Fälle:

Fall 1: O_i ist "billig" (ohne Reallokation) ... $T_i \leq A$

Potenzial ändert sich: um höchstens 1

Damit $\Phi_i - \Phi_{i-1} \geq -1$ und also $T_i \leq A \leq A + \underbrace{1 - 1}_{\leq \Phi_i - \Phi_{i-1}} \leq \underbrace{A+1}_{=:A'} + \Phi_i - \Phi_{i-1}$

Fall 2: O_i ist "teuer" (mit Reallokation) ... $T_i \leq A + B \cdot s_i$

Potenzial vor Reallokation: 0

siehe Folie X < 18

Potenzial nach Reallokation: $\min\{ s_i , s_i/2 \} \geq s_i/2$

Damit $\Phi_i - \Phi_{i-1} \geq s_i / 2$ $\iff s_i \leq 2(\Phi_i - \Phi_{i-1})$

Also $T_i \leq A + B \cdot s_i$ $\leq \underbrace{A}_{\leq A'} + \underbrace{2B}_{=:B'} \cdot (\Phi_i - \Phi_{i-1})$

Beweis mit Potenzialfunktion 5/5

- Vergleich der beiden Beweise
  - Für die dynamischen Felder war der "zu Fuß" Beweis nicht wirklich einfacher, aber direkter
  - Der Beweis über die Potenzialmethode war intuitiver, weil das Potenzial eine intuitive Bedeutung hat:

    Und zwar: Mindestdauer bis zur nächsten Reallokation
  - Spätere Vorlesung: amortisierte Analyse, wo der Beweis über eine Potenzialfunktion einfacher **und** intuitiver ist

    ÜB6, Aufgabe 3: dort ebenfalls !

Literatur / Links

- Dynamische Felder: Laufzeitanalyse
  - In Mehlhorn/Sanders:

    3.2 Unbounded Arrays
  - In Wikipedia

    http://en.wikipedia.org/wiki/Dynamic_array
  - Potenzial vs. Potential

    http://www.duden.de/rechtschreibung/potenzial