

Letzte Änderung 19.12.13

Untere Schranken für das Sortier-Problem

Zur Präzisierung und Ergänzung der Vorlesung soll hier noch einmal kurz auf den Nachweis von *unteren* Schranken für den Sortier-Aufwand eingegangen werden.

Satz. *Es gibt eine Konstante c , so dass die worst-case-Zeitkomplexität $A_w(n)$ eines jeden vergleichsbasierten Sortierverfahrens für genügend große n nach unten beschränkt ist durch:*

$$(*) \quad A_w(n) \geq \log_2(n!) \geq c \cdot n \cdot \log_2(n).$$

Beweis. Wir betrachten hier nur die letzte Ungleichung. Der Einfachheit halber nehmen wir an, dass n eine gerade Zahl ist, so dass $\frac{n}{2}$ wieder eine natürliche Zahl ist. Dann ist

$$n! = \underbrace{n \cdot (n-1) \cdot \dots \cdot \left(\frac{n}{2} + 1\right)}_{n/2 \text{ Faktoren, jeder } > n/2} \cdot \frac{n}{2} \cdot \dots \cdot 2 \cdot 1 > \left(\frac{n}{2}\right)^{\left(\frac{n}{2}\right)}$$

und daher

$$\log_2(n!) > \log_2\left(\left(\frac{n}{2}\right)^{\left(\frac{n}{2}\right)}\right) = \frac{n}{2} \cdot \log_2\left(\frac{n}{2}\right) = \frac{n}{2} \cdot (\log_2(n) - 1).$$

Die Festlegung einer Konstanten c ist jetzt sicher nicht mehr schwer. Für $c = 1/4$ gilt die gewünschte Abschätzung etwa für $n \geq 5$. \square

Bemerkung. Unter Verwendung eines zur \mathcal{O} -Notation dualen Landau-Symbols Ω für untere Schranken lässt sich $(*)$ auch schreiben als

$$A_w(n) = \Omega(n \cdot \log(n)).$$

Dabei ist allgemein $f(n) = \Omega(g(n))$ für Funktionen $f, g : \mathbb{N} \rightarrow \mathbb{R}^+$ definiert durch

$$\exists n_0 \in \mathbb{N}, c \in \mathbb{R}^+, \text{ s.d. } \forall n \geq n_0 : f(n) \geq c \cdot g(n).$$

Graphen

Definition. Ein *Graph* $G = (V, E)$ besteht aus einer endlichen Menge V von *Knoten* (engl. vertex) und einer endlichen Menge E von *Kanten* (engl. edge).

- G heißt *gerichtet*, wenn $E \subseteq (V \times V) - \{(v, v) \mid v \in V\}$.
- G heißt *ungerichtet*, wenn jede Kante $e \in E$ die Form $e = \{v, w\}$ mit $v, w \in V, v \neq w$ hat.

Definition. Ein *Weg* oder *Pfad* in einem Graphen G ist eine Folge $\pi = (v_i)_{i=1}^n$ von Knoten, so dass für alle $i = 1, \dots, n-1$

- $(v_i, v_{i+1}) \in E$, falls G gerichtet ist,
- $\{v_i, v_{i+1}\} \in E$, falls G ungerichtet ist. Im ungerichteten Fall verlangen wir zudem, dass $v_i \neq v_{i+2}$, $i = 1, \dots, n - 2$.

Wir schreiben einen Weg oft in der Form $\pi : v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_n$.

Ein Knoten w ist aus einem Knoten v *erreichbar* (symbolisch $v \rightarrow^* w$), wenn es einen Weg $v = v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_n = w$ gibt. Man beachte, dass insbesondere auch $v \rightarrow^* v$ für jeden Knoten v . (Wege der Länge 1 sind durch die Definition nicht ausgeschlossen.)

Ein Weg ist ein *Kreis* oder *Zyklus*, wenn $v_1 = v_n$ und

- $n > 1$, falls G gerichtet ist,
- $n > 2$, falls G ungerichtet ist.

Oft interessiert man sich für wiederholungsfreie Wege und Kreise:

Definition. Ein *einfacher* Weg ist ein Weg $v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_n$, bei dem alle v_i , $i = 1, \dots, n$, paarweise verschieden sind.

Ein *einfacher* Kreis ist ein Kreis $v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_n = v_1$, bei dem alle v_i , $i = 1, \dots, n - 1$ paarweise verschieden sind.

Definition. Eine *Zusammenhangskomponente* eines ungerichteten Graphen $G = (V, E)$ ist eine Teilmenge $V' \subseteq V$, so dass

- je zwei Elemente in V' durch einen Weg verbunden sind, der nur Knoten aus V' enthält,
- und zusätzlich V' maximal ist mit dieser Eigenschaft ist, d.h. Knoten in $V - V'$ sind von keinem Knoten aus V' erreichbar.

Definition. Eine *starke Zusammenhangskomponente* in einem gerichteten Graphen $G = (V, E)$ ist eine Teilmenge $V' \subseteq V$, so dass

- Für je zwei Elemente in $v, w \in V'$ gilt: $v \rightarrow^* w$ und $w \rightarrow^* v$, (Man beachte, dass formal schon die Forderung $\forall v, w : v \rightarrow^* w$ reichen würde.)
- und zusätzlich V' maximal ist mit dieser Eigenschaft ist, d.h. für jedes $u \in V$ gilt: Falls $v \rightarrow^* u$ und $u \rightarrow^* v$, dann gehört u bereits zu V' .

Darstellungsformen

Bei der Darstellung von *ungerichteten* Graphen $G = (V, E)$ bietet es sich an, diese als *doppelt* gerichtete Graphen aufzufassen. Man stellt also statt $G = (V, E)$ den gerichteten Graphen $G' = (V, E')$ dar, bei dem

$$E' := \{(v, w) \mid \{v, w\} \in E\} \cup \{(w, v) \mid \{v, w\} \in E\}.$$

Tiefensuche

In dem Algorithmus zur topologischen Sortierung und im Tarjan-Algorithmus soll die Formulierung

(*) "Wähle w aus $U \dots$ "

nicht etwa bedeuten, dass wir es mit einem nicht-deterministischen Verfahren zu tun haben, sondern dass es für den korrekten Ablauf des Verfahrens nicht entscheidend ist, welcher der möglichen Knoten gewählt wird. Falls gewünscht, könnte man etwa so vorgehen: Man nummeriere die Knoten v_1, \dots, v_n durch, und ersetze (*) durch:

“Wähle den Knoten v_i aus U mit kleinstem Index i aus ...”

Beim topologischen Sortieren wird durch eine spezielle Auswahl möglicherweise auch die Reihenfolge in der Ausgabeliste beeinflusst, aber in jedem Fall eine korrekte Sortierung ausgegeben.

Bei dem Tarjan-Algorithmus hängt die spezifische Zerlegung der Knotenmenge in Bäume sicherlich von der Wahl der w ab, aber das Resultat, die einzelnen Zusammenhangskomponenten, dagegen nicht.

Breitensuche

In dem Algorithmus 2.25 zur Breitensuche kann die Terminierung im Fall, dass t nicht aus s erreichbar ist, gewährleistet werden, indem Zeile (1) ersetzt wird durch

while $S \neq \emptyset$ **do**

Der Dijkstra-Algorithmus

Der Korrektheitsbeweis:

Beweis. Wir zeigen die folgenden drei Behauptungen durch verschachtelte Induktion über den Aufbau der Knotenmenge R .

- (a) Für $v \in R$, $y \in V - R$ gilt stets $l(v) \leq l(y)$.
- (b) Für jeden Knoten $v \in R$ gibt es einen kürzesten Weg $\pi : s \rightarrow v$, der nur aus Knoten in R besteht, so dass

$$w(\pi) = l(v)$$

und

die letzte Kante von π ist $(p(v), v)$.

- (c) Für jeden Knoten $v \in V - R$ gibt es einen kürzesten Weg $\pi : s \rightarrow v$, der nur aus Knoten in $R \cup \{v\}$ besteht, so dass

$$w(\pi) = l(v), \quad p(v) \in R, \quad l(v) = l(p(v)) + w(p(v), v).$$

Hierbei ist (b) die zentrale Behauptung. Sobald nämlich der Zielknoten z in R eingefügt worden ist, folgt aus (b) nämlich, dass π der gesuchte kürzeste Weg von s nach z ist.

Wir kommen zum Induktionsbeweis. Die Behandlung des *Induktionsanfangs* vor der ersten Ausführung der while-Schleife sei dem Leser überlassen.

Wir nehmen jetzt als *Induktionsannahme* (IA) an, dass (a), (b), (c) bei Eintritt in den nächsten Scheifendurchlauf für die bereits festgelegten Werte von R , l und p gelten.

In der Schleife werden R und möglicherweise auch l verändert. Zur Unterscheidung schreiben wir die Werte nach Durchlauf als R' und l' .

Im *Induktionsschritt* (IS) zeigen wir zunächst, dass unter diesen Voraussetzungen die Behauptung (a) auch für $v \in R'$, $y \in V - R'$ gilt.

Wir erhalten

$$\begin{aligned} l'(v) &= l(v) \\ &\leq l(u) && \text{nach IA (a), da } v \in R, u \in V - R \\ &\leq l'(y) && \text{nach (1), (2) der while-Schleife.} \end{aligned}$$

Wir kommen zum IS (b) für R' . Es reicht natürlich, die Behauptung für den neu hinzugekommenen Knoten u zu beweisen.

Weiterhin reicht es wegen der IA (c) für R zu zeigen, dass

$$(\bullet) \text{ Es existiert kein Weg } \pi : s = \underbrace{v_1 \rightarrow \dots \rightarrow v_n}_{\in R} \rightarrow \underbrace{y}_{\notin R} \rightarrow \dots \rightarrow u \text{ mit } w(\pi) < l(u).$$

Um die Behauptung (\bullet) zu zeigen, nehmen wir an, ein solcher Weg existiert *doch*, und führen diese Annahme zum Widerspruch. Gäbe es nämlich einen solchen Weg, würde nach IA (c) angewandt auf y für das Anfangsstück $\pi' = v_1 \rightarrow \dots \rightarrow v_n \rightarrow y$ von π gelten:

$$l(y) \leq w(\pi') \leq w(\pi) < l(u).$$

Das steht aber ein Widerspruch zur Wahl von u in (1) der Schleife; (\bullet) gilt also.

Wir kommen schließlich zum IS (c). Betrachte ein $v \in V - R'$ mit $l(v) \neq \infty$. Falls keine Kante von u nach v existiert, wird $l(v)$ im Schleifendurchlauf nicht verändert, und die Behauptung ist bereits von der IA (c) abgedeckt.

Es bleibt der Fall, dass $(u, v) \in E$. Da wir den IS für (b) bereits vollzogen haben, können wir voraussetzen, dass (b) für $R' = R \cup \{u\}$ gilt. D.h. aber, es existiert ein Weg

$$(*) \quad \pi : s \rightarrow \dots \rightarrow u \rightarrow v \text{ mit } w(\pi) = l(u) + w(u, v).$$

Wir zeigen, dass π die Forderung (c) für R' erfüllt.

Dazu nehmen wir dagegen an, dass es innerhalb von $R' \cup \{v\}$ doch einen Weg ρ gibt der Form

$$(**) \quad \rho : s \rightarrow v \text{ mit } w(\rho) < w(\pi).$$

Diese Annahme soll zum Widerspruch geführt werden. Zunächst zeigen wir, dass in einem solchen ρ gelten müsste:

$$(* * *) \quad u \text{ kommt in } \rho \text{ vor.}$$

Dazu nehmen wir an, u kommt in ρ nicht vor, und führen diese Annahme ihrerseits zum Widerspruch. Wenn u in ρ nicht vorkommt, ist ρ ein Weg ganz innerhalb von $R \cup \{v\}$. Dann bekommen wir aber

$$\begin{array}{ll}
 l(v) \leq w(\rho) & \text{nach IA (c)} \\
 < w(\pi) & \text{wegen (**)} \\
 \leq l'(v) & \text{wegen (*)} \\
 \leq l(v) & \text{da die } l\text{-Werte höchstes abnehmen.}
 \end{array}$$

Daher kommt u in ρ vor; es gilt (* * *).

Sei jetzt x der Vorgänger von v in (zur Erinnerung: diesem *gedachten*) ρ , also insbesondere $x \in R$. Nach IA (a) für R gilt wegen $u \in R$:

$$(\dagger) \quad l(x) \leq l(u).$$

Wieder unter Beachtung, dass l -Werte nur kleiner werden können:

$$\begin{array}{ll}
 l'(v) \leq l(x) + w(x, v) & \text{Übung} \\
 \leq l(u) + w(x, v) & \text{wegen (\dagger)} \\
 \leq w(\rho) & \text{da der Knoten } u \text{ und die Kante } (x, v) \text{ in } \rho \text{ vorkommen} \\
 < w(\pi) & \text{wegen der Annahme (**)} \\
 \leq l'(v) & \text{wegen (*)}.
 \end{array}$$

Wir haben also wieder einen Widerspruch; ein Weg ρ wie in (**) kann nicht existieren. Daher ist (*) tatsächlich der gesuchte kürzeste Weg. Wenn wir jetzt noch $p(v) := u$ setzen, ist der IS (c) nachgewiesen und wir sind fertig. \square

Bipartite Graphen und das Matching-Problem

In Ergänzung zu Definition N23(1) verlangen wir:

Definition. (1) Ein ungerichteter Graph $G = (V, E)$ heißt *bipartit*, falls gilt:

- V hat die Form $V_L \cup V_R$ mit $V_L \cap V_R = \emptyset$. (V_L und V_R bilden also eine disjunkte Zerlegung der Knotenmenge V .)
- Für jede Kante $e = \{v, w\} \in E$ ist

$$e \cap V_L \neq \emptyset \neq e \cap V_R.$$

- V hat keine isolierten Elemente, d.h. für jeden Knoten $v \in V$ gibt es eine Kante $e \in E$ mit $v \in e$.

Erst die letzte Bedingung gewährleistet, dass in dem zugehörigen Netzwerk N_G tatsächlich das Kirchhoffsche Gesetz gilt.

Bei der Definition des zu einem bipartiten Graphen G zugehörigen Netzwerks $N_G = (G', s, t, c)$ werden die Kanten von V_L nach V_R gerichtet, so dass sich in partieller Korrektur zur Vorlesung ergibt, dass $G' := (V', E')$ mit $V' = V \cup \{s, t\}$, und (geändert!):

$$\begin{aligned} E' := & \{(v, w) \mid \{v, w\} \in E, v \in V_L, w \in V_R\} \\ & \cup \{(s, v) \mid v \in V_L\} \\ & \cup \{(v, t) \mid v \in V_R\}. \end{aligned}$$

Numerische Grundbegriffe

Neu!

Aus mehreren Rückmeldungen ist hervorgegangen, dass eine zusätzliche Erläuterung einiger numerischer Grundbegriffe nützlich sein könnte.

Kondition und Stabilität

Zur Unterscheidung ist es wichtig, sich klarzumachen, dass sich der Begriff *Kondition* auf ein Problem selbst bezieht, also auf das Berechnen einer Funktion $y = f(x)$. Die Frage nach *Stabilität* bezieht sich dagegen auf ein konkretes Berechnungsverfahren.

Das Aufgabe sei etwa die Berechnung von $y = 1 - \frac{1}{1+x}$ für $|x| \ll 1$.

Die Kondition des Problems wird ausgedrückt durch die absolute Konditionszahl

$$\kappa_{\text{abs}} = |f'(x)| = \frac{1}{(1+x)^2}$$

bzw. die relative Konditionszahl

$$\kappa_{\text{rel}} = \left| \frac{f'(x)x}{f(x)} \right| = \frac{1}{1+x}.$$

Die Konditionszahlen geben an, wie sich eine Störung in den Eingabedaten auf das Resultat auswirkt. In dem Beispiel sieht man, dass für $|x| \ll 1$ die Nenner in κ_{abs} bzw. κ_{rel} beide $\gg 0$ sind, κ_{abs} und κ_{rel} also klein bleiben, die Wirkung einer fehlerhaften Eingabe sich daher in Grenzen hält.

Die Stabilität bezieht sich dagegen auf konkrete Verfahren, z.B. auf die Berechnung von $f(x)$ durch

$$(a) \quad \xi_1 = 1 + x, \quad \xi_2 := 1/\xi_1, \quad y := \xi_3 = 1 - \xi_2$$

oder

$$(b) \quad \xi_1 := 1 + x, \quad y := \xi_2 := x/\xi_1.$$

Bei exakter Rechnung liefern beide dasselbe korrekte Resultat. Bei tatsächlicher Ausführung besteht bei (a) im letzten Schritt die Gefahr der Auslöschung, die bei (b) vermieden wird. Unter Stabilitätsgesichtspunkten ist also (b) vorzuziehen.

Eine enge Beziehung zwischen Kondition und Stabilität besteht dennoch. Fehler, die im Laufe einer Berechnung auftreten, können natürlich als gestörte Eingabe für den nächsten Schritt aufgefasst werden, die Fehlerfortpflanzung daher unter Beachtung der Konditionszahl der nun anstehenden Operation abgeschätzt werden.

Gleitkommaarithmetik

Führt man Gleitkommaarithmetik per Hand aus, kann man natürlich auch ohne Mantissen/Exponent-Darstellung auskommen.

Die einzigen Regeln, die (hier am Beispiel 3-stelliger Gleitkommaarithmetik zur Basis 10) zu beachten sind, sind:

- Nach jeder arithmetischen Operation wird gerundet, so dass jedes Zwischenergebnis die erste von Null verschiedene Ziffer und die beiden folgenden Dezimalstellen, von denen die letzte gerundet ist, enthält.
- Entsprechend werden Eingabedaten auf 3 Stellen gerundet.

Beispiel. Wir erhalten etwa

$$123 \oplus 1.9 = \text{rd}(124.9) = 125$$

und

$$0.00123 \oplus 0.000019 = \text{rd}(0.001249) = 0.00125.$$

Wenn es der Lesbarkeit dient, kann man natürlich $1.25 \cdot 10^{-3}$ oder $0.125 \cdot 10^{-2}$ statt 0.00125 schreiben.

Beispiel. Noch ein Beispiel: Der Umfang u eines Kreises mit Radius $\sqrt{2}$ soll in 3-stelliger Gleitkommaarithmetik berechnet werden. Mit den Eingabedaten π gerundet auf 3.14 und $\sqrt{2}$ auf 1.41 ergibt sich

$$u = 2 \odot 3.14 \odot 1.41.$$

Wir werten den Ausdruck von links nach rechts aus:

$$(2 \odot 3.14) \odot 1.41 = \text{rd}(6.28) \odot 1.41 = 6.28 \odot 1.41 = \text{rd}(8.8548) = 8.85.$$

Bemerkung. Das letzte Beispiel illustriert auch, dass die Gleitkommaoperationen i.a. nicht assoziativ sind: Auswertung von rechts nach links ergibt

$$2 \odot (3.14 \odot 1.41) = 2 \odot \text{rd}(4.4274) = 2 \odot 4.43 = \text{rd}(8.86) = 8.86.$$

Beispiel. Als größeres Beispiel kann man etwa das lineare Gleichungssystem

$$\begin{pmatrix} 2.00 & 1.01 & 2.52 \\ 0.40 & 0.203 & -1.80 \\ 1.20 & -1.10 & 1.60 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 9.57 \\ -0.385 \\ -7.70 \end{pmatrix}$$

unter Benutzung der Gaußschen Elimination und 3-stelliger Gleitkommaarithmetik lösen, (1) ohne Vertauschung von Zeilen oder Spalten, oder auch (2) mit Spalten-Pivotwahl.

Normalisierte Darstellung

Die Betrachtung der Mantisse/Exponent-Darstellung wird erst dann wirklich wichtig, wenn man sich mit dem numerischen Gitter und der Zahlendarstellung im Rechner beschäftigt.

In der Vorlesung hatten wir eine Mantisse $m_1m_2 \dots m_t$ als die Zahl $0.m_1m_2m_t$ interpretiert.

Häufig findet man aber auch die Konvention, dass die Mantisse den Zahlenwert $m_1.m_2 \dots m_t$ repräsentiert. Letzlich sind natürlich beide Festlegungen gleichwertig; zur Umrechnung muss man nur den Exponenten in der Darstellung um 1 verändern.

Die *normalisierte Darstellung* fordert aber in jedem Fall, dass $m_1 \neq 0$. Festzulegen ist eben nur, ob damit $m_1.0$ oder $0.m_1$ gemeint ist.

Diese unterschiedliche Festlegung zieht sich bis in die Interpretation der Norm IEEE 754 hinein. Bei der $0.m_1$ -Auffassung laufen die Exponenten im doppelten Grundformat von $e_{\min} = -1021$ bis $e_{\max} = -1024$, bei der $m_1.0$ -Auffassung dagegen von $e_{\min} = -1022$ bis $e_{\max} = -1023$. Der darstellbare Bereich ist also identisch, der Unterschied der Auffassungen von außen gar nicht sichtbar.