

Vorlesungsskript
Theoretische Informatik III
Sommersemester 2008

Prof. Dr. Johannes Köbler
Humboldt-Universität zu Berlin
Lehrstuhl Komplexität und Kryptografie

10. Juli 2009

Inhaltsverzeichnis

1	Einleitung	1
2	Suchen und Sortieren	2
2.1	Suchen von Mustern in Texten	2
2.1.1	String-Matching mit endlichen Automaten . .	3
2.1.2	Der Knuth-Morris-Pratt-Algorithmus	4
2.2	Durchsuchen von Zahlenfolgen	6
2.3	Sortieralgorithmen	7
2.3.1	Sortieren durch Einfügen	7
2.3.2	Sortieren durch Mischen	8
2.3.3	Lösen von Rekursionsgleichungen	10
2.3.4	Eine untere Schranke für das Sortierproblem .	10
2.3.5	QuickSort	11
2.3.6	HeapSort	14
2.3.7	BucketSort	16
2.3.8	CountingSort	16
2.3.9	RadixSort	17
2.3.10	Vergleich der Sortierverfahren	17
2.4	Datenstrukturen für dynamische Mengen	18
2.4.1	Verkettete Listen	18
2.4.2	Binäre Suchbäume	19
2.4.3	Balancierte Suchbäume	20
3	Graphalgorithmen	24
3.1	Grundlegende Begriffe	24
3.2	Datenstrukturen für Graphen	26
3.3	Keller und Warteschlange	26

3.4	Durchsuchen von Graphen	28
3.4.1	Suchwälder	29
3.4.2	Klassifikation der Kanten eines (Di-)Graphen	29
3.4.3	Spannbäume und Spannwälder	31
3.4.4	Berechnung der Zusammenhangskomponenten	32
3.4.5	Breiten- und Tiefensuche	32
3.4.6	Starke Zusammenhangskomponenten	36
3.5	Kürzeste Pfade in Distanzgraphen	39
3.5.1	Der Dijkstra-Algorithmus	39
3.6	Negative Kantengewichte	42
3.6.1	Der Bellman-Ford-Algorithmus	42
3.6.2	Der Bellman-Ford-Moore-Algorithmus	43
3.7	Berechnung von allen kürzesten Wegen	45
3.7.1	Der Floyd-Warshall-Algorithmus	45
3.8	Flüsse in Netzwerken	46

1 Einleitung

In den Vorlesungen ThI 1 und ThI 2 standen folgende Themen im Vordergrund:

- Mathematische Grundlagen der Informatik, Beweise führen, Modellierung (Aussagenlogik, Prädikatenlogik)
- Welche Probleme sind lösbar? (Berechenbarkeitstheorie)
- Welche Rechenmodelle sind adäquat? (Automatentheorie)
- Welcher Aufwand ist nötig? (Komplexitätstheorie)

Dagegen geht es in der VL ThI 3 in erster Linie um folgende Frage:

- Wie lassen sich eine Reihe von praktisch relevanten Problemstellungen möglichst effizient lösen?
- Wie lässt sich die Korrektheit von Algorithmen beweisen und wie lässt sich ihre Laufzeit abschätzen?

Die Untersuchung dieser Fragen lässt sich unter dem Themengebiet **Algorithmik** zusammenfassen.

Der Begriff *Algorithmus* geht auf den persischen Gelehrten **Muhammed Al Chwarizmi** (8./9. Jhd.) zurück. Der älteste bekannte nicht-triviale Algorithmus ist der nach *Euklid* benannte Algorithmus zur Berechnung des größten gemeinsamen Teilers zweier natürlicher Zahlen (300 v. Chr.). Von einem Algorithmus wird erwartet, dass er jede *Problemeingabe* nach endlich vielen Rechenschritten löst (etwa durch Produktion einer **Ausgabe**). Ein Algorithmus ist ein „Verfahren“ zur Lösung eines Entscheidungs- oder Berechnungsproblems, das sich prinzipiell auf einer Turingmaschine implementieren lässt (**Church-Turing-These**).

Die Registermaschine

Bei Aussagen zur Laufzeit von Algorithmen beziehen wir uns auf die Registermaschine (engl. random access machine; RAM). Dieses Modell ist etwas flexibler als die Turingmaschine, da es den unmittelbaren Lese- und Schreibzugriff (**random access**) auf eine beliebige Speicherinheit (Register) erlaubt. Als Speicher stehen beliebig viele Register zur Verfügung, die jeweils eine beliebig große natürliche Zahl speichern können. Auf den Registerinhalten sind folgende arithmetische Operationen in einem Rechenschritt ausführbar: Addition, Subtraktion, abgerundetes Halbieren und Verdoppeln. Unabhängig davon geben wir die Algorithmen in Pseudocode an. Das RAM-Modell benutzen wir nur zur Komplexitätsabschätzung.

Die Laufzeit von RAM-Programmen wird wie bei TMs in der Länge der Eingabe gemessen. Man beachte, dass bei arithmetischen Problemen (wie etwa Multiplikation, Division, Primzahltests, etc.) die Länge einer Zahleingabe n durch die Anzahl $\lceil \log n \rceil$ der für die **Binärkodierung** von n benötigten Bits gemessen wird. Dagegen bestimmt bei nicht-arithmetischen Problemen (z.B. Graphalgorithmen oder Sortierproblemen) die Anzahl der gegebenen Zahlen die Länge der Eingabe.

Asymptotische Laufzeit und Landau-Notation

Definition 1. Seien f und g Funktionen von $\mathbb{N}$ nach $\mathbb{R}^+$. Wir schreiben $f(n) = \mathcal{O}(g(n))$, falls es Zahlen n_0 und c gibt mit

$$\forall n \geq n_0 : f(n) \leq c \cdot g(n).$$

Die Bedeutung der Aussage $f(n) = \mathcal{O}(g(n))$ ist, dass f „**nicht wesentlich schneller**“ als g wächst. Formal bezeichnet der Term $\mathcal{O}(g(n))$ die Klasse aller Funktionen f , die obige Bedingung erfüllen. Die Gleichung $f(n) = \mathcal{O}(g(n))$ drückt also in Wahrheit eine **Element-Beziehung** $f \in \mathcal{O}(g(n))$ aus. $\mathcal{O}$ -Terme können auch auf

der linken Seite vorkommen. In diesem Fall wird eine **Inklusionsbeziehung** ausgedrückt. So steht $n^2 + \mathcal{O}(n) = \mathcal{O}(n^2)$ für die Aussage $\{n^2 + f \mid f \in \mathcal{O}(n)\} \subseteq \mathcal{O}(n^2)$.

Beispiel 2.

- $7 \log(n) + n^3 = \mathcal{O}(n^3)$ ist **richtig**.
- $7 \log(n)n^3 = \mathcal{O}(n^3)$ ist **falsch**.
- $2^{n+\mathcal{O}(1)} = \mathcal{O}(2^n)$ ist **richtig**.
- $2^{\mathcal{O}(n)} = \mathcal{O}(2^n)$ ist **falsch** (siehe Übungen).

◁

Es gibt noch eine Reihe weiterer nützlicher Größenvergleiche von Funktionen.

Definition 3. Wir schreiben $f(n) = o(g(n))$, falls es für jedes $c > 0$ eine Zahl n_0 gibt mit

$$\forall n \geq n_0 : f(n) \leq c \cdot g(n).$$

Damit wird ausgedrückt, dass f „wesentlich langsamer“ als g wächst. Außerdem schreiben wir

- $f(n) = \Omega(g(n))$ für $g(n) = \mathcal{O}(f(n))$, d.h. f wächst **mindestens so schnell** wie g)
- $f(n) = \omega(g(n))$ für $g(n) = o(f(n))$, d.h. f wächst **wesentlich schneller** als g , und
- $f(n) = \Theta(g(n))$ für $f(n) = \mathcal{O}(g(n)) \wedge f(n) = \Omega(g(n))$, d.h. f und g wachsen **ungefähr gleich schnell**.

2 Suchen und Sortieren

2.1 Suchen von Mustern in Texten

In diesem Abschnitt betrachten wir folgende algorithmische Problemstellung.

String-Matching (STRINGMATCHING):

Gegeben: Ein Text $x = x_1 \cdots x_n$ und ein Muster $y = y_1 \cdots y_m$ über einem Alphabet Σ .

Gesucht: Alle Vorkommen von y in x .

Wir sagen y kommt in x an Stelle i vor, falls $x_{i+1} \cdots x_{i+m} = y$ ist. Typische Anwendungen finden sich in Textverarbeitungssystemen (emacs, grep, etc.), sowie bei der DNS- bzw. DNA-Sequenzanalyse.

Beispiel 4. Sei $\Sigma = \{A, C, G, U\}$.

Text $x = \text{AUGACGAUGAUGUAGGUAGCGUAGAUGAUGUAG}$,
Muster $y = \text{AUGAUGUAG}$.

Das Muster y kommt im Text x an den Stellen **6** und **24** vor. ◁

Bei naiver Herangehensweise kommt man sofort auf folgenden Algorithmus.

Algorithmus naive-String-Matcher(x, y)

```

1  Input: Text  $x = x_1 \cdots x_n$  und Muster  $y = y_1 \cdots y_m$ 
2     $V := \emptyset$ 
3  for  $i := 0$  to  $n - m$  do
```

```

4   if  $x_{i+1} \cdots x_{i+m} = y_1 \cdots y_m$  then
5        $V := V \cup \{i\}$ 
6   Output:  $V$ 

```

Die Korrektheit von **naive-String-Matcher** ergibt sich wie folgt:

- In der **for**-Schleife testet der Algorithmus alle potentiellen Stellen, an denen y in x vorkommen kann, und
- fügt in Zeile 4 genau die Stellen i zu V hinzu, für die $x_{i+1} \cdots x_{i+m} = y$ ist.

Die Laufzeit von **naive-String-Matcher** lässt sich nun durch folgende Überlegungen abschätzen:

- Die **for**-Schleife wird $(n - m + 1)$ -mal durchlaufen.
- Der Test in Zeile 4 benötigt maximal m Vergleiche.

Dies führt auf eine Laufzeit von $\mathcal{O}(nm) = \mathcal{O}(n^2)$. Für Eingaben der Form $x = a^n$ und $y = a^{\lfloor n/2 \rfloor}$ ist die Laufzeit tatsächlich $\Theta(n^2)$.

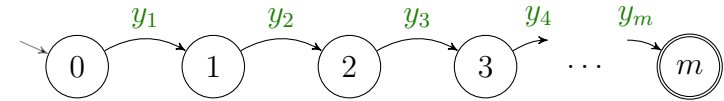
2.1.1 String-Matching mit endlichen Automaten

Durch die Verwendung eines endlichen Automaten lässt sich eine erhebliche Effizienzsteigerung erreichen. Hierzu konstruieren wir einen DFA M_y , der jedes Vorkommen von y in der Eingabe x durch Erreichen eines Endzustands anzeigt. M_y erkennt also die Sprache

$$L = \{x \in \Sigma^* \mid y \text{ ist Suffix von } x\}.$$

Konkret konstruieren wir $M_y = (Z, \Sigma, \delta, 0, m)$ wie folgt:

- M_y hat $m + 1$ Zustände, die den $m + 1$ Präfixen $y_1 \cdots y_k$, $k = 0, \dots, m$, von y entsprechen, d.h. $Z = \{0, \dots, m\}$.
- Liest M_y im Zustand k das Zeichen y_{k+1} , so wechselt M_y in den Zustand $k + 1$, d.h. $\delta(k, y_{k+1}) = k + 1$ für $k = 0, \dots, m - 1$:



- Falls das nächste Zeichen a nicht mit y_{k+1} übereinstimmt (engl. *mismatch*), wechselt M_y in den Zustand

$$\delta(k, a) = \max\{j \leq m \mid y_1 \cdots y_j \text{ ist Suffix von } y_1 \cdots y_k a\}.$$

Der DFA M_y speichert also in seinem Zustand die maximale Länge k eines Präfixes $y_1 \cdots y_k$ von y , das zugleich ein Suffix der gelesenen Eingabe ist:

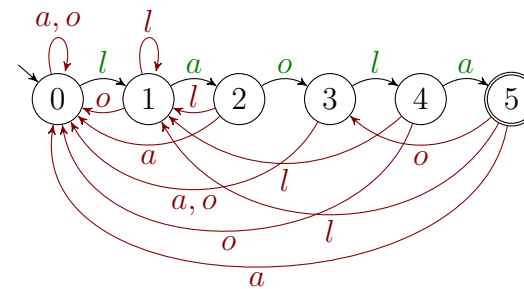
$$\hat{\delta}(0, x) = \max\{k \leq m \mid y_1 \cdots y_k \text{ ist Suffix von } x\}.$$

Die Korrektheit von M_y folgt aus der Beobachtung, dass M_y isomorph zum Äquivalenzklassenautomaten M_{R_L} für L ist. M_{R_L} hat die Zustände $[y_1 \cdots y_k]$, $k = 0, \dots, m$, von denen nur $[y_1 \cdots y_m]$ ein Endzustand ist. Die Überföhrungsfunktion ist definiert durch

$$\delta([y_1 \cdots y_k], a) = [y_1 \cdots y_j],$$

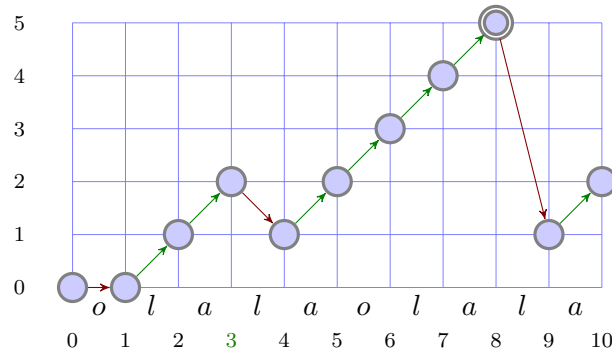
wobei $y_1 \cdots y_j$ das längste Präfix von $y = y_1 \cdots y_m$ ist, welches Suffix von $y_1 \cdots y_k a$ ist (siehe Übungen).

Beispiel 5. Für das Muster $y = laola$ hat M_y folgende Gestalt:



δ	0	1	2	3	4	5
a	0	2	0	0	5	0
l	1	1	1	4	1	1
o	0	0	3	0	0	3

M_y macht bei der Suche nach dem Muster $y = \text{laola}$ im Text $x = \text{olalaolala}$ folgende Übergänge:



◁

Insgesamt erhalten wir somit folgenden Algorithmus.

Algorithmus DFA-String-Matcher(x, y)

```

1 Input: Text  $x = x_1 \cdots x_n$  und Muster  $y = y_1 \cdots y_m$ 
2   konstruiere den DFA  $M_y = (Z, \Sigma, \delta, 0, m)$ 
3    $V := \emptyset$ 
4    $k := 0$ 
5   for  $i := 1$  to  $n$  do
6      $k := \delta(k, x_i)$ 
7     if  $k = m$  then  $V := V \cup \{i - m\}$ 
8 Output:  $V$ 

```

Die Korrektheit von **DFA-String-Matcher** ergibt sich unmittelbar aus der Tatsache, dass M_y die Sprache

$$L(M_y) = \{x \in \Sigma^* \mid y \text{ ist Suffix von } x\}$$

erkennt. Folglich fügt der Algorithmus genau die Stellen $j = i - m$ zu V hinzu, für die y ein Suffix von $x_1 \cdots x_i$ (also $x_{j+1} \cdots x_{j+m} = y$) ist.

Die Laufzeit von **DFA-String-Matcher** ist die Summe der Laufzeiten für die Konstruktion von M_y und für die Simulation von M_y bei Eingabe x , wobei letztere durch $\mathcal{O}(n)$ beschränkt ist. Für δ ist eine Tabelle mit $(m + 1)\|\Sigma\|$ Einträgen

$$\delta(k, a) = \max\{j \leq k + 1 \mid y_1 \cdots y_j \text{ ist Suffix von } y_1 \cdots y_k a\}$$

zu berechnen. Jeder Eintrag $\delta(k, a)$ ist in Zeit $\mathcal{O}(k^2) = \mathcal{O}(m^2)$ berechenbar. Dies führt auf eine Laufzeit von $\mathcal{O}(\|\Sigma\|m^3)$ für die Konstruktion von M_y und somit auf eine Gesamtlaufzeit von $\mathcal{O}(\|\Sigma\|m^3 + n)$. Tatsächlich lässt sich M_y sogar in Zeit $\mathcal{O}(\|\Sigma\|m)$ konstruieren.

2.1.2 Der Knuth-Morris-Pratt-Algorithmus

Durch eine Modifikation des Rücksprungmechanismus' lässt sich die Laufzeit von **DFA-String-Matcher** auf $\mathcal{O}(n + m)$ verbessern. Hierzu vergegenwärtigen wir uns folgende Punkte:

- Tritt im Zustand k ein Mismatch $a \neq y_{k+1}$ auf, so ermittelt M_y das längste Präfix p von $y_1 \cdots y_k$, das zugleich Suffix von $y_1 \cdots y_k a$ ist, und springt in den Zustand $j = \delta(k, a) = |p|$.
- Im Fall $j > 0$ hat p also die Form $p = p'a$, wobei $p' = y_1 \cdots y_{j-1}$ sowohl echtes Präfix als auch echtes Suffix von $y_1 \cdots y_k$ ist. Zudem gilt $y_j = a$.
- Die Idee beim KMP-Algorithmus ist nun, bei einem Mismatch unabhängig von a auf das nächst kleinere Präfix $\tilde{p} = y_1 \cdots y_i$ von $y_1 \cdots y_k$ zu springen, das auch Suffix von $y_1 \cdots y_k$ ist.
- Stimmt nach diesem Rücksprung das nächste Eingabezeichen a mit y_{i+1} überein, so wird dieses gelesen und der KMP-Algorithmus erreicht (nach einem kleinen Umweg über den Zustand i) den Zustand $i + 1 = j$, in den auch M_y wechselt.
- Andernfalls springt der KMP-Algorithmus nach derselben Methode solange weiter zurück, bis das nächste Eingabezeichen a

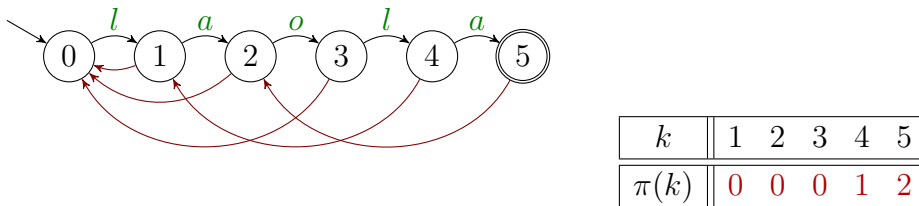
„passt“ (also $y_{i+1} = a$ und somit $\tilde{p}a$ ein Präfix von y ist) oder der Zustand 0 erreicht wird.

- In beiden Fällen wird a gelesen und der Zustand $\delta(k, a)$ angenommen.

Der KMP-Algorithmus besucht also alle Zustände, die auch M_y besucht, führt aber die Rücksprünge in mehreren Etappen aus. Die Sprungadressen werden durch die so genannte *Präfixfunktion* $\pi : \{1, \dots, m\} \rightarrow \{0, \dots, m-1\}$ ermittelt:

$$\pi(k) = \max\{0 \leq j \leq k-1 \mid y_1 \dots y_j \text{ ist Suffix von } y_1 \dots y_k\}.$$

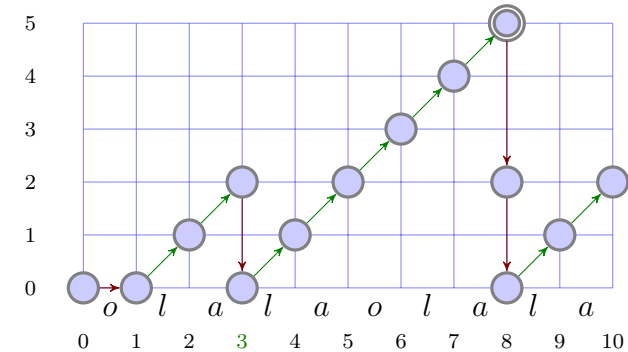
Beispiel 6. Für das Muster $y = \text{laola}$ ergibt sich folgende Präfixfunktion π :



Wir können uns die Arbeitsweise dieses Automaten wie folgt vorstellen:

1. Erlaubt das nächste Eingabezeichen einen Übergang vom aktuellen Zustand k nach $k+1$, so führe diesen aus.
2. Ist ein Übergang nach $k+1$ nicht möglich und $k \geq 1$, so springe in den Zustand $\pi(k)$ ohne das nächste Zeichen zu lesen.
3. Andernfalls (d.h. $k = 0$ und ein Übergang nach 1 ist nicht möglich) lies das nächste Zeichen und bleibe im Zustand 0.

Der KMP-Algorithmus macht bei der Suche nach dem Muster $y = \text{laola}$ im Text $x = \text{olalaolala}$ folgende Übergänge:



Auf die Frage, wie sich die Präfixfunktion π möglichst effizient berechnen lässt, werden wir später zu sprechen kommen. Wir betrachten zunächst das Kernstück des KMP-Algorithmus, das sich durch eine leichte Modifikation von **DFA-String-Matcher** ergibt.

DFA-String-Matcher(x, y)

```

1  Input: Text  $x_1 \dots x_n$ 
    und Muster  $y_1 \dots y_m$ 
2  konstruiere  $M_y$ 
3   $V := \emptyset$ 
4   $k := 0$ 
5  for  $i := 1$  to  $n$  do
6     $k := \delta(k, x_i)$ 
7
8  if  $k = m$  then
9     $V := V \cup \{i - m\}$ 
10 Output:  $V$ 
```

KMP-String-Matcher(x, y)

```

1  Input: Text  $x_1 \dots x_n$  und
    Muster  $y_1 \dots y_m$ 
2   $\pi := \text{KMP-Prefix}(y)$ 
3   $V := \emptyset$ 
4   $k := 0$ 
5  for  $i := 1$  to  $n$  do
6    while ( $k > 0 \wedge x_i \neq y_{k+1}$ ) do
7       $k := \pi(k)$ 
8    if  $x_i = y_{k+1}$  then  $k := k + 1$ 
9    if  $k = m$  then
10      $V := V \cup \{i - m\}, k := \pi(k)$ 
11 Output:  $V$ 
```

Die Korrektheit des Algorithmus **KMP-String-Matcher** ergibt sich einfach daraus, dass er den Zustand m an genau den gleichen Textstellen besucht wie **DFA-String-Matcher**, und somit wie dieser alle Vorkommen von y im Text x findet.

Für die Laufzeitanalyse von **KMP-String-Matcher** (ohne die Berechnung von **KMP-Prefix**) stellen wir folgende Überlegungen an.

- Die Laufzeit ist proportional zur Anzahl der Zustandsübergänge.
- Bei jedem Schritt wird der Zustand um maximal Eins erhöht.
- Daher kann der Zustand nicht öfter verkleinert werden als er erhöht wird (*Amortisationsanalyse*).
- Es gibt genau n Zustandsübergänge, bei denen der Zustand erhöht wird bzw. unverändert bleibt.
- Insgesamt finden also höchstens $2n = \mathcal{O}(n)$ Zustandsübergänge statt.

Nun kommen wir auf die Frage zurück, wie sich die Präfixfunktion π effizient berechnen lässt. Die Aufgabe besteht darin, für jedes Präfix $y_1 \cdots y_i$, $i \geq 1$, das längste echte Präfix zu berechnen, das zugleich Suffix von $y_1 \cdots y_i$ ist.

Die Idee besteht nun darin, mit dem KMP-Algorithmus das Muster y im Text $y_2 \cdots y_m$ zu suchen. Dann liefert der beim Lesen von y_i erreichte Zustand k gerade das längste Präfix $y_1 \cdots y_k$, das zugleich Suffix von $y_2 \cdots y_i$ ist (d.h. es gilt $\pi(i) = k$). Zudem werden bis zum Lesen von y_i nur Zustände kleiner als i erreicht. Daher sind die π -Werte für alle bis dahin auszuführenden Rücksprünge bereits bekannt und π kann in Zeit $\mathcal{O}(m)$ berechnet werden.

Prozedur **KMP-Prefix**(y)

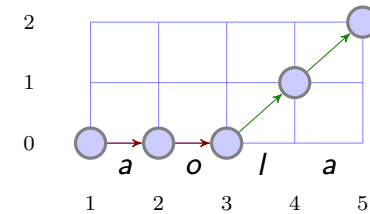
```

1   $\pi(1) := 0$ 
2   $k := 0$ 
3  for  $i := 2$  to  $m$  do
4      while ( $k > 0 \wedge y_i \neq y_{k+1}$ ) do  $k := \pi(k)$ 
5      if  $y_i = y_{k+1}$  then  $k := k + 1$ 
6       $\pi(i) := k$ 
7  return( $\pi$ )

```

Beispiel 7. Die Verarbeitung des Musters $y = \text{laola}$ durch

KMP-Prefix ergibt folgendes Ablaufprotokoll:



k	1	2	3	4	5
$\pi(k)$	0	0	0	1	2

◀

Wir fassen die Laufzeiten der in diesem Abschnitt betrachteten String-Matching Algorithmen in einer Tabelle zusammen:

Algorithmus	Vorverarbeitung	Suche	Gesamtlaufzeit
naiv	0	$\mathcal{O}(nm)$	$\mathcal{O}(nm)$
DFA (einfach)	$\mathcal{O}(\ \Sigma\ m^3)$	$\mathcal{O}(n)$	$\mathcal{O}(\ \Sigma\ m^3 + n)$
DFA (verbessert)	$\mathcal{O}(\ \Sigma\ m)$	$\mathcal{O}(n)$	$\mathcal{O}(\ \Sigma\ m + n)$
Knuth-Morris-Pratt	$\mathcal{O}(m)$	$\mathcal{O}(n)$	$\mathcal{O}(n)$

2.2 Durchsuchen von Zahlenfolgen

Als nächstes betrachten wir folgendes Suchproblem.

Element-Suche

Gegeben: Eine Folge $a_1, \dots, a_n$ von natürlichen Zahlen und eine Zahl a .

Gesucht: Ein Index i mit $a_i = a$ (bzw. eine Fehlermeldung, falls $a \notin \{a_1, \dots, a_n\}$ ist).

Typische Anwendungen finden sich bei der Verwaltung von Datensätzen, wobei jeder Datensatz über einen eindeutigen Schlüssel (z.B. *Matrikelnummer*) zugreifbar ist. Bei manchen Anwendungen können die Zahlen in der Folge auch mehrfach vorkommen. Gesucht sind dann

evtl. alle Indizes i mit $a_i = a$. Durch eine sequentielle Suche lässt sich das Problem in Zeit $\mathcal{O}(n)$ lösen.

Algorithmus Sequential-Search

```

1 Input: Eine Zahlenfolge  $a_1, \dots, a_n$  und eine Zahl  $a$ 
2    $i := 0$ 
3   repeat
4      $i := i + 1$ 
5   until  $(i = n \vee a = a_i)$ 
6 Output:  $i$ , falls  $a_i = a$  bzw. Fehlermeldung, falls
    $a_i \neq a$ 

```

Falls die Folge $a_1, \dots, a_n$ sortiert ist, d.h. es gilt $a_i \leq a_j$ für $i \leq j$, bietet sich eine *Binärsuche* an.

Algorithmus Binary-Search

```

1 Input: Eine Zahlenfolge  $a_1, \dots, a_n$  und eine Zahl  $a$ 
2    $l := 1$ 
3    $r := n$ 
4   while  $l < r$  do
5      $m := \lfloor (l + r) / 2 \rfloor$ 
6     if  $a \leq a_m$  then  $r := m$  else  $l := m + 1$ 
7 Output:  $l$ , falls  $a_l = a$  bzw. Fehlermeldung, falls
    $a_l \neq a$ 

```

Offensichtlich gibt der Algorithmus im Fall $a \notin \{a_1, \dots, a_n\}$ eine Fehlermeldung aus. Im Fall $a \in \{a_1, \dots, a_n\}$ gilt die *Schleifeninvariante* $a_l \leq a \leq a_r$. Daher muss nach Abbruch der **while**-Schleife $a = a_l$ sein. Dies zeigt die Korrektheit von **Binary-Search**.

Da zudem die Länge $l - r + 1$ des Suchintervalls $[l, r]$ in jedem Schleifendurchlauf mindestens auf $\lfloor (l - r) / 2 \rfloor + 1$ reduziert wird, werden höchstens $\lceil \log n \rceil$ Schleifendurchläufe ausgeführt. Folglich ist die Laufzeit von **Binary-Search** höchstens $\mathcal{O}(\log n)$.

2.3 Sortieralgorithmen

Wie wir im letzten Abschnitt gesehen haben, lassen sich Elemente in einer sortierten Folge sehr schnell aufspüren. Falls wir diese Operation öfters ausführen müssen, bietet es sich an, die Zahlenfolge zu sortieren.

Sortierproblem

Gegeben: Eine Folge $a_1, \dots, a_n$ von natürlichen Zahlen.

Gesucht: Eine Permutation $a_{i_1}, \dots, a_{i_n}$ dieser Folge mit $a_{i_j} \leq a_{i_{j+1}}$ für $j = 1, \dots, n - 1$.

Man unterscheidet *vergleichende Sortierverfahren* von den übrigen Sortierverfahren. Während erstere nur Ja-Nein-Fragen der Form „ $a_i \leq a_j$?“ oder „ $a_i < a_j$?“ stellen dürfen, können letztere auch die konkreten Zahlenwerte a_i der Folge abfragen. Vergleichsbasierte Verfahren benötigen im schlechtesten Fall $\Omega(n \log n)$ Vergleiche, während letztere unter bestimmten Zusatzvoraussetzungen sogar in Linearzeit arbeiten.

2.3.1 Sortieren durch Einfügen

Ein einfacher Ansatz, eine Zahlenfolge zu sortieren, besteht darin, sequentiell die Zahl a_i ($i = 2, \dots, n$) in die bereits sortierte Teilfolge $a_1, \dots, a_{i-1}$ einzufügen.

Algorithmus Insertion-Sort($a_1, \dots, a_n$)

```

1   for  $i := 2$  to  $n$  do  $z := a_i$ 
2      $j := i - 1$ 
3     while  $(j \geq 1 \wedge a_j > z)$  do
4        $a_{j+1} := a_j$ 
5        $j := j - 1$ 
6      $a_{j+1} := z$ 

```

Die Korrektheit von **Insertion-Sort** lässt sich induktiv durch den Nachweis folgender Schleifeninvarianten beweisen:

- Nach jedem Durchlauf der **for**-Schleife sind $a_1, \dots, a_i$ sortiert.
- Nach jedem Durchlauf der **while**-Schleife gilt $z < a_k$ für $k = j + 2, \dots, i$.

Zusammen mit der Abbruchbedingung der **while**-Schleife folgt hieraus, dass z in Zeile 5 an der jeweils richtigen Stelle eingefügt wird.

Da zudem die **while**-Schleife für jedes $i = 2, \dots, n$ höchstens $(i - 1)$ -mal ausgeführt wird, ist die Laufzeit von **Insertion-Sort** durch $\sum_{i=2}^n \mathcal{O}(i - 1) = \mathcal{O}(n^2)$ begrenzt.

Bemerkung 8.

- Ist die Eingabefolge $a_1, \dots, a_n$ bereits sortiert, so wird die **while**-Schleife niemals durchlaufen. Im besten Fall ist die Laufzeit daher $\sum_{i=2}^n \Theta(1) = \Theta(n)$.
- Ist die Eingabefolge $a_1, \dots, a_n$ dagegen absteigend sortiert, so wandert z in $i - 1$ Durchläufen der **while**-Schleife vom Ende an den Anfang der bereits sortierten Teilfolge $a_1, \dots, a_i$. Im schlechtesten Fall ist die Laufzeit also $\sum_{i=2}^n \Theta(i - 1) = \Theta(n^2)$.
- Bei einer zufälligen Eingabepermutation der Folge $1, \dots, n$ wird z im Erwartungswert in der Mitte der Teilfolge $a_1, \dots, a_i$ eingefügt. Folglich beträgt die (erwartete) Laufzeit im durchschnittlichen Fall ebenfalls $\sum_{i=2}^n \Theta(\frac{i-1}{2}) = \Theta(n^2)$.

2.3.2 Sortieren durch Mischen

Wir können eine Zahlenfolge auch sortieren, indem wir sie in zwei Teilfolgen zerlegen, diese durch rekursive Aufrufe sortieren und die sortierten Teilfolgen wieder zu einer Liste zusammenfügen.

Diese Vorgehensweise ist unter dem Schlagwort “Divide and Conquer” (auch “divide et impera”, also “teile und herrsche”) bekannt. Dabei wird ein Problem gelöst, indem man es

- in mehrere Teilprobleme aufteilt,

- die Teilprobleme rekursiv löst, und
- die Lösungen der Teilprobleme zu einer Gesamtlösung des ursprünglichen Problems zusammenfügt.

Die Prozedur **Mergesort**(A, l, r) sortiert ein Feld $A[l \dots r]$, indem sie

- es in die Felder $A[l \dots m]$ und $A[m + 1 \dots r]$ zerlegt,
- diese durch jeweils einen rekursiven Aufruf sortiert, und
- die sortierten Teilfolgen durch einen Aufruf der Prozedur **Merge**(A, l, m, r) zu einer sortierten Folge zusammenfügt.

Algorithmus Mergesort(A, l, r)

```

1  if  $l < r$  then
2     $m := \lfloor (l + r)/2 \rfloor$ 
3    Mergesort( $A, l, m$ )
4    Mergesort( $A, m + 1, r$ )
5    Merge( $A, l, m, r$ )

```

Die Prozedur **Merge**(A, l, m, r) mischt die beiden sortierten Felder $A[l \dots m]$ und $A[m + 1 \dots r]$ zu einem sortierten Feld $A[l \dots r]$.

Prozedur Merge(A, l, m, r)

```

1  allokiere Speicher fuer ein neues Feld  $B[l \dots r]$ 
2   $j := l$ 
3   $k := m + 1$ 
4  for  $i := l$  to  $r$  do
5    if  $j > m$  then
6       $B[i] := A[k]$ 
7       $k := k + 1$ 
8    else if  $k > r$  then
9       $B[i] := A[j]$ 
10      $j := j + 1$ 
11    else if  $A[j] \leq A[k]$  then
12       $B[i] := A[j]$ 

```

```

13     j := j + 1
14     else
15         B[i] := A[k]
16         k := k + 1
17     kopiere das Feld B[l...r] in das Feld A[l...r]
18     gib den Speicher fuer B wieder frei

```

Man beachte, dass **Merge** für die Zwischenspeicherung der gemischten Folge zusätzlichen Speicher benötigt. **Mergesort** ist daher kein “in place”-Sortiervverfahren, welches neben dem Speicherplatz für die Eingabefolge nur konstant viel zusätzlichen Speicher belegen darf. Zum Beispiel ist **Insertion-Sort** ein “in place”-Verfahren. Auch **Mergesort** kann als ein “in place”-Sortiervverfahren implementiert werden, falls die zu sortierende Zahlenfolge nicht als Array, sondern als mit Zeigern verkettete Liste vorliegt (hierzu muss allerdings auch noch die Rekursion durch eine Schleife ersetzt werden).

Unter der Voraussetzung, dass **Merge** korrekt arbeitet, können wir per Induktion über die Länge $n = r - l + 1$ des zu sortierenden Arrays die Korrektheit von **Mergesort** wie folgt beweisen:

$n = 1$: In diesem Fall tut **Mergesort** nichts, was offensichtlich korrekt ist.

$n \rightsquigarrow n + 1$: Um eine Folge der Länge $n + 1 \geq 2$ zu sortieren, zerlegt sie **Mergesort** in zwei Folgen der Länge höchstens n . Diese werden durch die rekursiven Aufrufe nach IV korrekt sortiert und von **Merge** nach Voraussetzung korrekt zusammengefügt.

Die Korrektheit von **Merge** lässt sich leicht induktiv durch den Nachweis folgender Invariante für die for-Schleife beweisen:

- Nach jedem Durchlauf enthält $B[l \dots i]$ die $i - l + 1$ kleinsten Elemente aus $A[l \dots m]$ und $A[m + 1 \dots r]$ in sortierter Reihenfolge.
- Hierzu wurden die ersten $j - 1$ Elemente von $A[l \dots m]$ und die ersten $k - 1$ Elemente von $A[m + 1 \dots r]$ nach B kopiert.

Nach dem letzten Durchlauf (d.h. $i = r$) enthält daher $B[l \dots r]$ alle $r - l + 1$ Elemente aus $A[l \dots m]$ und $A[m + 1 \dots r]$ in sortierter Reihenfolge, womit die Korrektheit von **Merge** bewiesen ist.

Um eine Schranke für die Laufzeit von **Mergesort** zu erhalten, schätzen wir zunächst die Anzahl $V(n)$ der Vergleiche ab, die **Mergesort** im schlechtesten Fall benötigt, um ein Feld $A[l \dots r]$ der Länge $n = r - l + 1$ zu sortieren. Offensichtlich erfüllt $V(n)$ die Rekursionsgleichung

$$V(n) = \begin{cases} 0, & \text{falls } n = 1, \\ V(\lfloor n/2 \rfloor) + V(\lceil n/2 \rceil) + M(n), & n \geq 2. \end{cases}$$

Dabei ist $M(n) = n - 1$ die Anzahl der Vergleiche, die **Merge** benötigt, um die beiden sortierten Felder $A[l \dots m]$ und $A[m + 1 \dots r]$ zu mischen. Falls n eine Zweierpotenz ist, erhalten wir also die Rekursion

$$V(1) = 0 \text{ und } V(n) = 2V(n/2) + n - 1, n \geq 2.$$

Für die Funktion $f(k) = V(2^k)$ gilt dann

$$f(0) = 0 \text{ und } f(k) = 2f(k - 1) + 2^k - 1, k \geq 1.$$

Aus den ersten Folgengliedern

$$\begin{aligned} f(0) &= 0, \\ f(1) &= 1, \\ f(2) &= 2 + 2^2 - 1 = 1 \cdot 2^2 + 1, \\ f(3) &= 2 \cdot 2^2 + 2 + 2^3 - 1 = 2 \cdot 2^3 + 1, \\ f(4) &= 2 \cdot 2 \cdot 2^3 + 2 + 2^4 - 1 = 3 \cdot 2^4 + 1 \end{aligned}$$

lässt sich vermuten, dass $f(k) = (k - 1) \cdot 2^k + 1$ ist. Dies lässt sich leicht durch Induktion über k verifizieren, so dass wir für V die Lösungsfunktion $V(n) = n \log_2 n - n + 1$ erhalten. Ist n keine Zweierpotenz, so können wir die Anzahl der Fragen durch $V(n) \leq V(n') \leq V(2n) = \mathcal{O}(V(n))$ abschätzen, wobei $n' < 2n$ die kleinste Zweierpotenz größer als n ist.

Da die Laufzeit $T(n)$ von **MergeSort** asymptotisch durch die Anzahl $V(n)$ der Vergleiche beschränkt ist, folgt $T(n) = \mathcal{O}(V(n))$.

Satz 9. *MergeSort ist ein vergleichendes Sortierverfahren mit einer Laufzeit von $\mathcal{O}(n \log n)$.*

2.3.3 Lösen von Rekursionsgleichungen

Im Allgemeinen liefert der “Divide and Conquer”-Ansatz einfach zu implementierende Algorithmen mit einfachen Korrektheitsbeweisen. Die Laufzeit $T(n)$ erfüllt dann eine Rekursionsgleichung der Form

$$T(n) = \begin{cases} \Theta(1), & \text{falls } n \text{ „klein“ ist,} \\ D(n) + \sum_{i=1}^{\ell} T(n_i) + C(n), & \text{sonst.} \end{cases}$$

Dabei ist $D(n)$ der Aufwand für das Aufteilen der Problemistanz und $C(n)$ der Aufwand für das Verbinden der Teillösungen. Um solche Rekursionsgleichungen zu lösen, kann man oft eine Lösung „raten“ und per Induktion beweisen. Mit Hilfe von Rekursionsbäumen lassen sich Lösungen auch „gezielt raten“. Eine asymptotische Abschätzung liefert folgender Hauptsatz der Laufzeitfunktionen (Satz von Akra & Bazzi).

Satz 10 (Mastertheorem). *Sei $T : \mathbb{N} \rightarrow \mathbb{N}$ eine Funktion der Form*

$$T(n) = \sum_{i=1}^{\ell} T(n_i) + f(n) \quad \text{mit } n_i \in \{\lfloor \alpha_i n \rfloor, \lceil \alpha_i n \rceil\},$$

wobei $0 < \alpha_i < 1$, $i = 1, \dots, \ell$, fest gewählte reelle Zahlen sind. Dann gilt im Fall $f(n) = \Theta(n^k)$ für ein $k \geq 0$:

$$T(n) = \begin{cases} \Theta(n^k), & \text{falls } \sum_{i=1}^{\ell} \alpha_i^k < 1, \\ \Theta(n^k \log n), & \text{falls } \sum_{i=1}^{\ell} \alpha_i^k = 1, \\ \Theta(n^c), & \text{falls } \sum_{i=1}^{\ell} \alpha_i^k > 1, \end{cases}$$

wobei c Lösung der Gleichung $\sum_{i=1}^{\ell} \alpha_i^c = 1$ ist.

Beispiel 11. *Die Anzahl $V(n)$ der Vergleiche von MergeSort erfüllt die Rekursion*

$$V(n) = V(\lfloor n/2 \rfloor) + V(\lceil n/2 \rceil) + n - 1,$$

d.h. $l = 2$, $\alpha_1 = \alpha_2 = 1/2$ und $f(n) = n - 1 = \Theta(n^k)$ für $k = 1$. Wegen $\sum_{i=1}^{\ell} \alpha_i^k = 1/2 + 1/2 = 1$ folgt daher $V(n) = \Theta(n \log n)$.

2.3.4 Eine untere Schranke für das Sortierproblem

Frage. *Wie viele Vergleichsfragen benötigt ein vergleichender Sortieralgorithmus A mindestens, um eine Folge $(a_1, \dots, a_n)$ von n Zahlen zu sortieren?*

Zur Beantwortung dieser Frage betrachten wir alle $n!$ Eingabefolgen $(a_1, \dots, a_n)$ der Form $(\pi(1), \dots, \pi(n))$, wobei $\pi \in S_n$ eine beliebige Permutation auf der Menge $\{1, \dots, n\}$ ist. Um diese Folgen korrekt zu sortieren, muss A solange Fragen der Form $a_i < a_j$ (bzw. $\pi(i) < \pi(j)$) stellen, bis höchstens noch eine Permutation $\pi \in S_n$ mit den erhaltenen Antworten konsistent ist. Damit A möglichst viele Fragen stellen muss, beantworten wir diese so, dass mindestens die Hälfte der verbliebenen Permutationen mit unserer Antwort konsistent ist (*Mehrheitsvotum*). Diese Antwortstrategie stellt sicher, dass nach i Fragen noch mindestens $n!/2^i$ konsistente Permutationen übrig bleiben. Daher muss A mindestens

$$\lceil \log_2(n!) \rceil = n \log_2 n - n \log_2 e + \frac{1}{2} \log n + \Theta(1) = n \log_2 n - \Theta(n)$$

Fragen stellen, um die Anzahl der konsistenten Permutationen auf Eins zu reduzieren.

Satz 12. *Ein vergleichendes Sortierverfahren benötigt mindestens $\lceil \log_2(n!) \rceil$ Fragen, um eine Folge $(a_1, \dots, a_n)$ von n Zahlen zu sortieren.*

Wir können das Verhalten von A auch durch einen Fragebaum B veranschaulichen, dessen Wurzel mit der ersten Frage von A markiert ist. Jeder mit einer Frage markierte Knoten hat zwei Kinder, die die Antworten ja und nein auf diese Frage repräsentieren. Stellt A nach Erhalt der Antwort eine weitere Frage, so markieren wir den entsprechenden Antwortknoten mit dieser Frage. Andernfalls gibt A eine Permutation π der Eingabefolge aus und der zugehörige Antwortknoten ist ein Blatt, das wir mit π markieren. Nun ist leicht zu sehen, dass die Tiefe von B mit der Anzahl $V(n)$ der von A benötigten Fragen im schlechtesten Fall übereinstimmt. Da jede Eingabefolge zu einem anderen Blatt führt, hat B mindestens $n!$ Blätter. Folglich können wir in B einen Pfad der Länge $\lceil \log_2(n!) \rceil$ finden, indem wir jeweils in den Unterbaum mit der größeren Blätterzahl verzweigen. Da also jedes vergleichende Sortierverfahren mindestens $\Omega(n \log n)$ Fragen benötigt, ist **Mergesort** asymptotisch optimal.

Korollar 13. *MergeSort ist ein vergleichendes Sortierverfahren mit einer im schlechtesten Fall asymptotisch optimalen Laufzeit von $\mathcal{O}(n \log n)$.*

2.3.5 QuickSort

Ein weiteres Sortierverfahren, das den “Divide and Conquer”-Ansatz benutzt, ist **QuickSort**. Im Unterschied zu **MergeSort** wird hier das Feld *vor* den rekursiven Aufrufen umsortiert. Als Folge hiervon bereitet die Zerlegung in Teilprobleme die Hauptarbeit, während das Zusammenfügen der Teillösungen trivial ist. Bei **MergeSort** ist es gerade umgekehrt.

Algorithmus QuickSort(A, l, r)

```

1  if  $l < r$  then  $m := \text{Partition}(A, l, r)$ 
2    QuickSort( $A, l, m - 1$ )
3    QuickSort( $A, m + 1, r$ )

```

Die Prozedur **QuickSort**(A, l, r) sortiert ein Feld $A[l \dots r]$ wie folgt:

- Zuerst wird die Funktion **Partition**(A, l, r) aufgerufen.
- Diese wählt ein *Pivotelement*, welches sich nach dem Aufruf in $A[m]$ befindet, und sortiert das Feld so um, dass gilt:

$$A[i] \leq A[m] \leq A[j] \text{ für alle } i, j \text{ mit } l \leq i < m < j \leq r. \quad (*)$$
- Danach werden die beiden Teilfolgen $A[l \dots m - 1]$ und $A[m + 1 \dots r]$ durch jeweils einen rekursiven Aufruf sortiert.

Die Funktion **Partition**(A, l, r) *pivotisiert* das Feld $A[l \dots r]$, indem sie

- $x = A[r]$ als Pivotelement wählt,
- die übrigen Elemente mit x vergleicht und dabei umsortiert und
- den neuen Index $i + 1$ von x zurückgibt.

Prozedur Partition(A, l, r)

```

1   $i := l - 1$ 
2  for  $j := l$  to  $r - 1$  do
3    if  $A[j] \leq A[r]$  then
4       $i := i + 1$ 
5      if  $i < j$  then
6        vertausche  $A[i]$  und  $A[j]$ 
7  if  $i + 1 < r$  then
8    vertausche  $A[i + 1]$  und  $A[r]$ 
9  return( $i + 1$ )

```

Unter der Voraussetzung, dass die Funktion **Partition** korrekt arbeitet, d.h. nach ihrem Aufruf gilt (*), folgt die Korrektheit von **QuickSort** durch einen einfachen Induktionsbeweis über die Länge $n = r - l + 1$ des zu sortierenden Arrays.

Die Korrektheit von **Partition** wiederum folgt leicht aus folgender Invariante für die for-Schleife:

$A[k] \leq A[r]$ für $k = l, \dots, i$ und $A[k] > A[r]$ für $k = i + 1, \dots, j$. (**)

Da nämlich nach Ende der for-Schleife $j = r - 1$ ist, garantiert die Vertauschung von $A[i + 1]$ und $A[r]$ die Korrektheit von **Partition**. Wir müssen also nur noch die Gültigkeit der Schleifeninvariante (**) nachweisen. Um eindeutig definierte Werte von j vor und nach jeder Iteration der for-Schleife zu haben, ersetzen wir diese durch eine semantisch äquivalente while-Schleife:

Prozedur Partition(A, l, r)

```

1   $i := l - 1$ 
2   $j := l - 1$ 
3  while  $j < r - 1$  do
4     $j := j + 1$ 
5    if  $A[j] \leq A[r]$  then
6       $i := i + 1$ 
7      if  $i < j$  then
8        vertausche  $A[i]$  und  $A[j]$ 
9  if  $i + 1 < r$  then
10    vertausche  $A[i + 1]$  und  $A[r]$ 
11 return( $i + 1$ )
  
```

Nun lässt sich die Invariante (**) leicht induktiv beweisen.

Induktionsanfang: Vor Beginn der while-Schleife gilt die Invariante, da i und j den Wert $l - 1$ haben.

Induktionsschritt: Zunächst wird j hochgezählt und dann $A[j]$ mit $A[r]$ verglichen.

Im Fall $A[j] \leq A[r]$ wird auch i hochgezählt (d.h. nach Zeile 6 gilt $A[i] > A[r]$). Daher gilt *nach* der Vertauschung in Zeile 8: $A[i] \leq A[r]$ und $A[j] > A[r]$, weshalb die Gültigkeit der Invariante erhalten bleibt.

Im Fall $A[j] > A[r]$ behält die Invariante ebenfalls ihre Gültigkeit, da nur j hochgezählt wird und i unverändert bleibt.

Als nächstes schätzen wir die Laufzeit von **QuickSort** im schlechtesten Fall ab. Dieser Fall tritt ein, wenn sich das Pivotelement nach jedem Aufruf von **Partition** am Rand von A (d.h. $m = l$ oder $m = r$) befindet. Dies führt nämlich dazu, dass **Partition** der Reihe nach mit Feldern der Länge $n, n - 1, n - 2, \dots, 1$ aufgerufen wird. Da **Partition** für die Umsortierung eines Feldes der Länge n genau $n - 1$ Vergleiche benötigt, führt **QuickSort** insgesamt die maximal mögliche Anzahl

$$V(n) = \sum_{i=1}^n (i - 1) = \binom{n}{2} = \Theta(n^2)$$

von Vergleichen aus. Dieser ungünstige Fall tritt insbesondere dann ein, wenn das Eingabefeld A bereits (auf- oder absteigend) sortiert ist.

Im *besten Fall* zerlegt das Pivotelement das Feld dagegen jeweils in zwei gleich große Felder, d.h. $V(n)$ erfüllt die Rekursion

$$V(n) = \begin{cases} 0, & n = 1, \\ V(\lfloor (n - 1)/2 \rfloor) + V(\lceil (n - 1)/2 \rceil) + n - 1, & n \geq 2. \end{cases}$$

Diese hat die Lösung $V(n) = n \log_2 n - \Theta(n)$ (vgl. die worst-case Abschätzung bei **MergeSort**).

Es gibt auch Pivotauswahlstrategien, die in linearer Zeit z.B. den Median bestimmen. Dies führt auf eine Variante von **QuickSort** mit einer Laufzeit von $\Theta(n \log n)$ bei *allen* Eingaben. Allerdings ist die Bestimmung des Medians für praktische Zwecke meist zu aufwändig. Bei der Analyse des Durchschnittsfalls gehen wir von einer zufälligen Eingabepermutation $A[1 \dots n]$ der Folge $1, \dots, n$ aus. Dann ist die Anzahl $V(n)$ der Vergleichsanfragen von **QuickSort** eine Zufallsvariable. Wir können $V(n)$ als Summe $\sum_{1 \leq i < j \leq n} X_{ij}$ folgender Indikatorvariablen darstellen:

$$X_{ij} = \begin{cases} 1, & \text{falls die Werte } i \text{ und } j \text{ verglichen werden,} \\ 0, & \text{sonst.} \end{cases}$$

Ob die Werte i und j verglichen werden, entscheidet sich beim ersten Aufruf von **Partition**(A, l, r), bei dem das Pivotelement $x = A[r]$ im Intervall

$$I_{ij} = \{i, \dots, j\}$$

liegt. Bis zu diesem Aufruf werden die Werte im Intervall I_{ij} nur mit Pivotelementen außerhalb von I_{ij} verglichen und bleiben daher im gleichen Teilfeld $A[l \dots r]$ beisammen. Ist das erste Pivotelement x in I_{ij} nun nicht gleich i oder j , dann werden i und j nicht miteinander verglichen. Das liegt daran dass im Fall $i < x < j$ die Werte i und j bei diesem Aufruf in zwei verschiedene Teilfelder getrennt werden ohne miteinander verglichen zu werden.

Die Werte i und j werden also genau dann verglichen, wenn das erste Pivotelement x im Intervall I_{ij} den Wert i oder j hat. Da die Eingabe eine Zufallsfolge ohne Mehrfachvorkommen ist, nimmt x jeden Wert in I_{ij} mit Wahrscheinlichkeit $1/(j - i + 1)$ an. Daher findet mit Wahrscheinlichkeit $p_{ij} = 2/(j - i + 1)$ ein Vergleich zwischen den Werten i und j statt.

Der Erwartungswert von $V(n) = \sum_{1 \leq i < j \leq n} X_{ij}$ berechnet sich nun zu

$$\begin{aligned} E[V(n)] &= \sum_{1 \leq i < j \leq n} \underbrace{E[X_{ij}]}_{p_{ij}} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{2}{j - i + 1} = \sum_{i=1}^{n-1} \sum_{k=2}^{n-i+1} \frac{2}{k} \\ &\leq \sum_{i=1}^{n-1} \sum_{k=2}^n \frac{2}{k} \leq 2 \sum_{i=1}^{n-1} \log n = \mathcal{O}(n \log n). \end{aligned}$$

Damit ist die durchschnittliche Laufzeit von **QuickSort** $\mathcal{O}(n \log n)$. Dass dies für vergleichende Sortierverfahren asymptotisch optimal ist, wird in den Übungen gezeigt.

Satz 14. *QuickSort ist ein vergleichendes Sortierverfahren mit einer im Durchschnitt asymptotisch optimalen Laufzeit von $\mathcal{O}(n \log n)$.*

Unabhängig davon nach welcher (deterministischen) Strategie das Pivotelement gewählt wird, wird es immer Eingabefolgen geben, für die

QuickSort $\binom{n}{2}$ Vergleiche benötigt. Eine Möglichkeit, die Effizienz von **QuickSort** im Durchschnittsfall auf den schlechtesten Fall zu übertragen, besteht darin, eine *randomisierte* Auswahlstrategie für das Pivotelement anzuwenden.

Die Prozedur **RandomQuickSort**(A, l, r) arbeitet ähnlich wie **QuickSort**. Der einzige Unterschied besteht darin, dass als Pivotelement ein zufälliges Element aus dem Feld $A[l \dots r]$ gewählt wird.

Algorithmus RandomQuickSort(A, l, r)

```

1  if  $l < r$  then
2     $m := \text{RandomPartition}(A, l, r)$ 
3    RandomQuickSort( $A, l, m - 1$ )
4    RandomQuickSort( $A, m + 1, r$ )
```

Prozedur RandomPartition(A, l, r)

```

1  guess randomly  $j \in \{l, \dots, r\}$ 
2  if  $j < r$  then
3    vertausche  $A[j]$  und  $A[r]$ 
4  return(Partition( $A, l, r$ ))
```

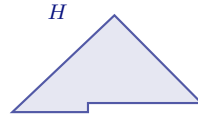
Es ist nicht schwer zu zeigen, dass sich **RandomQuickSort** bei *jeder* Eingabefolge $A[l, \dots, r]$ gleich verhält wie **QuickSort** bei einer zufälligen Permutation dieser Eingabefolge (siehe Übungen). Daher ist die erwartete Laufzeit von **RandomQuickSort** auch im *schlechtesten Fall* durch $\mathcal{O}(n \log n)$ beschränkt, falls die Zahlenwerte paarweise verschieden sind.

Satz 15. *RandomQuickSort ist ein randomisiertes vergleichendes Sortierverfahren mit einer im schlechtesten Fall asymptotisch optimalen erwarteten Laufzeit von $\mathcal{O}(n \log n)$.*

2.3.6 HeapSort

HeapSort benutzt als Datenstruktur einen so genannten *Heap*, um ein Feld zu sortieren.

Definition 16. Ein **Heap** H mit n Knoten ist ein geordneter Binärbaum nebenstehender Form. Das heißt,



- H hat in Tiefe $i = 0, 1, \dots, \lfloor \log_2 n \rfloor - 1$ jeweils die maximale Anzahl von 2^i Knoten und
- in Tiefe $\lfloor \log_2 n \rfloor$ sind alle Knoten linksbündig angeordnet.

Zudem ist jeder Knoten v mit einer Zahl $H[v]$ beschriftet, deren Wert mindestens so groß ist wie die Werte der Kinder von v (sofern vorhanden).

Ein Heap H mit n Knoten lässt sich in einem Feld $H[1, \dots, n]$ speichern. Dabei gilt:

- Das linke Kind von Knoten i hat den Index $\text{left}(i) = 2i$.
- Das rechte Kind von Knoten i hat den Index $\text{right}(i) = 2i + 1$.
- Der Elternknoten von Knoten i hat den Index $\text{parent}(i) = \lfloor i/2 \rfloor$.

Die Heap-Eigenschaft lässt sich nun wie folgt formulieren. Für alle Knoten $i \in \{1, \dots, n\}$ gilt

$$(2i \leq n \Rightarrow H[i] \geq H[2i]) \wedge (2i + 1 \leq n \Rightarrow H[i] \geq H[2i + 1]).$$

Da die Knoten im Intervall $\{\lfloor n/2 \rfloor + 1, \dots, n\}$ keine Kinder haben, ist für sie die Heap-Eigenschaft automatisch erfüllt.

Ist $H[1, \dots, n]$ ein Heap, dann repräsentiert auch jedes Anfangsstück $H[1, \dots, r]$, $1 \leq r \leq n$, einen Heap H_r mit r Knoten. Zudem ist für $1 \leq i \leq r \leq n$ der Teilbaum von H_r mit Wurzel i ein Heap, den wir mit $H_{i,r}$ bezeichnen.

Da die Wurzel $H[1]$ eines Heaps den größten Wert haben muss, können wir eine in einem Feld $H[1, \dots, n]$ gespeicherte Zahlenfolge sortieren, indem wir H zuerst zu einem Heap umsortieren und dann sukzessive

- die Wurzel $H[1]$ mit dem letzten Heap-Element vertauschen,
- den rechten Rand des Heaps um ein Feld nach links verschieben (also die vormalige Wurzel des Heaps herausnehmen) und
- die durch die Ersetzung der Wurzel verletzte Heap-Eigenschaft wieder herstellen.

Sei $H[1, \dots, n]$ ein Feld, so dass der Teilbaum $H_{i,r}$ die Heap-Eigenschaft in allen Knoten bis auf seine Wurzel i erfüllt. Dann stellt die Prozedur **Heapify**(H, i, r) die Heap-Eigenschaft im gesamten Teilbaum $H_{i,r}$ her.

Prozedur **Heapify**(H, i, r)

```

1  if  $(2i \leq r) \wedge (H[2i] > H[i])$  then
2       $x := 2i$ 
3  else
4       $x := i$ 
5  if  $(2i + 1 \leq r) \wedge (H[2i + 1] > H[x])$  then
6       $x := 2i + 1$ 
7  if  $x > i$  then
8      vertausche  $H[x]$  und  $H[i]$ 
9      Heapify( $H, x, r$ )
```

Unter Verwendung der Prozedur **Heapify** ist es nun leicht, ein Feld zu sortieren.

Algorithmus **HeapSort**($H, 1, n$)

```

1  for  $i := \lfloor n/2 \rfloor$  downto 1 do
2      Heapify( $H, i, n$ )
3  for  $r := n$  downto 2 do
4      vertausche  $H[1]$  und  $H[r]$ 
5      Heapify( $H, 1, r - 1$ )
```

Wir setzen zunächst voraus, dass die Prozedur **Heapify** korrekt arbeitet. D.h. **Heapify**(H, i, r) stellt die Heap-Eigenschaft im gesamten Teilbaum $H_{i,r}$ her, falls $H_{i,r}$ die Heap-Eigenschaft höchstens in seiner Wurzel i nicht erfüllt. Unter dieser Voraussetzung folgt die Korrektheit von **HeapSort** mittels folgender Schleifeninvarianten, die sich sehr leicht verifizieren lassen.

Invariante für die erste for-Schleife (Zeilen 1 – 2):

Für $j = i, \dots, n$ ist der Teilbaum $H_{j,n}$ ein Heap.

Nach Beendigung dieser Schleife (d.h. $i = 1$) ist demnach $H_{1,n}$ ein Heap.

Invariante für die zweite for-Schleife (Zeilen 3 – 5):

$H[r], \dots, H[n]$ enthalten die $n - r + 1$ größten Feldelemente in sortierter Reihenfolge und der Teilbaum $H_{1,r-1}$ ist ein Heap.

Am Ende der zweiten for-Schleife (d.h. $r = 2$) enthält also $H[2, \dots, n]$ die $n - 1$ größten Elemente in sortierter Reihenfolge, d.h. $H[1, \dots, n]$ ist sortiert.

Als nächstes zeigen wir die Korrektheit von **Heapify**. Sei also $H[1, \dots, n]$ ein Feld, so dass der Teilbaum $H_{i,r}$ die Heap-Eigenschaft in allen Knoten bis auf seine Wurzel i erfüllt. Dann müssen wir zeigen, dass **Heapify**(H, i, r) die Heap-Eigenschaft im gesamten Teilbaum $H_{i,r}$ herstellt.

Heapify(H, i, r) bestimmt den Knoten $x \in \{i, 2i, 2i + 1\}$ mit maximalem Wert $H(x)$. Im Fall $x = i$ erfüllt der Knoten i bereits die Heap-Eigenschaft. Ist x dagegen eines der Kinder von i , so vertauscht **Heapify** die Werte von i und x . Danach ist die Heap-Eigenschaft höchstens noch im Knoten x verletzt. Daher folgt die Korrektheit von **Heapify** durch einen einfachen Induktionsbeweis über die Rekursionstiefe.

Es ist leicht zu sehen, dass **Heapify**(H, i, r) maximal $2h(i)$ Vergleiche benötigt, wobei $h(i)$ die Höhe des Knotens i in $H_{1,r}$ ist. Daher ist die

Laufzeit von **Heapify**(H, i, r) durch $\mathcal{O}(h(i)) = \mathcal{O}(\log r)$ beschränkt. Für den Aufbau eines Heaps H der Tiefe $t = \lfloor \log_2 n \rfloor$ wird **Heapify** in der ersten for-Schleife für höchstens

- $2^0 = 1$ Knoten der Höhe $h = t$,
- $2^1 = 2$ Knoten der Höhe $h = t - 1$,
- $\vdots$
- 2^{t-1} Knoten der Höhe $h = t - (t - 1) = 1$

aufgerufen. Für $h = 1, \dots, t$ sind das also höchstens 2^{t-h} Knoten der Höhe h . Da **Heapify** für einen Knoten der Höhe h höchstens $2h$ Vergleichsfragen stellt, benötigt der Aufbau des Heaps maximal

$$V_1(n) \leq 2 \sum_{h=1}^t h 2^{t-h} \leq 2 \sum_{h=1}^t h \frac{n}{2^h} < 2n \sum_{h=1}^{\infty} \frac{h}{2^h} = 4n$$

Vergleiche. Für den Abbau des Heaps in der zweiten for-Schleife wird **Heapify** genau $(n - 1)$ -mal aufgerufen. Daher benötigt der Abbau des Heaps maximal

$$V_2(n) \leq 2(n - 1) \lfloor \log_2 n \rfloor \leq 2n \log_2 n$$

Vergleiche.

Satz 17. *HeapSort ist ein vergleichendes Sortierverfahren mit einer im schlechtesten Fall asymptotisch optimalen Laufzeit von $\mathcal{O}(n \log n)$.*

Die Floyd-Strategie

Die *Floyd-Strategie* benötigt beim Abbau des Heaps im Durchschnitt nur halb so viele Vergleiche wie die bisher betrachtete *Williams-Strategie*. Die Idee besteht darin, dass **Heapify**($H, 1, r$) beginnend mit der Wurzel $i_0 = 1$ sukzessive die Werte der beiden Kinder des aktuellen Knotens i_j vergleicht und jeweils zu dem Kind i_{j+1} mit dem

größeren Wert absteigt, bis nach $t \leq \lfloor \log_2 r \rfloor$ Schritten ein Blatt i_t erreicht wird.

Nun geht **Heapify** auf diesem Pfad bis zum ersten Knoten i_j mit $H[i_j] \geq H[1]$ zurück und führt auf den Werten der Knoten $i_j, i_{j-1}, \dots, i_0$ einen Ringtausch aus, um die Heap-Eigenschaft herzustellen. Dies erfordert

$$t + (t - j + 1) = 2t - j + 1$$

Vergleiche (im Unterschied zu $2j$ Vergleichen bei der Williams-Strategie). Da sich der Knoten i_j , an dessen Stelle der Wurzelknoten eingefügt wird, im Mittel sehr weit unten im Baum befindet (d.h. $t - j = \mathcal{O}(1)$), spart man auf diese Art asymptotisch die Hälfte der Vergleiche.

2.3.7 BucketSort

Die Prozedur **BucketSort** sortiert n Zahlen $a_1, \dots, a_n$ aus einem Intervall $[a, b]$ wie folgt (z.B. für $n = 10$, $a = 0$ und $b = 100$):

1. Erstelle für $j = 1, \dots, n$ eine Liste L_j für das halb offene Intervall $I_j = \left[a + (j-1) \frac{b-a}{n}, a + j \frac{b-a}{n} \right) = [10(j-1), 10j)$.
2. Bestimme zu jedem Element a_i das Intervall I_j , zu dem es gehört, und füge es in die entsprechende Liste L_j ein.
3. Sortiere jede Liste L_j .
4. Füge die sortierten Listen L_j wieder zu einer Liste zusammen.

Im schlechtesten Fall kommen alle Schlüssel in die gleiche Liste. Dann hat **BucketSort** dieselbe asymptotische Laufzeit wie das als Unteroutine verwendete Sortierverfahren. Sind dagegen die zu sortierenden Zahlenwerte im Intervall $[a, b]$ (annähernd) gleichverteilt, so ist die durchschnittliche Laufzeit von **BucketSort** $\Theta(n)$. Dies gilt sogar, wenn als Unteroutine ein Sortierverfahren der Komplexität $\mathcal{O}(n^2)$ verwendet wird.

Wir schätzen nun die erwartete Laufzeit von **BucketSort** ab, wobei wir annehmen, dass die Folgenglieder a_i im Intervall $[a, b]$ unabhängig gleichverteilt sind. Sei X_i die Zufallsvariable, die die Länge der Liste L_i beschreibt. Dann ist X_i binomialverteilt mit Parametern n und $p = 1/n$. Also hat X_i den Erwartungswert

$$E[X_i] = np = 1$$

und die Varianz

$$V[X_i] = np(1-p) = 1 - 1/n < 1.$$

Wegen $V[X_i] = E[X_i^2] - E[X_i]^2$ ist $E[X_i^2] = V[X_i] + E[X_i]^2 < 2$. Daher folgt für die erwartete Laufzeit $T(n)$ von **BucketSort**:

$$T(n) = \mathcal{O}(n) + E \left[\sum_{i=0}^{n-1} \mathcal{O}(X_i^2) \right] = \mathcal{O} \left(n + \sum_{i=0}^{n-1} E[X_i^2] \right) = \mathcal{O}(n).$$

2.3.8 CountingSort

Die Prozedur **CountingSort** sortiert eine Zahlenfolge, indem sie zunächst die Anzahl der Vorkommen jedes Wertes in der Folge und daraus die Rangzahlen $C[i] = \|\{j \mid A[j] \leq i\}\|$ der Zahlenwerte $i = 0, \dots, k$ bestimmt. Dies funktioniert nur unter der Einschränkung, dass die Zahlenwerte natürliche Zahlen sind und eine Obergrenze k für ihre Größe bekannt ist.

Algorithmus CountingSort($A, 1, n, k$)

```

1  for  $i := 0$  to  $k$  do  $C[i] := 0$ 
2  for  $j := 1$  to  $n$  do  $C[A[j]] := C[A[j]] + 1$ 
3  for  $i := 1$  to  $k$  do  $C[i] := C[i] + C[i-1]$ 
4  for  $j := 1$  to  $n$  do
5       $B[C[A[j]]] := A[j]$ 
6       $C[A[j]] := C[A[j]] - 1$ 
7  for  $j := 1$  to  $n$  do  $A[j] := B[j]$ 

```

Satz 18. *CountingSort* sortiert n natürliche Zahlen der Größe höchstens k in Zeit $\Theta(n + k)$ und Platz $\Theta(n + k)$.

Korollar 19. *CountingSort* sortiert n natürliche Zahlen der Größe $\mathcal{O}(n)$ in linearer Zeit und linearem Platz.

2.3.9 RadixSort

RadixSort sortiert d -stellige Zahlen $a = a_d \cdots a_1$ eine Stelle nach der anderen, wobei mit der niederwertigsten Stelle begonnen wird.

Algorithmus RadixSort($A, 1, n$)

```

1  for  $i := 1$  to  $d$  do
2    sortiere  $A[1, \dots, n]$  nach der  $i$ -ten Stelle

```

Hierzu sollten die Folgenglieder möglichst als Festkomma-Zahlen vorliegen. Zudem muss in Zeile 2 „stabil“ sortiert werden.

Definition 20. Ein Sortierverfahren heißt **stabil**, wenn es die relative Reihenfolge von Elementen mit demselben Wert nicht verändert.

Es empfiehlt sich, eine stabile Variante von **CountingSort** als Unter-routine zu verwenden. Damit **CountingSort** stabil sortiert, brauchen wir lediglich die for-Schleife in Zeile 4 in der umgekehrten Reihenfolge zu durchlaufen:

```

1  for  $j := 1$  to  $n$  do  $C[A[j]] := C[A[j]] + 1$ 
2  for  $i := 1$  to  $k$  do  $C[i] := C[i] + C[i - 1]$ 
3  for  $j := n$  downto 1 do
4     $B[C[A[j]]] := A[j]$ 
5     $C[A[j]] := C[A[j]] - 1$ 
6  for  $j := 1$  to  $n$  do  $A[j] := B[j]$ 

```

Satz 21. *RadixSort* sortiert n d -stellige Festkomma-Zahlen zur Basis b in Zeit $\Theta(d(n + b))$.

RadixSort sortiert beispielsweise n $\mathcal{O}(\log n)$ -stellige Binärzahlen in Zeit $\Theta(n \log n)$. Wenn wir r benachbarte Ziffern zu einer „Ziffer“ $z \in \{0, \dots, b^r - 1\}$ zusammenfassen, erhalten wir folgende Variante von **RadixSort**.

Korollar 22. Für jede Zahl $1 \leq r \leq d$ sortiert **RadixSort** n d -stellige Festkomma-Zahlen zur Basis b in Zeit $\Theta(d/r(n + b^r))$.

Wählen wir beispielsweise $r = \lceil \log_2 n \rceil$, so erhalten wir für $d = \mathcal{O}(\log n)$ -stellige Binärzahlen eine Komplexität von

$$\Theta(d/r(n + 2^r)) = \Theta(n + 2^r) = \Theta(n).$$

2.3.10 Vergleich der Sortierverfahren

Folgende Tabelle zeigt die Komplexitäten der betrachteten vergleichsbasierten Sortierverfahren.

	Insertion-Sort	MergeSort	Quick-Sort	Heap-Sort
worst-case	$\Theta(n^2)$	$\Theta(n \log n)$	$\Theta(n^2)$	$\Theta(n \log n)$
average-case	$\Theta(n^2)$	$\Theta(n \log n)$	$\Theta(n \log n)$	$\Theta(n \log n)$
Speicherplatz	$\Theta(1)$	$\Theta(n)$ bzw. $\Theta(1)$	$\Theta(\log n)$	$\Theta(1)$
stabil	ja	ja	nein	nein

Wir fassen auch die wichtigsten Eigenschaften der betrachteten Linearzeit-Sortierverfahren zusammen.

- **BucketSort:** Im Durchschnitt linearer Zeitverbrauch, falls die n Zahlen in einem Intervall $[a, b]$ gleichverteilt sind.
- **CountingSort:** Sogar im schlechtesten Fall lineare Zeit, falls die Werte natürliche Zahlen sind und $\mathcal{O}(n)$ nicht übersteigen.

- **RadixSort**: Bitweises Sortieren in linearer Zeit, falls die zu sortierenden Zahlen in Festkomma-Darstellung nicht mehr als $\mathcal{O}(\log n)$ Bit haben.

2.4 Datenstrukturen für dynamische Mengen

Viele Algorithmen benötigen eine Datenstruktur für dynamische Mengen. Eine solche Datenstruktur S sollte im Prinzip *beliebig viele* Elemente aufnehmen können. Die Elemente $x \in S$ werden dabei anhand eines *Schlüssels* $k = \text{key}(x)$ identifiziert. Auf die Elemente $x \in S$ wird meist nicht direkt, sondern mittels *Zeiger* (engl. *pointer*) zugegriffen.

Typische Operationen, die auf einer dynamische Mengen S auszuführen sind:

Insert(S, x): Fügt x in S ein.

Remove(S, x): Entfernt x aus S .

Search(S, k): Gibt für einen Schlüssel k (einen Zeiger auf) das Element $x \in S$ mit $\text{key}(x) = k$ zurück, falls ein solches Element existiert, und **nil** sonst.

Min(S): Gibt das Element in S mit dem kleinsten Schlüssel zurück.

Max(S): Gibt das Element in S mit dem größten Schlüssel zurück.

Prec(S, x): Gibt das Element in S mit dem nach x nächstkleineren Schlüssel zurück (bzw. **nil**, falls x das Minimum ist).

Succ(S, x): Gibt das Element in S mit dem nach x nächstgrößeren Schlüssel zurück (bzw. **nil**, falls x das Maximum ist).

2.4.1 Verkettete Listen

Die Elemente einer verketteten Liste sind in linearer Reihenfolge angeordnet. Das erste Element der Liste L ist **head**(L). Jedes Element x „kennt“ seinen Nachfolger **next**(x). Wenn jedes Element x auch

seinen Vorgänger **prev**(x) kennt, dann spricht man von einer *doppelt verketteten Liste*.

Die Prozedur **L-Insert**(L, x) fügt ein Element x in eine verkettete Liste L ein.

Prozedur L-Insert(L, x)

```
1 next(x) := head(L)
2 head(L) := x
```

Die Prozedur **DL-Insert**(L, x) fügt ein Element x in eine doppelt verkettete Liste L ein.

Prozedur DL-Insert(L, x)

```
1 next(x) := head(L)
2 if head(L) ≠ nil then
3   prev(head(L)) := x
4 head(L) := x
5 prev(x) := nil
```

Die Prozedur **DL-Remove**(L, x) entfernt wieder ein Element x aus einer doppelt verketteten Liste L .

Prozedur DL-Remove(L, x)

```
1 if x ≠ head(L) then
2   next(prev(x)) := next(x)
3 else
4   head(L) := next(x)
5 if next(x) ≠ nil then
6   prev(next(x)) := prev(x)
```

Die Prozedur **DL-Search**(L, k) sucht ein Element x mit dem Schlüssel k in der Liste L .

Prozedur DL-Search(L, k)

```

1   $x := \text{head}(L)$ 
2  while  $x \neq \text{nil}$  and  $\text{key}(x) \neq k$  do
3     $x := \text{next}(x)$ 
4  return( $x$ )

```

Es ist leicht zu sehen, dass **DL-Insert** und **DL-Remove** konstante Zeit $\Theta(1)$ benötigen, während **DL-Search** eine lineare (in der Länge der Liste) Laufzeit hat.

Bemerkung 23.

- Wird **DL-Remove** nur der Schlüssel übergeben, dann wäre die Laufzeit linear, da wir erst mit **DL-Search** das entsprechende Element suchen müssen.
- Für einfach verkettete Listen ist der Aufwand von **Remove** ebenfalls linear, da wir keinen direkten Zugriff auf den Vorgänger haben.
- Die Operationen **Max**, **Min**, **Prec** und **Succ** lassen sich ebenfalls mit linearer Laufzeit berechnen (siehe Übungen).
- Da sich **MergeSort** für Listen als „in place“-Verfahren implementieren lässt (siehe Übungen), können Listen in konstantem Platz und Zeit $\mathcal{O}(n \log n)$ sortiert werden.

2.4.2 Binäre Suchbäume

Ein Binärbaum B kann wie folgt durch eine Zeigerstruktur repräsentiert werden. Jeder Knoten x in B hat folgende drei Zeiger:

- $\text{left}(x)$ zeigt auf das linke Kind,
- $\text{right}(x)$ zeigt auf das rechte Kind und
- $\text{parent}(x)$ zeigt auf den Elternknoten.

Für die Wurzel $w = \text{root}(B)$ ist $\text{parent}(w) = \text{nil}$ und falls einem Knoten x eines seiner Kinder fehlt, so ist der entsprechende

Zeiger ebenfalls **nil**. Auf diese Art lassen sich beispielsweise Heaps für unbeschränkt viele Datensätze implementieren.

Definition 24. Ein binärer Baum B ist ein **binärer Suchbaum**, falls für jeden Knoten x in B folgende Eigenschaften erfüllt sind:

- Für jeden Knoten y im linken Teilbaum von x gilt $\text{key}(y) \leq \text{key}(x)$ und
- für jeden Knoten y im rechten Teilbaum von x gilt $\text{key}(y) \geq \text{key}(x)$.

Folgende Prozedur **ST-Search**(B, k) sucht ein Element x mit dem Schlüssel k im binären Suchbaum (engl. *search tree*) B .

Prozedur ST-Search(B, k)

```

1   $x := \text{root}(B)$ 
2  while  $x \neq \text{nil} \wedge \text{key}(x) \neq k$  do
3    if  $k \leq \text{key}(x)$  then
4       $x := \text{left}(x)$ 
5    else
6       $x := \text{right}(x)$ 
7  return( $x$ )

```

Die Prozedur **ST-Insert**(B, z) fügt ein neues Element z in B ein, indem sie den **nil**-Zeiger „sucht“, der eigentlich auf den Knoten z zeigen müsste.

Prozedur ST-Insert(B, z)

```

1  if  $\text{root}(B) = \text{nil}$  then
2     $\text{root}(B) := z$ 
3     $\text{parent}(z) := \text{nil}$ 
4  else
5     $x := \text{root}(B)$ 
6    repeat
7       $y := x$ 

```

```

8      if key(z) ≤ key(x) then
9          x := left(x)
10     else
11         x := right(x)
12     until x = nil
13     if key(z) ≤ key(y) then
14         left(y) := z
15     else
16         right(y) := z
17     parent(z) := y

```

Satz 25. Die Prozeduren *ST-Search* und *ST-Insert* laufen auf einem binären Suchbaum der Höhe h in Zeit $\mathcal{O}(h)$.

Bemerkung 26. Auch die Operationen *Min*, *Max*, *Succ*, *Prec* und *Remove* lassen sich auf einem binären Suchbaum der Höhe h in Zeit $\mathcal{O}(h)$ implementieren (siehe Übungen).

Die Laufzeiten der Operationen für binäre Suchbäume hängen von der Tiefe der Knoten im Suchbaum ab. Suchbäume können zu Listen entarten. Dieser Fall tritt z.B. ein, falls die Datensätze in sortierter Reihenfolge eingefügt werden. Daher haben die Operationen im schlechtesten Fall eine lineare Laufzeit.

Für die Analyse des Durchschnittsfalls gehen wir davon aus, dass die Einfügesequenz eine zufällige Permutation von n verschiedenen Zahlen ist. Dann lässt sich zeigen, dass der resultierende Suchbaum eine erwartete Tiefe von $\mathcal{O}(\log n)$ hat (siehe Übungen). Somit ist die erwartete Laufzeit der Operationen nur $\mathcal{O}(\log n)$.

2.4.3 Balancierte Suchbäume

Um die Tiefe des Suchbaums klein zu halten, kann er während der Einfüge- und Löschoperationen auch aktiv ausbalanciert werden. Hierfür gibt es eine ganze Reihe von Techniken. Die drei bekanntesten

sind Rot-Schwarz-Bäume, Splay-Bäume und die AVL-Bäume, mit denen wir uns im Folgenden etwas näher befassen möchten.

Definition 27. Ein **AVL-Baum** T ist ein binärer Suchbaum, der **höhenbalanciert** ist, d.h. für jeden Knoten x von T unterscheiden sich die Höhen des linken und rechten Teilbaumes von x höchstens um eins (die Höhe eines nicht existierenden Teilbaumes setzen wir mit -1 an).

Lemma 28. Die Höhe eines AVL-Baumes mit n Knoten ist $\mathcal{O}(\log n)$.

Beweis. Sei $M(h)$ die minimale Blattzahl eines AVL-Baumes der Höhe h . Dann gilt

$$M(h) = \begin{cases} 1, & h = 0 \text{ oder } 1, \\ M(h-1) + M(h-2), & h \geq 2. \end{cases}$$

$M(h)$ ist also die $(h+1)$ -te *Fibonacci-Zahl* F_{h+1} . Wir zeigen durch Induktion über h , dass $F_{h+1} \geq \phi^{h-1}$ für $h \geq 0$ ist, wobei $\phi = (1 + \sqrt{5})/2$ der *goldene Schnitt* ist. Der Induktionsanfang ($h = 0$ oder 1) ist klar, da $F_2 = F_1 = 1 = \phi^0 \geq \phi^{-1}$ ist. Unter der Induktionsannahme $F_{h'+1} \geq \phi^{h'-1}$ für $h' \leq h-1$ folgt wegen $\phi^2 = \phi + 1$

$$F_{h+1} = F_h + F_{h-1} \geq \phi^{h-2} + \phi^{h-3} = \phi^{h-3}(\phi + 1) = \phi^{h-1}.$$

Daher hat ein AVL-Baum der Höhe h mindestens

$$b \geq M(h) = F_{h+1} \geq \phi^{h-1}$$

Blätter. Da ein Binärbaum mit n Knoten höchstens $b \leq (n+1)/2$ Blätter hat, folgt

$$h \leq 1 + \log_\phi(b) < \log_\phi(n+1) = \mathcal{O}(\log_2 n).$$

Der konstante Faktor in $\mathcal{O}(\log_2 n)$ ist hierbei $\frac{1}{\log_2(\phi)} \approx 1,44$. ■

Für die Aufrechterhaltung der AVL-Eigenschaft eines AVL-Baums T benötigen wir folgende Information über jeden Knoten x . Seien h_l und h_r die Höhen des linken und des rechten Teilbaums von x . Dann heißt die Höhendifferenz

$$\text{bal}(x) = h_l - h_r$$

die *Balance* von x in T . T ist also genau dann ein AVL-Baum, wenn jeder Knoten x in T die Höhendifferenz 0, 1 oder -1 hat. Im Folgenden bezeichne $T(x)$ den Teilbaum von T mit der Wurzel x .

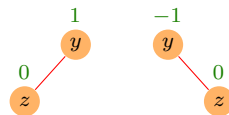
Wir fügen einen neuen Knoten z in einen AVL-Baum T ähnlich wie die Prozedur **ST-Insert** für binäre Suchbäume ein. D.h. wir „suchen“ den Schlüssel $k = \text{key}(z)$ in T bis wir einen Knoten y mit $k \leq \text{key}(y)$ und $\text{left}(y) = \text{nil}$ bzw. $k > \text{key}(y)$ und $\text{right}(y) = \text{nil}$ erreichen und fügen z an dieser Stelle als Kind von y ein. Da z ein Blatt ist, erhält z den Wert $\text{bal}(z) = 0$. Das Einfügen von z kann nur für Knoten auf dem Pfad von z zur Wurzel von T eine Änderung der Höhendifferenz bewirken. Daher genügt es, diesen Suchpfad zurückzugehen und dabei für jeden besuchten Knoten die AVL-Eigenschaft zu testen und nötigenfalls wiederherzustellen.

Wir untersuchen zuerst, ob y die AVL-Eigenschaft verletzt.

1. Falls der Wert von $\text{bal}(y)$ gleich -1 oder 1 ist, hatte $T(y)$ schon vor dem Einfügen von z die Höhe 1. Daher genügt es, $\text{bal}(y) = 0$ zu setzen.



2. Falls $\text{bal}(y) = 0$ ist, wurde z an ein Blatt gehängt, d.h. die Höhe von $T(y)$ ist um 1 gewachsen. Zunächst setzen wir $\text{bal}(y)$ auf den Wert 1 oder -1 ,



je nachdem ob z linkes oder rechtes Kind von y ist. Dann wird die rekursive Prozedur **AVL-Check-Insertion**(y) aufgerufen, die überprüft, ob weitere Korrekturen nötig sind.

Prozedur AVL-Insert(B, z)

```

1  if root( $B$ ) = nil then
2    root( $B$ ) :=  $z$ 
3    parent( $z$ ) := nil
4    bal( $z$ ) := 0
5  else
6     $x$  := root( $B$ )
7    repeat
8       $y$  :=  $x$ 
9      if key( $z$ ) ≤ key( $x$ ) then
10        $x$  := left( $x$ )
11     else
12        $x$  := right( $x$ )
13   until ( $x$  = nil)
14   if key( $z$ ) ≤ key( $y$ ) then
15     left( $y$ ) :=  $z$ 
16   else
17     right( $y$ ) :=  $z$ 
18   parent( $z$ ) :=  $y$ 
19   bal( $z$ ) := 0
20   if bal( $y$ ) ∈ {−1, 1} then
21     bal( $y$ ) := 0
22   else
23     if  $z$  = left( $y$ ) then
24       bal( $y$ ) := 1
25     else
26       bal( $y$ ) := −1
27   AVL-Check-Insertion( $y$ )

```

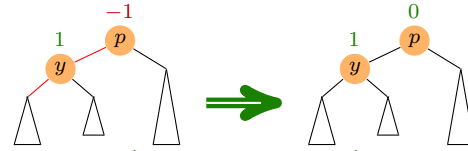
Als nächstes beschreiben wir die Prozedur **AVL-Check-Insertion**(y). Dabei setzen wir voraus, dass bei jedem Aufruf folgende Bedingung erfüllt ist:

Der Wert von $\text{bal}(y)$ wurde von 0 auf ± 1 aktualisiert, d.h.

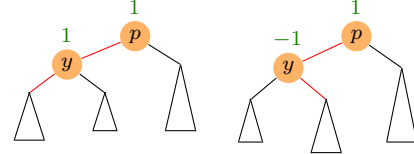
die Höhe von $T(y)$ ist um 1 gewachsen.

Falls y die Wurzel von T ist, ist nichts weiter zu tun. Andernfalls nehmen wir an, dass y linkes Kind von $p = \text{parent}(y)$ ist (der Fall $y = \text{right}(p)$ ist analog).

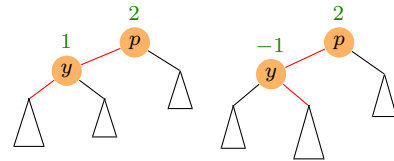
1. Im Fall $\text{bal}(p) = -1$ genügt es, $\text{bal}(p) = 0$ zu setzen.



2. Im Fall $\text{bal}(p) = 0$ setzen wir $\text{bal}(p) = 1$ und rufen **AVL-Check-Insertion**(p) auf.



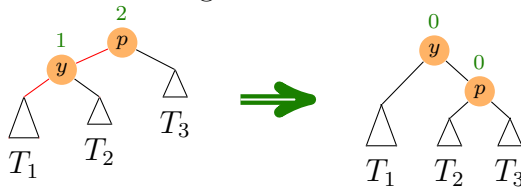
3. Im Fall $\text{bal}(p) = 1$ müssen wir T umstrukturieren, da die aktuelle Höhendifferenz von p gleich 2 ist.



- 3a. Im Fall $\text{bal}(y) = 1$ sei T_1 der linke und T_2 der rechte Teilbaum von y . Weiter sei T_3 der rechte Teilbaum von p und h sei die Höhe von $T(y)$. Dann gilt für die Höhen h_i der Teilbäume T_i :

$$h_1 = h - 1 \text{ und } h_2 = h_3 = h - 2.$$

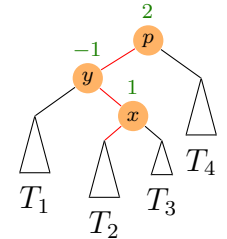
Wir führen nun eine so genannte *Rechts-Rotation* aus,



d.h. p wird rechtes Kind von y und erhält T_2 als linken und T_3 als rechten Teilbaum (d.h. $\text{bal}(p)$ erhält den Wert 0) und T_1 bleibt linker Teilbaum von y (d.h. $\text{bal}(y)$ erhält

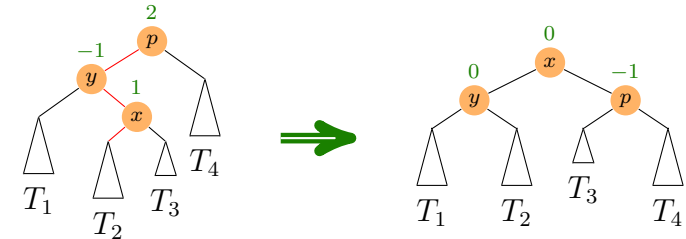
ebenfalls den Wert 0). Dann hat der rotierte Teilbaum wieder die gleiche Höhe wie vor dem Einfügen von z . Daher ist nichts weiter zu tun.

- 3b. Im Fall $\text{bal}(y) = -1$ sei T_1 der linke Teilbaum von y und T_4 der rechte Teilbaum von p . Weiter seien T_2 und T_3 der linke und rechte Teilbaum von $x = \text{right}(y)$. Die Höhe von $T(y)$ bezeichnen wir wieder mit h . Dann ist $h_3 = h_2 - \text{bal}(x)$, wobei $\text{bal}(x) = 0$ nur im Fall $h = 1$ möglich ist (d.h. $x = z$ und alle Teilbäume T_1, T_2, T_3 und T_4 sind leer). Weiter gilt



$$h_1 = h_4 = h - 2 \text{ und } h - 3 \leq h_2, h_3 \leq h - 2.$$

Daher genügt es, eine *Doppel-Rotation* (genauer: eine *Links-Rechts-Rotation*) auszuführen,



d.h. y wird linkes und p wird rechtes Kind von x , y erhält T_1 als linken und T_2 als rechten Teilbaum und p erhält T_3 als linken und T_4 als rechten Teilbaum. Die neuen Balance-Werte von p , y und x sind

$$\text{bal}(p) = \begin{cases} -1, & \text{bal}(x) = 1, \\ 0, & \text{sonst,} \end{cases} \quad \text{bal}(y) = \begin{cases} 1, & \text{bal}(x) = -1, \\ 0, & \text{sonst} \end{cases}$$

und $\text{bal}(x) = 0$. Der rotierte Teilbaum hat die gleiche Höhe wie der ursprüngliche Teilbaum an dieser Stelle und daher ist nichts weiter zu tun.

In Pseudocode lässt sich die Prozedur **AVL-Check-Insertion** demnach wie folgt implementieren.

Prozedur AVL-Check-Insertion(B, y)

```

1  $p := \text{parent}(y)$ 
2 if  $p = \text{nil}$  then return
3 if  $y = \text{left}(p)$  then
4   if  $\text{bal}(p) = -1$  then
5      $\text{bal}(p) := 0$ 
6   else if  $\text{bal}(p) = 0$  then
7      $\text{bal}(p) := 1$ 
8     AVL-Check-Insertion( $B, p$ )
9   else //  $\text{bal}(p) = 1$ 
10    if  $\text{bal}(y) = 1$  then
11      RightRotate( $B, y$ )
12    else //  $\text{bal}(y) = -1$ 
13      LeftRightRotate( $B, y$ )
14 else //  $y = \text{right}(p)$ 
15   if  $\text{bal}(p) = 1$  then
16      $\text{bal}(p) := 0$ 
17   else if  $\text{bal}(p) = 0$  then
18      $\text{bal}(p) := -1$ 
19     AVL-Check-Insertion( $B, p$ )
20   else //  $\text{bal}(p) = -1$ 
21     if  $\text{bal}(y) = -1$  then
22       LeftRotate( $B, y$ )
23     else //  $\text{bal}(y) = 1$ 
24       RightLeftRotate( $B, y$ )

```

Wir geben exemplarisch auch noch die Prozedur **RightRotate** in Pseudocode an. Die übrigen Rotationsprozeduren lassen sich ganz ähnlich implementieren (siehe Übungen).

Prozedur RightRotate(B, y)

```

1  $p := \text{parent}(y)$ 
2  $T_1 := \text{left}(y)$ ;  $T_2 := \text{right}(y)$ ;  $T_3 := \text{right}(p)$ 
3 // setze  $y$  an die Wurzel des Teilbaums
4  $T(p)$   $p' := \text{parent}(p)$ 
5 if  $p' = \text{nil}$  then
6    $\text{root}(B) := y$ 
7 else if  $\text{left}(p') = p$  then
8    $\text{left}(p') := y$ 
9 else //  $\text{right}(p') = p$ 
10     $\text{right}(p') := y$ 
11  $\text{parent}(y) := p'$ 
12 // setze  $p$  als rechtes Kind von  $y$ 
13  $\text{right}(y) := p$ 
14  $\text{parent}(p) := y$ 
15 // setze  $T_2$  als linken Teilbaum von  $p$ 
16  $\text{left}(p) := T_2$ 
17 if  $T_2 \neq \text{nil}$  then  $\text{parent}(T_2) := p$ 
18  $\text{bal}(y) := 0$ ;  $\text{bal}(p) := 0$  // aktualisiere die Balancen

```

Folgende Tabelle fasst die worst-case Komplexitäten der betrachteten Datenstrukturen für dynamische Mengen zusammen.

	Search	Min/Max	Prec/Succ	Insert	Remove
Heap	$\mathcal{O}(n)$	$\mathcal{O}(1)$	$\mathcal{O}(n)$	$\mathcal{O}(\log n)$	$\mathcal{O}(\log n)$
Liste (einfach)	$\mathcal{O}(n)$	$\mathcal{O}(n)$	$\mathcal{O}(n)$	$\mathcal{O}(1)$	$\mathcal{O}(n)$
Liste (doppelt)	$\mathcal{O}(n)$	$\mathcal{O}(n)$	$\mathcal{O}(n)$	$\mathcal{O}(1)$	$\mathcal{O}(1)$
Suchbaum	$\mathcal{O}(n)$	$\mathcal{O}(n)$	$\mathcal{O}(n)$	$\mathcal{O}(n)$	$\mathcal{O}(n)$
AVL-Baum	$\mathcal{O}(\log n)$	$\mathcal{O}(\log n)$	$\mathcal{O}(\log n)$	$\mathcal{O}(\log n)$	$\mathcal{O}(\log n)$

3 Graphalgorithmen

3.1 Grundlegende Begriffe

Definition 29. Ein (ungerichteter) **Graph** ist ein Paar $G = (V, E)$, wobei

- V - eine endliche Menge von **Knoten/Ecken** und
- E - die Menge der **Kanten** ist.

Hierbei gilt

$$E \subseteq \binom{V}{2} = \{\{u, v\} \subseteq V \mid u \neq v\}.$$

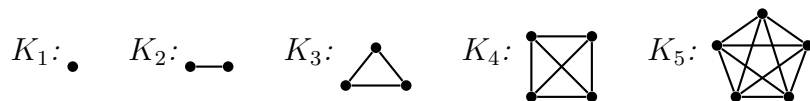
Sei $v \in V$ ein Knoten.

- a) Die **Nachbarschaft** von v ist $N_G(v) = \{u \in V \mid \{u, v\} \in E\}$.
- b) Der **Grad** von v ist $\deg_G(v) = \|N_G(v)\|$.
- c) Der **Minimalgrad** von G ist $\delta(G) = \min_{v \in V} \deg_G(v)$ und der **Maximalgrad** von G ist $\Delta(G) = \max_{v \in V} \deg_G(v)$.

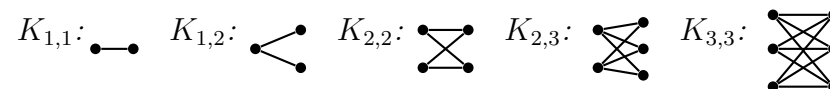
Falls G aus dem Kontext ersichtlich ist, schreiben wir auch einfach $N(v)$, $\deg(v)$, δ usw.

Beispiel 30.

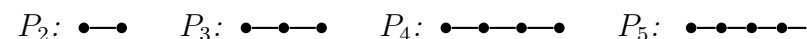
- Der vollständige Graph (V, E) auf n Knoten, d.h. $\|V\| = n$ und $E = \binom{V}{2}$, wird mit K_n und der leere Graph $(V, \emptyset)$ auf n Knoten wird mit E_n bezeichnet.



- Der vollständige bipartite Graph (A, B, E) auf $a+b$ Knoten, d.h. $A \cap B = \emptyset$, $\|A\| = a$, $\|B\| = b$ und $E = \{\{u, v\} \mid u \in A, v \in B\}$ wird mit $K_{a,b}$ bezeichnet.



- Der Pfad der Länge $n - 1$ wird mit P_n bezeichnet.



- Der Kreis der Länge n wird mit C_n bezeichnet.



Definition 31. Sei $G = (V, E)$ ein Graph.

- a) Eine Knotenmenge $U \subseteq V$ heißt **stabil**, wenn es keine Kante von G mit beiden Endpunkten in U gibt, d.h. es gilt $E \cap \binom{U}{2} = \emptyset$. Die **Stabilitätszahl** ist

$$\alpha(G) = \max\{\|U\| \mid U \text{ ist stabile Menge in } G\}.$$

- b) Eine Knotenmenge $U \subseteq V$ heißt **Clique**, wenn jede Kante mit beiden Endpunkten in U in E ist, d.h. es gilt $\binom{U}{2} \subseteq E$. Die **Cliquenzahl** ist

$$\omega(G) = \max\{\|U\| \mid U \text{ ist Clique in } G\}.$$

- c) Eine Abbildung $f: V \rightarrow \mathbb{N}$ heißt **Färbung** von G , wenn $f(u) \neq f(v)$ für alle $\{u, v\} \in E$ gilt. G heißt **k-färbbar**, falls eine Färbung $f: V \rightarrow \{1, \dots, k\}$ existiert. Die **chromatische Zahl** ist

$$\chi(G) = \min\{k \in \mathbb{N} \mid G \text{ ist } k\text{-färbbar}\}.$$

- d) Ein Graph heißt **bipartit**, wenn $\chi(G) \leq 2$ ist.

- e) Ein Graph $G' = (V', E')$ heißt **Sub-/Teil-/Untergraph** von G , falls $V' \subseteq V$ und $E' \subseteq E$ ist. Ein Subgraph $G' = (V', E')$ heißt (**durch V'**) **induziert**, falls $E' = E \cap \binom{V'}{2}$ ist. Hierfür schreiben wir auch $H = G[V']$.
- f) Ein **Weg** ist eine Folge von (nicht notwendig verschiedenen) Knoten $v_0, \dots, v_\ell$ mit $\{v_i, v_{i+1}\} \in E$ für $i = 0, \dots, \ell - 1$. Die **Länge** des Weges ist die Anzahl der Kanten, also ℓ . Im Fall $\ell = 0$ heißt der Weg **trivial**. Ein Weg $v_0, \dots, v_\ell$ heißt auch **v_0 - v_ℓ -Weg**.
- g) Ein Graph $G = (V, E)$ heißt **zusammenhängend**, falls es für alle Paare $\{u, v\} \in \binom{V}{2}$ einen u - v -Weg gibt.
- h) Ein **Zyklus** ist ein u - v -Weg der Länge $\ell \geq 2$ mit $u = v$.
- i) Ein Weg heißt **einfach** oder **Pfad**, falls alle durchlaufenen Knoten verschieden sind.
- j) Ein **Kreis** ist ein Zyklus $v_0, v_1, \dots, v_{\ell-1}, v_0$ der Länge $\ell \geq 3$, für den $v_0, v_1, \dots, v_{\ell-1}$ paarweise verschieden sind.
- k) Ein Graph $G = (V, E)$ heißt **kreisfrei**, **azyklisch** oder **Wald**, falls er keinen Kreis enthält.
- l) Ein **Baum** ist ein zusammenhängender Wald.
- m) Jeder Knoten $u \in V$ vom Grad $\deg(u) \leq 1$ heißt **Blatt** und die übrigen Knoten (vom Grad ≥ 2) heißen **innere Knoten**.

Es ist leicht zu sehen, dass die Relation

$$Z = \{(u, v) \in V \times V \mid \text{es gibt in } G \text{ einen } u\text{-}v\text{-Weg}\}$$

eine Äquivalenzrelation ist. Die durch die Äquivalenzklassen von Z induzierten Teilgraphen heißen die **Zusammenhangskomponenten** (engl. *connected components*) von G .

Definition 32. Ein **gerichteter Graph** oder **Digraph** ist ein Paar $G = (V, E)$, wobei

V - eine endliche Menge von **Knoten/Ecken** und
 E - die Menge der **Kanten** ist.

Hierbei gilt

$$E \subseteq V \times V = \{(u, v) \mid u, v \in V\},$$

wobei E auch Schlingen (u, u) enthalten kann. Sei $v \in V$ ein Knoten.

- a) Die **Nachfolgermenge** von v ist $N^+(v) = \{u \in V \mid (v, u) \in E\}$.
- b) Die **Vorgängermenge** von v ist $N^-(v) = \{u \in V \mid (u, v) \in E\}$.
- c) Die **Nachbarmenge** von v ist $N(v) = N^+(v) \cup N^-(v)$.
- d) Der **Ausgangsgrad** von v ist $\deg^+(v) = \|N^+(v)\|$ und der **Eingangsgrad** von v ist $\deg^-(v) = \|N^-(v)\|$. Der **Grad** von v ist $\deg(v) = \deg^+(v) + \deg^-(v)$.
- e) Ein **gerichteter v_0 - v_ℓ -Weg** ist eine Folge von Knoten $v_0, \dots, v_\ell$ mit $(v_i, v_{i+1}) \in E$ für $i = 0, \dots, \ell - 1$.
- f) Ein **gerichteter Zyklus** ist ein gerichteter u - v -Weg der Länge $\ell \geq 1$ mit $u = v$.
- g) Ein gerichteter Weg heißt **einfach** oder **gerichteter Pfad**, falls alle durchlaufenen Knoten verschieden sind.
- h) Ein **gerichteter Kreis** ist ein gerichteter Zyklus $v_0, v_1, \dots, v_{\ell-1}, v_0$ der Länge $\ell \geq 1$, für den $v_0, v_1, \dots, v_{\ell-1}$ paarweise verschieden sind.
- i) Ein Digraph $G = (V, E)$ heißt **kreisfrei** oder **azyklisch**, wenn er keinen gerichteten Kreis hat.
- j) Ein Digraph $G = (V, E)$ heißt **schwach zusammenhängend**, wenn es für jedes Paar $\{u, v\} \in \binom{V}{2}$ einen gerichteten u - v -Pfad oder einen gerichteten v - u -Pfad gibt.
- k) $G = (V, E)$ heißt **stark zusammenhängend**, wenn es für jedes Paar $\{u, v\} \in \binom{V}{2}$ sowohl einen gerichteten u - v -Pfad als auch einen gerichteten v - u -Pfad gibt.

3.2 Datenstrukturen für Graphen

Sei $G = (V, E)$ ein Graph bzw. Digraph und sei $V = \{v_1, \dots, v_n\}$. Dann ist die $(n \times n)$ -Matrix $A = (a_{ij})$ mit den Einträgen

$$a_{ij} = \begin{cases} 1, & \{v_i, v_j\} \in E \\ 0, & \text{sonst} \end{cases} \quad \text{bzw.} \quad a_{ij} = \begin{cases} 1, & (v_i, v_j) \in E \\ 0, & \text{sonst} \end{cases}$$

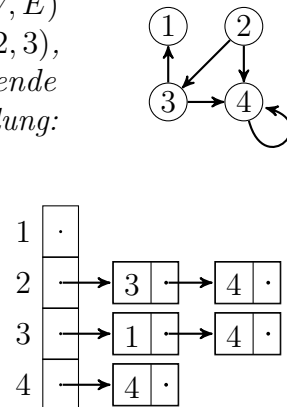
die *Adjazenzmatrix* von G . Für ungerichtete Graphen ist die Adjazenzmatrix symmetrisch mit $a_{ii} = 0$ für $i = 1, \dots, n$.

Bei der *Adjazenzlisten-Darstellung* wird für jeden Knoten v_i eine Liste mit seinen Nachbarn verwaltet. Im gerichteten Fall verwaltet man entweder nur die Liste der Nachfolger oder zusätzlich eine weitere für die Vorgänger. Falls die Anzahl der Knoten gleichbleibt, organisiert man die Adjazenzlisten in einem Feld, d.h. das Feldelement mit Index i verweist auf die Adjazenzliste von Knoten v_i . Falls sich die Anzahl der Knoten dynamisch ändert, so werden die Adjazenzlisten typischerweise ebenfalls in einer doppelt verketteten Liste verwaltet.

Beispiel 33.

Betrachte den gerichteten Graphen $G = (V, E)$ mit $V = \{1, 2, 3, 4\}$ und $E = \{(2, 3), (2, 4), (3, 1), (3, 4), (4, 4)\}$. Dieser hat folgende Adjazenzmatrix- und Adjazenzlisten-Darstellung:

	1	2	3	4
1	0	0	0	0
2	0	0	1	1
3	1	0	0	1
4	0	0	0	1



Folgende Tabelle gibt den Aufwand der wichtigsten elementaren Operationen auf Graphen in Abhängigkeit von der benutzten Datenstruktur

an. Hierbei nehmen wir an, dass sich die Knotenmenge V nicht ändert.

	Adjazenzmatrix		Adjazenzlisten	
	einfach	clever	einfach	clever
Speicherbedarf	$\mathcal{O}(n^2)$	$\mathcal{O}(n^2)$	$\mathcal{O}(n + m)$	$\mathcal{O}(n + m)$
Initialisieren	$\mathcal{O}(n^2)$	$\mathcal{O}(1)$	$\mathcal{O}(n)$	$\mathcal{O}(1)$
Kante einfügen	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(1)$
Kante entfernen	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(n)$	$\mathcal{O}(1)$
Test auf Kante	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(n)$	$\mathcal{O}(n)$

Bemerkung 34.

- Der Aufwand für die Initialisierung des leeren Graphen in der Adjazenzmatrixdarstellung lässt sich auf $\mathcal{O}(1)$ drücken, indem man mithilfe eines zusätzlichen Feldes B die Gültigkeit der Matrixeinträge verwaltet (siehe Übungen).
- Die Verbesserung beim Löschen einer Kante in der Adjazenzlistendarstellung erhält man, indem man die Adjazenzlisten doppelt verkettet und im ungerichteten Fall die beiden Vorkommen jeder Kante in den Adjazenzlisten der beiden Endknoten gegenseitig verlinkt (siehe Übungen).
- Bei der Adjazenzlistendarstellung können die Knoten auch in einer doppelt verketteten Liste organisiert werden. In diesem Fall können dann auch Knoten in konstanter Zeit hinzugefügt und in Zeit $\mathcal{O}(n)$ wieder entfernt werden (unter Beibehaltung der übrigen Speicher- und Laufzeitschranken).

3.3 Keller und Warteschlange

Für das Durchsuchen eines Graphen ist es vorteilhaft, die bereits besuchten (aber noch nicht abgearbeiteten) Knoten in einer Menge B zu speichern. Damit die Suche effizient ist, sollte die Datenstruktur für B folgende Operationen effizient implementieren.

Init(B): Initialisiert B als leere Menge.
Empty(B): Testet B auf Leerheit.
Insert(B, u): Fügt u in B ein.
Element(B): Gibt ein Element aus B zurück.
Remove(B): Gibt ebenfalls **Element**(B) zurück und entfernt es aus B .

Andere Operationen wie z.B. **Remove**(B, u) werden nicht benötigt. Die gewünschten Operationen lassen sich leicht durch einen *Keller* (auch *Stapel* genannt) (engl. *stack*) oder eine *Warteschlange* (engl. *queue*) implementieren. Falls maximal n Datensätze gespeichert werden müssen, kann ein Feld zur Speicherung der Elemente benutzt werden. Andernfalls können sie auch in einer einfach verketteten Liste gespeichert werden.

Stack S – Last-In-First-Out

Top(S): Gibt das oberste Element von S zurück.
Push(S, x): Fügt x als oberstes Element zum Keller hinzu.
Pop(S): Gibt das oberste Element von S zurück und entfernt es.

Queue Q – Last-In-Last-Out

Enqueue(Q, x): Fügt x am Ende der Schlange hinzu.
Head(Q): Gibt das erste Element von Q zurück.
Dequeue(Q): Gibt das erste Element von Q zurück und entfernt es.

Die Kelleroperationen lassen sich wie folgt auf einem Feld $S[1 \dots n]$ implementieren. Die Variable **size**(S) enthält die Anzahl der im Keller gespeicherten Elemente.

Prozedur **StackInit**(S)

```
1  size( $S$ ) := 0
```

Prozedur **StackEmpty**(S)

```
1  return(size( $S$ ) = 0)
```

Prozedur **Top**(S)

```
1  if size( $S$ ) > 0 then
2    return( $S$ [size( $S$ )])
3  else
4    return(nil)
```

Prozedur **Push**(S, x)

```
1  if size( $S$ ) <  $n$  then
2    size( $S$ ) := size( $S$ ) + 1
3     $S$ [size( $S$ )] :=  $x$ 
4  else
5    return(nil)
```

Prozedur **Pop**(S)

```
1  if size( $S$ ) > 0 then
2    size( $S$ ) := size( $S$ ) - 1
3    return( $S$ [size( $S$ ) + 1])
4  else
5    return(nil)
```

Es folgen die Warteschlangenoperationen für die Speicherung in einem Feld $Q[1 \dots n]$. Die Elemente werden der Reihe nach am Ende der Schlange Q (zyklisch) eingefügt und am Anfang entnommen. Die Variable **head**(Q) enthält den Index des ersten Elements der Schlange und **tail**(Q) den Index des hinter dem letzten Element von Q befindlichen Eintrags.

Prozedur QueueInit(Q)

```

1 head( $Q$ ) := 1
2 tail( $Q$ ) := 1
3 size( $Q$ ) := 0

```

Prozedur QueueEmpty(Q)

```

1 return(size( $Q$ ) = 0)

```

Prozedur Head(Q)

```

1 if QueueEmpty( $Q$ ) then
2   return(nil)
3 else
4   return $Q$ [head( $Q$ )]

```

Prozedur Enqueue(Q, x)

```

1 if size( $Q$ ) =  $n$  then
2   return(nil)
3 size( $Q$ ) := size( $Q$ ) + 1
4  $Q$ [tail( $Q$ )] :=  $x$ 
5 if tail( $Q$ ) =  $n$  then
6   tail( $Q$ ) := 1
7 else
8   tail( $Q$ ) := tail( $Q$ ) + 1

```

Prozedur Dequeue(Q)

```

1 if QueueEmpty( $Q$ ) then
2   return(nil)
3 size( $Q$ ) := size( $Q$ ) - 1
4  $x$  :=  $Q$ [head( $Q$ )]
5 if head( $Q$ ) =  $n$  then
6   head( $Q$ ) := 1

```

```

7 else
8   head( $Q$ ) := head( $Q$ ) + 1
9 return( $x$ )

```

Satz 35. *Sämtliche Operationen für einen Keller S und eine Warteschlange Q sind in konstanter Zeit $\mathcal{O}(1)$ ausführbar.*

Bemerkung 36. *Mit Hilfe von einfach verketteten Listen sind Keller und Warteschlangen auch für eine unbeschränkte Anzahl von Datensätzen mit denselben Laufzeitbeschränkungen implementierbar.*

Die für das Durchsuchen von Graphen benötigte Datenstruktur B lässt sich nun mittels Keller bzw. Schlange wie folgt realisieren.

Operation	Keller S	Schlange Q
Init(B)	StackInit(S)	QueueInit(Q)
Empty(B)	StackEmpty(S)	QueueEmpty(Q)
Insert(B, u)	Push(S, u)	Enqueue(Q, u)
Element(B)	Top(S)	Head(Q)
Remove(B)	Pop(S)	Dequeue(Q)

3.4 Durchsuchen von Graphen

Wir geben nun für die Suche in einem Graphen bzw. Digraphen $G = (V, E)$ einen Algorithmus **GraphSearch** mit folgenden Eigenschaften an:

GraphSearch benutzt eine Prozedur **Explore**, um alle Knoten und Kanten von G zu besuchen.

Explore(w) findet Pfade zu allen von w aus erreichbaren Knoten. Hierzu speichert **Explore(w)** für jeden über eine Kante $\{u, v\}$ bzw. (u, v) neu entdeckten Knoten $v \neq w$ den Knoten u in **parent(v)**. Wir nennen die bei der Entdeckung eines neuen Knotens v durchlaufenen Kanten **(parent(v), v)** *parent-Kanten*.

Algorithmus GraphSearch(V, E)

```

1  for all  $v \in V, e \in E$  do
2    visited( $v$ ) := false
3    parent( $v$ ) := nil
4    visited( $e$ ) := false
5  for all  $w \in V$  do
6    if visited( $w$ ) = false then Explore( $w$ )

```

Prozedur Explore(w)

```

1  visited( $w$ ) := true
2  Init( $B$ )
3  Insert( $B, w$ )
4  while  $\neg \text{Empty}(B)$  do
5     $u := \text{Element}(B)$ 
6    if  $\exists e = \{u, v\}$  bzw.  $e = (u, v) \in E : \text{visited}(e) = \text{false}$ 
       then
7      visited( $e$ ) := true
8      if visited( $v$ ) = false then
9        visited( $v$ ) := true
10       parent( $v$ ) :=  $u$ 
11       Insert( $B, v$ )
12  else
13    Remove( $B$ )

```

3.4.1 Suchwälder

Definition 37. Sei $G = (V, E)$ ein Digraph.

- Ein Knoten $w \in V$ heißt **Wurzel** von G , falls alle Knoten $v \in V$ von w aus erreichbar sind (d.h. es gibt einen gerichteten w - v -Weg in G).

- G heißt **gerichteter Wald**, wenn G kreisfrei ist und jeder Knoten $v \in V$ Eingangsgrad $\deg^-(v) \leq 1$ hat.
- Ein Knoten $u \in V$ vom Ausgangsgrad $\deg^+(u) = 0$ heißt **Blatt**.
- Ein Knoten $u \in V$ heißt **Nachfahre** von v , falls in G ein gerichteter v - u -Weg existiert. In diesem Fall ist v ein **Vorfahre** von u . Gilt zudem $u \neq v$, so sprechen wir auch von **echten** Nach- bzw. Vorfahren.
- Ein **gerichteter Wald**, der eine Wurzel hat, heißt **gerichteter Baum**.

In einem gerichteten Baum liegen die Kantenrichtungen durch die Wahl der Wurzel bereits eindeutig fest. Daher kann bei bekannter Wurzel auf die Angabe der Kantenrichtungen auch verzichtet werden. Man spricht dann von einem *Wurzelbaum*.

Betrachte den durch SearchGraph(V, E) erzeugten Digraphen $W = (V, E_{\text{parent}})$ mit

$$E_{\text{parent}} = \{(\text{parent}(v), v) \mid v \in V \text{ und } \text{parent}(v) \neq \text{nil}\}.$$

Da $\text{parent}(v)$ vor v markiert wird, ist klar, dass W kreisfrei ist. Zudem hat jeder Knoten v höchstens einen Vorgänger $\text{parent}(v)$. Dies zeigt, dass W tatsächlich ein gerichteter Wald ist. W heißt *Suchwald* von G und die Kanten $(\text{parent}(v), v)$ von W werden auch als *Baumkanten* bezeichnet.

W hängt zum einen davon ab, wie die Datenstruktur B implementiert ist (z.B. als Keller oder als Warteschlange). Zum anderen hängt W aber auch von der Reihenfolge der Knoten in den Adjazenzlisten ab.

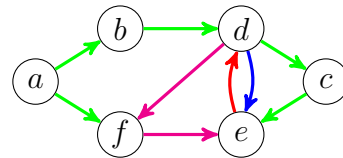
3.4.2 Klassifikation der Kanten eines (Di-)Graphen

Die Kanten eines Graphen $G = (V, E)$ werden durch den Suchwald $W = (V, E_{\text{parent}})$ in vier Klassen eingeteilt. Dabei erhält jede Kante die Richtung, in der sie bei ihrem ersten Besuch durchlaufen wird.

Neben den Baumkanten $(\text{parent}(v), v) \in E_{\text{parent}}$ gibt es noch *Rückwärts*-, *Vorwärts*- und *Querkanten*. *Rückwärtskanten* (u, v) verbinden einen Knoten u mit einem Knoten v , der auf dem **parent**-Pfad $P(u)$ von u liegt. Liegt dagegen u auf $P(v)$, so wird (u, v) als *Vorwärtskante* bezeichnet. Alle übrigen Kanten heißen *Querkanten*. Diese verbinden zwei Knoten, von denen keiner auf dem **parent**-Pfad des anderen liegt.

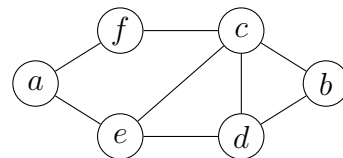
Die Klassifikation der Kanten eines Digraphen G erfolgt analog, wobei die Richtungen jedoch bereits durch G vorgegeben sind (dabei werden Schlingen der Kategorie der Vorwärtskanten zugeordnet).

Beispiel 38. Bei Aufruf mit dem Startknoten a generiert die Prozedur **Explore** nebenstehenden Suchwald.

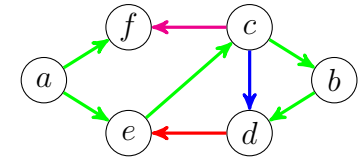


Menge B	bes. Knoten	bes. Kante	Typ	B	bes. Knoten	bes. Kante	Typ
$\{a\}$	a	(a, b)	B	$\{d, e, f\}$	d	(d, e)	V
$\{a, b\}$	a	(a, f)	B	$\{d, e, f\}$	d	(d, f)	Q
$\{a, b, f\}$	a	-	-	$\{d, e, f\}$	d	-	-
$\{b, f\}$	b	(b, d)	B	$\{e, f\}$	e	(e, d)	R
$\{b, d, f\}$	b	-	-	$\{e, f\}$	e	-	-
$\{d, f\}$	d	(d, c)	B	$\{f\}$	f	(f, e)	Q
$\{c, d, f\}$	c	(c, e)	B	$\{f\}$	f	-	-
$\{c, d, e, f\}$	c	-	-	$\emptyset$			

Bei einem Graphen durchläuft **Explore** die Knoten und Kanten in der gleichen Reihenfolge wie bei dem



Digraphen, der für jede ungerichtete Kante $\{u, v\}$ die beiden gerichteten Kanten (u, v) und (v, u) enthält.



Menge B	Knoten	Kante		B	Knoten	Kante	
$\{a\}$	a	$\{a, e\}$	B	$\{c, d, e, f\}$	c	$\{c, f\}$	Q
$\{a, e\}$	a	$\{a, f\}$	B	$\{c, d, e, f\}$	c	-	-
$\{a, e, f\}$	a	-	-	$\{d, e, f\}$	d	$\{d, b\}$	-
$\{e, f\}$	e	$\{e, a\}$	-	$\{d, e, f\}$	d	$\{d, c\}$	-
$\{e, f\}$	e	$\{e, c\}$	B	$\{d, e, f\}$	d	$\{d, e\}$	R
$\{c, e, f\}$	c	$\{c, b\}$	B	$\{d, e, f\}$	d	-	-
$\{b, c, e, f\}$	b	$\{b, c\}$	-	$\{e, f\}$	e	$\{e, d\}$	-
$\{b, c, e, f\}$	b	$\{b, d\}$	B	$\{e, f\}$	e	-	-
$\{b, c, d, e, f\}$	b	-	-	$\{f\}$	f	$\{f, a\}$	-
$\{c, d, e, f\}$	c	$\{c, d\}$	V	$\{f\}$	f	$\{f, c\}$	-
$\{c, d, e, f\}$	c	$\{c, e\}$	-	$\{f\}$	f	-	-

Satz 39. Falls der (un)gerichtete Graph G in Adjazenzlisten-Darstellung gegeben ist, durchläuft **GraphSearch** alle Knoten und Kanten von G in Zeit $\mathcal{O}(n + m)$.

Beweis. Offensichtlich wird jeder Knoten u genau einmal zu B hinzugefügt. Dies geschieht zu dem Zeitpunkt, wenn u zum ersten Mal „besucht“ und das Feld **visited** für u auf **true** gesetzt wird. Außerdem werden in Zeile 6 von **Explore** alle von u ausgehenden Kanten durchlaufen, bevor u wieder aus B entfernt wird. Folglich werden tatsächlich alle Knoten und Kanten von G besucht.

Wir bestimmen nun die Laufzeit des Algorithmus **GraphSearch**. Innerhalb von **Explore** wird die while-Schleife für jeden Knoten u genau

$(\deg(u) + 1)$ -mal bzw. $(\deg^+(u) + 1)$ -mal durchlaufen:

- einmal für jeden Nachbarn v von u und
- dann noch einmal, um u aus B zu entfernen.

Insgesamt sind das $n + 2m$ im ungerichteten bzw. $n + m$ Durchläufe im gerichteten Fall. Bei Verwendung von Adjazenzlisten kann die nächste von einem Knoten v aus noch nicht besuchte Kante e in konstanter Zeit ermittelt werden, falls man für jeden Knoten v einen Zeiger auf (den Endpunkt von) e in der Adjazenzliste von v vorsieht. Die Gesamtlaufzeit des Algorithmus **GraphSearch** beträgt somit $\mathcal{O}(n + m)$. ■

Als nächstes zeigen wir, dass **Explore**(w) zu allen von w aus erreichbaren Knoten v einen (gerichteten) w - v -Pfad liefert. Dieser lässt sich mittels **parent** wie folgt zurückverfolgen. Sei

$$u_i = \begin{cases} v, & i = 0, \\ \text{parent}(u_{i-1}), & i > 0 \text{ und } u_{i-1} \neq \text{nil} \end{cases}$$

und sei $\ell = \min\{i \geq 0 \mid u_{i+1} = \text{nil}\}$. Dann ist $u_\ell = w$ und $p = (u_\ell, \dots, u_0)$ ein w - v -Pfad. Wir nennen P den **parent**-Pfad von v und bezeichnen ihn mit $P(v)$.

Satz 40. *Falls beim Aufruf von **Explore** alle Knoten und Kanten als unbesucht markiert sind, berechnet **Explore**(w) zu allen erreichbaren Knoten v einen (gerichteten) w - v -Pfad $P(v)$.*

Beweis. Wir zeigen zuerst, dass **Explore**(w) alle von w aus erreichbaren Knoten besucht. Hierzu führen wir Induktion über die Länge ℓ eines kürzesten w - v -Weges.

$\ell = 0$: In diesem Fall ist $v = w$ und w wird in Zeile 1 besucht.

$\ell \rightsquigarrow \ell + 1$: Sei v ein Knoten mit Abstand $\ell + 1$ von w . Dann hat ein Nachbarknoten $u \in N(v)$ den Abstand ℓ von w . Folglich wird u

nach IV besucht. Da u erst dann aus B entfernt wird, wenn alle seine Nachbarn (bzw. Nachfolger) besucht wurden, wird auch v besucht.

Es bleibt zu zeigen, dass **parent** einen Pfad $P(v)$ von w zu jedem besuchten Knoten v liefert. Hierzu führen wir Induktion über die Anzahl k der vor v besuchten Knoten.

$k = 0$: In diesem Fall ist $v = w$. Da **parent**(w) = **nil** ist, liefert **parent** einen w - v -Pfad (der Länge 0).

$k - 1 \rightsquigarrow k$: Sei $u = \text{parent}(v)$. Da u vor v besucht wird, liefert **parent** nach IV einen w - u -Pfad $P(u)$. Wegen $u = \text{parent}(v)$ ist u der Entdecker von v und daher mit v durch eine Kante verbunden. Somit liefert **parent** auch für v einen w - v -Pfad $P(v)$. ■

3.4.3 Spannbäume und Spannwälder

In diesem Abschnitt zeigen wir, dass der Algorithmus **GraphSearch** für jede Zusammenhangskomponente eines (ungerichteten) Graphen G einen Spannbaum berechnet.

Definition 41. *Sei $G = (V, E)$ ein Graph und $H = (U, F)$ ein Untergraph.*

- H heißt **spannend**, falls $U = V$ ist.
- H ist ein **spannender Baum** (oder **Spannbaum**) von G , falls $U = V$ und H ein Baum ist.
- H ist ein **spannender Wald** (oder **Spannwald**) von G , falls $U = V$ und H ein Wald ist.

Es ist leicht zu sehen, dass für G genau dann ein Spannbaum existiert, wenn G zusammenhängend ist. Allgemeiner gilt, dass die Spannbäume für die Zusammenhangskomponenten von G einen Spannwald

bilden. Dieser ist bzgl. der Subgraph-Relation maximal, da er in keinem größeren Spannwald enthalten ist. Ignorieren wir die Richtungen der Kanten im Suchwald W , so ist der resultierende Wald W' ein maximaler Spannwald für G .

Da **Explore**(w) alle von w aus erreichbaren Knoten findet, spannt jeder Baum des (ungerichteten) Suchwaldes $W' = (V, E'_{\text{parent}})$ mit

$$E'_{\text{parent}} = \{ \{ \text{parent}(v), v \} \mid v \in V \text{ und } \text{parent}(v) \neq \text{nil} \}$$

eine Zusammenhangskomponente von G .

Korollar 42. *Sei G ein (ungerichteter) Graph.*

- Der Algorithmus **GraphSearch**(V, E) berechnet in Linearzeit einen Spannwald W' , dessen Bäume die Zusammenhangskomponenten von G spannen.
- Falls G zusammenhängend ist, ist W' ein Spannbaum für G .

3.4.4 Berechnung der Zusammenhangskomponenten

Folgende Variante von **GraphSearch** bestimmt die Zusammenhangskomponenten eines (ungerichteten) Eingabegraphen G .

Algorithmus $\text{CC}(V, E)$

```

1   $k := 0$ 
2  for all  $v \in V, e \in E$  do
3       $\text{cc}(v) := 0$ 
4       $\text{cc}(e) := 0$ 
5  for all  $w \in V$  do
6      if  $\text{cc}(w) = 0$  then
7           $k := k + 1$ 
8          ComputeCC( $k, w$ )
```

Prozedur $\text{ComputeCC}(k, w)$

```

1   $\text{cc}(w) := k$ 
2  Init( $B$ )
3  Insert( $B, w$ )
4  while  $\neg \text{Empty}(B)$  do
5       $u := \text{Element}(B)$ 
6      if  $\exists e = \{u, v\} \in E : \text{cc}(e) = 0$  then
7           $\text{cc}(e) := k$ 
8          if  $\text{cc}(v) = 0$  then
9               $\text{cc}(v) := k$ 
10             Insert( $B, v$ )
11      else
12          Remove( $B$ )
```

Korollar 43. *Der Algorithmus $\text{CC}(V, E)$ bestimmt für einen Graphen $G = (V, E)$ in Linearzeit $\mathcal{O}(n + m)$ sämtliche Zusammenhangskomponenten $G_k = (V_k, E_k)$ von G , wobei $V_k = \{v \in V \mid \text{cc}(v) = k\}$ und $E_k = \{e \in E \mid \text{cc}(e) = k\}$ ist.*

3.4.5 Breiten- und Tiefensuche

Wie wir gesehen haben, findet **Explore**(w) sowohl in Graphen als auch in Digraphen alle von w aus erreichbaren Knoten. Als nächstes zeigen wir, dass **Explore**(w) zu allen von w aus erreichbaren Knoten sogar einen kürzesten Weg findet, falls wir die Datenstruktur B als Warteschlange Q implementieren.

Die Benutzung einer Warteschlange Q zur Speicherung der bereits entdeckten, aber noch nicht abgearbeiteten Knoten bewirkt, dass zuerst alle Nachbarknoten $u_1, \dots, u_k$ des aktuellen Knotens u besucht werden, bevor ein anderer Knoten aktueller Knoten wird. Da die Suche also zuerst in die Breite geht, spricht man von einer *Breitensuche* (kurz *BFS*, engl. *breadth first search*). Den hierbei berechneten Suchwald bezeichnen wir als *Breitensuchwald*.

Bei Benutzung eines Kellers wird dagegen u_1 aktueller Knoten, bevor die übrigen Nachbarknoten von u besucht werden. Daher führt die Benutzung eines Kellers zu einer *Tiefensuche* (kurz *DFS*, engl. *depth first search*). Der berechnete Suchwald heißt dann *Tiefensuchswald*.

Die Breitensuche eignet sich eher für Distanzprobleme wie z.B. das Finden

- kürzester Wege in Graphen und Digraphen,
- längster Wege in Bäumen (siehe Übungen) oder
- kürzester Wege in Distanzgraphen (Dijkstra-Algorithmus).

Dagegen liefert die Tiefensuche interessante Strukturinformationen wie z.B.

- die zweifachen Zusammenhangskomponenten in Graphen,
- die starken Zusammenhangskomponenten in Digraphen oder
- eine topologische Sortierung bei azyklischen Digraphen (s. Übungen).

Wir betrachten zuerst den Breitensuchalgorithmus.

Algorithmus $\text{BFS}(V, E)$

```

1  for all  $v \in V, e \in E$  do
2     $\text{visited}(v) := \text{false}$ 
3     $\text{parent}(v) := \text{nil}$ 
4     $\text{visited}(e) := \text{false}$ 
5  for all  $w \in V$  do
6    if  $\text{visited}(w) = \text{false}$  then  $\text{BFS-Explore}(w)$ 
```

Prozedur $\text{BFS-Explore}(w)$

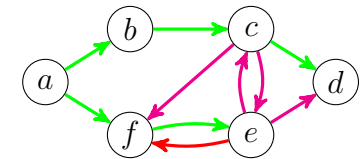
```

1   $\text{visited}(w) := \text{true}$ 
2   $\text{QueueInit}(Q)$ 
3   $\text{Enqueue}(Q, w)$ 
4  while  $\neg \text{QueueEmpty}(Q)$  do
```

```

5     $u := \text{Head}(Q)$ 
6    if  $\exists e = \{u, v\}$  bzw.  $e = (u, v) \in E : \text{visited}(e) = \text{false}$ 
7      then
8         $\text{visited}(e) := \text{true}$ 
9        if  $\text{visited}(v) = \text{false}$  then
10          $\text{visited}(v) := \text{true}$ 
11          $\text{parent}(v) := u$ 
12          $\text{Enqueue}(Q, v)$ 
13    else
14       $\text{Dequeue}(Q)$ 
```

Beispiel 44. *BFS-Explore* generiert bei Aufruf mit dem Startknoten a nebenstehenden Breitensuchswald.



Schlange Q	bes. Knoten	bes. Kante	Typ	Q	bes. Knoten	bes. Kante	Typ
$\leftarrow a \leftarrow$	a	(a, b)	B	c, e, d	c	(c, e)	Q
a, b	a	(a, f)	B	c, e, d	c	(c, f)	Q
a, b, f	a	-	-	c, e, d	c	-	-
b, f	b	(b, c)	B	e, d	e	(e, c)	Q
b, f, c	b	-	-	e, d	e	(e, d)	Q
f, c	f	(f, e)	B	e, d	e	(e, f)	R
f, c, e	f	-	-	e, d	e	-	-
c, e	c	(c, d)	B	d	d	-	-

Satz 45. Sei G ein Graph oder Digraph und sei w Wurzel des von $\text{BFS-Explore}(w)$ berechneten Suchbaumes T . Dann liefert **parent** für jeden Knoten v in T einen kürzesten w - v -Weg $P(v)$.

Beweis. Wir führen Induktion über die kürzeste Weglänge ℓ von w nach v in G .

$\ell = 0$: Dann ist $v = w$ und **parent** liefert einen Weg der Länge 0.

$\ell \rightsquigarrow \ell + 1$: Sei v ein Knoten, der den Abstand $\ell + 1$ von w in G hat. Dann existiert ein Knoten $u \in N^-(v)$ (bzw. $u \in N(v)$) mit Abstand ℓ von w in G hat. Nach IV liefert also **parent** einen w - u -Weg $P(u)$ der Länge ℓ . Da u erst aus Q entfernt wird, nachdem alle Nachfolger von u entdeckt sind, wird v von u oder einem bereits zuvor in Q eingefügten Knoten z entdeckt. Da Q als Schlange organisiert ist, ist $P(u)$ nicht kürzer als $P(z)$. Daher folgt in beiden Fällen, dass $P(v)$ die Länge $\ell + 1$ hat. ■

Wir werden später noch eine Modifikation der Breitensuche kennen lernen, die kürzeste Wege in Graphen mit nichtnegativen Kantenlängen findet (Algorithmus von Dijkstra).

Als nächstes betrachten wir den Tiefensuchalgorithmus.

Algorithmus DFS(V, E)

```

1  for all  $v \in V, e \in E$  do
2    visited( $v$ ) := false
3    parent( $v$ ) := nil
4    visited( $e$ ) := false
5  for all  $w \in V$  do
6    if visited( $w$ ) = false then DFS-Explore( $w$ )
```

Prozedur DFS-Explore(w)

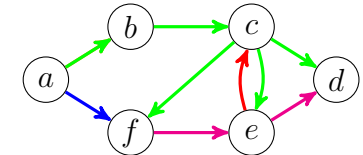
```

1  visited( $w$ ) := true
2  StackInit( $S$ )
3  Push( $S, w$ )
4  while  $\neg$ StackEmpty( $S$ ) do
5     $u :=$  Head( $S$ )
```

```

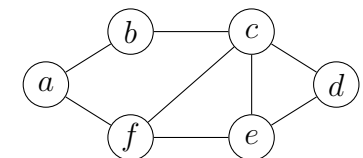
6  if  $\exists e = \{u, v\}$  bzw.  $e = (u, v) \in E : \text{visited}(e) = \text{false}$ 
   then
7    visited( $e$ ) := true
8    if visited( $v$ ) = false then
9      visited( $v$ ) := true
10     parent( $v$ ) :=  $u$ 
11     Push( $S, v$ )
12  else
13    Pop( $S$ )
```

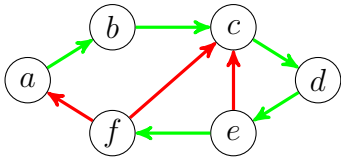
Beispiel 46. Bei Aufruf mit dem Startknoten a generiert die Prozedur **DFS-Explore** nebenstehenden Tiefensuchwald.



	bes.	bes.			bes.	bes.	
Keller S	Knoten	Kante	Typ	S	Knoten	Kante	Typ
$a \leftrightarrow$	a	(a, b)	B	a, b, c	c	(c, f)	B
a, b	b	(b, c)	B	a, b, c, f	f	(f, e)	Q
a, b, c	c	(c, d)	B	a, b, c, f	f	-	-
a, b, c, d	d	-	-	a, b, c	c	-	-
a, b, c	c	(c, e)	B	a, b	b	-	-
a, b, c, e	e	(e, c)	R	a	a	(a, f)	V
a, b, c, e	e	(e, d)	Q	a	a	-	-
a, b, c, e	e	-	-				

Die Tiefensuche auf nebenstehendem Graphen führt auf folgende Klassifikation der Kanten (wobei wir annehmen,





dass die Nachbarknoten in den Adjazenzlisten alphabetisch angeordnet sind):

Keller S	Kante	Typ	Keller S	Kante	Typ
$a \leftrightarrow$	$\{a, b\}$	B	a, b, c, d, e, f	$\{f, c\}$	R
a, b	$\{b, a\}$	-	a, b, c, d, e, f	$\{f, e\}$	-
a, b	$\{b, c\}$	B	a, b, c, d, e, f	-	-
a, b, c	$\{c, b\}$	-	a, b, c, d, e	-	-
a, b, c	$\{c, d\}$	B	a, b, c, d	-	-
a, b, c, d	$\{d, c\}$	-	a, b, c	$\{c, e\}$	-
a, b, c, d	$\{d, e\}$	B	a, b, c	$\{c, f\}$	-
a, b, c, d, e	$\{e, c\}$	R	a, b, c	-	-
a, b, c, d, e	$\{e, d\}$	-	a, b	-	-
a, b, c, d, e	$\{e, f\}$	B	a	$\{a, f\}$	-
a, b, c, d, e, f	$\{f, a\}$	R	a	-	-

<

Die Tiefensuche lässt sich auch rekursiv implementieren. Dies hat den Vorteil, dass kein (expliziter) Keller benötigt wird.

Prozedur DFS-Explore-rec(w)

```

1  visited( $w$ ) := true
2  while
     $\exists e = \{u, v\}$  bzw.  $e = (u, v) \in E : \text{visited}(e) = \text{false}$ 
    do
3    visited( $e$ ) := true
4    if visited( $v$ ) = false then
5      parent( $v$ ) :=  $w$ 
6      DFS-Explore-rec( $v$ )

```

Da **DFS-Explore-rec**(w) zu **parent**(w) zurückspringt, kann auch das Feld **parent**(w) als Keller fungieren. Daher lässt sich die Prozedur auch nicht-rekursiv ohne zusätzlichen Keller implementieren, indem die Rücksprünge explizit innerhalb einer Schleife ausgeführt werden (siehe Übungen).

Bei der Tiefensuche lässt sich der Typ jeder Kante algorithmisch leicht bestimmen, wenn wir noch folgende Zusatzinformationen speichern.

- Ein neu entdeckter Knoten wird bei seinem ersten Besuch grau gefärbt. Sobald er abgearbeitet ist, also bei seinem letzten Besuch, wird er schwarz. Zu Beginn sind alle Knoten weiß.
- Zudem merken wir uns die Reihenfolge, in der die Knoten entdeckt werden, in einem Feld **k**.

Dann lässt sich der Typ jeder Kante $e = (u, v)$ bei ihrem ersten Besuch wie folgt bestimmen:

Baumkante: $\text{farbe}(v) = \text{weiß}$,

Vorwärtskante: $\text{farbe}(v) \neq \text{weiß}$ und $k(v) \geq k(u)$,

Rückwärtskante: $\text{farbe}(v) = \text{grau}$ und $k(v) < k(u)$,

Querkante: $\text{farbe}(v) = \text{schwarz}$ und $k(v) < k(u)$.

Nun lässt sich der Typ jeder Kante $e = (u, v)$ bei ihrem Besuch in Zeile 6 anhand der Werte von **farbe**(v) und **k**(v) wie folgt bestimmen:

Baumkante: $\text{farbe}(v) = \text{weiß}$,

Vorwärtskante: $\text{farbe}(v) \neq \text{weiß}$ und $k(v) \geq k(u)$,

Rückwärtskante: $\text{farbe}(v) = \text{grau}$ und $k(v) < k(u)$,

Querkante: $\text{farbe}(v) = \text{schwarz}$ und $k(v) < k(u)$.

Die folgende Variante von **DFS** berechnet diese Informationen.

Algorithmus DFS(V, E)

```

1   $k := 0$ 
2  for all  $v \in V, e \in E$  do
3     $\text{farbe}(v) := \text{weiß}$ 
4    visited( $e$ ) := false

```

```

5  for all  $u \in V$  do
6    if farbe( $u$ ) = weiß then DFS-Explore( $u$ )

```

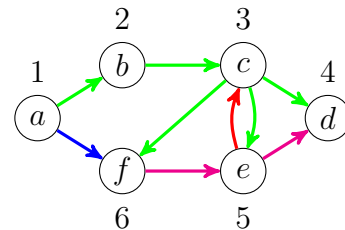
Prozedur DFS-Explore(u)

```

1  farbe( $u$ ) := grau
2   $k := k + 1$ 
3   $k(u) := k$ 
4  while  $\exists e = (u, v) \in E : \text{visited}(e) = \text{false}$  do
5     $\text{visited}(e) := \text{true}$ 
6    if farbe( $v$ ) = weiß then
7      DFS-Explore( $v$ )
8  farbe( $u$ ) := schwarz

```

Beispiel 47. Bei Aufruf mit dem Startknoten a werden die Knoten im nebenstehenden Digraphen von der Prozedur **DFS-Explore** wie folgt gefärbt (die Knoten sind mit ihren k -Werten markiert).



Keller	Farbe	Kante	Typ	Keller	Farbe	Kante	Typ
a	a : grau	(a, b)	B	a, b, c, e	e : schwarz	-	-
a, b	b : grau	(b, c)	B	a, b, c	-	(c, f)	B
a, b, c	c : grau	(c, d)	B	a, b, c, f	f : grau	(f, e)	Q
a, b, c, d	d : grau	-	-	a, b, c, f	f : schwarz	-	-
	d : schwarz			a, b, c	c : schwarz	-	-
a, b, c	-	(c, e)	B	a, b	b : schwarz	-	-
a, b, c, e	e : grau	(e, c)	R	a	-	(a, f)	V
a, b, c, e	-	(e, d)	Q	a	a : schwarz	-	-

Bei der Tiefensuche in ungerichteten Graphen können weder Quer- noch Vorwärtskanten auftreten. Da v beim ersten Besuch einer solchen Kante (u, v) nicht weiß ist und alle grauen Knoten auf dem **parent**-Pfad $P(u)$ liegen, müsste v nämlich bereits schwarz sein. Dies ist aber nicht möglich, da die Kante $\{u, v\}$ in v - u -Richtung noch gar nicht durchlaufen wurde. Folglich sind alle Kanten, die nicht zu einem neuen Knoten führen, Rückwärtskanten. Das Fehlen von Quer- und Vorwärtskanten spielt bei manchen Anwendungen eine wichtige Rolle, etwa bei der Zerlegung eines Graphen G in seine *zweifachen Zusammenhangskomponenten*.

3.4.6 Starke Zusammenhangskomponenten

Sei $G = (V, E)$ ein Digraph. Dann ist leicht zu sehen, dass die Relation

$$S = \{(u, v) \in V \times V \mid \text{es gibt in } G \text{ einen } u\text{-}v\text{-Weg und einen } v\text{-}u\text{-Weg}\}$$

eine Äquivalenzrelation ist. Für $(u, v) \in S$ schreiben wir auch kurz $u \sim v$.

Definition 48. Die durch die Äquivalenzklassen $U_1, \dots, U_k$ von S induzierten Teilgraphen $G[U_1], \dots, G[U_k]$ heißen die **starken Zusammenhangskomponenten** (engl. strongly connected components) von G .

Satz 49. Sei $G = (V, E)$ ein Digraph mit den starken Zusammenhangskomponenten $G[U_1], \dots, G[U_k]$. Dann ist der Digraph (C, D) mit $C = \{1, \dots, k\}$ und

$$D = \{(i, j) \mid 1 \leq i \neq j \leq k \wedge \exists u \in U_i, v \in U_j : (u, v) \in E\}$$

azyklisch.

Beweis. Da der Digraph (C, D) schlingenfrei ist, müsste ein Zyklus mindestens zwei verschiedene Knoten $i \neq j$ enthalten. Dann wären

◁

aber alle Knoten in den beiden Komponenten $G[U_i]$ und $G[U_j]$ gegenseitig erreichbar, d.h. alle Knoten in $U_i \cup U_j$ müssten in derselben Komponente liegen (Widerspruch). ■

Sei $G = (V, E)$ ein Digraph mit zugehörigem Tiefensuchwald W und starken Zusammenhangskomponenten $G[U_1], \dots, G[U_k]$. Den im Suchwald W durch alle Nachfahren von v induzierten Baum mit der Wurzel v bezeichnen wir mit $T(v)$.

Für $i = 1, \dots, k$ sei s_i der erste bei der Tiefensuche innerhalb von U_i besuchte Knoten. Wir bezeichnen s_i als den *Startknoten* von U_i . Dann bilden die Knoten $s_1, \dots, s_k$ ein Repräsentantensystem für die Äquivalenzklassen $U_1, \dots, U_k$.

Sei V_i die Menge aller Knoten in $T(s_i)$, die für keinen echten Nachfahren s_j von s_i in $T(s_j)$ enthalten sind. Es ist klar, dass die Mengen $V_1, \dots, V_k$ eine Partition von V bilden. Wir behaupten, dass die Mengen V_i in U_i enthalten sind und somit $U_i = V_i$ ist.

Angenommen, ein Knoten $u \in V_i$ wäre nicht in U_i , sondern in U_j für ein $j \neq i$ enthalten. Da s_j erst schwarz wird, nachdem alle von s_j aus erreichbaren Knoten besucht wurden, enthält $T(s_j)$ alle Knoten in U_j und somit auch u . Wegen $u \in V_i$ könnte dann s_j kein Nachfahre von s_i sein. Da u in $T(s_i)$ und $T(s_j)$ liegt, müsste s_j also ein Vorfahre von s_i sein. Wegen $u \sim s_j$ impliziert dies jedoch $u \sim s_i$. $\nmid$

Die Mengen U_i lassen sich also leicht bestimmen, falls wir die Startknoten s_i während der Tiefensuche bei ihrem letzten Besuch identifizieren können. In diesem Fall gehören alle seit dem ersten Besuch von s_i besuchten Knoten zu U_i , die nicht schon einem anderen s_j zugeordnet wurden. Um die Startknoten erkennen zu können, betrachten wir die Funktion

$$\mathbf{low}(u) = \min\{\mathbf{k}(v) \mid v \in P(u) \text{ und es gibt einen } u\text{-}v\text{-Weg in } G\}.$$

Da der zu u gehörige Startknoten $s_i \sim u$ von u aus erreichbar ist und auf dem *parent*-Pfad $P(u)$ von u liegt, ist $\mathbf{low}(u) \leq \mathbf{k}(s_i) \leq \mathbf{k}(u)$. Da

es zudem keinen u - v -Weg zu einem Knoten v auf $P(u)$ mit $\mathbf{k}(v) < \mathbf{k}(s_i)$ geben kann (sonst wäre v Startknoten von U_i), ist $\mathbf{low}(u) = \mathbf{k}(s_i)$. Folglich ist

$$\mathbf{low}(u) = \min\{\mathbf{k}(s_i) \mid s_i \in P(u) \text{ und es gibt einen } u\text{-}s_i\text{-Weg in } G\}$$

und nur für die Startknoten nehmen $\mathbf{low}$ und $\mathbf{k}$ den gleichen Wert an. Einfacher als $\mathbf{low}$ lässt sich die Funktion

$$\mathbf{l}(u) = \min\{\mathbf{k}(v) \mid v = u \vee \exists u' \in T(u), v \in T(s_i) : (u', v) \in E\},$$

berechnen, wobei s_i der eindeutig bestimmte Startknoten mit $u \sim s_i$ ist.

Dann gilt $\mathbf{low}(u) \leq \mathbf{l}(u) \leq \mathbf{k}(u)$, wobei $\mathbf{l}(u) = \mathbf{k}(u)$ mit $\mathbf{low}(u) = \mathbf{k}(u)$ (also mit $u = s_i$) gleichbedeutend ist. Ist nämlich $u \neq s_i$, so liegt s_i nicht in $T(u)$ und daher muss jeder u - s_i -Weg eine (Rückwärts- oder Quer-) Kante (u', v) mit $u' \in T(u)$ und $v \in T(s_i) - T(u)$ enthalten. Da jedoch alle von u aus erreichbaren Knoten, deren $\mathbf{k}$ -Wert größer als der von u ist, in $T(u)$ enthalten sind, muss $\mathbf{k}(v) < \mathbf{k}(u)$ und somit $\mathbf{l}(u) < \mathbf{k}(u)$ sein.

Algorithmus SCC(V, E)

```

1  k := 0
2  StackInit( $S$ )
3  for all  $v \in V, e \in E$  do
4    visited( $e$ ) := false
5     $\mathbf{k}(v)$  := 0
6    onStack( $v$ ) := false
7  for all  $u \in V$  do
8    if  $\mathbf{k}(u) = 0$  then Compute-SCC( $u$ )
```

Prozedur Compute-SCC(u)

```

1  k := k + 1
2   $\mathbf{k}(u)$  := k
```

```

3   $\ell(u) := k$ 
4  Push( $S, u$ )
5  while  $\exists e = (u, v) \in E : \text{visited}(e) = \text{false}$  do
6    visited( $e$ ) := true
7    if  $k(v) = 0$  then
8      Compute-SCC( $v$ )
9       $\ell(u) := \min\{\ell(u), \ell(v)\}$ 
10   else if onStack( $v$ ) = true then
11      $\ell(u) := \min\{\ell(u), k(v)\}$ 
12   if  $\ell(u) = k(u)$  then
13     Output-SCC( $u$ )

```

Prozedur Output-SCC(u)

```

1  write(Neue Komponente: )
2  repeat
3     $v := \text{Pop}(S)$ 
4    onStack( $v$ ) := false
5    write( $v$ )
6  until( $v = u$ )

```

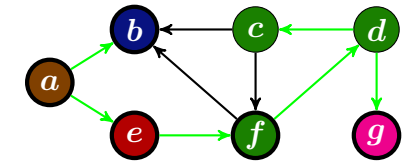
Der Algorithmus **SCC** berechnet für jeden Knoten u den Wert $\ell(u)$ und gibt der Reihe nach die Mengen U_i aus. **SCC** speichert alle entdeckten Knoten, die noch keiner Menge U_i zugeordnet werden konnten, in einem Keller S . Das Feld **onStack** speichert die Information, welche Knoten sich aktuell in S befinden. Besitzt ein Knoten u bei seinem letzten Besuch den Wert $\ell(u) = k(u)$, so wird die Prozedur **Output-SCC(u)** aufgerufen. **Output-SCC(u)** leert den Keller S bis einschließlich u und gibt diese Knoten als neu entdeckte Menge U_i aus.

Dass **SCC** bis zum letzten Besuch eines Knotens u tatsächlich den Wert $\ell(u)$ korrekt berechnet, lässt sich leicht induktiv über die Anzahl der vor u in $T(u)$ schwarz gewordenen Knoten zeigen.

Wird u als erster Knoten in $T(u)$ schwarz, so enthält S genau die Nachbarn von u , die in $T(s_i)$ liegen. Folglich werden die k -Werte dieser Nachbarn bei der Berechnung von $\ell(u)$ in Zeile 11 berücksichtigt.

Sind dagegen vor u bereits andere Knoten in $T(u)$ schwarz geworden, so können wir nach IV annehmen, dass die ℓ -Werte dieser Knoten richtig berechnet wurden. Da diese ℓ -Werte in Zeile 9 an die Elternknoten übermittelt werden, wird dann aber auch $\ell(u)$ korrekt berechnet.

Beispiel 50. Rufen wir **SCC** mit dem Startknoten a für nebenstehenden Digraphen auf, so werden die folgenden starken Zusammenhangskomponenten berechnet (die Startknoten der Komponenten sind dick umrandet).



Keller S	Knoten u	$\ell(u)$	$k(u)$	Kante (u, v)	output
a	a	1	1	(a, b)	-
a, b	b	2	2	-	b
a	a	1	1	(a, e)	-
a, e	e	3	3	(e, f)	-
a, e, f	f	4	4	(f, b)	-
a, e, f	f	4	4	(f, d)	-
a, e, f, d	d	5	5	(d, c)	-
a, e, f, d, c	c	6	6	(c, b)	-
a, e, f, d, c	c	6	6	(c, f)	-
a, e, f, d, c	c	4	6	-	-
a, e, f, d, c	d	4	5	(d, g)	-
a, e, f, d, c, g	g	7	7	-	g
a, e, f, d, c	d	4	5	-	-
a, e, f, d, c	f	4	4	-	f, d, c
a, e	e	3	3	-	e
a	a	1	1	-	a

<

3.5 Kürzeste Pfade in Distanzgraphen

In vielen Anwendungen tritt das Problem auf, einen kürzesten Weg von einem Startknoten s zu einem Zielknoten t in einem Digraphen zu finden, dessen Kanten (u, v) vorgegebene *Längen* $l(u, v)$ haben. Die Länge eines Weges $W = (v_0, \dots, v_\ell)$ ist

$$l(W) = \sum_{i=0}^{\ell-1} l(v_i, v_{i+1}).$$

Die kürzeste Pfadlänge von s nach t wird als *Distanz* $dist(s, t)$ von s zu t bezeichnet,

$$dist(s, t) = \inf\{l(W) \mid W \text{ ist ein } s\text{-}t\text{-Weg}\}.$$

Falls kein s - t -Weg existiert, setzen wir $dist(s, t) = \infty$. Man beachte, dass die Distanz auch den Wert $-\infty$ annehmen kann, falls Kreise mit negativer Länge existieren. In vielen Fällen haben jedoch alle Kanten in E eine nichtnegative Länge $l(u, v) \geq 0$. In diesem Fall nimmt $dist$ keine negativen Werte an und $D = (V, E, l)$ wird dann auch *Distanzgraph* genannt.

3.5.1 Der Dijkstra-Algorithmus

Der Dijkstra-Algorithmus findet einen kürzesten Weg $P(u)$ von s zu allen erreichbaren Knoten u (single-source shortest-path problem). Hierzu führt der Algorithmus eine modifizierte Breitensuche aus. Dabei werden die in Bearbeitung befindlichen Knoten in einer Prioritätswarteschlange U verwaltet. Genauer werden alle Knoten u , zu denen bereits ein s - u -Weg $P(u)$ bekannt ist, zusammen mit der Weglänge g solange in U gespeichert bis $P(u)$ optimal ist. Auf der Datenstruktur U sollten folgende Operationen (möglichst effizient) ausführbar sein.

Init(U): Initialisiert U als leere Menge.

Update(U, u, g): Erniedrigt den Wert von u auf g (nur wenn der aktuelle Wert größer als g ist). Ist u noch nicht in U enthalten, wird u mit dem Wert g zu U hinzugefügt.

RemoveMin(U): Gibt ein Element aus U mit dem kleinsten Wert zurück und entfernt es aus U (ist U leer, wird der Wert **nil** zurückgegeben).

Voraussetzung für die Korrektheit des Algorithmus' ist, dass alle Kanten eine nichtnegative Länge haben. Während der Suche werden bestimmte Kanten $e = (u, v)$ daraufhin getestet, ob $g(u) + l(u, v) < g(v)$ ist. Da in diesem Fall die Kante e auf eine Herabsetzung von $g(v)$ auf den Wert $g(u) + l(u, v)$ „drängt“, wird diese Wertzuweisung als **Relaxation** von e bezeichnet. Welche Kanten auf Relaxation getestet werden, wird beim Dijkstra-Algorithmus durch eine einfache Greedystrategie bestimmt: Wähle u unter allen in Bearbeitung befindlichen Knoten mit minimalem g -Wert.

Algorithmus Dijkstra(V, E, l, s)

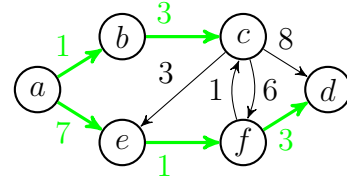
```

1  for all  $v \in V$  do
2     $g(v) := \infty$ 
3     $parent(v) := \text{nil}$ 
4     $done(v) := \text{false}$ 
5   $g(s) := 0$ 
6  Init( $P$ )
7  Update( $P, s, 0$ )
8  while  $u := \text{RemoveMin}(P) \neq \text{nil}$  do
9     $done(u) := \text{true}$ 
10   for all  $v \in N^+(u)$  do
11     if  $done(v) = \text{false} \wedge g(u) + l(u, v) < g(v)$  then
12        $g(v) := g(u) + l(u, v)$ 
13       Update( $P, v, g(v)$ )
14        $parent(v) := u$ 

```

Der Algorithmus speichert die aktuelle Länge des Pfades $P(u)$ in $\mathbf{g}(u)$. Knoten außerhalb des aktuellen Breitensuchbaums T haben den Wert $\mathbf{g}(u) = \infty$. In jedem Schleifendurchlauf wird in Zeile 8 ein Knoten u mit minimalem g -Wert aus U entfernt und als abgearbeitet markiert. Anschließend werden alle von u wegführenden Kanten $e = (u, v)$ auf Relaxation getestet sowie g , U und T gegebenenfalls aktualisiert.

Beispiel 51. Betrachte den nebenstehenden Distanzgraphen G . Bei Ausführung des Dijkstra-Algorithmus' mit dem Startknoten a werden die folgenden kürzesten Wege berechnet.



Inhalt von P	entfernt	besuchte Kanten	Update-Op.
$(a, 0)$	$(a, 0)$	$(a, b), (a, e)$	$(b, 1), (e, 7)$
$(b, 1), (e, 7)$	$(b, 1)$	(b, c)	$(c, 4)$
$(c, 4), (e, 7)$	$(c, 4)$	$(c, d), (c, e), (c, f)$	$(d, 12), (f, 10)$
$(e, 7), (f, 10), (d, 12)$	$(e, 7)$	(e, f)	$(f, 8)$
$(f, 8), (d, 12)$	$(f, 8)$	$(f, c), (f, d)$	$(d, 10)$
$(d, 10)$	$(d, 10)$	—	—

Als nächstes beweisen wir die Korrektheit des Dijkstra-Algorithmus'.

Satz 52. Sei $D = (V, E, l)$ ein Distanzgraph und sei $s \in V$. Dann berechnet **Dijkstra** (V, E, l, s) im Feld **parent** für alle von s aus erreichbaren Knoten $t \in V$ einen kürzesten s - t -Weg $P(t)$.

Beweis. Wir zeigen zuerst, dass alle von s aus erreichbaren Knoten $t \in V$ zu U hinzugefügt werden. Dies folgt aus der Tatsache, dass s zu U hinzugefügt wird, und spätestens dann, wenn ein Knoten u in Zeile 8 aus U entfernt wird, sämtliche Nachfolger von u zu U hinzugefügt werden.

Zudem ist klar, dass $\mathbf{g}(u) \geq \text{dist}(s, u)$ ist, da $P(u)$ im Fall $\mathbf{g}(u) < \infty$ ein s - u -Weg der Länge $\mathbf{g}(u)$ ist. Es bleibt also nur noch zu zeigen,

dass $P(u)$ für jeden aus U entfernten Knoten u ein kürzester s - u -Weg ist, d.h. es gilt $\mathbf{g}(u) \leq \text{dist}(s, u)$.

Hierzu zeigen wir induktiv über die Anzahl k der vor u aus U entfernten Knoten, dass $\mathbf{g}(u) \leq \text{dist}(s, u)$ ist.

$k = 0$: In diesem Fall ist $u = s$ und $P(u)$ hat die Länge $g(u) = 0$.

$k - 1 \rightsquigarrow k$: Sei $W = v_0, \dots, v_\ell = u$ ein kürzester s - u -Weg in G und sei v_i der Knoten mit maximalem Index i auf diesem Weg, der vor u aus P entfernt wird.

Nach IV gilt dann

$$\mathbf{g}(v_i) = \text{dist}(s, v_i). \quad (3.1)$$

Zudem ist

$$\mathbf{g}(v_{i+1}) \leq \mathbf{g}(v_i) + l(v_i, v_{i+1}). \quad (3.2)$$

Da u im Fall $u \neq v_{i+1}$ vor v_{i+1} aus P entfernt wird, ist

$$\mathbf{g}(u) \leq \mathbf{g}(v_{i+1}). \quad (3.3)$$

Daher folgt

$$\begin{aligned} \mathbf{g}(u) &\stackrel{(3.3)}{\leq} \mathbf{g}(v_{i+1}) \stackrel{(3.2)}{\leq} \mathbf{g}(v_i) + l(v_i, v_{i+1}) \\ &\stackrel{(3.1)}{=} \text{dist}(s, v_i) + l(v_i, v_{i+1}) \\ &= \text{dist}(s, v_{i+1}) \leq \text{dist}(s, u). \quad \blacksquare \end{aligned}$$

Um die Laufzeit des Dijkstra-Algorithmus' abzuschätzen, überlegen wir uns zuerst, wie oft die einzelnen Operationen auf der Datenstruktur P ausgeführt werden. Sei $n = \|V\|$ die Anzahl der Knoten und $m = \|E\|$ die Anzahl der Kanten des Eingabegraphen.

- Die **Init**-Operation wird nur einmal ausgeführt.
- Da die while-Schleife für jeden erreichbaren Knoten höchstens einmal durchlaufen wird, wird die **RemoveMin**-Operation höchstens $\min\{n, m\}$ -mal ausgeführt.

- Wie die Prozedur **BFS-Explore** besucht der Dijkstra-Algorithmus jede Kante maximal einmal. Daher wird die **Update**-Operation höchstens m -mal ausgeführt.

Beobachtung 53. Bezeichne $\text{Init}(n)$, $\text{RemoveMin}(n)$ und $\text{Update}(n)$ den Aufwand zum Ausführen der Operationen **Init**, **RemoveMin** und **Update** für den Fall, dass P nicht mehr als n Elemente aufzunehmen hat. Dann ist die Laufzeit des Dijkstra-Algorithmus' durch

$$\mathcal{O}(n + m + \text{Init}(n) + \min\{n, m\} \cdot \text{RemoveMin}(n) + m \cdot \text{Update}(n))$$

beschränkt.

Die Laufzeit hängt also wesentlich davon ab, wie wir die Datenstruktur U implementieren. Falls alle Kanten die gleiche Länge haben, wachsen die Distanzwerte der Knoten monoton in der Reihenfolge ihres (ersten) Besuchs. D.h. wir können U als Warteschlange implementieren. Dies führt wie bei der Prozedur **BFS-Explore** auf eine Laufzeit von $\mathcal{O}(n + m)$.

Für den allgemeinen Fall, dass die Kanten unterschiedliche Längen haben, betrachten wir folgende drei Möglichkeiten.

1. Da die Felder **g** und **done** bereits alle zur Verwaltung von U benötigten Informationen enthalten, kann man auf die (explizite) Implementierung von U auch verzichten. In diesem Fall kostet die **RemoveMin**-Operation allerdings Zeit $\mathcal{O}(n)$, was auf eine Gesamtlaufzeit von $\mathcal{O}(n^2)$ führt.

Dies ist asymptotisch optimal, wenn G relativ dicht ist, also $m = \Omega(n^2)$ Kanten enthält. Ist G dagegen relativ dünn, d.h. $m = o(n^2)$, so empfiehlt es sich, U als Prioritätswarteschlange zu implementieren.

2. Es ist naheliegend, U in Form eines Heaps H zu implementieren. In diesem Fall lässt sich die Operation **RemoveMin** in Zeit $\mathcal{O}(\log n)$ implementieren. Da die Prozedur **Update** einen

linearen Zeitaufwand erfordert, ist es effizienter, sie durch eine **Insert**-Operation zu simulieren. Dies führt zwar dazu, dass derselbe Knoten evtl. mehrmals mit unterschiedlichen Werten in H gespeichert wird. Die Korrektheit bleibt aber dennoch erhalten, wenn wir nur die erste Entnahme eines Knotens aus H beachten und die übrigen ignorieren.

Da für jede Kante höchstens ein Knoten in H eingefügt wird, erreicht H maximal die Größe n^2 und daher sind die Heap-Operationen **Insert** und **RemoveMin** immer noch in Zeit $\mathcal{O}(\log n^2) = \mathcal{O}(\log n)$ ausführbar. Insgesamt erhalten wir somit eine Laufzeit von $\mathcal{O}(n + m \log n)$.

Die Laufzeit von $\mathcal{O}(n + m \log n)$ bei Benutzung eines Heaps ist zwar für dünne Graphen sehr gut, aber für dichte Graphen schlechter als die implizite Implementierung von U mithilfe der Felder **g** und **done**.

3. Als weitere Möglichkeit kann U auch in Form eines so genannten *Fibonacci-Heaps* F implementiert werden. Dieser benötigt nur eine konstante amortisierte Laufzeit $\mathcal{O}(1)$ für die **Update**-Operation und $\mathcal{O}(\log n)$ für die **RemoveMin**-Operation. Insgesamt führt dies auf eine Laufzeit von $\mathcal{O}(m + n \log n)$. Allerdings sind Fibonacci-Heaps erst bei sehr großen Graphen mit mittlerer Dichte schneller.

Eine offene Frage ist, ob es auch einen Algorithmus mit linearer Laufzeit $\mathcal{O}(n + m)$ gibt. Wir fassen die verschiedenen Möglichkeiten zur Implementation der Datenstruktur U in folgender Tabelle zusammen.

	implizit	Heap	Fibonacci-Heap
Init	$\mathcal{O}(1)$	$\mathcal{O}(1)$	$\mathcal{O}(1)$
Update	$\mathcal{O}(1)$	$\mathcal{O}(\log n)$	$\mathcal{O}(1)$
RemoveMin	$\mathcal{O}(n)$	$\mathcal{O}(\log n)$	$\mathcal{O}(\log n)$
Gesamtlaufzeit	$\mathcal{O}(n^2)$	$\mathcal{O}(n + m \log n)$	$\mathcal{O}(m + n \log n)$

3.6 Negative Kantengewichte

In manchen Anwendungen treten negative Kantengewichte auf. Geben die Kantengewichte beispielsweise die mit einer Kante verbundenen Kosten wider, so kann ein Gewinn durch negative Kosten modelliert werden. Gesucht ist dann meist ein kürzester Pfad (also ein kürzester Weg, der jeden Knoten höchstens einmal besucht). Auf diese Weise lassen sich auch längste Pfade in Distanzgraphen berechnen, indem man alle Kantenlängen $\ell(u, v)$ mit -1 multipliziert und in dem resultierenden Graphen einen kürzesten Pfad bestimmt.

Die Komplexität des Problems hängt wesentlich davon ab, ob man (gerichtete) Kreise mit negativer Länge zulässt oder nicht. Falls negative Kreise zugelassen werden, ist das Problem NP-hart. Andernfalls existieren effiziente Algorithmen wie z.B. der Bellman-Ford-Algorithmus (BF-Algorithmus) oder der Bellman-Ford-Moore-Algorithmus (BFM-Algorithmus). Diese Algorithmen lösen das single-source shortest-path Problem mit einer Laufzeit von $\mathcal{O}(nm)$ im schlechtesten Fall.

Der Ford-Algorithmus arbeitet ganz ähnlich wie der Dijkstra-Algorithmus, betrachtet aber jede Kante nicht wie dieser nur einmal, sondern eventuell mehrmals. In seiner einfachsten Form sucht der Algorithmus wiederholt eine Kante $e = (u, v)$ mit

$$g(u) + \ell(u, v) < g(v)$$

und aktualisiert den Wert von $g(v)$ auf $g(u) + \ell(u, v)$ (Relaxation). Die Laufzeit hängt dann wesentlich davon ab, in welcher Reihenfolge die Kanten auf Relaxation getestet werden. Im besten Fall lässt sich eine lineare Laufzeit erreichen (z.B. wenn der zugrunde liegende Graph azyklisch ist). Bei der Bellman-Ford-Variante wird in $\mathcal{O}(nm)$ Schritten ein kürzester Weg von s zu allen erreichbaren Knoten gefunden (sofern keine negativen Kreise existieren).

Wir zeigen induktiv über die Anzahl k der Kanten eines kürzesten s - u -Weges, dass $g(u) = \text{dist}(s, u)$ gilt, falls g für alle Kanten (u, v)

die Dreiecksungleichung $g(v) \leq g(u) + \ell(u, v)$ erfüllt (also keine Relaxationen mehr möglich sind).

Im Fall $k = 0$ ist nämlich $u = s$ und somit $g(s) = 0 = \text{dist}(s, s)$. Im Fall $k > 0$ sei v ein Knoten, dessen kürzester s - v -Weg W aus k Kanten besteht. Dann gilt nach IV für den Vorgänger u von v auf W $g(u) = \text{dist}(s, u)$. Aufgrund der Dreiecksungleichung folgt dann

$$g(v) \leq g(u) + \ell(u, v) = \text{dist}(s, u) + \ell(u, v) = \text{dist}(s, v).$$

Aus dem Beweis folgt zudem, dass nach Relaxation aller Kanten eines kürzesten s - v -Weges W (in der Reihenfolge, in der die Kanten in W durchlaufen werden) den Wert $\text{dist}(s, v)$ hat. Dies gilt auch für den Fall, dass zwischendurch noch weitere Kantenrelaxationen stattfinden.

3.6.1 Der Bellman-Ford-Algorithmus

Diese Variante prüft in $n - 1$ Iterationen jeweils alle Kanten auf Relaxation. Sind in der n -ten Runde noch weitere Relaxationen möglich, muss ein negativer Kreis existieren. Die Laufzeit ist offensichtlich $\mathcal{O}(nm)$ und die Korrektheit folgt leicht durch Induktion über die minimale Anzahl von Kanten eines kürzesten s - t -Weges. Zudem wird bei jeder Relaxation einer Kante (u, v) der Vorgänger u im Feld **parent**(v) vermerkt, so dass sich ein kürzester Pfad von s zu allen erreichbaren Knoten (bzw. ein negativer Kreis) rekonstruieren lässt.

Algorithmus BF(V, E, ℓ, s)

```

1  for all  $v \in V$  do
2       $g(v) := \infty$ 
3      parent( $v$ )  $:= \text{nil}$ 
4   $g(s) := 0$ 
5  for  $i := 1$  to  $n - 1$  do
6      for all  $(u, v) \in E$  do
7          if  $g(u) + \ell(u, v) < g(v)$  then
8               $g(v) := g(u) + \ell(u, v)$ 
```

3.6.2 Der Bellman-Ford-Moore-Algorithmus

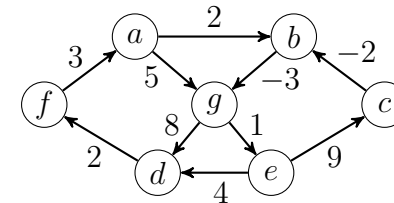
Algorithmus $\text{BFM}(V, E, \ell, s)$

```

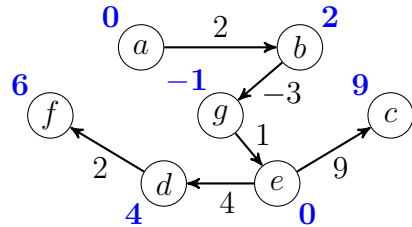
12         inQueue(v) := true
13     if  $i = n$  then
14         error(es gibt einen negativen Kreis)

```

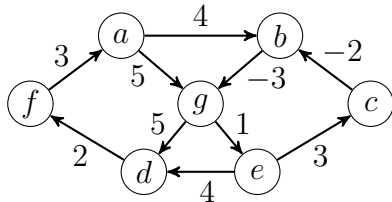
Beispiel 54. Betrachte untenstehenden kantenbewerteten Digraphen mit dem Startknoten a .


$$\begin{array}{ccccccc}
 & \uparrow & & & & & \\
 & | (0, a, \textit{nil}, 0) | & & & & & \\
 & \uparrow & & & & & \\
 & | (1, b, a, 2) | & & & & & \\
 & | (1, g, a, 5) | & | (1, g, b, -1) | & & & & \\
 & & & & & & | (2, d, g, 7) | \\
 & & & & & & | (2, e, g, 0) | \\
 & | (2, e, g, 0) | & & & & & \\
 & | (3, f, d, 9) | & | (3, f, d, 9) | & & & & \\
 & & | (3, c, d, 9) | & | (3, c, d, 9) | & & & \\
 & & | (3, d, e, 4) | & | (3, d, e, 4) | & | (3, d, e, 4) | & & \\
 & & & & & & | (4, f, d, 6) |
 \end{array}$$

Die berechneten Entfernungen mit den zugehörigen **parent**-Pfaden sind in folgendem Suchbaum wiedergegeben:

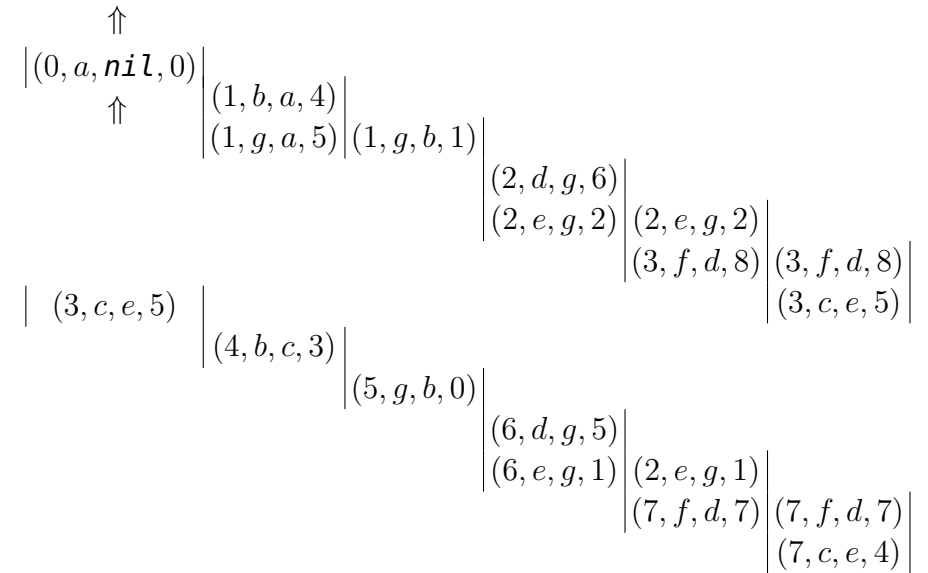


Als nächstes betrachten wir den folgenden Digraphen:

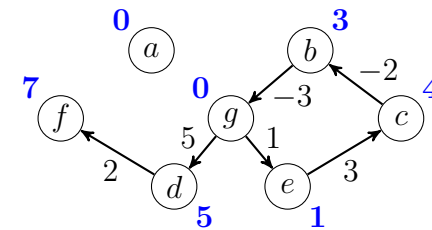


Da dieser einen negativen Kreis enthält, der vom Startknoten aus erreichbar ist, lassen sich die Entfernungen zu allen Knoten, die von

diesem Kreis aus erreichbar sind, beliebig verkleinern.



Da nun der Knoten f mit der Rundenzahl $i = n = 7$ aus der Schlange entnommen wird, bricht der Algorithmus an dieser Stelle mit der Meldung ab, dass negative Kreise existieren. Ein solcher Kreis (im Beispiel: g, e, c, b, g) lässt sich bei Bedarf anhand der **parent**-Funktion aufspüren, indem wir den **parent**-Weg zu f zurückverfolgen: f, d, g, b, c, e, g .



3.7 Berechnung von allen kürzesten Wegen

Sei $G = (V, E)$ ein endlicher Digraph mit $V = \{1, \dots, n\}$. Weiter sei ℓ eine Funktion, die jeder Kante $(i, j) \in E$ einen Kostenwert $\ell(i, j)$ zuweist. Für einen Weg $W = (v_1, v_2, \dots, v_{k-1}, v_k)$ heißen $v_2, \dots, v_{k-1}$ *innere Knoten* von W . Bezeichne $\mathbf{d}_k(i, j)$ die Länge eines kürzesten Weges von i nach j , der mindestens eine Kante durchläuft und dessen inneren Knoten alle aus der Menge $\{1, \dots, k\}$ stammen. Dann gilt

$$\mathbf{d}_0(i, j) = \begin{cases} \ell(i, j), & (i, j) \in E, \\ \infty, & \text{sonst.} \end{cases}$$

und

$$\mathbf{d}_k(i, j) = \min \left\{ \mathbf{d}_{k-1}(i, j), \mathbf{d}_{k-1}(i, k) + \mathbf{d}_{k-1}(k, j) \right\}.$$

3.7.1 Der Floyd-Warshall-Algorithmus

Der Algorithmus von Floyd-Warshall berechnet nichttriviale kürzeste Wege zwischen je zwei Knoten unter der Voraussetzung, dass kein Kreis eine negative Länge hat.

Algorithmus Floyd-Warshall(V, E, ℓ)

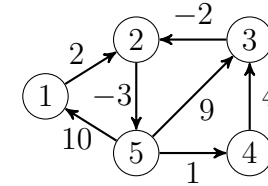
```

1  for  $i := 1$  to  $n$  do
2    for  $j := 1$  to  $n$  do
3      if  $(i, j) \in E$  then  $\mathbf{d}_0(i, j) := \ell(i, j)$  else  $\mathbf{d}_0(i, j) := \infty$ 
4  for  $k := 1$  to  $n$  do
5    for  $i := 1$  to  $n$  do
6      for  $j := 1$  to  $n$  do
7         $\mathbf{d}_k(i, j) = \min \left\{ \mathbf{d}_{k-1}(i, j), \mathbf{d}_{k-1}(i, k) + \mathbf{d}_{k-1}(k, j) \right\}$ 

```

Die Laufzeit ist offenbar $\mathcal{O}(n^3)$. Da die $\mathbf{d}_k$ -Werte nur von den $\mathbf{d}_{k-1}$ -Werten abhängen, ist der Speicherplatzbedarf $\mathcal{O}(n^2)$. Die Existenz negativer Kreise lässt sich daran erkennen, dass mindestens ein Diagonalelement $\mathbf{d}_k(i, i)$ einen negativen Wert erhält.

Beispiel 55. Betrachte folgenden kantenbewerteten Digraphen:



$\mathbf{d}_0$	1	2	3	4	5
1	∞	2	∞	∞	∞
2	∞	∞	∞	∞	-3
3	∞	-2	∞	∞	∞
4	∞	∞	4	∞	∞
5	10	∞	9	1	∞

$\mathbf{d}_1$	1	2	3	4	5
1	∞	2	∞	∞	∞
2	∞	∞	∞	∞	-3
3	∞	-2	∞	∞	∞
4	∞	∞	4	∞	∞
5	10	12	9	1	∞

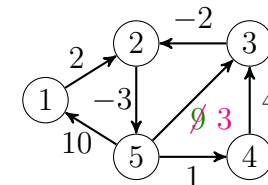
$\mathbf{d}_2$	1	2	3	4	5
1	∞	2	∞	∞	-1
2	∞	∞	∞	∞	-3
3	∞	-2	∞	∞	-5
4	∞	∞	4	∞	∞
5	10	12	9	1	9

$\mathbf{d}_3$	1	2	3	4	5
1	∞	2	∞	∞	-1
2	∞	∞	∞	∞	-3
3	∞	-2	∞	∞	-5
4	∞	2	4	∞	-1
5	10	7	9	1	4

$\mathbf{d}_4$	1	2	3	4	5
1	∞	2	∞	∞	-1
2	∞	∞	∞	∞	-3
3	∞	-2	∞	∞	-5
4	∞	2	4	∞	-1
5	10	3	5	1	0

$\mathbf{d}_5$	1	2	3	4	5
1	9	2	4	0	-1
2	7	0	2	-2	-3
3	5	-2	0	-4	-5
4	9	2	4	0	-1
5	10	3	5	1	0

Als nächstes betrachten wir folgenden Digraphen:



d_0	1	2	3	4	5
1	∞	2	∞	∞	∞
2	∞	∞	∞	∞	-3
3	∞	-2	∞	∞	∞
4	∞	∞	4	∞	∞
5	10	∞	3	1	∞

d_2	1	2	3	4	5
1	∞	2	∞	∞	-1
2	∞	∞	∞	∞	-3
3	∞	-2	∞	∞	-5
4	∞	∞	4	∞	∞
5	10	12	3	1	9

d_4	1	2	3	4	5
1	∞	2	∞	∞	-1
2	∞	∞	∞	∞	-3
3	∞	-2	∞	∞	-5
4	∞	2	4	∞	-1
5	10	1	3	1	-2

d_1	1	2	3	4	5
1	∞	2	∞	∞	∞
2	∞	∞	∞	∞	-3
3	∞	-2	∞	∞	∞
4	∞	∞	4	∞	∞
5	10	12	3	1	∞

d_3	1	2	3	4	5
1	∞	2	∞	∞	-1
2	∞	∞	∞	∞	-3
3	∞	-2	∞	∞	-5
4	∞	2	4	∞	-1
5	10	1	3	1	-2

d_5	1	2	3	4	5
1	9	0	2	0	-3
2	7	-2	0	-2	-5
3	5	-4	-2	-4	-7
4	9	0	2	0	-3
5	8	-1	1	-1	-4

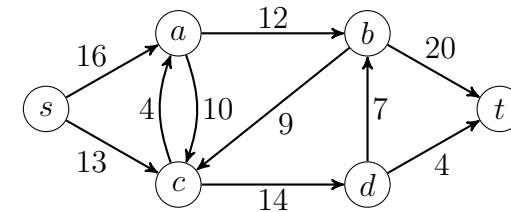
<

3.8 Flüsse in Netzwerken

Definition 56.

- Ein **Netzwerk** $N = (V, E, s, t, c)$ besteht aus einem gerichteten Graphen $G = (V, E)$ mit einer **Quelle** $s \in V$ und einer **Senke** $t \in V$ sowie einer **Kapazitätsfunktion** $c : V \times V \rightarrow \mathbb{N}$.
- Alle Kanten $(u, v) \in E$ müssen positive Kapazität $c(u, v) > 0$ und alle Nichtkanten $(u, v) \notin E$ müssen die Kapazität $c(u, v) = 0$ haben.

Die folgende Abbildung zeigt ein Netzwerk N .

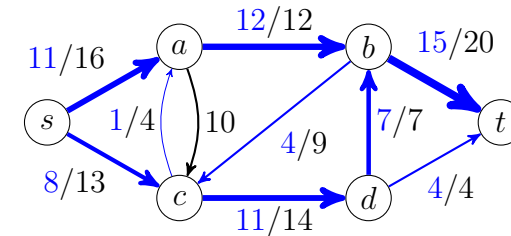


Definition 57. Ein **Fluss** in N ist eine Funktion $f : V \times V \rightarrow \mathbb{Z}$ mit

- $f(u, v) \leq c(u, v)$, „Kapazitätsbedingung“
- $f(u, v) = -f(v, u)$, „Symmetriebedingung“
- Für alle $u \in V - \{s, t\} : \sum_{v \in V} f(u, v) = 0$. „Kontinuitätsbedingung“

Die **Größe** von f ist $|f| = \sum_{v \in V} f(s, v)$.

Die Symmetriebedingung impliziert, dass $f(u, u) = 0$ für alle $u \in V$ ist, d.h. wir können annehmen, dass G schlingenfrei ist. Die folgende Abbildung zeigt einen Fluss f in N .



Wie lässt sich für einen Fluss f in einem Netzwerk N entscheiden, ob er vergrößert werden kann? Diese Frage lässt sich leicht beantworten, falls f der konstante Nullfluss $f = 0$ ist: In diesem Fall genügt es, in $G = (V, E)$ einen Pfad von s nach t zu finden. Andernfalls können wir zu N und f ein Netzwerk N_f konstruieren, so dass f genau dann vergrößert werden kann, wenn sich in N_f der Nullfluss vergrößern lässt.

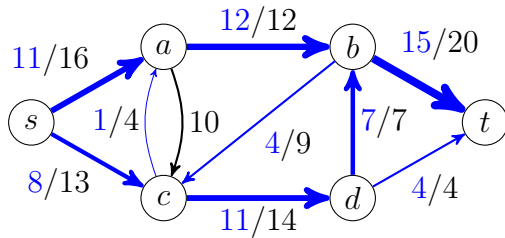
Definition 58. Sei $N = (V, E, s, t, c)$ ein Netzwerk und sei f ein Fluss in N . Das zugeordnete **Restnetzwerk** ist $N_f = (V, E_f, s, t, c_f)$ mit der Kapazität

$$c_f(u, v) = c(u, v) - f(u, v)$$

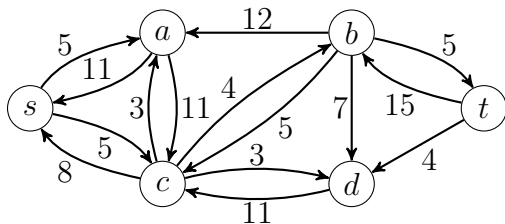
und der Kantenmenge

$$E_f = \{(u, v) \in V \times V \mid c_f(u, v) > 0\}.$$

Zum Beispiel führt der Fluss



auf das folgende Restnetzwerk N_f :



Definition 59. Sei $N_f = (V, E_f, s, t, c_f)$ ein Restnetzwerk. Dann heißt jeder s - t -Pfad P in (V, E_f) **Zunahmepfad** in N_f . Die **Kapazität von P in N_f** ist

$$c_f(P) = \min\{c_f(u, v) \mid (u, v) \text{ liegt auf } P\}$$

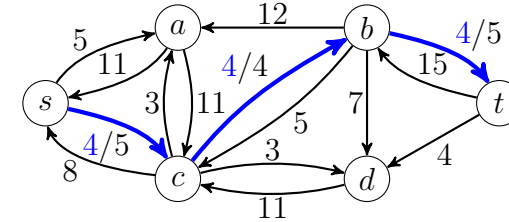
und der zu P gehörige **Fluss in N_f** ist

$$f_P(u, v) = \begin{cases} c_f(P), & (u, v) \text{ liegt auf } P, \\ -c_f(P), & (v, u) \text{ liegt auf } P, \\ 0, & \text{sonst.} \end{cases}$$

$P = (u_0, \dots, u_k)$ ist also genau dann ein Zunahmepfad in N_f , falls

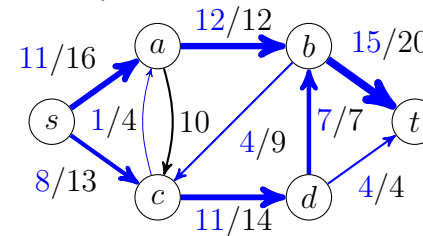
- $u_0 = s$ und $u_k = t$ ist,
- die Knoten $u_0, \dots, u_k$ paarweise verschieden sind
- und $c_f(u_i, u_{i+1}) > 0$ für $i = 0, \dots, k-1$ ist.

Die folgende Abbildung zeigt den zum Zunahmepfad $P = s, c, b, t$ gehörigen Fluss f_P in N_f . Die Kapazität von P ist $c_f(P) = 4$.

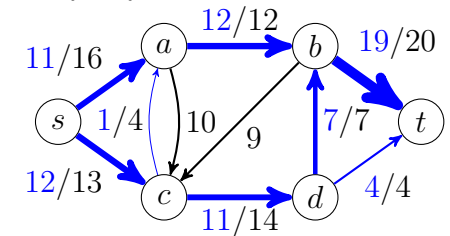


Es ist leicht zu sehen, dass f_P tatsächlich ein Fluss in N_f ist. Durch Addition der beiden Flüsse f und f_P erhalten wir einen größeren Fluss $f' = f + f_P$ in N mit dem Wert $|f'| = |f| + |f_P|$.

Fluss f :



Fluss $f + f_P$:



Nun können wir den **Ford-Fulkerson-Algorithmus** angeben.

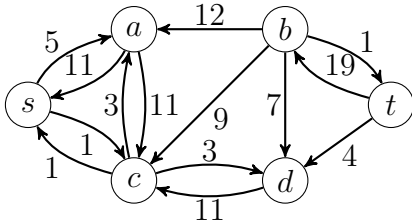
Algorithmus Ford-Fulkerson(V, E, s, t, c)

```

1  for all  $(u, v) \in V \times V$  do
2     $f(u, v) := 0$ 
3  while es gibt einen Zunahmepfad  $P$  fuer  $N_f$  do
4     $f := f + f_P$ 

```

Beispiel 60. Für den neuen Fluss erhalten wir nun folgendes Restnetzwerk:



In diesem existiert kein Zunahmepfad mehr. $\triangleleft$

Um zu beweisen, dass der Algorithmus von Ford-Fulkerson tatsächlich einen Maximalfluss berechnet, zeigen wir, dass es nur dann im Restnetzwerk N_f keinen Zunahmepfad mehr gibt, wenn der Fluss f maximal ist. Hierzu benötigen wir den Begriff des Schnitts.

Definition 61. Sei $N = (V, E, s, t, c)$ ein Netzwerk und seien $S, T \subseteq V$ mit $S \cap T = \emptyset$ und $S \cup T = V$. Dann heißt (S, T) ein **Schnitt** durch N , falls $s \in S$ und $t \in T$ ist. Die **Kapazität** eines Schnittes (S, T) ist

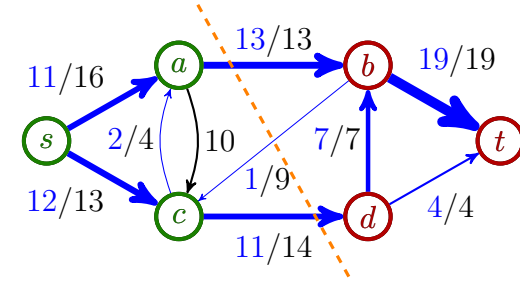
$$c(S, T) = \sum_{u \in S, v \in T} c(u, v).$$

Ist f ein Fluss in N , so heißt

$$f(S, T) = \sum_{u \in S, v \in T} f(u, v)$$

der **Fluss durch den Schnitt** (S, T) .

Beispiel 62. Betrachte den Schnitt (S, T) mit $S = \{s, a, c\}$ und $T = \{b, d, t\}$ in folgendem Netzwerk N mit dem Fluss f :



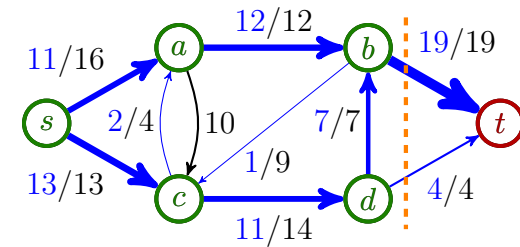
Dieser Schnitt hat die Kapazität

$$c(S, T) = c(a, b) + c(c, d) = 13 + 14 = 27$$

und der Fluss $f(S, T)$ durch diesen Schnitt ist

$$f(S, T) = f(a, b) + f(c, b) + f(c, d) = 13 - 1 + 11 = 23.$$

Dagegen hat der Schnitt (S', T') mit $S' = \{s, a, b, c, d\}$ und $T' = \{t\}$



die Kapazität

$$c(S', T') = c(b, t) + c(d, t) = 19 + 4 = f(b, t) + f(d, t) = |f|,$$

die mit dem Fluss durch diesen Schnitt übereinstimmt. $\triangleleft$

Lemma 63. Für jeden Schnitt (S, T) und jeden Fluss f gilt

$$|f| = f(S, T) \leq c(S, T).$$

Beweis. Die Gleichheit $|f| = f(S, T)$ zeigen wir durch Induktion über $k = \|S\|$.

$k = 1$: In diesem Fall ist $S = \{s\}$ und $T = V - \{s\}$. Daher folgt

$$|f| = \sum_{v \in V - \{s\}} f(s, v) = \sum_{u \in S, v \in T} f(u, v) = f(S, T).$$

$k - 1 \rightsquigarrow k$: Sei (S, T) ein Schnitt mit $\|S\| = k > 1$ und sei $w \in S - \{s\}$. Betrachte den Schnitt (S', T') mit $S' = S - \{w\}$ und $T' = T \cup \{w\}$. Dann gilt

$$f(S, T) = \sum_{u \in S, v \in T} f(u, v) = \sum_{u \in S', v \in T} f(u, v) + \sum_{v \in T} f(w, v)$$

und

$$f(S', T') = \sum_{u \in S', v \in T'} f(u, v) = \sum_{u \in S', v \in T} f(u, v) + \sum_{u \in S'} f(u, w).$$

Wegen $f(w, w) = 0$ ist $\sum_{u \in S'} f(u, w) = \sum_{u \in S} f(u, w)$ und daher

$$\begin{aligned} f(S, T) - f(S', T') &= \sum_{v \in V - S} f(w, v) - \sum_{u \in S} f(u, w) \\ &= \sum_{v \in V} f(w, v) = 0. \end{aligned}$$

Nach Induktionsvoraussetzung folgt somit $f(S, T) = f(S', T') = |f|$.

Schließlich folgt wegen $f(u, v) \leq c(u, v)$ die Ungleichung

$$f(S, T) = \sum_{u \in S, v \in T} f(u, v) \leq \sum_{u \in S, v \in T} c(u, v) = c(S, T). \quad \blacksquare$$

Satz 64 (Min-Cut-Max-Flow-Theorem). Sei f ein Fluss in einem Netzwerk $N = (V, E, s, t, c)$. Dann sind folgende Aussagen äquivalent:

1. f ist maximal.
2. In N_f existiert kein Zunahmepfad.
3. Es gibt einen Schnitt (S, T) mit $c(S, T) = |f|$.

Beweis. Die Implikation „1 $\Rightarrow$ 2“ ist klar, da die Existenz eines Zunahmepfads zu einer Vergrößerung von f führen würde.

Für die Implikation „2 $\Rightarrow$ 3“ betrachten wir den Schnitt (S, T) mit

$$S = \{u \in V \mid u \text{ ist in } N_f \text{ von } s \text{ aus erreichbar}\} \text{ und } T = V - S.$$

Da in N_f kein Zunahmepfad existiert, gilt dann

- $s \in S, t \in T$ und
- $c_f(u, v) = 0$ für alle $u \in S$ und $v \in T$.

Wegen $c_f(u, v) = c(u, v) - f(u, v)$ folgt somit

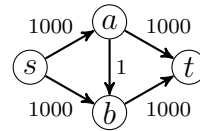
$$|f| = f(S, T) = \sum_{u \in S, v \in T} f(u, v) = \sum_{u \in S, v \in T} c(u, v) = c(S, T).$$

Die Implikation „3 $\Rightarrow$ 1“ ist wiederum klar, da im Fall $c(S, T) = |f|$ für jeden Fluss f' die Abschätzung $|f'| = f'(S, T) \leq c(S, T) = |f|$ gilt. $\blacksquare$

Sei $c_0 = \sum_{v \in V} c(s, v)$ die Kapazität des Schnittes (S, T) mit $S = \{s\}$. Dann durchläuft der Ford-Fulkerson-Algorithmus die while-Schleife höchstens c_0 -mal. Bei jedem Durchlauf ist zuerst das Restnetzwerk N_f und danach ein Zunahmepfad in N_f zu berechnen.

Die Berechnung des Zunahmepfads P kann durch Breitensuche in Zeit $\mathcal{O}(n + m)$ erfolgen. Da sich das Restnetzwerk nur entlang von P ändert, kann es in Zeit $\mathcal{O}(n)$ aktualisiert werden. Jeder Durchlauf benötigt also Zeit $\mathcal{O}(n + m)$, was auf eine Gesamtlaufzeit von $\mathcal{O}(c_0(n + m))$ führt. Da der Wert von c_0 jedoch exponentiell in der Länge der Eingabe (also der Beschreibung des Netzwerkes N) sein kann, ergibt dies keine polynomielle Zeitschranke.

Bei nebenstehendem Netzwerk benötigt Ford-Fulkerson zur Bestimmung des Maximalflusses abhängig von der Wahl der Zunahmepfade zwischen 2 und 2000 Schleifendurchläufe.



Im günstigsten Fall wird zuerst der Zunahmepfad (s, a, t) und dann der Pfad (s, b, t) gewählt. Im ungünstigsten Fall werden abwechselnd die beiden Zunahmepfade (s, a, b, t) und (s, b, a, t) gewählt:

Restnetzwerk	Zunahmepfad	neuer Fluss

Nicht nur in diesem Beispiel lässt sich die exponentielle Laufzeit wie folgt vermeiden:

- Man betrachtet nur Zunahmepfade mit einer geeignet gewählten Mindestkapazität. Dies führt auf eine Laufzeit, die polynomiell in n , m und $\log c_0$ ist.
- Man bestimmt in jeder Iteration einen kürzesten Zunahmepfad im Restnetzwerk. Diese Vorgehensweise wird als *Edmonds-Karp-Strategie* bezeichnet und führt auf eine Laufzeit von $\mathcal{O}(nm^2)$ (unabhängig von der Kapazitätsfunktion).

- Man bestimmt in jeder Iteration einen *blockierenden Fluss* (Algorithmus von Dinic). Dies führt auf eine Laufzeit von $\mathcal{O}(n^2m)$. Ein blockierender Fluss im Restnetzwerk ist ein Fluss mit der Eigenschaft, dass es auf jedem kürzesten s - t -Pfad mindestens eine Kante e gibt, die gesättigt ist (d.h. der Fluss durch e schöpft die Kapazität von e voll aus).