

Vorlesungsskript
Einführung in die
Komplexitätstheorie

Wintersemester 2014/15

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

9. Januar 2015

Inhaltsverzeichnis

1	Einführung	1
2	Rechenmodelle	3
2.1	Deterministische Turingmaschinen	3
2.2	Nichtdeterministische Berechnungen	4
2.3	Zeitkomplexität	5
2.4	Platzkomplexität	6
3	Grundlegende Beziehungen	7
3.1	Robustheit von Komplexitätsklassen	7
3.2	Deterministische Simulationen von nichtdeterministischen Berechnungen	9
3.3	Der Satz von Savitch	10
3.4	Der Satz von Immerman und Szelepcsényi	11
4	Hierarchiesätze	15
4.1	Unentscheidbarkeit mittels Diagonalisierung	15
4.2	Das Gap-Theorem	16
4.3	Zeit- und Platzhierarchiesätze	17
5	Reduktionen	20
5.1	Logspace-Reduktionen	20
5.2	P-vollständige Probleme und polynomielle Schaltkreiskomplexität	22
5.3	NP-vollständige Probleme	24
5.4	NL-vollständige Probleme	28

6	Probabilistische Berechnungen	29
6.1	Die Klassen PP, BPP, RP und ZPP	29
6.2	Anzahl-Operatoren	31
6.3	Reduktion der Fehlerwahrscheinlichkeit	33
6.4	Abschlusseigenschaften von Anzahl-Klassen	35
7	Die Polynomialzeithierarchie	38
8	Turing-Operatoren	40

1 Einführung

In der Komplexitätstheorie werden algorithmische Probleme daraufhin untersucht, welche Rechenressourcen zu ihrer Lösung benötigt werden. Naturgemäß bestehen daher enge Querbezüge zu

- Algorithmen (obere Schranken)
- Automatentheorie (Rechenmodelle)
- Berechenbarkeit (Was ist überhaupt algorithmisch lösbar?)
- Logik (liefert viele algorithmische Probleme, mit ihrer Hilfe kann auch die Komplexität von Problemen charakterisiert werden)
- Kryptografie (Wieviel Rechenressourcen benötigt ein Gegner, um ein Kryptosystem zu brechen?)

Zur weiteren Motivation betrachten wir eine Reihe von konkreten algorithmischen Problemstellungen.

Erreichbarkeitsproblem in Digraphen (Reach):

Gegeben: Ein gerichteter Graph $G = (V, E)$ mit $V = \{1, \dots, n\}$ und $E \subseteq V \times V$.

Gefragt: Gibt es in G einen Weg von Knoten 1 zu Knoten n ?

Zur Erinnerung: Eine Folge (v_1, \dots, v_k) von Knoten heißt **Weg** in G , falls für $j = 1, \dots, k - 1$ gilt: $(v_j, v_{j+1}) \in E$.

Da als Antwort nur “ja” oder “nein” möglich ist, handelt es sich um ein **Entscheidungsproblem**. Ein solches lässt sich formal durch eine Sprache beschreiben, die alle positiven (mit “ja” zu beantwortenden) Problemeingaben enthält:

$$\text{REACH} = \{G \mid \text{in } G \text{ ex. ein Weg von } 1 \text{ nach } n\}.$$

Hierbei setzen wir eine Kodierung von Graphen durch Wörter über einem geeigneten Alphabet Σ voraus. Wir können G beispielsweise durch eine Binärfolge der Länge n^2 kodieren, die aus den n Zeilen der Adjazenzmatrix von G gebildet wird.

Wir entscheiden REACH durch einen Wegsuche-Algorithmus. Dieser markiert nach und nach alle Knoten, die vom Knoten 1 aus erreichbar sind. Hierzu speichert er jeden markierten Knoten solange in einer Menge S bis er sämtliche Nachbarknoten markiert hat. Genauer ist folgendem Algorithmus zu entnehmen:

Algorithmus suche-Weg(G)

```

1  Input: Gerichteter Graph  $G = (V, E)$  mit  $V = \{1, \dots, n\}$ 
2   $S := \{1\}$ 
3  markiere Knoten 1
4  repeat
5      wähle einen Knoten  $u \in S$ 
6       $S := S - \{u\}$ 
7      for all  $(u, v) \in E$  do
8          if  $v$  ist nicht markiert then
9              markiere  $v$ 
10              $S := S \cup \{v\}$ 
11 until  $S = \emptyset$ 
12 if  $n$  ist markiert then accept else reject

```

Es ist üblich, den Ressourcenverbrauch von Algorithmen (wie z.B. Rechenzeit oder Speicherplatz) in Abhängigkeit von der Größe der Problemeingabe zu messen. Falls die Eingabe aus einem Graphen besteht, kann beispielsweise die Anzahl n der Knoten (und/oder die Anzahl m der Kanten) als Bezugsgröße dienen. Der Verbrauch hängt auch davon ab, wie wir die Eingabe kodieren.

Komplexitätsbetrachtungen:

- REACH ist in Zeit n^3 entscheidbar.

1 Einführung

- REACH ist nichtdeterministisch in Platz $\log n$ entscheidbar (und daher deterministisch in Platz $\log^2 n$; Satz von Savitch).

Als nächstes betrachten wir das Problem, einen maximalen Fluss in einem Netzwerk zu bestimmen.

Maximaler Fluß (MaxFlow):

Gegeben: Ein gerichteter Graph $G = (V, E)$ mit $V = \{1, \dots, n\}$, $E \subseteq V \times V$ und einer Kapazitätsfunktion $c : E \rightarrow \mathbb{N}$.

Gesucht: Ein Fluss $f : E \rightarrow \mathbb{N}$ von 1 nach n in G , d.h.

- $\forall e \in E : f(e) \leq c(e)$ und
- $\forall v \in V - \{1, n\} : \sum_{(v,u) \in E} f(v, u) = \sum_{(u,v) \in E} f(u, v)$,

mit maximalem Wert $w(f) = \sum_{(1,v) \in E} f(1, v)$.

Da hier nach einer Lösung (Fluss) mit optimalem Wert gesucht wird, handelt es sich um ein **Optimierungsproblem** (genauer: Maximierungsproblem). Im Gegensatz hierzu wird bei vielen Entscheidungsproblemen nach der Existenz einer Lösung (mit gewissen Eigenschaften) gefragt.

Komplexitätsbetrachtungen:

- MAXFLOW ist in Zeit n^5 lösbar.
- MAXFLOW ist in Platz n^2 lösbar.

Das folgende Problem scheint zwar auf den ersten Blick nur wenig mit dem Problem MAXFLOW gemein zu haben. In Wirklichkeit entpuppt es sich jedoch als ein Spezialfall von MAXFLOW.

Perfektes Matching in bipartiten Graphen (Matching):

Gegeben: Ein bipartiter Graph $G = (U, W, E)$ mit $U \cap W = \emptyset$ und $e \cap U \neq \emptyset \neq e \cap W$ für alle Kanten $e \in E$.

Gefragt: Besitzt G ein perfektes Matching?

Zur Erinnerung: Eine Kantenmenge $M \subseteq E$ heißt **Matching**, falls für alle Kanten $e, e' \in M$ mit $e \neq e'$ gilt: $e \cap e' = \emptyset$. Gilt zudem $\|M\| = n/2$, so heißt M **perfekt** (n ist die Knotenzahl von G).

Komplexitätsbetrachtungen:

- MATCHING ist in Zeit n^3 entscheidbar.
- MATCHING ist in Platz n^2 entscheidbar.

Die bisher betrachteten Probleme können in deterministischer Polynomialzeit gelöst werden und gelten daher als effizient lösbar. Zum Schluss dieses Abschnitts betrachten wir ein Problem, für das vermutlich nur ineffiziente Algorithmen existieren. Wie üblich bezeichnen wir die Gruppe aller Permutationen auf der Menge $\{1, \dots, n\}$ mit S_n .

Travelling Salesman Problem (TSP):

Gegeben: Eine symmetrische $n \times n$ -Distanzmatrix $D = (d_{ij})$ mit $d_{ij} \in \mathbb{N}$.

Gesucht: Eine kürzeste Rundreise, d.h. eine Permutation $\pi \in S_n$ mit minimalem Wert $w(\pi) = \sum_{i=1}^n d_{\pi(i), \pi(i+1)}$, wobei wir $\pi(n+1) = \pi(1)$ setzen.

Komplexitätsbetrachtungen:

- TSP ist in Zeit $n!$ lösbar (Ausprobieren aller Rundreisen).
- TSP ist in Platz n lösbar (mit demselben Algorithmus).
- Durch dynamisches Programmieren* lässt sich TSP in Zeit $n^2 \cdot 2^n$ lösen, der Platzverbrauch erhöht sich dabei jedoch auf $n \cdot 2^n$ (siehe Übungen).

*Hierzu berechnen wir für alle Teilmengen $S \subseteq \{2, \dots, n\}$ und alle $j \in S$ die Länge $l(S, j)$ eines kürzesten Pfades von 1 nach j , der alle Städte in S genau einmal besucht.

2 Rechenmodelle

2.1 Deterministische Turingmaschinen

Definition 1 (Mehrband-Turingmaschine).

Eine **deterministische k -Band-Turingmaschine** (**k -DTM** oder einfach **DTM**) ist ein 5-Tupel $M = (Q, \Sigma, \Gamma, \delta, q_0)$. Dabei ist

- Q eine endliche Menge von **Zuständen**,
- Σ eine endliche Menge von Symbolen (das **Eingabealphabet**) mit $\sqcup, \triangleright \notin \Sigma$ (\sqcup heißt **Blank** und \triangleright heißt **Anfangssymbol**,
- Γ das **Arbeitsalphabet** mit $\Sigma \cup \{\sqcup, \triangleright\} \subseteq \Gamma$,
- $\delta : Q \times \Gamma^k \rightarrow (Q \cup \{q_h, q_{ja}, q_{nein}\}) \times (\Gamma \times \{L, R, N\})^k$ die **Überföhrungsfunktion** (q_h heißt **Haltezustand**, q_{ja} **akzeptierender** und q_{nein} **verwerfender Endzustand**
- und q_0 der **Startzustand**.

Befindet sich M im Zustand $q \in Q$ und stehen die Schreib-Lese-Köpfe auf Feldern mit den Inschriften a_1, \dots, a_k (a_i auf Band i), so geht M bei Ausführung der Anweisung $\delta : (q, a_1, \dots, a_k) \mapsto (q', a'_1, D_1, \dots, a'_k, D_k)$ in den Zustand q' über, ersetzt auf Band i das Symbol a_i durch a'_i und bewegt den Kopf gemäß D_i (im Fall $D_i = L$ um ein Feld nach links, im Fall $D_i = R$ um ein Feld nach rechts und im Fall $D_i = N$ wird der Kopf nicht bewegt).

Außerdem verlangen wir von δ , dass für jede Anweisung $(q, a_1, \dots, a_k) \mapsto (q', a'_1, D_1, \dots, a'_k, D_k)$ mit $a_i = \triangleright$ die Bedingung $a'_i = \triangleright$ und $D_i = R$ erfüllt ist (d.h. das Anfangszeichen \triangleright darf nicht durch ein anderes Zeichen überschrieben werden und der Kopf muss nach dem Lesen von \triangleright immer nach rechts bewegt werden).

Definition 2. Eine **Konfiguration** ist ein $(2k + 1)$ -Tupel $K = (q, u_1, v_1, \dots, u_k, v_k) \in Q \times (\Gamma^* \times \Gamma^+)^k$ und besagt, dass

- q der momentane Zustand und
- $u_i v_i \sqcup \sqcup \dots$ die Inschrift des i -ten Bandes ist, und dass
- sich der Kopf auf Band i auf dem ersten Zeichen von v_i befindet.

Definition 3. Eine Konfiguration $K' = (q', u'_1, v'_1, \dots, u'_k, v'_k)$ heißt **Folgekonfiguration** von $K = (q, u_1, a_1 v_1, \dots, u_k, a_k v_k)$ (kurz: $K \xrightarrow[M]{}$ K'), falls eine Anweisung

$$(q, a_1, \dots, a_k) \mapsto (q', a'_1, D_1, \dots, a'_k, D_k)$$

in δ und $b_1, \dots, b_k \in \Gamma$ existieren, so dass für $i = 1, \dots, k$ jeweils eine der folgenden drei Bedingungen gilt:

1. $D_i = N$, $u'_i = u_i$ und $v'_i = a'_i v_i$,
2. $D_i = L$, $u_i = u'_i b_i$ und $v'_i = b_i a'_i v_i$,
3. $D_i = R$, $u'_i = u_i a'_i$ und $v'_i = \begin{cases} \sqcup, & v_i = \varepsilon, \\ v_i, & \text{sonst,} \end{cases}$

Wir schreiben $K \xrightarrow[M]{t} K'$, falls Konfigurationen K_0, \dots, K_t existieren mit $K_0 = K$ und $K_t = K'$, sowie $K_i \xrightarrow[M]{}$ K_{i+1} für $i = 0, \dots, t - 1$. Die reflexive, transitive Hülle von $\xrightarrow[M]{}$ bezeichnen wir mit $\xrightarrow[M]{*}$, d.h. $K \xrightarrow[M]{*} K'$ bedeutet, dass ein $t \geq 0$ existiert mit $K \xrightarrow[M]{t} K'$.

Definition 4. Sei $x \in \Sigma^*$ eine Eingabe. Die zugehörige **Startkonfiguration** ist

$$K_x = (q_0, \varepsilon, \underbrace{\triangleright x, \varepsilon, \triangleright, \dots, \varepsilon, \triangleright}_{(k-1)\text{-mal}}).$$

Definition 5. Eine Konfiguration $K = (q, u_1, v_1, \dots, u_k, v_k)$ mit $q \in \{q_h, q_{ja}, q_{nein}\}$ heißt **Endkonfiguration**. Im Fall $q = q_{ja}$ (bzw. $q = q_{nein}$) heißt K **akzeptierende** (bzw. **verwerfende**) **Endkonfiguration**.

Definition 6.

Eine DTM M **hält** bei Eingabe $x \in \Sigma^*$ (kurz: $M(x)$ hält), falls es eine Endkonfiguration $K = (q, u_1, v_1, \dots, u_k, v_k)$ gibt mit

$$K_x \xrightarrow[M]{*} K.$$

Weiter definieren wir das **Resultat** $M(x)$ der Rechnung von M bei Eingabe x ,

$$M(x) = \begin{cases} \text{ja,} & M(x) \text{ hält im Zustand } q_{\text{ja}}, \\ \text{nein,} & M(x) \text{ hält im Zustand } q_{\text{nein}}, \\ y, & M(x) \text{ hält im Zustand } q_h, \\ \uparrow \text{ (undefiniert),} & \text{sonst.} \end{cases}$$

Dabei ergibt sich y aus $u_k v_k$, indem das erste Symbol \triangleright und sämtliche Blanks am Ende entfernt werden, d. h. $u_k v_k = \triangleright y \sqcup^i$ für ein $i \geq 0$. Für $M(x) = \text{ja}$ sagen wir auch „ $M(x)$ akzeptiert“ und für $M(x) = \text{nein}$ „ $M(x)$ verwirft“.

Definition 7. Die von einer DTM M **akzeptierte Sprache** ist

$$L(M) = \{x \in \Sigma^* \mid M(x) \text{ akzeptiert}\}.$$

Eine DTM, die eine Sprache L akzeptiert, darf also bei Eingaben $x \notin L$ unendlich lange rechnen. In diesem Fall heißt L **semi-entscheidbar** (oder **rekursiv aufzählbar**). Dagegen muss eine DTM, die eine Sprache L entscheidet, bei jeder Eingabe halten.

Definition 8. Sei $L \subseteq \Sigma^*$. Eine DTM M **entscheidet** L , falls für alle $x \in \Sigma^*$ gilt:

$$\begin{aligned} x \in L &\Rightarrow M(x) \text{ akz.} \\ x \notin L &\Rightarrow M(x) \text{ verw.} \end{aligned}$$

In diesem Fall heißt L **entscheidbar** (oder **rekursiv**).

Definition 9. Sei $f : \Sigma^* \rightarrow \Sigma^*$ eine Funktion. Eine DTM M **berechnet** f , falls für alle $x \in \Sigma^*$ gilt:

$$M(x) = f(x).$$

f heißt dann **berechenbar** (oder **rekursiv**).

Aus dem Grundstudium wissen wir, dass eine nichtleere Sprache $L \subseteq \Sigma^*$ genau dann semi-entscheidbar ist, wenn eine berechenbare Funktion $f : \Sigma^* \rightarrow \Sigma^*$ existiert, deren Bild $\text{range}(f) = \{f(x) \mid x \in \Sigma^*\}$ die Sprache L ist.

2.2 Nichtdeterministische Berechnungen

Anders als eine DTM, für die in jeder Konfiguration höchstens eine Anweisung ausführbar ist, hat eine nichtdeterministische Turingmaschine in jedem Rechenschritt die Wahl unter einer endlichen Anzahl von Anweisungen.

Definition 10. Eine **nichtdeterministische k -Band-Turingmaschine** (kurz **k -NTM** oder einfach **NTM**) ist ein 5-Tupel $M = (Q, \Sigma, \Gamma, \delta, q_0)$, wobei Q, Σ, Γ, q_0 genau wie bei einer k -DTM definiert sind und

$$\delta : Q \times \Gamma^k \rightarrow \mathcal{P}(Q \cup \{q_h, q_{\text{ja}}, q_{\text{nein}}\} \times (\Gamma \times \{R, L, N\})^k)$$

die Eigenschaft hat, dass für $(q', a'_1, D_1, \dots, a'_k, D_k) \in \delta(q, a_1, \dots, a_k)$ im Fall $a_i = \triangleright$ immer $a'_i = \triangleright$ und $D_i = R$ gilt.

Die Begriffe **Konfiguration**, **Start-** und **Endkonfiguration** übertragen sich unmittelbar von DTMs auf NTMs. Der Begriff der **Folgekonfiguration** lässt sich übertragen, indem wir $\delta(q, a_1, \dots, a_k) = (q', a'_1, D_1, \dots, a'_k, D_k)$ durch $(q', a'_1, D_1, \dots, a'_k, D_k) \in \delta(q, a_1, \dots, a_k)$ ersetzen (in beiden Fällen schreiben wir auch oft

$$\delta : (q, a_1, \dots, a_k) \mapsto (q', a'_1, D_1, \dots, a'_k, D_k)$$

oder einfach $(q, a_1, \dots, a_k) \mapsto (q', a'_1, D_1, \dots, a'_k, D_k)$.

Wir werden NTMs nur zum Erkennen von Sprachen (d.h. als Akzeptoren) und nicht zum Berechnen von Funktionen benutzen.

Definition 11. Sei M eine NTM.

- Wir sagen $M(x)$ **akzeptiert**, falls $M(x)$ nur endlich lange Rechnungen ausführt und eine akzeptierende Endkonfiguration K existiert mit $K_x \rightarrow^* K$.
- Akzeptiert $M(x)$ nicht und hat $M(x)$ nur endlich lange Rechnungen, so **verwirft** $M(x)$.
- Falls $M(x)$ unendlich lange Rechnungen ausführt, ist $M(x) = \uparrow$ (undefiniert).
- Die von M **akzeptierte Sprache** ist

$$L(M) = \{x \in \Sigma^* \mid M(x) \text{ akzeptiert}\}.$$

- M **entscheidet** $L(M)$, falls M alle Eingaben $x \notin L(M)$ verwirft.

2.3 Zeitkomplexität

Der Zeitverbrauch $time_M(x)$ einer Turingmaschine M bei Eingabe x ist die maximale Anzahl an Rechenschritten, die M ausgehend von der Startkonfiguration K_x ausführen kann (bzw. undefiniert oder ∞ , falls unendlich lange Rechnungen existieren).

Definition 12.

- Sei M eine TM (d.h. eine DTM oder NTM) und sei $x \in \Sigma^*$ eine Eingabe. Dann ist

$$time_M(x) = \max\{t \geq 0 \mid \exists K : K_x \rightarrow^t K\}$$

die **Rechenzeit** von M bei Eingabe x , wobei $\max \mathbb{N} = \infty$ ist.

- Sei $t : \mathbb{N} \rightarrow \mathbb{N}$ eine Funktion. Dann ist M **$t(n)$ -zeitbeschränkt**, falls für alle $n \geq 0$ und alle $x \in \Sigma^*$ mit $|x| \leq n$ gilt:

$$time_M(x) \leq t(n).$$

Alle Sprachen, die in (nicht-)deterministischer Zeit $t(n)$ entscheidbar sind, fassen wir in den Komplexitätsklassen

$$DTIME(t(n)) = \{L(M) \mid M \text{ ist eine } t(n)\text{-zeitbeschränkte DTM}\}$$

bzw.

$$NTIME(t(n)) = \{L(M) \mid M \text{ ist eine } t(n)\text{-zeitbeschränkte NTM}\}$$

zusammen. Ferner sei

$$FTIME(t(n)) = \left\{ f \mid \begin{array}{l} f \text{ wird von einer } t(n)\text{-zeitbe-} \\ \text{schränkten DTM berechnet} \end{array} \right\}.$$

Für eine Klasse F von Funktionen $t : \mathbb{N} \rightarrow \mathbb{N}$ sei $DTIME(F) = \bigcup_{t \in F} DTIME(t(n))$. $NTIME(F)$ und $FTIME(F)$ sind analog definiert. Die Klasse $\mathcal{O}(n^{\mathcal{O}(1)})$ aller polynomiell beschränkten Funktionen bezeichnen wir mit $\text{poly}(n)$. Die wichtigsten Zeitkomplexitätsklassen sind

$$\begin{aligned} \text{LINTIME} &= DTIME(\mathcal{O}(n)) &= \bigcup_{c \geq 1} DTIME(cn + c) && \text{„Linearzeit“}, \\ \text{P} &= DTIME(\text{poly}(n)) &= \bigcup_{c \geq 1} DTIME(n^c + c) && \text{„Polynomialzeit“}, \\ \text{E} &= DTIME(2^{\mathcal{O}(n)}) &= \bigcup_{c \geq 1} DTIME(2^{cn+c}) && \text{„Lineare Exponentialzeit“}, \\ \text{EXP} &= DTIME(2^{\text{poly}(n)}) &= \bigcup_{c \geq 1} DTIME(2^{n^c+c}) && \text{„Exponentialzeit“}. \end{aligned}$$

Die Klassen NP, NE, NEXP und FP, FE, FEXP sind analog definiert.

2.4 Platzkomplexität

Zur Definition von Platzkomplexitätsklassen verwenden wir so genannte Offline-Turingmaschinen und Transducer. Diese haben die Eigenschaft, dass sie das erste Band nur als Eingabeband (also nur zum Lesen) bzw. das k -te Band nur als Ausgabeband (also nur zum Schreiben) benutzen. Der Grund für diese Einschränkungen liegt darin, sinnvolle Definitionen für Komplexitätsklassen mit einem sublinearen Platzverbrauch zu erhalten.

Definition 13. Eine TM M heißt **Offline-TM**, falls für jede Anweisung $(q, a_1, \dots, a_k) \mapsto (q', a'_1, D_1, \dots, a'_k, D_k)$ die Bedingung

$$a'_1 = a_1 \wedge [a_1 = \sqcup \Rightarrow D_1 = L]$$

gilt. Gilt weiterhin immer $D_k \neq L$ und ist M eine DTM, so heißt M **Transducer**.

Dies bedeutet, dass eine Offline-TM nicht auf das Eingabeband schreiben darf (*read-only*). Beim Transducer dient das letzte Band als Ausgabeband, auch hier können keine Berechnungen durchgeführt werden (*write-only*).

Der Zeitverbrauch $time_M(x)$ von Offline-TMs und von Transducern ist genauso definiert wie bei DTMs. Als nächstes definieren wir den Platzverbrauch einer TM als die Anzahl aller während der Rechnung besuchten Bandfelder.

Definition 14.

- a) Sei M eine TM und sei $x \in \Sigma^*$ eine Eingabe mit $time_M(x) < \infty$. Dann ist

$$space_M(x) = \max\{s \geq 1 \mid \exists K = (q, u_1, v_1, \dots, u_k, v_k) \\ \text{mit } K_x \rightarrow^* K \text{ und } s = \sum_{i=1}^k |u_i v_i|\}$$

der **Platzverbrauch** von M bei Eingabe x . Für eine Offline-TM ersetzen wir $\sum_{i=1}^k |u_i v_i|$ durch $\sum_{i=2}^k |u_i v_i|$ und für einen Transducer durch $\sum_{i=2}^{k-1} |u_i v_i|$.

- b) Sei $s : \mathbb{N} \rightarrow \mathbb{N}$. Dann ist M **$s(n)$ -platzbeschränkt**, falls für alle $n \geq 0$ und alle $x \in \Sigma^*$ mit $|x| \leq n$ gilt:

$$space_M(x) \leq s(n) \text{ und } time_M(x) < \infty.$$

D.h., $space_M(x)$ ist undefiniert, falls $time_M(x) = \infty$ ist.

Alle Sprachen, die in (nicht-) deterministischem Platz $s(n)$ entscheidbar sind, fassen wir in den Komplexitätsklassen

$$DSPACE(s(n)) = \left\{ L(M) \mid \begin{array}{l} M \text{ ist eine } s(n)\text{-platzbe-} \\ \text{schränkte Offline-DTM} \end{array} \right\}$$

bzw.

$$NSPACE(s(n)) = \left\{ L(M) \mid \begin{array}{l} M \text{ ist eine } s(n)\text{-platzbe-} \\ \text{schränkte Offline-NTM} \end{array} \right\}$$

zusammen. Ferner sei

$$FSPACE(s(n)) = \left\{ f \mid \begin{array}{l} f \text{ wird von einem } s(n)\text{-platzbe-} \\ \text{schränkten Transducer berechnet} \end{array} \right\}.$$

Die wichtigsten Platzkomplexitätsklassen sind

$$\begin{aligned} L &= LOGSPACE = DSPACE(O(\log n)) \\ L^c &= DSPACE(O(\log^c n)) \\ Linspace &= DSPACE(O(n)) \\ PSPACE &= DSPACE(\text{poly}(n)) \\ ESPACE &= DSPACE(2^{\mathcal{O}(n)}) \\ EXPSPACE &= DSPACE(2^{\text{poly}(n)}) \end{aligned}$$

Die Klassen NL, NLinspace und NPSPACE, sowie FL, FLinspace und FPSPACE sind analog definiert, wobei NPSPACE mit PSPACE zusammenfällt (wie wir bald sehen werden).

3 Grundlegende Beziehungen

In diesem Kapitel leiten wir die wichtigsten Inklusionsbeziehungen zwischen deterministischen und nichtdeterministischen Platz- und Zeitkomplexitätsklassen her. Zuerst befassen wir uns jedoch mit Robustheitseigenschaften dieser Klassen.

3.1 Robustheit von Komplexitätsklassen

Wir zeigen zuerst, dass platzbeschränkte TMs nur ein Arbeitsband benötigen.

Lemma 15 (Bandreduktion).

Zu jeder $s(n)$ -platzbeschränkten Offline-DTM M ex. eine $s(n)$ -platzbeschränkte Offline-2-DTM M' mit $L(M') = L(M)$.

Beweis. Sei $M = (Q, \Sigma, \Gamma, \delta, q_0)$ eine Offline- k -DTM mit $k \geq 3$. Betrachte die Offline-2-DTM $M' = (Q', \Sigma, \Gamma', \delta', q'_0)$ mit $\Gamma' = \Gamma \cup (\Gamma \cup \hat{\Gamma})^{k-1}$, wobei $\hat{\Gamma}$ für jedes $a \in \Gamma$ die markierte Variante \hat{a} enthält. M' hat dasselbe Eingabeband wie M , speichert aber die Inhalte von $(k-1)$ übereinander liegenden Feldern der Arbeitsbänder von M auf einem Feld ihres Arbeitsbandes. Zur Speicherung der Kopfpositionen von M werden Markierungen benutzt.

Initialisierung: In den ersten beiden Rechenschritten erzeugt M' auf ihrem Arbeitsband (Band 2) $k-1$ Spuren, die jeweils mit dem markierten Anfangszeichen $\hat{\triangleright}$ initialisiert werden:

$$K_x = (q'_0, \varepsilon, \triangleright x, \varepsilon, \triangleright) \xrightarrow{M'} (q'_1, \triangleright, x, \triangleright, \sqcup) \xrightarrow{M'} (q'_2, \varepsilon, \triangleright x, \triangleright, \begin{pmatrix} \hat{\triangleright} \\ \vdots \end{pmatrix})$$

Simulation: M' simuliert einen Rechenschritt von M , indem sie den Kopf auf dem Arbeitsband soweit nach rechts bewegt, bis sie alle $(k-1)$ markierten Zeichen a_2, \dots, a_k gefunden hat. Diese speichert sie neben dem aktuellen Zustand q von M in ihrem Zustand. Während M' den Kopf wieder nach links bewegt, führt M' folgende Aktionen durch: Ist a_1 das von M' (und von M) gelesene Eingabezeichen und ist $\delta(q, a_1, a_2, \dots, a_k) = (q', a_1, D_1, a'_2, D_2, \dots, a'_k, D_k)$, so bewegt M' den Eingabekopf gemäß D_1 , ersetzt auf dem Arbeitsband die markierten Zeichen a_i durch a'_i und verschiebt deren Marken gemäß D_i , $i = 2, \dots, k$.

Akzeptanzverhalten: M' akzeptiert genau dann, wenn M akzeptiert.

Offenbar gilt nun $L(M') = L(M)$ und $space_{M'}(x) \leq space_M(x)$. ■

In den Übungen wird gezeigt, dass die Sprache der Palindrome durch eine 2-DTM zwar in Linearzeit entscheidbar ist, eine 1-DTM hierzu jedoch Zeit $\Omega(n^2)$ benötigt. Tatsächlich lässt sich jede $t(n)$ -zeitbeschränkte k -DTM M von einer 1-DTM M' in Zeit $O(t(n)^2)$ simulieren. Bei Verwendung einer 2-DTM ist die Simulation sogar in Zeit $O(t(n) \log t(n))$ durchführbar (siehe Übungen). Als nächstes wenden wir uns wichtigen Robustheitseigenschaften von Platz- und Zeitkomplexitätsklassen zu.

Satz 16 (Lineare Platzkompression und Beschleunigung).

Für alle $c > 0$ gilt

- i) $DSPACE(s(n)) \subseteq DSPACE(2 + cs(n))$, (lin. space compression)
- ii) $DTIME(t(n)) \subseteq DTIME(2 + n + c \cdot t(n))$. (linear speedup)

Beweis. i) Sei $L \in DSPACE(s(n))$ und sei $M = (Q, \Sigma, \Gamma, \delta, q_0)$ eine $s(n)$ -platzbeschränkte Offline- k -DTM mit $L(M) = L$. Nach vorigem Lemma können wir $k = 2$ annehmen. O.B.d.A. sei $c < 1$. Wähle $m = \lceil 1/c \rceil$ und betrachte die Offline-2-DTM

$$M' = (Q \times \{1, \dots, m\}, \Sigma, \Sigma \cup \{\sqcup, \triangleright\} \cup \Gamma^m, \delta', (q_0, m))$$

mit

$$\delta'((q, i), a, b) = \begin{cases} ((q', 1), a, D_1, \triangleright, R), \\ \text{falls } b = \triangleright \text{ und } \delta(q, a, \triangleright) = (q', a, D_1, \triangleright, R), \\ ((q', j), a, D_1, (b_1, \dots, b_{i-1}, b'_i, b_{i+1}, \dots, b_m), D'_2), \\ \text{falls } [b = (b_1, \dots, b_m) \text{ oder } b = \sqcup = b_1 = \\ \dots = b_m] \text{ und } \delta(q, a, b_i) = (q', a, D_1, b'_i, D_2), \end{cases}$$

wobei

$$j = \begin{cases} i, & D_2 = N \\ i + 1, & D_2 = R, i < m \\ 1, & D_2 = R, i = m \\ m, & D_2 = L, i = 1 \\ i - 1, & D_2 = L, i > 1 \end{cases} \quad \text{und} \quad D'_2 = \begin{cases} L, & D_2 = L, i = 1 \\ R, & D_2 = R, i = m \\ N, & \text{sonst} \end{cases}$$

ist. Identifizieren wir die Zustände (q_{ja}, i) mit q_{ja} und (q_{nein}, i) mit q_{nein} , so ist leicht zu sehen, dass $L(M') = L(M) = L$ gilt. Zudem gilt

$$\begin{aligned} space_{M'} &\leq 1 + \lceil (space_M(x) - 1)/m \rceil \\ &\leq 2 + space_M(x)/m \\ &\leq 2 + c \cdot space_M(x) \quad (\text{wegen } m = \lceil 1/c \rceil \geq 1/c). \end{aligned}$$

ii) Sei $L \in \text{DTIME}(t(n))$ und sei $M = (Q, \Sigma, \Gamma, \delta, q_0)$ eine $t(n)$ -zeitbeschränkte k -DTM mit $L(M) = L$, wobei wir $k \geq 2$ annehmen. Wir konstruieren eine k -DTM M' mit $L(M') = L$ und $time_{M'}(x) \leq 2 + |x| + c \cdot time_M(x)$. M' verwendet das Alphabet $\Gamma' = \Gamma \cup \Gamma^m$ mit $m = \lceil 8/c \rceil$ und simuliert M wie folgt.

Initialisierung: M' kopiert die Eingabe $x = x_1 \dots x_n$ in Blockform auf das zweite Band. Hierzu fasst M' je m Zeichen von x zu einem Block $(x_{im+1}, \dots, x_{(i+1)m})$, $i = 0, \dots, l = \lceil n/m \rceil - 1$, zusammen, wobei der letzte Block $(x_{lm+1}, \dots, x_n, \sqcup, \dots, \sqcup)$ mit

$(l+1)m - n$ Blanks auf die Länge m gebracht wird. Sobald M' das erste Blank hinter der Eingabe x erreicht, ersetzt sie dieses durch das Zeichen \triangleright , d.h. das erste Band von M' ist nun mit $\triangleright x \triangleright$ und das zweite Band mit

$$\triangleright (x_1, \dots, x_m) \dots (x_{(l-1)m+1}, \dots, x_{lm}) (x_{lm+1}, \dots, x_n, \sqcup, \dots, \sqcup)$$

beschriftet. Hierzu benötigt M' genau $n+2$ Schritte. In weiteren $l+1 = \lceil n/m \rceil$ Schritten kehrt M' an den Beginn des 2. Bandes zurück. Von nun an benutzt M' das erste Band als Arbeitsband und das zweite als Eingabeband.

Simulation: M' simuliert jeweils eine Folge von m Schritten von M in 6 Schritten:

M' merkt sich in ihrem Zustand den Zustand q von M vor Ausführung dieser Folge und die aktuellen Kopfpositionen $i_j \in \{1, \dots, m\}$ von M innerhalb der gerade gelesenen Blöcke auf den Bändern $j = 1, \dots, k$. Die ersten 4 Schritte verwendet M' , um die beiden Nachbarblöcke auf jedem Band zu erfassen ($LRRL$). Mit dieser Information kann M' die nächsten m Schritte von M vorausberechnen und die entsprechende Konfiguration in 2 weiteren Schritten herstellen.

Akzeptanzverhalten: M' akzeptiert genau dann, wenn M dies tut.

Es ist klar, dass $L(M') = L$ ist. Zudem gilt für jede Eingabe x der Länge $|x| = n$

$$\begin{aligned} time_{M'}(x) &\leq n + 2 + \lceil n/m \rceil + 6 \lceil t(n)/m \rceil \\ &\leq n + 2 + 7 \lceil t(n)/m \rceil \\ &\leq n + 2 + 7ct(n)/8 + 7 \\ &\leq n + 2 + ct(n), \text{ falls } c \cdot t(n)/8 \geq 7. \end{aligned}$$

Da das Ergebnis der Rechnung von $M(x)$ im Fall $t(n) < 56/c$ nur von konstant vielen Eingabezeichen abhängt, kann M' diese Eingaben schon während der Initialisierungsphase (durch table-lookup) in Zeit $n+2$ entscheiden. ■

Korollar 17.

- i) $\text{DSPACE}(O(s(n))) = \text{DSPACE}(s(n))$, falls $s(n) \geq 2$.
- ii) $\text{DTIME}(O(t(n))) = \text{DTIME}(t(n))$, falls $t(n) \geq (1 + \varepsilon)n + 2$ für ein $\varepsilon > 0$ ist.
- iii) $\text{DTIME}(O(n)) = \bigcap_{\varepsilon > 0} \text{DTIME}((1 + \varepsilon)n + 2)$.

Beweis. i) Sei $L \in \text{DSPACE}(cs(n) + c)$ für eine Konstante $c \geq 0$. Ist $s(n) < 6$ für unendlich viele n , so folgt $L \in \text{DSPACE}(O(1)) = \text{DSPACE}(0)$. Gilt dagegen $s(n) \geq 6$ für fast alle n , so existiert für $c' = 1/2c$ eine Offline- k -DTM M , die L für fast alle Eingaben in Platz $2 + c'cs(n) + c'c \leq 3 + s(n)/2 \leq s(n)$ entscheidet. Wegen $s(n) \geq 2$ können wir M leicht so modifizieren, dass sie auch die endlich vielen Ausnahmen in Platz $s(n)$ entscheidet.

ii) Sei $L \in \text{DTIME}(ct(n) + c)$ für ein $c \geq 0$. Nach vorigem Satz existiert für $c' = \varepsilon/(2 + 2\varepsilon)c$ eine DTM M , die L in Zeit $2 + n + c'ct(n) + c'c$ entscheidet. Wegen $c'ct(n) = \varepsilon t(n)/(2 + 2\varepsilon)$ und da wegen $t(n) \geq (1 + \varepsilon)n$ für fast alle n gilt, dass $(2 + \varepsilon)t(n)/(2 + 2\varepsilon) \geq (2 + \varepsilon)n/2 \geq 2 + c'c + n$ ist, folgt $2 + c'c + n + c'ct(n) \leq t(n)$ für fast alle n . Wegen $t(n) \geq n + 2$ können wir M leicht so modifizieren, dass sie auch die endlich vielen Ausnahmen in Zeit $t(n)$ entscheidet.

iii) Klar, da $\text{DTIME}(O(n)) = \text{DTIME}(O((1 + \varepsilon)n + 2))$ und diese Klasse nach ii) für jedes $\varepsilon > 0$ gleich $\text{DTIME}((1 + \varepsilon)n + 2)$ ist. ■

3.2 Deterministische Simulationen von nichtdeterministischen Berechnungen

In diesem Abschnitt betrachten wir möglichst platz- und zeiteffiziente deterministische Simulationen von nichtdeterministischen TMs.

Satz 18.

- i) $\text{NTIME}(t(n)) \subseteq \text{DSPACE}(O(t(n)))$,

- ii) $\text{NSPACE}(s(n)) \subseteq \text{DTIME}(2^{O(s(n) + \log n)})$.

Beweis. i) Sei $L \in \text{NTIME}(t(n))$ und sei $N = (Q, \Sigma, \Gamma, \Delta, q_0)$ eine k -NTM, die L in Zeit $t(n)$ entscheidet. Weiter sei

$$d = \max_{(q, \vec{a}) \in Q \times \Gamma^k} \|\delta(q, \vec{a})\|$$

der maximale Verzweigungsgrad von N . Dann ist jede Rechnung

$$K_x = K_0 \xrightarrow{N} K_1 \xrightarrow{N} \dots \xrightarrow{N} K_t$$

der Länge t von $N(x)$ eindeutig durch eine Folge $(i_1, \dots, i_t) \in \{1, \dots, d\}^t$ beschreibbar. Betrachte die Offline- $(k+2)$ -DTM M , die auf ihrem 2. Band für $t = 1, 2, \dots$ der Reihe nach alle Folgen $(i_1, \dots, i_t) \in \{1, \dots, d\}^t$ generiert. Für jede solche Folge kopiert M die Eingabe auf Band 3 und simuliert die zugehörige Rechnung von $N(x)$ auf den Bändern 3 bis $k + 2$. M akzeptiert, sobald N bei einer dieser Simulationen in den Zustand q_{ja} gelangt. Wird dagegen ein t erreicht, für das alle d^t Simulationen von N im Zustand q_{nein} oder q_{h} enden, so verwirft M . Nun ist leicht zu sehen, dass $L(M) = L(N)$ und der Platzverbrauch von M durch

$$\text{space}_M(x) \leq \text{time}_N(x) + \text{space}_N(x) \leq (k + 1)(\text{time}_N(x) + 1)$$

beschränkt ist.

ii) Sei $L \in \text{NSPACE}(s(n))$ und sei $N = (Q, \Sigma, \Gamma, \delta, q_0)$ eine Offline-2-NTM, die L in Platz $s(n)$ entscheidet. Bei einer Eingabe x der Länge n kann N

- die Köpfe des Eingabe- bzw. Arbeitsbandes auf höchstens $n + 2$ bzw. $s(n)$ verschiedenen Bandfeldern positionieren,
- das Arbeitsband mit höchstens $\|\Gamma\|^{s(n)}$ verschiedenen Beschriftungen versehen und
- höchstens $\|Q\|$ verschiedene Zustände annehmen.

D.h. ausgehend von der Startkonfiguration K_x kann N in Platz $s(n)$ höchstens

$$t(n) = (n + 2)s(n) \|\Gamma\|^{s(n)} \|Q\| \leq c^{s(n)+\log n}$$

verschiedene Konfigurationen erreichen, wobei c eine von N abhängige Konstante ist. Um N zu simulieren, testet M für $s = 1, 2, \dots$, ob $N(x)$ eine akzeptierende Endkonfiguration $K = (q_{ja}, u_1, v_1, u_2, v_2)$ der Größe $|u_2v_2| = s$ erreichen kann. Ist dies der Fall, akzeptiert M . Erreicht dagegen s einen Wert, so dass $N(x)$ keine Konfiguration der Größe s erreichen kann, verwirft M . Hierzu muss M für $s = 1, 2, \dots, s(n)$ jeweils alle von der Startkonfiguration K_x erreichbaren Konfigurationen der Größe s bestimmen, was in Zeit $(c^{s(n)+\log n})^{O(1)} = 2^{O(s(n)+\log n)}$ möglich ist. ■

Korollar 19. $s(n) \geq \log n \Rightarrow \text{NSPACE}(s(n)) \subseteq \text{DTIME}(2^{O(s(n))})$.

Es gilt somit für jede Funktion $s(n) \geq \log n$,

$$\text{DSPACE}(s) \subseteq \text{NSPACE}(s) \subseteq \text{DTIME}(2^{O(s)})$$

und für jede Funktion $t(n) \geq n + 2$,

$$\text{DTIME}(t) \subseteq \text{NTIME}(t) \subseteq \text{DSPACE}(t).$$

Insbesondere erhalten wir somit die Inklusionskette

$$\begin{aligned} L &\subseteq NL \subseteq P \subseteq NP \subseteq \text{PSPACE} \subseteq \text{NSPACE} \\ &\subseteq \text{EXP} \subseteq \text{NEXP} \subseteq \text{EXPSPACE} \subseteq \dots \end{aligned}$$

Des weiteren impliziert Satz 16 für $t(n) \geq n + 2$ und $s(n) \geq \log n$ die beiden Inklusionen

$$\text{NTIME}(t) \subseteq \text{DTIME}(2^{O(t)}) \text{ und } \text{NSPACE}(s) \subseteq \text{DSPACE}(2^{O(s)}),$$

wovon sich letztere noch erheblich verbessern lässt, wie wir im nächsten Abschnitt sehen werden.

3.3 Der Satz von Savitch

Praktisch relevante Komplexitätsklassen werden durch Zeit- und Platzschranken $t(n)$ und $s(n)$ definiert, die sich mit relativ geringem Aufwand berechnen lassen.

Definition 20. Eine monotone Funktion $f : \mathbb{N} \rightarrow \mathbb{N}$ heißt **echte** (engl. proper) **Komplexitätsfunktion**, falls es einen Transducer M gibt mit

- $M(x) = 1^{f(|x|)}$,
- $\text{space}_M(x) = O(f(|x|))$ und
- $\text{time}_M(x) = O(f(|x|) + |x|)$.

Beispiele für echte Komplexitätsfunktionen sind k , $\lceil \log n \rceil$, $\lceil \log^k n \rceil$, $\lceil n \cdot \log n \rceil$, $n^k + k$, 2^n , $n! \cdot \lfloor \sqrt{n} \rfloor$ (siehe Übungen).

Satz 21 (Savitch, 1970).

Für jede echte Komplexitätsfunktion $s(n) \geq \log n$ gilt

$$\text{NSPACE}(s) \subseteq \text{DSPACE}(s^2).$$

Beweis. Sei $L \in \text{NSPACE}(s)$ und sei N eine Offline-2-NTM, die L in Platz $s(n)$ entscheidet. Wie im Beweis von Satz 18 gezeigt, kann N bei einer Eingabe x der Länge n höchstens $c^{s(n)}$ verschiedene Konfigurationen einnehmen. Daher muss im Fall $x \in L$ eine akzeptierende Rechnung der Länge $\leq c^{s(n)}$ existieren. Zudem können wir annehmen, dass $N(x)$ höchstens eine akzeptierende Endkonfiguration \hat{K}_x erreichen kann.

Sei $K_1, \dots, K_{c^{s(n)}}$ eine Aufzählung aller Konfigurationen von $N(x)$ die Platz höchstens $s(n)$ benötigen. Dann ist leicht zu sehen, dass für je zwei solche Konfigurationen K, K' und jede Zahl i folgende Äquivalenz gilt:

$$K \xrightarrow[N]{\leq 2^i} K' \Leftrightarrow \exists K_j : K \xrightarrow[N]{\leq 2^{i-1}} K_j \wedge K_j \xrightarrow[N]{\leq 2^{i-1}} K'.$$

Diese Beobachtung führt sofort auf folgende Prozedur $\text{reach}(K, K', i)$, um die Gültigkeit von $K \xrightarrow[N]{\leq 2^i} K'$ zu testen.

Prozedur $\text{reach}(K, K', i)$

```

1  if  $i = 0$  then return( $K = K'$  or  $K \xrightarrow[N]{>} K'$ )
2  for each Konfiguration  $K_j$  do
3    if  $\text{reach}(K, K_j, i - 1)$  and  $\text{reach}(K_j, K', i - 1)$  then
4      return(true)
5  return(false)

```

Nun können wir N durch folgende Offline- β -DTM M simulieren. M benutzt ihr 2. Band als Laufzeitkeller zur Verwaltung der Inkarnationen der rekursiven Aufrufe von reach . Hierzu speichert M auf ihrem 2. Band eine Folge von Tripeln der Form (K, K', i) . Das 3. Band wird zum Kopieren von Tripeln auf dem 2. Band und zur Berechnung von K_{j+1} aus K_j benutzt wird.

Initialisierung: $M(x)$ schreibt das Tripel $(K_x, \hat{K}_x, \lceil s(|n|) \log c \rceil)$ auf das 2. Band, wobei für das Eingabeband nur die Kopfposition, nicht jedoch die Beschriftung notiert wird (also z.B. $K_x = (q_0, 1, \varepsilon, \triangleright)$).

Simulation: Sei (K, K', i) das am weitesten rechts auf dem 2. Band stehende Tripel (also das oberste Kellerelement).

Im Fall $i = 0$ testet M direkt, ob $K \xrightarrow[N]{\leq 1} K'$ gilt und gibt die Antwort zurück.

Andernfalls fügt M beginnend mit $j = 1$ das Tripel $(K, K_j, i - 1)$ hinzu und berechnet (rekursiv) die Antwort für dieses Tripel.

Ist diese negativ, so wird das Tripel $(K, K_j, i - 1)$ durch das nächste Tripel $(K, K_{j+1}, i - 1)$ ersetzt (solange $j < c^{s(n)}$ ist, andernfalls erfährt das Tripel (K, K', i) eine negative Antwort).

Erhält $(K, K_j, i - 1)$ eine positive Antwort, so ersetzt M das Tripel $(K, K_j, i - 1)$ durch das Tripel $(K_j, K', i - 1)$ und berechnet die zugehörige Antwort. Bei einer negativen Antwort

fährt M mit dem nächsten Tripel $(K, K_{j+1}, i - 1)$ fort. Bei einer positiven Antwort erhält auch (K, K', i) eine positive Antwort.

Akzeptanzverhalten: M akzeptiert, falls die Antwort auf das Starttripel $(K_x, \hat{K}_x, \lceil s(|n|) \log c \rceil)$ positiv ist.

Da sich auf dem 2. Band zu jedem Zeitpunkt höchstens $\lceil s(|n|) \log c \rceil$ Tripel befinden und jedes Tripel $O(s(|x|))$ Platz benötigt, besucht M nur $O(s^2(|x|))$ Felder. ■

Korollar 22.

- i) $\text{NL} \subseteq \text{L}^2$,
- ii) $\text{NPSpace} = \bigcup_{k>0} \text{NSpace}(n^k) \subseteq \bigcup_{k>0} \text{DSpace}(n^{2k}) = \text{PSPACE}$,
- iii) NPSpace ist unter Komplement abgeschlossen,
- iv) $\text{CSL} = \text{NSpace}(n) \subseteq \text{DSpace}(n^2) \cap \text{E}$.

Eine weitere Folgerung aus dem Satz von Savitch ist, dass das Komplement \bar{L} einer Sprache $L \in \text{NSpace}(s)$ in $\text{DSpace}(s^2)$ und somit auch in $\text{NSpace}(s^2)$ liegt. Wir werden gleich sehen, dass \bar{L} sogar in $\text{NSpace}(s)$ liegt, d.h. die nichtdeterministischen Platzklassen $\text{NSpace}(s)$ sind unter Komplementbildung abgeschlossen.

3.4 Der Satz von Immerman und Szelepcsényi

Wie wir gesehen haben, impliziert der Satz von Savitch den Abschluss von NPSpace unter Komplementbildung. Dagegen wurde die Frage ob auch die Klasse $\text{CSL} = \text{NSpace}(n)$ der kontextsensitiven Sprachen unter Komplementbildung abgeschlossen ist, erst in den 80ern von Neil Immerman und unabhängig davon von Robert Szelepcsényi gelöst.

Definition 23. Für eine Sprachklasse \mathcal{C} bezeichne $\text{co-}\mathcal{C} = \{\bar{L} \mid L \in \mathcal{C}\}$ die zu \mathcal{C} **komplementäre Sprachklasse**. Dabei bezeichnet $\bar{L} = \Sigma^* - L$ das **Komplement** einer Sprache $L \subseteq \Sigma^*$.

Die zu NP komplementäre Klasse ist $\text{co-NP} = \{L | \bar{L} \in \text{NP}\}$. Ein Beispiel für ein co-NP-Problem ist TAUT:

Gegeben: Eine boolesche Formel F über n Variablen x_1, \dots, x_n .

Gefragt: Ist F eine Tautologie, d. h. gilt $f(\vec{a}) = 1$ für alle Belegungen $\vec{a} \in \{0, 1\}^n$?

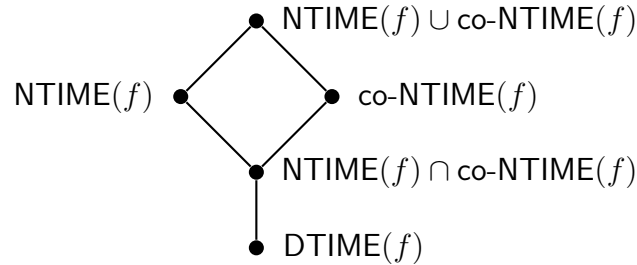
Die Frage ob NP unter Komplementbildung abgeschlossen ist (d. h., ob $\text{NP} = \text{co-NP}$ gilt), ist ähnlich wie das $\text{P} \stackrel{?}{=} \text{NP}$ -Problem ungelöst.

Da sich deterministische Rechnungen leicht komplementieren lassen (durch einfaches Vertauschen der Zustände q_{ja} und q_{nein}), sind deterministische Komplexitätsklassen unter Komplementbildung abgeschlossen.

Proposition 24.

- i) $\text{co-DSPACE}(s(n)) = \text{DSPACE}(s(n))$,
- ii) $\text{co-DTIME}(t(n)) = \text{DTIME}(t(n))$.

Damit ergibt sich folgende Inklusionsstruktur:



Dagegen lassen sich nichtdeterministische Berechnungen nicht ohne weiteres komplementieren; es sei denn, man fordert gewisse Zusatzeigenschaften.

Definition 25. Eine NTM N heißt **strong** bei Eingabe x , falls es entweder akzeptierende oder verwerfende Rechnungen bei Eingabe x gibt (aber nicht beides zugleich).

Satz 26 (Immerman und Szelepcsényi, 1987).

Für jede echte Komplexitätsfunktion $s(n) \geq \log n$ gilt

$$\text{NSPACE}(s) = \text{co-NSPACE}(s).$$

Beweis. Sei $L \in \text{NSPACE}(s)$ und sei N eine $s(n)$ -platzbeschränkte Offline-NTM mit $L(N) = L$. Wir konstruieren eine $O(s(n))$ -platzbeschränkte Offline-NTM N' mit $L(N') = L$, die bei allen Eingaben strong ist. Hierzu zeigen wir zuerst, dass die Frage, ob $N(x)$ eine Konfiguration K in höchstens t Schritten erreichen kann, durch eine $O(s(n))$ -platzbeschränkte Offline-NTM N_0 entscheidbar ist, die bei Kenntnis der Anzahl

$$r(x, t - 1) = \|\{K | K_x \xrightarrow[N]{\leq t-1} K\}\|$$

aller in höchstens $t - 1$ Schritten erreichbaren Konfigurationen strong ist. Betrachte die Sprache

$$L_0 = \{(x, r, t, K) | 1 \leq r, t \leq c^{s(n)} \text{ und } K_x \xrightarrow[N]{\leq t} K\},$$

wobei $K_1, \dots, K_{c^{s(n)}}$ wie im Beweis des Satzes von Savitch eine Aufzählung aller Konfigurationen von $N(x)$ ist, die Platz höchstens $s(n)$ benötigen.

Behauptung 27. Es existiert eine Offline-NTM N_0 mit $L(N_0) = L_0$, die bei jeder Eingabe $O(s(|x|))$ Felder besucht und auf allen Eingaben der Form $(x, r(x, t - 1), t, K)$ strong ist.

Beweis der Behauptung. $N_0(x, r, t, K)$ benutzt einen mit dem Wert 0 initialisierten Zähler z und rät der Reihe nach für jede Konfiguration K_i eine Rechnung von $N(x)$ der Länge $\leq t - 1$, die in K_i endet. Falls dies gelingt, erhöht N_0 den Zähler z um 1 und testet, ob $K_i \xrightarrow[N]{\leq 1} K$ gilt. Falls ja, so hält N_0 im Zustand q_{ja} . Nachdem N_0 alle Konfigurationen K_i durchlaufen hat, hält N_0 im Zustand q_{nein} , wenn z den Wert r hat, andernfalls im Zustand q_{h} .

Pseudocode für $N_0(x, r, t, K)$

```

1  if  $t = 0$  then halte im Zustand  $q_{\text{nein}}$ 
2   $z := 0$ 
3  for each Konfiguration  $K_i$  do
4    rate eine Rechnung  $\alpha$  der Laenge  $\leq t - 1$  von  $N(x)$ 
5    if  $\alpha$  endet in  $K_i$  then
6       $z := z + 1$ 
7      if  $K_i \xrightarrow{N}^{\leq 1} K$  then
8        halte im Zustand  $q_{\text{ja}}$ 
9  if  $z = r$  then
10   halte im Zustand  $q_{\text{nein}}$ 
11 else
12   halte im Zustand  $q_{\text{h}}$ 

```

Da N_0 genau dann eine akzeptierende Rechnung hat, wenn eine Konfiguration K_i mit $K_x \xrightarrow{N}^{\leq t-1} K_i$ und $K_i \xrightarrow{N}^{\leq 1} K$ existiert, ist klar, dass N_0 die Sprache L_0 entscheidet. Da N_0 zudem $O(s(n))$ -platzbeschränkt ist, bleibt nur noch zu zeigen, dass N_0 bei Eingaben der Form $x_0 = (x, r(x, t - 1), t, K)$, $t \geq 1$, strong ist, also $N_0(x_0)$ genau im Fall $x_0 \notin L_0$ eine verwerfende Endkonfiguration erreichen kann.

Um bei Eingabe x_0 eine verwerfende Endkonfiguration zu erreichen, muss N_0 $r = r(x, t - 1)$ Konfigurationen K_i finden, für die zwar $K_x \xrightarrow{N}^{\leq t-1} K_i$ aber nicht $K_i \xrightarrow{N}^{\leq 1} K$ gilt. Dies bedeutet jedoch, dass K von keiner der $r(x, t - 1)$ in $t - 1$ Schritten erreichbaren Konfigurationen in einem Schritt erreichbar ist und somit x_0 tatsächlich nicht zu L_0 gehört. Die Umkehrung folgt analog. \square

Betrachte nun folgende NTM N' , die für $t = 1, 2, \dots$ die Anzahl $r(x, t)$ der in höchstens t Schritten erreichbaren Konfigurationen in der Variablen r berechnet (diese Technik wird induktives Zählen, engl. *inductive counting*, genannt) und mit Hilfe dieser Anzahlen im

Fall $x \notin L$ verifiziert, dass keine der erreichbaren Konfigurationen akzeptierend ist.

Pseudocode für $N'(x)$

```

1   $t := 0$ 
2   $r := 1$ 
3  repeat
4     $t := t + 1$ 
5     $r^- := r$ 
6     $r := 0$ 
7    for each Konfiguration  $K_i$  do
8      simuliere  $N_0(x, r^-, t, K_i)$ 
9      if  $N_0$  akzeptiert then
10        $r := r + 1$ 
11       if  $K_i$  ist akzeptierende Endkonfiguration then
12         halte im Zustand  $q_{\text{ja}}$ 
13       if  $N_0$  haelt im Zustand  $q_{\text{h}}$  then
14         halte im Zustand  $q_{\text{h}}$ 
15  until ( $r = r^-$ )
16  halte im Zustand  $q_{\text{nein}}$ 

```

Behauptung 28. *Im t -ten Durchlauf der repeat-Schleife wird r^- in Zeile 5 auf den Wert $r(x, t - 1)$ gesetzt. Folglich wird N_0 von N' in Zeile 8 nur mit Eingaben der Form $(x, r(x, t - 1), t, K_i)$ aufgerufen.*

Beweis der Behauptung. Wir führen Induktion über t :

$t = 1$: Im ersten Durchlauf der repeat-Schleife erhält r^- den Wert $1 = r(x, 0)$.

$t \rightsquigarrow t + 1$: Da r^- zu Beginn des $(t + 1)$ -ten Durchlaufs auf den Wert von r gesetzt wird, müssen wir zeigen, dass r im t -ten Durchlauf auf $r(x, t)$ hochgezählt wird. Nach Induktionsvoraussetzung wird N_0 im t -ten Durchlauf nur mit Eingaben der Form $(x, r(x, t - 1), t, K_i)$ aufgerufen. Da N_0 wegen Beh. 1 auf all

diesen Eingaben *strong* ist und keine dieser Simulationen im Zustand q_h endet (andernfalls würde N' sofort stoppen), werden alle in $\leq t$ Schritten erreichbaren Konfigurationen K_i als solche erkannt und somit wird r tatsächlich auf den Wert $r(x, t)$ hochgezählt. ■

Behauptung 29. Bei Beendigung der *repeat*-Schleife in Zeile 15 gilt $r = r^- = \|\{K \mid K_x \xrightarrow{N^*} K\}\|$.

Beweis der Behauptung. Wir wissen bereits, dass im t -ten Durchlauf der *repeat*-Schleife r den Wert $r(x, t)$ und r^- den Wert $r(x, t - 1)$ erhält. Wird daher die *repeat*-Schleife nach t_e Durchläufen verlassen, so gilt $r = r^- = r(x, t_e) = r(x, t_e - 1)$.

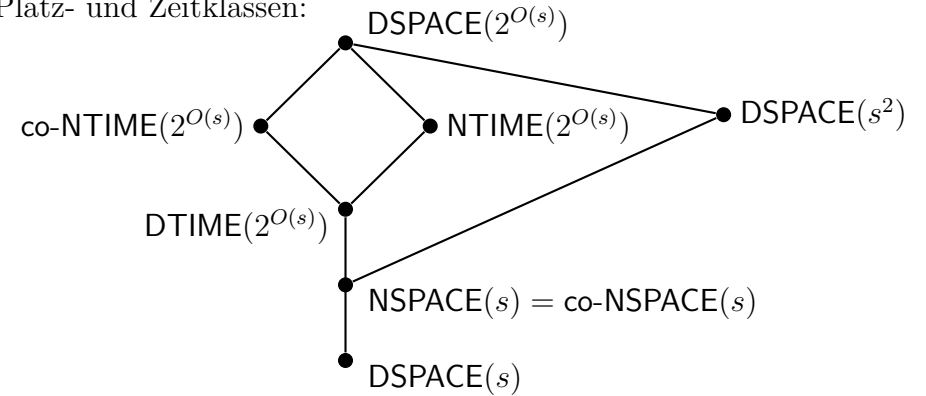
Angenommen $r(x, t_e) < \|\{K \mid K_x \xrightarrow{N^*} K\}\|$. Dann gibt es eine Konfiguration K , die für ein $t' > t_e$ in t' Schritten, aber nicht in t_e Schritten erreichbar ist. Betrachte eine Rechnung $K_x = K_0 \xrightarrow{N} K_1 \xrightarrow{N} \dots \xrightarrow{N} K_{t'} = K$ minimaler Länge, die in K endet. Dann gilt $K_x \xrightarrow{N}^{t_e} K_{t_e}$, aber nicht $K_x \xrightarrow{N}^{\leq t_e - 1} K_{t_e}$ und daher folgt $r(x, t_e) > r(x, t_e - 1)$. Widerspruch! ■

Da N' offenbar die Sprache L in Platz $O(s(n))$ entscheidet, bleibt nur noch zu zeigen, dass N' bei allen Eingaben *strong* ist. Wegen Behauptung 29 hat $N'(x)$ genau dann eine verwerfende Rechnung, wenn im letzten Durchlauf der *repeat*-Schleife alle erreichbaren Konfigurationen K als solche erkannt werden und darunter keine akzeptierende Endkonfiguration ist. Dies impliziert $x \notin L$. Umgekehrt ist leicht zu sehen, dass $N'(x)$ im Fall $x \in L$ eine verwerfende Rechnung hat. ■

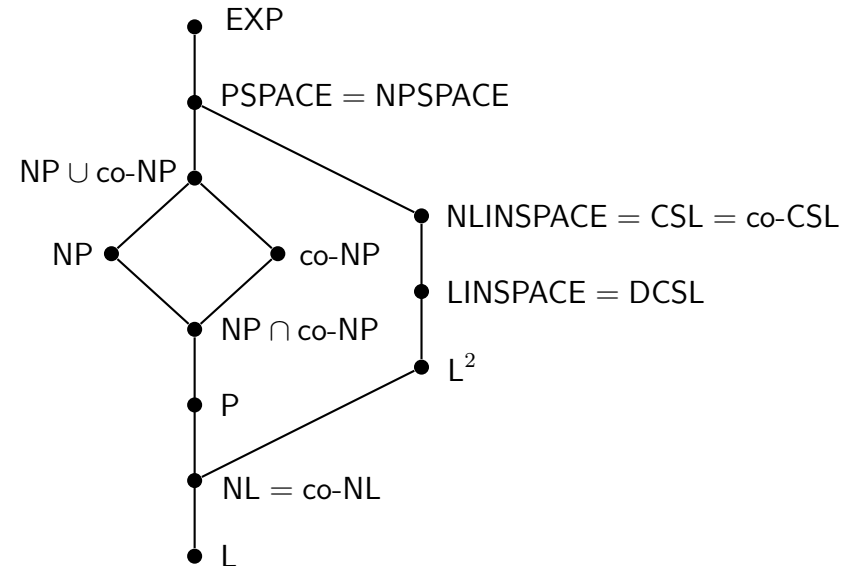
Korollar 30.

1. $NL = co-NL$,
2. $CSL = NLINSPACE = co-CSL$.

Damit ergibt sich folgende Inklusionsstruktur für (nicht)deterministische Platz- und Zeitklassen:



Angewandt auf die wichtigsten bisher betrachteten Komplexitätsklassen erhalten wir folgende Inklusionsstruktur:



Eine zentrale Fragestellung der Komplexitätstheorie ist, welche dieser Inklusionen echt sind. Dieser Frage gehen wir im nächsten Kapitel nach.

4 Hierarchiesätze

4.1 Unentscheidbarkeit mittels Diagonalisierung

Wir benutzen folgende Kodierung (Gödelisierung) von 1-DTMs $M = (Q, \Sigma, \Gamma, \delta, q_0)$. O.B.d.A. sei $Q = \{q_0, q_1, \dots, q_m\}$, $\{0, 1, \#\} \subseteq \Sigma$ und $\Gamma = \{a_1, \dots, a_l\}$ (also z.B. $a_1 = \sqcup$, $a_2 = \triangleright$, $a_3 = 0$, $a_4 = 1$ etc.). Dann kodieren wir jedes $\alpha \in Q \cup \Gamma \cup \{q_h, q_{ja}, q_{nein}, L, R, N\}$ wie folgt durch eine Binärzahl $c(\alpha)$ der Länge $b = \lceil \log_2(\|Q\| + \|\Gamma\| + 6) \rceil = \lceil \log_2(m + l + 7) \rceil$:

α	$c(\alpha)$
$q_i, i = 0, \dots, m$	$bin_b(i)$
$a_j, j = 1, \dots, l$	$bin_b(m + j)$
$q_h, q_{ja}, q_{nein}, L, R, N$	$bin_b(m + l + 1), \dots, bin_b(m + l + 6)$

M wird nun durch eine Folge von Binärzahlen, die durch $\#$ getrennt sind, kodiert:

$$\begin{aligned} &\#c(q_0) \#c(a_1) \#c(p_{0,1}) \#c(b_{0,1}) \#c(D_{0,1}) \# \\ &\#c(q_0) \#c(a_2) \#c(p_{0,2}) \#c(b_{0,2}) \#c(D_{0,2}) \# \\ &\quad \vdots \\ &\#c(q_m) \#c(a_l) \#c(p_{m,l}) \#c(b_{m,l}) \#c(D_{m,l}) \# \end{aligned}$$

wobei

$$\delta(q_i, a_j) = (p_{i,j}, b_{i,j}, D_{i,j})$$

für $i = 1, \dots, m$ und $j = 1, \dots, l$ ist. Kodieren wir die Zeichen $0, 1, \#$ binär (z.B. $0 \mapsto 00$, $1 \mapsto 11$, $\# \mapsto 10$), so gelangen wir zu einer

Binärkodierung von M . Diese Kodierung lässt sich auch auf k -DTM's und k -NTM's erweitern. Die Kodierung einer TM M bezeichnen wir mit $\langle M \rangle$. Umgekehrt können wir jedem Binärstring w eine TM M_w wie folgt zuordnen:

$$M_w = \begin{cases} M, & \langle M \rangle = w \\ M', & \text{sonst,} \end{cases}$$

wobei M' eine beliebig aber fest gewählte TM ist. Für M_w schreiben wir auch M_i , wobei i die Zahl mit der Binärdarstellung $1w$ ist. Ein Paar (M, x) bestehend aus einer TM M und einer Eingabe $x \in \{0, 1\}^*$ kodieren wir durch das Wort $\langle M, x \rangle = \langle M \rangle \# x$.

Satz 31. *Die Diagonalsprache*

$$D = \{\langle M_i \rangle \mid M_i \text{ ist eine DTM, die die Eingabe } x_i \text{ akzeptiert}\}.$$

ist semi-entscheidbar, aber nicht entscheidbar. Hierbei ist $x_1 = \varepsilon$, $x_2 = 0$, $x_3 = 1$, $x_4 = 00, \dots$ die Folge aller Binärstrings in lexikografischer Reihenfolge.

Beweis. Es ist klar, dass D semi-entscheidbar ist, da es eine DTM gibt, die bei Eingabe $\langle M_i \rangle$ die Berechnung von M_i bei Eingabe x_i simuliert und genau dann akzeptiert, wenn dies $M_i(x_i)$ tut.

Dass \bar{D} nicht semi-entscheidbar (und damit D nicht entscheidbar) ist, liegt daran, dass die charakteristische Funktion von \bar{D} „komplementär“ zur Diagonalen der Matrix ist, deren Zeilen die charakteristischen Funktionen aller semi-entscheidbaren Sprachen $L(M_i)$ auflisten. Wir zeigen durch einen einfachen Widerspruchsbeweis, dass keine Zeile der Matrix mit dem Komplement ihrer Diagonalen übereinstimmen kann. Wäre \bar{D} semi-entscheidbar, gäbe es also eine DTM M_d , die \bar{D} akzeptiert,

$$L(M_d) = \bar{D} \quad (**),$$

	x_1	x_2	x_3	x_4	\dots
M_1	1	0	0	0	\dots
M_2	0	1	0	0	\dots
M_3	1	0	0	0	\dots
M_4	0	0	0	1	\dots
\vdots	\vdots	\vdots	\vdots	\vdots	\ddots

M_d	0	0	1	0	\dots
-------	----------	----------	----------	----------	---------

so führt dies wegen

$$\begin{aligned} x_d \in D &\stackrel{(*)}{\Rightarrow} M_d(x_d) \text{ akzeptiert} \stackrel{(**)}{\Rightarrow} x_d \notin D \quad \text{!} \\ x_d \notin D &\stackrel{(*)}{\Rightarrow} M_d(x_d) \text{ akz. nicht} \stackrel{(**)}{\Rightarrow} x_d \in D \quad \text{!} \end{aligned}$$

zu einem Widerspruch. ■

Satz 32. Für jede berechenbare Funktion $g : \mathbb{N} \rightarrow \mathbb{N}$ existiert eine entscheidbare Sprache $D_g \notin \text{DTIME}(g(n))$.

Beweis. Betrachte die Diagonalsprache

$$D_g = \{ \langle M_i \rangle \mid M_i(x_i) \text{ akzeptiert in } \leq g(|x_i|) \text{ Schritten} \} \quad (*)$$

Offensichtlich ist D_g entscheidbar. Unter der Annahme, dass $D_g \in \text{DTIME}(g(n))$ ist, existiert eine $g(n)$ -zeitbeschränkte DTM M_d , die das Komplement von D_g entscheidet, d.h.

$$L(M_d) = \bar{D}_g \quad (**)$$

Dies führt jedoch auf einen Widerspruch:

$$\begin{aligned} x_d \in D_g &\stackrel{(*)}{\Rightarrow} M_d(x_d) \text{ akzeptiert} \stackrel{(**)}{\Rightarrow} x_d \notin D_g \quad \text{!} \\ x_d \notin D_g &\stackrel{(*)}{\Rightarrow} M_d(x_d) \text{ verwirft} \stackrel{(**)}{\Rightarrow} x_d \in D_g \quad \text{!} \end{aligned} \quad \blacksquare$$

Eine interessante Frage ist nun, wieviel Zeit eine DTM benötigt, um die Sprache D_g zu entscheiden. Im nächsten Abschnitt werden wir sehen, dass D_g i.a. sehr hohe Komplexität haben kann.

4.2 Das Gap-Theorem

Satz 33 (Gap-Theorem).

Es gibt eine berechenbare Funktion $g : \mathbb{N} \rightarrow \mathbb{N}$ mit

$$\text{DTIME}(2^{g(n)}) = \text{DTIME}(g(n)).$$

Beweis. Wir definieren $g(n) \geq n+2$ so, dass für jede $2^{g(n)}$ -zeitb. DTM M gilt:

$$\text{time}_M(x) \leq g(|x|) \text{ für fast alle Eingaben } x.$$

Betrachte hierzu das Prädikat

$$\begin{aligned} P(k, t) : &t \geq k + 2 \text{ und für } i = 1, \dots, k \text{ und alle } x \in \Sigma_i^k \text{ gilt:} \\ &\text{time}_{M_i}(x) \notin [t + 1, 2^t]. \end{aligned}$$

Hierbei bezeichnet Σ_i das Eingabealphabet von M_i . Da für jedes k alle

$$t \geq \max\{\text{time}_{M_i}(x) < \infty \mid 1 \leq i \leq k, x \in \Sigma_i^k\}$$

das Prädikat $P(k, t)$ erfüllen, können wir $g(n)$ wie folgt induktiv definieren:

$$g(n) = \begin{cases} 2, & n = 0, \\ \min\{t \geq g(n-1) + n \mid P(n, t)\}, & n > 0. \end{cases}$$

Da P entscheidbar ist, ist g berechenbar.

Um zu zeigen, dass jede Sprache $L \in \text{DTIME}(2^{g(n)})$ bereits in $\text{DTIME}(g(n))$ enthalten ist, sei M_k eine beliebige $2^{g(n)}$ -zeitbeschränkte DTM mit $L(M_k) = L$. Dann muss M_k alle Eingaben x mit $|x| \geq k$ in Zeit $\text{time}_{M_k}(x) \leq g(n)$ ($n = |x|$) entscheiden, da andernfalls $P(n, g(n))$ verletzt wäre. Folglich ist $L \in \text{DTIME}(g(n))$, da die endlich vielen Eingaben x mit $|x| < k$ durch table-lookup in Zeit $|x| + 2$ entscheidbar sind. ■

Es ist leicht zu sehen, dass der Beweis des Gap-Theorems für jede berechenbare Funktion h eine berechenbare Zeitschranke g liefert, so dass $\text{DTIME}(h(g(n))) = \text{DTIME}(g(n))$ ist. Folglich ist D_g nicht in Zeit $h(g(n))$ entscheidbar.

4.3 Zeit- und Platzhierarchiesätze

Um D_g zu entscheiden, müssen wir einerseits die Zeitschranke $g(|x_i|)$ berechnen und andererseits $M_i(x_i)$ simulieren. Wenn wir voraussetzen, dass g eine echte Komplexitätsfunktion ist, lässt sich $g(|x_i|)$ effizient berechnen. Für die zweite Aufgabe benötigen wir eine möglichst effiziente universelle TM.

Satz 34. *Es gibt eine universelle 3-DTM U , die für jede DTM M und jedes $x \in \{0, 1\}^*$ bei Eingabe $\langle M, x \rangle$ eine Simulation von M bei Eingabe x in Zeit $O(|\langle M \rangle|(time_M(x))^2)$ und Platz $O(|\langle M \rangle|space_M(x))$ durchführt und dasselbe Ergebnis liefert:*

$$U(\langle M, x \rangle) = M(x)$$

Beweis. Betrachte folgende Offline-3-DTM U :

Initialisierung: U überprüft bei einer Eingabe $w \# x$ zuerst, ob w die Kodierung $\langle M \rangle$ einer k -DTM $M = (Q, \Sigma, \Gamma, \delta, q_0)$ ist. Falls ja, erzeugt U die Startkonfiguration K_x von M bei Eingabe x , wobei sie die Inhalte von k übereinander liegenden Feldern der Bänder von M auf ihrem 2. Band in je einem Block von kb , $b = \lceil \log_2(\|Q\| + \|\Gamma\| + 6) \rceil$, Feldern speichert und den aktuellen Zustand von M zusammen mit den gerade von M gelesenen Zeichen auf ihrem 3. Band notiert (letztere werden zusätzlich auf dem 2. Band markiert). Hierfür benötigt M' Zeit $O(kbn) = O(n^2)$.

Simulation: U simuliert jeden Rechenschritt von M wie folgt: Zunächst inspiziert U die auf dem 1. Band gespeicherte Kodierung

von M , um die durch den Inhalt des 3. Bandes bestimmte Aktion von M zu ermitteln. Diese führt sie sodann auf dem 2. Band aus und aktualisiert dabei auf dem 3. Band den Zustand und die gelesenen Zeichen von M . Insgesamt benötigt U für die Simulation eines Rechenschrittes von M Zeit $O(kbg(n)) = O(n \cdot g(n))$.

Akzeptanzverhalten: Sobald die Simulation von M zu einem Ende kommt, hält U im gleichen Zustand wie M .

Nun ist leicht zu sehen, dass $U(\langle M, x \rangle)$ genau dann akzeptiert, wenn dies $M(x)$ tut, und dass $U(\langle M, x \rangle)$ $O(|\langle M \rangle|(time_M(x))^2)$ Rechenschritte macht und auf ihren beiden Arbeitsbändern $O(|\langle M \rangle|space_M(x))$ Bandfelder besucht. ■

Korollar 35. *(Zeithierarchiesatz)*

Für jede echte Komplexitätsfunktion $g(n) \geq n + 2$ gilt

$$\text{DTIME}(n \cdot g(n)^2) - \text{DTIME}(g(n)) \neq \emptyset$$

Beweis. Es genügt zu zeigen, dass D_g für jede echte Komplexitätsfunktion $g(n) \geq n + 2$ in Zeit $O(n^2g(n))$ entscheidbar ist. Betrachte folgende 4-DTM M' . M' überprüft bei einer Eingabe x der Länge n zuerst, ob x die Kodierung $\langle M \rangle$ einer k -DTM M ist. Falls ja, erzeugt M' auf dem 4. Band den String $1^{g(n)}$ in Zeit $O(g(n))$ und simuliert $M(x)$ wie im Beweis von Theorem 34. Dabei vermindert M' die Anzahl der Einsen auf dem 4. Band nach jedem simulierten Schritt von $M(x)$ um 1. M' bricht die Simulation ab, sobald M stoppt oder der Zähler auf Band 4 den Wert 0 erreicht. M' hält genau dann im Zustand q_{ja} , wenn die Simulation von M im Zustand q_{ja} endet. Nun ist leicht zu sehen, dass M' $O(n \cdot g(n)^2)$ -zeitbeschränkt ist und die Sprache D_g entscheidet. ■

Korollar 36.

$$P \subsetneq E \subsetneq EXP$$

Beweis.

$$\begin{aligned} P &= \bigcup_{c>0} \text{DTIME}(n^c + c) \subseteq \text{DTIME}(2^n) \\ &\subsetneq \text{DTIME}(n2^{2n}) \subseteq E = \bigcup_{c>0} \text{DTIME}(2^{cn}) \subseteq \text{DTIME}(2^{n^2}) \\ &\subsetneq \text{DTIME}(n2^{2n^2}) \subseteq \bigcup_{c>0} \text{DTIME}(2^{n^c+c}) = \text{EXP} \end{aligned}$$

■

Wir bemerken, dass sich mit Hilfe einer aufwändigeren Simulationstechnik von $g(n)$ -zeitbeschränkten k -DTMs durch eine 2 -DTM in Zeit $\mathcal{O}(g(n) \cdot \log g(n))$ folgende schärfere Form des Zeithierarchiesatzes erhalten lässt (ohne Beweis).

Satz 37. *Sei $f(n) \geq n + 2$ eine echte Komplexitätsfunktion und gelte*

$$\liminf_{n \rightarrow \infty} \frac{g(n) \cdot \log g(n)}{f(n)} = 0.$$

Dann ist

$$\text{DTIME}(f(n)) \setminus \text{DTIME}(g(n)) \neq \emptyset.$$

Für $g(n) = n^2$ erhalten wir beispielsweise die echten Inklusionen $\text{DTIME}(g(n)) \subsetneq \text{DTIME}(f(n))$ für die Funktionen $f(n) = n^3, n^2 \log^2 n$ und $n^2 \log n \log \log n$. In den Übungen zeigen wir, dass die Inklusion

$$\text{DTIME}(n^k) \subsetneq \text{DTIME}(n^k \log^a n)$$

tatsächlich für alle $k \geq 1$ und $a > 0$ echt ist. Für Platzklassen erhalten wir sogar eine noch feinere Hierarchie (siehe Übungen).

Satz 38 (Platzhierarchiesatz). *Sind $g(n), f(n) \geq 2$ und ist f eine echte Komplexitätsfunktion mit*

$$\liminf_{n \rightarrow \infty} \frac{g(n)}{f(n)} = 0,$$

dann ist

$$\text{DSPACE}(f(n)) \setminus \text{DSPACE}(g(n)) \neq \emptyset.$$

Damit lässt sich im Fall $g(n) \leq f(n)$ die Frage, ob die Inklusion von $\text{DSPACE}(g(n))$ in $\text{DSPACE}(f(n))$ echt ist, eindeutig beantworten: Sie ist genau dann echt, wenn $\liminf_{n \rightarrow \infty} g(n)/f(n) = 0$ ist, da andernfalls $f(n) = O(g(n))$ ist und somit beide Klassen gleich sind.

Korollar 39.

$$L \subsetneq L^2 \subsetneq \text{DCSL} \subseteq \text{CSL} \subsetneq \text{PSPACE} \subsetneq \text{ESPACE} \subsetneq \text{EXPSPACE}.$$

Durch Kombination der Beweistechnik von Satz 38 mit der Technik von Immerman und Szelepcsényi erhalten wir auch für nichtdeterministische Platzklassen eine sehr fein abgestufte Hierarchie.

Satz 40 (Nichtdeterministischer Platzhierarchiesatz). *Sind $g(n), f(n) \geq 2$ und ist f eine echte Komplexitätsfunktion mit*

$$\liminf_{n \rightarrow \infty} \frac{g(n)}{f(n)} = 0,$$

dann ist

$$\text{NSPACE}(f(n)) \setminus \text{NSPACE}(g(n)) \neq \emptyset.$$

Ob sich auch der Zeithierarchiesatz auf nichtdeterministische Klassen übertragen lässt, ist dagegen nicht bekannt. Hier gilt jedoch folgender Hierarchiesatz.

Satz 41 (Nichtdeterministischer Zeithierarchiesatz). *Sei $f(n) \geq n + 2$ eine echte Komplexitätsfunktion und gelte*

$$g(n + 1) = o(f(n)).$$

Dann ist

$$\text{NTIME}(g(n)) \subsetneq \text{NTIME}(f(n)).$$

Beweis. Sei M_1, M_2, \dots eine Aufzählung aller 2-NTMs. Für $x \in \{0, 1, \#\}^*$ sei

$$i(x) = \begin{cases} i, & x = 0^k \# \langle M_i \rangle \\ 1, & \text{sonst} \end{cases}$$

und x^+ (x^-) sei der lexikografische Nachfolger (bzw. Vorgänger) von x in $\{0, 1, \#\}^*$. Wir ordnen jedem $x \in \{0, 1, \#\}^*$ ein Intervall $I_x = [s(x), s(x^+) - 1]$ zu, wobei die Funktion s induktiv durch

$$s(x) = \begin{cases} 0, & x = \varepsilon \\ h(s(x^-) + |x^-|), & \text{sonst} \end{cases}$$

definiert ist. Hierbei ist $h(n) \geq 2^n$ eine monotone Funktion mit folgenden Eigenschaften:

- die Sprache

$$D = \{0^s \# \langle M_i \rangle \mid M_i(0^s) \text{ akz. nicht in } \leq f(s) \text{ Schritten}\}$$

ist von einer NTM in Zeit $h(n)$ entscheidbar.

- die Funktion $0^n \rightarrow 0^{h(n)}$ ist von einem Transducer T in Zeit $h(n) + 1$ berechenbar, d.h. $T(0^n)$ schreibt in jedem Rechenschritt (außer dem ersten) eine weitere Null auf's Ausgabeband.

Betrachte folgende NTM M :

```

1  input  $0^n$ 
2     $x := \varepsilon$ 
3     $s := 0$ 
4    while  $h(s + |x|) \leq n$  do
5       $s := h(s + |x|)$ 
6       $x := x^+$ 
7    if  $n < h(s + |x|) - 1$  then (*  $s = s(x) \leq n < s(x^+) - 1$  *)
8      akz. falls  $M_{i(x)}(0^{n+1})$  in  $\leq \frac{f(n)}{|\langle M_{i(x)} \rangle|}$  Schritten akz.
9    else (*  $n = s(x^+) - 1$  *)
10   akz. falls  $0^s \# \langle M_{i(x)} \rangle \in D$  ist

```

Es ist leicht zu sehen, dass M $\mathcal{O}(f(n))$ -zeitb. und somit $L = L(M) \in \text{NTIME}(f(n))$ enthalten ist. Dies liegt daran, dass

- die Berechnung von x und $s = s(x)$ mit $n \in I_x$ in der while-Schleife wegen $h(n) \geq 2^n$ und der Eigenschaften von T in Zeit $\mathcal{O}(n)$ ausführbar, sowie
- die Frage, ob $M_{i(x)}(0^{n+1})$ in $\leq \frac{f(n)}{|\langle M_{i(x)} \rangle|}$ Schritten akz., in Zeit $\mathcal{O}(f(n))$ und
- im Fall $n = s(x^+) - 1$ die Frage, ob $0^s \# \langle M_{i(x)} \rangle \in D$ enthalten ist, in Zeit $h(|0^s \# \langle M_{i(x)} \rangle|) \leq h(s + |x|) = n + 1$ entscheidbar ist.

L kann aber nicht in $\text{NTIME}(g(n))$ enthalten sein, da sonst eine Konstante c und eine 2-NTM M_i ex. würden mit $L(M_i) = L$ und $\text{time}_{M_i}(0^n) \leq cg(n)$ für fast alle n (siehe Übungen; Simulation von NTMs durch 2-NTMs). Wählen wir nun $k \geq 0$ so groß, dass für $x = 0^k \# \langle M_i \rangle$ und alle $n \geq s(x)$ die Ungleichung $|\langle M_i \rangle| \text{time}_{M_i}(0^{n+1}) \leq f(n)$ gilt, so folgt für alle $n \in [s(x), s(x^+) - 2]$:

$$0^n \in L(M) \Leftrightarrow 0^{n+1} \in L(M_i),$$

was $0^{s(x)} \in L \Leftrightarrow 0^{s(x^+)-1} \in L$ impliziert. Zudem gilt

$$0^{s(x^+)-1} \in L(M) \Leftrightarrow 0^{s(x)} \# \langle M_i \rangle \in D \Leftrightarrow 0^{s(x)} \notin L(M_i),$$

was wegen $L(M) = L = L(M_i)$ ein Widerspruch ist. ■

Satz 41 liefert für langsam wachsende Zeitschranken eine feinere Hierarchie als Satz 37. Beispielsweise impliziert Satz 41, dass $\text{NTIME}(n^k)$ für jede unbeschränkte monotone Funktion h echt in der Klasse $\text{NTIME}(n^k h(n))$ enthalten ist, da $(n+1)^k = \mathcal{O}(n^k) = o(n^k h(n))$ ist.

Für schnell wachsende Zeitschranken liefert dagegen Satz 37 eine feinere Hierarchie. So impliziert Satz 37 zum Beispiel, dass die Klasse $\text{DTIME}(2^{2^n})$ für jede unbeschränkte monotone Funktion h echt in $\text{DTIME}(h(n)2^n 2^{2^n})$ enthalten ist, während sich $\text{NTIME}(2^{2^n})$ mit Satz 41 nur von $\text{NTIME}(h(n)2^{2^n+1}) = \text{NTIME}(h(n)2^{2^n} 2^{2^n})$ separieren lässt.

5 Reduktionen

5.1 Logspace-Reduktionen

Oft können wir die Komplexitäten zweier Probleme A und B vergleichen, indem wir die Frage, ob $x \in A$ ist, auf eine Frage der Form $y \in B$ zurückführen. Lässt sich y leicht aus x berechnen, so kann jeder Algorithmus für B in einen Algorithmus für A verwandelt werden, der vergleichbare Komplexität hat.

Definition 42. Seien A und B Sprachen über einem Alphabet Σ . A ist auf B **logspace-reduzierbar** (in Zeichen: $A \leq_m^{\log} B$ oder einfach $A \leq B$), falls eine Funktion $f \in \text{FL}$ existiert, so dass für alle $x \in \Sigma^*$ gilt,

$$x \in A \Leftrightarrow f(x) \in B.$$

Lemma 43. $\text{FL} \subseteq \text{FP}$.

Beweis. Sei $f \in \text{FL}$ und sei M ein logarithmisch platzbeschränkter Transducer (kurz: FL-Transducer), der f berechnet. Da M bei einer Eingabe der Länge n nur $2^{O(\log n)}$ verschiedene Konfigurationen einnehmen kann, ist M dann auch polynomiell zeitbeschränkt. ■

Beispiel 44. Wir reduzieren das Hamiltonkreisproblem auf das Erfüllbarkeitsproblem SAT für aussagenlogische Formeln.

Hamiltonkreisproblem (Ham):

Gegeben: Ein Graph $G = (V, E)$.

Gefragt: Hat G einen Hamiltonkreis?

Erfüllbarkeitsproblem für boolesche Formeln (Sat):

Gegeben: Eine boolesche Formel F über n Variablen.

Gefragt: Ist F erfüllbar?

Hierzu benötigen wir eine Funktion $f \in \text{FL}$, die einen Graphen $G = (V, E)$ so in eine Formel $f(G) = F_G$ transformiert, dass F_G genau dann erfüllbar ist, wenn G hamiltonsch ist. Wir konstruieren F_G über den Variablen $x_{1,1}, \dots, x_{n,n}$, wobei $x_{i,j}$ für die Aussage steht, dass Knoten $j \in V = \{1, \dots, n\}$ in der Rundreise an i -ter Stelle besucht wird. Betrachte nun folgende Klauseln.

i) An der i -ten Stelle wird mindestens ein Knoten besucht:

$$x_{i,1} \vee x_{i,2} \vee \dots \vee x_{i,n}, \quad i = 1, \dots, n.$$

ii) An der i -ten Stelle wird höchstens ein Knoten besucht:

$$\neg x_{i,j} \vee \neg x_{i,k}, \quad i = 1, \dots, n, \quad 1 \leq j < k \leq n.$$

iii) Jeder Knoten j wird mindestens einmal besucht:

$$x_{1,j} \vee \dots \vee x_{n,j}, \quad j = 1, \dots, n.$$

iv) Für $(i, j) \notin E$ wird Knoten j nicht unmittelbar nach Knoten i besucht:

$$\neg x_{1,i} \vee \neg x_{2,j}, \dots, \neg x_{n-1,i} \vee \neg x_{n,j}, \neg x_{n,i} \vee \neg x_{1,j}, \quad (i, j) \notin E.$$

Die Klauseln in a) und b) stellen sicher, dass die Relation $\pi = \{(i, j) \mid x_{i,j} = 1\}$ eine Funktion $\pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ ist. Bedingung c) besagt, dass π surjektiv (und damit auch bijektiv) ist, und d) sorgt dafür, dass der durch π beschriebene Kreis entlang der Kanten von G verläuft. Bilden wir daher $F_G(x_{1,1}, \dots, x_{n,n})$ als Konjunktion dieser

$$n + n \binom{n}{2} + n + n \left[\binom{n}{2} - \|E\| \right] = O(n^3)$$

Klauseln, so ist leicht zu sehen, dass die Reduktionsfunktion f in FL berechenbar ist und G genau dann einen Hamiltonkreis besitzt, wenn F_G erfüllbar ist. ◁

Ein zentraler Begriff in der Komplexitätstheorie ist die Vollständigkeit einer Sprache für eine Komplexitätsklasse.

Definition 45.

- a) Sei C eine Sprachklasse. Eine Sprache L heißt **C-hart** (bzgl. \leq), falls für alle Sprachen $A \in C$ gilt, $A \leq L$.
- b) Eine C-harte Sprache, die zur Klasse C gehört, heißt **C-vollständig**.
- c) C heißt **abgeschlossen** unter \leq , falls gilt:

$$B \in C, A \leq B \Rightarrow A \in C.$$

Lemma 46.

- 1. Die \leq_m^{\log} -Reduzierbarkeit ist reflexiv und transitiv.
- 2. Die Klassen $L, NL, NP, co-NP, PSPACE, EXP$ und $EXPSPACE$ sind unter \leq abgeschlossen.
- 3. Sei L vollständig für eine Klasse C , die unter \leq abgeschlossen ist. Dann gilt

$$C = \{A \mid A \leq L\}.$$

Beweis. Siehe Übungen. ■

Definition 47. Ein **boolescher Schaltkreis** c mit n Eingängen ist eine Folge (g_1, \dots, g_m) von **Gattern**

$$g_l \in \{0, 1, x_1, \dots, x_n, (\neg, j), (\wedge, j, k), (\vee, j, k)\}$$

mit $1 \leq j, k < l$. Der am Gatter g_l berechnete Wert bei Eingabe $a = a_1 \cdots a_n$ ist induktiv wie folgt definiert.

g_l	0	1	x_i	(\neg, j)	(\wedge, j, k)	(\vee, j, k)
$g_l(a)$	0	1	a_i	$1 - g_j(a)$	$g_j(a)g_k(a)$	$g_j(a) + g_k(a) - g_j(a)g_k(a)$

Der Schaltkreis c berechnet die boolesche Funktion $c(a) = g_m(a)$. Er heißt **erfüllbar**, wenn es eine Eingabe $a \in \{0, 1\}^n$ mit $c(a) = 1$ gibt.

Bemerkung: Die Anzahl der Eingänge eines Gatters g wird als **Fan-in** von g bezeichnet, die Anzahl der Ausgänge (also die Anzahl der Gatter, die g als Eingabe benutzen) als **Fanout**. Boolesche Formeln entsprechen also den booleschen Schaltkreisen mit (maximalem) Fanout 1 und umgekehrt.

Ähnlich wie bei booleschen Formeln sind auch für Schaltkreise die beiden folgenden Entscheidungsprobleme von Interesse.

Auswertungsproblem für boolesche Schaltkreise (CirVal):

Gegeben: Ein boolescher Schaltkreis c mit n Eingängen und eine Eingabe $a \in \{0, 1\}^n$.

Gefragt: Ist der Wert von $c(a)$ gleich 1?

Erfüllbarkeitsproblem für boolesche Schaltkreise (CirSat):

Gegeben: Ein boolescher Schaltkreis c mit n Eingängen.

Gefragt: Ist c erfüllbar?

Im folgenden Beispiel führen wir die Lösung des Erreichbarkeitsproblems in gerichteten Graphen auf die Auswertung von booleschen Schaltkreisen zurück.

Beispiel 48. Für die Reduktion $REACH \leq CIRVAL$ benötigen wir eine Funktion $f \in FL$ mit der Eigenschaft, dass für alle Graphen G gilt:

$$G \in REACH \Leftrightarrow f(G) \in CIRVAL.$$

Der Schaltkreis $f(G)$ besteht aus den Gattern

$$g_{i,j,k'} \text{ und } h_{i,j,k} \text{ mit } 1 \leq i, j, k \leq n \text{ und } 0 \leq k' \leq n,$$

wobei die Gatter $g_{i,j,0}$ für $1 \leq i, j \leq n$ die booleschen Konstanten

$$g_{i,j,0} = \begin{cases} 1, & i = j \text{ oder } (i, j) \in E, \\ 0, & \text{sonst} \end{cases}$$

sind und für $k = 1, 2, \dots, n$ gilt,

$$\begin{aligned} h_{i,j,k} &= g_{i,k,k-1} \wedge g_{k,j,k-1}, \\ g_{i,j,k} &= g_{i,j,k-1} \vee h_{i,j,k}. \end{aligned}$$

Dann folgt

$$\begin{aligned} g_{i,j,k} = 1 &\Leftrightarrow \text{es existiert in } G \text{ ein Pfad von } i \text{ nach } j, \text{ der} \\ &\text{nur Zwischenknoten } l \leq k \text{ durchläuft,} \\ h_{i,j,k} = 1 &\Leftrightarrow \text{es existiert in } G \text{ ein Pfad von } i \text{ nach } j, \text{ der} \\ &\text{den Knoten } k, \text{ aber keinen Knoten } l > k \\ &\text{durchläuft.} \end{aligned}$$

Wählen wir also $g_{1,n,n}$ als Ausgabegatter, so liefert der aus diesen Gattern aufgebaute Schaltkreis c genau dann den Wert 1, wenn es in G einen Weg von Knoten 1 zu Knoten n gibt. Es ist auch leicht zu sehen, dass die Reduktionsfunktion f in FL berechenbar ist. \triangleleft

Der in Beispiel 48 konstruierte Schaltkreis hat Tiefe $2n$. In den Übungen werden wir sehen, dass sich REACH auch auf die Auswertung eines Schaltkreises der Tiefe $O(\log^2 n)$ reduzieren lässt. Als nächstes leiten wir Vollständigkeitsresultate für CIRVAL und CIRSAT her.

5.2 P-vollständige Probleme und polynomielle Schaltkreiskomplexität

Satz 49. CIRVAL ist P-vollständig.

Beweis. Es ist leicht zu sehen, dass CIRVAL \in P ist. Um zu zeigen, dass CIRVAL hart für P ist, müssen wir für jede Sprache $L \in$ P eine Funktion $f \in$ FL finden, die L auf CIRVAL reduziert, d.h. es muss für alle Eingaben x die Äquivalenz $x \in L \Leftrightarrow f(x) \in$ CIRVAL gelten.

Zu $L \in$ P existiert eine 1-DTM $M = (Q, \Sigma, \Gamma, \delta, q_0)$, die L in Zeit $n^c + c$ entscheidet. Wir beschreiben die Rechnung von $M(x)$, $|x| = n$, durch

eine Tabelle $T = (T_{i,j})$, $(i, j) \in \{1, \dots, n^c + c\} \times \{1, \dots, n^c + c + 2\}$, mit

$$T_{i,j} = \begin{cases} (q_i, a_{i,j}), & \text{nach } i \text{ Schritten besucht } M \text{ das } j\text{-te Bandfeld,} \\ a_{i,j}, & \text{sonst,} \end{cases}$$

wobei q_i der Zustand von $M(x)$ nach i Rechenschritten ist und $a_{i,j}$ das nach i Schritten an Position j befindliche Zeichen auf dem Arbeitsband ist. $T = (T_{i,j})$ kodiert also in ihren Zeilen die von $M(x)$ der Reihe nach angenommenen Konfigurationen. Dabei

- überspringen wir jedoch alle Konfigurationen, bei denen sich der Kopf auf dem ersten Bandfeld befindet (zur Erinnerung: In diesem Fall wird der Kopf sofort wieder nach rechts bewegt) und
- behalten die in einem Schritt $i < n^c + c$ erreichte Endkonfiguration bis zum Zeitpunkt $i = n^c + c$ bei.

Da M in $n^c + c$ Schritten nicht das $(n^c + c + 2)$ -te Bandfeld erreichen kann, ist $T_{i,1} = \triangleright$ und $T_{i,n^c+c+2} = \sqcup$ für $i = 1, \dots, n^c + c$. Außerdem nehmen wir an, dass M bei jeder Eingabe x auf dem zweiten Bandfeld auf einem Blank hält, d.h. es gilt

$$x \in L \Leftrightarrow T_{n^c+c,2} = (q_{ja}, \sqcup).$$

Da T nicht mehr als $l = \|\Gamma\| + \|(Q \cup \{q_h, q_{ja}, q_{nein}\}) \times \Gamma\|$ verschiedene Tabelleneinträge besitzt, können wir jeden Eintrag $T_{i,j}$ durch eine Bitfolge $t_{i,j,1} \cdots t_{i,j,m}$ der Länge $m = \lceil \log_2 l \rceil$ kodieren.

Da der Eintrag $T_{i,j}$ im Fall $i \in \{2, \dots, n^c + c\}$ und $j \in \{2, \dots, n^c + c + 1\}$ eine Funktion $T_{i,j} = g(T_{i-1,j-1}, T_{i-1,j}, T_{i-1,j+1})$ der drei Einträge $T_{i-1,j-1}$, $T_{i-1,j}$ und $T_{i-1,j+1}$ ist, existieren für $k = 1, \dots, m$ Schaltkreise c_k mit

$$\begin{aligned} t_{i,j,k} &= \\ &c_k(t_{i-1,j-1,1} \cdots t_{i-1,j-1,m}, t_{i-1,j,1} \cdots t_{i-1,j,m}, t_{i-1,j+1,1} \cdots t_{i-1,j+1,m}). \end{aligned}$$

Die Reduktionsfunktion f liefert nun bei Eingabe x folgenden Schaltkreis c_x mit 0 Eingängen.

- Für jeden der $n^c + c + 2 + 2(n^c + c - 1) = 3(n^c + c)$ Randeinträge $T_{i,j}$ mit $i = 1$ oder $j \in \{1, n^c + c + 2\}$ enthält c_x m konstante Gatter $c_{i,j,k} = t_{i,j,k}$, $k = 1, \dots, m$, die diese Einträge kodieren.
- Für jeden der $(n^c + c - 1)(n^c + c)$ übrigen Einträge $T_{i,j}$ enthält c_x für $k = 1, \dots, m$ je eine Kopie $c_{i,j,k}$ von c_k , deren $3m$ Eingänge mit den Ausgängen der Schaltkreise $c_{i-1,j-1,1} \cdots c_{i-1,j-1,m}, c_{i-1,j,1} \cdots c_{i-1,j,m}, c_{i-1,j+1,1} \cdots c_{i-1,j+1,m}$ verdrahtet sind.
- Als Ausgabegatter von c_x fungiert das Gatter $c_{n^c+c,2,1}$, wobei wir annehmen, dass das erste Bit der Kodierung von (q_{ja}, \sqcup) eine Eins und von (q_{nein}, \sqcup) eine Null ist.

Nun lässt sich induktiv über $i = 1, \dots, n^c + c$ zeigen, dass die von den Schaltkreisen $c_{i,j,k}$, $j = 1, \dots, n^c + c$, $k = 1, \dots, m$ berechneten Werte die Tabelleneinträge $T_{i,j}$, $j = 1, \dots, n^c + c$, kodieren. Wegen

$$x \in L \Leftrightarrow T_{n^c+c,2} = (q_{ja}, \sqcup) \Leftrightarrow c_x = 1$$

folgt somit die Korrektheit der Reduktion. Außerdem ist leicht zu sehen, dass f in logarithmischem Platz berechenbar ist, da ein $O(\log n)$ -platzbeschränkter Transducer existiert, der bei Eingabe x

- zuerst die $3(n^c + c)$ konstanten Gatter von c_x ausgibt und danach
- die $m(n^c + c - 1)(n^c + c)$ Kopien der Schaltkreise c_1, \dots, c_k erzeugt und diese Kopien richtig verdrahtet. ■

Eine leichte Modifikation des Beweises von Satz 49 liefert folgendes Resultat.

Korollar 50. Sei $L \subseteq \{0,1\}^*$ eine beliebige Sprache in P . Dann existiert eine Funktion $f \in FL$, die bei Eingabe 1^n einen Schaltkreis c_n mit n Eingängen berechnet, so dass für alle $x \in \{0,1\}^n$ gilt:

$$x \in L \Leftrightarrow c_n(x) = 1.$$

Korollar 50 besagt insbesondere, dass es für jede Sprache $L \subseteq \{0,1\}^*$ in P eine Schaltkreisfamilie $(c_n)_{n \geq 0}$ polynomieller Größe gibt, so dass

c_n für alle Eingaben $x \in \{0,1\}^n$ die charakteristische Funktion von L berechnet.

Die Turingmaschine ist ein **uniformes** Rechenmodell, da alle Instanzen eines Problems von einer einheitlichen Maschine entschieden werden. Im Gegensatz hierzu stellen Schaltkreise ein **nichtuniformes** Berechnungsmodell dar, da für jede Eingabegröße n ein anderer Schaltkreis c_n verwendet wird. Um im Schaltkreis-Modell eine unendliche Sprache entscheiden zu können, wird also eine unendliche Folge c_n , $n \geq 0$, von Schaltkreisen benötigt. Probleme, für die Schaltkreisfamilien polynomieller Größe existieren, werden zur Klasse PSK zusammengefasst.

Definition 51.

- a) Eine Sprache L über dem Binäralphabet $\{0,1\}$ hat **polynomielle Schaltkreiskomplexität** (kurz: $L \in PSK$), falls es eine Folge von booleschen Schaltkreisen c_n , $n \geq 0$, mit n Eingängen und $n^{O(1)} + O(1)$ Gattern gibt, so dass für alle $x \in \{0,1\}^*$ gilt:

$$x \in L \Leftrightarrow c_{|x|}(x) = 1.$$

- b) Eine Sprache L über einem Alphabet $\Sigma = \{a_0, \dots, a_{k-1}\}$ hat **polynomielle Schaltkreiskomplexität** (kurz: $L \in PSK$), falls die Binärcodierung

$$\text{bin}(L) = \{\text{bin}(x_1) \cdots \text{bin}(x_n) \mid x_1 \cdots x_n \in L\}$$

von L polynomielle Schaltkreiskomplexität hat. Hierbei kodieren wir a_i durch die m -stellige Binärdarstellung $\text{bin}(i) \in \{0,1\}^m$ von i , wobei $m = \max\{1, \lceil \log_2 k \rceil\}$ ist.

Korollar 52 (Savage 1972).

Es gilt $P \subseteq PSK$.

Ob auch alle NP-Sprachen polynomielle Schaltkreiskomplexität haben, ist ein berühmtes offenes Problem. Gelingt es nämlich, für ein NP-Problem superpolynomielle untere Schranken für die Schaltkreisgröße zu zeigen, so folgt mit dem Resultat von Savage $P \neq NP$.

Selbst für NEXP ist die Inklusion in PSK offen. Dagegen zeigt ein einfaches Diagonalisierungsargument, dass in EXPSPACE Sprachen mit superpolynomieller Schaltkreiskomplexität existieren. Wir werden später sehen, dass bereits die Annahme $\text{NP} \subseteq \text{PSK}$ schwerwiegende Konsequenzen für uniforme Komplexitätsklassen hat.

Es ist nicht schwer zu sehen, dass die Inklusion $\text{P} \subseteq \text{PSK}$ echt ist. Hierzu betrachten wir Sprachen über einem einelementigen Alphabet.

Definition 53. Eine Sprache T heißt **tally** (kurz: $T \in \text{TALLY}$), falls jedes Wort $x \in T$ die Form $x = 1^n$ hat.

Es ist leicht zu sehen, dass alle tally Sprachen polynomielle Schaltkreiskomplexität haben.

Proposition 54. $\text{TALLY} \subseteq \text{PSK}$.

Andererseits wissen wir aus der Berechenbarkeitstheorie, dass es tally Sprachen T gibt, die nicht einmal semi-entscheidbar sind (etwa wenn T das Komplement des Halteproblems unär kodiert). Folglich sind in PSK beliebig schwierige Sprachen (im Sinne der Berechenbarkeit) enthalten.

Korollar 55. $\text{PSK} \not\subseteq \text{RE}$.

5.3 NP-vollständige Probleme

Wir wenden uns nun der NP-Vollständigkeit von CIRSAT zu. Hierbei wird sich folgende Charakterisierung von NP als nützlich erweisen.

Definition 56. Sei $B \subseteq \Sigma^*$ eine Sprache und sei p ein Polynom.

- B heißt **p -balanciert**, falls B nur Strings der Form $x\#y$ mit $|y| = p(|x|)$ enthält.
- Die Sprache $\exists B$ ist definiert durch

$$\exists B = \{x \in \Sigma^* \mid \exists y \in \{0, 1\}^* : x\#y \in B\}.$$

c) Für eine Sprachklasse C seien

$$\begin{aligned} \exists \cdot C &= \{\exists B \mid B \in C\} \text{ und} \\ \exists^p \cdot C &= \{\exists B \mid B \in C \text{ ist polyn. balanciert}\}. \end{aligned}$$

Jeder String $y \in \{0, 1\}^*$ mit $x\#y \in B$ wird auch als **Zeuge** (engl. *witness, certificate*) für die Zugehörigkeit von x zu $\exists B$ bezeichnet.

Satz 57. $\text{NP} = \exists^p \cdot \text{P}$.

Beweis. Zu jeder NP-Sprache $A \subseteq \Sigma^*$ existiert eine NTM M , die A in Zeit $p(n)$ für ein Polynom p entscheidet. Dabei können wir annehmen, dass jede Konfiguration höchstens zwei Folgekonfigurationen hat, die entsprechend der zugehörigen Anweisungen angeordnet sind. Folglich lässt sich jede Rechnung von $M(x)$ durch einen Binärstring y der Länge $p(n)$ eindeutig beschreiben. Das Ergebnis der durch y beschriebenen Rechnung von $M(x)$ bezeichnen wir mit $M_y(x)$. Nun ist leicht zu sehen, dass

$$B = \{x\#y \mid |y| = p(|x|) \text{ und } M_y(x) = \text{ja}\}$$

eine p -balancierte Sprache in P mit $L = \exists B$ ist.

Gilt umgekehrt $A = \exists B$ für eine p -balancierte Sprache $B \in \text{P}$, dann kann A in Polynomialzeit durch eine NTM M entschieden werden, die bei Eingabe x einen String $y \in \{0, 1\}^{p(|x|)}$ geeigneter Länge rät und testet, ob $x\#y \in B$ ist. Diese Vorgehensweise von nichtdeterministischen Algorithmen wird im Englischen auch als “guess and verify” bezeichnet. ■

Satz 58. CIRSAT ist NP-vollständig.

Beweis. Es ist leicht zu sehen, dass $\text{CIRSAT} \in \text{NP}$ ist. Um zu zeigen, dass CIRSAT hart für NP ist, müssen wir für jede Sprache $L \in \text{NP}$ eine Funktion $f \in \text{FL}$ finden, die L auf CIRSAT reduziert, d.h. es

muss für alle Eingaben x die Äquivalenz $x \in L \Leftrightarrow f(x) \in \text{CIRSAT}$ gelten.

Im Beweis von Satz 57 haben wir gezeigt, dass für jede NP-Sprache $A \subseteq \Sigma^*$ eine p -balancierte Sprache $B \in \mathbf{P}$ mit $A = \exists B$ existiert,

$$x \in A \Leftrightarrow \exists y \in \{0, 1\}^{p(|x|)} : x\#y \in B.$$

Sei $m = \lceil \log_2 \|\Sigma \cup \{\#\}\| \rceil$. Da die Binärsprache

$$B' = \{ \text{bin}_m(x_1) \cdots \text{bin}_m(x_n) \text{bin}_m(\#)y \mid x_1 \cdots x_n \#y \in B \}$$

in \mathbf{P} entscheidbar ist, existiert nach Korollar 50 eine FL-Funktion f , die für 1^n einen Schaltkreis c_n mit $m(n+1) + p(n)$ Eingängen berechnet, so dass für alle $z \in \{0, 1\}^{m(n+1)+p(n)}$ gilt:

$$z \in B' \Leftrightarrow c_n(z) = 1.$$

Betrachte nun die Funktion g , die bei Eingabe x den Schaltkreis c_x ausgibt, der sich aus c_n dadurch ergibt, dass die ersten $m(n+1)$ Input-Gatter durch konstante Gatter mit den durch $\text{bin}_m(x_1) \cdots \text{bin}_m(x_n) \text{bin}_m(\#)$ vorgegebenen Werten ersetzt werden. Dann ist auch g in FL berechenbar und es gilt für alle Eingaben x , $|x| = n$,

$$\begin{aligned} x \in A &\Leftrightarrow \exists y \in \{0, 1\}^{p(n)} : x\#y \in B \\ &\Leftrightarrow \exists y \in \{0, 1\}^{p(n)} : c_n(\text{bin}_m(x_1) \cdots \text{bin}_m(x_n) \text{bin}_m(\#)y) = 1 \\ &\Leftrightarrow \exists y \in \{0, 1\}^{p(n)} : c_x(y) = 1 \\ &\Leftrightarrow c_x \in \text{CIRSAT}. \quad \blacksquare \end{aligned}$$

Als nächstes zeigen wir, dass auch SAT NP-vollständig ist, indem wir CIRSAT auf SAT reduzieren. Tatsächlich können wir CIRSAT sogar auf ein Teilproblem von SAT reduzieren.

Definition 59. Eine boolesche Formel F über den Variablen x_1, \dots, x_n ist in **konjunktiver Normalform** (kurz **KNF**), falls F eine Konjunktion

$$F = \bigwedge_{i=1}^m C_i$$

von Disjunktionen $C_i = \bigvee_{j=1}^{k_i} l_{ij}$ von **Literalen** $l_{ij} \in \{x_1, \dots, x_n, \bar{x}_1, \dots, \bar{x}_n\}$ ist. Hierbei verwenden wir \bar{x} als abkürzende Schreibweise für $\neg x$. Gilt $k_i \leq k$ für $i = 1, \dots, m$, so heißt F in **k -KNF**.

Eine Disjunktion $C = \bigvee_{j=1}^k l_j$ von Literalen wird auch als **Klausel** bezeichnet. Klauseln werden meist als Menge $C = \{l_1, \dots, l_k\}$ der zugehörigen Literale und KNF-Formeln als Menge $F = \{C_1, \dots, C_m\}$ ihrer Klauseln dargestellt.

Erfüllbarkeitsproblem für k -KNF Formeln (k -Sat):

Gegeben: Eine boolesche Formel in k -KNF.

Gefragt: Ist F erfüllbar?

Beispiel 60. Die Formel $F = (x_1 \vee \bar{x}_2) \wedge (\bar{x}_1 \vee x_3) \wedge (x_2 \vee \bar{x}_3 \vee x_4)$ ist in 3-KNF und lässt sich in Mengennotation durch $F = \{\{x_1, \bar{x}_2\}, \{\bar{x}_1, x_3\}, \{x_2, \bar{x}_3, x_4\}\}$ beschreiben. F ist offensichtlich erfüllbar, da in jeder Klausel ein positives Literal vorkommt. \triangleleft

Satz 61. 3-SAT ist NP-vollständig.

Beweis. Es ist leicht zu sehen, dass $3\text{-SAT} \in \text{NP}$ ist. Um 3-SAT als hart für NP nachzuweisen, reicht es aufgrund der Transitivität von \leq CIRSAT auf 3-SAT zu reduzieren.

Idee: Wir transformieren einen Schaltkreis $c = \{g_1, \dots, g_m\}$ mit n Eingängen in eine 3-KNF-Formel F_c mit $n + m$ Variablen $x_1, \dots, x_n, y_1, \dots, y_m$, wobei y_i den Wert des Gatters g_i wiedergibt.

Konkret enthält F_c für jedes Gatter g_i folgende Klauseln:

Gatter g_i	zugeh. Klauseln	Semantik
0	$\{\bar{y}_i\}$	$y_i = 0$
1	$\{y_i\}$	$y_i = 1$
x_j	$\{\bar{y}_i, x_j\}, \{\bar{x}_j, y_i\}$	$y_i \leftrightarrow x_j$
(\neg, j)	$\{\bar{y}_i, \bar{y}_j\}, \{y_j, y_i\}$	$y_i \leftrightarrow \bar{y}_j$
(\wedge, j, k)	$\{\bar{y}_i, y_j\}, \{\bar{y}_i, y_k\}, \{\bar{y}_j, \bar{y}_k, y_i\}$	$y_i \leftrightarrow y_j \wedge y_k$
(\vee, j, k)	$\{\bar{y}_j, y_i\}, \{\bar{y}_k, y_i\}, \{\bar{y}_i, y_j, y_k\}$	$y_i \leftrightarrow y_j \vee y_k$

Außerdem fügen wir noch die Klausel $\{y_m\}$ zu F_c hinzu. Nun ist leicht zu sehen, dass für alle $x \in \{0, 1\}^n$ die Äquivalenz

$$c(x) = 1 \Leftrightarrow \exists y \in \{0, 1\}^m : F_c(x, y) = 1$$

gilt. Dies bedeutet jedoch, dass der Schaltkreis c und die 3-KNF-Formel F_c erfüllbarkeitsäquivalent sind, d.h.

$$c \in \text{CIRSAT} \Leftrightarrow F_c \in \text{3-SAT}.$$

Zudem ist leicht zu sehen, dass die Reduktion $c \mapsto F_c$ in FL berechenbar ist. ■

3-SAT ist also nicht in Polynomialzeit entscheidbar, außer wenn $P = NP$ ist. Am Ende dieses Abschnitts werden wir sehen, dass dagegen 2-SAT effizient entscheidbar ist. Zunächst betrachten wir folgende Variante von 3-SAT.

Not-All-Equal-Satisfiability (NaeSat):

Gegeben: Eine Formel F in 3-KNF.

Gefragt: Existiert eine Belegung für F , unter der in jeder Klausel beide Wahrheitswerte angenommen werden?

Satz 62. $\text{NAESAT} \in \text{NPC}$.

Beweis. $\text{NAESAT} \in \text{NP}$ ist klar. Wir zeigen $\text{CIRSAT} \leq \text{NAESAT}$ durch eine leichte Modifikation der Reduktion $C(x_1, \dots, x_n) \mapsto F_c(x_1, \dots, x_n, y_1, \dots, y_m)$ von CIRSAT auf 3-SAT:

Sei $F'_c(x_1, \dots, x_n, y_1, \dots, y_m, z)$ die 3-KNF Formel, die aus F_c dadurch entsteht, dass wir zu jeder Klausel mit ≤ 2 Literalen die neue Variable z hinzufügen.

Dann ist die Reduktion $f : c \mapsto F'_c$ in FL berechenbar. Es bleibt also nur noch die Korrektheit von f zu zeigen, d.h.

$$c \in \text{CIRSAT} \Leftrightarrow F'_c \in \text{NAESAT}.$$

Ist $c = (g_1, \dots, g_m) \in \text{CIRSAT}$, so existiert eine Eingabe $x \in \{0, 1\}^n$ mit $c(x) = 1$. Wir betrachten die Belegung $a = xyz \in \{0, 1\}^{n+m+1}$ mit $y = y_1 \dots y_m$, wobei $y_i = g_i(x)$ und $z = 0$. Da $F_c(xy) = 1$ ist, enthält jede Klausel von F_c (und damit auch von F'_c) mindestens ein wahres Literal. Wegen $z = 0$ müssen wir nur noch zeigen, dass nicht alle Literale in den Dreierklauseln von F_c unter a wahr werden. Da a jedoch für jedes oder-Gatter $g_i = (\vee, j, k)$ die drei Klauseln

$$\{\bar{y}_i, y_j, y_k\}, \{\bar{y}_j, y_i\}, \{\bar{y}_k, y_i\}$$

und für jedes und-Gatter $g_i = (\wedge, j, k)$ die drei Klauseln

$$\{y_i, \bar{y}_j, \bar{y}_k\}, \{y_j, \bar{y}_i\}, \{y_k, \bar{y}_i\}$$

erfüllt, kann weder $y_i = 0$ und $y_j = y_k = 1$ noch $y_i = 1$ und $y_j = y_k = 0$ gelten, da im ersten Fall die Klausel $\{\bar{y}_j, y_i\}$ und im zweiten Fall die Klausel $\{y_j, \bar{y}_i\}$ falsch wäre.

Ist umgekehrt $F'_c \in \text{NAESAT}$, so existiert eine Belegung $xyz \in \{0, 1\}^{n+m+1}$ unter der in jeder Klausel von F'_c beide Wahrheitswerte vorkommen. Da dies dann auch auf die Belegung $\bar{x}\bar{y}\bar{z}$ zutrifft, können wir $z = 0$ annehmen. Dann erfüllt aber die Belegung xy die Formel F_c . ■

Definition 63. Sei $G = (V, E)$ ein ungerichteter Graph.

- a) Eine Menge $C \subseteq V$ heißt **Clique** in G , falls für alle $u, v \in C$ mit $u \neq v$ gilt: $\{u, v\} \in E$.
- b) $I \subseteq V$ heißt **unabhängig** (oder **stabil**), falls für alle $u, v \in I$ gilt: $\{u, v\} \notin E$.
- c) $K \subseteq V$ heißt **Kantenüberdeckung**, falls für alle $e \in E$ gilt: $e \cap K \neq \emptyset$.

Für einen gegebenen Graphen G und eine Zahl k betrachten wir die folgenden Fragestellungen:

Clique: Besitzt G eine Clique der Größe k ?

Independent Set (IS): Besitzt G eine stabile Menge der Größe k ?

Vertex Cover (VC): Besitzt G eine Kantenüberdeckung der Größe k ?

Satz 64. IS ist NP-vollständig.

Beweis. Wir reduzieren 3-SAT auf IS. Sei

$$F = \{C_1, \dots, C_m\} \text{ mit } C_i = \{l_{i,1}, \dots, l_{i,k_i}\} \text{ und } k_i \leq 3 \text{ für } i = 1, \dots, m$$

eine 3-KNF-Formel über den Variablen x_1, \dots, x_n . Betrachte den Graphen $G = (V, E)$ mit

$$\begin{aligned} V &= \{v_{ij} \mid 1 \leq i \leq m, 1 \leq j \leq k_i\} \\ E &= \{\{v_{ij}, v_{ij'}\} \mid 1 \leq i \leq m, 1 \leq j < j' \leq k_i\} \\ &\quad \cup \{\{v_{s,t}, v_{u,v}\} \mid l_{st} \text{ und } l_{uv} \text{ sind komplementär}\}. \end{aligned}$$

Dabei heißen zwei Literale **komplementär**, wenn das eine die Nega-

tion des anderen ist. Nun gilt

$$\begin{aligned} F \in 3\text{-SAT} &\Leftrightarrow \text{es gibt eine Belegung, die in jeder Klausel } C_i \text{ mindestens ein Literal wahr macht} \\ &\Leftrightarrow \text{es gibt } m \text{ Literale } l_{1,j_1}, \dots, l_{m,j_m}, \text{ die paarweise nicht komplementär sind} \\ &\Leftrightarrow \text{es gibt } m \text{ Knoten } v_{1,j_1}, \dots, v_{m,j_m}, \text{ die nicht durch Kanten verbunden sind} \\ &\Leftrightarrow G \text{ besitzt eine stabile Knotenmenge der Größe } m. \end{aligned}$$

Korollar 65. CLIQUE ist NP-vollständig.

Beweis. Es ist leicht zu sehen, dass jede Clique in einem Graphen $G = (V, E)$ eine stabile Menge in dem zu G komplementären Graphen $\bar{G} = (V, E')$ mit $E' = \binom{V}{2} \setminus E$ ist und umgekehrt. Daher lässt sich IS mittels der Reduktionsfunktion

$$f : (G, k) \mapsto (\bar{G}, k)$$

auf CLIQUE reduzieren. ■

Korollar 66. VC ist NP-vollständig.

Beweis. Offensichtlich ist eine Menge I genau dann stabil, wenn ihr Komplement $V \setminus I$ eine Kantenüberdeckung ist. Daher lässt sich IS mittels der Reduktionsfunktion

$$f : (G, k) \mapsto (G, n - k)$$

auf VC reduzieren, wobei $n = \|V\|$ die Anzahl der Knoten in G ist. ■

5.4 NL-vollständige Probleme

In diesem Abschnitt präsentieren wir einen effizienten Algorithmus für das 2-SAT-Problem.

Satz 67. 2-SAT \in NL.

Beweis. Sei F eine 2-KNF-Formel über den Variablen x_1, \dots, x_n . Betrachte den Graphen $G = (V, E)$ mit

$$V = \{x_1, \dots, x_n, \bar{x}_1, \dots, \bar{x}_n\},$$

der für jede Zweierklausel $\{l_1, l_2\}$ von F die beiden Kanten (\bar{l}_1, l_2) und (\bar{l}_2, l_1) und für jede Einerklausel $\{l\}$ die Kante (\bar{l}, l) enthält. Hierbei sei $\bar{x}_i = x_i$. Aufgrund der Konstruktion von G ist klar, dass

- (*) eine Belegung α genau dann F erfüllt, wenn für jede Kante $(l, l') \in E$ mit $\alpha(l) = 1$ auch $\alpha(l') = 1$ ist, und
- (**) l' von l aus genau dann erreichbar ist, wenn \bar{l} von \bar{l}' aus erreichbar ist.

Behauptung 68. F ist genau dann erfüllbar, wenn für keinen Knoten x_i in G ein Pfad von x_i über \bar{x}_i zurück nach x_i existiert.

Eine NL-Maschine kann bei Eingabe einer 2-KNF Formel F eine Variable x_i und einen Pfad von x_i über \bar{x}_i zurück nach x_i raten. Daher folgt aus der Behauptung, dass das Komplement von 2-SAT in NL ist. Wegen NL = co-NL folgt auch 2-SAT \in NL.

Nun zum Beweis der Behauptung. Wenn in G ein Pfad von x_i über \bar{x}_i nach x_i existiert, kann F nicht erfüllbar sein, da wegen (*) jede erfüllende Belegung, die x_i (bzw. \bar{x}_i) den Wert 1 zuweist, auch \bar{x}_i (bzw. x_i) diesen Wert zuweisen müsste. Existiert dagegen kein derartiger Pfad, so lässt sich für F wie folgt eine erfüllende Belegung α konstruieren:

- 1) Wähle einen beliebigen Knoten l aus G , für den $\alpha(l)$ noch undefiniert ist. Falls \bar{l} von l aus erreichbar ist, ersetze l durch \bar{l} (dies garantiert, dass \bar{l} von l aus nun nicht mehr erreichbar ist).

- 2) Weise jedem von l aus erreichbaren Knoten l' den Wert 1 (und \bar{l}' den Wert 0) zu.

- 3) Falls α noch nicht auf allen Knoten definiert ist, gehe zu 1).

Wegen (**) treten bei der Ausführung von 2) keine Konflikte auf:

- Hätte l' in einer früheren Runde den Wert 0 erhalten, dann hätte in dieser Runde \bar{l}' und somit auch \bar{l} den Wert 1 erhalten, was der Wahl von l widerspricht.
- Wäre von l aus auch \bar{l}' erreichbar, dann würde ein Pfad von l über \bar{l}' nach \bar{l} existieren, was durch die Wahl von l ebenfalls ausgeschlossen ist.

Zudem erfüllt α die Formel F , da für jede Kante $(l, l') \in E$ mit $\alpha(l) = 1$ auch $\alpha(l') = 1$ ist. ■

In den Übungen werden wir sehen, dass 2-SAT und REACH NL-vollständig sind.

6 Probabilistische Berechnungen

Eine **probabilistische Turingmaschine (PM)** M ist genau so definiert wie eine NTM. Es wird jedoch ein anderes Akzeptanzkriterium benutzt. Wir stellen uns vor, dass M in jedem Rechenschritt zufällig einen Konfigurationsübergang wählt. Dabei wird jeder mögliche Übergang $K \rightarrow_M K'$ mit derselben Wahrscheinlichkeit

$$\Pr[K \rightarrow_M K'] = \begin{cases} \|\{K'' | K \rightarrow_M K''\}\|^{-1}, & K \rightarrow_M K' \\ 0, & \text{sonst.} \end{cases}$$

gewählt. Eine Rechnung $\alpha = (K_1, K_2, \dots, K_m)$ wird also mit der Wahrscheinlichkeit

$$\Pr[\alpha] = \Pr[K_1 \rightarrow_M K_2 \rightarrow_M \dots \rightarrow_M K_m] = \prod_{i=1}^{m-1} \Pr[K_i \rightarrow_M K_{i+1}]$$

ausgeführt. Die **Akzeptanzwahrscheinlichkeit** von $M(x)$ ist

$$\Pr[M(x) \text{ akz.}] = \sum_{\alpha} \Pr[\alpha],$$

wobei sich die Summation über alle akzeptierenden Rechnungen α von $M(x)$ erstreckt. Wir vereinbaren für PMs, dass $M(x)$ am Ende jeder haltenden Rechnung entweder akzeptiert (hierfür schreiben wir kurz $M(x) = 1$) oder verwirft ($M(x) = 0$) oder „?“ ausgibt ($M(x) = ?$).

Die von einer PM M akzeptierte Sprache ist

$$L(M) = \{x \in \Sigma^* \mid \Pr[M(x) = 1] \geq 1/2\}.$$

6.1 Die Klassen PP, BPP, RP und ZPP

Eine Sprache $L \subseteq \Sigma^*$ gehört zur Klasse PP (probabilistic polynomial time), falls eine polynomiell zeitbeschränkte PM (PPTM) M mit $L(M) = L$ existiert.

Satz 69. $\text{co-NP} \subseteq \text{PP}$.

Beweis. Sei $L \in \text{co-NP}$ und sei N eine polynomiell zeitbeschränkte NTM mit $L(N) = \bar{L}$. Ersetzen wir in N den Zustand q_{ja} durch q_{nein} bzw. q_{nein} und q_{h} durch q_{ja} und fassen wir N als PPTM auf, so gilt für alle $x \in \Sigma^*$,

$$\begin{aligned} x \in L &\Rightarrow \Pr[N(x) = 1] = 1, \\ x \notin L &\Rightarrow \Pr[N(x) = 1] < 1. \end{aligned}$$

Betrachte folgende PPTM M , die bei Eingabe x zufällig eine der beiden folgenden Möglichkeiten wählt:

- M verwirft sofort,
- M simuliert N bei Eingabe x .

Dann gilt für alle $x \in \Sigma^*$,

$$\Pr[M(x) = 1] = \Pr[N(x) = 1]/2$$

und somit

$$\begin{aligned} x \in L &\Rightarrow \Pr[M(x) = 1] = 1/2, \\ x \notin L &\Rightarrow \Pr[M(x) = 1] < 1/2. \end{aligned} \quad \blacksquare$$

Als nächstes zeigen wir, dass PP unter Komplementbildung abgeschlossen ist. Das folgende Lemma zeigt, wie sich eine PPTM, die sich bei manchen Eingaben indifferent verhält (also genau mit Wahrscheinlichkeit $1/2$ akzeptiert) in eine äquivalente PPTM verwandeln lässt, die dies nicht tut.

Lemma 70. Für jede Sprache $L \in \text{PP}$ existiert eine PPTM M mit $L(M) = L$, die bei keiner Eingabe $x \in \Sigma^*$ mit Wahrscheinlichkeit $1/2$ akzeptiert, d.h. für alle x gilt

$$\Pr[M(x) \neq L(x)] < 1/2,$$

wobei $L(x)$ für alle $x \in L$ gleich 1 und für alle $x \notin L$ gleich 0 ist.

Beweis. Sei $L \in \text{PP}$ und sei N eine PPTM mit $L(N) = L$. Weiter sei p eine polynomielle Zeitschranke und $c \geq 2$ der maximale Verzweigungsgrad von N . Da $\Pr[N(x) = 1]$ nur Werte der Form $i/k^{-p(|x|)}$ für $k = \text{kgV}(2, \dots, c)$ annehmen kann, folgt für $\epsilon(x) = k^{-p(|x|)}$,

$$x \in L \Rightarrow \Pr[N(x) = 1] \geq 1/2,$$

$$x \notin L \Rightarrow \Pr[N(x) = 1] \leq 1/2 - \epsilon(x).$$

Sei N' eine PPTM mit $\Pr[N'(x) = 1] = 1/2 + \epsilon(x)/2$ und betrachte die PPTM M , die bei Eingabe x zufällig wählt, ob sie N oder N' bei Eingabe x simuliert. Dann gilt

$$\Pr[M(x) = 1] = \frac{\Pr[N(x) = 1] + \Pr[N'(x) = 1]}{2}$$

und somit

$$x \in L \Rightarrow \Pr[M(x) = 1] \geq \frac{1/2 + 1/2 + \epsilon(x)/2}{2} > 1/2$$

$$x \notin L \Rightarrow \Pr[M(x) = 1] \leq \frac{1/2 - \epsilon(x) + 1/2 + \epsilon(x)/2}{2} < 1/2. \quad \blacksquare$$

Eine direkte Folgerung von Lemma 70 ist der Komplementabschluss von PP.

Korollar 71. $\text{PP} = \text{co-PP}$.

Tatsächlich liefert Lemma 70 sogar den Abschluss von PP unter symmetrischer Differenz.

Satz 72. PP ist unter symmetrischer Differenz abgeschlossen, d.h.

$$L_1, L_2 \in \text{PP} \Rightarrow L_1 \Delta L_2 = (L_1 \setminus L_2) \cup (L_2 \setminus L_1) \in \text{PP}.$$

Beweis. Nach obigem Lemma existieren PPTMs M_1 und M_2 mit

$$x \in L_i \Rightarrow \Pr[M_i(x) = 1] = 1/2 + \epsilon_i,$$

$$x \notin L_i \Rightarrow \Pr[M_i(x) = 1] = 1/2 - \epsilon_i.$$

wobei $\epsilon_1, \epsilon_2 > 0$ sind und von x abhängen dürfen. Dann hat die PPTM M , die bei Eingabe x zunächst $M_1(x)$ und dann $M_2(x)$ simuliert und nur dann akzeptiert, wenn dies genau eine der beiden Maschinen tut, eine Akzeptanzwzk von

$$\Pr[M_1(x) = 1] \cdot \Pr[M_2(x) = 0] + \Pr[M_1(x) = 0] \cdot \Pr[M_2(x) = 1].$$

Folglich akzeptiert M alle Eingaben $x \in (L_1 \setminus L_2) \cup (L_2 \setminus L_1)$ mit Wahrscheinlichkeit

$$\begin{aligned} \Pr[M(x) = 1] &= (1/2 + \epsilon_1)(1/2 - \epsilon_2) + (1/2 - \epsilon_1)(1/2 + \epsilon_2) \\ &= (1/2 + 2\epsilon_1\epsilon_2) > 1/2 \end{aligned}$$

und alle Eingaben $x \in (L_1 \cap L_2) \cup \overline{L_1 \cup L_2}$ mit Wahrscheinlichkeit

$$\begin{aligned} \Pr[M(x) = 1] &= (1/2 + \epsilon_1)(1/2 - \epsilon_2) + (1/2 - \epsilon_1)(1/2 + \epsilon_2) \\ &= (1/2 - 2\epsilon_1\epsilon_2) < 1/2. \quad \blacksquare \end{aligned}$$

Anfang der 90er Jahre konnte auch der Abschluss von PP unter Schnitt und Vereinigung bewiesen werden. In den Übungen werden wir sehen, dass folgendes Problem PP-vollständig ist.

MajoritySat (MajSat):

Gegeben: Eine boolesche Formel $F(x_1, \dots, x_n)$.

Gefragt: Wird F von mindestens der Hälfte aller 2^n Belegungen erfüllt?

Definition 73. Sei M eine PPTM und sei $L = L(M)$. M heißt

- BPPTM, falls für alle x gilt: $\Pr[M(x) \neq L(x)] \leq 1/3$,
- RPTM, falls für alle $x \notin L$ gilt: $\Pr[M(x) = 1] = 0$,
- ZPPTM, falls für alle x gilt: $\Pr[M(x) = \bar{L}(x)] = 0$ und $\Pr[M(x) = ?] \leq 1/2$.

Die Klasse **BPP** (bounded error probabilistic polynomial time) enthält alle Sprachen, die von einer BPPTM akzeptiert werden. Entsprechend sind die Klassen **RP** (random polynomial time) und **ZPP** (zero error probabilistic polynomial time) definiert.

Man beachte, dass wir im Falle einer RPTM oder BPPTM M o.B.d.A. davon ausgehen können, dass M niemals ? ausgibt. Allerdings ist nicht ausgeschlossen, dass M ein falsches Ergebnis $M(x) = \bar{L}(x)$ liefert. Probabilistische Algorithmen mit dieser Eigenschaft werden auch als **Monte Carlo Algorithmen** bezeichnet. Im Unterschied zu einer BPPTM, die bei allen Eingaben x „lügen“ kann, ist dies einer RPTM nur im Fall $x \in L$ erlaubt. Man spricht hier auch von **ein-** bzw. **zweiseitigem Fehler**. Eine ZPPTM M darf dagegen überhaupt keine Fehler machen. Algorithmen von diesem Typ werden als **Las Vegas Algorithmen** bezeichnet.

Satz 74. $ZPP = RP \cap \text{co-RP}$.

Beweis. Die Inklusion von links nach rechts ist klar. Für die umgekehrte Richtung sei L eine Sprache in $RP \cap \text{co-RP}$. Dann existieren RPTMs M_1 und M_2 für L und \bar{L} , wobei wir annehmen, dass M_1 und M_2 niemals ? ausgeben. Weiter sei \bar{M}_2 die PPTM, die aus M_2 durch Vertauschen von q_{ja} und q_{nein} hervorgeht. Dann gilt

$$\begin{aligned} x \in L &\Rightarrow \Pr[M_1(x) = 1] \geq 1/2 \wedge \Pr[\bar{M}_2(x) = 1] = 1, \\ x \notin L &\Rightarrow \Pr[M_1(x) = 0] = 1 \wedge \Pr[\bar{M}_2(x) = 0] \geq 1/2. \end{aligned}$$

M_1 kann also nur Eingaben $x \in L$ akzeptieren, während \bar{M}_2 nur Eingaben $x \notin L$ verwerfen kann. Daher kann die Kombination $M_1(x) = 1$

und $\bar{M}_2(x) = 0$ nicht auftreten. Weiter ergeben sich hieraus für die PPTM M , die bei Eingabe x die beiden PPTMs $M_1(x)$ und $\bar{M}_2(x)$ simuliert und sich gemäß der Tabelle

	$M_1(x) = 1$	$M_1(x) = 0$
$\bar{M}_2(x) = 1$	1	?
$\bar{M}_2(x) = 0$	–	0

verhält, folgende Äquivalenzen:

$$\begin{aligned} M(x) = 1 &\Leftrightarrow M_1(x) = 1, \\ M(x) = 0 &\Leftrightarrow \bar{M}_2(x) = 0. \end{aligned}$$

Nun ist leicht zu sehen, dass folgende Implikationen gelten:

$$\begin{aligned} x \in L &\Rightarrow \Pr[M(x) = 1] = \Pr[M_1(x) = 1] \geq 1/2 \text{ und} \\ &\Pr[M(x) = 0] = \Pr[\bar{M}_2(x) = 0] = 0 \\ x \notin L &\Rightarrow \Pr[M(x) = 0] = \Pr[\bar{M}_2(x) = 0] \geq 1/2 \text{ und} \\ &\Pr[M(x) = 1] = \Pr[M_1(x) = 1] = 0. \end{aligned}$$

Dies zeigt, dass M eine ZPPTM für L ist, da für alle Eingaben x gilt:

$$\Pr[M(x) = \bar{L}(x)] = 0 \text{ und } \Pr[M(x) = ?] \leq 1/2. \quad \blacksquare$$

6.2 Anzahl-Operatoren

Definition 75 (Anzahlklassen). Sei C eine Sprachklasse und sei $B \subseteq \Sigma^*$ eine p -balancierte Sprache in C . Durch B werden die Funktion $\#B : \Sigma^* \rightarrow \mathbb{N}$ mit

$$\#B(x) = \|\{y \in \{0, 1\}^* \mid x\#y \in B\}\|$$

und folgende Sprachen definiert (n bezeichnet die Länge von x):

$$\begin{aligned}\exists^p B &= \{x \in \Sigma^* \mid \#B(x) > 0\}, \\ \forall^p B &= \{x \in \Sigma^* \mid \#B(x) = 2^{p(n)}\}, \\ \exists^{\geq 1/2} B &= \{x \in \Sigma^* \mid \#B(x) \geq 2^{p(n)-1}\} \\ \oplus B &= \{x \in \Sigma^* \mid \#B(x) \text{ ist ungerade}\}.\end{aligned}$$

Für $\text{Op} \in \{\#, \exists^p, \forall^p, \exists^{\geq 1/2}, \oplus\}$ sei

$$\text{Op} \cdot \mathbf{C} = \{\text{Op}B \mid B \in \mathbf{C} \text{ ist polynomiell balanciert}\}$$

die durch Anwendung des Operators Op auf \mathbf{C} definierte Funktionen- bzw. Sprachklasse. Zudem definieren wir die Operatoren \mathbf{R} und \mathbf{BP} durch

$$\mathbf{R} \cdot \mathbf{C} = \{\exists^{\geq 1/2} B \mid B \in \mathbf{C} \text{ ist einseitig}\}$$

und

$$\mathbf{BP} \cdot \mathbf{C} = \{\exists^{\geq 1/2} B \mid B \in \mathbf{C} \text{ ist zweiseitig}\}.$$

Dabei heißt eine p -balancierte Sprache B **einseitig**, falls $\#B(x)$ für keine Eingabe x einen Wert in $[1, 2^{p(n)-1} - 1]$ und **zweiseitig**, falls $\#B(x)$ für kein x einen Wert in $(2^{p(n)}/3, 2^{p(n)+1}/3)$ annimmt.

Wir wissen bereits, dass $\exists^p \cdot \mathbf{P} = \mathbf{NP}$ und $\forall^p \cdot \mathbf{P} = \mathbf{co-NP}$ ist. Als nächstes zeigen wir, dass $\mathbf{R} \cdot \mathbf{P} = \mathbf{RP}$, $\exists^{\geq 1/2} \cdot \mathbf{P} = \mathbf{PP}$ und $\mathbf{BP} \cdot \mathbf{P} = \mathbf{BPP}$ ist.

Lemma 76. Sei M eine PPTM und sei $\beta(x) = \Pr[M(x) = 1]$. Dann existieren ein Polynom q und q -balancierte Sprachen $B, B' \in \mathbf{P}$, so dass für alle Eingaben x , $|x| = n$, gilt:

$$\beta(x)/2 \leq \Pr_{y \in_R \{0,1\}^{q(n)}}[x\#y \in B] \leq \beta(x) \text{ und} \quad (6.1)$$

$$(\beta(x) - 1/2)/2 \leq \Pr_{z \in_R \{0,1\}^{q(n)}}[x\#z \in B'] - 1/2 \leq \beta(x) - 1/2. \quad (6.2)$$

Beweis. Sei p eine polynomielle Zeitschranke für M und sei $k = \text{kgV}(2, 3, \dots, c)$, wobei $c \geq 1$ der maximale Verzweigungsgrad von M ist. Sei x eine Eingabe der Länge n . Wir ordnen jeder Folge $r = r_1 \cdots r_{p(n)}$ aus der Menge $R_n = \{0, \dots, k-1\}^{p(n)}$ eindeutig eine Rechnung von $M(x)$ zu, indem wir im i -ten Rechenschritt aus den c_i zur Auswahl stehenden Folgekonfigurationen K_0, \dots, K_{c_i-1} die $(r_i \bmod c_i)$ -te wählen. Bezeichnen wir das Ergebnis der so beschriebenen Rechnung mit $M_r(x)$, so gilt

$$\beta(x) = \Pr_{r \in_R R_n}[M_r(x) = 1].$$

Sei nun $q(n) = \lceil \log_2(k^{p(n)}) \rceil$ und $\alpha(n) = k^{p(n)}/2^{q(n)}$. Dann können wir die ersten $k^{p(n)}$ Strings $y \in \{0, 1\}^{q(n)}$ zur Kodierung der Folgen $r \in R_n$ verwenden und für $M_r(x)$ auch $M_y(x)$ schreiben. Sei D_n die Menge der ersten $k^{p(n)}$ Strings in $\{0, 1\}^{q(n)}$ und sei $D'_n = \{0, 1\}^{q(n)} \setminus D_n$. Betrachte nun die Sprachen

$$\begin{aligned}B &= \{x\#y \mid y \in D_n \text{ und } M_y(x) = 1\} \text{ und} \\ B' &= B \cup \{x\#y \mid y \in D'_n \text{ und } y \text{ endet mit } 0\}.\end{aligned}$$

Dann gilt für ein zufällig aus $\{0, 1\}^{q(n)}$ gewähltes y :

$$\Pr[x\#y \in B] = \Pr[y \in D_n] \cdot \Pr[M_y(x) = 1 \mid y \in D_n] = \alpha(n)\beta(x).$$

Da D_n mehr als die Hälfte aller Strings der Länge $q(n)$ enthält, ist $\alpha(n) \in (1/2, 1]$, woraus (6.1) folgt. Da zudem $2^{q(n)} - k^{p(n)}$ gerade ist, enthält die Sprache B' für genau die Hälfte der Strings $y \in D'_n$ das Wort $x\#y$. Daher ist

$$\begin{aligned}\Pr[x\#y \in B'] &= \Pr[x\#y \in B] + \Pr[y \in D'_n] \cdot 1/2 \\ &= \alpha(n)\beta(x) + (1 - \alpha(n))1/2 \\ &= \alpha(n)(\beta(x) - 1/2) + 1/2,\end{aligned}$$

woraus auch (6.2) folgt. ■

Satz 77. $\mathbf{R} \cdot \mathbf{P} = \mathbf{RP}$, $\exists^{\geq 1/2} \cdot \mathbf{P} = \mathbf{PP}$ und $\mathbf{BP} \cdot \mathbf{P} = \mathbf{BPP}$.

Beweis. Die Inklusionen $R \cdot P \subseteq RP$, $\exists^{\geq 1/2} \cdot P \subseteq PP$ und $BP \cdot P \subseteq BPP$ sind klar.

Zum Nachweis von $RP \subseteq R \cdot P$ sei $L \in RP$ und M eine RPTM für L . Nach Lemma 76 existieren ein Polynom q und eine q -balancierte Sprache $B \in P$, so dass für alle Eingaben x , $|x| = n$, gilt:

$$\begin{aligned} x \in L &\Rightarrow \Pr[M(x) = 1] \geq 1/2 \Rightarrow \Pr_{y \in_R \{0,1\}^{q(n)}}[x\#y \in B] \geq 1/4, \\ x \notin L &\Rightarrow \Pr[M(x) = 1] = 0 \Rightarrow \Pr_{y \in_R \{0,1\}^{q(n)}}[x\#y \in B] = 0. \end{aligned}$$

Betrachte nun die $3q$ -balancierte Sprache

$$B' = \{x\#uvw \mid u, v, w \in \{0,1\}^{q(n)}, \exists y \in \{u, v, w\} : x\#y \in B\}.$$

Da für alle Eingaben x , $|x| = n$, die Implikationen

$$\begin{aligned} x \in L &\Rightarrow \Pr_{z \in_R \{0,1\}^{3q(n)}}[x\#z \in B'] \geq 1 - (3/4)^3 > 1/2, \\ x \notin L &\Rightarrow \Pr_{z \in_R \{0,1\}^{3q(n)}}[x\#z \in B'] = 0 \end{aligned}$$

gelten, ist B' einseitig und es folgt $L = \exists^{\geq 1/2} B' \in R \cdot P$.

Zum Nachweis von $PP \subseteq \exists^{\geq 1/2} \cdot P$ sei $L \in PP$ und M eine PPTM für L . Nach Lemma 76 existieren ein Polynom q und eine q -balancierte Sprache $B' \in P$, so dass für alle Eingaben x , $|x| = n$, gilt:

$$\begin{aligned} x \in L &\Rightarrow \Pr[M(x) = 1] - 1/2 \geq 0 \Rightarrow \Pr_{y \in_R \{0,1\}^{q(n)}}[x\#y \in B'] \geq 1/2, \\ x \notin L &\Rightarrow \Pr[M(x) = 1] - 1/2 < 0 \Rightarrow \Pr_{y \in_R \{0,1\}^{q(n)}}[x\#y \in B'] < 1/2, \end{aligned}$$

also $L = \exists^{\geq 1/2} B' \in \exists^{\geq 1/2} \cdot P$.

Zum Nachweis von $BPP \subseteq BP \cdot P$ sei $L \in BPP$ und M eine BPPTM für L . Nach Lemma 76 existieren ein Polynom q und eine q -balancierte Sprache $B' \in P$, so dass für alle Eingaben x , $|x| = n$, gilt:

$$\begin{aligned} x \in L &\Rightarrow \Pr[M(x) = 1] - 1/2 \geq 1/6 \Rightarrow \Pr_{y \in_R \{0,1\}^{q(n)}}[x\#y \in B'] \geq 7/12, \\ x \notin L &\Rightarrow \Pr[M(x) = 1] - 1/2 \leq 1/6 \Rightarrow \Pr_{y \in_R \{0,1\}^{q(n)}}[x\#y \in B'] \leq 1/3, \end{aligned}$$

also $L = \exists^{\geq 1/2} B' \in BP' \cdot P = BP \cdot P$ (siehe Übungen). \blacksquare

6.3 Reduktion der Fehlerwahrscheinlichkeit

In diesem Abschnitt zeigen wir, wie sich für RP-, ZPP- und BPP-Maschinen M die Wahrscheinlichkeit $\Pr[M(x) \neq A(x)]$ für ein inkorrektes oder indifferentes Ergebnis für ein beliebiges Polynom q auf einen exponentiell kleinen Wert $2^{-q(|x|)}$ reduzieren lässt.

Definition 78. A ist auf B **disjunktiv reduzierbar** (in Zeichen: $A \leq_d B$), falls eine Funktion $f \in FL$ existiert, die für jedes Wort x eine Liste $y_1\#\dots\#y_m$ von Wörtern y_i berechnet mit

$$x \in A \Leftrightarrow \exists i \in \{1, \dots, m\} : y_i \in B.$$

Gilt dagegen

$$x \in A \Leftrightarrow \|\{i \in \{1, \dots, m\} \mid y_i \in B\}\| \geq m/2,$$

so heißt A **majority-reduzierbar** auf B , wofür wir auch kurz $A \leq_{maj} B$ schreiben.

Es ist leicht zu sehen, dass die Klassen P , NP und $co-NP$ unter beiden Typen von Reduktionen abgeschlossen sind.

Satz 79. Falls C unter disjunktiven Reduktionen abgeschlossen ist, existieren für jede Sprache $A \in R \cdot C$ und jedes Polynom q eine Sprache $B \in C$ und ein Polynom r , so dass für alle x , $|x| = n$, gilt:

$$\begin{aligned} x \in A &\Rightarrow \Pr_{z \in_R \{0,1\}^{r(n)}}[B(x\#z) = 1] \geq 1 - 2^{-q(|x|)}, \\ x \notin A &\Rightarrow \Pr_{z \in_R \{0,1\}^{r(n)}}[B(x\#z) = 0] = 1. \end{aligned}$$

Beweis. Sei $A \in R \cdot C$ und sei $B \in C$ eine p -balancierte Sprache mit

$$\begin{aligned} x \in A &\Rightarrow \Pr_{y \in_R \{0,1\}^{p(n)}}[B(x\#y) = 1] \geq 1/2, \\ x \notin A &\Rightarrow \Pr_{y \in_R \{0,1\}^{p(n)}}[B(x\#y) = 0] = 1. \end{aligned}$$

Dann ist die Sprache

$$B = \{x\#y_1 \cdots y_{q(n)} \mid y_1, \dots, y_{q(n)} \in \{0, 1\}^{p(n)}, \exists i : x\#y_i \in B\}$$

disjunktiv auf B reduzierbar und somit in \mathbf{C} . Außerdem ist leicht zu sehen, dass B für $r(n) = p(n)q(n)$ die im Satz genannten Eigenschaften besitzt. ■

Korollar 80. Für jedes Polynom q und jede Sprache $L \in \mathbf{RP}$ existiert eine RPTM M mit

$$\begin{aligned} x \in L &\Rightarrow \Pr[M(x) = 1] \geq 1 - 2^{-q(|x|)}, \\ x \notin L &\Rightarrow \Pr[M(x) = 0] = 1. \end{aligned}$$

Ganz analog lässt sich die Zuverlässigkeit einer ZPPTM verbessern.

Korollar 81. Für jedes Polynom q und jede Sprache $L \in \mathbf{ZPP}$ existiert eine ZPPTM M mit $L(M) = L$ und $\Pr[M(x) = ?] \leq 2^{-q(|x|)}$ für jede Eingabe x .

Für die Reduktion der Fehlerwahrscheinlichkeit von BPPTMs benötigen wir das folgende Lemma.

Lemma 82. Sei E ein Ereignis, das mit Wahrscheinlichkeit $1/2 - \epsilon$, $\epsilon > 0$, auftritt. Dann ist die Wahrscheinlichkeit, dass sich E bei $m = 2t + 1$ unabhängigen Wiederholungen mehr als t -mal ereignet, höchstens $1/2(1 - 4\epsilon^2)^t$.

Beweis. Für $i = 1, \dots, m$ sei X_i die Indikatorvariable

$$X_i = \begin{cases} 1, & \text{Ereignis } E \text{ tritt beim } i\text{-ten Versuch ein,} \\ 0, & \text{sonst} \end{cases}$$

und X sei die Zufallsvariable $X = \sum_{i=1}^m X_i$. Dann ist X binomial verteilt mit Parametern m und $p = 1/2 - \epsilon$. Folglich gilt für $i > m/2$,

$$\begin{aligned} \Pr[X = i] &= \binom{m}{i} (1/2 - \epsilon)^i (1/2 + \epsilon)^{m-i} \\ &= \binom{m}{i} (1/2 - \epsilon)^{m/2} (1/2 + \epsilon)^{m/2} \left(\frac{1/2 - \epsilon}{1/2 + \epsilon}\right)^{i-m/2} \\ &\leq \binom{m}{i} \underbrace{(1/2 - \epsilon)^{m/2} (1/2 + \epsilon)^{m/2}}_{(1/4 - \epsilon^2)^{m/2}}. \end{aligned}$$

Wegen

$$\sum_{i=t+1}^m \binom{m}{i} \leq 2^{m-1} = \frac{4^{m/2}}{2}$$

erhalten wir somit

$$\begin{aligned} \sum_{i=t+1}^m \Pr[X = i] &\leq (1/4 - \epsilon^2)^{m/2} \sum_{i=t+1}^m \binom{m}{i} \\ &\leq \frac{(1 - 4\epsilon^2)^{m/2}}{2} \leq \frac{(1 - 4\epsilon^2)^t}{2}. \end{aligned}$$

Satz 83. Falls \mathbf{C} unter majority-Reduktionen abgeschlossen ist, existieren für jede Sprache $A \in \mathbf{BP} \cdot \mathbf{C}$ und jedes Polynom q eine Sprache $B \in \mathbf{C}$ und ein Polynom r , so dass für alle x , $|x| = n$, gilt:

$$\Pr_{z \in_R \{0,1\}^{r(n)}} [A(x) = B(x\#z)] \geq 1 - 2^{-q(n)}.$$

Beweis. Sei $A \in \mathbf{BP} \cdot \mathbf{C}$ und sei $B' \in \mathbf{C}$ eine p -balancierte Sprache mit

$$\Pr_{y \in_R \{0,1\}^{p(n)}} [A(x) \neq B'(x\#y)] \leq 1/3 = 1/2 - 1/6.$$

Dann ist die Sprache

$$B = \left\{ x\#y_1 \cdots y_{2t(n)+1} \mid \begin{array}{l} y_1, \dots, y_{2t(n)+1} \in \{0, 1\}^{p(n)}, \\ \|\{i : x\#y_i \in B\}\| > t(n) \end{array} \right\},$$

wobei $t(n) = \lceil (q(n) - 1) \log_2(9/8) \rceil$ ist, auf B' majority-reduzierbar und somit in \mathcal{C} . Weiter folgt nach Lemma 82 für $r(n) = p(n)(2t(n)+1)$ und für ein zufällig gewähltes $z = y_1 \cdots y_{2t(n)+1} \in_R \{0, 1\}^{r(n)}$

$$\begin{aligned} \Pr[A(x) \neq B(x\#z)] &= \Pr[\|\{i : A(x) \neq A(x\#y_i)\}\| > t(n)] \\ &\leq \frac{1}{2} \underbrace{\left(1 - \frac{4}{36}\right)^{t(n)}}_{8/9} \leq 2^{-q(|x|)}. \end{aligned} \quad \blacksquare$$

Korollar 84. Für jede Sprache $L \in \text{BPP}$ und jedes Polynom q ex. eine BPPTM M mit

$$\Pr[M(x) = L(x)] \geq 1 - 2^{-q(|x|)}.$$

Satz 85. $\text{BPP} \subseteq \text{PSK}$.

Beweis. Sei $A \in \text{BPP}$. Dann ist auch die Sprache $A' = \text{bin}(A)$ in $\text{BPP} = \text{BP} \cdot \text{P}$. Nach Satz 83 existieren für das Polynom $q(n) = n + 1$ ein Polynom r und eine r -balancierte Sprache $B \in \mathcal{C}$, so dass für alle n und alle $x \in \{0, 1\}^n$ gilt:

$$\Pr_{z \in_R \{0, 1\}^{r(n)}} [B(x\#z) \neq A'(x)] \leq 2^{-q(n)}.$$

Daher folgt für ein zufällig gewähltes $z \in_R \{0, 1\}^{r(n)}$,

$$\Pr[\exists x \in \{0, 1\}^n : B(x\#z) \neq A'(x)] \leq \sum_{x \in \{0, 1\}^n} \Pr[B(x\#z) \neq A'(x)] < 1.$$

Also muss für jede Eingabelänge n ein String z_n mit $A'(x) = B(x\#z_n)$ für alle $x \in \{0, 1\}^n$ existieren. Da $B \in \text{P}$ ist, können die Funktionen $f_{z_n} : x \mapsto B(x\#z_n)$ nach Korollar 50 durch Schaltkreise polynomieller Größe berechnet werden. \blacksquare

6.4 Abschlusseigenschaften von Anzahl-Klassen

In diesem Abschnitt gehen wir der Frage nach, unter welchen Reduktionen und Operatoren Anzahl-Klassen abgeschlossen sind.

Lemma 86. Sei \mathcal{C} eine Sprachklasse, die unter \leq_m^{\log} abgeschlossen ist. Dann gilt

- (i) $\text{co-}\exists^p \cdot \mathcal{C} = \forall^p \cdot \text{co-}\mathcal{C}$,
- (ii) $\text{co-BP} \cdot \mathcal{C} = \text{BP} \cdot \text{co-}\mathcal{C}$,
- (iii) $\text{co-}\oplus \cdot \mathcal{C} = \oplus \cdot \text{co-}\mathcal{C} = \oplus \cdot \mathcal{C}$.

Beweis. Wir zeigen nur die Inklusionen von links nach rechts. Die umgekehrten Inklusionen folgen analog.

- (i) Sei $A \in \text{co-}\exists^p \cdot \mathcal{C}$. Dann existieren ein Polynom p und eine p -balancierte Sprache $B \in \mathcal{C}$ mit $\bar{A} = \exists^p B$. Definiere die Sprache

$$\hat{B} = \{x\#y \mid x\#y \notin B \text{ und } |y| = p(|x|)\}.$$

Dann ist \hat{B} eine ebenfalls p -balancierte Sprache in $\text{co-}\mathcal{C}$ mit $\#\hat{B}(x) = 2^{p(n)} - \#B(x)$. Daher folgt

$$x \in A \Leftrightarrow \#B(x) = 0 \Leftrightarrow \#\hat{B}(x) = 2^{p(n)},$$

also $A = \forall^p \hat{B} \in \forall^p \cdot \text{co-}\mathcal{C}$.

- (ii) Sei $A \in \text{co-BP} \cdot \mathcal{C}$ und sei $B \in \mathcal{C}$ eine p -balancierte zweiseitige Sprache mit $\bar{A} = \exists^{\geq 1/2} B$. Dann ist die in (i) definierte p -balancierte Sprache $\hat{B} \in \text{co-}\mathcal{C}$ ebenfalls zweiseitig und es gilt $A = \exists^{\geq 1/2} \hat{B} \in \text{BP} \cdot \text{co-}\mathcal{C}$.

- (iii) Sei $A \in \oplus \cdot \mathcal{C}$ und sei $B \in \mathcal{C}$ eine p -balancierte Sprache mit $A = \oplus B$. Dann hat die in (i) definierte p -balancierte Sprache $\hat{B} \in \text{co-}\mathcal{C}$ die gleiche Parität wie $\#B(x)$ und daher gilt $A = \oplus \hat{B} \in \oplus \cdot \text{co-}\mathcal{C}$. Zudem ist

$$B' = \{x\#0y \mid x\#y \in B\} \cup \{x\#1^{p(|x|)+1}\}$$

eine $(p+1)$ -balancierte Sprache in \mathcal{C} mit $\bar{A} = \oplus B'$, d.h. $A \in \text{co-}\oplus \cdot \mathcal{C}$. \blacksquare

Lemma 87. Sei $B \in \mathcal{C}$ eine p -balancierte Sprache und sei \mathcal{C} unter \leq_m^{\log} abgeschlossen. Dann existieren für jede Funktion $f \in \text{FL}$ ein Polynom q und q -balancierte Sprachen $B', B'' \in \mathcal{C}$ mit

$$\#B'(x) = \#B(f(x)) \text{ und } \#B''(x)/2^{q(|x|)} = \#B(f(x))/2^{p(|f(x)|)}.$$

Beweis. Sei q ein Polynom mit $p(|f(x)|) \leq q(|x|)$ für alle x . Betrachte die Sprachen

$$B' = \{x\#y'y'' \mid |y'y''| = q(n), f(x)\#y' \in B \text{ und } y'' \in \{0\}^*\} \text{ und} \\ B'' = \{x\#y'y'' \mid |y'y''| = q(n), f(x)\#y' \in B \text{ und } y'' \in \{0, 1\}^*\}.$$

Dann gilt $B', B'' \leq_m^{\log} B$. Da jedes Präfix y' mit $f(x)\#y' \in B$ genau eine Verlängerung y'' mit $x\#y'y'' \in B'$ und genau $2^{q(|x|)-p(|f(x)|)}$ Verlängerungen y'' mit $x\#y'y'' \in B''$ hat, folgt

$$\#B'(x) = \#B(f(x)) \text{ und } \#B''(x) = \#B(f(x))2^{q(|x|)-p(|f(x)|)}. \blacksquare$$

Mit obigem Lemma ist es nun leicht, folgende Abschlusseigenschaften der Anzahlklassen $\exists^p \cdot \mathcal{C}$, $\forall^p \cdot \mathcal{C}$, $\mathcal{R} \cdot \mathcal{C}$, $\text{BP} \cdot \mathcal{C}$ und $\oplus \cdot \mathcal{C}$ zu zeigen.

Satz 88. Sei \mathcal{C} eine unter \leq_m^{\log} abgeschlossene Sprachklasse.

- (i) Für $\text{Op} \in \{\exists^p, \forall^p, \mathcal{R}, \text{BP}, \exists^{\geq 1/2}, \oplus\}$ ist auch $\text{Op} \cdot \mathcal{C}$ unter \leq_m^{\log} abgeschlossen.
- (ii) $\exists^p \cdot \exists^p \cdot \mathcal{C} = \exists^p \cdot \mathcal{C}$, $\forall^p \cdot \forall^p \cdot \mathcal{C} = \forall^p \cdot \mathcal{C}$ und $\oplus \cdot \oplus \cdot \mathcal{C} = \oplus \cdot \mathcal{C}$.

Beweis.

- (i) Sei $A \in \text{Op} \cdot \mathcal{C}$ mittels einem Polynom p und einer p -balancierten Sprache $B \in \mathcal{C}$. Weiter gelte $A' \leq A$ mittels einer FL-Funktion f . Nach obigem Lemma existieren ein Polynom q und eine q -balancierte Sprache $B' \in \mathcal{C}$ mit

$$\#B'(x)/2^{q(|x|)} = \#B(f(x))/2^{p(|f(x)|)}.$$

Nun folgt z.B. für $\text{Op} = \exists^p$,

$$x \in A' \Leftrightarrow f(x) \in A \Leftrightarrow \#B(f(x)) > 0 \Leftrightarrow \#B'(x) > 0,$$

weshalb $A' = \exists^p B' \in \exists^p \cdot \mathcal{C}$ ist. Die übrigen Fälle folgen analog.

- (ii) Siehe Übungen. ■

Das nächste Satz zeigt, dass mit \mathcal{C} auch $\text{BP} \cdot \mathcal{C}$ unter majority-Reduktionen abgeschlossen ist.

Satz 89. Sei \mathcal{C} eine unter majority-Reduktionen abgeschlossene Sprachklasse.

- (i) Für $\text{Op} \in \{\exists^p, \forall^p, \text{BP}\}$ ist auch $\text{Op} \cdot \mathcal{C}$ unter majority-Reduktionen abgeschlossen.
- (ii) Für $\text{Op} \in \{\exists^p, \forall^p, \exists^{\geq 1/2}, \oplus\}$ gilt $\text{Op} \cdot \text{BP} \cdot \mathcal{C} \subseteq \text{BP} \cdot \text{Op} \cdot \mathcal{C}$.

Beweis. (i) Sei $A \in \text{BP} \cdot \mathcal{C}$ und gelte $A' \leq_{\text{maj}} A$ mittels einer FL-Funktion $f : x \mapsto y_1\#\dots\#y_{m(|x|)}$, wobei wir annehmen können, dass $m(n) \equiv_2 1$ für alle n ist. Da \mathcal{C} unter majority-Reduktionen (und damit auch unter \leq_m^{\log}) abgeschlossen ist, ist nach Satz 88 auch $\text{BP} \cdot \mathcal{C}$ unter \leq_m^{\log} abgeschlossen. Daher können wir annehmen, dass alle $y_i \in \{0, 1\}^{r(n)}$ für ein Polynom r sind.

Zudem existieren in den Fällen $\text{Op} \in \{\exists^p, \forall^p\}$ ein Polynom p und eine p -balancierte Sprache $B \in \mathcal{C}$ mit $A = \text{Op}B$. Dann ist die $p(n)m(n)$ -balancierte Sprache

$$B' = \left\{ x\#z_1 \dots z_m \left| \begin{array}{l} z_1, \dots, z_m \in \{0, 1\}^{p(n)}, f(x) = y_1\#\dots\#y_m \\ \text{und } \|\{i \mid y_i\#z_i \in B\}\| \geq m/2 \end{array} \right. \right\}$$

auf B majority-reduzierbar und somit in \mathcal{C} , was wiederum wegen $A' = \text{Op}B'$ impliziert, dass $A' \in \text{Op} \cdot \mathcal{C}$ ist.

Da \mathcal{C} unter majority-Reduktionen abgeschlossen ist, können wir im Fall $\text{Op} = \text{BP}$ annehmen, dass $A = \exists^{\geq 1/2}B$ für eine p -balancierte Sprache $B \in \mathcal{C}$ mit

$$\Pr_{z \in_R \{0,1\}^{p(n)}} [B(y\#z) \neq A(y)] \leq 2^{-m(n)-1}$$

für alle $y \in \{0, 1\}^{r(n)}$ ist. Nun folgt wegen

$$\Pr_{z \in_R \{0,1\}^{p(n)m(n)}} [B'(x\#z) \neq A'(x)] \leq m(n)2^{-m(n)-1} < \frac{1}{3},$$

dass $A' \in \text{BP} \cdot \text{C}$ ist.

(ii) Für $\text{Op} \in \{\exists^p, \forall^p, \exists^{\geq 1/2}, \oplus\}$ sei $A' \in \text{Op} \cdot \text{BP} \cdot \text{C}$. Dann existieren ein Polynom r und eine r -balancierte Sprache $A \in \text{BP} \cdot \text{C}$ mit $A' = \text{Op}A$. Zu r und A existieren ein Polynom p und eine Sprache $B \in \text{C}$ mit

$$\Pr_{z \in_R \{0,1\}^{p(n)}} [A(x\#y) \neq B(x\#y\#z)] \leq 2^{-r(n)-2}.$$

Dann ist auch die Sprache

$$B' = \{x\#z\#y \mid x\#y\#z \in B\}$$

in C und wegen

$$\begin{aligned} & \Pr_{z \in_R \{0,1\}^{p(n)}} [\text{Op}B'(x\#z) \neq A'(x)] \\ & \leq \Pr_{z \in_R \{0,1\}^{p(n)}} [\exists y \in \{0, 1\}^{r(n)} : B'(x\#z\#y) \neq A(x\#y)] \\ & \leq 2^{r(n)} 2^{-r(n)-2} \\ & < \frac{1}{3} \end{aligned}$$

folgt $A' \in \text{BP} \cdot \text{Op} \cdot \text{C}$. ■

Durch Kombination von Satz 89 und Satz 88 erhalten wir sofort folgende Abschlusseigenschaften.

Korollar 90. *Sei C eine unter majority-Reduktionen abgeschlossene Sprachklasse. Dann gilt*

- (i) $\exists^p \cdot \text{BP} \cdot \exists^p \cdot \text{C} = \text{BP} \cdot \exists^p \cdot \text{C}$,
- (ii) $\forall^p \cdot \text{BP} \cdot \forall^p \cdot \text{C} = \text{BP} \cdot \forall^p \cdot \text{C}$ und
- (iii) $\oplus \cdot \text{BP} \cdot \oplus \cdot \text{C} = \text{BP} \cdot \oplus \cdot \text{C}$.

Nun folgt auch leicht der Abschluss von $\text{BP} \cdot \text{C}$ unter dem BP-Operator, falls C unter majority-Reduktionen abgeschlossen ist.

Lemma 91. *Für jede Klasse C , die unter majority-Reduktionen abgeschlossen ist, gilt*

$$\text{BP} \cdot \text{BP} \cdot \text{C} = \text{BP} \cdot \text{C}.$$

Beweis. Sei $A \in \text{BP} \cdot \text{BP} \cdot \text{C}$. Da mit C auch $\text{BP} \cdot \text{C}$ unter majority-Reduktionen abgeschlossen ist, existieren eine Sprache $B \in \text{BP} \cdot \text{C}$ und ein Polynom p , so dass für alle x , $|x| = n$, gilt

$$\Pr_{y \in_R \{0,1\}^{p(n)}} [B(x\#y) \neq A(x)] \leq 1/6.$$

Zudem existieren eine Sprache $B' \in \text{C}$ und ein Polynom q , so dass für alle x und $y \in \{0, 1\}^{p(n)}$ gilt

$$\Pr_{z \in_R \{0,1\}^{q(n)}} [B(x\#y) \neq B'(x\#y\#z)] \leq 1/6.$$

Mit B' ist auch die $(p+q)$ -balancierte Sprache

$$B'' = \{x\#yz \mid y \in \{0, 1\}^{p(n)}, z \in \{0, 1\}^{q(n)}, x\#y\#z \in B'\}$$

in C und wegen

$$\Pr_{u \in_R \{0,1\}^{p(n)+q(n)}} [A(x) \neq B''(x\#u)] \leq 1/6 + 1/6 = 1/3$$

folgt $A \in \text{BP} \cdot \text{C}$. ■

Folglich sind die Klassen BPP und $\text{BP} \cdot \text{NP}$ unter dem BP-Operator abgeschlossen. Analog folgt für jede Sprachklasse C , die unter disjunktiven Reduktionen abgeschlossen ist, die Gleichheit $\text{R} \cdot \text{R} \cdot \text{C} = \text{R} \cdot \text{C}$.

7 Die Polynomialzeithierarchie

Die Polynomialzeithierarchie extrapoliert den Übergang von P zu $\exists^p \cdot P = NP$ und besteht aus den Stufen Σ_k^p und Π_k^p , $k \geq 0$, welche induktiv wie folgt definiert sind:

$$\begin{aligned} \Sigma_0^p &= P, & \Pi_0^p &= P, \\ \Sigma_{k+1}^p &= \exists^p \cdot \Pi_k^p, & \Pi_{k+1}^p &= \forall^p \cdot \Sigma_k^p, \quad k \geq 0. \end{aligned}$$

Die Vereinigung aller Stufen der Polynomialzeithierarchie bezeichnen wir mit PH ,

$$PH = \bigcup_{k \geq 0} \Sigma_k^p = \bigcup_{k \geq 0} \Pi_k^p.$$

Es ist leicht zu sehen, dass $\Sigma_k^p = \text{co-}\Pi_k^p$ ist. Es ist nicht bekannt, ob die Polynomialzeithierarchie echt ist, also $\Sigma_k^p \neq \Sigma_{k+1}^p$ für alle $k \geq 0$ gilt. Die Annahme $\Sigma_k^p = \Sigma_{k+1}^p$ ist mit einem Kollaps von PH auf die k -te Stufe äquivalent. Es gilt allerdings als unwahrscheinlich, dass die Polynomialzeithierarchie kollabiert, erst recht nicht auf eine kleine Stufe.

Satz 92. *Für alle $k \geq 1$ gilt: $\Sigma_k^p = \Sigma_{k+1}^p \Leftrightarrow \Sigma_k^p = \Pi_k^p \Leftrightarrow PH = \Sigma_k^p$.*

Beweis. Wir zeigen die drei Implikationen $\Sigma_k^p = \Sigma_{k+1}^p \Rightarrow \Sigma_k^p = \Pi_k^p \Rightarrow PH = \Sigma_k^p \Rightarrow \Sigma_k^p = \Sigma_{k+1}^p$. Wegen $\Pi_k^p \subseteq \Sigma_{k+1}^p$ impliziert die Gleichheit $\Sigma_k^p = \Sigma_{k+1}^p$ sofort $\Pi_k^p \subseteq \Sigma_k^p$, was mit $\Sigma_k^p = \Pi_k^p$ gleichbedeutend ist. Für die zweite Implikation sei $\Sigma_k^p = \Pi_k^p$ angenommen. Wir zeigen durch Induktion über l , dass dann $\Sigma_{k+l}^p = \Sigma_k^p$ für alle $l \geq 0$ gilt. Der Induktionsanfang $l = 0$ ist klar. Für den Induktionsschritt setzen wir die Gleichheit $\Sigma_{k+l}^p = \Sigma_k^p$ (bzw. $\Pi_{k+l}^p = \Pi_k^p$) voraus und folgern

$$\Sigma_{k+l+1}^p = \exists^p \cdot \Pi_{k+l}^p = \exists^p \cdot \Pi_k^p = \exists^p \cdot \Sigma_k^p = \Sigma_k^p.$$

Die Implikation $PH = \Sigma_k^p \Rightarrow \Sigma_k^p = \Sigma_{k+1}^p$ ist klar. ■

Als Folgerung hieraus ergibt sich, dass eine NP -vollständige Sprache nicht in P (bzw. $\text{co-}NP$) enthalten ist, außer wenn PH auf P (bzw. NP) kollabiert. In den Übungen werden wir sehen, dass unter der Voraussetzung $PH \neq \Sigma_2^p$ keine NP -vollständige Sprache in PSK enthalten ist. Allgemeiner liefert die Polynomialzeithierarchie eine Folge von stärker werdenden Hypothesen der Form $PH \neq \Sigma_k^p$ für $k = 0, 1, 2, \dots$, die für $k = 0$ mit $P \neq NP$ und für $k = 1$ mit $NP \neq \text{co-}NP$ äquivalent sind.

Als nächstes zeigen wir, dass BPP in der zweiten Stufe der Polynomialzeithierarchie enthalten ist.

Satz 93 (Lautemann 1983, Sipser 1983). *Für jede Klasse C , die unter majority-Reduktionen abgeschlossen ist, gilt*

$$BP \cdot C \subseteq R \cdot \forall^p \cdot C.$$

Beweis. Sei $A \in BP \cdot C$. Dann existieren ein Polynom $p(n) \geq 1$ und eine p -balancierte Sprache $B \in C$, so dass für alle x der Länge n gilt:

$$\Pr_{y \in_R \{0,1\}^{p(n)}} [A(x) \neq B(x\#y)] \leq 2^{-n}.$$

Setzen wir $B_x = \{y \in \{0,1\}^{p(n)} \mid x\#y \in B\}$, so enthält B_x für $x \in A$ mindestens $(1 - 2^{-n})2^{p(n)}$ Wörter und für $x \notin A$ höchstens $2^{p(n)-n}$ Wörter. Sei \oplus die bitweise XOR-Operation auf $\{0,1\}^n$, d.h.

$$x_1 \cdots x_n \oplus y_1 \cdots y_n = (x_1 \oplus y_1) \cdots (x_n \oplus y_n).$$

Für $u_1, \dots, u_k \in \{0,1\}^{p(n)}$ sei

$$B_x(u_1 \cdots u_k) = \{y \oplus u_i \mid y \in B_x \text{ und } 1 \leq i \leq k\}.$$

Wir zeigen für alle x mit $|x| = n \geq 2$ und $2^n > p(n)$ die Implikationen

$$x \in A \Rightarrow \exists^{\geq 1/2} u \in \{0,1\}^{p(n)^2} : B_x(u) = \{0,1\}^{p(n)}$$

$$x \notin A \Rightarrow \forall u \in \{0,1\}^{p(n)^2} : B_x(u) \neq \{0,1\}^{p(n)}$$

7 Die Polynomialzeithierarchie

Dies beweist $A \in R \cdot \forall^p \cdot C$, da die p^2 -balancierte Sprache

$$B' = \{x\#u \mid u \in \{0,1\}^{p(n)^2}, B_x(u) = \{0,1\}^{p(n)}\}$$

in $\forall^p \cdot C$ entscheidbar ist. Sei also $x \in A$ mit $|x| = n \geq 2$ und sei v ein beliebiger String der Länge $p(n)$. Da B_x mehr als $(1-2^{-n})2^{p(n)}$ Wörter enthält, ist v für zufällig gewählte Strings $u_1, \dots, u_{p(n)} \in_R \{0,1\}^{p(n)}$ mit Wahrscheinlichkeit

$$\Pr[v \notin B_x(u_1 \cdots u_{p(n)})] = \Pr[\forall i : v \oplus u_i \notin B_x] \leq 2^{-np(n)}$$

nicht in $B_x(u_1 \cdots u_{p(n)})$ enthalten. Daher gilt für einen zufällig gewählten String $u \in_R \{0,1\}^{p(n)^2}$,

$$\begin{aligned} \Pr[B_x(u) \neq \{0,1\}^{p(n)}] &= \Pr[\exists v \in \{0,1\}^{p(n)} : v \notin B_x(u)] \\ &\leq 2^{p(n)-np(n)} \leq 1/2. \end{aligned}$$

Für den Nachweis der 2. Implikation nehmen wir an, dass für ein x der Länge n mit $2^n > p(n)$ Wörter $u_1, \dots, u_{p(n)}$ existieren mit $B_x(u_1 \cdots u_{p(n)}) = \{0,1\}^{p(n)}$. Da die Mengen $B_x(u_i)$ die gleiche Größe wie B_x haben, muss dann

$$\|B_x\| \geq 2^{p(n)}/p(n) > 2^{p(n)-n}$$

sein und somit x zu A gehören. ■

Insbesondere liefert Satz 92 folgende Inklusionen, indem wir $C = P$ bzw. $C = \text{co-NP}$ setzen (beachte, dass $R \cdot \forall^p \cdot \text{co-NP} = R \cdot \text{co-NP} \subseteq \text{BP} \cdot \text{co-NP}$ gilt).

Korollar 94. $\text{BPP} \subseteq \Sigma_2^p \cap \Pi_2^p$ und $\text{BP} \cdot \text{co-NP} \subseteq R \cdot \text{co-NP} \subseteq \Sigma_2^p$.

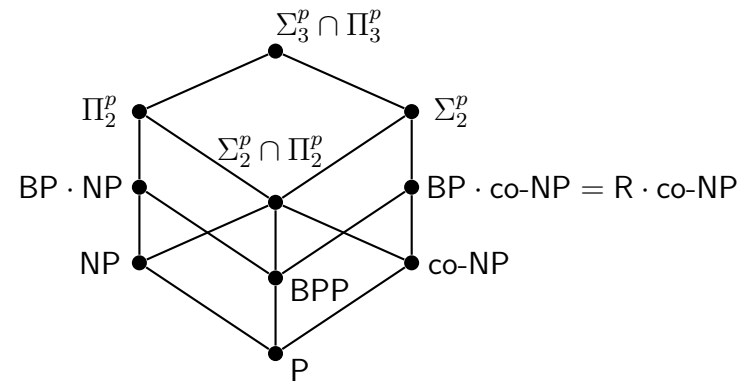
Eine weitere Folgerung ist die Gleichheit $\exists^p \cdot \forall^p \cdot \text{BP} \cdot C = \exists^p \cdot \forall^p \cdot C$ für alle Klassen C , die unter majority-Reduktionen abgeschlossen sind. Mit

C ist nämlich auch $\forall^p \cdot C$ unter majority-Reduktionen abgeschlossen und daher folgt mit den Sätzen 89 und 93

$$\exists^p \cdot \forall^p \cdot \text{BP} \cdot C \subseteq \exists^p \cdot \text{BP} \cdot \forall^p \cdot C \subseteq \exists^p \cdot R \cdot \forall^p \cdot \forall^p \cdot C \subseteq \exists^p \cdot \forall^p \cdot C.$$

Insbesondere gilt also $\exists^p \cdot \forall^p \cdot \text{BP} \cdot \text{co-NP} = \Sigma_2^p$ und somit ist NP nicht in $\text{BP} \cdot \text{co-NP}$ enthalten, außer wenn $\text{PH} = \Sigma_2^p = \text{BP} \cdot \text{NP}$ ist.

Korollar 95. $\text{NP} \subseteq \text{BP} \cdot \text{co-NP} \Rightarrow \text{PH} = \text{BP} \cdot \text{NP}$.



Zum Schluss dieses Kapitels fassen wir die bekannten Kollapskonsequenzen unter unterschiedliche Annahmen der Zugehörigkeit von NP -vollständigen Mengen A in bestimmten Komplexitätsklassen zusammen.

Annahme	Konsequenz
$A \in P$	$\text{PH} = P$ (Satz 92)
$A \in \text{co-NP}$	$\text{PH} = \text{NP}$ (Satz 92)
$A \in \text{BPP}$	$\text{PH} = \Sigma_2^p = \text{BPP}$ (siehe Übungen)
$A \in \text{BP} \cdot \text{co-NP}$	$\text{PH} = \Sigma_2^p = \text{BP} \cdot \text{NP}$ (Korollar 95)
$A \in \text{PSK}$	$\text{PH} = \Sigma_2^p = \text{ZPP}^{\text{NP}}$ (siehe Übungen)

8 Turing-Operatoren

In diesem Kapitel betrachten wir Berechnungen, die Zugriff auf eine Orakelsprache A haben, d.h., die Information, ob bestimmte Wörter in A enthalten sind oder nicht, kann durch eine Orakelanfrage, deren Beantwortung nur einen Rechenschritt kostet, abgerufen werden. Auf diese Weise erhalten wir zu jeder Komplexitätsklasse C eine relativierte Version C^A , in der alle Probleme enthalten sind, die relativ zum Orakel A innerhalb der durch C vorgegebenen Ressourcen lösbar sind.

Definition 96. Eine **deterministische Orakelmaschine (DOM)** M ist eine DTM, die mit einem speziellen write-only **Orakelfrageband** ausgerüstet ist. Außerdem besitzt M drei spezielle Zustände $q_?, q_+, q_-$. Als Orakel kann eine beliebige Sprache $A \subseteq \Sigma^*$ verwendet werden. Geht M in den **Fragezustand** $q_?$, so hängt der Folgezustand q' davon ab, ob das aktuell auf dem Orakelband stehende Wort y zu A gehört (in diesem Fall ist $q' = q_+$) oder nicht ($q' = q_-$). In beiden Fällen wird das Orakelband gelöscht und der Kopf an den Anfang zurückgesetzt. All dies geschieht innerhalb eines einzigen Rechenschrittes. Die unter einem Orakel A arbeitende DOM wird mit M^A bezeichnet und die von M^A **akzeptierte Sprache** ist $L(M^A)$.

Anstelle eines Sprachorakels A benutzen wir zuweilen auch ein funktionales Orakel f . In diesem Fall ist M mit einem speziellen **Orakelantwortband** ausgestattet, auf dem jede Orakelfrage y innerhalb eines Rechenschrittes mit $f(y)$ beantwortet wird.

Wir sagen, M ist **nichtadaptiv**, falls die Fragen von M nicht von den Antworten auf zuvor gestellte Fragen abhängen.

Wir verlangen von einer Orakel-Turingmaschine, dass sie vorgegebene Ressourcenschranken unabhängig vom benutzten Orakel einhält.

Definition 97. Die **Rechenzeit** einer DOM M bei Eingabe $x \in \Sigma^*$ ist

$$time_M(x) = \sup_{A \subseteq \Sigma^*} time_{M^A}(x).$$

M ist $t(n)$ -**zeitbeschränkt**, falls für alle Eingaben x gilt:

$$time_M(x) \leq t(|x|).$$

Wir fassen alle Sprachen, die von einer DOM M mit Orakel A in Zeit $t(n)$ entscheidbar sind, in der relativierten Komplexitätsklasse

$$DTIME^A(t(n)) = \{L(M^A) \mid M \text{ ist eine } t(n)\text{-zeitbeschränkte DOM}\}$$

zusammen. $DTIME^A(t(n))$ wird auch als die **Relativierung** der Klasse $DTIME(t(n))$ zum Orakel A bezeichnet.

Beispielsweise enthält die Klasse

$$P^A = \{L(M^A) \mid M \text{ ist eine PDOM}\}$$

alle von einer polynomiell zeitbeschränkten DOM (kurz PDOM) M akzeptierten Sprachen. Für eine Sprachklasse C sei

$$P^C = \bigcup_{A \in C} P^A.$$

Für P^C (bzw. P^A) schreiben wir auch $P(C)$ (bzw. $P(A)$). Ebenso wie DTMs lassen sich auch NTMs und PMs oder Transducer mit einem Orakelmechanismus ausstatten, wodurch wir NOMs und POMs oder Orakeltransducer erhalten. Ist die Rechenzeit dieser Maschinen polynomiell beschränkt, so bezeichnen wir sie als NPOMs bzw. PPOMs. Entsprechend erhalten wir dann die relativierten Komplexitätsklassen NP^A , PP^A , FP^A , BPP^A , RP^A , ZPP^A usw. Lassen wir nur nichtadaptive Orakelmaschinen zu, so notieren wir dies durch den Index \parallel und schreiben P_{\parallel}^A , FP_{\parallel}^A usw. Falls wir dagegen die Anzahl der Fragen durch eine Funktion $g(n)$ begrenzen, wobei n die Eingabelänge bezeichnet, so schreiben wir $P^{A[g(n)]}$ usw.

Satz 98.

- (i) $P^P = P$ und $NP^P = NP$,
- (ii) $P^{NP \cap co-NP} = NP \cap co-NP$ und $NP^{NP \cap co-NP} = NP$,
- (iii) $NP^{NP} = \Sigma_2^P$ und $NP^{\Sigma_k^P} = \Sigma_{k+1}^P$ für $k \geq 0$.

Beweis. (i) Die Inklusion $P \subseteq P(P)$ ist klar. Für die umgekehrte Richtung sei L eine Sprache in $P(P)$. Dann existiert eine PDOM M und ein Orakel $A \in P$ mit $L(M^A) = L$. Sei M' eine PDTM mit $L(M') = A$. Betrachte die DTM $M''(x)$, die $M(x)$ simuliert und jedesmal, wenn M eine Orakelfrage y stellt, $M'(y)$ simuliert, um die Zugehörigkeit von y zu A zu entscheiden. Dann gilt

$$L(M'') = L(M^A) = L$$

und da die Beantwortung einer Orakelfrage höchstens Zeit

$$\max_{y, |y| \leq \text{time}_M(x)} \text{time}_{M'}(y) = |x|^{O(1)}$$

erfordert, ist M'' polynomiell zeitbeschränkt. Die Gleichheit von NP^P und NP lässt sich vollkommen analog zeigen.

- (ii) Wir zeigen zuerst die Inklusion $NP^{NP \cap co-NP} \subseteq NP$. Sei $L = L(M^A)$ für eine NPOM M und sei A ein Orakel in $NP \cap co-NP$. Dann existieren NPTMs M' und M'' mit $L(M') = A$ und $L(M'') = \bar{A}$. Betrachte folgende NPTM M^* :

$M^*(x)$ simuliert $M(x)$ und jedesmal wenn M eine Orakelfrage y stellt, entscheidet sich M^* nichtdeterministisch dafür, entweder $M'(y)$ oder $M''(y)$ zu simulieren. Falls $M'(y)$ (bzw. $M''(y)$) akzeptiert, führt M^* die Simulation von M im Zustand q_+ (bzw. q_-) fort. Anderfalls bricht M^* die Simulation von M ab und verwirft x .

Nun gilt $L(M^*) = L(M^A) = L$ und daher ist $L \in NP$. Dies zeigt $NP^{NP \cap co-NP} \subseteq NP$. Da $P^{NP \cap co-NP}$ unter Komplementbildung

abgeschlossen ist, folgt auch sofort $P^{NP \cap co-NP} \subseteq NP \cap co-NP$. Die umgekehrten Inklusionen sind trivial.

- (iii) Wir zeigen zuerst die Inklusion von Σ_2^P in NP^{NP} . Zu jeder Sprache $L \in \Sigma_2^P = \exists^P \cdot co-NP$ existieren eine Sprache $A \in co-NP$ und ein Polynom p mit

$$x \in L \Leftrightarrow \exists y \in \{0, 1\}^{p(|x|)} : x \# y \in A.$$

Dann ist $\bar{A} \in NP$ und L wird von der NPOM M relativ zum Orakel \bar{A} akzeptiert, die bei Eingabe x ein Wort $y \in \{0, 1\}^{p(|x|)}$ rät und bei negativer Antwort auf die Orakelfrage $x \# y$ akzeptiert. Mit demselben Argument folgt auch Σ_k^P in $NP^{\Sigma_{k-1}^P}$. Der einzige Unterschied ist, dass nun $A \in \Pi_{k-1}^P$ bzw. $\bar{A} \in \Sigma_{k-1}^P$ und folglich $L = L(M^{\bar{A}}) \in NP^{\Sigma_{k-1}^P}$ ist.

Für die umgekehrte Richtung sei $L = L(M^A)$ für ein NP-Orakel A und eine NPOM M , deren Rechenzeit durch ein Polynom p und deren Verzweigungsgrad durch 2 beschränkt ist. Dann existieren ein Polynom q und eine Sprache $B \in P$ mit

$$y \in A \Leftrightarrow \exists z \in \{0, 1\}^{q(|y|)} : y \# z \in B.$$

Nun können wir jede Rechnung α von $M(x)$ durch ein Wort $r = r_1 \cdots r_{p(|x|)} \in \{0, 1\}^{p(|x|)}$ kodieren, wobei r_i im Fall, dass α im i -ten Rechenschritt nichtdeterministisch verzweigt, die Richtung, und im Fall, dass α im i -ten Rechenschritt eine Orakelfrage stellt, die Antwort angibt. Nun gilt

$$x \in L \Leftrightarrow \exists r \in \{0, 1\}^{p(|x|)} \exists z_1, \dots, z_{p(|x|)} \in \{0, 1\}^{q(p(|x|))} : x \# r z_1, \dots, z_{p(|x|)} \in C,$$

wobei C alle Strings $x \# r z_1, \dots, z_{p(|x|)}$ enthält mit

- r kodiert eine akzeptierende Rechnung von $M(x)$, während der m Orakelfragen y_1, \dots, y_m gestellt und mit a_1, \dots, a_m beantwortet werden, und
- für $i = 1, \dots, m$ gilt $(a_i = 1 \wedge y_i \# (z_i)_{\leq q(|y_i|)} \in B) \vee (a_i = 0 \wedge y_i \notin A)$,

8 Turing-Operatoren

wobei $(z)_{\leq k}$ das Präfix der Länge k von z bezeichnet. Da C in co-NP liegt, zeigt diese Charakterisierung, dass L in $\exists^p \cdot \text{co-NP}$ enthalten ist.

Mit diesem Argument lässt sich auch die Inklusion $\text{NP}^{\Sigma_k^p} \subseteq \Sigma_{k+1}^p$ zeigen. Der einzige Unterschied ist, dass nun A in Σ_k^p (bzw. A in Π_k^p) und B in Π_{k-1}^p liegen. Daher ist leicht zu sehen, dass C nun in Π_k^p liegt und somit $L = L(M^A) = \exists^p C$ in $\exists^p \cdot \Pi_k^p = \Sigma_{k+1}^p$ enthalten ist. ■

Der vorige Satz liefert folgende Charakterisierung für die Stufen der Polynomialzeithierarchie:

$$\Sigma_k^p = \text{NP}^{\Sigma_{k-1}^p} = \text{NP}^{\text{NP}^{\dots^{\text{NP}}}},$$

$$\Pi_k^p = \text{co-NP}^{\Sigma_{k-1}^p} = \text{co-NP}^{\text{NP}^{\dots^{\text{NP}}}},$$

wobei die Türme die Höhe k haben. Zudem erhalten wir weitere Stufen Δ_k^p :

$$\Delta_k^p = \text{P}^{\Sigma_{k-1}^p} = \text{P}^{\text{NP}^{\dots^{\text{NP}}}},$$

Offensichtlich gilt $\Delta_k^p \subseteq \Sigma_k^p \cap \Pi_k^p \subseteq \Sigma_k^p \cup \Pi_k^p \subseteq \Delta_{k+1}^p$:

