

Komplexität von Suchproblemen und Beweissystemen

Anton Ehrmanntraut

6. November 2023

Inhaltsverzeichnis

1	Einleitung	2
2	Grundlagen	3
2.1	Notation	3
2.2	Maschinenmodell	4
2.3	Komplexitätsklassen	5
2.4	Orakel und Relativierungen	7
2.5	Beweissysteme	8
3	Zur Konzeptionalisierung und Ordnung von Suchproblemen	10
3.1	Definition von Suchproblemen	10
3.2	Suchprobleme vs. Entscheidungsprobleme	13
3.3	Levin-Reduzierbarkeit	16
3.4	Zur gemeinsamen Struktur von vollständigen Suchproblemen	19
4	Suchprobleme und die Hypothese Q im Kontext des Pudlák'schen Programms	24
4.1	Welche Suchprobleme sind paddable?	26
4.2	Hypothese Q und die Vollständigkeit von Suchproblemen	28
4.3	Karp-Vollständigkeit vs. Levin-Vollständigkeit	30
4.4	Bekannte Implikationen, Offene Orakel	32
5	Orakel	35

1 Einleitung

Viele andere Beweise der Komplexitätstheorie funktionieren auch in relativierten Umgebungen, also wenn alle beteiligten TM zu OTM ausgewechselt werden. Das Diagonalargument in einem typischen Beweis von $P \subsetneq E$ relativiert beispielsweise, sodass auch $P^O \subsetneq E^O$ für jedes beliebige Orakel O gilt. Wir sagen dann auch, dass diese Aussagen bzw. Beweise *relativieren*.

Die typischen Beweistechniken der Komplexitätstheorie relativieren. Das macht eigens konstruierte Orakel zu Indizien, dass gewisse Aussagen schwer zu beweisen sind. Beispielsweise konstruieren Baker, Gill und Solovay (1975) ein Orakel A sodass $P^A \neq NP^A$. Mit diesem Fakt ist die Aussage „ $P = NP$ “ nicht mit relativierbaren Methoden beweisbar, da sonst ja auch $P^A = NP^A$ gelten würde.

Die algorithmische Komplexitätstheorie fokussiert sich primär auf Entscheidungsprobleme, und weniger auf Suchprobleme. Dabei sind Suchprobleme mindestens genauso natürlich wie Entscheidungsprobleme, und in einigen Fällen auch (vor allem für Nicht-Fachleute) intuitiver: wir sind nicht daran interessiert *ob* eine Dreifärbung eines Graphen existiert, sondern sind an der konkreten Färbung interessiert. Diese Fokussierung auf Entscheidungsprobleme ist durchaus fundiert: es lässt sich argumentieren, dass für viele relevante Instanzen das Suchproblem auf das entsprechende Entscheidungsproblem „reduziert“ werden kann, womit aus algorithmischer Perspektive das Suchproblem nicht schwerer ist als das Entscheidungsproblem. Die Konzentration auf Suchprobleme kommt dann unter anderem auch mit dem Vorteil, dass viele theoretische Konzepte einfacher zu fassen sind und kompakter zu formulieren sind; so z.B. der Begriff von *Sprache* und Turing-Maschinen die als „Ausgabe“ nur akzeptieren bzw. ablehnen können. Das hat auch historische Gründe: Entscheidungsprobleme waren der primäre Fokus innerhalb der Berechenbarkeitstheorie einerseits und in der Theorie der formalen Sprachen und Grammatiken andererseits.

2 Grundlagen

Dieses Kapitel legt die definitorischen Grundlagen für die folgenden Kapitel fest. In Abschnitt 2.1 legen wir mathematische Notationen für diese Arbeit fest. Abschnitt 2.2 spezifiziert das Maschinenmodell. Abschnitt 2.3 wiederholt einige Standarddefinitionen aus der Komplexitätstheorie. Abschnitt 2.4 setzt das hier verwendete Verständnis von Relativierungen fest. Abschließend geht Abschnitt 2.5 kurz auf Beweissysteme im Sinne von Cook und Reckhow (1979) ein.

2.1 Notation

Sei Σ das standardmäßige Alphabet mit $\Sigma = \{0, 1\}$. Elemente von Σ^* nennen wir Wörter, sind also endliche Sequenzen von Zeichen aus Σ . Teilmengen von Σ^* nennen wir auch Sprachen. Wir bezeichnen die Länge eines Wortes $w \in \Sigma^*$ mit $|w|$. Das leere Wort bezeichnen wir mit ε . Das i -te Zeichen eines Wortes w für $0 \leq i < |w|$ identifizieren wir mit $w[i]$. Diese Notation erweitern wir auf Sequenzen von Indizes: für $0 \leq i_1, i_2, \dots, i_k < |w|$ und $\alpha = (i_1, i_2, \dots, i_k)$ sei $w[\alpha] = w[i_1]w[i_2] \dots w[i_k]$. Insbesondere ist damit $w[0, 1, 2, \dots, |w| - 1] = w$. Falls w ein (echter) Präfix von v ist dann schreiben wir $w \sqsubseteq v$ (bzw. $w \sqsubset v$).

Die Menge aller natürlichen (nicht-negativen) Zahlen wird mit \mathbb{N} bezeichnet. Die leere Menge notieren wir wie üblich als \emptyset . Die Kardinalität einer Menge A notieren wir wie üblich als $|A|$. Für eine Menge $A \subseteq \Sigma^*$ und $n \in \mathbb{N}$ definieren wir $A^{\leq n} = \{w \in A \mid |w| \leq n\}$. Analog definieren wir $A^{< n}, A^{=n}$, usw. Außerdem bezeichnet $\ell(A) = \sum_{w \in A} |w|$. Für solche Teilmengen A von Σ^* verstehen wir das Komplement \bar{A} als $A - \Sigma^*$.

Zweistellige bzw. binäre Relationen $R \subseteq A \times B$ können wir mit den üblichen Eigenschaften beschreiben: die Relation R ist

- *(links-)total* wenn jedes Element aus A mit mindestens einem Element aus B reliert,
- *rechtstotal* bzw. *surjektiv* wenn jedes Element aus B mit mindestens einem Element aus A reliert,
- *linkseindeutig* bzw. *injektiv* wenn jedes Element aus B mit höchstens einem Element aus A reliert,
- *(rechts-)eindeutig* bzw. *funktional* wenn jedes Element aus A mit höchstens einem Element aus B reliert,
- *bijektiv* wenn jedes Element aus A mit genau einem Element aus B reliert und umgekehrt, also genau dann wenn R funktional, surjektiv und injektiv ist.

Binäre Relationen nennen wir eine (partielle) *Funktion* wenn diese Relation funktional ist. Eine Funktion sei also im Folgenden im Allgemeinen nicht total. Sollte (Links-)Totalität explizit gefordert sein, sprechen wir von *totalen Funktionen*. Binäre Relationen über Wörtern aus Σ^* , welche nicht unbedingt Funktionen sind, verstehen wir manchmal auch aus historischen Gründen als (partielle) Multifunktionen, dem Begriff der „*partial multivalued function*“ nachempfunden.

Für eine binäre Relation $R \subseteq \Sigma^* \times \Sigma^*$ schreiben wir $\text{Proj}(R)$ für die Menge $\{x \mid (x, y) \in R\}$. Für ein Wort $x \in \Sigma^*$ schreiben wir $\text{set-}R(x) = \{y \mid (x, y) \in R\}$ für die Bildmenge von x auf R . Manchmal werden wir binäre Relationen auch über die Spezifikation der jeweiligen Bildmengen definieren, also z.B. $\text{set-}Q(n) = \{0, 1, \dots, n\}$ schreiben um die Relation $Q = \{(a, b) \mid b \leq a\}$ zu definieren. Falls f eine Funktion bzw. funktional ist, meinen wir mit $f(x)$ wie üblich das *Bildelement* der Funktion f meinen, und nicht die *Bildmenge*.

Für eine Funktion f bezeichnen wir die Urbild- bzw. Bildmenge (domain und range) mit $\text{dom}(f)$ und $\text{ran}(f)$. (Beachte dass $\text{Proj}(f) = \text{dom}(f)$. Wir führen diese Unterscheidung nur wegen den Gewohnheiten dieser zwei Notationen ein.) Ist f eine injektive Funktion, dann bezeichnen wir mit f^{-1} dessen Umkehrfunktion. Beobachte dass f^{-1} funktional ist, ist ja f injektiv. Ist f sogar bijektiv, dann ist die Umkehrfunktion f^{-1} eine totale Funktion.

Eine Funktion $f: \Sigma^* \rightarrow \Sigma^*$ nennen wir *verlängernd* wenn $|f(x)| \geq |x|$ für alle $x \in \text{dom}(f)$. Die Funktion f nennen wir *polynomiell längenbeschränkt* wenn ein Polynom p existiert sodass $|x| \leq p(|x|)$ für alle $x \in \text{dom}(f)$. Die Funktion f nennen wir *ehrllich* wenn ein Polynom q existiert sodass $q(|f(x)|) \geq |x|$ für alle $x \in \text{dom}(f)$.

Beachte dass Funktionen nur spezielle Relationen sind. Wenn also f eine Funktion ist, meinen wir mit „ $f \in P$ “ dass *der Graph* von f in Polynomialzeit entschieden werden kann (i.e., gegeben Tupel (x, y) , gilt $f(x) = y$?). Das ist eine schwächere Aussage als „ $f \in \text{FP}$ “ die wie in üblicher Interpretation besagen soll, dass aus x das Bild $f(x)$ in Polynomialzeit berechnet werden kann.

Im Folgenden definieren wir noch den Begriff der *Verfeinerung*. Seien F, G zwei Multifunktionen. Wir nennen G eine *Verfeinerung* von F wenn $\text{Proj}(F) = \text{Proj}(G)$ und $\text{set-}G(x) \subseteq \text{set-}F(x)$ für alle $x \in \text{Proj}(F)$ (bzw. äquivalent $\in \text{Proj}(G)$). Ist F eine Multifunktion, und \mathcal{G} eine Klasse von Multifunktionen, schreiben wir $F \in_c \mathcal{G}$ wenn \mathcal{G} eine Verfeinerung $G \in \mathcal{G}$ von F enthält. Für zwei Klassen \mathcal{F}, \mathcal{G} von Multifunktionen schreiben wir $\mathcal{F} \subseteq_c \mathcal{G}$ falls für jede Multifunktion $F \in \mathcal{F}$ auch $F \in_c \mathcal{G}$ gilt.

Die endlichen Wörter Σ^* können über ihre quasi-lexikographischen Ordnung \prec_{lex} linear geordnet werden. Diese ist eindeutig definiert indem wir $0 \preceq_{\text{lex}} 1$ fordern. Unter dieser Definition existiert ein Ordnungsisomorphismus zwischen $(\Sigma^*, \preceq_{\text{lex}})$ und $(\mathbb{N}, <)$, welcher insbesondere eine Bijektion zwischen Σ^* und \mathbb{N} induziert, der sowohl in Polynomialzeit berechenbar als auch invertierbar ist. (Eine solcher Isomorphismus wird zum Beispiel durch eine dyadische Codierung realisiert.) Durch diese Identifikation können wir Wörter aus Σ^* als Zahlen aus \mathbb{N} behandeln und umgekehrt. Es können also auch Notationen, Beziehungen und Operationen für Σ^* auf \mathbb{N} übertragen werden und umgekehrt. Insbesondere können wir dann von einer Länge $|n|$ des Wortes sprechen, welches von $n \in \mathbb{N}$ repräsentiert wird. Insbesondere meint dieser Ausdruck nicht den Betrag von n . Ebenso bezeichnet die Ordnung \leq sowohl die Kleiner-oder-gleich-Ordnung auf den natürlichen Zahlen als auch der quasi-lexikographischen Ordnung \prec_{lex} auf den endlichen Wörtern. Diese Übereinstimmung ist nach den Eigenschaften des Ordnungsisomorphismus auch kompatibel mit der Identifikation von Wörtern mit Zahlen. Beachte dass der Längenoperator $|\cdot|$ ordnungserhaltend ist: wenn $a \leq b$ (oder eben äquivalent $a \preceq_{\text{lex}} b$) für zwei Wörter $a, b \in \Sigma^*$ dann ist auch $|a| \leq |b|$, bzw. ist das Wort a höchstens so lang wie das Wort b . Mit den Ausdrücken 0^n und 1^n meinen wir immer die zwei Wörter $000\cdots$ und $111\cdots$ aus Σ^n .

Wir definieren mit $\langle \cdots \rangle$ eine Paarungsfunktion von $\bigcup_{i \geq 0} (\Sigma^*)^i \rightarrow \Sigma^*$, welche injektiv und in Polynomialzeit sowohl berechenbar als auch invertierbar ist, und die im folgenden Sinne längeneffizient ist: $|\langle u_1, \dots, u_n \rangle| = 2(|u_1| + \cdots + |u_n| + n)$. Eine solche Paarungsfunktion kann beispielsweise über $\langle u_1, \dots, u_n \rangle \mapsto f(\#u_1\# \cdots \#u_n)$ realisiert werden, wobei f eine Codierung vom Alphabet $\{0, 1, \#\}$ auf Σ^* mittels $\{0 \mapsto 00, 1 \mapsto 11, \# \mapsto 01\}$ ist. Diese Paarungsfunktion werden wir häufig verwenden, um Tupel an Wörtern zu codieren, z.B. damit eine Turing-Maschine ein Tupel an Wörtern als Eingabe entgegen nehmen kann. Auf die konkrete Angabe dieser Paarungsfunktion wird aber im Folgenden meist verzichtet und sie wird nur implizit mitgedacht. So meinen wir mit dem Tupel (a, b) für $a, b \in \Sigma^*$ je nach Kontext entweder mathematisch präzise das Element aus dem Produkt $\Sigma^* \times \Sigma^*$, oder das Wort $\langle a, b \rangle \in \Sigma^*$. Ebenso verstehen wir je nach Kontext eine binäre Relation $R \subseteq \Sigma^* \times \Sigma^*$ auch als eine Sprache im Sinne einer Teilmenge von Σ^* , die bspw. von einer Turing-Maschine entschieden werden kann. Algorithmen und Turing-Maschinen verarbeiten nicht nur Wörter, sondern auch andere Objekte wie z.B. Graphen oder Turing-Maschinen. Daher werden wir die obige implizit mitgedachte Codierung auch auf andere Objekte ausweiten. Hierbei seien die jeweiligen Codierungen angemessen effizient, in dem Sinne dass die Codierungen kompakt sind und entsprechende Operationen auf den codierten Objekten in Polynomialzeit zulassen. Zum Beispiel lässt sich ein Graph mit Knotenmenge V und Kantenmenge E in polynomieller Länge abh. von $|V|$ und $|E|$ codieren, und auf der entsprechenden Codierung kann z.B. die Nachbarschaft eines ausgezeichneten Knotens ebenso in Polynomialzeit aufgezählt werden.

2.2 Maschinenmodell

Diese Arbeit baut auf dem Berechnungsmodell der Turing-Maschine (TM) auf. Wir betrachten hierbei sowohl die deterministische als auch die nichtdeterministische Variante. In dieser Arbeit haben TM sowohl ausgezeichnete Zustände zum Akzeptieren bzw. Ablehnen, ein Eingabeband, ein Arbeitsband, und ein Ausgabeband. Im Folgenden betrachten wir nur TM die immer terminieren. (Es ist einer TM im Allgemeinen nicht ansehbar, ob diese immer terminiert. Im Verlauf dieser Arbeit werden die TM aber so beschaffen sein, dass diese offensichtlich immer terminieren.)

Wir betrachten zunächst deterministische TM. Sei M eine deterministische TM, und x eine Eingabe. Dann induziert eine Berechnung $M(x)$ einen Rechenweg α , der in einem ausgezeichnetem Zustand q terminiert. Wir sagen dann auch, dass α der *Rechenweg von Berechnung* $M(x)$ ist. Wenn der terminierende Zustand q dieses Rechenwes α ein akzeptierender Zustand ist, dann sagen wir auch dass $M(x)$ mit Ausgabe y akzeptiert oder kurz $M(x)$ akzeptiert wobei y jenes Wort ist, welches auf dem Ausgabeband steht.

Eine solche deterministische TM M setzt nun gleichzeitig zwei unterschiedliche Funktions-

weisen um. Einerseits die eines Akzeptors einer Menge, und andererseits die einer Funktion:

- Die von M *entschiede Sprache* ist die Menge $L(M) = \{x \in \Sigma^* \mid M(x) \text{ akzeptiert}\}$.
- Die von M *berechnete Funktion* ist die Funktion $f_M: \Sigma^* \rightarrow \Sigma^*$ mit

$$f_M(x) = \begin{cases} y & \text{wenn } M(x) \text{ mit Ausgabe } y \text{ akzeptiert,} \\ \perp & \text{sonst.} \end{cases}$$

Wenn wir M im Kontext der zweiten Funktionsweise verstehen, dann sprechen wir auch von einem Turing-Transduktor. Wir kürzen dann auch „die von M berechnete Funktion“ durch „die Funktion M “ ab und verstehen den Turing-Transduktor M als genuine Funktion, und schreiben dann z.B. $M(x) = y$ anstelle $f_M(x) = y$.

Diese zwei Arten von Funktionsweisen einer TM erweitern wir nun auf nichtdeterministische TM. Sei N eine nichtdeterministische TM, und x eine Eingabe. Dann induziert analog eine Berechnung $N(x)$ nicht nur eine, sondern ggf. mehrere terminierende Rechenwege, die wir ebenso die Rechenwege von Berechnung $N(x)$ nennen. Terminiert ein solcher Rechenweg von $N(x)$ in einem akzeptierenden Zustand, nennen wir diesen Rechenweg auch einen *akzeptierenden Rechenweg*. Ähnlich wie im deterministischen Fall sagen wir dass $N(x)$ *auf Rechenweg α (mit Ausgabe y) akzeptiert* wenn α ein akzeptierender Rechenweg von $N(x)$ ist (und y auf dem Eingabeband steht). Beachte dass die Angabe eines Rechenwegs zwingend notwendig ist, da zu einer Berechnung $N(x)$ ja mehrere Rechenwege mit je unterschiedlichen Akzeptierverhalten und Ausgaben existieren. Im Sinne eines existentiellen Akzeptierverhaltens sagen wir dass $N(x)$ *akzeptiert* wenn *mindestens* ein akzeptierender Rechenweg α auf $N(x)$ existiert.

Analog ergeben sich nun wieder zwei Funktionsweisen, einerseits als Akzeptor, andererseits als Multifunktion:

- Die von N *entschiede Sprache* ist die Menge

$$L(N) = \{x \in \Sigma^* \mid N(x) \text{ akzeptiert}\} = \{x \in \Sigma^* \mid \text{ex. akz. Rechenweg auf } N(x)\}.$$

- Die von N *berechnete Multifunktion* ist die Multifunktion $f_N \subseteq \Sigma^* \times \Sigma^*$ mit

$$f_N(x) = \{y \mid N(x) \text{ akz. auf einem Rechenweg mit Ausgabe } y\}$$

Die berechnete Multifunktion kann in anderen Worten auch so verstanden werden, dass x den Ausgaben von $N(x)$ zugeordnet wird, wobei jeder akzeptierende Rechenweg eine Ausgabe macht, nämlich das Wort was auf dem Ausgabeband steht. Wie im deterministischen Fall können wir von nichtdeterministischen Turing-Transduktoren sprechen, wenn wir die zweite Funktionsweise betonen wollen. Ebenso können wir wieder abkürzend von „der Multifunktion N “ sprechen.

In sowohl dem deterministischen und nichtdeterministischen Fall können wir Berechnungen eine *Laufzeit* zuordnen: für eine TM M sei

$$\text{time}_M(x) = \max\{\text{Anz. Rechenschritte in } \alpha \mid \alpha \text{ ist ein Rechenweg von } M(x)\}.$$

Ist $\text{time}_M(x)$ durch ein Polynom in Abhängigkeit von $|x|$ beschränkt, und M eine deterministische (bzw. nichtdeterministische) TM, sagen wir auch dass M eine *deterministische* (bzw. *nichtdeterministische*) *Polyomialzeit-Turing-Maschine* (PTM bzw. NPTM) ist.

2.3 Komplexitätsklassen

Auf Basis der Turing-Maschinen als Berechnungsmodell können die üblichen Komplexitätsklassen der Entscheidungsprobleme bzw. Sprachen P, NP, coNP usw. definiert werden:

$$\begin{aligned} P &= \{L \subseteq \Sigma^* \mid \text{ex. PTM } M \text{ die } L \text{ entscheidet}\} \\ NP &= \{L \subseteq \Sigma^* \mid \text{ex. NPTM } M \text{ die } L \text{ entscheidet}\} \\ \text{coNP} &= \{L \subseteq \Sigma^* \mid \bar{L} \in NP\} \end{aligned}$$

Die Einfach- und Doppelt-Exponentialzeitklassen definieren wir wie folgt:

$$\begin{aligned} E &= \{L \subseteq \Sigma^* \mid \text{ex. TM } M \text{ die } L \text{ entscheidet, und ex. } c > 0 \text{ mit } \text{time}_M(x) \leq 2^{c|x|} \text{ für alle } x\} \\ EE &= \{L \subseteq \Sigma^* \mid \text{ex. TM } M \text{ die } L \text{ entscheidet, und ex. } c > 0 \text{ mit } \text{time}_M(x) \leq 2^{2^{c|x|}} \text{ für alle } x\} \end{aligned}$$

Die nichtdeterministischen Varianten NE, NEE und Komplementklassen coNE, coNEE sind analog definiert.

Die Funktioneklassen FP, NPMV, NPSV ist analog definiert (Selman 1994):

$$\begin{aligned} \text{FP} &= \{f : \Sigma^* \rightarrow \Sigma^* \mid f \text{ ist eine Funktion und ex. PTM-Transduktor } M \text{ der } f \text{ berechnet}\} \\ \text{NPSV} &= \{f : \Sigma^* \rightarrow \Sigma^* \mid f \text{ ist eine Funktion und ex. NPTM-Transduktor } M \text{ der } f \text{ berechnet}\} \\ \text{NPMV} &= \{f \subseteq \Sigma^* \times \Sigma^* \mid f \text{ ist eine Multifunktion und ex. NPTM-Transduktor } M \text{ der } f \text{ berechnet}\} \end{aligned}$$

Wir definieren NPMV_t als die Teilmenge von NPMV der Multifunktionen, die linkstotal sind. Analog NPSV_t . Ist für eine injektive Funktion $f \in \text{FP}$ auch die Umkehrfunktion $f^{-1} \in \text{FP}$ sagen wir auch, dass f *p-invertierbar* ist. Beachte, dass die Ehrlichkeit von f eine notwendige Bedingung für die p-Invertierbarkeit von f ist.

Wie üblich können wir mittels Reduktionen die Sprachen der Komplexitätsklassen ordnen. Seien A, B zwei Sprachen:

- $A \leq_T^p B$ wenn $A \in \text{FP}^B$ (Turing- bzw. Cook-Reduzierbarkeit).
- $A \leq_m^p B$ wenn eine Funktion $f \in \text{FP}$ existiert mit $x \in A \iff f(x) \in B$ (Many-one- bzw. Karp-Reduzierbarkeit).
- $A \leq_1^p B$ wenn eine injektive Funktion $f \in \text{FP}$ existiert mit $x \in A \iff f(x) \in B$ (One-one-Reduzierbarkeit).
- $A \leq_{1,i}^p B$ wenn eine injektive und p-invertierbare Funktion $f \in \text{FP}$ existiert mit $x \in A \iff f(x) \in B$.

Für die Funktioneklassen hat sich folgender sehr starke Begriff von Many-one-Reduzierbarkeit herausgebildet (Köbler und Messner 2000; Beyersdorff, Köbler und Messner 2009; Pudlák 2017). Seien g, h zwei Multifunktionen:

- $g \leq_m^p h$ wenn eine Funktion $f \in \text{FP}$ existiert mit $\text{set-}g(x) = \text{set-}h(f(x))$.

Jede dieser Ordnungsrelationen ist eine Quasiordnung, i.e. reflexiv und transitiv. Beachte, dass (auf Mengen) $\leq_{1,i}^p$ feiner als \leq_1^p ist, und diese feiner als \leq_m^p , und diese feiner als \leq_T^p ist. Beachte auch, dass P und NP auf \leq_m^p (und \leq_T^p) nach unten abgeschlossen sind:

$$\begin{aligned} A \leq_m^p B \text{ und } B \in \text{NP} &\implies A \in \text{NP} \\ A \leq_m^p B \text{ und } B \in \text{P} &\implies A \in \text{P} \end{aligned}$$

Sei \mathcal{C} eine Komplexitätsklasse und \preceq eine der obigen Reduktionsordnungen. Wie üblich nennen wir nun eine Sprache A *\preceq -hart für \mathcal{C}* wenn A eine obere Schranke für \mathcal{C} geordnet über \preceq ist (d.h. $B \preceq A$ für alle $B \in \mathcal{C}$). Wir nennen A *\preceq -vollständig für \mathcal{C}* wenn $A \in \mathcal{C}$ ein größtes Element von \mathcal{C} geordnet über \preceq ist (d.h. $B \preceq A$ für alle $B \in \mathcal{C}$ und $A \in \mathcal{C}$). Auf Grundlage der Existenz universeller effizienter Turing-Maschinen können für die Klassen P und NP jeweils eine kanonische \leq_m^p -vollständige Menge angegeben werden. Für NP ist diese

Definition 2.1.

$$\text{KAN} = \{(N, x, 1^n) \mid N \text{ ist eine NTM und akz. } x \text{ auf einem RW mit } \leq n \text{ vielen Schritten}\}.$$

Lemma 2.2. Die Menge KAN ist $\leq_{1,i}^p$ -vollständig.

Beweis. Sei $A \in \text{NP}$. Wir wollen zeigen dass $A \leq_{1,i}^p \text{KAN}$. Sei hierfür N eine NPTM welche A entscheidet. Es gibt also auch ein Polynom p welches die Laufzeit von N beschränkt. Definiere nun die Funktion $f(x) = (N, x, 1^{p(|x|)})$. Es gilt nun

$$\begin{aligned} x \in A &\iff N(x) \text{ akz. auf RW mit } \leq p(|x|) \text{ Schritten} \\ &\iff (N, x, 1^{p(|x|)}) \in \text{KAN} \iff f(x) \in \text{KAN}. \end{aligned}$$

Ferner ist leicht zu sehen, dass $f \in \text{FP}$, dass f injektiv und auch p-invertierbar ist. □

Die obigen Reduktionsordnungen erzeugen je eine kanonische Äquivalenzordnung („Duplikatrelation“):

- $A \equiv_m^p B \iff A \leq_m^p B \text{ und } B \leq_m^p A$.
- $A \equiv_1^p B \iff A \leq_1^p B \text{ und } B \leq_1^p A$.
- $A \equiv_{1,i}^p B \iff A \leq_{1,i}^p B \text{ und } B \leq_{1,i}^p A$.

Wir definieren auch noch die *p-Isomorphie* als eine Verfeinerung von $\equiv_{1,i}^p$:

- $A \equiv^p B \iff$ es existiert eine bijektive und p-invertierbare Funktion $f \in \text{FP}$ mit $x \in A \iff f(x) \in B$.

Gilt $A \equiv^p B$ dann sagen wir auch dass A und B *p-isomorph* sind. Im Folgenden werden noch die wichtigsten bekannten Aussagen bezüglich p-Isomorphie zusammengefasst:

Hartmanis und Berman (1976) zeigen dass aus $\equiv_{1,i}^p$ -äquivalente Sprachen dann p-isomorph sind, wenn die jeweiligen Reduktionsfunktionen verlängernd sind.

Satz 2.3 (Hartmanis und Berman 1976). *Gilt $A \leq_{1,i}^p B$ via f und $B \leq_{1,i}^p A$ via g , und f und g sind verlängernd, dann gilt $A \equiv^p B$.*

Um die Voraussetzungen vom vorigen Satz 2.3 zu vereinfachen, führen sie den Begriff der *paddability* ein.

Definition 2.4. Eine Sprache $A \neq \emptyset$ heißt (Berman–Hartmanis-) *paddable* genau dann wenn eine injektive und p-invertierbare Funktion $g \in FP$ existiert sodass für alle $x, y \in \Sigma^*$ gilt:

$$x \in A \iff g(x, y) \in A.$$

Das heißt, g fügt einen beliebigen String y zur „Problemistanz“ x hinzu, sodass die Mitgliedschaft zu A unverändert bleibt, und die beiden originalen Strings x und y wieder rekonstruiert werden können. \triangleleft

Es gilt:

Satz 2.5. (1) *Ist A paddable so gibt es für jedes B mit $B \leq_m^p A$ eine injektive p-invertierbare verlängernde Funktion die $B \leq_{1,i}^p$ realisiert.*

(2) *Sind A, B paddable, so folgt aus $A \equiv_m^p$ stets $A \equiv_p$.*

Alle bekannten \leq_m^p -vollständigen Mengen für NP sind zueinander p-isomorph. Berman und Hartmanis vermuteten, dass das für *alle* \leq_m^p -vollständigen Mengen gilt:

Vermutung 2.6 (IC). *Alle \leq_m^p -vollständigen Mengen für NP sind p-isomorph. In anderen Worten: die \leq_m^p -Äquivalenzklasse der vollständigen Mengen ist gleich der \equiv^p -Äquivalenzklasse von SAT.*

Mit obigem Begriff von Paddability lässt sich die \equiv^p -Äquivalenzklasse von SAT folgendermaßen charakterisieren:

Satz 2.7. *Eine Menge $A \in NP$ ist genau dann p-isomorph zu SAT wenn $A \leq_m^p$ -vollständig und paddable ist.*

Als Konsequenz ergibt sich hieraus, dass die *bekannten* \leq_m^p -vollständigen Mengen alle paddable sind. (Das ist die eigentliche empirische Beobachtung von Berman und Hartmanis, auf welcher diese IC vermuteten.)

2.4 Orakel und Relativierungen

Wie in der Einleitung schon angesprochen, ist die Idee hinter Orakel-Berechnungen zu untersuchen, welche Probleme B effizient(er) durch einen Algorithmus gelöst werden können, wenn die Algorithmen eine (fiktive) Möglichkeit haben, ein (ggf. sehr schweres) Problem A ohne Rechenaufwand zu lösen. Der Zugriff auf A kann also wie ein „Nachschlagewerk“ eine „Blackbox-Funktion“ verstanden werden, die auf magische Weise A augenblicklich löst.

Diese Idee wird im Berechnungsmodell der Orakel-Turing-Maschine (OTM) formalisiert. Orakel-Turing-Maschinen sind eine Erweiterung der (deterministischen und nichtdeterministischen) Turing-Maschinen, die zum Eingabe-, Arbeits- und Ausgabeband auch noch ein separates Orakelband haben. Ferner existieren drei ausgezeichnete Zustände $q_?$, q_{yes} , q_{no} .

Gegeben ein Orakel $A \subseteq \Sigma^*$ können OTM nun Fragen der Form $x \in A$ an das Orakel stellen, indem sie ein Wort x auf das Frageband schreiben, und in den Zustand $q_?$ übergeht. Im unmittelbar nächsten deterministischen Schritt der Berechnung wird der Zustand q_{yes} eingenommen falls $x \in A$, sonst den Zustand q_{no} .

Aus dieser Beschreibung wird klar, dass eine Berechnung einer OTM sowohl von der Eingabe x abhängig ist, als auch vom Orakel A , *relativ* zu diesem $M(x)$ rechnet. Wir schreiben dann auch kurz M^A wenn wir die OTM M mit festem Orakel A meinen, und $M^A(x)$ die Berechnung der OTM M auf Eingabe x mit Orakel A . Entsprechend können wir auch die Laufzeit $\text{time}_M^A(x)$ definieren, und von (deterministischen bzw. nichtdeterministischen) Polynomialzeit-Orakel-Turing-Maschinen (POTM, NPOTM) sprechen, wenn die Laufzeit auf allen Eingaben und allen Orakeln polynomiell durch die Eingabelänge beschränkt ist.

Wir können nun die relativierten Komplexitätsklassen P^O , NP^O , FP^O , $NPMV^O$, ... relativ zu einem gegebenen Orakel O definieren, wobei in der jeweiligen Definition die TM mit OTM ersetzt werden, die Zugriff auf das Orakel O haben. Diese Relativierung überträgt sich auch auf unsere weiteren Definitionen, wie z.B. Reduktion und Vollständigkeit. Wir schreiben z.B. $A \leq_m^{p,O} B$ wenn eine Funktion $f \in FP^O$ existiert mit $x \in A \leftrightarrow f(x) \in B$. Die kanonische

NP-vollständige Menge KAN kann ebenso zu KAN^O relativiert werden. Für *natürliche* Mengen wie SAT usw. werden wir dagegen keine relativierte Variante definieren. In diesem Sinne ist SAT im Allgemeinen *nicht* $\leq_m^{p,O}$ -vollständig, in dem Sinne dass ein Orakel O existiert und eine Menge $A \in \text{NP}^O$ sodass $A \not\leq_m^{p,O} \text{SAT}$.

In allgemeineren Beweisen, die nicht konkrete natürliche Mengen betreffen, lassen sich üblicherweise alle beteiligten TM mit OTM auswechseln, ohne die Gültigkeit der Aussage zu verändern. Aussagen bzw. Beweise, die in solchen relativierten Umgebungen relativ zu jedem beliebigen Orakel O gelten, nennen wir *relativierende* Aussagen bzw. Beweise. Das Diagonalargument in einem typischen Beweis von $P \subsetneq E$ relativiert beispielsweise, sodass auch $P^O \subsetneq E^O$ für jedes beliebige Orakel O gilt. Ebenso relativiert die Aussage „ $\text{KAN} \in \text{NP}$ ist \leq_m^p -vollständig“ (zu „ $\text{KAN}^O \in \text{NP}^O$ ist $\leq_m^{p,O}$ -vollständig“).

Im Folgenden soll jede Aussage als relativierbar verstanden werden, es sei denn es wird auf die Nichtrelativierbarkeit hingewiesen, oder von konkreten natürlichen Mengen gesprochen, welche ohnehin nicht relativieren.

2.5 Beweissysteme

Beweissysteme wurden in der Einleitung schon kurz definiert. In diesem Abschnitt wird die präzise Definition von Cook und Reckhow (1979) wiedergegeben.

Definition 2.8. Eine Funktion $f \in \text{FP}$ ist ein *Beweissystem* für L wenn $\text{ran}(f) = L$. Ist $f(w) = x$ schreiben wir auch, dass w ein f -Beweis für x ist.

Existiert zudem ein Polynom q sodass für jedes $x \in L$ ein f -Beweis x der Länge $\leq q(|x|)$ existiert, sagen wir dass f *kurzen Beweise* hat. \triangleleft

Hieraus stellt sich die erste Frage, welche Mengen Beweissysteme mit kurzen Beweisen haben.

Ein einfaches Beweissystem für die Menge $\text{SAT} \in \text{NP}$ wäre z.B. das *Standardbeweissystem* std für SAT:

$$std(\varphi, w) = \begin{cases} \varphi & \text{wenn } w \text{ eine erfüllende Belegung für } \varphi \text{ ist,} \\ \perp & \text{sonst.} \end{cases}$$

Dieses Beweissystem hat kurze Beweise.

Cook und Reckhow machen dagegen die Beobachtung im Fall von der Menge TAUT der aussagenlogischen Tautologien die Beobachtung, dass TAUT genau dann kein Beweissystem mit kurzen Beweisen hat, wenn $\text{NP} \neq \text{coNP}$. Diese Einsicht motivierte das sogenannte *Cook-Reckhow-Programm*: man nähert sich der Frage $\text{NP} \neq \text{coNP}$ mittels Untersuchung immer stärkerer Beweissysteme. Um nun die relative Stärke unterschiedlicher Beweissysteme zu vergleichen, führen Cook und Reckhow den Begriff der Simulation ein.

Definition 2.9. Seien f, g zwei Beweissysteme für L . Wir sagen dass f das Beweissystem g *simuliert* wenn eine (nicht notwendigerweise effiziente) polynomiell längenbeschränkte Funktion π existiert sodass

$$f(\pi(w)) = g(w).$$

Heißt, für jeden g -Beweis w für x existiert auch ein f -Beweis $\pi(w)$ für das gleiche x , und dieser f -Beweis $\pi(w)$ ist nur polynomiell länger als w .

Ist zusätzlich $\pi \in \text{FP}$, dann ist sagen wir, dass f das Beweissystem g *p-simuliert*. Das ist äquivalent zur Aussage $g \leq_m^p f$. \triangleleft

Wenn f das Beweissystem g p-simuliert, kürzen wir das entsprechend der Beobachtung in der Definition auch kurz mit $g \leq_m^p f$ ab.

Beobachte dass Beweissysteme mit kurzen Beweisen unter Simulation abgeschlossen sind: wenn Beweissystem f das Beweissystem g simuliert, und g kurze Beweise hat, dann hat auch f kurze Beweise. Auf ähnliche Beweise sind die Beweissysteme unter p-Simulation abgeschlossen, die einfach auffindbare Beweise haben: wenn Beweissystem f für L das Beweissystem g p-simuliert, und für jedes $x \in L$ in Polynomialzeit ein g -Beweis w gefunden werden kann, dann kann auch ein f -Beweis in Polynomialzeit gefunden werden.

Allgemein generiert die Relation der (p-)Simulation wieder eine Quasiordnung, die nach der Existenz von größten Elementen untersucht werden kann. Hieraus ergibt sich der Begriff der (p-)Optimalität.

Definition 2.10. Ein Beweissystem f für L ist *(p-)optimal* wenn es jedes Beweissystem g für L (p-)simulieren kann.

Die p-Optimalität von f ist äquivalent zur \leq_m^p -Vollständigkeit von f für die Teilmenge der Funktionen aus FP mit Bildmenge L . \triangleleft

Es ist leicht zu sehen, dass jedes Beweissystem mit kurzen Beweisen auch optimal ist. Im Zusammenhang mit dem Cook–Reckhow-Programm weisen Krajíček und Pudlák (1989) darauf hin, dass die Existenz eines optimalen Beweissystems für TAUT wahrscheinlich schwächer ist, als die Existenz eines Beweissystems mit kurzen Beweisen, denn Ersteres folgt schon aus $NE = coNE$. (Köbler, Messner und Torán, 2003, schwächen die Voraussetzung auf $NEE = coNEE$ ab.)

Für die Mengen aus P bzw. NP existieren p-optimale bzw. optimale Beweissysteme:

Beobachtung 2.11. (1) Ist $A \in P$, dann existiert ein p-optimales Beweissystem für A mit kurzen Beweisen.

(2) Ist $A \in NP$, dann existiert ein optimales Beweissystem für A mit kurzen Beweisen.

Beweis. 1. Zu (1): Betrachte die Funktion

$$h(x) = \begin{cases} x & \text{wenn } x \in A \\ \perp & \text{sonst} \end{cases}.$$

Diese Funktion ist definitiv ein Beweissystem für A . Sie ist in FP, ist ja der Test „ $x \in A$ “ in Polynomialzeit möglich. Klar ist auch dass h kurze Beweise hat. Dieses Beweissystem ist p-optimal, denn wenn g ein weiteres Beweissystem für A ist, und wenn w ein g -Beweis für x ist, dann ist $h(g(w)) = h(x) = x$, also $g(w)$ ein h -Beweis für x , wie gewünscht.

2. Zu (2): Kann durch die bekannte Zertifikats-Definition von NP gezeigt werden (was im späteren Teil der Arbeit auch geschieht), zur Vollständigkeit lässt sich an dieser Stelle aber auch ein Beweis über die NPTM-Definition von oben angeben. Sei N eine NPTM, die A mit polynomieller Laufzeit p entscheidet. Definiere nun

$$h(x, \alpha) = \begin{cases} x & N(x) \text{ akz. auf Rechenweg } \alpha \\ \perp & \text{sonst} \end{cases}$$

Nachdem $L(N) = A$ ist h definitiv ein Beweissystem für A . Es ist leicht zu sehen dass $h \in FP$, ist der Test „ α ist ein gültiger Rechenweg und akzeptiert“ in Polynomialzeit möglich. Ferner existiert für jedes $x \in A$ auch ein akzeptierender Rechenweg α auf $N(x)$ der Länge $\leq p(|x|)$, womit der Beweis (x, α) auch nur polynomiell länger als x ist. Da Beweissystem h hat also kurze Beweise, und ist damit auch optimal. \square

Insbesondere mit dem letzten Punkt können die optimalen Beweissysteme der Mengen aus NP auch als Beweissysteme mit kurzen Beweisen charakterisiert werden:

Beobachtung 2.12. Sei $A \in NP$ und f ein Beweissystem für A . Das Beweissystem f ist optimal genau dann wenn jedes $x \in A$ einen f -Beweis der Länge $\leq q(|x|)$ hat, wobei q ein Polynom ist.

Beweis. Richtung von rechts nach links klar, das gilt schon im Allgemeinen Fall. Für die andere Richtung sei f ein optimales Beweissystem für A . Dann muss f auch das Beweissystem h mit kurzen Beweisen aus voriger Beobachtung simulieren. Für jeden kurzen h -Beweis w für x existiert dann ein höchstens polynomiell längerer f -Beweis $\pi(w)$. \square

3 Zur Konzeptionalisierung und Ordnung von Suchproblemen

TODO: Einleitung und Übersicht über das Kapitel. Was passiert hier?

3.1 Definition von Suchproblemen

Wir geben hier noch einmal die Definition von Suchproblemen wieder, welche schon in der Einleitung erarbeitet wurde. Als Suchprobleme verstehen wir das algorithmische Problem, gegeben eine Probleminstanz x , eine entsprechende positive Lösungsinstanz y zu berechnen, oder negativ abzulehnen. Hier noch einmal ein Beispiel aus der Einleitung: gegeben eine aussagenlogische Formel φ , berechne entweder eine Belegung y welche φ erfüllt, oder gebe „unerfüllbar“ aus. Die wesentliche Einschränkung, welche wir auch schon in der Einleitung festgelegt haben, ist die Einschränkung auf „NP-Suchprobleme“. Wir meinen damit, dass

- die Lösungen nur polynomiell länger als die Probleminstanzen sind, und
- effizient in Polynomialzeit verifiziert werden kann, ob zu einer gegebenen Probleminstanz x ein beliebiges Wort y tatsächlich eine (positive) Lösung im Sinne des Suchproblems darstellt oder nicht.

(Wir fordern im Übrigen nicht, dass negatives Ablehnen effizient verifiziert werden kann.) Um das Beispiel wieder aufzugreifen: Zum einen haben Formeln φ , welche überhaupt erfüllbar sind, eine erfüllende Belegung in Länge von φ . Zum anderen kann effizient geprüft werden, ob y tatsächlich eine erfüllbare Belegung von φ ist.

Diese Einschränkung wird durch die empirische Einsicht gestützt, dass viele natürliche Suchprobleme, für die momentan kein effizienter Algorithmus bekannt ist, genau in eine solche Einschränkung fallen. Also Suchprobleme, die „verifizierbar“ sind und „kurze Lösungen“ haben. Einige weitere Beispiele werden wir im Folgenden noch betrachten.

Zunächst werden wir die beiden obigen Punkte noch einmal in eine formale Definition gießen:

Definition 3.1 (NP-Relation, FNP). Eine *NP-Relation* ist eine zweistellige Relation $R \subseteq \Sigma^* \times \Sigma^*$, sodass diese

- (1) in Polynomialzeit entscheidbar ist, d.h. $R \in P$, und
- (2) p -balanciert ist, d.h. es existiert ein Polynom q , sodass

$$(x, y) \in R \implies |y| \leq q(|x|) \quad \text{für alle } x, y \in \Sigma^*. \quad (3.1)$$

Die Wörter der ersten Komponente nennen wir *Probleminstanzen* oder *Instanzen* oder *Probleme* von R , die Wörter der zweiten Komponente nennen wir die *Zertifikate* (oder manchmal *Lösungen*) von R . Wir sagen dann für $(x, y) \in R$, dass y ein *Zertifikat* für x ist. In diesem Sinne sagt (3.1) aus, dass Zertifikate y für x nicht superpolynomiell länger als x sein dürfen. Das Polynom q nennen wir auch die *Zertifikatsschranke* zu R .

Wir schreiben FNP für die Klasse aller NP-Relationen. ◁

Das oben diskutierte Suchproblem zu einer NP-Relation R kann jetzt wie folgt formal formuliert werden:

Suchproblem zur Relation R :

Gegeben: Instanz x .

Gesucht: Zertifikat y mit $(x, y) \in R$ falls ein solches y überhaupt existiert, sonst „keine Lösung“ ausgeben.

Zur Erinnerung:

$$\text{Proj}(R) = \{x \mid \exists y \in \Sigma^*, (x, y) \in R\} \in \text{NP}.$$

Die Menge $\text{Proj}(R)$ ist also die Menge der Probleminstanzen, für welche ein zugehöriges Zertifikat existiert; damit entspricht $\text{Proj}(R)$ derjenigen Menge, die üblicherweise bei algorithmischen Entscheidungsproblemen betrachtet wird. Um die beiden Varianten noch einmal gegenüberzustellen: das entsprechende Entscheidungsproblem einer Relation R lautet

Entscheidungsproblem zur Relation R :

Gegeben: Instanz x .

Gesucht: Akzeptieren falls ein Zertifikat y mit $(x, y) \in R$ existiert, sonst ablehnen.

Das entspricht also dem Entscheiden der Sprache $\text{Proj}(R)$. Damit wird auch klar, dass das entsprechende Entscheidungsproblem bzw. die Sprache $\text{Proj}(R)$ nicht von der konkreten Relation R abhängig ist. Vielmehr: es existieren zur Sprache L ggf. unendlich viele NP-Relationen R mit $\text{Proj}(R) = L$. Für eine Sprache L sagen wir dann auch, dass R eine *NP-Relation für L* ist.

Die Zugehörigkeit des entsprechenden Suchproblems zu NP folgt hierbei unmittelbar aus der Definition von NP-Relationen. (Rate nichtdeterministisch ein Zertifikat und akzeptiere wenn dieses korrekt ist.) Im nächsten Abschnitt wird die Beziehung zwischen Suchproblemen bzw. NP-Relationen einerseits, und Entscheidungsproblemen bzw. Mengen aus NP andererseits, weiter behandelt. Festhalten können wir hier aber schon, dass das Suchproblem offenbar „schwieriger“ ist als das alleinige Entscheidungsproblem.

Im Folgenden werden einige Beispiele von natürlichen NP-Relationen angegeben. Um diese von den ansonsten üblicherweise verwendeten Labels für Mengen bzw. Suchprobleme abzugrenzen, sind im Verlauf dieser Arbeit *NP-Relationen* zu natürlichen Suchproblemen immer mit einem r am Anfang gekennzeichnet.

- $r\text{PERFECTMATCHING} = \{(G, M) \mid G \text{ ist ein Graph, } M \text{ ein perfektes Matching auf } G\}$.
- $r\text{SAT} = \{(\varphi, w) \mid \varphi \text{ ist eine aussagenlogische Formel, } w \text{ erfüllende Belegung für } \varphi\}$.
- $r\text{VC} = \{((G, k), C) \mid G \text{ ist ein Graph, } C \text{ eine Knotenüberdeckung, und } |C| \leq k\}$.
- $r\text{HAMCYCLE} = \{(G, P) \mid G \text{ ist ein Graph, } P \text{ ein Zyklus der jeden Knoten genau einmal berührt}\}$.
- $r\text{ANOTHERHAMCYCLE} = \{((G, P), P') \mid G \text{ ist ein Graph, } P, P' \text{ je ein Zyklus der jeden Knoten genau einmal berührt, } P \neq P'\}$.
- $r\text{FACTORIZATION} = \{(n, (p_1, p_2, \dots, p_k)) \mid n \in \mathbb{N}, n > 1, \text{ alle } p_i \text{ Primzahlen ungleich } 2 \text{ oder } n, \text{ und } n = p_1 \cdots p_k\}$.
- $r\text{FACTOR} = \{(n, p) \mid n \in \mathbb{N} \text{ ist nicht prim, und } p \text{ ist ein nichttrivialer Faktor von } n\}$.
- $r\text{SMALLFACTOR} = \{((n, a), p) \mid n \in \mathbb{N} \text{ ist nicht prim, und } p \text{ ist ein nichttrivialer Faktor von } n \text{ und } p \leq a\}$.
- $r\text{GI} = \{((G, H), \sigma) \mid G, H \text{ sind Graphen mit gleicher Knotenmenge, und } \sigma \text{ ist ein Graphisomorphismus von } G \text{ nach } H\}$.

Es ist leicht zu sehen, dass jede dieser Relationen auch eine NP-Relation ist. Beachte dass die Menge der Primzahlen in Polynomialzeit entscheidbar ist Agrawal, Kayal und Saxena (2004). Bei jeder der obigen natürlichen Relationen gilt, dass die Projektion auch der üblichen Sprache aus NP entspricht. Wir haben z.B.

$$\text{Proj}(r\text{VC}) = \{(G, k) \mid \text{ex. Knotenüberdeckung } C \text{ von Graph } G \text{ mit } |C| \leq k\}.$$

Die Definition Suchproblemen als NP-Relationen lässt es zu, Suchprobleme bzw. NP-Relationen als „partielle Multifunktionen“ zu verstehen. Selman (1994) definiert in seiner Taxonomie der Funktionsklassen die Klasse NPMV_g als die Klasse derjenigen Multifunktionen $f \in \text{NPMV}$, für die (der Graph) f in P liegt. Es lässt sich leicht sehen, dass die hier definierte Klasse FNP identisch zu Selman definierten Klasse NPMV_g ist, solange man Multifunktionen mit binäre Relationen identifiziert.

Mit dieser Perspektivierung ist auch einfach zu definieren, was mit „Suchproblem lösen“ gemeint ist. Wir machen hierbei Gebrauch von Verfeinerungen (von Multifunktionen). Wir sagen, dass das Suchproblem zur NP-Relation R in *Polynomialzeit lösbar ist*, wenn $R \in_c \text{FP}$. Diese Aussage bedeutet ja, dass eine Verfeinerung f von R existiert, und f ist dabei eine (partielle) Funktion. Für eine Eingabeinstanz x wird also entweder $f(x)$ einen Wert y ausgeben für den $y \in \text{set-}R(x)$ gilt, bzw. in anderen Worten, ein y für das $(x, y) \in R$ und damit ist die Ausgabe y eine Lösung für x . Oder, falls $f(x)$ ablehnt, dann ist $x \notin \text{dom}(f) = \text{Proj}(R)$, heißt „ $f(x)$ lehnt ab“ bedeutet dass x keine Lösung hat.

Wir können damit auch schon die obige intuitive Aussage beweisen, dass das Suchproblem „schwieriger“ ist als das entsprechende Entscheidungsproblem, in dem Sinne dass sich das Entscheidungsproblem auf das Suchproblem reduzieren lässt:

Beobachtung 3.2. Sei R eine NP-Relation. Falls $R \in_c \text{FP}$, dann gilt $\text{Proj}(R) \in \text{P}$.

Beweis. Sei $f \in \text{FP}$ die Verfeinerung von R nach Voraussetzung. Teste ob $f(x) \neq \perp$. Falls ja, dann ist $f(x) \in \text{set-}R(x)$ und damit hat x eine Lösung; akzeptiere. Falls nicht, dann ist $x \notin \text{dom}(f) = \text{Proj}(x)$; lehne ab. \square

Der aktuelle Stand zur Lösbarkeit der oben genannten natürlichen Suchprobleme ist:

- $\text{rPERFECTMATCHING} \in_c \text{FP}$.
- $\text{NP} = \text{P} \iff \text{rSAT} \in_c \text{FP} \iff \text{rVC} \in_c \text{FP} \iff \text{rHAMCYCLE} \in_c \text{FP} \iff \text{rANOTHERHAMCYCLE} \in_c \text{FP}$.
- Unklar ob $\text{rSMALLFACTOR}, \text{rFACTOR}, \text{rFACTORIZATION} \in_c^? \text{FP}$. Wir haben aber $\text{UP} \cap \text{coUP} = \text{P} \implies \text{rSMALLFACTOR} \in_c \text{FP} \iff \text{rFACTOR} \in_c \text{FP} \iff \text{rFACTORIZATION} \in_c \text{FP}$.
- Unklar ob $\text{rGI} \in_c^? \text{FP}$.

Bevor nun im nächsten Abschnitt die Suchprobleme den Entscheidungsproblemen näher gegenübergestellt werden, schließen wir diesen Abschnitt noch mit einer kurzen Diskussion zu *totalen* Suchproblemen ab.

Die oben formulierte Definition von FNP ist genau diejenige, die von Megiddo und Papadimitriou (1991) als erstes in dieser Form und Bezeichnung definiert wurde. Ihre Motivation war, hierbei insbesondere die *totalen* Suchprobleme in den Blick zu nehmen. Also solche Suchprobleme, bei der jede Proleminstanz immer mindestens ein Zertifikat bzw. Lösung hat. Die Faktorisierung ist beispielsweise ein solches totales Suchproblem, da ja jede natürliche Zahl sich faktorisieren lässt.

Das sind – entsprechend dieser Definition von FNP bzw. Konzeptionalisierung von Suchproblemen – genau jene NP-Relationen welche (links-)total sind: für jedes $x \in \Sigma^*$ existiert ein $y \in \Sigma^*$ mit $(x, y) \in R$. Die Relationen rFACTORIZATION und rFACTOR wie oben definiert sind nicht total, nachdem die negativen Instanzen aber besonders „einfach“ sind, können für beide NP-Relationen effektiv äquivalente Relationen angegeben werden, die total sind:

- $\text{rFACTORIZATION}' = \text{rFACTORIZATION} \cup \{(n, \text{„ungültig“}) \mid n \leq 1\}$.
- $\text{rFACTOR}' = \text{rFACTOR} \cup \{(n, \text{„ungültig“}) \mid n \leq 1 \text{ oder } n \text{ ist prim}\}$.

Megiddo und Papadimitriou (1991) fassen diese totalen NP-Relationen zur Klasse TFNP zusammen:

Definition 3.3 (TFNP). Die Klasse TFNP ist die Teilmenge von FNP derjenigen NP-Relationen R , welche linkstotal sind, heißt zu jedem $x \in \Sigma^*$ existiert ein $y \in \Sigma^*$ mit $(x, y) \in R$. \triangleleft

Hierzu gehören die oben genannten Varianten $\text{rFACTORIZATION}'$ und $\text{rFACTOR}'$. Für Megiddo und Papadimitriou befinden sich in TFNP eine Vielzahl von interessanten und schwierigen Suchproblemen, bei denen die Frage der Lösbarkeit in Polynomialzeit noch offen ist. Das betrifft u.a. zahlentheoretische Probleme aus der Kryptographie wie Faktorisierung, diskreter Logarithmus. Beachte dass TFNP nicht identisch ist zur Klasse NPMV_t ; es macht sich hier die gleiche Unterscheidung wie bei FNP vs. NPMV auf: Die Klasse TFNP ist eine Teilmenge von NPMV_t jener totalen Multifunktionen $f \in \text{NPMV}_t$, für die der (der Graph) f in P liegt. Beachte dass TFNP sogar eine echte Teilmengen von NPMV_t ist, außer $\text{P} = \text{NP}$:

Beobachtung 3.4 (vgl. Fenner u. a. 2003, Prop. 5). Wenn für alle $f \in \text{NPMV}_t$ auch (der Graph) $f \in \text{P}$ ist, dann gilt $\text{P} = \text{NP}$.

Beweis. Betrachte folgenden NPTM-Transduktor N auf Eingabe $\varphi \in \Sigma^*$: zunächst spaltet sich die Berechnung nichtdeterministisch auf. In der ersten Rechnung wird sofort 1 ausgegeben. In der zweiten Rechnung wird eine Belegung w für die aussagenlogische Formel φ geraten, und 2 ausgegeben wenn w die Formel φ erfüllt. Sei f die Multifunktion, welche von N berechnet wird. Damit gilt:

$$\text{set-}f(x) = \{1, 2\} \text{ falls } x \in \text{SAT}, \text{ und } \{1\} \text{ sonst}$$

und $f \in \text{NPMV}_t$. Nach Annahme ist $f \in \text{P}$. Nun kann aber SAT in Polynomialzeit entschieden werden, denn $\varphi \in \text{SAT}$ genau dann wenn $(\varphi, 2) \in f$.

Die Aussage relativiert, wenn anstelle SAT z.B. das kanonische vollständige Problem gewählt wird. \square

Mit der Beschäftigung mit TFNP-Problemen kam es zu einer umfassenden Theoriebildung. So kam z.B. eine verfeinerte Betrachtung durch Unterklassen von TFNP hinzu. Jede dieser Unterklassen verinnerlicht hierbei jeweils das kombinatorische Prinzip, „warum“ ein Suchproblem total ist. Exemplarisch werden hier zwei Unterklassen skizziert:

- Die Unterklasse PLS („polynomial local search“) umfasst die Suchprobleme, welche in die Form eines Suchgraphen polynomiellen Grads gebracht werden können, worauf ein lokales Optimum gesucht ist. Das zugrunde liegende kombinatorische Prinzip zur Totalität wäre „endliche Suchgraphen haben immer ein lokales Optimum“ oder allgemeiner „Jeder endliche gerichtete azyklische Graph hat eine Senke“.

Ein Beispiel hierfür wäre die Suche nach einem lokal optimalen Schnitt in einem Graphen; hier meint „lokal optimal“ dass kein Flip eines Knotens zu mehr Kantenschnitten führt. Nachdem es nur exponentiell viele Schnitte gibt, muss mindestens einer davon lokal optimal sein.

- Die Unterklasse PPP („polynomial pigeon principle“) umfasst Suchprobleme, welche aufgrund des kombinatorischen Schubfachprinzip total sind.

Ein Beispiel hierfür ist das Gleiche-Summe-Suchproblem: gegeben n positive ganze Zahlen die sich zu $< 2^n - 1$ aufsummieren, finde zwei unterschiedliche nichtleere Teilmengen dieser Zahlen welche die gleiche Summe haben. (Existiert nach Schubfachprinzip: es existieren $2^n - 1$ viele nichtleere Teilmengen, jede davon mit Summe $< 2^n - 1$, die Summen können also nicht alle unterschiedlich sein.)

Auf die weitere Theorie der TFNP-Probleme wird in dieser Arbeit nicht weiter eingegangen. Wir werden aber in Abschnitt ?? noch Reduktionen auf NP-Relationen definieren; dieser Reduktionsbegriff ist der identische wie auf den TFNP-Problemen Megiddo und Papadimitriou (1991).

3.2 Suchprobleme vs. Entscheidungsprobleme

Wie in der Einleitung schon ausgeführt, konzentriert sich die algorithmische Komplexitätstheorie primär auf die Entscheidungsprobleme und weniger auf die Suchprobleme. Das ist durchaus fundiert: es kommt mit einer Vereinfachung der Konzepte, Definitionen und Theorien, und gleichzeitig lässt sich für viele relevante Instanzen das Suchproblem auf das entsprechende Entscheidungsproblem „reduzieren“. Dieses Argument wird gerne als *search reduces to decision* beschrieben.

In diesem Abschnitt werden detailliert Suchprobleme und Entscheidungsprobleme gegenübergestellt und Forschungsergebnisse hierzu aus der Literatur präsentiert. Zum einen wird die eben genannte Reduzierbarkeit und das *search-reduces-to-decision*-Argument ausgeführt, und zum anderen werden Ergebnisse vorgestellt, die darauf hinweisen dass genau dieses Argument nicht für alle Suchprobleme zutrifft.

Wir wollen zunächst auf die Beziehungen zwischen NP-Relationen und NP-Sprachen hinweisen. Über die Projektionen von NP-Relationen können wir genau die Klasse NP von Entscheidungsproblemen charakterisieren: zu jeder Sprache bzw. Entscheidungsproblem $L \in \text{NP}$ existiert (mind.) eine NP-Relation R mit $L = \text{Proj}(R)$, und zu jeder NP-Relation bzw. NP-Suchproblem R ist $\text{Proj}(R) \in \text{NP}$. Das ist die übliche „Zertifikats-Charakterisierung“ von NP aus den Lehrbüchern.

Beobachtung 3.5 (Zertifikats-Definition von NP). $\text{NP} = \{\text{Proj}(R) \mid R \text{ ist eine NP-Relation}\}.$

Beweis. Für die Inklusion von links nach rechts starten wir mit einer Sprache L und einer NPTM N die L entscheidet, wobei die Laufzeit durch das Polynom p beschränkt ist. Definiere nun die Relation

$$R_N = \{(x, \alpha) \mid N(x) \text{ akz. mit RW } \alpha, \alpha \text{ hat } \leq p(|x|) \text{ viele Schritte}\}$$

Diese Relation ist eine NP-Relation. Der Test ist offenbar in Polynomialzeit möglich, und die Relation ist p-balanciert, ist $|\alpha| \in O(\# \text{ Schritte von } \alpha) \in O(p(|x|))$. Aus Definition geht hervor dass $L(N) = \text{Proj}(R_N)$.

Für die Inklusion von rechts nach links konstruieren wir uns zu einer gegebenen NP-Relation R mit Zertifikatschranke q eine NPTM N_R die $\text{Proj}(R)$ entscheidet. Die NPTM N_R rät auf Eingabe x zunächst ein $y \in \Sigma^{\leq q(|x|)}$ und akzeptiert genau dann wenn $(x, y) \in R$. Es ist wegen $R \in \text{P}$ klar, dass diese NTM in Polynomialzeit läuft. Wieder geht aus Definition hervor dass $L(N) = \text{Proj}(R_N)$. \square

Damit ist im Übrigen die obige Definition von NP-Relationen auch nicht „neu“ sondern schon immer mitgedacht. Die eben formulierte Charakterisierung findet sich in allen üblichen Einführungswerken zur Komplexitätstheorie. Dagegen machen die unterschiedlichen Einführungswerke ihren Zugang manchmal stärker von der Perspektive der Suchprobleme abhängig, und manchmal stärker von der typischen Herangehensweise über Entscheidungsproblemen. Vgl. z.B. Goldreich (2008) welcher in seinem Lehrbuch die P-vs.-NP-Frage zunächst als die äquivalente Frage der Beziehung zwischen den „efficiently solvable search problems“ und den „search problems with efficiently checkable solutions“ (letzteres sind genau die NP-Relationen) formuliert. Erst später wird mittels *search-reduces-to-decision*-Argumenten dafür argumentiert, NP-Entscheidungsprobleme als die zentralen Untersuchungsobjekte der Komplexitätstheorie anzusehen.

Zumindest für die P-NP-Frage ist es irrelevant, ob man sich auf Suchprobleme von NP-Relationen oder auf Entscheidungsproblemen von NP-Mengen bezieht. Jedes NP-Suchproblem ist in Polynomialzeit lösbar genau dann wenn jede Menge in NP in deterministischer Polynomialzeit entscheidbar ist.

Lemma 3.6. $\text{FNP} \subseteq_c \text{FP} \iff \text{P} = \text{NP}.$

Beweis. Die Richtung von links nach rechts ist klar, sind ja Suchprobleme schwieriger als Entscheidungsprobleme (Beobachtung 3.2 mit Beobachtung 3.5.)

Die Richtung von rechts nach links zeigen wir mittels Präfixsuche. Sei R eine beliebige NP-Relation mit Zertifikatsschranke q . Wir zeigen dass $R \in_c \text{FP}$. Betrachte folgende Menge

$$A = \{(x, z) \mid \exists z' \in \Sigma^{\leq q(|x|)}, (x, zz') \in R\}$$

Es ist leicht zu sehen dass $A \in \text{NP}$. Also gilt nach Annahme aus $A \in \text{P}$. Nun kann gegeben eine Instanz x iterativ ein Präfix eines Zertifikats verlängert werden:

```

1   $z \leftarrow \varepsilon$ 
2  solange  $|z| \leq q(|x|)$  tue (Invariante: wenn ein Zertifikat  $y$  für  $x$  ex., dann  $z \sqsubseteq y$ )
3      wenn  $(x, z) \in R$  dann
4          akzeptiere mit  $z$ 
5      sonst wenn  $(x, z0) \in A$  dann
6           $z \leftarrow z0$ 
7      sonst wenn  $(x, z1) \in A$  dann
8           $z \leftarrow z1$ 
9      sonst
10         ablehnen
11 ablehnen

```

Korrektheit klar. Unter Annahme $A \in \text{P}$ ist auch klar, dass diese Funktion von einem PTM-Transduktor berechnet werden kann. Damit $R \in_c \text{FP}$. \square

Es ist leicht zu sehen, dass der Suchalgorithmus von obigem Beweis so geändert werden kann, dass anstelle der Entscheidung von A auch Orakelfragen an ein (externes) Orakel A gestellt werden können, d.h. das Suchproblem von R kann *à la Cook* auf das Entscheidungsproblem von A reduziert werden. In anderen Worten, $R \in_c \text{FP}^A$. Das generalisiert sogar, wenn statt A irgendein beliebiges für NP-vollständiges Orakel gewählt wird. Ist also $\text{Proj}(R) \leq_m^{\text{P}}$ -vollständig für NP, dann gilt trivialerweise der Spezialfall $R \in_c \text{FP}^{\text{Proj}(R)}$. Das ist genau das *search-reduces-to-decision*-Argument: ist das Entscheidungsproblem zu R effizient lösbar, dann auch das Suchproblem zu R effizient lösbar.

Korollar 3.7 (*Search reduces to decision* für die NP-Vollständigen). *Sei R eine NP-Relation, für die $\text{Proj}(R)$ auch \leq_m^{P} -vollständig für NP ist. Dann gilt $R \in_c \text{FP}^{\text{Proj}(R)}$.*

Beweis. Wir zeigen die Aussage mit einem Relativierbarkeits-Argument.

Relativ zum Orakel $\text{Proj}(R)$ gilt $\text{P} = \text{NP}$, ist ja $\text{Proj}(R)$ vollständig für NP. Damit gilt mit vorigem Lemma 3.6 auch $\text{FNP} \subseteq_c \text{FP}$ relativ zu $\text{Proj}(R)$. Da $R \in \text{FNP}$, gilt also auch $R \in_c \text{FP}$ relativ zu $\text{Proj}(R)$. \square

Für die NP-Intermediates, also Entscheidungsprobleme aus NP, die weder in P liegen, noch NP-vollständig sind, ist aber unklar, ob immer das Suchproblem auf das Entscheidungsproblem reduziert werden kann.

Wie beim Suchalgorithmus aus obigem Beweis ist aber klar, dass Suchprobleme immer auf eine Präfix- bzw. Bisektion-Entscheidungsvariante reduziert werden können. Im allgemeinen Fall: für jede NP-Relation R mit Laufzeitschranke q gilt

$$R \in_c \text{FP}^{L_R} \text{ wobei } L_R = \{(x, z) \mid \exists z' \in \Sigma^{\leq q(|x|)} \mid (x, zz') \in R\}$$

und

$$R \in_c \text{FP}^{L'_R} \text{ wobei } L'_R = \{(x, z) \mid \exists y \in \Sigma^{\leq q(|x|)} \mid (x, y) \in R, y \leq z\}$$

Konkret ist das zum Beispiel der Fall bei der NP-Relation rSMALLFACTOR . Zur Erinnerung, wir haben

$$\text{Proj}(\text{rSMALLFACTOR}) = \{(n, a) \mid n \text{ nicht prim, ex. nichttrivialer Faktor } p \text{ von } n \text{ mit } p \leq a\}.$$

Durch Orakelfragen an $\text{Proj}(\text{rSMALLFACTOR})$ kann dann mit binärer Suche ein solcher Faktor auch gefunden werden.

Das *search-reduces-to-decision*-Argument hat aber auch Grenzen: Diese Technik scheitert insbesondere, wenn wir wirklich immer die exakte Projektion als Entscheidungsproblem verstehen. Betracht zum Beispiel die NP-Relation zur linearen Teilbarkeit:

$$\text{rLINDIV} = \{(a, b, k) \mid a, b, k \in \mathbb{N}, a \cdot k + 1 \text{ teilt } b\}.$$

Wir wissen dass $\text{Proj}(\text{rLINDIV}) \notin \text{P}$ außer $\text{NP} = \text{coNP}$ (Adleman und Manders 1977); ob $\text{Proj}(\text{rLINDIV})$ NP-vollständig ist, bleibt unklar. Bei dieser NP-Relation wäre nun nicht ersichtlich, wie das Suchproblem auf das Entscheidungsproblem reduziert werden könnte; eine triviale binäre Suche wie oben ist ja nicht möglich.

Für andere Suchprobleme existieren aber nichttriviale Möglichkeiten das Suchproblem auf das (natürliche) Entscheidungsproblem zu reduzieren, auch wenn das Entscheidungsproblem nicht in der Form einer Bisektion/Präfixsuche ist. Hierbei wird die spezifische Struktur

des Problems ausgenutzt. Ein Beispiel ist rSAT : Gegeben Formel φ , teste mittels dem Orakel, ob $\varphi[x_1/0] \in \text{SAT}$ oder $\varphi[x_1/1] \in \text{SAT}$. Hier meint $\varphi[x_1/0]$ die Formel sein soll, die entsteht wenn alle Vorkommen von Variable x_1 in φ mit 0 ersetzt werden, $\varphi[x_1/1]$ analog. Sollte jetzt $\varphi[x_1/0] \in \text{SAT}$ stimmen, dann wissen wir dass es eine Belegung für φ existiert die φ erfüllt und gleichzeitig x_1 auf 0 setzt. Wir können dann iterativ auf dem gleichen Weg eine Belegung für die nächste Variable x_2 bestimmen usw. (Der Fall dass $\varphi[x_1/1] \in \text{SAT}$ ist analog.) Es gilt $\text{rSAT} \in \text{FP}^{\text{Proj}(\text{rSAT})}$.

Beachte aber, dass $\text{Proj}(\text{rSAT}) = \text{SAT}$ schon NP-vollständig ist. Damit folgt $\text{rSAT} \in \text{FP}^{\text{SAT}}$ schon aus Korollar 3.7.

Ein nichttriviales Beispiel für ein Suchproblem, deren Projektion (mutmaßlich) nicht NP-vollständig ist, wäre die NP-Relation rGI . Zur Erinnerung: dieses Suchproblem sucht nach einem Graphisomorphismus zwischen zwei gegebenen Graphen. Hier gilt $\text{rGI} \in_c \text{FP}^{\text{Proj}(\text{rGI})}$: es lässt sich ein Graphisomorphismus zwischen G und H bestimmen, indem mehrmals mittels des Orakels bei (anderen) Paaren von Graphen getestet wird, ob diese isomorph sind (vgl. Goldreich 2008, S. 65, 100). Ob eine nichttriviale Reduktion für rLINDIV möglich ist, scheint in der Literatur nicht untersucht.

Abschließend wollen wir noch theoretische Resultate präsentieren. Die ersten zwei plausibilisieren, dass wahrscheinlich eine NP-Relation existiert, für die das Suchproblem nicht auf das entsprechende Entscheidungsproblem reduziert werden kann (also wie bei rLINDIV vermutet).

Satz 3.8 (Impagliazzo und Sudan 1991; Borodin und Demers 1976, Thm. 5). *Angenommen $E \neq \text{NE}$ oder $P \neq \text{NP} \cap \text{coNP}$. Dann existiert eine NP-Relation mit $\text{Proj}(R) \in \text{NP} - P$ für die $R \notin_c \text{FP}^{\text{Proj}(R)}$ gilt.*

Unter stärkeren Bedingungen kann sogar zeigen, dass sogar Mengen $L \in \text{NP}$ existieren, für die das Suchproblem jeder NP-Relation für L nicht auf das Entscheidungsproblem reduziert werden kann. In anderen Worten, unabhängig davon wie das „Zertifikatssystem“ für L aussieht, ist keins so einfach dass Zertifikate mit Hilfe eines Orakels für das Suchproblem gefunden werden können.

Satz 3.9 (Bellare und Goldwasser 1994, Thm. 1.1; Impagliazzo und Sudan 1991). *Angenommen $\text{EE} \neq \text{NEE}$ oder $\text{NE} \neq \text{coNE}$. Dann existiert eine Menge $L \in \text{NP} - P$ sodass $R \notin_c \text{FP}^L$ für jede NP-Relation R für L , i.e. für die $\text{Proj}(R) = L$ gilt.*

Beachte dass zu keiner dieser Relationen R aus den beiden vorigen Sätzen die Projektion $\text{Proj}(R)$ eine NP-Intermediate ist; $\text{Proj}(R)$ kann nicht \leq_m^P -vollständig sein, denn das wäre ein Widerspruch zu Korollar 3.7.

Folgendes Resultat charakterisiert diejenigen Sprachen $L \in \text{NP}$ die zumindest eine NP-Relation R für L haben, sodass das Suchproblem (bzgl. R) auf das Entscheidungsproblem reduzierbar ist.

Definition 3.10. (1) Eine deterministische OTM heißt *robust für A* falls $L(M^O) = A$ für alle Orakel O .

(2) Eine Menge A heißt *selbsthelfend* falls eine für A robuste OTM M existiert, für die $\text{time}_M^A(x)$ polynomiell in $\text{abh. von } |x|$ wächst, i.e. M^A ist eine POTM (zumindest mit dem Orakel A angeschlossen). \triangleleft

Balcázar fasst die Intuition hinter dieser Definition wie folgt zusammen: man will die Situation abbilden, dass ein Entscheidungsalgorithmus existiert der, mit genug Zeit, immer zu einem korrekten Ergebnis kommt, aber auch mit einem externen „Helfer“ interagieren darf, welcher dem Algorithmus helfen kann, schneller fertig zu rechnen.

Satz 3.11 (Balcázar 1989). *Sei $A \in \text{NP}$. Folgende Aussagen sind äquivalent:*

- (1) *A ist selbsthelfend.*
- (2) *es existiert eine NP-Relation R sodass $\text{Proj}(R) = A$ und $R \in_c \text{FP}^A$.*

Für totale Suchprobleme (i.e. aus TFNP) kann nicht sinnvoll gefragt werden, ob hier das Suchproblem auf das Entscheidungsproblem reduziert werden kann, ist ja für $R \in \text{TFNP}$ das entsprechende Entscheidungsproblem $\text{Proj}(R) = \Sigma^*$ trivial.

Stattdessen können wir uns aber fragen, ob das Suchproblem eines Zertifikats zu x einfacher wird, wenn wir Lösungen zu „kleineren“ Instanzen x' gratis abfragen dürfen. Hierzu können wir folgenden Begriff von Reduzierbarkeit definieren: Betrachte hierbei folgende Variante eine POTM-Transduktors relativ zu $R \in \text{TFNP}$: Dieser Transduktor ist wie ein üblicher PTM-Transduktor, hat zusätzlich aber Zugriff auf ein *funktionales* Orakel, in dem Sinne dass er Orakelfragen der Form „gib mir ein Zertifikat y für x' “ stellen kann. Das Orakel antwortet dann mit einem solchen Zertifikat y mit $(x', y) \in R$. Das existiert, ist ja R total.

Es ist klar, dass mit einem solchen Transduktor relativ zu R auch das Suchproblem zu R lösbar ist. (Gegeben x , stelle einfach die Frage „gib mir Zertifikat für x “.) Deshalb nehmen wir folgende Einschränkung vor: der Transduktor darf bei Eingabe x in den Orakelfragen nur nach Zertifikaten für x' fragen, die kürzer sind als x . Falls selbst unter dieser Einschränkung der Fragen das Suchproblem durch einen solchen Transduktor relativ zu R gelöst werden kann, sagen wir, dass R *nach unten selbstreduzierbar* ist.

Zum Verständnis: Wäre die TFNP-Relation $\text{rFACTOR}' \in \text{TFNP}$ (i.e., suche einen nicht-triviale Faktor für n , oder gebe „prim“ aus) nach unten selbstreduzierbar, dann würde das bedeuten dass ein Faktor von n effizient gefunden werden kann, wenn wir nach Faktoren von Zahlen $\leq n/2$ fragen dürfen.

Im Bezug auf die Faktorisierung zeigen Harsha, Mitropolsky und Rosen, dass diese wahrscheinlich nicht nach unten selbstreduzierbar ist.

Satz 3.12 (Harsha, Mitropolsky und Rosen 2023). *Die NP-Relation $\text{rFACTOR}' \in \text{TFNP}$ ist nicht nach unten selbstreduzierbar, außer $\text{rFACTOR}' \in \text{PLS}$.*

TODO: Hier wäre schön ein Beispiel eines Problems (muss in PLS liegen) zu haben, das nach unten selbstreduzierbar ist

Es ist offen ob $\text{rFACTOR}' \stackrel{?}{\in} \text{PLS}$ und zumindest unplausibel, weil unklar ist wie Faktorisierung als lokales Suchproblem repräsentiert werden kann. Tatsächlich zeigen die Autorinnen sogar die Konsequenz $\text{rFACTOR}' \in \text{UEOPL}$, was auch noch $\text{rFACTOR}' \in \text{PPAD}$ zur Folge hätte. Auch die Frage ob $\text{rFACTOR}' \stackrel{?}{\in} \text{PPAD}$ ist offen, und wurde breit untersucht. Eine positive Antwort wäre zumindest sehr überraschend (vgl. Harsha, Mitropolsky und Rosen 2023, 67:15; siehe ebd. auch für eine Def. von UEOPL, PPAD). Insgesamt ist die Forschung bezüglich Selbstreduzierbarkeit nach unten für TFNP-Probleme noch sehr klein, und es bedarf auf jeden Fall weiterer Untersuchungen, unter anderem auch in Richtung einer Konzeptionalisierung von „kleinerer Instanz“, die robuster als „kürzerer String“ ist. Vergleiche mit der *disjunktiver Selbstreduzierbarkeit* (Balcázar 1989, vgl. Selman 1988, vgl. Wechsung 2000, Abschn. 9.5) – einem ähnlichen Begriff der Selbstreduzierbarkeit auf der Ebene der Entscheidungsprobleme – liegen nahe.

3.3 Levin-Reduzierbarkeit

Ähnlich wie auf den üblichen Entscheidungsproblemen können wir auch von Reduzierbarkeiten zwischen verschiedenen Suchproblemen sprechen. In der Literatur hat sich folgender Begriff von Reduzierbarkeit zwischen NP-Relationen herausgebildet (vgl. Papadimitriou 1994, S. 229; Goldreich 2008, S. 61; Arora und Barak 2009, S. 50):

Definition 3.13 (Levin-Reduzierbarkeit). Seien Q, R zwei NP-Relationen. Wir sagen dass Q *sich auf R (Polynomialzeit-)Levin-reduzieren lässt*, bzw. $Q \leq_L^p R$ wenn zwei Funktionen $f, g \in \text{FP}$ existieren sodass

- (1) $x \in \text{Proj}(Q) \iff f(x) \in \text{Proj}(R)$,
- (2) $(f(x), y) \in R \implies (x, g(x, y)) \in Q$.

Punkt (1) sagt also nur aus, dass f eine Many-one-Polynomialzeit-Reduktion zwischen den entsprechenden Entscheidungsproblemen ist. Punkt (2) sagt nun aus, dass wenn y ein Zertifikat für die Instanz $f(x)$ aus R ist, dann lässt sich aus y wieder ein Zertifikat $g(x, y)$ für die originale Instanz x berechnen.

Die Funktion f nennen wir *Reduktionsfunktion*, die Funktion g nennen wir *Translationsfunktion*.

Wir schreiben $Q \leq_{L,1}^p R$ falls f zusätzlich injektiv ist. Wir schreiben $Q \leq_{L,1,\text{inv}}^p R$ falls f zusätzlich injektiv und p-invertierbar ist. Klar ist:

$$Q \leq_{L,1,\text{inv}}^p R \implies Q \leq_{L,1}^p R \implies Q \leq_L^p R \implies \text{Proj}(Q) \leq_m^p \text{Proj}(R).$$

Definiere die Duplikatrelation \equiv_L^p entsprechend. ◁

Die Bezeichnung *Levin-Reduktion* ist hier in Anlehnung an bisherige Verwendung gewählt, und bezieht sich darauf, dass in der Etablierung der NP-Vollständigkeit durch Karp (1972), Cook (1971) und Levin (1973) gerade Levin die Suchprobleme in den Blick genommen hat, während Karp und Cook sich auf Entscheidungsprobleme konzentriert haben. Die Formalisierung von NP-Suchproblemen durch NP-Relationen (Definition 3.1) findet sich in Grundzügen schon in Levins Präsentation. Es sei aber darauf hingewiesen, dass sich die hier genannte Definition der Levin-Reduzierbarkeit (Definition 3.13) eine schwächere Form der Reduzierbarkeit ist als die eigentliche von Levin vorgeschlagene. Die hier genannte Definition ist jedoch hinreichend für alle relevanten Eigenschaften, sowie für die Aussagen aus Levins eigener Publikation.

Beachte dass \leq_L^p -Reduktionen eine Verstärkung von \leq_m^p -Reduktionen auf den jeweiligen Projektionen darstellt:

Beobachtung 3.14. Wenn $R \leq_L^P Q$ dann gilt $\text{Proj}(R) \leq_m^P \text{Proj}(Q)$.

Die Relationen \leq_L^P , $\leq_{L,1}^P$ und $\leq_{L,1,i}^P$ sind reflexiv und transitiv, bilden also eine Quasiordnung. Intuitiv formt die Levin-Reduktion \leq_L^P auf den Suchproblemen das Analog der Many-One-Reduktion \leq_m^P auf den Entscheidungsproblemen.

Genau so wie wir es bei der üblichen \leq_m^P -Reduktion auf den Suchproblemen gewohnt sind, ordnet \leq_L^P die Suchprobleme der NP-Relationen nach ihrer „Schwierigkeit“: wenn $Q \leq_L^P R$ dann ist Q höchstens so „schwer“ wie R : gegeben einen Lösungsalgorithmus für R lässt sich auch Q effizient lösen, und das sogar mit nur einer Anfrage an den Lösungsalgorithmus. Damit folgt: wenn das Suchproblem zu R effizient gelöst werden kann, dann kann auch das Suchproblem zu Q gelöst werden. Formal ausgedrückt ist FP nach unten abgeschlossen unter der \leq_L^P -Ordnung:

Lemma 3.15. Wenn $Q \leq_L^P R$ und $R \in_c \text{FP}$ dann ist $Q \in_c \text{FP}$.

Beweis. Seien f, g die Reduktions- bzw. Translationsfunktion, welche $Q \leq_L^P R$ realisieren, und sei $r \in \text{FP}$ eine Verfeinerung von R . Definiere nun

$$q(x) = \begin{cases} g(x, r(f(x))) & \text{falls } r(f(x)) \neq \perp \\ \perp & \text{sonst.} \end{cases}$$

Offenbar ist $q \in \text{FP}$. Wir zeigen nun dass q eine Verfeinerung von Q ist. Zum einen gilt $\text{dom}(q) = \text{Proj}(Q)$:

$$x \in \text{Proj}(Q) \iff f(x) \in \text{Proj}(R) \iff f(x) \in \text{dom}(r) \iff q(x) \neq \perp \iff x \in \text{dom}(q).$$

Hier folgt die erste Äquivalenz nach Definition 3.13(1). Zum anderen haben wir

$$x \in \text{Proj}(Q) \implies f(x) \in \text{Proj}(R) \implies (f(x), r(f(x))) \in R \implies (x, g(x, r(f(x)))) \in Q \implies q(x) \in \text{set-}Q(x),$$

i.e. $q(x)$ ist ein Zertifikat für x , wie gewünscht. Hier folgt die dritte Implikation nach Definition 3.13(2). \square

Genau so wie bei der Many-one-Reduktion auf den Suchproblemen können wir nach größten Elementen auf der \leq_L^P -Ordnung fragen.

Definition 3.16. Sei \mathcal{F} eine Klasse von Multifunktionen (z.B. FNP oder TFNP). Wir nennen $R \in \mathcal{F}$ \leq_L^P -vollständig für \mathcal{F} wenn R ein größtes Element von \mathcal{F} geordnet über \leq_L^P ist: Für alle $Q \in \mathcal{F}$ gilt $Q \leq_L^P R$.

Die $\leq_{L,1}^P$ - und $\leq_{L,1,i}^P$ -Vollständigkeit ist analog definiert. \triangleleft

Es existiert eine \leq_L^P -vollständige NP-Relation für FNP. Diese ist im Wesentlichen die natürliche Erweiterung der kanonischen \leq_m^P -vollständigen Menge KAN:

$$\text{rKAN} = \{((N, x, 1^n), \alpha) \mid \alpha \text{ ist ein akz. Rechenweg auf } N(x) \text{ und } |\alpha| \leq n\}.$$

Beachte dass $\text{Proj}(\text{rKAN}) = \text{KAN}$.

Satz 3.17. Die kanonische NP-Relation rKAN ist $\leq_{L,1,\text{inv}}^P$ -vollständig für FNP.

Beweis. Sei R eine beliebige NP-Relation mit Zertifikatsschranke r , i.e. $(x, y) \in R \implies |y| \leq r(|x|)$. Sei M die PTM welche R entscheidet, mit Laufzeitschranke p . Sei N eine NPTM welche auf Eingabe x zunächst ein Zertifikat y , $|y| \leq r(|x|)$ rät, und dann testet ob $M(x, y)$ akzeptiert. Die Laufzeit von N ist beschränkt auf $p(|(x, y)|) \in O(p(r(|x|)))$ (hier nutzen wir die effiziente Listencodierung von ?? aus). Sei daher q ein Polynom, welches die Laufzeit von N beschränkt.

Definiere die Reduktionsfunktion $f(x) = (N, x, 1^{q(|x|)})$. Wir zeigen zunächst dass

$$x \in \text{Proj}(R) \iff f(x) \in \text{Proj}(\text{rKAN}).$$

Wenn $x \in \text{Proj}(R)$, dann existiert ein y , $|y| \leq r(|x|)$ sodass $(x, y) \in R$. Dann wird auch $N(x)$ akzeptieren, nämlich auf jenem Pfad welcher y rät. Es existiert also ein Rechenweg α mit $|\alpha| \leq q(|x|)$ sodass $N(x)$ auf α akzeptiert. Dann gilt aber auch $(f(x), \alpha) = ((N, x, 1^{q(|x|)}), \alpha) \in \text{rKAN}$. Die Rückrichtung $x \notin \text{Proj}(R) \implies f(x) \notin \text{Proj}(\text{rKAN})$ folgt analog. Es ist klar, dass f injektiv ist, dass f Polynomialzeit-berechenbar und -invertierbar ist.

Es lässt sich außerdem einfach eine Translationsfunktion $g \in \text{FP}$ angeben, die für $g(f(x), \alpha) = y$ aus α das entsprechende geratene Zertifikat y aus α berechnen kann. \square

Die Ordnung \leq_L^P und dessen Vollständigkeits-Begriff verhält sich auch sonst wie bei dem Analog \leq_m^P gewohnt.

Lemma 3.18. Sei R eine \leq_L^P -vollständige NP-Relation für FNP. Es gelten folgende Aussagen:

- (1) $\text{Proj}(R)$ ist eine \leq_m^P -vollständige Menge für NP.
- (2) Wenn Q eine NP-Relation ist und $R \leq_L^P Q$, dann ist auch $Q \leq_L^P$ -vollständig.
- (3) $R \in_c \text{FP} \iff \text{FNP} \subseteq_c \text{FP} \iff \text{NP} = \text{P}$.

Es existieren auch natürliche NP-Relationen, die \leq_L^P -vollständig für FNP sind. Das Bekannteste ist rSAT . Zur Erinnerung:

$$\text{rSAT} = \{(\varphi, w) \mid \varphi \text{ ist eine aussagenlogische Formel, } w \text{ erfüllende Belegung für } \varphi\}.$$

Die Aussage „ rSAT ist \leq_L^P -vollständig“ entspricht dem üblichen Cook-Levin-Satz, und wird hier nur wiederholt:

Satz 3.19 (Satz von Cook und Levin). *Die NP-Relation rSAT ist $\leq_{\text{L},1,\text{inv}}^{\text{P}}$ -vollständig für FNP.*

Skizze. Wir zeigen nur, dass $\text{rCSAT} = \{(C, w) \mid C \text{ ist SAT-Schaltkreis und } C(w) = 1\}$, i.e. Schaltkreiserfüllbarkeit, vollständig ist. Es ist leicht zu sehen dass $\text{rCSAT} \leq_{\text{L}}^{\text{P}} \text{rSAT}$ mit den gleichen Argumenten wie $\text{CSAT} \leq_{\text{m}}^{\text{P}} \text{SAT}$.

Ein üblicher Beweis des Satzes von Cook und Levin (welcher auf der Seite der Entscheidungsprobleme operiert) zeigt als erstes, dass eine PTM M und eine „Eingabegröße“ n als Schaltkreis C repräsentiert werden kann, sodass $C_{M,n}(x) = 1 \iff M(x)$ akzeptiert, für alle $x \in \Sigma^n$.

Sei nun R eine beliebige NP-Relation mit Laufzeitschranke q . Ohne Beschränkung können wir in diesem Beweis annehmen, dass alle Zertifikate für x bezüglich R genau die Länge $q(|x|)$ haben. Dann existiert eine PTM M die R entscheidet. Sei $x \in \Sigma^*$ gegeben. Für geeignetes n haben wir

$$\forall y \in \Sigma^{q(|x|)}. \quad C_{M,n}(x, y) = 1 \iff M(x, y) \text{ akz.} \iff (x, y) \in R.$$

Wenn wir jetzt x in $C_{M,n}$ „hart verdrahten“ erhalten wir einen Schaltkreis $C_{M,n,x}$ und es gilt

$$\forall y \in \Sigma^{q(|x|)}. \quad C_{M,n,x}(y) = 1 \iff C_{M,n}(x, y) = 1 \iff (x, y) \in R.$$

Und damit ist $f(x) = C_{M,n,x}$ eine Reduktionsfunktion von R nach rCSAT :

$$\begin{aligned} x \in \text{Proj}(R) &\iff \exists y \in \Sigma^{q(|x|)}. (x, y) \in R \iff \exists y \in \Sigma^{q(|x|)}. C_{M,n,x}(y) = 1 \\ &\iff C_{M,n,x} \in \text{Proj}(\text{rCSAT}) \iff f(x) \in \text{Proj}(\text{rCSAT}). \end{aligned}$$

Eine Translationsfunktion g kann auch einfach angegeben werden, wenn $(f(x), y) \in \text{rCSAT}$ dann gilt $C_{M,n,x}(y) = 1$ und damit (modulo einer ggf. notwendigen Umcodierung) $(x, y) \in R$.

Es ist leicht zu sehen dass f injektiv ist, denn wenn die Schaltkreise $f(x), f(x')$ identisch sind, dann muss auch $x = x'$. Genauso ist klar, dass aus $f(x) = C_{M,n,x}$ einfach das „hineincodierte“ x wieder ausgelesen werden kann. \square

Die bekannten NP-Relationen R mit $\leq_{\text{m}}^{\text{P}}$ -vollständiger Projektion $\text{Proj}(R)$ sind auch Levin-vollständig. Typische Präsentationen von $\leq_{\text{m}}^{\text{P}}$ -vollständigkeit, z.B. $\text{SAT} \leq_{\text{m}}^{\text{P}} \text{VC}$ geben uns nicht nur eine $\leq_{\text{m}}^{\text{P}}$ -Reduktionsfunktion f von Instanzen x der einen Menge (i.e. SAT) zu Instanzen $f(x)$ der anderen Menge (i.e. VC), sondern beinhalten meist im Beweis eine (implizit mitgedachte) effiziente Übersetzung von Zertifikaten von x nach $f(x)$ und umgekehrt. Die Reduktionsfunktion f mit der Rückübersetzung der (natürlichen) Zertifikate für $f(x)$ nach Zertifikaten für x reichen dann aus, um eine $\leq_{\text{L}}^{\text{P}}$ -Reduktion zu realisieren, und damit Levin-Vollständigkeit zu zeigen.

Nach Goldreich (2008, S. 104) sind die folgenden NP-Relationen jedenfalls definitiv $\leq_{\text{L},1,\text{i}}^{\text{P}}$ -vollständig: rSAT , rSETCOVER , rVC , rCLIQUE , r3COLORABILITY . Die von Papadimitriou (1994, S. 193–198) angegebene Reduktion $\text{SAT} \leq_{\text{m}}^{\text{P}} \text{rHAMCYCLE}$ lässt sich leicht zu $\text{rSAT} \leq_{\text{L}}^{\text{P}} \text{rHAMCYCLE}$ erweitern, womit rHAMCYCLE auch $\leq_{\text{L}}^{\text{P}}$ -vollständig ist. Ebenso ist rANOTHERHAMCYCLE auch $\leq_{\text{L}}^{\text{P}}$ -vollständig (1994, S. 232). Damit gilt mit Lemma 3.18(3) auch die in Abschnitt ?? angegebene Äquivalenz $\text{NP} = \text{P}$ genau dann wenn rSAT , rVC , ... \in_{c} FP.

NP-Relationen R für welche die Projektion zwar $\leq_{\text{m}}^{\text{P}}$ -vollständig, aber R nicht $\leq_{\text{L}}^{\text{P}}$ -vollständig sind, scheinen nicht bekannt zu sein. Aus dieser empirischen Beobachtung ergibt sich die Frage, ob das auch für *alle* NP-Relationen R gilt: wenn $\text{Proj}(R)$ eine $\leq_{\text{m}}^{\text{P}}$ -vollständige Menge für NP ist, ist dann auch R eine $\leq_{\text{L}}^{\text{P}}$ -vollständige NP-Relation für FNP? Aus Präsentationsgründen werden wir die pessimistische negative Antwort auf diese Frage als Vermutung formulieren:

Vermutung 3.20 (Karp-vs-Levin-Vermutung; KvL). *Es existiert eine NP-Relation R sodass $\text{Proj}(R) \leq_{\text{m}}^{\text{P}}$ -vollständig für NP ist, aber R ist nicht $\leq_{\text{L}}^{\text{P}}$ -vollständig für FNP.*

Obwohl diese Frage erstaunlich auf natürlich scheint, und zwei umfassende Reduktionsbegriffe der Komplexitätstheorie in Beziehung setzen versucht, gibt es erstaunlicherweise kaum Forschung welche sich dieser Hypothese annähert. Ein Beweis von KvL ist jedenfalls mindestens so schwer wie die P-NP-Frage, denn $\text{KvL} \Rightarrow \text{P} \neq \text{NP}$. In Abschnitt ?? werden wir diese Hypothese und dessen Beziehung zu anderen Hypothesen erarbeiten. Dort werden dann auch Argumente geliefert, die für diese negative Antwort sprechen.

Die oben genannte Frage lässt sich auf natürliche Weise abschwächen, indem man von der konkreten NP-Relation abstrahiert: Wenn L eine $\leq_{\text{m}}^{\text{P}}$ -vollständige Menge für NP ist, existiert dann zumindest *eine* NP-Relation R_L für L (i.e., $\text{Proj}(R) = L$) sodass $R_L \leq_{\text{L}}^{\text{P}}$ -vollständig ist. In anderen Worten, existiert ein hinreichend ausdrucksstarkes „Zertifikatssystem“ R für L sodass $R \leq_{\text{L}}^{\text{P}}$ -vollständig ist? Diese abgeschwächte Frage lässt sich positiv beantworten, falls man die Berman–Hartmanis-Vermutung IC annimmt:

Beobachtung 3.21 (Buhrman, Kadin und Thierauf 1998). *Für jede Menge $L \in \text{NP}$ die p -isomorph zu SAT ist, existiert eine NP-Relation R_L sodass $\text{Proj}(R_L) = L$ und R_L auch $\leq_{\text{L}}^{\text{P}}$ -vollständig ist.*

Beweis. Nach Voraussetzung haben wir eine bijektive p -invertierbare Funktion $h \in \text{FP}$ mit $x \in L \iff h(x) \in \text{SAT}$. Definiere nun

$$R_L = \{(x, w) \mid (h(x), w) \in \text{rSAT}\}.$$

Es ist leicht zu sehen, dass R_L eine NP-Relation ist. Es ist auch leicht zu sehen dass $\text{Proj}(R_L) = L$.

Wir zeigen nun, dass R_L auch \leq_L^p -vollständig ist. Sei hierfür Q eine beliebige NP-Relation. Nachdem rSAT ja \leq_L^p -vollständig ist, existieren Reduktions- und Translationsfunktionen f, g die $Q \leq_L^p \text{rSAT}$ realisieren. Definiere nun

$$f'(x) = h^{-1}(f(x)).$$

Insbesondere ist $h^{-1}(\cdot)$ wohldefiniert, ist ja h surjektiv. Damit gilt zum einen für f'

$$x \in \text{Proj}(Q) \iff f(x) \in \text{SAT} \iff \underbrace{h(h^{-1}(f(x)))}_{f'(x)} \in \text{SAT} \iff f'(x) \in \text{Proj}(R_L),$$

und zum anderen gilt

$$(f'(x), w) \in R_L \implies (h(h^{-1}(f(x))), w) \in \text{rSAT} \implies (f(x), w) \in \text{rSAT} \implies (x, g(x, w)) \in Q.$$

Damit erfüllen also f' und g die Voraussetzungen an eine Reduktions- bzw. Translationsfunktion und $Q \leq_L^p R_L$, wie gewünscht. \square

(Es ist leicht zu sehen, dass diese Aussage relativiert, wenn anstelle rSAT eine andere beliebige \leq_L^p -vollständige Relation R gewählt wird.)

Damit haben (im unrelativierten Fall) insbesondere alle *bekannten* \leq_m^p -vollständigen Mengen, i.e. zu SAT p-isomorphen Mengen, eine entsprechende \leq_L^p -vollständige NP-Relation. Es muss aber gleichzeitig darauf hingewiesen werden, dass die Zertifikate in den entsprechenden Relationen R_L nicht natürlich sind; die Zertifikate sind nur Belegungen für die Formeln $h(x)$ und haben an sich keinen Bezug bzw. Interpretierbarkeit gegenüber der Instanz x .

Wir schließen mit einer Diskussion zur Vollständigkeit bezüglich TFNP ab. Zum einen ist klar, dass unter dem hier vorgeschlagenen Reduktionsbegriff keine Reduktion von einer nicht-totalem NP-Relation auf eine (totale) TFNP-Relation möglich ist:

$$R \not\leq_L^p Q, \text{ für alle } Q \in \text{TFNP}, R \in \text{FNP}, \text{Proj}(R) \neq \Sigma^*,$$

denn jede potentielle Reduktion würde Definition 3.13(1) verletzen, i.e. die Many-one-Reduktion auf den jeweiligen Projektionen.

Andererseits kann gefragt werden, ob TFNP-Relationen existieren die \leq_L^p -vollständig für TFNP sind. In der Literatur (vgl. Pudlák 2017) wird hierauf eine negative Antwort vermutet:

Vermutung 3.22 (TFNP). *Es existiert keine NP-Relation $R \in \text{TFNP}$ die \leq_L^p -vollständig für TFNP ist.*

Auch hier ist ein Beweis für diese Vermutung mindestens so schwer wie ein Beweis für $\text{NP} \neq \text{coNP}$. Die Beziehung dieser Vermutung mit weiteren Vermutungen betreffend Promise-Problemen wird in Abschnitt ?? erarbeitet.

3.4 Zur gemeinsamen Struktur von vollständigen Suchproblemen

Aus den Erfahrungen in den Entdeckungen von NP-vollständigen Mengen bzw. der Entwicklung der hierfür notwendigen Reduktionen wurde intuitiv deutlich, dass die NP-vollständigen Mengen sich viele nicht-offensichtliche Eigenschaften teilen, die über „Equi-Lösbarkeit“ ($A \in P$ genau dann wenn $B \in P$ für \leq_T^p -vollständige Mengen A, B) hinaus geht. Das beginnt schon bei der sonst vorgezogenen und üblichen Many-one-Reduktion. Hier wird auf die fundamental ähnliche Struktur der \leq_m^p -vollständigen Mengen A, B hingewiesen: eine Instanz x von A kann als ein „äquivalenter“ Fall $f(x)$ von B repräsentiert werden.

Die intuitive Beobachtung, dass sämtlichen Beweise der \leq_m^p -Vollständigkeit zu Beweisen der $\leq_{1,1}$ -Vollständigkeit verstärkt werden können, führte schlussendlich zur (Berman-Hartmanis-)Isomorphievermutung IC, in der postuliert wird, dass es im Wesentlichen nur *eine* NP-vollständige Menge gibt, und die verschiedenen Ausprägungen unterschiedlicher NP-vollständiger Mengen nur triviale Umcodierungen des selben Problems sind. Obwohl die Forschung zu NP-Suchproblemen im Vergleich zu den entsprechenden Entscheidungsproblemen im Hintergrund blieb, wurde in der Forschung aufgrund der oben genannten Erfahrungen und Intuitionen die Beobachtung gemacht, dass viele der entsprechenden Suchprobleme eine inhärente strukturelle Ähnlichkeit untereinander haben (vgl. auch die Diskussion von Hemaspaandra 1998). Im Folgenden werden einige dieser Arbeiten kurz skizziert.

Simon (1975, S. 83) machte beispielsweise die Beobachtung, dass die ihm bekannten Reduktionsfunktionen $f: A \rightarrow B$ in den Beweisen zur NP-Vollständigkeit so gebaut sind, dass die Instanz x genau k „Lösungen“ bezüglich A hat genau dann wenn $f(x)$ genau k „Lösungen“ bezüglich B hat. „Lösung“ hier in Anführungszeichen weil auf Mengen überhaupt kein Begriff von Lösungen bzw. Zertifikaten existiert; Simon dachte in seinen Überlegungen die zugrunde liegende kombinatorischen (Such-)Probleme zu A und B nur mit.

Auf NP-Relationen lässt sich sein Reduktionsbegriff aber formal präzise formulieren:

Definition 3.23 (Geizige Reduktionen). Seien Q, R NP-Relationen. Wir sagen dass sich Q auf R (in Polynomialzeit) *geizig* („parsimonious“) reduzieren lässt, bzw. $Q \leq_{\text{pars}}^p R$ wenn eine Funktion $f \in \text{FP}$ existiert mit

$$|\text{set-}Q(x)| = |\text{set-}R(f(x))|.$$

◁

Beachte dass geizige Reduktionen immer eine Many-one-Reduktion realisieren: wir haben $x \in \text{Proj}(Q)$ genau dann wenn mind. ein Zertifikat für x bzgl. Q existiert, also genau dann wenn eines für $f(x)$ bzgl. R existiert, bzw. $f(x) \in \text{Proj}(R)$.

Geizige Reduktionen wurden insbesondere in der Komplexitätstheorie des Zählens aufgegriffen (Simon 1975; Valiant 1979). Typische algorithmische Probleme sind z.B. „wie viele Belegungen w erfüllen die aussagenlogische Formel φ “ oder, kanonischer, „Auf wievielen Rechenwegen akzeptiert die Berechnung $N(x)$ der NPTM $N^{\mathbb{Z}}$ “. Es ist einfach zu sehen, dass sich geizige Reduktionen $\text{rSAT} \leq_{\text{pars}}^p \text{rKAN}$ und $\text{rKAN} \leq_{\text{pars}}^p \text{rSAT}$ angeben lassen können. Damit kann das Zählproblem zur rSAT -Instanz x als rKAN -Instanz $f(x)$ repräsentiert werden und umgekehrt – die beiden Zählprobleme sind relativ zum jeweils anderem gleich schwer (wobei letztere Aussage auch mit schwächeren Reduktionsbegriffen gezeigt werden kann). Auf eine weitere Präsentation der Komplexitätstheorie des Zählens muss hier verzichtet werden.

Beachte, dass geizige Reduktionen nicht mit Levin-Reduktionen vergleichbar sind. Levin-Reduktionen erhalten im Allgemeinen nicht die Anzahl an Zertifikaten, während umgekehrt geizige Reduktionen keine effektive Übersetzung zwischen den Zertifikaten für $f(x)$ auf Zertifikate für x zulassen.

Lynch und Lipton (1978) verstärkt den Reduktionsbegriff der sparsamen Reduktionen und setzt voraus, dass Zertifikate für x in Zertifikate für $f(x)$ effizient umgerechnet werden können. Beachte aber, dass diese „Vorwärtsübersetzung“ nicht hinreichend ist um Levin-Reduzierbarkeit zu zeigen (die ja umgekehrt Zertifikate für $f(x)$ in Zertifikate für x umrechnet).

Fischer, Hemaspaandra und Torenvliet 1995 gingen noch einen Schritt weiter und definieren *zertifikats-isomorphe* Reduktionen („witness-isomorphic reduction“) zwischen zwei NP-Relationen. Hier erhält die Reduktionsfunktion $A \rightarrow B$ auf den Instanzen nicht nur die Anzahl der Zertifikate, sondern es werden zusätzlich die Zertifikate für $x \in A$ mit den Zertifikateion für $f(x)$ in B in eine effiziente in Polynomialzeit berechenbare Eins-zu-Eins-Korrespondenz gesetzt. Dieses Vorgehen ist intendiert als eine Generalisierung der p-Isomorphie (Hartmanis und Berman 1976) nicht nur auf Instanzen sondern auch auf Zertifikaten.

Definition 3.24 (Zertifikats-Isomorphie; Fischer, Hemaspaandra und Torenvliet 1995). Seien Q, R NP-Relationen. Wir sagen dass sich Q auf R (in Polynomialzeit) *zertifikats-isomorph* reduzieren lässt, bzw. $Q \leq_{\text{wi}}^p R$, wenn Funktion $f : \Sigma^* \rightarrow \Sigma^*, f \in \text{FP}$ und Funktion $g : \Sigma^* \times \Sigma^* \rightarrow \Sigma^* \in \text{FP}$ existieren, die p-invertierbar (also auch injektiv) sind, und

- (1) $(x, y) \in Q \implies (f(x), g(x, y)) \in R$ (Vorwärts-Translation von Zertifikaten),
- (2) $(f(x), z) \in R \implies \exists y. (x, y) \in R \wedge g(x, y) = z$ (g ist quasi „surjektiv“),
- (3) Falls $y_1, y_2 \in \text{set-}Q(x)$ und $y_1 \neq y_2$, dann ist auch $g(x, y_1) \neq g(x, y_2)$ (g ist quasi „injektiv“).

◁

(Der oben skizzierte Reduktionsbegriff von Lynch und Lipton (1978) geht aus der Zertifikats-Isomorphie hervor, wenn „p-invertierbar“ in der Definition gestrichen wird.)

Die Punkte (1)–(3) in der Definition der Zertifikats-Isomorphie können alternativ auch folgendermaßen äquivalent formuliert werden:

$$\forall x. |\text{set-}Q(x)| = |\text{set-}R(f(x))| \wedge \{g(x, y) \mid y \in \text{set-}Q(x)\} = \text{set-}R(f(x)). \quad (3.2)$$

Damit ist auch leicht zu sehen, dass für alle $x \in \text{Proj}(Q)$ die Funktion $g_x(y) = g(x, y)$ eine Bijektion zwischen den Zertifikaten für x und den Zertifikaten für $f(x)$ ist, sowie dass f eine sparsame Reduktion realisiert (und damit auch eine Many-one-Reduktion).

Um die Analogie zur p-Isomorphie abzuschließen: Fischer, Hemaspaandra und Torenvliet definieren einen *witness-isomorphic isomorphism* zwischen zwei NP-Relationen Q, R wenn $Q \leq_{\text{wi}}^p R$ via f, g und $R \leq_{\text{wi}}^p Q$ via f^{-1}, g^{-1} . Wieder analog zur Berman–Hartmanis ist für einen *witness-isomorphic isomorphism* ausreichend, wenn $Q \leq_{\text{wi}}^p R$ und $R \leq_{\text{wi}}^p Q$ jeweils über verlängernde Funktionen, à la Schröder–Bernstein.

Dieser Reduktionsbegriff stellt eine Verstärkung der Levin-Reduktion und sparsamen Reduktion dar:

Lemma 3.25. Seien Q, R zwei NP-Relationen. Falls $Q \leq_{\text{wi}}^p R$, dann ist $Q \leq_{\text{pars}}^p R$ und $Q \leq_{\text{L}}^p R$ und $Q \leq_{\text{m}}^p R$.

Beweis. Seien $f, g \in \text{FP}$ die zwei Funktionen, welche $Q \leq_{\text{wi}}^p R$ realisieren. Wir haben in der vorhergehenden Diskussion über Gleichung 3.2 schon gesehen, dass f eine sparsame (Many-one-)Reduktion darstellt.

Wir müssen für Levin-Reduzierbarkeit nur noch zeigen, dass wir Zertifikate z für $f(x)$ wieder zu Zertifikaten für x rückübersetzen können. Das wegen der p -Invertierbarkeit von g möglich: Mit Definition 3.24(2) gilt

$$(f(x), z) \in R \implies g^{-1}(z) = (x, y) \text{ und } (x, y) \in R.$$

Insbesondere ist mit Definition 3.24(2) auch $g^{-1}(z) = (x, y)$ definiert, und nachdem $g^{-1} \in \text{FP}$, lässt sich leicht eine entsprechende Translationsfunktion für die Levin-Reduktion angeben. \square

Wir werden später zeigen, dass natürliche NP-Relationen die mutmaßlich nicht \leq_{pars}^p -vollständig sind, und damit auch nicht zertifikats-isomorph zu rSAT sein können.

Einen anderen Weg gingen Agrawal und Biswas (1992). Sie sind konkret daran interessiert, wie die natürlichen NP-Vollständigen Mengen/Relationen konkret strukturiert sind, bzw. welche sie sich teilen. Die Intuition ist am verständlichsten, wenn man sich in Erinnerung ruft, wie übliche Beweise der NP-Vollständigkeit auf Mengen funktionieren. Zum Beispiel beinhaltet ein Beweis der NP-Vollständigkeit von VC eine Many-one-Reduktion von 3SAT auf VC . Eine übliche Strategie für diese Reduktion ist es nun, „Gadgets“ auf VC (hier: Teilgraphen) zu definieren, welche Klausen und Variablen simulieren. Diese werden dann zu einer Instanz $f(\varphi)$ zusammengesetzt, um ganze SAT-Formeln φ zu simulieren; damit kann dann die Reduktion von Formeln auf VC realisiert werden.

Agrawal und Biswas formalisieren diese Eigenschaft nun, und nennen NP-Relationen *universell*, wenn diese die Konstruktion eines *building blocks* (ungefähr eine Klausel) zulässt, und die *joinable* (entspricht ungefähr Disjunktion) und die *couplable* (entspricht ungefähr Konjunktion) sind. Damit lässt sich die NP-Vollständigkeit vieler Probleme bzw. Relationen in einer uniformen und allgemeinen Weise aus rein strukturellen Eigenschaften ableiten. Die Eigenschaft *universell* stellt eine Verstärkung der Levin-Vollständigkeit da, und ist eine der stärksten strukturellen Eigenschaften, die (zumindest von Agrawal und Biswas) bei allen natürlichen Levin-vollständigen NP-Relationen (modulo einer trivialen Umcodierung der Zertifikate) vermutet wird. Daher wird deren Arbeit hier auch extensiver ausgeführt.

Eine zentrale Methode von Agrawal und Biswas ist, oft relevante Informationen direkt aus dem „String“ der Zertifikate auslesen. Daher beschränken wir auf folgende Teilmenge der NP-Relationen, bei der die Zertifikatsstring möglichst uniform sind:

Definition 3.26 (Strenge NP-Relation). Sei R eine NP-Relation mit Zertifikatsschranke q . Wir nennen R *strenge* wenn für R gilt, dass $(x, y) \in R \implies |y| = q(|x|) > 0$. In anderen Worten, jedes Zertifikat y für x ist nicht ε und hat genau die Länge $q(|x|)$. \triangleleft

Dies sollte keine Einschränkung darstellen. zu jeder natürlichen NP-Relation R kann eine strenge NP-Relation R' angegeben werden mit $\text{Proj}(R) = \text{Proj}(R')$ und $R \equiv_L^p R'$, wobei R' noch einigermaßen natürlich bleibt.

Nun können wir, wie oben schon angedeutet, *joinable*, *couplable* und *building block* definieren.

Definition 3.27 (Universelle Relation; Agrawal und Biswas 1992). Sei R eine strenge NP-Relation mit Zertifikatsschranke q . Die Relation R ist *universell* wenn alle drei folgenden Eigenschaften erfüllt sind:

- (1) Die Relation R hat einen *building block*: es gibt ein Element $block \in \text{Proj}(R)$, sowie paarweise verschiedene $b_1, b_2, b_3 \in \mathbb{N}$ sodass

$$\{y[b_1, b_2, b_3] \mid y \in \text{set-}R(block)\} = \Sigma^3 - \{000\}$$

- (2) Die Relation R ist *joinable*: es gibt eine Funktion $join \in \text{FP}$ sodass $join(x_1, \dots, x_n) = (z, \alpha)$, wobei $x_1, \dots, x_n, z \in \Sigma^*$, $\sum_{k=1}^n q(|x_k|) = |\alpha| \leq q(|z|)$, und wobei $\alpha \in \mathbb{N}^*$ eine Sequenz von paarweise verschiedenen Indizes $< q(|z|)$ ist, und

$$\{y'[\alpha] \mid y' \in \text{set-}R(z)\} = \{y_1 \circ y_2 \circ \dots \circ y_n \mid (\forall k \leq n). y_k \in \text{set-}R(x_k)\}.$$

- (3) Die Relation R ist *couplable*: es gibt eine Funktion $cpl \in \text{FP}$ sodass

$$cpl(x, (i_1, \dots, i_n), (j_1, \dots, j_n)) = (z, \alpha)$$

wobei $x \in \Sigma^*$, $i_1, \dots, i_n, j_1, \dots, j_n \leq q(|x|) - 1$ und $|\alpha| = q(|x|)$, und wobei wieder $\alpha \in \mathbb{N}^*$ eine Sequenz von paarweise verschiedenen Indizes $< q(|z|)$ ist, und

$$\{y'[\alpha] \mid y' \in \text{set-}R(z)\} = \{y \mid y \in \text{set-}R(x) \text{ und } (\forall k \leq n)(y[i_k] \neq y[j_k])\}. \quad \triangleleft$$

Wir illustrieren diese Definition anhand von rSAT . Wir nehmen dafür an, dass rSAT als strenge NP-Relation gegeben ist. Die Formeln φ seien so codiert, dass in φ nur die Variablen

$x_0, \dots, x_{|\varphi|-1}$ vorkommen. Zertifikate für φ sind dann Strings w der Form $\Sigma^{|\varphi|}$ sodass φ durch die Variablenbelegung $x_0 = w[0], x_1 = w[1], \dots$ erfüllt wird. Beachte, dass wenn x_j nicht in φ vorkommt, dass dann $(\varphi, w) \in \text{rSAT}$ genau dann wenn $(\varphi, w') \in \text{rSAT}$, wobei w und w' sich nur in der j -ten Stelle unterscheiden.

Der *building block* ist im Wesentlichen dazu da, frische Klauseln einzuführen. Definiere zum Beispiel *block* als $(x_0 \vee x_1 \vee x_2)$. Wir haben dann z.B. $y = 001000 \dots \in \text{set-rSAT}(\varphi)$, welches der Variablenbelegung $x_0 = 0, x_1 = 0, x_2 = 1, x_3 = 0, \dots$ entspricht. Mit der Projektion über $b_1 = 0, b_2 = 1, b_3 = 2$ erhalten wir dann die Menge $\{y[b_1, b_2, b_3] \mid y \in \text{set-rSAT}(\varphi)\} = \Sigma^3 - \{000\}$. Das Wort 000 fehlt, denn (erweitert auf ein Zertifikat) erfüllt diese Belegung nicht φ .

Die Funktion $\text{join}(\varphi_1, \varphi_2, \dots) = (\psi, \alpha)$ definieren wir als die Konjunktion $\psi = \varphi_1 \wedge \varphi_2 \wedge \dots$ zusammen mit der Sequenz α . Hierbei nummerieren wir dabei aber die Variablen davor um sodass die Variablen in φ_i und φ_j disjunkt sind. Die Sequenz α reflektiert diese Umbenennung, sodass aus einem Zertifikat y für ψ die entsprechenden erfüllende Belegungen für jedes φ_i zusammensetzt.

Die Funktion $\text{cpl}(\varphi, (i), (j))$ sorgt dafür, dass zwei Variablen x_i, x_j „gekoppelt“ werden in dem Sinn, dass, unter allen erfüllenden Belegungen, $x_i \neq x_j$ gilt. Wir können das umsetzen indem wir $\text{cpl}(\varphi, (i), (j)) = (\psi, \beta)$ setzen mit $\psi = \varphi \wedge (x_i \vee x_j) \wedge (\neg x_i \vee \neg x_j)$. Die Sequenz β muss dann nur die $q(|\varphi|)$ ersten Bits eines Zertifikats auslesen; der Suffix ist redundant und entsteht nur weil ψ (und damit die Zertifikate für ψ) länger als φ ist. Es lässt sich leicht cpl auf längere Eingabesequenzen erweitern.

Mit dem *building block*, der Funktion join und der Funktion cpl ist es nun möglich, beliebige 3SAT-Formeln zusammenzubauen. Das gilt nicht nur für die hier gewählten Definition konkret für rSAT , sondern sogar für alle universellen NP-Relationen. Ferner ist es sogar möglich, aus den Zertifikaten wieder eine erfüllende Belegung für die 3SAT-Formel projektiv „herauszulesen“. Intuitiv wird damit ersichtlich, dass die universellen NP-Relationen auch \leq_L^P -vollständig sind. Die spezielle starke Form der Reduzierbarkeit, welche das einfache „Auslesen“ aus den Zertifikaten zulässt, formalisieren Agrawal und Biswas wie folgt:

Definition 3.28 (Projektive Levin-Reduktion). Seien Q und R zwei strenge NP-Relationen. Die NP-Relation Q lässt sich über eine *projektive Levin-Reduktion* auf R reduzieren, wenn eine Funktion $f \in \text{FP}$ existiert, welche die folgenden Bedingungen erfüllen:

- (1) $f(x) = (z, \alpha)$ wobei $x, z \in \Sigma^*$ und $\alpha \in \mathbb{N}_{>0}^{q(|x|)}$ ist eine Sequenz von positiven paarweise verschiedenen Indizes der Länge $q(|x|)$.
- (2) $\{y[\alpha] \mid y \in \text{set-}R(z)\} = \text{set-}Q(x)$. \triangleleft

Es ist leicht zu sehen, dass aus einer projektiven Levin-Reduktion von Q auf R auch $Q \leq_L^P R$ folgt.

Abschließend können universelle NP-Relationen als harte Relationen bezüglich projektiver Levin-Reduktion charakterisiert werden:

Satz 3.29 (Agrawal und Biswas 1992). *Sei R eine strenge NP-Relation. Folgende Aussagen sind äquivalent:*

- (1) R ist eine universelle Relation, i.e. hat einen building block, ist joinable und ist couplable.
- (2) jede NP-Relation lässt sich mittels einer projektiven Levin-Reduktion auf R reduzieren.

Diese Äquivalenz gilt nur im unrelativierten Fall.

Agrawal und Biswas (1992) haben explizit verifiziert, dass rSAT , rHAMCYCLE , rINDSET , rKNAPSACK , rMAXCUT alle universell sind.

Da aus projektiver Levin-Reduktion auch Levin-Reduktion folgt, sehen wir dass Universalität eine Verstärkung von \leq_L^P -Vollständigkeit ist.

Korollar 3.30. *Wenn R eine universelle NP-Relation ist, dann ist $R \leq_L^P$ -vollständig für FNP. Diese Aussage gilt nur im unrelativierten Fall.*

TODO: Erkläre warum die Reduktion/join/cpl mit Projektionen arbeitet. An dieser Stelle sei auch noch einmal darauf hingewiesen, dass Agrawal und Biswas (1992) das Ziel verfolgten, eine gemeinsame Struktur aller natürlichen NP-vollständigen Probleme zu erfassen. Im Speziellen vermuten sie, dass tatsächlich für alle natürlichen NP-vollständigen Entscheidungsprobleme L auch mindestens eine NP-Relation R für L existiert, die „hinreichend“ natürlich ist. Sie machen das an zwei Gründen fest: zum einen die intuitive Einsicht, dass *joinability* und *couplability* sehr natürlich wirkende strukturelle Eigenschaften sind. Zum anderen anhand empirischer Belege: In ihrer Arbeit präsentierten sie fünf konkrete Beispiele aus verschiedenen Problemgebieten, darunter Logik (Erfüllbarkeit), Zahlentheorie (Knapsack)

4 Suchprobleme und die Hypothese Q im Kontext des Pudlák'schen Programms

- Will Q in den Pudlák-Baum einordnen: dafür ist es notwendig, diese ordentlich zu relativieren. Insb. will ich zeigen, dass einige bisherige Resultate natürlicherweise auf „Standardbeweissysteme“ vollständiger Mengen übertragen (nicht nur das Standardbeweissystem für SAT).

Definition 4.1 (Levin-Paddability). Eine NP-Relation R ist *Levin-paddable* wenn Funktionen $pad \in \text{FP}$ und $padsol \in \text{FP}$ existieren, sowie ein Polynom r sodass

- (1) $x \in \text{Proj}(R) \iff pad(x, 1^n) \in \text{Proj}(R)$,
- (2) $(pad(x, 1^n), y) \in R \implies (x, padsol(x, 1^n, y)) \in R$,
- (3) $r(|pad(x, 1^n)|) \geq n$. (Funktion pad ist ehrlich bzgl. der zweiten Komponente.) \triangleleft

Definition 4.2 (Standardbeweissystem). Sei R eine NP-Relation. Wir definieren bezüglich R das *Standardbeweissystem* std_R für $\text{Proj}(R)$ wie folgt:

$$std_R(w) = \begin{cases} x & \text{wenn } w = (x, y) \text{ und } (x, y) \in R, \\ \perp & \text{sonst.} \end{cases} \quad \triangleleft$$

Beobachtung 4.3. Für jede NP-Relation R ist das Standardbeweissystem std_R ehrlich.

Beweis. Sei q die Zertifikatsschranke von R , und sei $w = (x, y)$ gegeben sodass $std_R(x, y) = x$. An dieser Stelle müssen wir auf die konkrete Codierung von Beweisen $w = (x, y)$ eingehen. Wie in ?? beschrieben, codieren wir Tupel in einer solchen Weise sodass

$$|w| = |(x, y)| = 2(|x| + |y| + 2) = 2|x| + 2|y| + 4.$$

Da $(x, y) \in R$ gilt für y auch $|y| \leq q(|x|)$. Damit also

$$|w| \leq 2|x| + 2q(|x|) + 4 \leq q'(|x|) = q'(|std_R(w)|).$$

für ein geeignetes Polynom q' , wie gewünscht. \square

Folgendes Lemma ist eine Generalisierung von Messner (2000, Thm. 5.2)

Lemma 4.4. Sei R eine NP-Relation die Levin-paddable ist. Folgende Aussagen sind äquivalent:

- (1) Das Standardbeweissystem std_R bzgl. R ist p -optimal.
- (2) Für alle NTM N (ohne Laufzeitbeschränkung) mit $L(N) = \text{Proj}(R)$ lassen sich akzeptierende Rechenwege von N in Zertifikate umrechnen: es existiert eine Funktion $h \in \text{FP}$ sodass

$$N(x) \text{ akz. mit Rechenweg } \alpha \implies (x, h(x, \alpha)) \in R.$$

- (3) Für alle NPTM N mit $L(N) = \text{Proj}(R)$ lassen sich akzeptierende Rechenwege von N in Zertifikate umrechnen: es existiert eine Funktion $h \in \text{FP}$ sodass

$$N(x) \text{ akz. mit Rechenweg } \alpha \implies (x, h(x, \alpha)) \in R.$$

Beweis. (1) \implies (2): Für NTM f_N können wir das assoziierte Beweissystem

$$f_N(x, \alpha) = \begin{cases} x & N(x) \text{ akz. mit Rechenweg } \alpha \\ \perp & \text{sonst} \end{cases}$$

angeben. Es ist klar, dass f_N ein Beweissystem für $\text{Proj}(R)$ ist. Aus der p -Optimalität von std_R gilt nun $std_R \leq^p f_N$, bzw. p -simuliert das Standardbeweissystem das Beweissystem f_N . Damit existiert eine Funktion $h \in \text{FP}$ sodass

$$std_R(h(x, \alpha)) = f_N(x, \alpha).$$

Diese Funktion h erfüllt genau die Eigenschaften von (2): Wir haben

$$\begin{aligned} N(x) \text{ akz. mit Rechenweg } \alpha &\implies h(x, \alpha) = x \\ &\implies std_R(h(x, \alpha)) = x \\ &\implies (x, h(x, \alpha)) \in R, \end{aligned}$$

wie gewünscht.

(2) \implies (3): Klar.

(3) \implies (1): Angenommen (3) gilt. Seien pad , $padsol$ die entsprechenden Funktionen, welche die Levin-Paddability von R realisieren. Das Polynom r sei so gewählt dass $r(|pad(x, 1^n)|) \geq n$ (vgl. 4.1(3)).

Wir wollen nun zeigen, dass std_R auch p-optimal ist. Sei hierfür f ein beliebiges Beweissystem für $\text{Proj}(R)$. Wir zeigen nun, dass $std_R \leq^p f$. Seien pad , $padsol$ die entsprechenden Padding-Funktionen von R . Definiere nun

$$f'(w) = \begin{cases} pad(x, 1^{|w|}) & \text{falls } w = 1z \text{ und } f(z) = x, \\ x & \text{falls } w = 0z \text{ und } std_R(z) = x, \\ \perp & \text{sonst.} \end{cases}$$

Es ist leicht zu sehen, dass f' ehrlich ist: es ist ehrlich für Eingaben $0z$, denn das Standardbeweissystem std_R ist ehrlich nach Beobachtung 4.3. Es ist ehrlich für Eingaben $w = 1z$, denn

$$|1z| = |w| \leq r(\underbrace{|pad(x, 1^{|w|})|}_{f'(1z)}) = r(|f'(|w|)|).$$

Sei im Folgenden dann das Polynom r' so gewählt, dass $|w| \leq r'(|f'(w)|)$ gilt.

Definiere nun die NPTM $N_{f'}$ welche auf Eingabe x erst nichtdeterministisch einen Beweis w , $|w| \leq r'(|x|)$ rät, und genau dann akzeptiert falls $f'(w) = x$. Es ist klar, dass $L(N_{f'}) = \text{Proj}(R)$. Nach Voraussetzung (3) gibt es also nun eine Funktion $h \in \text{FP}$ sodass

$$N_{f'}(x) \text{ akz. mit Rechenweg } \alpha \implies (x, h(x, \alpha)) \in R. \quad (4.1)$$

Jetzt können wir $std_R \leq^p f$ zeigen: sei z ein f -Beweis für x , d.h. $f(z) = x$. Wir wissen, dass $f'(1z) = pad(x, 1^{|1z|}) = x'$. Daher können wir aus z einen Rechenweg α_z konstruieren, sodass $N_{f'}(x')$ akzeptiert, nämlich jener der den f' -Beweis $1z$ rät. Die Abbildung $z \mapsto \alpha_z$ lässt sich in Polynomialzeit leisten.

Nun gilt

$$\begin{aligned} N_{f'}(x') \text{ akz. mit } \alpha_z &\implies (x', \underbrace{h(x', \alpha_z)}_{y'}) \in R \text{ nach (4.1)} \\ &\implies (pad(x, 1^{|1z|}), y') \in R \text{ mit } y' = h(x', \alpha_z) \text{ und obiger Def. von } x' \\ &\implies (x, \underbrace{padsol(x, 1^{|1z|}, y')}_{y}) \in R \\ &\implies std(x, y) = x \text{ mit } y = padsol(x, 1^{|1z|}, y') \end{aligned}$$

und wir haben aus dem f -Beweis z für x einen std_R -Beweis (x, y) für x bestimmt. Es ist klar, dass die Übersetzung $z \mapsto (x, y)$ in Polynomialzeit möglich ist. \square

Folgendes Lemma ist eine Generalisierung von Fenner u. a. (2003, Thm. 2)

Lemma 4.5. *Sei R eine \leq_L^p -vollständige NP-Relation, mit der zusätzlichen Eigenschaft dass für die jeweilige entsprechende Problem-Reduktionsfunktion $f: Q \rightarrow R$ für $Q \leq_L^p R$ immer gilt, dass f ehrlich ist. Folgende Aussagen sind äquivalent:*

- (1) *Für alle NPTM N mit $L(N) = \text{Proj}(R)$ lassen sich akzeptierende Rechenwege von N in Zertifikate umrechnen: es existiert eine Funktion $h \in \text{FP}$ sodass*

$$N(x) \text{ akz. mit Rechenweg } \alpha \implies (x, h(x, \alpha)) \in R.$$

- (2) *Für alle NPTM N mit $L(N) = \Sigma^*$ lassen sich aus Eingabe x Rechenwege von $N(x)$ effizient bestimmen: es existiert $r \in \text{FP}$ sodass $N(x)$ auf Rechenweg $r(x)$ akzeptiert. (Das ist die Aussage Q.)*

Beweis. (2) \implies (1): Sei R eine beliebige NP-Relation mit Zertifikatsschranke q , und sei N eine beliebige NPTM mit $L(N) = \text{Proj}(R)$. Definiere nun die NPTM $N'(w)$ wie folgt:

- 1 **wenn** w nicht von der Form (x, α) **dann** akzeptiere
- 2 $(x, \alpha) \leftarrow w$
- 3 **wenn** $N(x)$ akzeptiert **nicht** auf Rechenweg α **dann** akzeptiere
- 4 **sonst**
 - 5 Rate nichtdeterministisch $y \in \Sigma^{\leq q(|x|)}$
 - 6 Akzeptiere genau dann wenn $(x, y) \in A$.

Es ist nun leicht zu sehen dass $L(N') = \Sigma^*$. Nach Voraussetzung (2) existiert eine Funktion $r \in \text{FP}$ sodass für alle x die Maschine $N'(w)$ auf Rechenweg $r(w)$ akzeptiert. Nun gilt

$$\begin{aligned} N(x) \text{ akz. mit Rechenweg } \alpha &\implies N'(x, \alpha) \text{ akz. in Z. 6} \\ &\implies N'(x, \alpha) \text{ akz. mit Rechenweg } r(x, \alpha) \text{ in Z. 6,} \end{aligned}$$

und aus diesem Rechenweg $r(x, \alpha)$ kann effizient der geratene Zeuge $y \in \text{set-}R(x)$ aus Z. 5 ausgelesen werden. Da $r \in \text{FP}$ existiert also auch ein $h \in \text{FP}$ sodass $h(x, \alpha)$ genau diesen geratenen Zeugen y berechnet. Wir haben dann also

$$\implies (x, h(x, \alpha)) \in R,$$

wie gewünscht.

(1) \implies (2): Sei R eine \leq_L^p -vollständige NP-Relation unter ehrlichen Problem-Reduktionsfunktionen, und Zertifikatsschranke p . Sei nun N eine NPTM mit $L(N) = \Sigma^*$. Betrachte die entsprechende NP-Relation

$$R_N = \{(x, \alpha) \mid N(x) \text{ akz. mit Rechenweg } \alpha\}$$

Da R ja vollständig ist, gilt $R_N \leq_L^p R$ via $f, g \in \text{FP}$ und (nach Voraussetzung) ist f ehrlich; es existiert ein Polynom q sodass $q(|f(x)|) \geq |x|$.

Definiere nun die folgende NPTM $N'(w)$:

- 1 Rate nichtdeterministisch $x \in \Sigma^{\leq q(|w|)}$
- 2 **wenn** $f(x) = w$ **dann** akzeptiere
(Ab hier kann man x wegwerfen)
- 3 Rate nichtdeterministisch $y \in \Sigma^{\leq p(|w|)}$
- 4 Akzeptiere genau dann wenn $(w, y) \in R$.

Wir zeigen nun, dass $L(N') = \text{Proj}(R)$. Wir müssen hierfür nur die Fälle betrachten, wenn $N'(w)$ in Z. 2 akzeptiert. In diesem Fall gilt $f(x) = w$, und wir haben

$$x \in \Sigma^* \implies x \in \text{Proj}(R_N) \implies f(x) \in \text{Proj}(R) \implies w \in \text{Proj}(R),$$

wie gewünscht.

Nach Voraussetzung (1) gilt nun also, dass eine Funktion $h \in \text{FP}$ existiert sodass

$$N'(w) \text{ akz. mit Rechenweg } \alpha \implies (w, h(w, \alpha)) \in R.$$

Beobachte wie für $N'(f(x))$ immer ein trivialer akzeptierender Rechenweg α_x existiert: nämlich jener, welcher in Z. 1 das Urbild x rät. Beobachte dass die Umformung $x \mapsto \alpha_x$ in Polynomialzeit möglich ist.

Um nun (2) zu zeigen müssen wir aus $x \in \Sigma^*$ effizient einen akzeptierenden Rechenweg für N bestimmen. Wir haben

$$\begin{aligned} N'(f(x)) \text{ akz. mit Rechenweg } \alpha_x &\implies (f(x), h(f(x), \alpha_x)) \in R \\ &\implies (x, \underbrace{g(h(f(x), \alpha_x))}_{r(x)}) \in R_N \text{ nach Translationsfunktion } g \\ &\implies N(x) \text{ akz. mit Rechenweg } r(x) \end{aligned}$$

mit $r \in \text{FP}$, $r(x) = g(h(f(x), \alpha_x))$, wie gewünscht. \square

Lemma 4.6. Die in Lemma 4.4 und 4.5 genannten Voraussetzungen an die NP-Relation R werden von allen solchen R erfüllt, die \leq_L^p -vollständig sind und Levin-paddable sind.

Beweis. Es ist sofort klar, dass R die Voraussetzungen von Lemma 4.4 erfüllt. Es bleibt nur zu zeigen, dass für jede NP-Relation Q eine \leq_L^p -Reduktion angegeben werden kann, bei dem die Problem-Reduktionsfunktion ehrlich ist. Wir nutzen hierbei aus, dass R eine Levin-paddable Relation ist.

Nachdem R vollständig ist, gilt $Q \leq_L^p R$; sei $f, g \in \text{FP}$ die Reduktions- bzw. Translationsfunktion welche diese Reduktion realisieren. Wir werden nun Funktionen $f', g' \in \text{FP}$ angeben, welche die gleiche Reduktion realisieren, aber f' ehrlich, wie gewünscht.

Sei $\text{pad}, \text{padsol}$ die zu R zugehörigen Padding-Funktionen. Definiere

$$f'(x) = \text{pad}(f(x), 1^{|x|}).$$

Es gilt

$$x \in \text{Proj}(Q) \iff f(x) \in \text{Proj}(R) \iff \text{pad}(f(x), 1^{|x|}) = f'(x) \in \text{Proj}(R),$$

wobei erste Implikation die Eigenschaft der Reduktionsfunktion f ist, und die zweite aus der Definition von Levin-Paddability folgt. Aus der Definition von Levin-Paddability folgt auch $r(|f'(x)|) \geq |x|$ für ein geeignetes Polynom r , und damit ist auch f' ehrlich.

Definiere

$$g'(x, z) = g(x, \text{padsol}(f(x), 1^{|x|}, z)).$$

Sei nun $(f'(x), z) \in R$. Die Funktion g' berechnet nun ein Zertifikat y für x : Wir haben $(\text{pad}(f(x), 1^{|x|}), z) \in R$, also gilt nach Levin-Paddability dass

$$(f(x), \text{padsol}(f(x), 1^{|x|}, z)) \in R,$$

und nach Definition der Translationsfunktion g gilt dann

$$(x, g(x, \text{padsol}(f(x), 1^{|x|}, z))) \in Q,$$

und das ist genau $(x, g'(x, z)) \in Q$, wie gewünscht. \square

4.1 Welche Suchprobleme sind paddable?

Beobachtung 4.7. Die kanonische Levin-vollständige NP-Relation rKAN ist Levin-paddable.

Beobachtung 4.8. (1) Gilt $\text{rKAN} \leq_L^p R$, und ist die zugehörige Reduktionsfunktion f ehrlich, dann ist R Levin-paddable

(2) Jede $\leq_{L, \text{inv}}^p$ -vollständige NP-Relation R ist auch Levin-paddable.

Korollar 4.9. Jede $\leq_{L, \text{inv}}^p$ -vollständige Relation R erfüllt die in Lemma 4.4 und 4.5 genannten Voraussetzungen an die NP-Relation R .

Das sind im unrelativierten Fall u.a. rSAT , rSETCOVER , rVERTEXCOVER , rCLIQUE , r3COLORABILITY .

Wichtig: Ehrlichkeit ist hier notwendig, denn wir *müssen* das Urbild raten, und können das nicht als zweite Eingabe mitführen, damit $L(N') = \text{Proj}(R)$. Auch alternative Wege über Beweissysteme laufen darauf hinaus, Levin-Paddability voraussetzen, woraus ohnehin wieder Ehrlichkeit folgt.

So das Textbook von Goldreich (2008).

Beweis zu Beobachtung 4.8. Aussage (2) folgt unmittelbar aus (1): Wir haben $\text{rKAN} \leq_{L, \text{inv}}^p R$ und damit ist die entsprechende Reduktionsfunktion f p-invertierbar, und damit ehrlich.

Für (1) nutzen wir die Levin-Paddability von rKAN aus: übersetze Instanz x von R nach rKAN , padde dort hoch, und überetze zu R -Instanz x' zurück. Ist dann y' ein Zertifikat für x' , dann lässt sich dies auf ähnlichem Weg wieder zu einem Zertifikat für x zurückrechnen.

Seien f, g die Reduktions- bzw. Translationsfunktion, welche $\text{rKAN} \leq_L^p R$ bezeugen, und seien analog f', g' jene Funktionen, welche $R \leq_L^p \text{rKAN}$ bezeugen. Erstere existieren nach Voraussetzung, letztere existieren weil $\text{rKAN} \leq_L^p$ -vollständig ist. Nach Voraussetzung ist f ehrlich. Und nach Beobachtung 4.7 existieren für rKAN Padding-Funktionen $\text{pad}_{\text{rKAN}}, \text{padsol}_{\text{rKAN}}$. Sei q ein entsprechendes Polynom mit $q(|\text{pad}_{\text{rKAN}}(x, 1^n)|) \geq n$, $q(|f(x)|) \geq |x|$.
Definiere nun

$$\text{pad}_R(x, 1^n) = f(\text{pad}_{\text{rKAN}}(f'(x), 1^n)).$$

Die Zugehörigkeit zu $\text{Proj}(R)$ bleibt erhalten:

$$\begin{aligned} x \in \text{Proj}(R) &\iff f'(x) \in \text{KAN} \iff \text{pad}_{\text{rKAN}}(f'(x), 1^n) \in \text{KAN} \\ &\iff f(\text{pad}_{\text{rKAN}}(f'(x), 1^n)) \in \text{Proj}(R) \iff \text{pad}_R(x, 1^n) \in \text{Proj}(R). \end{aligned}$$

Ferner gilt

$$\begin{aligned} &q(q(|\text{pad}_R(x, 1^n)|)) \\ &= q(q(|f(\text{pad}_{\text{rKAN}}(f'(x), 1^n)|))) \\ &\geq q(|\text{pad}_{\text{rKAN}}(f'(x), 1^n)|) \\ &\geq n. \end{aligned}$$

und damit ist pad_R wie gewünscht ehrlich bzgl. n (mit Polynom $q \circ q$).

Es verbleibt noch die Funktion padsol_R . Nehme hierfür an dass wir ein y' haben mit $(\text{pad}_R(x, 1^n), y') \in R$. Wir können über g, g' das Zertifikat y' zu Zertifikat y mit $(x, y) \in R$ zurück übersetzen: Sei $p = \text{pad}_{\text{rKAN}}(f'(x), 1^n)$, dann gilt

$$(f(p), y') \in R \implies (p, \underbrace{g(p, y')}_{z}) \in \text{rKAN}.$$

Definiere $z = g(p, y')$. Nun haben wir

$$\begin{aligned} (p, z) &= (\text{pad}_{\text{rKAN}}(f'(x), 1^n), z) \in \text{rKAN} \\ &\implies (f'(x), \underbrace{\text{padsol}_{\text{rKAN}}(f'(x), 1^n, z)}_{z'}) \in \text{rKAN} \end{aligned}$$

und mit $z' = \text{padsol}_{\text{rKAN}}(f'(x), 1^n, z)$ gilt

$$(f'(x), z') \in \text{rKAN} \implies (x, \underbrace{g'(x, z')}_{y}) \in R.$$

Es ist leicht zu sehen, dass sich eine Funktion $\text{padsol}_R \in \text{FP}$ angeben kann, die aus $x, 1^n$ dieses entsprechende y berechnen kann. \square

Beobachtung 4.10. Jede NP-Relation mit einem building block und die joinable ist, ist auch Levin-paddable.

Korollar 4.11. Im unrelativierten Fall erfüllt jede universelle Relation R die in Lemma 4.4 und 4.5 genannten Voraussetzungen an die NP-Relation R .

Das sind u.a. rSAT , rHAM , rINDSET , rKNAPSACK , rMAXCUT .

Beweis zu Beobachtung 4.10. Sei R eine NP-Relation, mit zugehörigem Polynom q , welches die Zertifikatsgröße spezifiziert. Zur Erinnerung, dieses Polynom ist streng monoton steigend, und aus $(x, y) \in R$ folgt $|y| = q(|x|)$. Wir zeigen zunächst, wie wir für beliebige Instanz x und $n \in \mathbb{N}$ auf eine Instanz x' hochpadding, in dem Sinne dass $q(|x'|) \geq n$.

Nach Voraussetzung hat die Relation R einen building block. Es lässt sich leicht aus der Definition eines building block ableiten, dass $|block| > 0$ und $block \in \text{Proj}(R)$. Damit gilt auch dass die Zertifikate y zu $block$ die Länge $l = q(|block|) \geq |block| \geq 1$ haben.

Nach Voraussetzungen ist die Relation R auch joinable, das heißt wir haben eine Funktion $join \in \text{FP}$. Sei

$$(x', \delta) = \text{join}(x, \underbrace{block, block, \dots, block}_{n \text{ mal}}).$$

Wir werden nun über die Länge $|\delta|$ auf die Länge von Zertifikaten zu x' schließen, und damit $|x'|$ beschränken. Nach Definition ?? gilt

$$|\delta| = q(|x|) + q(n) \cdot q(|block|) = q(|x|) + q(n) \cdot l \geq n.$$

Beob. dass unter Definition ?? alle Zertifikate y' für x' die feste Länge $q(|x'|)$ haben. Zur Erinnerung: wir haben

$$\{y'[\delta] \mid y' \in \Sigma^{q(|x'|)}, (x', y') \in R\} = \{y_1 y_2 \dots y_n \mid y \in \Sigma^{q(|x|)}, y_1, y_2, \dots \in \Sigma^l, (x, y), (block, y_1), (block, y_2), \dots \in R\} \quad (4.2)$$

Die Sequenz δ besteht nach Definition aus paarweise verschiedenen Indizes, daher können wir argumentieren, dass auch alle Zertifikate y' (mit vorgegebener Länge $q(|x'|)$) mindestens die Länge $|\delta|$ haben. Damit gilt

$$q(|x'|) \geq |\delta| \geq n$$

wie gewünscht.

Sei nun pad genau jene polynomialzeit-berechenbare Funktion, die aus x und 1^n die Instanz x' konstruiert:

$$\text{pad}(x, 1^n) = x' \quad \text{wobei } (x', \delta) = \text{join}(x, \underbrace{block, block, \dots, block}_{q(n) \text{ mal}}).$$

Dann gilt schon sofort, dass $q(|\text{pad}(x, 1^n)|) = q(|x'|) \geq n$ wie gewünscht.

Wir zeigen jetzt, dass die Zugehörigkeit zu $\text{Proj}(R)$ erhalten bleibt: Gilt $x \notin \text{Proj}(R)$, dann ist die rechte Menge in (4.2) leer, also auch die linke Menge und damit $x' = \text{pad}(x, 1^n) \notin \text{Proj}(R)$. Falls anders herum $x \in \text{Proj}(R)$,

Es sieht nicht danach aus, dass wir über die Anzahl an Zertifikaten für x' die Länge abschätzen können: ist $x \notin \text{Proj}(R)$ dann hätte auch x' keine Zertifikate, und könnte sich erlauben entsprechend kurz zu sein.

dann ist die rechte Menge nicht leer, existiert ja ein Zertifikat y für x und je ein weiteres y_i für $block$. Also ist auch die linke Menge nicht leer, damit $pad(x, 1^n) \in \text{Proj}(R)$.

Die noch verbleibende Funktion $padsol$ ist durch die bitweise Projektion durch δ leicht möglich:

$$padsol(x, 1^n, y') = y'[\delta[1 : q(|x|)]] \quad \text{wobei } (\cdot, \delta) = \text{join}(x, \underbrace{block, block, \dots, block}_{n \text{ mal}}).$$

Wir verifizieren: Sei $(pad(x, 1^n), y') \in R$, dann ist nach (4.2) $y'[\delta] = yy_1y_2 \dots$ wobei $y \in \Sigma^{q(|x|)}$, $(x, y) \in R$. Wir haben

$$padsol(x, 1^n, y') = y'[\delta[1 : q(|x|)]] = (yy_1y_2 \dots)[1 : q(|x|)] = y$$

und damit $(x, padsol(x, 1^n, y')) = (x, y) \in R$, wie gewünscht. \square

4.2 Hypothese Q und die Vollständigkeit von Suchproblemen

Satz 4.12 (Äquivalente Formulierungen der Hypothese Q; Fenner u. a. 2003; Messner 2000).
Folgende Aussagen sind äquivalent:

- (1) *Hypothese Q: Für jede NPTM N mit $L(N) = \Sigma^*$ existiert eine Funktion $g \in \text{FP}$ sodass für alle x das Bild $g(x)$ eine akzeptierende Berechnung von $N(x)$ ist.*
- (2) $\text{NPMV}_t \subseteq_c \text{FP}$
- (3) $\text{P} = \text{NP} \cap \text{coNP}$ und $\text{NPMV}_t \subseteq_c \text{NPSV}_t$
- (4) *Jede surjektive ehrliche Funktion $f \in \text{FP}$ ist p -invertierbar.*
- (5) *Für jede Menge $L \in \text{P}$ und jede NPTM N mit $L(N) = L$ existiert eine Funktion $h \in \text{FP}$ mit*

$$x \in L \implies N(x) \text{ akz. mit Rechenweg } h(x).$$

- (6) *Für jedes Paar von NP-Relationen A, B und jede Funktion $f \in \text{FP}$ gilt:*

$$\text{Proj}(A) \leq_m^p \text{Proj}(B) \text{ via } f \iff A \leq_L^p B \text{ via Reduktionsfunktion } f.$$

- (7) *Für jedes Beweissystem h gilt: h ist optimal $\iff h$ ist p -optimal.*
- (8) *Es existiert eine \leq_L^p -vollständige Levin-paddable NP-Relation R sodass für alle NPTM N mit $L(N) = \text{Proj}(R)$ gilt: es existiert eine Funktion $h \in \text{FP}$ mit*

$$N(x) \text{ akz. mit Rechenweg } \alpha \implies (x, h(x, \alpha)) \in R.$$

- (9) *Es existiert eine \leq_L^p -vollständige Levin-paddable NP-Relation R für welche das Standardbeweissystem std_R p -optimal ist.*
- (10) *Es existiert eine $\leq_{L,1,\text{inv}}^p$ -vollständige NP-Relation R sodass für jede Menge $S \in \text{P}$ mit $S \subseteq \text{Proj}(R)$ gilt: es existiert eine Funktion $g \in \text{FP}$ sodass*

$$x \in S \implies (x, g(x)) \in R.$$

Beweis. 1. (1) \iff (2) \iff (3) \iff (4) \iff (5): nach Fenner u. a. (2003, Thm. 2).

2. (1) \iff (8) \iff (9): nach Lemma 4.5 und 4.4.

3. (5) \implies (10): Wir zeigen eine stärkere Variante von (10), welche sich über *alle* NP-Relationen R erstreckt (und damit auch über $\leq_{L,1,\text{inv}}^p$ -vollständige rKAN , wie von (10) gefordert). Sei R eine beliebige NP-Relation, wobei Polynom q die Zertifikatsgröße beschränkt. Sei nun $S \subseteq \text{Proj}(R)$ mit $S \in \text{P}$. Definiere die NPTM N , welche auf Eingabe x folgendes leistet: teste zuerst ob $x \in S$; falls nicht, lehne sofort ab. Rate dann ein $y \in \Sigma^{\leq q(|x|)}$ und akzeptiere genau dann wenn $(x, y) \in R$.

Klar ist, dass $L(N) = S$. Nach (5) existiert nun eine Funktion $h \in \text{FP}$, die für $x \in S$ einen akzeptierenden Rechenweg $h(x)$ von $N(x)$ ausgibt. Wir können sogar aus $h(x)$ das geratene Zertifikat y extrahieren. Es ist daher leicht eine Funktion $g \in \text{FP}$ anzugeben für die $(x, g(x)) \in R$ für alle $x \in S$.

4. (10) \implies (5): Sei $L \in \text{P}$ und sei N eine NPTM mit $L(N) = L$, wobei das Polynom q die Laufzeit beschränkt. Wir wollen eine Funktion $h \in \text{FP}$ definieren sodass $h(x)$ ein akzeptierender Rechenweg von $N(x)$ für $x \in L$ ist. Definiere die NP-Relation

$$Q = \{(x, y) \mid N(x) \text{ akzeptiert mit Rechenweg } y \in \Sigma^{\leq q(|x|)}\}.$$

Nachdem (10) gilt, haben wir eine $\leq_{L,1,\text{inv}}^p$ -vollständige NP-Relation R . Damit gilt $Q \leq_L^p R$ mittels Reduktions- bzw. Translationsfunktion $f, k \in \text{FP}$. Insbesondere existiert eine Inverse $f^{-1} \in \text{FP}$ zu f .

Sei $S = f(L)$ die Bildmenge der Elemente aus L , also

$$S = \{f(x) \mid x \in L\}.$$

Es ist leicht zu sehen dass $S \subseteq \text{Proj}(R)$. Außerdem ist $S \in \text{P}$: teste $z \in S$ indem getestet wird ob $f^{-1}(z) = x \neq \perp$ und ob $x \in L$.

Damit sind die Voraussetzungen von (10) erfüllt, und es existiert eine Funktion $g \in \text{FP}$ sodass $(z, g(z)) \in R$ für

alle $z \in S$. Damit gilt

$$\begin{aligned} x \in L &\implies f(x) \in S \implies (f(N, x), g(f(x))) \in R \\ &\implies ((x), k(g(f(x)))) \in Q \\ &\implies N(x) \text{ akz. mit Rechenweg } \underbrace{k(g(f(x)))}_{h(x)}. \end{aligned}$$

Definiere nun die gesuchte Funktion $h \in \text{FP}$ mit $h(x) = k(g(f(x)))$. Damit gilt für alle $x \in L$ dass $N(x)$ mit Rechenweg $h(x)$ akzeptiert, wie gewünscht.

5. (1) \implies (6): Die Richtung von rechts nach links ist klar. Für die andere Richtung sei $\text{Proj}(A) \leq_m^p \text{Proj}(B)$ mit A, B NP-Relationen. Sei q hierbei das Polynom was die Zertifikatslänge in A begrenzt. Wir wollen nun eine Levin-Reduktion von A auf B angeben. Sei $f \in \text{FP}$ die Funktion, welche die Reduktion $\text{Proj}(A) \leq_m^p \text{Proj}(B)$ realisiert.

Definiere folgende NPTM N , die wie folgt auf Eingabe w arbeitet:

```

1 wenn  $w$  nicht von der Form  $(x, y')$  dann akzeptiere
2  $(x, y') \leftarrow w$ 
3 wenn  $(f(x), y') \notin B$  dann akzeptiere
4 sonst
5    $\left[ \begin{array}{l} \text{(Ab hier gilt } f(x) \in \text{Pr}(B) \text{ und } x \in \text{Pr}(A)) \\ \text{Rate nichtdeterministisch } y \in \Sigma^{q(|x|)} \\ \text{Akzeptiere genau dann wenn } (x, y) \in A. \end{array} \right.$ 

```

Es ist leicht zu sehen dass $L(N) = \Sigma^*$. Nach (1) existiert nun also eine Funktion g sodass, für alle w , $g(w)$ eine akzeptierender Rechenweg von $N(w)$ ist.

Damit lässt die Levin-Reduktion von A auf B angeben: wähle f als Reduktionsfunktion, und definiere g' als Translationsfunktion, welche aus dem akzeptierenden Rechenweg $g(x, y')$ das geratene Zertifikat y von Zeile 5 ausliest. Dann gilt

$$\begin{aligned} (f(x), y') \in B &\implies N(x, y') \text{ akz. auf einem Rechenweg in Z. 6, ratet } y \\ &\implies (x, y) = (x, g'(x, y')) \in A \end{aligned}$$

wie gewünscht. Wir haben $A \leq_L^p$ via f, g' .

6. (6) \implies (8): Sei R eine beliebige \leq_L^p -vollständige und Levin-paddable NP-Relation (diese existiert immer) und sei N eine NPTM N mit $L(N) = \text{Proj}(R)$, wobei das Polynom q die Laufzeit von N beschränkt. Definiere

$$Q = \{(x, y) \mid N(x) \text{ akzeptiert mit Rechenweg } y \in \Sigma^{\leq q(|x|)}\}.$$

Es ist klar, dass $\text{Proj}(R) \leq_m^p \text{Proj}(Q) = L(N) = \text{Proj}(R)$ über die Identitätsfunktion. Nach (6) gilt nun auch $R \leq_L^p Q$, mit Identitätsfunktion als Reduktionsfunktion und $h \in \text{FP}$ als Translationsfunktion. Damit gilt

$$N(x) \text{ akz. mit RW } \alpha \implies (x, \alpha) \in Q \implies (x, h(x, \alpha)) \in R$$

wie gewünscht.

7. (2) \implies (7): Die Richtung von rechts nach links ist klar. Sei für die andere Richtung h ein optimales Beweissystem für eine Menge L . Wir wollen zeigen, dass h auch p-optimal ist. Sei dafür g ein weiteres Beweissystem für L . Nach Voraussetzung haben wir $g \leq h$, das heißt es existiert eine (nicht notwendigerweise effiziente) Funktion f sodass $g(w) = h(f(w))$, und gleichzeitig ist $|f(w)| \leq q(|w|)$ für ein geeignetes Polynom q .

Betrachte folgende Multifunktion f' :

$$f'(w) \mapsto y \iff \exists y \in \Sigma^{\leq q(|w|)}, g(w) = h(y).$$

Es lässt sich leicht zeigen, dass $f' \in \text{NPMV}$, über einen geeigneten NPTM-Transduktor. Es ist sogar $f' \in \text{NPMV}_t$, denn für jedes w mindestens $f(w) \in \text{set-}f'(w)$.

Nach (2) gilt also $f' \in \text{NPMV}_t \subseteq_c \text{FP}$, also existiert eine Funktion $f'' \in \text{FP}$ welche eine Verfeinerung von f' ist. Diese Funktion übersetzt g -Beweise w für x effizient in h -Beweise für x : Sei $g(w) = x$, dann gilt

$$f''(w) = y \quad \text{mit } y \in \Sigma^{\leq q(|w|)}, x = g(w) = h(y)$$

also ist $h(f''(w)) = x$ bzw. $f''(w)$ ein h -Beweis für x , wie gewünscht.

8. (7) \implies (9): klar, denn rKAN ist \leq_L^p -vollständig, ist Levin-paddable, und das Standardbeweissystem std_{rKAN} ist (wie jedes Standardbeweissystem einer NP-Relation) optimal. Zusammen mit (7) ist es also auch p-optimal. \square

Satz 4.13 (Formulierung Hypothese Q durch NP-Relationen, \forall -Variante). *Folgende Aussagen sind äquivalent:*

(1) *Hypothese Q*

(8') *Für jede \leq_L^p -vollständige Levin-paddable NP-Relation R , und für alle NPTM N mit $L(N) = \text{Proj}(R)$ gilt: es existiert eine Funktion $h \in \text{FP}$ mit*

$$N(x) \text{ akz. mit RW } \alpha \implies (x, h(x, \alpha)) \in R.$$

(9') *Für jede \leq_L^p -vollständige Levin-paddable NP-Relationen R ist das Standardbeweissystem std_R p-optimal.*

(10') *Für jede NP-Relation P und jede Menge $S \in \text{P}$ mit $S \subseteq \text{Proj}(P)$ gilt: es existiert eine Funktion $g \in \text{FP}$ sodass*

$$x \in S \implies (x, g(x)) \in P.$$

Beweis. 1. $(8') \Rightarrow (1)$, $(9') \Rightarrow (1)$, $(10') \Rightarrow (1)$: Die NP-Relation $rKAN$ ist $\leq_{L,1,inv}^P$ -vollständig (und damit auch Levin-paddable). Gilt also $(8')$, dann auch die existentielle Variante (8) von vorigem Satz 4.12, und damit (1). Beweise der anderen Implikationen sind analog.

2. $(1) \Rightarrow (8')$: Folgt aus Lemma 4.5.

3. $(8') \Rightarrow (9')$: Sei R eine beliebige \leq_L^P -vollständige Levin-paddable NP-Relation. Nach Voraussetzung können wir für jede NPTM N , $L(N) = \text{Proj}(R)$ eine Funktion $h \in FP$ angeben sodass

$$N(x) \text{ akz. mit RW } \alpha \Rightarrow (x, h(x, \alpha)) \in R.$$

Das ist genau die erste der äquivalenten Aussagen in Lemma 4.4. Damit gilt auch bezüglich *diesem* gewählten R auch die zweite der äquivalenten Aussagen, nämlich dass std_R p-optimal ist.

4. $(1) \Rightarrow (10')$: Im Beweis von Satz 4.12 wurde $(1) \Rightarrow (5) \Rightarrow (10)$ gezeigt. Beachte insbesondere dass im Beweis zur zweiten Implikation sogar eine stärkere Aussage bewiesen wurde, welche die Aussage (10) für alle NP-Relationen (und nicht nur vollständige) zeigt. Das entspricht genau der hier aufgestellten Aussage $(10')$, wie gewünscht. \square

4.3 Karp-Vollständigkeit vs. Levin-Vollständigkeit

Satz 4.14. $KvL \Rightarrow \neg Q$.

Beweis. Wir zeigen die Kontraposition, und starten mit der Voraussetzung Q . Wir wollen nun $\neg KvL$ zeigen. Sei hierfür R eine beliebige NP-Relation sodass $\text{Proj}(R) \leq_m^P$ -vollständig ist. Damit gilt also schon für alle weiteren NP-Relationen A , dass $\text{Proj}(A) \leq_m^P \text{Proj}(R)$. Nach Satz 4.12 gilt also auch die Aussage 4.12(6), und damit $A \leq_L^P R$. Also ist R auch \leq_L^P -vollständig, wie gewünscht und wir haben $\neg KvL$ gezeigt. \square

Was sind natürlich notwendige Bedingungen für die Hypothese KvL ? Diese Frage erscheint tatsächlich wesentlich schwieriger als gedacht. Insbesondere scheint es unklar, ob aus irgend einer von Pudlák's Hypothesen die Aussage KvL folgt.

Besonders interessant erscheint aber die Beziehung zur Hypothese $\neg Q$, also genau die Umkehrung von Satz 4.14. Betrachten wir hier exemplarisch den Fall für Relationen, die SAT bezeugen.

Starten wir mit $\neg Q$, dann haben wir über Satz 4.12 auch die Negation von Aussage 4.12(8). Nachdem $rSAT$ eine \leq_L^P -vollständige Levin-paddable NP-Relation ist, muss auch eine NPTM N existieren, für die $L(N) = SAT$, aber für alle Funktionen $h \in FP$ gilt

$$N(\varphi) \text{ akz. mit Rechenweg } w \not\Rightarrow (\varphi, h(\varphi, w)) \in rSAT. \quad (4.3)$$

In anderen Worten: es existiert zwar eine NPTM N welche die (\leq_m^P -vollständige Menge) SAT entscheidet, aber aus den akzeptierenden Rechenwegen w von $N(x)$ auf $x \in SAT$ kann nicht effizient eine akzeptierende Belegung für x abgeleitet werden.

Wir können N äquivalent als NP-Relation R_N repräsentieren, mit $(\varphi, w) \in R_N$ genau dann wenn $N(x)$ mit Rechenweg w akzeptiert. Damit kann Gleichung 4.3 so verstanden werden, dass $rSAT \not\leq_L^P R_N$ falls die Reduktionsfunktion f die Identitätsfunktion ist. An dieser Stelle muss erneut hervorgehoben werden, dass im Allgemeinen $rSAT \leq_L^P R_N$ mit Funktionen f, g gelten kann, notwendig hierfür ist aber dass $f \neq id$.

Gleichzeitig wäre die Existenz einer solchen Reduktion überraschend. Angenommen $rSAT \leq_L^P R_N$ wird von f, g ($f \neq id$) realisiert, dann haben wir nach Definition

$$N(f(\varphi)) \text{ akz. mit Rechenweg } w \Rightarrow \varphi \text{ wird von Belegung } g(\varphi, w) \text{ erfüllt.}$$

Einerseits ist es also nicht möglich, aus w effizient eine akzeptierende Belegung für $f(\varphi) \neq \varphi$ zu bestimmen. Andererseits reicht der „Beweis“ w aber aus, um (zusammen mit der Information φ) effizient wieder eine erfüllende Belegung für φ zu berechnen.

Ich vermute, dass solche Funktionen f, g nicht jeweils für alle NPTM N mit $L(N) = SAT$ existieren können. Tatsächlich können wir die eben formulierte Vermutung auch in der Theorie der Beweissystemen formulieren. Wir definieren zunächst eine abgeschwächte Variante der p-Simulation.

Definition 4.15. Seien h, h' Beweissysteme für L . Das Beweissystem h *p-simuliert effektiv* h' falls Funktionen $f, g \in FP$ existieren sodass

- (1) $x \in L \Rightarrow f(x) \in L$,
- (2) $h'(w) = f(x) \Rightarrow h(g(x, w)) = x$.

Wir schreiben in diesem Fall auch $h \leq_{\text{eff}}^p h'$.

In anderen Worten, falls $h \leq_{\text{eff}}^p h'$, dann kann h zwar nicht *jeden* h' -Beweis w für $x \in L$ in einen h -Beweis für (das gleiche) x effizient umrechnen, es kann aber zumindest alle *relevanten* h' -Beweise effizient umrechnen, nämlich für jedes $x \in L$ die h' -Beweise für $f(x)$ in h -Beweise für x .

Klar ist: p-Simulation impliziert effektive p-Simulation impliziert Simulation unter Beweissystemen.

\triangleleft

Nicht vergessen: es ist offenbar unklar, wie das Symbol \leq bzgl. Simulation gebraucht wird. Dose/Gläser würden in der Def. das Ordnungszeichen spiegeln, Krajíček/Pudlák dagegen genau so wie hier, Krajíček im Sammelband dagegen wieder

Vermutung 4.16 (KvL formuliert unter Beweissystemen). *Das Standardbeweissystem sat für SAT kann nicht alle anderen optimalen Beweissysteme für SAT effektiv p -simulieren.*

In anderen Worten, es existiert optimales Beweissystem h sodass $std \not\leq_{\text{eff}}^p h$.

Dass die Formulierung der Vermutungen 3.20 und 4.16 äquivalent sind, zeigt folgende Beobachtung:

Beobachtung 4.17. *Folgende Aussagen sind äquivalent:*

- (1) *Jede NP-Relation R mit \leq_m^p -vollständigem $\text{Proj}(R)$, ist auch \leq_L^p -vollständig.*
- (2) *Für alle optimalen Beweissysteme h für SAT gilt $std \leq_{\text{eff}}^p h$.*

Beweis. (1) \Rightarrow (2): Wir zeigen, dass sat jedes andere optimale Beweissystem h effektiv p -simulieren kann. Nachdem h optimal ist, hat es auch kurze Beweise: für jedes $\varphi \in \text{SAT}$ einen h -Beweis w mit $|w| \leq q(|\varphi|)$. Definiere

$$R_h = \{(\varphi, w) \mid |w| \leq q(|\varphi|), h(w) = \varphi\}.$$

Diese Relation ist offenbar eine NP-Relation und $\text{Proj}(R_h) = \text{SAT}$ und damit ist $\text{Proj}(R_h)$ auch \leq_m^p -vollständig.

Nach Voraussetzung ist also R_h auch \leq_L^p -vollständig. Insbesondere gilt also auch $\text{rSAT} \leq_L^p R_h$. Damit existieren also Funktionen $f, g \in \text{FP}$ sodass $x \in \text{SAT} \Leftrightarrow f(x) \in \text{SAT}$ und

$$(f(\varphi), w) \in R_h \implies (\varphi, g(\varphi, w)) \in \text{rSAT}.$$

Nach Definition gilt also

$$h(w) = f(\varphi) \implies sat(g(\varphi, w)) = \varphi,$$

und damit ist $sat \leq_{\text{eff}}^p h$.

(2) \Rightarrow (1): Sei R eine NP-Relation wobei $\text{Proj}(R) \leq_m^p$ -vollständig ist. Wir zeigenen nun, dass R auch \leq_L^p -vollständig ist. Aus der \leq_m^p -Vollständigkeit folgt unmittelbar die Existenz einer Reduktionsfunktion f mit

$$\varphi \in \text{SAT} \Leftrightarrow f(\varphi) \in \text{Proj}(R).$$

Definiere

$$h(w) = \begin{cases} \varphi & \text{falls } w = (x, y, \varphi) \text{ und } f(\varphi) = x \text{ und } (x, y) \in R \\ \perp & \text{sonst.} \end{cases}$$

Wir zeigen, dass h ein Beweissystem für SAT ist. Es ist offenbar dass $h \in \text{FP}$. Die Funktion h ist korrekt: wenn $h(x, y, \varphi) = \varphi$ dann ist $f(\varphi) = x \in \text{Proj}(R)$ und nach Eigenschaft von f auch $\varphi \in \text{SAT}$. Die Funktion h ist vollständig: Sei $\varphi \in \text{SAT}$. Dann ist schon $f(\varphi) \in \text{Proj}(R)$ und es gibt ein y mit $(f(\varphi), y) \in R$. Also ist $(f(\varphi), y, \varphi)$ ein h -Beweis für φ .

Außerdem ist klar, dass h ehrlich ist. Definiere

$$R_h = \{(\varphi, w) \mid h(w) = \varphi\},$$

diese Relation ist damit eine NP-Relation. Wir wollen nun zeigen dass $\text{rSAT} \leq_L^p R_h \leq_L^p Q$, womit dann Q auch \leq_L^p -vollständig ist, wie gewünscht.

Wir starten mit der zweiten Reduktion. Es gilt $R_h \leq_L^p Q$, denn es gilt

$$(f(\varphi), y) \in Q \implies h(\underbrace{f(\varphi), y}_{g(\varphi, y)}) = \varphi \implies (\varphi, g(\varphi, y)) \in R_h,$$

wobei $g(\varphi, y) = (f(\varphi), y, \varphi)$.

Für die erste Reduktion nutzen wir die Voraussetzung. Es gilt nach Voraussetzung $std \leq_{\text{eff}}^p h$. Damit existieren also Funktionen $f', g' \in \text{FP}$ mit

$$\begin{aligned} \varphi \in \text{SAT} &\Leftrightarrow f'(\varphi) \in \text{SAT} \\ h(w) = f'(\varphi) &\implies sat(g'(\varphi, w)) = \varphi. \end{aligned}$$

Jetzt ist aber auch klar, dass f', g' die Reduktion $\text{rSAT} \leq_L^p R_h$ realisieren, denn nun gilt

$$(f'(\varphi), w) \in R_h \implies (\varphi, g'(\varphi, w)) \in \text{rSAT}.$$

□

Mit der Definition der effektiven p -Simulation und der eben bewiesenen äquivalenten Formulierung der KvL-Vermutung lässt sich nun zumindest die Hypothese SAT so verstärken, dass diese hinreichend für KvL ist.

Vermutung 4.18 (SAT^{eff}). *Kein optimales Beweissystem für SAT kann alle anderen optimalen Beweissysteme für SAT effektiv p -simulieren. In anderen Worten, für jedes optimales Beweissystem h existiert ein optimales Beweissysteme h' sodass $h \not\leq_{\text{eff}}^p h'$.*

Satz 4.19. (1) $\text{SAT}^{\text{eff}} \implies \text{SAT}$

(2) $\text{SAT}^{\text{eff}} \implies \text{KvL}$

Beweis. 1. Zu (1): Klar aus Kontraposition. Wenn ein p -optimales Beweissystem h für SAT existiert, dann kann dieses (optimale) h auch alle anderen Beweissysteme p -simulieren, und damit insbesondere auch alle optimalen Beweissysteme h' effektiv p -simulieren.

2. Zu (2): Wieder klar aus Kontraposition. Unter $\neg \text{KvL}$ folgt mit der Formulierung aus Vermutung 4.16 dass das (optimale) Standardbeweissystem sat alle optimalen Beweissysteme effektiv p -simulieren kann. Dann existiert also

4.4 Bekannte Implikationen, Offene Orakel

Der folgende Satz ist eine Generalisierung von Dingel (2022).

Satz 4.20. *Wenn $\text{NP} = \text{coNP}$ dann existiert eine \leq_m^P -vollständige Multifunktion f für NPMV_t .*

Beweis. Nach Voraussetzung können wir in NP testen, ob ein Wort x im Urbild einer beliebigen NPMV-Multifunktion liegt. Es gilt $\text{KAN} \in \text{NP}$ und damit $\text{KAN} \in \text{coNP}$. Insbesondere ist dann die Menge

$$U = \{(i, x, 1^n) \mid T_i \text{ akz. auf keinem Rechenweg der Länge } \leq n\} \in \text{NP},$$

und wird von der NPTM N_u in Laufzeit $q(|(i, x, 1^n)|)$ entschieden.

Betrachte nun die Multifunktion f , die durch folgenden nichtdeterministischen Transduktor $T'(i, x, 1^n)$ berechnet wird:

```

1 wenn  $T$  kein Transduktor ist oder  $n \neq |x|^i + i$  dann
2   | akzeptiere
3 Rate nichtdeterministisch einen Rechenweg  $\alpha$  von  $T_i$  der Länge  $\leq n$ 
4 Rate nichtdeterministisch einen Rechenweg  $\beta$  von  $N_u$  der Länge  $\leq q(|(i, x, 1^n)|)$ 
5 wenn Falls  $T_i(x)$  mit  $\alpha$  akzeptiert dann
6   |  $y \leftarrow$  Ausgabe von  $T_i(x)$  auf  $\alpha$ 
7   | Gebe  $y$  aus
8 sonst wenn Falls  $N_u(i, x, 1^n)$  mit  $\beta$  akzeptiert dann
9   | Gebe  $\varepsilon$  aus
10 sonst
11   | Lehne ab

```

Es ist leicht zu sehen dass T' in Polynomialzeit arbeitet. Wir betrachten nun Eingaben $(i, x, 1^n)$, $n = |x|^i + i$. Es gilt nun:

- Entweder ist $\text{set-}T_i(x) \neq \emptyset$, dann existiert für jedes $y \in \text{set-}T_i(x)$ ein akzeptierender Rechenweg α der Länge $\leq n$ auf T_i der y ausgibt, und damit wird auch f dieses y in Z. 5 ausgeben. Gleichzeitig ist damit $(i, x, 1^n) \notin U$ und Z. 7 niemals erreicht. Es gilt also $\text{set-}f(i, x, 1^n) = \text{set-}T_i(x)$.
- Oder es gilt $\text{set-}T_i(x) = \emptyset$. Dann wird jeder Rechenweg der Länge $\leq n$ von $T_i(x)$ ablehnen, und f definitiv nicht in Z. 5 akzeptieren. Andererseits gilt dann $(i, x, 1^n) \in U$ und Z. 7 wird auf mindestens einem Rechenweg von f erreicht. Es gilt also $\text{set-}f(i, x, 1^n) = \{\varepsilon\}$.

Damit ist klar, dass $f \in \text{NPMV}_t$. Wir zeigen nun, dass f auch NPMV_t -vollständig ist. Sei hierfür g eine beliebige Multifunktion aus NPMV_t . Dann existiert auch ein i sodass der nichtdeterministische Transduktor T_i diese Multifunktion g in berechnet, und dabei terminiert $T_i(x)$ in $\leq |x|^i + i$ vielen Schritten.

Nun gilt nach obiger Beobachtung schon dass

$$\text{set-}g(x) = \text{set-}T_i(x) \neq \emptyset \implies \text{set-}f(\underbrace{(i, x, 1^{|x|^i+i})}_{h(x)}) = \text{set-}T_i(x) = \text{set-}g(x)$$

und $h(x) = (i, x, 1^{|x|^i+i})$ realisiert die Reduktion von g auf f , wie gewünscht. □

Satz 4.21. *Es gelten die in Abbildung 2 abgebildeten Implikationen und Äquivalenzen.*

Beweis. Es gelten die notierten Äquivalenzen:

1. $\neg Q \Leftrightarrow \text{sat}$ ist p-optimal, nach Satz 4.12.
2. $\text{NP} \neq \text{coNP} \Leftrightarrow$ existiert keine NP-harte Funktion in TFNP, nach Satz ??.
3. $\neg Q' \Leftrightarrow \exists$ P-inseparierbares DisjCoNP-Paar, nach Fortnow und Rogers (2002).
4. $\text{NP} \cap \text{coNP} \neq P \Leftrightarrow \text{NPSV}_t \not\subseteq \text{FP}$, nach Selman (1994).
5. $\text{UP} \neq P \Leftrightarrow \exists$ Einwegfunktionen, nach Grollmann und Selman (1988, Thm. 10).
6. $\text{UP} \cap \text{coUP} \neq P \Leftrightarrow \exists$ Einwegpermutationen, nach Homan und Thakur (2003).
7. $\text{NP} \cap \text{coNP} \Leftrightarrow \text{NPSV}_t$ hat keine vollständige Funktion, nach Beyersdorff, Köbler und Messner (2009, Prop. 3).
8. $\text{DisjNP} \Leftrightarrow \text{NPSV}$ hat keine vollständige Funktion, nach Glaßer, Selman und Sengupta (2005, Thm. 9).

Es gelten die eingezeichneten Implikationen:

1. $\text{DisjNP} \Rightarrow \text{CON}^N$ nach Köbler, Messner und Torán (2003).
2. $\text{UP} \Rightarrow \text{TAUT}$ nach Köbler, Messner und Torán (2003).
3. $\text{CON}^N \Rightarrow \text{NEE} \neq \text{coNEE}$ nach Köbler, Messner und Torán (2003).
4. $\text{NP} \cap \text{coNP} \neq P \Rightarrow \neg Q' \Rightarrow \text{NPMV}_t \not\subseteq_c \text{TFNP} \Rightarrow \neg Q$ nach Fenner u. a. (2003).
5. $E \neq \text{NE} \Rightarrow \exists$ NP-Relation die nicht auf Entscheidung reduzierbar ist, nach Impagliazzo und Sudan (1991).
6. $\text{UP} \neq P \Rightarrow \exists$ P-inseparierbares DisjNP-Paar, nach Grollmann und Selman (1988, Thm. 5).
7. $\text{NP} \cap \text{coNP} \Rightarrow \text{TAUT} \vee \text{SAT}$ nach Beyersdorff, Köbler und Messner (2009).
8. NPMV_t hat keine vollständige Funktion $\Rightarrow \text{SAT}$ nach Beyersdorff, Köbler und Messner (2009, Thm. 25).
9. NPMV_t hat keine vollständige Funktion $\Rightarrow \text{NP} \neq \text{coNP}$ nach Satz ??.
10. $\text{SAT}^{\text{eff}} \Rightarrow \text{SAT}, \text{SAT}^{\text{eff}} \Rightarrow \text{KvL}$, nach Satz 4.19.
11. $\text{KvL} \Rightarrow \neg Q$, nach Satz 4.14.
12. $\neg Q \Rightarrow \exists$ NP-Relation die nicht auf Entscheidung reduzierbar ist, denn unter $\neg Q$ gilt mit Satz 4.12 auch die Negation von 4.12(1), also eine NPTM N mit $L(N) = \Sigma^*$ wobei keine Funktion $g \in \text{FP}$ existiert, welche für alle x durch $g(x)$ einen akzeptierenden Rechenweg von $N(x)$ bestimmt. Definiere die NP-Relation R_N mit $(x, \alpha) \in R_N$ genau dann wenn $N(x)$ mit Rechenweg α existiert. Nun gilt nach Vorigem auch $R_N \not\subseteq_c \text{FP} = \text{FP}^{\Sigma^*} = \text{FP}^{L(R)}$.
13. $\text{DisjCoNP} \Rightarrow \text{TFNP} \Rightarrow \text{NPMV}_t$ hat keine vollständige Funktion, nach Pudlák (2017).
14. $\text{NP} \cap \text{coNP} \neq P \Rightarrow \exists$ P-inseparierbares DisjNP-Paar, denn wenn alle DisjNP-Paare P-separierbar, dann ist auch für jede Menge $L \in \text{NP} \cap \text{coNP}$ jeweils das DisjNP-Paar (L, \bar{L}) P-separierbar und damit $L \in P$.
15. $\text{DisjNP} \Rightarrow \exists$ P-inseparierbares DisjNP-Paar; ist klar, denn wenn alle DisjNP-Paare P-separierbar wären, dann

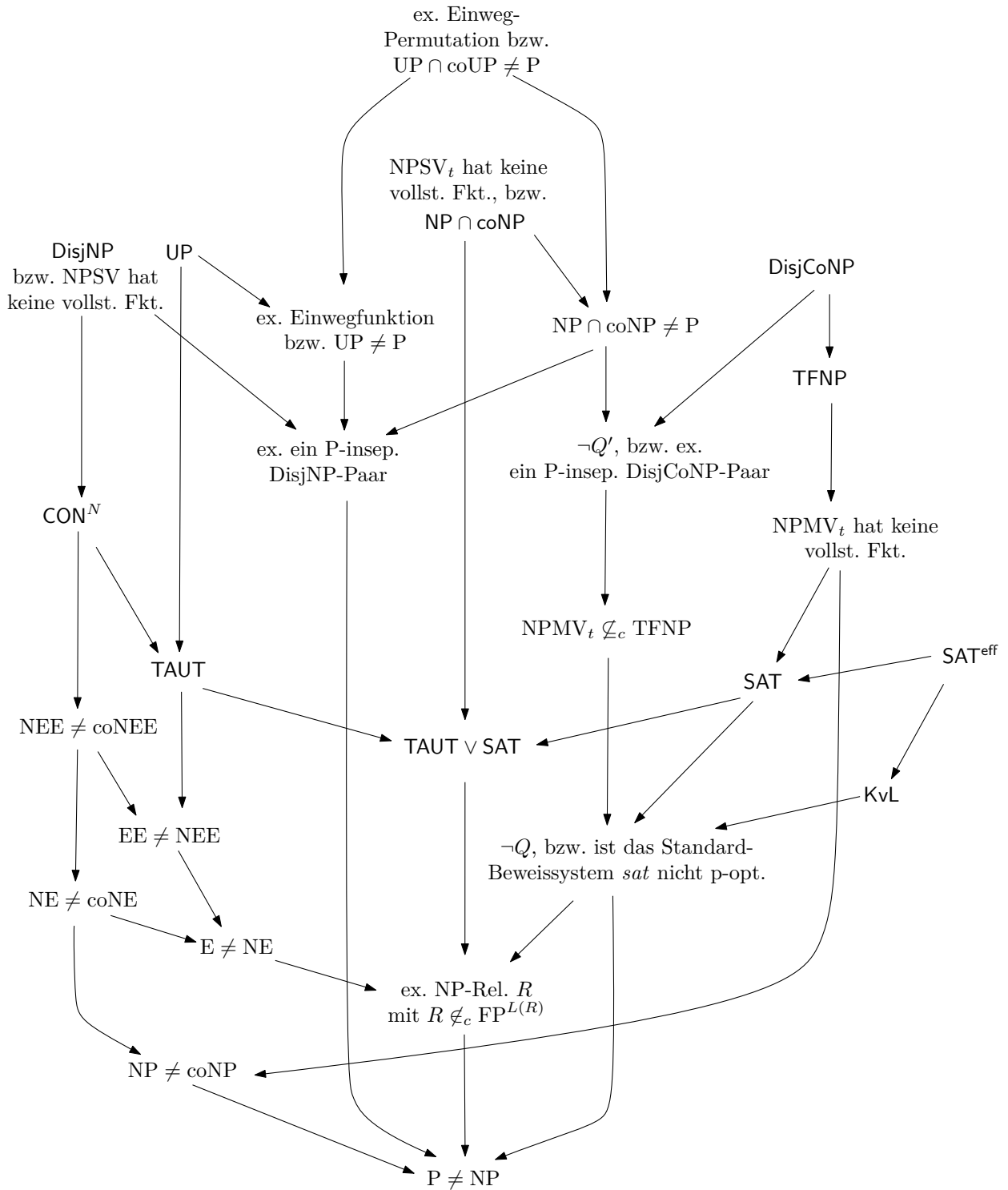


Abbildung 2: Bekannte (relativierenden) Implikationen zwischen den betrachteten Hypothesen und weiteren Aussagen. Satz 4.21 gibt Belegstellen für jede dieser Implikationen an.

wären auch alle Paare \leq_m^{PP} -vollständig.

16. $\text{DisjCoNP} \Rightarrow \exists$ P-inseparierbares DisjCoNP-Paar; ist aus selben Gründen klar.

17. $\text{CON}^N \Rightarrow \text{TAUT}$ klar, weil aus p-Optimalität auch Optimalität folgt.

18. $\text{SAT} \Rightarrow \neg Q$ klar, denn wenn *sat* p-optimal ist, dann existiert *ein* p-optimales Beweissystem für SAT.

19. $\text{UP} \Rightarrow \text{UP} \neq \text{P}$ klar.

20. $\text{NP} \cap \text{coNP} \Rightarrow \text{NP} \cap \text{coNP} \neq \text{P}$ klar.

21. \exists P-inseparierbares DisjNP-Paar $\Rightarrow \text{P} \neq \text{NP}$ klar.

22. $\text{UP} \cap \text{coUP} \Rightarrow \text{UP} \neq \text{P}$, $\text{UP} \cap \text{coUP} \Rightarrow \text{NP} \cap \text{coNP} \neq \text{P}$ klar.

23. $\text{NEE} \neq \text{coNEE} \Rightarrow \text{NE} \neq \text{coNE} \Rightarrow \text{NP} \neq \text{coNP} \Rightarrow \text{P} \neq \text{NP}$ klar.

24. $\text{NEE} \neq \text{coNEE} \Rightarrow \text{EE} \neq \text{NEE} \Rightarrow \text{E} \neq \text{NE}$ klar. □

5 Orakel

Abschließend erweitern wir noch unsere Notation, die uns insbesondere bei der Orakelkonstruktion helfen wird. Diese folgt Überlegungen von Dose und Glaßer (2019). Anstelle von Orakeln als Menge zu verstehen, können wir äquivalent Orakel auch als unendlich lange Wörter $u \in \Sigma^\omega$ formulieren, die wir als das Orakel $\{i \mid u[i] = 1\} \subseteq \mathbb{N}$ interpretieren. Mit der obigen Identifikation von Wörtern und natürlichen Zahlen beschreibt nun u sowohl ein Orakel über \mathbb{N} als auch über Σ^* ; wir können also z.B. von der relativen Berechnung $M^w(x)$ sprechen. Analog fassen wir endlich lange Wörter $w \in \Sigma^*$ als *partielles* Orakel $\{i \mid w[i] = 1\}$, welches die Zugehörigkeit der Wörter $x < |w|$ festlegt, aber die Zugehörigkeit aller Wörter $y \geq |w|$ noch nicht endgültig festlegt. Auf dieser Idee der endgültigen bzw. noch nicht endgültigen Zugehörigkeit aufbauend können wir auch von *definiten* Berechnungen sprechen: Eine Rechnung $M^w(x)$ ist *definit* wenn auf allen Rechenwegen von $M^w(x)$ nur Orakelfragen gestellt werden, welche eine Länge $< |w|$ haben.

Literatur

- Adleman, Leonard und Kenneth Manders. 1977. „Reducibility, Randomness, and Intractability (Abstract)“. In: *Proceedings of the Ninth Annual ACM Symposium on Theory of Computing (STOC '77)*. Boulder, Colorado, USA: Association for Computing Machinery, 1977, S. 151–163. DOI: 10.1145/800105.803405.
- Agrawal, Manindra und Somenath Biswas. 1992. „Universal relations“. In: *Proceedings of the Seventh Annual Structure in Complexity Theory Conference*. Seventh Annual Structure in Complexity Theory Conference. Juni 1992, S. 207–220. DOI: 10.1109/SCT.1992.215395.
- Agrawal, Manindra, Neeraj Kayal und Nitin Saxena. 2004. „PRIMES Is in P“. In: *Annals of Mathematics* 160.2 (2004), S. 781–793. DOI: 10.4007/annals.2004.160.781.
- Arora, Sanjeev und Boaz Barak. 2009. *Computational complexity: a modern approach*. Cambridge: Cambridge University Press, 2009. ISBN: 978-0-521-42426-4.
- Baker, Theodore, John Gill und Robert Solovay. 1975. „Relativizations of the P=?NP Question“. In: *SIAM Journal on Computing* 4.4 (Dez. 1975), S. 431–442. DOI: 10.1137/0204037.
- Balcázar, José L. 1989. „Self-reducibility structures and solutions of NP problems.“ In: *Revista Matemática de la Universidad Complutense de Madrid* 2.2-3 (1989), S. 175–184. URL: <http://eudml.org/doc/43531>.
- Bellare, Mihir und Shafi Goldwasser. 1994. „The Complexity of Decision Versus Search“. In: *SIAM Journal on Computing* 23.1 (Feb. 1994), S. 97–119. DOI: 10.1137/S0097539792228289.
- Beyersdorff, Olaf, Johannes Köbler und Jochen Messner. 2009. „Nondeterministic functions and the existence of optimal proof systems“. In: *Theoretical Computer Science* 410.38 (6. Sep. 2009), S. 3839–3855. DOI: 10.1016/j.tcs.2009.05.021. (Besucht am 23.06.2019).
- Borodin, Allan B. und Alan J. Demers. 1976. *Some Comments on Functional Self-Reducibility and the NP Hierarchy*. Techn. Ber. 76-284. Ithaca, New York, USA: Department of Computer Science, Cornell University, 1976. URL: <https://hdl.handle.net/1813/6540>.
- Buhrman, H., J. Kadin und T. Thierauf. 1998. „Functions Computable with Nonadaptive Queries to NP“. In: *Theory of Computing Systems* 31.1 (1. Feb. 1998), S. 77–92. DOI: 10.1007/s002240000079. (Besucht am 30.09.2022).
- Cook, Stephen A. 1971. „The Complexity of Theorem-Proving Procedures“. In: *Proceedings of the Third Annual ACM Symposium on Theory of Computing (STOC'71)*. Shaker Heights, Ohio, USA: Association for Computing Machinery, 1971, S. 151–158. DOI: 10.1145/800157.805047.
- Cook, Stephen A. und Robert A. Reckhow. 1979. „The relative efficiency of propositional proof systems“. In: *The Journal of Symbolic Logic* 44.1 (März 1979), S. 36–50. ISSN: 0022-4812, 1943-5886. DOI: 10.2307/2273702.
- Dingel, David. 2022. „Separation der relativierten Vermutungen SAT und TFNP“. Bachelorarbeit. Universität Würzburg, 22. Okt. 2022.
- Dose, Titus und Christian Glaßer. 2019. *NP-Completeness, Proof Systems, and Disjoint NP-Pairs*. 050. 2019. URL: <https://eccc.weizmann.ac.il/report/2019/050/> (besucht am 03.04.2019).
- Fenner, Stephen A., Lance Fortnow, Ashish V. Naik und John D. Rogers. 2003. „Inverting onto functions“. In: *Information and Computation* 186.1 (Okt. 2003), S. 90–103. DOI: 10.1016/S0890-5401(03)00119-6. (Besucht am 04.01.2022).
- Fischer, Sophie, Lane A. Hemaspaandra und Leen Torenvliet. 1995. „Witness-isomorphic reductions and the local search problem“. In: *Mathematical Foundations of Computer Science 1995*. Hrsg. von Jiří Wiedermann und Petr Hájek. Bd. 969. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 1995, S. 277–287. DOI: 10.1007/3-540-60246-1_134. (Besucht am 03.10.2022).
- Fortnow, Lance und John D. Rogers. 2002. „Separability and one-way functions“. In: *Computational Complexity* 11.3 (1. Juni 2002), S. 137–157. DOI: 10.1007/s00037-002-0173-4. (Besucht am 13.09.2022).
- Glaßer, Christian, Alan L. Selman und Samik Sengupta. 2005. „Reductions between disjoint NP-Pairs“. In: *Information and Computation* 200.2 (1. Aug. 2005), S. 247–267. ISSN: 0890-5401. DOI: 10.1016/j.ic.2005.03.003.

- Goldreich, Oded. 2008. *Computational Complexity: a Conceptual Perspective*. Cambridge: Cambridge University Press, 2008. 606 S. ISBN: 978-0-521-88473-0.
- Grollmann, Joachim und Alan L. Selman. 1988. „Complexity Measures for Public-Key Cryptosystems“. In: *SIAM Journal on Computing* 17.2 (Apr. 1988), S. 309–335. DOI: 10.1137/0217018. (Besucht am 05.03.2023).
- Harsha, Prahladh, Daniel Mitropolsky und Alon Rosen. 2023. „Downward Self-Reducibility in TFNP“. In: *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*. Hrsg. von Yael Tauman Kalai. Bd. 251. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023, 67:1–67:17. DOI: 10.4230/LIPIcs.ITCS.2023.67.
- Hartmanis, J. und L. Berman. 1976. „On isomorphisms and density of NP and other complete sets“. In: *Proceedings of the eighth annual ACM symposium on Theory of computing*. STOC '76. New York: Association for Computing Machinery, 3. Mai 1976, S. 30–40. DOI: 10.1145/800113.803628. (Besucht am 23.08.2023).
- Hemaspaandra, Lane A. 1998. „Complexity Theory Column 20: Take-home complexity“. In: *ACM SIGACT News* 29.2 (Juni 1998), S. 9–13. DOI: 10.1145/288079.288080. (Besucht am 03.10.2022).
- Homan, Christopher M. und Mayur Thakur. 2003. „One-way permutations and self-witnessing languages“. In: *Journal of Computer and System Sciences* 67.3 (1. Nov. 2003), S. 608–622. DOI: 10.1016/S0022-0000(03)00068-0. (Besucht am 30.09.2023).
- Impagliazzo, Russel und Mahdu Sudan. 1991. Private Kommunikation. Nicht Publiziert. Berichtet von Bellare und Goldwasser 1994, S. 102. Mai 1991.
- Karp, Richard M. 1972. „Reducibility among Combinatorial Problems“. In: *Complexity of Computer Computations*. Hrsg. von Raymond E. Miller, James W. Thatcher und Jean D. Bohlinger. Boston, MA: Springer, 1972, S. 85–103. DOI: 10.1007/978-1-4684-2001-2_9.
- Köbler, Johannes und Jochen Messner. 2000. „Is the Standard Proof System for SAT P-Optimal?“. In: *FST TCS 2000: Foundations of Software Technology and Theoretical Computer Science*. Hrsg. von Sanjiv Kapoor und Sanjiva Prasad. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2000, S. 361–372. DOI: 10.1007/3-540-44450-5_29.
- Köbler, Johannes, Jochen Messner und Jacobo Torán. 2003. „Optimal proof systems imply complete sets for promise classes“. In: *Information and Computation* 184.1 (10. Juli 2003), S. 71–92. DOI: 10.1016/S0890-5401(03)00058-0. (Besucht am 23.06.2019).
- Krajíček, Jan und Pavel Pudlák. 1989. „Propositional proof systems, the consistency of first order theories and the complexity of computations“. In: *Journal of Symbolic Logic* 54.3 (Sep. 1989), S. 1063–1079. DOI: 10.2307/2274765.
- Levin, Leonid A. 1973. „Универсальные задачи перебора [Universelle Suchprobleme]“. In: *Проблемы Передачи Информации [Problemy Peredachi Informatsii]* 9.3 (1973), S. 115–116. URL: <https://www.mathnet.ru/ppi914>. Eine Übersetzung erscheint bei Trakhtenbrot (1984, S. 399–400).
- Lynch, Nancy und Richard J. Lipton. 1978. „On Structure Preserving Reductions“. In: *SIAM Journal on Computing* 7.2 (Mai 1978), S. 119–126. DOI: 10.1137/0207010. (Besucht am 25.08.2023).
- Megiddo, Nimrod und Christos H. Papadimitriou. 1991. „On total functions, existence theorems and computational complexity“. In: *Theoretical Computer Science* 81.2 (30. Apr. 1991), S. 317–324. DOI: 10.1016/0304-3975(91)90200-L.
- Messner, Jochen. 2000. „On the simulation order of proof systems“. Diss. Universität Ulm, 2000. URL: <https://citeseerx.ist.psu.edu/pdf/bec3958d845653cfa73493258d3b550a17e8defd> (besucht am 03.07.2022). Archivierte Fassung des Originals.
- Papadimitriou, Christos H. 1994. *Computational complexity*. Reading, Massachusetts, USA: Addison-Wesley, 1994. ISBN: 978-0-201-53082-7.
- Pudlák, Pavel. 2017. „Incompleteness in the finite domain“. In: *The Bulletin of Symbolic Logic* 23.4 (2017), S. 405–441. DOI: 10.1017/bsl.2017.32. (Besucht am 01.04.2019).
- Selman, Alan L. 1988. „Natural Self-Reducible Sets“. In: *SIAM Journal on Computing* 17.5 (Okt. 1988), S. 989–996. DOI: 10.1137/0217062.
- Selman, Alan L. 1994. „A taxonomy of complexity classes of functions“. In: *Journal of Computer and System Sciences* 48.2 (Apr. 1994), S. 357–381. DOI: 10.1016/S0022-0000(05)80009-1. (Besucht am 08.01.2022).
- Simon, Janos. 1975. *On Some Central Problems in Computational Complexity*. Techn. Ber. 75-224. Ithaca, New York, USA: Department of Computer Science, Cornell University, 1975. URL: <https://hdl.handle.net/1813/6975>.

- Trakhtenbrot, B.A. 1984. „A Survey of Russian Approaches to Perebor (Brute-Force Searches) Algorithms“. In: *Annals of the History of Computing* 6.4 (Okt. 1984), S. 384–400. DOI: 10.1109/MAHC.1984.10036.
- Valiant, Leslie G. 1979. „The complexity of computing the permanent“. In: *Theoretical Computer Science* 8.2 (1979), S. 189–201. DOI: 10.1016/0304-3975(79)90044-6.
- Wechsung, Gerd. 2000. *Vorlesungen zur Komplexitätstheorie*. Bd. 32. Teubner-Texte zur Informatik. Wiesbaden: Vieweg+Teubner Verlag, 2000. DOI: 10.1007/978-3-322-80024-4.