Das klassische Entscheidungsproblem der mathematischen Logik

Verschiedene äquivalente Formulierungen:

Erfüllbarkeit: Man konstruiere einen Algorithmus, welcher zu jeder vorgelegten Formel der Prädikatenlogik entscheidet, ob sie erfüllbar ist oder nicht.

Gültigkeit: Man finde einen Algorithmus, welcher zu jeder Formel ψ der Prädikatenlogik entscheidet, ob sie allgemeingültig ist, d.h. ob jede zu ψ passende Interpretation ein Modell von ψ ist.

Beweisbarkeit: Man konstruiere einen Algorithmus, welcher zu jeder Formel $\psi \in \mathrm{FO}$ entscheidet, ob ψ (aus der leeren Hypothesenmenge) ableitbar ist. (Hier wird ein fester, vollständiger Beweiskalkül für die Prädikatenlogik zugrunde gelegt, z.B. der Sequenzenkalkül).

D. Hilbert, W. Ackermann, *Grundzüge der theoretischen Logik*, 1. Auflage, Berlin 1928, S. 73ff:

Das Entscheidungsproblem ist gelöst, wenn man ein Verfahren kennt, das bei einem vorgelegten logischen Ausdruck durch endlich viele Operationen die Entscheidung über die Allgemeingültigkeit bzw. Erfüllbarkeit erlaubt. (...) Das Entscheidungsproblem muss als das Hauptproblem der mathematischen Logik bezeichnet werden.

Turingmaschinen

Definition. Eine *Turingmaschine* ist gegeben durch ein 6-Tupel

$$M = (Q, \Sigma, \Gamma, q_0, F, \delta),$$

bestehend aus

- einer endlichen Menge Q von *Zuständen*,
- einem endlichen *Arbeitsalphabet* Σ mit einem ausgezeichneten *Blank*-Symbol $\Box$,
- einem *Inputalphabet* $\Gamma \subseteq \Sigma - \{\Box\}$,
- einem *Anfangszustand* $q_0 \in Q$,
- einer Menge $F \subseteq Q$ von *Endzuständen*, und
- einer (partiellen) *Übergangsfunktion*

$$\delta : (Q - F) \times \Sigma \to Q \times \Sigma \times \{-1, 0, 1\}.$$

Konfigurationen: $C = (q, p, w) \in Q \times \mathbb{N} \times \Sigma^*$

Inputkonfiguration auf Input x: $C_0(x) := (q_0, 0, x)$.

Endkonfiguration: $C = (q, p, w)$ mit $q \in F$.

Die Übergangsfunktion δ induziert eine partielle Übergangsfunktion auf Konfigurationen : $C \mapsto \mathrm{Next}(C)$.

Berechnung von M auf x: Folge $C_0, C_1, \ldots$ von Konfigurationen mit $C_0 := C_0(x)$ und $C_{i+1} = \mathrm{Next}(C_i)$ für alle i.

M **hält auf x:** die Berechnung von M auf x endet in einer Endkonfiguration.

$L(M) := \{x : M \text{ hält auf } x\}$.

M berechnet eine partielle Funktion $f_M : \Gamma^* \to \Sigma^*$.
Der Definitionsbereich von f_M ist $L(M)$, und $f_M(x) = y$ gdw. die Berechnung von M auf x in einer Endkonfiguration (q, p, y) endet.

Turing-Akzeptor: Turingmaschine mit Menge F^+ von *akzeptierenden* und Menge F^- von *verwerfenden Zuständen.* $F = F^+ \cup F^-$.
M *akzeptiert* (bzw. *verwirft*) x, wenn die Berechnung von M auf x in einer akzeptierenden (bzw. verwerfenden) Konfiguration endet.

Definition.

(1) $L \subseteq \Gamma^*$ ist *rekursiv aufzählbar (r.e.)*, wenn eine Turingmaschine M existiert, so dass $L = L(M)$.

(2) $L \subseteq \Gamma^*$ ist *co-r.e.* wenn das Komplement $\overline{L} := \Gamma^* - L$ r.e. ist.

(3) Eine (partielle) Funktion ist *(turing-)berechenbar*, wenn eine Turingmaschine M existiert, so dass $f = f_M$.

(4) $L \subseteq \Gamma^*$ ist *entscheidbar* oder *rekursiv*, wenn ein Turing-Akzeptor M existiert, so dass für alle $x \in \Gamma^*$:

- $x \in L \quad \Longrightarrow \quad M$ akzeptiert x.
- $x \notin L \quad \Longrightarrow \quad M$ verwirft x.

Äquivalente Definition: Eine Sprache $L \subseteq \Gamma^*$ ist *entscheidbar*, wenn ihre charakteristische Funktion $\chi_L : \Gamma^* \to \{0,1\}$ turingberechenbar ist.

Satz. *Eine Sprache L ist entscheidbar genau dann, wenn L und $\overline{L}$ rekursiv aufzählbar sind.*

Definition. Sei $A \subseteq \Gamma^*$, $B \subseteq \Sigma^*$. A ist *reduzierbar* auf B (kurz: $A \leq B$), wenn eine totale berechenbare Funktion $f : \Gamma^* \to \Sigma^*$ existiert, so dass für alle $x \in \Gamma^*$ gilt: $x \in A \quad \Longleftrightarrow \quad f(x) \in B$.

Lemma.

(i) $A \leq B$, B *entscheidbar* $\quad \Longrightarrow \quad A$ *entscheidbar.*

(ii) $A \leq B$, B *r.e.* $\quad \Longrightarrow \quad A$ *r.e.*

(iii) $A \leq B$, A *unentscheidbar* $\quad \Longrightarrow \quad B$ *unentscheidbar.*

Halteprobleme

Allgemeines Halteproblem: $H := \{\rho(M)\#\rho(x) : x \in L(M)\}$.

$\rho(M)$ und $\rho(x)$ sind Kodierungen von M und x über dem festen Alphabet $\{0,1\}$, so dass die Berechnung von M auf x aus den Kodierungen effektiv rekonstruiert werden kann. Das heisst: Es gibt eine *universelle Turingmaschine* U, welche, gegeben $\rho(M)\#\rho(x)$, die Berechnung von M auf x simuliert. Es folgt: H ist rekursiv aufzählbar.

Selbstanwendungsproblem: $H_0 := \{\rho(M) : \rho(M) \in L(M)\}$.

Halteproblem auf leerem Wort: $H_\lambda := \{\rho(M) : \lambda \in L(M)\}$.

Satz. *H_0, H und H_λ sind unentscheidbar.*

(1) H_0 ist nicht co-r.e. (und also unentscheidbar).
Anderfalls wäre $\overline{H_0} = L(M_0)$ für eine Turingmaschine M_0. Dann wäre aber

$$\rho(M_0) \in \overline{H_0} \iff M_0 \text{ hält auf } \rho(M_0) \iff \rho(M_0) \in H_0.$$

(2) H_0 Spezialfall von H (insbesondere $H_0 \leq H$)

(3) $H \leq H_\lambda$

Ein allgemeines Unentscheidbarkeitstheorem. Jede nicht-triviale Aussage über das Berechnungsverhalten von Turingmaschinen ist unentscheidbar. Sei $\mathcal{R}$ die Klasse aller turingberechenbaren Funktionen.

Satz (Rice). *Sei $\mathcal{S}$ eine Menge von turingberechenbaren Funktionen mit $\mathcal{S} \neq \varnothing$ und $\mathcal{S} \neq \mathcal{R}$. Dann ist* $\mathrm{code}(\mathcal{S}) := \{\rho(M) : f_M \in \mathcal{S}\}$ *unentscheidbar.*

(Reduziere H_λ oder $\overline{H_\lambda}$ auf $\mathrm{code}(\mathcal{S})$.)

Rekursive Untrennbarkeit

Definition. Seien $A, B \subseteq \Gamma^*$ mit $A \cap B = \varnothing$.
A und B sind *rekursiv untrennbar*, wenn es keine entscheidbare Menge $C \subseteq \Gamma^*$ gibt mit $A \subseteq C$ und $B \cap C = \varnothing$.

(Insbesondere sind A und B dann unentscheidbar.)

Seien A, B disjunkte Teilmengen von Γ^*, X, Y disjunkte Teilmengen von Σ^*.

$(A, B) \leq (X, Y)$: es existiert eine totale berechenbare Funktion $f : \Gamma^* \to \Sigma^*$ mit $f(A) \subseteq X$ und $f(B) \subseteq Y$.

Lemma. *Seien A, B rekursiv untrennbar und $(A, B) \leq (X, Y)$. Dann sind auch X, Y rekursiv untrennbar.*

- $A \leq B \iff (A, \overline{A}) \leq (B, \overline{B})$.
- A unentscheidbar gdw. $(A, \overline{A})$ rekursiv untrennbar.

Rekursiv untrennbare Halteprobleme.

$$
\begin{aligned}
H_\lambda^+ &:= \{\rho(M) : M \text{ Akzeptor, } M \text{ akzeptiert } \lambda\} \\
H_\lambda^- &:= \{\rho(M) : M \text{ Akzeptor, } M \text{ verwirft } \lambda\} \\
H_\lambda^\infty &:= \{\rho(M) : M \text{ Akzeptor, die Berechnung von } M \text{ auf } \lambda \\
&\qquad\qquad \text{ist unendlich und nicht-periodisch}\}
\end{aligned}
$$

Satz. *Die Probleme H_λ^+, H_λ^- und H_λ^∞ sind paarweise rekursiv untrennbar.*

FO: Prädikatenlogik erster Stufe (mit Gleichheit)

Abzählbare Signatur $\tau_\infty := \{R^i_j : i, j \in \mathbb{N}\} \cup \{f^i_j : i, j \in \mathbb{N}\}$.

R^i_j Relationssymbol, f^i_j Funktionssymbol der Stelligkeit i.

Wir betrachten nur Formelklassen mit Signatur $\tau \subseteq \tau_\infty$.

Für algorithmische Probleme müssen Formeln als Wörter über festem, endlichem Alphabet aufgefasst (kodiert) werden:

$$\Gamma := \{R, f, x, 0, 1, [,]\} \cup \{=, \neg, \wedge, \vee. \rightarrow, \leftrightarrow, \exists, \forall.(,)\}.$$

Relationssymbole:	R^i_j	$\longmapsto$	$R[\text{bin } i][\text{bin } j]$
Funktionssymbole:	f^i_j	$\longmapsto$	$f[\text{bin } i][\text{bin } j]$
Variablen:	x_j	$\longmapsto$	$x[\text{bin } j]$

$\Longrightarrow$: FO-Formeln beschrieben durch Wörter in Γ^*.

Entscheidungsprobleme für Formelklassen $X \subseteq \mathrm{FO}$:

$$\begin{aligned}
\mathit{Sat}(X) &:= \{\psi \in X : \psi \text{ ist erfüllbar}\} \\
\mathit{Fin\text{-}sat}(X) &:= \{\psi \in X : \psi \text{ hat ein endliches Modell}\} \\
\mathit{Val}(X) &:= \{\psi \in X : \psi \text{ ist eine Tautologie}\} \\
\mathit{Non\text{-}sat}(X) &:= \{\psi \in X : \psi \text{ ist unerfüllbar}\} \\
\mathit{Inf\text{-}axioms}(X) &:= \{\psi \in X : \psi \text{ ist Unendlichkeitsaxiom}\}
\end{aligned}$$

Unendlichkeitsaxiome: erfüllbare Formeln ohne endliche Modelle.

Definition. $X \subseteq \mathrm{FO}$ hat die *Endliche-Modell-Eigenschaft (EME)* wenn X keine Unendlichkeitsaxiome enthält.

D.h.: jede erfüllbare Formel aus X hat auch ein endliches Modell.

Satz. *Sei $X \subseteq \mathrm{FO}$ ist entscheidbar. Dann gilt:*

(1) $\mathit{Val}(X)$ ist r.e.

(2) $\mathit{Non\text{-}sat}(X)$ ist r.e.

(3) $\mathit{Fin\text{-}sat}(X)$ ist r.e.

(4) $\mathit{Sat}(X)$ ist co-r.e.

(5) $\mathit{Inf\text{-}axioms}(X)$ ist co-r.e.

Folgerung. $X \subseteq \mathrm{FO}$ *entscheidbar, X hat EME*
$\Longrightarrow \quad \mathit{Sat}(X), \mathit{Val}(X), \mathit{Fin\text{-}sat}(X), \ldots$ *sind entscheidbar.*

Satz von Trakhtenbrot: *Es gibt eine endliche Signatur $\tau \subsetneq \tau_\infty$, so dass Fin-sat*(FO($\tau$)), *Non-sat*(FO($\tau$)) *und Inf-axioms*(FO($\tau$)) *(paarweise) rekursiv untrennbar sind.*

Effektive Konstruktion $\rho(M) \mapsto \psi_M$ so dass:

(a) M akzeptiert $\lambda \implies \psi_M$ hat ein endliches Modell.

(b) M verwirft $\lambda \implies \psi_M$ ist unerfüllbar.

(c) Die Berechnung von M auf λ ist unendlich und nicht periodisch $\implies \psi_M$ ist ein Unendlichkeitsaxiom.

M Turing-Akzeptor mit Zustandsmenge $Q = \{q_0, \dots, q_r\}$, Anfangszustand q_0, mit Alphabet $\Sigma = \{a_0, \dots, a_s\}$ (wobei $a_0 = \Box$), mit Endzustandsmenge $F = F^+ \cup F^-$ und mit Übergangsfunktion δ.

ψ_M hat Signatur $\tau = \{0, f, q, p, w\}$ wobei 0 Konstante, f, q, p einstellige Funktionen und w zweistellige Funktion. Sei $\boldsymbol{k}$ der Term $f^k 0$.

Intendiertes Modell $\mathfrak{A}_M$ von ψ_M:

- Universum $A = \{0, 1, 2, \dots, n\}$ oder $A = \mathbb{N}$.
- $f(t) = t + 1$, wenn $t + 1 \in A$, und $f(t) = t$, wenn t das letzte Element von A ist.
- $q(t) = i$ genau dann, wenn M zur Zeit t im Zustand q_i.
- $p(t)$ ist die Kopfposition von M zur Zeit t.
- $w(s, t) = i$ genau dann, wenn zur Zeit t auf Feld s das Symbol a_i steht.

$$\psi_M := \text{START} \wedge \text{COMPUTE} \wedge \text{END}$$

$$\text{START} := (q0 = 0 \wedge p0 = 0 \wedge \forall x\ w(x,0) = 0).$$

(M ist zur Zeit 0 in der Inputkonfiguration auf λ)

$$\text{COMPUTE} = \varphi \wedge \bigwedge_{\delta:(q_i,a_j)\mapsto(q_k,a_\ell,m)} \forall y(\alpha_{i,j} \to \beta_{k,\ell,m}).$$

$$\varphi := \forall x \forall y (py \neq x \to w(x, fy) = w(x,y))$$

(Die Bandinschrift auf nicht bearbeiteten Feldern bleibt unverändert)

$\alpha_{i,j} \to \beta_{k,\ell,m}$ formalisiert die von der Übergangsregel
$\delta : (q_i, a_j) \mapsto (q_k, a_\ell, m)$ induzierten Änderungen.

$$\alpha_{ij} := (qy = \boldsymbol{i} \wedge w(py, y) = \boldsymbol{j})$$

(M ist zur Zeit y im Zustand q_i und liest das Symbol a_j)

$$\beta_{k,\ell,m} := (qfy = \boldsymbol{k} \wedge w(py, fy) = \boldsymbol{\ell} \wedge \text{MOVE}_m$$

$$\text{wobei MOVE}_m := \begin{cases} pfy = py & \text{für } m = 0 \\ pfy = fpy & \text{für } m = 1 \\ \exists z(fz = py \wedge pfy = z) & \text{für } m = -1. \end{cases}$$

(Zur Zeit $y + 1$ sind die entsprechenden Änderungen realisiert).

$$\text{END} = \bigwedge_{\substack{\delta(q_i,a_j) \text{ undef.} \\ q_i \notin F^+}} \forall y\, \neg\alpha_{ij}$$

(Die Berechnung kann nur durch eine akzeptierende Konfiguration beendet werden)

- ψ_M ist aus $\rho(M)$ effektiv konstruierbar.
- Wenn M das leere Wort akzeptiert, dann ist das intendierte Modell $\mathfrak{A}_M$ endlich, und $\mathfrak{A}_M \models \psi_M$.
- Wenn M auf λ unendlich lange rechnet, dann ist das intendierte Modell $\mathfrak{A}_M$ unendlich, und $\mathfrak{A}_M \models \psi_M$.

Sei $\mathfrak{B} = (B, 0, f, q, p, w) \models \psi_M$.

Definition. $\mathfrak{B}$ *erzwingt die Konfiguration* (q_i, j, w) *zur Zeit* t (wobei $w = a_{i_0} \dots a_{i_m} \in \Sigma^*$, wenn

(a) $\mathfrak{B} \models q\boldsymbol{t} = \boldsymbol{i}$,

(b) $\mathfrak{B} \models p\boldsymbol{t} = \boldsymbol{j}$,

(c) für alle $k \leq m$, $\mathfrak{B} \models w(\boldsymbol{k}, \boldsymbol{t}) = \boldsymbol{i_k}$ und für alle $k > m$, $\mathfrak{B} \models w(\boldsymbol{k}, \boldsymbol{t}) = 0$.

Da $\mathfrak{B} \models \psi_M$ gilt:

- $\mathfrak{B}$ erzwingt $C_0 = (q_0, 0, \lambda)$ zur Zeit 0
(da $\mathfrak{B} \models$ START.)
- Wenn $\mathfrak{B}$ zur Zeit t die Konfiguration C_t erzwingt, welche *nicht* eine akzeptierrende Konfiguration ist, dann erzwingt $\mathfrak{B}$ zur Zeit $t+1$ die Konfiguration $C_{t+1} = \mathrm{Next}(C_t)$.
- Insbesondere kann die Berechnung von M keine verwerfende Konfiguration erreichen.

Damit ist gezeigt:

Wenn M λ verwirft, ist ψ_M unerfüllbar.

Wenn die Berechnung von M auf unendlich und nicht-periodisch ist, dann ist ψ_M ein Unendlichkeitsaxiom.