4.2. Undecidability

A decision problem is called undecidable, if there does not exist a TM that answers each instance correctly after finitely many steps.

If $L \subseteq \Sigma^*$, then the Membership Problem for L is the following decision problem:

INSTANCE: A word $w \in \Sigma^*$.
QUESTION: Is $w \in L$?

Thus, this problem is decidable iff the language L is recursive.

In what follows we are interested in the Halting Problem for TMs:

INSTANCE: A TM M and an input word $w \in \Sigma^*$.
QUESTION: When starting with input w, will M halt eventually?

In order to study this problem we must encode the instance (M, w) in some way.

Let $M = (Q, \{0, 1\}, \{0, 1, \Box\}, \Box, \delta, q_0, q_n)$ be a 1-TM on $\Sigma = \{0, 1\}$, and let $Q = \{q_0, q_1, \ldots, q_n\}$.

We will encode M through a word $c(M) \in \Sigma^+$.

Let $\delta = \{(q_{i_1}, a_{i_1}, q_{j_1}, a_{j_1}, m_{j_1}), \ldots, (q_{i_m}, a_{i_m}, q_{j_m}, a_{j_m}, m_{j_m})\}$, where $q_{i_e}, q_{j_e} \in Q$, $a_{i_e}, a_{j_e} \in \Sigma \cup \{\Box\}$, and $m_{j_e} \in \{L, 0, R\}$.

Each 5-tuple $(q_{i_e}, a_{i_e}, q_{j_e}, a_{j_e}, m_{j_e})$ is encoded as

$$c(q_{i_e}, a_{i_e}, q_{j_e}, a_{j_e}, m_{j_e}) := 0^{i_e+1}10^{e(a_{i_e})}10^{j_e+1}10^{e(a_{j_e})}10^{e(m_{j_e})},$$

where $e(a_i) := \left\{ \begin{array}{ll} 1, & \text{if } a_i = 0, \\ 2, & \text{if } a_i = 1, \\ 3, & \text{if } a_i = \Box, \end{array} \right\}$ and $e(m) := \left\{ \begin{array}{ll} 1, & \text{if } m = L, \\ 2, & \text{if } m = 0, \\ 3, & \text{if } m = R. \end{array} \right\}$

The function δ is interpreted as a sequence of 5-tuples. Assuming that this sequence is sorted in lexicographical order, we take

$$c(M) \;:=\; 1110^{n+1}11111 \cdot c(q_{i_1}, a_{i_1}, q_{j_1}, a_{j_1}, m_{j_1}) \cdot 11 \cdot \\ \ldots \cdot 11 \cdot c(q_{i_m}, a_{i_m}, q_{j_m}, a_{j_m}, m_{j_m}) \cdot 111.$$

Lemma 4.9

The set

$$\{\, c(M) \mid M \text{ is a 1-TM on } \Sigma = \{0,1\} \text{ and } \Gamma = \{0,1,\Box\} \,\}$$

is recursive.

Proof of Lemma 4.9.

Let $w \in \{0,1\}^*$. If w is not an element of the regular language

$$1^3 \cdot 0^{n+1} \cdot 1^5 \cdot (0^{1 \le i \le n} \cdot 1 \cdot 0^{1 \le i \le 3} \cdot 1 \cdot 0^{1 \le i \le n+1} \cdot 1 \cdot 0^{1 \le i \le 3} \cdot 1 \cdot 0^{1 \le i \le 3} \cdot 11)^{\le 3 \cdot n} \cdot 1,$$

then w is not the encoding of a 1-TM.

If, however, w is an element of the above regular language, then one can try to reconstruct M from w. This reconstruction is successful iff w describes a function

$$\delta : \{q_0, \ldots, q_{n-1}\} \times \{0, 1, \Box\} \rightsquigarrow \{q_0, \ldots, q_n\} \times \{0, 1, \Box\} \times \{L, 0, R\}.$$

This function δ then yields the TM M satisfying $c(M) = w$. □

By M_∞ we denote the following TM, which does not halt on any input:

$$(\{q_0, q_1\}, \{0, 1\}, \{0, 1, \Box\}, \Box, \{(q_0, a, q_0, a, 0) \mid a \in \{0, 1, \Box\}\}, q_0, q_1).$$

With each word $w \in \Sigma^*$, we now associate a TM M_w:

$$M_w := \begin{cases} M, & \text{if } c(M) = w, \\ M_\infty, & \text{if } w \text{ is not the encoding of any TM.} \end{cases}$$

By Lemma 4.9, the TM M_w can be reconstructed from w.

Now let $K \subseteq \{0, 1\}^*$ be the following language:

$$K := \{\, w \in \{0, 1\}^* \mid M_w \text{ halts on input } w \,\}.$$

Theorem 4.10

The language K is not recursive.

Proof of Theorem 4.10.

Assume to the contrary that K is recursive. Then there exists a 1-TM M_0 that decides membership in K, that is,

$$q_0^{(0)} w \vdash_{M_0}^{*} q_1^{(0)} 1, \text{ if } w \in K,$$

and

$$q_0^{(0)} w \vdash_{M_0}^{*} q_1^{(0)} 0, \text{ if } w \notin K.$$

By modifying M_0 we obtain a new TM M_1 that behaves as follows:

$$q_0^{(0)} w \vdash_{M_0}^{*} q_1^{(0)} a \vdash_{M_1} \begin{cases} q_2 a \vdash_{M_1} q_2 a \vdash_{M_1} \cdots, & \text{if } a = 1, \\ q_1 0, & \text{if } a = 0. \end{cases}$$

Proof of Theorem 4.10 (cont.)

Hence, for all $w \in \Sigma^*$: M_1 halts on input w iff $q_0^{(0)} w \vdash_{M_0}^* q_1^{(0)} 0$, that is, iff $w \notin K$.

Now let $u := c(M_1)$. Then $M_u = M_1$, and we have the following sequence of equivalent statements:

M_1 halts on input u iff $u \notin K$
iff M_u does not halt on input u
iff M_1 does not halt on input u, a contradiction!

This contradiction shows that the language K is not recursive. □

Corollary 4.11

$K \in \mathrm{RE} \smallsetminus \mathrm{REC}$ and $K^c \notin \mathrm{RE}$.

Corollary 4.12

The Halting Problem for TMs is undecidable.

Proof.

Let H be the following language:

$$H := \{ (w, u) \mid M_w \text{ halts on input } u \}.$$

Then $w \in K$ iff $(w, w) \in H$.

If H were recursive, then K would be recursive, too.

Thus, H is not recursive, that is,
the Halting Problem for (1-)TMs is undecidable. □

Let $\mathcal{S}$ be a set of recursively enumerable languages on $\{0, 1\}$.
We interpret $\mathcal{S}$ as a property of recursively enumerable languages.

We say that a language L has property $\mathcal{S}$, if $L \in \mathcal{S}$.

The property $\mathcal{S}$ is called trivial, if $\mathcal{S} = \emptyset$ or $\mathcal{S} = \text{RE}(\{0, 1\})$.

Finally, let $L_{\mathcal{S}} := \{\, c(M) \mid L(M) \in \mathcal{S} \,\}$.

Theorem 4.13 (Rice 1953)

The language $L_{\mathcal{S}}$ is non-recursive for each non-trivial property $\mathcal{S}$ of recursively enumerable languages, that is, given a TM M, it is in general undecidable whether the language $L(M)$ has property $\mathcal{S}$.

Proof of Theorem 4.13.

W.l.o.g. we can assume that $\emptyset \notin \mathcal{S}$, as otherwise we could consider the set $\mathcal{S}^c := \text{RE}(\{0,1\}) \smallsetminus \mathcal{S}$ instead of $\mathcal{S}$.

As $\mathcal{S}$ is non-trivial, there exists a language $\emptyset \neq L \in \mathcal{S}$.

Let M_L be a TM such that $L(M_L) = L$.

Assume that the language $\mathcal{S}$ is decidable, that is, $L_{\mathcal{S}} \in \text{REC}(\{0,1\})$.
Then there is a TM $M_{\mathcal{S}}$ for deciding $L_{\mathcal{S}}$.
From M_L and $M_{\mathcal{S}}$, we now construct a TM for the halting problem H.

Let M be a TM, and let $w \in \{0,1\}^*$ be an input word.

From M and w, we can construct a TM $M'_{M,w}$ that, on input $x \in \{0,1\}^*$, executes the following program:

(1) simulate M on input w;

(2) **if** M halts on input w **then** simulate M_L on input x.

Proof of Theorem 4.13 (cont.)

Then $L(M'_{M,w}) = \begin{cases} \emptyset, & \text{if } w \notin L(M), \\ L, & \text{if } w \in L(M). \end{cases}$

By our hypothesis, $\emptyset \notin \mathcal{S}$ and $L \in \mathcal{S}$.
Hence, $c(M'_{M,w}) \in L_{\mathcal{S}}$ iff $w \in L(M)$.
Thus, the TM $M_{\mathcal{S}}$ accepts on input $c(M'_{M,w})$ iff $w \in L(M)$, and otherwise, $M_{\mathcal{S}}$ rejects this input.
It follows that the TM $M_{\mathcal{S}}$ decides membership in H.
As H is undecidable, this is a contradiction!
Hence, $L_{\mathcal{S}}$ is non-recursive. □

Corollary 4.14

The following properties are undecidable for recursively enumerable languages:

- *emptiness,*
- *finiteness,*
- *regularity,*
- *context-freeness.*

Let $M = (Q, \Sigma, \Gamma, \Box, \delta, q_0, q_1)$ be a 1-TM, and
let $\Delta := \Gamma \,\dot\cup\, Q \,\dot\cup\, \{\#\}$, where $\#$ is an additional symbol.

A valid computation of M is a word of the form

$$w = w_1 \# w_2^R \# w_3 \# w_4^R \cdots \# w_{2m}^R \# (w_{2m+1} \#)^\mu \in \Delta^+,$$

where $\mu \in \{0, 1\}$ and $n := \left\{ \begin{array}{ll} 2m, & \text{if } \mu = 0 \\ 2m+1, & \text{if } \mu = 1 \end{array} \right\}$,

that satisfies the following conditions:

(1) $\forall i = 1, 2, \ldots, n :\; w_i \in \Gamma^* \cdot Q \cdot \Gamma^*$, where w_i does not end with the symbol $\Box$;

(2) $w_1 = q_0 x$ for some $x \in \Sigma^*$, that is,
w_1 is an initial configuration of M;

(3) $w_n \in \Gamma^* \cdot q_1 \cdot \Gamma^*$, that is, w_n is a halting configuration of M;

(4) $\forall i = 1, 2, \ldots, n-1 : w_i \vdash_M w_{i+1}$.

By $\mathrm{GB}(M)$ we denote the language on Δ that consists of all valid computations of M.

Lemma 4.15

From a given 1-TM M, one can effectively construct two context-free grammars G_1 and G_2 such that $L(G_1) \cap L(G_2) = \mathrm{GB}(M)$.

Proof.

Let L_3 be the language

$$L_3 := \{\, y\#z^R \mid y, z \in \Gamma^* \cdot Q \cdot \Gamma^* \text{ such that } y \vdash_M z \,\}.$$

From M one can easily construct a PDA that accepts L_3.

From L_3 we obtain the language L_1:

$$L_1 := (L_3 \cdot \#)^* \cdot (\{\varepsilon\} \cup (\Gamma^* \cdot q_1 \cdot \Gamma^* \cdot \#)).$$

From M we can construct a context-free grammar for the language L_1 (Theorem 3.20, Theorem 3.22).

Proof of Lemma 4.15 (cont.)

Further, let L_4 be the language

$$L_4 := \{ y^R\#z \mid y, z \in \Gamma^* \cdot Q \cdot \Gamma^* \text{ such that } y \vdash_M z \},$$

and let L_2 be obtained from L_4 as follows:

$$L_2 := q_0\Sigma^* \cdot \# \cdot (L_4 \cdot \#)^* \cdot (\{\varepsilon\} \cup (\Gamma^* \cdot q_1 \cdot \Gamma^* \cdot \#)).$$

From M we can construct a context-free grammar for L_2.

Claim.

$L_1 \cap L_2 = \mathrm{GB}(M)$.

Proof of Lemma 4.15 (cont.)

Proof of Claim.

Let $w = w_1 \# w_2^R \# \cdots \# w_n \#$ such that $n \equiv 1 \bmod 2$.

If $w \in \mathrm{GB}(M)$, then properties (1) to (4) imply that $w \in L_1 \cap L_2$.

Conversely, if $w \in L_1 \cap L_2$, then we see from the definitions of L_1 and L_2 that w satisfies (1) and (4).

As $w \in L_2$, $w_1 = q_0 x$ for some $x \in \Sigma^*$, and

as $w \in L_1$, $w_n \in \Gamma^* \cdot q_1 \cdot \Gamma^*$, that is, $w \in \mathrm{GB}(M)$.

For $w = w_1 \# \cdots \# w_n^R \#$ such that $n \equiv 0 \bmod 2$, the proof is analogous.

Thus, $L_1 \cap L_2 = \mathrm{GB}(M)$. □

□

Let M be a 1-TM.

Then $L(M) \neq \emptyset$ iff $\mathrm{GB}(M) \neq \emptyset$.

Now let G_1 and G_2 be two context-free grammars such that $L(G_1) \cap L(G_2) = \mathrm{GB}(M)$.

Then $L(M) \neq \emptyset$ iff $L(G_1) \cap L(G_2) \neq \emptyset$.

As emptiness is undecidable for $L(M)$, this yields the following result.

Corollary 4.16

The following Intersection Emptiness Problem is undecidable:

INSTANCE: *Two context-free grammars G_1 and G_2.*
QUESTION: *Is $L(G_1) \cap L(G_2) = \emptyset$?*

The set $\Delta^* \setminus \mathrm{GB}(M) = \mathrm{GB}(M)^c$ is called the set of invalid computations of M.

Lemma 4.17

For each 1-TM M, $\mathrm{GB}(M)^c \in \mathsf{CFL}(\Delta)$.

As $L(M) = \emptyset$ iff $\mathrm{GB}(M)^c = \Delta^*$, we obtain the following undecidability result.

Corollary 4.18

The following Universality Problem is undecidable:

INSTANCE: A context-free grammar G on Δ.
QUESTION: Is $L(G) = \Delta^*$?

Theorem 4.19

The following problems are undecidable:

(1) *INSTANCE:* *Two context-free grammars* G_1 *and* G_2.

- *QUESTION:* *Is* $L(G_1) = L(G_2)$?
- *QUESTION:* *Is* $L(G_1) \subseteq L(G_2)$?
- *QUESTION:* *Is* $L(G_1) \cap L(G_2)$ *context-free?*
- *QUESTION:* *Is* $L(G_1) \cap L(G_2)$ *regular?*

(2) *INSTANCE:* *A context-free grammar* G *and a regular set* R.

- *QUESTION:* *Is* $L(G) = R$?
- *QUESTION:* *Is* $R \subseteq L(G)$?

(3) *INSTANCE:* *A context-free grammar* G.

- *QUESTION:* *Is* $L(G)^c$ *context-free?*
- *QUESTION:* *Is* $L(G)^c$ *regular?*

Proof.

Let G_1 be a context-free grammar s.t. $L(G_1) = R = \Sigma^*$.
Then the following holds for each context-free grammar G_2:

$$R = L(G_1) = L(G_2) \text{ iff } R = L(G_1) \subseteq L(G_2) \text{ iff } L(G_2) = \Sigma^*.$$

It follows from Corollary 4.18 that the first two problems of (1) and the two problems of (2) are undecidable.

The language $\mathrm{GB}(M)$ is finite and therewith regular, if $L(M)$ is finite; on the other hand, if $L(M)$ is infinite, then $\mathrm{GB}(M)$ is not even context-free, which can be shown by the Pumping Lemma (Theorem 3.14), if M makes at least 3 steps on each input.

Proof of Theorem 4.19 (cont.)

Let M be an arbitrary 1-TM. From M one can construct a 1-TM M' that accepts the same language as M, but that executes at least 3 steps on each input.

Now $L(M)$ is finite iff $\mathrm{GB}(M') = (\mathrm{GB}(M')^c)^c$ is context-free (regular).

Further, from M' we obtain two context-free grammars G_1 and G_2 such that $L(G_1) \cap L(G_2) = \mathrm{GB}(M')$.

As finiteness of $L(M)$ is undecidable, it follows that the questions of whether $(\mathrm{GB}(M')^c)^c$ or $L(G_1) \cap L(G_2)$ are context-free (regular) are undecidable, too. □