Automata and Grammars

SS 2018

Remarks 1: How to Prove $L = L(G)$

Seminary: Thursday, February 22, 2018.

Let $L \subseteq \Sigma^*$ be a language, and let $G = (N, \Sigma, P, S)$ be a grammar. We want to prove that $L = L(G)$. How do we do that in a systematic way? Here we present such a way using some simple examples. However, we first show how to prove a property concerning words.

Lemma 1. [Words]
Let $u, v, z \in \Sigma^*$ be three words such that $u \neq \varepsilon$ and $uv = vz$. Prove that there exist two words $x, y \in \Sigma^*$ and an integer $k \geq 0$ such that $u = xy$, $v = (xy)^k x$, and $z = yx$.

Proof. We proceed by induction on $|v|$:

- Basis of the induction: $|v| \leq |u|$: Then $uv = vz$ implies that $u = vw$ and $z = wv$ for some word $w \in \Sigma^*$. Choosing $x = v$, $y = w$, and $k = 0$, we obtain that $u = vw = xy$, $v = (xy)^k x$, and $z = wv = yx$.
- Induction hypothesis: Assume that the statement holds for all v up to some positive length $n \geq |u|$.
- Inductive claim: The statement also holds for all v of length $n + 1$.
- Inductive step: As $|v| = n + 1 > n \geq |u|$, we see that $v = ur$ and $v = rz$ for some $r \in \Sigma^+$, and so $uur = uv = urz$. This yields that $ur = rz$. Since $u \neq \varepsilon$, we see that $|r| < |v|$. Hence, the induction hypothesis applies to the equation $ur = rz$. It implies that $u = xy$, $r = (xy)^k x$, and $z = yx$ for some words $x, y \in \Sigma^*$ and an integer $k \geq 0$. Hence, $v = rz = (xy)^k xyx = (xy)^{k+1} x$. $\square$

Now we turn to the announced problem of proving $L(G) = L$ for a given grammar G and a language L.

Example 1. Let $G = (N, \Sigma, P, S)$, where $N = \{A, B\}$, $\Sigma = \{a\}$, $S = A$, and P contains the following productions: $A \to a, A \to aB, B \to aA$.

Theorem. $L(G) = \{\, a^{2n+1} \mid n \geq 0 \,\}$.

Let $L = \{\, a^{2n+1} \mid n \geq 0 \,\}$. Thus, we need to prove that $L = L(G)$. As L and $L(G)$ are languages, that is, sets of words, we do this by showing the two inclusions $L \subseteq L(G)$ and $L(G) \subseteq L$.

Proof of $L(G) \subseteq L$. We will proceed by induction on the length n of a derivation $S \to_G^n w$. Actually, we will show a stronger statement than just that $w \in L$, because such a stronger statement yields a stronger induction hypothesis that we can then use in the inductive step. Here we consider the set $\hat{L}(G)$ of all sentential forms generated by G, that is,

$$\hat{L}(G) = \{\, \alpha \in (N \cup \Sigma)^* \mid S \to_G^* \alpha \,\},$$

and we claim that $\hat{L}(G) \subseteq L'$, where

$$L' = \{\, a^{2n+1} \mid n \geq 0 \,\} \cup \{a^{2n} A \mid n \geq 0\,\} \cup \{a^{2n+1} B \mid n \geq 0\,\}.$$

This claim immediately yields $L(G) \subseteq L$, as $L(G) = \hat{L}(G) \cap \Sigma^* \subseteq L' \cap \Sigma^* = L$.

Claim 1. $\hat{L}(G) \subseteq L'$.

Proof of Claim 1: By induction on n, we show that $A \to_G^n \alpha$ implies that $\alpha \in L'$.

- Basis of induction. $n = 0$: $A \to_G^0 \alpha$ implies $\alpha = A$, which satisfies $A = a^{2\cdot 0}A \in L'$.
- Induction hypothesis: For some $n \geq 0$, if $A \to_G^n \alpha$, then $\alpha \in L'$.
- Inductive claim: If $A \to_G^{n+1} \alpha$, then $\alpha \in L'$.
- Inductive step: Assume that $A \to_G^{n+1} \alpha$. Hence, there exists a sentential form γ such that $A \to_G^n \gamma \to_G \alpha$. By the induction hypothesis, $\gamma \in L'$. Based on the structure of the set L', we now distinguish between three cases:
  1. $\gamma = a^{2m+1}$ for some $m \geq 0$. Then γ is a terminal word, and hence, no production can be applied to it. Thus, this case cannot occur.
  2. $\gamma = a^{2m}A$ for some $m \geq 0$. Then $\gamma \to_G \alpha$ implies that $\alpha = a^{2m+1}$ or $\alpha = a^{2m+1}B$. In either case, $\alpha \in L'$.
  3. $\gamma = a^{2m+1}B$ for some $m \geq 0$. Then $\gamma \to_G \alpha$ implies that $\alpha = a^{2m+2}A$, which belongs to L'.

This completes the proof of Claim 1. □

Claim 2. $L \subseteq L(G)$.

Proof of Claim 2: By induction on n, we show that $a^{2n+1} \in L(G)$.

- Basis of induction: $n = 0$: The word $a^{2n+1} = a^1 = a \in L(G)$, as $A \to_G a$ holds.
- Induction hypothesis: For some $n \geq 0$, $a^{2n+1} \in L(G)$.
- Inductive claim: $a^{2(n+1)+1} = a^{2n+3} \in L(G)$.
- Inductive step: As $a^{2n+1} \in L(G)$, we have a derivation $A \to_G^* a^{2n+1}$. Hence, we obtain a derivation $A \to_G aB \to_G aaA \to_G^* aaa^{2n+1} = a^{2n+3}$, which shows that $a^{2n+3} \in L(G)$.

This complete the proof of Claim 2, and therewith the proof of the theorem. □

In more involved cases it is often useful to consider all sets of the form

$$L(G, C) = \{\, w \in \Sigma^* \mid C \to_G^* w \,\} \quad (C \in N).$$

For the example above, we would get the claim that,

$$\text{for all } n \geq 0,\ a^{2n+1} \in L(G, A) \text{ and } a^{2n+2} \in L(G, B).$$

We prove this by induction on n:

- Basis of induction: $n = 0$: The word $a^{2n+1} = a^1 = a \in L(G, A)$, as $A \to_G a$ holds, and $a^{2n+2} = a^2 \in L(G, B)$, as $B \to_G aA \to_G aa$.
- Induction hypothesis: For some $n \geq 0$, $a^{2n+1} \in L(G, A)$ and $a^{2n+2} \in L(G, B)$.
- Inductive claim: $a^{2(n+1)+1} = a^{2n+3} \in L(G, A)$ and $a^{2(n+1)+2} = a^{2n+4} \in L(G, B)$.
- Inductive step: As $a^{2n+1} \in L(G, A)$, we have a derivation $A \to_G^* a^{2n+1}$, and as $a^{2n+2} \in L(G, B)$, we have a derivation $B \to_G^* a^{2n+2}$. Hence, we obtain a derivation $A \to_G aB \to_G^* aa^{2n+2} = a^{2n+3}$, which shows that $a^{2n+3} \in L(G, A)$, and we obtain a derivation $B \to_G aA \to_G^* aa^{2n+3} = a^{2n+4}$, which shows that $a^{2n+4} \in L(G, B)$.

This complete the proof of the above claim. □

Example 2. Let $L = \{ w \in \{a,b,c\}^* \mid |w|_a \equiv 0 \bmod 2 \}$. Determine a regular grammar that generates the language L!

(a) We first present a regular grammar $G = (N, \Sigma, P, S)$:

- $N = \{S, E, U\}$, $\Sigma = \{a, b, c\}$, and
- P contains the following productions:

$$\begin{array}{l} S \to aU, S \to b, S \to bE, S \to c, S \to cE, S \to \varepsilon, \\ E \to aU, E \to b, E \to bE, E \to c, E \to cE, \\ U \to a, U \to aE, U \to bU, U \to cU. \end{array}$$

(b) **Claim.** $L(G) = L$.

Proof. We will prove the following two statements:

$$\begin{array}{lcl} L(G,E) & = & \{ w \in \Sigma^+ \mid |w|_a \equiv 0 \mod 2 \} \text{ and} \\ L(G,U) & = & \{ w \in \Sigma^+ \mid |w|_a \equiv 1 \mod 2 \}. \end{array}$$

Assume first that the above statements have already been shown. Then

$$\begin{array}{rcl} L(G) = L(G,S) & = & \{\varepsilon, b, c\} \cup \{ aw \mid w \in L(G,U) \} \cup \{bw, cw \mid w \in L(G,E) \} \\ & = & \{\varepsilon, b, c\} \cup \{ aw \mid |w|_a \equiv 1 \mod 2 \} \cup \{ bw, cw \mid |w|_a \equiv 0 \mod 2 \} \\ & = & \{\varepsilon, b, c\} \cup \{ aw \mid |aw|_a \equiv 0 \mod 2 \} \cup \{ bw \mid |bw|_a \equiv 0 \mod 2 \} \\ & & \cup \{ cw \mid |cw|_a \equiv 0 \mod 2 \} \\ & = & \{ w \in \Sigma^* \mid |w|_a \equiv 0 \mod 2 \} = L. \end{array}$$

Now we prove that $L(G,E) \subseteq \{ w \in \Sigma^+ \mid |w|_a \equiv 0 \bmod 2 \}$ and that $L(G,U) \subseteq \{ w \in \Sigma^+ \mid |w|_a \equiv 1 \bmod 2 \}$. Since the empty word cannot be derived from E or from U, it suffices to consider non-empty words. Thus, we actually prove the following statement:

$$\begin{array}{ll} \forall n \geq 1 : (\forall w \in \Sigma^+ : & (\text{if } E \to_G^n w, \text{ then } |w|_a \equiv 0 \mod 2) \text{ and} \\ & (\text{if } U \to_G^n w, \text{ then } |w|_a \equiv 1 \mod 2)). \end{array}$$

We proceed by induction on n.

- Basis of induction: $n = 1$: If $E \to_G w$, then $w = b$ or $w = c$. Hence, $|w|_a \equiv 0 \bmod 2$. Further, if $U \to_G w$, then $w = a$. Hence, $|w|_a \equiv 1 \bmod 2$.
- Induction hypothesis: Assume that the above statement holds for some $n \geq 1$.
- Inductive claim: The statement also holds for $n + 1$.
- Inductive step: Assume that $E \to_G^{n+1} w$. Then there exists a sentential form $\alpha \in (N \cup \Sigma)^+ \smallsetminus \Sigma^+$ such that $E \to_G \alpha \to_G^n w$, and $|w| = n + 1$. We distinguish between three cases based on the production used in the first step $E \to_G \alpha$.
  (a) If $\alpha = aU$, then $w = av$ and $U \to_G^n v$. From the induction hypothesis we see that $|v|_a \equiv 1 \bmod 2$, which implies that $|w|_a = |av|_a \equiv 0 \bmod 2$.
  (b) If $\alpha = bE$, then $w = bv$ and $E \to_G^n v$. From the indction hypothesis we see that $|v|_a \equiv 0 \bmod 2$, which implies that $|w|_a = |bv|_a \equiv 0 \bmod 2$.
  (c) If $\alpha = cE$, then it follows analogously to (b) that $|w|_a \equiv 0 \bmod 2$.

Assume that $U \to_G^{n+1} w$. Then there exists a sentential form α such that $U \to_G \alpha \to_G^n w$. Here we again distinguish between three cases based on the production used in the first step $U \to_G \alpha$. Here $\alpha = aE$, $\alpha = bU$, or $\alpha = cU$, and these cases are dealt with analogously.

Now we prove that $\{\, w \in \Sigma^+ \mid |w|_a \equiv 0 \bmod 2 \,\} \subseteq L(G, E)$ and that $\{\, w \in \Sigma^+ \mid |w|_a \equiv 1 \bmod 2 \,\} \subseteq L(G, U)$. Thus, we actually prove the following statement:

$$\begin{aligned} \forall n \geq 1 : (\forall w \in \Sigma^n : \ & (\text{if } |w|_a \equiv 0 \mod 2, \text{ then } E \to_G^* w) \text{ and} \\ & (\text{if } |w|_a \equiv 1 \mod 2, \text{ then } U \to_G^* w)). \end{aligned}$$

We proceed by induction on n.

- Basis of induction: $n = 1$: Then $w \in \Sigma$. If $|w|_a \equiv 0 \bmod 2$, then $w = b$ or $w = c$, and we see that $E \to_G w$. If $|w|_a \equiv 1 \bmod 2$, then $w = a$, and we see that $U \to_G w$.
- Induction hypothesis: The statement above holds for some $n \geq 1$.
- Inductive claim: The statement holds also for $n + 1$.
- Inductive step: Let $w \in \Sigma^{n+1}$. Assume first that $|w|_a \equiv 0 \bmod 2$. We must show that $w \in L(G, E)$. We distinguish between three cases based on the first letter of w.

  (a) $w = av$: Then $|v| = n$ and $|v|_a \equiv 1 \bmod 2$. By the induction hypothesis we have $U \to_G^* v$. Hence, we obtain a derivation $E \to_G aU \to_G^* av = w$.

  (b) $w = bv$: Then $|v| = n$ and $|v|_a \equiv 0 \bmod 2$. By the induciton hypothesis we have $E \to_G^* v$. Hence, we obtain a derivation $E \to_G bE \to_G^* bv = w$.

  (c) $w = cv$: Analogously to (b) it follows that $E \to_G^* cv = w$.

  Assume now that $|w|_a \equiv 1 \bmod 2$. We must show that $w \in L(G, U)$. This is done analogously to the previous case. Thus, we have shown that $L(G, E) = \{\, w \in \Sigma^+ \mid |w|_a \equiv 0 \bmod 2 \,\}$ and that $L(G, U) = \{\, w \in \Sigma^+ \mid |w|_a \equiv 1 \bmod 2 \,\}$. □