4.5: Proving the Correctness of Grammars

In this section, we consider techniques for proving the correctness of grammars, i.e., for proving that grammars generate the languages we want them to.

Definition of Π

Suppose G is a grammar and $a \in Q_G \cup \textbf{alphabet}\, G$. Then $\Pi_{G,a} = \{\, w \in (\textbf{alphabet}\, G)^* \mid w \text{ is parsable from } a \text{ using } G \,\}$.

If it's clear which grammar we are talking about, we often abbreviate $\Pi_{G,a}$ to Π_a.

Clearly, $\Pi_{G,s_G} = L(G)$.

For example, if G is the grammar

$$\mathsf{A} \to \% \mid \mathsf{0A1}$$

then $\Pi_0 = \{0\}$, $\Pi_1 = \{1\}$ and $\Pi_{\mathsf{A}} = \{\, 0^n 1^n \mid n \in \mathbb{N} \,\} = L(G)$.

Properties of Π

Proposition 4.5.1

Suppose G is a grammar.

(1) *For all $a \in$ **alphabet** G, $a \in \Pi_{G,a}$.*

(2) *For all $q \in Q_G$, if $q \to \% \in P_G$, then $\% \in \Pi_{G,q}$.*

(3) *For all $q \in Q_G$, $n \in \mathbb{N} - \{0\}$, $a_1, \ldots, a_n \in$ **Sym** and $w_1, \ldots, w_n \in$ **Str**, if $q \to a_1 \cdots a_n \in P_G$ and $w_1 \in \Pi_{G,a_1}$, $\ldots$, $w_n \in \Pi_{G,a_n}$, then $w_1 \cdots w_n \in \Pi_{G,q}$.*

Main Example

Define $\mathbf{diff} \in \{0,1\}^* \to \mathbb{Z}$ by: for all $w \in \{0,1\}^*$,

$$\mathbf{diff}\, w = \text{the number of } 1\text{'s in } w - \text{the number of } 0\text{'s in } w.$$

Then:

- $\mathbf{diff}\, \% = 0$;
- $\mathbf{diff}\, 1 = 1$;
- $\mathbf{diff}\, 0 = -1$; and
- for all $x, y \in \{0,1\}^*$, $\mathbf{diff}(xy) = \mathbf{diff}\, x + \mathbf{diff}\, y$.

Main Example

Our main example will be the grammar G:

$$\mathsf{A} \to \% \mid \mathsf{0BA} \mid \mathsf{1CA},$$
$$\mathsf{B} \to \mathsf{1} \mid \mathsf{0BB},$$
$$\mathsf{C} \to \mathsf{0} \mid \mathsf{1CC}.$$

Let

$$X = \{\, w \in \{0,1\}^* \mid \textbf{diff}\ w = 0 \,\},$$
$$Y = \{\, w \in \{0,1\}^* \mid \textbf{diff}\ w = 1 \text{ and, for all proper prefixes } v \text{ of } w, \textbf{diff}\ v \le 0 \,\},$$
$$Z = \{\, w \in \{0,1\}^* \mid \textbf{diff}\ w = -1 \text{ and, for all proper prefixes } v \text{ of } w, \textbf{diff}\ v \ge 0 \,\}.$$

We will prove that $L(G) = \Pi_{G,\mathsf{A}} = X$, $\Pi_{G,\mathsf{B}} = Y$ and $\Pi_{G,\mathsf{C}} = Z$.

Main Example

Lemma 4.5.2

Suppose $x \in \{0,1\}^*$.

(1) *If* $\mathbf{diff}\, x \geq 1$, *then* $x = yz$ *for some* $y, z \in \{0,1\}^*$ *such that* $y \in Y$ *and* $\mathbf{diff}\, z = \mathbf{diff}\, x - 1$.

(2) *If* $\mathbf{diff}\, x \leq -1$, *then* $x = yz$ *for some* $y, z \in \{0,1\}^*$ *such that* $y \in Z$ *and* $\mathbf{diff}\, z = \mathbf{diff}\, x + 1$.

Proof. We show the proof of (1), the proof of (2) being similar.

Let $y \in \{0,1\}^*$ be the shortest prefix of x such that $\textbf{diff}\, y \geq 1$, and let $z \in \{0,1\}^*$ be such that $x = yz$.

Because $\textbf{diff}\, y \geq 1$, we have that $y \neq \%$. Thus $y = y'a$ for some $y' \in \{0,1\}^*$ and $a \in \{0,1\}$. By the definition of y, we have that $\textbf{diff}\, y' \leq 0$. Suppose, toward a contradiction, that $a = 0$. Since $\textbf{diff}\, y' + -1 = \textbf{diff}\, y \geq 1$, we have that $\textbf{diff}\, y' \geq 2$, contradicting the definition of y. Thus $a = 1$, so that $y = y'1$.

Because $\textbf{diff}\, y' + 1 = \textbf{diff}\, y \geq 1$, we have that $\textbf{diff}\, y' \geq 0$. Thus $\textbf{diff}\, y' = 0$, so that $\textbf{diff}\, y = \textbf{diff}\, y' + 1 = 1$. By the definition of y, every prefix of y' has a diff that is ≤ 0. Thus $y \in Y$.

Finally, because $\textbf{diff}\, x = \textbf{diff}\, y + \textbf{diff}\, z = 1 + \textbf{diff}\, z$ we have that $\textbf{diff}\, z = \textbf{diff}\, x - 1$. $\square$

Proving that Enough is Generated

First we study techniques for showing that everything we want a grammar to generate is really generated.

Since $X, Y, Z \subseteq \{0,1\}^*$, to prove that $X \subseteq \Pi_{G,\mathsf{A}}$, $Y \subseteq \Pi_{G,\mathsf{B}}$ and $Z \subseteq \Pi_{G,\mathsf{C}}$, it will suffice to use strong string induction to show that, for all $w \in \{0,1\}^*$:

(A) if $w \in X$, then $w \in \Pi_{G,\mathsf{A}}$;

(B) if $w \in Y$, then $w \in \Pi_{G,\mathsf{B}}$; and

(C) if $w \in Z$, then $w \in \Pi_{G,\mathsf{C}}$.

Enough is Generated in Example

We proceed by strong string induction. Suppose $w \in \{0,1\}^*$, and assume the inductive hypothesis: for all $x \in \{0,1\}^*$, if x is a proper substring of w, then:

- *(A)* if $x \in X$, then $x \in \Pi_{\mathsf{A}}$;
- *(B)* if $x \in Y$, then $x \in \Pi_{\mathsf{B}}$; and
- *(C)* if $x \in Z$, then $x \in \Pi_{\mathsf{C}}$.

We must prove that:

- *(A)* if $w \in X$, then $w \in \Pi_{\mathsf{A}}$;
- *(B)* if $w \in Y$, then $w \in \Pi_{\mathsf{B}}$; and
- *(C)* if $w \in Z$, then $w \in \Pi_{\mathsf{C}}$.

Enough is Generated in Example

(A) Suppose $w \in X$. We must show that $w \in \Pi_{\mathsf{A}}$. There are three cases to consider.

- Suppose $w = \%$. Because $\mathsf{A} \to \% \in P$, we have that $w = \% \in \Pi_{\mathsf{A}}$.
- Suppose $w = 0x$, for some $x \in \{0, 1\}^*$. Because $-1 + \textbf{diff}\, x = \textbf{diff}\, w = 0$, we have that $\textbf{diff}\, x = 1$. Thus, by Lemma 4.5.2(1), we have that $x = yz$, for some $y, z \in \{0, 1\}^*$ such that $y \in Y$ and $\textbf{diff}\, z = \textbf{diff}\, x - 1 = 1 - 1 = 0$. Thus $w = 0yz$, $y \in Y$ and $z \in X$. We have $0 \in \Pi_0$. Because $y \in Y$ and $z \in X$ are proper substrings of w, parts (B) and (A) of the inductive hypothesis tell us that $y \in \Pi_{\mathsf{B}}$ and $z \in \Pi_{\mathsf{A}}$. Thus, because $\mathsf{A} \to 0\mathsf{BA} \in P$, it follows that that $w = 0yz \in \Pi_{\mathsf{A}}$.
- Suppose $w = 1x$, for some $x \in \{0, 1\}^*$. The proof is analogous to the preceding case.

Enough is Generated in Example

(B) Suppose $w \in Y$. We must show that $w \in \Pi_\mathsf{B}$. Because **diff** $w = 1$, there are two cases to consider.

- Suppose $w = 1x$, for some $x \in \{0, 1\}^*$. Because all proper prefixes of w have diffs ≤ 0, we have that $x = \%$, so that $w = 1$. Since $\mathsf{B} \rightarrow 1 \in P$, we have that $w = 1 \in \Pi_\mathsf{B}$.
- Suppose $w = 0x$, for some $x \in \{0, 1\}^*$. Thus **diff** $x = 2$. Because **diff** $x \geq 1$, by Lemma 4.5.2(1), we have that $x = yz$, for some $y, z \in \{0, 1\}^*$ such that $y \in Y$ and **diff** $z =$ **diff** $x - 1 = 2 - 1 = 1$. Hence $w = 0yz$. To finish the proof that $z \in Y$, suppose v is a proper prefix of z. Thus $0yv$ is a proper prefix of w. Since $w \in Y$, it follows that **diff** $v =$ **diff**$(0yv) \leq 0$, as required. Since $y, z \in Y$, part (B) of the inductive hypothesis tell us that $y, z \in \Pi_\mathsf{B}$. Thus, because $\mathsf{B} \rightarrow 0\mathsf{BB} \in P$ we have that $w = 0yz \in \Pi_\mathsf{B}$.

(C) Suppose $w \in Z$. We must show that $w \in \Pi_\mathsf{C}$. The proof is analogous to the proof of part (B).

A Problem with Unit Productions

Suppose H is the grammar

$$\mathsf{A} \to \mathsf{B} \mid \mathsf{0A3}, \qquad \mathsf{B} \to \% \mid \mathsf{1B2},$$

and let

$$X = \{\, 0^n 1^m 2^m 3^n \mid n, m \in \mathbb{N} \,\} \quad \text{and} \quad Y = \{\, 1^m 2^m \mid m \in \mathbb{N} \,\}.$$

We can prove that $X \subseteq \Pi_{H,\mathsf{A}} = L(H)$ and $Y \subseteq \Pi_{H,\mathsf{B}}$ using the above technique, but the production $\mathsf{A} \to \mathsf{B}$, which is called a *unit production* because its right side is a single variable, makes part (A) tricky. If $w = 0^0 1^m 2^m 3^0 = 1^m 2^m \in Y$, we would like to use part (B) of the inductive hypothesis to conclude $w \in \Pi_{\mathsf{B}}$, and then use the fact that $\mathsf{A} \to \mathsf{B} \in P$ to conclude that $w \in \Pi_{\mathsf{A}}$. But w is not a proper substring of itself, and so the inductive hypothesis in not applicable. Instead, we must split into cases $m = 0$ and $m \geq 1$, using $\mathsf{A} \to \mathsf{B}$ and $\mathsf{B} \to \%$, in the first case, and $\mathsf{A} \to \mathsf{B}$ and $\mathsf{B} \to \mathsf{1B2}$, as well as the inductive hypothesis on $1^{m-1} 2^{m-1} \in Y$, in the second case.

A Problem with Unit Productions

Because there are no productions from B back to A, we could also first use strong string induction to prove that, for all $w \in \{0,1\}^*$,

(B) if $w \in Y$, then $w \in \Pi_B$,

and then use the result of this induction along with strong string induction to prove that for all $w \in \{0,1\}^*$,

(A) if $w \in X$, then $w \in \Pi_A$.

This works whenever two parts of a grammar are not mutually recursive.

With this grammar, we could also first use mathematical induction to prove that, for all $m \in \mathbb{N}$, $1^m2^m \in \Pi_B$, and then use the result of this induction to prove, by mathematical induction on n, that for all $n, m \in \mathbb{N}$, $0^n1^m2^m3^n \in \Pi_A$.

A Problem with %*-Productions*

Note that %-productions, i.e., productions of the form $q \rightarrow$ %, can cause similar problems to those caused by unit productions. E.g., if we have the productions

$$\mathsf{A} \rightarrow \mathsf{BC} \quad \text{and} \quad \mathsf{B} \rightarrow \%,$$

then $\mathsf{A} \rightarrow \mathsf{BC}$ behaves like a unit production.

Proving that Everything Generated is Wanted

To prove that everything generated by a grammar is wanted, we introduce a new induction principle that we call induction on Π.

Principle of Induction on Π

Theorem 4.5.3 (Principle of Induction on Π)
Suppose G *is a grammar,* $P_q(w)$ *is a property of a string* $w \in \Pi_{G,q}$*, for all* $q \in Q_G$*, and* $P_a(w)$*, for* $a \in \textbf{alphabet}\, G$*, says "*$w = a$*".*

If

(1) *for all* $q \in Q_G$*, if* $q \rightarrow \% \in P_G$*, then* $P_q(\%)$*, and*

(2) *for all* $q \in Q_G$*,* $n \in \mathbb{N} - \{0\}$*,* $a_1, \ldots, a_n \in Q_G \cup \textbf{alphabet}\, G$*, and* $w_1 \in \Pi_{G,a_1}, \ldots, w_n \in \Pi_{G,a_n}$*, if* $q \rightarrow a_1 \cdots a_n \in P_G$ *and (†)* $P_{a_1}(w_1), \ldots, P_{a_n}(w_n)$*, then* $P_q(w_1 \cdots w_n)$*,*

then

$$\text{for all } q \in Q_G, \text{ for all } w \in \Pi_{G,q}, \ P_q(w).$$

We refer to (†) as the inductive hypothesis.

Principle of Induction on Π

Proof. It suffices to show that, for all $pt \in \textbf{PT}$, for all $q \in Q_G$ and $w \in (\textbf{alphabet}\, G)^*$, if pt is valid for G, $\textbf{rootLabel}\, pt = q$ and $\textbf{yield}\, pt = w$, then $P_q(w)$. We prove this using the principle of induction on parse trees. □

When proving part (2), we can make use of the fact that, for $a_i \in \textbf{alphabet}\, G$, $\Pi_{a_i} = \{a_i\}$, so that $w_i \in \Pi_{a_i}$ will be a_i. Hence it will be unnecessary to assume that $P_{a_i}(a_i)$, since this says "$a_i = a_i$", and so is always true.

Using Induction on Π *in Example*

Consider, again, our main example grammar G:

$$\begin{aligned}
\mathsf{A} &\to \% \mid \mathsf{0BA} \mid \mathsf{1CA}, \\
\mathsf{B} &\to \mathsf{1} \mid \mathsf{0BB}, \\
\mathsf{C} &\to \mathsf{0} \mid \mathsf{1CC}.
\end{aligned}$$

Let

$$\begin{aligned}
X &= \{\, w \in \{0,1\}^* \mid \textbf{diff}\, w = 0 \,\}, \\
Y &= \{\, w \in \{0,1\}^* \mid \textbf{diff}\, w = 1 \text{ and}, \\
&\qquad \text{for all proper prefixes } v \text{ of } w, \textbf{diff}\, v \le 0 \,\}, \text{ and} \\
Z &= \{\, w \in \{0,1\}^* \mid \textbf{diff}\, w = -1 \text{ and}, \\
&\qquad \text{for all proper prefixes } v \text{ of } w, \textbf{diff}\, v \ge 0 \,\}.
\end{aligned}$$

Using Induction on Π *in Example*

We have already proven that $X \subseteq \Pi_\mathsf{A} = L(G)$, $Y \subseteq \Pi_\mathsf{B}$ and $Z \subseteq \Pi_\mathsf{C}$. To complete the proof that $L(G) = \Pi_\mathsf{A} = X$, $\Pi_\mathsf{B} = Y$ and $\Pi_\mathsf{C} = Z$, we will use induction on Π to prove that $\Pi_\mathsf{A} \subseteq X$, $\Pi_\mathsf{B} \subseteq Y$ and $\Pi_\mathsf{C} \subseteq Z$.

We use induction on Π to show that:

- *(A)* for all $w \in \Pi_\mathsf{A}$, $w \in X$;
- *(B)* for all $w \in \Pi_\mathsf{B}$, $w \in Y$; and
- *(C)* for all $w \in \Pi_\mathsf{C}$, $w \in Z$.

Formally, this means that we let the properties $P_\mathsf{A}(w)$, $P_\mathsf{B}(w)$ and $P_\mathsf{C}(w)$ be "$w \in X$", "$w \in Y$" and "$w \in Z$", respectively, and then use the induction principle to prove that, for all $q \in Q_G$, for all $w \in \Pi_q$, $P_q(w)$. But we will actually work more informally.

Using Induction on Π in Example

There are seven productions to consider.

- **(**$A \to \%$**)** We must show that $\% \in X$ (as "$w \in X$" is the property of part (A)). And this holds since $\textbf{diff}\ \% = 0$.
- **(**$A \to 0BA$**)** Suppose $w_1 \in \Pi_B$ and $w_2 \in \Pi_A$ (as $0BA$ is the right-side of the production, and 0 is in G's alphabet), and assume the inductive hypothesis, $w_1 \in Y$ (as this is the property of part (B)) and $w_2 \in X$ (as this is the property of part (A)). We must show that $0w_1w_2 \in X$, as the production shows that $0w_1w_2 \in \Pi_A$. Because $w_1 \in Y$ and $w_2 \in X$, we have that $\textbf{diff}\ w_1 = 1$ and $\textbf{diff}\ w_2 = 0$. Thus $\textbf{diff}(0w_1w_2) = -1 + 1 + 0 = 0$, showing that $0w_1w_2 \in X$.

Using Induction on Π *in Example*

- ($B \to 0BB$) Suppose $w_1, w_2 \in \Pi_B$, and assume the inductive hypothesis, $w_1, w_2 \in Y$. Thus w_1 and w_2 are nonempty. We must show that $0w_1w_2 \in Y$. Clearly, $\textbf{diff}(0w_1w_2) = -1 + 1 + 1 = 1$. So, suppose v is a proper prefix of $0w_1w_2$. We must show that $\textbf{diff}\ v \leq 0$. There are three cases to consider.
    - Suppose $v = \%$. Then $\textbf{diff}\ v = 0 \leq 0$.
    - Suppose $v = 0u$, for a proper prefix u of w_1. Because $w_1 \in Y$, we have that $\textbf{diff}\ u \leq 0$. Thus $\textbf{diff}\ v = -1 + \textbf{diff}\ u \leq -1 + 0 \leq 0$.
    - Suppose $v = 0w_1u$, for a proper prefix u of w_2. Because $w_2 \in Y$, we have that $\textbf{diff}\ u \leq 0$. Thus $\textbf{diff}\ v = -1 + 1 + \textbf{diff}\ u = \textbf{diff}\ u \leq 0$.
- The remaining productions are handled similarly.