CS/ECE 374 Spring 2017
Homework 3 Problem 3

Renheng Ruan (rruan2)
Lanxiao Bai (lbai5)

Let $L = \{0^i 1^j 2^k \mid k = 2(i+j)\}$.

(a) Prove that L is context free by describing a grammar for L.

(b) Prove that your grammar is correct. You need to prove that if $L \subseteq L(G)$ and $L(G) \subseteq L$ where G is your grammar from the previous part.

Solution:

(a) $S \to \varepsilon \mid SSS \mid \mathtt{0}S\mathtt{22} \mid \mathtt{1}S\mathtt{22} \mid \mathtt{0}S\mathtt{1}S\mathtt{2222}$

(b) We separately prove $L \subseteq L(G)$ and $L(G) \subseteq L$ as follows:

Claim 1. *$L(G) \subseteq L$, that is, every string in $L(G)$ has exactly twice as many* $\mathtt{2}$s *as the sum of* $\mathtt{0}$s *and* $\mathtt{1}$s.

Proof: For any string u, let $\Delta(u) = \#(\mathtt{2}, u) - 2(\#(\mathtt{0}, u) + \#(\mathtt{1}, u))$. We need to prove that $\Delta(w) = 0$ for every string $w \in L(G)$.

Let w be an arbitrary string in $L(G)$. Assume that $\Delta(x) = 0$ for every string $x \in L(G)$s. Consider the *shortest* derivation of w, and assume $\Delta(x) = 0$ for every string $x \in L(G)$ such that $|x| < |w|$. There are five cases to consider, depending on the first production in the derivation of w.

- If $w = \varepsilon$, then $\#(\mathtt{2}, w) = 2(\#(\mathtt{0}, w) + \#(\mathtt{1}, w)) = 0$ by definition, so $\Delta(w) = 0$.
- Suppose the derivation begins $S \rightsquigarrow SSS \rightsquigarrow^* w$. Then $w = xyz$ for some strings $x, y, z \in L(G)$, each of which can be derived with fewer than k productions. The inductive hypothesis implies $\Delta(x) = \Delta(y) = 0$. It immediately follows that $\Delta(w) = 0$.[1]
- Suppose the derivation begins $S \rightsquigarrow \mathtt{0}S\mathtt{22} \rightsquigarrow^* w$. Then $w = \mathtt{0}x\mathtt{22}$ for some string $x \in L(G)$. The inductive hypothesis implies $\Delta(x) = 0$. It immediately follows that $\Delta(w) = 0$.
- Suppose the derivation begins $S \rightsquigarrow \mathtt{1}S\mathtt{00} \rightsquigarrow^* w$. Then $w = \mathtt{1}x\mathtt{22}$ for some string $x \in L(G)$. The inductive hypothesis implies $\Delta(x) = 0$. It immediately follows that $\Delta(w) = 0$.
- Suppose the derivation begins $S \rightsquigarrow \mathtt{0}S\mathtt{1}S\mathtt{2222} \rightsquigarrow^* w$. Then $w = \mathtt{0}x\mathtt{1}y\mathtt{2222}$ for some strings $x, y \in L(G)$. The inductive hypothesis implies $\Delta(x) = \Delta(y) = 0$. It immediately follows that $\Delta(w) = 0$.

In all cases, we conclude that $\Delta(w) = 0$, as required. □

[1] Alternatively: Suppose the *shortest* derivation of w begins $S \rightsquigarrow SSS \rightsquigarrow^* w$. Then $w = xy$ for some strings $x, y \in L(G)$. Neither x or y can be empty, because otherwise we could shorten the derivation of w. Thus, x and y are both shorter than w, so the induction hypothesis implies.... We need some way to deal with the decompositions $w = \varepsilon \bullet w$ and $w = w \bullet \varepsilon$, which are both consistent with the production $S \to SSS$, without falling into an infinite loop.

Claim 2. *$L \subseteq L(G)$; that is, G generates every binary string with exactly twice as many* `2`*s as the sum of* `0`*s and* `1`*s.*

Proof: For any string u, let $\Delta(u) = 2\#(\texttt{1}, u) - (\#(\texttt{2}, u) - 2\#(\texttt{0}, u))$. For any string u and any integer $0 \le i \le |u|$, let $\boldsymbol{u_i}$ denote the ith symbol in u, and let $\boldsymbol{u_{\le i}}$ denote the prefix of u of length i.

Let w be an arbitrary binary string with twice as many `2`s as the sum of `0`s and `1`s. Assume that G generates every binary string x that is shorter than w and has twice as many `2`s as the sum of `0`s and `1`s. There are two cases to consider:

- If $w = \varepsilon$, then $\varepsilon \in L(G)$ because of the production $S \to \varepsilon$.
- Suppose w is non-empty. To simplify notation, let $\Delta_i = \Delta(w_{\le i})$ for every index i, and observe that $\Delta_0 = \Delta_{|w|} = 0$. There are several subcases to consider:
  - Suppose $\Delta_i = 0$ for some index $0 < i < |w|$. Then we can write $w = xyz$, where x, y and z are non-empty strings with $\Delta(x) = \Delta(y) = \Delta(z) = 0$. The induction hypothesis implies that $x, y, z \in L(G)$, and thus the production rule $S \to SSS$ implies that $w \in L(G)$.
  - Suppose $\Delta_i > 0$ for all $0 < i < |w|$. Then w must begin with `0`, since otherwise $\Delta_1 = -2$ or $\Delta_2 = -1$, and the last two symbol in w must be `22`, since otherwise $\Delta_{|w|-1} = -1$. Thus, we can write $w = \texttt{0}x\texttt{22}$ for some binary string x. We easily observe that $\Delta(x) = 0$, so the induction hypothesis implies $x \in L(G)$, and thus the production rule $S \to \texttt{0}S\texttt{22}$ implies $w \in L(G)$.
  - Suppose $\Delta_i < 0$ for all $0 < i < |w|$. A symmetric argument to the previous case implies $w = \texttt{1}x\texttt{22}$ for some binary string x with $\Delta(x) = 0$. The induction hypothesis implies $x \in L(G)$, and thus the production rule $S \to \texttt{1}S\texttt{22}$ implies $w \in L(G)$.
  - Finally, suppose none of the previous cases applies: $\Delta_i < 0$ and $\Delta_j > 0$ for some indices i and j, but $\Delta_i \ne 0$ for all $0 < i < |w|$.

    Let i be the smallest index such that $\Delta_i < 0$. Because Δ_j either increases by 1 or decreases by 2 when we increment j, for all indices $0 < j < |w|$, we must have $\Delta_j > 0$ if $j < i$ and $\Delta_j < 0$ if $j \ge i$.

    In other words, there is a *unique* index i such that $\Delta_{i-1} > 0$ and $\Delta_i < 0$. In particular, we have $\Delta_1 > 0$ and $\Delta_{|w|-1} < 0$. Thus, we can write $w = \texttt{0}x\texttt{1}y\texttt{2222}$ for some binary strings x and y, where $|\texttt{1}y\texttt{22}| = i$.

    We easily observe that $\Delta(x) = \Delta(y) = 0$, so the inductive hypothesis implies $x, y \in L(G)$, and thus the production rule $S \to \texttt{0}S\texttt{1}S\texttt{2222}$ implies $w \in L(G)$.

In all cases, we conclude that G generates w. □

Idea comes from the question 4 and use its form.