Autumn 1997–98, Prof. Boneh
Solution Set
CS 154, Problem Set 3

3.1 (1 point)

PROBLEM: **Construct a CFG accepting the language $L = \{w : w$ starts and ends with the same symbol$\}$. The alphabet is $\Sigma = \{0, 1\}$.**

Note that this language is actually regular: it's just $1(0 \cup 1)^*1 \cup 0(0 \cup 1)^*0 \cup 0 \cup 1$. According to this definition, ϵ is not in the language, thought the question as stated is ambiguous on this point. The CFG is very easy to eyeball:

$$\begin{aligned} S &\rightarrow 0A0 \mid 1A1 \mid 0 \mid 1 \\ A &\rightarrow 0A \mid 1A \mid \epsilon \end{aligned}$$

3.2 (3 points)

PROBLEM: **Prove that the following languages are not CFLs:**
(1) $\{a\#b : a$ is a substring of b where $w, x \in \{0,1\}^*\}$; (2) $\{0^i 1^j 0^i 1^j : i, j \geq 0\}$.

(1) (Note that I changed variable names to avoid confusion with the pumping string w and the pumping part x). I'm going to use basically the same string for both problems. As in all pumping proofs, the adversary gives us m, and I pick $w = 0^m 1^m \# 0^m 1^m$. The adversary can split up w into five parts, $uvxyz$, in basically four ways:

1. Either v or y include $\#$. Pumping a zillion times causes a zillion $\#$s, which is clearly not in the language. This is the uninteresting technical case.
2. vxy is entirely in the first half of the string (ie the first $0^m 1^m$). Then pumping twice times (to get uv^2xy^2z) causes $|a| > |b|$, so clearly a is not a substring of b. This works because vxy can't be empty (either v or y is non-zero) and it's entirely in the a half of the string.
3. vxy is entirely in the second half (the b part) of the string. Then pumping zero times causes $|b| < |a|$, so again a can't be a substring of b.
4. v is in the first half, and y is the second half (either can be empty). (Note each must be on one side of the halfway mark, or we're in case 1.) Once again we want to get $|a| > |b|$, so we need to deal with cases $|v| < |y|$, $|v| > |y|$, $|v| = |y|$. In the first case, we pump 0 times; since b shrinks more than a, $|b| < |a|$ after pumping. In the second case, we pump 2 times, making $|a| > |b|$.

   $|v| = |y|$ is the hard case, and we finally make use of the form of w. Since $|vxy| < m$ and crosses the middle of w, the start of vxy is at most m away from the middle, as is the end. Thus, vxy must be entirely in the middle $1^m 0^m$. After we pump, then we get $0^m 1^{m+i} 0^{m+i} 1^m$. (Remember that neither v or y cross the middle because that was dealt with in the first case.) Now i can't be 0 (why?), so a is not a substring of b because it has too many 1s.

(2) We use a similar w, $0^m 1^m 0^m 1^m$, and the same argument as before works. This is because if a is not a substring of b, it certainly isn't equal to b, as required by this language! The only difference is the first case. Now, if v or y cross the midpoint, the crosser will be of the form $0^i 1^j$, and if we pump a zillion times we get a lot of alternating 0s and 1s, which is not in the language.

3.3 (3 points)

PROBLEM: **Prove that CFLs are not closed under complement. That is, prove there exists a CFL L such that $L^c = \Sigma^* \setminus L$ is not regular.**

I use the terminology $\overline{L} = L^c$. By deMorgan's law,

$$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}.$$

Since we know CFLs are closed under union, if they were closed under complement then they would be closed under intersection too: to intersect two CFLs, we'd take their complements, union them, and take the complement again. But CFLs are *not* closed under intersection: $a^n b^n c^* \cap a^* b^n c^n = a^n b^n c^n$. Thus our initial assumption, that CFLs are closed under complement, is wrong.

Just for fun, prove that $L = \{a^i b^j c^k : i \neq j \text{ or } j \neq k\}$ is a CFL but L^c is not. Hint: consider the grammar

$$S \to AK \mid IB \,\Big|\, A \to aAb \mid aI \mid bJ \,\Big|\, B \to bBc \mid bJ \mid cK \,\Big|\, I \to aI \mid \epsilon \,\Big|\, J \to bJ \mid \epsilon n \,\Big|\, K \to cK \mid \epsilon$$

Then prove $L^c \cap a^* b^* c^*$ is not a CFL and appeal to the (contrapositive of the) next problem.

3.4 (3 points)

PROBLEM: **Let L be a CFL and R be a regular language. Prove that $L \cap R$ is a CFL.**

The following proof is from Hopcroft and Ullman, pp. 135–136. Let M_l, a PDA accepting L, be $\{Q_l, \Sigma, \Gamma, \delta_l, q_0, F_l\}$, and let a DFA M_r accepting R be $\{Q_r, \Sigma, \delta_r, q_0', F_r\}$. (It is a problem to assume the Σs are the same?)

The construction of the new PDA is straightforward: we consider the cross product of the states. Every time we read a letter, we move on both M_l and M_r at once. Whenever we read from the stack, the R half of the state stays still, while the L half moves according to the PDA transition function. Formally:

$$M' = (Q_l \times Q_r, \Sigma, \Gamma, \delta', \{q_0, q_0'\}, F_l \times F_r)$$

where $\delta'(\{q_l, q_r\}, a, s) = \{(\{q_l', q_r'\}, s') : q_r' = \delta_r(q_r, a) \text{ and } (q_l', s') \in \delta_l(q_l, a, s)\}$. Since DFAs are not defined for $a = \epsilon$ but we need to handle that case, we define $\delta_r(q, \epsilon) = q$ — that is, the DFA part does not move if the input is ϵ.

To prove the construction works, we must show δ' behaves like we think: M' can have Instantaneous Description $(\{q_l, q_r\}, w_{1\ldots i}, w_{i+1\ldots n}, S)$ iff M_l can have ID $(q_l, w_{1\ldots i}, w_{i+1\ldots n}, S)$ and M_r is in state q_r after reading $w_{1\ldots i}$. This is enough since it means M' lands in F' on w iff M_l lands in F_l and M_r lands in F_r.

We induct. The base case — $q_l = q_0$, $q_r = q_0'$ — is trivial. If M' has ID $(\{q_l, q_r\}, w_{1\ldots i-1}, w_{i\ldots n}, sS)$, then for the next transition we have two possibilities: we read w_i or we read ϵ. In the first case, δ' forces us into an ID of the form $(\{q_l', \delta_r(q_r, w_i)\}, w_{1\ldots i}, w_{i+1\ldots n}, s'S)$, where $(q_l', s') \in \delta_l(q_l, w_i, s)$. Since, upon reading w_i, M_r is in state $\delta_r(q_r, w_i)$ (by induction we know M_r was in state q_r after reading $w_{1\ldots i-1}$) and M_l has ID $(q_l', w_{1\ldots i}, w_{i+1\ldots n}, s'S)$, the induction holds.

If we read ϵ instead of w_i, then δ' causes M' to have an ID of the form $(\{q_l', q_r\}, w_{1\ldots i-1}, w_{i\ldots n}, s'S)$, where $(q_l', s') \in \delta_l(q_l, w_i, s)$. We are consistent with M_l for the same reason as before, and we are consistent with M_r by induction: we were consistent with M_r before this last transition, but since we read none of w and did not change q_r, we must still be consistent with M_r. Consistency follows because consistency with M_r is determined solely by w and q_r.

We must also induct in the other direction ("However M_r and M_l move, M' moves the same way") but the proof is very similar and I omit it.