FOUNDATIONS OF COMPUTATION 2019
EXAMPLES OF NON-CONTEXT-FREE LANGUAGES. PROOF BY PUMPING LEMMA

1. $L := \{a^k ba^k ba^k b \in \{a,b\}^* | \, k \geq 0\}$

Suppose, contrary to what we want to prove, that the language L is context-free. Applying the Pumping Lemma (for context-free languages) to L we get the number n. Consider the word $w = a^n ba^n ba^n b$ having the length $3n+3 > n$. Then, by the lemma, w can be represented as $w = uvxyz$. We have $|vxy| \leq n$. Consider two cases.

(1) vy consists only of letters a (it may happen that x contains the letter b, but no more than one). Then the word uv^0xy^0z contains more letters a between some pair of letters b than between some other pair, which contradicts to uv^0xy^0z being in L. Thus, our assumption that L is context-free led to contradiction with the Pumping Lemma. Hence L is not context-free.

(2) vy contains a letter b (it can't contain more than one b). Then the word uv^0xy^0z contains just two letters b, which contradicts to it being in L. Again we got a contradiction.

2. $L := \{a^p \in \{a\}^* | \, p \text{ is a prime number}\}$

Suppose that, contrary to what we want to prove, the language is L is context-free. Applying the Pumping Lemma (for context-free languages) to L we get the number n. Consider $a^p \in L$ for a prime number $p \geq n$. Then a^p is represented in a form $a^p = uvxyz$. Denote $|uxz| = s$ and $|vy| = r$. Then $p = s + r$ and $|uv^ixy^iz| = s + ir$. By Pumping Lemma, the word a^{s+ir} belongs to L, in other words, $s + ir$ is a prime number, for every $i \geq 0$.

Take $i = s + 2r + 2$. Then $s + ir = s + (s + 2r + 2)r = (r + 1)(s + 2r)$. Hence, for this choice of i, the number $s + ir$ is a product of two numbers greater than 1, which means that $s + ir$ is not prime. We get a contradiction, showing that L is not context-free.

3. $L := \{ww | \, w \in \{a,b\}^*\}$

Suppose that, contrary to what we want to prove, the language is L is context-free. Applying the Pumping Lemma (for context-free languages) to L we get the number n. Consider the word $\gamma = a^{n+1}b^{n+1}a^{n+1}b^{n+1}$ having the length $4n+4 > n$. Then, by the lemma, γ can be represented as $\gamma = uvxyz$, and the word uv^0xy^0z belongs to L, i.e., $uv^0xy^0z = w_1w_1$ for some $w_1 \in \{a,b\}^*$. It is clear that the first half of uv^0xy^0z begins with an a, while the second half of uv^0xy^0z ends with a b (since $|vxy| \leq n$). Therefore w_1 begins with an a and ends with a b, in other words, is of the kind $w_1 = a^\ell b^k$. Consider three cases.

(1) If vxy lies in the first subword $a^{n+1}b^{n+1}$ of γ then w_1 will end with an a (because the center of the word uv^0xy^0z will move to the right), which is a contradiction.

(2) If vxy lies in the second subword $a^{n+1}b^{n+1}$ of γ then w_1 will begin with a b (because the center of the word uv^0xy^0z will move to the left), which is a contradiction.

(3) If vxy lies in the subword $b^{n+1}a^{n+1}$, and neither entirely in the first nor in the second subword $a^{n+1}b^{n+1}$, then either the number of letters b in the first half of uv^0xy^0z will be less than the number of letters b in the second half of uv^0xy^0z, or the number of letters a in the second half of uv^0xy^0z will be less than the number of letters a in the first half of uv^0xy^0z, or both. We got a contradiction.