

Solution to Homework 4

Task: Prove that the following language is not regular

$$\{b^n a^{n+2}, n = 0, 1, 2, \dots\} = \{aa, baaa, bbaaaa, bbbaaaaa, \dots\}.$$

Solution. We will prove this result by contradiction. Let us assume that the language L is regular, and let us show that this assumption leads to a contradiction.

Since this language is regular, according to the Pumping Lemma, there exists an integer p such that every word from L whose length $\text{len}(w)$ is at least p can be represented as a concatenation $w = xyz$, where:

- y is non-empty;
- the length $\text{len}(xy)$ does not exceed p , and
- for every natural number i , the word $xy^i z \stackrel{\text{def}}{=} xy \dots yz$, in which y is repeated i times, also belongs to the language L .

Let us take the word

$$w = b^p a^{p+2} = b \dots ba \dots a,$$

in which first the letter b is repeated p times and then the letter a is repeated $p + 2$ times. The length of this word is $p + p + 2 = 2p + 2 > p$. So, by pumping lemma, this word can be represented as $w = xyz$ with $\text{len}(xy) \leq p$. The word $w = xyz$ starts with xy , and the length of xy is smaller than or equal to p . Thus, xy is among the first p symbols of the word w – and these symbols are all b 's. So, the word y only has b 's.

In the original word $w = xyz$, we had b repeated p times and a repeated $p + 2$ times. When we go from the word $w = xyz$ to the word $xyyz$, we add b 's, and we do not add any a 's. Thus, we still have a repeated $p + 2$ times, but the number of times b is repeated is now larger than p . In any word from the language L , we should have exactly two more a 's than b 's. So, in the word $xyyz$, this balance is disrupted. Thus, the word $xyyz$ cannot be in the language L .

On the other hand, by Pumping Lemma, the word $xyyz$ must be in the language L . So, we proved two opposite statements:

- that this word *is not* in L and

- that this word *is* in L .

This is a contradiction.

The only assumption that led to this contradiction is that L is a regular language. Thus, this assumption is false, so L is not regular.