Repetitions in Words

Thesis submitted to the
Indian Institute of Technology Guwahati
for the award of the degree

of

Doctor of Philosophy

Submitted by
Maithilee Patawar

Under the guidance of
Dr. Kalpesh Kapoor
Dr. Benny George K

Department of Computer Science and Engineering
Indian Institute of Technology Guwahati
April 2024

Abstract

Repetitions are fundamental properties of words, and different types of repetitions have been explored in the area of word combinatorics. This thesis investigates two types of repetitions: squares and antisquares. We investigate the square conjecture that anticipates the number of distinct squares in a word is less than its length. It is known that any location of a word can be mapped to at most two rightmost squares, and a pair of these squares was referred to as an FS-double square. For simplicity, we will refer to the longer square in this pair as an FS-double square throughout this thesis. We examine the properties of words containing FS-double squares and explore the consecutive locations starting with FS-double squares. We observe that FS-double squares introduce no-gain locations where no rightmost squares occur. The count of these no-gain locations in words with a sequence of FS-double squares demonstrates that the square density of such words is less than $\frac{11}{6}$. Furthermore, we investigate words that possess FS-double squares and maintain an equivalent number of such squares when reversed. We prove that the maximum number of FS-double squares in such a word is less than $\frac{1}{11}^{\text{th}}$ of the length of the word. Another aspect of our research involves counting squares in a repetition. A non-primitive word has a form u^k for some non-empty word u and some positive integer k such that $k > 2$. With no-gain locations and FS-double squares in these words, we conclude that the square density of such words approaches $\frac{1}{2}$ as k increases. Also, we work on the lower bound of the square conjecture. The previous lower bound is obtained using a structure that generates words containing a high number of distinct squares. We identify similar structures but produce words with more distinct squares. We also study antisquare, a special repetition of the form $u\bar{u}$ where u is a binary word, and $\bar{u}$ is its complement. We show that a word w can contain at most $\frac{|w|(|w|+2)}{8}$ antisquares, and the lower bound for the number of distinct antisquares in w is $|w| - 1$.