PART I

Problems

Damian Niwiński and Wojciech Rytter

Cambridge University Press & Assessment
978-1-316-51346-0 — 200 Problems on Languages, Automata, and Computation
Edited by Filip Murlak , Damian Niwiński , Wojciech Rytter
Excerpt
More Information

1

Words, Numbers, Graphs

Let us fix a finite set Σ; we shall refer to it as the *alphabet*. The elements of Σ are called *letters* or *symbols*. A *word* w over Σ is a finite sequence $a_1a_2\dots a_n$ of letters from Σ. The length of $w = a_1a_2\dots a_n$, denoted by $|w|$, is n. The *empty word*, denoted by ε, is the empty sequence; it has length 0. We write Σ^* for the set of all words over Σ, and Σ^+ for the set of non-empty words over Σ. The *concatenation* of words $u = a_1a_2\dots a_m$ and $v = b_1b_2\dots b_n$, denoted by $u \cdot v$ or simply uv, is the word $a_1a_2\dots a_mb_1b_2\dots b_n$. We write v^n for the word $\underbrace{vv\dots v}_{n}$.

For a word $w \in \Sigma^*$ and $1 \le i,j \le |w|$, we write $w[i]$ for the ith letter of w and $w[i..j]$ for the infix starting at the ith letter and ending at the jth letter of w; that is, if $w = a_1a_2\dots a_n$, then $w[i..j] = a_ia_{i+1}\dots a_j$. In particular $w[i..i] = w[i]$ and $w[1..|w|] = w$. For $j < i$ we let $w[i..j] = \varepsilon$.

A *language* (over Σ) is a set of words (over Σ). The *concatenation* of languages L and K is the language $LK = \{uv : u \in L \wedge v \in K\}$. We write L^n for the language $\{w_1w_2\dots w_n : w_1, w_2, \dots, w_n \in L\}$. The *Kleene star* of a language L is the language $L^* = \bigcup_{n=0}^{\infty} L^n$. Note the case of $n = 0$ in the definition of the Kleene star, which means that the empty word ε belongs to the Kleene star of every language, even the empty one. We write L^+ for $\bigcup_{n=1}^{\infty} L^n$.

Problem 1 ☆
Prove that all languages L and M satisfy

$$(L^*M^*)^* = (L \cup M)^*.$$

Problem 2. PARENTHESIS EXPRESSIONS ☆
Show that the following two ways of defining the set of balanced sequences of parentheses are equivalent:

- The least set L such that the empty sequence ε is in L and if $w, v \in L$ then $(w), wv \in L$.
- The set K of words over the alphabet $\{(\,,\,)\}$ in which the number of occurrences of) is the same as the number of occurrences of (, and in each prefix the number of occurrences of) is not greater than the number of occurrences of (.

Problem 3. GAME GRAPH
Consider the following game between a barman and a customer. Between the players there is a revolving tray with four glasses forming the vertices of a square. Each glass is either right-side up or upside down, but the barman is blindfolded and wears gloves, so he has no way of telling which of the two cases holds. In each round, the barman chooses one or two glasses and reverses them. Afterwards, the customer rotates the tray by a multiple of 90 degrees. The barman wins if at any moment all glasses are in the same position (he is to be informed about this immediately). Can the barman win this game, starting from an unknown initial position? If so, how many moves are sufficient? Would you play this game for money against the barman? What about an analogous game with three or five glasses?

Problem 4. SEMI-LINEAR SETS
For any fixed $a, b \in \mathbb{N}$, the set of natural numbers $\{a + bn : n \in \mathbb{N}\}$ is called *linear*. A *semi-linear* set is a finite union of linear sets. (The empty set is obtained as the union of the empty family of linear sets.)

(1) Prove that a set A of natural numbers is semi-linear if and only if it is *ultimately periodic*; that is, there exist $c \in \mathbb{N}$ and $d \in \mathbb{N} - \{0\}$ such that for all $x > c$,

$$x \in A \text{ if and only if } x + d \in A.$$

(2) Let $A \subseteq \mathbb{N}$ be arbitrary. Show that the following set is semi-linear:

$$A^* = \{a_1 + a_2 + \cdots + a_k \ : \ k \in \mathbb{N}, a_1, \ldots, a_k \in A\}\,.$$

HINT: *Use congruence* mod m *for a suitably chosen* m.

(3) Prove that the set $A = \{a + b_1 n_1 + \cdots + b_k n_k \ : \ n_1, \ldots, n_k \in \mathbb{N}\}$ is semi-linear for all fixed k and $a, b_1, \ldots, b_k \in \mathbb{N}$.

(4) Prove that the family of all semi-linear sets is closed under finite unions, finite intersections, and complement with respect to $\mathbb{N}$.

(5) Prove that (a) any semi-linear set can be represented as the set of lengths of directed paths between two fixed sets of vertices of a finite graph, and (b) any set of numbers obtained in this way is semi-linear.
(6) Give an example of a subset of $\mathbb{N}$ that is not semi-linear.

Problem 5. PRIMITIVE WORDS ★
A word $w \in \Sigma^*$ is *primitive* if it cannot be presented as $w = v^n$ for any $n > 1$.

(1) Prove that for each non-empty word w there is *exactly one* primitive word v such that $w = v^n$ for some $n \geq 1$. We call n the *exponent* of the word w.
(2) For any words w and v, we say that wv is a *cyclic shift* of vw and that wv and vw are *conjugate*. Prove that being conjugate is an equivalence relation and all conjugate words have the same exponent. What is the cardinality of the equivalence class of a word of length m and exponent n?

Problem 6. CODES ★
A language $C \subseteq \Sigma^+$ is a *code* if each word $w \in C^+$ can be *decoded*; that is, w admits exactly one factorization with respect to C: there is exactly one way to present w as $v_1 v_2 \ldots v_n$ with $v_1, v_2, \ldots, v_n \in C$ and $n \in \mathbb{N}$.

(1) Let $\Sigma = \{a, b\}$. Prove that the set $\{aa, baa, ba\}$ is a code and the set $\{a, ab, ba\}$ is not a code.
(2) For a finite set C that is not a code, give an upper bound on the length of the shortest word that admits two different factorizations. Can you find an example meeting your bound?
(3) Show that $\{u, v\}$ is a code if and only if $uv \neq vu$.

See also Problem 75.

Problem 7. THUE–MORSE WORD ★★
(1) Show that the following definitions of the Thue–Morse word are equivalent:
  - the infinite sequence of 0's and 1's obtained by starting with 0 and successively appending the sequence obtained so far with all bits flipped;
  - the infinite word $s_0 s_1 s_2 \ldots$ such that $s_n = 0$ if the number of 1's in the binary representation of n is even, and $s_n = 1$ if it is odd;

- the infinite word $t_0t_1t_2\ldots$, whose letters satisfy the recurrence relation: $t_0 = 0$, $t_{2n} = t_n$, and $t_{2n+1} = 1 - t_n$ for all n.

(2) Show that the Thue–Morse word is *cube-free*; that is, it contains no infix of the form www with $w \neq \varepsilon$. In fact, it is *strongly cube-free*; that is, it contains no infix of the form $bwbwb$ for $b \in \{0, 1\}$.

HINT: *First show that it contains neither* 000 *nor* 111 *as an infix, but each infix of length five contains* 00 *or* 11 *as an infix.*

(3) Construct an infinite word over a four-letter alphabet that is *square-free*; that is, it contains no infix of the form ww with $w \neq \varepsilon$.

(4) Can it be done with three letters? And two letters?