EXERCISES COMBINATORICS ON WORDS
(UNDER CONSTRUCTION)

1. Basics

(1) Prove that a real number is rational if and only if its base-b expansion is eventually periodic.

(2) Let p/q where $\gcd(p,q) = 1$ and $\gcd(q, 10) = 1$. Prove that the length of the period of the decimal expansion of $[p/q]_{10}$ is the multiplicative order of 10 modulo q, i.e. $10^e \equiv 1 \pmod{q}$.

(3) Suppose $\mathbf{w}$ is an eventually periodic infinite word. Show that the frequency of each letter in $\mathbf{w}$ exists and is rational. Does the converse hold?

(4) The *run lengths* of an infinite word $\mathbf{x} = (x_n)_{n\in\mathbb{N}}$ is the (finite or infintie) word $(r_n)_n \in \mathbb{N}_0^{\mathbb{N}} \cup \mathbb{N}_0^*\{\omega\}$ defined by

$$\mathbf{x} = a_0^{r_0} a_1^{r_1} a_2^{r_2} \cdots$$

where a_i's are letters and $a_i \neq a_{i+1}$ for all i.
a) if $\mathbf{w}$ is eventually periodic, then the sequence of run lengths of elements of $\mathbf{w}$ is finite or eventually periodic;
b) the converse is true if $\mathbf{w}$ is over an alphabet with at most 2 letters, but is false for alphabets of size at least 3.

(5) Prove that an infinite ultimately periodic and recurrent word is (purely) periodic.

(6) What is the factor complexity of $1234567891011\cdots$, the concatenation of the decimal expansions of the positive integers?

(7) Show that $235711131719232931\cdots$, the concatenation of the decimal expansions of the prime numbers, has factor complexity 10^n. Suggestion: make use of the following result. Let a, b be integers with $1 \leq a < b$ and $\gcd(a, b) = 1$. Then there exists a prime $p \equiv a \pmod{b}$ with $p = O(b^{11/2})$.

(8) Let $\mathbf{w}$ be an infinite word over an alphabet A. Prove that

$$p_{\mathbf{w}}(n) \leq p_{\mathbf{w}}(n+1) \leq (\#A)\, p_{\mathbf{w}}(n)$$

for all $n \geq 0$.

(9) Let $\mathbf{w}$ be an infinite word over an alphabet A. Prove that

$$p_{\mathbf{w}}(n+1) - p_{\mathbf{w}}(n) \leq (\#A)\,(p_{\mathbf{w}}(n) - p_{\mathbf{w}}(n-1))$$

for all $n \geq 1$.

(10) Let $\mathbf{w}$ be an infinite word over an alphabet A and $h : A^* \to B^*$ be a non-erasing morphism. Prove that $p_{h(\mathbf{w})}(n) \leq D\, p_{\mathbf{w}}(n)$ where $D = \max_{a\in A} |h(a)|$.

(11) What is the factor complexity of the word

$$1101000100000001\cdots,$$

the characteristic sequence of the powers of 2?

2. Finite words

(1) Prove the following result. Let w and x be non-empty words. Let $\mathbf{y} \in w\{w,x\}^\omega$ and $\mathbf{z} \in x\{w,x\}^\omega$ be two infinite words. Then the following conditions are equivalent:
   (a) $\mathbf{y}$ and $\mathbf{z}$ agree on a prefix of length at least $|w| + |x| - gcd(|w|, |x|)$;
   (b) $wx = xw$;
   (c) $\mathbf{y} = \mathbf{z}$.

(2) Prove that two finite words x and y are powers of the same word if and only if there exists integers $i, j \geq 0$ such that $x^i = y^j$.

(3) Show that a word $w \in A^*$ is primitive if and only if its period is not a proper divisor of its length.

(4) Show that if a primitive word $w \in A^*$ is the product of two non-empty palindromes, then this factorization is unique.

(5) Let Γ and Δ be the application from $\{0,1\}^* \times \{0,1\}^*$ into itself by

$$\Gamma(u,v) = (u, uv) \quad \text{and} \quad \Delta(u,v) = (vu, v)$$

A pair of words (u,v) is said to be *standard* if it is obtained from the pair $(0,1)$ by applying a finite composition of Γ and Δ. For instance, the pair $(10, 10101)$ is standard as $(10, 10101) = \Gamma^2(\Delta(0,1))$. A word $w \in \{0,1\}^*$ is *standard* if it is a component of a standard pair.
   (a) Show that every standard word is primitive.
   (b) Show that if (u,v) is a standard pair with $|u|, |v| \geq 2$, there exist palindromes p, q, r such that

$$x = p10 = qr \quad \text{and} \quad y = q01$$

   or

$$x = q10 = qr \quad \text{and} \quad y = p01 = qr.$$

   (c) A word w is *central* if $w01$ (or equivalently $w10$) is a standard word. Show that if w is a central word, then it has two periods k and ℓ such that $|w| = k + \ell - 2$ and $\gcd(k, \ell) = 1$.

3. Topology on $A^{\mathbb{N}}$

(1) Show that the distances d on $A^{\mathbb{N}}$ defined by

$$d(\mathbf{x}, \mathbf{y}) = 2^{-\inf\{n \in \mathbb{N} | x_n \neq y_n\}}$$

generates the product topology on $A^{\mathbb{N}}$ (where each copy of A is endowed with the discrete topology).

(2) Show that for every real number $\alpha > 1$, the distances d_α on $A^{\mathbb{N}}$ defined by

$$d_\alpha(\mathbf{x}, \mathbf{y}) = \alpha^{-\inf\{n \in \mathbb{N} | x_n \neq y_n\}}$$

generates the same topology.

(3) Show that the set of periodic words is dense in $A^{\mathbb{N}}$.

(4) Show that $(\mathbf{x}^{(n)})_n$ is a Cauchy sequence in $A^{\mathbb{N}}$ if and only if $d(\mathbf{x}^{(n)}, \mathbf{x}^{(n+1)})$ goes to 0 as n goes to infinity.

(5) A topological space is *connected* if it is not the union of two disjoint non-empty open set. A subset Y of a topological space is connected if, endowed with the subspace topology, it is a connected space. A subset Y of a topological space is a *connected component* if it is connected and maximal for the inclusion (among connected subsets).
What are the connected components of $A^{\mathbb{N}}$?

4. Automatic and morphic words

(1) Consider the *period-doubling sequence* which is the fixed point pf the morphism $0 \mapsto 01$ and $1 \mapsto 00$. Prove that d_n is equal to $\nu_2(n+1) \mod 2$ where $\nu_2(k)$ is the exponent of the largest power of 2 dividing k.

(2) Define a sequence of words $v_0 = ab$ and $v_{i+1} = av_0v_1 \cdots v_ib$ for all $i \geq 0$.
- Show that $h^i(v_0) = v_i$, where h is the morphism defined by $a \mapsto aab$, $b \mapsto b$.
- Show that
$$h^i(a) = ab^{\nu_2(1)}ab^{\nu_2(2)}ab^{\nu_2(3)} \cdots ab^{\nu_2(2^i)}$$
for $i \geq 0$, where ν_2 is defined as in the previous exercise.

(3) Consider the 4-uniform morphism given by $f : \mathtt{a} \mapsto \mathtt{abcd}, \mathtt{c} \mapsto \mathtt{cdcd}, \mathtt{d} \mapsto \mathtt{cdcd}, \mathtt{b} \mapsto \mathtt{eeee}, \mathtt{e} \mapsto \mathtt{bbbb}$ and the coding $g : \mathtt{a} \mapsto 1, \mathtt{b} \mapsto 0, \mathtt{c} \mapsto 1, \mathtt{d} \mapsto 0, \mathtt{e} \mapsto 1$. The first few symbols in $f^\omega(\mathtt{a})$ are
$$\mathtt{abcde}^4(\mathtt{cd})^4\mathtt{b}^{16}(\mathtt{cd})^{16}\mathtt{e}^{64} \cdots .$$
Study the frequencies of occurrence of the symbols in $f^\omega(\mathtt{a})$ and $g(f^\omega(\mathtt{a}))$ respectively.

(4) Define a sequence $(u_n)_{n\geq 0}$ of words over $\{0, 1\}$ as follows: $u_0 = 0$ and
$$u_{n+1} = u_n\tau(u_n), \quad \forall n \geq 0$$
where the morphism τ is defined by $\tau(j) = 1 - j$, $j \in \{0, 1\}$. Prove that $(u_n)_{n\geq 0}$ is converging to the Thue-Morse word.

(5) Let $\mathbf{t}$ be the Thue–Morse word obtained as $\mu^\omega(0) = 0110 \cdots$ where $\mu(0) = 01$ and $\mu(1) = 10$. We define two sequences $(u_n)_{n\geq 0}$ and $(v_n)_{n\geq 0}$ of words as follows: $u_0 = 0$ and $v_0 = 1$, and $u_{n+1} = u_nv_n$ and $v_{n+1} = v_nu_n$ for all $n \geq 0$. Prove that
- $u_n = \mu^n(0)$ and $v_n = \mu^n(1)$;
- $v_n = \tau(u_n)$ and $u_n = \tau(v_n)$, τ defined in the previous example;
- for n even, u_n and v_n are palindromes;
- for n odd, the reversal of u_n is equal to v_n.

(6) Let $n \geq 2$. Define the *generalized Thue-Morse word* over $\{0, \ldots, n-1\}$ obtained as $\gamma^\omega(0)$ where the images of the letters are cyclic permutations of $012 \cdots n$,
$$\gamma(i) = i\,(i+1)\,(i+2) \cdots n\,0\,1\, \cdots\, (i-1).$$
Prove that w_k is the sum-of-digits of the base-n expansion of k, modulo n.

(7) Consider the sequence $\mathbf{a} = (a_n)_{n\geq 0} = 1264224288 \cdots$ that gives the least significant non-zero digit in the base-10 expansion of $n!$.
- Prove that for $n \geq 2$, this digit is even.
- Show that $\mathbf{a}$ is a 5-automatic sequence.
- Give a 5-uniform morphism h and a coding g such that $\mathbf{a} = g(h^\omega(b))$.
- Show that $\mathbf{a}$ is not eventually periodic.

(8) Suppose that a word $\mathbf{w}$ is generated by a k-uniform morphism but also by a ℓ-uniform morphism. Show that $\mathbf{w}$ is generated by a $k\ell$-uniform morphism (and possibly an extra coding).

(9) Show that the Rudin–Shapiro sequence is not the fixed point of any non-trivial morphism.

(10) Define a sequence $(u_n)_{n\geq 0}$ of words as follows: $u_0 = a$, $u_1 = ab$ and
$$u_{n+2} = u_{n+1}u_n, \quad \forall n \geq 0.$$
Prove that $(u_n)_{n\geq 0}$ is converging to the Fibonacci word (fixed point of $a \mapsto ab$, $b \mapsto a$).