Binary Amiable Words [1]

Adrian ATANASIU
Faculty of Mathematics and Computer Science
Bucharest University
Str. Academiei 14, sector 1
010014 Bucharest, Romania
E-mail: `aadrian@gmail.com`

[1] This paper is dedicated to my colleague and friend Alexandru Mateescu, the first author of Parikh matrix mapping.

1 Notations

- The set of all positive integers is denoted by $\mathcal{N}$.
- Let Σ be an alphabet. The set of all words over Σ is Σ^*; if λ is the empty word, then the set of nonempty sequences is $\Sigma^+ = \Sigma^* \setminus \{\lambda\}$.
  For $\alpha \in \Sigma^*$, $|\alpha|$ denotes the length of α.
- For any finite set A we denote $|A|$ the number of elements contained by A.
- The mirror image of a word $\alpha \in \Sigma^*$, denoted $mi(\alpha)$, is defined as: $mi(\lambda) = \lambda, \quad mi(a_1 a_2 \dots a_n) = a_n \dots a_2 a_1$, where $a_i \in \Sigma$, $1 \leq i \leq n$.
- A word α is a "palindrome" iff $\alpha = mi(\alpha)$.
- An ordered alphabet Σ is an alphabet $\Sigma = \{a_1, a_2, \dots, a_k\}$ with a relation of order ("$<$") on it. In this paper we use only a binary ordered alphabet $\Sigma = \{a < b\}$.
- Let $x \in \Sigma$ be a letter. The number of occurences of x in a word $\alpha \in \Sigma^*$ is denoted by $|\alpha|_x$.
- Let $u, v \in \Sigma^+$. The word u is a scattered subword of v if there exists a word w such that $v \in shuffle(u, w)$.

  We denote by $|\alpha|_u$ the number of occurences of u in $v = \alpha$ as a scattered subword.

2 Parikh mapping and Parikh matrix mapping

In *R.J. Parikh - On the context-free languages* (J. Assoc. Comput. Mach., 13, 1966, pp 570-581) the notion of Parikh mapping is defined.

Let $\Sigma = \{a_1 < a_2 < \ldots < a_k\}$ be an ordered alphabet. The Parikh mapping is a mapping

$$\Psi : \Sigma^* \longrightarrow \mathcal{N}^k$$

defined as $\Psi(\alpha) = (|\alpha|_{a_1}, |\alpha|_{a_2}, \ldots, |\alpha|_{a_k})$.

The Parikh vector of α is $(|\alpha|_{a_1}, |\alpha|_{a_2}, \ldots, |\alpha|_{a_k})$.

References:

- Al. Mateescu, A. Salomaa, K. Salomaa, S. Yu - *On the extension of the Parikh mapping* (2000);
- A. Atanasiu, C. Martin - Vide, Al. Mateescu: *On the injectivity of Parikh matrix mapping* (2001);
- A. Atanasiu, C. Martin - Vide, Al. Mateescu: *Codifiable languages and Parikh matrix mapping* (2001);
- S. Fosse, G. Richmomme - *Some characterisations of Parikh matrix equivalent binary words* (2004).

A triangle matrix is a square matrix $m = (m_{ij})_{1\le i,j\le k}$ such that:

- $m_{ij} \in \mathcal{N} \quad (1 \le i, j \le k)$,
- $m_{ij} = 0$ for all $1 \le j < i \le k$,
- $m_{ii} = 1$ $(1 \le i \le k)$.

The set of all triangle matrices of dimension k is denoted by $\mathcal{M}_k$.

The set $\mathcal{M}_k$ is a monoid with respect to multiplication of matrices and has a unit which is the unit matrix of dimension k.

Definition 2.1 *Let $\Sigma = \{a_1 < a_2 < \ldots < a_k\}$ be an ordered alphabet, where $k \ge 1$. The Parikh matrix mapping is the morphism*

$$\Psi_{M_k} : \Sigma^* \longrightarrow \mathcal{M}_{k+1}$$

defined as follows:

If $\Psi_{M_k}(a_q) = (m_{ij})_{1\le i,j\le k+1}$ then $m_{ii} = 1,\ m_{q,q+1} = 1$ and all other elements are zero.

The basic property of Parikh matrix mapping is:
If $\Psi_{M_k}(\alpha) = (m_{ij})_{1\le i,j\le k+1}$ then

- $m_{ij} = 0$ for all $1 \le j < i \le k + 1$,
- $m_{ii} = 1$ for all $1 \le i \le k + 1$,
- $m_{i,j+1} = |\alpha|_{a_i \ldots a_j}$ for all $1 \le i \le j \le k$.

In particular, for $k = 2$, the Parikh matrix mapping

$$\Psi_{M_2} : \Sigma^* \longrightarrow \mathcal{M}_3$$

is defined as follows:
for each $\alpha \in \{a < b\}^*$

$$\Psi_{M_2}(\alpha) = \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}$$

where $x = |\alpha|_a,\ y = |\alpha|_b,\ z = |\alpha|_{ab}$.

Because in this paper we are going to work only with binary sequences, we can denote $\Psi_{M_2}(\alpha)$ with M_α without loss of generality.

A matrix $M \in \mathcal{M}_3$ with the property $M = M_\alpha$ for a particular word $\alpha \in \Sigma^*$ is called *Parikh matrix*.

If $\Sigma = \{a, b\}$ then

$$M_a = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad M_b = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

and (the property of morphism – Definition 2.1):

$$M_{\alpha\beta} = M_\alpha M_\beta, \quad \forall \alpha, \beta \in \Sigma^*$$

3 Amiable words

Definition 3.1 *Two words* $\alpha, \beta \in \Sigma^*$ $(\alpha \neq \beta)$ *are called "amiabile" iff* $M_\alpha = M_\beta$.

Let denote by $\alpha \sim_a \beta$ the property that α and β are amiable words.

The relation $\sim_a$ is an equivalence relation (S. Fosse and G. Richmomme defined a congruence relation $\equiv_2$ very close to $\sim_a$).

"Palindromic amiable" words: the words α, β are amiables and palindromes.

In *On the injectivity of Parikh matrix mapping* another equivalence relation is defined. Namely:

$x \equiv_{pa} y$ iff there are $\alpha, \beta \in \Sigma^+$ palindromic amiables words so that $x = u\alpha v,\ y = u\beta v$.

$\equiv_{pa}^*$ is the reflexive and transitive closure of $\equiv_{pa}$.

Example 3.1 $aabbabaaa \equiv_{pa}^* babaaaab$. *Indeed,*
$a\underbrace{abba}baaa \equiv_{pa} ab\underbrace{aabbaa}a \equiv_{pa} \underbrace{abba}aaaba \equiv_{pa} ba\underbrace{abaaaba} \equiv_{pa} babaaaaab$.

The next result can be proved using Theorem 3.9 (the reference above):

Proposition 3.1 *For any* $\alpha, \beta \in \Sigma^*$, $\alpha \sim_a \beta \iff \alpha \equiv_{pa}^* \beta$.

4 Classes of binary amiable words

Let $\Sigma = \{a < b\}$ be a binary ordered alphabet.

Lemma 4.1

1. *If* $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \Sigma^*, \quad \alpha_1 \sim_a \beta_1,\ \alpha_2 \sim_a \beta_2$*, then* $\alpha_1\alpha_2 \sim_a \beta_1\beta_2$.
2. *If* $\alpha, \beta, \gamma \in \Sigma^*$*, then* $\alpha ab\beta ba\gamma \sim_a \alpha ba\beta ab\gamma$.

Proof: The first assertion is obvious.

If

$$M_{ab} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad M_{ba} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad M_{\beta} = \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}$$

then

$$M_{ab}M_{\beta}M_{ba} = M_{ba}M_{\beta}M_{ab} = \begin{pmatrix} 1 & x+2 & x+y+z+2 \\ 0 & 1 & y+2 \\ 0 & 0 & 1 \end{pmatrix}$$

q.e.d.

Remark 4.1 *abba* $\sim_a$ *baab.*

Moreover, abba and baab are the binary amiable words of minimal length.

For a word $\alpha \in \Sigma^*$ we denote by C_α its equivalence class:

$$C_\alpha = \{\beta \in \Sigma^* | \beta \sim_a \alpha\}$$

There is an isomorphism between the monoid of 3×3 Parikh matrices and the monoid $(C_\alpha, \circ)$ where the rule $\circ$ is defined

$$C_\alpha \circ C_\beta = C_{\alpha\beta}.$$

Having a Parikh matrix

$$M_\alpha = \begin{pmatrix} 1 & p & q \\ 0 & 1 & n \\ 0 & 0 & 1 \end{pmatrix}$$

with $n, p, q \in \mathcal{N}$, how many words contains C_α ?

In particular, in which conditions $C_\alpha = \{\alpha\}$?

The last question characterizes the injectivity of the Parikh matrix mapping Ψ_{M_2}.

Let $\alpha \in \Sigma^+$ be a binary sequence (the case $\alpha = \lambda$ is trivial and will be ignored); α can be represented in the following form (by detailing the appareances of letter b):

$$\alpha = a^{x_1} b a^{x_2} b \ldots a^{x_n} b a^{x_{n+1}} \tag{1}$$

The Parikh matrix

$$M_\alpha = \begin{pmatrix} 1 & p & q \\ 0 & 1 & n \\ 0 & 0 & 1 \end{pmatrix}$$

corresponds to this word if and only if

$$(x_1, x_2, \ldots, x_{n+1}) \in \mathcal{N}^{n+1}$$

is a solution of the system

$$\left\{ \begin{array}{ccccccccccccc} x_1 & + & x_2 & + & \ldots & + & x_{n-1} & + & x_n & + & x_{n+1} & = & p \\ nx_1 & + & (n-1)x_2 & + & \ldots & + & 2x_{n-1} & + & x_n & & & = & q \end{array} \right. \tag{2}$$

The number of solutions of the system (2) equals $|C_\alpha|$.

Remark 4.2 *The number of solutions of the system* (2) *equals also the number of solutions of the equation* $x_1 + x_2 + \ldots + x_n = q$ *where* $x_i \in \mathcal{N}$ $(1 \leq i \leq n)$ *and* $0 \leq x_1 \leq \ldots \leq x_n \leq p$

(AMM - On the injectivity of Parikh matrix mapping).

Example 4.1 *Let*

$$M = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{pmatrix}$$

be a Parikh matrix with $p = 2,\ q = 4,\ n = 4.$

The associated system is

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 + x_5 &= 2 \\ 4x_1 + 3x_2 + 2x_3 + x_4 &= 4 \end{aligned}$$

This system with 5 *variables has three solutions in* $\mathcal{N}^5$. *Namely:*

1. $(x_1, x_2, x_3, x_4, x_5) = (1, 0, 0, 0, 1)$ *which corresponds to the word* $\beta_1 = abbbba$*;*

2. $(x_1, x_2, x_3, x_4, x_5) = (0, 1, 0, 1, 0)$ *which corresponds to the word* $\beta_2 = babbab$*;*

3. $(x_1, x_2, x_3, x_4, x_5) = (0, 0, 2, 0, 0)$ *which corresponds to the word* $\beta_3 = bbaabb$*.*

Thus the set of sequences with the Parikh matrix M *is*

$$C = \{abbbba, babbab, bbaabb\}$$

Theorem 4.1 *Let*

$$M = \begin{pmatrix} 1 & p & q \\ 0 & 1 & n \\ 0 & 0 & 1 \end{pmatrix}$$

be a matrix with $p, q, n \in \mathcal{N}$.
M is a Parikh matrix iff $q \in [0, np\,]$.

Proof: If M is a Parikh matrix, there is at least a binary word α of the type (1) which verifies the system (2).

We have to prove that for this word the relations $x_1+x_2+\ldots x_n+x_{n+1} = p$ and $nx_1 + (n-1)x_2 + \ldots + x_n \leq np$ are fulfilled.

The equality results directly from the construction of (1).

To prove the inequality, let us evaluate
$\sum_{i=1}^{n}(n+1-i)x_i = n\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}(i-1)x_i = n(p-x_{n+1}) - \sum_{i=1}^{n}(i-1)x_i = np - \sum_{i=1}^{n+1}(i-1)x_i \leq n \cdot p$

because $x_i \in \mathcal{N}$ (therefore their values are at least 0).

Assume that $n, p, q \in \mathcal{N}$ such that $q \leq n \cdot p$.
A solution $(x_1, x_2, \ldots, x_{n+1}) \in \mathcal{N}^{n+1}$ of the system (2) can be

$$\begin{aligned} x_1 &= \left\lfloor \frac{q}{n} \right\rfloor, \\ x_{i+1} &= \left\lfloor \frac{q - \Sigma_{j=1}^{i}(n+1-j)x_j}{n-i} \right\rfloor \ (1 \leq i \leq n-1), \\ x_{n+1} &= p - \sum_{i=1}^{n} x_i \end{aligned} \tag{3}$$

Therefore we found a binary word $\alpha \in \Sigma^*$ of the form (1) so that $M = M_\alpha$. q.e.d.

Remark 4.3 *In the cases $n = 0$ or $p = 0$ we immediately conclude that $q = 0$. So, in the following there will be treated only the nontrivial situations $n \cdot p \neq 0$.*

Example 4.2 *Let's consider the values $n = 3,\ p = 3,\ q = 4$.*

$$x_1 = \left\lfloor \frac{q}{n} \right\rfloor = 1,$$

$$x_2 = \left\lfloor \frac{q - nx_1}{n-1} \right\rfloor = 0,$$

$$x_3 = \left\lfloor \frac{q - nx_1 - (n-1)x_2}{n-2} \right\rfloor = 1,$$

$$x_4 = p - x_1 - x_2 - x_3 = 1$$

Indeed, $(1, 0, 1, 1)$ is a solution of the system

$$\begin{array}{lcl} x_1 + x_2 + x_3 + x_4 & = & 3 \\ 3x_1 + 2x_2 + x_3 & = & 4 \end{array}$$

Obviously, this is not the only solution.

Other solutions of the system are $(0, 2, 0, 1)$ and $(0, 1, 2, 0)$.

Lemma 4.2 *If* $M_\alpha = \begin{pmatrix} 1 & p & q \\ 0 & 1 & n \\ 0 & 0 & 1 \end{pmatrix}$ *then* $M_{mi(\alpha)} = \begin{pmatrix} 1 & p & np-q \\ 0 & 1 & n \\ 0 & 0 & 1 \end{pmatrix}$.

Proof:

Let us consider $\alpha = a^{x_1}ba^{x_2}b\ldots a^{x_n}ba^{x_{n+1}}$, which corresponds to the system (2).

Then for $mi(\alpha) = a^{x_{n+1}}ba^{x_n}b\ldots a^{x_2}ba^{x_1}$ will correspond a system in which the first equation is the same, but the second equation is

$$nx_{n+1} + (n-1)x_n + \ldots + x_2 = A$$

where A is a value to be found.

We evaluate
$np = n(x_1 + x_2 + \ldots + x_{n+1}) = [nx_1 + (n-1)x_2 + \ldots + x_n] + x_2 + 2x_3 + \ldots + nx_{n+1} = q + A$.

Thus $A = np - q$ and the system built for the word $mi(\alpha)$ corresponds to the Parikh matrix $M_{mi(\alpha)}$ defined above. q.e.d.

Theorem 4.2 *Let $\alpha \in \Sigma^+$ be a binary word with the Parikh matrix*

$$M_\alpha = \begin{pmatrix} 1 & p & q \\ 0 & 1 & n \\ 0 & 0 & 1 \end{pmatrix}, \quad p, q, n \in \mathcal{N}$$

For each of the cases

1. $n \leq 1$;
2. $p \leq 1$;
3. $q \in \{0, 1, np-1, np\}$,

   $|C_\alpha| = 1$ *(thus $C_\alpha = \{\alpha\}$) holds.*

Proof:

1. For $n = 0$ (thus $q = 0$) there is only one sequence: $\alpha = a^p$. For $n = 1$ (therefore, accordingly with the Theorem 4.1, $p \leq q$) there is again only one solution: $\alpha = a^q b a^{p-q}$.
2. Similarly.
3. We have to prove that for the values of q given above, the system (2) has only one solution.

   Using Lemma 4.2 we conclude that $|C_\alpha| = |C_{mi(\alpha)}|$; it is enough to prove the assertion only for values $q = 0$ and $q = 1$.

   - $\mathbf{q = 0}$: The diophantic equation

     $$nx_1 + (n-1)x_2 + \ldots + x_n = 0$$

     has only one solution: $x_i = 0$ $(i = 1, \ldots, n)$.
     It results $x_{n+1} = p$.
     The solution $(0, \ldots, 0, p)$ is unique and corresponds to the binary word $\alpha = b^n a^p$.

- **q = 1**: The diophantic equation

$$nx_1 + (n-1)x_2 + \ldots + x_n = 1$$

has the unique solution

$$x_1 = x_2 = \ldots = x_{n-1} = 0,\ x_n = 1.$$

From the first equation of the system (2) results
$x_{n+1} = p - 1$.
The constraint $1 \leq n \cdot p$ (obtained from Theorem 4.1 with $q = 1$) assures $n \cdot p \neq 0$.
In peculiar, $p \geq 1,\ n \geq 1$; thus the solution

$$(0, \ldots, 0, 1, p-1)$$

can be constructed.
It is unique and corresponds to the word $\alpha = b^{n-1}aba^{p-1}$.

q.e.d.

Theorem 4.3 *Let $\alpha \in \Sigma^+$ with the Parikh matrix*

$$M_\alpha = \begin{pmatrix} 1 & p & q \\ 0 & 1 & n \\ 0 & 0 & 1 \end{pmatrix}$$

If $n \geq 2,\ p \geq 2 \quad$ *and* $\quad q \in [2, np-2]$,
then $|C_\alpha| \geq 2$.

Proof: Two cases are possible:

1. $\mathbf{p} \geq \mathbf{q}$: Let us denote $p = q + s$ $(s \geq 0)$.

   Using the hypothesis, $(0, \ldots, 0, t, q-2t, t+s)$ (with 0 in the first $n-2$ positions) is a solution of the system (2).

   It corresponds to the word $\alpha = b^{n-2}a^t ba^{q-2t} ba^{t+s}$.

   The constraint $q - 2t \geq 0$ assures a distinct solution for each $t \in \left[0, \dfrac{q}{2}\right]$.

   Because $q \geq 2$, there are at least two solutions.

2. $\mathbf{p} < \mathbf{q}$: (3) is a solution of the system (2).

   For another (distinct) solution, let us consider $q = sp + r$ with $1 \leq s < n,\ 0 \leq r < p$. Then

   $$x_{n-s} = r,\ x_{n-s+1} = p - r,\ x_i = 0\ (i \neq n-s, n-s+1)$$

   is a solution for the system (it corresponds to the sequence $b^{n-s-1}a^r ba^{p-r} b^s$).

   Indeed, the first equation is verified with $r + (p-r) = p$, and the second with $(s+1)r + s(p-r) = sp + r = q$.

   The fact that these two solutions are distinct can be easily proven.

q.e.d.

Theorems 4.2 and 4.3 cover all cases. Therefore all the cases when $C_\alpha = \{\alpha\}$ are defined by the Theorem 4.2.

Unfortunately it is not easy to check an upper limit for the number of words which are amiable with a given binary word α.

In *AMM - On the injectivity of the Parikh matrix mapping* the next result holds (Theorem 4.7):

For a Parikh matrix

$$M_\alpha = \begin{pmatrix} 1 & p & q \\ 0 & 1 & n \\ 0 & 0 & 1 \end{pmatrix}$$

we have $|C_\alpha| = \phi(q, p, n)$
where $\phi(q, p, n)$ is recursively defined as follows:

$$\phi(q,p,n) = \sum_{i=0}^{min\{p,q\}} \phi(q-i, i, n-1), \qquad \phi(q,p,1) = \begin{cases} 1 & , \quad q \le p \\ 0 & , \quad q > p \end{cases}$$

The problem is that this function $\phi(q, p, n)$ is very difficult to be evaluated.

The next Theorem assures a lower bound of this limit:

Theorem 4.4 *There is at least a word $\alpha \in \Sigma^+$ with*

$$|C_\alpha| \ge \frac{1}{np+1} \begin{pmatrix} n+p \\ p \end{pmatrix}$$

Proof:

For two given values $n, p \in \mathcal{N}$ exists $\begin{pmatrix} n+p \\ p \end{pmatrix}$ binary sequences with p $a's$ and n $b's$.

For each $q \in [0, np]$ exists (Theorem 4.1) a Parikh matrix, thus it exists at least a word $\alpha \in \Sigma^*$.

There are $np + 1$ possible classes of amiable words, thus at least one of them will contain a number of binary words greater than the median. q.e.d

5 Theorem of characterization for amiable words

Let us consider an equivalence class C, corresponding to a given Parikh matrix M.

We define the unoriented graph $\Gamma_M = (V, E)$ as follows:

- $V = C$;
- $(\alpha, \beta) \in E \Longleftrightarrow \exists \gamma_1, \gamma_2, \gamma_3 \in \{a, b\}^*, \ \alpha = \gamma_1 ab \gamma_2 ba \gamma_3,$
  $\beta = \gamma_1 ba \gamma_2 ab \gamma_3.$

From Lemma 4.1 it results that the sequences α, β are amiable, thus they belong to the same equivalence class.

Theorem 5.1 *The graph Γ_M is connected.*

Proof:

The assertion is trivial for $|C| = 1$.

Let us consider only the situation studied by the Theorem 4.3.

We have to show that for every $\alpha, \beta \in C$ there exists a path between the nodes α and β; hence there is a sequence of transformations of the type $ab - ba$ by which the word β can be obtained from α (and viceversa).

On the set C we define the next derivation rule:

$$\forall \alpha, \beta \in C$$
$$\{\alpha \implies \beta\} \Longleftrightarrow \{\exists \gamma_1, \gamma_2, \gamma_3 \in \{a, b\}^*, \ \alpha = \gamma_1 ab \gamma_2 ba \gamma_3, \ \beta = \gamma_1 ba \gamma_2 ab \gamma_3\}$$

Lemma 4.1 assures that this derivation is well defined on C.

For $\alpha \in C$ we can apply this rule as long as it is possible.

Because, sometimes, several variants can appear in every step, we will define a constraint in order to choose only one continuation.

Namely

(i) *Let* $\alpha = a^{x_1}ba^{x_2}b\dots a^{x_n}ba^{x_{n+1}} \in C$ *and* x_i, x_j *the first two positive exponents such that* $j > i+1$.

Then we can write $\alpha \Longrightarrow \beta$, *where* $\alpha = \gamma_1 a^{x_i} b\gamma_2 ba^{x_j}\gamma_3$ *and* $\beta = \gamma_1 a^{x_i-1}ba\gamma_2 aba^{x_j-1}\gamma_3$.

This rule cannot be applied when there remain at most two positive exponents.

By detalying:

- If it remains only one positive exponent, then α has the form $\alpha = b^* a^p b^*$;
- If it remains two consecutive positive exponents, then α has the form $\alpha = b^* a^+ b a^+ b^*$

For each of the two cases discussed in the proof of Theorem 4.3, the word (denoted α_0) when the derivation rule cannot be applied accordingly to the constraint (i) is unique.

Let $q = sp + r$ where $0 \le s < n,\ 0 \le r < p$ (the case $s = 0$ covers the variant $q < p$). Then

$$\alpha_0 = b^{n-s-1}a^r ba^{p-r}b^s$$

Therefore, for each $\alpha \in C$ we have a (possible empty) sequence $\alpha \Longrightarrow \dots \Longrightarrow \alpha_0$.

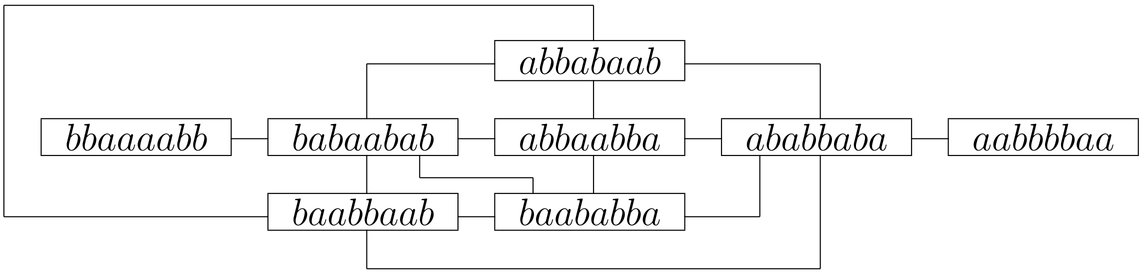
In the graph Γ_M associated to C, this assures (at least) a path from the node α to the node α_0.

The theorem results now from the fact that the graph Γ_M is unoriented: for each $\alpha, \beta \in C$, $(\alpha \neq \beta)$ there is a path between α and β which go through α_0. q.e.d.

Example 5.1 *Let $n = p = 4,\ q = 8$. The system (2) has eight solutions which correspond to the set of amiable words*

$C = \{aabbbbaa, ababbaba, abbaabba, baabbaab, babaabab, bbaaaabb,$
$baababba, abbabaab\}$

Using the derivation rule defined above, we can construct the graph Γ_M:



The word α_0 considered by the constraint (i) is $\alpha_0 = bbaaaabb$.

Some considerations regarding palindromes.

Lemma 5.1 *If* $p = n$ *then* $C_{mi(\alpha)} = C_{\overline{\alpha}}$*, where* $\overline{\alpha}$ *is obtained from the word* α *by replacing the letter* a *with* b *and viceversa.*

Proof:

We have to show the equality between Parikh matrices

$$M_{mi(\alpha)} = M_{\overline{\alpha}}$$

The only element which can differ is the number of scattered ab.

By computing,

$$|\overline{\alpha}|_{ab} = |\alpha|_{ba} = |mi(\alpha)|_{ab}$$

q.e.d.

Lemma 5.2 *Let* $\alpha, \beta \in \Sigma^*$ *be two binary palindromes with the same Parikh vector. Then* $\alpha \sim_a \beta$*.*

Proof: If α is palindrom, then $\alpha = mi(\alpha)$; thus $q = np - q$, that is (Lemma 4.2) $2q = np$. q.e.d.

As a result of this lemma, any binary palindrom α with $(|\alpha|_a, |\alpha|_b)$ fixed will have the same value for $|\alpha|_{ab}$, therefore the same Parikh matrix.

Remark 5.1 *From Lemma 5.2 it results that all palindromes are amiable and thus they are in the same equivalence class* C*.*

But the reverse is not true: not all words from C *are palindromes.*

For example, if $n = p = 4$ *we have* $q = 8$*. However*

$$baababba \sim_a aabbbbaa$$

although the second word is a palindrom and the first word is not a palindrom (Example 5.1).

This remark corrects Corollary 3.4 *(On the injectivity of Parikh matrix mapping).*

6 Distances defined on classes of amiable words

Let C be a class of binary amiable words defined by the Parikh matrix M.

We will consider three distances on C and then we will compare the words via these distances.

Let $\alpha, \beta \in C,\ \alpha = a_1 a_2 \ldots a_n,\ \beta = b_1 b_2 \ldots b_n\ (a_i, b_i \in \{a, b\})$ be two distinct words. Then:

1. The Hamming distance $d_H(\alpha, \beta)$ equals the number of different characters between α and β. Formally

   $$d_H(\alpha, \beta) = \sum_{i=1}^{n} (a_i \oplus b_i)$$

   where $\oplus : \Sigma \times \Sigma \longrightarrow \{0, 1\}$ is defined by

   $$a \oplus a = b \oplus b = 0, \quad a \oplus b = b \oplus a = 1$$

2. The Rank distance $d_R(\alpha, \beta)$ counts the "ecart" between the similar characters from α and β. Formally (L.P. Dinu, A. Sgarro - *A low complexity distance for DNA strings*, manuscript):

   $$d_R(\alpha, \beta) = \sum_{x \in \alpha} |ord_\alpha(x) - ord_\beta(x)|$$

   where $ord_u(x)$ represents the position of the character x in the string u, counted from the left to right.

3. The amiable distance $d_A(\alpha, \beta)$ defined as follows: if

   $$\alpha = a^{x_1} b a^{x_2} b \ldots a^{x_n} b a^{x_{n+1}}, \beta = a^{y_1} b a^{y_2} b \ldots a^{y_n} b a^{y_{n+1}}$$

   then

   $$d_A(\alpha, \beta) = \frac{1}{2} \sum_{i=1}^{n+1} |x_i - y_i|$$

Proposition 6.1 *d_A is a distance and $d_a(\alpha, \beta) \in \mathcal{N} \quad (\alpha, \beta \in C)$.*

Theorem 6.1 *Let $\alpha, \beta \in C$. Then*

1. *$d_R(\alpha, \beta) = 4k$, where k is the shortest path length in Γ_M between α and β.*

2. *$d_A(\alpha, \beta) \leq d_H(\alpha, \beta) \leq d_R(\alpha, \beta)$.*

Proof:
A complete proof has many cases to take into acount; we sketch only the basic ideas.
For two words $\alpha, \beta \in C$ we consider the shortest path beetween them.

Below are treated only paths of the length 1 and 2.

First, let $\alpha, \beta \in C$ be two words connected by an edge; thus $\alpha = \gamma_1 ab\gamma_2 ba\gamma_3$, $\beta = \gamma_1 ba\gamma_2 ab\gamma_3$ for some $\gamma_1, \gamma_2, \gamma_3 \in \Sigma^*$.
By applying the definitions it results that
$d_R(\alpha, \beta) = d_H(\alpha, \beta) = 4$, $d_A(\alpha, b) = 2$.

Now, if $\alpha, \beta \in C$ are connected by a 2-edges path, there is $\delta \in C$ such that $\alpha \Longrightarrow \delta \Longrightarrow \beta$.
Therefore we can write $\alpha = \gamma_1 ab\gamma_2 ba\gamma_3$, $\delta = \gamma_1 ba\gamma_2 ab\gamma_3 = \gamma_1' ab\gamma_2' ba\gamma_3'$, $\beta = \gamma_1' ba\gamma_2' ab\gamma_3'$ for some $\gamma_i, \gamma_i' \in \Sigma^*$ $(1 \leq i \leq 3)$.

Three main cases can appear:

1. Every letter from these eight letters used in 2-step derivation is moving once. Then
$d_R(\alpha, \beta) = d_R(\alpha, \delta) + d_R(\delta, \beta) = 4 + 4 = 8$
$d_H(\alpha, \beta) = d_H(\alpha, \delta) + d_H(\delta, \beta) = 4 + 4 = 8$
$d_A(\alpha, \beta) \leq d_A(\alpha, \delta) + d_A(\delta, \beta) = 2 + 2 = 4$
The value of d_A depends of the relative positions of pairs $ab - ba$ used at each step: For example $d_A(ababbaba, babaabab) = 2$, $d_A(abbabaab, baababba) = 3$, $d_A(abbaabba, baabbaab) = 4$.

2. One of these eight letters is moving twice. Two subcases (via reflexion) can appear:

   (a) $\alpha = uabbvbawbax,\ \delta = ubabvabwbax,\ \beta = ubbavabwabx.$ Then $d_R(\alpha,\beta) = 2+1+1+2+2 = 8,\ d_H(\alpha,\beta) = 6,$ $d_A(\alpha,\beta) = \frac{(1+1)+(1+1)+(1+1)}{2} = 3$

   (b) $\alpha = uaabvbawbax,\ \delta = uabavabwbax,\ \beta = ubaavabwabx.$ Then $d_R(\alpha,\beta) = 8,\ d_H(\alpha,\beta) = 6,\ d_A(\alpha,\beta) = \frac{(2+2)+(1+1)+(1+1)}{2}.$

3. Two letters are moving twice. Four cases (via reflexion) can appear:

   (a) $\alpha = uaabvbaaw,\ \delta = uabavabaw,\ \beta = ubaavaabw.$ Then $d_R(\alpha,\beta) = 4{+}4 = 8,\ d_H(\alpha,\beta) = 4,\ d_A(\alpha,\beta) = \frac{(2+2)+(2+2)}{2} = 4.$

   (b) $\alpha = uaabvbbaw,\ \delta = uabavbabw,\ \beta = ubaavabbw.$ Then $d_R(\alpha,\beta) = 4{+}4 = 8,\ d_H(\alpha,\beta) = 4,\ d_A(\alpha,\beta) = \frac{(2+2)+(1+1)}{2} = 3.$

   (c) $\alpha = uabbvbbaw,\ \delta = ubabvbabw,\ \beta = ubbavabbw.$ Then $d_R(\alpha,\beta) = 8,\ d_H(\alpha,\beta) = 4,\ d_A(\alpha,\beta) = \frac{(1+1)+(1+1)}{2} = 2.$

   (d) $\alpha = uabbvbaaw,\ \delta = ubabvabaw,\ \beta = ubbavaabw.$ Then $d_R(\alpha,\beta) = 8,\ d_H(\alpha,\beta) = 4,\ d_A(\alpha,\beta) = 4.$

q.e.d.

Corollary 6.1 $d_R(\alpha,\beta) = 4 \quad \Longrightarrow \quad d_H(\alpha,\beta) = 4.$

Example 6.1 *Using the class C defined in Example 5.1, we can construct a table containing these three distances*

$$(d_A(\alpha,\beta), d_H(\alpha,\beta), d_R(\alpha,\beta))$$

(corresponding to row α and column β):

	aabbbbaa	*ababbaba*	*abbaabba*	*baababba*	*baabbaab*	*babaabab*	*abbabaab*	*bbaaaabb*
aabbbbaa	$(0,0,0)$	$(2,4,4)$	$(2,4,8)$	$(3,4,8)$	$(4,4,8)$	$(4,4,12)$	$(3,4,12)$	$(4,8,16)$
ababbaba	$(2,4,4)$	$(0,0,0)$	$(2,4,4)$	$(2,4,4)$	$(2,4,4)$	$(2,8,8)$	$(2,4,4)$	$(4,4,12)$
abbaabba	$(2,4,8)$	$(2,4,4)$	$(0,0,0)$	$(2,4,4)$	$(4,8,8)$	$(2,4,4)$	$(2,4,4)$	$(2,4,8)$
baababba	$(3,4,8)$	$(2,4,4)$	$(2,4,4)$	$(0,0,0)$	$(2,4,4)$	$(2,4,4)$	$(3,8,8)$	$(3,4,8)$
baabbaab	$(4,4,8)$	$(2,4,4)$	$(4,8,8)$	$(2,4,4)$	$(0,0,0)$	$(2,4,4)$	$(2,4,4)$	$(4,4,8)$
babaabab	$(4,4,12)$	$(2,8,8)$	$(2,4,4)$	$(2,4,4)$	$(2,4,4)$	$(0,0,0)$	$(2,4,4)$	$(2,4,4)$
abbabaab	$(3,4,12)$	$(2,4,4)$	$(2,4,4)$	$(3,8,8)$	$(2,4,4)$	$(2,4,4)$	$(0,0,0)$	$(3,4,8)$
bbaaaabb	$(4,8,16)$	$(4,4,12)$	$(2,4,8)$	$(3,4,8)$	$(4,4,8)$	$(2,4,4)$	$(3,4,8)$	$(0,0,0)$

7 Conclusions

- There remain some open problems if some of above results can be extended to alphabets with more than two letters.
- The characterisation theorem offers more information than it was mentioned here.
- A separate investigation about the characterisation of amiable properties using the graph will be necessary.
- Also the distance d_A is a quite new notion and a through study about its properties remains to be conducted.