Amiable words

Adrian Atanasiu

Bucharest University

1. R.J. Parikh - *On context-free languages*, JACM, 13 (1966), 570-581.

2. Al. Mateescu, A. Salomaa, K. Salomaa, S. Yu - *On the extension of the Parikh mapping*, Theoret. Informatics Appl. 35 (2001), 551-564.

3. A. Atanasiu, C. Martin - Vide, Al. Mateescu - *On the injectivity of Parikh matrix mapping*, Fund. Inf. 49 (2001), 166-180.

4. A. Atanasiu - *Binary amiable words*, IJFCS 18, 2(2007), 387-400.

5. A. Atanasiu, R. Atanasiu, I. Petre - *Parikh Matrices and Amiable Words*, TCS 390, no. 1 (2008), pp. 102-109

6. A. Atanasiu - *Parikh Matrices and the Istrail Morphism*, submitted to IJFCS.

Let Σ be a finite 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.

$|\alpha|_a$ – the number of occurences of the letter $a \in \Sigma$ in the word $\alpha \in \Sigma^*$.

The alphabet used is

$$\Sigma = \{a_1, a_2, \ldots, a_s\}$$

Many results are only for $s = 2$.

Let u, v be words over Σ.

The word u is a scattered subword of v if

$$u = \beta_1\beta_2 \ldots \beta_r, \quad v = \gamma_0\beta_1\gamma_1 \ldots \gamma_{r-1}\beta_r\gamma_r$$

($r \geq 1$ and $\beta_i, \gamma_j \in \Sigma^*$).

$|\alpha|_u$ – the number of occurences of u in α as a scattered subword.

For instance $|a_1a_2a_1a_2|_{a_1a_2} = 3$.

Definition 1. *Let* $\Sigma = \{a_1, a_2, \ldots, a_s\}$ *be an ordered alphabet and* $\mathcal{M}_{s+1}$ *be the set of* $(s+1)$ *- dimensional upper-triangular matrices with nonnegative integral entries and unit diagonal.*

The Parikh matrix mapping is the morphism

$$\Psi_s : \Sigma^* \longrightarrow \mathcal{M}_{s+1}$$

defined: if $k = 1, \ldots, s$ *and*

$$\Psi_s(a_k) = (m_{i,j})_{1 \leq i,j \leq s+1}$$

then

$$m_{i,i} = 1 \ (1 \leq i \leq s+1), \qquad m_{k,k+1} = 1$$

all other elements of the matrix $\Psi_s(a_k)$ *being* 0.

If $|\Sigma| = s$ is fixed, we will denote $\Psi_s(\alpha)$ also by M_α.

A matrix $M \in \mathcal{M}_{s+1}$ with the property $M = M_\alpha$ for a particular word $\alpha \in \Sigma^*$ is called *Parikh matrix*.

Theorem 1. *Consider* $\Sigma = \{a_1, \ldots, a_s\}$ *and* $\alpha \in \Sigma^*$. *The matrix*

$$M_\alpha = \Psi_s(\alpha) = (m_{i,j})_{1 \le i,j \le s+1}$$

has the following properties

- $m_{i,j} = 0$ *for all* $1 \le j < i \le s+1$,
- $m_{i,i} = 1$ *for all* $1 \le i \le s+1$,
- $m_{i,j+1} = |\alpha|_{a_i \ldots a_j}$ $(1 \le i \le j \le s)$.

So, the second diagonal of the Parikh matrix of α gives the Parikh vector of α:

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

Example 1. *For the alphabet* $\Sigma = \{a, b, c, d\}$ *and* $\alpha \in \Sigma^*$*:*

$$M_\alpha = \begin{pmatrix} 1 & |\alpha|_a & |\alpha|_{ab} & |\alpha|_{abc} & |\alpha|_{abcd} \\ 0 & 1 & |\alpha|_b & |\alpha|_{bc} & |\alpha|_{bcd} \\ 0 & 0 & 1 & |\alpha|_c & |\alpha|_{cd} \\ 0 & 0 & 0 & 1 & |\alpha|_d \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Corollary 1. $\forall \alpha, \beta \in \Sigma^*$,

$$M_{\alpha\beta} = M_\alpha M_\beta.$$

Definition 2. *Two words* $\alpha, \beta \in \Sigma^*$ *are called "amiable" iff* $M_\alpha = M_\beta$.

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

The relation $\sim_a$ is obviously an equivalence relation.

Let

$$[\alpha] = \{\beta \mid \alpha \sim_a \beta\}$$

be the equivalence class defined by the non-empty word $\alpha \in \Sigma^*$.

Properties of the classes of amiable words

Let be the alphabet

$$\Sigma = \{a_1, a_2, \ldots, a_s\} \quad (s \geq 2)$$

Lemma 1.

1. *$\sim_a$ is left/right invariant.*

2. $a_i a_j \sim_a a_j a_i, \quad |i-j| \geq 2, \quad (1 \leq i, j \leq s)$;

3. $$a_i a_{i+1} x a_{i+1} a_i \sim_a a_{i+1} a_i x a_i a_{i+1},$$
$\forall x \in \left(\Sigma \setminus \{a_{i-1}, a_{i+2}\}\right)^* \quad (1 \leq i \leq s-1)$
(we consider $a_k = \lambda$ for any $k > s$ or $k < 1$).

Definition 3. *Let us consider* $\alpha, \beta \in \Sigma^*$.

1. α *transforms into* β
   using a type (1) transformation if

$$\alpha = x a_i a_j y, \qquad \beta = x a_j a_i y,$$

   where $x, y \in \Sigma^*,\ a_i, a_j \in \Sigma,\ |i - j| \geq 2$.

2. α *transforms into* β
   using a type (2) transformation if

$$\alpha = x a_i a_{i+1} y a_{i+1} a_i z, \quad \beta = x a_{i+1} a_i y a_i a_{i+1} z,$$

   where $a_i, a_{i+1} \in \Sigma, \quad x, y, z \in \Sigma^*,$
   $|y|_{a_{i-1}} = |y|_{a_{i+2}} = 0$.

- If α transforms into β, then obviously β transforms into α as well (no matter the transformation type).

  We will denote by $\alpha \overset{(i)}{\longleftrightarrow} \beta$ the fact that words α and β transforms one from each other using a type (i) transformation ($i \in \{1, 2\}$).

- If $|\Sigma| = 2$ then only transformations of type (2) are possible.

Lemma 2. *If* $\alpha \overset{(i)}{\longleftrightarrow} \beta$ $(i = 1, 2)$ *then* $\alpha \sim_a \beta$.

For the binary alphabet $\Sigma = \{a, b\}$, the next theorem was proved:

Theorem 2. *(***Theorem of characterization of the classes of amiable words***):*
Let $\alpha, \beta \in \Sigma^*$, $\alpha \sim_a \beta$. *There exists a sequence of type* (2) *transformations so that*

$$\alpha \overset{(2)}{\longleftrightarrow} \quad \dots \quad \overset{(2)}{\longleftrightarrow} \beta.$$

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:

- $V = C$;

- $(\alpha, \beta) \in E \quad \Longleftrightarrow \quad \exists \gamma_1, \gamma_2, \gamma_3 \in \{a, b\}^*,$

$$\alpha = \gamma_1 ab\gamma_2 ba\gamma_3, \quad \beta = \gamma_1 ba\gamma_2 ab\gamma_3.$$

The binary words α, β are amiable, thus they belong to the same equivalence class.

Theorem 3. *The graph Γ_M is connected.*

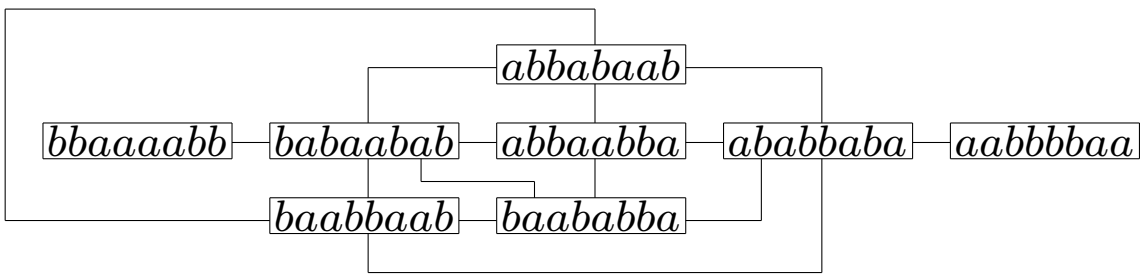
Example 2. *Let be the matrix*

$$M = \begin{pmatrix} 1 & 4 & 8 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{pmatrix}$$

There are eight binary words which have this matrix as Parikh matrix:

$$C = \left\{ \begin{array}{c} aabbbbaa, ababbaba, abbaabba, baabbaab, \\ babaabab, bbaaaabb, baababba, abbabaab \end{array} \right\}$$

The connected graph Γ_M *attached is:*



If $|\Sigma| > 2$ this result is not true.

Example 3. *Let* $\Sigma = \{a, b, c\}$ *and*

$$\alpha = abcbabcbabcbab, \quad \beta = bacabbcabbcbba$$

These words are amiable, having the Parikh matrix

$$M = \begin{pmatrix} 1 & 4 & 16 & 14 \\ 0 & 1 & 7 & 9 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

However, there is no chain of transformations (1) + (2) *between* α *and* β *; therefore, from this point of view, the Theorem* 2 *is not true.*

The problem is issued due to the transformation (2):

$$\alpha = xa_ia_{i+1}ya_{i+1}a_iz, \quad \beta = xa_{i+1}a_iya_ia_{i+1}z,$$

where $a_i, a_{i+1} \in \Sigma, \quad x, y, z \in \Sigma^*$,
$|y|_{a_{i-1}} = |y|_{a_{i+2}} = 0$.

The subword y may contain characters a_{i-1} or a_{i+2} that lead to changes of the Parikh matrix mapping if a single permutation $a_ia_j - a_ja_i$ is made, but which may be compensated through a parallel application of multiple transformations of type (2).

Morphisms and Parikh Maatrices

A remarkable improvement seems to be the use of some morphisms which distinguish the amiable binary words by their Parikh matrices.

If Σ_1 and Σ_2 are two finite nonempty sets, a *morphism on* Σ_1 is an application

$$\phi : \Sigma_1^* \longrightarrow \Sigma_2^*$$

such that

$$\phi(uv) = \phi(u)\phi(v)$$

for all $u, v \in \Sigma_1^*$.
It is uniquely determined by its values on the alphabet Σ_1.

We shall work using the assumption:

$$(\forall x \in \Sigma_2,\ \exists a \in \Sigma_1)\ [|\phi(a)|_x > 0]$$

Theorem 4. *If* $|\Sigma_2| = 1$ *then*

$$(\forall \alpha, \beta \in \Sigma_1^*)[\alpha \sim_a \beta \Longrightarrow \phi(\alpha) \sim_a \phi(\beta)]$$

Theorem 5. *If* $|\Sigma_1| = |\Sigma_2| = 2$, *then*

$$(\forall \alpha, \beta \in \Sigma_1^*)[\alpha \sim_a \beta \quad \Longrightarrow \quad \phi(\alpha) \sim_a \phi(\beta)]$$

Theorem 6. *If* $|\Sigma_1| = 2$, $|\Sigma_2| = 3$, *then*

$$(\forall \alpha, \beta \in \Sigma_1^*)\left[\alpha \sim_a \beta \Longrightarrow \right.$$

$$\left. M_{\phi(\alpha)} - M_{\phi(\beta)} = \begin{pmatrix} 0 & 0 & 0 & r \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ r \in \mathcal{Z} \right]$$

Using the values of $r \in \mathcal{Z}$ from Theorem 6 many amiable words can be separated.

Example 4. $\Sigma_1 = \{a, b\}$, $\Sigma_2 = \{a, b, c\}$ *and the Istrail morphism defined*

$$\phi(a) = abc, \quad \phi(b) = ac.$$

Let us consider all words with $\Psi = (19, 2)$.
After applying the Istrail morphism, there are no amiable words α, β *with* $\phi(\alpha) \sim_a \phi(\beta)$.

$\lvert\alpha\rvert_{ab}$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\lvert[\alpha]\rvert$	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8
$\#_\phi$	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8
$\#_{max}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

For every value of $q = |\alpha|_{ab}$, *the second row of the table shows the number of amiable words from the set* $[\alpha]$: *the words having the Parikh matrix*

$$M = \begin{pmatrix} 1 & 19 & q \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$$

The third row gives the number of classes of amiable words in which the set $X = \{\phi(w) \mid w \in [\alpha]\}$ *is divided, and the last row of the table represents the greatest number of components from such a class.*

Example 5. *Let us take all the binary words with* $\Psi = (10, 10)$.

The separation in amiable classes is

$\lvert\alpha\rvert_{ab}$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
$\lvert[\alpha]\rvert$	1	1	2	3	5	7	11	15	22	30	42	54	73	93
$\#_{\phi}$	1	1	2	3	5	7	9	15	19	25	31	45	41	67
$\#_{max}$	1	1	1	1	1	1	2	1	2	2	3	2	5	3

$\lvert\alpha\rvert_{ab}$	14	15	16	17	18	19	20	21	22	23
$\lvert[\alpha]\rvert$	121	152	193	237	295	356	433	515	615	720
$\#_{\phi}$	65	65	79	99	93	107	107	121	125	139
$\#_{max}$	6	6	9	6	10	8	14	17	16	13

$\lvert\alpha\rvert_{ab}$	24	25	26	27	28	29	30	31	32
$\lvert[\alpha]\rvert$	847	978	1131	1289	1420	1652	1860	2065	2293
$\#_{\phi}$	135	135	147	159	155	163	157	171	171
$\#_{max}$	23	24	25	28	34	25	47	35	42

$\lvert\alpha\rvert_{ab}$	33	34	35	36	37	38	39	40	41
$\lvert[\alpha]\rvert$	2517	2761	2994	3246	3481	3729	3956	4192	4397
$\#_{\phi}$	183	181	175	187	195	189	195	187	199
$\#_{max}$	49	52	47	70	56	59	60	84	59

$\lvert\alpha\rvert_{ab}$	42	43	44	45	46	47	48	49	50
$\lvert[\alpha]\rvert$	4609	4784	4959	5095	5226	5311	5392	5424	5448
$\#_{\phi}$	197	207	203	195	205	211	203	207	197
$\#_{max}$	83	70	74	89	92	65	98	78	89

As an example, for the Parikh matrix

$$M = \begin{pmatrix} 1 & 10 & 20 \\ 0 & 1 & 10 \\ 0 & 0 & 1 \end{pmatrix}$$

$[\alpha]$ *contains* 433 *amiable words.*

After applying the Istrail morphism, the maximal set of amiable words is

$$C = \{\alpha \in \Sigma^* \mid M_\alpha = M,\ |\phi(\alpha)|_{abc} = 750\}$$

It has 14 *elements. Namely:*

$abbbbbbbabbaaaaaaaba, babbbbbabbabaaaaabaa,$
$babbbbbabaaabaabaaa, bbabbabbbbbaaaaaabaa,$
$bbabbbabbbaabaaabaaa, bbabbbbaabbbaaaabaaa,$
$bbabbbbababaababaaaa, bbabbbbbaaaabbbaaaaa,$
$bbbababbbbaaababaaaa, bbbabbababbabaabaaaa,$
$bbbabbabbaababbaaaaa, bbbbaabbabbaabbaaaaa,$
$bbbbabaabbabbabaaaaa, bbbbbaaaabbbbbaaaaaa$

Other weaknesses:

Theorem 7. *Let* $\phi : \{a,b\}^* \longrightarrow \{x,y,z\}^*$ *be a morphism. If* $\alpha, \beta \in \{a,b\}^*$ *are amiable, then*

$$\phi(\alpha\beta) \sim_a \phi(\beta\alpha)$$

Also, the reciprocals of Theorems 5 and 6 are not true. So, regarding Theorem 5:

Let us consider the morphism $\phi(a) = xy$, $\phi(b) = yx$, and the words $\alpha = ab$, $\beta = ba$.

We have $\alpha \not\sim_a \beta$ but $\phi(\alpha) \sim_a \phi(\beta)$.

The assertions from the Theorems 5, 6 are true only for $|\Sigma_1| = 2$.

Let be $\Sigma_1 = \{a_1, \ldots, a_s\}$ $(s \geq 2)$ and
$w = c_1 c_2 \ldots c_n$, $(c_i \in \Sigma_1)$ a nonempty word.

Let

$$\Psi(w) = (n_1, \ldots, n_s)$$

be the Parikh vector of w.
We define the alphabet $\Sigma_2 = \{x_1, \ldots, x_n\}$ and
a morphism $\phi : \Sigma_1^* \longrightarrow \Sigma_2^*$ as follows:
For $p = 1, \ldots, s$:

$$\phi(a_p) = x_{k_1} \ldots x_{k_{n_p}},$$

where $c_{k_i} = a_p$, $(i = 1, \ldots, n_p)$

The word w is called "**control word**".

Example 6. *If* $\Sigma_1 = \{a, b, c\}$ *and the control word* $w = abccb$.

So, $n = 5$ *and* $\Psi(w) = (1, 2, 2)$.

Accordingly with this construction, we define

$$\Sigma_2 = \{x_1, x_2, x_3, x_4, x_5\}$$

and

$$\phi(a) = x_1, \quad \phi(b) = x_2 x_5, \quad \phi(c) = x_3 x_4.$$

Theorem 8. *Using the construction above, for a word* $\alpha \in \Sigma_1^*$, *the Parikh matrix of* $\phi(\alpha)$ *is*

$$M_{\phi(\alpha)} = \begin{pmatrix} 1 & |\alpha|_{c_1} & |\alpha|_{c_1 c_2} & \dots & |\alpha|_w \\ 0 & 1 & |\alpha|_{c_2} & \dots & |\alpha|_{c_2 \dots c_n} \\ & & \vdots & & \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

A morphism based on a cyclic control word

Let $p \geq 2$ be an integer, and $w = (a_1 \dots a_s)^p$ the control word.

Therefore the morphism

$$\phi : \Sigma_1^* \longrightarrow \Sigma_2^*$$

is defined by

$$\phi(a_i) = x_i x_{s+i} x_{2s+i} \cdots x_{(p-1)s+i} \quad (1 \leq i \leq s)$$

For $\alpha \in \Sigma_1^*$ the Parikh matrix is

$$M_{\phi(\alpha)} = \begin{pmatrix} 1 & |\alpha|_{a_1} & \dots & |\alpha|_{a_1 \dots a_s} & |\alpha|_{a_1 \dots a_s a_1} & \dots & |\alpha|_{(a_1 \dots a_s)^p} \\ 0 & 1 & \dots & |\alpha|_{a_2 \dots a_s} & |\alpha|_{a_2 \dots a_s a_1} & \dots & |\alpha|_{a_2 \dots a_s (a_1 \dots a_s)^{p-1}} \\ & & & \vdots & & & \\ 0 & 0 & \dots & |\alpha|_{a_s} & |\alpha|_{a_s a_1} & \dots & |\alpha|_{a_s (a_1 \dots a_s)^{p-1}} \\ 0 & 0 & \dots & 1 & |\alpha|_{a_1} & \dots & |\alpha|_{(a_1 \dots a_s)^{p-1}} \\ & & & \vdots & & & \\ 0 & 0 & \dots & 0 & 0 & \dots & 1 \end{pmatrix}$$

All informations are provided by α and the first s rows.

So, we can keep from the Parikh Matrix $M_{\phi(\alpha)}$ only the first s rows:

$$\overline{M}_{\phi(\alpha)} = \begin{pmatrix} 1 & |\alpha|_{a_1} & \dots & |\alpha|_{a_1 \dots a_s} & |\alpha|_{a_1 \dots a_s a_1} & \dots & |\alpha|_{(a_1 \dots a_s)^p} \\ 0 & 1 & \dots & |\alpha|_{a_2 \dots a_s} & |\alpha|_{a_2 \dots a_s a_1} & \dots & |\alpha|_{a_2 \dots a_s (a_1 \dots a_s)^{p-1}} \\ & & & \vdots & & & \\ 0 & 0 & \dots & |\alpha|_{a_s} & |\alpha|_{a_s a_1} & \dots & |\alpha|_{a_s (a_1 \dots a_s)^{p-1}} \end{pmatrix}$$

Let us denote by $row_1(\alpha), \dots, row_s(\alpha)$ the rows of the matrix $\overline{M}_{\phi(\alpha)}$.

In general the value of p is the highest integer such that

$$|\alpha| = |\beta| > s^p$$

Theorem 9. *The* $s \times (sp+1)$ *matrix* $\overline{M}_{\phi(\alpha)}$ *can be recursively generated as follows:*

1. $\overline{M}_{\phi(\epsilon)} = \begin{pmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ & & & \vdots & & \\ 0 & 0 & \dots & 1 & \dots & 0 \end{pmatrix}$

2. $\overline{M}_{a_i\alpha}$ *can be obtained from* $\overline{M}_{\alpha}$ *as follows:*

   (a) $row_j(a_i\alpha) \longleftarrow row_j(\alpha)$
   for $j \in \{1, 2, \dots, s\} \setminus \{i\}$;

   (b) $row_i(a_i\alpha) \longleftarrow row_i(\alpha) + row_{i+1}(\alpha)$
   if $1 \leq i < s$,
   $row_s(a_s\alpha) \longleftarrow row_s(\alpha) + Shift_s(row_1(\alpha))$.

Example 7. *Let be* $s = 2, p = 3$. *Then*

$$\overline{M}_{\phi(\alpha)} = \begin{pmatrix} 1 & |\alpha|_a & |\alpha|_{ab} & |\alpha|_{aba} & |\alpha|_{abab} & |\alpha|_{ababa} & |\alpha|_{ababab} \\ 0 & 1 & |\alpha|_b & |\alpha|_{ba} & |\alpha|_{bab} & |\alpha|_{baba} & |\alpha|_{babab} \end{pmatrix}$$

If we take the binary amiable words

$$abbabaab \text{ and } baababba$$

we have

$$\overline{M}_{\phi(abbabaab)} = \begin{pmatrix} 1 & 4 & 8 & 10 & \mathbf{12} & 4 & \mathbf{4} \\ 0 & 1 & 4 & 8 & 10 & \mathbf{4} & 4 \end{pmatrix}$$

and

$$\overline{M}_{\phi(baababba)} = \begin{pmatrix} 1 & 4 & 8 & 10 & \mathbf{4} & 4 & \mathbf{0} \\ 0 & 1 & 4 & 8 & 10 & \mathbf{12} & 4 \end{pmatrix}$$

Therefore these words are separated by 3 elements (from 14*):* $\#abab$, $\#baba$, $\#ababab$.

Unfortunately, this cyclic morphism does not solve entirely the separation of any amiable words.

Example 8. *The words* $\alpha = abccb$ *and* $\beta = acbbc$ *have the same Parikh matrix:*

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

Their images $\phi(\alpha)$ *and* $\phi(\beta)$ *constructed with* $p = 2$ *(but any value* $p > 2$ *gives the same result) have also the same Parikh matrix:*

$$M_{\phi(\alpha)} = \begin{pmatrix} 1 & 1 & 2 & 2 & 0 & 0 & 0 \\ 0 & 1 & 2 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Theorem 10. *Let* $\Sigma_1 = \{a_1, \ldots, a_s\}$ *be an ordered alphabet, and* $\alpha, \beta \in \Sigma_1^*$ *be two amiable words.*
Then there is an ordered alphabet Σ_2 *and a morphism* $\phi : \Sigma_1^* \longrightarrow \Sigma_2^*$ *such that*

$$\phi(\alpha) \not\sim_a \phi(\beta)$$

The construction is based on the control word $w = \alpha$.

Using Theorem 8, the last entry of the first row in a Parikh matrix for an arbitrary word $\gamma \in \Sigma_1^*$ is $|\gamma|_\alpha$.

Now, because $|\alpha| = |\beta| = |w|$ we have $|\alpha|_\alpha = 1$ and $|\beta|_\alpha = 1$ if and only if $\alpha = \beta$.

Example 9. *The words* $\alpha = acbbc$ *and* $\beta = abccb$ *from Example 8 can be separated using the construction from Theorem 10.*
For the control word $w = \alpha = acbbc$*:*

$$M_{\phi(acbbc)} = \begin{pmatrix} 1 & 1 & 2 & 2 & 1 & 1 \\ 0 & 1 & 2 & 2 & 1 & 1 \\ 0 & 0 & 1 & 2 & 1 & 1 \\ 0 & 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$M_{\phi(abccb)} = \begin{pmatrix} 1 & 1 & 2 & 2 & 0 & 0 \\ 0 & 1 & 2 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Parikh matrix mappings of languages

Let us consider an alphabet Σ and a language $L \subseteq \Sigma^*$, such that

$$\forall a \in \Sigma,\ \exists w \in L,\ \ |w|_a \neq 0$$

Theorem 11. *Let $\Sigma = \{a, b\}$ be an alphabet and $M \in \mathcal{M}_3$ a fixed Parikh matrix mapping, where $M = (m_{ij})_{1 \leq i,j \leq 3}$. Then the language*

$$L = \{\alpha \mid \alpha \in \Sigma^*,\ |\alpha|_{ab} = m_{13}\}$$

is regular.

Theorem 12. *Let $\Sigma = \{a_1, a_2, \ldots, a_s\}$ an ordered alphabet and $M \in \mathcal{M}_{s+1}$ a fixed Parikh matrix mapping. Then the following language is regular:*

$$L = \{\alpha \mid \alpha \in \Sigma^*, M_\alpha \text{ and } M$$

$$\text{are identical on the third diagonal}\}$$

Theorem 13. *Let $L \in \mathcal{L}_1$ over the alphabet $\Sigma = \{a_1, a_2, \ldots, a_s\}$ and $M \in \mathcal{M}_{s+1}$ a Parikh matrix mapping. It is decidable whether exists $w \in L$ having $M_w = M$.*

1. From M we extract the Parikh vector

$$\Psi = (x_1, x_2, \ldots, x_s)$$

2. Let $n = \sum_{i=1}^{s} x_i, \quad V_0 = \{S\}, \quad i := 1.$
3. $V_i = V_{i-1} \cup \{\alpha \mid \exists \beta \in V_{i-1},\ \beta \overset{G}{\Longrightarrow} \alpha,$
   $|\alpha| \leq n,\ |\alpha|_{a_i} \leq x_i,\ 1 \leq i \leq s\}.$
4. If $V_i \neq V_{i-1}$ then $i := i + 1$ and goto 3.
5. $W = V_i \cap \Sigma^*$.
6. If $W = \emptyset$ then NO, Stop.
7. Let $x \in W$. If $M_x = M$ then YES, Stop.
8. $W := W \setminus \{x\}$ and goto 6.

Amiable extensions of Chomsky languages

Let $L \subseteq \Sigma^*$. We will extend this language by adding all the words that are amiable with the elements of L. Formally,

$$[L] = \{w \mid \exists \alpha \in L, \;\; M_\alpha = M_w\}.$$

or

$$[L] = \bigcup_{\alpha \in L} [\alpha]$$

$[L]$ is the **amiable extension** of L.

Properties:

1. $[[L]] = [L]$;

2. $[L_1 \cup L_2] = [L_1] \cup [L_2]$;

3. $[L_1][L_2] \subseteq [L_1 L_2]$;

4. $L_1 \subseteq L_2 \quad \Longrightarrow \quad [L_1] \subseteq [L_2]$,

5. $[L]^* \subseteq [L^*]$.

6. $[L_1 \cap L_2] \subseteq [L_1] \cap [L_2]$.

7. $L \subset [L] \Longleftrightarrow [L] \cap [\overline{L}] \neq \emptyset$, where $\overline{L} = \Sigma^* \setminus L$.

8. $L = [L] \quad \Longleftrightarrow \quad \overline{L} = [\overline{L}]$.

Proposition 1. $L_\Sigma = \{[L] | L \subseteq \Sigma^*\}$ *forms a Boolean algebra (with the* $\bigcup$ *and* $\bigcap$ *operators).*

Proposition 2. *The set of languages characterized by the same set of Parikh matrices (given by a language L) is a join-semilattice.*

Theorem 14. *Let* $\Sigma = \{a, b\}$ *be an alphabet and* $L \subseteq \Sigma^*$ *a language.*

1. *If* $L \in \mathcal{L}_i$*, with* $1 \leq i \leq 3$*, then* $[L] \in \mathcal{L}_1$*.*

2. $\mathcal{L}_0$ *is closed under the* [ ] *extension.*

Definition 4. *Let n be a non-negative integer, and Σ a fixed alphabet.*
The following class of languages is defined $\mathcal{APL}_n(\Sigma)$ (Amiable Parikh Languages of n order over the alphabet Σ) in the following way:

A language $L \in \mathcal{APL}_n(\Sigma)$ if

1. *All the words of $L \subseteq \Sigma^*$ are amiable (they have the same Parikh matrix M);*

2. *$\exists w_1, w_2 \in L$ such that $w_1 \overset{(2.n)}{\longleftrightarrow} w_2$ where the transformation $(2.n)$ is irreducible.*

- $\mathcal{APL}_0(\Sigma)$ contains languages whose words are amiable using only transformations of type (1).
  For instance, if $\Sigma = \{a, b, c\}$, the languages $L = [\alpha]$, $(\alpha \in \{a, c\}^*)$ are in $\mathcal{APL}_0(\Sigma)$.

- If Σ is a binary alphabet, then all the languages $L = [\alpha]$ $(\alpha \in \Sigma^*)$ are in $\mathcal{APL}_1(\Sigma)$.

Theorem 15. *If Σ is an alphabet with $|\Sigma| > 2$, then the hierarchy $\mathcal{APL}_n(\Sigma)$ is infinite.*

Consider $|\Sigma| = 3$.
Let be the alphabet $\Sigma = \{a < b < c\}$ and n a fixed arbitrary non-negative integer.

$$w = (abcba)^{n-1}bac^{n-1}ab$$

$$w' = (bacab)^{n-1}abc^{n-1}ba$$

Statements:

1. $M_w = M_{w'} = \begin{pmatrix} 1 & 2n & 2n^2 & (n-1)(2n^2-3n+2) \\ 0 & 1 & 2n & (n-1)(3n-2) \\ 0 & 0 & 1 & 2n-2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

2. The transformation $w \overset{(2.n)}{\longleftrightarrow} w'$ is irreducible of order n.

The language

$$L = \{\alpha \mid \alpha \in \Sigma^*,\ M_\alpha = M_w\} \in \mathcal{APL}_n(\Sigma)$$

($w, w' \in L$ and between them there exists a (2.n) irreducible transformation).

If the non-binary alphabet Σ is fixed, we can use the notation $\mathcal{APL}_n$ instead of $\mathcal{APL}_n(\Sigma)$.

Proposition 3. *Let M be a Parikh matrix and $L \in \mathcal{APL}_1$ a language defined by M.*
Then there exists an unique integer n, and a sequence of languages $L_1, \ldots, L_n$ such that

1. $L = L_1 \subset L_2 \subset \ldots \subset L_n$,
   $L_i \in \mathcal{APL}_i$, $(1 \leq i \leq n)$.

2. $\forall w \in \Sigma^* \, [w \in L_n \quad \Longleftrightarrow \quad M_w = M]$.

In other words, $L_n = [L_n]$.

The following statements are obvious:

- $L_1, L_2 \in \mathcal{APL}_n \quad \Longrightarrow \quad L_1 \cup L_2 \in \mathcal{APL}_n$.

- $L_1, L_2 \in \mathcal{APL}_n,\ L_1 \cap L_2 \neq \emptyset \Longrightarrow L_1 = L_2$.

Proposition 4. *If* $L_1, L_2 \in \mathcal{APL}_n,\ L_1 \cap L_2 = \emptyset$ *and both languages are defined by the same Parikh matrix mapping* M*, then*

1. $L_i \subset [L_i],\ i = 1, 2$;

2. $\exists L_3 \in \mathcal{APL}_{n+1}$ *defined by the matrix* M*, where* $L_1 \cup L_2 \subseteq L_3$.