INFORMATIQUE THÉORIQUE ET APPLICATIONS

C. CHOFFRUT

Bijective sequential mappings of a free monoid onto another

Informatique théorique et applications, tome 28, nº 3-4 (1994), p. 265-276.

<http://www.numdam.org/item?id=ITA_1994__28_3-4_265_0>

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 28, n° 3-4, 1994, p. 265 à 276)

BIJECTIVE SEQUENTIAL MAPPINGS OF A FREE MONOID ONTO ANOTHER

by C. CHOFFRUT

Abstract. – *Given two free monoids Σ^* and Δ^* we show that there exists a bijective sequential mapping of Σ^* onto Δ^* if and only if the cardinality of Σ is greater or equal to that of Δ.*

1. INTRODUCTION

Given two finite alphabets Σ, Δ and Σ^*, Δ^* the free monoids they generate, it is well known that there exists a bijective morphism of Σ^* onto Δ^* if and only if Σ and Δ have the same cardinality. Here, we consider the more general class of sequential mappings of Σ^* into Δ^*, *i. e.*, of g-s-m mappings in the sense of Ginsburg and Rose [5] or generalized sequential total functions in the sense of Eilenberg [4] Chapter IX, p. 296. Such functions are defined by finite deterministic automata provided with an output function: given a input word $x \in \Sigma^*$, its image is the output word $y \in \Delta^*$ that is obtained by concatenating the outputs along the path labeled by x in the automaton. We address the question of determining conditions on the cardinalities of Σ and Δ under which there exists a bijective sequential mapping of Σ^* onto Δ^*. The prefix-preservation of such functions is a strong property since it leaves very little leeway of how the images are produced. It is not even clear whether there exists a solution at all, but surprisingly enough an example is mentioned in [4], p. 305, in the special case where $\text{Card}(\Sigma) = 3$ and $\text{Card}(\Delta) = 2$. We prove that this can be generalized to arbitrary free monoids with the necessary and sufficient condition that $\text{Card}(\Sigma) = \text{Card}(\Delta)$ or $1 < \text{Card}(\Delta) < \text{Card}(\Sigma)$ holds.

The problem is concerned with properties of sequential functions for their own sake. Though sequential mappings were introduced some thirty

[1] Université Paris-VII, LITP, Tour 55-56, 1^er^ étage, 75251 Paris Cedex.

0988-3754/94/03-04/$ 4.00/© AFCET-Gauthier-Villars

years ago, they were somehow neglected when the more general notion of rational transduction was discovered and applied as a means for investigating formal languages. The concept of rational transduction had the great merit of perfectly suiting the nature of contex-free languages and thus permitting a tremendous development of the theory of the so-called abstract families of languages, *cf.* [1]. Nevertheless, some authors have used the more restrictive notion of sequential functions with the same purpose. More precisely, using slightly different definitions of sequential mappings, algorithms for determinng under which conditions there exists a sequential function mapping a given rational language onto another have been devised, *cf.*, [7] and [8]. In [6] the same problem is considered where bijective rational transductions are used in place of sequential functions.

2. PRELIMINARIES

2.1. Free monoids - Rational ℕ-subsets - Transducers

Given a set Σ (an *alphabet*) consisting of *letters*, Σ^* denotes the free monoid it generates. The elements of Σ^* are *words*. The *length* of a word w is denoted by $|w|$. The *empty* word, denoted by 1, is the word of length 0. We refer to [5] Chapters VI and VII, for all standard definitions on *ℕ-subsets* and more specifically on *rational* ℕ-subsets. Intuitively, ℕ-subsets are ordinary subsets with multiplicity, *i. e.* every word is associated with an integer recording not only whether it belongs to the subset but also how many times it does so. We recall that given two ℕ-subsets X, Y, their *sum* is denoted by $X+Y$, their *contacenation* or *product* is denoted by XY, and the star of X is denoted by X^*.These are the usual *rational operations* in the *ℕ-algebra* of the rational ℕ-subsets. In particular a ℕ-subset is *unambiguous* if the multiplicity associated with an arbitrary word is either 0 or 1. More generally, a rational ℕ-subset is recognized by a (non-necessarily deterministic) finite automaton for which every word labels at most one successful path. The rational ℕ-subsets are also known as *rational power series* over the integers in the non commuting indeterminates Σ where the multiplicity is known as the *coefficient* of a word (*cf.* [3]). The *support* of a ℕ-subset L is the (ordinary) subset of all the words in L whose multiplicity is different from 0. When the support is finite we say L is a *polynomial* and when it is reduced to one element we say L is a *monomial.* E. g., if $L = 3x + y + 5xy$ then the support is the subset $\{x, y, xy\}$.

Sequential functions are mappings of a free monoid into another defined by finite sequential transducers. Given an *input* and an *output* alphabets Σ and

Δ, a *finite sequential transducer* (abbreviated transducer) $\mathcal{T}$ is a complete finite deterministic automaton whose states are all accepting, equipped with an *output function* from the set of transitions into Δ^*. Formally, we have $\mathcal{T} = (Q, i, \lambda)$ where Q is the set of states and i the initial state; the next state function of the underlying automaton associates with all pairs (q, a) a next state $q \cdot a$ and the output function λ associates with all pairs $(q, a) \in Q \times \Sigma$ a word in $\lambda(q, a) \in \Delta^*$. These two functions are extended to Σ^* by induction on the length of the words:

- for all $q \in Q$, $q \cdot 1 = q$
- for all $q \in Q$, $a \in \Sigma$, $u \in \Sigma^*$, $q \cdot ua = (q \cdot u) \cdot a$
- for all $q \in Q$, $\lambda(q, 1) = 1$
- for all $q \in Q$, $a \in \Sigma$, $u \in \Sigma^*$, $\lambda(q, ua) = \lambda(q, u)\, \lambda(q \cdot u, a)$

The cardinality of Q is the *dimension* of the transducer. A function $f : \Sigma^* \rightarrow \Delta^*$ is *sequential* if there exists a sequential transducer $\mathcal{T}$ such that $f(u) = \lambda(i, u)$ holds for all $u \in \Sigma^*$ (observe that all states are final). We say the function f is *realized* by the transducer $\mathcal{T}$.

EXAMPLE 2.1: Consider the following sequential transducer where $\Sigma = \{a, b, c\}$ and $\Delta = \{x, y\}$. The initial state is indicated as usual by an entering arrow (it is the state labeled by 0). Then, e. g.,

$$\lambda(1, cc) = (yxy)(yy) = yxyyy.$$

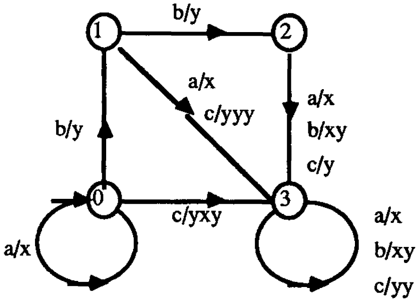


Figure 1. – A sequential function of dimension 4.

The question we ask is the following:

Given Σ and Δ does there exist a bijective sequential function of Σ^* onto Δ^*?

Clearly, we must have

$$\text{Card}(\Delta) = \text{Card}(\Sigma) \quad \text{or} \quad 1 < \text{Card}(\Delta) < \text{Card}(\Sigma).$$

Indeed, assume $\text{Card}(\Delta) > \text{Card}(\Sigma)$ holds and let Δ' be the proper subalphabet of Δ consisting of all the letters appearing as first occurrences of the words $\lambda(i, a)$ where $a \in \Sigma$. Then all words in $f(\Sigma^*)$ start with a letter in Δ' and the function can not be onto. Moreover, the case $\text{Card}(\Sigma) = \text{Card}(\Delta)$ is easily treated:

PROPOSITION 2.1: *Let $\mathcal{T}$ be a sequential transducer realizing a sequential function of Σ^* onto itself. Then it is a bijection if and only if for each state $q \in Q$, $a \to \lambda(q, a)$ is a bijection of the alphabet Σ.*

The proof of this assertion is a direct consequence of the following observation:

LEMMA 2.2: *Let f be a prefix-preserving function of Σ^* onto itself. Then it is bijective if and only if for all integers $n > 0$ it defines a bijective mapping on the set of words of length n.*

Proof: Indeed, assume by contradiction, there exists a maximum integer n such that f defines a bijection of Σ_n onto itself, where Σ_n denotes the set of all words of length n. It suffices to prove that $f^{-1}(\Sigma_{n+1}) \subseteq \Sigma_{n+1}$. Consider $ub \in \Sigma_{n+1}$, with $b \in \Sigma$, *i. e.*, $|u| = n$. Then there exists $x \in \Sigma^*$ and $a \in \Sigma$ such that $f(xa) = ub$. By hypothesis, we have $|u| \geqq n$. Since $f(x)$ is a proper prefix of ub thus a prefix of u, by induction hypothesis we have $|x| \leqq n$, completing the verification. ■

Now assume $\text{Card}(\Delta) < \text{Card}(\Sigma)$ holds and for all integers $k > 0$ set:

$$\alpha(k) = \max\{|f(u)|\,|u| \leqq k\}$$

Since f is sequential, $\alpha(k) < ak$ for some constant a. If f is bijective then it maps the set of all words of length less than or equal to k in Σ^* into the set of all words of length less than or equal to ak in Δ^*. The former set has cardinality $k^{\text{Card}(\Sigma)}$ and the latter $(ak)^{\text{Card}(\Delta)}$. This implies $\text{Card}(\Delta) > 1$. As a result, in the sequel we make the following assumption:

$$1 < \text{Card}(\Delta) < \text{Card}(\Sigma)$$

2.2. Normal forms

It is desirable to recognize the solutions that are artificially different. Distinct transducers may define the same sequential function but given a sequential function it can be realized by a so-called *minimal* transducer that is unique to within a renaming of the states (cf., *e. g.*, [5], Chapter XII. 3). This automaton is determined by the condition:

for all $q, q' \in Q$ the functions $u \rightarrow \lambda(q, u)$ and $u \rightarrow \lambda(q', u)$ are equal if and only if $q = q'$.

There is a further observation that classifies the possible bijections $f : \Sigma^* \rightarrow \Delta^*$. Consider two sequential bijections $\phi : \Sigma^* \rightarrow \Sigma^*$ and $\theta : \Delta^* \rightarrow \Delta^*$. Then the composition $\theta \circ f \circ \phi$ is again a sequential bijection from Σ^* to Δ^*. It is reasonable to consider the two sequential functions f and g as basically identical and we will say that they are *equivalent.*

2.3. Matrix interpretation

Actually, for our purpose there is no need of specifying the input alphabet which in a sense is immaterial. E. g., by ignoring the input letters, the previous transducer can be represented as a finite regular graph of degree equal to the cardinality of Σ and labelled by words in Δ^*:

Conversely, such a graph is solution of our problem if and only if the ℕ-subset of the labels of all paths starting in the initial mode i is equal to Δ^*. This again is computed as follows. For each node $q \in Q$, denote by L_q the ℕ-subset of all the words that label some path starting in q (each word occurring as many times as the number of paths it labels). Provided no image of the output function λ is equal to the empty word, L_i is the unique solution of the following system of linear equations where the unknowns are the L_q's and $q \in Q$ (*cf.* [5], Propositions VII.6.2)

$$L_q = 1 + \sum_{q \in Q} \gamma_{qp} L_p \tag{1}$$

with the notation that for all $p, q \in Q$, γ_{qp} is the ℕ-subset of the labels of an edge from q to p.

Example 2.2: (continued):

$$\begin{aligned} L_0 &= 1 + x L_0 + y L_1 + yxy L_3 \\ L_1 &= 1 + y L_2 + (x + yyy) L_3 \\ L_2 &= 1 + (x + xy + y) L_3 \\ L_3 &= 1 + (x + xy + yy) L_3 \qquad \blacksquare \end{aligned}$$

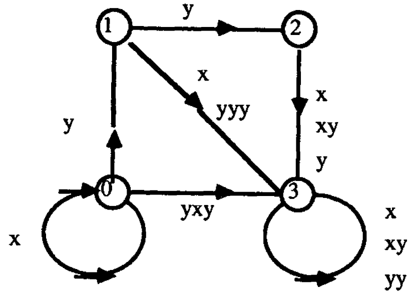


Figure 2. – **Labelled graph associated with the sequential function of Figure 1.**

Equivalently, set $n = \operatorname{Card} Q$, and consider the n-column vector $L = (L_q)_{q \in Q}$ and the $n \times n$-matrix $M = (\gamma_{qp})_{qp}$. Denote by E the n-column vector whose entries are all equal to 1. Then the system (1) has a unique solution which is the vector $M^* \, E$, where $M^* = I + M \cdots + M^k + \cdots$ (*cf.*, [5], Proposition VII, 6.2). Determining all possible bijective sequential mappings of the free monoid Σ^* onto the free monoid Δ^* reduces to determining all possible square matrices M with entries in the polynomials in the indeterminates Δ subject to the conditions

(2) the sum of all the entries in each row of M is an ambiguous polynomial with $\operatorname{Card}(\Sigma)$ monomials

(3) the sum of all the entries in the first row of the matrix M^* is the $\mathbb{N}$-rational subset Δ^*

Example 2.3: (continued):

$$M = \begin{pmatrix} x & y & 0 & yxy \\ 0 & 0 & y & x + yyy \\ 0 & 0 & 0 & x + xy + y \\ 0 & 0 & 0 & x + xy + yy \end{pmatrix}$$

The sums of the different rows are the polynomials $x + y + yxy$, $y + x + yyy$, $x + xy + y$, and $x + xy + yy$ respectively and the sum of the entries of the first row in M^* equals

$$x^* \left(yx + yxy + y^2 + y^3 + y^2\, xy + y^4\right) (x + xy + yy)^* \qquad \blacksquare$$

3. SOME EXAMPLES

In the previous paragraph we saw how the present problem could be regarded as equating the free monoid Δ^* with an unambigous rational subset subject to some constraints related to the structure of the underlying deterministic automaton. A natural source of inspiration is the case of groups where each group can be expressed as a disjoint union of cosets. For free monoids this situation is most closely reproduced by using maximal suffix codes, *cf.* [2], p. 98. Indeed, if X is such a code and S is the set of all the non empty words that are proper suffixes of X, then we have the following equality of $\mathbb{N}$-rational subsets:

$$\Delta^* = (1 + S) X^* \tag{4}$$

To fix ideas assume Σ has 3 elements and $\Delta = \{x, y\}$. Then the only maximal suffix code of cardinality 3 (up to a renaming of the letters) is $X = \{x, xy, yy\}$ and the set of its non empty proper suffixes is reduced to $\{y\}$. Using the previous equality (4) and the ordinary axioms of $\mathbb{N}$-algebras (*cf.* [5] Chapter VI, 3 or [4] Chapters 3, 4, 12 and 13), we have the following equalities:

$$\begin{aligned}\Delta^* &= (1 + y) X^* = 1 + (X = y) X^* = 1 + (x + xy + yy + y) X^* \\ &= 1 + x(1 + y) X^* + (y + yy) X^* \\ &= 1 + x\,\Delta^* + (y + yy) X^*\end{aligned}$$

which leads to the following bijective sequential mapping:

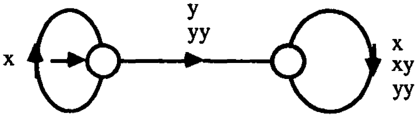


Figure 3. – A sequential bijection of dimension 2.

We can build on this example by modifying the labelled graph, *i. e.*, by creating new nodes or/and rearranging existing nodes without changing the $\mathbb{N}$-rational subset. In order to formalize this notion of rearrangement we need a definition that helps describe the structure of the set of labels associated with the paths of a uniform tree such as

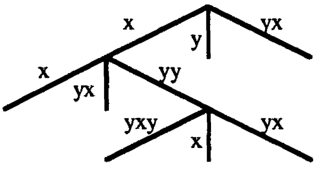


Figure 4. – Labelled complete tree of degree 3.

Then, *e. g.*, the polynomial whose monomials are all the possible labels of a path starting at the root in the above tree can be factored as:

$$x\,(1 + (x + yx + yy\,(1 + yxy + x + yx))) + y + yx$$

A *prefix-factorization* of an unambiguous $\mathbb{N}$-polynomial p in the indeterminates Δ, when it exists, is inductively defined by:

- $p = 0$ (the zero polynomial)
- otherwise $p = w_1\,(1 + p_1) + w_2\,(1 + p_2) + w_3\,(1 + p_3)$ where for $i = 1, 2, 3$ $w_i \neq 1$ is a monomial and p_i is a prefix-factorization.

Clearly, the following labelled graph obtained by creating two new states defines the same bijection as Figure 3:

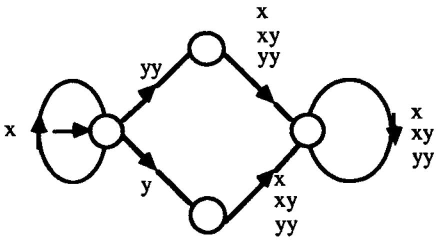
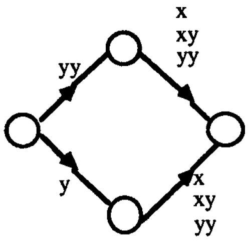


Figure 5. – Labelled graph defining the same sequential bijection as Figure 3.

In order to obtain new bijections it suffices to rearrange the transitions:



i. e., since the initial state has already a loop, it suffices to factorize in all possible ways the polynomial $Z = (y + yy)\,(1 + x + xy + yy)$ into $w_1\,(1 + p_1) + w_2\,(1 + p_2)$ where for $i = 1, 2$, $w_i \neq 1$ is a monomial and p_i a prefix-factorization. Here are the solutions:

$$y\,(1 + y\,(1 + x + xy + yyy) + xy + yy) + yx \tag{5}$$

$$y\,(1 + y\,(1 + x + xy + yy) + x + yy) + yxy \tag{6}$$

$$y(1+y(1+x+xy+y)+x+yyy)+yxy \tag{7}$$

$$y(1+y(1+x+y+yy)+x+yxy)+yxy \tag{8}$$

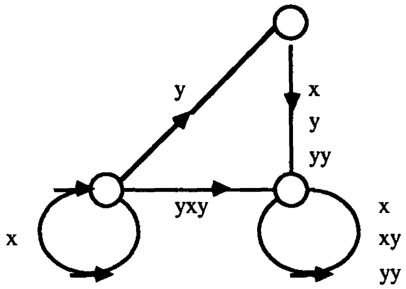
$$y(1+y(1+xy+y+yy)+x+yx)+yxy \tag{9}$$

$$y(1+y(1+xy+y+yy)+x+xy)+yyx \tag{10}$$

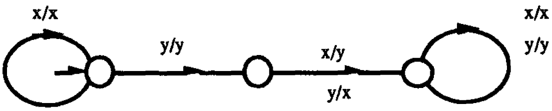
$$y(1+y(1+x+xy+y)+x+xy)+yyyy \tag{11}$$

$$y(1+y(1+x+xy+yy)+x+xy)+yyy \tag{12}$$

Expression (6) yields the following solution of dimension 3 (since two nodes can be merged)



The solution g corresponding to expression (12) is equivalent to the previous solution f under the sequential bijection $\theta : \Delta^* \longrightarrow \Delta^*$ whose transducer is:



Factorizations (7) and (8) also yield two sequential bijections that are equivalent.

4. THE GENERAL CASE

We are now in a position to prove the main result.

THEOREM 4.1: *Let Σ and Δ be two alphabets of cardinality m and n respectively. Then there exists a bijective sequential mapping of Σ^* onto Δ^**

if and only if $m > n > 1$ or $m = n$. Furthermore, in this case there exists a mapping of dimension 2.

Proof: Because of the observations of section 2.1, the condition is clearly necessary. Conversely, because of Proposition 2.1 it suffices to consider the case $m > n > 1$.

Assume first that $k(n-1) = m-1$ holds for some integer $k > 1$. Let X be a complexe suffix of Δ^* of cardinality m and let S be the set of its non empty proper suffixes. We have $\mathrm{Card}(S) = k$ and furthermore:

$$\Delta^* = (1+S)\,X^* \tag{13}$$

i. e., $(1-\Delta)^{-1} = (1+S)(1-X)^{-1}$. Multiplying by $1-\Delta$ to the left and by $1-X$ to the right we get: $\Delta(1+S) = S+X$. If b is an arbitrary letter of Δ we have:

$$S+X = (\Delta - b)(1+S) + b(1+S)$$

which shows that $S+X-b(1+S)$ is an unambiguous polynomial whose support has $m+k-(k+1) = m-1$ elements. Starting from (13) we get $\Delta^* = (1+S)X^* = 1+(S+X)X^*$ and further:

$$\Delta^* \;=\; 1 + (S+X-b(1+S)+b(1+S))\,X^*.$$

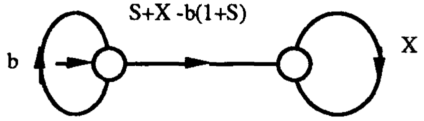
This yields:

$$\begin{aligned}\Delta^* &= 1 + b(1+S)X^* + (S+X-b(1+S))\,X^*\\ &= 1 + b\,\Delta^* + (S+X-b(1+S))\,X^*\end{aligned}$$

which shows that the pair $\begin{pmatrix}\Delta^*\\ X^*\end{pmatrix}$ is the unique solution of the linear system:

$$\begin{aligned}L_0 &= 1 + b\,L_0 + (S+X-b(1+S))\,L_1\\ L_1 &= 1 + XL_1\end{aligned}$$

yielding the following sequential bijection:



In terms of matrices as interpreted in paragraph 2.3, this means that Δ^* is the first component of the solution of the vector equation $L = I + ME$, where E is 2-column vector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and M is the matrix

$$M = \begin{pmatrix} b & S + X - b(1+S) \\ 0 & X \end{pmatrix}$$

Let us now turn to the general case. There always exists an integer $1 < p < n$ such that $n-p-1$ divides $m-p-1$: $m-p-1 = k(n-p-1)$ for some $k > 1$. Consider two partitions $\Sigma = \Sigma_1 \cup \Sigma_2$ and $\Delta = \Delta_1 \cup \Delta_2$ where $\mathrm{Card}(\Sigma_1) = m - p$, $\mathrm{Card}(\Delta_1) = n - p$, and therefore $\mathrm{Card}(\Sigma_2) = \mathrm{Card}(\Delta_2) = p$. Because of the previous discussion there exists a bijective sequential mapping of the free monoid Σ_1^* onto the free monoid Δ_1^*. Let M_1 be the corresponding 2×2-matrix with entries in Δ_1^* and satisfying the conditions (2) and (3) of paragraph 2.3. Now consider the matrix $M = M_1 + M_2$ where

$$M_2 = \begin{pmatrix} \Delta_2 & 0 \\ \Delta_2 & 0 \end{pmatrix}$$

Clearly the sum of all the entries in each row of the matrix M is an unambiguous polynomial with $\mathrm{Card}(\Sigma)$ monomials. Let us verify that the sum of all the entries in the first row of the matrix M^* is the $\mathbb{N}$-rational subset Δ^*. First we set

$$M_1 = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} \quad \text{and} \quad M_1^* = \begin{pmatrix} F & G \\ 0 & H \end{pmatrix}$$

where $\Delta_1^* = A^* + CB^* = F + G$. Using the identity $M^* = (M_1 + M_2)^* = (M_1^* M_2^*)^* M_1^*$ we obtain

$$M_1^* M_2 = \begin{pmatrix} F & G \\ 0 & H \end{pmatrix} \begin{pmatrix} \Delta_2 & 0 \\ \Delta_2 & 0 \end{pmatrix} = \begin{pmatrix} \Delta_1^* \Delta_2 & 0 \\ H \Delta_2 & 0 \end{pmatrix}$$

Thus

$$(M_1^* M_2)^* M_1^* = \begin{pmatrix} (\Delta_1^* \Delta_2)^* & 0 \\ (\Delta_1^* \Delta_2)^* H \Delta_2 & 0 \end{pmatrix} \begin{pmatrix} F & G \\ 0 & H \end{pmatrix}$$

The first row of this last matrix is equal to

$$(\Delta_1^* \Delta_2)^* (F + G) = (\Delta_1^* \Delta_2)^* \Delta_1^* = (\Delta_1 + \Delta_2)^* = \Delta^* \qquad \blacksquare$$

REFERENCES

1. J. Berstel and L. Boasson, *Context-Free Languages*, Elsevier, 1990, chapter 2, pp. 61-100.

2. J. Berstel and D. Perrin, *Theory of Codes*, Academic Press, 1985.

3. J. Berstel and C. Reutenauer, *Rational Series and Their Languages*, volume 12 of *EATCS Monograph on Theoretical Computer Science*, Academic Press, 1988.

4. J. H. Conway, *Regular Algebra and Finite Machines*, Chapman and Hall, London, 1971.

5. S. Eilenberg, *Automata, Languages and Machines*, volume A. Academic Press, 1974.

6. S. Ginsburg and G. F. Rose, A characterization of machine mappings, *Can. J. Math.*, 1966, *18*, pp. 381-388.

7. H. A. Maurer and M. Nivat, Rational bijections of rational sets, *Acta Informatica*, 1980, *13*, pp. 365-378.

8. R. Mac Naughton, *A decision procedure for generalized mappability-onto of regular sets*, manuscript.

9. K. B. Samolon, The decidability of a mapping problem for generalized sequential machines with final states, *J. of Comput. and Sys. Sci.*, 1975, *10*, 2, pp. 200-218.