WORDS, LANGUAGES AND COMBINATORICS

Kyoto, Japan **28 – 31 August 1990**

Editor

Masami Ito
Faculty of Science
Kyoto Sangyo University
Japan

BIBLIOTHÈQUE INTERUNIVERSITAIRE
INFORMATIQUE
RECHERCHE

World Scientific
Singapore • New Jersey • London • Hong Kong

RATIONAL WORD FUNCTIONS: CHARACTERIZATION AND MINIMIZATION

CHRISTOPHE REUTENAUER
Mathématiques et Informatique
Université du Québec à Montréal
Montréal, Canada H3C 3P8

and

MARCEL-PAUL SCHÜTZENBERGER
Académie des Sciences
Paris, France

ABSTRACT

A function from a free monoid is rational if its graph is a rational subset of the product monoid. We show that a function is rational if and only some congruence, canonically attached to it, is of finite index, and if the function preserves rationality of languages under inverse image. We construct a canonical bimachine computing a given rational function. This bimachine is minimal, in some precise sense, when the function is total.

1.Rational languages

It is well-known that a language $L \subseteq A^*$ is rational if it satisfies the Nerode criterion. This characterization may be equivalently expressed using the Hankel matrix of the language. This matrix is an A^* by A^* matrix, with a 1 or a 0 in the (u, v) - entry, depending whether uv is in L or not. For example, for $L = \{a, ba, bb\}^*$, the Hankel matrix is the following

	1	a	b	aa	ab	ba	bb	aaa	aab	aba	abb	baa	bab	bba	bbb ...
1	1	1	0	1	0	1	1	1	0	1	1	1	0	1	0
a	1	1	0	1	0	1	1	1	0	1	1	1	0	1	0
b	0	1	1	1	0	1	0	1	0	1	1	1	0	1	1
aa	1	1	0	1	0	1	1	1	0	1	1	1	0	1	0
ab	0	1	1	1	0	1	0	1	0	1	1	1	0	1	1
ba	1	1	0	1	0	1	1	1	0	1	1	1	0	1	0
bb	1	1	0	1	0	1	1	1	0	1	1	1	0	1	0
aaa	1	1	0	1	0	1	1	1	0	1	1	1	0	1	0
aab	0	1	1	1	0	1	0	1	0	1	1	1	0	1	1
aba	1	1	0	1	0	1	1	1	0	1	1	1	0	1	0
abb	1	1	0	1	0	1	1	1	0	1	1	1	0	1	0
baa	1	1	0	1	0	1	1	1	0	1	1	1	0	1	0
bab	0	1	1	1	0	1	0	1	0	1	1	1	0	1	1
bba	1	1	0	1	0	1	1	1	0	1	1	1	0	1	0
bbb	0	1	1	1	0	1	0	1	0	1	1	1	0	1	1

A language is rational if and only if its Hankel matrix has only finitely many distinct lines; this also holds with columns. In the example, there are 2 different rows, and 4 different columns.

The terminology "Hankel matrix" stems from classical analysis: it is well-known that a series

$$\sum_{n\geq 0} a_n x^n$$

whith coefficients in a field is rational (i.e. quotient of two polynomials) if and only if its Hankel matrix

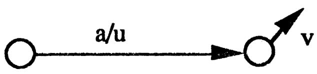
$$(a_{i+j})_{i,j\geq o}$$

has finite rank; this rank is equal to the dimension of the vector space generated by the columns (resp. of the rows), or to the maximum order of a non-vanishing subdeterminant of the matrix.

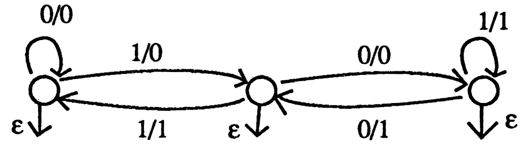
2. Subsequential functions

a. Hankel characterization

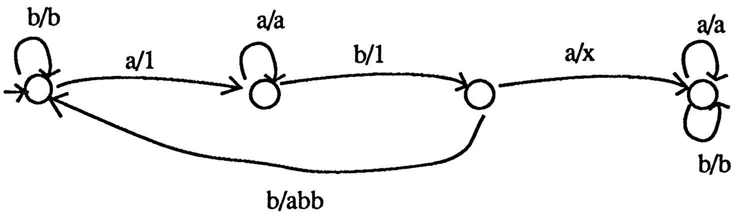
A (partial) function x: $A^* \longrightarrow B^*$ is (left-to-right) **subsequential** if is is computed by some subsequential transducer (i.e. a generalized sequential machine with endmarker). Locally, a subsequential transducer looks like this:



When input letter a is read, than output word u is produced, and in case a is the last letter of the input word, a final output v is produced.
Examples of subsequential transducers are integer division in a given base by a given integer, and pattern substitution.



integer division by 3 in base 2



pattern substitution: first occurrence of aba is replaced by x

Define the Hankel matrix of the function α: $A^* \longrightarrow B^*$ to be the $A^* \times A$ - matrix

$$(\alpha(u\ v))_{u,v \in A^*}$$

The Hankel matrix of the pattern substitution above is

1	a	b	aa	ab	ba	bb	aaa	aab	x	abb	baa
a	aa	ab	aaa	aab	x	abb	aaaa	aaab	ax	aabb	xa
b	ba	bb	baa	bab	bba	bbb	baaa	baab	bx	babb	bbaa
aa	aaa	aab	aaaa	aaab	ax	aabb	aaaaa	aaaab	aax	aaabb	axa
ab	x	abb	xa	xb	abba	abbb	xaa	xab	xba	xbb	abbaa
ba	baa	bab	baaa	baab	bx	babb	baaaa	baaab	bax	baabb	bxa
bb	bba	bbb	bbaa	bbab	bbba	bbbb	bbaaa	bbaab	bbx	bbabb	bbbaa
aaa	aaaa	aaab	aaaaa	aaaab	aax	aaabb	aaaaaa	aaaaab	aaax	aaaabb	aaxa
aab	ax	aabb	axa	axb	aabba	aabbb	axaa	axab	axba	axbb	aabbaa
x	xa	xb	xaa	xab	xba	xbb	xaaa	xaab	xaba	xabb	xbaa
abb	abba	abbb	abbaa	abbab	abbba	abbbb	abbaaa	abbaab	abbx	abbabb	abbbaa
baa	baaa	baab	baaaa	baaab	bax	baabb	baaaaa	baaaab	baax	baaabb	baxa
bab	bx	babb	bxa	bxb	babba	babbb	bxaa	bxab	bxba	bxbb	babbaa
bba	bbaa	bbab	bbaaa	bbaab	bbx	bbabb	bbaaaa	bbaaab	bbax	bbaabb	bbxa
bbb	bbba	bbbb	bbbaa	bbbab	bbbba	bbbbb	bbbaaa	bbbaab	bbbx	bbbabb	bbbbaa

There are not a finite number of distinct lines. However, a function is subsequential if and only if there exist a finite number of lines $L_1, .., L_n$ such that each line of its Hankel matrix is of the form uL_i for some word u.

The analogue property for the columns holds only when the function is subsequential from right to left; for example, multiplication by a given integer in a given base.

b.Choffrut's characterization of subsequential functions

Define the (left) distance between two words by

$$\| u, v, \| = |u'| + |v'|$$

where $u = xu'$, $v = xv'$ and x is the longest common left factor of u and v. We say that a function α: $A^* \longrightarrow B^*$ is uniformly bounded if for any integer k, there exists an integer K such that

$$\forall\ u, v \in \text{dom}(\alpha),\ \| u, v \| \leq k \Longrightarrow \| \alpha u, \alpha v \| \leq K$$

Intuitively, this means that if u, v have a long common prefix, then so are also α u and α v. Choffrut's theorem states that a function α is subsequential if and only if α is uniformly bounded and α^{-1} preserves rationality of languages. This theorem admits as a corollary the theorem of Ginsburg and Rose characterizing sequential functions.

3.Rational functions

Rational functions are obtained by closure under composition of left-to-right and right-to-left subsequential functions. Elgot and Mezei have shown that actually one composition is enough (for a direct proof, see Arnold and Latteux). Equivalently, a function α: $A^* \longrightarrow B^*$ is rational if its graph is a rational subset of the monoid $A^* \times B^*$ (i.e α is a functional rational transduction). Equivalently, α is obtained by a matrix representation, that is, there exists an homromrorphism u: $A^* \longrightarrow M_n(2^{B^*})$ (the matrix semiring over the subsets of B^*, with union and product), a lign matrix λ and column matrix γ of size n, such that for any word

$$\alpha\,(w) = \lambda\,\mu\,(w)\,\gamma$$

(where we identify a word u and the singleton {u}).

4.Schützenberger's Hankel characterization of rational functions

This theorem characterizes rational functions through a Hankel property. In order to justify the terminology "Hankel property", recall that a Hankel matrix $(a_{i+j})_{i,j\geq 1}$ over a field has finite rank if and only if there exist functions $\beta_1,\ldots\beta_n,\ldots\gamma_1,\ldots,\gamma_n$ from N into the field such that for any integers i, j, one has

$$a_{i+j} = \sum_{r=1}^{n} \beta_r\,(i)\,\gamma_r\,(j)$$

Schützenberger's theorem is the following: a function $\alpha: A^* \longrightarrow B^*$ is rational if and only if there exist functions $\beta_1,.., \beta_n, \gamma_1,..., \gamma_n: A^* \longrightarrow B^*$ such that for any words u, v

$$\alpha (u\, v) = \bigcup_{1 \leq r \leq n} \beta_r (u)\, \gamma_r (v)$$

We call this latter property the Hankel property. The direct part of the theorem follows directly from a matrix representation for α. The converse is more difficult, and we give a new proof through our new characterization of rational functions (next section).

5.A Nerode-like characterization of rational functions

Define the syntactic left conquence of a function $\alpha: A^* \longrightarrow B^*$ by : $u \sim v$ iff

$$\sup \{ \, \| \, \alpha (fu), \alpha (fv) \, \| \, , f \in A^* \} < \infty$$

(the convention with $\varnothing$ is that $\| \, \varnothing, w \, \| = \infty$ for any word w and $\| \, \varnothing,\varnothing \, \| = 0$). As an example take the function x: $\{a, b\}^* \longrightarrow \{a, b\}^*$ which removes odd runs in a word, e.g. α (a b b a a a b a a) = b b a a (the odd runs a, aaa, b are removed). Then $a^2 \sim 1$, because $\alpha (f\, a^2) = \alpha (f)\, 1\, a^2$ or $\alpha (f)$, depending on the parity of number of a's at the end of f; so $\sup \{ \| \, \alpha (fa^2), \alpha (f) \, \| , f \in A^* \} = 2 < \infty$. On the other hand, we don't have $a \sim 1$, because for even n, $\alpha (a^n\, a) = 1$ and $\alpha (a^n\, 1) = a^n$, so that $\| \, \alpha (fa), \alpha (f) \, \|$ is unbounded.

Note that if α is the characteristic function of a language L (α (w) = 1 if w $\in$ L, and = $\varnothing$ otherwise), then $\sim$ is the syntactic left congruence of L.
Theorem 1 A function α is rational if and only if its syntactic left congruence has finite index and if α^{-1} preserves rationality.

One part of the proof is to show that the Hankel property (see previous section) is preserved under composition, implies that dom (α) is rational, and more generally that α^{-1} preserves rationality. Another part relies on Choffrut's theorem.

6. Effectiveness

Our constructions are effective. To see it, let us say that two functions α,β: $A^* \longrightarrow B^*$ are adjacent if $\sup \{ \| \alpha(f), \beta(f) \|, f \in \mathrm{dom}(\alpha) \cap \mathrm{dom}(\beta) \} < \infty$

Proposition One can decide if two rational funcitons α, β are adjacent. In case they are, the function

$$\alpha \wedge \beta: \mathrm{dom}(\alpha) \cap \mathrm{dom}(\beta) \longrightarrow B^*$$
$$w \longrightarrow \text{longest common prefix of } \alpha(w) \text{ and } \beta(w)$$

is rational and may be effectively constructed.

As an application, note that for the syntactic left congruence $\sim$ of a function α, one has $u \sim v$ if and only if the two functions $f \longrightarrow \alpha(fu)$ and $f \longrightarrow \alpha(fv)$ have same domain and are adjacent.

7. Bimachines

A bimachine reads the input word in some sense simultaneously from left to right and from right to left. Formally, it means that there are two sets of states, left and right states L and R (which are respectively right and left A^* - modules), with initial states l_o, r_o, an output function ω: $L \times A \times R \longrightarrow B^*$, and left and right final functions λ and ρ. The function α computed by the bimachine is given by:

$$\alpha(a_1...a_n) = \lambda\,(a_1...a_n\, r_o) \prod_{i=1}^{n} \omega\,(l_o a_1...a_{i-1}, a_i, a_{i+1}...a_n r_o)\, \rho\,(l_o\, a_1...a_n)$$

When R is a singleton, then the bimachine is a (left-to-right) subsequential transducer. A bimachine in the sense of Eilenberg has $\lambda = \rho = 1$ (constant function).

As is well-known, a function is rational if and only if it is ocmputed by some bimachine.

8. A canonical bimachine

The syntatic (left) adjacency of α is the relation on A^* defined by $u \longleftrightarrow v$ if $\sup \{ \| \alpha(fu), \alpha(fv) \|, f \in A^*, \alpha(fu) \neq \emptyset \neq \alpha(fv)\} < \infty$

If α is a total function, then $\longleftrightarrow$ is the syntactic left congruence, and α is uniformly bounded if and only if u $\longleftrightarrow$ for any words u and v. Note that $\longleftrightarrow$ is however not transitive in general.

It is verified that if R is the set of right states of some bimachine computing α, then the left congruence on A^* defined by r_0 and the left action of A^* on R is compatible with the syntactic adjacency relation of α, i.e:

$$u\, r_0 = v\, r_0 \Longrightarrow u \longleftrightarrow v$$

The next result is a converse.

Theorem 2 let $\sim$ be a left congruence which is compatible with $\longleftrightarrow$. Then there exists a bimachine computing α with set of right states $R = A^* / \sim$.
The bimachine which we construct is canonical, once $\sim$ is chosen. If α is a total function, then it has the minimum number of left states among all bimachines computing α with set of right states R. The construction is long and technical.

9.Open problems

a.Find enough "morphisms" between bimachines such that the following result holds: there are only a finite number of "minimal" bimachines computing a given rational function.

b.Characterize the rational functions which may be computed by a bimachine whose transition monoid are both aperiodic. Tentative conjecture: α is as above $\Longleftrightarrow$ the period of α^{-1} (L) divides that of L, for any rational language L.

c.Characterize the rational functions which are both left-to-right and right-to-left subsequential. In the case of numerical functions (i.e. α (A^*) is contained in a cyclic submonoid), this has been done by Choffrut and Schützenberger.

Bibliography

A. Arnold, M. Latteux, A new proof of two theorems about rational transductions, Theor. Comput. Sci.8 (1979) 261-263.

J. Berstel, Transductions and context-free languages, Teubner (1979).

C. Choffrut, A generalization of Ginsburg and Rose's characterization of g. - s.- m. mappings, Lecture Notes Comput. Sci. 71 (1979) 88-103.

C. Choffrut, K. Culik, Properties of finite and pushdown transducers, Siam J. Comput. 12 (1983) 300-315.

C. Choffrut, M.-P. Schützenberger, Counting with rational functions, Theor. Comput. Sci. 58 (1988) 81-101.

S. Eilenberg, Automata, languages and machines, vol.A, Acad. Press (1974).

S. Ginsburg, An introduction to mathematical machine theory, Addison Wesley (1962).

G.N. Raney, Sequential functions, J. Assoc. Comput. Mach.5 (1958) 177-180.

C. Reutenauer, Subsequential functions: characterizations, minimization, examples, Intern. Meeting of Young Comput. Scientists, Proc., J. Kelemen ed., Lecture Notes Comput. Sci., to appear.

M.-P. Schützenberger, A remark on finite transducers, Information adn Control 4 (1961) 185-196.

M.-P. Schützenberger, Sur une variante des fonctions séquentielles, Theor. Comput. Sci. 4 (1977) 47-57.

M.-P. Schützenberger, Une propriété de Hankel des relations fonctionnelles entre monoïdes libres, Advances Maths 24 (1977) 274-280.