On Fatou properties of rational languages [1]

Christian Choffrut
L.I.A.F.A., Université Paris 7,
Tour 55-56, 1 er étage,
2 Pl. Jussieu – 75 251 Paris Cedex – France,
email: cc@liafa.jussieu.fr

Juhani Karhumäki
Dept. of Mathematics and
Turku Centre for Computer Science
University of Turku, SF 20014, Turku, Finland,
email: karhumak@cs.utu.fi

Abstract

We consider so-called Fatou-type properties of rational languages and relations. For example, if the rational expression $E(X_1, \ldots, X_n)$ over arbitrary languages $X_1, \ldots, X_n$ represents a rational language, then each language X_i can be replaced by a rational language Y_i such that $E(X_1, \ldots, X_n) = E(Y_1, \ldots, Y_n)$ holds and moreover $X_i \subseteq Y_i$ for $i = 1, \ldots, n$. Some similar results for rational relations are established and an open problem on rational relations is stated.

1 Introduction

A classical result of Fatou, when translated in the terminology of formal series, states that if the coefficients of a $\mathbb{Q}$-rational series in one variable are in $\mathbb{Z}$ then the series is $\mathbb{Z}$-rational, [7]. That is to say, that $\mathbb{Z}$-rational series possess the *Fatou property* with respect to $\mathbb{Q}$. This has been subsequently extended to other semirings and for series in several variables, cf. e.g. [14] or [2]. For example, it is well-known that $\mathbb{N}$-rational series possess the Fatou property with respect to $\mathbb{Q}_+$ but not respect to $\mathbb{Z}$, see [6] or [14].

Though this paper can be viewed as a contribution to the Fatou property in a very strict sense, as explained in the conclusion, we will understand here "Fatou-type" properties in a more general way as, loosely speaking, those claiming that if something constructed from "complicated" parts is actually "simple" then it can be constructed from "simple" parts.

[1] Supported by the Academy of Finland under the grant # 44087

At this point and in order to explain more specifically what we are after, an illustration will suffice. It was observed in [5, Chap. VII] that whenever the product of two languages, say $X \cdot Y$, is rational, then there exist rational languages $\overline{X}$ and $\overline{Y}$ such that $\overline{X} \cdot \overline{Y} = X \cdot Y$. Moreover, $\overline{X}$ and $\overline{Y}$ are obtained in a monotone way in the sense that $X \subseteq \overline{X}$ and $Y \subseteq \overline{Y}$ hold. Later, this property was reproved several times, cf. e.g. [9] and [12].

The goal of this note is to consider the above property and its extensions in a systematic way. First we point out that the result holds in a much broader form, namely if for some languages $X_1, \ldots, X_n$, the rational expression $E(X_1, \ldots, X_n)$ denotes a rational language, then the X_i's can be "saturated" in a unique way into subset-maximal rational languages $Y_1, \ldots, Y_n$ such that $E(Y_1, \ldots, Y_n)$ denotes the same language as $E(X_1, \ldots, X_n)$ and that $X_i \subseteq Y_i$ for all $i = 1, \ldots, n$. The use of the syntactic monoids makes the proof particularly simple.

Although in the above canonical construction from X_i to Y_i, Y_i is unique, this need not be true if the condition $X_i \subseteq Y_i$ is relaxed. Indeed, we give an example of a unary finite (and hence rational) language X having different maximal square roots, i.e., there exist two incomparable subset-maximal subsets Y and Z satisfying $Y^2 = Z^2 = X$.

Our main contribution is the generalization of this property to finite systems of rational expressions $E_j(X_1, \ldots, X_n)$ for $j = 1, \ldots, m$ over the common languages $X_1, \ldots, X_n$. This allows us to establish a Fatou property for finite transducers in the particular case where the transducer realizing a relation from a free monoid to another one has a deterministic underlying input automaton and arbitrary unary outputs for each transition. We call such transducers "input deterministic unary transducers", and we show for those that if the unary outputs associated with all inputs are rational then the outputs can be replaced, transition by transition, by rational languages. In proving the latter result we show that the equivalence problem for input deterministic unary transducers is decidable. This should be compared with two important undecidability results which state that if either one of the restrictions, i.e., input deterministic or unary, is relaxed, then the equivalence becomes undecidable, see [8, 10] and [11]. In the conclusions we propose the Fatou problem for rational relations in its broadest sense.

2 Preliminaries

In this section we recall necessary basics on automata theory, see [6], and we fix our notation.

2.1 Rational subsets

For a finite alphabet A we denote by A^* (resp. A^+) the free monoid (resp. free semigroup) generated by A. The identity of A^* referred to as the empty word, is denoted by 1. The *rational operations* on subsets of A^* (on languages) are the set union, the concatenation (product) and the star operation (also known as the Kleene star operation). The family of *rational* subsets of A^* constitutes of those subsets of A^* which are obtained from singletons of A^* by a finite number of applications of the operations of union, concatenation and star, and is denoted by $\mathrm{Rat}A^*$.

We recall that rational languages have several algebraic characterizations. It is a consequence of Kleene theorem establishing the equivalence between rational and *recognizable* (or *regular*) languages that a language is rational if and only if there exist a finite monoid M, a morphism $\phi : A^* \to M$ and a finite subset $F \subseteq M$ such that $L = \phi^{-1}(F)$. The monoid M is said to *recognize* L. The *syntactic monoid* of L is the smallest monoid recognizing L. It is uniquely defined up to an isomorphism.

2.2 Rational relations

We refer the interested reader to the monographs [1, 6] for a more comprehensive exposition of the theory of rational relations. For our purpose the following will suffice.

A rational transducer is a tuple $\mathcal{T} = (Q, A, B, E, Q_-, Q_+)$ where Q is the finite set of *states*, $Q_- \subseteq Q$ is the set of *initial* states, Q_+ is the set of *final* states, A and B are the *input* and *output* alphabets and $E \subseteq Q \times A^* \times B^* \times Q$ is the finite set of *transitions*. A *path* is a sequence

$$p = (q_0, u_1, v_1, q_1), (q_1, u_2, v_2, q_2), \dots , (q_{n-1}, u_n, v_n, q_n) \tag{1}$$

with $(q_{i-1}, u_i, v_i, q_i) \in E, i = 1, \dots , n$. Its *label* is the pair $(u_1u_2 \dots u_n, v_1v_2 \dots v_n)$ and $u_1u_2 \dots u_n$ and $v_1v_2 \dots v_n$ are respectively its *input* and *output* labels.

The path is *successful* if it leads from one initial state to some final state, i.e., $q_0 \in Q_-$ and $q_n \in Q_+$. The transducer $\mathcal{T}$ *defines* the relation $||\mathcal{T}|| \subseteq A^* \times B^*$ which is the set of labels of all successful paths. It is also said that $||\mathcal{T}||$ is the *behavior* of $\mathcal{T}$. The extension of the Kleene theorem for relations in $A^* \times B^*$ asserts that a relation is the behavior of a rational transducer if and only if it is *rational*, i.e. obtainable from singletons of $A^* \times B^*$ by a finite number of the operations of union, product and star.

The following notion is not standard but is quite natural. An *input deterministic transducer* is like a rational transducer except that Q_- is reduced to a unique state q_- and that the set of transitions E defines a partial function from $Q \times A$ into $\mathcal{P}(B^*) \times Q$ where $\mathcal{P}(B^*)$ is the power set of B^*. Observe that the behavior (defined in the obvious way) of an input deterministic transducer is not necessarily a rational relation.

A transducer, whether rational or not, is *output unary* whenever the output monoid has a single generator and can thus be identified to the additive monoid of the non-negative integers.

2.3 Rational expressions

A *rational expression* over a set $\{\xi_1, \ldots, \xi_n\}$ of variables, is an element of the free algebra with generators $\xi_1, \ldots, \xi_n$ and signature $\Sigma = \{+, \cdot, *\}$ where $+$ and $\cdot$ are two binary and $*$ is a unary symbol. Since the symbols will be interpreted as the three rational operations, we adopt the infix instead of the prefix notation, e.g., $(\xi_1 + \xi_3^*) \cdot \xi_2$. The variables may be interpreted in any subalgebra of subsets of an arbitrary monoid closed under the rational operations but our favorite interpretations are $\mathcal{P}(A^*)$, $\mathcal{P}(A^* \times B^*)$, $\mathrm{Rat}(A^*)$ and $\mathrm{Rat}(A^* \times B^*)$

3 A rediscovered property

In this section we consider the most basic Fatou property of rational languages stated in the next Theorem. It was first proved in [5] by means of quotients, later reproved among other things in [9] and finally explained in [12] through a direct construction based on automata.

THEOREM 1 *Assume that L is rational and has the decomposition $L = X \cdot Y$ where $X, Y \subseteq A^*$. Then there exist two rational languages $\overline{X}$ and $\overline{Y}$ such that $L = \overline{X} \cdot \overline{Y}$, and moreover $X \subseteq \overline{X}$ and $Y \subseteq \overline{Y}$.*

We will prove a more general result. Intuitively, we will show that whenever some rational expression involving arbitrary languages represents a rational language, then each language occurring in the expression can be replaced by some rational language. The question is, how do we define these rational languages? If we deal with the star operation, it is well known that whenever X^* is rational then there exits a unique subset-minimal language Y such that $Y^* = X^*$ and that this language is rational. In other words, if we view the problem as trying to solve the equation $K = X^*$ for some fixed rational subset K, taking the subset-minimal solution seems to be a good candidate. However, this no longer holds if we consider the concatenation product. Indeed, we have $a^* = (a^n)^* \cdot (1+a)^{n-1}$ and thus there are infinitely minimal (and rational) solutions below, say (a^*, a^*). But it is even worse with the third rational operation since the free monoid can be represented as the union of two nonrecursive sets which are the complements of one another. We are more lucky with maximal solutions as we turn to prove.

We consider the set of n-tuples of subsets of a free monoid ordered componentwise, i.e., $(X_1, \ldots, X_n) \leq (Y_1, \ldots, Y_n)$ if and only if for each $i = 1, \ldots, n$, we have $X_i \subseteq Y_i$. Since the three rational operations are monotone in each variable, so is each rational expression $E(\xi_1, \ldots, \xi_n)$ interpreted as mapping from $(\mathcal{P}(A^*))^n$ to $\mathcal{P}(A^*)$.

Theorem **2** [saturation] *Let $E_j(\xi_1, \ldots, \xi_n)$ for $j = 1, \ldots, m$ be m rational expressions in the variables $\xi_1, \ldots, \xi_n$ and let $K_1, \ldots, K_m \subseteq A^*$ be rational subsets. If the system*

$$E_j(X_1, \ldots, X_n) = K_j, \; j = 1, \ldots, m \tag{2}$$

is satisfiable then it has maximal solutions. Furthermore, the components of all these maximal solutions are rational.

Observe that Theorem 1 corresponds to $E(\xi_1, \xi_2) = \xi_1 \xi_2$.

Proof. For all $K_j = E_j(X_1, \ldots, X_n)$, $j = 1, \ldots, m$ let M_j be its syntactic monoid and $\phi_j : A^* \longrightarrow M_j$ its syntactic morphism. In particular, we have $K_j = \phi_j^{-1}(\phi_j(K_j)$. We let $\phi : A^* \longrightarrow M_1 \times \ldots \times M_m$ be the product morphism defined by $\phi(w) = (\phi_1(w), \ldots, \phi_m(w))$ and for all $i = 1, \ldots, n$ we set $F_i = \phi(X_i)$.

Denote by $\mathcal{M}$ the algebra whose elements are the subsets $\phi(Z)$ for some $Z \subseteq A^*$. Observe that $\mathcal{M}$ is closed under set union, concatenation and

Kleene star and that we may interpret any rational expression in $\mathcal{M}$. We order the direct product $\mathcal{M}^n$ componentwise where on each component the ordering is the subset inclusion. Finally let $(G_1, \ldots, G_n) \in \mathcal{M}^n$ be a maximal element greater than or equal to $(F_1, \ldots, F_n)$ satisfying $E_j(G_1, \ldots, G_n) = E_j(F_1, \ldots, F_n)$ for all $j = 1, \ldots m$ where E_j is interpreted in the finite algebra $\mathcal{M}$.

Set $Y_i = \phi^{-1}(G_i)$ for all $1 \leq i \leq n$. Clearly, no n-tuple satisfying the m equalities of (2) is greater than $(Y_1, \ldots, Y_n)$. Indeed, if say $(Z_1, \ldots, Z_n)$ is greater than $(Y_1, \ldots, Y_n)$ then by the maximality of $(G_1, \ldots, G_n)$ and the monotonicity of the rational operations, we have $\phi(Z_i) = G_i$ for all $i = 1, \ldots, n$ and thus $Z_i \subseteq \phi^{-1}(G_i) = Y_i$.

At this point since all Y_i's are rational, we are left with showing that indeed, the Y_i's satisfy (2). Since $X_i \subseteq Y_i$ holds and since the rational operations are monotone we have $K_j \subseteq E_j(Y_1, Y_2, \ldots, Y_n)$. Conversely, we have

$$\begin{aligned}\phi(E_j(Y_1, Y_2, \ldots, Y_n)) &= E_j(\phi(Y_1), \phi(Y_2), \ldots, \phi(Y_n))\\ &= E_j(F_1, F_2, \ldots, F_n) = E_j(\phi(X_1), \phi(X_2), \ldots, \phi(X_n))\\ &= \phi(E_j(X_1, X_2, \ldots, X_n)) = \phi(K_j)\end{aligned}$$

which proves the opposite inclusion and which completes the proof.

■

The previous result shows that if the set of solutions of an equation of the form $K = E(X_1, \ldots, X_n)$ where K is a given rational subset, E a rational expression, and the X_i's are unknowns is not empty, then it has maximal elements which are rational. Furthermore, there is a unique maximal solution above a given solution but the maximal solutions need not be unique. Indeed, set $X = \{0, 2, 3, 7, 10, 12, 14, 15\}$, $Y = \{0, 2, 3, 7, 12, 13, 14, 15\}$ and $Z = \{0 \leq i \leq 30 \mid i \neq 1, 8, 11, 23\}$. Then $X^2 = Y^2 = Z$ (the square is meant "additively") and X and Y are maximal with this property.

In [3] the expression $E(\xi) = \xi^2$ is studied in the free monoids and the extra condition is assumed that the product is unambiguous. It is shown that, given a rational subset K, the solution if it exists is unique i.e., there is at most one subset X such that $K = X^2$ holds. It is also proven that it is decidable whether or not the root is rational. Moreover, it is conjectured that this root is necessarily rational. Previous researches in this direction were also done in [13].

4 An application to input deterministic transducers

In this section we apply the above observations to the input deterministic transducers. We first start with a straightforward Lemma whose proof is left to the reader.

LEMMA 1 *Let M be a commutative monoid. For all elements $x, \bar{x}, y, \bar{y}$, $u, \bar{u}, v, \bar{v}$ we have*

$$\left.\begin{array}{rcl} xy & = & \bar{x}\bar{y} \\ xuy & = & \bar{x}\bar{u}\bar{y} \\ xvy & = & \bar{x}\bar{v}\bar{y} \end{array}\right\} \Rightarrow xuvy = \bar{x}\bar{u}\bar{v}\bar{y}.$$

Given a transducer, whether rational or not, we define its *underlying automaton* as the finite automaton obtained by ignoring the output component of each transition. We may then speak of *useful* states which are visited in at least one succesful path. We say that a transducer is *trimmed* if all its states are useful. We have the following simple but interesting result:

THEOREM 3 [equivalence] *Let $\mathcal{T}$ and $\mathcal{T}'$ be a two input deterministic transducers with unary outputs and let M be the number of useful states (q, q') in the direct product of their underlying automata. If $||\mathcal{T}||(u) = ||\mathcal{T}'||(u)$ holds for all $u \in A^*$ of length less than $2M$ then $||\mathcal{T}|| = ||\mathcal{T}'||$.*

Proof. Assume by contradiction that $||\mathcal{T}||(u) \neq ||\mathcal{T}'||(u)$ for some shortest input $u \in A^*$. Clearly $|u| \geq 2M$. Consider the two unique successful paths in $\mathcal{T}$ and $\mathcal{T}'$ with input label u and assume they are traversed synchronously in $\mathcal{T}$ and $\mathcal{T}'$ when reading the different prefixes of u. The combined sequence thus obtained contains $|u| + 1$ pairs of states of $Q \times Q'$ and thus some pair, say (q, q'), is visited at least three times. This means that the paths can be factored in $\mathcal{T}$ and in $\mathcal{T}'$ as

$$q_- \xrightarrow[\mathcal{T}]{u_1|X_1} q \xrightarrow[\mathcal{T}]{u_2|X_2} q \xrightarrow[\mathcal{T}]{u_3|X_3} q \xrightarrow[\mathcal{T}]{u_4|X_4} q_+ \quad \text{and} \quad {q'}_- \xrightarrow[\mathcal{T}']{u_1|{X'}_1} q' \xrightarrow[\mathcal{T}']{u_2|{X'}_2} q' \xrightarrow[\mathcal{T}']{u_3|{X'}_3} q' \xrightarrow[\mathcal{T}']{u_4|{X'}_4} {q'}_+,$$

where u_2 and u_3 are non empty and $q_+ \in Q_+$ and ${q'}_+ \in {Q'}_+$. By induction, we have the equalities $X_1X_4 = {X'}_1{X'}_4$, $X_1X_2X_4 = {X'}_1{X'}_2{X'}_4$ and

$X_1X_3X_4 = X'_1X'_3X'_4$, which by the preceding lemma implies $X_1X_2X_3X_4 = X'_1X'_2X'_3X'_4$, a contradiction.

■

Now, we are ready to a Fatou-type property of rational relations:

THEOREM **4** [fatou] *Let $\mathcal{T}$ be a trimmed input deterministic transducer with unary outputs such that $||\mathcal{T}||(u)$ is rational for each $u \in A^*$ of length less than $2N$, where N is the number of states of $\mathcal{T}$. Then $||\mathcal{T}||$ is rational.*

Proof. The proof follows the general "pumping property" pattern. Indeed, consider the outputs of the transitions as unknowns X_i for $i = 1, \dots, m$, where m is the number of transitions, and for each input u of length $\ell < 2N$ consider the equality $||\mathcal{T}||(u) = X_{i_1} \dots X_{i_\ell}$. We obtain thus a finite system of equalities indexed by all inputs of length less than ℓ, and we know, by Theorem 2, that it is possible to substitute a rational subset $X'_i \subseteq \mathbb{N}$ for each output label X_i in such a way that the finite system of equalities still holds. By the previous theorem, the rational transducer $\mathcal{T}'$ thus obtained defines the same relation as $\mathcal{T}$. This completes the proof.

■

Despite of the simplicity of Theorem 3, it is theoretically interesting. Indeed, if either condition "input deterministic" or "unary outputs" is relaxed, the problem becomes undecidable as shown in [10, 8] and [11], respectively.

As yet another consequence we obtain a decision result:

COROLLARY **1** *Let $\mathcal{F}$ be a family of subsets of $\mathbb{N}$ satisfying the following properties:*

i) *$\mathcal{F}$ is closed under set union and concatenation;*

ii) *There is an algorithm which given an element $X \in \mathcal{F}$ determines whether or not X is rational.*

Then there exists an algorithm which given an input deterministic transducer with outputs in $\mathcal{F}$ decides whether or not $||\mathcal{T}||$ is rational.

We conclude this section by noting that we can drop the assumptions on both the input determinism and the unarity in Theorem 4 if we assume that the domain of a transducer is finite, e.g. the transducer is *acyclic*:

THEOREM **5** *Let $\mathcal{T}$ be an acyclic transducer such that for each $u \in dom(\mathcal{T})$ the output $||\mathcal{T}||(u)$ is rational. Then $||\mathcal{T}||$ is rational as well.*

Proof. Follows directly from Theorem 2. ■

5 Conclusions

The original Fatou problem can be abstracted away by considering two semirings $K' \subseteq K$. Given a K-rational series whose coefficients are in K' is it true that it is K'-rational? The classical Fatou Theorem corresponds to the case when $K' = \mathbb{Z}$ and $K = \mathbb{Q}$.

We recall that a rational relation on $A^* \times B^*$ can be interpreted, in a natural way, as a rational series over A^* with coefficients in $\mathrm{Rat}B^*$, this is as an element of $\mathrm{Rat}_{\mathrm{Rat}B^*}(A^*)$ in terms of [6]. Consequently, the above problem can be formulated for rational transducers as a question whether or not the Fatou property holds when K' is the semiring of all rational subsets of a free monoid B^* and K is the power set of all subsets of B^*. To our knowledge this important problem is still open.

In order to comment on this, we need to recall some background on rational series. It can be proven that given a semiring K, a K-series $s : A^* \to K$ is rational if and only if there exist an integer n and a triple (λ, μ, γ) where μ is matrix representation of A^* in the monoid $K^{n\times n}$ of square matrices of dimension n with entries in K, and $\lambda \in K^{n\times 1}$ and $\gamma \in K^{1\times n}$ are two matrices such that $s(u) = \lambda\mu(u)\gamma$ holds, see [2, p.8]. Now, consider the mapping σ of K onto $\{0, 1\}$ defined by $\sigma(k) = 0$ if and only if $k = 0$ and extend it to any $K^{r\times s}$ matrix where r and s are arbitrary integers. We say that two triples (λ, μ, γ) and $(\lambda', \mu', \gamma')$ have the same *support* if $\sigma(\lambda) = \sigma(\lambda')$, $\sigma(\gamma) = \sigma(\gamma')$ and for all $a \in A$, $\sigma(\mu(a)) = \sigma(\mu'(a))$ holds.

Let now K' be a subsemigroup of K. A second (and stronger) version of the Fatou question asks whether the condition that $s(u) = \lambda\mu(u)\gamma$ is in K' for all $u \in A^*$ implies that there is a triple $(\lambda', \mu', \gamma')$ with the same support as (λ, μ, γ) and with entries in K' such that $s(u) = \lambda'\mu'(u)\gamma'$ holds. A third (and even stronger) version assumes that there is a natural ordering on K and that the triple $(\lambda', \mu', \gamma')$ satisfies the further condition that for all positions $\lambda_i \leq {\lambda'}_i$, $\gamma_i \leq {\gamma'}_i$ and $\mu_{i,j}(a) \leq {\mu'}_{i,j}(a)$ hold. The reader can easily observe that Theorem 4 is a partial answer to the third version of the Fatou property with $K' = \mathrm{Rat}B^*$ and $K = \mathcal{P}(B^*)$ and $\leq$ is the subset inclusion.

For more Fatou properties concerning some rational relations we refer the reader to [4].

References

[1] J. Berstel. *Transductions and context-free languages.* B. G. Teubner, 1979.

[2] J. Berstel and C. Reutenauer. *Rational Series and Their Languages.* Springer Verlag, 1988.

[3] A. Bertoni and P. Massazza. On the square root of regular languages. submitted for publication.

[4] Christian Choffrut. Rational relations and rational series. *Theoret. Comput. Sci.*, 98:5–13, 1992.

[5] J. H. Conway. *Regular Algebra and Finite Machines.* Chapman Hall, 1971.

[6] S. Eilenberg. *Automata, Languages and Machines*, volume A. Academic Press, 1974.

[7] Fatou. Sur les séries entières à coefficients entiers. *C. R. Acad. Sci. Paris*, pages 342–344, 1904.

[8] O. H. Ibarra. The unsolvability of the equivalence problem for ϵ-free ngsm's with unary input (output) alphabets and application. 7:520–532, 1978.

[9] L. Kari. *On insertions and deletions in formal languages.* PhD thesis, Turku University, 1991.

[10] L. P. Lisovik. The identity problem of regular events over cartesian products of free and cyclic semigroups. *Doklady of Academy of Sciences of Ukraine*, 6:480–497, 1979.

[11] L. P. Lisovik. The euqivalence problem for finite substitutions on regular languages. *Doklady of Academy of Sciences of Russia*, 357:293–301, 1997.

[12] A. Mateescu, A. Salomaa, and S. Yu. On ther decomposition of finite languages. Technical Report 222, TUCS, 1988.

[13] A. Restivo. Some decision results for recognizable stes of arbitrary monoids. In *ICALP'78*, volume 62, pages 363–371. Lecture Notes in Computer Science, 1978.

[14] A. Salomaa and M. Soittola. *Automata-Theoretic Aspects of Formal Power Series*. Springer Verlag, 1978.