CAHIERS DE
TOPOLOGIE ET GÉOMÉTRIE DIFFÉRENTIELLE
CATÉGORIQUES

S. KASANGIAN
R. ROSEBRUGH

Decomposition of automata and enriched category theory

Cahiers de topologie et géométrie différentielle catégoriques, tome 27, nº 4 (1986), p. 137-143
<http://www.numdam.org/item?id=CTGDC_1986__27_4_137_0>

CAHIERS DE TOPOLOGIE
ET GÉOMÉTRIE DIFFÉRENTIELLE
CATÉGORIQUES

Vol. XXVII-4 (1986)

DECOMPOSITION OF AUTOMATA AND ENRICHED CATEGORY THEORY

BY S. KASANGIAN and R. ROSEBRUGH *

RÉSUMÉ. On étend un résultat de la théorie des automates finis concernant la décomposition concaténative de langages réguliers (Paz & Pelag) aux automates à arbres. On utilise dans cet article la théorie catégorielle enrichie des automates, où les automates à arbres se laissent décrire comme catégories enrichies sur une bicatégorie.

1. INTRODUCTION.

The study of non-deterministic dynamics viewed as categories enriched in a biclosed monoidal category constructed from the input monoid [1, 2, 4], and its extension to tree automata [3], is here applied to decomposition of the associated behaviours using subsets of the state spaces. Our main result is related to the concatenative decompositions of regular events defined by Paz & Peleg [5]. They showed that the behaviour of a deterministic finite automaton (with a free input monoid) is decomposable exactly when there is a subset of the state set through which every computation passes and which, together with an associated subset, defines a decomposition.

We give a decomposition of the behaviour of (= set of trees accepted by) a deterministic tree automaton in the sense of [3]. The decomposition involves a set of tuples of trees substitutable into a final set of operations of fixed arity. The result of Paz & Peleg reappears essentially as a special case of the result just quoted.

----

*) This research was partially supported by a grant from NSERC Canada.

We recall briefly some definitions relative to tree automata viewed as enriched categories. For further details, see [3].

Given an algebraic theory T (a category whose objects are finite sets $[n] = \{1, \ldots, n\}$, $n = 0, 1, \ldots$ and which admits the category of finite sets as a subcategory, with $[0] = \emptyset$ as initial object and $[1] = \{1\}$, such that $[m]$ is the m-fold coproduct of [1]), we construct a bicategory B(T) which has the same objects as T, the 1-cells from $[n]$ to $[m]$ in B(T) are the subsets of $T([n], [m])$ and 2-cells are inclusions. Composition of 1-cells and identities are the obvious ones. B(T) is locally partially ordered, locally complete and cocomplete and also biclosed. The arrows of T seen as 1-cells of B(T) are called *atoms*.

Let X be a B(T)-category with only one object (say $*$) over [0]. We call an object b over $[n]$ *reachable* if it is the *cotensor of* $*$ *along an atom*. We call a skeletal B(T)-category *reachable* if all the objects are reachable.

Reachable B(T)-categories correspond to *non-deterministic T-algebras* (T-dynamics). Further, a reachable B(T)-category X corresponds to a *deterministic* reachable (i.e., definable) T-algebra if its underlying category is *discrete* and if it is *cotensored along the atoms*.

We denote by $X_{[n]}$ the "fibre" of X over $[n]$. Then $X_{[1]}$ is the carrier of the algebra and $X_{[0]} = *$. Denoting by $[n^\frown]$ the trivial, one object category over $[n]$, a *tree automaton* (i.e., a T-dynamics with a subset $F \subset X_{[1]}$ of *final states*) can be described as a triple (X, i, τ) as follows:

X is a reachable (possibly deterministic) B(T)-category;

$i\colon X \to [0^\frown]$ is the *initial module*, given by $i(b) = X(b, *)$, where $b = (b_1 \ldots b_n)$ is an X-object over $[n]$;

$\tau\colon [1^\frown] \to X$ is the *final module*, given by

$$\tau(b) = \{\, g \in T([1], [n]) \mid \text{there exists } a \in F \subset X_{[1]} \text{ and } g \in X(a, b) \,\}.$$

Notice that if the automaton is deterministic, the definition above can be stated using the cotensor, namely

$$\tau(b) = \{\, g \in T([1], [n]) \mid \text{there exists } a \in F \subset X_{[1]} \text{ and } b \pitchfork g = a \,\}.$$

Thus, the module i provides sets of tuples of trees (terms), whereas τ selects sets of operations which are "successful" if performed on those trees. The composite module $i.\tau$ is the *behaviour* of the automaton and consists of the set of trees, i.e., 1-cells at $T([1], [0])$, which are recognizable. Henceforth we assume that the tree automata considered are *deterministic*.

2. THE DECOMPOSITION THEOREM.

DEFINITION 2.1. We call a set of states $R \subseteq \text{obj}\, X$ a *decomposition set* for the automaton (X, i, τ) if $i.\tau = \Sigma_{r \in R}\, i(r).\tau(r)$.

DEFINITION 2.2. Given a set of states $S \subseteq \text{obj}\, X$, we call *associated with* S the set of states

$$S^\wedge = \{ s^\wedge \in \text{obj}\, X \mid \Pi_{s \in S}\, \tau(s) \subseteq \tau(s^\wedge) \}.$$

REMARK 2.3. Notice that if b is in X_m and c is in X_n, with $n \neq m$, it is always the case that $\tau(b) \cap \tau(c) = \emptyset$. Hence, the notion of associated set trivializes unless there is an n with $S \subseteq X_n$. Henceforth we assume this whenever we mention associated sets. Notice also that $S \subseteq S^\wedge$ from the definition.

Recall that the behaviour $i.\tau$ of an automaton (X, i, τ) is a module $i.\tau : [1^\wedge] \to [0^\wedge]$ and hence a 1-cell from [1] to [0] in B(T). Notice that it may admit a decomposition into two 1-cells of B(T). We have the following:

DEFINITION 2.4. Let (X, i, τ) be an automaton with behaviour $i.\tau$. We say that 1-cells $D: [1] \to [n]$ and $C: [n] \to [0]$ ($n \neq 0$ and if $n = 1$, then $D \neq 1_{[1]}$) in B(T) are a *decomposition* if $i.\tau = C.D$. Behaviours which admit a decomposition are said to be *decomposable*.

Notice that C is a set of n-tuples of *trees* and D is a set of n-ary operations which is performed successfully on these trees. Hence the definition ensures that the last branching of any tree of the behaviour is n-ary.

We are now able to prove the *Decomposition Theorem:*

THEOREM 2.5. Given an automaton (X, i, τ) *its behaviour is decomposable iff it admits a decomposition set* S *such that*:

$$i.\tau = \Sigma_{q \in S^\wedge}\, i(q) \,.\, \Pi_{q' \in S}\, \tau(q').$$

PROOF. Observe first that, by Remark 2.3, there exists an n such that

$$S \subseteq S^\wedge \subseteq X_n, \quad \Sigma_{q \in S^\wedge}\, i(q) : [n] \to [0] \quad \text{and} \quad \Pi_{q' \in S}\, \tau(q') : [1] \to [n],$$

so sufficiency is obvious. To prove necessity, assume the behaviour is decomposable, i.e.,

$$\text{i}.\tau = ([1] \xrightarrow[D]{} [n] \xrightarrow[C]{} [0], \quad \text{with} \quad n \neq 0.$$

We define $S \subset X_n$ as follows:

$$S = \{ s \in X_n \mid \text{i}(s) \cap C \neq \emptyset \}.$$

We show first that S is a decomposition set. For any 1-cell h in the behaviour $\text{i}.\tau$, there exists an n-tuple of trees $c_1 : [n] \to [0]$ in C and an n-ary operation $d_1 : [1] \to [n]$ in D such that $h = c_1 d_1$. Since X is reachable, there is a $q = *\phi c_1$ in X_n, so that $c_1 \in \text{i}(q) \cap C$ and hence $q \in S$. Since $c_1 d_1$ is in $\text{i}.\tau$, $d_1 \in \tau(q)$ and so we have

$$\text{i}.\tau \subset \Sigma_{q' \in S}\ \text{i}(q').\tau(q').$$

Hence S is a decomposition set. Further,

$$C \subset \Sigma_{q' \in S}\ \text{i}(q') \subset \Sigma_{q \in S^\wedge}\ \text{i}(q)$$

since $S \subset S^\wedge$. Next we show that $D \subset \Pi_{q' \in S}\ \tau(a')$. Given $d \in D$, we know that, for all $c \in C$, $cd \in \text{i}.\tau$ and there is a $q = *\phi c$ in S such that $c \in \text{i}(q)$ and $d \in \tau(q)$. Since S is a decomposition set, *for all* $q' \in S$, there exists a $c^- \in \text{i}(q')$ such that $q' = *\phi c^-$ and $d \in \tau(q')$. Thus it follows that $d \in \Pi_{q' \in S}\ \tau(a')$. Therefore $\Sigma_{q \in S^\wedge}\ \text{i}(q)$ and $\Pi_{q' \in S}\ \tau(q')$ are non-empty and moreover

$$\Sigma_{q \in S^\wedge}\ \text{i}(q)\ .\ \Pi_{q' \in S}\ \text{t}(q') \supset C.D = \text{i}.\tau.$$

To finish to show the reverse inclusion, let

$$z = xy, \quad \text{with} \quad x \in \Sigma_{q \in S^\wedge}\ \text{i}(q) \quad \text{and} \quad y \in \Pi_{q' \in S}\ \tau(q').$$

Now there exists a q^- in $S^\wedge$ with $x \in \text{i}(q)$, i.e., $q^- = *\phi x$. Further, by the definition of an associated set, $y \in \Pi_{q' \in S}\ \tau(q')$ implies $y \in \tau(q)$ *for all* $q \in S^\wedge$. Hence

$$y \in \tau(q^-) \quad \text{and} \quad z = xy \in \text{i}(q^-).\tau(q^-) = \Sigma_{q \in X}\ \text{i}(q).\tau(q) = \text{i}.\tau.$$

REMARK 2.6. Observe that the proof of Theorem 2.5 ensures that the decomposition decribed above is *maximal* with respect to the obvious partial order on the set of pairs of 1-cells which decompose the behaviour. Recall that, for any n, $B(T)([n], [n])$ is a monoid, with identity $1_{[n]}$ and multiplication given by composition of 1-cells.

PROPOSITION 2.7. *Let* (X, i, τ) *be an automaton and* $S \subset X_n$ *a decomposition set. Define*

$$L = \Pi_{q' \in S} \Sigma_{q \in S^{\wedge}} X(q', q).$$

L *is a submonoid of* B(T)([n], [n]).

PROOF. It is immediate that $1_{[n]} \in L$, since $S \subset S^{\wedge}$ and $1_{[n]} \subset X(s, s)$ for all $s \in S$. Notice also that, since the automaton is deterministic, an equivalent definition of the associated set of states (Definition 2.2) is

$$S^{\wedge} = \{ s^{\wedge} \in \text{obj } X \mid s\phi z \in F \text{ for all } s \in S \text{ implies } s^{\wedge}\phi z \in F \}.$$

Now we need to show that if $x, y \in L$, then $xy \in L$. Given a z : $[n] \to [n]$ such that $s\phi z \in F$ for all $s \in S$ and observing that $s\phi y \in S^{\wedge}$, we have that

$$(s\phi y)\phi z = s\phi(yz) \quad \text{for all} \quad s \in S.$$

By the same argument, $(s\phi x)\phi yz \in F$ for all $s \in S$. But

$$(s\phi x)\phi yz = s\phi(xyz) = (s\phi xy)\phi z,$$

so that $s\phi xy \in s^{\wedge}$ for all $s \in S$. Thus $xy \in L$.

The 1-cells of B(T)([n], [n]) are n-tuples of n-ary operations and composition is substitution. If we take $n = 1$, the 1-cells of B(T)([1], [1]) are unary operations so that giving a decomposition set $S \subset X_1$ amounts to considering actions of the monoid L above (a submonoid of B(T)([1], [1])) on a set of trees.

In the next section we will see an interpretation of these results in the more special context of sequential automata.

3. APPLICATIONS TO SEQUENTIAL AUTOMATA.

The B-categorical approach to tree automata admits a straightforward specialization to sequential automata. However, we will follow the lines of [1, 2, 4] giving a slightly different (though obviously equivalent) description in terms of categories enriched in a monoidal biclosed category, i.e., in a biclosed category with one object.

The input monoid X yields a monoidal biclosed category $X^{\sim} = 2^{X}$, where the tensor product is just the Frobenius product of subsets of X and the internal homs are given by left and right quotients. A (not necessarily deterministic) dynamics is then an $X^{\sim}$-category Q where objects q, q' in Q are the states and $Q(q, q')$ is the set of monoid elements which act on q (possibly in a non-deterministic way), carrying it to q'. A deterministic dynamics is an $X^{\sim}$-category which is tensored along the "atoms", i.e., the elements of X, and whose underlying category is discrete. An X-automaton is then a triple (Q, i, τ) as in the following diagram

$$1 \xrightarrow[\tau]{} Q \xrightarrow[i]{} 1$$

where 1 is the trivial, one-object $X^{\sim}$-category and i and τ are the *initial* and *final* modules. The behaviour is again i.τ and it is the subset of X (i.e., a *language*) recognized by the automaton. Modulo a "normalization" described in [4], we can think of these modules, as given by

$$i(q) = \Sigma_{j\in J}\, Q(j, q) \quad \text{and} \quad \tau(q) = \Sigma_{t\in F}\, Q(q, t),$$

where J and F are the sets of initial and final states. A deterministic automaton has a deterministic dynamics and further the initial module is required to be I_* for some $X^{\sim}$-functor I from 1 to Q, i.e., $Q(i, q) = I_*(q)$.

As for reachability, here it means that for all q in Q, $i(q) \neq \emptyset$. The definitions of decomposition set (2.1) and associated set (2.2) apply straightforwardly to this context. The decomposition of a behaviour still exhibits it as the composite of two 1-cells of the (one-object) bicategory $X^{\sim}$. Given a language A in $X^{\sim}$, a decomposition for it is a pair of languages B, C such that A = B.C, B ≠ {e}, C ≠ {e}. This definition applies of course to behaviours and yields the notion of *decomposable behaviour*. The following is the analogue of Proposition 2.5.

PROPOSITION 3.1. *Given an automaton* (Q, i, τ) *with deterministic dynamics, its behaviour is decomposable iff it admits a decomposition set* S *such that* $i.\tau = \Sigma_{q\in S^{-}}\, i(q)\, \Pi_{q'\in S}\, \tau(q')$.

The proof of Proposition 3.1 is nearly identical to that of Proposition 2.5 provided some attention is paid to different interpretations. In particular, recall the different meanings of i and τ and that now the decomposition set and its associated set are obviously constructed without the restrictions of Remark 2.3: Q is all in one "fibre". Further, in the proof the tensor in Q (rather than the cotensor) is used because no contravariance is involved. The same interchange of tensor and cotensor provides the adjustments necessary to prove the analogue of Proposition 2.7.

PROPOSITION 3.2. *Let* (Q, i, τ) *be an automaton with deterministic dynamics and* $S \subset Q$ *a decomposition set. Define*

$$L = \Pi_{q' \in S} \Sigma_{q \in S^\frown} Q(q', q).$$

L *is a monoid.*

Restricting ourselves to deterministic automata (that is with *one* initial state) and observing that the initial state is a decomposition set, we get the following:

PROPOSITION 3.3. *Let* (Q, i, τ) *be a deterministic reachable automaton. Then the behaviour is a monoid iff* $i_o^\frown = F$, *where* i_o *is the initial state.*

Notice finally that by specializing further to *finite state* deterministic automata on a *free* monoid X, we get the results of Paz & Peleg (see [5], Theorem 1, Lemma 3, Theorem 3).

REFERENCES.

1. R. BETTI, Automata e categoria chiuse, *Bol. Un. Mate. Ital.* 17-B (1980), 44.
2. R. BETTI & S. KASANGIAN, A quasi-universal realization of automata, *Rend. Inst. Mat. Univ. Trieste* (to appear).
3. R. BETTI & S. KASANGIAN, Tree automata and enriched categories, *Rendi. Inst. Mat. Univ. Trieste* (to appear).
4. S. KASANGIAN, G. M. KELLY & F. ROSSI, Cofibrations and the realization of non-deterministic automata, *Cahiers Top. et Géom. Diff.* XXIV-1 (1983), 23-46.
5. A. PAZ & B. PELEG, On concatenative decompositions of regular events, *I.E.E.E. Trans. on Computers* C-17, n° 3 (1968), 229-237.

S. KASANGIAN,
Dipto di Matematica "F. Enriques",
Università di Milano
MILANO, ITALIA.

R. ROSEBRUGH
Department of Mathematics
and Computer Science
Mount Allison University
U.S.A.