Notations for Sets of Sequences

Nachum Dershowitz
School of Computer Science
Tel Aviv University
Tel Aviv, Israel
nachumd@tau.ac.il

February 9, 2010

1 Purpose

We are interested in notations for characterizing finite or infinite sequences of elements as compositions of (binary) relations.

2 Sequences

Let Σ be some finite or infinite set of elements, called *points*, and $\Pi = \Sigma \times \Sigma$ all possible pairs of points. Think vertices of a finite or infinite graph and possible edges between them. A *path* (in the complete graph for Σ) corresponds to a sequence of points. We use Π^n for paths with n edges (sequences of $n+1$ points); $\diamond$ for all finite paths; Ω for infinite paths; and Ξ for all paths, finite or infinite.

We list sequences $s = s_1, s_2, \ldots$ with commas as punctuation (and parentheses only when needed), indexing from 1. The length $|s|$ of a sequence s is the number of its items, or ∞ if there are infinitely many. (We do not deal with longer transfinite sequences.) More generally, a comma acts to concatenate sequences: $r, s = r_1, \ldots, r_{|r|}, s_1, s_2, \ldots$, unless $|r| = \infty$, in which case $r, s = r$.

The empty sequence, with $|s| = 0$, is denoted ϕ (rather than ε, to avoid confusion with properties of strings). A point $a \in \Sigma$ may be viewed also as a single-item sequence. It will be convenient to imagine a fictitous "wildcard" element, which we will denote 1. For infinite sequences, let $r_{|r|} = r_\infty = 1$.

By the same token, we now have a special sequence, consisting of the wildcard element 1, also denoted 1, of length 1.

The *join* of two sequences (in the sense of database theory, not mere concatenation) is defined as follows, for $r, s \in \Xi$:

$$rs \quad = \quad \begin{cases} r_1, \ldots, r_{|r|}, s_2, s_3, \ldots & |r| < \infty,\ r_{|r|} = s_1 \\ \phi & |r| < \infty,\ r_{|r|} \neq s_1 \\ r & |r| = \infty \end{cases}$$

When it says $r_{|r|} = s_1$, this entails that r and s are nonempty, so both points exist and are equal. The join of an infinite sequence r with anything is just r because we are not interested in transfinite sequences of length beyond ω. This join operation is associative.

The wildcard sequence acts like a multiplicative unit, in that $1r = r1 = r$ for all r. The empty sequence acts almost as a *zero* (absorbing) element: $\phi s = \phi$, while $r\phi = \phi$ unless r is infinite, in which case $r\phi = r$. (So we have a monoid with a left zero.)

We can use exponentiation for repeated joins: $r^0 = 1$, $r^1 = r$, $r^2 = rr$, etc. Also, $r^\infty = rrr\cdots$.

We *collapse* a nonempty sequence r into the pair of its endpoints $\langle r_1, r_{|r|}\rangle$, denoted $[\![r]\!]$. For the empty sequence, $[\![\phi]\!]$ is undefined. When r is infinite, we have $[\![r]\!] = \langle r_1, 1\rangle$, pairing the initial point with 1, which acts like *any* point of the underlying set Σ.

3 Sets of Sequences

A binary relation B over points Σ is a set of pairs taken from Ξ, which can also be viewed as edges colored B. Similarly, an n-ary relation is a set of sequences of exactly n points. We refer to such sets of sequences as "relations". The pairs in a binary relation can be called "steps", and sequences in a general relation, "paths".

Suppose A and B are binary relations. (Think colorings of edges in the graph.) Rather than viewing their juxtaposition AB as the binary relation $A \circ B$, obtained by composing A and B (namely, $\{\langle x, z\rangle \colon \exists y.(xAy \wedge yBz)\}$), or the 4-letter strings obtained by concatenation ($\{u, v \colon u \in A, v \in B\}$), we view it as the ternary relation obtained by joining an A-step with a B-step, that is, $AB = \{\langle x, y, z\rangle : xAy, yBz\}$.

More generally, let $R, S \in \Xi$ be sets of (finite and/or infinite) sequences over Σ. Define their *join* as follows:

$$RS \quad = \quad \{rs : r \in R, s \in S\}$$

that is, an R-path followed by an S-path. Joining of relations is associative. The empty set (relation) $\varnothing$ acts as a zero element: $\varnothing S = S\varnothing = \varnothing$, for all S. (So, we have a monoid with a zero.) The almost empty relation $\{\phi\}$ behaves almost the same, except that $S\{\phi\}$ will contain all infinite sequences in S.

Let $\mathbf{1} = \{1\}$. This relation acts as a unit: $\mathbf{1}S = S\mathbf{1} = S$, for all S. One can identify $\mathbf{1}$ with the full set of singleton sequences, Σ, since it too acts as an identity element. Sets of singleton sequences are monadic predicates, and act as filters. In particular, $P\Sigma = \Sigma P = PP = P$ for all $P \subseteq \Sigma$. The binary relation $P\Pi$ is the set of edges whose source is in P, while ΠP are edges incident on P. For predicate P and relation R, PR and RP are those sequences in R that begin or end, respectively, with a point in P.

In analogy with regular expressions for strings, we avoid set formers for singletons. So aR for point a and relation R, is those sequences in R that begin with a.

Exponentiating works as expected: $R^0 = \mathbf{1}$, $R^1 = R$, $R^2 = RR$, etc., for any relation $R \subseteq \Xi$. So, R^i, for $i \in \mathbb{N}$, is the set of i-fold joins of paths in R. Define also the following:

1. R^- contains all finite sequences in R in reverse order.

2. $R^* = \bigcup_{i \geq 0} R^i$ is the set of all joins of finitely many sequences of R.

3. $R^+ = \bigcup_{i > 0} R^i$, so $R^* = R^+ + \mathbf{1}$, where $+$ is being used for set union.

4. R^∞ is the set of joins of infinitely many sequences of R.

5. $R^\sim$ is the set $R^* + R^\infty$ of finite and infinite sequences.

Referring back to the definitions at the outset, $\Diamond = \Pi^*$, $\Omega = \Pi^\infty$, and $\Xi = \Pi^\sim$. In the degenerate cases, $\varnothing^+ = \varnothing^\infty = \varnothing$. For any predicate, $P^+ = P^\infty = P$.

To avoid confusion, we indicate by $[\![R]\!]$ the binary relation that relates endpoints of sequences in R, recalling that the initial points of infinite sequences relate to the wildcard element. This *collapsed* relation is

$$[\![R]\!] \quad = \quad \{[\![r]\!] : r \in R, r \neq \phi\}$$

Unfortunately, $[\![RS]\!]$ need not be equal to $[\![R]\!] \circ [\![S]\!]$, on account of the infinite case. For example, let α be an infinite sequence of a's, that is, $\alpha = [\![a]\!]^\infty$. Then, $[\![\alpha b]\!] = [\![\alpha]\!] = \langle a, 1\rangle$, whereas $[\![\alpha]\!] \circ [\![b]\!] = \langle a, b\rangle$.

Now, let $\langle\!\langle R\rangle\!\rangle$ contain all finite sequences whose beginning and end points are also endpoints of a sequence in R, the intermediate points not mattering (symbolized $\diamond$). So

$$\langle\!\langle R\rangle\!\rangle \quad = \quad \left\{r_1 \diamond r_{|r|} : r \in R\right\}$$

Any infinite path r in R contributes all finite paths with the same head, that is, $r_1 \diamond 1$. It follows that

$$s \in \langle\!\langle R\rangle\!\rangle \text{ iff } [\![s]\!] \in [\![R]\!]$$

as long as the fictitious 1 does not appear in the sequences.

Viewed as tuples, $[\![R]\!]$ and R are the same for purely binary R. One may think of composition $A \circ B$ of binary relations as their collapsed join, $[\![AB]\!]$.

4 Properties of Sets of Sequences

Let $R, S \subseteq \Xi$ be sets of sequences. The notation

$$S \models R$$

means that each individual sequence in S is also a sequence in R. (I am using this notation, instead of containment of sets, to stress that this is a comparison of arbitrary, not necessarily binary, relations.) We also allow singletons in place of S, as in $aaa\cdots \models a\,\Omega$

For example, suppose A and B are binary relations. The following is what I call "escaping":

$$B^\infty \models \diamond[\![A(A+B)^\infty]\!]B^\infty$$

This means that there is a point in every infinite B-chain such that an A-step out of that point leads to a potentially "immortal" element (a point that heads at least one infinite path in their union).

5 An Example

Suppose Q is a quasi-order over Σ. Call an infinite sequence $s \in \Omega$ *good* if

$$s \models \Diamond \langle\langle Q \rangle\rangle \Omega$$

that is, if it contains two (not necessarily adjacent) points that are (quasi-) ordered, the earlier one less than or equivalent to the later one. Call a sequence *bad* if it is not good, and *very good* if all its infinite subsequences are good. Clearly, all subsequences of a bad sequence are bad. (A bad sequence is by its very nature very bad!) By (an easy instance of) Ramsey's Theorem, if a sequence s is very good, then it has a subsequence that is a Q-chain:

$$s \models \Diamond \langle\langle Q \rangle\rangle^{\infty}$$

A quasi-order Q is a *well-quasi-order (wqo)* if

$$\Omega \models \Diamond \, \langle\langle Q \rangle\rangle \, \Omega \tag{1}$$

meaning that every infinite sequence is good. It goes without saying that this implies

$$\Omega \models (\Diamond \, \langle\langle Q \rangle\rangle)^{\infty}$$

But, as we just saw, it is actually equivalent to the stronger

$$\Omega \models \Diamond \langle\langle Q \rangle\rangle^{\infty} \tag{1$'$}$$

meaning that infinite sequences are all very good and hence possess infinite chains.

Let $A_1, A_2, \ldots, A_n$ be a fixed finite number ($n \geq 0$) of binary relations over Σ. (For terms, say, A_i might give the ith immediate subterm.) Let $A = \bigcup_i A_i$ denote their union, and A^- its inverse. Let $\bar{A} \subseteq \Sigma \times \Sigma^n$ denote the set of pairs $(a, \langle a_1, \ldots, a_n \rangle)$, such that $a A_i a_i$ for all i, and let $\bar{A}^-$ be its inverse. Call a point that has an A-successor (that is, a neighbor at the head of an outgoing "arrow") a *parent* and its successor, a *child*. So, the relation A gives any child a_i of a parent $a \in \Sigma$, while $\bar{A}$ gives the tuple $\langle a_1, \ldots, a_n \rangle$ of all its children. Finally, let

$$\bar{Q} = \underbrace{Q \times \cdots \times Q}_{n \text{ times}}$$

denote the n-fold direct (Cartesian) product of Q, which is the componentwise comparison of n elements in Q.

Suppose A is well-founded:

$$A^\infty = \varnothing \tag{2}$$

Suppose further all the following:

$$Q \circ A^- \subseteq Q \tag{3}$$
$$\bar{A} \circ \bar{Q} \circ \bar{A}^- \subseteq Q \tag{4}$$
$$\Omega \models \Diamond [\![A\Pi]\!] \Omega + \Diamond \langle\!\langle Q \rangle\!\rangle \Omega \tag{5}$$

Supposition (3) means that what is smaller than a child is smaller than its parent. It is satisfied, as a special case, whenever $A^- \subseteq Q$ as sets of pairs, since Q is transitive. Supposition (4) means that two parents are ordered if *all* their children are. The last supposition (5) means that every infinite sequence includes a parent or else is good. By Ramsey's Theorem, again, this is equivalent to

$$\Omega \models \Diamond \langle\!\langle A\Pi \rangle\!\rangle^\infty + \Diamond \langle\!\langle Q \rangle\!\rangle^\infty \tag{5$'$}$$

That is, every infinite sequence includes infinitely many parents, or else is very good.

It turns out that Q must be a well-quasi-ordering of Σ if all four suppositions (2–5) hold. But that is another story.

Acknowledgements

Thank you Ori Brost and Bernhard Gramlich for your suggestions.