Representing real numbers in a generalized numeration system

Émilie Charlier[1] Marion Le Gonidec[2] Michel Rigo[1]

[1]Université de Liège

[2]Université du Sud Toulon Var

Numeration, Leiden 2010, June 7

By enumerating the words of a language L with respect to some order, we define a numeration system.

For $L = \{w \in \{a,b\}^* \mid w \text{ does not contain } aa\}$ and the radix order induced by $a < b$, the first few representations are

ε	0	aba	6
a	1	abb	7
b	2	bab	8
ab	3	bba	9
ba	4	bbb	10
bb	5	$abab$	11

An abstract numeration system (ANS) is given by a triplet $S = (L, \Sigma, <)$ where L is a regular language over a totally ordered alphabet $(\Sigma, <)$.

By enumerating the words in L with respect to the radix order induced by $<$, we define a one-to-one correspondence:

$$\mathrm{rep}_S : \mathbb{N} \to L \qquad \mathrm{val}_S = \mathrm{rep}_S^{-1} : L \to \mathbb{N}.$$

Most numeration systems satisfy the relation:

$$\begin{array}{cl} m < n & \text{(usual order on the naturals)} \\ \Updownarrow & \\ rep(m) < rep(n) & \text{(radix order)} \end{array}$$

Examples:

- Binary numeration system: $\mathcal{L}_2 = \{\varepsilon\} \cup 1\{0,1\}^*$ and $0 < 1$
- Fibonacci numeration system: $L = \{\varepsilon\} \cup 1\{0,01\}^*$ and $0 < 1$
- Non-standard numeration systems
- Rational base number systems

QUESTION: How to represent real numbers in an ANS ?

The decimal representation of $\frac{11}{13}$ is $0.(846153)^\omega$:

$$\frac{8}{10}, \frac{84}{100}, \frac{846}{1000}, \frac{8461}{10000}, \frac{84615}{100000}, \dots$$

$$n\text{-th fraction} = \frac{\mathrm{val}_{10}(\text{prefix of length } n \text{ of } (846153)^\omega)}{10^n}$$

$$\forall L \subseteq \Sigma^*,\ \mathbf{u}_L(n) = \mathrm{Card}(L \cap \Sigma^n);$$

$$\mathbf{v}_L(n) = \mathrm{Card}(L \cap \Sigma^{\leq n}) = \sum_{i=0}^{n} \mathbf{u}_L(i).$$

For the integer base $b \geq 2$:

$$\mathcal{L}_b = \{\varepsilon\} \cup \{1, \ldots, b-1\}\{0, \ldots, b-1\}^*$$

$$\mathbf{v}_{\mathcal{L}_b}(n) = \sum_{i=0}^{n} \mathbf{u}_{\mathcal{L}_b}(i) = b^n.$$

The decimal representation of $\frac{11}{13}$ is $0.(846153)^\omega$:

$$\frac{8}{10}, \frac{84}{100}, \frac{846}{1000}, \frac{8461}{10000}, \frac{84615}{100000}, \ldots$$

$$n\text{-th fraction} = \frac{\mathsf{val}_{10}(\text{prefix of length } n \text{ of } (846153)^\omega)}{\mathbf{v}_{\mathcal{L}_{10}}(n)}$$

The binary representation of $\frac{11}{13}$ is $0.(110110001001)^\omega$:

$$\frac{1}{2}, \frac{3}{4} = \frac{6}{8}, \frac{13}{16}, \frac{27}{32} = \frac{54}{64} = \frac{108}{128} = \frac{216}{256}, \frac{433}{512} = 0.845703125, \ldots$$

$$n\text{-th fraction} \ = \frac{\mathsf{val}_2(\text{prefix of length } n \text{ of } (110110001001)^\omega)}{\mathbf{v}_{\mathcal{L}_2}(n)}$$

7-th fraction: $108 = 64 + 32 + 8 + 4 = \mathsf{val}_2(1101100)$
$128 = 2^7 = \mathbf{v}_{\mathcal{L}_2}(7)$.

- $S = (L, \Sigma, <)$
- $w \in \Sigma^\omega$
- $(w^{(n)})_{n\geq 0} \in L^{\mathbb{N}}, \qquad w^{(n)} \to w$ as $n \to +\infty$

POINT: To show that, under certain hypotheses, the limit

$$\lim_{n\to+\infty} \frac{\mathrm{val}_S(w^{(n)})}{\mathbf{v}_L(|w^{(n)}|)}$$ exists and only depends on w.

In that case, w is an S-representation of the corresponding real.

QUESTION: And when L is not regular?

Example

The $\frac{3}{2}$-number system introduced by Akiyama, Frougny and Sakarovitch (2008) has a numeration language which is not context-free.

AIM: To provide a unified approach for representing real numbers

- Arbitrary infinite language L (not necessarily regular)
- Minimal automaton of L: $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$
- "Generalized" ANS: $S = (L, \Sigma, <)$

For all $x \in L$, the numerical value $\mathrm{val}_S(x)$ of x is given by

$$\mathbf{v}_L(|x|-1) + \sum_{i=0}^{|x|-1} \sum_{a<x[i]} \mathbf{u}_{L_{\delta(q_0, x[0,i-1]a)}}(|x|-i-1),$$

where $x[0, i-1]$ = prefix of length i of x
and L_q = language accepted from q in $\mathcal{A}$.

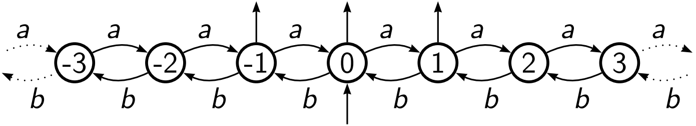
- w = limit of words in $L \Leftrightarrow \mathrm{Pref}(w) \subseteq \mathrm{Pref}(L)$
  $\Leftrightarrow w \in \mathrm{Adh}(L)$

Remark

Since $\mathrm{Adh}(L) = \mathrm{Adh}(\mathrm{Pref}(L))$, *there is no new representation if we assume that* L *is prefix-closed.*

Example: $L = \{w \in \{a,b\}^* \mid ||w|_a - |w|_b| \leq 1\}$

$= \{\varepsilon, a, b, ab, ba, aab, aba, abb, baa, bab, bba, aabb, \ldots\}$



For $S = (L, \{a,b\}, a < b)$, we can compute

$$\lim_{n\to+\infty} \frac{\mathrm{val}_S((ab)^n)}{\mathbf{v}_L(2n)} = \frac{3}{4} \quad \text{and} \quad \lim_{n\to+\infty} \frac{\mathrm{val}_S((ab)^n a)}{\mathbf{v}_L(2n+1)} = \frac{3}{5}$$

which shows that $\lim_{n\to+\infty} \frac{\mathrm{val}_S((ab)^\omega[0,n-1])}{\mathbf{v}_L(n)}$ does not exist.

L not prefix-closed: $\mathrm{Pref}(L) = \{a,b\}^*$

Hypotheses needed?

- (H1) L is prefix-closed
- (H2) $\mathrm{Adh}(L)$ is uncountable

QUESTION: What conditions must L satisfy so that

the limits $\lim_{n \to +\infty} \frac{\mathrm{val}_S(w[0, n-1])}{\mathbf{v}_L(n)}$ exist for all $w \in \mathrm{Adh}(L)$?

AIM: Define some approximation intervals of reals.

Their length should decrease as the prefix that is read becomes larger and larger.

$$\forall x \in L \cap \Sigma^n,\ \underbrace{\frac{\mathbf{v}_L(n-1)}{\mathbf{v}_L(n)}}_{=1-\frac{\mathbf{u}_L(n)}{\mathbf{v}_L(n)}} \leq \frac{\mathsf{val}_S(x)}{\mathbf{v}_L(n)} \leq \underbrace{\frac{\mathbf{v}_L(n)}{\mathbf{v}_L(n)}}_{=1}$$

If $\lim_{n\to+\infty} \frac{\mathbf{u}_L(n)}{\mathbf{v}_L(n)}$ exists , then it is denoted by r_ε and the represented interval is $I_\varepsilon = [1 - r_\varepsilon, 1]$.

Hypotheses needed?

Recall that, for all $x \in L$,

$$\mathrm{val}_S(x) = \mathbf{v}_L(|x|-1) + \sum_{i=0}^{|x|-1} \sum_{a<x[i]} \mathbf{u}_{L_{\delta(q_0, x[0,i-1]a)}}(|x|-i-1)$$

- (H3) $\forall x \in \Sigma^*,\ \exists r_x \geq 0,\ \lim_{n\to+\infty} \frac{\mathbf{u}_{L_{\delta(q_0,x)}}(n-|x|)}{\mathbf{v}_L(n)} = r_x$

Hypotheses needed?

In general, $|I_x| = r_x$

- (H4) $\forall w \in \mathrm{Adh}(L)$, $\lim_{\ell \to +\infty} r_{w[0,\ell-1]} = 0$

Remark

Let $\mathrm{Center}(L) = \mathrm{Pref}(\mathrm{Adh}(L))$. *Then* $x \notin \mathrm{Center}(L) \Leftrightarrow r_x = 0$.

Theorem (C.-Le G.-R. 2010)

The limits $\lim_{n \to +\infty} \frac{\mathrm{val}_S(w[0, n-1])}{\mathbf{v}_L(n)}$ *exist when* L *satisfies the following conditions:*

- (H1) L *prefix-closed*
- (H2) $\mathrm{Adh}(L)$ *uncountable*
- (H3) $\forall x \in \Sigma^*, \ \exists r_x \geq 0, \lim_{n \to +\infty} \frac{\mathbf{u}_{L_{\delta(q_0,x)}}(n-|x|)}{\mathbf{v}_L(n)} = r_x$
- (H4) $\forall w \in \mathrm{Adh}(L), \lim_{\ell \to +\infty} r_{w[0,\ell-1]} = 0$

For all $w \in \mathrm{Adh}(L)$, $\mathrm{val}_S(w) = \lim_{n \to +\infty} \frac{\mathrm{val}_S(w[0, n-1])}{\mathbf{v}_L(n)}$

is the numerical value of w.

The infinite word w is an S-representation of $\mathrm{val}_S(w)$.

Proposition (C.-Le G.-R. 2010)

Let $L \subseteq \Sigma^*$, $S = (\mathrm{Pref}(L), \Sigma, <)$ *be a (generalized) ANS.*
If $\mathrm{Pref}(L)$ *satisfies (H1), (H2) and (H3), then for all sequences* $(w^{(n)})_{n\geq 0} \in L^{\mathbb{N}}$ *converging to a word* $w \in \mathrm{Adh}(L)$, *we have*

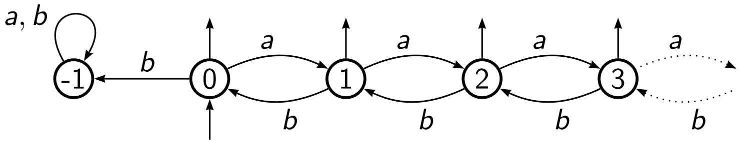
$$\lim_{n\to+\infty} \frac{\mathrm{val}_S(w^{(n)})}{\mathbf{v}_{\mathrm{Pref}(L)}(|w^{(n)}|)} = \mathrm{val}_S(w).$$

Example: Prefixes of Dyck words

$D = \{w \in \{a,b\}^* \mid |w|_a = |w|_b \text{ and } \forall u \in \mathrm{Pref}(w),\ |u|_a \geq |u|_b\}$

not prefix-closed $\longrightarrow$ we consider $S = (\mathrm{Pref}(D), \{a,b\}, a < b)$

$$\begin{aligned}\mathrm{Pref}(D) &= \{w \in \{a,b\}^* \mid \forall u \in \mathrm{Pref}(w),\ |u|_a \geq |u|_b\} \\ &= \{\varepsilon, a, aa, ab, aaa, aab, aba, aaaa, aaab, aaba, aabb, \ldots\}.\end{aligned}$$



Example: Prefixes of Dyck words (continued)

$$\lim_{n\to+\infty} \frac{\mathrm{val}_S((aab)^\omega[0,n-1])}{\mathbf{v}_L(n)} = \frac{39}{49} = 0.795918\ldots$$

x	$\mathrm{val}_S(x)$	$\mathbf{v}_L(\|x\|)$	$\frac{\mathrm{val}_S(x)}{\mathbf{v}_L(\|x\|)}$
a	1	2	0.50000
aa	2	4	0.50000
aab	5	7	0.71429
$aaba$	9	13	0.69231
$aabaa$	17	23	0.73913
$aabaab$	32	43	0.74419
$aabaaba$	60	78	0.76923
$aabaabaa$	112	148	0.75676
$aabaabaab$	213	274	0.77737
$aabaabaaba$	404	526	0.76806
$aabaabaabaa$	771	988	0.78036

Since $\lim_{n\to+\infty} \frac{\mathbf{v}_L(n-1)}{\mathbf{v}_L(n)} = \frac{1}{2}$, we represent the interval $I_\varepsilon = [\frac{1}{2}, 1]$

Center(Pref(D)) = Pref(D):

- $I_a = [1/2, 1]$
- $I_{aa} = [1/2, 7/8]$ $I_{ab} = [7/8, 1]$
- $I_{aaa} = [1/2, 3/4]$ $I_{aab} = [3/4, 7/8]$ $I_{aba} = [7/8, 1]$
- ...

Example: Prefixes of Dyck words (continued)

$\forall x \in [\frac{1}{2}, 1]$, Q_x designates the set of representations of x.

We have $Q_{1/2} = \{a^\omega\}$ and $Q_1 = \{(ab)^\omega\}$.

If $x \in]1/2, 1[$ and $x = \sup I_w = \inf I_z$ then $Q_x = \{\bar{w}(ab)^\omega, za^\omega\}$,
where $\bar{w}$ = the least Dyck word having w as a prefix.

Proposition (C.-Le G.-R. 2010)

If L is context-free, then the representations of the endpoints of the intervals are ultimately periodic.

- Characterize the automata recognizing a language L such that the corresponding ω-language $\mathrm{Adh}(L)$ is uncountable.

Open problems

Theorem (Boasson-Nivat 1980)

For every context-free language L, there exists a sequential mapping f such that $f(\mathrm{Adh}(D)) = \mathrm{Adh}(L)$, where D is the Dyck language.

- Let S and T be abstract numeration systems built respectively on $\mathrm{Pref}(D)$ and $\mathrm{Pref}(L)$. Give a mapping g such that the following diagram commutes.

$$\begin{array}{ccc} \mathrm{Adh}(D) & \xrightarrow{\ f\ } & \mathrm{Adh}(L) \\ {\scriptstyle \mathrm{val}_S}\downarrow & & \downarrow{\scriptstyle \mathrm{val}_T} \\ [s_0, 1] & \xrightarrow[\ g\ ]{} & [t_0, 1] \end{array}$$