Representing real numbers in a generalized numeration system

Émilie Charlier\textsuperscript{1}  Marion Le Gonidec\textsuperscript{2}  Michel Rigo\textsuperscript{1}

\textsuperscript{1}Université de Liège

\textsuperscript{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

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon$</td>
<td>0</td>
<td>$aba$</td>
<td>6</td>
</tr>
<tr>
<td>$a$</td>
<td>1</td>
<td>$abb$</td>
<td>7</td>
</tr>
<tr>
<td>$b$</td>
<td>2</td>
<td>$bab$</td>
<td>8</td>
</tr>
<tr>
<td>$ab$</td>
<td>3</td>
<td>$bba$</td>
<td>9</td>
</tr>
<tr>
<td>$ba$</td>
<td>4</td>
<td>$bbb$</td>
<td>10</td>
</tr>
<tr>
<td>$bb$</td>
<td>5</td>
<td>$abab$</td>
<td>11</td>
</tr>
</tbody>
</table>
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:

$$\text{rep}_S : \mathbb{N} \rightarrow L \quad \text{and} \quad \text{val}_S = \text{rep}_S^{-1} : L \rightarrow \mathbb{N}.$$
Most numeration systems satisfy the relation:

\[ m < n \quad \text{(usual order on the naturals)} \]

\[ \iff \]

\[ \text{rep}(m) < \text{rep}(n) \quad \text{(radix order)} \]

Examples:

- **Binary numeration system:** \( L_2 = \{\varepsilon\} \cup 1\{0, 1\}^* \) 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 $0.(846153)\omega$:

\[
\begin{align*}
8 & \quad 84 & \quad 846 & \quad 8461 & \quad 84615 \\
\frac{1}{{10}'} & \quad \frac{4}{{100}'} & \quad \frac{46}{{1000}'} & \quad \frac{61}{{10000}'} & \quad \frac{15}{{100000}'} \\
\end{align*}
\cdots
\]

\[n\text{-th fraction } = \frac{\text{val}_{10}(\text{prefix of length } n \text{ of } (846153)^\omega)}{10^n}\]
∀ L ⊆ Σ*, \( u_L(n) = \text{Card}(L \cap \Sigma^n) \);

\[ v_L(n) = \text{Card}(L \cap \Sigma^{\leq n}) = \sum_{i=0}^{n} u_L(i). \]

For the integer base \( b \geq 2 \):

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

\[ v_{\mathcal{L}_b}(n) = \sum_{i=0}^{n} u_{\mathcal{L}_b}(i) = b^n. \]
The decimal representation of $\frac{11}{13}$ is $0.(846153)^\omega$:

\[
\begin{array}{cccccc}
8 & 84 & 846 & 8461 & 84615 \\
10' & 100' & 1000' & 10000' & 100000' \\
\end{array}
\ldots
\]

\[n\text{-th fraction } = \frac{\text{val}_{10}(\text{prefix of length } n \text{ of } (846153)^\omega)}{v_{\mathcal{L}_{10}}(n)}\]
The binary representation of \( \frac{11}{13} \) is \(0.(110110001001)_{\omega} \):

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

\[n\text{-th fraction} = \frac{\text{val}_2(\text{prefix of length } n \text{ of } (110110001001)_{\omega})}{v_{L_2}(n)}\]

7-th fraction: \(108 = 64 + 32 + 8 + 4 = \text{val}_2(1101100)\)

\(128 = 2^7 = v_{L_2}(7)\).
\[
S = (L, \Sigma, <)
\]
\[
w \in \Sigma^\omega
\]
\[
(w^{(n)})_{n \geq 0} \in L^\mathbb{N}, \quad w^{(n)} \to w \text{ as } n \to +\infty
\]

**POINT:** To show that, under certain hypotheses, the limit
\[
\lim_{n \to +\infty} \frac{\text{val}_S(w^{(n)})}{v_L(|w^{(n)}|)}
\]
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
Generalization to non-regular languages

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

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

$$\text{val}_L(|x| - 1) + \sum_{i=0}^{\lfloor \frac{|x|}{2} \rfloor} \sum_{a \prec x[i]} u_{L_\delta(q_0, x[0,i-1]a)}(|x| - i - 1),$$

where $x[0, i - 1] = \text{prefix of length } i \text{ of } x$ and $L_q = \text{language accepted from } q \text{ in } A$. 
\[ w = \text{limit of words in } L \iff \text{Pref}(w) \subseteq \text{Pref}(L) \iff w \in \text{Adh}(L) \]

**Remark**

Since \( \text{Adh}(L) = \text{Adh}(	ext{Pref}(L)) \), 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

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

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

$L$ not prefix-closed: $\text{Pref}(L) = \{a, b\}^*$
Hypotheses needed?

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

**QUESTION:** What conditions must \(L\) satisfy so that the limits \(\lim_{n \to +\infty} \frac{\text{val}_S(w[0, n-1])}{v_L(n)}\) exist for all \(w \in \text{Adh}(L)\)?
Hypotheses needed?

**AIM**: Define some approximation intervals of reals.

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

\[
\forall x \in L \cap \Sigma^n, \quad \frac{v_L(n - 1)}{v_L(n)} \leq \frac{\text{val}_S(x)}{v_L(n)} \leq \frac{v_L(n)}{v_L(n)} = 1 - \frac{u_L(n)}{v_L(n)} = 1
\]

If \( \lim_{n \to +\infty} \frac{u_L(n)}{v_L(n)} \) exists, then it is denoted by \( r_\varepsilon \) and the represented interval is \( l_\varepsilon = [1 - r_\varepsilon, 1] \).
Recall that, for all $x \in L$,

$$\text{val}_S(x) = v_L(|x| - 1) + \sum_{i=0}^{|x|-1} \sum_{a<x[i]} u_{L_{\delta(q_0,x[0,i-1])}}(|x| - i - 1)$$

$\blacktriangleright$ (H3) $\forall x \in \Sigma^*, \exists r_x \geq 0, \lim_{n \to +\infty} \frac{u_{L_{\delta(q_0,x)}}(n-|x|)}{v_L(n)} = r_x$
Hypotheses needed?

In general, $|l_x| = r_x$

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

**Remark**

Let $\text{Center}(L) = \text{Pref}(\text{Adh}(L))$. Then $x \not\in \text{Center}(L) \iff r_x = 0$. 
The limits $\lim_{n \to +\infty} \frac{\text{val}_S(w[0, n-1])}{v_L(n)}$ exist when $L$ satisfies the following conditions:

- (H1) $L$ prefix-closed
- (H2) $\text{Adh}(L)$ uncountable
- (H3) $\forall x \in \Sigma^*, \exists r_x \geq 0, \lim_{n \to +\infty} \frac{u_L(q_0, x)(n-|x|)}{v_L(n)} = r_x$
- (H4) $\forall w \in \text{Adh}(L), \lim_{\ell \to +\infty} r_w[0, \ell-1] = 0$
For all $w \in \text{Adh}(L)$, \( \text{val}_S(w) = \lim_{n \to +\infty} \frac{\text{val}_S(w[0, n - 1])}{v_L(n)} \)

is the numerical value of \( w \).

The infinite word \( w \) is an \( S \)-representation of \( \text{val}_S(w) \).
**Proposition (C.-Le G.-R. 2010)**

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

$$\lim_{n \to +\infty} \frac{\text{val}_S(w^{(n)})}{\mathbf{v}_{\text{Pref}(L)}(|w^{(n)}|)} = \text{val}_S(w).$$
Example: Prefixes of Dyck words

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

not prefix-closed \( \rightarrow \) we consider \( S = (\text{Pref}(D), \{ a, b \}, a < b) \)

\[ \text{Pref}(D) = \{ w \in \{ a, b \}^* \mid \forall u \in \text{Pref}(w), \ |u|_a \geq |u|_b \} \]
\[ = \{ \varepsilon, a, aa, ab, aaa, aab, aba, aaaa, aaab, aaba, aabb, \ldots \}. \]
Example: Prefixes of Dyck words (continued)

\[
\lim_{n \to +\infty} \frac{\text{val}_S((aab)^{\omega}[0, n-1])}{v_L(n)} = \frac{39}{49} = 0.795918 \ldots
\]

| x          | val$_S$(x) | v$_L$(|x|) | val$_S$(x) / 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                |
| aabaabaaba | 771        | 988      | 0.78036                |
Example: Prefixes of Dyck words (continued)

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

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

- \( I_a = [1/2, 1] \)
- \( I_{aa} = [1/2, 7/8] \)  \( I_{ab} = [7/8, 1] \)
- \( I_{aaa} = [1/2, 3/4] \)  \( I_{aab} = [3/4, 7/8] \)  \( I_{aba} = [7/8, 1] \)
- . . .
\[ \forall x \in [\frac{1}{2}, 1], \ Q_x \text{ designates the set of representations of } x. \]

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.
Open problems

- Characterize the automata recognizing a language $L$ such that the corresponding $\omega$-language $\text{Adh}(L)$ is uncountable.
Open problems

**Theorem (Boasson-Nivat 1980)**

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

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

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