Appendix: Decadic Notation

There is a simple and natural one-to-one correspondence between the set of all strings over a finite alphabet and the set of natural numbers. The key to the correspondence is to use base-n notation, where n is the size of the alphabet, but without a 0 digit.

Suppose that Σ is a finite set of size n. We refer to Σ as the *alphabet*, we refer to the members of Σ as *letters*, and we refer to finite strings of letters as *words* (over Σ). Let Σ^* be the infinite set of all words (including the empty word λ). For example, if $\Sigma = \{1, 2\}$, then

$$\Sigma^* = \{\lambda, 1, 2, 11, 12, 21, 22, 111, 112, 121, \ldots\}.$$

We assume that the members of the alphabet Σ are ordered in some way (referred to as *alphabetic order*). Thus we have a function $v : \Sigma \to \{1, \ldots, n\}$ where

$$\begin{aligned} v(\text{the first letter}) &= 1 \\ v(\text{the second letter}) &= 2 \\ &\cdots \\ v(\text{the last letter}) &= n \end{aligned}$$

and we refer to $v(a)$ as the *value* of the letter a.

We obtain a one-to-one map from Σ^* onto the set $\mathbb{N} = \{0, 1, 2, \ldots\}$ by mapping the three-letter word *abc* to the number

$$(abc)_{n\text{-adic}} = v(a)n^2 + v(b)n + v(c)$$

and in general,

$$(a_k \cdots a_2 a_1)_{n\text{-adic}} = v(a_k)n^{k-1} + \cdots + v(a_2)n + v(a_1)$$

and $(\lambda)_{n\text{-adic}} = 0$ (the empty sum).

In what follows, the properties of this map will be developed in the special case where $n = 10$ and the ten letters are $\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}, \mathbf{7}, \mathbf{8}, \mathbf{9}, \boldsymbol{X}$, in that order. But all of the arguments generalize immediately to alphabets of any size 2 or more.

(An alphabet of size 1 is a somewhat special and somewhat boring case. Where the alphabet is the singleton $\{|\}$, a typical word is the string $|||||||$, and the above equation reduces to

$$(a_k \cdots a_2 a_1)_{1\text{-adic}} = 1 + \cdots + 1 + 1 = k$$

so that $(w)_{1\text{-adic}}$ equals the length of the word w. This obviously produces a one-to-one map onto $\mathbb{N}$.)

In the $n = 10$ case, we will refer to $(w)_{10\text{-adic}}$ as $(w)_{\text{decadic}}$, the number denoted by the word w in *decadic* notation. (The corresponding words for $n = 1, 2, 3, \ldots$ would be monadic, dyadic, triadic, ...) For example,

$$(\mathbf{415})_{\text{decadic}} = 4 \cdot 100 + 1 \cdot 10 + 5 = 415$$

and

$$(\boldsymbol{4X5})_{\text{decadic}} = 4 \cdot 100 + 10 \cdot 10 + 5 = 505.$$

As the first of these equations exemplifies, for any word w not containing the $\boldsymbol{X}$ digit, $(w)_{\text{decadic}}$ is simply the number denoted by the numeral w in the usual base-10 notation. But the alphabet contains no zero digit, and some words contain the $\boldsymbol{X}$ digit (the "ten" digit). For example,

$$(\boldsymbol{XX})_{\text{decadic}} = 10 \cdot 10 + 10 = 110$$

and

$$(\boldsymbol{XXX})_{\text{decadic}} = 10 \cdot 100 + 10 \cdot 10 + 10 = 1,110.$$

First consider the problem of how to add 1 in decadic notation. Here are some examples:

$$\begin{array}{rcl} (\boldsymbol{3X8})_{\text{decadic}} + 1 & = & (\boldsymbol{3X9})_{\text{decadic}} \\ (\boldsymbol{XXX})_{\text{decadic}} + 1 & = & (\boldsymbol{1111})_{\text{decadic}} \\ (\boldsymbol{2XXX})_{\text{decadic}} + 1 & = & (\boldsymbol{3111})_{\text{decadic}} \end{array}$$

In general, a word w consists of a word u (possibly empty) not ending in $\boldsymbol{X}$, followed by a string of k $\boldsymbol{X}$'s (where $k \geq 0$). Thus

$$(w)_{\text{decadic}} = (u)_{\text{decadic}} \cdot 10^k + 10^k + 10^{k-1} + \cdots + 10.$$

In order to add 1, we use the word w^+ consisting of u^+ followed by a string of k $\mathbf{1}$'s, where u^+ is obtained by incrementing u's rightmost digit (and $\lambda^+ = \mathbf{1}$). This works because

$$\begin{array}{rcl} (w^+)_{\text{decadic}} & = & ((u)_{\text{decadic}} + 1) \cdot 10^k + 10^{k-1} + \cdots + 10 + 1 \\ & = & (u)_{\text{decadic}} \cdot 10^k + 10^k + 10^{k-1} + \cdots + 10 + 1 \\ & = & (w)_{\text{decadic}} + 1. \end{array}$$

Now that we know how to add 1, we obtain a theorem:

Theorem. For every natural number n, we can find a word w for which $(w)_{\text{decadic}} = n$.

Proof. We prove the existence of w by induction on n. The basis, $n = 0$, holds because $(\lambda)_{\text{decadic}} = 0$. For the inductive step, we add 1 as above.

And not only does w exist, but we know how to calculate it by adding 1 many times (n times). (Of course, there are much faster ways to find w, if we are in a hurry.) $\dashv$

Next consider the problem of comparing decadic numerals to see which one denotes a larger number. Among 4-digit numbers (i.e., among the numbers denoted by 4-digit words), the smallest is obviously $(\mathbf{1111})_{\text{decadic}}$. Any other 4-digit word will give us a larger number. Similarly, the largest number denoted by a 3-digit word is $(\boldsymbol{XXX})_{\text{decadic}}$. Any other 3-digit word will give us a smaller number. Moreover,

$$(\boldsymbol{XXX})_{\text{decadic}} < (\mathbf{1111})_{\text{decadic}}$$

because adding 1 to the left side gives the right side.

We can conclude from this that any number m denoted by a 3-digit word is less than any number denoted n by a 4-digit word:

$$m \leq (\boldsymbol{XXX})_{\text{decadic}} < (\mathbf{1111})_{\text{decadic}} \leq n$$

And the argument generalizes: Any number denoted by a k-digit word is less than any number denoted by a $(k+1)$-digit word. And then by iterating this argument, we see that any number denoted by a k-digit word is less than any number denoted by word of more than k digits. That is, shorter words always give smaller numbers.

What then about two words of the same length? For example, take the four-letter words $\mathbf{2}{*}{*}{*}$ and $\mathbf{3}\star\star\star$, where in both cases we don't know the last three digits. Even so, we know that

$$\begin{array}{rcll} \mathbf{2}{*}{*}{*} & \leq & \mathbf{2}\boldsymbol{XXX} & \\ & < & \mathbf{3111} & \text{by 1} \\ & \leq & \mathbf{3}\star\star\star. & \end{array}$$

Similary, for the six-letter words $\mathbf{472}{*}{*}{*}$ and $\mathbf{473}\star\star\star$, we have

$$\begin{array}{rcll} \mathbf{472}{*}{*}{*} & \leq & \mathbf{472}\boldsymbol{XXX} & \\ & < & \mathbf{473111} & \text{by 1} \\ & \leq & \mathbf{473}\star\star\star. & \end{array}$$

In general, we need to look only at the first (i.e., leftmost) digit where two words of the same length disagree. The larger digit produces a larger number, no matter what the later digits are. That is, for words of the same length, we simply use lexicographic order. We can summarize these ideas as follows:

Theorem. (a) For two words of different lengths, the shorter word denotes a smaller number than does the longer word.

(b) For two words u and w of the same length, $(u)_{\text{decadic}} < (w)_{\text{decadic}}$ iff u lexicographically precedes w.

In particular, two *different* words must denote different numbers, because either they will have different lengths (and clause (a) of the theorem will apply) or else there will be a first digit where they differ (and clause (b) will apply). That is, our map from words to numbers is one-to-one. Together with the previous theorem, we now have the following:

Theorem. Decadic notation yields a one-to-one correspondence between the set of all words over our 10-letter alphabet and the set $\mathbb{N}$ of all natural numbers.

This theorem illustrates a property of decadic notation which standard decimal notation lacks. In decimal notation, there is the problem of "leading zeros"; the words **3** and **03** denote the same number (or else **03** needs to be declared an illegal word).

Roman numerals are sometimes criticized for lacking a numeral for zero. But the real difficulty with Roman numerals is the lack of place-value notation. Zero itself is nothing.