Math 295. Handout on Integer Digit Expansions

The purpose of this handout is to prove the existence of a good theory of "digit expansions" for natural numbers. We first record a mild strenthening of the division algorithm from class.

Lemma 0.1. *Let $a, b \in \mathbb{Z}$, $b \neq 0$. There exist unique integers q and r with $0 \leq r < |b|$ such that $a = bq + r$. Moreover, when $a \geq 0$, $b > 0$ then $q \geq 0$ and when $a \geq b > 1$ then $1 \leq q < a$.*

Proof. In class we proved the existence of such q and r. We did not proof the uniqueness, so we first settle that point. Suppose $bq + r = bq' + r'$ with $q, q', r, r' \in \mathbb{Z}$ and $0 \leq r, r' < |b|$. We want to prove that $q = q'$ and $r = r'$. Without loss of generality, we may suppose $r \leq r'$. Since $bq + r = bq' + r'$, we have $b(q - q') = r' - r$. Taking absolute value of both sides,

$$|b||q - q'| = |r' - r| = r' - r \leq r' < |b|.$$

But $|b| \neq 0$, so $|q - q'| < 1$. But since $|q - q'| \in \mathbb{Z}$ is a non-negative integer less than 1, it must be 0! Hence, $q = q'$ and so $r' - r = b(q - q') = 0$ so $r' = r$ also. This establishes the desired uniqueness.

When $a \geq 0$ and $b > 0$ then $q \geq 0$ because otherwise $q < 0$ and so $q \leq -1$, which forces

$$a = bq + r < b(-1) + r = r - b < 0,$$

a contradiction (recall $0 \leq r < |b| = b$). Finally, when $a \geq b > 1$ then we want to show $1 \leq q < a$. We know that at least $q \geq 0$. Certainly $q \neq 0$, for otherwise $a = r < b$, contrary to hypothesis. Thus, $q > 0$, so $q \geq 1$. Meanwhile, we must have $q < a$ for otherwise $q \geq a$ and hence

$$a = bq + r \geq ba + r \geq ba > a$$

since $b > 1$, and this is absurd.

■

Lemma 0.2. *If $b > 1$ and $n, m \in \mathbb{N}$ with $n > m$, then $b^n > b^m \geq b$.*

Proof. Once we show $b^r \geq b$ for all $r \in \mathbb{N}$, then $b^n = b^{n-m}b^m \geq b \cdot b^m > b^m$, so we're done. In order to prove $b^r \geq b$ for all $r \in \mathbb{N}$, we induct on r. The case $r = 1$ is clear and $b^{r+1} = b \cdot b^r \geq b$ if $b^r \geq b > 1$.

■

Now we are ready for the main result. Fix $b \in \mathbb{N}$ with $b > 1$. This is going to be our "base" (e.g., $b = 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1$). Let $D = \{k \in \mathbb{N} \mid k < b\}$. This is going to be our "digit set". Note that $0, 1 \in D$. We use the choice $b = 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1$ for biological reasons (it happens to correspond to the number of fingers and thumbs which most people have), and for this choice of b it is also common to use the "Arabic notation" $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ for the elements of D arranged in increasing order. However, other societies have made other choices. For example, the ancient Babylonians, Sumerians, and Indians used the choice denoted $b = 60$ in Arabic notation (probably because this base is divisible by lots of small numbers). Computers use the mathematically "simplest" choice $b = 1 + 1$ (with $D = \{0, 1\}$).

Theorem 0.3. *Every $a \in \mathbb{N}$ can be written in the form*

$$a = \sum_{i=0}^{n} d_i b^i$$

for some $d_i \in D$ and some $n \geq 0$ with $d_n \neq 0$. Moreover, such n and d_i's are uniquely determined by a.

The element a is customarily written "to the base b" by writing $d_n d_{n-1} \ldots d_0$, with the d_i's and n as in the theorem. In this notation, $b = 1 \cdot b + 0$ is written as 10. Thus, the use of such notation for non-negative integers requires specifying what choice has been made for b. That is, if I don't tell you

what b is, then you can't determine what is meant by the notation "10" other than that it is the base for the digit system being used. For example, when working in base $1+1+1$ the symbol 10 is used to denote what you write as 3 in Arabic notation, while with base $1+1+1+1+1+1+1+1+1+1+1$ the symbol 10 is used to denote what you write as 11 in Arabic notation. Of course, if you choose $b \leq 1+1+1+1+1+1+1+1+1+1$ then it is customary to denote the elements of the digit set by the corresponding subset of $\{0,1,2,3,4,5,6,7,8,9\}$, while for other b's (e.g., in ancient India) it is necessary to use more notation due to the additional elements of the digit set.

One typically spends much of elementary school mathematics learning intricate algorithms for calculating sums, products, and differences in $\mathbb{Z}$ entirely in terms of Arabic notation. You should convince yourself now that these algorithms are nothing other than massive applications of the commutative and associative laws for addition and multiplication, together with the distributive law. The algorithm for "computing" the q and r from the division algorithm in terms of this "digit notation" is a somewhat more complicated example of the same ideas.

Proof. (of Theorem). We first establish existence by strong induction on a, assuming that every $k \in \mathbb{N}$ with $k < a$ can be written in the asserted form. We want to prove the same property for a, namely that it can be written in the asserted form (once this induction is settled, we return to address uniqueness). If $a < b$ then taking $n = 0$ and $d_0 = a$ does the job. Now suppose $a \geq b$. By Lemma 0.1 $a = bq + r$ with $q, r \in \mathbb{Z}$, $0 \leq r < b$, and $1 \leq q < a$. Strong induction therefore lets us write

$$q = \sum_{i=0}^{m} \delta_i b^i$$

with $\delta_i \in D$ and $\delta_m \neq 0$. Then

$$a = bq + r = \sum_{i=0}^{m+1} d_i b^i$$

with $d_0 = r \in D$ and $d_i = \delta_{i-1} \in D$ for $0 < i \leq m+1$ (with $d_{m+1} = \delta_m \neq 0$). This finishes the strong induction on a.

For uniqueness, we first consider the case $a < b$. If $a = \sum_{i=0}^{m} d_i b^i$ with $d_m \neq 0$ then $b > a \geq d_m b^m \geq b^m$, so $m = 0$ by Lemma 0.2. Then $a = d_0$, so uniqueness is established for such a. Now we argue by strong induction, assuming $a \geq b$ without loss of generality. If $a = \sum_{i=0}^{m} d_i b^i$ then necessarily $m > 0$ (as otherwise $a = d_0 < b$). Thus,

$$a = \sum_{i=0}^{m} d_i b^i = b\left(\sum_{i=0}^{m-1} d_{i+1} b^i\right) + d_0 = bq + r$$

with $r \stackrel{\text{def}}{=} d_0 \in D$ and $q \stackrel{\text{def}}{=} \sum_{i=0}^{m-1} \delta_i b^i$ where $\delta_i = d_{i+1} \in D$ and $\delta_{m-1} \neq 0$. The uniqueness in the division algorithm implies that this q and r are uniquely determined by a (i.e., these must be *exactly* the same "q" and "r" that arise in the division algorithm for our a and b). This implies the uniqueness of the "units' digit" d_0 and by Lemma 0.1 we have $1 \leq q < a$. Thus, strong induction on a gives uniqueness for the expression for q, from which we obtain that $m-1$ (and hence m) is uniquely determined by a, as are the d_i's for $1 \leq i \leq m$.

■