Representation of integers in base b

Integers are usually expressed by decimal notation. For instance 27182818 means

$$2 \cdot 10^7 + 7 \cdot 10^6 + 1 \cdot 10^5 + 8 \cdot 10^4 + 2 \cdot 10^3 + 8 \cdot 10^2 + 1 \cdot 10^1 + 8$$

The theorem that we prove below provides a rigorous justification for this, and other, representations of integers.

In this handout b is a positive integer such that $b > 1$. By $\mathbb{D}$ we denote the set $\{0, 1, \dots, b-1\}$. By $\mathbb{N}_0$ we denote the set $\{0\} \cup \mathbb{N}$.

Lemma 1. Let k be a nonnegative integer and let $d_0, \dots, d_k \in \mathbb{D}$. Then

$$d_k\, b^k + \cdots + d_1\, b + d_0 \ \leq \ b^{k+1} - 1. \tag{1}$$

Proof. In the following calculation we use that $d_j \leq b - 1$ for all $j = 0, \dots, k$:

$$\begin{aligned} b^k d_k + b^{k-1} d_{k-1} + \cdots + b\, d_1 + d_0 &\leq b^k (b-1) + b^{k-1}(b-1) + \cdots + b\,(b-1) + (b-1) \\ &= b^{k+1} - b^k + b^k - b^{k-1} + \cdots + b^2 - b + b - 1 \\ &= b^{k+1} - 1. \end{aligned}$$

This proves (1). □

Theorem 2. Let a be a positive integer. Then there exist unique nonnegative integer m and unique $d_0, \dots, d_m \in \mathbb{D}$ with $d_m > 0$ such that

$$a = d_m b^m + \cdots + d_1 b + d_0.$$

Proof. Set

$$S = \bigl\{ x \,:\, x \leq a,\ x = b^k \text{ with } k \in \mathbb{N}_0 \bigr\}. \tag{2}$$

Since $1 = b^0$ and $1 \leq a$, the set S is not empty. By definition (2), S is bounded above by a. Hence, the well ordering principle implies that S has the maximum. Set $y = \max S$ and let $m \in \mathbb{N}_0$ be such that $y = b^m$. Notice that the definition of m implies that

$$b^m \leq a < b^{m+1}. \tag{3}$$

Now set $q_0 = a$ and apply the division algorithm (dividing with b) exactly $m + 1$ times:

$$\begin{aligned} a = q_0 &= b\, q_1 + d_0, & &\text{where} & q_1 &\in \mathbb{Z}, & d_0 &\in \mathbb{D}, \\ q_1 &= b\, q_2 + d_1, & &\text{where} & q_2 &\in \mathbb{Z}, & d_1 &\in \mathbb{D}, \\ q_2 &= b\, q_3 + d_2, & &\text{where} & q_3 &\in \mathbb{Z}, & d_2 &\in \mathbb{D}, \\ &\vdots \\ q_{m-1} &= b\, q_m + d_{m-1}, & &\text{where} & q_m &\in \mathbb{Z}, & d_{m-1} &\in \mathbb{D}, \\ q_m &= b\, q_{m+1} + d_m, & &\text{where} & q_{m+1} &\in \mathbb{Z}, & d_m &\in \mathbb{D}. \end{aligned}$$

Consecutive substitution, starting from the last equation, yields the following expression for a:

$$a = b^{m+1} q_{m+1} + b^m \, d_m + \cdots + b \, d_1 + d_0. \tag{4}$$

By Lemma 1

$$b^m d_m + b^{m-1} \, d_{m-1} + \cdots + b \, d_1 + d_0 \leq b^{m+1} - 1. \tag{5}$$

Substituting (5) in (4) we get

$$a \leq b^{m+1} q_{m+1} + b^{m+1} - 1 = b^{m+1} \big(q_{m+1} + 1 \big) - 1.$$

Since $a \geq 1$,

$$b^{m+1} \big(q_{m+1} + 1 \big) \geq 2.$$

Therefore,

$$q_{m+1} \geq 0. \tag{6}$$

By (3) and (4),

$$b^{m+1} q_{m+1} \leq a < b^{m+1}.$$

Consequently,

$$q_{m+1} < 1. \tag{7}$$

Inequalities (6) and (7) $q_{m+1} = 0$. Thus (4) becomes

$$a = b^m d_m + \cdots + b \, d_1 + d_0. \tag{8}$$

By Lemma 1 we have $b^{m-1} d_{m-1} + \cdots + b \, d_1 + d_0 \leq b^m - 1$. Therefore (8) and (3) imply

$$b^m \leq a \leq b^m d_m + b^m - 1.$$

Hence, $b^m d_m \geq 1$, and consequently $d_m \geq 1$. This proves the existence part of the theorem.

To prove the uniqueness, suppose that $k \in \mathbb{N}_0$ and $c_0, \ldots, c_k \in \mathbb{D}$ with $c_k > 0$ are such that

$$a = b^k c_k + \cdots + b \, c_1 + c_0. \tag{9}$$

Then $b^k \leq c_k b^k \leq a$. Therefore, $b^k \in S$. By Lemma 1, $a \leq b^{k+1} - 1 < b^{k+1}$. Thus, $b^k \leq b^m \leq a < b^{k+1}$. Consequently, $1 \leq b^{m-k} < b^1$, and therefore $k = m$. Now, subtracting (9) from (8), yields

$$b^m \big(d_m - c_m \big) + \cdots + b \, \big(d_1 - c_1 \big) + d_0 - c_0 = 0. \tag{10}$$

That is,

$$c_0 - d_0 = b \Big(b^{m-1} \big(d_m - c_m \big) + \cdots + \big(d_1 - c_1 \big) \Big). \tag{11}$$

Since by assumption $-b < c_0 - d_0 < b$, we get

$$-1 < b^{m-1} \big(d_m - c_m \big) + \cdots + \big(d_1 - c_1 \big) < 1.$$

Therefore,

$$b^{m-1} \big(d_m - c_m \big) + \cdots + \big(d_1 - c_1 \big) = 0. \tag{12}$$

Hence, by (11), $c_0 = d_0$. Now, starting from (12) instead of (10) and using $-b < c_1 - d_1 < b$, yields $c_1 = d_1$. Repeating this process $m+1$ times proves that $c_j = d_j$ for all $j = 0, 1, \ldots, m$.

This completes the proof of the theorem. $\square$