ON BINARY REPRESENTATIONS OF INTEGERS WITH DIGITS -1, 0, 1

HELMUT PRODINGER

ABSTRACT. Güntzer and Paul introduced a number system with base 2 and digits $-1, 0, 1$ which is characterized by separating nonzero digits by at least one zero. We find an explicit formula that produces the digits of the expansion of an integer n which leads us to many generalized situations. Syntactical properties of such representations are also discussed.

1. A binary number system

Integers n can be written in many ways as $n = \sum_{k\ge0} a_k 2^k$, where the "digits" a_k are taken from the set $\{-1, 0, 1\}$ (see section 10 for a more precise statement). Let us write $\bar{1} = -1$ for convenience. Güntzer and Paul have shown [6], however, that the representation is unique if the product of any two adjacent digits must be zero, or, in other words, if any nonzero digits are separated by (at least one) zero. (For more historial background, see section 10 .)

In this paper we want to understand this system (and some generalizations) in more detail.

It turns out that the Paul system is obtained by writing $3n/2$ in binary and subtracting (bitwise) the binary representation of $n/2$. For example, $25 = (11001)_2$, and $\frac{3}{2} \cdot 25 = 37.5 = (100101.1)_2$, $\frac{1}{2} \cdot 25 = 12.5 = (1100.1)_2$, and the bitwise difference is $(10\bar{1}001)_P$, which is the representation of 25 in the Paul system.

Clearly, the result is a representation of n in base 2 using digits $-1, 0, 1$. An obvious generalization is to consider $n = (\alpha+1)n - \alpha n$. We analyzed only the case that α is either an integer or an integer divided by a power of 2 (a dyadic rational number).

2. Experimental mathematics

Let us write n in the Paul system as $(n)_P = \dots a_2(n)a_1(n)a_0(n)$. Empirically we find that

---

Date: June 14, 2000.

I started this work when standing in (a long!) line, waiting for a visa for the U.S. When I arrived at Purdue in February, Prof. Szpankowski kindly allowed me to go on with this. Thanks, Wojtek!

$$
\begin{aligned}
\big(a_0(n)\big) &= (010\bar{1})^\omega,\\
\big(a_1(n)\big) &= (00100\bar{0}\bar{1}0)^\omega,\\
\big(a_2(n)\big) &= (000111000000\bar{1}\bar{1}\bar{1}00)^\omega,\\
\big(a_3(n)\big) &= (000000111111000000000000\bar{1}\bar{1}\bar{1}\bar{1}\bar{1}\bar{1}00000)^\omega,
\end{aligned}
$$

etc., which suggests that

$$a_k(n) = 1 \Longleftrightarrow \left\lfloor \frac{n - \frac12 + \frac{(-1)^k}{6}}{2^{k+2}} + \frac56 \right\rfloor - \left\lfloor \frac{n - \frac12 + \frac{(-1)^{k-1}}{6}}{2^{k+2}} + \frac46 \right\rfloor = 1$$

and

$$a_k(n) = \bar{1} \Longleftrightarrow \left\lfloor \frac{n - \frac12 + \frac{(-1)^k}{6}}{2^{k+2}} + \frac26 \right\rfloor - \left\lfloor \frac{n - \frac12 + \frac{(-1)^{k-1}}{6}}{2^{k+2}} + \frac16 \right\rfloor = 1.$$

These forms are modelled to achieve the jump between the different digits. They would give the correct digits, even for n being an arbitrary real number. However, for integers, the simpler

$$a_k(n) = 1 \Longleftrightarrow \left\lfloor \frac{n}{2^{k+2}} + \frac56 \right\rfloor - \left\lfloor \frac{n}{2^{k+2}} + \frac46 \right\rfloor = 1$$

and

$$a_k(n) = \bar{1} \Longleftrightarrow \left\lfloor \frac{n}{2^{k+2}} + \frac26 \right\rfloor - \left\lfloor \frac{n}{2^{k+2}} + \frac16 \right\rfloor = 1$$

also work! (If the expressions on the right hand sides are not 1, they must be 0.)

Consequently we have the formula

$$n = \sum_{k\ge0} \left(\left\lfloor \frac{n}{2^{k+2}} + \frac56 \right\rfloor - \left\lfloor \frac{n}{2^{k+2}} + \frac46 \right\rfloor - \left\lfloor \frac{n}{2^{k+2}} + \frac26 \right\rfloor + \left\lfloor \frac{n}{2^{k+2}} + \frac16 \right\rfloor \right) 2^k.$$

It would be possible to turn these results into a (clumsy) formal proof.

3. The explanation

The key formula (see [9, 1.2.4, Ex. 38]) is

$$\lfloor x \rfloor + \lfloor x + \tfrac1q \rfloor + \cdots + \lfloor x + \tfrac{q-1}{q} \rfloor = \lfloor qx \rfloor, \tag{3.1}$$

where q is an integer ≥ 1 and x a real number.

We can now consider the binary representation of λn, for a natural number λ: The kth digit is given by

$$a_k(\lambda n) = \sum_{i=0}^{2\lambda-1} \left\lfloor \frac{n}{2^{k+1}} + \frac{i}{2\lambda} \right\rfloor (-1)^{i-1};$$

this follows from

$$\lambda n = \sum_{k\ge 0} 2^k \left(\left\lfloor \frac{\lambda n}{2^k} \right\rfloor - 2 \left\lfloor \frac{\lambda n}{2^{k+1}} \right\rfloor \right)$$

and the formula (3.1).

From that one can get the digits for the $n = (\alpha+1)n - \alpha n$ representation. We explain it for $\alpha = 3$, from which the general instance should be clear. Multiplication by 4 involves the additive terms in the floor brackets

$$\frac{7}{8}, \frac{6}{8}, \frac{5}{8}, \frac{4}{8}, \frac{3}{8}, \frac{2}{8}, \frac{1}{8}, \frac{0}{8},$$

while multiplication by 3 involves the terms

$$\frac{5}{6}, \frac{4}{6}, \frac{3}{6}, \frac{2}{6}, \frac{1}{6}, \frac{0}{6}.$$

These two sequence must now be merged, so that the numbers are in decreasing order. Then appropriate signs are attached (which e. g. makes the terms with additive term 0 disappear). The sequence is

$$\frac{7}{8}, \frac{5}{6}, \frac{6}{8}, \frac{4}{6}, \frac{5}{8}, \frac{4}{8}, \frac{3}{6}, \frac{3}{8}, \frac{2}{6}, \frac{2}{8}, \frac{1}{6}, \frac{1}{8}.$$

Thus the digits are given by

$$\begin{aligned} a_k(4n) - a_k(3n) = {} & \left\lfloor \frac{n}{2^{k+1}} + \frac{7}{8} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{5}{6} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{6}{8} \right\rfloor + \left\lfloor \frac{n}{2^{k+1}} + \frac{4}{6} \right\rfloor \\ & + \left\lfloor \frac{n}{2^{k+1}} + \frac{5}{8} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{4}{8} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{3}{6} \right\rfloor + \left\lfloor \frac{n}{2^{k+1}} + \frac{3}{8} \right\rfloor \\ & + \left\lfloor \frac{n}{2^{k+1}} + \frac{2}{6} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{2}{8} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{1}{6} \right\rfloor + \left\lfloor \frac{n}{2^{k+1}} + \frac{1}{8} \right\rfloor . \end{aligned}$$

This principle works also for $\alpha = \frac{\lambda}{2^s}$, a positive dyadic fraction. The kth digit is given by

$$a_k(n) = \sum_{i=0}^{2\lambda-1} \left\lfloor \frac{n}{2^{k+1+s}} + \frac{i}{2\lambda} \right\rfloor (-1)^{i-1}.$$

Observe that now there are also digits after the binary point; we don't need them, however, since we take appropriate differences, these digits will necessarily cancel out.

Thus the kth digit of the $(\alpha+1)n-\alpha n$ representation is obtained by taking the difference of them, once for $\lambda'=2^s+\lambda$, and once for λ. If one want to count digits separately, it is required to arrange the terms

$$\pm\left\lfloor\frac{n}{2^{k+1+s}}+d\right\rfloor$$

is decreasing order (w.r.t. d).

In the Paul case, $\lambda=s=1$, $\lambda'=3$, and thus the kth digit is given by

$$a_k(n)=\sum_{i=0}^{5}\left\lfloor\frac{n}{2^{k+2}}+\frac{i}{6}\right\rfloor(-1)^{i-1}-\sum_{i=0}^{1}\left\lfloor\frac{n}{2^{k+2}}+\frac{i}{2}\right\rfloor(-1)^{i-1},$$

which upon simplification is

$$a_k(n)=\left\lfloor\frac{n}{2^{k+2}}+\frac{5}{6}\right\rfloor-\left\lfloor\frac{n}{2^{k+2}}+\frac{4}{6}\right\rfloor-\left\lfloor\frac{n}{2^{k+2}}+\frac{2}{6}\right\rfloor+\left\lfloor\frac{n}{2^{k+2}}+\frac{1}{6}\right\rfloor.$$

4. Syntactical properties

As mentioned in the introduction, the Paul system is characterized by the property that two nonzero digits are always separated.

In [6] the rewriting rules

$$1\bar{1}\longrightarrow 01\qquad \bar{1}1\longrightarrow 0\bar{1}\qquad 011\longrightarrow 10\bar{1}\qquad 0\bar{1}\bar{1}\longrightarrow \bar{1}01$$

are presented, that can be (repeatedly) applied in any order to transform the standard binary representation of $n\in\mathbb{N}$ into the Paul representation. We say "repeatedly" because of a problem with carries; the following example explains what happens:

$$01111\longrightarrow 10\bar{1}11\longrightarrow 100\bar{1}1\longrightarrow 1000\bar{1}.$$

A more "algorithmic" version is by the following *transducer.* It reads the binary representation of n from right to left (leading zeros must be added if needed). Of course, the output is also produced from right to left. The basic idea is to transform 01^k into $10^{k-1}\bar{1}$ for $k\ge 2$[1]. However, if the next group (to the left) starts immediately with 1, then we can't output the leading 1; it is then a carry, belonging to the next group. To make sure that the transduction ends in the starting state (as it should), we can add two leading zeros (see Figure 3).

When dealing with number systems, it is usually very instructive to see how the number 1 can be added. This can normally be done by an automaton. Here it is (see Figure 2). The word is processed from right to left. If no continuation is defined, the rest of the word is left unchanged. Again, possibly leading zeros are needed to lead to one of the accepting states to the right. The output is also recorded from right to left, usually replacing only a suffix of the word (representation of a number), when reaching one of the two states to the right.

Now let us consider a few properties of the system induced by writing $n=5n/4-n/4$:

[1]In a previous version, I did that also for $k=1$; thanks to a referee this error has been removed.

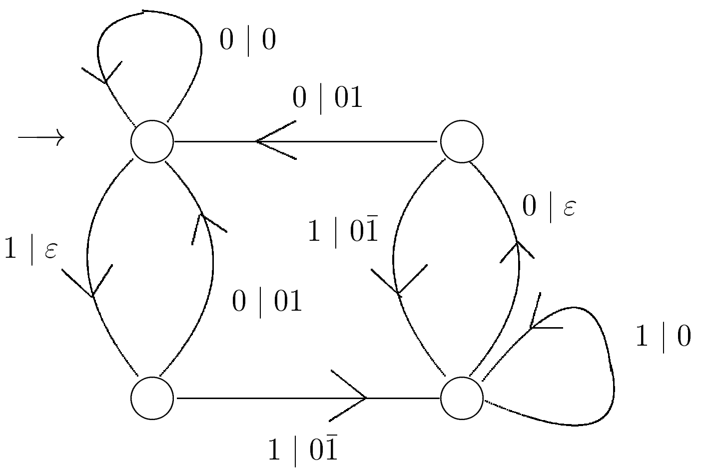

Figure 1. The binary → Paul transformer

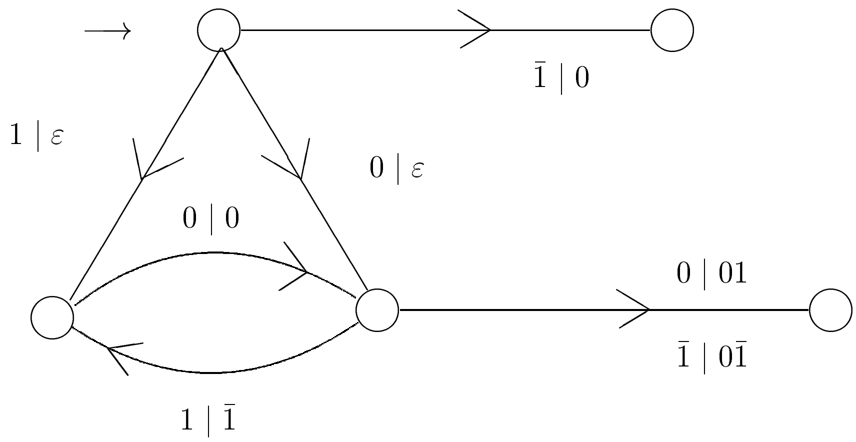

Figure 2. Adding 1 in the Paul representation

First, we describe the syntactically correct words. There are special sequences of nonzero digits, separated by at least two zeros, except perhaps for the right border. A bit more formally, if

$$B = 1 + \bar{1} + 11(\bar{1}1)^*(\varepsilon + \bar{1}) + \bar{1}\bar{1}(1\bar{1})^*(\varepsilon + 1),$$

then the possible representations are given by

$$\varepsilon + (B000^*)^*B0^*.$$

Here are rewriting rules (that should be applied from the rightmost possible position) to obtain the representation in question:

$$\begin{aligned} 101 &\longrightarrow 11\bar{1} \\ 10\bar{1} &\longrightarrow 011 \\ \bar{1}0\bar{1} &\longrightarrow \bar{1}\bar{1}1 \\ \bar{1}01 &\longrightarrow 0\bar{1}\bar{1} \end{aligned}$$

$$
\begin{aligned}
&01\bar{1} \longrightarrow 001 && \\
&0\bar{1}1 \longrightarrow 00\bar{1} && \\
&01^k \longrightarrow 10^{k-1}\bar{1} && k \ge 3 \\
&\bar{1}1^k \longrightarrow 0^k\bar{1} && k \ge 3 \\
&0\bar{1}^k \longrightarrow \bar{1}0^{k-1}1 && k \ge 3 \\
&1\bar{1}^k \longrightarrow 0^k1 && k \ge 3
\end{aligned}
$$

We also show an automaton, producing this transformation, by processing the word from right to left. Input are natural numbers, written in ordinary binary notation. The computation ends in the starting state.

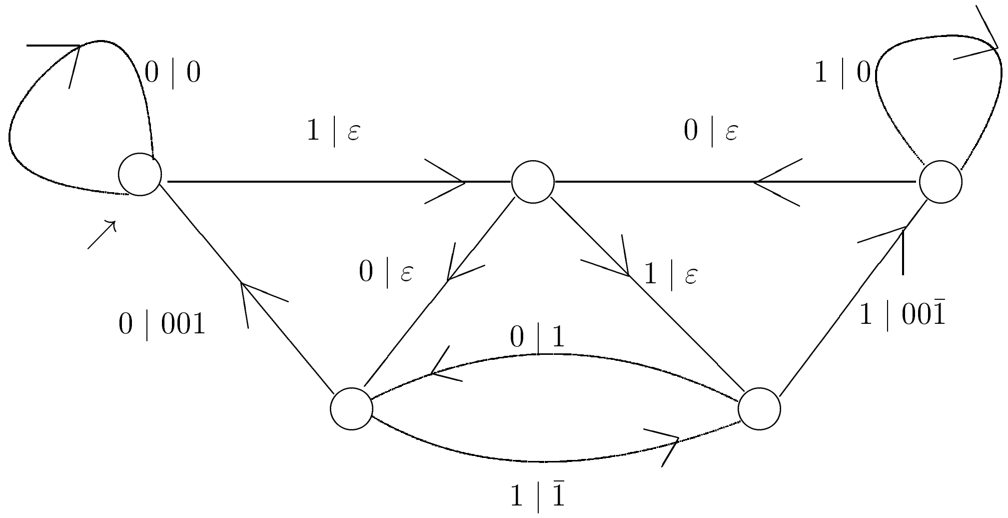

FIGURE 3. The binary → "1 = 5/4 − 1/4" transformer

Instead of giving the automaton to add 1 in this case, we give a complete list of possible situations. A few rules are recursive, which is caused by carries. Notation: w is the admissible representation of an integer n, and $T(w)$ the the admissible representation of $n+1$.

$$
\begin{aligned}
T(w000) &= w001 && \\
T(w11(\bar{1}1)^k00) &= w11(\bar{1}1)^{k-1}00, && k \ge 1 \\
T(w11(\bar{1}1)^k\bar{1}00) &= w11(\bar{1}1)^{k-1}000\bar{1}\bar{1}, && k \ge 1 \\
T(w\bar{1}\bar{1}(1\bar{1})^k00) &= w\bar{1}\bar{1}(1\bar{1})^k1\bar{1}, && k \ge 0 \\
T(w\bar{1}\bar{1}(1\bar{1})^k100) &= w\bar{1}\bar{1}(1\bar{1})^{k-1}100\bar{1}\bar{1}, && k \ge 1 \\
T(w\bar{1}\bar{1}100) &= w\bar{1}00\bar{1}\bar{1} && \\
T(w11\bar{1}00) &= w11\bar{1}1\bar{1} &&
\end{aligned}
$$

$$
\begin{aligned}
T(w1100) &= T(w)00\bar{1}\bar{1} \\
T(w00100) &= w0011\bar{1} \\
T(w00\bar{1}00) &= w000\bar{1}\bar{1} \\
T(w11(\bar{1}1)^k0) &= w11(\bar{1}1)^{k-1}00\bar{1}, && k \ge 1 \\
T(w110) &= T(w)00\bar{1} \\
T(w11(\bar{1}1)^k\bar{1}0) &= w11(\bar{1}1)^{k+1}, && k \ge 0 \\
T(w\bar{1}\bar{1}(1\bar{1})^k0) &= w\bar{1}\bar{1}(1\bar{1})^k1, && k \ge 0 \\
T(w\bar{1}\bar{1}(1\bar{1})^k10) &= w\bar{1}\bar{1}(1\bar{1})^{k-1}100\bar{1}, && k \ge 1 \\
T(w\bar{1}\bar{1}10) &= w\bar{1}00\bar{1} \\
T(w0010) &= w0011 \\
T(w00\bar{1}0) &= w000\bar{1} \\
T(w11(\bar{1}1)^k) &= w11(\bar{1}1)^{k-1}00, && k \ge 1 \\
T(w11) &= T(w)00 \\
T(w11(\bar{1}1)^k\bar{1}) &= w11(\bar{1}1)^k0, && k \ge 0 \\
T(\bar{1}\bar{1}(1\bar{1})^k) &= w\bar{1}\bar{1}(1\bar{1})^{k-1}10, && k \ge 1 \\
T(w\bar{1}\bar{1}) &= w\bar{1}0 \\
T(w001) &= T(w00)0 \\
T(w00\bar{1}) &= w000
\end{aligned}
$$

5. Path length

Let us first reconsider the Paul case with generating functions. Güntzer and Paul arranged all numbers with length $\le n$ in the tree, where the length means the number of digits, starting with no leading zeros. The number 0 is also in the tree, and its length is zero; it serves as the root. Node y is a child[2] of node x, if the least significant nonzero digit in y is replaced by 0, resulting in x. Thus, the depth of a node is the number of nonzero digits. Now consider

$$\varepsilon + (X0^+)^* X0^*, \qquad \text{where } X = 1 + \bar{1},$$

which describes the set of admissible representations in a unique way. Translating accordingly (we mark the digits $\bar{1}, 0, 1$ by a variable z, and also each X by u, to keep track of the depth), we obtain a generating function such that the coefficient of $z^n u^k$ is the number of representations (words) of length n and depth k, viz.

$$1 + \frac{1}{1 - \frac{2z^2u}{1-z}} \frac{2zu}{1-z}.$$

[2]The recent third edition of [9] has already this politically correct form.

We must divide by $1-z$, because we need the number of representations (words) of length $\le n$ and depth k. So we get

$$\frac{1}{1-z}+\frac{2zu}{(1-z)(1-z-2z^2u)};$$

differentiate with respect to u and set $u=1$ to get the generating function of the total path lengths:

$$\frac{2z}{(1+z)^2(1-2z)^2}.$$

The coefficient of z^n in that is

$$\frac{1}{9}n2^{n+2}+\frac{1}{27}2^{n+3}-\frac{2}{9}n(-1)^n-\frac{8}{27}(-1)^n,$$

which is a result of [6].

In this way, we also get the number of nodes in the tree; we must consider (simply replace u by 1)

$$\frac{1}{1-z}+\frac{2z}{(1-z)(1-z-2z^2)}=\frac{1+2z}{(1+z)(1-2z)}=\frac{4}{3}\frac{1}{1-2z}-\frac{1}{3}\frac{1}{1+z},$$

whence we get $2^{n+2}/3-1/3(-1)^n$; this was already reported in [6].

Dividing by the total number of nodes, we see that from the n digits approximately $\frac{n}{3}$ are nonzero digits. In fact, it was proved in [6] that this representation has the least number of nonzero digits of any possible representation!

Now let us consider the next (the $n=5n/4-n/4$) case; from

$$\varepsilon+(B000^*)^*B0^* \qquad \text{with} \qquad B=1+\bar{1}+11(\bar{1}1)^*(\varepsilon+\bar{1})+\bar{1}\bar{1}(1\bar{1})^*(\varepsilon+1)$$

we find

$$1+\frac{2z}{(1+z^2)(1-2z)},$$

which is the generating function of admissible words of length n.

The generating function of all numbers with a representation of length $\le n$ is thus given by

$$\frac{1}{1-z}+\frac{2z}{(1+z^2)(1-2z)(1-z)}=\frac{8}{5}\cdot\frac{1}{1-2z}-\frac{1}{5}\cdot\frac{3+z}{1+z^2}$$

and the coefficients (of z^n) are

$$\frac{2^{n+3}}{5}-\begin{cases}\frac{3}{5}(-1)^{n/2} & n \text{ even}\\ \frac{1}{5}(-1)^{(n-1)/2} & n \text{ odd}\end{cases}$$

or

$$2\left\lfloor\frac{2^{n+2}}{5}\right\rfloor+1.$$

This form has the advantage that the "+1" is responsible for the number 0, and then everything comes twice, because of positive and negative numbers.

This time, the depth of the node is the length of the number except for the trailing zeros.

We get the generating function for the total path length as

$$\frac{2z(1-2z+z^2+4z^3-2z^4)}{(1-z)^2(1-2z)^2(1+z^2)^2};$$

the coefficients are given by

$$\frac{n}{5}2^{n+3}-\frac{1}{5}2^{n+4}+O(n^2).$$

Dividing this by the number of nodes in the tree we get approximately n, which means that the tree is not at all useful. A related question is the total number of nonzero digits in all numbers with a representation of length $\le n$: We start from the generating function

$$1+\frac{2z}{(1-2z)^2(1+z^2)^2}$$

and get coefficients

$$\frac{1}{25}n2^{n+4}-\frac{1}{125}2^{n+6}+O(n^2).$$

Therefore, roughly speaking, from the n letters there are $\frac{2n}{5}$ nonzero digits and $\frac{3n}{5}$ zeros. We will however see more precise statements in the next section.

Here, we have a proportion of $\frac{2}{5}$, which is higher that the $\frac{1}{3}$ from [6].

6. Counting digits

We want to count the number of nonzero digits in the representations produced by $n=(1+1/2^t)n-n/2^t$, since this can be done it a clean and attractive way and leave more general instances as a problem for the interested readers. It is possible (and advisable) to count digits 1 and $\bar{1}$ separately.

The representation of n in this number system is given by

$$\begin{aligned}
n&=\sum_{k\ge0}2^k\sum_{\substack{i=1\\2^t+1\nmid i}}^{2(2^t+1)}\left\lfloor\frac{n}{2^{k+t+1}}+\frac{i}{2(2^t+1)}\right\rfloor(-1)^{i-1}\\
&=\sum_{k\ge0}2^k\sum_{i=1}^{2^t}\left\lfloor\frac{n}{2^{k+t+1}}+\frac{i}{2(2^t+1)}\right\rfloor(-1)^{i-1}+\sum_{k\ge0}2^k\sum_{i=2^t+2}^{2^{t+1}+1}\left\lfloor\frac{n}{2^{k+t+1}}+\frac{i}{2(2^t+1)}\right\rfloor(-1)^{i-1}
\end{aligned}$$

We want to count the average number of digits 1 (resp. $\bar{1}$) in all the integers $n=0,1,\dots,m-1$.

To give a flavour of what is going on we note that a term like

$$\left\lfloor x+\frac{i+1}{q}\right\rfloor-\left\lfloor x+\frac{i}{q}\right\rfloor$$

is responsible for the digit 1, and

$$-\left\lfloor x+\frac{i+1}{q}\right\rfloor+\left\lfloor x+\frac{i}{q}\right\rfloor$$

for the digit $\bar{1}$. The pattern of signs is like $+-+-\dots$, but because there is one index missing, it switches to $-+-+\dots$. So the counting function for the digit 1 in the number n is given by

$$\sum_{k\ge0}\sum_{i=2^t+2}^{2^{t+1}+1}\left\lfloor\frac{n}{2^{k+t+1}}+\frac{i}{2(2^t+1)}\right\rfloor(-1)^{i-1},$$

and the counting function for the digit $\bar{1}$ is given by

$$\sum_{k\ge0}\sum_{i=1}^{2^t}\left\lfloor\frac{n}{2^{k+t+1}}+\frac{i}{2(2^t+1)}\right\rfloor(-1)^{i}.$$

We are not repeating derivations that apply here and can be found in [8]; they are based on a method due to Delange [5].

A typical result is this:

$$\sum_{n=0}^{m-1}\sum_{k\ge0}\left(\left\lfloor\frac{n}{2^{k+2}}+\frac{5}{6}\right\rfloor-\left\lfloor\frac{n}{2^{k+2}}+\frac{4}{6}\right\rfloor\right)=\frac{1}{6}m\log_2 m+m\delta(\log_2 m)+E,$$

where $\delta(x)$ is a certain periodic function of period 1 that could be computed explicitly, as well as its Fourier coefficients, and E is a certain error term, that could also be computed explicitly. (We use δ and E as generic names; they usually vary with the parameters.)

A bit more generally, we have

$$\sum_{n=0}^{m-1}\sum_{k\ge0}\left(\left\lfloor\frac{n}{2^{k+t+1}}+\frac{i+1}{2(2^t+1)}\right\rfloor-\left\lfloor\frac{n}{2^{k+t+1}}+\frac{i}{2(2^t+1)}\right\rfloor\right)$$
$$=\frac{1}{2(2^t+1)}m\log_2 m+m\delta(\log_2 m)+E.$$

Now since in our digit counting problem we have 2^t such terms, the average number of nonzero digits amongst the numbers $0,1,\dots,m-1$ is given by

$$\frac{2^{t-1}}{2^t+1}\log_2 m+\delta(\log_2 m)+\frac{E}{m}.$$

Note that $t=1$ produces the factor $\frac{1}{3}$ from the Paul fame [6]; $t=2$ gives $\frac{2}{5}$, which we have seen previously, and so on; the factor in front of the logarithmic term approaches $\frac{1}{2}$ as t gets large.

7. More examples

Consider the system $\alpha = 1$, i. e. $n = 2n - n$. Then the admissible words are

$$\varepsilon + (10^*\bar{1}0^*)^*,$$

and the digits are produced by

$$a_k(n) = \left\lfloor \frac{n}{2^{k+1}} + \frac{3}{4} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{2}{4} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{2}{4} \right\rfloor + \left\lfloor \frac{n}{2^{k+1}} + \frac{1}{4} \right\rfloor .$$

(The first two terms are responsible for the digit 1, the remaining ones for $\bar{1}$. The average number of nonzero digits in the integers $0, 1, \ldots, m-1$ is thus given by (leading term only)

$$\frac{1}{2} \log_2 m.$$

For $\alpha = 2$ the digits are given by

$$\begin{aligned} a_k(n) = {} & \left\lfloor \frac{n}{2^{k+1}} + \frac{5}{6} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{3}{4} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{2}{3} \right\rfloor + \left\lfloor \frac{n}{2^{k+1}} + \frac{1}{2} \right\rfloor \\ & + \left\lfloor \frac{n}{2^{k+1}} + \frac{1}{2} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{1}{3} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{1}{4} \right\rfloor + \left\lfloor \frac{n}{2^{k+1}} + \frac{1}{6} \right\rfloor ; \end{aligned}$$

we get the Paul digits (shifted by one position) minus the digits from the system $\alpha = 2$. After all, this is not too unexpected, regarding what multiplication by 3 (resp. 3/2) does to a binary representation of an integer.

The average number of nonzero digits in the integers $0, 1, \ldots, m-1$ is thus given by (leading term only)

$$\frac{1}{2} \log_2 m.$$

Observe that the constant $\frac{1}{2}$ is computed here as

$$\frac{1}{2} = \left(\frac{5}{6} - \frac{3}{4} \right) + \left(\frac{2}{3} - \frac{1}{2} \right) + \left(\frac{1}{2} - \frac{1}{3} \right) + \left(\frac{1}{4} - \frac{1}{6} \right) .$$

Our last example is the instance $\alpha = \frac{3}{4}$. The digits are given by

$$\begin{aligned} a_k(n) = {} & \left\lfloor \frac{n}{2^{k+1}} + \frac{13}{14} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{12}{14} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{5}{6} \right\rfloor + \left\lfloor \frac{n}{2^{k+1}} + \frac{11}{14} \right\rfloor \\ & - \left\lfloor \frac{n}{2^{k+1}} + \frac{10}{14} \right\rfloor + \left\lfloor \frac{n}{2^{k+1}} + \frac{4}{6} \right\rfloor + \left\lfloor \frac{n}{2^{k+1}} + \frac{9}{14} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{8}{14} \right\rfloor \\ & - \left\lfloor \frac{n}{2^{k+1}} + \frac{6}{14} \right\rfloor + \left\lfloor \frac{n}{2^{k+1}} + \frac{5}{14} \right\rfloor + \left\lfloor \frac{n}{2^{k+1}} + \frac{2}{6} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{4}{14} \right\rfloor \\ & + \left\lfloor \frac{n}{2^{k+1}} + \frac{3}{14} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{1}{6} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{2}{14} \right\rfloor + \left\lfloor \frac{n}{2^{k+1}} + \frac{1}{14} \right\rfloor . \end{aligned}$$

The average number of nonzero digits in the integers $0, 1, \dots, m-1$ is thus given by (leading term only)

$$\frac{10}{21} \log_2 m.$$

8. Coquet

Coquet [4] dealt with the sum–of–digits function of $3n$. Probably it was never noted that the digits of $3n$ are given by

$$\begin{aligned} a_k(3n) = &\left\lfloor \frac{n}{2^{k+1}} + \frac{5}{6} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{4}{6} \right\rfloor + \left\lfloor \frac{n}{2^{k+1}} + \frac{3}{6} \right\rfloor \\ &- \left\lfloor \frac{n}{2^{k+1}} + \frac{2}{6} \right\rfloor + \left\lfloor \frac{n}{2^{k+1}} + \frac{1}{6} \right\rfloor - \left\lfloor \frac{n}{2^{k+1}} + \frac{0}{6} \right\rfloor . \end{aligned}$$

However, it is not clear whether one can derive his main theorem from that.

9. Conclusion

We have demonstrated that an idea about an exotic data structure (jump interpolation search trees) leads to interesting problems in elementary number theory.

Further research could concentrate on more general values of α, as well as the general (syntactical) study of admissible representations. Also, one could start from different systems than the binary one, and replace the simple difference $(\alpha+1)n - \alpha n$ by more fancy expressions.

10. Shallit

After this paper was finished, I learnt about an unfinished and unpublished draft of Shallit[3] [12] that is of some relevance here. For instance, it shows that each natural number n has *infinitely* many representations as $\sum_k a_k 2^k$ with $a_k \in \{-1, 0, 1\}$.

This draft contains also a substantial list of references that are otherwise not easy to find; the subset of the more relevant ones is [2, 3, 7, 11]. In particular, the Paul system goes back at least to Reitwiesner [11]. Some more recent references are [13, 10].

The draft also contains a transducer from "binary" to "Paul" and recursion formulæ for the sum–of–digits function, connecting it with k–regular sequences in the sense of Allouche and Shallit [1].

Furthermore, it also has the $\frac{2^n - (-1)^n}{3}$ formula, which relates to the number of nodes in the tree, see section 5, as well as an algorithm to generate the Paul representation.

[3] Thanks, Jeffrey for providing it and for several helpful remarks.

References

[1] J.-P. Allouche and J. Shallit. The ring of k–regular sequences. *Theoretical Computer Science*, 98:163–197, 1992.
[2] A. Avizienis. Signed–digit number representation for fast parallel arithmetic. *IRE Trans. Electron. Comput.*, 10:389–400, 1961.
[3] A. Booth. A signed binary multiplication technique. *Quart. J. Mech. Appl. Math.*, 4:236–240, 1951.
[4] J. Coquet. A summation formula related to the binary digits. *Inventiones Mathematicæ*, 73:107–115, 1983.
[5] H. Delange. Sur la fonction sommatoire de la fonction somme des chiffres. *Enseignement Mathématique*, 21:31–47, 1975.
[6] U. Güntzer and M. Paul. Jump interpolation search trees and symmetric binary numbers. *Information Processing Letters*, 26:193–204, 1987/88.
[7] J. Jedwab and C. Mitchell. Minimum weight modified signed–digit representations and fast exponentiation. *Electronics Letters*, 25:1171–1172, 1989.
[8] P. Kirschenhofer and H. Prodinger. Subblock occurrences in positional number systems and Gray code representation. *Journal of Information and Optimization Sciences*, 5:29–42, 1984.
[9] D. E. Knuth. *The Art of Computer Programming*, volume 1: Fundamental Algorithms. Addison-Wesley, 1968. Third edition, 1997.
[10] L. O'Connor. An analysis of exponentiation based on formal languages. *Advances in cryptography–EUROCRYPT '99 (Prague), Lecture Notes in Computer Science*, 1592:375–388, 1999.
[11] G. Reitwiesner. Binary arithmetic. *Vol. 1 of Advances in Computers, Academic Press*, pages 231–308, 1960.
[12] J. Shallit. A primer on balanced binary representations (7 pages). 1992.
[13] J. Thuswaldner. Summatory functions of digital sums occurring in Cryptography. *Periodica Math. Hungarica*, 38:111–130, 1999.

Helmut Prodinger
The John Knopfmacher Centre for Applicable Analysis and Number Theory
Department of Mathematics
University of the Witwatersrand, P. O. Wits
2050 Johannesburg, South Africa
email: `helmut@gauss.cam.wits.ac.za`,
homepage: `http://www.wits.ac.za/helmut/index.htm`