A Generalization to Bellman and Shapiro's Method on the Sum of Digital Sum Functions

C. Gadd and K.L. Wong

Abstract - Let $\alpha(n)$ be the sum of digits of n written in base 2. In 1948, R. Bellman and H.N. Shapiro proved that $\sum_{n\le x}\alpha(n) = \frac{x\log x}{2\log 2} + \mathcal{O}(x\log\log x)$. However, there was a mistake in their proof. We are able to correct the mistake and walk in their footsteps while retaining their idea and method to prove the theorem. We also generalize their method to prove the same theorem for general base.

Keywords : digital sums; base q representation

Mathematics Subject Classification (2020) **:** 11A63;

1 Introduction

Let $q > 1$ be a fixed integer and denote by $\alpha_q(n)$ the sum of digits of n written in base q. Let $f(x)$ and $g(x)$ be real-valued functions and $g(x)$ be positive for all large enough values of x. We write $f(x) = \mathcal{O}(g(x))$ if there exists a positive real number M and a real number x_0 such that $|f(x)| \le Mg(x)$ for all $x \ge x_0$. It was L.E. Bush [3] who first showed in 1940 that

$$A_q(x) := \sum_{n\le x}\alpha_q(n) \sim \frac{q-1}{2\log q}x\log x \text{ as } x\to\infty. \tag{1}$$

In 1949, L. Mirsky [6] proved that

$$A_q(x) = \frac{q-1}{2\log q}x\log x + \mathcal{O}(x), \tag{2}$$

using a special way to count numbers with a certain representation in base q. It was shown that $\mathcal{O}(x)$ is the best possible error term. A weaker error term, namely $\mathcal{O}(x\log\log x)$, for base 2 was discovered a year earlier by Bellman and Shapiro [2]. Many other authors have proved (2) later on using different methods (see [1], [4] and [5]). For example, Ballot in [1] defined a special function $B_q(N)$ which is equal to $A_q(N)$ at a power of q, but different at other numbers. By bounding the difference $A_q(N) - B_q(N)$, the author was able to show that $A_q(N) - B_q(N) = \mathcal{O}(N)$. After proving an asymptotic formula for $B_q(N)$, (2) was established.

This paper stemmed from noticing what seemed — at first — a major mistake in Bellman and Shapiro's proof. The mistake came in the formula [2, (3.10)] with a wrong evaluation of $A_2(N)$. Here, we evaluate $A_2(N)$ properly while, nevertheless, following their proof and idea to obtain the same error term. The corrected formula appears in equation (3) here. We mostly use their notation too. What distinguishes the Bellman–Shapiro method from other approaches is the introduction of a parameter $\mu(N)$ for where to split a sum S into two subsums involved in expressing $A_2(N)$. They then choose $\mu(N)$ so the two partial subsums have the same order of magnitude. In Section 1, we revise Bellman and Shapiro's proof and prove the following theorem.

Theorem 1.1 *We have*

$$A_2(x) = \frac{x \log x}{2 \log 2} + \mathcal{O}(x \log \log x).$$

In Section 3, we proved the same theorem for general bases.

Theorem 1.2 *We have*

$$A_q(x) = \frac{q-1}{2 \log q} x \log x + \mathcal{O}(x \log \log x).$$

2 Proof of Theorem 1.1

Proof. Let $\alpha(n) = \alpha_2(n)$. From the dyadically additive property of $\alpha(n)$, it follows that

$$\begin{aligned} A_2(2^n) &= \sum_{1 \le k \le 2^{n-1}} \alpha(k) + \sum_{1 \le r < 2^{n-1}} \alpha(2^{n-1} + r) + \alpha(2^n) \\ &= A_2(2^{n-1}) + \left[(2^{n-1} - 1) + A_2(2^{n-1}) - \alpha(2^{n-1})\right] + 1 \\ &= 2A_2(2^{n-1}) + 2^{n-1} - 1. \end{aligned}$$

Iterating this relationship, we obtain

$$\begin{aligned} A_2(2^n) &= 2\big(2A_2(2^{n-2}) + 2^{n-2} - 1\big) + 2^{n-1} - 1 = 2^2 A_2(2^{n-2}) + 2 \cdot 2^{n-1} - 3 \\ &= 2^2\big(2A_2(2^{n-3}) + 2^{n-3} - 1\big) + 2 \cdot 2^{n-1} - 3 = 2^3 A_2(2^{n-3}) + 3 \cdot 2^{n-1} - 7 \\ &= 2^3\big(2A_2(2^{n-4}) + 2^{n-4} - 1\big) + 3 \cdot 2^{n-1} - 7 = 2^4 A_2(2^{n-4}) + 4 \cdot 2^{n-1} - 15 \\ &= \ldots = 2^n A_2(2^{n-n}) + n \cdot 2^{n-1} - (2^n - 1) = n2^{n-1} + 1. \end{aligned}$$

If $N = 2^{n_1} + 2^{n_2} + \cdots + 2^{n_t}$, with $n_1 > n_2 > \cdots > n_t$, then

$$\begin{aligned} A_2(N) &= \sum_{k \le 2^{n_1}} \alpha(k) + \sum_{2^{n_1} < k \le 2^{n_1} + 2^{n_2}} \alpha(k) + \cdots + \sum_{2^{n_1} + 2^{n_2} + \cdots + 2^{n_{t-1}} < k \le 2^{n_1} + 2^{n_2} + \cdots + 2^{n_t}} \alpha(k) \\ &= \sum_{1 \le k \le 2^{n_1}} \alpha(k) + \sum_{1 \le k \le 2^{n_2}} \alpha(2^{n_1} + k) + \sum_{1 \le k \le 2^{n_3}} \alpha(2^{n_1} + 2^{n_2} + k) + \cdots \end{aligned}$$

$$
\begin{aligned}
&= A_2(2^{n_1}) + \big(2^{n_2} + A_2(2^{n_2})\big) + \big(2 \cdot 2^{n_3} + A_2(2^{n_3})\big) + \cdots + \big((t-1)\cdot 2^{n_t} + A_2(2^{n_t})\big)\\
&= \sum_{i=1}^{t} A_2(2^{n_i}) + \sum_{i=1}^{t}(i-1)2^{n_i} = \sum_{i=1}^{t}(n_i 2^{n_i-1}+1) + \sum_{i=1}^{t}(i-1)2^{n_i}\\
&= \frac{1}{2}\sum_{i=1}^{t} n_i 2^{n_i} + \sum_{i=1}^{t}(i-1)2^{n_i} + t.
\end{aligned}
$$

Thus, the corrected formula for $A_2(N)$ is

$$A_2(N) = \frac{1}{2}\sum_{i=1}^{t} n_i 2^{n_i} + \sum_{i=1}^{t}(i-1)2^{n_i} + t, \tag{3}$$

instead of

$$\frac{1}{2}\sum_{i=1}^{t} n_i 2^{n_i} + N,$$

in the Bellman–Shapiro paper [2].

Let $S = \sum_{i=1}^{t} n_i 2^{n_i} = \sum_{j=0}^{n_1} a_j (n_1 - j) 2^{n_1-j}$, where $a_j = 0$ or 1 depending upon the presence or absence of 2^{n_1-j} in the sum $N = 2^{n_1} + 2^{n_2} + \cdots + 2^{n_t}$. Then

$$S = \sum_{j=0}^{\mu(N)} a_j (n_1 - j) 2^{n_1-j} + \sum_{j=\mu(N)+1}^{n_1} a_j (n_1 - j) 2^{n_1-j}$$

where $n_1 \geq \mu(N) + 1 \geq 1$, and $\mu(N)$ will be chosen advantageously below.

$$
\begin{aligned}
S &= n_1 \sum_{j=0}^{\mu(N)} a_j 2^{n_1-j} - \sum_{j=0}^{\mu(N)} a_j j 2^{n_1-j} + O(n_1 2^{n_1-\mu(N)})\\
&= n_1 \sum_{j=0}^{n_1} a_j 2^{n_1-j} - n_1 \sum_{j=\mu(N)+1}^{n_1} a_j 2^{n_1-j} - \sum_{j=0}^{\mu(N)} a_j j 2^{n_1-j} + O(n_1 2^{n_1-\mu(N)})\\
&= n_1 N + O(n_1 2^{n_1-\mu(N)}) + O(\mu(N) 2^{n_1}) + O(n_1 2^{n_1-\mu(N)})\\
&= n_1 N + O(\mu(N) 2^{n_1}) + O(n_1 2^{n_1-\mu(N)}).
\end{aligned}
$$

Choose $\mu(N)$ to make the two error terms of comparable size, i.e., such that

$$
\begin{aligned}
n_1 2^{n_1-\mu(N)} &= \mu(N) 2^{n_1},\\
\text{or} \quad n_1 2^{-\mu(N)} &= \mu(N).
\end{aligned}
$$

Additionally,

$$
\begin{aligned}
N = 2^{n_1} + 2^{n_2} + \cdots + 2^{n_t} &= 2^{n_1}(1 + 2^{n_2-n_1} + \cdots + 2^{n_t-n_1})\\
\log_2 N &= n_1 + \log_2(1 + 2^{n_2-n_1} + \cdots + 2^{n_t-n_1}).
\end{aligned}
$$

Since $1 < 1 + 2^{n_2-n_1} + \cdots + 2^{n_t-n_1} < 2$, we have

$$0 < \log_2(1 + 2^{n_2 - n_1} + \cdots + 2^{n_t - n_1}) < 1$$
$$n_1 < \log_2 N < n_1 + 1.$$

Thus, $n_1 = \lfloor \log_2 N \rfloor = \left\lfloor \dfrac{\log N}{\log 2} \right\rfloor$.

Therefore,

$$\mu(N) = \left\lfloor \frac{\log N}{\log 2} \right\rfloor 2^{-\mu(N)}.$$

It follows that $\mu(N) \sim \dfrac{\log(\log N)}{\log 2}$ as $N \longrightarrow \infty$. Thus,

$$S = \frac{N \log N}{\log 2} + \mathcal{O}(N \log \log N).$$

Now, let $R = \sum_{i=1}^{t}(i-1)2^{n_i} + t$. To prove the theorem it suffices to show that $R = \mathcal{O}(N)$. But we can show more precisely that $R \leq N$.

A table of $N \leq 32$ values, binary representation, R values, and the increase in R values is presented below. The table shows some repeating patterns and suggests that $R \leq N$. A graph of R versus N is also included below.

N	Binary	R	R	Increase of R
1	1	$1 + 0 \cdot 2^0$	1	
2	10	$1 + 0 \cdot 2^1$	1	0
3	11	$2 + 0 \cdot 2^1 + 1 \cdot 2^0$	3	2
4	100	$1 + 0 \cdot 2^2$	1	-2
5	101	$2 + 0 \cdot 2^2 + 1 \cdot 2^0$	3	2
6	110	$2 + 0 \cdot 2^2 + 1 \cdot 2^1$	4	1
7	111	$3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 2 \cdot 2^0$	7	3
8	1000	$1 + 0 \cdot 2^3$	1	-6
9	1001	$2 + 0 \cdot 2^3 + 1 \cdot 2^0$	3	2
10	1010	$2 + 0 \cdot 2^3 + 1 \cdot 2^1$	4	1
11	1011	$3 + 0 \cdot 2^3 + 1 \cdot 2^1 + 2 \cdot 2^0$	7	3
12	1100	$2 + 0 \cdot 2^3 + 1 \cdot 2^2$	6	-1
13	1101	$3 + 0 \cdot 2^3 + 1 \cdot 2^2 + 2 \cdot 2^0$	9	3
14	1110	$3 + 0 \cdot 2^3 + 1 \cdot 2^2 + 2 \cdot 2^1$	11	2
15	1111	$4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 2 \cdot 2^1 + 3 \cdot 2^0$	15	4
16	10000	$1 + 0 \cdot 2^4$	1	-14
17	10001	$2 + 0 \cdot 2^4 + 1 \cdot 2^0$	3	2
18	10010	$2 + 0 \cdot 2^4 + 1 \cdot 2^1$	4	1
19	10011	$3 + 0 \cdot 2^4 + 1 \cdot 2^1 + 2 \cdot 2^0$	7	3

20	10100	$2+0\cdot 2^4+1\cdot 2^2$	6	-1
21	10101	$3+0\cdot 2^4+1\cdot 2^2+2\cdot 2^0$	9	3
22	10110	$3+0\cdot 2^4+1\cdot 2^2+2\cdot 2^1$	11	2
23	10111	$4+0\cdot 2^4+1\cdot 2^2+2\cdot 2^1+3\cdot 2^0$	15	4
24	11000	$2+0\cdot 2^4+1\cdot 2^3$	10	-5
25	11001	$3+0\cdot 2^4+1\cdot 2^3+2\cdot 2^0$	13	3
26	11010	$3+0\cdot 2^4+1\cdot 2^3+2\cdot 2^1$	15	2
27	11011	$4+0\cdot 2^4+1\cdot 2^3+2\cdot 2^1+3\cdot 2^0$	19	4
28	11100	$3+0\cdot 2^4+1\cdot 2^3+2\cdot 2^2$	19	0
29	11101	$4+0\cdot 2^4+1\cdot 2^3+2\cdot 2^2+3\cdot 2^0$	23	4
30	11110	$4+0\cdot 2^4+1\cdot 2^3+2\cdot 2^2+3\cdot 2^1$	26	3
31	11111	$5+0\cdot 2^4+1\cdot 2^3+2\cdot 2^2+3\cdot 2^1+4\cdot 2^0$	31	5
32	100000	$1+0\cdot 2^5$	1	-32

Table 1: R vs N

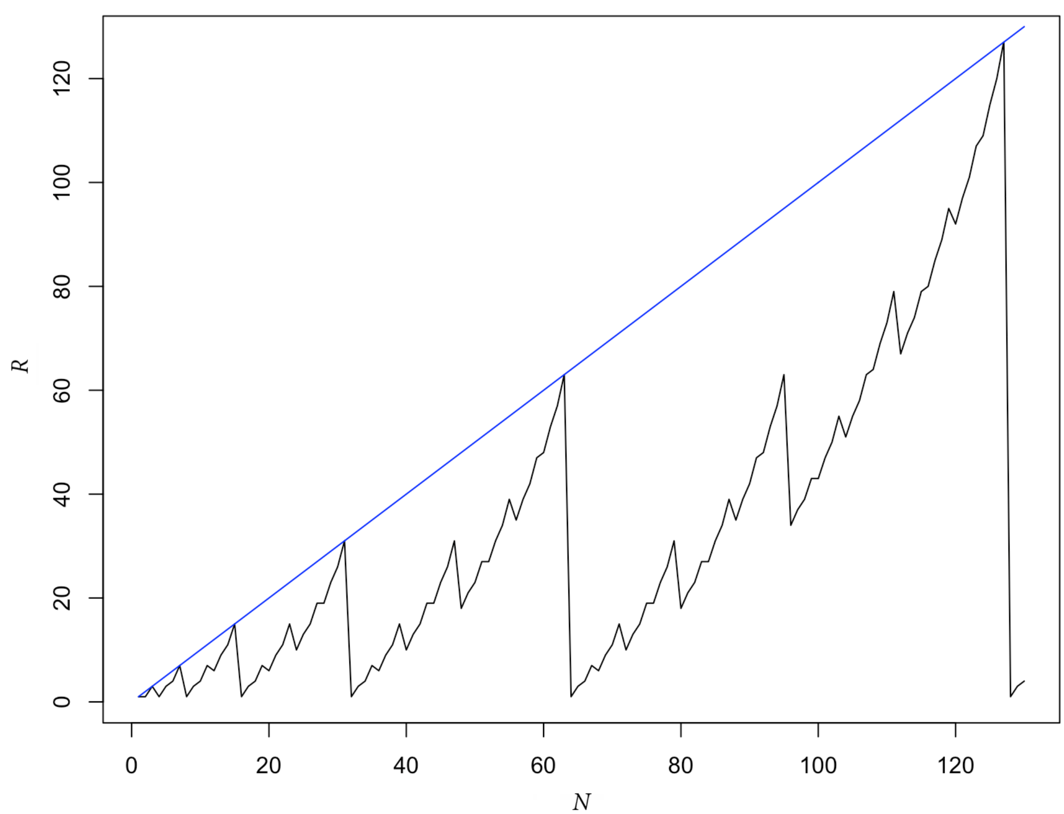

Figure 1: R vs N

Here we prove $R \leq N$. Consider

$$N - R = \sum_{i=1}^{t} 2^{n_i} - \left(\sum_{i=1}^{t} (i-1)2^{n_i} + t \right)$$
$$= 2^{n_1} - (2^{n_3} + 2 \cdot 2^{n_4} + 3 \cdot 2^{n_5} + \cdots + (t-2)2^{n_t} + t).$$

Note that

$$\begin{aligned} 2^{n_3} + 2^{n_4} + 2^{n_5} + \cdots + 2^{n_t} &\leq 2^{n_2} - 1, \\ 2^{n_4} + 2^{n_5} + \cdots + 2^{n_t} &\leq 2^{n_3} - 1, \\ 2^{n_5} + \cdots + 2^{n_t} &\leq 2^{n_4} - 1, \\ &\vdots \\ 2^{n_{t-1}} + 2^{n_t} &\leq 2^{n_{t-2}} - 1, \\ 2^{n_t} &\leq 2^{n_{t-1}} - 1. \end{aligned}$$

If we sum up all the inequalities above, we have

$$\begin{aligned} 2^{n_3} + 2 \cdot 2^{n_4} + 3 \cdot 2^{n_5} + \cdots + (t-2)2^{n_t} &\leq 2^{n_2} + 2^{n_3} + \cdots + 2^{n_{t-1}} - (t-2) \\ 2^{n_3} + 2 \cdot 2^{n_4} + 3 \cdot 2^{n_5} + \cdots + (t-2)2^{n_t} + t &\leq 2^{n_2} + 2^{n_3} + \cdots + 2^{n_{t-1}} + 2 \\ &\leq 2^{n_1} - 1 - 2^{n_t} + 2 \\ &\leq 2^{n_1} - (2^{n_t} - 1) \\ &\leq 2^{n_1}. \end{aligned}$$

Therefore, $N - R \geq 0$ and $R \leq N$. It is interesting to note that we only have an equality when all n_i are consecutive and $n_t = 0$. In other words, $R = N$ only when $N = 2^n - 1$ for some n. This can be seen from the table and the graph above.

Now, we have proved the following:

$$A_2(N) = \frac{N \log N}{2 \log 2} + O(N \log \log N).$$

□

3 Proof of Theorem 1.2

Let $q \geq 2$ be an arbitrary base. To use the same method as in the base 2 case, we first derive a formula for $A_q(q^n)$. This formula is proved in [1] using a different method. The author used a clever way to pair terms in the sum. Here we use a recursive method to prove the formula.

Lemma 3.1 *Let $A_q(x) = \sum_{k\le x} \alpha_q(k)$ where $\alpha_q(k)$ is the sum of digits of k written in base q. Then*

$$A_q(q^n) = \sum_{k\le q^n} \alpha_q(k) = 1 + \frac{q-1}{2} n q^n.$$

Proof. Following the idea from [2], we split the sum first at q^{n-1}. A difference here is that we also need to run through different coefficients of q^{n-1}.

$$\begin{aligned}
A_q(q^n) &= \sum_{k\le q^n} \alpha(k) = \sum_{k\le q^{n-1}} \alpha(k) + \sum_{m=1}^{q-1}\left(\sum_{r=1}^{q^{n-1}-1} \alpha(m\cdot q^{n-1}+r) + \alpha((m+1)q^{n-1})\right)\\
&= \sum_{k\le q^{n-1}} \alpha(k) + \sum_{m=1}^{q-1}\left(\sum_{r=1}^{q^{n-1}-1} \alpha(m\cdot q^{n-1}+r)\right) + \sum_{m=1}^{q-1} \alpha\big((m+1)q^{n-1}\big)\\
&= \sum_{k\le q^{n-1}} \alpha(k) + \sum_{m=1}^{q-1}\left(\sum_{r=1}^{q^{n-1}-1} \alpha(m\cdot q^{n-1})\right) + \sum_{m=1}^{q-1}\left(\sum_{r=1}^{q^{n-1}-1} \alpha(r)\right)\\
&\quad + \sum_{m=1}^{q-1} \alpha\big((m+1)q^{n-1}\big)\\
&= \sum_{k\le q^{n-1}} \alpha(k) + \sum_{m=1}^{q-1} m(q^{n-1}-1) + (q-1)\sum_{r=1}^{q^{n-1}-1} \alpha(r) + \sum_{m=1}^{q-1} m\\
&= \sum_{k\le q^{n-1}} \alpha(k) + \sum_{m=1}^{q-1} m(q^{n-1}-1) + (q-1)\sum_{r=1}^{q^{n-1}} \alpha(r) + \sum_{m=1}^{q-2} m\\
&= qA_q(q^{n-1}) + q^{n-1}\sum_{m=1}^{q-1} m - \sum_{m=1}^{q-1} m + \sum_{m=1}^{q-2} m\\
&= qA_q(q^{n-1}) + q^{n-1}\left(\frac{q(q-1)}{2}\right) - (q-1)\\
&= qA_q(q^{n-1}) + \frac{q-1}{2} q^n - (q-1)\\
&= q\left[qA_q(q^{n-2}) + \frac{q-1}{2}q^{n-1} - (q-1)\right] + \frac{q-1}{2}q^n - (q-1)\\
&= q^2A_q(q^{n-2}) + 2\cdot\frac{q-1}{2}q^n - q(q-1) - (q-1)\\
&= q^3A_q(q^{n-3}) + 3\cdot\frac{q-1}{2}q^n - q^2(q-1) - q(q-1) - (q-1)\\
&\ \ \vdots\\
&= q^nA_q(q^{n-n}) + n\cdot\frac{q-1}{2}q^n - (q-1)(1+q+q^2+\cdots+q^{n-1})
\end{aligned}$$

$$
\begin{aligned}
&= q^n + \frac{q-1}{2} n q^n - (q^n - 1) \\
&= 1 + \frac{q-1}{2} n q^n.
\end{aligned}
$$

□

Now let $N = c_1 q^{n_1} + c_2 q^{n_2} + c_3 q^{n_3} + \cdots + c_t q^{n_t}$, where $n_1 > n_2 > \cdots > n_t \geq 0$. Note that in the base q expansion of N, the coefficients are not necessarily 1. Thus, we need a formula for $A_q(c_n q^n)$ while the base 2 case only needs a formula for $A_2(2^n)$. Fortunately, when we split the sum $A_q(c_n q^n)$ at different multiples of q^n, we recover many copies of $A_q(q^n)$. Thus, we have

$$
\begin{aligned}
A_q(c_n q^n) &= \sum_{k \leq c_n q^n} \alpha(k) \\
&= \sum_{1 \leq k \leq q^n} \alpha(k) + \sum_{1 \leq r \leq q^n - 1} \alpha(q^n + r) + \alpha(2 \cdot q^n) + \cdots + \sum_{1 \leq r \leq q^n - 1} \alpha((c_n - 1) q^n + r) \\
&\quad + \alpha(c_n q^n) \\
&= A_q(q^n) + (q^n - 1) + \sum_{1 \leq r \leq q^n - 1} \alpha(r) + \alpha(2 \cdot q^n) + \cdots + (c_n - 1)(q^n - 1) \\
&\quad + \sum_{1 \leq r \leq q^n - 1} \alpha(r) + \alpha(c_n q^n) \\
&= c_n A_q(q^n) + q^n \big(1 + 2 + \cdots + (c_n - 1)\big) \\
&= c_n \left(1 + \frac{q-1}{2} n q^n\right) + q^n \left(\frac{c_n (c_n - 1)}{2}\right) \\
&= c_n + \left(\frac{c_n (q-1)}{2} n + \frac{c_n (c_n - 1)}{2}\right) q^n.
\end{aligned}
$$

This formula generalizes the formula in Lemma 3.1. Now we can look at $A_q(N)$. Similar to the base 2 case, we split the sum $A_q(N)$ at sums of different numbers of the first few summands of N. Thus, we have

$$
\begin{aligned}
A_q(N) &= \sum_{k \leq c_1 q^{n_1}} \alpha(k) + \sum_{c_1 q^{n_1} < k \leq c_1 q^{n_1} + c_2 q^{n_2}} \alpha(k) + \cdots \\
&\quad + \sum_{c_1 q^{n_1} + c_2 q^{n_2} + \cdots + c_{t-1} q^{n_{t-1}} < k \leq c_1 q^{n_1} + c_2 q^{n_2} + \cdots + c_t q^{n_t}} \alpha(k) \\
&= \sum_{k \leq c_1 q^{n_1}} \alpha(k) + \sum_{1 \leq r \leq c_2 q^{n_2}} \alpha(c_1 q^{n_1} + r) + \cdots \\
&\quad + \sum_{1 \leq r \leq c_t q^{n_t}} \alpha(c_1 q^{n_1} + c_2 q^{n_2} + \cdots + c_{t-1} q^{n_{t-1}} + r)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{k \le c_1 q^{n_1}} \alpha(k) + \sum_{1 \le r \le c_2 q^{n_2}} \alpha(c_1 q^{n_1}) + \sum_{1 \le r \le c_2 q^{n_2}} \alpha(r) + \cdots + \\
&\quad \sum_{1 \le r \le c_t q^{n_t}} \alpha(c_1 q^{n_1} + c_2 q^{n_2} + \cdots + c_{t-1} q^{n_{t-1}}) + \sum_{1 \le r \le c_t q^{n_t}} \alpha(r) \\
&= A_q(c_1 q^{n_1}) + c_1 \cdot c_2 q^{n_2} + A_q(c_2 q^{n_2}) + \cdots + (c_1 + c_2 + \cdots + c_{t-1}) c_t q^{n_t} + A_q(c_t q^{n_t}) \\
&= \sum_{i=1}^{t} A_q(c_i q^{n_i}) + \sum_{i=1}^{t-1} \left(c_{i+1} \sum_{j=1}^{i} c_j \right) q^{n_{i+1}} \\
&= \sum_{i=1}^{t} \left[c_i + \left(\frac{c_i(q-1)}{2} n_i + \frac{c_i(c_i-1)}{2} \right) q^{n_1} \right] + \sum_{i=1}^{t-1} \left(c_{i+1} \sum_{j=1}^{i} c_j \right) q^{n_{i+1}}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{t} \left(\frac{c_i(q-1)}{2} n_i + \frac{c_i(c_i-1)}{2} \right) q^{n_i} + \sum_{i=1}^{t} c_i + \sum_{i=1}^{t-1} \left(c_{i+1} \sum_{j=1}^{i} c_j \right) q^{n_{i+1}} \\
&= \sum_{i=1}^{t} \left(\frac{c_i(q-1)}{2} n_i + \frac{c_i(c_i-1)}{2} \right) q^{n_i} + \sum_{i=1}^{t} c_i + \sum_{i=2}^{t} (c_1 + c_2 + \cdots + c_{i-1}) c_i q^{n_i}.
\end{aligned}
$$

Similar to the base 2 case, there are two parts in the formula of $A_q(N)$. They both behave in the same ways as the two parts in the base 2 case respectively. To show that, we first let

$$
S_q = \sum_{i=1}^{t} \left(\frac{c_i(q-1)}{2} n_i + \frac{c_i(c_i-1)}{2} \right) q^{n_i} \text{ and } R_q = \sum_{i=1}^{t} c_i + \sum_{i=2}^{t} (c_1 + c_2 + \cdots + c_{i-1}) c_i q^{n_i}.
$$

We will prove the general results, namely

$$
S_q = \frac{q-1}{2} \cdot \frac{N \log N}{\log q} + \mathcal{O}(N \log \log N) \text{ and } R_q \le N.
$$

Then, Theorem 1.2 will follow.

Now we proceed the $\mu(N)$ approach to S_q. We have

$$
\begin{aligned}
S_q &= \sum_{i=1}^{t} \left(\frac{c_i(q-1)}{2} n_i + \frac{c_i(c_i-1)}{2} \right) q^{n_i} = \sum_{j=0}^{n_1} a_j \left(\frac{d_j(q-1)}{2} (n_1 - j) + \frac{d_j(d_j-1)}{2} \right) q^{n_1 - j} \\
&= \sum_{j=0}^{\mu(N)} a_j \left(\frac{d_j(q-1)}{2} (n_1 - j) + \frac{d_j(d_j-1)}{2} \right) q^{n_1 - j} \\
&\quad + \sum_{j=\mu(N)+1}^{n_1} a_j \left(\frac{d_j(q-1)}{2} (n_1 - j) + \frac{d_j(d_j-1)}{2} \right) q^{n_1 - j}
\end{aligned}
$$

where $n_1 \geq \mu(N) + 1 \geq 1$, and $\mu(N)$ will be chosen advantageously below. Then

$$\begin{aligned}
S_q =& n_1 \sum_{j=0}^{\mu(N)} a_j \left(\frac{d_j(q-1)}{2}\right) q^{n_1-j} - \sum_{j=0}^{\mu(N)} a_j \left(\frac{d_j(q-1)}{2}\right) j q^{n_1-j} + \sum_{j=0}^{\mu(N)} a_j \left(\frac{d_j(d_j-1)}{2}\right) q^{n_1-j} \\
& + \mathcal{O}(n_1 q^{n_1-\mu(N)}) \\
=& n_1 \sum_{j=0}^{n_1} a_j \left(\frac{d_j(q-1)}{2}\right) q^{n_1-j} - n_1 \sum_{j=\mu(N)+1}^{n_1} a_j \left(\frac{d_j(q-1)}{2}\right) q^{n_1-j} + \mathcal{O}(\mu(N) q^{n_1}) + \\
& \mathcal{O}(q^{n_1-\mu(N)}) + \mathcal{O}(n_1 q^{n_1-\mu(N)}) \\
=& \frac{(q-1)}{2} n_1 \sum_{j=0}^{n_1} a_j d_j q^{n_1-j} + \mathcal{O}(n_1 q^{n_1-\mu(N)}) + \mathcal{O}(\mu(N) q^{n_1}) + \mathcal{O}(n_1 q^{n_1-\mu(N)}) \\
=& \frac{(q-1)}{2} n_1 N + \mathcal{O}(\mu(N) q^{n_1}) + \mathcal{O}(n_1 q^{n_1-\mu(N)}).
\end{aligned}$$

Consider that,

$$\begin{aligned}
N &= c_1 q^{n_1} + c_2 q^{n_2} + c_3 q^{n_3} + \cdots + c_t q^{n_t} \\
\log_q N &= \log_q(q^{n_1}(c_1 + c_2 q^{n_2-n_1} + c_3 q^{n_3-n_1} + \cdots + c_t q^{n_t-n_1})) \\
\log_q N &= n_1 + \log_q(c_1 + c_2 q^{n_2-n_1} + c_3 q^{n_3-n_1} + \cdots + c_t q^{n_t-n_1}).
\end{aligned}$$

Since

$$1 \leq c_1 + c_2 q^{n_2-n_1} + c_3 q^{n_3-n_1} + \cdots + c_t q^{n_t-n_1} < (q-1) \sum_{i=0}^{\infty} \frac{1}{q^i} = q,$$

we have

$$\begin{aligned}
0 &\leq \log_q(c_1 + c_2 q^{n_2-n_1} + c_3 q^{n_3-n_1} + \cdots + c_t q^{n_t-n_1}) < 1 \\
n_1 &\leq \log_q N < n_1 + 1.
\end{aligned}$$

Thus, $n_1 = \lfloor \log_q N \rfloor = \left\lfloor \frac{\log N}{\log q} \right\rfloor$.

Now choose $\mu(N)$ so the two error terms are of the same magnitude:

$$\begin{aligned}
\mu(N) q^{n_1} &= n_1 q^{n_1-\mu(N)} \\
\mu(N) &= n_1 q^{-\mu(N)} \\
\mu(N) &= \left\lfloor \frac{\log N}{\log q} \right\rfloor q^{-\mu(N)}.
\end{aligned}$$

Taking logarithms on both sides, we see that, $\mu(N) \sim \dfrac{\log(\log N)}{\log q}$ as $N \longrightarrow \infty$.

Thus,

$$S_q = \frac{q-1}{2} \cdot \frac{N \log N}{\log q} + \mathcal{O}(N \log \log N).$$

Now we prove $R_q \leq N$. Although each coefficient of q^{n_i} in R_q looks more complicated with a sum, we are still able to split them in the same way as in the base 2 case. We first write out the sum,

$$R_q = c_1 \cdot c_2 \cdot q^{n_2} + (c_1 + c_2)c_3 \cdot q^{n_3} + \cdots + (c_1 + c_2 + \cdots + c_{t-1})c_t \cdot q^{n_t} + \sum_{i=1}^{t} c_i.$$

Note that,

$$\begin{aligned} c_1 \cdot c_2 \cdot q^{n_2} + c_1 \cdot c_3 \cdot q^{n_3} + c_1 \cdot c_4 \cdot q^{n_4} + \cdots + c_1 \cdot c_t \cdot q^{n_t} &\leq c_1(q^{n_1} - 1), \\ c_2 \cdot c_3 \cdot q^{n_3} + c_2 \cdot c_4 \cdot q^{n_4} + \cdots + c_2 \cdot c_t \cdot q^{n_t} &\leq c_2(q^{n_2} - 1), \\ c_3 \cdot c_4 \cdot q^{n_4} + \cdots + c_3 \cdot c_t \cdot q^{n_t} &\leq c_3(q^{n_3} - 1), \\ &\vdots \\ c_{t-1} \cdot c_t \cdot q^{n_t} &\leq c_{t-1}(q^{n_{t-1}} - 1). \end{aligned}$$

If we sum all the inequalities above, we have

$$R_q \leq c_1 q^{n_1} + c_2 q^{n_2} + \cdots + c_{t-1} q^{n_{t-1}} + c_t \leq N.$$

Note that we only have an equality if all n_i are consecutive, $n_t = 0$, and all $c_i = q - 1$, i.e., $N = q^n - 1$ for some n.

Therefore, we have now proved

$$A_q(N) = \frac{q-1}{2} \cdot \frac{N \log N}{\log q} + \mathcal{O}(N \log \log N).$$

Acknowledgment

We are thankful to Dr. Christian Ballot for his generous help and suggestions to make this paper better.

References

[1] C. Ballot, On Zeckendorf and base b digit sums, *Fibonacci Quart.*, **51** (2013), 319–325.

[2] R. Bellman and H.N. Shapiro, On a problem in additive number theory, *Ann. of Math.*, **49** (1948), 333–340.

[3] L.E. Bush, An asymptotic formula for the average sum of the digits of integers, *Amer. Math. Monthly*, **47** (1940), 154–156.

[4] P. Cheo and S. Yien, A problem on the k-adic representation of positive integers, *Acta Math. Sinica*, **5** (1955), 433–438

[5] J. Pihko, An algorithm for the additive representation of positive integers, *Ann. Acad. Sci. Fenn. Ser. A I Math. Dissertationes*, No. 46 (1983), 54 pp.

[6] L. Mirsky, A theorem on representations of integers in the scale of r, *Scripta Math.*, **15** (1949), 11–12.

Cora Gadd
Brigham Young University - Hawaii
55-220 Kulanui Street
Laie, HI, 96762
E-mail: coragadd@go.byuh.edu

Ka Lun Wong
Brigham Young University - Hawaii
55-220 Kulanui Street
Laie, HI, 96762
E-mail: kalun.wong@byuh.edu

Received: September 15, 2022 **Accepted**: October 1, 2022
Communicated by Steven J. Miller