Difference of the Digital Sums of an Integer Base $b$ and Its Prime Factors

WAYNE L. McDaniel

Department of Mathematics and Computer Science, University of Missouri, St. Louis, Missouri 63121

Communicated by H. Zassenhaus
Received February 25, 1987

Let $p_1, \ldots, p_k$ be primes and $m = \prod_{i=1}^{k} p_i$ be an integer base $b$; $S(b, m)$ denotes the sum of digits function base $b$, and $S_d(b, m) = \sum_{i=1}^{k} S(b, p_i)$. If $m$ is composite and $c$ is defined by $S_d(b, m) - S(b, m) = c$, we shall say that $m$ is an element of the set $I_b(c)$. We show that $I_b(c)$ is an infinite set for all integers $c$ and all integers $b \geq 8$, and prove that, for $c \geq 0$, the values of $S(b, m)/S_d(b, m)$ are dense in the interval $[0, 1]$. This provides the answer, for an arbitrary base $b \geq 8$ and integer $c$, to a question posed by A. Wilansky about the set $I_{10}(0)$, the elements of which Wilansky termed "Smith" numbers.

1. INTRODUCTION

Let $b$ be an integer $\geq 2$. If $m$ is an integer (base 10), the function

$$S(b, m) = m - \sum_{i=1}^{\infty} (b - 1)[m/b^i], \quad (1)$$

where $[ ]$ denotes the greatest integer function, is called the sum of digits function for the integer $m$ base $b$ [6]. If for some positive integer $n$, integers $a_i$, for $0 \leq i \leq n$, are known such that $0 \leq a_i < b$, and $m = \sum_{i=0}^{n} a_i b^i$, then $S(b, m)$ is, of course, simply $\sum_{i=0}^{n} a_i$.

The sum of digits function has been investigated by a number of authors and has been encountered in areas as diverse as topology, combinatorics, and probability theory. We refer the reader to Stolarsky's paper [10] for a list of over 60 references to relevant papers.

Possibly because the irregularities of the function $S(b, m)$ preclude the possibility of obtaining a "nice" analytic approximation, research efforts have been directed toward understanding the properties of the function

$$A(m) = \sum_{k < m} S(b, k).$$
Bush [1] obtained an asymptotic formula for $A(m)/m$ and Mirsky [7] showed that, in fact,

$$A(m) = [(b - 1)/2]m \cdot \log_b m + O(m).$$

An explicit formula for the error term is due, for $b = 2$, to Trollope [12], and, for $b > 2$, to Delange [2]. These formulas do not require a knowledge of the digits of $m$. An explicit formula for $A(m)$, for $b = 10$, requiring a knowledge of the digits of $m$, was published much earlier by d'Ocagne ([8], or see [3, p. 457]). During the last two decades, researchers have turned their attention to various generalizations of the sum of digits function (see [4, 5, 9, 10, 11]).

Recently, further interest in $S(b, m)$ has been generated by a question posed by A. Wilansky. Let $m = \prod_{i=1}^{k} p_i$, $p_i$ prime, and $S_p(b, m) = \sum_{i=1}^{k} S(b, p_i)$. If $m$ is composite and $c$ is defined by $S_p(b, m) - S(b, m) = c$, we shall say that $m$ is an element of the set $I_\lambda(c)$. Wilansky named the elements of $I_{10}(0)$ "Smith" numbers and asked whether $I_{10}(0)$ is infinite [13]. The question has focused attention on the relative size of $S(b, m)$ and $S_p(b, m)$.

Heuristically, one might expect that if $m$ ranges over all positive integers, $S(b, m)$ and $S_p(b, m)$ will be, on the average, approximately equal: If $m$ has $k$ prime factors and $q$ is the prime such that $q^k$ is "closest" to $m$, the average value of $S(b, x)$, where $x$ ranges over all nonnegative integers < $m$, i.e., of $A(m)/m$, will be approximately equal to the average value of $S(b, x)$, where $x$ ranges over all nonnegative integers < $q^k$, i.e., of $A(q^k)/q^k$. Since $S_p(b, q^k) = k \cdot S_p(b, q) = k \cdot S(b, q)$, the relation

$$A(q^k)/q^k = [(b - 1)/2]k \cdot \log_b q + O(1) = k \cdot A(q)/q$$

lends some credence to the conjecture.

It is our purpose, in this paper, to examine the sets $I_\lambda(c)$. We shall show that for all integers $c$ and all bases $b \geq 8$, $I_\lambda(c)$ is indeed infinite. We show, further, that if $c \geq 0$, the set of values of $S(b, m)/S_p(b, m)$, $m$ an integer base $b$ ($b \geq 8$), is dense in the interval $[0, 1]$. This latter result implies the existence, for $b - 1$ composite, of infinitely many integers $m$ base $b$ such that for $n_1/n_2 \leq 1$, $S(b, m) = (n_1/n_2) \cdot S_p(b, m)$.

2. PRELIMINARY OBSERVATIONS AND NOTATION

All our references to integers are to integers base $b$ unless otherwise stated. We will, however, occasionally write $m_{(b)}$ to clarify that $m$ is an integer base $b$. We shall write $S(m)$ and $S_p(m)$ for $S(b, m)$ and $S_p(b, m)$, respectively, and we denote the number of digits, $n + 1$, of $m = \sum_{i=0}^{n} a_i b^i$ by $N(m)$. 
We observe that \( S_p(m) = S(m) \) if \( m \) is prime, and that
\[
S_p(m_1 m_2) = S_p(m_1) + S_p(m_2), \quad \text{if } m_1, m_2 > 1.
\]

It might be remarked that, in addition to Smith numbers, certain other integers may be distinguished according to their appearance in the sets \( I_b(c) \). Simple examples include the sets \( \{2M \mid M \text{ is a Mersenne prime} \} \), and \( \{F^2_m \mid F_m - 2^m + 1 \text{, a Fermat prime, } m \geq 1 \} \), each of whose elements are contained in \( I_2(1) \).

3. An Upper Bound on \( S_p(m) \)

Our approach in showing that \( I_b(c) \) is infinite involves obtaining, for a large class of integers \( m \) base \( b \) (actually, all integers when \( b-1 \) is composite), an upper bound on \( S_p(m) \) in terms of \( N(m) \), and then applying this bound to an infinite collection of integers \( m' \) for which \( S(m') \) can be obtained.

Let \( m = p_1 p_2 \cdots p_k \) be an integer base \( b \), \( b \geq 8 \), with \( p_1, \ldots, p_k \) primes not necessarily distinct; let \( B_i = N(p_i) - 1 \), for \( i = 1, 2, \ldots, k \), and \( B = B_1 + \cdots + B_k \). We observe that since an integer is congruent to its digit sum, modulo \( b-1 \), \( (b-1) \mid S(p_i) \) only if \( p_i = b-1 \). Hence, if \( p_i = b-1 \),
\[
S(p_i) = b-1 = (b-1) N(p_i) = (b-1) B_i + (b-1),
\]
and, if \( p_i \neq b-1 \),
\[
S(p_i) \leq (b-1) N(p_i) - 1 = (b-1) B_i + (b-2).
\]
Accordingly, we partition the prime factors of \( m \) into \( b \) disjoint classes by means of the following: Let \( r_i \) be defined by
\[
S(p_i) = (b-1) B_i - r_i, \quad r_i \geq -(b-1). \tag{1}
\]
One class consists of those primes \( p_i \) for which \( r_i \) is positive and the other classes of those primes \( p_i \) for which \( r_i = j, -(b-1) \leq j \leq -1 \). (Note that \( r_i \neq 0 \) for any \( i \).)

Let \( n_j \) be the number of integers \( i \) \((1 \leq i \leq k) \) such that \( r_i = j \), \(- (b-1) \leq j \leq -1\), and let \( A = \{ r_i \mid r_i > 0, 1 \leq i \leq k \} \). In light of the discussion above, it should be noted that \( n_{b-1} \) is the number of times \( b-1 \) occurs as a factor of \( m \) if \( b-1 \) is prime, and is 0 if \( b-1 \) is composite.

We will need the following lemma which is readily proved using elementary calculus:

**Lemma 1.** If \( b \geq 8 \) and \( x \) is a real variable, the function \( f(x) = (b-1) \log_b x - x \) is positive for \( 1 + \lfloor (b-1)/b \rfloor \leq x \leq b-2 \).
**Theorem 1.** If (i) $b - 1$ is composite, or (ii) $b - 1$ is prime and $n_{b-1} = 1$, then

$$S_p(m) \leq (b - 1) N(m) - \sum_A r_i.$$

**Proof.** We observe, first, that

$$S_p(m) = \sum_{i=1}^k S(p_i) = \sum_{i=1}^k [(b - 1) B_i - r_i]$$

$$= (b - 1) B + \sum_{j=1}^{b-1} jn_j - \sum_A r_i. \quad (2)$$

Now, for $-(b - 1) \leq r_i \leq -1$, $S(p_i) = (b - 1) B_i - r_i$ implies that $p_i \geq (-r_i + 1) b^{B_i - 1}$ if $B_i > 0$, and $p_i = -r_i b^{B_i}$ if $B_i = 0$. For our purposes, it is sufficient to use the estimates

$$p_i \geq -r_i b^{B_i}, \quad \text{for } -(b - 1) \leq r_i < -1$$

and

$$p_i \geq [-r_i + (b - 1)/b] b^{B_i}, \quad \text{for } r_i = -1.$$  

For the remaining prime factors (i.e., those $p_i$ for which $r_i > 0$), $p_i \geq b^{B_i}$. It follows that

$$m = p_1 p_2 \cdots p_k \geq [1 + (b - 1)/b]^{n_1} \cdot 2^{n_2} \cdot 3^{n_3} \cdots (b - 1)^{n_{b-1}} \cdot b^B.$$  

Rewriting $m$ as $a \cdot b^{N(m)-1}$, for some rational number $1 \leq a < b$, and taking logarithms, base $b$, we have

$$\log_b a + N(m) - 1 \geq n_1 \log_b [1 + (b - 1)/b] + \sum_{j=2}^{b-1} n_j \log_b j + B.$$  

Multiplying by $b - 1$, applying Lemma 1, and rearranging the terms gives us

$$(b - 1) B + \sum_{j=1}^{b-2} jn_j < (b - 1) [N(m) - 1 + \log_b a - n_{b-1} \log_b (b - 1)]. \quad (3)$$

Substituting (3) in (2),

$$S_p(m) < (b - 1) [N(m) + (n_{b-1} - 1) + \log_b a - n_{b-1} \log_b (b - 1)] - \sum_A r_i.$$

We now assume either that $b - 1$ is composite, in which case $n_{b-1} = 0$, or
DIFFERENCE OF DIGITAL SUMS

that \( b - 1 \) is prime and \( n_{b-1} = 1 \). The theorem is clearly proved if \( n_{b-1} = 0 \), and
is immediate if \( n_{b-1} = 1 \) since \( a < b \) implies that \( \log_b a - \log_b(b - 1) \leq 0 \).

An example to illustrate that Theorem 1 does not hold for all bases less
than 8 is afforded by \( m = 2^9 = 512_{(10)} = 4022_{(5)} \). In base 5, \( S_p(m) = 33_{(5)} \),
but \( (b - 1) N(m) = 4 \cdot 4 = 31_{(5)} \).

4. SOME ADDITIONAL PROPERTIES OF \( S_p(m) \) AND \( S(m) \)

We now obtain two results needed in proving the Main Theorem. We
assume in this section that \( b \geq 2 \).

**Lemma 2.** If \( m \) base \( b \) is a positive integer \( > 1 \), there exists an integer \( t \)
such that \( S_p(t) = m \).

*Proof.* If \( b = 2 \), let \( t = 10^n \). If \( b > 2 \), there exist integers \( q \) and \( r \) such
that \( m = 2^q + r, r = 0 \) or 1. If \( r = 0 \), let \( t = 2^q \). If \( r = 1 \), let \( t = 2^q \cdot 10^n \) if
\( b = 3 \), and let \( t = 2^{q-1} \cdot 3 \) if \( b \geq 4 \).

**Corollary 1.** There exists a finite set \( T \) of integers base \( b \) such that the
set \( U = \{ S_p(t) \mid t \in T \} \) is \( \{ b, 3, ..., S_p(b) + 1 \} \).

**Lemma 3.** Let \( n \) and \( t \) be positive integers such that \( t \leq b^n - 1 \). If \( v \) is a
nonnegative integer and \( m = t(b^n - 1) b^v \), then \( S(m) = (b - 1)n \).

*Proof.* Let \( t = \sum_{j=1}^n a_j b^j \), \( 0 \leq a_j \leq b - 1 \), \( a_k > 0 \), \( a_r > 0 \), \( 0 \leq k \leq r \). We
adopt the convention that \( \sum_{j=k+1}^{n} f(j) = 0 \) if \( k = r \). If, in the product

\[
(t(b^n - 1)) = a_k b^{n+k} + \sum_{j=k+1}^{r} a_j b^{n+j} - a_k b^k - \sum_{j=k+1}^{r} a_j b^j,
\]

we replace \( a_k b^{n+k} \) by

\[
(a_k - 1) b^{n+k} + \sum_{j=k+1}^{n+k-1} (b - 1) b^j + \sum_{j=k+1}^{r} (b - 1) b^j + b \cdot b^k,
\]

we obtain

\[
t(b^n - 1) b^v = \left[ \sum_{j=k+1}^{r} a_j b^{n+j} + (a_k - 1) b^{n+k} + \sum_{j=k+1}^{n+k-1} (b - 1) b^j \right.
\]

\[
+ \sum_{j=k+1}^{r} (b - 1 - a_j) b^j + (b - a_k) b^k \] \cdot b^v.
\]

The coefficient of each power of \( b \) is a nonnegative integer less than \( b \), and
the digital sum \( S(m) \) is now readily seen to be \( (b - 1)n \).
5. THE SET $I_b(c)$ AND THE VALUES OF $S(m)/S_p(m)$

**Main Theorem.** Let $c$ be any integer base $b$, $b \geq 8$. The set $I_b(c)$ is infinite.

**Proof.** Let $u$ be a positive integer such that $c \geq 2 - u$. By Bertrand’s Postulate, there exists a prime $q$ such that $b < q < 2b - 2$. Let $n \equiv 0 \pmod{\varphi(q^u)}$ ($\varphi$ is the Euler phi-function), $n \equiv 0 \pmod{(b - 1)}$ (this is possible, since $(b - 1 \nmid \varphi(q^u)$), and $n$ be such that $b^n - 1$ exceeds the maximum element of $T$ (recalling that $T$ is finite). Let $M = b^n - 1$. The hypothesis of Theorem 1 is clearly satisfied: $(b - 1) \mid M$, but $(b - 1)^2 \nmid M$ since

$$M/(b - 1) = n \not\equiv 0 \pmod{(b - 1)}$$

(whether $b - 1$ is prime or not). Now, since $n \equiv 0 \pmod{\varphi(q^u)}$, $q^u$ is a prime power factor of $M$. The inequality

$$b + 1 \leq q \leq 2b - 3 = b^1 + (b - 3)$$

implies that

$$2 \leq S(q) \leq 1 + (b - 3) = b - 2;$$

setting $q = p_i$ in (1), for $i = 1, 2, ..., u$, we have

$$r_i = (b - 1) \cdot B_i - S(p_i) \geq (b - 1) \cdot 1 - (b - 2) = 1.$$ 

Thus, by Theorem 1, $S_p(M) \leq (b - 1)n - u$. Let $h = (b - 1)n - S_p(M) \geq u$. Since $h + c \geq u + c \geq 2$, and since the set $U$ of Corollary 1 is a complete residue system, modulo $S_p(b)$, there exists an integer $t \in T$ such that, for some nonnegative integer $v$,

$$S_p(t) = h + c - v \cdot S_p(b).$$

Let $m = tMb^v = t(b^n - 1) b^v$. Since $t < b^n - 1$, the hypothesis of Lemma 3 is satisfied and we have $S(m) = (b - 1)n$. Thus,

$$S_p(m) = S_p(t) + S_p(b^n - 1) + S_p(b^v)$$

$$= [h + c - v \cdot S_p(b)] + [(b - 1)n - h] + v \cdot S_p(b)$$

$$= (b - 1)n + c$$

$$= S(m) + c.$$ 

Therefore, $m \in I_b(c)$. Now, infinitely many choices for $n$ exist and each
determines a unique \( m \), which is clearly composite since \( b^n - 1 \) is composite, so \( I_p(c) \) is infinite.

We now define \( \alpha(m) = \frac{S(m)}{S_p(m)} \), for \( m \) base \( b \) an integer \( > 1 \).

**Theorem 2.** The set \( D = \{ \alpha(m) \mid m \text{ base } b \text{ any integer } > 1 \} \cap [0, 1] \), \( b \geq 8 \), is dense in the interval \( [0, 1] \).

**Proof.** Let \( 0 < x < y < 1 \). Let \( q \) be a prime such that \( b < q < 2b - 2 \). There exists a rational number \( \frac{n_1}{n_2} \) between \( x \) and \( y \) such that \( (b - 1) \frac{q - 1}{n_1} \leq y \). (If \( x \leq m_1/m_2 < m_3/m_4 \leq y \),

\[ \frac{n_1}{n_2} = \frac{[(b - 1) m_1 m_4 + 1]}{[(b - 1) m_2 m_4]} \]

is between \( m_1/m_2 \) and \( m_3/m_4 \).) Let \( m \) be defined as in the Main Theorem with \( u = 1 \), \( n = (q - 1)n_1 \), and \( c = (b - 1)(q - 1)(n_2 - n_1) \). Then

\[ \alpha(m) = \frac{S(m)}{S_p(m)} = \frac{S(m)}{[S(m) + c]} = \frac{n_1}{n_2}. \]

It follows that \( D \) is dense in \( [0, 1] \).

**Corollary 2.** If \( b - 1 \) is composite, there exists an infinitude of integers \( m \) base \( b \) such that, for any positive rational number \( \frac{n_1}{n_2} \leq 1 \), \( S(m) = (n_1/n_2) S_p(m) \).

**Proof.** Let \( k \) be any positive integer and \( \frac{n_1}{n_2} \) be any rational number \( \leq 1 \). Since the restriction that \( (b - 1) \frac{n}{k} \) is not necessary when \( b - 1 \) is composite, we obtain \( \alpha(m) = \frac{n_1}{n_2} \) by letting \( n = k(q - 1)n_1 \) and \( c = k(b - 1)(q - 1)(n_2 - n_1) \).

**Acknowledgment**

The author expresses his appreciation for the referee's helpful comments and suggestions.

**References**