Collisions of the binary and ternary sum-of-digits functions

Lukas Spiegelhofer
Montanuniversität Leoben, Austria

Abstract

We prove a folklore conjecture concerning the sum-of-digits functions in bases two and three: there are infinitely many positive integers \( n \) such that the binary sum of digits of \( n \) equals its ternary sum of digits.

1 Introduction and the main result

Representations of the same number \( x \) in two or more multiplicatively independent integer bases apparently look very different. This topic is far from being understood, and the relation of the base-\( q_1 \) and the base-\( q_2 \) expansion to each other is a source of difficult problems.

The base-\( q \) expansion is intimately connected to powers of \( q \). In order to understand the relation of different bases \( q_1 \) and \( q_2 \) to each other better we consider, as a start, the arrangement of powers of 2 and 3. Assume that the set containing all powers of two and three (with nonnegative exponents) is ordered in ascending order:

\[(a_n)_{n\geq0} = (1, 2, 3, 4, 8, 9, 16, 27, 32, 64, 81, 128, 243, 256, 512, 729, 1024, \ldots)\]

(this is sequence \( \text{A006899} \) in the OEIS [52]). In which way are the powers of two and three interleaved? Taking logarithms, we see that the question is encoded in the Sturmian word

\[w = (\lfloor (n+1)\alpha \rfloor - \lfloor n\alpha \rfloor)_{n\geq0},\]

where \( \alpha = \log 3/\log 2 = \log_2(3) \), as follows: start with \( 3^0 = 1 \), append the first \( w_0 = 1 \) powers of two — that is, the integer 2 — append \( 3^1 \), then \( w_1 = 2 \) powers of two, followed by \( 3^2 \) and \( w_2 = 1 \) powers of two, and so on. Our question is therefore equivalent to understanding the continued fraction expansion of \( \alpha \) (consult, for example, Berthé [5] for an explanation of this connection). However, it is not even known whether the sequence of partial quotients of \( \alpha \) is bounded, that is, whether \( \alpha \) is badly approximable; any system in this sequence has yet to be found. The number \( \alpha \) is transcendental by the Gelfond–Schneider theorem [22, 23]; by Baker’s theorem [2, 3, 4] we obtain

\[\left|\frac{\log 3}{\log 2} - \frac{p}{q}\right| \geq \frac{c}{q^\rho}\]

for all integers \( q > 0 \) and \( p \) and some effective positive constants \( c \) and \( \rho \). More precisely, a bound for the irrationality measure \( \mu(\alpha) \) of \( \alpha \), which is the infimum of \( \rho \) for which there exists \( c \) such that this estimate holds for all \( p, q \), was given by Rhin [46] Equation (8): we have \( \mu(\alpha) \leq 8.616 \). We also note that Wu and Wang [58] obtained the bound \( \mu(\log 3) \leq 5.1163051 \). Note that...
badly approximable numbers have irrationality measure 2. We would also like to mention the interesting blog entry by Tao\footnote{https://terrytao.wordpress.com/2011/08/21/hilberts-seventh-problem-and-powers-of-2-and-3/} on the topic.

In view of the above problem we have to expect major difficulties when we try to mix different bases. In this context, we mention the following unsolved conjecture of Furstenberg\cite{furstenberg} concerning multiplicatively independent integer bases $p, q \geq 2$ (that is, such that $p^k \neq q^\ell$ for all $k, \ell \geq 1$):

\[ O_p(x) := \{a^k x \mod 1 : k \in \mathbb{N}\} \]

and let $\dim_H(A)$ be the Hausdorff measure of a set $A \subseteq [0, 1]$. Then

\[ \dim_H(O_p(x)) + \dim_H(O_q(x)) \geq 1 \tag{1} \]

for all irrational $x \in [0, 1]$. This conjecture underlines the idea stated before: different bases should produce very different representations of the same number. We note the papers\cite{Adamczewski2011,Faverjon2012} for recent progress on this conjecture, and the recent preprint\cite{Adamczewski2021} by Adamczewski and Faverjon, where related independence results can be found.

Erdős\cite{erdos} conjectured that the only powers of two having no digit 2 in its ternary expansion are 1, 4, and 8 (see also Lagarias\cite{lagarias}). This conjecture is open, and Erdős wrote “[…] as far as I can see, there is no method at our disposal to attack this conjecture.”\cite{erdos}. Meanwhile, there is a close connection to Erdős’ squarefree conjecture\cite{erdos-squarefree}, stating that the central binomial coefficient \( \binom{2n}{n} \) is never squarefree for $n \geq 5$. This latter conjecture was proved for all large $n$ by Sárközy\cite{sarkozy}, and solved completely by Granville and Ramaré\cite{granville-ramare}. In fact, the connection between these two conjectures can be understood by the identities

\[ \nu_2 \left( \binom{2n}{n} \right) = s_2(n) \quad \text{and} \quad \nu_3 \left( \binom{2n}{n} \right) = s_3(n) - \frac{s_3(2n)}{2}, \]

where $s_q$ is the sum-of-digits function in base $q$, and $\nu_p$ is the $p$-adic valuation of an integer $\geq 1$. That is, $\binom{2n}{n}$ is divisible by the square 4 if $n \geq 1$ is not a power of two, and so the (already proved) squarefree conjecture would follow from a proof of the observation that $s_3(2^k) - s_3(2^{k+1})/2 \geq 2$ for $k \geq 9$. This in turn would follow if we could prove that $2^k$ contains at least two digits equal to 2 in ternary for $k \geq 9$: in this case at least two carries appear in the addition $2^k + 2^k$ in ternary. We also would like to note the recent preprint\cite{dimitrov-howe} by Dimitrov and Howe on this topic.

The main objects in this paper are the sum-of-digits functions $s_2$ and $s_3$. For a nonnegative integer $n$ and a base $q$, the integer $s_q(n)$ is in fact the minimal number of powers of $q$ needed to represent $n$ as their sum. (This small fact can be proved using that the $q$-ary expansion is the lexicographically largest representation of $n$ as sum of powers of $q$.)

Senge and Straus\cite{senge-sinus} proved the important theorem that for coprime integers $p, q \geq 2$ and arbitrary $c > 0$, there are only finitely many integers $n \geq 0$ such that

\[ s_p(n) \leq c \quad \text{and} \quad s_q(n) \leq c. \tag{2} \]

This statement is, at least heuristically, close to Furstenberg’s conjecture\cite{furstenberg}. This result was extended by Stewart\cite{stewart}, Mignotte\cite{mignotte}, Schlickewei\cite{schlickewei, schlickewei-2}, Pethő–Tichy\cite{petho-tichy}, and Ziegler\cite{ziegler}. See also\cite{stewart, mignotte, schlickewei, schlickewei-2, petho-tichy, ziegler} for related results.

Gelfond\cite{gelfond} proposed to prove that

\[ \# \{ n \leq x : s_{q_1}(n) \equiv \ell_1 \mod m_1 \text{ and } s_{q_2}(n) \equiv \ell_2 \mod m_2 \} = \frac{x}{m_1m_2} + O(x^\delta) \tag{3} \]
for some \( \delta < 1 \), where \( q_1, q_2 \geq 2 \) are coprime bases, \( m_1, m_2 \) are integers satisfying \( \gcd(m_1, q_1 - 1) = \gcd(m_2, q_2 - 1) = 1 \), and \( \ell_1, \ell_2 \in \mathbb{Z} \). A weak error term \( o(1) \) for this problem was proved by Bésineau [6], while the full statement was obtained by D.-H. Kim [26].

Drmota [13, Theorem 4] proved (among other things) an asymptotic formula for the numbers

\[ \frac{1}{x} \# \{ n < x : s_p(n) = k, s_q(n) = \ell \}, \]

where \( p, q \geq 2 \) are coprime bases, with an error term \((\log x)^{-1}\). This may be called a local limit theorem for the joint sum-of-digits function \( n \mapsto (s_p(n), s_q(n)) \). Note that this includes Bésineau’s result as a special case, as the two sum-of-digits functions on \([0, x)\) are concentrated around their expected values (compare (52) below).

We also wish to note the recent paper by Drmota, Mauduit, and Rivat [16], who proved a result on the sum of digits of prime numbers in two different bases.

The starting point for the present paper is the article [11] by Deshouillers, Habsieger, Landreau, and Laishram.

“[. . . ] it seems to be unknown whether there are infinitely many integers \( n \) for which \( s_2(n) = s_3(n) \) or even for which \( |s_2(n) - s_3(n)| \) is significantly small.” [11]

They prove the following result.

**Theorem.** For sufficiently large \( N \), we have

\[ \# \{ n \leq N : |s_3(n) - s_2(n)| \leq 0.1457205 \log n \} > N^{0.970359}. \]

Note that the difference \( s_3(n) - s_2(n) \) is expected to have a value around \( C \log n \), where

\[ C = \frac{1}{\log 3} - \frac{1}{\log 4} = 0.18889 \pm 10^{-5}; \]

by the above theorem there exist indeed many integers \( n \) such the difference \( |s_2(n) - s_3(n)| \) is “significantly small”.

This result was extended by de la Bretèche, Stoll, and Tenenbaum [10], who proved in particular that

\[ \{ s_p(n)/s_q(n) : n \geq 1 \} \]

is dense in \( \mathbb{R}^+ \) for all multiplicatively independent integer bases \( p, q \geq 2 \).

We also wish to mention the papers [35] by Mauduit and Sárközy, and Mauduit, Pomerance, and Sárközy [33]. In these papers, integers with a fixed sum of digits and corresponding asymptotic formulas are studied, and possible extensions to several bases are mentioned.

The question on the infinitude of collisions of \( s_2(n) \) and \( s_3(n) \), mentioned in [11], is not a new one; M. Drmota (private communication) received a hand-written letter from A. Hildebrand more than twenty years ago, in which the very same problem was presented.

In the present paper, we give a definite answer to this question.

**Theorem 1.1.** There exist infinitely many nonnegative integers \( n \) such that

\[ s_2(n) = s_3(n). \]

More precisely, for all \( \varepsilon > 0 \) we have

\[ \# \{ n < N : s_2(n) = s_3(n) \} \gg N^{\frac{\log 3}{\log 4} - \varepsilon}. \]

Note that \( \log 3/\log 4 = 0.792 \pm 10^{-3} \).
1 INTRODUCTION AND THE MAIN RESULT

The difficulty in proving this theorem lies in the separation of the values of \( s_2(n) \) and \( s_3(n) \). The sum-of-digits functions can be thought of as a sum of independent, identically distributed random variables, and they concentrate (according to Hoeffding’s inequality, for example) around the values \( \frac{1}{2} \log_2 N \) and \( \log_3 N \) respectively, where \( 0 \leq n < N \). More precisely, the variances are of order \( \log N \), and the tails of these distributions decay as fast as \( \exp(-C(x - \mu)^2/\sigma^2) \), where \( \mu \) is the expected value, and \( \sigma^2 \) the variance. Since the gap \( \log N(1/\log 3 - 1/\log 4) \) comprises \( (\log N)^{1/2} \) standard deviations, we can only expect a number \( \ll N^{\delta} \) of collisions, where \( \delta < 1 \) is some constant. In light of this argument, our result is certainly not too far from the true number of collisions.

The increasing sequence \( s_{2,3} \) of nonnegative integers \( n \) such that \( [6] \) holds is listed as entry A037301 in the OEIS [52]. The question whether this sequence is infinite had to remain open there. The first few collisions are as follows:

\[
\begin{align*}
n & \text{in binary} \quad 0 & 1 & 110 & 111 & 1010 & 1011 & 1100 & 1101 & 10010 & 10011 & 10101 & 100100 \\
n & \text{in ternary} \quad 0 & 1 & 20 & 21 & 101 & 102 & 110 & 111 & 200 & 201 & 210 & 1100 \\
n & \text{in decimal} \quad 0 & 1 & 6 & 7 & 10 & 11 & 12 & 13 & 18 & 19 & 21 & 36.
\end{align*}
\]

Note the subsequence \( (10, 11, 12, 13) \); contiguous subsequences of \( N \) of length greater than four do not appear in \( s_{2,3} \), since \( s_3 \) on such a subsequence contains two consecutive up-steps, while \( s_2 \) decreases or stays constant after one up-step.

Remark. We expect that it is possible to extend our proof to patterns in \( s_{2,3} \): for example, we expect that there are infinitely many \( n \) such that

\[ s_2(n + v) = s_3(n + v) \quad \text{for} \quad v \in \{0, 1, 2, 3\}, \quad (8) \]

and more generally, every pattern that occurs at all should occur infinitely often. To this end, we will have to study certain residue classes modulo \( 2^k 3^l \) — note that for \( n \equiv 2 \mod 8 \) and \( n \equiv 1 \mod 9 \), for example, we have \( s_2(n + v) - s_3(n + v) = c \) for some \( c \) and all \( v \in \{0, 1, 2, 3\} \). The next step would be to scan these “candidate residue classes” for collisions, using our method. But residue classes of this form are used in our proof anyway, so we are optimistic that the main work has already been done. (Note that a suitable replacement for Proposition 2.1 below will have to be found. This takes care of the parity restriction \( s_3(n + t) - s_3(n) \equiv s_3(t) \mod 2 \).

We would like to note that our proof of Theorem 1.1 is not a constructive one. We do not give an algorithm that allows us to find integers \( n \) such that \( s_2(n) = s_3(n) \). We leave it as an open problem to find a construction method for such integers \( n \).

Also, it would be very interesting to obtain \( s_2(p) = s_3(p) \) for infinitely many prime numbers \( p \). We believe that this question is a difficult one; as in the case of missing digit problems, the set \( A = \{ n : s_2(n) = s_3(n) \} \) is sparse (that is, \( \#(A \cap [1, N]) \ll N^\delta \) for some \( \delta < 1 \)). Maynard [39], in an important and difficult paper, could prove that infinitely many primes excluding any given decimal digit exist.

Plan of the paper. The main body of the paper concerns the proof of our auxiliary statement, Proposition 2.1 which directly leads to the main theorem. This proof is organized into three stages, considered in Sections 2.1, 2.2 and 2.3. In Section 2.4, these steps are combined. In the same section, we finish the proofs with an argument leading to the quantitative statement (7). At the end of the paper, we present (mostly difficult) research questions.

Notation. The symbol \( \log \) denotes the natural logarithm, and \( \log_a = \frac{1}{\log_a} \log \) is the logarithm in base \( a > 1 \). We use Landau notation, employing the symbols \( O, \ll, \) and \( o \). The symbol \( f(n) \asymp g(n) \) abbreviates the statement \( (f(n) = O(g(n))) \) and \( g(n) = O(f(n)) \), while \( f(n) \sim g(n) \) means that \( f(n)/g(n) \) converges to 1 as \( n \to \infty \).
2 Proofs

Our main theorem follows from the following proposition.

**Proposition 2.1.** There exists infinitely many positive integers \( n \) such that \( n \equiv 9 \mod 12 \) and
\[
s_2(n) - s_3(n) \in \{0, 1\}. \tag{9}
\]

It is easy to derive Theorem 2.4 from this result. If there are infinitely many solutions of \( s_2(n) - s_3(n) = 0 \) such that \( n \equiv 9 \mod 12 \), we are done. Otherwise, we note that \( n \equiv 9 \mod 12 \) is equivalent to \( (n \equiv 0 \mod 3 \text{ and } n \equiv 1 \mod 4) \), therefore \( s_3(n + 1) = s_3(n) + 1 \) and \( s_2(n + 1) = s_2(n) \). The existence of infinitely many solutions of \( s_2(n) - s_3(n) = 1 \) on \( 9 + 12\mathbb{Z} \) therefore implies infinitely many collisions on the residue class \( 10 + 12\mathbb{Z} \).

Let \( N \geq 4 \) be an integer. We are going to find many collisions in the interval \([N, 2N] \) for all large enough \( N \), which will prove Theorem 2.4. Let \( \varepsilon > 0 \) be arbitrary. It will be used as exponent of \( \log N \), and the precise choice will be irrelevant for the proof. Set
\[
\tilde{\lambda} = \log N, \quad \tilde{\eta} = \tilde{\lambda}^{3/4}, \quad \tilde{f} = (\log \tilde{\lambda})^{1/2 + \varepsilon}, \quad \tilde{m} = \tilde{\lambda}^{1/2} / \tilde{f}, \quad \tilde{J} = \tilde{f}^2, 
\[
\lambda = \lfloor \tilde{\lambda} \rceil, \quad \eta = 4 \lfloor \tilde{\eta} / 4 \rceil + 2, \quad f = \lfloor \tilde{f} \rfloor, \quad m = \lfloor \tilde{m} \rfloor, \quad J = \lfloor \tilde{J} \rfloor. \tag{10}
\]

The search for collisions will consist of three main steps.

1. Study constant differences \( f(n + t) - f(n) \) for \( n \) in a certain residue class, where \( f(n) = s_2(n) - s_3(n) \);
2. Concentrate the values of \( f(n) \) into the interval \([−Jm, Jm] \) by rarefying and truncating the residue class;
3. Select only those \( n \) such that \( f(n) \) lies in \( m\mathbb{Z} \).

Steps 2 and 3 yield many values of \( n \) such that \( f(n) \in \{-Jm, (-J + 1)m, \ldots, Jm\} \). The purpose of Step 1 is to find in advance a residue class \( A' = L + 2^\eta 3^\beta \mathbb{Z} \) and shifts \( d_j \) such that \( f(n + d_j) - f(n) = jm + \xi_j \) for some \( \xi_j \in \{0, 1\} \) and all \( j \in \{-J, \ldots, J\} \). This procedure yields many \( n \) such that \( f(n) \in \{0, 1\} \), by selecting for each good index \( n \) the appropriate shift \( d_j \). The harmless argument presented after Proposition 2.4 allows us to get rid of the unpleasant correction term \( \xi_j \).

2.1 Constant differences of the sum-of-digits function

We will use blocks in ternary, whose lengths are given by the integer \( \eta \). Let us choose nonnegative integers \( d_{−J}, d_{−J+1}, \ldots, d_J \) by concatenating such blocks of ternary digits. Set
\[
b = (01^{\eta−2}0)_3 = \frac{3^{\eta−1}−3}{2}
\]
(where \( 1^{\eta−1} \) denotes \( \eta − 1 \) repetitions of the digit 1) and
\[
d_{−J} = b, 
\]
\[
d_j = d_{j−1} + 3^{(j+J)\eta}b \quad \text{for} \quad −J < j \leq J. \tag{11}
\]

The idea is that the zero between the blocks of ones avoids carry propagation between the blocks, while the \( \eta − 2 \) ones allow for obtaining any even integer \( d \in [−\eta/2 + 1, \eta/2 − 1] \) as a difference
\[
d = s_3(a + b) − s_3(a) \tag{12}
\]
with some $0 \leq a < 3^\nu$ (see the calculation leading to (13) below). Moreover, since the ternary expansion of $d_j$ consists of blocks 1111 and 0, and ends with 0, we have $d_j \equiv 0 \mod 12$ (note that $4 \mid (1111)_3 = 40$). Choose the integer $\nu \geq 1$ minimal so that
\[ 2^{\nu-1} \geq 3^{(2j+1)\eta}. \] (13)

In particular,
\[ d_j < 2^{\nu-1}. \] (14)

The next important step consists in choosing a certain integer $a \in \{1, \ldots, 2^{\nu-1} - 1\}$; its meaning will become clear in a moment. The size restrictions imply $d_j + a < 2^\nu$ for all $j \in \{-J, \ldots, J\}$. This means in particular that no carry from the $(\nu - 1)$th to the $\nu$th digit occurs in the addition $d_j + a$, which implies the simple but important identity
\[ s_2(2^\nu n + a + d_j) - s_2(2^\nu n + a) = s_2^{(\nu)}(a + d_j) - s_2^{(\nu)}(a) \] (15)
for all $n \geq 0$. The function defined by $s_2^{(\nu)}(n) = s_2(n \mod 2^\nu)$ is the truncated binary sum-of-digits function. Note that the right hand side of (15) is independent of $n$.

Our next goal is to use a concentration argument (Chebychev’s inequality) for choosing a value $a$ such that the right hand side of (15) is small for all $j \in \{-J, \ldots, J\}$.

We adapt parts from [54]. For integers $t, L \geq 0$ and $j$, we set
\[ \varphi(j, t, L) = \frac{1}{2^L} \#\{0 \leq n < 2^L : s_2^{(L)}(n + t) = s_2^{(L)}(n) + 1\}, \] (16)
and we define the characteristic function
\[ \omega_t(\vartheta, L) = \sum_{j \in \mathbb{Z}} \varphi(j, t, L) e(\vartheta j) = \frac{1}{2^L} \sum_{0 \leq n < 2^L} e(\vartheta s_2^{(L)}(n + t) - \vartheta s_2^{(L)}(n)), \] (17)
where $e(x) = \exp(2\pi i x)$. Noting that
\[ s_2^{(L+1)}(2n) = s_2^{(L)}(n) \quad \text{and} \quad s_2^{(L+1)}(2n + 1) = s_2^{(L)}(n) + 1, \] (18)
the proof of the following statement is not difficult and left to the reader.

**Lemma 2.2.** For all $t, L \geq 0$ and $j \in \mathbb{Z}$ we have
\[ \varphi(1, j, L) = \begin{cases} 2^{j-2}, & 2 - L \leq j \leq 1; \\ 2^{-L}, & j = -L; \\ 0, & \text{otherwise,} \end{cases} \] (19)
\[ \varphi(j, 2t, L + 1) = \varphi(j, t, L), \]
\[ \varphi(j, 2t + 1, L + 1) = \frac{1}{2} \varphi(j - 1, t, L) + \frac{1}{2} \varphi(j + 1, t + 1, L). \]

The characteristic function satisfies
\[ |\omega_t(\vartheta, L)| \leq 1, \quad \omega_{2t}(\vartheta, 2L) = \omega_t(\vartheta, L), \]
\[ \omega_{2t+1}(\vartheta, 2L) = \frac{e(\vartheta)}{2} \omega_t(\vartheta, L) + \frac{e(-\vartheta)}{2} \omega_{t+1}(\vartheta, L) \quad \text{for } t \geq 1. \] (20)
The recurrence (20) leads to a recurrence for the moments $m_k(t, L)$ of the probability mass function $\varphi(t, \nu, L)$.

For all $k \geq 0$, $t \geq 1$, and $L \geq 0$, we have

\begin{align}
m_1(t, L) &= 0, \\
m_2(1, L) &= 2 - 2^{-L+1}, \\
m_k(2t, L + 1) &= m_k(t, L), \\
m_2(2t + 1, L + 1) &= 1 + \frac{m_2(t, L) + m_2(t + 1, L)}{2}. \\
\end{align}

(21)

Of course, recurrences for the higher moments exist too, but we only need the first and second moments.

In analogy to Corollary 2.3 in the preprint [55] we obtain the following result.

**Lemma 2.3.** There is a constant $C$, independent of $t$ and $L$, such that

$$m_2(t, L) \leq CB,$$

(22)

where $B$ is the number of blocks of 1s in $t$.

However, we only need a weaker version, which follows directly from (21): we have

$$m_2(t, \nu) \leq 2\nu \text{ for all } t, \nu \geq 1 \text{ such that } t < 2\nu.$$  

(23)

In particular, this holds for $t = d_j$ defined in (11), and for this estimate we do not need to know what $d_j$ looks like in binary. We are interested in the differences on the right hand side of (15).

By Chebychev’s inequality and (23), the number of integers $a$, $0 \leq a < 2\nu$, such that

$$\left| s^{(\nu)}_2(a + d_j) - s^{(\nu)}_2(a) \right| \leq R_2(2\nu)^{1/2},$$

(24)

is bounded below by

$$2^\nu(1 - 1/R_2^2).$$

Intersecting $2J + 1$ sets, we obtain the set of $a < 2\nu$ that satisfy (24) for all $j \in \{-J, \ldots, J\}$, having cardinality $\geq 2^\nu(1 - (2J + 1)/R_2^2)$. We choose $R_2 = \lambda/(2\nu)$, which is asymptotically $\asymp \lambda^{1/8}/(\log \lambda)^{1/2+\varepsilon}$. It follows that the set of $a \in \{0, \ldots, 2\nu - 1\}$ satisfying

$$\left| s^{(\nu)}_2(a + d_j) - s^{(\nu)}_2(a) \right| \leq \lambda^{1/2} \text{ for all } j \in \{-J, \ldots, J\}$$

(25)

has at least

$$2^\nu(1 - \mathcal{O}((\log \lambda)^{2+4\varepsilon}\lambda^{-1/4}))$$

elements, by the definitions (19). Since powers win against logarithms for large $N$, we obtain some integer $a$ with the properties that

\begin{align}
a &\equiv 1 \mod 4, \\
0 &\leq a < 2^{\nu-1}, \text{ and} \\
|\delta_j| &\leq \lambda^{1/2} \text{ for all } j \in \{-J, \ldots, J\},
\end{align}

(26)

where

$$\delta_j := s^{(\nu)}_2(a + d_j) - s^{(\nu)}_2(a).$$

(27)
Note that the first two restrictions in (28) will pose no problem since asymptotically almost all \( a < 2^\nu \) satisfy the third.

By (13) we have therefore found a good arithmetic progression

\[ A = a + 2^\nu N \]  

such that each of the sequences

\[ \sigma_j = (s_2(m + d_j) - s_2(m))_{m \in A}, \]

for \(-J \leq j \leq J\), is constant, and attains a value \( \delta_j \) bounded by \( \lambda^{1/2} \) in absolute value.

In the next step, the ternary sum of digits will come into play, and we rarefy the progression \( A \) by a factor \( 3^\beta \), where

\[ \beta = (2J + 1)\eta. \]  

(29)

Note that \( \eta \approx \lambda^{3/4} \) has been used in the definition (11) of the values \( d_j \) before. The selection of this subsequence has to be done with care, so that certain differences

\[ f(n) = s_2(n) - s_3(n), \]  

(30)

are attained on this rarefied progression for \(-J \leq j \leq J\). Sure enough, in order to obtain these differences we will have to “repair” the deviation \( \delta_j \) from 0 caused by the differences of binary sums of digits. We are going to select a residue class \( B = K + 3^\beta N \), where \( K < 3^\beta \), on which certain differences

\[ s_3(n + d_j) - s_3(n) \]  

(31)

occur for \( n \in B \). This process will be carried out step by step, thinning out the current residue class by a factor \( 3^\beta \) for each \( j \in \{-J, \ldots, J\} \). There are many ones in the ternary expansion of \( d_j \); we will see that these digits equal to 1 can be used to obtain the needed variation of the base-3 sum-of-digits function (31). We have found a certain arithmetic progression \( A \) in (28); a sub-progression \( A' \) of \( A \) having the desired boundedness property \( \| \delta_j \| \leq \lambda^{-1/2} \) in base 2 and difference property in base 3 given by (30) below will be obtained by the intersection

\[ A \cap B = (a + 2^\nu N) \cap (K + 3^\beta N) = L + 2^\nu 3^\beta N, \]  

(32)

where \( 0 \leq L < 2^\nu 3^\beta \). We need to find \( K \). This number will in fact be divisible by 3 — together with \( a \equiv 1 \mod 4 \) this leads to \( L \equiv 9 \mod 12 \). The construction is similar to the definition of \( d_j \), where we concatenated ternary expansions of length \( \eta \), given by \( b = (01^\nu 20)^3 \). We begin with the integer \( k_{-J} \). By our preparation, the quantity \( Jm + \delta_{-J} \) (of size \( \lambda^{1/2} \) times a logarithmic factor) is considerably smaller than \( \eta \) (of size \( \lambda^{3/4} \)).

We will see in a moment that the large number of 1s in \( b \) can be used to find some \( a \in \{0, \ldots, 3^{\nu - 1} - 1\} \) divisible by 3 and \( \xi \in \{0, 1\} \) such that

\[ s_3(a + b) - s_3(a) = Jm + \delta_{-J} - \xi. \]  

(33)

In fact, such an integer \( a \) is found by assembling blocks of length four of ternary digits (where no carry between these blocks occurs), using the following addition patterns in base 3:

\[
\begin{array}{c c c c}
0202 & 0200 & 0112 \\
+1111 & +1111 & +1111 \\
\hline
2020, & 2011, & 2000.
\end{array}
\]

We see that each block of length four can be used to obtain a variation \( \in \{-2, 0, 2\} \) of the ternary sum of digits; there are \( (\eta - 2)/4 \gg \lambda^{3/4} \) such blocks, while the needed variation is
\( \approx \lambda^{1/2}(\log \lambda)^{1/2+\varepsilon} \) and thus much smaller. Moreover, by construction (10), the integer \( \eta - 2 \) is divisible by four, so there are no phenomena due to trailing digits. Using any \( \xi \in \{0, 1\} \) and \( a < 3^{\eta-1} \) satisfying \( 3 \mid a \) and (39), we set

\[
k_{-j} := a \quad \text{and} \quad \xi_{-j} := \xi.
\]

By our construction using the addition patterns displayed above, the zeroth ternary digit of \( a \) will indeed be zero. This is the case since \( b \) is divisible by three, too. Trivially, we obtain

\[
s_3(k_{-j} + d_{-j}) - s_3(k_{-j}) = Jm + \delta_{-j} - \xi_{-j}.
\]

Since \( a < 3^{\eta-1} \), there does not appear a carry to the \( n \)th ternary digit in the addition \( k_{-j} + d_{-j} \). In this way we obtain independence between the contributions of the blocks.

Assume that \( k_{j-1} \) has already been defined, for some \(-J < j \leq J\). In analogy to the above, choose \( a \in \{0, \ldots, 3^{\eta-1} - 1\} \) and \( \xi \in \{0, 1\} \) in such a way that

\[
s_3(a + b) - s_3(a) = -m - \delta_{j-1} + \xi_{j-1} + \delta_j - \xi,
\]

and set

\[
k_j = k_{j-1} + 3^{(j+1)} a \quad \text{and} \quad \xi_j := \xi.
\]

Note that the target value satisfies \(-m - \delta_{j-1} + \xi_{j-1} + \delta_j - \xi \ll \lambda^{1/2} \), which is again small compared to the number of 1s in \( b \). Since carry propagation between blocks of length \( \eta \) is not possible by construction, we obtain, by concatenation of blocks of length \( \eta \) and a telescoping sum,

\[
s_3(k_j + d_j) - s_3(k_j) = -jm + \delta_j - \xi_j \quad \text{for all } j \in \{-J, \ldots, J\}.
\]

Finally, set \( K = k_j \) and note that \( \beta = (2J + 1)\eta \) according to (29), so that \( K < 3^2 \). By construction (note that the ternary digits of \( d_j \) from \((j + 1)\eta\) on are zero) we have

\[
s_3(K + d_j) - s_3(K) = -jm + \delta_j - \xi_j \quad \text{for all } j \in \{-J, \ldots, J\}.
\]

Similar to (15), noting that there is no carry propagation in base three to the \( \beta \)th digit in the addition \( K + d_j \), we have in fact

\[
s_3(n + d_j) - s_3(n) = -jm + \delta_j - \xi_j
\]

for all \( n \in K + 3^\beta \mathbb{N} \). Defining \( L \) by (32), we obtain the key property for the function \( f(n) = s_3(n) - s_2(n) \):

\[
f(n + d_j) - f(n) = jm + \xi_j \quad \text{for all } J \in \{-J, \ldots, J\} \text{ and } n \in A' = L + 2^n 3^\beta \mathbb{N}.
\]

By construction, the residue class \( L + 2^n 3^\beta \mathbb{Z} \) is a subset of both \( 3 \mathbb{Z} \) and \( 1 + 4 \mathbb{Z} \), therefore \( L \equiv 9 \mod 12 \). This small restriction will be used for a final correction of the function \( f \) by a value \( \in \{0, -1\} \) — clearly, we have \( s_2(n+1) - s_2(n) = 0 \) for \( n \equiv 1 \mod 4 \) and \( s_3(n+1) - s_3(n) = 1 \) for \( n \equiv 0 \mod 3 \).

### 2.2 Small values of \( f(n) \)

In this section, we rarefy \( A' \) once more in order to obtain concentration of the values \( f(n) \) around a value near zero. Such a concentration property can only be satisfied for a finite portion of any arithmetic progression, since the expectation of \( f \) along any arithmetic progression will diverge to infinity. The fact that the expectation can be shifted close to zero is an essential point. It
is based on the consideration that $3^\tau n$ has the same ternary sum of digits as $n$ for all integers $\tau \geq 0$, while the binary sum of digits — usually — increases considerably under multiplication by $3^\tau$. This small remark is in fact the main idea that started the research on the present paper.

By our key property (11) it is sufficient to prove the existence of (many) elements $n \in \mathcal{A}'$ such that

$$f(n) \in Q, \quad \text{where } Q = \{jm : -J \leq j \leq J\}. \quad (42)$$

After all, for each $n$ satisfying (42), we can correct the value of $f$, up to a correction term $\in \{0, 1\}$, by any amount $c \in Q$ using the appropriate shift $D(n) \in \{d_{-J}, d_{-J+1}, \ldots, d_J\}$ in order to arrive at $f(n + D(n)) \in \{0, 1\}$. That is, there are many solutions to $s_2(n) - s_3(n) \in \{0, 1\}$, for certain $n \in [N + \min Q, 2N + \max Q]$. Since for each given $N$, the constructed quantities $d_j$ are smaller than half the common difference of $\mathcal{A}'$ — by (11), we have $d_j < 2^\nu - 1 \leq \frac{1}{2}2^\nu 3^\beta$ — this will show that there are infinitely many solutions to $s_2(n) - s_3(n) \in \{0, 1\}$, and in fact we will give a quantitative lower bound. Proving that (42) has many solutions in $\mathcal{A}'$ will be the subject of this and the following section (Section 2.3), constituting the second and third main steps in our proof. In the present section we are concerned with restricting our residue class $\mathcal{A}'$ in order to obtain $f(n) \in [-Jm, Jm]$ for many integers $n$ in the new class $\mathcal{A}'$. The third step will consist in the study of the property $f(n) \in m\mathbb{Z}$, which will be carried out in Section 2.3.

We set

$$\mathcal{A}'' = L + 2^\nu 3^\beta + \zeta \mathbb{Z} \quad (43)$$

for a natural number $\zeta$ that will be chosen in due course. Appropriate choice of $\zeta$ will shift the center of mass of $f(n)$ into the interval $[-Jm, Jm]$. At this point we only note that $3^\zeta$ will be much larger than $2^\nu$ and $3^\beta$. In orders of magnitude we have $\nu \asymp \beta \asymp \lambda^{3/4}(\log \lambda)^{1+2\epsilon}$, while $\zeta \asymp \lambda$. Trivially, (11) is satisfied on the subsequence $\mathcal{A}''$ too. We are therefore interested in the expression

$$f(L + 2^\nu 3^\beta + \zeta k) = s_2(L + 2^\nu 3^\beta + \zeta k) - s_3(L + 2^\nu 3^\beta + \zeta k), \quad (44)$$

where $k$ varies in the interval $I$ defined by

$$I = \{k \in \mathbb{N} : N \leq 2^\nu 3^\beta + \zeta k < 2N\}. \quad (45)$$

We can decompose (44) in the form

$$f(L + 2^\nu 3^\beta + \zeta k) = s_2(b_2 + 3^\beta + \zeta k) - s_3(b_3 + 2^\nu k) + s_2(r_2) - s_3(r_3), \quad (46)$$

where

$$b_2 = [2^{-\nu}L] \quad \text{and} \quad b_3 = [3^{-\beta - \zeta}L],$$

$$r_2 = L \mod 2^\nu \quad \text{and} \quad r_3 = L \mod 3^\beta + \zeta.$$

Let us choose

$$\zeta := \log_3(N) \left(1 - \frac{\log 3}{\log 4}\right) + s_3(L) - s_2(r_2) + \frac{\nu}{2} - \beta, \quad \text{and} \quad \zeta = \lceil \zeta \rceil. \quad (47)$$

We have $r_2 < 2^\nu$, and $L < 2^\nu 3^\beta$; moreover, it follows from the definitions that $\nu = o(\log N)$ and $\beta = o(\log N)$. Therefore $\zeta \sim C \log_3 N$, where the constant equals

$$C = 1 - \frac{\log 3}{\log 4} = 0.2075 \pm 10^{-4}. \quad (48)$$

In particular, we have $3^\zeta \geq 2^\nu$ for all large $N$. Since $L < 2^\nu 3^\beta$, the integer $b_3$ is in fact equal to zero, and $r_3 = L$ (note that we have already replaced $r_3$ by $L$ in the definition of $\zeta$ in order to avoid a circular definition). The deviation caused by $s_2(r_2) - s_3(r_3)$ in (46) will therefore
be bounded by \(\ll J\lambda^{3/4}\). Increasing \(\zeta\) further does not change \(r_2\) and \(r_3\), and we can exploit this freedom in order to shift the expected value of \(J\zeta\) near zero. This is the purpose of the definition \((17)\).

We study the values
\[
f_2(k) = s_2(b_2 + 3^{\beta+\zeta}k) \quad \text{and} \quad f_3(k) = s_3(2^\nu k)
\]
separately, as \(k\) varies in \(I\). Note that \(b_2 < 3^\beta\).

Sure enough the study of \((19)\) will be infeasible in general using current techniques. This is the case because we encounter problems arising from powers of 2 and 3, as considered in the introduction. In our application however, the interval \(I\) is of the form
\[
I = [M, 2M + O(1)]
\]
for some \(M\) considerably larger than \(2^\nu\) and \(3^{\beta+\zeta}\), which enables us to prove a nontrivial statement on the distributions of \(f_2(k)\) and \(f_3(k)\).

In the following, we use the abbreviation \(\alpha = \beta + \zeta\). Let us partition the binary expansion of \(b_2 + 3^\nu k\) into two parts, using the integer \(\kappa_2 = \min\{m : 2^m \geq 3^\alpha\}\). For all integers \(k \geq 0\), we have
\[
s_2(b_2 + 3^\nu k) = s_2\left(\left\lfloor \frac{3^\nu}{2^{\kappa_2}} + \sigma \right\rfloor \right) + s_2((b_2 + 3^\nu k) \mod 2^{\kappa_2}),
\]
where \(\sigma = b_2 2^{-\kappa_2} < 1\).

The values of \(\lfloor k3^{\nu}/2^{\kappa_2} + \sigma \rfloor\) start at \(\tilde{M} + O(1)\), where \(\tilde{M} = \rho M\) and \(\rho = 3^\nu/2^{\kappa_2} \in (1/2, 1)\), increase step by step as \(k\) runs through \(I\), and remain on the same integer for at most two consecutive values of \(k\). Consequently, the distribution of the first summand for \(k \in I\) originates from the distribution of \(s_2(k')\) for \(k' \in I'\), where
\[
I' = [\tilde{M} - 1, 2\tilde{M} + 1],
\]
and each number of occurrences is multiplied by a value \(\in \{0, 1, 2\}\). Therefore, using the binomial distribution, the first summand in \((51)\) can be found within a short interval containing \(\frac{1}{2} \log_2 M\) most of the time. More precisely, we apply Hoeffding’s inequality. Considering the binary sum-of-digits function on \([0, 2^K]\) as a sum of independent random variables with mean \(1/2\), we obtain for all integers \(T \geq 0\) and real \(t \geq 0\)
\[
\frac{1}{2T} \{0 \leq n < 2^T : |s_2(n) - T/2| \geq t\} \leq 2 \exp(-2t^2/T).
\]
We apply this for \(t = Jm/5\) and \(T\) minimal such that \(2T \geq 2\tilde{M} + 1\). Note that
\[
T \sim \log_2 \left(\frac{N}{2^T 3^\beta + \zeta}\right) \geq \lambda.
\]
Note that we used the definition of \(\zeta\) for the latter asymptotics. From \((52)\) we obtain
\[
\{k \in I : |s_2((k3^\nu/2^{\kappa_2} + \sigma)) - T/2| \geq t\} \leq 2 \{k' \in I' : |s_2(k') - T/2| \geq t\} \\
\leq 2 \{0 \leq k' < 2^T : |s_2(k') - T/2| \geq t\} \\
\ll \exp(-2\lambda(\log \lambda)^{1+2\epsilon}/(25T)) \\
\ll \exp(-C(\log \lambda)^{1+2\epsilon}) \\
\ll \lambda^{-D}
\]
for all $D > 0$ and some $C$, as $N \to \infty$. Meanwhile, the second summand in (51) also follows a binomial distribution, with mean $\kappa_2/2$ and a corresponding concentration property. After forming an intersection, it is still the case that for all but $O$ and also considering the rounding error coming from the floor function $D > 2 \ PROOFS$

following important conclusion. 

The contribution of $f_3(k) = s_3(2^r k)$ can be handled in an analogous fashion. We obtain that the value of $f_3(k) = s_3(2^r k)$ is $2Jm/5$-close to the value

$$E_4 = \log_3 \left( \frac{N}{2^\nu \beta + \zeta} \right) + \log_3(2^\nu) = \log_3(N) - \beta - \zeta$$

for all but $O(|I| \lambda^{-D})$ integers $k \in I$. Again, $D > 0$ is arbitrary. Including the term $s_2(r_2) - s_3(r_3)$ from from (46) leads to the definition of $\zeta$ in (47). Joining the preceding statements and (46), noting that the allowed deviation $Jm$ is not surpassed when adding two times the error $2Jm/5$ and also considering the rounding error coming from the floor function $\zeta = \lfloor \zeta \rfloor$, we obtain the following important conclusion.

For all but $O(|I| \lambda^{-D})$ integers $n \in [N, 2N) \cap A''$, the quantity $f(n)$ is $Jm$-close to 0. (54)

2.3 The critical expression modulo $m$

The final piece in the puzzle, before we proceed to the assembly of these pieces, is the study of the function $f(n) \mod m = (s_2(n) - s_3(n)) \mod m$ along arithmetic progressions.

We are going to adopt the Mauduit–Rivat method for digital problems [13, 15, 16, 18, 29, 30, 31, 32, 34, 35, 36, 37], also applied in the papers [17, 41, 42, 43, 44, 53]. This will be used in order to obtain a statement concerning the number of $n$ such that $f(n) \in m\mathbb{Z}$, where $n$ varies in an arithmetic progression. Combining this with (51), we will obtain many values of $n$ in $A''$ such that $f(n)$ is close to zero and an element of $m\mathbb{Z}$. Taken together, these statements imply $f(n) \in Q$ for many values of $n$ in $A''$.

Let us define

$$S_0 = S_0(\vartheta) = \sum_{k \in I} e(\vartheta s_2(b_2 + 3^\beta + \zeta k) - \vartheta s_3(2^\nu k)).$$

In order to study the distribution in residue classes modulo $m$, we are going to choose $\vartheta = b/m$, where $b \in \{0, \ldots, m - 1\}$. By orthogonality relations,

$$P = \frac{|I|}{m} + \frac{1}{m} \sum_{1 \leq b < m} S_0 \left( \frac{b}{m} \right),$$

where

$$P = \# \{ k \in I : s_2(b_2 + 3^\beta + \zeta k) - s_3(2^\nu k) \equiv 0 \mod m \}.$$  

It is therefore sufficient to find an estimate for $S_0(\vartheta)$. We apply van der Corput’s inequality (for example, [35, Lemme 4]):

$$|S_0|^2 \leq \frac{|I| + R - 1}{R} \sum_{-R < r < R} \left( 1 - \frac{|r|}{R} \right) \sum_{\substack{k \in I \\ k + r \in I}} e \left( \vartheta (s_2(b_2 + 3^\beta + \zeta (k + r)) - s_2(b_2 + 3^\beta + \zeta k)) - \vartheta (s_3(2^\nu (k + r)) - s_3(2^\nu k)) \right).$$
Next, we apply a suitable carry propagation lemma in order to “cut off digits”. See [53, Lemma 4.5] for the base-2 version used here; an analogous statement holds for all bases, and we also need the completely analogous base-3 variant (compare the original appearance in [35, Lemme 5]). We treat the summand \( r = 0 \) separately, discard the condition \( n + r \in I \), and join the cases \( r \) and \( -r \). For \( \mu_2 \) and \( \mu_3 \) to be chosen later, we obtain

\[
|S_0|^2 \leq |I|^2 C \left( \frac{R}{|I|} + \frac{1 + 3^{\gamma + \zeta} R}{2^{\mu_2}} + \frac{2^\nu R}{3^{\mu_3}} \right) + \frac{2 |I|}{R} \sum_{1 \leq r < R} |S_1|, \tag{58}
\]

where

\[
S_1 = \sum_{k \in I} \left( \vartheta s_2^{(\mu_2)} (3^{\gamma + \zeta} k + b_2 + 3^{\gamma + \zeta} r) - \vartheta s_2^{(\mu_2)} (3^{\gamma + \zeta} k + b_2) \\
- \vartheta s_3^{(\mu_3)} (2^\nu k + 2^\nu r) + \vartheta s_3^{(\mu_3)} (2^\nu k) \right). \tag{59}
\]

As we had in base 2 before, we use truncated sum-of-digits functions:

\[
\begin{align*}
  s_2^{(\mu_2)} (n) &= s_2 (n \mod 2^{\mu_2}), \\
  s_3^{(\mu_3)} (n) &= s_3 (n \mod 3^{\mu_3}).
\end{align*}
\]

Note that the lowest \( \mu_2 \) binary digits of \( 3^{\gamma + \zeta} k + b_2 \) and the lowest \( \mu_3 \) ternary digits of \( 2^\nu k \) are visited uniformly and independently — this is just the Chinese remainder theorem.

We obtain

\[
S_1 = \frac{|I|}{2^{\mu_2} 3^{\mu_3}} \sum_{0 \leq n_2 < 2^{\mu_2}} \sum_{0 \leq n_3 < 3^{\mu_3}} e \left( \vartheta s_2^{(\mu_2)} (n_2 + 3^{\gamma + \zeta} r) - \vartheta s_2^{(\mu_2)} (n_2) \right) \\
\times \sum_{0 \leq n_3 < 3^{\mu_3}} e \left( \vartheta s_3^{(\mu_3)} (n_3 + 2^\nu r) - \vartheta s_3^{(\mu_3)} (n_3) \right) + O (2^{\mu_2} 3^{\mu_3}). \tag{60}
\]

For this estimate to be relevant, it is important that \( C \) defined in (15) is smaller than \( 1/2 \): the interval \( I \) has length \( \asymp N/(2^\nu 3^{\gamma + \zeta}) \), and we need to run through \( 2^\nu 3^{\gamma + \zeta} \) many integers \( n \in I \) in order to apply the Chinese remainder theorem. Comparing the bases 2 and 5, the corresponding constant will already be greater than \( 1/2 \), so we will need new ideas! Meanwhile, adjacent bases \( b \) and \( b + 1 \) can certainly be handled; the sequence of constants for this case converges to \( 1/2 \) from below as \( b \to \infty \).

It is sufficient to estimate the first factor. We are concerned with the correlation (the characteristic function) we had in (17):

\[
\omega_t (\vartheta, L) = \frac{1}{2L} \sum_{0 \leq n < 2L} e \left( \vartheta s_2^{(L)} (n + t) - \vartheta s_2^{(L)} (n) \right).
\]

Reusing the bound given in [34], and Lemma 2.2 we obtain the following result.

**Lemma 2.4.** Assume that integers \( B \geq 0 \) and \( L, t \geq 1 \) are given such that \( t \) contains at least \( 2B + 1 \) blocks of 1s, and \( t < 2^L \). Then

\[
|\omega_t (\vartheta, L)| \leq \left( 1 - \frac{1}{2} \| \vartheta \|^2 \right)^B.
\]
Our focus therefore lies on the number \( B \) of blocks of 1s in the binary expansion of \( 3^{\beta + \epsilon} r \).

The only thing we need to know about powers of three in this context is the fact that they are odd integers — we exploit in an essential way the summation over \( r \) instead. The parameter \( R \) will be a certain power of \( N \); in this way, the expected size of \( B \) is \( \gg \lambda \). For the estimation of the number of exceptions we use the binomial distribution again.

Note that counting the number of blocks of 1s in binary amounts to counting the number of occurrences of 01 (where the 0 corresponds to the more significant digit), up to an error \( \mathcal{O}(1) \). For simplicity, we only count such occurrences where the digit 1 in the block 01 occurs at an even index. For example, in the binary expansion 10110110 the corresponding number is 1, whereas there exist three blocks of 1s. This simplification will, on average, give 1/2 of the actual expected value. We are therefore concerned with the number of 1s occurring in the base-4 expansion. Note that the number of blocks of 0 \( n < 4^k \) having \( \ell \) ones in base 4 is given by

\[
4^K \binom{K}{\ell} \left( \frac{1}{4}\right)^\ell \left( \frac{3}{4}\right)^{K-\ell}.
\]

Suppose that we have \( R = 4^K \). Note that

\[
r \mapsto r3^{\beta + \epsilon} \mod 4^K
\]

is a bijection of the set \( \{0, \ldots, 4^K - 1\} \). We abbreviate \( \alpha = 1 - \|\vartheta\|^2/2 \), and obtain by Lemma 2.4

\[
S_2 := \sum_{1 \leq r < R} \left| \frac{1}{2^{\mu_2}} \sum_{0 \leq n_2 < 2^{\mu_2}} e(\vartheta s_2^{(r)} (n_2 + 3^{\beta+\epsilon} r) - \vartheta s_2^{(r)} (n_2)) \right| \\
\leq \sum_{0 \leq \ell \leq K} \sum_{1 \leq r < 4^K \atop r \text{ in base } 4 \text{ has } \ell \text{ ones}} \alpha^{\frac{\ell}{2}} = 4^K \alpha^{-1/2} \sum_{0 \leq \ell \leq K} \binom{K}{\ell} \left( \frac{1}{4}\right)^\ell \left( \frac{3}{4}\right)^{K-\ell} \alpha^{\ell/2} \\
= 4^K \alpha^{-1/2} \left( \sqrt{\alpha/4} + 3/4 \right)^K.
\]

We note that \( \|\vartheta\| \geq 1/m \asymp \lambda^{-1/2}(\log \lambda)^{1/2+\epsilon} \) and therefore

\[
\alpha^{1/2} = \left( 1 - \|\vartheta\|^2/2 \right)^{1/2} \asymp 1 - \lambda^{-1}(\log \lambda)^{1+2\epsilon}.
\]

(61)

By the estimate

\[
(1 + x)^K = \exp(K \log(1 + x)) \leq \exp(Kx)
\]

it follows that

\[
S_2 \ll 4^K \exp \left( \frac{1}{4} K \lambda^{-1}(\log \lambda)^{1+2\epsilon} \right).
\]

(62)

Translating this back to \( S_0 \), we obtain

\[
|S_0|^2 \ll |I|^2 \left( \frac{R}{|I|} + \frac{1}{R} + \frac{3^{\beta+\epsilon} R}{2^{\mu_2}} + \frac{2^{\nu} R}{3^{\mu_3}} + \exp \left( \frac{1}{4} K \lambda^{-1}(\log \lambda)^{1+2\epsilon} \right) \right).
\]

(63)

We see that the last term yields a contribution to \( S_0 \) that is is smaller than the fair share \( |I|m^{-1} \asymp |I|\lambda^{-1/2}(\log \lambda)^{1/2+\epsilon} \) as soon as \( K \asymp \lambda \), due to the presence of the power \( (\log \lambda)^{1+2\epsilon} \) in the exponent. For this, we need to choose \( R = 4^K \) as large as some positive (fixed) power of \( N \). At the same time we have to take care of the other error terms in (63). It is obvious that it is possible to choose as \( R \) a (small) power \( N^\nu \) of \( N \), and \( 2^{\mu_2} \) resp. \( 3^{\mu_3} \) larger than \( R^{3^{\beta+\epsilon}} \) resp. \( R^{2^\nu} \) (by some small power of \( N \)), in such a way that \( 2^{\mu_2} 3^{\mu_3} \) is still smaller than \( |I| \) (by
another power of \(N\). This is possible by the observation that \(\zeta < 1/2\), and we commented on this after \(\text{(60)}\). With \(R = N^\epsilon\) for some \(\epsilon > 0\) we therefore obtain from \(\text{(59)}\)
\[
P = \frac{|I|}{m} (1 + o(1))
\]
as \(N \to \infty\), where \(P\) was defined in \(\text{(64)}\). That is, the residue class \(m\mathbb{Z}\) receives the expected ratio \(\lambda^{-1/2}(\log \lambda)^{1/2+\epsilon}\) of the values of \(f(n) = s_2(n) - s_3(n)\) along the finite arithmetic progression \(A'' \cap [N, 2N]\), where \(A'' = L + 2^{\nu}3^{\beta+\epsilon}\mathbb{Z}\).

2.4 Finishing the proof of Theorem 1.1

We obtained the expected number of \(n \in A'' \cap [N, 2N]\) such that \(f(n) \in m\mathbb{Z}\) in the previous section. At the same time, \(\text{(52)}\) states that \(f(n)\) lies in the interval \([-Jm, Jm]\) for \(|I| (1-\mathcal{O}(\lambda^{-D}))\) many integers \(n \in A'' \cap [N, 2N]\). Consequently, any choice \(D > 1/2\) will yield an integer \(n \in A'' \cap [N, 2N]\) such that \(s_2(n) - s_3(n) = jm\) for some \(j \in \{-J, \ldots, J\}\). By \(\text{(61)}\) the integer \(n' = n + d_j\) satisfies \(s_2(n') - s_3(n') \in \{0, 1\}\). As we noted before, the shifts \(d_j\) are smaller than the halved common difference \(2^{\nu-1}3^{\beta+\epsilon}\) of \(A''\). Varying \(N\), we get many almost-collisions (as in Proposition 2.1) in each large enough interval \([N, 2N]\) and thus the qualitative statement in Theorem 1.1.

It is easy to see that the interval \(I\) from section 2.2 (note the definition of \(\zeta\) and \(\text{(17)}\)) is in fact of size \(\gg N^{\log 3/\log 2-\epsilon}\) for all \(\epsilon > 0\). Most \(k \in I\) yield a value \(\tilde{f}(k) = f(L + 2^\nu 3^{\beta+\epsilon}k) \in [-Jm, Jm]\) (see \(\text{(61)}\)), and the expected proportion \(\sim m^{-1} \gg (\log N)^{-1/2}\) of them satisfy \(\tilde{f}(k) \in m\mathbb{Z}\), see \(\text{(64)}\). These \(k\) yield pairwise different values \(k + d_j(k)\) as before. Here the integer \(j = j(k)\) is chosen suitably from \(\{-J, \ldots, J\}\) in order to yield a collision. Also, the argument connecting Proposition 2.1 and Theorem 1.1 is easily adapted. This proves \(\text{(7)}\).

In the last step of our proof — choosing \(j \in \{-J, \ldots, J\}\) suitably — we can clearly see the “element of non-constructiveness” in our proof. Currently we do not have any control over the choice of \(j\).

3 Open problems

1. Find a construction method for collisions.

2. Prove that there are infinitely many prime numbers \(p\) such that
\[
s_2(p) = s_3(p).
\]

3. Prove or disprove the asymptotic formula
\[
\#\{n < N : s_2(n) = s_3(n)\} \sim cN^\eta
\]
for some real constants \(c\) and \(\eta\).

4. Prove an asymptotic formula (in \(k\)) for the number of solutions of the equation
\[
2^{\mu_1} + \cdots + 2^{\mu_k} = 3^{\nu_1} + \cdots + 3^{\nu_k}
\]
in not necessarily distinct natural numbers \(\mu_j, \nu_j\). In particular, the equation should have a positive, finite number of solutions for all but finitely many \(k\). Also, prove such a formula for \(s_2(n) = s_3(n) = k\) (where finiteness of the number of solutions in this case follows from Senge and Straus \[50\]).
5. Generalize the theorems in this paper and Problems 1–4 to arbitrary families $(q_1, \ldots, q_K)$ of pairwise coprime bases $\geq 2$. It would also be interesting to prove the existence of infinitely many Catalan numbers exactly divisible by some power of $a$, were $a \geq 2$ is an arbitrary integer. This property can be defined by

$$a^k \parallel n \Leftrightarrow (a^k \mid n \text{ and } \gcd(na^{-k}, a) = 1).$$  \ (68)

Acknowledgements.

The author is grateful to his teachers Michael Drmota and Joël Rivat, who sparked his interest in the kind of research problems presented in this paper. Moreover, he thanks Jean-Marc Deshouillers for pointing out the article [11] to him, which was the starting point for the work on the present paper.

References

[1] B. Adamczewski and C. Faverjon, Mahler’s method in several variables and finite automata, 2020.

[2] A. Baker, Linear forms in the logarithms of algebraic numbers, Mathematika, 13 (1966), pp. 204–216.

[3] A. Baker, Linear forms in the logarithms of algebraic numbers. II, Mathematika, 14 (1967), pp. 102–107.

[4] —, Linear forms in the logarithms of algebraic numbers. III, Mathematika, 14 (1967), pp. 220–228.

[5] V. Berthé, Autour du système de numération d’Ostrowski, Bull. Belg. Math. Soc. Simon Stevin, 8 (2001), pp. 209–239. Journées Montoises d’Informatique Théorique (Marne-la-Vallée, 2000).

[6] J. Bésineau, Indépendance statistique d’ensembles liés à la fonction “somme des chiffres”, Acta Arith., 20 (1972), pp. 401–416.

[7] E. F. Bravo and J. J. Bravo, Powers of two as sums of three Fibonacci numbers, Lith. Math. J., 55 (2015), pp. 301–311.

[8] J. J. Bravo and F. Luca, On the Diophantine equation $F_n + F_m = 2^n$, Quaest. Math., 39 (2016), pp. 391–400.

[9] Y. Bugeaud, M. Cipu, and M. Mignotte, On the representation of Fibonacci and Lucas numbers in an integer base, Ann. Math. Qué., 37 (2013), pp. 31–43.

[10] R. de la Bretèche, T. Stoll, and G. Tenenbaum, Somme des chiffres et changement de base, Ann. Inst. Fourier, 69 (2019), pp. 2507–2518.

[11] J.-M. Deshouillers, L. Habsieger, S. Laishram, and B. Landreau, Sums of the digits in bases 2 and 3, in Number theory — Diophantine problems, uniform distribution and applications, Springer, 2017, pp. 211–217.

[12] V. S. Dimitrov and E. W. Howe, Powers of 3 with few nonzero bits and a conjecture of Erdős, 2021.
REFERENCES

[13] M. Drmota, The joint distribution of q-additive functions, Acta Arith., 100 (2001), pp. 17–39.

[14] M. Drmota, C. Mauduit, and J. Rivat, Primes with an average sum of digits, Compos. Math., 145 (2009), pp. 271–292.

[15] ———, Normality along squares, J. Eur. Math. Soc. (JEMS), 21 (2019), pp. 507–548.

[16] M. Drmota, C. Mauduit, and J. Rivat, Prime numbers in two bases, Duke Math. J., 169 (2020), pp. 1809–1876.

[17] M. Drmota and J. F. Morgenbesser, Generalized Thue-Morse sequences of squares, Isr. J. Math., 190 (2012), pp. 157–193.

[18] M. Drmota, J. Rivat, and T. Stoll, The sum of digits of primes in Z[i], Monatsh. Math., 155 (2008), pp. 317–347.

[19] P. Erdős, Some unconventional problems in number theory, Math. Mag., 52 (1979), pp. 67–70.

[20] P. Erdős and R. L. Graham, Old and new problems and results in combinatorial number theory, vol. 28, L’Enseignement Mathématique, Université de Genève, Genève, 1980.

[21] H. Furstenberg, Intersections of Cantor sets and transversality of semi-groups. Probl. Analysis, Sympos. in Honor of Salomon Bochner, Princeton Univ. 1969, 41-59 (1970), 1970.

[22] A. Gelfond, Sur le septième problème de D. Hilbert, C. R. (Dokl.) Acad. Sci. URSS, n. Ser., 1934 (1934), pp. 1–6.

[23] ———, Sur le septième Problème de Hilbert, Bull. Acad. Sci. URSS, 1934 (1934), pp. 623–634.

[24] A. O. Gel’fond, Sur les nombres qui ont des propriétés additives et multiplicatives données, Acta Arith., 13 (1967/68), pp. 259–265.

[25] A. Granville and O. Ramaré, Explicit bounds on exponential sums and the scarcity of squarefree binomial coefficients, Mathematika, 43 (1996), pp. 73–107.

[26] D.-H. Kim, On the joint distribution of q-additive functions in residue classes, J. Number Theory, 74 (1999), pp. 307–336.

[27] J. C. Lagarias, Ternary expansions of powers of 2, J. Lond. Math. Soc., II. Ser., 79 (2009), pp. 562–588.

[28] F. Luca, On the Diophantine equation $p^{x_1} - p^{x_2} = q^{y_1} - q^{y_2}$, Indag. Math., New Ser., 14 (2003), pp. 207–222.

[29] B. Martin, C. Mauduit, and J. Rivat, Théorème des nombres premiers pour les fonctions digitales, Acta Arith., 165 (2014), pp. 11–45.

[30] ———, Fonctions digitales le long des nombres premiers, Acta Arith., 170 (2015), pp. 175–197.

[31] ———, Nombres premiers avec contraintes digitales multiples, Bull. Soc. Math. Fr., 147 (2019), pp. 259–287.
REFERENCES

[32] ———, Propriétés locales des chiffres des nombres premiers, J. Inst. Math. Jussieu, 18 (2019), pp. 189–224.

[33] C. Mauduit, C. Pomerance, and A. Sárközy, On the distribution in residue classes of integers with a fixed sum of digits, Ramanujan J., 9 (2005), pp. 45–62.

[34] C. Mauduit and J. Rivat, La somme des chiffres des carrés, Acta Math., 203 (2009), pp. 107–148.

[35] ———, Sur un problème de Gelfond: la somme des chiffres des nombres premiers, Ann. of Math. (2), 171 (2010), pp. 1591–1646.

[36] ———, Prime numbers along Rudin-Shapiro sequences, J. Eur. Math. Soc. (JEMS), 17 (2015), pp. 2595–2642.

[37] C. Mauduit and J. Rivat, Rudin-Shapiro sequences along squares, Trans. Am. Math. Soc., 370 (2018), pp. 7899–7921.

[38] C. Mauduit and A. Sárközy, On the arithmetic structure of the integers whose sum of digits is fixed, Acta Arith., 81 (1997), pp. 145–173.

[39] J. Maynard, Primes with restricted digits, Invent. Math., 217 (2019), pp. 127–218.

[40] M. Mignotte, Sur les entiers qui s’écrivent simplement en différentes bases. (On integers simply represented in different bases), Eur. J. Comb., 9 (1988), pp. 307–316.

[41] J. F. Morgenbesser and T. Stoll, On a problem of Chen and Liu concerning the prime power factorization of n!, Proc. Am. Math. Soc., 141 (2013), pp. 2289–2297.

[42] C. Müllner, Automatic sequences fulfill the Sarnak conjecture, Duke Math. J., 166 (2017), pp. 3219–3290.

[43] C. Müllner, The Rudin-Shapiro sequence and similar sequences are normal along squares, Can. J. Math., 70 (2018), pp. 1096–1129.

[44] N. Ouled Azaiez, M. Mkaouar, and J. M. Thuswaldner, Sur les chiffres des nombres premiers translatés, Funct. Approximatio, Comment. Math., 51 (2014), pp. 237–267.

[45] A. Pethő and R. F. Tichy, S-unit equations, linear recurrences and digit expansions, Publ. Math., 42 (1993), pp. 145–154.

[46] G. Rhin, Approximants de Padé et mesures effectives d’irrationalité. Théorie des Nombres, Sémin. Paris 1985/86, Prog. Math. 71, 155-164 (1987), 1987.

[47] A. Sárközy, On divisors of binomial coefficients. I, J. Number Theory, 20 (1985), pp. 70–80.

[48] H. P. Schlickewei, Linear equations in integers with bounded sum of digits, J. Number Theory, 35 (1990), pp. 335–344.

[49] ———, S-unit equations over number fields, Invent. Math., 102 (1990), pp. 95–107.

[50] H. G. Senge and E. G. Straus, PV-numbers and sets of multiplicity, Period. Math. Hung., 3 (1973), pp. 93–100.

[51] P. Shmerkin, On Furstenberg’s intersection conjecture, self-similar measures, and the $L^q$ norms of convolutions, Ann. Math. (2), 189 (2019), pp. 319–391.
[52] N. J. A. SLOANE, *The On-Line Encyclopedia of Integer Sequences*, 2021. Published electronically at [https://oeis.org](https://oeis.org).

[53] L. SPIEGELHOFER, *The level of distribution of the Thue–Morse sequence*, Compos. Math., 156 (2020), pp. 2560–2587.

[54] ——, *A lower bound for Cusick’s conjecture on the digits of n + t*, Math. Proc. Cambridge Philos. Soc., (2020). Published online by Cambridge University Press: 24 February 2021, pp. 1-23.

[55] L. SPIEGELHOFER AND M. WALLNER, *The digits of n + t*, 2020. Preprint. arXiv:2005.07167v2.

[56] C. L. STEWART, *On the representation of an integer in two different bases*, J. Reine Angew. Math., 319 (1980), pp. 63–72.

[57] M. WU, *A proof of Furstenberg’s conjecture on the intersections of χp- and χq-invariant sets*, Ann. Math. (2), 189 (2019), pp. 707–751.

[58] Q. WU AND L. WANG, *On the irrationality measure of log 3*, J. Number Theory, 142 (2014), pp. 264–273.

[59] V. ZIEGLER, *Effective results for linear equations in members of two recurrence sequences*, Acta Arith., 190 (2019), pp. 139–169.