



# The sum-of-digits function on arithmetic progressions

Lukas Spiegelhofer, Thomas Stoll

## ► To cite this version:

Lukas Spiegelhofer, Thomas Stoll. The sum-of-digits function on arithmetic progressions. Moscow Journal of Combinatorics and Number Theory, 2020, Vol. 9 (No. 1), pp.43-49. 10.2140/moscow.2020.9.43 . hal-03283478

**HAL Id: hal-03283478**

**<https://hal.science/hal-03283478>**

Submitted on 10 Jul 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THE SUM-OF-DIGITS FUNCTION ON ARITHMETIC PROGRESSIONS

LUKAS SPIEGELHOFFER AND THOMAS STOLL

ABSTRACT. Let  $s_2$  be the sum-of-digits function in base 2, which returns the number of non-zero binary digits of a nonnegative integer  $n$ . We study  $s_2$  along arithmetic subsequences and show that — up to a shift — the set of  $m$ -tuples of integers that appear as an arithmetic subsequence of  $s_2$  has full complexity.

## 1. RESULTS

The binary sum-of-digits function  $s_2$  is an elementary object studied in number theory. It is defined by the equation

$$s_2(\varepsilon_\nu 2^\nu + \cdots + \varepsilon_0 2^0) = \varepsilon_\nu + \cdots + \varepsilon_0,$$

where  $\varepsilon_i \in \{0, 1\}$  for  $0 \leq i \leq \nu$ . Despite the simplicity of definition, the behaviour of  $s_2$  on arithmetic progressions is not fully understood. Cusick's conjecture on the sum-of-digits function [3, 9] concerns this area of research: for an integer  $t \geq 0$ , we define the limit

$$c_t = \lim_{N \rightarrow \infty} \frac{1}{N} |\{n : 0 \leq n < N, s_2(n+t) \geq s_2(n)\}|.$$

(The limit exists, see for example Bésineau [2]. In fact, the set in this definition is periodic with period  $2^k$  for some  $k$ .) Cusick's conjecture states that

$$(1.1) \quad c_t > 1/2$$

for all  $t \geq 0$ . Drmota, Kauers, and the first author [3] proved that  $c_t > 1/2$  for *almost all*  $t$  in the sense of asymptotic density; we also wish to note the works by Emme and Prihod'ko [6] and Emme and Hubert [4, 5], and the recent partial result by the first author [9].

In the current note, motivated by Cusick's conjecture, we are concerned with the  $(m+1)$ -tuple  $(s_2(n), s_2(n+t), \dots, s_2(n+mt))$ , where  $t \geq 0$  and  $m \geq 1$  are integers. We aim to understand the set of tuples that can occur, as  $n$  and  $t$  run. In fact, our theorem states that, up to a shift, all tuples occur.

**Theorem 1.1.** *Assume that  $k_1, \dots, k_m \in \mathbb{Z}$ . There exist  $n$  and  $t$  such that for  $1 \leq \ell \leq m$ ,*

$$k_\ell = s_2(n + \ell t) - s_2(n).$$

---

2010 *Mathematics Subject Classification.* 11A63, 11B25.

The authors acknowledge support by the project MuDeRa, which is a joint project between the FWF (Austrian Science Fund) and the ANR (Agence Nationale de la Recherche). Moreover, the first author was supported by the FWF project F5502-N26, which is a part of the Special Research Program “Quasi Monte Carlo methods: Theory and Applications”; the second author was supported by the project ANR-18-CE40-0018.

This is a generalization of the statement that the Thue–Morse sequence  $\mathbf{t}$  has full *arithmetic complexity*, meaning that every finite word  $\omega \in \{0, 1\}^L$  occurs as an arithmetic subsequence of  $\mathbf{t}$ . This was first proved in [1] and also follows from Müllner and the first named author [8], and Konieczny [7].

Theorem 1.1 is not hard to prove for  $m = 1$ . We present three arguments leading to this fact.

- (1) Assume first that  $k \geq 0$ . Set  $n = 2^{k+1}$  and  $t = 2^k - 1$ . Then  $s_2(n+t) = k+1$  and  $s_2(n) = 1$ , yielding  $k = s_2(n+t) - s_2(n)$ . If  $k < 0$ , we set  $n = 2^{-k+1} - 1$  and  $t = 1$ . Then  $s_2(n) = -k+1$  and  $s_2(n+t) = 1$ , which yields  $s_2(n+t) - s_2(n) = k$ . Alternatively, we may also write, as in the case  $m = 2$  presented below,  $t = 2^c - 1$  and  $n = 2^{c-1}(2^a - 1)$ , for positive integers  $a$  and  $c$ . We obtain  $s_2(n+t) = c$  and  $s_2(n) = a$ , and clearly the difference  $c - a$  runs through all integers.
- (2) We have  $s_2(n+1) - s_2(n) = 1 - \nu_2(n+1) \leq 1$ , where  $\nu_2(m) = \max\{k \geq 0 : 2^k \mid m\}$  for  $m \geq 1$  is the 2-adic valuation of  $m$ . This formula follows by considering the number of 1s with which the binary expansion of  $n$  ends. Since  $s_2(2^\ell) = 1$  and  $s_2(2^{\ell+1} - 1) = \ell + 1$ , we obtain the fact that  $s_2(n)$  attains all values in  $\{1, \dots, \ell + 1\}$  as  $n$  varies in  $\{2^\ell, \dots, 2^{\ell+1} - 1\}$ . Let  $k \in \mathbb{Z}$  be given and set  $\ell = 2|k|$ . Choose  $n \in \{2^\ell, \dots, 2^{\ell+1} - 1\}$  such that  $s_2(n) = |k| + 1$  and  $n' \in \{2^{\ell+1}, \dots, 2^{\ell+2} - 1\}$  such that  $s_2(n') = |k| + 1 + k$ . Then  $s_2(n') - s_2(n) = k$ , which implies the statement.
- (3) Consider the densities

$$\delta(k, t) = \lim_{N \rightarrow \infty} \frac{1}{N} |\{n : 0 \leq n < N, s_2(n+t) - s_2(n) = k\}|$$

(as it was the case for  $c_t$ , this asymptotic density exists [2]). These quantities satisfy the following recurrence [3]:

$$\begin{aligned} \delta(k, 1) &= \begin{cases} 2^{k-2}, & k \leq 1; \\ 0 & \text{otherwise;} \end{cases} \\ \delta(k, 2t) &= \delta(k, t); \\ \delta(k, 2t+1) &= \frac{1}{2}\delta(k-1, t) + \frac{1}{2}\delta(k+1, t+1). \end{aligned}$$

From this, it is very easy to show that  $\delta(k, t) > 0$  for all  $k \leq s_2(t)$ . For  $k$  given, choose  $t$  in such a way that  $s_2(t) \geq k$ ; the positivity of the density  $\delta(k, t)$  implies that there exists an  $n$  such that  $s_2(n+t) - s_2(n) = k$ .

For  $m = 2$ , it is also possible to obtain the statement by elementary considerations: consider integers  $a, c \geq 1, b, d \geq 0$  and choose the integers  $n$  and  $t$  in such a way that the binary expansions look as follows:

$$\begin{array}{rcl} n : & \overbrace{1 \cdots 1}^a 1 0 \cdots 0 & \overbrace{1 \cdots 1}^b 0 \cdots 0 \\ t : & & \underbrace{1 1 \cdots 1}_c 0 \cdots 0 \underbrace{1 \cdots 1}_d. \end{array}$$

The sums of digits of  $n$ ,  $n+t$  and  $n+2t$  respectively are  $a+b$ ,  $b+c+d$  and  $c+d$  respectively. By varying the variables, we can obtain the statement for all integers  $k_1$

and  $k_2$  such that  $k_2 \leq k_1$ . For the case  $k_1 < k_2$ , we use the following configuration of the integers  $n$  and  $t$ , where  $a, d \geq 1$  and  $c \geq 0$ :

$$\begin{array}{rcccl} & & \overbrace{1 \cdots 1}^a 1 0 \cdots 0 & & \\ n : & & & & \\ & & \underbrace{1 \cdots 1}_c 0 \cdots 0 & \underbrace{1 1 \cdots 1}_d & \\ t : & & & & \end{array}$$

The sums of digits of  $n$ ,  $n + t$  and  $n + 2t$  are  $a$ ,  $d$  and  $c + d$  respectively, and we see that we obtain all pairs  $(k_1, k_2) \in \mathbb{Z}^2$  such that  $k_1 \leq k_2$ .

However, the method quickly experiences difficulties, as multiplication by 3 is not a shift of the binary digits anymore. While we believe that the case  $m = 3$  can be made work by some effort, a general principle is not apparent. Therefore we choose a different approach.

We prove Theorem 1.1 by induction on  $m$ , the cases  $m = 1, 2$  having been discussed above. Assume that  $m \geq 3$  and let  $k_1, \dots, k_m \in \mathbb{Z}$  be given. By induction hypothesis, there exist  $t_0$  and  $n_0$  such that  $k_\ell = s_2(n_0 + \ell t_0) - s_2(n_0 + (\ell - 1)t_0)$  for  $1 \leq \ell < m$ . Set  $k'_m = s_2(n_0 + m t_0) - s_2(n_0 + (m - 1)t_0)$ . We are going to show that we may vary  $k'_m$  by steps of  $\pm 1$ , thus yielding the full statement.

By concatenation of binary expansions, it is sufficient to show the following statement.

$$(1.2) \quad \begin{array}{l} \text{There exist } t_1, n_1 \text{ such that } s_2(n_1 + \ell t_1) - s_2(n_1 + (\ell - 1)t_1) = 0 \text{ for } 1 \leq \ell < m \\ \text{and } s_2(n_1 + m t_1) - s_2(n_1 + (m - 1)t_1) = \pm 1. \end{array}$$

This concatenation is straightforward and summarized in the following lemma, which we will also use again in a moment.

**Lemma 1.2.** *Let  $\ell \geq 1$ ,  $m \geq 1$ ,  $n_0, \dots, n_{k-1}$  and  $t_0, \dots, t_{k-1}$  be nonnegative integers. There exist nonnegative integers  $n$  and  $t$  such that*

$$s_2(n + \ell t) - s_2(n + (\ell - 1)t) = \sum_{0 \leq j < k} (s_2(n_j + \ell t_j) - s_2(n_j + (\ell - 1)t_j))$$

for  $1 \leq \ell \leq m$ .

*Proof.* The base case  $k = 1$  is trivial; it is sufficient to prove the statement for  $k = 2$ , the general case following easily from repeated application of this case.

Let  $N$  be so large that  $n_0 + m t_0 < 2^N$ , and set  $n = 2^N n_1 + n_0$  and  $t = 2^N t_1 + t_0$ . Since no carry propagation between the digits below and above  $N$  occurs, we can add up the contribution of the two blocks in order to yield the statement.  $\square$

We reduce the problem further, using this block representation again: choose  $t_j = 1$  for all  $0 \leq j < k$ ; it is sufficient to find a  $k \geq 1$  and nonnegative integers  $n_j$  for  $0 \leq j < k$  such that

$$(1.3) \quad \sum_{0 \leq j < k} (s_2(n_j + \ell) - s_2(n_j + \ell - 1)) = \begin{cases} 0, & \text{if } 1 \leq \ell < m; \\ \pm 1, & \text{if } \ell = m. \end{cases}$$

In order to show (1.3), we use the telescoping sum

$$\sum_{a \leq j < a+2^L} g(j) = s_2(a + 2^L) - s_2(a) = g(\lfloor a/2^L \rfloor),$$

where  $g(j) = s_2(j+1) - s_2(j)$ . This representation yields for  $1 \leq \ell \leq m$ , where  $L$  is chosen such that  $2^L \leq m < 2^{L+1}$ ,

$$\begin{aligned} \sum_{2 \cdot 2^L - m + \ell \leq j < 3 \cdot 2^L - m + \ell} g(j) &= g(2 + \lfloor (-m + \ell)/2^L \rfloor) = \begin{cases} g(1) = 0, & \text{if } 1 \leq \ell < m; \\ g(2) = 1, & \text{if } \ell = m; \end{cases} \\ \sum_{2^L - m + \ell \leq j < 2 \cdot 2^L - m + \ell} g(j) &= g(1 + \lfloor (-m + \ell)/2^L \rfloor) = \begin{cases} g(0) = 1, & \text{if } 1 \leq \ell < m; \\ g(1) = 0, & \text{if } \ell = m; \end{cases} \\ \sum_{3 \cdot 2^{L+1} + \ell \leq j < 4 \cdot 2^{L+1} + \ell} g(j) &= g(3) = -1 \text{ for } 1 \leq \ell \leq m. \end{aligned}$$

The first of these three identities yields the “+”-part of (1.3) by choosing  $k = 2^L$  and  $n_j = 2 \cdot 2^L - m + j$  for  $0 \leq j < k$ .

The “-”-part is obtained from the second and third identities: by considering the disjoint union  $J = [2^L - m, 2 \cdot 2^L - m) \cup [3 \cdot 2^{L+1}, 4 \cdot 2^{L+1})$ , we have

$$\sum_{j \in J} (s_2(j + \ell) - s_2(j + \ell - 1)) = \begin{cases} 0, & \text{if } 1 \leq \ell < m; \\ -1, & \text{if } \ell = m. \end{cases}$$

The statement follows by merging the two intervals and choosing  $n_j$  accordingly. This finishes the proof of our theorem.

## 2. POSSIBLE EXTENSIONS

From our proof, it is possible to effectively construct integers  $n$  and  $t$  such that  $s_2(n + \ell t) - s_2(n) = k_\ell$  for  $1 \leq \ell \leq m$ . In particular, this yields integers  $n$  and  $t$  such that  $\mathbf{t}_{n+\ell t} = \omega_\ell$  for  $1 \leq \ell \leq m$ , where  $(\omega_1, \dots, \omega_m) \in \{0, 1\}^m$  and  $\mathbf{t}$  is the Thue–Morse sequence on  $\{0, 1\}$ . (Note that we also used  $\mathbf{t}(2^\lambda + n) = 1 - \mathbf{t}(n)$  for  $n < 2^\lambda$ .) This gives a constructive result concerning the problem of full arithmetic complexity of the Thue–Morse sequence considered in [1, 7, 8].

As an extension of the presented line of research, we are interested in the proportion of cases in which  $s_2(n + \ell t) - s_2(n) = k_\ell$  occurs (for  $1 \leq \ell \leq m$ ). For this, we define more generally

$$\delta(\mathbf{k}, \boldsymbol{\varepsilon}, t) = \text{dens} \{n \in \mathbb{N} : s_2(n + \ell t + \varepsilon_\ell) - s_2(n) = k_\ell \text{ for } 1 \leq \ell \leq m\},$$

where  $\mathbf{k} = (k_1, \dots, k_m) \in \mathbb{Z}^m$  and  $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_m) \in \mathbb{N}^m$ . This generalizes the array  $\delta$  defined before. As in the one-dimensional case, the densities in this definition actually exist, and they satisfy the following recurrence relation:

$$\delta(\mathbf{k}, \boldsymbol{\varepsilon}, 2t) = \frac{1}{2} \delta(\mathbf{k}', \boldsymbol{\varepsilon}', t) + \frac{1}{2} \delta(\mathbf{k}'', \boldsymbol{\varepsilon}'', t),$$

where  $k'_\ell = k_\ell - (\varepsilon_\ell \bmod 2)$ ,  $k''_\ell = k_\ell + 1 - ((\varepsilon_\ell + 1) \bmod 2)$ ,  $\varepsilon'_\ell = \lfloor \varepsilon_\ell / 2 \rfloor$  and  $\varepsilon''_\ell = \lfloor (\varepsilon_\ell + 1) / 2 \rfloor$ ; moreover,

$$\delta(\mathbf{k}, \boldsymbol{\varepsilon}, 2t + 1) = \frac{1}{2} \delta(\mathbf{k}', \boldsymbol{\varepsilon}', t) + \frac{1}{2} \delta(\mathbf{k}'', \boldsymbol{\varepsilon}'', t),$$

where  $k'_\ell = k_\ell - ((\varepsilon_\ell + \ell) \bmod 2)$ ,  $k''_\ell = k_\ell + 1 - ((\varepsilon_\ell + \ell + 1) \bmod 2)$ ,  $\varepsilon'_\ell = \lfloor (\varepsilon_\ell + \ell) / 2 \rfloor$  and  $\varepsilon''_\ell = \lfloor (\varepsilon_\ell + \ell + 1) / 2 \rfloor$ . This recurrence is the reason for the introduction of  $\boldsymbol{\varepsilon}$  in the

definition of  $\delta(\mathbf{k}, \varepsilon, t)$ . To our knowledge, these identities are new; we give the idea of proof. We have

$$\begin{aligned} \delta(\mathbf{k}, \varepsilon, 2t) &= \frac{1}{2} \text{dens}\{n \in \mathbb{N} : s(2n + 2\ell t + \varepsilon_\ell) - s(2n) = k_\ell \text{ for } 1 \leq \ell \leq m\} \\ &\quad + \frac{1}{2} \text{dens}\{n \in \mathbb{N} : s(2n + 2\ell t + \varepsilon_\ell + 1) - s(2n + 1) = k_\ell \text{ for } 1 \leq \ell \leq m\}, \end{aligned}$$

and using the identities  $s(2n) = s(n)$  and  $s(2n + 1) = s(n) + 1$ , we obtain

$$s(2n + 2\ell t + \varepsilon_\ell) = s(2(n + \ell t + \lfloor \varepsilon_\ell/2 \rfloor) + (\varepsilon_\ell \bmod 2)) = s(n + \ell t + \varepsilon'_\ell) + (\varepsilon_\ell \bmod 2),$$

therefore  $s(2n + 2\ell t + \varepsilon_\ell) - s(2n) = k_\ell$  is equivalent to  $s(n + \ell t + \varepsilon'_\ell) - s(n) = k'_\ell$ . Analogous computations are valid for  $\varepsilon_\ell$  replaced by  $\varepsilon_\ell + 1$ ,  $\varepsilon_\ell + \ell$  and  $\varepsilon_\ell + \ell + 1$  respectively, which yields the claim.

This recurrence can be used to prove statements on the densities  $\delta(\mathbf{k}, \varepsilon, t)$ . The general intuitive idea is that the differences  $s(n + jt) - s(n)$ , for  $1 \leq j \leq m$ , should be *almost independent for most  $t$* ; in the light of this consideration, we consider generalizations of Cusick's conjecture and of Emme and Hubert's result [4]. On the one hand, we may ask for multidimensional generalizations of (1.1), relating the relative sizes of the values  $s_2(n), s_2(n + t), \dots, s_2(n + mt)$  to one another. We propose the following conjecture, extending (1.1).

**Conjecture 1.** Assume that  $m \geq 1$  is an integer. For an integer  $t \geq 0$ , define

$$c_t^{(m)} = \text{dens}\{n \in \mathbb{N} : s(n) \leq s(n + jt) \text{ for } 1 \leq j \leq m\}.$$

Then for all  $t \geq 0$ ,

$$c_t^{(m)} > \frac{1}{2^m}.$$

The statement is wrong for any larger constant in place of  $1/2^m$ . Also, define

$$C_t^{(m)} = \text{dens}\{n \in \mathbb{N} : s(n) \leq s(n + t) \leq s(n + 2t) \leq \dots \leq s(n + mt)\}.$$

Then for all  $t \geq 0$ ,

$$C_t^{(m)} > \frac{1}{2^m m!}.$$

The constant  $1/(2^m m!)$  is maximal.

On the other hand, we could ask for the overall shape of the  $m$ -dimensional probability distribution defined by  $\delta(\cdot, \varepsilon, t)$ .

**Problem 1.** Prove a multidimensional generalization of the theorem by Emme and Hubert [4]: for most  $t$ , the densities  $\text{dens}\{n \in \mathbb{N} : s_2(n + \ell t) - s_2(n) = k_\ell \text{ for } 1 \leq \ell \leq m\}$  should define a probability distribution that is close to a multivariate Gaussian law.

We can now understand the intuition behind the constants in Conjecture 1: a bivariate normal distribution with mean  $(0, 0)$  has one quarter of its total weight in the quadrant  $\{(x, y) \in \mathbb{R}^2 : 0 \leq x, 0 \leq y\}$ , and analogous considerations hold for higher dimensions. Concerning the values  $C_t^{(2)}$ , a bivariate normal distribution with mean  $(0, 0)$  has one *eighth* of its total weight in the octant  $\{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq y\}$ , which corresponds to the complex closed region  $\{re^{2\pi i x} : r \geq 0, 1/8 \leq x \leq 1/4\}$ . Noting the fact that  $s(n + 2t) \geq s(n + t)$  if and only if  $s(n + 2t) - s(n) \geq s(n + t) - s(n)$ , we see

the link between the densities  $\delta(\mathbf{k}, \varepsilon, t)$  with  $\mathbf{k}$  lying in this octant and the values  $C_t^{(2)}$ . Concerning higher dimensions, we note that the  $m!$  sets

$$\{x \in \mathbb{R}^m : \|m\| \leq 1, 0 \leq x_{\sigma(1)} \leq x_{\sigma(2)} \leq \cdots \leq x_{\sigma(m)}\},$$

where  $\sigma$  is a permutation of  $\{1, \dots, m\}$ , are unions of line segments  $[0, z]$ , where  $\|z\| = 1$ ; they are pairwise congruent (using a linear transformation that is a permutation matrix in the canonical base); an intersection of two distinct sets of this form has measure zero, and their union comprises a fraction  $1/2^m$  of the unit ball in  $m$  dimensions.

We implemented the computation of the densities  $\delta(\mathbf{k}, \varepsilon, t)$  for two dimensions in the Sage computer algebra system [11], and the resulting Sage worksheet is available from the website of the first named author [10]. We obtain  $c_t^{(2)} \geq c_{951}^{(2)} = 94299/262144 = 0.3597 \dots > 1/4$  for  $t \leq 2^{10}$  and  $C_t^{(2)} \geq c_{991}^{(2)} = 43947/262144 = 0.1676 \dots > 1/8$  for  $t \leq 2^{10}$ . The implementation involves nine two-dimensional arrays of rational numbers (corresponding to the nine possibilities  $(\varepsilon_1, \varepsilon_2) \in \{0, 1, 2\}^2$ ) and each of these calculations took about 5 minutes on a standard machine. We note that we did not optimize the Sage code, and certainly this computation can be sped up significantly.

Other conjectures similar to Conjecture 1 are conceivable: what about the other octants and quadrants  $A$  in the plane (including the borders)? Is it always true that  $1/8$  resp.  $1/4$  is a lower bound for the density  $\text{dens}\{n \in \mathbb{N} : (s(n+t)-s(n), s(n+2t)-s(n)) \in A\}$ ? We leave this question open for future discussion, but we note that the analogous problem in one dimension is true for almost all  $t$ : we also have

$$\tilde{c}_t = \text{dens}\{n \in \mathbb{N} : s(n+t) \leq s(n)\} > 1/2$$

for  $t$  in a set having asymptotic density 1 [3]. In other words, usually the median of the probability distribution defined by  $\delta(\cdot, t)$  is very close to the mean value (which is 0). We believe that  $\tilde{c}_t \geq 1/2$  is true for all  $t$ , which complements Cusick's conjecture (1.1) (note that  $\tilde{c}_1 = 1/2$ , thus there is no strict inequality).

We expect that nontrivial statements on both Conjecture 1 and Problem 1, at least for small  $m$ , can be obtained by extending the study of moments set forward by Emme and Hubert [4]. This is certainly not easy and will introduce technical difficulties that have to be surmounted.

## REFERENCES

- [1] S. V. AVGUSTINOVICH, D. G. FON-DER-FLAASS, AND A. E. FRID, *Arithmetical complexity of infinite words*, in Words, languages & combinatorics, III (Kyoto, 2000), World Sci. Publ., River Edge, NJ, 2003, pp. 51–62.
- [2] J. BÉSINEAU, *Indépendance statistique d'ensembles liés à la fonction "somme des chiffres"*, Acta Arith., 20 (1972), pp. 401–416.
- [3] M. DRMOTA, M. KAUERS, AND L. SPIEGELHOFER, *On a Conjecture of Cusick Concerning the Sum of Digits of  $n$  and  $n+t$* , SIAM J. Discrete Math., 30 (2016), pp. 621–649.
- [4] J. EMME AND P. HUBERT, *Central Limit Theorem for Probability Measures Defined by Sum-of-digits Function in Base 2*, 2018. To appear in Annali della Scuola Normale Superiore.
- [5] ———, *Normal distribution of correlation measures of binary sum-of-digits functions*, Preprint, 2018. <http://arxiv.org/abs/1810.11234>.
- [6] J. EMME AND A. PRIKHOD'KO, *On the Asymptotic Behavior of Density of Sets Defined by Sum-of-digits Function in Base 2*, Integers, 17 (2017), pp. A58, 28.
- [7] J. KONIECZNY, *Gowers norms for the Thue-Morse and Rudin-Shapiro sequences*, Preprint, 2017. <http://arxiv.org/abs/1611.09985>.

- [8] C. MÜLLNER AND L. SPIEGELHOFFER, *Normality of the Thue–Morse sequence along Piatetski-Shapiro sequences, II*, Israel J. Math., 220 (2017), pp. 691–738.
- [9] L. SPIEGELHOFFER, *Approaching Cusick’s conjecture on the sum-of-digits function*, 2019. Accepted for publication in Integers; <http://arxiv.org/abs/1904.08646>.
- [10] ———, <https://dmg.tuwien.ac.at/spiegelhofer/>, 2019.
- [11] THE SAGE DEVELOPERS, *SageMath, the Sage Mathematics Software System (Version 8.1)*, 2017. <http://www.sagemath.org>.

INSTITUTE OF DISCRETE MATHEMATICS AND GEOMETRY, VIENNA UNIVERSITY OF TECHNOLOGY,  
VIENNA, AUSTRIA

INSTITUT ÉLIE CARTAN DE LORRAINE, UNIVERSITÉ DE LORRAINE, VANDŒUVRE-LÈS-NANCY, FRANCE