Institut für Geometrie
Technische Universität Wien
Ao.Univ.Prof.Dr. Michael Drmota

Statistical Properties of Number Systems

Proposal for a
Project within the research area
"Numbertheoretical Algorithms"

1 Introduction

Number systems naturally appear in number theoretical algorithms since every integer number is usually represented in the binary number system. However, the binary number system is not only used for representing numbers.

For example, the RSA enciphering algorithms (see [40]) needs the computation of powers $a^e \bmod m$, where a, e, m are large integers. By using the binary representation of

$$e = 2^{j_1} + 2^{j_2} + \cdots + 2^{j_k} \qquad (j_1 > j_2 > \cdots > j_k \geq 0)$$

it is possible to compute $a^e \bmod m$ extremely fast in $\mathcal{O}((\log e)(\log m)^2)$ elementary operations. First one computes $a^{2^j} \bmod m$ inductively by consecutive squaring and then by multiplying $a^{2^{j_1}} a^{2^{j_2}} \cdots a^{2^{j_k}} \bmod m$. (Every multiplication needs - in the simplest case - $\mathcal{O}((\log m)^2)$ elementary operations.)

A completely different application of binary or q-ary representations of integers is the construction of (at the moment best known) low discrepancy sequences (see [22, 48]). The Van der Corput sequence, the Halton sequence, and many examples of net sequences are based on these expansions. For example, if a natural number n has the q-ary expansion

$$n = \epsilon_k q^k + \epsilon_{k-1} q^{k-1} + \cdots + \epsilon_1 q + \epsilon_0 \qquad (0 \leq \epsilon_j < q)$$

then the n-th element $\phi_q(n)$ of Van der Corput sequence (with respect to base q) is given by

$$\phi_q(n) = \epsilon_0 q^{-1} + \epsilon_1 q^{-2} + \cdots + \epsilon_{k-1} q^{-k} + \epsilon_k q^{-k-1} \in [0, 1).$$

It is well known ([22, 41, 48]) that the discrepancy[1] of this sequence is bounded by

$$D_N(\phi_q(n)) = \mathcal{O}\left(\frac{\log N}{N}\right)$$

which is optimal (with respect to the order of magnitude).

Moreover, number systems appear in the theoretical analysis of various algorithms, too. For example, the analysis of the Weyl sequence $n\alpha$ mod 1, the prototype of random numbers and of low discrepancy sequences, respectively, essentially relies on the Ostrowski expansion (see Schoissengeier [56]), where the denominators q_m of the convergents $\frac{p_m}{q_m} = [a_0; a_1, \ldots, a_m]$ of $\alpha = [a_0; a_1, a_2, \ldots]$ are used instead

[1] The notion of discrepancy is defined in the next section.

of the binary basis 2^k. One (crude) estimate reads as follows. Suppose that N has the expansion

$$N = \epsilon_k q_k + \epsilon_{k-1} q_{k-1} + \cdots + \epsilon_1 q_1 + \epsilon_0 q_0 \qquad (0 \le \epsilon_j \le a_{j+1})$$

with $\epsilon_k \neq 0$ then the discrepancy of the sequence $n\alpha$ mod 1 is bounded below and above by

$$c_1 \left(\sum_{j=1}^{k} a_k + \epsilon_k \right) \le \max_{M \le N} \left(M D_M(n\alpha) \right) \le c_2 \left(\sum_{j=1}^{k} a_k + \epsilon_k \right)$$

with two absolute constants $c_1, c_2 > 0$.

2 Status of Research

A possible approach to a better understanding of number theoretical algorithms and of specific random number generators and low discrepancy sequences, which are related to number systems, is to look at the statistical distribution of specific parameters of the involved number system.

First of all let us introduce the following **general notion of a number system**. Let $G = (G_j)_{j\ge0}$ be a strictly increasing sequence of integers with $G_0 = 1$. Then every non-negative integer n has a (unique) proper G-ary digital expansion

$$n = \sum_{j\ge0} \epsilon_j(n) G_j$$

with integer digits $\epsilon_j(n) \ge 0$ provided that

$$\sum_{j<k} \epsilon_j(n) G_j < G_k$$

for all $k \ge 0$.

Of course this notion covers all examples mentioned above. The sequence $G_j = 2^j$ generates the binary expansion and more generally $G_j = q^j$ (with q a positive integer) the q-ary expansion. Furthermore, the sequence $G_j = q_j$ (with q_m the denominiators of the convergents $\frac{p_m}{q_m} = [a_0; a_1, \ldots, a_m]$ of an irrational $\alpha = [a_0; a_1, a_2, \ldots])$ induces the Ostrowski expansion. Further examples are Cantor representations, where G_j is of the form $G_j = b_0 b_1 \cdots b_j$ (with positive integers b_j), and the Zeckendorf representations, where $G_j = F_{j+2}$ are the Fibonacci numbers. More generally, one also consideres G_j which are solutions of a linear recurrence with constant coefficients.

A very significant functional of number systems is the sum-of-digits function $s_G(n)$ which is given by

$$s_G(n) = \sum_{j\ge0} \epsilon_j(n).$$

We use this functional to explain in detail what kinds of **statistical properties of number systems** could be of interest. The first aim is to get an insight into the distribution of $s_G(n)$, i.e. into the behaviour of the numbers

$$a_{Nk} = |\{n < N : s_G(n) = k\}|.$$

It is convenient to consider a related sequence of (discrete) random variables X_N, $N \ge 0$, defined by

$$\mathbf{Pr}[X_N = k] = \frac{a_{Nk}}{N}.$$

Expected value and variance of X_N are given by

$$\mathbf{E}X_N = \frac{1}{N}\sum_{n<N} s_G(n) \quad \text{and by} \quad \mathbf{V}X_N = \frac{1}{N}\sum_{n<N} (s_G(n) - \mathbf{E}X_N)^2 \tag{1}$$

There is a vast literature concerning asymptotic properties of $\mathbf{E}X_N$ and $\mathbf{V}X_N$ and on the distribution of X_N.

Asymptotic and exact formulas for $\mathbf{E}X_N$ are due to Bush [6], Bellman and Shapiro [3], Delange [12], and Trollope [62] for q-ary digital expansions and due to Pethö and Tichy [50] and Grabner and Tichy [35, 36] for G-ary digital expansions with respect to linear recurrences. Corresponding formulas for higher moments $\mathbf{E}X_N^d$ and for the variance $\mathbf{V}X_N$ can be found in Coquet [10], Kirschenhofer [39], Kennedy and Cooper [37], Grabner, Kirschenhofer, Prodinger, and Tichy [34], and in Dumont and Thomas [23]. All known results are of the form

$$\mathbf{E}X_N \sim C_1 \log N \quad \text{and} \quad \mathbf{V}X_N \sim C_2 \log N$$

with constants $C_1, C_2 > 0$ depending on the number system $G = (G_j)$.

The asymptotic distribution of X_N and related problems are discussed in Schmidt [53], Schmid [52], Bassily and Katai [2], in Dumont and Thomas [24] and in Drmota and Gajdosik [16], where a new appoach using generating functions is presented. Again we always have the same result, a central limit theorem of the form

$$\frac{X_N - \mathbf{E}X_N}{\sqrt{\mathbf{V}X_N}} \to N(0,1).$$

It is also very interesting to study not only the sum-of-digits function of all non-negative integers but to restrict to a subsequence, e.g. to an arithmetic sequence $an + b$ (see Schmid [52]) or to values of a polynomial $f(n)$ (see Bassily and Kátai [2]). Other sequences of interest are the sequence of primes p_n (or the values of a polynomial $f(p_n)$ with prime argument), see [2]. In the case of the q-ary number system it is possible to prove corresponding asymptotic expansions for mean value and variance of the above type and a central limit law.

Another aspect is the distribution of the sum-of-digits function in residue classes, e.g. the number of $n < N$ with even $s_G(n)$ and the number of $n < N$ with odd $s_G(n)$. It was already known to Gelfond [29] that the sum-of-digits function (with respect to the q-ary number system) is always uniformly distributed in residue classes $\bmod m$ if m and $q-1$ are coprime, i.e.

$$\#\{n < N : s_q(n) \equiv c \bmod m\} = \frac{N}{m} + o(N).$$

The more interesting problem is consider the error term. It is known that this error term is (asmptotically) of the form $N^\alpha F(\log N)$ with $0 < \alpha < 1$ and a periodic continuous function F, see [31, 21].

Gelfond [29] also stated a conjecture on the correlation between the sum-of-digit functions with respect to different bases. Suppose that $q_1, q_2, \ldots, q_k$ are multiplicatively independent and $m_1, m_2, \ldots, m_k$ are positive integers such that m_j and $q_j - 1$ are coprime. Then he conjectured that

$$\#\{n < N : s_{q_1}(n) \equiv c_1 \bmod m_1, \ldots, s_{q_k}(n) \equiv c_k \bmod m_k\} = \frac{N}{m_1 \cdots m_k} + o(N).$$

with an error term of order $\mathcal{O}(N^\alpha)$ with $\alpha < 1$. Besinéau [4] could prove that the error term is really $o(N)$, i.e. the vector sequence $(s_{q_1}(n), \ldots, s_{q_k}(n))$ is really uniformly distributed in residue classes modulo $(m_1, \ldots, m_k)$. Later Kim [38] could also give an effective upper bound of the conjectured form $\mathcal{O}(N^\alpha)$ with $\alpha < 1$.

A similar theorem holds if we consider only an arithmetic sequence $n = ak + b$. Here we have

$$\#\{n < N : n \equiv b \bmod a,\ s_q(n) \equiv c \bmod m\} = \frac{N}{am} + o(N).$$

This was also known to Gelfond [29]. As above the error term is (asmptotically) of the form $N^\alpha F(\log N)$ for some $0 < \alpha < 1$ and a periodic continuous function F.

Interestingly, there are specific cases, where this function F is strictly positive or strictly negative, e.g. Newman [47] proved that the difference

$$\begin{aligned}\sum_{n<N, n\equiv 0 \bmod 3} (-1)^{s_2(n)} &= \#\{n < N : n \equiv 0 \bmod 3,\ s_2(n) \equiv 0 \bmod 2\} \\ &\quad - \#\{n < N : n \equiv 0 \bmod 3,\ s_2(n) \equiv 0 \bmod 2\}\end{aligned}$$

is always positive. Graber [32] and Grabner, Herendi, and Tichy [33] proved that there is a similar phenomenon if the module 3 is replaced by 5 or 17. Later Drmota and Skałba [21] proved that all modules which are multiples of 3 or of the form $2^m + 1$ have this property. This approach was generalized by Leinfellner [44, 45] to modules of the form $(2^m + 1)/3$. This shows that a *bias* in the error term appears quite frequently. Conversely, Drmota and Skałba [19, 20] showed that the number of prime modules $\le x$ with this property is $o(x/\log x)$, i.e. there are only *few* primes.

Finally we want to mention that more general expansions, e.g. the Zeckendorf expansion (where $G_j = F_{j+2}$ are the Fibonacci numbers) lead to similar results (see [21]). Furthermore, the error term

$$\sum_{n<N} (-1)^{s_F(n)} = \#\{n < N : s_F(n) \equiv 0 \bmod 2\} - \#\{n < N : s_F(n) \equiv 0 \bmod 2\}$$

satisfies a central limit law [21, 28].

The sum-of-digits function can also be used to generate uniformly distributed sequences. A sequence (x_n) of real numbers in the unit interval $[0, 1)$ is called *uniformly distributed modulo* 1 if the discrepancy

$$D_N(x_n) = \sup_{0\le a<b<1} \left| \frac{1}{N} \sum_{n=0}^{N-1} \chi_{[a,b]}(x_n) - (b-a) \right|,$$

satisfies

$$\lim_{N\to\infty} D_N(x_n) = 0.$$

(Here χ_E denotes the characteristic function of the set E.) It has been observed by Coquet [9] that the sequence $s(n)\alpha \bmod 1$ is uniformly distributed modulo 1. Upper and lower bounds for the discrepancy can be found in Larcher [42], Larcher and Tichy [43], Tichy and Turnwald [60, 61], Pethő and Tichy [50], Grabner and Tichy [35, 36], and in Drmota [13].

3 Project aims

The general aim of this project is to investigate the **statistical properties of specific functionals of general number systems and related problems** in order to get an better insight into **number theoretical algorithms** and into specific **random number generators** resp. **low discrepancy sequences**.

In particular it is planned to consider the following topics:

1. **Correlations between digits**
2. **The sum-of-digits function and prime numbers**
3. **Digital expansions and random sequences**

ad 1. There are many possibilities to analyze correlations between digits, e.g. in [1, 14] the number of digit skips for specific number systems is discussed. We want to find methods for more general number systems, e.g. Parry expansions.

Another point is to find a stochastic process modelling digital expansions. One possiblily was introduced by Manstavicius [46]. We want to to discuss an alternate approach.

Finally we want to treat to correlation between different digital expansions.

ad 2. Prime numbers play a central role in number theoretical algorithms. Therefore it is a natural question whether the statistical properties of number systems change if only the expansions of prime numbers are considered instead of all non-negative integers, e.g. there is no difference in the distribution of the sum-of-digits funcion (see [2]). However, it is an unsolved problem whether there are infinitely many prime numbers with even binary digits or not.

Furthermore we want to get a better insight into those modulues q for which the sum-of-digits function the sum $S_q(N) = \sum_{j<N/q}(-1)^{s(jq)}$ is always positive.

ad 3. It was already mentioned that q-ary expansions are used to construct random numbers and low discrepancy sequences, e.g. the Van der Corput sequence and the $s(n)\alpha \bmod 1$ are uniformly distributed modulo 1.

In the third part of the project we want to discuss the discrepancy of (multi dimensional) sequences of the kind $(s_{q_1}(n)\alpha_1, \dots, s_{q_k}(n)\alpha_k) \bmod 1$ in terms of the continued fraction expansions of α_j etc.

We also want to discuss the influence of statistical properties of number systems to the construction of net sequences (which generalize the Van der Corput sequence) which are the best known low discrepancy sequences.

4 Detailled list of topics

4.1 Correlations between digits

4.1.1 Blocks in digital expansions

In section 2 we introduced a general notion of a number system $G = (G_j)_{j\ge 0}$. Now assume that all digits are uniformly bounded by some (integer) bound Q (i.e. the sequence of fractions G_{j+1}/G_j is bounded) and let $F : \{0,1,\dots,Q\}^{L+1} \to \mathbf{R}$ be any given function (for some $L \ge 0$) with $F(0,0,\dots,0) = 0$. Furthermore, set

$$s_{G,F}(n) = \sum_{j\ge 0} F\left(\epsilon_{j+L}(n), \epsilon_{j+L-1}(n), \dots, \epsilon_j(n)\right).$$

This means that we consider a weighted sum over all subsequent digitel patterns of length $L+1$ of the digital expansion of n. For example, for $L = 0$ and $F(\epsilon) = \epsilon$

we just obtain the sum-of-digits function, or if $L = 1$ and $F(\epsilon, \eta) = 1 - \delta_{\epsilon,\eta}$ ($\delta_{x,y}$ denoting the Kronecker delta) then $s_{G,F}(n)$ is just counting the number of times that a digit is different from the preceding one etc.

As above (see section 2 concerning the sum-of-digits function) it is convenient to consider a related sequence of random variables $X_{N,F}$, $N \geq 1$, defined by

$$\mathbf{Pr}[X_{N,F} \leq x] = \frac{1}{N} \#\{n < N : s_{G,F}(n) \leq x\}$$

Expected value and variance of X_N are given by

$$\mathbf{E}X_{N,F} = \frac{1}{N} \sum_{n<N} s_{G,F}(n) \quad \text{and by} \quad \mathbf{V}X_{N,F} = \frac{1}{N} \sum_{n<N} (s_{G,F}(n) - \mathbf{E}X_{N,F})^2. \qquad (2)$$

Kirschenhofer [39], Cateland [7], Tenenbaum [59], and Barat, Tichy, and Tijdeman [1] obtained asymptotic expansions for the expected value $\mathbf{E}X_{N,F}$. Furthermore, Drmota [14] proved that $X_{N,F}$ satisfies a central limit theorem of the form

$$\frac{X_{N,F} - c_1 \log N}{\sqrt{c_2 \log N}} \to N(0,1)$$

if the number system G is either q-ary or G_j satisfies a linear recurrene of finite degree with constant coefficients.

A first aim is to generalize these results to more general number systems, e.g. to the Parry system (see Parry [49] and Grabner and Tichy [36]).

Another aspect is discussed by Barat, Tichy, and Tijdeman [1]. They provide lower bounds for the number of times that a digit is different from the preceding one of the numbers a^k, $k \geq 0$, in the q-ary number system if a and q are coprime. (They also treat more general number systems.) However, nothing is known about the *mean value*

$$\frac{1}{K} \sum_{k \leq K} s_{q,F}(a^k)$$

or the distribution of $s_{q,F}(a^k)$. We will try to treat this problem, too.

4.1.2 Sum-of-digits function

If we just consider the sum-of-digits function we already introduced the random variables X_N given by

$$\mathbf{Pr}\,[X_N = k] = \frac{1}{N} \#\{n < N : s_G(n) = k\}.$$

It was mentioned in section 2 that (in many cases) this random variable (suitable normalized) satisfies a central limit theorem

$$\frac{X_N - c_1 \log N}{\sqrt{c_2 \log N}} \to N(0,1)$$

with constants $c_1, c_2 > 0$ depending on the number system G.

Now let $Y_N(t)$, $t \geq 0$, be a sequence of stochastic processes defined by

$$Y_N(t) = \frac{X_{N^t} - E(X_{N^t})}{\sqrt{c_2 \log(N)}}$$

which describes the *variation* of the distribution of $s_G(n)$ if we consider *more or less* numbers. Our conjecture is that $Y_N(t)$ converges (weakly) to Brownian motian $B(t)$

$$(Y_N(t), t \geq 0) \to (B(t), t \geq 0).$$

It is planned to use Prohorov's theorem (see [5]) which states that we have to prove

1) finitely dimensional convergence and

2) tightness.

It seems that both questions can be treated by generalizing the generation function approach of Drmota and Gajdosik [16] (compare also with Drmota and Skałba [20] and Drmota [14]). It should be mentioned that the generating function approach was already very successful for proving functional limit theorems for random trees, see Drmota and Gittenberger [17, 18] and Gittenberger [30].

We also want to note that another approach leading to a functional limit theorem involving Brownian motion was introduced by Manstavicius [46]. He uses the truncated sum-of-digits function

$$s_{G;l}(n) = \sum_{j=0}^{l} \epsilon_j(n)$$

where only the first $l+1$ digits are considered. Let $X_{N;j}$ denote the random variables defined by

$$\mathbf{Pr}\left[X_{N;j} = k\right] = \frac{1}{N}\#\{n < N : s_{G;j}(n) = k\}$$

and $Z_N(t)$, $t \geq 0$, the sequence of stochastic processes given by

$$Z_N(t) = \frac{X_{N;Nt} - E(X_{N;Nt})}{\sqrt{c_2 \log(N)}}.$$

Then Manstavicius [46] proved in the case of Cantor representations that

$$(Z_N(t), t \geq 0) \to (B(t), t \geq 0).$$

4.1.3 Sums of digits concerning different bases

As already mentioned in section 2 there are some results concering the common distribution of the sum-of-digits functions $s_{q_1}(n), \dots, s_{q_k}(n)$ with respect to multiplicatively independent q_j. It is known that sequences of this kind are uniformly distributed in residue classes (see Gelfond [29], Besineau [4],and Kim [38])

Another problem has been considered by Senge and Straus [57]. They proved that if q_1 and q_2 are coprime and c in any given positive constant then there are only finitely many $n \geq 0$ such that

$$s_{q_1}(n) \leq c \text{ and } s_{q_2}(n) \leq c.$$

This result was later generalized and sharpended by Stewart [58], Schlickewei [54, 55] and by Pethő and Tichy [51]. The proofs use Baker's method on linear forms of logarithms and the p-adic version of Schmidt's subspace theorem by Schlickewei applied to S-unit equations.

Another question which seems to be open is to determine the joint distribution of

$$s_{q_1}(n) \text{ and } s_{q_2}(n)$$

if q_1 and q_2 are coprime. Let X_N by $X_N = (X_{1,N}, X_{2,N})$ denote the random vector defined by

$$\mathbf{Pr}\left[(X_{1,N}X_{2,N}) = (k_1, k_2)\right] = \frac{1}{N}\#\left\{n < N : s_{q_1}(n) = k_1, s_{q_2}(n) = k_2\right\}.$$

The conjecture is that X_N (suitable normalized) satisfies a central limit theorem. It seems that those methods presented in Kim [38] (maybe combined with Baker's or Schlickewei's theorems) will lead to a positive answer of this question. Kim essentially gives bounds for exponential sums involving $s_{q_1}(n)$ and $s_{q_2}(n)$. Such sums are similar to those sums which are needed to prove a weak limit theorem via characteristic functions.

4.2 The sum-of-digits function and prime numbers

4.2.1 Sum-of-digits function

Most of the above questions on properties of the sum-of-digits function or distributional problems can also be stated in a form where only primes $p < N$ are considered instead of all non-negative integers $n < N$.

Firstly, let Z_N be a random variable defined by

$$\mathbf{Pr}\left[Z_N = k\right] = \frac{1}{\pi(N)} \# \left\{p < N : p \in \mathbf{P}, s_G(n) = k\right\}$$

It is known (see Bassily and Kátai [2]) that in the q-ary case there is a weak limit theorem of the form

$$\frac{Z_N - c_1 \cdot \log N}{\sqrt{c_2 \cdot \log N}} \to N\left(0, 1\right) \tag{3}$$

The first question is to generalize this property to more general number systems (which seems to be the easy part) and the second question is to provide more detailed information on the numbers

$$\# \left\{p < N : p \in \mathbf{P}, s_G(n) = k\right\},$$

e.g. an asymptotic expansion (which seems to be very hard) or at least some local infomation of the form

$$\frac{1}{\pi(N)} \sum_{k=a}^{a+c_n} \# \left\{p < N : p \in \mathbf{P}, s_G(n) = k\right\}$$
$$\sim \int_a^{a+c_n} \frac{1}{\sqrt{2\pi c_2 \log N}} \exp\left(-\frac{(x - c_1 \log N)^2}{2c_2 \log N}\right) dx$$

for some sequence $c_n \to \infty$ with $c_n = o(\sqrt{\log N})$.

4.2.2 Parity of the sum-of-digits functions

By examining the sum of (binary) digits function $s_2(p)$ for primes p it seems that it is more likely that $s_2(p)$ is odd than even. However, this is just an heuristic observation. Almost nothing is known about the set of primes $p \le x$ with an even sum of binary digits $s_2(p)$ resp. even sum of binary digits $s_2(p)$. (Of course, if it where possible to prove asymptotic relations for the numbers $\# \left\{p < N : p \in \mathbf{P},\ s_G(n) = k\right\}$, as proposed in the previous section, then one would get some information.)

The first question is whether both sets

$$\{p \in \mathbf{P} : s_2(p) \equiv 0 \bmod 2\}$$

and

$$\{p \in \mathbf{P} : s_2(p) \equiv 1 \bmod 2\}$$

are infinite? This problem is unsolved. However it is known that the set of numbers $n \le x$ that are prime or the product of two primes and that have an even sum of digits has size

$$\# \left\{n \le x : n \in \mathbf{P} \vee (n = n_1 \cdot n_2 \wedge n_1, n_2 \in \mathbf{P}), s_2(n) \equiv 0 \bmod 2\right\} \ge c' n$$

with some constant $c' > 0$. A similar estimate holds for the corresponing set with an odd sum of digits

$$\# \left\{n \le x : n \in \mathbf{P} \vee (n = n_1 \cdot n_2 \wedge n_1, n_2 \in \mathbf{P}), s_2(n) \equiv 1 \bmod 2\right\} \ge c'' n,$$

see Fouvry and Mauduit [26, 27]. A much stronger assertion is the following:

$$S(N) = \sum_{p\le N} (-1)^{s_2(p)} = o(N) \qquad (N\to\infty)$$

It would imply that $\#\{p \le x : s_2(p) \equiv 0 \bmod 2\} \sim \frac{1}{2}\pi(x)$. (By the way the above observation says that $S(N) < 0$.)

It is clear that the presented problems are surely extremely difficult to solve. However, we hope that a *local version* of the above weak limit theorem (3) maybe gives some new light on these kinds of problems.

4.2.3 Rarefied sums of digital sums

In section 2 we reported on some properties of the sums

$$S_q(n) = \sum_{j<n,\ j\equiv 0 \bmod q} (-1)^{s_2(j)}$$

If q are positive integers divisible by 3 or of the form $p = 2^m+1$ resp. $p = (2^m+1)/3$ then $S_p(n) > 0$ (for sufficiently large n), see Newman [47], Grabner [32, 33], Drmota and Skałba [21]. Let us call the set of positive integers q with the property that $S_q(n) > 0$ (for sufficiently large n) **Pos**, i.e. $q \in \mathbf{Pos}$ means that there are more multiples of q with an even binary sum of digits than with an odd one.

Drmota and Skałba [21] showed that

$$\#\,\{p \le x : p \in \mathbf{P},\ p \in \mathbf{Pos}\} = o(\pi(x)) \qquad (x\to\infty), \tag{4}$$

i.e. there are only *few* primes with this positivity property. The main ingrediences for the proof of (4) are that $p \in \mathbf{P}\cap\mathbf{Pos}$ implies that the multiplicative order $\mathrm{ord}_p(2)$ of $2 \bmod p$ is bounded by

$$\mathrm{ord}_p(2) < 16\pi^2\sqrt{p}\,\log p$$

and secondly that for any sequence $\varepsilon_p \to 0$ (as $p\to\infty$)

$$\#\{p \le x :\ s = \mathrm{ord}_p(2) < p^{1/2+\varepsilon_p}\} = o\left(\pi(x)\right) \qquad (x\to\infty). \tag{5}$$

Of course, these two facts prove (4) The first property is proved in [21] and the second one is a theorem by Erdős [25] which has never been improved. It is therefore a natural question to improve (5) which directly sharpens (4).

It is a nice observation that any Fermat prime p is in **Pos**. So if $\mathbf{P}\cap\mathbf{Pos}$ were finite, there would exist only finitely many Fermat primes. However, it is very likely that the set $\mathbf{P}\cap\mathbf{Pos}$ is infinite.

4.2.4 Classification

The function $S_q(n) = \sum_{j<n,\ j\equiv 0 \bmod n} (-1)^{s(j)}$ and the set **Pos** can be studied a little further. We already mentioned that all multiples of 3 and numbers of the form $q = 2^m+1$ resp. $q = (2^m+1)/3$ are contained in **Pos**.

It is also clear (see [44, 45]) that *many* multiples of 5 are in **Pos** but surely not all of them. It is an interesting problem to decide which multiples of 5 are in **Pos** and which are not. Of course, similar question can be formulated for the multiples of 17 etc.

It is surely a very difficult problem to find a complete description but (in view of the known results) there is some hope to obtain (at least) more information about the multiples of 5.

4.3 Digital expansions and random sequences

4.3.1 Discrepancy

As mentioned in section 2 the discrepancy $D_N(x_n)$ of a real sequence $x_n \in [0,1)$ is defined by

$$D_N(x_n) = \sup_{0 \le a < b < 1} \left| \frac{1}{N} \sum_{n=0}^{N-1} \chi_{[a,b]}(x_n) - (b-a) \right|,$$

see Kuipers and Niederreiter [41] and Drmota and Tichy [22]. The discrepancy $D_N(x_n)$ measures the deviation of the (normalized) counting measure induced by the sequence $x_0, x_1, \ldots, x_{N-1}$ from the uniform distribution. A sequence (x_n) is called uniformly distributed modulo 1, if

$$\lim_{N \to \infty} D_N(x_n) = 0.$$

Explicitly known sequences with *small discrepancy* $D_N(x_n)$ can be used for numerical integration (especially in higher dimensions) and for random number generators (for details see [22, 41]).

One of the best known uniformly distributed sequence is the Weyl sequence $x_n = n\alpha \bmod 1$ (with an irrational α). Surprisingly, sequences of this kind remain uniformly distributed if n is replaced by the sum-of-digits function $s_G(n)$ (for various kinds of number systems G), see Coquet [9], Delange [11], Drmota [13], Larcher [42], Larcher and Tichy [43], and Tichy and Turnwald [60, 61].

There are lots of discrepancy estimates for sequences of the kind $x_n = s_G(n)\alpha \bmod 1$, mostly depending on the approximation type of α. This indicates that the continued fraction expansion should be of dominant importance when one wants to get more precise information about the magnitude of the discrepancy of $s_G(n)\alpha \bmod 1$. (For the Weyl sequence this problem has been completely solved by Schoissengeier [56].) Hence, we are looking for approximations for the discrepancy depending on the continued fraction expansion of α.

4.3.2 Common distribution

Now let $q_1, \ldots, q_k$ be multiplicatively independent positive integers and $\alpha_1, \ldots, \alpha_k$ irrationals. Then it is very likely that the k-dimensional sequence

$$(s_{q_1}(n)\alpha_1, \ldots, s_{q_k}(n)\alpha_k) \bmod 1$$

is uniformly distributed (in the k-dimensional unit cube $[0,1)^k$). It seems that the methods of Kim [38] directly generalize to prove such a property. As mentioned above, Kim gives estimates for exponential sums involving $s_{q_1}(n), \ldots, s_{q_k}(n)$. These exponential sums are almost the same to those which are needed to prove bounds for the discrepancy via Erdős-Turan-Koksma's inequality or to prove uniform distribution via Weyl's criterion (see [22, 41]).

It is also an interesting problem to obtain non-trivial lower bounds for the discrepancy.

4.3.3 Net Sequences

Net sequences are sequences in the unit cube $[0,1)^s$ which satisfy certain combinatorial properties (see [22, 48]). Net sequences have a very small discrepancy (depeding on the parameters (t,s)) and are therefore well suited for numerical integration (even if one uses only few points in high dimensional integrals).

The usual construction of net sequences uses the digital expansions of n in the q-ary number system, e.g. the Van der Corput sequence $\phi_q(n)$ introduced in section 2 is a net sequence in the one dimensional case. Although there are general constructions (via digital expansions) it is completely non-trivial problem to construct net sequences for arbitrary parameters (t, s), especially for small t.

It seems that detailed information on the statistical properties of digital expansions may be useful for constructing net sequences with small discrepancy. Therefore we also want to focus on construction methods for net sequences from this point of view.

5 Position in the FSP and international relations

The applicant has not only worked in the proposed fields [13, 14, 15, 16, 17, 18, 19, 20, 21, 22] but there are lots of relations to the work of other applicants of the planned FSP [1, 22, 34, 35, 36, 39, 42, 43, 50, 60, 61].

Moreover there are direct connections to the planned projects by Kirschenhofer (Toolkits for the Average Case-Analysis of Algorithms), by Larcher (Digital Problems in Number-theoretical Numerics) and by Tichy and Kirschenhofer (Algorithmic Solution of Diophantine Equations).

Kirschenhofer wants to exploit tools for the average case-analysis of algorithms. It is apparent that the runtime of many algorithms (e.g. Quick Sort) depend on the binary representation of the number of inputs. This property is reflected in the average case-analysis of these algorithms, they involve exponential sums over the sum-of-digits function etc. Therefore it is seriously expected that parts of the methods which will be developed within the present project will apply to problems of Kirschenhofer's project and vice versa.

Larcher will discuss point sets which are generated with help of digital expansions, e.g. net sequences. For example, he wants to introduce a Walsh series approch to the analysis of generalized sum-of-digis function. Hence, there surely is an interaction in the kind of problems of these two projects.

Finally, Tichy and Kirschenhofer want to discuss Diophantine problems where digital expansions are involved, e.g. they want to exploit the methods of Barat, Tichy, and Tijdeman [1] from a little bit different view as it is proposed here. However, they want to use the same kind of number systems $G = (G_j)_{j\geq 0}$ and of course, similiar methods.

We will also stay in frequent contact with the other groups of the FSP and will try to apply to apply our methods to their problems and vice versa. For example, Hellekalek's test batteries for the computational testing of stochastic properties of pseudo-random-numbers may be applied in the present project etc.

There also exist various international relations e.g. to M. Skałba (Warsaw, Poland) (see [19, 20, 21]) or to E. Manstavicius (Vilnius, Lithuania). Although there is no official (institutional) international cooperation it is planned to cooperate with them within the topics of the proposed project.

6 Research Plan for the first two years and the research institution

Within the first two years it is planned to consider mainly the topics **Correlation between digits** and **Random sequences**.

We want to start with the problems concerning the correlation between the sum-of-digits function $s_{q_j}(n)$, $1 \leq j \leq k$, with respect to multiplicatively independent bases. The methods of Kim [38] are the kind that it should not be too difficult to adapt them to settle the questions of section 4.3.2 and maybe those of section 4.1.3. This could be done within half a year.

In the second half of the first year we want to settle the questions of section 4.3.1 concerning the discrepancy of the sequence $s_G(n)\alpha \bmod 1$ in term of the continued fraction expansion of the irrational α.

The main topics of the second year should be the problems described in sections 4.1.1 and 4.1.2. Starting with the problems of central limit laws we will turn to functional limit theorems.

In order to do all this we need **one full and one half position** of research assistents and **temporary contracts**.

We will also stay in frequent contact with the other groups of the FSP, especially with the groups of Kirschenhofer, Larcher, and Tichy.

The applicant works in the **Department of Geometry** of the **TU Wien**. There is enough room and proper computer equipement to realize the project there. Of course, we need additional software licences etc.

7 Requested Amount

One full position and one half position for a research assistent	ATS 756 000 per year
Travelling	ATS 50 000 per year
Temporary contracts, software	ATS 50 000 per year

For visiting positions we will apply for ATS 50.000.- per year within the coordination project.

References

[1] G. Barat, R. F. Tichy, and R. Tijdeman, Digital blocks in linear numeration systems, *Proceedings of the Number Theory Conference*, Zakopane, 1997, to appear.

[2] N. L. Bassily and I. Kátai, Distribution of the values of q-additive functions on polynomial sequences, *Acta Math. Hung.* **68** (1995), 353–361.

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

[4] J. Bésineau, Indépendence statistique d'ensembles liés à la fonction "sommes des chiffres", *Acta Arith.* **20** (1972), 401–416.

[5] P. Billingsley, Convergence of Probability Measures, John Wiley & Sons, New York, 1968.

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

[7] D. Cateland, Suites digitales et suites k-régulières, *Thèse de doctorat*, Université de Bordeaux 1 (1992).

[8] J. Coquet, *A Summation Formula Related to the Binary Digits*, Invent. math. **73**, 107-115 (1983)

[9] J. Coquet, , Sur certaines suites uniformément équiréparties modulo 1, *Acta Arith.* **36** (1980), 157–162.

[10] J. Coquet, Power sums of digital sums, *J. Number Th.* **22** (1986), 161–176.

[11] H. Delange, On some arithmetic function, *Illinois J. Math.* **2** (1958), 81–87.

[12] H. Delange Sur la fonction sommatoire de la fonction "Somme de Chiffres", *L'Enseignement math.* **21** (1975), 31–77.

[13] M. Drmota, q-Additive functions and well distribution, *Demonst. Math.* **30** (1998), 883–896.

[14] M. Drmota, The Distribution of Patterns in Digital Expansions, *Proceedings Number Theory Conference.*, Graz, 1998, to appear.

[15] M. Drmota and J. Gajdosik, The parity of the sum-of-digits-function of generalized Zeckendorf representations, *Fibonacci Q.* **36** (1998), 3–19.

[16] M. Drmota and J. Gajdosik, The Distribution of the Sum-of-Digits Function, *J. Theor. Nombres Bordx.* **10** (1998), 17–32.

[17] M. Drmota and B. Gittenberger, On the profile of random trees, *Random Struct. Algorithms* **10** (1997), 421–451.

[18] M. Drmota and B. Gittenberger, Strata of random mappings – a combinatorial approach, *Stoch. Proc. Appl.*, to appear.

[19] M. Drmota and M. Skałba, Sign-changes of the Thue-Morse sequence and Dirichlet L-series, *manuscripta math.* **86** (1995), 519–541.

[20] M. Drmota and M. Skałba, The parity of the Zeckendorf sum-of-digits-function, *manuscipta math.*, submitted.

[21] M. Drmota and M. Skałba, Rarified sums of the Thue-Morse sequence, *Trans. Am. Math. Soc.*, to appear.

[22] M. Drmota and R. F. Tichy, , Sequences, Discrepancies, and Applications, *Lecture Notes Math.* **1651**, Springer, Berlin, 1997.

[23] J. M. Dumont and A. Thomas, Digital sum moments and substitutions, *Acta Arith.* **64** (1993), 205–225.

[24] J. M. Dumont and A. Thomas, Gaussian asymptotic properties of the sum-of-digits functions, *J. Number Th.* **62** (1997), 19–38.

[25] P. Erdős, Bemerkungen zu einer Aufgabe von Elementen, *Archiv Math.* **27** (1976), 159–163.

[26] E. Fouvry and C. Mauduit, Sommes des chiffres et nombres presque premiers. *Math. Ann.* **305** (1996), 571-599.

[27] E. Fouvry and C. Mauduit, Methodes de crible et fonctions sommes des chiffres. *Acta Arith.* **77** (1996), 339-351.

[28] J. Gajdosik, Kombinatorische Faktorisierungen und Ziffernentwicklungen, *Dissertation*, TU Wien, 1996.

[29] A. O. Gelfond, Sur les nombres qui ont des propriétés additives et multiplicatives données, *Acta Arith.* **13** (1968), 259–265.

[30] B. Gittenberger, On the contour of random trees, *SIAM J. Discrete Math.*, to appear.

[31] S. Goldstein, K. A. Kelly, and E. R. Speer, The fractal structure of rarefied sums of the Thue-Morse sequence, *J. Number Th.* **42** (1992), 1–19.

[32] P. J. Grabner, A note on the parity of the sum-of-digits function, *Actes* 30$^{\mathrm{e}}$ *Séminaire Lotharingien de Combinatoire*, 1993, 35–42.

[33] P. J. Grabner, T. Herendi, and R. F. Tichy, Fractal digital sums and codes, *Appl. Algebra Engin. Comm. Comput.* **1** (1997), 33–39.

[34] P. J. Grabner, P. Kirschenhofer, H. Prodinger, and R. F. Tichy, On the moments of the sum-of-digits function, in: *Applications of Fibonacci Numbers* **5** (1993), 263–271

[35] P. J. Grabner and R. F. Tichy, Contributions to digit expansions with respect to linear recurrences, *J. Number Th.* **36** (1990), 160–169.

[36] P. Grabner and R. F. Tichy, α-Expansions, linear recurrences, and the sum-of-digits function, *manuscripta math.* **70** (1991), 311–324.

[37] R. E. Kennedy and C. N. Cooper, An extension of a theorem by Cheo and Yien concerning digital sums, *Fibonacci Q.* **29** (1991), 145–149.

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

[39] P. Kirschenhofer, On the variance of the sum of digits function, *Lecture Notes Math.* **1452** (1990), 112–116.

[40] N. Koblitz, *A Course in Number Theory and Cryptography*, Springer, New York, 1987.

[41] L. Kuipers and H. Niederreiter, , Uniform Distribution of Seqences, Wiley, New York, 1974.

[42] G. Larcher, Zur Diskrepanz verallgemeinerter Ziffernsummenfolgen, *Österr. Akad. Wiss. SB II* **202** (1993), 179–185.

[43] G. Larcher and R. F. Tichy, Some number-theoretical properties of generalized sum-of-digit functions, *Acta Arith.* **52** (1989), 183–196.

[44] H. Leinfellner, New Results on Rarefied Sums of the Thue-Morse Sequence, *thesis*, TU Wien, 1998.

[45] H. Leinfellner, New Results on Rarefied Sums of the Thue-Morse Sequence, *Grazer Math. Berichte*, to appear.

[46] E. Manstavicius, Functional approach in the divisor distribution problem, *Acta Math. Hung.* **66** (1995), 343–359.

[47] D. J. Newman, On the number of binary digits in a multiple of three, *Proc. Am. Math. Soc.* **21** (1969), 719–721.

[48] H. Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, *SIAM Conf. Ser. Appl. Math.* **63**, Philadelphia, 1992.

[49] W. Parry, On the β-expansion of real numbers, *Acta Math. Acad. Sci. Hung.*, **12** (1961), 401–416.

[50] A. Pethö and R. F. Tichy, On digit expansions with respect to linear recurrences, *J. Number Th.* **33** (1989), 243–256.

[51] A. Pethö and R. F. Tichy, S-unit equations, linear recurrences and digit expansions, *Publ. Math.* **42** (1993), 145-154.

[52] J. Schmid, The joint distribution of the binary digits of integer multiples, *Acta Arith.* **43** (1984), 391–415.

[53] W. M. Schmidt, The joint distribution the digits of certain integer s-tuples, *Studies in pure mathematics, Mem. of P. Turan* (1983), 605–622.

[54] H. P. Schlickewei, S-unit equations over number fields, *Invent. Math.* **102**(1990), 95-107.

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

[56] J. Schoissengeier, On the discrepancy of $(n\alpha)$, *Acta Arith.*, **44** (1984), 241–279.

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

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

[59] G. Tenenbaum, Sur la non-dérivaibité de fonctions périodiques associées à certaines formules sommatoires, In: *The mathematics of Paul Erdős, I* (R. L. Graham and J. Nesestril eds.), Springer, Berlin, 1997, 117–128.

[60] R. F. Tichy and G. Turnwald, On the discrepancy of some special sequences, *J. Number Theory* **26** (1986), 351–366.

[61] R. F. Tichy and G. Turnwald, Gleichmäßige Diskrepanzabschätzungen für Ziffernsummen, *Anz.Österr. Akad. Wiss.* (1986), 17–21.

[62] H. Trollope, An explicit expression for binary digital sums, *Meth. Mag.* **41** (1968), 21–25.