Marseille 2018

On the digits of primes and squares

Joël RIVAT

Institut de Mathématiques de Marseille, Université d'Aix-Marseille.

in collaboration with

Michael DRMOTA (TU Wien)
Christian MAUDUIT (Marseille)

Background, problems and Results

A.O. Gelfond

Gelfond's paper

In base $q \geqslant 2$ any $n \in \mathbb{N}$ can be written $n = \sum_{j \geqslant 0} \varepsilon_j(n)\, q^j$ where $\varepsilon_j(n) \in \{0, \ldots, q-1\}$.

Gelfond, 1968: The sum of digits $\mathsf{s}(n) = \sum_{j \geqslant 0} \varepsilon_j(n)$ is well distributed in arithmetic progressions: given $m \geqslant 2$ with $(m, q-1) = 1$, there exists an explicit $\sigma_m > 0$ such that

$$\forall m' \in \mathbb{N}^*,\ \forall (n',\ s) \in \mathbb{Z}^2, \qquad \sum_{\substack{n \leqslant x \\ n \equiv n' \bmod m' \\ \mathsf{s}(n) \equiv s \bmod m}} 1 = \frac{x}{mm'} + O(x^{1-\sigma_m}).$$

Gelfond's problems, 1968:

1. Evaluate the number of prime numbers $p \leqslant x$ such that $\mathsf{s}(p) \equiv a \bmod m$.

2. Evaluate the number of integers $n \leqslant x$ such that $\mathsf{s}(P(n)) \equiv a \bmod m$, where P is a suitable polynomial *[for example $P(n) = n^2$]*.

Gelfond's conjecture for primes

Mauduit-Rivat, 2010: If $(q-1)\alpha \in \mathbb{R} \setminus \mathbb{Z}$, there exists $C_q(\alpha) > 0$ and $\sigma_q(\alpha) > 0$,

$$\left| \sum_{p \leqslant x} \exp(2i\pi\alpha\, \mathsf{s}(p)) \right| \leqslant C_q(\alpha)\; x^{1-\sigma_q(\alpha)}.$$

Hence

- For $q \geqslant 2$ the sequence $(\alpha\, \mathsf{s}(p_n))_{n\geqslant 1}$ is equidistributed modulo 1 if and only if $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ (here $(p_n)_{n\geqslant 1}$ denotes the sequence of prime numbers).

- For $q \geqslant 2$, $m \geqslant 2$ such that $(m, q-1) = 1$ and $a \in \mathbb{Z}$,

$$\sum_{\substack{p \leqslant x \\ \mathsf{s}(p) \equiv a \bmod m}} 1 \;\sim\; \frac{1}{m} \sum_{p \leqslant x} 1 \quad (x \to +\infty).$$

Local result for primes

Drmota-Mauduit-Rivat, 2009: uniformly for all integers $k \geqslant 0$ with $(k, q-1) = 1$

$$\#\{p \leqslant x : \mathsf{s}(p) = k\}$$
$$= \frac{q-1}{\varphi(q-1)} \frac{\pi(x)}{\sqrt{2\pi\sigma_q^2 \log_q x}} \left(\exp\left(\frac{-(k - \mu_q \log_q x)^2}{2\sigma_q^2 \log_q x} \right) + O((\log x)^{-\frac{1}{2}+\varepsilon}) \right),$$

where

$$\mu_q = \frac{q-1}{2}, \quad \sigma_q^2 = \frac{q^2-1}{12}$$

and $\varepsilon > 0$ is arbitrary but fixed.

Such a local result was considered by **Erdős** as "hopelessly difficult".

Gelfond's conjecture for squares

Mauduit-Rivat, 2009: if $(q-1)\alpha \in \mathbb{R} \setminus \mathbb{Z}$, there exist $C_q(\alpha) > 0$ and $\sigma_q(\alpha) > 0$,

$$\left| \sum_{n \leqslant x} \exp(2i\pi\alpha\, \mathsf{s}(n^2)) \right| \leqslant C_q(\alpha)\ x^{1-\sigma_q(\alpha)}.$$

Hence

- For $q \geqslant 2$ the sequence $(\alpha\, \mathsf{s}(n^2))_{n \geqslant 1}$ is equidistributed modulo 1 if and only if $\alpha \in \mathbb{R} \setminus \mathbb{Q}$.

- For $q \geqslant 2$, $m \geqslant 2$ such that $(m, q-1) = 1$ and $a \in \mathbb{Z}$,

$$\sum_{\substack{n \leqslant x \\ \mathsf{s}(n^2) \equiv a \bmod m}} 1 \ \sim\ \frac{x}{m} \quad (x \to +\infty).$$

The Rudin-Shapiro sequence

Let

$$f(n) = \mathrm{e}\left(\frac{1}{2}\sum_{j\geqslant 1}\varepsilon_{j-1}(n)\varepsilon_j(n)\right) = (-1)^{\sum_{j\geqslant 1}\varepsilon_{j-1}(n)\varepsilon_j(n)}$$

$\widehat{f_\lambda}$ are flat polynomials:

$$\left\|\widehat{f_\lambda}\right\|_\infty \leqslant 2^{\frac{1-\lambda}{2}}.$$

Since $\left\|\widehat{f_\lambda}\right\|_2 = 1$, by Cauchy-Schwarz it is easy to deduce that

$$2^{\frac{\lambda-1}{2}} \leqslant \left\|\widehat{f_\lambda}\right\|_1 \leqslant 2^{\frac{\lambda}{2}}$$

The proof for the sum of digits function requires $\left\|\widehat{f_\lambda}\right\|_1 = O(2^{\eta\lambda})$ with $\eta < \frac{1}{2}$.

This is not satisfied for the Rudin-Shapiro sequence !!!

Rudin-Shapiro sequences of order δ

Let $\delta \in \mathbb{N}$ and $\beta_\delta(n)$ the number of occurencies of patterns $1\underbrace{*\cdots*}_{\delta}1$, *i.e.* of the form $1w1$ (where $w \in \{0,1\}^\delta$) in the representation of n:

$$\beta_\delta(n) = \sum_{k \geqslant \delta+1} \varepsilon_{k-\delta-1}(n)\,\varepsilon_k(n).$$

Mauduit-Rivat, 2014: for any $\delta \in \mathbb{N}$, $\alpha \in \mathbb{R}$, $\vartheta \in \mathbb{R}$ and $x \geqslant 2$, there exists explicit constants $C(\delta)$ and $\sigma(\alpha) > 0$ such that

$$\left|\sum_{n \leqslant x} \Lambda(n)\,\mathsf{e}\left(\beta_\delta(n)\alpha + \vartheta n\right)\right| \leqslant C(\delta)\,(\log x)^{\frac{11}{4}}\,x^{1-\sigma(\alpha)}$$

and

$$\left|\sum_{n \leqslant x} \mu(n)\,\mathsf{e}\left(\beta_\delta(n)\alpha + \vartheta n\right)\right| \leqslant C(\delta)\,(\log x)^{\frac{11}{4}}\,x^{1-\sigma(\alpha)}.$$

Rudin-Shapiro sequences of degree d

Let $d \in \mathbb{N}$ with $d \geqslant 2$ and $b_d(n)$ denote the number of occurencies of $\underbrace{1 \cdots 1}_{d}$ *i.e.* blocks of d consecutive 1's in the representation of n in base 2:

$$b_d(n) = \sum_{k \geqslant d-1} \varepsilon_{k-d+1}(n) \cdots \varepsilon_k(n).$$

Mauduit-Rivat, 2014: for any $d \in \mathbb{N}$ with $d \geqslant 2$, $\alpha \in \mathbb{R}$, $\vartheta \in \mathbb{R}$ and $x \geqslant 2$ there exist an explicit constant $\sigma(d, \alpha) > 0$ such that

$$\left| \sum_{n \leqslant x} \Lambda(n)\, \mathsf{e}\left(b_d(n)\alpha + \vartheta n\right) \right| \ll (\log x)^{\frac{11}{4}}\, x^{1-\sigma(d,\alpha)},$$

$$\left| \sum_{n \leqslant x} \mu(n)\, \mathsf{e}\left(b_d(n)\alpha + \vartheta n\right) \right| \ll (\log x)^{\frac{11}{4}}\, x^{1-\sigma(d,\alpha)}.$$

General result – Definitions

Let $\mathbb{U} = \{z \in \mathbb{C},\ |z| = 1\}$.

Definition 1 *A function $f : \mathbb{N} \to \mathbb{U}$ has the* carry property *if, uniformly for $(\lambda, \kappa, \rho) \in \mathbb{N}^3$ with $\rho < \lambda$, the number of integers $0 \leqslant \ell < q^\lambda$ such that there exists $(k_1, k_2) \in \{0, \dots, q^\kappa - 1\}^2$ with*

$$f(\ell q^\kappa + k_1 + k_2)\,\overline{f(\ell q^\kappa + k_1)} \neq f_{\kappa+\rho}(\ell q^\kappa + k_1 + k_2)\,\overline{f_{\kappa+\rho}(\ell q^\kappa + k_1)}$$

is at most $O(q^{\lambda-\rho})$ where the implied constant may depend only on q and f.

Definition 2 *Given a non decreasing function $\gamma : \mathbb{R} \to \mathbb{R}$ satisfying $\lim_{\lambda\to+\infty} \gamma(\lambda) = +\infty$ and $c > 0$ we denote by $\mathcal{F}_{\gamma,c}$ the set of functions $f : \mathbb{N} \to \mathbb{U}$ such that for $(\kappa, \lambda) \in \mathbb{N}^2$ with $\kappa \leqslant c\lambda$ and $t \in \mathbb{R}$:*

$$\left| q^{-\lambda} \sum_{0 \leqslant u < q^\lambda} f(uq^\kappa)\, \mathsf{e}\,(-ut) \right| \leqslant q^{-\gamma(\lambda)}.$$

General result

Let $\gamma : \mathbb{R} \to \mathbb{R}$ be a non decreasing function satisfying $\lim_{\lambda \to +\infty} \gamma(\lambda) = +\infty$, $c \geqslant 10$ and $f : \mathbb{N} \to \mathbb{U}$ be a function satisfying Definition 1 and $f \in \mathcal{F}_{\gamma,c}$ in Definition 2. Then for any $\theta \in \mathbb{R}$ we have

$$\left| \sum_{n \leqslant x} \Lambda(n) f(n) \, \mathsf{e}\,(\theta n) \right| \ll c_1(q) (\log x)^{c_2(q)} \, x \, q^{-\gamma(2\lfloor (\log x)/80 \log q \rfloor)/20},$$

with explicit $c_1(q)$ and $c_2(q)$.

Of course the same estimate holds if we replace the von Mangoldt function Λ by the Möbius function μ.

Müllner has recently extended this result to all automatic sequences !

Gelfond's conjecture for polynomials

Drmota-Mauduit-Rivat, 2011: let $d \geqslant 2$, $q \geqslant q_0(d)$, and $P \in \mathbb{Z}[X]$ of degree d such that $P(\mathbb{N}) \subset \mathbb{N}$ for which the leading coefficient a_d is co-prime to q. If $(q-1)\alpha \in \mathbb{R} \setminus \mathbb{Z}$ then there exists $c = c(q,d) > 0$ with

$$\sum_{n<x} \exp(2i\pi\alpha\, s(P(n))) \ll x^{1-c\|(q-1)\alpha\|^2}.$$

Furthermore

$$q_0(d) \leqslant e^{67d^3(\log d)^2}.$$

With a more technical proof the assumptions that q is prime and that $(a_d, q) = 1$ can be relaxed.

The case $q < q_0(d)$, remains an open problem.

Primes in two bases

Drmota-Mauduit-Rivat, 2018:

If f is a strongly q_1-multiplicative function and g a strongly q_2-multiplicative function such that $(q_1, q_2) = 1$ and f is is not of the form $n \mapsto \mathsf{e}(kn/(q_1 - 1))$ with $k \in \mathbb{Z}$, then we have uniformly for $\vartheta \in \mathbb{R}$

$$\left| \sum_{n \leqslant x} \Lambda(n) f(n) g(n) \, \mathsf{e}(\vartheta n) \right| \ll x \exp\left(-c \frac{\log x}{\log \log x} \right)$$

for some positive constant c.

The proof uses a variant of Baker's theorem on linear forms due to Waldschmidt, which does not permit to save a power of x.

0110100110010110100101100110100110010110011010010110100110010110
1001011001101001011010011001011001101001100101101001011001101001
1001011001101001011010011001011001101001100101101001011001101001
0110100110010110100101100110100110010110011010010110100110010110
1001011001101001011010011001011001101001100101101001011001101001
0110100110010110100101100110100110010110011010010110100110010110
0110100110010110100101100110100110010110011010010110100110010110
1001011001101001011010011001011001101001100101101001011001101001

The Thue-Morse sequence

1001011001101001011010011001011001101001100101101001011001101001
0110100110010110100101100110100110010110011010010110100110010110
0110100110010110100101100110100110010110011010010110100110010110
1001011001101001011010011001011001101001100101101001011001101001
0110100110010110100101100110100110010110011010010110100110010110
1001011001101001011010011001011001101001100101101001011001101001
1001011001101001011010011001011001101001100101101001011001101001
0110100110010110100101100110100110010110011010010110100110010110

The Thue-Morse sequence

The Thue-Morse sequence $\mathbf{t} = (t_n)_{n\in\mathbb{N}}$ can be defined by induction:

$$t_0 = 0,\ t_{2n} = t_n,\ t_{2n+1} = 1 - t_n.$$

It is easy to check that

$$t_n \equiv \mathsf{s}_2(n) \bmod 2.$$

It is the fixed point of the substitution $t_0 = 0,\ 0 \mapsto 01,\ 1 \mapsto 10$:

0

01

0110

01101001

0110100110010110

01101001100101101001011001101001

0110100110010110100101100110100110010110100101101001100101101001011001101001

Symbolic complexity

Definition 3 *The symbolic complexity of a sequence* $\mathbf{u} \in \{0,1\}^{\mathbb{N}}$ *is the function* $p_{\mathbf{u}}$ *defined for any integer* $k \geqslant 1$ *by*

$$p_{\mathbf{u}}(k) = \text{card}\{(b_0, \ldots, b_{k-1}) \in \{0,1\}^k,\ \exists i \,/\, u_i = b_0, \ldots, u_{i+k-1} = b_{k-1}\}$$

*(*i.e. $p_{\mathbf{u}}(k)$ *is the number of distinct factors of length* k *that occur in the sequence* $\mathbf{u}$*).*

Let $(X(\mathbf{u}), T)$ be the dynamical system where T is the shift on $\{0,1\}^{\mathbb{N}}$ and $X(\mathbf{u})$ the closure of the orbit of $\mathbf{u}$ under the action of T (for the product topology of $\{0,1\}^{\mathbb{N}}$) .

The topological entropy of $(X(\mathbf{u}), T)$ can be shown to be equal to $\lim\limits_{k\to\infty} \dfrac{\log p_{\mathbf{u}}(k)}{k}$.

In that sense $p_{\mathbf{u}}$ constitutes a measure for the pseudorandomness of the sequence $\mathbf{u}$.

The Thue Morse sequence is very "simple"

$\mathbf{t}$ is not periodic and cubeless.

$\mathbf{t}$ is almost periodic: any subword occuring in $\mathbf{t}$ occurs infinitely often with bounded gaps.

The symbolic complexity of $\mathbf{t}$ is very low: there exist $c_1 > 0$ and $c_2 > 0$ such that, for all $k \geqslant 1$, $c_1 k \leqslant p_{\mathbf{t}}(k) \leqslant c_2 k$.

Zero topological entropy of the corresponding dynamical system: $h = \lim\limits_{k\to\infty} \dfrac{\log p_{\mathbf{t}}(k)}{k} = 0.$

For any fixed $(a, b) \in \mathbb{N}^2$ it is easy to check that the sequence $\mathbf{t_{a,b}} = (t_{an+b})_{n\in\mathbb{N}}$ is also obtained by a simple algorithm (it is generated by a finite 2-automaton).

It follows that its symbolic complexity is also sublinear: $p_{\mathbf{t_{a,b}}}(k) = O_a(k)$ and that any symbolic dynamical system $(X(\mathbf{t}_{a,b}), T)$ obtained by extracting a subsequence of $\mathbf{t}$ along arithmetic progressions still has zero topological entropy.

The Thue-Morse sequence along squares

Picking the values at square positions

$$0110100110010110100101100110100110010110011010010110100,$$

we get

$$0110110111110010111110110100110111111011110110100111$$
$$0001101100101111101110011111101001111101100101101111 0.$$

Moshe, 2007 (conjectured by **Allouche** and **Shallit** in 2003): the subword complexity of $(t_{n^2})_{n\geqslant 0}$ is $p_k^{(2)} = 2^k$, *i.e.* for $k \geqslant 1$, every word $b_1 \cdots b_k$ with $b_j \in \{0, 1\}$ appears in $(t_{n^2})_{n\geqslant 0}$.

Mauduit-Rivat, 2009: Both letters 0 and 1 have frequency $\frac{1}{2}$.

Question: what is the frequency of a given word ?

The Thue-Morse sequence along squares is normal

A sequence $\mathbf{u} \in \{0,1\}^{\mathbb{N}}$ is normal if, for any $k \in \mathbb{N}$ and any $(b_0, \ldots, b_{k-1}) \in \{0,1\}^k$:

$$\lim_{N\to\infty} \frac{1}{N} \text{card}\{i < N,\ u_i = b_0, \ldots, u_{i+k-1} = b_{k-1}\} = \frac{1}{2^k}.$$

Notion introduced by **Borel** in 1909. First explicit construction by **Champernowne** in 1933.

Only few explicit constructions are known.

Drmota-Mauduit-Rivat: The sequence $(t_{n^2})_{n\in\mathbb{N}}$ is normal.

This theorem provides a new method to construct normal numbers in a given base.

The real number $\alpha = \sum_{n=0}^{\infty} \frac{t_{n^2}}{2^n}$ is normal in base 2.

Ideas and tools

Approach to digital problem

1. Reduce the problem to an exponential sum,

2. apply several times the van der Corput inequality to remove the upper and lower digits,

3. separate into a discrete Fourier transform part and an analytic part,

4. handle the analytic part to see which Fourier estimates are needed,

5. obtain the corresponding Fourier estimates.

Introduction of exponential sums

For any $(b_0, \dots, b_{k-1}) \in \{0,1\}^k$ we have

$$\begin{aligned}
&\mathrm{card}\{n < N : (t_{n^2}, \dots, t_{(n+k-1)^2}) = (b_0, \dots, b_{k-1})\} \\
&= \sum_{n<N} \mathbf{1}_{[t_{n^2}=b_0]} \cdots \mathbf{1}_{[t_{(n+k-1)^2}=b_{k-1}]} \\
&= \sum_{n<N} \prod_{\ell=0}^{k-1} \left(\frac{1}{2} \sum_{\alpha_\ell=0}^{1} (-1)^{\alpha_\ell (\mathrm{s}((n+\ell)^2) - b_\ell)} \right) \\
&= \frac{1}{2^k} \sum_{\alpha_0,\dots,\alpha_{k-1}} (-1)^{\alpha_0 b_0 + \cdots + \alpha_{k-1} b_{k-1}} \sum_{n<N} \exp\left(i\pi \sum_{\ell=0}^{k-1} \alpha_\ell \, \mathrm{s}((n+\ell)^2) \right)
\end{aligned}$$

so that

$$\begin{aligned}
&\left| \mathrm{card}\{n < N : (t_{n^2}, \dots, t_{(n+k-1)^2}) = (b_0, \dots, b_{k-1})\} - \frac{N}{2^k} \right| \\
&\leqslant \frac{1}{2^k} \sum_{(\alpha_0,\dots,\alpha_{k-1}) \in \{0,1\}^k \setminus \{(0,\dots,0)\}} \left| \sum_{n<N} \exp\left(i\pi \sum_{\ell=0}^{k-1} \alpha_\ell \, \mathrm{s}((n+\ell)^2) \right) \right|.
\end{aligned}$$

Estimate of the exponential sum

We have to prove that for any $k \geqslant 1$ and $(\alpha_0, \ldots, \alpha_{k-1}) \in \{0,1\}^k \setminus \{(0,\ldots,0)\}$ there exists $\eta > 0$ for which

$$\sum_{n<N} \exp\left(i\pi \sum_{\ell=0}^{k-1} \alpha_\ell \, \mathsf{s}((n+\ell)^2)\right) \ll N^{1-\eta}.$$

The case $k = 1$ corresponds to our previous result on $s(n^2)$.

But: the method used there fails when $k \geqslant 2$ for many reasons.

Main difficulty: huge size and large number of variables.

We introduce a new approach to control the Fourier transform of correlations of any order.

A variant of van der Corput's inequality

For all complex numbers $z_1, \dots, z_L$ and integers $k \geqslant 1$, $R \geqslant 1$ we have

$$\left|\sum_{\ell=1}^{L} z_\ell\right|^2 \leqslant \frac{L+kR-k}{R}\left(\sum_{\ell=1}^{L}|z_\ell|^2 + 2\sum_{r=1}^{R-1}\left(1-\frac{r}{R}\right)\sum_{\ell=1}^{L-kr}\Re\left(z_{\ell+kr}\overline{z_\ell}\right)\right).$$

(for $k=1$ this is the classical van der Corput's inequality.)

Idea: Taking $z_\ell = \exp(2i\pi\varphi(\ell))$ for some function φ, since r is small we can take advantage a better control of the difference $\varphi(\ell+kr)-\varphi(\ell)$ instead of the more general $\varphi(\ell')-\varphi(\ell)$.

Here $\varphi(l) = \sum_j f((\ell+d_j)^2)$ with $f(n) = t_n$.

More generally in base q this is useful when f is q-additive.

With $k=1$, this will permits to remove the upper digits.
With $k = q^\mu$, this will permits to remove the lower digits.

Removing the upper digits

f is q-additive if for all $k \geqslant 0$ and all $(a_0, \ldots, a_k) \in \{0, \ldots, q-1\}^{k+1}$,

$$f(a_0 + a_1 q + \cdots + a_k q^k) = f(a_0) + f(a_1 q) + \cdots + f(a_k q^k).$$

Consider the difference $f(a+b) - f(a)$, with $b \asymp q^\beta$ much smaller that $a \asymp q^\alpha$. Example:

$$a = \overbrace{351167907809998065465234754734623368576435 65}^{\alpha},$$
$$b = \underbrace{396576345354568797095646467570}_{\beta},$$

In the sum $a+b$ the digits of index $\geqslant \beta$ may change only by carry propagation. The proportion of pairs (a,b) for which the carry propagation exceeds $\beta + \rho$ is likely to be $O(q^{-\rho})$. If so we can ignore these exceptional pairs and replace $f(a+b) - f(a)$ by $f_{\beta+\rho}(a+b) - f_{\beta+\rho}(a)$ where $f_{\beta+\rho}$ is the truncated f function which considers only the digits of index $< \beta + \rho$:

$$f_{\beta+\rho}(n) := f(n \bmod q^{\beta+\rho})$$

which is periodic of period $q^{\beta+\rho}$.

Removing the lower digits

Now if $a \asymp q^\alpha$, $c \asymp q^\gamma$ with $\gamma + \mu \leqslant \alpha$, consider the difference $f_{\beta+\rho}(a + q^\mu c) - f_{\beta+\rho}(a)$.
Example:

$$a = \overbrace{351167907809998065465234754734623368576435 65}^{\alpha},$$
$$q^\mu c = \underbrace{396576345354568797095646467571}_{\gamma}\underbrace{000000000000}_{\mu},$$

In the sum $a + q^\mu c$ the digits of index $< \mu$ are not modified. We have

$$f_\mu(a + q^\mu c) = f_\mu(a)$$

so

$$f_{\beta+\rho}(a + q^\mu c) - f_{\beta+\rho}(a) = (f_{\beta+\rho} - f_\mu)(a + q^\mu c) - (f_{\beta+\rho} - f_\mu)(a)$$

and $f_{\beta+\rho} - f_\mu$ depends only on the digits of index $\in \{\mu, \ldots, \beta + \rho - 1\}$.

If f is a more general digital function (not q-additive) these arguments need to be adapted.

Detection of digits

Let $\mathsf{r}_{\kappa_1,\kappa_2}(a)$ be the integer obtained using the digits of a of indexes $\kappa_1, \ldots, \kappa_2 - 1$. We have

$$\mathsf{r}_{\kappa_1,\kappa_2}(a) = u \iff \frac{a}{q^{\kappa_2}} \in \left[\frac{u}{q^{\kappa_2-\kappa_1}}, \frac{u+1}{q^{\kappa_2-\kappa_1}}\right) + \mathbb{Z}.$$

It remains to detect which points belong to an interval modulo 1.

For $0 \leqslant \alpha < 1$ let $\chi_\alpha(x) = \lfloor x \rfloor - \lfloor x - \alpha \rfloor$. For any integer $H \geqslant 1$ there exist real valued trigonometric polynomials such that for all $x \in \mathbb{R}$,

$$\left|\chi_\alpha(x) - A_{\alpha,H}(x)\right| \leqslant B_{\alpha,H}(x)$$

(using Vaaler's kernel derived from the Beurling-Selberg function) with

$$A_{\alpha,H}(x) = \sum_{|h|\leqslant H} a_{\alpha,H}(h)\exp(2i\pi hx)$$
$$B_{\alpha,H}(x) = \sum_{|h|\leqslant H} b_{\alpha,H}(h)\exp(2i\pi hx),$$

with $a_{\alpha,H}(0) = \alpha$, $\left|a_{\alpha,H}(h)\right| \leqslant \min\left(\alpha, \frac{1}{\pi|h|}\right)$, $\left|b_{\alpha,H}(h)\right| \leqslant \frac{1}{H+1}$.

Muldimensional approximation

How to detect points in a small d-dimensional box (modulo 1) ?

For $(\alpha_1, \ldots, \alpha_d) \in [0,1)^d$ and $(H_1, \ldots, H_d) \in \mathbb{N}^d$ with $H_1 \geqslant 1, \ldots, H_d \geqslant 1$, we have for all $(x_1, \ldots, x_d) \in \mathbb{R}^d$

$$\left| \prod_{j=1}^{d} \chi_{\alpha_j}(x_j) - \prod_{j=1}^{d} A_{\alpha_j, H_j}(x_j) \right| \leqslant \sum_{\emptyset \neq J \subseteq \{1,\ldots,d\}} \prod_{j \notin J} \chi_{\alpha_j}(x_j) \prod_{j \in J} B_{\alpha_j, H_j}(x_j)$$

where $A_{\alpha,H}(.)$ and $B_{\alpha,H}(.)$ are (**Vaaler**'s) real valued trigonometric polynomials.

This reminds to **Koksma**'s inequality

- $\alpha_1, \ldots, \alpha_d$ are small here,

- $\chi_{\alpha_1}, \ldots, \chi_{\alpha_d}$ on the right hand side can be used non trivially.

After two van der Corput's inequalities

We have removed the upper and lower digits and are led to consider

$$\sum_n \exp\Bigg(i\pi \sum_{\ell=0}^{k-1} \alpha_\ell \Big(f_{\mu,\lambda}((n+\ell)^2) - f_{\mu,\lambda}((n+r+\ell)^2) \\ - f_{\mu,\lambda}((n+s2^\mu+\ell)^2) + f_{\mu,\lambda}((n+s2^\mu+r+\ell)^2) \Big) \Bigg),$$

This is where the Fourier analysis becomes difficult !!!

Fourier analysis

We are now working modulo $2^{\lambda-\mu}$. Consider the Discrete Fourier Transform

$$F_{\mu,\lambda}(t) = \frac{1}{2^{\lambda-\mu}} \sum_{0 \leqslant u < 2^{\lambda-\mu}} \mathrm{e}\left(f_{\mu,\lambda}(2^\mu u) - \frac{ut}{2^{\lambda-\mu}} \right)$$

with

$$\mathrm{e}(t) = \exp(2i\pi t).$$

In our previous works, by Fourier inversion formula and exchanges of summations we could separate the "Fourier part" and the "exponential sum" part.

Then the properties of the Fourier transforms permitted to complete the proof.

It is not sufficient to prove the normality of the Thue-Morse sequence along squares.

Estimates of the discrete Fourier transform

It is enough to provide good estimates for the Fourier terms

$$G_\lambda^I(h,d) = \frac{1}{2^\lambda} \sum_{0 \leqslant u < 2^\lambda} \mathrm{e}\left(\frac{1}{2} \sum_{\ell=0}^{k-1} \alpha_\ell s_\lambda(u + \ell d + i_\ell) - \frac{h}{2^\lambda} \right),$$

where $s_\lambda(n) = \sum_{j<\lambda} n_j$ and $I = (i_0, \ldots, i_{k-1}) \in \mathbb{N}^k$.

These Fourier terms can be interpreted as coefficients of products of quite involved matrices and the second step is to study the combinatorial properties of the graphs associated to these matrices in order to provide these estimates.

For any $k \in \mathbb{N}$, we denote by $\mathcal{I}_k$ the set of integer vectors $I = (i_0, \ldots, i_{k-1})$ with $i_0 = 0$ and $i_{\ell-1} \leqslant i_\ell \leqslant i_{\ell-1} + 1$ for $1 \leqslant \ell \leqslant k-1$ (note that $\mathcal{I}_k$ consists of 2^{k-1} elements).

Proposition 1 *If $\alpha_0 + \ldots + \alpha_{k-1}$ is even, then there exists $\eta > 0$ such that for any $I \in \mathcal{I}_k$ we have*

$$\frac{1}{2^{\lambda'}} \sum_{0 \leqslant d < 2^{\lambda'}} |G_\lambda^I(h,d)|^2 \ll 2^{-\eta\lambda}$$

uniformly for all integers h, where $\frac{1}{2}\lambda \leqslant \lambda' \leqslant \lambda$.

Proposition 2 *If $\alpha_0 + \ldots + \alpha_{k-1}$ is odd, then there exists $\eta > 0$ such that for any $I \in \mathcal{I}_k$ we have*

$$\left|G_\lambda^I(h,d)\right| \ll 2^{-\eta L} \max_{J \in \mathcal{I}_k} \left|G_{\lambda-L}^J(h, \lfloor d/2^L \rfloor)\right|$$

uniformly for all non-negative integers h, d and L.

Estimate of the matrix product

The vector $\mathbf{G}_\lambda(h,d) = (G^I_\lambda(h,d))_{I\in\mathcal{I}_k}$ can be determined recursively:

$$\mathbf{G}_\lambda(h,d) = \frac{1}{2}\mathbf{M}^{\varepsilon_0(d)}\left(\mathrm{e}(-h/2^\lambda)\right)\mathbf{G}_{\lambda-1}(h,\lfloor d/2\rfloor),$$

where for any $\varepsilon\in\{0,1\}$, $\mathbf{M}^\varepsilon(z)$ is a $2^{k-1}\times 2^{k-1}$ matrix whose coefficients are of the form $a+bz$ with $(a,b)\in\{-1,0,1\}$.

It follows by induction that for any integer $n\geqslant 1$, we have

$$\mathbf{G}_\lambda(h,d) = \frac{1}{2^m}\mathbf{M}^{\varepsilon_0(d)\ldots\varepsilon_{m-1}(d)}\left(\mathrm{e}(-h/2^\lambda)\right)\mathbf{G}_{\lambda-m}(h,\lfloor d/2^m\rfloor),$$

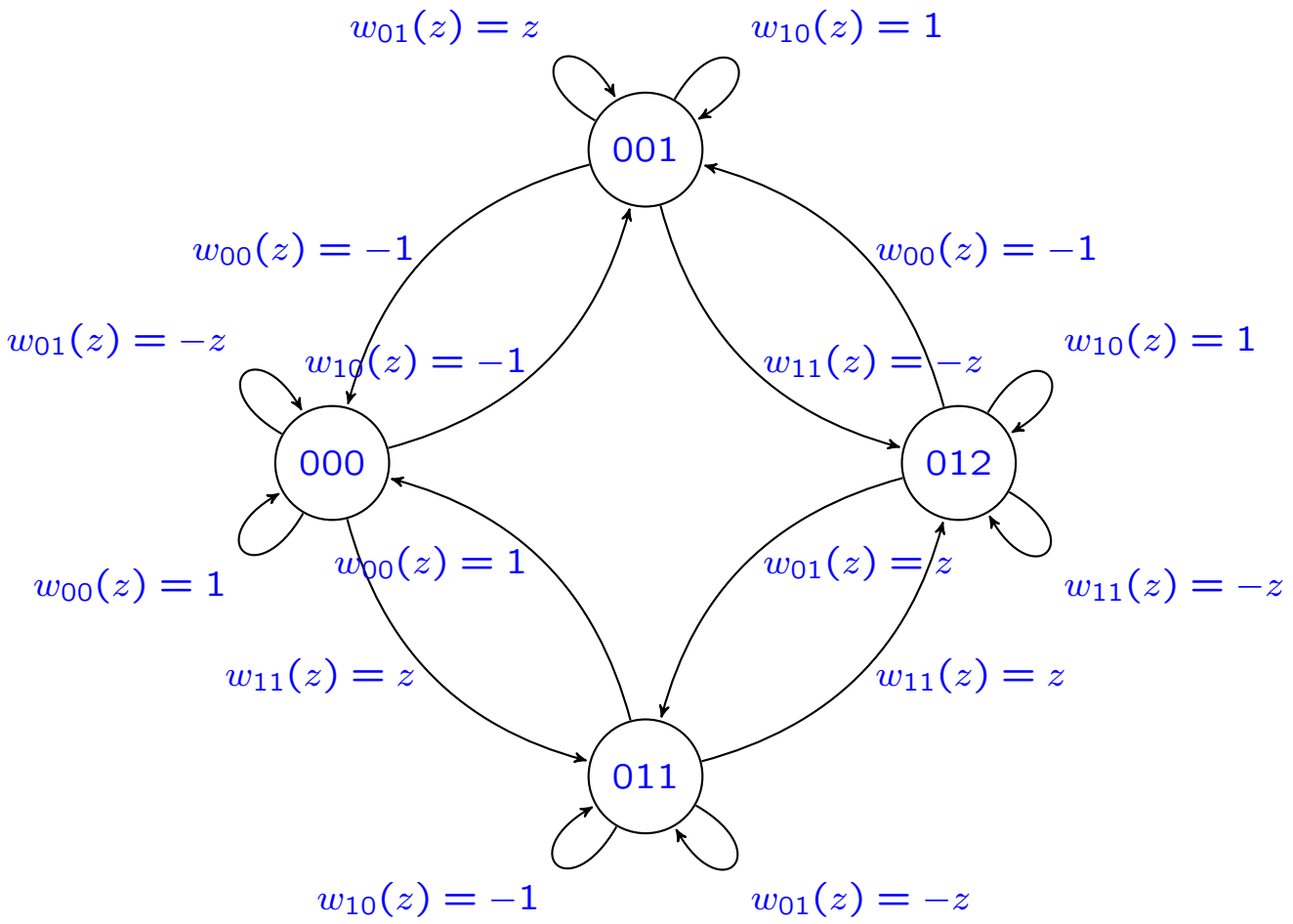
where for any $\mathbf{d}=(d_0,\ldots,d_{m-1})\in\{0,1\}^m$ we put

$$\mathbf{M}^{\mathbf{d}}(z) = \mathbf{M}^{d_0\ldots d_{m-1}}(z) = \mathbf{M}^{d_0}(z)\cdots\mathbf{M}^{d_{m-1}}(z^{2^{m-1}}).$$

For example when $k = 3$ we have

$$\mathbf{M}^0(z) = \begin{pmatrix} 1-z & 0 & 0 & 0 \\ -1 & z & 0 & 0 \\ 1 & 0 & -z & 0 \\ 0 & -1 & z & 0 \end{pmatrix}, \quad \mathbf{M}^1(z) = \begin{pmatrix} 0 & -1 & z & 0 \\ 0 & 1 & 0 & -z \\ 0 & 0 & -1 & z \\ 0 & 0 & 0 & 1-z \end{pmatrix}$$

and for any $\mathbf{d} = (d_0, \ldots, d_{m-1}) \in \{0,1\}^m$ we interpret the coefficients of the matrix $\mathbf{M}^{\mathbf{d}}(z)$ as coding of paths of length m with, for $j \in \{0, \ldots, m-1\}$, step j in the graph $\mathcal{G}(z^{2^j})$:


001
000
012
011
$w_{01}(z) = z$
$w_{10}(z) = 1$
$w_{00}(z) = -1$
$w_{00}(z) = -1$
$w_{01}(z) = -z$
$w_{10}(z) = -1$
$w_{11}(z) = -z$
$w_{10}(z) = 1$
$w_{00}(z) = 1$
$w_{00}(z) = 1$
$w_{01}(z) = z$
$w_{11}(z) = -z$
$w_{11}(z) = z$
$w_{11}(z) = z$
$w_{10}(z) = -1$
$w_{01}(z) = -z$

Open problems

For any polynomial of degree $\geqslant 3$ taking values in $\mathbb{N}$, is it true that $(t_{P(n)})_{n\in\mathbb{N}}$ is normal ?

Mauduit-Rivat, 2010: the frequencies of 0 and 1 in the sequence $(t_{p_n})_{n\in\mathbb{N}}$ are equal to $\frac{1}{2}$ (where $(p_n)_{n\in\mathbb{N}}$ denote the sequence of prime numbers).

For any non constant polynomial taking values in $\mathbb{N}$, is it true that $(t_{P(p_n)})_{n\in\mathbb{N}}$ is normal ?

Moreover it would be interesting to find some other almost periodic sequences $\mathbf{u}$ with the same property and also to understand this phenomenon from the dynamical system point of view.