On Gelfond's conjecture on the sum-of-digits function

Joël RIVAT

work in collaboration with

Christian MAUDUIT

Institut de Mathématiques de Luminy CNRS-UMR 6206,

Aix-Marseille Université, France.

rivat@iml.univ-mrs.fr

The sum of digits function

Let $q \in \mathbb{N}$ with $q \geqslant 2$. All $n \in \mathbb{N}$ can be written uniquely in basis q:

$$n = \sum_{k \geqslant 0} n_k q^k \quad \text{where } n_k \in \{0, \ldots, q-1\}.$$

The sum of digits function is defined by:

$$\mathsf{s}(n) = \sum_{k \geqslant 0} n_k.$$

The sum of digits function has many aspects that have been studied, for instance ergodicity, finite automata, dynamical systems, number theory.

Mahler introduced this function in the context of harmonic analysis:

Theorem A (Mahler, 1927) *For $q = 2$, the sequence*

$$\left(\frac{1}{N} \sum_{n<N} (-1)^{s(n)} (-1)^{s(n+k)} \right)_{N \geqslant 1}$$

converges for all $k \in \mathbb{N}$ and its limit is different from zero for infinitely many k's.

Gelfond's paper

The origin of our work is the following result of Gelfond:

Theorem B (Gelfond, 1968) *Let* $m \geqslant 2$, $(m, q-1) = 1$. *Then there exists* $\lambda < 1$ *such that for all* $d \in \mathbb{N}^*$, $a,\ r \in \mathbb{Z}$,

$$\sum_{\substack{n<N \\ n\equiv r \bmod d \\ \mathsf{s}(n)\equiv a \bmod m}} 1 = \frac{N}{md} + O(N^{\lambda}).$$

In the same paper Gelfond pose the following two problems:

Problem A (Gelfond, 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 $P(n) = n^2$].

Digits and primes - Historical background

Until recently, very little was known concerning the digits of prime numbers. We can mention a result of Sierpiński (1959), recently generalized by Wolke (2005) and then by Harman (2006), on prime numbers with some prescribed digits. Concerning Gelfond's question, no progress was made in its original form. Let us mention the two following variants:

Theorem C (Fouvry–Mauduit, 1996) *For $m \geqslant 2$ with $(m, q-1) = 1$, there exists $C(q,m) > 0$ such that for all $a \in \mathbb{Z}$ and $x > 0$,*

$$\sum_{\substack{n \leqslant x \\ n=p \text{ or } n=p_1p_2 \\ \mathsf{s}(n) \equiv a \bmod m}} 1 \geqslant \frac{C(q,m)}{\log\log x} \sum_{\substack{n \leqslant x \\ n=p \text{ or } n=p_1p_2}} 1.$$

Theorem D (Dartyge–Tenenbaum, 2005) *For $m \geqslant 2$ with $(m, q-1) = 1$ and $r \geqslant 2$, there exists $C(q,m,r) > 0$ such that for all $a \in \mathbb{Z}$ and $x > 0$,*

$$\sum_{\substack{n \leqslant x \\ n=p_1 \dots p_r \\ \mathsf{s}(n) \equiv a \bmod m}} 1 \geqslant \frac{C(q,m,r)}{\log\log x \log\log\log x} \sum_{\substack{n \leqslant x \\ n=p_1 \dots p_r}} 1.$$

Digits and primes - Results

Theorem 1 *For $\alpha \in \mathbb{R}$ such that $(q-1)\alpha \in \mathbb{R}\setminus\mathbb{Z}$, there exists $C(q,\alpha) > 0$ and $\sigma_q(\alpha) > 0$,*

$$\left|\sum_{p\leqslant x} \mathrm{e}(\alpha\, \mathrm{s}(p))\right| \leqslant C(q,\alpha)\; x^{1-\sigma_q(\alpha)}$$

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

Corollary 1 *The sequence $(\alpha\, \mathrm{s}(p_n))_{n\geqslant 1}$ 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).*

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

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

Sum over prime numbers

We want to estimate a sum of the form

$$\sum_{p \leqslant x} g(p)$$

where the function g detects the property under consideration. A classical process (Vinogradov, Vaughan, Heath-Brown) remains (using some more technical details), for some $0 < \beta_1 < 1/3$ and $1/2 < \beta_2 < 1$, to estimate uniformly the sums

$$S_I := \sum_{m \sim M} \left| \sum_{n \sim N} g(mn) \right| \quad \text{for } M \leqslant x^{\beta_1} \text{ (type I)}$$

where $MN = x$ (which implies that the "easy" sum over n is long) and for all complex numbers a_m, b_n with $|a_m| \leqslant 1$, $|b_n| \leqslant 1$ the sums

$$S_{II} := \sum_{m \sim M} \sum_{n \sim N} a_m b_n \, g(mn) \quad \text{for } x^{\beta_1} < M \leqslant x^{\beta_2} \text{ (type II)},$$

(which implies that both sums have a significant length).

Sums of type I

For the sums of type I we might expect that the knowledge of the function g permits to get a satisfactory estimate of the sum

$$\sum_{n\sim N} g(mn).$$

Indeed in our case where $g(n) = \mathsf{e}(\alpha\, \mathsf{s}(n))$ we were able to adapt successfully arguments from Fouvry and Mauduit (1996).

Sums of type II - Smoothing the sums

By Cauchy-Schwarz inequality:

$$|S_{II}|^2 \leqslant M \sum_{m\sim M} \left| \sum_{n\sim N} b_n \, \mathsf{e}(\alpha\, \mathsf{s}(mn)) \right|^2 .$$

Here, expanding the square and exchanging the summations, we would get a smooth sum over m, but also two free variables n_1 and n_2. However, we can get a useful control by using van der Corput's inequality: for $z_1, \dots, z_L \in \mathbb{C}$ and $R \in \mathbb{N}^*$ we have

$$\left| \sum_{1\leqslant \ell \leqslant L} z_\ell \right|^2 \leqslant \frac{L+R-1}{R} \sum_{|r|<R} \left(1 - \frac{|r|}{R}\right) \sum_{\substack{1\leqslant \ell \leqslant L \\ 1\leqslant \ell+r\leqslant L}} z_{\ell+r}\overline{z_\ell}.$$

The interest of this inequality is that now we have $n_1 = n + r$ and $n_2 = n$ so that the size of $n_1 - n_2 = r$ is under control.

Now in fact we can take $M = q^\mu$, $N = q^\nu$ and $R = q^\rho$ where μ, ν and ρ are integers such that $\rho/(\mu+\nu)$ is "very small". It remains to estimate non trivially

$$\sum_{q^{\nu-1}<n\leqslant q^\nu} b_{n+r}\,\overline{b_n} \sum_{q^{\mu-1}<m\leqslant q^\mu} \mathsf{e}(\alpha\, \mathsf{s}(m(n+r)) - \alpha\, \mathsf{s}(mn)).$$

Sums of type II - Truncated sum of digits function

We want to take advantage of the fact that in the difference $\mathsf{s}(m(n+r))-\mathsf{s}(mn)$, the product mr is much smaller that mn. In the example:

$$mn = \overbrace{351167907809998065465234754734623368576435 65}^{\mu+\nu},$$
$$mr = \underbrace{396576345354568797095646467570}_{\mu+\rho},$$

we see that in the sum $mn+mr$ the digits after index $\mu+\rho$ may change only by carry propagation.

Proving that the number of pairs (m,n) for which the carry propagation exceeds

$$\lambda := \mu + 2\rho$$

is bounded by $O(q^{\mu+\nu-\rho})$, we can ignore them and replace $\mathsf{s}(m(n+r))-\mathsf{s}(mn)$ by $\mathsf{s}_\lambda(m(n+r))-\mathsf{s}_\lambda(mn)$ where s_λ is the truncated sum of digits function

$$\mathsf{s}_\lambda(n) := \sum_{k<\lambda} n_k,$$

which is periodic of period q^λ.

Sums of type II - Fourier analysis

The periodicity of s_λ enables us to write

$$\sum_{q^{\mu-1}<m\leqslant q^\mu} \mathsf{e}(\alpha\,\mathsf{s}_\lambda(m(n+r)) - \alpha\,\mathsf{s}_\lambda(mn))$$
$$= \sum_{0\leqslant u<q^\lambda} \sum_{0\leqslant v<q^\lambda} \mathsf{e}(\alpha\,\mathsf{s}_\lambda(u) - \alpha\,\mathsf{s}_\lambda(v)) \sum_{\substack{q^{\mu-1}<m\leqslant q^\mu \\ m(n+r)\equiv u \bmod q^\lambda \\ mn\equiv v \bmod q^\lambda}} 1.$$

The orthogonality formula

$$\frac{1}{q^\lambda} \sum_{0\leqslant h<q^\lambda} \mathsf{e}\left(\frac{h\ell}{q^\lambda}\right) = \begin{cases} 1 & \text{if } \ell \equiv 0 \bmod q^\lambda, \\ 0 & \text{if } \ell \not\equiv 0 \bmod q^\lambda, \end{cases}$$

leads us to introduce the discrete Fourier transform of $u \mapsto \mathsf{e}(\alpha\,\mathsf{s}_\lambda(u))$:

$$F_\lambda(h) = q^{-\lambda} \sum_{0\leqslant u<q^\lambda} \mathsf{e}\left(\alpha\,\mathsf{s}_\lambda(u) - \frac{hu}{q^\lambda}\right),$$

Sums of type II - Exponential sums

Summing over n and taking absolute values we must show that the quantity

$$\sum_{0\leqslant h<q^\lambda}\sum_{0\leqslant k<q^\lambda}\left|F_\lambda(h)\,\overline{F_\lambda(-k)}\right|\sum_{q^{\nu-1}<n\leqslant q^\nu}\left|\sum_{q^{\mu-1}<m\leqslant q^\mu}\mathsf{e}\left(\frac{hm(n+r)+kmn}{q^\lambda}\right)\right|$$

is estimated by $O(q^{\mu+\nu-\rho})$.

Here we observe that the summations over m (geometric sum !) and n can be handled by classical arguments from analytic number theory, while we hope that the digital structure hidden in F_λ will produce a huge saving.

For instance for $q=2$ we have

$$|F_\lambda(h)|=\prod_{i=1}^{\lambda}\left|\cos\pi\left(\alpha-\frac{h}{2^i}\right)\right|.$$

Heuristic end of the proof

On average, for fixed (h,k), the geometric sum over m is small so that the sum over n should be $O(q^{\nu+\varepsilon})$. Hence after many technical steps to handle the exceptions, we will need to get the crucial upper bound

$$\sum_{0\leqslant h<q^{\lambda}} |F_{\lambda}(h)| = O\left(q^{\eta_q \lambda}\right) \quad \text{with } \eta_q < 1/2,$$

which means that we need an upper bound sharper than the square root of the trivial estimate.

Indeed suppose this has been done, then we get

$$\sum_h \sum_k \sum_n \sum_m \cdots = O(q^{2\eta_q\lambda+\nu+\varepsilon}),$$

and since $\lambda = \mu + 2\rho$, we have

$$2\eta_q\lambda + \nu + \varepsilon \leqslant \mu + \nu - \rho$$

for μ, ν large enough.

Digits and squares - Historical background

Until recently, very little was known concerning the digits of squares. We can mention a result of Davenport and Erdős (1952), later improved by Peter (2002).

Theorem E (Consequence of Davenport-Erdős, 1952)

$$\sum_{n \leqslant x} s_q(n^2) \sim (q-1)\, x\, \frac{\log x}{\log q} \quad (x \to +\infty).$$

Erdős considered that passing from such a mean result to a local result like the question of Gelfond *"hopelessly difficult"*.

Concerning Gelfond's question, Dartyge and Tenenbaum (2005) obtained a positive density:

Theorem F (Dartyge-Tenenbaum, 2005) *For $m \geqslant 2$ such that $(m, q-1) = 1$, there exists $C(q,m) > 0$ and $x_0(q,m) \geqslant 1$ such that for all $a \in \mathbb{Z}$ and $x \geqslant x_0(q,m)$, we have*

$$\sum_{\substack{n \leqslant x \\ s(n^2) \equiv a \bmod m}} 1 \geqslant C(q,m)\, x.$$

Digits and squares - Results

Theorem 2 *For $\alpha \in \mathbb{R}$ such that $(q-1)\alpha \in \mathbb{R}\setminus\mathbb{Z}$, there exists $C(q,\alpha) > 0$ and $\sigma_q(\alpha) > 0$,*

$$\left|\sum_{n\leqslant x} \mathsf{e}(\alpha\, \mathsf{s}(n^2))\right| \leqslant C(q,\alpha)\; x^{1-\sigma_q(\alpha)}.$$

Corollary 3 *The sequence $(\alpha\, \mathsf{s}(n^2))_{n\geqslant 1}$ is equidistributed modulo 1 if and only if $\alpha \in \mathbb{R}\setminus\mathbb{Q}$.*

Corollary 4 *For $m \geqslant 2$ such that $(m, q-1) = 1$ and $a \in \mathbb{Z}$,*

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

Truncated sum of digits function

By van der Corput's inequality

$$\left|\sum_{n\sim q^\nu} \mathsf{e}(\alpha\,\mathsf{s}(n^2))\right|^2 \leqslant q^{\nu-\rho} \sum_{|r|<q^\rho} \left|\sum_{n\sim q^\nu} \mathsf{e}(\alpha\,\mathsf{s}((n+r)^2)-\alpha\,\mathsf{s}(n^2))\right|.$$

We want to take advantage of the fact that in the difference $\mathsf{s}((n+r)^2)-\mathsf{s}(n^2)$, the term $2nr+r^2$ is much smaller that n^2. In the example:

$$n^2 = \overbrace{97546107376558473380082563433318536692575814662 4}^{2\nu},$$
$$2nr+r^2 = \underbrace{39262222544221525318072970 8185}_{\nu+\rho+1},$$

we see that in the sum $n^2+2nr+r^2$ the digits after index $\nu+\rho+1$ may change only by carry propagation. The number of integers n for which the carry propagation exceeds $\lambda := \nu+2\rho+1$ is bounded by $O(q^{\nu-\rho})$, we can ignore them and replace the difference $\mathsf{s}((n+r)^2)-\mathsf{s}(n^2)$ by $\mathsf{s}_\lambda((n+r)^2)-\mathsf{s}_\lambda(n^2)$ where s_λ is the truncated sum of digits function $\mathsf{s}_\lambda(n) := \sum_{k<\lambda} n_k$,

which is periodic of period q^λ.

Heuristic considerations

It remains to estimate

$$\sum_{n\sim q^\nu} \mathsf{e}(\alpha\, \mathsf{s}_\lambda((n+r)^2) - \alpha\, \mathsf{s}_\lambda(n^2)).$$

The natural approach as for the primes would be to introduce again the discrete Fourier transform F_λ and show the estimate

$$\sum_{0\leqslant h<q^\lambda} \sum_{0\leqslant k<q^\lambda} \left|F_\lambda(h)\ \overline{F_\lambda(-k)}\right| \left|\sum_{q^{\nu-1}<n\leqslant q^\nu} \mathsf{e}\left(\frac{h(n+r)^2+kn^2}{q^\lambda}\right)\right| = O(q^{\nu-\rho}).$$

Remember that $\displaystyle\sum_{0\leqslant h<q^\lambda} |F_\lambda(h)| = O(q^{\eta_q\lambda})$ and the quadratic Gauss sum is usually $O(\lambda\, q^{\lambda/2})$.

From $2\eta_q + \frac{1}{2} < 1$ we realize that we need $\eta_q < 1/4$, which is not true for q small.

Conclusion: this method would give at most a *partial result*.

Variant of van der Corput

Lemma 1 *For $z_1, \dots, z_L \in \mathbb{C}$ and integers $k \geqslant 1$, $S \geqslant 1$ we have*

$$\left| \sum_{1 \leqslant \ell \leqslant L} z_\ell \right|^2 \leqslant \frac{L + (S-1)k}{S} \sum_{|s| < S} \left(1 - \frac{|s|}{S}\right) \sum_{\substack{1 \leqslant \ell \leqslant L \\ 1 \leqslant \ell + sk \leqslant L}} z_{\ell + sk} \overline{z_\ell}.$$

We apply this inequality with $S = q^{2\rho}$ and $k = q^\mu$ with $\mu = \nu - 2\rho - 1$. It remains to estimate non trivially

$$\sum_{n \sim q^\nu} \mathsf{e}\left(\alpha\left(\mathsf{s}_\lambda((n + r + sq^\mu)^2) - \mathsf{s}_\lambda((n+r)^2) - \mathsf{s}_\lambda((n+sq^\mu)^2) + \mathsf{s}_\lambda(n^2)\right)\right).$$

When we add sq^μ the digits of index below μ are not modified. Hence we may replace $s_\lambda(n)$ by

$$s_{\mu,\lambda}(n) = s_\lambda(n) - s_\mu(n) = \sum_{\mu \leqslant j < \lambda} n_j$$

In conclusion we have eliminated all the digits except a small interval of them: those between $\mu = \nu - 2\rho - 1$ and $\lambda = \nu + 2\rho + 1$.

Fourier analysis

We introduce the discrete Fourier transform of $u \mapsto \mathsf{e}(\alpha\, \mathsf{s}_{\mu,\lambda}(u))$:

$$F_{\mu,\lambda}(h) = q^{-\lambda} \sum_{0 \leqslant u < q^\lambda} \mathsf{e}\left(\alpha\, \mathsf{s}_{\mu,\lambda}(u) - \frac{hu}{q^\lambda}\right).$$

We need to estimate

$$\sum_{0 \leqslant h_1,h_2,h_3,h_4 < q^\lambda} \left|F_{\mu,\lambda}(h_1)F_{\mu,\lambda}(-h_2)F_{\mu,\lambda}(h_3)F_{\mu,\lambda}(-h_4)\right|\ G(h_1,h_2,h_3,h_4)$$

where $G(h_1,h_2,h_3,h_4)$ is the quadratic Gauss sum

$$G(h_1,h_2,h_3,h_4) = \sum_{n \in I(\nu,s,\mu)} \mathsf{e}\left(\frac{h_1(n+r+sq^\mu)^2 + h_2(n+r)^2 + h_3(n+sq^\mu)^2 + h_4 n^2}{q^\lambda}\right)$$

Heuristic end of the proof

On average, for fixed (h_1, h_2, h_3, h_4), the quadratic Gauss sum should be $O(\lambda q^{\frac{\lambda}{2}})$. Considering the crude estimate

$$\sum_{0 \leqslant h < q^\lambda} \left| F_{\mu,\lambda}(h) \right| = O(\mu\, q^{\lambda-\mu}) = O(\mu q^{4\rho})$$

we observe that

$$\frac{\lambda}{2} + 4\rho = \frac{\nu + 2\rho + 1}{2} + 4\rho < \nu - 2\rho$$

for ρ small enough, so there is some hope.

However, big technical problems occur from degenerate cases (*e.g.* $h_1 + h_2 + h_3 + h_4 = 0$) for which the quadratic Gauss sum are huge, but we have an additional condition.

The divisor of q also play a rôle.

Fortunately we could handle all these technical problems, and prove the result.