PRIMES WITH RESTRICTED DIGITS

VIVIAN KUPERBERG

1. Introduction and Problem Statement

This seminar talk discusses James Maynard's result [1] (which he then improved to [2]) on counting primes with restricted digits.

Let's start with the question setup; we will then get into the circle method, how it works in this case, and why this case is special. We will fix an integer q, our "base"; think of q as being very large. We will also fix $a_0 \in \{0, \dots, q-1\}$.

Definition 1.1. Let $\mathcal{A} := \left\{\sum_{i \geq 0} n_i q^i : n_i \in \{0, \dots, q-1\} \setminus \{a_0\}\right\}$ be the set of numbers with no a_0 in their base-q expansion.

Question 1.2. Does $\mathcal{A}$ contain infinitely many primes?

Remark 1.3.

$$|\mathcal{A} \cap [1, X]| \approx (q-1)^{\log X / \log q} = X^{\log(q-1)/\log q}$$

So, this set is *very* thin! For some more natural thin sets, like those involving short intervals or arithmetic progressions, information about primes in them would tell us about zero-free regions of L-functions. In particular, for many thin sets, finding primes is hard.

Theorem 1.4 (Maynard)**.** *Let $q > 2000000$. For any constant $R > 0$,*

$$\sum_{n < q^k} \Lambda(n) \mathbb{1}_{\mathcal{A}}(n) = \kappa_q(a_0)(q-1)^k + O_R\left(\frac{(q-1)^k}{(\log q^k)^R}\right),$$

with

$$\kappa_q(a_0) = \begin{cases} \frac{q}{q-1} & \text{if } (a_0, q) \neq 1 \\ \frac{q}{q-1}\frac{\phi(q)-1}{\phi(q)} & \text{if } (a_0, q) = 1 \end{cases}$$

We'll prove this using the circle method, since the Fourier transform of the indicator function $\mathbb{1}_{\mathcal{A}}$ has particularly nice properties.

Specifically, let $\widehat{F}_{q^k}$ be the Fourier transform of $\mathbb{1}_{\mathcal{A} \cap [1, q^k]}$. Then

$$\mathbb{1}_{\mathcal{A}}(n) = \frac{1}{q^k} \sum_{0 \leq a < q^k} \widehat{F}_{q^k}\left(\frac{a}{q^k}\right) e\left(\frac{-an}{q^k}\right)$$

$$\Rightarrow \sum_{n \leq q^k} \mathbb{1}_{\mathcal{A}}(n)\Lambda(n) = \frac{1}{q^k} \sum_{0 \leq a < q^k} \widehat{F}_{q^k}\left(\frac{a}{q^k}\right) S_{q^k}\left(\frac{-a}{q^k}\right),$$

with

$$S_{q^k}(\theta) = \sum_{n \leq q^k} \Lambda(n) e(n\theta)$$

Both $\widehat{F}_{q^k}$ and S_{q^k} are larger when $\frac{a}{q^k}$ are close to some $\frac{\ell}{d}$, with d small. So, we can split the problem into the major arcs, i.e. terms when $\frac{a}{q^k} \approx \frac{\ell}{d}$ with d small, and the minor arcs, i.e. terms when $\frac{a}{q^k} \approx \frac{\ell}{d}$ for d big. We'll bound the contribution of $\widehat{F}_{q^k} S_{q^k}$ on minor arcs to get something smaller than the contribution from the major arcs.

2. Interlude: The Circle Method, and Why You Might Not Expect It To Work Here

Say we have some sequence a_n whose asymptotic size we would like to know, or bound below. By Fourier inversion,

$$a_n = \int_0^1 F(\theta) e(-n\theta) \mathrm{d}\theta,$$

with $F(\theta) = \sum a_k e(k\theta)$. We then divide the circle up into major arcs (θ near $\frac{\ell}{d}$ for d small) and minor arcs (everything else).

Let's consider the example of the ternary Goldbach problem. Here we say a_n is a weighted count of the number of ways to write n as a sum of three primes, so $a_n = \sum_{k_1+k_2+k_3=n} \Lambda(k_1)\Lambda(k_2)\Lambda(k_3)$, and

$$a_n = \int_0^1 \left(\sum_{k \leq n} \Lambda(k) e(k\theta) \right)^3 e(-n\theta) \mathrm{d}\theta,$$

where we'll say here that $F(\theta) = \sum_{k \leq n} \Lambda(k) e(k\theta)$, which in this case is the cube root of the Fourier transform.

By the PNT, $F(0) = F(1) = n + o(n)$. At the same time, by Parseval and partial summation,

$$\int_0^1 |F(\theta)|^2 \mathrm{d}\theta = \sum_{k \leq n} \Lambda(k)^2 = n \log n + o(n \log n).$$

So, the average size of F is about the *square root* of the sum length, a phenomenon known as *square root cancellation*.

The major arcs here are small isolated intervals around points $\frac{\ell}{d}$ with d small, i.e. points where $F(\theta)$ is big. So, their contribution is something like

$$\sum_{d \leq \log n} \sum_{\substack{a=1 \\ (a,d)=1}}^{d} F\left(\frac{a}{d}\right)^3 e\left(-n\frac{a}{d}\right) \frac{2 \log n}{n} \approx n^2 \mathfrak{S}.$$

The minor arcs are everything else, where f is small. Vinogradov's crucial contribution here was the result that $\max_{\theta \in \mathfrak{m}} F(\theta) \ll \frac{n}{\log^D n}$, getting a log power saving. Then

$$\int_{\mathfrak{m}} |F(\theta)|^3 \mathrm{d}\theta \ll \frac{n}{\log^D n} \int_0^1 |F(\theta)|^2 \mathrm{d}\theta \ll \frac{n^2}{\log^{D-1} n}.$$

The minor arc contribution is small, so it provides an error term. Let's contrast this situation with the situation of binary Goldbach, where instead we want to consider

numbers a_n that are a sum of two primes. Here we have

$$a_n = \int_0^1 F(\theta)^2 \mathrm{d}\theta,$$

for the same $F(\theta)$ as above. In this case by the same argument, the major arc contribution is of size $\approx n\mathfrak{S}$. But $\int_0^1 |F|^2 \mathrm{d}\theta \approx n \log n$, which is bigger! Square root cancellation is not enough in this case, so the minor arcs don't have a smaller contribution.

This problem as well is a binary problem, in the sense that our integral has two terms. If we only saw square root cancellation in the minor arcs, again that would not be enough to help us. However, the special structure of our problem allows us to use an L^1 bound to get enough savings.

3. Minor arcs

Let's return to our sum

$$\sum_{n \leq q^k} \mathbb{1}_{\mathcal{A}}(n)\Lambda(n) = \frac{1}{q^k} \sum_{0 \leq a < q^k} \widehat{F}_{q^k}\left(\frac{a}{q^k}\right) S_{q^k}\left(\frac{-a}{q^k}\right)$$

with

$$S_{q^k}(\theta) = \sum_{n \leq q^k} \Lambda(n) e(n\theta)$$

Let's get some bounds on $\widehat{F}_{q^k}$ and S_{q^k}.

Lemma 3.1 (L^1 bound)**.** *There exists a constant $C_q \in [1/\log q, 1 + 3/\log q]$ such that*

$$\sup_{\theta \in \mathbb{R}} \sum_{0 \leq a < q^k} \left| \widehat{F}_{q^k}\left(\theta + \frac{a}{q^k}\right) \right| \ll (C_q q \log q)^k.$$

Proof outline. Let's expand our definition of $\widehat{F}_{q^k}$, with $n = \sum_{i=0}^{k-1} n_i q^i$ the base-q expansion of n. We have

$$\widehat{F}_{q^k}(t) = \sum_{n < q^k} \mathbb{1}_{\mathcal{A}}(n) e(tn) = \prod_{i=0}^{k-1} \left(\sum_{n_i=0}^{q-1} \mathbb{1}_{\mathcal{A}}(n_i) e(n_i q^i t) \right),$$

with the inner sum a sum over all values in $\{0, \ldots, q-1\} \setminus \{a_0\}$. The inner sum is a geometric series, so we can bound it by

$$\left| \frac{e(q^{i+1}t) - 1}{e(q^i t) - 1} - e(a_0 q^i t) \right| \leq \min\left(q, 1 + \frac{1}{2||q^i t||}\right).$$

Let's expand $t \in [0,1)$ in base q, to get $t = \sum_{i=1}^{k} t_i q^{-i} + \varepsilon$, with $\varepsilon \in [0, 1/q^k)$. Then $||q_i t||^{-1} = ||t_{i+1}/q + \varepsilon_i||^{-1}$, with $\varepsilon_i \in [0, 1/q)$. Using this, we can bound

$$\begin{aligned}\sup_{\theta \in \mathbb{R}} \sum_{0 \leq a < q^k} \left| \widehat{F}_{q^k}\left(\theta + \frac{a}{q^k}\right) \right| &\ll \sum_{t_1, \ldots, t_k < q} \prod_{i=1}^{k} \min\left(q, 1 + \max\left(\frac{q}{2t_i}, \frac{q}{2(q-1-t_i)}\right)\right) \\ &\ll (3q + q\log q)^k.\end{aligned}$$

This completes the proof. □

This is very good! We can also get the following bound using both this and the large sieve.

Lemma 3.2 (Hybrid estimate)**.** *Let $B, D \gg 1$. Then*

$$\sum_{D\le d\le 2D}\sum_{\substack{\ell<d\\(\ell,d)=1}}\sum_{\substack{|\eta|<B\\ q^k\ell/d+\eta\in\mathbb{Z}}}\left|\widehat{F}_{q^k}\left(\frac{\ell}{d}+\frac{\eta}{q^k}\right)\right| \ll (q-1)^k(D^2B)^{\alpha_q}+D^2B(C_q\log q)^k,$$

where C_q is the same constant from before, and

$$\alpha_q=\frac{\log(C_q\frac{q}{q-1}\log q)}{\log q}.$$

Remark 3.3. Note that we had $C_q\in[1/\log q, 1+3/\log q]$, so that

$$\alpha_q\le\frac{\log\left(\frac{q}{q-1}\log q+\frac{3q}{q-1}\right)}{\log q}.$$

As $q\to\infty$, this approaches 0, so eventually it is as small as we like. In particular, $\alpha_q<1/5$ when $q>2000000$.

If $B>q^k$ it follows immediately from the L^1 bound. If $B<q^k$, the strategy is to decompose the k digits as k_1 digits with $B\approx q^{k_1}$, and then k_2 additional digits. The product decomposition for $\widehat{F}_{q^k}$ means that we can very nicely separately apply bounds for the first k_1 digits and the last k_2 digits; the first of these are addressed again by our L^1 bound, and the remainder using a large sieve bound.

Now let's consider a bound for the other half of this sum, namely S_{q^k}. For this, we have a more standard exponential sum bound:

Lemma 3.4. *Let $\alpha=\frac{a}{d}+\beta$ with $(a,d)=1$, $|\beta|<1/d^2$. Then*

$$S_x(\alpha)=\sum_{n<x}\Lambda(n)e(n\alpha)\ll\left(x^{4/5}+\frac{x^{1/2}}{|d\beta|^{1/2}}+x|d\beta|^{1/2}\right)(\log x)^4.$$

We can glue these together to show that when α is far away from a rational with small denominator, $\widehat{F}_{q^k}(\alpha)S_{q^k}(-\alpha)$ is typically small.

Proposition 3.5. *Let $1\ll B\ll q^k/D_0D$ and $1\ll D\ll D_0\ll q^{k/2}$. Then*

$$\sum_{D\le d\le 2D}\sum_{\substack{0<\ell<d\\(\ell,d)=1}}\sum_{\substack{B\le|\eta|\le 2B\\ q^k\ell/d+\eta\in\mathbb{Z}}}\left|\widehat{F}_{q^k}\left(\frac{\ell}{d}+\frac{\eta}{q^k}\right)S_{q^k}\left(-\frac{\ell}{d}-\frac{\eta}{q^k}\right)\right|$$
$$\ll k^4(q-1)^kq^k\left(\frac{1}{(DB)^{1/5-\alpha_q}}+\frac{q^{k\alpha_q}}{D_0^{1/2}}\right),$$

and

$$\sum_{D\le d\le 2D}\sum_{\substack{0<\ell<d\\(\ell,d)=1}}\sum_{\substack{|\eta|\ll 1\\ q^k\ell/d+\eta\in\mathbb{Z}}}\left|\widehat{F}_{q^k}\left(\frac{\ell}{d}+\frac{\eta}{q^k}\right)S_{q^k}\left(-\frac{\ell}{d}-\frac{\eta}{q^k}\right)\right|$$
$$\ll k^4(q-1)^k q^k\left(\frac{1}{D^{1/5-\alpha_q}}+\frac{D_0^{1/2+2\alpha_q}}{q^{k/2}}\right).$$

Proof. We'll show just the first statement. We have from our assorted lemmas that if $D^2B\ll q^k$,

$$\sum_{D\le d\le 2D}\sum_{\substack{\ell<d\\(\ell,d)=1}}\sum_{\substack{|\eta|<B\\ q^k\ell/d+\eta\in\mathbb{Z}}}\left|\widehat{F}_{q^k}\left(\frac{\ell}{d}+\frac{\eta}{q^k}\right)\right|\ll (q-1)^k(D^2B)^{\alpha_q}$$

and

$$\sup_{\substack{D\le d\le 2D\\(\ell,d)=1\\B\le|\eta|\le 2B}}\left|S_{q_k}\left(-\frac{\ell}{d}-\frac{\eta}{q^k}\right)\right|\ll\left(q^{4k/5}+\frac{(DB)^{1/2}}{q^{k/2}}+\frac{q^k}{(DB)^{1/2}}\right)(k\log q)^4.$$

Combining these gives that what we want is

$$\ll k^4q^k(q-1)^k\left(\frac{(D^2B)^{\alpha_q}}{q^{k/5}}+\frac{(D^2B)^{\alpha_q}}{(DB)^{1/2}}+\frac{(DB)^{1/2}(D^2B)^{\alpha_q}}{q^{k/2}}\right)$$
$$\ll k^4q^k(q-1)^k\left((D^2B)^{\alpha_q-1/5}+(D^2B)^{\alpha_q-1/4}+\frac{q^{k\alpha_q}}{D_0^{1/2}}\right),$$

using for the second line that $D^2B<q^k$ and $DB<q^k/D_0$. □

Let's see what bound this gives us. In particular, our Fourier expansion gives us

$$\sum_{n<q^k}\Lambda(n)\mathbb{1}_{\mathcal{A}}(n)=\frac{1}{q^k}\sum_{0\le a<q^k}\widehat{F}_{q^k}\left(\frac{a}{q^k}\right)S_{q^k}\left(\frac{-a}{q^k}\right).$$

We fix D_0 and D, to be determined later. By Dirichlet's approximation theorem, there exist $(\ell,d)=1$ with $d<D$ and $|\beta|<1/DD_0$ such that $\frac{a}{q^k}=\frac{\ell}{d}+\beta$. We'll define minor arcs as the values where $\max(d,q^k|\beta|)\ge(\log q^k)^R$. These terms are

$$\ll O_b\left(\frac{1}{q^k}\left(k^4q^k(q-1)^k\left(\frac{1}{(\log q^k)^{R(1/5-\alpha_q)}}+\frac{kD_0^{1/2+2\alpha_q}}{q^{k/2}}+\frac{kq^{k\alpha_q}}{D_0^{1/2}}\right)\right)\right).$$

If we take $D_0=q^{k/2}$, and $q>2000000$ so that $\alpha_q<1/5$, and choose R such that $R>(R'+5)/(1/5-\alpha_q)$, then this is

$$\ll O_{R'}((q-1)^k(\log q^k)^{-R'}),$$

which is the error term we're aiming for.

4. Major Arcs

Lemma 4.1. *Let $R > 0$. For $D, B < (\log q^k)^R$, we have*

$$\frac{1}{q^k} \sum_{\substack{d<D \\ p|d \Rightarrow p|q}} \sum_{\substack{0\le \ell<d \\ (\ell,d)=1}} \sum_{|b|<B} \widehat{F}_{q^k}\left(\frac{\ell}{d}+\frac{b}{q^k}\right) S_{q^k}\left(-\frac{\ell}{d}-\frac{b}{q^k}\right)$$
$$= \kappa_q(a_0)(q-1)^k + O_R\left(\frac{(q-1)^k}{(\log q^k)^R}\right),$$

with

$$\kappa_q(a_0) = \begin{cases} \frac{q}{q-1} & \text{if } (a_0,q) \neq 1 \\ \frac{q}{q-1}\frac{\phi(q)-1}{\phi(q)} & \text{if } (a_0,q) = 1. \end{cases}$$

Proof. First, we can discard all terms with $b \neq 0$. If $b \neq 0$, then by the prime number theorem in arithmetic progressions and partial summation,

$$S_{q^k}\left(-\frac{\ell}{d}-\frac{b}{q^k}\right) = \sum_{n\le q^k} \Lambda(n) e\left(-\frac{n\ell}{d}-\frac{nb}{q^k}\right) \ll_R \frac{q^k}{(\log q^k)^{4R}},$$

so in total these terms contribute

$$\frac{(\log q^k)^{3R}}{q^k} \sup_{0<a<q^k} \left|\widehat{F}_{q^k}\left(\frac{a}{q^k}\right)\right| \frac{q^k}{(\log q^k)^{4R}} \ll \frac{(q-1)^k}{(\log q^k)^R}.$$

So let's look at the terms with $b = 0$. Again using the prime number theorem in arithmetic progressions,

$$S_{q^k}\left(\frac{-\ell}{d}\right) = \frac{q^k}{\phi(d)} \sum_{\substack{0<c<d \\ (c,d)=1}} e\left(\frac{-\ell c}{d}\right) + O_R\left(\frac{q^k}{(\log q^k)^{4R}}\right) = \frac{\mu(d)q^k}{\phi(d)} + O_R\left(\frac{q^k}{(\log q^k)^{4R}}\right)$$

Since $d|q^n$ by our assumption on primes, either $d|q$ or d is not square-free. So we can restrict to $d|q$, and take ℓ' so that $\ell'/q = \ell/d$. These terms then contribute

$$\frac{1}{q^k} \sum_{0\le \ell'<q} \widehat{F}_{q^k}\left(\frac{\ell'}{q}\right) S_{q^k}\left(\frac{-\ell'}{q}\right) = \frac{1}{q^{k-1}} \sum_{\substack{n,m<q^k \\ n\equiv m \pmod q}} \Lambda(n)\mathbb{1}_{\mathcal{A}}(m)$$
$$= \frac{q}{\phi(q)} \sum_{\substack{1<a<q \\ (a,q)=1}} \sum_{\substack{m<q^k \\ m\equiv a \pmod q}} \mathbb{1}_{\mathcal{A}}(m) + O_R\left(\frac{q^k}{(\log q^k)^{4R}}\right)$$

If $a \neq a_0$, there are $(q-1)$ choices for each digit of m apart from the last one, which must be a. Thus the inner sum is $(q-1)^{k-1}$. If $a = a_0$, the sum is empty. Thus

$$\frac{q}{\phi(q)} \sum_{\substack{1<a<q \\ (a,q)=1}} \sum_{\substack{m<q^k \\ m\equiv a \pmod q}} \mathbb{1}_{\mathcal{A}}(m) = \begin{cases} q(q-1)^{k-1} & \text{if } (a_0,q) \neq 1 \\ \frac{\phi(q)-1}{\phi(q)} q(q-1)^{k-1} & \text{if } (a_0,q) = 1 \end{cases}$$

□

The last step here is to show that the terms with d not dividing any power of q are also in the error term; this follows from an L^∞ bound on these terms of the Fourier transform $\widehat{F}_{q^k}$, which finally yields the theorem.

REFERENCES

[1] Maynard, James. Primes and polynomials with restricted digits. https://arxiv.org/abs/1510.0771
[2] Maynard, James. Primes with restricted digits. *Invent. Math.* 217, no. 1, 127–218. 2019.
[3] Miller, Steven, and Takloo-Bighash, Ramin. The Circle method. https://web.williams.edu/Mathematics/sjmiller/public_html/BrownClasses/1/circlemethod.pdf