J. Maynard (2021) "Primes and Polynomials With Restricted Digits,"
International Mathematics Research Notices, Vol. 00, No. 0, pp. 1–23
https://doi.org/10.1093/imrn/rnab002

Primes and Polynomials With Restricted Digits

James Maynard*
Mathematical Institute, Woodstock Road, Oxford, UK, OX2 6GG

**Correspondence to be sent to: e-mail: james.alexander.maynard@gmail.com*

Let q be a sufficiently large integer, and $a_0 \in \{0, \dots, q-1\}$. We show there are infinitely many prime numbers that do not have the digit a_0 in their base q expansion. Similar results are obtained for values of a polynomial (satisfying the necessary local conditions) and if multiple digits are excluded.

1 Introduction

Let $a_0 \in \{0, \dots, q-1\}$ and let

$$\mathcal{A} = \Big\{ \sum_{i \geq 0} n_i q^i : n_i \in \{0, \dots, q-1\} \backslash \{a_0\} \Big\}$$

be the set of numbers that have no digit equal to a_0 when written in base q. For fixed q, the number of elements of $\mathcal{A}$ that are less than x is $O(x^{1-\epsilon_q})$, where $\epsilon_q = \log(q/(q-1))/\log q > 0$. In particular, $\mathcal{A}$ is a sparse subset of the natural numbers. A set being sparse in this way presents several analytic difficulties if one tries to answer arithmetic questions such as whether the set contains infinitely many primes. Typically we can only show that sparse sets contain infinitely many primes when the set in question possesses some additional multiplicative structure.

The set $\mathcal{A}$ has unusually nice structure in that its Fourier transform has a convenient explicit analytic description and is often unusually small in size. There has been much previous work [1, 2, 4–6, 10, 13] studying $\mathcal{A}$ and related sets by exploiting this Fourier structure. In particular the work by Dartyge and Mauduit [7, 8] shows

Received December 12, 2020; Revised December 12, 2020; Accepted December 31, 2020

the existence of infinitely many integers in $\mathcal{A}$ with at most 2 prime factors, this result relying on the fact that $\mathcal{A}$ is well distributed in arithmetic progressions [7, 12, 14]. We also mention the related work by Mauduit–Rivat [15] who showed the sum of digits of primes was well distributed, and the work by Bourgain [3] that showed the existence of primes in the sparse set created by prescribing a positive proportion of the digits.

We show that there are infinitely many primes in $\mathcal{A}$, and any polynomial P satisfying suitable local conditions takes infinitely many values in $\mathcal{A}$ provided the base q is sufficiently large (i.e., provided $\mathcal{A}$ is not too sparse). Our proof is based on the circle method, and in particular makes key use of the Fourier structure of $\mathcal{A}$, in the same spirit as the aforementioned works. Somewhat surprisingly, the Fourier structure is sufficient to deduce the existence of primes in $\mathcal{A}$ using only existing exponential sum estimates for the primes, and without having to investigate further bilinear sums.

Theorem 1.1. Let $q > 2000000, a_0 \in \{0, \dots, q-1\}$ and $\mathcal{A} = \{\sum_{i \geq 0} n_i q^i : n_i \in \{0, \dots, q-1\} \backslash \{a_0\}\}$ be the set of numbers with no digit in base q equal to a_0. Then for any constant $A > 0$ we have

$$\sum_{n<q^k} \Lambda(n) \mathbf{1}_{\mathcal{A}}(n) = \kappa_q(a_0)(q-1)^k + O_A\Big(\frac{(q-1)^k}{(\log q^k)^A}\Big),$$

where

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

Thus there are infinitely many primes with no digit a_0 when written in base q. There is nothing special about the fact we sum up to a power of q; one could sum n up to x instead of q^k and have $\sum_{n<x} \mathbf{1}_{\mathcal{A}}(n)$ instead of $(q-1)^k$ in the statement.

We have made no particular effort to optimize the lower bound on q; it is likely that it could be improved significantly. In particular, a more involved calculation shows that $q > 2500$ is sufficient by the same method, while it appears that the method of bilinear sums, Harman's sieve and zero density estimates all have the potential to show the existence of primes missing digits when the base is noticeably smaller. One might conjecture that the result would remain true for all $q > 2$.

As presented here the bound is ineffective due to the reliance on estimates for primes in arithmetic progressions. However, since these estimates are only used when

the modulus is highly composite, in fact Siegel zeros do not play a role, and so the error terms could be replaced by effective ones of size $O((q-1)^k \exp(-ck^{1/2}))$ if desired.

An analysis of our method reveals that in fact one can choose digits $a_0, \dots, a_{k-1} \in \{0, \dots, q-1\}$, and we obtain the same statement for primes $p = \sum_{i=0}^{k-1} p_i q^i$ with $p_i \in \{0, \dots, q-1\} \backslash \{a_i\}$ uniformly over all such choices of $a_0, \dots a_{k-1}$.

Our results hold for q sufficiently large not only because we require $\mathcal{A}$ to be not too sparse, but also because we separately get superior L^1 control on the Fourier transform of $\mathcal{A}$ as $q \to \infty$. A similar feature was present in the earlier work [11].

Theorem 1.2. Let $q > \exp(\exp(2r))$, and $P \in \mathbb{Z}[X]$ be a polynomial of degree r with lead coefficient a_r. Then for any $A > 0$ we have

$$\sum_{P(n)<q^k} 1_{\mathcal{A}}(P(n)) = a_r^{1/r} \mathfrak{S}(P) \frac{q^{k/r}(q-1)^k}{q^k} + O_{P,A}\Big(\frac{q^{k/r}(q-1)^k}{q^k (\log q^k)^A}\Big),$$

where

$$\mathfrak{S}(P) = \lim_{J \to \infty} \frac{\#\{(n,m) : 0 \le n, m < q^J,\ m \in \mathcal{A},\ P(n) \equiv m \pmod{q^J}\}}{(q-1)^J}.$$

Given a polynomial P it is a straightforward computation to determine whether $\mathfrak{S}(P) > 0$, in which case it takes infinitely many values in $\mathcal{A}$, or whether $\mathfrak{S} = 0$ in which case it takes finitely many values in $\mathcal{A}$. (This is because $P(\mathbb{Z}_p)$ is a disjoint union of open balls and a finite set of points in the p-adic topology.) In particular, by Hensel lifting, we see that Theorem 1.2 shows that there are infinitely many ℓ^{th} powers in $\mathcal{A}$, provided that $q > \exp(\exp(2\ell))$.

Again we have made no particular effort to optimize the lower bound on q. It is clear that the statement must require q to grow with r, since the main term $q^{k/r}(q-1)^k/q^k$ is only larger than 1 if q is large enough in terms of r. Presumably this bound would be improved if one used stronger bounds of Vinogradov type for the Weyl sums that appear rather than bounds based on Weyl differencing for large r, and by less crude numerical bounds. We note that although the implied constant in the error term in the statement of the theorem depends on the coefficients of P, the lower bound on q depends only on the degree.

Theorem 1.3. Let $\epsilon > 0$, $0 < s < q^{1/5-\epsilon}$ and let q be sufficiently large in terms of $\epsilon > 0$. Let $b_1, \dots, b_s \in \{0, \dots, q-1\}$ be distinct and let $\mathcal{B} = \{\sum_{i=0}^{k-1} n_i q^i : n_i \in$

$\{0,\dots,q-1\}\backslash\{b_0,b_1,\dots,b_s\}\}$ be the set of k-digit numbers in base q with no digit in the set $\{b_1,\dots,b_s\}$. Then we have

$$\sum_{n<q^k}\Lambda(n)\mathbf{1}_{\mathcal{B}}(n)=\frac{q(\phi(q)-s')}{(q-1)\phi(q)}(q-s)^k+O_A\Big(\frac{(q-s)^k}{(\log q^k)^A}\Big),$$

where $s'=\#\{1\le i\le s:(b_i,q)=1\}$.

Moreover, if $b_1,\dots,b_s$ are consecutive integers, then the same result holds provided only that $q-s\ge q^{4/5+\epsilon}$ and q is sufficiently large in terms of ϵ.

In the case of $b_1,\dots,b_s$ consecutive with $q-s=q^{4/5+\epsilon}$ we see that Theorem 1.3 shows the existence of primes in a set containing $x^{4/5+\epsilon}$ elements less than x. The exponent $4/5$ is ultimately related to the $4/5$ exponent of Lemma 4.2 for an exponential sum over primes and represents a limit of our basic method. As with Theorem 1.1, one would hope that utilizing Type I–II sums and Harman's sieve would extend this to sets of smaller density.

The conclusion of Theorem 1.3 holds in the case $q=10^8$ and $s=10$, so one can choose $\{b_1,\dots,b_{10}\}=\{0,11111111,22222222,\dots,99999999\}$. Thus there are infinitely many prime numbers with no string of 15 consecutive base 10 digits being the same. (Again, we expect 15 to be able to be reduced with slightly more effort.)

An analogous statement for the set $\mathcal{B}$ for polynomial values also holds, but in the more restrictive region $0<s<q^{1/r2^r-\epsilon}$ for arbitrary $b_1,\dots,b_s$ or $q-s\ge q^{1/r2^r+\epsilon}$ for consecutive $b_1,\dots,b_s$.

2 Notation

We use $e(x)=e^{2\pi ix}$ as the complex exponential and $\|x\|=\inf_{n\in\mathbb{Z}}|x-n|$ to denote the distance to the largest integer. We will use various expressions of the form $\min(A,\|\alpha\|^{-1})$, which are interpreted to take the value A if $\|\alpha\|=0$. We use $n\sim N$ to abbreviate $n\in[N,2N)$. Any implied constants in asymptotic notation $\ll$ or $O(\cdot)$ are allowed to depend on the base q and when dealing with polynomials as in Theorem 1.2, the polynomial P, but on no other quantity unless explicitly indicated by a subscript. Outside of Section 4 all quantities should be thought of as $k\to\infty$. In particular, k will implicitly be assumed to be larger than any fixed constant.

3 Outline

We give an informal sketch the overall outline of the proof, which is essentially an application of the Hardy–Littlewood circle method. We let $\hat{F}_X$ be the Fourier transform (over $\mathbb{Z}$) of the set $\mathcal{A}$ restricted to $\{1, \dots, X\}$. Thus for $X = q^k$ we have

$$\hat{F}_{q^k}(\theta) = \sum_{n \le q^k} \mathbf{1}_A(n) e(n\theta) = \prod_{i=0}^{k-1} \Big(\sum_{0 \le n_i \le q-1} \mathbf{1}_{\mathcal{A}}(n_i) e(n_i q^i \theta) \Big).$$

Here we have written $n = \sum_{i=0}^{k-1} n_i q^i$. It is this factorization of $\hat{F}_{q^k}$ and the fact that the sum over n_i is almost a geometric series, which allows us very good Fourier control over $\mathcal{A}$. By Fourier inversion on $\mathbb{Z}/q^k\mathbb{Z}$

$$\mathbf{1}_{\mathcal{A}}(n) = \frac{1}{q^k} \sum_{0 \le a < q^k} \hat{F}_{q^k}\Big(\frac{a}{q^k}\Big) e\Big(\frac{-an}{q^k}\Big).$$

Thus

$$\sum_{n \le q^k} \Lambda(n) \mathbf{1}_{\mathcal{A}}(n) = \frac{1}{q^k} \sum_{0 \le a < q^k} \hat{F}_{q^k}\Big(\frac{a}{q^k}\Big) S_{\Lambda, q^k}\Big(\frac{-a}{q^k}\Big),$$

where

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

We split the contribution up depending on whether a/q^k is close to a rational with small denominator or not. This distinguishes between those a when $S_{\Lambda,q^k}(a/q^k)$ is large or not. It turns out that $\hat{F}_{q^k}(a/q^k)$ is large if a is "close" to a number with few non-zero base q-digits, but these are somewhat rare and "spread out" except when a/q^k close to a rational with denominator being a small power of q, and so it turns out decomposition is adequate for describing $\hat{F}_{q^k}$ as well as S_{Λ,q^k}.

If $\mathcal{D}$ is the set of a such that $a/q^k = \ell/d + \beta$ for some integers $(\ell, d) = 1$ of and some $\beta \in \mathbb{R}$ with $d|\beta|$ of size D, we use a L^∞-L^1 bound to show their contribution is at most

$$\sup_{a \in \mathcal{D}} \Big| S_{\Lambda, q^k}\Big(\frac{a}{q^k}\Big) \Big| \sum_{a \in \mathcal{D}} \frac{1}{q^k} \Big| \hat{F}_{q^k}\Big(\frac{a}{q^k}\Big) \Big|.$$

One can save a small power of D over the trivial bound on $S_{\Lambda,q^k}(a/q^k)$ for $a \in \mathcal{D}$. By using a large-sieve type argument (and the analytic description of $\hat{F}$) we show equidistribution for a truncated version of $\hat{F}_J$ of $\hat{F}_{q^k}$

$$\sum_{a/q^k=\ell/d+\beta} \left|\hat{F}_J\Big(\frac{a}{q^k}\Big)\right| \approx J\int_0^1 |\hat{F}_J(\theta)|\mathrm{d}\theta,$$

where $J = \#\mathcal{D}$. We then use the explicit analytic description of $\hat{F}_{q^k}$ to obtain a final bound that is unusually strong. In particular, we importantly make use of the averaging over different β. This bound loses only a small power of D over the size of the largest individual terms in the sum. Crucially this power decreases to 0 as $q \to \infty$, while the power saving in S_{Λ,q^k} was independent of q, and so we have an overall saving of a small power of D if q is sufficiently large. This saving shows that these "minor arc" contributions when D is large are negligible.

Thus only those a/q^k that are very close to a rational (i.e., $d|\beta|$ is small) make a noticeable contribution. In this case the problem simply reduces to estimating primes and elements of $\mathcal{A}$ separately in short intervals and arithmetic progressions. For primes this is well known, while for the set $\mathcal{A}$ this follows from a suitable L^∞ bound on $\hat{F}$.

After writing this paper, the author discovered that very similar ideas appeared earlier in the literature, notably in [3, 12, 14, 15]. For simplicity we give an essentially self-contained proof but emphasize to the reader that many lemmas appearing are not new. It appears possible that (at least in the case when the base q is large) an argument similar to the one here might simplify or extend other arguments in the study of digit related functions.

Much of the previous work relied on estimating correlations of primes with digit-related functions relied on exploiting a certain property of the Fourier transform described in [16] as the "carry property," which often allowed one to simplify bilinear expressions so the Fourier transform only relied on the lower-order digits. This feature is not present in our work.

4 Exponential Sums for Primes and Polynomials

We first collect some results for exponential sums for primes and polynomials. The bounds here are well known, but we give a essentially complete proofs since they differ slightly from some standard references.

Lemma 4.1. Let $\alpha = a/d+\beta$ with a, d coprime integers and $\beta \in \mathbb{R}$ satisfying $|\beta| < 1/d^2$. Then we have

$$\sum_{n=1}^{N} \min\Big(M, \|\alpha n\|^{-1}\Big) \ll \Big(N + NMd|\beta| + \frac{1}{d|\beta|} + d\Big)\log N.$$

Proof. If $Nd|\beta| < 1/2$ then we let $n = n_0 + dn_1$ for non-negative integers n_0, n_1 with $n_0 < d$ and $n_1 < N/d$. If $n_0 \neq 0$ then

$$\|\alpha n\| = \|n_0 a/d + \beta n\| \geq \|n_0 a/d\| - \|\beta n\| \geq \|n_0 a/d\|/2$$

since $N|\beta| < 1/2$. We let $b \in \{0, \dots, d-1\}$ be such that $b \equiv m_0 a \pmod d$. Thus the terms with $n_0 \neq 0$ contribute a total

$$\ll \sum_{n_1 < N/d} \sum_{1 \leq b < \min(d,N)} \frac{d}{b} \ll \sum_{n_1 < N/d} d\log N \ll (N+d)\log N.$$

The terms with $n_0 = 0$ contribute

$$\ll \sum_{1 \leq n_1 < N/d} \min\Big(M, \|dn_1\beta\|^{-1}\Big) \ll \sum_{1 \leq n_1 < N/d} \frac{1}{dn_1|\beta|} \ll \frac{\log N}{d|\beta|}.$$

Here we have used the fact that since $n_0 = 0$ and we sum over $n \geq 1$ we must have $n_1 \geq 1$.

We now consider the case $Nd|\beta| > 1/2$. We let $n = n_0 + dn_1 + d\lfloor (d^2\beta)^{-1}\rfloor n_2$, with $0 \leq n_0 < d$, $0 \leq n_1 \leq (d^2\beta)^{-1}$ and $0 \leq n_2 \ll Nd\beta$. Thus we obtain

$$\sum_{n=1}^{N} \min\Big(M, \|\alpha n\|^{-1}\Big) \ll \sum_{\substack{n_1 \leq 1/d^2\beta \\ n_2 \ll N\beta/d}} \sum_{0 \leq n_0 < d} \min\Big(N, \Big\|\theta + n_1 d\beta + n_0(\ell/d + \beta)\Big\|^{-1}\Big)$$

where we have put $\theta = \beta d\lfloor (d^2\beta)^{-1}\rfloor m_2$ for convenience. The inner sum is of the form $\sum_i \min(N, \|\theta_i\|^{-1})$ for d points θ_i, which are $1/2d$ separated. Therefore the sum over m_0 is

$$\ll d\log d + \sup_{0 \leq m_0 < d} \min\Big(N, \|\theta + m_1 d\beta + m_0(\ell/d+\beta)\|^{-1}\Big)$$

$$\ll d\log N + \sup_{0 \leq \epsilon < 1} \min\Big(N, \frac{d}{\|d\theta + (m_1 + O(1))d^2\beta\|}\Big)$$

since $\|t\|^{-1} \le d\|dt\|^{-1}$ for all t. The term $d\log d$ contributes $\ll (M+d)\log N$ to the total sum, which is acceptable. Thus we are left to bound

$$\sum_{m_2 \ll Md\beta} \sup_{\theta\in\mathbb{R}} \sum_{m_1 \le 1/d\beta} \min\Big(N, \frac{d}{\|\theta + (m_1 + O(1))d^2\beta\|}\Big).$$

The inner sum is of the form ($O(1)$ copies of) $\sum_i \min(M, \|\theta_i\|^{-1})$ for $O(1/d^2\beta)$ points θ_i, which are $d^2\beta$-separated mod $ref1$. Therefore the inner sum is $\ll (M + d/d^2\beta)\log N$, and this gives a bound

$$\ll Nd|\beta|\Big(M + \frac{1}{d|\beta|}\Big)\log N \ll \Big(MNd|\beta| + N\Big)\log N.$$

Putting these bounds together gives

$$\sum_{n=1}^{N} \min\Big(M, \|\alpha n\|^{-1}\Big) \ll \Big(N + NMd|\beta| + \frac{1}{d|\beta|} + d\Big)\log N.$$

■

Lemma 4.2. Let $\alpha = a/d + \beta$ with $(a,d) = 1$ and $|\beta| < 1/d^2$. Then

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

Proof. From [9, (6), Page 142], taking $f(n) = e(n\alpha)$ we have that for any choice of $U, V \ge 2$ with $UV \le x$

$$\sum_{n<x} \Lambda(n)e(n\alpha) \ll U + (\log x) \sum_{1\le t<UV} \sup_{w} \Big|\sum_{w<r\le x/t} e(rt\alpha)\Big| + x^{1/2}(\log x)^3 \sup_{\substack{U\le M\le x/V \\ V\le j\le N/M}} \Big(\sum_{V<k<x/M} \Big|\sum_{\substack{M<m\le 2M \\ m\le x/k \\ m\le x/j}} e(\alpha m(j-k))\Big|\Big)^{1/2}.$$

The sum over r is clearly $\ll \min(x/t, \|t\alpha\|^{-1})$ and the sum over m is similarly $\ll \min(M, \|(j-k)\alpha\|^{-1})$. Putting t and $j-k$ into dyadic intervals and applying Lemma 4.1 to the resulting sums (or the trivial bound when $j = k$) gives a bound

$$\ll \Big(UV + xd|\beta| + \frac{1}{d|\beta|} + d + \frac{x}{U^{1/2}} + \frac{x}{V^{1/2}} + x|d\beta|^{1/2} + \frac{x^{1/2}}{|d\beta|^{1/2}} + x^{1/2}d^{1/2}\Big)(\log x)^4.$$

Choosing $U = V = x^{2/5}$ and simplifying the terms then gives the result. ■

Lemma 4.3. Let $P \in \mathbb{Z}[X]$ be an integer polynomial of degree $r \geq 2$ with lead coefficient a_r. Let $\alpha \in \mathbb{R}$ be such that $a_r r!\,\alpha = a/d + \beta$ with $(a,d) = 1$ and $|\beta| < 1/d^2$. Then for any constant $\epsilon > 0$ we have

$$S_{P,x}(\alpha) = \sum_{P(n)<x} e(\alpha P(n)) \ll_\epsilon x(\log x)\Big(\frac{1}{x} + \frac{1}{x^r d|\beta|} + d|\beta| + \frac{d}{x^r}\Big)^{1/2^r}.$$

Proof. If $\mathcal{I}$ is an interval contained in $[0,x]$ then

$$\Big|\sum_{n\in\mathcal{I}} e(\alpha P(n))\Big|^2 = \sum_{|h|<x} \sum_{\substack{n\in\mathcal{I}\\ n+h\in\mathcal{I}}} e(\alpha(P(n+h)-P(n))) \leq \sum_{|h|<x} \Big|\sum_{n\in\mathcal{I}(h)} e(\alpha Q_h(n))\Big|$$

where $\mathcal{I}(h) = \mathcal{I} \cap (\mathcal{I} - h)$ is an interval contained in $[0,x]$, and $Q_h(n) = P(n+h) - P(n)$ is a polynomial of degree $r-1$ with lead coefficient $a_r r h$. Applying this and Cauchy's inequality $r-1$ times gives

$$\begin{aligned}\Big|\sum_{n\in\mathcal{I}} e(\alpha P(n))\Big|^{2^{r-1}} &\leq (2x)^{2^{r-1}-r} \sum_{|h_1|,\dots,|h_{r-1}|<x} \Big|\sum_{n\in\mathcal{I}(h_1,\dots,h_{r-1})} e(\alpha r!\, h_1\dots h_{r-1} n)\Big| \\ &\ll x^{2^{r-1}-r} \sum_{H<x^{r-1}} \tau_{r-1}(H)\min(x, \|a_r r!\, H\|^{-1})\end{aligned}$$

where we have put $H = h_1 \dots h_r$. We split the sum depending on whether $\tau_{r-1}(H) > B$ or not, for some quantity B that we choose later. This shows that the inner sum is of size

$$\begin{aligned}&\ll \sum_{\substack{H<x^{r-1}\\ \tau_{r-1}(H)<B}} B\min(x, \|\alpha a_r r!\, H\|^{-1}) + \sum_{\substack{H<x^{r-1}\\ \tau_{r-1}(H)>B}} x\frac{\tau_{r-1}(H)^2}{B} \\ &\ll B\Big(x^{r-1} + x^r d|\beta| + \frac{1}{d|\beta|} + d\Big)\log x + \frac{x^r(\log x)^{(r-1)^2}}{B}\end{aligned}$$

by applying Lemma 4.1. Writing this bound as $x^r B/Z + x^r(\log x)^{(r-1)^2}/B$ and choosing $B = Z^{1/2}$ then gives the result, noting that $(\log x)^{(r-1)^2/2^{r-1}} < \log x$. ∎

5 Fourier Analysis

We now establish in turn several properties of the function $\hat{F}_{q^k}$, which are the key ingredient in our result.

Lemma 5.1 (L^1 bound). There exists a constant $C_q \in [1/\log q, 1+3/\log q]$ such that

$$\sup_{\theta\in\mathbb{R}} \sum_{0\le a<q^k} \Big|\hat{F}_{q^k}\Big(\theta+\frac{a}{q^k}\Big)\Big| \ll (C_q q\log q)^k.$$

Proof. We expand out the definition of $\hat{F}_{q^k}$, and let $n=\sum_{i=0}^{k-1} n_i q^i$ be the base-q expansion of n.

$$\hat{F}_{q^k}(t)=\sum_{n<q^k} \mathbf{1}_{\mathcal{A}}(n)e(tn)=\prod_{i=0}^{k-1}\Big(\sum_{n_i=0}^{q-1}\mathbf{1}_{\mathcal{A}}(n_i)e(n_i q^i t)\Big).$$

The sum over n_i is a sum over all values in $\{0,\dots,q-1\}\backslash\{a_0\}$, and so is bounded by

$$\Big|\frac{e(q^{i+1}t)-1}{e(q^i t)-1}-e(a_0 q^i t)\Big| \le \min\Big(q, 1+\frac{1}{2\|q^i t\|}\Big). \tag{5.1}$$

For $t\in[0,1)$, we write $t=\sum_{i=1}^{k} t_i q^{-i}+\epsilon$ with $t_1,\dots,t_k\in\{0,\dots,q-1\}$ and $\epsilon\in[0,1/q^k)$. We see that $\|q^i t\|^{-1}=\|t_{i+1}/q+\epsilon_i\|^{-1}$ for some $\epsilon_i\in[0,1/q)$. In particular, $\|q^i t\|^{-1}\le \max(q/t_{i+1}, q/(q-1-t_{i+1}))$ if $t_{i+1}\ne 0,q-1$. Thus we see that

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

Here we used a small computation to verify $\sum_{1\le t\le (q-1)/2} t^{-1}\le \log q$ for all integers $q<20$, while for $q\ge 20>2/(\log ref2-\gamma)$ (where γ is Euler's constant), we have $\sum_{1\le t\le (q-1)/2} t^{-1}\le \log q-\log ref2+\gamma+2/q\le \log q$. (This bound is only relevant to the final lower bound on q; for a qualitative statement a bound $O(\log q)$ suffices.) ■

Lemma 5.2 (Large sieve estimate). We have

$$\sup_{\theta\in\mathbb{R}}\sum_{d\sim D}\sum_{\substack{0<\ell<d\\(\ell,d)=1}}\sup_{|\epsilon|<\frac{1}{10D^2}}\Big|\hat{F}_{q^k}\Big(\frac{\ell}{d}+\theta+\epsilon\Big)\Big| \ll (D^2+q^k)(C_q\log q)^k.$$

Here C_q is the constant described in Lemma 5.1.

Proof. We have that

$$\hat{F}_{q^k}(t) = \hat{F}_{q^k}(u) + \int_t^u \hat{F}'_{q^k}(v)\mathrm{d}v.$$

Thus integrating over $u \in [t-\delta, t+\delta]$ we have

$$|\hat{F}_{q^k}(t)| \ll \frac{1}{\delta}\int_{t-\delta}^{t+\delta} |\hat{F}_{q^k}(u)|\mathrm{d}u + \int_{t-\delta}^{t+\delta} |\hat{F}'_{q^k}(u)|\mathrm{d}u.$$

We note that the fractions $\ell/d + \theta + \epsilon$ with $(\ell, d) = 1$, $d < 2D$ and $|\epsilon| < 1/10D^2$ are separated from one another by $\gg 1/D^2$. Thus

$$\sum_{d\sim D}\sum_{\substack{0<\ell<d\\(\ell,d)=1}} \sup_{|\epsilon|<1/10D^2} \Big|\hat{F}_{q^k}\Big(\frac{\ell}{d}+\theta+\epsilon\Big)\Big| \ll D^2\int_0^1 |\hat{F}_{q^k}(u)|\mathrm{d}u + \int_0^1 |\hat{F}'_{q^k}(u)|\mathrm{d}u.$$

We note that, writing $n = \sum_{i=0}^{k-1} n_i q^i$ we have

$$\begin{aligned}\hat{F}'_{q^k}(t) &= 2\pi i \sum_{n\le q^k} n\mathbf{1}_{\mathcal{A}}(n)e(nt)\\ &= 2\pi i \sum_{j=0}^{k-1} q^j \Big(\sum_{0\le n_j<q} n_j \mathbf{1}_{\mathcal{A}} e(n_j q^j t)\Big)\prod_{i\neq j}\Big(\sum_{0\le n_i<q}\mathbf{1}_{\mathcal{A}}(n_i)e(n_i q^i t)\Big).\end{aligned}$$

Thus, as in Lemma 5.1, we have

$$|\hat{F}'_{q^k}(t)| \ll \sum_{j=0}^{k-1} q^{j+1}\prod_{i\neq j}\min\Big(q, 1+\frac{1}{2\|q^i t\|}\Big) \ll q^k \prod_{i=0}^{k-1}\min\Big(q, 1+\frac{1}{2\|q^i t\|}\Big),$$

and we have the same bound for $|\hat{F}_{q^k}(t)|$ but without the q^k factor. We let $t = \sum_{i=1}^{k} t_i q^{-k} + \epsilon$ for some $t_1, \dots, t_k \in \{0, \dots, q-1\}$ and $\epsilon \in [0, 1/q^k)$. We see that, as in Lemma 5.1 we have

$$\begin{aligned}\int_0^1 \prod_{i=0}^{k-1}\min\Big(q, 1+\frac{1}{2\|q^i t\|}\Big)\mathrm{d}t &\ll \frac{1}{q^k}\sum_{t_1,\dots,t_k<q}\prod_{i=0}^{k-1}\Big(1+\min\Big(q, \frac{q}{2t_i}, \frac{q}{2(q-1-t_i)}\Big)\Big)\\ &\ll (C_q \log q)^k.\end{aligned}$$

Putting this all together then gives the result. ∎

Lemma 5.3 (Hybrid estimate). Let $B, D \gg 1$. Then we have

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

where C_q is the constant described in Lemma 5.1 and

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

Proof. The result follows immediately from Lemma 5.1 if $B > q^k$, so we may assume $B < q^k$. For any integer $k_1 \in [0,k]$ we have

$$\begin{aligned}\hat{F}_{q^k}(\alpha) &= \prod_{i=0}^{k-k_1-1}\Big(\sum_{n_i<q}\mathbf{1}_{\mathcal{A}}(n_i)e(n_iq^i\alpha)\Big)\prod_{i=k-k_1}^{k-1}\Big(\sum_{n_i<q}\mathbf{1}_{\mathcal{A}}(n_i)e(n_iq^i\alpha)\Big)\\ &= \hat{F}_{q^{k-k_1}}(\alpha)\hat{F}_{q^{k_1}}(q^{k-k_1}\alpha).\end{aligned}$$

Using this and the trivial bound $|\hat{F}_{q^j}(\theta)| \le (q-1)^j$, for $k_1+k_2 \le k$ we have that

$$\Big|\hat{F}_{q^k}\Big(\frac{\ell}{d}+\frac{\eta}{q^k}\Big)\Big| \le (q-1)^{k-k_1-k_2}\Big|\hat{F}_{q^{k_1}}\Big(\frac{q^{k-k_1}\ell}{d}+\frac{\eta}{q^{k_1}}\Big)\Big|\sup_{|\epsilon|\le B/q^k}\Big|\hat{F}_{q^{k_2}}\Big(\frac{\ell}{d}+\epsilon\Big)\Big|.$$

Substituting this bound gives

$$\begin{aligned}\sum_{d\sim D}\sum_{\substack{\ell<d\\(\ell,d)=1}}\sum_{\substack{|\eta|<B\\ q^k\ell/d+\eta\in\mathbb{Z}}}\Big|\hat{F}_{q^k}\Big(\frac{\ell}{d}+\frac{\eta}{q^k}\Big)\Big| &\ll (q-1)^{k-k_1-k_2}\\ &\times\sum_{d\sim D}\sum_{\substack{\ell<d\\(\ell,d)=1}}\sup_{|\epsilon|<B/q^k}\Big|\hat{F}_{q^{k_2}}\Big(\frac{\ell}{d}+\epsilon\Big)\Big|\sum_{\substack{|\eta|<B\\ q^k\ell/d+\eta\in\mathbb{Z}}}\Big|\hat{F}_{q^{k_1}}\Big(\frac{q^{k-k_1}\ell}{d}+\frac{\eta}{q^{k_1}}\Big)\Big|.\end{aligned}$$

We choose k_1 minimally such that $q^{k_1} > B$ and extend the inner sum to $|\eta| < q^{k_1}$. Applying Lemma 5.1 to the inner sum, and then Lemma 5.2 to the sum over d, ℓ gives

$$\sum_{d\sim D}\sum_{\substack{\ell<d\\(\ell,d)=1}}\sum_{\substack{|\eta|<B\\ q^k\ell/d+\eta\in\mathbb{Z}}}\Big|\hat{F}_{q^k}\Big(\frac{\ell}{d}+\frac{\eta}{q^k}\Big)\Big| \ll (q-1)^{k-k_1-k_2}q^{k_1}(q^{k_2}+D^2)(C_q\log q)^{k_1+k_2}.$$

We choose $k_2 = \min(k - k_1, \lfloor 2 \log D / \log q \rfloor)$. We see that

$$\Big(\frac{C_q q \log q}{q-1}\Big)^{k_1+k_2} \ll (D^2 B)^{\alpha_q},$$
$$D^2 q_1^k \Big(\frac{C_q \log q}{q-1}\Big)^{k_1+k_2} \ll \frac{D^2 B}{(q-1)^k}(C_q \log q)^k + (D^2 B)^{\alpha_q}.$$

Combining these bounds gives the result. ■

Lemma 5.4 (L^∞ bound). Let $d < q^{k/3}$ be of the form $d = d_1 d_2$ with $(d_1, q) = 1$ and $d_1 \neq 1$, and let $|\epsilon| < 1/2q^{2k/3}$. Then for any integer ℓ coprime with d we have

$$\Big|\hat{F}_{q^k}\Big(\frac{\ell}{d} + \epsilon\Big)\Big| \ll (q-1)^k \exp\Big(-c_q \frac{k}{\log d}\Big)$$

for some constant $c_q > 0$ depending only on q.

Proof. We have that

$$|e(n\theta) + e((n+1)\theta)|^2 = 2 + 2\cos(2\pi\theta) < 4\exp(-2\|\theta\|^2).$$

This implies that

$$\Big|\sum_{n_i<q} \mathbf{1}_{\mathcal{A}}(n_i) e(n_i\theta)\Big| \leq q - 3 + 2\exp(-\|\theta\|^2) \leq (q-1)\exp\Big(-\frac{\|\theta\|^2}{q}\Big).$$

We substitute this bound into our expression for $\hat{F}$, which gives

$$\Big|\hat{F}_{q^k}\Big(\frac{\ell}{d}\Big)\Big| = \prod_{i=0}^{k-1}\Big|\sum_{n_i<q} \mathbf{1}_{\mathcal{A}}(n_i) e(n_i q^i t)\Big|$$
$$\leq (q-1)^k \exp\Big(-\frac{1}{q}\sum_{i=0}^{k-1} \|q^i t\|^2\Big).$$

If $\|q^i t\| < 1/2q$ then $\|q^{i+1} t\| = q\|q^i t\|$. If $t = \ell/d_1 d_2$ with $d_1 > 1$, $(d_1, q) = 1$ and $(\ell, d_1) = 1$ then $\|q^i t\| \geq 1/d$ for all i. Similarly, if $t = \ell/d_1 d_2 + \epsilon$ with ℓ, d_1, d_2 as above $|\epsilon| < q^{-2k/3}/2$ and $d = d_1 d_2 < q^{k/3}$ then for $i < k/3$ we have $\|q^i t\| \geq 1/d - q^i|\epsilon| \geq 1/2d$. Thus, for any

interval $\mathcal{I} \subseteq [0, k/3]$ of length $\log d/\log q$, there must be some integer $i \in \mathcal{I}$ such that $\|q^i(\ell/d + \epsilon)\| > 1/2q^2$. This implies that

$$\sum_{i=0}^{k} \Big\| q^i\Big(\frac{\ell}{d} + \epsilon\Big)\Big\|^2 \geq \frac{1}{4q^4}\Big\lfloor \frac{k\log q}{3\log d}\Big\rfloor.$$

Substituting this into the bound for $\hat{F}$, and recalling we assume $d < q^{k/3}$ gives the result. ∎

6 Minor Arcs

We now use the exponential sum estimates from the previous sections to show that when α is "far" from a rational with small denominator the quantity $\hat{F}_{q^k}(\alpha)S_{\Lambda,q^k}(-\alpha)$ and $\hat{F}_{q^k}(\alpha)S_{P,q^k}(-\alpha)$ are typically small in absolute value.

Lemma 6.1. Let $1 \ll B \ll q^k/D_0D$ and $1 \ll D \ll D_0 \ll q^{k/2}$. Then we have

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

and

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

Here α_q is the constant described in Lemma 5.3.

Proof. By Lemma 5.3 we have that if $D^2B \ll q^k$ then

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

By Lemma 4.2 we have

$$\sup_{\substack{d\sim D\\(\ell,d)=1\\|\eta|\sim B}}\Big|\sum_{n<q^k}\Lambda(n)e\Big(-n\Big(\frac{\ell}{d}+\frac{\eta}{q^k}\Big)\Big)\Big|\ll\Big(q^{4k/5}+\frac{q^k}{(DB)^{1/2}}+\frac{(DB)^{1/2}}{q^{k/2}}\Big)(k\log q)^4.$$

Putting these together gives

$$\sum_{d\sim D}\sum_{\substack{0<\ell<d\\(\ell,d)=1}}\sum_{\substack{|\eta|\sim B\\q^k\ell/d+\eta\in\mathbb{Z}}}\Big|\hat{F}_{q^k}\Big(\frac{\ell}{d}+\frac{\eta}{q^k}\Big)S_{\Lambda,q^k}\Big(-\frac{\ell}{d}-\frac{\eta}{q^k}\Big)\Big|$$
$$\ll k^4q^k(q-1)^k\Big(\frac{(D^2B)^{\alpha_q}}{q^{k/5}}+\frac{(D^2B)^{\alpha_q}}{(DB)^{1/2}}+\frac{(DB)^{1/2}(D^2B)^{\alpha_q}}{q^{k/2}}\Big).$$

Recalling that $D^2B<q^k$ and $DB<q^k/D_0$ by assumption, we see that this is

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

and the 1st term clearly dominates the 2nd.

By partial summation we see that we obtain the same bound for $S_{\Lambda,q^k}(\alpha+O(1/q^k))$ as the bound for $S_{\Lambda,q^k}(\alpha)$ given in Lemma 4.2. Thus in the case $|\eta|\ll 1$ we obtain the well-known bound

$$\sup_{\substack{d\sim D\\(\ell,d)=1\\|\eta|\ll 1}}\Big|\sum_{n<q^k}\Lambda(n)e\Big(-n\Big(\frac{\ell}{d}+\frac{\eta}{q^k}\Big)\Big)\Big|\ll\Big(q^{4k/5}+\frac{q^k}{D^{1/2}}+\frac{D^{1/2}}{q^{k/2}}\Big)(k\log q)^4.$$

This gives

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

Recalling that $1\ll D\ll D_0\ll q^{k/2}$ then gives the result. ■

Lemma 6.2. Let $DB \ll q^k/D_0$ and $D \ll D_0 \ll q^{k/2}$. Then we have

$$\sum_{d\sim D}\sum_{\substack{0<\ell<da_r r!\\(\ell,d)=1}}\sum_{\substack{|\eta|\sim B\\ q^k\ell/d+\eta\in\mathbb{Z}}}\Big|\hat{F}_{q^k}\Big(\frac{\ell}{da_r r!}+\frac{\eta}{q^k a_r r!}\Big)S_{P,q^k}\Big(\frac{-\ell}{da_r r!}+\frac{-\eta}{q^k a_r r!}\Big)\Big| \ll k(q-1)^k q^{k/r}\Big(\frac{1}{(DB)^{1/r2^r-\alpha_q}}+\frac{q^{k\alpha_q}}{D_0^{1/2^r}}\Big),$$

and

$$\sum_{d\sim D}\sum_{\substack{0<\ell<da_r r!\\(\ell,d)=1}}\sum_{\substack{|\eta|\ll 1\\ q^k\ell/d+\eta\in\mathbb{Z}}}\Big|\hat{F}_{q^k}\Big(\frac{\ell}{da_r r!}+\frac{\eta}{q^k a_r r!}\Big)S_{P,q^k}\Big(\frac{-\ell}{da_r r!}+\frac{-\eta}{q^k a_r r!}\Big)\Big| \ll k(q-1)^k q^{k/r}\Big(\frac{1}{D^{1/r2^r-\alpha_q}}+\frac{D_0^{2\alpha_q+1/2^r}}{q^{k/2^r}}\Big).$$

Here α_q is the constant described in Lemma 5.3.

Proof. By Lemma 5.3 we have that if $D^2B \ll q^k$ then

$$\sum_{d\sim a_r r!D}\sum_{\substack{\ell<d\\(\ell,d)=1}}\sum_{\substack{|\eta|<B\\ q^k\ell/d+\eta\in\mathbb{Z}}}\Big|\hat{F}_{q^k}\Big(\frac{\ell}{d}+\frac{\eta}{q^k}\Big)\Big| \ll (q-1)^k(D^2B)^{\alpha_q}.$$

By Lemma 4.3 we have

$$\sup_{\substack{d\sim D\\(\ell,d)=1\\|\eta|\sim B}}\Big|\sum_{P(n)<q^k}e\Big(\frac{-P(n)}{a_r r!}\Big(\frac{\ell}{d}+\frac{\eta}{q^k}\Big)\Big)\Big| \ll kq^{k/r}\Big(\frac{1}{q^{k/r2^r}}+\frac{1}{(DB)^{1/2^r}}+\frac{(DB)^{1/2^r}}{q^{k/2^r}}\Big).$$

Putting these together gives

$$\sum_{d\sim D}\sum_{\substack{0<\ell<da_r r!\\(\ell,d)=1}}\sum_{\substack{|\eta|\sim B\\ q^k\ell/d+\eta\in\mathbb{Z}}}\Big|\hat{F}_{q^k}\Big(\frac{\ell}{da_r r!}+\frac{\eta}{q^k a_r r!}\Big)\sum_{P(n)<q^k}e\Big(P(n)\Big(\frac{-\ell}{da_r r!}+\frac{-\eta}{q^k a_r r!}\Big)\Big)\Big| \ll kq^{k/r}(q-1)^k\Big(\frac{(D^2B)^{\alpha_q}}{q^{k/r2^r}}+\frac{(D^2B)^{\alpha_q}}{(DB)^{1/2^r}}+\frac{(DB)^{1/2^r}(D^2B)^{\alpha_q}}{q^{k/2^r}}\Big). \tag{6.1}$$

Recalling that $D^2B < q^k$ and $DB < q^k/D_0$ by assumption, we see that this is

$$\ll kq^{k/r}(q-1)^k\Big((D^2B)^{\alpha_q-1/r2^r}+(D^2B)^{\alpha_q-1/2^{r+1}}+\frac{q^{k\alpha_q}}{D_0^{1/2^r}}\Big),$$

and the 1st term clearly dominates the 2nd. As in Lemma 6.1, in the case we instead sum over $|\eta| \ll 1$, we obtain the same bound as (6.1) with B replaced by 1, since by partial summation we obtain the bound of Lemma 4.3 for $S_{P,q^k}(\alpha)$ as $S_{P,q^k}(\alpha + O(1/q^k))$. This gives

$$\sum_{d\sim D}\sum_{\substack{0<\ell<da_r r!\\(\ell,d)=1}}\sum_{\substack{|\eta|\ll 1\\ q^k\ell/d+\eta\in\mathbb{Z}}}\Big|\hat{F}_{q^k}\Big(\frac{\ell}{da_r r!}+\frac{\eta}{q^k a_r r!}\Big)\sum_{P(n)<q^k}e\Big(P(n)\Big(\frac{-\ell}{da_r r!}+\frac{-\eta}{q^k a_r r!}\Big)\Big)\Big|$$

$$\ll kq^{k/r}(q-1)^k\Big(\frac{D^{2\alpha_q}}{q^{k/r2^r}}+\frac{D^{2\alpha_q}}{D^{1/2^r}}+\frac{D^{1/2^r+2\alpha_q}}{q^{k/2^r}}\Big).$$

Recalling $1 \ll D \ll D_0 \ll q^{k/2}$ gives the result. ■

7 Major Arcs

We now consider $\hat{F}_{q^k}(\alpha)S_{\Lambda,q^k}(-\alpha)$ and $\hat{F}_{q^k}(\alpha)S_{P,q^k}(-\alpha)$ when α is close to a rational with small denominator.

Lemma 7.1. Let $D, B \ll \exp(c_q^{1/2}k^{1/2}/3)$ where c_q is the constant from Lemma 5.4. Then we have

$$\sum_{\substack{d<D\\ \exists p|d,p\nmid q}}\sum_{\substack{0<\ell<d\\(\ell,d)=1}}\sum_{\substack{|\eta|\ll B\\ q^k\ell/d+\eta\in\mathbb{Z}}}\Big|\hat{F}_{q^k}\Big(\frac{\ell}{d}+\frac{\eta}{q^k}\Big)S_{\Lambda,q^k}\Big(\frac{-\ell}{d}+\frac{-\eta}{q^k}\Big)\Big| \ll \frac{q^k(q-1)^k}{\exp(c_q^{1/2}k^{1/2})},$$

and

$$\sum_{\substack{d<D\\ \exists p|d,p\nmid q}}\sum_{\substack{0<\ell<d\\(\ell,d)=1}}\sum_{\substack{|\eta|\ll B\\ q^k\ell/d+\eta\in\mathbb{Z}}}\Big|\hat{F}_{q^k}\Big(\frac{\ell}{d}+\frac{\eta}{q^k}\Big)S_{P,q^k}\Big(\frac{-\ell}{d}+\frac{-\eta}{q^k}\Big)\Big| \ll \frac{q^{k/r}(q-1)^k}{\exp(c_q^{1/2}k^{1/2})}.$$

Proof. This follows immediately from Lemma 5.4, using the trivial bound for the exponential sum involving primes or polynomials. ■

Lemma 7.2. Let $A > 0$. Then for $D, B < (\log q^k)^A$ and $d > q$ we have

$$\frac{1}{q^k}\sum_{\substack{d<D\\ p|d\Rightarrow p|q}}\sum_{\substack{0\le \ell<d\\ (\ell,d)=1}}\sum_{|b|<B}\hat{F}_{q^k}\Big(\frac{\ell}{d}+\frac{b}{q^k}\Big)S_{\Lambda,q^k}\Big(\frac{-\ell}{d}+\frac{-b}{q^k}\Big)$$
$$= \kappa_q(a_0)(q-1)^k + O_A\Big(\frac{(q-1)^k}{(\log q^k)^A}\Big),$$

where

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

Proof. If $b\neq 0$ then by the prime number theorem in arithmetic progressions in short intervals and partial summation we have

$$S_{\Lambda,q^k}\Big(\frac{-\ell}{d}+\frac{-b}{q^k}\Big) \ll_A \frac{q^k}{(\log q^k)^{4A}}.$$

Thus the terms with $b\neq 0$ contribute

$$\ll \frac{(\log q^k)^{3A}}{q^k}\sup_{0<a<q^k}\Big|\hat{F}_{q^k}\Big(\frac{a}{q^k}\Big)\Big|\frac{q^k}{(\log q^k)^{4A}} \ll \frac{(q-1)^k}{(\log q^k)^A}.$$

Here we used the trivial bound that $|\hat{F}_{q^k}(\theta)|\le (q-1)^k$ for all θ.

Using the prime number theorem in arithmetic progressions again, we see that

$$S_{\Lambda,q^k}\Big(\frac{-\ell}{d}\Big) = \frac{q^k}{\phi(d)}\sum_{\substack{0<c<d\\ (c,d)=1}} e\Big(\frac{-lc}{d}\Big) + O_A\Big(\frac{q^k}{(\log q^k)^{4A}}\Big) = \frac{\mu(d)q^k}{\phi(d)} + O_A\Big(\frac{q^k}{(\log q^k)^{4A}}\Big).$$

Thus we may restrict to $d|q$, since all other such d are not square free. Letting $\ell'/q = \ell/d$, we see terms with $b=0$ and $d|q$ contribute

$$\frac{1}{q^k}\sum_{0\le \ell'<q}\hat{F}_{q^k}\Big(\frac{\ell'}{q}\Big)S_{\Lambda,q^k}\Big(\frac{-\ell'}{q}\Big) = \frac{1}{q^{k-1}}\sum_{\substack{n,m<q^k\\ n\equiv m \pmod q}}\Lambda(n)\mathbf{1}_{\mathcal{A}}(m)$$
$$= \frac{q}{\phi(q)}\sum_{\substack{1<a<q\\ (a,q)=1}}\sum_{\substack{m<q^k\\ m\equiv a \pmod q}}\mathbf{1}_{\mathcal{A}}(m) + O_A\Big(\frac{q^k}{(\log q^k)^{4A}}\Big).$$

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

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

■

Lemma 7.3. For B, $q^J < \exp(qk^{1/2})$ we have

$$\frac{1}{q^k} \sum_{d|q^J} \sum_{\substack{0\le \ell<d \\ (\ell,d)=1}} \sum_{|b|<B} \hat{F}_{q^k}\Big(\frac{\ell}{d}+\frac{b}{q^k}\Big) S_{P,q^k}\Big(\frac{-\ell}{d}+\frac{-b}{q^k}\Big) = \mathfrak{S}_J \frac{a_r^{1/r} q^{k/r} (q-1)^k}{q^k} + O\Big(\frac{q^{k/r}}{q^{1/2r}}\Big),$$

where

$$\mathfrak{S}_J = \frac{\#\{(n,m):\ 0 \le n,m < q^J,\ m \in \mathcal{A},\ P(n) \equiv m \pmod{q^J}\}}{(q-1)^J}.$$

Proof. For $y \ll x^{1-1/2r}$ we have

$$\begin{aligned} &\#\{n : P(n) \in [x, x+y], n \equiv n_0 \pmod d\} \\ &\quad = \#\{n : a_r n^r + O(x^{1-1/r}) \in [x, x+y], n \equiv n_0 \pmod d\} \\ &\quad = \frac{y}{d a_r x^{1-1/r}} + O(1). \end{aligned}$$

Thus the values of $P(n) < q^k$ are well-distributed arithmetic progressions modulo $d < q^J$ and in short intervals of length $\gg q^{k(1-1/2r)}$. Therefore by partial summation we have that for $b \neq 0$ and $d < D$

$$\sum_{P(n)\le q^k} e\Big(-P(n)\Big(\frac{\ell}{d}+\frac{b}{q^k}\Big)\Big) \ll q^{k/r-1/2r}.$$

In particular, using the trivial bound $|\hat{F}_{q^k}(\theta)| \le (q-1)^k$, we have

$$\frac{1}{q^k} \sum_{d|q^J} \sum_{\substack{0\le \ell<d \\ (\ell,d)=1}} \sum_{0<|b|\ll B} \hat{F}_{q^k}\Big(\frac{\ell}{d}+\frac{b}{q^k}\Big) \sum_{P(n)\le q^k} e\Big(-P(n)\Big(\frac{\ell}{d}+\frac{b}{q^k}\Big)\Big) \ll \frac{(q-1)^k q^{k/r}}{q^k q^{1/2r}}.$$

Thus we may restrict our attention to $b = 0$. Rewriting ℓ/d as ℓ'/q^J the sum we see that these terms are equal to

$$\frac{1}{q^k}\sum_{0\le \ell'<q^J}\hat{F}_{q^k}\Big(\frac{\ell'}{q^J}\Big)\sum_{P(n)<q^k} e\Big(\frac{-P(n)\ell'}{q^J}\Big) = \frac{1}{q^{k-J}}\sum_{m<q^k}\mathbf{1}_{\mathcal{A}}(m)\sum_{\substack{P(n)<q^k\\ p(n)\equiv m \pmod{q^J}}} 1.$$

Putting n, m into residue classes $\pmod{p^J}$ then gives the result. ∎

8 Proof of Theorems 1.1 and 1.2

Proof of Theorem 1.1. By Fourier expansion we have

$$\sum_{n<q^k}\Lambda(n)\mathbf{1}_{\mathcal{A}}(n) = \frac{1}{q^k}\sum_{0\le a<q^k}\hat{F}_{q^k}\Big(\frac{a}{q^k}\Big)S_{\Lambda,q^k}\Big(\frac{-a}{q^k}\Big).$$

By Dirichlet's approximation theorem, for any choice of $0 < D_0$ and any $0 \le a < q^k$ there exists integers $(\ell, d) = 1$ with $d < D$ and a real $|\beta| < 1/DD_0$ such that

$$\frac{a}{q^k} = \frac{\ell}{d} + \beta.$$

We see that $q^k\ell/d + q^k\beta \in \mathbb{Z}$. We use Lemmas 7.2 and 7.1 to estimate the contribution when $\max(d, q^k|\beta|) < (\log q^k)^A$ and use Lemma 6.1 for the remaining cases. This gives

$$\frac{1}{q^k}\sum_{0\le a<q^k}\hat{F}_{q^k}\Big(\frac{a}{q^k}\Big)S_{\Lambda,q^k}\Big(\frac{-a}{q^k}\Big) = \kappa_q(a_0)(q-1)^k$$
$$+ O_A\Big((q-1)^k\Big(\frac{1}{(\log q^k)^A} + \frac{k^4}{(\log q^k)^{A(1/5-\alpha_q)}} + \frac{k^5q^{k\alpha_q}}{D_0^{1/2}} + \frac{k^5D_0^{1/2+2\alpha_q}}{q^{k/2}}\Big)\Big).$$

Choosing $D_0 = q^{k/2}$ we see that the error term is $O_B((q-1)^k(\log q^k)^{-B})$ provided $\alpha_q < 1/5$ and A is chosen such that $A > (B+5)/(1/5-\alpha_q)$. We recall from Lemmas 5.3 and 5.1 that

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

This clearly tends to zero as $q \to \infty$. A calculation shows that $\alpha_q < 0.198$ for $q > 2000000$. This gives the result. ∎

Proof of Theorem 1.2. The proof is essentially identical to that of Theorem 1.1 above. We choose $D_0 = q^{k/2}$ and split our summation according to ℓ, d, β such that

$$\frac{a}{a_r r!\, q^k} = \frac{\ell}{d} + \beta.$$

We use Lemma 6.2 in place of 6.1 for $\max(d, |\beta| q^k) > q^J$ and Lemma 7.3 instead of 7.2 along with Lemma 7.1 to deal with $\max(d, q^k|\beta|) < q^J$. For any choice of J with $q^J < \exp(qk^{1/2})$ we obtain

$$\frac{1}{q^k} \sum_{0 \le a < q^k} \hat{F}_{q^k}\Big(\frac{a}{q^k}\Big) S_{P,q^k}\Big(\frac{-a}{q^k}\Big) = \mathfrak{S}_J \frac{a_r^{1/r} q^{k/r} (q-1)^k}{q^k}$$

$$+ O_A\Big(\frac{q^{k/r}(q-1)^k}{q^k}\Big(\frac{1}{\exp(c_q^{1/2} k^{1/2})} + \frac{k^4}{q^{J(1/r2^r - \alpha_q)}} + \frac{k q^{k\alpha_q}}{D_0^{1/2^r}} + \frac{D_0^{1/2^r + 2\alpha_q}}{q^{k/2^r}}\Big)\Big).$$

Since $D_0 = q^{k/2}$, we see that provided $\alpha_q < 2^{-r}/r$ the error term is small. In particular there is some quantity $\mathfrak{S}$ such that for any such choice of $J < c_q^{1/2} k^{1/2}$

$$\mathfrak{S} = \mathfrak{S}_J + O\Big(\frac{k^4}{q^{J(1/r2^r - \alpha_q)}}\Big).$$

Thus, if $\alpha_q < 2^{-r}/r$, we see that $\mathfrak{S}_J$ converges to $\mathfrak{S}$ as $J \to \infty$ and that

$$\frac{1}{q^k} \sum_{0 \le a < q^k} \hat{F}_{q^k}\Big(\frac{a}{q^k}\Big) S_{P,q^k}\Big(\frac{-a}{q^k}\Big) = \mathfrak{S} \frac{a_r^{1/r} q^{k/r} (q-1)^k}{q^k} + O\Big(\frac{q^{k/r}}{\exp(c_q^{1/2} k^{1/2})}\Big).$$

Since $\alpha_q \to 0$ as $q \to \infty$, we see that $\alpha_q < 2^{-r}/r$ for $q > q_0(r)$. From the bound on $C_q \le 1 + 3/\log q$, we see that this holds for $q \ge \exp(\exp(2r))$. This completes the proof. ∎

9 Modifications for Theorem 1.3

In this section we sketch the modifications required to establish Theorem 1.3, leaving the precise details to the interested reader. The results of Section 4 remain unchanged. In Lemma 5.1, instead of equation (1), we have

$$\Big|\frac{e(q^{i+1}t) - 1}{e(q^i t) - 1} - \sum_{i=1}^{s} e(b_i q^i t)\Big| \le \min\Big(q, s + \frac{1}{2\|q^i t\|}\Big).$$

If the b_i are consecutive integers then this can be improved to $\min(2q, 1/\|q^i t\|)$. Thus we can instead take $C_q = C_{q,s} = 1 + (2+s)/\log q$ in general, or $C_{q,s} = 2 + 2/\log q$ if the b_i are consecutive. Lemma 5.2 remains unchanged while in Lemma 5.3 all occurrences of $q-1$ should be replaced by $q-s$. In particular, we have

$$\alpha_q = \alpha_{q,s} = \frac{\log\left(C_{q,s}\frac{q}{q-s}\log q\right)}{\log q}.$$

With these values of $\alpha_{q,s}$ and $C_{q,s}$ in place of α_q and C_q, the arguments and statements of Lemmas 5.4, 6.1, 6.2, 7.1, 7.2, and 7.3 all go through as before, except that any occurrence of $q-1$ must be replaced by $q-s$. In Lemma 7.1 we made use of the fact that there were two consecutive digits that were not excluded; clearly this still holds in the cases considered by Theorem 1.3.

The final proofs of Theorems 1.1 and 1.2 then work as before, provided that the constraints $\alpha_{q,s} < 1/5$ or $\alpha_{q,s} < r^{-1}2^{-r}$ hold. If the b_i are not necessarily consecutive then we take $C_{q,s} = 1 + (2+s)/\log q$ and see that if q is sufficiently large in terms of ϵ and $s < q/2$ then

$$\alpha_{q,s} \leq \frac{\log s}{\log q} + \epsilon.$$

In particular, if $s < q^{1/5-\epsilon}$ then $\alpha_{q,s} < 1/5$, as required. A computation reveals that if $q = 10^8$ and $s = 10$ then $\alpha_{q,s} < 1/5$, justifying the remark made after Theorem 1.3.

If the b_i are consecutive then we can take $C_{q,s} = 2 + 2/\log q$, and see that for q sufficiently large in terms of ϵ we have

$$\alpha_{q,s} \leq \frac{\log q/(q-s)}{\log q} + \epsilon.$$

Thus $\alpha_{q,s} < 1/5$ provided $q - s > q^{4/5+\epsilon}$.

Funding

This work was supported by a Clay research fellowship and a fellowship by examination of the Magdalen College, Oxford. The later stages were supported by a Royal Society Wolfson Merit Award and funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 851318).

Acknowledgments

We thank Ben Green for introducing the author to this problem.

References

[1] Banks, W. D., A. Conflitti, and I. E. Shparlinski. "Character sums over integers with restricted g-ary digits." *Illinois J. Math.* 46, no. 3 (2002): 819–36.

[2] Banks, W. D. and I. E. Shparlinski. "Arithmetic properties of numbers with restricted digits." *Acta Arithmetica* 112, no. 4 (2004): 313–32.

[3] Bourgain, J. "Prescribing the binary digits of primes, II." *Israel J. Math.* 206, no. 1 (2015): 165–82.

[4] Col, S. "Diviseurs des nombres ellipséphiques." *Period. Math. Hungar.* 58, no. 1 (2009): 1–23.

[5] Coquet, J. "On the uniform distribution modulo one of some subsequences of polynomial sequences." *J. Number Theory* 10, no. 3 (1978): 291–6.

[6] Coquet, J. "On the uniform distribution modulo one of subsequences of polynomial sequences. II." *J. Number Theory* 12, no. 2 (1980): 244–50.

[7] Dartyge, C. and C. Mauduit. "Nombres presque premiers dont l'écriture en base r ne comporte pas certains chiffres." *J. Number Theory* 81, no. 2 (2000): 270–91.

[8] Dartyge, C. and C. Mauduit. "Ensembles de densité nulle contenant des entiers possédant au plus deux facteurs premiers." *J. Number Theory* 91, no. 2 (2001): 230–55.

[9] Davenport, H. "Multiplicative number theory." *Graduate Texts in Mathematics*, vol. 74, 3rd ed. New York: Springer, 2000. Revised and with a preface by Hugh L. Montgomery.

[10] Drmota, M. and C. Mauduit. "Weyl sums over integers with affine digit restrictions." *J. Number Theory* 130, no. 11 (2010): 2404–27.

[11] Drmota, M., C. Mauduit, and J. Rivat. "The sum-of-digits function of polynomial sequences." *J. Lond. Math. Soc. (2)* 84, no. 1 (2011): 81–102.

[12] Erdoős, P., C. Mauduit, and A. Sárközy. "On arithmetic properties of integers with missing digits. I. Distribution in residue classes." *J. Number Theory* 70, no. 2 (1998): 99–120.

[13] Erdoős, P., C. Mauduit, and A. Sárközy. "On arithmetic properties of integers with missing digits. II. Prime factors. *Discrete Math.* 200, no. 1–3 (1999): 149–64. Paul Erdoős memorial collection.

[14] Konyagin, S. "Arithmetic properties of integers with missing digits: distribution in residue classes." *Period. Math. Hungar.* 42, no. 1–2 (2001): 145–62.

[15] Mauduit, C. and J. Rivat. "Sur un problème de Gelfond: la somme des chiffres des nombres premiers." *Ann. Math. (2)* 171, no. 3 (2010): 1591–646.

[16] Mauduit, C. and J. Rivat. "Prime numbers along Rudin Shapiro sequences." *J. Eur. Math. Soc.* 17 (2015): 2595–642.