ARITHMETIC PROPERTIES OF INTEGERS IN CHAINS AND REFLECTIONS OF g-ARY EXPANSIONS

DOMINGO GÓMEZ-PÉREZ AND IGOR E. SHPARLINSKI

ABSTRACT. Recently, there has been a sharp rise of interest in properties of digits primes. Here we study yet another question of this kind. Namely, we fix an integer base $g \geqslant 2$ and then for every infinite sequence

$$\mathcal{D} = \{d_i\}_{i=0}^{\infty} \in \{0, \ldots, g-1\}^{\infty}$$

of g-ary digits we consider the counting function $\varpi_{\mathcal{D},g}(N)$ of integers $n \leqslant N$ for which $\sum_{i=0}^{n-1} d_i g^i$ is prime. We construct sequences $\mathcal{D}$ for which $\varpi_{\mathcal{D},g}(N)$ grows fast enough, and show that for some constant $\vartheta_g < g$ there are at most $O(\vartheta_g^N)$ initial elements $(d_0, \ldots, d_{N-1})$ of $\mathcal{D}$ for which $\varpi_{\mathcal{D},g}(N) = N + O(1)$. We also discuss joint arithmetic properties of integers and mirror reflections of their g-ary expansions.

1. INTRODUCTION

We fix an integer base $g \geqslant 2$ and then for every infinite sequence

$$\mathcal{D} = \{d_i\}_{i=0}^{\infty} \in \{0, \ldots, g-1\}^{\infty}$$

of g-ary digits.

We say that $\mathcal{D}$ is of *length* N if d_{N-1} is the last non-zero element of $\mathcal{D}$ if such N exists; otherwise we say that $\mathcal{D}$ is of *infinite length*.

We form the sequence of integers

$$(1.1) \qquad u_{\mathcal{D},g}(n) = \sum_{i=0}^{n-1} d_i g^i$$

and define the counting function $\varpi_{\mathcal{D},g}(N)$ of integers $n \leqslant N$, for which $u_{\mathcal{D},g}(n)$ is prime. The question of the distribution of prime values in the sequences (1.1) has been introduced by Angell and Godwin [1], see also [17]. More precisely, both papers [1, 17] study sequences $\mathcal{D}$ such that the elements of (1.1) are all primes. Analogues of this question for polynomials over finite fields have been considered by Chou and Cohen [7] and more recently by Gómez-Pérez, Ostafe and Sha [9], which

---

have in fact motivated this work. The rest of our motivation comes from a series of recent striking results about primes with restricted digits [3, 5, 14, 15].

It is easy to see that for almost all sequences $\mathcal{D}$ (in the sense of the Lebesgue measure in one interpret $\mathcal{D}$ as a g-ary expansion of a real number in $[0, 1]$) we have $u_{\mathcal{D},g}(n) = g^{n+o(n)}$ (and in fact $g^{n-1} \leqslant u_{\mathcal{D},g}(n) < g^n$ if $\mathcal{D}$ does not contain zero digits). Hence, the standard heuristic suggests that $\varpi_{\mathcal{D},g}(N)$ has to grow as

$$\sum_{n=1}^{N} \frac{1}{\log u_{\mathcal{D},g}(n)} = \sum_{n=1}^{N} \frac{1}{n \log g + o(n)} \sim \frac{\log N}{\log g}$$

(clearly there are also some local conditions which we have ignored as we are only interested in the rate of growth).

On the other hand, one can clearly guarantee that $\varpi_{\mathcal{D},g}(N) \geqslant N - 1$ by simply taking $d_0 = 0$, $d_1 = 1$ if $g = 2$ and $d_0 = 2$, if $g \geqslant 3$, setting all other elements to zero. However, we are interested in prime values of $u_{\mathcal{D},g}(n)$ for nontrivial sequences $\mathcal{D}$ of large or infinite length.

For instance, in Section 2 we construct a sequence $\mathcal{D} \in \{0, \ldots, g - 1\}^{\infty}$ with infinitely many non-zero digits for which

$$(1.2) \qquad\qquad \varpi_{\mathcal{D},g}(N) \gg \log N$$

where, as usual, the expressions $A \ll B$, $B \gg A$ and $A = O(B)$ are each equivalent to the statement that $|A| \leqslant cB$ for some positive constant c. Throughout the paper the implied constants may depend on g.

Let $P_g(N)$ be the number of sequences $\mathcal{D}$ of length N (that is, with $d_{N-1} \neq 0$) such that $\varpi_{\mathcal{D},g}(N) = N - \eta$, where

$$\eta = \begin{cases} 1, & \text{if } g = 2, \\ 0, & \text{if } g \geqslant 3. \end{cases}$$

In particular, from the prime number theorem we immediately obtain the following trivial bound

$$P_g(N) \ll \frac{g^N}{N},$$

which we use as a benchmark for our improvements in Section 3.

We also use this opportunity to introduce another question about digits of primes. Namely, given a g-ary expansion

$$s = \sum_{i=0}^{n-1} d_i g^i, \qquad d_i \in \{0, \ldots, g - 1\}, \ i = 1, \ldots, n,$$

we denote by s_g^* the "mirror" reflection of s, that is,

$$s_g^* = \sum_{i=0}^{n} d_{n-1-i} g^i.$$

We denote by $M_g(N)$ the number of primes $p \in [g^{N-1}, g^N - 1]$, for which p_g^* is also prime. For example, if p is a Fermat prime, then p_2^* is also a prime. Although we have not been able to obtain any nontrivial bounds on $M_g(N)$, in Section 4 we give some other results about the simultaneous arithmetic structure of p and p_g^*. In passing, we note that corresponding question for polynomials is trivial as the "mirror" polynomial $X^N f(1/X)$ of a polynomial $f(X)$ of degree N has the same arithmetic structure as f.

2. Constructing Sequences With Many Primes

Using the bound of Chang [6, Corollary 11] on the smallest prime in an arithemtic progression modulo an integer composed out of small primes (see also [12]), we obtain the following more precise form of (1.2).

Theorem 2.1. *There is a sequence*

$$\mathcal{D} = \{d_i\}_{i=0}^{\infty} \in \{0, \ldots, g-1\}^{\infty}$$

that has infinitely many non-zero elements, for which

$$\varpi_{\mathcal{D},g}(N) \geqslant \left(\frac{1}{\log(12/5)} + o(1) \right) \log N$$

as $N \to \infty$.

Proof. We choose d_0 and d_1 in such a way that either $u_{\mathcal{D},g}(1)$ or $u_{\mathcal{D},g}(2)$ is prime. Now, assume that $d_0, \ldots, d_{n-1}$ have already been chosen for the initial segment of $\mathcal{D}$. Using a result of Chang [6, Corollary 11] (see also [12]) we see that for any $\varepsilon > 0$ there exists a constant c such that for every $n = 1, 2, \ldots$ there exists a prime p with

$$p \equiv u_{\mathcal{D},g}(n) \pmod{g^n} \qquad \text{and} \qquad p \leqslant c g^{n(12/5 + \varepsilon)}$$

(note that $\gcd(u_{\mathcal{D},g}(n), g) = 1$), see also for much larger class of moduli than g^n. We now define the next

$$m = (7/5 + \varepsilon)n + O(1)$$

elements of $\mathcal{D}$ as the g-ary digits of $(p - u_{\mathcal{D},g}(n))/g^n$. This implies the inductive inequality $\varpi_{\mathcal{D},g}(n + m) \geqslant \varpi_{\mathcal{D},g}(n) + 1$. Thus for $N_k = \lceil (12/5 + 2\varepsilon)^k \rceil$, $k = 1, 2, \ldots$, we obtain $\varpi_{\mathcal{D},g}(N_k) \geqslant k + O(1)$. Since ε is arbitrary, the result now follows. $\square$

3. Bounding the number of sequences with all primes

We now use a version of the Brun–Titchmarsh inequality, due to Montgomery and Vaughan [16, Theorem 2] to improve the trivial upper bound (1.2).

For $g \geqslant 3$ we define

$$(3.1) \qquad \gamma_g = g \min_{m=1,2,\dots} \left(\frac{2g}{m\varphi(g)\log g} \right)^{1/m},$$

where $\varphi(q)$ is the Euler function of the integer $q \geqslant 1$.

Clearly $\gamma_g < g$ for any g and also when g is large enough then $m = 1$ is the optimal value and thus

$$\gamma_g = \frac{2g^2}{\varphi(g)\log g}.$$

The standard bound on the Euler function (see [11, Theorem 328]) guarantees that

$$\gamma_g = O\left(\frac{g \log \log g}{\log g} \right)$$

as $g \to \infty$.

Theorem 3.1. *For a sufficiently large N, we have*

$$P_g(N) \ll \gamma_g^N,$$

where the implied constant is absolute.

Proof. We start with deriving an inductive inequality between $P_g(n)$ and $P_g(n+m)$ for an appropriately chosen m.

We first observe that the first n digits of any $n + m$ digit sequence $(d_0, \dots, d_{n+m-1})$ counted in $P_g(n + m)$ must come from a sequence counted in $P_g(n)$. Now, assume that $m > \eta$. Then, all such extensions of a n digit sequence to a $n + m$ digit sequence counted in $P_g(n + m)$ generates a prime $p \leqslant g^{n+m}$ in a fixed arithmetic progression modulo g^n. We now recall the upper bound from [16, Theorem 2]

$$(3.2) \qquad \pi(x; q, a) \leqslant \frac{2x}{\varphi(q)\log(x/q)},$$

on the number of primes $p \leqslant x$ in arithmetic progressions $p \equiv a \pmod{q}$, (see also [13, Theorem 6.6] for a slightly weaker result, which is still sufficient for our purposes). Therefore, we obtain

$$P_g(n + m) \leqslant P_g(n)\frac{2g^{n+m}}{\varphi(g^n)\log g^m} = P_g(n)\frac{2g^{m+1}}{m\varphi(g)\log g}.$$

We now conclude that for any fixed integer $m \geqslant 1$, denoting by r the remainder of N on division by m, and using the trivial bound $P_g(r) \leqslant g^r$, we have

$$P_g(N) \leqslant g^r \left(\frac{2g^{m+1}}{m\varphi(g)\log g} \right)^{\lfloor N/m \rfloor} = g^N \left(\frac{2g}{m\varphi(g)\log g} \right)^{\lfloor N/m \rfloor}$$
$$\ll g^N \left(\frac{2g}{m\varphi(g)\log g} \right)^{N/m}.$$

with an absolute implied constant. Simple calculus shows that

$$\lim_{z\to\infty} \left(\frac{2g}{z\varphi(g)\log g} \right)^{1/z} = 1.$$

Hence there is integer m_0, depending only on g, on which the minimum in (3.1) is achieved, and the result now follows. $\square$

We note that for the values of q in the medium range, for example, for $x^\vartheta \leqslant q \leqslant 2x^\vartheta$ for some fixed real $\vartheta \in (0,1)$, there are various improvements of (3.2), see [4, 8] and references therein. However these results do not seem to be useful in our context.

On the other hand, for smaller values of g one can obtain better values of γ_g via an application of the sieve of Eratosthenes instead of a direct application of [16, Theorem 2] (in fact implicitly this is a part of the argument of the proof of [16, Theorem 2], see [16, Lemma 3]).

For positive integers q and U we define the function

$$\varphi(q, U) = \max_{1\leqslant h\leqslant q} \sum_{\substack{u=1 \\ \gcd(u+h,q)=1}}^{U} 1.$$

In particular, $\varphi(q, q) = \varphi(q)$ is the classical Euler function. We also note that it can be defined in a more general but equivalent form

$$\varphi(q, U) = \max_{\substack{1\leqslant a,b\leqslant q \\ \gcd(a,q)=1}} \sum_{\substack{u=1 \\ \gcd(au+b,q)=1}}^{U} 1.$$

Using the Möbius function $\mu(d)$ over the divisors of q to detect the co-primality condition, see [13, Equation (1.18)] and interchanging the order of summation, we derive

$$\sum_{\substack{u=1 \\ \gcd(u+h,q)=1}}^{U} 1 = \sum_{d|q} \mu(d) \sum_{\substack{u=1 \\ d|u+h}}^{U} 1 = \sum_{d|q} \mu(d) \left(\frac{U}{d} + \xi_d \right),$$

where $|\xi_d| \leqslant (d-1)/d$ (since the condition $d \mid u + h$ puts u in a prescribed arithmetic progression modulo d). Hence, using s to denote the number of prime divisors of q we obtain

$$(3.3) \qquad \sum_{\substack{u=1 \\ \gcd(u+h,q)=1}}^{U} 1 \leqslant \sum_{d|q} \mu(d)\frac{U-1}{d} + \sum_{d|q} |\mu(d)| = \frac{\varphi(q)}{q}(U-1) + 2^s$$

by [13, Equation (1.36)]. However for our purposes below, we work with rather small values of q and U, so we can always compute $\varphi(q, U)$ explicitly.

We can now use the above argument to improve the values of γ_g of Theorem 3.1 for $g = 2, 3, 5, 10$, and show that

$$\gamma_2 = 1.876\ldots, \quad \gamma_3 = 2.622\ldots, \quad \gamma_5 = 3.947\ldots, \quad \gamma_{10} = 8.441\ldots$$

(corresponding to the $m = 16, 7, 4, 6$, respectively, in (3.1)). However, instead of using the bound (3.3) directly, we simply evaluate $\varphi(q, U)$ for concrete values of q and U that optimize our results.

Theorem 3.2. *For $g = 2, 3, 5, 10$, we have,*

$$P_g(N) \ll \vartheta_g^N,$$

where

$$\vartheta_2 = 5^{1/3} = 1.709\ldots, \qquad \vartheta_3 = 2, \qquad \vartheta_5 = 2, \qquad \vartheta_{10} = 6.$$

Proof. We present the argument in a rather generic form suitable for further generalizations. Let s be an appropriately chosen integer and let q_s be a product of first s primes that are relatively prime to g.

We proceed inductively as in the proof of Theorem 3.1. We assume that n is large enough so that we always have $u_{\mathcal{D},g}(n) > q_s$. However, now, instead of requesting that the extended sequence $\mathcal{D}_{n+m} = (d_0, \ldots, d_{n+m-1})$ corresponds to prime values of $u_{\mathcal{D},g}(n+m)$ we merely request that $\gcd(u_{\mathcal{D},g}(n+m), q_s) = 1$. Hence,

$$P_g(n+m) \leqslant P_g(n)\varphi(q_s, g^m).$$

A simple inductive argument implies that for any fixed m we have

$$P_g(N) \ll g^N \left(g^{-m}\varphi(q_s, g^m)\right)^{N/m} = \varphi(q_s, g^m)^{N/m}.$$

Now, for $g = 2$ we take $s = 2$ (so $q_s = 15$) and $m = 3$. For $g = 3$ we take $s = 1$ (so $q_s = 2$) and $m = 1$. For $g = 5$ we take $s = 2$ (so $q_s = 6$) and $m = 1$. Finally, for $g = 10$ we take $s = 2$ (so $q_s = 21$) and $m = 1$. $\qquad \square$

We remark that it is quite possible that the elementary method of the proof of Theorem 3.2 always improves on Theorem 3.1, but it seems to be more difficult to analyze.

One can also obtain similar results for sequences generating square free integers. More precisely, we can proceed exactly as in Theorem 3.2 but instead count integers in short intervals which fall in residues classes $a \pmod{q_s^2}$, where $\gcd(a, q_s^2)$ is square free.

Since we allow zero digits, one expects that $P_g(N)$ is a growing function of N as for any prime "ending" p one expects to find N such that $g^N + p$ is prime again.

This expectation is based on the standard heuristic predicting primes in increasing sequences of integers (without any local obstructions). Namely, since the series

$$\sum_{N=1}^{\infty} \frac{1}{\log(g^N + p)} = \infty$$

is diverging, for any p with $\gcd(g, p) = 1$ there are probably infinitely many positive integers N for which $g^N + p$ is prime (we need only one such N). Our numerical tests suggest that in fact $P_g(N)$ grows exponentially, see Figure 2.

4. Prime Mirrors in Arithmetic Progressions

For positive integers N and m and an arbitrary integer a, we denote by $R_g(N, m, a)$ the number of primes $p \in [g^{N-1}, g^N - 1]$ such that

$$p_g^* \equiv a \pmod{m}.$$

For $R_g(N, m, a)$ we have the following two trivial bounds

$$(4.1) \qquad R_g(N, m, a) \ll \frac{g^N}{N} \qquad \text{and} \qquad R_g(N, m, a) \ll \frac{g^N}{m} + 1.$$

We now obtain a bound which improves (4.1) in the medium range.

Theorem 4.1. *For any integer $m \geqslant 1$ we have*

$$R_g(N, m, a) \ll \frac{g^N \gcd(g^N, m)}{N m^{1/2}}.$$

Proof. We choose some integer parameter $r \geqslant 1$ and consider the integers

$$1 \leqslant b_1 < \ldots < b_t \leqslant g^r - 1,$$

formed by the top $r \leqslant N$ g-ary digits of primes $p \in [g^{N-1}, g^N - 1]$. Clearly, for at least $(t-1)/2$ values of $i = 1, \ldots, t-1$, we have

$$b_{i+1} - b_i \leqslant H,$$

where
$$H = 2\frac{g^r}{t-1}.$$

Let $Q(N, h)$ be the number of primes $p \leqslant g^N$ such that $p + h$ is also prime. Then we see that

(4.2)
$$\frac{t-1}{2} \leqslant \sum_{1 \leqslant h \leqslant H} Q(N, h).$$

Using a very special case of the classical result of Halberstam and Richert [10, Theorem 3.12], we see that

(4.3)
$$Q(N, h) \ll \frac{g^N}{N^2} \prod_{\substack{\ell \mid h,\ \ell \nmid g \\ \ell \geqslant 3 \text{ prime}}} \left(\frac{\ell-1}{\ell-2}\right).$$

It is easy to show that

(4.4)
$$\sum_{\substack{1 \leqslant h \leqslant H}} \prod_{\substack{\ell \mid h \\ \ell \text{ odd prime}}} \left(\frac{\ell-1}{\ell-2}\right) \ll H.$$

For example, using the elementary inequality
$$\frac{z-1}{z-2} \leqslant \left(\frac{z}{z-1}\right)^2$$

that holds for $z \geqslant 3$, we obtain
$$\prod_{\substack{\ell \mid h,\ \ell \nmid g \\ \ell \geqslant 3 \text{ prime}}} \left(\frac{\ell-1}{\ell-2}\right) \leqslant \prod_{\substack{\ell \mid h,\ \ell \nmid g \\ \ell \geqslant 3 \text{ prime}}} \left(\frac{\ell}{\ell-1}\right)^2 \leqslant \left(\frac{h}{\varphi(h)}\right)^2$$

and (4.4) follows immediately from the general results of Balakrishnan and Pétermann [2].

Thus, assuming that $t \geqslant 2$, substituting (4.3) and (4.4) in (4.2) we obtain
$$t \ll H\frac{g^N}{N^2} + 1 \ll \frac{g^{N+r}}{N^2 t} + 1.$$

Hence

(4.5)
$$t \ll \frac{g^{(N+r)/2}}{N}.$$

Now, writing a prime p as $p = g^{N-r}v + u$, with integers $u \in [0, g^{N-r}-1]$ and $v \in [g^{r-1}, g^r - 1]$, we see that $p_g^* = g^r u_g^* + v_g^*$. Clearly, if we define r by the condition

(4.6)
$$g^{N-r} \leqslant m < g^{N-r+1}$$

then for each value of v, and thus of v_g^*, the congruence

$$g^r u_g^* + v_g^* \equiv a \pmod{m}$$

defines u_g^* and thus u in at most $\gcd(m, g^r) \leqslant \gcd(m, g^N)$ ways. Hence

$$R_g(N, m, a) \leqslant t \gcd(m, g^N),$$

which together (4.5) and (4.6) implies the desired bound. $\qquad\square$

We now give an arithmetic application of Theorem 4.1. In particular, we show that the sum of divisors function

$$\sigma(k) = \sum_{d|k} d$$

grows linearly on average over the sequence p_g^* for primes p in the interval $p \in [g^{N-1}, g^N - 1]$.

Corollary 4.2. *We have,*

$$\frac{g^N}{N} \ll \sum_{p \in [g^{N-1}, g^N - 1]} \frac{\sigma(p_g^*)}{p_g^*} \ll \frac{g^N}{N}.$$

Proof. We set $R_g(N, m) = R_g(N, m, 0)$. We also write

$$\sum_{p \in [g^{N-1}, g^N - 1]} \frac{\sigma(p_g^*)}{p_g^*} = \sum_{p \in [g^{N-1}, g^N - 1]} \sum_{d | p_g^*} \frac{d}{p_g^*} = \sum_{p \in [g^{N-1}, g^N - 1]} \sum_{d | p_g^*} \frac{1}{d}$$

$$= \sum_{d < g^N} \sum_{\substack{p \in [g^{N-1}, g^N - 1] \\ p_g^* \equiv 0 \pmod{d}}} \frac{1}{d} = \sum_{d < g^N} \frac{1}{d} R_g(N, d).$$

Using Theorem 4.1 we obtain

$$\sum_{p \in [g^{N-1}, g^N - 1]} \frac{\sigma(p_g^*)}{p_g^*} \ll \frac{g^N}{N} \sum_{d < g^N} \frac{\gcd(g^N, d)}{d^{3/2}}$$

$$\ll \frac{g^N}{N} \sum_{r=0}^{N} g^r \sum_{e < g^{N-r}} \frac{1}{(g^r e)^{3/2}} \ll \frac{g^N}{N}.$$

The lower bound follows from the prime number theorem (see [13, Corollary 5.29]), as we trivially have $\sigma(n)/n \geqslant 1$. $\qquad\square$

We now denote by $\omega(k)$ the number of distinct prime divisors of an integer $k \geqslant 1$.

Corollary 4.3. *We have,*

$$\omega \left(\prod_{p \in [g^{N-1}, g^N - 1]} p_g^* \right) \gg \frac{N^2}{(\log N)^2}.$$

Proof. Let us consider only the primes with the first digit equal to 1. That is primes $p \in [g^{N-1}, 2g^{N-1} - 1]$. It is enough to consider

$$W_g(N) = \prod_{p \in [g^{N-1}, 2g^{N-1} - 1]} p_g^*.$$

and estimate

$$w_g(N) = \omega \left(W_g(N) \right).$$

For a prime ℓ we denote by $\nu_g(\ell, N)$ the largest power ℓ that divides $W_g(N)$. We have

$$\nu_g(\ell, N) \leqslant \sum_{j=1}^{\infty} R_g(N, \ell^j, 0).$$

Clearly $R_g(N, \ell^j, 0) = 0$ if $j \geqslant N \log g / \log \ell$. We also note that for $p \in [g^{N-1}, 2g^{N-1} - 1]$ we have $p_g^* \equiv 1 \pmod{g}$, hence only primes ℓ with $\gcd(\ell, g) = 1$ have to be considered.

Hence, using Theorem 4.1 we obtain

(4.7) $$\nu_g(\ell, N) \ll \frac{g^N}{N \ell^{1/2}}.$$

By the prime number theorem,

(4.8) $$\log W_g(N) \geqslant \log \left(\left(g^{N-1} \right)^{(1+o(1))g^{N-1}/(N \log g)} \right) = (1 + o(1))g^{N-1}.$$

On the other hand, using (4.7) we obtain

(4.9) $$\log W_g(N) = \sum_{\ell | W_g(N)} \log \left(\ell^{\nu_g(\ell, N)} \right) \leqslant \frac{g^N}{N} \sum_{\ell | W_g(N)} \frac{\log \ell}{\ell^{1/2}},$$

where the sum is over all primes $\ell \mid W_g(N)$. It is easy to see that by the prime number theorem and partial summation, for any real $L \geqslant 3$, we have

$$\sum_{\ell \leqslant L} \frac{\log \ell}{\ell^{1/2}} \ll L^{1/2}.$$

Hence

$$\sum_{\ell | W_g(N)} \frac{\log \ell}{\ell^{1/2}} \ll (w_g(N) \log w_g(N))^{1/2}$$

and combining this with (4.8) and (4.9) we obtain

$$(w_g(N) \log w_g(N))^{1/2} \gg N$$

and the result follows.   □

One can also derive from Theorem 4.1 that for all but $o(g^N/N)$ primes $p \in [g^{N-1}, g^N - 1]$, the cube-full part of p_g^* is at most $\psi(N)$ for any function $\psi(N)$ with $\psi(N) \to \infty$ as $N \to \infty$. The bounds (4.1) give only $\psi(N)N^{1/2}$.

It is also possible to derive other arithmetic applications of Theorem 4.1, for example to show that

$$g^N/N \gg \sum_{p \in [g^{N-1}, g^N - 1]} \varphi(p_g^*) \gg g^N/N.$$

Obtaining better bounds on $R_g(N, m, a)$, in particular, improving those in (4.1) and Theorem 4.1, is also an interesting problem as well, with many potential applications. For example, one can conjecture that if $\gcd(g, m) = 1$ then we have

$$R_g(N, m, a) \ll \frac{g^N}{mN}$$

in a wide range of parameters m and N. It is also possible that there is an asymptotic formula for $R_g(N, m, a)$, but it has to take into account some local conditions of the same type which are used for

$$(m, g) = (2, 2),\ (3, 2)$$

in Section 5.

5. Heuristics and Numerical Tests

Theorem 3.2 motivates us to define

$$\rho_g = \limsup_{N \to \infty} P_g(N)^{1/N}$$

thus $\rho_g \leqslant \gamma_g$. We believe that $P_g(N)$ grows exponentially and thus $\rho_g > 1$ but the growth is rather slow and thus ρ_g is much smaller than g. Figure 1 shows the growth of $P_g(N)$ for $N \leqslant 50$ and the following estimates seems to be more accurate:

$$\rho_2 \approx \rho_3 \approx 1.045, \quad \rho_5 \approx 1.05, \quad \rho_{10} \approx 2.25.$$
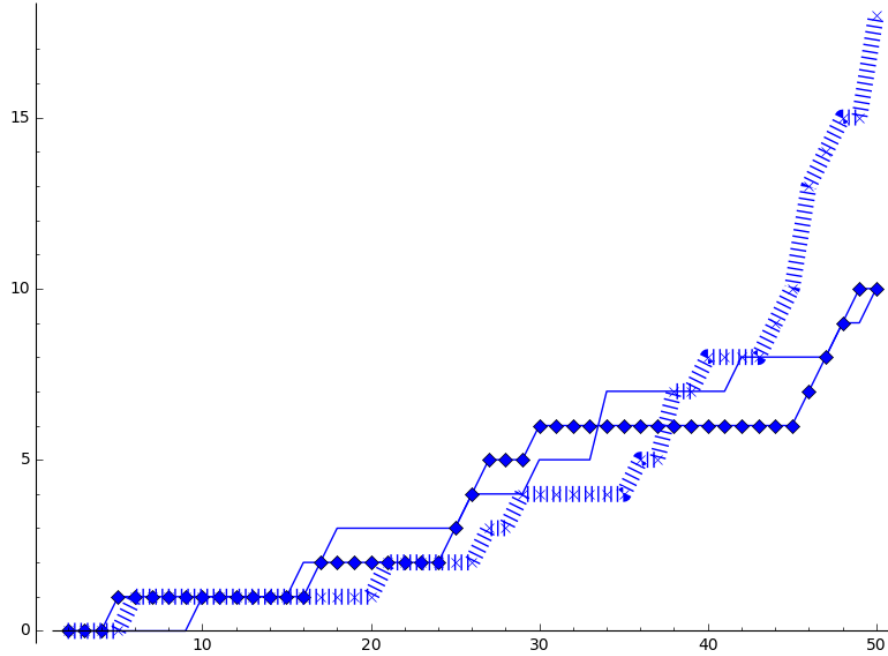
FIGURE 1. Growth of $P_g(N)$ for $g = 2$, 3 and 5 in solid, diamonds and dashed lines, respectively.

On the other hand, it seems that the arithmetic structure of g has to be reflected in any good approximation for ρ_g. For example, our computation seems to point that $\rho_{11} \approx 1.6$ and $\rho_{12} \approx 3.6$.

It is also natural to forbid zero digits and denote by $P_g^*(N)$ the number of prime generating sequences $\mathcal{D} = \{d_i\}_{i=0}^{N-1} \in \{1, \ldots, g - 1\}^{N-1}$ of length N that consist only of non-zero digits (in particular the definition is only interesting for $g \geqslant 3$). Heuristically, from each value $u_{\mathcal{D},g}(N-1)$ contributing to $P_g^*(N-1)$ we seek through $g-1$ values for $d_{N-1} \in \{1, \ldots, g-1\}$ such that $u_{\mathcal{D},g}(N) = d_{N-1}g^{N-1} + u_{\mathcal{D},g}(N-1)$ is prime. For $N \geqslant 2$, a naive approximation to the number of primes

$$p < g^N \qquad \text{and} \qquad p \equiv u_{\mathcal{D},g}(N - 1) \pmod{g^{N-1}}$$

is

$$\frac{g^N}{\varphi\left(g^{N-1}\right)\log(g^N)} = \frac{g^2}{N\varphi\left(g\right)\log g}.$$

However for d_{N-1} only $g - 1$ out g values are admissible. Hence, we are led to the approximate recursive relation

$$P_g^*(N) \approx \frac{g(g-1)}{N\varphi(g)\log g}P_g^*(N-1),$$

with the inital value $P_g^*(1) = \pi(g-1)$, which in turn leads us to the approximation

$$P_g^*(N) \approx A_g \frac{(g(g-1))^N}{N! \left(\varphi(g) \log g\right)^N},$$

where the coefficient A_g depends on the actual values of $P_g^*(N)$ for small values of N. Using the Stirling formula in the very crude form $N! \approx (N/e)^N$, we rewrite this as

$$(5.1) \qquad\qquad P_g^*(N) \approx A_g \left(\frac{eg(g-1)}{N\varphi(g) \log g} \right)^N$$

for some factor A_g depending only on g. We remark that it is hard to get any explicit formula for A_g, which depends on the initial behaviour of the sequence $P_g^*(N)$. In particular we see that (5.1), ignoring the presence of the factor A_g, suggests that $P_g^*(N) < 1$ (and thus $P_g^*(N) = 0$) for $N > N_g$ where

$$(5.2) \qquad\qquad N_g = \left\lfloor \frac{eg(g-1)}{\varphi(g) \log g} \right\rfloor .$$

Quite naturally, the approximation (5.1) is better when N is bigger. Figure 2 shows the values of the relative error of the ratio between $P_g^*(N)$ and the term of the approximation (5.1) that varies with N, that is,

$$\alpha_g(N) = P_g^*(N)^{1/N} \frac{N\varphi(g) \log g}{eg(g-1)}.$$

We expect that $\alpha_g(N)$ approximates $A_g^{1/N}$ from below (as the density of primes in the initial intervals $[1, x]$ is a little higher than $1/\log x$, especially for small values of x). So if the constant A_g in (5.1) is not too large, it is natural to expect that $\alpha_g(N)$ is close to 1 in the middle range of N (when the values of $P_g^*(N)$ are large). Figure 2 demonstrates the validity of this heuristic prediction.
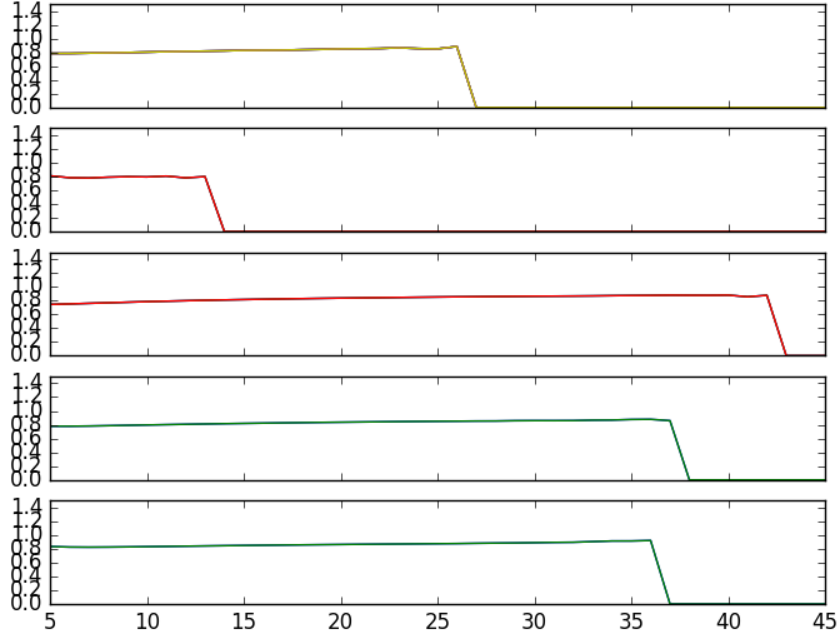
FIGURE 2. Values of $\alpha_g(N)$ for $g = 16, 17, 18, 20, 22$

We note that

$$N_{16} = 29, \qquad N_{17} = 16, \qquad N_{18} = 47, \qquad N_{20} = 43, \qquad N_{22} = 40,$$

while Figure 2 shows that the smallest values of N for which $\alpha_g(N) = 0$ for $g = 16$, 17, 18, 20 and 22 are 27, 14, 43, 38 and 37, respectively. We also see that in the middle range of N the function $\alpha_g(N)$ behaves as almost a constant function, until it suddenly drops to zero. So if the constant A_g in (5.1) is not too large, it is natural to expect that $\alpha_g(N)$ is close to 1 in the middle range of N (when the values of $P_g^*(N)$ are large). Certainly, using the precise values of $N!$ instead of the Stirling formula can also produce to a more precise numerical prediction of $P_g^*(N)$. For example, we always have $N! > (N/e)^N$ which leads to slight overestimation of N_g in (5.2).

Another consequence of the approximation (5.1) is that one expects $P_g^*(N) = 0$ for a sufficiently large N (by the Stirling formula, of size about $e(g - 1)/\log g$). This has been tested for $g < 40$ using the computer resources provided by the Santander Supercomputing and, in all the cases $P_g^*(N) = 0$ for sufficiently large N.

We now make some comments on the expected growth of $M_g(N)$. We concentrate on the case of $g = 2$. Clearly a mirror of a prime $p \in [2^{N-1}, 2^N - 1]$ is always odd, which we write as

$$(5.3) \qquad\qquad p_2^* \not\equiv 0 \pmod 2.$$

Furthermore considering the digit expansion of $s \in [2^{N-1}, 2^N - 1]$, we derive

$$s = \sum_{i=0}^{N-1} d_i 2^i \equiv \sum_{i=0}^{N-1} d_i (-1)^i$$

$$\equiv (-1)^{N-1} \sum_{i=0}^{N-1} d_i (-1)^{N-i-1} \equiv \pm s_2^* \pmod 3$$

we see that for any prime $p \in [2^{N-1}, 2^N - 1]$ we also have

$$(5.4) \qquad\qquad p_2^* \not\equiv 0 \pmod 3.$$

The local conditions (5.3) and (5.4) (there seems to be no other local conditions) coupled with the standard heuristic suggest that

$$
\begin{aligned}
M_2(N) &\approx \left(1 - \frac{1}{2}\right)^{-1} \cdot \left(1 - \frac{1}{3}\right)^{-1} \\
&\qquad \cdot \left(\pi(2^N) - \pi(2^{N-1})\right) \cdot \frac{\pi(2^N) - \pi(2^{N-1})}{2^{N-1}} \\
&= 6 \cdot \frac{\left(\pi(2^N) - \pi(2^{N-1})\right)^2}{2^N}.
\end{aligned}
\tag{5.5}
$$

We have done some computer experiments for $g = 2$ and we have plotted the results of $6 \cdot \left(\pi(2^N) - \pi(2^{N-1})\right)^2 / \left(2^N M_2(N)\right)$ in Figure 3, which in general seems to be consistent with (5.5). However, Figure 3 also indicates that there is some small positive bias.
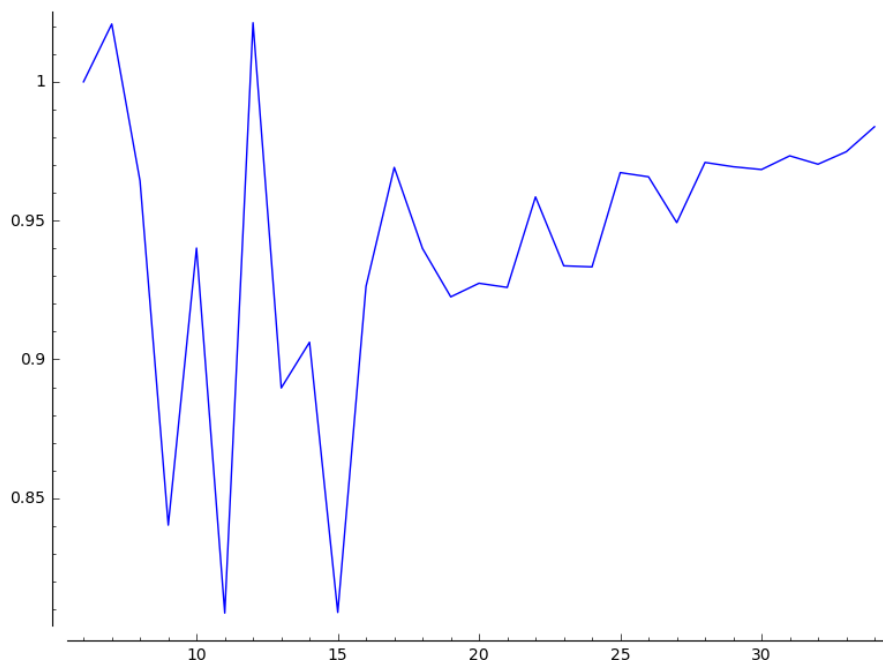
FIGURE 3. Growth of $6 \cdot (\pi(2^N) - \pi(2^{N-1}))^2 / (2^N M_2(N))$

ACKNOWLEDGEMENTS

REFERENCES

[1] I. O. Angell and H. J. Godwin 'On truncatable primes', *Math. Comp*, **31**, (1977), 265–267.

[2] U. Balakrishnan and Y.-F. S. Pétermann, 'The Dirichlet series of $\zeta(s)\zeta^\alpha(s+1)f(s+1)$: On an error term associated with its coefficients', *Acta Arith.*, **75** (1996), 39–69.

[3] J. Bourgain, 'Prescribing the binary digits of primes, II', *Israel J. Math.*, **206** (2015), 165–182.

[4] J. Bourgain and M. Z. Garaev, 'Sumsets of reciprocals in prime fields and multilinear Kloosterman sums', *Izvestiya: Mathematics*, **78** (2014), 656–707.

[5] M. Drmota, C. Mauduit and J. Rivat, 'Primes with an average sum of digits', *Compos. Math.*, **145** (2009), 271–292.

[6] M.-C. Chang, 'Short character sums for composite moduli', *J. d'Analyse Math.*, **123** (2014), 1–33.

[7] W.-S. Chou and S. D. Cohen, Polynomial distribution and sequences of irreducible polynomials over finite fields', *J. Number Theory*, **75** (1999), 145–159.

[8] J. Friedlander and H. Iwaniec, 'The Brun–Titchmarsh theorem', *Analytic Number Theory*, Lond. Math. Soc. Lecture Note Series **247**, 1997, 363–372.

[9] D. Gómez-Pérez, A. Ostafe and M. Sha, 'The arithmetic of consecutive polynomial sequences over finite fields', *Preprint*, 2015, http://arxiv.org/abs/1509.01936.

[10] H. Halberstam and H.-E. Richert, *Sieve Methods*, Academic Press, London, 1974.

[11] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, Oxford Univ. Press, Oxford, 1979.

[12] H. Iwaniec, 'On zeros of Dirichlet's L series', *Invent. Math.*, **23** (1974), 97–104.

[13] H. Iwaniec and E. Kowalski, *Analytic number theory*, Amer. Math. Soc., Providence, RI, 2004.

[14] C. Mauduit and J. Rivat, 'Sur un probléme de Gelfond: La somme des chiffres des nombres premiers', *Ann. of Math.*, **171** (2010), 1591–1646.

[15] J. Maynard, 'Primes with restricted digits', *Preprint*, 2016, http://arxiv.org/abs/1604.01041.

[16] H. L. Montgomery and R. C. Vaughan, 'The large sieve', *Mathematika*, **20** (1973), 119–134.

[17] A. J. van der Poorten, 'A quote', *Math. Intelligencer*, **7**(2), (1985), 40.

Department of Mathematics, Statistics and Computation, Universidad de Cantabria, Santander, 39005 Cantabria, Spain
   E-mail address: domingo.gomez@unican.es

Department of Pure Mathematics, University of New South Wales, Sydney, NSW 2052, Australia
   E-mail address: igor.shparlinski@unsw.edu.au