Rev. Mat. Iberoam. (online first)
DOI 10.4171/RMI/1234

On Cilleruelo's conjecture for the least common multiple of polynomial sequences

Zeév Rudnick and Sa'ar Zehavi

Abstract. A conjecture due to Cilleruelo states that for an irreducible polynomial f with integer coefficients of degree $d \geq 2$, the least common multiple $L_f(N)$ of the sequence $f(1), f(2), \dots, f(N)$ has asymptotic growth $\log L_f(N) \sim (d-1)N \log N$ as $N \to \infty$. We establish a version of this conjecture for almost all shifts of a fixed polynomial, the range of N depending on the range of shifts.

1. Introduction

1.1. Background

It is a well known and elementary fact that the least common multiple of all integers $1, 2, \dots, N$ is exactly given by

$$\log \operatorname{lcm}\{1, 2, \dots, N\} = \psi(N) := \sum_{n \leq N} \Lambda(n),$$

with $\Lambda(n)$ being the von Mangoldt function, and hence by the prime number theorem,

$$\log \operatorname{lcm}\{1, 2, \dots, N\} \sim N.$$

For a polynomial $f \in \mathbb{Z}[X]$, set

$$L_f(N) := \operatorname{lcm}\{f(n) : n = 1, \dots, N\}.$$

The goal is to understand the asymptotic growth of $\log L_f(N)$ as $N \to \infty$.

In the linear case $\deg f = 1$, we still have $\log L_f(N) \sim c_f N$ from the prime number theorem in arithmetic progressions, see e.g. [1]. A similar growth occurs for products of linear polynomials, see [3]. However, in the case of irreducible polynomials of higher degree, Cilleruelo [2] conjectured that the growth is faster than linear, precisely:

Mathematics Subject Classification (2010): 11N37.
Keywords: Least common multiple, irreducible polynomial, primes.

Conjecture 1.1. *If f is an irreducible polynomial with* $\deg f \geq 2$*, then*

$$\log L_f(N) \sim (\deg f - 1)N \log N, \quad N \to \infty.$$

Cilleruelo proved Conjecture 1.1 for quadratic polynomials. Moreover, in that case there is a secondary main term (see also [6]):

$$\log L_f(N) = N \log N + b_f N + o(N).$$

No other case of Conjecture 1.1 is known to date. We do know that for any irreducible f of degree $d \geq 3$, we have an upper bound precisely compatible with the conjecture: $\log L_f(N) \lesssim (d-1)N \log N$, and a lower bound of the correct order of magnitude: $\log L_f(N) \gg N \log N$, see [4].

We will show that Conjecture 1.1 holds *for almost all f* in a suitable sense.

1.2. General setup

We fix a polynomial $f_0(x) \in \mathbb{Z}[x]$ of degree $d \geq 3$, which we assume is monic, and for $a \in \mathbb{Z}$ we set

$$f_a(x) = f_0(x) - a.$$

It is known that $f_a(x)$ is generically irreducible. Set

$$L_a(N) = \operatorname{lcm}\{f_0(n) - a : n = 1, \dots, N\}.$$

We want to show that:

Theorem 1.2. *For almost all $|a| \leq T$, and for all N satisfying*

$$T^{1/(d-1)} < N < \frac{T}{\log T},$$

we have

$$\log L_a(N) \sim (d-1)N \log N. \tag{1.1}$$

Remark. What one would like to show is that (1.1) holds for *all* $N > N_0(a)$, for all but $o(T)$ values of $|a| \leq T$. At this time we do not know how to do this.

1.3. Plan

Let

$$P_a(N) = \prod_{n \leq N} |f_0(n) - a|.$$

We write down the prime power factorization

$$P_a(N) = \prod_p p^{\alpha_p(a;N)}.$$

If $T \ll N^{d-1}$, then $\alpha_p(a; N) = 0$ for $p \gg N^d$, and (see Lemma 2.3 below)

$$\log P_a(N) = dN \log N + O(N).$$

We also write the prime power factorization of $L_a(N)$ as

$$L_a(N) = \prod_p p^{\beta_p(N)}.$$

Let

$$D(a) = \mathrm{disc}(f_0(x) - a)$$

be the discriminant of $f_0(x) - a$. It is a polynomial in a of degree $d-1$, with integer coefficients.

We will show (see Proposition 2.2) that

$$\log L_a(N) = dN \log N - \mathrm{Bad}_N(a) - \Delta_N(a) - NC_N(a) + O(N), \tag{1.2}$$

where

$$\begin{aligned} \mathrm{Bad}_N(a) &= \sum_{\substack{p \le N \\ p \mid D(a)}} \alpha_p(N) \log p, \\ \Delta_N(a) &= \sum_{N < p \ll N^d} (\alpha_p(N) - \beta_p(N)) \log p, \\ C_N(a) &= \sum_{\substack{p \le N \\ p \nmid D(a)}} \frac{\log p}{p-1} \rho(a; p), \end{aligned}$$

with

$$\rho(a; d) = \#\{n \bmod d : f_0(n) - a = 0 \bmod d\}.$$

We will show that for almost all $|a| \le T$, with $N \log N < T < N^{d-1}$, we have

$$C_N(a) \sim \log N, \tag{1.3}$$

$$\mathrm{Bad}_N(a) \ll N(\log \log N)^{1+o(1)}, \tag{1.4}$$

$$\Delta_N(a) \ll N(\log \log N)^{1+o(1)}. \tag{1.5}$$

Inserting these into (1.2) will prove Theorem 1.2.

To prove (1.3), (1.4) and (1.5) we use averaging: denoting by $\langle \bullet \rangle$ the average over all $|a| \le T$ such that $f_0(x) - a$ is irreducible, we show that for $N \log N < T < N^{d-1}$,

$$\left\langle |C_N(a) - \log N|^2 \right\rangle \ll (\log \log N)^2, \tag{1.6}$$

$$\langle \mathrm{Bad}_N(a) \rangle \ll N \log \log N, \tag{1.7}$$

$$\langle \Delta_N(a) \rangle \ll N \log \log N. \tag{1.8}$$

Noting that $\Delta_N(a), \mathrm{Bad}_N(a) \ge 0$ are non-negative, we obtain (1.3), (1.4), (1.5) from the Chebyshev/Markov inequality.

Remark. In the deterministic case (a fixed, $N \to \infty$), the quantities Bad_N and C_N can be handled easily, as in the quadratic case $d = 2$, see [2]. It is the quantity $\Delta_N(a)$ which we do not know how to show is $o(N \log N)$ individually (though the upper bound $O(N \log N)$ is easy). This is why we need to average over a. However, letting a grow with N introduces new problems, in particular for the study of C_N, which may need the generalized Riemann hypothesis to overcome individually. The results (1.6) and (1.7) for random a are much easier and this is the method that we use.

Acknowledgements. We thank Shaofang Hong and Guoyou Qian for introducing the problem during a visit to Chengdu in 2017, and Lior Bary Soroker and James Maynard for discussions.

2. Background

2.1. Generic irreducibility

Fix $f_0(x) \in \mathbb{Z}[x]$ monic, of degree $d \geq 2$. It is known that $f_0(x) - a$ is generically irreducible; in fact (see Section 9.7 of [8]):

Lemma 2.1. *Fix $f_0(x) \in \mathbb{Z}[x]$ of degree $d \geq 2$. Then the number of $|a| \leq T$ for which $f_0(x) - a$ is reducible is $O(\sqrt{T})$.*

This is sharp in this generality, since for even degree $d = 2m$, for the polynomial $f_0(x) = x^{2m}$ we have $x^{2m} - a$ is reducible whenever $a = b^2$ is a perfect square.

Denote

$$D(a) = \mathrm{disc}(f_0(x) - a)$$

the discriminant of $f_a(x)$, which is a polynomial in a of degree $\leq d-1$ with integer coefficients (depending on the coefficients of f_0). We assume that a is such that $f_0(x) - a$ is irreducible, and therefore $D(a)$ is not zero, i.e., f_a has no multiple roots.

Examples:
i) $f_0(x) = x^3$, then $\mathrm{disc}(f_0(x) - a) = \mathrm{disc}(x^3 - a) = -27a^2$.
ii) $f_0(x) = x^3 - 3x$, then $\mathrm{disc}(f_0(x) - a) = -27(a - 2)(a + 2)$.

2.2. A decomposition

Proposition 2.2. *For $|a| \leq N^{d-1}$ such that $f_0(x) - a$ is irreducible, we have*

$$\log L_a(N) = d \log N - \mathrm{Bad}_N(a) - NC_N(a) - \Delta_N(a) + O(N).$$

For $a \in \mathbb{Z}$ such that $f_a(x) = f_0(x) - a$ is irreducible, let

$$P_a(N) := \prod_{n \leq N} |f_a(n)|,$$

which is nonzero since f_a has no rational roots, and write the prime power decomposition as

$$P_a(N) = \prod_p p^{\alpha_p(N)}$$

so that

$$\alpha_p(N) = \sum_{n \le N} \nu_p(f_a(n)),$$

where $\nu_p(m) := \max(k \ge 0 : p^k \mid m)$.

Following Cilleruelo [2], we want to relate $\log L_a(N)$ to $\log P_a(N)$, which is clearly bigger. We write the prime power decomposition of $L_a(N)$ as

$$L_a(N) = \prod_p p^{\beta_p(N)}, \quad \text{with} \quad \beta_p(N) = \max\{\nu_p(f_a(n)) : n \le N\}.$$

Using the prime factorization of $L_a(N)$ and $P_a(N)$, we have

$$\begin{aligned} \log L_a(N) = \log P_a(N) - \sum_{p \le N} \alpha_p(N) \log p + \sum_{p \le N} \beta_p(N) \log p \\ - \sum_{p > N} (\alpha_p(N) - \beta_p(N)) \log p, \end{aligned} \tag{2.1}$$

where we have separated out the contribution of primes $p \le N$, and the larger ones. We further break off the contribution of primes $p \le N$ which divide the discriminant $D(a) = \mathrm{disc}(f_a)$, by setting

$$\mathrm{Bad}_N(a) := \sum_{\substack{p \le N \\ p \mid D(a)}} \alpha_p(N) \log p,$$

and abbreviate the contribution of big primes $p > N$ as

$$\Delta_N(a) := \sum_{p > N} (\alpha_p(N) - \beta_p(N)) \log p.$$

Note that Bad_N and $\Delta_N \ge 0$ are both non-negative. We obtain the expression

$$\begin{aligned} \log L_a(N) = \log P_a(N) + \sum_{p \le N} \beta_p(N) \log p \\ - \mathrm{Bad}_N(a) - \sum_{\substack{p \le N \\ p \nmid D(a)}} \alpha_p(N) \log p - \Delta_N(a). \end{aligned} \tag{2.2}$$

2.3. The quantity $P_a(N)$

Lemma 2.3. *For $f_0(x) \in \mathbb{Z}[x]$ monic of degree d, and for $|a| \ll N^{d-1}$ so that $f_0(x) - a$ is irreducible, we have*

$$\log P_a(N) = dN \log N + O(N).$$

Proof. Write

$$\log P_a(N) = \sum_{n\leq N} \log|f_0(n) - a|.$$

Since we assume $f_0(x) - a$ is irreducible, non of the factors $f_0(n) - a$ can vanish, so that $\log P_a(N)$ is well defined. If $f_0(x) = x^d + c_{d-1}x^{d-1} + \cdots$, we have for $n \leq N$,

$$f_0(n) - a = n^d\Big(1 + \frac{c_{d-1}}{n} + \frac{c_{d-2}}{n^2} + \cdots - \frac{a}{n^d}\Big).$$

Consider first the n's satisfying $N/\log N < n \leq N$, for which we use (recall that $|a| \leq N^{d-1}$)

$$\log\Big(1 + \frac{c_{d-1}}{n} + \frac{c_{d-2}}{n^2} + \cdots - \frac{a}{n^d}\Big) = O\Big(\frac{(\log N)^d}{N}\Big),$$

so that

$$\sum_{\frac{N}{\log N}<n\leq N} \log|f_0(n)-a| = \sum_{\frac{N}{\log N}<n\leq N} \Big(d\log n + O\Big(\frac{(\log N)^d}{N}\Big)\Big) = dN\log N + O(N).$$

For $1 \leq n \leq N/\log N$, we just use $1 \leq |f_0(n)-a| \ll N^d$ so that $0 \leq \log|f_0(n)-a| \ll \log N$, and

$$\sum_{n\leq N/\log N} \log|f_0(n) - a| \ll \sum_{n\leq N/\log N} \log N \ll N.$$

Hence

$$\log P_a(N) = dN\log N + O(N)$$

as claimed. □

2.3.1. Dealing with $\beta_p(N)$. For a such that $f_0(x) - a$ is irreducible, we have

$$\beta_p(N) \ll \frac{\log N}{\log p}$$

because

$$\beta_p(N) = \max_{n\leq N} \max(k \geq 0 : p^k \mid f_0(n) - a),$$

and since $f_0(n) - a \neq 0$ for all n, if $p^k \mid f_0(n) - a \neq 0$, then

$$k \leq \frac{\log|f_0(n) - a|}{\log p} \ll \frac{\log n + \log|a|}{\log p}.$$

Hence, since $|a| \lll N^{d-1}$,

$$\beta_p(N) \ll \frac{\log N}{\log p},$$

and hence the contribution of primes $p \leq N$ to (2.2) is

$$\sum_{p\leq N} \beta_p(N)\log p \ll \sum_{p\leq N} \log N \ll N. \tag{2.3}$$

2.3.2. Dealing with $\alpha_p(N)$. Using Hensel's lemma, it is easy to check that (see [5], and also Lemma 4 in [2]):

Lemma 2.4. *For $p \nmid D(a) = \mathrm{disc}(f_0(x) - a)$, we have*

$$\alpha_p(N) = N\,\frac{\rho(a;p)}{p-1} + O\Big(\frac{\log N}{\log p}\Big),$$

where $\rho(a;p) = \#\{n \bmod p : f_0(n) - a = 0 \bmod p\}$.

Consequently, we find that in (2.2),

$$\sum_{\substack{p\le N\\ p\nmid D(a)}} \alpha_p(N)\log p = NC_N(a) + O(N), \quad \text{where} \quad C_N(a) := \sum_{\substack{p\le N\\ p\nmid D(a)}} \frac{\log p}{p-1}\,\rho(a;p).$$

Therefore we have proven Proposition 2.2.

3. Bounding Bad_N almost surely

Recall that we defined

$$\mathrm{Bad}_N(a) = \sum_{\substack{p\le N\\ p\mid D(a)}} \log p \sum_{n\le N} \#\{k \ge 1 : p^k \mid f_0(n) - a\}$$

(we assume that $f_0(x) - a$ is irreducible).

We denote the averaging operator over $|a| \le T$ such that $f_0(x) - a$ is irreducible by

$$\langle \bullet \rangle = \frac{1}{\#\{|a| \le T : f_0(x) - a \text{ is irreducible}\}} \sum_{\substack{|a|\le T\\ f_0(x)-a \text{ irreducible}}} \bullet$$

The number of $|a| \le T$ for which $f_0(x) - a$ is reducible is $O(\sqrt{T})$ (Lemma 2.1), so that

$$\langle \bullet \rangle = \frac{1}{2T + O(\sqrt{T})} \sum_{\substack{|a|\le T\\ f_0(x)-a \text{ irreducible}}} \bullet \tag{3.1}$$

Proposition 3.1. *If $T \ge N$ but $\log T \ll \log N$, then*

$$\langle \mathrm{Bad}_N \rangle \ll N \log\log N.$$

Proof. We separate out the contribution $B_1(a)$ of $k = 1$ and the contribution $B_2(a)$ of the remaining $k \ge 2$:

$$\mathrm{Bad}_N(A) = B_1(a) + B_2(a),$$

where

$$B_1(a) = \sum_{\substack{p\le N\\ p\mid D(a)}} \log p\, \#\{n \le N : f_0(n) = a \bmod p\}$$

and

$$B_2(a) = \sum_{\substack{p\leq N \\ p|D(a)}} \log p \sum_{n\leq N} \#\{k \geq 2 : p^k \mid f_0(n) - a\}.$$

We will show that

$$B_1(a) \ll N \log\log N \tag{3.2}$$

and that

$$\langle B_2 \rangle \ll N,$$

proving Proposition 3.1.

We first show that

$$B_1(a) \ll N \log\log |D(a)|,$$

which suffices for (3.2) since $\log |D(a)| \ll \log T \ll \log N$.

Indeed, for $p \leq N$ we have

$$\begin{aligned}\#\{n \leq N : f_0(n) = a \bmod p\} &= \Big(\frac{N}{p} + O(1)\Big) \#\{n \bmod p : f_0(n) = a \bmod p\} \\ &\ll \frac{N}{p}\, \rho(a;p)\,,\end{aligned}$$

where

$$\rho(a;p) := \#\{n \bmod p : f_0(n) = a \bmod p\},$$

which we see by dividing the interval $[1, N]$ into consecutive intervals of length p.

Since $f_0(x)$ is a monic polynomial of degree d, it is nonzero modulo p and still of degree $\leq d$, hence $\rho(a;p) \leq d$. Thus

$$B_1(a) \ll \sum_{\substack{p\leq N \\ p|D(a)}} \log p\, \frac{N}{p}\, \rho(a;p) \ll N \sum_{p|D(a)} \frac{\log p}{p}\,.$$

We use:

Lemma 3.2. *For $k > 1$,*

$$\sum_{p|k} \frac{\log p}{p} \ll \log\log k.$$

Proof. Indeed, splitting the sum into small primes $p \leq \log k$, and the rest (where the summands are at most $\log\log k/\log k$), we get

$$\begin{aligned}\sum_{p|k} \frac{\log p}{p} &\leq \sum_{\substack{p|k \\ p\leq \log k}} \frac{\log p}{p} + \sum_{\substack{p|k \\ p>\log k}} \frac{\log p}{p} \ll \sum_{p\leq \log k} \frac{\log p}{p} + \frac{\log\log k}{\log k} \sum_{p|k} 1 \\ &\ll \log\log k + \frac{\log\log k}{\log k} \cdot \frac{\log k}{\log\log k} \ll \log\log k,\end{aligned}$$

since the number of distinct prime divisors of k is $\ll \log k/\log\log k$. □

Therefore
$$\sum_{p\mid D(a)} \frac{\log p}{p} \ll \log\log|D(a)| \ll \log\log|a|,$$
and we obtain
$$B_1(a) \ll N\log\log|D(a)|.$$

Next we bound the mean value of $B_2(a)$:
$$\langle B_2\rangle = \frac{1}{2T+O(\sqrt{T})} \sum_{\substack{|a|\le T\\ f_0(x)-a\\ \text{irreducible}}} \sum_{\substack{p\le N\\ p\mid D(a)}} \log p \sum_{k\ge 2} \mathbf{1}(f_0(n)=a \bmod p^k).$$

Now if $f_0(x)-a$ is irreducible, then $f_0(n)-a\neq 0$, and so if $p^k \mid f_0(n)-a$ with $n\le N$, then $k \ll \log N/\log p$, so we restrict the summation to $2\le k\ll \log N/\log p$. Moreover, given n, the condition $f_0(n) = a \bmod p^k$ determines a modulo p^k, so there are $\ll T/p^k+1$ choices for a. Hence we may bound
$$\begin{aligned}\langle B_2\rangle &\ll \frac{1}{T}\sum_{p\le N}\log p \sum_{n\le N}\sum_{2\le k\ll \log N/\log p}\Big(\frac{T}{p^k}+1\Big)\\ &= \frac{N}{T}\sum_{p\le N}\log p \sum_{2\le k\ll \log N/\log p}\Big(\frac{T}{p^k}+1\Big).\end{aligned}$$

We have
$$\frac{N}{T}\sum_{p\le N}\log p \sum_{2\le k\ll \log N/\log p}\frac{T}{p^k} \ll N\sum_{p\le N}\log p\sum_{k\ge 2}\frac{1}{p^k} \ll N\sum_{p\le N}\frac{\log p}{p^2} \ll N$$
and
$$\frac{N}{T}\sum_{p\le N}\log p \sum_{2\le k\ll \log N/\log p} 1 \ll \frac{N}{T}\sum_{p\le N}\log p\cdot\frac{\log N}{\log p} \ll \frac{N^2}{T}.$$
Altogether we find
$$\langle B_2\rangle \ll N+\frac{N^2}{T},$$
which is $O(N)$ if $T\ge N$. □

4. Averaging $\Delta_N(a)$

Let
$$\Delta_N(a) = \sum_{p>N}\log p\big(\alpha_p(N)-\beta_p(N)\big).$$
Then clearly $\Delta_N\ge 0$, and we want to show:

Proposition 4.1. *Assume that $T\ge N\log N$, but $\log T\ll \log N$. Then*
$$\langle\Delta_N\rangle \ll_{f_0} N\log\log N.$$

4.1. Preparations

Let

$$G(m,n)=\frac{f_0(m)-f_0(n)}{m-n},$$

which, given n, is a (nonzero) polynomial in m, of degree $\le d-1$. If f_0 is monic then so is $G(m,n)$, so its degree is exactly $d-1$.

Lemma 4.2. *There is some $C_1=C_1(f_0)$ so that if $m,n\ge 1$ and $\max(m,n)>C_1$, then $G(m,n)\ne 0$.*

Proof. We have

$$G(m,n)=\sum_{j=1}^{d} c_j\,\frac{m^j-n^j}{m-n},$$

and if $j\ge 2$ then, for $n=\max(m,n)$,

$$\frac{m^j-n^j}{m-n}=n^{j-1}+n^{j-2}m+\cdots+m^{j-1}\le jn^{j-1},$$

while

$$\frac{m^d-n^d}{m-n}=n^{d-1}+n^{d-2}m+\cdots+m^{d-1}>n^{d-1},$$

so that (assuming f_0 monic, so $c_d=1$)

$$G(m,n)\ge\frac{m^d-n^d}{m-n}-\sum_{j=2}^{d-1}|c_j|\,\frac{m^d-n^d}{m-n}-|c_1|>n^{d-1}-\sum_{j=1}^{d-1}|c_j|\,j\,n^{j-1},$$

which is clearly positive once n is sufficiently large in terms of the coefficients $c_1,\dots,c_{d-1}$ of f_0. □

Lemma 4.3. *There is some $C(d)>0$ so that for all $|a|\le N^d$ such that $f_a(x)=f_0(x)-a$ is irreducible, we have $\alpha_p(N)\le C(d)$ if $p>N$. Moreover, $\alpha_p(N)=0$ unless $p\ll N^d+|a|$.*

Proof. We have, by definition,

$$\alpha_p(N)=\sum_{n\le N}\sum_{k\ge1}\mathbf{1}(f_0(n)=a \bmod p^k)=\sum_{k\ge1}\#\{n\le N: f_0(n)=a \bmod p^k\}.$$

Since we assume that $f_a(x)=f_0(x)-a$ is irreducible, hence has no rational zeros, we must have, if $p\mid f_a(n)$, that $p\le|f_a(n)|\ll N^d+|a|\ll N^d$ uniformly in $|a|\le T$ (recall $T\le N^d$). Hence $\alpha_p(N)=0$ for $p\gg N^d$.

Given n so that $p\mid f_a(n)$, with $p>N$, we claim that there are at most d such integers:

$$\#\{m\le N: f_a(m)=f_a(n) \bmod p\}\le d.$$

Indeed, for any $c \in \mathbb{Z}/p\mathbb{Z}$, the number of solutions $m \bmod p$ of $f_a(m) = c \bmod p$ is at most d, and since $p > N$, this certainly applies to those $m \le N$ which solve $f_a(m) = c$ with $c = f_a(n)$.

Moreover, if $p > N$, the maximal k so that $p^k \mid f_0(n) - a$ for some $n \le N$ is, because we assume $f_a(n) \ne 0$,

$$\ll \frac{\log(N^d + |a|)}{\log p} = O_d(1)$$

since we assume that $|a| \le T$ with $\log T \ll \log N$.

Therefore

$$\alpha_p(N) = \sum_{k \ge 1} \#\{n \le N : f(n) = 0 \bmod p^k\} \le \sum_{1 \le k \ll O_d(1)} d = O_d(1)$$

as claimed. □

4.2. A preliminary bound on $\Delta_N(a)$

Lemma 4.4. *If a is such that $f_0(x) - a$ has no rational zeros, and* $\log |a| \ll \log N$, *then*

$$\Delta_N(a) \ll \sum_{\substack{1 \le m < n \le N \\ G(m,n) \ne 0}} \sum_{\substack{N < p \ll N^d \\ p \mid f_0(m) - a \\ p \mid G(m,n)}} \log p + O(\log N). \tag{4.1}$$

Proof. We have $\alpha_p(N) \ne \beta_p(N)$ if and only if there are two distinct integers $m, n \le N$ so that $p \mid f_a(m)$ and $p \mid f_a(n)$. Using Lemma 4.3, we see that $\alpha_p(N) - \beta_p(N) = O_d(1)$ for $p > N$, and hence applying a union bound we obtain, if a is such that $f_a(x)$ has no rational zeros,

$$\Delta_N(a) \ll_d \sum_{1 \le m < n \le N} \sum_{\substack{N < p \ll N^d \\ p \mid f_0(m) - a \\ p \mid f_0(n) - a}} \log p.$$

Note that if $p \mid f_a(m)$ and $p \mid f_a(n)$ then $p \mid f_a(m) - f_a(n) = (m-n)G(m,n)$, and so since $p \nmid m - n$ (because $1 \le n - m \le N - 1 < p$), we must have $p \mid G(m,n)$. Thus

$$\Delta_N(a) \ll \sum_{1 \le m < n \le N} \sum_{\substack{N < p \ll N^d \\ p \mid f_0(m) - a \\ p \mid G(m,n)}} \log p. \tag{4.2}$$

We break off the terms corresponding to $G(m,n) = 0$. According to Lemma 4.2, the condition $G(m,n) = 0$ forces $m, n \le C_1$ to be bounded. Hence the contribution of such pairs to (4.2) is bounded by

$$\ll \sum_{m,n \le C_1} \sum_{\substack{N < p \ll N^d \\ p \mid f_0(m) - a}} \log p \ll \log N \max_{m \le C_1} \#\{p > N : p \mid a - f_0(m)\}.$$

Note that $0 < |f_0(m) - a| \ll |a| + 1$ if $m \leq C_1$ (we assume that a is such that $f_0(x) - a$ has no rational zeros, hence $f_0(m) - a \neq 0$), and hence the number of primes $p > N$ dividing $f_0(m) - a$ is at most $\ll \log|a|/\log N$. Hence the contribution of pairs $m < n$ with $G(m,n) = 0$ to (4.2) is at most $\ll \log|a|$. Thus

$$\Delta_N(a) \ll \sum_{\substack{1 \leq m < n \leq N \\ G(m,n) \neq 0}} \sum_{\substack{N < p \ll N^d \\ p | f(m) \\ p | G(m,n)}} \log p + O(\log|a|).$$

Finally, the assumption $\log|a| \ll \log N$ gives (4.1). □

4.3. Proof of Proposition 4.1

Now to average over $|a| \leq T$ (such that $f_0(x) - a$ is irreducible). Using (4.1), noting that $\log|a| \ll \log T \ll \log N$ gives

$$\langle \Delta_N \rangle \ll \sum_{\substack{1 \leq m < n \leq N \\ G(m,n) \neq 0}} \sum_{\substack{N < p \ll N^d \\ p | G(m,n)}} \log p \, \frac{1}{T} \#\{|a| \leq T : p \mid a - f_0(m)\} + O(\log N).$$

Given $1 \leq m < N$ and $N < p \ll N^d$, the number of $|a| \leq T$ with $a = f_0(m) \bmod p$ is $\ll T/p + 1$. Hence

$$\begin{aligned}\langle \Delta_N \rangle \ll & \sum_{\substack{1 \leq m < n \leq N \\ G(m,n) \neq 0}} \sum_{\substack{N < p \ll N^d \\ p | G(m,n)}} \frac{\log p}{p} + \frac{1}{T} \sum_{\substack{1 \leq m < n \leq N \\ G(m,n) \neq 0}} \sum_{\substack{N < p \ll N^d \\ p | G(m,n)}} \log p + O(\log N) \\ =: & \, I + II + O(\log N).\end{aligned}$$

To treat the sum II, we note if $m, n \leq N$, then $|G(m,n)| \leq C(f_0) N^{d-1}$ and so there are at most $d - 1$ distinct primes $p > N$ which divide $G(m,n)$ (which we assume is non-zero), and for these, $\log p \ll \log N$. Therefore

$$II \ll \frac{\log N}{T} \sum_{1 \leq m < n \leq N} (d-1) \ll \frac{N^2 \log N}{T},$$

which is $O(N)$ if $T > N \log N$.

To treat the sum I, we separate the prime sum into primes with $N < p \leq N \log N$ and the remaining large primes $N \log N < p \ll N^{d-1}$ to get

$$I \ll \sum_{\substack{1 \leq m < n \leq N \\ G(m,n) \neq 0}} \sum_{\substack{N < p < N \log N \\ p | G(m,n)}} \frac{\log p}{p} + \sum_{\substack{1 \leq m < n \leq N \\ G(m,n) \neq 0}} \sum_{\substack{N \log N < p \ll N^d \\ p | G(m,n)}} \frac{\log p}{p}.$$

We treat the sum over small primes by switching the order of summation:

$$\sum_{\substack{1\le m<n\le N\\ G(m,n)\neq 0}} \sum_{\substack{N<p<N\log N\\ p\mid G(m,n)}} \frac{\log p}{p} \le \sum_{N<p<N\log N} \frac{\log p}{p} \#\{1\le m<n\le N : G(m,n)=0 \bmod p\}.$$

Now given m, the congruence $G(m,n)=0 \bmod p$ (if solvable) determines $n \bmod p$ up to $d-1$ possibilities, since $G(m,n)$ is a monic polynomial of degree $d-1$ in n, and since $n\le N<p$ means that n is determined as an integer up to $d-1$ possibilities. Hence

$$\#\{1\le m<n\le N : G(m,n)=0 \bmod p\} \le (d-1)N,$$

and the sum over small primes is bounded by

$$\ll \sum_{N<p<N\log N} \frac{\log p}{p} N = N\big\{(\log(N\log N)+O(1))-(\log N+O(1))\big\} \sim N\log\log N$$

on using Mertens' theorem.

The sum over large primes is treated by using $\log p/p \ll 1/N$ for $p>N\log N$, giving

$$\sum_{\substack{1\le m<n\le N\\ G(m,n)\neq 0}} \sum_{\substack{N\log N<p\ll N^d\\ p\mid G(m,n)}} \frac{\log p}{p} \ll \frac{1}{N} \sum_{\substack{1\le m<n\le N\\ G(m,n)\neq 0}} \#\{p>N\log N : p\mid G(m,n)\}.$$

Now given $1\le m<n\le N$ with $G(m,n)\neq 0$, there are at most $d-1$ primes $p>N\log N$ dividing $G(m,n)\ll N^{d-1}$, so that the contribution of large primes is bounded by

$$\ll \frac{1}{N}\sum_{1\le m<n\le N}(d-1) \ll N.$$

This gives $I\ll N\log\log N$, and hence,

$$\langle \Delta_N\rangle \ll N\log\log N,$$

as claimed. □

5. Almost sure behaviour of C_N

5.1. An identity involving $C_N(a)$

Let $f\in\mathbb{Z}[x]$ be an irreducible polynomial, and let $\rho_f(p)$ be the number of distinct roots of the polynomial f modulo a prime p. It is well known [5] that for fixed f, the mean value of $\rho_f(p)$ over all primes is 1:

$$\frac{1}{\pi(x)}\sum_{p\le x}\rho_f(p) = 1+o_f(1).$$

We write

$$\rho_f(p) = 1 + \sigma_f(p)\,,$$

where $\sigma_f(p)$ is a fluctuating quantity, having mean zero.

Now fix

$$f_0(x) = x^d + c_{d-1}x^{d-1} + \cdots + c_1 x \in \mathbb{Z}[x],$$

a monic polynomial of degree d, and for $a \in \mathbb{Z}$ set

$$f_a(x) = f_0(x) - a.$$

Write $\rho(a;p) = \rho_{f_a}(p)$ and $\sigma(a;p) = \sigma_{f_a}(p)$. Note that $0 \le \rho(a;p) \le d$.

We write

$$C_N(a) := \sum_{\substack{p \le N \\ p \nmid \operatorname{disc}(f)}} \frac{\log p}{p-1}\,\rho(a;p) = \sum_{p \le N} \frac{\log p}{p} - E_N(a) + D_N(a) + O(1),$$

where

$$D_N(a) := \sum_{\substack{p \le N \\ p \nmid \operatorname{disc}(f_a)}} \frac{\log p}{p}\,\sigma(a;p) \quad \text{and} \quad E_N(a) := \sum_{\substack{p \le N \\ p \mid D(a)}} \frac{\log p}{p}\,.$$

By Mertens' theorem,

$$\sum_{p \le N} \frac{\log p}{p} = \log N + O(1).$$

The contribution $E_N(a)$ of primes dividing the discriminant $D(a) = \operatorname{disc}(f_0(x) - a)$ can be bounded individually, for $|a| \le T \ll N^d$, using Lemma 3.2 (assuming $D(a) \neq 0$):

$$E_N(a) = \sum_{\substack{p \le N \\ p \mid D(a)}} \frac{\log p}{p} \le \sum_{p \mid D(a)} \frac{\log p}{p} \ll \log\log |D(a)|.$$

Since $D(a)$ is a polynomial of degree $d-1$ in a, and $|a| \le T \ll N^d$, we find

$$\sum_{\substack{p \le N \\ p \mid D(a)}} \frac{\log p}{p} \ll \log\log N\,,$$

which is negligible relative to the main term. Hence

$$C_N(a) = \log N + D_N(a) + O(\log\log N).$$

In the following part, we will establish the following upper bound on the second moment of $D_N(a)$.

Proposition 5.1. *For $T \ge N \log N$, the second moment of $D_N(a)$ satisfies*

$$\left\langle |D_N|^2 \right\rangle \ll 1.$$

Using the triangle inequality and Cauchy–Schwartz, we obtain:

Proposition 5.2.

$$\left\langle |C_N - \log N|^2 \right\rangle \ll (\log\log N)^2.$$

As a consequence, we deduce our main objective for this section.

Proposition 5.3. *For almost all* $|a| \le T$ *(with* $N\log N \le T \ll N^{d-1}$*),*

$$C_N(a) = \log N + O(\log\log N).$$

5.2. Proof of Proposition 5.1

Proof. Expanding, we have

$$\left\langle (D_N)^2 \right\rangle = \sum_{p\le N}\sum_{q\le N} \frac{\log p \log q}{pq} \left\langle \sigma(a;p)\,\sigma(a;q) \right\rangle .$$

The diagonal contribution $p = q$ gives

$$\sum_{p\le N} \frac{(\log p)^2}{p^2} \left\langle \sigma(a;p)^2 \right\rangle .$$

Now note that

$$-1 \le \sigma(a;p) \le d-1$$

is uniformly bounded. This is because the polynomial $f_0(x)-a$ is monic of degree d, hence has at most d zeros modulo p, so that $0 \le \rho(a;p) \le d$ and so $-1 \le \sigma(a;p) \le d-1$. Thus we obtain a bound for the diagonal sum:

$$\sum_{p\le N} \frac{(\log p)^2}{p^2} \left\langle \sigma(a;p)^2 \right\rangle \ll \sum_{p\le N} \frac{(\log p)^2}{p^2} \ll 1.$$

For the off-diagonal terms, we use:

Lemma 5.4. *For distinct primes* $p \ne q$*,*

$$\left| \left\langle \sigma(\bullet;p)\,\sigma(\bullet;q) \right\rangle \right| \ll \frac{\sqrt{pq}\log(pq)}{T} + \frac{1}{\sqrt{T}}.$$

Therefore, given Lemma 5.4, we obtain

$$\begin{aligned}
\sum_{p\ne q\le N} \frac{\log p \log q}{pq} \left| \left\langle \sigma(a;p)\,\sigma(a;q) \right\rangle \right| &\ll \sum_{p\ne q\le N} \frac{\log p\log q}{pq}\left(\frac{\sqrt{pq}\log(pq)}{T} + \frac{1}{\sqrt{T}} \right) \\
&\ll \frac{\log N}{T}\Big(\sum_{p\le N} \frac{\log p}{\sqrt{p}} \Big)^2 + \frac{1}{\sqrt{T}} \Big(\sum_{p\le N} \frac{\log p}{p} \Big)^2 \\
&\ll \frac{N\log N}{T} + \frac{(\log N)^2}{\sqrt{T}},
\end{aligned}$$

which is $O(1)$ if $T \ge N\log N$, proving Proposition 5.1. □

5.3. Proof of Lemma 5.4

For the argument, it will be important to have a run over an interval. So we first remove the restriction on a in the averaging, that $f_0(x)-a$ is irreducible. Since $-1\leq \sigma(a;p)\leq d-1$, this introduces an error bounded by

$$\ll \frac{1}{T}\sum_{\substack{|a|\leq T\\ f_0(x)-a \text{ reducible}}}(d-1)^2 \ll \frac{1}{T}\#\{|a|\leq T: f_0(x)-a \text{ reducible}\}\ll \frac{1}{\sqrt{T}},$$

and so

$$\langle\sigma(a;p)\sigma(a;q)\rangle = \frac{1}{2T+O(\sqrt{T})}\sum_{|a|\leq T}\sigma(a;p)\,\sigma(a;q)+O\Big(\frac{1}{\sqrt{T}}\Big).$$

We express $\rho(a;p)$ as an exponential sum:

$$\rho(a;p)=\#\{x \bmod p: f_0(x)-a=0 \bmod p\} = \sum_{x \bmod p}\frac{1}{p}\sum_{t \bmod p} e\Big(\frac{t(f_0(x)-a)}{p}\Big).$$

The term $t=0$ contributes the main term of 1, and we obtain the following expression for $\sigma(a;p)=\rho(a;p)-1$:

$$\sigma(a;p)=\frac{1}{p}\sum_{t\neq 0 \bmod p} e\Big(-\frac{at}{p}\Big)\sum_{x \bmod p} e\Big(\frac{tf_0(x)}{p}\Big), \tag{5.1}$$

where $e(z):=e^{2\pi i z}$. Set

$$\mathcal{S}_{f_0}(b,n):=\sum_{x \bmod n} e\Big(\frac{bf_0(x)}{n}\Big).$$

Using (5.1), we have on switching orders of summation,

$$\begin{aligned}\frac{1}{2T+O(\sqrt{T})}&\sum_{|a|\leq T}\sigma(a;p)\,\sigma(a;q)\\ &=\frac{1}{2T+O(\sqrt{T})}\frac{1}{pq}\sum_{\substack{0\neq t \bmod p\\ 0\neq s \bmod q}}\sum_{|a|\leq T} e\Big(-a\Big(\frac{t}{p}+\frac{s}{q}\Big)\Big)\,\mathcal{S}_{f_0}(t,p)\,\mathcal{S}_{f_0}(s,q).\end{aligned}$$

Weil's bound [9], [7] shows that there is a constant $c(d)>0$ so that, for all primes p and all b coprime to p,

$$|\mathcal{S}_{f_0}(b,p)|\leq c(d)\sqrt{p}. \tag{5.2}$$

In fact, for any $f_0\in\mathbb{Z}[x]$ with $f_0(x)$ primitive of degree d, if $p>d$ then

$$|S_{f_0}(b,p)|\leq (d-1)\sqrt{p}.$$

Hence we find

$$\begin{aligned} |\langle \sigma(\bullet;p)\,\sigma(\bullet;q)\rangle| &\ll_d \frac{1}{T\sqrt{pq}} \sum_{\substack{0\neq t \bmod p\\ 0\neq s \bmod q}} \Big| \sum_{|a|\le T} e\Big(-a\Big(\frac{t}{p}+\frac{s}{q}\Big)\Big)\Big| + O\Big(\frac{1}{\sqrt{T}}\Big)\\ &= \frac{1}{T\sqrt{pq}} \sum_{\substack{m \bmod pq\\ \gcd(m,pq)=1}} \Big| \sum_{|a|\le T} e\Big(-\frac{am}{pq}\Big)\Big| + O\Big(\frac{1}{\sqrt{T}}\Big), \end{aligned}$$

where we have used that if $p \neq q$ are distinct primes, then as t and s vary over all invertible residues modulo p (resp., modulo q), $tq + sp \bmod pq$ covers all invertible residues modulo pq exactly once.

We sum the geometric progression

$$\Big| \sum_{|a|\le T} e\Big(-\frac{am}{pq}\Big)\Big| \ll \min\Big(T, \Big\|\frac{m}{pq}\Big\|^{-1}\Big)$$

where $\|\alpha\| = \operatorname{dist}(\alpha,\mathbb{Z})$. We may take $1 \le m < pq/2$, and then the bound is

$$\ll pq/m.$$

This gives

$$|\langle \sigma(\bullet;p)\,\sigma(\bullet;q)\rangle| \ll \frac{1}{T\sqrt{pq}} \sum_{\substack{1\le m\le pq/2\\ \gcd(m,pq)=1}} \frac{pq}{m} + O\Big(\frac{1}{\sqrt{T}}\Big) \ll \frac{\sqrt{pq}\log(pq)}{T} + O\Big(\frac{1}{\sqrt{T}}\Big),$$

proving Lemma 5.4. □

References

[1] Bateman, P., Kalb, J. and Stenger, A.: Problem 10797: A limit involving least common multiples. *Amer. Math. Monthly* **109** (2002), no. 4, 393–394.

[2] Cilleruelo, J.: The least common multiple of a quadratic sequence. *Compos. Math.* **147** (2011), no. 4, 1129–1150.

[3] Hong, S., Qian, G. and Tan, Q.: The least common multiple of sequence of product of linear polynomials. *Acta Math. Hungar.* **135** (2012), no. 1-2, 160–167.

[4] Maynard, J. and Rudnick, Z.: A lower bound on the least common multiple of polynomial sequences. To appear in *Riv. Math. Univ. Parma (N.S.)*.

[5] Nagel, T.: Généralization d'un théorème de Tchebycheff. *J. Math. Pures Appl. (8)* **4** (1921) 343–356.

[6] Rué, J., Šarka, P. and Zumalacárregui, A.: On the error term of the logarithm of the lcm of a quadratic sequence. *J. Théor. Nombres Bordeaux* **25** (2013), no. 2, 457–470.

[7] Schmidt, W. M.: *Equations over finite fields: an elementary approach.* Second edition. Kendrick Press, Heber City, UT, 2004.

[8] SERRE, J.-P.: *Lectures on the Mordell–Weil theorem*. Third edition. Aspects of Mathematics, Friedr. Vieweg & Sohn, Braunschweig, 1997.

[9] WEIL, A.: On some exponential sums. *Proc. Nat. Acad. Sci. U. S. A.* **34** (1948), 204–207.

Received May 18, 2019; revised October 30, 2019. Published online November 3, 2020.

ZEÉV RUDNICK: School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel.
E-mail: rudnick@tauex.tau.ac.i

SA'AR ZEHAVI: School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel.
E-mail: saarzehavi@mail.tau.ac.il

The research was supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreements no. 320755 and 786758.