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

Ze‘ev 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 \( f(1), f(2), \ldots, f(N) \) 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 \( 1, 2, \ldots, N \) is exactly given by

\[
\log \text{lcm}\{1, 2, \ldots, N\} = \psi(N) := \sum_{n \leq N} \Lambda(n),
\]

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

\[
\log \text{lcm}\{1, 2, \ldots, N\} \sim N.
\]

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

\[
L_f(N) := \text{lcm}\{f(n) : n = 1, \ldots, N\}.
\]

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:

\[ \text{Mathematics Subject Classification (2010): 11N37.} \]

\[ \text{Keywords: Least common multiple, irreducible polynomial, primes.} \]
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: \( \log L_f(N) \ll (d-1)N \log N \), 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
\[
L_a(N) = \text{lcm}\{ f_0(n) - a : n = 1, \ldots, N \}.
\]

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
\[
(1.1) \quad \log L_a(N) \sim (d-1)N \log N.
\]

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

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

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

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

where

$$\text{Bad}_N(a) = \sum_{p \leq N \atop p \mid D(a)} \alpha_p(N) \log p,$$

$$\Delta_N(a) = \sum_{N < p \ll N^d} \left( \alpha_p(N) - \beta_p(N) \right) \log p,$$

$$C_N(a) = \sum_{p \leq N \atop p \nmid D(a)} \frac{\log p}{p - 1} \rho(a; p),$$

with

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

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

$$(1.3) \quad C_N(a) \sim \log N,$$

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

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

Inserting these into (1.2) will prove Theorem 1.2.

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

$$(1.6) \quad \langle |C_N(a) - \log N|^2 \rangle \ll (\log \log N)^2,$$

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

$$(1.8) \quad \langle \Delta_N(a) \rangle \ll N \log \log N.$$ 

Noting that $\Delta_N(a), \text{Bad}_N(a) \geq 0$ 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 $\text{Bad}_N$ 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

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

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 $\text{disc}(f_0(x) - a) = \text{disc}(x^3 - a) = -27a^2$.

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

2.2. A decomposition

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

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

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
\[
\alpha_p(N) = \sum_{n \leq N} \nu_p(f_a(n)),
\]
where \( \nu_p(m) := \max(k \geq 0 : p^k \mid m) \).

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

Using the prime factorization of \( L_a(N) \) and \( P_a(N) \), we have
\[
\log L_a(N) = \log P_a(N) - \sum_{p \leq N} \alpha_p(N) \log p + \sum_{p \leq N} \beta_p(N) \log p
- \sum_{p > N} (\alpha_p(N) - \beta_p(N)) \log p,
\]
where we have separated out the contribution of primes \( p \leq N \), and the larger ones. We further break off the contribution of primes \( p \leq N \) which divide the discriminant \( D(a) = \text{disc}(f_a) \), by setting
\[
\text{Bad}_N(a) := \sum_{p \leq N, p \mid D(a)} \alpha_p(N) \log p,
\]
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 \( \text{Bad}_N \) and \( \Delta_N \geq 0 \) are both non-negative. We obtain the expression
\[
\log L_a(N) = \log P_a(N) + \sum_{p \leq N} \beta_p(N) \log p
- \text{Bad}_N(a) - \sum_{p \leq N, p \mid D(a)} \alpha_p(N) \log p - \Delta_N(a).
\]

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, none 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 \( n \leq N \),
\[ f_0(n) - a = n^d \left( 1 + \frac{c_{d-1}}{n} + \frac{c_{d-2}}{n^2} + \cdots - \frac{a}{n^d} \right). \]

Consider first the \( n \)'s satisfying \( N/\log N < n \leq N \), for which we use (recall that \(|a| \leq N^{d-1}\))
\[ \log \left( 1 + \frac{c_{d-1}}{n} + \frac{c_{d-2}}{n^2} + \cdots - \frac{a}{n^d} \right) = O\left( \frac{(\log N)^d}{N} \right), \]
so that
\[ \sum_{N/\log N < n \leq N} \log |f_0(n) - a| = \sum_{N/\log N < n \leq N} \left( d \log n + O\left( \frac{(\log N)^d}{N} \right) \right) = dN \log N + O(N). \]

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. \( \square \)

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
\[ \beta_p(N) = \max_{n \leq N} \max_{k \geq 0} (k : p^k | f_0(n) - a), \]
and since \( f_0(n) - a \neq 0 \) for all \( n \), if \( p^k | f_0(n) - a \neq 0 \), then
\[ k \leq \frac{\log |f_0(n) - a|}{\log p} \ll \frac{\log n + \log |a|}{\log p}. \]

Hence, since \(|a| \ll N^{d-1}\),
\[ \beta_p(N) \ll \frac{\log N}{\log p}, \]
and hence the contribution of primes \( p \leq N \) to (2.2) is
\[ (2.3) \quad \sum_{p \leq N} \beta_p(N) \log p \ll \sum_{p \leq N} \log N \ll N. \]
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 $p \nmid D(a) = \text{disc}(f_0(x) - a)$, we have

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

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

Consequently, we find that in (2.2),

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

Therefore we have proven Proposition 2.2.

3. Bounding $\text{Bad}_N$ almost surely

Recall that we defined

$$\text{Bad}_N(a) = \sum_{\substack{p \leq N \atop p \mid D(a)}} \log p \sum_{n \leq N} \#\{ k \geq 1 : p^k \mid f_0(n) - a \}$$

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

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

$$(\ast) = \frac{1}{\#\{|a| \leq T : f_0(x) - a \text{ is irreducible}\}} \sum_{\substack{|a| \leq T \atop f_0(x) - a \text{ irreducible}}} \bullet$$

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

$$(3.1) \quad (\ast) = \frac{1}{2T + O(\sqrt{T})} \sum_{|a| \leq T \atop f_0(x) - a \text{ irreducible}} \bullet$$

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

$$\text{Bad}_N \ll N \log \log N.$$  

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

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

where

$$B_1(a) = \sum_{\substack{p \leq N \atop p \mid D(a)}} \log p \#\{ n \leq N : f_0(n) = a \mod p \}$$
and

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

We will show that

\[ (3.2) \quad B_1(a) \ll N \log \log N \]

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

\[ \# \{ n \leq N : f_0(n) = a \text{ mod } p \} = \left( \frac{N}{p} + O(1) \right) \# \{ n \text{ mod } p : f_0(n) = a \text{ mod } p \} \ll \frac{N}{p} \rho(a;p), \]

where

\[ \rho(a;p) := \# \{ n \text{ mod } p : f_0(n) = a \text{ mod } p \}, \]

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

\[ B_1(a) \ll \sum_{p \leq N \atop p | D(a)} \log p \frac{N}{p} \rho(a;p) \ll N \sum_{p | D(a)} \frac{\log p}{p}. \]

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

\[ \sum_{p | k} \frac{\log p}{p} \leq \sum_{p | k \atop p \leq \log k} \frac{\log p}{p} + \sum_{p | k \atop p > \log k} \frac{\log p}{p} \ll \sum_{p \leq \log k} \frac{\log p}{p} + \frac{\log \log k}{\log k} \sum_{p | k} 1 \]

\[ \ll \log \log k + \frac{\log \log k}{\log k} \cdot \frac{\log k}{\log \log k} \ll \log \log k, \]

since the number of distinct prime divisors of \( k \) is \( \ll \log k/\log \log k \). \qed
Therefore
\[
\sum_{p | D(a)} \frac{\log p}{p} \ll \log \log |D(a)| \ll \log \log |a|,
\]
and we obtain
\[
B_1(a) \ll N \log \log |D(a)|.
\]

Next we bound the mean value of \(B_2(a)\):
\[
\langle B_2 \rangle = \frac{1}{2T + O(\sqrt{T})} \sum_{|a| \leq T} \sum_{\substack{p \leq N \\text{irreducible} \\text{ } p | D(a)}} \log p \sum_{k \geq 2} 1(f_0(n) = a \mod p^k).
\]

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 \(n \leq N\), then \(k \ll \log N/\log p\), so we restrict the summation to \(2 \leq k \ll \log N/\log p\). Moreover, given \(n\), the condition \(f_0(n) = a \mod p^k\) determines \(a\) modulo \(p^k\), so there are \(\ll T/p^k + 1\) choices for \(a\). Hence we may bound
\[
\langle B_2 \rangle \ll \frac{1}{T} \sum_{p \leq N} \log p \sum_{n \leq N} \sum_{2 \leq k \ll \log N/\log p} \left( \frac{T}{p^k} + 1 \right)
\]
\[
= \frac{N}{T} \sum_{p \leq N} \log p \sum_{2 \leq k \ll \log N/\log p} \left( \frac{T}{p^k} + 1 \right).
\]

We have
\[
\frac{N}{T} \sum_{p \leq N} \log p \sum_{2 \leq k \ll \log N/\log p} \frac{T}{p^k} \ll N \sum_{p \leq N} \log p \sum_{k \geq 2} \frac{1}{p^k} \ll N \sum_{p \leq N} \frac{\log p}{p^2} \ll N
\]
and
\[
\frac{N}{T} \sum_{p \leq N} \log p \sum_{2 \leq k \ll \log N/\log p} 1 \ll \frac{N}{T} \sum_{p \leq N} \log p \cdot \frac{\log N}{\log p} \ll \frac{N^2}{T}.
\]

Altogether we find
\[
\langle B_2 \rangle \ll N + \frac{N^2}{T},
\]
which is \(O(N)\) if \(T \geq N\).

\[\square\]

4. Averaging \(\Delta_N(a)\)

Let
\[
\Delta_N(a) = \sum_{p > N} \log p \left( \alpha_p(N) - \beta_p(N) \right).
\]

Then clearly \(\Delta_N \geq 0\), and we want to show:

**Proposition 4.1.** Assume that \(T \geq N \log N\), 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 \( \leq d - 1 \). 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 \( m, n \geq 1 \) and \( \max(m, n) > C_1 \), then \( G(m, n) \neq 0 \).

**Proof.** We have

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

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

\[ \frac{m^j - n^j}{m - n} = n^{j-1} + n^{j-2}m + \cdots + m^{j-1} \leq jn^{j-1}, \]

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 \( c_d = 1 \))

\[ G(m, n) \geq \frac{m^d - n^d}{m - n} - \sum_{j=2}^{d-1} |c_j| \frac{m^d - n^d}{m - n} - |c_1| > n^{d-1} - \sum_{j=1}^{d-1} |c_j| jn^{j-1}, \]

which is clearly positive once \( n \) is sufficiently large in terms of the coefficients \( c_1, \ldots, c_{d-1} \) of \( f_0 \). \( \square \)

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

**Proof.** We have, by definition,

\[ \alpha_p(N) = \sum_{n \leq N} \sum_{k \geq 1} 1(f_0(n) = a \mod p^k) = \sum_{k \geq 1} \#\{n \leq N : f_0(n) = a \mod p^k\}. \]

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 \( p \leq |f_a(n)| \ll N^d + |a| \ll N^d \) uniformly in \( |a| \leq T \) (recall \( T \leq N^d \)). 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:

\[ \#\{m \leq N : f_a(m) = f_a(n) \mod p\} \leq d. \]
Indeed, for any \( c \in \mathbb{Z}/p\mathbb{Z} \), the number of solutions \( m \mod p \) of \( f_a(m) = c \mod p \) is at most \( d \), and since \( p > N \), this certainly applies to those \( m \leq N \) 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 \( n \leq N \) is, because we assume \( f_a(n) \neq 0 \),

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

since we assume that \( |a| \leq T \) with \( \log T \ll \log N \).

Therefore

\[
\alpha_p(N) = \sum_{k \geq 1} \# \{ n \leq N : f(n) = 0 \mod p^k \} \leq \sum_{1 \leq k \ll O_d(1)} d = O_d(1)
\]
as claimed. \( \square \)

### 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

\[
\Delta_N(a) \ll \sum_{1 \leq m < n \leq N} \sum_{N < p \leq N^d \atop p \nmid f_0(m) - a, \ p \nmid (m,n)} \log p + O(\log N).
\]

**Proof.** We have \( \alpha_p(N) \neq \beta_p(N) \) if and only if there are two distinct integers \( m, n \leq N \) 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,

\[
\Delta_N(a) \ll \sum_{1 \leq m < n \leq N} \sum_{N < p \leq N^d \atop p \nmid f_0(m) - a, \ p \nmid (m,n)} \log p.
\]

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 \( 1 \leq n - m \leq N - 1 < p \)), we must have \( p \mid G(m,n) \). Thus

\[
\Delta_N(a) \ll \sum_{1 \leq m < n \leq N} \sum_{N < p \leq N^d \atop p \nmid f_0(m) - a} \log p.
\]

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

\[
\ll \sum_{m, n \leq C_1} \sum_{N < p \leq N^d} \log p \ll \log N \max_{m \leq C_1} \# \{ p > N : p \nmid f_0(m) \}.
\]
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

$$\Delta_N(a) \ll \sum_{1 \leq m < n \leq N} \sum_{N < p \leq N^d \atop p | f_0(m) \atop p | G(m, n)} \log p + O(\log |a|).$$

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

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

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

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

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

$$I \ll \sum_{1 \leq m < n \leq N} \sum_{N < p \leq N \log N \atop G(m, n) \neq 0 \atop p | G(m, n)} \frac{\log p}{p} + \sum_{1 \leq m < n \leq N} \sum_{N \log N < p \leq N^d \atop G(m, n) \neq 0 \atop p | G(m, n)} \frac{\log p}{p}.$$
We treat the sum over small primes by switching the order of summation:
\[
\sum_{1 \leq m < n \leq N} \sum_{p \mid G(m,n), N < p < N \log N} \frac{\log p}{p} \\
\leq \sum_{N < p < N \log N} \frac{\log p}{p} \#\{1 \leq m < n \leq N : G(m,n) = 0 \text{ mod } p\}.
\]

Now given \(m\), the congruence \(G(m,n) = 0 \text{ mod } p\) (if solvable) determines \(n \text{ mod } p\) up to \(d - 1\) possibilities, since \(G(m,n)\) is a monic polynomial of degree \(d - 1\) in \(n\), and since \(n \leq N < p\) means that \(n\) is determined as an integer up to \(d - 1\) possibilities. Hence
\[
\#\{1 \leq m < n \leq N : G(m,n) = 0 \text{ mod } p\} \leq (d - 1)N,
\]
and the sum over small primes is bounded by
\[
\ll \sum_{N < p < N \log N} \frac{\log p}{p} N = N\{\log(N \log N) + O(1)\} \sim N \log \log N
\]
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
\[
\sum_{1 \leq m < n \leq N} \sum_{N \log N < p < N^d} \frac{\log p}{p} \ll \frac{1}{N} \sum_{1 \leq m < n \leq N} \sum_{G(m,n) \neq 0} \#\{p > N \log N : p \mid G(m,n)\}.
\]

Now given \(1 \leq m < n \leq N\) 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
\[
\ll \frac{1}{N} \sum_{1 \leq m < n \leq N} (d - 1) \ll N.
\]

This gives \(I \ll N \log \log N\), and hence,
\[
\langle \Delta_N \rangle \ll N \log \log N,
\]
as claimed. \(\square\)

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:
\[
\frac{1}{\pi(x)} \sum_{p \leq x} \rho_f(p) = 1 + o_f(1).
\]
We write
\[ \rho_f(p) = 1 + \sigma_f(p), \]
where \( \sigma_f(p) \) is a fluctuating quantity, having mean zero.

Now fix
\[ f_0(x) = x^d + c_{d-1}x^{d-1} + \cdots + c_1x \in \mathbb{Z}[x], \]
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 \( 0 \leq \rho(a; p) \leq d \).

We write
\[ C_N(a) := \sum_{\substack{p \leq N \\atop p \mid \text{disc}(f_a)}} \frac{\log p}{p-1} \rho(a; p) = \sum_{\substack{p \leq N \\atop p \mid \text{disc}(f_a)}} \frac{\log p}{p} E_N(a) + D_N(a) + O(1), \]
where
\[ D_N(a) := \sum_{\substack{p \leq N \\atop p \mid \text{disc}(f_a)}} \frac{\log p}{p} \sigma(a; p) \quad \text{and} \quad E_N(a) := \sum_{\substack{p \leq N \\atop p \mid D(a)}} \frac{\log p}{p}. \]

By Mertens’ theorem,
\[ \sum_{\substack{p \leq N \\atop p \mid \text{disc}(f_a)}} \frac{\log p}{p} = \log N + O(1). \]

The contribution \( E_N(a) \) of primes dividing the discriminant \( D(a) = \text{disc}(f_0(x) - a) \) can be bounded individually, for \( |a| \leq T \ll N^d \), using Lemma 3.2 (assuming \( D(a) \neq 0 \)):
\[ E_N(a) = \sum_{\substack{p \leq N \\atop p \mid D(a)}} \frac{\log p}{p} \leq \sum_{\substack{p \leq N \\atop p \mid D(a)}} \frac{\log p}{p} \ll \log \log |D(a)|. \]

Since \( D(a) \) is a polynomial of degree \( d - 1 \) in \( a \), and \( |a| \leq T \ll N^d \), we find
\[ \sum_{\substack{p \leq N \\atop p \mid D(a)}} \frac{\log p}{p} \ll \log \log N, \]
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 \( T \geq N \log N \), the second moment of \( D_N(a) \) satisfies
\[ \langle |D_N|^2 \rangle \ll 1. \]
Using the triangle inequality and Cauchy–Schwartz, we obtain:

**Proposition 5.2.**
\[ \langle |C_N - \log N|^2 \rangle \ll (\log \log N)^2. \]

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

**Proposition 5.3.** For almost all \(|a| \leq T\) (with \(N \log N \leq T \ll N^{d-1}\),
\[ C_N(a) = \log N + O(\log \log N). \]

### 5.2. Proof of Proposition 5.1

**Proof.** Expanding, we have
\[ \langle (D_N)^2 \rangle = \sum_{p \leq N} \sum_{q \leq N} \frac{\log p \log q}{pq} \langle \sigma(a; p) \sigma(a; q) \rangle. \]

The diagonal contribution \(p = q\) gives
\[ \sum_{p \leq N} \frac{(\log p)^2}{p^2} \langle \sigma(a; p)^2 \rangle. \]

Now note that
\[ -1 \leq \sigma(a; p) \leq d - 1 \]

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 \(0 \leq \rho(a; p) \leq d\) and so \(-1 \leq \sigma(a; p) \leq d - 1\). Thus we obtain a bound for the diagonal sum:
\[ \sum_{p \leq N} \frac{(\log p)^2}{p^2} \langle \sigma(a; p)^2 \rangle \ll \sum_{p \leq N} \frac{(\log p)^2}{p^2} \ll 1. \]

For the off-diagonal terms, we use:

**Lemma 5.4.** For distinct primes \(p \neq q\),
\[ |\langle \sigma(\bullet; p) \sigma(\bullet; q) \rangle| \leq \frac{\sqrt{pq \log(pq)}}{T} + \frac{1}{\sqrt{T}}. \]

Therefore, given Lemma 5.4, we obtain
\[ \sum_{p \neq q \leq N} \frac{\log p \log q}{pq} |\langle \sigma(a; p) \sigma(a; q) \rangle| \ll \sum_{p \neq q \leq N} \frac{\log p \log q}{pq} \left( \frac{\sqrt{pq \log(pq)}}{T} + \frac{1}{\sqrt{T}} \right) \]
\[ \ll \frac{1}{T} \left( \sum_{p \leq N} \frac{\log p}{\sqrt{p}} \right)^2 + \frac{1}{\sqrt{T}} \left( \sum_{p \leq N} \frac{\log p}{p} \right)^2 \]
\[ \ll \frac{N \log N}{T} + \frac{(\log N)^2}{\sqrt{T}}, \]

which is \(O(1)\) if \(T \geq N \log N\), proving Proposition 5.1. \(\square\)
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

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

and so

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

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

$$
\rho(a; p) = \#\{x \mod p : f_0(x) - a = 0 \mod p\} = \sum_{x \mod p} \frac{1}{p} \sum_{t \mod p} e\left(\frac{tf_0(x) - a}{p}\right).
$$

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

$$
(5.1) \quad \sigma(a; p) = \frac{1}{p} \sum_{t \neq 0 \mod p} e\left(-\frac{at}{p}\right) \sum_{x \mod p} e\left(\frac{tf_0(x)}{p}\right),
$$

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

$$
S_{f_0}(b, n) := \sum_{x \mod n} e\left(\frac{bf_0(x)}{n}\right).
$$

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

$$
\frac{1}{2T + O(\sqrt{T})} \sum_{|a| \leq T} \sigma(a; p) \sigma(a; q)
$$

$$
= \frac{1}{2T + O(\sqrt{T})} \frac{1}{p^q} \sum_{0 \neq t \mod p} \sum_{|a| \leq T} e\left(-a\left(\frac{t}{p} + \frac{s}{q}\right)\right) S_{f_0}(t, p) S_{f_0}(s, q).
$$

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 $p$,

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

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
\[
|\langle \sigma(\bullet; p) \sigma(\bullet; q) \rangle| \ll_d \frac{1}{T^{\sqrt{pq}}} \sum_{\substack{0 \neq t \mod p \\ 0 \neq s \mod q}} \sum_{|a| \leq T} e\left(-a\left(\frac{t}{p} + \frac{s}{q}\right)\right) + O\left(\frac{1}{\sqrt{T}}\right)
\]
\[
= \frac{1}{T^{\sqrt{pq}}} \sum_{m \mod pq \atop \gcd(m,pq)=1} \sum_{|a| \leq T} e\left(-\frac{am}{pq}\right) + O\left(\frac{1}{\sqrt{T}}\right),
\]
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 \( q \)), \( tq + sp \mod pq \) covers all invertible residues modulo \( pq \) exactly once.

We sum the geometric progression
\[
\left| \sum_{|a| \leq T} e\left(-\frac{am}{pq}\right) \right| \ll \min\left(T, \|a\|^{-1}\right)
\]
where \( \|a\| = \text{dist}(a, \mathbb{Z}) \). We may take \( 1 \leq m < pq/2 \), and then the bound is
\[
\ll pq/m.
\]
This gives
\[
|\langle \sigma(\bullet; p) \sigma(\bullet; q) \rangle| \ll \frac{1}{T^{\sqrt{pq}}} \sum_{1 \leq m \leq pq/2 \atop \gcd(m,pq)=1} \frac{pq}{m} + O\left(\frac{1}{\sqrt{T}}\right) \ll \frac{\sqrt{pq} \log(pq)}{T} + O\left(\frac{1}{\sqrt{T}}\right),
\]
proving Lemma 5.4. \( \square \)

References


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

Ze`ev Rudnick: School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel.
E-mail: rudnick@tauex.tau.ac.il

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.