Linear independent prime numbers

Orges Leka and ChatGPT 5.1

November 19, 2025

Abstract

We study the arithmetic and linear-algebraic structure of the primes through the successor map

$$\Phi(p) \;=\; \min\{\, q \text{ prime} : q \equiv 1 \pmod{p} \,\},$$

and the associated valuation vectors $\varphi(q) = (v_2(q-1), v_3(q-1), \dots)$. A prime q is called *linearly independent* (LI) if $\varphi(q)$ does not lie in the $\mathbb{Q}$-span of the vectors $\varphi(p)$ for smaller primes $p < q$; otherwise it is *linearly dependent* (LD).

On the dynamical side, we show that the primes decompose disjointly into infinite successor chains

$$C(d) = \{d, \Phi(d), \Phi^2(d), \dots\},$$

indexed by LD primes d: the starting points are precisely the LD primes, and the successors $\Phi^k(d)$ for $k \geq 1$ are LI. This gives a canonical partition of the prime set by Φ-orbits. On the linear-algebraic side, we prove that the vectors $\varphi(\Phi(p))$ form, in an appropriate sense, a $\mathbb{Z}$-basis for the valuation data of all primes: every $\varphi(q)$ admits a unique integral expansion in terms of $\varphi(\Phi(p))$, and correspondingly $q-1$ admits a unique multiplicative factorization in terms of the numbers $\Phi(p) - 1$.

We then investigate the distribution of LI and LD primes. Using only the elementary inequality $\Phi(p) > 2p$ and the prime number theorem, we show that the number $\nu(x)$ of LD primes up to x satisfies $\nu(x) \gg x/(\log x)^2$, so there are infinitely many LD primes. Under the additional hypothesis $\Phi(p) \leq p^2 - 1$ (a Linnik-type assumption), we obtain a much sharper picture: successor chains grow so rapidly that they have length only $O(\log\log x)$ below height x, and this forces

$$\nu(x) \;=\; \pi(x) - \xi(x) \;=\; \pi(x) - O\!\left(\frac{\sqrt{x}}{\log x}\right),$$

where $\xi(x)$ counts LI primes and $\pi(x)$ is the prime counting function. Thus LD primes form a set of asymptotic density 1 among all primes, while LI primes constitute a comparatively thin but structurally crucial subset.

This leads to a striking but ultimately coherent dichotomy. From the vector-space viewpoint, LI primes are the "basis": their valuation vectors carry all independent information, and every LD prime is a $\mathbb{Z}$-linear combination of LI ones. From the dynamical viewpoint, however, LD primes are the "sources": they are the unique starting points of successor chains, from which LI primes emerge as successors. The apparent paradox—that a sparse set of LI primes generates, in the linear sense, almost all primes, while LD primes themselves make up nearly 100% of the primes—disappears once one clearly separates these two roles. The successor chains control the dynamical genealogy of primes, while the valuation vectors organize their multiplicative exponents into a highly redundant but rigid linear structure.

Contents

1 Introduction

Let $(p_k)_{k\geq 1}$ denote the increasing sequence of prime numbers,

$$p_1 = 2, \quad p_2 = 3, \quad p_3 = 5, \quad p_4 = 7, \quad \ldots.$$

For each prime p_k we consider the exponents of the prime factors in the factorization of $p_k - 1$. Collecting these exponents for the earlier primes $p_1, \ldots, p_{k-1}$ gives a natural integer vector associated to p_k.

This leads to a notion of "linear independence" for primes: roughly speaking, p_k is called *linearly independent* if the exponent vector of $p_k - 1$ is not in the linear span of the corresponding vectors for earlier primes. In this paper we formalize this definition, relate it to the OEIS sequence A071349, and show that there are infinitely many linearly independent primes.

The proof of infinitude given here uses Dirichlet's theorem on primes in arithmetic progressions. At the end of Section 5 we briefly comment on the question whether there is a proof avoiding Dirichlet's theorem or comparable analytic input.

We then attach to the linearly independent primes a natural $\mathbb{Z}$-lattice generated by their exponent vectors and study its basic invariants. In particular, we show that this "phi-lattice" is an odd unimodular lattice and formulate a theta-series conjecture suggesting that it might be isometric to a standard cubic lattice. Using the classification of unimodular lattices in small rank, we verify this conjecture when the rank is at most 7.

2 Related Work and Context

The framework introduced in this paper, which connects the p-adic valuations of $p-1$ to linear algebra and lattice theory, builds upon several classical streams of number theory. While the specific synthesis and the resulting conjectures (e.g., Conjecture 8.1) appear to be new, the components themselves have deep roots in the literature.

2.1 The Multiplicative Structure of $p-1$

The study of the prime factors of $p-1$ is a cornerstone of classical number theory. It forms the basis for:

- **Artin's Conjecture** on primitive roots, which is deeply reliant on the factorization of $p-1$.
- **Pratt certificates** for primality, which require a full factorization of $p-1$.
- **Factoring algorithms**, most notably Pollard's $p-1$ algorithm [1], which is efficient precisely when $p-1$ is smooth (i.e., composed of small prime factors).

While this classical work focuses on the *properties* of the set of prime factors of $p-1$ (e.g., its largest element), our work introduces a novel *structural* and *cumulative* perspective. We ask not just what the factors are, but whether their exponent vector $\varphi(p_k)$ represents "new information" by lying outside the $\mathbb{Q}$-span of all preceding vectors. The primes that are "smooth" (and thus useful for Pollard's algorithm, such as Fermat primes) are, in our framework, highly linearly *dependent*.

2.2 A Parallel with Dirichlet's Unit Theorem

The strongest conceptual parallel to our construction comes from algebraic number theory, specifically the proof of Dirichlet's Unit Theorem [2].

1. **Classical Construction (Dirichlet):** To understand the multiplicative structure of the group of units $\mathcal{O}_K^\times$ in a number field K, one uses a *logarithmic embedding*. This map sends a unit $\varepsilon \in \mathcal{O}_K^\times$ to a vector of its real and complex log-valuations, e.g.,
$$L(\varepsilon) := \big(\log|\sigma_1(\varepsilon)|, \ldots, \log|\sigma_r(\varepsilon)|\big) \in \mathbb{R}^r.$$

The image $L(\mathcal{O}_K^\times)$ forms a lattice in a hyperplane, and the theorem determines its *rank*.

2. **This Paper's Construction:** We, too, use a valuation-based embedding. We map an integer $n = p_k - 1$ to a vector of its p-adic valuations:

$$\varphi(p_k) := (v_{p_1}(p_k - 1), \ldots, v_{p_{k-1}}(p_k - 1)) \in \mathbb{Q}^{k-1}.$$

We then form a lattice Λ_N from the $\mathbb{Z}$-span of these vectors.

The analogy is striking: Dirichlet's theorem uses archimedean valuations to find the *rank* of a structural lattice (the units). Our work uses non-archimedean (p-adic) valuations to study the *isometry class* of a cumulatively-defined lattice Λ_N (Question 7.2). Our framework can thus be seen as applying the "log-embedding" philosophy not to the structural units of a field, but to the combinatorial sequence of integers $p - 1$.

2.3 Record-Setters and Arithmetic Functions

The first major result of this paper, Theorem 4.6, proves that our algebraic definition of linear independence is exactly equivalent to the analytic definition of k being a record-setter for the sequence $g(k) = \gcd(q(k), f(k))$. This sequence is noted in the OEIS as A071349 [7].

The study of such arithmetic functions, particularly the gcd of n and $\phi(n)$ (Euler's totient function), is also a classical topic, with notable contributions from Erdős and Pomerance [3]. However, the discovery that the record-setters for this specific $g(k)$ sequence correspond perfectly to a change in rank of a p-adic valuation matrix appears to be a new connection, bridging the analytic/combinatorial behavior of $g(k)$ with the algebraic structure of the φ-vectors.

3 Definitions

3.1 Notation and basic objects

Let $(p_k)_{k\geq 1}$ denote the k-th prime, so $p_1 = 2$, $p_2 = 3$, and so on.

For an integer $n \geq 1$ and a prime p, let $v_p(n)$ be the usual *p-adic valuation* of n, i.e. the unique integer $e \geq 0$ such that

$$p^e \mid n \quad \text{and} \quad p^{e+1} \nmid n.$$

For each integer $n \geq 1$ we consider the $n \times n$ matrix $E^{(n)} = (e_{i,k})_{1\leq i,k\leq n}$ with entries

$$e_{i,k} := v_{p_i}(p_k - 1).$$

We view the k-th column of $E^{(n)}$ as a vector in $\mathbb{Q}^n$ (or $\mathbb{R}^n$), and write

$$V_k^{(n)} := \begin{pmatrix} v_{p_1}(p_k - 1) \\ v_{p_2}(p_k - 1) \\ \vdots \\ v_{p_n}(p_k - 1) \end{pmatrix}.$$

Note that if p_i is a prime divisor of $p_k - 1$, then necessarily $p_i < p_k$, so $i < k$. In particular, the diagonal entries of $E^{(n)}$ are all zero:

$$v_{p_k}(p_k - 1) = 0.$$

3.2 Phi-vectors and linear independence

We will work with the following slight variant of $V_k^{(n)}$.

Definition 3.1 (Phi-vector of a prime)**.** For $k \geq 1$ we define the *phi-vector* of p_k by

$$\varphi(p_k) := (v_{p_1}(p_k - 1), v_{p_2}(p_k - 1), \ldots, v_{p_{k-1}}(p_k - 1)) \in \mathbb{Z}^{k-1}.$$

Equivalently, $\varphi(p_k)$ is the k-th column of $E^{(k-1)}$.

We now define linear independence for primes in terms of these vectors.

Definition 3.2 (Linearly independent prime)**.** For $k \geq 1$ we call the prime p_k *linearly independent* if $\varphi(p_k)$ does *not* lie in the $\mathbb{Q}$-linear span of the vectors

$$\varphi(p_1), \varphi(p_2), \ldots, \varphi(p_{k-1}).$$

Otherwise we call p_k (or its index k) *linearly dependent*.

By convention the first prime $p_1 = 2$ is linearly independent, since there are no earlier vectors.

It is often convenient to phrase this in terms of ranks of matrices.

Remark 3.3. Fix $n \geq 1$ and consider the matrix $E^{(n)}$. Its k-th column $V_k^{(n)}$ is the extension of $\varphi(p_k)$ to length n by adding zeros in positions $i \geq k$. For $k \leq n$ the following are equivalent:

1. p_k is linearly independent.
2. The column $V_k^{(n)}$ is not in the span of $V_1^{(n)}, \ldots, V_{k-1}^{(n)}$.
3. The rank of $E^{(k)}$ is strictly larger than the rank of $E^{(k-1)}$.

In practice we will often work with the columns $V_k^{(n)}$ and the matrices $E^{(n)}$, because they have a convenient triangular structure, but all statements can be translated back to the phi-vectors $\varphi(p_k)$.

3.3 GCD record indices and OEIS A071349

We now introduce the sequence on the GCD of the primorial and its "totient part", as in the OEIS entry A071349.

Definition 3.4 (Primorial and totient product)**.** For $n \geq 1$ define

$$q(n) := \prod_{i=1}^{n} p_i \quad \text{and} \quad f(n) := \prod_{i=1}^{n} (p_i - 1).$$

We also set

$$g(n) := \gcd(q(n), f(n)).$$

The OEIS sequence A071349 consists of those $n \geq 1$ for which $g(n)$ sets a new record, i.e. for which $g(n) > g(k)$ for all $k < n$.

The following simple observation will be important.

Lemma 3.5. *For each $n \geq 1$ we have*

$$g(n) = \prod_{\substack{1 \leq i \leq n \\ \exists\, k \leq n : p_i \mid (p_k - 1)}} p_i,$$

and each prime p_i appears with exponent either 0 or 1 in $g(n)$. In particular the sequence $g(n)$ is non-decreasing in n.

Proof. For each prime p_i with $i \leq n$ we have $v_{p_i}(q(n)) = 1$, since $q(n)$ is the product of distinct primes $p_1, \ldots, p_n$. On the other hand

$$v_{p_i}(f(n)) = v_{p_i}\left(\prod_{k=1}^{n}(p_k - 1)\right) = \sum_{k=1}^{n} v_{p_i}(p_k - 1).$$

Thus

$$v_{p_i}(g(n)) = \min\bigl(v_{p_i}(q(n)), v_{p_i}(f(n))\bigr) = \begin{cases} 1, & \text{if } v_{p_i}(f(n)) \geq 1, \\ 0, & \text{otherwise.} \end{cases}$$

The condition $v_{p_i}(f(n)) \geq 1$ is equivalent to $p_i \mid (p_k - 1)$ for some $k \leq n$, and the claimed product formula follows. The exponent of each p_i in $g(n)$ is thus either 0 or 1, and since $v_{p_i}(f(n))$ is non-decreasing in n, each $v_{p_i}(g(n))$ is non-decreasing, hence so is $g(n)$. □

As a consequence, the condition "$g(n)$ sets a new record" is equivalent to $g(n) > g(n-1)$.

4 Equality of the two characterizations

In this section we show that the indices of linearly independent primes are exactly the indices n where $g(n)$ sets a new record, i.e. those listed in OEIS A071349.

4.1 First occurrence indices and the rank of $E^{(n)}$

For each $i \geq 1$ we introduce the first occurrence index of p_i as a divisor of some $p_k - 1$.

Definition 4.1 (First occurrence index)**.** For $i \geq 1$ define

$$t(i) := \min\{\, k \geq 1 : v_{p_i}(p_k - 1) \geq 1 \,\},$$

if this set is non-empty, and set $t(i) := \infty$ otherwise.

Note that if $p_i \mid (p_k - 1)$ then $p_i < p_k$ and hence $i < k$, so in fact $t(i) \geq i + 1$ whenever $t(i) < \infty$.

For fixed $n \geq 1$ consider the matrix $E^{(n)} = (e_{i,k})_{1 \leq i,k \leq n}$ with $e_{i,k} = v_{p_i}(p_k - 1)$. We now relate the rank of $E^{(n)}$ to the first occurrence indices $t(i)$.

Proposition 4.2. *For each $n \geq 1$ we have*

$$\operatorname{rank} E^{(n)} = \#\{\, i \in \{1, \ldots, n\} : t(i) \leq n \,\}.$$

Proof. Let

$$I_n := \{\, i \in \{1, \ldots, n\} : t(i) \leq n \,\}, \quad r(n) := \#I_n.$$

Lower bound. List the elements of I_n as $i_1, \ldots, i_{r(n)}$ in such a way that

$$t(i_1) < t(i_2) < \cdots < t(i_{r(n)}).$$

Consider the $r(n) \times r(n)$ submatrix M of $E^{(n)}$ with rows indexed by $i_1, \ldots, i_{r(n)}$ and columns indexed by $t(i_1), \ldots, t(i_{r(n)})$.

By definition of $t(i_\ell)$, in row i_ℓ all entries in columns $k < t(i_\ell)$ are zero, and the entry in column $t(i_\ell)$ is $v_{p_{i_\ell}}(p_{t(i_\ell)} - 1) \geq 1$. Since $t(i_1) < \cdots < t(i_{r(n)})$, this means precisely that M is a triangular matrix (upper or lower, depending on convention) with nonzero diagonal entries. Hence $\det M \neq 0$ and M has full rank $r(n)$.

Therefore $\operatorname{rank} E^{(n)} \geq r(n)$.

Upper bound. If $i \notin I_n$, i.e. $t(i) > n$, then by definition $v_{p_i}(p_k - 1) = 0$ for all $k \leq n$, hence the i-th row of $E^{(n)}$ is the zero row. Thus $E^{(n)}$ has at most $r(n)$ nonzero rows, so its rank is at most $r(n)$.

Combining the two bounds shows $\operatorname{rank} E^{(n)} = r(n) = \#I_n$, as claimed. □

4.2 Rank growth and linear independence

We now express linear independence of p_n in terms of the first occurrence indices $t(i)$.

Lemma 4.3. *For each $n \geq 1$ the following are equivalent:*

1. *p_n is linearly independent.*
2. *The rank of $E^{(n)}$ is strictly larger than the rank of $E^{(n-1)}$.*
3. *There exists $i \in \{1, \ldots, n\}$ with $t(i) = n$.*

Proof. The equivalence of (1) and (2) is the standard linear algebra fact that adding a column to a matrix increases the rank if and only if the new column is not in the span of the previous columns.

For the equivalence of (2) and (3), note that by Proposition 4.2 we have

$$\operatorname{rank} E^{(n)} = \#\{\, i \leq n : t(i) \leq n \,\},$$

and similarly

$$\operatorname{rank} E^{(n-1)} = \#\{\, i \leq n-1 : t(i) \leq n-1 \,\}.$$

The difference $\operatorname{rank} E^{(n)} - \operatorname{rank} E^{(n-1)}$ is thus exactly the number of indices $i \leq n$ with $t(i) = n$. Therefore the rank increases from $n-1$ to n if and only if there exists at least one i with $t(i) = n$. □

4.3 Comparison with the GCD record indices

We now connect the first occurrence indices $t(i)$ with the gcd sequence $g(n) = \gcd(q(n), f(n))$.

Proposition 4.4. *For each $n \geq 1$ we have*

$$g(n) = \prod_{\substack{1 \leq i \leq n \\ t(i) \leq n}} p_i.$$

Equivalently, the set of prime divisors of $g(n)$ is exactly

$$\{p_i : 1 \leq i \leq n, \quad t(i) \leq n\}.$$

Proof. By definition of $t(i)$ we have $t(i) \leq n$ if and only if there exists $k \leq n$ such that $p_i \mid (p_k - 1)$. Thus the set on the right-hand side is exactly the set described in Lemma 3.5. Together with the fact that each p_i appears with exponent 0 or 1 in $g(n)$, this implies the claimed product formula. □

Combining this with Lemma 3.5 we obtain:

Corollary 4.5. *The sequence $(g(n))_{n\geq 1}$ is non-decreasing, and*

$$g(n) > g(n-1) \quad \Longleftrightarrow \quad \exists\, i \in \{1, \ldots, n\} \text{ with } t(i) = n.$$

Proof. Non-decreasingness was already noted in Lemma 3.5. Since each prime p_i appears in $g(n)$ with exponent 0 or 1, the value of $g(n)$ increases from $n-1$ to n if and only if at least one new prime divisor appears, i.e. if and only if there is i such that $t(i) = n$. □

Now we can state and prove the main equality.

Theorem 4.6 (Equality of characterizations)**.** *For each $n \geq 1$ the following are equivalent:*

1. *The prime p_n is linearly independent.*
2. *The rank of $E^{(n)}$ is strictly larger than the rank of $E^{(n-1)}$.*
3. *$g(n) > g(n-1)$.*
4. *$g(n) > g(k)$ for all $k < n$.*

In particular, the indices of linearly independent primes coincide with the indices listed in OEIS A071349.

Proof. The equivalence (1) ⇔ (2) is part of Lemma 4.3. The equivalence (2) ⇔ (3) follows from Lemma 4.3 and Corollary 4.5, since both conditions are equivalent to the existence of some i with $t(i) = n$.

Finally, the equivalence (3) ⇔ (4) follows from the fact that $g(n)$ is non-decreasing in n (Lemma 3.5). If $g(n) > g(n-1)$ then automatically $g(n) > g(k)$ for all $k < n$, and conversely if $g(n)$ is strictly larger than all previous values then it is in particular larger than $g(n-1)$. □

5 Infinitely many linearly independent primes

In this section we show that there exist infinitely many linearly independent primes. The argument uses Dirichlet's theorem on primes in arithmetic progressions.

5.1 Dirichlet's theorem

We state the standard form of Dirichlet's theorem that we will use.

Theorem 5.1 (Dirichlet)**.** *Let a and q be coprime positive integers. Then there exist infinitely many primes p such that*

$$p \equiv a \pmod{q}.$$

In particular, for each integer $q \geq 2$ there exist infinitely many primes p with $p \equiv 1 \pmod{q}$.

5.2 Infinitude of linearly independent primes

We now prove the desired infinitude result.

Theorem 5.2. *There exist infinitely many linearly independent primes. Equivalently, there exist infinitely many indices n such that p_n is linearly independent.*

Proof. Suppose for contradiction that only finitely many primes are linearly independent. Then there exists an integer N such that for all $n > N$ the prime p_n is linearly dependent, i.e. $\varphi(p_n)$ lies in the span of $\varphi(p_1), \ldots, \varphi(p_N)$.

Consider the finite set of integers

$$p_1 - 1,\ p_2 - 1,\ \ldots,\ p_N - 1.$$

Let R be the largest prime divisor of the product

$$M := \prod_{k=1}^{N} (p_k - 1).$$

Then every prime divisor of any $p_k - 1$ with $k \leq N$ is at most R.

Choose a prime $q > R$. By Dirichlet's theorem (Theorem 5.1) there exist infinitely many primes p such that

$$p \equiv 1 \pmod{q}.$$

In particular we can choose such a prime p with $p > q$ and $p > p_N$. Since $p \equiv 1 \pmod{q}$ we have $q \mid (p-1)$, so the exponent $v_q(p-1)$ is at least 1.

On the other hand, for each $k \leq N$ we have $q > R$, so q is not a prime divisor of $p_k - 1$. Therefore $v_q(p_k - 1) = 0$ for $k \leq N$.

Now consider the phi-vector $\varphi(p)$ of the prime p:

$$\varphi(p) = \big(v_{p_1}(p-1), \ldots, v_{p_m}(p-1)\big),$$

where m is the index of p in the sequence of primes. Since $q < p$ the prime q appears among the earlier primes $p_1, \ldots, p_{m-1}$; let j be the index such that $p_j = q$.

Then the j-th coordinate of $\varphi(p)$ is $v_q(p-1) \geq 1$, whereas the j-th coordinate of each $\varphi(p_k)$ with $k \leq N$ is $v_q(p_k - 1) = 0$. Consequently the vector $\varphi(p)$ cannot lie in the $\mathbb{Q}$-linear span of $\varphi(p_1), \ldots, \varphi(p_N)$, because any linear combination of those vectors has j-th coordinate equal to 0, while $\varphi(p)$ has j-th coordinate nonzero.

Thus p is linearly independent, contradicting the assumption that all primes with index $> N$ are linearly dependent. This contradiction shows that there must be infinitely many linearly independent primes. □

5.3 A question about Dirichlet-free proofs

The proof of Theorem 5.2 relied on Dirichlet's theorem to guarantee, for a given large prime q, the existence of primes p with $p \equiv 1 \pmod{q}$.

Question 5.3. Is there a proof of Theorem 5.2 that does not use Dirichlet's theorem (or any other result of comparable strength on the distribution of primes in arithmetic progressions)?

At present the author is not aware of such a proof. It would be interesting to know whether a more elementary argument exists, or whether the existence of infinitely many linearly independent primes is in some sense equivalent to a Dirichlet-type statement.

6 Examples of phi-vectors

In this section we give a small table of $\varphi(p)$ for some primes p. We list the first few primes and the exponent vectors of $p-1$ with respect to the primes $2, 3, 5, 7, 11$.

Definition 6.1 (Truncated phi-vector)**.** Fix the list of primes $(2, 3, 5, 7, 11)$. For a prime p we define the truncated phi-vector

$$\varphi_{(2,3,5,7,11)}(p) := \big(v_2(p-1),\, v_3(p-1),\, v_5(p-1),\, v_7(p-1),\, v_{11}(p-1)\big).$$

p	factorization of $p-1$	$\varphi_{(2,3,5,7,11)}(p)$
2	1	$(0,0,0,0,0)$
3	2	$(1,0,0,0,0)$
5	2^2	$(2,0,0,0,0)$
7	$2\cdot 3$	$(1,1,0,0,0)$
11	$2\cdot 5$	$(1,0,1,0,0)$
13	$2^2\cdot 3$	$(2,1,0,0,0)$
17	2^4	$(4,0,0,0,0)$
19	$2\cdot 3^2$	$(1,2,0,0,0)$
23	$2\cdot 11$	$(1,0,0,0,1)$

Table 1: Some values of $\varphi_{(2,3,5,7,11)}(p)$ for small primes p.

For instance, for $p = 23$ we have $p-1 = 22 = 2\cdot 11$, so $v_2(22) = 1$, $v_{11}(22) = 1$, and the other valuations are zero, leading to the vector $\varphi_{(2,3,5,7,11)}(23) = (1,0,0,0,1)$.

These examples illustrate how new prime factors of $p-1$ create new nonzero coordinates in the phi-vectors, which in turn is related to the linear independence phenomenon studied above.

7 Theta series of the phi-lattices and sums of squares

Recall that for a fixed integer $N \geq 2$ we write

$$\mathcal{P}(N) = \{\, p_1, p_2, \ldots, p_m \,\}$$

for the set of primes $\leq N$ in increasing order ($p_1 = 2$, $p_2 = 3$, ...). For each prime $p \in \mathcal{P}(N)$ we define the *phi-vector*

$$\varphi_N(p) := \big(v_{p_1}(p-1), v_{p_2}(p-1), \ldots, v_{p_m}(p-1)\big) \in \mathbb{Z}^m,$$

where $v_q(\cdot)$ denotes the q-adic valuation. A prime $p_k \in \mathcal{P}(N)$ is called *linearly independent (up to N)* if $\varphi_N(p_k)$ does not lie in the $\mathbb{Q}$-span of $\{\varphi_N(p_j) : p_j < p_k\}$.

Let

$$\{q_1, \ldots, q_r\} \subset \mathcal{P}(N)$$

be the set of linearly independent primes up to N (in increasing order), and set

$$b_j := \varphi_N(q_j) \in \mathbb{Z}^m, \qquad \xi(N) := r.$$

Definition 7.1 (The lattice Λ_N)**.** The *phi-lattice* Λ_N is the $\mathbb{Z}$-span

$$\Lambda_N := \mathrm{Span}_{\mathbb{Z}}\{b_1, \ldots, b_r\} \subseteq \mathbb{Z}^m \subset \mathbb{R}^m,$$

equipped with the standard Euclidean inner product $\langle x, y\rangle = \sum_{i=1}^m x_i y_i$ and norm $\|x\|^2 = \langle x, x\rangle$. Let G_N be the $r \times r$ Gram matrix $(\langle b_i, b_j\rangle)_{i,j}$.

The theta series of Λ_N is

$$\Theta_{\Lambda_N}(q) := \sum_{v \in \Lambda_N} q^{\|v\|^2} = \sum_{n=0}^{\infty} a_N(n)\, q^n,$$

where $a_N(n)$ counts the vectors $v \in \Lambda_N$ with $\|v\|^2 = n$.

On the other hand, for $m \geq 1$ let $s_m(n)$ denote the number of ways of writing n as a sum of m squares,

$$s_m(n) := \#\{(x_1, \ldots, x_m) \in \mathbb{Z}^m : x_1^2 + \cdots + x_m^2 = n\}.$$

The classical theta series of the standard cubic lattice $\mathbb{Z}^m$ is then

$$\Theta_{\mathbb{Z}^m}(q) = \sum_{n=0}^{\infty} s_m(n)\, q^n.$$

7.1 The theta-series conjecture

Motivated by numerical experiments (see below), we formulate the following question.

Question 7.2 (Theta-series conjecture)**.** For each $N \geq 2$, is it true that

$$\Theta_{\Lambda_N}(q) = \sum_{n=0}^{\infty} s_{\xi(N)}(n)\, q^n = \Theta_{\mathbb{Z}^{\xi(N)}}(q)\,?$$

Equivalently, is the lattice Λ_N isometric to the standard cubic lattice $\mathbb{Z}^{\xi(N)}$?

Below we give some structural properties of Λ_N and explain why an argument based only on minimal vectors is not sufficient in general.

7.2 Unimodularity of Λ_N

Proposition 7.3. *For every $N \geq 2$ the lattice Λ_N is an odd unimodular lattice of rank $\xi(N)$; in particular*

$$\det(G_N) = 1.$$

Sketch of proof. Consider the $m \times r$ matrix

$$B_N := [\, b_1 \ \cdots \ b_r \,]$$

with columns $b_j = \varphi_N(q_j)$. The lattice Λ_N is exactly the image $B_N \mathbb{Z}^r \subset \mathbb{Z}^m$. The Gram matrix is $G_N = B_N^\top B_N$, so $\det(G_N) = (\operatorname{vol} \Lambda_N)^2$ is the square of the covolume of Λ_N in $\mathbb{R}^r$.

From the construction of the linearly independent primes one sees that we can order a suitable subset of the prime indices $i_1, \ldots, i_r \in \{1, \ldots, m\}$ such that the $r \times r$ submatrix of B_N formed by rows $i_1, \ldots, i_r$ is upper triangular with diagonal entries all equal to 1. (Informally: for each new linearly independent prime q_j there is some new prime divisor of $q_j - 1$ that has not appeared in any $q_k - 1$ with $k < j$, and this gives a 1 on a new diagonal position; compare the triangular argument used earlier for the rank computation.)

In particular this $r \times r$ minor has determinant ± 1, hence the greatest common divisor of all $r \times r$ minors of B_N is 1. This implies that Λ_N has index 1 in the full lattice generated by these rows, so Λ_N is unimodular; equivalently, $\det(G_N) = 1$.

The fact that Λ_N is *odd* (i.e. not all norms are even) follows already from the vector $b_1 = \varphi_N(3)$, which is $(1, 0, \ldots, 0)$ and has squared norm 1. □

7.3 Minimal vectors and low-dimensional cases

The minimal norm of Λ_N is $\min\{\|v\|^2 : v \in \Lambda_N \setminus \{0\}\}$.

Lemma 7.4. *For every $N \geq 3$ the minimal norm of Λ_N is 1.*

Proof. For $p = 3$ we have $3 - 1 = 2$, so

$$\varphi_N(3) = (1, 0, \ldots, 0)$$

whenever $N \geq 3$. Thus $\|\varphi_N(3)\|^2 = 1$, so the minimal norm is at most 1. Since all lattice coordinates are integers, any nonzero vector has squared norm at least 1, so the minimal norm is exactly 1. □

The minimal vectors of Λ_N are therefore all $v \in \Lambda_N$ with $\|v\|^2 = 1$. Numerically, one finds that for all $N \leq 200$ the number of minimal vectors is $2\,\xi(N)$, exactly as for the standard lattice $\mathbb{Z}^{\xi(N)}$. However, in general the set of minimal vectors does not determine the lattice uniquely when the rank becomes larger.

Using Proposition 7.3, we can nevertheless prove the theta-series identity in small ranks.

Theorem 7.5 (Low-dimensional case)**.** *Let $N \geq 2$ and set $r = \xi(N)$. If $r \leq 7$, then*

$$\Theta_{\Lambda_N}(q) = \Theta_{\mathbb{Z}^r}(q) = \sum_{n=0}^{\infty} s_r(n)\, q^n.$$

Proof. By Proposition 7.3 the lattice Λ_N is an odd unimodular lattice of rank r. A theorem of Griess and others (see e.g. Chenevier, *Unimodular hunting*, Theorem 1.1) asserts that for $r \leq 7$ the standard cubic lattice

$$I_r \cong \mathbb{Z}^r$$

is the *unique* positive definite unimodular lattice of rank r up to isometry. Hence Λ_N is isometric to $\mathbb{Z}^r$, and therefore their theta series coincide:

$$\Theta_{\Lambda_N}(q) = \Theta_{\mathbb{Z}^r}(q) = \sum_{n=0}^{\infty} s_r(n)\, q^n.$$ □

Example 7.6 (The case $N = 24$)**.** For $N = 24$ the linearly independent primes are

$$q_1 = 3, \quad q_2 = 7, \quad q_3 = 11, \quad q_4 = 23,$$

so $\xi(24) = 4$. The corresponding basis vectors are

$$\begin{aligned} b_1 &= (1,0,0,0,0,0,0,0,0),\\ b_2 &= (1,1,0,0,0,0,0,0,0),\\ b_3 &= (1,0,1,0,0,0,0,0,0),\\ b_4 &= (1,0,0,0,1,0,0,0,0), \end{aligned}$$

with Gram matrix

$$G_{24} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}, \qquad \det G_{24} = 1.$$

A computation of the theta series gives

$$\Theta_{\Lambda_{24}}(q) = 1 + 8q + 24q^2 + 32q^3 + 24q^4 + 48q^5 + \cdots,$$

which coincides term by term with the theta series of the 4-dimensional cubic lattice $\mathbb{Z}^4$, i.e. with the generating function for the number of representations of n as a sum of four squares. This is consistent with Theorem 7.5, since here $r = 4 \leq 7$.

7.4 Why minimal vectors are not enough in general

Lemma 7.4 shows that Λ_N has minimal norm 1, and numerical evidence suggests that the number of minimal vectors is $2\,\xi(N)$, exactly as for the standard lattice $\mathbb{Z}^{\xi(N)}$. One might hope to prove Question 7.2 by showing that the minimal vectors of Λ_N "look like" those of $\mathbb{Z}^{\xi(N)}$.

However, for unimodular lattices of larger rank (e.g. rank ≥ 8) it is well known that many non-isometric lattices can share the same minimal norm and even the same number of minimal vectors. In particular, knowing only the set of minimal vectors does not in general determine the entire theta series. Additional invariants—for example, information about vectors of slightly larger norm, or structural properties of the lattice such as its *shadow*—are needed to distinguish different unimodular lattices.

Therefore, while minimal vectors provide necessary conditions for the identity $\Theta_{\Lambda_N}(q) = \Theta_{\mathbb{Z}^{\xi(N)}}(q)$, they are not sufficient to settle Question 7.2 in general. At present the full answer to this question (for all N) remains open.

8 A Cholesky–type conjecture for the Gram matrices

We keep the notation of the previous section. For each fixed $N \geq 2$ we have linearly independent primes

$$q_1, \ldots, q_r \in \mathcal{P}(N), \qquad r = \xi(N),$$

their exponent vectors

$$b_j := \varphi_N(q_j) \in \mathbb{Z}^m \quad (1 \leq j \leq r),$$

and the associated Gram matrix

$$G_N := (\langle b_i, b_j \rangle)_{1 \leq i,j \leq r} \in M_r(\mathbb{Z}),$$

which is symmetric positive definite and unimodular (Proposition 7.3).

Over $\mathbb{R}$, every such G_N admits a unique Cholesky decomposition

$$G_N = L_N L_N^\top$$

with L_N lower triangular and positive diagonal entries. In general L_N need not have integer entries. The following conjecture asserts that, for the specific Gram matrices coming from linearly independent primes, we can choose L_N to be *integral* and unimodular.

Conjecture 8.1 (Integral Cholesky basis change)**.** For every $N \geq 2$ let G_N be the Gram matrix of the phi-lattice Λ_N as above, with rank $r = \xi(N)$. Then there exists a lower triangular matrix

$$L_N \in \mathrm{GL}_r(\mathbb{Z})$$

with positive diagonal entries such that

$$G_N = L_N L_N^\top.$$

Equivalently, if we set

$$U_N := L_N^{-1} \in \mathrm{GL}_r(\mathbb{Z}),$$

then

$$U_N^\top G_N U_N = I_r,$$

so that the lattice $(\Lambda_N, \langle\cdot,\cdot\rangle)$ is isometric (as an integral quadratic lattice) to the standard cubic lattice ($\mathbb{Z}^r$, usual dot product), and the isometry is given by the explicit basis change

$$[b_1 \ \cdots \ b_r] \ = \ [e_1 \ \cdots \ e_r]\, U_N^{\top},$$

where $(e_1, \ldots, e_r)$ is the standard basis of $\mathbb{Z}^r$.

In other words: the conjecture says that the *real* Cholesky factor of G_N can always be chosen with integer entries and determinant 1, and that its inverse $U_N = L_N^{-1}$ is the desired unimodular basis change between the cubic lattice and the phi-lattice.

Remark 8.2 (Equivalent arithmetic formulation and meaning)**.**

1. Let $b_1, \ldots, b_r$ be the exponent vectors of the linearly independent primes up to N and G_N their Gram matrix. The existence of $U_N \in \mathrm{GL}_r(\mathbb{Z})$ with $U_N^{\top} G_N U_N = I_r$ is equivalent to the existence of an *orthonormal $\mathbb{Z}$-basis* of the φ-lattice: there are vectors

   $$e_1, \ldots, e_r \in \Lambda_N$$

   such that

   $$\langle e_i, e_j \rangle = \delta_{ij} \quad \text{and} \quad e_j = \sum_{k=1}^{r} u_{kj}\, b_k \quad (u_{kj} \in \mathbb{Z}),$$

   i.e. the e_j are integer linear combinations of the exponent vectors b_k and form an orthonormal basis.

2. Arithmetically, each vector b_k encodes the p-adic exponents of $q_k - 1$ for primes $p \le N$; the entries of the Gram matrix are

   $$G_{ij} = \langle b_i, b_j \rangle = \sum_{p \le N} v_p(q_i - 1)\, v_p(q_j - 1).$$

   The conjecture is therefore equivalent to the following statement:

   There exist integer coefficients u_{kj} such that, for every integer n whose prime factors are $\le N$, the quadratic form

   $$\sum_{p \le N} v_p(n)^2$$

   can be written as a sum of r squares of integer linear forms *in the exponents $v_p(n)$, namely*

   $$\sum_{p \le N} v_p(n)^2 \ = \ \sum_{j=1}^{r} \left(\sum_{k=1}^{r} u_{kj}\, c_k(n) \right)^2,$$

   where $(c_1(n), \ldots, c_r(n))$ are the coordinates of n in the basis $(b_1, \ldots, b_r)$.

   In more informal terms: after an integer change of variables among the exponent vectors attached to the linearly independent primes, the Euclidean length

   $$\| \varphi_N(n) \|^2 = \sum_{p \le N} v_p(n)^2$$

   becomes literally a *sum of squares of r independent integer coordinates.* The Cholesky factor L_N encodes this change of variables, and its inverse U_N gives an explicit arithmetic basis in which the quadratic form "sum of squared valuations" is completely diagonal.

9 Equivalence of the Cholesky conjecture and the arithmetic reformulation

In this section we explain and prove the linear-algebraic statement behind the remark:

> The conjecture $G = LL^\top$ with $L \in \mathrm{GL}_r(\mathbb{Z})$ is equivalent to the existence of integer coefficients u_{kj} such that, for every integer n whose prime factors are $\le N$, the quadratic form
> $$\sum_{p \le N} v_p(n)^2$$
> can be written as a sum of r squares of integer linear forms in the coordinates $c_k(n)$ of n in the basis $(b_1, \dots, b_r)$:
> $$\sum_{p \le N} v_p(n)^2 = \sum_{j=1}^{r} \left(\sum_{k=1}^{r} u_{kj}\, c_k(n) \right)^2.$$

9.1 Linear algebra set-up

Fix $N \ge 2$. Let $p_1, \dots, p_m$ be the primes $\le N$ and consider the Euclidean space $\mathbb{R}^m$ with standard basis $e_{p_1}, \dots, e_{p_m}$ and standard inner product. For each linearly independent prime $q_j \le N$ we have its exponent vector

$$b_j := \varphi_N(q_j) = \big(v_{p_1}(q_j - 1), \dots, v_{p_m}(q_j - 1)\big) \in \mathbb{Z}^m, \qquad 1 \le j \le r.$$

Let B be the $m \times r$ matrix with columns b_j:

$$B := [\, b_1 \ \cdots \ b_r \,].$$

The associated Gram matrix is

$$G := G_N := B^\top B \in M_r(\mathbb{Z}),$$

so its entries are

$$G_{ij} = \langle b_i, b_j \rangle = \sum_{p \le N} v_p(q_i - 1)\, v_p(q_j - 1).$$

Now take any integer n whose prime factors are $\le N$. Its exponent vector is

$$v(n) := \big(v_{p_1}(n), \dots, v_{p_m}(n)\big) \in \mathbb{Z}^m.$$

Assume (as in the lattice set-up) that the b_j form a $\mathbb{Z}$-basis of the relevant exponent lattice, so $v(n)$ can be uniquely written as an integer linear combination of the b_j:

Definition 9.1 (Coordinates of n in the b-basis)**.** For such n we define $c(n) = (c_1(n), \dots, c_r(n))^\top \in \mathbb{Z}^r$ by

$$v(n) = \sum_{j=1}^{r} c_j(n)\, b_j = B\, c(n).$$

We now express the Euclidean norm of $v(n)$ in terms of $c(n)$.

Lemma 9.2. *For every n with prime factors $\le N$,*

$$\sum_{p \le N} v_p(n)^2 = \|v(n)\|^2 = c(n)^\top G\, c(n).$$

Proof. By definition of the norm in $\mathbb{R}^m$,

$$\|v(n)\|^2 = v(n)^\top v(n) = \sum_{p \le N} v_p(n)^2.$$

On the other hand, $v(n) = B\,c(n)$, so

$$\|v(n)\|^2 = (Bc(n))^\top (Bc(n)) = c(n)^\top (B^\top B)\, c(n) = c(n)^\top G\, c(n).$$

This is exactly the desired equality. □

So *arithmetically* the quadratic form

$$Q(n) := \sum_{p \le N} v_p(n)^2$$

is, in the coordinates $c(n)$, just the quadratic form with matrix G:

$$Q(n) = c(n)^\top G\, c(n).$$

9.2 From $U^\top G U = I$ to "sum of squares of linear forms"

Now suppose our Cholesky/basis-change conjecture holds in the following form:

$$\exists\, U \in \mathrm{GL}_r(\mathbb{Z}) \quad \text{such that} \quad U^\top G U = I_r. \tag{1}$$

(If you prefer the Cholesky factor L with $G = LL^\top$ and L unimodular, just take $U = L^{-1}$; the two formulations are equivalent.)

Define new integer coordinates by

$$z(n) := U^{-1} c(n) \in \mathbb{Z}^r.$$

Since U is invertible over $\mathbb{Z}$, this is an integer change of variables on $\mathbb{Z}^r$.

Then, using Lemma 9.2 and (1),

$$\begin{aligned} Q(n) &= c(n)^\top G\, c(n) \\ &= c(n)^\top (U^{-T} I_r U^{-1}) c(n) \qquad \text{(because } G = U^{-T} I_r U^{-1} \text{ is equivalent to } U^\top G U = I_r\text{)} \\ &= (U^{-1}c(n))^\top (U^{-1}c(n)) \\ &= z(n)^\top z(n) \\ &= \sum_{j=1}^{r} z_j(n)^2. \end{aligned}$$

Now write out $z_j(n)$ as an integer linear form in the $c_k(n)$. If we denote the entries of U^{-1} by

$$U^{-1} = (u_{jk})_{1 \le j,k \le r},$$

then

$$z_j(n) = \sum_{k=1}^{r} u_{jk}\, c_k(n).$$

Hence

$$Q(n) = \sum_{j=1}^{r} z_j(n)^2 = \sum_{j=1}^{r} \left(\sum_{k=1}^{r} u_{jk}\, c_k(n) \right)^2.$$

If you prefer the indices in the other order, just rename $u_{kj} := u_{jk}$; what matters is: *each $z_j(n)$ is an integer linear form in the coordinates $c_k(n)$, and the norm is the sum of their squares.*

In words:

> Assuming $U^\top G U = I_r$ with $U \in \mathrm{GL}_r(\mathbb{Z})$, the Euclidean length
>
> $$\|v(n)\|^2 = \sum_{p \le N} v_p(n)^2$$
>
> is equal to a sum of r squares of integer linear forms in the coordinate vector $c(n)$:
>
> $$\sum_{p \le N} v_p(n)^2 = \sum_{j=1}^{r} \left(\sum_{k=1}^{r} u_{jk}\, c_k(n) \right)^2 .$$

That is exactly the formula you quoted.

9.3 Conversely: from a sum of squares representation back to $U^\top G U = I$

Now let me show the converse: if such a representation exists *for all* coordinate vectors $c \in \mathbb{Z}^r$, then the matrix equality $U^\top G U = I_r$ holds.

Suppose there are integers u_{jk} such that for all integer vectors $c = (c_1, \dots, c_r)^\top \in \mathbb{Z}^r$ we have

$$c^\top G c = \sum_{j=1}^{r} \left(\sum_{k=1}^{r} u_{jk}\, c_k \right)^2 . \tag{2}$$

(Here we are now thinking of this as a *purely linear algebra* statement about the quadratic form with matrix G.)

Define the matrix $U^{-1} = (u_{jk})$ and $U := (U^{-1})^{-1}$. Then the right-hand side of (2) can be written as

$$\sum_{j=1}^{r} \left(\sum_{k=1}^{r} u_{jk}\, c_k \right)^2 = \sum_{j=1}^{r} z_j^2 = z^\top z,$$

where $z = U^{-1} c$. In matrix form this is

$$z^\top z = c^\top (U^{-1})^\top U^{-1} c.$$

Thus (2) is equivalent to

$$c^\top G c = c^\top (U^{-1})^\top U^{-1} c \qquad \text{for all } c \in \mathbb{Z}^r.$$

Since two symmetric matrices that give the same quadratic form on all $c \in \mathbb{R}^r$ must be equal, we conclude

$$G = (U^{-1})^\top U^{-1},$$

which is the same as

$$U^\top G U = I_r.$$

So the existence of such an integer-coefficient sum-of-squares decomposition *for all* coordinates c is *equivalent* to the integral Cholesky-type condition $U^\top G U = I_r$.

9.4 Summary in words

Putting it all together:

- Each integer n with primes $\leq N$ is encoded by its exponent vector $v(n) = (v_{p_1}(n), \ldots, v_{p_m}(n))$.
- The "length" of this vector is exactly the quadratic form
$$\|v(n)\|^2 = \sum_{p \leq N} v_p(n)^2.$$
- When you express $v(n)$ in the basis of exponent vectors $b_1, \ldots, b_r$ of the linearly independent primes,
$$v(n) = \sum_{j=1}^{r} c_j(n)\, b_j,$$
the same quadratic form becomes
$$\|v(n)\|^2 = c(n)^\top G\, c(n).$$
- The conjecture $U^\top G U = I_r$ with U unimodular is exactly the statement that there is an *integer* change of variables $z = U^{-1}c$ such that
$$c^\top G c = z^\top z = \sum_{j=1}^{r} z_j^2.$$
Writing $z_j = \sum_k u_{jk} c_k$ gives the "sum of squares of integer linear forms" description.

10 Infinitely many linearly independent primes: Wojowu's argument

In this section we present in detail a short and elegant proof, due to Wojowu [4], that there must exist infinitely many linearly independent primes in the sense of the φ–vectors.

We work in the infinite-dimensional rational vector space
$$V := \bigoplus_{q \text{ prime}} \mathbb{Q}\, e_q,$$
whose elements are sequences indexed by primes with only finitely many nonzero entries. For a prime number p we define its *φ–vector*
$$\varphi(p) := \sum_{q \text{ prime}} v_q(p-1)\, e_q \in V,$$
where $v_q(p-1)$ is the usual q–adic valuation of $p-1$. By construction each $\varphi(p)$ has only finitely many nonzero coordinates, namely at the primes dividing $p-1$.

Definition 10.1. A prime p is called *linearly independent* (LI) if the vector $\varphi(p)$ does not lie in the $\mathbb{Q}$–linear span of $\{\varphi(q) : q < p\}$. Equivalently, the family of vectors
$$\{\varphi(p) : p \text{ prime}\} \subset V$$
has a subset of linearly independent vectors, and those primes p whose $\varphi(p)$ belong to some fixed basis of the span are called LI primes.

We now state Wojowu's theorem.

Theorem 10.2 (Wojowu [4])**.** *There exist infinitely many linearly independent primes.*

Proof. Assume for contradiction that there are only finitely many linearly independent primes. Let these be

$$q_1, q_2, \ldots, q_r.$$

By definition, the vectors $\varphi(q_1), \ldots, \varphi(q_r)$ form a $\mathbb{Q}$–basis of the linear span of all $\varphi(p)$ with p prime. In particular, for every prime p there exist rational numbers $c_1(p), \ldots, c_r(p)$ such that

$$\varphi(p) = \sum_{i=1}^{r} c_i(p)\, \varphi(q_i).$$

Step 1: a finite set of possible prime divisors of $p-1$. For each $i = 1, \ldots, r$, the vector $\varphi(q_i)$ has nonzero coordinates only at those primes q that divide $q_i - 1$. Let

$$P := \bigcup_{i=1}^{r} \{\, q \text{ prime} : q \mid (q_i - 1) \,\}$$

be the finite set of all primes that divide at least one integer $q_i - 1$.

We claim that for every prime p and every prime $q \notin P$ we have

$$v_q(p-1) = 0,$$

i.e. $p-1$ has no prime divisors outside the fixed finite set P.

Indeed, fix a prime p and a prime $q \notin P$. Taking the q–th coordinate of the identity

$$\varphi(p) = \sum_{i=1}^{r} c_i(p)\, \varphi(q_i)$$

gives

$$v_q(p-1) = \sum_{i=1}^{r} c_i(p)\, v_q(q_i - 1).$$

But by definition of P, for $q \notin P$ we have $q \nmid (q_i - 1)$ for all i, so $v_q(q_i - 1) = 0$ for every i. Hence the right-hand side is 0, and therefore $v_q(p-1) = 0$ as claimed.

Thus we have shown:

> There exists a finite set of primes P such that, for every prime p, all prime divisors of $p-1$ lie in P.

Equivalently, for every prime p the integer $p-1$ belongs to the set

$$S := \{\, n \geq 1 : \text{all prime factors of } n \text{ lie in } P \}.$$

Step 2: the set S is too small to contain all $p-1$. We now compare the "size" of S with the set of all $p-1$ as p ranges over the primes, using Dirichlet series.

By construction, every $n \in S$ has the form

$$n = \prod_{q \in P} q^{k_q}$$

for some exponents $k_q \in \mathbb{Z}_{\geq 0}$. Therefore

$$\sum_{n \in S} \frac{1}{n} = \prod_{q \in P} \left(\sum_{k=0}^{\infty} \frac{1}{q^k} \right) = \prod_{q \in P} \frac{1}{1 - 1/q}.$$

Since P is finite and each factor $(1 - 1/q)^{-1}$ is finite, the Euler product on the right converges, and hence the sum

$$\sum_{n \in S} \frac{1}{n}$$

is finite.

On the other hand, the set $\{p - 1 : p \text{ prime}\}$ is contained in S, so

$$\sum_{p \text{ prime}} \frac{1}{p-1} \leq \sum_{n \in S} \frac{1}{n} < \infty.$$

We now compare this with the well-known divergence of the sum of reciprocals of primes (Euler):

$$\sum_{p \text{ prime}} \frac{1}{p} = \infty.$$

For all sufficiently large primes p we have $p - 1 \geq \frac{1}{2}p$, so

$$\frac{1}{p-1} \leq \frac{2}{p}.$$

Thus, apart from finitely many small p,

$$\sum_{p \text{ prime}} \frac{1}{p-1} \geq \frac{1}{2} \sum_{p \text{ prime},\ p>2} \frac{1}{p} = \infty.$$

The series $\sum_p 1/(p-1)$ thus diverges just like $\sum_p 1/p$.

We have thus reached a contradiction:

- Above we have shown that $\sum_p 1/(p-1)$ *converges*, because all $p - 1$ lie in S and $\sum_{n \in S} 1/n < \infty$.
- From the divergence of $\sum_p 1/p$ it follows that $\sum_p 1/(p-1)$ *diverges*.

This contradiction shows that our assumption was false. Thus the set of linearly independent primes cannot be finite; there must be infinitely many LI primes. □

11 Existence of a prime $q \equiv 1 \pmod{p}$ via Dirichlet

Fix a prime number p. We want to show that there exists a prime q of the form

$$q = kp + 1$$

for some integer $k \geq 1$. In other words,

$$q \equiv 1 \pmod{p}.$$

Dirichlet's theorem

We recall the classical result of Dirichlet on primes in arithmetic progressions:

Theorem 11.1 (Dirichlet)**.** *Let a, q be coprime positive integers, i.e.* $\gcd(a, q) = 1$*. Then there exist infinitely many prime numbers r such that*

$$r \equiv a \pmod{q}.$$

In particular, for any integer $q \geq 2$ we may take $a = 1$, and since $\gcd(1, q) = 1$, there exist infinitely many primes r with

$$r \equiv 1 \pmod{q}.$$

Application to the progression $1 \bmod p$

Now fix a prime p. Apply Dirichlet's theorem with modulus p and residue $a = 1$. Since $\gcd(1, p) = 1$, there are infinitely many primes q with

$$q \equiv 1 \pmod{p}.$$

Equivalently, each such prime can be written as

$$q = kp + 1$$

for some integer $k \geq 1$.

Thus we have proved:

Proposition 11.2. *For every prime p there exist infinitely many primes q of the form $q = kp + 1$ with $k \geq 1$.*

In particular, since there are infinitely many such primes, there is a *smallest* one. Let

$$q_{\min}(p)$$

denote the smallest prime with $q_{\min}(p) \equiv 1 \pmod{p}$. Then $q_{\min}(p)$ is well-defined, and we may write

$$q_{\min}(p) = k_{\min} p + 1$$

for some minimal integer $k_{\min} \geq 1$.

12 Minimal primes in the progression $1 \bmod p$ are linearly independent

We work with the notion of linear independence for primes defined via the exponent vectors of $p - 1$.

12.1 Set-up and notation

Let $(p_i)_{i \geq 1}$ denote the increasing sequence of primes,

$$p_1 = 2,\ p_2 = 3,\ p_3 = 5, \dots.$$

For each integer $n \geq 2$ we consider the (infinite) exponent vector

$$\varphi(n) := (v_{p_1}(n-1), v_{p_2}(n-1), v_{p_3}(n-1), \dots) \in \prod_{i \geq 1} \mathbb{Z},$$

where $v_{p_i}(\cdot)$ is the p_i-adic valuation. Only finitely many coordinates of $\varphi(n)$ are nonzero.

Definition 12.1 (Linear independence for primes)**.** A prime q is called *linearly independent* (LI) if the vector $\varphi(q)$ does not lie in the $\mathbb{Q}$-linear span of the vectors $\varphi(r)$ with r prime and $r < q$. Otherwise q is called *linearly dependent* (LD).

Equivalently, if we fix some $N \geq q$ and truncate the vectors to the first $\pi(N)$ coordinates, we can regard all $\varphi(r)$ with $r \leq q$ as vectors in the same finite-dimensional space $\mathbb{Q}^{\pi(N)}$, and the definition is unchanged.

12.2 Statement

Fix a prime p, and suppose there exists a prime q in the arithmetic progression $1 \bmod p$ which is *minimal* in the following sense:

$$q \equiv 1 \pmod{p}, \quad \text{and} \quad \nexists \text{ prime } r < q \text{ with } r \equiv 1 \pmod{p}. \tag{3}$$

(In your algorithm, this is exactly: choose the smallest $k \geq 1$ such that $q = kp + 1$ is prime; then q is the smallest prime $\equiv 1 \pmod{p}$.)

Proposition 12.2. *If q satisfies* (3)*, then q is linearly independent in the above sense.*

12.3 Proof

Let p be a fixed prime, and let q be a prime satisfying (3). We must show that $\varphi(q)$ does not lie in the $\mathbb{Q}$-span of the vectors $\varphi(r)$ with r prime and $r < q$.

Step 1: work in a common finite-dimensional space.

Take $N := q$ and consider the truncated exponent vectors

$$\varphi_N(n) := \big(v_{p_1}(n-1), \ldots, v_{p_{\pi(N)}}(n-1)\big) \in \mathbb{Z}^{\pi(N)}.$$

For primes $r \leq q$ we regard $\varphi_N(r)$ as vectors in $\mathbb{Q}^{\pi(N)}$. Linear (in)dependence of $\varphi(q)$ from the previous $\varphi(r)$ is equivalent to that of $\varphi_N(q)$ from the $\varphi_N(r)$ with $r < q$.

Let $p = p_i$ for some index i (i.e. $p_i = p$). Then the i-th coordinate of $\varphi_N(n)$ is $v_p(n-1)$.

Step 2: the p-coordinate of $\varphi_N(q)$.

Since $q \equiv 1 \pmod{p}$, we have $p \mid (q-1)$ and hence

$$v_p(q-1) \;\geq\; 1.$$

Therefore the p-coordinate (i.e. the i-th coordinate) of $\varphi_N(q)$ is nonzero:

$$\big(\varphi_N(q)\big)_i = v_p(q-1) \neq 0.$$

Step 3: the p-coordinate of earlier vectors.

Now let r be any prime with $r < q$. We claim that

$$v_p(r-1) = 0.$$

Indeed, if $v_p(r-1) \geq 1$, then $p \mid (r-1)$, so

$$r \equiv 1 \pmod{p}.$$

But this is impossible by the minimality assumption (3), which says that there is *no* prime $r < q$ with $r \equiv 1 \pmod{p}$.

Hence, for all primes $r < q$, the p-coordinate of $\varphi_N(r)$ is zero:

$$\big(\varphi_N(r)\big)_i = v_p(r-1) = 0 \qquad \text{for all primes } r < q.$$

Step 4: linear independence from the p-coordinate.

Suppose, for contradiction, that $\varphi_N(q)$ lies in the $\mathbb{Q}$-span of the previous vectors $\varphi_N(r)$, $r < q$. Then there exist rational numbers c_r (only finitely many nonzero) such that

$$\varphi_N(q) = \sum_{\substack{r<q \\ r \text{ prime}}} c_r\, \varphi_N(r).$$

Comparing the p-coordinate (the i-th coordinate) on both sides, we obtain

$$v_p(q-1) = \big(\varphi_N(q)\big)_i = \sum_{r<q} c_r \,\big(\varphi_N(r)\big)_i = \sum_{r<q} c_r\, v_p(r-1).$$

But we have just shown that $v_p(r-1) = 0$ for all primes $r < q$, so the right-hand side is 0, while the left-hand side is $v_p(q-1) \geq 1$. This is a contradiction.

Therefore $\varphi_N(q)$ cannot lie in the $\mathbb{Q}$-span of the earlier vectors $\{\varphi_N(r) : r < q\}$, and hence q is linearly independent.

13 Equivalence of the prime version and the general version

Fix the set of all prime numbers $\mathcal{P}$, and consider integer coefficients $(a_q)_{q\in\mathcal{P}}$ with only finitely many nonzero entries.

We compare the following two statements:

(P) For every prime number p there exist integers a_q (almost all equal to 0) such that

$$\log p \;=\; \sum_{q\in\mathcal{P}} a_q \,\log(q-1).$$

(N) For every integer $n \geq 1$ there exist integers a_q (almost all equal to 0) such that

$$\log n \;=\; \sum_{q\in\mathcal{P}} a_q \,\log(q-1).$$

We show that (P) and (N) are equivalent.

(N) implies (P)

This direction is immediate: primes are just a special case of natural numbers. If (N) holds for all $n \geq 1$, then in particular for every prime p we have such a representation with suitable integers a_q, and hence (P) holds.

(P) implies (N)

Assume now that (P) holds. Let $n \geq 2$ be an arbitrary integer with prime factorisation

$$n = \prod_{i=1}^{k} p_i^{e_i},$$

where $p_1, \ldots, p_k$ are distinct primes and $e_i \geq 1$.

By (P), for each prime factor p_i there exist integers $a_q^{(i)}$ (almost all equal to 0) such that

$$\log p_i \;=\; \sum_{q \in \mathcal{P}} a_q^{(i)} \, \log(q-1).$$

Using the usual rules for the logarithm, we have

$$\log n \;=\; \log\Bigl(\prod_{i=1}^{k} p_i^{e_i}\Bigr) \;=\; \sum_{i=1}^{k} e_i \, \log p_i.$$

Substituting the representations of the $\log p_i$ gives

$$\begin{aligned} \log n &= \sum_{i=1}^{k} e_i \left(\sum_{q \in \mathcal{P}} a_q^{(i)} \, \log(q-1) \right) \\ &= \sum_{q \in \mathcal{P}} \left(\sum_{i=1}^{k} e_i \, a_q^{(i)} \right) \log(q-1). \end{aligned}$$

Define new integer coefficients

$$b_q := \sum_{i=1}^{k} e_i \, a_q^{(i)} \in \mathbb{Z}.$$

Since for each i only finitely many of the $a_q^{(i)}$ are nonzero and there are only finitely many indices i, it follows that only finitely many b_q are nonzero. Hence we obtain

$$\log n \;=\; \sum_{q \in \mathcal{P}} b_q \, \log(q-1),$$

which is precisely the representation required in (N).

For $n = 1$ we have $\log 1 = 0$, which is trivially obtained by taking $b_q = 0$ for all q.

Conclusion

We have shown:

$$\text{(P) holds} \iff \text{(N) holds.}$$

That is, the statement

$$\forall p \text{ prime} : \; \log p = \sum_{q} a_q \log(q-1)$$

is equivalent to the general statement

$$\forall n \in \mathbb{N}_{\geq 1} : \; \log n = \sum_{q} a_q \log(q-1),$$

provided that in both cases only finitely many of the a_q are nonzero.

14 From Hypothesis (H) to logarithmic representations

We fix the following hypothesis on primes in arithmetic progressions.

(H) For every integer $n \geq 2$ there exists a prime number q and an integer k with $1 \leq k \leq n-1$ such that

$$q = kn + 1.$$

In other words, for each n there is a prime congruent to $1 \pmod{n}$ whose multiplier k is at most $n-1$.

Under this hypothesis we can prove the following representation theorem.

Proposition 14.1. *Assume Hypothesis* (H)*. Then for every integer $n \geq 2$ there exist integers a_q (only finitely many nonzero) such that*

$$\log n \;=\; \sum_{q \text{ prime}} a_q \log(q-1).$$

Proof. We argue by induction on n.

Base case $n = 2$. We have

$$\log 2 = \log(3-1),$$

so the statement holds with $a_3 = 1$ and $a_q = 0$ for $q \neq 3$.

Induction step. Assume that the claim holds for all integers m with $2 \leq m \leq n-1$, i.e. for each such m there exist integers $a_q^{(m)}$, almost all zero, such that

$$\log m = \sum_{q \text{ prime}} a_q^{(m)} \log(q-1).$$

We now prove the claim for n.

By Hypothesis (H) there exist a prime q and an integer k with $1 \leq k \leq n-1$ such that

$$q = kn + 1.$$

In particular

$$q - 1 = kn,$$

and by the logarithm laws,

$$\log(q-1) = \log(kn) = \log k + \log n.$$

Since $1 \leq k \leq n-1$ and $n \geq 3$ in the induction step, we have $k \geq 2$ and $k \leq n-1$, so k lies in the range of the induction hypothesis. Hence there exist integers $a_p^{(k)}$ (almost all zero) such that

$$\log k = \sum_{p \text{ prime}} a_p^{(k)} \log(p-1).$$

Substituting this into the identity for $\log(q-1)$ yields

$$\begin{aligned} \log(q-1) &= \log k + \log n \\ &= \sum_{p \text{ prime}} a_p^{(k)} \log(p-1) + \log n. \end{aligned}$$

Rearranging, we obtain

$$\log n = \log(q-1) - \sum_{p \text{ prime}} a_p^{(k)} \log(p-1).$$

This has the desired form: if we define new integer coefficients b_r by

$$b_q := 1, \qquad b_p := -a_p^{(k)} \text{ for all primes } p \neq q,$$

then only finitely many b_r are nonzero (because only finitely many $a_p^{(k)}$ are nonzero), and

$$\log n = \sum_{r \text{ prime}} b_r \log(r-1).$$

Thus the statement holds for n, completing the induction.

By induction, for every integer $n \geq 2$ there exist integers a_q, almost all zero, with

$$\log n = \sum_{q \text{ prime}} a_q \log(q-1),$$

as claimed. □

Remark 14.2. Hypothesis (H) is a very strong conjecture about the size of the smallest prime in the arithmetic progression $1 \bmod n$: it asserts the existence of a prime $q \equiv 1 \pmod{n}$ with $q \leq n^2 - n + 1$ for every $n \geq 2$. This is far beyond what is currently known unconditionally about primes in arithmetic progressions; the argument above therefore shows that *if* such a strong distribution property holds, then every $\log n$ lies in the $\mathbb{Z}$-span of the numbers $\log(q-1)$ for primes q.

15 A recursive formal decomposition of $\log p$

We are interested in representing the logarithm of a prime number p as an integer linear combination of logarithms of numbers of the form $(q-1)$, where q runs over primes. This matches the shape of conjecture (N):

$$\log n \;=\; \sum_{q \text{ prime}} a_q \log(q-1), \qquad a_q \in \mathbb{Z}, \text{ finitely many nonzero.}$$

The recursive procedure suggested by the Sage code can be formulated purely formally, without assuming that such representations always exist.

15.1 Formal symbols and ambient module

Let $\mathcal{P}$ denote the set of all prime numbers. We consider the free $\mathbb{Z}$–module

$$\mathcal{F} := \bigoplus_{q \in \mathcal{P}} \mathbb{Z}\, \Psi_q,$$

whose basis elements Ψ_q are formal symbols intended to represent $\log(q-1)$.

A typical element of $\mathcal{F}$ is a finite sum

$$\sum_{q \in \mathcal{P}} c_q \, \Psi_q, \qquad c_q \in \mathbb{Z}.$$

The guiding idea is:

For each prime p, we want to construct a formal expression

$$R(p) \in \mathcal{F}$$

such that, after the substitution $\Psi_q \mapsto \log(q-1)$, the real number $R(p)$ evaluates to $\log p$.

The printed output of the Sage code

$$\log 2 = \Psi_3, \quad \log 3 = -\Psi_3 + \Psi_7, \quad \log 5 = \Psi_{11} - \Psi_3, \ \ldots$$

should be thought of as the formal objects $R(p)$ written in this basis.

15.2 Choice of auxiliary primes q

Fix a prime number p. The algorithm searches for an *auxiliary prime* q of the shape

$$q = kp + 1,$$

with an integer k satisfying

$$1 \le k \le p.$$

(Concretely, the code tries $k = 1, 2, \ldots, p$ and chooses the first k for which $kp+1$ is prime.)

For such q we have the elementary identity

$$q - 1 = kp \quad \Rightarrow \quad \log(q-1) = \log(kp) = \log k + \log p.$$

Rearranging gives

$$\log p \ = \ \log(q-1) - \log k. \tag{4}$$

The recursion is based exactly on (4), together with a way to express $\log k$ in terms of the formal symbols Ψ_q for smaller integers.

15.3 Inductive hypothesis for smaller integers

To describe the recursion cleanly, we phrase everything in terms of *all* positive integers, but we will only apply it to primes.

Definition 15.1 (Formal representation of $\log n$)**.** Let $n \ge 1$ be an integer. A *formal representation* of $\log n$ is an element $R(n) \in \mathcal{F}$ of the form

$$R(n) = \sum_{q \in \mathcal{P}} c_q(n)\, \Psi_q, \qquad c_q(n) \in \mathbb{Z},$$

with only finitely many $c_q(n) \neq 0$, such that

$$\log n \ = \ R(n)\big|_{\Psi_q = \log(q-1)} \ = \ \sum_{q \in \mathcal{P}} c_q(n)\, \log(q-1).$$

The recursive scheme assumes:

For all integers $m < p$, the values $R(m)$ have already been defined in $\mathcal{F}$ and satisfy

$$\log m = R(m)\big|_{\Psi_q = \log(q-1)}.$$

In particular, if $k \le p$ is the coefficient in $q = kp + 1$, then we will use $R(k)$ as the formal representation of $\log k$.

15.4 Recursive definition of $R(p)$ for primes

We now describe the recursion that corresponds to the Sage code.

Definition 15.2 (Formal recursion for $R(p)$)**.** We define $R(p) \in \mathcal{F}$ for primes p as follows.

1. *Base step.* For $p = 2$ choose the auxiliary prime $q = 3$, so $q - 1 = 2 \cdot 1$. Then

$$\log 2 = \log(3 - 1),$$

and we set

$$R(2) := \Psi_3.$$

2. *Inductive step.* Let $p > 2$ be a prime, and assume that $R(m)$ is defined for every integer $m < p$.

Choose an auxiliary prime $q = kp + 1$ with $1 \le k \le p$. Using (4),

$$\log p = \log(q - 1) - \log k.$$

By the inductive hypothesis we have a formal representation

$$R(k) = \sum_{r \in \mathcal{P}} c_r(k)\, \Psi_r \quad \text{with} \quad \log k = \sum_{r \in \mathcal{P}} c_r(k)\, \log(r - 1).$$

We then *define* the formal element

$$R(p) \;:=\; \Psi_q \;-\; R(k) \;=\; \Psi_q \;-\; \sum_{r \in \mathcal{P}} c_r(k)\, \Psi_r \;\in\; \mathcal{F}. \tag{5}$$

By construction, after substitution $\Psi_q \mapsto \log(q - 1)$, the expression (5) evaluates to

$$\log(q - 1) - \log k = \log p,$$

so $R(p)$ is indeed a formal representation of $\log p$.

15.5 Relation to the Sage output

The Sage function `repr_(n)` is used with $n = p + 1$ for primes p, and prints expressions of the shape

$$\log(p) \;\sim\; \text{(integer combination of symbols } \Psi_q),$$

which, in your notation, were written as

$$\phi(p + 1) = \text{combination of } \phi_q.$$

Translated into the present logarithmic language, those lines are exactly the formal identities

$$\log p = R(p)\big|_{\Psi_q = \log(q-1)},$$

with $R(p)$ built recursively via the rule

$$R(p) = \Psi_{kp+1} - R(k), \qquad 1 \le k \le p,\ q = kp + 1 \text{ prime}.$$

In other words, the algorithm constructs step by step a representation of each $\log p$ as an integer linear combination of $\log(q - 1)$ for various primes q, and the coefficients are encoded symbolically in the formal expressions $R(p) \in \mathcal{F}$.

16 Uniqueness of representations using linearly independent primes

We keep the notation from the previous sections. Let

$$V := \bigoplus_{p \text{ prime}} \mathbb{Q}\, e_p$$

be the $\mathbb{Q}$–vector space of finitely supported sequences indexed by primes, and for each prime q let

$$\varphi(q) := \sum_p v_p(q-1)\, e_p \in V$$

be the exponent vector of $q-1$.

Definition 16.1 (Linearly independent primes revisited)**.** A set of primes $\mathcal{L}$ is called a set of *linearly independent primes* if the vectors

$$\{\varphi(q) : q \in \mathcal{L}\} \subset V$$

are $\mathbb{Q}$–linearly independent. We additionally assume that $\{\varphi(q) : q \in \mathcal{L}\}$ spans the exponent lattice generated by all $\varphi(p)$ (so that every relevant exponent vector can be expressed as an integer linear combination of the $\varphi(q)$ with $q \in \mathcal{L}$).

In the logarithmic formulation, we are interested in representations of the form

$$\log n = \sum_{q \in \mathcal{L}} a_q \log(q-1), \qquad a_q \in \mathbb{Z},$$

where only finitely many a_q are nonzero. Equivalently,

$$n = \prod_{q \in \mathcal{L}} (q-1)^{a_q}.$$

The following proposition shows that once we restrict to linearly independent primes, such representations (if they exist) are automatically unique.

Proposition 16.2 (Uniqueness of representation via LI primes)**.** *Let $\mathcal{L}$ be a set of linearly independent primes as above. Suppose for some integer $n \geq 1$ we have two representations*

$$n = \prod_{q \in \mathcal{L}} (q-1)^{a_q} = \prod_{q \in \mathcal{L}} (q-1)^{b_q},$$

with $a_q, b_q \in \mathbb{Z}$ and only finitely many of them nonzero. Then

$$a_q = b_q \quad \text{for all } q \in \mathcal{L}.$$

Equivalently, if

$$\log n = \sum_{q \in \mathcal{L}} a_q \log(q-1) = \sum_{q \in \mathcal{L}} b_q \log(q-1),$$

then $a_q = b_q$ for all $q \in \mathcal{L}$.

Proof. From the two factorizations we obtain

$$1 = \frac{n}{n} = \prod_{q \in \mathcal{L}} (q-1)^{a_q - b_q}.$$

Set

$$c_q := a_q - b_q \in \mathbb{Z},$$

again with only finitely many nonzero entries. Then

$$\prod_{q \in \mathcal{L}} (q-1)^{c_q} = 1.$$

Taking p–adic valuations on both sides for each rational prime p gives

$$0 = v_p(1) = v_p\Big(\prod_{q \in \mathcal{L}} (q-1)^{c_q}\Big) = \sum_{q \in \mathcal{L}} c_q \, v_p(q-1).$$

In vector notation, this says precisely that

$$\sum_{q \in \mathcal{L}} c_q \, \varphi(q) = 0 \quad \text{in } V.$$

By the defining property of $\mathcal{L}$, the family $\{\varphi(q) : q \in \mathcal{L}\}$ is $\mathbb{Q}$–linearly independent. Hence the only rational solution of

$$\sum_{q \in \mathcal{L}} c_q \, \varphi(q) = 0$$

is $c_q = 0$ for all $q \in \mathcal{L}$. In particular, there is no nontrivial integer relation, so all c_q must vanish:

$$c_q = 0 \quad \text{for all } q \in \mathcal{L}.$$

Therefore $a_q - b_q = 0$ for all q, i.e. $a_q = b_q$ for all $q \in \mathcal{L}$. This proves uniqueness of the representation. □

Remark 16.3. Without restricting to a linearly independent set of primes, uniqueness need not hold: different combinations of dependent vectors $\varphi(p)$ can produce the same exponent vector, hence the same product $\prod (p-1)^{a_p}$ and the same logarithmic identity $\log n = \sum a_p \log(p-1)$. The restriction to LI primes removes exactly these hidden linear relations.

17 Related works on least primes in arithmetic progressions

The hypotheses and constructions above are closely related to the classical and modern theory of primes in arithmetic progressions, especially to results and conjectures about the *least* prime in a given progression.

Dirichlet, Linnik, and the least prime in an arithmetic progression

Dirichlet's theorem guarantees infinitely many primes in each reduced arithmetic progression $a \bmod q$, but gives no quantitative control on the size of the smallest such prime.

Linnik's theorem asserts that there exists an absolute constant $L > 0$ such that, for every $q \geq 1$ and every a coprime to q, the least prime $p \equiv a \pmod q$ satisfies

$$p \ll q^L,$$

with an ineffective implied constant and some explicit bounds known for L (currently $L < 5.2$ is known). Under the Generalized Riemann Hypothesis for Dirichlet L-functions one expects (and can prove) much stronger bounds $p \ll q^{2+\varepsilon}$ for every $\varepsilon > 0$, and various conditional results of this shape are classical; see, for instance, Montgomery and Vaughan, *Multiplicative Number Theory I. Classical Theory.*

Connections with Hypothesis (H)

Hypothesis (H) from the previous section asserts that for every $n \geq 2$ there is a prime $q \equiv 1 \pmod{n}$ with

$$q = kn + 1, \qquad 1 \leq k \leq n - 1,$$

i.e. $q \leq n^2 - n + 1$. In terms of the least prime $p(1, n)$ in the progression $1 \bmod n$, this says

$$p(1, n) \leq n^2 - n + 1 \quad \text{for all } n \geq 2.$$

Thus (H) is a very strong uniform upper bound of the Linnik-type form $p(1, n) \ll n^2$, and in fact with an explicit constant 1 and a precise polynomial $n^2 - n + 1$. Unconditionally, no such general bound is known: Linnik's theorem gives $p(1, n) \ll n^L$ for some $L > 2$, while under GRH one can show (roughly)

$$p(1, n) \ll n^{2+\varepsilon}$$

for every $\varepsilon > 0$, but not with a fixed exponent 2 and no ε.

In this sense, Hypothesis (H) is at least as strong as the "GRH-level" conjectures about the least prime in an arithmetic progression, and probably strictly stronger. The consequences derived above—namely, that every $\log n$ can be written as an integer linear combination of $\log(q - 1)$, and hence (P) and (N) hold—give a different, structural way to interpret such strong bounds.

Montgomery-type conjectures and distribution of primes

Montgomery and Vaughan, and later Montgomery and Soundararajan, formulated a number of conjectures about the distribution of primes in arithmetic progressions and short intervals, often phrased in terms of the error term in the prime number theorem for arithmetic progressions and in terms of equidistribution of primes among residue classes. These conjectures typically imply very strong bounds for the least prime $p(a, q)$ in a given residue class $a \bmod q$, often of the form

$$p(a, q) \ll q^{2+\varepsilon}$$

or even closer to the "square-root barrier" in certain averaged senses.

While Hypothesis (H) is not obviously equivalent to any of these conjectures, it is philosophically in the same spirit: it postulates the existence of very small primes in the progression $1 \bmod n$ *uniformly* for all n, and the derivation of (N) shows that such a strong distribution of primes forces the logarithms of all integers to lie in the integer span of the special set $\{\log(q - 1)\}_{q \text{ prime}}$.

It would be interesting to understand whether some averaged or weakened form of (H) follows from Montgomery-type conjectures, or conversely whether the logarithmic representation properties of Sections above can be used to reformulate aspects of the least-prime problem in linear-algebraic or lattice terms.

18 Recovering the H–conjecture from the bounded log–representation

In this subsection we show that a suitable bounded log–representation actually *implies* the H–conjecture. Roughly speaking, if every prime p admits an expression of $\log p$ as an integer linear combination of $\log(q-1)$ with $q \le p^2$, then for each such p there must exist a prime q of the form $q = kp+1$ with $1 \le k \le p-1$.

We formalise this as follows.

Hypothesis 18.1 (Bounded log–representation)**.** For every integer $n \ge 2$ there exists a finite set of primes $\mathcal{Q}(n) \subset \{q \text{ prime} : q \le P_1(n)^2\}$ and integers $a_{n,q} \in \mathbb{Z}$ such that

$$\log n \;=\; \sum_{q \in \mathcal{Q}(n)} a_{n,q} \log(q-1). \tag{6}$$

Here $P_1(n)$ denotes the largest prime factor of n (so in particular $P_1(p) = p$ if p is prime). Note that we do *not* assume the recursive construction from Definition **??** in this hypothesis; we only assume the existence of a representation with a range restriction $q \le P_1(n)^2$.

We now prove that Hypothesis 18.1 implies the H–conjecture.

Proposition 18.2 (Reverse implication)**.** *Assume Hypothesis 18.1. Then for every prime p there exists a prime q and an integer k with*

$$q \;=\; kp+1, \qquad 1 \le k \le p-1.$$

In other words, the H–conjecture holds for all primes p.

Proof. Fix a prime p. By Hypothesis 18.1 applied to $n = p$, there exists a finite set of primes

$$\mathcal{Q}(p) \;\subset\; \{q \text{ prime} : q \le P_1(p)^2\} \;=\; \{q \text{ prime} : q \le p^2\}$$

and integers $a_q := a_{p,q} \in \mathbb{Z}$ such that

$$\log p \;=\; \sum_{q \in \mathcal{Q}(p)} a_q \log(q-1). \tag{7}$$

Since $p \ge 2$, we have $\log p > 0$, so at least one of the coefficients a_q must be nonzero.

Step 1: Passing from logs to a multiplicative identity.

Exponentiating (7) gives

$$p = \exp(\log p) = \exp\Big(\sum_q a_q \log(q-1)\Big) = \prod_{q \in \mathcal{Q}(p)} \exp\big(a_q \log(q-1)\big) = \prod_{q \in \mathcal{Q}(p)} (q-1)^{a_q}.$$

Thus we obtain an exact multiplicative identity

$$p \;=\; \prod_{q \in \mathcal{Q}(p)} (q-1)^{a_q}, \tag{8}$$

where the product is taken over finitely many primes $q \le p^2$ with integer exponents a_q (which may be positive, zero, or negative).

Step 2: Separating positive and negative exponents.

Write each integer exponent as

$$a_q = a_q^+ - a_q^-,$$

where $a_q^+ := \max(a_q, 0) \in \mathbb{Z}_{\geq 0}$ and $a_q^- := \max(-a_q, 0) \in \mathbb{Z}_{\geq 0}$. Then

$$(q-1)^{a_q} = (q-1)^{a_q^+ - a_q^-} = \frac{(q-1)^{a_q^+}}{(q-1)^{a_q^-}}.$$

Substituting into (8), we get

$$p = \prod_q \frac{(q-1)^{a_q^+}}{(q-1)^{a_q^-}} = \frac{\prod_q (q-1)^{a_q^+}}{\prod_q (q-1)^{a_q^-}}.$$

Multiplying both sides by the (positive integer) denominator $\prod_q (q-1)^{a_q^-}$, we obtain

$$p \cdot \prod_{q \in \mathcal{Q}(p)} (q-1)^{a_q^-} = \prod_{q \in \mathcal{Q}(p)} (q-1)^{a_q^+}. \tag{9}$$

Both sides of this identity are positive integers. The right–hand side is a product of positive powers of integers of the form $q-1$, with $q \in \mathcal{Q}(p)$ and $q \leq p^2$.

Step 3: Ensuring a nontrivial positive part.

We next observe that at least one exponent a_q must be strictly positive. Suppose, for contradiction, that $a_q \leq 0$ for all $q \in \mathcal{Q}(p)$. Then $a_q^+ = 0$ for all q, and (8) reduces to

$$p = \prod_q (q-1)^{a_q} = \prod_q (q-1)^{-a_q^-} = \frac{1}{\prod_q (q-1)^{a_q^-}}.$$

The right–hand side is a positive rational number of the form $1/D$ with $D \in \mathbb{N}$, while the left–hand side is the integer $p \geq 2$. This is impossible. Therefore, there exists at least one prime $q \in \mathcal{Q}(p)$ such that $a_q^+ > 0$, i.e. $a_q > 0$.

Consequently, the right–hand side of (9), $\prod_q (q-1)^{a_q^+}$, is a nontrivial product of integers $(q-1) \geq 1$, with at least one factor strictly greater than 1.

Step 4: Divisibility by p and existence of a prime $q \equiv 1 \pmod{p}$.

Consider the integer identity (9):

$$p \cdot \prod_q (q-1)^{a_q^-} = \prod_q (q-1)^{a_q^+}.$$

Denote

$$D := \prod_q (q-1)^{a_q^-}, \qquad N := \prod_q (q-1)^{a_q^+}.$$

Then the identity reads

$$p \cdot D = N.$$

In particular, p divides N. Since p is a prime, the fact that p divides the product N implies that

$$p \text{ divides } (q-1) \text{ for at least one } q \text{ with } a_q^+ > 0.$$

Thus, there exists a prime $q \in \mathcal{Q}(p)$ with $a_q > 0$ such that

$$p \mid (q-1).$$

Equivalently, there exists an integer $k \geq 1$ such that

$$q - 1 = kp, \qquad \text{i.e.} \qquad q = kp + 1.$$

Moreover, by Hypothesis 18.1, each such q satisfies $q \leq p^2$ (since $P_1(p) = p$). Therefore

$$kp + 1 = q \;\leq\; p^2 \quad \Longrightarrow \quad kp \;\leq\; p^2 - 1 \quad \Longrightarrow \quad k \;\leq\; \frac{p^2 - 1}{p} = p - \frac{1}{p}.$$

Since k is an integer, this gives

$$1 \;\leq\; k \;\leq\; p - 1.$$

Thus we have exhibited, for the fixed prime p, a prime q and an integer k with

$$q = kp + 1, \qquad 1 \leq k \leq p - 1,$$

as claimed. □

Remark 18.3. The proof shows slightly more: if Hypothesis 18.1 holds for all integers $n \geq 2$, then for any prime divisor p of n, there must exist a prime $q \leq P_1(n)^2$ with $q \equiv 1 \pmod{p}$. Specialising to $n = p$ itself recovers exactly the H–conjecture. In this sense, the existence of sufficiently "short" log–representations with the range restriction $q \leq P_1(n)^2$ is essentially equivalent to the H–conjecture.

18.1 Montgomery's conjecture and Hypothesis (H)

The hypothesis (H) we use in the recursive logarithmic construction is a very strong uniform statement about primes in the progression 1 mod n. In its prime-modulus form it says:

Hypothesis 18.4 (Prime-modulus version of (H))**.** For every prime p there exist an integer k with $1 \leq k \leq p - 1$ and a prime q such that

$$q = kp + 1.$$

Equivalently, the least prime $q \equiv 1 \pmod{p}$ satisfies $q \leq p^2 - p + 1$.

Unconditionally this is far out of reach; even the weaker bound $\min\{q \equiv 1 \bmod p\} \ll p^{2+\varepsilon}$ for every $\varepsilon > 0$ is not known in general. However, Hypothesis 21.3 fits very naturally into the framework of conjectures on the distribution of primes in arithmetic progressions, particularly a conjecture of Montgomery on the size of the error term in the prime number theorem for arithmetic progressions.

We recall one convenient formulation (see Iwaniec–Kowalski, *Analytic Number Theory*, §17.3, conjecture (17.5), and Exercise 17.6). Roughly speaking, Montgomery's conjecture predicts square-root cancellation in the error term of the prime number theorem in arithmetic progressions, uniformly in the modulus:

Conjecture 18.5 (Montgomery)**.** For each $\varepsilon > 0$ there is a constant C_ε such that for all $x \geq 2$, all integers $q \geq 1$ and residues a mod q with $(a, q) = 1$, we have

$$\pi(x; q, a) \;=\; \frac{\mathrm{Li}(x)}{\varphi(q)} \;+\; O_\varepsilon\!\left(C_\varepsilon \, x^{\varepsilon} \left(\frac{x}{q} \right)^{1/2} \right),$$

uniformly for $q \leq x$.

Here $\pi(x;q,a)$ counts primes $p \le x$ with $p \equiv a \pmod q$, φ is Euler's totient function, and $\mathrm{Li}(x)$ is the logarithmic integral. The exact normalisation varies in the literature, but the key point is the square-root size $(x/q)^{1/2}$ of the error term, which is what one expects heuristically from a "random model" for primes.

Under Conjecture 18.5 one can prove very strong bounds for the least prime in an arithmetic progression. A precise formulation convenient for us is the so-called *PKD conjecture*, which concerns primes of the shape $km + d$ with $(d, m) = 1$. Goldston and Heath-Brown (in a MathOverflow answer, building on Iwaniec–Kowalski (17.5)) show that Montgomery's conjecture implies the following statement with $f(N) = N - 1$: for all sufficiently large moduli m, and every residue d coprime to m, there exists a prime of the form

$$km + d \qquad \text{with} \qquad 1 < k \le m - 1.$$

Specialising to the progression $1 \bmod p$ with p prime, and taking $m = p$, $d = 1$, this says:

Corollary 18.6 (Montgomery ⇒ prime-modulus (H) for large p)**.** *Assume Conjecture 18.5. Then there exists p_0 such that for every prime $p \ge p_0$ there is a prime q of the form*

$$q = kp + 1 \qquad \textit{with} \qquad 1 \le k \le p - 1.$$

In other words, for all sufficiently large primes p the prime-modulus version of Hypothesis 21.3 holds.

For a fixed finite list of small primes $p < p_0$, the existence of $q = kp + 1$ with $k \le p - 1$ can be verified by computation, so that Conjecture 18.5 plus a finite check would in fact imply Hypothesis 21.3 *for all* primes p.

From the point of view of this paper, this provides a natural justification for regarding Hypothesis 21.3 (and the stronger version used in our global log–representation) as "Montgomery-level" assumptions: they are in line with what one expects from the conjectural square-root behaviour of primes in arithmetic progressions, and they follow (for prime moduli) from one of the central conjectures in analytic number theory about the distribution of primes in residue classes.

Conversely, in Section 18.2 we observed that a sufficiently strong bounded log–representation with the range restriction $q \le P_1(n)^2$ actually implies Hypothesis 21.3. Thus our logarithmic framework sits precisely at the interface between deep conjectures on primes in arithmetic progressions (such as Montgomery's) and structural constraints on representing $\log n$ in terms of the "basis logs" $\log(q - 1)$.

18.2 A Euclidean proof of infinitely many LI primes, conditional on the log–conjecture

In this subsection we explain how the bounded log–representation hypothesis implies the existence of infinitely many linearly independent primes, in a way directly reminiscent of Euclid's original proof of the infinitude of primes.

From the log–conjecture to the H–conjecture

Recall Hypothesis 18.1 (bounded log–representation):

Hypothesis 18.7 (Bounded log–representation)**.** For every integer $n \geq 2$ there exists a finite set of primes $\mathcal{Q}(n) \subset \{q \text{ prime} : q \leq P_1(n)^2\}$ and integers $a_{n,q} \in \mathbb{Z}$ such that

$$\log n \;=\; \sum_{q \in \mathcal{Q}(n)} a_{n,q} \log(q-1).$$

As shown in Proposition 18.2, this hypothesis implies the H–conjecture:

Conjecture 18.8 (H–conjecture (prime case))**.** For every prime p there exists a prime q and an integer k such that

$$q = kp + 1, \qquad 1 \leq k \leq p-1.$$

For convenience we restate the conclusion we need.

Proposition 18.9. *Assume Hypothesis 18.7. Then for every prime p there exists a prime q and an integer k with*

$$q = kp + 1, \qquad 1 \leq k \leq p-1.$$

In particular, for each prime p there exists at least one prime $q \equiv 1 \pmod{p}$.

Proof. This is exactly Proposition 18.2 specialised to the prime case $n = p$. For completeness, we sketch the argument.

Fix a prime p. By Hypothesis 18.7 we can write

$$\log p = \sum_{q \leq p^2} a_q \log(q-1)$$

with integer coefficients a_q, only finitely many nonzero. Exponentiating gives

$$p = \prod_{q \leq p^2} (q-1)^{a_q}.$$

Separating positive and negative exponents (writing $a_q = a_q^+ - a_q^-$ with $a_q^\pm \geq 0$) leads to an integer identity

$$p \cdot \prod_{q \leq p^2} (q-1)^{a_q^-} \;=\; \prod_{q \leq p^2} (q-1)^{a_q^+}.$$

Since p divides the right-hand side, there must be some q with $a_q^+ > 0$ such that $p \mid (q-1)$, i.e. $q \equiv 1 \pmod{p}$. Writing $q - 1 = kp$ gives an integer $k \geq 1$ with $q = kp + 1$. The bound $q \leq p^2$ then implies $k \leq p-1$. This is exactly the statement of the H–conjecture for p. □

Minimal primes in 1 mod p **are LI**

We proved the following purely combinatorial fact about the φ–vectors; we recall it here.

Proposition 18.10 (Minimal 1 mod p primes are linearly independent)**.** *Let p be a fixed prime, and let q be the* smallest *prime satisfying*

$$q \equiv 1 \pmod{p}.$$

Then q is linearly independent in the sense of the φ–vectors; that is, the vector $\varphi(q)$ does not lie in the $\mathbb{Q}$–span of $\{\varphi(r) : r < q\}$.

The proof uses only p–adic valuations: since $p \mid (q-1)$, the p–coordinate of $\varphi(q)$ is nonzero, whereas by minimality of q no earlier prime $r < q$ satisfies $p \mid (r-1)$, so all $\varphi(r)$ for $r < q$ have p–coordinate 0. Hence $\varphi(q)$ cannot be a $\mathbb{Q}$–linear combination of them.

Euclidean argument: from any finite set to a larger LI prime

We now combine Proposition 18.9 and Proposition 18.10 to give a Euclid-style proof of the infinitude of linearly independent primes, conditional on the bounded log–conjecture.

Theorem 18.11 (Infinitely many LI primes under the log–conjecture)**.** *Assume Hypothesis 18.7. Then there exist infinitely many linearly independent primes.*

Proof. Suppose, for contradiction, that there are only finitely many linearly independent primes. Let

$$q_1 < q_2 < \cdots < q_r$$

be the complete list of LI primes, and set

$$Q := q_r$$

to be the largest LI prime.

Since there are infinitely many primes in total, there exists at least one prime strictly larger than Q; let p be the *smallest* prime with

$$p > Q.$$

By definition of Q, this prime p is linearly dependent.

Now apply Proposition 18.9 to this prime p. It gives us at least one prime q such that

$$q = kp + 1, \qquad 1 \le k \le p - 1,$$

in particular $q \equiv 1 \pmod{p}$.

Since there is at least one prime in the progression $1 \bmod p$, by the well-ordering principle there is a *smallest* such prime; denote it by $q_{\min}$. By definition,

$$q_{\min} \equiv 1 \pmod{p}, \qquad q_{\min} \le q.$$

But $q = kp + 1$ with $k \ge 1$ implies $q > p$, and hence

$$q_{\min} \ge p + 1 > p > Q.$$

In particular, $q_{\min}$ is a prime larger than Q.

By Proposition 18.10, the smallest prime $q_{\min} \equiv 1 \pmod{p}$ is linearly independent. Thus $q_{\min}$ is a linearly independent prime strictly larger than Q, contradicting the maximality of Q.

This contradiction shows that our assumption was false: there cannot be only finitely many linearly independent primes. Therefore there exist infinitely many linearly independent primes. □

An explicit Euclidean chain of LI primes

The proof above can be made into an explicit infinite sequence of linearly independent primes, entirely analogous to Euclid's iterative construction.

Assume again Hypothesis 18.7, so that Proposition 18.9 and Proposition 18.10 hold. Define sequences $(p_j)_{j\ge1}$ and $(q_j)_{j\ge1}$ inductively as follows:

- Start with $p_1 := 3$.
- Given p_j, let q_j be the smallest prime congruent to 1 modulo p_j.

- Let p_{j+1} be the smallest prime strictly larger than q_j.

By Proposition 18.9, for each p_j the progression 1 mod p_j contains at least one prime, so q_j is well defined. By Proposition 18.10, each q_j is linearly independent. Moreover we have

$$3 = p_1 < q_1 < p_2 < q_2 < p_3 < q_3 < \dots,$$

so in particular the sequence $(q_j)_{j\geq 1}$ is an infinite strictly increasing sequence of linearly independent primes.

This provides, conditional on the bounded log–conjecture, a third proof of the infinitude of LI primes, alongside the Dirichlet-based argument and Wojowu's combinatorial proof. Conceptually it is closest to Euclid's original proof: from any finite "initial segment" of LI primes, one produces a strictly larger LI prime by passing to the smallest prime in a suitably chosen congruence class.

19 An injective successor map and infinite chains of LI primes

In this section we work under the H–hypothesis in its prime form, and show that it induces an *injective* successor map on the set of primes. If we additionally choose, for each prime p, the *smallest* prime $q \equiv 1 \pmod{p}$, then every successor is linearly independent, and each starting prime generates an infinite chain of linearly independent primes.

19.1 Hypothesis H and the successor map

Hypothesis 19.1 (H for primes)**.** For every prime number p there exist an integer k and a prime q such that

$$q = kp + 1, \qquad 1 \leq k \leq p - 1.$$

Given Hypothesis 21.3, there may be many primes of the form $kp+1$ with $1 \leq k \leq p-1$. For definiteness we choose the smallest one.

Definition 19.2 (Successor map on primes)**.** Assume Hypothesis 21.3. For each prime p define $q(p)$ to be the smallest prime of the form

$$q(p) = kp + 1, \qquad 1 \leq k \leq p - 1,$$

and set

$$\Phi(p) := q(p).$$

We call $\Phi\colon \{\text{primes}\} \to \{\text{primes}\}$ the *successor map*.

By construction $\Phi(p) > p$ for every prime $p \geq 2$, since $\Phi(p) \geq p + 1$.

19.2 Injectivity of the successor map

We now show that the map $p \mapsto \Phi(p)$ is injective. The key point is that different primes cannot share the same representation $q = kp + 1$ with $1 \leq k \leq p - 1$.

Proposition 19.3 (Injectivity of Φ)**.** *Assume Hypothesis 21.3. Then the successor map $\Phi\colon p \mapsto q(p)$ is injective on the set of primes. In other words, if $\Phi(p) = \Phi(p')$ for primes p, p', then $p = p'$.*

Proof. Let p and p' be primes, and suppose $\Phi(p) = \Phi(p') = q$. By definition of Φ there exist integers k, k' with

$$1 \le k \le p-1, \quad 1 \le k' \le p'-1,$$

such that

$$q = kp + 1 = k'p' + 1.$$

Subtracting 1 from both sides gives

$$kp = k'p'.$$

Since p and p' are prime numbers, there are only two possibilities:

- either $p = p'$, in which case we are done;
- or $p \neq p'$, in which case p divides k' and p' divides k.

Assume for contradiction that $p \neq p'$. Without loss of generality suppose $p < p'$. Then p is a proper divisor of k', so $k' \ge p$. But by definition of $\Phi(p')$ we also have $k' \le p' - 1$, hence

$$p \;\le\; k' \;\le\; p' - 1.$$

On the other hand, from $kp = k'p'$ we obtain

$$k = \frac{k'p'}{p} \;\ge\; \frac{p \cdot p'}{p} = p',$$

so $k \ge p'$. Combining this with $k \le p - 1$ gives

$$p' \;\le\; k \;\le\; p - 1,$$

which is impossible since $p' < p$ by assumption. This contradiction shows that the case $p \neq p'$ cannot occur. Therefore we must have $p = p'$.

Hence Φ is injective on the set of primes. □

Remark 19.4. The proof did not use any minimality of $q(p)$ beyond $1 \le k \le p - 1$; it relies only on the shape $q = kp + 1$ with $k < p$ and the primality of p, p', q.

Since $\Phi(p) > p$ for every prime p, injectivity immediately rules out cycles:

Corollary 19.5 (No cycles)**.** *Under Hypothesis 21.3, the iterates of Φ have no cycles. More precisely, for any prime p_0 the sequence*

$$p_0, \quad p_1 := \Phi(p_0), \quad p_2 := \Phi(p_1), \ \ldots$$

is strictly increasing and consists of pairwise distinct primes.

Proof. We have $\Phi(p) > p$ for every prime p, hence $p_{n+1} > p_n$ for all $n \ge 0$, so the sequence is strictly increasing and in particular has no repetitions. □

Thus each prime p_0 lies at the start of an infinite chain

$$p_0 < p_1 < p_2 < \ldots, \qquad p_{n+1} = \Phi(p_n),$$

of pairwise distinct primes.

19.3 Linear independence of successors and infinite LI chains

We now recall the notion of linear independence for primes, as defined earlier via the vectors

$$\varphi(p_k) := \big(v_{p_1}(p_k - 1), \dots, v_{p_{k-1}}(p_k - 1)\big) \in \mathbb{Z}^{k-1},$$

where $p_1 = 2, p_2 = 3, \dots$ denote the primes in increasing order, and $v_{p_i}(\cdot)$ is the p_i-adic valuation. A prime p_k is called *linearly independent* (LI) if $\varphi(p_k)$ does not lie in the $\mathbb{Q}$-span of

$$\varphi(p_1), \dots, \varphi(p_{k-1}),$$

and *linearly dependent* (LD) otherwise.

The following lemma, proved earlier in a slightly more general form, shows that *minimal* primes in the progression $1 \pmod p$ are always linearly independent.

Lemma 19.6 (Minimal primes 1 mod p are LI)**.** *Let p be a prime, and let q be the smallest prime such that*

$$q \equiv 1 \pmod p.$$

Then q is linearly independent.

Proof. Write $q = p_m$ in the increasing sequence of primes, so $p_m = q$ and $\varphi(q) \in \mathbb{Z}^{m-1}$. Let $p_i = p$ for some $i < m$. Since $q \equiv 1 \pmod p$, we have $p \mid (q - 1)$ and hence

$$v_p(q - 1) = v_{p_i}(q - 1) \geq 1,$$

so the i-th coordinate of $\varphi(q)$ is nonzero.

If $r < p_m$ is any smaller prime, then $r \not\equiv 1 \pmod p$ by the minimality of q. Hence $p \nmid (r - 1)$ and thus $v_p(r - 1) = 0$, so the i-th coordinate of $\varphi(r)$ is zero.

Suppose for contradiction that $\varphi(q)$ lies in the $\mathbb{Q}$-span of $\{\varphi(p_k) : k < m\}$. Then there exist rationals c_k such that

$$\varphi(q) = \sum_{k<m} c_k \, \varphi(p_k).$$

Comparing the i-th coordinates of both sides gives

$$v_p(q - 1) = \sum_{k<m} c_k \, v_p(p_k - 1) = \sum_{k<m} c_k \cdot 0 = 0,$$

which contradicts $v_p(q - 1) \geq 1$. Thus $\varphi(q)$ is not in the span of the preceding vectors, and q is linearly independent. □

By Definition 19.2, for each prime p the successor $q(p) = \Phi(p)$ is the *smallest* prime of the form $kp + 1$, so Lemma 20.2 applies.

Corollary 19.7 (Successors are linearly independent)**.** *Assume Hypothesis 21.3. For every prime p, the successor $q(p) = \Phi(p)$ is a linearly independent prime.*

Proof. By Definition 19.2, $q(p)$ is the smallest prime with $q(p) \equiv 1 \pmod p$, so Lemma 20.2 gives that $q(p)$ is linearly independent. □

Combining Corollary 19.5 and Corollary 19.7 we obtain the desired statement about infinite chains of LI primes.

Corollary 19.8 (Every prime generates an infinite LI chain)**.** *Assume Hypothesis 21.3. Let p_0 be any prime number (linearly independent or linearly dependent). Define recursively*

$$p_{n+1} := \Phi(p_n) \qquad (n \geq 0).$$

Then:

1. *The sequence $(p_n)_{n\geq 0}$ is strictly increasing and consists of pairwise distinct primes.*
2. *For every $n \geq 1$, the prime p_n is linearly independent.*

In particular, each prime p_0 gives rise (under Hypothesis 21.3) to an infinite chain

$$p_0 < p_1 < p_2 < \dots$$

of linearly independent primes.

Proof. Part (1) is exactly Corollary 19.5. For (2), we have $p_{n+1} = \Phi(p_n)$, so by Corollary 19.7 each p_{n+1} is linearly independent. Thus all terms p_n with $n \geq 1$ are LI. □

Remark 19.9. Corollary 19.8 gives, conditional on the H–hypothesis, yet another proof of the infinitude of linearly independent primes: from any starting prime p_0 (for instance $p_0 = 2$) we obtain an infinite strictly increasing sequence of primes, all of which are linearly independent from the second term onward.

20 Prime chains under Hypothesis (H)

In this section we assume the following strengthening of Dirichlet's theorem.

Hypothesis 20.1 (H)**.** For every prime number p there exists a prime q and an integer k with

$$q = kp + 1, \qquad 1 \leq k \leq p - 1,$$

such that q is the *smallest* prime congruent to $1 \pmod p$. Equivalently: for each prime p, let

$$q(p) := \min\{\, q \text{ prime} : q \equiv 1 \pmod p \,\};$$

then Hypothesis 20.1 asserts that

$$q(p) = kp + 1 \quad \text{with} \quad 1 \leq k \leq p - 1.$$

Thus we obtain a well-defined *successor map*

$$\varphi : \{\text{primes}\} \longrightarrow \{\text{primes}\}, \qquad \varphi(p) := q(p),$$

sending a prime p to the minimal prime congruent to $1 \pmod p$.

We continue to use the notion of linearly independent / dependent primes introduced earlier: a prime q is called *linearly independent* (LI) if its exponent vector $v(q-1) = (v_r(q-1))_{r \text{ prime}}$ does not lie in the $\mathbb{Q}$–span of the vectors $v(r-1)$ for smaller primes $r < q$; otherwise q is *linearly dependent* (LD).

20.1 Successors are always linearly independent

We begin by recalling a key lemma from the earlier discussion.

Lemma 20.2 (Minimal primes modulo p are LI)**.** *Let p be a prime, and let q be the smallest prime with $q \equiv 1 \pmod{p}$. Then q is linearly independent.*

Proof. Since $q \equiv 1 \pmod{p}$ we have $p \mid (q-1)$ and hence $v_p(q-1) \geq 1$. By minimality of q, no smaller prime $r < q$ satisfies $r \equiv 1 \pmod{p}$, so $p \nmid (r-1)$ for all $r < q$ and therefore $v_p(r-1) = 0$ for all such r.

Consider the exponent vectors $v(r-1)$ as elements of the vector space $V = \bigoplus_{s \text{ prime}} \mathbb{Q} e_s$. The p-coordinate of $v(q-1)$ is $v_p(q-1) \geq 1$, whereas the p-coordinate of every $v(r-1)$ with $r < q$ is zero. Hence $v(q-1)$ cannot lie in the $\mathbb{Q}$–span of $\{v(r-1) : r < q\}$, and q is linearly independent. □

Immediate consequence:

Proposition 20.3 (Linearly dependent primes are never successors)**.** *Under Hypothesis 20.1, if a prime p is linearly dependent, then p is not in the image of φ. Equivalently, there is no prime r with*

$$\varphi(r) = p.$$

Proof. Suppose, for contradiction, that p is linearly dependent and that there is a prime r with $\varphi(r) = p$. By definition of φ, this means that p is the smallest prime with $p \equiv 1 \pmod{r}$, so p is a minimal prime in the progression $1 \bmod r$. By Lemma 20.2, p is then linearly independent, contradicting our assumption that p is LD. Thus no LD prime lies in the image of φ. □

This proves the informal statement:

A linearly dependent prime can never appear as a successor $\varphi(r)$.

20.2 Every linearly independent prime has a predecessor

We now show the converse: every linearly independent prime q is in the image of φ, i.e. it is the successor of some (unique) smaller prime.

To make this precise, recall the following notation from the matrix formalism. Let $(p_k)_{k \geq 1}$ be the increasing sequence of primes, and for each prime p_i let

$$t(i) := \min\{\, k \geq 1 : p_i \mid (p_k - 1) \,\}$$

be the index of the first prime p_k whose predecessor $p_k - 1$ is divisible by p_i (if no such k exists, set $t(i) := \infty$).

Lemma 20.4. *Let $q = p_n$ be a linearly independent prime. Then there exists at least one prime p_i dividing $q-1$ such that*

$$t(i) = n,$$

i.e. q is the first *prime for which $p_i \mid (p_k - 1)$.*

Proof. As recalled earlier, p_n is linearly independent if and only if the rank of the valuation matrix increases from $n-1$ to n, equivalently if at least one new p_i appears as a divisor of some $p_k - 1$ for the first time at $k = n$. Concretely, this means exactly that there is a prime p_i with $t(i) = n$, and such a prime p_i must divide $p_n - 1 = q - 1$. □

For such a prime p_i we have:

Lemma 20.5 (Predecessor of an LI prime)**.** *Let $q = p_n$ be linearly independent, and choose a prime $p = p_i$ with $t(i) = n$ as in Lemma 20.4. Then q is the smallest prime with $q \equiv 1 \pmod{p}$, i.e.*

$$\varphi(p) = q.$$

Proof. Since $p \mid (q-1)$ we have $q \equiv 1 \pmod{p}$. If there were a smaller prime $r < q$ with $r \equiv 1 \pmod{p}$, then $p \mid (r-1)$ and hence $t(i)$, the first occurrence index of p, would satisfy $t(i) \leq \mathrm{index}(r) < n$, contradicting $t(i) = n$. Thus no smaller prime $r < q$ satisfies $r \equiv 1 \pmod{p}$, and q is indeed the minimal prime in the residue class $1 \bmod p$. By the definition of φ, we have $\varphi(p) = q$. □

Combining Lemmas 20.4 and 20.5 we obtain:

Proposition 20.6 (Every LI prime is a successor)**.** *Under Hypothesis 20.1, for every linearly independent prime q there exists a prime $p < q$ such that*

$$\varphi(p) = q.$$

In particular, every LI prime lies in the image of φ.

Remark 20.7. In many computed examples, the "predecessor prime" p given by Lemma 20.5 coincides with the largest prime factor $P_1(q-1)$ of $q-1$. It is natural to ask whether this always holds, i.e. whether one always has

$$q \text{ LI} \quad \Longrightarrow \quad \varphi(P_1(q-1)) = q.$$

At present we do *not* know how to prove this stronger statement, and we therefore work only with the (provable) existence of some predecessor prime p dividing $q-1$.

For the purposes of this section, Propositions 20.3 and 20.6 are exactly what we need: they identify the image of φ as the set of linearly independent primes, and its complement as the set of linearly dependent primes.

20.3 Decomposition into disjoint infinite prime chains

We now combine Hypothesis 20.1 with the previous propositions to obtain a global structural picture of the primes.

Theorem 20.8 (Prime chains)**.** *Assume Hypothesis 20.1. Then:*

1. *The map φ sends primes to strictly larger primes:*

$$\varphi(p) > p \quad \text{for all primes } p.$$

2. *Every linearly independent prime lies in the image of φ, and no linearly dependent prime lies in the image of φ.*

3. *The set of all primes decomposes into disjoint infinite chains*

$$p_0 \longmapsto p_1 := \varphi(p_0) \longmapsto p_2 := \varphi(p_1) \longmapsto p_3 := \varphi(p_2) \longmapsto \cdots,$$

where the starting primes p_0 are exactly *the linearly dependent primes, and all subsequent primes p_i in the chain are linearly independent.*

Proof. (1) For each prime p, the definition of $\varphi(p)$ as the smallest prime congruent to 1 (mod p) implies $\varphi(p) \geq p+1$, hence $\varphi(p) > p$.

(2) This is precisely the content of Propositions 20.3 and 20.6:

- Proposition 20.3 says that no LD prime lies in the image of φ.
- Proposition 20.6 says that every LI prime does lie in the image of φ.

Therefore the image of φ is exactly the set of linearly independent primes, and the set of linearly dependent primes is exactly the complement of the image.

(3) Since $\varphi(p) > p$ for all primes p, the forward orbit

$$p,\ \varphi(p),\ \varphi(\varphi(p)), \dots$$

is a strictly increasing sequence of primes and hence cannot contain any cycles. Starting from any prime p_0 and iterating φ we therefore obtain a chain

$$p_0 \to p_1 \to p_2 \to \cdots$$

in which all p_i are distinct and $p_i < p_{i+1}$.

By (2), the primes that *do not* lie in the image of φ are exactly the linearly dependent primes. These are the primes that can only occur as the starting points p_0 of chains. All other primes occur somewhere as $\varphi(p)$ for a smaller prime p, hence appear as non-initial elements of a chain.

It remains to see that each chain is infinite. Suppose for contradiction that there were a finite chain

$$p_0 \to p_1 \to \cdots \to p_T$$

with no successor $\varphi(p_T)$ defined. Under Hypothesis 20.1, however, $\varphi(p_T)$ is defined for *every* prime p_T, so the chain can always be extended one step further. Thus each chain must be infinite in the forward direction.

Since every prime either (i) is linearly dependent and hence a unique starting point of such a chain, or (ii) is linearly independent and hence occurs exactly once as a successor $\varphi(p)$ inside some chain, the set of all primes is partitioned into disjoint infinite chains as claimed. □

Corollary 20.9 (Every LD prime generates infinitely many LI primes)**.** *Under Hypothesis 20.1, each linearly dependent prime p_0 is the start of an infinite chain*

$$p_0 \to p_1 \to p_2 \to \cdots$$

in which all p_i for $i \geq 1$ are linearly independent. In particular, each LD prime gives rise to an infinite sequence of distinct LI primes.

Proof. If p_0 is LD, then by Proposition 20.3 it is not in the image of φ, so it can only occur as a starting point of a chain. By Theorem 20.8, this chain is infinite and all successors $p_i = \varphi(p_{i-1})$ are LI. □

Remark 20.10. Theorem 20.8 gives, under Hypothesis 20.1, an *Euclid-style* proof of the infinitude of linearly independent primes: pick any LD prime (for instance the smallest LD prime) as p_0; the chain starting at p_0 then consists of infinitely many distinct LI primes $p_1, p_2, \dots$. This provides a third, conditional, proof of the infinitude of LI primes, complementing the unconditional arguments based on Dirichlet's theorem and on Wojowu's p-adic counting argument.

21 Successor inequalities and a lower bound for $\xi(x)$

In this section we formalize the inequalities

$$\xi\Big(\max_{1\le k\le r}\varphi(p_k)\Big)\ \ge\ r \tag{10}$$

and

$$\xi\Big(\max_{p\le x}\varphi(p)\Big)\ \ge\ \pi(x), \tag{11}$$

where p_k is the k-th prime, $\pi(x)$ is the prime-counting function, and

$$\varphi(p) := \min\{\, q \text{ prime} : q \equiv 1 \pmod p \,\}$$

is the least prime in the progression $1 \bmod p$. We then show how, under the strong Hypothesis (H), these inequalities imply the pointwise bound

$$\xi(p^2 - p + 1)\ \ge\ \pi(p)$$

for every prime p, and how the prime number theorem yields a global lower bound for $\xi(x)$ of order $\sqrt{x}/\log x$.

21.1 Equivalence of the two formulations

We first record the equivalence between (10) and (11).

Proposition 21.1. *The following are equivalent:*

(i) For every integer $r \ge 1$,

$$\xi\Big(\max_{1\le k\le r}\varphi(p_k)\Big)\ \ge\ r.$$

(ii) For every real number $x \ge 2$,

$$\xi\Big(\max_{p\le x}\varphi(p)\Big)\ \ge\ \pi(x).$$

Proof. (i) $\Rightarrow$ (ii). Let $x \ge 2$ and put $r := \pi(x)$, so that $p_r \le x < p_{r+1}$ and

$$\{p : p \le x\} = \{p_1, \dots, p_r\}.$$

Hence

$$\max_{p\le x}\varphi(p) = \max_{1\le k\le r}\varphi(p_k),$$

and by (i) we obtain

$$\xi\Big(\max_{p\le x}\varphi(p)\Big) = \xi\Big(\max_{1\le k\le r}\varphi(p_k)\Big)\ \ge\ r = \pi(x).$$

(ii) $\Rightarrow$ (i). Conversely, let $r \ge 1$ and set $x := p_r$. Then $\pi(x) = r$ and

$$\max_{p\le x}\varphi(p) = \max_{p\le p_r}\varphi(p) = \max_{1\le k\le r}\varphi(p_k).$$

Applying (ii) with this x gives

$$\xi\Big(\max_{1\le k\le r}\varphi(p_k)\Big) = \xi\Big(\max_{p\le x}\varphi(p)\Big) \ge \pi(x) = r,$$

which is (i). □

Thus it suffices to work with either formulation; we shall mainly use (10).

21.2 Minimal primes in $1 \bmod p$ are linearly independent

The key structural input is that the least prime in the progression $1 \bmod p$ is always linearly independent in the sense of the φ-vectors. This is already proved earlier in the paper; we restate it here for convenience.

Theorem 21.2 (Minimal primes in $1 \bmod p$ are LI)**.** *Let p be a prime, and let*

$$\varphi(p) := \min\{\, q \text{ prime} : q \equiv 1 \pmod{p} \,\}$$

be the smallest prime congruent to $1 \bmod p$. Then $\varphi(p)$ is linearly independent: its exponent vector $\phi(\varphi(p))$ does not lie in the $\mathbb{Q}$-span of the vectors $\phi(r)$ for primes $r < \varphi(p)$.

Sketch of proof. Since $p \mid (\varphi(p) - 1)$, the p-coordinate of $\phi(\varphi(p))$ is non-zero. By minimality, no smaller prime $r < \varphi(p)$ satisfies $r \equiv 1 \pmod{p}$, so $p \nmid (r-1)$ and the p-coordinate of $\phi(r)$ is zero for all such r. Hence $\phi(\varphi(p))$ cannot lie in the $\mathbb{Q}$-span of $\{\phi(r) : r < \varphi(p)\}$. □

In particular, for every prime p the successor prime $\varphi(p)$ is an LI prime.

21.3 Successor map, injectivity and inequality (10)

We now work under the strengthened least-prime hypothesis introduced earlier.

Hypothesis 21.3 (Prime form of (H))**.** For every prime p there exist an integer k and a prime q such that

$$q = kp + 1, \qquad 1 \le k \le p - 1,$$

and q is the smallest prime congruent to $1 \pmod{p}$. Equivalently,

$$\varphi(p) = q = kp + 1 \quad \text{with } 1 \le k \le p - 1.$$

Under Hypothesis 21.3 the successor map

$$\Phi : \{\text{primes}\} \longrightarrow \{\text{primes}\}, \qquad \Phi(p) := \varphi(p),$$

is particularly well behaved.

Proposition 21.4 (Injectivity of the successor map)**.** *Assume Hypothesis 21.3. Then the map Φ is injective on the set of primes: if $\Phi(p) = \Phi(p')$ for primes p, p', then $p = p'$.*

Proof. By Hypothesis 21.3, if $\Phi(p) = q$ then $q = kp + 1$ with $1 \le k \le p - 1$. Suppose $\Phi(p) = \Phi(p') = q$ for primes p, p'. Then

$$q = kp + 1 = k'p' + 1$$

for some integers k, k' with $1 \le k \le p - 1$ and $1 \le k' \le p' - 1$. Thus

$$kp = k'p'.$$

Since p and p' are prime, either $p = p'$, in which case we are done, or $p \neq p'$, in which case $p \mid k'$ and $p' \mid k$.

Assume $p \neq p'$ and, without loss of generality, $p < p'$. Then $p \mid k'$ forces $k' \ge p$, so $p \le k' \le p' - 1$. From $kp = k'p'$ we get

$$k = \frac{k'p'}{p} \ge \frac{p \cdot p'}{p} = p',$$

so $k \ge p'$, contradicting $k \le p - 1 < p'$. Hence the case $p \neq p'$ is impossible, and we must have $p = p'$. □

Note that Theorem 21.2 and Proposition 21.4 together imply that the map $p \mapsto \varphi(p)$ sends primes to *distinct* LI primes.

We can now prove (10) under Hypothesis 21.3.

Theorem 21.5. *Assume Hypothesis 21.3. Then for every integer $r \geq 1$,*

$$\xi\Big(\max_{1\leq k\leq r} \varphi(p_k)\Big) \;\geq\; r. \tag{12}$$

Equivalently, by Proposition 21.1, for every $x \geq 2$,

$$\xi\Big(\max_{p\leq x} \varphi(p)\Big) \;\geq\; \pi(x).$$

Proof. Fix $r \geq 1$. For each $1 \leq k \leq r$ consider the prime

$$q_k := \varphi(p_k).$$

By Theorem 21.2, each q_k is linearly independent. By Proposition 21.4, the primes $q_1, \ldots, q_r$ are pairwise distinct. Let

$$Q_r := \max_{1\leq k\leq r} q_k = \max_{1\leq k\leq r} \varphi(p_k).$$

Then all q_k lie in the set of LI primes $\leq Q_r$, so

$$\xi(Q_r) = \#\{\text{LI primes } \leq Q_r\} \;\geq\; \#\{q_1, \ldots, q_r\} = r.$$

This is exactly (12). The equivalence with the x-version follows from Proposition 21.1. □

21.4 A corollary of Hypothesis (H): $\xi(p^2 - p + 1) \geq \pi(p)$

Hypothesis 21.3 gives a very concrete upper bound for the successor prime $\varphi(p)$, which in turn yields a pointwise corollary for $\xi(x)$ at quadratic arguments.

Corollary 21.6. *Assume Hypothesis 21.3. Then for every prime p,*

$$\xi(p^2 - p + 1) \;\geq\; \pi(p). \tag{13}$$

Proof. Fix a prime p and consider all primes $r \leq p$. By Hypothesis 21.3, for each such r we have

$$\varphi(r) = k_r r + 1$$

with $1 \leq k_r \leq r - 1$. In particular,

$$\varphi(r) \;\leq\; (r-1)r + 1 = r^2 - r + 1 \;\leq\; p^2 - p + 1,$$

since $r \leq p$ and the polynomial $t^2 - t$ is increasing for $t \geq 1$. Thus every successor prime $\varphi(r)$ with $r \leq p$ lies $\leq p^2 - p + 1$.

By Theorem 21.2, each $\varphi(r)$ is LI, and by Proposition 21.4 the map $r \mapsto \varphi(r)$ is injective on the set of primes. Hence we obtain at least $\pi(p)$ *distinct* LI primes $\varphi(r)$ in the interval $[2, p^2 - p + 1]$. Therefore

$$\xi(p^2 - p + 1) = \#\{\text{LI primes } \leq p^2 - p + 1\} \;\geq\; \pi(p),$$

which is (13). □

In words: under Hypothesis (H), the quadratic window $[2, p^2 - p + 1]$ contains at least as many LI primes as there are ordinary primes $\leq p$.

21.5 A global lower bound via the prime number theorem

We finally use Corollary 21.6 together with the prime number theorem to deduce a global lower bound on $\xi(x)$.

Theorem 21.7. *Assume Hypothesis 21.3 and the prime number theorem. Then there exists a constant $C > 0$ and $x_0 \geq 2$ such that for all $x \geq x_0$,*

$$\xi(x) \;\geq\; C\,\frac{\sqrt{x}}{\log x}. \tag{14}$$

More precisely, for every $\varepsilon > 0$ there exists $x_0(\varepsilon)$ such that

$$\xi(x) \;\geq\; (1-\varepsilon)\,\frac{\sqrt{x}}{\log x} \quad \textit{for all } x \geq x_0(\varepsilon).$$

Proof. Let $x \geq 4$ and set $y := \sqrt{x}$. Choose p to be the largest prime $\leq y$; then $p \to \infty$ as $x \to \infty$ and $p \sim y$ by the prime number theorem. For all sufficiently large x we have

$$p^2 - p + 1 \;\leq\; y^2 = x,$$

so $p^2 - p + 1 \leq x$. By Corollary 21.6,

$$\xi(x) \;\geq\; \xi(p^2 - p + 1) \;\geq\; \pi(p).$$

Now apply the prime number theorem in the form

$$\pi(t) \sim \frac{t}{\log t} \qquad (t \to \infty).$$

In particular, for every $\varepsilon > 0$ there exists $T(\varepsilon)$ such that

$$\pi(t) \;\geq\; (1-\varepsilon)\,\frac{t}{\log t} \quad \text{for all } t \geq T(\varepsilon).$$

Taking $t = p$ and using $p \sim y = \sqrt{x}$ and $\log p \sim \tfrac{1}{2}\log x$, we obtain, for x large enough (depending on ε),

$$\pi(p) \;\geq\; (1-\varepsilon)\,\frac{p}{\log p} \;\gg\; (1-\varepsilon)\,\frac{\sqrt{x}}{\log x},$$

where the implicit constant can be absorbed into $(1-\varepsilon)$ by further increasing $x_0(\varepsilon)$ if necessary.

Combining this with $\xi(x) \geq \pi(p)$ we obtain

$$\xi(x) \;\geq\; (1-\varepsilon)\,\frac{\sqrt{x}}{\log x}$$

for all $x \geq x_0(\varepsilon)$, which is the claimed bound. □

Thus, conditional on Hypothesis (H), the set of linearly independent primes has asymptotic density zero among all primes, but it is still large enough that

$$\xi(x) \;\gg\; \frac{\sqrt{x}}{\log x},$$

which is a non-trivial global lower bound for the growth of $\xi(x)$.

22 A fourth unconditional proof via Murthy's theorem

In this section we give a fourth proof that there are infinitely many linearly independent primes. In contrast to the Dirichlet-based argument, this proof is entirely elementary and rests on a result of Murthy on prime divisors of certain repunit numbers.

Throughout we keep the notation and definition of linearly independent primes from the previous sections: if $p_1 < p_2 < \cdots$ denotes the sequence of primes, then to each prime p_k we associate the valuation vector

$$\varphi(p_k) := (v_{p_1}(p_k - 1), v_{p_2}(p_k - 1), \ldots, v_{p_{k-1}}(p_k - 1)) \in \mathbb{Z}^{k-1},$$

and we call p_k *linearly independent* (LI) if $\varphi(p_k)$ does not lie in the $\mathbb{Q}$-span of $\varphi(p_1), \ldots, \varphi(p_{k-1})$.

22.1 Murthy's theorem on primes congruent to 1 modulo a prime

We begin by recalling a result of Murthy [10]. For a positive integer m let

$$I(m) := \underbrace{11\ldots1}_{m \text{ digits}} = \frac{10^m - 1}{9}$$

be the repunit with m digits 1. Let P_n be the n-th prime, so that $P_1 = 2$, $P_2 = 3$, $P_3 = 5$, and so on, and define

$$u(n) := I(P_n).$$

Murthy proves the following.

Theorem 22.1 (Murthy)**.** *For every prime p there exists a prime q such that*

$$q \equiv 1 \pmod{p}.$$

More precisely, if $p = P_n$ is the n-th prime and q is any prime divisor of $u(n) = I(P_n)$, then $q \equiv 1 \pmod{p}$.

For completeness we recall the short proof.

Proof (after Murthy). Let $n \geq 1$ and $p = P_n$. Let q be a prime divisor of $u(n) = I(P_n) = (10^{P_n} - 1)/9$. Then $q \mid 10^{P_n} - 1$, hence $10^{P_n} \equiv 1 \pmod{q}$. In particular $\gcd(q, 10) = 1$, so by Fermat's little theorem also $10^{q-1} \equiv 1 \pmod{q}$, i.e. $q \mid 10^{q-1} - 1$. Again using $\gcd(q, 9) = 1$ we obtain $q \mid I(q-1) = (10^{q-1} - 1)/9$.

Thus q divides both $I(P_n)$ and $I(q-1)$. It is well known (and easy to check) that

$$\gcd(10^a - 1, 10^b - 1) = 10^{\gcd(a,b)} - 1 \quad \text{for all } a, b \geq 1,$$

and therefore, after dividing by 9,

$$\gcd\big(I(a), I(b)\big) = I\big(\gcd(a, b)\big).$$

Applying this with $a = P_n$ and $b = q - 1$ we deduce

$$q \mid \gcd\big(I(P_n), I(q-1)\big) = I\big(\gcd(P_n, q-1)\big).$$

Since P_n is prime, we have $\gcd(P_n, q-1) \in \{1, P_n\}$. If $\gcd(P_n, q-1) = 1$, then $q \mid I(1) = 1$, which is impossible. Hence $\gcd(P_n, q-1) = P_n$ and therefore $P_n \mid (q-1)$, i.e. $q \equiv 1 \pmod{P_n}$.

Finally, for any given prime p we can write $p = P_n$ and choose q as above. This yields the stated existence of a prime q with $q \equiv 1 \pmod{p}$. □

In particular, combining Theorem 22.1 with the trivial cases $p = 2, 3$, we obtain:

Corollary 22.2. *For every prime p there exists at least one prime $q > p$ with*

$$q \equiv 1 \pmod{p}.$$

22.2 Minimal primes $q \equiv 1 \pmod{p}$ are linearly independent

For each prime p let

$$\Phi(p)$$

denote the smallest prime q such that $q \equiv 1 \pmod{p}$, whenever such a prime exists. By Corollary 22.2, $\Phi(p)$ is well-defined for every prime p.

The next proposition shows that these minimal successors are always linearly independent in the sense of our valuation vectors.

Proposition 22.3. *Let p be a prime and let $q = \Phi(p)$ be the smallest prime with $q \equiv 1 \pmod{p}$. Then q is linearly independent.*

Proof. Write $q = p_k$ in the global ordering $p_1 < p_2 < \dots$ of the primes. By definition of the valuation vector we have

$$\varphi(q) = \big(v_{p_1}(q-1), \dots, v_{p_{k-1}}(q-1)\big).$$

Since $p \mid (q-1)$, the p-coordinate of $\varphi(q)$ is non-zero:

$$v_p(q-1) \geq 1.$$

On the other hand, if r is any prime with $r < q$, then by minimality of q we have $r \not\equiv 1 \pmod{p}$. In particular $p \nmid (r-1)$, so

$$v_p(r-1) = 0.$$

Thus for every $r < q$ the p-coordinate of $\varphi(r)$ is zero, whereas the p-coordinate of $\varphi(q)$ is non-zero.

Now suppose for contradiction that $\varphi(q)$ were in the $\mathbb{Q}$-span of the previous vectors, i.e.

$$\varphi(q) = \sum_{r<q} \lambda_r \, \varphi(r) \quad \text{with } \lambda_r \in \mathbb{Q}.$$

Comparing the p-coordinate on both sides, the right-hand side has p-coordinate

$$\sum_{r<q} \lambda_r \, v_p(r-1) = 0,$$

while the left-hand side has p-coordinate $v_p(q-1) \geq 1$. This is a contradiction. Hence $\varphi(q)$ is not in the $\mathbb{Q}$-span of $\{\varphi(r) : r < q\}$, and q is linearly independent. □

22.3 Infinitely many linearly independent primes

We can now combine Murthy's theorem with Proposition 22.3 to obtain another unconditional proof of the infinitude of linearly independent primes.

Theorem 22.4. *There are infinitely many linearly independent primes.*

Proof. Let p_0 be any fixed prime (for instance $p_0 = 5$) and define inductively a sequence of primes $(p_n)_{n \geq 0}$ by

$$p_{n+1} := \Phi(p_n),$$

where $\Phi(p)$ is the smallest prime q with $q \equiv 1 \pmod{p}$.

By Corollary 22.2, for each prime p_n the successor $p_{n+1} = \Phi(p_n)$ exists, so this recursion defines p_n for every $n \geq 0$. Moreover $p_{n+1} > p_n$ for all n, since $p_n \equiv 1 \pmod{p_n}$ and p_n is itself the smallest positive multiple of p_n, so p_n cannot be congruent to 1 modulo p_n.

Thus $(p_n)_{n\geq 0}$ is a strictly increasing sequence of primes. By Proposition 22.3, each $p_{n+1} = \Phi(p_n)$ is linearly independent. Hence the sequence (p_n) gives an infinite family of distinct linearly independent primes.

Therefore there exist infinitely many linearly independent primes. □

Remark 22.5. This argument uses only Murthy's existence result for primes $q \equiv 1 \pmod{p}$ and the very elementary valuation-based observation of Proposition 22.3. In particular, it does not appeal to Dirichlet's theorem on primes in arithmetic progressions and is therefore logically independent of the Dirichlet-based proof given earlier in the paper.

23 Prime chains, the map $\nu(p)$, and linearly dependent primes

In this section we work under Hypothesis (H). Under (H) we have the successor map

$$\Phi(p) \;=\; \min\{\, q \text{ prime} : q \equiv 1 \pmod{p} \,\},$$

well–defined for every prime p, and satisfying the bound $q = \Phi(p) \leq p^2 - 1$. The structural results on prime chains (Theorem 20.8) are proved conditional on (H). We recall the parts that we shall use and then introduce the map $\nu(p)$.

23.1 Successor chains and LD starting points (under (H))

For a prime p_0 we define its (successor) chain

$$C(p_0) \;:=\; \{\, p_0,\ \Phi(p_0),\ \Phi^2(p_0), \dots \}.$$

Because $\Phi(p) > p$ for all primes p, each chain $C(p_0)$ is strictly increasing and infinite.

The following is a reformulation of Theorem 20.8, proved in Section 20 under Hypothesis (H).

Theorem 23.1 (Prime chain decomposition under (H))**.** *Assume Hypothesis (H). Then the set of all primes can be written as a disjoint union of infinite chains*

$$C(d) \;=\; \{\, d,\ \Phi(d),\ \Phi^2(d), \dots \},$$

indexed by the linearly dependent primes d. More precisely:

1. *A prime d is linearly dependent if and only if it is not in the image of Φ. Equivalently, the linearly dependent primes are precisely the starting points of chains.*
2. *If p is linearly independent, then there exists a unique prime $d < p$ such that $p \in C(d)$, that is, $p = \Phi^k(d)$ for a unique $k \geq 1$.*
3. *The chains $C(d)$, as d runs over all linearly dependent primes, are pairwise disjoint and cover all primes.*

In particular, under (H) every prime belongs to a unique chain, whose smallest element is a unique linearly dependent prime.

23.2 The map $\nu(p)$ and its basic property

For a given prime p we consider the forward orbits

$$S(r) \;:=\; \{\Phi^k(r) : k = 0, 1, 2, \dots\} \;=\; C(r), \qquad r \text{ prime},$$

and form the union of chains started not above p:

$$U(p) \;:=\; \bigcup_{r \le p} S(r) \;=\; \bigcup_{r \le p} C(r).$$

We are interested in primes that are *not* contained in $U(p)$, i.e. primes whose entire chain starts strictly above p.

Definition 23.2. Let p be a prime. If there exists at least one prime q not contained in $U(p)$, we set

$$\nu(p) \;:=\; \min\{\, q \text{ prime} : q \notin U(p)\}.$$

Thus $\nu(p)$, when it exists, is the smallest prime that does not lie in any of the chains started at primes $\le p$.

We can describe $U(p)$ more intrinsically using Theorem 23.1.

Lemma 23.3. *Assume (H). For every prime p we have*

$$U(p) \;=\; \bigcup_{\substack{d \text{ LD prime} \\ d \le p}} C(d),$$

that is, $U(p)$ is the union of all chains whose (linearly dependent) starting point is at most p.

Proof. Let q be any prime. By Theorem 23.1, there exists a unique linearly dependent prime d such that $q \in C(d)$, and d is the smallest element of $C(d)$.

If $d \le p$, then d is one of the primes $r \le p$, and we have $C(d) = S(d) \subset U(p)$. Hence $q \in U(p)$. Conversely, if $d > p$, then $C(d)$ contributes nowhere to $U(p)$, since we only union chains $C(r)$ with $r \le p$. Thus no element of $C(d)$ lies in $U(p)$ in that case.

Collecting these observations for all q proves the stated identity. □

From this description we obtain a very simple characterization of $\nu(p)$ and, in particular, its linear dependence.

Proposition 23.4. *Assume Hypothesis (H) and suppose that $\nu(p)$ is defined. Then $\nu(p)$ is a linearly dependent prime. More precisely,*

$$\nu(p) \;=\; \min\{\, d \text{ LD prime} : d > p\}.$$

Proof. By Lemma 23.3, a prime q lies in $U(p)$ if and only if its chain starts from some LD prime $d \le p$, i.e. $q \in C(d)$ for an LD prime $d \le p$.

Thus a prime q fails to lie in $U(p)$ if and only if it belongs to a chain $C(d)$ whose starting point d satisfies $d > p$. In such a chain $C(d)$ the smallest prime is d itself (because the chain is strictly increasing). Hence:

- For every chain $C(d)$ with $d > p$ the smallest prime outside $U(p)$ coming from this chain is exactly d.

- Among all primes outside $U(p)$, the smallest one is therefore the smallest LD prime $d > p$.

By Definition 23.2 we have

$$\nu(p) \;=\; \min\{\, q \text{ prime} : q \notin U(p)\},$$

and combining this with the discussion above yields

$$\nu(p) \;=\; \min\{\, d \text{ LD prime} : d > p\}.$$

In particular $\nu(p)$ is itself a starting point of a chain, hence is not in the image of Φ and thus is linearly dependent by Theorem 23.1(1). □

24 An algorithmic characterization of linearly independent primes

In this section we justify the correctness of the following simple algorithm which decides whether a given prime q is linearly independent.

24.1 The algorithm

For a positive integer n we define

$$\text{next\_li\_prime}(n)$$

to be the smallest prime q of the form $q = kn + 1$ with $k \geq 1$, i.e. the smallest prime q such that $q \equiv 1 \pmod{n}$. In pseudocode:

```
def next_li_prime(n):
    k = 0
    while True:
        k += 1
        q = k*n + 1
        if is_prime(q):
            return q
```

(Dirichlet's theorem on primes in arithmetic progressions, or Murthy's repunit argument, guarantees that this procedure eventually terminates for every prime input n.)

Given a prime q, we then define

$$\text{is\_li\_prime}(q)$$

by looping over the prime divisors p of $q-1$ and checking whether q is the smallest prime $\equiv 1 \pmod{p}$:

```
def is_li_prime(q):
    for p in prime_divisors(q-1):
        if q == next_li_prime(p):
            return True
    return False
```

The claim is that, apart from the trivial small cases, this procedure returns `True` if and only if q is linearly independent in the sense of the valuation vectors introduced earlier.

24.2 A predecessor characterization of LI primes

Recall that for the increasing sequence of primes

$$p_1 = 2 < p_2 = 3 < p_3 = 5 < \dots$$

we associated to each p_k its valuation vector

$$\varphi(p_k) := \big(v_{p_1}(p_k - 1), \dots, v_{p_{k-1}}(p_k - 1)\big) \in \mathbb{Z}^{k-1},$$

and we called p_k *linearly independent* (LI) if $\varphi(p_k)$ does not lie in the $\mathbb{Q}$-span of $\varphi(p_1), \dots, \varphi(p_{k-1})$.

The algorithm above is based on the following purely arithmetical characterization of LI primes.

Proposition 24.1 (Successor characterization)**.** *Let q be a prime > 2. Then the following are equivalent:*

1. *q is linearly independent (in the sense of valuation vectors).*
2. *There exists a prime divisor p of $q - 1$ such that q is the smallest prime with*

$$q \equiv 1 \pmod{p}.$$

In other words, a prime q is LI if and only if it can be written in the form

$$q = \Phi(p)$$

for some prime $p < q$, where $\Phi(p)$ denotes the smallest prime $\equiv 1 \pmod{p}$.

Proof. (2) $\Rightarrow$ (1). Suppose there is a prime $p \mid (q - 1)$ such that q is the smallest prime $1 \bmod p$. Let $q = p_k$ and note that $p < p_k = q$, so p is one of $p_1, \dots, p_{k-1}$. By definition of $\varphi(p_k)$,

$$v_p(q - 1) = v_p(p_k - 1) \geq 1,$$

so the p–coordinate of $\varphi(p_k)$ is non-zero.

On the other hand, if r is any prime with $r < q$, then by minimality of q we have $r \not\equiv 1 \pmod{p}$, so $p \nmid (r - 1)$ and hence $v_p(r - 1) = 0$. Thus for every $r < q$ the p–coordinate of $\varphi(r)$ is zero, whereas the p–coordinate of $\varphi(q)$ is non-zero.

If we assume for contradiction that $\varphi(q)$ lies in the $\mathbb{Q}$-span of $\{\varphi(r) : r < q\}$, then the p–coordinate of $\varphi(q)$ would have to be a $\mathbb{Q}$-linear combination of zeros, which is impossible. Hence $\varphi(q)$ is not in the $\mathbb{Q}$-span of the previous vectors, and q is linearly independent.

(1) $\Rightarrow$ (2). Now assume that q is linearly independent. Then, by the structural results of Section 20, there exists a (unique) *predecessor* prime $p < q$ such that

$$q = \Phi(p),$$

i.e. q is the smallest prime with $q \equiv 1 \pmod{p}$. Moreover, by construction p divides $q - 1$, so p is one of the prime divisors of $q - 1$.

This gives the desired prime divisor p of $q - 1$ with the stated minimality property. $\square$

Thus for primes $q > 2$ we have:

$$q \text{ is LI} \quad \Longleftrightarrow \quad \exists \text{ prime } p \mid (q - 1) \text{ with } q = \Phi(p).$$

24.3 Correctness of the algorithm

We now show that is_li_prime(q) implements exactly the criterion of Proposition 24.1.

Theorem 24.2. *Let $q > 2$ be a prime. Then the procedure `is_li_prime(q)` returns `True` if and only if q is a linearly independent prime.*

Proof. By definition, `is_li_prime(q)` returns `True` if and only if there exists a prime divisor p of $q-1$ such that

$$q = \text{next\_li\_prime}(p),$$

i.e. such that q is the smallest prime of the form $kp+1$ with $k \geq 1$. This is exactly the statement that q is the smallest prime $1 \bmod p$.

Thus, for $q > 2$,

$$\texttt{is\_li\_prime(q) == True} \iff \exists \text{ prime } p \mid (q-1) \text{ with } q \text{ minimal } \equiv 1 \pmod{p}.$$

By Proposition 24.1, this condition is equivalent to q being linearly independent. This proves the theorem. □

Remark 24.3. The case $q = 2$ can be treated separately, depending on the chosen convention. Since $2 - 1 = 1$ has no prime divisors, the loop in `is_li_prime(2)` does not execute and the function returns `False`. If one wishes to declare 2 (and possibly 3) to be linearly independent by convention, one may simply add a special case at the beginning of the function.

25 Additive and multiplicative structure of primes via successors

In this section we do *not* assume Hypothesis (H). The only input we need is that for every prime p there exists at least one prime q with $q \equiv 1 \pmod{p}$ (for instance by Dirichlet's theorem on primes in arithmetic progressions, or by Murthy's repunit argument). This allows us to define the successor map

$$\Phi(p) := \min\{\, q \text{ prime} : q \equiv 1 \pmod{p} \,\}$$

for every prime p.

Throughout, let $p_1 = 2 < p_2 = 3 < p_3 = 5 < \dots$ be the increasing sequence of primes. For a prime q we define the valuation vector

$$\varphi(q) := (v_{p_1}(q-1), v_{p_2}(q-1), v_{p_3}(q-1), \dots) \in \mathbb{Z}^{(\mathcal{P})},$$

i.e. as the infinite vector indexed by all primes p_i, with $v_{p_i}(q-1)$ in the p_i–coordinate. (Equivalently, we may think of $\varphi(q)$ as the formal sum $\sum_{p \text{ prime}} v_p(q-1)\, e_p$ in the free abelian group on the primes.)

We say that a prime q is *linearly independent* (LI) if $\varphi(q)$ does not lie in the $\mathbb{Q}$–span of the vectors $\varphi(r)$ with $r < q$, and *linearly dependent* (LD) otherwise.

25.1 Additive expansion in terms of successors

We first show that the vectors $\varphi(\Phi(p))$ form, in a precise sense, an integral basis of the valuation data of all primes.

Theorem 25.1 (Additive expansion via successors)**.** *For every prime q there exist unique integers a_p (only finitely many of them nonzero) such that*

$$\varphi(q) \;=\; \sum_{p \text{ prime}} a_p \, \varphi(\Phi(p)).$$

Proof. We use two structural facts about LI primes, both proved purely from the definition of φ and the successor map Φ:

(i) The vectors $\varphi(r)$ with r LI form a $\mathbb{Z}$–basis of the lattice generated by all $\varphi(q)$. In particular, for each prime q there exist unique integers c_r (only finitely many nonzero) such that

$$\varphi(q) = \sum_{r \text{ LI}} c_r \, \varphi(r).$$

(ii) A prime r is LI if and only if there exists a prime $p \mid (r-1)$ such that r is the smallest prime q with $q \equiv 1 \pmod{p}$. In other words, every LI prime r has a unique *predecessor* p with

$$r = \Phi(p),$$

and conversely every prime of the form $r = \Phi(p)$ is LI.

Fact (i) is just the usual statement that when we process the primes in increasing order, each LI prime contributes one new $\mathbb{Q}$–independent vector, and the set of all such vectors forms an integral basis of the lattice generated by the $\varphi(q)$. Fact (ii) is the predecessor characterization of LI primes, proved by looking at the p–adic coordinate in the valuation vectors.

Using (i), we can write, for any fixed prime q,

$$\varphi(q) \;=\; \sum_{r \text{ LI}} c_r \, \varphi(r), \quad c_r \in \mathbb{Z}, \tag{$*$}$$

with only finitely many nonzero c_r.

By (ii), for each LI prime r there exists a unique prime $p_r \mid (r-1)$ such that $r = \Phi(p_r)$. Substituting $r = \Phi(p_r)$ into $(*)$ gives

$$\varphi(q) \;=\; \sum_{r \text{ LI}} c_r \, \varphi(\Phi(p_r)).$$

Now we simply reindex the sum by p instead of r: for each prime p set

$$a_p := \begin{cases} c_r, & \text{if there is an LI prime } r \text{ with } r = \Phi(p), \\ 0, & \text{otherwise.} \end{cases}$$

Because each LI prime r has exactly one predecessor p_r, every LI prime contributes to exactly one a_p, and we obtain

$$\varphi(q) \;=\; \sum_{p} a_p \, \varphi(\Phi(p)),$$

with only finitely many nonzero a_p. This proves existence.

For uniqueness, suppose we have two expansions

$$\varphi(q) \;=\; \sum_p a_p\,\varphi(\Phi(p)) \;=\; \sum_p b_p\,\varphi(\Phi(p)).$$

Subtracting, we get

$$0 \;=\; \sum_p (a_p - b_p)\,\varphi(\Phi(p)).$$

The primes of the form $\Phi(p)$ are exactly the LI primes r, and by (i) the vectors $\varphi(r)$ with r LI form a $\mathbb{Z}$–basis, hence are linearly independent over $\mathbb{Q}$. Thus all coefficients $a_p - b_p$ must vanish, i.e. $a_p = b_p$ for all primes p. This proves uniqueness. □

25.2 A multiplicative structural theorem for primes

The additive expansion of Theorem 25.1 can be turned into a multiplicative factorization of the integer $q-1$ in terms of the numbers $\Phi(p)-1$.

Corollary 25.2 (Multiplicative structure of primes)**.** *For every prime q there exist unique integers a_p (only finitely many nonzero) such that*

$$q - 1 \;=\; \prod_p (\Phi(p)-1)^{a_p},$$

and hence

$$q \;=\; 1 \;+\; \prod_p (\Phi(p)-1)^{a_p}.$$

Proof. Let q be a prime. By Theorem 25.1 there are unique integers a_p with

$$\varphi(q) \;=\; \sum_p a_p\,\varphi(\Phi(p)).$$

Define

$$F \;:=\; \prod_p (\Phi(p)-1)^{a_p} \;\in \mathbb{Q}^\times.$$

We compare the r–adic valuations of F and $q-1$ for all primes r.

By definition of φ, the r–coordinate of $\varphi(q)$ is $v_r(q-1)$, and the r–coordinate of $\varphi(\Phi(p))$ is $v_r(\Phi(p)-1)$. Thus, from the equality of vectors

$$\varphi(q) = \sum_p a_p\,\varphi(\Phi(p)),$$

we read off, coordinatewise, that for every prime r,

$$v_r(q-1) \;=\; \sum_p a_p\,v_r(\Phi(p)-1).$$

On the other hand,

$$v_r(F) \;=\; v_r\!\left(\prod_p (\Phi(p)-1)^{a_p}\right) \;=\; \sum_p a_p\,v_r(\Phi(p)-1).$$

Hence

$$v_r(F) = v_r(q-1) \quad \text{for every prime } r.$$

It follows that the quotient $F/(q-1)$ has r–adic valuation 0 for every prime r, i.e. $F/(q-1)$ is a unit in $\mathbb{Z}$. Therefore $F/(q-1) = \pm 1$, and since both F and $q-1$ are positive, we must have $F = q-1$.

This is exactly the claimed factorization

$$q - 1 \;=\; \prod_{p} (\Phi(p) - 1)^{a_p},$$

and adding 1 to both sides gives the equivalent form for q.

Uniqueness of the exponents a_p follows in exactly the same way as in Theorem 25.1: if two families (a_p) and (b_p) give the same product, then comparing valuations at all primes forces $a_p = b_p$ for every p. □

26 Decomposition via chains from linearly dependent primes

In this section we do *not* assume Hypothesis (H). We only use that for each prime p there exists at least one prime $q \equiv 1 \pmod{p}$, so that the successor map

$$\Phi(p) := \min\{\, q \text{ prime} : q \equiv 1 \pmod{p} \,\}$$

is well-defined.

Recall:

- For each prime q we define the valuation vector
$$\varphi(q) := \big(v_r(q-1)\big)_{r \text{ prime}} \in \mathbb{Z}^{(\mathcal{P})}.$$
- A prime q is called *linearly independent* (LI) if $\varphi(q)$ is not in the $\mathbb{Q}$-span of $\{\varphi(r) : r < q\}$, and *linearly dependent* (LD) otherwise.
- Every LI prime r has a unique *predecessor* prime $p \mid (r-1)$ with $r = \Phi(p)$, and conversely every $r = \Phi(p)$ is LI.
- Each prime lies in a unique successor chain
$$C(d) := \{d,\ \Phi(d),\ \Phi^2(d), \dots\},$$
whose starting point d is linearly dependent, and all further elements $\Phi^k(d)$, $k \geq 1$, are LI.

Thus every LI prime can be written uniquely as $\Phi^k(d)$ with d LD and $k \geq 1$.

26.1 Additive and multiplicative decomposition along LD chains

We now refine the additive and multiplicative structure theorems by organising all LI primes along the chains starting at LD primes.

Theorem 26.1 (Chain-based additive expansion)**.** *For every prime q there exist unique integers $a_{d,k}$ (only finitely many nonzero), indexed by linearly dependent primes d and integers $k \geq 1$, such that*

$$\varphi(q) \;=\; \sum_{\substack{d \text{ LD prime} \\ k \geq 1}} a_{d,k}\, \varphi\big(\Phi^k(d)\big).$$

Proof. We first use the standard LI/LD structure:

(i) The vectors $\varphi(r)$ with r LI form a $\mathbb{Z}$-basis of the lattice generated by all $\varphi(q)$. Hence for each prime q there exist unique integers c_r (finite support) such that

$$\varphi(q) = \sum_{r \text{ LI}} c_r \, \varphi(r).$$

(ii) Every LI prime r can be written uniquely as $r = \Phi^k(d)$ with d LD and $k \geq 1$ (the unique starting point and position in its successor chain).

Using (ii), rewrite the sum in (i) as

$$\varphi(q) \;=\; \sum_{r \text{ LI}} c_r \, \varphi(r) \;=\; \sum_{\substack{d \text{ LD} \\ k \geq 1}} c_{\Phi^k(d)} \, \varphi\big(\Phi^k(d)\big),$$

where we understand $c_{\Phi^k(d)} = 0$ if the LI prime $\Phi^k(d)$ does not appear in the original expansion.

Now define

$$a_{d,k} := c_{\Phi^k(d)}.$$

Then

$$\varphi(q) \;=\; \sum_{\substack{d \text{ LD} \\ k \geq 1}} a_{d,k} \, \varphi\big(\Phi^k(d)\big),$$

with only finitely many $a_{d,k}$ nonzero (since only finitely many LI primes appear in the original sum). This proves existence.

For uniqueness, suppose we had another family $(b_{d,k})$ with

$$\varphi(q) \;=\; \sum_{d,k} a_{d,k} \, \varphi(\Phi^k(d)) \;=\; \sum_{d,k} b_{d,k} \, \varphi(\Phi^k(d)).$$

Subtracting,

$$0 \;=\; \sum_{d,k} (a_{d,k} - b_{d,k}) \, \varphi(\Phi^k(d)).$$

The set $\{\Phi^k(d) : d \text{ LD}, \, k \geq 1\}$ is exactly the set of all LI primes r, and by (i) the vectors $\varphi(r)$ with r LI form a $\mathbb{Z}$-basis, hence are linearly independent over $\mathbb{Q}$. Therefore all coefficients $a_{d,k} - b_{d,k}$ must vanish, so $a_{d,k} = b_{d,k}$ for all d, k. This proves uniqueness. □

As before, we can translate this additive statement into a multiplicative factorisation of $q - 1$.

Corollary 26.2 (Chain-based multiplicative structure for primes)**.** *For every prime q there exist unique integers $a_{d,k}$ (only finitely many nonzero), with d running over linearly dependent primes and $k \geq 1$, such that*

$$q - 1 \;=\; \prod_{\substack{d \text{ LD prime} \\ k \geq 1}} \big(\Phi^k(d) - 1\big)^{a_{d,k}},$$

and hence

$$q \;=\; 1 \;+\; \prod_{\substack{d \text{ LD prime} \\ k \geq 1}} \big(\Phi^k(d) - 1\big)^{a_{d,k}}.$$

Proof. Let q be a prime and let $(a_{d,k})$ be as in Theorem 26.1. Define

$$F \;:=\; \prod_{\substack{d \text{ LD} \\ k\geq 1}} \bigl(\Phi^k(d)-1\bigr)^{a_{d,k}} \;\in \mathbb{Q}^\times.$$

By definition of φ, the r–th coordinate of $\varphi(q)$ is $v_r(q-1)$, and the r–th coordinate of $\varphi(\Phi^k(d))$ is $v_r(\Phi^k(d)-1)$. From

$$\varphi(q) = \sum_{d,k} a_{d,k}\,\varphi(\Phi^k(d))$$

we obtain, for every prime r,

$$v_r(q-1) \;=\; \sum_{d,k} a_{d,k}\, v_r(\Phi^k(d)-1).$$

On the other hand,

$$v_r(F) \;=\; v_r\left(\prod_{d,k}(\Phi^k(d)-1)^{a_{d,k}}\right) \;=\; \sum_{d,k} a_{d,k}\, v_r(\Phi^k(d)-1).$$

Hence $v_r(F) = v_r(q-1)$ for all primes r.

Therefore $F/(q-1)$ has r–adic valuation 0 for every prime r, so $F/(q-1)$ is a unit in $\mathbb{Z}$, i.e. $F/(q-1) = \pm 1$. Since $F > 0$ and $q-1 > 0$, we must have $F = q-1$, proving the desired factorisation.

Uniqueness of the exponents $a_{d,k}$ follows in the same way as in Theorem 26.1: if two families $(a_{d,k})$ and $(b_{d,k})$ give the same product, then comparing valuations at all primes forces $a_{d,k} = b_{d,k}$ for every d, k. □

Remark 26.3. If q itself is linearly independent, then in fact all $a_{d,k}$ are zero except for a single pair (d_0, k_0) with $\Phi^{k_0}(d_0) = q$, and we recover the trivial identity $q-1 = (\Phi^{k_0}(d_0)-1)^1$. If q is linearly dependent, then all nonzero $a_{d,k}$ necessarily correspond to LI primes $\Phi^k(d)$ with $\Phi^k(d) < q$.

27 Growth of successor chains

In this section we study the growth of orbits under the successor map Φ, both unconditionally (using only elementary tools such as Bertrand's postulate and the Prime Number Theorem) and under Hypothesis (H). Throughout, we write

$$\Phi(p) \;:=\; \min\{\, q \text{ prime} : q \equiv 1 \pmod{p} \,\}$$

for the successor of a prime p, and we consider the iterates

$$q_0 := d, \qquad q_{k+1} := \Phi(q_k) \quad (k \geq 0),$$

so that $q_k = \Phi^k(d)$ is the k-th element in the successor chain starting at the prime d. When d is linearly dependent (LD), all q_k for $k \geq 1$ are linearly independent (LI), but for the growth estimates below we only use that all q_k are odd primes ≥ 3.

We write $\pi(x)$ for the prime counting function, and we recall Bertrand's postulate: for every integer $n > 1$ there exists a prime r with $n < r < 2n$.

27.1 A lower exponential growth bound via Bertrand

We begin with a simple but robust unconditional growth estimate.

Lemma 27.1. *For every odd prime p we have*

$$\Phi(p) > 2p.$$

Proof. Any integer $q \equiv 1 \pmod{p}$ can be written as $q = kp + 1$ with $k \geq 1$. The only possibilities with $q \leq 2p$ are

$$p + 1 \quad \text{and} \quad 2p + 1.$$

For $p > 2$ the number $p + 1$ is even and greater than 2, hence not prime. Thus there is no prime $q \equiv 1 \pmod{p}$ with $q \leq 2p$, so the smallest such prime $\Phi(p)$ must satisfy $\Phi(p) > 2p$. □

Iterating this inequality along a successor chain yields an exponential lower bound.

Lemma 27.2. *Let $d \geq 3$ be a prime and let $q_k = \Phi^k(d)$ be its successor chain. Then for all $k \geq 1$ we have*

$$q_k > 2^k\, d.$$

Proof. By Lemma 27.1 we have $q_{i+1} = \Phi(q_i) > 2q_i$ for every $i \geq 0$. Iterating gives $q_k > 2^k q_0 = 2^k d$ for all $k \geq 1$. □

This can be inverted to give an upper bound for the level k in terms of the size of q_k.

Corollary 27.3. *Let $d \geq 3$ be a prime and $q = \Phi^k(d)$ for some $k \geq 1$. Then*

$$k \;\leq\; \log_2\!\Big(\frac{q}{d}\Big).$$

Proof. Lemma 27.2 gives $q = q_k > 2^k d$, hence $2^k < q/d$ and thus $k < \log_2(q/d)$. Since k is an integer, this yields the stated inequality. □

27.2 Counting primes along a chain

Bertrand's postulate also shows that each step of the chain forces at least one new prime in a disjoint interval.

Lemma 27.4. *Let $d \geq 3$ be a prime and $q_k = \Phi^k(d)$ its successor chain. For each $i \geq 1$ there exists a prime r_i such that*

$$q_{i-1} < r_i < 2q_{i-1} < q_i,$$

and the primes r_i are pairwise distinct.

Proof. Since $q_{i-1} > 1$, Bertrand's postulate yields a prime $r_i \in (q_{i-1}, 2q_{i-1})$. By Lemma 27.1 we have $q_i = \Phi(q_{i-1}) > 2q_{i-1}$, so indeed $q_{i-1} < r_i < 2q_{i-1} < q_i$.

The intervals $(q_{i-1}, 2q_{i-1})$ are disjoint for different i, because

$$2q_{i-1} < q_i \leq q_i < 2q_i \leq \cdots,$$

so the corresponding primes r_i are all distinct. □

Combining this with the chain primes themselves gives a crude but useful lower bound on $\pi(q_k)$.

Theorem 27.5. *Let $d \geq 3$ be a prime and $q_k = \Phi^k(d)$ its successor chain. Then*

$$\pi(q_k) \;\geq\; 2k+1.$$

In particular, the index of q_k in the increasing sequence of primes satisfies

$$\text{prime\_pi}(q_k) \;\geq\; 2k+1.$$

Proof. For each $i = 1, \ldots, k$, Lemma 27.4 provides a prime r_i with $q_{i-1} < r_i < q_i$, and these primes are all distinct. Together with the $k+1$ chain primes $q_0, \ldots, q_k$ this gives a strictly increasing sequence

$$q_0 < r_1 < q_1 < r_2 < q_2 < \cdots < r_k < q_k$$

of $2k+1$ different primes all $\leq q_k$. Hence $\pi(q_k) \geq 2k+1$. □

Remark 27.6. Using the Prime Number Theorem $\pi(x) \sim x/\log x$ together with Lemma 27.2 suggests the heuristic growth

$$\pi(q_k) \;\approx\; \frac{2^k d}{k \log 2},$$

so that k should be of order $\log_2 \pi(q_k)$ and also of order $\log_2 q_k$. The rigorous bounds from Lemmas 27.2 and 27.5 are much weaker but entirely elementary.

27.3 An upper growth bound under Hypothesis (H)

We now assume Hypothesis (H), which states that for every prime p the successor $\Phi(p)$ satisfies the upper bound

$$\Phi(p) \leq p^2 - 1 < p^2.$$

Equivalently,

$$p < \Phi(p) \leq p^2.$$

Iterating this along a successor chain gives a double-exponential *upper* bound for $q_k = \Phi^k(d)$.

Lemma 27.7 (Upper growth bound under (H))**.** *Assume Hypothesis (H). Let $d \geq 3$ be a prime and $q_k = \Phi^k(d)$. Then for all $k \geq 1$ we have*

$$q_k \;\leq\; d^{2^k}.$$

Proof. We argue by induction on k. For $k = 1$ we have

$$q_1 = \Phi(d) \leq d^2 = d^{2^1}$$

by Hypothesis (H). Suppose now $q_k \leq d^{2^k}$. Then

$$q_{k+1} = \Phi(q_k) \leq q_k^2 \leq \left(d^{2^k}\right)^2 = d^{2^{k+1}},$$

again using (H) in the first inequality. This proves the claim. □

This can be inverted to give a nontrivial *lower* bound for the level k in terms of the size of q_k.

Theorem 27.8 (Lower bound for the level under (H))**.** *Assume Hypothesis (H). Let $d \geq 3$ be a prime and $q = \Phi^k(d)$ for some $k \geq 1$. Then*

$$k \;\geq\; \log_2\Big(\frac{\log q}{\log d}\Big).$$

Equivalently,

$$\log\log q \;\geq\; k\log 2 + \log\log d.$$

Proof. From Lemma 27.7 we have $q \leq d^{2^k}$, hence

$$\log q \;\leq\; 2^k \log d.$$

Thus

$$2^k \;\geq\; \frac{\log q}{\log d},$$

and taking binary logarithms yields

$$k \;\geq\; \log_2\Big(\frac{\log q}{\log d}\Big).$$

The equivalent inequality for $\log\log q$ is obtained by applying log to both sides and rearranging. □

27.4 Combined picture

Putting the unconditional and conditional estimates together, we obtain the following qualitative picture for the position k of a prime $q_k = \Phi^k(d)$ in the successor chain of a fixed LD starting prime d:

- Unconditionally, by Lemma 27.2, $q_k > 2^k d$, so $k \leq \log_2(q_k/d)$, and by Theorem 27.5 we also have $\pi(q_k) \geq 2k + 1$.
- Under Hypothesis (H), by Lemma 27.7 and Theorem 27.8, we have $q_k \leq d^{2^k}$, hence

$$k \;\geq\; \log_2\Big(\frac{\log q_k}{\log d}\Big).$$

- Heuristically, combining these with the Prime Number Theorem $\pi(x) \sim x/\log x$ suggests that k should grow like $\log_2 q_k$ (or equivalently $\log_2 \pi(q_k)$) as $q_k \to \infty$.

Even without relying on such heuristics, the inequalities above show that along any successor chain the level k grows at most logarithmically and, under Hypothesis (H), at least doubly-logarithmically in the size of q_k.

28 Infinitude and lower density of linearly dependent primes

In this section we show that there exist infinitely many linearly dependent (LD) primes, and in fact that they occur with a positive lower order of magnitude. The argument is unconditional: it does not use Hypothesis (H), but relies on the existence of the successor map Φ (via Dirichlet or Murthy), the chain decomposition, the elementary inequality $\Phi(p) > 2p$, and the Prime Number Theorem.

28.1 Counting LI and LD primes

Recall that for each prime q we defined its valuation vector $\varphi(q)$, and called q *linearly independent* (LI) if $\varphi(q)$ is not in the $\mathbb{Q}$-span of $\{\varphi(r) : r < q\}$, otherwise *linearly dependent* (LD). We denote by $\mathcal{P}$ the set of all primes.

We introduce the following counting functions:

- $\pi(x) := \#\{q \in \mathcal{P} : q \le x\}$, the usual prime counting function.
- $\xi(x) := \#\{q \in \mathcal{P} : q \le x,\ q \text{ LI}\}$, the number of LI primes $\le x$.
- $\eta(x) := \pi(x) - \xi(x) = \#\{q \in \mathcal{P} : q \le x,\ q \text{ LD}\}$, the number of LD primes $\le x$.

Our goal is to show that $\eta(x) \to \infty$ and in fact $\eta(x) \gg x/(\log x)^2$ as $x \to \infty$.

28.2 Chain decomposition of the primes

We recall the chain decomposition (proved earlier, without using (H)):

- For each prime p we define its successor
$$\Phi(p) := \min\{\, q \text{ prime} : q \equiv 1 \pmod{p} \,\}.$$
- For any prime d, its successor chain is
$$C(d) := \{d,\ \Phi(d),\ \Phi^2(d), \dots\}.$$
- Each LD prime d is *not* in the image of Φ and serves as the unique starting point of a chain $C(d)$.
- Every LI prime q is in the image of Φ, and in fact lies in exactly one chain $C(d)$ with LD starting point d and $q = \Phi^k(d)$ for a unique $k \ge 1$.
- The chains $C(d)$, as d ranges over LD primes, are pairwise disjoint and cover all primes:
$$\mathcal{P} \;=\; \bigsqcup_{d \text{ LD}} C(d).$$

Thus linearly dependent primes index the chains, while linearly independent primes are the successors along these chains.

28.3 Growth along a chain: an exponential lower bound

We first recall the elementary inequality
$$\Phi(p) > 2p$$
for any odd prime p: among integers $\equiv 1 \pmod{p}$, the only candidates $\le 2p$ are $p+1$ and $2p+1$, and $p+1$ is not prime. Hence $\Phi(p) > 2p$.

Applying this along a successor chain gives:

Lemma 28.1. *Let $d \ge 3$ be a prime and let*
$$q_0 := d, \qquad q_{k+1} := \Phi(q_k) \quad (k \ge 0)$$
be its successor chain. Then for all $k \ge 1$ we have
$$q_k > 2^k\, d.$$

Proof. For each $i \geq 0$ we have $q_{i+1} = \Phi(q_i) > 2q_i$. Iterating,

$$q_k > 2^k q_0 = 2^k d$$

for all $k \geq 1$. □

Corollary 28.2. *Let $d \geq 3$ be a prime. Then the number of elements of the chain $C(d)$ that are $\leq x$ is bounded by*

$$\#(C(d) \cap [2, x]) \;\leq\; 1 + \Big\lfloor \log_2\Big(\frac{x}{d}\Big) \Big\rfloor \;\leq\; 1 + \log_2 x.$$

Proof. If $q_k \in C(d)$ and $q_k \leq x$, then by Lemma 28.1

$$2^k d < q_k \leq x \quad \Rightarrow \quad k \leq \log_2\Big(\frac{x}{d}\Big).$$

Thus there are at most $k + 1 \leq 1 + \lfloor \log_2(x/d) \rfloor$ indices $0 \leq k$ with $q_k \leq x$. The inequality $\log_2(x/d) \leq \log_2 x$ yields the final bound. □

In particular, each LD starting prime $d \leq x$ contributes at most $O(\log x)$ primes $\leq x$ through its chain.

28.4 Counting primes via chains

Let $\mathcal{D}(x)$ denote the set of LD primes $\leq x$. Then $\eta(x) = \#\mathcal{D}(x)$.

Every prime $q \leq x$ lies in exactly one chain $C(d)$ with LD starting point d. If $d > x$, then all elements of $C(d)$ are $\geq d > x$, so $C(d) \cap [2, x] = \emptyset$. Hence all primes $\leq x$ arise from chains $C(d)$ with $d \leq x$. Therefore

$$\pi(x) \;=\; \sum_{\substack{d \text{ LD} \\ d \leq x}} \#(C(d) \cap [2, x]).$$

By Corollary 28.2, for each LD prime $d \leq x$ we have

$$\#(C(d) \cap [2, x]) \;\leq\; 1 + \log_2 x.$$

Thus

$$\pi(x) \;\leq\; \sum_{\substack{d \text{ LD} \\ d \leq x}} (1 + \log_2 x) \;=\; \eta(x)\,(1 + \log_2 x).$$

Rearranging gives the lower bound

$$\eta(x) \;\geq\; \frac{\pi(x)}{1 + \log_2 x}. \tag{15}$$

28.5 Lower order of magnitude using the Prime Number Theorem

Finally we invoke the Prime Number Theorem, which states that

$$\pi(x) \sim \frac{x}{\log x} \quad \text{as } x \to \infty.$$

In particular, there exists x_0 and a constant $c_1 > 0$ such that

$$\pi(x) \;\geq\; c_1 \frac{x}{\log x} \quad \text{for all } x \geq x_0.$$

Combining this with (15) and observing that $1 + \log_2 x \asymp \log x$, we obtain: for all sufficiently large x,

$$\eta(x) \;\geq\; \frac{\pi(x)}{1+\log_2 x} \;\geq\; c_2\,\frac{x/\log x}{\log x} \;=\; c_2\,\frac{x}{(\log x)^2},$$

for some constant $c_2 > 0$.

We summarize this as:

Theorem 28.3 (Infinitude and lower density of LD primes)**.** *Unconditionally (without Hypothesis* (H)*) we have*

$$\eta(x) := \#\{q \leq x : q \textit{ is linearly dependent}\} \;\gg\; \frac{x}{(\log x)^2}.$$

In particular, there exist infinitely many linearly dependent primes.

Proof. The inequality $\eta(x) \gg x/(\log x)^2$ follows from (15) and the Prime Number Theorem as explained above. Since the right-hand side tends to infinity with x, the set of LD primes must be infinite. □

Remark 28.4. The proof uses only:

1. the existence of the successor map $\Phi(p)$ (via Dirichlet's theorem or Murthy's argument);
2. the chain decomposition of the primes into disjoint successor chains $C(d)$ indexed by LD starting primes d;
3. the elementary inequality $\Phi(p) > 2p$ for odd primes p;
4. the Prime Number Theorem.

No form of Hypothesis (H) is required. The result shows that LD primes are not a sparse curiosity: they occupy at least order $x/(\log x)^2$ among the primes up to x.

28.6 Average number of LI primes per LD chain

In this subsection we make the dependence on the counting functions explicit. Recall that

$$\pi(x) := \#\{q \leq x : q \text{ prime}\}$$

is the usual prime counting function. We denote by

$$\nu(x) := \#\{q \leq x : q \text{ linearly dependent}\}$$

the number of LD primes up to x, and by

$$\xi(x) := \#\{q \leq x : q \text{ linearly independent}\}$$

the number of LI primes up to x. Thus

$$\pi(x) = \xi(x) + \nu(x), \qquad \xi(x) = \pi(x) - \nu(x).$$

As shown in Theorem 28.3, we have the lower bound

$$\nu(x) \;\geq\; c\,\frac{x}{(\log x)^2}$$

for some absolute constant $c > 0$ and all sufficiently large x. On the other hand, the Prime Number Theorem gives

$$\pi(x) \sim \frac{x}{\log x} \qquad (x \to \infty).$$

The average number of LI primes contributed by a single LD chain up to height x can be expressed in terms of these counting functions as follows. Each LD prime $d \le x$ generates a chain

$$C(d) = \{d,\ \Phi(d),\ \Phi^2(d), \dots\},$$

in which exactly one element (namely d itself) is LD, and all other elements (if any) are LI. Thus

$$\xi(x) = \sum_{\substack{d \text{ LD} \\ d \le x}} \#\big(C(d) \cap (d, x]\big), \qquad \nu(x) = \#\{d \text{ LD} : d \le x\}.$$

We define the (global) average number of LI primes per LD chain up to x by

$$A(x) := \frac{\xi(x)}{\nu(x)} = \frac{\pi(x) - \nu(x)}{\nu(x)} = \frac{\pi(x)}{\nu(x)} - 1.$$

Using the asymptotics for $\pi(x)$ and the lower bound for $\nu(x)$, we obtain heuristically

$$\frac{\pi(x)}{\nu(x)} \ \approx\ \frac{\dfrac{x}{\log x}}{c\,\dfrac{x}{(\log x)^2}} = \frac{1}{c}\log x,$$

and therefore

$$A(x) = \frac{\xi(x)}{\nu(x)} = \frac{\pi(x)}{\nu(x)} - 1 \ \sim\ \frac{1}{c}\log x \qquad (x \to \infty).$$

In words: on average, an LD chain contributes on the order of $\log x$ LI primes up to height x. More precisely, the mean number of LI elements in the set $\{q \in C(d) : d < q \le x\}$, averaged over all LD starting points $d \le x$, grows logarithmically with x.

Remark 28.5. This is a *global* statement about the ensemble of all LD chains: it controls only the average

$$A(x) = \frac{1}{\nu(x)} \sum_{\substack{d \text{ LD} \\ d \le x}} \#\big(C(d) \cap (d, x]\big),$$

and does not preclude large fluctuations between individual chains. In particular, it remains entirely compatible with the possibility that a fixed chain such as $C(5)$ contributes only very few (or even finitely many) LI primes, provided that other chains compensate by containing correspondingly more LI primes so that the global average still grows like $\asymp \log x$.

29 Successor chain factorization of integers and complexity bounds

In this section we exploit the successor chains

$$C(d) \ =\ \{\, d,\ \Phi(d),\ \Phi^2(d), \dots\}$$

introduced in Section **??**. Recall that $\Phi(p)$ denotes the smallest prime congruent to 1 modulo p, and that by Theorem 23.1 the primes decompose disjointly into these chains, indexed by the linearly dependent (LD) primes d. We first show that this chain structure extends in a canonical way to all positive integers and then derive upper and lower bounds for a natural "successor complexity" attached to an integer.

29.1 Factorization into iterated successors

For a prime p there is a unique LD prime d and a unique integer $k \geq 0$ such that

$$p \;=\; \Phi^k(d),$$

namely d is the LD starting point of the unique chain containing p, and k is the level of p in that chain.

We now extend this to arbitrary integers.

Theorem 29.1 (Successor chain factorization of integers)**.** *Let $n \geq 2$ be an integer. Then there exist unique nonnegative integers $e_{d,k}(n) \in \mathbb{Z}_{\geq 0}$, indexed by LD primes d and integers $k \geq 0$, such that*

$$n \;=\; \prod_{d \;LD} \prod_{k \geq 0} \Phi^k(d)^{\,e_{d,k}(n)}. \tag{16}$$

All but finitely many of the exponents $e_{d,k}(n)$ are zero.

Proof. Let

$$n \;=\; \prod_{i=1}^{r} p_i^{e_i}$$

be the usual prime factorization of n, with distinct primes p_i and exponents $e_i \geq 1$. By Theorem 23.1, each prime p_i lies in a unique chain $C(d_i)$ with LD starting point d_i, and there is a unique $k_i \geq 0$ such that

$$p_i \;=\; \Phi^{k_i}(d_i).$$

Define

$$e_{d,k}(n) \;:=\; \sum_{\substack{1 \leq i \leq r \\ d_i = d,\; k_i = k}} e_i,$$

and set $e_{d,k}(n) = 0$ whenever there is no i with $(d_i, k_i) = (d, k)$. Then by construction

$$\prod_{d \;\mathrm{LD}} \prod_{k \geq 0} \Phi^k(d)^{\,e_{d,k}(n)} = \prod_{i=1}^{r} \Phi^{k_i}(d_i)^{\,e_i} \;=\; \prod_{i=1}^{r} p_i^{e_i} \;=\; n.$$

This shows the existence of a representation of the form (16).

For uniqueness, suppose that

$$n = \prod_{d,k} \Phi^k(d)^{\,e_{d,k}} = \prod_{d,k} \Phi^k(d)^{\,e'_{d,k}}$$

are two such representations with $e_{d,k}, e'_{d,k} \in \mathbb{Z}_{\geq 0}$ and only finitely many nonzero coefficients. Since the numbers $\Phi^k(d)$ run through the primes without repetition, a comparison of the usual prime factorizations of n immediately yields $e_{d,k} = e'_{d,k}$ for all pairs (d, k). This proves uniqueness. □

For $n = 1$ we adopt the convention that all $e_{d,k}(1) = 0$, so that the empty product in (16) has value 1.

29.2 A weighted successor complexity and a lower bound under (H)

The factorization (16) suggests a natural measure of the "complexity" of an integer with respect to the successor chains.

Definition 29.2 (Successor complexity)**.** For $n \geq 2$ we define the *successor complexity*

$$S(n) \;:=\; \sum_{\substack{d \text{ LD} \\ k \geq 0}} e_{d,k}(n)\, 2^k \log d,$$

where the exponents $e_{d,k}(n)$ are those from Theorem 29.1. For $n = 1$ we put $S(1) = 0$.

The weights $2^k \log d$ reflect both the size of the starting prime d and the depth k in the chain.

Under Hypothesis (H) we obtain a simple but central lower bound.

Proposition 29.3 (Lower bound for $\log n$ under (H))**.** *Assume Hypothesis (H), i.e. $\Phi(p) \leq p^2 - 1$ for all primes p. Then for every integer $n \geq 2$ we have*

$$\log n \;\leq\; S(n).$$

Proof. First observe that (H) iterates as follows: for every prime p and every $k \geq 1$,

$$\Phi^k(p) \;=\; \Phi\big(\Phi^{k-1}(p)\big) \;\leq\; \big(\Phi^{k-1}(p)\big)^2 \;\leq\; p^{2^k},$$

and hence

$$\log\big(\Phi^k(p)\big) \;\leq\; 2^k \log p.$$

Now let $n \geq 2$ and write

$$n = \prod_{d,k} \Phi^k(d)^{\,e_{d,k}(n)}$$

as in Theorem 29.1. Then

$$\begin{aligned} \log n = \log\left(\prod_{d,k} \Phi^k(d)^{\,e_{d,k}(n)}\right) \;&=\; \sum_{d,k} e_{d,k}(n)\, \log\big(\Phi^k(d)\big) \\ &\leq \sum_{d,k} e_{d,k}(n)\, 2^k \log d \;=\; S(n), \end{aligned}$$

as claimed. □

29.3 A general upper bound for the successor complexity

Remarkably, one can also obtain a coarse but completely unconditional upper bound for $S(n)$. For this we use only the elementary growth $\Phi(p) > 2p$ for odd primes p.

Lemma 29.4 (Growth along a chain)**.** *Let d be an odd prime and $k \geq 1$. Then*

$$\Phi^k(d) \;>\; 2^k d.$$

Proof. For an odd prime p, the number $p + 1$ is even and hence not prime. The smallest prime $q \equiv 1 \pmod{p}$ can therefore not be $p + 1$ and must satisfy $q \geq 2p + 1$, in particular $\Phi(p) > 2p$. Iterating this inequality gives

$$\Phi^{j+1}(d) > 2\,\Phi^j(d) \quad \text{for all } j \geq 0,$$

and hence by induction $\Phi^k(d) > 2^k d$ for all $k \geq 1$. □

For a prime factor $p = \Phi^k(d)$ in the chain of d we obtain from Lemma 29.4

$$2^k \;\le\; \frac{\Phi^k(d)}{d} \;\le\; \Phi^k(d), \qquad \log d \le \log \Phi^k(d),$$

and therefore

$$S(p) \;=\; 2^k \log d \;\le\; 2^k \log \Phi^k(d) \;\le\; \Phi^k(d)\, \log \Phi^k(d) \;=\; p \log p. \tag{17}$$

For the exceptional prime $p = 2$ we have $\Phi(2) = 3$, and (17) is easily checked directly.

Proposition 29.5 (Unconditional upper bound for $S(n)$)**.** *For every integer $n \ge 2$ we have*

$$S(n) \;\le\; n \log n.$$

Proof. Write

$$n = \prod_{i=1}^{r} p_i^{e_i}$$

for the usual prime factorization of n. Since S is additive with respect to multiplication,

$$S(n) = \sum_{i=1}^{r} e_i\, S(p_i).$$

From (17) we obtain

$$S(p_i) \;\le\; p_i \log p_i,$$

and hence

$$S(n) \;\le\; \sum_{i=1}^{r} e_i\, p_i \log p_i.$$

Set $n_i := p_i^{e_i}$, so that $n = \prod_i n_i$. Then $e_i p_i \log p_i \le n_i \log n_i$, since $n_i \log n_i = p_i^{e_i}\, e_i \log p_i \ge e_i p_i \log p_i$ for $e_i \ge 1$. Thus

$$S(n) \;\le\; \sum_{i=1}^{r} n_i \log n_i.$$

We now claim that for positive real numbers $n_i \ge 1$ with product $n = \prod_i n_i$ one always has

$$\sum_i n_i \log n_i \;\le\; n \log n.$$

For two factors $a, b \ge 1$ with fixed product $ab = K$, the sum $a \log a + b \log b$ is maximized under this constraint when $a = b = \sqrt{K}$; at this point

$$a \log a + b \log b = 2\sqrt{K} \log \sqrt{K} = \sqrt{K} \log K \le K \log K = ab \log(ab).$$

By induction on the number of factors the inequality $\sum_i n_i \log n_i \le n \log n$ follows for general n_i. We therefore obtain

$$S(n) \;\le\; \sum_i n_i \log n_i \;\le\; n \log n,$$

as claimed. □

Remark 29.6. Combining Proposition 29.3 and Proposition 29.5, we obtain under Hypothesis (H) for all $n \ge 2$ the two-sided estimate

$$\log n \;\le\; S(n) \;\le\; n \log n.$$

The size of $S(n)$ thus measures quantitatively how far the factorization of n reaches "into the depth" of the successor chains: large starting primes d, large depths k, or many factors $e_{d,k}(n)$ produce large values of $S(n)$.

30 The Multiplicative Extension of the Successor Map

We already have defined for primes $\Phi(p) = \min\{q \in \mathbb{P} : q \equiv 1 \pmod{p}\}$. We now introduce the function $\Phi(n)$ as the **completely multiplicative extension** of this map to the natural numbers.

Definition 30.1. Let $n = p_1^{e_1} \cdots p_k^{e_k}$ be the prime factorization of an integer $n \geq 1$. We define:

$$\Phi(n) := \prod_{i=1}^{k} \Phi(p_i)^{e_i}.$$

Under **Hypothesis (H)** from the manuscript (which asserts that $\Phi(p) \leq p^2 - 1$ and that Φ is injective on primes), this function exhibits several strong properties.

30.1 Properties of $\Phi(n)$ under Hypothesis (H)

1. **Injectivity:** Since the map $p \mapsto \Phi(p)$ is injective on the set of primes (as proven in the manuscript) and $\Phi(p)$ is always prime, the Fundamental Theorem of Arithmetic implies that $\Phi(n)$ is injective on $\mathbb{N}$. There are no collisions; $n \neq m \implies \Phi(n) \neq \Phi(m)$.

2. **Quadratic Upper Bound:** Hypothesis (H) implies $\Phi(p) \leq p^2$. Due to complete multiplicativity, this bound extends to composite numbers:

$$\Phi(n) = \prod \Phi(p_i)^{e_i} \leq \prod (p_i^2)^{e_i} = \left(\prod p_i^{e_i}\right)^2 = n^2.$$

   Thus, the successor of any number n is strictly bounded by its square.

3. **Exponential Lower Bound:** Unconditionally, $\Phi(p) > 2p$ for odd primes. Letting $\Omega(n)$ denote the number of prime factors of n counted with multiplicity, we have:

$$\Phi(n) > 2^{\Omega(n)} \cdot n.$$

4. **Action as a Shift Operator:** The manuscript introduces the *Successor Chain Factorization*, where every integer is uniquely represented as a product over linearly dependent (LD) starting primes d:

$$n = \prod_{d,k} \left(\Phi^k(d)\right)^{e_{d,k}}.$$

   Applying Φ to n acts as a **shift operator** on this factorization. Since $\Phi(\Phi^k(d)) = \Phi^{k+1}(d)$, the application of Φ simply increments the depth k for every factor:

$$\Phi(n) = \prod_{d,k} \left(\Phi^{k+1}(d)\right)^{e_{d,k}}.$$

31 Relationship with Successor Complexity $S(n)$

The manuscript defines the *Successor Complexity* $S(n)$ as an additive function measuring the "depth" of a number within the successor chains:

$$S(n) := \sum_{p|n} S(p) = \sum_{d,k} e_{d,k}(n) \cdot 2^k \log d.$$

There exists an exact and elegant relationship between the structural metric $S(n)$ and the dynamic map $\Phi(n)$.

Theorem 31.1 (The Eigenfunction Property)**.** *For any integer $n \geq 1$, the following identity holds:*

$$S(\Phi(n)) = 2 \cdot S(n).$$

Proof. Since S is additive and Φ is multiplicative, it suffices to show this for a single prime factor $p = \Phi^k(d)$. The complexity of p is given by $S(p) = 2^k \log d$. Applying the successor map yields $\Phi(p) = \Phi^{k+1}(d)$. The complexity of the successor is:

$$S(\Phi(p)) = 2^{k+1} \log d = 2 \cdot (2^k \log d) = 2 \cdot S(p).$$

Summing over all prime factors proves the theorem for general n. □

31.1 Interpretation: Φ as a Structural Squaring

This relationship offers a powerful theoretical justification for the definition of $S(n)$.

- **Size vs. Structure:** Under Hypothesis (H), the magnitude of the numbers behaves like a squaring operation: $\Phi(n) \approx n^2$ (implying $\log \Phi(n) \approx 2 \log n$).
- **Exact Scaling:** On the structural level measured by $S(n)$, this scaling is exact: $S(\Phi(n)) = 2S(n)$.

Therefore, $S(n)$ can be interpreted as a **"structural logarithm"** that measures the dynamic size of a number. Under iteration of Φ, while the values of the numbers depend on the irregular distribution of primes, their structural complexity S grows purely deterministically and exponentially ($S(\Phi^k(n)) = 2^k S(n)$).

32 Factorials and the density of linearly dependent primes

In this section we apply the successor chain machinery to factorials $n = m!$. Since $m!$ is the product of all integers up to m, its prime factors are precisely the primes $\leq m$. This makes $m!$ a convenient test object for translating the chain decomposition of the primes into lower bounds for the number of linearly dependent (LD) primes.

Throughout this section we write

$$\pi(x) := \#\{p \leq x : p \text{ prime}\}, \qquad \nu(x) := \#\{d \leq x : d \text{ LD prime}\}.$$

32.1 Chains and the prime factors of $m!$

By the chain decomposition theorem (Theorem 23.1), every prime belongs to a unique successor chain

$$C(d) \;=\; \{\, d,\ \Phi(d),\ \Phi^2(d), \dots \},$$

where d is an LD prime and all further elements $\Phi^k(d)$, $k \geq 1$, are linearly independent. The chains $C(d)$ are pairwise disjoint and cover all primes.

Fix $m \geq 2$. Then the prime divisors of $m!$ are exactly the primes $\leq m$, and each of these lies in a unique chain $C(d)$ with LD starting point $d \leq m$. If we let

$$L_d(m) := \#\big(C(d) \cap [2, m]\big)$$

denote the number of elements of the chain $C(d)$ which are $\leq m$, we obtain the exact identity

$$\pi(m) \;=\; \sum_{\substack{d \text{ LD} \\ d \leq m}} L_d(m). \tag{18}$$

In other words, the total number of primes $\le m$ is the sum over all LD starting points $d \le m$ of the lengths of their chains up to height m.

Our aim is to bound $L_d(m)$ from above, first unconditionally and then under Hypothesis (H), and insert these bounds into (18) to obtain lower bounds for $\nu(m)$.

32.2 Unconditional bound: logarithmic chain length

Unconditionally we have the elementary growth estimate $\Phi(p) > 2p$ for every odd prime p: the integer $p+1$ is even and hence not prime, so the smallest prime congruent to 1 modulo p is at least $2p+1$. Iterating this along a chain gives the following.

Lemma 32.1 (Unconditional growth along a chain)**.** *Let d be an odd prime and write*

$$C(d) = \{q_0, q_1, q_2, \dots\}, \qquad q_0 = d,\ q_{k+1} = \Phi(q_k).$$

Then for all $k \ge 0$ we have

$$q_k \ > \ 2^k d.$$

Proof. The inequality $\Phi(p) > 2p$ for odd primes p implies $q_{k+1} = \Phi(q_k) > 2q_k$ for all $k \ge 0$. The claim follows by induction: $q_0 = d > 2^0 d$, and if $q_k > 2^k d$ then $q_{k+1} > 2q_k > 2^{k+1} d$. □

If $q_k \le m$, then Lemma 32.1 gives $2^k d < q_k \le m$, hence

$$k \ \le \ \log_2\!\Big(\frac{m}{d}\Big) \ \le \ \log_2 m.$$

Thus any chain $C(d)$ can contribute at most

$$L_d(m) \ \le \ 1 + \Big\lfloor \log_2\!\Big(\frac{m}{d}\Big) \Big\rfloor \ \le \ 1 + \log_2 m$$

primes $\le m$. Inserting this bound into (18) yields

$$\pi(m) \ = \sum_{\substack{d \text{ LD} \\ d \le m}} L_d(m) \ \le \ \nu(m)\,(1 + \log_2 m),$$

and therefore

$$\nu(m) \ \ge \ \frac{\pi(m)}{1 + \log_2 m}. \tag{19}$$

Using the prime number theorem $\pi(m) \sim m/\log m$ we obtain in particular the unconditional asymptotic lower bound

$$\nu(m) \ \gg \ \frac{m}{(\log m)^2}.$$

Thus there exist infinitely many LD primes, and in fact they are at least as numerous as a positive constant times $m/(\log m)^2$ up to height m.

32.3 Improved bound under Hypothesis (H): doubly logarithmic chain length

Under Hypothesis (H) the chains grow much faster, and the argument above can be sharpened considerably.

Hypothesis 32.2 ((H))**.** For every prime p we have $\Phi(p) \le p^2 - 1$.

Iterating (H) along a chain shows that, starting from an LD prime d, the successors grow at most doubly exponentially in the level.

Lemma 32.3 (Upper growth under (H))**.** *Assume Hypothesis (H). Let d be a prime and define $q_0 = d$, $q_{k+1} = \Phi(q_k)$. Then for every $k \ge 1$,*

$$q_k \;\le\; d^{2^k}.$$

Proof. For $k = 1$ we have $q_1 = \Phi(d) \le d^2 - 1 < d^2$. Suppose inductively that $q_k \le d^{2^k}$. Then by (H),

$$q_{k+1} = \Phi(q_k) \le q_k^2 \le \left(d^{2^k}\right)^2 = d^{2^{k+1}},$$

which proves the claim. □

Fix again $m \ge 2$ and let q_k be the elements of $C(d)$ as in Lemma 32.3. If $q_k \le m$, then $d^{2^k} \ge q_k \ge 2$ and

$$2^k \log d \;\ge\; \log q_k \;\ge\; 0, \qquad 2^k \log d \;\le\; \log m.$$

Equivalently,

$$2^k \;\le\; \frac{\log m}{\log d}, \qquad k \;\le\; \log_2\!\Big(\frac{\log m}{\log d}\Big).$$

In particular, for every LD prime $d \le m$ we obtain the bound

$$L_d(m) \;\le\; 1 + \max\{k \ge 0 : q_k \le m\} \;\ll\; \log\log m,$$

where the implied constant is absolute (using that $\log d \ge \log 2$ for all primes d).

Inserting this into (18) gives under (H)

$$\pi(m) \;=\; \sum_{\substack{d \text{ LD} \\ d \le m}} L_d(m) \;\ll\; \nu(m)\,\log\log m,$$

and hence

$$\nu(m) \;\gg\; \frac{\pi(m)}{\log\log m}. \tag{20}$$

Using again $\pi(m) \sim m/\log m$ we conclude

$$\nu(m) \;\gg\; \frac{m}{\log m\,\log\log m}.$$

Compared to the unconditional lower bound $\nu(m) \gg m/(\log m)^2$, the additional factor of $\log\log m$ in the denominator shows that, under Hypothesis (H), LD primes are forced to be significantly more numerous: the rapid (essentially quadratic) growth of the successors compresses each chain $C(d)$ into only $O(\log\log m)$ elements below m, so that many more LD starting points are required to account for all primes up to m.

32.4 Consequences for the distribution of LD primes

The bounds (19) and (20) can be summarized as follows:

- Unconditionally, each chain $C(d)$ contributes at most $O(\log m)$ primes up to m, and one obtains
$$\nu(m) \gg \frac{m}{(\log m)^2}.$$
- Under Hypothesis (H), each chain $C(d)$ contributes at most $O(\log\log m)$ primes up to m, and one obtains the stronger lower bound
$$\nu(m) \gg \frac{m}{\log m \, \log\log m}.$$

In particular, under (H) the average number of primes contributed by a single LD chain up to height m drops from order $\log m$ to order $\log\log m$. To compensate, the number of LD starting primes must increase correspondingly. This suggests that, in the presence of (H), linearly independent primes (the successors within the chains) form a relatively thin subset of the primes, while linearly dependent primes account for the vast majority of primes up to m.

32.5 An upper bound for $\nu(n)$ under Hypothesis (H)

Recall that

$$\pi(n) = \#\{p \le n : p \text{ prime}\}, \quad \nu(n) = \#\{p \le n : p \text{ LD}\}, \quad \xi(n) = \#\{p \le n : p \text{ LI}\},$$

so that

$$\pi(n) = \nu(n) + \xi(n).$$

Thus any upper bound for $\nu(n)$ is equivalent to a lower bound for $\xi(n)$.

A trivial bound

Since LD primes are a subset of all primes, we trivially have

$$\nu(n) \le \pi(n),$$

which asymptotically yields $\nu(n) \lesssim n/\log n$. This, however, does not reflect any of the additional structure coming from the successor chains.

A refined bound under Hypothesis (H)

Under Hypothesis (H), the manuscript establishes a concrete lower bound for the number of LI primes. In particular, Corollary 21.6 shows that for every prime p we have

$$\xi(p^2 - p + 1) \ge \pi(p),$$

i.e. up to height $p^2 - p + 1$ there are at least as many LI primes as there are primes up to p.

Writing $n \approx p^2$, so that $p \approx \sqrt{n}$, this implies a lower bound for $\xi(n)$ of the shape

$$\xi(n) \ge \pi(\sqrt{n}).$$

Using the prime number theorem $\pi(x) \sim x/\log x$, we obtain

$$\xi(n) \;\gtrsim\; \frac{\sqrt{n}}{\log\sqrt{n}} \;=\; \frac{2\sqrt{n}}{\log n}.$$

Since $\nu(n) = \pi(n) - \xi(n)$, this yields the corresponding upper bound

$$\nu(n) \;\leq\; \pi(n) - \pi(\sqrt{n}),$$

and, at the level of asymptotics,

$$\nu(n) \;\lesssim\; \frac{n}{\log n} - \frac{2\sqrt{n}}{\log n}.$$

More precisely, under (H) there exists a constant $C > 0$ such that

$$\nu(n) \;\leq\; \pi(n) - C\,\frac{\sqrt{n}}{\log n} \tag{21}$$

for all sufficiently large n.

Interpretation

The inequality (21) has several structural consequences for the distribution of LD and LI primes.

- Since $\pi(\sqrt{n})$ is negligible compared to $\pi(n)$ as $n \to \infty$, the ratio $\nu(n)/\pi(n)$ tends to 1. In other words, LD primes form a set of asymptotic density 1 among all primes, while LI primes form a relatively thin subset.
- The underlying reason is the rapid growth of the successor map Φ under (H). With $\Phi(p) \leq p^2$, the elements of a chain

  $$d,\; \Phi(d),\; \Phi^2(d), \ldots$$

  leave the interval $[2, n]$ after only about $\log\log n$ steps. Thus each chain contributes very few LI primes below n.
- To account for all $\pi(n)$ primes up to n, one therefore needs many distinct chains, that is, many LD starting points d. This forces $\nu(n)$ to be very close to $\pi(n)$, with a deficit of size at least of order $\sqrt{n}/\log n$ corresponding to the LI primes.

In summary, under Hypothesis (H) one has both a strong *lower* bound for $\nu(n)$ (from the chain length estimates in Section 32) and the upper bound (21), showing that LD primes dominate the prime spectrum in a quantitative sense, while LI primes occupy a significantly sparser layer.

33 Successor Complexity and the Logarithmic Bound

In this section, we establish a fundamental connection between the growth rate of the successor map Φ and the successor complexity $S(n)$. We show that a global logarithmic lower bound for S is tightly linked to the fact that Φ does not grow super–quadratically along prime chains.

33.1 The Successor Complexity Metric

Recall the definition of the successor complexity $S(n)$ from the previous section. For a prime p, which can be written uniquely as

$$p \;=\; \Phi^k(d)$$

for some linearly dependent (LD) prime d and integer $k \geq 0$, we defined

$$S(p) := 2^k \log d. \tag{22}$$

This function is extended to all integers $n \geq 1$ by complete additivity, i.e.

$$S(nm) = S(n) + S(m) \qquad (n, m \in \mathbb{N}),$$

and thus

$$S(n) = \sum_{p^e \| n} e\, S(p),$$

where the sum runs over all primes p.

By construction, S satisfies the following eigenfunction property with respect to the successor map Φ on primes:

$$S(\Phi(p)) = 2\, S(p) \qquad \text{for all primes } p. \tag{23}$$

Together with complete additivity, (22)–(23) determines the function S uniquely.

33.2 Hypothesis (H) implies $\log n \leq S(n)$

We first show that if the successor map is bounded quadratically (Hypothesis (H)), then the structural cost of generating an integer via successor chains is always at least as large as its logarithmic size.

Theorem 33.1 (The Complexity Lower Bound)**.** *Assume Hypothesis* (H)*, i.e.*

$$\Phi(p) \leq p^2 - 1 \qquad \textit{for all primes } p.$$

Then for all integers $n \geq 1$ we have

$$\log n \;\leq\; S(n).$$

Proof. Both $\log n$ and $S(n)$ are completely additive functions of n, so it suffices to prove the inequality for primes.

Let p be an arbitrary prime. By the chain decomposition theorem, there exists a unique LD prime d and an integer $k \geq 0$ such that

$$p = \Phi^k(d).$$

We prove by induction on the chain depth k that

$$\log p \;\leq\; S(p).$$

Base case ($k = 0$): If $k = 0$, then $p = d$ is itself LD. By definition (22),

$$S(d) = 2^0 \log d = \log d.$$

Thus $\log d = S(d)$, and the desired inequality holds with equality.

Inductive step: Assume the inequality holds for a prime $q = \Phi^k(d)$, i.e.

$$\log q \ \le \ S(q).$$

Let $p = \Phi(q) = \Phi^{k+1}(d)$ be its successor in the chain.

By Hypothesis (H) we have

$$p = \Phi(q) \ \le \ q^2 - 1 \ < \ q^2.$$

Taking logarithms on both sides yields

$$\log p \ \le \ \log(q^2) \ = \ 2 \log q.$$

Using the induction hypothesis $\log q \le S(q)$, we obtain

$$\log p \ \le \ 2S(q).$$

On the other hand, by the eigenfunction property (23),

$$2S(q) = S(\Phi(q)) = S(p).$$

Substituting this into the previous inequality gives

$$\log p \ \le \ S(p).$$

By induction on k, the inequality $\log p \le S(p)$ holds for every prime p in every chain. Finally, for a general integer $n \ge 1$ with prime factorization $n = \prod p^e$ we have

$$\log n = \sum_{p^e \| n} e \, \log p \ \le \ \sum_{p^e \| n} e \, S(p) = S(n),$$

since S is completely additive. This proves the theorem. □

33.3 The converse direction: super–quadratic growth forces violations

The inequality $\log n \le S(n)$ is not merely a consequence of Hypothesis (H); it is in fact incompatible with any *persistent* super–quadratic growth of the successor map along a prime chain. We make this precise in terms of growth exponents along chains.

Let d be an LD prime and consider its chain

$$q_0 = d, \quad q_{k+1} = \Phi(q_k) \qquad (k \ge 0).$$

By definition,

$$q_k = \Phi^k(d) \qquad \text{and} \qquad S(q_k) = 2^k \log d \quad \text{for all } k \ge 0.$$

We want to understand what happens if, along this chain, the successor map grows strictly faster than quadratically in a uniform way.

Proposition 33.2 (Super–quadratic chains violate the logarithmic bound)**.** *Let d be an LD prime, and let $(q_k)_{k \ge 0}$ be its chain defined by $q_0 = d$ and $q_{k+1} = \Phi(q_k)$. Suppose there exist constants $\alpha > 2$ and $k_0 \ge 0$ such that*

$$q_{k+1} \ \ge \ q_k^{\alpha} \qquad \textit{for all } k \ge k_0. \tag{24}$$

Then

$$\lim_{k\to\infty} \frac{S(q_k)}{\log q_k} = 0.$$

In particular, for all sufficiently large k we have

$$S(q_k) \;<\; \log q_k,$$

so the global inequality $\log n \le S(n)$ fails.

Proof. By iterating the growth condition (24), we obtain for all $k \ge k_0$:

$$q_k \;\ge\; q_{k_0}^{\alpha^{\,k-k_0}}.$$

Taking logarithms,

$$\log q_k \;\ge\; \alpha^{\,k-k_0}\, \log q_{k_0}.$$

On the other hand, by definition of S along the chain we have

$$S(q_k) = 2^k \log d.$$

Therefore,

$$\frac{S(q_k)}{\log q_k} \;\le\; \frac{2^k \log d}{\alpha^{\,k-k_0}\, \log q_{k_0}} \;=\; \Big(\frac{2}{\alpha}\Big)^k \cdot \alpha^{k_0}\, \frac{\log d}{\log q_{k_0}}.$$

Since $\alpha > 2$, the base $2/\alpha$ lies strictly between 0 and 1, and hence

$$\Big(\frac{2}{\alpha}\Big)^k \longrightarrow 0 \qquad (k \to \infty).$$

The prefactor $\alpha^{k_0}\, \frac{\log d}{\log q_{k_0}}$ is a fixed constant independent of k, so we conclude

$$\lim_{k\to\infty} \frac{S(q_k)}{\log q_k} = 0.$$

In particular, there exists K such that for all $k \ge K$,

$$\frac{S(q_k)}{\log q_k} < 1,$$

i.e. $S(q_k) < \log q_k$. This contradicts the global inequality $\log n \le S(n)$, and the proposition follows. □

Corollary 33.3 (A quadratic growth barrier along chains)**.** *The global inequality*

$$\log n \;\le\; S(n) \qquad \textit{for all } n \ge 1$$

is equivalent to the following constraint on the growth of the successor map along chains:

> *For every LD prime d and its chain (q_k), and for every $\varepsilon > 0$, there exists $k_0 = k_0(d, \varepsilon)$ such that*
>
> $$q_{k+1} \;\le\; q_k^{2+\varepsilon} \qquad \textit{for all } k \ge k_0.$$

Equivalently: no prime chain can satisfy a uniform super–quadratic growth condition of the form (24) *for some $\alpha > 2$.*

Proof. The implication

$$\text{“super–quadratic growth as in (24)”} \;\Rightarrow\; \text{“}\log n \le S(n) \text{ fails”}$$

is precisely Proposition 33.2.

Conversely, assume $\log n \le S(n)$ holds for all $n \ge 1$, and fix an LD prime d with chain (q_k). If there existed some $\varepsilon > 0$ and an infinite subsequence of indices k with

$$q_{k+1} \;\ge\; q_k^{2+\varepsilon},$$

then by possibly passing to a tail of the sequence we could arrange a uniform growth condition (24) with some $\alpha > 2$, contradicting Proposition 33.2. Hence for each $\varepsilon > 0$ the inequality $q_{k+1} \le q_k^{2+\varepsilon}$ must hold for all sufficiently large k, as claimed. □

Remark 33.4. The combination of Theorem 33.1 and Corollary 33.3 shows that:

- the strong pointwise bound $\Phi(p) \le p^2 - 1$ for all primes p immediately implies the global logarithmic lower bound $\log n \le S(n)$;
- conversely, any systematic attempt of Φ to grow like $p^{2+\varepsilon}$ (or faster) along a chain inevitably forces $S(n) < \log n$ for infinitely many n, and thus contradicts the global bound.

In this sense, the inequality $\log n \le S(n)$ encodes a *quadratic growth barrier* for the successor map Φ along prime chains.

34 Conclusion

The starting point of this work is the successor map

$$\Phi(p) \;=\; \min\{\, q \text{ prime} : q \equiv 1 \pmod{p} \,\},$$

together with the decomposition of the primes into disjoint successor chains

$$C(d) \;=\; \{\, d,\; \Phi(d),\; \Phi^2(d), \dots \},$$

indexed by linearly dependent (LD) primes d. On the one hand, this yields a purely multiplicative and dynamical picture: every prime belongs to a unique chain, whose starting point is an LD prime and whose higher elements are linearly independent (LI). On the other hand, the valuation vectors

$$\varphi(q) = (v_2(q-1), v_3(q-1), v_5(q-1), \dots)$$

encode an additive, linear-algebraic structure on the set of primes, and LI primes are precisely those for which $\varphi(q)$ does not fall into the rational span of the $\varphi(p)$ for smaller primes.

At first sight this leads to a striking tension. Unconditionally, we proved that the number $\nu(x)$ of LD primes up to x satisfies

$$\nu(x) \;\gg\; \frac{x}{(\log x)^2},$$

so there are infinitely many LD primes and they already occupy a positive proportion of the primes in a crude sense. Under Hypothesis (H) (which asserts $\Phi(p) \le p^2 - 1$), the

picture becomes even more extreme: combining chain-length estimates for the factorial $m!$ with the prime number theorem, we obtain

$$\frac{m}{(\log m)^2} \ll \nu(m) \ll \frac{m}{\log m \log\log m},$$

and in particular

$$\nu(x) = \pi(x) - \xi(x) = \pi(x) - O\left(\frac{\sqrt{x}}{\log x}\right),$$

where $\xi(x)$ counts LI primes. Thus LD primes form a set of asymptotic density 1 among all primes, while LI primes occupy a much thinner layer.

At the same time, from the linear-algebraic point of view, the LI primes play the role of a basis: their valuation vectors $\varphi(q)$ span the lattice generated by all $\varphi(p)$, and every LD prime has $\varphi(q)$ in the $\mathbb{Z}$-span of the LI vectors. This gives rise to the following apparent paradox:

> A relatively small, sparse set of LI primes ("few" on the scale of $\pi(x)$) generates, in the vector-space sense, almost all other primes, while at the same time the LD primes themselves make up asymptotically almost 100% of the prime numbers.

The resolution of this paradox lies in a clear separation of two different notions of "generation" that run through the paper:

1. *Vector-space generation.* In the additive setting of valuation vectors, LI primes carry the essential "information". Each new LI prime contributes a genuinely new direction in the space of exponent patterns of $q-1$, whereas LD primes contribute only linear combinations of these directions. In this sense, the set of LI primes forms a (sparse) basis for the valuation data of all primes. There can be infinitely many LD primes, all lying in the span of a comparatively small set of LI primes, without any contradiction: this is exactly analogous to a vector space where a thin basis supports a very large (or even infinite) collection of dependent vectors.

2. *Dynamical generation via successor chains.* In the dynamical setting of the successor map Φ, the direction of generation is reversed. LD primes are precisely the *starting points* of the chains; they are never images of Φ. LI primes are the *successors* in these chains: elements of the form $\Phi^k(d)$ with $k \geq 1$ and d LD. Under Hypothesis (H), the successors grow extremely rapidly (essentially like iterated squaring), and the chains therefore leave any fixed interval $[2, x]$ after only $O(\log\log x)$ steps. Each chain is very short below x, often contributing only a handful of primes, but there are many such chains—one for each LD starting point. To populate the set of all primes up to x, the number of LD starting primes $\nu(x)$ must itself be very large, asymptotically comparable to $\pi(x)$.

In other words, LI primes are sparse but informationally rich (*basis vectors*), while LD primes are numerous but informationally redundant (*span elements*); at the same time, dynamically, LD primes act as the sources from which LI primes originate via the successor map. The apparent contradiction disappears once one separates these two roles: linear-algebraic span versus dynamical starting points.

Beyond this conceptual clarification, the paper develops several structural tools that may be of independent interest:

- A canonical factorization of every integer n into powers of iterated successors $\Phi^k(d)$ of LD primes d (Theorem 29.1), and the associated successor complexity $S(n)$, satisfying $\log n \leq S(n) \leq n \log n$ (under (H) for the lower bound, unconditionally for the upper bound).
- An additive expansion of valuation vectors $\varphi(q)$ in terms of $\varphi(\Phi(p))$, and a corresponding multiplicative factorization of $q-1$ in terms of the numbers $\Phi(p)-1$, showing that the successor map controls not only the positions of primes but also the fine structure of their predecessors.
- Density bounds for LD and LI primes, both unconditional and under (H), which together paint a two-layered picture of the prime spectrum: a thin, structurally indispensable layer of LI primes, supported by a thick "sea" of LD primes that act as chain sources.

Taken together, these results suggest a new way to organize the prime universe: not only by size or congruence class, but by the successor dynamics and the linear structure of the exponent vectors of $p-1$. This dual viewpoint—dynamical and linear-algebraic—appears to be rich enough to generate further questions, for instance about the distribution of special primes (such as Sophie Germain primes) along successor chains, or about refined analogues of classical density conjectures in the LI/LD dichotomy.

References

[1] J. M. Pollard. *Theorems on factorization and primality testing.* Proceedings of the Cambridge Philosophical Society **76** (1974), no. 3, 521–528.

[2] J. Neukirch. *Algebraic Number Theory.* Grundlehren der mathematischen Wissenschaften **322**, Springer-Verlag (1999).

[3] P. Erdős and C. Pomerance. *On the normal number of prime factors of $\phi(n)$.* Rocky Mountain Journal of Mathematics **15** (1985), no. 2, 343–352.

[4] W. Wojowu, *Linear independent prime numbers*, MathOverflow answer (2025), `https://mathoverflow.net/a/503855/165920`.

[5] O. Leka. *Ehrhart polynomials and prime numbers.* 2025. https://www.orges-leka.de/Erhart_polynomials_and_prime_numbers.pdf.

[6] G. Chenevier. *Unimodular Hunting.* Experimental Mathematics **21** (2012), no. 2, 153–163.

[7] N. J. A. Sloane (ed.). *The On-Line Encyclopedia of Integer Sequences*, A071349. `https://oeis.org/A071349`.

[8] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, AMS Colloquium Publications, vol. 53, American Mathematical Society, 2004.

[9] D. Goldston and D. R. Heath-Brown, *Comments on primes in the progression* 1 mod p, MathOverflow answer (2025), `https://mathoverflow.net/questions/503943/given-a-prime-p-and-by-dirichlet-a-prime-q-k-cdot-p1-minimal-of-this-fo`.

[10] A. Murthy, "On the divisors of Smarandache unary sequence", *Smarandache Notions Journal* **11** (2000), no. 1–2–3, 184–185.

[11] O. Leka, *Counting primes with polynomials*, preprint, September 26, 2025. Available at `https://www.orges-leka.de/counting_primes_with_polynomials.pdf`.

[12] Z. Zihan and D. Han, An improved asymptotic formula for the distribution of irreducible polynomials in arithmetic progressions over $\mathbb{F}_q$, preprint (2019), arXiv:1911.05295.