The Prime-Factorization Fractal

Orges Leka
Limburg

June 24, 2025

1. The lexicographic order on prime factorizations

Write every positive integer in its *prime-factor list* form

$$m = p_1\, p_2 \cdots p_r, \quad 1 < p_1 \leq p_2 \leq \cdots \leq p_r,$$

and similarly

$$n = q_1\, q_2 \cdots q_s, \quad 1 < q_1 \leq q_2 \leq \cdots \leq q_s.$$

Then define

$$m \trianglelefteq n \iff (p_1, \ldots, p_r) \leq_{\text{lex}} (q_1, \ldots, q_s),$$

where "$\leq_{\text{lex}}$" is the standard lexicographic order on finite tuples.

Here are as an example the first 100 natural numbers sorted by $\trianglelefteq$:

$$1, 2, 4, 8, 16, 32, 64, 96, 48, 80, 24, 72, 40, 56, 88, 12, 36, 60, 84, 20, 100, 28, 44, 52, 68,$$
$$76, 92, 6, 18, 54, 90, 30, 42, 66, 78, 10, 50, 70, 14, 98, 22, 26, 34, 38, 46, 58, 62, 74, 82, 86,$$
$$94, 3, 9, 27, 81, 45, 63, 99, 15, 75, 21, 33, 39, 51, 57, 69, 87, 93, 5, 25, 35, 55, 65, 85, 95, 7,$$
$$49, 77, 91, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97$$

2. The "gcd" matrix and its embedding

1. **Function.** For any $a, b \in \{1, 2, \ldots, N\}$, set

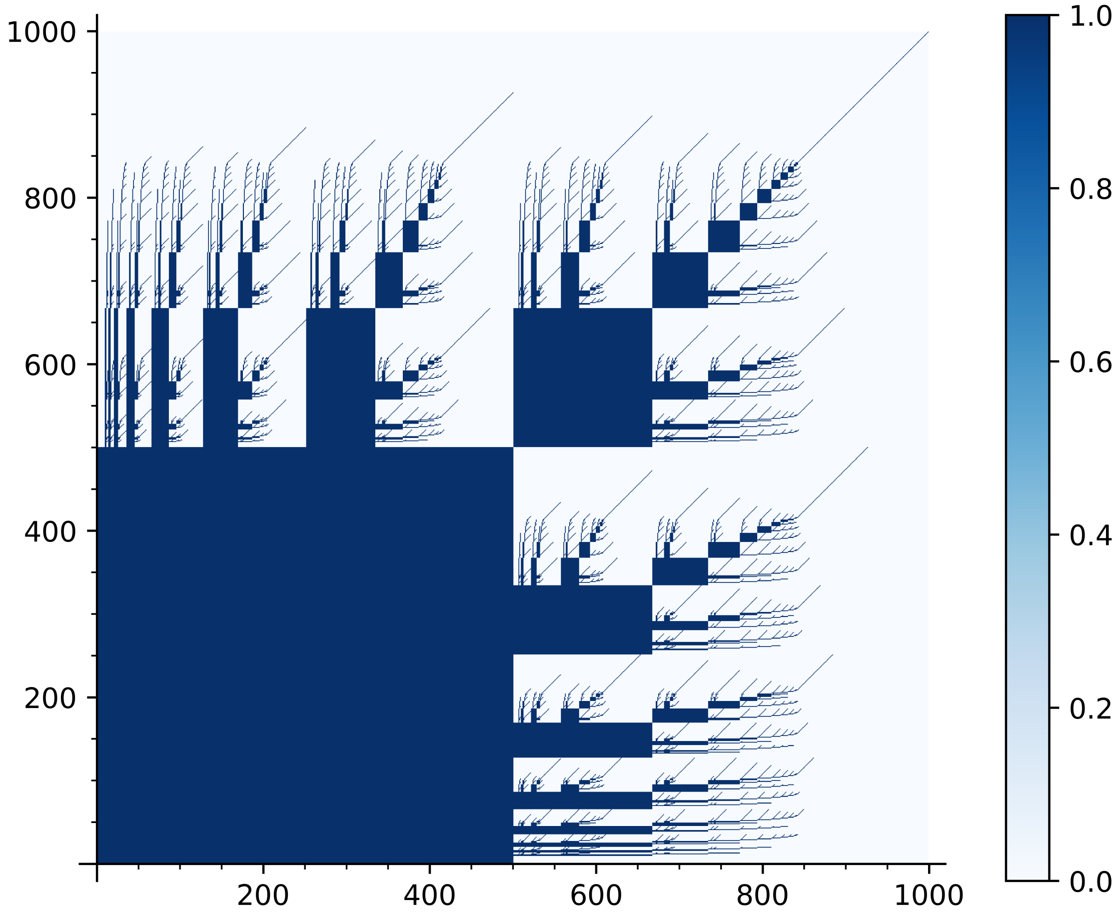
$$k(a, b) = \begin{cases} 1, & \gcd(a, b) > 1, \\ 0, & \gcd(a, b) = 1. \end{cases}$$

2. **Sorted list.** Let $\{s_1, \ldots, s_N\}$ be the numbers $1, \ldots, N$ arranged in increasing order under $\trianglelefteq$.

3. **Matrix.** Form the $N \times N$ binary matrix

$$M_N(i,j) = k(s_i, s_j).$$

4. **Fractal picture.** Plot the points $\{(i/N, j/N) \mid M_N(i,j) = 1\}$ in blue. As $N \to \infty$, this set visually converges to a self-similar fractal:



1 Transition from F_n to F_{n+1}

Informal description

We start with the unit square $[0,1]^2$ subdivided into an $n \times n$ grid of little squares

$$F_n \;=\; \bigcup_{i,j=1}^{n} S_{i,j}^{(n)},$$

where each

$$S_{i,j}^{(n)} = \left[\tfrac{i-1}{n},\, \tfrac{i}{n}\right] \times \left[\tfrac{j-1}{n},\, \tfrac{j}{n}\right].$$

These squares are arranged in *prime-factor lex order*: we list $1, \dots, n$ as

$$s_1 \preceq s_2 \preceq \cdots \preceq s_n,$$

and set

$$\operatorname{rank}(i, n) \;=\; \text{position of } i \text{ in } \{s_1, \dots, s_n\}.$$

Each square $S_{i,j}^{(n)}$ is then positioned at

$$\left(\tfrac{\operatorname{rank}(i,n)-1}{n},\ \tfrac{\operatorname{rank}(j,n)-1}{n}\right)$$

and colored by

$$\kappa(i, j) = \begin{cases} 1, & \gcd(i, j) > 1 \text{ (black)}, \\ 0, & \gcd(i, j) = 1 \text{ (white)}. \end{cases}$$

To obtain F_{n+1}:

1. *Scale down each original cell* by

   $$T_n(x, y) = \left(\tfrac{n}{n+1}x,\ \tfrac{n}{n+1}y\right),$$

   so the original cell preserves its color $\kappa(i, j)$.

2. *Insert* n new thin *vertical* strips at columns $\big(\operatorname{rank}(i, n+1) - 1\big)/(n+1)$ for $i = 1, \dots, n$, and n new thin *horizontal* strips at rows $\big(\operatorname{rank}(j, n+1) - 1\big)/(n+1)$ for $j = 1, \dots, n$. Each new strip corresponding to index i is colored black if $\gcd(i, n + 1) > 1$, white otherwise.

3. *Place* the single new cell $B_{n+1,n+1}^{(n+1)}$ at

   $$\big(\tfrac{\operatorname{rank}(n+1,n+1)-1}{n+1},\ \tfrac{\operatorname{rank}(n+1,n+1)-1}{n+1}\big),$$

   which by lex order usually lies centrally. It is always black since $\gcd(n+1, n+1) > 1$.

Exact mathematical formulation

(1) Define the ordered list $s_1 \preceq s_2 \preceq \cdots \preceq s_{n+1}$ of $\{1, \dots, n + 1\}$ under prime-factor lex order, and set

$$\operatorname{rank}(i, n + 1) = \text{index } k \text{ with } s_k = i.$$

(2) For each $1 \leq i, j \leq n + 1$ define the color-indicator

$$\kappa(i, j) = \begin{cases} 1, & \gcd(i, j) > 1, \\ 0, & \gcd(i, j) = 1. \end{cases}$$

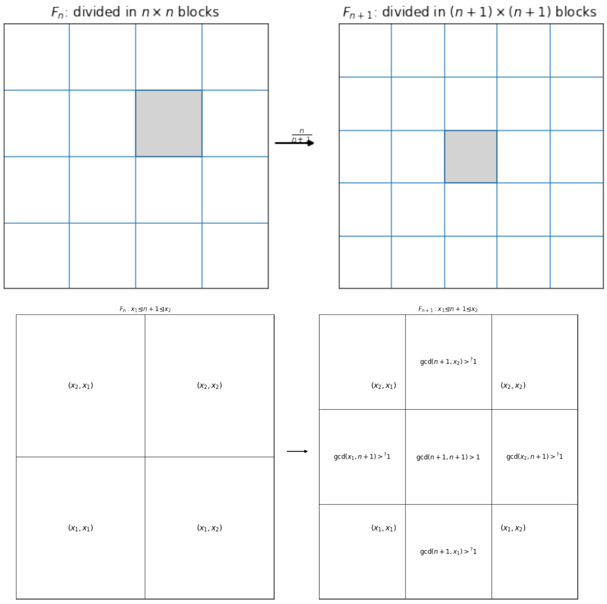


Figure 1: Left: F_n in lex-order ranking, colored by $\kappa(i,j)$. Right: F_{n+1} obtained by scaling, inserting new strips, and adding the central cell.

(3) In $[0,1]^2$ set

$$B_{i,j}^{(n+1)} = \left[\tfrac{i-1}{n+1}, \tfrac{i}{n+1}\right] \times \left[\tfrac{j-1}{n+1}, \tfrac{j}{n+1}\right], \quad i,j = 1,\dots,n+1.$$

Then

$$F_{n+1} = \bigcup_{i,j=1}^{n+1} B_{i,j}^{(n+1)},$$

where each $B_{i,j}^{(n+1)}$ is placed at

$$\left(\tfrac{\mathrm{rank}(i,n+1)-1}{n+1}, \tfrac{\mathrm{rank}(j,n+1)-1}{n+1}\right)$$

and colored black if $\kappa(i,j) = 1$, white if $\kappa(i,j) = 0$.

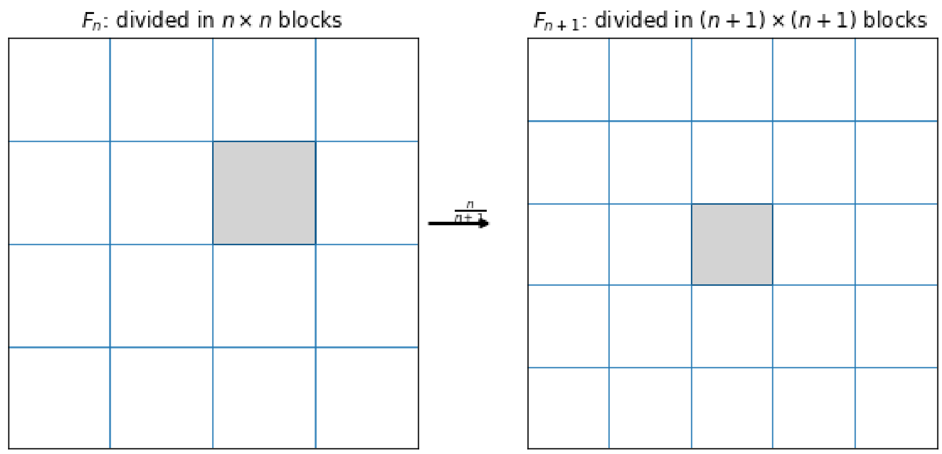
(4) Equivalently,

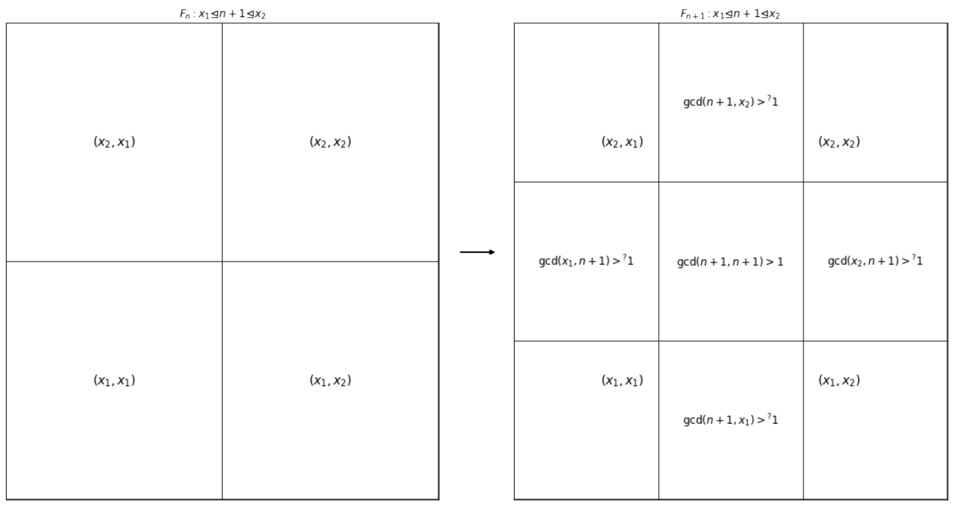
$$F_{n+1} = T_n(F_n) \cup \bigcup_{i=1}^{n} B_{i,n+1}^{(n+1)} \cup \bigcup_{j=1}^{n} B_{n+1,j}^{(n+1)} \cup B_{n+1,n+1}^{(n+1)},$$

where:

- $T_n(F_n)$ is the scaled-down copy of F_n,
- each new strip $B_{i,n+1}$ or $B_{n+1,j}$ has color $\kappa(i,n+1)$ or $\kappa(n+1,j)$,

– and $B^{(n+1)}_{n+1,n+1}$ is always black.





2 Area of the fractal

In the limit as $N \to \infty$, the proportion of black squares in F_N equals the probability that two integers $1 \leq m, n \leq N$ are *not* coprime, i.e. $\gcd(m, n) > 1$. To see this, set

$$S(N) = \sum_{\substack{1 \leq m,n \leq N \\ \gcd(m,n)=1}} 1.$$

By Möbius inversion,

$$S(N) = \sum_{m,n \leq N} \sum_{d \mid \gcd(m,n)} \mu(d) = \sum_{d \leq N} \mu(d) \left\lfloor \frac{N}{d} \right\rfloor^2.$$

Dividing by N^2 and letting $N \to \infty$ gives

$$\lim_{N\to\infty} \frac{S(N)}{N^2} = \sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} = \frac{1}{\zeta(2)} = \frac{6}{\pi^2},$$

in agreement with the classical result that two random integers are coprime with probability $6/\pi^2$[2]. Hence the complementary proportion

$$1 - \frac{6}{\pi^2}$$

is the area of the black (non-coprime) fractal set in the unit square.

3 Asymptotic Density of Numbers by Minimal Prime Divisor

Let $a_q(n) = a_{n,q}$ denote the number of integers $1 \le k \le n$ whose minimal prime divisor is q. We claim that

$$\lim_{n\to\infty} \frac{a_p(n)}{n} = \frac{1}{p} \prod_{\substack{q<p \\ q \text{ prime}}} \left(1 - \frac{1}{q}\right). \tag{1}$$

Proof via the Chinese Remainder Theorem. An integer k has minimal prime divisor p precisely when

- $k \equiv 0 \pmod p$, and
- $k \not\equiv 0 \pmod q$ for every prime $q < p$.

There are

$$p \prod_{q<p} q$$

total residue classes modulo $p \prod_{q<p} q$, of which exactly

$$1 \times \prod_{q<p} (q-1) = \prod_{q<p} (q-1)$$

are *admissible* (zero mod p, nonzero mod each $q < p$). We obtain (1) by dividing the number of admissible classes by the total number of classes:

$$\frac{\prod_{q<p}(q-1)}{p \prod_{q<p} q} = \frac{1}{p} \prod_{q<p} \frac{(q-1)}{q} = \frac{1}{p} \prod_{q<p} (1 - \frac{1}{q})$$

□

Moreover, since every integer in $\{1, 2, \ldots, n\}$ has a unique minimal prime divisor (or else is 1), we have

$$1 \;=\; \lim_{n\to\infty} \Big(\tfrac{1}{n} + \sum_{p \le n} \frac{a_p(n)}{n} \Big),$$

and hence in the limit

$$1 \;=\; \sum_{p \text{ prime}} \frac{1}{p} \prod_{\substack{q<p \\ q \text{ prime}}} \Big(1 - \tfrac{1}{q}\Big).$$

Writing the primes in ascending order $p_1 = 2, p_2 = 3, \ldots$, this becomes

$$1 \;=\; \sum_{n=1}^{\infty} \frac{1}{p_n} \prod_{q<p_n} \Big(1 - \tfrac{1}{q}\Big). \tag{2}$$

Equivalently, one may "unfold" this into the nested form

$$1 = \frac{1}{2} \;+\; \Big(1 - \tfrac{1}{2}\Big)\Big(\frac{1}{3} \;+\; \Big(1 - \tfrac{1}{3}\Big)\Big(\frac{1}{5} \;+\; \Big(1 - \tfrac{1}{5}\Big)(\cdots)\Big)\Big),$$

and so on through all primes $2, 3, 5, 7, \ldots$.

4 Prime-Based Series Representation of Real Numbers in $[0, 1]$

We shall show that for every real number $x \in [0, 1]$ there exists a (finite or infinite) strictly increasing sequence of primes

$$q_1 < q_2 < q_3 < \cdots$$

such that

$$x \;=\; \sum_{n=1}^{\infty} \frac{1}{q_n} \prod_{m=1}^{n-1} \Big(1 - \frac{1}{q_m}\Big).$$

The proof proceeds by exhibiting two algorithms: one that computes the partial sums given a prime sequence, and one greedy procedure that, given x, constructs the required prime sequence.

1. From a Prime Sequence to the Value

Define

$$x_{\text{from\_seq}}(q_1, q_2, \ldots, q_N) \;=\; \sum_{n=1}^{N} \frac{1}{q_n} \prod_{m=1}^{n-1} \Big(1 - \frac{1}{q_m}\Big).$$

In Python-style pseudocode:

```
def x_from_seq(pp):
    x = 0
    for n in range(len(pp)):
        pn = pp[n]
        pr = 1
        for q in pp[:n]:
            pr *= (1 - 1/q)
        x += 1/pn * pr
    return x
```

First observe that if $S = \mathbb{P}$ is the full sequence of all primes, then

$$\sum_{p\in\mathbb{P}} \frac{p-1}{p} \prod_{q<p}\Big(1-\frac{1}{q}\Big) = 1,$$

since this has been shown previously. Now let $S \subset \mathbb{P}$ be any (finite or infinite) subsequence. The same telescoping argument shows

$$\sum_{p\in S} \frac{p-1}{p} \prod_{\substack{q<p\\ q\in S}}\Big(1-\frac{1}{q}\Big) \;\le\; \sum_{p\in\mathbb{P}} \frac{p-1}{p} \prod_{q<p}\Big(1-\frac{1}{q}\Big) = 1.$$

Hence the series converges absolutely and its sum $x = \sum_{p\in S} \frac{p-1}{p} \prod_{q<p, q\in S} (1-\frac{1}{q})$ satisfies $0 \le x \le 1$.

2. Greedy Construction of the Prime Sequence

Given $x \in [0,1]$, we construct a sequence (q_n) by the following greedy algorithm: at step k we have a residual $x_k = x - x_{\text{from\_seq}}(q_1, \dots, q_{k-1})$, and we choose the smallest prime p for which

$$\frac{1}{p} \prod_{m=1}^{k-1}\Big(1-\frac{1}{q_m}\Big) \;\le\; x_k.$$

In Python-style pseudocode:

```
def seq_from_x(x, eps=1e-12, verbose=False):
    pp = []
    pr = 1              # product of (1 - 1/q_i) so far
    xk = x              # current residual
    k  = 0
    while xk > eps:
        k += 1
        pk = nth_prime(k)
        term = 1/pk * pr
```

```
        if xk >= term:
            pp.append(pk)
            xk -= term
            pr *= (1 - 1/pk)
            if verbose:
                print("residual:", xk, "sequence:", pp)
        # otherwise skip pk and try next
    return pp
```

3. Proof of Correctness

1. *Termination or Infinite Continuation.* At each step the residual x_k decreases by at least $\frac{1}{q_k}\prod_{m<k}(1-\frac{1}{q_m})$, so $\{x_k\}$ is a nonincreasing nonnegative sequence. If it ever falls below the chosen tolerance ε, the algorithm may terminate with a finite prime list; otherwise it produces an infinite strictly increasing sequence of primes.

2. *Exactness of the Sum.* By construction,
$$x \;=\; \sum_{n=1}^{N} \frac{1}{q_n} \prod_{m=1}^{n-1}\Bigl(1-\frac{1}{q_m}\Bigr) \;+\; x_{N+1}.$$
In the infinite-sequence case, monotone convergence gives $\lim_{N\to\infty} x_{N+1} = 0$, whence the series sums exactly to x. In the finite-sequence case, one may check that the terminal residual $x_{N+1} < \varepsilon$ can be made arbitrarily small by choosing $\varepsilon \to 0$.

3. *Uniqueness of the Greedy Choice.* At each step k, the requirement that $\frac{1}{p}\prod_{m<k}(1-\frac{1}{q_m}) \le x_k$ determines uniquely the next prime q_k as the least prime satisfying the inequality. This enforces strict increase.

Conclusion

Thus every $x \in [0,1]$ admits the desired expansion in terms of a (possibly finite) increasing sequence of primes. The function `x_from_seq` computes the value from the sequence, while `seq_from_x` recovers the sequence from the value via a greedy algorithm, completing the proof.

5 Boolean Operations on Primes and the Unit Interval

We identify
$$\mathbf{0} \;\leftrightarrow\; \varnothing, \qquad \mathbf{1} \;\leftrightarrow\; \{\text{all primes}\},$$

and for $x, y \in [0,1]$ set

$$\begin{aligned}
\neg x \;&:=\; 1 - x,\\
x \wedge y \;&:=\; \text{x_from_seq}\big(\, S_x \cap S_y \,\big),\\
x \vee y \;&:=\; \text{x_from_seq}\big(\, S_x \cup S_y \,\big).
\end{aligned}$$

Theorem 5.1. *These operations on* $[0,1]$ *satisfy the axioms of a Boolean algebra, and in particular obey De Morgan's laws*

$$\neg\big(x \wedge y\big) = \neg x \;\vee\; \neg y, \qquad \neg\big(x \vee y\big) = \neg x \;\wedge\; \neg y.$$

Proof. Under the bijection $x \leftrightarrow S_x \subset \{\text{primes}\}$, addition of sets corresponds to union, multiplication to intersection, and set-complement to $1 - x$. Since $\wp(\{\text{primes}\})$ is a Boolean algebra under $\cup, \cap, \setminus$, the transported operations on $[0,1]$ satisfy all Boolean identities, including De Morgan's laws. $\square$

6 Binary prime-digits

Let p_n denote the n-th prime, and let $x \in [0,1]$. Recall that the "prime-digit" expansion of x is given by a (possibly infinite) sequence of primes

$$[q_1, q_2, \ldots, q_n, \ldots]$$

such that

$$x \;=\; \sum_{k=1}^{\infty} \frac{1}{q_k} \prod_{m=1}^{k-1} \Big(1 - \tfrac{1}{q_m}\Big).$$

Define, for each $n \in \mathbb{N}$, the indicator

$$\varepsilon_n(x) \;:=\; \begin{cases} 1, & \text{if } p_n \text{ appears among the } q_k,\\ 0, & \text{otherwise.} \end{cases}$$

We claim that this induces the alternative expansion

$$x \;=\; \sum_{n=1}^{\infty} \frac{\varepsilon_n(x)}{p_n} \prod_{m=1}^{n-1} \Big(1 - \tfrac{1}{p_m}\Big)^{\varepsilon_m(x)},$$

so that the sequence $\big(\varepsilon_n(x)\big)_{n\geq 1}$ is precisely the "binary prime-digit" expansion of x.

Proof. By definition, each prime p_n can appear at most once in the list (q_k). If p_n does appear, let $k(n)$ be the unique index with $q_{k(n)} = p_n$; if it does not appear, set $k(n) = \infty$. Then the original expansion of x may be regrouped by collecting the single term corresponding to each prime:

$$x = \sum_{k=1}^{\infty} \frac{1}{q_k} \prod_{m=1}^{k-1} \left(1 - \tfrac{1}{q_m}\right) = \sum_{n=1}^{\infty} \begin{cases} \dfrac{1}{p_n} \displaystyle\prod_{m=1}^{k(n)-1} \left(1 - \tfrac{1}{q_m}\right), & k(n) < \infty, \\ 0, & k(n) = \infty. \end{cases}$$

But by construction, the set $\{ q_1, \ldots, q_{k(n)-1} \}$ is exactly the set of those primes p_m for which $m < n$ and $\varepsilon_m(x) = 1$. Hence

$$\prod_{m=1}^{k(n)-1} \left(1 - \tfrac{1}{q_m}\right) = \prod_{\substack{1 \le m < n \\ \varepsilon_m(x)=1}} \left(1 - \tfrac{1}{p_m}\right) = \prod_{m=1}^{n-1} \left(1 - \tfrac{1}{p_m}\right)^{\varepsilon_m(x)}.$$

Inserting this into the regrouped sum gives exactly

$$x = \sum_{n=1}^{\infty} \frac{\varepsilon_n(x)}{p_n} \prod_{m=1}^{n-1} \left(1 - \tfrac{1}{p_m}\right)^{\varepsilon_m(x)},$$

completing the proof. □

3. The factorization tree T_n

Define the infinite rooted, ordered tree T by growing one integer at a time:

- **Root:** the node 1.
- Suppose you've built T_n containing nodes $1, 2, \ldots, n$. To insert $n+1$, factor it as $n + 1 = p_1 \cdots p_r$. Let

  $$P_1(m) = \begin{cases} 1, & m = 1, \\ \max\{q : q \mid m,\ q \text{ prime}\}, & m > 1, \end{cases}$$

  and attach $n + 1$ as a child of the unique node $m \in T_n$ satisfying $P_1(m) \le p_1$ and $mp_1 \cdots p_r \le n+1$, in the position preserving preorder.

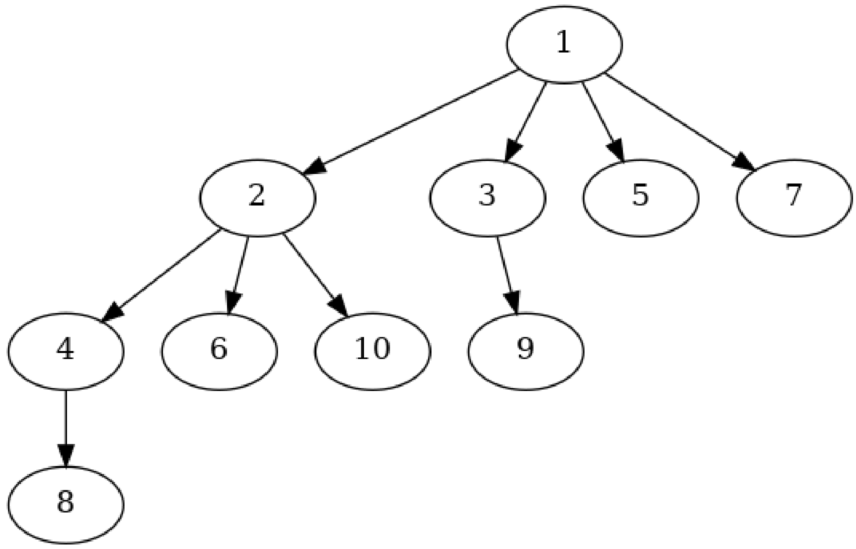
Equivalently, for each $1 \le m \le n$, define $T_{n,m}$ whose root is m, and whose children are all mp with p prime, $P_1(m) \le p$, and $mp \le n$. Then $T_n = T_{n,1}$.

When you list the vertices of T_n in *preorder*, you recover exactly the lex-order $\trianglelefteq$ on $\{1, \ldots, n\}$.

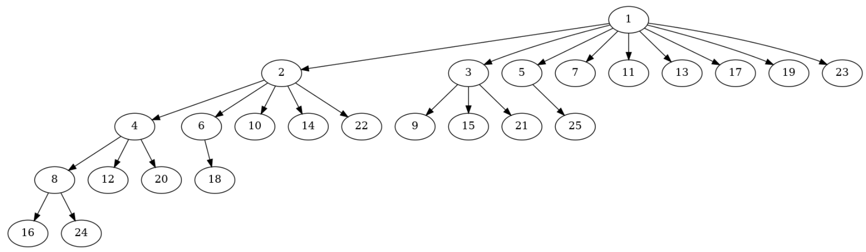
4. Example of T_n

For $n = 10, 25$ the trees look like:

$n = 10$



$n = 25$



7 Encoding and Decoding of Factorizations Using Only Primality Tests

We store the entire factorization tree T_n of the integers $\{1, 2, \ldots, n\}$ as a plane rooted tree and encode its shape in a $2n$-bit *balanced-parentheses* (BP) string.

Tree Construction

- Build T_n incrementally: start at the root 1, and for each $m = 2, \ldots, n$ attach m as a child of the unique node u whose path from 1 reproduces the prime-factor list of m.
- Each node stores its prime factor p so that the ordered list of children at u corresponds to primes $\geq P_1(u)$.

Balanced-Parentheses Encoding

1. Perform a preorder traversal of T_n.
2. On *entering* a node emit "`1`"; on *leaving* emit "`0`".

Since there are exactly n enters and n leaves, the result is a $2n$-bit string w.

Decoding Algorithm

Given $w \in \{0,1\}^{2n}$:

1. Reconstruct the tree shape from the BP-string via the standard stack-based method.
2. Precompute all primes up to n by any primality sieve.
3. Traverse the rebuilt tree in preorder, keeping track of the minimal descendant prime $P_1(u)$. For the k-th child of u, select the k-th prime $\geq P_1(u)$ as the factor.
4. The product along the path to a node yields its value m, thus recovering all factorizations.

This procedure uses only primality testing (for the sieve) plus simple BP parsing, and achieves

$$\{\text{all factorizations of } 1, \ldots, n\} \longleftrightarrow \{0,1\}^{2n}.$$

Example codings for $1 \le m \le 19$

$$
\begin{aligned}
1 &\mapsto 10,\\
2 &\mapsto 1100,\\
3 &\mapsto 110100,\\
4 &\mapsto 11100100,\\
5 &\mapsto 1110010100,\\
6 &\mapsto 111010010100,\\
7 &\mapsto 11101001010100,\\
8 &\mapsto 1111001001010100,\\
9 &\mapsto 111100100110010100,\\
10 &\mapsto 11110010100110010100,\\
11 &\mapsto 1111001010011001010100,\\
12 &\mapsto 111101001010011001010100,\\
13 &\mapsto 11110100101001100101010100,\\
14 &\mapsto 1111010010101001100101010100,\\
15 &\mapsto 111101001010100110100101010100,\\
16 &\mapsto 11111001001010100110100101010100,\\
17 &\mapsto 1111100100101010011010010101010100,\\
18 &\mapsto 111110010011001010011010010101010100,\\
19 &\mapsto 11111001001100101001101001010101010100.
\end{aligned}
$$

8 A Probabilistic Interpretation of the Prime-Divisibility Identity

Let X be a "randomly chosen" positive integer, in the sense that we consider the natural density over $\{1, 2, \ldots, N\}$ and let $N \to \infty$. For each prime p_k we ask for the event

$$E_k = \big\{\, X \text{ is divisible by } p_k \text{ but by no smaller prime } q < p_k \big\}.$$

Since every integer $n > 1$ has a unique smallest prime divisor, the events $E_1, E_2, \ldots$ form a partition of $\{2, 3, 4, \ldots\}$. (We may assign the value $k = 0$ to $n = 1$, if desired.)

For a fixed prime p_k, the probability that a random integer is divisible by p_k is

$$P\big(p_k \mid X\big) = \lim_{N\to\infty} \frac{\lfloor N/p_k \rfloor}{N} = \frac{1}{p_k}.$$

Similarly, the probability that X is *not* divisible by any smaller prime $q < p_k$ is

$$\prod_{q<p_k} \Big(1 - P(q \mid X)\Big) = \prod_{q<p_k} \Big(1 - \tfrac{1}{q}\Big).$$

By independence of divisibility by distinct primes in the natural-density sense, the probability of the event E_k is therefore

$$P(E_k) = \frac{1}{p_k} \prod_{q<p_k} \Big(1 - \tfrac{1}{q}\Big).$$

On the other hand, since the events $\{E_k : k \geq 1\}$ exhaust all integers ≥ 2 (up to the negligible singleton $\{1\}$), we have

$$\sum_{k=1}^{\infty} P(E_k) = \lim_{N\to\infty} \frac{N-1}{N} = 1.$$

Hence the identity

$$\sum_{k=1}^{\infty} \frac{1}{p_k} \prod_{q<p_k} \Big(1 - \tfrac{1}{q}\Big) \; = \; 1$$

shows at once that

$$P(X = k) \; := \; \frac{1}{p_k} \prod_{q<p_k} \Big(1 - \tfrac{1}{q}\Big)$$

defines a valid probability mass function on the indices $k = 1, 2, \ldots$.

9 Rank Invariance under Diagonal Normalization of Gram Matrices

Proposition 9.1. *Let $x_1, \ldots, x_n \neq 0$ be vectors in a Hilbert space $\mathcal{H}$, and define*

$$G = \big(\langle x_i, x_j\rangle\big)_{i,j=1}^{n}, \qquad H = \Big(\frac{\langle x_i, x_j\rangle}{\|x_i\| \, \|x_j\|}\Big)_{i,j=1}^{n}.$$

Then

$$\operatorname{rank}(G) \; = \; \operatorname{rank}(H).$$

Proof. Set

$$D = \operatorname{diag}\big(\|x_1\|^2, \ldots, \|x_n\|^2\big), \qquad D^{1/2} = \operatorname{diag}\big(\|x_1\|, \ldots, \|x_n\|\big).$$

Since each $\|x_i\| > 0$, the diagonal matrix D is invertible. A direct computation shows

$$\big(D^{-1/2}\, G\, D^{-1/2}\big)_{ij} = \frac{1}{\|x_i\|} \, \langle x_i, x_j\rangle \, \frac{1}{\|x_j\|} = H_{ij}.$$

Hence

$$H = D^{-1/2}\, G\, D^{-1/2}.$$

Multiplying a matrix by an invertible matrix on the left or right does not change its rank. Therefore

$$\operatorname{rank}(H) = \operatorname{rank}\bigl(D^{-1/2}\, G\, D^{-1/2}\bigr) = \operatorname{rank}(G),$$

as claimed. □

10 Application: The $\omega(\text{gcd})$–Kernel and Its Rank

Let $\omega(n) = \bigl|\{p \mid n : p \text{ prime}\}\bigr|$ denote the number of distinct prime divisors of n. For $2 \le a, b \le N$ define the kernel

$$k(a,b) \;=\; \omega\bigl(\gcd(a,b)\bigr).$$

Proposition 10.1. *The matrix*

$$G_N = \bigl[k(a,b)\bigr]_{a,b=2}^{N}$$

is positive semidefinite, and its rank is

$$\operatorname{rank}(G_N) \;=\; \pi(N),$$

where $\pi(N) = \#\{p \le N : p \text{ prime}\}$.

Proof. For each prime $p \le N$ define the feature function

$$\varphi_p(a) = \begin{cases} 1, & p \mid a, \\ 0, & \text{otherwise}, \end{cases} \qquad a = 2, \dots, N.$$

Then the vector

$$\Phi(a) = \bigl(\varphi_p(a)\bigr)_{p \le N} \;\in\; \{0,1\}^{\pi(N)}$$

satisfies

$$\langle \Phi(a),\, \Phi(b)\rangle = \sum_{p \le N} \varphi_p(a)\, \varphi_p(b) = \bigl|\{\, p \le N : p \mid a \text{ and } p \mid b\}\bigr| = \omega\bigl(\gcd(a,b)\bigr).$$

Thus G_N is a Gram matrix of the vectors $\Phi(a)$, hence positive semidefinite. Moreover, since each prime $p \le N$ appears as a coordinate in some $\Phi(a)$, these $\pi(N)$ coordinates are linearly independent over $\{a = 2, \dots, N\}$. Therefore

$$\operatorname{rank}(G_N) = \dim\bigl(\operatorname{span}\{\Phi(2), \dots, \Phi(N)\}\bigr) = \pi(N).$$

□

As an immediate corollary of the diagonal–normalization invariance (see Proposition 2.1), the entrywise–normalized kernel

$$h(a,b) = \frac{\omega\big(\gcd(a,b)\big)}{\sqrt{\omega(a)\,\omega(b)}},$$

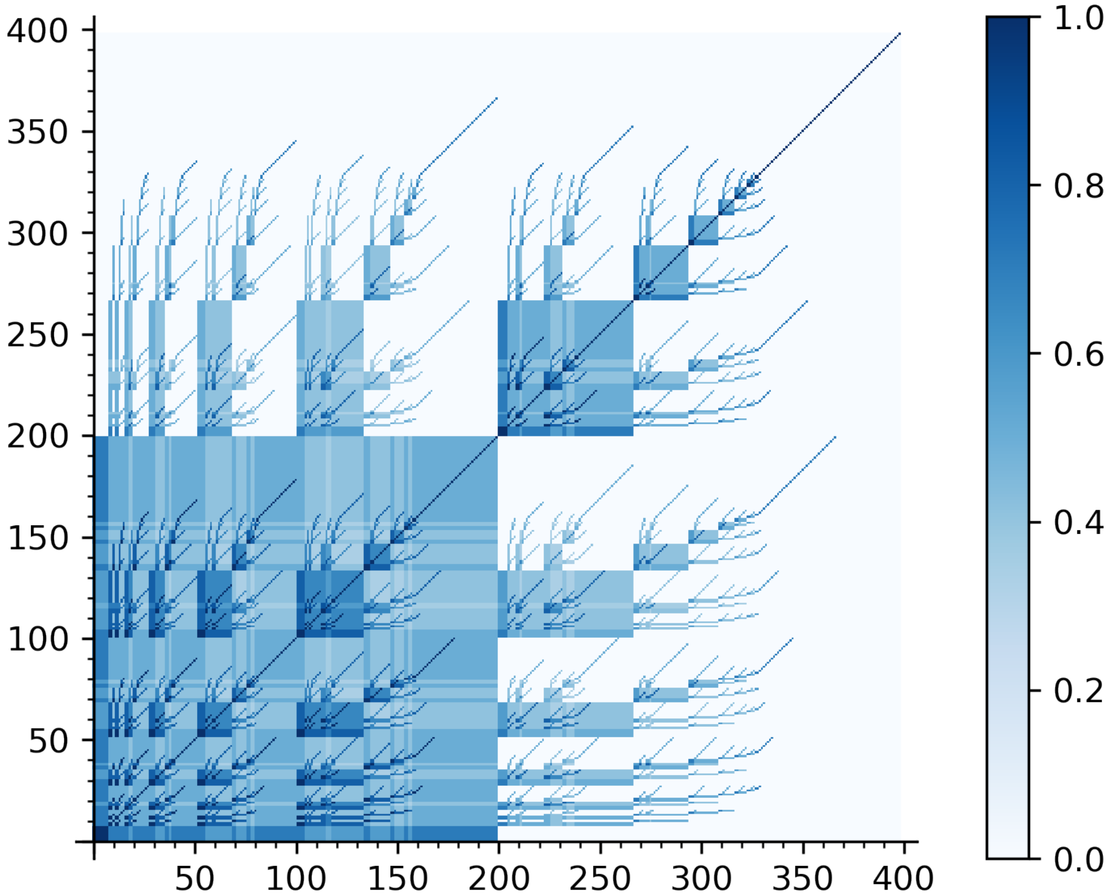
with matrix

$$H_N = \big[h(a,b)\big]_{a,b=2}^{N},$$

satisfies

$$\operatorname{rank}(H_N) = \operatorname{rank}(G_N) = \pi(N).$$

Thus both the raw and the diagonally normalized $\omega(\gcd)$–kernels yield finite adjacency (Gram) matrices of rank exactly $\pi(N)$.



Equivalence of the shifted kernel K_N and H_N

Recall that for $N \geq 2$ we defined the $(N-1) \times (N-1)$ matrix

$$H_N = \big[h(a,b)\big]_{a,b=2}^{N}, \qquad h(a,b) = \frac{\omega\big(\gcd(a,b)\big)}{\sqrt{\omega(a)\,\omega(b)}}.$$

Now introduce the shifted kernel

$$K_N = \big[k(a,b)\big]_{a,b=1}^{N-1}, \qquad k(a,b) = \frac{\omega\big(\gcd(a+1,b+1)\big)}{\sqrt{\omega(a+1)\,\omega(b+1)}}.$$

Observe that the map $a \mapsto a+1$ is a bijection from $\{1,\ldots,N-1\}$ onto $\{2,\ldots,N\}$. Hence for all $1 \le a,b \le N-1$ we have

$$k(a,b) = \frac{\omega\big(\gcd(a+1,b+1)\big)}{\sqrt{\omega(a+1)\,\omega(b+1)}} \;=\; h\big(a+1,b+1\big).$$

In other words,

$$K_N(a,b) \;=\; H_N\big(a+1,b+1\big),$$

so K_N coincides exactly with the submatrix of H_N obtained by reindexing rows and columns $2,\ldots,N$ as $1,\ldots,N-1$. Therefore the two matrices have identical entries (up to relabeling) and hence the same rank:

$$\operatorname{rank}(K_N) = \operatorname{rank}(H_N) = \pi(N).$$

11 Rank Relation between the Non-Coprimality and Coprimality Matrices

Definitions

1. For each $n \ge 1$, define the *non-coprimality matrix*

$$(M_n)_{i,j} = \begin{cases} 1, & \gcd(i,j) > 1, \\ 0, & \gcd(i,j) = 1, \end{cases} \qquad 1 \le i,j \le n.$$

2. Define the *coprimality matrix*

$$(A_n)_{i,j} = \begin{cases} 1, & \gcd(i,j) = 1, \\ 0, & \gcd(i,j) > 1, \end{cases} \qquad 1 \le i,j \le n.$$

3. Let $a(m)$ be the integer sequence **A013928** from OEIS, which satisfies

$$a(m) = \big|\{\, k \le m : k \text{ is squarefree}\}\big|.$$

Equivalently, $a(n+1)$ is the number of squarefree integers in $\{1,\ldots,n\}$.

Main Theorem

Theorem 11.1. *For every $n \ge 1$,*

$$\operatorname{rank}\big(M_n\big) \;=\; a(n+1) \;-\; 1.$$

Proof.

(1) Relation between A_n and M_n. Observe that the all-ones matrix J_n satisfies

$$A_n \;+\; M_n \;=\; J_n.$$

(2) Column spaces and ranks. Let $u = (1, 1, \ldots, 1)^T \in \mathbb{R}^n$. The first column of A_n is

$$A_n(:, 1) \;=\; u,$$

so $u \in \mathrm{Col}(A_n)$. On the other hand, the first column of M_n is

$$M_n(:, 1) = \left[\mathbf{1}_{\gcd(i,1)>1}\right]_{i=1}^{n} = \mathbf{0},$$

so $u \notin \mathrm{Col}(M_n)$.

From $J_n = A_n + M_n$ we get

$$\mathrm{Col}(J_n) \;\subseteq\; \mathrm{Col}(A_n) + \mathrm{Col}(M_n).$$

But $\mathrm{Col}(J_n) = \mathrm{span}\{u\} \subset \mathrm{Col}(A_n)$, hence $\mathrm{Col}(J_n) + \mathrm{Col}(M_n) = \mathrm{Col}(A_n)$. Conversely, $\mathrm{Col}(A_n) = \mathrm{Col}(J_n) + \mathrm{Col}(M_n) \subseteq \mathrm{Col}(A_n) + \mathrm{Col}(M_n)$, so

$$\mathrm{Col}(A_n) = \mathrm{Col}(J_n) + \mathrm{Col}(M_n) = \mathrm{Col}(A_n).$$

It follows that

$$\mathrm{Col}(A_n) = \mathrm{Col}(M_n) \oplus \mathrm{span}\{u\},$$

a direct sum because $u \notin \mathrm{Col}(M_n)$. Therefore

$$\mathrm{rank}(A_n) = \dim \mathrm{Col}(A_n) = \dim \mathrm{Col}(M_n) + 1 = \mathrm{rank}(M_n) + 1,$$

i.e.

$$\mathrm{rank}(M_n) = \mathrm{rank}(A_n) \;-\; 1.$$

(3) Rank of the coprimality matrix A_n. By standard Möbius-inversion arguments (or OEIS A013928), one shows

$$\mathrm{rank}(A_n) = \#\{\, d \le n : d \text{ is squarefree}\} = a(n+1).$$

(4) Conclusion. Combining the above,

$$\mathrm{rank}(M_n) = \mathrm{rank}(A_n) - 1 = a(n+1) - 1,$$

as claimed. $\square$

References

[1] "Probability that two random numbers are coprime is $6/\pi^2$", Math.StackExchange, Question 64498.

[2] "How to define a fractal from the lexicographic sorting on the prime factorization of natural numbers?", MathOverflow, Question 458845.

[3] L. Lovász, *Large Networks and Graph Limits*, American Mathematical Society, Providence, RI, 2012. `hombook-almost.final.pdf`.

[4] OEIS Foundation, *Number of squarefree numbers up to n*, OEIS Sequence A013928, https://oeis.org/A013928.