Introducing Cancelling Sequences with Applications to Sieve Theory

Suryansh S. Shekhawat

February 11, 2025

Abstract

This paper considers prime numbers as a sequence that can be described by arithmetic progressions missing on a few terms. This tool (cancelling sequences) can be used to generalise the sieves of Eratosthenes, Sundaram, etc, and resolve them into generating formulae with a few unknowns.

Introduction

Prime numbers, if viewed from the perspective of Sieving, are basically a set of natural numbers that have some missing terms. These terms are removed on the basis of different criteria for different sieves.

Consider the simplest sieve, the sieve of Eratosthenes [Vrd24]. What the sieve of Eratosthenes does is establish an interval from 1 to n, and starts cancelling (crossing) all multiples of 2, up to n; then 3, up to n, and so on until we get to $\sqrt{n}$. Once this happens, we can be sure that all the remaining numbers in the interval from $\sqrt{n}$ to n are prime numbers.

I have included a visual to explain this better; see Figure 1

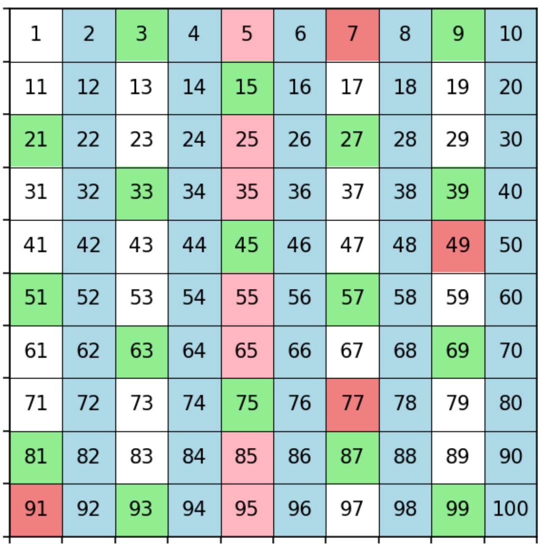

Figure 1: Sieve of Eratosthenes applied over the first 100 natural numbers

In this figure, since n = 100, we consider all the numbers (excluding 1) up to $\sqrt{100}$, that is, 10. We cross out their multiples. In this case, that would be the coral crossings of 2, green crossings of 3 pink crossings of 5 and red crossings of 7. As you can see then the numbers left in the interval [10, 100] are all primes.

Here onwards, we will build a tool (cancelling sequences) that describes the progression of these terms as a modified arithmetic progression. Once we have this, we can apply this to simplify problems or achieve exact forms for $\pi(n)$ & other formulae.

Cancelling Sequences

0.1 Proposed Definitions

Def.I : A "cancelling factor", $\mathbf{c} \in \mathbb{N}$ $(\neq 1)$, is a number whose multiples cancel the terms of an arithmetic progression.

Def.II : An arithmetic progression that excludes all terms which are multiples of $\mathbf{c} \in \mathbb{N}$ $(\neq 1)$ is called a "cancelled arithmetic sequence".

Note : Throughout this paper, the terms "sequence" and "progression" have been used interchangeably.

Def.III : A "natural arithmetic progression" is an arithmetic progression with $a_1 = 1$ and d = 1.

0.2 Formula for Cancelling Sequence and Example

If a_c denotes the general term of a natural arithmetic progression (d = 1, $a_0 = 1 \to \mathbb{N}$),
if we then reassign indices to the non-cancelled term we get,

a_c	:	1, 2, . . . , C - 1, ~~C~~, C + 1, C + 2, . . . ~~2C~~, 2C + 1, . . .
n'(old index)	:	1, 2, . . . , C - 1, C, C + 1, C + 2, . . . 2C, 2C + 1, . . .
n (new index)	:	1, 2, . . . , C - 1, C, C + 1, . . . 2C - 1, 2C, . . .

Note : The new index lags behind the old one by the number of cancellations that have occurred.

Using 6 months worth of trial, I have found the following relation to hold between n' and n. This was initially guessed and its proof was done later using hard case-wise induction, Appendix [A]

$$n' = n + \left\lceil \frac{\frac{\frac{n}{c}+n}{c} + n}{c} \ddots_{\infty} - 1 \right\rceil$$

Where $\lceil \ \rceil$ is the ceil/smallest integer function [Natar]

If we take common divisors from the top of the fraction, this can be further simplified as,

$$n' = n + \left\lceil n \left(\frac{1}{c} + \frac{1}{c^2} + \cdots + \frac{1}{c^{\infty}} \right) - 1 \right\rceil$$

Since the parentheses house a simple geometric progression with the common ratio of $\frac{1}{c}$, which is less than 1, we can find the complete sum and substitute it such as.

$$n' = n + \left\lceil \frac{n}{c-1} - 1 \right\rceil$$

To make this less abstract, let us consider an example;
Consider the arithmetic progression, with $a_0 = 1$, d = 1 (natural numbers) with a cancelling factor, c = 3.
In simpler words count from 1 upto infinity while skipping every 3rd term.
Then,

$$n' = n + \left\lceil \frac{n}{3-1} - 1 \right\rceil = n + \left\lceil \frac{n-2}{2} \right\rceil$$

To get the sequence 1, 2, 4, 5, 7, 8, 10, . . . (skipping every multiple of 3),
we can use the formula,

$$a_n = n + \left\lceil \frac{n-2}{2} \right\rceil$$

I have provided the proof for the general case via induction in the Appendix [A].

Combined Cancelling Sequences

0.3 Proposed Definitions

Def.IV : A "combined cancelling sequence", is an arithmetic progression, which has the multiples of more than one cancelling factor missing.

Note : In the context of a combined cancelling sequence, we can consider a set of these cancelling factors, appropriately named: "set of cancelling factors", $\mathbf{C} = \{ c_1, c_2, \dots \}$.

Def.V : An "Exclusive Combined Cancelling Sequence" (ECCS) is a combined cancelling sequence such that all the cancelling factors are co-prime to each other.

Note : In this paper for the sake of conciseness and consequently relevance we will only be considering Exclusive Combined Cancelling Sequences.

0.4 Example

Consider cancelling factors of $c_1 = 2$ and $c_2 = 3$ applied on a natural arithmetic progression;
For $c_1 = 2$,

AP terms :	1,	~~2~~,	3,	~~4~~,	5,	~~6~~,	7,	~~8~~,	9,	~~10~~,	11,	~~12~~,	...
n (old index) :	1,	2,	3,	4,	5,	6,	7,	8,	9,	10,	11,	12,	...
n'(new index):	1,		2,		3,		4,		5,		6,	...	

For $c_2 = 3$,

AP terms :	1,	2,	~~3~~,	4,	5,	~~6~~,	7,	8,	~~9~~,	10,	11,	~~12~~,	...
n (old index) :	1,	2,	3,	4,	5,	6,	7,	8,	9,	10,	11,	12,	...
n'(new index):	1,	2,		3,	4,		5,	6,		7,	8,	...	

Combining the two,

AP terms :	1,	~~2~~,	~~3~~,	~~4~~,	5,	~~6~~,	7,	~~8~~,	~~9~~,	~~10~~,	11,	~~12~~,	...
n (old index) :	1,	2,	3,	4,	5,	6,	7,	8,	9,	10,	11,	12,	...
n'(new index):	1,				2,		3,				4,	...	

The terms for the cancelled sequence with factor 2,

$$a_{n,c=2} = n + \left\lceil \frac{n_2}{2-1} - 1 \right\rceil$$

The terms for the cancelled sequence with factor 3,

$$a_{n,c=3} = n + \left\lceil \frac{n_3}{3-1} - 1 \right\rceil$$

Note : n_2 and n_3 here refer to the new index based on the individual cancelled arithmetic progressions of 2 and 3 respectively.

Since the sequences have now been combined, we account for cancellations from both the cancelling sequences. Consider the new index $= 2$, which gives 5, this is because there are 3 cancellations before 5,

$$2 \ \ (index) + 3 \ \ (number \ \ of \ \ cancellations) = 5 \ \ (term).$$

To describe the combined cancelling sequence we can consider a trivial extension of this, which would be the following formula,

$$a_{n,c=2,3} = n + \left\lceil \frac{n_2}{2-1} - 1 \right\rceil + \left\lceil \frac{n_3}{3-1} - 1 \right\rceil$$

Note : that the new index by n is different for both cancelling factors, i.e., $n_{2,3}$

Note : This formula works perfectly until 6, but after that it start lagging by 1, from 12 onwards by 2 and so on... It appears this is because of common multiples of the cancelling factor 2 and 3 as they have been cancelled twice at 6, 12, ... To account for this we have to remove over accounting taking place in our formula.

$$a_{n,c=2,3} = n + \left\lceil \frac{n_2}{2-1} - 1 \right\rceil + \left\lceil \frac{n_3}{3-1} - 1 \right\rceil \;- \left\lceil \frac{n_{2.3}}{2.3-1} - 1 \right\rceil$$

Note : 2.3 is presented as a subscript to n for better understanding only, it is a product and therefore equivalent to 6.

0.5 Formula for Combined Cancelling Sequence

Before we generalise this, consider a case where we had 3 cancelling factors: 2, 3 and 5. Now when we get the general term we will have cancelled 30 3 times, that is once by 2, then once by 3 and then once by 5. We are also removing all pairwise cancellations (2, 3), (2, 5) and (3, 5) and will therefore remove 3 cancellations. By doing this we have overshot our goal of maintaining at least 1 cancellations (3 - 3), and must add 1 back.

$$\begin{aligned} a_{n,c=2,3,5} = n &+ \left\lceil \frac{n_2}{2-1} - 1 \right\rceil + \left\lceil \frac{n_3}{3-1} - 1 \right\rceil + \left\lceil \frac{n_5}{5-1} - 1 \right\rceil \\ &- \left\lceil \frac{n_{2,3}}{2.3-1} - 1 \right\rceil - \left\lceil \frac{n_{3,5}}{2.5-1} - 1 \right\rceil - \left\lceil \frac{n_{2,5}}{3.5-1} - 1 \right\rceil \\ &+ \left\lceil \frac{n_{2,3,5}}{2.3.5-1} - 1 \right\rceil \end{aligned}$$

Such kind of accounting for over and under cancelling should give us the formula for a_n,

$$a_{n,c_1,c_2,\dots c_l} = n + \sum_{i=1}^{l} \lceil \frac{n_{c_i}}{c_i - 1} - 1 \rceil - \sum_{\substack{i,j=1,2 \\ 1 \le i < j \le l}}^{l} \lceil \frac{n_{c_{i,j}}}{lcm(c_i, c_j) - 1} - 1 \rceil + \sum_{\substack{i,j,k=1,2,3 \\ 1 \le i < j < k \le l}}^{l} \lceil \frac{n_{c_{i,j,k}}}{lcm(c_i, c_j, c_k) - 1} - 1 \rceil - \dots$$

This notation seems very complicated, and the sums can also be expressed using single-operation-sigmas,

$$\begin{aligned} a_{n,c_1,\dots c_l} = n &+ \sum_{k=1}^{m} \left\lceil \frac{n}{c_k - 1} - 1 \right\rceil - \sum_{k_1=1}^{m} \sum_{k_2=k_1+1}^{m} \left\lceil \frac{n}{\mathrm{lcm}(c_{k_1}, c_{k_2}) - 1} - 1 \right\rceil \\ &+ \sum_{k_1=1}^{m} \sum_{k_2=k_1+1}^{m} \sum_{k_3=k_2+1}^{m} \left\lceil \frac{n}{\mathrm{lcm}(c_{k_1}, c_{k_2}, c_{k_3}) - 1} - 1 \right\rceil - \cdots \end{aligned}$$

The proof for the Exclusive Combined Cancelling Sequence (ECCS, 0.3) is given in the Appendix [B] and is just a consequence, as you shall see, of the inclusion-exclusion principle.

0.6 Formula for the Sieve of Eratosthenes

Here,
P is the set of cancelling factors (seed primes up to $\lfloor\sqrt{n}\rfloor$ for primes until n considered),
n is the nth prime needed,
n_p is the new index based on the individual cancelled sequence of primes in n using cancelling factor p $\in$ **P**.
l is the cardinality of the set **P**, i.e., $|\mathbf{P}| = $ l,
p_i is the i^{th} prime that is a member of the ordered set **P**.

$$a_{n,\mathbf{P}} = n + \sum_{\substack{p_i=2 \\ 1 \le i < l}}^{l} \lceil \frac{n_{p_i}}{p_i - 1} - 1 \rceil - \sum_{\substack{p_i,p_j=2,3 \\ 1 \le i < j \le l}}^{l} \lceil \frac{n_{p_{i,j}}}{p_i . p_j - 1} - 1 \rceil + \sum_{\substack{p_i,p_j,p_k=2,3,5 \\ 1 \le i < j < k \le l}}^{l} \lceil \frac{n_{p_{i,j,k}}}{p_i . p_j . p_k - 1} - 1 \rceil - \dots$$

Here the set P is the set of cancelling factors, and is therefore the set of first l prime numbers. p_i becomes a cancelling factor, it is the ith prime and lastly n_{p_i} is the new index we get when the natural progression is cancelled by the cancelling factor p_i.

0.6.1 Number of unknowns and knowns

Since the n_{p_i} term is an unknown from the perspective of just n and other prime numbers until $\lfloor\sqrt{n}\rfloor$. We have $\binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n} = \sum_{k=1}^{n} \binom{n}{k} = 2^n$ - 1 (sum of binomal terms) many unknown terms. These terms can be found analytically, but as we will see we are only interested in approximations, as there are far faster ways to compute primes than this approach seems to suggest.

0.6.2 Computational Efficiency

The Big-O notation complexity [Knu98] when using the Prime Number Theorem [HW08] and some approximations turns out to be,

$$O(\frac{\sqrt{n}^{\sqrt{n}}}{\ln\sqrt{n}})$$

I have included the proof for this in the Appendix [C]

0.7 Formula for the Sieve of Sundaram

Before we dive into the application of ECCS (0.3) to the Sieve of Sundaram, let's briefly revise what it is [Kha19]:

Algorithm 1 Generate primes using the Sieve of Sundaram

Step 1: Create a list of integers from 1 to n (inclusive).
Step 2: Eliminate numbers from the list that can be expressed as i + j + 2ij where,
* i, j are integers,
* 1 ≤ i ≤ j,
* i + j + 2ij ≤ n
Step 3: The numbers left in the list are transformed to generate primes using the formula 2x + 1, where x is the number in the list.
Step 4: The resulting numbers are all primes less than 2n + 2 with the exception of 2.

0.7.1 Important facts about the Sieve of Sundaram

1. There are no terms cancelled more than once

Proof. In the Sieve of Sundaram, the terms are generated by the expression $i+j+2ij$ for pairs of integers i, j such that $1 \leq i \leq j$ and $i+j+2ij \leq n$. The term $i+j+2ij$ is strictly increasing with respect to j for a fixed i, since for any $j_1 < j_2$, we have:

$$i + j_1 + 2ij_1 < i + j_2 + 2ij_2.$$

This strict monotonicity ensures that for each i, the values of j produce distinct terms. Moreover, since $j \geq i$, no term can be counted more than once because the indices (i, j) and (j, i) represent different pairs when $i \neq j$, and the ordering $i \leq j$ prevents symmetry from leading to duplicate terms. Thus, each pair (i, j) generates a unique term, ensuring no repetitions in the list of terms.

□

2. Given an n, we know how many numbers we need to go up for i to get i + j + 2ij ≤ n.

Proof. We can comment on how high i can go, given n, so we don't waste computations at values of i and j, where we are going to overshoot over n. We know that for each i, the corresponding j must satisfy the inequality $i+j+2ij \leq n$. To find the maximum value of i, consider the case when j is at its smallest value, which is equal to i. In this case, the inequality becomes:

$$i + j + 2ij \leq n$$

Simplifying this inequality, (i = j):

$$2i^2 + 2i - n \leq 0$$

Using the quadratic formula,

$$i \leq \frac{-2 \pm \sqrt{2^2 - 4(2)(-n)}}{2(2)}$$

Since i must be non-negative,

$$i \leq \frac{-1 + \sqrt{2n+1}}{2}$$

Thus, the highest value i can take is:

$$i_{\max} = \left\lfloor \frac{-1 + \sqrt{2n+1}}{2} \right\rfloor$$

Note : ⌊⌋ is the greatest integer function/ floor function [Natar]

This is the largest value of i that needs to be considered when running the Sieve of Sundaram for a given n. For each value of i from 1 to $i_{\max}$, we compute the corresponding values of j such that $i + j + 2ij \leq n$. □

3. Given an i & n, we know how many numbers we need to go up for j to get i + j + 2ij ≤ n.

Proof. The cancellation formula is:

$$i + j + 2ij \leq n$$

Rearranging for j, we get:

$$j + 2ij \leq n - i$$

Factoring out j:

$$j(1 + 2i) \leq n - i$$

Solving for j, we get:

$$j \leq \frac{n - i}{1 + 2i}$$

Thus, for a given i, the largest integer j that satisfies the inequality is:

$$j = \left\lfloor \frac{n - i}{1 + 2i} \right\rfloor$$

the last term that needs to be considered before incrementing i. □

0.7.2 Modelling Sieve of Sundaram as a Cancelling Sequence for modelling

Since there are no conflicting cancellations in the "Sieve of Sundaram", we will need an approach where for each i + j + 2ij only 1 cancellation occurs, not all multiples until n.

This can be done in one of two ways;

(1) Use the naive approach, that is account for all cancellations by (i + j + 2ij) cancelling factors and then remove the ones after the first cancellation.

$$p = \left\lfloor \frac{n+1}{n} \right\rfloor + 2\left((n-1) + \sum_{i=1}^{\left\lfloor \frac{-1+\sqrt{2(n-1)+1}}{2} \right\rfloor} \sum_{j=i}^{\left\lfloor \frac{(n-1)-i}{1+2i} \right\rfloor} \left\lceil \frac{n_{i+j+2ij}}{(i+j+2ij)-1} - 1 \right\rceil \right)$$
$$- 2 \sum_{i_2=1}^{\left\lfloor \frac{-1+\sqrt{2(n-1)+1}}{2} \right\rfloor} \sum_{j=i_2}^{\left\lfloor \frac{(n-1)-i_2}{1+2i_2} \right\rfloor} \sum_{i_1=1}^{\left\lfloor \frac{-1+\sqrt{2(n-1)+1}}{2} \right\rfloor} \sum_{j_1=i_1}^{\left\lfloor \frac{n-1-i_1}{1+2i_1} \right\rfloor} \left\lceil \frac{n_{i_1+j_1+2i_1j_1, i_2+j+2i_2j}}{\operatorname{lcm}(i_1+j_1+2i_1j_1, i_2+j+2i_2j)-1} - 1 \right\rceil$$

(2) Use a special function which moves forward if the cancellation occurs once

We achieve this with a simple algebraic function that returns 1 if y is greater than or equal to x and 0 if x otherwise. Let's consider one such function.

$$D(x,y) = \left\lfloor \frac{y-x}{|y-x|+1} \right\rfloor + 1 = \begin{cases} 1 & \text{if } x \geq y, \\ 0 & \text{if } x \not\geq y. \end{cases}$$

$$p = \left\lfloor \frac{n+1}{n} \right\rfloor + 2\left((n-1) + \sum_{i=1}^{\left\lfloor \frac{-1+\sqrt{2(n-1)+1}}{2} \right\rfloor} \sum_{j=i}^{\left\lfloor \frac{(n-1)-i}{1+2i} \right\rfloor} D(i+j+2ij, (n_{i+j+2ij} - 1)) \right)$$

Note : It would be most useful to consider analytical or algebraic functions in D, i.e., functions with congruence moduli used or free of ceils and floors such that summands can be simplified.

This statement can be made more algebraically pleasing, if we realise that we actually always know the number of cancellations given an n. For example, if we have n, we know that i $\leq \left\lfloor \frac{-1+\sqrt{2n+1}}{2} \right\rfloor$. For each of these i's we have to take $\lfloor \frac{n-i}{1+2i} \rfloor$ many j's. Each of these has a contribution of 1.

The total number of cancellations is then,

$$\lfloor\frac{n-1}{1+2(1)}\rfloor+\lfloor\frac{n-2}{1+2(2)}\rfloor+...+\lfloor\frac{n-(\lfloor\frac{-1+\sqrt{2n+1}}{2}\rfloor)}{1+2(\lfloor\frac{-1+\sqrt{2n+1}}{2}\rfloor)}\rfloor$$

Note : The use of the floor function introduces discrete jumps, preventing algebraic simplification of the summation without explicitly evaluating it.

References

[HW08] G. H. Hardy and E. M. Wright. An Introduction to the Theory of Numbers. Oxford University Press, 2008.

[Kha19] Nurul Khairina. The comparison of methods for generating prime numbers between the sieve of eratosthenes, atkins, and sundaram. SinkrOn, 3:293, 04 2019.

[Knu98] Donald E. Knuth. The Art of Computer Programming, Volume 2: Seminumerical Algorithms. Addison-Wesley, 3rd edition, 1998.

[Natar] National Council of Educational Research and Training. NCERT Textbook for Class 11. NCERT, New Delhi, India, Edition Year.

[Sta11] R. P. Stanley. Enumerative Combinatorics. Cambridge University Press, 2011.

[Vrd24] Anton Vrdoljak. Sieve of eratosthenes. ResearchGate, 2024. Accessed: December 24, 2024.

Appendix

A Proof for Cancelling Sequence Formula

Proof. For a sequence A_c, a natural arithmetic progression with c as the cancelling factor:

a_c : 1, 2, . . . , C - 1, ~~C~~, C + 1, C + 2, . . . ~~2C~~, 2C + 1, . . .
n'(old index) : 1, 2, . . . , C - 1, C, C + 1, C + 2, . . . 2C , 2C + 1, . . .
n (new index) : 1, 2, . . . , C - 1, C, C + 1, . . . 2C - 1, 2C, . . .

Let $P(n)$ be the statement:

$$P(n) : n' = n + \lceil \frac{n}{c-1} - 1 \rceil, \quad c \in \mathbb{N}, c \geq 2.$$

A.1 Base Case: $n = 1$

For n' = 1, n = 1 (must), because C cannot cancel a number less than 2,

$$P(1) : n' = 1 + \lceil \frac{1-c+1}{c-1} \rceil = \lceil \frac{1}{c-1} \rceil = 1, \quad \text{true for all}\, c \in \mathbb{N}, c \geq 2.$$

Note : n'_k is n' when n = k

A.2 Inductive Hypothesis

Assume that $P(n = k)$ is true, i.e.,

$$n'_{n=k} = k + \lceil \frac{k-c+1}{c-1} \rceil.$$

A.3 Inductive Step

For the conditions mentioned before, We need to prove that $P(n = k + 1)$ is also true, i.e.,

$$n'_{n=k+1} = (k+1) + \lceil \frac{(k+1)-c+1}{c-1} \rceil.$$

The recurrence that we can establish amongst terms of the old index is:

$$\begin{cases} n'_{k+1} = n'_k + 1 & \text{if } (c-1) \nmid k, \\ n'_{k+1} = n'_k + 2 & \text{if } (c-1) \mid k. \end{cases}$$

What follows is the description of why this recurrence holds,

A.3.1 Why does this recurrence hold ?

The reason we split into 2 cases is that, the skipped numbers create a "gap" and how much we jump depends if (k + 1) [an index further than the one we are assuming] falls into it.

Only multiples of c are skipped; so when, moving from $n'_{n=k}$ to $n'_{n=k+1}$ we need to figure out:
1) Does (k + 1) land on a valid number ?
 If (k + 1) is not a multiple of c, then it is valid.
2) Does (k + 1) land on a skipped number ?
 If (k + 1) is a multiple of c it gets skipped ahead to next valid number.

Example of recursive jumps

a_c : 1, 2, ~~3~~, 4, 5, ~~6~~, 7, 8, ~~9~~, 10, 11, ~~12~~, . . .
n'(old index) : 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, . . .
n (new index) : 1, 2, 3, 4, 5, 6, 7, 8, . . .

Consider:
The jump from $n'_{n=k=3} \rightarrow n'_{n=k+1=3+1=4}$
In this case, $n'_{k+1} = n'_k + 1$ as no term was skipped in this form.

The jump from $n'_{n=k=4} \rightarrow n'_{n=k+1=4+1=5}$
In this case, $n'_{k+1} = n'_k + 2$ as one term was skipped in this form.

Why (c - 1)?
Whenever we remove numbers divisible by c (c, 2c, 3c, ...) the numbers left behind form a pattern.
The gap between consecutive c's (cancelled) is always (c - 1),
For example:

$$\underbrace{1, 2} \quad \underbrace{4, 5} \quad \underbrace{7, 8} \quad \underbrace{10, 11}$$

Note : The number of terms before the gaps occur are, (c - 1) i.e., 3 - 1 = 2.

Why (c - 1) | k ?
When figuring out the relationship between k to skipped number. If k is just before the number skipped by c, then k + 1 must land on a mulitple of c,

$$k \mod (c-1) = 0$$

As the sequence resets every (c - 1) terms.
Otherwise,

$$k \mod (c-1) \neq 0$$

term described by (k + 1) is safe.

A.4 Case-Wise Proof

Case I: $(c-1) \nmid k$
To prove case I, the following must be true,

$$n'_{k+1} = k + \lceil \frac{k-c+1}{c-1} \rceil + 1.$$

and we know the following to be true,

$$n'_{k+1} = (k+1) + \lceil \frac{(k+1)}{c-1} - 1 \rceil$$

Substituting these as equal and noting if the equality holds should suffice.

$$\implies k + \lceil \frac{k-c+1}{c-1} \rceil + 1 = (k+1) + \lceil \frac{(k+1)}{c-1} - 1 \rceil$$

$$\implies \lceil \frac{k-c+1}{c-1} \rceil = \lceil \frac{(k+1)}{c-1} - 1 \rceil$$

$$\implies \lceil \frac{k}{c-1} - 1 \rceil = \lceil \frac{k+1}{c-1} - 1 \rceil$$

We can take an integer out of the ceil function

$$\implies \lceil \frac{k}{c-1} \rceil - 1 = \lceil \frac{k+1}{c-1} \rceil - 1$$

$$\implies \lceil \frac{k}{c-1} \rceil = \lceil \frac{k+1}{c-1} \rceil$$

Since (c - 1) $\nmid k$, $\frac{k}{c-1}$ should be a fraction when ceiled, that is $= \lceil \frac{k+1}{c-1} \rceil$.

Case II: $(c-1) \mid k$
To prove case II, the following must be true,

$$n'_{k+1} = k + \lceil \frac{k-c+1}{c-1} \rceil + 2.$$

and we know the following to be true,

$$n'_{k+1} = (k+1) + \lceil \frac{(k+1)}{c-1} - 1 \rceil$$

Substituting these as equal and noting if the equality holds should suffice.

$$\implies k + \lceil \frac{k - c + 1}{c - 1} \rceil + 2 = (k + 1) + \lceil \frac{(k + 1)}{c - 1} - 1 \rceil$$

$$\implies \lceil \frac{k}{c - 1} - 1 \rceil + 1 = \lceil \frac{(k + 1)}{c - 1} - 1 \rceil$$

We can take an integer out of the ceil function

$$\implies \lceil \frac{k}{c - 1} \rceil = \lceil \frac{(k + 1)}{c - 1} \rceil - 1$$

Since (c - 1) | k, $k \mod (c - 1) = [(k + 1) \mod (c - 1)] - 1$, the following equality stands,

$$\implies \lceil \frac{k}{c - 1} \rceil = \lceil \frac{k}{c - 1} \rceil + 1 - 1$$

Since Case I & Case II are true, by the Principle of Mathematical Induction, $P(n)$ is true for all $n \in \mathbb{N}, c > 1$.

□

B Proof for Combined Cancelling Sequence Formula

B.1 Definitions and Notation

Let $C = \{c_1, c_2, \ldots, c_\ell\}$ be a set of pairwise coprime cancelling factors. For any subset $S \subseteq C$, we define:

- $lcm(S)$ as the least common multiple of elements in S
- $|S|$ as the cardinality of S
- a_n as the nth term of the resulting sequence
- A_i as the set of multiples of c_i up to n
- $A_{i_1,\ldots,i_k}$ as the intersection of sets $A_{i_1}, \ldots, A_{i_k}$

B.2 Main Results

Theorem B.1 (Combined Cancelling Sequence Formula)**.** For a set $C = \{c_1, c_2, \ldots, c_\ell\}$ of pairwise coprime cancelling factors acting on a natural arithmetic progression, the nth term of the resulting sequence is given by:

$$\begin{aligned} a_n = n &+ \sum_{i=1}^{\ell} \lceil \frac{n}{c_i - 1} - 1 \rceil \\ &- \sum_{1 \le i < j \le \ell} \lceil \frac{n}{c_i c_j - 1} - 1 \rceil \\ &+ \sum_{1 \le i < j < k \le \ell} \lceil \frac{n}{c_i c_j c_k - 1} - 1 \rceil \\ &+ \cdots + (-1)^{\ell+1} \lceil \frac{n}{c_1 c_2 \cdots c_\ell - 1} - 1 \rceil \end{aligned} \tag{1}$$

B.3 Supporting Lemmas

Lemma B.2 (Coprime Intersection Property)**.** For any subset $S \subseteq C$ of pairwise coprime cancelling factors, a number is divisible by all factors in S if and only if it is divisible by $lcm(S)$, which equals the product of all elements in S.

Proof. Since the elements of S are pairwise coprime, by the fundamental theorem of arithmetic, their least common multiple equals their product. Therefore, if a number is divisible by $lcm(S)$, it is divisible by each factor in S, and conversely, if it is divisible by each factor, it must be divisible by their product. □

Lemma B.3 (Counting Formula for Intersections)**.** For any subset $S \subseteq C$ of size k, the number of integers up to n that are divisible by all elements in S is:

$$|A_S| = \left\lfloor \frac{n}{\prod_{c \in S} c} \right\rfloor \tag{2}$$

Proof. By Lemma B.2, we need to count multiples of $\prod_{c \in S} c$ up to n. This is precisely $\left\lfloor \frac{n}{\prod_{c \in S} c} \right\rfloor$. □

Lemma B.4 (Correction Term Property)**.** The correction term for the overlap of k cancelling factors $c_{i_1}, \ldots, c_{i_k}$ is:

$$\left\lceil \frac{n}{c_{i_1} \cdots c_{i_k} - 1} - 1 \right\rceil \tag{3}$$

Proof. Consider k cancelling factors. By Lemma B.3, their multiples intersect at positions that are multiples of their product. The correction term accounts for the shift in indexing caused by previous cancellations, hence the ceiling function and the subtraction of 1. □

B.4 Main Proof

We prove Theorem B.1 through structural induction on $|C|$.

B.5 Base Cases

Proposition B.5 (Single Cancelling Factor)**.** For $|C| = 1$, the formula reduces to:

$$a_n = n + \lceil \frac{n}{c_1 - 1} - 1 \rceil \tag{4}$$

which correctly counts cancellations of multiples of c_1 as shown in Appendix [A].

Proposition B.6 (Two Cancelling Factors)**.** For $|C| = 2$, the formula:

$$a_n = n + \lceil \frac{n}{c_1 - 1} - 1 \rceil + \lceil \frac{n}{c_2 - 1} - 1 \rceil - \lceil \frac{n}{c_1 c_2 - 1} - 1 \rceil \tag{5}$$

correctly accounts for all cancellations.

Proof. For two coprime cancelling factors c_1 and c_2:

- $|A_1| = \lceil \frac{n}{c_1 - 1} - 1 \rceil$ counts cancellations by c_1
- $|A_2| = \lceil \frac{n}{c_2 - 1} - 1 \rceil$ counts cancellations by c_2
- $|A_{1,2}| = \lceil \frac{n}{c_1 c_2 - 1} - 1 \rceil$ corrects for double-counting

By the inclusion-exclusion principle [Sta11], $|A_1 \cup A_2| = |A_1| + |A_2| - |A_{1,2}|$, which matches our formula. □

B.6 Inductive Step

Theorem B.7 (Inductive Case)**.** If the formula holds for all sets of size k, it holds for sets of size $k + 1$.

Proof. Let the formula hold for $|C| = k$ and consider $C' = C \cup \{c_{k+1}\}$. By the inclusion-exclusion principle:
1. Add cancellations from c_{k+1}: $\lceil \frac{n}{c_{k+1} - 1} - 1 \rceil$
2. Subtract overlaps with each previous factor:

$$-\sum_{i=1}^{k} \lceil \frac{n}{c_i c_{k+1} - 1} - 1 \rceil \tag{6}$$

3. Add triple overlaps:

$$\sum_{1 \le i < j \le k} \lceil \frac{n}{c_i c_j c_{k+1} - 1} - 1 \rceil \tag{7}$$

4. Continue alternating according to the inclusion-exclusion principle.
This exactly matches the form of our formula for $k + 1$ factors. □

B.7 Verification of Properties

Theorem B.8 (Essential Properties)**.** The formula satisfies:

1. Each cancelled term is counted exactly once.
2. Non-cancelled terms remain unchanged.
3. The sequence preserves order between remaining terms.

Proof. :

1. By the inclusion-exclusion principle [Sta11] and Lemma B.2, overlaps are accounted for exactly once.
2. For non-cancelled terms, all ceiling terms sum to zero as no cancellations occur.
3. The monotonicity of the original sequence is preserved for non-cancelled terms as the formula only adds positive integers to account for cancellations.

□

C Complexity Calculations

The computational efficiency [Knu98] of ECCS (0.3) depends on the sum of individual ceiled terms that are being added and subtracted. We parameterise ECCS with n, as it reflects the scaling factor of the problem perfectly. This is because once the n is decided, $\lfloor\sqrt{n}\rfloor$ is fixed, and thereby all the primes being used less than $\lfloor\sqrt{n}\rfloor$ are also determined as $\pi(\lfloor\sqrt{n}\rfloor)$.

Note : The first ceil terms (single primes being added), $\sum_{i=1}^{l}\lceil\frac{n_{c_i}}{c_i-1}-1\rceil$ are basically $O(\pi(\lfloor\sqrt{n}\rfloor))$.

Note : $\pi(n)$ is the prime counting function [HW08].

The second ceil terms (double primes being subtracted), $\sum_{i,j=1,2}^{l}\lceil\frac{n_{c_i}}{lcm(c_i,c_j)-1}-1\rceil$ is again number of primes being considered up to $\lfloor\sqrt{n}\rfloor$, however we are now also considering their combinations, and therefore get $O(\pi(\binom{\lfloor\sqrt{n}\rfloor}{2}))$, where the inner most ( ) pair is a binomial.

Note : We will have to take lcm(s) of pairs, triplets, and so on until $\lfloor\sqrt{n}\rfloor$. It is because of this that we require $O(2.\pi(\binom{\lfloor\sqrt{n}\rfloor}{2}))$

As we would expect this would continue for the remaining summands, giving us the following worst-case complexity for this formula,

$$O(\pi\left[\binom{\lfloor\sqrt{n}\rfloor}{1}\right])+O(2.\pi\left[\binom{\lfloor\sqrt{n}\rfloor}{2}\right])+O(3.\pi\left[\binom{\lfloor\sqrt{n}\rfloor}{3}\right])+...+O(\lfloor\sqrt{n}\rfloor.\pi\left[\binom{\lfloor\sqrt{n}\rfloor}{\lfloor\sqrt{n}\rfloor}\right])$$

We can open up the binomials, however because of lack of algebraic expressions dealing with $\pi(n)$ we really can't be more precise about the asymptotic complexity than this,

$$\implies O\left[\pi(\frac{\lfloor\sqrt{n}\rfloor!}{1!(\lfloor\sqrt{n}\rfloor-1)!})+2.\pi(\frac{\lfloor\sqrt{n}\rfloor!}{2!(\lfloor\sqrt{n}\rfloor-2)!})+3.\pi(\frac{\lfloor\sqrt{n}\rfloor!}{3!(\lfloor\sqrt{n}\rfloor-3)!})+...+\lfloor\sqrt{n}\rfloor.\pi(\frac{\lfloor\sqrt{n}\rfloor!}{\lfloor\sqrt{n}\rfloor!(\lfloor\sqrt{n}\rfloor-\lfloor\sqrt{n}\rfloor)!})\right]$$

Since this is not very informative let us instead make some approximations knowing $\pi(n)\approx\frac{n}{\ln n}$ we can approximate our complexity analysis to reduce each $O(\pi(x))$ terms to $O(\frac{x}{\ln x})$, doing this for each term will help us simplify the expression for complexity even more.

Note : To reduce the pain this notation is causing us let us also assume, $\lfloor\sqrt{n}\rfloor=x$. This will simplify the above statement,

$$\implies O\left[\pi(\frac{x!}{1!(x-1)!})+2.\pi(\frac{x!}{2!(x-2)!})+3.\pi(\frac{x!}{3!(x-3)!})+...+x.\pi(\frac{x!}{x!(x-x)!})\right]$$

Each of the term is of the form, $k.\pi(\frac{x!}{k!(x-k)!})$ where k = 1, 2, 3, ..., x.
Now using the Prime Number Theorem [HW08], we can substitute to get,

$$k.\pi(\binom{x}{k})\approx k.\frac{\binom{x}{k}}{\ln\binom{x}{k}}$$

So, the entire sum becomes,

$$O(\sum_{k=1}^{x}k.\frac{\binom{x}{k}}{\ln\binom{x}{k}})$$

For large x, $\binom{x}{k}\approx\frac{x^k}{k!}$ when k is relatively small. The leading term simplifies to $\frac{x^k}{\ln x}$

Therefore, the asymptotic complexity simplifies to,

$$O(\frac{x^x}{\ln x})=O(\frac{\sqrt{n}^{\sqrt{n}}}{\ln\sqrt{n}})$$

Note : The ceils on the $\sqrt{n}$ have been dropped for the sake of simplicity.
Given how quickly $(\sqrt{n})^{\sqrt{n}}$ grows and the formula becomes infeasible at values as low as n = 24, it is quite obvious that this formula will never be used for any actual evaluation. It is still useful as we shall see for expressing primes within different contexts, important formulae and solutions to conjectures in future papers.