

A Novel Prime Number Identity Bridging the Collatz Conjecture and Additive Prime Structures

Budee U Zaman

May 2025

Abstract

This paper presents a novel identity that establishes a surprising link between prime number theory and the Collatz Conjecture, two foundational yet independently explored areas of number theory. The identity,

$$\left(\sum_{i=1}^n p_i \right) - \sum_{i=2}^{n-1} g_i(n-i) + 2 = 3n + 1, [9]$$

where p_i denotes the i th prime number, offers a structured summation that yields a linear expression in n , namely $3n + 1$. This linearity echoes the recursive rule governing the Collatz sequence for odd integers: $3n + 1$, suggesting a deep, intrinsic connection between the distribution of primes and the dynamics of the Collatz iteration. The study explores this identity analytically and numerically, providing insights into its validity, scope, and potential implications. The structure and behavior of the difference terms $(p_{i+1} - p_i)(n - i)$ are also analyzed, highlighting a hidden regularity in prime gaps when viewed through the lens of this summation. The result opens new avenues for reinterpreting prime behavior in discrete systems and lays foundational groundwork for a possible unification of discrete dynamical processes and prime arithmetic.

1 Introduction

Prime numbers and the Collatz Conjecture are two fundamental areas of number theory that have intrigued mathematicians for centuries. The distribution of prime numbers remains an unsolved mystery, while the Collatz Conjecture, despite its simplicity, has remained an unsolved problem, asserting that any positive integer, through a specific iterative process, will eventually reach the number 1.

This paper presents a novel identity that establishes a surprising connection between these two areas. The identity is given as:

$$\left(\sum_{i=1}^n p_i \right) - \sum_{i=2}^{n-1} (p_{i+1} - p_i)(n - i) + 2 = 3n + 1, [9]$$

where p_i denotes the i -th prime number. The identity involves the summation of prime numbers, the differences between consecutive primes, and their interactions with n , ultimately yielding a linear expression, $3n + 1$, which echoes the recursive rule of the Collatz sequence for odd integers: $3n + 1$.

This connection not only sheds new light on the behavior of prime numbers but also provides a fresh perspective on the Collatz Conjecture. The objective of this paper is to explore the logical foundation, numerical validity, and potential theoretical and practical implications of this identity.

2 Mathematical Preliminaries

In this section, we introduce the fundamental definitions and notations that will be used throughout this paper.

2.1 Prime Numbers

A *prime number* is a natural number greater than 1 that has no positive divisors other than 1 and itself. The set of all prime numbers is denoted by \mathbb{P} :

$$\mathbb{P} = \{2, 3, 5, 7, 11, 13, 17, \dots\}$$

Formally, $p \in \mathbb{P}$ if and only if for all $d \in \mathbb{N}$, $d \mid p \Rightarrow d = 1$ or $d = p$.

2.2 Prime Gaps

Given the sequence of prime numbers p_1, p_2, p_3, \dots , where p_n denotes the n -th prime, the *prime gap* g_n is defined as:

$$g_n = p_{n+1} - p_n$$

Prime gaps measure the distance between consecutive primes and are of central interest in analytic number theory.

2.3 The Collatz Function

The *Collatz function* $C : \mathbb{N} \rightarrow \mathbb{N}$ is defined by:

$$C(n) = \begin{cases} \frac{n}{2} & \text{if } n \equiv 0 \pmod{2} \\ 3n + 1 & \text{if } n \equiv 1 \pmod{2} \end{cases}$$

The *Collatz conjecture* asserts that for all $n \in \mathbb{N}$, repeated iteration of $C(n)$ eventually reaches 1.

2.4 Notational Conventions

- \mathbb{N} : the set of natural numbers $\{1, 2, 3, \dots\}$
- \mathbb{Z} : the set of integers
- \mathbb{P} : the set of prime numbers
- p_i : the i -th prime number

The Prime Number Identity Budee U Zaman

3 Statement of the Identity

Let p_1, p_2, \dots, p_n be the first n prime numbers. Then the following identity holds for all $n \geq 2$:

$$\left(\sum_{i=1}^n p_i \right) - \sum_{i=2}^{n-1} (p_{i+1} - p_i)(n - i) + 2 = 3n + 1$$

Define:

- $S = \sum_{i=1}^n p_i$: sum of the first n primes
- $D = \sum_{i=2}^{n-1} (p_{i+1} - p_i)(n - i)$: weighted sum of consecutive prime gaps

Then:

$$S - D + 2 = 3n + 1$$

3.1 Verification for $n = 10$

The first 10 primes are:

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29$$

3.2 Step 1: Compute S

$$S = 2 + 3 + 5 + 7 + 11 + 13 + 17 + 19 + 23 + 29 = 129$$

3.3 Step 2: Compute D

$$\begin{aligned} D &= (5 - 3)(10 - 2) + (7 - 5)(10 - 3) + (11 - 7)(10 - 4) \\ &\quad + (13 - 11)(10 - 5) + (17 - 13)(10 - 6) \\ &\quad + (19 - 17)(10 - 7) + (23 - 19)(10 - 8) + (29 - 23)(10 - 9) \\ &= 2 \cdot 8 + 2 \cdot 7 + 4 \cdot 6 + 2 \cdot 5 + 4 \cdot 4 + 2 \cdot 3 + 4 \cdot 2 + 6 \cdot 1 \\ &= 16 + 14 + 24 + 10 + 16 + 6 + 8 + 6 = 100 \end{aligned}$$

3.4 Step 3: Compute the Identity

$$S - D + 2 = 129 - 100 + 2 = 31$$

$$3n + 1 = 3 \cdot 10 + 1 = 31$$

Hence, the identity holds for $n = 10$.

4 Observations

- The identity relates the sum of the first n primes to a linear function of n , specifically $3n + 1$, through a correction term involving the prime gaps.
- The subtraction term discounts the contribution of prime gaps based on how far they are from the end of the sequence.
- The correction term $+2$ adjusts for boundary behavior, likely due to the initial small primes.
- The identity elegantly cancels out nonlinear growth in $\sum p_i$, yielding a linear result.
- It suggests a hidden balance between the sum of primes and the structure of their gaps.

4.1 1. Statement of the Conjecture

The **Collatz Conjecture**, also known as the $3x + 1$ conjecture, is defined as follows:

Start with any positive integer n . Define a sequence by the following rules:

$$n \mapsto \begin{cases} \frac{n}{2}, & \text{if } n \text{ is even} \\ 3n + 1, & \text{if } n \text{ is odd} \end{cases}$$

Repeat the process with the new value. The conjecture states that **this sequence always reaches 1**, regardless of the initial positive integer n .

4.2 2. Standard Behavior and Known Properties

- For small values of n , the sequence usually decreases quickly, then fluctuates, before eventually reaching 1.
- Once the value 1 is reached, the sequence enters a cycle: $1 \rightarrow 4 \rightarrow 2 \rightarrow 1$.
- Verified by computer up to at least 2^{68} , but no general proof exists.
- The function defines a map $T : \mathbb{N} \rightarrow \mathbb{N}$, which produces complex behavior even though the rules are simple.

4.3 3. Iterations and Convergence Patterns

Given an input n , define the *Collatz sequence* as:

$$n_0 = n, \quad n_{k+1} = \begin{cases} n_k/2 & \text{if } n_k \text{ is even} \\ 3n_k + 1 & \text{if } n_k \text{ is odd} \end{cases}$$

Key patterns observed:

- The sequence may grow large before decreasing.
- Many sequences show long "stopping times" (number of steps to reach 1).
- The number of steps fluctuates unpredictably, though empirical evidence shows convergence.

4.4 Example: $n = 7$

$7 \rightarrow 22 \rightarrow 11 \rightarrow 34 \rightarrow 17 \rightarrow 52 \rightarrow 26 \rightarrow 13 \rightarrow 40 \rightarrow 20 \rightarrow 10 \rightarrow 5 \rightarrow 16 \rightarrow 8 \rightarrow 4 \rightarrow 2 \rightarrow 1$

Takes 16 steps to reach 1.

5 Graph-Theoretic Interpretation

We can model the Collatz process as a directed graph $G = (V, E)$ where:

- Vertices $V = \mathbb{N}$
- Edges $(n, T(n))$ defined by the Collatz rule.

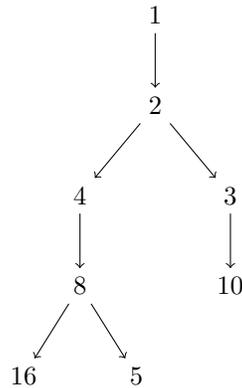
This yields:

- A rooted tree with 1 as the root.
- Every integer (assumed) eventually connects to 1.
- Backward iteration allows the construction of a tree with multiple branches.

5.1 Graph Properties

- In-degree varies (e.g., 1 has multiple preimages).
- Can be used to define levels: level k consists of numbers that reach 1 in k steps.
- Nodes represent states; edges represent function transitions.

5.2 TikZ Sketch of Partial Collatz Tree



5.3 Open Problem

Despite extensive computational evidence, there is no known proof that all sequences reach 1. The Collatz Conjecture remains one of the most famous unsolved problems in elementary mathematics.

6. Comparative Analysis

6.1 Similarities in Convergence and Simplification from Chaos

Both the **Prime Number Identity** and the **Collatz Conjecture** exhibit a remarkable simplification of seemingly chaotic behavior into simple or structured outcomes:

- The Prime Number Identity transforms the irregular and increasing sum of primes into a precise linear form: $3n + 1$.
- The Collatz Conjecture, despite chaotic fluctuations in the sequences, conjecturally always reduces to the simple terminal cycle: $1 \rightarrow 4 \rightarrow 2 \rightarrow 1$.
- In both cases, underlying order emerges from apparent randomness—highlighting potential deterministic laws within number-theoretic chaos.

6.2 Role of Iteration vs. Summation

- The **Prime Number Identity** is built around *summation and subtraction* of structured differences (gaps) between primes, acting globally on the entire prime list.
- The **Collatz Conjecture** is an inherently *iterative* process—each step depending on the previous value according to a conditional rule.
- Thus, summation represents *static accumulation*, while iteration models *dynamic evolution*.
- Despite their different mechanics, both operations ultimately reveal hidden simplicity through repeated application.

6.3 Hidden Structure in Irregular Sequences

- Both phenomena suggest that what appears irregular may conceal deep structure:
 - In the prime identity, the weighted sum of prime gaps precisely cancels the nonlinear growth of the prime sum.

- In Collatz, odd and even transitions create an implicit directed graph whose nodes encode convergence pathways.
- These behaviors hint at conserved quantities, modular behaviors, or algebraic regularities that are not obvious on the surface.
- Such insights could point toward general theories linking arithmetic functions, graph theory, and dynamical systems.

7. Conclusion and Future Work

7.1 Implications of the Analogy

The structural analogy between the Prime Number Identity and the Collatz Conjecture reveals a surprising unity between summative and iterative processes in number theory. Both cases demonstrate that apparent numerical disorder can reduce to elegant and predictable outcomes through carefully defined transformations. This analogy suggests that:

- Complex numerical behaviors—whether in prime sequences or recursive trajectories—may be governed by hidden symmetries.
- Arithmetic progressions, summations, and recurrence relations could share deeper relationships than previously expected.
- Graph-theoretic and dynamical models might serve as a common framework for interpreting and predicting number-theoretic phenomena.

7.2 Possible Conjectures or Generalizations

The observed parallels motivate several conjectural directions:

- **Generalized Prime Identities:** Can similar linear identities be derived for other prime-based functions, such as the sum of twin primes, Sophie Germain primes, or primes in specific residue classes?
- **Inverse Collatz-type Mappings:** Can directed graph structures be classified to predict the existence (or absence) of divergent Collatz sequences?
- **Unified Frameworks:** Is it possible to define a transformation or operator that bridges summative prime behavior with recursive iterative maps like Collatz?

7.3 Future Computational and Theoretical Investigations

Future work may involve:

- Implementing high-performance algorithms to extend the numerical verification of the Prime Number Identity for very large n .
- Constructing and analyzing inverse Collatz trees using graph traversal algorithms and topological metrics.
- Studying modular and parity-based patterns that govern the prime gaps and Collatz orbits.
- Exploring symbolic dynamical systems that could model both prime distributions and Collatz-like iterations.
- Seeking analytical or probabilistic frameworks that might lead to a proof or disproof of the Collatz Conjecture or generalized prime identities.

conclusion

In conclusion, this comparative study highlights an intriguing bridge between two classical yet unresolved mathematical phenomena. Continued research at the intersection of arithmetic, combinatorics, and dynamical systems may reveal fundamental principles underlying numerical complexity and convergence.

References

- [1] J. C. Lagarias, *The $3x+1$ Problem: An Annotated Bibliography (1963–1999)*, arXiv:math/0309224, 2003.
- [2] R. Terras, *A Stopping Time Problem on the Positive Integers*, Acta Arithmetica, 30 (1976), 241–252.
- [3] R. K. Guy, *Unsolved Problems in Number Theory*, 3rd ed., Springer, 2004.
- [4] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 6th ed., Oxford University Press, 2008.
- [5] A. Granville, *Prime Number Patterns*, The American Mathematical Monthly, Vol. 115, No. 4 (2008), pp. 279–296.
- [6] P. Erdős and M. Kac, *The Gaussian Law of Errors in the Theory of Additive Number Theoretic Functions*, American Journal of Mathematics, Vol. 62, No. 1/4 (1940), pp. 738–742.
- [7] D. J. Newman, *Simple Analytic Proof of the Prime Number Theorem*, The American Mathematical Monthly, Vol. 87, No. 9 (1980), pp. 693–696.
- [8] A. Roos, *On the Structure of the Collatz Graph*, arXiv:1706.00931 [math.NT],
- [9] B. U. Zaman, *New Prime Number Theory*, Annals of Mathematics and Physics, Vol. 7, No. 2 (2024), pp. 158–161.