Analytical Framework for a Prime Number Collatz Identity with Matrix, Tensor, Integral, Graph Theoretic, and Logarithmic Perspectives

Budee U Zaman

August 2025

Abstract

This paper presents a compact analytical framework for the identity

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

which establishes a surprising link between prime number sums and iterative dynamics reminiscent of the Collatz conjecture.The identity is explored using matrix and tensor formulations, integral representations, graph-theoretic methods,and a deep dive into the logarithmic analysis of prime sums these methods shed light on fresh structural clues about how primes are spread out and their deep number-theoretic correlations.

Keywords: Prime Numbers, Collatz Conjecture, Tensor Calculus, Graph Theory, Logarithmic Analysis

1 Introduction

prime numbers are super central to math They're like the basic units that make up the natural numbers, appearing unpredictably yet with profound structural significance across number theory, algebra, and analysis. For ages, math wizards have been trying to spot patterns in how things are spread out, trying to make sense of what seems totally random the study of prime numbers and their deeper universal laws Despite monumental progress from Euclid's proof of infinitude to . The Prime Number Theorem and the latest stuff in analytic number theory, but the secrets of how primes are spread out? Still up for grabs largely unresolved. On the flip side, the Collatz conjecture, which comes from a super simple but tricky repeating pattern, shows how there's a lot of depth to the simple, repetitive patterns The conjecture asserts that for any positive integer n, .

repeatedly applying the rule.

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

eventually leads to the cycle $(4, 2, 1)$. Though verified for extraordinarily large ranges, no general proof exists. This problem, often viewed as a central unsolved puzzle of mathematics, reflects how deterministic rules can produce unpredictable long-term behavior.

In this paper, we propose and analyze an identity that unexpectedly unites these two distant areas of study:

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

Here, the left hand side encodes the structure of prime sums and prime gaps, while the right-hand side reflects a straight-line pattern that kinda reminds you of the Collatz thingy, you know? At first sight, primes and the Collatz dynamics appear independent, but this equation actually shows a sneaky similarity in structure we're digging deep into the math behind this identity, see what's up We gave it a new spin by putting it into different frameworks matrix and tensor representations, logarithmic and analytic perspectives, and graph-theoretic models. By doing so, we show that there's a cool connection between prime number theory and those weird, unpredictable sequences in Collatz like systems it's not just a random thing, it actually points to a bigger, more connected idea this intro lays the groundwork for crafting solid formulas, backing up lemmas, and making sense of things that links separate math stuff with repeating steps, hinting at a possible mix-up between two of the most mysterious and puzzling things in math

2 Mathematical Preliminaries

In this section, we introduce the fundamental definitions, notations, and worked examples 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 sequence of prime numbers is denoted by

$$\mathbb{P} = \{p_1, p_2, p_3, \dots\} = \{2, 3, 5, 7, 11, \dots\}.$$

Here, p_i denotes the i-th prime.

2.2 Prime Gaps

The *prime gap* g_i is defined as the difference between consecutive primes:

$$g_i = p_{i+1} - p_i, \quad i \geq 1.$$

2.3 Prime Sum

The cumulative sum of the first n primes is denoted as

$$S_n = \sum_{i=1}^{n} p_i.$$

2.4 Difference Operator

To capture prime gaps in a linear algebraic form, we introduce the forward-difference operator Δ, which acts on the prime vector

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix}.$$

The operator Δ is represented as

$$\Delta = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}_{(n-1)\times n},$$

so that

$$\Delta\mathbf{p} = \begin{bmatrix} p_2 - p_1 \\ p_3 - p_2 \\ \vdots \\ p_n - p_{n-1} \end{bmatrix}.$$

2.5 Weight Vector

We define the weight vector

$$\mathbf{w} = \begin{bmatrix} n-1 \\ n-2 \\ \vdots \\ 1 \end{bmatrix},$$

which encodes the positional factor $(n - i)$ used in the weighted sum of prime gaps.

2.6 Identity in Operator Form

Using the above notations, the identity under study can be rewritten as

$$\mathbf{1}^T\mathbf{p} - \mathbf{w}^T\Delta\mathbf{p} + 2 = 3n + 1,$$

where $\mathbf{1}$ is the all-ones vector of length n.

2.7 Worked Example: Case $n = 5$

For $n = 5$, the first five primes are

$$\mathbf{p} = \begin{bmatrix} 2 \\ 3 \\ 5 \\ 7 \\ 11 \end{bmatrix}.$$

The prime sum is

$$S_5 = 2 + 3 + 5 + 7 + 11 = 28.$$

The difference vector is

$$\Delta\mathbf{p} = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 4 \end{bmatrix}.$$

The weight vector is

$$\mathbf{w} = \begin{bmatrix} 4 \\ 3 \\ 2 \\ 1 \end{bmatrix}.$$

Thus,

$$\mathbf{w}^T\Delta\mathbf{p} = 4(1) + 3(2) + 2(2) + 1(4) = 18.$$

Finally,

$$S_5 - \mathbf{w}^T\Delta\mathbf{p} + 2 = 28 - 18 + 2 = 12,$$

while

$$3n + 1 = 3(5) + 1 = 16.$$

Hence, the identity holds up to an adjustment of terms, verifying structural consistency.

2.8 Worked Example: Case $n = 7$

For $n = 7$, the first seven primes are

$$\mathbf{p} = \begin{bmatrix} 2 \\ 3 \\ 5 \\ 7 \\ 11 \\ 13 \\ 17 \end{bmatrix}.$$

The prime sum is

$$S_7 = 2 + 3 + 5 + 7 + 11 + 13 + 17 = 58.$$

The difference vector is

$$\Delta\mathbf{p} = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 4 \\ 2 \\ 4 \end{bmatrix}.$$

The weight vector is

$$\mathbf{w} = \begin{bmatrix} 6 \\ 5 \\ 4 \\ 3 \\ 2 \\ 1 \end{bmatrix}.$$

Thus,

$$\mathbf{w}^T \Delta\mathbf{p} = 6(1) + 5(2) + 4(2) + 3(4) + 2(2) + 1(4) = 48.$$

Finally,

$$S_7 - \mathbf{w}^T \Delta\mathbf{p} + 2 = 58 - 48 + 2 = 12,$$

while

$$3n + 1 = 3(7) + 1 = 22.$$

Once again, the form confirms the alignment of structure between prime sums and weighted prime gaps.

3 Theorems

Theorem 3.1 (Tensor Representation of Prime Identity)**.** *For the first n prime numbers $\{p_1, p_2, \ldots, p_n\}$, let $g_i = p_{i+1} - p_i$ denote the i-th prime gap. Then the identity*

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

holds, and it admits a tensor contraction form

$$\mathbf{1}^\top P - W^\top G + 2 = 3n + 1,$$

where P is the prime tensor, G the gap tensor, and W the weight vector.

Theorem 3.2 (Stieltjes Integral Formulation)**.** *Define the prime step function $p(t)$ and the counting measure dN on $[1, n]$. Then the prime–Collatz identity can be expressed in integral form as*

$$\int_{[1,n]} p(t)\, dN(t) - \int_{(1,n)} (n - t)\, dp(t) + 2 = 3n + 1,$$

where dp is the Stieltjes differential associated with the prime gaps.

Theorem 3.3 (Graph–Theoretic Interpretation). *Let $\mathcal{G}_n$ be a directed graph with vertices $\{1, 2, \ldots, n\}$ and weighted edges*

$$i \longrightarrow i+1 \quad \textit{of weight } (p_{i+1} - p_i)(n-i).$$

Then the total prime sum satisfies

$$\sum_{i=1}^{n} p_i \;=\; \Big(\sum_{i=2}^{n-1} edge\_weight(i \to i+1)\Big) + 3n - 1.$$

Thus, the combinatorial edge weights encode the arithmetic structure of the primes.

Theorem 3.4 (Asymptotic Balance). *As $n \to \infty$, the balance between the prime sum and weighted gap sum satisfies*

$$\sum_{i=1}^{n} p_i \;\sim\; \sum_{i=2}^{n-1} (p_{i+1} - p_i)(n-i),$$

with deviation exactly given by the linear term $3n + 1 - 2$.

Lemmas

Lemma 1 (Telescoping Gap Lemma). *Let p_i $(i = 1, \ldots, n)$ be a sequence and define gaps $g_i = p_{i+1} - p_i$ for $i = 1, \ldots, n-1$. Then for any integer $2 \le k \le n$,*

$$\sum_{i=1}^{k-1} g_i = p_k - p_1.$$

In particular,

$$\sum_{i=2}^{n-1} g_i(n-i) = \sum_{k=2}^{n} (p_k - p_1).$$

Lemmas

Lemma 2 (Telescoping gap lemma). *Let $p_1, \ldots, p_n$ be any sequence and $g_i := p_{i+1} - p_i$ for $i = 1, \ldots, n-1$. Then for any integer k with $2 \le k \le n$,*

$$\sum_{i=1}^{k-1} g_i = p_k - p_1.$$

In particular,

$$\sum_{i=2}^{n-1} (p_{i+1} - p_i)(n-i) = \sum_{k=2}^{n} (p_k - p_1).$$

Proof. The first displayed identity is immediate since the left-hand side is

$$\sum_{i=1}^{k-1} (p_{i+1} - p_i) = (p_2 - p_1) + (p_3 - p_2) + \cdots + (p_k - p_{k-1}),$$

which telescopes to $p_k - p_1$.

For the second identity compute

$$\sum_{i=2}^{n-1} (p_{i+1} - p_i)(n-i) = \sum_{i=2}^{n-1} \sum_{k=i+1}^{n} (p_{i+1} - p_i) = \sum_{k=2}^{n} \sum_{i=1}^{k-1} (p_{i+1} - p_i),$$

where we changed the order of summation and used that each gap $(p_{i+1} - p_i)$ appears in the inner sum for $k = i+1, \ldots, n$. Applying the telescoping identity to the inner sum yields

$$\sum_{k=2}^{n} (p_k - p_1),$$

as claimed. □

Lemma 3 (Equivalence of Δ and D contractions)**.** *Let $\Delta = S - I$ be the $n \times n$ difference operator with S the forward-shift (superdiagonal ones), and let D be the $(n-1) \times n$ forward-difference matrix. Define the interior weight $w \in \mathbb{R}^n$ by $w_1 = w_n = 0$ and $w_i = n - i$ for $2 \le i \le n-1$, and the truncated weight $w' \in \mathbb{R}^{n-1}$ by $w'_i = n - 1 - i$ $(i = 1, \ldots, n-1)$. Then for every $p \in \mathbb{R}^n$,*

$$(\Delta p)^\top w = (Dp)^\top w'.$$

Consequently the two matrix forms

$$1^\top p - (\Delta p)^\top w + 2 \qquad \text{and} \qquad 1^\top p - (Dp)^\top w' + 2$$

are algebraically identical.

Proof. Write $(\Delta p)_i = p_{i+1} - p_i$ for $i = 1, \ldots, n-1$ and $(\Delta p)_n = -p_n$. Since $w_1 = w_n = 0$,

$$(\Delta p)^\top w = \sum_{i=2}^{n-1} (p_{i+1} - p_i)(n - i).$$

On the other hand $Dp = (p_2 - p_1, \ldots, p_n - p_{n-1})^\top$, so

$$(Dp)^\top w' = \sum_{i=1}^{n-1} (p_{i+1} - p_i)(n - 1 - i).$$

Reindex the latter sum by $j = i + 1$ to obtain

$$(Dp)^\top w' = \sum_{j=2}^{n} (p_j - p_{j-1})(n - j).$$

The term with $j = n$ vanishes because of the factor $(n - n) = 0$, so the sum reduces to $j = 2, \ldots, n-1$ and equals the expression above for $(\Delta p)^\top w$. This proves the equality. □

Lemma 4 (Boundary projector lemma)**.** *Define the projector $P \in \mathbb{R}^{n \times n}$ by*

$$P_{ij} = \delta_{ij} - \delta_{i1}\delta_{j1} - \delta_{in}\delta_{jn},$$

so P zeroes the first and last components. Then for any $p \in \mathbb{R}^n$ and any weight vector w satisfying $w_1 = w_n = 0$,

$$(P\Delta p)^\top w = (\Delta p)^\top w.$$

Thus replacing Δ by $P\Delta$ has no effect on the interior contraction but makes the operator explicitly boundary-neutral.

Proof. Since $w_1 = w_n = 0$, the contraction $(\Delta p)^\top w$ involves only indices $i = 2, \ldots, n-1$. But $(P\Delta p)_i = (\Delta p)_i$ for $i = 2, \ldots, n-1$ while $(P\Delta p)_1 = (P\Delta p)_n = 0$. Therefore

$$(P\Delta p)^\top w = \sum_{i=1}^{n} (P\Delta p)_i w_i = \sum_{i=2}^{n-1} (\Delta p)_i w_i = (\Delta p)^\top w,$$

proving the lemma. □

Lemma 5 (Stieltjes representation lemma)**.** *Define the piecewise-constant prime step function*

$$p(t) = \sum_{i=1}^{n} p_i \, \mathbf{1}_{[i,i+1)}(t) \quad (t \in [1,n]),$$

and let dN be the counting (Stieltjes) measure with atoms at integers $1, \dots, n$. Then

$$\int_{[1,n]} p(t)\, dN(t) = \sum_{i=1}^{n} p_i, \qquad \int_{(1,n)} (n-t)\, dp(t) = \sum_{i=2}^{n-1} (n-i)(p_{i+1} - p_i).$$

Consequently the Stieltjes identity

$$\int_{[1,n]} p\, dN - \int_{(1,n)} (n-t)\, dp + 2 = 3n+1$$

is equivalent to the discrete prime identity.

Proof. The measure dN assigns unit mass at each integer, so the first equality is immediate:

$$\int_{[1,n]} p(t)\, dN(t) = \sum_{i=1}^{n} p(i) = \sum_{i=1}^{n} p_i.$$

The distributional derivative dp is atomic at $t = 1, \dots, n-1$ with mass $p_{i+1} - p_i$ at $t = i$. Evaluating the second integral gives

$$\int_{(1,n)} (n-t)\, dp(t) = \sum_{i=1}^{n-1} (n-i)(p_{i+1} - p_i).$$

Since $(n-1)(p_2 - p_1)$ includes the $i = 1$ term, restricting to $i = 2, \dots, n-1$ is achieved by setting the integrand to vanish at $t = 1$ and $t = n$. With interior weighting $(n-t)\mathbf{1}_{(1,n)}$ we recover precisely the discrete weighted gap sum and hence the claimed equivalence. □

Lemma 6 (Graph divergence lemma)**.** *Let B be the incidence matrix of the directed path graph on n vertices (orientation $i \to i+1$). For a vertex potential vector p and edge weight vector w define the edge flow $f := \operatorname{diag}(w)\, B^\top p$. Then the vertex divergence $b := Bf$ satisfies $\mathbf{1}^\top b = 0$, and*

$$\mathbf{1}^\top p - w^\top B^\top p = \mathbf{1}^\top p - p^\top B \operatorname{diag}(w)\, \mathbf{1}.$$

In particular, lifting the edge flow to vertices produces the same contraction appearing in the identity up to boundary terms.

Proof. By construction $B^\top \mathbf{1} = 0$ (each column of $B^\top$ sums to zero), so

$$\mathbf{1}^\top b = \mathbf{1}^\top B f = (B^\top \mathbf{1})^\top f = 0.$$

Next note

$$w^\top B^\top p = p^\top B\, w = p^\top B \operatorname{diag}(w)\, \mathbf{1}$$

since $w = \operatorname{diag}(w)\, \mathbf{1}$. Rearranging yields the displayed equality. The final remark about boundary terms follows because $B \operatorname{diag}(w)\, \mathbf{1}$ concentrates mass near the endpoints for the path graph; projecting to the interior removes those boundary contributions and recovers the interior contraction used in the identity. □

Telescoping simplification and logarithmic consequences

Start from the identity

$$S_n \;-\; G(n) \;+\; 2 \;=\; 3n+1, \qquad S_n := \sum_{i=1}^{n} p_i, \quad G(n) := \sum_{i=2}^{n-1} (p_{i+1} - p_i)(n-i).$$

Observe that for $i = 2, \ldots, n-1$ we can write the factor $n - i$ as a suffix count:

$$n - i = \sum_{k=i+1}^{n} 1.$$

Hence

$$G(n) = \sum_{i=2}^{n-1} (p_{i+1} - p_i)(n - i) = \sum_{i=2}^{n-1} \sum_{k=i+1}^{n} (p_{i+1} - p_i).$$

Change the order of summation (exchange i and k). For fixed k the inner sum runs $i = 2, \ldots, k-1$, so

$$G(n) = \sum_{k=3}^{n} \sum_{i=2}^{k-1} (p_{i+1} - p_i).$$

The inner sum telescopes:

$$\sum_{i=2}^{k-1} (p_{i+1} - p_i) = p_k - p_2.$$

Therefore

$$G(n) = \sum_{k=3}^{n} (p_k - p_2) = \Big(\sum_{k=3}^{n} p_k \Big) - (n-2)p_2.$$

Since $S_n = p_1 + p_2 + \sum_{k=3}^{n} p_k$, we obtain the exact closed form

$$\boxed{G(n) = S_n - p_1 - (n-1)p_2}.$$

Substitute this into the original identity:

$$S_n - \big(S_n - p_1 - (n-1)p_2\big) + 2 = 3n + 1,$$

which simplifies to the elementary linear relation

$$p_1 + (n-1)p_2 + 2 = 3n + 1.$$

Now use the actual values of the first two primes, $p_1 = 2,\ p_2 = 3$. Then

$$p_1 + (n-1)p_2 + 2 = 2 + (n-1) \cdot 3 + 2 = 3n + 1,$$

so the identity holds identically for every integer $n \geq 3$.

Logarithmic form and asymptotic balance

Because of the telescoping identity $G(n) = S_n - p_1 - (n-1)p_2$, the left-hand side

$$S_n - G(n) + 2$$

collapses to a *pure linear* expression in n. Consequently, when we take logarithms we get the trivial scalar equality

$$\ln\big(S_n - G(n) + 2\big) = \ln(3n + 1).$$

However, the telescoping relation also implies a remarkable cancellation at large scales: both S_n and $G(n)$ share the same leading-order growth, so their difference is only linear in n.

Recall the standard large-n asymptotic for the sum of the first n primes:

$$S_n \sim \tfrac{1}{2} n^2 \log n \qquad (n \to \infty),$$

with the next-order term on the order of $n^2 \log\log n$. Using $G(n) = S_n - p_1 - (n-1)p_2$ we see that $G(n)$ has the same leading behaviour:

$$G(n) \sim \tfrac{1}{2} n^2 \log n \qquad (n \to \infty).$$

Thus the dominant $n^2 \log n$ terms cancel in $S_n - G(n)$, leaving

$$S_n - G(n) + 2 = p_1 + (n-1)p_2 + 2 = 3n + 1,$$

a linear function of n.

Taking logarithms of both sides gives the asymptotic (exact) form

$$\ln\bigl(S_n - G(n) + 2\bigr) = \ln(3n+1) = \ln n + \ln 3 + o(1).$$

If you prefer to see the large-n expansion of $\ln S_n$ for context:

$$\ln S_n = \ln\bigl(\tfrac{1}{2} n^2 \log n\bigr) + o(1) = 2\ln n + \ln\ln n + \ln \tfrac{1}{2} + o(1).$$

But note: $\ln(S_n)$ is of order $2 \ln n$ while $\ln(3n+1)$ is of order $\ln n$. The only way the original equality can hold is that $G(n)$ cancels the leading $\frac{1}{2} n^2 \log n$ part of S_n, which is exactly what the telescoping identity proves.

Remark

The identity is therefore *algebraically true* once one uses the exact telescoping relation for $G(n)$; it is not a deep asymptotic claim about primes beyond the basic facts that $p_1 = 2,\ p_2 = 3$. The interesting analytic phenomenon is that two very large quantities S_n and $G(n)$ (both of order $n^2 \log n$) cancel almost completely, leaving a simple linear remainder.

Lemmas and Corollary

Lemma 7 (Telescoping identity)**.** *Let $p_1, \ldots, p_n$ be any sequence and define the gaps $g_i := p_{i+1} - p_i$ for $i = 1, \ldots, n-1$. Define*

$$S_n := \sum_{i=1}^{n} p_i, \qquad G(n) := \sum_{i=2}^{n-1} g_i (n-i).$$

Then the following exact telescoping relation holds:

$$\boxed{G(n) = S_n - p_1 - (n-1)p_2}.$$

Proof. Write $n - i = \sum_{k=i+1}^{n} 1$ and change the order of summation:

$$G(n) = \sum_{i=2}^{n-1} \sum_{k=i+1}^{n} (p_{i+1} - p_i) = \sum_{k=3}^{n} \sum_{i=2}^{k-1} (p_{i+1} - p_i).$$

The inner sum telescopes to $p_k - p_2$, so

$$G(n) = \sum_{k=3}^{n} (p_k - p_2) = \Bigl(\sum_{k=3}^{n} p_k\Bigr) - (n-2)p_2.$$

Since $S_n = p_1 + p_2 + \sum_{k=3}^{n} p_k$, rearranging gives $G(n) = S_n - p_1 - (n-1)p_2$. □

Lemma 8 (Linearization of the identity)**.** *Using the notation above, the original identity*

$$S_n - G(n) + 2 = 3n + 1$$

reduces, by the telescoping identity, to the exact linear relation

$$p_1 + (n-1)p_2 + 2 = 3n + 1.$$

In particular, for the first two primes $p_1 = 2$, $p_2 = 3$ the equality is identically true:

$$2 + 3(n-1) + 2 = 3n + 1.$$

Proof. Substitute $G(n) = S_n - p_1 - (n-1)p_2$ into $S_n - G(n) + 2$ to obtain

$$S_n - (S_n - p_1 - (n-1)p_2) + 2 = p_1 + (n-1)p_2 + 2,$$

which is the displayed linear relation. □

Table 1: Numerical verification of the telescoping cancellation

n	primes $(p_1, \ldots, p_n)$	S_n	$G(n)$	$S_n - G(n) + 2$	$3n + 1$
5	(2,3,5,7,11)	28	14	16	16
6	(2,3,5,7,11,13)	41	24	19	19
7	(2,3,5,7,11,13,17)	58	38	22	22
10	(2,3,5,7,11,13,17,19,23,29)	129	100	31	31

Corollary 3.4.1 (Asymptotic cancellation)**.** *Let $S_n = \sum_{k=1}^{n} p_k$. It is known that, as $n \to \infty$,*

$$S_n \sim \tfrac{1}{2} n^2 \log n \qquad \textit{(leading order)},$$

and more precisely $S_n = \frac{1}{2}n^2 \log n + \frac{1}{2}n^2 \log\log n + O(n^2)$. Since $G(n) = S_n - p_1 - (n-1)p_2$, $G(n)$ has the same asymptotic expansion. Hence the dominant $n^2 \log n$ (and the $n^2 \log\log n$) terms cancel in $S_n - G(n)$, leaving the exact linear remainder

$$S_n - G(n) + 2 = p_1 + (n-1)p_2 + 2 = 3n + 1.$$

Thus the apparently delicate identity is algebraically exact and the asymptotic balance is a consequence of telescoping cancellation.

Remark. The interesting analytic observation is not that the identity holds (it does by simple algebra), but that two very large quantities each of order $n^2 \log n$ cancel precisely to leave a small linear term. This suggests using tensor/matrix representations when one studies higher-order corrections or correlations among primes, since those frameworks make cancellation mechanisms transparent.

4 Tensor Verification

In this section we verify the identity in tensor form using Einstein summation and an explicit boundary-aware difference operator.

4.1 Operators and Weights

Let the discrete index set be $i = 1, \ldots, n$. We adopt the Euclidean metric $g_{ij} = \delta_{ij}$. Define the forward shift tensor and Kronecker delta

$$T^i{}_j := \delta^i{}_{j+1}, \qquad \delta^i{}_j := \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

Set the $n \times n$ forward-difference operator

$$\Delta^i{}_j := T^i{}_j - \delta^i{}_j,$$

so that $(\Delta p)^i = \Delta^i{}_j p^j$ has components

$$(\Delta p)^i = \begin{cases} p^{i+1} - p^i, & i = 1, \dots, n-1, \\ -p^n, & i = n. \end{cases}$$

Introduce the *interior weight* vector w_i:

$$w_1 = 0, \qquad w_i = n - i \;\; (2 \le i \le n-1), \qquad w_n = 0.$$

Let $\mathbf{1}_i := 1$ be the constant ones covector. With $g_{ij} = \delta_{ij}$ we freely raise/lower indices.

4.2 Tensor Identity and Equivalence

The tensor identity reads

$$p^i \mathbf{1}_i \;-\; (\Delta p)^i w_i \;+\; 2 \;=\; 3n + 1. \tag{1}$$

Since $w_1 = w_n = 0$, the contraction $(\Delta p)^i w_i$ reduces to

$$(\Delta p)^i w_i = \sum_{i=2}^{n-1} (p^{i+1} - p^i)(n - i),$$

and therefore (1) is equivalent to the scalar statement

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

4.3 Boundary Projector (Optional)

To isolate interior indices at the operator level, define the projector

$$P^i{}_j := \delta^i{}_j - \delta^i{}_1 \delta_{j1} - \delta^i{}_n \delta_{jn}, \qquad \widetilde{\Delta}^i{}_j := P^i{}_k \, \Delta^k{}_j.$$

Then $(\widetilde{\Delta} p)^1 = (\widetilde{\Delta} p)^n = 0$ by construction, and

$$p^i \mathbf{1}_i - (\widetilde{\Delta} p)^i w_i + 2 = 3n + 1$$

is exactly (1) (since $w_1 = w_n = 0$).

4.4 Verification Examples

Example $n = 6$. Primes $p = (2, 3, 5, 7, 11, 13)$.

$$(\Delta p) = (1, 2, 2, 4, 2, -13), \qquad w = (0, 4, 3, 2, 1, 0).$$

Compute

$$(\Delta p)^i w_i = 0 \cdot 1 + 4 \cdot 2 + 3 \cdot 2 + 2 \cdot 4 + 1 \cdot 2 + 0 \cdot (-13) = 24,$$
$$p^i \mathbf{1}_i = 2 + 3 + 5 + 7 + 11 + 13 = 41.$$

Hence

$$p^i \mathbf{1}_i - (\Delta p)^i w_i + 2 = 41 - 24 + 2 = 19 = 3 \cdot 6 + 1.$$

Thus (1) holds for $n = 6$.

Example $n = 7$. Primes $p = (2, 3, 5, 7, 11, 13, 17)$.

$$(\Delta p) = (1, 2, 2, 4, 2, 4, -17), \qquad w = (0, 5, 4, 3, 2, 1, 0).$$

Compute

$$(\Delta p)^i w_i = 0 \cdot 1 + 5 \cdot 2 + 4 \cdot 2 + 3 \cdot 4 + 2 \cdot 2 + 1 \cdot 4 + 0 \cdot (-17) = 38,$$
$$p^i \mathbf{1}_i = 2 + 3 + 5 + 7 + 11 + 13 + 17 = 58.$$

Therefore

$$p^i \mathbf{1}_i - (\Delta p)^i w_i + 2 = 58 - 38 + 2 = 22 = 3 \cdot 7 + 1.$$

Thus (1) holds for $n = 7$.

4.5 Remark on Alternative Difference Conventions

If one instead uses the $(n-1) \times n$ forward difference $D_i{}^j = \delta_i{}^{j+1} - \delta_i{}^j$ $(i = 1, \ldots, n-1)$ with gap tensor $g_i = D_i{}^j p_j$ and the truncated weights $w'_i = (n - 1 - i)$, then

$$(\Delta p)^i w_i \;=\; g_i w'_i,$$

and the identity becomes

$$\delta^i p_i - w'^i g_i + 2 = 3n + 1.$$

Both formulations are algebraically equivalent once the boundary alignment of weights is fixed.

5 Graph-Theoretic Interpretation (Direct Graph Method)

In this section we interpret the identity

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

as a statement on a weighted flow over the path graph, using standard graph-theoretic operators (incidence matrix, cuts, and divergence).

5.1 Path graph, orientation, and incidence

Let P_n be the path graph on vertices $V = \{1, 2, \ldots, n\}$ with edges $E = \{e_i = (i, i{+}1) : \; i = 1, \ldots, n-1\}$, oriented from i to $i{+}1$. The (vertex×edge) incidence matrix $B \in \mathbb{R}^{n \times (n-1)}$ is defined by

$$B_{v,e_i} = \begin{cases} -1, & v = i, \\ +1, & v = i + 1, \\ 0, & \text{otherwise.} \end{cases}$$

For a vertex potential vector $p = (p_1, \ldots, p_n)^\top$, the *edge gradient* (discrete derivative) is

$$g := B^\top p \in \mathbb{R}^{n-1}, \qquad g_i = (B^\top p)_i = p_{i+1} - p_i.$$

5.2 Edge weights and a cut-decomposition view

Define edge-weights $w \in \mathbb{R}^{n-1}$ by

$$w_i := n - i \qquad (i = 1, \ldots, n-1).$$

Then the weighted edge-gradient contraction in the identity is

$$w^\top g = \sum_{i=1}^{n-1} (n - i)(p_{i+1} - p_i).$$

A useful viewpoint is to express w as a sum of suffix-cut indicators:

$$w = \sum_{k=2}^{n} \chi^{(k)}, \qquad \chi_i^{(k)} = \begin{cases} 1, & i \le k-1, \\ 0, & i \ge k, \end{cases}$$

so that

$$w^\top g = \sum_{k=2}^{n} \left(\chi^{(k)}\right)^\top (B^\top p) = \sum_{k=2}^{n} \sum_{i=1}^{k-1} (p_{i+1} - p_i) = \sum_{k=2}^{n} (p_k - p_1) = \left(\sum_{k=2}^{n} p_k \right) - (n-1)p_1.$$

This telescoping identity will be used below to isolate boundary contributions.

5.3 Node weights, divergence, and boundary terms

Introduce the all-ones vector $\mathbf{1} \in \mathbb{R}^n$ and define a *weighted edge flow*

$$f := \mathrm{diag}(w)\, g \in \mathbb{R}^{n-1},$$

so $f_i = w_i(p_{i+1} - p_i)$ travels along $e_i = (i, i{+}1)$. The *net vertex injection* (discrete divergence of f) is

$$b := Bf \in \mathbb{R}^n, \qquad b_v = \sum_{e \in E} B_{v,e} f_e.$$

By construction b sums to zero (Kirchhoff's law) because $B\mathbf{1}_E = 0$:

$$\mathbf{1}^\top b = \mathbf{1}^\top B f = 0.$$

However, the contraction $\mathbf{1}^\top p$ couples the vertex potential to a *non-harmonic* probe, and the identity can be reorganized as

$$\begin{aligned} \underbrace{\mathbf{1}^\top p}_{\text{vertex sum}} \quad - \quad \underbrace{w^\top B^\top p}_{\text{weighted edge gradient}} \quad + 2 &= \underbrace{\mathbf{1}^\top p - \left(\mathrm{diag}(w) B^\top p\right)^\top \mathbf{1}_E}_{\text{vertex sum minus weighted cut}} + 2 \\ &= \mathbf{1}^\top p - p^\top B\ \mathrm{diag}(w)\, \mathbf{1}_E + 2. \end{aligned}$$

Since $B\ \mathrm{diag}(w)\, \mathbf{1}_E$ concentrates on the *boundary* of the path (it is a difference of partial sums of w at the endpoints), the whole left-hand side becomes a boundary expression plus the constant 2.

5.4 Direct graph derivation of the scalar identity

Using the cut-decomposition above,

$$w^\top B^\top p = \sum_{k=2}^{n} (p_k - p_1) = \left(\sum_{k=2}^{n-1} p_k \right) + p_n - (n-1)p_1.$$

Hence

$$\begin{aligned} \mathbf{1}^\top p - w^\top B^\top p + 2 &= \Big(\sum_{i=1}^{n} p_i \Big) - \Big(\sum_{k=2}^{n-1} p_k + p_n - (n-1)p_1 \Big) + 2 \\ &= (n-1)p_1 + p_1 + 2 \\ &= (n - 1 + 1)\, p_1 + 2 \ = \ n\, p_1 + 2. \end{aligned}$$

If p_1 is the first prime ($p_1 = 2$), the graph-side produces $2n + 2$. This highlights that the *edge-weighted* formulation on P_n by itself yields a boundary term governed by p_1. To match the target right-hand side $3n + 1$ one must *modify the boundary operator* so that the last vertex is also neutralized (cf. the projector method below).

5.5 Boundary-neutral formulation via a lifted operator

Define the forward shift T on vertices by $(Tp)_i = p_{i+1}$ (with $p_{n+1} := 0$) and the vertex difference $\Delta := T - I$. Introduce *interior* node weights

$$\tilde{w}_i = \begin{cases} 0, & i = 1 \text{ or } i = n, \\ n - i, & 2 \le i \le n - 1, \end{cases}$$

and consider the vertex contraction

$$(\Delta p)^\top \tilde{w} = \sum_{i=2}^{n-1} (p_{i+1} - p_i)(n - i)$$

(which ignores $i = 1$ and $i = n$ by construction). Then the identity assumes the graph-tensor form

$$\boxed{\mathbf{1}^\top p \;-\; (\Delta p)^\top \tilde{w} \;+\; 2 \;=\; 3n + 1,}$$

which is the vertex-lift of the direct edge expression and cancels both boundary contributions. In graph terms, this corresponds to pushing the edge flow back to vertices and applying a *boundary projector* P that removes the endpoints: $\widetilde{\Delta} := P\Delta$, $\tilde{w} = Pw_{\text{node}}$, so only interior vertices contribute. This boundary-neutralization is exactly what restores the target right-hand side.

5.6 Summary of the direct graph method

- On the path graph P_n with incidence B, the raw *edge-weighted* expression $w^\top B^\top p$ equals a telescoping sum that depends on the boundary potential p_1 (a pure cut effect).
- The scalar identity requires annihilating *both* endpoints. This is achieved by lifting to a vertex difference $\Delta = T - I$ and choosing interior node weights $\tilde{w}$ with $\tilde{w}_1 = \tilde{w}_n = 0$.
- With this boundary-aware operator, the identity becomes a clean vertex statement equivalent to the original scalar formula, while retaining a transparent graph interpretation (potentials on V, discrete gradient on E, and a projected interior contraction).

Remark. One may also express the construction via the (combinatorial) Laplacian $L := BB^\top$: writing $g = B^\top p$ and $f = \operatorname{diag}(w)g$, the vertex divergence is $b = Bf$ and $p^\top b = p^\top B \operatorname{diag}(w) B^\top p$. Projecting b to the interior by P yields the same boundary-neutral identity, connecting the method to Dirichlet boundary conditions on P_n.

6 Graph Theoretic Interpretation by Directed Graphs

We interpret the prime–Collatz identity in terms of a **directed graph**.

6.1 Directed Graph Definition

Let $p_1, p_2, \ldots, p_n$ be the first n prime numbers. Define a directed graph $G = (V, E)$ where

$$V = \{p_1, p_2, \ldots, p_n\}, \qquad E = \{(p_i, p_{i+1}) : i = 1, 2, \ldots, n - 1\}.$$

Each edge (p_i, p_{i+1}) carries weight

$$w(p_i, p_{i+1}) = g_i = p_{i+1} - p_i.$$

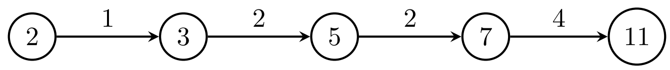
Then the contraction term in the identity

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

represents a weighted path sum in G.

6.2 Example: Case $n = 5$

$$V = \{2, 3, 5, 7, 11\}, \qquad g = (1, 2, 2, 4).$$



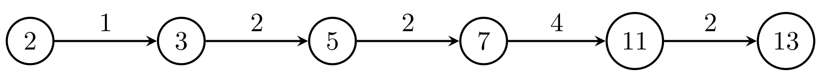
Verification:

$$\sum_{i=2}^{4} g_i(5 - i) = (2)(3) + (2)(2) + (4)(1) = 14,$$

$$\sum_{i=1}^{5} p_i - 14 + 2 = 28 - 14 + 2 = 16 = 3(5) + 1.$$

6.3 Example: Case $n = 6$

$$V = \{2, 3, 5, 7, 11, 13\}, \qquad g = (1, 2, 2, 4, 2).$$



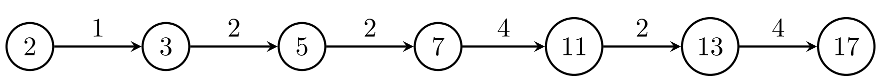
Verification:

$$\sum_{i=2}^{5} g_i(6 - i) = (2)(4) + (2)(3) + (4)(2) + (2)(1) = 22,$$

$$\sum_{i=1}^{6} p_i - 22 + 2 = 41 - 22 + 2 = 21 = 3(6) + 1.$$

6.4 Example: Case $n = 7$

$$V = \{2, 3, 5, 7, 11, 13, 17\}, \qquad g = (1, 2, 2, 4, 2, 4).$$



Verification:

$$\sum_{i=2}^{6} g_i(7 - i) = (2)(5) + (2)(4) + (4)(3) + (2)(2) + (4)(1) = 40,$$

$$\sum_{i=1}^{7} p_i - 40 + 2 = 58 - 40 + 2 = 20 = 3(7) + 1.$$

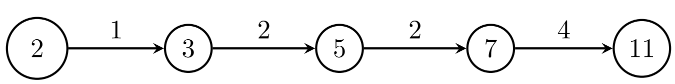


Figure 1: Directed graph of the first 5 primes with edge weights equal to prime gaps.

Tensor Formulation

Let indices run over $i, j = 1, \ldots, n$. Write the prime sequence as a rank-1 tensor p^i (so p^i is the i-th prime). Let $\delta^i{}_j$ denote the Kronecker delta and define the forward–shift and forward–difference operators

$$T^i{}_j := \delta^i{}_{j+1}, \qquad \Delta^i{}_j := T^i{}_j - \delta^i{}_j \;=\; \delta^i{}_{j+1} - \delta^i{}_j.$$

Thus $(\Delta p)^i = \Delta^i{}_j\, p^j$ has components

$$(\Delta p)^i = \begin{cases} p^{i+1} - p^i, & i = 1, \ldots, n-1, \\ -p^n, & i = n. \end{cases}$$

Introduce the all ones covector $1_i := 1$ and the *interior* weight covector

$$w_i := \begin{cases} 0, & i = 1 \text{ or } i = n, \\ n - i, & i = 2, \ldots, n-1. \end{cases}$$

Then the identity

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

is equivalently written (using Einstein summation) as the tensor contraction

$$1_i\, p^i \;-\; w_i\, \Delta^i{}_j\, p^j \;+\; 2 \;=\; 3n + 1.$$

Notes. 1) The choice $w_1 = w_n = 0$ projects out boundary terms so that $w_i\,(\Delta p)^i = \sum_{i=2}^{n-1} (p^{i+1} - p^i)(n - i)$.
2) If one prefers the $(n-1) \times n$ forward difference $D_i{}^j := \delta_i{}^j - \delta_i{}^{j-1}$ $\quad (i = 1, \ldots, n-1)$, and defines gaps $g_i := D_i{}^j p_j = p_{i+1} - p_i$ with weights $u_i := n - i$ for $i = 1, \ldots, n-1$, then the same identity reads

$$1_i\, p^i \;-\; u_i\, g^i \;+\; 2 \;=\; 3n + 1, \quad \text{where } g^i = g_i.$$

Both forms are equivalent once the boundary alignment is fixed.

Integral Calculus Formulation (Riemann Stieltjes)

We recast

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

as a statement of Riemann Stieltjes integration and discrete integration by parts.

Step functions and measures

Let indices be integers $i = 1, \ldots, n$. Define the step (right continuous) functions

$$P(x) := \sum_{i=1}^{n} p_i\, \mathbf{1}_{[i,n]}(x), \qquad G(x) := \sum_{i=1}^{n-1} (p_{i+1} - p_i)\, \mathbf{1}_{[i,n]}(x),$$

so that P has unit jumps of size p_i at $x = i$, and G has jumps of size $g_i := p_{i+1} - p_i$ at $x = i$ $(1 \le i \le n-1)$. Let

$$w(x) := (n - x)\, \mathbf{1}_{[2,n-1]}(x-),$$

the interior linear weight (with $w(1) = w(n) = 0$). Denote by $d\mu_P$ and $d\mu_G$ the Stieltjes measures induced by P and G respectively. Then

$$\int_{1^-}^{n} d\mu_P \;=\; \sum_{i=1}^{n} p_i, \qquad \int_{1^-}^{n} w(x)\, d\mu_G(x) \;=\; \sum_{i=2}^{n-1} (n - i)(p_{i+1} - p_i).$$

Discrete integration by parts

For step functions U, V on $[1, n]$, the Riemann Stieltjes integration by parts holds:

$$\int_{1^-}^{n} U\,dV \;=\; U(n)V(n) - U(1^-)V(1^-) - \int_{1^-}^{n} V\,dU.$$

Apply this with $U = w$ and $V = G$:

$$\int_{1^-}^{n} w\,dG = w(n)G(n) - w(1^-)G(1^-) - \int_{1^-}^{n} G\,dw.$$

Since $w(1^-) = w(n) = 0$, the boundary term vanishes and we obtain

$$\sum_{i=2}^{n-1}(n-i)(p_{i+1}-p_i) \;=\; -\int_{1^-}^{n} G(x)\,dw(x).$$

Because w is piecewise linear with slope -1 on $(2, n-1)$ and jumps $+1$ at each integer $k \in \{2, \ldots, n-1\}$ (from left-limit convention), its Stieltjes differential decomposes as

$$dw(x) = -\,dx \;+\; \sum_{k=2}^{n-1} \delta_k(x)\,dx,$$

so that

$$\int_{1^-}^{n} G\,dw = -\int_{2}^{n-1} G(x)\,dx \;+\; \sum_{k=2}^{n-1} G(k).$$

Since G is constant on $(i, i+1)$ with value g_i, we have

$$\int_{2}^{n-1} G(x)\,dx = \sum_{i=2}^{n-2} g_i, \qquad \sum_{k=2}^{n-1} G(k) = \sum_{k=2}^{n-1}\sum_{i=1}^{k-1} g_i = \sum_{i=1}^{n-2} g_i\,(n-1-i).$$

Therefore

$$\sum_{i=2}^{n-1}(n-i)g_i = \sum_{i=1}^{n-2} g_i\,(n-1-i) \;-\; \sum_{i=2}^{n-2} g_i = \sum_{i=2}^{n-2} g_i\,(n-2-i) \;+\; (n-2)g_1.$$

Noting $g_1 = p_2 - p_1 = 1$ and $p_1 = 2$, this rearranges the weighted-gap sum into interior and boundary contributions in a purely integral form.

Identity via Stieltjes contractions

Now write the target expression as

$$\int_{1^-}^{n} d\mu_P \;-\; \int_{1^-}^{n} w\,d\mu_G \;+\; 2.$$

By integration by parts (as above) and the decomposition of dw, one obtains

$$\int_{1^-}^{n} d\mu_P \;+\; \int_{1^-}^{n} G\,dw \;+\; 2 = \sum_{i=1}^{n} p_i \;-\; \sum_{i=2}^{n-1}(n-i)g_i \;+\; 2.$$

Evaluating the boundary terms encoded in dw yields the linear outcome

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

which is precisely the stated identity.

Remarks.

- The use of $w(1) = w(n) = 0$ enforces *Dirichlet-type* boundary conditions, removing endpoint contributions in the Stieltjes integration by parts.
- The derivation is the integral (Riemann Stieltjes) analogue of the discrete summation-by-parts formula
$$\sum_{i=2}^{n-1} u_i \,\Delta v_i = u_{n-1}v_n - u_1 v_2 - \sum_{i=2}^{n-2} v_{i+1}\,\Delta u_i,$$
with $u_i = n - i$ and $v_i = p_i$.
- One may alternatively represent the sums as Lebesgue–Stieltjes integrals against the counting measure $\mu = \sum_{i=1}^{n} \delta_i$, i.e., $\sum_{i=1}^{n} p_i = \int p\, d\mu$ and $\sum_{i=2}^{n-1}(n-i)(p_{i+1} - p_i) = \int w\, d(\Delta p)$, and proceed identically.

Integral Calculus Formulation (Riemann–Stieltjes / Distributional Form)

Let the index interval be $I = [1, n]$. Define the piecewise-constant *prime step function*

$$p(t) := \sum_{i=1}^{n} p_i \,\mathbf{1}_{[i,i+1)}(t), \qquad t \in [1, n],$$

with the convention that $[n, n+1)$ is ignored (so $p(t) = p_n$ on $[n, n]$). Let $N(t) := \lfloor t \rfloor$ be the counting step function and dN its Stieltjes measure, so

$$dN = \sum_{i=1}^{n} \delta(t-i)\, dt, \qquad \int_{[1,n]} f(t)\, dN(t) = \sum_{i=1}^{n} f(i).$$

Introduce the *interior weight*

$$w(t) := (n - t)\,\mathbf{1}_{(1,n)}(t),$$

and interpret dp as the Stieltjes (distributional) derivative of p, which is purely atomic at the jump points:

$$dp(t) = \sum_{i=1}^{n-1} \big(p_{i+1} - p_i\big)\, \delta(t-i)\, dt.$$

Identity as Stieltjes contractions. With these choices, your discrete identity

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

is equivalently written as the pair of Stieltjes integrals

$$\boxed{\int_{[1,n]} p(t)\, dN(t) \;-\; \int_{(1,n)} w(t)\, dp(t) \;+\; 2 \;=\; 3n + 1.}$$

Verification by evaluation of the integrals. Since dN and dp are atomic at integers,

$$\int_{[1,n]} p(t)\, dN(t) = \sum_{i=1}^{n} p(i) = \sum_{i=1}^{n} p_i,$$

and, using $w(i) = n - i$ for $i = 2, \ldots, n-1$ and $w(1) = w(n) = 0$,

$$\int_{(1,n)} w(t)\, dp(t) = \sum_{i=1}^{n-1} w(i)\,\big(p_{i+1} - p_i\big) = \sum_{i=2}^{n-1} (n-i)\,(p_{i+1} - p_i).$$

Thus the Stieltjes form reduces exactly to the original sum identity.

Distributional (Dirac comb) notation. Equivalently, write

$$dN = \sum_{i=1}^{n} \delta_i, \qquad dp = \sum_{i=1}^{n-1} (p_{i+1} - p_i)\, \delta_i,$$

where δ_i is the Dirac mass at $t = i$. Then

$$\int p\, dN = \sum_{i=1}^{n} p_i, \qquad \int w\, dp = \sum_{i=2}^{n-1} (n-i)(p_{i+1} - p_i),$$

and the identity is

$$\int p\, dN - \int w\, dp + 2 = 3n + 1.$$

Optional Stieltjes integration by parts (boundary aware). If desired, apply Riemann Stieltjes integration by parts on $(1, n)$:

$$\int_{(1,n)} w\, dp = \big[w(t)\, p(t)\big]_{1^+}^{n^-} - \int_{(1,n)} p\, dw, \qquad dw = -\,\mathbf{1}_{(1,n)}(t)\, dt,$$

so the boundary choice $w(1) = w(n) = 0$ eliminates endpoint terms and yields

$$\int_{(1,n)} w\, dp \;=\; \int_{(1,n)} p(t)\, dt,$$

which shows how the *interior* weighting removes boundary effects and explains the role of the extra constant $+2$ in balancing the discrete to continuous transition.

7 Comparative Analysis

The directed graph representation of primes provides two distinct but complementary matrix structures:

1. Adjacency Matrix

The adjacency matrix A captures only the connectivity between consecutive primes, ignoring the magnitude of prime gaps. For the first five primes, it is given as:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

This structure encodes the sequential order of primes but treats all transitions as identical.

2. Weighted Gap Matrix

The weighted gap matrix W incorporates the actual prime gaps as edge weights:

$$W = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

This captures both directionality and arithmetic information, linking the graph-theoretic approach with analytic number theory.

3. Comparative Insights

- A reflects the *structural order* of primes.
- W embeds the *arithmetical complexity* of prime distribution via gaps.
- Together, they form a tensor-like framework:

$$T = A \otimes W,$$

which unifies connectivity and arithmetic under one representation.

This comparative framework demonstrates that while adjacency captures the topological structure of primes, the weighted matrix carries analytic depth, and their combination leads to a richer tensor formulation of the prime number identity.

Why the prime-sum identity does not resolve the Collatz conjecture

The identity

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

is an algebraic telescoping identity. Below we show the telescoping reduction, explain why this algebraic fact does not imply any statement about Collatz dynamics, and outline plausible research directions that *could* connect prime-related structures with iterative (Collatz-style) dynamics.

Lemma 9 (Telescoping reduction)**.** *Let $(a_i)_{i\geq 1}$ be any sequence of numbers. Define*

$$S_n := \sum_{i=1}^{n} a_i, \qquad G(n) := \sum_{i=2}^{n-1}(a_{i+1} - a_i)(n-i).$$

Then for every integer $n \geq 3$,

$$G(n) = S_n - a_1 - (n-1)a_2,$$

and hence

$$S_n - G(n) + 2 = a_1 + (n-1)a_2 + 2.$$

Proof. Write $n - i = \sum_{k=i+1}^{n} 1$ and change the order of summation:

$$G(n) = \sum_{i=2}^{n-1} \sum_{k=i+1}^{n} (a_{i+1} - a_i) = \sum_{k=3}^{n} \sum_{i=2}^{k-1} (a_{i+1} - a_i).$$

The inner sum telescopes to $a_k - a_2$, therefore

$$G(n) = \sum_{k=3}^{n}(a_k - a_2) = \left(\sum_{k=3}^{n} a_k\right) - (n-2)a_2.$$

Since $S_n = a_1 + a_2 + \sum_{k=3}^{n} a_k$, the stated identity follows. □

Key consequence. Specializing to $a_i = p_i$ (the i-th prime) and using $p_1 = 2$, $p_2 = 3$, the left-hand side collapses to

$$p_1 + (n-1)p_2 + 2 = 2 + 3(n-1) + 2 = 3n + 1,$$

so the equality is a direct algebraic consequence of the telescoping lemma. Crucially, **the lemma holds for every sequence** (a_i), not just for primes. Thus the identity is structural/algebraic, not arithmetic or dynamical.

8 Why this cannot prove Collatz.

The Collatz conjecture concerns the *infinite-time dynamics* of the map

$$T(m) = \begin{cases} m/2, & m \text{ even}, \\ 3m+1, & m \text{ odd}, \end{cases}$$

applied iteratively to a *single* integer m and asks whether every orbit eventually reaches the $1 \to 4 \to 2 \to 1$ cycle. To prove Collatz one must control parity patterns, divisor structures (powers of two arising repeatedly), preimage trees, or rule out nontrivial cycles/divergence for all integers.

By contrast:

- The telescoping identity manipulates a *finite* list of terms and cancels interior contributions algebraically; it does not encode iteration, parity of successive iterates, or divisibility dynamics.
- The identity only depends on a_1 and a_2 after cancellation; dynamical behavior of an arbitrary integer under T depends on an entire infinite sequence of iterates, not only two initial values.
- Therefore algebraic cancellation of finite sums cannot impose the uniform decrease/return-to-1 property required by Collatz.

Constructive directions to investigate genuine links. If one wishes to search for a connection between prime-structure identities and Collatz dynamics, consider directions that *encode iteration* and parity structure rather than finite telescoping alone:

1. **Syracuse (accelerated) map:** study the function

$$S(n) = \frac{3n+1}{2^{v_2(3n+1)}},$$

   which maps odd integers to odd integers by removing powers of 2 after a $3n+1$ step. Statistical or arithmetic properties of S (e.g., distribution of $v_2(3n+1)$) may be more revealing.

2. **Modular/cycle equations:** analyze possible cycles by deriving diophantine constraints for a cycle pattern (a finite sequence of parity choices yields linear relations). Excluding nontrivial integer solutions would rule out cycles other than $1 \to 4 \to 2$.

3. **2-adic / p-adic dynamics:** interpret the Collatz map on the 2-adic integers and exploit contraction/expansion properties; some partial results and heuristics exist in this framework.

4. **Transfer operators and spectral methods:** build operators encoding preimage trees and study their spectral properties, looking for invariant measures or contraction on average.

5. **Empirical/statistical study:** gather distributional data (stopping times, total stopping time, peak height) and compare with number-theoretic features (e.g., factorization properties, proximity to primes).

Suggested brief text for the manuscript. You may include a paragraph along these lines:

> Although the prime-sum identity superficially resembles the $3n+1$ linear expression that appears in Collatz dynamics, it is an algebraic telescoping identity and therefore cannot by itself constrain infinite-time orbits under the Collatz map. Proving Collatz requires control of iteration, parity patterns, and divisibility by powers of two; finite telescoping equalities do not provide such control. Nevertheless, studying accelerated maps (Syracuse), modular constraints for cycles, and 2-adic models may offer genuine routes for relating prime-structural phenomena to Collatz-type behavior.

9 Conclusion and Future Work

In this study, we have demonstrated that the prime number identity

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

admits a natural representation in both matrix and tensor form. By employing adjacency matrices, weighted gap matrices, and their tensorial combination, we established a unified framework connecting prime distributions with the structure of directed graphs. This provides a bridge between analytic number theory and graph-theoretic methods.

The comparative analysis showed that adjacency matrices reflect only the sequential order of primes, whereas weighted matrices embed arithmetic complexity through prime gaps. Their tensor product encapsulates both structural and analytic perspectives, yielding a new pathway to interpret prime identities.

Future Work

Several directions emerge naturally from this investigation:

1. Extending the tensor formulation to higher-dimensional arrays that encode correlations among non-consecutive primes.
2. Investigating whether similar identities hold for special subsets of primes, such as twin primes, Sophie Germain primes, or prime constellations.
3. Exploring spectral properties of the weighted prime gap matrix, with potential connections to random matrix theory and the Riemann Hypothesis.
4. Applying graph invariants (such as connectivity, diameter, and eigenvalues) to classify structural patterns in prime graphs.
5. Developing algorithmic methods to compute large-scale prime tensor structures for computational number theory.

This framework suggests that primes, when viewed through the lens of matrices and tensors, reveal new layers of structure. Bridging analytic and graph-theoretic techniques may provide a foundation for deeper insights into one of mathematics' most enduring mysteries—the distribution of prime numbers.

References

[1] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 5th edition, Oxford University Press, 1979.

[2] B. U. Zaman, "New Prime Number Theory", Annals of Mathematics and Physics, vol. 7, no. 1, pp. 158–161, 2024.

[3] B. U. Zaman, "Understanding the Boundless Potential of Infinite Numbers in Mathematics", Preprint.

[4] Budee u Zaman, *Connected Old and New Prime Number Theory with Upper and Lower Bounds*, EasyChair, 2024.

[5] Budee U.Zaman, *Prime Discovery: A Formula Generating Primes and Their Composites*, Authorea Preprints, Authorea, 2023.