TWO PROOFS OF THE PRIME NUMBER THEOREM

DESMOND SAUNDERS

Abstract. The Prime Number Theorem (PNT) is perhaps the simplest and most elegant way to characterize the assymptotic behavior of prime numbers (in particular, their frequency among the natural numbers). In this paper we present two very different proofs of the PNT: one using complex analysis and the Riemann zeta function, and the other using only elementary methods.

Contents

1. Introduction

Theorem 1.1 (Prime Number Theorem)**.** *Let*

$$\pi(x) = \textit{the number of primes less than } x.$$

Then $\lim_{x\to\infty} \frac{\pi(x)}{x/\log x} = 1$*. Equivalently,* $\pi(x)$ *is* asymptotic *to* $\frac{x}{\log x}$*, or* $\pi(x) \sim \frac{x}{\log x}$*.*

The goal of this paper is simply to prove this theorem, and we will do so in two very different ways. The first proof, much like the original proof of the theorem as done independently by Hadamard and de la Vallée Poussin in 1896, relies on complex analysis.[1] The second proof is elementary (in that it only uses basic notions of real analysis, and could even be formulated without them) and is due to Selberg and Erdős in the late 1940's.[2] By juxtaposing these two proofs we can

Date: August 28, 2024.

[1]Our proof will more closely follow D.J. Newman's simplified version from 1980.

[2]Selberg and Erdős generated a few different proofs. Although it may not be the simplest, we will follow the very first one as published by Erdős in 1949.

understand why complex analysis is helpful in proving this non-complex result, and also why the complex proof was discovered over 50 years before the (very convoluted) elementary one.

One of the few things that these two proofs have in common is the focus on the Chebyshev function

$$\vartheta(x) = \sum_{p \leq x} \log p$$

where p is prime (throughout the paper p and q will always mean primes and n will always mean natural numbers greater than or equal to 1). This function is much nicer to study than $\pi(x)$, and so both proofs actually show that $\vartheta(x) \sim x$. This result is entirely equivalent to the PNT, but for our purposes we only need to show that one implies the other:

Theorem 1.2. *$\vartheta \sim x$ implies that $\pi(x) \sim \frac{x}{\log x}$.*

Proof. We would like to bound $\pi(x) \log x$ on both sides by $\vartheta(x)$. One side is easy,

$$\vartheta(x) = \sum_{p \leq x} \log p \leq \sum_{p \leq x} \log x = \pi(x) \log x.$$

For the other side, let $\epsilon > 0$. Then

$$\begin{aligned} \vartheta(x) \geq \sum_{x^{1-\epsilon} \leq p \leq x} \log p \geq \sum_{x^{1-\epsilon} \leq p \leq x} \log x^{1-\epsilon} &= (1-\epsilon) \log x \left[\pi(x) - \pi(x^{1-\epsilon})\right] \\ &\geq (1-\epsilon) \log x \left[\pi(x) - x^{1-\epsilon}\right]. \end{aligned}$$

Thus we have

$$\vartheta(x) \leq \pi(x) \log x \leq \frac{\vartheta(x)}{1-\epsilon} + x^{1-\epsilon} \log x.$$

We then divide both sides by x and take the limit as x goes to infinity, using the standard fact that $\lim\limits_{x \to \infty} \frac{\log x}{x^\epsilon} = 0$ for any $\epsilon > 0$,

$$1 = \lim_{x \to \infty} \frac{\vartheta(x)}{x} \leq \lim_{x \to \infty} \frac{\pi(x) \log x}{x} \leq \lim_{x \to \infty} \frac{1}{1-\epsilon} \frac{\vartheta(x)}{x} + \frac{\log x}{x^\epsilon} = \frac{1}{1-\epsilon}$$

and so indeed making ϵ arbitrarily small and using the given that $\vartheta \sim x$, it must be that $\lim\limits_{x \to \infty} \dfrac{\pi(x)}{x / \log x} = 1$. □

Before showing that $\vartheta(x) \sim x$, it is helpful to find a basic upper bound on the growth rate of $\vartheta(x)$. To make this notion of "growth rate" precise, we use the following two pieces of notation: let f and g be functions of a variable x with g everywhere positive. Then

$$f = O(g)$$

means that there exist some constants x_0 and $C > 0$ such that $|f(x)| \leq Cg(x)$ for all $x \geq x_0$. Similarly we say

$$f = o(g)$$

if, for all $C > 0$, there exists an x_C such that $|f(x)| \leq Cg(x)$ for all $x \geq x_C$.[3] Note that the equals sign as used in this notation does not denote any sort of equality of

[3] We can also modify both of these definitions to describe the behavior as f approaches a point other than infinity.

functions, and therefore one must be very careful to avoid abuses of this notation. This allows us to state the following theorem:

Theorem 1.3 (Chebyshev). $\vartheta(x) = O(x)$

Proof. Using standard facts about binomial coefficients we have

$$2^{2n} = \sum_{j=1}^{2n} \binom{2n}{j} \geq \binom{2n}{n} = \frac{(2n)(2n-1)\dots(n+1)}{n!}$$

and since we know $\binom{2n}{n}$ is an integer, the denominator must divide the numerator. However, any primes in the numerator (i.e primes between $n+1$ and $2n$) will not divide out and so remain in the numerator. Thus

$$2^{2n} \geq \frac{(2n)(2n-1)\dots(n+1)}{n!} \geq \prod_{n<p\leq 2n} p = e^{\vartheta(2n)-\vartheta(n)}$$

and so we have $\vartheta(2n) - \vartheta(n) \leq 2n \log 2$. We would like to show a similar result for $2n+1$. We have $\vartheta(2n+1) \leq \vartheta(2n) + \log(2n+1)$, and so

$$\vartheta(2n+1) - \vartheta\left(\frac{2n+1}{2}\right) = \vartheta(2n+1) - \vartheta(n) \leq 2n \log 2 + \log(2n+1).$$

Therefore, regardless of which case we are in there exist constants $C > 0$ and x_0 such that, for all $x \geq x_0$, we have

$$\vartheta(x) - \vartheta\left(\frac{x}{2}\right) \leq Cx$$

Now consider adding together the inequalities

$$\vartheta(x) - \vartheta\left(\frac{x}{2}\right) \leq Cx, \quad \vartheta\left(\frac{x}{2}\right) - \vartheta\left(\frac{x}{4}\right) \leq C\frac{x}{2}, \quad \vartheta\left(\frac{x}{4}\right) - \vartheta\left(\frac{x}{8}\right) \leq C\frac{x}{4}, \quad \dots$$

Once this goes below x_0 these inequalities no longer hold, but that just adds some bounded finite error for all x. Thus taking this sum we have

$$\vartheta(x) \leq C\left(x + \frac{x}{2} + \frac{x}{4} + \dots\right) + O(1) = 2Cx + O(1)$$

and so indeed $\vartheta(x) = O(x)$. □

2. Complex Analytic Proof

2.1. **Overview and Prerequisites.** This proof has three major players:

$$\vartheta(x) = \sum_{p\leq x} \log p$$

$$\Phi(s) = \sum_{p} \frac{\log p}{p^s}$$

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

where s is used to represent complex numbers (this is convention for the ζ function).

First we will give a very brief overview of the proof. It has four major steps:

(1) $\zeta(s)$ has an analytic continuation with no zeros on the line $\Re(s) = 1$.
(2) $\Phi(s)$ can be written in terms of $\zeta(s)$, such that the above features of the zeta function provide an analytic continuation for $\Phi(s)$.

(3) We can show that an integral converges if we can show that the Laplace transform of the function inside the integral has an analytic continuation to the line $\Re(z) = 0$.
(4) The statement $\vartheta(x) \sim x$ is equivalent to the convergence of the integral

$$\int_1^\infty \frac{\vartheta(x) - x}{x^2} dx$$

and indeed the analytic continuation for $\Phi(s)$ gives us an analytic continuation for the Laplace transform of this integrand.

The essential connection between the PNT and complex analysis comes in step (3), where the niceties of complex analysis provide a relatively simple way of proving the convergence of a real-valued integral.

All that is needed to follow this proof is a basic understanding of complex analysis (any reader with such an understanding can skip to the next section). Here, assuming knowledge of complex numbers, we will provide four basic results of complex analysis without proof to aid the unfamiliar reader.

Definition 2.1. A function $f : \mathbb{C} \to \mathbb{C}$ is called *analytic* on an open set U if one of two equivalent statements hold:
(i) For all $z_0 \in U$, f is differentiable at z_0.
(ii) For all $z_0 \in U$, there exists a power series

$$\sum_{n=0}^{\infty} a_n (z - z_0)^n$$

and a radius $r > 0$ such that for all z, if $|z - z_0| < r$, then the power series converges at z and equals $f(z)$.

Complex analysis frequently concerns itself with the study of these analytic functions. While these may seem like relatively weak conditions for real-valued functions, in the complex numbers analytic functions are highly constrained. A good example of this is the crucial phenomenon of analytic continuation:

Theorem 2.2. *Let U and V be connected open sets such that $U \cap V \neq \emptyset$, and let f be an analytic function on U and g be an analytic function on V. If $f(z) = g(z)$ for all $z \in U \cap V$, then g is the unique analytic function on V with this property. We call g the* analytic continuation of f to V.

Typically, when we show that such an analytic continuation exists, we will not even bother to give it a separate name on its new domain.

It may not be immediately obvious that functions defined as series or integrals are analytic. However, one can safely assume that any sufficiently well-behaved function (e.g. uniformly convergent) is analytic.

Finally, complex analysis is frequently concerned with integrals along paths. We will not provide the rigorous definition of such integrals, however we will use two key results of complex analysis, both due to Augustin-Louis Cauchy.

Theorem 2.3 (Cauchy's Theorem)**.** *Let γ and η be paths in an open, simply-connected set U having the same beginning point and ending point. Let f be analytic on U. Then*

$$\int_\gamma f = \int_\eta f.$$

In other words, all that matters are the starting and ending points of a path. Furthermore, the value of a function at any point can be determined by integrating along (essentially) any curve that surrounds that point:

Theorem 2.4 (Local Cauchy Formula)**.** *Let $\overline{D}$ be a closed disk of positive radius (or in general any closed region with sufficiently simple boundary) and let ∂D be the boundary of $\overline{D}$. Suppose f is analytic on $\overline{D}$. Then for all z_0 in the interior of $\overline{D}$ we have*

$$f(z_0) = \frac{1}{2\pi i}\int_{\partial D}\frac{f(z)}{z-z_0}dz.$$

Aside from these results, everything should be understandable to a reader with some understanding of real analysis and a willingness to accept that most results of real analysis extend perfectly to the complex plane.

2.2. **Function 1: $\zeta(s)$, the Riemann Zeta Function.** First we show that the zeta function is well defined on the complex plane with $\Re(s) > 1$.

Theorem 2.5. $\zeta(s) = \sum_{n=1}^{\infty}\frac{1}{n^s}$ *converges absolutely for $\Re(s) > 1$ and uniformly on $\Re(s) \geq 1+\delta$ for any $\delta > 0$.*

Proof. This is a simple consequence of the fact that the magnitude of a real number raised to a complex power depends only on the real part of the power,

$$\left|\frac{1}{n^{\sigma+it}}\right| = \left|e^{\log\frac{1}{n}(\sigma+it)}\right| = \left|e^{\log\frac{1}{n}\sigma}\right|\left|e^{i(t\log\frac{1}{n})}\right| = \frac{1}{n^\sigma}$$

Since $\sum_{n=1}^{\infty}\frac{1}{n^\sigma}$ converges for all $\sigma > 1$, we have that $\zeta(s)$ is absolutely convergent if $\Re(s) > 1$. Furthermore, for all s in the region $\Re(s) \geq 1+\delta$ for any $\delta > 0$ we have

$$\left|\frac{1}{n^s}\right| \leq \frac{1}{n^{1+\delta}}$$

and so this convergence is uniform. □

Now we can state a key result of Euler's that begins to show the connection between the zeta function and prime numbers.

Theorem 2.6 (The Euler product)**.** *The infinite product*

$$\prod_p \frac{1}{1-\frac{1}{p^s}}$$

converges absolutely when $\Re(s) > 1$, uniformly on the region $\Re(s) \geq 1+\delta$ for $\delta > 0$, and we have

$$\zeta(s) = \prod_p \frac{1}{1-\frac{1}{p^s}}.$$

Proof. Let us first take it as given that this product converges absolutely and uniformly on the desired regions. Then we can expand with the geometric series

$$\frac{1}{1-\frac{1}{p^s}} = 1 + \left(\frac{1}{p^s}\right) + \left(\frac{1}{p^s}\right)^2 + \cdots = 1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \ldots.$$

Furthermore, from the Fundamental Theorem of Arithmetic, every number $n \in \mathbb{N}$ can be uniquely factored into primes, and thus n^s can be uniquely factored into primes all raised to a multiple of s. Therefore when we further expand

$$\prod_p \left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \dots\right)$$

each number $\frac{1}{n^s}$ appears exactly once, and so indeed $\zeta(s) = \prod_p \frac{1}{1-\frac{1}{p^s}}$.

To show that this product does in fact converge, consider taking the product of the first N terms. Expanding with the geometric series as above, this gives us an infinite sum of the form $\sum \frac{1}{n^s}$ with the restriction that the prime factorization of n only contains the first N primes (for example, this includes all numbers less than the $(N+1)^{th}$ prime). By choosing larger and larger N, we can ensure that this sum is arbitrarily close to $\zeta(s)$, which we just showed converges in the desired manner. □

Corollary 2.7. *$\zeta(s)$ has no zeros on $\Re(s) > 1$.*

Proof. The Euler product has no zeros, since no term is zero and the terms approach 1, and so neither does $\zeta(s)$. □

While this is certainly a useful result, we would really like to show that $\zeta(s)$ has no zeros on the region $\Re(s) \geq 1$. To do so we must first show that the zeta function has an analytic continuation. In general it is possible to show that the zeta function has a continuation to the entire complex plane (except at the point $s = 1$), however this is unnecessary for our purposes. Thus we will only show that the zeta function has a continuation to $\Re(s) > 0$.

Theorem 2.8. *$\zeta(s)$ has the following continuation to $\Re(s) > 0$ which is analytic except for a pole at $s = 1$:*

$$\zeta(s) = \frac{s}{s-1} - s\int_1^\infty \frac{x - \lfloor x \rfloor}{x^{s+1}} dx$$

where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

Proof. For $\Re(s) > 1$ we have

$$\begin{aligned}
\frac{s}{s-1} - s\int_1^\infty \frac{x - \lfloor x \rfloor}{x^{s+1}} dx &= s\int_1^\infty \frac{1}{x^s} dx - s\int_1^\infty \frac{x - \lfloor x \rfloor}{x^{s+1}} dx \\
&= s\int_1^\infty \frac{\lfloor x \rfloor}{x^{s+1}} dx \\
&= \sum_{n=1}^\infty \int_n^{n+1} \frac{sn}{x^{s+1}} dx \\
&= \sum_{n=1}^\infty \frac{n}{n^s} - \sum_{n=1}^\infty \frac{n}{(n+1)^s} \\
&= \sum_{n=1}^\infty \frac{n}{n^s} - \sum_{n=1}^\infty \frac{n-1}{n^s} \\
&= \sum_{n=1}^\infty \frac{1}{n^s} = \zeta(s)
\end{aligned}$$

Other than the pole at $s = 1$, we have $x - \lfloor x \rfloor \leq 1$ and so

$$\int_1^\infty \frac{x - \lfloor x \rfloor}{x^{s+1}} dx \leq \int_1^\infty \frac{1}{x^{1+s}} dx$$

thus the integral is absolutely convergent. Therefore this expression is well-defined and indeed analytic on $\Re(s) > 0$. Thus we have found the desired analytic continuation for the zeta function. □

Corollary 2.9. *As $s \to 1$ we have $\zeta(s) = O(\frac{1}{s-1})$.*

Proof. Since we showed that this integral is analytic whenever $\Re(s) > 0$, we know we can find the following Taylor series:

$$\int_1^\infty \frac{x - \lfloor x \rfloor}{x^{s+1}} dx = a_0 + a_1(s-1) + a_2(s-1)^2 + \dots$$

for some constants $a_0, a_1, a_2, \dots$. Using our previous result we can thus find the Laurent series for $\zeta(s)$ around $s = 1$,

$$\zeta(s) = \frac{s}{s-1} + \int_1^\infty \frac{x - \lfloor x \rfloor}{x^{s+1}} dx = \frac{1}{s-1} + 1 + \left(a_0 + a_1(s-1) + a_2(s-1)^2 + \dots\right)$$

The only negative power term in this series is exactly $\frac{1}{s-1}$, and in particular we can conclude that $\zeta(s) = O(\frac{1}{s-1})$ as s approaches 1. □

Now that we have defined $\zeta(s)$ on the line $\Re(s) = 1$, it remains to show that there are no zeros on this line.

Theorem 2.10. *For $\Re(s) \geq 1$, we have $\zeta(s) \neq 0$.*

Proof. If $\Re(s) > 1$, then this has already been shown in Corollary 2.7. Now we use Theorem 2.6 and the typical Taylor series for $\log(1 - x)$ to expand

$$\begin{aligned}
\zeta(s) &= \prod_p \left(1 - \frac{1}{p^s}\right)^{-1} \\
&= \exp \log \prod_p \left(1 - \frac{1}{p^s}\right)^{-1} \\
&= \exp \sum_p -\log\left(1 - \frac{1}{p^s}\right) \\
&= \exp \sum_p \sum_{m=1}^\infty \frac{1}{mp^{ms}}
\end{aligned}$$

so long as $\Re(s) > 1$. Since the magnitude of such an expression only depends on the real part of the exponent, we have

$$|\zeta(s)| = \exp \sum_p \sum_{m=1}^\infty \Re\left(\frac{1}{mp^{ms}}\right) = \exp \sum_p \sum_{m=1}^\infty \frac{\cos(mt \log p)}{mp^{m\sigma}}$$

where $s = \sigma + it$. In particular we have

$$\zeta^3(\sigma)|\zeta(\sigma + it)|^4 |\zeta(\sigma + 2it)| = \exp \sum_p \sum_{m=1}^\infty \frac{3 + 4\cos(mt \log p) + \cos(2mt \log p)}{mp^{m\sigma}}.$$

We can then factor the numerator on the right, since

$$3 + 4\cos\theta + \cos 2\theta = 3 + 4\cos\theta + (2\cos^2\theta - 1) = 2(1+\cos\theta)^2 \geq 0.$$

Thus for $\sigma > 1$, $\zeta^3(\sigma)|\zeta(\sigma+it)|^4|\zeta(\sigma+2it)|$ is equal to e taken to the sum of infinitely many non-negative terms, so

$$\zeta^3(\sigma)|\zeta(\sigma+it)|^4|\zeta(\sigma+2it)| \geq e^0 = 1. \tag{2.11}$$

Now suppose for contradiction the zeta function had a zero on the line $\Re(s) = 1$, or in other words there exists some $t^* \neq 0$ such that $\zeta(1+it^*) = 0$.[4] Consider $\zeta(\sigma+it^*)$ as σ approaches 1 from the right. Since $\zeta(s)$ is analytic around $1+it^*$, we know it has a Taylor series with zero constant term, so as $\sigma \to 1$, $\zeta(\sigma+it^*) = O(\sigma-1)$.

Similarly, ζ is analytic around $1+2it$, possibly with non-zero constant term, so as $\sigma \to 1$, $\zeta(\sigma+2it^*) = O(1)$. Finally, by Corollary 2.9 we have $\zeta(\sigma) = O(\frac{1}{\sigma-1})$ as $\sigma \to 1$. Thus

$$\zeta^3(\sigma)|\zeta(\sigma+it)|^4|\zeta(\sigma+2it)| = O\left(\frac{1}{(\sigma-1)^3}(\sigma-1)^4\right) = O(\sigma-1).$$

For σ sufficiently close to 1 this expression must get arbitrarily close to zero, and in particular become smaller than 1. This contradicts (2.11), and so the proof is finished. □

2.3. Function 2: $\Phi(s)$.

Recall that $\Phi(s) = \sum_p \frac{\log p}{p^s}$. As with the zeta function, we first show that it converges on $\Re(s) > 1$.

Theorem 2.12. *$\Phi(s)$ converges absolutely on $\Re(s) > 1$ and uniformly on the region $\Re(s) \geq 1+\delta$ for $\delta > 0$.*

Proof. Let $\delta > 0$. By a standard fact of calculus, there exists n_0 such that for all $n \geq n_0$ we have $\log n \leq n^{\frac{\delta}{2}}$. Let $\Re(s) \geq 1+\delta$. For all but the finite number of terms less than n_0 we have

$$\sum_{p \geq n_0} \left|\frac{\log p}{p^s}\right| \leq \sum_{n=n_0}^{\infty} \left|\frac{\log n}{n^s}\right| \leq \sum_{n=n_0}^{\infty} \frac{n^{\frac{\delta}{2}}}{n^{1+\delta}} = \sum_{n=n_0}^{\infty} \frac{1}{n^{1+\frac{\delta}{2}}}$$

and so indeed the convergence is absolute and uniform as desired. □

Now we are ready to connect this function to the zeta function.

Theorem 2.13. *For $\Re(s) > 1$ we have*

$$\Phi(s) = -\frac{\zeta'(s)}{\zeta(s)} - h(s)$$

where $h(s)$ is a function that is analytic for $\Re(s) > \frac{1}{2}$.

Proof. We first take the logarithmic derivative of the zeta function. Since we showed in Theorem 2.6 that the Euler product is absolutely and uniformly convergent on

[4] If $t^* = 0$ then we would be assuming $\zeta(1) = 0$ which we know to be incorrect.

$\Re(s) > 1$, we can bring the derivative inside the sum as so:

$$\begin{aligned}\frac{\zeta'(s)}{\zeta(s)} = \frac{d}{ds}\log\zeta(s) &= \frac{d}{ds}\sum_p \log\left(1-\frac{1}{p^s}\right)^{-1} \\ &= \sum_p \frac{d}{ds}\log\left(1-\frac{1}{p^s}\right)^{-1} \\ &= -\sum_p \frac{\log p}{p^s-1}\end{aligned}$$

Furthermore, from the geometric series we have

$$\frac{1}{p^s-1} = \frac{1}{p^s}\frac{1}{1-\frac{1}{p^s}} = \frac{1}{p^s}\left(1+\frac{1}{p^s}+\frac{1}{p^{2s}}+\dots\right) = \frac{1}{p^s}+\frac{1}{p^{2s}}+\frac{1}{p^{3s}}+\dots$$

and so

$$\begin{aligned}-\frac{\zeta'(s)}{\zeta(s)} &= \sum_p \log p\left(\frac{1}{p^s}+\frac{1}{p^{2s}}+\dots\right) \\ &= \sum_p \frac{\log p}{p^s} + \sum_p \log p\left(\frac{1}{p^{2s}}+\frac{1}{p^{3s}}+\dots\right) \\ &= \Phi(s) + \sum_p h_p(s)\end{aligned}$$

where

$$\begin{aligned}|h_p(s)| &= \left|\log p\left(\frac{1}{p^{2s}}+\frac{1}{p^{3s}}+\dots\right)\right| \\ &= \left|\frac{\log p}{p^{2s}}\frac{1}{1-\frac{1}{p^s}}\right| \\ &\le C\frac{\log p}{|p^{2s}|}\end{aligned}$$

for some constant C that is the same for all primes p on the whole region $\Re(s) > \frac{1}{2}$. Thus on this region we have

$$\sum_p |h_p(s)| \le \sum_{n=1}^{\infty}|h_n(s)| \le C\sum_{n=1}^{\infty}\frac{\log n}{p^{2n}}$$

which converges absolutely and uniformly on the region $\Re(s) > \frac{1}{2}+\delta$ for $\delta > 0$ by the same argument as in Theorem 2.12. Thus, if we let $h(s) = \sum_p h_p(s)$ then h is analytic for $\Re(s) > \frac{1}{2}$ and indeed

$$\Phi(s) = -\frac{\zeta'(s)}{\zeta(s)} - h(s)$$

as desired. □

Theorem 2.14. *Let* $g(z) = \frac{\Phi(z+1)}{z+1} - \frac{1}{z}$. *Then* $g(z)$ *extends to a function that is analytic for* $\Re(z) \ge 0$.

Proof. Let $z = s-1$. We would equivalently like to show that

$$\frac{\Phi(s)}{s} - \frac{1}{s-1}$$

extends to an analytic function on $\Re(s) \geq 1$ (we already know this is analytic for $\Re(s) > 1$). From Theorem 2.13 we have

$$g(s) = -\frac{\zeta'(s)}{s\zeta(s)} - \frac{1}{s}h(s) - \frac{1}{s-1}$$

where $\frac{1}{s}h(s)$ is indeed analytic on the desired region. Thus we only need to show that $-\frac{\zeta'(s)}{s\zeta(s)} - \frac{1}{s-1}$ extends to an analytic function for $\Re(s) = 1$. To do so we use our continuation of the zeta function. By Theorem 2.10 we know $\zeta(s) \neq 0$ on this whole line, which shows that this expression must be analytic everywhere except $s = 1$. At $s = 1$ we can analyze the Laurent series of our functions. By Corollary 2.9 we know that the Laurent series of $\zeta(s)$ has a single term with negative power, $\frac{1}{s-1}$. Since $\frac{s}{s-1} = \frac{1}{s-1} + 1$, the function $s\zeta(s)$ shares this pole. Thus for some constants $a_0, a_1, a_2, \dots$ we have

$$s\zeta(s) = \frac{1}{s-1} + a_0 + a_1(s-1) + a_2(s-1)^2 + \dots.$$

We can similarly differentiate the Laurent series of $\zeta(s)$ to get

$$-\zeta'(s) = \frac{1}{(s-1)^2} - b_0 - b_1(s-1) - b_2(s-1)^2 - \dots.$$

For some constants $b_0, b_1, b_2, \dots$. Thus

$$\begin{aligned} \frac{-\zeta'(s)}{s\zeta(s)} - \frac{1}{s-1} &= \frac{\frac{1}{(s-1)^2} - b_0 - b_1(s-1) - b_2(s-1)^2 - \dots}{\frac{1}{s-1} + a_0 + a_1(s-1) + a_2(s-1)^2 + \dots} - \frac{1}{s-1} \\ &= \frac{(s-1)^2}{(s-1)^2} \frac{\left(\frac{1}{(s-1)^2} - b_0 - b_1(s-1) - \dots\right)}{\left(\frac{1}{s-1} + a_0 + a_1(s-1) + \dots\right)} - \frac{1}{s-1} \\ &= \frac{1 - b_0(s-1)^2 - b_1(s-1)^3 - \dots}{(s-1)(1 + a_0(s-1) + a_1(s-1)^2 + \dots)} - \frac{1}{s-1}. \end{aligned}$$

When we incorporate the $\frac{1}{s-1}$ term, the constant term in the numerator disappears,

$$\begin{aligned} \frac{-\zeta'(s)}{s\zeta(s)} - \frac{1}{s-1} &= \frac{-a_0(s-1) - (a_0+b_0)(s-1)^2 - \dots}{(s-1)(1 + a_0(s-1) + a_1(s-1)^2 + \dots)} \\ &= \frac{-a_0 - (a_0+b_0)(s-1) - \dots}{1 + a_0(s-1) + a_1(s-1)^2 + \dots} \end{aligned}$$

which must be analytic around 1 since the denominator is non-zero. Thus indeed every component of $\Phi(s)$ extends to an analytic function on $\Re(s) = 1$, and therefore so does $\Phi(s)$. □

Finally, it will help to have an integral representation of $\Phi(s)$. Recall that

$$\vartheta(x) = \sum_{p \leq x} \log p.$$

Theorem 2.15. *For $\Re(s) > 1$ we have*

$$\Phi(s) = s \int_1^\infty \frac{\vartheta(x)}{x^{s+1}} dx.$$

Proof. Note that between any two prime numbers ϑ is a constant function. Thus, if we define the sequence (p_n) where each p_n is the n^{th} prime (and for convenience we can set $p_0 = 1$), we have

$$\begin{aligned} s \int_1^\infty \frac{\vartheta(x)}{x^{s+1}} dx &= s \sum_{n=0}^\infty \int_{p_n}^{p_{n+1}} \frac{\vartheta(p_n)}{x^{s+1}} dx \\ &= s \sum_{n=0}^\infty \vartheta(p_n) \frac{1}{s} \left(\frac{1}{p_n^s} - \frac{1}{p_{n+1}^s} \right) \\ &= \sum_{n=0}^\infty \frac{\vartheta(p_n)}{p_n^s} - \frac{\vartheta(p_{n+1}) - \log p_{n+1}}{p_{n+1}^s} \\ &= \sum_{n=0}^\infty \left(\frac{\vartheta(p_n)}{p_n^s} - \frac{\vartheta(p_{n+1})}{p_{n+1}^s} \right) + \sum_p \frac{\log p}{p^s} \\ &= \Phi(s). \end{aligned}$$

In the second to last line, the sum telescopes, leaving only $\frac{\vartheta(1)}{1^s} = 0$. □

2.4. **Function 3: $\vartheta(x)$.** The key result about $\vartheta(x)$ is that the integral

$$\int_1^\infty \frac{\vartheta(x) - x}{x^2} dx$$

converges. However, we will prove this result somewhat indirectly using the Laplace transform.

Definition 2.16. Let $f : \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R}$ be bounded and piecewise continuous.[5] Then for $\Re(z) > 0$ we define the Laplace transform of f to be

$$g(z) = \int_0^\infty f(t) e^{-zt} \, dt.$$

Our central idea is that evaluating the integral

$$\int_0^\infty f(t) \, dt$$

is essentially equivalent to evaluating the Laplace transform at 0. While our restrictions on f are sufficient to show that the Laplace Transform is well defined (i.e. the integral converges) whenever $\Re(z) > 0$, we cannot naively assume convergence when $\Re(z) = 0$. This is where analytic continuation comes into play. Formally, the key step is encapsulated in this lemma:

Lemma 2.17. *Let f and g be as in Definition 2.16. If g extends to an analytic function on $\Re(z) \geq 0$, then $\int_0^\infty f(t) dt$ exists and equals $g(0)$.*

[5]These restrictions on f are actually stronger than necessary, but are sufficient for our purposes.

Proof. For $T > 0$, let

$$g_T(z) = \int_0^T f(t)e^{-zt}\, dt.$$

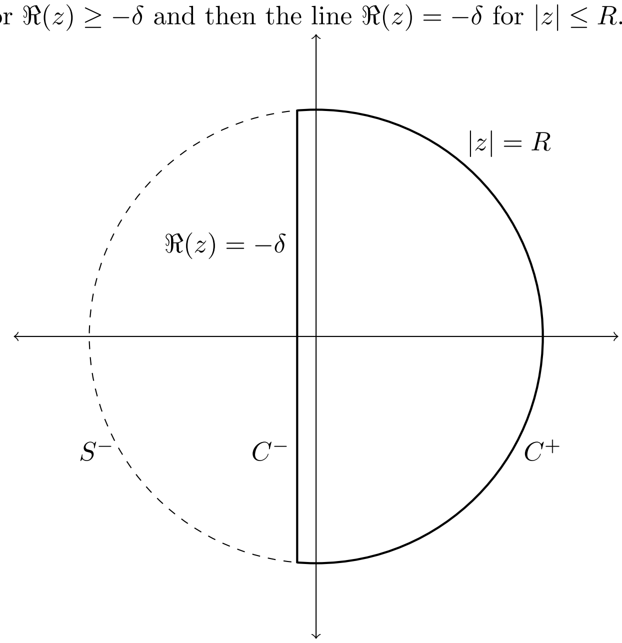
Since it is a finite integral of a well-behaved function, g_T is analytic everywhere. Furthermore we have that

$$\lim_{T\to\infty} g_T(0) = \lim_{T\to\infty} \int_0^T f(t)e^{-0t}\, dt = \int_0^\infty f(t)\, dt$$

so it is equivalent to our lemma to show that

$$\lim_{T\to\infty} g_T(0) = g(0).$$

This statement may seem obvious, but since we are dealing with analytic continuations we must tread carefully. Let C be the solid closed path in the diagram below, consisting of an arc that is just a tiny bit longer than a semi-circle and a vertical line. It is defined by two parameters, R and δ, and composed of the circle $|z| = R$ for $\Re(z) \geq -\delta$ and then the line $\Re(z) = -\delta$ for $|z| \leq R$.



We define C^+ to be the part of this path to the right of the imaginary axis, and C^- to be to the left (so C^- includes two tiny peices of the arc).

By our assumption that g extends to an analytic function on $\Re(z) \geq 0$, we actually know that it is analytic on some open set containing this region, and so for any R we can choose some δ small enough that g is analytic on all of C and its interior. Thus, since $g_T(z)$ is analytic everywhere, we have that

$$H_T(z) = (g(z) - g_T(z))\, e^{Tz} \left(1 + \frac{z^2}{R^2}\right)$$

is analytic on this whole region. Thus we can use Theorem 2.4 to get

$$g(0) - g_T(0) = H_T(0) = \frac{1}{2\pi i} \int_C \frac{H_T(z)}{z}\, dz.$$

We would like to show that this difference must vanish for large enough T. To do so, we bound the integral with strategic choice of R. Since we are given that f is bounded, there exists some $B \geq 0$ such that $|f(t)| \leq B$ for all $t > 0$.

Let us start with C^+. We would like to show that

$$\left| \frac{1}{2\pi i} \int_{C^+} \frac{H_T(z)}{z} \, dz \right| \leq \frac{B}{R}. \tag{2.18}$$

For $\Re(z) > 0$ we have

$$\begin{aligned} |g(z) - g_T(z)| = \left| \int_T^\infty f(t) e^{-zt} \, dt \right| \leq \int_T^\infty |f(t)| \left| e^{-zt} \right| \, dt \leq B \int_T^\infty \left| e^{-zt} \right| \, dt \\ = \frac{B}{\Re(z)} e^{-\Re(z)T} \end{aligned}$$

and

$$\begin{aligned} \left| e^{Tz} \left(1 + \frac{z^2}{R^2} \right) \frac{1}{z} \right| = e^{\Re(z)T} \frac{|R^2 + z^2|}{|R^2|} \frac{1}{R} = e^{\Re(z)T} \frac{|R^2 + z^2|}{|Rz\overline{z}/\overline{z}|} \frac{1}{R} = e^{\Re(z)T} \left| \frac{\overline{z}}{R} + \frac{z}{R} \right| \frac{1}{R} \\ = e^{\Re(z)T} \frac{2\Re(z)}{R^2}. \end{aligned}$$

Thus by taking the product of these two estimates, we get an upper bound for $\frac{H_T(z)}{z}$. The integral therefore must be less than the length of the path times this upper bound, or

$$\left| \frac{1}{2\pi i} \int_{C^+} \frac{H_T}{z} \, dz \right| \leq \frac{1}{2\pi} \left(\frac{B}{\Re(z)} e^{-\Re(z)T} \right) \left(e^{\Re(z)T} \frac{2\Re(z)}{R^2} \right) (\pi R) = \frac{B}{R}$$

as desired.

Now we would like to estimate the integral on C^-. We have

$$\begin{aligned} \left| \int_{C^-} \frac{H_T(z)}{z} \, dz \right| \leq \int_{C^-} \left| g(z) e^{Tz} \left(1 + \frac{z^2}{R^2} \right) \frac{1}{z} \right| dz \\ + \int_{C^-} \left| g_T(z) e^{Tz} \left(1 + \frac{z^2}{R^2} \right) \frac{1}{z} \right| dz. \end{aligned}$$

We again estimate each of these integrals separately. First, similar to the positive case, we have

$$\int_{C^-} \left| g_T(z) e^{Tz} \left(1 + \frac{z^2}{R^2} \right) \frac{1}{z} \right| dz \leq \frac{B}{R}. \tag{2.19}$$

Since g_T is analytic everywhere, this integrand is analytic everywhere on $\Re(z) < 0$ (in fact, everywhere except $z = 0$). Therefore, by Theorem 2.3, we can take this integral over any path in the half-plane $\Re(z) < 0$ as long as it has the same start and end points. In particular, let S^- be the part of the circle $|z| = R$ to the left of the imaginary axis (the dashed line in the diagram). This path satisfies our criteria, so

$$\int_{C^-} \left| g_T(z) e^{Tz} \left(1 + \frac{z^2}{R^2} \right) \frac{1}{z} \right| dz = \int_{S^-} \left| g_T(z) e^{Tz} \left(1 + \frac{z^2}{R^2} \right) \frac{1}{z} \right| dz.$$

and we can make essentially the same estimations as above

$$|g_T(z)| = \left| \int_0^T f(t) e^{-zt} \, dt \right| \leq B \int_0^T \left| e^{-zt} \right| \, dt = B \frac{e^{-\Re(z)T} - 1}{-\Re(z)} < B \frac{e^{-\Re(z)T}}{-\Re(z)}$$

and as before

$$\left| e^{Tz} \left(1 + \frac{z^2}{R^2}\right) \frac{1}{z} \right| = e^{\Re(z)T} \frac{2\Re(z)}{R^2}$$

so we have

$$\int_{S^-} \left| g_T(z) e^{Tz} \left(1 + \frac{z^2}{R^2}\right) \frac{1}{z} \right| \le \left(B \frac{e^{-\Re(z)T}}{-\Re(z)} \right) \left(e^{\Re(z)T} \frac{2\Re(z)}{R^2} \right) (\pi R) = \frac{B}{R}.$$

Finally we would like to show that

$$\lim_{T\to\infty} \int_{C^-} \left| g(z) e^{Tz} \left(1 + \frac{z^2}{R^2}\right) \frac{1}{z} \right| dz = 0. \tag{2.20}$$

We can write this integrand as e^{Tz} times some analytic function $h(t)$ that is independent of T. Since we are considering $h(t)$ on a compact set where it is analytic, it must have some upper bound M. Thus we have

$$\int_{C^-} \left| g(z) e^{Tz} \left(1 + \frac{z^2}{R^2}\right) \frac{1}{z} \right| dz \le M \int_{C^-} \left| e^{Tz} \right| dz$$

and we can certainly choose large enough T such that this integral is arbitrarily small (for $\Re(z) < 0$).

We can now prove the lemma. We would like to show that

$$\lim_{T\to\infty} |g(0) - g_T(0)| = 0.$$

Let $\epsilon > 0$ and choose R large enough such that $\frac{B}{R} < \frac{\epsilon}{3}$. We can also choose sufficiently large T such that the expression in (2.20) is also less than $\frac{\epsilon}{3}$. We can then add together our results from (2.18), (2.19), and (2.20) to get

$$|g(0) - g_T(0)| \le \frac{B}{R} + \frac{B}{R} + \frac{\epsilon}{3} < \epsilon$$

which proves the lemma. □

With this lemma we can show the convergence of the desired integral using our previous work on analytic continuations:

Theorem 2.21. *The integral*

$$\int_1^\infty \frac{\vartheta(x) - x}{x^2} dx$$

converges.

Proof. Let $f(t) = \dfrac{\vartheta(e^t) - e^t}{e^t}$. Then making the substitution $x = e^t$ we have

$$\int_1^\infty \frac{\vartheta(x) - x}{x^2} dx = \int_0^\infty \frac{\vartheta(e^t) - e^t}{e^{2t}} e^t dt = \int_0^\infty f(t)\, dt.$$

By Theorem 1.3 we know that f is bounded, and it is also piecewise continuous. Therefore if the Laplace transform of f has an analytic continuation we are done by Lemma 2.17. Specifically if we can show that

$$\int_0^\infty f(t) e^{-zt} dt = \frac{\Phi(z+1)}{z+1} - \frac{1}{z}$$

then applying Theorem 2.14 would complete the proof. By Theorem 2.15 we have

$$\frac{\Phi(s)}{s} - \frac{1}{s-1} = \int_1^\infty \frac{\vartheta(x)}{x^{s+1}}\,dx - \int_1^\infty \frac{1}{x^s}\,dx = \int_1^\infty \frac{\vartheta(x)-x}{x^{s+1}}\,dx$$

and so indeed letting $z = s-1$ we get

$$\frac{\Phi(z+1)}{z+1} - \frac{1}{z} = \int_1^\infty \frac{\vartheta(x)-x}{x^{z+2}}\,dx = \int_0^\infty \frac{\vartheta(e^t)-e^t}{e^{2t+z}}e^t\,dt = \int_0^\infty f(t)e^{-zt}\,dt$$

as desired. □

This last result was the key application of complex analysis, and we can immediately prove the final theorem:

Theorem 2.22. *$\vartheta(x) \sim x$.*

Proof. Recall that this statement really means that $\lim_{x\to\infty} \frac{\vartheta(x)}{x} = 1$. This is equivalent to the following:

(i) For any $\lambda > 1$, there exists x_0 such that for all $x \geq x_0$, $\vartheta(x) < \lambda x$, and
(ii) For any $0 < \lambda < 1$, there exists x_0 such that for all $x \geq x_0$, $\vartheta(x) > \lambda x$.

Let $\lambda > 1$ and suppose for contradiction that the set of x such that $\vartheta(x) \geq \lambda x$ is unbounded. Let x be some such point and consider the interval from x to λx. Since ϑ is non-decreasing, for all t in this interval we have $\vartheta(t) \geq \vartheta(x) \geq \lambda x$. Thus

$$\int_x^{\lambda x} \frac{\vartheta(t)-t}{t^2}\,dt \geq \int_x^{\lambda x} \frac{\lambda x - t}{t^2}\,dt = \left[-\frac{\lambda x}{t} - \log t\right]_x^{\lambda x} = (\lambda - 1) - \log\lambda$$

which is positive and independent of x. By assumption, we can find arbitrarily large x where this is the case. Thus for any possible value of the integral, there will always be an x such that the integral will leave any sufficiently small ϵ-neighborhood, and so the integral does not converge. However this contradicts Theorem 2.21.

Similarly, let $0 < \lambda < 1$ and suppose that the set of x such that $\vartheta(x) \leq \lambda x$ is unbounded. Then for any such x,

$$\int_{\lambda x}^{x} \frac{\vartheta(t)-t}{t^2}\,dt \leq \int_{\lambda x}^{x} \frac{\lambda x - t}{t^2}\,dt = \int_\lambda^1 \frac{\lambda - t}{t^2}\,dt < 0$$

which is negative and independent of x, and so again the integral does not converge, contradicting Theorem 2.21. □

Right at the beginning, in Theorem 1.2, we showed that $\vartheta(x) \sim x$ implies that $\pi(x) \sim \frac{x}{\log x}$. Thus Theorem 2.22 completes the proof of the PNT.

3. Elementary Proof

3.1. **Overview and Assumptions.** This proof involves three steps, which happen to correspond chronologically to the contributions of Selberg and Erdős. While this proof requires no advanced methods, we shall still assume some basic techniques of elementary number theory without proof. The most notable of these is the Möbius Inversion Formula. To understand this formula, we must first define the Möbius function:

Definition 3.1. We define $\mu : \mathbb{N} \longrightarrow \{-1, 0, 1\}$ by

$$\mu(n) = \begin{cases} 1 & n = 1 \\ 0 & n \text{ is not square-free} \\ (-1)^r & n = p_1 p_2 \dots p_r \text{ with each } p_j \text{ distinct} \end{cases}$$

where square-free means that no prime in the number's prime factorization has a power greater than 1. This is an essential function in elementary number theory and will be used throughout our proof, primarily due to its role in the following formula:

Theorem 3.2 (Möbius Inversion Formula)**.** *Let f and g be functions from the natural numbers to the complex numbers such that*

$$g(n) = \sum_{d|n} f(d).$$

Then

$$f(n) = \sum_{d|n} \mu(d) g\left(\frac{n}{d}\right).$$

We will also make use of this alternate version:

Theorem 3.3. *Let f and g be functions from the interval $[1, \infty)$ to the complex numbers such that*

$$g(x) = \sum_{n \le x} f\left(\frac{x}{n}\right)$$

then

$$f(x) = \sum_{n \le x} \mu(n) g\left(\frac{x}{n}\right).$$

This elementary proof does also use calculus, although it requires nothing beyond the high-school level other than a familiarity with lim sup, lim inf, and asymptotic notation.

In addition to this, we will use three basic estimates without proof. Let

$$a = \liminf_{x \to \infty} \frac{\vartheta(x)}{x} \quad \text{and} \quad A = \limsup_{x \to \infty} \frac{\vartheta(x)}{x}.$$

In Theorem 1.3 we showed that $\vartheta(x) = O(x)$, and thus there exists some constant C such that $\vartheta(x) \le Cx$ for large enough x. We will assume specifically that

$$A \le 1.5. \tag{3.4}$$

We take this as a reasonable assumption because one can actually show much stronger bounds with elementary methods. Indeed, before Hadamard and de la Vallée Poussin proved the PNT in 1896, Chebyshev had found bounds $c_1 < 1 < c_2$ very close to 1 such that

$$c_1 < \lim_{x \to \infty} \frac{\pi(x) \log x}{x} < c_2.$$

(recall that $\frac{\vartheta(x)}{x}$ and $\frac{\pi(x) \log x}{x}$ are closely related, see Theorem 1.2). This also motivates the next assumption:

$$a > 0 \tag{3.5}$$

which can also be shown with purely elementary methods. Finally we assume

$$\sum_{p \le x} \frac{\log p}{p} = (1 + o(1)) \log x. \tag{3.6}$$

Again, one can actually show a much stronger result with purely elementary methods, namely $\sum_{p \le x} \frac{\log p}{p} = \log x + o(1)$.

3.2. **The Selberg Symmetry Formula.** The process of finding the elementary proof of the PNT truly began when Selberg mentioned to Erdős that he had found an elementary proof of the following formula:

$$\sum_{p \le x} \log^2 p + \sum_{pq \le x} \log p \log q = 2x \log x + O(x). \tag{3.7}$$

Although not nearly as nice a statement as the PNT, this formula does speak to the asymptotic behavior of primes and so it becomes an essential tool that will be used repeatedly throughout this proof.

Definition 3.8. Define the 1^{st} *and* 2^{nd} *von Mangoldt functions* as

$$\Lambda(n) = \sum_{d|n} \mu(d) \log\left(\frac{n}{d}\right)$$
$$\Lambda_2(n) = \sum_{d|n} \mu(d) \log^2\left(\frac{n}{d}\right).$$

We will start by showing that the 2^{nd} von Mangoldt function is an approximate indicator for primes and the product of two primes:

Theorem 3.9. *For distict primes p and q we have*

$$\Lambda_2(1) = 0$$
$$\Lambda_2(p) = \log^2 p$$
$$\Lambda_2(p^2) = 3 \log^2 p$$
$$\Lambda_2(pq) = 2 \log p \log q$$

and for any n with three or more prime factors, $\Lambda_2(n) = 0$.

Proof. The first four statements follow from simple calculation.[6] Let $n = p_1^{e_1} \dots p_k^{e_k}$ with $k \ge 3$. To show that $\Lambda_2(n) = 0$, we need to show that

$$\sum_{d|n} \mu(d) \log^m d = 0$$

so long as $m < k$, which we do with induction. Our base case is that

$$\sum_{d|n} \mu(d) = 0$$

whenever n has one or more prime factors (i.e. $n \ge 1$). This is a standard fact about the Möbius function which we assume. Now suppose we are given that

$$\sum_{d|n} \mu(d) \log^m d = 0$$

[6] For example, $\Lambda_2(pq) = (\log pq)^2 - \log p - \log q = 2 \log p \log q$.

for some $m > 1$ so long as n has more than m prime factors. We have

$$\sum_{d|n} \mu(d) \log^{m+1} d = \sum_{d|n} (\mu(d) \log^m d) \left(\log p_{i_1}^{e_{i_1}} + \cdots + \log p_{i_l}^{e_{i_l}} \right)$$

where $p_{i_1}^{e_{i_1}} \dots p_{i_l}^{e_{i_l}}$ is the prime factorization of d. We can then break up this sum over the powers of prime factors of n to get

$$\sum_{p^r|n} \left(\sum_{\substack{d|\frac{n}{p^r} \\ p\nmid d}} \mu(dp^r) \log^m dp^r \right) \log p^r.$$

If $r \geq 2$ then $\mu(dp^r) = 0$. Otherwise $r = 1$ and since $p \nmid d$ we know $\mu(dp) = -\mu(d)$. Therefore

$$\begin{aligned} \sum_{d|\frac{n}{p}} \mu(dp) \log^m dp &= -\sum_{d|\frac{n}{p}} \mu(d) (\log d + \log p)^m \\ &= -\sum_{i=1}^{m} \binom{m}{i} \log^i p \sum_{d|\frac{n}{p}} \mu(d) \log^{m-i} d \\ &= 0 \end{aligned}$$

using strong induction. Note that certain elements of this last sum may have one fewer prime factor, so we can only apply our assumption if n has more than $m+1$ prime factors, which is as expected.

Thus we have

$$\begin{aligned} \Lambda_2(n) &= \sum_{d|n} \mu(d) \log^2 \left(\frac{n}{d}\right) \\ &= \sum_{d|n} \mu(d) (\log n - \log d)^2 \\ &= \log^2 n \sum_{d|n} \mu(d) - 2 \log n \sum_{d|n} \mu(d) \log d + \sum_{d|n} \mu(d) \log^2 d. \end{aligned}$$

By our above result each of these terms is zero, so indeed $\Lambda_2(n) = 0$. □

In order to show the Selberg Symetry Formula, (3.7), we will show separately that $\sum_{n \leq x} \Lambda_2(n)$ equals both sides of the equation. First, however, we need the following lemma:

Lemma 3.10. $\Lambda_2(n) = \Lambda(n) \log n + \sum_{d|n} \Lambda(d) \Lambda\left(\frac{n}{d}\right).$

Proof. If $n = p^k$, then the only divisors of n such that $\mu(d) \neq 0$ are $d = 1$ and $d = p$, so $\Lambda(n) = k \log p - (k-1) \log p = \log p$. If n has two or more prime factors then by our previous argument

$$\Lambda(n) = \sum_{d|n} \mu(d) \log\left(\frac{n}{d}\right) = \log n \sum_{d|n} \mu(d) + \sum_{d|n} \mu(d) \log d = 0.$$

Thus we have $\Lambda(n) = \sum_{p^r \leq x} \log p$. If we have $n = p_1^{e_1} \dots p_k^{e_k}$ then

$$\sum_{d|n} \Lambda(d) = \sum_{p^r|n} \log p = \sum_{i=1}^{k} e_i \log p_i = \log n.$$

We apply this equality in the first and fourth lines of the following manipulations:

$$\begin{aligned}
\log^2 n &= \sum_{d|n} \Lambda(d) \log n \\
&= \sum_{d|n} \Lambda(d) \left(\log d + \log \frac{n}{d}\right) \\
&= \sum_{d|n} \Lambda(d) \log \frac{n}{d} + \sum_{d|n} \Lambda(d) \log d \\
&= \sum_{d|n} \left(\Lambda(d) \sum_{a|\frac{n}{d}} \Lambda(a)\right) + \sum_{d|n} \Lambda(d) \log d \\
&= \sum_{ad|n} \Lambda(a)\Lambda(d) + \sum_{d|n} \Lambda(d) \log d \\
&= \sum_{b|n} \left(\Lambda(b) \log b + \sum_{d|b} \Lambda(d)\Lambda\left(\frac{b}{d}\right)\right).
\end{aligned}$$

Applying Theorem 3.2 to $\log^2 n = \sum_{b|n} \left(\Lambda(b) \log b + \sum_{d|b} \Lambda(d)\Lambda\left(\frac{b}{d}\right)\right)$ completes the proof. □

Now we are ready for the first half of the symmetry formula.

Proposition 3.11. $\sum_{n \leq x} \Lambda_2(x) = \sum_{p \leq x} \log^2 p + \sum_{pq \leq x} \log p \log q + O(x)$.

Proof. We break this proof into three steps. First, using Lemma 3.10 we have

$$\begin{aligned}
\sum_{n \leq x} \Lambda_2(n) &= \sum_{n \leq x} \left(\Lambda(n) \log n + \sum_{d|n} \Lambda(d)\Lambda\left(\frac{n}{d}\right)\right) \\
&= \sum_{n \leq x} \Lambda(n) \log n + \sum_{mn \leq x} \Lambda(m)\Lambda(n). \qquad (3.12)
\end{aligned}$$

We now analyze each of these sums indidually. We have

$$\sum_{n \leq x} \Lambda(n) \log n = \sum_{p^k \leq x} \log p \log p^k = \sum_{p \leq x} \log^2 p + \sum_{\substack{p^k \leq x \\ k \geq 2}} \log p \log p^k.$$

We would like to show that this last term is negligible, so we estimate

$$\sum_{\substack{p^k \leq x \\ k \geq 2}} \log p \log p^k \leq \sum_{\substack{p^k \leq x \\ k \geq 2}} \log^2 x \leq \sqrt{x} \log^2 x \implies \sum_{\substack{p^k \leq x \\ k \geq 2}} \log p \log p^k = O\left(\sqrt{x} \log^2(x)\right)$$

and so

$$\sum_{n\le x}\Lambda(n)\log n=\sum_{p\le x}\log^2 p+O\left(\sqrt{x}\log^2(x)\right). \tag{3.13}$$

Before we turn to the other sum, we need to show that $\sum_{n\le x}\Lambda(n)=O(x)$, which comes straight from Theorem 1.3 and the estimate we just did,

$$\sum_{n\le x}\Lambda(n)=\sum_{p\le x}\log p+\sum_{\substack{p^k\le x\\ k\ge 2}}\log p=O(x)+O\left(\sqrt{x}\log x\right)=O(x).$$

To complete the proof, we only need to show that

$$\sum_{mn\le x}\Lambda(m)\Lambda(n)=\sum_{pq\le x}\log p\log q+O(x). \tag{3.14}$$

Moving $\sum_{pq\le x}\log p\log q$ to the left side we get the equivalent form

$$\sum_{\substack{mn\le x\\ m,n\text{ not both prime}}}\Lambda(m)\Lambda(n)=O(x)$$

or by symmetry

$$\sum_{\substack{mn\le x\\ m\text{ not prime}}}\Lambda(m)\Lambda(n)=O(x).$$

Using the above estimate we have

$$\begin{aligned}\sum_{\substack{mn\le x\\ m\text{ not prime}}}\Lambda(m)\Lambda(n)&=\sum_{\substack{m\le x\\ m\text{ not prime}}}\Lambda(m)\sum_{n\le\frac{x}{m}}\Lambda(n)\\&=O\left(\sum_{\substack{m\le x\\ m\text{ not prime}}}\Lambda(m)\frac{x}{m}\right)\\&=O\left(x\cdot\sum_{\substack{p^k\le x\\ k\ge 2}}\frac{\log p}{p^k}\right)\end{aligned}$$

We would like to show that $\sum_{\substack{p^k\le x\\ k\ge 2}}\frac{\log p}{p^k}=O(1)$. It suffices to show it converges,

$$\begin{aligned}\sum_{\substack{p^k\le x\\ k\ge 2}}\frac{\log p}{p^k}&\le\sum_{p\le x}\log p\sum_{k=2}^{\infty}\frac{1}{p^k}\\&\le\sum_{n\le x}\log n\sum_{k=2}^{\infty}\frac{1}{n^k}\\&\le\sum_{n\le x}\log n\frac{1}{n(n-1)}\end{aligned}$$

which converges by the integral test. Thus we have shown (3.14), and combining (3.12), (3.13), and (3.14) compeletes the proof. □

All that remains is to show $\sum_{n\le x}\Lambda_2(n) = 2x\log x + O(x)$.

Lemma 3.15. *We have the following three estimates:*

$$\sum_{n\le x}\frac{\mu(n)}{n} = O(1) \tag{3.16}$$

$$\sum_{n\le x}\frac{\mu(n)}{n}\log\left(\frac{x}{n}\right) = O(1) \tag{3.17}$$

$$\sum_{n\le x}\frac{\mu(n)}{n}\log^2\left(\frac{x}{n}\right) = 2\log x + O(1) \tag{3.18}$$

Proof. By Theorem 3.3 we have

$$\begin{aligned} x = \sum_{n\le x} 1 \implies 1 &= \sum_{n\le x}\mu(n)\left(\frac{x}{n} + O(1)\right) \\ &\implies \sum_{n\le x}\frac{\mu(n)}{n} = \frac{1}{x} + O(1) \end{aligned}$$

which shows (3.16). Now we use an integral to estimate

$$\sum_{n\le x}\frac{1}{n} = \int_1^x \frac{1}{n}dn + \int_1^x\left(\frac{1}{\lfloor n\rfloor} - \frac{1}{n}\right)dn$$

where, since the function $\frac{1}{n}$ is decreasing,

$$\left|\int_1^x\left(\frac{1}{\lfloor n\rfloor} - \frac{1}{n}\right)dn\right| \le \sum_{n\le x}\frac{1}{n} - \frac{1}{n+1} = 1 - \frac{1}{\lfloor x\rfloor + 1}.$$

Thus for some constant C we have

$$\sum_{n\le x}\frac{1}{n} = \log x + C + O\left(\frac{1}{x}\right).$$

Multiplying by x, we get $\sum_{n\le x}\frac{x}{n} = x\log x + Cx + O(1)$, and applying Theorem 3.3

$$\begin{aligned} x &= \sum_{n\le x}\mu(n)\left(\frac{x}{n}\log\frac{x}{n} + C\frac{x}{n} + O(1)\right) \\ &= x\sum_{n\le x}\frac{\mu(n)}{n}\log\frac{x}{n} + Cx\sum_{n\le x}\frac{\mu(n)}{n} + O(x) \end{aligned}$$

applying (3.16) then gives

$$x\sum_{n\le x}\frac{\mu(n)}{n}\log\frac{x}{n} = O(x)$$

and dividing by x completes the proof of (3.17). Now we use the same strategy to prove (3.18) starting with the integral approximation

$$\begin{aligned}\sum_{n\le x}\frac{1}{n}\log\frac{x}{n} &= \int_1^x \frac{1}{n}\log\frac{x}{n}dn + \int_1^x\left(\frac{1}{\lfloor n\rfloor}\log\frac{x}{\lfloor n\rfloor}-\frac{1}{n}\log\frac{x}{n}\right)dn\\ &= \frac{1}{2}\log^2 x + \int_1^\infty\left(\frac{1}{\lfloor n\rfloor}-\frac{1}{n}\right)\log x\,dn + \int_1^x\left(\frac{1}{n}\log n - \frac{1}{\lfloor n\rfloor}\log\lfloor n\rfloor\right)dn\\ &= \frac{1}{2}\log^2 x + C_1\log x + O\left(\frac{\log x}{x}\right) + C_2 + O\left(\frac{\log x}{x}\right).\end{aligned}$$

For some constants C_1 and C_2. Again multiplying both sides by x and using Theorem 3.3 gives

$$x\log x = \frac{x}{2}\sum_{n\le x}\frac{\mu(n)}{n}\log^2\frac{x}{n} + C_1 x\sum_{n\le x}\frac{\mu(n)}{n}\log\frac{x}{n} - C_2 x\sum_{n\le x}\frac{\mu(n)}{n} + \sum_{n\le x}O\left(\log\frac{x}{n}\right).$$

We then apply (3.16), (3.17) and the following integral approximation

$$\sum_{n\le x}\log\frac{x}{n} = x + O(\log x)$$

to get

$$x\log x = \frac{x}{2}\sum_{n\le x}\frac{\mu(n)}{n}\log^2\frac{x}{n} + O(x) + O(x) + O(x) + O(\log x)$$

and dividing by $\frac{x}{2}$ completes the proof of (3.18). □

Finally we can complete the proof of the Selberg Symmetry formula:

Theorem 3.19. $\sum_{p\le x}\log^2 p + \sum_{pq\le x}\log p\log q = 2x\log x + O(x).$

Proof. First we need two more integral approximations

$$\sum_{n\le x}\log^2\frac{x}{n} = O(x) \quad\text{and}\quad \sum_{n\le x}\log^2 n = x\log^2 x - 2x\log x + 2x + O(\log^2 x)$$

of which we will derive the second,

$$\begin{aligned}\sum_{n\le x}\log^2 n &= \int_1^x \log^2 n\,dn + \int_1^x\left(\log^2\lfloor n\rfloor - \log^2 n\right)dn\\ &= x\log^2 x - 2x\log x + 2x + C + \int_1^x\left(\log^2\lfloor n\rfloor - \log^2 n\right)dn\end{aligned}$$

where, since the natural log is increasing,

$$\left|\int_1^x\left(\log^2\lfloor n\rfloor - \log^2 n\right)dn\right| \le \sum_{n\le x}\log^2(n+1) - \log^2(n) = O(\log^2 x).$$

Thus we have

$$\begin{aligned}\sum_{n\leq x}\Lambda_2(n) = \sum_{n\leq x}\sum_{d|n}\mu(d)\log^2\frac{n}{d} &= \sum_{ad\leq x}\mu(d)\log^2(a)\\ &= \sum_{d\leq x}\mu(d)\sum_{a\leq\frac{x}{d}}\log^2 a\\ &= x\sum_{d\leq x}\frac{\mu(d)}{d}\log^2\frac{x}{d} - 2x\sum_{d\leq x}\frac{\mu(d)}{d}\log\frac{x}{d}\\ &\quad + 2x\sum_{d\leq x}\frac{\mu(d)}{d} + O\left(\sum_{d\leq x}\log^2\frac{x}{d}\right)\end{aligned}$$

Applying all three estimates from Lemma 3.15 and the above mentioned approximation for $\sum_{d\leq x}\log^2\frac{x}{d}$, we get

$$\sum_{n\leq x}\Lambda_2(x) = 2x\log x + O(x) + O(x) + O(x) + O(x)$$

which is indeed the right hand side of our formula. Thus, combining with Proposition 3.11, the proof is complete. □

3.3. **Erdős's Contribution.** After Selberg showed him the symmetry formula, Erdős managed to prove the result of Theorem 3.21, after which it only took Selberg two days to prove the PNT. Before we present Erdős's result, we need the following lemma:

Lemma 3.20. *If $x < y$, then $\vartheta(y) - \vartheta(x) \leq 2(y-x) + o(y)$.*

Proof. Slightly weakening the result of Theorem 3.19, we have

$$\sum_{p\leq x}\log^2 p + \sum_{pq\leq x}\log p\log q = 2x\log x + o(x\log x).$$

Evaluating at y and x and subtracting yields

$$\sum_{x<p\leq y}\log^2 p \leq 2(y-x)\log y + o(y\log y).^7$$

Now we must break into cases. For the first case, suppose that $x \geq \frac{y}{\log^2 y}$. Thus

$$\log x \geq \log y - 2\log(\log y) \implies \log x = \log y + o(\log y)$$

and so for all $p \in (x,y]$ we have $\log p = (1+o(1))\log y$. Thus

$$\begin{aligned}\sum_{x<p\leq y}\log^2 p = (1+o(1))\log y\sum_{x<p\leq y}\log p &\leq 2(y-x)\log y + o(y\log y)\\ \implies (1+o(1))\sum_{x<p\leq y}\log p &\leq 2(y-x) + o(y)\\ \implies \sum_{x<p\leq y}\log p + o(y) &\leq 2(y-x) + o(y)\end{aligned}$$

[7] Note how $f(x) = g(x) + o(h(x))$ and $f(x) \leq g(x) + o(h(x))$ are different. For our purposes, the little-o can be interpreted as follows: there exists a function $h'(x)$ such that $h'(x) = o(h(x))$ and in the first case $f(x) = g(x) + h'(x)$ or in the second $f(x) \leq g(x) + h'(x)$.

where the last step is by Theorem 1.3. This completes case 1. Case 2, where $x < \frac{y}{\log^2 y}$, can be reduced to the first case as follows:

$$\begin{aligned}\vartheta(y) - \vartheta(x) &= \vartheta(y) - \vartheta\left(\frac{y}{\log^2 y}\right) + \vartheta\left(\frac{y}{\log^2 y}\right) - \vartheta(x)\\ &= \vartheta(y) - \vartheta\left(\frac{y}{\log^2 y}\right) + O\left(\frac{y}{\log^2 y}\right) + O\left(\frac{y}{\log^2 y}\right)\\ &< 2\left(y - \frac{y}{\log^2 y}\right) + o(y)\\ &< 2\left(y - x\right) + o(y)\end{aligned}$$

as desired. ☐

Theorem 3.21. *For every $c > 0$, there exists $\delta_c > 0$ and $x_0 > 0$ such that for all $x \geq x_0$ we have $\vartheta(x(1+c)) - \vartheta(x) > \delta_c x$.*

Proof. We will proceed by contradiction. Suppose there exists some c such that for all $\delta > 0$, we have $\vartheta(x(1+c)) - \vartheta(x) \leq \delta x$ for arbitrarily large values of x. Let X be an unbounded subset of $\mathbb{R}_{\geq 0}$ such that, for all $\delta > 0$, there exists an $x_0 \in X$ such that $\vartheta(x(1+c)) - \vartheta(x) \leq \delta x$ for all $x \in X$ where $x \geq x_0$.[8] In other words, if we restrict x' to run through X and not all of $\mathbb{R}$, we have $\vartheta(x'(1+c)) - \vartheta(x') = o(x')$.[9]

Let S be the set of all c where this is the case, and let $C = \sup S$. By (3.4) and (3.5) we have $0 < a \leq A < \infty$, so $C < \infty$ as well. The first step of our proof is to show that $C \in S$, i.e. $\vartheta(x(1+C)) - \vartheta(x) = o(x)$ for some values of x.

Let $\epsilon > 0$ and let $c \in S$ such that $c > C - \frac{\epsilon}{2}$. We have that $\vartheta(x'(1+c)) - \vartheta(x') = o(x')$. Thus, using Lemma 3.20,

$$\begin{aligned}\vartheta(x'(1+C)) - \vartheta(x') &= \vartheta(x'(1+C)) - \vartheta(x'(1+c)) + \vartheta(x'(1+c)) - \vartheta(x')\\ &\leq 2(C-c)x' + o(x')\\ &< \epsilon x' + o(x')\end{aligned}$$

and since ϵ was arbitrary, this is simply $o(x')$. Now we can move towards our contradiction. We would like to use C and its associated x' values to give a lower bound on A. This bound should then contradict (3.4). In fact, throughout the rest of this proof we will only be concerned with the $x \in X$, and therefore we will simply denote these values x rather than x'.

First we do some manipulations. Subtracting two versions of (3.7) yields

$$\sum_{x<p\leq x(1+C)} \log^2 p + \sum_{x<pq\leq x(1+C)} \log p \log q = 2Cx\log x + o(x\log x).$$

[8] For an explicit construction of X we can inductively define a sequence (x_n), where each x_n is chosen such that $\vartheta(x_n(1+c)) - \vartheta(x_n) \leq \frac{1}{n}x$ and $x_n > x_{n-1} + 1$.

[9] This will be a very useful tool for us: when we do not know about the asymptotic behavior of some function f in general but we do know about its behavior on some specific unbounded set X, then we can restrict x' to only run through X and so state $f(x') = o(g(x'))$. Indeed, in the future we will do so with much less explicit construction and notation. Note that we cannot a priori conclude anything about $f(ax')$ for $a \neq 1$.

Since $\log p \le \log(x(1+C))$ we have

$$\begin{aligned}\sum_{x<p\le x(1+C)} \log^2 p &\le \log(x(1+C)) \sum_{x<p\le x(1+C)} \log p \\ &= (\vartheta(x(1+C)) - \vartheta(x)) \log(x(1+C)) \\ &= o(x \log x).\end{aligned}$$

Subtracting these two results yields

$$\begin{aligned}2Cx\log x + o(x\log x) &= \sum_{x<pq\le x(1+C)} \log p \log q \\ &= \sum_{\le x(1+C)} \log p \sum_{\frac{x}{p}<q\le \frac{x}{p}(1+C)} \log q \\ &= \sum_{p\le x(1+C)} \left(\vartheta\left(\frac{x}{p}(1+C)\right) - \vartheta\left(\frac{x}{p}\right)\right) \log p. \end{aligned} \tag{3.22}$$

Now we need a lemma to understand what happens when our restricted x values are divided by primes.

Lemma 3.23. *Consider running through the values of x such that $\vartheta(x(1+C)) - \vartheta(x) = o(x)$. Then*

$$\vartheta\left(\frac{x}{p}(1+C)\right) - \vartheta\left(\frac{x}{p}\right) = 2C\frac{x}{p} + o\left(\frac{x}{p}\right) \tag{3.24}$$

for all primes p other than some set $\mathcal{P}$ which satisfies

$$\sum_{p\in\mathcal{P}_x} \frac{\log p}{p} = o(\log x)$$

where $\mathcal{P}_x = \{p \in \mathcal{P} \mid p \le x\}$.

Proof. We will once again use contradiction. For all $p \le x(1+C)$ that fail the first equation, by Lemma 3.20 it must be that

$$\vartheta\left(\frac{x}{p}(1+C)\right) - \vartheta\left(\frac{x}{p}\right) < 2C\frac{x}{p} + o\left(\frac{x}{p}\right)$$

or in other words there exists some $b_1 > 0$ such that

$$\vartheta\left(\frac{x}{p}(1+C)\right) - \vartheta\left(\frac{x}{p}\right) < (2C - b_1)\frac{x}{p} \tag{3.25}$$

for large x. Since $\mathcal{P}_x$ is finite, we can take the minimum such b_1 which works for all $p \in \mathcal{P}_x$. Suppose for contradiction that

$$\sum_{p\in\mathcal{P}_x} \frac{\log p}{p} \ne o(\log x).$$

However, from (3.6) we know that $\sum_{p\in\mathcal{P}_x} \frac{\log p}{p} = O(\log x)$. Thus there must exist some constant b_2 such that

$$\sum_{p\in\mathcal{P}_x} \frac{\log p}{p} \sim b_2 \log x.$$

Using (3.22) we have

$$\begin{aligned}
2Cx\log x + o(x\log x) &= \sum_{p\le x(1+C)} \left(\vartheta\left(\frac{x}{p}(1+C)\right) - \vartheta\left(\frac{x}{p}\right)\right)\log p \\
&= \sum_{p\in\mathcal{P}_x} \left(\vartheta\left(\frac{x}{p}(1+C)\right) - \vartheta\left(\frac{x}{p}\right)\right)\log p \\
&\quad + \sum_{\substack{p\notin\mathcal{P}_x \\ p\le x(1+C)}} \left(\vartheta\left(\frac{x}{p}(1+C)\right) - \vartheta\left(\frac{x}{p}\right)\right)\log p.
\end{aligned}$$

Applying (3.24) and (3.25) then gives

$$\begin{aligned}
2Cx\log x + o(x\log x) &< \sum_{p\in\mathcal{P}_x} (2C-b_1)\, x\frac{\log p}{p} + \sum_{\substack{p\notin\mathcal{P}_x \\ p\le x(1+C)}} 2Cx\frac{\log p}{p} + o\left(x\frac{\log p}{p}\right) \\
&\le (2C-b_1)xb_2\log x + 2Cx\left(\sum_{p\le x(1+C)} \frac{\log p}{p} - \sum_{p\in\mathcal{P}_x}\frac{\log p}{p}\right) + o(\dots) \\
&= b_2(2C-b_1)x\log x + 2Cx\left((1+o(1))\log x - b_2\log x\right) + o(\dots) \\
&= (2C-b_1b_2)x\log x + o(x\log x)
\end{aligned}$$

where the unwritten steps to evaluate the little-o were identical to the steps taken in the right-most sum. This is a contradiction, so the lemma is proven. $\square$

Now we return to the proof of Theorem 3.21. We will label all primes that satisfy (3.24) as good, and all others (primes in $\mathcal{P}$) as bad. Now for $\epsilon > 0$ suppose we have a list of good primes $p_1 < \cdots < p_k$ such that $(1+\epsilon)p_i < p_{i+1} < (1+C)(1+\epsilon)^2 p_i$ for all $1 \le i < k$, and the list is sufficiently long so that $p_k > 10p_1$. We will show that we can actually construct such a list at the end of this proof. First we would like to show that

$$\vartheta\left(\frac{x}{p_i}(1+C)\right) - \vartheta\left(\frac{x}{p_{i+1}}\right) > 1.9\left(\frac{x}{p_i}(1+C) - \frac{x}{p_{i+1}}\right) \tag{3.26}$$

for all $1 \le i < k$. We will need to break into two cases. Consider the intervals

$$I_i = \left[\frac{x}{p_i}, \frac{x}{p_i}(1+C)\right].$$

For our first case, suppose $I \cap I_{i+1} \ne \emptyset$. Thus, using the given restriction that $p_{i+1} > p_i(1+\epsilon)$, we have

$$\frac{x}{p_{i+1}}(1+\epsilon) < \frac{x}{p_i} < \frac{x}{p_{i+1}}(1+C).$$

Our claim is that, in this case,

$$\vartheta\left(\frac{x}{p_i}\right) - \vartheta\left(\frac{x}{p_{i+1}}\right) = 2\left(\frac{x}{p_i} - \frac{x}{p_{i+1}}\right) + o\left(\frac{x}{p_i}\right). \tag{3.27}$$

To show this, suppose for contradiction there exists a c_1 such that

$$\vartheta\left(\frac{x}{p_i}\right) - \vartheta\left(\frac{x}{p_{i+1}}\right) < (2-c_1)\left(\frac{x}{p_i} - \frac{x}{p_{i+1}}\right)$$

(where Lemma 3.20 shows that this is in fact the negation of our desired result). Since p_{i+1} is good we have

$$\vartheta\left(\frac{x}{p_{i+1}}(1+C)\right)-\vartheta\left(\frac{x}{p_{i+1}}\right)=2C\frac{x}{p_{i+1}}+o\left(\frac{x}{p_{i+1}}\right).$$

Taking the difference of these two equations we get

$$\begin{aligned}\vartheta\left(\frac{x}{p_{i+1}}(1+C)\right)-\vartheta\left(\frac{x}{p_i}\right)&>2C\frac{x}{p_{i+1}}+o\left(\frac{x}{p_{i+1}}\right)-(2-c_1)\left(\frac{x}{p_i}-\frac{x}{p_{i+1}}\right)\\&=2\left(\frac{x}{p_{i+1}}(1+C)-\frac{x}{p_i}\right)+c_1\left(\frac{x}{p_i}-\frac{x}{p_{i+1}}\right)+o\left(\frac{x}{p_{i+1}}\right)\\&>2\left(\frac{x}{p_{i+1}}(1+C)-\frac{x}{p_i}\right)+o\left(\frac{x}{p_{i+1}}\right)\end{aligned}$$

which contradicts Lemma 3.20, thus confirming (3.27). To show (3.26), all we need to do is add (3.27) and (3.24) evaluated at p_i (since p_i is good),

$$\begin{aligned}\vartheta\left(\frac{x}{p_i}(1+C)\right)-\vartheta\left(\frac{x}{p_{i+1}}\right)&=2\left(\frac{x}{p_i}(1+C)-\frac{x}{p_{i+1}}\right)+o\left(\frac{x}{p_i}\right)\\&>1.9\left(\frac{x}{p_i}(1+C)-\frac{x}{p_{i+1}}\right).\end{aligned}$$

Now, for our second case, suppose that $I_i\cap I_{i+1}=\emptyset$, and so

$$\frac{x}{p_{i+1}}(1+C)<\frac{x}{p_i}<\frac{x}{p_{i+1}}(1+C)(1+\epsilon)^2.$$

Similarly to our first case, it suffices to show that

$$\vartheta\left(\frac{x}{p_i}\right)-\vartheta\left(\frac{x}{p_{i+1}}\right)>1.9\left(\frac{x}{p_i}-\frac{x}{p_{i+1}}\right)$$

since then we can use the fact that p_i is good and add (3.24) to get (3.26). Using the above inequalities, the monotonicity of ϑ, and that p_{i+1} is good gives us

$$\begin{aligned}\vartheta\left(\frac{x}{p_i}\right)-\vartheta\left(\frac{x}{p_{i+1}}\right)&>\vartheta\left(\frac{x}{p_{i+1}}(1+C)\right)-\vartheta\left(\frac{x}{p_{i+1}}\right)\\&=2\left(\frac{x}{p_{i+1}}(1+C)-\frac{x}{p_{i+1}}\right)+o\left(\frac{x}{p_{i+1}}\right)\\&>2\left(\frac{x}{p_i(1+\epsilon)^2}-\frac{x}{p_{i+1}}\right)+o\left(\frac{x}{p_{i+1}}\right)\\&>\frac{2}{(1+\epsilon)^2}\left(\frac{x}{p_i}-\frac{x}{p_{i+1}}\right)+o\left(\frac{x}{p_{i+1}}\right)\\&>1.9\left(\frac{x}{p_i}-\frac{x}{p_{i+1}}\right)\end{aligned}$$

for sufficiently small ϵ. Thus we have established (3.26) in both cases. Adding these inequalities together for all $1\leq i<k$ we get

$$\begin{aligned}&\vartheta\left(\frac{x}{p_1}(1+C)\right)-\vartheta\left(\frac{x}{p_k}\right)>1.9\left(\frac{x}{p_1}(1+C)-\frac{x}{p_k}\right)\\&\implies\vartheta\left(\frac{x}{p_1}(1+C)\right)>1.9\left(\frac{x}{p_1}(1+C)-\frac{x}{p_k}\right).\end{aligned}$$

Since $p_k > 10p_1$ we have $-1.9\frac{x}{p_k} > -0.19\frac{x}{p_1} > -0.19\frac{x}{p_1}(1+C)$ and so

$$\vartheta\left((1+C)\frac{x}{p_1}\right) > 1.6\frac{x}{p_1}(1+C).$$

Recall that, although it is restricted, x still becomes arbitrarily large, and thus so does $(1+C)\frac{x}{p_1}$. Therefore, this sequence of values provides a lower bound on A. In particular, $A = \limsup_{x\to\infty} \frac{\vartheta(x)}{x} \geq 1.6$ and this contradicts (3.4).

All that remains is to show that we can construct the necessary sequence of primes. For this purpose, let $\epsilon > 0$ and let B be some sufficiently large number.[10] For each natural number r up to $\lfloor \frac{\log x}{2\log B} \rfloor - 1$, let $I_r = \left(B^{2r}, B^{2r+1}\right)$. By construction we have that $I_r \subset (0, x)$ for all r.

Our first step is to show that for all but $o(\log x)$ many values of r, I_r contains a good prime. We defined C as $\sup S$, thus so long as $B > 1 + C$ we know $B - 1 \notin S$ and so in particular there must exist a constant $\delta_B > 0$ such that

$$\vartheta(Bx) - \vartheta(x) \geq \delta_B Bx.$$

Thus if we let $x_r = B^{2r}$ we have

$$\sum_{p\in I_r} \frac{\log p}{p} = \sum_{x_r < p \leq Bx_r} \frac{\log p}{p} \geq \frac{1}{Bx_r} \sum_{x_r < p \leq Bx_r} \log p \geq \frac{1}{Bx_r}\delta_B Bx_r = \delta_B.$$

Let $R(x)$ be the number r values such that I_r does not contain a good prime. By Lemma 3.23, we have

$$o(\log x) = \sum_{\substack{p\leq x \\ p \text{ is bad}}} \frac{\log p}{p} \geq \sum_{\substack{p\in I_r \\ I_r \text{ has no good } p}} \frac{\log p}{p} \geq R(x)\delta_B \implies R(x) = o(\log x)$$

as desired.

For each r such that I_r does contain a good prime, let $p_1(r)$ be the smallest good prime in I_r. We then iteratively choose good primes $p_2(r), \ldots, p_j(r)$ such that $(1+\epsilon)p_i(r) < p_{i+1}(r) < (1+C)(1+\epsilon)^2 p_i(r)$ for $1 \leq i < j$. If any of these lists are long enough such that $p_j(r) > 10p_1(r)$ then we are done. Thus suppose for contradiction that $p_j(r) \leq 10p_1(r)$ for all r and none of these lists can be extended to any more good primes. Let $J_r(j) = [p_j(r)(1+\epsilon), p_j(r)(1+C)(1+\epsilon)^2]$. Then by this assumption, $J_r(j)$ only contains bad primes for all r. As before, since $(1+C)(1+\epsilon) > 1 + C$, there exists c_1 such that

$$\vartheta((1+\epsilon)(1+C)x) - \vartheta(x) > c_1 x$$

and so

$$\sum_{p\in J_r(j)} \frac{\log p}{p} > c_1.$$

Let $S = \{J_r(j) \mid I_r \text{ contains a good prime}\}$. Since we showed earlier that $R(x) = o(\log x)$, for sufficiently large x we know $R(x) < \frac{\log x}{4\log B}$. Thus

$$|S| = \left\lfloor \frac{\log x}{2\log B} \right\rfloor - 1 - R(x) > \frac{\log x}{4\log B}.$$

[10] An explicit lower bound could be calculated, but is not necessary.

If each $J_r(j)$ is disjoint, then we have

$$\sum_{\substack{p\le x \\ p \text{ is bad}}} \frac{\log p}{p} \ge \sum_{\substack{p\in J_r(j) \\ \text{for some } r}} \frac{\log p}{p} > c_1 \frac{\log x}{4\log B}$$

which contradicts Lemma 3.23. Thus, all that is needed to complete the proof is to show that we can make each of these intervals disjoint. Let $B > 10(1+C)(1+\epsilon)^2$. We assumed that $p_j(r) \le 10p_1(r)$ for all r, thus we have

$$\begin{aligned}(1+C)(1+\epsilon)^2 p_j(r) &\le 10(1+C)(1+\epsilon)^2 p_1(r) \\ &\le 10(1+C)(1+\epsilon)^2 B^{2r+1} \\ &< B^{2r+2} \\ &\le p_1(r+1) \\ &< (1+\epsilon)p_{j'}(r+1)\end{aligned}$$

where j' is the j value associated with $r+1$. Therefore the upper bound of $J_r(j)$ is less than the lower bound of $J_{r+1}(j')$, and the proof is complete. □

3.4. **Proof of the Prime Number Theorem.** As in the complex proof, our goal is to show that $\vartheta(x) \sim x$ so that Theorem 1.2 completes the proof of the PNT. Recall that

$$A = \limsup_{x\to\infty} \frac{\vartheta(x)}{x} \quad \text{and} \quad a = \liminf_{x\to\infty} \frac{\vartheta(x)}{x}.$$

Thus if we can show $A = a = 1$, then the fact that $\lim_{x\to\infty} \frac{\vartheta(x)}{x} = 1$ follows.

We will do so in two steps: first we show that $A + a = 2$ and then $A = a$. Before that, however, we need the following lemma:

Lemma 3.28. $\displaystyle\sum_{p\le x} \log^2 p = \vartheta(x)\log x + O(x)$.

Proof. Let

$$\psi(x) = \sum_{n\le x} \Lambda(n) = \sum_{n\le x} \sum_{p^m \le n} \log p.$$

We know that for all $p \in (x^{1/2}, x)$, $\log p$ will appear in our sum exactly once, while all $p \in (x^{1/3}, x^{1/2})$, $\log p$ will appear in our sum twice, and so on. Thus

$$\psi(x) = \vartheta(x) + \vartheta\left(x^{1/2}\right) + \vartheta\left(x^{1/3}\right) + \cdots = \sum_{2^m < x} \vartheta\left(x^{1/m}\right).$$

For all $x \ge 2$ we have $\vartheta(x) \le x\log x$, so for $m \ge 2$ we have

$$\vartheta\left(x^{1/m}\right) < x^{1/m}\log x \le x^{1/2}\log x \implies \vartheta\left(x^{1/m}\right) = O\left(x^{1/2}\log x\right).$$

Thus, since there are approximately $\log x$ values of m such that $2^m < x$,

$$\psi(x) = \vartheta(x) + \sum_{\substack{m\ge 2 \\ 2^m < x}} \vartheta\left(x^{1/m}\right) = \vartheta(x) + O\left(\log x\left(x^{1/2}\log x\right)\right). \tag{3.29}$$

By (3.13) we have

$$\sum_{p\le x} \log^2 p = \sum_{n\le x} \Lambda(n)\log n + O\left(\sqrt{x}\log x\right)$$

so we manipulate this sum, letting $N = \lfloor x \rfloor$,

$$\begin{aligned}\sum_{n \leq x} \Lambda(n) \log n &= \psi(1)\log(1) + (\psi(2) - \psi(1))\log 2 + \cdots + (\psi(N-1) - \psi(N))\log N \\ &= \psi(1)(\log 1 - \log 2) + \cdots + \psi(N-1)(\log(N-1) - \log N) + \psi(N)\log N.\end{aligned}$$

For all $1 \leq n \leq N-1$, we know that ψ is constant on $(n, n+1)$. Thus each of our terms can be written

$$\psi(n)(\log n - \log(n+1)) = -\int_n^{n+1} \frac{\psi(t)}{t} dt$$

and so we have

$$\sum_{n \leq x} \Lambda(n) \log n = \psi(x) \log x - \int_1^x \frac{\psi(t)}{t} dt.$$

By (3.29), we have $\psi(t) = O(t)$, and so this integral is $O(x)$. Thus we have

$$\begin{aligned}\sum_{p \leq x} \log^2 p &= \sum_{n \leq x} \Lambda(n) \log n + O\left(\sqrt{x} \log x\right) \\ &= \psi(x) \log x + O(x) + O\left(\sqrt{x} \log x\right) \\ &= \vartheta(x) \log x + O\left(\sqrt{x} \log^3 x\right) + O(x) \\ &= \vartheta(x) \log x + O(x)\end{aligned}$$

as desired. ☐

Theorem 3.30. $A + a = 2$.

Proof. By the definition of A, we can restrict x to only run through values where $\vartheta(x)$ is "large" compared to x, and so $\vartheta(x) = Ax + o(x)$. Thus, by Lemma 3.28

$$\sum_{p \leq x} \log^2 p = \vartheta(x) \log x + o(x \log x) = Ax \log x + o(x \log x).$$

Subtracting this from (3.7) gives

$$\sum_{p \leq x} \vartheta\left(\frac{x}{p}\right) \log p = \sum_{pq \leq x} \log p \log q = (2 - A) x \log x + o(x \log x). \tag{3.31}$$

Given that we are only dealing with values of x such that $\vartheta(x) = Ax + o(x)$, for large enough x we know that $\vartheta(x) \geq ax$. Thus, by (3.6) we have

$$\sum_{p \leq x} \vartheta\left(\frac{x}{p}\right) \log p \geq \sum_{p \leq x} a \frac{x}{p} \log p = ax \log x + o(x \log x).$$

Combining these two equations we have

$$\begin{aligned}ax \log x + o(x \log x) &\leq (2 - A) x \log x + o(x \log x) \\ &\implies 2 - A - a + o(1) \geq 0\end{aligned}$$

and so $2 - A - a \geq 0$. If we then repeat this argument with values of x such that $\vartheta(x) = ax + o(x)$ and so $\vartheta(x) \leq Ax$ for large x, we get $2 - A - a \leq 0$. Thus we conclude $A + a = 2$. ☐

Before showing that $a = A$, we need some lemmas. From here through the rest of the proof we are going to assume that x is restricted to values such that $\vartheta(x) = Ax + o(x)$.

Lemma 3.32. *For our x such that $\vartheta(x) = Ax + o(x)$, we also have*

$$\vartheta\left(\frac{x}{p}\right) = a\frac{x}{p} + o\left(\frac{x}{p}\right) \tag{3.33}$$

for all primes outside a set $\mathcal{P}$ which satisfies

$$\sum_{p\in\mathcal{P}_x} \frac{\log p}{p} = o(\log x)$$

where $\mathcal{P}_x = \{p \in \mathcal{P} \mid p \leq x\}$.

Proof. As in Lemma 3.23, suppose for contradiction that there exist constants b_1 and b_2 such that for all $p \in \mathcal{P}$

$$\vartheta\left(\frac{x}{p}\right) > (a + b_1)\frac{x}{p} \quad \text{and} \quad \sum_{p\in\mathcal{P}_x} \frac{\log p}{p} \sim b_2 \log x.$$

From (3.31) and Theorem 3.30 we have

$$\sum_{p\leq x} \vartheta\left(\frac{x}{p}\right) \log p = (2 - A)x \log x + o(x \log x) = ax \log x + o(x \log x)$$

and so

$$\begin{aligned} ax \log x + o(x \log x) &= \sum_{p\leq x} \vartheta\left(\frac{x}{p}\right) \log p \\ &= \sum_{p\in\mathcal{P}_x} \vartheta\left(\frac{x}{p}\right) \log p + \sum_{\substack{p\leq x \\ p\notin\mathcal{P}_x}} \vartheta\left(\frac{x}{p}\right) \log p \\ &> \sum_{p\in\mathcal{P}_x} (a + b_1)\frac{x}{p} \log p + \sum_{\substack{p\leq x \\ p\notin\mathcal{P}_x}} a\frac{x}{p} \log p \\ &> b_2(a + b_1)x \log x + (1 - b_2)ax \log x + o(x \log x) \\ &= ax \log x + b_1 b_2 x \log x + o(x \log x) \end{aligned}$$

which is a contradiction. □

Lemma 3.34. *Let p_1 be the smallest prime such that*

$$\vartheta\left(\frac{x}{p_1}\right) = a\frac{x}{p_1} + o\left(\frac{x}{p_1}\right)$$

as in Lemma 3.32. We have

$$\vartheta\left(\frac{x}{p_1 p}\right) = a\frac{x}{p_1 p} + o\left(\frac{x}{p_1 p}\right) \tag{3.35}$$

for all primes p outside a set $\mathcal{P}$ which satisfies

$$\sum_{p\in\mathcal{P}_x} \frac{\log p}{p} = o(\log x)$$

where $\mathcal{P}_x = \{p \in \mathcal{P} \mid p \leq x\}$.

Proof. The proof is identical to the proof of Lemma 3.32, simply replacing x with $\frac{x}{p_1}$ and switching A and a. □

We are now ready to prove the Prime Number Theorem:

Theorem 3.36. $\vartheta(x) \sim x$.

Proof. Let p_1 as in Lemma 3.34. Let p_i denote a prime satisfying (3.33), and p_j one satisfying (3.35). For any $0 < c < \frac{A}{a} - 1$,[11] let

$$I_i = \left[\frac{p_i}{p_1}, (1+c)\,\frac{p_i}{p_1}\right].$$

We would like to show that there is some choice of p_i and p_j such that $p_j \in I_i$. Suppose for contradiction that $p_j \notin I_i$ for all p_i and p_j. Thus all primes in I_i fail to satisfy (3.35). Now we come to the key application of Theorem 3.21:

$$\begin{aligned}\sum_{p_i \in I_i} \frac{\log p}{p} &\geq \frac{p_1}{p_i(1+c)} \sum_{\frac{p_i}{p_1} < p < \frac{p_i}{p_1}(1+c)} \log p \\ &> \frac{p_1}{p_i(1+c)} \cdot \delta_c \frac{p_i}{p_1}\end{aligned}$$

which is independent of p_i. Thus we can find a constant η_1 such that $\sum_{p_i \in I_i} \frac{\log p}{p} \geq \eta_1$ for all sufficiently large p_i. As in the proof of Theorem 3.21, for large x we can construct $\eta_2 \log x$ many disjoint intervals I_i all less than x. Thus we have

$$\sum_{\substack{p \leq x \\ p \text{ does not satisfy (3.35)}}} \frac{\log p}{p} \geq \sum_{p \in \bigcup I_i} \frac{\log p}{p} > \eta_1 \eta_2 \log x$$

and this contradicts Lemma 3.34. Thus there must exist some p_i and p_j such that

$$\frac{p_i}{p_1} < p_j < \frac{p_i}{p_1}(1+c)$$

and so by rearranging, we get

$$\frac{x}{p_1 p_j} < \frac{x}{p_i} < \frac{x}{p_1 p_j}(1+c).$$

For any $\delta > 0$, we use the definitions of p_i and p_j to get

$$(A-\delta)\frac{x}{p_1 p_j} < \vartheta\left(\frac{x}{p_1 p_j}\right) < \vartheta\left(\frac{x}{p_i}\right) < (a+\delta)\frac{x}{p_i} \leq (a+\delta)(1+c)\frac{x}{p_1 p_j}$$

and in particular $A - \delta < (a+\delta)(1+c)$. Making both δ and c arbitrarily small yields $A \leq a$. Since we know by definition that $A \geq a$, we have $A = a$. This result, together with Theorem 3.30, shows that $A = a = 1$ and proves the theorem. □

Acknowledgements

It is a pleasure to thank my mentor, Micah Gay, for her help and guidance throughout. I would also like to thank J. Peter May for organizing the REU and Daniil Rudenko for the apprentice lecture series, and Nick Fiori for originally introducing me to this topic. Finally I would like to thank my friends and family for their constant support.

[11] We implicitly break into cases: one where $A = a$, in which case the proof is done, and one in which $A > a$, and so this range is well-defined and we can proceed.

References

[1] Sergei Lang. Complex Analysis. Springer. 1998.

[2] D. J. Newman. Simple Analytic Proof of the Prime Number Theorem. The American Mathematical Monthly, Vol. 87, No. 9 (Nov., 1980), pp. 693-696. https://www.jstor.org/stable/2321853.

[3] Anup Rao. A Proof of the Prime Number Theorem. https://www.youtube.com/watch?v=4N-tMaqjEy4.

[4] Guangqiu Liang. Prime Number Theorem. https://www.math.emory.edu/~gliang7/PNT.pdf.

[5] Theodore W. Gamelin. Complex Analysis. Springer. 2001.

[6] Ashvin A. Swaminathan. On the Selberg-Erdős Proof of the Prime Number Theorem.

[7] Paul Erdős. On a New Method In Elementary Number Theory Which Leads To an Elementary Proof of the Prime Number Theorem. Proc Natl Acad Sci U S A. 1949 Jul;35(7):374-84. doi: 10.1073/pnas.35.7.374. PMID: 16588909; PMCID: PMC1063042. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1063042/.

[8] Atle Selberg. An Elementary Proof of the Prime-Number Theorem. Annals of Mathematics, Second Series, Vol. 50, No. 2 (Apr., 1949), pp. 305-313. https://www.jstor.org/stable/1969455.

[9] G.H. Hardy and E.M. Wright. An Introduction to the Theory of Numbers. Fourth Edition. Oxford University Press. 1960. https://blngcc.wordpress.com/wp-content/uploads/2008/11/hardy-wright-theory_of_numbers.pdf.