

An Elementary Proof of the Prime Number Theorem

Nishant Patil

Received January 23, 2024

Accepted June 03, 2024

Electronic access June 30, 2024

This paper explores the depth of the Prime Number Theorem and proves the theorem using an array of arithmetical functions such as the Möbius function, Von Mangoldt function, Chebyshev Functions and Dirichlet Multiplication. This account attempts to simplify and smoothen rather temperamental (or fluctuating) functions, by replacing them with their integrals or using big-oh approximation. Inspired by previous works in the field, this is an adaptation of various elementary accounts that culminates in a proof accessible to even high school students. The proof walks readers through each of the arithmetical functions and their natures and gradually begins proving crucial theorems and lemmas, namely Selberg's Identity, Mertens Theorem as well as several inequalities that describe some of these functions even further, until these can be used to describe the nature of prime number distribution itself. In its essence, the paper captures intriguing relationships between functions that count divisors, prime factors and their logarithms from their behaviours and limits in order to display the growth behaviour of prime numbers on the number line.

Introduction

In the vast scope of number theory, prime numbers have long been the focus of mathematical intrigue. Their unpredictable distribution along the number line has challenged mathematicians for centuries. As mathematicians dealt with the puzzling patterns of the prime numbers as we move along the number line, a rich tapestry of mathematical inquisition unfolded, setting the stage for the Prime Number Theorem. From the early findings of reverent mathematicians Euler and Legendre to the significant contributions of Riemann and Dirichlet in making their own functions towards proving this theorem, the narrative unfolds with focus on the collaborative efforts that brought the PNT to fruition. The Prime Number Theorem states that the number of prime numbers not exceeding x is asymptotic to $x/\log x$, as $x \rightarrow \infty$.

Several articles and papers, notably Newman's Short Proof¹, seek to demonstrate this theorem through the use of the Riemann's infinite series, $\zeta(x)$, and finding its real roots for $\text{Re}(X) = 1$. Its rigorous use of complex analysis and logarithmic integrals demands a strong background. While this adaptation of findings from Norman Levinson² and other elementary proofs appears long and technical, its thorough and exhaustive explanation of every function and notation used makes this an extremely comprehensive account of the proof.

Defining Functions

Prime Number Counter

The concept of prime numbers, such as 2, 3, 5, 7, 11, 13, dates back to ancient times, and Euclid provided a proof demonstrating their infinitude. The count of prime numbers less than or equal to a given value x is denoted as $\pi(x)$. This function can be represented by:

$$\pi(x) := \sum_{p \leq x} 1 \quad (1)$$

where p runs over all prime numbers less than x .

Theorem 2.1: The Fundamental Theorem of Arithmetic states that every $n \in \mathbf{N}$ can be uniquely expressed in the form $p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}$, where p_1, p_2, \cdots, p_k are distinct primes and m_1, m_2, \cdots, m_k are powers of primes.

Möbius Function

The Möbius function, denoted as $\mu(n)$, has an alternating behavior, taking the value of 1 if n is a square-free positive integer with an even number of distinct prime factors, -1 if it has an odd number of distinct prime factors, and 0 if n has a squared prime factor. In essence:

$$\mu(n) := \begin{cases} 1 & \text{if } n = 1 \\ (-1)^k & \text{if } n = p_1 p_2 \cdots p_k \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

An important theorem relating to the Möbius function involves its divisor sum, the partial sum of the function over

the divisors of n . Divisor sums will be expressed as a sum with over $d|n$ throughout the paper.

Theorem 2.2:

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Von Mangoldt Function

The von Mangoldt function, denoted as $\Lambda(n)$, is defined such that $\Lambda(n)$ equals the natural logarithm of a prime power when n is a power of a prime and is zero otherwise. This can be shown as:

$$\Lambda(n) := \begin{cases} \log(n) & \text{if } n = p^k, \text{ where } k \in \mathbf{N} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Theorem 2.3: An important formula related to the partial sum of the function states that:

$$\sum_{d|n} \Lambda(d) = \log(n) \quad (5)$$

Due to the Fundamental Theorem of Arithmetic, that states that every $n \in \mathbf{N}$ can be expressed in the form $p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}$, the partial sum of the von Mangoldt function can be expressed as (5).

Dirichlet Multiplication

The Dirichlet product is a mathematical operation defined for two arithmetic functions, returning another arithmetic function, typically denoted as $f * g(n)$, where n is a positive integer³. It is computed by:

$$f * g(n) := \sum_{d|n} f(d)g\left(\frac{n}{d}\right) \quad (6)$$

Theorem 2.4: Some important identities that will be used throughout this paper in relation to Dirichlet multiplication, or Dirichlet convolution are:

1. For an identity function, $I(n)$, $n \in \mathbf{Z}$, where

$$I(n) := \left\lfloor \frac{1}{n} \right\rfloor = \begin{cases} 1 & n = 1 \\ 0 & n > 1 \end{cases}$$

it can be stated that:

$$f * I(n) = \sum_{d|n} f(d)I(n/d) = f(n) \left\lfloor \frac{n}{n} \right\rfloor = f(n) \quad (7)$$

2. Dirichlet multiplication is commutative as well as associative:

$$f * g = g * f \quad (\text{commutative law}) \quad (8)$$

$$f * g * h = (f * g) * h = f * (g * h) \quad (\text{associative law}) \quad (9)$$

3. For every arithmetical function, $f(n)$, for which $f(1) \neq 0$, there exists a function $f^{-1}(n)$, the Dirichlet inverse of f , where

$$f * f^{-1} = I \quad (10)$$

Theorem 2.5: Möbius Inversion If f and g are arithmetical functions, such that

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

Then

$$\sum_{d|n} f(d)\mu(n/d) = g(n) \quad (11)$$

By (6), we have that for a unit function, $u(n) = 1$ for all $n \in \mathbf{Z}$,

$$g * u = f$$

Taking the Dirichlet product of both sides with μ , we obtain

$$\begin{aligned} \mu * g * u &= f * \mu \\ &= (\mu * u) * g = f * \mu \end{aligned}$$

The convolution of μ and u can be written as:

$$\sum_{d|n} \mu(d)u\left(\frac{n}{d}\right) = \sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{otherwise} \end{cases} = I(n)$$

Now substituting this result back into the original $(\mu * u)$,

$$\begin{aligned} (\mu * u) * g &= f * \mu \\ &= I * g = f * \mu \\ &= g = f * \mu \end{aligned}$$

which proves the inversion formula.

Asymptotic Relationships and Elementary Results

Big-Oh Notation and Asymptotic Equality

Definition 3.1: For an arithmetical function, $g(x)$, if there exists a constant, M , for which another function, $f(x)$ can be expressed as

$$|f(x)| \leq M(g(x))$$

for $x \geq a$, any constant, then the relationship between f and g is said to be:

$$f(x) = O(g(x)) \quad (12)$$

This essentially states that the function $g(x)$ grows faster than or as fast as $f(x)$. In this paper, we use big-oh to approximate one function as equal to another, to obtain smoother relationships and functions. We establish these from the idea that as we approach ∞ on the number line, these functions are virtually identical. This paper deals with functions that share these relations as asymptotics.

Definition 3.2: If

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 1$$

Then the relationship between f and g can be written as

$$f(x) \sim g(x) \quad \text{as } x \rightarrow \infty \quad (13)$$

Identities and Properties of Big-Oh

Some properties of the big-oh function that will be useful looking through this paper are listed below:

1. For arithmetical functions f, g , where $O(g(x)) > O(f(x))$,

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 0 \quad (14)$$

2. Since the $O(g(x))$ function gives us an idea of the growth rate of $g(x)$, any slower growing function, $f(x)$ than g can also take the form $O(g(x))$, or in summary, if

$$O(g(x)) > O(f(x))$$

Then:

$$O(f(x)) = O(g(x)) \quad (15)$$

Although this is a worse approximation of the function, f , we may need to use such relationships in order to obtain "nicer" functions to handle in our proof.

3. Adding or subtracting two big-oh functions just yields the dominating function as follows:

$$O(g(x)) > O(f(x))$$

Then:

$$O(g(x)) + O(f(x)) = O(g(x)) \quad (16)$$

Order of complexity of functions

The below table represents the order of complexity of different functions, from lowest to highest, or fastest to slowest-growing functions. Knowledge of this table will help in replacing more complex functions with less complex ones for the sake of approximation.

$f(x)$
$x!$
n^x
x^n
$x \log(x)$
$\log(x)$
1

for $n > 1$.

Euler's Summation Formula

Euler summation is a technique used to approximate the partial sum of a function by connecting it to its integral. It is going to be a very useful tool in our proof, where we deal extensively with partial sums.

Lemma 3.1: (Abel's Summation Formula) Let $f(t)$ have a continuous derivative, $f'(t)$, for $t \geq 1$. Let $c_n, n \geq 1$, be constants and let $C(u) = \sum_{n \leq u} c_n$. Then

$$\sum_{n \leq x} c_n f(n) = f(x)C(x) - \int_1^x f'(t)C(t)dt \quad (17)$$

$C(n) - C(n-1) = c_n$ and $C(u) = C(\lfloor u \rfloor)$ since $C(u)$ is a step function, as $u \in \mathbf{N}$. Thus if $\lfloor x \rfloor = K$,

$$\begin{aligned} \sum_{n \leq x} c_n f(n) &= \sum_{n \leq x} (C(n) - C(n-1))f(n) \\ &= \sum_{n \leq x-1} C(n)(f(n) - f(n+1)) + C(x)f(K) \\ &= - \sum_{n \leq x-1} C(n) \int_n^{n+1} f'(t)dt + C(x)f(K) \\ &= - \int_1^K C(t)f'(t)dt + C(x)f(K). \end{aligned}$$

Since $C(t)$ is constant for $K \leq t < x$,

$$\int_K^x C(t)f'(t)dt = C(x)(f(x) - f(K)).$$

Adding this to the previous equation and transposing the integral on the left side to the right proves (17).

Theorem 3.1: (Euler's Summation Formula). Let f be a function with Riemann integrable derivative defined on the interval

$[1, x]$. Then:

$$\sum_{n \leq x} f(n) = \int_1^x f(t) dt + \int_1^x f'(t)(t - [t]) dt + f(x)([x] - x) + f(1) \quad (18)$$

This result follows from Abel's Summation Formula. If $c_n = 1$, $C(x) = [x]$, and so, (17) leads to:

$$\begin{aligned} \sum_{n \leq x} f(n) &= [x]f(x) - \int_1^x tf'(t) dt \\ &= [x]f(x) - \int_1^x tf'(t) dt + \int_1^* f'(t)(t - [t]) dt \end{aligned}$$

Integrating the first integral on the right by parts proves (18).

We will use these findings to deduce information about the asymptotic tendencies of specific arithmetic functions.

Theorem 3.2:

$$\sum_{n \leq x} \frac{1}{n} = \log x + \gamma + O\left(\frac{1}{x}\right) \quad (19)$$

where γ is a constant (Euler's constant).

Applying Euler Summation to $f(t) = 1/t$

$$\begin{aligned} \sum_{n \leq x} \frac{1}{n} &= \int_1^x \frac{dt}{t} - \frac{x - [x]}{x} + 1 - \int_1^x \frac{t - [t]}{t^2} dt. \\ &= \sum_{n \leq x} \frac{1}{n} = \log x - \frac{x - [x]}{x} + 1 - \int_1^x \frac{t - [t]}{t^2} dt. \end{aligned}$$

If

$$\gamma = 1 - \int_1^{\infty} \frac{t - [t]}{t^2} dt \quad (20)$$

then $\sum_{n \leq x} 1/n = \log x + \gamma + \sigma$, where

$$\sigma = \int_x^{\infty} \frac{t - [t]}{t^2} dt - \frac{x - [x]}{x} = O\left(\frac{1}{x}\right)$$

since $t - [t] < 1$ is bounded, which proves (19).

Chebyshev's Functions

In this subsection, we introduce 3 closely-related functions and analyse their asymptotic behaviours. These functions are similar to the prime-number counter and will be used extensively in this proof of the Prime Number Theorem.

Definition 3.3: The Chebyshev-Theta function, $\vartheta(x)$, is defined as:

$$\vartheta(x) := \sum_{p \leq x} \log(p) \quad (21)$$

where p runs over all primes not exceeding x .

Definition 3.4: The Chebyshev-Psi function, $\psi(x)$, is defined as:

$$\psi(x) := \sum_{n \leq x} \Lambda(n) \quad (22)$$

Since $\Lambda(n) = 0$ unless n is a prime power we can write the definition of $\psi(x)$ as follows:

$$\psi(x) = \sum_{n \leq x} \Lambda(n) = \sum_{p^m \leq x} \sum_p \Lambda(p^m) = \sum_{m=1}^{\infty} \sum_{p \leq x^{1/m}} \log p,$$

where m runs over all positive integers.

The sum on m is actually a finite sum. In fact, the sum on p is empty if $x^{1/m} < 2$, that is, if $(\log x)/m < \log 2$, or if

$$m > \frac{\log x}{\log 2} = \log_2 x.$$

Therefore, we have:

$$\psi(x) = \sum_{m \leq \log_2 x} \sum_{p \leq x^{1/m}} \log p$$

The above formula for $\psi(x)$ can be restated using the ϑ function as:

$$\psi(x) = \sum_{m \leq \log_2 x} \vartheta(x^{1/m}) \quad (23)$$

Lemma 3.2: For $x > 0$ we have

$$0 \leq \frac{\psi(x)}{x} - \frac{\vartheta(x)}{x} \leq \frac{(\log x)^2}{2\sqrt{x} \log 2}. \quad (24)$$

From (23) we find

$$0 \leq \psi(x) - \vartheta(x) = \sum_{2 \leq m \leq \log_2 x} \vartheta(x^{1/m}).$$

From the definition of $\vartheta(x)$, we obtain the following inequality:

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

This implies that:

$$\begin{aligned} 0 \leq \psi(x) - \vartheta(x) &\leq \sum_{2 \leq m \leq \log_2 x} x^{1/m} \log(x^{1/m}) \leq (\log_2 x) \sqrt{x} \log \sqrt{x} \\ &= \frac{\log x}{\log 2} \cdot \frac{\sqrt{x}}{2} \log x = \frac{\sqrt{x}(\log x)^2}{2 \log 2}. \end{aligned}$$

Now divide by x to obtain the theorem.

Note: This inequality implies that

$$\lim_{x \rightarrow \infty} \left(\frac{\psi(x)}{x} - \frac{\vartheta(x)}{x} \right) = 0.$$

In other words, if one of $\psi(x)/x$ or $\vartheta(x)/x$ tends to a limit then so does the other, and the two limits are equal.

Definition 3.5: Rewriting (5) from Chapter 2, we can get:

$$\sum_{i|j=x} \Lambda(j) = \log(x)$$

We define a third Chebyshev function, T , that is expressed as:

$$T(x) := \sum_{n \leq x} \log(n) \tag{25}$$

or, using (5),

$$T(x) = \sum_{i|j \leq x} \Lambda(j) \tag{26}$$

If the double sum (26) is treated as a repeated sum, summing first on j ,

$$T(x) = \sum_{i \leq x} \sum_{j \leq x/i} \Lambda(j) = \sum_{i \leq x} \psi(x/i)$$

This identity was discovered by Chebyshev and will be rewritten as

$$T(x) = \sum_{n \leq x} \psi(x/n). \tag{27}$$

Chebyshev Functions and Asymptotics

Theorem 3.3: The above identity is a transform relationship that takes the form similar to f, g in the below definition:

$$g(x) = \sum_{n \leq x} f(x/n)$$

According to Möbius inversion (11), the above relationship can be inverted as:

$$f(x) = \sum_{n \leq x} \mu(n)g(x/n)$$

Replacing f with ψ and g with T , we can obtain the inversion formula:

$$\psi(x) = \sum_{n \leq x} \mu(n)T(x/n) \tag{28}$$

Applying Euler's summation formula to $f(t) = \log t$:

$$T(x) = x \log x - x + O(\log x) \tag{29}$$

Lemma 3.3: For large x ,

$$\psi(x) = O(x). \tag{30}$$

Using (27),

$$T(x) - 2T(x/2) = \psi(x) - \psi(x/2) + \psi(x/3) - \psi(x/4)$$

$$+ \dots \geq \psi(x) - \psi(x/2)$$

because $\psi(x/(2n-1)) - \psi(x/2n) \geq 0$ since ψ is a nondecreasing function.

From (29), $\psi(x) - \psi(x/2) \leq x \log 2 + C \log x$, $x \geq 2$, for a constant C . Applying the above with x replaced by $x/2^j$,

$$\psi\left(\frac{x}{2^j}\right) - \psi\left(\frac{x}{2^{j+1}}\right) \leq \frac{x}{2^j} \log 2 + C \log x$$

so long as $x/2^j \geq 2$ which implies $j < \log x / \log 2$. Recalling that $\psi(t) = 0, t < 2$, and adding the above for $0 \leq j < \log x / \log 2$,

$$\begin{aligned} \psi(x) &\leq x \log 2 \left(1 + \frac{1}{2} + \frac{1}{2^2} + \dots\right) + \frac{\log x}{\log 2} C \log x \\ &= 2x \log 2 + C \log^2 x / \log 2. \end{aligned}$$

Since $\log 2$ is a constant this proves (29), showing that ψ is bounded by a function that is equivalent to $O(x)$.

Lemma 3.4:⁴

$$\sum_{n \leq x} \frac{\Lambda(n)}{n} = \log x + O(1). \tag{31}$$

We rearrange the double sum $T(x) = \sum_{i \leq x} \sum_{j \leq x/i} \Lambda(j)$, summing over i first and then j , we obtain:

$$\begin{aligned} T(x) &= \sum_{j \leq x} \Lambda(j) \sum_{i \leq x/j} 1 = \sum_{j \leq x} \Lambda(j) \left\lfloor \frac{x}{j} \right\rfloor \\ &= x \sum_{j \leq x} \frac{\Lambda(j)}{j} + \sum_{j \leq x} \Lambda(j) \left(\left\lfloor \frac{x}{j} \right\rfloor - \frac{x}{j} \right) \end{aligned}$$

This can be rewritten as:

$$x \sum_{j \leq x} \frac{\Lambda(j)}{j} + \sum_{j \leq x} \Lambda(j) \left(\frac{x}{j} - \left\lfloor \frac{x}{j} \right\rfloor \right) \tag{32}$$

Moreover, we can infer

$$0 \leq \sum_{j \leq x} \Lambda(j) \left(\frac{x}{j} - \left\lfloor \frac{x}{j} \right\rfloor \right) \leq \sum_{j \leq x} \Lambda(j) = \psi(x) = O(x) \tag{33}$$

from (30). Now using (29),

$$\begin{aligned} x \log x - x + O(\log x) &= x \sum_{j \leq x} \frac{\Lambda(j)}{j} + O(x) \\ &= \log x - 1 + O(\log x / x) = \sum_{j \leq x} \frac{\Lambda(j)}{j} + O(1) \end{aligned}$$

The terms -1 , $O(\log x/x)$ and $-O(1)$ can be contained in $O(1)$, which proves the lemma.⁵

Theorem 3.3: The following 4 statements are equivalent:

$$\lim_{x \rightarrow \infty} \frac{\pi(x) \log x}{x} = 1 \quad (34)$$

$$\lim_{x \rightarrow \infty} \frac{\vartheta(x)}{x} = 1 \quad (35)$$

$$\lim_{x \rightarrow \infty} \frac{\psi(x)}{x} = 1 \quad (36)$$

$$\lim_{x \rightarrow \infty} \frac{\psi(x) - x}{x} = 0 \quad (37)$$

Note: Throughout the rest of this paper, the function $\psi(x) - x$ is denoted by $R(x)$ and will be known as remainder function.

We prove the four statements of the theorem later on in chapter 5.

Definition 3.6:

$$R(x) := \psi(x) - x \quad (38)$$

Lemma 3.5:

$$\psi(x) = \pi(x) \log x + O\left(\frac{x \log \log x}{\log x}\right), \quad (39)$$

so that (36) is equivalent to the prime number theorem, since $O\left(\frac{x \log \log x}{\log x}\right)$ grows slower than x^6 . The remainder of this paper seeks to obtain a relationship between $\psi(x)$ and $\pi(x)$ so that we can tie back to this statement.

Using the ψ and Λ function definitions,

$$\psi(x) = \vartheta(x) + \vartheta(x^{1/2}) + \vartheta(x^{1/3}) + \dots \quad (40)$$

where the sums for $p \leq x^{1/r}$ above are zero when $x^{1/r} < 2$ and when $r > \log x / \log 2$. Hence, we approximate ψ using the inequality:

$$\psi(x) \leq \sum_{p \leq x} \log p + \frac{\log x}{\log 2} \sum_{p \leq x^{1/2}} \log p. \quad (41)$$

From the definition of $\pi(x)$, this gives

$$\psi(x) \leq \pi(x) \log x + \frac{\log x}{\log 2} \pi\left(x^{1/2}\right) \log x^{1/2}.$$

Since $\pi(x^{1/2}) < x^{1/2}$ for large x , the above statement yields

$$\psi(x) \leq \pi(x) \log x + \frac{1}{2} \frac{x^{1/2} \log^2 x}{\log 2}.$$

By (40),

$$\begin{aligned} \psi(x) &\geq \vartheta(x) - \vartheta(x/\log^2 x) \\ &= \sum_{x/\log^2 x < p \leq x} \log p \end{aligned}$$

Since the partial sum is greater than its lower limit,

$$\begin{aligned} \psi(x) &\geq \log\left(\frac{x}{\log^2 x}\right) \sum_{x/\log^2 x < p < x} 1 \\ &= \log\left(\frac{x}{\log^2 x}\right) \left(\pi(x) - \pi\left(\frac{x}{\log^2 x}\right)\right). \end{aligned}$$

Since

$$\pi\left(\frac{x}{\log^2 x}\right) \leq \frac{x}{\log^2 x}$$

and

$$\log\left(\frac{x}{\log^2 x}\right) = \log x - 2 \log \log x$$

this gives:

$$\frac{\psi(x)}{\log x - 2 \log \log x} \geq \pi(x) - \frac{x}{\log^2 x},$$

which can then be rewritten as

$$\begin{aligned} \pi(x) \log x &\leq \psi(x) \frac{\log x}{\log x - 2 \log \log x} + \frac{x}{\log x} \\ &= \psi(x) + \psi(x) \frac{2 \log \log x}{\log x - 2 \log \log x} + \frac{x}{\log x} \end{aligned} \quad (42)$$

Using Lemma 3.3, we know that $\psi(x) \leq Gx$, for a constant, G . Replacing this inequality in (42),

$$\pi(x) \log x \leq \psi(x) + \frac{Gx * 2 \log \log x}{\log x - 2 \log \log x} + \frac{x}{\log x}$$

The second and third terms on the RHS can be contained in $O\left(\frac{x \log \log x}{\log x}\right)$. This proves the Lemma.

Selberg's Identity

Theorem 4.1: (Selberg's Asymptotic Formula). For all positive x we have the following:

$$\psi(x) \log x + \sum_{n \leq x} \Lambda(n) \psi\left(\frac{x}{n}\right) = 2x \log x + O(x) \quad (43)$$

This entire section will be a proof of this asymptotic formula, extracted from one of Selberg's lecture series⁷. However, in this application, it is modified to use big-oh approximation instead of the more complex little-oh function.

Lemma 4.2: Let $f(x)$ be a function defined on $x \in \mathbf{R}^+$ and let g be given by:

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

For this relationship, the following statement is true:

$$f(x) \log x + \sum_{n \leq x} f\left(\frac{x}{n}\right) \Lambda(n) = \sum_{d \leq x} \mu(d) g\left(\frac{x}{d}\right) \quad (44)$$

We first rewrite $f(x) \log x$ as a sum using (7). We have:

$$f(x) \log x = \sum_{n \leq x} I(n) f\left(\frac{x}{n}\right) \log \frac{x}{n} = \sum_{n \leq x} \sum_{d|n} \mu(d) f\left(\frac{x}{n}\right) \log \frac{x}{n} \quad (45)$$

Now, we can use the Möbius inversion formula of (5) in Chapter 2 to say that:

$$\Lambda(n) = \sum_{d|n} \mu(d) \log \frac{n}{d}$$

Using both forms of (5) in (43):

$$\sum_{n \leq x} f\left(\frac{x}{n}\right) \Lambda(n) = \sum_{n \leq x} f\left(\frac{x}{n}\right) \sum_{d|n} \mu(d) \log \frac{n}{d} \quad (46)$$

Adding (45) and (46) we have:

$$\sum_{n \leq x} f\left(\frac{x}{n}\right) \sum_{d|n} \mu(d) \left[\log\left(\frac{x}{n}\right) + \log\left(\frac{n}{d}\right) \right]$$

This is equal to:

$$\sum_{n \leq x} f\left(\frac{x}{n}\right) \sum_{d|n} \mu(d) \log \frac{x}{d}$$

Plugging this back into (45):

$$f(x) \log x + \sum_{n \leq x} f\left(\frac{x}{n}\right) \Lambda(n) = \sum_{n \leq x} f\left(\frac{x}{n}\right) \sum_{d|n} \mu(d) \log \frac{x}{d}$$

In the last sum, we write $n = qd$ and obtain:

$$\sum_{n \leq x} f\left(\frac{x}{n}\right) \sum_{d|n} \mu(d) \log \frac{x}{d} = \sum_{d \leq x} \mu(d) \log \frac{x}{d} \sum_{q \leq \frac{x}{d}} f\left(\frac{x}{qd}\right)$$

And by our initial definition of g , we can write the RHS of the above as:

$$\sum_{d \leq x} \mu(d) G\left(\frac{x}{d}\right)$$

as needed.

We now have to prove Theorem 4.1 using this Lemma.⁸

Taking $F(x) = \psi(x)$ and $F_1(x) = x - \gamma - 1$, the respective G would be:

$$G(x) = \log(x) \sum_{n \leq x} \psi\left(\frac{x}{n}\right)$$

which, according to (29) and, gives

$$G(x) = \log(x)T(x) = x \log^2 x - x \log x + O(\log^2 x) \quad (47)$$

and the corresponding G_1 can be expressed as:

$$\begin{aligned} G_1(x) &= \log(x) \sum_{n \leq x} \left(\frac{x}{n}\right) - \gamma - 1 \\ &= x \log(x) \sum_{n \leq x} \left(\frac{1}{n}\right) - (\gamma + 1) \log x \sum_{n \leq x} 1 \end{aligned}$$

Using Theorem 3.2:

$$\begin{aligned} G_1(x) &= \log(x) \sum_{n \leq x} \left(\frac{x}{n}\right) - \gamma - 1 \\ &= x \log x (\log x + \gamma + O(1/x)) - (\gamma + 1) \log x (x + O(1)) \\ &= x \log^2 x - x \log x + O(\log x) \end{aligned}$$

by containing all $\log x$ and constant terms in $O(\log x)$.

Now that we have G and G_1 in similar terms, we use a weak estimate of their difference

$$G(x) - G_1(x) = O(\log^2 x) = O(x^{1/2}) \quad (48)$$

Using Lemma 4.2, we can break this down to:

$$\psi(x) \log x + \sum_{n \leq x} \psi\left(\frac{x}{n}\right) \Lambda(n) = \sum_{d \leq x} \mu(n) G\left(\frac{x}{n}\right) \quad (49)$$

and for $x - \gamma - 1$:

$$(x - \gamma - 1) \log x + \sum_{n \leq x} (x/n - \gamma - 1) \Lambda(n) = \sum_{d \leq x} \mu(n) G_1\left(\frac{x}{n}\right) \quad (50)$$

Subtracting RHS of (48) and (49),

$$\sum_{d \leq x} \mu(n) \left[G\left(\frac{x}{n}\right) - G_1\left(\frac{x}{n}\right) \right]$$

Using our $O(x^{1/2})$ estimate for the difference as our argument, we have:

$$\begin{aligned} \sum_{d \leq x} \mu(n) \left[G\left(\frac{x}{n}\right) - G_1\left(\frac{x}{n}\right) \right] &= O\left[\sum_{d \leq x} (x/n)^{1/2} \right] \\ &= O(x) \sqrt{x} \int_1^x \frac{1}{t} dt = O(x) \end{aligned}$$

Using Lemma 4.2 for $\psi(x) - (x - \gamma - 1)$, we have:

$$\{\psi(x) - (x - \gamma - 1)\} \log x + \sum_{n \leq x} \left\{ \psi\left(\frac{x}{n}\right) - \left(\frac{x}{n} - \gamma - 1\right) \right\} \Lambda(n) = O(x). \quad (51)$$

Rearranging terms and using Lemma 3.4 we find that

$$\psi(x) \log x + \sum_{n \leq x} \psi\left(\frac{x}{n}\right) \Lambda(n) = (x - \gamma - 1) \log x + \sum_{n \leq x} \left(\frac{x}{n} - \gamma - 1\right)$$

$$\Lambda(n) + O(x) = 2x \log x + O(x),$$

proving Selberg's identity. With $c_n = \Lambda(n)$, Abel's Summation and Lemma 3.3 yield:

$$\sum_{n \leq x} \Lambda(n) \log n = \psi(x) \log x - \int_1^x \frac{\psi(t)}{t} dt = \psi(x) \log x + O(x) \quad (52)$$

Also

$$\sum_{n \leq x} \Lambda(n) \psi\left(\frac{x}{n}\right) = \sum_{n \leq x} \Lambda(n) \sum_{k \leq x/n} \Lambda(k) = \sum_{nk \leq x} \Lambda(n) \Lambda(k). \quad (53)$$

Thus if

$$H(n) := \Lambda(n) \log n + \sum_{jk=n} \Lambda(j) \Lambda(k), \quad (54)$$

then (50) and (51) in (41) yield $\sum_{n \leq x} H(n) = 2x \log x + O(x)$ as an equal to Selberg's formula. By (28), $\sum_{n \leq x} \log n = x \log x + O(x)$. Combining the above two inequalities,

$$Q(x) := \sum_{n \leq x} (H(x) - 2 \log n) = O(x), \quad \text{for } n \geq 2, \quad (55)$$

and

$$Q(1) = 0.$$

Proving the Prime Number Theorem

We use Selberg's Identity to prove the prime number theorem. We will mainly be working with the function $R(x)$ as it is more convenient to obtain inequalities that demonstrate its decreasing nature.

Lemma 5.1: Selberg's identity can be restated as:

$$R(x) \log x + \sum_{n \leq x} \Lambda(n) R\left(\frac{x}{n}\right) = O(x)$$

Definition 5.1: Unfortunately, due to the nature of $\psi(x)$, $R(x)$ tends to fluctuate, and does is the quotient, $\frac{R(x)}{x}$. For convenience, we will use the smoother function:

$$S(x) := \int_2^x \frac{R(t)}{t} dt \quad (56)$$

Lemma 5.2: There exists a constant M such that

$$|S(y)| \leq My, \quad y \geq 2 \quad (57)$$

and

$$|S(y_2) - S(y_1)| \leq M|y_1 - y_2|. \quad (58)$$

For this condition, $S(y)$ is considered Lipschitz.

Moreover a consequence of (42) is

$$S(y) \log y + \sum_{j \leq y} \Lambda(j) S\left(\frac{y}{j}\right) = O(y). \quad (59)$$

From (29), $-x \leq \psi(x) - x \leq \frac{1}{2}x$ for large x . Hence

$$\limsup_{x \rightarrow \infty} \frac{|R(x)|}{x} \leq 1$$

suggesting that $|R(x)|$ is bounded for finite x . Hence, there must exist a constant M such that

$$|R(x)| \leq Mx, \quad x \geq 2.$$

From the definition of $S(y)$, $S'(y) = R(y)/y$ except at $y = p^k$ where $R(y)$ is discontinuous. By the above statement then:

$$|S'(y)| \leq M, \quad y \neq p^k.$$

Hence, first for the case where the interval $a < y < b$ contains no p^k , (58) is true.

However since $S(y)$ is continuous, it allows (58) to be extended for all a and b . The result (57) follows from (58) with $a = 2$. Since $||x_1| - |x_2|| \leq |x_1 - x_2|$, (58) yields

$$||S(b)| - |S(a)|| \leq M|b - a|. \quad (60)$$

To prove (59), divide the LHS of Lemma 5.1 by x and integrate to get

$$\int_2^y \frac{R(x)}{x} \log x dx + \sum \Lambda(n) \int_2^y R\left(\frac{x}{n}\right) \frac{dx}{x} = O(y).$$

If we use integration by parts, we get

$$\int_2^y \frac{R(x)}{x} \log x dx = \log y S(y) - \int_2^y \frac{S(x)}{x} dx = \log y S(y) + O(y),$$

which proves (59).

Lemma 5.3: With (54) and Z_1 a constant

$$\log^2 y |S(y)| \leq \sum H(m) |S(y/m)| + Z_1 y \log y. \quad (61)$$

Replace y in (59) by y/k , multiply by $\Lambda(k)$ and sum to get:

$$\sum_{k \leq y} \Lambda(k) S\left(\frac{y}{k}\right) \log \frac{y}{k} + \sum_{m \leq y} \sum_{jk=m} \Lambda(k) \Lambda(j) S\left(\frac{y}{jk}\right) = O(y) \sum_{k \leq y} \frac{\Lambda(k)}{k}.$$

Setting $jk = m$ in the second sum and summing on m , and setting $\log y/k = \log y - \log k$ in the first sum and replacing this latter k by m ,

$$\log y \sum_{k \leq y} \Lambda(k) S\left(\frac{y}{k}\right) - \sum_{m \leq y} S\left(\frac{y}{m}\right) \left[\Lambda(m) \log m - \sum_{jk=m} \Lambda(j) \Lambda(k) \right]$$

$$= O(y \log y)$$

where (30) is used to get the RHS. The first sum above is now replaced by use of (59) to give

$$S(y) \log^2 y = - \sum_{m \leq y} S\left(\frac{y}{m}\right) \left[\Lambda(m) \log m - \sum_{jk=m} \Lambda(j) \Lambda(k) \right] + O(y \log y).$$

Replacing all terms in the sum on the right by their absolute values from the Lipschitz inequalities gives (61).⁹

Lemma 5.4: There exists a constant Z_2 such that:

$$\log^2 y |S(y)| \leq 2 \sum_{m \leq y} \left| S\left(\frac{y}{m}\right) \right| \log m + Z_2 y \log y \quad (62)$$

We define a second remainder function $C(y)$, as given by:

$$C(y) := \sum_{m \leq y} \left| S\left(\frac{y}{m}\right) \right| (H(m) - 2 \log m) \quad (63)$$

Now, recall the function Q defined by:

$$Q(y) = \sum_{m \leq y} (H(m) - 2 \log m)$$

We see that:

$$H(m) - 2 \log m = Q(m) - Q(m-1)$$

Substituting this in (62), we have:

$$\sum_{m \leq y} \left| S\left(\frac{y}{m}\right) \right| (Q(m) - Q(m-1)) = C(y) \quad (64)$$

We can ignore all terms with $m < 2$ as $S(m) = 0$ for those terms. Through examination, we can re-index the RHS as:

$$C(y) = \sum_{2 \leq m \leq y} \left(\left| S\left(\frac{y}{m}\right) \right| - \left| S\left(\frac{y}{m+1}\right) \right| \right) Q(m)$$

Applying the inequality $|x_1 - x_2| \geq ||x_1| - |x_2||$, we have:

$$C(y) \leq \sum_{2 \leq m \leq y} \left(\left| S\left(\frac{y}{m}\right) - S\left(\frac{y}{m+1}\right) \right| \right) Q(m)$$

Because S is Lipschitz, we have a constant Z_3 , such that:

$$\begin{aligned} & \sum_{2 \leq m \leq y} \left(\left| S\left(\frac{y}{m}\right) - S\left(\frac{y}{m+1}\right) \right| \right) Q(m) \\ & \leq \sum_{2 \leq m \leq y} \left(Z_3 \left| \left(\frac{y}{m}\right) - \left(\frac{y}{m+1}\right) \right| \right) Q(m) \end{aligned} \quad (65)$$

Moreover, we can bound $Q(m)$ by some linear function of m as $Q(m) = O(m)$ by (55). So, making Z_3 large enough and substituting (64) we have:

$$C(y) \leq Z_3 m \sum_{2 \leq m \leq y} \left(\left| \left(\frac{y}{m}\right) - \left(\frac{y}{m+1}\right) \right| \right)$$

Factoring the y and simplifying on RHS, we have:

$$C(y) \leq y Z_3 m \sum_{2 \leq m \leq y} \left(\left| \left(\frac{1}{m}\right) - \left(\frac{1}{m+1}\right) \right| \right) = y Z_3 \sum_{2 \leq m \leq y} \left| \left(\frac{1}{m+1}\right) \right|$$

We now obtain:

$$C(y) \leq y Z_3 \sum_{2 \leq m \leq y} \left| \left(\frac{1}{m+1}\right) \right| \leq y Z_3 \int_1^y \frac{1}{m} dm = Z_3 y \log y$$

Combining the above with our expression for $C(y)$ proves the lemma.

Lemma 5.5: There is a constant Z_4 such that:

$$\log^2 y |S(y)| \leq 2 \int_2^y \left| S\left(\frac{y}{u}\right) \right| \log u du + Z_4 y \log y \quad (66)$$

Given that $\log u$ is increasing:

$$\log m |S(y/m)| \leq \int_m^{m+1} \log u |S(y/u)| du.$$

On the RHS, use the inequality $|S(y/m) + S(y/u)| + |S(y/u)| \geq |S(y/m)|$

$$\log m |S(y/m)| \leq \int_m^{m+1} \log u |S(y/u)| du + J_m \quad (67)$$

for

$$J_m := \int_m^{m+1} \log u |S(y/m) - S(y/u)| du$$

$$J_m = \int_m^{m+1} \left| S\left(\frac{y}{m}\right) - S\left(\frac{y}{u}\right) \right| \log u du \leq \int_m^{m+1} M_1 \left| \frac{y}{m} - \frac{y}{u} \right| \log u du$$

Due to S being Lipschitz, we can bound this by:

$$J_m \leq \int_m^{m+1} M_1 \left| \frac{y}{m} - \frac{y}{u} \right| \log u du \leq M_1 \left(\frac{y}{m} - \frac{y}{m+1} \right) \int_m^{m+1} \log u du$$

Simplifying and bounding the final integral on the LHS we have:

$$J_m \leq M_1 \left(\frac{y}{m} - \frac{y}{m+1} \right) \int_m^{m+1} \log u du \leq M_1 y \frac{\log(m+1)}{m(m+1)}$$

Because $m \geq \log(m+1)$ we have:

$$J_m \leq \frac{M_1 y}{m+1}$$

Now, returning to our original expression, we see:

$$\log m \left| S\left(\frac{y}{m}\right) \right| \leq \int_m^{m+1} \left| S\left(\frac{y}{u}\right) \right| \log u du + \frac{M_1 y}{m+1} \quad (68)$$

Using $Z_4 = Z_2 + M_1$ and applying (68) to Lemma 5.4, we have the desired result.

The inequality of (66) assumes a simpler form with a change of variables. Letting $x = \log y$ and letting $v = \log \frac{y}{u}$, we can rewrite (66) as:

$$x^2 |S(e^x)| \leq 2 \int_0^{x-\log 2} |S(e^v)|(x-v)e^{(x-v)} dv + Z_4 x e^x \quad (69)$$

Definition 5.2: We now introduce another function, $W(x)$, for which:

$$W(x) := e^{-x} S(x) \quad (70)$$

From this, we obtain

$$\begin{aligned} x^2 \frac{|W(x)|}{e^{-x}} &\leq \int_0^{x-\log 2} |S(e^v)|(x-v)e^{(x-v)} dv + Z_4 x e^x \\ &= x^2 \frac{|W(x)|}{e^{-x}} \leq 2 \int_0^{x-\log 2} |W(v)|e^v(x-v) \frac{e^x}{e^v} dv + Z_4 x e^x \\ &= |W(x)|e^x \leq \frac{2}{x^2} \int_0^{x-\log 2} |W(v)|(x-v)e^x dv + Z_4 \frac{x e^x}{x^2} \end{aligned}$$

Getting the statement:

$$|W(x)| \leq \frac{2}{x^2} \int_0^x |W(v)|(x-v)dv + Z_4 \frac{1}{x} \quad (71)$$

The transformations above were done to yield a function W which is essentially dominated by a weighted average of $|W|$. We now prove two lemmas about $|W(x)|$.

Lemma 5.6: $\limsup_{x \rightarrow \infty} f(x)$ returns the value for which $f(x)$ is bounded by as x approaches ∞

$$\limsup_{x \rightarrow \infty} |W(x)| = \alpha \leq 1$$

We also define another constant, γ (NOT Euler's constant), as

$$\limsup_{x \rightarrow \infty} \frac{1}{x} \int_0^x |W(x/n)| = \gamma$$

By the statement $S'(x) = O(1)$ we know that:

$$\limsup_{x \rightarrow \infty} \frac{R(x)}{x} \leq 1$$

By definition of S , it follows that:

$$\limsup_{y \rightarrow \infty} \frac{|S(y)|}{y} \leq 1$$

From this, it is clear that

$$\limsup_{x \rightarrow \infty} |W(x)| \leq 1$$

Lemma 5.7: Let

$$\limsup_{x \rightarrow \infty} \frac{1}{x} \int_0^x |W(t)| dt = \beta$$

Then $\beta \geq \alpha = 1$

$$|W(x)| \leq \frac{2}{x^2} \int_0^x u du \left(\frac{1}{u} \int_0^u |W(v)| dv \right) + Z_4 \frac{1}{x} \quad (72)$$

which can be verified by reversing the order of integration. Note that:

$$\frac{2}{x^2} \int_0^x u du = 1$$

Thus we have the integral on the RHS of the form:

$$\frac{1}{u} \int_0^u |W(v)| dv = \frac{1}{u} \int_0^u |e^{-v} S(e^v)| dv$$

And this is bounded by M , the Lipschitz constant from lemma 5.2, hence:

$$\frac{1}{u} \int_0^u |W(v)| dv = \frac{1}{u} \int_0^u |e^{-v} S(e^v)| dv \leq M$$

Thus, similar to the approach taken by Tom M. Apostol¹⁰, if we fix x_1 , and take any $x > x_1$, we have:

$$\begin{aligned} I(x) &:= \frac{2}{x^2} \int_0^x u du \left(\frac{1}{u} \int_0^u |W(v)| dv \right) \\ &\leq \frac{2M}{x^2} \int_0^{x_1} u du + \frac{2}{x^2} \int_{x_1}^x u du \left(\frac{1}{u} \int_0^u |W(v)| dv \right) \end{aligned}$$

as we separate the integrals at x_1 . Given the constant $K > 0$, if we take a sufficiently large x_1 , we have by the definition of β that:

$$\frac{1}{u} \int_0^{x_1} |W(v)| dv \leq \beta + K$$

for all $u \geq x_1$. Thus, substituting into our inequality for $I(x)$ we have:

$$I(x) \leq \frac{Mx_1^2}{x^2} + (\beta + K) \left(1 - \frac{x_1^2}{x^2} \right)$$

Thus, for large x , we have by (71) that:

$$|W(x)| \leq \frac{Mx_1^2}{x^2} + (\beta + K) \left(1 - \frac{x_1^2}{x^2} \right) + \frac{Z_4}{x}$$

For $x \rightarrow \infty$, we have $\alpha \leq \beta + K$, and the inequality holds as it is true for a constant K . Now, we attempt to show $\alpha = 0$

Lemma 5.8: For $c = 2M$,

$$|W(x_1) - W(x_2)| \leq c|x_1 - x_2| \quad (73)$$

Since $W(x) = e^{-x} S(x)$

$$|W'(x)| \leq e^{-x}|S(x)| + |S'(x)|$$

Hence, $|W'(x)| \leq 2M = c$. This leads to (72)

Lemma 5.9: If $W(r) \neq 0$ for $r_1 < r < r_2$ then there exists a constant L such that:

$$\int_{r_1}^{r_2} W(v)dv \leq L \quad (74)$$

We see through an application of Abel Summation formula (Lemma 3.1) that:

$$\int_2^x \frac{\psi(t)}{t} = \log x + O(1)$$

Or, using the remainder function:

$$\int_2^x \frac{R(t)}{t^2} dt = O(1) \quad (75)$$

Taking $\frac{S(y)}{y^2}$, we do another double integral decomposition:

$$\begin{aligned} \int_2^x \frac{S(y)}{y^2} dy &= \int_2^x \frac{dy}{y^2} \int_2^y \frac{R(t)}{t} \\ &= \int_2^x \frac{R(t)}{t} \left(\int_t^x \frac{dy}{y^2} \right) dt \\ &= \int_2^x \frac{R(t)}{t^2} dt - \frac{1}{x} \int_2^x \frac{R(t)}{t} dt \end{aligned}$$

Using $S(x) = O(x)$ and (74), it follows that $\int_2^x \frac{S(y)}{y^2} dy = O(1)$.

By changing variables with $y = e^u, x = e^v$ we have:

$$\int_{\log 2}^v W(u)du = O(1)$$

Writing the same for $v = r_1$ and $v = r_2$ and subtracting the integrals with these as endpoints, we have that the result integral is bounded, and thus there exists an L such that:

$$\int_{r_1}^{r_2} W(u)du \leq L.$$

Lemma 5.10: A function $W(x)$ must in fact have $\alpha = 0$.

Take $\omega > \alpha$. Then from the definition of α there exists an x_ω such that

$$|W(x)| \leq \omega \quad x \geq x_\omega. \quad (76)$$

If $W(x) \neq 0$ for all large x it follows from (74) that $\gamma = 0$ and hence that $\alpha = 0$. Suppose then that $W(x)$ has arbitrarily large zeros. Let a and b be successive zeros of $W(x)$ for $x > x_\omega$.

Case 1: $b - a \geq 2L/\omega$. By (74), since $W(x) \neq 0, a < x < b$,

$$\int_a^b |W(x)| \leq L \leq \frac{1}{2}(b - a)\omega.$$

(Hence the average of $|W|$ on (a, b) is less than $\frac{1}{2}\omega$.)

Case 2: $b - a \leq 2\omega/k$. In this case it follows from (72) that if the graph of $|W(x)|$ rises as rapidly as possible going right from $x = a$ and left from $x = b$, it cannot lie above a triangle with altitude $k(b - a)/2 \leq \omega$ and hence

$$\int_a^b |W(x)|dx \leq \frac{1}{2}(b - a)\omega.$$

Case 3: $2\omega/k < b - a < 2L/\omega$. Reasoning as in Case 2 for a distance ω/k from each endpoint and using (76) otherwise,

$$\begin{aligned} \int_a^b |W(x)| &\leq \frac{\beta^2}{k} + \left(b - a - \frac{2\beta}{k}\right)\beta \\ &= (b - a)\beta \left(1 - \frac{\beta}{k(b - a)}\right) \leq (b - a)\beta \left(1 - \frac{\beta^2}{2Mk}\right) \\ &< (b - a)\beta \left(1 - \frac{\alpha^2}{2Mk}\right). \end{aligned}$$

Since $Mk > 1$ and since $\alpha \leq 1$, the above is valid in Cases 1 and 2 also. If x_1 is the first zero of $W(x)$ to the right of x_β and \bar{x} the largest zero to the left of y , then (73) and the above imply that

$$\int_0^y |W(x)|dx \leq \int_0^{x_1} |W(x)|dx + (\bar{x} - x_1)\beta \left(1 - \frac{\alpha^2}{2Lk}\right) + L.$$

Dividing by y and noting that $x \leq y$,

$$\frac{1}{y} \int_0^y |W(x)|dx \leq \frac{1}{y} \int_0^{x_1} |W(x)|dx + \beta \left(1 - \frac{\alpha^2}{2Mk}\right) + \frac{M}{y}.$$

Letting $y \rightarrow \infty, \gamma \leq \beta \left(1 - \frac{\alpha^2}{2Lk}\right)$, and since $\gamma \geq \alpha$,

$$\alpha \leq \beta \left(1 - \frac{\alpha^2}{2Lk}\right). \quad (77)$$

Since this holds for all $\beta > \alpha$ it must hold for $\beta = \alpha$. Hence $\alpha^3 \leq 0$, and since $\alpha \geq 0$, this implies $\alpha = 0$. Since $W(x) = e^{-x}S(e^x)$, this implies that $|S(y)|/y \rightarrow 0$ as $y \rightarrow \infty$. Hence if given $K > 0$, if y is large enough,

$$|S(y)| \leq \frac{1}{3}K^2y.$$

Thus $S(y(1 + K)) - S(y) \leq \frac{1}{3}K^2(y(1 + K) + y) < K^2y$, or

$$\int_y^{y(1+K)} \frac{R(u)}{u} du \leq K^2y.$$

Since $R(u) = \psi(u) - u$ and ψ is nondecreasing,

$$\frac{\psi(y)}{y(1 + K)} \int_y^{y(1+K)} du - \int_y^{y(1+K)} du \leq K^2y.$$

Hence $\psi(y)/y \leq (1 + K)^2$. Similarly $S(y) - S(y(1 - K)) \geq -K^2y$ for large enough y leads to $\psi(y)/y \geq (1 - K)^2$.

End of the Proof: Since K is an arbitrary constant, by integrating and eliminating K , we can prove that (37) is true. As established in Theorem 3.3, (37) is equivalent to (36). Now that the asymptotic relationship of $\psi(x)$ has been obtained, we can directly state that (34) is true by (39). By definition, (34) is indeed the prime number theorem and thus, it can be concluded that the prime number theorem does hold true.

Acknowledgements

I express my sincere gratitude to everyone who has contributed to the completion of this project. Special thanks to my mentor, Lukas Kofler, for his guidance, support, and invaluable insights throughout the research process. I am also thankful to Lumiere Education, for providing this platform to me, and my program manager, Srishti Bansal, for organising and managing this research experience.

References

- 1 D. Zagier, *The American Mathematical Monthly*, 1997, **104**, 705–708.
- 2 N. Levinson, *The American Mathematical Monthly*, 1969, **76**, 225–245.
- 3 J. Binder, 2008.
- 4 M. B. Villarino, *Depto. de Matemática, Universidad de Costa Rica*, 2005.
- 5 J.-M. D. Koninck and F. Luca, *Analytic Number Theory: Exploring the Anatomy of Integers*, The American Mathematical Society, 2012, vol. 134.
- 6 P. Erdős, *Proceedings of the National Academy of Sciences*, 1949, **35**, 374 – 384.
- 7 author, 1998.
- 8 A. Choudhary, *REUPapers*, 2017.
- 9 G. Hardy and E. Wright, *An Introduction to the Theory of Numbers*, Oxford at the Clarendon Press, 4th edn, 1971.
- 10 T. M. Apostol, *Introduction to Analytic Number Theory*, Springer-Verlag, 1976.