Selberg's Elementary Proof of the Prime Number Theorem

A report submitted by
Sayan Das
BSc 3rd year, Jadavpur University
as part of a
Summer Internship

advised by
Prof. Satadal Ganguly
Associate Professor,
Theoretical Statistics and Mathematics Division,
Indian Statistical Institute Kolkata

Abstract

This report provides an exposition of Selberg's proof of the Prime Number Theorem (PNT), $\pi(x) \sim \frac{x}{\log x}$, or rather the equivalent statement $\psi(x) \sim x$. Selberg's proof is remarkable due to it being the first proof of the PNT not requiring any complex analytic techniques. The key is Selberg's asymptotic formula $\psi(x)\log x + \sum_{n \leqslant x} \Lambda(n)\psi\left(\frac{x}{n}\right) = 2x\log x + \mathrm{O}(x)$. A generalisation to prime ideals in a number field, the Prime Ideal Theorem, is also discussed briefly.

1. Introduction

The Prime Number Theorem (PNT) is the pinnacle of classical analytic number theory, and a fundamental result about the distribution of primes. The original proof of the PNT, due to Hadamard and de la Valée Poussin, uses complex analytic techniques and properties of Riemann's ζ-function. It was not known whether an elementary proof was possible (Hardy even thought it to be impossible), until Selberg's proof [Sel49] in 1949; Erdős also published a proof around the same time.

[Apo89, Chapter 4] sketches an outline of Selberg's proof. My aim here is to fill in the technical details and present a complete proof, following very closely the proof given in [HW08]. I also briefly discuss a generalisation to prime ideals in a number field, the so-called Prime Ideal Theorem.

1.1. Acknowledgments

This report is a culmination of a month of working through Apostol's *Introduction to Analytic Number Theory* [Apo89] under the supervision of Professor Satadal Ganguly as part of a summer internship. Thus, I am extremely grateful to Professor Satadal Ganguly for his kind mentorship and guidance throughout the entire program.

2. Preliminaries

2.1. Notation

I use the $\mathrm{O}(\cdot)$, $\mathrm{o}(\cdot)$ and $\sim$ notations. For functions $f\colon \mathbb{R} \to \mathbb{C}$, $g : \mathbb{R} \to [0, \infty)$, $f(x) = \mathrm{O}\big(g(x)\big)$ means $|f| \leqslant Cg$ for some constant C, $f(x) = \mathrm{o}\big(g(x)\big)$ means $|f/g| \to 0$ as $x \to \infty$, and $f(x) \sim g(x)$ means $|f/g| \to 1$ as $x \to \infty$. The notation $\lfloor x \rfloor$ refers to the floor function, the largest integer n such that $n \leqslant x$. Note that I *never* use $\{x\}$ to refer to the fractional part of x.

Finally, I always use the letter p is to denote a prime, and $\mathbb{N}$ to denote the positive integers starting from 1.

2.2. Basic Notions

The PNT is usually stated in terms of the prime counting function $\pi(x)$,

Definition 2.1. The *prime counting function* $\pi(x) : (0, \infty) \to \mathbb{C}$ is given by

$$\pi(x) = \#\{p : p \leqslant x\}.$$

Theorem 2.2 (Prime Number Theorem)**.**

$$\pi(x) \sim \frac{x}{\log x}.$$

This is the main result that we will prove, and for this we need a few more preliminary definitions and results.

Definition 2.3. The *Euler-Mascheroni* constant is defined as

$$\gamma = \lim_{n\to\infty} \left(1 + \frac{1}{2} + \cdots + \frac{1}{n} - \log n\right).$$

Definition 2.4. The *Riemann ζ-function* is defined by

$$\zeta(s) = \begin{cases} \displaystyle\sum_{n=1}^{\infty} \frac{1}{n^s}, & \text{if } s > 1 \\ \displaystyle\lim_{x\to\infty} \left(\sum_{n\leqslant x} \frac{1}{n^s} - \frac{x^{1-s}}{1-s}\right), & \text{if } 0 < s < 1. \end{cases}$$

Definition 2.5. The *von Mangoldt function* $\Lambda : \mathbb{N} \to \mathbb{C}$ is given by

$$\Lambda(n) = \begin{cases} \log(p), & \text{if } n = p^k, \text{ where } k \geqslant 1 \text{ is an integer} \\ 0, & \text{otherwise.} \end{cases}$$

Definition 2.6. The *Chebyshev ϑ-function* $\vartheta : (0, \infty) \to \mathbb{C}$ is given by

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

Definition 2.7. The *Chebyshev ψ-function* $\psi : (0, \infty) \to \mathbb{C}$ is given by

$$\psi(x) = \sum_{n \leqslant x} \Lambda(n).$$

We will prove the PNT using Selberg's method, by proving the equivalent PNT in Chebyshev form.

Theorem 2.8 (PNT in Chebyshev form [Apo89, Theorem 4.4])**.** *We have*

$$\pi(x) \sim \frac{x}{\log x} \iff \vartheta(x) \sim x \iff \psi(x) \sim x.$$

The following theorem due to Tatuzawa and Iseki [TI51] will be used to prove Selberg's asymptotic formula ((3.1)).

Theorem 2.9 ([Apo89, Theorem 4.17])**.** *Let $F : (0, \infty) \to \mathbb{C}$ and $G(x) = \log x \sum_{n \leqslant x} F\left(\frac{x}{n}\right)$. Then*

$$F(x) \log x + \sum_{n \leqslant x} F\left(\frac{x}{n}\right) \Lambda(n) = \sum_{d \leqslant x} \mu(d) G\left(\frac{x}{d}\right).$$

Proof. First write $F(x) \log x$ as a sum

$$F(x) \log x = \sum_{n \leqslant x} \left[\frac{1}{n}\right] F\left(\frac{x}{n}\right) \log \frac{x}{n} = \sum_{n \leqslant x} F\left(\frac{x}{n}\right) \log \frac{x}{n} \sum_{d|n} \mu(d)$$

and using the identity $\Lambda(n) = \sum_{d|n} \mu(d) \log \frac{n}{d}$ we can write

$$\sum_{n \leqslant x} F\left(\frac{x}{n}\right) \Lambda(n) = \sum_{n \leqslant x} F\left(\frac{x}{n}\right) \sum_{d|n} \mu(d) \log \frac{n}{d}.$$

Now adding these two equations we get

$$\begin{aligned} F(x) \log x + \sum_{n \leqslant x} F\left(\frac{x}{n}\right) \Lambda(n) &= \sum_{n \leqslant x} F\left(\frac{x}{n}\right) \sum_{d|n} \mu(d) \left\{\log \frac{x}{n} + \log \frac{n}{d}\right\} \\ &= \sum_{n \leqslant x} F\left(\frac{x}{n}\right) \sum_{d|n} \mu(d) \log \frac{x}{d} \\ &= \sum_{n \leqslant x} \sum_{d|n} F\left(\frac{x}{n}\right) \mu(d) \log \frac{x}{d} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{q \leqslant x/d} \sum_{d \leqslant x} F\left(\frac{x}{qd}\right) \mu(d) \log \frac{x}{d} \qquad \text{write } n = qd \\
&= \sum_{d \leqslant x} \mu(d) \log \frac{x}{d} \sum_{q \leqslant x/d} F\left(\frac{x}{qd}\right) \\
&= \sum_{d \leqslant x} \mu(d) G\left(\frac{x}{d}\right).
\end{aligned}
$$

□

The following technique called *Abel summmation* will be used several times.

Theorem 2.10 (Abel summmation [Apo89, Theorem 4.2]). *For any arithmetic function* $a : \mathbb{N} \to \mathbb{C}$ *let*

$$A(x) = \sum_{n \leqslant x} a(n),$$

where $A(x) = 0$ *for all* $x < 1$. *Assume* f *has a continuous derivative on the interval* $[y, x]$, *where* $0 < y < x$. *Then*

$$\sum_{y < n \leqslant x} a(n) f(n) = A(x) f(x) - A(y) f(y) - \int_y^x A(t) f'(t) \mathrm{d}t.$$

Proof. As $A(x)$ is a step function with jump $f(n)$ at each integer n, we can express the sum on the left as a Riemann-Stieltjes integral

$$\sum_{y \leqslant n \leqslant x} a(n) f(n) = \int_y^x f(t) \mathrm{d}A(t).$$

Integrating by parts, we get

$$
\begin{aligned}
\sum_{y \leqslant n \leqslant x} a(n) f(n) &= f(x) A(x) - f(y) A(y) - \int_y^x A(t) \mathrm{d}f(t) \\
&= f(x) A(x) - f(y) A(y) - \int_y^x A(t) f'(t) \mathrm{d}t.
\end{aligned}
$$

In particular, for $y < 1$, we have $\sum_{n \leqslant x} a(n) f(n) = A(x) f(x) - \int_1^x A(t) f'(t) \mathrm{d}t.$ □

The following estimates will also be useful.

Theorem 2.11 ([Apo89, Theorem 3.2(a)]). *If $x \geqslant 1$, then*

$$\sum_{n \leqslant x} \frac{1}{n} = \log x + \gamma + \mathrm{O}\left(\frac{1}{x}\right).$$

Theorem 2.12 ([Apo89, Theorem 3.2(b)]). *If $x \geqslant 1,\ s > 0$ and $s \neq 1$ then*

$$\sum_{n \leqslant x} \frac{1}{n^s} = \frac{x^{1-s}}{1-s} + \zeta(s) + \mathrm{O}(x^{-s}).$$

Theorem 2.13 ([Apo89, Theorem 4.9]).

$$\sum_{n \leqslant x} \frac{\Lambda(n)}{n} = \log x + \mathrm{O}(1).$$

Theorem 2.14 ([Apo89, Theorem 4.11]). *For all $x \geqslant 1$ we have*

$$\sum_{n \leqslant x} \psi\left(\frac{x}{n}\right) = x \log x - x + \mathrm{O}(\log x).$$

Theorem 2.15 ([HW08, Theorem 414]).

$$\psi(x) = \mathrm{O}(x).$$

Proof. Using Theorem (2.14), and the fact that $\log x < x$, we get

$$\psi(x) - \psi\left(\frac{x}{2}\right) = x \log x + \mathrm{O}(x) - 2\left(\frac{x}{2} \log \frac{x}{2} + \mathrm{O}(x)\right) = \mathrm{O}(x).$$

So there exists a constant $K > 0$ such that

$$\psi(x) - \psi\left(\frac{x}{2}\right) \leqslant Kx \quad \forall x \geqslant 1.$$

Replacing x successively by $x/2$, $x/4$, ... we obtain

$$\psi\left(\frac{x}{2}\right) - \psi\left(\frac{x}{4}\right) \leqslant K\frac{x}{2}$$

$$\psi\left(\frac{x}{4}\right) - \psi\left(\frac{x}{8}\right) \leqslant K\frac{x}{4}$$

and so forth. Note that $\psi(x/2^n) = 0$ when $2^n > x$. Adding these inequalities yields

$$\psi(x) \leqslant Kx\left(1 + \frac{1}{2} + \frac{1}{4} + \cdots\right) = 2Kx.$$

Hence, $\psi(x) \leqslant Bx$ with $B = 2K$, so $\psi(x) = \mathrm{O}(x)$. □

2.3. Outline of the argument

The key lemma is Selberg's asymptotic formula

$$\psi(x)\log x + \sum_{n\leqslant x}\Lambda(n)\psi\left(\frac{x}{n}\right) = 2x\log x + \mathrm{O}(x).$$

It is natural (it will become clear later why) to define a function

$$\sigma(x) = e^{-x}\psi(e^x) - 1,$$

then Selberg's formula implies the inequality

$$x^2\left|\sigma(x)\right| \leqslant 2\int_0^x\int_0^y\left|\sigma(u)\right|\mathrm{d}u\ \mathrm{d}y + \mathrm{O}(x). \tag{2.1}$$

The PNT is then equivalent to the statement: $\sigma(x)\to 0$ as $x\to\infty$. Hence, if we let

$$C = \limsup_{x\to\infty}\left|\sigma(x)\right|,\ K = \limsup_{x\to\infty}\frac{1}{x}\int_0^x\left|\sigma(u)\right|\mathrm{d}u$$

then the PNT is equivalent to showing that $C = 0$. This is proved as follows by assuming towards a contradiction that $C > 0$. From the definition of C and K,

$$\left|\sigma(x)\right| \leqslant C + \mathrm{o}(1),\ \left|\sigma(x)\right| \leqslant K + \mathrm{o}(1) \tag{2.2}$$

with $C \leqslant K$. If $C > 0$, then this inequality along with (2.1) yields $K < C$, which is absurd. So $C = 0$.

3. Proof of the main result

3.1. Selberg's asymptotic formula

The following asymptotic formula of Selberg is the key lemma in this proof

Theorem 3.1 (Selberg's theorem). *For $x > 0$ we have*

$$\psi(x)\log x + \sum_{n\leqslant x}\Lambda(n)\psi\left(\frac{x}{n}\right) = 2x\log x + \mathrm{O}(x)$$

and

$$\sum_{n\leqslant x}\Lambda(n)\log n + \sum_{mn\leqslant x}\Lambda(m)\Lambda(n) = 2x\log x + \mathrm{O}(x).$$

Proof. The two statements above are equivalent as

$$\sum_{n\leqslant x}\Lambda(n)\psi\left(\frac{x}{n}\right) = \sum_{n\leqslant x}\Lambda(n)\sum_{m\leqslant x/n}\Lambda(m) = \sum_{mn\leqslant x}\Lambda(m)\Lambda(n)$$

and using Abel summmation with $a(n) = \Lambda(n)$ and $f(t) = \log t$ we get

$$\sum_{n\leqslant x}\Lambda(n)\log n = \psi(x)\log x - \int_2^x\frac{\psi(t)}{t}\mathrm{d}t = \psi(x)\log x + \mathrm{O}(x)$$

using Theorem (2.15). Now, we apply Theorem (2.9) to the functions $F_1 = \psi(x)$ as well as $F_2(x) = x - \gamma - 1$, where γ is the Euler-Mascheroni constant. For F_1, using Theorem (2.14), we have

$$G_1(x) = \log x \sum_{n \leqslant x} \psi\left(\frac{x}{n}\right) = x \log^2 x - x \log x + \mathrm{O}(\log^2 x).$$

For F_2, using Theorem (2.11) we have

$$\begin{aligned} G_2(x) &= \log x \sum_{n \leqslant x} \left(\frac{x}{n} - \gamma - 1\right) \\ &= x \log x \sum_{n \leqslant x} \frac{1}{n} - (\gamma + 1) \log x \sum_{n \leqslant x} 1 \\ &= x \log x \left(\log x + \gamma + \mathrm{O}\left(\frac{1}{x}\right)\right) - (\gamma + 1)\big(x + \mathrm{O}(1)\big) \log x \\ &= x \log^2 x - x \log x + \mathrm{O}(\log x). \end{aligned}$$

Hence, $G_1(x) - G_2(x) = \mathrm{O}(\log^2 x)$. Only the weaker estimate $G_1(x) - G_2(x) = \mathrm{O}(\sqrt{x})$ is needed in fact. Now applying Theorem (2.9) to F_1 and F_2 yields

$$\begin{aligned} &\{F_1(x) - F_2(x)\} \log x + \sum_{n \leqslant x} \left\{F_1\left(\frac{x}{n}\right) - F_2\left(\frac{x}{n}\right)\right\} \Lambda(n) \\ =& \sum_{d \leqslant x} \mu(d) \left\{G_1\left(\frac{x}{d}\right) - G_2\left(\frac{x}{d}\right)\right\} = \mathrm{O}\left(\sum_{d \leqslant x} \sqrt{\frac{x}{d}}\right). \end{aligned}$$

Then applying Theorem (2.12) to the above yields

$$\{\psi(x) - (x - \gamma - 1)\} \log x + \sum_{n \leqslant x} \left\{\psi\left(\frac{x}{n}\right) - \left(\frac{x}{n} - \gamma - 1\right)\right\} \Lambda(n) = \mathrm{O}\left(\sqrt{x} \sum_{d \leqslant x} \frac{1}{\sqrt{d}}\right) = \mathrm{O}(x).$$

Hence, using Theorem (2.13), we get

$$\begin{aligned} \psi(x) \log x + \sum_{n \leqslant x} \Lambda(n) \psi\left(\frac{x}{n}\right) &= (x - \gamma - 1) \log x + \sum_{n \leqslant x} \left(\frac{x}{n} - \gamma - 1\right) \Lambda(n) + \mathrm{O}(x) \\ &= x \log x + x \sum_{n \leqslant x} \frac{\Lambda(n)}{n} - (\gamma + 1) \left\{\log x + \sum_{n \leqslant x} \Lambda(n)\right\} + \mathrm{O}(x) \\ &= 2x \log x + \mathrm{O}(1) - 2(\gamma + 1) \log x + \mathrm{O}(x) \\ &= 2x \log x + \mathrm{O}(x) \end{aligned}$$

where the last step is due to the fact that $\log x < x$. □

3.2. Proof of the PNT in Chebyshev form

We now prove the PNT in Chebyshev form (Theorem (2.8)). Set $\psi(x) = x + R(x)$; the aim is to show that $R(x) = \mathrm{o}(x)$. From Theorem (3.1) we get

$$x \log x + R(x) \log x + \sum_{n \leqslant x} \Lambda(n) \left(\frac{x}{n}\right) + \sum_{n \leqslant x} \Lambda(n) R\left(\frac{x}{n}\right) = 2x \log x + \mathrm{O}(x).$$

Then using Theorem (2.13) we get

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

Replacing n by m, x by x/n

$$R\left(\frac{x}{n}\right) \log\left(\frac{x}{n}\right) + \sum_{m \leqslant x/n} \Lambda(m) R\left(\frac{x}{mn}\right) = \mathrm{O}\left(\frac{x}{n}\right).$$

Hence, using Theorem (2.13) again

$$\begin{aligned}
&\log x \left\{ R(x) \log x + \sum_{n \leqslant x} \Lambda(n) R\left(\frac{x}{n}\right) \right\} \\
&- \sum_{n \leqslant x} \Lambda(n) \left\{ R\left(\frac{x}{n}\right) \log\left(\frac{x}{n}\right) + \sum_{m \leqslant x/n} \Lambda(m) R\left(\frac{x}{mn}\right) \right\} \\
&= \mathrm{O}(x \log x) + \mathrm{O}\left(x \sum_{n \leqslant x} \frac{\Lambda(n)}{n} \right) = \mathrm{O}(x \log x).
\end{aligned}$$

Distributing the first and second terms, and using $\log(x/n) = \log x - \log n$

$$\begin{aligned}
R(x) \log^2 x &= - \sum_{n \leqslant x} \Lambda(n) R\left(\frac{x}{n}\right) \log n \\
&\quad + \sum_{mn \leqslant x} \Lambda(m) \Lambda(n) R\left(\frac{x}{mn}\right) + \mathrm{O}(x \log x) \\
&= - \sum_{n \leqslant x} \Lambda(n) R\left(\frac{x}{n}\right) \log n \\
&\quad + \sum_{n \leqslant x} \sum_{hk = n} \Lambda(h) \Lambda(k) R\left(\frac{x}{hk}\right) + \mathrm{O}(x \log x) \\
\implies |R(x)| \log^2 x &\leqslant \sum_{n \leqslant x} \Lambda(n) \left| R\left(\frac{x}{n}\right) \right| \log n \\
&\quad + \sum_{n \leqslant x} \sum_{hk = n} \Lambda(h) \Lambda(k) \left| R\left(\frac{x}{hk}\right) \right| + \mathrm{O}(x \log x) \\
&\leqslant \sum_{n \leqslant x} \left\{ \Lambda(n) \log n + \sum_{hk = n} \Lambda(h) \Lambda(k) \right\} \left| R\left(\frac{x}{n}\right) \right| + \mathrm{O}(x \log x)
\end{aligned}$$

from where, noting that $\sum\limits_{mn\leqslant x}\Lambda(m)\Lambda(n)R\left|\left(\frac{x}{mn}\right)\right| = \sum\limits_{\ell\leqslant x}\Lambda(m)\Lambda\left(\frac{\ell}{m}\right)R\left|\left(\frac{x}{\ell}\right)\right|$, we get

$$|R(x)|\log^2 x \leqslant \sum_{n\leqslant x} a_n \left|R\left(\frac{x}{n}\right)\right| + \mathrm{O}(x\log x) \tag{3.1}$$

where $a_n = \Lambda(n)\log n + \sum\limits_{hk=n}\Lambda(h)\Lambda(k)$ and $\sum\limits_{n\leqslant x} a_n = 2x\log x + \mathrm{O}(x)$. Now we replace the sum with an integral.

Lemma 3.2.

$$|R(x)|\log^2 x \leqslant 2\int_1^x \left|R\left(\frac{x}{t}\right)\right|\log t\mathrm{d}t + \mathrm{O}(x\log x). \tag{3.2}$$

Proof. If $t > t' \geqslant 0$, and $F(t) := \psi(t) + t = \mathrm{O}(t)$ be an increasing function, then

$$\begin{aligned} ||R(t)| - |R(t')|| &\leqslant |R(t) - R(t')| = |\psi(t) - \psi(t') - (t - t')| \\ &\leqslant \psi(t) - \psi(t') + t - t' = F(t) - F(t') \end{aligned}$$

$$\begin{aligned} \implies \sum_{n\leqslant x-1} n\left\{F\left(\frac{x}{n}\right) - F\left(\frac{x}{n+1}\right)\right\} &= \sum_{n\leqslant x} F\left(\frac{x}{n}\right) - [x]F\left(\frac{x}{[x]}\right) \\ &= \mathrm{O}\left(x\sum_{n\leqslant x}\frac{1}{n}\right) = \mathrm{O}(x\log x). \end{aligned}$$

Let $c(1) = 0,\ c(n) = a_n - 2\int_{n-1}^n \log t\mathrm{d}t,\ f(n) = \left|R(\frac{x}{n})\right|,\ C(x) = \sum\limits_{n\leqslant x} c(n)$,

then $C(x) = \sum\limits_{n\leqslant x} a_n - 2\int_1^{[x]}\log t\mathrm{d}t = \mathrm{O}(x)$ and using

$$\sum_{n\leqslant x} c(n)f(n) = \sum_{n\leqslant x-1} C(n)\left\{f(n) - f(n+1)\right\} + C(x)f\left([x]\right)$$

we have

$$\begin{aligned} \sum_{n\leqslant x} a_n \left|R\left(\frac{x}{n}\right)\right| &- 2\sum_{2\leqslant n\leqslant x}\left|R\left(\frac{x}{n}\right)\right|\int_{n-1}^n \log t\mathrm{d}t \\ &= \sum_{n\leqslant x-1} C(n)\left\{\left|R\left(\frac{x}{n}\right)\right| - \left|R\left(\frac{x}{n+1}\right)\right|\right\} + C(x)\left|R\left(\frac{x}{[x]}\right)\right| \\ &= \mathrm{O}\left(\sum_{n\leqslant x-1} n\left\{F\left(\frac{x}{n}\right) - F\left(\frac{x}{n+1}\right)\right\}\right) + \mathrm{O}(x) \\ &= \mathrm{O}(x\log x). \end{aligned} \tag{3.3}$$

Now

$$\begin{aligned} \left|\left|R\left(\frac{x}{n}\right)\right|\int_{n-1}^n \log t\mathrm{d}t - \int_{n-1}^n \left|R\left(\frac{x}{t}\right)\right|\log t\mathrm{d}t\right| &\leqslant \int_{n-1}^n \left|\left|R\left(\frac{x}{n}\right)\right| - \left|R\left(\frac{x}{t}\right)\right|\right|\log t\mathrm{d}t \\ \leqslant \int_{n-1}^n \left\{F\left(\frac{x}{t}\right) - F\left(\frac{x}{n}\right)\right\}\log t\mathrm{d}t &\leqslant (n-1)\left\{F\left(\frac{x}{n-1}\right) - F\left(\frac{x}{n}\right)\right\} \end{aligned}$$

so that

$$\sum_{2\leqslant n\leqslant x}\left|R\left(\frac{x}{n}\right)\right|\int_{n-1}^{n}\log t\mathrm{d}t-\int_{1}^{x}\left|R\left(\frac{x}{t}\right)\right|\log t\mathrm{d}t$$

$$=\mathrm{O}\left(\sum_{n\leqslant x-1}n\left\{F\left(\frac{x}{n}\right)-F\left(\frac{x}{n+1}\right)\right\}\right)+\mathrm{O}(x\log x)$$

$$=\mathrm{O}(x\log x). \tag{3.4}$$

Adding the equations (3.3) and 2×(3.4) yields (3.2). □

Hence, we can rewrite (3.1) as

$$\log^2(z)\,|R(z)|\leqslant 2\int_1^z\left|R\left(\frac{z}{t}\right)\right|\log t\mathrm{d}t+\mathrm{O}(z\log z). \tag{3.5}$$

Showing $R(x)=\mathrm{o}(x)$ directly using (3.5) is hard because the behaviour of the von Mangoldt function $\Lambda(n)$ depends on the location of primes which is exactly what we're trying to find. Hence it is natural to define a smoother function $\sigma(x)=e^{-x}R(e^x)=e^{-x}\psi(e^x)-1$. Substitute $z=e^x,\ t=ze^{-u} \implies \mathrm{d}t=-ze^{-u}\mathrm{d}u$. Then for the integral's limits in (3.5), $t=z$ when $u=0$ and $t=1$ when $u=x$. Also $|R(z/t)|=|R(e^u)|=e^u|\sigma(u)|$ and $\log t=x-u$. So the integral becomes simplified as

$$\int_1^z\left|R\left(\frac{z}{t}\right)\right|\log t\mathrm{d}t=-z\int_x^0 e^u|\sigma(u)|(x-u)e^{-u}\mathrm{d}u$$

$$=z\int_0^x|\sigma(u)|(x-u)\mathrm{d}u=e^x\int_0^x|\sigma(u)|\int_u^x\mathrm{d}y\mathrm{d}u$$

$$\text{(change the order of integration)}\ =e^x\int_0^x\int_0^y|\sigma(u)|\mathrm{d}u\mathrm{d}y.$$

Hence we may rewrite (3.5) as the simpler

$$x^2|\sigma(x)|\leqslant 2\int_0^x\int_0^y|\sigma(u)|\mathrm{d}u\mathrm{d}y+\mathrm{O}(x). \tag{3.6}$$

As $\psi(x)=\mathrm{O}(x)$ by Theorem (2.15), by definition $\sigma(x)$ is bounded for large x. So the upper limits

$$C=\limsup_{x\to\infty}|\sigma(x)|\quad\text{and}\quad K=\limsup_{x\to\infty}\frac{1}{x}\int_0^x|\sigma(u)|\mathrm{d}u \tag{3.7}$$

exist. Then

$$\sigma(x)\leqslant C+\mathrm{o}(1)\quad\text{and}\quad\int_0^x|\sigma(u)|\mathrm{d}u\leqslant Kx+\mathrm{o}(x) \tag{3.8}$$

so using (3.6) we get

$$\sigma(x)\leqslant K+\mathrm{o}(1). \tag{3.9}$$

Hence $C\leqslant K$. Now $R(x)=\mathrm{o}(x)\iff\sigma(x)=\mathrm{o}(1)$ so our aim is to show $C=0$. So we assume for contradiction that $C>0$, then show that $K<C$ which is absurd.

We need the following two lemmas.

Lemma 3.3. *There is a fixed $A_1 > 0$ such that for all $x_1, x_2 > 0$ we have* $\left|\int_{x_1}^{x_2} \sigma(u)\mathrm{d}u\right| < A_1$.

Proof. From Theorem (2.13), and using Abel summmation with $a(n) = \Lambda(n), f(t) = \frac{1}{t}$ we get the estimate $\int_2^z \frac{\psi(t)}{t^2}\mathrm{d}t = \log z + \mathrm{O}(1)$. Substituting $z = e^x, t = e^u$ then yields

$$\int_0^x \sigma(u)\mathrm{d}u = \int_0^x \left\{e^{-u}\psi(e^u) - 1\right\} \mathrm{d}u = \int_1^z \left\{\frac{\psi(t)}{t^2} - \frac{1}{t}\right\} \mathrm{d}t = \mathrm{O}(1).$$

To prove the lemma it suffices to show that the integral is $\mathrm{O}(1)$, so we note that

$$\int_{x_1}^{x_2} \sigma(u)\mathrm{d}u = \int_0^{x_2} - \int_0^{x_1} \sigma(u)\mathrm{d}u = \mathrm{O}(1).$$ □

Lemma 3.4. *If $\sigma(u_0) = 0$ for some $u_0 > 0$ then* $\int_0^C |\sigma(u_0 + t)| \,\mathrm{d}t \leqslant \frac{C^2}{2} + \mathrm{O}(u_0^{-1})$.

Proof. We rewrite Selberg's formula (Theorem (3.1)) as

$$\psi(x)\log x + \sum_{mn \leqslant x} \Lambda(m)\Lambda(n) = 2x\log x + \mathrm{O}(x).$$

If $x > x_0 \geqslant 1$, the same holds for x_0 in place of x. Subtracting the two yields

$$\psi(x)\log x - \psi(x_0)\log x_0 + \sum_{x_0 < mn \leqslant x} \Lambda(m)\Lambda(n) = 2\left(x\log x - x_0\log x_0\right) + \mathrm{O}(x).$$

Since $\Lambda(n) \geqslant 0$, we have $0 \leqslant \psi(x)\log x - \psi(x_0)\log x_0 \leqslant 2\left(x\log x - x_0\log x_0\right) + \mathrm{O}(x)$. This implies $|R(x)\log x - R(x_0)\log x_0| \leqslant x\log x - x_0\log x_0 + \mathrm{O}(x)$. Put $x = e^{u_0+t}$, $x_0 = u_0$ so that $R(x_0) = 0$. Then for $0 \leqslant t \leqslant C$, we have

$$\begin{aligned} |\sigma(u_0+t)| &\leqslant 1 - \left(\frac{u_0}{u_0+t}\right)e^{-t} + \mathrm{O}\left(\frac{1}{u_0}\right) \\ &= 1 - e^{-t} + \mathrm{O}\left(\frac{1}{u_0}\right) \leqslant t + \mathrm{O}\left(\frac{1}{u_0}\right). \end{aligned}$$

Hence, $\displaystyle\int_0^C |\sigma(u_0+t)|\mathrm{d}t \leqslant \int_0^C t\mathrm{d}t + \mathrm{O}(u_0^{-1}) = \frac{C^2}{2} + \mathrm{O}(u_0^{-1}).$ □

Now let $\delta = \dfrac{3C^2 + 4A_1}{2C} > C > 0$ and let $y > 0$ be arbitrary. We study the behaviour of $\sigma(u)$ on the interval $[y, y+\delta-C]$. By its definition, $\sigma(u) = e^{-u}\psi(e^u) - 1$ is monotone increasing only at the jump discontinuities $u = \log p^k$ where it increases by $\log p$ and between any two jump discontinuities $\sigma(u)$ decreases monotonically as $\psi(e^u)$ remains constant whilst e^{-u} decreases. This means that either $\sigma(u)$ vanishes at some point $u = u_0$ or $\sigma(u)$ changes sign at most once.

Case I: As $\sigma(u_0) = 0$ for some $u_0 \in [y, y+\delta-C]$, we use (3.8) and Lemma (3.4) to obtain

$$\int_y^{y+\delta} |\sigma(u)|\mathrm{d}u = \int_y^{u_0} + \int_{u_0}^{u_0+C} + \int_{u_0+C}^{y+\delta} |\sigma(u)|\mathrm{d}u$$

$$\begin{aligned}&\leqslant C(u_0 - y) + \frac{C^2}{2} + C\big(y + \delta - u_0 - C\big) + \mathrm{o}(1)\\&= C\left(\delta - \frac{C}{2}\right) = \mathrm{o}(1) = C'\delta + \mathrm{o}(1)\end{aligned}$$

for all y sufficiently large, where we took $C' = C\left(1 - \dfrac{C}{2\delta}\right) < C$.

<u>Case II:</u> If $\sigma(u)$ changes sign exactly once at some point $u = u_1 \in [y, y + \delta - C]$, then by Lemma (3.3)

$$\int_y^{y+\delta-C} |\sigma(u)|\mathrm{d}u = \left|\int_y^{u_1} \sigma(u)\mathrm{d}u\right| + \left|\int_{u_1}^{y+\delta-C} \sigma(u)\mathrm{d}u\right| < 2A_1.$$

If $\sigma(u)$ does not change sign at all in the interval, then by Lemma (3.3) again

$$\int_y^{y+\delta-C} |\sigma(u)|\mathrm{d}u = \left|\int_y^{y+\delta-C} \sigma(u)\mathrm{d}u\right| < A_1 < 2A_1.$$

Hence,

$$\begin{aligned}\int_y^{y+\delta} |\sigma(u)|\mathrm{d}u &= \int_y^{y+\delta-C} + \int_{y+\delta-C}^{y+\delta} |\sigma(u)|\mathrm{d}u\\&< 2A_1 + \int_{y+\delta-C}^{y+\delta} |C + \mathrm{o}(1)|\mathrm{d}u\\&= 2A_1 + C^2 + \mathrm{o}(1) = C''\delta + \mathrm{o}(1)\end{aligned}$$

where we took $C'' = \dfrac{2A_1 + C^2}{\delta} = C\left(\dfrac{4A_1 + 2C^2}{4A_1 + 3C^2}\right) = C\left(1 - \dfrac{C}{2\delta}\right) = C'$.

In both cases we always have

$$\int_y^{y+\delta} |\sigma(u)|\mathrm{d}u \leqslant C'\delta + \mathrm{o}(1).$$

If $M = [x/\delta]$, then

$$\begin{aligned}\int_0^x |\sigma(u)|\mathrm{d}u &= \sum_{m=0}^{M-1} \int_{m\delta}^{(m+1)\delta} |\sigma(u)|\mathrm{d}u + \int_{M\delta}^x |\sigma(u)|\mathrm{d}u\\&\leqslant C'M\delta + \mathrm{o}(M) + \mathrm{O}(1) = C'x + \mathrm{o}(x).\end{aligned}$$

Hence,

$$K = \limsup_{x\to\infty} \frac{1}{x}\int_0^x |\sigma(u)|\mathrm{d}u \leqslant C' < C.$$

An absurdity. Hence, we must have $C = 0$. This proves the PNT in Chebyshev form.

4. Further Generalisations

It is possible to generalise Theorem (2.2) to prime ideals in a number field. Let K be a number field and $\pi_K(X)$ denote the number of prime ideals in K of norm at most X,

$$\pi_K(X) = \#\{\mathfrak{p} : \mathrm{N}_K(\mathfrak{p}) \leqslant X\}.$$

In 1903, Landau [Lan03] proved the following with an asymptotic analogous to the PNT

Theorem 4.1 (Prime Ideal Theorem)**.**

$$\pi_K(X) \sim \frac{X}{\log X}.$$

Landau's original proof involves complex analysis and properties of the Riemann ζ-function, but an elementary proof in the spirit of Selberg's proof of the PNT is also possible as shown by Shapiro [Sha49] in 1949.

References

[Apo89] Tom M. Apostol. *Introduction to Analytic Number Theory*. Springer International Student Edition. Springer-Verlag, 1989.

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

[Lan03] Edmund Landau. "Neuer Beweis des Primzahlsatzes und Beweis des Primidealsatzes". In: *Mathematische Annalen* 56.4 (1903).

[Sel49] Atle Selberg. "An Elementary Proof of the Prime-Number Theorem". In: *Annals of Mathematics* 50.2 (1949).

[Sha49] Harold Shapiro. "An elementary proof of the prime ideal theorem". In: *Communications on Pure and Applied Mathematics* 2.4 (1949).

[TI51] Tikao Tatuzawa and Kanesiroo Iseki. "On Selberg's elementary proof of the prime-number theorem". In: *Proceedings of the Japan Academy, Series A, Mathematical Sciences* 27.7 (1951).

Sayan Das
Jadavpur University
dassayan0013@gmail.com