An Elementary Proof of the Prime Number Theorem

Nicholas Brusilow, William Dudarov, James Yan

Spring 2021

1 Introduction

The Prime Number Theorem (PNT) is one of the foundational results in the mathematical subfield of analytic number theory, first conjectured by Gauss and Legendre, with partial complex-analytic progress made by Riemann, who tied this conjecture to behavior of what we now call the Riemann Zeta Function, and with partial progress made by Chebyshev, who used properties of the binomial coefficients $\binom{2n}{n}$ to show that if indeed the limit in the conjecture (in Theorem 1) exists, then it must converge to 1.[3] Much to many mathematicians' surprise, Atle Selberg and Paul Erdős independently produced proofs of the PNT which did not require any complex analysis, but their disputes over the attribution of the proof and whether the proof should be published jointly led to bad blood between Selberg and Erdős which would continue throughout their lives. Selberg had proved an asymptotic formula from which Erdős quickly derived the PNT while Selberg was abroad for a short while.[2]

Our goal is to demonstrate a proof the Prime Number Theorem through purely elementary means (i.e. not involving the methods of complex analysis). This paper is based on Levinson's wonderful 1969 publication in the American Mathematical Monthly, and the goal of this paper is to provide an exposition to Levinson's proof, expanding on some especially important or interesting lemmas of the proof.

Recall that the Prime Number Theorem states:

Theorem 1 (The Prime Number Theorem)**.**

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\log x} = 1,$$

where

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

for p prime, is the prime counting function.

There are equivalent statements of the Prime Number Theorem that are easier to work with, and in particular, we will be working with the equivalent statement, whose equivalence to the PNT is proved in Theorem 4.4 in Apostol:

Theorem 2 (The Prime Number Theorem, equivalent statement)**.**

$$\lim_{x\to\infty} \frac{\psi(x)}{x} = 1,$$

where $\psi(x) = \sum_{j\leq x} \Lambda(j)$, and $\Lambda(n) = \log p$ if and only if $n = p^i$ for some prime number p.

2 Setup

Let's consider Selberg's result which was known to both Selberg and Erdős when they were both working towards finding the elementary proof of the PNT.

The main result we wish to prove in this section is Selberg's asymptotic formula, which is covered as Theorem 4.18 in Section 4.11 in Apostol's fantastic text. Here we will call it a lemma, since it is an important stepping stone in the elementary proof of the Prime Number Theorem.

Lemma 3. *For $x > 0$, we have*

$$\psi(x)\log x + \sum_{n\le x}\Lambda(n)\psi\left(\frac{x}{n}\right) = 2x\log x + O(x). \tag{1}$$

Although Selberg did not have it at his disposal, in order to prove this important lemma more easily we can first prove the following result, which is an inversion-type formula of Tatuzawa and Iseki proven in 1951, after Selberg's and Erdős's proof had already been established:

Lemma 4. *For F a real- or complex-valued function defined on $(0,\infty)$, and*

$$G(x) = \log x \sum_{n\le x} F\left(\frac{x}{n}\right),$$

we have

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

Proof. Consider $F(x)\log(x)$.

We can write this expression as a sum:

$$F(x)\log x = \sum_{n\le x}\left[\frac{1}{n}\right]F\left(\frac{x}{n}\right)\log\frac{x}{n}.$$

Notice that because of the $\left[\frac{1}{n}\right]$ term, the only nonzero term is when $n = 1$.

Notice that

$$\sum_{n\le x}\left[\frac{1}{n}\right] = \sum_{n\le x}\sum_{d|n}\mu(d),$$

by Theorem 3.12 in Apostol, which we covered in our readings, so our expression is

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

Shifting gears to the other term in the sum we are trying to prove: with help from Theorem 2.11 in Apostol, which states that

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

we define

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

Then we add our $F(x)\log x$ terms to $\sum_{n\le x} F(x/n)\Lambda(n)$, we get:

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

Swapping the sums, this is equivalent to

$$\sum_{d\le x}\mu(d)\log\frac{x}{d}\sum_{q\le\frac{x}{d}} F\left(\frac{x}{qd}\right) = \sum_{d\le x}\mu(d)G\left(\frac{x}{d}\right),$$

by the definition of G, as desired.

□

We can now use this result to prove Selberg's asymptotic formula, lemma 3.

Proof. To use lemma 4 to prove Selberg's formula, we must first specify our real- or complex-valued function F and its respective G.

Consider $F = \psi(x) - x + C$, where C is a constant.

Then

$$\begin{aligned}
G &= \log x \sum_{n\le x}\psi\left(\frac{x}{n}\right) - \frac{x}{n} + C\\
&= \log x\left(\sum_{n\le x}\psi\left(\frac{x}{n}\right) - \sum_{n\le x}\frac{x}{n} + \sum_{n\le x} C\right)\\
&= \log x\left(x\log x - x + O(\log x) - \sum_{n\le x}\frac{x}{n} + \sum_{n\le x} C\right)
\end{aligned}$$

by Theorem 4.11 in Apostol, and furthermore,

$$\begin{aligned}
G &= x\log^2 x - x\log x + O(\log^2 x) - \log x\sum_{n\le x}\frac{x}{n} + \log x\sum_{n\le x} C\\
&= x\log^2 x - x\log x + O(\log^2 x) - x\log x\sum_{n\le x}\frac{1}{n} + C\log x\sum_{n\le x} 1\\
&= x\log^2 x - x\log x + O(\log^2 x) - x\log x\left(\log x + C + O\left(\frac{1}{x}\right)\right) + C\log x\sum_{n\le x} 1
\end{aligned}$$

by Theorem 3.2(a) in Apostol.

Turning G into an easy-to-work with big-O asymptotic term,

$$\begin{aligned}
G &= x\log^2 x - x\log x + O(\log^2 x) - x\log^2 x + Cx\log x + O(\log x) + C\log x\sum_{n\le x} 1 \\
&= (C-1)x\log x + O(\log^2 x) + C\log x\sum_{n\le x} 1 \\
&= O(\log^2 x) \\
&= O(\sqrt{x})
\end{aligned}$$

With this in mind, applying lemma 4 to this F and G yields

$$\begin{aligned}
(\psi(x) - x + C)\log x + \sum_{n\le x}\left(\psi\left(\frac{x}{n}\right) - \frac{x}{n} + C\right)\Lambda(n) &= \sum_{d\le x}\mu(d)O(\sqrt{\frac{x}{d}}) \\
&= O\left(\sum_{d\le x}\sqrt{\frac{x}{d}}\right) \\
&= O(x).
\end{aligned}$$

By Theorem 4.9 in Apostol, we get that

$$\begin{aligned}
O(x) &= (\psi(x) - x + C)\log x + \sum_{n\le x}\left(\psi\left(\frac{x}{n}\right) - \frac{x}{n} + C\right)\Lambda(n) \\
O(x) &= \psi(x) - x\log x + C\log x + \sum_{n\le x}\psi\left(\frac{x}{n}\right)\Lambda(n) - \frac{x}{n}\Lambda(n) + C\Lambda(n) \\
O(x) + x\log x - C\log x &= \psi(x)\log x + \sum_{n\le x}\psi\left(\frac{x}{n}\right)\Lambda(n) - \frac{x}{n}\Lambda(n) + C\Lambda(n) \\
O(x) + (x-c)\log x &= \psi(x)\log x + \sum_{n\le x}\psi\left(\frac{x}{n}\right)\Lambda(n) - \sum_{n\le x}\frac{x}{n}\Lambda(n) + C\Lambda(n) \\
(x-C) + \sum_{n\le x}\left(\frac{x}{n} - C\right)\Lambda(n) + O(x) &= \psi(x)\log x + \sum_{n\le x}\psi\left(\frac{x}{n}\right)\Lambda(n) \\
2x\log x + O(x) &= \psi(x)\log x + \sum_{n\le x}\psi\left(\frac{x}{n}\right)\Lambda(n)
\end{aligned}$$

as desired. □

Now that we have put a lot of work into proving Selberg's formula, we are in the same position Erdős and Selberg were back in the summer of 1948. How could we take what we know so far to prove the PNT without invoking complex analysis? As we shall demonstrate, the "elementary" proof is quite complicated, specifically because it can't invoke these powerful tools from complex analysis, but we will attempt make the rest of this paper a guide to navigating the logical steps from Selberg's asymptotic formula to a complete elementary proof of the PNT.

3 Proof of the Prime Number Theorem

Philosophy

As we have seen earlier, in order to prove the Prime Number Theorem, we can consider its equivalent statement $\lim_{x\to\infty} \frac{\psi(x)}{x} = 1$. With some manipulations,

$$\lim_{x\to\infty} \frac{\psi(x)}{x} = 1 \Leftrightarrow \lim_{x\to\infty} \frac{\psi(x)}{x} - 1 = 0 \Leftrightarrow \lim_{x\to\infty} \frac{\psi(x) - x}{x} = 0.$$

Set $R(x) = \psi(x) - 1$, it is sufficient to prove

$$\lim_{x\to\infty} \frac{R(x)}{x} = 0. \tag{2}$$

Also, we can replace $\psi(x)$ in Selberg's result (1) by $R(x)$,

$$R(x)\log x + \sum_{n\leq x} \Lambda(n)R(x/n) = O(n). \tag{3}$$

However, (2) is a long way from Selberg's result (3). For one, the weights $\Lambda(x)$ in the weighted sum in (3) depends on the distribution of the prime numbers, which is precisely what we are trying to answer. Thus, in order to use this result, we need to make it more tractable. The way we do it will be loosening up the equality by some "smoothing operations" in hoping to gain more tangibility over this result. But we will do it in a conservative manner because the "smoothing operations" are lossy, and we need to keep essential information in the inequality under each operation.
Also, working with $R(x)$ is restrictive, as it is not continuous – we see discontinuous "jumps" starting at points p^a where $R(x) = \psi(x) - x = \psi(p^i) - [x]$ where p is prime p^i is the largest such number at more x. To assist our smoothing operation, we need to make it more workable. As we will see, defining $S(x) = \int_2^y \frac{R(x)}{x} dy$ and $W(x) = e^{-x}S(e^x)$ help smoothing (3). As a consequence of equation 3

$$S(y)\log y + \sum_{j\leq y} \Lambda(j)S\left(\frac{y}{j}\right) = O(y). \tag{4}$$

and we shift our focus accordingly to showing

$$\lim_{y\to\infty} \frac{S(y)}{y} = 0 \tag{5}$$

and

$$\limsup_{x\to\infty} W(x) = 0. \tag{6}$$

The relation between equations 2, 5, and 6 will become clear as we tie them up at the end of the paper. We will first state 3 lemmas which will be useful in our derivation of the PNT.

Lemma 5.

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

Remark. *This result is Theorem 4.9 in Apostol, and the result is obtained by using Shapiro's theorem with* $a(n) = \Lambda(n)$, *but we did not cover Shapiro's theorem in this course, and will simply assume this result.*

Lemma 6. *Define* $\Lambda_2(n) = \Lambda(n)\log n + \sum_{jk=n} \Lambda(j)\Lambda(k)$. *Then*

$$\frac{1}{x}\sum_{n\leq x} \Lambda_2(n) = 2\log x + O(1).$$

Remark. *This lemma tells us that on average, $\Lambda_2(n)$ acts like $2\log n$. We omit the proof in this paper, but one can check the fact that the sum over $n \leq x$ of $\Lambda(n)\log n$ equals $\psi(x)\log x + O(1)$ by Euler summation, and the rest can be derived by applying Selberg's asymptotic formula. This is a great exercise to tackle for the student reader.*

We also assume the following result from Levinson's paper, without proof. See the Levinson paper for more detail.

Lemma 7.

$$| S(y) | \leq cy$$

$$y \geq 2$$

and

$$| S(y_2) - S(y_1) | \leq c \mid y_1 - y_2 \mid .$$

Improving on equation 4

Lemma 8. *With $\Lambda_2(n) = \Lambda(n) + \sum_{ij=n} \Lambda(i)\Lambda(j)$, and K_1 a constant*

$$\log^2(y) \mid S(y) \mid \leq \sum \Lambda_2(m) \mid S(y/m) \mid + K_1 y \log y.$$

Proof. First we relabel y in equation 4 by $\frac{y}{k}$, multiply by $\Lambda(k)$, and sum over $k \leq y$,

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

Let's consider the second sum here. We can combine the double sum into a single sum of all pairs of (j,k) where $jk \leq y$ since in $S\left(\frac{y}{jk}\right), jk \leq y$. Use lemma 5 on the right side; let $jk = m$ in the second sum,

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

where the second double sum holds because exactly the pairs (j,k) such that $jk = m \leq y$ are considered. Finally, $\log\frac{y}{k} = \log y - \log k$ so the first sum can be broken up to two sums. We relabel k by m the first sum becomes $\sum_{k\leq y} \Lambda(k) S\left(\frac{y}{k}\right) \log y - \sum_{m\leq y} \Lambda(m) S\left(\frac{y}{m}\right) \log m$. We substitute this in the equation to get

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

Observe that the first sum above appears in equation 4. We have

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

where $O(y)$ is absorbed into $O(y\log y)$. Lastly, with $\Lambda_2(n) = \Lambda(n)\log n + \sum_{jk=n} \Lambda(j)\Lambda(k)$ as in (4.14), we reach

$$\log^2(y) \mid S(y) \mid \leq \sum \Lambda_2(m) \mid S(y/m) \mid + K_1 y \log y$$

if we consider the magnitude of all terms in the sums. □

Remark. *The weights $\Lambda(n)$ pose a challenge when we try to derive PNT from equation 4. Lemma 6 tells us on average $\Lambda_2(n)$ acts like $2\log n$, which is more manageable. Thus, we should try to replace $\Lambda_2(m)$ with $2\log m$.*

Lemma 9. *There is a constant K_2 such that*

$$\log^2 y \mid S(y) \mid \leq 2 \sum \mid S(y/m) \mid \log m + K_2 y \log y.$$

Remark. *We omit the proof here. But note that we are able to replace $\Lambda_2(m)$ by $2 \log y$, and the error term is still of $O(y \log y)$.*

Lemma 10. *There is a constant K_4 such that*

$$\log^2 y \mid S(y) \mid \leq 2 \int_2^y \mid S(y/u) \mid \log u du + K_3 y \log y.$$

Proof. Notice that the logarithmic function $\log u$ is increasing,

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

On the other hand, $|S(y/m)| - |S(y/u)| \leq |S(y/m) - S(y/u)|$ by the Triangle Inequality. The right hand side can be broken up to two parts:

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

Set the last integral as J_m. Apply Lemma 6 to it:

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

in which the last inequality is also because $\log u$ is increasing. Moreover, use the Taylor expansion, we find $\log(m+1) = \sum_{i=1}^{\infty} (-1)^{n+1} \frac{m^n}{n} \leq ((-1)^{1+1} \frac{m^1}{1} = m$ the first term. We reach

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

Combining this inequality with Lemma 9,

$$\log^2 y |S(y)| \leq 2 \sum_{m \leq y} \left(\int_m^{m+1} \log u |S(y/u)| du + \frac{cy}{m+1} \right) + K_2 y \log y.$$

Notice that $S(y) = 0, y < 2$ and $m, m+1 \leq y$,

$$\log^2 y |S(y)| \leq 2 \left(\int_2^3 \log u |S(y/u)| du + \cdots + \int_{y-1}^{y} \log u |S(y/u)| du \right) + K_3 y \log y$$

with $K_3 = K_2 + c$.
By continuity of $\log m |S(y/m)|$ we can combine the integrals and complete the proof.

□

Bounding α

Lemma 11.

$$\alpha = \limsup_{x \to \infty} \mid W(x) \mid,$$

$$\gamma = \limsup_{x \to \infty} \frac{1}{x} \int_0^x \mid W(\zeta) \mid d\zeta;$$

then $\alpha \leq 1$ and $\alpha \leq \gamma$

Remark. *Recall that $W(x) = e^{-x}S(e^x)$. Our ultimate goal is $\alpha = 0$. It will be justified later this is enough to show PNT.*
Here we first show that α is bounded by constant 1 and also by γ. α, limit superior of $|W(x)|$, provides an upper bound of $|W(x)|$; $\alpha < \gamma$ tells us that the limit superior[1] of $|W(x)|$ is bounded by that of the average of $W(x)$. After being modified in Lemma 8,9, and 10, Selberg's inequality is finally used here.

Lemma 12. *If $k = 2c$ then*

$$| W(x_2) - W(x_1) | \leq k \mid x_2 - x_1 \mid,$$

and hence

$$|| W(x_2) | - | W(x_1) | \leq k \mid x_2 - x_1 \mid .$$

Remark. *This is just an analogue of lemma 7 for $W(x)$.*

Lemma 13. *If $W(v) \neq 0$ for $v_1 < v < v_2$, then there exists a number M such that*

$$\int_{v_1}^{v_2} | W(v) | \, dv \leq M,$$

$W(v) \neq 0$, $v_1 < v < v_2$.

Remark. *The proof of this involves showing that the integral (sans the absolute value) is bounded; the hypothesis that no zeroes exist within the bounds allows us to do this.*
The fact that $|R(x)|$ is bounded for finite x allows us to show the expression

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

from which the lemma follows.

Lemma 14. *A function $W(x)$ subject to the three conditions (which are the result of earlier lemmas):*
That

$$\lim_{x\to\infty} \sup |W(x)| \leqslant \lim_{x\to\infty} \sup \frac{1}{x}\int_0^x |W(z)| dz, \text{ or that}$$
$$\alpha \leqslant \gamma$$

from the definition that $\alpha = \lim_{x\to\infty} \sup |W(x)|$ and $\gamma = \lim_{x\to\infty} \sup \frac{1}{x}\int_0^x |W(z)|dz$;
That

$$||W(x_1)| - |W(x_2)|| \leqslant k|x_1 - x_2|$$

for some constant k independent of x_1 and x_2;
And that

$$\int_{v_1}^{v_2} |W(v)| dv \leqslant M$$

for some constant M independent of v_1 and v_2 so long as $W(v) \neq 0$ for $v_1 < v < v_2$, must in fact have $\alpha = 0$.

Proof. The proof of this lemma will be broken into cases depending on the zeroes of the function $W(x)$. First, we allow $\beta > \alpha$. From the definition of α, there exists an x_β such that for all $x \geq x_\beta$,

$$|W(x)| \leq \beta.$$

Case O: If $W(x) \neq 0$ for all $x \geq x_\beta$, then it follows from the fact that there exists a bound M so that

$$\lim_{x\to\infty} \int_{x_b}^{x} |W(v)| dv \leq M.$$

[1] For a bounded real sequence (x_n), define sequence (y_k), $y_k = \sup\{x_n, n \geq k\}$. Then $\limsup_{n\to\infty} x_n = \lim_{k\to\infty}(y_k)$.

The fact that the integral remains bounded for $x \to \infty$ tells us that $|W(x)| \to 0$ as $x \to \infty$ in this case. If it were not so, the integral could not possibly be bounded.

For the rest of the cases, we assume that $W(x)$ has arbitrarily large zeroes. Let $b > a > x_\beta$ such that a and b are successive zeroes of $W(x)$.

Case 1: $(b-a) \geq \frac{2M}{\beta}$, and thus

$$M \leq \frac{1}{2}(b-a)\beta.$$

Taking a and b as values between which $|W|$ is always positive,

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

Case 2: $(b-a) \leq \frac{2\beta}{k}$. We remark, that for $x \in (a,\, b)$, from the second property that

$$||W(a)| - |W(x)|| \leq k(x-a); |W(x)| \leq kx - ka,$$

or that for such an x, the graph of $|W(x)|$ cannot rise above the line $kx - ka$, and that even if it rises as fast as possible to the right of a (the same applies as rising as fast as possible to the left of b), the greatest value it can attain, traveling exactly half of the interval $[a,\, b]$, so traveling at most distance β/k at slope k, attaining a value of β (which itself is the upper bound of $|W(x)|$ here, and then falling as fast as it must down to b, the whole integral must have an area less than or equal to the triangle with base length $(b-a)$ and height β, that is,

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

Case 3: $\frac{2\beta}{k} < (b-a) < \frac{2M}{\beta}$. Reasoning as in Case 2, but this time with a trapezoid with sides sloping up to k, but then a middle section of width $(b-a-\frac{2\beta}{k})$ and height β, in this case we can establish the bound that

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

because from the assumption of Case 3, it is true that

$$\frac{\beta}{k(b-a)} \geq \frac{\beta^2}{2Mk}.$$

Because $\beta > \alpha$ now,

$$\int_a^b |W(x)|dx < (b-a)\beta\left(1 - \frac{\alpha^2}{2Mk}\right).$$

Because $\alpha \leq 1$, and $Mk > 1$, it must be the case that

$$\frac{1}{2} \leq \left(1 - \frac{\alpha^2}{2Mk}\right),$$

so this Case 3 bound on the integral is valid in Cases 1 and 2 as well.

We remark that, while this bound was established explicitly for *successive* zeroes of the W function, the fact that $(b-a)$ is a factor in the bound means it holds for arbitrary zeroes a and b with $a<b$. Let $a<a_1<a_2<\cdots<a_n<b$ be the successive zeroes between a and b. Then, allowing $B=\beta\left(1-\frac{\alpha^2}{2Mk}\right)$ for clarity:

$$\begin{aligned}\int_a^b |W(x)|dx &= \int_a^{a_1} |W(x)|dx + \int_{a_1}^{a_2} |W(x)dx + \cdots + \int_{a_n}^{b} |W(x)|dx \\ &< (a_1-a)B + (a_2-a_1)B + \cdots + (b-a_n)B \\ &= a_1B - aB + a_2B - a_1B + \cdots + bB - a_nB \\ &= (b-a)B.\end{aligned}$$

Having established this bound for all cases in which $W(x)$ has arbitrarily large zeroes, we let x_1 be the first zero to the right of x_β, and for a variable y, let $\overline{x}$ be the largest zero to the left of y. Then by the established bounds,

$$\begin{aligned}\int_0^y |W(x)|dx &= \int_0^{x_1} |W(x)|dx + \int_{x_1}^{\overline{x}} |W(x)|dx + \int_{\overline{x}}^{y} |W(x)|dx \\ &\le \int_0^{x_1} |W(x)|dx + (\overline{x}-x_1)\beta\left(1-\frac{\alpha^2}{2Mk}\right) + M.\end{aligned}$$

By the definition, $\overline{x}\le y$, and thus $(\overline{x}-x_1)\le y$, so

$$\frac{(\overline{x}-x_1)}{y}\le 1.$$

We now divide through the larger inequality by y and establish another inequality from this one involving 1 we just proved;

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

We now let $y\to\infty$, and from the definition of γ,

$$\gamma\le\beta\left(1-\frac{\alpha^2}{2Mk}\right).$$

Because $\alpha\le\gamma$, and $\beta=\alpha+\epsilon$ for some $\epsilon>0$,

$$\alpha\le(\alpha+\epsilon)\left(1-\frac{\alpha^2}{2Mk}\right).$$

Manipulating:

$$\begin{aligned}\alpha &\le \alpha-\frac{\alpha^3}{2Mk}+\epsilon\left(1-\frac{\alpha^2}{2Mk}\right); \\ \frac{\alpha^3}{2Mk} &\le \epsilon\left(1-\frac{\alpha^2}{2Mk}\right).\end{aligned}$$

We see from this that, because this inequality holds for all values of $\epsilon>0$, it is impossible that $\alpha^3>0$, therefore $\alpha^3\le 0$, but we know that $\alpha\ge 0$, therefore $\alpha=0$, as desired. $\square$

Having concluded that $\lim_{x\to\infty}\sup|W(x)|=0$, and from the fact that

$$W(x)=\frac{S(e^x)}{e^x},$$

we conclude that

$$\frac{|S(y)|}{y} \to 0 \text{ as } y \to \infty.$$

Therefore, for some arbitrary $\epsilon > 0$, there exists a y which is large enough to satisfy

$$\frac{|S(y)|}{y} \leq \frac{1}{2}\epsilon^2,$$

and hence

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

This large y can also be chosen so that

$$|S(y(1+\epsilon))| + |S(y)| \leq \frac{1}{2}\epsilon^2 y(1+\epsilon) + \frac{1}{2}\epsilon^2 y.$$

Overmore,

$$\frac{1}{2}\epsilon^2 y(1+\epsilon) + \frac{1}{2}\epsilon^2 y = \frac{1}{2}\epsilon^2(2y + y\epsilon) = \epsilon^2 y + \frac{1}{2}\epsilon^3 y < \epsilon^2 y.$$

We remark that because

$$S(y(1+\epsilon)) - S(y) \leq |S(y(1+\epsilon))| + |S(y)|,$$

it must be true that

$$S(y(1+\epsilon)) - S(y) \leq \frac{1}{2}\epsilon^2(y(1+\epsilon) + y) < \epsilon^2 y.$$

Recall the definition that

$$S(y) = \int_2^y \frac{R(u)}{u} du,$$

where $R(u) = \psi(u) - u$. Therefore

$$S(y(1+\epsilon)) - S(y) = \int_2^{y(1+\epsilon)} \frac{R(u)}{u} du - \int_2^y \frac{R(u)}{u} du = \int_y^{y(1+\epsilon)} \frac{R(u)}{u} du,$$

and by the earlier inequality

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

Therefore simply by the definition of R,

$$\int_y^{y(1+\epsilon)} \frac{\psi(u) - u}{u} du < \epsilon^2 y.$$

We split up this integral and get

$$\int_y^{y(1+\epsilon)} \frac{\psi(u) - u}{u} du = \int_y^{y(1+\epsilon)} \frac{\psi(u)}{u} du - \int_y^{y(1+\epsilon)} 1 \, du.$$

Notice that $\psi(u)$ is nondecreasing and u is also obviously nondecreasing, so the first integral can be bounded below as such:

$$\frac{\psi(y)}{y(1+\epsilon)} \int_y^{y(1+\epsilon)} 1 \, du \leqslant \int_y^{y(1+\epsilon)} \frac{\psi(u)}{u} du,$$

and hence

$$\frac{\psi(y)}{y(1+\epsilon)}\int_y^{y(1+\epsilon)} 1\,du - \int_y^{y(1+\epsilon)} 1\,du = \left(\frac{\psi(y)}{y(1+\epsilon)} - 1\right)\int_y^{y(1+\epsilon)} 1\,du < \epsilon^2 y.$$

Evaluating this integral gives us that

$$\left(\frac{\psi(y)}{y(1+\epsilon)} - 1\right)(y + y\epsilon - y) = y\epsilon\left(\frac{\psi(y)}{y(1+\epsilon)} - 1\right) < \epsilon^2 y.$$

Manipulating more yields

$$\begin{aligned}
\frac{\psi(y)}{y(1+\epsilon)} - 1 &< \epsilon; \\
\frac{\psi(y)}{y(1+\epsilon)} &< 1+\epsilon; \\
\frac{\psi(y)}{y} &< (1+\epsilon)^2.
\end{aligned}$$

Having shown this inequality resulting from

$$S(y(1+\epsilon)) - S(y) < \epsilon^2 y,$$

we remark that an equivalent "beginning point" would have been

$$S(y) - S(y(1-\epsilon)) > -\epsilon^2 y,$$

and that the above steps could be repeated to yield

$$\frac{\psi(y)}{y} > (1-\epsilon)^2.$$

Therefore, because ϵ was chosen arbitrarily,

$$\lim_{y\to\infty} \frac{\psi(y)}{y} = 1,$$

which proves the Prime Number Theorem.

References

[1] Apostol, Tom M. (1976). Introduction to Analytic Number Theory. Undergraduate Texts in Mathematics. Springer.

[2] Goldfeld, Dorian. The Elementary Proof of the Prime Number Theorem: an Historical Perspective. `https://www.math.columbia.edu/~goldfeld/ErdosSelbergDispute.pdf`.

[3] Lemmermeyer, Franz. The Prime Number Theorem. `http://www.fen.bilkent.edu.tr/~franz/nt/cheb.pdf`.

[4] Levinson, Norman. A Motivated Account of an Elementary Proof of the Prime Number Theorem. (1969). The American Mathematical Monthly, Vol. 76, No. 3., pp. 225-245. `http://www.math.stonybrook.edu/~moira/mat331-spr10/papers/1969%20LevinsonA%20Motivated%20Account%20of%20an.pdf`.