

# Annotation: A discussion of the fundamental ideas behind Selberg's "Elementary proof of the prime-number theorem"

Steve Balady

August 11, 2006

## 1 Introduction

The study of prime numbers (non-unit numbers whose only proper divisors are units) is both ancient and important: the prime numbers form the core of number theory, necessary for many of the most fundamental results of algebra and cryptography; and it was Euclid who proved their first basic property, their infinitude. But the question of any sort of *pattern* to these numbers went unsolved for two millenia after Euclid. Shortly before 1800, Gauss and Legendre both stated (though they could not prove) versions of what is today called the prime number theorem, concerning how the primes are distributed among the integers. Gauss claimed that, for large  $x$ ,  $\int_2^x \frac{du}{\log u}$  is a good approximation to  $\pi(x)$ , where  $\pi(x)$  is the number of primes less than or equal to  $x$  and 'good approximation' is taken to mean that the percentage error of this estimate goes to 0 as  $x \rightarrow \infty$ ; Legendre suggested  $\frac{x}{\log x+B}$  for some numerical constant  $B$ . Today, we state the prime number theorem as

$$\lim_{x \rightarrow \infty} \frac{\pi(x)}{\frac{x}{\log x}} = 1. \quad (1.1)$$

In the 1850s, the first major progress on a proof of the prime number theorem was made when Chebyshev proved that the limit above must be equal to 1 if it exists at all, and also proved that  $\frac{x}{\log x}$  always lies between two positive constants. These results involved only methods of real variables; but in 1896, using the techniques of complex analysis, Hadamard and de la Vallée Poussin independently proved the prime number theorem. At the time, the proof seemed so thoroughly dependent on complex variables that it was doubted that any techniques of real analysis would be powerful enough to establish the result. But in 1946, Selberg and Erdős (working independently) gave proofs using only real variables, the so-called 'elementary' proofs. As it happens, Erdős published first, but his paper was dependent on a fundamental idea due to Selberg; this complication led to a bitter rivalry between Selberg and Erdős for years.

Who deserves the credit is not the point of this paper: instead, my goal is to organize the basic ideas behind Selberg's formula and to indicate how it can be used to prove the prime number theorem. I have called this paper an annotation because its key results follow those of Selberg [2]. My sources for the rather unfortunate history of this proof are Andrews [1] and Shapiro [3]; the basic ideas in this paper are due to Landau, Chebyshev, Möbius, and Selberg; the proofs that I give are usually syntheses of proofs in [1]-[3]. I take no credit for any of these ideas (a good thing, especially since they are all relatively well-known); instead, it is my hope that this organization of available information will help to make Selberg's proof [2] more readily comprehensible to the diligent mathematical tourist. Frankly, my main attemptive proof is that Selberg's 'elementary proof' is in fact both fairly elementary and well-motivated.

## 2 Asymptotics

Before we begin to study the properties of  $\pi(x)$ , we introduce some notation first proposed by Landau which will make our discussion significantly easier to follow.

1. If  $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 1$ , we call  $f$  asymptotic to  $g$  and write  $f(x) \sim g(x)$ .
2. If  $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 0$ , we write  $f(x) = o(g(x))$  (read " $f$  is little-oh of  $g$ ").
3. If  $\exists k \in \mathbb{R}$  s.t.  $\lim_{x \rightarrow \infty} \left| \frac{f(x)}{g(x)} \right| < k$ , we write  $f(x) = O(g(x))$  (read " $f$  is big-oh of  $g$ ").

If we write  $O(g(x))$  alone, we take it to stand for some function which is  $O(g(x))$ ; for example,  $O(1)$  stands for a function which is uniformly bounded, and  $\sum_{n \leq x} = O(x)$  if the various  $O(1)$  functions themselves are uniformly bounded.

In addition, we have

**Lemma 2.1.**

$$f(x) \sim g(x) \Leftrightarrow f(x) = g(x) + o(g(x)).$$

*Proof.* If  $f(x) = g(x) + o(g(x))$ , then  $\frac{f(x)}{g(x)} = 1 + \frac{o(g(x))}{g(x)}$ ; letting  $x \rightarrow \infty$  gives  $f(x) \sim g(x)$ . Conversely,  $f(x) \sim g(x) \Rightarrow \frac{f(x)}{g(x)} - 1 = o(1) \Rightarrow f(x) - g(x) = o(g(x))$ .  $\square$

We can now restate the prime number theorem as

$$\pi(x) \sim \frac{x}{\log x}, \tag{2.1}$$

and indeed this is the form in which it is most commonly expressed.

### 3 Equivalent formulations of the prime number theorem

It turns out that we have very little control over  $\pi(x)$  itself. To remedy this, we introduce first the von Mangoldt function  $\Lambda(x)$  and then study the asymptotics of the Chebyshev functions  $\theta(x)$  and  $\psi(x)$  which are related to  $\pi(x)$  but prove more susceptible to direct approximation because of the property  $\log x + \log y = \log xy$ . We note here that for the rest of the paper (though we consider this to be a good idea in general),  $p$  will always be taken to denote a prime.

$$\Lambda(x) \equiv \begin{cases} 0 & \text{if } x = 1 \\ \log p & \text{if } \exists \alpha \in \mathbb{N} \text{ s.t. } x = p^\alpha \\ 0 & \text{otherwise} \end{cases} \quad (3.1)$$

$$\theta(x) \equiv \sum_{p \leq x} \log p \quad (3.2)$$

$$\psi(x) \equiv \sum_{n \leq x} \Lambda(n) \quad (3.3)$$

Since  $\log(x)$  is an increasing function, we have the trivial estimates

$$\theta(x) = O(x \log x) \text{ and} \quad (3.4)$$

$$\psi(x) = O(x \log x). \quad (3.5)$$

Our first goal is to verify that  $\theta(x)$  and  $\psi(x)$  have any bearing on the prime number theorem.

**Theorem 3.1.** *The following expressions are equivalent:*

$$\pi(x) \sim \frac{x}{\log x} \quad (3.6)$$

$$\theta(x) \sim x \quad (3.7)$$

$$\psi(x) \sim x \quad (3.8)$$

*Proof.* First we show that (3.7)  $\Leftrightarrow$  (3.8).

$$\psi(x) = \sum_{p^\alpha \leq x} \log p = \sum_{\alpha} \sum_{p \leq x^{\frac{1}{\alpha}}} \log p. \quad (3.9)$$

Since  $2^\alpha \leq x$ , or  $\alpha \leq \frac{\log x}{\log 2}$ , (3.9) implies

$$\psi(x) = \sum_{\alpha=1}^{\frac{\log x}{\log 2}} \sum_{p \leq x^{\frac{1}{\alpha}}} \log p = \sum_{\alpha=1}^{\frac{\log x}{\log 2}} \theta(x^{\frac{1}{\alpha}}) = \theta(x) + \theta(x^{\frac{1}{2}}) + \theta(x^{\frac{1}{3}}) + \dots$$

Since  $\alpha \leq \frac{\log x}{\log 2}$ , we can trivially bound this sum, apply our weak approximation (3.4), and conclude that

$$\begin{aligned}\psi(x) &= \theta(x) + O(\theta(\sqrt{x}) \log x), \\ &= \theta(x) + O(\sqrt{x} \log^2 x), \\ \frac{\psi(x)}{x} &= \frac{\theta(x)}{x} + O\left(\frac{\log^2 x}{\sqrt{x}}\right).\end{aligned}\tag{3.10}$$

As  $x \rightarrow \infty$ ,  $\frac{\log^2 x}{\sqrt{x}} \rightarrow 0$ , so letting  $x \rightarrow \infty$  in (3.10),

$$\lim_{x \rightarrow \infty} \frac{\psi(x)}{x} = \lim_{x \rightarrow \infty} \frac{\theta(x)}{x},\tag{3.11}$$

from which it follows immediately that  $\psi(x) \sim x$  iff  $\theta(x) \sim x$ . Now we prove the other half of the theorem, that (3.6)  $\Leftrightarrow$  (3.7).

$$\begin{aligned}\theta(x) &= \sum_{p \leq x} \log p \leq \sum_{p \leq x} \log x = \log x \sum_{p \leq x} 1 = \pi(x) \log x, \\ \frac{\theta(x)}{x} &\leq \frac{\pi(x)}{\frac{x}{\log x}}.\end{aligned}\tag{3.12}$$

Fix  $\alpha \in [0, 1]$ . Then

$$\begin{aligned}\theta(x) - \theta(x^\alpha) &= \sum_{p \leq x} \log p - \sum_{p^\alpha \leq x} \log p \\ &= \sum_{x^\alpha \leq p \leq x} \log p \\ &\geq \alpha \log x \sum_{x^\alpha \leq p \leq x} 1,\end{aligned}$$

so that

$$\theta(x) - \theta(x^\alpha) \geq \alpha \log x (\pi(x) - \pi(x^\alpha)).\tag{3.13}$$

Using our weak approximation (3.4) on (3.13), and noting that  $\pi(x) = O(x^\alpha)$  for the same reason,

$$\begin{aligned}\theta(x) + O(x^\alpha \log x) &\geq \alpha \pi(x) \log x + O(x^\alpha \log x) \\ \theta(x) &\geq \alpha \pi(x) \log x + O(x^\alpha \log x) \\ \frac{\theta(x)}{x} &\geq \frac{\alpha \pi(x) \log x}{x} + O\left(\frac{\log x}{x^{1-\alpha}}\right).\end{aligned}\tag{3.14}$$

Now assume that  $\theta(x) \sim x$ . Then by (3.12),

$$\underline{\lim} \frac{\pi(x) \log x}{x} \geq 1,\tag{3.15}$$

Likewise, by (3.14),

$$1 \geq \alpha \overline{\lim} \frac{\pi(x) \log x}{x}; \quad (3.16)$$

this holds for all  $\alpha \in [0, 1]$ . Letting  $\alpha \rightarrow 1$  from below and applying (3.15),

$$1 \geq \overline{\lim} \frac{\pi(x) \log x}{x} \geq \underline{\lim} \frac{\pi(x) \log x}{x} \geq 1, \quad (3.17)$$

so that  $\lim_{x \rightarrow \infty} \frac{\pi(x) \log x}{x}$  exists and is equal to 1.

We proceed similarly if we assume that  $\pi(x) \sim \frac{x}{\log x}$  : (3.12) gives

$$\overline{\lim} \frac{\theta(x)}{x} \leq 1,$$

while (3.14) gives

$$\underline{\lim} \frac{\theta(x)}{x} \geq \alpha.$$

Taking  $\alpha \rightarrow 1$  and combining these last two expressions, we conclude that the limit exists and is 1.  $\square$

It follows from this theorem that it is enough to prove any of these three forms to establish the prime number theorem, and thus that Selberg is justified in stating (1.1) and (1.2) in [2]. In what follows we will mainly be concerned with  $\psi(x)$ , and so will be interested in the properties of  $\Lambda(x)$ . (We note that [2] proves that  $\theta(x) \sim x$ ; however, we feel that the properties of  $\Lambda(x)$  are more intuitive, and the theorems we prove in the following are almost identical to those in [2].) First, however, we take a detour to prove a series of formulas which will aid us in analyzing  $\Lambda(x)$ .

## 4 The Möbius Inversion Formula

Continuing our journey through the Greek alphabet, we first introduce the Möbius function  $\mu(x)$  and then prove a basic result about it.

$$\mu(n) \equiv \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } \exists p \text{ s.t. } p^2 | n \\ (-1)^r & \text{if } n = p_1 p_2 \dots p_r \text{ for } p_i \text{ distinct primes} \end{cases} \quad (4.1)$$

**Lemma 4.1.**

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n > 1. \end{cases}$$

*Proof.* If  $n = 1$ , this is trivial; we proceed by induction on the number of prime factors of  $n$ . If  $n$  has precisely one prime factor (i.e.,  $n = p^n$ ),

$$\sum_{d|n} \mu(d) = \mu(1) + \mu(p) + \mu(p^2) + \dots + \mu(p^n) = 1 - 1 + 0 + \dots + 0 = 0.$$

Now assume that the lemma is true for any number  $n'$  with  $k$  distinct prime factors; then any  $n$  with  $k + 1$  distinct prime factors is of the form  $n'p^n$  for  $(p, n') = 1$ . But then

$$\begin{aligned}\sum_{d|n} \mu(d) &= \sum_{d|n'} \mu(d) + \sum_{d|n'} \mu(pd) + \sum_{d|n'} \mu(p^2d) + \dots + \sum_{d|n'} \mu(p^nd) \\ &= \sum_{d|n'} \mu(d) - \sum_{d|n'} \mu(d) + 0 + \dots + 0 = 0.\end{aligned}$$

□

The Möbius function  $\mu(x)$  will be essential in analyzing the Chebyshev functions in light of the

**Theorem 4.2.** (*Möbius Inversion Theorem*) For any  $f(n), g(n)$ , the following expressions are equivalent:

$$f(n) = \sum_{d|n} g(d) \tag{4.2}$$

$$g(n) = \sum_{d|n} \mu(d) f\left(\frac{n}{d}\right) \tag{4.3}$$

*Proof.* If  $f(n) = \sum_{d|n} g(d)$ , then we have that

$$\begin{aligned}\sum_{d|n} \mu(d) f\left(\frac{n}{d}\right) &= \sum_{dd'=n} \mu(d) f(d') \\ &= \sum_{dd'=n} \mu(d) \sum_{e|d'} g(e) = \sum_{deh=n} \mu(d) g(e) = \sum_{eh'=n} g(e) \sum_{d|h'} \mu(d) = g(n),\end{aligned}$$

where the last step follows by applying (4.1) to  $\sum_{d|h'} \mu(d)$ .

Conversely, assuming that  $g(n) = \sum_{d|n} \mu(d) f\left(\frac{n}{d}\right)$ ,

$$\begin{aligned}\sum_{d|n} g(n) &= \sum_{d|n} \sum_{d'|d} \mu(d') f\left(\frac{d}{d'}\right) \\ &= \sum_{d'eh=n} \mu(d') f(e) = \sum_{eh'=n} f(e) \sum_{d'|h'} \mu(d) = f(n),\end{aligned}$$

with the last step following, as before, by applying (4.1) to  $\sum_{d'|h'} \mu(d)$ . □

Several results that follow depend on well-known asymptotic formulas for certain functions of  $\log x$ ; to prove these here would interrupt the flow of ideas. We state without proof these formulas here: excellent derivations are given in [3].

For some fixed constants  $C > 0$  and  $C' > 0$ ,

$$\sum_{n \leq x} \frac{1}{n} = \log x + C + O\left(\frac{1}{x}\right), \quad (4.4)$$

$$\sum_{n \leq x} \frac{\log n}{n} = \frac{1}{2} \log^2 x + C' + O\left(\frac{\log x}{x}\right), \quad (4.5)$$

$$\sum_{n \leq x} \frac{\log \frac{x}{n}}{n} = \frac{1}{2} \log^2 x + C \log x - C' + O\left(\frac{\log x}{x}\right), \quad (4.6)$$

$$\sum_{n \leq x} \log \frac{x}{n} = x + O(\log x). \quad (4.7)$$

We now apply the Möbius inversion formula to some specific sums which will be useful in proving our core result (6.1).

1. Take  $g(x) \equiv 1$ . Then, by (4.2),

$$f(x) = \sum_{n \leq x} 1 = [x] = x + O(1);$$

applying (4.2) and using that  $|\mu(n)| \leq 1$ , we obtain

$$\begin{aligned} 1 &= \sum_{n \leq x} \mu(n) \left( \frac{x}{n} + O(1) \right) = x \sum_{n \leq x} \frac{\mu(n)}{n} + \sum_{n \leq x} O(1) = \left( x \sum_{n \leq x} \frac{\mu(n)}{n} \right) + O(x), \\ x \sum_{n \leq x} \frac{\mu(n)}{n} &= O(x) \\ \sum_{n \leq x} \frac{\mu(n)}{n} &= O(1). \end{aligned} \quad (4.8)$$

2. Take  $g(x) \equiv x$ .

$$f(x) = \sum_{n \leq x} \frac{x}{n} = x \sum_{n \leq x} \frac{1}{n} = x \log x + Cx + O(1) \text{ (by (4.4))}.$$

Applying the inversion formula,

$$\begin{aligned} x &= \sum_{n \leq x} \mu(n) \left( \frac{x}{n} \log \frac{x}{n} + C \frac{x}{n} + O(1) \right) \\ &= x \sum_{n \leq x} \frac{\mu(n)}{n} \log \frac{x}{n} + Cx \sum_{n \leq x} \frac{\mu(n)}{n} + \sum_{n \leq x} O(1) \\ &= x \sum_{n \leq x} \frac{\mu(n)}{n} \log \frac{x}{n} + O(x) + O(x) \text{ (by (4.8))}, \end{aligned}$$

so that

$$\begin{aligned} x \sum_{n \leq x} \frac{\mu(n)}{n} \log \frac{x}{n} &= O(x), \\ \sum_{n \leq x} \frac{\mu(n)}{n} \log \frac{x}{n} &= O(1). \end{aligned} \tag{4.9}$$

3. Finally, take  $g(x) \equiv x \log x$ .

$$f(x) = \sum_{n \leq x} \frac{x}{n} \log \frac{x}{n} = x \sum_{x \leq n} \frac{\log \frac{x}{n}}{n} = \frac{x}{2} \log^2 x + Cx \log x - C'x + O(\log x) \text{ (by (4.6))}.$$

By the Möbius inversion formula (collecting terms in one step this time),

$$x \log x = \frac{x}{2} \sum_{n \leq x} \frac{\mu(n)}{n} \log^2 \frac{x}{n} + Cx \sum_{n \leq x} \frac{\mu(n)}{n} \log \frac{x}{n} - C'x \sum_{n \leq x} \frac{\mu(n)}{n} + \sum_{n \leq x} O\left(\log \frac{x}{n}\right).$$

Finally, applying (4.9) to the second sum, (4.8) to the third sum, and (4.7) to the fourth sum,

$$\begin{aligned} x \log x &= \frac{x}{2} \sum_{n \leq x} \frac{\mu(n)}{n} \log^2 \frac{x}{n} + O(x) + O(x) + O(x), \\ \sum_{n \leq x} \frac{\mu(n)}{n} \log^2 \frac{x}{n} &= 2 \log x + O(1). \end{aligned} \tag{4.10}$$

## 5 Properties of $\Lambda(x)$

We now return to our main goal: improving our asymptotic formula for  $\pi(x)$  (equivalently,  $\theta(x)$  or  $\psi(x)$ ). As we said, we will focus on  $\psi(x)$ : to do this we need the following two lemmas.

**Lemma 5.1.**

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

*Proof.* Letting  $n = \prod p_i^{a_i}$ , we have that

$$\log n = \sum a_i \log p_i.$$

On the other hand, there are precisely  $a_i$  divisors of  $n$  for which  $\Lambda(d) = \log p_i$ , so that

$$\sum_{d|n} \Lambda(d) = \sum a_i \log p_i = \log n.$$

□



The motivation for the next lemma, and indeed the rest of this paper, is the following. The prime numbers themselves are very difficult to work with, which is precisely why the prime number theorem eluded proof for a century after it was first hypothesized. Therefore, we will have a much better chance of proving statements about  $\pi(x)$  which do not explicitly require knowledge of the distribution of primes; Selberg's main insight was to construct an asymptotic formula which related a sum over  $\theta(p)$  to a function involving only  $\log(x)$  which allowed him to extend. We proceed in effectively the same manner to find a similar asymptotic formula relating  $\psi(x)$  (that is, a sum over  $\Lambda(n)$ ) to a function of  $\log(x)$ .

**Lemma 5.2.**

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

*Proof.*

$$\begin{aligned} \log^2 n &= (\log n)(\log n) = \log n \sum_{d|n} \Lambda(d) \text{ (by (5.1))} \\ &= \sum_{d|n} \Lambda(d) \log \frac{dn}{d} \\ &= \sum_{d|n} \Lambda(d) \log \frac{n}{d} + \sum_{d|n} \Lambda(d) \log d \\ &= \sum_{d|n} \Lambda(d) \sum_{d'|\frac{n}{d}} \Lambda(d') + \sum_{n|d} \Lambda(d) \log d \text{ (by (5.1))} \\ &= \sum_{dd'|n} \Lambda(d) \Lambda(d') + \sum_{d|n} \Lambda(d) \log d. \end{aligned}$$

Letting  $c = dd'$  and replacing  $d$  with  $c$  in the second sum,

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

Now, applying (4.1) to this sum with respect to  $c$ , we conclude that

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

Changing  $c$  back to  $d$  in the first sum, we obtain (5.2).  $\square$

## 6 The symmetry formula

Lemma 5.2 gives us fundamental knowledge about  $\Lambda(n)$  for individual  $n$ : it follows that if we sum this information over all  $n \leq x$ , we should obtain infor-

mation about  $\psi(n)$ . Selberg [2] calls attention to this as the “basic new thing in his proof”; [3] calls it the symmetry formula.

**Theorem 6.1.** (*The symmetry formula*)

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

*Proof.* Since (5.2) is true for all  $n$ , we may sum it over  $n \leq x$  to obtain

$$\sum_{n \leq x} \sum_{d|n} \mu(d) \log^2 \frac{n}{d} = \sum_{n \leq x} \Lambda(n) \log n + \sum_{n \leq x} \sum_{d|n} \Lambda(d) \Lambda\left(\frac{n}{d}\right). \quad (6.1)$$

Let  $n = mn'$ ; then the third term in (6.1) becomes

$$\sum_{mn' \leq x} \Lambda(m) \Lambda(n');$$

replacing  $n'$  with  $n$  gives us the second term in (6.1), so we need only show

$$\sum_{n \leq x} \sum_{d|n} \mu(d) \log^2 \frac{n}{d} = 2x \log x + O(x).$$

Letting  $n = dd'$ ,

$$\begin{aligned} \sum_{n \leq x} \sum_{d|n} \mu(d) \log^2 \frac{n}{d} &= \sum_{dd' \leq x} \mu(d) \log^2 d' \\ &= \sum_{d \leq x} \mu(d) \sum_{d' \leq \frac{x}{d}} \log^2 d' \\ &= \sum_{d \leq x} \frac{\mu(d)}{d} \log^2 \frac{x}{d} - 2x \sum_{d \leq x} \frac{\mu(d)}{d} \log \frac{x}{d} + 2x \sum_{d \leq x} \frac{\mu(d)}{d} + O\left(\sum_{d \leq x} \log^2 \frac{x}{d}\right), \end{aligned} \quad (6.2)$$

with the last step following by (4.5). Now, by (4.10), the first term is  $2x \log x + O(x)$ . By (4.9), the second term is  $O(x)$ . By (4.8), the third term is  $O(x)$ . And by (4.7), the last term is  $O(x)$ . Then, finally, (6.2) reduces to  $2x \log x + O(x)$ , which is what we needed.  $\square$

It can become easy to get bogged down in the specific mechanics of this proof, which is why the bulk of the heavy asymptotic lifting was done in Section 4. The main idea here should be clear: we have now expressed a sum over  $\Lambda(x)$  asymptotically in terms of  $\log(x)$ , which is a huge step forward. If we take for granted the formula in [3] that  $\sum_{n \leq x} \Lambda(n) \log n = \psi(x) \log x + O(x)$ , then we can express Theorem 6.1 in the following way, which corresponds almost exactly to (1.3) in [2]:

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

## 7 Toward the prime number theorem

With the symmetry formula established, a sketch of the remainder of Selberg's proof follows. First we define the error term  $R(x)$  by the equation  $\psi(x) = x + R(x)$ ; then the theorem is proved if we show that  $R(x) = o(x)$ , since this will imply by Lemma 2.1 that  $\psi(x) \sim x$ . We establish, relying heavily on the symmetry formula, the inequality

$$|R(x)| \log^2 x \leq 2 \sum_{n \leq x} \log n \left| R\left(\frac{x}{n}\right) \right| + O(x \log x) \quad (7.1)$$

which has the virtue of not relating directly to the primes and is consequently easier to deal with. From there, we use  $R(x)$  to find intervals farther than any arbitrary distance from 0 on which the prime number theorem is "approximately true:" that is, there exists a constant  $K$  such that for any  $\delta < 1$  and  $x > 4$ , there exists  $y \in [x, e^{\frac{K}{\delta}} x]$  for which

$$\left| \frac{R(y)}{y} \right| < \delta.$$

From there, we apply the fundamental inequality (7.1) iteratively over these "patches," each time giving us an estimate  $\delta_n$  such that  $\left| \frac{R(y)}{y} \right| < \delta_n$ , where  $\delta_i = \delta_{i-1} - a\delta_{i-1}^3$  for some fixed  $a > 0$ , so that  $\delta_i \rightarrow \beta$  for some  $\beta$ . Taking the limit as  $i \rightarrow \infty$ , we find that  $\beta = \beta - a\beta^3$ , and thus that  $\beta = 0$ , establishing the theorem.

## References

- [1] 1994. Andrews, George E. *Number Theory*. New York: Dover.
- [2] Selberg, A. "An elementary proof of the prime-number theorem." *The Annals of Mathematics*, 2nd Ser., Vol 50, No. 2. (Apr., 1949), pp. 205-313.
- [3] 1948. Shapiro, Harold N. *Lectures on the theory of numbers*. New York: New York University Institute for Mathematics and Mechanics.