The prime number theorem: Analytic and elementary proofs

Ciarán O'Rourke

Thesis submitted for the degree of
M.Sc.

Roinn na Matamaitice
Ollscoil na h-Éireann, Má Nuad
Feabhra 2013

Head of Department: Professor Stephen Buckley
Supervisor: Dr. Detta Dickinson

Abstract

Three proofs of the prime number theorem are presented. The first is a heavily analytic proof based on early accounts. Cauchy's residue theorem and various results relating to the Riemann zeta function play a vital role. A weaker result than the prime number theorem is used for the proof, namely Chebyshev's theorem. The second proof is elementary in the sense that it involves no complex analysis. Instead, mainly number-theoretic results are used, in particular, Selberg's formulas. The third proof, like the first, relies heavily on the Riemann zeta function, but is considerably shorter for the use of the Laplace transform and the analytic theorem.

Contents

Acknowledgements

To my family, your constant love and support during good times and bad gets me through. Thank you for believing in me, and for me at times. To my friends, I am glad to have every one of you. Thank you for your loyalty and integrity. And there are many other people from outside mathematics to whom I owe a debt of gratitude, too numerous to mention here.

I would like to thank two of my former teachers, Pádraig O'Súilleabháin, who first mentioned variables, and Fr. Canice Murphy, who, among other things, encouraged my interest in the Olympiad at Maynooth.

To all the staff of the Mathematics and Statistics Department, thank you for further fostering my interest. I would particularly like to thank the head of the Department, Professor Stephen Buckley, his predecessor, Professor Anthony O'Farrell and Dr. David Wraith, who was practically a mentor to me during my final undergraduate years. To Gráinne, Janice and Tony, I owe you so much more than I can ever give.

Thank you to the external examiner, Professor Simon Kristensen, and Dr. James O'Shea for their most helpful suggestions and comments. I hope that this thesis is the better for them.

Special mention also of Dr. Pat McCarthy and Dr. Ciarán Mac an Bhaird, two extraordinary gentlemen who supervised me during the early part of my thesis, and have continued to support me since then. But it is to Detta that I give my final word of thanks. Without your wisdom and guidance, this ship would surely have hit the rocks. I consider myself very fortunate to have had this experience, and have come to enjoy our meetings immensely. As a chapter closes, I trust that your door will remain open.

Thank you

Ciarán O'Rourke

Introduction

Since the dawn of mathematics, the prime numbers and their distribution have been a source of fascination. Over the years, people have sought, mainly in vain, for patterns among the primes. Motivating many of these attempts was a fundamental question: how many primes are there less than or equal to a given number X? With no discernible patterns, an exact answer, in the form of an expression involving X, is not available. However, it is natural to ask whether some form of approximation exists. Indeed, around 1800, Carl Friedrich Gauss was among those to suggest that $X/\log X$ was a good approximation. To be more precise, Gauss conjectured that

$$\lim_{X \longrightarrow \infty} \frac{\pi(X)}{X/\log X} = 1$$

where $\pi(X)$ denotes the number of primes less than or equal to X.

This report focuses on three proofs of this conjecture, a result known as the prime number theorem, which is one of the most important in all of number theory. Regarding the initial conjecture, it was based more on solid numerical evidence than on much solid mathematics. However, it did provide mathematicians with something "definite" to prove. Yet, in an indication of its difficulty, little progress was made until the mid 19th century, when Pafnuty Lvovich Chebyshev published two significant papers on the subject. With the aid of Chebyshev's work, the prime number theorem was finally proved independently by Jacques Hadamard and Charles-Jean de la Vallée Poussin in 1896.

In chapter 1, Chebyshev's theorem is proved, a result that proves the existence of upper and lower bounds for $\frac{\pi(X)}{X/\log X}$. Chebyshev's theorem is vital for the work of Chapter 2, in which is outlined a proof of the prime number theorem based on many of the methods of Hadamard and de la Vallée Poussin. These first proofs rely on many results from a seemingly unrelated area of mathematics, namely complex analysis, and perhaps this partially explains why it was nearly one hundred years from conjecture to proof. Of course, nowadays, every number theorist knows a thing or two about analysis.

After the initial proofs, the next question was whether or not the prime number theorem could be proved by (mainly) number-theoretic methods. Note that as the theorem is a statement about limits, some real analysis must be involved. Such proofs do indeed exist, due independently in the first instances to Erdős and Selberg in 1949. The proof outlined in

Chapter 3 is of the same nature, based on a proof due to Norman Levinson in 1969. When segmented, there is actually little in this proof that a final year undergraduate could not pick up easily. To choose the segments and then put them together is the trick.

The third proof, outlined in Chapter 4, is an interpretation of Donald J. Newman's short proof of the prime number theorem, dating to 1980. Like the proof outlined in Chapter 2, it is heavily analytic. In both proofs, similar results concerning the famous Riemann zeta function are used, but after that, very different approaches are used. In this sense, and compared to all other previous proofs as well, Newman's proof of the prime number theorem is a new one, and not a case of improving on someone else's argument.

During the course of all the proofs, the author's aim is to give the reader some flavour of the motives for doing things. Some of the reasons given may not be what fully motivated the likes of Chebyshev and Newman; indeed, in some cases, the reasons are tenuous, and in some cases it was felt best to offer no reasons at all. In a final remark, the author can only conclude that inspiration is the heart of any proof of the prime number theorem, and thus the prime number theorem lies at the heart of mathematics.

Chapter 1

Chebyshev's theorem, an important forerunner to the prime number theorem

In two papers from 1848 and 1850, Pafnuty Lvovich Chebyshev attempted to prove the prime number theorem, namely that

$$\lim_{X \longrightarrow \infty} \frac{\pi(X)}{X/\log X} = 1,$$

where $\pi(X)$ denotes the number of primes less than or equal to X. For more on Chebyshev's work, see [3] and [4]. Although he did not succeed, he was able to prove that if this limit exists at all, then it equals one. He also proved that the ratio — as opposed to the then unproven limit — is bounded above and below by two positive constants near to 1 for all x. This result became known as Chebyshev's theorem, although some other results he proved carry the same name. Clearly, it is a weaker result than our ultimate interest, the prime number theorem. It should be no surprise then that it features in many proofs of the prime number theorem, including the analytic proof that follows. We begin by stating Chebyshev's theorem, and aim thereafter to obtain a proof.

1.1 Chebyshev's theorem

Theorem 1.1.1 (Chebyshev's theorem) *There exist positive constants c_1 and c_2 such that for every real number $X \geq 2$,*

$$c_1 \frac{X}{\log X} < \pi(X) < c_2 \frac{X}{\log X} .$$

Analysing this for a moment, one may ask, how does one start? Indeed, even familiarity with the proof to be outlined does not point out an obvious starting place, or natural route. To elaborate, Chebyshev's theorem is not a step-by-step exercise, but rather the result of some fine mathematical thinking. Thus, the author's aim is to stretch logic as far as it goes, but ultimately, to leave the reader with an appreciation of Chebyshev's work. We begin with a definition.

Definition 1.1.1 *Define the function* $\theta : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ *by*

$$\theta(X) = \sum_{p \leq X} \log p$$

where the sum is over all primes p *less than or equal to* X*.*

Keeping in mind that Chebyshev's interest was in proving the prime number theorem outright, for which there was much numerical evidence, and which involved a log term, his motivation for working with $\theta(X)$ was that he could introduce $\theta(X)$ in place of X, provided that it could be shown that

$$\lim_{X \longrightarrow \infty} \frac{\theta(X)}{X} = 1.$$

This would reduce the proof of the prime number theorem to proving that

$$\lim_{X \longrightarrow \infty} \frac{\pi(X)}{\theta(X)/\log X} = 1.$$

Although the numerical evidence lent support to the first statement, Chebyshev was unable to prove it. He did find an upper estimate for $\theta(X)$, and this is proved in Chapter 3, Theorem 3.2.3. However, he was unable to prove that there existed a positive constant c_1 such that $c_1 X \leq \theta(X)$ for all $X \geq 2$, which would have immediately implied Chebyshev's lower bound, given the fact that $\theta(X) \leq \sum_{p \leq X} \log X = \pi(X) \log X$. Therefore, Chebyshev also worked with a better lower bound for $\pi(X) \log X$ than $\theta(X)$. Before introducing this bound, the von Mangoldt function is defined, named after Hans Carl Friedrich von Mangoldt.

Definition 1.1.2 *Define the von Mangoldt function* $\Lambda : \mathbb{N} \longrightarrow \mathbb{R}$ *by*

$$\Lambda(n) = \begin{cases} \log p, & \text{if } n = p^r \text{ with } p \text{ prime, } r \in \mathbb{N} \\ 0, & \text{otherwise.} \end{cases}$$

Note that this is an *arithmetic function,* that is, a function from the natural numbers to the real or complex numbers. Chebyshev's ψ function is now introduced.

Definition 1.1.3 *Define* $\psi : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ *by*

$$\psi(X) = \sum_{n \leq X} \Lambda(n) = \sum_{\substack{p,n \\ p^n \leq X}} \log p$$

where p *denotes a prime and* $n \in \mathbb{N}$*.*

Clearly, $\theta(X) \le \psi(X)$, the latter having extra "prime power" terms. Such terms are relatively rare among the integers, so this together with the fact that $\theta(X) \le \pi(X)\log X$ suggests that $\psi(X)$ is a lower bound for $\pi(X)\log X$. In any event, this is easily shown. From the above definition, where p denotes a prime and $n \in \mathbb{N}$,

$$\psi(X) = \sum_{p \le X} \log p \left(\sum_{n,\, p^n \le X} 1 \right).$$

Now $p^n \le X, n \in \mathbb{N}$ if and only if $1 \le n \le \log_p X$, equivalently, $1 \le n \le \frac{\log X}{\log p}$. Therefore,

$$\begin{aligned} \psi(X) &= \sum_{p \le X} \log p \left(\sum_{1 \le n \le \frac{\log X}{\log p}} 1 \right) = \sum_{p \le X} \log p \left[\frac{\log X}{\log p} \right] \\ &\le \sum_{p \le X} \log X = \pi(X) \log X. \end{aligned} \tag{1.1}$$

In light of (1.1), Chebyshev's lower bound is proved if it can be shown that for all $X \ge 2$, there exists a positive constant c_1 such that

$$c_1 X \le \psi(X) \tag{1.2}$$

for all $X \ge 2$.

Turning our attention to Chebyshev's upper bound, the idea is to find an upper bound for $\pi(X)$ involving $\log X$ as the denominator, after which it is required that the numerator is bounded by $c_2 X$, some positive constant c_2. Write

$$\pi(X) = \sum_{p \le X^{1/2}} 1 + \sum_{X^{1/2} < p \le X} 1.$$

This may seem strange until one considers the following inequalities:

$$\sum_{X^{1/2} < p \le X} 1 \le \sum_{X^{1/2} < p \le X} \frac{\log p}{\log X^{1/2}} = \frac{2}{\log X} \sum_{X^{1/2} < p \le X} \log p$$

and

$$\sum_{p \le X^{1/2}} 1 \le X^{1/2} = \frac{X}{X^{1/2}} \le \frac{X}{\log X^{1/2}} = \frac{2X}{\log X}.$$

Clearly then, for the upper bound, it is sufficient to show that there exists a positive constant c_3 such that

$$\sum_{X^{1/2} < p \le X} \log p \le c_3 X.$$

It is therefore sufficient to show that for all $X \geq 2$, there exists a positive constant c_3 such that

$$\sum_{X^{1/2} < n \leq X} \Lambda(n) \leq c_3 X,$$

and the latter statement proves easier to justify. For the moment, consider the following argument. Suppose that $X \geq 2$ and let the integer $k \geq 0$ be defined so that $2^k < X^{1/2} \leq 2^{k+1}$. Thus if

$$X^{1/2} < n \leq X,$$

it follows that

$$\frac{X}{2^{k+1}} < n \leq X.$$

Hence, for some $j \in \{0 \ldots k\}$,

$$\frac{X}{2^{j+1}} < n \leq \frac{X}{2^j}$$

and

$$\sum_{X^{1/2} < n \leq X} \Lambda(n) \leq \sum_{j=0}^{k} \sum_{\frac{X}{2^{j+1}} < n \leq \frac{X}{2^j}} \Lambda(n).$$

Thus for the upper bound, it is sufficient to show that

$$\sum_{j=0}^{k} \sum_{\frac{X}{2^{j+1}} < n \leq \frac{X}{2^j}} \Lambda(n) < c_3 X, \tag{1.3}$$

for some positive constant c_3. In light of this, we shall attempt to find a suitable upper bound for the term

$$\sum_{Y/2 < n \leq Y} \Lambda(n)$$

for $Y > 0$. The floor function is now introduced, which has a number of properties that will prove vital to our argument.

Definition 1.1.4 *Let $X \in \mathbb{R}$. The floor of X, denoted $[X]$, is defined to be the largest integer not greater than X. The fractional part of X, denoted $\{X\}$, is then defined by the formula $\{X\} = X - [X]$.*

Note that if $Y/2 < n \leq Y$, then $[Y/n] = 1$ and $[Y/2n] = 0$, thus

$$\sum_{Y/2 < n \leq Y} \Lambda(n) = \sum_{Y/2 < n \leq Y} \Lambda(n) \left(\left[\frac{Y}{n} \right] - c \left[\frac{Y}{2n} \right] \right)$$

for any constant c. Next, note that

$$\left[\frac{k}{n} \right] \leq \frac{[k]}{n}$$

for all $n \in \mathbb{N}$. (Should the reader wish to verify this, write $k = [k] + \{k\}$ and use the Euclidean algorithm.) Thus,

$$\frac{[Y/n]}{2} \geq \left[\frac{Y/n}{2}\right]$$

so that

$$\left[\frac{Y}{n}\right] - 2\left[\frac{Y}{2n}\right] \geq 0.$$

It follows that

$$\begin{aligned}
\sum_{Y/2<n\leq Y} \Lambda(n) &= \sum_{Y/2<n\leq Y} \Lambda(n)\left(\left[\frac{Y}{n}\right] - 2\left[\frac{Y}{2n}\right]\right) \\
&\leq \sum_{n\leq Y} \Lambda(n)\left(\left[\frac{Y}{n}\right] - 2\left[\frac{Y}{2n}\right]\right) \\
&= \sum_{n\leq Y} \Lambda(n)\left[\frac{Y}{n}\right] - 2\sum_{n\leq Y/2} \Lambda(n)\left[\frac{Y/2}{n}\right], \qquad (1.4)
\end{aligned}$$

with the last equality coming from the fact that

$$\left[\frac{X/2}{n}\right] = 0$$

whenever $n > X/2$. Thus, if an upper bound is obtained for

$$\sum_{n\leq Y} \Lambda(n)\left[\frac{Y}{n}\right], \qquad (1.5)$$

then an upper bound for $\sum_{Y/2<n\leq Y} \Lambda(n)$ will follow. The idea is to re-express (1.5) in terms of logs, and then obtain a bound. Such a re-expression is found using the standard result below.

Theorem 1.1.2 *For every $m \in \mathbb{N}$,*

$$\sum_{n|m} \Lambda(n) = \log m. \qquad (1.6)$$

Proof. This is trivial for $m = 1$, so it remains to consider the case $m \geq 2$. Let $m = {p_1}^{u_1} \ldots {p_r}^{u_r}$ be the canonical decomposition of m. Then, the only non-zero contribution to the left hand side of (1.6) occurs when $n = {p_i}^{v_i}$, where $i = 1 \ldots r$, $1 \leq v_i \leq u_i$. Therefore,

$$\begin{aligned}
\sum_{n|m} \Lambda(n) &= \sum_{i=1}^{r}\sum_{v_i=1}^{u_i} \Lambda({p_i}^{v_i}) = \sum_{i=1}^{r}\sum_{v_i=1}^{u_i} \log p_i = \sum_{i=1}^{r} u_i \log p_i \\
&= \sum_{i=1}^{r} \log {p_i}^{u_i} = \log\left({p_1}^{u_1} \ldots {p_r}^{u_r}\right) = \log m.
\end{aligned}$$

Q.E.D.

We are now in a position to re-express (1.5).

Proposition 1.1.3 *For all* $X \in \mathbb{R}$,

$$\sum_{m \le X} \log m = \sum_{n \le X} \Lambda(n) \left[\frac{X}{n} \right].$$

Proof. It follows from Theorem 1.1.2 that

$$\sum_{m \le X} \log m = \sum_{m \le X} \sum_{n|m} \Lambda(n).$$

Now, for $m \le X$, if $n|m$, then $n \le X$. Therefore,

$$\begin{aligned}
\sum_{m \le X} \log m &= \sum_{n \le X} \sum_{m \le X} \sum_{n|m} \Lambda(n) \\
&= \sum_{n \le X} \left(\Lambda(n) \sum_{m \le X} \sum_{\substack{m=kn \\ k \in \mathbb{N}}} 1 \right) \\
&= \sum_{n \le X} \Lambda(n) \sum_{kn \le X} 1 \\
&= \sum_{n \le X} \Lambda(n) \sum_{k \le X/n} 1 \\
&= \sum_{n \le X} \Lambda(n) \left[\frac{X}{n} \right].
\end{aligned}$$

Q.E.D.

The next theorem provides the required bound for (1.5).

Theorem 1.1.4 *As* $X \longrightarrow \infty$,

$$\sum_{n \le X} \Lambda(n) \left[\frac{X}{n} \right] = X \log X - X + O(\log X).$$

Proof. In view of Proposition 1.1.3, the proof is completed if it can be shown that as $X \longrightarrow \infty$,

$$\sum_{m \le X} \log m - X \log X + X = O(\log X). \tag{1.7}$$

The fact that $\log X$ increases with X is used throughout. Begin by noting that

$$\log m \le \int_m^{m+1} \log u du \le \log(m+1)$$

for all $m \in \mathbb{N}$. Thus

$$\begin{aligned}\sum_{m \leq X} \log m &\leq \sum_{m \leq X} \int_m^{m+1} \log u du = \int_1^{[X]+1} \log u du \\ &= \int_1^X \log u du + \int_X^{[X]+1} \log u du. \qquad (1.8)\end{aligned}$$

Now

$$\int_X^{[X]+1} \log u du \leq \int_X^{X+1} \log u du \leq \log (X+1). \qquad (1.9)$$

Also, if $f(u) = u \log u - u$, then $f'(u) = \log u$; hence

$$\int_1^X \log u du = [u \log u - u]_1^X = X \log X - X + 1. \qquad (1.10)$$

Given (1.9) and (1.10), the inequality

$$\sum_{m \leq X} \log m - (X \log X - X + 1) \leq \log(X+1) \qquad (1.11)$$

can be derived from (1.8). On the other hand, for every $m \in \mathbb{N}$,

$$\log m \geq \int_{m-1}^m \log u du.$$

With a similar argument to the first part of the proof, we can also derive the inequality

$$-\log X \leq \sum_{m \leq X} \log m - (X \log X - X + 1). \qquad (1.12)$$

Thus by (1.11) and (1.12),

$$\left| \sum_{m \leq X} \log m - (X \log X - X + 1) \right| \leq \log (X+1).$$

For $X \geq 2$, $\log (X+1) \leq \log X^2 = 2 \log X$. Therefore, as $X \longrightarrow \infty$,

$$\sum_{m \leq X} \log m - (X \log X - X + 1) = O(\log X).$$

The result easily follows.

Q.E.D.

An upper bound for $\sum_{Y/2<n\le Y}\Lambda(n)$ is now obtained. To elaborate, given (1.4), as $Y \longrightarrow \infty$,

$$\begin{aligned}
\sum_{Y/2<n\le Y}\Lambda(n) &\le \sum_{n\le Y}\Lambda(n)\left[\frac{Y}{n}\right] - 2\sum_{n\le Y/2}\Lambda(n)\left[\frac{Y/2}{n}\right] \\
&= Y\log Y - Y + O(\log Y) - 2\left(\frac{Y}{2}\log\frac{Y}{2} - \frac{Y}{2} + O(\log Y)\right) \\
&= Y\left(\log Y - \log\frac{Y}{2}\right) + O(\log Y) \\
&= Y\log 2 + O(\log Y) \qquad (1.13) \\
&= O(Y). \qquad (1.14)
\end{aligned}$$

Having found a bound for $\sum_{Y/2<n\le Y}\Lambda(n)$, and recalling that (1.3) is a sufficient condition for Chebyshev's upper bound, we now attempt to prove this sufficient condition. Note that although a bound for $\sum_{Y/2<n\le Y}\Lambda(n)$ has been found, it has only been shown to apply for sufficiently large Y. As it is required that (1.3) hold for all $X \ge 2$, during the course of the proof, we extend this bound to all $Y > 0$.

Theorem 1.1.5 *For all $X \ge 2$,*

$$\sum_{j=0}^{k}\sum_{\frac{X}{2^{j+1}}<n\le\frac{X}{2^j}}\Lambda(n) < c_2X$$

for some positive constant c_2.

Proof. By (1.14), there exists a positive constant c_3 and $Y_0 > 0$ such that for all $Y > Y_0$,

$$\sum_{Y/2<n\le Y}\Lambda(n) < c_3Y.$$

To extend this to all Y such that $0 < Y \le Y_0$, observe that in this range, $[Y]$ takes finitely many values, as does $[Y/2]$. Therefore, by noting that

$$\sum_{Y/2<n\le Y}\Lambda(n) = \sum_{[Y/2]<n\le[Y]}\Lambda(n),$$

it follows that

$$\max\left\{\sum_{Y/2<n\le Y}\Lambda(n) : 0 < Y \le Y_0\right\}$$

is defined. Call this maximum c_4. Thus

$$\sum_{Y/2<n\le Y}\Lambda(n) \le c_4$$

whenever $0 \le Y \le Y_0$. Clearly then,

$$\sum_{Y/2<n\le Y} \Lambda(n) \le c_4 Y \tag{1.15}$$

whenever $1 \le Y \le Y_0$, and as

$$\sum_{Y/2<n\le Y} \Lambda(n) = 0$$

if $0 \le Y < 1$, (1.15) extends to all positive $Y \le Y_0$. Letting $c_5 = \max\{c_3, c_4\}$, it follows that

$$\sum_{Y/2<n\le Y} \Lambda(n) \le c_5 Y$$

for all $Y > 0$. Hence, for $X \ge 2$ we have that

$$\sum_{j=0}^{k} \sum_{\frac{X}{2^{j+1}}<n\le\frac{X}{2^j}} \Lambda(n) \le \sum_{j=0}^{k} c_5 \frac{X}{2^j} = c_5 X \sum_{j=0}^{k} \frac{1}{2^j} < 2c_5 X,$$

so that $2c_5$ is the desired constant c_2.

Q.E.D.

Thus, Chebyshev's upper bound is proved. To complete Chebyshev's theorem, it remains to prove the lower bound. Recall it is sufficient to prove (1.2) for all $X \ge 2$. As it happens, some of the work has already been done in the course of proving the upper bound. As we already know much about expressions involving the floor function, the idea is to use it to obtain a lower bound for $\psi(X) = \sum_{n\le X} \Lambda(n)$.

Theorem 1.1.6 *For all $X \ge 2$, there exists a positive constant c_1 such that*

$$c_1 X \le \psi(X).$$

Proof. First, consider the function $[\alpha] - 2[\alpha/2]$. Clearly, $[\alpha] \le \alpha$ and $2[\alpha/2] > 2(\alpha/2 - 1)$. Therefore,

$$[\alpha] - 2[\alpha/2] < \alpha - 2(\alpha/2 - 1) = 2.$$

Since $[\alpha] - 2[\alpha/2]$ is an integer, $[\alpha] - 2[\alpha/2] \le 1$. Therefore,

$$\left[\frac{X}{n}\right] - 2\left[\frac{X}{2n}\right] \le 1$$

so that

$$\sum_{n\le X} \Lambda(n) \left(\left[\frac{X}{n}\right] - 2\left[\frac{X}{2n}\right] \right) \le \sum_{n\le X} \Lambda(n) = \psi(X).$$

Thus a lower bound for the left-hand side is also a lower bound for the right. Now, as $X \longrightarrow \infty$, it follows from (1.4) and (1.13) that

$$\sum_{n \le X} \Lambda(n) \left(\left[\frac{X}{n} \right] - 2 \left[\frac{X}{2n} \right] \right) = X \log 2 + O(\log X).$$

To obtain a bound, rewrite this as

$$\frac{1}{2} X \log 2 = \sum_{n \le X} \Lambda(n) \left(\left[\frac{X}{n} \right] - 2 \left[\frac{X}{2n} \right] \right) - \frac{1}{2} X \log 2 + O(\log X).$$

Hence, there exists a positive constant c_6 such that

$$\frac{1}{2} X \log 2 \le \sum_{n \le X} \Lambda(n) \left(\left[\frac{X}{n} \right] - 2 \left[\frac{X}{2n} \right] \right) \le \psi(X)$$

for all $X \ge c_6$. Thus, it suffices to find a positive constant c_7 such that $c_7 X \le \psi(X)$ for all X such that $2 \le X \le c_6$. Note that $\psi(X)$ only takes a finite number of values in this domain, say $\psi(X_1) < \cdots < \psi(X_n)$, where $X_n = c_6$. Consider the constants $C_m, m = 1 \ldots n$ such that $C_m X_{m+1} = \psi(X_m)$. Then if $X_m \le X < X_{m+1}$, it follows that $c_m X \le \psi(X)$. A sufficient c_7 is the smallest of these constants, and the smaller of c_7 and $\frac{1}{2} \log 2$ is the desired constant c_1.

Q.E.D.

This completes the proof of Chebyshev's Theorem.

Chapter 2

An analytic proof of the prime number theorem

2.1 Sufficient conditions for the prime number theorem

This chapter outlines an analytic proof of the prime number theorem, as based on a set of lectures given by WWL Chen in 1981 [5]. The proof is analytic in the sense that it relies heavily on some of the most important results in complex analysis, and indeed the first proofs of the prime number theorem, due independently to Jacques Hadamard and Charles-Jean de la Vallée Poussin in 1896, were analytic. For full accounts of their proofs, see [9], [10] and [16]. Chen's approach is based on these classical proofs, and throughout the proof, a number of standard results from complex analysis are used. In most cases, these results are stated, rather than proved.

The prime number theorem has already been stated in Chapter 1 in the form

$$\lim_{X\longrightarrow\infty}\frac{\pi(X)}{X/\log X}=1.$$

We now introduce some new notation, so as to write this in a different way.

Definition 2.1.1 *If*

$$\lim_{X\to\infty}\frac{f(X)}{g(X)}=1,$$

then we write

$$f(X)\sim g(X)\quad as\quad X\longrightarrow\infty,$$

and say that the functions f and g are asymptotic.

Thus the prime number theorem may equivalently be written as

$$\pi(X)\sim\frac{X}{\log X}\ \text{ as }\ X\longrightarrow\infty,$$

and this is also known as the "asymptotic law of the distribution of prime numbers". Provided $g(X) \geq 0$ for all sufficiently large X, it is easy to show that the statement

$$f(x) \sim g(x) \quad \text{as} \quad X \longrightarrow \infty$$

is equivalent to stating that for an arbitrary $\epsilon > 0$,

$$1 - \epsilon < \frac{f(X)}{g(X)} < 1 + \epsilon \tag{2.1}$$

for all sufficiently large X.

Before the prime number theorem was proved, all of the numerical evidence suggested that it was true. Indeed, from Chapter 1, Chebyshev's upper and lower bounds were being proved to be closer and closer to one. Needless to say, such evidence is not a proof. But recall from (1.1) that

$$\psi(X) = \sum_{p \leq X} \log p \left[\frac{\log X}{\log p} \right].$$

When one considers even the possibility that

$$\sum_{n \leq X} \log n \sim X \log X \quad \text{as} \quad X \longrightarrow \infty,$$

and consequently, the greater possibility that

$$\sum_{n \leq X} \log n \left[\frac{\log X}{\log n} \right] \sim X \log X \quad \text{as} \quad X \longrightarrow \infty,$$

it seems plausible that

$$\psi(X) \sim \pi(X) \log X \quad \text{as} \quad X \longrightarrow \infty,$$

although at this stage we can assume nothing about the distribution of primes. But in fact, the numerical evidence backs up this claim. If it is indeed true, then it is a straightforward matter to check that the statement

$$\psi(X) \sim X \quad \text{as} \quad X \longrightarrow \infty \tag{2.2}$$

implies the prime number theorem. Of course, it might not even be true, but again, the numerical evidence suggests otherwise. Before attempting to prove (2.2), we first prove it is a sufficient condition for the prime number theorem. Indeed, (2.2) and the prime number theorem are equivalent statements, the proof of which is not necessary here.

Theorem 2.1.1 *If*

$$\psi(X) \sim X \quad \textit{as} \quad X \longrightarrow \infty,$$

then

$$\pi(X) \sim \frac{X}{\log X} \quad \textit{as} \quad X \longrightarrow \infty.$$

Proof. Assume that

$$\psi(X) \sim X \text{ as } X \longrightarrow \infty,$$

that is, by (2.1), assume that given any $\epsilon > 0$, we have

$$1 - \epsilon < \frac{\psi(X)}{X} < 1 + \epsilon \tag{2.3}$$

for all sufficiently large X. It follows from (1.1) that

$$1 - \epsilon < \frac{\psi(X)}{X} \leq \frac{\pi(X)}{X/\log X}$$

for all sufficiently large X. It is clear then that if it can be shown that

$$\frac{\pi(X)}{X/\log X} < 1 + \epsilon$$

for all sufficiently large X, then the prime number theorem is implied. The idea is to obtain a suitable inequality involving $\pi(X)/(X/\log X)$ and $\psi(X)/X$, and again use (2.3). Begin by choosing any $\alpha \in (0, 1)$. Then, it follows that

$$\begin{aligned}
\psi(X) &\geq \sum_{p \leq X} \log p \geq \sum_{X^\alpha < p \leq X} \log p \\
&\geq \sum_{X^\alpha < p \leq X} \log X^\alpha \\
&= \alpha(\pi(X) - \pi(X^\alpha)) \log X.
\end{aligned}$$

From this, it easily follows that

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

By (2.3), given any $\delta > 0$, we have

$$\frac{\pi(X)}{X/\log X} < \frac{1}{\alpha}(1 + \delta) + \frac{\pi(X^\alpha)}{X/\log X}$$

for all sufficiently large X. Choosing $\alpha = 1/(1 + \delta)$, we then have

$$\frac{\pi(X)}{X/\log X} \leq (1 + \delta)^2 + \frac{\pi(X^{1/(1+\delta)})}{X/\log X}$$

for all sufficiently large X. This holds for any $\delta > 0$, so choose δ such that $(1+\delta)^2 < 1+\epsilon/2$. It is clear then that the prime number theorem is implied if, for any $\alpha \in (0, 1)$,

$$\lim_{X \longrightarrow \infty} \frac{\pi(X^\alpha)}{X/\log X} = 0.$$

Since

$$\frac{\pi(X^\alpha)}{X/\log X} \geq 0,$$

it suffices to show that for any constant $c > 0$

$$\frac{\pi(X^\alpha)}{X/\log X} < c$$

for all sufficiently large X. This inequality may be rewritten as

$$\pi(X^\alpha) < c\frac{X}{\log X} \ .$$

Recalling Chebyshev's theorem, (Theorem 1.1.1) there exists a positive constant c_2 such that whenever $X^\alpha \geq 2$,

$$\pi(X^\alpha) < c_2\frac{X^\alpha}{\log X^\alpha} \ .$$

(Note that since $\alpha > 0$, such X exist.) It suffices then to show that, given the choices of c and α,

$$c_2\frac{X^\alpha}{\log X^\alpha} \leq c\frac{X}{\log X} \tag{2.4}$$

for all sufficiently large X. It is easily shown that (2.4) is equivalent to the inequality

$$\sqrt[1-\alpha]{\frac{c_2}{c\ \alpha}} \geq X,$$

which, as $\alpha < 1$, is clearly true for all sufficiently large X. The theorem is therefore completed.

Q.E.D.

Thus (2.2) is a sufficient condition for the prime number theorem. (Note that the previous proof did not contain any complex analysis, therefore (2.2) is also a sufficient condition for the elementary proof of the prime number theorem outlined in Chapter 3.) This is difficult to prove directly, to a large extent because ψ is discontinuous, so instead, we consider a smooth average of the function $\psi(X)$. For $X > 0$, let

$$\psi_1(X) = \int_0^X \psi(x)\mathrm{d}x.$$

Now if $\psi(X) \sim X$ as $X \longrightarrow \infty$, the following is suggested:

$$\int_0^X \psi(x)\mathrm{d}x \sim \int_0^X x\mathrm{d}x \ \text{ as } \ X \longrightarrow \infty,$$

that is,

$$\psi_1(X) \sim \frac{1}{2}X^2 \ \text{ as } \ X \longrightarrow \infty.$$

If the converse of this statement is true, we have another sufficient condition for the prime number theorem.

Theorem 2.1.2 *Suppose that $\psi_1(X) \sim \frac{1}{2}X^2$ as $X \longrightarrow \infty$. Then $\psi(X) \sim X$ as $X \longrightarrow \infty$.*

Proof. Since $\psi(X)$ increases as X increases, it follows by the mean value theorem that for every $X > 0$ and $\beta > 1$,

$$\begin{aligned} \psi(X) &\leq \frac{1}{\beta X - X} \int_X^{\beta X} \psi(x) \mathrm{d}x \\ &= \frac{1}{(\beta - 1)X} \left(\psi_1(\beta X) - \psi_1(X)\right), \end{aligned}$$

which relates ψ and ψ_1. It follows that

$$\frac{\psi(X)}{X} \leq \frac{\psi_1(\beta X) - \psi_1(X)}{(\beta - 1)X^2} \; . \tag{2.5}$$

In a similar way, it can be shown that whenever $X > 0$ and $\alpha \in (0, 1)$,

$$\frac{\psi(X)}{X} \geq \frac{\psi_1(X) - \psi_1(\alpha X)}{(1 - \alpha)X^2} \; . \tag{2.6}$$

Assume that $\psi_1(X) \sim \frac{1}{2}X^2$ as $X \longrightarrow \infty$. Therefore,

$$\psi_1(\beta X) \sim \frac{1}{2}\beta^2 X^2 \text{ as } X \longrightarrow \infty.$$

It follows that

$$\psi_1(\beta X) - \psi_1(X) \sim \frac{1}{2}((\beta^2 - 1)X^2) \text{ as } X \longrightarrow \infty.$$

Equivalently,

$$\frac{\psi_1(\beta X) - \psi_1(X)}{(\beta - 1)X^2} \sim \frac{1}{2}(\beta + 1) \text{ as } X \longrightarrow \infty.$$

In the same way,

$$\frac{\psi_1(X) - \psi_1(\alpha X)}{(1 - \alpha)X^2} \sim \frac{1}{2}(1 + \alpha) \text{ as } X \longrightarrow \infty.$$

Choose $\epsilon \in (0, 1)$. Then, since $\psi_1(X)$ is an increasing function, by (2.1) it follows that

$$\frac{\psi_1(\beta X) - \psi_1(X)}{(\beta - 1)X^2} < \frac{1}{2}(\beta + 1)(1 + \epsilon)$$

and

$$\frac{1}{2}(1 + \alpha)(1 - \epsilon) < \frac{\psi_1(X) - \psi_1(\alpha X)}{(1 - \alpha)X^2}$$

for all sufficiently large X. Then by (2.5) and (2.6),

$$\frac{1}{2}(1 + \alpha)(1 - \epsilon) < \frac{\psi(X)}{X} < \frac{1}{2}(\beta + 1)(1 + \epsilon)$$

for all sufficiently large X. The result follows on letting $\alpha = 1 - 2\epsilon$ and $\beta = 1 + 2\epsilon$.

Q.E.D.

Therefore, to prove the prime number theorem, it suffices to prove:

Theorem 2.1.3

$$\psi_1(X) \sim \frac{1}{2}X^2 \text{ as } X \longrightarrow \infty.$$

The rest of our efforts are dedicated to this. The next proposition gives an alternative expression for $\psi_1(X)$, though its usefulness may be not yet be clear.

Proposition 2.1.4 *For $X \geq 1$,*

$$\psi_1(X) = \sum_{n=1}^{[X]} (X - n)\Lambda(n).$$

Proof. Note that for $0 < x < 1$,

$$\psi(x) = \sum_{n \leq x} \Lambda(n) = 0.$$

Hence,

$$\begin{aligned}
\psi_1(X) &= \int_1^X \psi(x)\mathrm{d}x \\
&= \int_1^X \left(\sum_{n \leq x} \Lambda(n) \right) \mathrm{d}x \\
&= \sum_{m=1}^{[X]-1} \int_m^{m+1} \left(\sum_{n \leq x} \Lambda(n) \right) \mathrm{d}x + \int_{[X]}^X \left(\sum_{n \leq x} \Lambda(n) \right) \mathrm{d}x \\
&= \sum_{m=1}^{[X]-1} \sum_{n \leq m} \Lambda(n) + (X - [X]) \sum_{n \leq [X]} \Lambda(n). \qquad (2.7)
\end{aligned}$$

Claim: For all $r \in \mathbb{N}$,

$$\sum_{m=1}^{r-1} \sum_{n \leq m} \Lambda(n) = \sum_{n=1}^{r} (r - n)\Lambda(n).$$

The claim is clearly true for $r = 1$. Assume it is true for $r = k$. Then

$$\begin{aligned}
\sum_{m=1}^{k} \sum_{n \le m} \Lambda(n) &= \sum_{m=1}^{k-1} \sum_{n \le m} \Lambda(n) + \sum_{n \le k} \Lambda(n) \\
&= \sum_{n=1}^{k} (k-n)\Lambda(n) + \sum_{n=1}^{k} \Lambda(n) \\
&= \sum_{n=1}^{k} (k+1-n)\Lambda(n) \\
&= \sum_{n=1}^{k+1} (k+1-n)\Lambda(n).
\end{aligned}$$

Inductively, the claim is proved true. Hence, given (2.7),

$$\begin{aligned}
\psi_1(X) &= \sum_{n=1}^{[X]} ([X]-n)\Lambda(n) + \sum_{n=1}^{[X]} (X-[X])\Lambda(n) \\
&= \sum_{n=1}^{[X]} (X-n)\Lambda(n).
\end{aligned}$$

Q.E.D.

A variant of this result shall be used later on. Complex analysis now enters our argument for the first time with a discussion on the Riemann zeta function.

2.2 The Riemann zeta function

We begin this section with the definition of a Dirichlet series, and then define the Riemann zeta function, it being a specific case of such a series.

Definition 2.2.1 *A Dirichlet series is a series of the type*

$$F(s) = \sum_{n=1}^{\infty} f(n) n^{-s} \tag{2.8}$$

where $f : \mathbb{N} \longrightarrow \mathbb{C}$ *is an arithmetic function and* $s \in \mathbb{C}$.

Definition 2.2.2 *The Dirichlet series*

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}, \quad \text{Re } s > 1 \tag{2.9}$$

is called the Riemann zeta function, or the zeta function.

It is perhaps unclear how complex analysis might enter into our present argument, and when one considers the definition of the zeta function, it hardly seems likely that it has anything to offer. However, those familiar with the zeta function may recall that it is related to the von Mangoldt function, which lies at the heart of our argument. This relationship is first stated, but before a proof is outlined, some auxiliary results are needed. As in nearly every area where complex analysis is used, the zeta function plays a hugely significant role in this analytic proof.

Throughout the rest of this thesis, we shall write $s = \sigma + it$, where σ and t are real. Also note that

$$|n^{it}| = |e^{\log n^{it}}| = |e^{it \log n}| = |\cos(t \log n) + i \sin(t \log n)| = 1, \tag{2.10}$$

and this fact is used throughout.

Theorem 2.2.1 *Suppose that $\sigma > 1$. Then*

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{n=1}^{\infty} \Lambda(n) n^{-s}. \tag{2.11}$$

If a proof is to be obtained, note that we require, for $\sigma > 1$, that $\zeta(s) \neq 0$. Obviously, this is equivalent to showing that $|\zeta(s)| \neq 0$. Due to the fact that the absolute value of a product equals the product of the absolute values, it is desirable to write the Riemann zeta function as a product. A result concerning the convergence properties of Dirichlet series is first stated, with a proof available in [5]. A relevant theorem to such a re-expression is then proved, although its usefulness is not immediately apparent. In actual fact, it shall be used twice to prove Theorem 2.2.1.

Theorem 2.2.2 *Suppose that the series* (2.8) *converges for some $s \in \mathbb{C}$. Then there exist unique real numbers $\sigma_0, \sigma_1, \sigma_2$ satisfying $-\infty \leq \sigma_0 \leq \sigma_1 \leq \sigma_2 \leq +\infty$ such that the following statements hold:*

1) *The series* (2.8) *converges for every $s \in \mathbb{C}$ with $\sigma > \sigma_0$. Furthermore, for every $\epsilon > 0$, the series* (2.8) *diverges for some $s \in \mathbb{C}$ with $\sigma_0 - \epsilon < \sigma \leq \sigma_0$.*

2) *For every $\eta > 0$, the series* (2.8) *converges uniformly on the set $\{s \in \mathbb{C} : \sigma > \sigma_1 + \eta\}$ and does not converge uniformly on the set $\{s \in \mathbb{C} : \sigma_0 > \sigma_1 - \eta\}$.*

3) *The series* (2.8) *converges absolutely for every $s \in \mathbb{C}$ with $\sigma > \sigma_2$. Furthermore, for every $\epsilon > 0$, the series* (2.8) *does not converge absolutely for some $s \in \mathbb{C}$ with $\sigma_2 - \epsilon < \sigma \leq \sigma_2$.*

Theorem 2.2.3 *Suppose that for each $j = 1, 2, 3$,*

$$F_j(s) = \sum_{n=1}^{\infty} f_j(n) n^{-s},$$

where $f_j : \mathbb{N} \longrightarrow \mathbb{C}$ *is an arithmetic function and* $s \in \mathbb{C}$*. Suppose further that for every* $n \in \mathbb{N}$*,*

$$f_3(n) = \sum_{\substack{x,\, y \\ xy=n}} f_1(x) f_2(y).$$

Then, $F_3(s) = F_1(s)F_2(s)$*, provided that* $\sigma > \max\{\sigma_1, \sigma_2\}$*, where the series* $F_j(s)$ *converges absolutely for every* $s \in \mathbb{C}$ *with* $\sigma > \sigma_j, j = 1, 2$*.*

Proof. The idea is to obtain a partial sum expression mirroring $F_1(s)F_2(s) = F_3(s)$, and take limits. We have,

$$\begin{aligned}
\sum_{n=1}^{N} f_3(n)n^{-s} &= \sum_{n=1}^{N} \left(\sum_{\substack{x,\, y \\ xy=n}} f_1(x) f_2(y) \right) n^{-s} \\
&= \sum_{n=1}^{N} \left(\sum_{\substack{x,\, y \\ xy=n}} f_1(x) f_2(y) n^{-s} \right) \\
&= \sum_{n=1}^{N} \left(\sum_{\substack{x,\, y \\ xy=n}} f_1(x) x^{-s} f_2(y) y^{-s} \right).
\end{aligned}$$

Note that for $1 \le n \le N,\ xy = n$ if and only if $1 \le xy \le N$ or, equivalently, $1 \le x \le N, y \le N/x$. Therefore,

$$\begin{aligned}
\sum_{n=1}^{N} f_3(n)n^{-s} &= \sum_{x \le N} \sum_{y \le N/x} f_1(x)x^{-s} f_2(y) y^{-s} \\
&= \sum_{x \le N} \left(f_1(x)x^{-s} \sum_{y \le N/x} f_2(y)y^{-s} \right) \\
&= \sum_{x \le \sqrt{N}} \left(f_1(x)x^{-s} \sum_{y \le N/x} f_2(y)y^{-s} \right) + \sum_{\sqrt{N} < x \le N} \left(f_1(x)x^{-s} \sum_{y \le N/x} f_2(y)y^{-s} \right).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\sum_{n=1}^{N} f_3(n)n^{-s} - \sum_{x \le \sqrt{N}} f_1(x)x^{-s} \sum_{y \le \sqrt{N}} f_2(y)y^{-s} \\
&= \sum_{x \le \sqrt{N}} \left(f_1(x)x^{-s} \sum_{\sqrt{N} < y \le N/x} f_2(y)y^{-s} \right) + \sum_{\sqrt{N} < x \le N} \left(f_1(x)x^{-s} \sum_{y \le N/x} f_2(y)y^{-s} \right).
\end{aligned}$$

It follows that

$$
\begin{aligned}
&\left|\sum_{n=1}^{N} f_3(n)n^{-s} - \sum_{x\le\sqrt{N}} f_1(x)x^{-s} \sum_{y\le\sqrt{N}} f_2(y)y^{-s}\right| \\
\le\ & \sum_{x\le\sqrt{N}} \left(\left|f_1(x)x^{-s}\right| \sum_{\sqrt{N}<y\le N/x} \left|f_2(y)y^{-s}\right|\right) + \sum_{\sqrt{N}<x\le N} \left(\left|f_1(x)x^{-s}\right| \sum_{y\le N/x} \left|f_2(y)y^{-s}\right|\right) \\
\le\ & \sum_{x=1}^{\infty} \left|f_1(x)x^{-s}\right| \sum_{y>\sqrt{N}} \left|f_2(y)y^{-s}\right| + \sum_{x>\sqrt{N}} \left|f_1(x)x^{-s}\right| \sum_{y=1}^{\infty} \left|f_2(y)y^{-s}\right|. \qquad (2.12)
\end{aligned}
$$

Suppose that $\sigma > \max\{\sigma_1, \sigma_2\}$. Then the series $F_1(s)$ and $F_2(s)$ converge absolutely. In other words, the series

$$\sum_{x=1}^{\infty} \left|f_1(x)x^{-s}\right| \quad \text{and} \quad \sum_{y=1}^{\infty} \left|f_2(y)y^{-s}\right|$$

converge. From this, it is clear that

$$\sum_{x>\sqrt{N}} \left|f_1(x)x^{-s}\right| \quad \text{and} \quad \sum_{y>\sqrt{N}} \left|f_2(y)y^{-s}\right|$$

converge to zero as $N \longrightarrow \infty$. It follows that the right hand side of (2.12) converges to zero as $N \longrightarrow \infty$. Therefore,

$$\lim_{N\longrightarrow\infty} \left|\sum_{n=1}^{N} f_3(n)n^{-s} - \sum_{x\le\sqrt{N}} f_1(x)x^{-s} \sum_{y\le\sqrt{N}} f_2(y)y^{-s}\right| = 0.$$

Hence,

$$
\begin{aligned}
\lim_{N\longrightarrow\infty} \sum_{n=1}^{N} f_3(n)n^{-s} &= \lim_{N\longrightarrow\infty} \left(\sum_{x\le\sqrt{N}} f_1(x)x^{-s} \sum_{y\le\sqrt{N}} f_2(y)y^{-s}\right) \\
&= \lim_{N\longrightarrow\infty} \sum_{x\le\sqrt{N}} f_1(x)x^{-s} \lim_{N\longrightarrow\infty} \sum_{y\le\sqrt{N}} f_2(y)y^{-s}.
\end{aligned}
$$

In other words,

$$\sum_{n=1}^{\infty} f_3(n)n^{-s} = \sum_{x=1}^{\infty} f_1(x)x^{-s} \sum_{y=1}^{\infty} f_2(y)y^{-s};$$

that is,

$$F_3(s) = F_1(s)F_2(s),$$

as required.

Q.E.D.

This theorem can be generalised to state the following:

Theorem 2.2.4 *Suppose that for each* $j = 1, 2, \ldots, k+1$,

$$F_j(s) = \sum_{n=1}^{\infty} f_j(n) n^{-s},$$

where $f_j : \mathbb{N} \longrightarrow \mathbb{C}$ *is an arithmetic function and* $s \in \mathbb{C}$. *Suppose further that for every* $n \in \mathbb{N}$,

$$f_{k+1}(n) = \sum_{\substack{x_1, \ldots, x_k \\ x_1 \ldots x_k = n}} f_1(x_1) \ldots f_k(x_k).$$

Then, $F_{k+1}(s) = F_1(s) F_2(s) \ldots F_k(s)$, *provided that* $\sigma > \max\{\sigma_1, \ldots, \sigma_k\}$, *where the series* $F_j(s)$ *converges absolutely for every* $s \in \mathbb{C}$ *with* $\sigma > \sigma_j, j = 1, 2, \ldots, k$.

This generalisation is used to prove the following theorem, which takes a Dirichlet series and equates it to a product. In this case, the arithmetic function involved is multiplicative. A further result assumes a *completely multiplicative function,* that is, an arithmetic function such that $a(n)a(m) = a(nm)$ for all $n, m \in \mathbb{N}$. This is the case for the zeta function, with $f(n) = 1$ for all $n \in \mathbb{N}$.

Theorem 2.2.5 *Suppose that the function* $f : \mathbb{N} \longrightarrow \mathbb{C}$ *is multiplicative. Suppose further that* $F(s) = \sum_{n=1}^{\infty} f(n) n^{-s}$ *is absolutely convergent. Then*

$$F(s) = \prod_p \left(\sum_{h=0}^{\infty} f(p^h) p^{-hs} \right) .$$

Proof. Let p_j be the j^{th} prime in increasing order and choose any $k \in \mathbb{N}$. Define arithmetic functions $g_j : \mathbb{N} \longrightarrow \mathbb{C}$ by

$$g_j(n) = \begin{cases} f(n) & \text{if } n = {p_j}^h \text{ for some } h \in \mathbb{N} \\ 0 & \text{otherwise.} \end{cases}$$

Since these are arithmetic functions, it follows by Theorem 2.2.4 that

$$\sum_{n=1}^{\infty} \left(\sum_{\substack{x_1, \ldots, x_k \\ x_1 \ldots x_k = n}} g_1(x_1) \ldots g_k(x_k) \right) n^{-s} = \prod_{j=1}^{k} \left(\sum_{n=1}^{\infty} g_j(n) n^{-s} \right) . \tag{2.13}$$

Denoting $\mathbb{N} \cup \{0\}$ by $\mathbb{N}_0$, it follows by the definitions of g_j for $j \in \mathbb{N}$ that

$$\begin{aligned}\sum_{n=1}^{\infty} g_j(n)n^{-s} &= \sum_{\substack{n={p_j}^h \\ h\in\mathbb{N}_0}} g_j(n)n^{-s} + \sum_{\substack{n\neq{p_j}^h \\ h\in\mathbb{N}_0}} g_j(n)n^{-s} \\ &= \sum_{\substack{n={p_j}^h \\ h\in\mathbb{N}_0}} f(n)n^{-s} \\ &= \sum_{h=0}^{\infty} f({p_j}^h)(p_j)^{-hs}. \qquad (2.14)\end{aligned}$$

Next, for $n \in \mathbb{N}$, we consider the sum

$$S_n = \sum_{\substack{x_1,\ldots,x_k \\ x_1\ldots x_k=n}} g_1(x_1)\ldots g_k(x_k)$$

by examining two possibilities for n. The first is that the unique prime factorisation of n is expressible in terms of powers of the primes $p_1, \ldots, p_k$. The second is that it is not.

Case 1. Suppose $n = {p_1}^{h_1} \ldots {p_k}^{h_k}$, where $h_j \in \mathbb{N}_0$ for all $j \in \{1, \ldots, k\}$. Letting $x_1 = {p_1}^{h_1}, \ldots, x_k = {p_k}^{h_k}$, so that $n = x_1 \ldots x_k$, it follows that

$$g_1(x_1)\ldots g_k(x_k) = f({p_1}^{h_1})\ldots f({p_k}^{h_k}) = f({p_1}^{h_1}\ldots {p_k}^{h_k}) = f(n),$$

where we use the fact that f is multiplicative.

Now, since there is only one prime factorisation of n, it follows by the pigeonhole principle that there is no other way to express each x_j as a prime power. Therefore, if the x_js are not defined as above, we have $g_j(x_j) = 0$ for some $j \in \{1, \ldots, k\}$. Hence, $g_1(x_1)\ldots g_k(x_k) = 0$. It follows that $S_n = f(n)$.

Case 2. Suppose p is a prime factor of n such that $p \neq p_j$, any $j \in \{1, \ldots, k\}$. Then if $n = x_1 \ldots x_k$, p is a factor of x_j for some $j \in \{1, \ldots, k\}$. Therefore, x_j is not a power of any of the first k primes. It easily follows that $S_n = 0$.

Therefore, in both cases,

$$\sum_{\substack{x_1,\ldots,x_k \\ x_1\ldots x_k=n}} g_1(x_1)\ldots g_k(x_k) = \theta_k(n)f(n), \qquad (2.15)$$

where

$$\theta_k(n) = \begin{cases} 1 & \text{if all the prime factors of } n \text{ are among } p_1, \ldots, p_k \\ 0 & \text{otherwise.} \end{cases}$$

It easily follows by (2.13), (2.14) and (2.15) that

$$\prod_{j=1}^{k}\left(\sum_{h=0}^{\infty} f({p_j}^h){p_j}^{-hs}\right) = \sum_{n=1}^{\infty} \theta_k(n) f(n) n^{-s}.$$

Thus

$$\prod_{j=1}^{k}\left(\sum_{h=0}^{\infty} f({p_j}^h){p_j}^{-hs}\right) - \sum_{n=1}^{\infty} f(n)n^{-s} = \sum_{n=1}^{\infty}(\theta_k(n)-1) f(n) n^{-s}. \qquad (2.16)$$

Now for $n \in \{1, \ldots, k\}$, if p is a prime factor of n, then $p \leq n \leq k \leq p_k$. Therefore, $\theta_k(n) = 1$. Also, $|\theta_k(n) - 1| \leq 1$. Using these two facts, along with (2.16), it follows that

$$\begin{aligned}\left|\prod_{j=1}^{k}\left(\sum_{h=0}^{\infty} f({p_j}^h){p_j}^{-hs}\right) - F(s)\right| &= \left|\sum_{n=k+1}^{\infty}(\theta_k(n)-1) f(n) n^{-s}\right| \\ &\leq \sum_{n=k+1}^{\infty} |\theta_k(n)-1||f(n)|n^{-\sigma} \\ &= \sum_{n=k+1}^{\infty} |f(n)|n^{-\sigma}. \qquad (2.17)\end{aligned}$$

Since $F(s) = \sum_{n=1}^{\infty} f(n)n^{-s}$ is absolutely convergent, that is, since $\sum_{n=1}^{\infty} |f(n)|n^{-\sigma}$ is convergent, it follows that

$$\sum_{n=k+1}^{\infty} |f(n)|n^{-\sigma} \longrightarrow 0 \text{ as } k \longrightarrow \infty.$$

This result, together with (2.17) imply that

$$\prod_{j=1}^{k}\left(\sum_{h=0}^{\infty} f({p_j}^h){p_j}^{-hs}\right) - F(s) \longrightarrow \infty \text{ as } k \longrightarrow \infty.$$

The result follows.

Q.E.D.

Using Theorem 2.2.5, the following theorem allows for the re-expression of the zeta function.

Theorem 2.2.6 *Suppose that the function $f : \mathbb{N} \longrightarrow \mathbb{C}$ is completely multiplicative and not identically zero. Suppose further that $F(s) = \sum_{n=1}^{\infty} f(n)n^{-s}$ is absolutely convergent. Then*

$$F(s) = \prod_{p}(1 - f(p)p^{-s})^{-1}.$$

Proof. Choosing any prime p, consider the series $G(s) = \sum_{h=0}^{\infty} f(p^h)p^{-hs}$. Since

$$\sum_{h=0}^{\infty} f(p^h)p^{-hs} = \sum_{h=0}^{\infty} f(p^h)(p^h)^{-s} = \sum_{\substack{n=p^h \\ h\geq 0}} f(n)n^{-s},$$

it follows by comparison with $F(s)$ that the series $G(s)$ is absolutely convergent. Therefore, it is convergent.

Also, using the fact that f is completely multiplicative and not identically zero, it is straightforward to check that the series $G(s)$ is a geometric series with the first term being 1 and the ratio being $f(p)p^{-s}$. Therefore,

$$\sum_{h=0}^{\infty} f(p^h)p^{-hs} = \sum_{n=0}^{\infty}(f(p)p^{-s})^n = \frac{1}{1-f(p)p^{-s}}\ .$$

The result follows upon application of Theorem 2.2.5.

Q.E.D.

Notice for the zeta function — which is absolutely convergent — that $f(n) = 1$ for all $n \in \mathbb{N}$. It follows immediately by Theorem 2.2.6 that

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

It is now proved that $\zeta(s) \neq 0$, by examining the individual terms of its re-expression as a product.

Theorem 2.2.7 *Suppose that $\sigma > 1$. Then $\zeta(s) \neq 0$.*

Proof. By our previous result,

$$|\zeta(s)| = \left|\prod_p \frac{1}{1-p^{-s}}\right| = \prod_p \frac{1}{|1-p^{-s}|}\ . \tag{2.18}$$

Now

$$\begin{aligned}
|1-p^{-s}| &= |1-p^{-\sigma}p^{-it}| \\
&= |1-p^{-\sigma}(\cos\log p^{-t} + i\sin\log p^{-t})| \\
&= |(1-p^{-\sigma}\cos\log p^{-t}) - ip^{-\sigma}\sin\log p^{-t}| \\
&= \sqrt{(1-p^{-\sigma}\cos\log p^{-t})^2 + (p^{-\sigma}\sin\log p^{-t})^2} \\
&= \sqrt{1-2p^{-\sigma}\cos\log p^{-t} + p^{-2\sigma}}
\end{aligned}$$

and, since $-\cos\log p^{-t} \leq 1$, it follows that

$$-2p^{-\sigma}\cos\log p^{-t} \leq 2p^{-\sigma}.$$

Therefore,

$$|1-p^{-s}| \leq \sqrt{1+2p^{-\sigma}+p^{-2\sigma}} = 1+p^{-\sigma}.$$

In other words,

$$\frac{1}{|1-p^{-s}|} \geq \frac{1}{1+p^{-\sigma}} \ .$$

It follows then from (2.18) that

$$|\zeta(s)| \geq \prod_p \frac{1}{1+p^{-\sigma}} \ . \tag{2.19}$$

Now $(1+p^{-\sigma})(1-p^{-\sigma}) = 1-p^{-2\sigma}$, where, since $\sigma > 1$, neither of the terms on the left are zero. In other words,

$$\frac{1}{1+p^{-\sigma}} = \frac{(1-p^{-2\sigma})^{-1}}{(1-p^{-\sigma})^{-1}} \ .$$

It follows then from (2.19) that

$$|\zeta(s)| \geq \prod_p \frac{(1-p^{-2\sigma})^{-1}}{(1-p^{-\sigma})^{-1}} = \frac{\prod_p (1-p^{-2\sigma})^{-1}}{\prod_p (1-p^{-\sigma})^{-1}} = \frac{\zeta(2\sigma)}{\zeta(\sigma)} \ ,$$

provided the denominator is not zero. Noting that $\zeta(2\sigma) = \sum_{n=1}^{\infty} n^{-2\sigma} > 0$, the proof is completed on also noting that $\zeta(\sigma) = \sum_{n=1}^{\infty} n^{-\sigma} > 0$.

Q.E.D.

Perhaps the next obvious step is to obtain $\zeta'(s)$. If term by term differentiation is possible, then the solution is obvious. The following proposition gives conditions for when this is possible. The proof can be found in [5] and is omitted here.

Proposition 2.2.8 *Suppose that the series* (2.8) *converges to $F(s)$ for $\sigma > \sigma_0$. Then $F(s)$ is analytic for such s, with derivative given by*

$$F'(s) = -\sum_{n=1}^{\infty} \frac{f(n)\log n}{n^s} \ .$$

Now, given the fact that $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ converges absolutely for every $s \in \mathbb{C}$ with $\sigma > 1$ and diverges for every real $s \leq 1$, it follows by Theorem 2.2.2 that $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ converges on the set $\{s \in \mathbb{C} : \sigma > 1\}$. Hence, by Proposition 2.2.8, $\zeta(s)$ is analytic in the domain $\sigma > 1$, and

$$\zeta'(s) = -\sum_{n=1}^{\infty} (\log n) n^{-s}.$$

Thus, given that $\zeta(s) \neq 0$, proving Theorem 2.2.1 amounts to proving that

$$\sum_{n=1}^{\infty} (\log n) n^{-s} = \sum_{n=1}^{\infty} \Lambda(n) n^{-s} \sum_{n=1}^{\infty} n^{-s},$$

as is duly done below.

Proof of Theorem 2.2.1. Recall Theorem 1.1.2,

$$\log n = \sum_{x|n} \Lambda(x) = \sum_{\substack{x,\, y \\ xy=n}} \Lambda(x)u(y)$$

where $u(y) = 1$ for all $y \in \mathbb{N}$. Thus by Theorem 2.2.3, for $\sigma > 1$,

$$\begin{aligned} \sum_{n=1}^{\infty} (\log n) n^{-s} &= \sum_{n=1}^{\infty} \Lambda(n) n^{-s} \sum_{n=1}^{\infty} u(n) n^{-s} \\ &= \sum_{n=1}^{\infty} \Lambda(n) n^{-s} \sum_{n=1}^{\infty} n^{-s}, \end{aligned} \tag{2.20}$$

provided that the series (2.9) and (2.11) converge absolutely for $\sigma > 1$. This is true for the series (2.9). Also, since $0 \leq \Lambda(n) \leq \log n$ for all $n \in \mathbb{N}$,

$$\sum_{n=1}^{\infty} \left| \frac{\Lambda(n)}{n^s} \right| = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^{\sigma}} \leq \sum_{n=1}^{\infty} \frac{\log n}{n^{\sigma}} < \infty \tag{2.21}$$

whenever $\sigma > 1$. Hence (2.20) holds for $\sigma > 1$.

Q.E.D.

2.3 $\psi_1(X)$ as a complex integral

Next, we prove a result that shall help us write $\psi_1(X)$ as a complex integral involving the Riemann zeta function. Cauchy's residue theorem is central to the argument. Before the residue theorem is stated (without proof), some definitions and results are first required. Along with the Cauchy residue theorem, these results can all be found in [1].

Definition 2.3.1 *A region is defined to be an open connected set.*

Definition 2.3.2 *A region D is simply connected if for any $z_0 \in \tilde{D}$ (the complement of D) and $\epsilon > 0$, there is a continuous curve $\gamma(t), 0 \leq t < \infty$ such that*

1) $d(\gamma(t), \tilde{D}) < \epsilon$ for all $t \geq 0$,

2) $\gamma(0) = z_0$,

3) $\lim_{t \longrightarrow \infty} \gamma(t) = \infty$,

where d is the shortest Euclidean distance between the curve and the region.

Definition 2.3.3 *A curve C is closed if its initial and terminal points coincide. C is a simple closed curve if no other points coincide.*

Definition 2.3.4 *Suppose that* γ *is a closed curve and that* $a \notin \gamma$*. Then*

$$n(\gamma, a) = \frac{1}{2\pi i} \int_\gamma \frac{dz}{z-a}$$

is called the winding number of γ *around* a*.*

The Jordan curve theorem asserts that any simple closed curve divides the plane into exactly two components — one bounded, the other unbounded. Furthermore, it can be shown that for these "Jordan" curves, if a is in the bounded component, then $n(\gamma, a) = \pm 1$. Otherwise, $n(\gamma, a) = 0$. If the curve is traversed so that the bounded component lies to the left of the curve, then $n(\gamma, a) = 0$ or 1.

Definition 2.3.5 *A complex function* f *is analytic at* z *if* f *is complex differentiable in a neighbourhood of* z*. Similarly, a complex function* f *is analytic on a set* S *if* f *is complex differentiable at all points of some open set containing* S*.*

Hereafter, the term *deleted neighbourhood of* z_0 denotes a set of the form $\{z : 0 < |z - z_0| < d\}$.

Definition 2.3.6 *A function* f *is said to have an isolated singularity at* z_0 *if* f *is analytic in a deleted neighbourhood* D *of* z_0 *but is not analytic at* z_0*.*

Definition 2.3.7 *Suppose that* f *has an isolated singularity at* z_0*.*

1) *If there exists a function* g*, analytic at* z_0 *and such that* $f(z) = g(z)$ *for all* z *in some deleted neighbourhood of* z_0, f *is said to have a removable singularity at* z_0*.*

2) *If, for* $z \neq z_0, f$ *can be written in the form* $f(z) = A(z)/B(z)$ *where* A *and* B *are analytic at* z_0*,* $A(z_0) \neq 0$*, and* $B(z_0) = 0, f$ *is said to have a pole at* z_0*.*

3) *If* f *has neither a removable singularity nor a pole at* z_0, f *is said to have an essential singularity at* z_0*.*

Definition 2.3.8 *If* $f(z) = \sum_{k=-\infty}^{\infty} C_k (z - z_0)^k$ *in a deleted neighbourhood of* z_0*, where* $C_k \in \mathbb{C}$ *for all* $k \in \mathbb{Z}$*,* C_{-1} *is called the residue of* f *at* z_0*. We use the notation* $C_{-1} = Res(f; z_0)$*.*

We are now in a position to state Cauchy's residue theorem. A proof can be found in [1].

Theorem 2.3.1 (Cauchy's residue theorem) *Suppose that* f *is analytic in a simply connected domain* D *except for isolated singularities at* $z_1, z_2, \ldots, z_m$*. Let* γ *be a closed curve not intersecting any of the singularities. Then*

$$\int_\gamma f(s)\mathrm{d}s = 2\pi i \sum_{k=1}^{m} n(\gamma, z_k) Res(f; z_k).$$

Also crucial to writing $\psi_1(X)$ as a complex integral are the following two standard results, again stated without proof.

Proposition 2.3.2 *For a curve C partitioned into n sections $C_1, \ldots, C_n$,*

$$\int_C f(s)\mathrm{d}s = \sum_{i=1}^{n} \int_{C_i} f(s)\mathrm{d}s.$$

Also,

$$-\int_C f(s)\mathrm{d}s = \int_{-C} f(s)\mathrm{d}s.$$

Theorem 2.3.3 (The M-L formula) *Suppose that C is a smooth curve of length L, that f is continuous on C, and that $|f(z)| \leq M$ for all $z \in C$. Then*

$$\left| \int_C f(z)\mathrm{d}z \right| \leq ML.$$

Note that a curve determined by $z : [a, b] \longrightarrow \mathbb{C}$ is said to be *smooth* if it is piecewise differentiable and $z'(t) \neq 0$ except at a finite number of points. We are now in a position to prove a result which will enable us to write $\psi_1(X)$ as a complex integral.

Theorem 2.3.4 *Suppose that $Y > 0$ and $c > 1$. Then*

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{Y^s}{s(s+1)} \, \mathrm{d}s = \begin{cases} 0 & \text{if } Y \leq 1, \\ 1 - \frac{1}{Y} & \text{if } Y \geq 1, \end{cases}$$

where the path of integration is the straight line $\sigma = c$.

Proof. Let $T > 1$, and write

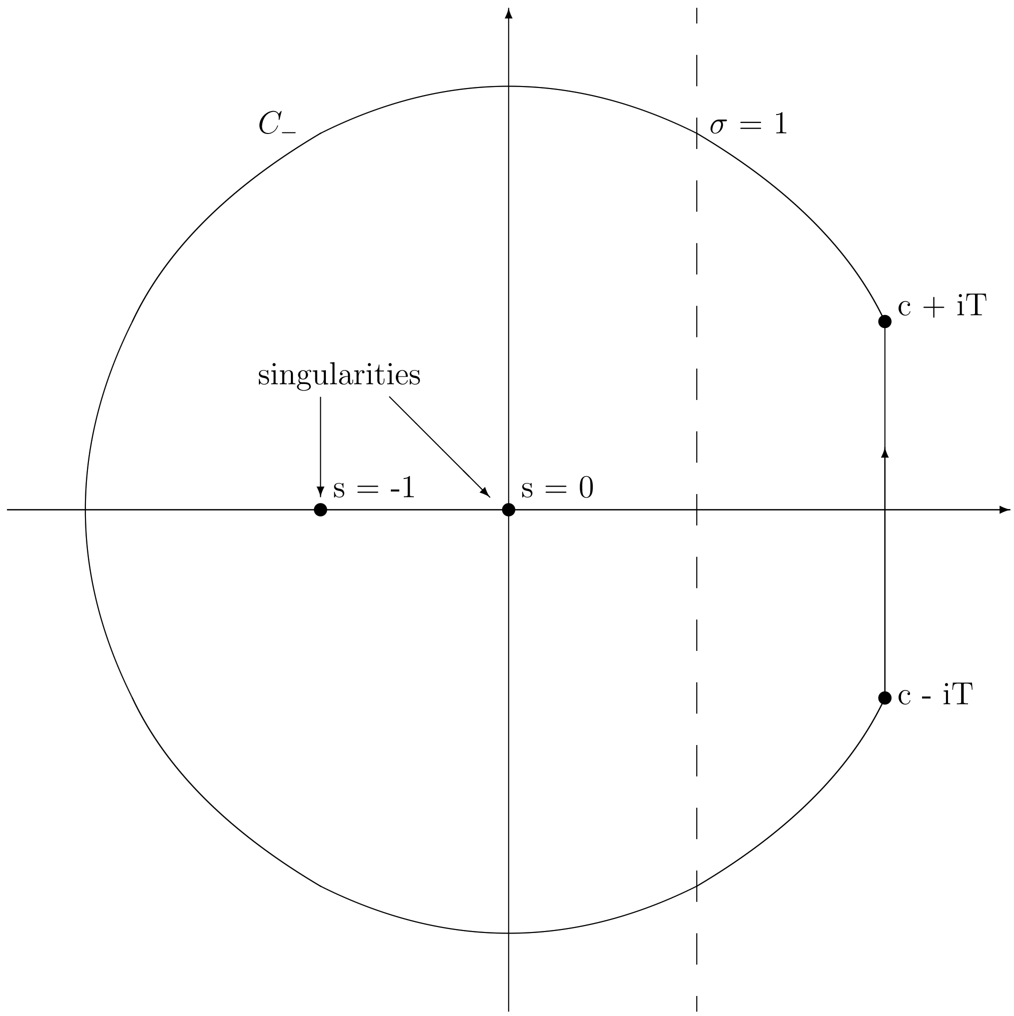
$$I_T = \frac{1}{2\pi i} \int_{c-iT}^{c+iT} \frac{Y^s}{s(s+1)} \, \mathrm{d}s$$

so that

$$\lim_{T \longrightarrow \infty} I_T = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{Y^s}{s(s+1)} \, \mathrm{d}s.$$

It is tempting to try and obtain this limit directly, but the integration is far from straightforward. Instead, notice that $Y^s/(s(s+1))$ is analytic in $\mathbb{C}$ except for isolated singularities at 0 and -1. Therefore, Cauchy's residue theorem may be applied to $Y^s/(s(s+1))$ and any closed curve in $\mathbb{C}$ not intersecting 0 or -1. Combined with Proposition 2.3.2, the integral I_T may be expressed in terms of other suitable integrals, winding numbers and residues. In general, the benefit of such an alternative expression is that the other integrals may be easier to evaluate, or in our case, where a limit is required, the limit of the alternative may be easier to obtain.

For the case $Y \geq 1$, take the simple closed curve C_- consisting of the line segment $[c - iT, c + iT]$, and the arc $A^-(c, T)$ of the circle centred at $s = 0$ with radius $\sqrt{c^2 + T^2}$ passing clockwise from $c - iT$ to $c + iT$, as the following diagram demonstrates.



Let

$$J_T^- = \frac{1}{2\pi i} \int_{A^-(c,T)} \frac{Y^s}{s(s+1)}\,\mathrm{d}s.$$

The singularities 0 and -1 both lie inside the bounded region. The curve C_- is simple, and the bounded region can be assumed to lie to the left of the curve, so the winding numbers are both 1. To find the residues, let $A(s) = Y^s$ and $B(s) = s(s+1)$. These functions are analytic at 0. Also, $A(0) = 1 \neq 0$ and $B(0) = 0$, while $B'(0) = 1 \neq 0$. Therefore, the singularity of $Y^s/(s(s+1))$ at $s = 0$ is a simple pole. Without proof, we state the following: If f has a simple pole at z_0, that is, if

$$f(z) = \frac{A(z)}{B(z)},$$

where A and B are analytic at z_0, $A(z_0) \neq 0$ and B has a simple zero at z_0, then

$$C_{-1} = \lim_{z \longrightarrow z_0} (z - z_0) f(z) = \frac{A(z_0)}{B'(z_0)} \ . \tag{2.22}$$

Hence,

$$Res\left(\frac{Y^s}{s(s+1)}, 0\right) = \frac{A(0)}{B'(0)} = 1.$$

Likewise, it can be shown that the singularity at -1 is a simple pole. In this case,

$$Res\left(\frac{Y^s}{s(s+1)}, -1\right) = -\frac{1}{Y} \ .$$

Therefore, by Cauchy's residue theorem and Proposition 2.3.2,

$$I_T = J_T^- + 1 - \frac{1}{Y} \ ,$$

so that

$$\lim_{T \longrightarrow \infty} I_T = \lim_{T \longrightarrow \infty} J_T^- + 1 - \frac{1}{Y} \ .$$

To complete the proof for the case $Y \geq 1$, it remains to show that $\lim_{T \longrightarrow \infty} J_T^- = 0$, or, equivalently, that $\lim_{T \longrightarrow \infty} |J_T^-| = 0$. Begin by noting that

$$|J_T^-| \leq \frac{1}{2\pi} \int_{A^-(c,T)} \frac{|Y^s|}{|s||s+1|} \, \mathrm{d}s.$$

Now for $s \in A^-(c,T)$, since $Y \geq 1$ and $\sigma \leq c$, we have

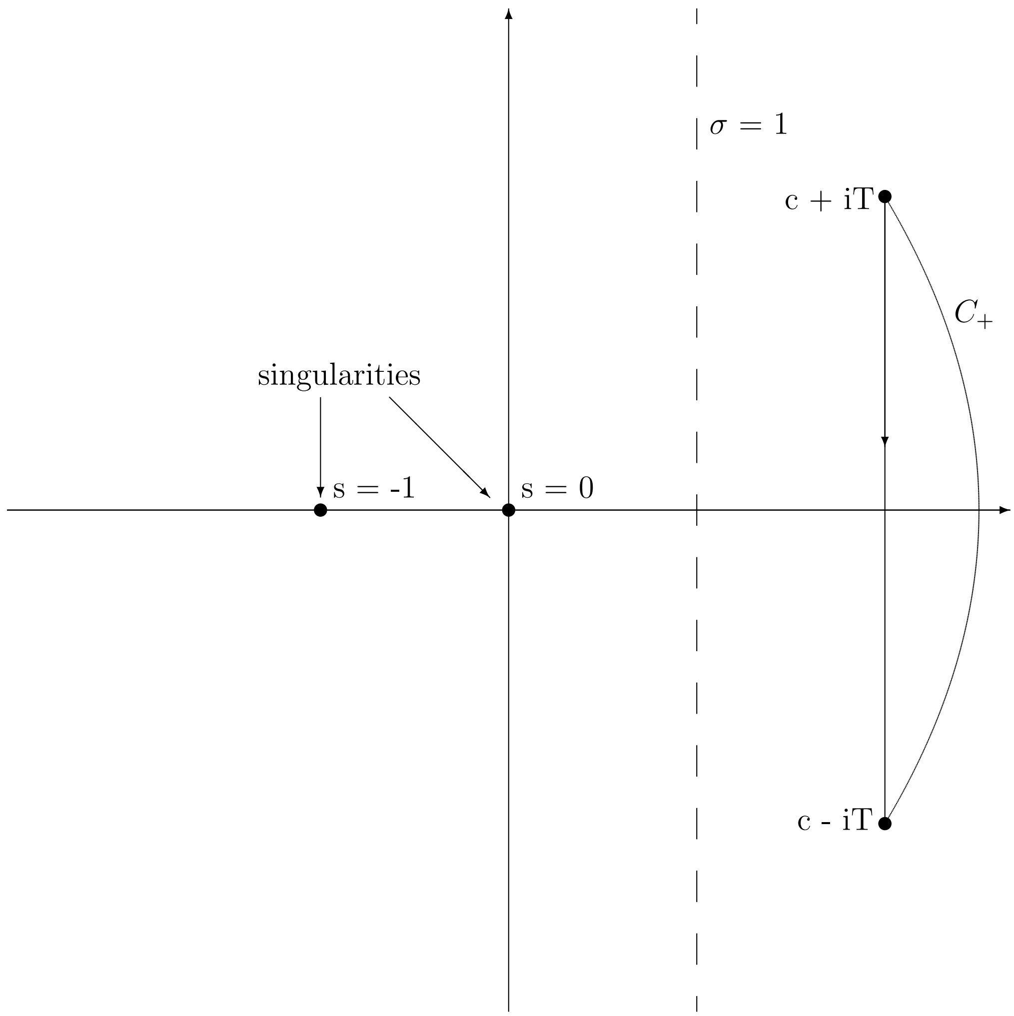
$$|Y^s| = Y^\sigma \leq Y^c.$$

Next, $|s| = R$, where R is the radius of $A^-(c,T)$, namely $\sqrt{c^2 + T^2}$. (This is why an arc around the origin is used.) Also, it is easy to show that $|s+1| \geq R - 1$ $(R - 1 > 0)$, or alternatively put, that $|s+1| \geq |s| - 1$. Combining these results, it follows that

$$\begin{aligned}
|J_T^-| &\leq \frac{1}{2\pi} \int_{A^-(c,T)} \frac{Y^c}{R(R-1)} \, \mathrm{d}s \\
&\leq \frac{1}{2\pi} \frac{Y^c}{R(R-1)} 2\pi R \\
&= \frac{Y^c}{R-1} \\
&\leq \frac{Y^c}{T-1}
\end{aligned}$$

since $R > T$, and where the M-L formula (Theorem 2.3.3) is used to obtain the second inequality. It follows that $\lim_{T \longrightarrow \infty} |J_T^-| = 0$, as required for the case $Y \geq 1$.

A very similar argument proves the case $Y \leq 1$. Take the simple closed curve C_+ consisting of the line segment $[c - iT, c + iT]$ and the arc $A^+(c, T)$ of the circle centred at $s = 0$ with radius $\sqrt{c^2 + T^2}$ passing anti-clockwise from $c - iT$ to $c + iT$, as the following diagram demonstrates.



Let

$$J_T^+ = \frac{1}{2\pi i} \int_{A^+(c,T)} \frac{Y^s}{s(s+1)} \, \mathrm{d}s.$$

There are no singularities in the associated bounded region. Therefore, by Cauchy's residue theorem and Proposition 2.3.2,

$$I_T = J_T^+,$$

so that

$$\lim_{T \longrightarrow \infty} I_T = \lim_{T \longrightarrow \infty} J_T^+.$$

Clearly then, the proof is completed if it is shown that $\lim_{T \longrightarrow \infty} J_T^+ = 0$, or, equivalently, that $\lim_{T \longrightarrow \infty} |J_T^+| = 0$. Proving this is very similar to proving that $\lim_{T \longrightarrow \infty} |J_T^-| = 0$. The details are left to the reader.

Q.E.D.

We now state and prove the result alluded to at the beginning of the section. Theorem 2.1.4, linking $\psi_1(X)$ to the von Mangoldt function, and Theorem 2.2.1, linking the von Mangoldt function to the Riemann zeta function are both used, with the von Mangoldt function serving in a sense as an intermediary function.

Theorem 2.3.5 *Suppose that $X > 0$ and $c > 1$. Then*

$$\psi_1(X) = -\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{X^{s+1}}{s(s+1)}\frac{\zeta'(s)}{\zeta(s)}\,\mathrm{d}s,$$

where the path of integration is the straight line $\sigma = c$.

Proof. Given Proposition 2.1.4, it follows that

$$\frac{\psi_1(X)}{X} = \sum_{n=1}^{[X]}\left(1 - \frac{1}{X/n}\right)\Lambda(n).$$

The sum on the right is over $n \leq X$, hence the sum is over n such that $X/n \geq 1$. Therefore, given Theorem 2.3.4, it follows that

$$\frac{\psi_1(X)}{X} = \sum_{n=1}^{[X]}\frac{\Lambda(n)}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{(X/n)^s}{s(s+1)}\,\mathrm{d}s.$$

Furthermore, applying the same theorem,

$$\sum_{n=[X]+1}^{\infty}\frac{\Lambda(n)}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{(X/n)^s}{s(s+1)}\,\mathrm{d}s = 0,$$

since this sum is over n such that $X/n \leq 1$. It follows that

$$\begin{aligned}\frac{\psi_1(X)}{X} &= \sum_{n=1}^{\infty}\frac{\Lambda(n)}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{(X/n)^s}{s(s+1)}\,\mathrm{d}s \\ &= \frac{1}{2\pi i}\sum_{n=1}^{\infty}\int_{c-i\infty}^{c+i\infty}\frac{X^s}{s(s+1)}\frac{\Lambda(n)}{n^s}\,\mathrm{d}s. \qquad (2.23)\end{aligned}$$

The order of summation and integration can be changed if

$$\sum_{n=1}^{\infty}\left|\int_{c-i\infty}^{c+i\infty}\frac{X^s}{s(s+1)}\frac{\Lambda(n)}{n^s}\,\mathrm{d}s\right| < \infty.$$

First note that

$$\sum_{n=1}^{\infty}\left|\int_{c-i\infty}^{c+i\infty}\frac{X^s}{s(s+1)}\frac{\Lambda(n)}{n^s}\,\mathrm{d}s\right| \le \sum_{n=1}^{\infty}\int_{c-i\infty}^{c+i\infty}\frac{|X^s|}{|s||s+1|}\frac{|\Lambda(n)|}{|n^s|}\,\mathrm{d}s.$$

Since $c > 1$, it follows that $|s+1| > |s|$. Thus $|s+1||s| > |s|^2 = c^2+t^2$. Also, $|X^s| = X^c, |n^s| = n^c$ and $|\Lambda(n)| = \Lambda(n)$. Therefore,

$$\begin{aligned}\sum_{n=1}^{\infty}\left|\int_{c-i\infty}^{c+i\infty}\frac{X^s}{s(s+1)}\frac{\Lambda(n)}{n^s}\,\mathrm{d}s\right| &\le X^c\sum_{n=1}^{\infty}\frac{\Lambda(n)}{n^c}\int_{c-i\infty}^{c+i\infty}\frac{\mathrm{d}s}{c^2+t^2}\\ &= \left(X^c\int_{-\infty}^{+\infty}\frac{\mathrm{d}t}{c^2+t^2}\right)\sum_{n=1}^{\infty}\frac{\Lambda(n)}{n^c}\\ &\le \left(X^c\int_{-\infty}^{+\infty}\frac{\mathrm{d}t}{t^2}\right)\sum_{n=1}^{\infty}\frac{\Lambda(n)}{n^c}\ .\end{aligned}$$

Note that the integral is finite, and by (2.21), so too is the series. Therefore,

$$\sum_{n=1}^{\infty}\left|\int_{c-i\infty}^{c+i\infty}\frac{X^s}{s(s+1)}\frac{\Lambda(n)}{n^s}\,\mathrm{d}s\right| < \infty.$$

Hence, from (2.23), it follows that

$$\begin{aligned}\frac{\psi_1(X)}{X} &= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\sum_{n=1}^{\infty}\frac{X^s}{s(s+1)}\frac{\Lambda(n)}{n^s}\,\mathrm{d}s\\ &= \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{X^s}{s(s+1)}\sum_{n=1}^{\infty}\frac{\Lambda(n)}{n^s}\,\mathrm{d}s\end{aligned}$$

and the result follows by Theorem 2.2.1.

Q.E.D.

2.4 An analytic continuation of $\zeta(s)$

Let

$$G(s) = \frac{-1}{s(s+1)}\frac{\zeta'(s)}{\zeta(s)}\ .$$

Then by Theorem 2.3.5, the statement $\psi_1(X) \sim \frac{1}{2}X^2$ as $X \longrightarrow \infty$ is implied for any $c > 1$ by the statement

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}G(s)X^{s-1}\,\mathrm{d}s \sim \frac{1}{2} \quad \text{as} \quad X \longrightarrow \infty.$$

In light of the proof of Theorem 2.3.4, Cauchy's residue theorem may prove useful in finding an alternative expression for

$$\frac{1}{2\pi i}\int_{c-iU}^{c+iU} G(s)X^{s-1}\,\mathrm{d}s,$$

and thus enable us to investigate its limit

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} G(s)X^{s-1}\,\mathrm{d}s.$$

Again in light of Theorem 2.3.4, to find this limit, bounds for $G(s)$ and $G(s)X^{s-1}$ may be required. First, simply connected domains must be found where $G(s)X^{s-1}$ is analytic, except for a finite number of singularities, so that the residue theorem can be applied. Recall that $\zeta(s)$ is analytic in the region $\sigma > 1$. The following two results allow one to conclude that $\zeta'(s)$ is also analytic for $\sigma > 1$. The first is merely stated, with a proof available in [1].

Theorem 2.4.1 *Let f be analytic at z. Then f is infinitely differentiable at z.*

Theorem 2.4.2 *Let f be analytic on an open set S. Then f is infinitely differentiable on S.*

Proof. Choose any point $z \in S$. By Definition 2.3.5, f is analytic at z. Therefore, by the previous result, f is infinitely differentiable at z. Since z was an arbitrary point in S, f is infinitely differentiable on S.

Q.E.D.

Therefore, as $\zeta(s)$ is analytic on the region $\sigma > 1$, $\zeta'(s)$ is differentiable on the region $\sigma > 1$. In other words, $\zeta'(s)$ is analytic on the region. Now as $\zeta(s) \neq 0$ when $\sigma > 1$, $G(s)X^{s-1}$ is analytic on the whole of the region $\sigma > 1$. However, working with the residue theorem in this region proves difficult, so it is preferable to work in an extended domain where $G(s)X^{s-1}$ is also analytic.

It is sometimes possible to obtain an *analytic continuation* of a function, defined as follows:

Definition 2.4.1 *Suppose that f is an analytic function defined on an open subset U of the complex plane $\mathbb{C}$. If V is a larger open subset of $\mathbb{C}$ containing U, and F is an analytic function defined on V such that*

$$F(z) = f(z) \ \text{for all} \ z \in U,$$

then F is called an analytic continuation of f.

Thus, a function is extended beyond an analytic domain in such a way that the function is also analytic in the extended domain. This section establishes an analytic continuation of $\zeta(s)$ to the set $D = \{s \in \mathbb{C} : \sigma > 0, s \neq 1\}$, so that, by Theorem 2.4.2, $\zeta'(s)$ is analytic on D. Thus $G(s)X^{s-1}$ is analytic on the half-plane $\sigma > 0$, except for an isolated singularity at $s = 1$ and at the zeros of $\zeta(s)$.

Theorem 2.4.3 *The function $\zeta(s)$, analytic for $\sigma > 1$, admits an analytic continuation to the set $\{s \in \mathbb{C} : \sigma > 0, s \neq 1\}$, with $s = 1$ being a simple pole with residue 1.*

A lemma is first outlined, and two results, used in the proof of Theorem 2.4.3, but whose proofs would take us too far afield, are stated.

Lemma 2.4.4 *For any function f,*

$$\sum_{n=1}^{r}\sum_{m=n}^{r}\int_{m}^{m+1} f(x)\mathrm{d}x = \sum_{m=1}^{r} m \int_{m}^{m+1} f(x)\mathrm{d}x.$$

Proof. That the statement holds for $r = 1$ is trivial. We will proceed by induction. Assume then that

$$\sum_{n=1}^{k}\sum_{m=n}^{k}\int_{m}^{m+1} f(x)\mathrm{d}x = \sum_{m=1}^{k} m \int_{m}^{m+1} f(x)\mathrm{d}x.$$

Thus

$$\begin{aligned}
\sum_{n=1}^{k+1}\sum_{m=n}^{k+1}\int_{m}^{m+1} f(x)\mathrm{d}x &= \sum_{n=1}^{k+1}\sum_{m=n}^{k}\int_{m}^{m+1} f(x)\mathrm{d}x + \sum_{n=1}^{k+1}\int_{k+1}^{k+2} f(x)\mathrm{d}x \\
&= \sum_{n=1}^{k}\sum_{m=n}^{k}\int_{m}^{m+1} f(x)\mathrm{d}x + (k+1)\int_{k+1}^{k+2} f(x)\mathrm{d}x \\
&= \sum_{m=1}^{k} m\int_{m}^{m+1} f(x)\mathrm{d}x + (k+1)\int_{k+1}^{k+2} f(x)\mathrm{d}x \\
&= \sum_{m=1}^{k+1} m\int_{m}^{m+1} f(x)\mathrm{d}x.
\end{aligned}$$

So by induction, the statement is proved true for all $r \in \mathbb{N}$.

Q.E.D.

Theorem 2.4.5 *Suppose that the path Γ is defined by $w(t) = u(t)+iv(t)$, where $u(t), v(t) \in \mathbb{R}$ for every $t \in [0, 1]$. Suppose further that $u'(t)$ and $v'(t)$ are continuous on $[0, 1]$. Let D be a domain in $\mathbb{C}$. For every $s \in D$, let*

$$F(s) = \int_{\Gamma} f(s, w)\mathrm{d}w,$$

where

1) $f(s, w)$ is continuous for every $s \in D$ and every $w \in \Gamma$;

and

2) for every $w \in \Gamma$, the function $f(s,w)$ is analytic in D.

Then, $F(s)$ is analytic in D.

Theorem 2.4.6 *If a series of analytic functions is uniformly convergent on a domain D, then the series is analytic on D.*

Theorem 2.4.5 can be readily verified. For the reader interested in a proof of Theorem 2.4.6, see [7].

Proof of Theorem 2.4.3. Since $\zeta(s)$ diverges whenever $0 < \sigma \leq 1$, extending $\zeta(s)$ in the natural way is not an analytic continuation. Therefore, for $\sigma > 1$, we seek to express $\zeta(s)$ in a different manner, and prove that an analytic continuation is possible. The idea is to obtain an expression for the partial sums of the zeta function, and then take limits. Begin by noting that

$$\begin{aligned}
\sum_{n\leq X} n^{-s} - \sum_{n\leq X} X^{-s} &= \sum_{n\leq X}(n^{-s} - X^{-s}) = \sum_{n=1}^{[X]} \int_n^X sx^{-s-1}\mathrm{d}x \\
&= \sum_{n=1}^{[X]-1} \int_n^X sx^{-s-1}\mathrm{d}x + \int_{[X]}^X sx^{-s-1}\mathrm{d}x \\
&= \sum_{n=1}^{[X]-1} \int_n^{[X]} sx^{-s-1}\mathrm{d}x + \sum_{n=1}^{[X]-1} \int_{[X]}^X sx^{-s-1}\mathrm{d}x \\
&\qquad + \int_{[X]}^X sx^{-s-1}\mathrm{d}x \\
&= \sum_{n=1}^{[X]-1} \sum_{m=n}^{[X]-1} \int_m^{m+1} sx^{-s-1}\mathrm{d}x + \sum_{n=1}^{[X]} \int_{[X]}^X sx^{-s-1}\mathrm{d}x.
\end{aligned}$$

Given Lemma 2.4.4, it follows that

$$\begin{aligned}
\sum_{n\leq X} n^{-s} - \sum_{n\leq X} X^{-s} &= \sum_{m=1}^{[X]-1} m \int_m^{m+1} sx^{-s-1}\mathrm{d}x + [X] \int_{[X]}^X sx^{-s-1}\mathrm{d}x \\
&= \sum_{m=1}^{[X]-1} \int_m^{m+1} [x]sx^{-s-1}\mathrm{d}x + \int_{[X]}^X [x]sx^{-s-1}\mathrm{d}x \\
&= s\int_1^X [x]x^{-s-1}\mathrm{d}x.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sum_{n\leq X} n^{-s} &= [X]X^{-s} + s\int_1^X [x]x^{-s-1}\,\mathrm{d}x \\
&= (X-\{X\})X^{-s} + s\int_1^X x^{-s}\,\mathrm{d}x - s\int_1^X \{x\}x^{-s-1}\,\mathrm{d}x \\
&= \frac{1}{X^{s-1}} - \frac{\{X\}}{X^s} + s\left(\frac{-x^{-s+1}}{s-1}\right)\Bigg|_1^X - s\int_1^X \frac{\{x\}}{x^{s+1}}\,\mathrm{d}x \\
&= \frac{1}{X^{s-1}} - \frac{\{X\}}{X^s} + \frac{s}{s-1} - \frac{s}{(s-1)X^{s-1}} - s\int_1^X \frac{\{x\}}{x^{s+1}}\,\mathrm{d}x. \qquad (2.24)
\end{aligned}$$

Thus for $\sigma > 1$,

$$\zeta(s) = \lim_{X\longrightarrow\infty} \sum_{n\leq X} n^{-s} = \frac{s}{s-1} - s\int_1^\infty \frac{\{x\}}{x^{s+1}}\,\mathrm{d}x. \qquad (2.25)$$

Next, define a function

$$g(s) = \begin{cases} \frac{s}{s-1} - s\int_1^\infty \{x\}x^{-s-1}\mathrm{d}x & \text{if } s \neq 1 \\ 0 & \text{otherwise} \end{cases}$$

on the half-plane $\sigma > 0$, so that $g(s) = \zeta(s)$ for $\sigma > 1$. Assume, for the moment, that

$$\int_1^\infty \{x\}x^{-s-1}\mathrm{d}x \qquad (2.26)$$

is an analytic function for $\sigma > 0$. Then g is analytic on the half-plane $\sigma > 0$, except at $s = 1$. Let

$$A(s) = s - s(s-1)\int_1^\infty \{x\}x^{-s-1}\mathrm{d}x$$

and

$$B(s) = s - 1$$

so that

$$g(s) = \frac{A(s)}{B(s)},\ s \neq 1.$$

Note that A and B are analytic at 1. Also, $A(1) = 1$ and $B(1) = 0$, while $B'(1) = 1$. Therefore, g has a simple pole at $s = 1$, and

$$Res\,(g,1) = \frac{A(1)}{B'(1)} = 1.$$

Hence, g is an analytic continuation of ζ to the half-plane $\sigma > 0$ omitting the point $s = 1$, which is a simple pole with residue 1. Thus, it remains to prove (2.26). Note that we can write

$$\int_1^\infty \{x\}x^{-s-1}\mathrm{d}x = \sum_{n=1}^\infty F_n(s),$$

where for every $n \in \mathbb{N}$,

$$F_n(s) = \int_n^{n+1} \{x\} x^{-s-1} \mathrm{d}x.$$

Therefore, by Theorem 2.4.6, if

1) for every $n \in \mathbb{N}, F_n(s)$ is analytic in $\mathbb{C}$,

   and

2) for every $\delta > 0$, the series $\sum_{n=1}^{\infty} F_n(s)$ converges uniformly for $\sigma > \delta$,

then the series $\sum_{n=1}^{\infty} F_n(s)$ is analytic for $\sigma > \delta$. The series would then be analytic on the half-plane $\sigma > 0$, since $\delta > 0$ is arbitrary. Therefore, it remains to prove 1 and 2.

Proof of 1. Note that

$$\begin{aligned} F_n(s) &= \int_0^1 \{n+t\}(n+t)^{-s-1} \mathrm{d}t \\ &= \int_0^1 t(n+t)^{-s-1} \mathrm{d}t \\ &= \int_0^1 t e^{\log(n+t)^{-s-1}} \mathrm{d}t \\ &= \int_0^1 t e^{-(s+1)\log(n+t)} \mathrm{d}t. \end{aligned}$$

Write $f(s,w) = we^{-(s+1)\log(n+w)}$, and define Γ by $w(t) = t$ for every $t \in [0,1]$. Thus $w(t) = u(t) + iv(t)$, where $u(t) = t, v(t) = 0$. Hence, $u'(t)$ and $v'(t)$ are continuous on $[0,1]$. By definition,

$$\begin{aligned} \int_\Gamma f(s,w)dw &= \int_0^1 f(s,w(t))w'(t)\mathrm{d}t \\ &= \int_0^1 f(s,t)\mathrm{d}t \\ &= F_n(s). \end{aligned}$$

Note that $f(s,w)$ is continuous in $\mathbb{C}$ for every $s \in \mathbb{C}$ and every $w \in \Gamma$, and that for every $w \in \Gamma, f(s,w)$ is analytic in $\mathbb{C}$. Thus by Theorem 2.4.5, $F_n(s)$ is analytic in $\mathbb{C}$ for all $n \in \mathbb{N}$.

Proof of 2. Choose $\delta > 0$. Then for $\sigma > \delta$,

$$\begin{aligned} |F_n(s)| &\leq \int_n^{n+1} |\{x\}x^{-s-1}|\,\mathrm{d}x \\ &\leq \int_n^{n+1} |x^{-s-1}|\,\mathrm{d}x \\ &= \int_n^{n+1} x^{-\sigma-1}\mathrm{d}x \\ &\leq n^{-\sigma-1} \\ &< n^{-\delta-1}. \end{aligned}$$

Now choose $\epsilon > 0$. Since $\sum_{n=1}^{\infty} n^{-\delta-1}$ converges, it follows that there exists $N \in \mathbb{N}$ such that for all $m \geq N$,

$$\left| \sum_{n=m+1}^{\infty} F_n(s) \right| \leq \sum_{n=m+1}^{\infty} |F_n(s)| < \sum_{n=m+1}^{\infty} n^{-\delta-1} < \epsilon$$

for all $\sigma > \delta$. Thus the series $\sum_{n=1}^{\infty} F_n(s)$ converges uniformly for $\sigma > \delta$. This completes the proof.

Q.E.D.

2.5 A bound for $G(s)$, the zeros of $\zeta(s)$

In the previous section, it was stated that if the residue theorem could be used to find an alternative expression for

$$\frac{1}{2\pi i} \int_{c-iU}^{c+iU} G(s)X^{s-1}\,\mathrm{d}s$$

where $c > 1$, then bounds for $G(s)$ and $G(s)X^{s-1}$ might be required. To establish if the residue theorem can be used, recall also that $G(s)X^{s-1}$ was shown to be analytic on the half-plane $\sigma > 0$, except for an isolated singularity at $s = 1$ and at the zeros of $\zeta(s)$. Clearly then, the zeros are of interest, and the next two lemmas are about these zeros. The proof of the first is omitted, as it would take us too far afield.

Definition 2.5.1 *A zero of an analytic function f is said to be isolated if there exists a deleted neighbourhood of f that contains no zeros.*

Lemma 2.5.1 *The zeros of any non-zero analytic function are isolated.*

Lemma 2.5.2 *Suppose that f is analytic on a bounded region D, and on its boundary B. If f is not identically zero on $D \cup B$, then f has finitely many zeros in D.*

Proof. Assuming the contrary, there is an infinite sequence of distinct points $s_k \in D$ with $f(s_k) = 0$. This sequence has a convergent subsequence $\{w_k\}$ with

$$\lim_{k \longrightarrow \infty} w_k = w,$$

where $w \in D \cup B$. Since $f(w_k) = 0$ and f is continuous, $f(w) = 0$. Clearly, w is a zero that is not isolated, so by the last lemma, f is identically zero on $D \cup B$.

Q.E.D.

In light of all this, for any $\delta > 0$, any bounded region in the region $\sigma > \delta$ contains only a finite number of zeros of $\zeta(s)$. Clearly then, regions within bounded regions can be found with no zeros of $\zeta(s)$, which is desirable, as we have made no attempt to locate these zeros, much less find their residues. Therefore, in these regions, $G(s)X^{s-1}$ is analytic, except at 1, if it is in the region. Furthermore, as $G(s)X^{s-1}$ is analytic in the region $\sigma > 1$, regions can be found containing the line segment $[c - iU, c + iU]$ and $s = 1$ for which $G(s)X^{s-1}$ is analytic, except at 1. Hence, the residue theorem may be used, and bounds for $G(s)$ and $G(s)X^{s-1}$ may indeed prove useful. Clearly, bounds for $1/(s(s+1)), 1/\zeta(s)$ and $\zeta'(s)$ will provide a bound for $G(s)$.

It was also stated in the previous section that it is desirable to work outside of the region $\sigma > 1$, and in doing so, at least some segment of the line $\sigma = 1$ must be within the region we choose to work in. For this proof of the prime number theorem, the segment length is arbitrarily large, requiring that $\zeta(s)$ is examined along the line $\sigma = 1$. In this section, it is shown that $\zeta(s)$ has no zeros on the line $\sigma = 1$, allowing for the possibility of a bound on $1/\zeta(s)$ in the region $\sigma \geq 1$. First, bounds are obtained for $\zeta(s)$ and $\zeta'(s)$, where a variant of Cauchy's integral formula, known as the generalised Cauchy integral formula, is relied upon. This result is stated without proof.

Theorem 2.5.3 (The generalised Cauchy integral formula) Suppose that f is analytic in a simply connected domain D and that γ is a regular closed curve contained in D. Then for each z inside γ and $k = 0, 1, 2, ...$,

$$f^{(k)}(z) = \frac{k!}{2\pi i} \int_\gamma \frac{f(w)}{(w-z)^{k+1}} \mathrm{d}w.$$

(Indeed, the case $k = 0$, with D a disc and γ a circle within D with the same centre, is precisely Cauchy's integral formula.)

Theorem 2.5.4 *If $\sigma \geq 1$ and $t \geq 2$, then*

1) $\zeta(s) = O(\log t)$;

and

2) $\zeta'(s) = O(\log^2 t)$.

3) Suppose further that $0 < \delta < 1$. *If* $\sigma \geq \delta$ *and* $t \geq 1$, *then* $\zeta(s) = O_\delta(t^{1-\delta})$.

Notation: In 3, by $O_\delta(t^{1-\delta})$ we mean that $|\zeta(s)| \le Ct^{1-\delta}$ where C depends on δ.

Proof. If $\sigma > 0, t \ge 1$ and $X \ge 1$, then by (2.24) and (2.25),

$$\begin{aligned}\zeta(s) - \sum_{n\le X} \frac{1}{n^s} &= \frac{\{X\}}{X^s} - \frac{1}{X^{s-1}} + \frac{s}{(s-1)X^{s-1}} - s\int_X^\infty \frac{\{x\}}{x^{s+1}}\,\mathrm{d}x \\ &= \frac{\{X\}}{X^s} + \frac{1}{(s-1)X^{s-1}} - s\int_X^\infty \frac{\{x\}}{x^{s+1}}\,\mathrm{d}x. \qquad (2.27)\end{aligned}$$

Note that $X \ge 1$ so the sum is not empty. It follows that

$$|\zeta(s)| \le \left|\sum_{n\le X}\frac{1}{n^s}\right| + \left|\frac{\{X\}}{X^s}\right| + \left|\frac{1}{(s-1)X^{s-1}}\right| + \left|s\int_X^\infty \frac{\{x\}}{x^{s+1}}\,\mathrm{d}x\right|. \qquad (2.28)$$

Next, note that

$$\left|\sum_{n\le X}\frac{1}{n^s}\right| \le \sum_{n\le X}\left|\frac{1}{n^s}\right| = \sum_{n\le X}\frac{1}{n^\sigma}\ .$$

Also

$$\left|\frac{\{X\}}{X^s}\right| \le \left|\frac{1}{X^s}\right| = \frac{1}{X^\sigma}\ ,$$

$$\left|\frac{1}{(s-1)X^{s-1}}\right| = \frac{1}{|\ s-1|}\frac{1}{|X^{s-1}|} \le \frac{1}{tX^{\sigma-1}}$$

and

$$\begin{aligned}\left|s\int_X^\infty \frac{\{x\}}{x^{s+1}}\,\mathrm{d}x\right| &\le |s|\int_X^\infty \left|\frac{\{x\}}{x^{s+1}}\right|\mathrm{d}x \\ &\le \sqrt{\sigma^2+t^2}\int_X^\infty \frac{1}{x^{\sigma+1}}\,\mathrm{d}x \\ &\le (\sigma+t)\frac{1}{X^\sigma\sigma} \\ &= \left(1+\frac{t}{\sigma}\right)\frac{1}{X^\sigma}\ .\end{aligned}$$

Hence, it follows from (2.28) that

$$|\zeta(s)| \le \sum_{n\le X}\frac{1}{n^\sigma} + \frac{1}{tX^{\sigma-1}} + \left(2+\frac{t}{\sigma}\right)\frac{1}{X^\sigma}\ . \qquad (2.29)$$

If $\sigma \ge 1$, then

$$|\zeta(s)| \le \sum_{n\le X}\frac{1}{n} + 1 + \frac{2+t}{X}\ .$$

Now

$$\sum_{n \le X} \frac{1}{n} \le \log X + 1,$$

a proof of which is outlined in Chapter 3, Proposition 3.5.1. Hence

$$|\zeta(s)| \le (\log X + 1) + 3 + \frac{t}{X}\,.$$

This inequality holds for all $X \ge 1$. Therefore, choosing $X = t$, it follows that

$$|\zeta(s)| \le \log t + 5,$$

proving the first statement. On the other hand, supposing that $0 < \delta < 1$, if $\sigma \ge \delta$, $t \ge 1$ and $X \ge 1$, it follows from (2.29) that

$$|\zeta(s)| \le \sum_{n \le X} \frac{1}{n^{\delta}} + \frac{1}{X^{\delta-1}} + \left(2 + \frac{t}{\delta}\right)\frac{1}{X^{\delta}}\,.$$

Now,

$$\sum_{n \le X} \frac{1}{n^{\delta}} = \sum_{n=0}^{[X]-1} \frac{1}{(n+1)^{\delta}} \le \sum_{n=0}^{[X]-1} \int_n^{n+1} \frac{\mathrm{d}x}{x^{\delta}} \le \int_0^X \frac{\mathrm{d}x}{x^{\delta}} = \frac{X^{1-\delta}}{1-\delta}$$

and as $t > \delta$, it follows that $2 + t/\delta < 3t/\delta$. Therefore,

$$|\zeta(s)| \le \frac{X^{1-\delta}}{1-\delta} + X^{1-\delta} + \frac{3t}{\delta X^{\delta}},$$

and again choosing $X = t$, it follows that

$$|\zeta(s)| \le t^{1-\delta}\left(\frac{1}{1-\delta} + 1 + \frac{3}{\delta}\right), \tag{2.30}$$

proving the third statement.

To deduce the second statement, one could differentiate each term in (2.27) and proceed. However, this proves difficult. Instead, the idea is to choose an arbitrary point $s_0 = \sigma_0 + it_0$, where $\sigma_0 \ge 1$ and $t_0 \ge 2$, and use Theorem 2.5.3 to equate $\zeta'(s_0)$ to an integral involving $\zeta(s)$. The path of integration is the circle C centred at s_0 with radius $\rho < 1/2$. To begin, as $\zeta(s)$ is analytic within the circle C, it follows from Theorem 2.5.3 that

$$|\zeta'(s_0)| = \left|\frac{1}{2\pi i}\int_C \frac{\zeta(s)}{(s-s_0)^2}\,\mathrm{d}s\right| = \frac{1}{2\pi}\left|\int_C \frac{\zeta(s)}{(s-s_0)^2}\,\mathrm{d}s\right|.$$

Clearly, if $M = \sup_{s \in C} |\zeta(s)|$ exists, then

$$\left|\frac{\zeta(s)}{(s-s_0)^2}\right| \le \frac{M}{\rho^2}$$

and, as the length of $C = 2\pi\rho$, it follows by the M-L formula that

$$|\zeta'(s_0)| \le \frac{M}{\rho} . \tag{2.31}$$

We next show that M exists and find a bound. Note first that $0 < 1-\rho < 1$. Next, for every $s \in C$, we have $\sigma \ge \sigma_0 - \rho \ge 1-\rho$ and $t > t_0 - \rho > 1$. Also note that $t \le t_0 + 1/2 < 2t_0$. Substituting $1-\rho$ for δ in (2.30), it follows for every $s \in C$ that

$$\begin{aligned} |\zeta(s)| \le t^\rho\left(\frac{1}{\rho} + 1 + \frac{3}{1-\rho}\right) &\le 2^\rho t_0^\rho\left(\frac{1}{\rho} + \frac{1}{\rho} + \frac{3}{\rho}\right) \\ &\le 2t_0^\rho\left(\frac{5}{\rho}\right) \\ &= \frac{10t_0^\rho}{\rho} . \end{aligned}$$

Therefore, $M \le 10t_0^\rho/\rho$. By (2.31), it follows that

$$|\zeta'(s_0)| \le \frac{10t_0^\rho}{\rho^2} .$$

Let $\rho = (\log t_0 + 2)^{-1}$, so that $\rho < 1/2$ and $\rho < 1/\log t_0$. Then $t_0^\rho = e^{\rho \log t_0} < e$, hence

$$|\zeta'(s_0)| \le 10e(\log t_0 + 2)^2.$$

The second statement now follows.

Q.E.D.

The following proof must count as the most skillful encountered to date, and, in this author's opinion, surpasses the final argument for the proof of the prime number theorem. There seems little motive for much of what is done, but in the end, very different ideas are beautifully woven together.

Theorem 2.5.5 *The function $\zeta(s)$ has no zeros on the line $\sigma = 1$. Furthermore, as $t \longrightarrow \infty$, we have, for $\sigma \ge 1$, that*

$$\frac{1}{\zeta(s)} = O((\log t)^7).$$

Proof. Begin by noting that

$$\begin{aligned} 3 + 4\cos\theta + \cos 2\theta &= 3 + 4\cos\theta + \cos^2\theta - \sin^2\theta \\ &= 3 + 4\cos\theta - 1 + 2\cos^2\theta \\ &= 2(1 + 2\cos\theta + \cos^2\theta) \\ &= 2(1+\cos\theta)^2 \\ &\ge 0. \end{aligned} \tag{2.32}$$

Second, for $\sigma > 1$, and where the log of a complex number is taken to be the principal value, one might suspect that since

$$\log \zeta(s) = \log \left(\prod_p (1 - p^{-s})^{-1} \right),$$

it would follow that

$$\log \zeta(s) = \sum_p \log{(1 - p^{-s})}^{-1}. \tag{2.33}$$

Indeed, this is the case. To see why, it can be easily verified that for a complex sequence $\{a_n\}$ and some $m \in \mathbb{Z}$, and where $p[N]$ denotes the set of the first N primes,

$$\log \left(\prod_{p \in P[N]} a_p \right) = \sum_{p \in P[N]} \log a_p + (2\pi i)m.$$

for all $N \in \mathbb{N}$. Now assume that

$$\lim_{N \longrightarrow \infty} \left(\prod_{p \in P[N]} a_p \right) = L,$$

for some $L \in \mathbb{C}$. Then

$$\lim_{N \longrightarrow \infty} \log \left(\prod_{p \in P[N]} a_p \right) = \log L = \log \left(\lim_{N \longrightarrow \infty} \left(\prod_{p \in P[N]} a_p \right) \right),$$

or equivalently,

$$\lim_{N \longrightarrow \infty} \left(\sum_{p \in P[N]} \log a_p + (2\pi i)m \right) = \log \left(\prod_p a_p \right).$$

In other words,

$$\sum_p \log a_p + (2\pi i)m = \log \left(\prod_p a_p \right).$$

Letting $a_p = (1 - p^{-s})^{-1}$, and considering the case $s = 2$, where m equals zero, (2.33) is proved to be true.

We next consider the terms $\log(1 - p^{-s})^{-1}$. Using the substitution $x = p^{-s}$, note that

$$\log{(1 - x)}^{-1} = \int \frac{1}{1 - x} \mathrm{d}x + C_1 = \int \sum_{m=0}^{\infty} x^m \mathrm{d}x + C_1 = \sum_{m=0}^{\infty} \frac{x^{m+1}}{m + 1} + C_2 = \sum_{m=1}^{\infty} \frac{x^m}{m} + C_2,$$

where C_1 and C_2 are constants of integration. Considering the case $s = 2$ and any prime p, it follows that $C_2 = 0$. The last equality can therefore be written as

$$\log\left(1 - p^{-s}\right)^{-1} = \sum_{m=1}^{\infty} \frac{(p^{-s})^m}{m} . \tag{2.34}$$

for any s such that $\sigma > 1$ and any prime p. Hence, given (2.33),

$$\log \zeta(s) = \sum_{p} \sum_{m=1}^{\infty} \frac{(p^m)^{-s}}{m} .$$

By letting $\log(a + bi) = c + di$, it is easily verifiable that $\log|a + bi| = Re(\log(a + bi))$. By this result,

$$\log|\zeta(\sigma + it)| = Re(\log \zeta(\sigma + it)) = Re\left(\sum_{p} \sum_{m=1}^{\infty} \frac{(p^m)^{-\sigma - it}}{m}\right).$$

Now

$$\sum_{p} \sum_{m=1}^{\infty} \frac{(p^m)^{-\sigma - it}}{m} = \sum_{\substack{p \\ n = p^m}} \frac{n^{-\sigma - it}}{m} .$$

Thus, if

$$c_n = \begin{cases} \frac{1}{m} & \text{if } n = p^m, p \text{ prime}, m \in \mathbb{N} \\ 0 & \text{otherwise,} \end{cases}$$

we have

$$\sum_{p} \sum_{m=1}^{\infty} \frac{(p^m)^{-\sigma - it}}{m} = \sum_{n=2}^{\infty} c_n n^{-\sigma - it}.$$

Therefore,

$$\begin{aligned} \log|\zeta(\sigma + it)| &= Re\left(\sum_{n=2}^{\infty} c_n n^{-\sigma - it}\right) \\ &= \sum_{n=2}^{\infty} Re\left(c_n n^{-\sigma} n^{-it}\right) \\ &= \sum_{n=2}^{\infty} c_n n^{-\sigma} Re(n^{-it}) \\ &= \sum_{n=2}^{\infty} c_n n^{-\sigma} \cos(t \log n). \end{aligned} \tag{2.35}$$

Combining (2.32) and (2.35), it follows that

$$\begin{aligned}
\log|\zeta^3(\sigma)\zeta^4(\sigma+it)\zeta(\sigma+2it)| &= 3\log|\zeta(\sigma)| + 4\log|\zeta(\sigma+it)| + \log|\zeta(\sigma+2it)| \\
&= 3\sum_{n=2}^{\infty} c_n n^{-\sigma}\cos 0 + 4\sum_{n=2}^{\infty} c_n n^{-\sigma}\cos(t\log n) \\
&\qquad + \sum_{n=2}^{\infty} c_n n^{-\sigma}\cos(2t\log n) \\
&= \sum_{n=2}^{\infty} c_n n^{-\sigma}(3 + 4\cos(t\log n) + \cos(2t\log n)) \\
&\geq 0.
\end{aligned}$$

It follows that $|\zeta^3(\sigma)\zeta^4(\sigma+it)\zeta(\sigma+2it)| \geq 1$, hence for $\sigma > 1$,

$$|(\sigma-1)\zeta(\sigma)|^3 \left|\frac{\zeta(\sigma+it)}{\sigma-1}\right|^4 |\zeta(\sigma+2it)| \geq \frac{1}{\sigma-1}\ . \tag{2.36}$$

Assume that $s = 1+it$ is a zero of $\zeta(s)$, i.e. assume that $\zeta(1+it) = 0$. Since $\zeta(s)$ is undefined at $s = 1$, clearly $t \neq 0$. As $\zeta(s)$ is analytic elsewhere on the half-plane $\sigma > 0$, $\zeta(s)$ is analytic at the points $s = 1+it$ and $s = 1+2it$. Thus, as analyticity at a point implies continuity,

$$\lim_{\sigma \longrightarrow 1+} \zeta(\sigma+2it) = \zeta(1+2it). \tag{2.37}$$

Also, by the analyticity of $\zeta(s)$ at $1+it$,

$$\lim_{\substack{h \longrightarrow 0 \\ h \in \mathbb{C}}} \frac{\zeta(1+it+h) - \zeta(1+it)}{h}$$

exists. Since $\zeta(1+it) = 0$, it follows that

$$\lim_{h \longrightarrow 0+} \frac{\zeta(1+it+h)}{h}$$

exists. In other words,

$$\lim_{\sigma \longrightarrow 1+} \frac{\zeta(\sigma+it)}{\sigma-1} \tag{2.38}$$

exists. Furthermore, $\zeta(s)$ has a simple pole at $s = 1$ with residue $C_{-1} = 1$. Therefore, recalling (2.22),

$$\lim_{\sigma \longrightarrow 1+} (\sigma-1)\zeta(\sigma) = C_{-1} = 1. \tag{2.39}$$

By (2.37), (2.38) and (2.39), the left hand side of (2.36) must converge to a finite limit as $\sigma \longrightarrow 1+$, meaning that the right hand side must also converge as $\sigma \longrightarrow 1+$, which is not the case. Hence, $s = 1+it$ cannot be a zero of $\zeta(s)$.

To prove the second assertion, recall from Theorem 2.2.7 that for $\sigma > 1$,

$$\left|\frac{1}{\zeta(s)}\right| \leq \frac{\zeta(\sigma)}{\zeta(2\sigma)} .$$

Thus as $\zeta(2\sigma) > 1$, it follows that for $\sigma \geq 2$,

$$\left|\frac{1}{\zeta(s)}\right| < \zeta(\sigma) \leq \zeta(2),$$

and as $\zeta(2)$ is a constant, it will now be assumed, without loss of generality, that $1 \leq \sigma < 2$. First suppose that $1 < \sigma < 2$ and $t \geq 2$. It follows from (2.36) that

$$(\sigma - 1)^3 \leq |(\sigma - 1)\zeta(\sigma)|^3 |\zeta(\sigma + it)|^4 |\zeta(\sigma + 2it)|.$$

Recall that $\lim_{\sigma \longrightarrow 1+}(\sigma - 1)\zeta(\sigma) = 1$ and note that when $\sigma = 2$, we have $(\sigma - 1)\zeta(\sigma) = \zeta(2)$. This being the case, as $(\sigma - 1)\zeta(\sigma)$ is continuous, $(\sigma - 1)\zeta(\sigma)$ is bounded by a constant. Together with Theorem 2.5.4 (1), it follows that

$$(\sigma - 1)^3 \leq A_1 |\zeta(\sigma + it)|^4 \log 2t \leq A_1 |\zeta(\sigma + it)|^4 2 \log t$$

for some positive constant A_1. It easily follows that

$$|\zeta(\sigma + it)| \geq \frac{(\sigma - 1)^{3/4}}{A_2 (\log t)^{1/4}}, \tag{2.40}$$

where A_2 is a positive constant. Clearly, (2.40) is also true when $\sigma = 1$.

Next, suppose that $1 < \eta < 2$. If $1 \leq \sigma \leq \eta$ and $t \geq 2$, then it follows from Theorem 2.5.4 (2) that

$$\begin{aligned} |\zeta(\sigma + it) - \zeta(\eta + it)| &= \left| \int_\sigma^\eta \zeta'(x + it) \mathrm{d}x \right| \\ &\leq \int_\sigma^\eta |\zeta'(x + it)| \, \mathrm{d}x \\ &\leq \int_\sigma^\eta A_3 \log^2 t \, \mathrm{d}x \\ &= A_3 (\eta - \sigma) \log^2 t \\ &\leq A_3 (\eta - 1) \log^2 t, \end{aligned}$$

where A_3 is a positive constant. Combining this with (2.40), it follows that

$$|\zeta(\sigma + it)| \geq |\zeta(\eta + it)| - A_3(\eta - 1)\log^2 t \geq \frac{(\eta - 1)^{3/4}}{A_2 (\log t)^{1/4}} - A_3 (\eta - 1) \log^2 t. \tag{2.41}$$

On the other hand, if $\eta \le \sigma \le 2$ and $t \ge 2$, then

$$\frac{(\sigma-1)^{3/4}}{A_2(\log t)^{1/4}} \ge \frac{(\eta-1)^{3/4}}{A_2(\log t)^{1/4}},$$

and in view of (2.40), (2.41) must also hold. Therefore, (2.41) holds if $1 \le \sigma \le 2, t \ge 2$. Choose $\eta < 2$ so that

$$\frac{(\eta-1)^{3/4}}{A_2(\log t)^{1/4}} = 2A_3(\eta-1)\log^2 t,$$

in other words, choose

$$\eta = 1 + (2A_2A_3)^{-4}(\log t)^{-9},$$

where $(2A_2A_3)^{-4}(\log t)^{-9} < 1$. This is true for all sufficiently large t, say $t > t_0 \ge 2$. Thus for $1 \le \sigma \le 2$ and $t \ge t_0$,

$$|\zeta(\sigma+it)| \ge A_3(\eta-1)\log^2 t = A_4(\log t)^{-7},$$

where A_4 is a positive constant. The result follows.

Q.E.D.

We are now in a position to obtain a bound for

$$G(s) = \frac{1}{s(s+1)}\zeta'(s)\frac{1}{\zeta(s)} .$$

Begin by noting that

$$\frac{1}{s(s+1)} = O(t^{-2}).$$

Also, by Theorem 2.5.4 (2), $\zeta'(s) = O(\log^2 t)$ for $\sigma \ge 1$ and $t > t_0$, since $t_0 \ge 2$. Next, by Theorem 2.5.5,

$$\frac{1}{\zeta(s)} = O((\log t)^7),$$

for $\sigma \ge 1$ and $t > t_0$. Thus

$$\begin{aligned} G(s) &= O(t^{-2}(\log t)^2(\log t)^7) \\ &= O(t^{-2}(\log t)^9) \end{aligned}$$

for $\sigma \ge 1$ and $t > t_0$. Since $(\log t)^9 = O(t^{1/2})$ for sufficiently large t, it follows that

$$G(s) = O(t^{-3/2}).$$

2.6 Completion of the analytic proof

As proved in section 2.1, to prove the prime number theorem, it suffices to prove Theorem 2.1.3, namely that

$$\psi_1(X) \sim \frac{1}{2}X^2 \quad \text{as} \quad X \longrightarrow \infty.$$

It was later shown that the statement

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} G(s)X^{s-1}\,\mathrm{d}s \sim \frac{1}{2} \quad \text{as} \quad X \longrightarrow \infty$$

implied Theorem 2.1.3 for any $c > 1$, and Cauchy's residue theorem was mentioned as a possible way to investigate the expression on the left via an alternative expression of the term

$$\frac{1}{2\pi i}\int_{c-iU}^{c+iU} G(s)X^{s-1}\mathrm{d}s$$

for $U \in \mathbb{R}$. Regions containing the line segment $[c-iU, c+iU]$ and $s = 1$ were proved to exist where $G(s)X^{s-1}$ is analytic, except at 1. In the last section it was proved that the line $\sigma = 1$ has no zeros, so in fact, regions containing $[c-iU, c+iU]$, $s = 1$, and a segment of $\sigma = 1$ exist where $G(s)X^{s-1}$ is analytic, except at 1. To prove Theorem 2.1.3, the idea is to take such a region and apply the residue theorem to $G(s)X^{s-1}$ and a simple closed curve within that region, where the curve includes the line segment $[c - iU, c + iU]$. A curve composed of line segments, and more specifically, vertical and horizontal line segments, is easiest to work with.

Proof of Theorem 2.1.3.

First, let $\epsilon > 0$ be given, and consider a simple closed curve γ, with direction such that the bounded region lies to the left of γ, made up of the straight line segments

$$\begin{cases} L_1 = [c - iU, 1 - iU], \\ L_2 = [1 - iU, 1 - iT], \\ L_3 = [1 - iT, \alpha - iT], \\ L_4 = [\alpha - iT, \alpha + iT], \\ L_5 = [\alpha + iT, 1 + iT], \\ L_6 = [1 + iT, 1 + iU], \\ L_7 = [1 + iU, c + iU], \\ L_8 = [c + iU, c - iU], \end{cases}$$

where $T = T(\epsilon) > t_0, \alpha = \alpha(T) = \alpha(\epsilon) \in (1/2, 1), U$ and $\beta = \beta(U) \in (0, 1)$ satisfy the following conditions:

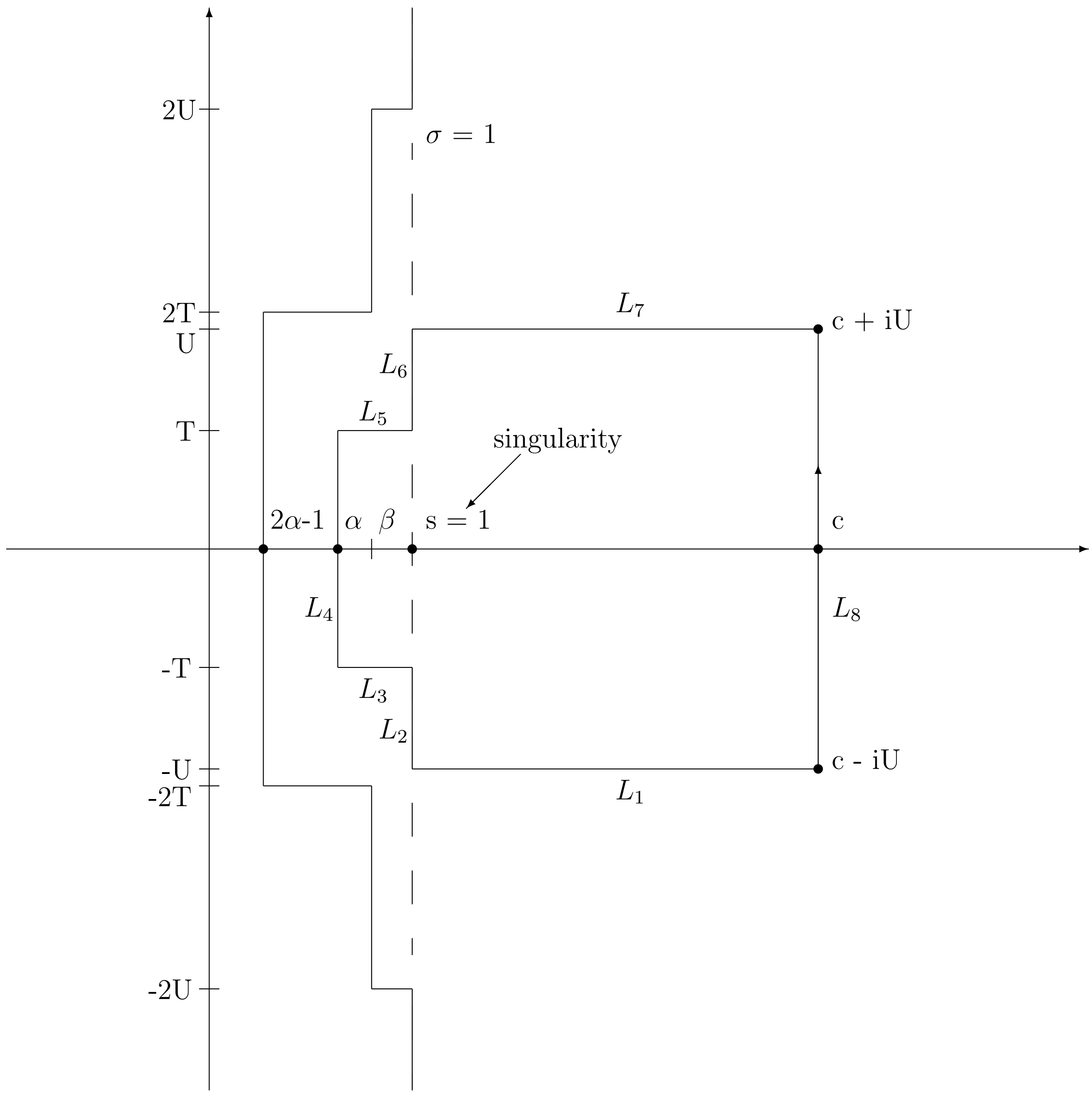
1) $\int_T^\infty |G(1 + it)|\mathrm{d}t < \epsilon$;

2) the set $A = (2\alpha - 1, 1] \times (-2T, 2T)$ contains no zeros of $\zeta(s)$;

3) $U > T$;

4) the set $B = (\beta, 1] \times (-2U, 2U)$ contains no zeros of $\zeta(s)$.

The following diagram demonstrates this setup.



Some nuances of the setup are best discussed after the proof. The first and third conditions allow us to estimate integrals along the line $\sigma = 1$. To see that such a T exists, recall that $G(s) = O(t^{-3/2})$ for all $t > t_0$. In other words, there exists $R \in \mathbb{R}$ such that for all $t > t_0$,

$$|G(s)| < Rt^{-3/2}.$$

Therefore, it suffices to show that there exists $T > t_0$ such that $\int_T^\infty t^{-3/2}\mathrm{d}t < \epsilon$, which is a straightforward exercise. The second condition is possible since $\zeta(s)$ has no zeros on the line $\sigma > 1$, and α can be chosen so that the bounded region $(\alpha/2, 1) \times (-2T, 2T)$ contains no zeros of $\zeta(s)$. The fourth condition is possible in the same way. It follows that $G(s)X^{s-1}$ is analytic on the union of A, B and the region $\sigma > 1$, except for an isolated singularity at $s = 1$. This union, call it D, is a simply connected region, and D contains γ, with the singularity $s = 1$ within the region bounded by γ. Therefore, by Cauchy's residue theorem,

$$\frac{1}{2\pi i}\int_{c-iU}^{c+iU} G(s)X^{s-1}\mathrm{d}s = \sum_{j=1}^{7} \frac{1}{2\pi i}\int_{L_j} G(s)X^{s-1}\mathrm{d}s + Res(G(s)X^{s-1}, 1).$$

To calculate the residue, note that by Theorem 2.4.3, $\zeta(s)$ has a simple pole at $s = 1$ with residue 1. Therefore,

$$\zeta(s) = \frac{A(s)}{B(s)},\ s \neq 1,$$

where A and B are analytic, $A(1) \neq 0, B(1) = 0$ and $B'(1) \neq 0$. Thus for $X > 1$,

$$G(s)X^{s-1} = -\frac{1}{s(s+1)}\frac{\zeta'(s)}{\zeta(s)}X^{s-1} = \frac{(B'(s)A(s) - A'(s)B(s))X^{s-1}}{(s^2+s)B(s)A(s)},\ s \neq 1.$$

Letting $C(s) = (B'(s)A(s) - A'(s)B(s))X^{s-1}$ and $D(s) = (s^2 + s)B(s)A(s)$, it follows that

$$G(s)X^{s-1} = \frac{C(s)}{D(s)}, s \neq 1,$$

where clearly, C and D are analytic, and it is easy to show that $C(1) = B'(1)A(1) \neq 0,\ D(1) = 0$ and $D'(1) = 2B'(1)A(1) \neq 0$. Therefore, $G(s)X^{s-1}$ has a simple pole at $s = 1$ with

$$Res(G(s)X^{s-1}, 1) = \frac{C(1)}{D'(1)} = 1/2.$$

It follows that, for every $X > 1$,

$$\left|\frac{1}{2\pi i}\int_{c-iU}^{c+iU} G(s)X^{s-1}\mathrm{d}s - \frac{1}{2}\right| \leq \sum_{j=1}^{7}\left|\frac{1}{2\pi i}\int_{L_j} G(s)X^{s-1}\mathrm{d}s\right|. \tag{2.42}$$

Notice the symmetry of the curve about the real axis. Choose any simple closed curve in D that includes both L_1 and L_7 in its composition, and for which $s = 1$ lies outside the simply connected region D' bounded by the curve. Then, $G(s)X^{s-1}$ is analytic in D'. A relevant theorem is stated, the proof of which can be found in [1].

Theorem 2.6.1 *If a function f is analytic in a simply connected region D, there exists a function F, analytic in D such that $F' = f$.*

Thus there exists a function F, analytic in D' such that $F'(s) = G(s)X^{s-1}$. Since L_1 and L_7 are smooth curves, a generalisation of the Fundamental Theorem of Calculus applies. The result is merely stated here, with a proof to be found in [1].

Proposition 2.6.2 *If C is a smooth curve with initial point α and terminal point β, and F is an analytic function on C such that $F' = f$, then $\int_C f(z)\mathrm{d}z = F(\beta) - F(\alpha)$.*

It follows that

$$\begin{aligned} \left|\int_{L_1} G(s)X^{s-1}\mathrm{d}s\right| &= \left|\int_{c-iU}^{1-iU} F'(s)\mathrm{d}s\right| \\ &= |F(1-iU) - F(c-iU)|. \end{aligned}$$

Note that since F is an analytic function, $\overline{F(s)} = F(\overline{s})$. Hence

$$\begin{aligned} \left|\int_{L_1} G(s)X^{s-1}\mathrm{d}s\right| &= |F(\overline{1+iU}) - F(\overline{c+iU})| \\ &= |\overline{F(1+iU)} - \overline{F(c+iU)}| \\ &= |\overline{F(1+iU) - F(c+iU)}| \\ &= |F(1+iU) - F(c+iU)| \\ &= \left|\int_{L_7} G(s)X^{s-1}\mathrm{d}s\right|. \end{aligned}$$

Likewise,

$$\left|\int_{L_2} G(s)X^{s-1}\mathrm{d}s\right| = \left|\int_{L_6} G(s)X^{s-1}\mathrm{d}s\right|$$

and

$$\left|\int_{L_3} G(s)X^{s-1}\mathrm{d}s\right| = \left|\int_{L_5} G(s)X^{s-1}\mathrm{d}s\right|.$$

Therefore,

$$\begin{aligned} \left|\int_{L_2} G(s)X^{s-1}\mathrm{d}s\right| &= \left|\int_{L_6} G(s)X^{s-1}\mathrm{d}s\right| \\ &= \left|\int_{1+iT}^{1+iU} G(s)X^{s-1}\mathrm{d}s\right| \\ &= \left|\int_{T}^{U} G(1+it)X^{it}\mathrm{d}t\right| \\ &\leq \int_{T}^{U} |G(1+it)|\mathrm{d}t \\ &< \epsilon. \end{aligned} \tag{2.43}$$

On the other hand,

$$\begin{aligned} \left|\int_{L_3} G(s)X^{s-1}\mathrm{d}s\right| &= \left|\int_{L_5} G(s)X^{s-1}\mathrm{d}s\right| \\ &\leq \int_{\alpha+iT}^{1+iT} |G(s)||X^{s-1}|\mathrm{d}s. \end{aligned}$$

But as G is analytic on $L_3 \cup L_4 \cup L_5$,

$$\sup_{L_3 \cup L_4 \cup L_5} |G(s)|$$

is defined. Call this supremum M. Thus,

$$\begin{aligned}
\left|\int_{L_3} G(s)X^{s-1}\mathrm{d}s\right| &= \left|\int_{L_5} G(s)X^{s-1}\mathrm{d}s\right| \\
&\leq M\int_{\alpha}^{1} |X^{\sigma+iT-1}|\mathrm{d}\sigma \\
&= M\int_{\alpha}^{1} |X^{\sigma-1}|\mathrm{d}\sigma \\
&\leq \frac{M}{\log X},
\end{aligned} \tag{2.44}$$

the last inequality being a trivial matter once it is recalled that $X > 0$. Also,

$$\begin{aligned}
\left|\int_{L_4} G(s)X^{s-1}\mathrm{d}s\right| &= \left|\int_{\alpha-iT}^{\alpha+iT} G(s)X^{s-1}\mathrm{d}s\right| \\
&\leq \int_{\alpha-iT}^{\alpha+iT} |G(s)||X^{s-1}|\mathrm{d}s \\
&\leq M\int_{-T}^{T} |X^{\alpha+it-1}|\mathrm{d}t \\
&= M\int_{-T}^{T} X^{\alpha-1}\mathrm{d}t \\
&= 2MTX^{\alpha-1}.
\end{aligned} \tag{2.45}$$

Furthermore, since $U > t_0$,

$$\left|\int_{L_1} G(s)X^{s-1}\mathrm{d}s\right| = \left|\int_{L_7} G(s)X^{s-1}\mathrm{d}s\right| \tag{2.46}$$

$$\begin{aligned}
&\leq \int_{1+iU}^{c+iU} |G(s)||X^{s-1}|\mathrm{d}s \\
&\leq RU^{-3/2}\int_{1}^{c} |X^{\sigma-iU-1}|\mathrm{d}\sigma \\
&= RU^{-3/2}\int_{1}^{c} X^{\sigma-1}\mathrm{d}\sigma \\
&\leq RU^{-3/2}\frac{X^{c-1}}{\log X}\ .
\end{aligned} \tag{2.47}$$

Combining the results (2.42) through to (2.47), it follows that

$$\left|\frac{1}{2\pi i}\int_{c-iU}^{c+iU} G(s)X^{s-1}\mathrm{d}s - \frac{1}{2}\right| \leq \frac{1}{2\pi}\left(2\epsilon + \frac{2M}{\log X} + \frac{2TM}{X^{1-\alpha}} + 2RU^{-3/2}\frac{X^{c-1}}{\log X}\right).$$

Note that

$$\lim_{U \longrightarrow \infty} \left(\frac{1}{2\pi i} \int_{c-iU}^{c+iU} G(s) X^{s-1} \mathrm{d}s - \frac{1}{2} \right) = \frac{\psi_1(X)}{X^2} - \frac{1}{2} \; .$$

Using the fact that if $\lim_{U \longrightarrow \infty} f(U) = L$, then $\lim_{U \longrightarrow \infty} |f(U)| = |L|$, it follows that

$$\left| \frac{\psi_1(X)}{X^2} - \frac{1}{2} \right| = \lim_{U \longrightarrow \infty} \left| \frac{1}{2\pi i} \int_{c-iU}^{c+iU} G(s) X^{s-1} \mathrm{d}s - \frac{1}{2} \right| \leq \frac{\epsilon}{\pi} + \frac{M}{\pi \log X} + \frac{TM}{\pi X^{1-\alpha}} \; . \tag{2.48}$$

It is clear then that for sufficiently large X, we have

$$\left| \frac{\psi_1(X)}{X^2} - \frac{1}{2} \right| \leq \epsilon.$$

Finally, since $\epsilon > 0$ is arbitrary, and the left hand side is independent of ϵ, it follows that

$$\lim_{X \longrightarrow \infty} \frac{\psi_1(X)}{X^2} = 1/2,$$

thus completing the proof of Theorem 2.1.3.

Q.E.D.

As mentioned at the start of the section, this also proves the prime number theorem.

Remark Regarding the setup: while it gives a suitable closed curve within a region for which $G(s)X^{s-1}$ is analytic except at $s = 1$, there are other such setups that would do the same. Perhaps the most basic one is to discard A and "push" L_4 far enough to the right so that γ lies in the union of B and the region $\sigma > 1$, with $s = 1$ still bounded by γ. However, the desired result would not follow, for taking limits as U tends to infinity means that B may also change (become narrower), thus L_4 would have to be pushed further to the right. But then α tends to 1, and it is clear that we cannot have (2.48) for fixed $\alpha \in (1/2, 1)$. Note however, that such a setup is viable if the Riemann hypothesis, which states that all the zeros of $\zeta(s)$ lie on the line $\sigma = 1/2$, is true.

Chapter 3

An "elementary" proof of the prime number theorem

3.1 Introduction to the elementary proof

The title of this chapter suggests an "elementary" proof of the prime number theorem exists. The word is used here in the sense that the proof to be outlined does not use any complex analysis, but instead uses number-theoretic ideas and some real analysis. Note that real analysis must play some part due to fact that the prime number theorem is a statement about limits. In the usual sense, the proof is far from elementary, quite the opposite in fact, and unlike the analytic proof of Chapter 2, the motivations for doing things are far less clear. The first elementary proofs of the prime number theorem were published independently by Paul Erdős and Atle Selberg in 1949, although in actual fact, they were in collaboration until a dispute arose. For an interesting account of the history of the elementary proof, including this dispute, see [8]. Erdős' and Selberg's proofs are outlined in [6] and [15] respectively. The proof outlined here follows a method set out by Norman Levinson in 1969 [12], as outlined with minor simplifications by G.J.O. Jameson in [11].

There is some common ground between the elementary and the previous analytic proof, namely a portion of the analytic proof that did not require complex analysis. To specify, it was proved by "elementary" means that it is sufficient to show that $\psi(x)/x \longrightarrow 1$ as $x \longrightarrow \infty$, where Chebyshev's ψ function was defined as follows:

Definition 3.1.1 *Define* $\psi : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ *by*

$$\psi(x) = \sum_{n \leq x} \Lambda(n) = \sum_{\substack{p,n \\ p^n \leq x}} \log p$$

where p *denotes a prime and* $n \in \mathbb{N}$.

In this chapter, we prove this sufficient condition without resorting to complex analysis. Before a proof is given, some investigation of ψ is undertaken. Begin by recalling from Chapter 1 Chebyshev's θ function, defined as follows:

Definition 3.1.2 *Define the function $\theta : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ by*

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

where the sum is over all primes p less than or equal to x.

The ψ function was introduced by Chebyshev as a function "close" to θ, "close" meaning that the two functions differ only in the relatively rare "prime power" terms, so anything said about θ is likely to have implications for ψ. In particular, an upper bound for θ should help to obtain an upper bound for ψ. In the next section of this chapter, upper estimates of both θ and ψ are obtained.

3.2 Chebyshev's upper estimates for $\theta(x)$ and $\psi(x)$.

Before obtaining an upper estimate of $\theta(x)$, two results from elementary number theory are stated.

Lemma 3.2.1 *If p is prime and $p|ab$, then $p|a$ or $p|b$.*

Lemma 3.2.2 *If $p_1|a$ and $p_2|a$ and $gcd(p_1, p_2) = 1$, then $p_1p_2|a$.*

Theorem 3.2.3 *For all $x \geq 1$, we have $\theta(x) \leq (\log 4)x$.*

Proof. If the result holds for the integers n, then for $x \geq 1, n \leq x < n + 1$ for some n. We have $\theta(x) = \theta(n) \leq (\log 4)n \leq (\log 4)x$. Therefore, it is sufficient to prove the result for integers n. It is clear that the result is true for $n = 1, n = 2$. Suppose for induction that $\theta(m) \leq m \log 4$ for all $m \leq 2n$, for some $n \geq 1$. In particular, we assume that $\theta(n + 1) \leq (n + 1) \log 4$. Let $p_{k+1}, \ldots, p_r$ denote the primes p such that $n + 2 \leq p \leq 2n + 1$. Clearly,

$$\theta(2n + 1) = \theta(n + 1) + \sum_{j=k+1}^{r} \log p_j \leq (n + 1) \log 4 + \log \left(\prod_{j=k+1}^{r} p_j \right).$$

The result can thus be proved for $m = 2n+1$ if it can be shown that $\log \left(\prod_{j=k+1}^{r} p_j \right) \leq n \log 4$, or equivalently, if it can be shown that $\prod_{j=k+1}^{r} p_j \leq 4^n = 2^{2n}$. Motivated by the fact that the primes $p_{k+1}, \ldots, p_r$ are between $n + 2$ and $2n + 1$, observe that

$$\binom{2n+1}{n} + \binom{2n+1}{n+1} \leq (1 + 1)^{2n+1}.$$

Let

$$N = \binom{2n+1}{n} = \frac{(2n+1)(2n)\dots(n+2)}{n!}.$$

Since $\binom{2n+1}{n} = \binom{2n+1}{n+1}$, it follows that

$$N \leq 2^{2n}.$$

Hence, it is sufficient to show that

$$\prod_{j=k+1}^{r} p_j \leq N.$$

By Lemma 3.2.1, the primes $p_{k+1}, \dots p_r$ do not divide $n!$. However they do divide $(2n+1)(2n)\dots(n+2) = n!N$. By Lemma 3.2.1 again, they therefore divide N. As they are all distinct primes, their product $\prod_{j=k+1}^{r} p_j$ divides N, and so is not greater than N, which establishes the result for $m = 2n+1$. To complete the proof, note that $\theta(2n+2) = \theta(2n+1)$, as $2n+2$ is not prime. Hence $\theta(2n+2) \leq (2n+1)\log 4 \leq (2n+2)\log 4$, thus completing the inductive step.

Q.E.D.

As noted at the beginning of the chapter, the definitions of ψ and θ are quite similar. In fact, ψ can be nicely expressed in terms of θ, as the following proposition demonstrates.

Proposition 3.2.4 *Choose $x \geq 2$. Let m be the largest integer such that $2^m \leq x$. Then*

$$\psi(x) = \sum_{n \leq m} \theta(x^{1/n}).$$

Proof.

$$\psi(x) = \sum_{n\in\mathbb{N}} \sum_{\substack{p\in P \\ p^n \leq x}} \log p = \sum_{n\in\mathbb{N}} \sum_{p \leq x^{1/n}} \log p = \sum_{n\in\mathbb{N}} \theta(x^{1/n}).$$

Choose $r \geq 1$. Then $2^{m+r} > x$, or in other words, $2 > x^{1/(m+r)}$. It follows that $\theta(x^{1/(m+r)}) = \theta(1) = 0$. The result clearly follows.

Q.E.D.

Of course, given the upper estimate for $\theta(x)$, this last identity immediately provides an upper estimate for $\psi(x)$. However, it is an awkward one, and therefore, it is more practical to see just how "close" $\theta(x)$ and $\psi(x)$ are, that is, obtain a bound for $\psi(x) - \theta(x)$. This will then be used in conjunction with Chebyshev's upper estimate for $\theta(x)$ to provide an upper estimate for $\psi(x)$.

Proposition 3.2.5 *For all $x > 1$, we have $\psi(x) - \theta(x) \leq 6x^{1/2}$. Also, given $\epsilon > 0$, we have $\psi(x) - \theta(x) \leq (\log 4 + \epsilon)x^{1/2}$ for sufficiently large x.*

Proof. Whenever $1 < x < 2$, we have $\psi(x) = \theta(x)$, thus both inequalities hold. Whenever $x \geq 2$, by Proposition 3.2.4,

$$\psi(x) - \theta(x) = \sum_{2 \leq n \leq m} \theta(x^{1/n}),$$

where m is the largest integer such that $2^m \leq x$. It follows that

$$\psi(x) - \theta(x) \leq \theta(x^{1/2}) + m\theta(x^{1/3}),$$

where $m \leq (\log x / \log 2)$. Therefore, by Chebyshev's upper estimate for θ,

$$\psi(x) - \theta(x) \leq x^{1/2} \log 4 + \frac{\log x}{\log 2} x^{1/3} \log 4 = x^{1/2} \log 4 + 2x^{1/3} \log x. \tag{3.1}$$

Clearly, the second term is not greater than $\epsilon x^{1/2}$ once x is sufficiently large, thus for sufficiently large x,

$$\psi(x) - \theta(x) \leq x^{1/2} \log 4 + \epsilon x^{1/2} = x^{1/2}(\log 4 + \epsilon).$$

This proves the second assertion. Also, one finds by differentiation that for positive α, the greatest value of $\log x / x^\alpha$ is $1/(\alpha e)$. Hence

$$x^{1/3} \log x = x^{1/2} \frac{\log x}{x^{1/6}} \leq 6e^{-1} x^{1/2}.$$

It follows from (3.1) that

$$\psi(x) - \theta(x) \leq x^{1/2} \log 4 + 12e^{-1} x^{1/2} = (\log 4 + 12e^{-1}) x^{1/2} < 6x^{1/2}.$$

Q.E.D.

Given the proposition, Chebyshev's upper estimate for $\psi(x)$ follows with relatively little trouble.

Theorem 3.2.6 (Chebyshev's upper estimate for ψ(x)) *For all $x > 1$, we have $\psi(x) < 2x$. Also, given $\epsilon > 0, \psi(x) \leq (\log 4 + \epsilon)x$ for sufficiently large x.*

Proof. By Proposition 3.2.5, for all $x > 1$,

$$\frac{\psi(x)}{x} \leq \frac{\theta(x)}{x} + \frac{6}{x^{1/2}} \leq \log 4 + \frac{6}{x^{1/2}}.$$

Clearly, $\psi(x)/x \leq \log 4 + \epsilon$ once x is sufficiently large. This proves the second assertion. Note that $\log 4 < 1.4$ and for $x \geq 100$, we have $\log 4 + 6/x^{1/2} < 2$. Thus for $x \geq 100$, it follows that $\psi(x) < 2x$. For numbers below 100, first write

$$\frac{\psi(x)}{x} = \frac{\theta(x)}{x} + \frac{1}{x}(\psi(x) - \theta(x)) < 1.4 + \frac{1}{x}(\psi(x) - \theta(x)).$$

Focusing on the last term, from Proposition 3.2.4, $\psi(x) - \theta(x)$ is increasing, and applying Proposition 3.2.4,

$$\begin{aligned}\psi(100) - \theta(100) &= \theta(10) + \theta(4) + \theta(3) + \theta(2) + \theta(2) \\ &= \log 7 + \log 5 + 3\log 3 + 5\log 2 \\ &< 10.4.\end{aligned}$$

Therefore, $\psi(x) - \theta(x) < 10.4$ for all $x \leq 100$ and it follows that $\frac{1}{x}(\psi(x) - \theta(x)) < 0.6$ for $17 \leq x \leq 100$. Hence, $\psi(x)/x \leq 2$ for all $x \geq 17$. Finally, one checks easily that $\psi(n) < 2n$ for each integer $n \leq 16$. Then if $n \leq x < n+1$, we have $\psi(x) = \psi(n) < 2n \leq 2x$.

Q.E.D.

3.3 A sufficient condition

Chebyshev's upper estimate for $\psi(x)$ is clearly very far from the stated objective of proving that

$$\psi(x)/x \longrightarrow 1 \text{ as } x \longrightarrow \infty. \tag{3.2}$$

A bound like this implies nothing about limits, and so it is necessary to either find a sufficient condition for (3.2), or obtain more knowledge about the behaviour of ψ. Considering the first, roughly speaking, it is required that $\psi(x)/x$ be close to 1 for sufficiently large x. If this were true, then

$$\frac{1}{x}\int_0^x \frac{\psi(t)}{t}\mathrm{d}t \approx \frac{1}{x}\int_0^x \mathrm{d}t = 1$$

for sufficiently large x. Again speaking roughly, the converse of this gives a sufficient condition for (3.2). The following argument formally gives a new sufficient condition, although at this point, its usefulness is not clear.

Proposition 3.3.1 *Suppose that $A(x)$ is non-negative and increasing for $x \geq 1$, and that*

$$\frac{1}{x}\int_1^x \frac{A(t)}{t}\,\mathrm{d}t \longrightarrow 1 \quad \textit{as } x \longrightarrow \infty.$$

Then

$$\frac{A(x)}{x} \longrightarrow 1 \quad \textit{as } x \longrightarrow \infty.$$

Proof. The idea is to establish upper and lower bounds on $A(x)/x$ for sufficiently large x. Write

$$F(x) = \int_1^x \frac{A(t)}{t}\,\mathrm{d}t.$$

Thus

$$\frac{F(x)}{x} \longrightarrow 1 \quad \text{as } x \longrightarrow \infty.$$

Hence for $\lambda > 0$, we have

$$\frac{F(\lambda x)}{x} \longrightarrow \lambda \quad \text{as} \;\; x \longrightarrow \infty. \tag{3.3}$$

Choose $\epsilon > 0$ and $x \geq 1$. Suppose we have t such that $x \leq t \leq (1+\epsilon)x$. Then

$$\frac{A(t)}{t} \geq \frac{A(x)}{(1+\epsilon)x}$$

since A is increasing. Hence

$$F[(1+\epsilon)x] - F(x) = \int_x^{(1+\epsilon)x} \frac{A(t)}{t}\, \mathrm{d}t \geq \frac{\epsilon}{1+\epsilon} A(x). \tag{3.4}$$

Also, by (3.3),

$$\frac{1}{x}\{F[(1+\epsilon)x] - F(x)\} \longrightarrow \epsilon \quad \text{as} \;\; x \longrightarrow \infty.$$

Therefore, for all sufficiently large x,

$$F[(1+\epsilon)x] - F(x) \leq \epsilon(1+\epsilon)x. \tag{3.5}$$

By (3.4) and (3.5), for such x it follows that

$$\frac{A(x)}{x} \leq (1+\epsilon)^2. \tag{3.6}$$

In the same way,

$$\epsilon(1-\epsilon)x \leq F(x) - F[(1-\epsilon)x] \leq \frac{\epsilon}{(1-\epsilon)} A(x)$$

for all sufficiently large x. Hence for such x, we have $(1-\epsilon)^2 \leq A(x)/x$. The proposition immediately follows by using (3.6).

Q.E.D.

Since ψ is non-negative and increasing for $x \geq 1$, the prime number theorem is implied if it can be shown that

$$\frac{1}{x}\int_1^x \frac{\psi(t)}{t}\, \mathrm{d}t \longrightarrow 1 \quad \text{as} \;\; x \longrightarrow \infty. \tag{3.7}$$

In the next two sections, more is said about the ψ function.

3.4 Abel summation; Dirichlet convolution

Recall that $\psi(x) = \sum_{n \leq x} \Lambda(n)$. By a method known as Abel summation — due to Niels Henrik Abel — the partial sums of a function can be related to the function itself and any other function f. Thus, results involving Λ often give information about ψ. The first proposition, which is the basic result in Abel summation, is based on the difference between consecutive partial sums. The details of the proof are left to the reader, and more can be found on Abel summation in [11].

Proposition 3.4.1 *Suppose that a is an arithmetic function, and let $A(x) = \sum_{r\le x} a(r)$. Then for integers $n > m \ge 0$,*

$$\sum_{r=m+1}^{n} a(r)f(r) = \left(\sum_{r=m}^{n-1} A(r)(f(r) - f(r+1))\right) + A(n)f(n) - A(m)f(m).$$

In particular,

$$\sum_{r=1}^{n} a(r)f(r) = \left(\sum_{r=1}^{n-1} A(r)(f(r) - f(r+1))\right) + A(n)f(n).$$

The next proposition introduces an integral, which is useful, bearing in mind (3.7).

Theorem 3.4.2 *Let $y < x$, and let f be a function having a continuous derivative on $[y, x]$. Then*

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

Proof. Let m, n be integers such that $n \le x < n+1$ and $m \le y < m+1$. Then, by Proposition 3.4.1,

$$\begin{aligned}
\sum_{y<r\le x} a(r)f(r) &= \sum_{r=m+1}^{n} a(r)f(r) \\
&= \sum_{r=m}^{n-1} A(r)[f(r) - f(r+1)] + A(n)f(n) - A(m)f(m) \\
&= \sum_{r=m+1}^{n-1} A(r)[f(r) - f(r+1)] + A(n)f(n) - A(m)f(m+1) \\
&= -\sum_{r=m+1}^{n-1} A(r)\int_r^{r+1} f'(t)\mathrm{d}t + A(n)f(n) - A(n)f(m+1).
\end{aligned}$$

Noting that $A(t) = A(r)$ for t such that $r \le t < r+1$ for some $r \in \mathbb{N}$, it follows that

$$A(r)\int_r^{r+1} f'(t)\mathrm{d}t = \int_r^{r+1} A(r)f'(t)\mathrm{d}t = \int_r^{r+1} A(t)f'(t)\mathrm{d}t.$$

Also, $A(x) = A(n)$ and $A(y) = A(m)$, so that

$$\begin{aligned}
\sum_{y<r\le x} a(r)f(r) &= -\sum_{r=m+1}^{n-1}\int_r^{r+1} A(t)f'(t)\mathrm{d}t + A(x)f(n) - A(y)f(m+1) \\
&= -\int_{m+1}^{n} A(t)f'(t)\mathrm{d}t + A(x)f(n) - A(y)f(m+1) \\
&= -\int_y^x A(t)f'(t)\mathrm{d}t + \int_n^x A(t)f'(t)\mathrm{d}t + \int_y^{m+1} A(t)f'(t)\mathrm{d}t \\
&\qquad + A(x)f(n) - A(y)f(m+1).
\end{aligned}$$

Recalling that $x < n+1$, it follows that

$$\int_n^x A(t)f'(t)\mathrm{d}t = \int_n^x A(n)f'(t)\mathrm{d}t = \int_n^x A(x)f'(t)\mathrm{d}t = A(x)(f(x) - f(n)).$$

Likwise, $\int_y^{m+1} A(t)f'(t)\mathrm{d}t = A(y)(f(m+1) - f(y))$. The result quickly follows.

Q.E.D.

A final result on Abel summation is stated. The first part is a special case of Theorem 3.4.2. The details of the proof are again left to the reader.

Proposition 3.4.3 *Let f have a continuous derivative on $[1, x]$. Then*

$$\sum_{r \le x} a(r)f(r) = A(x)f(x) - \int_1^x A(t)f'(t)\mathrm{d}t$$

and

$$\sum_{r \le x} a(r)[f(x) - f(r)] = \int_1^x A(t)f'(t)\mathrm{d}t.$$

Letting $a(n) = \Lambda(n)$ and $f(n) = 1/n$ in the first equation, it follows that

$$\sum_{n \le x} \frac{\Lambda(n)}{n} = \frac{\psi(x)}{x} + \int_1^x \frac{\psi(t)}{t^2}\mathrm{d}t.$$

While it would be ideal to obtain an equation involving $\int_1^x \psi(t)/t \, \mathrm{d}t$, investigating the left-hand side will provide more information about the behaviour of ψ. To obtain an alternative expression for the left-hand side, a technique known as Dirichlet convolution is used, the definition of which follows:

Definition 3.4.1 *If f and g are arithmetic functions, the Dirichlet convolution of f and g, denoted $f * g$, is defined to be*

$$(f * g)(n) = \sum_{jk=n} f(j)g(k) = \sum_{j|n} f(j)g\left(\frac{n}{j}\right),$$

where the sum extends over all positive divisors j of n, or equivalently, over all ordered pairs (j, k) of positive integers whose product is n.

Two standard results on Dirichlet convolution are now proved. For the reader interested in Dirichlet convolution, see [11].

Proposition 3.4.4 *Let a and b be arithmetic functions. Let $A(x) = \sum_{n \le x} a(n)$ and $B(x) = \sum_{n \le x} b(n)$. Then*

$$\sum_{n \le x} (a * b)(n) = \sum_{jk \le x} a(j)b(k) = \sum_{j \le x} a(j)B\left(\frac{x}{j}\right) = \sum_{k \le x} A\left(\frac{x}{k}\right)b(k).$$

Proof. By definition,

$$\sum_{n\le x}(a*b)(n) = \sum_{n\le x}\sum_{jk=n} a(j)b(k).$$

It follows that

$$\begin{aligned}\sum_{n\le x}(a*b)(n) &= \sum_{jk\le x} a(j)b(k) = \sum_{j\le x}\sum_{k\le x/j} a(j)b(k)\\ &= \sum_{j\le x} a(j)\sum_{k\le x/j} b(k) = \sum_{j\le x} a(j)B\left(\frac{x}{j}\right).\end{aligned}$$

The last expression is obtained by interchanging a and b.

Q.E.D.

We have a special case of Proposition 3.4.4 when b is taken to be the following function u:

Definition 3.4.2 *Let u be the arithmetic function defined by $u(n) = 1$ for all $n \in \mathbb{N}$.*

Corollary 3.4.5 *For any arithmetic function a,*

$$\sum_{n\le x}(a*u)(n) = \sum_{j\le x} a(j)\left[\frac{x}{j}\right] = \sum_{k\le x} A\left(\frac{x}{k}\right).$$

Proof. By Proposition 3.4.4,

$$\sum_{n\le x}(a*u)(n) = \sum_{j\le x} a(j)\sum_{k\le x/j} u(k) = \sum_{j\le x} a(j)\left[\frac{x}{j}\right].$$

Also by Proposition 3.4.4,

$$\sum_{n\le x}(a*u)(n) = \sum_{k\le x} A\left(\frac{x}{k}\right)u(k) = \sum_{k\le x} A\left(\frac{x}{k}\right).$$

Q.E.D.

The significance of this corollary is that

$$\sum_{n\le x}\Lambda(n)\left[\frac{x}{n}\right] = \sum_{n\le x}(\Lambda*u)(n).$$

Furthermore, given Theorem 1.1.2, note that

$$(\Lambda*u)(n) = \sum_{j|n}\Lambda(j)u\left(\frac{n}{j}\right) = \sum_{j|n}\Lambda(j) = \log n, \tag{3.8}$$

so that

$$x\sum_{n\le x}\frac{\Lambda(n)}{n}=\sum_{n\le x}\Lambda(n)\left\{\frac{x}{n}\right\}+\sum_{n\le x}\log n. \tag{3.9}$$

By estimating the terms on the right-hand side, a bound can be obtained for

$$\sum_{n\le x}\frac{\Lambda(n)}{n}=\frac{\psi(x)}{x}+\int_1^x\frac{\psi(t)}{t^2}\mathrm{d}t. \tag{3.10}$$

To obtain an estimate for $\sum_{n\le x}\log n$, results relating this sum to the integral $\int_1^x \log t \,\mathrm{d}t$ are used. The next section provides these results, and thus the bound.

3.5 Estimation of sums by integrals; Mertens' estimates

When a function is always increasing or always decreasing, much can be said about how the partial sums of this function and the corresponding integral relate. For $x \ge 1$, write

$$S(x)=\sum_{1\le r\le x} f(r) \quad \text{and} \quad I(x)=\int_1^x f(t)\mathrm{d}t.$$

It is easy to show that when $f(t)$ is decreasing,

$$S(n)-f(1)\le I(n)\le S(n-1), \tag{3.11}$$

or equivalently,

$$I(n+1)\le S(n)\le I(n)+f(1),$$

with opposite inequalities if $f(t)$ is increasing. The next two propositions relate $S(x)$ to $I(x)$ for $x \ge 1$.

Proposition 3.5.1 *Suppose that $f(t)$ is non-negative and decreasing for all $t \ge 1$. Then for all $x \ge 1$,*

$$I(x)\le S(x)\le I(x)+f(1).$$

Proof. Suppose that $n \le x < n+1$, where n is an integer. Then $S(x) = S(n)$ and $I(n) \le I(x) \le I(n+1)$. Recalling that $I(n+1) \le S(n) \le I(n) + f(1)$, the result follows.

Q.E.D.

For example, let $S(x) = \sum_{r\le x}\frac{1}{r}$. Here, $f(r) = 1/r$, non-negative and decreasing. Thus, $f(1) = 1$ and

$$I(x)=\int_1^x\frac{1}{t}\,\mathrm{d}t=\log x,$$

so that $\log x \le \sum_{r\le x}\frac{1}{r} \le \log x + 1$. This result shall be of use at a later stage of this chapter.

Proposition 3.5.2 *Suppose that $f(t)$ is non-negative and increasing for $t \geq 1$. Define $S(x), I(x)$ as above. Then for all $x \geq 1$, we have $S(x) = I(x) + r(x)$, where $|r(x)| \leq f(x)$.*

Proof. As $f(t)$ is increasing, the opposite inequality to (3.11) applies, that is, $S(n-1) \leq I(n) \leq S(n) - f(1)$, or equivalently, $S(n) - f(n) \leq I(n) \leq S(n) - f(1)$. Let $n \leq x < n+1$. Then

$$S(x) = S(n) \leq I(n) + f(n) \leq I(x) + f(x). \tag{3.12}$$

Since $f(t)$ is increasing and $x - n < 1$, we have $I(x) - I(n) = \int_n^x f(t)\mathrm{d}t \leq f(x)$, and hence

$$I(x) \leq I(n) + f(x) \leq S(n) - f(1) + f(x) \leq S(x) + f(x). \tag{3.13}$$

By (3.12) and (3.13), $-f(x) \leq S(x) - I(x) \leq f(x)$, and the result follows.

Q.E.D.

As the log function is non-negative and increasing, this last result allows us to relate $\sum_{n \leq x} \log n$ to $\int_1^x \log t\mathrm{d}t$, or to be more precise, shows that

$$\sum_{n \leq x} \log n = \int_1^x \log t\mathrm{d}t + b(x), \tag{3.14}$$

where $|b(x)| \leq \log x$. We are now in a position to obtain a bound on

$$\sum_{n \leq x} \frac{\Lambda(n)}{n} = \frac{\psi(x)}{x} + \int_1^x \frac{\psi(t)}{t^2}\mathrm{d}t;$$

the rest of the results in this section are estimates due to Franz Mertens. For more on Mertens' work, see [11].

Theorem 3.5.3 *For all $x > 1$,*

$$\sum_{n \leq x} \frac{\Lambda(n)}{n} = \log x + r(x),$$

where $|r(x)| \leq 2$.

Proof. Recall (3.9), namely that

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

Given (3.14), this can be rewritten as

$$x \sum_{n \leq x} \frac{\Lambda(n)}{n} = a(x) + \int_1^x \log t\mathrm{d}t + b(x),$$

where $|b(x)| \leq \log x$ and $a(x) = \sum_{n \leq x} \Lambda(n) \left\{ \frac{x}{n} \right\}$. It follows after integration that

$$x \sum_{n \leq x} \frac{\Lambda(n)}{n} = a(x) + x \log x - x + 1 + b(x),$$

To prove the statement, it is clearly enough to show that $a(x) - x + 1 + b(x)$ is bounded by $2x$. Indeed, by Chebyshev's upper estimate for $\psi(x)$, it follows that

$$0 \leq a(x) \leq \sum_{n \leq x} \Lambda(n) = \psi(x) < 2x,$$

thus $|a(x) - x| < x$. Also, $|b(x) + 1| \leq |b(x)| + 1 \leq \log x + 1 \leq x$, for $x > 1$. The statement follows.

Q.E.D.

The following corollary tells us more about the behaviour of the function, as was sought at the end of section 3.3.

Corollary 3.5.4 *The statements*

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

and

$$\int_1^x \frac{\psi(t)}{t^2} \mathrm{d}t = \log x + O(1)$$

are equivalent.

Proof. The result is immediate from (3.10) and the fact that $\psi(x)/x$ is bounded by 2.

Q.E.D.

Therefore, since the first statement of Corollary 3.5.4 is true, so too is the second. Recall that the prime number theorem roughly states that $\psi(x)$ is "close" to x for large enough x. If this is true, then for sufficiently large x we have that

$$\int_1^x \frac{\psi(t)}{t^2} \mathrm{d}t \approx \int_1^x \frac{1}{t} \mathrm{d}t = \log x.$$

The second condition of Corollary 3.5.4 can clearly be restated as

$$\int_1^x \frac{\psi(t) - t}{t^2} \, \mathrm{d}t = O(1).$$

To complete the section, the function R is defined as

$$R(x) = \begin{cases} 0 & \text{if } x < 1 \\ \psi(x) - x & \text{if } x \geq 1, \end{cases}$$

and as $0 \leq \psi(x) \leq 2x, R$ has the following properties:

Proposition 3.5.5 *Let R be defined as above. Then,*

1) *for all $x \geq 0, |R(x)| \leq x$;*

   and

2) *$\int_1^x \frac{R(t)}{t^2}\,\mathrm{d}t$ is bounded for all $x \geq 1$.*

3.6 A further sufficient condition

The aim of this section is to establish further sufficient conditions for the proof of the prime number theorem. It should first be noted that a sufficient condition for the prime number theorem is $R(x)/x \longrightarrow 0$ as $x \longrightarrow \infty$, as this is equivalent to stating that $\psi(x)/x \longrightarrow 1$ as $x \longrightarrow \infty$. While the properties of R are heavily used later in crucial results known as Selberg's formulas, at this stage, what is known of R seems unlikely to be of use. On the other hand, recall (3.7): it is sufficient to show that

$$\frac{1}{x}\int_1^x \frac{\psi(t)}{t}\,\mathrm{d}t \longrightarrow 1 \quad \text{as } x \longrightarrow \infty.$$

It is easy to show that this is equivalent to the statement

$$\frac{1}{x}\int_1^x \frac{R(t)}{t}\,\mathrm{d}t \longrightarrow 0 \quad \text{as } x \longrightarrow \infty.$$

Again, the properties of R do not appear helpful. However, letting

$$S(x) = \int_1^x \frac{R(t)}{t}\,\mathrm{d}t \quad \text{for all } x \geq 1,$$

the statement becomes $S(x)/x \longrightarrow 0$ as $x \longrightarrow \infty$. The properties of S will be investigated in Proposition 3.6.3 after the proofs of two auxiliary results on reversing the order of integration.

Proposition 3.6.1 *Let $f, g : [x, y] \longrightarrow \mathbb{R}$ be differentiable functions. Letting $F = f'$ and $G = g'$, it follows that*

$$\int_y^x G(t)\int_y^t F(u)\mathrm{d}u\,\mathrm{d}t = \int_y^x F(u)\int_u^x G(t)\mathrm{d}t\,\mathrm{d}u.$$

Proof.

$$\begin{aligned}
\int_y^x G(t)\int_y^t F(u)\mathrm{d}u\ \mathrm{d}t &- \int_y^x F(u)\int_u^x G(t)\mathrm{d}t\ \mathrm{d}u \\
&= \int_y^x g'(t)\int_y^t f'(u)\mathrm{d}u\ \mathrm{d}t - \int_y^x f'(u)\int_u^x g'(t)\mathrm{d}t\ \mathrm{d}u \\
&= \int_y^x g'(t)(f(t)-f(y))\mathrm{d}t - \int_y^x f'(u)(g(x)-g(u))\mathrm{d}u \\
&= -f(y)\int_y^x g'(t)\mathrm{d}t + \int_y^x [g'(t)f(t)+f'(t)g(t)]\mathrm{d}t - g(x)\int_y^x f'(t)\mathrm{d}t \\
&= -f(y)(g(x)-g(y)) + f(x)g(x) - f(y)g(y) - g(x)(f(x)-f(y)) \\
&= 0.
\end{aligned}$$

Q.E.D.

Note in particular that Proposition 3.6.1 is valid whenever F and G are continuous, as antiderivatives of continuous functions are guaranteed to exist. Now $R(u)/u = \psi(u)/u - 1$ is clearly discontinuous; however, it is also clear that $R(u)/u$ is continuous between consecutive integers, and a modified version of Proposition 3.6.1, namely Proposition 3.6.2, is therefore used in the proof of Proposition 3.6.3.

Proposition 3.6.2 *Let F be a function defined on $\mathbb{R}$ such that F is continuous on $(a, a+1)$ for all $a \in \mathbb{Z}$. Let $g : [x, y] \longrightarrow \mathbb{R}$ be a differentiable function, and let $G = g'$. Then*

$$\int_y^x G(t)\int_y^t F(u)\mathrm{d}u\ \mathrm{d}t = \int_y^x F(u)\int_u^x G(t)\mathrm{d}t\ \mathrm{d}u.$$

Proof.

$$\begin{aligned}
\int_y^x G(t)\int_y^t F(u)\mathrm{d}u\ \mathrm{d}t &= \sum_{n=[y]+1}^{[x]-1}\int_n^{n+1} G(t)\int_y^t F(u)\mathrm{d}u\ \mathrm{d}t \\
&+ \int_y^{[y]+1} G(t)\int_y^t F(u)\mathrm{d}u\ \mathrm{d}t + \int_{[x]}^x G(t)\int_y^t F(u)\mathrm{d}u\ \mathrm{d}t. \qquad (3.15)
\end{aligned}$$

For $n \in \mathbb{N}$,

$$\int_n^{n+1} G(t)\int_y^t F(u)\mathrm{d}u\ \mathrm{d}t = \int_n^{n+1} G(t)\int_y^n F(u)\mathrm{d}u\ \mathrm{d}t + \int_n^{n+1} G(t)\int_n^t F(u)\mathrm{d}u\ \mathrm{d}t.$$

By Proposition 3.6.1,

$$\int_n^{n+1} G(t)\int_n^t F(u)\mathrm{d}u\ \mathrm{d}t = \int_n^{n+1} F(u)\int_u^{n+1} G(t)\mathrm{d}t\ \mathrm{d}u,$$

since $F(u)$ is continuous between n and the variable t, t being less than or equal to $n+1$. Also,

$$\int_n^{n+1} G(t) \left(\int_y^n F(u)\mathrm{d}u \right) \mathrm{d}t = \int_y^n F(u)\mathrm{d}u \left(\int_n^{n+1} G(t)\mathrm{d}t \right) = \int_y^n F(u) \int_n^{n+1} G(t)\mathrm{d}t.$$

It follows that

$$\int_n^{n+1} G(t) \int_y^t F(u)\mathrm{d}u \;\mathrm{d}t = \int_y^n F(u) \int_n^{n+1} G(t)\mathrm{d}t \;\mathrm{d}u + \int_n^{n+1} F(u) \int_u^{n+1} G(t)\mathrm{d}t \;\mathrm{d}u.$$

Similarly,

$$\int_{[x]}^x G(t) \int_y^t F(u)\mathrm{d}u \;\mathrm{d}t = \int_y^{[x]} F(u) \int_{[x]}^x G(t)\mathrm{d}t \;\mathrm{d}u + \int_{[x]}^x F(u) \int_u^x G(t)\mathrm{d}t \;\mathrm{d}u$$

and

$$\int_y^{[y]+1} G(t) \int_y^t F(u)\mathrm{d}u \;\mathrm{d}t = \int_y^{[y]+1} F(u) \int_u^{[y]+1} G(t)\mathrm{d}t \;\mathrm{d}u.$$

Letting $A_n = \int_n^{n+1} G(t) \int_y^t F(u)\mathrm{d}u \;\mathrm{d}t$ it follows by (3.15) that

$$\begin{aligned} \int_y^x G(t) \int_y^t F(u)\mathrm{d}u \;\mathrm{d}t &= \sum_{n=[y]+1}^{[x]-1} A_n + \int_y^{[y]+1} F(u) \int_u^{[y]+1} G(t)\mathrm{d}t \;\mathrm{d}u \\ &+ \int_y^{[x]} F(u) \int_{[x]}^x G(t)\mathrm{d}t \;\mathrm{d}u + \int_{[x]}^x F(u) \int_u^x G(t)\mathrm{d}t \;\mathrm{d}u. \quad (3.16) \end{aligned}$$

Inductively, one can easily show that

$$\sum_{n=1}^{[x]-1} A_n = \int_y^{[x]} F(u) \int_u^{[x]} G(t)\mathrm{d}t \;\mathrm{d}u.$$

Hence,

$$\begin{aligned} \sum_{n=[y]+1}^{[x]-1} A_n &= \sum_{n=1}^{[x]-1} A_n - \sum_{n=1}^{[y]} A_n \\ &= \int_y^{[x]} F(u) \int_u^{[x]} G(t)\mathrm{d}t \;\mathrm{d}u - \int_y^{[y]+1} F(u) \int_u^{[y]+1} G(t)\mathrm{d}t \;\mathrm{d}u, \end{aligned}$$

and by (3.16), it follows that

$$\begin{aligned}
\int_y^x G(t)\int_y^t F(u)\mathrm{d}u\,\mathrm{d}t &= \int_y^{[x]} F(u)\int_u^{[x]} G(t)\mathrm{d}t\,\mathrm{d}u \\
&+ \int_y^{[x]} F(u)\int_{[x]}^x G(t)\mathrm{d}t\,\mathrm{d}u + \int_{[x]}^x F(u)\int_u^x G(t)\mathrm{d}t\,\mathrm{d}u \\
&= \int_y^{[x]} F(u)\int_u^x G(t)\mathrm{d}t\,\mathrm{d}u + \int_{[x]}^x F(u)\int_u^x G(t)\mathrm{d}t\,\mathrm{d}u \\
&= \int_y^x F(u)\int_u^x G(t)\mathrm{d}t\,\mathrm{d}u.
\end{aligned}$$

Q.E.D.

Proposition 3.6.3 *Let S be defined as above. Then,*

1) *for $x_2 > x_1 > 0, |S(x_2)-S(x_1)| \leq x_2-x_1$ (i.e. S is Lipschitz); in particular, $|S(x)| \leq x$ and S is continuous;*

   also,

2) $\int_1^x \frac{S(t)}{t^2}\,\mathrm{d}t$ *is bounded for all $x \geq 1$.*

Proof of 1**.** By Proposition 3.5.5 (1), $|R(t)/t| \leq 1$. Thus whenever $x_2 > x_1 > 0$,

$$\begin{aligned}
|S(x_2) - S(x_1)| &= \left|\int_{x_1}^{x_2} \frac{R(t)}{t}\,\mathrm{d}t\right| \leq \int_{x_1}^{x_2} \left|\frac{R(t)}{t}\right| \mathrm{d}t \\
&\leq \int_{x_1}^{x_2} \mathrm{d}t = x_2 - x_1.
\end{aligned}$$

It follows that $|S(x)| = |S(x) - S(0)| \leq x$, with the continuity of S being obvious. Note that R is not continuous.

Proof of 2**.** By the definition of S,

$$\int_1^x \frac{S(t)}{t^2}\,\mathrm{d}t = \int_1^x \frac{1}{t^2}\int_1^t \frac{R(u)}{u}\,\mathrm{d}u\,\mathrm{d}t.$$

In light of Proposition 3.6.2, one can re-express this as

$$\begin{aligned}
\int_1^x \frac{S(t)}{t^2}\,\mathrm{d}t &= \int_1^x \frac{R(u)}{u}\int_u^x \frac{1}{t^2}\,\mathrm{d}t\,\mathrm{d}u \\
&= \int_1^x \frac{R(u)}{u}\left(\frac{1}{u} - \frac{1}{x}\right)\mathrm{d}u \\
&= \int_1^x \frac{R(u)}{u^2}\,\mathrm{d}u - \frac{1}{x}\int_1^x \frac{R(u)}{u}\,\mathrm{d}u \\
&= \int_1^x \frac{R(u)}{u^2}\,\mathrm{d}u - \frac{S(x)}{x}.
\end{aligned}$$

This is bounded, given Proposition 3.5.5 and the fact that $|S(x)| \leq x$.

Q.E.D.

Again, the properties of S do not seem helpful in proving that $S(x)/x \longrightarrow 0$ as $x \longrightarrow \infty$. Therefore, for motivations which are far from obvious, define

$$W(x) = \frac{S(e^x)}{e^x} \quad \text{for all } \ x \geq 0.$$

If it could be shown that $W(x) \longrightarrow 0$ as $x \longrightarrow \infty$, then $S(x)/x \longrightarrow 0$ as $x \longrightarrow \infty$, and our elementary proof would be complete. Some properties of W will now be deduced from the properties of S. Begin by recalling that $|S(x)| \leq x$ for all $x \geq 1$. Hence, $|W(x)| \leq 1$ for all $x \geq 0$. Further properties are proved in the next proposition, and are the main motivation for the definition of W.

Proposition 3.6.4 *Let W be defined as above. Then,*

1) for $x_2 > x_1 > 0, |W(x_2) - W(x_1)| \leq 2(x_2 - x_1)$;

and

2) $\int_0^x W(t)\,\mathrm{d}t$ is bounded for all $x \geq 0$.

Proof of 1.

$$\begin{aligned} |W(x_2) - W(x_1)| &= \left| \frac{S(e^{x_2})}{e^{x_2}} - \frac{S(e^{x_1})}{e^{x_1}} \right| \\ &= \frac{1}{e^{x_1}e^{x_2}} \left| e^{x_1}S(e^{x_2}) - e^{x_2}S(e^{x_1}) \right| \\ &= \frac{1}{e^{x_1}e^{x_2}} \left| e^{x_1}(S(e^{x_2}) - S(e^{x_1})) + S(e^{x_1})(e^{x_1} - e^{x_2}) \right| \\ &\leq \frac{1}{e^{x_1}e^{x_2}} \left(e^{x_1} \left| S(e^{x_2}) - S(e^{x_1}) \right| + |S(e^{x_1})||e^{x_1} - e^{x_2}| \right). \end{aligned}$$

By Proposition 3.6.3 (1),

$$\begin{aligned} |W(x_2) - W(x_1)| &\leq \frac{1}{e^{x_1}e^{x_2}}(e^{x_1}(e^{x_2} - e^{x_1}) + e^{x_1}(e^{x_2} - e^{x_1})) \\ &= 2(1 - e^{x_1 - x_2}). \end{aligned}$$

Since $e^x \geq 1 + x$, it follows that $x_2 - x_1 \geq 1 - e^{x_1 - x_2}$, and the result follows.

Proof of 2. We have

$$\int_0^x W(t)\,\mathrm{d}t = \int_0^x \frac{S(e^t)}{e^t}\,\mathrm{d}t.$$

Let $t = \log u$, then $u = e^t$ and $\mathrm{d}t/\mathrm{d}u = 1/u$. Also, when $t = 0, u = 1$ and when $t = x, u = e^x$. Therefore,

$$\int_0^x W(t)\ \mathrm{d}t = \int_1^{e^x} \frac{S(u)}{u^2}\ \mathrm{d}u,$$

which is bounded, by Proposition 3.6.3 (2).

Q.E.D.

As with S and R before, these properties of W seem insufficient to prove the prime number theorem. However,

$$\limsup_{x \longrightarrow \infty} |W(x)| = 0 \tag{3.17}$$

implies that $\lim_{x \longrightarrow \infty} W(x) = 0$. Thus (3.17) is another sufficient condition for the prime number theorem. The proposition that follows establishes a further sufficient condition, before we turn our attention to establishing more about the functions R, S and ultimately W. Some remarks are now made for the reader unfamiliar with limsups.

Limsups. Let f be any bounded real-valued function on (a, ∞) for some a. Define $s_f(x) = \sup_{y \geq x} f(y)$. Then $s_f(x)$ is clearly a decreasing function of x. It is also bounded below, so it tends to a limit as $x \longrightarrow \infty$. This limit is called the *limsup* of $f(x)$ as $x \longrightarrow \infty$, and is denoted $\limsup_{x \longrightarrow \infty} f(x)$. In other words,

$$\limsup_{x \longrightarrow \infty} f(x) = \lim_{x \longrightarrow \infty} s_f(x) = \lim_{x \longrightarrow \infty} \sup_{y \geq x} f(y).$$

Note that while not all bounded functions tend to a limit, they all have a limsup. The above statement concerning limsups and any such further statements about limsups can be proved using the limsup definition together with limit definitions.

Proposition 3.6.5 *Let $W(x)$ be defined as above, and let*

$$\alpha = \limsup_{x \longrightarrow \infty} |W(x)|$$

and

$$\beta = \limsup_{x \longrightarrow \infty} \frac{1}{x} \int_0^x |W(t)|\ \mathrm{d}t.$$

Then either $\alpha = 0$ or $\beta < \alpha$.

Proof. By Proposition 3.6.4 (2), there exists $M \in \mathbb{R}$ such that

$$\left| \int_0^x W(t) \mathrm{d}t \right| \leq \frac{M}{2}$$

for all $x > 0$. For any $x_1, x_2 > 0$, it easily follows that

$$\left| \int_{x_1}^{x_2} W(t) \mathrm{d}t \right| \leq M. \tag{3.18}$$

Clearly, $\alpha \geq 0$. Assume that $\alpha > 0$. The proof is complete if it can be shown that $\beta < \alpha$. Begin by choosing a such that $\alpha < a \leq 2\alpha$. Then since

$$\lim_{x \longrightarrow \infty} \sup_{y \geq x} |W(y)| = \alpha,$$

it follows that $\sup_{y \geq x} |W(y)| \leq a$ for large enough x. Suppose this is true for x_0. Then $|W(x)| \leq a$ for all $x > x_0$. Choosing $x_1 > x_0$, consider the integral of $|W(x)|$ on the interval $I = [x_1, x_1 + h]$, where $h \geq 2\alpha$ is to be chosen. With regard to (3.18), if either $W(x) \geq 0$ or $W(x) \leq 0$ for all $x \in I$, then clearly

$$\int_{x_1}^{x_1+h} |W(x)| \mathrm{d}x \leq M. \tag{3.19}$$

Otherwise, since W is continuous by Proposition 3.6.4 (1), it is zero at some point z of I. Again by Proposition 3.6.4 (2), $|W(x)| = |W(x) - W(z)| < 2|x - z|$ for all x. In particular, if $|x - z| \leq a/2$, then $|W(x)| \leq 2|x - z|$ is better than the estimate $|W(x)| \leq a$ that applies generally on I. Therefore, if we wish to get a better estimate for

$$\int_{x_1}^{x_1+h} |W(x)| \mathrm{d}x$$

than ah, it is of interest as to how much of the interval $[z - a/2, z + a/2]$ lies in I. On observing that $a < h$, it is easy to show that at least one of $z \pm a/2$ (say w.l.o.g. $z + a/2$) is in I, and

$$\int_{z}^{z+a/2} |W(x)| \mathrm{d}x \leq \int_{z}^{z+a/2} 2(x - z) \mathrm{d}x = \frac{a^2}{4} .$$

Combining this with the estimate $|W(x)| \leq a$ on the rest of I, it follows that

$$\int_{x_1}^{x_1+h} |W(x)| \mathrm{d}x \leq \frac{a^2}{4} + \left(h - \frac{a}{2}\right) a = a \left(h - \frac{a}{4}\right) < a \left(h - \frac{\alpha}{4}\right).$$

For this inequality to hold in the case where $W(x)$ does not change sign, recall (3.19) and choose h to be greater than both $M/\alpha + \alpha/4$ and 2α. Note that h does not vary with a. Write $h - \alpha/4 = \rho h$, where clearly, $\rho < 1$. Then,

$$\int_{x_1}^{x_1+h} |W(x)| \mathrm{d}x < \rho a h. \tag{3.20}$$

Now take any $x \geq x_1 + h$, and let $n \in \mathbb{N}$ be such that $x_0 + nh \leq x < x_0 + (n + 1)h$. Write $\int_0^{x_0} |W(x)| \mathrm{d}x = C$. Then

$$\begin{aligned} \int_0^x |W(x)| \mathrm{d}x &\leq \int_0^{x_0} |W(x)| \mathrm{d}x + \int_{x_0}^{x_0+(n+1)h} |W(x)| \mathrm{d}x \\ &= C + \sum_{k=0}^{n} \int_{x_0+kh}^{(x_0+kh)+h} |W(x)| \mathrm{d}x. \end{aligned}$$

For any $k \in \{0, \ldots, n\}, x_0 + kh \geq x_0$. Hence by (3.20),

$$\int_{x_0+kh}^{(x_0+kh)+h} |W(x)| \mathrm{d}x \leq \rho a h.$$

It follows that

$$\int_0^x |W(x)| \mathrm{d}x \leq C + (n+1)\rho a h,$$

and since $x \geq nh$, we have

$$\frac{1}{x} \int_0^x |W(x)| \mathrm{d}x \leq \frac{C}{x} + \left(1 + \frac{1}{n}\right) \rho a.$$

Noting that $n \longrightarrow \infty$ as $x \longrightarrow \infty$, it follows that $C/x + \rho a/n \longrightarrow 0$ as $x \longrightarrow \infty$. Hence

$$\lim_{x \longrightarrow \infty} \frac{1}{x} \int_0^x |W(x)| \mathrm{d}x \leq \rho a,$$

which implies that $\beta \leq \rho a$. This is true for all a such that $\alpha < a \leq 2\alpha$, so in fact $\beta \leq \rho\alpha < \alpha$.

Q.E.D.

Proposition 3.6.5 shows that if $\alpha > 0$, then $\beta < \alpha$. Therefore, if $\alpha \leq \beta$, then $\alpha \leq 0$. Again by Proposition 3.6.5, this would imply that $\alpha = 0$, that is, $\limsup_{x \longrightarrow \infty} |W(x)| = 0$, thus the elementary proof would also be complete. Proving $\alpha \leq \beta$ depends on work of a completely different nature, the main prerequisites being Möbius inversion and Euler's summation formula. These are presented in the next section.

3.7 The Möbius function; Möbius inversion

For the purposes of this elementary proof of the prime number theorem, it would suffice to simply define the Möbius function, and then proceed. However, the definition is not an obviously useful one, and it is worthwhile briefly considering the motivation behind it. We begin by stating the Euler product, a remarkable identity and a generator of many further ideas. The proof is omitted, as the Möbius function arises naturally from the statement. Recall from Chapter 2 that a *completely multiplicative function* $a(n)$ is an arithmetic function such that $a(n)a(m) = a(nm)$ for all $n, m \in \mathbb{N}$. Throughout, P denotes the primes, with $P[N]$ denoting the set of primes less than or equal to a real number N.

Theorem 3.7.1 *Suppose that $a(n)$ is a completely multiplicative function that is not identically zero such that $\sum_{n=1}^{\infty} |a(n)|$ is convergent. Then $\sum_{n=1}^{\infty} a(n) \neq 0$, and*

$$\sum_{n=1}^{\infty} a(n) = \prod_{p \in P} \frac{1}{1 - a(p)} \,.$$

(Note that if $a(p) = 1$ for any $p \in P$, then $a(p^k) = 1$ for all $k \in \mathbb{N}$, and the sum would not be convergent.) Thus an infinite series is equated to a product involving only the primes, encapsulating the fact that the integers are uniquely expressible as products of primes. Also note that the identity

$$\zeta(s) = \prod_{p \in P} (1 - p^{-s})^{-1},$$

as proved in Theorem 2.2.7 for $\sigma > 1$, is a special case of the Euler product.

It is clear that

$$\frac{1}{\sum_{n=1}^{\infty} a(n)} = \prod_{p \in P} (1 - a(p)).$$

In an attempt to write the product as a sum, consider the finite product

$$Q_N = \prod_{p \in P[N]} (1 - a(p)). \tag{3.21}$$

The idea here is to write Q_N as a *sum,* the identification of terms and non-terms leading us to define the *Möbius function.* By a *term,* we mean a term of the infinite product expansion, and not a factor of the same. Clearly, $a(1) = 1$ is a term. Otherwise, x is a term if and only if $x = (-1)^k a(p_1)a(p_2)\ldots a(p_k)$, where $p_1, p_2, \ldots p_k$ are distinct elements of $P[N]$. Since a is completely multiplicative, this is equivalent to stating that x is a term if and only if $x = (-1)^k a(p_1 p_2 \ldots p_k)$, where $p_1, p_2, \ldots, p_k$ are distinct elements of $P[N]$, or stated equivalently again, x is a term if and only if $x = (-1)^k a(n)$, where $n = p_1 p_2 \ldots p_k$ for distinct elements $p_1, p_2, \ldots, p_k$ of $P[N]$. Hence, if E_N is the set of all positive integers less than or equal to the product of all primes in $P[N]$, then

$$Q_N = \sum_{n \in E_N} \mu(n) a(n), \tag{3.22}$$

where

$$\begin{aligned} &\mu(1) = 1, && \\ &\mu(n) = (-1)^k && \text{if } n = p_1 p_2 \ldots p_k, \text{ a product of } k \text{ distinct primes,} \\ &\mu(n) = 0 && \text{otherwise. (that is, if } p^2 | n \text{ for some prime } p.) \end{aligned}$$

This defines the *Möbius function* $\mu(n)$. For completeness, considering (3.21) and (3.22), one might guess that

$$\frac{1}{\sum_{n=1}^{\infty} a(n)} = \prod_{p \in P} (1 - a(p)) = \sum_{n=1}^{\infty} \mu(n) a(n),$$

and indeed, that is the case. The proof — outlined in [11] — is omitted. The main point is that the Möbius function is clearly motivated. We shall show that some equalities can be re-expressed to involve the Möbius function, in a method known as Möbius inversion. The relevance to our argument is that bounds for functions which might otherwise not be obtained are often found in this manner. This will be clearer after Möbius inversion is actually outlined. We start with a definition.

Definition 3.7.1 *Define the arithmetic function* $e_1 : \mathbb{N} \longrightarrow \mathbb{R}$ *by*

$$e_1(n) = \begin{cases} 1 & \text{if } \; n = 1, \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to check that e_1 is the identity with respect to Dirichlet convolution, as discussed in Section 3.4. Recall also from that section the definition of u. The following theorem shows that μ is the inverse of u with respect to convolution.

Theorem 3.7.2 *For all* $n \in \mathbb{N}$, *we have* $(u * \mu)(n) = e_1(n)$. *Equivalently, for* $n > 1$, *we have* $\sum_{i|n} \mu(i) = 0$.

Proof. Clearly, $(u * \mu)(1) = u(1)\mu(1) = 1 = e_1(1)$. The equivalence of the statements now follows on noting that for $n > 1$, $(u * \mu)(n) = \sum_{i|n} \mu(i)$ and $e_1(n) = 0$.

It is now proved that for $n > 1$, we have $\sum_{i|n} \mu(i) = 0$. Let n have prime factorisation

$$n = {p_1}^{r_1} \dots {p_k}^{r_k}.$$

Suppose that $i|n, i > 1$ and $\mu(i) \neq 0$. Then $i = q_1 \dots q_j$, where $q_1, \dots, q_j$ are a choice of j numbers from $p_1, \dots, p_k$. For a fixed $j \leq k$, there are $\binom{k}{j}$ choices of j numbers from $p_1, \dots, p_k$. Clearly then, where the qs are prime,

$$\sum_{\substack{i|n \\ i=q_1 \dots q_j}} \mu(i) = (-1)^j \binom{k}{j}.$$

Thus

$$\begin{aligned} \sum_{i|n} \mu(i) &= \mu(1) + \sum_{\substack{i|n \\ i \geq 2}} \mu(i) = 1 + \sum_{j=1}^{k} (-1)^j \binom{k}{j} \\ &= \sum_{j=0}^{k} (-1)^j \binom{k}{j} = (1-1)^k = 0. \end{aligned}$$

Q.E.D.

Building on this result, a first form of Möbius inversion can now be introduced.

Corollary 3.7.3 (Möbius inversion, first form) *For arithmetic functions* a, b, *the following statements are equivalent:*

1) $b(n) = \sum_{i|n} a(i)$*; that is,* $b = a * u$*,*

and

2) $a(n) = \sum_{i|n} \mu(i) b(n/i)$*; that is,* $a = b * \mu$*.*

Proof. If $b = a * u$, then

$$b * \mu = a * u * \mu = a * e_1 = a.$$

Conversely, if $a = b * \mu$, then $a * u = b * \mu * u = b * e_1 = b$.

Q.E.D.

Suppose then that a bound for b existed. Then if the first statement is true, so too is the second. As μ is bounded by 1, it is clear that a bound for a can also be found. A second form of Möbius inversion is introduced after two propositions.

Proposition 3.7.4 *If* a, b, c *are arithmetic functions and* c *is completely multiplicative, then* $(ac) * (bc) = (a * b)c$*.*

Proof. Let $f(n) = a(n)c(n)$ and let $g(n) = b(n)c(n)$ for all $n \in \mathbb{N}$. Then for any $n \in \mathbb{N}$, it follows that

$$\begin{aligned}[(ac) * (bc)](n) &= (f * g)(n) = \sum_{jk=n} f(j)g(k) = \sum_{jk=n} a(j)c(j)b(k)c(k) \\ &= \sum_{jk=n} a(j)b(k)c(jk) = c(n) \sum_{jk=n} a(j)b(k) = c(n)(a * b)(n).\end{aligned}$$

Q.E.D.

Proposition 3.7.5 *If* a *is completely multiplicative, then the inverse of* a *with respect to convolution is* μa*.*

Proof. Note that $a(1) = 1$, so that $a(1)e_1(1) = e_1(1)$. Also, for $n > 1$, we have $e_1(n) = 0$. Therefore, $ae_1 = e_1$. Since $u * \mu = e_1$, it follows that $a(u * \mu) = e_1$. Applying Proposition 3.7.4, it follows that $au * a\mu = e_1$. Since $au = a$, the result follows.

Q.E.D.

Proposition 3.7.6 (Möbius Inversion, second form) *Let* F *be a function on* $(0, \infty)$ *that takes the value* 0 *on some interval* $(0, a)$*. For* $x > 0$*, let*

$$G(x) = \sum_{n=1}^{\infty} F\left(\frac{x}{n}\right).$$

Then

$$F(x) = \sum_{n=1}^{\infty} \mu(n) G\left(\frac{x}{n}\right).$$

Proof. Note that $F(x/n) = 0$ for $n > x/a$, so the sum defining $G(x)$ is really a finite one, and there are no convergence problems. In the following, sums to infinity are written with this understanding. Since

$$e_1(j) = (\mu * u)(j) = \sum_{n|j} \mu(n)$$

and

$$F(x) = \sum_{j=1}^{\infty} e_1(j) F\left(\frac{x}{j}\right),$$

it follows that

$$F(x) = \sum_{j=1}^{\infty} \sum_{n|j} \mu(n) F\left(\frac{x}{j}\right).$$

Since stating that n divides j is equivalent to stating that j is a multiple of n, it follows that

$$\begin{aligned} F(x) &= \sum_{n=1}^{\infty} \sum_{j=kn} \mu(n) F\left(\frac{x}{j}\right) = \sum_{n=1}^{\infty} \mu(n) \sum_{j=kn} F\left(\frac{x}{j}\right) \\ &= \sum_{n=1}^{\infty} \mu(n) \sum_{k=1}^{\infty} F\left(\frac{x}{kn}\right) = \sum_{n=1}^{\infty} \mu(n) G\left(\frac{x}{n}\right). \end{aligned}$$

Q.E.D.

This concludes our present discussion on Möbius inversion. Its usefulness will become clearer later.

3.8 Selberg's formulas and completion of the proof

So far, it has been shown that if

$$\limsup_{x \longrightarrow \infty} |W(x)| \leq \limsup_{x \longrightarrow \infty} \frac{1}{x} \int_0^x |W(t)| \mathrm{d}t, \tag{3.23}$$

then the prime number theorem is implied. The function W is dependent on S, S being dependent on R. Therefore, the above statement is effectively about R. Thus to prove (3.23), obtaining more information about the behaviour of R seems the natural course. However, the precise direction is not an obvious one, and as before, there is often seemingly little motive for the results proved below. The entire proof can be considered as a particularly outstanding linkage of seemingly disparate ideas, many of which are brilliant in themselves.

Among these are Selberg's formulas, results that when stated will not obviously relate to the prime number theorem. The first of Selberg's formulas states that

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

The proof of this lies partially in an identity very close to Möbius inversion, with a logarithmic factor introduced. Once stated, the link to Selberg's first formula is clear. A lemma used in the proof of the identity — called the Tatuzawa-Iseki identity — is first outlined.

Lemma 3.8.1 *For all $n \ge 1$, $\Lambda(n) = -\sum_{k|n} \mu(k)\log k$.*

Proof. By (3.8), $\log = \Lambda * u$. By the first form of Möbius inversion, this implies that

$$\begin{aligned}\Lambda(n) &= (\log * \mu)(n) = \sum_{n|k} \log\left(\frac{n}{k}\right)\mu(k) \\ &= \log n \sum_{k|n} \mu(k) - \sum_{k|n} \mu(k)\log k.\end{aligned}$$

The result follows on recalling Theorem 3.7.2.

Q.E.D.

Proposition 3.8.2 (the Tatuzawa-Iseki identity) *Let F be a function on $[1,\infty)$ and let $G(x) = \sum_{n\le x} F(x/n)$. Then for $x \ge 1$,*

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

Proof. Expanding the left-hand side,

$$\sum_{k\le x} \mu(k)\log\frac{x}{k}\, G\left(\frac{x}{k}\right) = \log x \sum_{k\le x} \mu(k) G\left(\frac{x}{k}\right) - \sum_{k\le x} \mu(k)\log k\; G\left(\frac{x}{k}\right).$$

Previously, if F_0 was a function on $(0,\infty)$ with $F_0 = 0$ on $(0,1)$, and $G(x) = \sum_{n=1}^{\infty} F_0(x/n)$, then Möbius inversion could be applied. Here F is only defined on $[1,\infty)$. However, if F_0 is taken to be the extension of F to $(0,\infty)$ with $F_0 = 0$ on $(0,1)$, then

$$G(x) = \sum_{n\le x} F_0\left(\frac{x}{n}\right) = \sum_{n=1}^{\infty} F_0\left(\frac{x}{n}\right).$$

Möbius inversion may now be applied to F_0, that is,

$$F_0(x) = \sum_{n=1}^{\infty} \mu(n) G\left(\frac{x}{n}\right)$$

for $x > 0$. It follows that

$$F(x) = \sum_{n=1}^{\infty} \mu(n) G\left(\frac{x}{n}\right)$$

for $x \geq 1$. It is easy to check that $G(x/n) = 0$ whenever $n > x$, hence

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

for $x \geq 1$. It suffices then to show that

$$-\sum_{k \leq x} \mu(k) \log k \; G\left(\frac{x}{k}\right) = \sum_{n \leq x} \Lambda(n) F\left(\frac{x}{n}\right)$$

for $x \geq 1$. By the definition of G,

$$\begin{aligned}
\sum_{k \leq x} \mu(k) \log k \; G\left(\frac{x}{k}\right) &= \sum_{k \leq x} \mu(k) \log k \sum_{j \leq x/k} F\left(\frac{x}{jk}\right) \\
&= \sum_{k \leq x} \sum_{j \leq x/k} F\left(\frac{x}{jk}\right) \mu(k) \log k \\
&= \sum_{jk \leq x} F\left(\frac{x}{jk}\right) \mu(k) \log k \\
&= \sum_{n \leq x} \sum_{jk = n} F\left(\frac{x}{jk}\right) \mu(k) \log k \\
&= \sum_{n \leq x} F\left(\frac{x}{n}\right) \sum_{k | n} \mu(k) \log k.
\end{aligned}$$

The result follows on recalling Lemma 3.8.1.

Q.E.D.

In light of this identity, it is tempting to prove Selberg's first formula by letting $F(x) = R(x)$ and proving that

$$\sum_{k \leq x} \mu(k) \log \frac{x}{k} \; G\left(\frac{x}{k}\right) = O(x) \tag{3.24}$$

where $G(x) = \sum_{n \leq x} R(x/n)$. A suitable bound is then required for $G(x/k)$ with $k \leq x$. While ultimately, the proof shall be derived in a different way, proceeding in this manner will cast much light on what is eventually required. Begin by using the fact that $R(x) = \psi(x) - x$ to write

$$G(x) = \sum_{n \leq x} \left(\psi\left(\frac{x}{n}\right) - \frac{x}{n}\right) = \sum_{n \leq x} \psi\left(\frac{x}{n}\right) - x \sum_{n \leq x} \frac{1}{n} \,. \tag{3.25}$$

Recall that $\psi(x) = \sum_{n \leq x} \Lambda(n)$ and, by (3.8), that $\Lambda * u = \log$. Thus by Corollary 3.4.5,

$$\sum_{n \leq x} \psi\left(\frac{x}{n}\right) = \sum_{n \leq x} (\Lambda * u)(n) = \sum_{n \leq x} \log n.$$

Since $\log x$ is non-negative and increasing for $x \geq 1$, it follows by Proposition 3.5.2 that

$$\sum_{n \leq x} \psi\left(\frac{x}{n}\right) = \int_1^x \log t \, \mathrm{d}t + r_1(x) = x \log x - x + 1 + r_1(x) \tag{3.26}$$

where $|r_1(x)| \leq \log x$ for all $x \geq 1$. At the same time, by Proposition 3.5.1,

$$x \log x \leq x \sum_{n \leq x} \frac{1}{n} \leq x \log x + x,$$

an inequality that can be rewritten as

$$x \sum_{n \leq x} \frac{1}{n} = x \log x + \frac{x}{2} + r_2(x) \tag{3.27}$$

where $|r_2(x)| \leq x/2$. Recalling (3.25), it follows by (3.26) and (3.27) that

$$G(x) = -\frac{3x}{2} + r_1(x) + 1 - r_2(x)$$

where $|r_1(x)| \leq \log x$ and $|r_2(x)| \leq x/2$. Therefore,

$$-2x - \log x + 1 \leq G(x) \leq -x + \log x + 1. \tag{3.28}$$

Since $x \geq 1$, $|G(x)| \leq 2x + \log x - 1$. As the reader may care to verify, it therefore seems improbable that (3.24) holds with such a bound for G, and it is the x term that causes the difficulty. If one were to redefine $G(x) = \sum_{n \leq x} R(x/n) + cx$ for some constant c, an x term would still remain on at least one side of (3.28). Thus either one or both of the estimates for $\sum_{n \leq x} \psi(x/n)$ and $x \sum_{n \leq x} 1/n$ are not accurate enough. As it turns out, a more accurate estimate of $x \sum_{n \leq x} 1/n$ is available, due to Leonhard Euler. In an aside to Selberg's first formula, this estimate is verified. It is derived from results known as Euler's summation formulas.

Proposition 3.8.3 (Euler's summation formula, version 1) *Let m, n be integers and let f be a differentiable function on the interval $[m, n]$. Then*

$$\sum_{r=m+1}^{n} f(r) - \int_m^n f(t)\mathrm{d}t = \int_m^n (t - [t]) f'(t)\mathrm{d}t.$$

Proof. As $[t]$ is constant between integers, the idea here is to consider the integral

$$\int_{r-1}^{r} (t - [t]) f'(t)\mathrm{d}t = \int_{r-1}^{r} (t - r + 1) f'(t)\mathrm{d}t$$

where $r \in \mathbb{N}$. Let $u = t - r + 1$ and let $\mathrm{d}v/\mathrm{d}t = f'(t)$. Then $\mathrm{d}u/\mathrm{d}t = 1$ and $v = f(t) + c$ for some constant c. It follows that

$$\begin{aligned} \int_{r-1}^{r} (t - [t]) f'(t)\mathrm{d}t &= [(t - r + 1) f(t)]_{r-1}^{r} - \int_{r-1}^{r} f(t)\mathrm{d}t \\ &= f(r) - \int_{r-1}^{r} f(t)\mathrm{d}t. \end{aligned}$$

The result follows upon summing each term from $m+1$ to n.

Q.E.D.

Proposition 3.8.4 (version 2) *Let m be an integer and let x be real. Let f be differentiable on $[m, x]$. Then*

$$\sum_{m<r\leq x} f(r) - \int_m^x f(t)\mathrm{d}t = \int_m^x (t-[t])f'(t)\mathrm{d}t - (x-[x])f(x).$$

Proof. Let $[x] = n$. Then

$$\int_m^x (t-[t])f'(t)\mathrm{d}t = \int_m^n (t-[t])f'(t)\mathrm{d}t + \int_n^x (t-n)f'(t)\mathrm{d}t.$$

The result follows by Proposition 3.8.3 and integration by parts of the second term.

Q.E.D.

The second version of Euler's summation formula allows the estimate of $x\sum_{n\leq x} 1/n$ to be derived. It is in fact a special case of the following proposition.

Proposition 3.8.5 *Suppose that $f(t)$ is decreasing, differentiable and tends to 0 as $t \longrightarrow \infty$. Define*

$$S(x) = \sum_{1\leq r\leq x} f(r)$$

and

$$I(x) = \int_1^x f(t)\mathrm{d}t.$$

Then $S(x) - I(x) \longrightarrow L$ as $x \longrightarrow \infty$, where

$$L = f(1) + \int_1^\infty (t-[t])f'(t)\mathrm{d}t.$$

Also, $0 \leq L \leq f(1)$ and for all $x \geq 1$,

$$S(x) = I(x) + L + q(x),$$

where $|q(x)| \leq f(x)$. If x is an integer, then $0 \leq q(x) \leq f(x)$.

Proof. Take $m = 1$ in Proposition 3.8.4 and add $f(1)$ to both sides to obtain

$$S(x) - I(x) = f(1) + \int_1^x (t-[t])f'(t)\mathrm{d}t - (x-[x])f(x). \tag{3.29}$$

Clearly, for the first statement, one must show that the right-hand side has limit L as $x \longrightarrow \infty$. Begin by noting that as $f(x) \geq 0$ for all x, it follows that $0 \leq (x-[x])f(x) \leq f(x)$ for all x. Therefore, since $\lim_{x\longrightarrow\infty} f(x) = 0$, it follows that $\lim_{x\longrightarrow\infty}(x-[x])f(x)$ exists and

$$\lim_{x\longrightarrow\infty}(x-[x])f(x) = 0.$$

Next, note that since $f'(t) \leq 0$ for all t, it follows that $f'(t) \leq (t-[t])f'(t) \leq 0$. Now

$$\int_1^\infty f'(t)\mathrm{d}t = \lim_{R \longrightarrow \infty} \int_1^R f'(t)\mathrm{d}t = \lim_{R \longrightarrow \infty} (f(R) - f(1)) = -f(1),$$

so that

$$-f(1) \leq \int_1^\infty (t-[t])f'(t)\mathrm{d}t \leq 0;$$

or, in other words, $0 \leq L \leq f(1)$. To complete the proof, note that

$$S(x) - I(x) = L - \int_x^\infty (t-[t])f'(t)\mathrm{d}t - (x-[x])f(x),$$

and write

$$q(x) = -\int_x^\infty (t-[t])f'(t)\mathrm{d}t - (x-[x])f(x)$$

so that $S(x) = I(x) + L + q(x)$. Now

$$0 \leq -\int_x^\infty (t-[t])f'(t)\mathrm{d}t \leq -\int_x^\infty f'(t)\mathrm{d}t = f(x).$$

Also, $-f(x) \leq -(x-[x])f(x) \leq 0$. Clearly then, $|q(x)| \leq f(x)$, and if x is an integer, then $(x-[x])f(x) = 0$ so that $0 \leq q(x) \leq f(x)$.

Q.E.D.

The following proposition is a special case of the above with $f(t) = 1/t$, so that $S(x) = \sum_{r \leq x} 1/r$. The details are left to the reader.

Proposition 3.8.6 *The expression* $\sum_{r=1}^n \frac{1}{r} - \log n$ *tends to the limit* γ *as* $n \longrightarrow \infty$*, where*

$$\gamma = 1 - \int_1^\infty \frac{t-[t]}{t^2}\,\mathrm{d}t$$

(γ is called Euler's constant). Furthermore, $0 < \gamma < 1$ *and*

$$\sum_{1 \leq r \leq x} \frac{1}{r} = \log x + \gamma + q(x),$$

where $|q(x)| \leq 1/x$.

(It is worth mentioning that it is an open problem as to whether or not γ is rational). Clearly then,

$$x \sum_{n \leq x} \frac{1}{n} = x\log x + \gamma x + r_3(x) \tag{3.30}$$

where $|r_3(x)| \leq 1$. This ends our aside on Euler summation.

With a better estimate of $x\sum_{n\leq x} 1/n$, we can obtain a better estimate of

$$G(x) = \sum_{n\leq x} \psi\left(\frac{x}{n}\right) - x\sum_{n\leq x}\frac{1}{n}\,.$$

To be precise,

$$G(x) = -(\gamma+1)x + 1 + r_1(x) + r_3(x)$$

where $|r_1(x)| \leq \log x$ and $|r_3(x)| \leq 1$. We then have the following:

$$-(\gamma+1)x - \log x \leq G(x) \leq -(\gamma+1)x + \log x + 2. \tag{3.31}$$

The temptation may be to define a new function

$$G_1(x) = G(x) + (\gamma+1)x = \sum_{n\leq x} R\left(\frac{x}{n}\right) + (\gamma+1)x$$

so that $|G_1(x)| \leq \log x + 2$. However, recall that we want to prove Selberg's first formula by proving (3.24) where $G_1(x) = \sum_{n\leq x} F(x/n)$ for some function F. Clearly, G_1 is not of this form. Instead, let $G_2(x) = G(x) + (\gamma+1)[x]$, so that

$$G_2(x) = \sum_{n\leq x}\left(R\left(\frac{x}{n}\right) + \gamma + 1\right).$$

Thus, if $F(x) = R(x) + \gamma + 1$, then $G_2(x) = \sum_{n\leq x} F(x/n)$. It follows from (3.31) that

$$-(\gamma+1)\{x\} - \log x \leq G_2(x) \leq -(\gamma+1)\{x\} + \log x + 2;$$

Clearly this implies that $|G_2(x)| \leq \log x + 2$. Hence, the following lemma has been proved.

Lemma 3.8.7 *Let $F(x) = R(x) + \gamma + 1$ and $G(x) = \sum_{n\leq x} F(x/n)$. Then $|G(x)| \leq \log x + 2$ for all $x \geq 1$.*

Two equivalent versions of Selberg's formula are now outlined.

Proposition 3.8.8 *For all $x \geq 1$,*

$$R(x)\log x + \sum_{n\leq x}\Lambda(n) R\left(\frac{x}{n}\right) = O(x) \tag{3.32}$$

and

$$\psi(x)\log x + \sum_{n\leq x}\Lambda(n)\psi\left(\frac{x}{n}\right) = 2x\log x + O(x). \tag{3.33}$$

Proof. The equivalence of (3.32) and (3.33) is first established. Since $R(x) = \psi(x) - x$, it is clear that (3.32) can be rewritten as

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

Noting, by Theorem 3.5.3 that $x\sum_{n\le x}\frac{\Lambda(n)}{n} = x\log x + O(x)$, it follows that (3.32) can further be rewritten as

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

which is equivalent to (3.33). The proof of (3.32) is now outlined. Write $c = \gamma + 1$, where γ is Euler's constant, and write $F(x) = R(x) + c$. Define

$$J(x) = \sum_{k\le x}\mu(k)\log\frac{x}{k}\,G\left(\frac{x}{k}\right)$$

where $G(x) = \sum_{n\le x} F(x/n)$. The idea is that by Proposition 3.8.2, J can be re-expressed in terms of F, and ultimately R, while on the other hand, the fact that G is bounded is used to get a bound on J. More specifically,

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

where the last equality is clear once one recalls that $0 \le \psi(x) \le 2x$. We show that $J(x) = O(x)$ to complete the proof. By Lemma 3.8.7, $|G(x)| \le \log x + 2$. Therefore,

$$\begin{aligned}
|J(x)| &\le \sum_{k\le x}|\mu(k)|\left|\log\frac{x}{k}\right|\left|G\left(\frac{x}{k}\right)\right|\\
&\le \sum_{k\le x}\log\frac{x}{k}\left(\log\frac{x}{k}+2\right).
\end{aligned}$$

There exists a constant C such that

$$\log x(\log x + 2) \le Cx^{1/2}$$

for all $x \geq 1$. (The proof follows easily from the fact that $(\log x + 2)/x^{1/4} \longrightarrow 0$ as $x \longrightarrow \infty$, and is thus bounded.) Hence

$$\begin{aligned} |J(x)| &\leq \sum_{k \leq x} C \left(\frac{x}{k}\right)^{1/2} \\ &= Cx^{1/2} \sum_{k \leq x} \frac{1}{k^{1/2}}. \end{aligned}$$

The proof is clearly complete if there exist constants C_1 and x_0 such that for all $x > x_0$,

$$\sum_{k \leq x} \frac{1}{k^{1/2}} \leq C_1 x^{1/2}.$$

Using (3.11), and the fact that $1 \leq \int_1^0 t^{-1/2} \mathrm{d}t$, it is easy to show that when $C_1 = 2$, this inequality holds for all $x \geq 1$.

Q.E.D.

These formulas of Selberg are claimed to be vital to nearly all elementary proofs of the prime number theorem. However, taking (3.32), it seems at this stage that, bar R appearing in both terms, there is no apparent link to our ultimate aim of proving that, in the notation of the first section, we have $\alpha \leq \beta$. The same can be said of (3.33). However, one can tenuously suggest that the desired inequality $\alpha \leq \beta$ involves W on both sides, W of course being ultimately derived from R. This being the case, it is worth investigating what can be said of S in light of Selberg's formulas, and thus of W. Indeed, as the next result shows, R can be replaced by S in (3.32).

Proposition 3.8.9 *For all $x \geq 1$, we have*

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

Proof. Dividing through by t in (3.32), we have

$$\frac{R(t) \log t}{t} + \frac{1}{t} \sum_{n \leq t} \Lambda(n) R\left(\frac{t}{n}\right) = O(1).$$

This means that the left-hand side is bounded by some constant M. It follows that

$$\begin{aligned} \left| \int_1^x \left[\frac{R(t) \log t}{t} + \frac{1}{t} \sum_{n \leq t} \Lambda(n) R\left(\frac{t}{n}\right) \right] \mathrm{d}t \right| &\leq \int_1^x \left| \frac{R(t) \log t}{t} + \frac{1}{t} \sum_{n \leq t} \Lambda(n) R\left(\frac{t}{n}\right) \right| \mathrm{d}t \\ &\leq \int_1^x M \mathrm{d}t < Mx. \end{aligned}$$

Hence,

$$\int_1^x \frac{R(t)\log t}{t}\,\mathrm{d}t + \int_1^x \frac{1}{t}\sum_{n\le t}\Lambda(n)R\left(\frac{t}{n}\right)\mathrm{d}t = O(x).$$

As presently written, the integral and the sum in the second term on the left clearly cannot be interchanged. But $R(t/n) = 0$ whenever $n > t$. Therefore, when $1 \le t \le x$, it follows that

$$\sum_{n\le x}\Lambda(n)R\left(\frac{t}{n}\right) = \sum_{n\le t}\Lambda(n)R\left(\frac{t}{n}\right).$$

Thus

$$\int_1^x \frac{R(t)\log t}{t}\,\mathrm{d}t + \int_1^x \frac{1}{t}\sum_{n\le x}\Lambda(n)R\left(\frac{t}{n}\right)\mathrm{d}t = O(x);$$

that is,

$$\int_1^x \frac{R(t)\log t}{t}\,\mathrm{d}t + \sum_{n\le x}\Lambda(n)\int_1^x R\left(\frac{t}{n}\right)\frac{1}{t}\,\mathrm{d}t = O(x), \tag{3.34}$$

where equality of the left-hand sides of the last two statements follows upon reordering the terms. Considering the second term on the left (with fixed n), let $u = t/n$. Then $u = 1/n$ when $t = 1$ and $u = x/n$ when $t = x$. Also, $\mathrm{d}u/\mathrm{d}t = 1/n$ so that

$$\int_1^x R\left(\frac{t}{n}\right)\frac{1}{t}\,\mathrm{d}t = \int_1^x \frac{R(t/n)}{t/n}\frac{1}{n}\,\mathrm{d}t = \int_{1/n}^{x/n}\frac{R(u)}{u}\,\mathrm{d}u = \int_1^{x/n}\frac{R(u)}{u}\,\mathrm{d}u = S\left(\frac{x}{n}\right).$$

Note that due to R being constant over certain intervals, $R(u)/u$ has an anti-derivative when confined to any of these intervals, allowing the substitution method to be used. Given this result, it is enough to show that

$$\int_1^x \frac{R(u)\log u}{u}\,\mathrm{d}u = S(x)\log x + O(x),$$

or in other words, that the difference is $O(x)$. Thus, recalling the definition of S,

$$\begin{aligned} S(x)\log x - \int_1^x \frac{R(u)\log u}{u}\,\mathrm{d}u &= \int_1^x \frac{R(u)}{u}(\log x - \log u)\,\mathrm{d}u \\ &= \int_1^x \frac{R(u)}{u}\int_u^x \frac{1}{t}\,\mathrm{d}t\,\mathrm{d}u \\ &= \int_1^x \frac{1}{t}\int_1^t \frac{R(u)}{u}\,\mathrm{d}u\,\mathrm{d}t \\ &= \int_1^x \frac{S(t)}{t}\,\mathrm{d}t, \end{aligned}$$

where the second last equality comes from Proposition 3.6.2. We finish by noting that Proposition 3.6.3 (1) states that $|S(t)/t| \le 1$ for all t so that

$$\left|\int_1^x \frac{S(t)}{t}\,\mathrm{d}t\right| \le x.$$

Q.E.D.

Note the appearance of Λ in Proposition 3.8.9, and recall that, for the prime number theorem, we are concerned with the behaviour of the partial sums of Λ (namely ψ). However, if working with Λ and its partial sums were easy, one would imagine that the prime number theorem would be quickly arrived at. Therefore, it seems wiser to define a function q – in terms of Λ – for which a good estimate of the partial sums can be found. In fact, the following estimate is an equivalent version of Selberg's formula.

Corollary 3.8.10 *Let* $q(n) = \Lambda(n)\log n + (\Lambda * \Lambda)(n)$. *Then*

$$\sum_{n \le x} q(n) = 2x \log x + O(x)$$

and

$$\sum_{n \le x} (q(n) - 2\log n) = O(x).$$

Proof. The proof is derived from (3.33). Note first that by Abel's summation formula, that is, by Proposition 3.4.3,

$$\sum_{n \le x} \Lambda(n) \log n = \psi(x) \log x - \int_1^x \frac{\psi(t)}{t} \, \mathrm{d}t = \psi(x) \log x + O(x),$$

since $0 \le \psi(t) \le 2t$. Also, by Proposition 3.4.4,

$$\sum_{n \le x} (\Lambda * \Lambda)(n) = \sum_{n \le x} \Lambda(n) \psi\left(\frac{x}{n}\right).$$

Thus (3.33) may be rewritten as

$$\sum_{n \le x} \Lambda(n) \log n + \sum_{n \le x} (\Lambda * \Lambda)(n) = 2x \log x + O(x)$$

which, given the definition of $q(n)$, is the first statement. To establish the equivalence of the first and second statements, the reader may care to verify, using Abel summation, that

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

Thus the first statement may be rewritten as

$$\sum_{n \le x} q(n) = 2 \sum_{n \le x} \log n + O(x),$$

which is clearly the second statement. To finish, as (3.33) is true, so too are the equivalent statements.

Q.E.D.

The first result states that the partial sums of $q(n)$ are within a constant times x of $2x \log x$. As x gets bigger, this difference is relatively small. By the second (equivalent) result, the difference between the partial sums of $q(n)$ and $2 \log n$ is also relatively small. Having thus got "good" estimates for the partial sums of $q(n)$, the next result, derived from Proposition 3.8.9, provides the platform for the use of those estimates.

Proposition 3.8.11 *There is a constant K_1 such that for all $x \geq 1$,*

$$|S(x)|(\log x)^2 \leq \sum_{n \leq x} q(n) \left| S\left(\frac{x}{n}\right) \right| + K_1 x \log x.$$

Proof. It is easily verified that $a(x) + b(x) = O(x)$ implies $|a(x)| = |b(x)| + O(x)$. Applying this to Proposition 3.8.9, and multiplying through by $\log x$, it follows that, for all $x \geq 1$,

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

The rest of the proof consists of showing that, for all $x \geq 1$,

$$\left| \log x \sum_{n \leq x} \Lambda(n) S\left(\frac{x}{n}\right) \right| = \sum_{n \leq x} q(n) \left| S\left(\frac{x}{n}\right) \right| + O(x \log x).$$

The idea is to restructure the left-hand side with an eye on the definition of $q(n)$. To begin, introduce $\Lambda(n) \log n$ by rewriting the left-hand side as

$$\left| \sum_{n \leq x} \Lambda(n) S\left(\frac{x}{n}\right) \log \frac{x}{n} + \sum_{n \leq x} \Lambda(n) \log n \; S\left(\frac{x}{n}\right) \right|.$$

This in turn is equal to

$$\left| \sum_{n \leq x} \Lambda(n) S\left(\frac{x}{n}\right) \log \frac{x}{n} + \sum_{n \leq x} q_1(n) \; S\left(\frac{x}{n}\right) + \sum_{n \leq x} (\Lambda * \Lambda)(n) \; S\left(\frac{x}{n}\right) \right|,$$

where $q_1(n) = \Lambda(n) \log n - (\Lambda * \Lambda)(n)$. Note that $q(n)$ is not introduced in a similar manner. This would lead to difficulties later on. Now $|q_1(n)| \leq q(n)$, a useful fact given that the proof is completed on showing that, for all $x \geq 1$,

$$\left| A + \sum_{n \leq x} q_1(n) \; S\left(\frac{x}{n}\right) + B \right| = \sum_{n \leq x} q(n) \left| S\left(\frac{x}{n}\right) \right| + O(x \log x),$$

where

$$A = \sum_{n \leq x} \Lambda(n) S\left(\frac{x}{n}\right) \log \frac{x}{n}$$

and

$$B = \sum_{n \le x} (\Lambda * \Lambda)(n) \; S\left(\frac{x}{n}\right).$$

Clearly, it is sufficient to show that, for all $x \ge 1$,

$$|A + B| + \left| \sum_{n \le x} q_1(n) \; S\left(\frac{x}{n}\right) \right| = \sum_{n \le x} q(n) \left| S\left(\frac{x}{n}\right) \right| + O(x \log x),$$

and since $|q_1(n)| \le q(n)$, it suffices to show that $|A+B| = O(x \log x)$ for all $x \ge 1$. Begin by noting that

$$\begin{aligned}
B &= \sum_{n \le x} (\Lambda * \Lambda)(n) \; S\left(\frac{x}{n}\right) \\
&= \sum_{n \le x} \sum_{jk=n} \Lambda(j)\Lambda(k) S\left(\frac{x}{jk}\right) \\
&= \sum_{jk \le x} \Lambda(j)\Lambda(k) S\left(\frac{x}{jk}\right) \\
&= \sum_{k \le x} \sum_{j \le x/k} \Lambda(j)\Lambda(k) S\left(\frac{x}{jk}\right) \\
&= \sum_{k \le x} \Lambda(k) \sum_{j \le x/k} \Lambda(j) S\left(\frac{x}{jk}\right).
\end{aligned}$$

Therefore,

$$|A + B| = \sum_{k \le x} \Lambda(k) \left[S\left(\frac{x}{k}\right) \log \frac{x}{k} + \sum_{j \le x/k} \Lambda(j) S\left(\frac{x}{jk}\right) \right],$$

so that

$$|A + B| \le \sum_{k \le x} \Lambda(k) \left| S\left(\frac{x}{k}\right) \log \frac{x}{k} + \sum_{j \le x/k} \Lambda(j) S\left(\frac{x}{jk}\right) \right|. \tag{3.35}$$

Recalling Proposition 3.8.9,

$$S\left(\frac{x}{k}\right) \log \frac{x}{k} + \sum_{j \le x/k} \Lambda(j) S\left(\frac{x}{jk}\right) = O\left(\frac{x}{k}\right)$$

whenever $k \le x$. Therefore, there exists a constant C such that

$$\left| S\left(\frac{x}{k}\right) \log \frac{x}{k} + \sum_{j \le x/k} \Lambda(j) S\left(\frac{x}{jk}\right) \right| \le C \frac{x}{k}$$

whenever $k \le x$. It follows from (3.35) that

$$|A+B| \le Cx \sum_{k \le x} \frac{\Lambda(k)}{k} \; .$$

Thus, to complete the proof, it suffices to show that there exists a constant ρ such that $\sum_{k\le x} \Lambda(k)/k \le \rho \log x$ for all $x \ge 1$. Recall from Theorem 3.5.3 that $\sum_{k\le x} \Lambda(k)/k \le \log x + 2$ for all $x > 1$. It is easy to show that there exists a constant ρ satisfying $\log x + 2 \le \rho \log x$ for all $x \ge 2$, meaning that $\sum_{k\le x} \Lambda(k)/k \le \rho \log x$ for all $x \ge 2$. Since $\Lambda(1) = 0$, this bound actually holds for all $x \ge 1$.

Q.E.D.

Thus an inequality featuring q and S is established, and the next argument shows, using Corollary 3.8.10, that $q(n)$ can be replaced by $2\log n$. Recall that a good estimate for $\sum_{n\le x} q(n)$ is $\sum_{n\le x} 2\log n$.

Proposition 3.8.12 *There is a constant K_2 such that for all sufficiently large x,*

$$\sum_{n\le x} q(n) \left| S\left(\frac{x}{n}\right)\right| = 2\sum_{n\le x} \log n \left| S\left(\frac{x}{n}\right)\right| + K_2 x \log x.$$

Proof. Write $a(n) = q(n) - 2\log n$. In the new notation, we require

$$\sum_{n\le x} a(n) \left| S\left(\frac{x}{n}\right)\right| = O(x\log x). \tag{3.36}$$

The first thing one might consider is obtaining the absolute values of the terms of (3.36), but we have no bound for $a(n)$. However, by Corollary 3.8.10, where $A(n) = \sum_{j\le n} a(j)$, $A(n)$ is bounded by n. Therefore, it is desirable to re-express the left-hand side of (3.36) in terms of $A(n)$, and then Abel summation (Proposition 3.4.1) gives

$$\sum_{n\le x} a(n) \left| S\left(\frac{x}{n}\right)\right| = \sum_{n\le x-1} A(n) \left[\left| S\left(\frac{x}{n}\right)\right| - \left| S\left(\frac{x}{n+1}\right)\right|\right] + A([x]) S\left(\frac{x}{[x]}\right).$$

In fact, since $S(x/([x]+1)) = 0$,

$$\sum_{n\le x} a(n) \left| S\left(\frac{x}{n}\right)\right| = \sum_{n\le x} A(n) \left[\left| S\left(\frac{x}{n}\right)\right| - \left| S\left(\frac{x}{n+1}\right)\right|\right].$$

It follows that

$$\left|\sum_{n\le x} a(n) \left| S\left(\frac{x}{n}\right)\right|\right| \le \sum_{n\le x} |A(n)| \left|\left| S\left(\frac{x}{n}\right)\right| - \left| S\left(\frac{x}{n+1}\right)\right|\right|.$$

Noting Proposition 3.6.3, we have

$$\left|\left| S\left(\frac{x}{n}\right)\right| - \left| S\left(\frac{x}{n+1}\right)\right|\right| \le \left| S\left(\frac{x}{n}\right) - S\left(\frac{x}{n+1}\right)\right| \le \frac{x}{n} - \frac{x}{n+1} = x\left(\frac{1}{n(n+1)}\right),$$

and since there exists C such that $|A(n)| \leq Cn$ for all n, it follows that

$$\left|\sum_{n\leq x} a(n)\left|S\left(\frac{x}{n}\right)\right|\right| \leq Cx\sum_{n\leq x}\frac{1}{n+1} \leq Cx(\log x+1),$$

where the last inequality is clear when one considers the example directly after Proposition 3.5.1. The result follows.

Q.E.D.

Next, the discrete sum of Proposition 3.8.12 is related to the corresponding integral, which in light of what must be proved, namely (3.23), must at some stage make an entrance.

Proposition 3.8.13 *There exists a constant C' such that, for all $x \geq 1$,*

$$\sum_{n\leq x}\log n\left|S\left(\frac{x}{n}\right)\right| \leq \int_1^x \log t\left|S\left(\frac{x}{t}\right)\right|\,\mathrm{d}t + C'x.$$

Proof. Let $n \leq x$ and take t such that $n \leq t < n+1$. By Proposition 3.6.3, we have

$$\left|S\left(\frac{x}{n}\right) - S\left(\frac{x}{t}\right)\right| \leq \frac{x}{n} - \frac{x}{t}\,.$$

Hence,

$$\left|S\left(\frac{x}{n}\right)\right| \leq \left|S\left(\frac{x}{t}\right)\right| + \frac{x}{n(n+1)}\,.$$

Therefore,

$$\begin{aligned}\log n\left|S\left(\frac{x}{n}\right)\right| &\leq \log n\int_n^{n+1}\left[\left|S\left(\frac{x}{t}\right)\right| + \frac{x}{n(n+1)}\right]\mathrm{d}t\\ &\leq \int_n^{n+1}\log t\left|S\left(\frac{x}{t}\right)\right|\mathrm{d}t + x\frac{\log n}{n(n+1)}\,.\end{aligned}$$

Summing over all $n \leq x$, it should be clear that

$$\sum_{n\leq x}\log n\left|S\left(\frac{x}{n}\right)\right| \leq \int_1^{[x]+1}\log t\left|S\left(\frac{x}{t}\right)\right|\mathrm{d}t + x\sum_{n\leq x}\frac{\log n}{n(n+1)}\,.$$

The proof is completed on noting that the series

$$\sum_{n=1}^{\infty}\frac{\log n}{n(n+1)}$$

is convergent and that $S(x/t) = 0$ when $t > x$.

Q.E.D.

Combining the last three results, we obtain:

Proposition 3.8.14 *There exists a constant K_3 such that for all x large enough,*

$$|S(x)|(\log x)^2 \leq 2\int_1^x \log t \left|S\left(\frac{x}{t}\right)\right| \mathrm{d}t + K_3 x \log x.$$

At this point, W is re-introduced. Recall that $W(y) = e^{-y}S(e^y)$.

Corollary 3.8.15 *For K_3 as above and for all sufficiently large y,*

$$|W(y)| \leq \frac{2}{y^2}\int_0^y (y-u)|W(u)|\mathrm{d}u + \frac{K_3}{y}\ .$$

Proof. In view of how W is defined, it is clearly a good idea to let $x = e^y$ (thus $y = \log x$) in Proposition 3.8.14. With this and the substitution $t = e^u$, Proposition 3.8.14 can be rewritten to state

$$|S(e^y)|y^2 \leq 2\int_0^y u\left|S\left(e^{y-u}\right)\right| e^u \mathrm{d}u + K_3 y e^y$$

for all sufficiently large y. With the further substitution $S(e^y) = e^y W(y)$, it follows that

$$y^2 e^y |W(y)| \leq 2\int_0^y u e^{y-u}\left|W(y-u)\right| e^u \mathrm{d}u + K_3 y e^y$$

for all sufficiently large y, or in other words,

$$|W(y)| \leq \frac{2}{y^2}\int_0^y u|W(y-u)|\mathrm{d}u + \frac{K_3}{y}$$

for all sufficiently large y. Finally, with the substitution $v = y - u$, one can easily check that

$$\int_0^y u|W(y-u)|\mathrm{d}u = \int_0^y (y-v)|W(v)|\mathrm{d}v.$$

Q.E.D.

This result gives a bound for $|W(y)|$ in terms of an integral involving W. Thus one may reasonably hope that the proof of (3.23), and thus of the prime number theorem itself, is now within reach. The next proposition is indeed that proof.

Proposition 3.8.16 *The inequality* (3.23) *holds; or equivalently, in the notation of Proposition 3.6.5, $\alpha \leq \beta$.*

Proof. The rough idea is to establish inequalities involving α and β on either side of the inequality in Corollary 3.8.15. Note first that, by definition,

$$\alpha = \lim_{y\longrightarrow\infty} \sup_{x\geq y} |W(x)|,$$

where the supremum function is decreasing. Therefore, given any $y > 0$,

$$\alpha \leq \sup_{x\geq y} |W(x)|.$$

Thus, given $\delta > 0$ and $y > 0$, there exists $x \geq y$ such that

$$\alpha - \delta < |W(x)|. \tag{3.37}$$

Second, note that

$$\begin{aligned} \int_0^x (x-u)|W(u)|\mathrm{d}u &= \int_0^x |W(u)| \int_u^x \mathrm{d}v \, \mathrm{d}u \\ &= \int_0^x \int_0^v |W(u)|\mathrm{d}u \, \mathrm{d}v, \end{aligned} \tag{3.38}$$

where the second equality follows from Proposition 3.6.2 and the continuity of $|W|$, which follows from Proposition 3.6.4. Now, by definition,

$$\beta = \lim_{x \longrightarrow \infty} \sup_{v \geq x} \frac{1}{v} \int_0^v |W(u)|\mathrm{d}u.$$

Thus, given $\epsilon > 0$, there exists x_1 such that

$$\sup_{v \geq x_1} \frac{1}{v} \int_0^v |W(u)|\mathrm{d}u \leq \beta + \epsilon.$$

In other words, given $\epsilon > 0$, there exists x_1 such that for all $v \geq x_1$,

$$\frac{1}{v} \int_0^v |W(u)|\mathrm{d}u \leq \beta + \epsilon.$$

Hence, for $x \geq x_1$, it follows that

$$\int_{x_1}^x \int_0^v |W(u)|\mathrm{d}u \, \mathrm{d}v \leq \int_{x_1}^x (\beta + \epsilon) v \, \mathrm{d}v = \frac{1}{2}(\beta + \epsilon)(x^2 - {x_1}^2).$$

Also, since $|W(u)| \leq 1$ for all u,

$$\int_0^{x_1} \int_0^v |W(u)|\mathrm{d}u \, \mathrm{d}v \leq \int_0^{x_1} \int_0^v \mathrm{d}u \, \mathrm{d}v = \int_0^{x_1} v \, \mathrm{d}v = \frac{1}{2}{x_1}^2.$$

Combining these last two statements with (3.38), we obtain

$$\int_0^x (x-u)|W(u)|\mathrm{d}u \leq \frac{1}{2}(\beta + \epsilon)x^2 + \frac{1}{2}{x_1}^2. \tag{3.39}$$

Choosing $y > x_1$, (3.37) and (3.39) hold for some $x \geq y$. Since y can be arbitrarily large, this is true for arbitrarily large x, and it easily follows, by Corollary 3.8.15, that for such x,

$$\alpha - \delta < \beta + \epsilon + \frac{{x_1}^2}{x^2} + \frac{K_3}{x} \; .$$

This being the case, it follows that $\alpha \leq \beta + \epsilon + \delta$, with the result following, since this is true for all $\epsilon > 0$ and $\delta > 0$.

Q.E.D.

Chapter 4

Newman's short proof of the prime number theorem

4.1 Sufficient conditions for the short proof

Throughout, we write $s = \sigma + it$, where σ and t are real. As with the classical analytic proof of the prime number theorem, outlined in Chapter 2, the Riemann zeta function plays a crucial role in this short proof, based on a proof by Donald Newman in 1980. For the reader interested in the original proof, see [13]. For this interpretation of Newman's proof, the author has heavily relied upon [2] and [17]. For our previous two proofs, proving that $\psi(x)$ is asymptotic to x was shown to be sufficient. Likewise, proving that $\theta(x)$ is asymptotic to x — as was suspected by Chebyshev and backed up by the numerical evidence — is sufficient, and the proof of this is the main point of Newman's proof, where the reader may recall the definition of θ from Chapter 1.

Definition 4.1.1 *Define the function $\theta : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ by*

$$\theta(x) = \sum_{p \le x} \log p$$

where the sum is over all primes p less than or equal to x.

First, it must be shown that proving that $\theta(x)$ is asymptotic to x is sufficient for the prime number theorem.

Lemma 4.1.1 *If*

$$\theta(x) \sim x \quad as \quad x \longrightarrow \infty, \tag{4.1}$$

then the prime number theorem is true.

Proof. This is shown by relating $\theta(x)$ to $\pi(x)\log x$, and given the assumption that $\theta(x) \sim x$ as $x \longrightarrow \infty$, we hope to relate x to $\pi(x)\log x$. Begin by noting that

$$\theta(x) \le \sum_{p \le x} \log x = \pi(x) \log x.$$

Then, for any $\epsilon > 0$, (without loss of generality, $\epsilon < 1/2$) and for sufficiently large x,

$$1 - \epsilon \leq \frac{\theta(x)}{x} \leq \frac{\pi(x)\log x}{x} .$$

Second,

$$\begin{aligned}
\theta(x) &\geq \sum_{x^{1-\epsilon}<p\leq x} \log p \geq \sum_{x^{1-\epsilon}<p\leq x} \log x^{1-\epsilon} \\
&= (1-\epsilon)\log x \sum_{x^{1-\epsilon}<p\leq x} 1 \\
&= (1-\epsilon)\log x(\pi(x) - \pi(x^{1-\epsilon})) \\
&\geq (1-\epsilon)\log x(\pi(x) - x^{1-\epsilon}).
\end{aligned}$$

Thus bounds have been established on $\theta(x)$ involving the term $\pi(x)\log x$. The above inequality can be restated as

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

Since

$$\lim_{x\longrightarrow\infty} -(1-\epsilon)\frac{\log x}{x^{\epsilon}} = 0,$$

then for any $\delta > 0$, and for sufficiently large x,

$$-\delta \leq -(1-\epsilon)\frac{\log x}{x^{\epsilon}} .$$

Clearly then,

$$(1-\epsilon)\frac{\pi(x)\log x}{x} - \delta \leq \frac{\theta(x)}{x} .$$

Since $\theta(x) \sim x$ as $x \longrightarrow \infty$, it follows that for sufficiently large x,

$$\frac{\theta(x)}{x} \leq \delta + 1.$$

By these last two inequalities, for sufficiently large x,

$$\frac{\pi(x)\log x}{x} \leq \frac{2\delta + 1}{1-\epsilon} .$$

Choosing $\delta = \frac{\epsilon(1-2\epsilon)}{2} > 0$, it easily follows that for sufficiently large x,

$$\frac{\pi(x)\log x}{x} \leq 1 + 2\epsilon,$$

so that

$$\lim_{x\longrightarrow\infty} \frac{\pi(x)\log x}{x} = 1.$$

Q.E.D.

It remains then to prove (4.1). The main idea is to acquire some knowledge about θ, and build upon it. In Theorem 3.2.3, it has already been shown that $\theta(x) \leq (\log 4)x$. An alternative proof that $\theta(x) = O(x)$ is now outlined, although the estimate is not as good.

Theorem 4.1.2 $\theta(x) = O(x)$.

Proof. Begin by noting that

$$\begin{aligned} e^{\theta(2n)-\theta(n)} &= \frac{e^{\sum_{p\leq 2n}\log p}}{e^{\sum_{p\leq n}\log p}} \\ &= \frac{\prod_{p\leq 2n} e^{\log p}}{\prod_{p\leq n} e^{\log p}} \\ &= \prod_{n<p\leq 2n} e^{\log p} \\ &= \prod_{n+1\leq p\leq 2n} p. \end{aligned}$$

During the course of Theorem 3.2.3, it was shown that

$$\prod_{n+2\leq p\leq 2n+1} p \leq 2^{2n}.$$

In the same way, it can be shown that

$$\prod_{n+1\leq p\leq 2n} p \leq 2^{2n},$$

that is, $e^{\theta(2n)-\theta(n)} \leq 2^{2n}$. It follows that $\theta(2n) - \theta(n) \leq \log 2^{2n} = 2n\log 2$. In particular, $\theta(2^r) - \theta(2^{r-1}) \leq 2^r \log 2, r \in \mathbb{N}$, so that given $m \in \mathbb{N}$,

$$\theta(2^m) = \sum_{r=1}^{m}(\theta(2^r) - \theta(2^{r-1})) \leq \sum_{r=1}^{m} 2^r \log 2 = (2^{m+1} - 2)\log 2 < 2^{m+1}\log 2.$$

Now, for any $x \geq 1$, there exists m such that $2^{m-1} \leq x < 2^m$. It follows that

$$\theta(x) \leq \theta(2^m) < 2^{m+1}\log 2 = (4\log 2)2^{m-1} \leq (4\log 2)x$$

for all $x \geq 0$, as required.

Q.E.D.

A series of sufficient conditions are now established for (4.1), with Theorem 4.1.2 being used more than once in their establishment. That the first of these conditions is sufficient seems reasonable at first glance, it being the convergence of $\int_1^\infty \frac{\theta(u)-u}{u^2}\mathrm{d}u$. This does suggest the relative closeness of $\theta(x)$ and x, although the fact that θ is a non-decreasing function is also required. However, many other convergent integrals suggest the same, so why is this one chosen? The reason lies in a concept known as the Laplace transform, and in a result known as the analytic theorem. Even to the reader familiar with these, the choice of integral may still be obscure. The Laplace transform and the analytic theorem will be discussed after the following theorem.

Theorem 4.1.3 *If*

$$\int_1^\infty \frac{\theta(u)-u}{u^2}\mathrm{d}u$$

is a convergent integral, then

$$\theta(x) \sim x \quad \textit{as} \quad x \longrightarrow \infty.$$

Proof. Suppose that

$$\int_1^\infty \frac{\theta(u)-u}{u^2}\mathrm{d}u$$

is a convergent integral. Assume then that

$$\lim_{x\longrightarrow\infty} \frac{\theta(x)}{x} \neq 1.$$

Therefore, for some $\epsilon > 0$, there exist arbitrarily large x such that

$$\left|\frac{\theta(x)}{x} - 1\right| \geq \epsilon.$$

It easily follows that there exist arbitrarily large x such that either $\theta(x) \leq (1-\epsilon)x$ or $\theta(x) \geq (1+\epsilon)x$.

Case 1: Suppose that there exist arbitrarily large x such that $\theta(x) \leq (1-\epsilon)x$. Letting $\lambda = 1-\epsilon$, we have $\theta(x) \leq \lambda x$, for some $\lambda < 1$. Since θ is non-decreasing,

$$\begin{aligned}
\int_{\lambda x}^{x} \frac{\theta(u)-u}{u^2}\mathrm{d}u &\leq \int_{\lambda x}^{x} \frac{\theta(x)-u}{u^2}\mathrm{d}u \\
&\leq \int_{\lambda x}^{x} \frac{\lambda x-u}{u^2}\mathrm{d}u \\
&= \lambda x \int_{\lambda x}^{x} \frac{1}{u^2}\mathrm{d}u - \int_{\lambda x}^{x} \frac{1}{u}\mathrm{d}u \\
&= \left[\frac{-\lambda x}{u} - \log u\right]_{\lambda x}^{x} \\
&= -\lambda - \log x + 1 + \log \lambda x \\
&= -\lambda + 1 + \log \lambda,
\end{aligned}$$

which is a constant. It is easily shown that

$$-\lambda + 1 + \log\lambda = \int_\lambda^1 \frac{\lambda - u}{u^2}\mathrm{d}u < 0.$$

Therefore,

$$\int_{\lambda x}^{x} \frac{\theta(u) - u}{u^2} \mathrm{d}u \leq -\lambda + 1 + \log \lambda < 0$$

for arbitrarily large x, which contradicts the initial supposition.

Case 2: Suppose that there exist arbitrarily large x such that $\theta(x) \geq (1+\epsilon)x$. Letting $\lambda = 1 + \epsilon$, we have $\theta(x) \geq \lambda x$, for some $\lambda > 1$. Since θ is non-decreasing,

$$\begin{aligned}\int_{x}^{\lambda x} \frac{\theta(u) - u}{u^2} \mathrm{d}u &\geq \int_{x}^{\lambda x} \frac{\theta(x) - u}{u^2} \mathrm{d}u \\ &\geq \int_{x}^{\lambda x} \frac{\lambda x - u}{u^2} \mathrm{d}u \\ &= \lambda - 1 - \log \lambda > 0,\end{aligned}$$

for arbitrarily large x. This again contradicts the initial supposition.

Q.E.D.

4.2 The Laplace transform

The Laplace transform is now defined.

Definition 4.2.1 *Let $h : [0, \infty) \longrightarrow \mathbb{R}$ be a piecewise continuous function. The Laplace transform of h is the complex function $(\mathcal{L}h)(s)$ given by*

$$(\mathcal{L}h)(s) = \int_{0}^{\infty} h(x) e^{-sx} \mathrm{d}x$$

whenever the integral converges.

At first glance, this integral seems unrelated to $\int_1^\infty \frac{\theta(u)-u}{u^2}\mathrm{d}u$, whose convergence is a sufficient condition for the prime number theorem. However, making the change of variables $u = e^x$, it is easy to show that

$$\int_{1}^{\infty} \frac{\theta(u) - u}{u^2} \mathrm{d}u = \int_{0}^{\infty} (\theta(e^x)e^{-x} - 1) \mathrm{d}x.$$

However, this is hardly a useful Laplace transform, with $h(x)$ from Definition 4.2.1 looking complicated! In fact, letting $H(x) = \theta(e^x)e^{-x} - 1$, we proceed indirectly by investigating the Laplace transform of $H(x)$. The analytic theorem below gives conditions on this transform which imply the convergence of $\int_0^\infty H(x)\mathrm{d}x$, as required for the prime number theorem. The analytic theorem is now stated and the proof is given later.

Theorem 4.2.1 (The analytic theorem) *Let $f : [0, \infty) \longrightarrow \mathbb{R}$ be a bounded and piecewise continuous function whose Laplace transform $g(s) = \int_0^\infty f(x)e^{-sx}\mathrm{d}x$, initially defined for $\sigma > 0$, extends to an analytic function for $\sigma \geq 0$. Then the improper integral $\int_0^\infty f(x)\mathrm{d}x$ converges, and its value is $g(0)$.*

Assuming that the analytic theorem is true, it remains to show that H satisfies the various conditions mentioned. Firstly, by Theorem 4.1.2, there exists a real constant $C > 0$ such that $\theta(e^x) \leq Ce^x$ for all $x \geq 0$. Thus $H(x) = \theta(e^x)e^{-x} - 1 \leq C - 1$ and is bounded. Also, e^x is continuous and θ is piecewise continuous, so H is piecewise continuous. The Laplace transform of H now needs to be computed and shown to be analytic for $\sigma > 0$. The computation is necessary if we are to extend the Laplace transform of H to an analytic function for $\sigma \geq 0$. To make the statement of the next theorem a little easier, we first define the function $\Phi(s)$ on the domain $\sigma > 1$ by the formula

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

Theorem 4.2.2 *For* $\sigma > 0$,

$$(\mathcal{L}H)(s) = \frac{\Phi(s+1)}{s+1} - \frac{1}{s}$$

and is analytic in this domain.

Proof. Begin by noting that

$$\begin{aligned}(\mathcal{L}H)(s) &= \int_0^\infty e^{-sx}(\theta(e^x)e^{-x} - 1)\mathrm{d}x \\ &= \int_0^\infty e^{-(s+1)x}\theta(e^x)\mathrm{d}x - \int_0^\infty e^{-sx}\mathrm{d}x.\end{aligned}$$

It is easy to check that $\int_0^\infty e^{-sx}\mathrm{d}x = 1/s$, which is convergent and an analytic function for $\sigma > 0$. It remains to show firstly that $\int_0^\infty e^{-(s+1)x}\theta(e^x)\mathrm{d}x$ is convergent and an analytic function for $\sigma > 0$, and secondly that

$$\int_0^\infty e^{-(s+1)x}\theta(e^x)\mathrm{d}x = \frac{\Phi(s+1)}{s+1} .$$

By Theorem 4.1.2, $\theta(e^x) \leq Ce^x$, hence

$$\int_0^T |e^{-(s+1)x}\theta(e^x)|\mathrm{d}x \leq \int_0^T Ce^{-\sigma x}\mathrm{d}x = \frac{C}{\sigma}(1 - e^{-\sigma T}).$$

Now $1 - e^{-\sigma T} < 1$, thus when $\sigma > 0$, we obtain the estimate

$$\int_0^\infty |e^{-(s+1)x}\theta(e^x)|\mathrm{d}x \leq \frac{C}{\sigma},$$

which means that $\int_0^\infty e^{-(s+1)x}\theta(e^x)\mathrm{d}x$ is convergent for $\sigma > 0$.

To show that $F(s) = \int_0^\infty e^{-(s+1)x}\theta(e^x)\mathrm{d}x$ is analytic for $\sigma > 0$, write $F_n(s) = \int_n^{n+1} e^{-(s+1)x}\theta(e^x)\mathrm{d}x$ for each $n \in \mathbb{N} \cup \{0\}$, so that

$$F(s) = \sum_{n=0}^\infty F_n(s).$$

Given Theorem 2.4.6, the idea is to show that for each $n \in \mathbb{N} \cup \{0\}, F_n(s)$ is analytic for $\sigma > 0$, and that, given $\delta > 0$, the series $\sum_{n=0}^{\infty} F_n(s)$ is uniformly convergent for all $\sigma > \delta$. Letting D be the half–plane $\sigma > 0$, note that $e^{-(s+1)x}\theta(e^x)$ is a piecewise continuous function of $x, n \leq x \leq n+1$, for fixed s and an analytic function of $s \in D$ for fixed x. Thus by the theorem stated below — which shall be proved later — $F_n(s)$ is analytic in D.

Theorem 4.2.3 *Let D be a simply connected domain of the complex plane. Suppose $g : D \times [a,b] \longrightarrow \mathbb{C}$ is a piecewise continuous function of $x, a \leq x \leq b$, for fixed s and an analytic function of $s \in D$ for fixed x. Then*

$$f(s) = \int_a^b g(s,x)\mathrm{d}x$$

is analytic in D.

To show that the series is uniformly convergent, choose $\delta > 0$. Then for $\sigma > \delta$,

$$|F_n(s)| \leq \int_n^{n+1} |e^{-(s+1)x}\theta(e^x)|\,\mathrm{d}x \leq \int_n^{n+1} |Ce^{-sx}|\,\mathrm{d}x$$

for some positive constant C, by Theorem 4.1.2. It follows that

$$|F_n(s)| \leq C\int_n^{n+1} e^{-\sigma x}\,\mathrm{d}x \leq Ce^{-\sigma n} \leq C(e^{-\delta})^n.$$

Note that $\sum_{n=1}^{\infty} (e^{-\delta})^n$ is a convergent geometric series. Thus by an analogous argument to proving uniform convergence in Theorem 2.4.3, the series $\sum_{n=1}^{\infty} F_n(s)$ converges uniformly for $\sigma > \delta$. Therefore, $F(s)$ is analytic for $\sigma > 0$.

To compute the integral, we take advantage of the fact that $\theta(e^x) = \theta(p_n)$ whenever $p_n < e^x < p_{n+1}$ (p_n denotes the nth prime number). In other words, $\theta(e^x) = \theta(p_n)$ whenever $\log p_n < x < \log p_{n+1}$. This is obvious from the definition of θ. Also, $\theta(e^x) = 0$ whenever $0 < x < \log 2 = \log p_1$. Thus

$$\begin{aligned}\int_0^{\infty} e^{-(s+1)x}\theta(e^x)\mathrm{d}x &= \sum_{n=1}^{\infty}\int_{\log p_n}^{\log p_{n+1}} e^{-(s+1)x}\theta(e^x)\mathrm{d}x \\ &= \sum_{n=1}^{\infty}\theta(p_n)\int_{\log p_n}^{\log p_{n+1}} e^{-(s+1)x}\mathrm{d}x = \frac{1}{s+1}\sum_{n=1}^{\infty}\theta(p_n)(p_n^{-(s+1)} - p_{n+1}^{-(s+1)}).\end{aligned}$$

Bearing in mind the fact that $\theta(p_n) - \theta(p_{n-1}) = \log p_n$, it follows that

$$\begin{aligned}\int_0^{\infty} e^{-(s+1)x}\theta(e^x)\mathrm{d}x &= \frac{1}{s+1}\left(\sum_{n=1}^{\infty}\theta(p_n)p_n^{-(s+1)} - \sum_{n=1}^{\infty}\theta(p_n)p_{n+1}^{-(s+1)}\right) \\ &= \frac{1}{s+1}\left(\theta(2)2^{-(s+1)} + \sum_{n=2}^{\infty}\theta(p_n)p_n^{-(s+1)} - \sum_{n=2}^{\infty}\theta(p_{n-1})p_n^{-(s+1)}\right) \\ &= \frac{1}{s+1}\left(\frac{\log 2}{2^{s+1}} + \sum_{n=2}^{\infty}\frac{\log p_n}{{p_n}^{s+1}}\right) = \frac{\Phi(s+1)}{s+1},\end{aligned}$$

as required.

Q.E.D.

Now that the Laplace transform of H has been computed and shown to be an analytic function for $\sigma > 0$, and assuming that the analytic theorem holds, it remains to show that the Laplace transform of H extends to an analytic function for $\sigma \geq 0$. Before this is attempted, it is worth listing the following results relating to the Laplace transform, which were proved in specific instances during the course of Theorem 4.2.2.

Lemma 4.2.4 *Let $g, h : [0, \infty) \longrightarrow \mathbb{R}$ be piecewise continuous functions such that, for real numbers C and B, h satisfies $|h(x)| \leq Ce^{Bx}$ for $0 \leq x < \infty$. Then,*

1) $\mathcal{L}(g + h) = \mathcal{L}g + \mathcal{L}h$.

2) If $g(x) = e^{at}h(x)$, then $(\mathcal{L}g)(s) = (\mathcal{L}h)(s - a)$ for $\sigma > B + a$.

3) $\mathcal{L}(1)(s) = 1/s$.

Lemma 4.2.5 *If there exist real constants B and C such that $|h(x)| \leq Ce^{Bx}$ for $0 \leq x < \infty$, then the integral defining the Laplace transform of $h(x)$ converges absolutely and is an analytic function for $\sigma > B$.*

4.3 Extending the Laplace transform of H

In this section, the Laplace transform of H, that is, $\frac{\Phi(s+1)}{s+1} - \frac{1}{s}$, is extended to an analytic function on the domain $\sigma \geq 0$. Clearly, this is equivalent to showing that

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

can be extended to an analytic function on the domain $\sigma \geq 1$. The key here is to link $\Phi(s)$ to $\zeta(s)$, as discussed in Section 2.2. It was shown — during Theorem 2.2.7 — that

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

for $\sigma > 1$, thus $\Phi(s)$ and $\zeta(s)$ can both be defined in terms of the primes. Since $\Phi(s)$ involves a log term, it is helpful to have an equation involving logs and $\zeta(s)$. Recall that Theorem 2.11 states that

$$-\zeta'(s)/\zeta(s) = \sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s}$$

where $\Lambda : \mathbb{N} \longrightarrow \mathbb{R}$ was defined in Chapter 1 by

$$\Lambda(n) = \begin{cases} \log p, & \text{if } n = p^r \text{ with } p \text{ prime, } r \in \mathbb{N} \\ 0, & \text{otherwise.} \end{cases}$$

It follows that

$$\begin{aligned}-\zeta'(s)/\zeta(s) &= \sum_p \sum_{r=1}^{\infty} \frac{\Lambda(p^r)}{(p^r)^s} = \sum_p \log p \sum_{r=1}^{\infty} \frac{1}{(p^r)^s} \\ &= \sum_p \frac{p^{-s}\log p}{1-p^{-s}} = \sum_p \frac{\log p}{p^s - 1}\ .\end{aligned}$$

Recall that $\Phi(s) = \sum_p \frac{\log p}{p^s}$. Therefore, using the identity

$$\frac{1}{p^s} = \frac{1}{p^s-1} - \frac{1}{p^s(p^s-1)},$$

it follows that

$$\Phi(s) = -\frac{\zeta'(s)}{\zeta(s)} - \sum_p \frac{\log p}{p^s(p^s-1)}\ . \tag{4.2}$$

Thus information about $\Phi(s)$ will follow from a study of the two terms on the right. To extend $\frac{\Phi(s)}{s} - \frac{1}{s-1}$ to an analytic function on the domain $\sigma \geq 1$, it is clearly useful to try and analytically extend $\Phi(s)$ to the domain $\sigma \geq 1$. To begin, consider the infinite sum appearing in (4.2). If we define $f : \mathbb{N} \longrightarrow \mathbb{R}$ by

$$f(n) = \begin{cases} \log n, & \text{if } n \text{ prime} \\ 0, & \text{otherwise,} \end{cases}$$

then

$$\sum_p \frac{\log p}{p^s(p^s-1)} = \sum_{n=2}^{\infty} \frac{f(n)}{n^s(n^s-1)},$$

which is clearly a series of analytic functions. Thus by Theorem 2.4.6, to show that this series converges to an analytic function on the domain $\sigma \geq 1$, it suffices to show that it converges uniformly on the domain $\sigma \geq 1$. First, for $\sigma \geq 1$, it is easy to show that

$$\left| \frac{f(n)}{n^s(n^s-1)} \right| \leq \frac{\log n}{n(n-1)}\ .$$

Now $\sum_{n=2}^{\infty} \frac{\log n}{n(n-1)}$ is convergent by comparison with $\sum_{n=2}^{\infty} \frac{1}{n^{3/2}}$, so choosing $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $m \geq N$,

$$\left| \sum_{n=m+1}^{\infty} \frac{f(n)}{n^s(n^s-1)} \right| \leq \sum_{n=m+1}^{\infty} \left| \frac{f(n)}{n^s(n^s-1)} \right| \leq \sum_{n=m+1}^{\infty} \frac{\log n}{n(n-1)} < \epsilon$$

whenever $\sigma \geq 1$.

An analytic extension of $\Phi(s)$ will follow if one exists for $-\zeta'(s)/\zeta(s)$. As a result of Theorem 2.4.2, this will be the case if $\zeta(s)$ extends to an analytic function on the domain $\sigma \geq 1$, and for this extension, $\zeta(s)$ has no zeros on the domain $\sigma \geq 1$. Closely related to an analytic function is the concept of a meromorphic function.

Definition 4.3.1 *A function f is said to be meromorphic on a domain D if f is analytic in D except at isolated poles.*

In fact, Theorem 2.4.3 states that $\zeta(s)$ extends meromorphically to the domain $\sigma > 0$, with only a simple pole at $s = 1$ which is a simple pole with residue 1. In this section, an alternative proof is outlined. The idea is to determine what the statement of Theorem 2.4.3 implies. If the converse is true, it is sufficient to prove the implied statement. Begin by noting that if Theorem 2.4.3 is true, then

$$\lim_{s \longrightarrow 1}(s-1)\zeta(s) = Res(\zeta, 1) = 1. \tag{4.3}$$

This equation involves $\zeta(s)$, which we are trying to show is meromorphic. Motivated by the fact that analytic functions are easier to work with than meromorphic ones, note that if a function f has an isolated singularity at s_0 and if

$$\lim_{s \longrightarrow s_0}(s-s_0)f(s) = 0,$$

then f has a removable singularity at s_0. For the reader interested in a proof of this result, see [1]. If this is the only singularity in a domain D, then by suitably "removing" the singularity, f can be extended to an analytic function on D. If Theorem 2.4.3 is true, then by (4.3),

$$\lim_{s \longrightarrow 1}(s-1)\left(\zeta(s) - \frac{1}{s-1}\right) = 0,$$

so that $\zeta(s) - \frac{1}{s-1}$ has a removable singularity at $s = 1$. Indeed, if Theorem 2.4.3 is true, it can be shown that this is the only singularity on the half-plane $\sigma > 0$. Thus $\zeta(s) - \frac{1}{s-1}$ can be extended to an analytic function on the half-plane $\sigma > 0$. It is not necessary to prove this is the only singularity. Of more importance is asking, if $\zeta(s) - \frac{1}{s-1}$ extends to an analytic function on the half-plane $\sigma > 0$, does Theorem 2.4.3 follow? Indeed, it follows easily upon writing $\zeta(s) = (\zeta(s) - \frac{1}{s-1}) + \frac{1}{s-1}$ and finding poles and residues with regard to Definition 2.3.7 and (2.22). Therefore, Theorem 2.4.3 amounts to proving that $\zeta(s) - \frac{1}{s-1}$, analytic on the domain $\sigma > 1$, extends to an analytic function on the half-plane $\sigma > 0$.

We now analytically extend $\zeta(s) - 1/(s-1)$ to the half-plane $\sigma > 0$, as required for Theorem 2.4.3. This is achieved by initially expressing $\zeta(s) - 1/(s-1)$ as a series for $\sigma > 1$. Theorem 2.4.6 is then applied to this series.

Theorem 4.3.1 *The function $\zeta(s) - 1/(s-1)$ extends analytically to the half-plane $\sigma > 0$.*

Proof. First, the function $1/(s-1)$ is rewritten as an integral. It is easy to show that, for $\sigma > 1$,

$$\int_1^\infty \frac{1}{x^s}\,\mathrm{d}x = \frac{1}{s-1}\ .$$

It follows that, for $\sigma > 1$,

$$\begin{aligned}
\zeta(s) - \frac{1}{s-1} &= \sum_{n=1}^{\infty} \frac{1}{n^s} - \int_1^{\infty} \frac{1}{x^s}\,\mathrm{d}x \\
&= \sum_{n=1}^{\infty} \frac{1}{n^s} - \sum_{n=1}^{\infty} \int_n^{n+1} \frac{1}{x^s}\,\mathrm{d}x \\
&= \sum_{n=1}^{\infty} \int_n^{n+1} \frac{1}{n^s}\,\mathrm{d}x - \sum_{n=1}^{\infty} \int_n^{n+1} \frac{1}{x^s}\,\mathrm{d}x \\
&= \sum_{n=1}^{\infty} \int_n^{n+1} \left(\frac{1}{n^s} - \frac{1}{x^s}\right)\mathrm{d}x. \qquad (4.4)
\end{aligned}$$

Let

$$F_n(s) = \int_n^{n+1} \left(\frac{1}{n^s} - \frac{1}{x^s}\right)\mathrm{d}x.$$

Then, by Theorem 2.4.6, if

1) for every $n \in \mathbb{N}, F_n(s)$ is analytic in $\mathbb{C}$,

   and

2) for every $\delta > 0$, the series (4.4) converges uniformly for $\sigma > \delta$,

then the series (4.4) is analytic for $\sigma > \delta$. The series would then be analytic on the half-plane $\sigma > 0$, since $\delta > 0$ is arbitrary, and thus $\zeta(s) - 1/(s-1)$ would extend analytically to the half-plane $\sigma > 0$. Therefore, it remains to prove 1 and 2. Proving 1 is a simple exercise. To prove 2, choose $\delta > 0$. Then for $\sigma > \delta$,

$$\begin{aligned}
\left|\int_n^{n+1}\left(\frac{1}{n^s}-\frac{1}{x^s}\right)\mathrm{d}x\right| &= \left|-\int_n^{n+1}(x^{-s}-n^{-s})\mathrm{d}x\right| \\
&= \left|\int_n^{n+1}\int_n^x su^{-s-1}\mathrm{d}u\,\mathrm{d}x\right| \\
&\leq \int_n^{n+1}\left|\int_n^x su^{-s-1}\mathrm{d}u\right|\mathrm{d}x \\
&\leq \max_{n\leq x\leq n+1}\left|\int_n^x su^{-s-1}\mathrm{d}u\right| \\
&\leq \max_{n\leq x\leq n+1}\int_n^x |su^{-s-1}|\mathrm{d}u \\
&= \int_n^{n+1}|su^{-s-1}|\mathrm{d}u \\
&= |s|\int_n^{n+1}u^{-\sigma-1}\mathrm{d}u \\
&\leq \frac{|s|}{n^{\sigma+1}} \leq \frac{|s|}{n^{\delta+1}}\ .
\end{aligned}$$

Since the series

$$\sum_{n=0}^{\infty}\frac{|s|}{n^{\delta+1}}$$

converges, it follows by an analogous argument to proving uniform convergence in Theorem 2.4.3 that the series (4.4) is uniformly convergent for $\sigma > \delta$. This completes the proof.

Q.E.D.

Thus we also have an alternative proof of Theorem 2.4.3.

Also of interest are the zeros of $\zeta(s)$ on the domain $\sigma \geq 1$. In Theorem 2.2.7 it was shown that no zeros lie in the domain $\sigma > 1$, while similarly no zeros lie on the line $\sigma = 1$ by Theorem 2.5.5. While Newman uses a different method to prove the non-vanishing of $\zeta(s)$ on the line $\sigma = 1$, one due in essence to Hadamard, it is hardly any shorter than the proof already outlined, and thus is not included here. For the reader interested in Hadamard's work, see [10]. Another alternative proof is to be found in [2], a proof which is extremely similar to that already outlined.

An analytic extension of $\zeta(s)$ was sought, and it has been established by Theorem 2.4.3 that $\zeta(s)$ is meromorphic on the domain $\sigma \geq 1$ except for a simple pole at $s = 1$. This together with Theorem 2.5.5 imply that $-\zeta'(s)/\zeta(s)$ extends to an analytic function on the domain $\sigma \geq 1$, except for a singularity at $s = 1$. The question is, what kind of singularity is it? Using the same method as for finding the residue of $G(s)X^{s-1}$ as outlined in section

2.6, it is not difficult to check that the singularity at $s = 1$ is a simple pole with residue 1. Thus $-\zeta'(s)/\zeta(s)$ extends meromorphically to the domain $\sigma \geq 1$, and in light of (4.2), $\Phi(s)$ is meromorphic on the same domain, with a simple pole at $s = 1$ with residue 1. Therefore, analogous to proving that $\zeta(s) - 1/(s-1)$ can be extended analytically to the domain $\sigma > 0$, $\Phi(s) - 1/(s-1)$ can be extended analytically to the domain $\sigma \geq 1$.

Recall the aim of this section was to show $\frac{\Phi(s)}{s} - \frac{1}{s-1}$ extends analytically to the domain $\sigma \geq 1$. The following lemma finally demonstrates this.

Lemma 4.3.2 *The function* $\frac{\Phi(s)}{s} - \frac{1}{s-1}$, *analytic for* $\sigma > 1$, *extends analytically to the domain* $\sigma \geq 1$.

Proof. The idea is to express $\frac{\Phi(s)}{s} - \frac{1}{s-1}$ in terms of $\Phi(s) - 1/(s-1)$, which has already been shown to extend analytically to the domain $\sigma \geq 1$. Write

$$\frac{\Phi(s)}{s} - \frac{1}{s-1} = \frac{1}{s}\left(\Phi(s) - \frac{s}{s-1}\right) = \frac{1}{s}\left(\Phi(s) - \frac{1}{s-1} - 1\right).$$

The required analytic extension clearly follows from this re-expression.

Q.E.D.

It has been shown that the Laplace transform of H extends to an analytic function on the domain $\sigma \geq 0$. All that remains for the prime number theorem is to prove the analytic theorem, namely Theorem 4.2.1. This is done in the next section.

4.4 The analytic theorem

A proof of the analytic theorem is now outlined. Two results, Morera's theorem and Fubini's theorem, from complex analysis and measure theory respectively, are needed to prove an auxiliary result, namely Theorem 4.2.3. Morera's theorem is merely stated here, as the proof would take us too far afield. For the interested reader, a proof is outlined in [1]. Fubini's theorem is not stated, but is well known and allows us to change the order of integration in the proof of Theorem 4.2.3. A proof of Fubini's theorem can be found in [14]. Theorem 4.2.3 is proved after the statement of Morera's theorem.

Theorem 4.4.1 (Morera's Theorem) *Let f be a continuous function on an open set D. If*

$$\int_\Gamma f(s)\mathrm{d}s = 0$$

whenever Γ is the boundary of a closed rectangle in D, then f is analytic on D.

Proof of Theorem 4.2.3.

Suppose $g(s,x)$ is discontinuous at the points $a_1, a_2, \ldots, a_n \in [a,b]$ where $a = a_0 < a_1 < \cdots < a_n < a_{n+1} = b$. If it can be shown that

$$f_m(s) = \int_{a_m}^{a_{m+1}} g(s,x)\mathrm{d}x$$

is analytic in D for $m \in \{0, \ldots, n\}$, then the result is clearly implied. Choose $m \in \{0, \ldots, n\}$. As f_m is a continuous function of s (which follows from the definition of continuity), then according to Morera's theorem, it need only be shown that $\int_\Gamma f(s)\mathrm{d}s = 0$ for any rectangle boundary $\Gamma \subset D$.

Now

$$\int_\Gamma f_m(s)\mathrm{d}s = \int_\Gamma \left(\int_{a_m}^{a_{m+1}} g(s,x)\mathrm{d}x \right) \mathrm{d}s,$$

at which point the order of integration can be reversed by application of Fubini's theorem. Thus

$$\int_\Gamma f_m(s)\mathrm{d}s = \int_{a_m}^{a_{m+1}} \left(\int_\Gamma g(s,x)\mathrm{d}s \right) \mathrm{d}x.$$

Since g is analytic in s, Cauchy's residue theorem implies that

$$\int_\Gamma f_m(s)\mathrm{d}s = \int_{a_m}^{a_{m+1}} 0 \ \mathrm{d}x = 0,$$

as required.

Q.E.D.

The analytic theorem will now be proved.

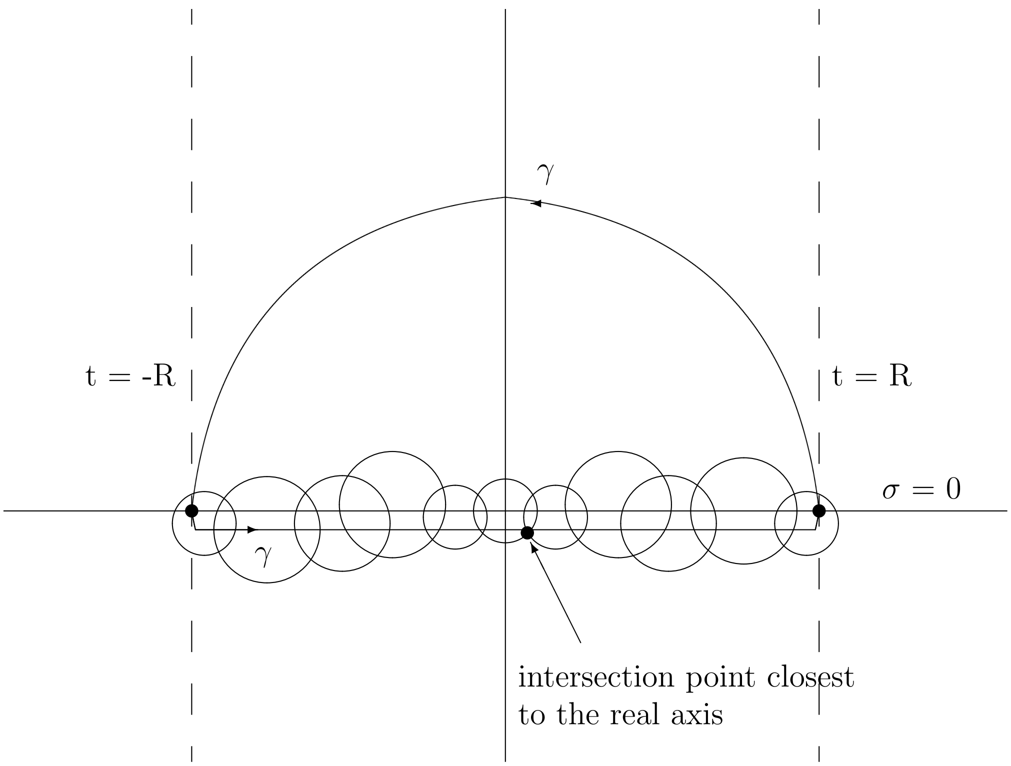
Proof of Theorem 4.2.1. (The analytic theorem)

First, the theorem is restated:

Let $f : [0,\infty) \longrightarrow \mathbb{R}$ be a bounded and piecewise continuous function whose Laplace transform $g(s) = \int_0^\infty f(t)e^{-sx}\mathrm{d}x$, initially defined for $\sigma > 0$, extends to an analytic function for $\sigma \geq 0$. Then, the improper integral $\int_0^\infty f(x)\mathrm{d}x$ converges, and its value is $g(0)$.

Proof. Consider the sequence of functions $g_T(s) = \int_0^T f(x)e^{-sx}\mathrm{d}x$. By Theorem 4.2.3, these functions are all entire, since $f(x)e^{-sx}$ is a piecewise continuous function of $x, 0 \leq x \leq T$, for fixed s and an analytic function of $s \in \mathbb{C}$ for fixed x. Also, $g_T(0) = \int_0^T f(x)\mathrm{d}x$. Thus we are trying to show that $\lim_{T\longrightarrow\infty} g_T(0)$ exists and equals $g(0)$.

Recall that we write $s = \sigma + it$, where σ and t are real. Choose a large real number R and consider the line segment $L := \{s \in \mathbb{C} : |s| \leq R, \sigma = 0\}$. As the Laplace transform g is analytic for $\sigma \geq 0, g$ is analytic along L. Therefore, taking any point on L, g is analytic on some open ball containing the point. Draw such a ball around each point of L. The union of these balls is an open cover of L and, as L is compact, a finite subcover exists. Let this finite subcover be $B_1, \ldots, B_n$. For $i \in \{1, \ldots, n\}$, the boundary of B_i intersects L at either one or two points. Because each ball is open, the points are contained in other balls in the finite subcover, so that B_i intersects the boundary of a finite number of balls in the subcover at points below the real axis. As the endpoints of L are contained in some open ball of the subcover, at least one ball's boundary intersects the line $t = -R$ at a point below the real axis and at least one ball's boundary intersects the line $t = R$ at a point below the real axis.



The number of intersection points between the boundaries of balls and between the boundaries of balls and the lines $t = -R$ and $t = R$ is finite. Taking the smallest distance between the intersection points and the real axis, it is clear that for some $\delta > 0$ (depending on R), g is analytic inside and on γ, where the curve γ (orientated counterclockwise) is the boundary of the region $\{s \in \mathbb{C} : |s| \leq R, \sigma > -\delta\}$.

To prove that $\lim_{T \longrightarrow \infty} g_T(0) = g(0)$, for all $T > 0$, an analytic function G_T will be defined inside and on γ such that $G_T(0) = g(0) - g_T(0)$, after which Theorem 2.5.3 is applied to obtain

$$g(0) - g_T(0) = G_T(0) = \frac{1}{2\pi i} \int_\gamma \frac{G_T(s)}{s} \mathrm{d}s. \tag{4.5}$$

The idea is to express the integral in (4.5), and thus $g(0) - g_T(0)$, in terms of other suitable integrals. The M-L formula (Theorem 2.3.3) is then used to find bounds on the limsups of the absolute values of these integrals, with a bound on $\limsup_{T\longrightarrow\infty}|g(0)-g_T(0)|$ following. We then use the fact that R is arbitrarily large to show that $\limsup_{T\longrightarrow\infty}|g(0)-g_T(0)|=0$. For the reader unfamiliar with limsups, see section 3.6. The question is, what is a suitable definition for G_T? A wrong choice will lead to an unsuitable bound on the integral in (4.5), or even lead to no bound at all. To get the required expression for $G_T(0)$, it is perhaps natural to write $G_T(s)=g(s)-g_T(s)$ for all $s\in\mathbb{C}$. However, note that, if $B=\sup_{t\geq 0}|f(x)|$, then

$$|g(s)-g_T(s)|=\left|\int_T^\infty f(t)e^{-sx}\mathrm{d}x\right|\leq B\int_T^\infty |e^{-sx}|\mathrm{d}t=B\int_T^\infty e^{-\sigma x}\mathrm{d}x=\frac{Be^{-\sigma T}}{\sigma}$$

for $\sigma>0$. Also, $1/|s|=1/R$ for $s\in\gamma_+$, where γ_+ is the semicircle $\{|s|=R,\sigma\geq 0\}$. Note that γ_+ is a piece of the curve γ. Together, these estimates establish that $|G_T(s)/s|\leq\frac{Be^{-\sigma T}}{\sigma R}$ for $s\in\gamma_+$. As no bound exists for $e^{-\sigma T}/\sigma$, one cannot apply the M-L formula to get a bound on

$$\frac{1}{2\pi i}\int_{\gamma_+}\frac{G_T(s)}{s}\mathrm{d}s.$$

This is not a totally wasted exercise. It is easy to see that by letting $G_T(s)=(g(s)-g_T(s))e^{sT}$ for all $s\in\mathbb{C}$, we also have $G_T(0)=g(0)-g_T(0)$ and $|G_T(s)/s|\leq\frac{B}{\sigma R}$ for $s\in\gamma_+$. As before, $G_T(s)/s$ is unbounded on γ_+. Indeed, it could be asked how this definition of G_T helps at all, since $e^{-\sigma T}<1$ for $\sigma>0$. The need for the exponential term is apparent when one considers the case $\sigma<0$.

To define G_T so that $G_T(0)=g(0)-g_T(0)$ and $G_T(s)/s$ is bounded on γ_+, it is sufficient to write $G_T(s)=(g(s)-g_T(s))e^{sT}F(s)$ where $F(0)=1$ and for all $s\in\gamma_+$, $|F(s)|\leq M\sigma$ for some constant M. Functions that obviously fulfill the first condition are $1+s$ and $1+s^2$. However, the imaginary part t of s cannot be eliminated from any upper bound on these functions. The function $1+s/N$ also proves unsuccessful for any real constant N. However, given a real constant N,

$$\left|1+\frac{s^2}{N}\right|=\left|\frac{N+(\sigma+it)^2}{N}\right|=\left|\frac{N+\sigma^2-t^2+2\sigma ti}{N}\right|.$$

Choosing $N=\sigma^2+t^2=R^2$, it follows that

$$\left|1+\frac{s^2}{N}\right|=\left|\frac{2\sigma^2+2\sigma ti}{R^2}\right|=\frac{2\sigma|\sigma+it|}{R^2}=\frac{2\sigma}{R}\ .$$

Therefore, if $G_T:\mathbb{C}\longrightarrow\mathbb{C}$ is defined by $(g(s)-g_T(s))e^{sT}(1+\frac{s^2}{R^2})$, then $G_T(0)=g(0)-g_T(0)$ and

$$|G_T(s)/s|\leq\frac{2B}{R^2}$$

for all $s\in\gamma_+$. Also, G_T is analytic inside and on γ_+. By the M-L formula,

$$\left|\frac{1}{2\pi i}\int_{\gamma_+}\frac{G_T(s)}{s}\mathrm{d}s\right|\leq B/R.$$

It is tempting to apply a similar argument to the remainder of the curve γ, that is, to $\gamma_- = \gamma \cap \{\sigma < 0\}$. However, due to the change in sign of σ, the best estimate for $|g(s) - g_T(s)|$ is ∞ if one proceeds in the manner above. The approach instead is to consider the functions $g(s)$ and $g_T(s)$ separately.

Since $g_T(s)$ is entire, so is $g_T(s)e^{sT}(1 + s^2/R^2)$. Therefore, if $\gamma_-' = \{|s| = R, \sigma < 0\}$ (a semicircle), then by path independence,

$$\int_{\gamma_-} (g(s) - g_T(s))e^{sT} \left(1 + \frac{s^2}{R^2}\right) \frac{\mathrm{d}s}{s} = \int_{\gamma_-'} (g(s) - g_T(s))e^{sT} \left(1 + \frac{s^2}{R^2}\right) \frac{\mathrm{d}s}{s} .$$

Similar to above, it is easy to show that

$$|g_T(s)| < \frac{Be^{-\sigma T}}{|\sigma|}$$

and that

$$\left| e^{sT} \left(1 + \frac{s^2}{R^2}\right) \frac{1}{s} \right| = e^{\sigma T} \frac{2|\sigma|}{R^2}$$

for $s \in \gamma_-'$. Thus by the M-L formula, it follows that

$$\left| \frac{1}{2\pi i} \int_{\gamma_-} g_T(s)e^{sT} \left(1 + \frac{s^2}{R^2}\right) \frac{\mathrm{d}s}{s} \right| \leq B/R. \tag{4.6}$$

Finally, to estimate

$$\left| \frac{1}{2\pi i} \int_{\gamma_-} g(s)e^{sT} \left(1 + \frac{s^2}{R^2}\right) \frac{\mathrm{d}s}{s} \right|,$$

begin by noting that as T tends to ∞, e^{sT} tends to 0 uniformly on compact subsets of γ_-. (If the reader wishes to verify this, use the fact that $\sigma < 0$ and σ takes a maximum on closed and bounded sets.) Note then that as $g(s)\left(1 + \frac{s^2}{R^2}\right)\frac{1}{s}$ is independent of T,

$$g(s)e^{sT} \left(1 + \frac{s^2}{R^2}\right) \frac{1}{s} \longrightarrow 0$$

uniformly as $T \longrightarrow \infty$ on compact subsets of γ_-. In other words, if C is any compact subset of $\gamma_-, L(C)$ is the length of the curve C and $g(s,T) = g(s)e^{sT}\left(1 + \frac{s^2}{R^2}\right)\frac{1}{s}$, then given $\epsilon > 0$, there exists $T_0 > 0$ such that for all $T > T_0, |g(s,T)| < \epsilon/L(C)$ for all $s \in C$. Using the M-L formula, it follows that

$$\int_C |g(s,T)|\mathrm{d}s < \epsilon,$$

and thus

$$\lim_{T \longrightarrow \infty} \int_C g(s,T)\mathrm{d}s = 0.$$

That

$$\lim_{T\longrightarrow\infty} \frac{1}{2\pi i}\int_{\gamma_-} g(s)e^{sT}\left(1+\frac{s^2}{R^2}\right)\frac{\mathrm{d}s}{s} = 0 \tag{4.7}$$

follows easily. Letting

$$L_T = \frac{1}{2\pi i}\int_{\gamma} \frac{G_T(s)}{s}\mathrm{d}s,$$

$$M_T = -\frac{1}{2\pi i}\int_{\gamma_-} g_T(s)e^{sT}\left(1+\frac{s^2}{R^2}\right)\frac{\mathrm{d}s}{s}$$

and

$$N_T = \frac{1}{2\pi i}\int_{\gamma_-} g(s)e^{sT}\left(1+\frac{s^2}{R^2}\right)\frac{\mathrm{d}s}{s},$$

we have $g(0) - g_T(0) = L_T + M_T + N_T$. To summarise our results, $|L_T| \leq B/R$ for all $T > 0$, $|M_T| \leq B/R$ for all $T > 0$, and $\lim_{T\longrightarrow\infty} N_T = 0$. It follows that

$$\limsup_{T\longrightarrow\infty} |L_T| + \limsup_{T\longrightarrow\infty} |M_T| + \limsup_{T\longrightarrow\infty} |N_T| \leq \frac{2B}{R},$$

or in other words,

$$\lim_{T\longrightarrow\infty}\left(\sup_{Y\geq T}|L_Y| + \sup_{Y\geq T}|M_Y| + \sup_{Y\geq T}|N_Y|\right) \leq \frac{2B}{R}\ .$$

Hence,

$$\lim_{T\longrightarrow\infty}\left(\sup_{Y\geq T}|L_Y + M_Y + N_Y|\right) \leq \frac{2B}{R},$$

that is,

$$\limsup_{T\longrightarrow\infty} |g(0) - g_T(0)| \leq \frac{2B}{R}\ .$$

Now $R > 0$ was arbitrary, so in fact, $\limsup_{T\longrightarrow\infty} |g(0) - g_T(0)| = 0$, and the required statement follows.

Q.E.D.

Thus the prime number theorem has been proved.

Bibliography

[1] JOSEPH BAK, DONALD J. NEWMAN. *Complex Analysis, 2nd ed.*, Springer-Verlag, (1995).

[2] MATT BAKER, DENNIS CLARK. *The Prime Number Theorem*, A paper, (2001).

[3] PAVNUTY LVOVICH CHEBYSHEV. Sur la fonction qui détermine la totalité des nombres premiers inférieurs á une limite donnée, Mémoires présentés á lAcadémie Impériale des Sciences de St.-Pétersbourg par divers Savants et lus dans ses Assemblées, Bd. 6, S. (1851), pp. 141-157.

[4] PAFNUTY LVOVICH CHEBYSHEV. Mémoire sur les nombres premiers, *J. de Math. Pures Appl.* (1) 17 (1852), pp. 366-390; reprinted in *Oeuvres 1* (1899), pp. 49-70.

[5] WWL CHEN. *Distribution of Prime Numbers*, A Series of Lectures, (1981).

[6] PAUL ERDŐS. On a new method in elementary number theory which leads to an elementary proof of the prime number theorem, *Proc. Nat. Acad. Scis. U.S.A.* 35 (1949), pp. 374-384.

[7] EBERHARD FREITAG, ROLF BUSAM. *Complex Analyis, 2nd ed.*, Springer (2005), pp. 106.

[8] DORIAN GOLDFELD. *The elementary proof of the prime number theorem: an historical perspective*, A paper. (YEAR?)

[9] JACQUES HADAMARD. Étude sur les proprietés des fonctions entiéres et en particulier dune fonction considérée par Riemann, *J. de Math. Pures Appl.* (4) 9 (1893), pp. 171-215; reprinted in *Oeuvres de Jacques Hadamard*, C.N.R.S., Paris, 1968, vol 1, pp. 103-147. 13

[10] JACQUES HADAMARD. Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques, *Bull. Soc. Math. France*, 24 (1896), pp. 199-220; reprinted in *Oeuvres de Jacques Hadamard*, C.N.R.S., Paris, 1968, vol 1, pp. 189-210.

[11] G.J.O. JAMESON. *The Prime Number Theorem*, Cambridge University Press, (2003).

[12] NORMAN LEVINSON. A motivated account of an elementary proof of the prime number theorem, *American Math. Monthly*, 76 (1969), pp. 225-245.

[13] DONALD J. NEWMAN. A Simple Analytic Proof of the Prime Number Theorem. *American Math. Monthly*, 87 (1980) pp. 693-697.

[14] WALTER RUDIN. *Real and Complex Analysis, 3rd ed.*, McGraw Hill, (1987).

[15] ATLE SELBERG. An elementary proof of the primenumber theorem, *Ann. of Math.* (2) 50 (1949), pp. 305-313; reprinted in *Atle Selberg Collected Papers*, SpringerVerlag, Berlin Heidelberg New York, 1989, vol 1, pp. 379-387.

[16] C.J. DE LA VALLÉE POUSSIN, Recherches analytiques sur la théorie des nombres premiers, *Ann. Soc. Sci. Bruxelles*, 20 (1896), pp. 183-256.

[17] DON ZAGIER. Newman's short proof of the prime number theorem, *American Math. Monthly*, 104 (1997), pp. 705-708.