An operator theoretic approach to the Riemann Hypothesis

Ramon van de Scheur

Master's thesis
2023:E74

Lund University

Faculty of Science
Centre for Mathematical Sciences
Mathematics

CENTRUM SCIENTIARUM MATHEMATICARUM

Abstract

In 2023 an operator theoretic approach to the Prime Number Theorem was introduced by Olsen in [12]. In this thesis this approach is examined and applied to give a new, operator theoretic, proof of a different version of the Prime Number Theorem and of the Prime Number Theorem for arithmetic progressions. This approach is then expanded and rearranged to focus on the approximation error in the Prime Number Theorem and then operator theoretic equivalents of the Riemann Hypothesis are derived. Lastly, further properties of the operators involved in these operator theoretic equivalents are discussed.

Popular scientific summary

Prime numbers are whole numbers greater than 1 which are divisible only by 1 and itself. The first few prime numbers are 2, 3, 5, 7, and 11. Not only are they very important in mathematics, but they are also widely used in our everyday lives. For example, you have probably surfed the internet securely today using encryption that is based on prime numbers (look up the RSA algorithm if you are interested).

It has been known since the ancient Greeks that there are infinitely many prime numbers (Euclid's theorem), but as we go further along the number line it gets harder and harder to find them, as they become rarer and more difficult to recognize. Mathematicians in the 19th century have found a formula[1] which tells us approximately where they are, but we have difficulty to this day in determining how accurate this formula is. Of course, we could check its accuracy with a computer for, say, the first several million numbers, but many results in number theory (the area of mathematics where this formula comes from) depend on its accuracy in the long run.

In the 19th century, the work of the great mathematician Bernhard Riemann led to the hypothesis that this formula is as accurate as it can be (still not accurate enough to worry encryption experts though). This is now known as the Riemann Hypothesis and many of the greatest minds of the past 150 years have tried to prove or disprove it. It was included by the famous mathematician David Hilbert in 1900 on his list of 23 important mathematical problems for the 20th century and in 2000 was also named by the Clay Mathematics Institute as one of its 10 Millennium Prize Problems (with a reward of one million dollars for anyone who solves one!). Because this problem is connected to so many results in number theory and has been able to withstand so many attacks for so long and was included on these lists it is considered to be one of the most important unsolved problems in mathematics.

While the formula comes from number theory, this thesis studies its accuracy using a different area of mathematics, called operator theory, and shows how this accuracy is related to certain operators and their properties. Hopefully, a different perspective will shed more light on this accuracy and the Riemann Hypothesis.

[1] For those who know about logarithms and integrals: the formula says that the number of primes less than N is approximately $\int_2^N \frac{1}{\ln(t)}dt$. It turns out that this can in turn be approximated by $\frac{N}{\ln N}$ and that this last approximation is the same as saying that the Nth prime number can be found near $N \ln N$.

Contents

1 Introduction

Prime numbers have been studied since at least the ancient Greeks. One of the most significant results of this time period was Euclid's theorem; that there are infinitely many primes. Since then, mathematicians, especially in the area of number theory, have wondered how these primes were distributed along the number line. More specifically, they started studying the prime counting function (see Definition 3.6):

$$\pi(x) := \sum_{p \leq x, p \text{ prime}} 1.$$

While this function looks like a rather random staircase when considering a small range of x (see Figure 1), the edges and randomness smooth out and a smooth curve seems to emerge when zooming out to a larger range of x (see Figure 2).

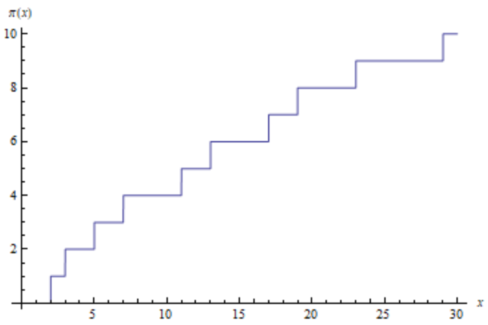

Figure 1: The prime-counting function for a small range of x.

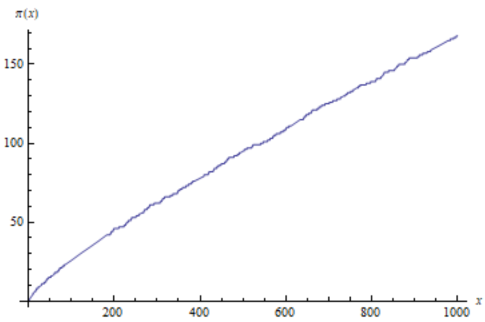

Figure 2: The prime-counting function for a larger range of x.

In the late 18th and early 19th century[2], this led Legendre and Gauss to independently conjecture that this curve is 'essentially' $\frac{x}{\log(x)}$, in the sense that the following limit holds:

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

Dirichlet soon published his own conjectured estimate, namely that

$$\lim_{x \to \infty} \frac{\pi(x)}{\mathrm{li}(x)} = 1,$$

where $\mathrm{li}(x)$, the logarithmic integral, is defined as

$$\mathrm{li}(x) := \int_2^x \frac{dt}{\log(t)}$$

(see also Definition 3.5). The logarithmic integral (which itself is 'essentially' $\frac{x}{\log(x)}$) turned out to be a 'better' estimate of $\pi(x)$, in the sense that the latter limit converges faster to 1 than the former.

These conjectured limits were proved in 1896 by Hadamard in [8] and de la Vallée-Poussin in [18] and this result is known as the Prime Number Theorem. For their proofs they drew upon ideas by Bernhard Riemann, who in 1859 published an explicit formula for $\pi(x)$ (see Section 3.3 and [15]), remarkably

[2]The historical context leading up to Riemann's 1859 paper can be found in for example H.M. Edwards' 1974 book "Riemann's Zeta function"

writing the number theoretic function $\pi(x)$ as a sum of terms derived from the locations of the zeros and pole of a complex analytic function: the Riemann zeta function (see Definition 3.10).

In the same paper, Riemann determined the location of the pole and the so-called trivial zeros, determined that the remaining nontrivial zeros are symmetric about the line $1/2+it$, and conjectured that they all lie exactly on this line. This is the famous Riemann hypothesis, which was shown by von Koch in [10] to be equivalent to the statement that

$$\pi(x) = \mathrm{li}(x) + O\left(\sqrt{x}\log(x)\right).$$

As the power of x in the big-oh term is equal to the supremum of the real parts of all nontrivial zeros, this error bound is the best possible bound of this approximation to the prime counting function, given the symmetry of the zeros about $1/2+it$.

In the more than 150 years since then, nobody has been able to prove or disprove the Riemann hypothesis, despite many results in number theory dependent on this error estimate (for example the growth of many number theoretic functions) and it being part of Hilbert's 1900 list of 23 important mathematical problems for the 20th century and the Clay Mathematics Institute's 2000 list of 10 Millennium Prize Problems.

In 2023 Olsen published a new proof of the Prime Number Theorem that uses an approach based on operator theory (see [12]). This in turn inspired this thesis, which aims to apply this approach to the Riemann Hypothesis in the hope of enabling further progress in the future.

The structure of the thesis is as follows. First, a brief introduction of relevant concepts and results of operator theory and number theory is given in Sections 2 and 3 respectively. Then, Olsen's approach to the Prime Number Theorem is explained and discussed in Section 4. Moreover, in this section this approach is applied to give new, operator theoretic, proofs of a different version of the Prime Number Theorem (Corollary 4.10 in Section 4.1) and the Prime Number Theorem for arithmetic progressions (Theorem 4.14 in Section 4.2). Also, a slight generalization of this approach in section 4.3 leads to an investigation of the approximation error in Section 5, where also the main results are proved: Theorems 5.3 and 5.4 in Section 5.1 give operator theoretic equivalents of the Riemann Hypothesis. Finally, in Section 5.2 some further properties of the operators involved in the main results are investigated.

2 Operators

Before we define the operators we use in this thesis we first briefly discuss the vector spaces on which they act. The thesis uses only Hilbert spaces and we assume that the reader is already familiar with them, especially the space $L^2(\mathbb{R})$. The spaces are all assumed to be separable (i.e., they have a countable basis).

We mostly focus on the $L^2(S)$ spaces (where S is usually some interval I):

Definition 2.1. For a given $S \subseteq \mathbb{R}$ we define $L^2(S)$ as the subspace of $L^2(\mathbb{R})$ of functions supported only on S.

Note that the definition of $L^2(S)$ differs slightly from most textbooks (where it denotes the space of all functions that are square-integrable on S, regardless of their support) and is chosen here as it is more convenient for the purposes of the thesis.

It is assumed in the next section and in 2.2 that the reader is not yet familiar with basic concepts and results from operator theory. A reader already familiar with these subjects can skip ahead to section 3.

2.1 Definition and basic properties of operators

The main objects in this thesis are linear operators. The qualifier 'linear' is generally omitted, as all operators in this thesis are linear.

Definition 2.2. A **(linear) operator** $A : X \to Y$ is a mapping between two vector spaces that preserves both vector operations (addition and scalar multiplication).
We call the two vector spaces respectively the **domain** and **codomain** of the operator, i.e.: an operator A is from the domain X to the codomain Y. If the domain and codomain are the same, we say that the operator is on the (co)domain.

We can view operators as matrices, since we can construct a (possibly infinite) matrix from an operator after choosing a basis for its domain. An important example is a diagonalizable operator:

Definition 2.3. A **diagonalizable operator** is an operator whose matrix representation in a certain basis is a diagonal matrix.

The operators in this thesis are in general bounded:

Definition 2.4. Let A be an operator from a normed space X to a normed space Y. If it is the case that $||Ax||_Y \leq c||x||_X$ for some $c > 0$ and all $x \in X$, then we say that the operator is **bounded**. The smallest such c is called the **operator norm** of A, which we denote by $||A||$.

Bounded operators are usually easier to work with than unbounded operators, as they more closely resemble operators on finite-dimensional spaces (which are also bounded) and we have a lot of results at our disposal to study them.

For example, a useful property of operators on complex finite-dimensional spaces is that they each have a unique conjugate transpose. The Riesz representation theorem (see e.g. Theorem 6.4 in [11]) allows us to generalize this notion to bounded operators between all Hilbert spaces. Only now this operator is called the adjoint, but it is also guaranteed to exist, to be unique, and to be bounded:

Definition 2.5. Let A be a bounded operator from a Hilbert space X to a Hilbert space Y. The bounded operator A^* from Y to X such that $\langle Ax, y\rangle_Y = \langle x, A^*y\rangle_X$ for all $x \in X$ and $y \in Y$ is called the **adjoint** of A.

Definition 2.6. Let A be a bounded operator on a Hilbert space X. If $AA^* = A^*A$, then the operator is called **normal** and if $A^* = A$, then it is called **self-adjoint** (or **Hermitian**).

Some operators in this thesis are compact:

Definition 2.7. An **open cover** of a subset S of a topological space is a family of open sets whose union contains S. A subset S of a topological space is called **compact** if every open cover has a finite subcover.

Definition 2.8. Let A be an operator from a normed space X to a normed space Y. If the image of every bounded subset of X has compact closure in Y, then it is called **compact**. Equivalently, A is compact if for every bounded sequence $\{x_n\}$ of vectors in X the sequence $\{Ax_n\}$ has a Cauchy subsequence.

It is important to know if an operator is compact since we have even more results available to study them than we do for bounded operators. This is because compact operators are always bounded:

Proposition 2.9. A compact operator from a normed space X to a normed space Y is bounded.

Proof. Let $A : X \to Y$ be a compact operator and let $C \subset Y$ be the closure of the image of the unit ball under A. For C we can find an open cover consisting of B_r: open balls of radius $r > 0$ centered at the origin, as they form an open cover of all of Y. Since this open cover of C must have a finite subcover of C, there is a finite M such that B_M covers C. From this we get that $||Ax|| \leq M||x||$ for all $x \in X$. ☐

Interestingly, the identity operator on X, which we denote by Id to prevent confusion with an interval I (i.e.: $\text{Id}x = x$ for all $x \in X$), is compact if and only if X is finite-dimensional (Theorem 5.6 in [11]), so this immediately gives an example of a bounded operator that is not compact. A good example of a useful result in the study of compact operators, which we will use in the next section, is the following: if A is a compact normal operator, then it is diagonalizable (Theorem 3.3.8 in [14]).

Lastly, we mention that an operator can also be bounded below, but generally this is not the case if it is compact:

Definition 2.10. Let A be an operator from a normed space X to a normed space Y. If it is the case that $||Ax||_Y \geq c||x||_X$ for some $c > 0$ and all $x \in X$, then we say that the operator is **bounded below**.

Proposition 2.11. A compact operator from an infinite-dimensional normed space X to a normed space Y is not bounded below.

Proof. We prove the contrapositive and let $A : X \to Y$ be an operator bounded from below. Since X is infinite-dimensional there exists a sequence of unit vectors $\{x_n\}$ without a Cauchy subsequence (e.g., such that $||x_i - x_j||_X > \frac{1}{2}$ whenever $i \neq j$). As $||Ax_i - Ax_j||_Y \geq c||x_i - x_j||_X$ the image of this sequence does not have a Cauchy subsequence in Y either, which means that A is not compact. ☐

2.2 The spectrum of a bounded operator

Definition 2.12. Let A be a bounded operator on a Banach space X. A complex number z is in $\rho(A)$, the **resolvent set** of A, if $A - z\text{Id}$ is a bijection. Conversely, a complex number z is in $\sigma(A)$, the **spectrum** of A, if $A - z\text{Id}$ is not a bijection.

Let (λ, x) be an eigenvalue-eigenvector pair of A, i.e.: $Ax = \lambda x$. We have that the operator $A - \lambda\text{Id}$ maps a nonzero x to 0, which means that this operator is not injective. Conversely, if $A - \lambda\text{Id}$ is not injective for some λ, then there is at least one pair of distinct vectors (x_1, x_2) that is mapped to the same vector, but this means that $x = x_1 - x_2 \neq 0$ is mapped to 0, so (λ, x) is an eigenvalue-eigenvector pair. Therefore, the set of all λ such that $A - \lambda\text{Id}$ is not injective is exactly the set of eigenvalues of A.

So every eigenvalue of A is part of $\sigma(A)$ and if X is finite-dimensional then an operator is injective if and only if it is surjective. Thus, the eigenvalues are all of $\sigma(A)$ then. However, if X is infinite-dimensional, then $A - \lambda\text{Id}$ can be injective, but not surjective, and we partition $\sigma(A)$ as follows:

Definition 2.13. Let A be a bounded operator on a Banach space and $\lambda \in \sigma(A)$.

- $\lambda \in \sigma_p(A)$, the **point spectrum** of A, if $A - \lambda\text{Id}$ is not injective
- $\lambda \in \sigma_c(A)$, the **continuous spectrum** of A, if $A - \lambda\text{Id}$ is injective, but its range is a dense proper subset of X (so it is 'almost' surjective)
- $\lambda \in \sigma_r(A)$, the **residual spectrum** of A, if $A - \lambda\text{Id}$ is injective, but its range is not dense in X (so it is 'far from' surjective)

In this way, the spectrum of A serves as a generalization of eigenvalues in a finite-dimensional context.

To see an example of this, suppose we have a bounded diagonalizable operator M on a Hilbert space with the sequence (m_n) on the diagonal in a matrix representation. The eigenvectors of M are precisely the diagonalizing basis vectors e_n ($Me_n = m_n e_n$), which means that $\sigma_p(M) = \{m_n\}$. It could be that $\sigma_c(M)$ is empty, e.g., if $M = \text{Id}$, then $M - \lambda\text{Id} = (1 - \lambda)\text{Id}$, which is surjective whenever it is injective (i.e., when $\lambda \neq 1$). An example where $\sigma_c(M)$ for a diagonalizable operator M is not empty will be shown later in this section. However, $\sigma_r(M)$ can only be empty:

Proposition 2.14. If M is a diagonalizable operator on a Hilbert space X, then $\sigma_r(M) = \emptyset$.

Proof. Let M be a diagonalizable operator on a Hilbert space and $\{e_n\}$ the diagonalizing basis. Notice that $R_\lambda = M - \lambda\text{Id}$ is also a diagonalizable operator (with the same diagonalizing basis) and denote with $(r_{\lambda,n})$ the sequence on the diagonal in the matrix representation of R_λ in this basis. Now suppose that R_λ is injective. This means that the e_n cannot be mapped to 0, so all $r_{\lambda,n}$ must be nonzero. Since $\frac{1}{r_{\lambda,n}} R_\lambda e_n = e_n$, we have that all e_n are in the range of R_λ, which means that their span is also in the range of R_λ. Thus, the range of R_λ is dense in X and $\sigma_r(M) = \emptyset$. □

Definition 2.15. Let A be a bounded operator on a Banach space and $\lambda \in \sigma(A)$. The **spectral radius**, $r(A)$, is the supremum of all $|\lambda|$.

In general, the spectrum of a bounded operator A on a Banach space is a nonempty compact subset of $\mathbb{C}$ and for all $\lambda \in \sigma(A)$ we have that $|\lambda| \leq ||A||$ (Theorem 17.4 in [11]). Thus, $r(A) \leq ||A||$.

If we know that A is a compact operator on a Banach space X, then the spectral theorem for compact operators (Theorem 21.6 in [11]) gives a relatively simple description of $\sigma(A)$. It is a countable set, which can have either no limit point or 0 as its only limit point, and if a $\lambda \in \sigma(A)$ is nonzero it must be in $\sigma_p(A)$. If $0 \in \sigma(A)$, which has to be the case if X is infinite-dimensional, it could be in either of the 3 partitions.

Using this description we can construct an example where $\sigma_c(M)$ for a diagonalizable operator M is not empty: a compact M on an infinite-dimensional Hilbert space with no zeros on the diagonal. This operator will have $0 \in \sigma(M)$, but $0 \notin \sigma_p(M)$. As M is diagonalizable, we have that $\sigma_r(M) = \emptyset$ by proposition 2.14, so it must be that $0 \in \sigma_c(M)$.

Depending on what else we know about an operator A on a Hilbert space we get further information about its spectrum from other spectral theorems. If A is a compact normal operator, then it is diagonalizable (Theorem 3.3.8 in [14]) and thus $\sigma_r(A) = \emptyset$ by Proposition 2.14. If A is a self-adjoint operator, then its spectrum consists only of real numbers (Theorem 32.5 in [11]). Finally, while the point spectrum can be empty in the case of a compact operator and also in the case of a self-adjoint operator, if A is both compact and self-adjoint, then it can be shown that at least one of $||A||$ and $-||A||$ is an eigenvalue of A (Theorem 2 in [2]).

3 The Prime Number Theorem and the Riemann Hypothesis

3.1 Asymptotic notations

While it is assumed that the reader is already familiar with the asymptotic notations $O(g(x))$ and $o(g(x))$ for some given function g, their definitions are given below so they can be compared with the definitions of two less commonly used asymptotic notations that are also employed in this thesis. Note that all asymptotic notations in the thesis are used exclusively to describe a given $f(x)$ as x tends to infinity.

Definition 3.1. Let f and g be real functions. A function $f(x)$ is **little-oh** of $g(x)$, i.e., $f(x) = o(g(x))$, if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$.

Definition 3.2. Let f and g be real functions, with $g(x) > 0$ for all $x \geq x_0$ for some $x_0 \in \mathbb{R}$. A function $f(x)$ is **big-oh** of $g(x)$, i.e., $f(x) = O(g(x))$, if there exists a positive constant C and a real constant x_0 such that $|f(x)| \leq Cg(x)$ for all $x \geq x_0$ or, equivalently, if $\limsup_{x\to\infty} \frac{|f(x)|}{g(x)} < \infty$.

We can see these notations as the asymptotic equivalents of '$<$' and '$\leq$' respectively. The first less commonly used asymptotic notation we introduce can be seen as the asymptotic equivalent of '$=$':

Definition 3.3. Let f and g be real functions. A function $f(x)$ is **on the order of** $g(x)$ or **asymptotically equal** to $g(x)$, i.e., $f(x) \sim g(x)$, if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$

The usage of a tilde here is not accidental, as this indeed defines an equivalence relation. This notation indicates that f grows or decays at approximately the same rate as g. The function g is in some sense a 'good' approximation of f. If we write $f(x) = g(x) + E(x)$, where $E(x)$ collects all the approximation errors, then $\frac{f(x)}{g(x)} = 1 + \frac{E(x)}{g(x)}$. Thus, $f(x) \sim g(x)$ implies $E(x) = o(g(x))$ (and similarly $E(x) = o(f(x))$). So, this is a 'good' approximation in the sense that the relative error decays.

The second less commonly used asymptotic notation we introduce is the negation of little-oh:

Definition 3.4. Let f and g be real functions. A function $f(x)$ is **big Omega** of $g(x)$, i.e., $f(x) = \Omega(g(x))$, if $\limsup_{x\to\infty} \frac{|f(x)|}{g(x)} > 0$

If we look at the very simple example $-5x^2+x+10 = O(x^2)$, then it is also the case that $-5x^2+x+10 = \Omega(x^2)$. In a way, this means that we cannot 'improve' our big-oh estimate. We call such an estimate a **sharp estimate**.

3.2 The Prime Number Theorem

Definition 3.5. The **logarithmic integral** is the function defined by

$$\mathrm{li}(x) := \int_2^x \frac{dt}{\log(t)}.$$

Using partial integration we see that this function can be approximated by $x/\log(x)$, in which case we make an error of $O(x/\log(x)^2)$, which is sharp. Note that this implies that $\mathrm{li}(x) \sim x/\log(x)$, but this last statement only tells us that the relative error decays to 0, not how fast. From the error estimate we can see that this decay is very slow, as it has the sharp estimate $O(1/\log(x))$.

Definition 3.6. For every $K \subseteq \mathbb{N}$ we can define a **counting function**, which counts the numbers in K less than or equal to x:

$$\pi_K(x) = \sum_{k\leq x, k\in K} 1.$$

Special cases are $\pi_{\mathbb{N}}(x) = \lfloor x \rfloor$ and the **prime-counting function**, which we simply write as $\pi(x)$.

It had long been an important goal of mathematicians to find a 'good' approximation of the prime-counting function, i.e., $\pi(x) \sim g(x)$ for some simpler function $g(x)$. Initially it was conjectured that $\pi(x) \sim \frac{x}{\log(x)}$, but the equivalent $\pi(x) \sim \operatorname{li}(x)$ is more accurate and theoretically more interesting (as discussed in the next section). This is considered of such importance that the eventual theorem that proved these conjectures is called the Prime Number Theorem, which we state as follows:

Theorem 3.7. *The Prime Number Theorem (Hadamard in [8] and de la Vallée-Poussin in [18])*

$$\pi(x) \sim \operatorname{li}(x).$$

The convergence of both approximations is demonstrated below:

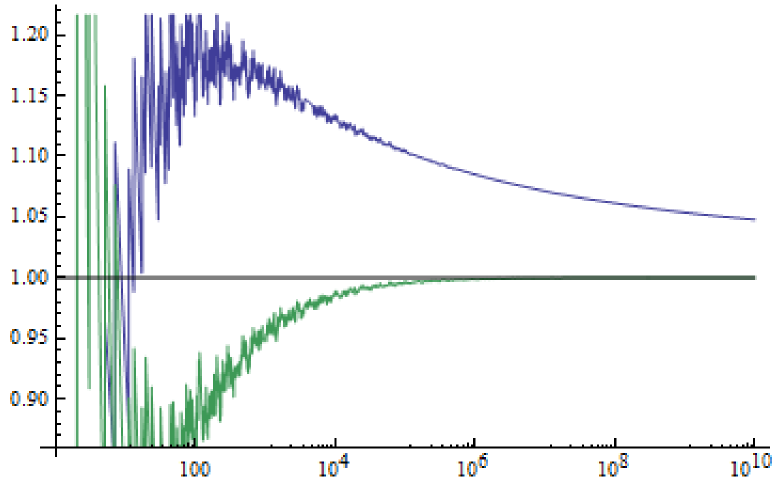

Figure 3: The ratio of the prime-counting function to two of its approximations: $\pi(x)/\frac{x}{\log(x)}$ (blue) and $\pi(x)/\operatorname{li}(x)$ (green).

Instead of working directly with the prime-counting function, it is common to work via the Chebyshev functions:

Definition 3.8. The **first Chebyshev function** is defined as

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

where p is prime.

Definition 3.9. The **second Chebyshev function** is defined as

$$\psi(x) = \sum_{k \in \mathbb{N}} \sum_{p^k \leq x} \log p,$$

where p is prime.

These functions are closely related to the prime-counting function, but it tends to be easier to work with them. Chebyshev showed in [3] that the Prime Number Theorem is equivalent to both $\theta(x) \sim x$ and $\psi(x) \sim x$ and the latter equivalence was used by Hadamard and de la Vallée-Poussin in their proofs of the Prime Number Theorem.

3.3 The Riemann Hypothesis

The Riemann Hypothesis concerns the zeros of the Riemann zeta function:

Definition 3.10. Let $K \subseteq \mathbb{N}$. The **zeta function** for K, $\zeta_K(s)$, is a complex function defined to be the analytic continuation of the series

$$\sum_{k \in K} k^{-s}.$$

Special cases are the **prime zeta function**, $\zeta_P(s)$, where P denotes the set of primes, and the **Riemann zeta function**, $\zeta(s)$, for $K = \mathbb{N}$.

As is the historical convention, we use the notation $s = \sigma + it$ for complex numbers when discussing the Riemann zeta function. The infinite sum used to define this function converges in the half-plane $\sigma > 1$, but can be analytically continued to all of $\mathbb{C}$ except for a pole at $s = 1$, which was shown by Riemann in [15]. One of the main ingredients in his construction of the analytic continuation was the following functional equation he found:

$$\zeta(s) = 2^s \pi^{s-1} \sin\left(\frac{\pi s}{2}\right) \Gamma(1-s)\zeta(1-s).$$

Here, $\Gamma(s)$ is the gamma function, which is the analytic continuation of:

$$\Gamma(s) = \int_0^\infty t^{s-1} e^{-t} dt.$$

Due to the sine in the functional equation we immediately get that $\zeta(s) = 0$ for all negative even integers, as the other zeros of the sine are cancelled by the pole of $\zeta(1-s)$ at $s = 0$ and the poles of $\Gamma(1-s)$ at positive even integers. These zeros of $\zeta(s)$ are called the trivial zeros. Riemann also showed in the same paper that the remaining, nontrivial, zeros must lie in the strip $0 \leq \sigma \leq 1$, that they are symmetric about $\sigma = 1/2$, and that there are infinitely many of them.

One of Riemann's most remarkable achievements is finding an explicit formula for $\pi(x)$ (also in [15]), based on the pole and the zeros of the Riemann zeta function (trivial or not). To understand this formula we first need to define the Möbius function:

Definition 3.11. The Möbius function is the function defined for all $n \in \mathbb{N}$ by

$$\mu(n) = \begin{cases} 0 & \text{if } n \text{ has a squared prime factor} \\ (-1)^k & \text{if } n \text{ is a product of } k \text{ distinct primes.} \end{cases}$$

We then define the following function:

$$R(x) = \sum_{n \in \mathbb{N}} \frac{\mu(n)}{n} \operatorname{li}(x^{1/n}).$$

Riemann partially proved that

$$\pi(x) = R(x) - \sum_{\rho} R(x^\rho),$$

where $\sum_\rho$ sums over all zeros of the Riemann zeta function. The gaps in the proof were later filled by Hadamard and von Mangoldt. Note that this equality holds almost everywhere. As is typical for approximations of functions with jumps, the right-hand side converges to the average of the left and right limits at each jump of $\pi(x)$.

This remarkable result shows that essentially all information about the prime-counting function is contained in the location of the pole (which determines the $R(x)$ term) and the zeros of $\zeta(s)$. The Prime

Number Theorem ($\pi(x) \sim \text{li}(x)$) then tells us that, for large x, the first term of $R(x)$ gives the biggest contribution in this formula for $\pi(x)$.

This formula is demonstrated below:

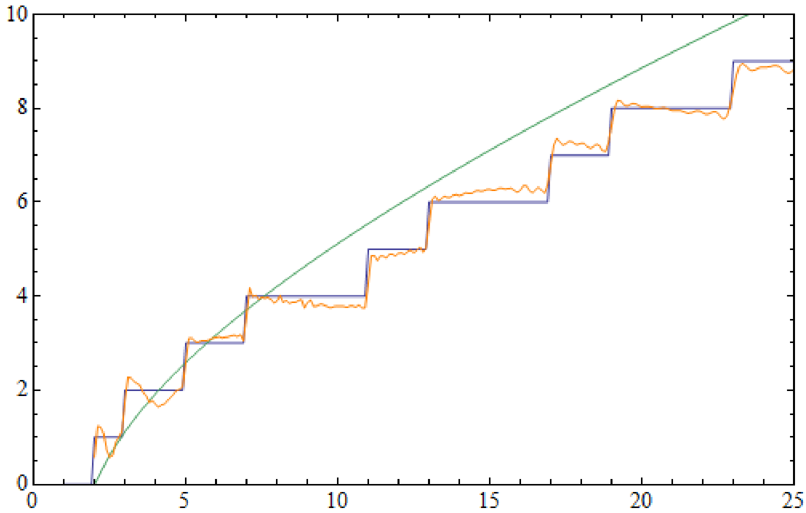

Figure 4: The prime-counting function (blue), li(x) (green), and Riemann's formula using all trivial zeros and the first 100 nontrivial zeros (orange).

Afterwards, both Hadamard and de la Vallée-Poussin independently proved that there are no zeros on the line $\sigma = 1$ and then used that result to prove the Prime Number Theorem. A few years later de la Vallée-Poussin was also able to give a bound on $\pi(x) - \text{li}(x)$ (see [19]):

$$\pi(x) = \text{li}(x) + O\left(x \exp\left(-a\sqrt{\log(x)}\right)\right).$$

Here, a is some positive constant and the notation $\exp(f(x)) = e^{f(x)}$ is used. This bound has been improved somewhat over the years. Right now, the state of the art is by Ford (see [6]):

$$\pi(x) = \text{li}(x) + O\left(x \exp\left(-0.2098\frac{\log(x)^{3/5}}{\log(\log(x))^{1/5}}\right)\right).$$

These bounds are constructed by first determining a zero-free region in the strip $0 < \sigma < 1$. Ingham showed how to translate zero-free regions of a certain type to error bounds, should they turn out to be true (see [9]). Broadly speaking, a 'wider' zero-free region would result in a better error term. A consequence of this is that if one were to prove that $\zeta(s)$ has no zeros in the domain $\sigma > 1 - b$, for some $0 < b \leq \frac{1}{2}$, then the error bound would be greatly improved to:

$$\pi(x) = \text{li}(x) + O\left(x^{1-b}\log(x)\right).$$

In particular, if $\sigma = \frac{1}{2}$ for all nontrivial zeros, then we get the best possible version of this bound:

$$\pi(x) = \text{li}(x) + O\left(\sqrt{x}\log(x)\right).$$

The conjecture that $\sigma = \frac{1}{2}$ for all nontrivial zeros of $\zeta(s)$ was already made by Riemann in [15] and is known as the Riemann Hypothesis. The Riemann Hypothesis has been verified empirically for many nontrivial zeros, but no proof (or counterexample) is known at the time of writing.

Interestingly, an explicit formula for the second Chebyshev function is also known, due to von Mangoldt:

$$\psi(x) = x - \sum_{\rho} \frac{x^{\rho}}{\rho} - \frac{\zeta'(0)}{\zeta(0)} = x - \sum_{\substack{\rho \\ 0<\mathrm{Re}\ \rho<1}} \frac{x^{\rho}}{\rho} - -\frac{1}{2}\log(1 - x^{-2}) - \log(2\pi),$$

where $\sum_{\rho}$ sums over all zeros of the Riemann zeta function. Again, the right-hand side converges to the value halfway between the values to the left and right of the jump discontinuities of $\psi(x)$, but is otherwise exact. This shows that the location of the pole and the zeros of $\zeta(s)$ also contain essentially all information about this function, with a slight correction from the logarithmic derivative evaluated at 0.

4 An operator theoretic approach to the Prime Number Theorem

We now turn to discuss the operator theoretic approach to the Prime Number Theorem that inspired this thesis. It is assumed that the reader is already familiar with the basics of Fourier Analysis and of the Laplace transform. To understand this operator theoretic approach we first need four additional definitions.

Definition 4.1. Let $I \subset \mathbb{R}$ be a bounded interval symmetric about the origin and $2I$ the same interval, but twice as wide. A **convolution operator** W on $L^2(I)$ is an operator of the form

$$Wf(t) = \chi_I(t) \int_I f(\tau) K(t-\tau) d\tau$$

for some function K integrable on $2I$ called the **kernel**.

Definition 4.2. A **net** is a generalization of the concept of a sequence. While the index set of a sequence is the natural numbers, the index set of a net can be uncountable, as long as it is a directed set: a nonempty set with a binary relation (denoted by '$\leq$') that is reflexive and transitive and has the property that for any a, b in the index set there exists a c in the index set such that both $a \leq c$ and $b \leq c$.

This generalization allows us to use subsets of the real numbers as index sets.

Definition 4.3. A net of vectors x_ϵ in a Hilbert space X with $(0, \infty)$ as index set **converges strongly** to a vector $x \in X$ if $\lim_{\epsilon \to 0^+} ||x_\epsilon - x|| = 0$. A net of operators W_ϵ on a Hilbert space X with $(0, \infty)$ as index set converges strongly to an operator W on X if $\lim_{\epsilon \to 0^+} ||W_\epsilon x_0 - W x_0|| = 0$ for all x_0 in X.

Definition 4.4. A net of vectors x_ϵ in a Hilbert space X with $(0, \infty)$ as index set **converges weakly** to a vector $x \in X$ if $\lim_{\epsilon \to 0^+} \langle x_\epsilon, x_0 \rangle = \langle x, x_0 \rangle$ for all x_0 in X. A net of operators W_ϵ on a Hilbert space X with $(0, \infty)$ as index set converges weakly to an operator W on X if $\lim_{\epsilon \to 0^+} \langle W_\epsilon x_1, x_2 \rangle = \langle W x_1, x_2 \rangle$ for all x_1, x_2 in X.

Weak convergence is called that because it is implied by strong convergence, but the other direction does not hold in general (see e.g. example 10.2 in [11]):

Proposition 4.5. Strong convergence of a net of vectors or operators implies weak convergence.

Proof. Let the net x_ϵ in a Hilbert space X with $(0, \infty)$ as index set converge to x strongly and x_0 be given. Then (using Cauchy-Schwarz)

$$\lim_{\epsilon \to 0+} |\langle x_\epsilon, x_0 \rangle - \langle x, x_0 \rangle| = \lim_{\epsilon \to 0+} |\langle x_\epsilon - x, x_0 \rangle| \leq \lim_{\epsilon \to 0+} ||x_\epsilon - x|| ||x_0|| = 0.$$

The proof for operators follows the same steps. □

The Prime Number Theorem was originally proved using techniques and results from complex analysis and number theory. In 2023, Olsen published a new proof that uses results and techniques from operator theory and Fourier analysis (see [12]). In this proof, a given function $S(x)$, which is non-decreasing on $[0, \infty)$ and $O(x)$, is considered. As $S(x) = O(x)$, the associated function $G(s) := \mathscr{L}(S(e^u))(s)$ (where $\mathscr{L}$ denotes the Laplace transform) exists for $\sigma > 1$. We can thus define the following net of convolution operators on $L^2(I)$ for positive ϵ (with $I \subset \mathbb{R}$ again a bounded interval symmetric about the origin):

$$W_{S,I,\epsilon} f(t) := \frac{\chi_I(t)}{\pi} \int_I f(\tau) \operatorname{Re}\ G(1 + \epsilon + i(t-\tau)) d\tau.$$

The key result of this new approach is

Theorem 4.6. *(Theorem 1.2 in [12])*

Let $S(x)$, non-decreasing on $[0,\infty)$ and $O(x)$, be given and the net of operators $W_{S,I,\epsilon}$ defined as above. Then

$$\lim_{\epsilon\to 0^+} W_{S,I,\epsilon} = A\mathrm{Id} + \Psi_{S,I}$$

weakly, for some constant $A \geq 0$ and compact operator $\Psi_{S,I}$ for all sufficiently large I if and only if

$$\lim_{u\to\infty} \frac{S(e^u)}{e^u} = A.$$

Or, in asymptotic notation: $S(x) = o(x)$ if $A = 0$ and $S(x) \sim Ax$ if $A > 0$. In other words, we can use this theorem to try to prove if the given function $S(x)$ can eventually be approximated 'well' by a line. Theorems like this one, where properties of an integral transform of a function S are used to prove asymptotic behavior of S, are called Tauberian theorems.

Note that this approach is not guaranteed to succeed. An example is $S(x) = x + \frac{x+1}{2}(\sin(\log(x+1)) + \cos(\log(x+1)))$. This function is non-decreasing on $[0,\infty)$, since its derivative is $1 + \cos(\log(x+1))$. However, the relative error, $\frac{x+1}{2x}(\sin(\log(x+1)) + \cos(\log(x+1)))$, does not converge.

To illustrate Theorem 4.6 and to also show how the Prime Number Theorem can be proved using this result, we set $S(x) = \pi(x)\log(x)$. This is indeed a non-decreasing function on $[0,\infty)$ that we expect to be approximated 'well' by x eventually.

We can use the following lemma to facilitate the calculation of the associated function $G(s)$. It gives a nice relation between the counting function and the zeta function of the same infinite set $K \subseteq \mathbb{N}$:

Lemma 4.7. *Let $K = (k_n)$ be an arbitrary infinite subset of $\mathbb{N}$, then*

$$\mathscr{L}(\pi_K(e^u))(s) = \frac{\zeta_K(s)}{s},$$

for $\sigma > 1$.

Proof. The following integral converges for $\sigma > 1$, as in the worst case we have $K = \mathbb{N}$ and $\pi_{\mathbb{N}}(e^u) \leq e^u$:

$$\begin{aligned}
\mathscr{L}(\pi_K(e^u))(s) &= \int_0^\infty \pi_K(e^u)e^{-su}du \\
&= \sum_{k_n\in K} \int_{\log k_n}^{\log k_{n+1}} ne^{-su}du \\
&= \sum_{k_n\in K} \frac{n}{s}(k_n^{-s} - k_{n+1}^{-s}) \\
&= \frac{1}{s}\sum_{k_n\in K} k_n^{-s} \\
&= \frac{\zeta_K(s)}{s}.
\end{aligned}$$

□

As a final ingredient for this proof we will use a lemma that will make it much easier to show that a certain class of convolution operators is compact:

Lemma 4.8. *If a convolution operator on $L^2(I)$ with I compact has a kernel that is continuous on I, then it is a compact operator (e.g., Theorem 22.3 in [11]).*

We are now in a position to prove the following version of the Prime Number Theorem:

Corollary 4.9. *(Corollary 1.3 in [12])*

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

Proof. After using Lemma 4.7 on $S(x) = \pi(x)\log(x)$ we encounter the prime zeta function, $\zeta_P(s)$. This function has a logarithmic singularity at $s = 1$, but is otherwise analytic in a neighborhood of $\sigma > 1$. That is, we can write $\frac{\zeta_P(s)}{s} = -\log(s-1) + \psi_P(s)$, where $\psi_P(s)$ is analytic in a neighborhood of $\sigma > 1$ (see e.g. (6) in [12]). We thus get that

$$\begin{aligned}
G(s) &= \mathscr{L}(S(e^u))(s) \\
&= \mathscr{L}(\pi(e^u)u)(s) \\
&= -\frac{d}{ds}\mathscr{L}(\pi(e^u))(s) \\
&= -\frac{d}{ds}\frac{\zeta_P(s)}{s} \\
&= \frac{d}{ds}\left(\log(s-1) - \psi_P(s)\right) \\
&= \frac{1}{s-1} - \psi_P'(s).
\end{aligned}$$

Plugging this in $W_{S,I,\epsilon}$ we get that

$$\begin{aligned}
W_{S,I,\epsilon}f(t) &= \frac{\chi_I(t)}{\pi}\int_I f(\tau)\,\mathrm{Re}\ G(1+\epsilon+i(t-\tau))d\tau \\
&= \frac{\chi_I(t)}{\pi}\int_I f(\tau)\,\mathrm{Re}\ \left(\frac{1}{\epsilon+i(t-\tau)} - \psi_P'(1+\epsilon+i(t-\tau))\right)d\tau \\
&= \frac{\chi_I(t)}{\pi}\int_I f(\tau)\frac{\epsilon}{\epsilon^2+(t-\tau)^2}d\tau - \frac{\chi_I(t)}{\pi}\int_I f(\tau)\,\mathrm{Re}\ \psi_P'(1+\epsilon+i(t-\tau))d\tau.
\end{aligned}$$

The left integral we recognize as the Poisson kernel on the upper half plane, which converges strongly to $f(t)$ (see e.g. Theorem 5.2.6 in [17] for a subset of functions that is dense in $L^2(\mathbb{R})$ by, e.g., Proposition 8.17 in [5]). The right integral converges strongly to an operator that is compact by Lemma 4.8. So $\lim_{\epsilon\to0^+} W_{S,I,\epsilon} = \mathrm{Id} + \Psi_{S,I}$ for some compact $\Psi_{S,I}$ strongly, hence also weakly, and thus by Theorem 4.6,

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

i.e.,

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

□

4.1 Operator theoretic proof of a different version of the Prime Number Theorem

Another example may be instructive and this allows us to also show a direct, operator theoretic, proof of a different version of the Prime Number Theorem:

Corollary 4.10 (The Prime Number Theorem for $\theta(x)$)**.**

$$\theta(x) \sim x.$$

Proof. Now the given function S is simply $S(x) = \theta(x)$, the first Chebyshev function, and we note that it is non-decreasing. Let p_n denote the nth prime number. As

$$\begin{aligned}
\theta(x) &:= \sum_{\substack{n\in\mathbb{N}\\ p_n\le x}} \log(p_n)\\
&\le \int_1^x \log(t)dt\\
&= x(\log(x)-1)+1,
\end{aligned}$$

we thus have that $\theta(e^u) \le e^u(u-1)+1$. Therefore, the following integral converges for $\sigma > 1$:

$$\begin{aligned}
G(s) &= \mathscr{L}\{\theta(e^u)\}(s)\\
&= \int_0^\infty \theta(e^u)e^{-su}du\\
&= \sum_{n\in\mathbb{N}} \int_{\log p_n}^{\log p_{n+1}} \sum_{j=1}^{n} \log(p_j)e^{-su}du\\
&= \sum_{n\in\mathbb{N}} \sum_{j=1}^{n} \log(p_j)\frac{1}{s}(p_n^{-s} - p_{n+1}^{-s})\\
&= \sum_{n\in\mathbb{N}} \frac{\log(p_n)}{s}p_n^{-s}\\
&= -\frac{1}{s}\zeta_P'(s)\\
&= \frac{1}{s}\frac{d}{ds}(\log(s-1) - \psi_P(s))\\
&= \frac{1}{s(s-1)} - \frac{\psi_P'(s)}{s}\\
&= \frac{1}{s-1} - \frac{1}{s} - \frac{\psi_P'(s)}{s}.
\end{aligned}$$

To simplify the computation we define $h(s) := \frac{1}{s} + \frac{\psi_P'(s)}{s}$, which is also analytic in a neighborhood of $\sigma > 1$. Plugging this in $W_{S,I,\epsilon}$ gives

$$\begin{aligned}
W_{S,I,\epsilon}f(t) &= \frac{\chi_I(t)}{\pi}\int_I f(\tau)\,\mathrm{Re}\;\, G(1+\epsilon+i(t-\tau))d\tau\\
&= \frac{\chi_I(t)}{\pi}\int_I f(\tau)\,\mathrm{Re}\;\left(\frac{1}{\epsilon+i(t-\tau)} - h(1+\epsilon+i(t-\tau))\right)d\tau\\
&= \frac{\chi_I(t)}{\pi}\int_I f(\tau)\frac{\epsilon}{\epsilon^2+(t-\tau)^2}d\tau - \frac{\chi_I(t)}{\pi}\int_I f(\tau)\,\mathrm{Re}\;\, h(1+\epsilon+i(t-\tau))d\tau.
\end{aligned}$$

Again we see the Poisson kernel on the upper half plane in the left integral. Also, the right integral again converges strongly to an operator that is compact by Lemma 4.8. Hence, $\theta(x) \sim x$. □

4.2 Operator theoretic proof of the Prime Number Theorem for arithmetic progressions

One final example of how this approach can be used to give short proofs of asymptotic estimates is a proof of the Prime Number Theorem for arithmetic progressions.

Definition 4.11. An **arithmetic progression** is a sequence of numbers with a constant difference between consecutive elements. If we let a denote the starting element and d the common difference, then the sequence is of the form $(a, a+d, a+2d, \dots)$ (i.e., they are all congruent to a modulo d).

Definition 4.12. Euler's totient function, $\phi(n)$, is the function that counts the number of positive integers up to the given positive integer n that are coprime to n. For example, $\phi(10) = 4$, as 1, 3, 7, and 9 are coprime to 10.

Definition 4.13. Let A be a subset of the set of primes P. If the limit

$$\lim_{s \to 1+} \frac{\sum_{p \in A} p^{-s}}{\sum_{p \in P} p^{-s}}$$

exists, then this number is called the **Dirichlet density** of A.

Theorem 4.14. *The Prime Number Theorem for arithmetic progressions (de la Vallée-Poussin in [18])*

Let $\pi_{d,a}(x)$ be the counting function of all primes less than x that are also part of the arithmetic progression defined by a and d. If d and a are coprime, then

$$\pi_{d,a}(x) \sim \frac{\mathrm{li}(x)}{\phi(d)}.$$

Proof. The above statement is equivalent to

$$\pi_{d,a}(x) \sim \frac{x}{\phi(d)\log(x)}.$$

So we set $S(x) = \phi(d)\pi_{d,a}(x)\log(x)$, which is again non-decreasing. By Dirichlet's theorem on arithmetic progressions there are infinitely many primes congruent to a modulo d if a and d are coprime (similar to and implying Euclid's theorem that there are infinitely many primes). If we let K denote this set of primes congruent to a modulo d we thus get by Lemma 4.7 that

$$\mathscr{L}(\pi_{d,a}(e^u))(s) = \frac{\zeta_K(s)}{s}.$$

We saw in the proof of Corollary 4.9 that the prime zeta function has a logarithmic singularity at $s = 1$ and is otherwise analytic in a neighborhood of $\sigma > 1$. The zeta functions we encounter here turn out to behave almost the same, except that we need to account for the Dirichlet density of K when approaching the logarithmic pole. This is proved in Theorem VI.2 in [16], where we get that in this case the Dirichlet density is $1/\phi(d)$ and that

$$\frac{\zeta_K(s)}{s} = -\frac{1}{\phi(d)}\log(s-1) + \psi_K(s),$$

where $\psi_K(s)$ is analytic in a neighborhood of $\sigma > 1$. Therefore, the associated function is

$$
\begin{aligned}
G(s) &= \mathscr{L}(\phi(d)\pi_{d,a}(e^u)u)(s) \\
&= -\phi(d)\frac{d}{ds}\frac{\zeta_K(s)}{s} \\
&= \frac{1}{s-1} - \phi(d)\psi_K'(s),
\end{aligned}
$$

and we can follow the same steps as in Corollary 4.9 to conclude the proof. □

Of course, we get back the 'ordinary' Prime Number Theorem if we set $d = 1$.

4.3 Examining and generalizing this approach

To study how this approach works and slightly generalize it at the same time, we set, for some nonzero $p \in \mathbb{R}$,

$$S(x) = Ax^p + E(x),$$

where $E(x) = O(x^p)$ captures the errors we make when approximating some non-decreasing (non-increasing if $p < 0$) $S(x)$ on $[0, \infty)$ with Ax^p. We then plug this in to the following net of convolution operators:

$$W_{S,I,\epsilon,p}f(t) = \frac{\chi_I(t)}{\pi}\int_I f(\tau)\,\mathrm{Re}\ G(p+\epsilon+i(t-\tau))d\tau.$$

Now the associated function G becomes (the integral converges for $\sigma > p$ as $S(x) = O(x^p)$):

$$
\begin{aligned}
G(s) &= \mathscr{L}\{Ae^{pu} + E(e^u)\}(s) \\
&= A\int_0^\infty e^{(p-s)u}du + \mathscr{L}\{E(e^u)\}(s) \\
&= \frac{A}{s-p} + \mathscr{L}\{E(e^u)\}(s).
\end{aligned}
$$

If we plug this in to $W_{S,I,\epsilon,p}$, we get

$$
\begin{aligned}
W_{S,I,\epsilon,p}f(t) &= \frac{\chi_I(t)}{\pi}\int_I f(\tau)\,\mathrm{Re}\ G(p+\epsilon+i(t-\tau))d\tau \\
&= \frac{\chi_I(t)}{\pi}\int_I f(\tau)\,\mathrm{Re}\ \left(\frac{A}{\epsilon+i(t-\tau)} + \mathscr{L}\{E(e^u)\}(p+\epsilon+i(t-\tau))\right)d\tau \\
&= \frac{A\chi_I(t)}{\pi}\int_I f(\tau)\frac{\epsilon}{\epsilon^2+(t-\tau)^2}d\tau + \frac{\chi_I(t)}{\pi}\int_I f(\tau)\,\mathrm{Re}\ \left(\mathscr{L}\{E(e^u)\}(p+\epsilon+i(t-\tau))\right)d\tau \\
&= \frac{A\chi_I(t)}{\pi}\int_I f(\tau)\frac{\epsilon}{\epsilon^2+(t-\tau)^2}d\tau + \frac{\chi_I(t)}{\pi}\int_I f(\tau)\,\mathrm{Re}\ \left(\int_0^\infty E(e^u)e^{-(p+\epsilon)u}e^{-i(t-\tau)u}du\right)d\tau.
\end{aligned}
$$

Just like in the previous examples for $p = 1$ we see that the left net of operators converges strongly to $A\mathrm{Id}$. The right net can be rewritten as follows (using Re $z = \frac{z+\bar{z}}{2}$ and the fact that $E(x) = O(x^p)$ allows changing the order of integration using Fubini-Tonelli):

$$
\begin{aligned}
\Psi_{S,I,\epsilon,p} f(t) &= \frac{\chi_I(t)}{\pi} \int_I f(\tau) \operatorname{Re} \left(\int_0^\infty E(e^u) e^{-(p+\epsilon)u} e^{-i(t-\tau)u} du \right) d\tau \\
&= \frac{\chi_I(t)}{2\pi} \int_I f(\tau) \int_0^\infty E(e^u) e^{-(p+\epsilon)u} e^{-i(t-\tau)u} du d\tau + \frac{\chi_I(t)}{2\pi} \int_I f(\tau) \int_0^\infty E(e^u) e^{-(p+\epsilon)u} e^{i(t-\tau)u} du d\tau \\
&= \frac{\chi_I(t)}{2\pi} \int_I f(\tau) \int_{-\infty}^0 E(e^{-v}) e^{(p+\epsilon)v} e^{i(t-\tau)v} dv d\tau + \frac{\chi_I(t)}{2\pi} \int_I f(\tau) \int_0^\infty E(e^u) e^{-(p+\epsilon)u} e^{i(t-\tau)u} du d\tau \\
&= \frac{\chi_I(t)}{2\pi} \int_I f(\tau) \int_{-\infty}^0 E(e^{|v|}) e^{-(p+\epsilon)|v|} e^{i(t-\tau)v} dv d\tau + \frac{\chi_I(t)}{2\pi} \int_I f(\tau) \int_0^\infty E(e^{|u|}) e^{-(p+\epsilon)|u|} e^{i(t-\tau)u} du d\tau \\
&= \frac{\chi_I(t)}{2\pi} \int_I f(\tau) \int_{\mathbb{R}} E(e^{|u|}) e^{-(p+\epsilon)|u|} e^{i(t-\tau)u} du d\tau \\
&= \frac{\chi_I(t)}{2\pi} \int_{\mathbb{R}} E(e^{|u|}) e^{-(p+\epsilon)|u|} \int_I f(\tau) e^{iu(t-\tau)} d\tau du \\
&= \frac{\chi_I(t)}{2\pi} \int_{\mathbb{R}} E(e^{|u|}) e^{-(p+\epsilon)|u|} \int_{\mathbb{R}} f(\tau) e^{-iu\tau} d\tau e^{iut} du \\
&= \frac{\chi_I(t)}{2\pi} \int_{\mathbb{R}} E(e^{|u|}) e^{-(p+\epsilon)|u|} \mathscr{F}\{f\}(u) e^{iut} du.
\end{aligned}
$$

In the previous examples for $p = 1$ we always used strong convergence to show that these operators converge weakly. It turns out that this strong convergence is always the case:

Theorem 4.15. *Let $S(x)$, non-decreasing on $[0, \infty)$ and $O(x^p)$, be given and the net of operators $\Psi_{S,I,\epsilon,p}$ defined as above. Then $\Psi_{S,I,\epsilon,p}$ converges strongly to the bounded operator*

$$
\Psi_{S,I,p} f(t) = \frac{\chi_I(t)}{2\pi} \int_{\mathbb{R}} E(e^{|u|}) e^{-p|u|} \mathscr{F}\{f\}(u) e^{iut} du.
$$

Proof. As $E(x) = O(x^p)$ we have for some constant $C > 0$ that (using first Plancherel's theorem and the Fourier inversion theorem and then the dominated convergence theorem):

$$
\begin{aligned}
2\pi \lim_{\epsilon \to 0+} ||\Psi_{S,I,p} f - \Psi_{S,I,\epsilon,p} f|| &= \lim_{\epsilon \to 0+} || \int_{\mathbb{R}} E(e^{|u|}) e^{-p|u|} (1 - e^{-\epsilon|u|}) \mathscr{F}\{f\}(u) e^{iut} du|| \\
&= \lim_{\epsilon \to 0+} ||E(e^{|t|}) e^{-p|t|} (1 - e^{-\epsilon|t|}) \mathscr{F}\{f\}(t)|| \\
&\leq C \lim_{\epsilon \to 0+} ||(1 - e^{-\epsilon|t|}) \mathscr{F}\{f\}(t)|| \\
&= C|| \lim_{\epsilon \to 0+} (1 - e^{-\epsilon|t|}) \mathscr{F}\{f\}(t)|| \\
&= 0.
\end{aligned}
$$

Note that the application of the dominated convergence theorem is justified, as, applying Plancherel's theorem again, we have for all $\epsilon > 0$ that

$$
||(1 - e^{-\epsilon|t|}) \mathscr{F}\{f\}(t)|| \leq ||\mathscr{F}\{f\}(t)|| = ||f(t)|| < \infty,
$$

from which we also obtain that

$$
||\Psi_{S,I,p}|| \leq \frac{C}{2\pi}.
$$

□

Since the Fourier transform of a convolution is the product of the Fourier transforms we have that $E(e^{|u|})e^{-p|u|}$ is the Fourier transform of Re $(\mathscr{L}\{E(e^u)\}(p + i(t - \tau)))$. If we assume that $E(x)$ is also $o(x^p)$, i.e., that $E(e^{|u|})e^{-p|u|}$ decays, then $\Psi_{S,I,p}$ is a compact operator due to the following lemma:

Lemma 4.16. *(Lemma 2 in [13] with slightly stricter hypotheses for convenience)*

Let $I \subset \mathbb{R}$ be a bounded interval symmetric about the origin and $2I$ the same interval, but twice as wide. Also let $(K_\epsilon)_{\epsilon\in(0,1)}$ be a net of functions integrable on $2I$ that converges weakly to a K integrable on $2I$ whose Fourier transform decays to 0 at infinity.

Then the operator

$$\Psi f(t) = \lim_{\epsilon\to 0} \chi_I(t) \int_I f(\tau) K_\epsilon(t-\tau) d\tau$$

is a compact operator on $L^2(I)$.

This leads to a proof of a slight generalization of Theorem 4.6 we essentially get for free:

Corollary 4.17. *(Generalization of Theorem 2 in this thesis, which is Theorem 1.2 in [12])*

Let $S(x)$ be $O(x^p)$ for some nonzero $p \in \mathbb{R}$ and non-decreasing on $[0,\infty)$ (non-increasing for $p<0$) and let the net of operators $W_{S,I,\epsilon,p}$ be defined as

$$W_{S,I,\epsilon,p} f(t) := \frac{\chi_I(t)}{\pi} \int_I f(\tau) \operatorname{Re}\ G(p+\epsilon+i(t-\tau)) d\tau,$$

with $G(s) := \mathscr{L}(S(e^u))(s)$.

Then the following 3 conditions are equivalent:

$$\lim_{\epsilon\to 0^+} W_{S,I,\epsilon,p} = A\mathrm{Id} + \Psi_{S,I,p}$$

weakly, for some constant $A \geq 0$ and compact operator $\Psi_{S,I,p}$ for all sufficiently large I,

$$\lim_{\epsilon\to 0^+} W_{S,I,\epsilon,p} = A\mathrm{Id} + \Psi_{S,I,p}$$

strongly, for some constant $A \geq 0$ and compact operator $\Psi_{S,I,p}$ for all sufficiently large I, and

$$\lim_{u\to\infty} \frac{S(e^u)}{e^{pu}} = A.$$

Proof. We can follow the same steps as the proof in [12] and the equivalence between weak and strong convergence follows from Proposition 4.5 and Theorem 4.15.

□

To summarize, we see here how a monotone function $S(x)$, that is known to be $O(x^p)$, can be analyzed using this approach by giving operator theoretic conditions that are equivalent to the condition that the relative error decays. If one of these conditions holds, this approach shows that S is then either also $o(x^p)$ (if $A=0$) or $\Omega(x^p)$, and in the latter case it separates S into the part that is $\sim x^p$ and the part that is $o(x^p)$.

5 The approximation error

Corollary 4.17 may at first not seem very useful. Instead of directly testing if an $S(x)$ is approximated 'well' by x^p we could instead simply multiply $S(x)$ by a suitable power of x and use Theorem 4.6. However, since Corollary 4.17 works if and only if $E(x) = S(x) - Ax^p = o(x^p)$, we immediately obtain:

Corollary 5.1. *$\Psi_{S,I,p}$ is a compact operator if and only if $E(x) = o(x^p)$.*

Therefore, if we consider $\Psi_{S,I,p}$ for lower and lower p we can investigate how fast the error grows (or even decays if $p < 0$) and the lowest p for which Corollary 5.1 works tells us how 'good' the approximation is. Essentially, we filter out the 'signal' x^p, so we can study what used to be the 'noise' as our new 'signal'.

We can probably even refine this method a bit by investigating if the image of $\Psi_{S,I,p}$ lies in C^k, with k a natural number or even a real number, if and only if $E(x) = o(x^p / \log(x)^k)$. Such a result would be similar to how Delange generalized the Wiener-Ikehara theorem in [4]. While some work in this direction was done, it was ultimately decided to be too far outside the scope of this thesis.

5.1 Application to the Riemann Hypothesis

We will now apply Corollary 5.1 to obtain an equivalent to the Riemann Hypothesis, but first we need to take care of a small detail:

Lemma 5.2. *Let $a \in \mathbb{R}$, then we have that $f(x) = o(x^{a+\epsilon})$ for all $\epsilon > 0$ if and only if $f(x) = O(x^{a+\epsilon})$ for all $\epsilon > 0$.*

Proof. From the definitions we immediately get for every $\epsilon_0 > 0$ that $f(x) = o(x^{a+\epsilon_0})$ implies $f(x) = O(x^{a+\epsilon_0})$, so that gives one direction.

For the other direction we assume $\limsup_{x\to\infty} |f(x)|x^{-a-\epsilon}$ is finite for all $\epsilon > 0$. We want to show that $\lim_{x\to\infty} f(x)x^{-a-\epsilon} = 0$ for all $\epsilon > 0$, so we fix an $\epsilon_0 > 0$. We know that $\limsup_{x\to\infty} |f(x)|x^{-a-\epsilon_0/2}$ is finite, so $\limsup_{x\to\infty} |f(x)|x^{-a-\epsilon_0}$ has to be 0. Since it holds for nonnegative functions $g(x)$ that

$$0 \le \liminf_{x\to\infty} g(x) \le \limsup_{x\to\infty} g(x),$$

we must also have that $\liminf_{x\to\infty} |f(x)|x^{-a-\epsilon_0} = 0$. Therefore,

$$\lim_{x\to\infty} |f(x)|x^{-a-\epsilon_0} = \lim_{x\to\infty} f(x)x^{-a-\epsilon_0} = 0.$$

Since ϵ_0 was arbitrary, this holds for all $\epsilon > 0$. □

With that taken care of, we can prove our first main result by setting $S(x) = \theta(x)$:

Theorem 5.3 (First operator theoretic equivalent of the Riemann Hypothesis)**.**

The operator on $L^2(I)$ defined by

$$\Psi_{S,I,p} f(t) := \frac{\chi_I(t)}{2\pi} \int_{\mathbb{R}} (\theta(e^{|u|})e^{-p|u|} - 1)\mathscr{F}\{f\}(u)e^{iut}du$$

is a compact operator for all $p > \frac{1}{2}$ if and only if the Riemann Hypothesis is true.

Proof. Von Koch showed that the Riemann Hypothesis is equivalent to (see section 7 in [10]):

$$\theta(x) = x + O(x^{\frac{1}{2}+\epsilon}) \quad \text{for all } \epsilon > 0.$$

By Lemma 5.2 this is equivalent to

$$E(x) = \theta(x) - x = o(x^{\frac{1}{2}+\epsilon}) \quad \text{for all } \epsilon > 0$$

and the result follows by Corollary 5.1. □

Von Koch showed in the same article that the Riemann Hypothesis is also equivalent to

$$\psi(x) = x + O(x^{\frac{1}{2}+\epsilon}) \quad \text{for all } \epsilon > 0.$$

Therefore, by setting $S(x) = \psi(x)$ and noting that $\theta(x) \sim \psi(x)$ we similarly obtain our second main result:

Theorem 5.4 (Second operator theoretic equivalent of the Riemann Hypothesis)**.**

The operator on $L^2(I)$ defined by

$$\Psi_{S,I,p} f(t) := \frac{\chi_I(t)}{2\pi} \int_{\mathbb{R}} (\psi(e^{|u|})e^{-p|u|} - 1)\mathscr{F}\{f\}(u)e^{iut}du$$

is a compact operator for all $p > \frac{1}{2}$ if and only if the Riemann Hypothesis is true.

These theorems allow the Riemann Hypothesis to be studied using methods and results in operator theory, combined with results from number theory of course. There are quite a few methods and results in operator theory dedicated to proving or disproving the compactness of operators, for example by trying to approach the operator in question with a net of operators that are more well-behaved or well-understood (maybe bounded finite-rank operators could help here). In the next section we discuss some further properties of the operators appearing in these theorems, which could give some initial lines of reasoning to investigate.

5.2 Further properties of $\Psi_{S,I,p}$

Given the connection to the Riemann Hypothesis and their potential usefulness in investigating other asymptotic estimates, we further investigate the operators $\Psi_{S,I,p}$.

Firstly, these operators are self-adjoint:

Theorem 5.5. *Let $S(x)$ be $O(x^p)$ for some nonzero $p \in \mathbb{R}$ and non-decreasing on $[0, \infty)$ (non-increasing for $p < 0$) and let the operator $\Psi_{S,I,p}$ be defined as*

$$\Psi_{S,I,p} f(t) := \frac{\chi_I(t)}{2\pi} \int_{\mathbb{R}} E(e^{|u|})e^{-p|u|}\mathscr{F}\{f\}(u)e^{iut}du,$$

with $E(x) = S(x) - Ax^p$ for some constant $A \geq 0$, then this operator is self-adjoint.

Proof. If we set

$$K_{S,p}(x) = \frac{1}{\pi} \int_0^{\infty} E(e^u)e^{-pu}\cos(xu)du$$

and working our way back from Theorem 4.15, then we note that $\Psi_{S,I,p}$ is a convolution over I with the kernel $K_{S,p}$:

$$\Psi_{S,I,p} f(t) = \chi_I(t) \int_I f(\tau)K_{S,p}(t - \tau)d\tau.$$

This form of $\Psi_{S,I,p}$ makes it easier to calculate the adjoint (which is of course also on $L^2(I)$):

$$\begin{aligned}\langle \Psi_{S,I,p}f, g\rangle &= \int_I \Psi_{S,I,p}f(t)g(t)dt \\ &= \int_I \chi_I(t) \int_I f(\tau)K_{S,p}(t-\tau)d\tau g(t)dt \\ &= \int_I f(\tau)\chi_I(\tau) \int_I g(t)K_{S,p}(t-\tau)dtd\tau \\ &= \langle f, \Psi^*_{S,I,p}g\rangle.\end{aligned}$$

Therefore,

$$\Psi^*_{S,I,p}g(t) = \chi_I(t) \int_I g(\tau)K(\tau - t)d\tau.$$

However, since $K_{S,p}$ is even, we have that $\Psi_{S,I,p}$ is self-adjoint. □

We thus immediately get from the results discussed in Section 2.2 that if $\Psi_{S,I,p}$ is compact then

- Its spectrum is real and countable
- It is diagonalizable and $\sigma_r(\Psi_{S,I,p}) = \emptyset$
- 0 is either in $\sigma_c(\Psi_{S,I,p})$ or is an eigenvalue
- At least one of $||\Psi_{S,I,p}||$ and $-||\Psi_{S,I,p}||$ is an eigenvalue
- The eigenvalues have 0 as the only possible limit point
- If an eigenvalue is nonzero, then it has finitely many linearly independent associated eigenvectors

As a compact operator $\Psi_{S,I,p}$ has at least one eigenvalue, it could be fruitful to study its eigenvectors. It may for example be possible to derive a commutation relation between $\Psi_{S,I,p}$ and the derivative operator and use that relation to study them (similar to, e.g., [1]). While some work in this direction was done, it was ultimately decided to be too far outside the scope of this thesis.

Moreover, this could be a very hard problem. If we set for example

$$S(x) = Ax^p + x^p\chi_{[1,e]}(x),$$

then

$$\Psi_{S,I,p}f(t) = \frac{\chi_I(t)}{\pi} \int_I f(\tau)\frac{\sin(t-\tau)}{t-\tau}d\tau.$$

In for example [7], the known results on the eigenvectors and eigenvalues of this operator are summarized. It turns out, that even in this simple case the eigenvectors are expressible in terms of so-called prolate spheroidal wave functions. Of course, this case falls slightly outside the scope of this thesis as this S violates the monotonicity assumption once, but it still gives a good indication of how difficult this problem could be.

We further note that, if $\Psi_{S,I,p}$ is compact, it is not bounded below (by Proposition 2.11) and not surjective either (a consequence of the open mapping principle, see e.g. Theorem 15.9 in [11]).

Lastly, if $\Psi_{S,I,p}$ is an integral operator with a smooth kernel, like we saw in the proofs of the various Prime Number Theorems, then it must be a so-called trace class operator (Theorem 30.13 in [11]), i.e.: an operator whose trace is finite and well-defined (just as in the finite-dimensional case). We can then even get a very nice formula for the trace:

Theorem 5.6. *Let $S(x)$ be $O(x^p)$ for some nonzero $p \in \mathbb{R}$ and non-decreasing on $[0, \infty)$ (non-increasing for $p < 0$) and let the operator $\Psi_{S,I,p}$ be defined as*

$$\Psi_{S,I,p} f(t) = \chi_I(t) \int_I f(\tau) K_{S,p}(t-\tau) d\tau,$$

with a smooth

$$K_{S,p}(x) = \frac{1}{\pi} \int_0^\infty E(e^u) e^{-pu} \cos(xu) du,$$

with $E(x) = S(x) - Ax^p$ for some constant $A \geq 0$, then this operator has the following trace:

$$\frac{|I|}{\pi} \int_0^\infty E(e^u) e^{-pu} du.$$

Proof. Let the set of e_n be the standard orthonormal exponential basis for $L^2(I)$. Following the same computations as in [12] we get a result similar to (8) in that article:

$$\langle \Psi_{S,I,p} e_n, e_n \rangle = \frac{|I|}{2\pi} \int_{\mathbb{R}} E(e^{|u|}) e^{-p|u|} \left(\frac{\sin(u|I|/2)}{u|I|/2 - \pi n} \right)^2 du.$$

And this gives us the formula for the trace (we can switch the order of summation and integration by Fubini-Tonelli as the sum is absolutely convergent since the trace is well-defined):

$$\begin{aligned}
\sum_{n \in \mathbb{Z}} \langle \Psi_{S,I,p} e_n, e_n \rangle &= \frac{|I|}{2\pi} \sum_{n \in \mathbb{Z}} \int_{\mathbb{R}} E(e^{|u|}) e^{-p|u|} \left(\frac{\sin(u|I|/2)}{u|I|/2 - \pi n} \right)^2 du \\
&= \frac{|I|}{2\pi^3} \int_{\mathbb{R}} E(e^{|u|}) e^{-p|u|} \sin(u|I|/2)^2 \sum_{n \in \mathbb{Z}} \frac{1}{(n - u|I|/2\pi)^2} du \\
&= \frac{|I|}{2\pi^3} \int_{\mathbb{R}} E(e^{|u|}) e^{-p|u|} \sin(u|I|/2)^2 \frac{\pi^2}{\sin(u|I|/2)^2} du \\
&= \frac{|I|}{\pi} \int_0^\infty E(e^u) e^{-pu} du.
\end{aligned}$$

□

So essentially, the trace is then equal to a multiple of the 'total' relative error and is also equal to $|I| K_{s,p}(0)$.

References

[1] Alexandru Aleman, Alfonso Montes-Rodríguez, and Andreea Sarafoleanu. "The Eigenfunctions of the Hilbert Matrix". In: *Constructive Approximation* 36 (2012), pp. 353–374.

[2] Jordan Bell. *The singular value decomposition of compact operators on Hilbert spaces*. 2014.

[3] P. L. Chebyshev. "Mémoire sur les nombres premiers". In: *J. de Math. Pures Appl.* 17 (1852), pp. 366–390.

[4] Hubert Delange. "Généralisation du théorème de ikehara". In: *Ann. Sci. Ec. Norm. Sup.* 71.3 (1954), pp. 213–242.

[5] Gerald B. Folland. *Real Analysis: Modern Techniques and Their Applications*. 1999.

[6] Kevin Ford. "Vinogradov's Integral and Bounds for the Riemann Zeta Function". In: *Proceedings of the London Mathematical Society* 85.3 (2002), pp. 565–633. DOI: 10.1112/S0024611502013655.

[7] W. H. J. Fuchs. "On the Eigenvalues of an Integral Equation Arising in the Theory of Band-Limited Signals". In: *Journal of Mathematical Analysis and Applications* 9 (1964), pp. 317–330.

[8] J.S. Hadamard. "Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques". In: *Bulletin de la Société Mathématique de France* 24 (1896), pp. 199–220.

[9] A.E. Ingham. *The Distribution of Prime Numbers*. 1932.

[10] N.F.H von Koch. "Sur la distribution des nombres premiers". In: *Acta Mathematica* 24 (1901), pp. 159–182.

[11] Peter D. Lax. *Functional Analysis*. John Wiley & Sons, 2002.

[12] Jan-Fredrik Olsen. "An Operator Theoretic Approach to the Prime Number Theorem". In: *Journal of Mathematical Physics, Analysis, Geometry* 19.1 (2023), pp. 1–6.

[13] Jan-Fredrik Olsen. "Modified zeta functions as kernels of integral operators". In: *Journal of Functional Analysis* 259.2 (2010), pp. 359–383. ISSN: 0022-1236. DOI: https://doi.org/10.1016/j.jfa.2010.04.009. URL: https://www.sciencedirect.com/science/article/pii/S0022123610001485.

[14] Gert K. Pedersen. *Analysis Now*. Springer New York, NY, 2012.

[15] Bernhard Riemann. "Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse". In: *Monatsberichte der Berliner Akademie* (Nov. 1859).

[16] J.-P. Serre. *A Course in Arithmetic*. 1973.

[17] Elias M. Stein and Rami Shakarchi. *Fourier Analysis: An Introduction*. Princeton University Press, 2003.

[18] C.J.E.G.N. de la Vallée Poussin. *Recherches analytiques sur la théorie des nombres premiers*. 1896.

[19] C.J.E.G.N. de la Vallée Poussin. "Sur la fonction $\zeta(s)$ de Riemann et le nombre des nombres premiers inférieurs a une limite donnée." In: *Mémoires couronnés de l'Académie de Belgique* 59 (1899), pp. 1–74.

Master's Theses in Mathematical Sciences 2023:E74
ISSN 1404-6342

LUNFMA-3143-2023

Mathematics
Centre for Mathematical Sciences
Lund University
Box 118, SE-221 00 Lund, Sweden

http://www.maths.lu.se/