

The Selberg Sieve and Prime Pairs

by

Martin Christopher Duarte

B.Sc., York University, 2001

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

The Faculty of Graduate Studies

Mathematics

The University Of British Columbia

April 21, 2005

© Martin Christopher Duarte, 2005

Abstract

This paper gives an exposition of the Selberg Sieve and its application to the sifting of prime pairs. In particular, Selberg's methods are applied to the sifting approach of Eratosthenes and to an alternate sifting approach presented by the author. The results obtained are then used to show that the sum of the reciprocals of the twin primes (and similarly for general prime pairs) is convergent or finite. This paper is intended to be mostly self-contained and will hopefully be readable to someone with little background in the area of Sieve Theory.

Contents

Abstract	ii
Contents	iii
Preface	v
Acknowledgements	vi
Dedication	vii
1 Introduction	1
2 Sieve Notations and Preliminaries	3
3 The Selberg Sieve	14
3.1 The Set-up	14
3.1.1 Selberg's Main Theorem	14
3.1.2 Bounding $\lambda(d)$	19
3.2 Estimates for $G(x)$ and $E(D, P)$	21
3.2.1 Using Rankin's Device	21
3.2.2 An Asymptotic Bound	25
3.2.3 The Error Term	33
4 Twin Primes	34
4.1 Sifting for Twin Primes	34
4.1.1 An example of the Sieve of Eratosthenes	34
4.1.2 An Alternate Sieve for Twin Primes	35
4.1.3 Proving the Alternate Sieve works	36
4.2 Applying Selberg's Results to Twin Primes	38
4.2.1 The Eratosthenes Approach	38
4.2.2 The Alternate Approach	39

4.3	Sum of the Reciprocals of Twin Primes	41
5	General Prime Pairs (G.P.P.)	45
5.1	Sifting General Prime Pairs	45
5.1.1	An example of the Sieve of Eratosthenes	45
5.1.2	The Alternate Sieve for General Prime Pairs	46
5.1.3	Proving the General Alternate Sieve works	48
5.2	Applying Selberg's Method	51
5.2.1	The Eratosthenes Approach	51
5.2.2	The Alternate Approach	53
5.3	Sum of the Reciprocals	55
6	Concluding Remarks	58
	Bibliography	60

Preface

I was in my senior year of high school when I first learned about twin primes and the famous Twin Prime Conjecture. Though I found the problem fascinating at the time, I had given it little thought and soon forgot about it. Then in the beginning of the second year of my studies at York University, I was once again introduced to the outstanding problem. I was once again fascinated by the fact that nobody could solve this problem over the course of approximately the last two thousand years. I asked a close friend of mine at the time if he would be interested to work on the problem with me. It soon became clear that he was not interested and I began to seriously think about the problem on my own.

I recall staying behind after one of my computer science lectures to ponder the mysteries behind twin primes. I began to notice some peculiarities among the twin primes and later would discover the alternate sifting approach to that of Eratosthenes, along with a quite cumbersome proof that it worked. I knew next to nothing about modular arithmetic at the time.

Over the last few years I made some progress toward cleaning up the proof and also learned some number theory and of the existence of Sieve Theory. I then met my supervisor, Greg Martin, at the University of British Columbia and he suggested that I learn more about Sieve Theory. It is thus that the thesis you are about to read came into being.

In particular, I have focused on the Selberg Sieve and much of the paper is dedicated to an exposition of it. I have then applied Selberg's results to the Eratosthenes approach for sifting twin primes and to my own approach as well. I have also discovered that the alternate approach works for general prime pairs as well. In fact, by making that discovery, I have come to a greater understanding of how the alternate sieve works. The proofs showing that it works for twin primes and general prime pairs have been greatly simplified making good use of modular arithmetic. Finally, I have also applied Selberg's results to both sifting approaches in the case of general prime pairs.

Acknowledgements

I would like to thank Greg Martin for all of his guidance with this thesis. Without his support and supervision, this paper would have been impossible. I would like to thank David Boyd for taking the time to be the second reader. Thanks to Lee Yupitun and Izabella Laba for providing me with the necessary information for ensuring that all requirements are met for submission of this thesis. Thanks to Michael Forbes for providing the necessary tools used to format the paper. I would also like to thank all my family, friends, peers, and especially Mariah Hamel and Abbas Moameni. Without their support, I could not have achieved this accomplishment.

To My Parents

This accomplishment was made possible by the sacrifices and hard work of my parents, Martin and Shira Duarte. I thank you both for your ever continuing love and support.

Chapter 1

Introduction

What is a Sieve /SIV/n.? The first definition that follows is provided by the Oxford Minidictionary, whereas the second is provided by the author.

1. Utensil with a wire mesh or gauze through which liquids or fine particles can pass. -v.t. put through a sieve.
2. A mathematical cancellation process by which a desired set of numbers/objects is yielded from a larger set of numbers/objects.

Eratosthenes, a mathematician of ancient Greece, is the first known person to have constructed a mathematical sieve (third century, B.C.). He started with a list of odd numbers and then proceeded to cancel 3^2 and every third number thereafter, then cancel 5^2 and every fifth number thereafter, and so forth. After going through this process, one is left with the number 1 and the prime numbers. Today, this process is generally referred to as the "Sieve of Eratosthenes" and makes use of the observation that a number, n , if not prime, has a prime factor less than or equal to \sqrt{n} . For example, one starts with a list of all integers from 1 to N , and then cancels the proper multiples of all the primes, p , between 1 and \sqrt{N} . The numbers yielded by this sieve are 1 and the primes between 1 and N .

In 1808, A.M. Legendre observed that sieves can be used to give useful information about the distribution of primes and provided a more analytical formulation of the concept. His work and the work of those that follow him, like Brun and Selberg, provide the necessary tools and notations of Sieve Theory.

The Selberg formulation and results can be used for a variety of problems where sieves provide useful information. One such problem is the so called "Twin Prime Problem". A Twin Prime is a pair of prime numbers that differ by 2; (5, 7) and (11, 13) are examples. They are usually sifted by simply picking them out after conducting the Sieve of Eratosthenes. While it is well known that prime numbers occur infinitely throughout the integers, it is still an open question as to whether or not twin primes occur infinitely

often. One may also wish to consider more general prime pairs, $(p, p + 2B)$. It is also unknown whether or not they occur infinitely often throughout the integers.

Selberg's method provides useful upper bounds for both problems: upper bounds on the number of twin primes less than X , and on the number of general prime pairs less than X . An alternate approach to the Sieve of Eratosthenes for sifting the twin primes, or general prime pairs, also leads to the same upper bounds. In all cases, one can deduce that the sum of the reciprocals of the twin primes is convergent (or possibly finite), and similarly for more general prime pairs.

Chapter 2

Sieve Notations and Preliminaries

In order to discuss the Selberg Sieve and applications of it, we first need to look at the notations and language of Sieve Theory. Several lemmas and identities that will be needed should also be mentioned.

The following are some identities that are relevant when discussing sieves; they can be found in many texts on Analytic Number Theory. Where p represents a prime and C is a constant,

$$\sum_{p \leq X} \frac{\log(p)}{p} = \log(X) + O(1) \quad (2.1)$$

$$\sum_{p \leq X} \frac{1}{p} = \log(\log(X)) + C + O\left(\frac{1}{\log(X)}\right); \quad (2.2)$$

$$\prod_{p < z} \left(1 - \frac{1}{p}\right)^{-1} = e^{\gamma} \log(z) + O(1) \quad (\gamma \text{ is Euler's constant}) \quad (2.3)$$

(2.3) is known as Mertens' product.

The Möbius function, μ , is important to Sieve Theory; we recall its definition:

$$\begin{aligned} \mu(1) &= 1 \\ \mu(p_1 p_2 \cdots p_r) &= (-1)^r \text{ if } p_1, \dots, p_r \text{ are distinct primes} \\ \mu(d) &= 0 \text{ if } p^2 \mid d \text{ for some prime } p, \end{aligned}$$

i.e., μ is the multiplicative function defined at prime powers, p^α , by

$$\mu(p) = -1, \quad \mu(p^\alpha) = 0 \text{ if } \alpha \geq 2.$$

We also recall a key identity involving the Möbius function:

$$\sum_{d \mid B} \mu(d) = \begin{cases} 1 & \text{if } B = 1 \\ 0 & \text{otherwise.} \end{cases} \quad (2.4)$$

A will denote a set of numbers we wish to sift. A typical example would be to let A be the set of natural numbers up to some bound, and then proceed to sift (choose via a cancellation process) this set for numbers not divisible by certain primes, i.e., cancel the numbers from A that are divisible by certain primes.

A_d will denote the set of numbers in A that are divisible by d .

P will denote a product of distinct primes, not necessarily containing all the primes up to some specified bound. However, in many cases,

$$P = \prod_{p < z} p . \quad (2.5)$$

P will usually be the product of primes we are sifting (cancelling) by.

In connection with P , we denote

$$P(z) = \prod_{\substack{p|P \\ p < z}} p . \quad (2.6)$$

In situations where no product P is implied or specified, we simply take $P(z)$ to be

$$P(z) = \prod_{p < z} p . \quad (2.7)$$

(a, b) will denote the g.c.d. of a and b .

Definition 2.1 We define $S(X) = \sum_{d|X} \mu(d)$. Also, abbreviate $S((a, P))$ to $S(a, P)$, so we have

$$S(a, P) = \sum_{d|(a, P)} \mu(d) = \begin{cases} 1 & \text{if } (a, P) = 1 \\ 0 & \text{otherwise.} \end{cases} \quad (2.8)$$

The second equality followed from identity (2.4). It is important to observe that the condition $(a, P) = 1$ says that a is not divisible by any "sifting" prime factor of P .

Definition 2.2 With $S(a, P)$ as in identity (2.8), we define

$$S(A, P) = \sum_{a \in A} S(a, P) . \quad (2.9)$$

This says that $S(A, P)$ is the number of a in A that are not divisible by any primes in the product P . Putting definitions (2.1) and (2.2) together, we get

$$S(A, P) = \sum_{a \in A} \sum_{\substack{d|a \\ d|P}} \mu(d) = \sum_{d|P} \mu(d) |A_d|, \quad (2.10)$$

which is known as “Legendre’s Identity”. The second equality followed by switching the order of summation.

It will be seen later that with the Selberg Sieve, identity (2.8) is replaced with an inequality where μ is replaced by another function, λ , where in some sense, it is easier to work with λ . Also, suitable estimates for $|A_d|$ are needed in order to attain a good calculation of Legendre’s Identity.

With the notations we have so far, conducting a sieve for twin primes can be thought of as a sieve on a set of numbers given by the polynomial expression: $f(n) = n(n+2)$. Likewise, for a more general prime pair, we would have $f(n) = n(n+2B)$. More generally, we take

$$A = \{f(n) : Y - X < n \leq Y\},$$

where $f(n)$ is a polynomial over the integers (for twin primes, $f(n) = n(n+2)$, and for general prime pairs, $f(n) = n(n+2B)$). Also, P is as in identity (2.5). To estimate $|A_d|$, we consider each residue class, mod d , separately. We get

$$|A_d| = \sum_{\substack{1 \leq l \leq d \\ f(l) \equiv 0 \pmod{d}}} \sum_{\substack{Y-X < n \leq X \\ n \equiv l \pmod{d}}} 1 = \sum_l \left(\frac{X}{d} + O(1) \right). \quad (2.11)$$

The inner sum over n was estimated by expressing n as $l + dm$, so m is confined to an interval of length X/d . Letting $\rho(d)$ represent the number of roots of the congruence $f(l) \equiv 0 \pmod{d}$, we then get

$$|A_d| = X \frac{\rho(d)}{d} + O(\rho(d)). \quad (2.12)$$

Using this in Legendre's Identity (2.10), we get

$$\begin{aligned}
 S(A, P) &= \sum_{d|P} \mu(d) |A_d| \\
 &= \sum_{d|P} \mu(d) \left(X \frac{\rho(d)}{d} + O(\rho(d)) \right) \\
 &= X \prod_{p|P} \left(1 - \frac{\rho(p)}{p} \right) + O \left(\sum_{d|P} \rho(d) \right). \quad (2.13)
 \end{aligned}$$

The function ρ just mentioned plays a very important role in Sieve Theory. It is related to another concept known as *sifting density*, or *dimension* of a sieve. The function ρ has a sifting density κ if a certain average of $\rho(p)$ at primes p is not greater than κ (and the sieve is said to have dimension κ). For twin primes, ρ has a sifting density of 2 since $f(n) = n(n+2) \equiv 0 \pmod{p}$ has 2 roots for primes $p \geq 3$ and 1 root for $p = 2$. Similarly, one can check that ρ has a sifting density of 2 for more general prime pairs, where $f(n) = n(n+2B)$. We will see later how the sifting density concept is defined for sieves in general. It also turns out that the following assumptions can be made about the function ρ :

- (1) $0 \leq \rho(p) < p$ when $p | P$
- (2) $\rho(p) = 0$ whenever $p \nmid P$.

Convention: $\rho(d) \neq 0$ only when $d | P$ means that the condition $d | P$ may be omitted from any sums involving $\rho(d)$.

Definition 2.3 When A_d , X , and $\rho(d)$ are specified, define the "remainder" $r_A(d)$ by

$$|A_d| = X \frac{\rho(d)}{d} + r_A(d) \quad \text{if } d | P, d \leq D. \quad (2.14)$$

Identity (2.14) will be applicable with ρ a multiplicative function and D to be specified. Apparently, in many sieve problems, including the applications to twin primes and general prime pairs,

$$|r_A(d)| \leq \rho(d). \quad (2.15)$$

Now recall we mentioned μ will be replaced by a different function λ . In fact, functions λ_D^+ , λ_D^- are constructed/chosen so that

$$\sum_{d|B} \lambda_D^-(d) \leq \sum_{d|B} \mu(d) \leq \sum_{d|B} \lambda_D^+(d) \text{ when } B \mid P. \quad (2.16)$$

The idea is to pick λ_D^+ , λ_D^- in a way that the inequalities are effective in a given sifting situation for relatively large products P . Also, D will have the property

$$\lambda_D^\pm(d) \neq 0 \Rightarrow d \leq D, \quad (2.17)$$

Definition 2.4 When inequalities (2.16) and (2.17) hold, we say that λ^+ , λ^- are (respectively) upper and lower sifting functions of level D for the product P .

Definition 2.5 Define the expression $V(P)$ by

$$V(P) = \sum_{d|P} \frac{\mu(d)\rho(d)}{d} = \prod_{p|P} \left(1 - \frac{\rho(p)}{p}\right). \quad (2.18)$$

V^+ and V^- are similarly defined where μ is replaced by λ^+ , λ^- respectively.

This along with identity (2.15) and $D/d \geq 1$ gives the following useful result:

$$\sum_{\substack{d|P \\ d \leq D}} |r_A(d)| \leq D \sum_{\substack{d|P \\ d \leq D}} \frac{\rho(d)}{d} \leq D \prod_{p|P} \left(1 - \frac{\rho(p)}{p}\right)^{-1} = \frac{D}{V(P)}. \quad (2.19)$$

Using all the information so far, we now get the essential Meta-theorem of Sieve Theory:

Theorem 2.1 When A satisfies identity (2.14) and λ^\pm satisfies inequality (2.16) we obtain

$$XV^-(D, P) + R^-(D, P) \leq S(A, P) \leq XV^+(D, P) + R^+(D, P), \quad (2.20)$$

where S is as in identities (2.8) and (2.9) and

$$V^\pm(D, P) = \sum_{d|P} \frac{\lambda_D^\pm(d)\rho(d)}{d}, \quad R^\pm(D, P) = \sum_{d|P} \lambda_D^\pm(d)r_A(d). \quad (2.21)$$

Proof:

Taking $B = (a, P)$ in inequality (2.16) and summing over $a \in A$ as in Legendre's Identity (2.10), we get

$$\sum_{d|P} \lambda_D^-(d) |A_d| \leq S(A, P) \leq \sum_{d|P} \lambda_D^+(d) |A_d|.$$

Making use of Identity (2.14) we now have

$$\sum_{d|P} \lambda_D^-(d) \left(X \frac{\rho(d)}{d} + r_A(d) \right) \leq S(A, P) \leq \sum_{d|P} \lambda_D^+(d) \left(X \frac{\rho(d)}{d} + r_A(d) \right),$$

from which the statement of the theorem follows.

Q.E.D.

The goal, therefore, is to find and choose λ^\pm and D so that R^\pm are relatively small in comparison to XV^\pm , and much work is done in estimating V^\pm and R^\pm . It turns out that things can be done in such a way that $\lambda_D^\pm(1) = 1$ (this is known as a normalizing constraint), and we will also get

$$|\lambda_D^\pm(d)| \leq 1 \quad \text{for all } d. \quad (2.22)$$

As a result of this, we have the following

$$\left| \sum_{d|P} \lambda_D^\pm(d) r_A(d) \right| \leq \sum_{\substack{d|P \\ d \leq D}} |r_A(d)|,$$

from which we get a Corollary of Theorem 2.1:

Corollary 2.1.1 *Suppose identity (2.14) and inequalities (2.16) and (2.22) hold. Then*

$$XV^-(D, P) - \sum_{\substack{d|P \\ d \leq D}} |r_A(d)| \leq S(A, P) \leq XV^+(D, P) + \sum_{\substack{d|P \\ d \leq D}} |r_A(d)|.$$

Sifting Density:

We mentioned earlier that the function ρ has a sifting density κ if a certain

average of $\rho(p)$ at primes p is not greater than κ . In many situations, the following is true:

$$\sum_{p < z} \frac{\rho(p) \log(p)}{p} = \kappa \log(z) + O(1). \quad (2.23)$$

When dealing with twin primes, for instance, we had $\rho(p) = 2$ for primes $p \geq 3$ and $\rho(p) = 1$ for $p = 2$. In this case, identity (2.23) follows from identity (2.1) with $\kappa = 2$ (as desired).

It turns out that in many sifting arguments, working with identity (2.23) can lead to annoying complications such as ρ depending on the size of A in some sense. Such complications are avoided, therefore, by using alternate formulations of the sifting density concept. However, the various formulations for sifting density are equivalent in the sense that one can be obtained from the others using the techniques of partial summation. Also, these formulations will not disagree with the $\kappa = 2$ density for twin primes and general prime pairs.

Definition 2.6 We say that κ is a sifting density for the function ρ if there exists a constant $L > 1$ (depending on κ) such that if $2 \leq w \leq z$,

$$\frac{V(P(w))}{V(P(z))} = \prod_{p|P(z)/P(w)} \left(1 - \frac{\rho(p)}{p}\right)^{-1} \leq K_w \left(\frac{\log(z)}{\log(w)}\right)^\kappa, \quad (2.24)$$

with

$$K_w \leq 1 + \frac{L}{\log(w)}. \quad (2.25)$$

In some cases a bit less is used: There exists a constant $K > 1$ such that

$$K_w \leq K \text{ when } w \geq 2. \quad (2.26)$$

Note: The choice $w = p$, $z = p + \epsilon$ shows that inequality (2.24) gives

$$\left(1 - \frac{\rho(p)}{p}\right)^{-1} \leq K_p, \quad (2.27)$$

so $\rho(p)$ is bounded away from p : $\rho(p) \leq p(1 - 1/K_p)$. Also, notice that κ is not uniquely defined. If inequality (2.24) holds for some κ_0 , then it holds for $\kappa \geq \kappa_0$. For some sieve techniques, the following definition is used, in which case, κ will be unique (provided it exists).

Definition 2.7 We say that κ is a two-sided sifting density for ρ if it satisfies Definition 2.6 and, additionally, there exists $L' \geq L$ (depending on ρ) such that

$$\frac{V(P(w))}{V(P(z))} \geq \left(1 - \frac{L'}{\log(w)}\right) \left(\frac{\log(z)}{\log(w)}\right)^\kappa \quad \text{if } 2 \leq w < z. \quad (2.28)$$

For the Selberg Sieve, it turns out that it is more convenient to use something slightly stronger than Definition 2.6.

Definition 2.8 Let $g(d)$ be the multiplicative function defined for squarefree d by specifying for primes p :

$$g(p) = \frac{\rho(p)}{p - \rho(p)} \quad (2.29)$$

Using this definition, we can replace inequality (2.24) by specifying the existence of a constant $A > 1$ such that

$$\sum_{w \leq p < z} g(p) \log(p) \leq \kappa \log\left(\frac{z}{w}\right) + A \quad \text{when } 2 \leq w < z. \quad (2.30)$$

Note: As stated by Lemma 2.4 to follow, inequality (2.30) implies the usual sifting density hypothesis/definition (2.24).

Many sifting situations, including that of twin primes and general prime pairs, satisfy both inequalities (2.24) and (2.30). In particular, the next lemma applies to the twin prime and general prime pair scenarios.

Lemma 2.1 Suppose $\rho(p) < p$ when $p \leq \kappa$ and $\rho(p) \leq \kappa$ for all p . Then ρ has sifting density κ , in the stronger sense expressed by inequality (2.25), and also satisfies inequality (2.30).

Proof:

The inequality $\rho(p) < p$ guarantees $\rho(p) \leq p(1 - 1/K)$ when $p \leq \kappa$, for a certain $K > 1$ depending only on κ , so that $(1 - \rho(p)/p)^{-1} < K$ for these p .

Mertens' product (2.3) shows that

$$\prod_{w \leq p < z} \left(1 - \frac{1}{p}\right)^{-\kappa} = \left(\frac{\log(z)}{\log(w)}\right)^\kappa \left(1 + O\left(\frac{1}{\log(w)}\right)\right),$$

where the O-constant may depend on κ . When $w \geq \kappa$ this gives

$$\prod_{w \leq p < z} \left(1 - \frac{\rho(p)}{p}\right)^{-1} \leq \left(\frac{\log(z)}{\log(w)}\right)^{\kappa} \left(1 + O\left(\frac{1}{\log(w)}\right)\right) \prod_{w \leq p < z} \frac{\left(1 - \frac{1}{p}\right)^{\kappa}}{\left(1 - \frac{\kappa}{p}\right)}.$$

The last product over primes p is

$$\begin{aligned} \exp\left(\sum_{w \leq p < z} \kappa \log\left(1 - \frac{1}{p}\right) - \log\left(1 - \frac{\kappa}{p}\right)\right) &= \exp\left(\sum_{w \leq p < z} O\left(\frac{1}{p^2}\right)\right) \\ &= \exp\left(O\left(\frac{1}{w}\right)\right) \\ &= 1 + O\left(\frac{1}{w}\right), \end{aligned} \quad (2.31)$$

so inequality (2.25) follows, with some value of the constant L . It also holds when $w < \kappa$, possibly with a larger value of L , because the contribution from primes $p < \kappa$ to the product in inequality (2.24) is bounded.

In a similar way observe when $w > \kappa$

$$\sum_{w \leq p < z} \frac{\rho(p) \log(p)}{p - \rho(p)} \leq \sum_{w \leq p < z} \frac{\kappa \log(p)}{p - \kappa} \leq \kappa \log\left(\frac{z}{w}\right) + O(1),$$

where we used identity (2.1). Thus inequality (2.30) follows for these w , and remains valid for smaller w because $g(p)$ is bounded when $p \leq \kappa$.

Q.E.D.

We are now finally ready to discuss the Selberg Sieve but first we record the next four lemmas for completion and referral. See [4, Section 1.3.5] for their proofs and more detailed information on sifting density. The first two lemmas deal with Partial Summation where use of the notation of Stieltjes integrals is convenient. Using integration by parts, one can take as a definition:

$$\int_w^z f(t) d(\bar{c}(t)) = c(z)f(z) - c(w)f(w) - \int_w^z f'(t)c(t) dt \quad (2.32)$$

when f' is piecewise continuous on $[w, z]$ and c is of bounded variation.

Lemma 2.2 Define

$$C(x, y) = \sum_{x \leq n < y} c_n$$

whenever the indicated numbers c_n are defined.

(i) If $w \leq z$ and f' is piecewise continuous on $[w, z]$, then

$$\sum_{w \leq n < z} c_n f(n) = \int_w^z f(t) d(C(w, t)) = - \int_w^z f(t) d(C(t, z)), \quad (2.33)$$

where the integrals with respect to t are given by identity (2.32).

(ii) Suppose, in addition, that f is positive and monotone in $[w, z]$ and that $C(x, y)$ is given by an expression

$$C(x, y) = \gamma(y) - \gamma(x) + e(x, y),$$

where $e(x, y) \leq E$ whenever $w \leq x \leq y \leq z$. Then

$$\sum_{w \leq n < z} c_n f(n) \leq \int_w^z f(t) d(\gamma(t)) + E \max\{f(w), f(z)\}.$$

(iii) If f and g both satisfy the conditions required of f in parts (i) and (ii) then

$$\sum_{w \leq n < z} c_n f(n) g(n) \leq \int_w^z f(t) g(t) d(\gamma(t)) + E \left(\max_{w \leq u \leq z} f(u) \right) \left(\max_{w \leq v \leq z} g(v) \right).$$

Lemma 2.3 Suppose that, when t and n are in an interval $[x, y]$, $f'(t)$ is continuous and $f(t)$ is positive and decreasing, $g(n) \geq 0$, and $E(t, n) \leq A$. Then

$$\int_{x < t < y} f(t) d \left(\sum_{x \leq n < t} g(n) E(t, n) \right) \leq A \sum_{x \leq n < y} f(n) g(n).$$

Lemma 2.4 Suppose ρ is such that the function g satisfies (2.30). Then the sifting density hypothesis follows in the form (2.25), with

$$K \leq e^{A/\log(w)} = 1 + \frac{L}{\log(w)}.$$

Lemma 2.5 Suppose that ρ has sifting density κ , and that $2 \leq w < z$.

(i) When inequality (2.26) holds we find

$$\sum_{w \leq p < z} \frac{\rho(p) \log(p)}{p} \leq (\kappa + \log(K)) \log(z). \quad (2.34)$$

(ii) If inequality (2.25) holds then

$$\sum_{t \leq p < z} \frac{\rho(p)}{p} \leq \kappa \log \left(\frac{\log(z)}{\log(t)} \right) + \frac{L}{\log(t)}, \quad (2.35)$$

$$\sum_{w \leq p < z} \frac{\rho(p) \log(p)}{p} \leq \kappa \log \left(\frac{z}{w} \right) + L \log \left(e \frac{\log(z)}{\log(w)} \right). \quad (2.36)$$

Furthermore the function $g(p)$ defined in identity (2.29) satisfies

$$g(p) \leq \frac{L}{\log(p)}, \quad (2.37)$$

$$\begin{aligned} \sum_{w \leq p < z} g(p) \log(p) &\leq \kappa \log \left(\frac{z}{w} \right) + L(1 + \kappa) \log \left(e \frac{\log(z)}{\log(w)} \right) \\ &\quad + \frac{L^2}{\log(w)}. \end{aligned} \quad (2.38)$$

Chapter 3

The Selberg Sieve

3.1 The Set-up

3.1.1 Selberg's Main Theorem

The main goal of this section is Theorem 3.1.1. It seems a bit artificial at first glance and one wonders how it was conceived, but keep in mind that Theorem 3.1.1 is the end result of a series of discoveries. The main thing to focus on here is the upper bound (3.1.9) for $S(A, P)$. $G(\sqrt{D})$ is as in identity (3.1.4) and it is being used in Theorem 3.1.1 because, as will be seen later, it can be estimated in such a way that we get nice results/estimates for the upper bound (3.1.9) for $S(A, P)$.

The rest of Theorem 3.1.1, i.e., identity (3.1.10) for the error term and identity (3.1.11) involving λ , are consequences of Lemma 3.1.1. In turn, Lemma 3.1.1 gives several identities, (3.1.6), (3.1.7), (3.1.8), that are consequences of identity (3.1.5) for $V^+(\lambda)$ and are relevant to the proof and result of Theorem 3.1.1. The identity (3.1.5) shows up in the proof of Theorem 3.1.1 as a result of Selberg's clever idea: inequality (3.1.14) involving the square of the sum of the λ values.

We first recall the following:

$$|A_d| = X \frac{\rho(d)}{d} + r_A(d) \text{ if } d \mid P, d < D \quad (3.1.1)$$

where $|A_d|$ is the number of a in A divisible by a squarefree number d , and ρ is multiplicative. The symbol $[d_1, d_2]$ denotes a least common multiple. The function ρ^* will be the multiplicative function "conjugate" to ρ in that its values at primes p are given as

$$\rho^* = p - \rho(p). \quad (3.1.2)$$

This defines $\rho^*(d)$ for squarefree d , the only d that will be relevant. Then the function from identity (2.29) is

$$g(n) = \begin{cases} \rho(n)/\rho^*(n) & \text{if } n \text{ is squarefree} \\ 0 & \text{otherwise.} \end{cases} \quad (3.1.3)$$

In Theorem 3.1.1 and elsewhere we shall denote

$$G(x) = \sum_{\substack{n \leq x \\ n|P}} g(n), \quad (3.1.4)$$

where P is a specified product of primes.

In Lemma 3.1.1, the numbers $\lambda(d)$ may be arbitrary, subject only to a restriction that they are supported on an interval $1 \leq d \leq \sqrt{D}$, so that all the sums appearing in Lemma 3.1.1 are finite. The restriction to $\rho(h) \neq 0$ in identity (3.1.6) is a natural one, because for other h the summand $x^2(h)/g(h)$ is a multiple of $\rho(h)$ as follows: identity (3.1.7) gives $\rho(h) \mid x(h)$, so that $x^2(h) = \rho^2(h) \cdot l$, for some l . Now

$$\frac{x^2(h)}{g(h)} = \frac{(x^2(h)\rho^*(h))}{\rho(h)} = \frac{(\rho^2(h) \cdot l \cdot \rho^*(h))}{\rho(h)},$$

so after cancellation of $\rho(h)$ from the numerator and denominator, the summand is a multiple of $\rho(h)$.

Lemma 3.1.1 *Define*

$$V^+(\lambda) = \sum_{d_1|P} \sum_{d_2|P} \frac{\lambda(d_1)\lambda(d_2)\rho([d_1, d_2])}{[d_1, d_2]}. \quad (3.1.5)$$

Then

$$V^+(\lambda) = \sum_{\substack{h < \sqrt{D} \\ \rho(h) \neq 0}} \frac{x^2(h)}{g(h)}, \quad (3.1.6)$$

where g is as in identity (3.1.3) and

$$x(h) = \sum_{d \equiv 0 \pmod{h}} \frac{\lambda(d)\rho(d)}{d}. \quad (3.1.7)$$

Moreover

$$\frac{\lambda(d)\rho(d)}{d} = \sum_k \mu(k)x(kd). \quad (3.1.8)$$

Proof:

Using $[d_1, d_2] = d_1 d_2 / (d_1, d_2)$ in equation (3.1.5) gives

$$V^+(\lambda) = \sum_{\substack{d_1 \\ \rho((d_1, d_2)) \neq 0}} \sum_{d_2} \frac{\lambda(d_1)\rho(d_1)}{d_1} \frac{\lambda(d_2)\rho(d_2)}{d_2} \frac{(d_1, d_2)}{\rho((d_1, d_2))}.$$

Since $f = (d_1, d_2)$ is squarefree, the last fraction may be expressed as

$$\frac{f}{\rho(f)} = \prod_{p|f} \frac{p}{\rho(p)} = \prod_{p|f} \left(1 + \frac{\rho^*(p)}{\rho(p)}\right) = \sum_{h|f} \frac{1}{g(h)}.$$

We therefore have,

$$V^+(\lambda) = \sum_{\substack{d_1 \\ \rho((d_1, d_2)) \neq 0}} \sum_{d_2} \frac{\lambda(d_1)\rho(d_1)}{d_1} \frac{\lambda(d_2)\rho(d_2)}{d_2} \sum_{h|(d_1, d_2)} \frac{1}{g(h)}.$$

Rearranging the order of summation gives

$$\begin{aligned} V^+(\lambda) &= \sum_{\substack{h < \sqrt{D} \\ \rho(h) \neq 0}} \frac{1}{g(h)} \sum_{d_1 \equiv 0 \pmod{h}} \frac{\lambda(d_1)\rho(d_1)}{d_1} \sum_{d_2 \equiv 0 \pmod{h}} \frac{\lambda(d_2)\rho(d_2)}{d_2} \\ &= \sum_{\substack{h < \sqrt{D} \\ \rho(h) \neq 0}} \frac{x^2(h)}{g(h)}, \end{aligned}$$

which is identity (3.1.6), where $x(h)$ is as in identity (3.1.7). Lastly, identity (3.1.7) inverts to give

$$\begin{aligned} \sum_k \mu(k)x(kN) &= \sum_{kN|d} \mu(k) \frac{\lambda(d)\rho(d)}{d} = \sum_{kNl=d} \mu(k) \frac{\lambda(d)\rho(d)}{d} \\ &= \sum_{mN=d} \frac{\lambda(d)\rho(d)}{d} \sum_{k|m} \mu(k) \\ &= \frac{\lambda(N)\rho(N)}{N}, \end{aligned}$$

which is identity (3.1.8). Here we used the characteristic property (2.4) of the Möbius function.

Q.E.D.

Theorem 3.1.1 Suppose identity (3.1.1) holds. Then

$$S(A, P) \leq \frac{X}{G(\sqrt{D})} + E(D, P), \quad (3.1.9)$$

where G is as in identity (3.1.4),

$$E(D, P) = \frac{1}{\lambda^2(1)} \sum_{\substack{d_1 < \sqrt{D} \\ d_1 | P}} \sum_{\substack{d_2 < \sqrt{D} \\ d_2 | P}} \lambda(d_1) \lambda(d_2) r_A([d_1, d_2]), \quad (3.1.10)$$

and the real numbers $\lambda(d)$ are given, for some $C \neq 0$, by

$$\frac{\lambda(d)\rho(d)}{d} = C\mu(d) \sum_{\substack{h \equiv 0 \pmod{d} \\ h < \sqrt{D}, h|P}} g(h) \quad \text{if } \rho(d) \neq 0, \quad (3.1.11)$$

with $\lambda(d) = 0$ if $\rho(d) = 0$.

Some Notes: Taking $d = 1$ in equation (3.1.11) shows

$$C = \frac{\lambda(1)}{G(\sqrt{D})}. \quad (3.1.12)$$

The usual normalization in Theorem 3.1.1 is to make $\lambda(1) = 1$.

The value of $\lambda(d)$ when $\rho(d) = 0$ is irrelevant to the coefficient of X in Theorem 3.1.1. The choice $\lambda(d) = 0$ for these d is the most efficient one when the error term $E(D, P)$ is considered.

Since g is as in identity (3.1.3), substituting $h = dk$ in equation (3.1.11) gives

$$\lambda(d) = C\mu(d) \frac{d}{\rho^*(d)} \sum_{\substack{(k, d)=1 \\ k < \sqrt{D}/d, k|P}} g(k). \quad (3.1.13)$$

Proof:

Suppose, for the moment, that $\lambda(d)$ are arbitrary real numbers supported on squarefree $d < \sqrt{D}$. We start with Selberg's brilliant idea

$$\lambda^2(1)S(a, P) \leq \left(\sum_{d|(a, P)} \lambda(d) \right)^2. \quad (3.1.14)$$

This holds because if $(a, P) = 1$, then $S(a, P) = 1$ and both sides of inequality (3.1.14) take the value $\lambda^2(1)$, but if $(a, P) > 1$, then $S(a, P) = 0$ and the right side is non-negative.

The right side of inequality (3.1.14) can be written

$$\sum_{d_1|(a, P)} \sum_{d_2|(a, P)} \lambda(d_1)\lambda(d_2) = \sum_{d_1|P} \sum_{d_2|P} \lambda(d_1)\lambda(d_2) \sum_{[d_1, d_2]|a} 1.$$

On summing over $a \in A$ and expressing the result in terms of ρ from identity (3.1.1), this gives

$$\lambda^2(1)S(A, P) \leq XV^+(\lambda) + \lambda^2(1)E(D, P), \quad (3.1.15)$$

with $V^+(\lambda)$ as in identity (3.1.5) and $E(D, P)$ as stated in identity (3.1.10). If we had $\lambda(1) = 1$, then we would have achieved the result of Theorem 2.1. The conclusions of Theorem 3.1.1 are achieved by minimizing $V^+(\lambda)$, thereby getting the best bound possible for $S(A, P)$. We now proceed to show how this is done.

In discussing $V^+(\lambda)$ we may streamline the notation by using the convention that $\rho(d) = 0$ if $d \nmid P$, together with a natural one that terms with $\rho(d) = 0$ are not to be included in summations over d . This will save a certain amount of repetition of the conditions $d \mid P$ or $\rho(d) \neq 0$, where these are implied. Similarly, the conditions $d < \sqrt{D}$, $\mu^2(d) = 1$ need not be explicit in a sum involving $\lambda(d)$.

Theorem 3.1.1 will follow by choosing $x(k)$ so as to minimize, for given $\lambda(1)$, the expression (3.1.6) for the quantity $V^+(\lambda)$. The constraint

$$\sum_{k < \sqrt{D}} \mu(k)x(k) = \lambda(1) \quad (3.1.16)$$

is now required by identity (3.1.8).

We will use Cauchy's inequality

$$\left(\sum_{h < H} a_h b_h \right)^2 \leq \sum_{h < H} a_h^2 \sum_{h < H} b_h^2,$$

which holds with equality when there is a constant C with $a_h = Cb_h$ whenever $h < H$. This is an immediate consequence of the identity

$$\sum_{h < H} a_h^2 \sum_{k < H} b_k^2 - \left(\sum_{h < H} a_h b_h \right)^2 = \frac{1}{2} \left(\sum_{h < H, k < H} a_h b_k - a_k b_h \right)^2.$$

Apply Cauchy's inequality to the condition (3.1.16) in a way that relates to identity (3.1.6) for $V^+(\lambda)$:

$$\begin{aligned}\lambda^2(1) &= \left(\sum_{k < \sqrt{D}} \frac{\mu(k)x(k)}{\sqrt{g(k)}} \cdot \sqrt{g(k)} \right)^2 \\ &\leq \left(\sum_{k < \sqrt{D}} \frac{x^2(k)}{g(k)} \right) \left(\sum_{k < \sqrt{D}} g(k) \right) \leq V^+(\lambda)G(\sqrt{D}),\end{aligned}\quad (3.1.17)$$

where $G(\sqrt{D})$ is as stated in identity (3.1.4). Equality occurs in (3.1.17) if

$$x(k) = C\mu(k)g(k) \text{ when } k < \sqrt{D}, \quad (3.1.18)$$

in which case inequality (3.1.15) gives the conclusion (3.1.9) required in Theorem 3.1.1. This situation is attained if we make $\lambda(d)$ satisfy equation (3.1.8) with these $x(k)$, so that

$$\frac{\lambda(d)\rho(d)}{d} = C \sum_{k < \sqrt{D}/d} \mu(k)\mu(kd)g(kd) = C\mu(d) \sum_{\substack{h \equiv 0 \pmod{d} \\ h < \sqrt{D}}} g(h).$$

This is the value stated in equation (3.1.11), in which the implied constraint $h \mid P$ was written explicitly.

Q.E.D.

3.1.2 Bounding $\lambda(d)$

The following theorem states that $|\lambda(d)| \leq |\lambda(1)|$. Keep in mind that the usual normalization in Theorem 3.1.1 is to make $\lambda(1) = 1$, so Theorem 3.1.2 provides a nice, convenient upper bound on $\lambda(d)$.

For given d , a squarefree number h may be written uniquely as $h = fn$, where $f \mid d$ and $(n, d) = 1$. Hence

$$\sum_{h < \sqrt{D}} g(h) \leq \sum_{f \mid d} g(f) \sum_{\substack{(n, d)=1 \\ n < \sqrt{D}}} g(n). \quad (3.1.19)$$

Observe from identity (3.1.2) that for squarefree d the sum over f in (3.1.19) is

$$\sum_{f|d} g(f) = \prod_{p|d} \left(1 + \frac{\rho(p)}{p - \rho(p)} \right) = \frac{d}{\rho^*(d)}. \quad (3.1.20)$$

As before, the fact that f , n and h divide P may be taken as implicit in the definition of the function ρ .

Theorem 3.1.2 *The numbers $\lambda(d)$ described in Theorem 3.1.1 satisfy the inequality*

$$|\lambda(d)| \leq |\lambda(1)|.$$

Proof:

Let C be the normalizing constant as in identity (3.1.12), so that

$$\lambda(1) = C \sum_{h < \sqrt{D}} g(h).$$

This leads to the following inequality which is in the opposite sense/direction to inequality (3.1.19):

$$|\lambda(1)| \geq |C| \sum_{f|d} g(f) \sum_{\substack{(n,d)=1 \\ n < \sqrt{D}/d}} g(n). \quad (3.1.21)$$

With the identity (3.1.20) this gives

$$|\lambda(1)| \geq |C| \frac{d}{\rho^*(d)} \sum_{\substack{(n,d)=1 \\ n < \sqrt{D}/d}} g(n) = |\lambda(d)|,$$

by the expression (3.1.13) for $\lambda(d)$.

Q.E.D.

Corollary 3.1.2.1 *In Theorem 3.1.1, the error term satisfies*

$$E(D, P) \leq \sum_{\substack{d_1 < \sqrt{D} \\ d_1 | P}} \sum_{\substack{d_2 < \sqrt{D} \\ d_2 | P}} |r_A([d_1, d_2])|.$$

Proof:

The result immediately follows from identity (3.1.10) for $E(D, P)$ and the previous theorem.

Q.E.D.

3.2 Estimates for $G(x)$ and $E(D, P)$

We would now like to get some estimates for $G(x)$ and $E(D, P)$. We address each of these in turn.

3.2.1 Using Rankin's Device

To get estimates for $G(x)$, we will define a useful, related expression, $G_z(x)$. Then the next theorem will provide upper and lower bounds for $G_z(\sqrt{D})$, where the only assumptions about ρ used are:

$$\begin{aligned} (1) \quad & 0 \leq \rho(p) < p \\ (2) \quad & \frac{1}{\log(z)} \sum_{p < z} \frac{\rho(p) \log(p)}{p} \leq B \quad (\text{by Lemma 2.5}), \end{aligned} \quad (3.2.22)$$

where B is a constant. The proof of the theorem relies on the next lemma, which provides an upper bound for $I_z(x) = G_z(\infty) - G_z(x)$. The proof of the lemma uses a technique known as Rankin's device; it will be pointed out.

In connection with $G(x)$, we have the "incomplete" sum

$$G_z(x) = \sum_{\substack{d \leq x \\ d|P(z)}} g(d). \quad (3.2.23)$$

This is considered "incomplete" because the sum is over all d dividing $P(z)$ (not P) whereas the sum for $G(x)$ is over all d dividing P . (Recall: $P(z) = \prod_{p|P, p < z} p$). For future reference recall that the equation $g(p) = \rho(p)/(p - \rho(p))$ introduced earlier can be rewritten as

$$\left(1 - \frac{\rho(p)}{p}\right) g(p) = \frac{\rho(p)}{p}, \quad 1 + g(p) = \left(1 - \frac{\rho(p)}{p}\right)^{-1} \quad (3.2.24)$$

Observe that $G_z(x)$ increases with z , so

$$G(x) = G_x(x) \geq G_z(x) \quad (3.2.25)$$

when $z \leq x$. In Theorem 3.1.1, a lower bound for $G(x)$ is required (with $x = \sqrt{D}$), so a lower bound for $G_z(x)$ is relevant.

It is natural to denote

$$G_z(\infty) = \sum_{d|P(z)} g(d),$$

so that $G_z(x) = G_z(\infty)$ whenever $x \geq P(z)$. Observe, referring to the identities in (3.2.24) where necessary,

$$G_z(x) \leq G_z(\infty) = \prod_{p|P(z)} (1 + g(p)) = \prod_{p|P(z)} \left(1 - \frac{\rho(p)}{p}\right)^{-1} = \frac{1}{V(P(z))}. \quad (3.2.26)$$

We see here a connection between $G_z(x)$ coming from Selberg's method and the product $V(P)$, but the inequality is in the wrong direction. The next theorem will give us what we seek, but first the lemma.

Lemma 3.2.1 *Suppose that the assumptions (3.2.22) about ρ hold, and denote*

$$I_z(x) = G_z(\infty) - G_z(x) = \sum_{\substack{d \geq x \\ d|P(z)}} g(d), \quad (3.2.27)$$

and write $x = z^v$. Then for each $c \geq 0$

$$I_z(x) \leq \frac{1}{V(P(z))} \exp(-cv + B(e^c - 1)). \quad (3.2.28)$$

In particular

$$I_z(x) \leq \frac{\exp(-\psi_B(v))}{V(P(z))}, \quad (3.2.29)$$

where, for each $v > 0$,

$$\psi_B(v) = \max \left\{ 0, v \log \left(\frac{v}{B} \right) - v + B \right\} = \int_{B < t < v} \log \left(\frac{t}{B} \right) dt.$$

Proof:

We use Rankin's device. Letting $\epsilon \geq 0$, we observe that when $d \geq x$, we have $(d/x)^\epsilon \geq 1^\epsilon = 1$, so we get

$$\begin{aligned}
I_z(x) &= \sum_{\substack{d \geq x \\ d|P(z)}} g(d) \leq \sum_{\substack{d \geq x \\ d|P(z)}} g(d) \left(\frac{d}{x}\right)^\epsilon \\
&\leq \sum_{\substack{d \geq x \\ d|P(z)}} g(d) \left(\frac{d}{x}\right)^\epsilon + \sum_{\substack{d < x \\ d|P(z)}} g(d) \left(\frac{d}{x}\right)^\epsilon \\
&= \sum_{d|P(z)} g(d) \left(\frac{d}{x}\right)^\epsilon = \frac{1}{x^\epsilon} \sum_{d|P(z)} g(d) d^\epsilon \\
&= \frac{1}{x^\epsilon} \prod_{p|P(z)} (1 + p^\epsilon g(p)). \tag{3.2.30}
\end{aligned}$$

Hence, by the definitions of $V(P(z))$ and $g(p)$:

$$\begin{aligned}
I_z(x)V(P(z)) &\leq \frac{1}{x^\epsilon} \prod_{p|P(z)} \left(\frac{p - \rho(p)}{p}\right) \left(1 + \frac{p^\epsilon \rho(p)}{p - \rho(p)}\right) \\
&= \frac{1}{x^\epsilon} \prod_{p|P(z)} \left(1 + \frac{\rho(p)}{p} (p^\epsilon - 1)\right) \\
&\leq \frac{1}{x^\epsilon} \exp \left(\sum_{p|P(z)} \frac{\rho(p)}{p} (p^\epsilon - 1) \right),
\end{aligned}$$

the last inequality following from the fact $1 + x \leq e^x$.

We may choose $\epsilon = c/\log(z)$, provided $c \geq 0$. Then $z^\epsilon = e^c$. Observing that $(e^t - 1)/t$ increases when $t > 0$, we have when $p < z$,

$$\frac{p^\epsilon - 1}{\epsilon \log(p)} = \frac{e^{c \log(p)/\log(z)} - 1}{c \log(p)/\log(z)} \leq \frac{e^c - 1}{c}.$$

It follows that

$$p^\epsilon - 1 \leq \frac{e^c - 1}{c} (\epsilon \log(p)) = \frac{e^c - 1}{c} \cdot \frac{c}{\log(z)} \cdot \log(p) = (e^c - 1) \frac{\log(p)}{\log(z)},$$

and when $x = z^v$, $x^\epsilon = (z^v)^\epsilon = (z^\epsilon)^v = (e^c)^v = e^{cv}$. Using these facts along with the assumptions (3.2.22) about ρ , and the fact that summing over $p < z$

is greater or equal than summing over $p \mid P(z)$, we now have

$$\begin{aligned}
 I_z(x)V(P(z)) &\leq \frac{1}{x^\epsilon} \exp \left(\sum_{p \mid P(z)} \frac{\rho(p)}{p} (p^\epsilon - 1) \right) \\
 &\leq \frac{1}{e^{cv}} \exp \left(\sum_{p < z} \frac{\rho(p)}{p} (e^c - 1) \frac{\log(p)}{\log(z)} \right) \\
 &= \frac{1}{e^{cv}} \exp \left(\frac{e^c - 1}{\log(z)} \sum_{p < z} \frac{\rho(p)}{p} \log(p) \right) \\
 &\leq \exp(-cv + B(e^c - 1)),
 \end{aligned}$$

as required for inequality (3.2.28).

The optimal choice of c satisfies $v = Be^c$, i.e. $c = \log(v) - \log(B)$, provided $v > B$. If $v \leq B$ then the best permissible choice is $c = 0$. This yields inequality (3.2.29) and the proof is complete.

Q.E.D.

Theorem 3.2.1 *Suppose that the assumptions (3.2.22) about ρ hold, where $z \geq 2$, and write $z = D^{1/s}$. Then $G_z(\sqrt{D})$, as defined by (3.2.23), satisfies*

$$\frac{1 - \exp(-\psi_B(\frac{1}{2}s))}{V(P(z))} \leq G_z(\sqrt{D}) \leq \frac{1}{V(P(z))}, \quad (3.2.31)$$

where ψ_B is as in the last lemma. In particular,

$$\begin{aligned}
 \psi_B\left(\frac{1}{2}s\right) &\sim \frac{1}{2}s \log(s) \text{ as } s \rightarrow \infty, \\
 \exp\left(-\psi_B\left(\frac{1}{2}s\right)\right) &\leq e^{7B-s} \text{ for all } s > 0.
 \end{aligned}$$

Proof:

The second inequality in (3.2.31) is just (3.2.26). The proof of (3.2.31) is completed by noting that inequality (3.2.26), identity (3.2.27), and inequality (3.2.29) give

$$G_z(x) = G_z(\infty) - I_z(x) \geq \frac{1 - \exp(-\psi_B(v))}{V(P(z))},$$

and taking $x = \sqrt{D}$, so that the notations $z^v = x$ and $z = D^{1/s}$ require $v = \frac{1}{2}s$. We get $\exp(\psi_B(\frac{1}{2}s)) \leq e^{7B-s}$ by taking the choice $c = 2$ in the last lemma and noting that $e^2 - 1 < 7$.

Q.E.D.

3.2.2 An Asymptotic Bound

Though the last theorem provided a lower bound for $G_z(\sqrt{D})$, it is not the sharpest result. It turns out that in many applications of Theorem 3.1.1, one wishes to estimate $G_z(x)$ from below in the case when $z = x$. The next theorem provides a much desired, sharp asymptotic lower bound. The proof of the theorem makes use of

$$\sum_{p < v} g(p) \log(p) = \kappa \log(v) + \eta(v) \quad (3.2.32)$$

where $\eta(t)$ is required to be bounded above. This is actually a little stronger than what is given by Lemma 2.5. It is possible to proceed just based on Lemma 2.5 but using (3.2.32) makes things a bit easier. For the sifting situations we're interested in, $\eta(t)$ will indeed be bounded above, allowing us to make use of Theorem 3.2.2. The proof of the theorem also relies on the next lemma, which gives an integral equation for $G_z(x)$. To facilitate the lemma we write

$$g_z(n) = \begin{cases} g(n) & \text{if } n \mid P(z) \\ 0 & \text{otherwise,} \end{cases}$$

so the expression (3.2.23) becomes

$$G_z(x) = \sum_{\substack{n < x \\ n \mid P(z)}} g(n) = \sum_{n < x} g_z(n). \quad (3.2.33)$$

Lemma 3.2.2 *When $z \leq x$ the expression $G_z(x)$ given in (3.2.33) satisfies*

$$G_z(x) \log(x) = \int_1^x G_z(t) \frac{dt}{t} + \kappa \int_{x/z}^x G_z(t) \frac{dt}{t} + \delta_z(x), \quad (3.2.34)$$

where $\delta_z(x)$ is expressible as

$$\delta_z(x) = \sum_{n < x} g_z(n) E_{z,n} \left(\frac{x}{n} \right) \quad \text{with} \quad E_{z,n}(t) \leq \eta(\min(t, z)) \quad (3.2.35)$$

Proof:

Since m is squarefree when $g_z(m) \neq 0$ we have

$$\begin{aligned} \sum_{m < x} g_z(m) \log(m) &= \sum_n \sum_{\substack{p \\ np < x}} g_z(np) \log(p) \\ &= \sum_{n < x} g_z(n) \sum_{\substack{p < x/n \\ p < z}} g(p) \log(p) \\ &\quad - \sum_{n < x} g_z(n) \sum_{\substack{p < x/n, p < z \\ p|n}} g(p) \log(p), \end{aligned}$$

where the second equality follows from g being multiplicative and m being squarefree.

Write $n = pk$ in the last sum. This gives

$$\begin{aligned} \sum_{m < x} g_z(m) \log(m) &= \sum_{n < x} g_z(n) \sum_{\substack{p < x/n \\ p < z}} g(p) \log(p) \\ &\quad - \sum_{k < x} g_z(k) \sum_{\substack{p < \sqrt{x/k} \\ p < z, p \nmid k}} g^2(p) \log(p) \\ &= \sum_{n < x} g_z(n) F_{z,n} \left(\frac{x}{n} \right), \end{aligned}$$

where the use of equation (3.2.32) shows

$$\begin{aligned} F_{z,n}(y) &= \sum_{\substack{p < y \\ p < z}} g(p) \log(p) - \sum_{\substack{p < \sqrt{y}, p < z \\ (p,n)=1}} g^2(p) \log(p) \\ &\leq \sum_{\substack{p < y \\ p < z}} g(p) \log(p) = \kappa \log(\min(y, z)) + \eta(\min(y, z)). \end{aligned}$$

Hence

$$\begin{aligned} \sum_{n < x} g_z(n) \log(n) &= \kappa \sum_{n < x} g_z(n) \log \left(\min \left(\frac{x}{n}, z \right) \right) + \delta_z(x) \\ &= \kappa \sum_{n < x} g_z(n) \log \left(\frac{x}{n} \right) - \kappa \sum_{n < x/z} g_z(n) \log \left(\frac{x/z}{n} \right) + \delta_z(x), \quad (3.2.36) \end{aligned}$$

where, setting $E_{z,n}(y) = F_{z,n}(y) - \kappa \log(y)$, we have

$$\delta_z(x) = \sum_{n < x} g_z(n) E_{z,n}\left(\frac{x}{n}\right) \quad \text{with} \quad E_{z,n}(t) \leq \eta(\min(t, z)), \quad (3.2.37)$$

the desired expression for δ_z . The last equality of (3.2.36) is best seen by looking at the two cases: when $x/n < z$, $x/z < n$ making the second sum equal to zero, and when $z < x/n$, the difference agrees with the expression after the first equality in (3.2.36).

Next, note the following, valid for $x \geq 1$,

$$\sum_{m < x} g(m) \log\left(\frac{x}{m}\right) = \sum_{m < x} g(m) \int_m^x \frac{dt}{t} = \int_1^x \sum_{m < t} g(m) \frac{dt}{t}.$$

Using this and the definition of $G_z(x)$, we can rearrange and rewrite (3.2.36) to get

$$G_z(x) \log(x) - (\kappa + 1) \int_1^x G_z(t) \frac{dt}{t} + \kappa \int_1^{x/z} G_z(t) \frac{dt}{t} = \delta_z(x).$$

This is equivalent to equation (3.2.34), so the proof is complete.

Q.E.D.

Theorem 3.2.2 *Let $G(z) = G_z(z)$ and $\eta(t)$ be as specified in (3.2.25), (3.2.23), and (3.2.32). Assume*

$$\eta(t) \leq A \text{ when } 2 \leq t < z \quad (3.2.38)$$

where $A > 1$. Then

$$G(z) \geq \frac{e^{-\gamma\kappa}}{\Gamma(\kappa+1)V(P(z))} \left(1 + O\left(\frac{A}{\log(z)}\right)\right), \quad (3.2.39)$$

where $V(P)$ is as defined earlier. The implied constant depends only on κ . In (3.2.39) we may use

$$\frac{e^{-\gamma\kappa}}{V(P(x))} = \log^\kappa x \left(1 + O\left(\frac{1}{\log(x)}\right)\right) \prod_{p < x} \left(1 - \frac{1}{p}\right)^\kappa \left(1 - \frac{\rho(p)}{p}\right)^{-1}. \quad (3.2.40)$$

Equation (3.2.40) follows from Mertens' product discussed earlier. The form on the right (with $x = z$) is what will be seen in the following proof.

Proof:

This theorem refers to $G_z(x)$ only in the case $z = x$, and Lemma 3.2.2 gives

$$G(x) \log(x) = (\kappa + 1) \int_1^x G(t) \frac{dt}{t} + \delta(x), \quad (3.2.41)$$

where

$$\delta(x) = \sum_{n < x} g(n) E_n \left(\frac{x}{n} \right). \quad (3.2.42)$$

The primes p exceeding z are irrelevant to the theorem since all functions involving primes are only concerned with primes less than or equal to z . Therefore for present purposes, re-define $\rho(p)$ for $p \geq z$ by specifying

$$1 - \frac{\rho(p)}{p} = \left(1 - \frac{1}{p} \right)^\kappa \quad \text{if } p \geq z. \quad (3.2.43)$$

This will make questions relating to the convergence (as $x \rightarrow \infty$) of the product in equation (3.2.40) become trivial. Also, when $p > z$ the function g as in identities (3.2.24) now satisfies

$$\begin{aligned} g(p) &= \frac{\rho(p)}{p - \rho(p)} = \frac{\rho(p)/p}{1 - \rho(p)/p} = \frac{(\rho(p)/p) + 1 - 1}{1 - \rho(p)/p} = \frac{-(1 - \rho(p)/p) + 1}{1 - \rho(p)/p} \\ &= -1 + \frac{1}{\left(1 - \frac{1}{p} \right)^\kappa} = -1 + \left(1 - \frac{1}{p} \right)^{-\kappa} = \frac{\kappa}{p} + O \left(\frac{1}{p^2} \right). \end{aligned} \quad (3.2.44)$$

The last equality follows from the binomial theorem and the two equalities before that follow from (3.2.43). Now inequality (3.2.38) extends to an inequality valid also when $t \geq z$,

$$\eta(t) \leq A_1 \ll A \quad \text{when } 2 \leq t, \quad (3.2.45)$$

where the constant A_1 may be somewhat larger than A .

Using Stieltjes integrals, equation (3.2.41) leads to

$$\int_2^x \frac{d\delta(t)}{\log^{\kappa+1}(t)} = \int_2^x \frac{1}{\log^{\kappa+1}(t)} \left(\log(t) dG(t) - \kappa \frac{G(t)}{t} dt \right) = \int_2^x d \frac{G(t)}{\log^\kappa(t)}. \quad (3.2.46)$$

This is best seen as follows. Re-arranging (3.2.41) gives

$$\delta(x) = G(x) \log(x) - (\kappa + 1) \int_1^x G(t) \frac{dt}{t}.$$

Differentiation gives

$$\begin{aligned} d\delta(t) &= \left(d(G(t)) \log(t) + G(t) \frac{dt}{t} \right) - (\kappa + 1) \frac{G(t)}{t} dt \\ &= d(G(t)) \log(t) - \kappa \frac{G(t)}{t} dt, \end{aligned}$$

thereby giving

$$\frac{d\delta(t)}{\log^{\kappa+1}(t)} = \frac{1}{\log^{\kappa+1}(t)} \left(d(G(t)) \log(t) - \kappa \frac{G(t)}{t} dt \right). \quad (3.2.47)$$

Now we also compute the following

$$\begin{aligned} d \left(\frac{G(t)}{\log^{\kappa}(t)} \right) &= d(G(t)) \log^{-\kappa}(t) + G(t) (-\kappa \log^{-\kappa-1}(t)) \frac{1}{t} dt \\ &= \frac{d(G(t)) \log(t)}{\log^{\kappa+1}(t)} - \kappa \frac{1}{\log^{\kappa+1}(t)} \frac{G(t)}{t} dt \\ &= \frac{1}{\log^{\kappa+1}(t)} \left(d(G(t)) \log(t) - \kappa \frac{G(t)}{t} dt \right). \end{aligned} \quad (3.2.48)$$

Equating (3.2.47) and (3.2.48) and integrating yields (3.2.46).

Denote

$$H(x) = \frac{G(x)}{\log^{\kappa}(x)} \quad \text{if } x \geq 2. \quad (3.2.49)$$

Then, for $2 \leq x < y$, (3.2.46) and (3.2.42) give

$$H(y) - H(x) = \int_x^y \frac{d\delta(t)}{\log^{\kappa+1}(t)} = \int_x^y \frac{1}{\log^{\kappa+1}(t)} d \sum_{x \leq n < t} g(n) E_n \left(\frac{t}{n} \right).$$

Here $E_n(t/n) \leq A_1$ because of identity (3.2.35) and inequality (3.2.45), so Lemma 2.3 applies, to give

$$H(y) - H(x) \leq A_1 \sum_{x \leq n < y} \frac{g(n)}{\log^{\kappa+1}(n)}. \quad (3.2.50)$$

The sum in (3.2.50) can be expressed in terms of the function H by a “condensation” procedure. Choose $c > 1$. When $U \geq 2$ we find

$$\sum_{U \leq n < U^c} \frac{g(n)}{\log^{\kappa+1}(n)} \leq \frac{1}{\log(U)} \frac{G(U^c)}{\log^{\kappa}(U)} = c^{\kappa} \frac{1}{\log(U)} H(U^c).$$

The last equality follows from (3.2.49) and basic rules of logarithms.

So we now have

$$\sum_{x \leq n < y} \frac{g(n)}{\log^{\kappa+1}(n)} = \sum_{r \geq 0} \sum_{\substack{x c^r \leq n < x c^{r+1} \\ n < y}} \frac{g(n)}{\log^{\kappa+1}(n)} \leq c^{\kappa} \sum_{r \geq 0} \frac{1}{c^r \log(x)} \sup_{x \leq t < y} H(t).$$

Using this in (3.2.50) gives

$$H(y) - H(x) \leq \frac{A_2}{\log(x)} \sup_{x \leq t < y} H(t), \quad (3.2.51)$$

with $A_2 = CA_1$. Here we have

$$C = c^{\kappa} \sum_{r \geq 0} \frac{1}{c^r}.$$

Now choosing $c = 1 + 1/(\kappa + 1)$ ensures $|1/c| < 1$ so that we get

$$C = c^{\kappa} \sum_{r \geq 0} \frac{1}{c^r} = c^{\kappa} \frac{1}{1 - (1/c)} = \frac{c^{\kappa+1}}{c - 1}.$$

Finally, using the fact that $(1 + (1/n))^n < e$ for all $n > 0$, (see [7, pgs. 63-64]), we have

$$C = \frac{c^{\kappa+1}}{c - 1} = \frac{\left(1 + \frac{1}{\kappa+1}\right)^{\kappa+1}}{1/(\kappa+1)} = (\kappa+1) \left(1 + \frac{1}{\kappa+1}\right)^{\kappa+1} \leq (\kappa+1)e.$$

Now let $0 < \epsilon < \frac{1}{2}$, and let $J = \limsup_{t \rightarrow \infty} H(t)$. Then J is finite, for otherwise there would exist a sequence y_n such that

$$H(y_n) \rightarrow \infty \text{ as } n \rightarrow \infty, \quad H(y_n) > (1 - \epsilon) \sup_{t \leq y_n} H(t),$$

and combining with inequality (3.2.51) then gives $1 - \epsilon \leq A_2/\log(x)$, which is false for large x .

If x is sufficiently large, then $\sup_{x \leq t} H(t) < J + \epsilon$. We can choose $y > x$ so that $H(y) > J - \epsilon$. We manipulate inequality (3.2.51) and use these inequalities to get

$$\begin{aligned} H(x) &\geq H(y) - \frac{A_2}{\log(x)} \sup_{x \leq t \leq y} H(t) \\ &\geq H(y) - \frac{A_2}{\log(x)} (J + \epsilon) \\ &\geq (J - \epsilon) - \frac{A_2}{\log(x)} \cdot J - \frac{A_2}{\log(x)} \cdot \epsilon \\ &= J \left(1 - \frac{A_2}{\log(x)} \right) - \epsilon \left(1 + \frac{A_2}{\log(x)} \right) \end{aligned}$$

Now take $0 < \epsilon < J/\log(x)$, so that with identity (3.2.49) we have

$$\frac{G(x)}{\log^\kappa(x)} = H(x) \geq J \left(1 + O \left(\frac{A_2}{\log(x)} \right) \right), \quad (3.2.52)$$

initially for sufficiently large x , and so for $x \geq 2$.

We need a partial identification of the constant J . For this purpose we have the following Dirichlet series:

$$Z_g(s) = \sum_{n=1}^{\infty} \frac{g(n)}{n^s} = \prod_p \left(1 + \frac{g(p)}{p^s} \right),$$

since g is supported on squarefree numbers. Recalling integration by parts and Lemma 2.2 with $C = G$ and $f(t) = 1/t^s$, when $s > 0$ we have

$$\begin{aligned} Z_g(s) &= \sum_{n=1}^{\infty} \frac{g(n)}{n^s} = \int_1^{\infty} \frac{1}{t^s} d(G(t)) \\ &= \left[\frac{G(t)}{t^s} \right]_1^{\infty} - \int_1^{\infty} G(t) d(t^{-s}) \end{aligned}$$

Now $G(1 - \epsilon) = 0$ for every $\epsilon > 0$ and $\lim_{t \rightarrow \infty} G(t)/t^s = 0$ so we just get

$$Z_g(s) = - \int_1^{\infty} G(t) d(t^{-s})$$

Appealing to inequality (3.2.52) and then using the substitution $u = \log(t)$, we get

$$\begin{aligned} Z_g(s) &= - \int_1^\infty G(t) d(t^{-s}) \\ &\leq - \int_1^\infty (J + o(1)) \log^\kappa(t) d(t^{-s}) \\ &= s \int_0^\infty (J + o(1)) u^\kappa e^{-su} du \\ &\sim \frac{J\Gamma(\kappa+1)}{s^\kappa} \text{ as } s \rightarrow 0^+, \end{aligned}$$

where the quantities $o(1)$ are as $t \rightarrow \infty$ and $u \rightarrow \infty$. The last approximation above follows from the definition of the Gamma function:

$$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx,$$

with the substitutions $r = \kappa + 1$ and $x = su$ (so $dx = s du$).

We now recall some facts about the Riemann ζ function:

- (1) $\frac{1}{\zeta(s+1)} = \prod_p \left(1 - \frac{1}{p^{s+1}}\right)$
- (2) $\zeta(s+1)$ has a simple pole at $s = 0$,
and $\zeta(s+1)s \rightarrow 1$ as $s \rightarrow 0$.

Using these facts along with the inequality we have for $Z_g(s)$, we get the following:

$$\limsup_{s \rightarrow 0^+} \prod_p \left(1 - \frac{1}{p^{s+1}}\right)^\kappa \left(1 + \frac{g(p)}{p^s}\right) = \limsup_{s \rightarrow 0^+} \frac{Z_g(s)}{\zeta^\kappa(s+1)} \leq J \cdot \Gamma(\kappa+1). \quad (3.2.53)$$

Our choice (3.2.43) of $\rho(p)$ when $p \geq z$ ensures that the product on the left converges when $s = 0$, being in fact finite for this s . Observe when $z \leq x < y$ the contribution to (3.2.53) from large p is

$$\begin{aligned} &\exp \sum_{x \leq p < y} \left(\kappa \log \left(1 - \frac{1}{p^{s+1}}\right) + \log \left(1 + \frac{g(p)}{p^s}\right) \right) \\ &= \exp \sum_{x \leq p < y} \left(\frac{g(p) - \kappa/p}{p^s} + O \left(g^2(p) + \frac{1}{p^2} \right) \right). \end{aligned}$$

Because of equation (3.2.44) the summand is $O(1/p^2)$ when $s \geq 0$, so the sum converges uniformly to a sum continuous in $s \geq 0$. That is, the product in (3.2.53) is continuous in $s \geq 0$. It is now easy to infer

$$J \cdot \Gamma(\kappa + 1) \geq \prod_{p < z} \left(1 - \frac{1}{p}\right)^\kappa \left(1 - \frac{\rho(p)}{p}\right)^{-1}.$$

Theorem 3.2.2 now follows in the form using identity (3.2.40) on taking $x = z$ in inequality (3.2.52).

Q.E.D.

3.2.3 The Error Term

The following theorem provides an upper bound estimate for the Error term, $E(D, P(z))$, in Theorem 3.1.1.

Theorem 3.2.3 *Suppose that equation (3.1.1) holds with*

$$|r_A(d)| \leq \rho(d), \quad \rho(p) \geq 1 \text{ when } p \mid P, \quad (3.2.54)$$

and that ρ has sifting density κ . Then

$$E(D, P(z)) \ll D \log^{2\kappa}(z).$$

Proof:

Corollary 3.1.2.1 and (3.2.54) give

$$E(D, P) \leq \sum_{\substack{d_1 < \sqrt{D} \\ d_1 \mid P}} \sum_{\substack{d_2 < \sqrt{D} \\ d_2 \mid P}} \frac{\rho(d_1)\rho(d_2)}{\rho((d_1, d_2))} \leq \left(\sum_{\substack{d < \sqrt{D} \\ d \mid P}} \rho(d) \right)^2.$$

Then

$$\sum_{\substack{d < \sqrt{D} \\ d \mid P(z)}} \rho(d) \leq \sqrt{D} \sum_{d \mid P(z)} \frac{\rho(d)}{d} \leq \sqrt{D} \prod_{p < z} \left(1 + \frac{\rho(p)}{p}\right) \leq \frac{\sqrt{D}}{V(P(z))}.$$

Here $V(P(z))^{-1} \ll K \log^\kappa(z)$ because ρ has sifting density κ , so the final result follows.

Q.E.D.

Chapter 4

Twin Primes

Recall that a twin prime is a pair of prime numbers that differ by 2. They can be represented as $(p, p + 2)$, where both p and $p + 2$ are prime. We will now take a look at two methods for finding twin primes and then explore how Selberg's results apply to these methods.

4.1 Sifting for Twin Primes

As mentioned in the introduction, after conducting the sieve of Eratosthenes to the set of numbers from 1 to N , one can then identify those primes that differ by 2, i.e. twin primes. There also exists a different sifting approach that yields all twin primes except $(3, 5)$. We provide examples of both approaches and also prove that the alternate sifting method works.

4.1.1 An example of the Sieve of Eratosthenes

For $N = 36$, we cancel all proper multiples of primes less than or equal to $\sqrt{N} = 6$ (i.e., 2, 3, 5) and obtain 1 and all primes p where $1 \leq p \leq N$. We then observe that $(3, 5)$, $(5, 7)$, $(11, 13)$, $(17, 19)$, $(29, 31)$ are twin primes.

Note: In the following display of numbers and in those to follow throughout the paper, there is no special significance to the direction of the strokes. What is of some importance is that some numbers are cancelled more than once.

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36

4.1.2 An Alternate Sieve for Twin Primes

It turns out that one can obtain the twin primes (except (3, 5)) without the other primes by using a slightly different sifting approach to that of Eratosthenes. This new sieve actually yields twin primes in an indirect way in that it gives the numbers k , where both $6k - 1$ and $6k + 1$ are prime; i.e. a twin prime. The sieve also makes use of the observation that if $6k - 1$ (or $6k + 1$) is not prime, then it has a prime factor less than or equal to $\sqrt{6k - 1}$ (or $\sqrt{6k + 1}$).

We start with all the integers from 1 to N . Since N is the largest possible “ k ”, the largest possible twin prime is $(6N - 1, 6N + 1)$. As such, we need only worry about prime factors less than or equal to $\sqrt{6N + 1}$, except for 2 and 3 since $2, 3 \nmid 6k - 1, 6k + 1$ for any k . For each prime p , where $5 \leq p \leq \sqrt{6N + 1}$, we define a_p, b_p by

$$\begin{aligned} 6a_p - 1 &\equiv 0 \pmod{p}, \quad 0 \leq a_p < p \\ 6b_p + 1 &\equiv 0 \pmod{p}, \quad 0 \leq b_p < p. \end{aligned}$$

We observe that $b_p \equiv -a_p \pmod{p}$. Then starting with $\frac{p^2-1}{6}$, we cancel all numbers that are $\pm a_p \pmod{p}$. All remaining numbers, k , yield a twin prime: $(6k - 1, 6k + 1)$. Also observe that the process only makes sense for $N \geq 4$. For $N = 1, 2, 3$, there is nothing to do since each of 1, 2, 3 yields a twin prime.

An example of the Alternate Sieve

For $N = 48$, we worry about prime factors p where $5 \leq p \leq \sqrt{6N + 1} = \sqrt{289} = 17$, i.e., 5, 7, 11, 13 and 17.

For 5, $a_p = 1$, and cancellation starts with $\frac{5^2-1}{6} = 4$.

For 7, $a_p = 1$, and cancellation starts with $\frac{7^2-1}{6} = 8$.

For 11, $a_p = 2$, and cancellation starts with $\frac{11^2-1}{6} = 20$.

For 13, $a_p = 2$, and cancellation starts with $\frac{13^2-1}{6} = 28$.

For 17, $a_p = 3$, and cancellation starts with $\frac{17^2-1}{6} = 48$.

We therefore cancel everything starting with 4 that is $\pm 1 \pmod{5}$, then cancel everything starting with 8 that is $\pm 1 \pmod{7}$, then cancel everything starting with 20 that is $\pm 2 \pmod{11}$, then cancel everything starting with

28 that is $\pm 2 \pmod{13}$, and finally, cancel everything starting with 48 that is $\pm 3 \pmod{17}$. We observe that some numbers are cancelled more than once.

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48		

One can easily verify that the remaining numbers, k , all yield twin primes $(6k - 1, 6k + 1)$.

4.1.3 Proving the Alternate Sieve works

The following theorem states that the Alternate Sieve does indeed yield twin primes. By observing that all primes $p \geq 5$ are either of the form $6m - 1$ or $6m + 1$, we see that all twin primes, except $(3, 5)$, are of the form $(6k - 1, 6k + 1)$. Therefore, this sieve gives us all twin primes, except $(3, 5)$.

Theorem 4.1.1 For every prime $p \geq 5$, define a_p, b_p by

$$\begin{aligned} 6a_p - 1 &\equiv 0 \pmod{p}, \quad 0 \leq a_p < p \\ 6b_p + 1 &\equiv 0 \pmod{p}, \quad 0 \leq b_p < p. \end{aligned}$$

Let

$$\begin{aligned} K &= \{k \mid k \in \mathbb{N} \text{ and } k \not\equiv a_p \pmod{p}, \forall p \text{ prime}, 5 \leq p \leq \sqrt{6k - 1} \\ &\quad (\text{when this makes sense})\} \\ &\cap \{k \mid k \in \mathbb{N} \text{ and } k \not\equiv b_p \pmod{p}, \forall p \text{ prime}, 5 \leq p \leq \sqrt{6k + 1} \\ &\quad (\text{when this makes sense})\} \end{aligned}$$

Then $k \in K$ if and only if $6k - 1, 6k + 1$ are both prime. That is, $(6k - 1, 6k + 1)$ is a twin prime.

Proof:

“ \Leftarrow ” (Contrapositive)

Suppose $k \notin K$. Then either

$$\begin{aligned} k &\equiv a_p \pmod{p}, \text{ for some prime } p, \quad 5 \leq p \leq \sqrt{6k - 1} < 6k - 1 \\ \text{or } k &\equiv b_p \pmod{p}, \text{ for some prime } p, \quad 5 \leq p \leq \sqrt{6k + 1} < 6k + 1 \end{aligned}$$

Therefore either

$$\begin{aligned} 6k-1 &\equiv 6a_p-1 \equiv 0 \pmod{p}, \quad p < 6k-1 \\ \text{or } 6k+1 &\equiv 6b_p+1 \equiv 0 \pmod{p}, \quad p < 6k+1 \end{aligned}$$

So one of $6k-1$, $6k+1$ is composite. That is, $(6k-1, 6k+1)$ is not a twin prime.

" \Rightarrow "

Suppose $k \in K$. So

$$\begin{aligned} k &\not\equiv a_p \pmod{p}, \quad \forall p \text{ prime}, 5 \leq p \leq \sqrt{6k-1} \\ \text{and } k &\not\equiv b_p \pmod{p}, \quad \forall p \text{ prime}, 5 \leq p \leq \sqrt{6k+1}. \end{aligned}$$

It follows that

$$\begin{aligned} 6k-1 &\not\equiv 6a_p-1 \equiv 0 \pmod{p}, \quad \forall p \text{ prime}, 5 \leq p \leq \sqrt{6k-1} \\ \text{and } 6k+1 &\not\equiv 6b_p+1 \equiv 0 \pmod{p}, \quad \forall p \text{ prime}, 5 \leq p \leq \sqrt{6k+1}. \end{aligned}$$

By observing 2, $3 \nmid 6k-1$, $6k+1$, we have

$$\begin{aligned} p &\nmid 6k-1 \quad \forall p \text{ prime}, p \leq \sqrt{6k-1} \\ \text{and } p &\nmid 6k+1 \quad \forall p \text{ prime}, p \leq \sqrt{6k+1}. \end{aligned}$$

Therefore $6k-1$, $6k+1$ are both prime. That is, $(6k-1, 6k+1)$ is a twin prime.

Additional Remarks:

1. If $k \not\equiv a_p \pmod{p}$, $\forall p$ prime, $5 \leq p \leq \sqrt{6k-1}$, thereby indicating $6k-1$ prime, then it is actually true that $k \not\equiv a_p \pmod{p}$, $\forall p$ prime, $p \geq 5$, except for $p = 6k-1$ itself.
2. If $k \not\equiv b_p \pmod{p}$, $\forall p$ prime, $5 \leq p \leq \sqrt{6k+1}$, thereby indicating $6k+1$ prime, then it is actually true that $k \not\equiv b_p \pmod{p}$, $\forall p$ prime, $p \geq 5$, except for $p = 6k+1$ itself.
3. We mentioned earlier that to check the primality of a number, n , we only had to be concerned about the primes $p \leq \sqrt{n}$. We now observe

that for every prime, p , the first composite number that is divisibly by p and has no prime factors $q < p$ is p^2 . This means that all possible $6k \pm 1$ that are divisible by p and not divisible by any prime $q < p$ must be greater than or equal to p^2 . This means that cancellation of numbers, k , where $k \equiv a_p \pmod{p}$, starts at $\lfloor \frac{p^2+1}{6} \rfloor$. Likewise, the cancellation of numbers, k , where $k \equiv b_p \pmod{p}$, starts at $\lfloor \frac{p^2-1}{6} \rfloor$. It is thus ensured that for all remaining numbers, k , both $6k - 1 \not\equiv 0 \pmod{p}$ and $6k + 1 \not\equiv 0 \pmod{p}$. We will refer to $\lfloor \frac{p^2-1}{6} \rfloor$ and $\lfloor \frac{p^2+1}{6} \rfloor$ as the starting points of cancellation and they are actually equal since for every $p \geq 5$, one has $p^2 \equiv 1 \pmod{6}$. We will see later on that for a more general prime pair, the starting points of cancellation may not necessarily coincide, and there could be complications if one decided to use only the lower cancellation point.

4. We observe that if $p \leq \sqrt{6k-1}$ ($\sqrt{6k+1}$), then $k \geq \lfloor \frac{p^2-1}{6} \rfloor$ ($\lfloor \frac{p^2+1}{6} \rfloor$), which further justifies the starting points of cancellation for each prime $p \geq 5$ when the sieve is performed.

Q.E.D.

4.2 Applying Selberg's Results to Twin Primes

4.2.1 The Eratosthenes Approach

For the next theorem we apply Selberg's method to the numbers $n(n+2)$. Let $\pi_2(X)$ denote the number of n with $1 \leq n < X$ such that both n and $n+2$ are prime.

Theorem 4.2.1 *Let*

$$C = \prod_{p>2} \frac{p(p-2)}{(p-1)^2}, \quad (4.2.1)$$

where the product is over all odd primes p . Then

$$\pi_2(X) \leq \frac{16CX}{\log^2 X} \left(1 + O\left(\frac{\log \log X}{\log X}\right) \right)$$

Proof:

Take $f(n) = n(n+2)$. Then the function ρ in this sifting situation is given by

$$\rho(p) = 2 \text{ if } p > 2, \quad \rho(2) = 1.$$

Thus $\rho(p) \leq 2$ for all p , so (as in Lemma 2.1) ρ has sifting density $\kappa = 2$, and (3.2.38) holds with $\eta(z, w) = O(1)$. Then Theorem 3.2.2 gives

$$\begin{aligned} G(\sqrt{D}) &\geq \frac{1}{2!} \left(\frac{e^{-2\gamma}}{V(P(\sqrt{D}))} \right) \left(1 + O\left(\frac{1}{\log D}\right) \right) \\ &= \frac{1}{2!} A_D \log^2 \sqrt{D} \left(1 + O\left(\frac{1}{\log D}\right) \right), \end{aligned}$$

where

$$\begin{aligned} A_D &= \frac{1}{2} \prod_{2 < p < \sqrt{D}} \left(1 - \frac{1}{p} \right)^2 \left(1 - \frac{2}{p} \right)^{-1} \\ &= \frac{1}{2} \prod_{2 < p < \sqrt{D}} \frac{(p-1)^2}{p(p-2)} \\ &= \frac{1}{2C} \prod_{p \geq \sqrt{D}} \frac{p(p-2)}{(p-1)^2} \quad (\text{C as in (4.2.1)}) \\ &= \frac{1}{2C} \left(1 + O\left(\frac{1}{\sqrt{D}}\right) \right) \quad (\text{see (2.31) with } w = \sqrt{D}). \end{aligned}$$

Applying Theorems 3.1.1 and 3.2.3, we now have

$$\pi_2(X) \leq S(A, P(\sqrt{D})) + \pi(\sqrt{D}) \leq \frac{2X}{A_D \log^2 \sqrt{D}} + O(D \log^{2\kappa} D).$$

Our result follows as stated on choosing $D = X/(\log X)^{2\kappa+3}$, where $\kappa = 2$.

Q.E.D.

4.2.2 The Alternate Approach

In the alternate approach presented for sifting the twin primes, if we sift the numbers n where $1 \leq n < X/6$, to retain only those for which both

$6n - 1, 6n + 1$ are prime, we obtain all the twin primes less than or equal to X . Correspondingly with the standard approach, if the numbers $N = 6n - 1$ are sifted where $1 \leq N < X$, to retain those where both $N = 6n - 1, N + 2 = 6n + 1$ are prime, we also get all the twin primes less than or equal to X . Therefore, in applying the Selberg sieve to the alternate approach, we must sift the numbers $(6n - 1)(6n + 1)$, just as the previous theorem sifted the numbers $N(N + 2)$.

If we let $\pi'_2(X)$ denote the number of n with $1 \leq n < X/6$ such that both $6n - 1$ and $6n + 1$ are prime, we observe that $\pi'_2(X) = \pi_2(X) - 1$; $(3, 5)$ is not counted by $\pi'_2(X)$. Hence we would expect $\pi'_2(X)$ to have a similar upper bound as $\pi_2(X)$ upon applying the Selberg sieve on the numbers $(6n - 1)(6n + 1)$. This is indeed the case as demonstrated by the next theorem, and we can actually achieve the same upper bound.

Theorem 4.2.2 *Let*

$$C = \prod_{p>2} \frac{p(p-2)}{(p-1)^2}, \quad (4.2.2)$$

where the product is over all odd primes p . Then

$$\pi'_2(X) \leq \frac{16CX}{\log^2 X} \left(1 + O\left(\frac{\log \log X}{\log X}\right) \right)$$

Proof:

Take $f(n) = (6n - 1)(6n + 1)$. Then the function ρ in this sifting situation is given by

$$\rho(p) = 2 \text{ if } p > 3, \quad \rho(2) = 0, \quad \rho(3) = 0.$$

Thus $\rho(p) \leq 2$ for all p , so (as in Lemma 2.1) ρ has sifting density $\kappa = 2$, and (3.2.38) holds with $\eta(z, w) = O(1)$. Then Theorem 3.2.2 gives

$$\begin{aligned} G(\sqrt{D}) &\geq \frac{1}{2!} \left(\frac{e^{-2\gamma}}{V(P(\sqrt{D}))} \right) \left(1 + O\left(\frac{1}{\log D}\right) \right) \\ &= \frac{1}{2!} A'_D \log^2 \sqrt{D} \left(1 + O\left(\frac{1}{\log D}\right) \right), \end{aligned}$$

where

$$\begin{aligned}
 A'_D &= \frac{1}{9} \prod_{3 < p < \sqrt{D}} \left(1 - \frac{1}{p}\right)^2 \left(1 - \frac{2}{p}\right)^{-1} \\
 &= \frac{1}{9} \prod_{3 < p < \sqrt{D}} \frac{(p-1)^2}{p(p-2)} \\
 &= \frac{1}{12C} \prod_{p \geq \sqrt{D}} \frac{p(p-2)}{(p-1)^2} \quad (\text{C as in (4.2.2)}) \\
 &= \frac{1}{12C} \left(1 + O\left(\frac{1}{\sqrt{D}}\right)\right) \quad (\text{see (2.31) with } w = \sqrt{D}).
 \end{aligned}$$

Applying Theorems 3.1.1 and 3.2.3, we have

$$\pi'_2(X) \leq S(A, P(\sqrt{D})) + \pi(\sqrt{D}) \leq \frac{2 \cdot \frac{X}{6}}{A'_D \log^2 \sqrt{D}} + O(D \log^{2\kappa} D).$$

Our result follows as stated on choosing $D = X/(\log X)^{2\kappa+3}$, where $\kappa = 2$.

Q.E.D.

As can be seen by comparing the results of the last two theorems, the alternate approach to sifting the twin primes leads to the same upper bound as the standard (Eratosthenes) approach.

4.3 Sum of the Reciprocals of Twin Primes

It has been known for quite some time that the sum of the reciprocals of the prime numbers is divergent. That is,

$$\sum_{i=1}^n \frac{1}{p_i} \rightarrow \infty \text{ as } n \rightarrow \infty,$$

where p_i is the i -th prime number. The proof of this fact can be found in many texts on elementary number theory (see [6, Theorem 1.19]). In particular, this result implies that there are infinitely many prime numbers (a fact earlier known to Euclid (see [6, Theorem 1.17])). So a good question to ask is "Can this strategy be used to show that there are infinitely many twin primes?".

Unfortunately, the answer is “No”, leaving open to this day the question as to whether or not there are infinitely many twin primes. Viggo Brun was the first to show that the sum of the reciprocals of twin primes is convergent (or finite if there are only finitely many twin primes) [2]. That is,

$$\sum_i \left(\frac{1}{p_i} + \frac{1}{p_i + 2} \right) \leq M,$$

for some constant M , where p_i is the i -th twin prime, counting the twin primes by their first members. We present a proof of this result using Partial Summation and the common result of the previous two theorems. In the proof, we will also provide an explicit constant M .

Theorem 4.3.1 *Let p_1, p_2, \dots be the sequence of prime numbers p such that $p + 2$ is also prime. Then*

$$\sum_i \left(\frac{1}{p_i} + \frac{1}{p_i + 2} \right) \leq M,$$

where M is a constant.

Proof:

We wish to use partial summation, in particular Lemma 2.2 which makes use of integration by parts (2.32). Let

$$a_n = \begin{cases} 1 & \text{if } n, n+2 \text{ are both prime,} \\ 0 & \text{otherwise.} \end{cases}$$

and

$$f(n) = \frac{1}{n} \quad (f' \text{ is continuous on } [2, \infty)).$$

Then let

$$A(x) = \sum_{2 \leq n < x} a_n = \pi_2(x) = \pi_2'(x),$$

and

$$B(x) = \sum_{2 \leq n < x} \frac{a_n}{n} = \sum_{\substack{2 \leq n < x \\ n, n+2 \text{ prime}}} \frac{1}{n}.$$

Note here that $B(x)$ is the sum of the reciprocals of only the first members of twin primes. This will not be a problem, however, because if $B(x)$ is shown to be bounded by a constant $\frac{M}{2}$, then the sum of all reciprocals would be bounded by M since $\frac{1}{n+2} < \frac{1}{n}$. Proceeding now with use of Lemma 2.2, we have

$$B(x) = \int_2^x f(t) d(A(t)),$$

and making use of integration by parts (2.32), this becomes

$$\begin{aligned} B(x) &= [A(t)f(t)]_2^x - \int_2^x A(t)f'(t)dt \\ &= A(x)f(x) - A(2)f(2) - \int_2^x A(t)f'(t)dt \\ &= A(x)f(x) - \int_2^x A(t)f'(t)dt \quad (\text{since } A(2) = 0) \\ &= A(x)\left(\frac{1}{x}\right) - \int_2^x A(t)\left(\frac{-1}{t^2}\right)dt \\ &= A(x)\left(\frac{1}{x}\right) + \int_2^x A(t)\left(\frac{1}{t^2}\right)dt \\ &\leq 16C \left[\left(\frac{x}{\log^2(x)}\right) \left(\frac{1}{x}\right) + \int_2^x \left(\frac{t}{\log^2(t)}\right) \left(\frac{1}{t^2}\right)dt \right] \\ &\quad (\text{by Theorems 4.2.1 and 4.2.2}), \end{aligned}$$

where C is the constant from Theorems 4.2.1 and 4.2.2. Continuing the calculation, we get

$$\begin{aligned} B(x) &\leq 16C \left[\frac{1}{\log^2(x)} + \int_2^x \frac{1}{t \log^2(t)} dt \right] \\ &= 16C \left[\frac{1}{\log^2(x)} - \left[\frac{1}{\log(t)} \right]_2^x \right] \\ &= 16C \left[\frac{1}{\log^2(x)} - \frac{1}{\log(x)} + \frac{1}{\log(2)} \right] \\ &\rightarrow \frac{16C}{\log(2)} \quad (\text{as } x \rightarrow \infty). \end{aligned}$$

From this it follows that

$$\sum_i \left(\frac{1}{p_i} + \frac{1}{p_i + 2} \right) \leq M,$$

with

$$M = \frac{32C}{\log(2)}.$$

Q.E.D.

Chapter 5

General Prime Pairs (G.P.P.)

The last section gave some good results concerning twin primes; prime pairs of the form $(p, p + 2)$, where p and $p + 2$ are both prime. Similar results can also be obtained for a more general prime pair, $(p, p + 2B)$, where $B \geq 1$ is an integer constant. We shall apply Selberg's results to both the traditional (Eratosthenes) method for finding general prime pairs, and to a general version of the alternate method that was used to find twin primes. As such, we first proceed to show that the alternate sifting method can be generalized to find general prime pairs, $(p, p + 2B)$.

5.1 Sifting General Prime Pairs

Recall in the introduction that after conducting the sieve of Eratosthenes to the set of numbers from 1 to N , one can then identify those primes that differ by 2, i.e., twin primes. Similarly, one could simply pick out those primes that differ by $2B$, i.e., G.P.P.s. We provide a simple example of the Eratosthenes approach and then demonstrate the generalized alternate sifting method. Also, we provide proof that the general version of the alternate sieve works.

5.1.1 An example of the Sieve of Eratosthenes

Using the same example from before, for $N = 36$, we cancel all proper multiples of primes less than or equal to $\sqrt{N} = 6$ (i.e., 2, 3, 5) and obtain 1 and all primes p where $1 \leq p \leq N$. We then observe that for $B = 4$, $(3, 11)$, $(5, 13)$, $(11, 19)$, $(23, 31)$ are G.P.P.s of the form $(p, p + 2B)$.

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30
31	32	33	34	35	36

5.1.2 The Alternate Sieve for General Prime Pairs

As for twin primes, one can obtain the general prime pairs (except possibly $(3, 3 + 2B)$) without the other primes by using a modified version of the previously described alternate approach. As before, this new sieve will actually yield general prime pairs in an indirect way. Since all primes, other than 2 and 3, are either of the form $6k - 1$ or $6k + 1$, we see that general prime pairs, $(p, p + 2B)$, are either of the form $(6k - 1, 6k + 2B - 1)$, or $(6k + 1, 6k + 2B + 1)$. Furthermore, we see that the existence of these forms depends on the residue class, modulo 3, of B . That is, general prime pairs are of the form $(6k + t, 6k + 2B + t)$ where

$$t = \begin{cases} -1 & \text{if } B \equiv 1 \pmod{3} \\ 1 & \text{if } B \equiv 2 \pmod{3} \\ -1 \text{ or } 1 & \text{if } B \equiv 0 \pmod{3} \end{cases}$$

Convention Note:

It will be convenient for us to always say that general prime pairs are of the form $(6k + t, 6k + 2B + t)$, where it is to be understood that t has been predetermined by B , and should $B \equiv 0 \pmod{3}$, we will say that general prime pairs have either the form $(6k + t, 6k + 2B + t)$ or $(6k - t, 6k + 2B - t)$. Also, in several situations to follow, we will use the phrase “(when applicable)” in connection with the general prime pairs of the form $(6k - t, 6k + 2B - t)$, to indicate the possibility of dealing with both types of general prime pairs should $B \equiv 0 \pmod{3}$.

We shall see that the general alternate sieve is indirect in that it gives the numbers k , where both $6k + t, 6k + 2B + t$ are prime, and/or possibly both $6k - t, 6k + 2B - t$ are prime (when applicable); i.e., we get general prime pairs. The sieve also makes use of the observation that if $6k \pm t$ (or $6k + 2B \pm t$) is not prime, then it has a prime factor less than or equal to

$\sqrt{6k \pm t}$ (or $\sqrt{6k + 2B \pm t}$).

If $B \equiv 0 \pmod{3}$, we actually conduct two separate applications of the alternate sieve, once for $t = 1$, and once for $t = -1$. Afterward, we combine the numbers, k , retrieved from both applications, making careful note of any repeated elements. We start with all the integers from 1 to N . Since N is the largest possible " k ", the largest possible general prime pair is $(6N + t, 6N + 2B + t)$. As such, we need only worry about prime factors less than or equal to $\sqrt{6N + 2B + t}$, except for 2 and 3 since $2, 3 \nmid 6k + t, 6k + 2B + t$ for any k (recall that t is predetermined specifically so that $3 \nmid 6k + t, 6k + 2B + t$). For each prime p , where $5 \leq p \leq \sqrt{6N + 2B + t}$, we define a_p, b_p by

$$\begin{aligned} 6a_p + t &\equiv 0 \pmod{p}, \quad 0 \leq a_p < p \\ 6b_p + 2B + t &\equiv 0 \pmod{p}, \quad 0 \leq b_p < p. \end{aligned}$$

Then starting with $\lfloor \frac{p^2 - t}{6} \rfloor$, we cancel all numbers that are $a_p \pmod{p}$, and starting with $\lfloor \frac{p^2 - 2B - t}{6} \rfloor$, we cancel all numbers that are $b_p \pmod{p}$. All remaining numbers, k , yield a general prime pair $(6k + t, 6k + 2B + t)$.

An example of the General Alternate Sieve

We Consider the case $B = 3$. Here, we have both types of general prime pairs, $(6k - 1, 6k + 5)$, and $(6k + 1, 6k + 7)$. We conduct the sieve for each type separately and respectively, and then combine the results.

Type 1: $(6k - 1, 6k + 5)$

For $N = 18$, we worry about prime factors p where $5 \leq p \leq \sqrt{6N + 5} = \sqrt{113} < 11$, i.e., 5 and 7.

$$\begin{aligned} \text{For } 5: \quad a_p &= 1, \text{ and cancellation starts with } \lfloor \frac{5^2 + 1}{6} \rfloor = 4. \\ b_p &= 0, \text{ and cancellation starts with } \lfloor \frac{5^2 - 5}{6} \rfloor = 3. \\ \text{For } 7: \quad a_p &= 6, \text{ and cancellation starts with } \lfloor \frac{7^2 + 1}{6} \rfloor = 8. \\ b_p &= 5, \text{ and cancellation starts with } \lfloor \frac{7^2 - 5}{6} \rfloor = 7. \end{aligned}$$

We therefore cancel everything starting with 4 that is $1 \pmod{5}$, and everything starting with 3 that is $0 \pmod{5}$. Then cancel everything starting with 8 that is $6 \pmod{7}$, and everything starting with 7 that is $5 \pmod{7}$.

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18		

It is easily verified that all the remaining numbers, k , yield general prime pairs, $(6k - 1, 6k + 5)$.

Type 2: $(6k + 1, 6k + 7)$

For $N = 18$, we worry about prime factors p where $5 \leq p \leq \sqrt{6N + 7} = \sqrt{115} < 11$, i.e., 5 and 7.

For 5: $a_p = 4$, and cancellation starts with $\lfloor \frac{5^2-1}{6} \rfloor = 4$.
 $b_p = 3$, and cancellation starts with $\lfloor \frac{5^2-7}{6} \rfloor = 3$.
 For 7: $a_p = 1$, and cancellation starts with $\lfloor \frac{7^2-1}{6} \rfloor = 8$.
 $b_p = 0$, and cancellation starts with $\lfloor \frac{7^2-7}{6} \rfloor = 7$.

We therefore cancel everything starting with 4 that is 4 (mod 5), and everything starting with 3 that is 3 (mod 5). Then cancel everything starting with 8 that is 1 (mod 7), and everything starting with 7 that is 0 (mod 7).

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18		

It is easily verified that all the remaining numbers, k , yield general prime pairs, $(6k + 1, 6k + 7)$. Combining together all the numbers, k , with multiplicity if needed, that we got from the double application of the sieve, we get all general prime pairs, $(p, p + 6)$, that are less than or equal to $6N + 7 = 6(18) + 7 = 115$.

5.1.3 Proving the General Alternate Sieve works

The following theorem states that the General Alternate Sieve does indeed yield general prime pairs. We have already seen that all general prime pairs, except possibly $(3, 3 + 2B)$, are of the form $(6k + t, 6k + 2B + t)$ and/or $(6k - t, 6k + 2B - t)$ (when applicable). Therefore, this sieve, or a double application of it (when applicable), gives us all general prime pairs, except possibly $(3, 3 + 2B)$.

Theorem 5.1.1 Let $B \geq 1$ be an integer constant and,

$$\text{Fix } t = \begin{cases} 1 & \text{if } 2B - 1 \equiv 0 \pmod{3} \text{ (i.e., } B \equiv 2 \pmod{3}) \\ -1 & \text{if } 2B + 1 \equiv 0 \pmod{3} \text{ (i.e., } B \equiv 1 \pmod{3}) \\ 1 \text{ or } -1 & \text{if } B \equiv 0 \pmod{3} . \end{cases}$$

Also, for every prime $p \geq 5$, define a_p, b_p by

$$\begin{aligned} 6a_p + t &\equiv 0 \pmod{p}, \quad 0 \leq a_p < p \\ 6b_p + 2B + t &\equiv 0 \pmod{p}, \quad 0 \leq b_p < p . \end{aligned}$$

Let

$$\begin{aligned} K &= \{k \mid k \in \mathbb{N} \text{ and } k \not\equiv a_p \pmod{p}, \forall p \text{ prime}, 5 \leq p \leq \sqrt{6k+t} \\ &\quad \text{(when this makes sense)}\} \\ &\cap \{k \mid k \in \mathbb{N} \text{ and } k \not\equiv b_p \pmod{p}, \forall p \text{ prime}, 5 \leq p \leq \sqrt{6k+2B+t} \\ &\quad \text{(when this makes sense)}\} \end{aligned}$$

Then $k \in K$ if and only if $6k+t, 6k+2B+t$ are both prime. That is, $(6k+t, 6k+2B+t)$ is a G.P.P.

Proof:

“ \Leftarrow ” (Contrapositive)

Suppose $k \notin K$. Then either

$$\begin{aligned} k &\equiv a_p \pmod{p}, \text{ for a prime } p, \quad 5 \leq p \leq \sqrt{6k+t} < 6k+t \\ \text{or } k &\equiv b_p \pmod{p}, \text{ for a prime } p, \quad 5 \leq p \leq \sqrt{6k+2B+t} < 6k+2B+t \end{aligned}$$

Therefore either

$$\begin{aligned} 6k+t &\equiv 6a_p+t \equiv 0 \pmod{p}, \quad p < 6k+t \\ \text{or } 6k+2B+t &\equiv 6b_p+2B+t \equiv 0 \pmod{p}, \quad p < 6k+2B+t \end{aligned}$$

So one of $6k+t, 6k+2B+t$ is composite. That is, $(6k+t, 6k+2B+t)$ is not a G.P.P.

" \Rightarrow "

Suppose $k \in K$. So

$$\begin{aligned} k &\not\equiv a_p \pmod{p}, \forall p \text{ prime}, 5 \leq p \leq \sqrt{6k+t} \\ \text{and } k &\not\equiv b_p \pmod{p}, \forall p \text{ prime}, 5 \leq p \leq \sqrt{6k+2B+t}. \end{aligned}$$

It follows that

$$\begin{aligned} 6k+t &\not\equiv 6a_p+t \equiv 0 \pmod{p}, \\ &\forall p \text{ prime}, 5 \leq p \leq \sqrt{6k+t} \\ \text{and } 6k+2B+t &\not\equiv 6b_p+2B+t \equiv 0 \pmod{p}, \\ &\forall p \text{ prime}, 5 \leq p \leq \sqrt{6k+2B+t}. \end{aligned}$$

By observing 2, $3 \nmid 6k+t$, $6k+2B+t$, we have

$$\begin{aligned} p &\nmid 6k+t \quad \forall p \text{ prime}, p \leq \sqrt{6k+t} \\ \text{and } p &\nmid 6k+2B+t \quad \forall p \text{ prime}, p \leq \sqrt{6k+2B+t}. \end{aligned}$$

Therefore $6k+t$, $6k+2B+t$ are both prime. That is, $(6k+t, 6k+2B+t)$ is a G.P.P.

Additional Remarks:

1. If $k \not\equiv a_p \pmod{p}$, $\forall p$ prime, $5 \leq p \leq \sqrt{6k+t}$, thereby indicating $6k+t$ prime, then it is actually true that $k \not\equiv a_p \pmod{p}$, $\forall p$ prime, $p \geq 5$, except for $p = 6k+t$ itself.
2. If $k \not\equiv b_p \pmod{p}$, $\forall p$ prime, $5 \leq p \leq \sqrt{6k+2B+t}$, thereby indicating $6k+2B+t$ prime, then it is actually true that $k \not\equiv b_p \pmod{p}$, $\forall p$ prime, $p \geq 5$, except for $p = 6k+2B+t$ itself.
3. As previously mentioned, the first composite number that is divisible by a prime, p , and has no prime factors $q < p$ is p^2 . So all possible $6k+t$, $6k+2B+t$ that are divisible by p and not divisible by any prime $q < p$ must be greater than or equal to p^2 . This means that cancellation of numbers, k , where $k \equiv a_p \pmod{p}$, starts at $\lfloor \frac{p^2-t}{6} \rfloor$. Likewise, the cancellation of numbers, k , where $k \equiv b_p \pmod{p}$, starts at $\lfloor \frac{p^2-2B-t}{6} \rfloor$. It is thus ensured that for all remaining numbers, k , both $6k+t \not\equiv 0 \pmod{p}$ and $6k+2B+t \not\equiv 0 \pmod{p}$.

4. If $p \leq \sqrt{6k+t}$ ($\sqrt{6k+2B+t}$) then $k \geq \lfloor \frac{p^2-t}{6} \rfloor$ ($\lfloor \frac{p^2-2B-t}{6} \rfloor$), which further justifies the starting points of cancellation for each prime $p \geq 5$ when the sieve is performed.
5. When sifting $\{1, \dots, N\}$, one has sifting conditions for all primes $p \leq \sqrt{6N+2B+t}$. There could be some k where $6k+t$, $6k+2B+t$ are both prime and $6k+t \leq \sqrt{6N+2B+t}$. If one decided to only use the smaller starting cancellation points, $\lfloor \frac{p^2-2B-t}{6} \rfloor$, then for these primes $p = 6k+t$, one could potentially have $\lfloor \frac{(6k+t)^2-2B-t}{6} \rfloor \leq k$, equivalently $6k+t \leq \sqrt{6k+2B+t}$. This means that these values of k would get cancelled incorrectly. However, when B is small, as in the case for twin primes, one can safely use just the lower of the two starting points, due to Remark 1 and the fact that $\sqrt{6k+2B+t} < 6k+t$ when B is small. For larger B , one can still use just the lower of the two starting points, but some care must be taken to not cancel those values of k where $6k+t$ and $6k+2B+t$ are prime, and $6k+t \leq \sqrt{6k+2B+t}$.

Q.E.D.

5.2 Applying Selberg's Method

5.2.1 The Eratosthenes Approach

Recall that for the Eratosthenes approach to sifting twin primes, we applied Selberg's method to the numbers $n(n+2)$. Similarly, we now apply the method to the numbers $n(n+2B)$. We define $\pi_{2B}(X)$ to be the number of n where $1 \leq n < X$, such that both n and $n+2B$ are prime.

Theorem 5.2.1 *Let*

$$C = \prod_{p>2} \frac{p(p-2)}{(p-1)^2} \quad \text{and} \quad C_B = \prod_{\substack{p|B \\ p \neq 2}} \frac{p-1}{p-2}, \quad (5.2.1)$$

where the products are over odd primes p . Then

$$\pi_{2B}(X) \leq \frac{16CC_B X}{\log^2 X} \left(1 + O\left(\frac{\log \log X}{\log X}\right) \right)$$

Proof:

Take $f(n) = n(n + 2B)$. Then the function ρ in this sifting situation is given by

$$\rho(p) = \begin{cases} 2 & \text{if } p \nmid 2B \\ 1 & \text{if } p \mid 2B \end{cases}.$$

Thus $\rho(p) \leq 2$ for all p , so (as in Lemma 2.1) ρ has sifting density $\kappa = 2$, and (3.2.38) holds with $\eta(z, w) = O(1)$. Then Theorem 3.2.2 gives

$$\begin{aligned} G(\sqrt{D}) &\geq \frac{1}{2!} \left(\frac{e^{-2\gamma}}{V(P(\sqrt{D}))} \right) \left(1 + O\left(\frac{1}{\log D}\right) \right) \\ &= \frac{1}{2!} A_{DB} \log^2 \sqrt{D} \left(1 + O\left(\frac{1}{\log D}\right) \right), \end{aligned}$$

where

$$\begin{aligned} A_{DB} &= \frac{1}{2} \prod_{\substack{p \mid B \\ p \neq 2}} \left(1 - \frac{1}{p} \right) \prod_{\substack{p < \sqrt{D} \\ p \nmid B, p \neq 2}} \left(1 - \frac{1}{p} \right)^2 \left(1 - \frac{2}{p} \right)^{-1} \\ &= \frac{1}{2} \prod_{\substack{p \mid B \\ p \neq 2}} \left(1 - \frac{1}{p} \right) \prod_{\substack{p < \sqrt{D} \\ p \nmid B, p \neq 2}} \frac{(p-1)^2}{p(p-2)} \\ &= \frac{1}{2} \left(\prod_{\substack{p \mid B \\ p \neq 2}} \left(1 - \frac{1}{p} \right) \right) \left(\prod_{\substack{p \mid B \\ p \neq 2}} \frac{p(p-2)}{(p-1)^2} \right) \frac{1}{C} \prod_{p \geq \sqrt{D}} \frac{p(p-2)}{(p-1)^2} \\ &\quad (C \text{ as in (5.2.1)}) \\ &= \frac{1}{2} \left(\prod_{\substack{p \mid B \\ p \neq 2}} \left(\frac{p-1}{p} \right) \right) \left(\prod_{\substack{p \mid B \\ p \neq 2}} \frac{p(p-2)}{(p-1)^2} \right) \frac{1}{C} \left(1 + O\left(\frac{1}{\sqrt{D}}\right) \right) \\ &\quad (\text{see (2.31) with } w = \sqrt{D}) \\ &= \frac{1}{2} \frac{1}{C_B} \frac{1}{C} \left(1 + O\left(\frac{1}{\sqrt{D}}\right) \right) \quad (C_B \text{ as in (5.2.1)}). \end{aligned}$$

Applying Theorems 3.1.1 and 3.2.3, we have

$$\pi_{2B}(X) \leq S(A, P(\sqrt{D})) + \pi(\sqrt{D}) \leq \frac{2X}{A_{DB} \log^2 \sqrt{D}} + O(D \log^{2\kappa} D).$$

Our result follows as stated on choosing $D = X/(\log X)^{2\kappa+3}$, where $\kappa = 2$.

Q.E.D.

5.2.2 The Alternate Approach

In the alternate approach presented for sifting general prime pairs, for $1 \leq n < X/6$, if we sift the numbers n to retain those where both $6n+t$, $6n+2B+t$ are prime (t fixed as either 1 or -1 , and $B \geq 1$ is an integer constant), along with those where both $6n-t$, $6n+2B-t$ are prime (when applicable), we obtain all the general prime pairs less than or equal to X . Correspondingly with the standard approach, for $1 \leq N < X$, if the numbers $N = 6n+t$ are sifted to retain those where both $N = 6n+t$, $N+2B = 6n+2B+t$ are prime, along with those where both $N = 6n-t$, $N+2B = 6n+2B-t$ are prime (when applicable), we also get all the general prime pairs less than or equal to X . Therefore, in applying the Selberg sieve to the alternate approach, we must sift the numbers $(6n+t)(6n+2B+t)$, along with the numbers $(6n-t)(6n+2B-t)$ (when applicable), just as the previous theorem sifted the numbers $N(N+2B)$. This makes sense and agrees with the fact that the Selberg sieve was applied to the numbers $(6n-1)(6n+1)$ in the case of the alternate approach for sifting twin primes.

We let $\pi'_{2B}(X)$ denote the number of n with $1 \leq n < X/6$ such that both $6n+t$ and $6n+2B+t$ are prime, and when applicable, $6n-t$ and $6n+2B-t$ are also prime. We observe that

$$\pi'_{2B}(X) = \begin{cases} \pi_{2B}(X) - 1 & \text{if } (3, 3+2B) \text{ is a G.P.P.} \\ \pi_{2B}(X) & \text{if } (3, 3+2B) \text{ is not a G.P.P.} \end{cases}$$

Hence we would expect $\pi'_{2B}(X)$ to have a similar upper bound as $\pi_{2B}(X)$ upon applying the Selberg sieve on the numbers $(6n+t)(6n+2B+t)$, along with the numbers $(6n-t)(6n+2B-t)$ (when applicable). This is indeed the case as demonstrated by the next theorem, and we can actually achieve the same upper bound.

Theorem 5.2.2 *Let*

$$C = \prod_{p>2} \frac{p(p-2)}{(p-1)^2} \quad \text{and} \quad C_B = \prod_{\substack{p|B \\ p \neq 2}} \frac{p-1}{p-2}, \quad (5.2.2)$$

where the products are over odd primes p . Then

$$\pi'_{2B}(X) \leq \frac{16CC_B X}{\log^2 X} \left(1 + O\left(\frac{\log \log X}{\log X}\right) \right)$$

Proof:

Recall that general prime pairs are of the form $(6n + t, 6n + 2B + t)$ where $B \geq 1$ is an integer constant and

$$t = \begin{cases} -1 & \text{if } B \equiv 1 \pmod{3} \\ 1 & \text{if } B \equiv 2 \pmod{3} \\ -1 \text{ or } 1 & \text{if } B \equiv 0 \pmod{3}. \end{cases}$$

Also recall that when $3 \mid B$, we apply Theorem 5.1.1 to each case separately ($t = 1$ and $t = -1$). So in the case when $3 \mid B$, it is convenient to think of t as being fixed (say $t = 1$) and so general prime pairs in this situation are either of the form $(6n + t, 6n + 2B + t)$ or of the form $(6n - t, 6n + 2B - t)$. When $3 \nmid B$, we just said that general prime pairs are of the form $(6n + t, 6n + 2B + t)$ (recall t is understood to be predetermined). As such, we take $f_1(n) = (6n + t)(6n + 2B + t)$, and when applicable (i.e., $3 \mid B$), $f_2(n) = (6n - t)(6n + 2B - t)$.

Then for each situation, f_1 and f_2 (when applicable), the function ρ is given by

$$\rho(p) = \begin{cases} 0 & \text{if } p = 2 \text{ or } p = 3 \\ 1 & \text{if } p \mid B \text{ and } p \neq 2 \text{ or } 3 \\ 2 & \text{otherwise.} \end{cases}$$

Thus $\rho(p) \leq 2$ for all p , so (as in Lemma 2.1) ρ has sifting density $\kappa = 2$, and (3.2.38) holds with $\eta(z, w) = O(1)$. Then Theorem 3.2.2 gives

$$\begin{aligned} G(\sqrt{D}) &\geq \frac{1}{2!} \left(\frac{e^{-2\gamma}}{V(P(\sqrt{D}))} \right) \left(1 + O\left(\frac{1}{\log D}\right) \right) \\ &= \frac{1}{2!} A'_{DB} \log^2 \sqrt{D} \left(1 + O\left(\frac{1}{\log D}\right) \right), \end{aligned}$$

where

$$\begin{aligned}
 A'_{DB} &= \frac{1}{9} \prod_{\substack{p|B \\ p \neq 2, 3}} \left(1 - \frac{1}{p}\right) \prod_{\substack{p < \sqrt{D} \\ p \nmid B, p \neq 2, 3}} \left(1 - \frac{1}{p}\right)^2 \left(1 - \frac{2}{p}\right)^{-1} \\
 &= \frac{1}{9} \prod_{\substack{p|B \\ p \neq 2, 3}} \left(1 - \frac{1}{p}\right) \prod_{\substack{p < \sqrt{D} \\ p \nmid B, p \neq 2, 3}} \frac{(p-1)^2}{p(p-2)} \\
 &= \frac{1}{9} \left(\prod_{\substack{p|B \\ p \neq 2, 3}} \left(1 - \frac{1}{p}\right) \right) \frac{3}{4} \left(\prod_{\substack{p|B \\ p \neq 2, 3}} \frac{p(p-2)}{(p-1)^2} \right) \frac{1}{C} \prod_{p \geq \sqrt{D}} \frac{p(p-2)}{(p-1)^2} \\
 &\quad (C \text{ as in (5.2.2)}) \\
 &= \frac{1}{12} \left(\prod_{\substack{p|B \\ p \neq 2, 3}} \left(\frac{p-1}{p}\right) \right) \left(\prod_{\substack{p|B \\ p \neq 2, 3}} \frac{p(p-2)}{(p-1)^2} \right) \frac{1}{C} \left(1 + O\left(\frac{1}{\sqrt{D}}\right)\right) \\
 &\quad (\text{see (2.31) with } w = \sqrt{D}) \\
 &= \begin{cases} \frac{1}{12} \frac{1}{(C_B/2)} \frac{1}{C} \left(1 + O\left(\frac{1}{\sqrt{D}}\right)\right) & \text{if } 3 \mid B \text{ and } C_B \text{ as in (5.2.2)} \\ \frac{1}{12} \frac{1}{C_B} \frac{1}{C} \left(1 + O\left(\frac{1}{\sqrt{D}}\right)\right) & \text{if } 3 \nmid B \text{ and } C_B \text{ as in (5.2.2)} \end{cases}
 \end{aligned}$$

Applying Theorems 3.1.1 and 3.2.3, and multiplying by 2 in the case when $3 \mid B$ (to combine both situations, f_1 and f_2), we have in all cases,

$$\pi'_{2B}(X) \leq S(A, P(\sqrt{D})) + \pi(\sqrt{D}) \leq \frac{2 \cdot \frac{X}{6}}{A'_{DB} \log^2 \sqrt{D}} + O(D \log^{2\kappa} D).$$

Our result follows as stated on choosing $D = X/(\log X)^{2\kappa+3}$, where $\kappa = 2$.

Q.E.D.

As can be seen by comparing the results of the last two theorems, the alternate approach to sifting the general prime pairs leads to the same upper bound as the standard (Eratosthenes) approach.

5.3 Sum of the Reciprocals

As for twin primes, the sum of the reciprocals for general prime pairs is convergent (or possibly finite). The proof of this result is practically identical

to that of Theorem 4.3.1 but with the obvious differences.

Theorem 5.3.1 *Let p_1, p_2, \dots be the sequence of prime numbers p such that $p + 2B$ is also prime. Then*

$$\sum_i \left(\frac{1}{p_i} + \frac{1}{p_i + 2B} \right) \leq M,$$

where M is a constant.

Proof:

We wish to use partial summation; in particular, Lemma 2.2 which makes use of integration by parts (2.32). Let

$$a_n = \begin{cases} 1 & \text{if } n, n + 2B \text{ are both prime,} \\ 0 & \text{otherwise.} \end{cases}$$

and

$$f(n) = \frac{1}{n} \quad (f' \text{ is continuous on } [2, \infty)).$$

Then let

$$A(x) = \sum_{2 \leq n < x} a_n = \pi_{2B}(x),$$

and

$$B(x) = \sum_{2 \leq n < x} \frac{a_n}{n} = \sum_{\substack{2 \leq n < x \\ n, n+2B \text{ prime}}} \frac{1}{n}.$$

Note here that $B(x)$ is the sum of the reciprocals of only the first members of the general prime pairs. This will not be a problem, however, because if $B(x)$ is shown to be bounded by a constant $\frac{M}{2}$, then the sum of all reciprocals would be bounded by M since $\frac{1}{n+2B} < \frac{1}{n}$. Proceeding now with use of Lemma 2.2, we have

$$B(x) = \int_2^x f(t) d(A(t)),$$

and making use of integration by parts (2.32), this becomes

$$\begin{aligned}
 B(x) &= [A(t)f(t)]_2^x - \int_2^x A(t)f'(t)dt \\
 &= A(x)f(x) - A(2)f(2) - \int_2^x A(t)f'(t)dt \\
 &= A(x)f(x) - \int_2^x A(t)f'(t)dt \quad (\text{since } A(2) = 0). \\
 &= A(x)\left(\frac{1}{x}\right) - \int_2^x A(t)\left(\frac{-1}{t^2}\right)dt \\
 &= A(x)\left(\frac{1}{x}\right) + \int_2^x A(t)\left(\frac{1}{t^2}\right)dt \\
 &\leq 16CC_B \left[\left(\frac{x}{\log^2(x)}\right)\left(\frac{1}{x}\right) + \int_2^x \left(\frac{t}{\log^2(t)}\right)\left(\frac{1}{t^2}\right)dt \right] \\
 &\quad (\text{by Theorems 5.2.1 and 5.2.2}),
 \end{aligned}$$

where C and C_B are the constants from Theorems 5.2.1 and 5.2.2. Continuing the calculation, we get

$$\begin{aligned}
 B(x) &\leq 16CC_B \left[\frac{1}{\log^2(x)} + \int_2^x \frac{1}{t \log^2(t)} dt \right] \\
 &= 16CC_B \left[\frac{1}{\log^2(x)} - \left[\frac{1}{\log(t)} \right]_2^x \right] \\
 &= 16CC_B \left[\frac{1}{\log^2(x)} - \frac{1}{\log(x)} + \frac{1}{\log(2)} \right] \\
 &\rightarrow \frac{16CC_B}{\log(2)} \quad (\text{as } x \rightarrow \infty).
 \end{aligned}$$

From this it follows that

$$\sum_i \left(\frac{1}{p_i} + \frac{1}{p_i + 2B} \right) \leq M,$$

with

$$M = \frac{32CC_B}{\log(2)}.$$

Q.E.D.

Chapter 6

Concluding Remarks

As we have seen, the methods of Selberg are quite good for providing upper bounds. This is good and leads to results like those of Theorems 4.3.1 and 5.3.1, where the sum of the reciprocals of the twin primes/general prime pairs are convergent (or finite). However, mathematicians would like to find some means of providing effective lower bounds. It is not yet clear if, and how, the methods of Selberg can be manipulated or reformulated in such a way so as to provide good lower bounds. This is an area of active research; [4] addresses/presents some plausible approaches to the problem.

The alternate approach we presented for sifting twin primes was quite elegant and easy to perform. The starting points of cancellation coincided and we saw that for every prime $p \geq 5$, we had $b_p = -a_p$. This meant that the cancellation step for each prime had a symmetrical aspect to it. While things were pleasant for twin primes, they got a bit uglier for more general prime pairs. We generally had two distinct starting cancellation points for each prime $p \geq 5$ and we no longer had symmetry in the form $b_p = -a_p$. Also, depending on whether or not $3 \mid B$, we may have had to perform a double application of the sieve. Despite all this, however, the alternate sifting approach was effective and easy to perform.

It must also be pointed out that the alternate sifting approach could be reformulated and applied toward finding very specific prime pairs. For instance, if one were specifically interested in finding prime pairs of the form $(Ak+l, Ak+2B+l)$, where A, B, l are fixed constants and $\text{g.c.d.}(A, l) = 1$ and $\text{g.c.d.}(A, 2B+l) = 1$, the alternate approach could be formulated to do just that. In addition to this, the alternate sifting approach can also be extended and used for finding prime n -tuples; that is, tuples each consisting of n primes, where it has been pre-determined how the primes are to be separated from each other (twin primes are 2-tuples that are separated by 2). Note also that one could decide to "shrink" the method and find 1-tuples; i.e., just find primes of some form $Ak+l$, where A, l are constants and $\text{g.c.d.}(A, l) = 1$.

We saw how the Selberg method was applied to twin primes, and it was mentioned earlier that it is still an open problem as to whether or not the twin primes occur infinitely often throughout the integers. The Twin Prime Conjecture states that they do occur infinitely often, and somewhat associated with this conjecture, is the famous Goldbach Conjecture, which states that every even number greater than 2 can be written as the sum of precisely two primes. Both of these problems have been open for many years, but the techniques of Sieve Theory have come remarkably close to solving them. Things seem to have culminated with the results of J.R. Chen, see [3]. He essentially proved, using similar techniques, the following two theorems.

Theorem 6.1 *There are infinitely many primes p where $p+2$ is either prime or is the product of 2 primes.*

Theorem 6.2 *Every sufficiently large even number is either the sum of two primes or is the sum of a prime and a product of 2 primes.*

One can also see [5, Chapter 11], where the entire chapter is devoted to Chen's theorem(s).

It is clear that Sieve Theory is quite powerful and useful, yet seems to be in need of new information in order to crack these age old problems. The Twin Prime Conjecture and Goldbach Conjecture are only two of many problems where sieve techniques can be applied.

Bibliography

- [1] Apostol, Tom M. *Introduction to Analytic Number Theory*, Undergraduate Texts in Mathematics. Springer, 1976.
- [2] Brun, Viggo. La série $1/5 + 1/7 + 1/11 + 1/13 + 1/17 + 1/19 + 1/29 + 1/31 + 1/41 + 1/43 + 1/59 + 1/61 + \dots$, où les dénominateurs sont "nombres premiers jumeaux" est convergente où finie. *Bulletin des Sciences Mathématiques*, 43, 1919. pgs. 100-104, 124-128.
- [3] Chen, J. On the representation of a large even integer as the sum of a prime and the product of at most two primes. (Chinese) *Kexue Tongbao*, 17, 1966. pgs. 385-386.
- [4] Greaves, George. *Sieves in Number Theory*, A Series of Modern Surveys in Mathematics, Volume 43. Springer, 2001.
- [5] Halberstam, H. and Richert, H.-E. *Sieve Methods*. Academic Press Inc. 1974. pgs. 320-338.
- [6] Niven, Ivan; Zuckerman, Herbert S. and Montgomery, Hugh L. *An Introduction to The Theory of Numbers*, fifth edition. John Wiley and Sons, Inc. 1991. pgs. 25-28.
- [7] Rudin, Walter. *Principles of Mathematical Analysis*, third edition. McGraw-Hill, Inc. 1976. pgs. 63-64.