REFINEMENTS OF SELBERG'S SIEVE

BY SARA ELIZABETH BLIGHT

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Mathematics

Written under the direction of

Henryk Iwaniec

and approved by

New Brunswick, New Jersey

May, 2010

ABSTRACT OF THE DISSERTATION

Refinements of Selberg's Sieve

by Sara Elizabeth Blight

Dissertation Director: Henryk Iwaniec

This thesis focuses on refinements of Selberg's sieve as well as new applications of the sieve. Sieve methods are addressed in four ways. First, we look at lower bound sieves. We will construct new lower bound sieves that give us non-trivial lower bounds for our sums. The lower bound sieves we construct will give better results than those previously known.

Second, we create an upper bound sieve and use it to bound the number of primes to improve Selberg's version of the Brun-Titchmarsh Theorem. We improve a constant in the bound of the number of primes in an arbitrary interval of fixed length.

Third, we construct an upper bound sieve to improve the large sieve inequality in special cases. Sieve methods allow us to improve this well-known bound of exponential sums.

Finally, we include some notes on the use of successive approximations to give a choice of an upper bound sieve that minimizes the main term and the remainder term simultaneously.

Acknowledgements

I would like to thank my advisor, Professor Iwaniec, for his invaluable guidance over the years. I have learned so much from his classes and from working with him on research. Also, I would like to thank the rest of my committee, Professors Chamizo, Miller, and Weibel for their help and feedback during this process. All of the Rutgers faculty have been very helpful, especially Professors Cohen, Greenfield, and Nussbaum.

I want to thank my friends and family for their support. My fellow graduate students have made the graduate experience very enjoyable. In particular, I want to thank Beth, Emilie, and Leigh for their wonderful friendship. Finally, I would like to thank my family for always listening, supporting me during the tough times, and celebrating with me during the good times.

Dedication

This thesis is dedicated to my family and friends, especially to my Mom for her invaluable advice of starting with a fresh piece of paper.

Table of Contents

Chapter 1

Introduction

In sifting theory, a quantity we are interested in is how many of the numbers $n \leq x$ have no prime factors p less than a parameter z. If we take $P(z)$ to be the product of all primes $p \leq z$, then this means that we want to count how many numbers $n \leq x$ are such that $(n, P(z)) = 1$. In summation form, this is

$$S(x, z) = \sum_{\substack{n \leq x \\ (n, P(z))=1}} 1. \tag{1.1}$$

Although this is an interesting problem by itself, we would like to look at an even more general problem. Instead of looking at the sum (1.1), we look at the weighted sum below.

$$S(\mathcal{A}, x, z) = \sum_{\substack{n \leq x \\ (n, P(z))=1}} a_n \tag{1.2}$$

where $\mathcal{A} = (a_n)$ is a sequence of non-negative real numbers. Now we would like to estimate this sum. We would like to remove the condition $(n, P(z)) = 1$ from the summation. One way we can do this is by using the Mobiüs function. We recall that

$$\sum_{d|m} \mu(d) = \begin{cases} 1 & \text{if } m = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1.3}$$

Therefore,

$$\begin{aligned} S(\mathcal{A}, x, z) &= \sum_{n \leq x} a_n \sum_{d|(n,P(z))} \mu(d) \\ &= \sum_{n \leq x} a_n \sum_{\substack{d|n \\ d|P(z)}} \mu(d). \end{aligned}$$

Unfortunately, the Mobiüs function does not have nice asymptotics, so this is not the best approach for estimations. Therefore, we look at more general functions. If we want an upper bound for $S(\mathcal{A}, x, z)$, we take a sequence of real numbers $\Lambda^+ = (\lambda_d^+)$ such that

$$\sum_{d|m} \lambda_d^+ \begin{cases} = 1 & \text{if } m = 1, \\ \geq 0 & \text{if } m \neq 1. \end{cases}$$

Then

$$S(\mathcal{A}, x, z) \quad \leq \sum_{n \leq x} a_n \sum_{\substack{d|n \\ d|P(z)}} \lambda_d^+.$$

We note that the two conditions we imposed on λ_d^+ are equivalent to the conditions that $\lambda_1^+ = 1$ and $\sum_{d|m} \lambda_d^+ \geq 0$ for all m. If these conditions are satisfied and $\lambda_d^+ = 0$ for $d > D$, then $\Lambda^+ = (\lambda_d^+)$ is called an upper bound sieve of level D. With a clever choice of λ_d^+, we will be able to construct good upper bounds for $S(\mathcal{A}, x, z)$.

Now we would like to construct a lower bound for $S(\mathcal{A}, x, z)$. Trivially, we know that the sum is non-negative because each $a_n \geq 0$. We would like to be able to construct non-trivial lower bounds as well. If we want a lower bound, we take a sequence of real numbers $\Lambda^- = (\lambda_d^-)$ such that

$$\sum_{d|m} \lambda_d^- \begin{cases} = 1 & \text{if } m = 1, \\ \leq 0 & \text{if } m \neq 1. \end{cases}$$

Then

$$S(\mathcal{A}, x, z) \quad \geq \sum_{n \leq x} a_n \sum_{\substack{d|n \\ d|P(z)}} \lambda_d^-.$$

We note that the two conditions we imposed on λ_d^- are equivalent to the conditions that $\lambda_1^- = 1$ and $\sum_{d|m} \lambda_d^- \leq 0$ for all $m \neq 1$. If these conditions are satisfied and $\lambda_d^- = 0$ for $d > D$, then $\Lambda^- = (\lambda_d^-)$ is called a lower bound sieve of level D.

For now, we let

$$S(\Lambda) = \sum_{n \leq x} a_n \sum_{\substack{d|n \\ d|P(z)}} \lambda_d \tag{1.4}$$

where $\Lambda = (\lambda_d)$ is a general sieve of level D, either an upper bound sieve or a lower bound sieve. Changing the order of summation, we have

$$S(\Lambda) \quad = \sum_{d|P(z)} \lambda_d \sum_{\substack{n \leq x \\ n \equiv 0(d)}} a_n = \sum_{d|P(z)} \lambda_d A_d(x)$$

where

$$A_d(x) = \sum_{\substack{n\le x \\ n\equiv 0(d)}} a_n. \tag{1.5}$$

In order to treat the summation, we shall assume some asymptotics of the partial sums $A_d(x)$. We write

$$A_d(x) = g(d)X + r(\mathcal{A}, d)$$

where $g(d)X$ is the expected main term and $r(\mathcal{A}, d)$ is an error term which we think of as being small. In the main term,

$$X \approx \sum_{n\le x} a_n$$

so $g(d)$ is the density of the masses a_n attached to $n \equiv 0(\text{mod } d)$. If we think of divisibility by distinct primes as independent events, we are led to assume that $g(d)$ is a multiplicative function with $0 < g(p) < 1$ if $p|P(z)$ and $g(p) = 0$ otherwise. Given these asymptotics of $A_d(x)$, we have

$$\begin{aligned} S(\Lambda) \quad &= X\sum_{d|P(z)} g(d)\lambda_d + \sum_{d|P(z)} \lambda_d r(\mathcal{A}, d) \\ &= XV(D,z) + R(\mathcal{A}, D) \\ \text{where} \quad & V(D,z) = \sum_{d|P(z)} g(d)\lambda_d,\ R(\mathcal{A}, D) = \sum_{d|P(z)} \lambda_d r(\mathcal{A}, d). \end{aligned}$$

In this thesis, we will address sieve methods in four ways. First, we will look at lower bound sieves. A fundamental problem in sieve theory is the sifting limit problem. We wish to construct lower bound sieves that give non-trivial lower bounds for our sums. Given a particular asymptotic for our density function $g(d)$, the problem is to find a lower bound sieve that gives a non-trivial lower bound. Details of the problem are described in chapter (2). Selberg established the asymptotic result for this problem. He also predicted a value for the sifting limit in general. This is still an open problem. In this thesis, we construct a lower bound sieve that gives better results than those previously known in many cases.

Second, we will create an upper bound sieve and use it to bound the number of primes to improve Atle Selberg's version of the Brun-Titchmarsh Theorem. We will improve a constant in the bound of the number of primes in an arbitrary interval of fixed length.

Third, we will construct an upper bound sieve to improve the large sieve inequality in special cases. Sieve methods will allow us to improve this well-known bound of exponential sums.

Finally, we make some notes on the use of successive approximations to give an upper bound sieve that simultaneously minimizes the main term and remainder term. We provide useful lemmas to solve systems of equations prevalent in sieve theory.

Chapter 2

Sifting Limit

2.1 Introduction

From chapter (1), we know that if $\Lambda = (\lambda_d)$ is a lower bound sieve of level D, then

$$S(\mathcal{A}, x, z) = \sum_{\substack{n \leq x \\ (n, P(z))=1}} a_n \geq XV(D, z) + R(\mathcal{A}, D)$$

where

$$V(D, z) = \sum_{d|P(z)} g(d)\lambda_d$$

and

$$R(\mathcal{A}, D) = \sum_{d|P(z)} \lambda_d r(\mathcal{A}, d).$$

Since $a_n \geq 0$ for all n, we know trivially that $S(\mathcal{A}, x, z) \geq 0$. We would like to find a choice of Λ that gives a nontrivial lower bound. To do this, we first make a couple definitions.

Definition 2.1.1. *Let $g(p)$ be a multiplicative function supported on square free numbers with $0 < g(p) < 1$ for $p|P(z)$ and $g(p) = 0$ for $p \nmid P(z)$ and $\kappa > 0$ is a number which satisfies*

$$\sum_{p \leq x} g(p) \log p = \kappa \log x + O(1). \tag{2.1}$$

Further assume for any $w \geq 2$ with $w < z$ that

$$\prod_{w \leq p \leq z} (1 - g(p))^{-1} \ll \left(\frac{\log z}{\log w}\right)^{\kappa}.$$

Then κ is called the sifting dimension.

To control the size of $g(p)$, we assume

$$\sum_p g(p)^2 \log p < \infty.$$

In this chapter, we will be examining a quantity β_κ, known as the sifting limit for a sifting dimension κ. A precise definition of sifting limit is very complicated. For such a definition, the reader is referred to Selberg's Lectures on Sieves [4, Section 14]. Loosely speaking, for a sifting dimension κ and a lower bound sieve Λ of level D, by the sifting limit $\beta_{\kappa,\Lambda}$ we mean the minimum of $\log D/\log z$ for which $V(D,z) > 0$ for $\log D/\log z > \beta_{\kappa,\Lambda}$ and $V(D,z) \leq 0$ for $\log D/\log z \leq \beta_{\kappa,\Lambda}$. By the sifting limit β_κ for a sifting dimension κ, we mean the greatest lower bound of the $\beta_{\kappa,\Lambda}$ over all possible lower bound sieves Λ of level D. For ease of notation, in this thesis we denote $\beta_{\kappa,\Lambda}$ by β_κ when the sieve Λ is understood.

Selberg proposed that the sifting limit is 2κ. He was able to prove this result asymptotically as κ approached infinity. For $1/2 < \kappa < 1$, Iwaniec and Rosser constructed a sieve with $\beta_\kappa < 2\kappa$. However, at this time, a lower bound sieve with a sieving limit of 2κ has not been found for $\kappa > 1$. The sifting limit problem is to find lower bound sieves that give $\beta_\kappa \leq 2\kappa$ for each $\kappa > 1$.

There are many types of lower bound sieves, each with its own advantages and disadvantages. We will mention a few of these sieves and the corresponding sifting limits. Then we will give an improvement on Selberg's lower bound sieve, which will provide significant improvement of the sifting limit for $\kappa \geq 3$.

2.2 Various Lower Bound Sieves

There are many choices of lower bound sieves that provide good sifting limits. Here we will focus on the beta-sieve, the Diamond-Halberstam sieve, and the Selberg sieve.

2.2.1 Beta Sieve

The beta sieve was created by Iwaniec and Rosser. The sieve works very well when the sifting dimension is small. We let β_κ denote the sifting limit for sifting dimension κ.

Using formulas (B.9), (11.42), and (11.57) of [2], we established the following numerical values of β_κ.

κ	β_κ
0.5	1.0000000000
0.55	1.0340771100
0.6	1.1042161305
0.65	1.1922077070
0.7	1.2912892849
0.75	1.3981115251
0.8	1.5107489225
0.85	1.6279798714
0.9	1.7489723058
0.95	1.8731283112
1	2.0000000000
1.05	2.1292406269
1.1	2.2605745188
1.15	2.3937776845
1.2	2.5286648100
1.25	2.6650802364
1.3	2.8028915201
1.35	2.9419847168
1.4	3.0822608556
1.45	3.2236332483
1.5	3.3660254038
2	4.8339865967

Asymptotically, the beta sieve gives

$$\beta_\kappa \sim c\kappa$$

where $c = 3.591\ldots$ is the number which solves the equation $(c/e)^c = e$. We also note that $\beta_\kappa < 2\kappa$ if $\frac{1}{2} < \kappa < 1$.

2.2.2 Diamond-Halberstam Sieve

The Diamond-Halberstam sieve works well for slightly larger sifting dimension. We have the following values of sifting limits [1, p.227].

κ	β_κ
1.0	2.000000
1.5	3.115821
2.0	4.266450
2.5	5.444068
3.0	6.640859
3.5	7.851463
4.0	9.072248
4.5	10.300628
5.0	11.534709
5.5	12.773074
6.0	14.014644
6.5	15.258588
7.0	16.504285
7.5	17.751146
8.0	18.998853
8.5	20.247056
9.0	21.495510
9.5	22.744013
10.0	23.992408

The Diamond-Halberstam sieve is an infinite iteration of the Ankeny-Onishi sieve[1]. Therefore, it is believed that the sifting limit

$$\beta_\kappa \sim c\kappa$$

as $\kappa \to \infty$, where $c = 2.445\ldots$.

2.2.3 Selberg Sieve

Finally, we turn to the Selberg sieve. The Selberg sieve does not give good sifting limits when the sifting dimension is very small. However, asymptotically,

$$\beta_\kappa \sim 2\kappa$$

which is better than any other sieve. By making different choices and better estimates, we have been able to modify the Selberg sieve to give good sifting limits for small κ. These sifting limits are smaller than the sifting limits of the other sieves for $\kappa \geq 3$. In this section, we explain Selberg's approach. In the next section, we will explain the modifications.

With straightforward calculations, we see that if Λ^+ is an upper bound sieve of level D_1 and Λ^- is a lower bound sieve of level D_2, then $\Lambda = \Lambda^+\Lambda^-$ is a lower bound sieve of level $D_1 D_2$, defined by

$$\sum_{\substack{d|n \\ d|P(z)}} \lambda_d = \Bigg(\sum_{\substack{d|n \\ d|P(z)}} \lambda_d^+ \Bigg) \Bigg(\sum_{\substack{d|n \\ d|P(z)}} \lambda_d^- \Bigg).$$

Applying this lower bound sieve, we have

$$S(\mathcal{A}, x, z) \;=\; \sum_{\substack{n \leq x \\ (n, P(z))=1}} a_n \geq \sum_{n \leq x} a_n \Bigg(\sum_{\substack{d|n \\ d|P(z)}} \lambda_d \Bigg) = \sum_{n \leq x} a_n \Bigg(\sum_{\substack{d|n \\ d|P(z)}} \lambda_d^- \Bigg) \Bigg(\sum_{\substack{d|n \\ d|P(z)}} \lambda_d^+ \Bigg).$$

For Λ^-, Selberg chose $\lambda_1^- = 1$, $\lambda_p^- = -1$ for $p \leq z$ and $\lambda_d^- = 0$ otherwise, so that Λ^- is a lower bound sieve of level z. Then,

$$\sum_{\substack{d|n \\ d|P(z)}} \lambda_d^- = 1 - \sum_{\substack{p|n \\ p|P(z)}} 1.$$

This is a crude lower bound sieve. However, with a good choice of Λ^+, the overall choice of $\Lambda = \Lambda^-\Lambda^+$ is still good. For Λ^+, Selberg chose his Λ^2 sieve, which is the convolution of Λ with itself, $\Lambda^+ = \Lambda\Lambda$ with $\Lambda = \{\rho_d\}$. That is he chose λ_d^+ such that

$$\sum_{d|m} \lambda_d^+ = \Bigg(\sum_{d|m} \rho_d \Bigg)^2$$

where $\{\rho_d\}$ is another sequence of real numbers with $\rho_1 = 1$ and $\rho_d = 0$ for $d > \sqrt{D/z} = Y$. Then Λ^+ is an upper bound sieve of level D/z and $\Lambda = \Lambda^- \Lambda^+$ is a lower bound sieve of level D. He kept the choice of ρ_d open. Applying these choices, we see

$$\begin{aligned} S(\mathcal{A}, x, z) &= \sum_{n \le x} a_n \Bigg(1 - \sum_{\substack{p|n \\ p|P(z)}} 1\Bigg)\Bigg(\sum_{\substack{d|n \\ d|P(z)}} \rho_d\Bigg)^2 \\ &= XV(D, z) + R(\mathcal{A}, D) \end{aligned}$$

where

$$V(D, z) = \sum_{d|P(z)} g(d)\lambda_d, \; R(\mathcal{A}, D) = \sum_{d|P(z)} \lambda_d r_d(\mathcal{A}),$$

in the notation of chapter 1.

We note that with the above definitions,

$$\lambda_d \; = \lambda_d^+ - \sum_{p|d} \left(\lambda_d^+ + \lambda_{d/p}^+\right).$$

Rewriting λ_d^+ in terms of ρ_d and noting that d is squarefree, and manipulating the result, we see that

$$\lambda_d \; = \sum_{[d_1,d_2]=d} \rho_{d_1}\rho_{d_2} - \sum_{[p,d_1,d_2]=d} \rho_{d_1}\rho_{d_2}.$$

Then

$$V(D, z) = \sum_{d_1}\sum_{d_2} g([d_1, d_2])\rho_{d_1}\rho_{d_2} - \sum_{p}\sum_{d_1}\sum_{d_2} g([p, d_1, d_2])\rho_{d_1}\rho_{d_2},$$

where p, d_1, d_2 run independently over divisors of $P(z)$, p prime.

In order to look at the first sum, we first define the multiplicative function $h(d)$ by

$$h(p) = \frac{g(p)}{1 - g(p)}.$$

We note that since $0 < g(p) < 1$ for $p|P(z)$, we have $h(p) > 0$ for all $p|P(z)$ and $h(d) > 0$ for all $d|P(z)$. Since $g(p) = 0$ for $p \nmid P(z)$, $h(p) = 0$ for $p \nmid P(z)$. Then we have

$$\sum_{d_1}\sum_{d_2} g([d_1, d_2])\rho_{d_1}\rho_{d_2} = \sum_{d} h(d)^{-1}\Bigg(\sum_{m \equiv 0 \pmod{d}} g(m)\rho_m\Bigg)^2.$$

In order to treat the second sum, we need to make some more definitions. We define $g_p(d) = g([p, d])/g(p)$ and $h_p(d)$ as $h_p(d) = \infty$ if $p|d$ and $h_p(d) = h(d)$ otherwise.

Finally, we define

$$G_p \quad := \sum_{d} h_p(d)^{-1}\Bigg(\sum_{m\equiv 0 (\mathrm{mod}\ d)} g_p(m)\rho_m\Bigg)^2.$$

By expanding the square and simplifying, we find

$$G_p \quad = \sum_{m_1}\sum_{m_2} \rho_{m_1}\rho_{m_2}\frac{g([p,m_1,m_2])}{g(p)}.$$

Therefore,

$$\begin{aligned}\textstyle\sum_p g(p)G_p \quad &= \sum_{p} g(p)\sum_{d} h_p(d)^{-1}\Bigg(\sum_{m\equiv 0 (\mathrm{mod}\ d)} g_p(m)\rho_m\Bigg)^2 \\ &= \sum_{m_1}\sum_{m_2} \rho_{m_1}\rho_{m_2} g([p,m_1,m_2]),\end{aligned}$$

which is the second sum in our expression for $V(D,z)$.

Rewriting g_p and h_p in terms of g and h, we find that

$$\textstyle\sum_{p|P(z)} g(p)G_p \quad = \displaystyle\sum_{pd|P(z)} \frac{g(p)}{h(d)}\Bigg(\sum_{\substack{m\equiv 0 (\mathrm{mod}\ d)\\ m|P(z)}} g(m)\rho_m + \frac{1}{h(p)}\sum_{\substack{m\equiv 0 (\mathrm{mod}\ pd)\\ m|P(z)}} g(m)\rho_m\Bigg)^2.$$

Finally,

$$\begin{aligned}V(D,z) \quad &= \sum_{d|P(z)} \frac{1}{h(d)}\Bigg(\sum_{\substack{m\equiv 0 (\mathrm{mod}\ d)\\ m|P(z)}} g(m)\rho_m\Bigg)^2 \\ &- \sum_{d|P(z)}\sum_{pd|P(z)} \frac{g(p)}{h(d)}\Bigg(\sum_{\substack{m\equiv 0 (\mathrm{mod}\ d)\\ m|P(z)}} g(m)\rho_m + \frac{1}{h(p)}\sum_{\substack{m\equiv 0 (\mathrm{mod}\ pd)\\ m|P(z)}} g(m)\rho_m\Bigg)^2.\end{aligned}$$

Since the sum

$$\sum_{\substack{m\equiv 0 (\mathrm{mod}\ d)\\ m|P(z)}} g(m)\rho_m$$

is prevalent, we make a change of variables to simplify the expression. Selberg chose

$$y_d = \frac{\mu(d)}{h(d)}\sum_{\substack{m|P(z)\\ m\equiv 0 (\mathrm{mod}\ d)}} g(m)\rho_m. \tag{2.2}$$

Making this substitution, we find

$$V(D,z) = \sum_{d|P(z)} h(d)y_d^2 - \sum_{pd|P(z)} g(p)h(d)\Big(y_d - y_{pd}\Big)^2. \tag{2.3}$$

We note that the original variables ρ_d can be found in terms of the new variables y_d by Mobiüs inversion. For the normalization, we note that $\rho_1 = 1$ means that

$$1 = \sum_{d|P(z)} h(d) y_d.$$

Also, the support of ρ_d being $d \leq Y$ is equivalent to the support of y_d being $d \leq Y$.

This is the point where two different paths may be taken. The first path is ideal for large sifting dimension because it illumines the asymptotics Selberg was able to achieve. However, in order to clearly see the asymptotics, some estimates are made which worsen the result for small sifting dimension. The second path is more computationally heavy, but yet gives better results for small sifting dimension. The work of this thesis expands upon the second path to give even more precise results for small sifting dimension. We will explore these paths in the following sections.

2.3 Selberg's Choice

In section (2.2.3), we gave the set-up of Selberg's sifting limit argument. We now continue with an explanation of his work. Since $h(p) \geq g(p)$, equation (2.3) gives

$$V(D,z) \geq \sum_{d|P(z)} h(d) y_d^2 - \sum_{pd|P(z)} h(pd)(y_d - y_{pd})^2. \tag{2.4}$$

Rewriting this, we find

$$\begin{aligned} V(D,z) &\geq \sum_{d|P(z)} h(d) y_d^2 - \sum_{pd|P(z)} h(pd)(y_d - y_{pd})^2 \\ &= \sum_{d|P(z)} h(d) \left\{ y_d^2 - \sum_{p|d} (y_{d/p} - y_d)^2 \right\} \\ &= \sum_{d|P(z)} h(d) \{ y_d^2 - l(d) \} \end{aligned}$$

where

$$l(d) = \sum_{p|d} (y_{d/p} - y_d)^2.$$

Selberg then chose

$$y_d = J^{-1} \begin{cases} \min\left\{1, \dfrac{\log Y/d}{\log z}\right\} & \text{if } 1 \leq d \leq Y \\ 0 & \text{otherwise} \end{cases}, \text{ where } J = \sum_{\substack{d|P(z) \\ d \leq Y}} h(d).$$

We recall that $Y = \sqrt{D/z}$. We note that with this choice of y_d, $l(d) = 0$ except for $x/z < d < xz$. In this range,

$$l(d) = \sum_{p|d} (y_{d/p} - y_d)^2 \le J^{-2} \sum_{p|d} \left(\frac{\log p}{\log z} \right)^2 \le J^{-2} \sum_{p|d} \frac{\log p}{\log z} = J^{-2} \frac{\log d}{\log z}. \tag{2.5}$$

By making these crude estimates, we lose some of our precision. The precision will not matter for the main term of the asymptotic result, but it does effect the result for small sifting dimension. We have

$$\sum_{\substack{d|P(z) \\ Y/z<d<Yz}} h(d)\{y_d^2 - l(d)\} \ge -J^{-2} \sum_{\substack{d|P(z) \\ Y/z<d<Y}} h(d) \frac{\log d}{\log z}. \tag{2.6}$$

Also,

$$\sum_{\substack{d|P(z) \\ d\le Y/z}} h(d)\{y_d^2 - l(d)\} = J^{-2} \sum_{\substack{d|P(z) \\ d\le Y/z}} h(d). \tag{2.7}$$

Therefore,

$$\begin{aligned} J^2 V(D,z) &\ge J^2 \sum_{d|P(z)} h(d)\{y_d^2 - l(d)\} \\ &\ge \sum_{\substack{d|P(z) \\ d\le Y/z}} h(d) - \sum_{\substack{d|P(z) \\ Y/z<d<Yz}} h(d) \frac{\log d}{\log z} \\ &\ge \sum_{\substack{d|P(z) \\ d\le Y/z}} h(d) - \frac{\log Yz}{\log z} \sum_{\substack{d|P(z) \\ d\ge Y/z}} h(d) \\ &= \sum_{d|P(z)} h(d) - \sum_{\substack{d|P(z) \\ Y/z<d<Yz}} h(d) - \frac{\log Yz}{\log z} \sum_{\substack{d|P(z) \\ Y/z<d<Yz}} h(d) \\ &= \sum_{d|P(z)} h(d) - \frac{\log Yz^2}{\log z} \sum_{\substack{d|P(z) \\ Y/z<d<Yz}} h(d). \end{aligned}$$

Definition 2.3.1. *We define $V(z)$ by*

$$V(z)^{-1} = \sum_{d|P(z)} h(d).$$

We define

$$I(X,z) = \sum_{\substack{d\ge X \\ d|P(z)}} h(d).$$

Then

$$\begin{aligned} J^2V(D,z) &\geq V(z)^{-1} - \frac{\log Yz^2}{\log z} I(Y/z,z) \\ &\text{so} \\ J^2V(D,z)V(z) &\geq 1 - \frac{\log Yz^2}{\log z} I(Y/z,z)V(z). \end{aligned}$$

From Opera de Cribro by Friedlander and Iwaniec [2, p.111], we have

$$I(X,z)V(z) \leq e^{-\kappa}\left(\frac{2e\kappa}{t}\right)^{t/2}$$

if $t = 2\log X/\log z > 2\kappa$. We note that $s = \log D/\log z = 2(\log Y/\log z)+1$. Therefore, $\log Y/\log z = (s-1)/2$. Letting $X = Y/z$, we have $t = s-3$. Therefore, if $s > 2\kappa+3$,

$$I(Y/z,z)V(z) \leq e^{-\kappa}\left(\frac{2e\kappa}{s-3}\right)^{(s-3)/2}.$$

Hence,

$$J^2V(D,z)V(z) \geq 1 - \frac{s+3}{2e^{\kappa}}\left(\frac{2e\kappa}{s-3}\right)^{(s-3)/2}.$$

Thus, $V(D,z) > 0$, if

$$\frac{s+3}{2e^{\kappa}}\left(\frac{2e\kappa}{s-3}\right)^{(s-3)/2} < 1$$

assuming $s > 2\kappa+3$. This occurs when $s > 2\kappa + 2\sqrt{2\kappa\log\kappa} + \log\kappa + 9$. This provides an upper bound for the sifting limit in the case of sifting dimension κ.

2.4 Choice for Small Sifting Dimension

In the previous section, we made some estimates to give a clear asymptotic. We now use better bounds to achieve good sifting limits for small sifting dimension. In equation (2.3) we choose our variables y_d as follows.

$$y_d = \begin{cases} J^{-1}F\left(\dfrac{\log d}{\log Y}\right) & \text{if } 1 \leq d \leq Y \\ 0 & \text{otherwise} \end{cases} \tag{2.8}$$

where F is a general continuous, piecewise smooth function which will be picked to optimize results for each sifting dimension. We define $\alpha = \log z/\log Y = 2/(s-1)$.

We note that Selberg's choice of y_d from the previous section corresponds to:

$$F(t) = \begin{cases} 1 & \text{if } 0 \le t \le 1-\alpha, \\ \frac{1}{\alpha}(1-t) & \text{if } 1-\alpha < t \le 1. \end{cases} \tag{2.9}$$

Now, we have

$$\begin{aligned} J^2V(D,z) &= \sum_{\substack{d|P(z) \\ d\le Y}} h(d)F^2\left(\frac{\log d}{\log Y}\right) \\ &- \sum_{d|P(z)} h(d) \sum_{p|P(z)} g(p) \left[F\left(\frac{\log d}{\log Y}\right) - F\left(\frac{\log pd}{\log Y}\right)\right]^2. \end{aligned}$$

We now treat the sum over primes in the second line. We assume that F is a continuous piecewise smooth function with $(F(0) - F(u))^2 \ll u$ for $0 \le u \le 1$. We now apply Lemma (2.8.3) with

$$\Phi\left(\frac{\log p}{\log Y}\right) = \left[F\left(\frac{\log d}{\log Y}\right) - F\left(\frac{\log d}{\log Y} + \frac{\log p}{\log Y}\right)\right]^2.$$

Therefore, we have

$$\begin{aligned} &\sum_{p|P(z)} g(p) \left[F\left(\frac{\log d}{\log Y}\right) - F\left(\frac{\log pd}{\log Y}\right)\right]^2 \\ &= \kappa \int_0^\alpha (F(v) - F(v+u))^2 \frac{du}{u} + O\left(\frac{\log\log z}{\log z}\right) \end{aligned}$$

where $\alpha = \log z/\log Y$ and $v = \log d/\log Y$. We apply Lemma (2.8.4) with

$$\Phi(v) = F^2(v) - \kappa \int_0^\alpha (F(v) - F(u+v))^2 \frac{du}{u}.$$

For the contribution from the error term of $O(\log\log z/\log z)$, we note that $\sum_{d|P(z)} h(d) = V(z)^{-1}$. Then we have

$$\begin{aligned} c^{-1}V(z)J^2V(D,z) &= \int_0^1 F(v)^2 d\mathfrak{f}(v/\alpha) \\ &-\kappa \int_0^1 \int_0^\alpha (F(v) - F(u+v))^2 \frac{du}{u} d\mathfrak{f}(v/\alpha) \\ &+O\left(c^{-1}\frac{\log\log z}{\log z}\right) \end{aligned}$$

where $c^{-1} = e^{\gamma\kappa}\Gamma(\kappa+1)$, γ is Euler's constant, $\mathfrak{f}$ is given by (2.16) and

$$V(z) = \prod_{p|P(z)} (1 + h(p))^{-1}.$$

Then we have

$$\begin{aligned} c^{-1}V(z)J^2V(D,z) &= \int_0^1 F(v)^2 d\mathfrak{f}(v/\alpha) \\ &-\kappa \int_0^1 \int_0^\alpha (F(v)-F(v+u))^2 \frac{du}{u} d\mathfrak{f}(v/\alpha) \\ &+O\left(c^{-1}\frac{\log\log z}{\log z}\right). \end{aligned}$$

Then

$$\begin{aligned} \tfrac{\alpha}{c}V(z)J^2V(D,z) &= \int_0^1 F^2(v)\mathfrak{f}'(v/\alpha)dv \\ &-\kappa \int_0^1 \int_0^\alpha (F(v)-F(u+v))^2 \frac{1}{u}\mathfrak{f}'(v/\alpha)dudv \\ &+O\left(V(z)^{-1}\alpha\frac{\log\log z}{\log z}\right). \end{aligned}$$

Applying the condition that $F(v) = 0$ for $v > 1$, we have

$$\begin{aligned} \tfrac{\alpha}{c}V(z)J^2V(D,z) &= \int_0^1 F^2(v)\mathfrak{f}'(v/\alpha)dv \\ &-\kappa \int_{1-\alpha}^1 \int_0^{1-v} (F(v)-F(u+v))^2 \frac{1}{u}\mathfrak{f}'(v/\alpha)dudv \\ &-\kappa \int_0^{1-\alpha} \int_0^\alpha (F(v)-F(u+v))^2 \frac{1}{u}\mathfrak{f}'(v/\alpha)dudv \\ &-\kappa \int_{1-\alpha}^1 \int_{1-v}^\alpha (F(v))^2 \frac{1}{u}\mathfrak{f}'(v/\alpha)dudv \\ &+O\left(V(z)^{-1}\alpha\frac{\log\log z}{\log z}\right). \end{aligned}$$

Definition 2.4.1. *For $\alpha = 2/(s-1)$, we define*

$$\begin{aligned} T_F(s) &= \int_0^1 F^2(v)\mathfrak{f}'(v/\alpha)dv \\ &-\kappa \int_{1-\alpha}^1 \int_0^{1-v} (F(v)-F(u+v))^2 \frac{1}{u}\mathfrak{f}'(v/\alpha)dudv \\ &-\kappa \int_0^{1-\alpha} \int_0^\alpha (F(v)-F(u+v))^2 \frac{1}{u}\mathfrak{f}'(v/\alpha)dudv \\ &-\kappa \int_{1-\alpha}^1 \int_{1-v}^\alpha (F(v))^2 \frac{1}{u}\mathfrak{f}'(v/\alpha)dudv. \end{aligned}$$

With this definition, we have

Proposition 2.4.2. *Let F be a continuous piecewise smooth function with $(F(0) - F(u))^2 \ll u$ for all $0 \leq u \leq 1$ and $F(v) = 0$ for $v > 1$. Assume $T_F(s)$ as defined above is positive. Then there is some z_0 such that if $z > z_0$, then $V(D,z)$ is also positive.*

Proof. As $z \to \infty$, the error term above approaches zero and the main term is positive as stated. $\square$

2.4.1 Choice of $F(t)$

We have $V(D, z)$ positive if $T_F(s)$ is positive. Now we wish to examine different choices for the function F. We would like to find the minimum of s such that $T_F(s)$ is positive for some function F. To do so, we will look at various families of functions. We recall that the requirement on F is that it is continuous, piecewise smooth and satisfies $(F(0) - F(u))^2 \ll u$ for all $0 \leq u \leq 1$. For a particular function F, we let $\beta_\kappa(F)$ be the minimum s such that $T_F(s)$ is positive in the case of sifting dimension κ. For ease of notation we will denote $\beta_\kappa(F)$ by β_κ where F is understood. We note that $F(t) = (1-t)^m + c$ for any $m > 1/2$ and constant c satisfies the conditions on F.

With the choice of $F(t) = 1 - t$, we have

$$\beta_5 < 10.76$$

$$\beta_4 < 8.7499$$

$$\beta_{3.5} < 7.81.$$

With the choice of $F(t) = (1-t)^{0.7}$ we have

$$\beta_3 < 6.6125.$$

All of these sifting limits are smaller than the sifting limits given by the Diamond-Halberstam sieve. We note that all computations were done using Maple math software.

We would also like to compare this result to the sieve given in section (2.3). To find asymptotics of the sifting limit, Selberg chose y_d corresponding to the choice of $F(t)$ given in (2.9). The sifting limit using $F(t) = 1 - t$ is much better than this choice of Selberg. For example for $\kappa = 3$, the function $F(t) = 1 - t$ gives us

$$\beta_3 < 6.75$$

while Selberg's choice gives us

$$\beta_3 < 7.24.$$

Although this choice of Selberg is not the best choice for small sifting dimension, his choice does give us some insight into the problem. He chose a piecewise defined function for F with a break at $t = 1-\alpha$. We note that this is a natural choice due to the limits of integration in the definition of $T_F(s)$. We now follow this example, but keep the definition of F very general to allow us more freedom. We define

$$F(t) = \begin{cases} F_1(t) & \text{if } 0 \le t \le 1-\alpha \\ \dfrac{F_1(1-\alpha)}{F_2(1-\alpha)} F_2(t) & \text{if } 1-\alpha < t \le 1 \end{cases} \tag{2.10}$$

where $F_1(t)$ and $F_2(t)$ are continuous, piecewise monotonic functions such that $(F_1(0) - F_1(w))^2 \ll w$ and $(F_2(0) - F_2(w))^2 \ll w$ for $0 \le w \le 1$. Then $F(t)$ is a continuous piecewise monotonic function that satisfies $(F(0) - F(w))^2 \ll w$ for $0 \le w \le 1$. We also note that the simple case of a single function F follows when $F_1 = F_2$. Using this definition of $F(t)$, we find that $T_F(s)$ is:

$$\begin{aligned} T_F(s) &= \int_0^{1-\alpha} F_1(v)^2 \mathfrak{f}'(v/\alpha) dv \\ &+ \frac{F_1(1-\alpha)^2}{F_2(1-\alpha)^2} \int_{1-\alpha}^1 F_2(v)^2 \mathfrak{f}'(v/\alpha) dv \\ &- \kappa \frac{F_1(1-\alpha)^2}{F_2(1-\alpha)^2} \int_{1-\alpha}^1 \int_{1-v}^{\alpha} (F_2(v))^2 \frac{1}{u} \mathfrak{f}'(v/\alpha) du dv \\ &- \kappa \frac{F_1(1-\alpha)^2}{F_2(1-\alpha)^2} \int_{1-\alpha}^1 \int_0^{1-v} (F_2(v) - F_2(u+v))^2 \frac{1}{u} \mathfrak{f}'(v/\alpha) du \\ &- \kappa \int_{1-2\alpha}^{1-\alpha} \int_{1-\alpha-v}^{\alpha} \left(F_1(v) - \frac{F_1(1-\alpha)}{F_2(1-\alpha)} F_2(u+v) \right)^2 \frac{1}{u} \mathfrak{f}'(v/\alpha) du dv \\ &- \kappa \int_{1-2\alpha}^{1-\alpha} \int_0^{1-\alpha-v} (F_1(v) - F_1(u+v))^2 \frac{1}{u} \mathfrak{f}'(v/\alpha) du dv \\ &- \kappa \int_0^{1-2\alpha} \int_0^{\alpha} (F_1(v) - F_1(u+v))^2 \frac{1}{u} \mathfrak{f}'(v/\alpha) du dv. \end{aligned}$$

This does give us more improvements. For $\kappa = 3$, we chose $F_1(t) = (1-t)^{0.83}$ and $F_2(t) = (1-t)^{0.57}$. With this choice, we have:

$$\beta_3 < 6.576.$$

With the choice of $F(t) = (1-t)^{0.7}$, we had

$$\beta_3 < 6.6125.$$

Another choice of F suggested by Selberg [5, p.482] is $F(t) = 1 - t + c$ where c is a constant dependent on κ.

With $c = 0.1$, we have

$$\beta_3 < 6.5206.$$

With $c = 0.07$, we have

$$\beta_4 < 8.53.$$

By using the piecewise defined F with $F_1(t) = (1-t)^{0.86}$ and $F_2(t) = (1-t)^{0.98} + 0.1$ we achieve

$$\beta_3 < 6.51998.$$

which is an improvement over Selberg's method presented in section (2.3).

2.5 Further Generality

In the previous section, we considered the sieve $\Lambda^-\Lambda^+$ where Λ^+ was Selberg's Λ^2 sieve and $\Lambda^- = (\lambda_q^-)$ was given by $\lambda_1^- = 1$, $\lambda_p^- = -1$ for $p \leq z$, and $\lambda_q^- = 0$ otherwise. Now, we would like to consider a more general lower bound sieve Λ^-. We let $\Lambda^- = (\lambda_q^-)$ supported on $q \leq z$ and we let $\Lambda^+ = \Lambda^2 = (\rho_d)^2$ be Selberg's Λ^2 sieve in terms of ρ_d with support $d \leq \sqrt{D/z} = Y$.

Proposition 2.5.1. *With $\Lambda = \Lambda^-\Lambda^2 = (\lambda_d)$, we have*

$$V(D,z) = \sum_{\substack{d \leq D \\ d|P(z)}} g(d)\lambda_d = \sum_{\substack{q \leq z \\ q|P(z)}} \lambda_q^- g(q) \sum_{\substack{d \leq \sqrt{D/z} \\ d|P(z)}} h(d) \left(\sum_{c|q} \mu(c) y_{cd} \right)^2 \tag{2.11}$$

where

$$y_d = \frac{\mu(d)}{h(d)} \sum_{\substack{m \equiv 0 (\mathrm{mod}\ d) \\ m|P(z)}} g(m)\rho_m. \tag{2.12}$$

Proof. We first write λ_d in terms of λ_q^- and ρ_d.

$$\begin{aligned}
\textstyle\sum_{n\le x} a_n \sum_{\substack{d|n\\ d|P(z)}} \lambda_d &= \sum_{n\le x} a_n \Bigg(\sum_{\substack{q|n\\ q|P(z)}} \lambda_q^-\Bigg)\Bigg(\sum_{\substack{d|n\\ d|P(z)}} \rho_d\Bigg)^2\\
&= \sum_{n\le x} a_n \sum_{\substack{q|n\\ q|P(z)}} \lambda_q^- \sum_{\substack{d_1|n\\ d_1|P(z)}} \rho_{d_1} \sum_{\substack{d_2|n\\ d_2|P(z)}} \rho_{d_2}\\
&= \sum_{\substack{d_1\le Y\\ d|P(z)}} \rho_{d_1} \sum_{\substack{d_2\le Y\\ d_2|P(z)}} \rho_{d_2} \sum_{\substack{q\le z\\ q|P(z)}} \lambda_q^- \sum_{n\equiv 0 (\text{mod } [q,d_1,d_2])} a_n.
\end{aligned}$$

Therefore,

$$\begin{aligned}
V(D,z) &= \sum_{\substack{d\le\sqrt{D}\\ d|P(z)}} g(d)\lambda_d\\
&= \sum_{\substack{d_1\le Y\\ d_1|P(z)}} \rho_{d_1} \sum_{\substack{d_2\le Y\\ d_2|P(z)}} \rho_{d_2} \sum_{\substack{q\le z\\ q|P(z)}} \lambda_q^- g([d_1,d_2,q])\\
&= \sum_{\substack{q\le z\\ q|P(z)}} \lambda_q^- g(q) \sum_{\substack{d_1\le Y\\ d_1|P(z)}} \rho_{d_1} \sum_{\substack{d_2\le Y\\ d_2|P(z)}} \rho_{d_2} g\left(\frac{[d_1,d_2,q]}{q}\right).
\end{aligned}$$

Now we turn to the right-hand side of equation (2.11). We only need to show that

$$\sum_{\substack{d_1\le Y\\ d_1|P(z)}} \rho_{d_1} \sum_{\substack{d_2\le Y\\ d_2|P(z)}} \rho_{d_2} g\left(\frac{[d_1,d_2,q]}{q}\right) = \sum_{\substack{d\le Y\\ d|P(z)}} h(d)\Bigg(\sum_{c|q} \mu(c) y_{cd}\Bigg)^2.$$

Applying the definition of y_{cd} (2.12), we find

$$\begin{aligned}
\textstyle\sum_{\substack{d\le Y\\ d|P(z)}} h(d)\Big(\sum_{c|q}\mu(c)y_{cd}\Big)^2 &= \sum_{\substack{d\le Y\\ d|P(z)}} \frac{1}{h(d)}\Bigg(\sum_{c|q}\frac{1}{h(c)} \sum_{\substack{m\equiv 0 (\text{mod } cd)\\ m|P(z)}} g(m)\rho_m\Bigg)^2\\
&= \sum_{\substack{d\le Y\\ d|P(z)}} \frac{1}{h(d)}\Bigg(\sum_{\substack{m\equiv 0 (\text{mod } d)\\ m|P(z)}} g(m)\rho_m \sum_{c|(q,m)} \frac{1}{h(c)}\Bigg)^2\\
&= \sum_{\substack{d\le Y\\ d|P(z)}} \frac{1}{h(d)}\Bigg(\sum_{\substack{m\equiv 0 (\text{mod } d)\\ m|P(z)}} g(m)\rho_m \frac{1}{g((q,m))}\Bigg)^2.
\end{aligned}$$

Now we expand the square.

$$\begin{aligned}
&\sum_{\substack{d\leq Y\\ d|P(z)}} h(d)\Big(\sum_{c|q}\mu(c)y_{cd}\Big)^2\\
&= \sum_{\substack{d\leq Y\\ d|P(z)}} \frac{1}{h(d)} \sum_{\substack{m\equiv 0 (\mathrm{mod}\ d)\\ m|P(z)}} g(m)\rho_m \frac{1}{g((q,m))} \sum_{\substack{n\equiv 0 (\mathrm{mod}\ d)\\ n|P(z)}} g(n)\rho_n \frac{1}{g((q,n))}\\
&= \sum_{\substack{m\leq Y\\ m|P(z)}} \rho_m \sum_{\substack{n\leq Y\\ n|P(z)}} \rho_n \frac{g(m)g(n)}{g((q,m))g((q,n))} \sum_{d|(m,n)} \frac{1}{h(d)}\\
&= \sum_{\substack{m\leq Y\\ m|P(z)}} \rho_m \sum_{\substack{n\leq Y\\ n|P(z)}} \rho_n \frac{g(m)g(n)}{g((q,m))g((q,n))g((m,n))}\\
&= \sum_{\substack{m\leq Y\\ m|P(z)}} \rho_m \sum_{\substack{n\leq Y\\ n|P(z)}} \rho_n g\Big(\frac{[m,n,q]}{q}\Big).
\end{aligned}$$

Therefore,

$$V(D,z) = \sum_{\substack{d\leq\sqrt{D}\\ d|P(z)}} g(d)\lambda_d = \sum_{\substack{q\leq z\\ q|P(z)}} \lambda_q^- g(q) \sum_{\substack{d\leq Y\\ d|P(z)}} h(d)\Big(\sum_{c|q}\mu(c)y_{cd}\Big)^2.$$

□

2.5.1 Choice of Λ^-

We now consider specific choices for Λ^-. Previously, we chose $\lambda_1^- = 1$, $\lambda_p^- = -1$ if $p|P(z)$ and $\lambda_d^- = 0$ otherwise. We can instead make the following choice:

Lemma 2.5.2. *Let $\lambda_1 = 1$, $\lambda_p = -1$ for $p|P(z)$, $\lambda_{p_1p_2} = 1$ for $p_2 < p_1 \leq z^{1/3}$, $\lambda_{p_1p_2p_3} = -1$ for $p_3 < p_2 < p_1 \leq z^{1/3}$, and $\lambda_d = 0$ otherwise, where $p_1p_2p_3|P(z)$. Then $\Lambda_d = \{\lambda_d\}$ is a lower bound sieve of level z.*

Proof. We first note that $\lambda_1 = 1$. Also, the sieve is of level z because $\lambda_d = 0$ for $d \geq z$.

We have

$$\sum_{\substack{d|n\\ d|P(z)}} \lambda_d \leq \sum_{\substack{d|n^*\\ d|P(z)}} \lambda_d$$

where n^* is the (squarefree) part of n with all prime divisors $\leq z^{1/3}$ since the rest contributes a non-positive amount. Let $n^* = p_1 \cdots p_m$. Then

$$\sum_{d|n^*} \lambda_d = 1 - \binom{m}{1} + \binom{m}{2} - \binom{m}{3} = -\binom{m-1}{3} \leq 0.$$

The last equality follows from the identities

$$\binom{m}{3} = \binom{m-1}{3} + \binom{m-1}{2}, \text{ and } \binom{m}{2} = \binom{m-1}{2} + \binom{m-1}{1}.$$

□

We let $Y = \sqrt{D/z}$. Then according to Proposition (2.5.1) and Lemma (2.5.2), we have

$$\begin{aligned} V(D,z) &= \sum_{\substack{d \leq Y \\ d|P(z)}} h(d) y_d^2 \\ &- \sum_{\substack{p_1 \leq z \\ p_1|P(z)}} g(p_1) \sum_{\substack{d \leq Y \\ d|P(z)}} h(d)(y_d - y_{p_1 d})^2 \\ &+ \sum_{\substack{p_2 < p_1 \leq z^{1/3} \\ p_1 p_2|P(z)}} g(p_1) g(p_2) \sum_{\substack{d \leq Y \\ d|P(z)}} h(d) \Big(y_d - y_{p_1 d} - y_{p_2 d} + y_{p_1 p_2 d} \Big)^2 \\ &- \sum_{\substack{p_3 < p_2 < p_1 \leq z^{1/3} \\ p_1 p_2 p_3|P(z)}} g(p_1) g(p_2) g(p_3) \sum_{\substack{d \leq Y \\ d|P(z)}} h(d) \begin{pmatrix} y_d - y_{p_1 p_2 p_3 d} \\ -y_{p_1 d} - y_{p_2 d} - y_{p_3 d} \\ +y_{p_1 p_2 d} + y_{p_1 p_3 d} + y_{p_2 p_3 d} \end{pmatrix}^2 . \end{aligned}$$

The above is a logical choice of lower bound sieve. However, Selberg presented another choice of lower bound sieve that will give more flexibility and in fact better results.

Lemma 2.5.3. *Let T be a positive integer. Let $\lambda_1 = 1$, $\lambda_p = -1$ for $p|P(z)$, $\lambda_{p_1 p_2} = (4T-2)/T(T+1)$ for $p_2 < p_1 \leq z^{1/3}$, $\lambda_{p_1 p_2 p_3} = -6/T(T+1)$ for $p_3 < p_2 < p_1 \leq z^{1/3}$ and $\lambda_d = 0$ otherwise, where $p_1 p_2 p_3 | P(z)$. Then $\Lambda_d = \{\lambda_d\}$ is a lower bound sieve of level z.*

Proof. We first note that $\lambda_1 = 1$. Also, the sieve is of level z because $\lambda_d = 0$ for $d \geq z$.

We have

$$\sum_{\substack{d|n \\ d|P(z)}} \lambda_d \leq \sum_{\substack{d|n^* \\ d|P(z)}} \lambda_d$$

where n^* is the (squarefree) part of n with all prime divisors $\leq z^{1/3}$ since the rest contributes a non-positive amount. Let $n^* = p_1 \cdots p_m$.

$$\begin{aligned}
\sum_{\substack{d|n^* \\ d|P(z)}} \lambda_d &= 1 - \binom{m}{1} + \frac{4T-2}{T(T+1)}\binom{m}{2} - \frac{6}{T(T+1)}\binom{m}{3} \\
&= -\frac{(m-1)(T-m)(T-(m-1))}{T(T+1)} \leq 0
\end{aligned}$$

since T is an integer and $m \geq 1$. We do note that if T is not an integer, this condition does not hold. □

With this choice of Λ^- for some integer T, we have

$$\begin{aligned}
&V(D,z) \\
&= \sum_{\substack{d\leq Y \\ d|P(z)}} h(d) y_d^2 \\
&- \sum_{\substack{p_1\leq z \\ p_1|P(z)}} g(p_1) \sum_{\substack{d\leq Y \\ d|P(z)}} h(d)(y_d - y_{p_1 d})^2 \\
&+ \frac{4T-2}{T(T+1)} \sum_{\substack{p_2<p_1\leq z^{1/3} \\ p_1p_2|P(z)}} g(p_1)g(p_2) \sum_{\substack{d\leq Y \\ d|P(z)}} h(d)(y_d - y_{p_1 d} - y_{p_2 d} + y_{p_1p_2 d})^2 \\
&- \frac{6}{T(T+1)} \sum_{\substack{p_3<p_2<p_1\leq z^{1/3} \\ p_1p_2p_3|P(z)}} g(p_1)g(p_2)g(p_3) \sum_{\substack{d\leq Y \\ d|P(z)}} h(d) \begin{pmatrix} y_d - y_{p_1p_2p_3 d} + y_{p_1p_2 d} \\ -y_{p_1 d} - y_{p_2 d} - y_{p_3 d} \\ +y_{p_1p_3 d} + y_{p_2p_3 d} \end{pmatrix}^2 .
\end{aligned}$$

2.5.2 Lower Bound Sieve with Three Primes

In this section, we present the sifting limit arguments for the lower bound sieve $\Lambda = \Lambda^-\Lambda^2$ where Λ^- is given by $\lambda_1 = 1$, $\lambda_p = -1$ for $p|P(z)$, $\lambda_{p_1p_2} = (4T-2)/T(T+1)$ for $p_2 < p_1 \leq z^{1/3}$ and $\lambda_{p_1p_2p_3} = -6/T(T+1)$ for $p_3 < p_2 < p_1 \leq z^{1/3}$, and $\lambda_d = 0$ otherwise, where $p_1p_2p_3|P(z)$. We again make the following choice of y_d.

$$y_d = \begin{cases} J^{-1} F\left(\dfrac{\log d}{\log Y}\right) & \text{if } 1 \leq d \leq Y \\ 0 & \text{otherwise} \end{cases} \tag{2.13}$$

where $Y = \sqrt{D/z}$.

To make notation easier, we introduce some new variables. We let $v = \log d/\log Y$, $u = \log p_1/\log Y$, $w = \log p_2/\log Y$ and $x = \log p_3/\log Y$. With this notation, we see that

$$\begin{aligned}
J^2V(D,z) &= \sum_{\substack{d\le Y\\ d|P(z)}} h(d)F(v)^2 \\
&- \sum_{p_1\le z} g(p_1)\sum_{d\le Y} h(d)\left(F(v)-F(u+v)\right)^2 \\
&+\frac{4T-2}{T(T+1)}\sum_{p_2<p_1\le z^{1/3}} g(p_1)g(p_2)\sum_{d\le Y} h(d)\left(\begin{smallmatrix} F(v)+F(u+w+v)\\ -F(u+v)-F(w+v)\end{smallmatrix}\right)^2 \\
&-\frac{6}{T(T+1)}\sum_{p_3<p_2<p_1\le z^{1/3}} g(p_1)g(p_2)g(p_3)\sum_{d\le Y}\begin{pmatrix} F(v)-F(u+v)\\ -F(w+v)-F(x+v)\\ +F(u+w+v)\\ +F(u+x+v)\\ +F(w+x+v)\\ -F(u+w+x+v)\end{pmatrix}^2 .
\end{aligned}$$

We now apply the same arguments as in section (2.4) to this expression in order to analyze the sums. We employ the notation that $\alpha = \log z/\log Y$. We also let $\beta = \alpha/3 = \log z^{1/3}/\log Y$, so that $x < w < u \le \beta$ in the second and third sums. Finally, we note that $d \le Y$ means $v \le 1$. Therefore,

$$\begin{aligned}
&\frac{\alpha}{c}V(z)J^2V(D,z)\\
&= \int_0^1 F(v)^2\mathfrak{f}'(v/\alpha)dv\\
&-\kappa\int_0^1\int_0^\alpha (F(v)-F(u+v))^2\mathfrak{f}'(v/\alpha)dudv\\
&+\kappa^2\frac{4T-2}{T(T+1)}\int_0^1\int_0^\beta\int_w^\beta\left(\begin{smallmatrix} F(v)+F(u+w+v)\\ -F(u+v)-F(w+v)\end{smallmatrix}\right)^2\mathfrak{f}'\left(\frac{v}{\alpha}\right)\frac{du}{u}\frac{dw}{w}dv\\
&-\kappa^3\frac{6}{T(T+1)}\int_0^1\int_0^\beta\int_x^\beta\int_w^\beta\begin{pmatrix} F(v)-F(u+w+x+v)\\ -F(u+v)+F(u+w+v)\\ -F(w+v)+F(u+x+v)\\ -F(x+v)+F(w+x+v)\end{pmatrix}^2\mathfrak{f}'\left(\frac{v}{\alpha}\right)\frac{du}{u}\frac{dw}{w}\frac{dx}{x}dv\\
&+O\left(\frac{\alpha}{c}\frac{\log\log z}{\log z}\right).
\end{aligned}$$

We again define $T_F(s)$ to be the right-hand side of the above expression without the error term. Then we have $V(D,z) > 0$ for large enough z if s is chosen such that $T_F(s) > 0$.

We now turn to the analysis of $T_F(s)$.

2.5.3 Analysis of $T_F(s)$

This section describes the computation of $T_F(s)$. The computations are tedious and long, so we only include highlights of the computations. The results from these computations can be found in the next section.

The first step in computing $T_F(s)$ is to apply the condition that $F(v) = 0$ for $v > 1$. We make the following definitions:

$$\begin{aligned} B &= \{0 \le v \le 1, 0 \le u \le \alpha\} \\ B_1 &= \{(u,v) \in B : u + v \le 1\} \\ B_2 &= \{(u,v) \in B : u + v > 1\} \end{aligned}$$

$$\begin{aligned} C &= \{0 \le v \le 1, 0 \le w \le \beta, w \le u \le \beta\} \\ C_1 &= \{(u,w,v) \in C : u + v + w \le 1\} \\ C_2 &= \{(u,w,v) \in C : u + v + w \ge 1, u + v \le 1\} \\ C_3 &= \{(u,w,v) \in C : u + v \ge 1, w + v \le 1\} \\ C_4 &= \{(u,w,v) \in C : w + v \ge 1\} \end{aligned}$$

$$\begin{aligned} D &= \{0 \le v \le 1, 0 \le x \le \beta, x \le w \le \beta, w \le u \le \beta\} \\ D_1 &= \{(u,w,x,v) \in D : x + v \ge 1\} \\ D_2 &= \{(u,w,x,v) \in D : x + v \le 1, w + v \ge 1\} \\ D_3 &= \{(u,w,x,v) \in D : w + v \le 1, u + v \ge 1, x + w + v \ge 1\} \\ D_4 &= \{(u,w,x,v) \in D : x + w + v \le 1, u + v \ge 1, x + w \le u\} \\ D_5 &= \{(u,w,x,v) \in D : u + v \le 1, x + w + v \ge 1, x + w \ge u\} \\ D_6 &= \{(u,w,x,v) \in D : u + v \le 1, x + w + v \le 1, x + u + v \ge 1\} \\ D_7 &= \{(u,w,x,v) \in D : x + u + v \le 1, w + u + v \ge 1\} \\ D_8 &= \{(u,w,x,v) \in D : w + u + v \le 1, x + w + u + v \ge 1\} \\ D_9 &= \{(u,w,x,v) \in D : x + w + u + v \le 1\}. \end{aligned}$$

Then we have

$$T_F(s) \quad + T_A - \kappa T_B + \kappa^2 \frac{4T-2}{T(T+1)} T_C - \kappa^3 \frac{6}{T(T+1)} T_D$$

where

$$T_A = \int_0^1 F(v)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) dv$$

$$\begin{aligned} T_B &= \int_{B_1} (F(v) - F(u+v))^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u} dv \\ &+ \int_{B_2} F(v)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u} dv \end{aligned}$$

$$\begin{aligned} T_C &= \int_{C_1} \big(F(v) - F(u+v) - F(w+v) + F(u+w+v)\big) \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u} \frac{dw}{w} dv \\ &+ \int_{C_2} \big(F(v) - F(u+v) - F(w+v)\big)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u} \frac{dw}{w} dv \\ &+ \int_{C_3} \big(F(v) - F(w+v)\big)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u} \frac{dw}{w} dv \\ &+ \int_{C_4} F(v)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u} \frac{dw}{w} dv \end{aligned}$$

$$\begin{aligned} T_D &= \int_{D_1} F(v)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u} \frac{dw}{w} \frac{dx}{x} dv \\ &+ \int_{D_2} \big(F(v) - F(x+v)\big)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u} \frac{dw}{w} \frac{dx}{x} dv \\ &+ \int_{D_3} \big(F(v) - F(x+v) - F(w+v)\big)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u} \frac{dw}{w} \frac{dx}{x} dv \\ &+ \int_{D_4} \big(F(v) - F(x+v) - F(w+v) + F(x+w+v)\big)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u} \frac{dw}{w} \frac{dx}{x} dv \\ &+ \int_{D_5} \big(F(v) - F(x+v) - F(w+v) - F(u+v)\big)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u} \frac{dw}{w} \frac{dx}{x} dv \\ &+ \int_{D_6} \big(F(v) - F(x+v) - F(w+v) - F(u+v) + F(x+w+v)\big)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u} \frac{dw}{w} \frac{dx}{x} dv \\ &+ \int_{D_7} \left(\begin{array}{l} F(v) - F(x+v) - F(w+v) - F(u+v) \\ +F(x+w+v) + F(x+u+v) \end{array} \right)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u} \frac{dw}{w} \frac{dx}{x} dv \\ &+ \int_{D_8} \left(\begin{array}{l} F(v) - F(x+v) - F(w+v) - F(u+v) \\ +F(x+w+v) + F(x+u+v) + F(w+u+v) \end{array} \right)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u} \frac{dw}{w} \frac{dx}{x} dv \\ &+ \int_{D_9} \left(\begin{array}{l} F(v) - F(x+v) - F(w+v) - F(u+v) + F(x+w+v) \\ +F(x+u+v) + F(w+u+v) - F(x+w+u+v) \end{array} \right)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u} \frac{dw}{w} \frac{dx}{x} dv. \end{aligned}$$

Now we can simply use the definition of $F(t)$ on the interval $0 \le t \le 1$ instead of a piecewise definition. The next step is to write the sets B_i, C_i and D_i as unions of sets such that the variables are in intervals. For example,

$$B_1 = \{1 - \alpha \le v \le 1, 0 \le w \le 1 - v\} \cup \{0 \le v \le 1 - \alpha, 0 \le w \le \alpha\}.$$

We do not include the decomposition of each set here, since the decompositions are quite complicated. For example, D_8 decomposes into 22 such sets.

Ideally, once we have the decompositions we would like to be able to input $T_F(s)$ into a math software program to compute its value and leave the choice of $F(t)$ open. Unfortunately, the integrals over the C_i and D_i are too difficult for typical math software to handle. Hence, we now make a general choice of $F(t) = 1-t+c$ where c is a constant to be chosen later. With this choice of F, we can now simplify some of the integrals by hand using changes of variables and by changing the order of integration. The math software does not use these techniques. We simplify the integrals as much as necessary in order for the software to be able to process the rest. In some cases, we estimate some of the integrals in order to make computation time feasible. In two cases, we use the estimate that $-\ln(1-u) \leq (4\ln 2 - 2)u^2 + u$ for $u \in [0, 1/2]$.

We note that our choice of $F(t) = 1 - t + c$ simplifies some of our computations quite a bit. For example, over C_1, $F(v) - F(u+v) - F(w+v) + F(u+w+v) = 0$. Over D_4, $F(v) - F(x+v) - F(w+v) + F(x+w+v) = 0$. Finally, over D_9, $F(v) - F(x+v) - F(w+v) - F(u+v) + F(x+w+v) + F(x+u+v) + F(w+u+v) - F(x+w+u+v) = 0$. Applying our definition of $F(t)$, we have

$$T_A = \int_0^1 (1+c-v)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) dv$$

$$\begin{aligned} T_B &= \int_{B_1} u\mathfrak{f}'\left(\frac{v}{\alpha}\right) du dv \\ &+ \int_{B_2} (1+c-v)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u} dv \end{aligned}$$

$$\begin{aligned} T_C &= \int_{C_2} (1+c-u-v-w)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u}\frac{dw}{w} dv \\ &+ \int_{C_3} w\mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u} dw dv \\ &+ \int_{C_4} (1+c-v)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right) \frac{du}{u}\frac{dw}{w} dv \end{aligned}$$

$$
\begin{aligned}
T_D &= \int_{D_1} (1+c-v)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right)\frac{du}{u}\frac{dw}{w}\frac{dx}{x}dv \\
&+ \int_{D_2} x\mathfrak{f}'\left(\frac{v}{\alpha}\right)\frac{du}{u}\frac{dw}{w}dxdv \\
&+ \int_{D_3} (1+c-x-w-v)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right)\frac{du}{u}\frac{dw}{w}\frac{dx}{x}dv \\
&+ \int_{D_5} (2+2c-2v-x-w-u)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right)\frac{du}{u}\frac{dw}{w}\frac{dx}{x}dv \\
&+ \int_{D_6} (1+c-v-u)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right)\frac{du}{u}\frac{dw}{w}\frac{dx}{x}dv \\
&+ \int_{D_7} x\mathfrak{f}'\left(\frac{v}{\alpha}\right)\frac{du}{u}\frac{dw}{w}dxdv \\
&+ \int_{D_8} (1+c-v-u-w-x)^2 \mathfrak{f}'\left(\frac{v}{\alpha}\right)\frac{du}{u}\frac{dw}{w}\frac{dx}{x}dv.
\end{aligned}
$$

Once we have the integrals in a simple enough form, we use Maple software to numerically integrate the functions with error less than 10^{-10}. The results of these computations are in the next section.

2.6 New Sifting Limit Results

Let $\Lambda_1 = \Lambda^2\Lambda^-$ where $\Lambda = \{\rho_d\}$ and $\Lambda^- = \{\lambda_q\}$. We define λ_q by $\lambda_1 = 1$, $\lambda_p = -1$ for $p|P(z)$, $\lambda_{p_1p_2} = (4T-2)/T(T+1)$ for $p_2 < p_1 \le z^{1/3}$, $\lambda_{p_1p_2p_3} = -6/T(T+1)$ for $p_3 < p_2 < p_1 \le z^{1/3}$ and $\lambda_d = 0$ otherwise, where T is an integer to be determined later and $p_1p_2p_3|P(z)$.

In the notation of the previous sections, with the choice of $F(t) = 1-t+c$, we have the following results:

For $\kappa = 2$, with $c = 0.2214971799$ and $T = 16$, we have

$$\beta_2 < 4.45.$$

For $\kappa = 2.5$ with $c = 0.17$ and $T = 19$, we have

$$\beta_{2.5} < 5.455.$$

For $\kappa = 3$ with $c = 0.13$ and $T = 24$ we have

$$\beta_3 < 6.458.$$

For $\kappa = 4$ with $c = 0.11$ and $T = 31$ we have

$$\beta_4 < 8.47.$$

The sieve of Diamond and Halberstam gives a smaller sifting limit for $\kappa = 2$ and $\kappa = 2.5$. However, our sifting limit is very close to that of Diamond and Halberstam for $\kappa = 2.5$. In addition, the sifting limits for $\kappa = 3, 4$ are much smaller than those given by the Diamond-Halberstam sieve.

2.7 Analysis of Choice of Sieve

There are many choices for a lower bound sieve. In the previous sections, we have made the choice of $\Lambda_1 = \Lambda^- \Lambda^2$, a convolution of a lower bound sieve and Selberg's Λ^2 upper bound sieve. Our original choice of Λ^- was $\lambda_1 = 1$, $\lambda_p = -1$ for $p|P(z)$ and $\lambda_d = 0$ otherwise. This is a pretty weak lower bound sieve and yet our results have been very good. This is due to the power of Selberg's Λ^2 sieve.

For example, with this choice of Λ^- and $\Lambda = \{\rho_d\}$ we have

$$S(\mathcal{A}, x, z) = \sum_{n \leq x} a_n \Bigg(1 - \sum_{\substack{p|n \\ p|P(z)}} 1\Bigg)\Bigg(\sum_{\substack{d|n \\ d|P(z)}} \rho_d\Bigg)^2. \tag{2.14}$$

We would like $S(\mathcal{A}, x, z)$ to be a good estimate for the number of $n \leq x$ such that $(n, P(z)) = 1$. If n does not have any prime factors less than z, then a_n is counted with weight 1. If n has one prime factor less than z then it is counted with weight zero. The sieve comes into play when n has more than one prime factor less than z. In such a case, the lower bound sieve Λ^- gives a negative weight to a_n. However, when n is highly composite, the upper bound sieve Λ^2 gives a weight close to zero. Therefore, the overall weight for a_n is negative but small. Therefore, the crudeness of the lower bound sieve Λ^- is counterbalanced by the power of the upper bound sieve Λ^2.

When we changed our Λ^- to also address n with up to three prime factors, we saw an improvement in the sieve. However, the improvement was relatively small, so it is likely that further improvements to Λ^- will not have a noticeable effect on the sifting limit results. The tradeoff between better results and computational difficulty seems to make a Λ^- addressing five prime factors inadvisable.

It is possible that a better choice of Λ^- that still address three prime factors could be used. So far, we have used the choice of Λ^- with $\lambda_1 = 1$, $\lambda_p = -1$ for $p|P(z)$, $\lambda_{p_1p_2} = (4T-2)/T(T+1)$ for $p_2 < p_1 \leq z^{1/3}$, $\lambda_{p_1p_2p_3} = -6/T(T+1)$ for $p_3 < p_2 < p_1 \leq z^{1/3}$ and $\lambda_d = 0$, where T is an integer and $p_1p_2p_3|P(z)$.

This choice of Λ^- makes computations feasible. However, it does not address the case when n has two or three prime factors with one of the factors $p > z^{1/3}$. This accounts for a small portion of n, so we do not lose much. We can take this case into account though by choosing Λ^- with $\lambda_1 = 1$, $\lambda_p = -1$ for $p|P(z)$, $\lambda_{p_1p_2} = f_1(T)$ for $p_2 < p_1 \leq z$ such that $p_2^2p_1 \leq z$ and $\lambda_{p_1p_2p_3} = -f_2(T)$ for $p_3 < p_2 < p_1 \leq z$ with $p_2^2p_1 \leq z$, where $p_1p_2p_3|P(z)$ and $f_1(T), f_2(T)$ are some functions that make Λ^- a lower bound sieve of level z.

2.8 Appendix

In this appendix, we give some useful general lemmas for sifting limit calculations. These lemmas are referenced in the previous chapter. Throughout this appendix, we let $\kappa > 0$ and let $g(p)$ be a multiplicative function with the three following properties.

1. $\sum_{p\leq x} g(p)\log p = \kappa \log x + \delta(x)$ where $\delta(x)$ is bounded for $x \geq 2$.
2. $\prod_{w\leq p\leq z}(1+|g(p)|) \ll \left(\frac{\log z}{\log w}\right)^\kappa$ if $z > w \geq 2$.
3. $\sum_p g(p)^2 \log p < \infty$.

We note that these three conditions are satisfied by the multiplicative function $g(d)$ that gives the density of the masses a_n attached to $n \equiv 0 \mod d$ in sifting theory. The three conditions are also satisfied by the multiplicative function $h(d)$ defined by $h(p) = g(p)/(1-g(p))$.

Lemma 2.8.1. *We have*

$$\sum_{p\leq x} g(p) = \kappa \log\log x + b + O\left(\frac{1}{\log x}\right)$$

where b is a constant given in the proof.

Proof. We first write $\sum_{p\le x} g(p)\log p = L(x) = \kappa\log x + R(x)$ with $R(x) \ll 1$. We then have

$$\begin{aligned}
\textstyle\sum_{p\le x} g(p) &= \int_{2^-}^{x} (\log u)^{-1} dL(u) \\
&= \kappa \int_{2^-}^{x} \frac{1}{\log u} d\log u + \int_{2^-}^{x} \frac{dR(u)}{\log u} \\
&= \kappa \int_{\log 2^-}^{\log x} \frac{du}{u} + R(u)\log u|_{2^-}^{x} + \int_{2^-}^{x} \frac{R(u)}{u(\log u)^2} du \\
&= \kappa \log\log x - \kappa\log\log 2 + \frac{R(x)}{\log x} + 1 \\
&+ \int_{2}^{\infty} \frac{R(u)}{u(\log u)^2} du - \int_{x}^{\infty} \frac{R(u)}{u(\log u)^2} du
\end{aligned}$$

where the 1 term appears because $R(2) = -\log 2/2$. We note that the third term $R(x)/\log x \ll 1/\log x$. Additionally, the final term is $O(1/\log x)$. Therefore,

$$\sum_{p\le x} g(p) = \kappa\log\log x + b + O(1/\log x)$$

where

$$b = 1 - \kappa\log\log 2 + \int_{2}^{\infty} \frac{R(u)}{u(\log u)^2} du$$

□

We now give a general theorem from Opera de Cribro [2, Theorem A.7]

Theorem 2.8.2.

$$\sum_{\substack{d|P(z)\\ d\le Y}} g(d) = W(z)\left\{c\mathfrak{f}(\log Y/\log z) + O((\log Y)^{-1})\right\} \tag{2.15}$$

where $\mathfrak{f}$ *is the solution to the differential difference equation:*

$$\begin{cases} \mathfrak{f}(t) = t^{\kappa} & \text{if } 0 < t \le 1, \\ t\mathfrak{f}'(t) = \kappa(\mathfrak{f}(t) - \mathfrak{f}(t-1)) & \text{if } t > 1, \end{cases} \tag{2.16}$$

$$W(z) = \prod_{p<z} (1 + g(p)),$$

and

$$c = e^{-\gamma\kappa}\Gamma(\kappa+1)^{-1}.$$

With the above theorem, we can prove the following general lemma:

Lemma 2.8.3. *Let $\Phi(w)$ be a continuous piecewise smooth function with $\Phi(w) = O(w)$. Let $\alpha = \log z/\log Y$. Then*

$$\sum_{p\le z} g(p)\Phi\left(\frac{\log p}{\log Y}\right) = \kappa\int_0^\alpha \Phi(w)\frac{dw}{w} + O\left(\frac{\log\log z}{\log z}\right).$$

Proof.

$$\begin{aligned} S &= \sum_{p\le z} g(p)\Phi\left(\frac{\log p}{\log Y}\right) \\ &= \int_1^z \Phi\left(\frac{\log w}{\log Y}\right) d\sum_{p\le w} g(p) \\ &= \int_1^z \Phi\left(\frac{\log w}{\log Y}\right) d\log\log w dw + \int_1^z \Phi\left(\frac{\log w}{\log Y}\right) dR(w) \end{aligned}$$

where $R(w) \ll 1/\log 2w$. By partial summation,

$$\begin{aligned} S &= \int_1^z \Phi\left(\frac{\log w}{\log Y}\right) d\log\log w dw \\ &+O\left(\Phi\left(\frac{\log w}{\log Y}\right) R(w)|_1^z + \int_1^z R(w) d\left|\Phi'\left(\frac{\log w}{\log Y}\right)\right|\right). \end{aligned}$$

We note that since $\Phi(u) \ll u$, we have $\Phi'(\log w/\log Y) \ll 1/(w\log Y)$

$$\begin{aligned} S &= \int_0^\alpha \Phi(w)\frac{1}{w}dw + O\left((\log Y)^{-1} + \int_1^z \frac{dw}{w\log 2w}\right) \\ &= \int_0^\alpha \Phi(w)\frac{1}{w}dw + O\left(\frac{\log\log z}{\log z}\right). \end{aligned}$$

□

We need one more final lemma:

Lemma 2.8.4. *Let $\Phi(v)$ be a continuous, piecewise smooth function for $0 \le v \le 1$ with $\Phi(v) = \Phi(0) + O(v)$. Then*

$$\begin{aligned} c^{-1}W(z)^{-1}\sum_{\substack{d|P(z)\\ 1\le d\le Y}} g(d)\Phi\left(\tfrac{\log d}{\log Y}\right) &= \int_1^Y \Phi\left(\frac{\log w}{\log Y}\right) d\mathfrak{f}\left(\frac{\log w}{\log z}\right) \\ &+O\left(c^{-1}\frac{\log\log Y}{\log Y}\right) \end{aligned}$$

where $c, W(z)$, and $\mathfrak{f}$ are as in Theorem (2.8.2).

Proof. Define $\Phi_1(v) = \Phi(v) - \Phi(0)$. Then $\Phi_1(v) = O(v)$. Suppose that the result is true for $\Phi_1(v)$. Then

$$\begin{aligned} S &= \sum_{\substack{d|P(z) \\ 1\le d\le Y}} g(d)\Phi\left(\frac{\log d}{\log Y}\right) \\ &= \Phi(0)\sum_{\substack{d|P(z) \\ 1\le d\le Y}} g(d) + \sum_{\substack{d|P(z) \\ 1\le d\le Y}} g(d)\Phi_1(v). \end{aligned}$$

From Theorem (2.8.2), we have

$$\sum_{\substack{d\le Y \\ d|P(z)}} g(d) = cW(z)\mathfrak{f}\left(\frac{\log Y}{\log z}\right) + O\left(W(z)\frac{1}{\log 2Y}\right).$$

Then

$$\begin{aligned} S &= cW(z)\mathfrak{f}\left(\frac{\log Y}{\log z}\right)\Phi(0) + cW(z)\int_1^Y \Phi_1\left(\frac{\log w}{\log Y}\right)d\mathfrak{f}\left(\frac{\log w}{\log z}\right) \\ &+O\left(W(z)\frac{\log\log Y}{\log Y}\right) \\ &= cW(z)\int_1^Y \Phi(0)d\mathfrak{f}\left(\frac{\log w}{\log z}\right) + cW(z)\int_1^Y \Phi_1\left(\frac{\log w}{\log Y}\right)d\mathfrak{f}\left(\frac{\log w}{\log z}\right) \\ &+O\left(W(z)\frac{\log\log Y}{\log Y}\right) \\ &= cW(z)\int_1^Y \Phi_1\left(\frac{\log w}{\log Y}\right)d\mathfrak{f}\left(\frac{\log w}{\log z}\right) + O\left(W(z)\frac{\log\log Y}{\log Y}\right). \end{aligned}$$

We have now reduced the question to $\Phi(v)$ with $\Phi(v) = O(v)$.

$$\begin{aligned} S &= \sum_{\substack{d|P(z) \\ 1\le d\le Y}} g(d)\Phi\left(\frac{\log d}{\log Y}\right) \\ &= \int_1^Y \Phi\left(\frac{\log w}{\log Y}\right)d\sum_{\substack{d|P(z) \\ d\le w}} g(d) \\ &= cW(z)\int_1^Y \Phi\left(\frac{\log w}{\log Y}\right)d\mathfrak{f}\left(\frac{\log w}{\log z}\right) \\ &+O\left(W(z)\frac{\Phi(1)}{\log 2Y} + W(z)\int_1^Y \frac{1}{\log 2w}\left|d\Phi\left(\frac{\log w}{\log Y}\right)\right|\right). \end{aligned}$$

We note that $\Phi(v) = O(v)$ and $\Phi(v)$ piecewise smooth gives $\Phi'(v) \ll 1$, so

$$\begin{aligned} S &= cW(z) \int_1^Y \Phi\left(\frac{\log w}{\log Y}\right) d\mathfrak{f}\left(\frac{\log w}{\log z}\right) \\ &+O\left(\frac{W(z)}{\log Y} + \frac{W(z)}{\log Y} \int_1^Y \frac{dw}{w \log 2w}\right) \\ &= cW(z) \int_0^1 \Phi(v) d\mathfrak{f}\left(\frac{v}{\alpha}\right) + O\left(W(z) \frac{\log\log Y}{\log Y}\right). \end{aligned}$$

$\square$

Chapter 3

Brun-Titchmarsh Theorem

3.1 Introduction

A historic problem in analytic number theory is to bound the number of primes in an interval of a fixed length. We define $\pi(x)$ to be the number of primes $p \leq x$. If we wish to bound the number of primes in an interval of length y, we want to bound $\pi(x+y)-\pi(x)$. We want to find some function $F(y)$ that is independent of x such that

$$\pi(x+y) - \pi(x) < F(y)$$

for all x, when y is sufficiently large. A theorem with a bound of this type is called a Brun-Titchmarsh Theorem named for Viggo Brun and Edward Charles Titchmarsh. Brun first established a bound of this type in 1915 with

$$F(y) = \frac{c_1 y}{\log y}$$

with a constant $c_1 > 2$.

Then Titschmarsh established a bound for primes in arithmetic progressions. We define $\pi(x; a, k)$ to be the number of primes $p \leq x$ such that p is congruent to a modulo k. Then we want a function $G(y, k)$ independent of x and a such that

$$\pi(x + y; a, k) - \pi(x; a, k) < G(y, k)$$

for y sufficiently large. Titchmarsh proved this for

$$G(y, k) = \frac{c_2 y}{\phi(k) \log (y/k)}$$

with a constant $c_2 > 2$, where ϕ is Euler's totient function.

Various improvements of these bounds have been made over the years. Our focus is on Selberg's improvement using sieve methods. In section 22 of his Lectures on Sieves [4], Selberg established the following two bounds:

$$\pi(x+y) - \pi(x) < \frac{2y}{\log y + 2.8}$$
$$\pi(x+y; a, k) - \pi(x; a, k) < \frac{2y}{\phi(k)\Big(\log(y/k) + 2.8\Big)}.$$

Following Selberg's notes, we are able to modify his constructions and create a computer program to establish the improved bounds:

$$\pi(x+y) - \pi(x) < \frac{2y}{\log y + 2.8168}$$
$$\pi(x+y; a, k) - \pi(x; a, k) < \frac{2y}{\phi(k)\Big(\log(y/k) + 2.8168\Big)}.$$

These bounds will be proven using sieve methods in the next section.

3.2 Improvement of Brun-Titchmarsh Theorem

3.2.1 Preliminaries

In order to establish the bounds we desire, we first introduce a function θ_q, following Selberg's work.

Definition 3.2.1. *For q prime, we define Q to be the product of all primes $p \le q$. Then we define*

$$\theta_q = \max_{a,x} \left| x \prod_{p \le q} \left(1 - \frac{1}{p}\right) - \sum_{\substack{a \le n \le a+x \\ (n,Q)=1}} 1 \right|.$$

We note that the maximum exists by periodicity. We also note that $\prod_{p \le q}(1 - p^{-1}) = \phi(Q)/Q$.

In his Lectures on Sieves [4, Section 22], Selberg proved the following Lemma regarding θ_q:

Lemma 3.2.2 (Selberg)**.** *With θ_q as defined in (3.2.1), we have that θ_q grows faster than any power of q. Also $\theta_1 = 1$, $\theta_2 = 1$, $\theta_3 = 4/3$, $\theta_5 = 28/15$ and $\theta_7 = 106/35$.*

We can extend this result further by computing values of θ_q for larger q. Selberg computed the above values, and with a simple computer program we compute the following values.

Lemma 3.2.3 (B.). *With θ_q as defined in (3.2.1), we have that $\theta_1 = 1$, $\theta_2 = 1$, $\theta_3 = 4/3$, $\theta_5 = 28/15$, $\theta_7 = 106/35$, $\theta_{11} = 388/77$, $\theta_{13} = 7102/1001$, $\theta_{17} \approx 10.87759$ and $\theta_{19} \geq 16.96824$.*

Proof. To determine θ_q we may assume $0 \leq a \leq Q$ and $0 \leq x \leq Q$ due to the periodicity mod Q. Therefore, θ_q can always be found by inspecting a finite number of cases. In this way, we are able to compute the values in the lemma.

For the proof of the growth of θ_q, we refer the reader to section 22 of Selberg's Lectures on Sieves [4, Lemma].

Selberg then established the following technical lemma:

Lemma 3.2.4 (Selberg). *Let $\sigma(\rho)$ denote the sum of the divisors of ρ. Let $\phi(\rho)$ be Euler's function and once again let Q be the product of all primes $p \leq q$. Then for $z > 1$, we write:*

$$\begin{aligned}
\sum\nolimits_1 &= \sum_{\substack{\sigma(\rho)\leq z \\ (\rho,Q)=1}} \mu^2(\rho)\frac{\sigma(\rho)}{\phi(\rho)}\left(1 - \frac{\sigma(\rho)}{z}\right), \\
\sum\nolimits_2 &= \sum_{\substack{\sigma(\rho)\leq z \\ (\rho,Q)=1}} \mu^2(\rho)\frac{1}{\phi(\rho)}\left(1 - \frac{\sigma(\rho)}{z}\right) \\
\sum\nolimits_3 &= \sum_{\substack{\sigma(\rho)\leq z \\ (\rho,Q)=1}} \mu^2(\rho)\frac{1}{\phi(\rho)}\left(1 - \frac{\sigma(\rho)}{z}\right)^2.
\end{aligned}$$

Then

$$\begin{aligned}
\sum\nolimits_1 &= \frac{z}{2}\prod_{p\leq q}\left(1 - \frac{1}{p}\right) + O\left(ze^{-\sqrt{\log z}}\right), \\
\sum\nolimits_2 &= \prod_{p\leq q}\left(1 - \frac{1}{p}\right)\{\log z + \gamma + \kappa_1 + \kappa'(q) - 1\} + O\left(e^{-\sqrt{\log z}}\right) \\
\sum\nolimits_3 &= \prod_{p\leq q}\left(1 - \frac{1}{p}\right)\left\{\log z + \gamma + \kappa_1 + \kappa'(q) - \frac{3}{2}\right\} + O\left(e^{-\sqrt{\log z}}\right).
\end{aligned}$$

Here γ is Euler's constant, while

$$\begin{aligned} \kappa_1 &= \sum_p \left(\frac{\log p}{p-1} - \frac{\log (p+1)}{p} \right) \\ \kappa'(q) &= \sum_{p \le q} \frac{\log (p+1)}{p}. \end{aligned}$$

For a proof of this lemma, the reader is again referred to Section 22 of Selberg's Lectures on Sieves [4]. In the next section, we will use these preliminaries to prove the Brun-Titchmarsh Theorem.

3.2.2 Brun-Titchmarsh Theorem

We now wish to apply the preliminaries established in the previous section to another problem. Consider the interval $I_x = [b, b+x]$. We exclude one residue class for each prime less than z. We then look for an upper bound of the number of integers left. We choose z just large enough to get the smallest possible upper bound that the method allows. Without loss of generality, we can assume that the residue class is always represented by zero. Then we look at the case where we are excluding all n which are divisible by p for $p \le z$. We assume that we have already excluded the n that are divisible by $p \le q$. We then apply a Λ^2 sieve with P, the product of all primes p with $q+1 \le p \le z$ and $\Lambda = (\rho_d)$. We also recall that $\phi(Q)$ is the product of all primes $p \le q$. Using Lemma (3.2.2), we find

$$\sum_{\substack{b \le n \le b+x \\ (n,P)=1}} 1 \le \frac{\phi(Q)}{Q} x \sum_{d_1, d_2} \frac{(d_1, d_2)}{d_1 d_2} \rho_{d_1} \rho_{d_2} + \theta_q \left(\sum_d |\rho_d| \right)^2. \tag{3.1}$$

We now introduce new variables y_ρ according to the equation

$$\sum_{d \equiv 0 (\mathrm{mod} \rho)} \frac{\rho_d}{d} = \mu(\rho) \frac{y_\rho}{\phi(\rho)},$$

which implies

$$\sum_{\rho \equiv 0 (\mathrm{mod} d)} \frac{y_\rho}{\phi(\rho)} = \mu(d) \frac{\rho_d}{d}.$$

The first term of (3.1) in terms of these new variables is then

$$\frac{\phi(Q)}{Q} x \sum_{d_1, d_2} \frac{(d_1, d_2)}{d_1 d_2} \rho_{d_1} \rho_{d_2} = \frac{\phi(Q)}{Q} x \sum_{(\rho, Q)=1} \frac{y_\rho^2}{\phi(\rho)}.$$

In addition, we have the bound

$$\sum_{d} |\rho_d| \leq \sum_{(\rho,Q)=1} \frac{\sigma(\rho)}{\phi(\rho)} |y_\rho|.$$

From the normalization of $\rho_1 = 1$, we have the side condition:

$$\sum_{(\rho,Q)=1} \frac{y_\rho}{\phi(\rho)} = 1. \tag{3.2}$$

We lose nothing by assuming that the $y_\rho \geq 0$, so we may write the upper bound as

$$\sum_{\substack{a \leq n \leq a+x \\ (n,P)=1}} 1 \leq \frac{\frac{\phi(Q)}{Q} x \sum_{(\rho,Q)=1} \frac{y_\rho^2}{\phi(\rho)} + \theta_q \left(\sum_{(\rho,Q)=1} \frac{\sigma(\rho)}{\phi(\rho)} y_\rho \right)^2}{\left(\sum_{(\rho,Q)=1} \frac{y_\rho}{\phi(\rho)} \right)^2} \tag{3.3}$$

and we drop the condition (3.2).

We now wish to choose $y_\rho \geq 0$ to minimize the expression (3.3) with z chosen as large as possible. We recall that P is the product of all primes p with $q + 1 \leq p \leq z$. We follow Selberg's approach and choose

$$y_\rho = \begin{cases} 1 - \dfrac{\sigma(\rho)}{z} & \text{if } \sigma(\rho) < z, \\ 0 & \text{if } \sigma(\rho) \geq z. \end{cases}$$

Then the upper bound (3.3) becomes

$$\sum_{\substack{a \leq n \leq a+x \\ (n,P)=1}} 1 \leq \frac{\frac{\phi(Q)}{Q} x \sum_3 + \theta_q \left(\sum_1 \right)^2}{\left(\sum_2 \right)^2},$$

where $\sum_1, \sum_2$ and $\sum_3$ are defined in Lemma (3.2.4). We then use the product expansions given in the same lemma and write $z = e^u \sqrt{x}$, where u will be chosen later, independent of x. After some manipulation, we obtain the upper bound

$$\frac{2x}{\log x + 2\gamma + 2\kappa_1 + 2\kappa'(q) - 1 + 2u - \frac{\theta_q}{2} e^{2u} + O\left(e^{-\frac{1}{2}\sqrt{\log x}} \right)}.$$

The optimal choice of u is given by

$$e^{2u} = \frac{2}{\theta_q}.$$

With this choice of u, we have the upper bound

$$\frac{2x}{\log x + 2\gamma + 2\kappa_1 + 2\kappa'(q) - 2 + \log \frac{2}{\theta_q} + O\left(e^{-\frac{1}{2}\sqrt{\log x}}\right)}.$$

We can now choose q so that

$$2\kappa'(q) + \log \frac{2}{\theta_q}$$

is maximized. We know there is an optimal choice because by the Lemma (3.2.2), this expression tends to negative infinity as q grows without bound. By maximizing this expression, we maximize the constant denominator term

$$f(q) = 2\gamma + 2\kappa_1 + 2\kappa'(q) - 2 + \log \frac{2}{\theta_q}.$$

We approximate κ_1 by restricting its defining sum to the primes less than 300. We note that κ_1 is a sum of positive terms, so we have

$$\kappa_1 \geq \sum_{p \leq 300} \left(\frac{\log p}{p-1} - \frac{\log (p+1)}{p} \right) \geq 0.368582.$$

We now calculate the values of $f(q)$ for $q \leq 19$. We have the following values:

q	f(q)
1	0.584743117
2	1.683355406
3	2.319869574
5	2.700101125
7	2.810290556
11	2.75298288
13	2.816814684
17	2.729530369
19	$<$ 2.600232097

We see that the maximum of $f(q)$ occurs for $q = 13$. In Selberg's lectures, he chose $q = 7$ because that was the highest value he calculated. By calculating larger values of θ_q, we were able to find a larger value of $f(q)$.

We find that

$$z = \sqrt{e^{2u}x} = \sqrt{\frac{2x}{\theta_{13}}} = \sqrt{\frac{1001}{3551}x} < \sqrt{x}.$$

Thus if we exclude one residue class from I_x for each prime $p < \sqrt{x}$, there remain at most

$$\frac{2x}{\log x + 2.8168}$$

numbers if $x > x_0$ for some x_0. We know simply that such an x_0 exists. Finding the value x_0 is a much more difficult problem without much benefit.

Theorem 3.2.5. *We have*

$$\pi(x+y) - \pi(x) < \frac{2y}{\log y + 2.8168} \tag{3.4}$$

for $y > x_0$ and all $x > 0$. Similarly for $(a,k) = 1$, if $\pi(x;a,k)$ denotes the number of primes $p \leq x$ which are congruent to a modulo k, we have

$$\pi(x+y;a,k) - \pi(x;a,k) < \frac{2y}{\phi(k)\big(\log(y/k) + 2.8168\big)}, \tag{3.5}$$

for $y/k > x_0$.

Proof. Equation (3.4) follows easily. For equation (3.5), we note that sifting the section of the arithmetic progression that lies in an interval of length y is equivalent to sifting an interval of length y/k. We also note that the sifting range for the arithmetic progression does not include the primes dividing k. □

Therefore, we have an improvement of Selberg's version of the Brun-Titchmarsh theorem.

Chapter 4

Large Sieve Inequality

4.1 Introduction

Large sieve inequalities are a very general problem in analytic number theory. The topic was first introduced to address the problem of least quadratic non-residues by Linnik. The problem has changed shape quite a bit since then. In general, we consider a finite set $\mathcal{X}$ of "harmonics" that will solve some interesting equation. For each element $x \in \mathcal{X}$, we associate a sequence $(x(n))$. In a way, these are "Fourier coefficients." The large sieve problem then is to find a constant $C = C(\mathcal{X}, N) \geq 0$ such that the following "large sieve inequality" holds:

$$\sum_{x \in \mathcal{X}} \left| \sum_{n \leq N} a_n x(n) \right|^2 \leq C(\mathcal{X}, N) \|a\|_2^2 \tag{4.1}$$

for any complex numbers a_n, where $\|a\|_2$ is the L^2 norm of (a_n). That is, $\|a\|_2^2 = \sum |a_n|^2$.

There are two main forms of the large sieve inequality, the additive and the multiplicative. In the additive large sieve inequality, the harmonics are additive characters of $\mathbb{Z}$ with $x(n) = e(\alpha n)$ for some $\alpha \in \mathbb{R}$, where $e(x) = e^{2\pi i x}$. We will use the notation $e(x)$ throughout this chapter. In the multiplicative large sieve inequality, we take the harmonics to be Dirichlet characters. We only address the additive case here.

For the additive large sieve inequality, we are interested in estimating the trigonometric polynomials

$$S(\alpha) = \sum_n a_n e(\alpha n) \tag{4.2}$$

where the support of a_n is $M < n \leq M + N$ for some M. For $y \in \mathbb{R}$, we define $\|y\|$ to be the distance to the nearest integer. If there is some $\delta > 0$ so that

$$\|\alpha_r - \alpha_s\| > \delta, \text{ if } r \neq s$$

then the α_r are well-spaced. In this case, the number of distinct α_r is less than or equal to $1 + \delta^{-1}$. We say that $\{\alpha_r\}$ is a set of δ-spaced points. Selberg, Montgomery and Vaughn proved the following theorem independently.

Theorem 4.1.1. *For any set of δ-spaced points $\alpha_r \in \mathbb{R}/\mathbb{Z}$ and any complex numbers a_n with $M < n \leq M + N$, where $0 < \delta \leq \frac{1}{2}$ and $N \geq 1$ an integer, we have*

$$\sum_r \left| \sum_{M<n\leq M+N} a_n e(\alpha_r n) \right|^2 \leq \left(\delta^{-1} + N - 1 \right) \|a\|_2^2. \tag{4.3}$$

We now look at a special case of the large sieve inequality. We take α_r to be rationals a/q with $1 \leq q \leq Q$ and $(a, q) = 1$. Then α_r are spaced by $\delta = Q^{-2}$. Hence we have the following:

Theorem 4.1.2. *For any complex numbers a_n with $1 \leq n \leq N$, where N is a positive integer, we have*

$$\sum_{q\leq Q} \sideset{}{^*}\sum_{a(\bmod q)} \left| \sum_{1\leq n\leq N} a_n e\left(\frac{an}{q}\right) \right|^2 \leq (Q^2 + N - 1)\|a\|_2^2 \tag{4.4}$$

where

$$\|a\|_2^2 = \sum_{1\leq n\leq N} |a_n|^2.$$

The notation $*$ denotes the restriction to $(a, q) = 1$ throughout.

For general a_n, (4.4) gives the best possible bound. However, we can improve the bound in special cases. We are interested when the a_n are sparsely supported and Q is much smaller than N. To make this improvement, we apply the techniques of Selberg's sieve, which is new in this context.

4.1.1 Improvement of Large Sieve

By utilizing Selberg's sieve, we will prove the following:

Theorem 4.1.3. *We let $\{\lambda_d\}$ be any finite sequence of real numbers such that $|\lambda_d| \leq 1$ for all d, $\lambda_1 = 1$ and*

$$\sum_{d|n} \lambda_d \geq 0, \text{ for all } n \geq 1.$$

We define

$$\mathcal{D} = \{d : \lambda_d \neq 0\},\ D = \max \mathcal{D} + 1,$$

and

$$\mathcal{Q} = \{q \leq Q : (q, d) = 1 \text{ for all } d \in \mathcal{D}\}.$$

Finally, we assume $3 \leq Q \leq N/D^2$*. Then, for any complex numbers* $a_n \in \mathbb{C}$ *with* $1 \leq n \leq N$*, we have*

$$\sum_{q \in \mathcal{Q}} \sideset{}{^*}\sum_{a (\mathrm{mod}\ q)} \left| \sum_{n \leq N} a_n e\left(\frac{an}{q}\right) \left(\sum_{d|n} \lambda_d\right) \right|^2 \leq D(Q, N) \sum_{n \leq N} |a_n|^2 \left(\sum_{d|n} \lambda_d\right).$$

where

$$D(Q, N) = 2N \sum_{d \leq D} \frac{\lambda_d}{d} + 2Q \left(\frac{N}{3} \sum_{d \leq D} \frac{\lambda_d}{d}\right)^2.$$

We now look at a special case:

Corollary 4.1.4. *Let* $\{a_n\}$ *be a sequence of complex numbers supported on* n *whose smallest prime factor is larger than some* z*. We define*

$$\mathcal{Q}_z = \{q \leq Q : \text{ smallest prime factor of } q \text{ is } > z\}.$$

We let

$$D = 1 + \prod_{p \leq z} p.$$

Then if $3 \leq Q^2 \leq N/D^2$*, we have*

$$\sum_{q \in \mathcal{Q}_z} \sideset{}{^*}\sum_{a (\mathrm{mod}\ q)} \left| \sum_{n \leq N} a_n e\left(\frac{an}{q}\right) \right|^2 \leq D(Q, N) \sum_{n \leq N} |a_n|^2,\ \forall a_n \in \mathbb{C}.$$

with

$$D(Q, N) = 2aN + 2Q \left(\frac{aN}{3}\right)^{1/2} \text{ with } a = a(z),$$

$a(2) = 1/2,\ a(3) = 1/3,\ a(5) = 4/15,\ a(7) = 8/35,\ a(11) = 16/77.$

In general,

$$a(z) = \prod_{p \leq z} (1 - p^{-1}).$$

Proof. We choose $\lambda_d = \mu(d)$ whenever the largest prime factor of d is $\leq p_r$. Then

$$\sum_{d \leq D}{}' \frac{\lambda_d}{d} = a(p_r).$$

We note that if n does not have any prime factors $\leq p_r$, then

$$\sum_{d|n} \lambda_d = \lambda_1 = 1.$$

Applying this choice of λ_d to the theorem gives the corollary. □

4.2 Proof of Theorem

In order to prove Theorem (4.1.3), we use the following lemma.

Lemma 4.2.1. *Given a set $\mathcal{Q}$ bounded by Q, consider the following three statements:*

1. *For all $a_n \in \mathbb{C}$,*

$$\sum_{q \in \mathcal{Q}} \sum_{a (\bmod q)}^{*} \left| \sum_{n \leq N} a_n e\left(\frac{an}{q}\right) \left(\sum_{d|n} \lambda_d\right) \right|^2 \leq D(Q,N) \sum_{n \leq N} |a_n|^2 \left(\sum_{d|n} \lambda_d\right).$$

2. *For all $b_n \in \mathbb{C}$,*

$$\sum_{q \in \mathcal{Q}} \sum_{a (\bmod q)}^{*} \left| \sum_{n \leq N} b_n X\left(n, \frac{a}{q}\right) \right|^2 \leq D(Q,N) \sum_{n \leq N} |b_n|^2$$

where the operator $X(n, a/q)$ is defined by

$$X\left(n, \frac{a}{q}\right) = \left(\sum_{d|n} \lambda_d\right)^{1/2} e\left(\frac{an}{q}\right).$$

3. *For all $c(a,q) \in \mathbb{C}$,*

$$\sum_{n \leq N} \sum_{d|n} \lambda_d \left| \sum_{q \in \mathcal{Q}} \sum_{a (\bmod q)}^{*} c(a,q) e\left(\frac{an}{q}\right) \right|^2 \leq D(Q,N) \|c\|_2^2$$

where

$$\|c\|_2^2 = \sum_{q \in \mathcal{Q}}^{*} \sum_{a (\bmod q)} |c(a,q)|^2.$$

Then statement (1) follows from (2) and (2) is equivalent to (3).

Proof. Statement (1) follows from statement (2) by letting

$$b_n = \left(\sum_{d|n} \lambda_d \right)^{1/2} a_n.$$

By duality, (2) is equivalent to (3). □

Now we wish to prove (3) in order to prove Theorem (4.1.3).

To continue our analysis, we introduce a continuous function $g : \mathbb{R} \to \mathbb{R}$ such that

$$g(x) \geq 1 \text{ for } |x| \leq N \text{ and } g(x) \geq 0 \text{ for all } x.$$

We will later pick this function to optimize the results. Let $\widehat{g}(y)$ be the Fourier transform of g, so that

$$\widehat{g}(y) = \int_{\mathbb{R}} g(x) e^{-2\pi i x y} dx.$$

We have

$$\begin{aligned} &\sum_{n \leq N} \left| \sum_{q \in \mathcal{Q}} \sideset{}{^*}\sum_{a (\mathrm{mod}\ q)} c(a,q) \left(\sum_{d|n} \lambda_d \right)^{1/2} e\left(\frac{an}{q} \right) \right|^2 \\ &\leq \sum_{n \in \mathbb{Z}} g(n) \left(\sum_{d|n} \lambda_d \right) \left| \sum_{q \in \mathcal{Q}} \sideset{}{^*}\sum_{a (\mathrm{mod}\ q)} c(a,q) e\left(\frac{an}{q} \right) \right|^2. \end{aligned}$$

The diagonal terms of this sum will be the main contribution, so we treat the diagonal and off-diagonal terms separately. We let A be the contribution of the diagonal terms and B be the contribution of the off-diagonal terms.

$$\begin{aligned} &\sum_{n \leq N} \left| \sum_{q \in \mathcal{Q}} \sideset{}{^*}\sum_{a (\mathrm{mod}\ q)} c(a,q) \left(\sum_{d|n} \lambda_d \right)^{1/2} e\left(\frac{an}{q} \right) \right|^2 \\ &= \sum_{n \in \mathbb{Z}} g(n) \left(\sum_{d|n} \lambda_d \right) \sum_{q \in \mathcal{Q}} \sideset{}{^*}\sum_{a (\mathrm{mod}\ q)} |c(a,q)|^2 \\ &+ \sum_{n \in \mathbb{Z}} g(n) \left(\sum_{d|n} \lambda_d \right) \sideset{}{^*}\sum_{\frac{a_1}{q_1} \neq \frac{a_2}{q_2}} c(a_1, q_1) \overline{c(a_2, q_2)} e\left(\left(\frac{a_1}{q_1} - \frac{a_2}{q_2} \right) n \right) \\ &= A + B. \end{aligned}$$

4.2.1 Preliminary Analysis

We now analyze the diagonal and off-diagonal terms. We first treat the diagonal terms.

$$\begin{aligned} A &= \sum_{n\in\mathbb{Z}} g(n)\Big(\sum_{d|n}\lambda_d\Big)\sum_{q\in\mathcal{Q}}\sum_{a(\text{mod } q)}^{*} |c(a,q)|^2 = \|c\|_2^2 \sum_{n\in\mathbb{Z}} g(n)\Big(\sum_{d|n}\lambda_d\Big) \\ &= \|c\|_2^2 \sum_{d\le D}\lambda_d \sum_{m\in\mathbb{Z}} g(dm) = \|c\|_2^2 \sum_{d\le D}\frac{\lambda_d}{d}\sum_{h\in\mathbb{Z}}\widehat{g}\Big(\frac{h}{d}\Big)\end{aligned}$$

by Poisson summation. We then have

$$A = \|c\|_2^2\left\{\widehat{g}(0)\sum_{d\le D}\frac{\lambda_d}{d} + 2\sum_{d\le D}\frac{\lambda_d}{d}\sum_{h=1}^{\infty}\widehat{g}\Big(\frac{h}{d}\Big)\right\}.$$

We cannot analyze the diagonal contribution further until g is chosen. It will be chosen later to optimize the result. Next, we treat the off-diagonal terms.

$$\begin{aligned} B &= \sum_{n\in\mathbb{Z}} g(n)\Big(\sum_{d|n}\lambda_d\Big)\sum_{\frac{a_1}{q_1}\neq\frac{a_2}{q_2}}^{*} c(a_1,q_1)\overline{c(a_2,q_2)}\,e\Big(\Big(\frac{a_1}{q_1}-\frac{a_2}{q_2}\Big)n\Big) \\ &= \sum_{\frac{a_1}{q_1}\neq\frac{a_2}{q_2}}^{*} c(a_1,q_1)\overline{c(a_2,q_2)}\sum_{n\in\mathbb{Z}} g(n)\Big(\sum_{d|n}\lambda_d\Big)e\Big(\Big(\frac{a_1}{q_1}-\frac{a_2}{q_2}\Big)n\Big) \\ &= \sum_{\frac{a_1}{q_1}\neq\frac{a_2}{q_2}}^{*} c(a_1,q_1)\overline{c(a_2,q_2)}\sum_{d\le D}\lambda_d\sum_{m\in\mathbb{Z}} g(dm)\,e\Big(\Big(\frac{a_1}{q_1}-\frac{a_2}{q_2}\Big)dm\Big) \\ &= \sum_{\frac{a_1}{q_1}\neq\frac{a_2}{q_2}}^{*} c(a_1,q_1)\overline{c(a_2,q_2)}\sum_{d\le D}\frac{\lambda_d}{d}\sum_{h\in\mathbb{Z}}\widehat{g}\Big(\frac{h}{d}-\Big(\frac{a_1}{q_1}-\frac{a_2}{q_2}\Big)\Big)\end{aligned}$$

also by Poisson summation.

We note that $\widehat{g}(x)$ is of manageable size when $x\neq 0$. Unfortunately, $\widehat{g}(0)$ is a rather large term that should be avoided. We have a $\widehat{g}(0)$ contribution in the diagonal terms, which is expected as part of the main term. However, we do not wish to have $\widehat{g}(0)$ in the off-diagonal contribution. Therefore, we wish to avoid the situation where

$$\frac{h}{d}-\Big(\frac{a_1}{q_1}-\frac{a_2}{q_2}\Big)=0, \text{ for some } h\in\mathbb{Z}.$$

That is, when

$$\Big(\frac{a_1}{q_1}-\frac{a_2}{q_2}\Big)d = \Big(\frac{a_1q_2-a_2q_1}{q_1q_2}\Big)d\in\mathbb{Z}.$$

Thus, we have a problem when (q_1q_2,d) is large. We have that Q is much larger than $D>d$, but small values of q_1,q_2 can cause problems. For example, this technique is not useful for $\mathcal{Q}=\{q:1\le q\le Q\}$.

We let

$$\mathcal{D} = \{d : \lambda_d \neq 0\}, \text{ and } D = \max \mathcal{D} + 1.$$

Finally, we let

$$\mathcal{Q} = \{q : (q, d) = 1 \text{ for all } d \in \mathcal{D}\}.$$

We note in particular that

$$\mathcal{Q}_1 = \{q : q \text{ prime}, q \geq D\} \subset \mathcal{Q},$$

so the bound for $\mathcal{Q}$ will also work for $\mathcal{Q}_1$.

Lemma 4.2.2. *We have*

$$dx := d\left(\frac{a_1 q_2 - a_2 q_1}{q_1 q_2}\right) \notin \mathbb{Z} \textit{ for all } q_1, q_2 \in \mathcal{Q},\ \frac{a_1}{q_1} \neq \frac{a_2}{q_2}.$$

Proof. We assume for contradiction that $dx \in \mathbb{Z}$. Let $r = (q_1, q_2)$, $q_1' = q_1/r$, and $q_2' = q_2/r$. Then $(q_1', q_2') = 1$. Then

$$dx = \frac{d}{r}\left(\frac{a_1 q_2' - a_2 q_1'}{q_1' q_2'}\right) \in \mathbb{Z}.$$

Consider $p | q_1'$. Then $p | a_2 q_1'$ but $p \nmid a_1 q_2'$ because $(a_1, q_1') = 1 = (q_1', q_2')$. Thus, $p \nmid (a_1 q_2' - a_2 q_1')$ and hence $p | d$. This implies that $q_1' | d$. Likewise, $q_2' | d$. Since $(q_1', q_2') = 1$, this shows that $q_1' = q_2' = 1$ and so $r = q_1 = q_2$. Then

$$dx = d\left(\frac{a_1 - a_2}{q_1}\right) \in \mathbb{Z}.$$

This would imply that $q_1 | (a_1 - a_2)$, but $0 < |a_1 - a_2| < q_1$, so this is impossible. Therefore, $dx \notin \mathbb{Z}$. $\square$

Thus, if we define $\|dx\|$ to be the distance of dx to the closest integer, we have $\|dx\| \neq 0$ for all d, x with $\lambda_d \neq 0$. Furthermore, due to the spacing of rationals, $\|dx\| \geq 1/Q^2$ for all d, x.

Now we return to the contribution of the off-diagonal terms. We have

$$\left|\sum_{d \leq D} \frac{\lambda_d}{d} \sum_{h \in \mathbb{Z}} \widehat{g}\left(\frac{h}{d} - x\right)\right| = \left|\sum_{d \leq D} \frac{\lambda_d}{d}\left(\widehat{g}\left(\frac{\|dx\|}{d}\right) + \sum_{k \neq 0}^{\infty} \widehat{g}\left(\frac{k + \|dx\|}{d}\right)\right)\right|.$$

This is the most analysis of B we can do until we choose a function g.

4.2.2 Choice of g

We have several constraints while choosing our function g. On the one hand, we would like $\widehat{g}(0)$ to be as small as possible so that A is small. However, we also need the sum

$$\sum_{k\neq 0}^{\infty} \widehat{g}\left(\frac{k+\|dx\|}{d}\right)$$

to converge, so we need $\widehat{g}(y) \leq O(1/y^2)$. Finally, we need $g(x) \geq 1$ for $|x| \leq N$ and $g(x) \geq 0$ for all x. A family of such functions is given by:

$$g_k(x) = \begin{cases} 1 & |x| \leq N \\ -k\left|\frac{x}{N}\right| + (k+1) & N \leq |x| \leq N\left(1+\frac{1}{k}\right) \end{cases}$$

where k is a parameter that will be chosen later. Then we have

$$\widehat{g}_k(y) = \begin{cases} N\left(2+\frac{1}{k}\right) & y = 0, \\ \dfrac{k}{2N(\pi y)^2}\left[\cos\left(2\pi N y) - \cos(2\pi N y\left(1+\frac{1}{k}\right)\right)\right] & \text{otherwise.} \end{cases}$$

We then note that $\widehat{g}_k(y)$ is majorized by $f_k(y)$ where f_k is given by

$$f_k(y) = \begin{cases} N\left(2+\frac{1}{k}\right) & y = 0, \\ \dfrac{k}{N(\pi y)^2} & \text{otherwise.} \end{cases}$$

4.2.3 Conclusion

With a choice of our function g, we can now continue our analysis of the diagonal and off-diagonal terms. For the diagonal terms, we have

$$\begin{aligned} A &= \|c\|_2^2 \left\{ \widehat{g}_k(0) \sum_{d\leq D} \frac{\lambda_d}{d} + 2 \sum_{d\leq D} \frac{\lambda_d}{d} \sum_{h=1}^{\infty} \widehat{g}_k\left(\frac{h}{d}\right) \right\} \\ &\leq \|c\|_2^2 \left\{ N\left(2+\frac{1}{k}\right) \sum_{d\leq D} \frac{\lambda_d}{d} + \sum_{d\leq D} \frac{|\lambda_d|}{d} \frac{2k}{N} \sum_{h=1}^{\infty} \frac{d^2}{\pi^2 h^2} \right\} \\ &= \|c\|_2^2 \left\{ N\left(2+\frac{1}{k}\right) \sum_{d\leq D} \frac{\lambda_d}{d} + \frac{k}{3N} \sum_{d\leq D} |\lambda_d| d \right\}. \end{aligned}$$

We now examine the contribution of the off-diagonal terms.

$$\begin{aligned}
&\left|\sum_{d\le D}\frac{\lambda_d}{d}\sum_{h\in\mathbb{Z}}\widehat{g}_k\left(\frac{h}{d}-x\right)\right| \\
&=\left|\sum_{d\le D}\frac{\lambda_d}{d}\left(\widehat{g}_k\left(\frac{\|dx\|}{d}\right)+\sum_{k\ne 0}^{\infty}\widehat{g}_k\left(\frac{k+\|dx\|}{d}\right)\right)\right| \\
&\le\sum_{d\le D}\frac{|\lambda_d|}{d}\left(f_k\left(\frac{\|dx\|}{d}\right)+\sum_{l\ne 0}^{\infty}f_k\left(\frac{l+\|dx\|}{d}\right)\right) \\
&=\sum_{d\le D}\frac{|\lambda_d|}{d}\left(f_k\left(\frac{\|dx\|}{d}\right)+\frac{k}{N\pi^2}\sum_{l\ne 0}^{\infty}\frac{d^2}{(l+\|dx\|)^2}\right).
\end{aligned}$$

To continue our analysis, we note that $0\le\|dx\|\le 1/2$. Splitting the summation over l, we find that

$$(l+\|dx\|)^2\ge\begin{cases} l^2 & \text{if } l\ge 1,\\ 1/4 & \text{if } l=-1,\\ (l+1)^2 & \text{if } l\le -2.\end{cases}$$

Therefore,

$$\sum_{l\ne 0}^{\infty}\frac{d^2}{(l+\|dx\|)^2}\le\sum_{l=1}^{\infty}\frac{d^2}{l^2}+4d^2+\sum_{-\infty}^{-2}\frac{d^2}{(l+1)^2}\le d^2\pi^2\frac{7}{9}.$$

Applying this estimate to our sum, we have

$$\left|\sum_{d\le D}\frac{\lambda_d}{d}\sum_{h\in\mathbb{Z}}\widehat{g}_k\left(\frac{h}{d}-x\right)\right|\le\sum_{d\le D}\frac{|\lambda_d|}{d}\left(f_k\left(\frac{\|dx\|}{d}\right)+\frac{7d^2k}{9N}\right).$$

Finally,

$$|B|\le\sum_{q_1\in\mathcal{Q}}\sum_{a_1\,(\mathrm{mod}\ q_1)}^{*}|c(a_1,q_1)|^2\sum_{x\ne 0}\sum_{d\le D}\frac{|\lambda_d|}{d}\left(f_k\left(\frac{\|dx\|}{d}\right)+\frac{7d^2k}{9N}\right).$$

We know that $\|dx\|\ne 0$ for all d such that $\lambda_d\ne 0$. In addition, we know that the a/q are δ-spaced where $\delta=Q^{-2}$. Therefore, so are the $\|dx\|$. Applying these facts, we see

$$\begin{aligned}
|B| &\le\sum_{q_1\in\mathcal{Q}}\sum_{a_1\,(\mathrm{mod}\ q_1)}^{*}|c(a_1,q_1)|^2\sum_{d\le D}\frac{|\lambda_d|}{d}\sum_{x\ne 0}\left(f_k\left(\frac{\|dx\|}{d}\right)+\frac{7kd^2}{9N}\right) \\
&\le\sum_{q_1\in\mathcal{Q}}\sum_{a_1\,(\mathrm{mod}\ q_1)}^{*}|c(a_1,q_1)|^2\left\{\sum_{d\le D}\frac{|\lambda_d|}{d}\left[\frac{7kQ^2d^2}{9N}+\sum_{l=1}^{\infty}f_k\left(\frac{l}{dQ^2}\right)\right]\right\} \\
&=\|c\|_2^2\sum_{d\le D}|\lambda_d|d\left(\frac{7kQ^2}{9N}+\frac{kQ^4}{6N}\right).
\end{aligned}$$

Simplifying the expression slightly, we find

$$A + B \leq \|c\|_2^2 \left(N \left(2 + \frac{1}{k} \right) \sum_{d \leq D} \frac{\lambda_d}{d} + \frac{kQ^4}{3N} \sum_{d \leq D} |\lambda_d| d \right).$$

In addition, since $|\lambda_d| \leq 1$ for all d,

$$\sum_{d \leq D} |\lambda_d| d \leq D^2.$$

Thus, if we assume that $Q^2 \leq N/D^2$, then

$$A + B \leq \|c\|_2^2 \left(N \left(2 + \frac{1}{k} \right) \sum_{d \leq D} \frac{\lambda_d}{d} + \frac{kQ^2}{3} \right).$$

Finally, we choose k as the geometric mean of $N \sum \lambda_d/d$ and $Q^2/3$. Then

$$A + B \leq \|c\|_2^2 \left(2N + 2Q \left(\frac{N}{3} \sum_{d \leq D} \frac{\lambda_d}{d} \right)^{1/2} \right).$$

Therefore, by Lemma (4.2.1), we have Theorem (4.1.3).

Chapter 5

Notes on Successive Approximations

This chapter consists of notes on the use of successive approximations to refine Selberg's upper bound sieve. We also provide some useful lemmas for solving systems of equations which are prevalent in sieve theory.

5.1 Introduction

Let $\mathcal{A} = (a_n)$, $a_n \geq 0$, $\Lambda = (\lambda_d)$, $d \leq D$. Assume

$$\sum_{d|n} \lambda_d \geq 0, \text{ for all } n \text{ and } \lambda_1 = 1.$$

Let $\rho_d \in \mathbb{R}$ such that

$$\left(\sum_{d|n} \rho_d\right)^2 = \sum_{d|n} \lambda_d$$

with $\rho_1 = 1$. Then we have

$$\lambda_d = \sum_{[d_1,d_2]=d} \rho_{d_1}\rho_{d_2}.$$

We now wish to look at the sum

$$S(\mathcal{A}) = \sum_{(n,P)=1} a_n.$$

We estimate this sum by

$$S(\mathcal{A}) \leq S^+(\mathcal{A}) = \sum_n a_n \left(\sum_{d|n} \rho_d\right)^2.$$

We would like to find variables ρ_d that give the minimum of $S^+(\mathcal{A})$. In general, this problem is impossible. However, we can find a choice of ρ_d which give a good approximation to the minimum of $S^+(\mathcal{A})$. Following the notation of Chapter (1), we write

$$\begin{aligned} S^+(\mathcal{A}) &= X \sum_{d|P(z)} g(d)\lambda_d + \sum_{d|P(z)} \lambda_d r(\mathcal{A}, d) \\ &= XV(D, z) + R(\mathcal{A}, D). \end{aligned}$$

In order to approximate $S(\mathcal{A})$, Selberg chose the variables ρ_d to minimize the main term $V(D, z)$. His choice of ρ_d also resulted in a remainder term that was under control. He chose

$$\rho_d = \frac{\mu(d)h(d)}{g(d)J} \sum_{l \leq \sqrt{D}/d(d,l)=1} h(l)$$

where $h(d)$ is the multiplicative function defined by $h(p) = g(p)/(1 - g(p))$ and

$$J = \sum_{\substack{d \leq \sqrt{D} \\ d|P}} h(d).$$

We then have

$$S^+(\mathcal{A}) \leq \frac{X}{J} + R(\mathcal{A}, D) \tag{5.1}$$

where

$$R(\mathcal{A}, D) = \sum_{d|P} \lambda_d r(\mathcal{A}, d).$$

With this choice of ρ_d, we have $|\rho_d| \leq 1$ so the remainder term is under control. However, we would like to find a choice of ρ_d that minimizes the remainder term at the same time. To do this, we will use the method of successive approximations. We will start by choosing our ρ_d to be the same as Selberg's choice. Then we will perturb this choice slightly in order to also minimize the remainder term. This process could be continued indefinitely. However, since Selberg's choice of ρ_d already gives such a good result, we will only execute this process once. In the following sections, we will describe this process.

5.1.1 Property of S^+

We note that S^+ is a quadratic form in variables ρ_d for $d \neq 1$. We know that S^+ must obtain a minimum because trivially $S^+ \geq 0$, so we want to find the (ρ_d) at which the minimum is attained. That is, where

$$\frac{\partial}{\partial \rho_k} S^+ = 0 \text{ for } k \neq 1.$$

For $k \neq 1$,

$$\frac{\partial}{\partial \rho_k} S^+ = 2 \sum_{n \equiv 0 (\bmod\ k)} a_n \left(\sum_{d|n} \rho_d \right) = 0. \tag{5.2}$$

We now note a special property of S^+_{min}, that is the minimum of S^+ given by the optimal choice of ρ_d.

Proposition 5.1.1. *Let ρ_d be chosen to minimize*

$$S^+(\mathcal{A}) = \sum_{n|P(z)} a_n \left(\sum_{d|n} \rho_d \right)^2.$$

Then

$$S^+_{min} = \sum_{n|P(z)} a_n \sum_{d|n} \rho_d.$$

Proof. If ρ_d give the minimum of $S^+(\mathcal{A})$, then we have

$$\sum_{n \equiv 0 (\bmod\ k)} a_n \left(\sum_{d|n} \rho_d \right) = 0$$

for all $k|P(z)$, $k \neq 1$. Then

$$\begin{aligned} S^+(\mathcal{A}) &= \sum_{n|P(z)} a_n \left(\sum_{d|n} \rho_d \right)^2 \\ &= \sum_{n|P(z)} a_n \left(\sum_{d_1|n} \rho_{d_1} \right) \left(\sum_{d_2|n} \rho_{d_2} \right) \\ &= \sum_{d_1|P(z)} \rho_{d_1} \sum_{\substack{n|P(z) \\ n \equiv 0 (\bmod\ d_1)}} a_n \sum_{d_2|n} \rho_{d_2} \\ &= \sum_{d_1|P(z)} \rho_{d_1} \begin{cases} 0 & \text{if } d \neq 1 \\ \sum_{n|P(z)} a_n \sum_{d|n} \rho_{d_2} & \text{if } d_1 = 1 \end{cases} \\ &= \sum_{n|P(z)} a_n \sum_{d|n} \rho_d \end{aligned}$$

since $\rho_1 = 1$. This concludes the proof. □

Now we return to finding the minimum ρ_d. We define

$$A_l = \sum_{n \equiv 0 (\bmod\ l)} a_n.$$

Then the condition for the derivatives (5.2) is equivalent to:

$$\sum_{d\neq 1} \rho_d A_{[d,k]} = -A_k \text{ for all } k \neq 1.$$

Let S^+_{min} be the minimum of $S^+(\mathcal{A})$. Then we would like to find ρ_d such that

$$\sum_{\substack{d|P \\ d\leq\sqrt{D}}} \rho_d A_{[d,k]} = \delta_{k1} S^+_{min}.$$

In order to solve this system of equations, we will need some lemmas, which we present in the following section.

5.2 Preliminaries

The following lemmas are true for all systems of equations where $g(d)$ is a multiplicative function with $g(d) = 0$ for $d \geq \sqrt{D}$ and $h(d)$ is the multiplicative function defined by $h(p) = g(p)/(1 - g(p))$.

Lemma 5.2.1. *Consider the system of equations:*

$$\sum_{\substack{d|P \\ d\leq\sqrt{D}}} \rho_d g([d,k]) = \begin{cases} J^{-1} & \text{if } k = 1 \\ 0 & \text{if } k \neq 1 \end{cases}$$

where $k|P$, $k \leq \sqrt{D}$ *and*

$$J = \sum_{\substack{d|P \\ d\leq\sqrt{D}}} h(d).$$

The solution to this system of equations is given by

$$\rho_d = \frac{\mu(d)}{g(d)J} \sum_{\substack{m\equiv 0 (\mathrm{mod}\ d) \\ m|P \\ m\leq\sqrt{D}}} h(m).$$

Proof. Let ρ_d be given by the formula in the lemma. Then

$$\begin{aligned}
\textstyle\sum_{\substack{d|P\\ d\le\sqrt{D}}} \rho_d g([d,k]) &= \frac{1}{J}\sum_{\substack{d|P\\ d\le\sqrt{D}}} \mu(d)\frac{g([d,k])}{g(d)} \sum_{\substack{m\equiv 0 \pmod d\\ m|P\\ m\le\sqrt{D}}} h(m)\\
&= \frac{1}{H}\sum_{\substack{m|P\\ m\le\sqrt{D}}} h(m)\sum_{d|m}\mu(d)\frac{g([d,k])}{g(d)}\\
&= \frac{1}{H}\sum_{\substack{m|P\\ m\le\sqrt{D}}} h(m)\sum_{d|m}\mu(d)g\left(\frac{k}{(d,k)}\right).
\end{aligned}$$

Write $d = d_1 d_2$ with $d_1|k$ and $(d_2,k)=1$. Then

$$\begin{aligned}
\textstyle\sum_{d|m}\mu(d)g\left(\frac{k}{(d,k)}\right) &= \sum_{d_1|(m,k)}\sum_{\substack{d_2|m\\ (d_2,k)=1}} \mu(d_1)\mu(d_2)g\left(\frac{k}{d_1}\right)\\
&= \sum_{d_1|(m,k)}\mu(d_1)g\left(\frac{k}{d_1}\right)\sum_{\substack{d_2|m\\ (d_2,k)=1}}\mu(d_2)\\
&= \sum_{d_1|(m,k)}\mu(d_1)g\left(\frac{k}{d_1}\right)\sum_{d_2|m/(m,k)}\mu(d_2)\\
&= \sum_{d_1|(m,k)}\mu(d_1)g\left(\frac{k}{d_1}\right)\begin{cases}1 & \text{if } m|k\\ 0 & \text{otherwise}\end{cases}.
\end{aligned}$$

Therefore, for $m\nmid k$, $\sum_{d|m}\mu(d)g(k/(d,k))=0$ and for $m|k$,

$$\begin{aligned}
\textstyle\sum_{d|m}\mu(d)g\left(\frac{k}{(d,k)}\right) &= \sum_{d_1|m}\mu(d_1)g\left(\frac{k}{d_1}\right)\\
&= g(k)\sum_{d_1|m}\frac{\mu(d_1)}{g(d_1)}\\
&= g(k)\prod_{p|m}\left(1-\frac{1}{g(p)}\right)\\
&= g(k)\frac{\mu(m)}{h(m)}.
\end{aligned}$$

Returning to the original sum, we have

$$\begin{aligned}
\sum_{\substack{d|P \\ d\le\sqrt{D}}} \rho_d g([d,k]) &= \frac{1}{J} \sum_{\substack{m|P \\ m\le\sqrt{D}}} h(m) \sum_{d|m} \mu(d) g\left(\frac{k}{(d,k)}\right) \\
&= \frac{1}{J} g(k) \sum_{\substack{m|k \\ m\le\sqrt{D}}} h(m) \frac{\mu(m)}{h(m)} \\
&= \frac{1}{J} g(k) \sum_{\substack{m|k \\ m\le\sqrt{D}}} \mu(m) \\
&= \begin{cases} \frac{1}{J} & \text{if } k = 1, \\ 0 & \text{if } k \neq 1. \end{cases}
\end{aligned}$$

□

Lemma 5.2.2. *Let $l|P$ and $l \le \sqrt{D}$. Consider the system of equations in variables $t_d^{(l)}$:*

$$\sum_{\substack{d\le\sqrt{D} \\ d|P}} t_d^{(l)} g([d,k]) = \begin{cases} 1 & \textit{if } k = l, \\ 0 & \textit{if } k \neq l. \end{cases}$$

The solution to this system of equations is given by:

$$t_d^{(l)} = \frac{\mu(d)\mu(l)}{g(d)g(l)} \sum_{\substack{m\equiv 0 (\mathrm{mod}\ [d,l]) \\ m|P \\ m\le\sqrt{D}}} h(m).$$

Proof. Let $t_d^{(l)}$ be given as in the Lemma.

$$\begin{aligned}
\sum_{\substack{d\le\sqrt{D} \\ d|P}} t_d^{(l)} g([d,k]) &= \frac{\mu(l)}{g(l)} \sum_{\substack{d\le\sqrt{D} \\ d|P}} \frac{\mu(d)}{g(d)} g([d,k]) \sum_{\substack{m\equiv 0 (\mathrm{mod}\ [d,l]) \\ m|P \\ m\le\sqrt{D}}} h(m) \\
&= \frac{\mu(l)}{g(l)} \sum_{\substack{m\equiv 0 (\mathrm{mod}\ l) \\ m|P \\ m\le\sqrt{D}}} h(m) \sum_{d|m} \mu(d) g\left(\frac{k}{(d,k)}\right).
\end{aligned}$$

We recall from the proof of Lemma (5.2.1) that

$$\sum_{d|m} \mu(d) g\left(\frac{k}{(d,k)}\right) = g(k) \frac{\mu(m)}{h(m)}$$

for $m|k$ and is zero otherwise. Then

$$\begin{aligned}
\textstyle\sum_{\substack{d\leq\sqrt{D}\\ d|P}} t_d^{(l)} g([d,k]) &= \frac{\mu(l)}{g(l)} \sum_{\substack{m\equiv 0(\text{mod } l)\\ m|P\\ m\leq\sqrt{D}}} h(m) \sum_{d|m} \mu(d) g\left(\frac{k}{(d,k)}\right) \\
&= \frac{\mu(l)}{g(l)} \sum_{\substack{m\equiv 0(\text{mod } l)\\ m|k}} h(m) g(k) \frac{\mu(m)}{h(m)} \\
&= g(k)\frac{\mu(l)}{g(l)} \sum_{\substack{m\equiv 0(\text{mod } l)\\ m|k}} \mu(m) \\
&= g(k)\frac{\mu(l)}{g(l)} \mu(l) \sum_{n|k/l} \mu(m) \\
&= \begin{cases} 1 & \text{if } k = l, \\ 0 & \text{if } k \neq l. \end{cases}
\end{aligned}$$

□

We can now use Lemma (5.2.2) to find the solution to a more general system of equations:

Lemma 5.2.3. *Consider the system of equations in variables ρ_d given by*

$$\sum_{\substack{d\leq\sqrt{D}\\ d|P}} \rho_d g([d,k]) = E_k$$

where E_k are general expressions depending on k where $k|P$ and $k \leq \sqrt{D}$. Then the solution to this system of equations is given by:

$$\rho_d = \sum_{\substack{l\leq\sqrt{D}\\ l|P}} t_d^{(l)} E_l$$

where $t_d^{(l)}$ are given by the formula in Lemma (5.2.2)

Proof. Let ρ_d be given by the formula in the statement of the Lemma. Then

$$\begin{aligned}\sum_{\substack{d|P\\ d\leq\sqrt{D}}} \rho_d g([d,k]) &= \sum_{\substack{d|P\\ d\leq\sqrt{D}}} g([d,k]) \sum_{\substack{l\leq\sqrt{D}\\ l|P}} t_d^{(l)} E_l \\ &= \sum_{\substack{l\leq\sqrt{D}\\ l|P}} E_l \sum_{\substack{d|P\\ d\leq\sqrt{D}}} t_d^{(l)} g([d,k]) \\ &= \sum_{\substack{l\leq\sqrt{D}\\ l|P}} E_l \delta_{lk}\end{aligned}$$

by Lemma (5.2.2). Therefore,

$$\sum_{\substack{d|P\\ d\leq\sqrt{D}}} \rho_d g([d,k]) = E_k.$$

□

5.3 Solution to System of Equations

We recall that we wish to solve the system of equations

$$\sum_{\substack{d|P\\ d\leq\sqrt{D}}} \rho_d A_{[d,k]} = \delta_{k1} S^+_{min}$$

where S^+_{min} is the minimum of $S^+(\mathcal{A})$. We let

$$A_m = g(m)X + r_m$$

where g is a multiplicative function. Then, we want to solve the system of equations

$$\delta_{k1} S^+_{min} = X \sum_{\substack{d|P\\ d\leq\sqrt{D}}} g(d)\rho_{[d,k]} + \sum_{\substack{d|P\\ d\leq\sqrt{D}}} \rho_d r_{[d,k]}.$$

We use the method of successive approximations in order to determine the best choice of ρ_d. We start first with Selberg's choice of ρ_d. Since Selberg's standard choice of ρ_d was made to minimize the main term, his choice corresponds to $r_m = 0$. We let $r_m = 0$ and $S^+ = X/J$ (the main term corresponding to Selberg's choice). We note that $S^+ = X/J$ is the definition of S^+ and is not equal to $S^+(\mathcal{A})$. We let $\rho_d^{(0)}$ be the variables such that

$$\sum_{\substack{d|P \\ d\leq\sqrt{D}}} \rho_d^{(0)} g([d,k]) = \delta_{k1}\frac{S^+}{X} = \delta_{k1}\frac{1}{J}.$$

Then from Lemma (5.2.1) of section (5.2) we find that

$$\rho_d^{(0)} = \frac{\mu(d)}{g(d)J} \sum_{\substack{m\equiv 0 (\text{mod } d) \\ m|P \\ m\leq\sqrt{D}}} h(m)$$

which is Selberg's choice of ρ_d. We would like to perturb this choice slightly in order to find a better choice of ρ_d when the error term is close to the main term. Therefore, we write

$$\rho_d = \rho_d^{(0)} + \rho_d'$$

where ρ_d is the optimal choice, which gives the minimum of $S^+(\mathcal{A})$, $\rho_d^{(0)}$ is Selberg's choice and ρ_d' is the difference, which we think of as small. We note that

$$S^+_{min} = \frac{X}{J} + \sum_{d|P(z)} g(d)\rho_d' + \sum_{d|P(z)} \rho_d^{(0)} r_d + \sum_{d|P(z)} \rho_d' r_d.$$

We also note that

$$\begin{aligned}
\sum_{\substack{d|P \\ d\leq\sqrt{D}}} \rho_d' A_{[d,k]} &= \delta_{k1}S^+_{min} - \sum_{\substack{d|P \\ d\leq\sqrt{D}}} \rho_d^{(0)} A_{[d,k]} \\
&= \delta_{k1}S^+_{min} - \delta_{k1}\frac{X}{J} - \sum_{\substack{d|P \\ d\leq\sqrt{D}}} \rho_d^{(0)} r_{[d,k]} \\
&= \delta_{k1}\left(S^+_{min} - \frac{X}{J}\right) - \sum_{\substack{d|P \\ d\leq\sqrt{D}}} \rho_d^{(0)} r_{[d,k]} \\
&\stackrel{def}{=} C_k.
\end{aligned}$$

We then have

$$\sum_{\substack{d|P \\ d\leq\sqrt{D}}} \rho_d' \left(g([d,k])X + r_{[d,k]}\right) = C_k.$$

We want to solve the system of equations:

$$\sum_{\substack{d|P \\ d\leq\sqrt{D}}} \rho_d' g([d,k]) = \delta_{k1}\left(\frac{S^+_{min}}{X} - \frac{1}{J}\right) - \frac{1}{X}\sum_{\substack{d|P \\ d\leq\sqrt{D}}} (\rho_d^{(0)} + \rho_d') r_{[d,k]}. \tag{5.3}$$

We let E_k be the right-hand side of (5.3) and we see from Lemma (5.2.3) of Section (5.2) that

$$\begin{aligned}
\rho'_d &= \sum_{\substack{l\leq\sqrt{D}\\ l|P}} t_d^{(l)} E_l \\
&= \frac{\mu(d)}{g(d)} \sum_{\substack{l|P\\ l\leq\sqrt{D}}} \frac{\mu(l)}{g(l)} E_l \sum_{\substack{m\equiv 0 (\mathrm{mod}\ [d,l])\\ m|P\\ m\leq\sqrt{D}}} h(m) \\
&= \frac{\mu(d)}{g(d)} \left(\frac{S^+_{min}}{X} - \frac{1}{J} \right) \sum_{\substack{m\equiv 0 (\mathrm{mod}\ d)\\ m|P\\ m\leq\sqrt{D}}} h(m) \\
&\quad - \frac{1}{X} \frac{\mu(d)}{g(d)} \sum_{\substack{m\equiv 0 (\mathrm{mod}\ d)\\ m|P\\ m\leq\sqrt{D}}} h(m) \sum_{l|m} \frac{\mu(l)}{g(l)} \sum_{\substack{t|P\\ t\leq\sqrt{D}}} \rho'_t r_{[t,l]} \\
&\quad - \frac{1}{XJ} \frac{\mu(d)}{g(d)} \sum_{\substack{m\equiv 0 (\mathrm{mod}\ d)\\ m|P\\ m\leq\sqrt{D}}} h(m) \sum_{l|m} \frac{\mu(l)}{g(l)} \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n) \sum_{t|n} \mu(t) g(t) r_{[t,l]}.
\end{aligned}$$

We note that the second term will be especially small since we think of both ρ'_d and r_d as small. Therefore, we ignore that term and define

$$\begin{aligned}
\rho_d^{(1)} &= \frac{\mu(d)}{g(d)} \left(\frac{S^+_{min}}{X} - \frac{1}{J} \right) \sum_{\substack{m\equiv 0 (\mathrm{mod}\ d)\\ m|P\\ m\leq\sqrt{D}}} h(m) \\
&\quad - \frac{1}{XJ} \frac{\mu(d)}{g(d)} \sum_{\substack{m\equiv 0 (\mathrm{mod}\ d)\\ m|P\\ m\leq\sqrt{D}}} h(m) \sum_{l|m} \frac{\mu(l)}{g(l)} \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n) \sum_{t|n} \mu(t) g(t) r_{[t,l]}.
\end{aligned}$$

This is our next approximation to the best choice of ρ_d. We now define

$$\rho_d = \rho_d^{(0)} + \rho_d^{(1)}. \tag{5.4}$$

We see then that

$$\begin{aligned}
\rho_d &= \frac{\mu(d)}{g(d)} \frac{S^+_{min}}{X} \sum_{\substack{m\equiv 0 (\mathrm{mod}\ d)\\ m|P\\ m\leq\sqrt{D}}} h(m) \\
&\quad - \frac{1}{XJ} \frac{\mu(d)}{g(d)} \sum_{\substack{m\equiv 0 (\mathrm{mod}\ d)\\ m|P\\ m\leq\sqrt{D}}} h(m) \sum_{l|m} \frac{\mu(l)}{g(l)} \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n) \sum_{t|n} \mu(t) g(t) r_{[t,l]}.
\end{aligned}$$

For this choice of ρ_d, we have the following:

Proposition 5.3.1. *Define*

$$\begin{aligned}\rho_d = \rho_d^{(0)} + \rho_d^{(1)} &= \frac{\mu(d)}{g(d)}\frac{S^+_{min}}{X} \sum_{\substack{m\equiv 0 (\bmod d)\\ m|P\\ m\le\sqrt{D}}} h(m) \\ &-\frac{1}{XJ}\frac{\mu(d)}{g(d)} \sum_{\substack{m\equiv 0 (\bmod d)\\ m|P\\ m\le\sqrt{D}}} h(m)\sum_{l|m}\frac{\mu(l)}{g(l)}\sum_{\substack{n|P\\ n\le\sqrt{D}}} h(n)\sum_{t|n}\mu(t)g(t)r_{[t,l]}.\end{aligned}$$

Then for all $k|P$, $k\le\sqrt{D}$, we have

$$\begin{aligned}&\sum_{\substack{d|P\\ d\le\sqrt{D}}} \rho_d A_{[d,k]} \\ &= S^+_{min}\left(\delta_{k1} + \frac{1}{X}\sum_{\substack{n|P\\ n\le\sqrt{D}}} h(n)\sum_{t|n}\frac{\mu(t)}{g(t)}r_{[t,k]}\right) - \frac{1}{J}\sum_{\substack{n|P\\ n\le\sqrt{D}}} h(n)\sum_{t|n}\frac{\mu(t)}{g(t)}r_{[t,k]} \\ &-\frac{1}{XJ}\sum_{\substack{d|P\\ d\le\sqrt{D}}}\frac{\mu(d)}{g(d)}r_{[d,k]}\sum_{\substack{m\equiv 0 (\bmod d)\\ m|P\\ m\le\sqrt{D}}} h(m)\sum_{l|m}\frac{\mu(l)}{g(l)}\sum_{\substack{n|P\\ n\le\sqrt{D}}} h(n)\sum_{t|n}\mu(t)g(t)r_{[t,l]}.\end{aligned}$$

To prove this statement, we will need some lemmas.

Lemma 5.3.2. *For all $n|k$,*

$$\sum_{\substack{m|P\\ m\le\sqrt{D}\\ (m,k)=n}} h(m)\sum_{d|m}\mu(d)\frac{g([d,k])}{g(d)} = \mu(n)g(k).$$

Proof.

$$\sum_{\substack{m|P\\ m\le\sqrt{D}\\ (m,k)=n}} h(m)\sum_{d|m}\mu(d)\frac{g([d,k])}{g(d)} = g(k)\sum_{\substack{m|P\\ m\le\sqrt{D}\\ (m,k)=n}} h(m)\sum_{d|m}\mu(d)\frac{1}{g((d,k))}.$$

If $d|m$ and $(m,k) = n$, then $(d,k) = (d,n)$. For each d, we write $t = (d,n)$ and

$l = d/(d,n)$ so $d = tl$. Then for m such that $(m,k) = n$, we have

$$\begin{aligned}
\textstyle\sum_{d|m} \mu(d)\frac{1}{g((d,k))} &= \sum_{t|n} \frac{1}{g(t)} \sum_{\substack{l|m \\ (l,n)=1}} \mu(tl) \\
&= \sum_{t|n} \frac{\mu(t)}{g(t)} \sum_{l|m/n} \mu(l) \\
&= \begin{cases} \mu(n)/h(n) & \text{if } m = n, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

Applying this to our sum, we find

$$\sum_{\substack{m|P \\ m\le\sqrt{D} \\ (m,k)=n}} h(m) \sum_{d|m} \mu(d)\frac{g([d,k])}{g(d)} = \mu(n)g(k).$$

□

We use the above lemma to prove the following two lemmas.

Lemma 5.3.3. *For all $k|P$, $k \le \sqrt{D}$, we have*

$$\delta_{k1} = \sum_{\substack{d|P \\ d\le\sqrt{D}}} \frac{\mu(d)g([d,k])}{g(d)} \sum_{\substack{m\equiv 0 (\text{mod } d) \\ m|P \\ m\le\sqrt{D}}} h(m).$$

Proof. We have

$$\begin{aligned}
&\sum_{\substack{d|P \\ d\le\sqrt{D}}} \frac{\mu(d)g([d,k])}{g(d)} \sum_{\substack{m\equiv 0 (\text{mod } d) \\ m|P \\ m\le\sqrt{D}}} h(m) = \sum_{\substack{m|P \\ m\le\sqrt{D}}} h(m) \sum_{d|m} \mu(d)\frac{g([d,k])}{g(d)} \\
&= \sum_{n|k} \sum_{\substack{m|P \\ m\le\sqrt{D} \\ (m,k)=n}} h(m) \sum_{d|m} \mu(d)\frac{g([d,k])}{g(d)} = g(k)\sum_{n|k}\mu(n) \\
&= \delta_{k1}.
\end{aligned}$$

□

Lemma 5.3.4. *For all $k|P$, $k \leq \sqrt{D}$, we have*

$$\sum_{\substack{d|P\\ d\leq\sqrt{D}}} \mu(d)\frac{g([d,k])}{g(d)} \sum_{\substack{m\equiv 0 (\mathrm{mod}\ d)\\ m|P\\ m\leq\sqrt{D}}} h(m)\sum_{l|m}\frac{\mu(l)}{g(l)} \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n)\sum_{t|n}\frac{\mu(t)}{g(t)} r_{[t,l]}$$
$$= \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n)\sum_{t|n}\frac{\mu(t)}{g(t)} r_{[t,k]}.$$

Proof.

$$\sum_{\substack{d|P\\ d\leq\sqrt{D}}} \mu(d)\frac{g([d,k])}{g(d)} \sum_{\substack{m\equiv 0 (\mathrm{mod}\ d)\\ m|P\\ m\leq\sqrt{D}}} h(m)\sum_{l|m}\frac{\mu(l)}{g(l)} \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n)\sum_{t|n}\frac{\mu(t)}{g(t)} r_{[t,l]}$$
$$= g(k)\sum_{\substack{m|P\\ m\leq\sqrt{D}}} h(m)\sum_{l|m}\frac{\mu(l)}{g(l)} \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n)\sum_{t|n}\frac{\mu(t)}{g(t)} r_{[t,l]} \sum_{d|m}\frac{\mu(d)}{g((d,k))}$$
$$= g(k)\sum_{\alpha|k}\sum_{\substack{m|P\\ m\leq\sqrt{D}\\ (m,k)=\alpha}} h(m)\sum_{l|m}\frac{\mu(l)}{g(l)} \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n)\sum_{t|n}\frac{\mu(t)}{g(t)} r_{[t,l]} \sum_{d|m}\frac{\mu(d)}{g((d,k))}$$
$$= g(k)\sum_{\alpha|k}\sum_{\substack{m|P\\ m\leq\sqrt{D}\\ (m,k)=\alpha}} h(m)\sum_{l|m}\frac{\mu(l)}{g(l)} \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n)\sum_{t|n}\frac{\mu(t)}{g(t)} r_{[t,l]} \left(\begin{cases} \dfrac{\mu(\alpha)}{h(\alpha)} & \text{if } m=\alpha \\ 0 & \text{otherwise} \end{cases}\right)$$
$$= g(k)\sum_{\alpha|k}\mu(\alpha)\sum_{l|\alpha}\frac{\mu(l)}{g(l)} \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n)\sum_{t|n}\frac{\mu(t)}{g(t)} r_{[t,l]}$$
$$= g(k)\sum_{l|k}\frac{\mu(l)}{g(l)}\mu(l) \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n)\sum_{t|n}\frac{\mu(t)}{g(t)} r_{[t,l]}\sum_{\beta|k/l}\mu(\beta)$$
$$= g(k)\sum_{l|k}\frac{1}{g(l)} \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n)\sum_{t|n}\frac{\mu(t)}{g(t)} r_{[t,l]}\begin{cases} 1 & \text{if } k=l \\ 0 & \text{otherwise} \end{cases}$$
$$= \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n)\sum_{t|n}\frac{\mu(t)}{g(t)} r_{[t,l]}.$$

□

Now we return to the proof of Proposition (5.3.1).

Proof.

$$\sum_{\substack{d|P\\ d\leq\sqrt{D}}} \rho_d A_{[d,k]}$$

$$= S^+_{min} \sum_{\substack{d|P\\ d\leq\sqrt{D}}} \frac{\mu(d)g([d,k])}{g(d)} \sum_{\substack{m\equiv 0 (\text{mod } d)\\ m|P\\ m\leq\sqrt{D}}} h(m)$$

$$+\frac{S^+_{min}}{X} \sum_{\substack{d|P\\ d\leq\sqrt{D}}} \frac{\mu(d)}{g(d)} r_{[d,k]} \sum_{\substack{m\equiv 0 (\text{mod } d)\\ m|P\\ m\leq\sqrt{D}}} h(m)$$

$$-\frac{1}{J} \sum_{\substack{d|P\\ d\leq\sqrt{D}}} \frac{\mu(d)g([d,k])}{g(d)} \sum_{\substack{m\equiv 0 (\text{mod } d)\\ m|P\\ m\leq\sqrt{D}}} h(m) \sum_{l|m} \frac{\mu(l)}{g(l)} \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n) \sum_{t|n} \frac{\mu(t)}{g(t)} r_{[t,l]}$$

$$-\frac{1}{XJ} \sum_{\substack{d|P\\ d\leq\sqrt{D}}} \frac{\mu(d)}{g(d)} r_{[d,k]} \sum_{\substack{m\equiv 0 (\text{mod } d)\\ m|P\\ m\leq\sqrt{D}}} h(m) \sum_{l|m} \frac{\mu(l)}{g(l)} \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n) \sum_{t|n} \frac{\mu(t)}{g(t)} r_{[t,l]}.$$

Rearranging the order of summation and applying our lemmas, we see

$$\sum_{\substack{d|P\\ d\leq\sqrt{D}}} \rho_d A_{[d,k]}$$

$$= S^+_{min}\left(\delta_{k1} + \frac{1}{X} \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n) \sum_{t|n} \frac{\mu(t)}{g(t)} r_{[t,k]}\right)$$

$$-\frac{1}{J} \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n) \sum_{t|n} \frac{\mu(t)}{g(t)} r_{[t,k]}$$

$$-\frac{1}{XJ} \sum_{\substack{d|P\\ d\leq\sqrt{D}}} \frac{\mu(d)}{g(d)} r_{[d,k]} \sum_{\substack{m\equiv 0 (\text{mod } d)\\ m|P\\ m\leq\sqrt{D}}} h(m) \sum_{l|m} \frac{\mu(l)}{g(l)} \sum_{\substack{n|P\\ n\leq\sqrt{D}}} h(n) \sum_{t|n} \mu(t) g(t) r_{[t,l]}.$$

$\square$

Definition 5.3.5. *For this choice of* $\rho_d = \rho_d^0 + \rho_d^{(1)}$, *define*

$$S^+(\mathcal{A}) = \sum_{n|P(z)} a_n \left(\sum_{d|n} \rho_d\right)^2.$$

Let ρ_d^{min} *be the optimal choice of* ρ_d *so that*

$$\sum_{n|P(z)} a_n \sum_{d|n} \rho_d^{min} = S^+_{min}.$$

Define

$$R_1(k) = \sum_{\substack{d|P \\ d\le\sqrt{D}}} (\rho_d - \rho_d^{min}) A_{[d,k]}.$$

We recall that

$$\sum_{\substack{d|P \\ d\le\sqrt{D}}} \rho_d^{min} A_{[d,k]} = \delta_{k1} S^+_{min}$$

so that

$$\sum_{\substack{d|P \\ d\le\sqrt{D}}} \rho_d A_{[d,k]} = \delta_{k1} S^+_{min} + R_1(k).$$

We think of the remainder term $R_1(k)$ as being very small. With these definitions, from Proposition (5.3.1) we have

Proposition 5.3.6. *For all* $k|P$, $k \le \sqrt{D}$,

$$\delta_{k1} S^+_{min} + R_1(k) \quad = S^+_{min}\left(\delta_{k1} + \frac{1}{X} A_k\right) - \frac{1}{J} A_k - \frac{1}{XJ} B_k$$

where

$$A_k = \sum_{\substack{n|P \\ n\le\sqrt{D}}} h(n) \sum_{t|n} \frac{\mu(t)}{g(t)} r_{[t,k]}$$

and

$$B_k = \sum_{\substack{d|P \\ d\le\sqrt{D}}} \frac{\mu(d)}{g(d)} r_{[d,k]} \sum_{\substack{m\equiv 0 (\mathrm{mod}\ d) \\ m|P \\ m\le\sqrt{D}}} h(m) \sum_{l|m} \frac{\mu(l)}{g(l)} \sum_{\substack{n|P \\ n\le\sqrt{D}}} h(n) \sum_{t|n} \mu(t) g(t) r_{[t,l]}.$$

Corollary 5.3.7. *For all* $k|P$, $k \le \sqrt{D}$, *if*

$$0 \ne A_k = \sum_{\substack{n|P \\ n\le\sqrt{D}}} h(n) \sum_{t|n} \frac{\mu(t)}{g(t)} r_{[t,k]}$$

then

$$S^+_{min} = \frac{X}{J} + \frac{1}{J}\frac{B_k}{A_k} + \frac{X}{A_k} R_1(k)$$

where

$$B_k = \sum_{\substack{d|P \\ d\le\sqrt{D}}} \frac{\mu(d)}{g(d)} r_{[d,k]} \sum_{\substack{m\equiv 0 (\mathrm{mod}\ d) \\ m|P \\ m\le\sqrt{D}}} h(m) \sum_{l|m} \frac{\mu(l)}{g(l)} \sum_{\substack{n|P \\ n\le\sqrt{D}}} h(n) \sum_{t|n} \mu(t) g(t) r_{[t,l]}.$$

Chapter 6

Conclusion

In conclusion, we have seen the refinement of Selberg's choice of sieve in order to improve the current results in the sifting limit problem. In addition, we have seen the use of Selberg's sieve to give an improvement of the Brun-Titchmarsh theorem as well as an improvement of the large sieve inequality in special cases.

6.1 Further Research

There is still much work to be done on the sifting limit problem. With further investigation and improvements, our techniques will provide even better results. The work is very general and provides quite a bit of flexibility. By altering parameters, we have already seen great improvement. By optimizing these parameters, we hope to see even greater improvement.

Most likely, future research will focus on the work with successive approximations. In the course of my research, I have developed several tools that will be useful in this problem. By minimizing the main term and remainder term simultaneously, I hope to give an improvement on the upper bound in certain cases. This improvement would be especially useful in the case when Selberg's result gives a remainder term which is close the main term. With this improvement, a refinement of the Brun-Titchmarsh theorem and other results could be possible.

Bibliography

[1] Harold G. Diamond and H. Halberstam, *A Higher-Dimensional Sieve Method*, Cambridge University Press, Cambridge, 2008.

[2] John Friedlander and Henryk Iwaniec, *Opera dé Cribro*, AMS, Providence, 2010.

[3] Henryk Iwaniec and Emmanuel Kowalski, *Analytic Number Theory*, American Mathematical Society, Providence, 2004.

[4] Atle Selberg, *Collected Papers: Volume II*, Springer-Verlag, Berlin, 1991.

[5] ———, *Sifting Problems, Sifting Density, and Sieves*, Number Theory, Trace Formulas and Discrete Groups (Oslo, Norway, 1987), Academic Press, Inc., Boston, 1989, pp. 467-484.

Vita

Sara Blight

2005 - 2010	Ph. D. in Mathematics, Rutgers University
2002 - 2005	B. Sc. in Mathematics from University of Arizona
2002	Graduated from Desert Vista High School, Phoenix, AZ

Work Experience

Jan-May 2010	Teaching assistant for calculus II for life and social sciences.
Sep-Dec 2009	Teaching assistant for calculus I for mathematical and physical sciences.
Jan-May 2009	Graduate assistant for Professor Henryk Iwaniec.
Sep-Dec 2008	Course instructor for calculus II for mathematical and physical sciences.
June-Aug 2008	Adjunct Research Staff Member with the IDA Center for Communications Research, Princeton, NJ.
Sep 2007-May 2008	Graduate assistant for Professor Henryk Iwaniec.
Jan-May 2007	Teaching assistant for calculus I for liberal arts majors.
Sep-Dec 2006	Teaching assistant for calculus I for mathematical and physical sciences.
June-Aug 2006	Adjunct Research Staff Member with the IDA Center for Communications Research, Princeton, NJ.
Sep 2005-May 2006	Grader for Rutgers University math department.
May-Aug 2005	NSA Director's Summer Program

Publications

Aug 2008 Internal paper at IDA Center for Communications Research, Princeton, NJ.

Aug 2006 Internal paper at IDA Center for Communications Research, Princeton, NJ.

Aug 2005 Internal paper at NSA.