SIEVES

1. Introduction

Let (a_n) be a sequence of positive real numbers. In many cases, this will be the characterisitic function of a set, such as numbers of the form n^2+1. We wish to sift the sequence by a set of primes S. This will usually be all the primes, but occasionally might be, say, only those primes which are 1 mod 4. We let S_z be the primes in S that are less than z, and T_z their product.

One goal of sieve theory is to estimate

$$A(x, S_z) = \sum_{\substack{n \le x \\ \gcd(n, T_z)=1}} a_n.$$

In particular, if S contains all the primes and $z > \sqrt{x}$, then the above sum is essentially $\sum_{p \le x} a_p$ over the primes. Of course, this is difficult to obtain in practise.

The typical inputs to the sieve are asymptotic estimates for (a_n) in residue classes. That is, we need information on the sums

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

The d-uniformity in such estimates plays a large rôle in determining the strength of our theorems. We write $A(x) = A_1(x)$, and usually assume that $A_d(x) \sim g(d)A(x)$ for some multiplicative function $g(d)$ for some range of d. Another way of considering this is to write

$$A_d(x) = g(d)A(x) + r_d(x)$$

and to assume some sort of bound on the error/remainder terms r_d.

The following upper-bound is often called the "fundamental lemma of the sieve":

$$A(x, S_z) \ll A(x) \prod_{p \in S_z} (1 - g(p)) \qquad \text{for } z \text{ some small power of } x.$$

Typically we are able to take z about of size $\sqrt{x}$, at least when $A(x)$ is linear. Of course we need more assumptions about the uniformity of (a_n) to be able to prove such a bound.

1.1. **Acknowledgments.** Much of this is based upon notes by H. Iwaniec. Notes of D. Charles, B. Green, and A. Kumchev were also used, and I thank H. A. Helfgott for useful conversations.

2. Basic tools

2.1. **Inclusion-exclusion and Möbius inversion.** The famous Möbius inversion formula is based on the simple fact that

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

This can easily be shown by first noting that we can replace n by its squarefree kernel $\hat{n}$; letting k be the number of distinct prime factors of $\hat{n}$, there are $\binom{k}{l}$ divisors of $\hat{n}$ with exactly l prime factors, and so the above sum is

$$\sum_{l=0}^{k}(-1)^l\binom{k}{l},$$

so that the result follows from the binomial theorem. An alternative proof can be derived from the expanding the reciprocal Euler products $\zeta(s) = \prod_p(1-1/p^s)^{-1}$ and $1/\zeta(s) = \prod_p(1-1/p^s) = \sum_n \mu(n)/n^s$ as sums and multiplying them together.

Dedekind's version of Möbius inversion is that

$$g(n) = \sum_{d|n} f(d) \iff f(n) = \sum_{d|n}\mu(d)g(n/d) = \sum_{de=n}\sum \mu(d)g(e),$$

while Möbius originally phrased it more in terms of

$$G(x) = \sum_{n\le x} F(x/n) \iff F(x) = \sum_{m\le x}\mu(m)G(x/m).$$

A third form of this idea is that

$$F(d) = \sum_e G(de) \iff G(m) = \sum_n \mu(n)F(mn). \tag{1}$$

This can be proven by writing $d = nm$ on the left, multiplying by $\mu(n)$ and summing over n, writing $f = en$, swapping the order of summation, and then applying the above fact:

$$\begin{aligned}\sum_n \mu(n)F(nm) &= \sum_n \mu(n)\sum_e G(emn)\\ &= \sum_n \mu(n)\sum_{f\equiv 0(n)} G(fm) = \sum_f G(fm)\sum_{n|f}\mu(n) = G(m).\end{aligned}$$

The first form of Möbius inversion can be formalised via the operation of Dirichlet convolution of arithmetic functions. Given two such functions f and g, we define their convolution as

$$(f\star g)(n) = \sum_{ab=n}\sum f(a)g(b).$$

Note that δ acts as an identity in this algebra, as $\delta\star f = f$ for all f. We see that the first fact of above can be re-written as $\mu\star 1 = \delta$, so that μ is inverse to the constant function 1. The Möbius inversion formula is equivalent to

$$g = 1\star f \iff f = \mu\star g,$$

which follows from applying μ to each side of the left and using $\mu\star 1\star f = \delta\star f = f$.

The logarithm function $L(n) = \log n$ has an interesting place in this algebra. In particular, we have the additive relation $\log(ab) = \log(a) + \log(b)$, and so we get $L\cdot(f\star g) = (L\cdot f)\star g + f\star(L\cdot g)$, implying that multiplication by L is a derivation.

We define the von Mangoldt function as $\Lambda = \mu\star L$, so that

$$\Lambda(n) = \sum_{ab=n}\sum \mu(a)\log b = \sum_{d|n}\mu(d)\log\frac{n}{d}.$$

We can then use the fact that $\log \frac{n}{d} = \log n - \log d$ to get

$$\Lambda(n) = -\sum_{d|n} \mu(d) \log d + \log n \sum_{d|n} \mu(d),$$

and the second term is $\delta(n) \log n$ and thus vanishes. Via inversion we get that

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

and from this it follows that

$$\Lambda(p^k) = \begin{cases} \log p & n = p^l, l > 0 \\ 0 & \text{otherwise.} \end{cases}$$

2.2. **Notation, partial summation, and Cauchy's inequality.** We use $\tau(n)$ to denote the number of prime divisors of n; these are also given by the Dirichlet series coefficients for $\zeta(s)^2$. In general, we let $\tau_k(n)$ be the Dirichlet series coefficients for $\zeta(s)^k$, and this counts the number of distinct ways writing n as a product of k positive factors. We let $\omega(n)$ be the number of prime divisors of n.

We use the notation $f = O(g)$ to denote that f/g is bounded in some region of interest, and $f \ll g$ means the same. For sums over dyadic intervals, we denote the sum over $x < n \leq 2x$ as $n \sim x$. The floor of x, that is, the greatest integer less than x, is denoted by $\lfloor x \rfloor$, and the fractional part is given by $\{x\} = x - \lfloor x \rfloor$.

We recall the partial summation; this is easily proved by integration by parts using Riemann-Stieltjes integrals. Letting (a_n) be a sequence and $F(x)$ a differentiable function, we have that

$$\sum_{n \leq x} a_n F(n) = F(x) \sum_{n \leq x} a_n - \int_1^x \left(\sum_{n \leq t} a_n \right) F'(t)\, dt.$$

We also recall Cauchy's inequality, which states that if (b_n) and (c_n) are sequences of complex numbers, then

$$\left| \sum_n b_n c_n \right|^2 \leq \left(\sum_n |b_n|^2 \right) \left(\sum_n |c_n|^2 \right).$$

2.3. **Consequences from prime number theory.** We define slightly atypical prime-counting functions, that will be used in our proof of the Bombieri-Vinogradov theorem, and then later for small gaps between primes. Rather than the usual ψ-functions that count prime powers, we define

$$\hat{\psi}(x) = \sum_{p \sim x} \log p \quad \text{and} \quad \hat{\psi}(x; q, a) = \sum_{\substack{p \sim x \\ p \equiv a \pmod q}} \log p \quad \text{and} \quad \hat{\psi}(x, \chi) = \sum_{p \sim x} \chi(p) \log p,$$

which count only primes and count them in dyadic intervals. The prime number theorem states, in particular, that there is some absolute constant $\hat{c} > 0$ such that

$$\psi(x) = x + O\left(\frac{x}{e^{3\hat{c}\sqrt{\log x}}} \right)$$

and for nonprincipal Dirichlet characters χ modulo q with $q \leq e^{\hat{c}\sqrt{\log x}}$ we have

$$\left| \hat{\psi}(x, \chi) \right| \ll \frac{x}{e^{2\hat{c}\sqrt{\log x}}}$$

though here we need to be careful about possible Siegel zeros if we want this latter statement to be effective.

In particular, we will say that a character is S-Siegel if its L-function has a real zero $\beta \geq 1 - \frac{1/3}{\log S}$. It is then a standard theorem that there is at most one such primitive character with $q \leq S$. For all other characters with $q \leq e^{\hat{c}\sqrt{\log x}}$ we obtain the effective result that

$$\left|\hat{\psi}(x,\chi)\right| \ll (\log x)^2 \cdot x \exp\left(-\frac{\log x}{3\log S}\right) + xe^{-2\hat{c}\sqrt{\log x}} \tag{2}$$

Taking $S = e^{\hat{c}\sqrt{\log x}}$, when $2\hat{c} \leq 1/3\hat{c}$ we recover the same bound as above (obviously we can make $\hat{c}$ sufficiently small to ensure that this inequality is true). Since the class number formula implies that we have the bound $\beta \leq 1 - 1/\sqrt{q}$ for zeros of real L-functions of modulus q, we see that a modulus $q \leq (3\log S)^2$ cannot be S-Siegel.

Finally, we recall the theorem of Mertens, which states

$$\prod_{p\leq x}\left(1-\frac{1}{p}\right)^{-1} \sim e^{\gamma}\log x.$$

2.4. **Bounds on multiplicative functions.** Let f be a multiplicative, nonnegative function supported on the squarefree numbers. We assume that

$$\sum_{p\leq t} f(p)\log p = \kappa \log t + O(1) \tag{3}$$

for some $\kappa > 0$. We shall call this the regularity hypothesis (which could be made weaker, with similar effect on the results), and by partial summation it implies that

$$\sum_{p\leq t} f(p) = \kappa \log\log t + O(1).$$

We also need the mild assumption

$$\sum_{p} f(p)^2 \log p < \infty, \tag{4}$$

This boundedness implies

$$\sum_{d\leq t} f(d) \leq \prod_{p\leq t}\big(1+f(p)\big) \leq \exp\left(\sum_{p\leq t} f(p)\right) \ll (\log t)^{\kappa}.$$

We can now use the theory of multiplicative functions from Wirsing to get that (f being nonnegative and supported on squarefree integers):

Theorem 2.1. *With the above assumptions, we have that*

$$\sum_{d\leq t} f(d) = c_f(\log t)^{\kappa} + O\big((\log t)^{\kappa-1}\big) \quad \textit{where} \quad c_f = \frac{1}{\Gamma(\kappa+1)}\prod_p\left(1-\frac{1}{p}\right)^{\kappa}\big(1+f(p)\big).$$

A similar theorem can be shown when restricting the divisors of d to a given set of primes, but we do not do this here (we could just modify f).

Proof. We write $M_f(x) = \sum_{m\leq x} f(m)$. First, following Tchebyshev, we have that

$$M_f^L(x) = \sum_{m\leq x} f(m)\log m = \sum_{p\leq x}\sum_{n\leq x/p} f(np)\log p,$$

as $f(m)$ appears with multiplicity $\sum_{p|m} \log p = \log m$ in the right-hand expression.

Writing $\sum_{p\le t} f(p)\log p = \kappa \log t + E(t)$, this then simplifies to

$$\begin{aligned}
M_f^L(x) &= \sum_{n\le x} f(n) \sum_{\substack{p\le x/n\\ \gcd(p,n)=1}} f(p)\log p \\
&= \sum_{n\le x} f(n) \sum_{p\le x/n} f(p)\log p - \sum_{n\le x} f(n) \sum_{\substack{p|n\\ p\le x/n}} f(p)\log p \\
&= \sum_{n\le x} f(n)\Big[\kappa \log\frac{x}{n} + E(x/n)\Big] - \sum_{p\le x} f(p)\log p \sum_{\substack{n\equiv 0(p)\\ n\le x/p}} f(n) \\
&= \kappa \sum_{n\le x} f(n)\log\frac{x}{n} + \sum_{n\le x} f(n)E(x/n) - \sum_{p\le x} f(p)\log p \sum_{mp\le x/p} f(mp) \\
&= \kappa M_f(x)\log x - \kappa M_f^L(x) + \sum_{n\le x} f(n)E(x/n) - \sum\sum_{mp^2\le x} f(p)f(mp)\log p.
\end{aligned}$$

We move $\kappa M_f(x)\log x$ and $\kappa M_f^L(x)$ to the left side to get

$$(\kappa+1)M_f^L(x) - \kappa M_f(x)\log x = N_f(x), \tag{5}$$

where, since $E(t) = O(1)$ and using the boundedness hypothesis (4), we get

$$\begin{aligned}
N_f(x) &= \sum_{n\le x} f(n)E(x/n) - \sum\sum_{mp^2\le x} f(p)f(mp)\log p \\
&\ll M_f(x) + \sum_{m\le x} f(m) \sum_{p\le x} f(p)^2 \log p \ll M_f(x) \ll (\log x)^\kappa.
\end{aligned}$$

The idea is now that we have some sort of a functional equation for $M_f(x)$, and the smallness of $N_f(x)$ will force the desired asymptotic for $M_f(x)$.

To this end, via partial summation we note that

$$M_f^L(x) = M_f(x)\log x - \int_1^x M_f(u)\frac{du}{u},$$

so that (5) becomes

$$M_f(x)\log x - (\kappa+1)\int_1^x M_f(u)\frac{du}{u} = N_f(x). \tag{6}$$

It is convenient to move the starting point of the integral from 1 to 2, and this introduces an extra factor of $(\kappa+1)\log 2$, which we add to $N_f(x)$, the result being denoted $N_f^\star(x)$. We are now faced with a standard problem; if the integral involving $M_f(u)$ involved $N_f^\star(u)$ instead, we would have a nice expression for $M_f(x)$.

The main idea here is to multiply both sides of (6) by some factor $K(x)$ and then integrate from 2 to y. The factor will be chosen so that we get cancellation on the left side upon swapping the order of integration. We then substitute the resulting expression for the integral back into (6). We get

$$\int_2^y K(x)M_f(x)\log x\,dx - (\kappa+1)\int_2^y K(x)\int_2^x M_f(u)\frac{du}{u}\,dx = \int_2^y K(x)N_f^\star(x)\,dx,$$

and so by reversing the order of integration we get

$$\int_2^y K(x)N_f^\star(x)\,dx = \int_2^y K(x)M_f(x)\log x\,dx - (\kappa+1)\int_2^y\Big(\int_u^y K(x)\,dx\Big)M_f(u)\frac{du}{u}$$
$$= \int_2^y K(x)M_f(x)\log x\,dx + (\kappa+1)\int_2^y \tilde{K}(u)M_f(u)\frac{du}{u} - (\kappa+1)\int_2^y \tilde{K}(y)M_f(u)\frac{du}{u},$$

where $\tilde{K}$ is an antiderivative of K. The third term is exactly the sort of thing we want, so as to be able to replace the integral in (6). The first and second terms cancel when we choose $K(x)\log x = -(\kappa+1)\tilde{K}(x)/x$, so that we have

$$\log\tilde{K}(x) = -(\kappa+1)\int\frac{dx}{x\log x} = -(\kappa+1)\log\log x + C,$$

and so $\tilde{K}(x) = e^C/(\log x)^{\kappa+1}$, where we can choose the value of C. So we can take $K(x) = 1/x(\log x)^{\kappa+2}$, and replace the integral in (6) over $M_f(u)$ with one over $N_f^\star(u)$ to get

$$M_f(x)\log x = -\frac{1}{\tilde{K}(x)}\int_2^x K(u)N_f^\star(u)\,du + N_f^\star(x)$$
$$= (\kappa+1)(\log x)^{\kappa+1}\int_2^x \frac{N_f^\star(u)}{(\log u)^{\kappa+2}}\frac{du}{u} + N_f^\star(x).$$

The bound on $N_f^\star(x)$ implies that the integral is $c_f + O(1/\log x)$, and so by dividing through by $\log x$ we get the desired asymptotic.

We are left to compute c_f. Again the technique is well-known, as we consider $\zeta_f(s) = \sum_m f(m)/m^s$ as $s\to 0$ and compare it to $\zeta(s+1)^\kappa$. Writing the $\zeta_f(s)$ sum as an integral and using integration by parts and then substituting $x = e^t$ we get

$$\zeta_f(s) = \int_1^\infty \frac{dM_f(x)}{x^s} = -\int_1^\infty M_f(x)\,dx^{-s} = -\int_0^\infty M_f(e^t)\,de^{-st}$$
$$= -\int_0^\infty\Big[c_f + O\Big(\frac{1}{t}\Big)\Big]t^\kappa de^{-st} = \int_0^\infty\Big[c_f + O\Big(\frac{s}{u}\Big)\Big]\Big(\frac{u}{s}\Big)^\kappa e^{-u}\,du$$
$$= \frac{c_f}{s^\kappa}\Gamma(\kappa+1) + O\Big(\frac{s}{s^\kappa}\Big) \quad \text{as } s\to 0 \text{ as we substituted } u = st.$$

Of course, we have $\zeta(s+1)^\kappa \sim 1/s^\kappa$ as $s\to 0$, and the product over primes

$$\frac{\zeta_f(s)}{\zeta(s+1)^\kappa} = \prod_p\Big(1-\frac{1}{p^{s+1}}\Big)^\kappa\Big(1+\frac{f(p)}{p^s}\Big)$$

has a limit as $s\to 0$ by the regularity hypothesis. This gives the result. □

Here is another useful fact, generalised from the harmonic series.

Lemma 2.1. *Suppose that* $\sum_{p\le t} f(p)\log p \ll t$. *Then*

$$\sum_{n\le x} f(n) \ll \frac{x}{\log x}\sum_{n\le x}\frac{f(n)}{n}. \tag{7}$$

Proof. As above, we have the Tchebyshev relation, and get

$$\sum_{m\le x} f(m)\log m = \sum_{n\le x}\sum_{p\le\frac{x}{n}} f(np)\log p \le \sum_{n\le x} f(n)\sum_{p\le\frac{x}{n}} f(p)\log p \ll \sum_{n\le x} f(n)\frac{x}{n}.$$

This is a relation between the logarithmic weighting and the reciprocal weighting. By partial summation we get

$$\begin{aligned}\sum_{m\le x} f(m) &= f(1) + \frac{1}{\log x}\sum_{m\le x} f(m)\log m + \int_2^x \Big(\sum_{m\le t} f(m)\log m\Big)\frac{dt}{t(\log t)^2}\\ &\ll \frac{x}{\log x}\sum_{n\le x}\frac{f(n)}{n} + \int_2^x\Big(\sum_{m\le t}\frac{f(m)}{m}\Big)\frac{dt}{(\log t)^2}\\ &\ll \frac{x}{\log x}\sum_{n\le x}\frac{f(n)}{n} + \Big(\sum_{m\le x}\frac{f(m)}{m}\Big)\int_2^x\frac{dt}{(\log t)^2} \ll \frac{x}{\log x}\sum_{n\le x}\frac{f(n)}{n}.\end{aligned}$$

□

3. Sieves: phrasing of sieve problems

3.1. **A typical setup.** We wish to estimate

$$A(x, S_z) = \sum_{\substack{n\le x\\ \gcd(n,T_z)=1}} a_n.$$

Using the Möbius relation, we have that

$$A(x, S_z) = \sum_{\substack{n\le x\\ \gcd(n,T_z)=1}} a_n = \sum_{n\le x} a_n \sum_{d\mid \gcd(n,T_z)} \mu(d) = \sum_{d\mid T_z}\mu(d)\sum_{\substack{d\mid n\\ n\le x}} a_n.$$

The inner sum is exactly the sort of congruential sum that, as input to our sieve mechanism, we expect to be able to estimate asymptotically. Denoting it by $A_d(x)$, we get that

$$A(x, S_z) = \sum_{d\mid T_z}\mu(d)A_d(x).$$

We write

$$A_d(x) = g(d)A(x) + r_d(x),$$

where $g(d)$ is a multiplicative function; the idea of this decomposition is to have the error/remainder term $r_d(x)$ as small as possible. With this decomposition, we get that

$$A(x, S_z) = \sum_{d\mid T_z}\mu(d)g(d)A(x) + \sum_{d\mid T_z}\mu(d)r_d(x) = M(S_z)A(x) + R(x, S_z)$$

where

$$M(S_z) = \sum_{d\mid T_z}\mu(d)g(d) = \prod_{p\mid T_z}\big(1 - g(p)\big) \quad\text{and}\quad R(x, S_z) = \sum_{d\mid T_z}\mu(d)r_d(x)$$

and so we want g to be sufficiently nice to be able to make estimates here. The bulk of the problem is in getting good estimates for the error term $R(x, S_z)$.

3.2. **Twin primes.** Let (a_n) be the characteristic function of numbers of the form $m(m+2)$. When $m \le x$, if $m(m+2)$ has no prime factors less than $\sqrt{x}$, then both m and $m+2$ are prime. We wish to sift the sequence by the residue classes $0, -2 \pmod p$, ideally for all primes up to $\sqrt{x}$. For each prime $p > 2$, we should have that $g(p) = 2/p$, and it is not too difficult to estimate the remainder terms. Unfortunately, we are unable to take z as large as $\sqrt{x}$; we can take it almost this large, and this will give an upper bound.

3.3. **Small and large sieves.** The main difference between small and large sieves is the average size of $g(p)$. In fact, $pg(p)$ measures the number of residue classes mod p that are sieved out by our process. If the average of $pg(p)$ exists, it is often denoted by κ, and called the sieve dimension. The case where $\kappa = 1$ is extremely important, and is called the linear sieve. Also important is the half-linear sieve when $\kappa = 1/2$, and $\kappa = 2$ comes up as above in the twin prime problem. The large sieve typically refers to the case where $g(p)$ is itself not small on average. For instance, sifting by the quadratic non-residues mod p eliminates $(p-1)/2$ residue classes for each p.

3.4. **An example.** We consider the characteristic function of the integers $a_n = 1$ for all n. We wish to estimate

$$\Phi(x, S_z) = \sum_{\substack{n \le x \\ \gcd(n, T_z)=1}} 1,$$

particularly in the case that T_z is the product of the primes up to z. As we eliminate the 0 mod p class for each prime, it is clear that $g(p) = 1/p$ best models the situation. To estimate the remainder term $r_d(x)$ we first note that

$$\lfloor x \rfloor = \sum_{e|T_z} \sum_{\substack{n \le x \\ \gcd(n, T_z)=e}} 1 = \sum_{e|T_z} \sum_{\substack{m \le x/e \\ \gcd(m, T_z)=1}} 1 = \sum_{e|T_z} \Phi(\lfloor x/e \rfloor, T_z),$$

and so by Möbius inversion we have

$$\Phi(\lfloor x \rfloor, T_z) = \sum_{d|T_z} \mu(d) \lfloor x/d \rfloor.$$

Next we use $\lfloor x/d \rfloor = (x/d) - \{x/d\}$ to get that

$$\Phi(x, S_z) = \sum_{d|T_z} \mu(d)(x/d) - \sum_{d|T_z} \mu(d)\{x/d\} = x \prod_{p|T_z} (1 - 1/p) + R(x, S_z),$$

where as a trivial estimate we have $|R(x, S_z)| \le 2^{\#S_z}$ while the main term is given by $x\phi(T_z)/T_z$. This gives an asymptotic as long as T_z does not have too many prime factors. In particular, it fails when $\#S_z > \log x$.

In general we cannot do much better than this, but in this specific example we can use a trick (in its general form attributed to Rankin) to estimate $\{x/d\}$ more sharply. In particular, this is the case when $d > x$, which should be true for many factors of T_z. The inequality that we use is

$$\{x/d\} \le \min(1, x/d) \le (x/d)^\alpha,$$

where α can be chosen with $0 \le \alpha \le 1$. This gives us that

$$|R(x, S_z)| = \left| \sum_{d|T_z} \mu(d)\{x/d\} \right| \le \sum_{d|T_z} (x/d)^\alpha = x^\alpha \prod_{p|T_z} (1 + 1/p^\alpha).$$

We now insert $\frac{\phi(T_z)}{T_z} = \prod_{p \in S_z} (1 - 1/p)$ into the picture, so as to be able to derive asymptotics more easily, and then note that the multiplicand is bigger than 1 but decreases with p, and so we can replace the product over primes dividing T_z by the product of primes up to z to get

$$|R(x, S_z)| \le x^\alpha \frac{\phi(T_z)}{T_z} \prod_{p|T_z} (1+1/p^\alpha)(1-1/p)^{-1} \le x^\alpha \frac{\phi(T_z)}{T_z} \prod_{p \le z} (1+1/p^\alpha)(1-1/p)^{-1}.$$

Choosing α so that $z^{1-\alpha} = C$ we have $C/p \geq 1/p^\alpha$ for all $p \leq z$, and so

$$|R(x, T_z)| \leq x \frac{\phi(T_z)}{T_z} C^{-\log x/\log z} \prod_{p \leq z} (1 + C/p)(1 - 1/p)^{-1},$$

where the product is bounded by $O_C\big((\log z)^{C+1}\big)$ by the formula of Mertens. By choosing $\log x/\log z = B/\log\log x$, we thus get that $\Phi(x, T_z) \sim \frac{\phi(T_z)}{T_z} x$ provided that $\#S_z \leq \pi(z) \leq z = x^{1/B \log\log x}$ and $C + 1 < C^B$. As C can be freely chosen, we can take B as close to 1 as desired. This is much superior to our previous range of uniformity.

3.5. **Upper-bound and lower-bound sieves.** The above example shows how important it is to have good control over the remainder terms $r_d(x)$. One idea is to try to drop information before using the asymptotics for $A_d(x)$. For instance, we might be able to estimate

$$\sum_{d|T_z} \lambda_d^- A_d(x) \leq A(x, S_z) \leq \sum_{d|T_z} \lambda_d^+ A_d(x)$$

using various weights $\lambda_d^\pm$. This would be particularly useful if the $\lambda_d^\pm$ were 0 for large d.

Due to positivity we have that $A_d(x) \geq 0$. Recalling the expansion of

$$A(x, T_z) = \sum_{d|T_z} \mu(d) A_d(x),$$

we are thus led to seek systems of weights such that

$$\sum_{d|m} \lambda_d^- \leq \sum_{d|m} \mu(d) \leq \sum_{d|m} \lambda_d^+$$

for all m dividing T_z. Of course, the middle sum is just $\delta(m)$. We call these choices of $\lambda_d^\pm$ to be upper and lower bound sieves (though we introduce them together, we can have an upper bound sieve without a corresponding lower bound sieve, and vice-versa). If we have $\lambda_d^\pm = 0$ for $d > D$, then we say that the sieve is of level D.

The first sieves to be constructed were restrictions of the Möbius function to integers which had their primes factors lying in given intervals. This was largely developed by Brun following work of Merlin, and then later by Rosser. These are usually called combinatorial sieves. Though we shall not consider them here, they can be quite powerful albeit clumsy to use.

3.6. **Some sieve facts.** We give some more basic facts about sieves. We write $\sigma^\pm = \lambda^\pm \star 1$, so that the sieve condition is that $\sigma^- \leq \delta \leq \sigma^+$, at least on divisors of T_z. Given any multiplicative function f supported on the divisors of T_z (in particular on squarefree numbers), we use $\lambda^\pm = \sigma^\pm \star \mu$ to get

$$\sum_{d|T_z} \lambda_d^\pm f(d) = \sum_{d|T_z} \sum_{b|d} \sigma_b^\pm \mu(d/b) f(b) f(d/b) = \sum_{ab|T_z} \sum \sigma_b^\pm \mu(a) f(a) f(b)$$
$$= \sum_{b|T_z} \sigma_b^\pm f(b) \prod_{\substack{p \in S_z \\ p \nmid b}} \big(1 - f(p)\big) = \sum_{b|T_z} \frac{\sigma_b^\pm f(b)}{\prod_{p|b}(1 - f(p))} \cdot \prod_{p \in S_z} \big(1 - f(p)\big).$$

We assume that $0 \le f(p) < 1$ for all p. It turns out that the multiplicative function defined by

$$\hat{f}(p) = \frac{f(p)}{1 - f(p)}$$

is quite important. For instance, we have $1/\hat{f} = (1/f) \star \mu$, but here we note that

$$\big(1 - f(p)\big)^{-1} = \big(1 + \hat{f}(p)\big) = \frac{\hat{f}(p)}{f(p)},$$

so that we can write the above as

$$\sum_{d|T_z} \lambda_d^{\pm} f(d) = \sum_{b|T_z} \sigma_b^{\pm} \hat{f}(b) \cdot \prod_{p \in S_z} \big(1 - f(p)\big).$$

Noting that $\sigma_1^{\pm} = \lambda_1^{\pm} = 1$, the positivity/negativity of $\sigma^{\pm}$ gives us that

$$\sum_{d|T_z} \lambda_d^{-} f(d) \le \prod_{p \in S_z} \big(1 - f(p)\big) \le \sum_{d|T_z} \lambda_d^{+} f(d),$$

where the rightmost sum is also thus seen to be nonnegative.

The partial converse statement that

$$\sum_{d|T_z} \lambda_d^{+} f(d) \ll \prod_{p \in S_z} \big(1 - f(p)\big)$$

is also quite often true, and is indeed related to the fundamental lemma. However, unlike the above, it depends on a specific relation between the multiplicative function f and the sieve. In particular, when f satisfies the regularity (3) and boundedness (4) hypotheses, the choice of λ^+ in the Selberg upper-bound sieve will imply this inequality. In many situations, this can then be used to get an upper bound on $A(x, S_z)$.

4. The Selberg upper-bound sieve

This is a quite general upper-bound sieve first described by Selberg, who called it the Λ^2 sieve. There is also a related $\Lambda^2\Lambda^-$ lower-bound sieve, but this is not so useful, and we do not describe it here. As in the last section, we want an upper-bound sieve (λ_d) of level D, which means that $\lambda_1 = 1$ and $\lambda_d = 0$ for $d > D$ and $\sigma_m = \sum_{d|m} \lambda_d \ge 0$ for all m.

Selberg eases this last condition by choosing λ_d with

$$\sigma_m = \sum_{d|m} \lambda_d = \Big(\sum_{d|m} \rho_d\Big)^2,$$

where the ρ_d are real numbers with $\rho_1 = 1$ and $\rho_d = 0$ for $d > \sqrt{D}$. Obviously the squaring ensures the desired inequality. From this λ-ρ relation, by inducting on the number of prime divisors of l we get that

$$\lambda_l = \sum_{\mathrm{lcm}(a,b)=l} \sum \rho_a \rho_b. \tag{8}$$

When we apply this to our situation with $A(x, S_z)$, we get that

$$A(x,S_z) = \sum_{\substack{k\le x\\ \gcd(k,T_z)=1}} a_k \le \sum_{k\le x} a_k \Bigg(\sum_{d|\gcd(k,T_z)} \rho_d\Bigg)^2 = \sum_{m|T_z}\rho_m \sum_{n|T_z}\rho_n \sum_{\substack{k\equiv 0(m)\\ k\equiv 0(n)}} a_k =$$
$$= \sum_{m,n|T_z}\sum \rho_m\rho_n A_l(x) = A(x)\sum_{m,n|T_z}\sum g(l)\rho_m\rho_n + \sum_{m,n|T_z}\sum \rho_m\rho_n r_l(x),$$

where $l = \operatorname{lcm}(m,n)$ throughout. We wish to choose the ρ in such a way to optimise this inequality. If we ignore the remainder term, we can note that the main term is just a quadratic form in the ρ. We thus diagonalise this, and minimise it. The ρ that we obtain will be sufficiently small, in particular $|\rho_d| \le 1$, so as to make estimation of the remainder not be our main worry in most cases.

We assume that $0 < g(p) < 1$ for $p|T_z$ and $g(p) = 0$ otherwise. We also take ρ to be supported on divisors of T_z, and in particular squarefree numbers. The main term in the above is now given by

$$G = \sum_{l|T_z}\lambda_l g(l) = \sum_{m,n|T_z}\sum g(l)\rho_m\rho_n = \sum_{a,b,c|T_z}\sum\sum g(abc)\rho_{ac}\rho_{bc},$$

where we wrote $m = ac$ and $n = bc$ with $c = \gcd(m,n)$ so $abc = mn/c = \operatorname{lcm}(m,n)$. Our goal is make this look like a quadratic form, and then to diagonalise it so as to compute its minimum subject to $\rho_1 = 1$.

We pull out the c to get (dropping the T_z-condition; it is in the ρ-support), and use the multiplicativity of g and the squarefree support of ρ to get

$$G = \sum_c g(c) \sum_{\gcd(a,b)=1}\sum g(a)g(b)\rho_{ac}\rho_{bc}.$$

We next pull the standard trick of inserting the Möbius relation for something the is equal to 1, in this case, the gcd of a and b, and then swapping the sums yields

$$G = \sum_c g(c)\sum_a\sum_b g(a)g(b)\sum_{e|\gcd(a,b)}\mu(e)\rho_{ac}\rho_{bc} = \sum_c g(c)\sum_e \mu(e)\sum_{a\equiv 0(e)} g(a)\rho_{ac}\sum_{b\equiv 0(e)} g(b)\rho_{bc},$$

whereupon we replace $m = a/e$ and $n = b/e$ and use g-multiplicativity again to get

$$G = \sum_c g(c)\sum_e \mu(e)\sum_m g(em)\rho_{emc}\sum_n g(en)\rho_{enc}$$
$$= \sum_c \frac{1}{g(c)}\sum_e \mu(e)\Bigg(\sum_m g(cem)\rho_{cem}\Bigg)^2.$$

The inner sum is looking good for our diagonalised quadratic form, and we can simplify it by via the multiplicative function.

$$\hat g(p) = \frac{g(p)}{1-g(p)} = \frac{1}{-1+1/g(p)} \qquad \text{so that} \quad 1/\hat g = \big(1/g\big)\star\mu,$$

and by writing $n = ce$ and $s = cem$ we get

$$G = \sum_{n|T_z}(1/g\star\mu)(n)\Bigg(\sum_{s\equiv 0(n)} g(s)\rho_s\Bigg)^2 = \sum_{n|T_z}\frac{1}{\hat g(n)}\Bigg(\sum_{s\equiv 0(n)} g(s)\rho_s\Bigg)^2.$$

This can now be diagonalised by making the linear change of variables

$$\xi_d = \mu(d) \sum_{m \equiv 0(d)} g(m)\rho_m = \mu(d) \sum_e g(de)\rho_{de}, \quad \text{so that} \quad G = \sum_{d|T_z} \frac{\xi_d^2}{\hat{g}(d)}.$$

We immediately see that $\xi_d = 0$ for $d > \sqrt{D}$. We can invert the ξ-ρ relation by using the third Möbius relation (1) to get

$$\rho_l g(l) = \sum_n \mu(n)\xi_{nl}\mu(nl)$$

so that

$$\rho_l = \frac{\mu(l)}{g(l)} \sum_{d \equiv 0(l)} \xi_d,$$

and with $l = 1$ this gives $1 = \rho_1 = \sum_d \xi_d$, where the sum is over divisors of T_z.

We want to minimise G on this hyperplane. Via Cauchy's inequality we have

$$GH = \Big(\sum_{d|T_z} \frac{\xi_d^2}{\hat{g}(d)}\Big)\Big(\sum_{\substack{d|T_z \\ d \le \sqrt{D}}} \hat{g}(d)\Big) \ge \Big(\sum_{d|T_z} \xi_d\Big)^2 = 1^2 = 1. \tag{9}$$

Incidentally, here we have that

$$H \le \sum_{d|T_z} \hat{g}(d) = \prod_{p \in S_z} \big(1 + \hat{g}(p)\big) = \prod_{p \in S_z} \big(1 - g(p)\big)^{-1}.$$

Equality occurs in the above when $\xi_d = \hat{g}(d)/H$ for $d \le \sqrt{D}$, and this gives

$$\rho_m = \frac{\mu(m)}{Hg(m)} \sum_{\substack{d \le \sqrt{D} \\ d \equiv 0(m)}} \hat{g}(d) = \frac{\mu(m)\hat{g}(m)}{Hg(m)} \sum_{\substack{e \le \sqrt{D}/m \\ \gcd(e,m)=1}} \hat{g}(e). \tag{10}$$

In particular, we can see that $|\rho_m| \le 1$ (the argument is due to van Lint and Richert) by grouping the terms of H according to the gcd of d and m to get

$$\begin{aligned} H &= \sum_{k|m} \sum_{\substack{d \le \sqrt{D} \\ \gcd(d,m)=k}} \hat{g}(d) = \sum_{k|m} \hat{g}(k) \sum_{\substack{e \le \sqrt{D}/k \\ \gcd(e,m)=1}} \hat{g}(e) \\ &\ge \sum_{k|m} \hat{g}(k) \cdot \sum_{\substack{e \le \sqrt{D}/m \\ \gcd(e,m)=1}} \hat{g}(e) = \frac{\hat{g}(m)}{g(m)} \cdot \mu(m)\rho_m \frac{g(m)}{\hat{g}(m)} H = \mu(m)\rho(m)H, \end{aligned}$$

where we used that

$$\sum_{k|m} \hat{g}(k) = \prod_{p|m} \big(1 + \hat{g}(p)\big) = \prod_{p|m} \frac{\hat{g}(p)}{g(p)} = \frac{\hat{g}(m)}{g(m)}.$$

By the formula (8) for λ_d, this implies that $|\lambda_d| \le \tau_3(d)$.

We can now recap most of the above:

Theorem 4.1. *We have*

$$A(x, S_z) = \sum_{\substack{n \le x \\ \gcd(n,T_z)=1}} a_n \le \frac{A(x)}{H_D} + R^D(x, S_z),$$

where

$$H_D = \sum_{\substack{d \le \sqrt{D} \\ d|T_z}} \hat{g}(d) \quad \textit{and} \quad R^D(x, S_z) = \sum_{m,n|T_z} \sum |\rho_m \rho_n r_l(x)|.$$

Of course, for this to be of use, we need a good lower bound for H_D, and some control over $R^D(x, S_z)$. For simplicity, we will now take S to be the set of all primes and $z > \sqrt{D}$.

In order to be able to lower-bound H_D, we need to assume that

$$\sum_{p\le t} g(p)\log p = \kappa \log t + O(1) \quad \text{and} \quad \sum_p g(p)^2 \log p < \infty,$$

so as to use the section on multiplicative functions. The relationship between the functions g and $\hat{g}$ implies that if the above assumptions hold for g, then since we have $\hat{g}(p) = g(p) + O\big(g(p)^2\big)$, they also hold for $\hat{g}$, though maybe with different implied constants. We get that

$$H_D = c_{\hat{g}}(\log\sqrt{D})^{\kappa}\left[1 + O\Big(\frac{1}{\log D}\Big)^{-1}\right].$$

A typical bound for the remainder terms would be something like $|r_d(x)| \le g(d)d$, or possibly with an extra $\tau(d)$ on the right. In many cases, the exact power of logarithm does not matter much. Under the above assumption, and assuming that $pg(p) \ge 1$ for $p \in S_z$ (so that $dg(d)$ is multiplicatively non-decreasing) we get

$$\begin{aligned} R^D(x, S_z) &\le \sum_{m,n|T_z}\sum |\rho_m\rho_n g(l)l| \le \sum_{m,n|T_z}\sum |\rho_m\rho_n g(mn)mn| = \Big(\sum_{d\le\sqrt{D}} dg(d)|\rho_d|\Big)^2 \\ &\le \Big(\frac{1}{H_D}\sum_{d\le\sqrt{D}} d\hat{g}(d)\sum_{\substack{u\le\sqrt{D}/d\\ \gcd(d,u)=1}}\hat{g}(u)\Big)^2 \le \Big(\frac{1}{H_D}\sum_{v\le\sqrt{D}}\sigma(v)\hat{g}(v)\Big)^2 \end{aligned}$$

multiplied and divided by H_D using (10), and then wrote $v = ud$, taking $\sigma(v)$ as the sum of the divisors of v. Using the bound (7) from multiplicative functions, we get a sum over $\sigma(v)\hat{g}(v)/v$, which is multiplicative and regular with the same κ as $\hat{g}$, and so we get:

$$\begin{aligned} R^D(x, S_z) &\le \Big(\frac{1}{H_D}\sum_{v\le\sqrt{D}}\sigma(v)\hat{g}(v)\Big)^2 \ll \Big(\frac{1}{H_D}\frac{\sqrt{D}}{\log D}\sum_{v\le\sqrt{D}}\frac{\sigma(v)\hat{g}(v)}{v}\Big)^2 \\ &\ll \Big(\frac{1}{H_D}\frac{\sqrt{D}}{\log D}\cdot H_D\Big)^2 = \frac{D}{(\log D)^2} \end{aligned}$$

Our final theorem thus comes as:

Theorem 4.2. *Suppose that (a_n) is a nonnegative sequence with*

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

for some $\kappa > 0$ and $|r_d(x)| \le dg(d)$ with $pg(p) \ge 1$ when it is nonzero and $\sum_p g(p)^2\log p < \infty$. Then for $z > D$ we have

$$A(x, S_z) = \sum_{\substack{n\le x\\ \gcd(n,T_z)=1}} a_n \ll \frac{A(x)}{(\log D)^{\kappa}} + \frac{D}{(\log D)^2},$$

and when $D(\log D)^{\kappa-2} \ll A(x)$ we have

$$A(x, S_z) \ll A(x)\prod_{p\le D}\big(1 - g(p)\big).$$

Here we again used the regularity hypothesis for $g(p)$ in the last step. Note that when $A(x)$ is essentially linear, we can take $\sqrt{D}$ to be almost as large as $\sqrt{x}$.

One can show that our choice in Selberg's sieve gives

$$\rho_d \sim \mu(d)\left(\frac{\log\sqrt{D}/d}{\log\sqrt{D}}\right)^{\kappa}$$

for d not too large, and the weights on the right side are sometimes called the quasi-optimal weights; they do not lose much, and are often easier to analyse.

We return to the idea that, for a given multiplicative function f supported on the squarefree numbers and satisfying the regularity (3) and boundedness (4) hypotheses, we should have some sieve λ_d^+ with

$$\sum_{d<D}\lambda_d^+ f(d) \ll \prod_{p<D}\big(1-f(p)\big)$$

The left sum is just G, and so the result follows from multiplicative function theory as above.

5. Applications of the Selberg sieve

5.1. **Twin primes.** First we recall the above setting of twin primes. We let (a_n) be the characteristic function of numbers of the form $m(m+2)$. For $m \le x$, we have that m and $m+2$ are both prime when the product has no prime factor as large as $\sqrt{x}$. Conversely, if $m \le \sqrt{x}$ with both m and $m+2$ prime, then it has no prime factor smaller than $\sqrt{x}$. Writing $\pi_2(x)$ for the number of $m \le x$ with m and $m+2$ both prime we get the bound

$$\pi_2(x) - \pi_2(\sqrt{x}) \le A(x^2+2x, S_z)$$

where $z = \sqrt{x}$ and S is the set of all primes.

We use the Selberg upper-bound sieve with $g(p) = 2/p$ for $p > 2$, so that we have the regularity condition with $\kappa = 2$. We have that $m(m+2) \equiv 0 \pmod p$ precisely when $m \equiv 0, -2 \pmod p$. Thus we have that

$$A_p(x^2+2x) = \sum_{\substack{m\le x\\ m\equiv 0,-2(p)}} 1 = \frac{2}{p}\sum_{m\le x} 1 + E,$$

where $|E| \le 2$. From this, arguing in the same manner for composite moduli, we readily obtain that $|r_d(x)| \le dg(d)$. The Selberg sieve gives that

$$A(x^2+2x, S_z) \le \frac{x}{H_D} + O\left(\frac{D}{(\log D)^2}\right),$$

and we choose $D = x/(\log x)$ so that

$$H_D \sim (\log\sqrt{D})^{\kappa}\frac{1}{\Gamma(\kappa+1)}\prod_p\left(1-\frac{1}{p}\right)^{\kappa}\big(1+\hat{g}(p)\big)$$

to get that

$$\pi_2(x) \le \frac{x\cdot 2^2\cdot 2!}{(\log x)^2}\prod_p\big(1-g(p)\big)\left(1-\frac{1}{p}\right)^{-\kappa}\cdot\left[1+O\left(\frac{\log\log x}{\log x}\right)\right].$$

The main term is larger by a factor of $\kappa!\cdot 2^{\kappa}$ than the expected amount. The argument works in the same manner for other polynomials, with $pg(p)$ being the number of roots mod p of the polynomial.

5.2. **The Brun-Titchmarsh theorem.** Here we let (a_n) be the characteristic function of an arithmetic progression in some given interval, namely we want that $n \equiv a \pmod q$ and $x < n \leq x + y$ where $\gcd(a, q) = 1$ and $q < y$. Our set S of primes shall be those with $p \nmid q$, and z will be taken as x. As everything with a prime divisor less than x is sieved out, we have the inequality

$$\pi(x + y; q, a) - \pi(x; q, a) \leq A(x + y, S_x),$$

where the π are the usual prime-counting functions in arithmetic progressions.

We clearly want $g(p) = 1/p$ when $p \nmid q$, and as we are just counting arithmetic progressions, we see that

$$A_p(x + y) = \sum_{\substack{n \equiv a(q) \\ n \equiv 0(p)}} a_n = \frac{y}{pq} + E$$

where $|E| \leq 1$, thus giving us the remainder condition. We could estimate H_D via elementary means, but instead just note that

$$H_D \sim \log \sqrt{D} \cdot \prod_{p|q} (1 - 1/p) = \frac{\phi(q)}{2q} \log D$$

using multiplicative functions. From the Selberg sieve, since $A(x + y) = y/q + E$, we thus get

$$\pi(x + y; q, a) - \pi(x; q, a) \leq \frac{2y}{\phi(q) \log D} + O\left(\frac{D}{(\log D)^2}\right),$$

and by choosing $D = y/q$ we get the standard bound

$$\pi(x + y; q, a) - \pi(x; q, a) \leq \frac{2y}{\phi(q) \log(y/q)} + O\left(\frac{y/q}{\big(\log(y/q)\big)^2}\right).$$

Thus we get an upper bound that is essentially twice the expected size. Montgomery and Vaughan have shown that the error term can be omitted. Typically in applications we will have y small compared to x, say of size $\log x$ maybe, while q will be fixed. There is also a relationship here with Siegel zeros, as if the factor of two could be lowered, then they do not exist (see the work of Selberg).

6. The large sieve

The large sieve developed in various parts of mathematics. It can be seen as a tool in functional analysis, or probability theory, but we shall look at its arithmetic applications. The main idea is to estimate the trigonometric polynomial

$$S(\alpha) = \sum_n a_n e(\alpha n),$$

where we have the standard notation $e(x) = e^{2\pi i x}$, and the a_n are complex numbers supported on $M < n \leq M + N$. Via Cauchy's inequality we get the bound

$$|S(\alpha)|^2 = \left|\sum_n a_n e(\alpha n)\right|^2 \leq \sum_{M<n\leq M+N} |e(\alpha n)|^2 \sum_n |a_n|^2 = N \cdot \sum_n |a_n|^2.$$

This must be best-possible for a given α, but if we have a sequence of well-spaced α's, then we might expect to get some cancellation.

The usual hypothesis is that

$$\|\alpha_r - \alpha_s\| \geq \delta \qquad \text{for } r \neq s,$$

where $\|x\|$ is the distance to the nearest integer (this is natural on the torus $\mathbf{R}/\mathbf{Z}$). We say that such a sequence is δ-spaced. The large sieve inequality (which we shall prove in its dual form) will state that

$$\sum_r |S(\alpha_r)|^2 \leq (\delta^{-1} + N) \sum_n |a_n|^2, \tag{11}$$

and indeed this is essentially best possible. Applications to number theory come about from, say, taking the Dirichlet characters mod q for $q \leq Q$, as for these the distinct character values differ by at least $1/Q^2$.

6.1. **Generalised Hilbert's inequality.** The first step in proving the large sieve inequality is to prove a generalised version of Hilbert's inequality. (It is related to the classical Hilbert transform.)

Lemma 6.1. *Let λ_r be real numbers with $\lambda_{r+1} - \lambda_r \geq \delta$. Then for any complex numbers (z_r) we have*

$$\left| \sum\sum_{r \neq s} \frac{z_r \bar{z}_s}{\lambda_r - \lambda_s} \right| \leq \frac{\pi}{\delta} \sum_r |z_r|^2.$$

Proof. We note that we can rescale any nonzero $\vec{z}$ to have norm 1. We shall estimate

$$W = \sum_r \left| \sum_{s \neq r} \frac{\bar{z}_s}{\lambda_r - \lambda_s} \right|^2$$

and then apply Cauchy's inequality. We note that for W we are estimating the norm of the matrix/operator $M = (\mu_{rs})$ with $\mu_{rs} = (\lambda_r - \lambda_s)^{-1}$ for $r \neq s$ and $\mu_{rr} = 0$ for all r. Indeed, we have that $\sum_{s \neq r} \frac{z_s}{\lambda_r - \lambda_s}$ is the rth component of $M\vec{z}$, and so we get $W = \|M\vec{z}\| \leq \|M\|$. We can also assume that $\vec{z}$ is extremal for M. The skew-symmetry (or more precisely, skew-Hermiticity) of the matrix/operator implies that such an extremal vector is an eigenvector, which says that there is some real number η for which we have

$$\sum_{k \neq u} \frac{z_k}{\lambda_k - \lambda_u} = i\eta z_u$$

for each u.

By expanding the square in W and isolating the diagonal we get that

$$\begin{aligned} W &= \sum_s \sum_t \bar{z}_s z_t \sum_{r \neq s,t} \frac{1}{\lambda_r - \lambda_s} \frac{1}{\lambda_r - \lambda_t} \\ &= \sum_s |z_s|^2 \sum_{r \neq s} \frac{1}{(\lambda_r - \lambda_s)^2} + \sum\sum_{s \neq t} \frac{\bar{z}_s z_t}{\lambda_s - \lambda_t} \sum_{r \neq s,t} \left[\frac{1}{\lambda_r - \lambda_s} - \frac{1}{\lambda_r - \lambda_t} \right]. \end{aligned}$$

We further simplify the last inner sum by using that

$$\sum_{r \neq s,t} \left[\frac{1}{\lambda_r - \lambda_s} - \frac{1}{\lambda_r - \lambda_t} \right] = \frac{2}{\lambda_s - \lambda_t} + \sum_{r \neq s} \frac{1}{\lambda_r - \lambda_s} - \sum_{r \neq t} \frac{1}{\lambda_r - \lambda_t}$$

so that we have

$$W = \sum_s |z_s|^2 \sum_{r\neq s} \frac{1}{(\lambda_r-\lambda_s)^2} + \sum\sum_{s\neq t} \frac{2\bar z_s z_t}{(\lambda_s-\lambda_t)^2}$$
$$+ \sum\sum_{s\neq t} \frac{\bar z_s z_t}{\lambda_s-\lambda_t} \sum_{r\neq s} \frac{1}{\lambda_r-\lambda_s} - \sum\sum_{s\neq t} \frac{\bar z_s z_t}{\lambda_s-\lambda_t} \sum_{r\neq t} \frac{1}{\lambda_r-\lambda_t}.$$

To simplify the second line, we use the eigenvector property for the t-sum in the first term and the s-sum in the second, to get that it is

$$\sum\sum_{r\neq s} \frac{-i\eta z_s \bar z_s}{\lambda_r-\lambda_s} - \sum\sum_{r\neq t} \frac{\overline{i\eta z_t} z_t}{\lambda_r-\lambda_t} = 0.$$

Thus for extremal vectors we can apply $2|z_s z_t| \le |z_s|^2 + |z_t|^2$ to get

$$W = \sum_s |z_s|^2 \sum_{r\neq s} \frac{1}{(\lambda_r-\lambda_s)^2} + \sum\sum_{s\neq t} \frac{2\bar z_s z_t}{(\lambda_s-\lambda_t)^2} \le 3\sum_s |z_s|^2 \sum_{r\neq s} \frac{1}{(\lambda_r-\lambda_s)^2}.$$

Finally we use the hypothesis that $|\lambda_r-\lambda_s| \ge \delta|r-s|$ to get that

$$W \le 3\sum_s |z_s|^2 \sum_{r\neq s} \frac{1}{(\lambda_r-\lambda_s)^2} \le 3\sum_s |z_s|^2 \sum_{r\neq s} \frac{1}{\delta^2(r-s)^2} \le 3\sum_s |z_s|^2 \cdot \frac{2\zeta(2)}{\delta^2}.$$

This gives us an estimate for W, and now we use Cauchy's inequality to get that

$$\left|\sum\sum_{r\neq s} \frac{z_r\bar z_s}{\lambda_r-\lambda_s}\right|^2 \le \sum_r |z_r|^2 \cdot \sum_r \left|\sum_{s\neq r} \frac{\bar z_s}{\lambda_r-\lambda_s}\right|^2 \le W\sum_r |z_r|^2 \le \left(\sum_r |z_r|^2\right)^2 \frac{\pi^2}{\delta^2},$$

from which the result follows. □

6.2. **The large sieve inequality.** Next we relate this generalised Hilbert inequality to exponential sums. We have the following lemma.

Lemma 6.2. *Let α_r be δ-spaced, and z_r be complex. Then*

$$\left|\sum\sum_{r\neq s} \frac{z_r\bar z_s}{\sin\pi(\alpha_r-\alpha_s)}\right| \le \delta^{-1}\sum_r |z_r|^2.$$

Proof. The idea is to apply the previous lemma to the doubly indexed sets given by $z_{m,r} = (-1)^m z_r$ and $\lambda_{m,r} = m+\alpha_r$ for $1 \le m \le K$. Of course, we will eventually let $K\to\infty$, as this is an averaging process over the m. This gives

$$\left|\sum\sum_{(r,m)\neq(s,n)} \frac{(-1)^{m-n} z_r\bar z_s}{m-n+\alpha_r-\alpha_s}\right| \le \frac{\pi}{\delta}\sum_r |z_r|^2 \sum_{m=1}^K |(-1)^m|^2.$$

Noting that when $r=s$ the (m,n) term cancels the (n,m) term for $n\neq m$, we change the sum-restriction to $r\neq s$, and write $k=m-n$ and divide by K to get

$$\left|\sum\sum_{r\neq s} z_r\bar z_s \sum_{k=-K}^K \frac{K-|k|}{K}\frac{(-1)^k}{k+\alpha_r-\alpha_s}\right| \le \frac{\pi}{\delta}\sum_r |z_r|^2,$$

as the Fejer kernel appears from counting how often k appears. By letting $K\to\infty$ and recalling that $\sum_k \frac{(-1)^k}{k+\alpha} = \frac{\pi}{\sin\pi\alpha}$ for α non-integral, we get the result. □

Corollary 6.1. *For any real x we have*

$$\left|\sum\sum_{r\neq s} z_r\bar{z}_s \frac{\sin 2\pi x(\alpha_r-\alpha_s)}{\sin\pi(\alpha_r-\alpha_s)}\right| \leq \delta^{-1}\sum_r |z_r|^2.$$

Proof. Apply the preceding lemma to $z_r e(x\alpha_r)$ and $z_r e(-x\alpha_r)$. □

We are now ready to prove the large sieve inequality. Since (in Banach spaces) a linear operator and its adjoint (conjugate transpose in the matrix setting) have the same norm, we shall show the dual inequality of (11).

Theorem 6.1. *For any complex numbers z_n and δ-spaced α_r we have*

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

Proof. By squaring out the left side, the diagonal gives $N\sum_r |z_r|^2$, and the rest is

$$\sum\sum_{r\neq s} z_r\bar{z}_s \sum_{M<n\leq M+N} e\big(n(\alpha_r-\alpha_s)\big) = \sum\sum_{r\neq s} z_r\bar{z}_s e\big(k(\alpha_r-\alpha_s)\big)\frac{\sin\pi N(\alpha_r-\alpha_s)}{\sin\pi(\alpha_r-\alpha_s)}$$

where $k = M+\frac{1}{2}(N+1)$. This exponential part can be inserted into the z_r and z_s, and so by the corollary we get a bound of $\delta^{-1}\sum_r |z_r|^2$. □

Corollary 6.2. *Let (a_n) be complex numbers supported on $M<n\leq M+N$. Then*

$$\sum_{q\leq Q}\sideset{}{^\star}\sum_{b\,(\mathrm{mod}\,q)} \big|S(b/q)\big|^2 = \sum_{q\leq Q}\sideset{}{^\star}\sum_{b\,(\mathrm{mod}\,q)}\left|\sum_{M<n\leq M+N} a_n e(bn/q)\right|^2 \leq (Q^2+N)\sum_n |a_n|^2.$$

Proof. The star restricts the sum to coprime residue classes, and the result is immediate as for distinct $\frac{a_1}{q_1}$ and $\frac{a_2}{q_2}$ we have $\left|\frac{a_1}{q_1}-\frac{a_2}{q_2}\right| \geq \frac{1}{q_1q_2} \geq \frac{1}{Q^2}$. □

Consider a sequence of complex numbers all of norm one. Then we might expect $S(b/q)$ to be size $\sqrt{N}$ typically, giving us a total of Q^2N when summing over b and q. The bound we would get is then $(Q^2+N)N$, and thus it is seemingly the diagonal contribution which scuppers our accounting. This might appear rather strange, because it says that taking N large, that is, having large inner sums, is bad, but recall that we square out and then swap the summations, so the cancellation is detected elsewise.

6.3. **Estimates on sifted sets.** Let (a_n) be a finitely supported sequence. We take a slightly different viewpoint than before, and consider the sifted sum

$$A(U) = \sum_{n\in U} a_n$$

where U is a set of squarefree integers which satisfy various congruence conditions modulo primes. For instance, we might take U to be the set of integers for which -1 is a square. For each prime p, we let $\varpi(p)$ be the number of residue classes modulo p which are not in U. Then we let $\hat{g}(p) = \frac{\varpi(p)}{p-\varpi(p)}$ and extend it multiplicatively to squarefree integers. We consider the exponential sum and congruence sums given by

$$S_U(\alpha) = \sum_{n\in U} a_n e(\alpha n) \quad\text{and}\quad X_q(b) = \sum_{\substack{n\equiv b\,(\mathrm{mod}\,q)\\ n\in U}} a_n.$$

Lemma 6.3. *Let q be squarefree. We have*

$$\hat{g}(q)|S_U(0)|^2 \le \sideset{}{^\star}\sum_{b \,(\mathrm{mod}\, q)} \big|S_U(b/q)\big|^2$$

We might first comment on the usefulness of this. Suppose that the reduction of U mod p only has one element for each prime p, so that $\hat{g}(p) = p-1$. As might be expected, this says that all the sums on the right are all the same size as each sum on the right is just a phase shift of $S_U(0)$. Conversely, if the reduction of U has $p-1$ elements, then $\hat{g}(p) = 1/(p-1)$, and the sum on the left can be estimated by the large sieve, even though this is essentially the small sieve case. Finally, when U mod p has about $p/2$ elements (quadratic residues for instance), then $\hat{g} \sim 1$ — this can be useful in the case where (a_n) is real and we can show cancellation in the sums on the right.

Proof. First consider $q = p$ a prime. Sorting by b mod p, we have that

$$|S_U(0)|^2 = \Big|\sum_{b \,(\mathrm{mod}\, p)} \sum_{\substack{n \equiv b \,(\mathrm{mod}\, p)\\ n \in U}} a_n\Big|^2 = \Big|\sum_{b \,(\mathrm{mod}\, p)} X_p(b)\Big|^2 \le \Big(\sum_{\substack{b \,(\mathrm{mod}\, p)\\ X_p(b)\neq 0}} 1^2\Big) \sum_{b \,(\mathrm{mod}\, p)} |X_p(b)|^2,$$

the last step by Cauchy's inequality. The number of congruence classes for which $X_p(b) = 0$ is at least $\varpi(p)$, while the last sum can be re-written via orthogonality of additive characters to get

$$|S_U(0)|^2 \le \big(p - \varpi(p)\big) \sum_{b \,(\mathrm{mod}\, p)} |X_p(b)|^2 = \big(p - \varpi(p)\big) \sum_{b \,(\mathrm{mod}\, p)} \sum\sum_{\substack{m,n \equiv b \,(\mathrm{mod}\, p)\\ m,n \in U}} a_m \bar{a}_n$$

$$= \big(p - \varpi(p)\big) \sum\sum_{\substack{m \equiv n \,(\mathrm{mod}\, p)\\ m,n \in U}} a_m \bar{a}_n = \big(p - \varpi(p)\big) \sum\sum_{m,n \in U} a_m \bar{a}_n \frac{1}{p} \sum_{b \,(\mathrm{mod}\, p)} e(bm/p) e(-bn/p)$$

$$= \frac{p - \varpi(p)}{p} \sum_{b \,(\mathrm{mod}\, p)} \Big|\sum_{m \in U} a_m e(bm/p)\Big|^2 = \frac{p - \varpi(p)}{p} \sum_{b \,(\mathrm{mod}\, p)} \big|S_U(b/p)\big|^2$$

which gives the result for prime moduli upon moving $|S_U(0)|^2$ to the left.

For $q = q_1 q_2$ with $\gcd(q_1, q_2) = 1$, we use the familiar Chinese Remainder trick

$$\sideset{}{^\star}\sum_{b \,(\mathrm{mod}\, q)} |S_U(b/q)|^2 = \sideset{}{^\star}\sum_{b_1 \,(\mathrm{mod}\, q_1)} \sideset{}{^\star}\sum_{b_2 \,(\mathrm{mod}\, q_2)} |S_U(b_1/q_1 + b_2/q_2)|^2.$$

Assuming the bound holds for q_1 and q_2, we change $a_n \to a_n e(nb_1/q_1)$ and estimate the q_2-sum and get

$$\sum_{b \,(\mathrm{mod}\, q)} |S_U(b/q)|^2 \ge \hat{g}(q_2) \sum_{b_1 \,(\mathrm{mod}\, q_1)} |S_U(b_1/q_1)|^2 \ge \hat{g}(q_2)\hat{g}(q_1)|S_U(0)|^2,$$

and the result follows by inducting on the number of prime factors of q. □

In particular, by summing over $q \le Q$ we get that

$$\sum_{q \le Q} \hat{g}(q) \cdot |S_U(0)|^2 \le \sum_{q \le Q} \sideset{}{^\star}\sum_{b \,(\mathrm{mod}\, q)} \big|S_U(b/q)\big|^2 \le (Q^2 + N) \sum_n |a_n|^2,$$

and by taking (a_n) to be the characteristic function of an interval, we get an upper bound on the sifted sum $S_U(0)$ similar to that from the Selberg upper-bound sieve.

7. The Bombieri-Vinogradov theorem

We now prove the Bombieri-Vinogradov theorem. For $Q \le x/e^{6\hat{c}\sqrt{\log x}}$ this says

$$\sum_{q\le Q}\max_{(a,q)=1}\left|\hat{\psi}(x;q,a)-\frac{x}{\phi(q)}\right| \ll \frac{x(\log x)^4}{e^{\hat{c}\sqrt{\log x}}}+Q\sqrt{x}(\log x)^4$$

which essentially gives us GRH on average. Indeed, the expected value of the error term in the prime number theorem for arithmetic progressions for a given a and q is about size $\sqrt{x/q}$, which would give us $\sqrt{Q}x$ when summing over Q. Sadly, we cannot quite achieve this — even GRH would only give us an error term of size $\sqrt{x}$ for each q and summing this over q gives $Q\sqrt{x}$ as indicated above. This has the impact of restricting the most efficacious use of the theorem to when $Q \le \sqrt{x}$.

One can also consider the square of the error term, and this is much easier to handle. In fact, one can handle other sequences in a similar manner, and this falls under the guise of the Barban-Davenport-Halberstam theorem.

7.1. Review of Gauss sums for Dirichlet characters.

Let χ be a primitive Dirichlet character modulo q. For any integer n we recall the Gauss sum

$$G(n,\chi)=\sum_{b\,(\mathrm{mod}\,q)}\chi(b)e(bn/q).$$

Lemma 7.1. *For primitive χ we have $G(n,\chi)=\bar{\chi}(n)G(1,\chi)$.*

Proof. When $\gcd(n,q)=1$ we have

$$G(n,\chi)=\bar{\chi}(n)\sum_{b\,(\mathrm{mod}\,q)}\chi(bn)e(bn/q)=\bar{\chi}(n)\sum_{c\,(\mathrm{mod}\,q)}\chi(c)e(c/q)=\bar{\chi}(n)G(1,\chi),$$

as the invertibility of n implies that bn runs over all the reduced residues.

When $g=\gcd(n,q)>1$, we wish to show that $G(n,\chi)=0$, and this follows from choosing (by primitivity) some c congruent to 1 mod q/g with $\chi(c)\neq 1$, as

$$\begin{aligned}\chi(c)G(n,\chi)&=\sum_{b\,(\mathrm{mod}\,q)}\chi(cb)e(bn/q)=\sum_{b\,(\mathrm{mod}\,q)}\chi(cb)e\left(\frac{b(n/g)}{(q/g)}\right)\\&=\sum_{b\,(\mathrm{mod}\,q)}\chi(cb)e\left(\frac{bc(n/g)}{(q/g)}\right)=\sum_{b\,(\mathrm{mod}\,q)}\chi(cb)e(bcn/q)=\sum_{d\,(\mathrm{mod}\,q)}\chi(d)e(dn/q)=G(n,\chi).\end{aligned}$$

□

Lemma 7.2. *For primitive χ we have $|G(1,\chi)|=\sqrt{q}$.*

Proof. Writing $\delta(c,d)=1,0$ depending on whether $c=d$, we have that

$$\begin{aligned}\phi(q)|G(1,\chi)|^2&=\sum_{b\,(\mathrm{mod}\,q)}|G(b,\chi)|^2=\sum_{b\,(\mathrm{mod}\,q)}\sum_{c\,(\mathrm{mod}\,q)}\chi(c)e(bc/q)\sum_{d\,(\mathrm{mod}\,q)}\bar{\chi}(d)e(-bd/q)\\&=\sum_{c\,(\mathrm{mod}\,q)}\chi(c)\sum_{d\,(\mathrm{mod}\,q)}\bar{\chi}(d)\sum_{b\,(\mathrm{mod}\,q)}e\big(b(c-d)/q\big)\\&=q\sum_{c\,(\mathrm{mod}\,q)}\chi(c)\sum_{d\,(\mathrm{mod}\,q)}\bar{\chi}(d)\delta(c,d)=q\sum_{c\,(\mathrm{mod}\,q)}|\chi(c)|^2=q\phi(q).\end{aligned}$$

□

Lemma 7.3 (Pólya-Vinogradov)**.** *Suppose that χ is primitive. Then*

$$\Big|\sum_{M<n<N}\chi(n)\Big| \le \sqrt{q}\log q.$$

Proof. We have that

$$G(1,\bar\chi)\sum_{M<n<N}\chi(n) = \sum_{M<n<N}G(n,\bar\chi) = \sum_{M<n<N}\sum_{c\,(\mathrm{mod}\,q)}\bar\chi(c)e(cn/q) = \sum_{c\,(\mathrm{mod}\,q)}\bar\chi(c)\sum_{M<n<N}e(cn/q).$$

So we get

$$\Big|\sum_{M<n<N}\chi(n)\Big| \le \frac{1}{\sqrt{q}}\sum_{c=1}^{q-1}\Big|\sum_{M<n<N}e(cn/q)\Big| \le \frac{1}{\sqrt{q}}\sum_{c=1}^{q-1}\csc(\pi c/q),$$

the last step from summing the geometric series to (essentially) get $\frac{e(cN/q)-e(cM/q)}{1-e(c/q)}$, and then trivially bounding the numerator by 2 and multiplying the denominator by $e(-c/2q)$. Recalling that $\csc(\pi x) \le 1/(2x)$ for $0 < x \le 1/2$, and using the symmetry of the cosecant to handle $m \ge q/2$, for even q we get

$$\sum_{c=1}^{q-1}\csc(\pi c/q) \le 1 + (q/2)\cdot 2\sum_{c=1}^{q/2-1}\frac{1}{c} \le 2 + q\log(q/2-1) \le q\log q$$

and for odd q we have

$$\sum_{c=1}^{q-1}\csc(\pi c/q) \le (q/2)\cdot 2\sum_{c=1}^{(q-1)/2}\frac{1}{c} \le 1 + q\log(q-1) \le q\log q.$$

The result now follows immediately. □

7.2. **Large sieve for multiplicative characters.**

Lemma 7.4. *Let (a_n) be a sequence of complex numbers. Then*

$$\sum_{q\le Q}\frac{q}{\phi(q)}\sideset{}{^\star}\sum_{\chi\,\mathrm{mod}\,q}\Big|\sum_{n\le N}a_n\chi(n)\Big|^2 \le (Q^2+N)\sum_{n\le N}|a_n|^2.$$

We only consider the inner sum over primitive characters; a more general statement is possible.

Proof. We call the sum A and recall the Gauss sum relation $G(n,\chi) = \bar\chi(n)G(1,\chi)$ and $|G(1,\chi)| = \sqrt{q}$ to get

$$\begin{aligned} A = \sum_{q\le Q}\frac{q}{\phi(q)}\sideset{}{^\star}\sum_{\chi\,\mathrm{mod}\,q}\Big|\sum_{n\le N}a_n\bar\chi(n)\Big|^2 &= \sum_{q\le Q}\frac{q}{\phi(q)}\frac{1}{|G(1,\chi)|^2}\sideset{}{^\star}\sum_{\chi\,\mathrm{mod}\,q}\Big|\sum_{n\le N}a_nG(n,\chi)\Big|^2 \\ &= \sum_{q\le Q}\frac{1}{\phi(q)}\sideset{}{^\star}\sum_{\chi\,\mathrm{mod}\,q}\Big|\sum_{b\,(\mathrm{mod}\,q)}\chi(b)S(b/q)\Big|^2 \end{aligned}$$

where

$$S(b/q) = \sum_{n\le N}a_ne(bn/q) \quad\text{and}\quad G(n,\chi) = \sum_{b\,(\mathrm{mod}\,q)}\chi(b)e(bn/q).$$

We next extend the character sum to include imprimitive characters, and expand the square to get an upper bound of

$$\begin{aligned} A &\leq \sum_{q\leq Q} \frac{1}{\phi(q)} \sum_{\chi \bmod q} \sum_{b \,(\mathrm{mod}\, q)} \chi(b) S(b/q) \sum_{c \,(\mathrm{mod}\, q)} \bar{\chi}(c) \bar{S}(c/q) \\ &= \sum_{q\leq Q} \sideset{}{^\star}\sum_{b \,(\mathrm{mod}\, q)} S(b/q) \sideset{}{^\star}\sum_{c \,(\mathrm{mod}\, q)} \bar{S}(c/q) \frac{1}{\phi(q)} \sum_{\chi \bmod q} \bar{\chi}(c)\chi(b) \\ &= \sum_{q\leq Q} \sideset{}{^\star}\sum_{b \,(\mathrm{mod}\, q)} |S(b/q)|^2 \leq (Q^2+N) \sum_{n\leq N} |a_n|^2. \end{aligned}$$

the penultimate step using orthogonality of characters, and then Corollary 6.2. □

7.3. **Combinatorial identities.** Let f be an arithmetic function. We wish to write a sum over primes involving f as a combination of congruence sums (to relatively small moduli) and bilinear terms (which will take care of the larger moduli). Of course, it is easier to use the von Mangoldt function than to sum over primes, and the difference herein can be neglected if our sequence is not too sparse — however, we will see below that this difference between primes and prime powers is already enough to limit the allowable range of moduli in the Bombieri-Vinogradov theorem (at least if we bound things in the most simple manner).

We start by noting that

$$\sum_{n\sim X} \Lambda(n) f(n) = \sum_{n\sim X} f(n) \sum_{c|n} \Lambda(c) \sum_{b|(n/c)} \mu(b),$$

as the inner-most sum is nonzero only when $c = n$, and by inverting the order of summation we get

$$\sum_{n\sim X} \Lambda(n) f(n) = \sum_{n\sim X} f(n) \sum_{b|n} \mu(b) \sum_{c|(n/b)} \Lambda(c).$$

We now split-off the small values of b at some parameter B, and use the summation formula for the von Mangoldt function to compute the c-sum in these parts. We get

$$\sum_{n\sim X} \Lambda(n) f(n) = \sum_{n\sim X} f(n) \sum_{\substack{b\leq B \\ b|n}} \mu(b) \log(n/b) + \sum_{n\sim X} f(n) \sum_{\substack{b|n \\ b>B}} \mu(b) \Bigg(\sum_{\substack{c|(n/b) \\ c\leq B}} + \sum_{\substack{c|(n/b) \\ c>B}} \Bigg) \Lambda(c)$$

We can hopefully estimate the first sum via congruential means due to the smallness of b, and we similarly split off small values of bc in the second sum. In the first part of this second sum, we can move the b-sum to the inside. If we take $B < X$ then we have $c \neq n$, so that the sum over all $b|(n/c)$ is zero; thus we can flip the $b > B$ condition to $b \leq B$ via negating the result. We get

$$\begin{aligned}
\sum_{n\sim X}\Lambda(n)f(n) &= \sum_{n\sim X} f(n)\sum_{\substack{b\le B\\ b|n}}\mu(b)\log(n/b) - \sum_{n\sim X} f(n)\sum_{\substack{c|n\\ c\le B}}\Lambda(c)\sum_{\substack{b|(n/c)\\ b\le B}}\mu(b) + \\
&\quad + \sum_{n\sim X} f(n)\sum_{\substack{b|n\\ b>B}}\mu(b)\sum_{\substack{c|(n/b)\\ c>B}}\Lambda(c) \\
&= \sum_{b\le B}\mu(b)\sum_{u\sim X/b} f(bu)\log u - \sum_{b\le B}\mu(b)\sum_{c\le B}\Lambda(c)\sum_{v\sim X/bc} f(bcv) + \\
&\quad + \sum_{b>B}\mu(b)\sum_{c>B}\Lambda(c)\sum_{v\sim X/bc} f(bcv) \\
&= \sum_{b\le B}\mu(b)\sum_{u\sim X/b} f(bu)\log u - \sum_{b\le B}\mu(b)\sum_{c\le B/b}\Lambda(c)\sum_{v\sim X/bc} f(bcv) - \\
&\quad - \sum_{b\le B}\mu(b)\sum_{B/b<c\le B}\Lambda(c)\sum_{v\sim X/bc} f(bcv) + \sum_{b>B}\mu(b)\sum_{c>B}\Lambda(c)\sum_{v\sim X/bc} f(bcv) \\
&= \sum_{b\le B}\mu(b)\sum_{u\sim X/b} f(bu)\log u - \sum_{b\le B}\mu(b)\sum_{c\le B/b}\Lambda(c)\sum_{v\sim X/bc} f(bcv) - \\
&\quad - \sum_{X/B^2<v<2X/B}\ \sum_{l\sim X/v} f(lv)\eta_l + \sum_{B<b<2X/B}\ \sum_{m\sim X/b} f(bm)\mu(b)\xi_m
\end{aligned}$$

In the third step we split c according to whether $bc \le B$ (desiring $v \gg X/B$), and then wrote the last two sums as bilinear forms with

$$\eta_l = \sum_{\substack{bc=l\\ bc\ge B \text{ and } b,c\le B}} \mu(b)\Lambda(c) \quad \text{and} \quad \xi_m = \sum_{\substack{cv=m\\ c\ge B}}\Lambda(c),$$

so that $|\eta_l| \le \log l$ and $|\xi_m| \le \log m$. In practise, the first two expressions (to the moduli b and bc) can be handled by congruence sums or the like (with $f = \chi$ we will detect cancellation in the innermost sum via Pólya-Vinogradov), while the third and fourth will be viewed as a bilinear form and estimated via some other method (such as the large sieve). This is a rather particular version of Vaughan's identity that will suffice for our applications.

7.4. **Separation of variables.** It is often useful to be able to take summation conditions such as $mn \sim x$ and relax this to $m \sim M$, $n \sim N$ with $MN \sim x$. The following lemma gives a method for doing this, and usually loses only a logarithm. The idea is to transform the summation condition to an expression with $(mn)^{it}$ appearing, and this can then be split into m and n separately. When used in conjunction with the large sieve, the terms like m^{it} will not matter, because we can use arbitrary sequences and changing $a_m \to a_m m^{it}$ will not change the norm of the sequence. As many of our bounds only depend on the coefficients of m and n being small in size, this will not affect the end result.

Lemma 7.5. *Let $x \ge 1$ be given. The function g_x in the theorem below has $\int_{-\infty}^{\infty}|g_x(t)|\,dt < \log 6x$ and for every positive integer l we have*

$$\int_{-\infty}^{\infty} g_x(t) l^{it}\,dt = \begin{cases} 1 & \text{if} \quad l \le x, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Let f be the function given by $f(u) = u$ for $0 \le u \le 1$, and $f(u) \equiv 1$ for $1 \le u \le \lfloor x \rfloor$, and $f(u) = \lfloor x+1 \rfloor - u$ for $\lfloor x \rfloor \le u \le \lfloor x+1 \rfloor$, and $f(u) \equiv 0$ elsewhere. Thus for positive integers l we have $f(l) = 1$ for $l \le x$ and $f(l) = 0$ otherwise. The inverse Mellin transform of f is given by

$$f(u) = \int_{(0)} \frac{h(s)}{u^s} \frac{ds}{2\pi i} = \int_{-\infty}^{\infty} \frac{h(it)}{u^{it}} \frac{dt}{2\pi}$$

where

$$\begin{aligned} h(s) &= \int_0^\infty f(u) u^{s-1}\, du = \int_0^1 u \cdot u^{s-1}\, du + \int_1^{\lfloor x \rfloor} u^{s-1}\, du + \int_{\lfloor x \rfloor}^{\lfloor x+1 \rfloor} \big(\lfloor x+1 \rfloor - u\big) u^{s-1}\, du \\ &= \frac{1}{s+1} + \frac{1}{s}\big(\lfloor x \rfloor^s - 1\big) + \frac{\lfloor x+1 \rfloor}{s}\big(\lfloor x+1 \rfloor^s - \lfloor x \rfloor^s\big) + \frac{1}{s+1}\big(\lfloor x \rfloor^{s+1} - \lfloor x+1 \rfloor^{s+1}\big) \\ &= \left(\frac{1}{s+1} - \frac{1}{s}\right) + \lfloor x \rfloor^s \left(\frac{1}{s} - \frac{\lfloor x \rfloor + 1}{s} + \frac{\lfloor x \rfloor}{s+1}\right) + \lfloor x+1 \rfloor^s \left(\frac{\lfloor x+1 \rfloor}{s} - \frac{\lfloor x+1 \rfloor}{s+1}\right) \\ &= \frac{1}{s(s+1)}\big(1 + \lfloor x \rfloor^{s+1} - \lfloor x+1 \rfloor^{s+1}\big) = \frac{1}{s}\int_0^1 u^s\, du - \frac{1}{s}\int_{\lfloor x \rfloor}^{\lfloor x+1 \rfloor} u^s\, du. \end{aligned}$$

Thus for imaginary $s = it$ the very first expression gives that $|h(it)| \le 1 + \log x$, while the very last gives $|h(it)| \le 2/t$ and the penultimate implies $|h(it)| \le \frac{2(x+1)}{t(t+1)}$. From these we obtain that

$$\begin{aligned} \int_{-\infty}^{\infty} |h(it)|\, dt &\le 2\int_0^1 \big(1 + \log x\big)\, dt + 2\int_1^x \frac{2}{t}\, dt + 2\int_x^\infty \frac{2(x+1)}{t^2}\, dt \\ &\le 2 + 2\log x + 4\log x + \frac{4(x+1)}{x} \le 2\pi \log 6x \end{aligned}$$

The result follows upon taking $g_x(t) = -h(-it)/2\pi$. □

We leave it to the interested reader to show a similar result for summation conditions of the type $m \le n$; here we want $(m/n)^{it}$ to appear and the integral of the absolute value of the g-function will be bounded by $\log 6x^2$. We can also, of course, restrict a summation to a dyadic interval by using the function $g_{2x} - g_x$.

7.5. **Proof of main theorem.** We recall our definition that

$$\hat{\psi}(x; q, a) = \sum_{\substack{p \sim x \\ p \equiv a \pmod q}} \log p.$$

Theorem 7.1. *Suppose that* $2 \le Q \le x/e^{6\hat{c}\sqrt{\log x}}$. *Then we have*

$$\sum_{q \le Q}^{\bullet} \max_{(a,q)=1} \left| \hat{\psi}(x; q, a) - \frac{x}{\phi(q)} \right| \ll \frac{x(\log x)^4}{e^{\hat{c}\sqrt{\log x}}} + Q\sqrt{x}(\log x)^4.$$

Here the sum excludes a possible primitive P-Siegel character for $P = e^{\hat{c}\sqrt{\log x}}$. Alternatively, such a character could be included, but at the cost of replacing the first term with $x/(\log x)^A$ for any $A > 0$ and making the result ineffective. Often the result is presented with a maximum also being taken over $y \le x$; this adds no difficulty to the proof, but this is unnecessary for our application.

The result is mainly useful when the first term dominates, and thus we can't take Q larger than $\sqrt{x}$ in size in applications. The Elliott-Halberstam conjecture implies (essentially) that we can ignore the second term for Q as large as $x^{1-\epsilon}$

for any $\epsilon > 0$. Breaking the $\sqrt{x}$ barrier would be a major result and have many consequences. Through the use of Kloosterman sums, this barrier can be passed when, say, we fix $|a|$ to be small (this is most definitively done in work of Bombieri, Friedlander, and Iwaniec), but the general case is not currently known.

Proof. By orthogonality of characters (when a and q are coprime) we have that

$$\hat{\psi}(x;q,a) = \sum_{\substack{p\sim x \\ p\equiv a\ (q)}} \log p = \frac{1}{\phi(q)} \sum_{\chi \bmod q} \bar{\chi}(a)\hat{\psi}(x,\chi)$$

where $\hat{\psi}(x,\chi) = \sum_{x<p\le 2x} \chi(p)\log p$. It has been noted by some authors that attempts to improve the Bombieri-Vinogradov will likely exploit the variation of $\bar{\chi}(a)$ here. Indeed, we simply bound it in absolute value, which certainly gives a bound for the maximal sum. Subtracting off $x/\phi(q)$ with the principal character, and writing $S(x,Q)$ for the sum in question, we have

$$S(x,Q) \le \sum_{q\le Q}^{\bullet} \frac{1}{\phi(q)} \sum_{\chi \bmod q} \left|\hat{\psi}(x,\chi) - \delta_\chi x\right| = T(x,Q) + U(x,Q),$$

where $\delta_\chi = 1$ when χ is principal and zero otherwise, and the T-U splitting is by whether the character is principal. The contribution from the principal characters is easily bounded via the prime number theorem by

$$T(x,Q) \ll \sum_{q\le Q} \frac{1}{\phi(q)} \cdot \frac{x}{e^{2\hat{c}\sqrt{\log x}}} \ll (\log x) \cdot \frac{x}{e^{2\hat{c}\sqrt{\log x}}}.$$

The q-sum is easily estimated either by elementary techniques or by comparing the residues of the corresponding Dirichlet series and $\zeta(s)$.

For the nonprincipal characters, we group the sum so as to combine the contributions from all characters induced by the same primitive character and get

$$U(x,Q) \ll \sum_{q\le Q}^{\bullet} \sum_{\chi \bmod q}^{\star} |\hat{\psi}(x,\chi)| \sum_{r\le Q/q} \frac{1}{\phi(qr)} \ll \sum_{q\le Q}^{\bullet} \frac{\log Q}{\phi(q)} \sum_{\chi \bmod q}^{\star} |\hat{\psi}(x,\chi)|,$$

where we estimated the r-sum as above (either by elementary means or by a residue comparison). In essence, we took each primitive character modulo q, and then considered the lift to the modulus qr; these have the same error term (we need not fiddle with primes dividing r, as they are less than x).

For small moduli, we will bound the inner sum by using results related to the prime number theorem for arithmetic progressions. From (2), for $q \le P = e^{\hat{c}\sqrt{\log x}}$ we have $|\hat{\psi}(x,\chi)| \ll x/e^{2\hat{c}\sqrt{\log x}}$, though this bound will not be effective unless we exclude a possible P-Siegel character. This gives that

$$U(x,Q) \ll \frac{xP\log Q}{e^{2\hat{c}\sqrt{\log x}}} + (\log x) \sum_{P<q\le Q} \frac{1}{\phi(q)} \sum_{\chi \bmod q}^{\star} |\hat{\psi}(x,\chi)| = U_1(x,Q) + U_2(x,Q).$$

As $U_1(x,Q)$ is already sufficiently small, we need only estimate $U_2(x,Q)$ here, and use the above combinatorial identity to dissect the $\hat{\psi}$-function. We write $\tilde{\psi}(x,\chi)$ for the dyadically-summed χ-twisted von Mangoldt function, and note that we have the bound $\left|\tilde{\psi}(x,\chi) - \hat{\psi}(x,\chi)\right| \ll \sqrt{x}(\log x)^2$, as the summands only differ on prime

powers. From the above version of Vaughan's identity with $f = \chi$ we have

$$\begin{aligned}\tilde{\psi}(x,\chi) &= \sum_{n\sim x}\Lambda(n)\chi(n)\\ &= \sum_{b\le B}\chi(b)\mu(b)\sum_{m\sim x/b}\chi(m)\log(m) - \sum_{b\le B}\mu(b)\sum_{c\le B/b}\Lambda(c)\chi(bc)\sum_{v\sim x/bc}\chi(v) -\\ &\quad - \sum_{x/B^2<v<2x/B}\sum_{l\sim x/v} f(lv)\eta_l + \sum_{B<w<2x/B}\sum_{m\sim x/w} f(bm)\mu(b)\xi_m.\end{aligned}$$

Here we choose $B = e^{3\hat{c}\sqrt{\log x}} \le \sqrt{x/Q}$.

Both of the first two sums can be estimated via the Pólya-Vinogradov inequality. The second is direct and, writing $l = bc$, we crudely estimate the triple sum as

$$\sum_{l\le B}\tau(l)(\log B)\sqrt{q}(\log q) \ll B\sqrt{q}(\log x)^3,$$

where q is the modulus of the character (and is less than x). For the first sum we first use partial summation and get the same result:

$$\sum_{b\le B}\Big|\sum_{m\sim x/b}\chi(m)\int_1^m \frac{dt}{t}\Big| = \sum_{b\le B}\Big|\int_1^{2x/b}\sum_{t\le m\le 2x/b}\chi(m)\frac{dt}{t}\Big| \ll B\sqrt{q}(\log q)(\log x).$$

Putting these back into the expression for $U_2(x,Q)$ and recalling the difference between $\tilde{\psi}$ and $\hat{\psi}$, we get a contribution of no more than

$$\ll (\log x)\sum_{P<q\le Q}\Big(B\sqrt{q}(\log x)^3 + \sqrt{x}(\log x)^2\Big) \ll BQ^{3/2}(\log x)^4 + Q\sqrt{x}(\log x)^3,$$

which is sufficient provided that $B \le \sqrt{x/Q}$.

To estimate the contributions from the other two parts of $\tilde{\psi}(x,\chi)$ we split the outer sums dyadically, which gives an extra factor of $\log x$, and get a bound of

$$\ll (\log x)^2\sum_{P<q\le Q}\frac{1}{\phi(q)}\sideset{}{^\star}\sum_{\chi \bmod q}\Bigg(\Big|\sum_{v\sim V}\sum_{l\sim x/v}\eta_l\chi(lv)\Big| + \Big|\sum_{u\sim U}\sum_{j\sim x/u}\mu(u)\xi_j\chi(ju)\Big|\Bigg),$$

for some V, U with $B \le V, U \le \frac{x}{B}$. We also split the q-sum into dyadic intervals indexed by 2^kP for $k \le \log_2(Q/P)$; this will allow us to induce $q/\phi(q)$ into our sums (as appears in the large sieve inequality) with little loss. We are desirous of using Cauchy's inequality, but first must use the variable-splitting trick of Lemma 7.5 with $h = g_{2x} - g_x$ to get

$$\text{a bound of} \quad \ll (\log x)^2\sum_{k=0}^{\log_2(Q/P)}\int_{-\infty}^{\infty}|h(t)|F(t)\,dt \qquad \text{where } F(t) \text{ is}$$

$$\sum_{q\sim 2^kP}\frac{q/2^kP}{\phi(q)}\sideset{}{^\star}\sum_{\chi \bmod q}\Bigg(\Big|\sum_{v\sim V}\chi(v)v^{it}\sum_{l\sim x/V}\eta_l\chi(l)l^{it}\Big| + \Big|\sum_{u\sim U}\mu(u)\chi(u)u^{it}\sum_{j\sim x/U}\eta_j\chi(j)j^{it}\Big|\Bigg),$$

and the integral of $|h|$ is less than $(\log x)$ in size. Next we apply Cauchy's inequality and the multiplicative version of the large sieve to bound $F(t)$ by

$$\Bigg(\sum_{q\le 2^{k+1}P}\frac{q/2^kP}{\phi(q)}\sideset{}{^\star}\sum_{\chi \bmod q}\Bigg|\sum_{v\sim V}\chi(v)v^{it}\Bigg|^2\Bigg)^{1/2}\Bigg(\sum_{q\le 2^{k+1}P}\frac{q/2^kP}{\phi(q)}\sideset{}{^\star}\sum_{\chi \bmod q}\Bigg|\sum_{l\sim x/V}\eta_l\chi(l)l^{it}\Bigg|^2\Bigg)^{1/2}+$$

$$+\Bigg(\sum_{q\le 2\cdot 2^kP}\frac{q/2^kP}{\phi(q)}\sideset{}{^\star}\sum_{\chi \bmod q}\Bigg|\sum_{u\sim U}\mu(u)\chi(u)u^{it}\Bigg|^2\Bigg)^{1/2}\Bigg(\sum_{q\le 2\cdot 2^kP}\frac{q/2^kP}{\phi(q)}\sideset{}{^\star}\sum_{\chi \bmod q}\Bigg|\sum_{j\sim x/U}\xi_j\chi(j)j^{it}\Bigg|^2\Bigg)^{1/2}$$

$$\ll \frac{1}{2^kP}\Big(V+(2^kP)^2\Big)^{1/2}\Big(\sum_{v\sim V}1^2\Big)^{1/2}\cdot\Big(\frac{x}{V}+(2^kP)^2\Big)^{1/2}\Big(\sum_{l\sim x/V}|\eta_l|^2\Big)^{1/2}+$$

$$+\frac{1}{2^kP}\Big(U+(2^kP)^2\Big)^{1/2}\Big(\sum_{u\sim U}1^2\Big)^{1/2}\cdot\Big(\frac{x}{U}+(2^kP)^2\Big)^{1/2}\Big(\sum_{j\sim x/U}|\xi_j|^2\Big)^{1/2}$$

$$\ll \frac{\sqrt{x}\log x}{2^kP}\Big[x+\Big(\frac{x}{V}+V+\frac{x}{U}+U\Big)(2^kP)^2+(2^kP)^4\Big]^{1/2}$$

$$\ll \frac{\sqrt{x}\log x}{2^kP}\Big[x^{1/2}+\sqrt{\frac{x}{B}}(2^kP)+(2^kP)^2\Big].$$

The last line here uses $B\le U,V\le \frac{x}{B}$. Recalling the sum over k, the integral of $|h|$, and the above parts we estimated via Pólya-Vinogradov, we get a bound of

$$U_2(x,Q)\ll(\log x)^3\cdot\Big(\frac{x\log x}{P}+\frac{x(\log x)^2}{\sqrt{B}}+Q\sqrt{x}(\log x)\Big)+BQ^{3/2}(\log x)^4,$$

which gives the result upon taking $B=e^{3\hat{c}\sqrt{\log x}}\le\sqrt{x/Q}$ and $P=e^{\hat{c}\sqrt{\log x}}$. □

8. Two lemmata

Lemma 8.1. *Suppose that $Z\ge 3$ and $1\le u\le\sqrt{\log Z}$ is an integer. Then*

$$\sum_{m\le Z}\frac{\tau_u(m)}{m}\ll\frac{1}{\sqrt{u}}\frac{(\log Z)^{u+1}}{u!}.$$

Proof. Recalling that $\Gamma(s)$ has e^{-x} as its inverse Mellin transform, we have

$$\sum_{m\le Z}\frac{\tau_u(m)}{m}\le e\sum_{m=1}^{\infty}\frac{\tau_u(m)}{m}e^{-m/Z}=e\sum_{m=1}^{\infty}\frac{\tau_u(m)}{m}\int_{(2)}\frac{\Gamma(s)}{(m/Z)^s}\frac{ds}{2\pi i}$$

$$=\frac{e}{2}\int_{(\delta)}\Gamma(s)\zeta(s+1)^uZ^s\frac{ds}{2\pi i}\ll Z^\delta\zeta(1+\delta)^u\int_{-\infty}^{\infty}|\Gamma(\delta+it)|\,dt$$

Recalling that $\zeta(1+\delta)\le(1+\delta)/\delta$ and taking $\delta=u/\log Z\le 1/\sqrt{\log Z}$, we get

$$\sum_{m\le Z}\frac{\tau_u(m)}{m}\ll\frac{Z^\delta}{\delta}\Big(\frac{1+\delta}{\delta}\Big)^u=\frac{(\log Z)^{u+1}}{u}\Big(\frac{e}{u}\Big)^u\Big(1+\frac{u}{\log Z}\Big)^u\ll\frac{1}{\sqrt{u}}\frac{(\log Z)^{u+1}}{u!},$$

where we noted that the integral is $\ll 1/\delta$ and used that $(e/u)^u\le\sqrt{u}/u!$ and $(1+1/v)^v\le e$ with $v=\sqrt{\log Z}$. □

Lemma 8.2. *Suppose that $Z \geq 3$ and $1 \leq u \leq \sqrt{\log Z}$ is an integer. Then*

$$\sum_{m \leq Z} \tau_u(m) \ll \sqrt{u} Z \frac{(\log Z)^u}{u!}.$$

Proof. We imitate the proof of the previous lemma, getting

$$\sum_{m \leq Z} \tau_u(m) \ll \int_{(1+\delta)} \Gamma(s)\zeta(s)^u Z^s \frac{ds}{2\pi i} \ll Z^{1+\delta}\left(\frac{1+\delta}{\delta}\right)^u \ll \sqrt{u} Z \frac{(\log Z)^u}{u!}.$$

□

9. Recent work on gaps between primes

We will describe recent work of Goldston, Yıldırım, and others that shows that there are small gaps between primes on average. This will use the Bombieri-Vinogradov theorem, some typical techniques from analytic number theory, and consideration of prime k-tuples on average. The "Small gaps between primes exist" preprint of Goldston, Motohashi, Pintz, and Yıldırım shows that

$$\liminf_{n \to \infty} \frac{p_{n+1} - p_n}{\log p_n} = 0,$$

and we include improvements that appear to yield

$$\liminf_{n \to \infty} \frac{p_{n+1} - p_n}{(\log p_n)^{7/9} (\log \log p_n)^{1/9}} < \infty.$$

Here p_n denotes the nth prime. It is claimed that an exponent of $\frac{1}{2}$ on $\log p_n$ can be obtained via similar methods. However, we should stress that the main thrust of the mechanism is not in reducing this exponent, but that an improvement in the Bombieri-Vinogradov theorem would imply the existence of a constant C such that there are infinitely many pairs of primes that differ by no more than C.

We aim to show that, for large N, there are two primes in the interval $[N, 3N]$ that differ by no more than $c(\log N)^{7/9}(\log \log N)^{1/9}$. We first recall that Heath-Brown has shown (effectively) that if there is a P-Siegel character with $P = e^{\sqrt{\log N}}$, then there is a pair of twin primes larger than $\log P = \sqrt{\log N}$. (In fact, his proof is an application of sieve methods, using the idea that $\mu \approx \chi$ for an exceptional character, which gives a sifting function that is essentially periodic.) Thus we can assume that there is no such exceptional character (this only affects effectivity).

9.1. Overview.

The main idea of the proof is in considering the sum

$$Y = \sum_{n \sim N} \left[\Big(\sum_{h \leq H} \hat{\Lambda}(n+h) \Big) - \log 3N \right] \Lambda^R_{\mathcal{H}}(n, k+l)^2.$$

Here $\hat{\Lambda}$ is the von Mangoldt function restricted to primes. If Y is positive, then we must have some integer $n \sim N$ with

$$\left[\Big(\sum_{h \leq H} \hat{\Lambda}(n+h) \Big) - \log 3N \right] > 0,$$

and this means that there is some subinterval of length H in $[N, 2N+H]$ that has at least two primes. Thus we want H to be as small possible.

If the Bombieri-Vinogradov theorem could be extended to handle moduli as large as x^θ for any $\theta > 1/2$, we would get a proof that there are bounded gaps between

primes (we do not explicitly derive this result here, though it follows in the same manner). The argument we give here uses averaging over prime k-tuples, though this would unnecessary with a Bombieri-Vinogradov extension. Indeed, much of our worry is in obtaining results uniform in k, whereas this problem would disappear if we could take $\theta > 1/2$ (we would only need k to be larger than some constant).

Note that we have yet to mention the term $\Lambda^R_{\mathcal{H}}(n, k+l)^2$ in the above. Indeed, its positivity is most important, and other than that, we want to be able to compute asymptotics for the two parts of Y. Here $\mathcal{H}$ is an admissible k-tuple contained in $[1, H]$ and $\Lambda^R_{\mathcal{H}}(n, k+l)^2$ is a truncated version of the generalised von Mangoldt function. It is related to the quasi-optimal Selberg weights for a sieve of dimension $k+l$, so we are detecting k-tuples that have no more than $k+l$ prime factors, and so expect many primes. We will take $l \sim \sqrt{k}$ in the end. The truncation parameter R is related to the sieving limit D of before; we want it as large as possible, but we need (essentially) that $R^2 \ll N^{1/2}$ to be able to apply the Bombieri-Vinogradov theorem.

The crux of the argument is in computing asymptotics for the two parts of Y. The part without $\hat{\Lambda}$ can be analysed with standard tools from analytic number theory (such as contour integration, though here a bivariate version), while the second part can be handled similarly, with the Bombieri-Vinogradov theorem used to control the error term. This can be seen as a version of the so-called weighted sieve, as we allow positive and negative contributions, and determine which dominates.

9.2. **Definitions.** Let $\mathcal{H}$ be a tuple (ordered set) of distinct integers taken from the interval $[1, H]$. For each prime p, let $\Omega_{\mathcal{H}}(p)$ be the set of residue classes modulo p given by the $-h$ for $h \in \mathcal{H}$, and extend $\Omega_{\mathcal{H}}(d)$ multiplicatively to squarefree numbers via the Chinese remainder theorem, so that $n \in \Omega_{\mathcal{H}}(d)$ when $n \in \Omega_{\mathcal{H}}(p)$ for all $p|d$. If we desire, we can invert the notation and note that

$$n \in \Omega_{\mathcal{H}}(d) \iff \quad d|P_{\mathcal{H}}(n) \text{ where } P_{\mathcal{H}}(n) = \prod_{h \in \mathcal{H}} (n+h).$$

We call $\mathcal{H}$ admissible if $\#\Omega_{\mathcal{H}}(p) < p$ for all primes p.

Recalling the quasi-optimal weights from the Selberg sieve, we define (for $a \geq 1$)

$$\lambda^R_a(d) = \frac{\mu(d)}{2\pi i} \int_{(1)} \left(\frac{R}{d}\right)^s \frac{ds}{s^{a+1}} = \begin{cases} \frac{1}{a!}\mu(d)\big(\log R/d\big)^a & \text{if } d \leq R, \\ 0 & \text{if } d > R, \end{cases}$$

where the integral is over the vertical line $\Re s = 1$. Our truncated approximation of a generalisation of the von Mangoldt function is then given by

$$\Lambda^R_{\mathcal{H}}(n, a) = \sum_{n \in \Omega_{\mathcal{H}}(d)} \lambda^R_a(d) = \frac{1}{a!} \sum_{\substack{d|P(n,\mathcal{H}) \\ d \leq R}} \mu(d)\big(\log R/d\big)^a. \tag{12}$$

The truncation parameter R will be chosen as large as possible; it turns out that we need $R^2 \leq Q$ where Q is the maximal modulus allowed in the Bombieri-Vinogradov theorem, so we will have R of size about $N^{1/4}$. We shall also choose $H \leq \log N$, as else the final theorem is of no value in any case. We declare that the final choices of our parameters will be (for some small constant $\dot{c} > 0$)

$$k \sim \dot{c}\frac{(\log N)^{4/9}}{(\log\log N)^{2/9}}, \quad l \sim \sqrt{k}, \quad R \sim N^{1/4}/e^{\sqrt{\log N}},$$

which will give us positivity of Y when $H \gg (\log N)^{7/9}(\log\log N)^{1/9}$. Much of our trouble is in attaining the uniformity in k; derivation of the result with k an arbitrarily large constant is comparatively much easier, and leads to the qualitative result that there are infinitely gaps smaller than any (constant) multiple of $\log N$.

9.3. **First evaluation.** First we will evaluate

$$W(\mathcal{H}) = \sum_{n\sim N} \Lambda_{\mathcal{H}}^R(n, k+l)^2,$$

where $\mathcal{H}$ is an admissible k-tuple. We expand the square in this, and note that with $m = \operatorname{lcm}(d_1, d_2)$ we have that $\Omega_{\mathcal{H}}(d_1) \cap \Omega_{\mathcal{H}}(d_2) = \Omega_{\mathcal{H}}(m)$, so that we get

$$\begin{aligned} W(\mathcal{H}) &= \sum_{d_1} \lambda_{k+l}^R(d_1) \sum_{d_2} \lambda_{k+l}^R(d_2) \sum_{\substack{n\sim N \\ n\in\Omega_{\mathcal{H}}(m)}} 1 \\ &= N \sum_{d_1} \lambda_{k+l}^R(d_1) \sum_{d_2} \lambda_{k+l}^R(d_2) \frac{\#\Omega_{\mathcal{H}}(m)}{m} + O\Big(\frac{(\log R)^{2(k+l)}}{(k+l)!^2} \sum_{d_1\le R} \sum_{d_2 \le R} \#\Omega_{\mathcal{H}}(m) \Big) \end{aligned}$$

since the n-sum is $N/m + O\big(\#\Omega_{\mathcal{H}}(m)\big)$. The number of ways of writing m as the lcm of d_1 and d_2 is bounded by $\tau_3(m)$, as can be seen from considering $m = (a/g)(b/g)g$ where $g = \gcd(a, b)$. Furthermore, since $\#\mathcal{H} = k$ we have that $\#\Omega_{\mathcal{H}}(m) \le \tau_k(m)$. Noting that $\tau_3(m)\tau_k(m) \le \tau_{3k}(m)$, we use Lemma 8.2 bounded the error term as

$$E_1 \ll \sqrt{k} R^2 \frac{(\log R^2)^{3k}}{(3k)!} \cdot \frac{(\log R)^{2(k+l)}}{(k+l)!^2} \ll \sqrt{N} \frac{(\log N)^{6k}}{(3k)!\, e^{ck}}.$$

The fact that R^2 is much smaller than N will imply that this is easily ignorable.

We are left to compute the double sum in the main term, and we note that, via using the integral representation for λ_a^R, this term is given by

$$\hat{W}(\mathcal{H}) = \frac{N}{(2\pi i)^2} \int_{(1)} \int_{(1)} F_{\mathcal{H}}(s_1, s_2) \frac{R^{s_1+s_2}}{(s_1 s_2)^{k+l+1}}\, ds_1\, ds_2,$$

where (multiplicativity and computing the 3 terms with $m = p$ gives the product)

$$F_{\mathcal{H}}(s_1, s_2) = \sum_{d_1} \sum_{d_2} \frac{\mu(d_1)}{d_1^{s_1}} \frac{\mu(d_2)}{d_2^{s_2}} \frac{\#\Omega_{\mathcal{H}}(m)}{m} = \prod_p \Big[1 - \frac{\#\Omega_{\mathcal{H}}(p)}{p} \Big(\frac{1}{p^{s_1}} + \frac{1}{p^{s_2}} - \frac{1}{p^{s_1+s_2}} \Big) \Big].$$

Because we have that $\Omega_{\mathcal{H}}(p) = k$ for $p > H$, we consider the function

$$G_{\mathcal{H}}(s_1, s_2) = F_{\mathcal{H}}(s_1, s_2) \Big(\frac{\zeta(s_1+1)\zeta(s_2+1)}{\zeta(s_1+s_2+1)} \Big)^k,$$

and below show that this is regular somewhat to the left of the imaginary axes. Note that we have that $G_{\mathcal{H}}(0, 0)$ is equal to the singular series for $\mathcal{H}$, as we have

$$G_{\mathcal{H}}(0, 0) = S(\mathcal{H}) = \prod_p \Big(1 - \frac{\#\Omega_{\mathcal{H}}(p)}{p} \Big) \Big(1 - \frac{1}{p} \Big)^{-k}.$$

The method to get an asymptotic for $\hat{W}(\mathcal{H})$ is sufficiently standard that here we just state the result, and postpone the details until later. We get that

$$\hat{W}(\mathcal{H}) = N \cdot S(\mathcal{H}) \cdot \binom{2l}{l} \frac{(\log R)^{k+2l}}{(k+2l)!} \Big[1 + O\Big(\frac{k^2\sqrt{\log k}}{\log N} \Big)^2 \Big].$$

Here the pole-order $k+2l$ comes from $2(k+l)-k$, in which the two $(k+l)$'s come from $1/s_1^{k+l+1}$ and $1/s_2^{k+l+1}$ and the subtracted term comes from the ζ-quotient in $G_{\mathcal{H}}(s_1,s_2)$. The factor of $\binom{2l}{l}$ appears from an integral like $\int(\xi+1)^{2l}/\xi^{l+1}\,d\xi$ around a small circle; this integrand comes about by multiplying the ζ-factor by $s_1s_2/(s_1+s_2)$ and changing variables. It is not entirely ridiculous to try to improve the error term, and this would then affect the final result. Oppositely, an asymptotic with the bracketed term as $\big[1+O\big(\frac{e^{ck}}{\log N}\big)\big]$ can be derived more easily. Finally, note that the bound on our previous error term E_1 fits in this error term, due to the fact (see below) that $S(\mathcal{H})\gg 1/e^{ck}$ for admissible $\mathcal{H}$.

9.4. **Second evaluation.** We now evaluate the sum

$$I_{\mathcal{H}}(h)=\sum_{n\sim N}\hat{\Lambda}(n+h)\Lambda_{k+l}^R(n,\mathcal{H})^2$$

first assuming that $h\notin\mathcal{H}$ and that $\mathcal{H}\cup\{h\}$ is admissible. Expanding, we get that

$$\begin{aligned}I_{\mathcal{H}}(h)&=\sum_{n\sim N}\tilde{\Lambda}(n+h)\sum_{\substack{d_1|P_{\mathcal{H}}(n)\\ d_1\le R}}\lambda_{k+l}^R(d_1)\sum_{\substack{d_2|P_{\mathcal{H}}(n)\\ d_2\le R}}\lambda_{k+l}^R(d_2)\\ &=\sum_{d_1\le R}\lambda_{k+l}^R(d_1)\sum_{d_2\le R}\lambda_{k+l}^R(d_2)\sum_{c\in\Omega_{\mathcal{H}}(m)}\sum_{\substack{n\sim N\\ n\equiv c \bmod m}}\tilde{\Lambda}(n+h)\\ &=\sum_{d_1\le R}\lambda_{k+l}^R(d_1)\sum_{d_2\le R}\lambda_{k+l}^R(d_2)\sum_{c\in\Omega_{\mathcal{H}}(m)}\hat{\psi}(N;m,c+h),\end{aligned}$$

where m is the lcm of d_1 and d_2. We now replace $\hat{\psi}$ by its asymptotic evaluation $N/\phi(m)$, calling the error term $\mathcal{E}$. We postpone the main term for below, and note that the error is

$$\begin{aligned}E_2&\ll\sum_{d_1\le R}\frac{(\log R)^{k+l}}{(k+l)!}\sum_{d_2\le R}\frac{(\log R)^{k+l}}{(k+l)!}\sum_{c\in\Omega_{\mathcal{H}}(m)}\mathcal{E}(N;m,c+h)\\ &\ll\frac{(\log R)^{2(k+l)}}{(k+l)!^2}\sum_{m\le R^2}\tau_3(m)\cdot\#\Omega_{\mathcal{H}}(m)\cdot\max_r\mathcal{E}(N;m,r).\end{aligned}$$

Here we used that the number of ways of writing an lcm is τ_3. We recall that $\#\Omega_{\mathcal{H}}(m)\le\tau_k(m)$ and $\tau_3\tau_k\le\tau_{3k}$, and next we split the sum according to whether $\tau_{3k}(m)\le(\log N)^A$ for some parameter A (a variant of Rankin's trick). Upon using the trivial bound that $\mathcal{E}\le N/m$ for large $\tau_{3k}(m)$ and Bombieri-Vinogradov when $\tau_{3k}(m)$ is small, this gives (for any $\alpha>0$) that

$$\begin{aligned}E_2&\ll\frac{(\log R)^{2(k+l)+A}}{(k+l)!^2}\sum_{m\le R^2}\max_r\mathcal{E}(N;m,r)+\sum_{m\le R^2}\tau_{3k}(m)\bigg(\frac{\tau_{3k}(m)}{(\log N)^A}\bigg)^{\alpha}\frac{N}{m}\\ &\ll\frac{(\log R)^{2(k+l)+A}}{(k+l)!^2}\bigg(R^2\sqrt{N}(\log N)^4+\frac{N}{e^{\hat{c}\sqrt{\log N}}}\bigg)+\frac{N}{(\log N)^{A/9}}\sum_{m\le R^2}\frac{\tau_{3k}(m)^{10/9}}{m}\\ &\ll\frac{(\log R)^{2(k+l)+A}}{(k+l)!^2}\frac{N}{e^{\hat{c}\sqrt{\log N}}}+\frac{N(\log N)^{4k^{10/9}-A/9}}{k^{10/9}!}\ll\frac{N}{(2k)!\,e^{ck}}.\end{aligned}$$

Here we chose $\alpha=1/9$ and $A=37k^{10/9}$, and then used that $k\le(\log N)^{4/9}$ and $(10/9)(4/9)<1/2$, while noting that $\tau_{3k}^{10/9}\le\tau_{4k^{10/9}}$ before using Lemma 8.1.

9.4.1. *Main term.* The main term in $I_{\mathcal{H}}(h)$ is given by

$$N \sum_{d_1 \le R} \lambda^R_{k+l}(d_1) \sum_{d_2 \le R} \lambda^R_{k+l}(d_2) \cdot \frac{1}{\phi(m)} \sum_{c \in \Omega_{\mathcal{H}}(m)} \delta\big(\gcd(c+h,p)\big).$$

Here we have $\delta(x) = 1$ for $x = 1$ and $\delta(x) = 0$ else. The inner sum is

$$\sum_{c \in \Omega_{\mathcal{H}}(m)} \delta\big(\gcd(c+h,m)\big) = \prod_{p|m} \Big[\sum_{c \in \Omega_{\mathcal{H}}(p)} \delta\big(\gcd(c+h,m)\big) \Big] = \prod_{p|m} \big(\#\Omega_{\mathcal{H}^h}(p) - 1\big).$$

Here $\mathcal{H}^h$ denotes the union of $\mathcal{H}$ with $\{h\}$, and we noted that the gcd is 1 except for $c = -h$, which occurs in the c-sum exactly when $h \in \mathcal{H}$.

The argument is now exactly analogous to the previous. Noting that $\phi(p) = p-1$, we want to estimate the integral

$$\hat{I}_{\mathcal{H}}(h) = \frac{N}{(2\pi i)^2} \int\int \prod_p \Big[1 - \frac{\#\Omega_{\mathcal{H}^h}(p) - 1}{p-1} \Big(\frac{1}{p^{s_1}} + \frac{1}{p^{s_2}} - \frac{1}{p^{s_1+s_2}} \Big) \Big] \frac{R^{s_1+s_2}\, ds_1\, ds_2}{(s_1 s_2)^{k+l+1}},$$

and we multiply/divide by the kth power of a ζ-quotient as before. For the evaluation at $(0,0)$ we get

$$\prod_p \Big(1 - \frac{\#\Omega_{\mathcal{H}^h}(p) - 1}{p-1} \Big) \Big(1 - \frac{1}{p} \Big)^{-k} = \prod_p \Big(\frac{p - 1 - \#\Omega_{\mathcal{H}^h}(p) + 1}{p-1} \Big) \Big(\frac{p-1}{p} \Big)^{-k}$$

$$= \prod_p \Big(\frac{p - \#\Omega_{\mathcal{H}^h}(p)}{p} \Big) \Big(\frac{p}{p-1} \Big) \Big(\frac{p-1}{p} \Big)^{-k} = \prod_p \Big(1 - \frac{\#\Omega_{\mathcal{H}^h}(p)}{p} \Big) \Big(\frac{p-1}{p} \Big)^{-(k+1)},$$

and this is exactly $S(\mathcal{H}^h)$.

So assuming that $h \notin \mathcal{H}$ and $\mathcal{H}^h$ is admissible, we get that

$$\hat{I}_{\mathcal{H}}(h) = N \cdot S(\mathcal{H} \cup \{h\}) \cdot \binom{2l}{l} \frac{(\log R)^{k+2l}}{(k+2l)!} \Big[1 + O\Big(\frac{k^2 \sqrt{\log k}}{\log N} \Big)^2 \Big], \quad h \notin \mathcal{H}.$$

Next we note that when $h \in \mathcal{H}$ we have that $\hat{\Lambda}(n+h)\Lambda^R_{\mathcal{H}}(n,k+l)^2$ is nonzero only when $(n+h)$ is prime, and in this case, writing $\mathcal{H}_h = \mathcal{H} \backslash \{h\}$, since $R < N$ we have that $d | P_{\mathcal{H}}(n) \iff d | P_{\mathcal{H}_h}(n)$ for $d \le R$, so that by (12) we have

$$\hat{\Lambda}(n+h)\Lambda^R_{\mathcal{H}}(n,k+l)^2 = \hat{\Lambda}(n+h)\Lambda^R_{\mathcal{H}_h}(n,k+l)^2$$

We can then apply the above to $\mathcal{H}_h$ upon making the change $k \to k-1$ and $l \to l+1$. So when $h \in \mathcal{H}$ (and $\mathcal{H}$ is admissible) we get that

$$\hat{I}_{\mathcal{H}}(h) = N \cdot S(\mathcal{H}) \cdot \binom{2l+2}{l+1} \frac{(\log R)^{k+2l+1}}{(k+2l+1)!} \Big[1 + O\Big(\frac{k^2 \sqrt{\log k}}{\log N} \Big)^2 \Big], \quad h \in \mathcal{H}.$$

Finally, we note that the error term E_2 from the Bombieri-Vinogradov theorem is much smaller than the error term here.

9.5. **Comparison of terms.** Let $T_k(H)$ be the set of all distinct k-tuples of integers taken from $[1,H]$. We wish to estimate

$$Y = \sideset{}{^\star}\sum_{\mathcal{H} \in T_k(H)} \sum_{n \sim N} \Big[\Big(\sideset{}{^\star}\sum_{h \le H} \hat{\Lambda}(n+h) \Big) - \log 3N \Big] \Lambda^R_{\mathcal{H}}(n,k+l)^2,$$

where the star on the h-sum restricts to h such that $\mathcal{H}^h$ is admissible, and the outer star similarly insists that $\mathcal{H}$ be admissible. We will then choose H so that

this is positive. This will then give two primes in some subinterval of length H in $[N, 2N+H]$. For the term involving $\hat{\Lambda}$, we split the h-sum depending on whether we have $h \in \mathcal{H}$. Factoring out the common factor of $N(\log R)^{k+2l}\binom{2l}{l}/(k+2l)!$, we get asymptotics of

$$Y_+ = \sum_{\mathcal{H}\in T_k(H)} \sideset{}{^\star}\sum_{\substack{h\le H\\ h\notin\mathcal{H}}} S(\mathcal{H}\cup\{h\}) + \left(\frac{\log R}{k+2l+1}\right)\left(\frac{(2l+2)(2l+1)}{(l+1)(l+1)}\right) \sideset{}{^\star}\sum_{\mathcal{H}\in T_k(H)} \sum_{\substack{h\le H\\ h\in\mathcal{H}}} S(\mathcal{H})$$

from the terms with $\bar{\Lambda}$, and

$$Y_- = \log 3N \sideset{}{^\star}\sum_{\mathcal{H}\in T_k(H)} S(\mathcal{H})$$

for the other term. The error terms give

$$Y_E \ll (Y_+ + Y_-)\left(\frac{k^2\sqrt{\log k}}{\log N}\right)^2.$$

We write

$$U_k(H) = \sum_{\mathcal{H}\in T_k(H)} S(\mathcal{H}),$$

(where it does not matter whether the sum is over admissible $\mathcal{H}$) and see that the second double sum in the above display is $kU_k(H)$, as the h-sum has k members each of which repeats the $\mathcal{H}$-sum. Meanwhile, the first double sum in Y_+ is $U_{k+1}(H)$; indeed, ignoring admissibility as we may, there is one-to-one map between $(\mathcal{H}, h)$ and $\mathcal{H}\cup\{h\}$, where this union puts h at the end of the tuple, and the latter enumerates $T_{k+1}(H)$.

We now use that the l-quotient $2\big(\frac{2l+1}{l+1}\big)$ is $\big(4+O(1/l)\big)$ combined with the estimate $(k+2l+1) = k\big(1+O(l/k)\big)$ and the choice $l \sim \sqrt{k}$ to get

$$\begin{aligned} Y_+ - Y_- - Y_E = U_{k+1}(H) + U_k(H)\Big[\log R\cdot\Big(4+O\Big(\frac{1}{\sqrt{k}}\Big)\Big) - \log 3N\Big] \\ - O\Big[\Big(\frac{k^2\sqrt{\log k}}{\log N}\Big)^2\big[U_{k+1}(H) + U_k(H)\log N\big]\Big] \end{aligned}$$

Recalling that $R = N^{1/4}/e^{\sqrt{\log N}}$, we get that $\hat{Y} = Y_+ - Y_- - Y_E$ is estimated by

$$\hat{Y} = U_k(H)\Big[\frac{U_k(H+1)}{U_k(H)}\Big(1 - O\Big(\frac{k^4\log k}{(\log N)^2}\Big)\Big) - O\Big(\frac{\log N}{\sqrt{k}} + \sqrt{\log N} + \frac{k^4\log k}{\log N}\Big)\Big]$$

Finally we use a fact (14) about $U_k(H)$ that we prove in Lemma 9.1 below, namely

$$\frac{U_{k+1}(H)}{U_k(H)} \ge H\cdot\Big[1 - O\Big(\frac{k\log H}{H}\Big)\Big],$$

at least when $k\log H \ll H$. This gives us that

$$\hat{Y} \ge U_k(H)\Big[H - O\Big(\frac{\log N}{\sqrt{k}} + \sqrt{\log N} + \frac{k^4\log k}{\log N} + k\log H\Big)\Big].$$

Our choice of $k \sim \dot{c}(\log N)^{4/9}/(\log\log N)^{2/9}$ implies that there is some constant $c > 0$ such that this is positive when $H \ge c(\log N)^{7/9}(\log\log N)^{1/9}$. This gives the desired result of Goldston and Yıldırım about small gaps between primes. We now turn to proving the technical lemmata about $G_{\mathcal{H}}(s_1, s_2)$ and $U_k(H)$, again noting

that if R can be taken with $\frac{\log R}{\log N} > 1/4 + \epsilon$ for some $\epsilon > 0$, then we would get bounded gaps.

9.6. **The singular series on average.** We wish to estimate

$$U_k(H) = \sum_{\mathcal{H} \in T_k(H)} S(\mathcal{H})$$

under a mild assumption such as $H \gg k(\log H)$. We let $z = \max(k^{100}, H^{10})$ and let $Z = e^{100z}$ (or take the limit as $Z \to \infty$). The parameter z is a sieving limit for which we detect quasi-primes, while the parameter Z introduces an averaging over an immense number of shifts of an interval of length H.

We take Q_z to be the set of integers which are coprime to all primes less than z. For each integer i from 1 to Z we define

$$f_i = \sum_{\substack{i+1 \le m \le i+H \\ m \in Q_z}} 1.$$

This counts z-quasiprimes in the interval $[i+1, i+H]$. We let $a_i(k) = \binom{f_i}{k} \cdot k!$ be the number of k-tuples of distinct z-quasiprimes from this interval.

We can note that $a_i(k+1) = a_i(k)(f_i - k)$ by binomial coefficients, and so

$$\sum_{i=1}^{Z} a_i(k+1) = \sum_{i=1}^{Z} f_i \cdot a_i(k) - k \sum_{i=1}^{Z} a_i(k).$$

We note that $(f_i - f_j)$ and $\big(a_i(k) - a_j(k)\big)$ have the same sign for any i, j, k, and thus we get that

$$0 \le \sum_{i=1}^{Z} \sum_{j=1}^{Z} (f_i - f_j)\big(a_i(k) - a_j(k)\big) = 2\Bigg(Z \sum_{i=1}^{Z} f_i \cdot a_i(k) - \sum_{i=1}^{Z} f_i \sum_{i=1}^{Z} a_i(k) \Bigg).$$

Replacing the right side of the display previous, we get that

$$\sum_{i=1}^{Z} a_i(k+1) \ge \frac{1}{Z} \sum_{i=1}^{Z} f_i \cdot \sum_{i=1}^{Z} a_i(k) - k \sum_{i=1}^{Z} a_i(k). \tag{13}$$

We use a familiar averaging technique to exploit the regularity in the f_i-sum to get

$$\sum_{i=1}^{Z} f_i = \sum_{i=1}^{Z} \sum_{\substack{i+1 \le m \le i+H \\ m \in Q_z}} 1 = \sum_{\substack{m=H+1 \\ m \in Q_z}}^{Z-H} H + \sum_{\substack{m=1 \\ m \in Q_z}}^{H} m + \sum_{\substack{m=Z-H+1 \\ m \in Q_z}}^{Z} (Z+1-m) = H \sum_{\substack{m=1 \\ m \in Q_z}}^{Z} 1 + O(H^2).$$

Writing $P(z) = \prod_{p \le z} p$ and $V_z = \prod_{p \le z} (1 - 1/p)^{-1} \sim e^{\gamma} \log z$, we have that the sum here is $Z/V_z + O\big(P(z)\big)$, and we note that $P(z) \le e^{3z/2} = O(Z^{1/50})$. Also, since $H \le z^{1/10}$ we have that $H^2 = O\big((\log Z)^{1/5}\big)$.

Thus we are left to estimate the sum $\sum_i a_i(k)$. The main fact we use is that

$$a_n(k) = \sum_{\substack{\mathcal{H} \in T_k(H) \\ P_{\mathcal{H}}(n) \in Q_z}} 1.$$

Indeed, there are f_n integers u in $[1, H]$ with $(n+u)$ coprime to $P(z)$, and the sum in the above display counts the the number of ways of choosing k of these (including

order), which gives $\binom{f_n}{k} \cdot k! = a_n(k)$ as desired. Now we sum this over n and switch the order of summation to get

$$\sum_{n=1}^{Z} a_n(k) = \sum_{n=1}^{Z} \sum_{\substack{\mathcal{H} \in T_k(H) \\ P_{\mathcal{H}}(n) \in Q_z}} 1 = \sum_{\mathcal{H} \in T_k(H)} \sum_{\substack{n=1 \\ P_{\mathcal{H}}(n) \in Q_z}}^{Z} 1 = \sum_{\mathcal{H} \in T_k(H)} R(\mathcal{H}).$$

We now proceed to estimate the inner sum $R(\mathcal{H})$. We have that

$$R(\mathcal{H}) = \sum_{\substack{n=1 \\ P_{\mathcal{H}}(n) \in Q_z}}^{Z} 1 = \sum_{n=1}^{Z} \sum_{\substack{d \mid P_{\mathcal{H}}(n) \\ d \mid P(z)}} \mu(d) = \sum_{d \mid P(z)} \mu(d) \sum_{\substack{n=1 \\ d \mid P_{\mathcal{H}}(n)}}^{Z} 1.$$

The inner sum is quite regular, and we get that

$$R(\mathcal{H}) = \sum_{d \mid P(z)} \mu(d) \#\Omega_{\mathcal{H}}(d) \left(\frac{Z}{d} + O(1) \right) = Z \prod_{p \le z} \left(1 - \frac{\#\Omega_{\mathcal{H}}(p)}{p} \right) + O\left(\prod_{p \le z} (1+k) \right).$$

This gives

$$R(\mathcal{H}) = \frac{Z}{V_z^k} \prod_{p \le z} \left(1 - \frac{\#\Omega_{\mathcal{H}}(p)}{p} \right) \left(1 - \frac{1}{p} \right)^{-k} + O\big((k+1)^{3z/2 \log z} \big).$$

Since we have $k \le z$ the error is $O(e^{3z/2}) = O(Z^{1/50})$, so we get

$$R(\mathcal{H}) = S(\mathcal{H}) \frac{Z}{V_z^k} \exp\left[O\left(\sum_{p > z} \frac{k^2}{p^2} \right) \right] + O(Z^{1/50}).$$

The exponential term here is $1 + O(k^2/z \log z)$. Plugging back into (13), we get that

$$\sum_{H \in T_{k+1}(\mathcal{H})} R(\mathcal{H}) \ge \frac{1}{Z} \left(\frac{HZ}{V_z} + O\big(Z^{1/5}\big) \right) \cdot \sum_{H \in T_k(\mathcal{H})} R(\mathcal{H}) - k \sum_{H \in T_k(\mathcal{H})} R(\mathcal{H}).$$

Evaluating $R(\mathcal{H})$, we get

$$\frac{Z}{V_z^{k+1}} \sum_{H \in T_{k+1}(\mathcal{H})} S(\mathcal{H}) \ge \left(\frac{H}{V_z} - 2k \right) \cdot \left[1 - O\left(\frac{k^2}{z \log z} \right) \right] \cdot \frac{Z}{V_z^k} \sum_{H \in T_k(\mathcal{H})} S(\mathcal{H}).$$

Summing and removing common factors gives us that

$$U_{k+1}(H) \ge U_k(H) \cdot (H - 10k \log z) \cdot \left[1 - O\left(\frac{k^2}{z \log z} \right) \right].$$

Finally, we use that $z = \max(k^{100}, H^{10})$ and $H \gg k(\log H)$ and get the following:

Lemma 9.1. *Suppose that $H \gg k(\log H)$. Then*

$$U_{k+1}(H) \ge H \cdot U_k(H) \cdot \left[1 + O\left(\frac{k \log H}{H} \right) \right]. \tag{14}$$

9.7. **Contour integrals.** First we outline the method we use. We will consider

$$T(\alpha) = \frac{N}{(2\pi i)^2}\int_{(1)}\int_{(1)} F_\alpha(s_1,s_2)\frac{R^{s_1+s_2}}{(s_1s_2)^{k+l+1}}\,ds_1\,ds_2$$
$$= \frac{N}{(2\pi i)^2}\int_{(1)}\int_{(1)} G_\alpha(s_1,s_2)\left(\frac{\zeta(s_1+s_2+1)}{\zeta(s_1+1)\zeta(s_2+1)}\right)^k\frac{R^{s_1+s_2}}{(s_1s_2)^{k+l+1}}\,ds_1\,ds_2,$$

where

$$F_\alpha(s_1,s_2) = \prod_p\left[1-\alpha_p\left(\frac{1}{p^{s_1}}+\frac{1}{p^{s_2}}-\frac{1}{p^{s_1+s_2}}\right)\right].$$

In our applications we will have that α_p is either $\frac{\#\Omega_{\mathcal{H}}(p)}{p}$ or $\frac{\#\Omega_{\mathcal{H}^h}(p)-1}{p-1}$, but here we shall only assume that $0\le\alpha_p\le 1-\frac{1}{p}$ for all p and $\alpha_p=\frac{k}{p}+O\big(\frac{k}{p^2}\big)$ for $p>H$. To evaluate $T(\alpha)$ we shall move the contours close (about $1/\sqrt{\log N}$ away) to the imaginary axes, and truncate them at a high height (this will be $J=e^{\sqrt{\log N}}$). The largeness of the denominator $(s_1s_2)^{k+l+1}$ will ensure that the truncations give negligible impact. Then we move the inner contour across the imaginary axis (the same distance away). The new integral is small because of the $1/R^{1/\sqrt{\log N}}$ term, the horizontal bits are negligible as before, and we get residues from poles at $s_1=0$ and $s_1=-s_2$, while the zero-free region for the ζ-function excludes other poles. The residue at $s_1+s_2=0$ is small mainly because $R^{s_1+s_2}\ll 1$ with everything else controlled, though an accounting of logarithms is necessary to ensure this. By moving the other contour similarly across the imaginary axis, the resulting integral is again small, and we get a residue at $s_2=0$. Thus the above expression for $T(\alpha)$ is essentially given from the residue at $s_1=s_2=0$. This double residue is then expanded in terms of G_α (actually a slight normalisation of it, which eases the calculation) and its logarithmic derivatives, for which we need good bounds, as we want the leading $G_\alpha(0,0)$ term to dominate.

9.7.1. *Bounds for G_α.* Recall the definition of $G_\alpha=G_\alpha(s_1,s_2)$ as an Euler product:

$$G_\alpha=\prod_p\left[1-\alpha_p\left(\frac{1}{p^{s_1}}+\frac{1}{p^{s_2}}-\frac{1}{p^{s_1+s_2}}\right)\right]\left(1-\frac{1/p}{p^{s_1}}\right)^{-k}\left(1-\frac{1/p}{p^{s_2}}\right)^{-k}\left(1-\frac{1/p}{p^{s_1+s_2}}\right)^{k}.$$

The first part of this is $F_\alpha(s_1,s_2)$, while the second is a ζ-quotient.

We write $\eta=\max(-\Re s_1,0)+\max(-\Re s_2,0)$, always taking $\eta\le 1/\log H$ for convenience. We first consider primes $p\ge H$, for which we have $\alpha_p=\frac{k}{p}+O\big(\frac{k}{p^2}\big)$. Their contribution to $\log G_\alpha$ is

$$\sum_{p\ge H}\left[-\left(\frac{k/p}{p^{s_1}}+\frac{k/p}{p^{s_2}}-\frac{k/p}{p^{s_1+s_2}}\right)+O\left(\frac{k^2p^{2\eta}}{p^2}\right)\right]+\left[\left(\frac{k/p}{p^{s_1}}+\frac{k/p}{p^{s_2}}+\frac{-k/p}{p^{s_1+s_2}}\right)+O\left(\frac{kp^{2\eta}}{p^2}\right)\right]$$

which gives a contribution B_1 bounded as

$$B_1\ll k^2H^{2\eta-1}\ll\frac{k^2}{H}\ll\frac{k}{\log H}.$$

We introduce the parameter $U=(6k)^{1/(1-2\eta)}\le(6k)^{1+3\eta}\le 12k$, and for primes with $U\le p\le H$ we use the fact that $3kp^{2\eta}/p\le 1/2$ to bound the contribution B_2 to $\log G_\alpha$ by

$$B_2\ll\sum_{U\le p\le H}\frac{3kp^{2\eta}}{p}\ll\frac{kH^{2\eta}}{\log H}\ll\frac{k}{\log H}.$$

For $p \le U$ we use the trivial upper-bound of $\log(1+3kp^{2\eta-1}) - 3k\log(1-p^{2\eta}/p)$, which is bounded in size by $\big(\log k + kp^{2\eta-1}\big)$. The p-sum of the second term is bounded in size by $kH^{2\eta}$ and the first by k so that, upon adding the above estimates for B_1 and B_2, we get that

$$\log|G_\alpha(s_1,s_2)| \ll k \quad \text{and so} \quad |G_\alpha(s_1,s_2)| \ll e^{ck} \quad \text{when} \quad \eta \le 1/\log H. \tag{15}$$

This is an upper bound for G_α; however, we could have $G_\alpha(s_1,s_2)=0$.

9.7.2. *Moving contours.* Now we return to the double integral representation

$$T(\alpha) = \frac{N}{(2\pi i)^2}\int_{L_2}\int_{L_1} G_\alpha(s_1,s_2)\bigg(\frac{\zeta(s_1+s_2+1)}{\zeta(s_1+1)\zeta(s_2+1)}\bigg)^k \frac{R^{s_1+s_2}}{(s_1s_2)^{k+l+1}}\,ds_1\,ds_2,$$

where we moved the contours to L_1 given by $\frac{1}{\log J}+it$ and L_2 given by $\frac{1}{2\log J}+it$ where $J=e^{\sqrt{\log N}}$ and we truncate the contours to $|t|\le J$. This truncation gives an error of no more than

$$E_3 \ll N\cdot\frac{e^{ck}(\log J)^{3k}R^{3/\log J}}{J^k} \ll N^{1+1/\sqrt{\log N}}\cdot\frac{(\log J)^{3k}}{J^k} \ll \frac{N}{(2k)!\,e^{ck}},$$

in which the denominator swamps the other contributions, the last step using (a ridiculously crude bound) that $J \ge k(\log k)^4$. Here we used that ζ and its reciprocal are both bounded in size by $\log J$ on the contours.

Now we move the contour L_1 to L_1^- which is given by $\frac{-1}{\log J}+it$, where again the contribution from the horizontal segments is easy to estimate. This gives us residues from the poles at $s_1=0$ and $s_1=-s_2$, and the zero-free region for the zeta-function implies (for large N) that there are no other poles (there are better zero-free regions, but we do not require them). The integral over L_1^- (and then over L_2) is small due to the dominant $R^{-1/\log J}$ factor, and indeed we can bound its contribution to $T(\alpha)$ as

$$E_4 \ll N\cdot e^{ck}(\log J)^{3k}R^{-1/2\log J} \ll N\cdot\frac{(c\log J)^{3k}}{e^{0.1\sqrt{\log N}}} \ll \frac{N}{(2k)!\,e^{ck}}.$$

Here we noted that ζ and its reciprocal are bounded by $\log J$ on the contour, both near the origin and with t near J; the final estimate again comes from comparing $k\log\log J + k\log k$ to $\sqrt{\log N}$, where the latter is much bigger since $k \le (\log N)^{4/9}$.

We are left with the estimation of

$$\tilde{T}(\alpha) = \frac{N}{2\pi i}\int_{L_2}\Big(\operatorname*{res}_{s_1=0} + \operatorname*{res}_{s_1=-s_2}\Big)\,ds_2.$$

We next note that the contribution from the residue at $s_1=-s_2$ is small. For this, we make a circle $C(s_2)$ of size $\frac{1}{9\log J}$ about s_2, noting that

$$\operatorname*{res}_{s_1=-s_2} = \frac{1}{2\pi i}\int_{C(s_2)} G_\alpha(s_1,s_2)\bigg(\frac{\zeta(s_1+s_2+1)}{\zeta(s_1+1)\zeta(s_2+1)}\bigg)^k \frac{R^{s_1+s_2}}{(s_1s_2)^{k+l+1}}\,ds_1,$$

Since we have the crude bound $\frac{1}{9\log J} \le \frac{1}{\log H}$, on $C(s_2)$ we can use the previous bound (15) of $|G_\alpha| \ll e^{ck}$, while we have $\big|\zeta(s_1+s_2+1)\big| \ll \log J$. Noting the

bounds $|s_1| \ll |s_2| \ll |s_1|$ and that $\big|1/s\zeta(s+1)\big| \ll 1$ for our contours, we get

$$\frac{N}{2\pi i}\int_{L_2} \operatorname*{res}_{s_1=-s_2} ds_2 \ll N\int_{L_2} e^{ck}(\log J)^k \frac{ds_2}{|s_2|^{2l+2}} \ll N\cdot(c\log J)^{k+2l+2}$$
$$\ll N(c\log N)^{k/2+l+1} \ll \frac{N(\log N)^k}{(k+2l+1)!\,e^{ck}},$$

where the last step as usual follows from making the comparison of logarithms: $(k/2)\log\log N + k\log k \le \big(17k/18\big)\log\log N \le k\log\log N$, recalling $k \le (\log N)^{4/9}$.

9.7.3. *The double residue.* Finally we want to estimate

$$\hat{T}(\alpha) = \frac{N}{2\pi i}\int_{L_2} \operatorname*{res}_{s_1=0} ds_2,$$

and for this we write

$$Z_\alpha(s_1,s_2) = G_\alpha(s_1,s_2)\left(\frac{(s_1+s_2)\zeta(s_1+s_2+1)}{s_1\zeta(s_1+1)s_2\zeta(s_2+1)}\right)^k = F_\alpha(s_1,s_2)\left(\frac{s_1+s_2}{s_1s_2}\right)^k, \tag{16}$$

which is regular (and nonzero) about $(0,0)$. This function has the advantage that the poles come from powers of s_1 and s_2 rather than from ζ-functions, and this will allow a more direct analysis. We put this into the integral to get

$$\hat{T}(\alpha) = \frac{N}{2\pi i}\int_{L_2} \operatorname*{res}_{s_1=0} \frac{Z_\alpha(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{l+1}}\, ds_2,$$

and move the integral to the contour L_2^- given by given by $-\frac{1}{2\log J}+it$. The only residue is at $s_2=0$, and the contributions from the horizontal parts at height J and the integral over L_2^- are seen to be small as before (the first as with E_3 since J is big, and the second as with E_4 as the R-exponent of is sufficiently negative).

Writing the residues as integrals we get

$$T(\alpha) = N\cdot \operatorname*{res}_{s_2=0}\operatorname*{res}_{s_1=0} \frac{Z_\alpha(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{l+1}} = \frac{N}{(2\pi i)^2}\int_{C_2}\int_{C_1} \frac{Z_\alpha(s_1,s_2)R^{s_1+s_2}}{(s_1+s_2)^k(s_1s_2)^{l+1}}\, ds_1\, ds_2,$$

where C_1 is a small circle and C_2 is twice as big. We substitute $s_1 = s$, $s_2 = s\xi$ and take $C_3 : |\xi| = 2$ to get

$$\hat{T}(\alpha) = \frac{N}{(2\pi i)^2}\int_{C_3}\int_{C_1} Z_\alpha(s,s\xi)R^{s(\xi+1)}\cdot\frac{ds}{s^{k+2(l+1)}}\frac{s\,d\xi}{(\xi+1)^k\xi^{l+1}}.$$

Now we expand Z_α in a bivariate Taylor series, getting that

$$Z_\alpha(x,y) = \sum_{i=0}^\infty\sum_{j=0}^\infty \frac{Z_\alpha^{(i,j)}(0,0)}{i!j!}x^iy^j,$$

and by expanding $R^{s(\xi+1)}$ via the exponential function, we get

$$T(\alpha) = \frac{N}{(2\pi i)^2}\sum_{i=0}^\infty\sum_{j=0}^\infty \frac{Z_\alpha^{(i,j)}(0,0)}{i!j!}\sum_{d=0}^\infty\frac{(\log R)^d}{d!}\int_{C_3}\int_{C_1}\frac{s^ds^{i+j}\,ds}{s^{k+2l+1}}\frac{(\xi+1)^d\xi^j\,d\xi}{(\xi+1)^k\xi^{l+1}}.$$

The s-integral picks out $d = k+2l-(i+j)$ to give

$$T(\alpha) = \frac{N}{2\pi i}\sum_{i=0}^\infty\sum_{j=0}^\infty \frac{Z_\alpha^{(i,j)}(0,0)}{i!j!}\frac{(\log R)^{k+2l-(i+j)}}{\big(k+2l-(i+j)\big)!}\int_{C_3}(\xi+1)^{2l-(i+j)}\xi^{j-l-1}\,d\xi.$$

By typical methods (for instance, the residue at infinity), the integral is $\binom{2l-(i+j)}{l-j}^{\star}$, where the star restricts the arguments to have the same sign (else it is zero), and where $\binom{-A}{-B} = \binom{B-1}{A-1}(-1)^{A+B}$ for $A, B > 0$.

9.7.4. *Lower bounds on $G_\alpha(0,0)$ and bounds on derivatives.* We now restrict ourselves to the bi-disc $\mathbf{D}$ given by $|s_1|, |s_2| \le 1/8\sqrt{k}\log k$. From our above bounds with (15) on B_1 and B_2, the primes $p \ge U$ contribute no more than $\frac{ck}{\log H}$ to $\log G_\alpha$ in the larger domain $\eta \le 1/\log H$, and so in particular in $\mathbf{D}$. For the primes with $p \le U$ we have that $X_p(s_1, s_2) = \big(p^{-s_1} + p^{-s_2} - p^{-s_1-s_2}\big)$ is (in $\mathbf{D}$) given by

$$\big[1 - s_1 \log p\big] + \big[1 - s_2 \log p\big] - \big[1 - (s_1 + s_2)\log p\big] + O\big((|s_1^2| + |s_2^2|)(\log p)^2\big),$$

and thus (calculating the O-constant) we get $|X_p(s_1, s_2)| \le 1 + \frac{1}{3p}$. We have assumed that $\alpha_p \le 1 - \frac{1}{p}$, so that $\big|\alpha_p X_p(s_1, s_2)\big| \le 1 - \frac{1}{2p}$ and we get

$$\left|\log\left[1 - \alpha_p\left(\frac{1}{p^{s_1}} + \frac{1}{p^{s_2}} - \frac{1}{p^{s_1+s_2}}\right)\right]\right| \le \log\left[1 - \left(1 - \frac{1}{2p}\right)\right] \ll \log p.$$

Thus summing this over $p \le U$ gives a contribution to $\log G_\alpha$ that is bounded by $U \ll k$, and so by evaluating at $(0,0)$ we get the claimed lower bound of $G_\alpha(0,0) \gg 1/e^{ck}$ on the singular series.

Next we turn to bounding $\partial_1 Z_\alpha$ at $(0,0)$, though in comparison to $Z_\alpha(0,0)$ itself. The ζ-part of Z_α contributes an amount bounded by ck. With G_α, we again note that for $p \ge U$ and $|s_1|, |s_2| \le 1/\log H$, from the previous estimates for B_1 and B_2 we have the bound $|\log G_\alpha^{\ge U}(s_1, s_2)| \ll \frac{k}{\log H}$. Thus by the Cauchy estimate we have that the derivative of this with respect to either s_1 or s_2 is bounded by $\frac{k}{\log H} \cdot \log H$ for $|s_1|, |s_2| \le 1/2\log H$, and so in particular in $\mathbf{D}$.

For the primes $p \le U$, we can compute the derivative directly and get

$$\left|\frac{\partial_1 F_\alpha^{\le U}(s_1, s_2)}{F_\alpha^{\le U}(s_1, s_2)}\right| \ll \sum_{p \le U} \frac{\alpha_p |p^{-s_1}[1 - p^{-s_2}]| \log p}{|1 - \alpha_p X_p(s_1, s_2)|} \ll \sum_{p \le U} \frac{(\log p)^2}{\frac{1}{p}\sqrt{k}\log k} \ll k^{3/2},$$

while the ζ-quotient contribution to $G_\alpha^{\le U}$ and the ζ-quotient in the definition (16) are bounded in size by k. Thus we have that

$$\left|\frac{Z_\alpha^{(1,0)}(s_1, s_2)}{Z_\alpha(s_1, s_2)}\right| \ll k^{3/2}$$

in the region $|s_1|, |s_2| \le 1/8\sqrt{k}\log k$, and similarly with the s_2-derivative.

We write $Z_\alpha^{(1,0)} = Z_\alpha \cdot \frac{\partial_1 Z_\alpha}{Z_\alpha}$ so that the product formula for derivatives gives

$$Z_\alpha^{(m+1,n)} = \sum_{u=0}^{m} \binom{m}{u} \sum_{v=0}^{n} \binom{n}{v} Z_\alpha^{(u,v)} \left(\frac{\partial_1 Z_\alpha}{Z_\alpha}\right)^{(m-u,n-v)}. \tag{17}$$

We note that $Z_\alpha^{(m,0)}(0,0) = 0$ for $m \ge 1$, and similarly when taking derivatives of the second variable. Indeed, by the above formula we need only show that $(\log Z_\alpha)^{(m,0)}$ vanishes at $(0,0)$, and then the result follows by induction. We have

$$\log G_\alpha = \sum_p \sum_{k=1}^{\infty} \frac{\alpha_p(-1)^k}{k} H_p(s_1, s_2)^k \quad \text{where} \quad H_p(s_1, s_2) = \left(\frac{1}{p^{s_1}} + \frac{1}{p^{s_2}} - \frac{1}{p^{s_1+s_2}}\right).$$

Since $\partial_1^m H_p(s_1, s_2)(0,0) = 0$ for any $m \ge 1$, using the product rule for derivatives gives the same evaluation of $(\log G_\alpha)^{(m,0)}$ at the bi-origin. For the ζ-quotient part

of $\log Z_\alpha$, we note it is of the form $V(s_1,s_2)=\log U(s_1+s_2)-\log U(s_1)-\log U(s_2)$ for some function U. Expanding $\log U(s)$ in a power series $\sum_l c_l s^l$, we get that the mth s_1-derivative of $V(s_1,s_2)$ at the bi-origin is $m!\,c_m - m!\,c_m = 0$ as desired.

By the Cauchy formula for derivatives, there is some constant c with

$$\left|\left(\frac{\partial_1 Z_\alpha}{Z_\alpha}\right)^{(i,j)}(0,0)\right| = i!\,j!\cdot\left|\int\!\!\int \frac{\frac{\partial_1 Z_\alpha}{Z_\alpha}(s_1,s_2)\,ds_1\,ds_2}{s_1^{i+1}s_2^{j+1}\,(2\pi i)^2}\right| \le ck^{3/2}\cdot i!\cdot j!\cdot\big(8\sqrt{k}\log k\big)^{i+j},$$

where we integrated over the small circles $|s_1|,|s_2| = 1/8\sqrt{k}\log k$.

We divide (17) by Z_α, and write $Y_\alpha^{(m,n)} = \frac{Z_\alpha^{(m,n)}(0,0)}{Z_\alpha(0,0)}$; upon noting that $v=n$ does not contribute due to the vanishing of derivatives, the above bound yields

$$Y_\alpha^{(m+1,n)} \le ck^{3/2}\sum_{u=0}^{m}\binom{m}{u}(m-u)!\sum_{v=0}^{n-1}\binom{n}{v}(n-v)!Y_\alpha^{(u,v)}\big(8\sqrt{k}\log k\big)^{m+n-u-v}.$$

When $m=1$ we have

$$Y_\alpha^{(1,n)} \le 2ck^{3/2}\cdot n!\cdot\big(8\sqrt{k}\log k\big)^n,$$

and by induction, we get (using the result when $m\le n$)

$$Y_\alpha^{(m,n)} \le (2ck^{3/2})^m\cdot m!\cdot 2^n n!\cdot\big(8\sqrt{k}\log k\big)^n, \tag{18}$$

as from this assumption we can compute that (the $u=m$ term dominates)

$$\begin{aligned} Y_\alpha^{(m+1,n)} &\le ck^{3/2}\cdot m!\cdot n!\cdot\big(8\sqrt{k}\log k\big)^n\left[\sum_{u=0}^{m}\big(8\sqrt{k}\log k\big)^{m-u}(2ck^{3/2})^u\sum_{v=0}^{n-1}2^v\right] \\ &\le \big(ck^{3/2}\cdot m!\cdot n!\big)\cdot\big(8\sqrt{k}\log k\big)^n\cdot(m+1)\cdot 2(2ck^{3/2})^m\cdot 2^n. \end{aligned}$$

Returning to the expression

$$\hat{T}(\alpha) = N\cdot\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{Z_\alpha^{(i,j)}(0,0)}{i!j!}\frac{(\log R)^{k+2l-(i+j)}}{\big(k+2l-(i+j)\big)!}\binom{2l-(i+j)}{l-j}^{\star},$$

the $i=j=0$ term gives the main contribution as

$$\tilde{T}(\alpha) = N\cdot Z_\alpha(0,0)\cdot\binom{2l}{l}\cdot\frac{(\log R)^{k+2l}}{(k+2l)!}.$$

For $2\le i+j\le 2l$ we group according to $i+j=r$, taking $j\le i$ by symmetry, and bound the binomial coefficient by $\binom{2l}{l}$, and then multiply and divide by $Z_\alpha(0,0)$ before using (18) to get a contribution bounded as

$$\begin{aligned} E_5 &\ll N\cdot Z_\alpha(0,0)\binom{2l}{l}\cdot\sum_{r=2}^{2l}\sum_{j=1}^{r/2}(2ck^{3/2})^j\big(16\sqrt{k}\log k\big)^{r-j}\frac{(\log R)^{k+2l-r}}{\big(k+2l-r\big)!}, \\ &\ll N\cdot Z_\alpha(0,0)\binom{2l}{l}\cdot\sum_{r=2}^{2l}\big(32ck^2\log k\big)^{r/2}\frac{(\log R)^{k+2l-r}}{\big(k+2l-r\big)!} \\ &\ll N\cdot Z_\alpha(0,0)\binom{2l}{l}\frac{(\log R)^{k+2l}}{(k+2l)!}\cdot\sum_{r=2}^{2l}\big(32ck\sqrt{\log k}\big)^r\left(\frac{2k}{\log R}\right)^r \\ &\ll N\cdot Z_\alpha(0,0)\binom{2l}{l}\frac{(\log R)^{k+2l}}{(k+2l)!}\cdot\left(\frac{k^2\sqrt{\log k}}{\log R}\right)^2. \end{aligned}$$

Here we pulled out the desired main term, noting that $k+2l \le 2k$, and executed the r-sum assuming that $k^2\sqrt{\log k} = o\big(\log R\big)$.

When $2l \le i+j \le k+l$ we switch the binomial coefficient around and bound it by $\binom{j-l-1}{r-2l-1} \le \binom{r}{l} \le r^l/l! \le (6\sqrt{k})^l$ due to $l \sim \sqrt{k}$, and get

$$\begin{aligned}
E_5 &\ll N \cdot Z_\alpha(0,0) \cdot (6\sqrt{k})^l \sum_{r=2l}^{k+l} \sum_{j=1}^{r/2} (2ck^{3/2})^j \big(16\sqrt{k}\log k\big)^{r-j} \frac{(\log R)^{k+2l-r}}{\big(k+2l-r\big)!}, \\
&\ll N \cdot Z_\alpha(0,0) \cdot (6\sqrt{k})^l \sum_{r=2l}^{k+l} \big(32ck^2 \log k\big)^{r/2} \frac{(\log R)^{k+2l-r}}{\big(k+2l-r\big)!} \\
&\ll N \cdot Z_\alpha(0,0) \frac{(\log R)^{k+2l}}{(k+2l)!} \cdot (6\sqrt{k})^l \left(\frac{32ck^2\sqrt{\log k}}{\log R}\right)^{2l} \\
&\ll N \cdot Z_\alpha(0,0) \frac{(\log R)^{k+2l}}{(k+2l)!} \cdot \frac{1}{2^{2l}} \ll N \cdot Z_\alpha(0,0) \binom{2l}{l} \frac{(\log R)^{k+2l}}{(k+2l)!} \cdot \left(\frac{k^2\sqrt{\log k}}{\log R}\right)^2.
\end{aligned}$$

The last line notes essentially that $192ck^{9/2}\log k/(\log R)^2 \le 1/2$ due to the asymptotic $k \sim \dot{c}(\log N)^{4/9}/(\log\log N)^{2/9}$, and then that $(1/2)^{2l} \le \big(\frac{k^2\sqrt{\log k}}{\log R}\big)^2$ as $l \sim \sqrt{k}$. Noting that $Z_\alpha(0,0) = G_\alpha(0,0)$, this gives the desired bound that

$$T(\alpha) = N \cdot G_\alpha(0,0) \cdot \binom{2l}{l} \cdot \frac{(\log R)^{k+2l}}{(k+2l)!} \left[1 + O\left(\frac{k^2\sqrt{\log k}}{\log N}\right)^2\right].$$

We can note that the main part of the error term is induced by

$$\frac{Z_\alpha^{(1,1)}(0,0)}{Z_\alpha(0,0)} = \sum_p \left[\frac{\alpha_p(\log p)^2}{1-\alpha_p} - \frac{kp(\log p)^2}{(p-1)^2}\right] - k(\gamma^2 + 2\gamma_1)$$

where $\zeta(s) = \frac{1}{s-1} + \gamma - \gamma_1(s-1) + O\big((s-1)^2\big)$, and the sum will indeed be of size $k^2\log k$ in the case when $\big(1-\alpha_p\big) \approx 1/p$ for most primes $p \le k$.

10. Bombieri and the asymptotic sieve

We now discuss the limitations of sieve methods. For instance, if we only use information about congruence sums, we cannot detect information about primes. This is one instance of the *parity problem*. This has been overcome in some situations through the use of bilinear forms (similar to Vaughan's identity), but as Selberg pointed out long ago, cannot be overcome in general. We describe some of Bombieri's work in this genre.

10.1. **Limits of sieves.** Suppose that we have a nonnegative sequence (a_n) with summatory function $A(x)$, and there is a multiplicative function g such that the congruence sums satisfy $A_d(x) = g(d)A(x) + r_d(x)$ with

$$\sum_{d \le x^{1-\epsilon}} |r_d(x)| \ll_{B,\epsilon} \frac{A(x)}{(\log x)^B}$$

for all $B, \epsilon > 0$. This is about as strong of a statement that could be expected in general. Can we then get an asymptotic for $\sum_{p \le x} a_p$?

10.1.1. *Selberg's example.* The answer to this is in the negative. Let $\lambda(n)$ be the Liouville function which is totally multiplicative and -1 on all primes. We have

$$F(s) = \sum_{n=1}^{\infty} \frac{\lambda(n)}{n^s} = \prod_p \left(1 - \frac{1}{p^s} + \frac{1}{p^{2s}} - + \cdots\right) = \prod_p \left(1 + \frac{1}{p^s}\right)^{-1} = \frac{\zeta(2s)}{\zeta(s)}.$$

By the zero-free region for $\zeta(s)$, as with the prime number theorem we obtain

$$\sum_{n \le x} \lambda(n) \ll \frac{x}{e^{c\sqrt{\log x}}}.$$

Now consider the sequence $a_n = 1 + \lambda(n)$, which is clearly nonnegative. We have

$$\begin{aligned} A_d(x) &= \sum_{\substack{n \le x \\ n \equiv 0 (d)}} \big(1 + \lambda(n)\big) = \sum_{m \le x/d} \big(1 + \lambda(m)\lambda(d)\big) \\ &= \frac{x}{d} + O(1) + \lambda(d) \sum_{m \le x/d} \lambda(m) = \frac{x}{d} + O\left(\frac{x/d}{e^{c\sqrt{\log x/d}}}\right). \end{aligned}$$

So with $g(d) = 1$ for all d, for all $B, \epsilon > 0$ we get

$$\sum_{d \le x^{1-\epsilon}} |r_d(x)| \ll \sum_{d \le x^{1-\epsilon}} \frac{1}{d} \frac{x}{e^{c\sqrt{\log x/d}}} \ll \frac{x \log x}{e^{c\sqrt{\epsilon \log x}}} \ll_{B,\epsilon} \frac{x}{(\log x)^B}.$$

However, it is clear that $\sum_p a_p = 0$. Contrariwise, by taking $a_n \equiv 1$, we get a sequence with the same g and the same remainder estimate, but $\sum_{p \le x} a_p \sim x/\log x$.

This exhibits the parity problem. As Bombieri's work will show, we can get results for sums supported on (say) numbers with either one or two prime factors, or five/six, etc., but the information solely from congruence sums cannot distinguish between an even and odd number of prime factors.

10.2. **Preliminaries.** We define a generalisation of the von Mangoldt function. This is given by $\Lambda_k = \mu \star L^k$ where $L(n) = \log n$, so that $L^k = 1 \star \Lambda_k$ and

$$\Lambda_k(n) = \sum_{d|n} \mu(d) \left(\log \frac{n}{d}\right)^k.$$

Thus we have that $\Lambda_1 = \Lambda$ and can note that

$$1 \star \Lambda_{k+1} = L^{k+1} = L \cdot L^k = L \cdot \big(1 \star \Lambda_k\big) = L \star \Lambda_k + 1 \star L\Lambda_k = \big(1 \star \Lambda\big) \star \Lambda_k + 1 \star L\Lambda_k$$

and then by grouping and convolving with μ on the left, we obtain the recurrence formula $\Lambda_{k+1} = L\Lambda_k + \Lambda \star \Lambda_k$. This gives that Λ_k is nonnegative, and thus from $L^k = 1 \star \Lambda_k$ we get that $0 \le \Lambda_k \le L^k$. We note that Λ_k is supported on numbers with at most k distinct prime factors. Finally, the additivity of the logarithm and the binomial theorem together imply

$$\Lambda_k(mn) = \sum_{j=0}^{k} \binom{k}{j} \Lambda_j(m) \Lambda_{k-j}(n) \quad \text{when} \quad \gcd(m,n) = 1. \tag{19}$$

It will be technically convenient to have a slightly modified version of this, namely

$$\Lambda_k^x(n) = \sum_{d|n} \mu(d) \left(\log \frac{x}{d}\right)^k = \sum_{d|n} \mu(d) \left(\log \frac{x}{n} + \log \frac{n}{d}\right)^k = \sum_{j=0}^{k} \binom{k}{j} \Lambda_j(n) \left(\log \frac{x}{n}\right)^{k-j}.$$

Taking only the $j = k$ term gives a lower bound for $n \le x$, while an upper bound follows from iterating the bound $L\Lambda_k \le \Lambda_{k+1}$ to get $L^{k-j}\Lambda_j \le \Lambda_k$ for $j \le k$, so

$$\Lambda_k^x(n)\left(\frac{\log n}{\log x}\right)^k \le \frac{\Lambda_k(n)}{(\log x)^k}\sum_{j=0}^{k}\binom{k}{j}(\log n)^j\left(\log\frac{x}{n}\right)^{k-j} = \Lambda_k(n) \le \Lambda_k^x(n). \tag{20}$$

We also have that $\Lambda_k^x(n) \le (\log x)^k$ for $n \le x$, which easily follows by using the bound $\Lambda_j(n) \le (\log n)^j$ in the definition of $\Lambda_k^x(n)$ and use the binomial theorem.

10.2.1. *More inequalities for the modified generalised von Mangoldt functions.* We next generalise (19) for Λ_k^x. Assuming that $\gcd(m,n) = 1$, we can write any divisor $d|mn$ uniquely as $d = ab$ with $a|m$ and $b|n$. Again from additivity of the logarithm we have $\log\frac{x}{d} = \log\frac{x}{bm} + \log\frac{m}{a}$, and so we get

$$\Lambda_k^x(mn) = \sum_{d|mn}\mu(d)\left(\log\frac{x}{d}\right)^k = \sum_{a|m}\mu(a)\sum_{b|n}\mu(b)\left(\log\frac{x}{bm} + \log\frac{m}{a}\right)^k =$$

$$= \sum_{j=0}^{k}\binom{k}{j}\sum_{b|n}\mu(b)\left(\log\frac{x}{bm}\right)^j\sum_{a|m}\mu(a)\left(\log\frac{m}{a}\right)^{k-j} = \sum_{j=0}^{k}\binom{k}{j}\Lambda_{k-j}(m)\Lambda_j^{x/m}(n).$$

Now note that when $m > 1$ we have that $\Lambda_0(m) = 0$, and so we can eliminate $j = k$ from the j-sum. We use $\Lambda_{k-j} \le L^{k-j}$ and $\Lambda_j^{x/m} \le \Lambda_j^x$, and assume that $mn \le x$ so that $\log m \le \log\frac{x}{n}$. Expanding the definition of Λ_j^x then (sloppily) gives us that

$$\begin{aligned}
\Lambda_k^x(mn) &\le k\sum_{j=0}^{k-1}\binom{k-1}{j}(\log m)^{k-j}\Lambda_j^x(n)\\
&\le k\log m\sum_{j=0}^{k-1}\binom{k-1}{j}\left[\sum_{i=0}^{j}\binom{j}{i}\Lambda_i(n)\left(\log\frac{x}{n}\right)^{j-i}\right]\left(\log\frac{x}{n}\right)^{k-j-1}\\
&\le k\log m\sum_{i=0}^{k-1}\left[\sum_{j=i}^{k-1}\binom{j}{i}\binom{k-1}{j}\right]\Lambda_i(n)\left(\log\frac{x}{n}\right)^{k-1-i}\\
&\le k2^k\log m\sum_{i=0}^{k-1}\binom{k-1}{i}\Lambda_i(n)\left(\log\frac{x}{n}\right)^{k-1-i} = k2^k\Lambda_{k-1}^x(n)\log m. \qquad (21)
\end{aligned}$$

Since this is true when $mn \le x$ and $\gcd(m,n) = 1$ and $m > 1$, we can induct on the number of prime divisiors. For squarefree $n = \prod_i p_i$ with r prime factors we get that $\Lambda_k\big(\prod_i p_i, x\big) \ll_k (\log x)^{k-r}\prod_i \log p_i$. Here $(\log x)^{k-r}$ is an upper bound for $\Lambda_{k-r}^x(1)$. Using that $\Lambda_k^x(n)$ is supported on integers with no more than k prime divisors, for a nonnegative mulitplicative function g supported on squarefree integers we get

$$\sum_{n\le x} g(n)\Lambda_k(n,x) \ll_k \Big(\log x + \sum_{p\le x} g(p)\log p\Big)^k. \tag{22}$$

10.2.2. *Some more sieve facts.* We continue with some comments like those from Section 3.6, as we are interested in monotonicity of sieves. Let (λ_d) be an upper

bound sieve of level z, and g be a nonnegative multiplicative function supported on squarefree integers. Writing $\sigma = 1 \star \lambda$ so that $\lambda = \mu \star \sigma$, for all q we have

$$\sum_{\gcd(d,q)=1} \lambda_d g(d) = \sum_{\substack{\gcd(ab,q)=1 \\ \gcd(a,b)=1}} \sum \mu(a)\sigma(b)g(a)g(b) = \sum_{\gcd(b,q)=1} \sigma(b)g(b) \prod_{p \nmid bq} \big(1 - g(p)\big)$$
$$= \frac{\hat{g}(q)}{g(q)} \sum_{\gcd(b,q)=1} \sigma(b)g(b) \prod_{p \nmid b} \big(1 - g(p)\big) \leq \frac{\hat{g}(q)}{g(q)} \sum_d \lambda_d g(d),$$

where $\hat{g}(p) = \frac{g(p)}{1-g(p)}$ as before (assuming $g(p) < 1$), and the inequality comes from dropping the gcd condition on b (exploiting the positivity of σ), which puts us in the $q = 1$ case, and the argument is then reversed. By subtracting to get the complementary sum, for any prime $p \leq z$ we get

$$\sum_{d \equiv 0(p)} \lambda_d g(d) \geq \left(1 - \frac{\hat{g}(p)}{g(p)}\right) \sum_d \lambda_d g(d) = -\hat{g}(p) \sum_d \lambda_d g(d).$$

Let α be multiplicative and β be additive, both nonnegative, and

$$\eta(d) = \sum_{q|d} \alpha(q)\beta(q) = \sum_{q|d} \alpha(q) \sum_{p|q} \beta(p) = \sum_{p|d} \beta(p) \sum_{\substack{q|d \\ q \equiv 0(p)}} \alpha(q) =$$
$$= \sum_{p|d} \beta(p)\alpha(p) \prod_{l|(d/p)} \big[1 + \alpha(l)\big] = \sum_{p|d} \frac{\beta(p)\alpha(p)}{1 + \alpha(p)} \cdot \prod_{l|d} \big[1 + \alpha(l)\big].$$

In the second line, l is prime. We let $f(p) = g(p)\big(1 + \alpha(p)\big)$, and make the assumption that $f(p) < 1$ for all $p \leq z$. We then use the above inequality with f for each prime $p \leq z$, and weight each contribution by $\alpha(p)\beta(p)\frac{g(p)}{f(p)}$ before summing to get

$$\sum_{p \leq z} \alpha(p)\beta(p)\frac{g(p)}{f(p)} \sum_{d \equiv 0(p)} \lambda_d f(d) \geq -\sum_{p \leq z} \frac{\hat{f}(p)}{f(p)} \alpha(p)\beta(p)g(p) \cdot \sum_d \lambda_d f(d).$$

We now proceed to manipulate the left side of this, while for the right side we simply define $\psi(z) = \sum_{p \leq z} \frac{\alpha(p)\beta(p)g(p)}{1-f(p)}$. Noting that $\frac{\hat{f}(p)}{f(p)} = \frac{1}{1-f(p)}$, we get the inequality

$$-\psi(z) \sum_{d \leq z} \lambda_d f(d) \leq \sum_{p \leq z} \alpha(p)\beta(p)\frac{g(p)}{f(p)} \sum_{d \equiv 0(p)} \lambda_d f(d) = \sum_{d \leq z} \lambda_d f(d) \sum_{p|d} \alpha(p)\beta(p)\frac{g(p)}{f(p)}$$
$$= \sum_{d \leq z} \lambda_d g(d) \frac{f(d)}{g(d)} \cdot \sum_{p|d} \frac{\alpha(p)\beta(p)}{1 + \alpha(p)} = \sum_{d \leq z} \lambda_d g(d) \cdot \eta(d). \tag{23}$$

10.2.3. *Multiplicative functions.* The following subsections can get quite technical. In the sequel, we will assume an asymptotic (for $u \geq 1$) on a nonnegative multiplicative function g such as

$$\sum_{d \leq u}^{\flat} g(d) = c_1 \log u + c_2 + O\left(\frac{1}{(\log 9u)^B}\right), \tag{24}$$

(for some large B) where the $\flat$ restricts to squarefree integers. Indeed, we shall often simply assume that g is supported on squarefree integers. Also, it is nice to

have $0 \le g(d) < 1$. Since the sum is asymptotically to a linear function in $\log u$, we are in the situation of the linear sieve. As with Theorem 2.1, we have that

$$c_1 = \prod_p \big(1 + g(p)\big)\Big(1 - \frac{1}{p}\Big),$$

which we want to be positive. Although the error term we assume in (24) is significantly stronger than that in Theorem 2.1, it can often be established in practise.

10.2.4. *Coprimality restrictions.* We show that (24) implies similar asymptotics when restricted by a coprimality condition. Indeed, for $q \ge 1$ we shall show

$$\sideset{}{^\flat}\sum_{\substack{m \le x \\ \gcd(m,q)=1}} g(m) = \alpha_q\big[c_1\big(\log x + \beta_q\big) + c_2\big] + O\Big(\frac{\tau(q)}{(\log 9x)^B}\Big). \tag{25}$$

where

$$\alpha_q = \prod_{p|q} \frac{1}{1+g(p)} \quad \text{and} \quad \beta_q = \sum_{p|q} \frac{g(p)\log p}{1+g(p)}.$$

To show this, we write

$$P(s) = \prod_{p|q}\Big(1 + \frac{g(p)}{p^s}\Big)^{-1} = \sum_n \frac{u_n}{n^s} \quad \text{so that} \quad \sum_{\gcd(m,q)=1} \frac{g(m)}{m^s} = P(s)\sum_m \frac{g(m)}{m^s},$$

implying that

$$\sum_{\substack{m \le x \\ \gcd(m,q)=1}} g(m) = \sum_{d \le x} \sum_{m \le x/d} u_d\, g(m) = \sum_{d \le x} u_d \Big[c_1 \log \frac{x}{d} + c_2 + O\Big(\frac{1}{(\log 9x/d)^B}\Big)\Big].$$

We also introduce

$$Q(s) = \prod_{p|q}\Big(1 - \frac{g(p)}{p^s}\Big)^{-1} = \sum_n \frac{|u_n|}{n^s} \quad \text{and} \quad R(s) = -\frac{Q'}{Q}(s) = \sum_{p|q}\sum_{l=1}^{\infty} \frac{g(p)^l \log p}{p^{sl}}.$$

The main terms are estimated (with Rankin's trick for $d \ge x$) as

$$\sum_{d \le x} u_d = P(0) - \sum_{d \ge x} u_d = P(0) + O\Big(\sum_d \frac{|u_d|(\log d)^B}{(\log x)^B}\Big) = \alpha_q + O\Big(\frac{Q^{(B)}(0)}{(\log 9x)^B}\Big),$$

$$\text{and} \quad -\sum_{d \le x} u_d \log d = P'(0) + O\Big(\sum_{d \ge x} \frac{|u_d|(\log d)^{B+1}}{(\log x)^B}\Big) = P(0)\beta_q + O\Big(\frac{Q^{(B+1)}(0)}{(\log 9x)^B}\Big).$$

The last simplifies since $P(0) = \alpha_q$. Turning to an estimate of the error term, we note that $(\log 9x)^B = \big(\log \frac{9x}{d} + \log d\big)^B \ll \big(\log \frac{9x}{d}\big)^B (\log 3d)^B$, so we get a bound of

$$\sum_{d \le x} \frac{|u_d|(\log 3d)^B}{(\log 9x)^B} = \frac{1}{(\log 9x)^B} \sum_{d \le x}\sum_{j=0}^{B} \binom{B}{j} |u_d|(\log d)^j (\log 3)^{B-j} \ll_B \frac{\sum_j \big|Q^{(j)}(0)\big|}{(\log 9x)^B}.$$

Differentiating the equation $-Q'(s) = Q(s)R(s)$ repeatedly $(j-1)$ times gives

$$-Q^{(j)}(0) = \sum_{k=0}^{j-1} \binom{j-1}{k} Q^{(k)}(0) R^{(j-1-k)}(0) \ll_B \max_{k<j} \big|Q^{(k)}(0)\big| \sum_{p|q}\sum_{l=1}^{\infty} g(p)^l (2l \log p)^j$$

The bound (24) implies that $g(p) \ll 1/(\log p)^B$, and combined with $g(p) < 1$ this gives that the last sum is bounded uniformly in p. So we get $|Q^j(0)| \ll_j Q(0)\omega(q)^j$ by induction, which gives the total error to be bounded as

$$\ll_B \frac{Q(0)\omega(q)^{B+1}}{(\log 9x)^B} \ll \frac{\alpha_q \tau(q)}{(\log 9x)^B},$$

where we noted that

$$Q(0)\cdot\omega(q)^{B+1} \ll \prod_{p|q} \frac{1}{1-g(p)} \cdot (4/3)^{\omega(q)} B^{B^2} \ll_{B,g} \prod_{p|q} \frac{2}{1+g(p)} = \alpha_q \tau(q),$$

since, again from (24), there are only finitely many primes with $\frac{1+g(p)}{1-g(p)} \geq 3/2$.

10.2.5. *Sums over primes.* We now show that (24) implies the expected asymptotic

$$\sum_{p\leq u} g(p)\log p = \log u + c + O\Big(\frac{\log 9u}{(\log 9u)^B}\Big). \tag{26}$$

for some constant c. To see this, we can use Tchebyshev's trick with $ng(n)$ to note

$$\sum_{n\leq x} g(n)\cdot n\log n = \sum_{mp\leq x}\sum g(mp)\cdot mp\log p = \sum_{p\leq x} g(p)\, p\log p \sum_{\substack{m\leq x/p\\ \gcd(m,p)=1}} mg(m).$$

By partial summation of (24) we have the asymptotic

$$\sum_{n\leq x} g(n)\cdot n\log n = c_1 x(\log x - 1) + O\Big(\frac{x(\log x)}{(\log 9x)^B}\Big),$$

and by partial summation of (25) we have

$$\sum_{\substack{n\leq x\\ \gcd(n,p)=1}} n\cdot g(n) = \frac{c_1}{1+g(p)}x + O\Big(\frac{x}{(\log 9x)^B}\Big).$$

Plugging this into the previous display, we get

$$c_1 x\log x - c_1 x = c_1 x \sum_{p\leq x} \frac{g(p)\log p}{1+g(p)} + O\Big(\frac{x(\log x)}{(\log 9x)^B}\Big).$$

To obtain (26), we use $g(p) < 1$ and $g(p) \ll 1/(\log p)^B$ to note that the sum here is

$$\sum_{p\leq x} g(p)\log p + \sum_{p\leq x}\sum_{k\geq 2}(-1)^k g(p)^k \log p = \sum_{p\leq x} g(p)\log p + \tilde{c} + O\Big(\frac{\log x}{(\log 9x)^B}\Big).$$

10.2.6. *Sums over almost primes.* Furthermore, once we have this relation for sums over primes, we can derive a relation for sums over almost primes, namely that

$$\sum_{n\leq u} g(n)\Lambda_k(n) = (\log u)^k + c\cdot k(\log u)^{k-1} + O\big((\log u)^{k-2+1/(B-1)}\big). \tag{27}$$

Writing $n = md$, by the recurrence for Λ_{k+1} we have

$$P_{k+1}(x) = \sum_{n\leq x} g(n)\Lambda_{k+1}(n) = \sum_{n\leq x} g(n)\Lambda_k(n)\log n + \sum_{m\leq x} g(m)\Lambda_k(m) \sum_{\substack{d\leq x/m\\ \gcd(m,d)=1}} g(d)\Lambda(d).$$

The inner sum is estimated from (26), and by putting this back into the above and turning $m \to n$, the logarithm simplifies and we get

$$P_{k+1}(x) = \sum_{n \le x} g(n)\Lambda_k(n)\Big[\log x + c + O\Big(\sum_{p|n} g(p)\log p\Big) + O\Big(\frac{(\log 9x/n)}{(\log 9x/n)^B}\Big)\Big].$$

We bound the contribution of the first error term as

$$\sum_{p \le x} g(p)\log p \sum_{m \le x/p} g(pm)\Lambda_k(pm) \ll_k \sum_{p \le x} g(p)^2\Lambda_k(p)\log p \sum_{m \le x} g(m)\Lambda_{k-1}(m)$$
$$\ll \sum_{p \le x} g(p)^2(\log p)^{k+1} \sum_{m \le x} g(m)\Lambda_{k-1}(m) \ll (\log x)^{k-1},$$

where we used (19) to note that $\Lambda_k(pm) \le \Lambda_k(p)\Lambda_{k-1}(m)$ as $p \ne 1$ (taking care when $k = 1$) and in the last step used induction before noting that the p-sum converges. Indeed, returning to the main term, induction now gives (27) upon taking y with $x/y = e^{(\log x)^{1/(B-1)}}$ so that $\frac{\log 9x/y}{(\log 9x/y)^B} \ll \frac{1}{\log x}$ and noting that then

$$\sum_{n \le x} g(n)\Lambda_k(n)\frac{\log 9x/n}{(\log 9x/n)^B} \ll \sum_{n \le y} g(n)\Lambda_k(n)\frac{\log 9x/y}{(\log 9x/y)^B} + \sum_{y \le n \le x} g(n)\Lambda_k(n),$$

and this is bounded by $(\log x)^{k-1} + (\log x)^{k-1}\log\frac{x}{y} \ll (\log x)^{k-1+1/(B-1)}$.

10.3. **Statement of results.** The first main result of Bombieri is the following:

Proposition 10.1. *Suppose that (a_n) is a sequence of nonnegative real numbers supported on* ***squarefree*** *integers in $[1, x]$. Suppose there is some multiplicative function g and some constants $c_1 > 0$, $B > 20$, and c_2 such that we have the estimate $A_d(x) = g(d)A(x) + r_d(x)$ with*

$$\sum_{d \le u}^{\flat} g(d) = c_1 \log u + c_2 + O\Big(\frac{1}{(\log 9u)^B}\Big)$$

for all u with $1 \le u \le x$ and, letting $D \ge x^{1-1/2k} \ge x^{3/4}$ be a parameter,

$$\sum_{d \le D}^{\flat} \tau_5(d)|r_d(x)| \ll \frac{A(x)}{(\log x)^2}.$$

Supposing a crude bound like $\sum_{n \le D} a_n \ll A(x)/(\log x)^2$, for $k \ge 2$ we have

$$S_k(x) = \sum_{n \le x} a_n\Lambda_k(n) = Hk \cdot A(x)(\log x)^{k-1}\Big[1 + O_{g,k}\Big(\frac{\log(x/D)}{\log x}\Big)\Big], \tag{28}$$

where $H = \prod_p \big(1 - g(p)\big)\big(1 - \frac{1}{p}\big)^{-1}$ [this corresponds to $1/c_{\hat{g}}$ in previous notation].

The condition that the sequence be supported on squarefree integers is technically convenient, and will be removed below. In the condition on the remainder term, the thrust of Bombieri's final result is that D can be taken as $x^{1-\epsilon}$ for any $\epsilon > 0$, so that the remainder term is quite well-behaved — the extra τ_5 in the sum should only perturb our estimate by a power of logarithm, which does not bother us much. The third condition just says a_n is not too spiky. The condition that (a_n) is supported on $[1, x]$ is commonly denoted as the "local sieve", and allows us to ignore moduli larger than x without comment.

We shall first approximate Λ_k with Λ_k^x, and here the crude bound implies that the induced error is small. The term $\big(\log x/d\big)^k$ from the convolution for Λ_k^x is small (when $k \geq 2$) for d near x, so the large moduli (larger than D^2/x) contribute little. The main term is then computed via congruence sums for the smaller moduli.

A consequence of Proposition 10.1 is the Selberg relation; taking $a_n = 1$ for all n with $1 \leq n \leq x$ and $k = 2$, we get

$$\sum_{p \leq x} (\log p)^2 + \sum_{p \leq x} \sum_{q \leq x/p} \log p \log q = \sum_{n \leq x} \Lambda_2(n) \sim 2x \log x,$$

from which the prime number theorem can be obtained via elementary means.

10.4. **First transformations.** By definition we have $S_k(x) = \sum_n a_n \Lambda_k(n)$. We define $S_k^x(x) = \sum_n a_n \Lambda_k^x(n)$, and from (20) we get

$$S_k^x(x) \geq S_k(x) \geq \sum_{n \leq x} a_n \Lambda_k^x(n) \left(\frac{\log n}{\log x}\right)^k \geq \left(\frac{\log D}{\log x}\right)^k \sum_{n \geq D} a_n \Lambda_k^x(n) + \sum_{n \leq D} a_n (\log n)^k.$$

We extend the first sum to $n \leq x$, and use the crude bound on the errors to get

$$S_k^x(x) \geq S_k(x) \geq (1-\epsilon)^k S_k^x(x) + O\bigg((\log x)^k \frac{A(x)}{(\log x)^2}\bigg),$$

where $\epsilon = \frac{\log(x/D)}{\log x}$. So if we can prove the asymptotic formula (28) for $S_k^x(x)$, then we will have $|S_k^x(x) - S_k(x)| \ll k\epsilon S_k^x(x)$, and so (28) will also be true for $S_k(x)$.

10.5. **The introduction of an upper-bound sieve.** The next technical reduction is quite important. We intend to split the convolution for $\Lambda_k^x(n)$ based on whether a divisor d is bigger than D^2/x. However, this splitting forgets the information that $\Lambda_k^x(n)$ is supported on integers with no more than k distinct prime factors. Bombieri introduces an upper-bound sieve λ_e of level z where we will have $z = x/D$ in the end. In particular, we have that

$$\lambda_1 = 1, \quad |\lambda_e| \leq \tau_3(e), \quad \lambda_e = 0 \quad \text{for } e > z,$$

(we could perhaps get $|\lambda_e| \leq 1$ by using a combinatorial sieve) and we have that

$$\sigma_n = 1 \star \lambda = \sum_{e|n} \lambda_e \geq 0 \quad \text{for all } n.$$

Let $P_z = \prod_{p \leq z} p$ and note $\sigma_n = 1$ when $\gcd(n, P_z) = 1$ (only $e = 1$ contributes).

Since $z \leq x^{1/2k}$, this is often true for things in the support of $\Lambda_k^x(n)$. We get

$$\hat{S}_k^x(x) = \sum_{n \leq x} a_n \sigma_n \Lambda_k^x(n) = \sum_{n \leq x} a_n \Lambda_k^x(n) + O\bigg(\sum_{\substack{n \leq x \\ \gcd(n, P_z) > 1}} a_n \big(1 + \sigma_n\big) \Lambda_k^x(n)\bigg).$$

Due to fact that $|\lambda_e| \leq \tau_3(n)$, we get that $|\sigma_n| \leq 4^k$ for integers n with at most k prime factors. Noting that $\Lambda_k^x(n)$ is supported on such integers, we break the sum at D to get a bound for the error term of

$$E_1 \ll 4^k \sum_{\substack{n \leq x \\ \gcd(n, P_z) > 1}} a_n \Lambda_k^x(n) \ll \sum_{\substack{D \leq n \leq x \\ \gcd(n, P_z) > 1}} a_n \Lambda_k^x(n) + (\log x)^k \sum_{n \leq D} a_n.$$

The second sum is small by the crude bound. By writing $n = mp$ where $p \geq n^{1/k}$ is the largest prime factor of n and $m \leq n^{1-1/k} \leq x^{1-1/k}$, by (21) we get that the remaining part of the error term is bounded by

$$\hat{E}_1 \ll \sum_{\substack{m \leq x^{1-1/k} \\ \gcd(m,P_z)>1}} \sum_{\substack{D/m \leq p \leq x/m \\ \gcd(p,m)=1}} a_{mp} \Lambda_k^x(mp) \ll \sum_{\substack{m \leq x^{1-1/k} \\ \gcd(m,P_z)>1}} k2^k \Lambda_{k-1}^x(m) \sum_{\substack{D/m \leq p \leq x/m \\ \gcd(p,m)=1}} a_{mp} \log p.$$

Estimating $\log p \leq \log x$, we now use a sieve inside a sieve (as it were) on the sum over p. That is, we let $(b_l) = (a_{lm})$, and note that for primes not dividing m, the distribution function g is the same for $\vec{a}$ and $\vec{b}$. We write $y = D/m$ and then estimate by Theorem 4.1 (Selberg's upper bound sieve) that

$$\sum_{\substack{p \leq x/m \\ \gcd(p,P_y)=1}} a_{mp} \leq \sum_{\substack{l \leq x/m \\ \gcd(l,P_y)=1}} b_l \ll B(x) \prod_{\substack{p \leq \sqrt{y} \\ p \nmid m}} \big(1 - g(p)\big) + \sideset{}{^\flat}\sum_{l \leq y} \tau_3(l) |r_{ml}(x)|.$$

Since $B(x) = A_m(x)$, upon accounting for primes $p|m$ the error estimate becomes

$$E_1 \ll \sum_{\substack{m \leq x^{1-1/k} \\ \gcd(m,P_z)>1}} \Lambda_{k-1}^x(m) \log x \left(\frac{\hat{g}(m)}{g(m)} \frac{A_m(x)}{\log D/m} + \sideset{}{^\flat}\sum_{l \leq D/m} \tau_3(l) |r_{ml}(x)| \right).$$

We note $D/m \geq x^{1/2k}$ so that $\log D/m \gg_k \log x$, and $\frac{\hat{g}(m)}{g(m)} \ll \tau(m)$ as there are finitely many p with $\hat{g}(p) \geq 2g(p)$. Using $A_m(x) = g(m)A(x) + r_m(x)$ we then get

$$E_1 \ll \sum_{\substack{m \leq x^{1-1/k} \\ \gcd(m,P_z)>1}} \Lambda_{k-1}^x(m) \Big(\hat{g}(m) A(x) + \tau(m) |r_m(x)| \Big) + \sum_{m \leq x^{1-1/k}} (\log x)^k \sideset{}{^\flat}\sum_{d \leq D} \tau_4(d) |r_d(x)|.$$

The remainder terms are bounded by $A(x)(\log x)^{k-2}$, and in the $\hat{g}$-sum we write $m = lp$ where $p \leq z$ is prime, and use (21) and then (22) and (26) to get

$$\sum_{\substack{m \leq x^{1-1/k} \\ \gcd(m,P_z)>1}} \hat{g}(m) \Lambda_{k-1}^x(m) \ll_k \left(\sum_{p \leq z} \hat{g}(p) \log p \right) \left(\sum_{l \leq x} \hat{g}(l) \Lambda_{k-2}^x(l) \right) \ll (\log z)(\log x)^{k-2}.$$

Since $\log z = \epsilon \log x$, this gives a total error estimate of

$$\hat{S}_k^x(x) = S_k^x(x) + O\big(\epsilon A(x) (\log x)^{k-1}\big).$$

10.5.1. *The choice of sieve.* Let $f(p) = g(p) \frac{2+g(p)}{1+g(p)}$ be multiplicative and let $\mathbf{S}_g$ be the (finite) set of primes p for which $f(p) \geq 1$. We wish to shape our sieve after f, but at the same time we must exclude the primes with $f(p) \geq 1$. From the Selberg sieve (9) we choose λ_e supported outside $\mathbf{S}_g$ in such a way to minimise

$$G = \sum_e \lambda_e f(e) = \Bigg(\sum_{\substack{e \leq \sqrt{z} \\ e \notin \mathbf{S}_g}} \hat{f}(e) \Bigg)^{-1}.$$

From (26) we see that f satisfies the regularity hypothesis for a sieve of dimension 2, and the crude bound follows since g satisfies the same [due to (24)]. Letting $\tilde{f}$ be the restriction of f outside $\mathbf{S}_g$, (and similarly with $\tilde{g}$), from Theorem 2.1 we get

$$\sum_e \lambda_e f(e) = \sum_e \lambda_e \tilde{f}(e) \ll \frac{1}{(\log z)^2}. \tag{29}$$

10.6. **Splitting and first estimation.** Now we split the sum $\hat{S}_k^x(x)$ according to the size of the divisor in the convolution for Λ_k^x. This gives that

$$\hat{S}_k^x(x) = \sum_{n\le x} a_n\sigma_n \sum_{\substack{d|n\\ d\le D^2/x}} \mu(d)\Big(\log\frac{x}{d}\Big)^k + \sum_{n\le x} a_n\sigma_n \sum_{\substack{d|n\\ d> D^2/x}} \mu(d)\Big(\log\frac{x}{d}\Big)^k = U_1 + U_2.$$

Here we estimate the secondary term U_2. Writing $n = dt$ we have that

$$\begin{aligned} U_2 &\ll \sum_{n\le x} a_n\sigma_n \sum_{\substack{d|n\\ d> D^2/x}} \Big(2\log\frac{x}{D}\Big)^k \le \Big(2\log\frac{x}{D}\Big)^k \sum_{n\le x} a_n\sigma_n \sum_{\substack{t|n\\ t<nx/D^2}} 1 \\ &\ll \Big(2\log\frac{x}{D}\Big)^k \sum_{e\le z}\lambda_e \sum_{t<x^2/D^2} \sum_{\substack{n\equiv 0(t)\\ n\equiv 0(e)}} a_n \\ &\ll (2\epsilon\log x)^k A(x)\cdot \sum_{e\le z}\lambda_e \sum_{t<x^2/D^2} g(m) + (\log x)^k \sum_{e\le z}\tau_3(e) \sum_{t<x^2/D^2} r_m(x), \end{aligned}$$

where $m = \operatorname{lcm}(e,t)$. Provided that $z(x^2/D^2)\le D$, the remainder term is

$$U_2^R \ll \sideset{}{^\flat}\sum_{m\le zx^2/D^2} \tau_5(m) r_m(x) \ll A(x)(\log x)^{k-2},$$

The choice of $z = x/D$ and requirement of $D\ge x^{3/4}$ ensure this.

We are left to deal with the primary term. Our choice of λ will save two logarithms here, and the t-sum will only lose one. Since g is supported on squarefree integers, We can assume that t is squarefree, and so we are left with

$$\hat{U}_2 \ll (\epsilon\log x)^k A(x)\cdot \sum_{e\le z}\lambda_e \sideset{}{^\flat}\sum_{t\le x^2/D^2} g(m) = (\epsilon\log x)^k A(x) \sum_{e\le z}\lambda_e g(e) \sum_{q|e} \sideset{}{^\flat}\sum_{\substack{l<x^2/qD^2\\ \gcd(l,q)=1}} g(l),$$

writing $q = \gcd(e,t)$ and $l = m/q$. The inner sum is now estimated by (25) to get

$$\sideset{}{^\flat}\sum_{\substack{l<x^2/qD^2\\ \gcd(l,q)=1}} g(l) = \alpha(q)\Big[c_1\log\frac{x^2}{qD^2} + c_2 + c_1\big[\log(q) - \hat{\beta}(q)\big]\Big] + O\Big(\frac{\tau(q)}{(\log x/D)^B}\Big)$$

where α is multiplicative and $\hat{\beta}$ is additive, given on primes by $\alpha(p) = \frac{1}{1+g(p)}$ and $\hat{\beta}(q) = \alpha(p)\log p$. We then sum over q and multiply by $g(e)$ to get

$$\hat{U}_2 \ll (\epsilon\log x)^k A(x)\cdot \sum_{e\le z}\lambda_e\Big[\Big(c_1\log\frac{x^2}{D^2} + c_2\Big) f(e) - c_1 g(e)\eta(e) + O\Big(\frac{g(e)\tau_3(e)}{(\log x/D)^B}\Big)\Big]$$

where $f(p) = g(p)\big(1+\alpha(p)\big) = g(p)\cdot\frac{2+g(p)}{1+g(p)}$ is multiplicative and $\eta(e) = \sum_{d|e}\alpha(d)\hat{\beta}(d)$.

Since we have $\lambda_e = 0$ when $(f,g)\ne(\tilde{f},\tilde{g})$, in the above expression we can replace the former by the latter. Then we have $\tilde{f}(p) < 1$ for all primes, and so by (23) the e-sum here is bounded above by

$$\Big(c_1\log\frac{x}{D} + c_2 + c_1\tilde{\psi}(z)\Big)\Big[\sum_{e\le z}\lambda_e\tilde{f}(e)\Big] + O\Big(\frac{1}{(\log z)^B}\cdot\prod_{p\le z}\big[1 + g(p)\tau_9(p)\big]\Big)$$

where

$$\tilde{\psi}(z) = \sum_{p\le z} \frac{\alpha(p)\hat{\beta}(p)\tilde{g}(p)}{1-\tilde{f}(p)} = \sum_{p\le z} \frac{\tilde{g}(p)\log p}{1-2g(p)^2-g(p)^3} \ll \log z.$$

Recalling our choice of sieve weights and (29), we have that $\sum_e \lambda_e \tilde{f}(e) \ll 1/(\log z)^2$, while we can bound $\prod_p \big[1+9g(p)\big] \le \prod_p \big(1+g(p)\big)^9 \ll (\log z)^9$. Recalling that $\log z = \epsilon \log x$, these then give the desired bound that (we use $k \ge 2$ here)

$$U_2 \ll (\epsilon \log x)^k A(x) \cdot c_1 \log z \cdot (\log z)^{-2} \ll A(x)(\epsilon \log x)^{k-1} \ll \epsilon A(x)(\log x)^{k-1}.$$

10.7. **Second estimation.** Next we turn to the estimation of U_1. We have

$$U_1 = \sum_{n\le x} a_n \sigma_n \sum_{\substack{d\mid n\\ d\le D^2/x}} \mu(d)\left(\log\frac{x}{d}\right)^k = \sum_{e\le z}\lambda_e \sum_{d\le D^2/x}\mu(d)\left(\log\frac{x}{d}\right)^k \cdot \sum_{\substack{n\le x\\ n\equiv 0(m)}} a_n,$$

where here we have $m = \operatorname{lcm}(d,e)$. The inner sum is $A_m(x)$, and we insert its approximation $g(m)A(x) + r_m(x)$, estimating the remainder term as

$$(\log x)^k \sideset{}{^\flat}\sum_{e\le z} \tau_3(e) \sum_{d\le D^2/x} |r_m(x)| \ll (\log x)^k \sideset{}{^\flat}\sum_{m\le D} \tau_5(m)|r_m(x)| \ll A(x)(\log x)^{k-2}.$$

We are left to evaluate the main term

$$\hat{U}_1 = A(x)\sum_{e\le z}\lambda_e \sum_{d\le D^2/x} g(m)\mu(d)\left(\log\frac{x}{d}\right)^k.$$

10.7.1. *A trick to evaluate the main term.* We shall next use a trick to avoid evaluation of the main term. We note that it is somewhat independent of the actual sequence, as it only depends on the multiplicative function g [and its size on $A(x)$]. In particular, we can choose a sequence (b_n) with the same g for which we can estimate the main term directly. We let $b_n = \mu(n)^2\hat{g}(n)$ for n with $x/e < n \le x$, where $\hat{g}$ is multiplicative with $\hat{g}(p) = \frac{g(p)}{1-g(p)}$, and $e \approx 2.718$ has $\log e = 1$.

Since g satisfies (24), we have that $\hat{g}$ does also, but with different constants. We can note that the leading coefficient is given by

$$c_3 = \prod_p \big(1+\hat{g}(p)\big)\left(1-\frac{1}{p}\right) = \prod_p \big(1-g(p)\big)^{-1}\left(1-\frac{1}{p}\right) = 1/H.$$

Thus for every squarefree d we have

$$B_d(x) = \sum_{\substack{x/e<n\le x\\ n\equiv 0(d)}} \hat{g}(n) = \hat{g}(d) \sum_{\substack{x/ed<m\le x/d\\ \gcd(d,m)=1}} \hat{g}(m) = c_3\hat{g}(d)\prod_{p\mid d}\frac{1}{1+\hat{g}(p)} + O\left(\frac{\tau(d)\hat{g}(d)}{(\log x/9d)^B}\right)$$

where we used (25) again, and exploited that $\log e = 1$.

Here we have that $B(x) = c_3$, and we note that $\frac{\hat{g}(p)}{1+\hat{g}(p)} = g(p)$, so that $\vec{b}$ is well-approximated by g. We are left to estimate the cumulative error term as

$$\sideset{}{^\flat}\sum_{d\le D} \frac{\tau(d)\hat{g}(d)}{(\log x/D)^B} \ll \frac{1}{(\epsilon\log x)^B}\prod_{p\le x}\big(1+2\hat{g}(p)\big) \ll \frac{1}{(\log x)^B}\prod_{p\le x}\big(1+\hat{g}(p)\big)^2 \ll \frac{(\log x)^2}{(\log x)^B}.$$

This gives the remainder estimate, but with $(B-2)$ instead of B; this is OK because we required that $B \ge 20$ to begin, and only used that $B \ge 15$ in the

proof. Thus $\vec{b}$ satisfies the conditions of Proposition 10.1 with the multiplicative function g. Finally we conclude by noting that for $\vec{b}$ by (27) we have

$$S_k(x) = \sum_{n\le x} b_n\Lambda_k(n) = \sum_{x/e<n\le x} \mu(n)^2\hat{g}(n)\Lambda_k(n) = k(\log x)^{k-1} + O\big((\log x)^{k-2}\big).$$

10.8. **Generalising from squarefree sequences.** Now we show that if Proposition 10.1 is true for (a_n) supported on squarefree integers, then it is true if we drop the squarefree restriction. To this end, we write $n = l\hat{n}$ where $\hat{n}$ is squarefree and $l|\hat{n}^\infty$; in fact, $\hat{n}$ is the squarefree kernel of n. We let

$$b_m = \mu(m)^2 \sum_{l|m^\infty} a_{lm},$$

so that for squarefree d we have

$$B_d(x) = \sum_{\substack{m\le x\\ d|m}} \mu(m)^2 \sum_{l|m^\infty} a_{lm} = \sum_{\substack{n\le x\\ d|n}} a_n \sum_{\substack{m|n\\ (n/m)|m^\infty}} \mu(m)^2.$$

Calling the last inner sum $T(n)$, we have that every prime that divides n must also divide m, but also that m must be squarefree for a nonzero contribution. Therefore only when $m = \hat{n}$ is there a contribution to the sum, so that $T(n) = 1$ and we get that $B_d(x) = A_d(x)$ for all squarefree d. Similarly, the definition of $\Lambda_k^x(n)$ implies that only squarefree divisors d contribute, and for such d we note that $d|n$ is equivalent to $d|\hat{n}$. This gives $\Lambda_k^x(n) = \Lambda_k^x(\hat{n})$, and so

$$\sum_{m\le x} b_m\Lambda_k^x(m) = \sum_{m\le x} \mu(m)^2 \sum_{(n/m)|m^\infty} a_n\Lambda_k^x(n) = \sum_{n\le x} a_n\Lambda_k^x(n)T(n) = \sum_{n\le x} a_n\Lambda_k^x(n).$$

We conclude that if (28) is true for (b_m) then it is also true for (a_n).

10.9. **The parity problem.** The above result of Bombieri's was only a first step in his work. Proposition 10.1 implies that if we have the remainder estimate

$$\sideset{}{^\flat}\sum_{d\le x^{1-\epsilon}} \tau_5(d)|r_d(x)| \ll_\epsilon \frac{A(x)}{(\log x)^2} \tag{30}$$

for all $\epsilon > 0$, then for all $k \ge 2$ we have the asymptotic

$$\sum_{n\le x} a_n\Lambda_k(n) \sim H\cdot A(x)\cdot k(\log x)^{k-1}.$$

Of course, it is natural to expect this asymptotic also to be true when $k = 1$. Bombieri shows that the asymptotic holds for $k = 1$ if and only if the sequence (a_n) gives equal weight to integers n with an even/odd number of prime divisors. In particular, Selberg's example puts all the weight on those with an even number of prime divisors. Also, note that the generalised von Mangoldt functions $\Lambda_k(n)$ for $k \ge 2$ give equal weight to those n with an even/odd number of prime factors.

It transpires that the generalised von Mangoldt functions Λ_k with $k \ge 2$ together with the multi-convolutions $\Lambda_{\vec{k}} = \Lambda_{k_1} \star \cdots \star \Lambda_{k_r}$ with some $k_i \ge 2$ form some sort of a basis for the functions that equally weight the parity of number of divisors. Bombieri shows that knowing (30) for all $\epsilon > 0$ also implies that

$$\sum_{n\le x} a_n\Lambda_{\vec{k}}(n) \sim H\cdot A(x)\cdot \frac{(\vec{k})!}{(|\vec{k}|-1)!}(\log x)^{|\vec{k}|-1}$$

assuming that some $k_i \geq 2$, where $|\vec{k}| = \sum_i k_i$ and $(\vec{k})! = \prod_i k_i!$.

Orthogonal to the functions that equally weight parity are the multi-convolutions of just Λ, namely $\vec{k} = (1, \ldots, 1)$. If we are able to show an asymptotic for such vectors, then we can obtain an asymptotic formula for a_n summed over integers with a fixed number of prime factors. Writing $\mathbf{P}_r$ for the set of squarefree integers with exactly r prime factors, an approximation theorem of Stone-Weierstrass type would extend the result to get

$$\sum_{\substack{n \leq x \\ n \in \mathbf{P}_r}} a_n F_r(n) \sim \delta_r \left(\int_{T_r} G_r \, d\mu_r \right) \cdot \frac{H \cdot A(x)}{\log x}, \tag{31}$$

for some constants δ_r, where $F_r(n) = G_r\big(\frac{\log p_1}{\log n}, \ldots, \frac{\log p_r}{\log n}\big)$ and G_r is smooth and compactly supported on

$$T_r = \Big\{ (u_1, \cdots, u_r), 0 < u_r < \cdots u_1 < 1, \sum_{i=1}^{r} u_i = 1 \Big\} \quad \text{with} \quad d\mu_r = \frac{du_1 \cdots du_{r-1}}{u_1 \cdots u_r}.$$

Knowing δ_r is essentially equivalent to estimating $\sum_n a_n \Lambda_{\vec{k}}(n)$ with $\vec{k} = (1)^r$, as the other contributants in the F_r-approximation are determined as indicated above.

What Bombieri realised is that the δ_r are determined by δ_1, where the latter is given by the asymptotic formula

$$\sum_{p \leq x} a_p \sim \delta_1 \cdot H \cdot \frac{A(x)}{\log x}.$$

Obviously we expect $\delta_1 = 1$ for any well-distributed sequence, while Selberg's example has $\delta_1 = 0$ (taking different linear combinations of the constant sequence and the Liouville function can give any value with $0 \leq \delta_1 \leq 2$). Bombieri's phrasing of the parity problem is that we have

$$\delta_r = \begin{cases} \delta_1 & \text{if } r \text{ is odd,} \\ 2 - \delta_1 & \text{if } r \text{ is even.} \end{cases}$$

We can note that we have $\Lambda_3 - L\Lambda_2 = \Lambda \star \Lambda_2 = \frac{L}{2}(\Lambda \star \Lambda) + (\Lambda \star \Lambda \star \Lambda)$, and we can get asymptotics for both terms on the left via Bombieri's method (using partial summation for the second). However, it is much more difficult to do so with either term on the right individually, as they are supported on integers with exactly two or three prime factors respectively. Similar ideas can be used to isolate a function supported on $(\Lambda\star)^r$ and $(\Lambda\star)^s$ for any r and s of opposite parity.

10.10. **Bilinear information.** The parity problem has been overcome in some instances via the use of bilinear information. We already saw the use of bilinear bounds in our proof of the Bombieri-Vinogradov theorem, but it was only in 1995 or so that Duke, Friedlander, and Iwaniec were able to give the first instance (though works on primes in short intervals by Jutila and Iwaniec and then later by Heath-Brown are closely related) of the breaking of the parity problem in the classical setting (they showed that any irreducible quadratic polynomial has equidistribution of its roots modulo p as $p \to \infty$ — the corresponding result [for all degrees of polynomials] when averaging over all moduli is due to Hooley in the 1960s). The

work of Fouvry and Iwaniec then considered the sequence given by the number of representations of n as $n = a^2 + b^2$, and showed

$$\sum_{a^2+b^2\le x}\sum \xi_b\Lambda(a^2+b^2)=\sum_{a^2+b^2\le x}\sum \xi_b\psi(b)+O_B\left(\frac{x}{(\log x)^B}\right) \text{ where } \psi(l)=\prod_{p\nmid l}\Big(1-\frac{\chi_4(p)}{p-1}\Big).$$

In particular, by supporting ξ only on the primes, this shows that there are infinitely many primes $p = a^2 + b^2$ with b prime.

Friedlander and Iwaniec then showed that the squares were sufficiently well-behaved that one could still get an asymptotic in the above when taking ξ to be the characteristic function of the squares — thus there are infinitely many primes of the form $p = a^2 + b^4$, and was the first time that such methods had been used on a sequence whose support was significantly less than linear (here $A(x) \sim cx^{3/4}$). In the wake of this, Heath-Brown showed that there are infinitely many primes of the form $p = a^3 + 2b^3$, and this was generalised with Moroz to other cubic forms.

The bilinear estimate used by Friedlander and Iwaniec is of the form (simplified)

$$\sum_m\Big|\sum_{\substack{n\sim N\\ mn\le x}} a_{mn}\mu(n)\Big| \ll \frac{A(x)}{(\log x)^{9^9}} \quad \text{for all } N \text{ with } A(x)^{1/2+\epsilon}\le N\le x^{1/2-\epsilon}.$$

where the exact power of logarithm is largely irrelevant, and m might be weighted by a (generalised) divisor function. Looking at the inner sum, its smallness says that we expect that there should be some μ-cancellation in the support of the sequence, and furthermore that this should be true when considering the sequence restricted to suitable arithmetic progressions. There is some hope that this might hold when $A(x) \gg x^{2/3}$, but for sparser sequences the inner sum has less than one term (on average) when $N \le x/A(x)$. In the case of Friedlander and Iwaniec, the proof of this bilinear bound proceeds by re-interpreting the problem in the domain of Gaussian integers, and then achieves the estimate via a use of Cauchy's inequality and much (about 80 pages) technical work to handle divisors (of the determinant of a pair of lattice points) in various ranges (the result is ineffective [as was Fouvry-Iwaniec] due to the handling of small moduli via zero-free regions). The method does not work *mutatis mutandis* for $a^2 + b^6$ because the use of Cauchy's inequality loses too much in that case (in a recent paper called "The illusory sieve", Friedlander and Iwaniec show that $a^2 + b^6$ is infinitely often prime, but under the [unlikely] assumption of Siegel zeros). On the other hand, Heath-Brown exploits a multi-dimensional version of the large sieve to show a corresponding bilinear bound (and the result is also ineffective).