1. Selberg's sieve

1.1. **The basic sieve.** Suppose we have a set of integers $\mathcal{A}$ of some number theoretic importance, e.g. the set of primes, or the set of squarefree numbers within sum interval. A fundamental problem is to estimate the cardinality of $\mathcal{A}$. If the set is defined by the conjunction of a large of individually simple conditions, the inclusion-exclusion principle is among the first things that come to mind.

Theorem 1. *Let X be a finite set, $A_1, \dots, A_k$ be subsets of X. Then*

$$|X \setminus \bigcup_{i=1}^{k} A_i| = |X| - \sum_{i=1}^{k} |A_i| + \sum_{i<j} |A_i \cap A_j| - \dots \pm |A_1 \cap \dots \cap A_k|. \tag{1}$$

In its pure form this principle is hardly ever useful, since on the right hand side there are 2^k terms. Usually each term can be computed with an error term which is at least $\mathcal{O}(1)$, so even in the best possible case the total error becomes $\mathcal{O}(2^k)$. If we e.g. express the number of primes in $[\sqrt{x}, x]$ in this way, we would obtain $\pi(x) - \pi(\sqrt{x}) = \frac{x}{\log x} + \mathcal{O}(2^{\sqrt{2x/\log x}})$, which is worse than trivial.

However, if the size of the sets A_i and of the intersections thereof decreases rapidly, we can get a better error term by not looking at each individual term, but by neglecting whole classes of summands in one step. This method is best explained with an example.

Example 2. *The number of squarefree integers in the interval $[x, x+y]$ is $\frac{6}{\pi^2} y + (x+y)^{1/2+\varepsilon}$.*

Proof. Let X be the set of all integers in $[x, x+y]$, and A_i be the subset of all integers in $[x, x+y]$ divisible by p_i^2, where p_i is the i-th prime number. Suppose we replace the right hand side of (1) by

$$|X| - \sum_{i}^{*} |A_i| + \sum_{i<j}^{*} |A_i \cap A_j| - \dots, \tag{2}$$

wher $\sum^*$ means that the sum is restricted to terms $|A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_\ell}|$, for which $p_{i_1}^2 \cdots p_{i_\ell}^2 \le t$, and t is a parameter to be determined later. All integers in X, which are not divisible by a square $> t$ are included and excluded in such a way that in the end squarefree numbers are counted once, and numbers divisible by a square are not counted. Numbers divisible by a square $a^2 > t$ are not counted correctly. Such an integer occurs in $\tau(a)$ sets, where τ denotes the number of divisors, hence the difference between (2) and the right hand side of (1) is

$$\le \max_{a \le \sqrt{x+y}} \tau(a) \sum_{\sqrt{t} < a \le \sqrt{x+y}} \left[\frac{x+y}{a^2} \right] - \left[\frac{x}{a^2} \right] \le (x+y)^{\varepsilon} \left(\frac{y}{\sqrt{t}} + \sqrt{x+y} \right).$$

Next

$$|A_{i_1} \cap \dots A_{i_e ll}| = \left[\frac{x+y}{p_{i_1}^2 \cdots p_{i_\ell}^2} \right] - \left[\frac{x}{p_{i_1}^2 \cdots p_{i_\ell}^2} \right] = \frac{y}{p_{i_1}^2 \cdots p_{i_\ell}^2} + \mathcal{O}(1),$$

thus

$$\begin{aligned} |X| - \sum_i{}^* |A_i| + \sum_{i<j}{}^* |A_i \cap A_j| - \dots &= |X| \sum_{a \le \sqrt{t}} \frac{\mu(a)}{a^2} + \mathcal{O}(\sqrt{t}) \\ &= \frac{6}{\pi^2} y + \mathcal{O}(y \sum_{a > \sqrt{t}} \frac{1}{a^2}) + \mathcal{O}(\sqrt{t}). \end{aligned}$$

Collecting the error terms we obtain that the number of squarefree integers in $[x, x+y]$ equals

$$\frac{6}{\pi^2} y + \mathcal{O}(\frac{y^{1+\varepsilon} x^{\varepsilon}}{\sqrt{t}} + (x+y)^{1/2+\varepsilon} + t^{1/2})$$

Taking $t = y$ we see that the second error term dominates the other two, and our claim follows. □

The reason that this method works is that the series $\sum_{a \ge \sqrt{t}} \frac{y}{a^2}$ gets small as t gets large. In other words, it relies on the fact that $\sum \frac{1}{a^2}$ converges. Therefore this argument is often referred to as the converging sieve. In the same way we can count the number of consecutive square free numbers, and even prove the k-tuple conjecture for square free numbers. However, since $\sum \frac{1}{p}$ diverges, it does not tell us anything about primes.

1.2. **Brun's sieve.** To obtain information about primes one needs another method to restrict the right hand side of (1). To do so recall where the signs in (1) come from. Each integer in one of the A_i's is removed by the first sum. Integers removed to often are added again by the second sum. Integers re-added to often are again removed by the third sum, and so on. From this reasoning the following is clear.

Theorem 3. *Let X be a finite set, $A_1, \dots, A_k$ be subsets of X. Then for all even ℓ we have*

$$|X \setminus \bigcup_{i=1}^{k} A_i| \le |X| - \sum_{i=1}^{k} |A_i| + \sum_{i<j} |A_i \cap A_j| - \dots + \sum_{i_1 < i_2 < \dots < i_\ell} |A_{i_1} \cap \dots \cap A_{i_\ell}|,$$

and for all odd ℓ we have

$$|X \setminus \bigcup_{i=1}^{k} A_i| \ge |X| - \sum_{i=1}^{k} |A_i| + \sum_{i<j} |A_i \cap A_j| - \dots - \sum_{i_1 < i_2 < \dots < i_\ell} |A_{i_1} \cap \dots \cap A_{i_\ell}|,$$

In other words, when truncating (1) the error always has the same sign as the first term ommitted.

In probability theory, this statement is usually referred to as *Bonferoni inequalities*, and they are obiously not very deep. Brun's deep discovery was that they actually yield interesting number theoretic results. In fact, by taking $X = [\sqrt{x}, x]$, $A_i = \{n \in X : n(n+2) \equiv 0 \pmod{p_i}\}$ he showed that there are at most $\frac{x}{\log^{2-\varepsilon} x}$ prime twins up to x.

In this form Brun's sieve is hardly used anymore, since most of the time Selberg's sieve is technically simpler and gives superior results. However, Iwaniec has shown that the combinatorial approach by Brun can be turned into a sifting method which in many cases is better than Selberg's sieve. This sieve is commonly referred to as the Rosser-Iwaniec sieve.

1.3. **Selberg's Λ^2-method.** In both the converging sieve as in Brun's sieve we started with (1), and modified the right hand side into a shorter sum at the cost of some error. In both cases we did so by completely ignoring most of the summands. From a combinatorial point of view this is quite natural, however, from an analytic point of view using an indicator function is discontinuous, which is pretty bad. It should be better to smooth things out, that is, we are led to the following problem:

In the setting of Theorem 1, find weights $\lambda_{i_1,\dots,i_\ell}$, $1 \leq \ell \leq k$, $i_1 < i_2 < \cdots < i_\ell$, such that

$$|X| - \sum_{i=1}^{k} \lambda_i |A_i| + \sum_{i<j} \lambda_{i,j} |A_i \cap A_j| - \cdots \pm \lambda_{1,2,\dots,k} |A_1 \cap \cdots \cap A_k|$$

can be estimated with a small error and approximates the right hand side of (1) well

Clearly the two goals conflict with each other, for the first we want the weights to decrease rapidly, whereas for the second we want them to stay close to 1. Moreover, we are trying to optimize a system containing 2^k variables, which is tremendously difficult. To describe Selberg's idea of overcoming these difficulties, let us consider the following problem: Given real numbers x, y, howe many prime numbers are there in the interval $[x, x+y]$? Moreover, as can already be seen from the example of prime twins, upper bounds are easier then lower bounds, therefore we only look at the problem of finding an upper bound.

Let X be the set of integers in $[x, x+y]$, A_i be the subset of integers divisible by p_i. We would like to consider all primes p_i up to $\sqrt{x+y}$, however, if $y < \sqrt{x}$ then even the first sum on the right contains more than y terms, which means we are in trouble. We therefore consider only primes up to some bound z, which we shall determine later. Note that if we count integers without prime factors $\leq z$, we get an upper bound for the number of primes, which is what we want.

Looking at (1) from the point of view of a single element of X, which is contained in m of the sets A_i, we can rewrite this equation as

$$1 - m + \binom{m}{2} - \cdots \pm 1 = \begin{cases} 1, & m = 0 \\ 0, & m > 0 \end{cases}$$

In the same way we can rewrite the Bonferoni inequality as

$$1 - m + \binom{m}{2} - \cdots + \binom{m}{\ell} = \begin{cases} 1, & m = 0 \\ 0, & 1 \leq m \leq \ell \\ \textit{something non-negative}, & m > \ell \end{cases}.$$

Hence the Bonferoni inequality is useful, since it is close to the indicator function we want, and if it differs from what we want, then the sign of the difference is predictable. So we can use this inequality for each single element of X, add things up, and obtain an upper bound.

Applying this reasoning to our hunt for weights we see that in the concrete problem of bounding the number of primes in an interval, we have to find real numbers λ_d, such that

$$\sum_{d|n} \lambda_d = \begin{cases} 1, & n = 1 \\ \textit{something non-negative}, & n > 1 \end{cases},$$

so that we are sure to overestimate the quantity we are looking for, while at the same time $\sum_d |\lambda_d|$ is as small as possible, so that the error term coming from the rounding errors does not explode.

The first condition can be satisfied by taking real numbers Λ_d with $\Lambda_1 = 1$, and putting $\sum_{d|n} \lambda_d = \left(\sum_{d|n} \Lambda_d\right)^2$. If we further assume that $\Lambda_d = 0$ for all d which are divisible by a prime $p > z$, we find that the number of integers $n \in [x, x+y]$ which are not divisible by a prime $\leq z$ is bounded above by

$$\begin{aligned}\sum_{n=x}^{x+y} \left(\sum_{d|n} \Lambda_d\right)^2 &= \sum_{d,e} \Lambda_d \Lambda_e \left(\left[\frac{x+y}{[d,e]}\right] - \left[\frac{x}{[d,e]}\right]\right) \\ &= y \sum_{d,e} \frac{\Lambda_d \Lambda_e}{[d,e]} + \mathcal{O}\left(\left(\sum_d |\Lambda_d|\right)^2\right).\end{aligned}$$

In applications we choose the Λ_d in such a way that the error term is of smaller magnitude then the main term. The reason is that the optimal choice usually leads to expressions of the form $\frac{y}{\log^A z} + \mathcal{O}(z^2)$, thus changing z by a small factor decreases the error term significantly while leaving the main term asymptotically unchanged. This reasoning also shows that when optimizing we should focus on the main term. The standard way to do so would be via Lagrange multipliers, however, here it is actually easier.

We begin by estimating some auxiliary sums, which occur during the computation.

Lemma 4. (1) *We have* $\sum_{n\leq z} \frac{\mu^2(n)}{\varphi(n)} \geq \log z$*;*
(2) *We have* $\sum_{n\leq z} \frac{n}{\varphi(n)} \ll z$*.*

Proof. Let $\gamma(n)$ be the squarefree kernel of n, that is, the largest squarefree integer dividing n. Then we have for n squarefree

$$\frac{1}{\varphi(n)} = \frac{1}{n} \prod_{p|n} \frac{p}{p-1} = \frac{1}{n} \prod_{p|n} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \dots\right) = \sum_{\substack{m \\ s(m)=n}} \frac{1}{m}.$$

Hence $\sum_{n\leq z} \frac{1}{\varphi(n)} = \sum_{s(m)\leq z} \frac{1}{m} \geq \sum_{m\leq z} \frac{1}{m} \geq \log z$.

For the second claim we have $\frac{n}{\varphi(n)} = \prod_{p|n} \frac{p}{p-1} = \sum_{d|n} \frac{\mu^2(d)}{\varphi(d)}$. Hence

$$\sum_{n\leq z} \frac{n}{\varphi(n)} = \sum_{n\leq z} \sum_{d|n} \frac{\mu^2(d)}{\varphi(d)} = \sum_{d\leq z} \frac{\mu^2(d)}{\varphi(d)} \left[\frac{z}{d}\right] \leq z \sum_{d\leq z} \frac{\mu^2(d)}{d\varphi(d)} \ll z,$$

since the sum clearly converges. □

We have

$$\frac{1}{[d,e]} = \frac{(d,e)}{de} = \frac{1}{de} \sum_{f|(d,e)} \varphi(f),$$

thus

$$\sum_{d,e} \frac{\Lambda_d \Lambda_e}{[d,e]} = \sum_f \varphi(f) \sum_{f|d, f|e} \frac{\Lambda_d}{d} \frac{\Lambda_e}{e} = \sum_f \varphi(f) \left(\sum_{f|d} \frac{\Lambda_d}{d}\right)^2 =: \sum_f \varphi(f) y_f^2.$$

By Möbius inversion the y_f determine the Λ_d as $\Lambda_d = d\sum_{d|f} y_f\mu(f/d)$. In particular the condition $\Lambda_d = 1$ yields $\sum_f y_f\mu(f) = 1$. Putting $L(z) = \sum_{n\le z}\frac{\mu^2(n)}{\varphi(n)}$ we find that the last equation implies

$$\sum_f \varphi(f)y_f^2 = \sum_f \left(y_f - \frac{\mu(f)}{\varphi(f)L(z)}\right)^2 + \frac{1}{L(z)}.$$

Clearly this expression is minimzed by taking $y_f = \frac{\mu(f)}{\varphi(f)L(z)}$, and for this choice we have in fact

$$\sum_f y_f\mu(f) = \frac{1}{L(z)}\sum_f \frac{\mu^2(f)}{\varphi(f)} = 1,$$

provided that the y_f are defined for f up to z. With this choice for the Λ_d the main term becomes $\frac{y}{L(z)}$.

We now turn to the error term. We have

$$\Lambda_d = d\sum_{\substack{d|f\\ f\le z}} y_f\mu(f/d) = d\sum_{\substack{d|f\\ f\le z}} \frac{\mu(f)\mu(f/d)}{\varphi(f)L(z)} = \frac{d\mu(d)}{\varphi(d)L(z)}\sum_{\substack{(t,d)=1\\ t\le z/d}}\frac{\mu(t)^2}{\varphi(t)}.$$

Using Lemma 4 (2) we conclude

$$\sum_{d\le z}|\Lambda_d| \le \sum_{d\le z}\frac{d}{\varphi(d)L(z)}\sum_{t\le z/d}\frac{1}{\varphi(t)} = \frac{1}{L(z)}\sum_{t\le z}\frac{1}{\varphi(t)}\sum_{d\le z/t}\frac{d}{\varphi(d)} \le \frac{1}{L(z)}\sum_{t\le z}\frac{z}{t\varphi(t)} \ll \frac{z}{L(z)}.$$

Putting things together we obtain that the number of integers $n \in [x, x+y]$, which are not divisible by any prime number $p < z$ is $\frac{y}{L(z)} + \mathcal{O}(\frac{z^2}{L(z)^2}$. Using Lemma 4 (1) we obtain that this number is bounded above by $\frac{y}{\log z} + \mathcal{O}(\frac{z^2}{\log^2 z})$. The error term becomes negligible for $z = \sqrt{y}$, in particular we have proven that the number of primes in $[x, x+y]$ is bounded above by $\frac{(2+o(1))y}{\log y}$.

1.4. **The number of prime twins below x.**

2. The large sieve

2.1. **Some linear algebra.** Let $(V, \langle\cdot,\cdot\rangle)$ be a hermitian vector space, and let $\phi_1, \dots, \phi_r$ be an orthonormal family of elements of V. Then the abstract Fourier expansion yields for every $\xi \in V$ the inequality $\sum_{i=1}^r |\langle\xi, \phi_i\rangle|^2 \le \|\xi\|^2$. A quite natural problem, apparently first considered in the context of functional analysis, is the question whether we can weaken the condition of orthonormality. Boas proved the following.

Theorem 5. *Let $(V, \langle\cdot,\cdot\rangle)$ be a hermitian vector space, and $\phi_1, \dots, \phi_r, \xi$ be elements of V. Then we have*

$$\sum_{i=1}^r |\langle\xi, \phi_i\rangle|^2 \le \max_{1\le i\le r}\sum_{j=1}^r |\langle\phi_i, \phi_j\rangle| \cdot \|\xi\|^2. \tag{3}$$

Proof. Put $A = \max_{1\le i\le r}\sum_{j=1}^r |\langle\phi_i, \phi_j\rangle|$. We first show that for any choice of the complex numbers u_i, $1 \le i \le r$, we have

$$\sum_{1\le i,j\le r} u_i\overline{u_j}\langle\phi_i, \phi_j\rangle \le A\sum_{i=1}^r |u_i|^2.$$

In fact, using the inequality between the quadratic and geometric mean the left hand side is bounded above by

$$\sum_{1\le i,j\le r} \frac{|u_i|^2+|u_j|^2}{2}\langle \phi_i,\phi_j\rangle = \sum_{i=1}^r |u_i|^2 \sum_{j=1}^r \langle \phi_i,\phi_j\rangle \le A\sum_{i=1}^r |u_i|^2.$$

We now deduce (3). Many of the elementary inequalities are obtained by manipulating the obvious inequality $x^2 \ge 0$. For example, the inequality $\frac{a^2+b^2}{2} \ge ab$ immediately follows from $(a-b)^2 \ge 0$. In the context of hermitian spaces this inequality is replaced by $\|\xi\|^2$, which motivates the first inequality. The remainder of the proof is just a calculation. For any choice of complex numbers u_i we have

$$\begin{aligned} 0 \le \left\| \xi - \sum_{i=1}^r u_i\phi \right\|^2 &= \|\xi\|^2 - 2\sum_{i=1}^r u_i\langle \xi,\phi_i\rangle + \sum_{1\le i,j\le r} u_i\overline{u_j}\langle \phi_i,\phi_j\rangle \\ &\le \|\xi\|^2 - 2\sum_{i=1}^r u_i\langle \xi,\phi_i\rangle + A\sum_{i=1}^r |u_i|^2. \end{aligned}$$

We now put $u_i = \frac{\overline{\langle \xi,\phi_i\rangle}}{A}$ and obtain

$$0 \le \|\xi\|^2 - \frac{1}{A}\sum_{i=1}^r |\langle \xi,\phi_i\rangle|^2,$$

which is our claim. □

The structure of the proof is worth being remembered. Sometimes one sets of from $x^2 \ge 0$ like here, but more often one first uses Cauchy-Schwarz inequality to separate a sum into an analytic and a number theoretic part. In any case one ends up with squares, which one expands. The resulting expansion is often so complicated that instead of evaluating the concrete example one is interested in one resorts to general inequalities for bilinear forms. We shall see more examples of this technique later.

2.2. **Passing to number theory.** In number theoretic applications the vector ξ encodes the arithmetic object we want to study, while each ϕ_i is one aspect of this object. For example, ξ could be the indicator function of the set of primes in an interval, while the ϕ_i are indicator functions of arithmetic progressions. In this way the large sieve was applied by Renyi. The problem with this approach is that the indicator functions of arithmetic progressions are far for orthogonal. Roth discovered that exponential sums give better orthogonality. Write $e(t) = e^{2\pi i t}$. Suppose that $\xi = (a_1, a_2, \dots, a_N)$, and let $t_1, \dots, t_r \in [0,1]$ be real numbers satisfying $\min_{i,j} |t_i - t_j| = \delta$. Here we identify $[0,1]$ with the circle , thus the distance of 0.99 and 0.01 is 0.02. Put $\phi_i = (e(t_i), e(2t_i), \dots, e(Nt_i))$. Then we have

$$|\langle \phi_i,\phi_j\rangle| = \left| \sum_{n=1}^N e((t_i-t_j)n) \right| \le \min\left(N, \frac{1}{|t_i-t_j|}\right),$$

thus

$$\max_i \sum_j |\langle \phi_i,\phi_j\rangle| \le N + \sum_{k=1}^{r/2} \frac{2}{\delta_k} \le N + \delta^{-1}\log r.$$

If we let the t_i be the list of all rational numbers in $[0,1]$ with denominator $\leq Q$, then for $t_i \neq t_j$ we have

$$\left|\frac{a_1}{q_1} - \frac{a_2}{q_2}\right| \geq \frac{1}{q_1 q_2} \geq \frac{1}{Q^2},$$

hence we obtain

$$\sum_{q=1}^{Q} \sum_{\substack{1\leq a\leq q\\(a,q)=1}} \left|\sum_{n=1}^{N} a_n e\left(\frac{an}{q}\right)\right|^2 \leq \left(N + 2Q^2 \log Q\right) \sum_{n=1}^{N} |a_n|^2.$$

By choosing the vectors ϕ_i a more careful, Selberg proved the following.

Theorem 6. *For integers N, Q and complex numbers $a_1, \ldots, a_n$ we have*

$$\sum_{q=1}^{Q} \sum_{\substack{1\leq a\leq q\\(a,q)=1}} \left|\sum_{n=1}^{N} a_n e\left(\frac{an}{q}\right)\right|^2 \leq \left(N + Q^2\right) \sum_{n=1}^{N} |a_n|^2. \tag{4}$$

2.3. **Reformulations of the large sieve inequality.** Let $\mathcal{N} \subseteq [1,N]$ be a set of integers, and put $a_n = 1 \Leftrightarrow n \in \mathcal{N}$. Then (4) implies that the distribution of $\mathcal{N}$ in residue classes modulo q cannot be too wild for too many q. Making this precise is difficult, because e.g. the exponential sum $\sum_{n} \in \mathcal{N} e(n/2)$ influences the distribution of $\mathcal{N}$ modulo all even modules. This problem can be avoided by restricting ourselves to the case q prime. We have the following.

Theorem 7. *Let $\mathcal{N}$ be a set of integers. Then we have*

$$\sum_{p\leq Q}{}^{*} p \sum_{a=1}^{q} \left|\#\{n \in \mathcal{N} : n \equiv a \pmod p\} - \frac{|\mathcal{N}|}{p}\right|^2 \leq (N + Q^2)|\mathcal{N}|,$$

where $\sum^{*}$ denotes summation over prime numbers only.

Proof. Put $S(\alpha) = \sum_{n\in\mathcal{N}} e(n\alpha)$, and $\mathcal{N}(p,a) = \#\{(n,m) \in \mathcal{N}^2 : n \equiv m \pmod p\}$. Then

$$\begin{aligned}
\sum_{a=1}^{p} \left|S(\frac{a}{p})\right|^2 &= \sum_{n,m\in\mathcal{N}} \sum_{a=1}^{q} e\left(\frac{a(n-m)}{p}\right) \\
&= p\#\{(n,m) \in \mathcal{N}^2 : n \equiv m \pmod p\} \\
&= p\sum_{a=1}^{p} \mathcal{N}(p,a)^2
\end{aligned}$$

Since $S(0) = |\mathcal{N}|$, we deduce

$$\begin{aligned}
p\sum_{a=1}^{p} \left|\mathcal{N}(p,a) - \frac{|\mathcal{N}|}{p}\right|^2 &= p\sum_{a=1}^{p} \mathcal{N}(p,a)^2 - 2\mathcal{N}\sum_{a=1}^{p} \mathcal{N}(p,a) + |\mathcal{N}|^2 \\
&= \sum_{a=1}^{p} \left|S\left(\frac{a}{p}\right)\right|^2 - S(0)^2 \\
&= \sum_{a=1}^{p-1} \left|S\left(\frac{a}{p}\right)\right|^2
\end{aligned}$$

Theorem 7 now follows from Theorem 6. □

We can also reformulate the large sieve in terms of characters.

Theorem 8. *Let $a_1, \dots, a_N$ be complex numbers. Then we have*

$$\sum_{q \le Q} \frac{q}{\varphi(q)} \sum_{\chi \pmod q}{}^{*} \left| \sum_{n=1}^{N} a_n \chi(n) \right|^2 \le (N + Q^2) \sum_{n=1}^{N} |a_n|^2,$$

where $\sum^*$ denotes summation over primitive characters only.

Proof. For χ primitive we have

$$\chi(n) = \frac{1}{\tau(\overline{\chi})} \sum_{a=1}^{q} \overline{\chi}(a) e\left(\frac{an}{q}\right),$$

thus

$$\sum_{n=1}^{N} a_n \chi(n) = \frac{1}{\tau(\overline{\chi})} \sum_{a=1}^{q} \overline{\chi}(a) \sum_{n=1}^{N} a_n e\left(\frac{an}{q}\right).$$

Using $|\tau(\chi)| = \sqrt{q}$ and putting $S(\alpha) = \sum_{n=1}^{N} a_n e(n\alpha)$ we obtain

$$\sum_{\chi \pmod q}{}^{*} \left| \sum_{n=1}^{N} a_n \chi(n) \right|^2 = \frac{1}{q} \sum_{\chi \pmod q}{}^{*} \left| \sum_{a=1}^{q} \overline{\chi}(a) S\left(\frac{a}{q}\right) \right|^2.$$

If we replace the sum on the right by a sum over all characters modulo q, we obtain an inequality, expanding the square we obtain

$$\begin{aligned} \sum_{\chi \pmod q}{}^{*} \left| \sum_{n=1}^{N} a_n \chi(n) \right|^2 &= \frac{1}{q} \sum_{1 \le a,b \le q} S\left(\frac{a}{q}\right) \overline{S\left(\frac{b}{q}\right)} \sum_{\chi \pmod q} \overline{\chi}(a)\chi(b) \\ &= \frac{\varphi(q)}{q} \sum_{(a,q)=1} \left| S\left(\frac{a}{q}\right) \right|^2. \end{aligned}$$

Inserting this bound into (4) our claim follows. □

The basic philosophy of the reformulations is that the space of functions $f : \mathbb{Z}/q\mathbb{Z} \to \mathbb{C}$ has three natural bases: The exponential functions, the character functions, and the δ-functions. Quite often exponential sums are the easiest basis to work with, but for the final results we are usually more interested in arithmetics progressions or characters.

In general the reformulation in terms of characters is pretty optimal. However, if one has some information on the support of the sequence a_n or if only a certain set of modules q is interesting, the reader should not try to apply Theorem 6 directly, but rather go back to Theorem 5 and use the extra information there. As an example, we prove the following, which is known as Halász inequality.

Theorem 9. *Let $\chi_1, \dots, \chi_r$ be distinct primitive character to modules $\le Q$, and let $a_1, \dots, a_N$ be copmlex numbers. Then we have*

$$\sum_{i=1}^{r} \left| \sum_{n=1}^{N} a_n \chi_i(n) \right|^2 \le (N + 2rQ \log Q) \sum_{n=1}^{N} |a_n|^2.$$

Proof. In view of Theorem 5 our aim is to compute $\max_i \sum_{j=1}^{r} \sum_{n=1}^{N} \chi_i(n)\overline{\chi_j}(n)$. If $\chi_i = \chi_j$, then $\chi_i\overline{\chi}_j$ is a principal character, while for $\chi \not= \chi_j$ we have that $\chi_i\overline{\chi_j}$ is a non-principal character to a module $\leq Q^2$. The first case contributes at most N, while for the second we use the Polya-Vinogradov-inequality to see that each term contributes at most $2Q\log Q$. Inserting these values our claim follows. □

Theorem 9 is superior to Theorem 8, if R is somewhat smaller than Q. Hence it is quite suitable to deal with a small set of really bad characters. For this reason Halász inequality is of great importance in the problem of bounding the number of zeros of L-series far off the critical line.

2.4. **Applications of the large sieve.** As first application we prove the following, which is due to Linnik. For a prime number p let $n(p)$ be the least quadratic non-residue modulo p. The best known estimates is $n(p) \ll p^{1/(4\sqrt{e}}\log p$, which follows by combining Burgess estimates for character sums with an elementary trick due to Vinogradov. Since Burgess estimates in turn depend on Weil's bounds for the number of points on algebraic curves, this result is pretty deep, still it is far from the expected bound $n(p) < \log^{1+\varepsilon} p$. Linnik showed that at least on average we can do a lot better.

Theorem 10. *For every $c > 0$ we have that the number of primes $p < x$ with $n(p) > p^c$ is $\mathcal{O}(\log\log x)$.*

For the proof we shall use the following result without proof.

Lemma 11. *For every $c > 0$ there exists some $\rho(c) > 0$, such that the number of integers $n \leq x$, which do not have a prime factor $\geq x^c$, is $\geq (\rho(c) + o(1))x$.*

Proof of Theorem 10. For $1 \leq n \leq N$ put $a_n = 1$, if all prime divisors of n are $\leq x^c$, and $a_n = 0$ otherwise. Let $p_1, \dots, p_r$ be the list of all prime numbers $p \leq Q$ with $n(p) > Q^c$, and let χ_i be the quadratic character modulo p_i. Since χ_i is multiplicative and equal to 1 for all $n \leq x^c$, we find that $\chi_i(n) = 1$ whenever $a_n = 1$, thus

$$\sum_{n\leq N} a_n\chi_i(n) = \sum_{n\leq N} a_n \geq \rho(c)N.$$

Inserting this into Theorem 9 we obtain

$$r\rho(c)^2N^2 \leq (N + 2rQ\log Q)\rho(c)N.$$

Since the parameter r appears on both sides of this inequality, we have to pick Q and N in such a way that the coefficient of r on the right is smaller than on the left. For example, we can take $Q = \frac{\rho(c)N}{4\log N}$, and obtain $r\rho(c)^2N^2 \leq 2\rho(c)N^2$, that is, $r \leq \frac{2}{\rho(c)}$.

We conclude that the number of primes $p \leq Q$ with $n(p) > Q^c$ is bounded independently of Q. Hence the number of primes $p \in [\sqrt{Q}, Q]$ with $n(p) > p^{2c}$ is also bounded, glueing together intervals of the form $[x, x^2]$ our claim follows. □

We could have proven the same result using Theorem 8 in place of Theorem 9. However, since in this case r is very small, Theorem 9 gives smaller constants. On the other hand one sees that for a sub-optimal choice of the paramters N and Q it might happen that Theorem 9 yields a trivial result, while Theorem 8 is more robust in this respect.

This application also explains where the term *large sieve* comes from. For each prime p with $n(p)$ large we have that the set of integers which has no large prime divisors is restricted to $\frac{p+1}{2}$ residue classes modulo p. We then compare the lower bound coming from Lemma 11 with an upper bound which is valid for any set of integers which is restricted to $\frac{p+1}{2}$ residue classes modulo p for many integers p. Selberg's sieve could not handle the removal of such a large number of classes, while the large sieve can.

A possibly more important difference between Selberg's sieve and the large sieve is the fact that while the former sets of with a set of integer, the latter is an analytic estimate for exponential sums or character sums. We can therefore apply the large sieve even if we are not sifting anything. We shall do so in the proof of the Bombieri-Vinogradov theorem, and this technique is central to studying the distribution of zeros of L-series. A much simpler application is the following, which is known as the Barban-Davenport-Hapberstam theorem.

Theorem 12. *For $\frac{x}{\log^A} x \leq Q \leq x$ we have*

$$\sum_{q \leq Q} \sum_{(a,q)=1} \left(\Psi(x,q,a) - \frac{x}{\varphi(q)} \right)^2 \ll Qx \log x. \tag{5}$$

Proof. Taking $a_n = \Lambda(n)$ in Theorem 8 we obtain

$$\sum_{q \leq Q} \frac{q}{\varphi(q)} {\sum_{\chi \pmod q}}^* |\Psi(x,\chi)|^2 \ll (x+Q^2) x \log x. \tag{6}$$

To deduce (5) from this bound, we write

$$\Psi(x,q,a) - \frac{x}{\varphi(q)} = \frac{1}{\varphi(q)} \sum_{\chi \neq \chi_0} \overline{\chi}(a) \Psi(x,\chi) + \frac{\Psi(x)-x}{\varphi(q)}.$$

To simplify notation we put $\Psi'(x,\chi) = \begin{cases} \Psi(x,\chi) - x, & \chi \text{ principal} \\ \Psi(x,\chi), & \chi \text{ not principal} \end{cases}$. If we insert our expression fo $\Psi(x,q,a)$ into (5), we cannot directly apply (6), because the latter inequality runs over primitive characters only. Instead we obtain for a single module q the equation

$$\sum_{(a,q)=1} \left(\Psi(x,q,a) - \frac{x}{\varphi(q)} \right)^2 = \frac{1}{\varphi(q)} \sum_{\chi \pmod q} |\Psi'(x,\chi)|^2$$

If the character $\chi \pmod q$ is induced by the primitive character $\chi' \pmod{q'}$, then $\Psi(x,\chi)$ differs from $\Psi(x,\chi')$ only in not counting prime powers which are coprime q', but not coprime q. Each prime divisor of q' induces $\ll \log x$ prime powers, hence we have $\Psi(x,\chi) - \Psi(x,\chi') \ll \log^2 Qx$. We conclude

$$\begin{aligned} \sum_{q \leq Q} \sum_{(a,q)=1} \left(\Psi(x,q,a) - \frac{x}{\varphi(q)} \right)^2 &= \sum_{q \leq Q} \frac{1}{\varphi(q)} \sum_{\chi \pmod q} |\Psi'(x,\chi')|^2 + \mathcal{O}(Q\sqrt{x} \log^3 Qx) \\ &= \sum_{q \leq Q} \sum_{\substack{d \leq Q/q \\ (d,q)=1}} \frac{1}{\varphi(dq)} {\sum_{\chi \pmod q}}^* |\Psi'(x,\chi')|^2 + \mathcal{O}(Q\sqrt{x} \log^3 Qx). \end{aligned}$$

Since φ is multiplicative, and $\sum_{n\le z}\frac{1}{\varphi(n)} \ll 1+\log z$, we find that it suffices to prove

$$\sum_{q\le Q}\frac{1+\log Q/q}{\varphi(q)}\sum_{\chi \pmod q}\left|\Psi'(x,\chi')\right|^2 \ll Qx\log x. \tag{7}$$

If we restrict the range for q to $[U,2U]$, we have

$$\begin{aligned}\sum_{U\le q\le 2U}\frac{1+\log Q/q}{\varphi(q)}\sum_{\chi \pmod q}\left|\Psi'(x,\chi')\right|^2 &\ll \frac{1+\log Q/U}{U}\sum_{U\le q\le 2U}\frac{q}{\varphi(q)}\sum_{\chi \pmod q}\left|\Psi'(x,\chi')\right|^2\\ &\ll (1+\log Q/U)(\frac{x^2}{U}+xU)\log x\end{aligned}$$

Summing things up we find that our estimate covers the range $[\frac{x\log x}{Q}, Q]$, but fails to prove anything for the smallest q.

For the smallest q we do not use the large sieve, but the Siegel-Walfisz theorem. For $q<\log^{A+1}x$ we have $\Psi'(x,\chi')\ll xe^{-c\sqrt{\log x}}$, and we find that the sum over small q is negligible. Hence we have established (7), and the proof is complete. □

3. Vaughan's identity

3.1. **The identity.** Suppose $f:\mathbb{N}\to\mathbb{C}$ is a function which is highly oscillating, and we want to estimate $\sum_{p\le x}f(p)$. Let P be the product of all prime numbers $\le\sqrt{x}$. Then we have

$$f(1)+\sum_{\sqrt{x}<p\le x}f(p)=\sum_{\substack{n\le x\\(n,P)=1}}f(n)=\sum_{\substack{t|P\\ t\le x}}\sum_{d\le x/t}f(dt).$$

If f is heavily oscillating we may expect that a sum of the form $\sum f(dt)$ can be estimated in a non-trivial way, and such a saving would carry over to a non-trivial bound for the whole sum over primes. However, in the present form this approach does not work, since in most terms of the first sum t is not much smaller than x, so the second sum contains most of the time only a few terms, and for a sum e.g. of length 3 we cannot expect significant cancelation. Vinogradov managed to rearrange these terms in such a way that they are covered by longer progressions, however, his method was extremely complicated. In 1977 Vaughan found a much simpler way of rearranging terms, which since became a standard way of evaluating sums over primes. We begin by the following simple fact.

Lemma 13. *Let F,G be meromorphic functions. Then we have*

$$-\frac{\zeta'}{\zeta}=F-\zeta FG-\zeta' G+\left(-\frac{\zeta'}{\zeta}-F\right)(1-\zeta G).$$

Proof. Just expand the product. □

The idea of Vaughan is the following: On the left hand side we have something connected to primes, while the right hand side contains 4 terms, three of which are simple. The last one is a complicated product, so we choose F and G in such a way that this last term becomes small. Hence F should be an approximation to $-\frac{\zeta'}{\zeta}$, and G should be an approximation to $\frac{1}{\zeta}$. We take the simplest choices, defining $F(s)=\sum_{n\le U}\Lambda(n)n^{-s}$, $G(s)=\sum_{n\le V}\mu(n)n^{-s}$. The parameters U,V are optimized depending on the application at the very end of the argument. In the

case mentioned above, where we want to estimate $\sum_{N\le n\le 2N}\Lambda(n)f(n)$, U and V are usually small powers of N.

Inserting these Dirichlet series into the identity we obtain the following.

Theorem 14. *For integers U, V we have*

$$\Lambda(n) = a_1(n) + a_2(n) + a_3(n) + a_4(n),$$

where

$$\begin{aligned}
a_1(n) &= \begin{cases}\Lambda(n), & n\le U\\ 0, & n>U\end{cases}\\
a_2(n) &= -\sum_{\substack{mdr=n\\ m\le U, d\le V}}\Lambda(m)\mu(d)\\
a_3(n) &= \sum_{\substack{dm=n\\ m\le V}}\mu(m)\log d\\
a_4(n) &= -\sum_{\substack{mk=n\\ m>U, k\neq 1}}\Lambda(m)\left(\sum_{\substack{d|k\\ d\le V}}\mu(d)\right)
\end{aligned}$$

Why this decomposition is helpful probably needs some explication. When interchanging the order of summation, $\sum_{n\le x}\sum_{d|n}$ becomes $\sum_d\sum_{d|n,n\le x}$. Such a sum is bad, since for d close to x the inner sum is too short to be estimated. However, if one has an upper bound for d, which is significantly smaller than x, the inner sum remains long and might be estimated in a good way. The reader should recall Dirichlet's hyperbola trick for the first application of this method.

3.2. **Application of the identity: The general case.** Suppose we want to estimate the sum $\sum_{n\le N}\Lambda(n)f(n)$. Then using Theorem 14 we see that we have to handle the 4 sums $S_i = \sum a_i(n)f(n)$. We now describe the strategy of doing so. We assume that the function f is well understood, bounded (or at least not too large), and heavily oscillating, so that sums of the form $\sum_n f(rn)$ are small. We further assume that U and V are neither very small nor very close to N, in particular, S_1 is always negligible.

In each step we want to separate a sum containing only f from the number theoretic functions μ and Λ. The latter will be estimated trivially. Write S_2 as

$$S_2 = -\sum_{t\le UV}\left(\sum_{\substack{md=t\\ m\le U, d\le V}}\mu_d\Lambda(t)\right)\sum_{r\le N/t}f(rt) \le \log UV\sum_{t\le UV}\left|\sum_{r\le N/t}f(rt)\right|.$$

If there is no cancelation in the inner sum, and f is bounded, then this becomes $\ll N\log^2 N$, so essentially we did not lose anything. The length of the inner sum is $\ge N/UV$, since we can hope for a saving of magnitude a small power of the length of the sum, this is fine.

The next sum is

$$S_3 = \sum_{d\le V}\mu(d)\sum_{h\le N/d} f(dh)\log h = \sum_{d\le V}\sum_{h\le N/d} f(N/d)\int_1^h \frac{dt}{t}$$
$$= \int_1^N \sum_{d\le V}\mu(d)\sum_{w\le h\le N/d} f(dh)\frac{dt}{t} \ll \log N \sum_{d\le V}\max_w \left|\sum_{w\le h\le N/d} f(dh)\right|$$

Note that for many choices of f a sum of the form $\sum f(dh)\log h$ is not more complicated than a sum of the form $\sum f(dh)$, however, since we have to deal with the latter anyway expressing the logarithm via an integral saves some time.

The sum S_4 is more complicated. To estimate it we use a trick we have already seen in the proof of Theorem 5: If we have coefficients like μ and Λ we do not understand, we estimate a bilinear form for arbitrary coefficients. Suppose for M, N fixed that Δ is a real number such that for all complex numbers b_m, c_k we have

$$\left|\sum_{M<m\le 2M} b_m \sum_{V<k\le N/M} c_k f(mk)\right| \le \Delta\left(\sum_{m=M}^{2M}|b_m|^2\right)^{1/2}\left(\sum_{k\le N/M}|c_k|^2\right)^{1/2}.$$

Then we have

$$\begin{aligned} S_4 &= \sum_{U<m\le N/V}\Lambda(m)\sum_{V<k\le N/m}\left(\sum_{\substack{d|k\\ d\le V}}\mu(d)\right) f(mk) \\ &\le \log N \max_{U\le M\le N/V}\Delta\left(\sum_{m=M}^{2M}\Lambda(m)^2\right)^{1/2}\left(\sum_{k\le N/M} d(k)^2\right)^{1/2} \\ &\le N^{1/2}\log^3 N \max_{U\le M\le N/V}\Delta. \end{aligned}$$

There is no general method for giving bounds for Δ. In particular, if f is completely multiplicative we have

$$\sum_{M<m\le 2M} b_m \sum_{V<k\le N/M} c_k f(mk) = \sum_{M<m\le 2M} b_m f(m) \sum_{V<k\le N/M} c_k f(k),$$

thus taking $b_m = \overline{f(m)}$, $c_k = \overline{f(k)}$ we see that $\Delta = \sum_{M<m\le 2M}\sum_{V<k\le N/M} 1 = N - MV$, that is, there is no non-trivial estimate. In particular, we cannot use Vaughan's identity to bound $\Psi(x,\chi)$.

If we apply Cauchy-Schwarz, we obtain

$$\left|\sum_{M<m\le 2M} b_m \sum_{V<k\le N/M} c_k f(mk)\right| \le \left(\sum_{m=M}^{2M}|b_m|^2\right)^{1/2}\left(\sum_{m=M}^{2M}\left|\sum_{V<k\le N/M} c_k f(mk)\right|^2\right)^{1/2}.$$

The first sum is what we want, the second we write as

$$\sum_{V<j\le N/M} c_j \sum_{V<k\le N/M} \overline{c_k} \sum_{\substack{M<m\le 2M\\ m\le \min(N/j,N/k)}} f(mj)\overline{f(mk)}$$

$$\le \sum_{V<j\le N/M} |c_j|^2 \sum_{V<k\le N/M} \left| \sum_{\substack{M<m\le 2M\\ m\le \min(N/j,N/k)}} f(mj)\overline{f(mk)} \right|$$

We conclude that

$$\Delta \le \left(\max_{V<j\le N/M} \sum_{V<k\le N/M} \left| \sum_{\substack{M<m\le 2M\\ m\le \min(N/j,N/k)}} f(mj)\overline{f(mk)} \right| \right)^{1/2}.$$

If f is bounded, and we bound the inner sum trivially, we obtain $\Delta \ll N^{1/2}$, so to obtain a non-trivial result in the end a little cancelation in the inner sum suffices.

Collecting the estimates for S_1–S_4 we obtain the following.

Theorem 15. *Suppose that $|f(n)| \le 1$, $U, V \ge 2$, $UV \le N$. Then we have*

$$\sum_{n\le N} f(n)\Lambda(n) \ll U + \log N \sum_{t\le UV} \max_w \left| \sum_{w\le r\le N/t} f(rt) \right|$$

$$+ N^{1/2}\log^3 N \max_{M\le M\le N/V} \max_{V\le j\le N/M} \left(\sum_{V<k\le N/M} \left| \sum_{\substack{M<m\le 2M\\ m\le \min(N/k,N/j)}} f(kj)\overline{f(mk)} \right| \right)^{1/2}$$

Quite often the sums $\sum f(rt)$ are referred to as type I sums, and the sums $\sum f(mj)\overline{f(mk)}$ as type II sums. Obviously type II sums are more complicated than type I sums.

3.3. **Application of the identity: An example.** Vinogradov first used his method to bound the sum $\sum_{n\le N} \Lambda(n)e(n\alpha)$. The estimation of this sum is the crucial part in proving the ternary Goldbach problem. Hardy and Ramanujan had expressed this sum using characters and used the generalized Riemann hypothesis to bound the resulting character sums, in this way they showed that GRH implies that every large odd integer is the sum of 3 prime numbers. Vinogradov gave the first unconditional proof of this result. Helfgott's proof that this statement is actually true for all integers follows the same lines, however, for small N a large power of $\log N$ is no longer negligible when compared to a small power of N, which makes the argument a lot more difficult. Here we use Vaughan's identity to prove the following.

Theorem 16. *Let $\alpha \in [0,1]$ be a real number, a, q be integers with $(a,q) = 1$ and $|\alpha - \frac{a}{q}| \le \frac{1}{q^2}$. Then we have*

$$\sum_{n\le N} \Lambda(n)e(n\alpha) \ll \log^4 N \left(Nq^{-1/2} + N^{4/5} + N^{1/2}q^{1/2} \right)$$

Note that this estimate is non-trivial for $q > \log^9 N$, and if α is very close to $\frac{a}{q}$ with q very small, we can use the Siegel-Walfisz theorem to approximately compute $S(\alpha)$.

We begin by estimating the type I sums.

Lemma 17. *We have*

$$\sum_{t\le T}\max_{w}\left|\sum_{w\le r\le N/t} e(rt\alpha)\right| \ll \left(\frac{N}{q}+T+q\right)\log 2qT.$$

Proof. We have $\sum_{n=N_1}^{N_2} e(n\beta) \ll \min(N_1-N_2, \frac{1}{\|\beta\|})$, where $\|\cdot\|$ denotes the minimal distance to the nearest integer. Hence

$$\sum_{t\le T}\max_{w}\left|\sum_{w\le r\le N/t} e(rt\alpha)\right| \ll \sum_{t\le T}\min\left(\frac{N}{t}, \frac{1}{\|t\alpha\|}\right)$$

Among $q/2$ consecutive values of t, there are at most k with $\|t\alpha\| < k/4$, hence, if we exclude the value closest to an integer, the remaining terms contribute at most $\mathcal{O}(q\log q)$ to the sum. We can cover $[1,T]$ by $\ll \frac{T}{q}+1$ such intervals, hence if we forget about one summand among $q/2$ consecutive ones, we can estimate the sum by $\mathcal{O}((T+q)\log q)$. Moreover, in the first interval there is no need for an exclusion. For the missing terms we use the bound $\frac{N}{t}$, the contribution of these terms becomes $\ll \sum_{d\le T/q}\frac{N}{dq/2+1} \ll \frac{N}{q}\log T$, and our claim follows. □

Hence Theorem 15 implies that

$$\sum_{n\le N}\Lambda(n)e(n\alpha) \ll \left(\frac{N}{q}+UV+q\right)\log^2 2qN$$
$$+ N^{1/2}\log^3 N \max_{U\le M\le N/V}\max_{V\le j\le N/M}\left(\sum_{V<k\le N/M}\min\left(M, \frac{1}{\|(k-j)\alpha\|}\right)\right)^{1/2}.$$

The last term is essentially independent of j, and we obtain that the second summand is bounded above by

$$N^{1/2}\log^3 N \max_{U\le M\le N/V}\left(M+\sum_{1\le m\le N/M}\min\left(\frac{N}{m}, \frac{1}{\|m\alpha\|}\right)\right)^{1/2}$$
$$\ll N^{1/2}\log^3 N \max_{U\le M\le N/V}\left(M+\frac{N}{M}+\frac{N}{q}+q\right)^{1/2}\log^{1/2} qN.$$

We may assume that $q < N$, since otherwise our estimate is trivial anyway, and we can take the maximum and the square root componentwise. Hence we obtain that the sum in question is bounded by

$$\left(UV+q+NU^{-1/2}+NV^{-1/2}+Nq^{-1/2}+(Nq)^{1/2}\right)\log^4 N.$$

Choosing $U = V = N^{2/5}$ makes the first, third and fourth term equal to $N^{4/5}$, and our claim follows.

4. The Bombieri-Vinogradov theorem

4.1. **The theorem.** The generalized Riemann hyptohesis is equivalent to the statement that for integers q, a with $(q, a) = 1$ we have $\Psi(x, q, a) = \frac{x}{\varphi(q)} + \mathcal{O}(x^{1/2+\varepsilon})$. For many applications of GRH it is not the quality of the error term that is important, but the fact that we have an asymptotic for x as small as $q^{2+\varepsilon}$. The best unconditional result known is the Siegel-Walfisz theorem, which applies for $q < \log^A x$ and $x > x_0(A)$. This result has not been improved upon for 80 years, in particular, there is still no effective version known. However, although we do not know much more concerning the non-existence of zeroes of L-series off the critical line, we do know that such zeroes are rare. At first such density results were just seen as lending credibility to the Riemann hypothesis, but Hoheisel showed that density estimates fo the ζ-function can be used to prove results about primes in short intervals. Turán showed that density estimates for L-series are in a similar way connected to the problem of estimating $\Psi(x, q, a)$ for relatively small values of x. Then Linnik proved zero-density estimates which were powerful enough to show that there exists a constant C, such that for all q, a with $(q, a) = 1$ there exists a prime number $p \equiv a \pmod q$ and $p \leq q^C$. After the development of the large sieve, Bombieri gave a density estimate, which allowed him to prove what is now known as the Bombieri-Vinogradov theorem. Later Vaughan discovered his identity and showed that it can be used to give a simple proof of the Bombieri-Vinogradov theorem.

Theorem 18. *For any fixed A we have*

$$\sum_{q \leq Q} \max_{y \leq x} \max_{a:(q,a)=1} \left| \Psi(y, q, a) - \frac{y}{\varphi(q)} \right| \ll x^{1/2} Q \log^5 x$$

uniformly in $\frac{x^{1/2}}{\log^A x} \leq Q \leq x^{1/2}$.

By the Brun-Titchmarsh inequality we have $\Psi(y, q, a) \ll \frac{y}{\varphi(q)}$, hence the left hand side of the inequality is trivially bounded above by $\sum_{q \leq Q} \frac{x}{\varphi(q)} \ll x \log Q$, that is, te Bombierei-Vinogradov theorem is non-trivial by some power of $\log x$. The formulation of the left hand side involving a double maximum might look a little strange, but it implies some flexibility which is useful in applications.

4.2. **Proof part I: An estimate for character sums.** The main part of the proof of Theorm 18 is the following.

Theorem 19. *We have for all $x, Q \geq 1$ the estimate*

$$\sum_{q \leq Q} \frac{q}{\varphi(q)} \sum_{\chi \pmod q}^{*} \max_{y \leq x} |\Psi(x, \chi)| \ll \left(x + x^{5/6} Q + x^{1/2} Q^2 \right) \log^4 Qx,$$

where $\sum^*$ denotes summation over primitive characters only.

We begin by turning the large sieve into a bilinear inequality.

Lemma 20. *For integers M, N, Q and complex numbers a_m, b_n, $m \le M, n \le N$ we have*

$$\sum_{q\le Q}\frac{q}{\varphi(q)}\sideset{}{^*}\sum_{\chi \pmod q}\max_u\left|\sum_{\substack{m\le M\\ mn\le u}}\sum_{n\le N}a_m b_n\chi(nm)\right| \ll (M+Q^2)^{1/2}(N+Q^2)^{1/2}\left(\sum_{m\le M}|a_m|^2\right)^{1/2}\left(\sum_{n\le N}|b_n|^2\right)^{1/2}\log 2MN$$

Proof. Applying Cauchy-Schwarz and then the large sieve in the form of Theorem 8 we obtain

$$\begin{aligned}(8)\quad &\sum_{q\le Q}\frac{q}{\varphi(q)}\sideset{}{^*}\sum_{\chi \pmod q}\max_u\left|\sum_{m\le M}\sum_{n\le N}a_m b_n\chi(nm)\right| \\ &\le\left(\sum_{q\le Q}\sideset{}{^*}\sum_{\chi}\left|\sum_{m\le M}a_m\chi(m)\right|^2\right)^{1/2}\left(\sum_{q\le Q}\sideset{}{^*}\sum_{\chi}\left|\sum_{n\le N}b_n\chi(n)\right|^2\right)^{1/2} \\ &\ll (M+Q^2)^{1/2}(N+Q^2)^{1/2}\left(\sum_{m\le M}|a_m|^2\right)^{1/2}\left(\sum_{n\le N}|b_n|^2\right)^{1/2}\log 2MN.\end{aligned}$$

To introduce the condition $mn \le u$ we use the Mellin transform formula

$$\int_{-T}^{T}\frac{\sin t\beta}{t}e^{it\alpha}\,dt = \begin{cases}\pi, & |\alpha|\le\beta\\ 0, & |\alpha|>\beta\end{cases} + \mathcal{O}\left(\frac{1}{T(|\beta-|\alpha||}\right).$$

If we put

$$A(t,\chi)=\sum_{m\le M}a_m\chi(m)m^{-it},\quad B(t,\chi)=\sum_{n\le N}b_n\chi(n)n^{-it},$$

and $\beta=\log u$, then

$$\sum_{\substack{m\le M\\ mn\le u}}\sum_{n\le N}a_m b_n\chi(nm)=\int_{-T}^{T}A(T,\chi)B(T,\chi)\frac{\sin(t\log u)}{\pi t}dt + \mathcal{O}\left(\frac{1}{T}\sum_{n,m}|a_m b_n|\log\left|\frac{mn}{u}\right|^{-1}\right).$$

We now apply (8) to $A(t,\chi)B(t,\chi)$ and find that the main term is bounded above by

$$(M+Q^2)^{1/2}(N+Q^2)^{1/2}\left(\sum_{m\le M}|a_m|^2\right)^{1/2}\left(\sum_{n\le N}|b_n|^2\right)^{1/2}\int_{-T}^{T}\min\left(\frac{1}{|t|},\log 2MN\right)dt$$
$$\ll (M+Q^2)^{1/2}(N+Q^2)^{1/2}\left(\sum_{m\le M}|a_m|^2\right)^{1/2}\left(\sum_{n\le N}|b_n|^2\right)^{1/2}\log 2MNT.$$

As long as T is bounded by some power of MN, this is of acceptable magnitude. If we take $T=MN$, and assume that $\|u\|=1/2$, which we may do, since shifting u within an interval of the form $[k,k+1]$ does not alter the summation conditions, we have $\log\left|\frac{mn}{u}\right|\ge\log\frac{u+1/2}{u}\gg\frac{1}{T}$, hence the error term is

$$\ll\sum_{m,n}|a_mb_n|\ll M^{1/2}N^{1/2}\left(\sum_{m\le M}|a_m|^2\right)^{1/2}\left(\sum_{n\le N}|b_n|^2\right)^{2},$$

which is also acceptable. □

We now insert Vaughan's identity in the form $\Lambda(n)=a_1(n)+\cdots+a_4(n)$ into the left hand side of Theorem 19. We shall choose the parameters U,V at the end. We put $S_i(\chi,y)=\sum_{n\le y}a_i(n)\chi(n)$. Then we have

$$\sum_{q\le Q}\frac{q}{\varphi(q)}\max_{y\le x}|S_1(y,\chi)|\ll U\sum_{q\le Q}\frac{q}{\varphi(q)}\ll QU.$$

Next we treat S_4. We have

$$S_4(y,\chi)=\sum_{U<m\le y/V}\Lambda(m)\sum_{V<k\le y/m}\left(\sum_{\substack{d|k\\ d\le V}}\mu(d)\right)\chi(mk).$$

It is convenient to deal with summation parameters of the same magnitude, therefore we cut the sum over m into parts of the form $M<m\le 2M$. Using Lemma 20 we have

$$\sum_{q\le Q}\frac{q}{\varphi(q)}\sum_{\chi\pmod q}{}^{*}\max_{y\le x}\left|\sum_{M<m\le 2M}\Lambda(m)\sum_{V<k\le y/m}\left(\sum_{\substack{d|k\\ d\le V}}\mu(d)\right)\chi(mk)\right|$$
$$\ll (Q^2+M)^{1/2}\left(Q^2+\frac{x}{M}\right)^{1/2}\left(\sum_{m=M}^{2M}\Lambda(m)^2\right)^{1/2}\left(\sum_{k\le x/M}d(k)^2\right)^{1/2}\log x$$
$$\ll (Q^2+Qx^{1/2}M^{-1/2}+QM^{1/2}+x^{1/2})(M\log M)^{1/2}\left(\frac{x\log^3 x}{M}\right)^{1/2}\log x$$
$$\le (Q^2x^{1/2}+QxM^{-1/2}+QM^{1/2}x^{1/2}+x)\log^3 x$$

Summing over intervals of the form $[M, 2m]$ we obtain that the contribution of S_4 to the whole sum is bounded above by

$$(Q^2x^{1/2} + QxU^{-1/2} + QxV^{-1/2} + x)\log^4 x$$

The sum over S_3 is bounded above by

$$\sum_{q\leq Q}\sideset{}{^*}\sum_{\chi}\max_{y}\log y\sum_{d\leq V}\max_{w}\left|\sum_{w\leq h\leq y/d}\chi(h)\right|.$$

If χ is a primitive character modulos $q \geq 2$, we can use the Polya-Vinogradov inequality to bound the inner sum by $\sqrt{q}\log q \leq \sqrt{Q}\log Q$. Since this expression does not depend on anything, we find that non-principal characters contribute $\ll Q^{5/2}V\log^2 Qx$. There is only one primitive principal character, which contributes $\log x\sum_{d\leq V}\frac{x}{d} \ll x\log^2 x$, hence we find that the S_3 terms yield $(Q^{5/2}V+x)\log^2 Qx$.

Finally to compute the sum over S_2 we cut

$$S_2 = -\sum_{t\leq UV}\left(\sum_{\substack{t=md\\ m\leq U, d\leq V}}\mu(d)\Lambda(m)\right)\sum_{r\leq y/t}\chi(rt)$$

into two parts depending on the size of t. More precisely, we write $\sum_{t\leq UV} = \sum_{t\leq U} + \sum_{U<t\leq UV}$. We treat the first sum like we did S_3, and the second like we did S_4, and obtain that the sum over S_2 can be bounded above by

$$(Q^{5/2}U + x)\log^2 Qx + (Q^2x^{1/2} + QxU^{-1/2} + Qx^{1/2}U^{1/2}V^{1/2} + x)\log^2 x.$$

Putting everything together we find that

$$\sum_{q\leq Q}\frac{q}{\varphi(q)}\sideset{}{^*}\sum_{\chi \pmod q}\max_{y\leq x}|\Psi(x,\chi)|$$

is bounded above by

$$(Q^2x^{1/2}+QxU^{-1/2}+QxV^{-1/2}+Qx^{1/2}U^{1/2}V^{1/2})\log^4 x+(Q^{5/2}U+Q^{5/2}V+x)\log^2 Qx$$

We now ahve to choose optimal values for U and V. Since for UV fixed both $U+V$ and $U^{-1/2}+V^{-1/2}$ are minimal for $U = V$, we take $U = V$. The expression now becomes

$$(Q^2x^{1/2} + QxU^{-1/2} + Qx^{1/2}U + Q^{5/2}U + x)\log^4 Qx.$$

We can choose U in such a way that the second and third term become equal to $Qx^{5/6}$, and we can choose U in such a way that the second and the last term are both equal to $Q^{3/2}x^{2/3}$. Hence, depending on the relative size of x and Q we choose one or the other optimization, but in any case the last expression is of magnitude at most

$$(Q^2x^{1/2} + Qx^{5/6} + Q^{3/2}x^{2/3} + x)\log^4 Qx.$$

If $x \leq Q^3$, the first term dominates the third one, while if $x > Q^3$, the second dominates the third. Hence we can neglect the third one, which completes the proof of the theorem.

4.3. **Proof part II: Passing from characters to residue classes.** We now show how to deduce Theorem 18 from Theorem 19. The proof will be similar to the proof of Theorem 12. As in the proof of that theorem put

$$\Psi'(x,\chi) = \begin{cases} \Psi(x,\chi) - x, & \chi \text{ principal} \\ \Psi(x,\chi), & \chi \text{ not principal} \end{cases}.$$

Then we have

$$\left|\Psi(x,q,a) - \frac{x}{\varphi(q)}\right| \leq \frac{1}{\varphi(q)} \sum_{\chi \pmod q} |\Psi'(x,\chi)|.$$

Since the right hand side does not depend on a, the ineqiality remains valid if on the left hand side we take the maximum over all a with $(q,a)=1$. Replacing a character χ by the primitive character χ_1 inducing this character gives a contribution $\ll \log^2 qx$. We obtain

$$\max_{y\leq x} \max_{a:(q,a)=1} \left|\Psi(y,q,a) - \frac{y}{\varphi(q)}\right| \ll \log^2 qx + \frac{1}{\varphi(q)} \max_{y\leq x} \sum_{\chi \pmod q} |\Psi'(y,\chi_1)|$$

Summing over $q \leq Q$ the first term becomes $Q\log^2 Qx$, which is completely negligible. Since a primitive character $\chi_1 \pmod q$ induces characters modulo each multiple of q, we obtain

$$\sum_{q\leq Q} \max_{y\leq x} \max_{a:(q,a)=1} \left|\Psi(y,q,a) - \frac{y}{\varphi(q)}\right| \ll Q\log^2 Qx + \sum_{q\leq Q} \sum_{d\leq Q/q} \frac{1}{\varphi(qd)} \max_{y\leq x} {\sum_{\chi \pmod q}}^{*} |\Psi'(y,\chi)|$$

Since $\varphi(qd) \geq \varphi(q)\varphi(d)$, and $\sum_{t\leq z} \frac{1}{\varphi(t)} \ll \log z$, the right hand side is bounded above by

$$Q\log^2 Qx + \log Q \sum_{q\leq Q} \frac{1}{\varphi(q)} \max_{y\leq x} {\sum_{\chi \pmod q}}^{*} |\Psi'(y,\chi)|$$

The first summand is fine. If we restrict the second to the range $U < q \leq 2U$ and apply Theorem 19 we obtain

$$\sum_{U<q\leq 2U} \frac{1}{\varphi(q)} \max_{y\leq x} {\sum_{\chi \pmod q}}^{*} |\Psi'(y,\chi)| \ll \left(\frac{x}{U} + x^{5/6} + x^{1/2}U\right) \log^4 x.$$

Intersecting $[\log^A x, Q]$ into intervals of the form $[U, 2U]$ we obtain

$$\sum_{\log^A x<q\leq Q} \frac{1}{\varphi(q)} \max_{y\leq x} {\sum_{\chi \pmod q}}^{*} |\Psi'(y,\chi)| \ll \frac{x}{\log^{A-4} x} + x^{5/6}\log^5 x + x^{1/2}Q\log^4 x,$$

which is what we need. Finally in the range $q \leq \log^A x$ we can use the Siegel-Walfisz theorem in the form $\Psi'(x,\chi) \ll xe^{-c\sqrt{\log x}}$, hence the contribution of this range is much smaller. Hence Theorem 18 follows.

4.4. **The Bobieri-Vinogradov theorem and Selberg's sieve.** Suppose we want to bound the number of prime twins below x. Then, as we did in section 1.4, we can consider the set of all integers n, such that $n(n+2) \not\equiv 0 \pmod q$ for all small prime numbers q.However, doing so created some complication, since the sum we are optimizing no longer consists of simple terms like $\frac{1}{[d,e]}$, but the denominator now contains some number theoretic function of $[d,e]$.

Using the Bombieri-Vinogradov theorem we can avoid this complication. Instead of starting with all integers and sifting out two residue classes modulo all small primes, we start with all integers of the form $p+2$ and sift out one residue class modulo all small prime numbers. We will not give the details here, since the argument is essentially the same as in section 1.3. There are only two changes to be made: First, we only know that primes are well distributed modulo integers q up to $x^{1/2}$, hence we must ensure that $[d,e] < x^{1/2}\log^{-A} x$. This forces us to take z slightly smaller than $x^{1/2}$. The second change is that even in that range we have a good estimate for $\pi(x,q,-2)$ only for most values of q. A common way of dealing with this problem is to split the set of modules q into a good set, where we can estimate $\pi(x,q,-2)$, and a bad set, where we use a trivial bound (e.g. the Brun-Titchmarsh inequality). Since the number of bad modules is $< \frac{Q}{\log^A x}$, and we usually only have to beat the trivial bound by some power of $\log x$, this usually suffices.