CHAPTER 4

Selberg's sieve

Given (a_n) an arithmetic sequence a major problem in analytic number theory is to obtain informations on the sum

$$S_{\mathcal{P}}(a;x)=\sum_{n\leq x}a_n1_{\mathcal{P}}(n)=\sum_{p\leq x}a_p$$

which evaluate how the function a correlate with the characteristic function of the primes. For instance if a is the characteristic function of a some set of integers, one would be counting the number of prime $\leq x$ in that set.

When a_n is the characterist function of an arithmetic progression, we used Dirichlet characters and their associated L-function. For more general sets (not admitting an evident group theoretic interpretation) other (more robust) methods have been developped under the name of *sieve methods.* Such methods where considered by Legendre but the real pioneer was V. Brun in the beginning of the 20-th century.

During te exercise sessions you have seen how the large sieve inequalities can be used to great effect to evaluate (in fact to give non-trivial upper bounds for) the sums $S_{\mathcal{P}}(a;x)$. In this chapter we will discuss another, in fact older approach, due to another pioneer of the Sieve, A. Selberg.

1. Legendre sieve

Let us start first with some basics concerning sieve methods. The basic idea of the sieve is the following sufficient criterion

An integer $n\leq x$ is prime if it is not divisible by any prime $\leq x^{1/2}$.

Indeed if n where not prime it would be divisible by some prime $x^{1/2}<p<x$ and then $1<n/p<x^{1/2}$; therefore n/p would be divisible by some prime $<x^{1/2}$ which would also divide n. □

If follows from this sufficient condition that

$$S_{\mathcal{P}}(a;x)=\sum_{p\leq x^{1/2}}a_p+\sum_{\substack{x^{1/2}<n\leq x\\(n,P_{x^{1/2}}))=1}}a_n$$

where we have set for any $w\geq 1$

$$P_w:=\prod_{p\leq w}p.$$

Let $\mathcal{P}' \subset \mathcal{P}$ be a finite set of primes and

$$P = \prod_{\substack{p \le z \\ p \in \mathcal{P}'}} p$$

a finite product of primes from $\mathcal{P}'$; we look to evaluate the sum

$$S(a; x, P) = \sum_{\substack{n \le x \\ (n,P)=1}} a_n$$

(as pointed out before if P is the product of primes $\le x^{1/2}$, that sum is the sum of a_p over the primes p satisfying $x^{1/2<p\le x}$). We have

$$\delta(m) = \begin{cases} 1 & \text{if } m = 1 \\ 0 & \text{else} \end{cases} = \sum_{d|m} \mu(d)$$

and therefore

$$\delta_{(n,P)=1} = \sum_{\substack{d|P \\ d|n}} \mu(d)$$

$$S(a; x, P) = \sum_{d|P} \mu(d) A(x; d), \;\; A(x; d) = \sum_{d|n\le x} a_n = \sum_{m \le x/d} a_{dm}.$$

We therefore need information on the average behavior of a_n in the 0 congruence class modulo d for d a divisor of P: let us assume that that the sum $A(x; d)$ is written in the following form

$$A(x; d) = g(d)X + r(x; d)$$

where X is a function of x (eg. $x \log x$) which does not depend on d, $g(d) > 0$ is a multiplicative function so that $g(d)X$ and $r(x; d)$ is an error term.

EXAMPLE 4.1. $a \equiv 1$, $X = x$, $g(d) = 1/d$, $|r(x; d)| \le 1$.

We obtain

$$S(a; x, P) = V(P)X + R(x, P)$$

$$V(P) = \prod_{p|P} (1 - g(p))$$

$$R(x, P) = \sum_{d|P} \mu(d) r(x; d).$$

2. Upper bound sieves

The problem is that $R(x,P)$ may have too many terms to make a good error term (a priori up to x terms).

Brun's idea was to find a variant of Legendre's identity to reduce the number of terms in that summation. This hower require making some compromise, for instance by looking only at upper or lower bound for the sum $S(a;x,P)$.

Suppose that (a_n) is non negative we seek non-trivial bounds for

$$S(a;x,P)$$

for this it is sufficient to find "weights" (λ_d) such that

$$w(m)=\sum_{d|m}\lambda_d=\begin{cases}\lambda_1=1 & m=1\\ \geq 0, & m>1\end{cases} \tag{4.1}$$

for one would obtain

$$\begin{aligned}S(a;x,P)\leq\sum_{n\leq x}w(n)a_n&=\sum_{n\leq x}(\sum_{d|(n,P)}\lambda_d)a_n\\ &=\sum_{d|P}\lambda_d A(x;d)=V_\lambda(P)X+R_\lambda(x,P)\end{aligned}$$

with

$$V_\lambda(P)=\sum_{d|P}\lambda_d g(d),\ R_\lambda(x,P)=\sum_{d|P}\lambda_d r(x;d).$$

One expect to be able to evaluate the "main term" $V_\lambda(P)x$ with not too much effort. As for the "remainder term" $R_\lambda(x,P)$ which is wilder because the $r(x;d)$ are error terms, one hope to be able to evaluate it when (λ_d) has relatively small support: for instance if $\mathrm{supp}(\lambda)\subset[1,z]$ for some $z\leq x$ the sum $R_\lambda(x,P)$ will have at most z terms and what we are then required to evaluate is

$$\sum_{d\leq z}|\lambda_d||r(x;d)|$$

which is a measure of the average quality of the distribution of (a_n) in the arithmetic progression $0\,(\mathrm{mod}\,d)$ for moduli $d\leq z$.

Brun was the first to find such efficient coefficients (λ_d) and used them to show that the set of "twin primes" (the set of pairs $(p,p+2)$ where both p and $p+2$ are primes) is small:

$$\sum_{p,p+2\text{ both primes}}\frac{1}{p}<\infty$$

(remember that

$$\sum_{p\text{ prime}}\frac{1}{p}=\infty.)$$

Brun construction (which is called a combinatorial sieve) was quite involved.

3. Selberg's upper bound sieve

A few years later, Selberg found a different and robust way to construct upperbound sieve coefficients: he has the simple but beautiful idea to use the fact that squares are non-negative to enforce the requirement (4.1): Selberg choose the λ_d such that

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

for (ρ_d) another arithmetic function normalized to that

$$\rho_1 = 1. \tag{4.2}$$

This gives

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

We obtain

$$S(a;x,P) \leq \sum_{n\leq x} a_n \sum_{d|(n,P)} \lambda_d = \sum_{n\leq x} a_n (\sum_{d|(n,P)} \rho_d)^2$$

$$= \sum\sum_{d_1,d_2|P} \rho_{d_1}\rho_{d_2} A_{[d_1,d_2]}(x) = GX + R(x,P)$$

where

$$G = \sum\sum_{d_1,d_2|P} \rho_{d_1}\rho_{d_2} g([d_1,d_2])$$

$$R = \sum\sum_{d_1,d_2|P} \rho_{d_1}\rho_{d_2} r(x;[d_1,d_2]).$$

Now, given some parameter $z \geq 1$, if we assume that

$$(\rho_d) \text{ is supported on the interval } [1, z^{1/2}], \tag{4.3}$$

then the sequence (λ_d) will be supported on $[1, z]$ (since $d = [d_1, d_2] \leq d_1 d_2 \leq z$ if $d_1, d_2 \leq z^{1/2}$) therefore the sum $R(x, P)$ will have at most z terms in its summation.

We would like to choose the (ρ_d) so as to minimize the value of the quadratic form $G = G((\rho_d))$ with respect to the constraints (4.2) and (4.3). For this we will diagonalize this quadratic form.

Since d will anyway range over divisors of P which is squarefree we may and will assume that g is supported on squarefree integers. We will also assume that

$$g(p) \in]0, 1] \text{ whenever } p|d.$$

Likewise we may and will assume that ρ_d is supported on squarefree divisors of P.

We have (writing $a=(d_1,d_2),\ d_1=ab,\ d_2=ac,\ b,c\leftrightarrow eb,ec,\ d=ae$ and using that $g(a)\neq 0$ for $a|P$)

$$G=\sum_{\substack{abc|P\\(b,c)=1}} g(abc)\rho_{ac}\rho_{bc}=\sum_a g(a)\sum_e \mu(e)g(e)^2(\sum_{aeb|P} g(b)\rho_{aeb})^2$$

$$=\sum_{a,e}\mu(e)g(a)^{-1}(\sum_{aeb|P} g(aeb)\rho_{aeb})^2=\sum_d h(d)^{-1}(\sum_{db|P} g(db)\rho_{db})^2,$$

where h denote the multiplicative function supported on squarefree integers and such that

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

We set

$$\xi_d=\mu(d)\sum_{bd|P} g(bd)\rho_{bd}$$

and in particular

$$\mathrm{supp}(\rho_d)\subset[1,z^{1/2}]\Rightarrow\mathrm{supp}(lambda_d)\subset[1,z^{1/2}].$$

With these notations, we obtain

$$G=\sum_{d|P} h(d)^{-1}\xi_d^2=G'((\xi_d))$$

say. We claim that the map

$$(\rho_d)\mapsto(\xi_d)$$

is bijective in the space of sequences indexed by the divisors of P; with this, we have diagonalized the quadratic form $G((\rho_d))$. Let usprove the claim: we have

$$\rho_d=\frac{\mu(d)}{g(d)}\sum_{dl|P}\xi_{dl}.$$

Indeed (recall that the terms in this sum range over divisors of P) for any $d|P$,

$$\sum_{dl|P}\xi_{dl}=\sum_{dbl|P}\mu(dl)g(dbl)\rho_{dbl}=\mu(d)\sum_{m|P/d} g(dm)\rho_{dm}\sum_{l|m}\mu(l)=\mu(d)g(d)\rho_d$$

(setting $m=bl$). In particular, under the change of variable $\rho\leftrightarrow\xi$, (4.2) becomes

$$\sum_{d|P}\xi_d=1. \tag{4.4}$$

What remains to do is to minimize the quadratic form $G'((\xi_d))$ under the linear constraint (4.4). By Cauchy-Schwarz, we have

$$1=(\sum_{d|P}\xi_d)=(\sum_{d|P} h(d)^{1/2}h(d)^{-1/2}\xi_d)\leq GH$$

with

$$H = \sum_{d \leq z^{1/2}} h(d)$$

so that

$$G \geq 1/H$$

with equality if

$$\xi_d = h(d)/H$$

and we choose ξ_d that way. We have therefore

$$\rho_d = \frac{\mu(d)}{g(d)} \sum_{l|P/d \leq z^{1/2}/d} h(dl) = \frac{\mu(d)h(d)}{g(d)} \sum_{l \leq z^{1/2}/d,\ (d,l)=1} h(l). \tag{4.5}$$

Proposition 4.1. *We have*

$$|\rho_d| \leq 1$$

and therefore

$$|R| \leq \sum_{d \leq z} d_3(d)|r(x;d)|.$$

Proof. We have for any $d|P$

$$H = \sum_{k|d} \sum_{\substack{l \leq z^{1/2},(d,l)=k \\ l|P}} h(l) = \sum_{k|d} h(k) \sum_{\substack{l' \leq z^{1/2}/k \\ kl'|P}} h(l')$$

$$\geq \sum_{k|d} h(k) \sum_{l' \leq z^{1/2}/d} dl'|Ph(l') = |\rho_d| H$$

since

$$\sum_{k|d} h(k) = \prod_{p|d} (1 + h(p)) = h(d)/g(d).$$

We have therefore

$$|R| \leq \sum_{d \leq z} (\sum_{[d_1,d_2]=d} 1)|r(x;d)| \leq \sum_{d \leq z} d_3(d)|r(x;d)|$$

□we have proven the

Theorem 4.1 (Selberg sieve). *Let $z \leq x$ and (a_n) be a non-negative arithmetic function. Write*

$$A_d = g(d)X + r(x;d)$$

where g is a multiplicative function supported on the divisors of P satisfying for $p|P|$, $g(p) \in]0,1]$. We have

$$S(a;x,P) = \sum_{n \leq x,\ (n,P)=1} a_n \leq \frac{X}{H} + R$$

where

$$H = \sum_{d\leq z^{1/2}, d|P} h(d)$$

with

$$h(d) = \prod_{p|d} \frac{g(p)}{1-g(p)}$$

and R satisfying

$$|R| \leq \sum_{d\leq z} d_3(d)|r(x;d)|.$$

4. Application to twin primes

We will use Selberg's sieve to bound the number of twin primes: the integers n such that n and $n+2$ are both primes; more generally, for any h even, we will bound the number of pairs $(n, n+h)$ where both entries are primes. Let

$$\pi_{(0,h)}(x) = |\{n \leq x,\ n,\ n+h \text{ both primes}\}| = \sum_{p\leq x} a_p$$

where we have set for $n \geq 1$

$$a_n = 1_{\mathcal{P}}(n+h).$$

Given $w > 1$ let

$$P_w = \prod_{\substack{p\leq w\\(p,h)=1}} p.$$

The sum $S(a; x, P)$ counts the number of integers $n \leq x$ whose smallest prime factor is $> w$ such that $n+h$ is prime (if $n+h$ is prime then n is automatically coprime with any prime divisor of h). We will take $w = z^{1/2}$ where $z = x^\alpha$ where $\alpha > 0$.

We have for d squarefree

$$\sum_{d|n} a_n = \pi(x+h; d, h) + O(1) = \pi(x; d, h) + O(1) = g(d)X + O(E(1_{\mathcal{P}}, x; d, h) + 1)$$

with

$$X = \pi(x),$$

$$g(p) = \begin{cases} 0 & p|h \\ \frac{1}{p-1} & p \not| h. \end{cases}$$

and

$$h(p) = \begin{cases} 0 & p|h \\ \frac{1}{p-2} & p \not| h. \end{cases}$$

Lemma 4.1. *One has for* $t \geq 2$

$$\sum_{d \leq t} h(d) = c_h (\log t)(1 + O(\log^{-1} t))$$

with

$$c_h = \prod_p (1 + h(p))(1 - \frac{1}{p}) = \prod_p \frac{(1 - \frac{1}{p})^2}{1 - \frac{2}{p}} = C_{2,h}^{-1}.$$

In particular

$$H = \sum_{d \leq z^{1/2}} h(d) = c_h (\log(z^{1/2}))(1 + O(\log^{-1} z))$$

Proof. Let

$$L(s) = \sum_d h(d)/d^{s-1}$$

where we have extended h to $\mathbb{N}_{\geq 1}$. One has

$$L(s) = \prod_p (1 + \frac{ph(p)}{p^s}) = \zeta(s) \prod_p (1 + \frac{ph(p)}{p^s})(1 - \frac{1}{p^s})$$

and since

$$(1 + \frac{ph(p)}{p^s})(1 - \frac{1}{p^s}) = (1 + \frac{(1 - \frac{1}{p})}{1 - \frac{1}{p-1}} p^{-s})(1 - \frac{1}{p^s})$$

$$= (1 + p^{-s} + O(p^{-s-1}))(1 - 1/p^s) = 1 + O(p^{-2s} + p^{-s-1})$$

the series $L(s)/\zeta(s)$ is absolutely converging for $\Re s > 1/2$ and

$$\frac{L}{\zeta}(1) = \prod_p (1 + \frac{1}{p-2})(1 - \frac{1}{p}) = c_h$$

Therefore writing $dh(d) = \alpha \star 1(d)$ one has

$$\sum_{d \leq x} h(d) = \sum_{ab \leq x} \frac{\alpha(a)}{ab} = \sum_{b \leq x} \frac{1}{b} \sum_{a \leq x/b} \frac{\alpha(a)}{a}$$

$$= \sum_{b \leq x} \frac{1}{b}(c_h + O(b/x)) = c_h \log x + O(1).$$

□

Applying Theorem 4.1 we obtain

$$\pi_{(0,h)}(x) = \sum_{p \leq x} 1_{\mathcal{P}}(p + h) \leq \sum_{\substack{n \leq x \\ (n, P_{z^{1/2}})=1}} 1_{\mathcal{P}}(n + h) + O(z^{1/2})$$

$$\leq 2C_{2,h} \frac{x}{\log z + O(1)} + O(z \log^2 z + \sum_{d \leq z} d_3(d) E(1_{\mathcal{P}}, x; d, h))$$

We now bound the error term: by CS we have

$$\sum_{d\leq z} d_3(d)E(1_{\mathcal{P}},x;d,h) \leq (\sum_{d\leq z}(d_3(d))^2E(1_{\mathcal{P}},x;d,h))^{1/2}(\sum_{d\leq z}E(1_{\mathcal{P}},x;d,h))^{1/2}$$

We have by a trivial estimate

$$E(1_{\mathcal{P}},x;d,h) \ll x/\varphi(d)$$

so that the first factor is bounded by

$$\ll x^{1/2}\log^{O(1)} x$$

while by the Bombieri-Vinogradov theorem, the second factor is bounded by

$$\ll_A x^{1/2}\log^{-A} x$$

for any $A > 0$ as long as

$$z \leq x^{1/2}/\log^{B(A)} x.$$

We have therefore obtained (setting $z = x^{1/2}/\log^{B(A)} x$)

$$\sum_{p\leq x} 1_{\mathcal{P}}(p+h) \leq 4C_{2,h}\frac{x}{\log x}(1+o(1)).$$

The precise form of the twin prime conjecture predict that

$$\sum_{p\leq x} 1_{\mathcal{P}}(p+h) = C_{2,h}\frac{x}{\log x}(1+o(1))$$

therefore our upper-bound is off by a factor 4.

Remark 4.1. On can prove a weaker result which avoid the BV theorem: by counting the number of n such that the polynomial value

$$P_h(n) = n(n+h)$$

is coprime with P_z for a suitable value of z (namely $z = x/\log^{B(A)} x$): one can then show that

$$\pi_{(0,h)}(x) \leq 8C_{2,h}\frac{x}{\log x}(1+o(1)).$$

In other terms, we have improved the upper bound by a factor 2. This difference in treatment is the essence of the superiority of Maynard's method over the Goldston-Yildirim-Pintz-Zhang's initial approach.