LECTURE NOTES 1 FOR CAMBRIDGE PART III COURSE ON "ELEMENTARY METHODS IN ANALYTIC NUMBER THEORY", LENT 2015

ADAM J HARPER

Abstract. These are rough notes covering the first block of lectures in the "Elementary Methods in Analytic Number Theory" course. In these first lectures we will explore the idea of sieving out by different primes, introduce Selberg's powerful method for doing this, and apply this to some prime counting problems. We will also discuss the limitations on sieve methods imposed by the Parity Phenomenon.

(No originality is claimed for any of the contents of these notes. In particular, they borrow from the book of Friedlander and Iwaniec [1].)

1. The idea of sieving

By (a slight reformulation of the) definition, a number $2 \leq n \leq x$ is prime if and only if it has no *prime* divisors strictly less than n. This leads to the following ancient inductive procedure for identifying all the primes $p \leq x$, which will probably be familiar to most people from primary school:

- $p_1 := 2$ is the first prime;
- given a list $p_1 < p_2 < ... < p_k$ of all primes up to some point p_k less than x, strike out (i.e. remove) all multiples of $p_1, ..., p_k$ from the numbers $2 \leq n \leq x$;
- set p_{k+1} to be the least n that has not been struck out, repeating the above until all numbers less than x are struck out.

The process of striking out multiples of p_j is usually described as *sieving by* p_j (or *sieving by the zero residue class modulo* p_j), since only those numbers that are not divisible by p_j survive through the sieve. The entire procedure described above is usually called the *Sieve of Eratosthenes*.

As we have described it, the Sieve of Eratosthenes is an algorithm for listing prime numbers, not a method for counting them (or quickly estimating the number of them). However, we can adapt things and obtain a counting version, which is usually called the *Sieve of Eratosthenes–Legendre*. This will be facilitated by introducing a bit of notation.

Date: 5th February 2015.

Definition 1.1. Let $\omega(n)$ denote the number of distinct prime factors of n, and define the *Möbius function*

$$\mu(n) := \begin{cases} 0 & \text{if } n \text{ has a non-trivial square divisor} \\ (-1)^{\omega(n)} & \text{if } n \text{ is squarefree.} \end{cases}$$

Note that, in the Sieve of Eratosthenes, a number will be struck out multiple times depending on the number of distinct prime factors that it has (below a certain point). The Möbius function keeps track of this in an inclusion-exclusion way, since we will remove (i.e. subtract) the multiples of all primes; then add back the multiples of all products of two primes to avoid over-subtracting; then remove all products of three primes to avoid over-compensating; etc..

Proposition 1.2 (Sieve of Eratosthenes–Legendre, Basic Version)**.** *For any $2 \le y \le x$ we have*

$$\#\{n \le x : p \mid n \Rightarrow p > y\} = \sum_{\substack{d \le x, \\ p \mid d \Rightarrow p \le y}} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor,$$

where $\lfloor \cdot \rfloor$ denotes integer part.

In particular, we have

$$\pi(x) - \pi(\sqrt{x}) + 1 = \sum_{d \le \sqrt{x}} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor,$$

and for any $2 \le y \le x$ we have

$$\pi(x) \le \pi(y) - 1 + \sum_{\substack{d \le x, \\ p \mid d \Rightarrow p \le y}} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor = x \prod_{p \le y} \left(1 - \frac{1}{p}\right) + O(2^{\pi(y)}).$$

Proof of Proposition 1.2. Note that for any integer n and any y,

$$\sum_{\substack{d \mid n, \\ p \mid d \Rightarrow p \le y}} \mu(d) = \prod_{\substack{p \mid n, \\ p \le y}} (1 + \mu(p)),$$

in view of the fact that $\mu(p_1)...\mu(p_j) = \mu(p_1...p_j)$ for any *distinct* primes $p_1, ..., p_j$. However, by definition of the Möbius function we always have $1 + \mu(p) = 0$, so the product is zero unless it has no terms, so actually

$$\sum_{\substack{d \mid n, \\ p \mid d \Rightarrow p \le y}} \mu(d) = \mathbf{1}_{p \mid n \Rightarrow p > y},$$

where $\mathbf{1}$ denotes the indicator function. It follows, as claimed, that

$$\#\{n \le x : p \mid n \Rightarrow p > y\} = \sum_{n \le x} \mathbf{1}_{p \mid n \Rightarrow p > y} = \sum_{n \le x} \sum_{\substack{d \mid n, \\ p \mid d \Rightarrow p \le y}} \mu(d) = \sum_{\substack{d \le x, \\ p \mid d \Rightarrow p \le y}} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor.$$

The second statement follows by taking $y = \sqrt{x}$ in the first statement. For if $n \leq x$ has two or more prime factors, at least one of them must be smaller than $\sqrt{x}$. Therefore $\{n \leq x : p \mid n \Rightarrow p > \sqrt{x}\}$ consists precisely of the primes strictly between $\sqrt{x}$ and x, together with the unit 1 (which has no prime divisors).

The first part of the third statement follows because $\{n \leq x : p \mid n \Rightarrow p > y\}$ certainly contains all the primes strictly between y and x (and the unit 1). For the second part, note that

$$\sum_{\substack{d \leq x, \\ p|d \Rightarrow p \leq y}} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor = \sum_{d:p|d \Rightarrow p \leq y} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor = \sum_{d:p|d \Rightarrow p \leq y} \mu(d) \left(\frac{x}{d} + O(1) \right)$$

$$= \sum_{d:p|d \Rightarrow p \leq y} \mu(d) \frac{x}{d} + O\left(\sum_{d:p|d \Rightarrow p \leq y} |\mu(d)| \right).$$

Then, using again the fact that $\mu(p_1)...\mu(p_j) = \mu(p_1...p_j)$ for distinct p_i, we have

$$\sum_{d:p|d \Rightarrow p \leq y} \mu(d) \frac{x}{d} = x \prod_{p \leq y} \left(1 + \frac{\mu(p)}{p} \right) = x \prod_{p \leq y} \left(1 - \frac{1}{p} \right),$$

and similarly the "big Oh" error term is $O(\prod_{p \leq y}(1+1)) = O(2^{\pi(y)})$. We clearly have $\pi(y) - 1 = O(2^{\pi(y)})$ as well. □

Let's take stock of what we have shown about primes in Proposition 1.2.

- We have an exact formula for $\pi(x) - \pi(\sqrt{x}) + 1$, so we have a formula for $\pi(x)$ up to an error $O(\sqrt{x})$.
- We have a family of upper bounds for $\pi(x)$, in terms of a parameter y.

Unfortunately, our exact formula involves the Möbius function $\mu(d)$, which changes sign mysteriously, and it involves the integer part function $\lfloor \cdot \rfloor$ which is hard to understand. Thus it is only the upper bound in terms of y that is really useful: indeed, if we invoke Fact 3 from Chapter 0 we can rewrite that upper bound as

$$\pi(x) \leq x \prod_{p \leq y} \left(1 - \frac{1}{p} \right) + O(2^{\pi(y)}) \ll \frac{x}{\log y} + 2^{\pi(y)}.$$

However, in order that the term $2^{\pi(y)}$ doesn't make the bound trivial we need to have $\pi(y) \leq \log x$ (say). This means that we can only choose y roughly as big as $\log x$, and so the strongest bound we can deduce is that

$$\pi(x) \ll \frac{x}{\log \log x}.$$

This is much weaker than the bound $\pi(x) \ll x/\log x$ of Chebychev, contained in Fact 1 of Chapter 0.

This is disappointing, but we can console ourselves that we have done better than the trivial bound $\pi(x) \leq x$. It also turns out that the Sieve of Eratosthenes–Legendre is rather more general than Chebychev's bounds, and can prove results that do not follow from them, such as:

Proposition 1.3 (Sieve of Eratosthenes–Legendre, Second Version)**.** *For any* $10 \leq z \leq x$ *we have*

$$\pi(x+z) - \pi(x) \ll \frac{z}{\log\log z}.$$

Proof of Proposition 1.3. Let $2 \leq y \leq z$ be a parameter, and note that again we have

$$\pi(x+z) - \pi(x) \leq \#\{x < n \leq x+z : p \mid n \Rightarrow p > y\} = \sum_{\substack{d \leq x+z, \\ p \mid d \Rightarrow p \leq y}} \mu(d) \left(\left\lfloor \frac{x+z}{d} \right\rfloor - \left\lfloor \frac{x}{d} \right\rfloor \right).$$

Arguing as in the proof of Proposition 1.2, the right hand side is

$$\sum_{d : p \mid d \Rightarrow p \leq y} \mu(d) \left(\left\lfloor \frac{x+z}{d} \right\rfloor - \left\lfloor \frac{x}{d} \right\rfloor \right) = \sum_{d : p \mid d \Rightarrow p \leq y} \mu(d) \left(\frac{z}{d} + O(1) \right) = z \prod_{p \leq y} \left(1 - \frac{1}{p} \right) + O(2^{\pi(y)}).$$

Choosing $y = \log z$ then yields the result. □

This is non-trivial (although also not sharp) even when z is very small compared with x. This is a powerful feature that we shall discuss again later.

The motivating problem of sieve theory is to find a way to improve the Sieve of Eratosthenes–Legendre. *More precisely, we want to find a substitute for the identity*

$$\sum_{\substack{d \mid n, \\ p \mid d \Rightarrow p \leq y}} \mu(d) = \boldsymbol{1}_{p \mid n \Rightarrow p > y}$$

that leads to sharper results. It turns out that considerable improvements are possible. However, some features of Propositions 1.2 and 1.3 will remain relevant throughout our work. It is generally much easier to obtain upper bounds than lower bounds using sieve methods (i.e. methods based on sums over divisors, loosely speaking). Obtaining useful asymptotic formulae is generally extremely difficult, but can be done in certain cases that constitute some of the crowning achievements of sieve theory.

2. Selberg's sieve

Our goal is to find a substitute for the identity $\sum_{\substack{d \mid n, \\ p \mid d \Rightarrow p \leq y}} \mu(d) = \mathbf{1}_{p \mid n \Rightarrow p > y}$ that leads to superior results in various applications, such as counting primes. Actually it is useful

to start from a more general position, namely the identity

$$\sum_{\substack{d|n,\\ p|d\Rightarrow p\in\mathcal{P}}} \mu(d) = \mathbf{1}_{p|n\Rightarrow p\notin\mathcal{P}},$$

where $\mathcal{P}$ is any set of primes that one wants to sieve by (sometimes called the *sifting range*). This identity can be proved exactly as in the previous section. The greater generality is important in some cases, and is also psychologically helpful because the sieve process is no longer tied up automatically with the size of the divisors d.

If we temporarily define $\lambda_d := \mu(d)\mathbf{1}_{p|d\Rightarrow p\in\mathcal{P}}$ for all $d \in \mathbb{N}$, then our identity asserts that

$$\sum_{d|n} \lambda_d = \mathbf{1}_{p|n\Rightarrow p\notin\mathcal{P}}.$$

The problem is that if $\mathcal{P}$ is a moderately large set of primes then the support of the sequence (λ_d) is very large, leading to many "big Oh" error terms in our arguments that we cannot control. *So we can try to look for alternative sequences of weights λ_d (depending on $\mathcal{P}$ and possibly on other features of the problem) that have smaller support*, and that still satisfy some approximate identity $\sum_{d|n} \lambda_d \approx \mathbf{1}_{p|n\Rightarrow p\notin\mathcal{P}}$. In upper bound sieve theory, the kind of "approximate identity" that we usually insist on is that

$$\sum_{d|n} \lambda_d \geq \mathbf{1}_{p|n\Rightarrow p\notin\mathcal{P}} \ \ \forall n \in \mathbb{N}.$$

The first major developments in sieve theory were due to Brun in the 1910s–1920s, who chose his weights λ_d to be the restriction of $\mu(d)$ to certain sets depending on the arithmetic structure of d. This approach can be very powerful, but is also rather complicated initially. Another famous approach originated in 1947 with Selberg, who found an amazingly simple way to select the λ_d that is also very powerful. An important feature is that Selberg's approach no longer constrains his weights to take the values $-1, 0, 1$, which introduces extra flexibility.

Lemma 2.1 (Λ^2 construction, following Selberg)**.** *Let $\mathcal{P}$ be any set of primes, let $2 \leq \sqrt{D}$, and let (ρ_d) be any sequence of real numbers subject to the following constraints:*

$$\rho_1 = 1, \ \text{and} \ \rho_d = 0 \, \text{unless} \, (p \mid d \Rightarrow p \in \mathcal{P}), \ \text{and} \ \rho_d = 0 \, \forall d > \sqrt{D}.$$

Define

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

where $[d_1, d_2]$ denotes least common multiple.

Then the weights λ_d satisfy

$$\sum_{d|n} \lambda_d \geq \mathbf{1}_{p|n\Rightarrow p\notin\mathcal{P}} \ \ \forall n \in \mathbb{N}.$$

Moreover the λ_d are supported on those $d \le D$.

Proof of Lemma 2.1. The first part of the proof follows from the observation that

$$\sum_{d|n} \lambda_d = \sum_{d|n} \sum_{[d_1,d_2]=d} \rho_{d_1}\rho_{d_2} = \sum_{\substack{d_1|n,\\ d_2|n}} \rho_{d_1}\rho_{d_2} = \left(\sum_{d|n} \rho_d\right)^2.$$

Since the ρ_d are real the right hand side is certainly ≥ 0. Moreover, if all the prime factors of n are $\notin \mathcal{P}$ then the only term in the sum on the right is $\rho_1 = 1$, so it is $\ge \mathbf{1}_{p|n \Rightarrow p \notin \mathcal{P}}$.

The statement about the support of λ_d follows because $[d_1, d_2] \le d_1 d_2$ for all d_1, d_2, and $\rho_d = 0$ whenever $d > \sqrt{D}$, so $\lambda_d = 0$ whenever $d > D$. □

Remark 2.2. If sieve weights (λ_d) are supported on those $d \le D$ we sometimes say the weights have *level* D.

In view of Lemma 2.1, for a given level D we have a very flexible way to construct upper bound sieve weights λ_d, since we have great freedom in choosing the constituent parts ρ_d. However, it is not clear what choice might lead to good results in applications. To explore this issue, let us think again about how we are ultimately going to use our sieve weights, this time in a bit more generality than when we just sieved for primes in intervals.

Lemma 2.3. *Let $\mathcal{P}$ be any set of primes and let $2 \le \sqrt{D}$. Let $\mathcal{A} = (a_n)$ be any finite sequence of non-negative numbers, let $X > 0$, and suppose we are given a function $g(d)$, supported on squarefree numbers, that satisfies:*

(i) *(multiplicativity) $g(ab) = g(a)g(b)$ whenever a, b are coprime;*
(ii) *$0 < g(p)$ for all primes p.*

Finally, define the remainder numbers $r(d) = r(d, \mathcal{A}, g, X)$ by

$$r(d) := \sum_{n:d|n} a_n - g(d)X.$$

Then for any upper bound sieve weights (λ_d) of level D (i.e. any sequence (λ_d) that satisfies $\sum_{d|n} \lambda_d \ge \mathbf{1}_{p|n \Rightarrow p \notin \mathcal{P}}$ for all n and is supported on those $d \le D$), we have

$$\sum_{n:p|n \Rightarrow p \notin \mathcal{P}} a_n \le X \sum_{d \le D} \lambda_d g(d) + \sum_{d \le D} \lambda_d r(d).$$

In particular, if the weights (λ_d) are obtained using the Λ^2 construction (Lemma 2.1) then we have

$$\sum_{n:p|n\Rightarrow p\notin\mathcal{P}} a_n \leq X \sum_{t\leq\sqrt{D}} \left(\sum_{\substack{m\leq\sqrt{D},\\ t|m}} g(m)\rho_m \right)^2 \prod_{p|t} \left(\frac{1}{g(p)} - 1 \right) + \sum_{d\leq D} \lambda_d r(d).$$

Remark 2.4. As stated, Lemma 2.3 applies to any sequence $\mathcal{A}$ for any X and $g(d)$, since the remainders $r(d)$ are unconstrained. However, in practice we will want the $r(d)$ to usually be small compared with $Xg(d)$, so that the more structured term $X\sum_{d\leq D}\lambda_d g(d)$ is hopefully the main term. Notice this is the case when sieving the interval $(x, x+z]$, say, in which case we can take $a_n = \mathbf{1}_{x<n\leq x+z}$, $X = z$, and $g(d) = 1/d$ for all squarefree d. (Later we will arrange for the λ_d to be supported on squarefree d with all their prime factors from $\mathcal{P}$, so we only need to care about the squarefree case.)

Proof of Lemma 2.3. By definition of upper bound sieve weights of level D, and by non-negativity of the a_n, we have

$$\sum_{n:p|n\Rightarrow p\notin\mathcal{P}} a_n \leq \sum_n a_n \sum_{d|n} \lambda_d = \sum_{d\leq D} \lambda_d \sum_{n:d|n} a_n.$$

Now using the fact that $\sum_{n:d|n} a_n = g(d)X + r(d)$, we obtain that

$$\sum_{n:p|n\Rightarrow p\notin\mathcal{P}} a_n \leq X \sum_{d\leq D} \lambda_d g(d) + \sum_{d\leq D} \lambda_d r(d),$$

as claimed.

In the special case of Λ^2 sieve weights, by definition we have

$$\sum_{d\leq D} \lambda_d g(d) = \sum_{d\leq D} g(d) \sum_{[d_1,d_2]=d} \rho_{d_1}\rho_{d_2} = \sum_{d_1,d_2\leq\sqrt{D}} \rho_{d_1}\rho_{d_2} g([d_1,d_2]).$$

((Here we would like to manipulate things to "decouple" the sums over d_1, d_2 as much as we can.)) On letting c denote the highest common factor of $d_1 = ac, d_2 = bc$ we can rewrite the above as

$$\Sigma := \sum_{\substack{a,b,c\leq\sqrt{D},\\ (a,b)=1}} \rho_{ac}\rho_{bc} g(abc).$$

Moreover, since g is supported on squarefree integers we may assume in the sum that a, b, c are all squarefree and that $(ab, c) = 1$, although we won't write this explicitly to ease the notation. Since g is multiplicative we then have that Σ is

$$\sum_{\substack{a,b,c\leq\sqrt{D},\\ (a,b)=1}} \rho_{ac}\rho_{bc} g(a)g(b)g(c) = \sum_{\substack{a,b,c\leq\sqrt{D},\\ (a,b)=1}} \rho_{ac}\rho_{bc}\frac{g(ac)g(bc)}{g(c)} = \sum_{c\leq\sqrt{D}} \frac{1}{g(c)} \sum_{a\leq\sqrt{D}} g(ac)\rho_{ac} \sum_{b\leq\sqrt{D}} g(bc)\rho_{bc}\mathbf{1}_{(a,b)=1}.$$

Finally, similarly as previously we can write the indicator function $\mathbf{1}_{(a,b)=1}$ as a divisor sum involving the Möbius function, namely

$$\mathbf{1}_{(a,b)=1} = \prod_{p|(a,b)} (1+\mu(p)) = \sum_{d|(a,b)} \mu(d) = \sum_{d|a \text{ and } b} \mu(d).$$

Consequently we have

$$\begin{aligned} \Sigma &= \sum_{c \le \sqrt{D}} \frac{1}{g(c)} \sum_d \mu(d) \left(\sum_{\substack{a \le \sqrt{D}, \\ d|a}} g(ac)\rho_{ac} \right) \left(\sum_{\substack{b \le \sqrt{D}, \\ d|b}} g(bc)\rho_{bc} \right) \\ &= \sum_{c \le \sqrt{D}} \frac{1}{g(c)} \sum_d \mu(d) \left(\sum_{\substack{m \le \sqrt{D}, \\ cd|m}} g(m)\rho_m \right)^2, \end{aligned}$$

and changing variables by letting $t = cd$ we obtain

$$\Sigma = \sum_{t \le \sqrt{D}} \left(\sum_{cd=t} \frac{\mu(d)}{g(c)} \right) \left(\sum_{\substack{m \le \sqrt{D}, \\ t|m}} g(m)\rho_m \right)^2.$$

Now if t is not squarefree then the sum over m vanishes (because $g(m)$ is supported on squarefree numbers), and if t is squarefree then it is easy to check by expanding the product that

$$\sum_{cd=t} \frac{\mu(d)}{g(c)} = \prod_{p|t} \left(\frac{1}{g(p)} - 1 \right),$$

so the lemma follows. □

In view of Lemma 2.3, we see a good way to choose the numbers ρ_d in any given problem will be such that

$$\Sigma := \sum_{t \le \sqrt{D}} \left(\sum_{\substack{m \le \sqrt{D}, \\ t|m}} g(m)\rho_m \right)^2 \prod_{p|t} \left(\frac{1}{g(p)} - 1 \right)$$

is minimised, subject to our standing constraints that $\rho_1 = 1$ and $\rho_d = 0$ when d has prime factors $\notin \mathcal{P}$ or $d > \sqrt{D}$. We emphasise that it is wrong to claim such a choice is automatically optimal, since it ignores the term $\sum_{d \le D} \lambda_d r(d)$ which might, depending on the sizes of D and $r(d)$, make a significant contribution. But we generally think of $\sum_{d \le D} \lambda_d r(d)$ as a term that is hard to understand, so usually in classical sieve theory we "prioritise" and settle for optimising the first term, and then choosing D to ensure that $\sum_{d \le D} \lambda_d r(d)$ is negligible.

Lemma 2.5 ("Optimal" Λ^2 construction, following Selberg)**.** *Let $\mathcal{P}$ be any set of primes, let $2 \le \sqrt{D}$, let $g(d)$ be a multiplicative function supported on squarefree integers, and*

suppose now that $0 < g(p) < 1$ *for all primes* p*. Define*

$$J = J(\mathcal{P}, g, D) := \sum_{\substack{d \le \sqrt{D}, \\ d \text{ squarefree}, \\ p|d \Rightarrow p \in \mathcal{P}}} \prod_{p|d} \frac{g(p)}{1-g(p)}.$$

Then if we set

$$\rho_d = (\mathbf{1}_{p|d \Rightarrow p \in \mathcal{P}})\mu(d) \left(\prod_{p|d} \frac{1}{1-g(p)} \right) \frac{1}{J} \sum_{\substack{t \le \sqrt{D}/d, \\ t \text{ squarefree}, (t,d)=1, \\ p|t \Rightarrow p \in \mathcal{P}}} \prod_{p|t} \frac{g(p)}{1-g(p)} \quad \forall d \le \sqrt{D},$$

we have

$$\Sigma = \sum_{t \le \sqrt{D}} \left(\sum_{\substack{m \le \sqrt{D}, \\ t|m}} g(m)\rho_m \right)^2 \prod_{p|t} \left(\frac{1}{g(p)} - 1 \right) = \frac{1}{J}.$$

Moreover, this is the smallest that the left hand side can be for any numbers ρ_d *such that* $\rho_1 = 1$*, and* $\rho_d = 0$ *when* d *has prime factors* $\notin \mathcal{P}$ *or* $d > \sqrt{D}$*.*

Proof of Lemma 2.5. Thanks to our work in Lemma 2.3, the sum over $t \le \sqrt{D}$ that we want to minimise is already a fairly nice looking quadratic form in the ρ_m. Moreover, we may assume without loss of generality that the "optimal" ρ_m are supported on squarefree m, since this is true for $g(m)$ and in our objective sum ρ_m is always multiplied by $g(m)$. Similarly, in the objective sum Σ we may restrict $t \le \sqrt{D}$ to squarefree values with all their prime factors from $\mathcal{P}$, because for all other t the sum $\left(\sum_{\substack{m \le \sqrt{D}, \\ t|m}} g(m)\rho_m \right)^2$ vanishes.

To find an "optimal" choice of the ρ_m we want to diagonalise our objective sum/quadratic form. To do this we shall use the following small lemma.

Lemma 2.6. *Suppose* (ρ_d) *is a sequence supported on squarefree* $d \le \sqrt{D}$ *with all their prime factors from* $\mathcal{P}$*. Define*

$$y_t := \mu(t) \left(\sum_{\substack{m \le \sqrt{D}, \\ t|m}} g(m)\rho_m \right) \prod_{p|t} \left(\frac{1}{g(p)} - 1 \right) \quad \forall t \in \mathbb{N}.$$

Then (y_t) *is a sequence supported on squarefree* $t \le \sqrt{D}$ *with all their prime factors from* $\mathcal{P}$*, and we have*

$$\rho_d = \frac{\mu(d)}{g(d)} \sum_{t : d|t} y_t \prod_{p|t} \frac{g(p)}{1-g(p)} \quad \forall d \in \mathbb{N}.$$

The reverse statement is also true (if one starts with a sequence (y_t) and then defines (ρ_d) in terms of it).

Proof of Lemma 2.6. The proof will be an exercise on the first Problem Sheet. (It uses a technique called *Möbius inversion*.) □

In terms of the y_t from Lemma 2.6, our task is to minimise the form

$$\sum_{t\le\sqrt{D}}\left(\sum_{\substack{m\le\sqrt{D},\\ t|m}} g(m)\rho_m\right)^2 \prod_{p|t}\left(\frac{1}{g(p)}-1\right) = \sum_t y_t^2 \prod_{p|t}\left(\frac{1}{g(p)}-1\right)^{-1} = \sum_t y_t^2 \prod_{p|t}\frac{g(p)}{1-g(p)}$$

over all sequences (y_t) that are supported on squarefree $t \le \sqrt{D}$ with all their prime factors from $\mathcal{P}$. The other constraint we must satisfy is that $\rho_1 = 1$, and using Lemma 2.6 we can rewrite that as

$$\sum_t y_t \prod_{p|t}\frac{g(p)}{1-g(p)} = \frac{\mu(1)}{g(1)}\sum_t y_t \prod_{p|t}\frac{g(p)}{1-g(p)} = \rho_1 = 1.$$

But now we can perform the minimisation just by completing the square, noting that

$$\begin{aligned}
&\sum_t (y_t - 1/J)^2 \prod_{p|t}\frac{g(p)}{1-g(p)} \\
= \;& \sum_t y_t^2 \prod_{p|t}\frac{g(p)}{1-g(p)} - (2/J)\sum_t y_t \prod_{p|t}\frac{g(p)}{1-g(p)} + (1/J)^2 \sum_t \prod_{p|t}\frac{g(p)}{1-g(p)} \\
= \;& \sum_t y_t^2 \prod_{p|t}\frac{g(p)}{1-g(p)} - 1/J,
\end{aligned}$$

by definition of J. Thus it is clear that the minimum possible value of our quadratic form is $1/J$, and this is achieved and the constraint $\sum_t y_t \prod_{p|t}\frac{g(p)}{1-g(p)} = 1$ is satisfied when $y_t = 1/J$ for all squarefree $t \le \sqrt{D}$ with all their prime factors from $\mathcal{P}$. One can check using Lemma 2.6 that this choice of y_t corresponds to the choice of ρ_d claimed in Lemma 2.5. □

By combining Lemmas 2.3 and 2.5, and doing a little bit more work to bound the error term, we finally obtain a version of Selberg's powerful upper bound sieve.

Theorem 2.7 (Selberg upper bound sieve)**.** *Let the situation be as in Lemma 2.3, with the restriction that $0 < g(p) < 1$ for all primes p. Then*

$$\sum_{n:p|n\Rightarrow p\notin\mathcal{P}} a_n \le \frac{X}{J} + \sum_{\substack{d\le D,\\ d\ squarefree,\\ p|d\Rightarrow p\in\mathcal{P}}} 3^{\omega(d)}|r(d)|,$$

where $J := \sum_{\substack{d\le\sqrt{D},\\ d\ squarefree,\\ p|d\Rightarrow p\in\mathcal{P}}} \prod_{p|d}\frac{g(p)}{1-g(p)}$ *and where* $\omega(d)$ *denotes the number of distinct prime factors of* d.

Proof of Theorem 2.7. After combining Lemmas 2.3 and 2.5, it only remains to show that $\sum_{d\le D}\lambda_d r(d) \le \sum_{\substack{d\le D,\\ d\,\text{squarefree},\\ p|d\Rightarrow p\in\mathcal{P}}} 3^{\omega(d)}|r(d)|$ when the λ_d are Λ^2 sieve weights and the constituents ρ_d are chosen as in Lemma 2.5. We recall this means that

$$\lambda_d := \sum_{[d_1,d_2]=d}\rho_{d_1}\rho_{d_2} \text{ and } \rho_d := (\mathbf{1}_{p|d\Rightarrow p\in\mathcal{P}})\mu(d)\left(\prod_{p|d}\frac{1}{1-g(p)}\right)\frac{1}{J}\sum_{\substack{t\le\sqrt{D}/d,\\ t\,\text{squarefree},\,(t,d)=1,\\ p|t\Rightarrow p\in\mathcal{P}}}\prod_{p|t}\frac{g(p)}{1-g(p)}.$$

In particular, the ρ_d are supported on squarefree d with all their prime factors from $\mathcal{P}$ (because the Möbius function $\mu(d)$ is supported on squarefree d), which implies the same is true for the λ_d. Therefore we do have $\sum_{d\le D}\lambda_d r(d) = \sum_{\substack{d\le D,\\ d\,\text{squarefree},\\ p|d\Rightarrow p\in\mathcal{P}}}\lambda_d r(d)$, so it will suffice to prove that $|\lambda_d| \le 3^{\omega(d)}$ for all such d.

However, for all squarefree d we can expand the product and find

$$\prod_{p|d}\frac{1}{1-g(p)} = \prod_{p|d}\left(1+\frac{g(p)}{1-g(p)}\right) = \sum_{k|d}\prod_{p|k}\frac{g(p)}{1-g(p)},$$

and therefore we have

$$\begin{aligned}\left(\prod_{p|d}\frac{1}{1-g(p)}\right)\sum_{\substack{t\le\sqrt{D}/d,\\ t\,\text{squarefree},\,(t,d)=1,\\ p|t\Rightarrow p\in\mathcal{P}}}\prod_{p|t}\frac{g(p)}{1-g(p)} &= \sum_{k|d}\prod_{p|k}\frac{g(p)}{1-g(p)}\sum_{\substack{t\le\sqrt{D}/d,\\ t\,\text{squarefree},\,(t,d)=1,\\ p|t\Rightarrow p\in\mathcal{P}}}\prod_{p|t}\frac{g(p)}{1-g(p)}\\ &\le \sum_{k|d}\prod_{p|k}\frac{g(p)}{1-g(p)}\sum_{\substack{t\le\sqrt{D}/k,\\ t\,\text{squarefree},\,(t,d)=1,\\ p|t\Rightarrow p\in\mathcal{P}}}\prod_{p|t}\frac{g(p)}{1-g(p)} = J,\end{aligned}$$

since the right hand side is just the sum of $\prod_{p|t}\frac{g(p)}{1-g(p)}$ over all $t\le\sqrt{D}$ divided up according to the highest common factor $k=(d,t)$.

It follows that $|\rho_d| \le J/J = 1$ for all d, and therefore $|\lambda_d| \le \sum_{[d_1,d_2]=d}1 = \sum_{abc=d}1 = 3^{\omega(d)}$ for all squarefree d. □

3. Selberg's sieve versus Riemann's zeta function

In the previous section we developed a version of Selberg's sieve (Theorem 2.7) that looks neat, and realises the idea of replacing the Möbius function with more general weights. However, it remains to be seen whether this actually leads to better results. First let us revisit the question of primes in intervals.

Corollary 3.1. *For any* $1000 \leq z \leq x$ *we have*

$$\pi(x) \ll \frac{x}{\log x},$$

and more generally

$$\pi(x+z) - \pi(x) \ll \frac{z}{\log z}.$$

Proof of Corollary 3.1. We will only prove the second statement, since the first one is easier (and actually follows from the second).

Similarly as before, if $\mathcal{P}$ is any subset of the primes $\leq x$ and $2 \leq \sqrt{D}$ is any parameter then

$$\pi(x+z) - \pi(x) \leq \#\{x < n \leq x+z : p \mid n \Rightarrow p \notin \mathcal{P}\}.$$

To upper bound the right hand side we apply Theorem 2.7, taking $a_n = \mathbf{1}_{x<n\leq x+z}$, $X = z$, and $g(d) = 1/d$ for all squarefree d. Here the remainders satisfy the very good bound

$$r(d) := \sum_{x<n\leq x+z : d|n} 1 - \frac{z}{d} = O(1),$$

and so Theorem 2.7 yields

$$\pi(x+z) - \pi(x) \leq \frac{z}{J} + O\bigg(\sum_{\substack{d\leq D,\\ d \text{ squarefree},\\ p|d\Rightarrow p\in\mathcal{P}}} 3^{\omega(d)} \bigg), \text{ where } J = \sum_{\substack{d\leq \sqrt{D},\\ d \text{ squarefree},\\ p|d\Rightarrow p\in\mathcal{P}}} \prod_{p|d} \frac{1/p}{1-1/p}.$$

Taking a crude approach to the "big Oh" term, we certainly always have $\omega(d) \leq (\log d)/\log 2$, so

$$\sum_{\substack{d\leq D,\\ d \text{ squarefree},\\ p|d\Rightarrow p\in\mathcal{P}}} 3^{\omega(d)} \leq \sum_{d\leq D} d^{(\log 3)/\log 2} \leq D^{1+(\log 3)/\log 2} \leq D^3,$$

say. If we choose $D = z^{1/4}$ then this is negligible. To maximise J we can choose $\mathcal{P}$ to consist of all primes less than x, and then

$$J = \sum_{\substack{d\leq \sqrt{D},\\ d \text{ squarefree}}} \frac{1}{d} \prod_{p|d} \frac{1}{1-1/p} = \sum_{\substack{d\leq \sqrt{D},\\ d \text{ squarefree}}} \frac{1}{d} \prod_{p|d} \sum_{k=0}^{\infty} \frac{1}{p^k} \geq \sum_{d\leq\sqrt{D}} \frac{1}{d} \gg \log(\sqrt{D}) \gg \log z,$$

which proves the corollary. □

Remark 3.2. We see that we have improved substantially on the Sieve of Eratosthenes–Legendre, since we now recover Chebychev's upper bound for $\pi(x)$ (Fact 1 from Chapter 0) and we have a similar-looking upper bound in all short intervals.

For comparison, the theory of the Riemann zeta function $\zeta(s)$ allows one to show that, for a certain absolute constant $c > 0$,

$$\pi(x) = \int_2^x \frac{dt}{\log t} + O\left(xe^{-c(\log^{3/5} x)/(\log\log x)^{1/5}}\right) \text{ as } x \to \infty.$$

When $1000 \le z \le x$ we then have

$$\begin{aligned}\pi(x+z) - \pi(x) &= \int_x^{x+z} \frac{dt}{\log t} + O\left(xe^{-c(\log^{3/5} x)/(\log\log x)^{1/5}}\right) \\ &= (1+o(1))\frac{z}{\log x} + O\left(xe^{-c(\log^{3/5} x)/(\log\log x)^{1/5}}\right) \text{ as } x \to \infty.\end{aligned}$$

Provided that z is a bit larger than $xe^{-c(\log^{3/5} x)/(\log\log x)^{1/5}}$ the first term will dominate the "big Oh" term, and we will have $\pi(x+z) - \pi(x) \sim z/\log x$. This is much stronger than Corollary 3.1, because we have $\log x$ rather than $\log z$ in the denominator, and because we have an asymptotic rather than just an upper bound.

However, if z is smaller than $xe^{-c(\log^{3/5} x)/(\log\log x)^{1/5}}$ then we only get the bound $\pi(x+z) - \pi(x) = O\left(xe^{-c(\log^{3/5} x)/(\log\log x)^{1/5}}\right)$, which is trivial! Even assuming the Riemann Hypothesis, which implies that

$$\pi(x) = \int_2^x \frac{dt}{\log t} + O\left(\sqrt{x}\log^2 x\right) \text{ as } x \to \infty,$$

we could only handle intervals of length about $\sqrt{x}$ using the zeta function, whereas Corollary 3.1 is non-trivial even for $z = x^{0.01}$, say.

Next we can apply Selberg's sieve to study primes in arithmetic progressions. This is an example where we *won't* choose $\mathcal{P}$ to consist of absolutely all the primes up to some point.

Define the counting function

$$\pi(x; q, a) := \#\{p \le x : p \equiv a \bmod q\},$$

where p denotes primes and a, q are any integers. If the highest common factor (a, q) is not 1 then there is at most one prime congruent to a modulo q, so we restrict attention to the coprime case.

Corollary 3.3. *For any $2 \le q \le x/1000$, and any $(a, q) = 1$, we have*

$$\pi(x; q, a) \ll \frac{x}{\phi(q)\log(x/q)},$$

where $\phi(q) := \#\{1 \le r \le q : (r, q) = 1\}$ denotes Euler's totient function.

Proof of Corollary 3.3. The proof is analogous to the proof of Corollary 3.1.

For any set $\mathcal{P}$ of primes and any parameter $2 \le \sqrt{D}$ we have

$$\pi(x; q, a) \le \pi(\max \mathcal{P}; q, a) + \#\{n \le x : n \equiv a \bmod q,\ \ p \mid n \Rightarrow p \notin \mathcal{P}\}.$$

The first term is trivially $\leq (\max \mathcal{P})/q + 1$, which will be negligible provided $\max \mathcal{P} \leq \sqrt{x}$, say. To upper bound the other term we apply Theorem 2.7, taking $a_n = \mathbf{1}_{\substack{n \leq x, \\ n \equiv a \bmod q}}$, $X = x/q$, and $g(d) = 1/d$ for all squarefree d. *Provided d is invertible modulo q*, the remainders satisfy

$$r(d) := \sum_{\substack{n \leq x,\, d|n, \\ n \equiv a \bmod q}} 1 - \frac{x}{qd} = \sum_{\substack{m \leq x/d, \\ md \equiv a \bmod q}} 1 - \frac{x}{qd} = O(1).$$

If d isn't invertible modulo q then we have a bad bound, so it makes sense to choose $\mathcal{P} = \{p \leq \sqrt{x} : (p, q) = 1\}$ so that we don't have to handle $r(d)$ for such d. Applying Theorem 2.7, and estimating the error term as in the proof of Corollary 3.1, yields

$$\pi(x; q, a) \leq \frac{\sqrt{x}}{q} + 1 + \frac{x}{qJ} + O(D^3), \text{ where } J = \sum_{\substack{d \leq \sqrt{D}, \\ d \text{ squarefree}, \\ p|d \Rightarrow p \in \mathcal{P}}} \prod_{p|d} \frac{1/p}{1 - 1/p}.$$

Finally, if we choose $D = (x/q)^{1/4}$ then the "big Oh" term is negligible, and as in the proof of Corollary 3.1 we have

$$J \geq \sum_{d \leq \sqrt{D}, (d,q)=1} \frac{1}{d} \geq \prod_{p|q} \left(1 - \frac{1}{p}\right) \sum_{d \leq \sqrt{D}} \frac{1}{d} \gg \prod_{p|q} \left(1 - \frac{1}{p}\right) \log(x/q).$$

Since $q \prod_{p|q} \left(1 - \frac{1}{p}\right) = \phi(q)$, this finishes the proof of the corollary. □

Remark 3.4. It seems reasonable to suppose the primes less than x are roughly equidistributed among the $\phi(q)$ coprime residue classes modulo q, for any $q \leq x^{1-\epsilon}$. (One clearly cannot have equidistribution if $q = 1000x$, say, but for q a bit smaller than x there are no clear obstructions.) More precisely, we would expect that for any $q \leq x^{1-\epsilon}$ and any $(a, q) = 1$ we should have

$$\pi(x; q, a) \sim \frac{x}{\phi(q) \log x} \text{ as } x \to \infty.$$

We only know how to prove this when $q \leq \log^A x$ for any fixed $A > 0$, a result called the *Siegel–Walfisz Theorem* whose usual proof requires Dirichlet L-functions $L(s, \chi)$, which generalise the Riemann zeta function. Assuming the Generalised Riemann Hypothesis for all L-functions to modulus q, we could prove the asymptotic when $q \leq \sqrt{x}/\log^{O(1)} x$. Once again, although it is only an upper bound Corollary 3.3 is non-trivial even when $q = x^{0.99}$, say, which is far beyond even the Generalised Riemann Hypothesis range.

Remark 3.5. Note that the totient function $\phi(q)$ that appears in Corollary 3.3 also appears in the asymptotic formula that we guess is the true answer. In the proof of the corollary, $\phi(q)$ emerged because we could only sieve by moduli coprime to q. This

shows that the choices we make (or are forced upon us) inside the sieve process are not always just technicalities, but may reflect real features of the problem.

As a final example, we turn to an ancient question that cannot be attacked using the zeta function at all, namely twin primes. We would like to show there are infinitely many twin primes, but this seems extremely difficult. However, even showing there are significantly *fewer* twin primes than primes is a difficult question, unsolved until Brun started to develop sieve theory. We can recover such results using Selberg's sieve.

Corollary 3.6 (following Brun, 1919)**.** *For all $x \geq 2$ we have*

$$\#\{p \leq x : p, p+2 \text{ are prime}\} \ll \frac{x}{\log^2 x}.$$

As a consequence, the series

$$\sum_{p:p,p+2 \text{ are prime}} \left(\frac{1}{p} + \frac{1}{p+2}\right)$$

is convergent ((to a value ≈ 1.902 that is traditionally called Brun's constant*)).*

Recall there are $\gg x/\log x$ primes less than x (Chebychev's Fact 1), and $\sum_{p\leq x} 1/p \sim \log\log x$ diverges, so Corollary 3.6 indeed asserts that there are significantly fewer twin primes than primes.

Proof of Corollary 3.6. Note that if the first bound claimed in the corollary is true then

$$\sum_{x/2<p\leq x:p,p+2 \text{ are prime}} \left(\frac{1}{p} + \frac{1}{p+2}\right) \leq \frac{2}{x/2}\#\{x/2 < p \leq x : p, p+2 \text{ are prime}\} \ll \frac{1}{\log^2 x}.$$

Applying this to the sequence of values $x = 2^j$ for $j \in \mathbb{N}$, (which is called *dyadic decomposition*), we deduce that

$$\sum_{p:p,p+2 \text{ are prime}} \left(\frac{1}{p} + \frac{1}{p+2}\right) \ll \sum_{j=1}^{\infty} \frac{1}{j^2},$$

which is convergent.

It remains to prove the first bound. As before, if $\mathcal{P}$ is any subset of the primes $\leq \sqrt{x}$ and $2 \leq \sqrt{D}$ is a parameter then

$$\#\{p \leq x : p, p+2 \text{ are prime}\} \leq \pi(\sqrt{x}) + \#\{m \leq x : p \mid m(m+2) \Rightarrow p \notin \mathcal{P}\}.$$

The first term is negligible, and we can estimate the second by applying Theorem 2.7 with $a_n = \mathbf{1}_{n=m(m+2) \text{ for some } m\leq x}$, with $X = x$, and with $g(d)$ the multiplicative function on squarefree integers defined by $g(2) = 1/2$, and $g(p) = 2/p$ for all other primes p. Then our remainders become

$$r(d) = \sum_{m\leq x : d \mid m(m+2)} 1 - xg(d) = \sum_{m\leq x : m\equiv 0 \text{ or } -2 \; \forall p \mid d} 1 - xg(d)$$

for all squarefree d. Here the condition on m is equivalent to saying that it should lie in one of $2^{\omega(d)} = dg(d)$ residue classes modulo d if d is odd, or in one of $2^{\omega(d)-1} = dg(d)$ residue classes modulo d if d is even (because 0 and -2 are the same residue class mod 2). Since the number of $m \leq x$ in any such residue class is $x/d + O(1)$, it follows that

$$r(d) = O(dg(d)) = O(2^{\omega(d)}).$$

Theorem 2.7 yields that

$$\#\{m \leq x : p \mid m(m+2) \Rightarrow p \notin \mathcal{P}\} \leq \frac{x}{J} + O\left(\sum_{d \leq D} 3^{\omega(d)} 2^{\omega(d)}\right) = \frac{x}{J} + O\left(D^4\right),$$

say, where we estimated the "big Oh" term similarly as before. If we choose $D = x^{1/5}$ then this is negligible. Meanwhile, if we choose $\mathcal{P}$ to consist of all primes less than $\sqrt{x}$ then

$$J = \sum_{\substack{d \leq \sqrt{D}, \\ d \text{ squarefree}}} \prod_{p \mid d} \frac{g(p)}{1-g(p)} = \sum_{\substack{d \leq \sqrt{D}, \\ d \text{ squarefree}}} g(d) \prod_{p \mid d} \sum_{k=0}^{\infty} g(p)^k \geq \sum_{d \leq \sqrt{D}, \, d \text{ odd}} \frac{2^{\Omega(d)}}{d},$$

where $\Omega(d)$ denotes the total number of prime factors of d (not just the distinct ones). One can check that $2^{\Omega(d)} \geq \sum_{a \mid d} 1$, and then by inserting this and changing the order of summation deduce that $J \gg \log^2 D \gg \log^2 x$. This finishes the proof. □

4. What can the sieve not do?

We have seen that Selberg's sieve massively improves on the Sieve of Eratosthenes–Legendre, and gives upper bounds that seem close to the truth in ranges where even the Riemann Hypothesis cannot help. Our construction of the Selberg weights started from the assumption that we should take $\lambda_d := \sum_{[d_1,d_2]=d} \rho_{d_1}\rho_{d_2}$ (i.e. that we should use a square to ensure positivity), and it isn't clear that this was the best way to proceed. So it is tempting to think that some even more efficient weights might exist, that would let us prove lower bounds or even asymptotics in our sieve problems.

In this section we will prove that, *in a certain limited sense*, there are no sieve weights that are capable of proving lower bounds when counting primes.

Theorem 4.1 (Parity Phenomenon, Selberg, 1949)**.** *Let x be large. There exist two sequences $\mathcal{A}^+, \mathcal{A}^-$ of non-negative numbers having the following properties: if we choose $X = x/2$ and $g(d) = 1/d$ for all squarefree d, then the remainder numbers $r^+(d) = r(d, \mathcal{A}^+, g, X)$ and $r^-(d) = r(d, \mathcal{A}^-, g, X)$ satisfy*

$$r^+(d), r^-(d) \ll \frac{x}{d \log^{10}(x/d+1)} \quad \forall d \leq x.$$

Moreover we have

$$\sum_{p\,prime} a_p^+ = \pi(x) \sim \frac{x}{\log x} \sim \frac{2X}{\log X}, \quad but \quad \sum_{p\,prime} a_p^- = 0.$$

The point of the parity phenomenon (often also called the *Parity Problem*) is that from the point of view of sieves the sequences $\mathcal{A}^+, \mathcal{A}^-$ look essentially the same, having quite small remainders all the way up to $d \leq x^{1-\epsilon}$. But one of the sequences contains twice as many primes as we might guess, and the other contains no primes at all, so *it is impossible for any procedure that only uses these bounds on remainders to prove the existence of primes, or an upper bound that is better than twice as large as the usual guess.*

Proof of Theorem 4.1. We just choose $\mathcal{A}^+$ to be the indicator function of those $n \leq x$ that have an odd number of prime factors, and $\mathcal{A}^-$ to be the indicator function of those $n \leq x$ that have an even number of prime factors. (This explains why this phenomenon is called the Parity Phenomenon!)

The claims about $\sum_{p\,\text{prime}} a_p^+, \sum_{p\,\text{prime}} a_p^-$ are trivial, so it remains to check the bounds on the remainders. We can write $a_n^\pm = (1/2)(1 \mp \lambda(n))$ for all $n \leq x$, where $\lambda(n) = (-1)^{\Omega(n)}$ and $\Omega(n)$ denotes the total number of prime factors of n. (This function $\lambda(n)$ is called the *Liouville function*, and is closely related to the Möbius function $\mu(n)$.) We then have

$$r^+(d) := (1/2) \sum_{n \leq x : d|n} (1 - \lambda(n)) - Xg(d) = -(1/2) \sum_{n \leq x : d|n} \lambda(n) + O(1),$$

similarly for $r^-(d)$, since $X = (1/2)x$ and $g(d) = 1/d$. Then

$$\sum_{n \leq x : d|n} \lambda(n) = \lambda(d) \sum_{m \leq x/d} \lambda(m),$$

and it is a fact (equivalent to the Prime Number Theorem) that $\sum_{m \leq M} \lambda(m) \ll M/\log^{10} M$. □

Theorem 4.1 implies that we cannot detect primes just using bounds on remainders $r(d)$, however cleverly we might choose sieve weights, but the crucial caveat is that we might be able to detect them if we insert some other information into the arguments. There are now several cases where this has been achieved, for example:

(i) One can use sieve weights to remove most terms from a problem (get an upper bound of the correct order) and organise the remaining terms as a double or triple sum (often called *Type II sums*). Such sums can sometimes be estimated using some kind of Fourier analysis. This was famously achieved in the "asymptotic sieve" of Friedlander and Iwaniec, who found an asymptotic for the number of prime values of the polynomial $X^2 + Y^4$, for example.

(ii) One can use sieve weights to remove most terms, and organise the remaining ones such that they cancel one another in some way. This kind of strategy is used as part of an iterative argument in some elementary proofs of the Prime Number Theorem.

(iii) One can use sieve weights in a situation where one knows in advance there are some primes in the sequence under study, and then compare the sieved sum with the sum over primes to show the primes must be distributed in certain ways. This is the overall strategy in showing the existence of infinitely many bounded gaps between primes.

These examples show that it is important to be aware of the limitations of our methods, but we shouldn't be misled into not trying new things because small variations on the setup can change the limitations completely.

References

[1] J. Friedlander and H. Iwaniec. *Opera de Cribro.* AMS Colloquium Publications, vol. 57. 2010

Jesus College, Cambridge, CB5 8BL
E-mail address: A.J.Harper@dpmms.cam.ac.uk